

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Instituto de Ciências Biológicas

Programa de Pós-Graduação em Genética

Anna Caroline Rossi

**Caracterização, localização cromossômica e evolução de DNAs satélites em
Drosophila gouveai e *D. borborema* (cluster *buzzatii*, grupo *repleta*)**

Belo Horizonte

2025

Anna Caroline Rossi

**Caracterização, localização cromossômica e evolução de DNAs satélites em
Drosophila gouveai e *D. borborema* (cluster *buzzatii*, grupo *repleta*)**

Dissertação apresentada ao programa de PósGraduação em Genética da Universidade Federal de Minas Gerais como pré-requisito obrigatório para obtenção do título de Mestre em Genética, área de concentração Genética de Populações e Evolução.

Orientador: Dr. Gustavo Campos e Silva Kuhn

Belo Horizonte

2025

043

Rossi, Anna Caroline.

Caracterização, localização cromossômica e evolução de DNAs satélites em *Drosophila gouveai* e *D. boborema* (cluster *buzzatii*, grupo *repleta*) [manuscrito] / Anna Caroline Rossi. – 2025.

103 f. : il. ; 29,5 cm.

Orientador: Gustavo Campos e Silva Kuhn.

Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Genética.

1. Genética. 2. DNA Satélite. 3. *Drosophila*. 4. Cromossomos. 5. Heterocromatina. I. Kuhn, Gustavo Campos e Silva. II. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III. Título.

CDU: 502.7



UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA

FOLHA DE APROVAÇÃO

**"CARACTERIZAÇÃO, LOCALIZAÇÃO CROMOSSÔMICA E EVOLUÇÃO DE DNAs SATÉLITES EM
DROSOPHILA GOUVEAI E *D. BORBOREMA* (CLUSTER BUZZATII, GRUPO REPLETA)"**

ANNA CAROLINE ROSSI

DISSERTAÇÃO APROVADA PELA BANCA EXAMINADORA CONSTITUÍDA PELOS PROFESSORES:

Prof. Gustavo Campos e Silva Kuhn - UFMG

Prof. Maria Bernadete Lovato - UFMG

Prof. Fernando Faria Franco - UFSCAR

Belo Horizonte, 27 de março de 2025.



Documento assinado eletronicamente por **Fernando de Faria Franco, Usuário Externo**, em 01/04/2025, às 10:23, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Gustavo Campos e Silva Kuhn, Professor(a)**, em 01/04/2025, às 11:21, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Maria Bernadete Lovato, Professora do Magistério Superior**, em 08/04/2025, às 11:00, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **4077380** e o código CRC **88B7542A**.



UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA

ATA DE DEFESA DE DISSERTAÇÃO DA ALUNA

ANNA CAROLINE ROSSI

Às nove horas do dia **27 de março de 2025**, reuniu-se a Comissão Examinadora de Dissertação, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: "**Caracterização, localização cromossômica e evolução de DNAs satélites em *Drosophila gouveai* e *D. borborema* (cluster *buzzatii*, grupo *repleta*)**", requisito para obtenção do grau de Mestre em **Genética**. Abrindo a sessão, o Presidente da Comissão, **Dr. Gustavo Campos e Silva Kuhn**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa da candidata. Logo após, a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Prof./Pesq.	Instituição	Indicação
Dr. Gustavo Campos e Silva Kuhn	UFMG	APROVADA
Dra. Maria Bernadete Lovato	UFMG	APROVADA
Dr. Fernando Faria Franco	UFSCAR	APROVADA

Pelas indicações, a candidata foi considerada: **APROVADA**.

O resultado final foi comunicado publicamente à candidata pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.

Belo Horizonte, 27 de março de 2025.

Dr. Gustavo Campos e Silva Kuhn

Dra. Maria Bernadete Lovato

Dr. Fernando Faria Franco

Assinatura dos membros da banca examinadora:



Documento assinado eletronicamente por **Fernando de Faria Franco, Usuário Externo**, em 01/04/2025, às 10:23, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Gustavo Campos e Silva Kuhn, Professor(a)**, em 01/04/2025, às 11:20, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Maria Bernadete Lovato, Professora do Magistério Superior**, em 08/04/2025, às 11:00, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **4077379** e o código CRC **5EE7461F**.

*Dedico este trabalho a todas as mulheres
que lutaram para conquistar nossos direitos
para que hoje eu pudesse estar aqui.*

Agradecimentos

Gostaria de agradecer meu orientador, Prof. Dr. Gustavo Campos e Silva Kuhn por todos os ensinamentos, conselhos, oportunidades e confiança durante esses dois anos de mestrado, foi fundamental na minha formação acadêmica e profissional, o senhor fez toda a diferença na minha experiência dentro da pós-graduação, o meu muito obrigada.

Agradeço aos meus pais, Ana Maria e Rogério, por todos os sacrifícios feitos para que eu pudesse chegar até aqui e por acreditarem em mim, por serem a base para onde sempre posso recorrer.

Aos amigos que fiz em BH, obrigada por estarem presentes e tornarem essa caminhada mais leve e divertida. Em especial, gostaria de agradecer a Carol por dividir as angustias e saudades de casa, por ser um momento de conforto e respiro em meio ao caos. Ao Luan, Pedro e Mayane uma amizade essencial nesses dois anos, obrigada por todos os momentos juntos, toda ajuda, trilhas, filmes e desabafos. A Ívian, que mesmo distante se fez e se faz presente, obrigada pelo acolhimento, apoio, amizade e carinho.

Aos amigos antigos, que estiveram ao meu lado dando apoio emocional e por toda a compreensão dos 1000km que nos separaram nesse período, obrigada por todo vinho, magic, jantas, conversas, chamadas e por se fazerem presentes, tenho muita sorte por tê-los, vocês são raros!

Ao grupo do laboratório, A Ana e o Rhuan, obrigada por toda ajuda, paciência, ensinamentos e pelas contribuições com meu trabalho, e claro, pela amizade.

Gostaria também de agradecer as agências de fomento, CAPES, FAPEMIG e CNPQ, responsáveis pelo financiamento da bolsa de pesquisa do meu trabalho e também dos outros projetos do LCev.

Aos professores e funcionários da UFMG e do Programa de Pós-Graduação em Genética. Em especial as meninas da secretaria e aos integrantes da CAE pelo acolhimento recebido.

E a todos aqueles que contribuíram direta e indiretamente para a realização deste trabalho e que fizeram parte desse processo.

Resumo

Os DNAs satélites (satDNAs) são extensas sequências de repetições em tandem, encontradas principalmente nas regiões heterocromáticas dos cromossomos. Eles não codificam proteínas, mas podem desempenhar papéis importantes, como regulação gênica, modulação da cromatina e manutenção de centrômeros funcionais. Este estudo teve como objetivo identificar e caracterizar o satelitoma de duas espécies próximas filogeneticamente do cluster *buzzatii*: *D. borborema* e *D. gouveai*. Utilizamos o pipeline RepeatExplorer2 para identificação de satDNAs *de novo* a partir dos genomas sequenciados por Illumina de *D. gouveai* (Altinópolis, SP) e de duas localidades de *D. borborema* (Morro do Chapéu e Milagres, BA). Identificamos quatro satDNAs compartilhados entre as duas espécies: CDSTR138, CDSTR230, DBC-150 e CDSTR16. Além disso, o satDNA pBuM-2 foi encontrado em *D. gouveai* e poucas cópias foram encontradas em um dos genomas de *D. borborema*. A proporção dos satDNAs de cada espécie foi inferior a 3%, valor considerado baixo para *Drosophila*, mas semelhante ao encontrado em outras espécies do cluster *buzzatii*. A abundância de satDNAs mostrou variação tanto interespecífica quanto intraespecífica. Para determinar a localização cromossômica dos satDNAs, realizamos FISH nos cromossomos metafásicos de *D. gouveai*. Encontramos que CDSTR230 está presente em dois *loci* no cromossomo Y, na heterocromatina distal do cromossomo X e nos microcromossomos. O CDSTR16 está localizado nos centrômeros de todos os autossomos, exceto nos microcromossomos, assim como o satDNA CDSTR138. A localização dos satDNAs pBuM-2 e DBC-150 já era conhecida através de estudos anteriores. As filogenias dos satDNAs encontrados em *D. borborema* e *D. gouveai* mostrou o padrão típico de evolução combinada. No entanto, evolução combinada não foi observada nas comparações entre os genomas das duas localidades de *D. borborema*. Filogenias construídas com os satDNAs compartilhados entre *D. gouveai*, *D. borborema*, *D. serido* e *D. seriema* mostraram topologias diferentes, impedindo um maior esclarecimento das relações filogenéticas entre estas espécies. Os resultados obtidos neste trabalho adicionam informações importantes sobre os satDNAs em *D. borborema* e *D. gouveai* e permitem uma melhor compreensão da organização, contribuição para o centrômero e evolução dos satDNAs nas espécies do cluster *buzzatii*.

Palavras chave: DNA satélite; elementos transponíveis; *Drosophila*; heterocromatina; citogenômica

Abstract

Satellite DNAs (satDNAs) are extensive arrays of tandem repeats, primarily found in the heterochromatic regions of chromosomes. They do not encode proteins but may play important roles, such as gene regulation, chromatin modulation, and maintenance of functional centromeres. This study aimed to identify and characterize the satellitome of two closely related species from the *buzzatii* cluster: *D. borborema* and *D. gouveai*. We used the RepeatExplorer2 pipeline for a *de novo* satDNA identification from the Illumina-sequenced genomes of *D. gouveai* (Altinópolis, SP) and from two localities of *D. borborema* (Morro do Chapéu and Milagres, BA). We identified four satDNAs shared between the two species: CDSTR138, CDSTR230, DBC-150, and CDSTR16. Additionally, the satDNA pBuM-2 was found in *D. gouveai* and in few copies in one of the genomes of *D. borborema*. The satellite proportion of the genomic DNA for each species was less than 3%, a value considered low for *Drosophila*, but similar to those found in other species of the *buzzatii* cluster. The abundance of satDNAs showed both interspecific and intraspecific variation. To determine the chromosomal localization of satDNAs, we performed FISH on metaphase chromosomes of *D. gouveai*. We found that CDSTR230 is present at two loci on the Y chromosome, in the distal heterochromatin of the X chromosome, and in the microchromosomes. CDSTR16 is located in the centromeres of all autosomes except the microchromosomes, as well as satDNA CDSTR138. The localization of the pBuM-2 and DBC-150 was already known from previous studies. Phylogenetic trees revealed a concerted evolution pattern for the satDNAs present in *D. borborema* and *D. gouveai*. However, concerted evolution was not observed between satDNA copies from the two localities of *D. borborema*. Phylogenetic trees with satDNA copies of *D. gouveai*, *D. borborema*, *D. serido* and *D. seriema* revealed different topologies, preventing a better understanding of the phylogenetic relationships among these species. The results obtained from this work add important information about the satDNAs of *D. borborema* and *D. gouveai* and provide a better understanding of the organization, centromeric contribution, and evolution of satDNAs in species of the *buzzatii* cluster.

Keywords: satellite DNA; transposable elements; *Drosophila*; heterochromatin; cytogenomics.

Lista de Figuras

Figura 1. Esquema representativo da localização cromossômica dos DNAs satélites em cromossomos

Figura 2. Esquema representativo do processo de evolução combinada em um arranjo de repetições em tandem

Figura 3. Esquema representativo da organização dos satDNAs

Figura 4. Arvore filogenética de espécies de *Drosophila* pertencentes ao cluster *buzzatii*, contendo a distribuição das famílias de satDNAs encontradas em cada espécie.

Figura 5. Distribuição geográfica das espécies do cluster *buzzatii* (extraído de Kuhn, 2003)

Figura 6. Representação esquemática da família de satDNA pBuM-1 contendo a porção *alpha* e pBuM-2 contendo as porções *alpha/beta*

Figura 7. Representação das sequências repetitivas encontradas em *D. borborema* (A)

Figura 8. Representação das sequências repetitivas encontradas em *D. borborema* (B)

Figura 9. Representação das sequências repetitivas encontradas em *D. gouveai*

Figura 10. Sequencia das cópias do satDNA pBuM-2 encontrados em *D. borborema*, em comparação com a sequência consenso de *D. antonietae*. Nos retângulos cor-de-rosa está destacada a porção *alpha*, e nos azuis a porção *beta*.

Figura 11. Esquema representativo de pBuM-2 encontrado no genoma de *D. borborema* (A)

Figura 12. Representação das variações das sequências dos monômeros (8pb) encontrados em diferentes contigs de *D. borborema*

Figura 13. Proporção (%) dos DNAs satélites no DNA genômico de *D. borborema* (linhagens A e B) e *D. gouveai*.

Figura 14. Idiograma representativo do cariótipo básico encontrado em espécies do cluster *buzzatii*

Figura 15. Hibridização *in situ* fluorescente (FISH) de pBuM2 (vermelho) e CDSTR16 (verde) em cromossomos metafásicos de *D. gouveai*. **a.** cromossomos metafásicos contracolorados com DAPI **b.** sinal fluorescente da hibridização das sondas **c.** cromossomos contracolorados com DAPI junto da hibridização

Figura 16. FISH de CDSTR230(vermelho) e CDSTR16 (verde) em cromossomos metafásicos de *D. gouveai*. **a. d. g.** cromossomos metafásicos contracolorados com DAPI **b. e. h.** sinal fluorescente da hibridização das sondas **c. f. i.** cromossomos contracolorados com DAPI junto da hibridização. **g. h. i.** sinal no cromossomo Y.

Figura 17. FISH de CDSTR230 (vermelho) e SAT109/104 (verde) em cromossomos metafásicos de *D. gouveai*. **a.** cromossomos metafásicos contracolorados com DAPI **b.** sinal fluorescente da hibridização das sondas **c.** cromossomos contracolorados com DAPI junto da hibridização

Figura 18. FISH de pBuM2 (vermelho) e CDSTR138 (verde) em cromossomos metafásicos de *D. gouveai*. **a.** cromossomos metafásicos contracolorados com DAPI **b.** sinal fluorescente da hibridização das sondas **c.** cromossomos contracolorados com DAPI junto da hibridização

Figura 19. FISH de CDSTR138(vermelho) e CDSTR16 (verde) em cromossomos metafásicos de *D. gouveai*. **a-f** cromossomos metafásicos femininos **g-i.** cromossomos metafásicos masculino, nenhum sinal observado no cromossomo Y.

Figura 20. Idiograma com a localização cromossômica dos satDNAs em *D. gouveai*

Figura 21. – Árvore de máxima verossimilhança (ML) contendo cópias de pBuM-2 de *D. borborema* A(vermelho) e *D. gouveai* (azul). A árvore foi inferida pelo modelo Tamura de 3 parâmetros com 1000 réplicas de bootstrap. O suporte mínimo de bootstrap é de 50%.

Figura 22 – Árvore de máxima verossimilhança (ML) contendo cópias de DBC-112 de *D. borborema* A (vermelho), *D. borborema* B (rosa) e *D. gouveai* (azul). A árvore foi inferida pelo modelo Jukes-Cantor com 1000 réplicas de bootstrap. O suporte mínimo de bootstrap é de 50%.

Figura 23 - Árvore de máxima verossimilhança (ML) contendo cópias de DBC-150 de *D. borborema* A (vermelho), *D. borborema* B (rosa) e *D. gouveai* (azul). A árvore foi inferida pelo modelo Hasegawa-Kishino-Yano, utilizando uma distribuição Gamma discreta (+G) com 5 categorias, e 1000 réplicas de bootstrap. O suporte mínimo de bootstrap é de 50%.

Figura 24 - Árvore de máxima verossimilhança (ML) contendo cópias de CDSTR138 de *D. borborema* A (vermelho), *D. borborema* B (rosa), e *D. gouveai* (azul). A árvore foi inferida pelo modelo Tamura de 3 parâmetros com 1000 réplicas de bootstrap. O suporte mínimo de bootstrap é de 50%.

Figura 25 - Árvore de máxima verossimilhança (ML) contendo cópias de CDSTR230 de *D. borborema* A (vermelho), *D. borborema* B (rosa) e *D. gouveai* (azul). A árvore foi inferida pelo modelo Hasegawa-Kishino-Yano, utilizando uma distribuição Gamma discreta (+G) com 5 categorias, e 1000 réplicas de bootstrap. O suporte mínimo de bootstrap é de 50%.

Figura 26 - Árvore de máxima verossimilhança (ML) contendo cópias de pBuM-2 de *D. borborema* (vermelho), *D. gouveai* (azul), *D. serido* (roxo), e *D. seriema* (verde). A árvore foi inferida pelo modelo Tamura de 3 parâmetros, utilizando uma distribuição Gamma discreta (+G) com 5 categorias, com 1000 réplicas de bootstrap. O suporte mínimo de bootstrap é de 50%.

Figura 27 - Árvore de máxima verossimilhança (ML) contendo cópias de CDSTR138 de *D. borborema* (vermelho), *D. gouveai* (azul), *D. serido* (roxo), e *D. seriema* (verde). A árvore foi inferida pelo modelo Tamura de 3 parâmetros, com 1000 réplicas de bootstrap. O suporte mínimo de bootstrap é de 50%.

Figura 28 - Árvore de máxima verossimilhança (ML) contendo cópias de CDSTR230 de *D. borborema* (vermelho), *D. gouveai* (azul), *D. serido* (roxo), e *D. seriema* (verde). A árvore foi inferida pelo modelo Tamura de 3 parâmetros, utilizando uma distribuição Gamma discreta (+G) com 5 categorias, com 1000 réplicas de bootstrap. O suporte mínimo de bootstrap é de 50%.

Figura 29 - Árvore de máxima verossimilhança (ML) contendo cópias de DBC-150 de *D. borborema* (vermelho), *D. gouveai* (azul), *D. serido* (roxo), e *D. seriema* (verde). A árvore foi inferida pelo modelo Kimura de 2 parâmetros, utilizando uma distribuição Gamma

discreta (+G) com 5 categorias, com 1000 réplicas de bootstrap. O suporte mínimo de bootstrap é de 50%.

Figura 30 – Árvore de máxima verossimilhança (ML) contendo cópias de DBC-112 de *D. borborema* (vermelho), *D. gouveai* (azul), *D. serido* (roxo), e *D. seriema* (verde). A árvore foi inferida pelo modelo Kimura de 2 parâmetros, utilizando uma distribuição Gamma discreta (+G) com 5 categorias, com 1000 réplicas de bootstrap. O suporte mínimo de bootstrap é de 50%.

Lista de tabelas

Tabela 1. Características gerais dos DNAs satéltes encontrados em *D. borborema* e *D. gouveai*

Lista de abreviaturas

BLAST – “Basic Local Alignment Search Tool” (Ferramenta de busca e alinhamento básico local)

DNA – ácido desoxirribonucleico (ADN)

FISH – Fluorescence in situ hybridization (Hibridização *in situ* fluorescente)

Gb - gigabases

HOR – “Higher-order repeats” (repetições de ordem superior)

Kbs – kilobases

MEGA - Molecular Evolutionary Genetics Analysis (software)

NCBI – National Center for Biotechnology Information (instituição)

NGS – “Next-generation sequencing” (sequenciamento de nova geração)

pb/bp – pares de base/“base pairs”

PCR – reação em cadeia da polimerase

Reads – fragmentos gerados após sequenciamento de DNA

satDNAs – DNAs satélites

TEs – Elementos transponíveis

R – Linguagem de programação

SUMÁRIO

1 INTRODUÇÃO	18
1.1 DNAs satélites.....	18
1.2 O gênero <i>Drosophila</i>	22
1.3 O grupo <i>repleta</i> e as espécies do cluster <i>buzzatii</i>	23
1.4 Revisão dos estudos de DNAs satélites em <i>D. borborema</i> e <i>D. gouveai</i>	27
1.5 Identificação de DNAs satélites	29
2 OBJETIVOS	31
2.1 OBJETIVO GERAL	31
2.2 OBJETIVOS ESPECÍFICOS	31
3 METODOLOGIA	31
3.1 GENOMAS SEQUENCIADOS	31
3.2 IDENTIFICAÇÃO DE DNAs SATÉLITES.....	32
3.3 FILOGENIA	33
3.4 FISH (Hibridização <i>in situ</i> por fluorescência)	33
4 RESULTADOS E DISCUSSÃO	34
4.1 DNAs REPETITIVOS EM <i>D. borborema</i> e <i>D. gouveai</i>	34
4.2 SATELITOMA DE <i>D. borborema</i> e <i>D. gouveai</i>	38
4.3 LOCALIZAÇÃO CROMOSSÔMICA DOS satDNAs	44
4.4 EVOLUÇÃO DOS satDNAs EM <i>D. borborema</i> e <i>D. gouveai</i>	49
4.5 DNAs SATÉLITES COMO MARCADORES FILOGENÉTICOS	55
5 CONCLUSÃO	58
REFERENCIAS	60
ANEXOS.....	71

1 INTRODUÇÃO

1.1 DNAs satélites

O genoma dos eucariotos tipicamente é composto por mais de 50% de DNAs repetitivos não codificantes (Garrido-Ramos 2017; Zhang e cols., 2019), sendo este muito variável entre espécies, tanto em organização, quanto em abundância (Elder & Turner 1995; Lower e cols., 2018). De acordo com sua organização no genoma, podem ser classificados em duas classes principais: as repetições que estão organizadas em tandem, como os microsátélites, minissátélites, DNAs satélites (satDNAs) e repetições teloméricas, e repetições dispersas, como elementos transponíveis (TEs) (Garrido-Ramos 2017). Atualmente, satDNAs são definidos como longos arranjos de milhares de cópias organizadas em tandem, com cópias que variam de 2 pb até vários kb, podendo representar mais da metade do DNA genômico (Tautz 1993; Thakur e cols., 2021; Satovic-Vuksic & Plohl 2023).

Os DNAs satélites foram descobertos por Kit (1961) durante experimentos bioquímicos com ultracentrifugação de gradiente de densidade de césio, utilizando preparações de DNA genômico. Esses experimentos revelaram duas bandas de densidades distintas no DNA do camundongo *Mus musculus*: a banda principal que corresponde ao DNA genômico típico, e uma segunda banda menor denominada "banda satélite". Kit escolheu o termo "satélite" provavelmente inspirado pela então recente "Era Espacial" e os primeiros satélites artificiais em órbita (Belyayev e cols., 2020; Altemose 2022).

Os satDNAs estão localizados principalmente na heterocromatina, em áreas de recombinação reduzida, como os centrômeros, regiões sub-teloméricas, e ao longo do cromossomo Y (Charlesworth e cols., 1994; Kuhn & Sene 2005) (Figura 1). Em alguns casos, é possível encontrar alguns blocos de satDNAs também na eucromatina, mas contendo um número reduzido de repetições (<20) (Kuhn e cols., 2012).

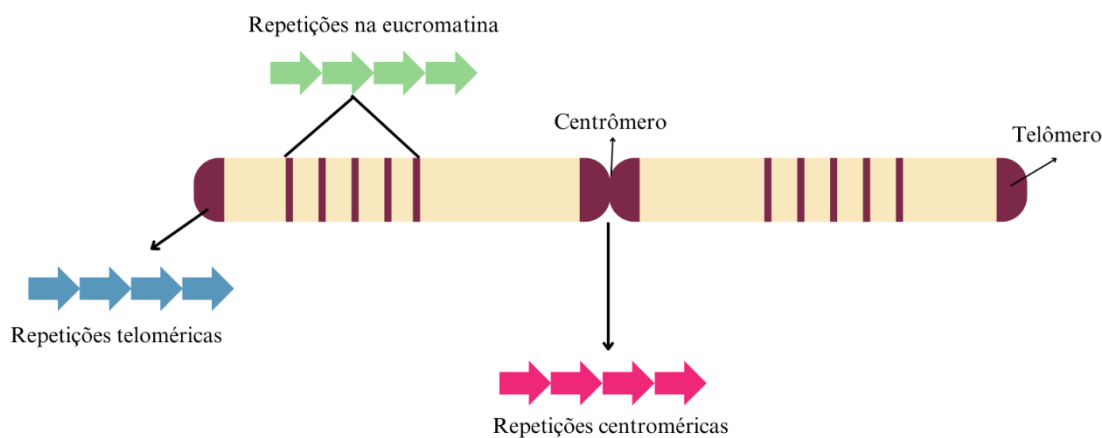


Figura 1. Esquema representativo da localização cromossômica dos DNAs satélites em cromossomos (adaptado de Kuhn e col., 2021)

Dentro de uma mesma espécie, é possível identificar diferentes famílias de DNAs satélites, como por exemplo, no genoma humano, que apresenta 9 famílias de satDNAs (Levy e cols., 2007; Miga, 2015). Normalmente, um satDNA específico é encontrado apenas em espécies que possuem proximidade filogenética. No entanto, em alguns casos, os satélites podem ser exclusivos de uma única espécie (Garrido-Ramos, 2017). Além disso, há uma extrema diversidade de famílias de DNA satélite, tanto em genomas animais quanto vegetais (Satović-Vukšić & Plohl, 2023). A variação também pode ser observada entre indivíduos de uma mesma espécie, como ocorre com o DNA satélite HSat3 em humanos, onde já foi relatada uma diferença de até 98 Mb no número de cópias do cromossomo Y entre diferentes pessoas. (Altemose e cols. 2014).

Os satDNAs apresentam uma rápida taxa de evolução, impulsionada por mecanismos moleculares como *crossing-over* desigual, conversão gênica, transposição e amplificação por círculo rolante (Plohl e cols., 2012; Camacho e cols., 2022). Esses mecanismos provocam uma homogeneização das repetições dentro da espécie, mas também provocam uma divergência entre espécies, uma vez que diferentes variantes podem se tornar fixas após a especiação. Esse processo é

conhecido como evolução combinada (Zimmer e cols., 1980; Elder & Turner 1995; Kuhn e cols., 2012; Belyayev e cols., 2019) (Figura 2).

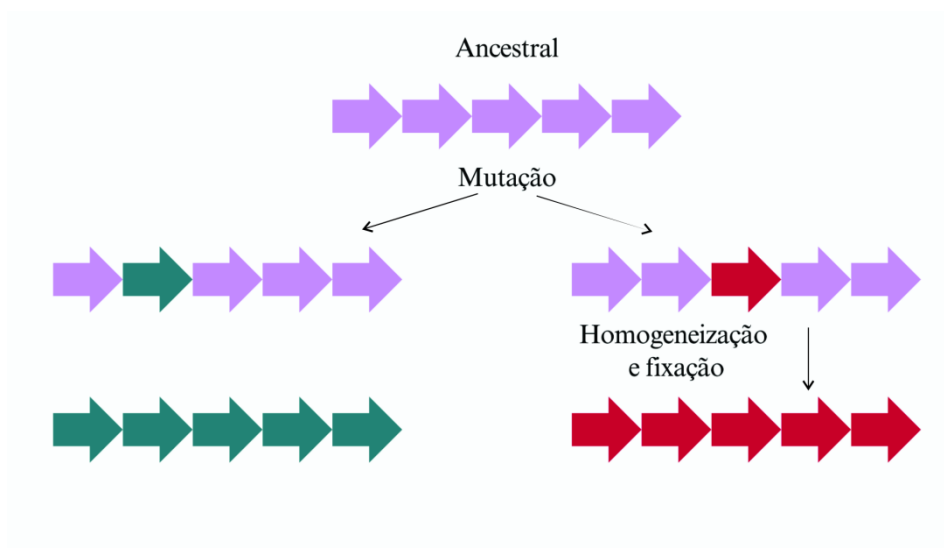


Figura 2. Esquema representativo do processo de evolução combinada em um arranjo de repetições em tandem (Adaptado de Plohl e col., 2012).

Ferree e Barbash (2009) demonstraram que a rápida taxa de evolução dos satDNAs pode provocar um desequilíbrio na segregação cromossômica dos satDNAs em híbridos entre *Drosophila melanogaster* e *D. simulans*, contribuindo para o estabelecimento de mecanismos de isolamento pós-zigóticos e, conseqüentemente, para o processo de especiação (Mestrovic e cols., 1998; Henikoff e cols., 2001; Satovic-Vuksic & Plohl 2023).

Embora os satDNAs não codifiquem proteínas, eles podem participar de funções celulares importantes, como na regulação e modulação da cromatina, formação e manutenção dos centrômeros, regulação gênica, organização espacial dos cromossomos no núcleo, e no emparelhamento e segregação cromossômica (John & Miklos 1979; Dawe & Henikoff 2006; Plohl 2012; Dias e cols., 2015).

Alguns satDNAs mostram uma organização complexa de repetições, resultando em repetições de ordem superior (HORs) (Figura 3) (Willard & Wayne, 1987). Como exemplos de HOR, um satDNA de 532 pb encontrado em um besouro das cavernas *Pholeuon proserpinae* é composto de dois monômeros de 266 pb (Pons e cols., 2003); o satDNA DBC-150 encontrado em espécies de *Drosophila* do cluster *buzzatii*, possui tamanho de 150pb e pode ser organizado em HOR, em arranjos de 3monomeros (3mer), como em *D. serido* (Kuhn e cols., 2009). O modelo de HOR mais estudado são os arranjos do satélite *alpha*, encontrado nos centrômeros dos cromossomos humanos (Lower e cols. 2018).

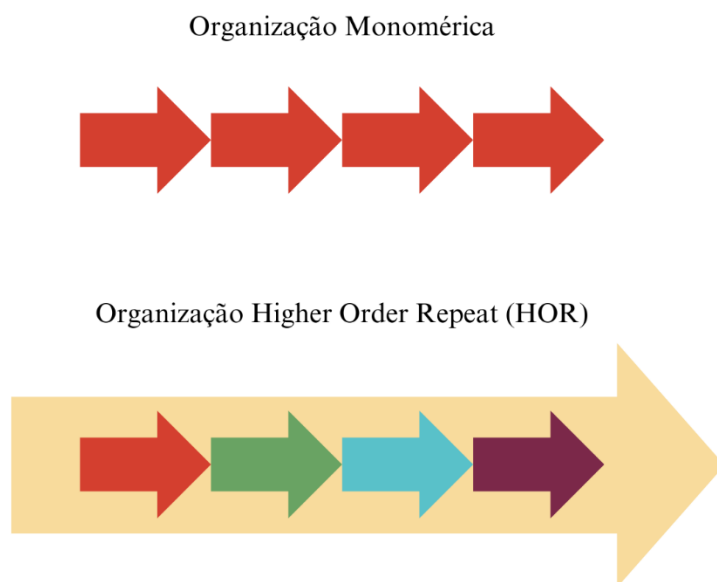


Figura 3. Esquema representativo da organização dos satDNAs

O tamanho das repetições varia consideravelmente entre satDNAs. Existem relatos de satDNA pequenos, como por exemplo, o satDNA de 7pb descrito em *Drosophila virilis*, ou o de 5pb e 12pb em *D. melanogaster* (Gall & Atherton, 1974; Lohe & Brutlag, 1986). Também existem satDNAs com repetições mais extensas, como os encontrados no besouro *Misolampus goudoti*, com cópias de 1.061pb (Pons e cols., 2004), de 1kb ou superior em cetáceos (Grétarsdóttir & Árnason 1992), e o SatIII identificado em *Capreolus capreolus* e em *Hydropotes inermis*, o

qual possui uma unidade de repetição de 2,2kb (Buntjer e cols., 1998; Lin e cols. 2006).

Os DNAs satélites desempenham um papel importante nos centrômeros da maioria dos animais e plantas, mas apresentam uma alta divergência entre espécies, podendo ser espécie-específicos (Hartley & O'Neill 2019). Melters e colaboradores (2013) sugeriram que o satDNA mais abundante de um genoma deve corresponder ao DNA centromérico. O que de fato determina um centrômero funcional é a presença de nucleossomos contendo a histona H3 variante CenH3, que atua como uma marca epigenética, tornando a cromatina desse DNA distinta de outras partes do genoma (Brown & O'Neill 2014; Dalal e cols., 2007).

Prytkova e colaboradores (2011) demonstraram que uma periodicidade de 10pb nos nucleotídeos AA contribui para a formação de nucleossomos, auxiliando na estabilização da histona específica do centrômero, cenH3. Já as sequências dinucleotídica WW (onde W pode ser A ou T), com o mesmo intervalo de 10pb, facilitam o processo de enovelamento dos nucleossomos (STRUHL & SEGAL, 2013). De fato, estudos com outras espécies mostraram que satDNAs centroméricos apresentam uma periodicidade de ~10pb de nucleotídeos A+T (Shang e cols., 2010; Gent e cols., 2011; Bracewell e cols., 2019).

1.2 O gênero *Drosophila*

Espécies de *Drosophila* são comumente conhecidas como “moscas-da-fruta”. Pertencem a ordem *Diptera* e família *Drosophilidae* (Powell, 1997). Constituem um dos principais organismos modelos de estudo em genética, desde o trabalho de Thomas Morgan e colaboradores (1910). O gênero *Drosophila* abrange mais de 1.600 espécies, que apresentam variações morfológicas, diferença em seu tamanho genômico, diversidade de habitats, divergência evolutiva, entre outros (Powell,

1997; Markow & O'Grady, 2005; Bosco, e cols. 2007; Singh, e col. 2009; Jennings, 2011).

As relações filogenéticas da família *Drosophilidae* ainda são debatidas, pois as primeiras árvores basearam-se em características morfológicas. Atualmente o gênero é considerado parafilético (Gao e cols, 2011; O'Grady & DeSalle, 2018; Finet, e cols. 2021)

O gênero *Drosophila* possui 15 subgêneros, onde é possível destacar o subgênero *Drosophila* que compõe mais de 80% das espécies do gênero (O'Grady & DeSalle, 2018), e *Sophophora*, o qual abriga a espécie mais conhecida do gênero, a *Drosophila melanogaster*. O tempo de divergência entre os dois subgêneros foi estimado entre 40 e 60 milhões de anos de acordo com dados fósseis (Powell e DeSalle 1995; Tamura e cols., 2004).

Pesquisas com drosófilas abrangem diversos campos de estudos (Cobb 2007; Bellen e cols., 2010; Ugur e cols. 2016; Trinder e cols., 2017). Nos anos 2000, *D. melanogaster* teve seu genoma sequenciado por Adams e colaboradores. Atualmente mais de 360 espécies de drosofilídeos possuem seus genomas sequenciado (Kim, e cols., 2023) permitindo estudos comparativos sobre evolução genômica, variação genética e adaptação.

1.3 O grupo *repleta* e as espécies do cluster *buzzatii*

O grupo *repleta* possui uma das maiores radiações de espécies do gênero *Drosophila*, sendo suas espécies consideradas um ótimo sistema modelo para estudos evolutivos e ecológicos (Patterson e Stone, 1952; Vilela, 1983; Markow e O'Grady, 2006). A maioria das espécies é cactofílica, ou seja, utiliza de tecidos de cactos em decomposição para o seu desenvolvimento (Ruiz & Heed, 1988; Markow & O'Grady, 2005).

O número exato de espécies pertencentes ao grupo *repleta* é desconhecido, pois alguns táxons foram propostos como similares. Além disso, o grupo possui um número elevado de espécies crípticas, onde o limite entre elas, não é bem esclarecido (Oliveira e cols., 2005 e 2008).

O grupo *repleta* é subdividido em seis subgrupos: *mulleri*, *hydei*, *mercatorum*, *repleta*, *fasciola* e *inca*, os quais são subdivididos em *clusters* e *subclusters*, tendo como referência a morfologia da genitália masculina, associações ecológicas e padrões de bandas cromossômicas (Patterson, 1943; Wharton, 1944; Wasserman, 1962; Rafael & Arcos, 1989).

O *cluster buzzatii* (grupo *repleta*) é um grupo monofilético formado por sete espécies: *D. borborema*, *D. serido*, *D. seriema*, *D. gouveai*, *D. buzzatii*, *D. antonietae* e *D. koepferae* (Manfrin & Sene, 2006). As espécies desse grupo possuem uma divergência recente, ocorrida entre menos de 2 milhões de anos a 4 milhões de anos atrás (Figura 4), o que permite estudar a divergência dos satélites em períodos curtos e analisar as mudanças evolutivas em seus estágios iniciais.

As hipóteses filogenéticas para o grupo de espécies *buzzatii* foram elaboradas com base em diferentes abordagens como a análise de sequências nucleotídicas do gene mitocondrial COI (Manfrin e col., 2001; Franco e Manfrin, 2013), inversões cromossômicas fixas e sequências do gene *period* ligado ao cromossomo X (Ruiz e cols., 2000; Oliveira e cols., 2011, 2013), bem como em dados sobre a morfologia do edeago (Tidon-Sklorz e Sene, 2001), arranjos de DNA satélite pBuM (Kuhn e Sene, 2005) e sequências nucleotídicas de genes como *Xdh* (Rodríguez-Trelles e cols., 2000), além de análises combinadas de genes nucleares e mitocondriais (Oliveira e col., 2012). Embora a filogenia entre as espécies do cluster tenha sido investigada por diversos marcadores, as relações filogenéticas entre as espécies *D. borborema*, *D. gouveai*, *D. seriema* e *D. serido* ainda não são compreendidas (Figura 4)

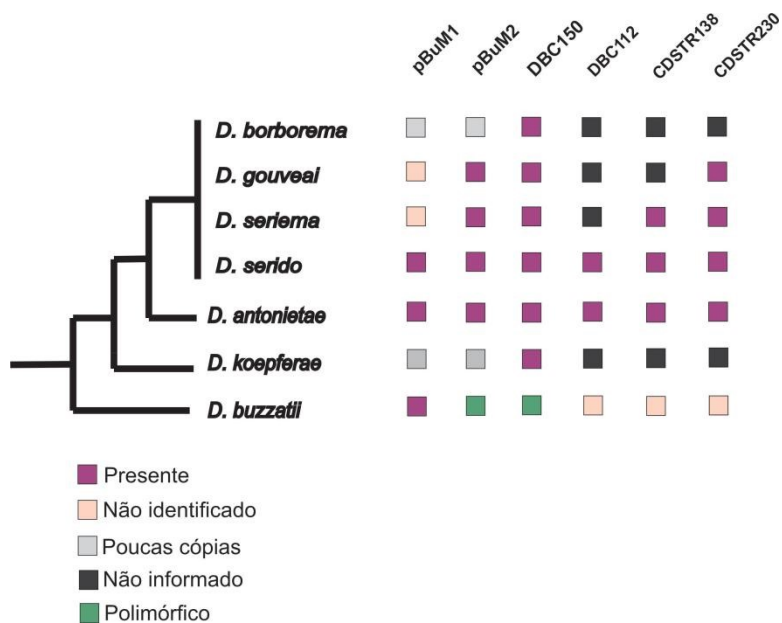


Figura 4. Arvore filogenética de espécies de *Drosophila* pertencentes ao cluster *buzzatii* (adaptada de Kim, 2024), contendo a distribuição das famílias de satDNAs encontradas em cada espécie.

Com exceção de *D. buzzatii*, que é uma espécie semi-cosmopolita, as demais espécies do cluster *buzzatii* são endêmicas da América do Sul (Hurtado *e cols.*, 2019), onde habitam áreas de regiões semiáridas e desérticas do continente (Figura 5) (Manfrin & Sene 2006). A distribuição das espécies está relacionada a distribuição de seus cactos hospedeiros, porque as larvas usam, exclusivamente, tecidos de cactos em decomposição para alimentação e desenvolvimento (Pereira *e cols.*, 1983).

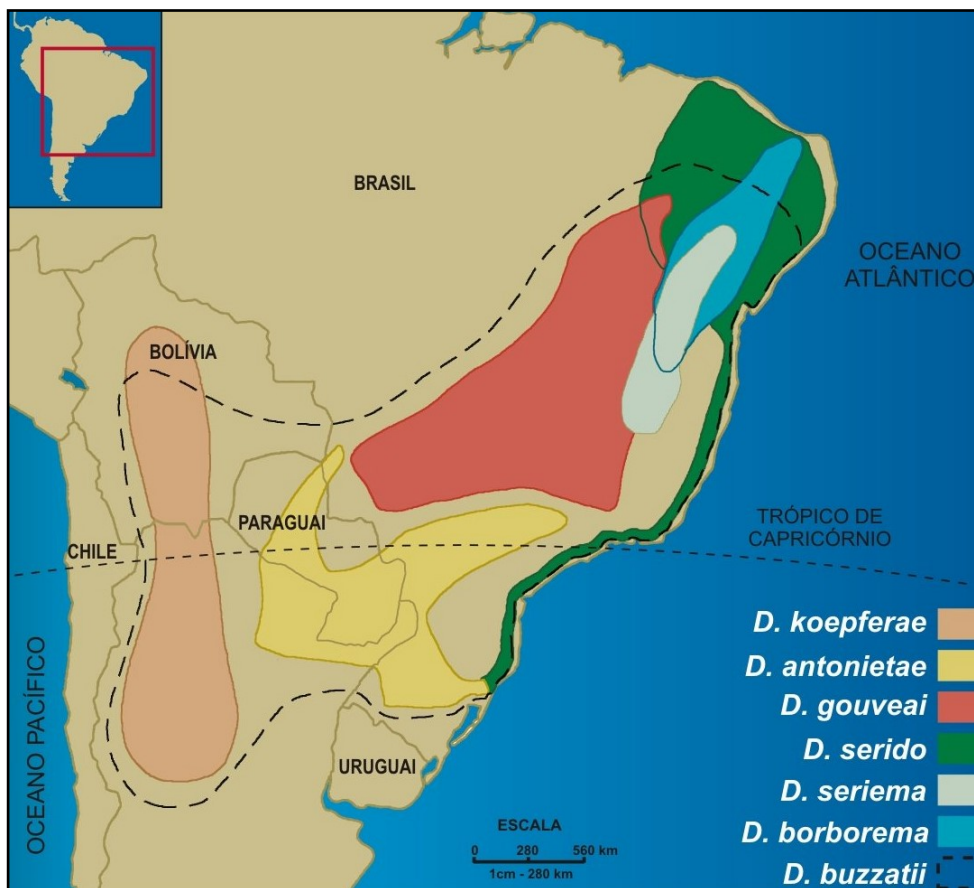


Figura 5. Distribuição geográfica das espécies do cluster *buzzatii* (extraído de Kuhn, 2003)

As espécies do cluster *buzzatii* possuem um cariótipo básico similar, constituído por seis pares de cromossomos: quatro autossomos telocêntricos de mesmo tamanho, um par de microcromossomos, e um par de cromossomos sexuais (Baimai e cols., 1983; Kuhn e cols., 1996).

O genoma das espécies do cluster possui um tamanho menor de ~150Mb enquanto outras espécies do gênero possuem um genoma de ~180Mb (Bosco e cols., 2007; Gregory & Johnston 2008). Além disso o grupo *repleta* apresenta uma proporção ocupada pelos satDNAs no DNA genômico total das espécies de menos de 5%, enquanto em outras espécies de *Drosophila* essa proporção está acima de 20% (Bosco e cols., 2007; Craddock e cols., 2016; Garrido-Ramos 2017). Por conta do tamanho de genoma menor, surge a hipótese que essa baixa contribuição dos

satDNAs seja por conta de uma pressão seletiva (Lima e cols., 2017; Kuhn e cols., 2021).

1.4 Revisão dos estudos de DNAs satélites em *D. borborema* e *D. gouveai*

Os primeiros satDNAs estudados no cluster *buzzatii* foram isolados experimentalmente a partir de digestões do DNA genômico com enzimas de restrição (revisado por Kuhn e cols., 2021). O primeiro satDNA identificado foi o pBuM (Kuhn e cols., 1999). Este satélite possui duas variantes principais: pBuM-1 que consiste em sequências *alpha* de 190 pb, e o pBuM-2, com sequencias *alpha/beta* de aproximadamente 370pb, que consistem em uma sequência *alpha* de 190 pb seguida de uma sequência não-homóloga *beta* de ~180pb. O pBuM foi localizado inicialmente nas regiões centroméricas dos cromossomos autossomos e no cromossomo X, além de possuir uma sequência rica em A+T (Figura 6) (Kuhn e cols., 2008).

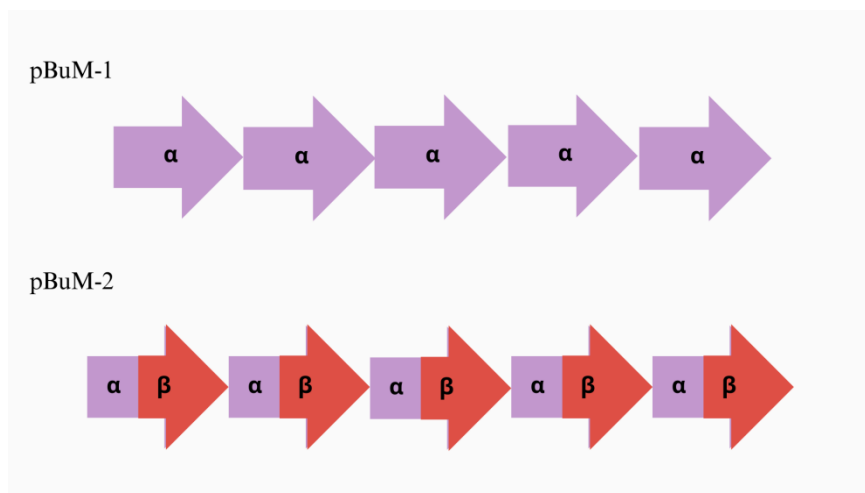


Figura 6. Representação esquemática da família de satDNA pBuM-1 contendo a porção *alpha* e pBuM-2 contendo as porções *alpha/beta*

Já o segundo satDNA identificado nas espécies do cluster, DBC-150, possui repetições de 150pb, está localizado nos microcromossomos e possui um conteúdo rico em C+G (Kuhn e cols., 2007).

Lima e colaboradores (2017) descreveram um novo satélite denominado de CDSTR138, que apresenta um tamanho de 138pb. Este satDNA foi localizado nos centrômeros dos autossomos de *D. seriema*. Este satDNA foi posteriormente encontrado em *D. gouveai* e localizado nos centrômeros dos quatro pares dos autossomos, no X e nos microcromossomos (Nascimento, 2023).

Estudos anteriores utilizando a técnica de enzimas de restrição e hibridização *in situ* em membrana, mostraram ausência do satDNA pBuM-1 mas presença do satélite pBuM-2 em *D. gouveai* (Kuhn & Sene, 2005). Experimentos de hibridização *in situ* fluorescente (FISH) mostraram que, nesta espécie, o satDNA pBuM-2 se localiza nos centrômeros de 1 par de autossomos, no centrômero do X e nos microcromossomos (ou *dots*) (Kuhn e cols., 2021). Já em *D. borborema* tanto pBuM-1 quanto pBuM-2 não foram detectados em experimentos hibridização *in situ* em membrana ou em experimentos FISH nos cromossomos (Kuhn & Sene, 2005; Kuhn e cols., 2008; Kuhn e cols., 2021). O satélite DBC-150 foi encontrado em ambas as espécies e em *D. gouveai* foi localizado exclusivamente nos microcromossomos (Kuhn e cols., 2007). Não existem informações sobre a localização cromossômica de DBC-150 em *D. borborema*.

No cluster *buzzatii*, quatro espécies já tiveram seus satelitomas estudados a partir de genomas sequenciados: *D. buzzatii*, *D. seriema*, *D. antonietae* e *D. serido* (Guillén e cols., 2015; Lima e cols., 2017; Nascimento 2023). Este trabalho visa ampliar esse conhecimento, ao investigar o satelitoma de duas outras espécies, do mesmo cluster: *D. gouveai* e *D. borborema* (Figura 4). Estudar o satelitoma dessas duas espécies permitirá comparar e contrastar com os dados já existentes, ajudando a identificar padrões que podem fornecer novos *insights* sobre a evolução e a diversidade genômica dentro do *cluster buzzatii*.

1.5 Identificação de DNAs satélites

Na era pré-genômica, as sequências heterocromáticas, incluindo os satDNAs, foram consideradas “DNA egoísta” ou “DNA lixo” (Podgornaya, 2022). Na era genômica surgiram os primeiros genomas montados como por exemplo: *Caenorhabditis elegans* (The *C. elegans* Sequencing Consortium, 1998), o de *Drosophila melanogaster* (Adams e cols., 2000) e o Projeto Genoma Humano (Lander e cols., 2001; Venter e cols. 2001).

As regiões repetitivas ainda não foram incluídas nesses sequenciamentos, pois as técnicas utilizadas apresentavam limitações práticas, dificultando a montagem das regiões repetitivas (Garrido-Ramos, 2017). Sob uma perspectiva computacional, as repetições causam ambiguidades no alinhamento e na montagem do genoma, o que pode resultar em erros na interpretação dos dados (Treangen & Salzberg, 2012).

A partir do desenvolvimento de novas ferramentas e *softwares* de bioinformática, junto do sequenciamento de nova geração (*Next Generation Sequencing*), tornou-se possível a identificação de todos os DNAs repetitivos de uma espécie, o que inclui os satDNAs (Kroupin e cols., 2019; Fu e cols., 2019; Ruiz-Ruano e cols., 2019). Algumas *pipelines* para a identificação de sequências repetitivas foram desenvolvidas, como é o caso do RepeatExplorer e o TAREAN (Novák e cols., 2020). O primeiro detecta de forma geral os elementos repetitivos, estejam eles dispersos ou em tandem, já o segundo identifica apenas sequências repetitivas em tandem.

Novák e colaboradores (2020) publicaram uma nova versão da pipeline RepeatExplorer, denominada RepeatExplorer2, que possui a ferramenta TAREAN integrada nessa versão, que identifica potenciais satDNAs baseado na topologia dos grafos de seus *clusters* (Novák e cols., 2017). Essas ferramentas possibilitam o descobrimento de novos satDNAs, conforme foi demonstrado em outros estudos do nosso grupo (Kuhn e cols., 2021; Silva e cols., 2021 e 2023).

Os pipelines RepeatExplorer e TAREAN utilizam como *input* um arquivo único FASTQ, formado por *single* ou *paired-end reads* geradas por sequenciamento *shotgun* do genoma. Para começar as análises, é fundamental que as *reads* tenham tamanho uniforme (entre 100 e 200pb), que o número total de *reads* seja inferior a 1x da cobertura do genoma (as diretrizes do pipeline recomendam uma cobertura entre 0,01x e 0,50x), que as *reads* sejam pré-filtradas por qualidade (com score ≥ 10 em 95% das bases, sem "Ns") e que apenas *reads* completas sejam incluídas na análise (Novák e col. 2017).

Ambos os *pipelines* foram projetados com o intuito de identificar elementos repetitivos no genoma. Para isso, foram desenvolvidos com base na montagem de grafos de Bruijn (Compeau e cols. 2011). Primeiramente, as *reads* são agrupadas em clusters com base na similaridade das sequências. Em seguida, são ajustadas para formar k-mers e, finalmente, são geradas sequências consenso. No caso do TAREAN, os grafos criados são avaliados quanto à presença de arranjos em tandem, e os clusters identificados são classificados como satélites de alta ou baixa confiança.

A disponibilidade recente dos genomas sequenciados de *D. gouveai* e *D. borborema* fornece uma oportunidade para acessar pela primeira vez o satelitoma destas espécies e ampliar nosso conhecimento sobre a evolução dos DNAs satélites nas espécies do *cluster buzzatii* e seus impactos. Além disso, o conhecimento sobre a variação destes DNAs satélites pode contribuir para um melhor entendimento das relações filogenéticas entre as espécies do *cluster buzzatii*.

2 OBJETIVOS

2.1 OBJETIVO GERAL

O presente trabalho teve como objetivo geral identificar e caracterizar o satelitoma de duas espécies próximas filogeneticamente do cluster *buzzatii* (grupo: *repleta*; gênero: *Drosophila*): *D. borborema* e *D. gouveai*.

2.2 OBJETIVOS ESPECÍFICOS

- Identificar e caracterizar todos os satDNAs de *D. borborema* e *D. gouveai*;
- Determinar a localização cromossômica dos satDNAs identificados;
- Analisar o padrão de evolução dos satDNAs encontrados;
- Estudar a variação inter-populacional dos satDNAs de *D. borborema*;
- Avaliar o uso dos satDNAs encontrados como marcadores filogenéticos para entender as relações evolutivas entre as espécies *D. borborema*, *D. gouveai*, *D. seriema* e *D. serido*.

3 METODOLOGIA

3.1 GENOMAS SEQUENCIADOS

Neste trabalho utilizamos genomas sequenciados (Illumina HiSeq 2500) de duas linhagens de *D. borborema*: Linhagem (A) proveniente de Morro do Chapéu, BA (3403.4), disponível publicamente no GenBank (SRR22206175), e Linhagem (B) proveniente de Milagres, BA, sequenciado pelo grupo da Profa. Maura Manfrim

(USP-Ribeirão Preto). Para algumas questões específicas, também utilizamos o sequenciamento PacBio de *D. borborema* (A), disponível publicamente no GenBank (SRR22206174). Para *D. gouveai*, utilizamos o genoma sequenciado (Illumina HiSeq 2500) de uma linhagem proveniente de Altinópolis, SP (H6).

3.2 IDENTIFICAÇÃO DE DNAs SATÉLITES

As análises *in silico* para identificação dos DNAs satélites foram conduzidas utilizando a plataforma *online* "Galaxy" (Afgan e cols., 2018) com o *pipeline* RepeatExplorer2 (Novák e cols., 2020). A qualidade das *reads* foi avaliada pelo FastQC (*Read Quality reports*), seguido pela conversão das sequências através do FASTQ Groomer (Sanger & Illumina 1.8+). Após a exclusão de *reads* de baixa qualidade e remoção de adaptadores com a ferramenta *Preprocessing of FASTQ paired-end reads*, as *paired-reads* foram intercaladas em um arquivo chamado *'Interlaced paired reads'*. Esse arquivo foi usado como entrada no pipeline RepeatExplorer2 > Repeat Explorer *clustering*, com os seguintes parâmetros: read sampling: no; select táxon and protein domain database version (REXdb): metazoa version 3.0; select queue: long. Os resultados, apresentados em um arquivo .zip, contém documentos gerados pelos *pipelines* Repeat Explorer e TAREAN.

A classificação do TAREAN utiliza os índices 'C' (Índice de Componente Conectado) que mostra os clusters formados por repetições em tandem e 'P' (Índice de Completude do Par) que determina o tamanho dos arranjos em tandem, para distinguir DNAs satélites de alta (HC) e baixa confiança (LC). Os índices 'C' e 'P' são importantes para a precisão dessa classificação. O TAREAN agrupa as leituras de sequenciamento em aglomerados com base na similaridade de sequência, formando grafos. Grafos com forma globular são sugestivos de satDNAs (Novák e cols., 2017; Silva e cols., 2023). A fim de classificar os satDNAs nas espécies, a partir das sequências consensos geradas, utilizamos o *software* Geneious Prime

2024.0.3, e procuramos por sequências homólogas em vários bancos de dados, como GenBank, RepBase e FlyBase, e entre os satDNAs que já foram descritos no *cluster buzzatii*: pBuM (Kuhn e cols., 1999), DBC-150 (Kuhn e cols., 2007), CDSTR130 e CDSTR138 (Lima e cols., 2017), CDSTR198 (Guillén e cols., 2015) e dois novos satélites SAT109/104 e CDSTR230 (Laborne e cols., 2024).

3.3 FILOGENIA

Usamos o software Geneious Prime 2024.0.3 para extrair cópias completas de cada satDNA dos contigs gerados pelo pipeline RepeatExplorer2. Organizamos as sequências de nucleotídeos para cada satDNA identificado, incluindo aqueles já descritos e novos, em arquivos no formato FASTA. Realizamos o alinhamento das cópias utilizando a ferramenta 'MUSCLE Alignment' Edgar (2004), disponível no software MEGA10 (Tamura e cols., 2021). Todos os alinhamentos foram checados e manualmente corrigidos quando necessário. Para a construção da filogenia de cada satDNA, selecionamos o melhor modelo de substituição nucleotídica através do programa MEGA10 (Tamura e cols., 2021). Foi utilizado o método de máxima verossimilhança para construir árvores filogenéticas, com bootstrap de 1000. As edições das árvores foram feitas utilizando TreeGraph 3 e FigTree v1.4.4 (Stöver & Müller 2010; disponível em <http://tree.bio.ed.ac.uk/software/figtree/>).

3.4 FISH (Hibridização *in situ* por fluorescência)

Cromossomos metafásicos foram obtidos através de gânglios cerebrais de larvas de terceiro instar de *D. gouveai*. Esse procedimento seguiu o protocolo descrito em Baimai (1997) e Ashburner (1989), com algumas modificações. Para a preparação dos cromossomos em metáfase, os gânglios cerebrais foram incubados em Colcemid por 12 minutos e em solução de citrato trissódico a 0,7% por 15-17

minutos. As larvas foram obtidas a partir de uma linhagem de *D. gouveai* (linhagem H6, Altinópolis - SP), que está sendo mantida em nosso laboratório. Não existem linhagens de *D. borborema* disponíveis, por se tratar de espécie muito difícil de se manter em laboratório. As preparações das sondas e os experimentos de FISH foram conduzidos conforme a metodologia descrita por Dias e cols., (2014) e Laborne e cols., (2025).

Para os experimentos de dupla-FISH, as sondas foram marcadas com antidigoxigenina-Rodamina e avidina-FITC (Roche Applied Science), e as hibridizações cromossômicas foram realizadas de acordo com o método descrito por Dias e cols. (2014), com algumas alterações, incluindo um tratamento de 1 minuto e 15 segundos com solução de NaOH 0,07M. Os cromossomos foram contracorados com DAPI (4,6-diamidino-2-fenilindol) em solução antifade (SlowFade, Invitrogen) e analisados em microscópio de epifluorescência Axio Imager A2, equipado com a câmera AxioCamMRm (Zeiss). As imagens foram capturadas com o software Axiovision (Zeiss) e posteriormente editadas no Adobe Photoshop.

4 RESULTADOS E DISCUSSÃO

4.1 DNAs REPETITIVOS EM *D. borborema* e *D. gouveai*

Utilizamos o pipeline RepeatExplorer2 (Novak e cols., 2020) com *short reads* obtidas por meio de sequenciamento Illumina para identificar os DNAs repetitivos de *D. borborema* e *D. gouveai*.

O sequenciamento de DNA gerou quantidades diferentes de dados para cada uma das espécies. Para *D. borborema* (A), o sequenciamento gerou aproximadamente 13,6 Gb, já para *D. borborema* (B), gerou cerca de 5,5 Gb. Para *D. gouveai* (H6), obteve-se aproximadamente 12,0 Gb. A pipeline RepeatExplorer2 analisou 4.074.818 *reads* de *D. borborema* A, 5.717.574 de *D. borborema* B e

4.156.234 *reads* de *D. gouveai*. Cada análise foi repetida 3 vezes, sendo que os resultados foram consistentes entre estas análises independentes.

As análises indicaram que a fração de DNA repetitivo nos genomas de *D. borborema* varia de 17,7% (linhagem A) a 21,8% (linhagem B), enquanto que em *D. gouveai*, a fração repetitiva estimada foi de aproximadamente 21%. As tabelas suplementares (no anexo) contêm todos os DNAs repetitivos identificados em ambas as espécies e a proporção genômica de cada elemento.

Com relação a contribuição genômica dos DNA satélites, em *D. borborema*, observou-se uma diferença entre as linhagens, com a linhagem (A) apresentando 2,83% e a linhagem (B) 1,81%. Em *D. gouveai*, essa contribuição foi de aproximadamente 2,58%. Esses resultados são consistentes com estudos anteriores, que também encontraram uma contribuição baixa de satDNA em outras espécies do grupo *repleta*, como *D. buzzatii* (~1,9%), *D. seriema* (~2,9%), *D. mojavensis* (2,5%), *D. serido* (~1,1 a ~2,9%) e *D. antonietae* (~2,6 a ~3,9%) (Lima e cols., 2017; Laborne e cols., 2025).

A contribuição de elementos repetitivos classificados como TEs pelo RepeatExplorer2 foi maior que a do satDNA em *D. borborema* e *D. gouveai*, assim como também anteriormente relatado por Lima e cols. (2017) para *D. buzzatii* e *D. serido*, e por Laborne e cols. (2025) para *D. antonietae* e *D. serido*. Em *D. borborema*, os TEs representaram aproximadamente 6,05% (A) e 6,88% (B) do DNA genômico, enquanto em *D. gouveai* foi de cerca de 8,25%. Ambas as espécies apresentaram contribuições menores de TEs para o DNA genômico em comparação com outras espécies do *cluster buzzatii*, como *D. buzzatii* (~12%) e *D. seriema* (~18%) (Lima e cols., 2017).

Tanto *D. borborema* quanto *D. gouveai* apresentaram padrões semelhantes na contribuição genômica de diferentes ordens de elementos transponíveis (TEs), com os retrotransposons de repetições longas terminais (LTRs) sendo os mais abundantes em ambas as espécies. Em *D. borborema*, os LTRs representaram cerca de 3,9% e 4,5% nas linhagens (A) e (B), respectivamente, enquanto em *D. gouveai* a contribuição foi de aproximadamente 5,08%. A segunda ordem de TE mais

abundante em ambas as espécies foi a dos LINEs, seguida pelos Helitrons. Resultados semelhantes também foram relatados em outras espécies do *cluster buzzatii* (Laborne e cols., 2025).

Nosso estudo, em conjunto com pesquisas anteriores sobre DNAs repetitivos em *Drosophila* (Silva e cols., 2019; Laborne e cols., 2025), destaca a variação no conteúdo repetitivo entre diferentes espécies. Além disso, as análises revelaram diferenças entre a fração repetitiva dos dois genomas de *D. borborema* examinados. Apesar dos perfis de sequenciamento exibirem tamanhos distintos, a qualidade dos dados foi semelhante, e o pipeline RepeatExplorer2 analisou um número comparável de leituras em ambas as linhagens. Isso sugere que a variação observada no conteúdo de DNA repetitivo entre as linhagens não é um artefato. De modo geral, nossos resultados evidenciam a variação no conteúdo de DNA repetitivo, tanto intraespecífica quanto interespecífica (Figuras 7, 8 e 9).

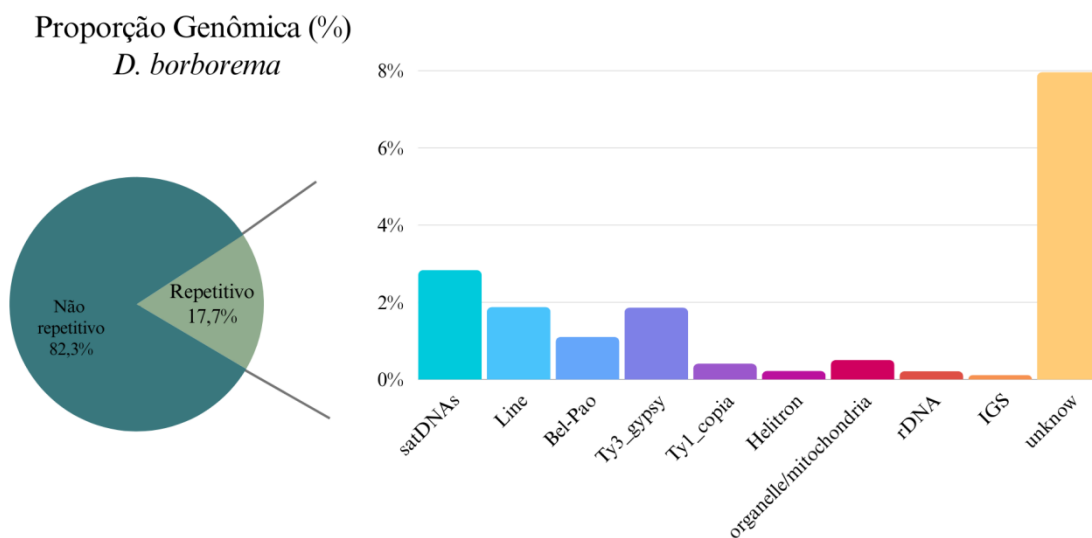


Figura 7. Representação das sequências repetitivas encontradas em *D. borborema* (A)

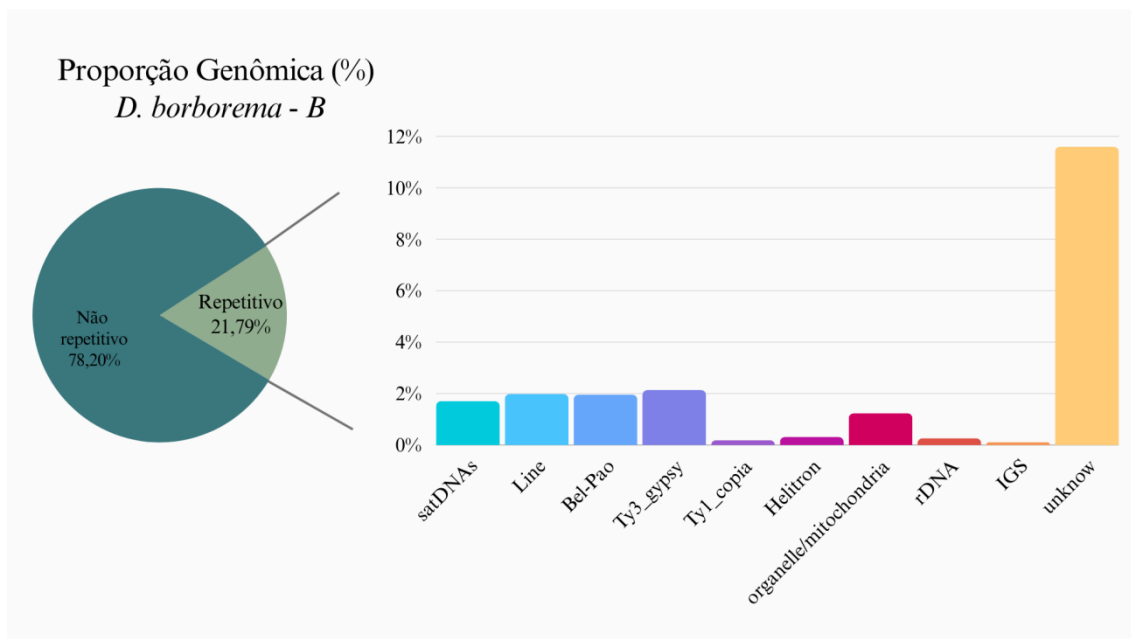


Figura 8. Representação das sequências repetitivas encontradas em *D. borborema* (B)

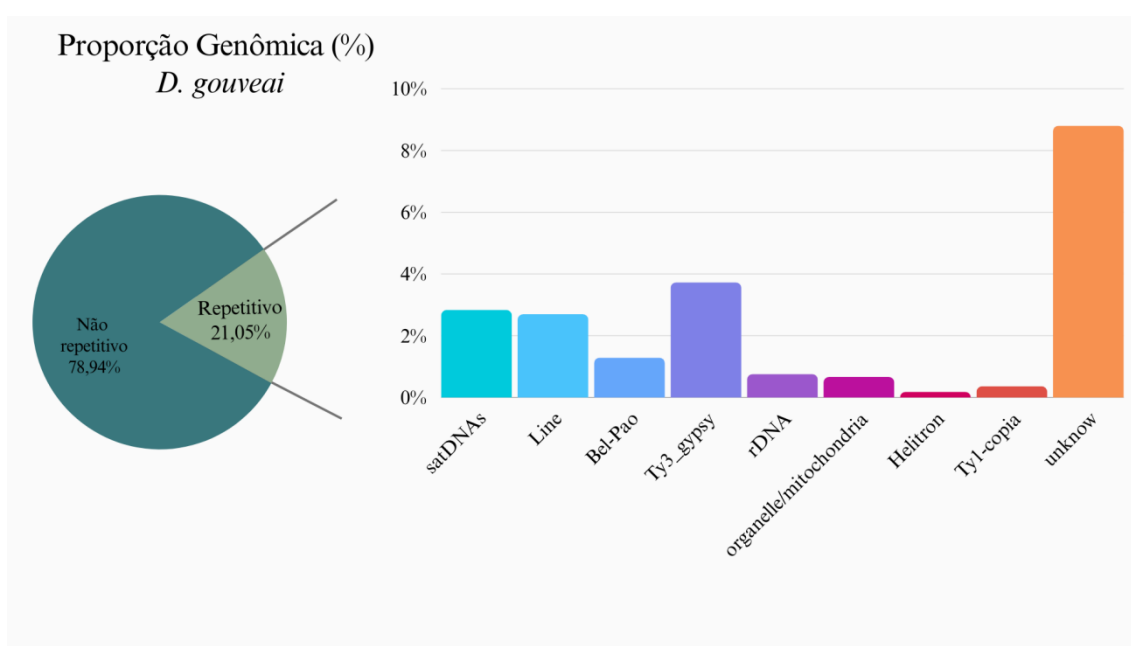


Figura 9. Representação das sequências repetitivas encontradas em *D. gouveai*

4.2 SATELITOMA DE *D. borborema* e *D. gouveai*

O pipeline RepeatExplorer2 inclui uma versão integrada do TAREAN, que possibilita a identificação de potenciais DNAs satélite de alta confiança. As sequências consenso geradas pela pipeline foram submetidas a inspeção manual para garantir a identificação precisa das sequências monoméricas dos satélites. Selecionamos para análise posterior os satDNAs com mais de 1.000 cópias em pelo menos uma das linhagens de *Drosophila* analisadas (Tabela 1). No total, identificamos cinco satDNAs em *D. borborema* e *D. gouveai*.

Nossa análise revelou que *D. borborema* e *D. gouveai* compartilham o mesmo conjunto de cinco DNAs satélites (Tabela 1). Desses, quatro já haviam sido descritos em outras espécies do cluster *buzzatii*. Os satDNAs pBuM e DBC-150 já haviam sido previamente identificados em ambas as espécies (Kuhn & Sene 2005; Kuhn e cols. 2007). O CDSTR138, inicialmente descrito em *D. seriema* (Lima e cols., 2017), foi encontrado aqui pela primeira vez em *D. borborema*. Os satDNAs CDSTR230 e SAT109/104, recentemente descritos em *D. antonietae* e *D. serido* (Laborne e cols., 2024), foram aqui identificados pela primeira vez em *D. borborema* e *D. gouveai*. Além destes, identificamos um novo satDNA, denominado CDSTR16, baseado no comprimento consenso de 16pb em cada espécie. Também relatamos, pela primeira vez, a presença do minissatélite CDSTR198 (Lima e cols., 2017; Kuhn e cols., 2021) em *D. borborema* e *D. gouveai*.

Tabela 1. Características gerais dos DNAs satélites encontrados em *D. borborema* e *D. gouveai*
* = minissatélite

Indivíduo	Cluster	satDNAs	Nº de cópias	Prop. do DNA Genômico(%)	Probab. de satélite (%)	Tamanho do consenso (bp)
<i>D. brb A</i>	1	CDSTR16	202500	1,792	0,979	16
	20	CDSTR230	1910	0,262	0,994	245
	32	CDSTR138	2348	0,177	0,994	138
	40	DBC-150	2089	0,133	0,985	112
	17	CDSTR198*	2337	0,269	0,815	208
<i>D. brb B</i>	2	CDSTR16	36563	0,650	0,985	32
	7	CDSTR230	2571	0,350	0,986	245
	24	SAT109/104	4395	0,209	0,853	86
	25	CDSTR138	2739	0,200	0,986	138
	30	DBC-150	3214	0,180	0,994	112
	38	CDSTR198*	1820	0,110	0,941	178
<i>D. gou</i>	1	CDSTR230	4461	0,578	0,992	230
	5	SAT109/104	9837	0,469	9,55E-01	86
	6	CDSTR138	6221	0,468	0,981	136
	8	CDSTR16	33000	0,442	0,933	24
	9	pBuM2	1995	0,412	0,992	370
	30	DBC-150	2537	0,216	0,933	149
	100	CDSTR198*	603	0,063	0,0661	188

O primeiro DNA satélite descrito no *cluster buzzatii* foi o pBuM. Na nossa análise utilizando a pipeline TAREAN, encontramos o pBuM apenas em *D. gouveai*, onde apresentou 370 pb, correspondendo as repetições alfa/beta (pBuM-2). Nas análises utilizando o TAREAN não foi possível identificar o pBuM em nenhum dos cluster gerados nas análises dos dois genomas de *D. borborema*. No entanto, identificamos o pBuM-2 no genoma montado utilizando a plataforma PACBIO (Moreyra e cols., 2023). Foi possível identificar sequências de pBuM2 em *D. borborema* (A) (Figura 10). Encontramos somente duas cópias parciais do satDNA, no *scaffold* JAUESP010000267|JAUESP010000267.1, onde identificamos uma sequência de 94pb, correspondendo ao final da sequência beta de pBuM-2, seguido de uma sequência de 230pb que compreende a porção alfa e o início da porção beta. Entre as duas cópias do pBuM-2, identificamos o TE Galileo (Figura 11). Estes

resultados mostram que o satDNA pBuM não foi perdido completamente do genoma de *D. borborema*.

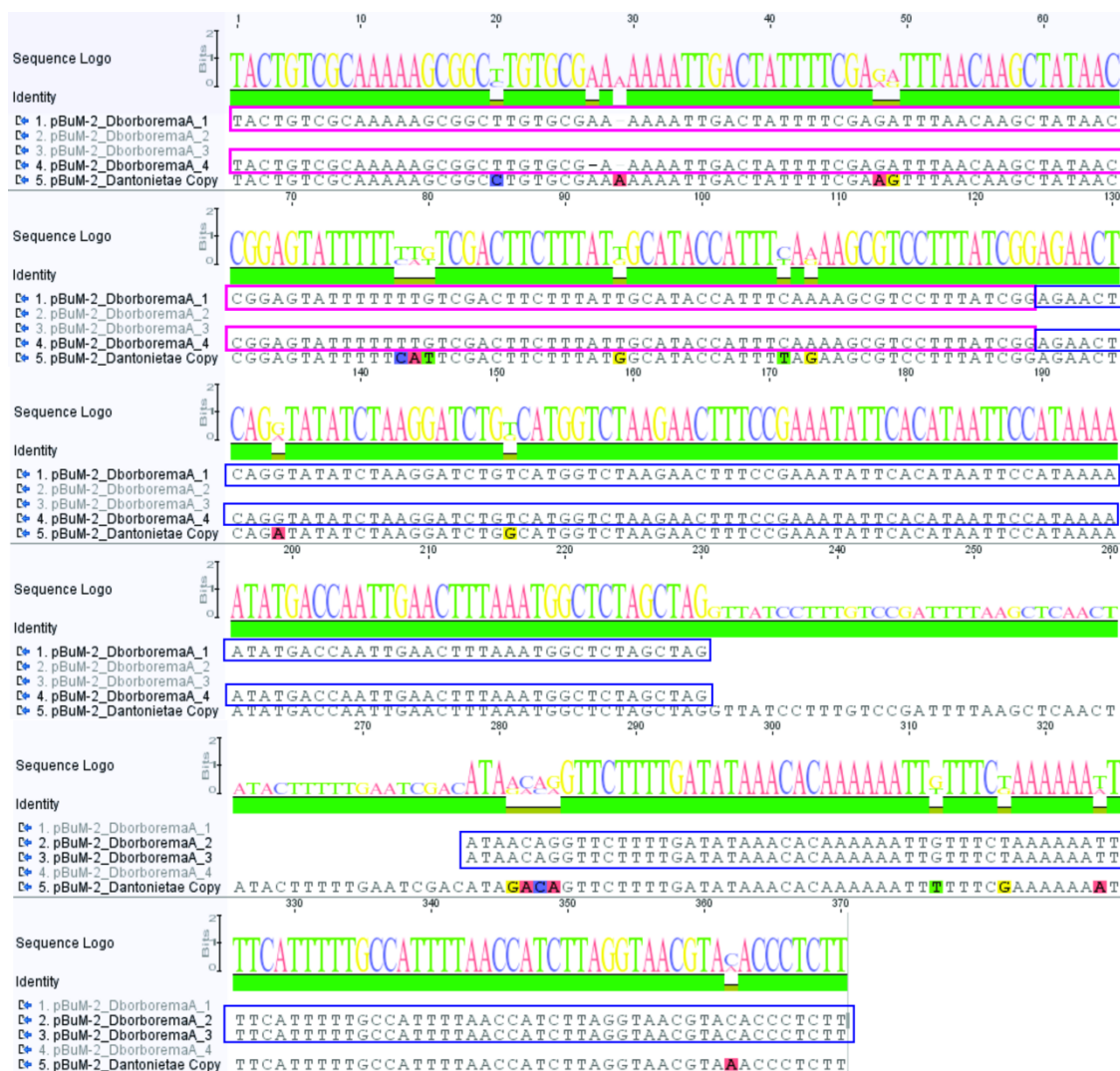


Figura 10. Sequencia das cópias do satDNA pBuM-2 encontrados em *D. borborema*, em comparação com a sequência consenso de *D. antonietae*. Nos retângulos cor-de-rosa está destacada a porção *alpha*, e nos azuis a porção *beta*.



Figura 11. Esquema representativo de pBuM-2 encontrado no genoma de *D. borborema* (A)

O satDNA DBC-150, com repetições de 150pb, foi previamente identificado em ambas as espécies (Kuhn e cols. 2007). Recentemente Laborne e colaboradores (2024) relataram variações desse satDNA com tamanhos de 112pb e 46pb. Em nossa análise, o TAREAN gerou um comprimento consenso de 112pb para ambas as populações de *D. borborema* (indicando que esta variante é a predominante na espécie) e de 149pb para *D. gouveai*, que difere dos comprimentos de repetições inicialmente relatados. No entanto, nas análises dos *contigs* gerados pelo TAREAN, pudemos observar a presença de cópias tanto de 112pb quanto de 150pb, em ambas as espécies. Em *D. borborema* o satDNA apresentou um conteúdo AT de ~45,4% e ~45,5% na população A e B respectivamente, já em *D. gouveai* a proporção foi de ~41,6%, um pouco diferente da encontrada por Kuhn e cols. (2007), onde apresentava ~44% de AT. Apesar das diferenças na sequência do satélite, ele apresentou características semelhantes ao satDNA inicialmente descrito (Kuhn e cols. 2008, 2021).

O satDNA CDSTR138 inicialmente foi descrito por Lima e cols. (2017) em *D. seriema* e, até recentemente, não havia sido descrito em outras espécies do *cluster buzzatii*. Laborne e cols. (2025) encontraram o satDNA CDSTR138 em *D. antonietae* e *D. serido*. No presente trabalho, encontramos este satDNA tanto em *D. borborema* (A) e (B), quanto em *D. gouveai*, sendo o segundo satDNA mais abundante nessa espécie. O CDSTR138 apresentou um tamanho de 138pb em *D. borborema* e 136pb em *D. gouveai*, com um conteúdo AT de 67,4% e 69,9% respectivamente, sendo similar aos encontrados em estudos anteriores (Lima e cols., 2017; Laborne e cols., 2025).

O DNA satélite CDSTR230 foi descrito recentemente por Laborne e cols. (2025) em *D. antonietae* e *D. serido*, onde apresentou um tamanho de 197pb e 230pb respectivamente. Em *D. borborema* ele apresentou um tamanho de 245pb, possui conteúdo AT de 71,8% e é o segundo mais abundante na espécie, tendo uma contribuição genômica de ~0,26% e de ~0,35% em *D. borborema* A e B, respectivamente. Em *D. gouveai* foi o satDNA mais abundante na espécie,

ocupando uma proporção genômica de ~0,57%, com conteúdo AT de 70,9% e apresentou um tamanho consenso de 230pb.

O SAT109/104, também descrito no trabalho de Laborne e cols. (2025), recebeu esse nome devido ao tamanho consenso encontrado em *D. antonietae* e *D. serido*. Entretanto análises posteriores revelaram que essa sequência é um espaçador intergênico (IGS) de gene ribossomal, não sendo considerado um satDNA.

O novo satDNA, CDSTR16, apresenta tamanhos diferentes em *D. borborema* (16 pb) e *D. gouveai* (24 pb), com conteúdo de AT de 68,7% e 75% em *D. borborema* A e B, respectivamente, e de 70,9% em *D. gouveai*. Esse satDNA é o mais abundante em ambas as populações de *D. borborema*, contribuindo com 1,79% (A) e 0,65% (B) do genoma, enquanto em *D. gouveai* sua contribuição é de 0,4%. Uma análise detalhada das cadeias contendo repetições em tandem revelou que este satDNA possui organização monomérica e também como *higher order repeat* (HOR). Nas análises manuais dos *contigs* gerados pelo TAREAN, encontramos variações nas sequências entre as duas populações de *D. borborema* (Figura 12, Figura suplementar 1 e 2).

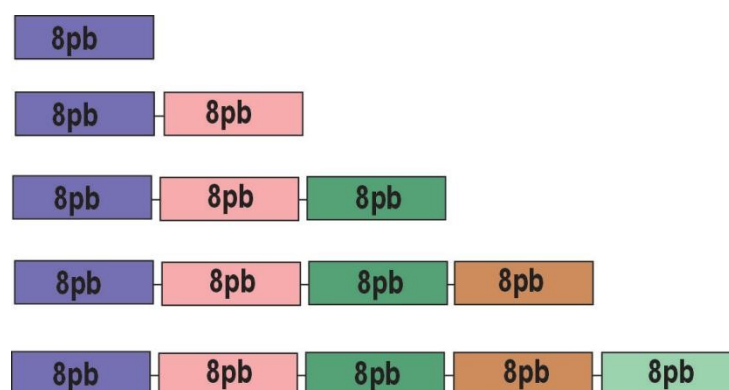


Figura 12. Representação das variações das sequências dos monômeros (8pb) encontrados em diferentes contigs de *D. borborema*

Também identificamos o minissatélite CDSTR198, previamente descrito em *D. buzzatii* (Guillén e cols., 2014), sendo aqui relatado pela primeira vez em ambas

as espécies. O CDSTR198 apresentou um conteúdo de AT de 64,4% em *D. borborema* A, 63,5% em *D. borborema* B e 65,4% em *D. gouveai*. Seu consenso foi de 208pb e 178pb em *D. borborema* A e B, respectivamente, e de 188pb em *D. gouveai*. Em *D. gouveai*, esse minissatélite apresentou menos de 1.000 cópias, com uma contribuição genômica de 0,07%. Em *D. borborema* (B), foram encontradas mais de 1.000 cópias, contribuindo com 0,110% do genoma, enquanto em *D. borborema* (A), é a segunda sequência repetitiva mais abundante, com aproximadamente 2.337 cópias e uma contribuição genômica de 0,26%. No entanto, em uma busca por similaridade no genoma montado de *D. borborema* (A) (Moreyra e cols., 2023), encontramos cópias dispersas do minissatélite. Considerando essa dispersão, ele provavelmente é um minissatélite na espécie, apesar do alto número de cópias.

Os DNAs satélites identificados em *D. borborema* e *D. gouveai* revelaram um tamanho de monômeros variando de 16 a 370pb, que são predominantemente ricos em AT, exceto pelo DBC-150, que apresentou um conteúdo médio de AT de 44%. Essas características parecem ser comuns aos satDNAs em espécies do grupo *repleta* (revisado por Kuhn e cols. 2021). Além disso, a contribuição total dos satDNAs para o DNA genômico é baixa em *D. borborema* (1,8-2,8%) e *D. gouveai* (2,5%) (Figura 13). Esses resultados estão de acordo com estudos anteriores em espécies do grupo *repleta* e sustentam a hipótese de que a baixa contribuição de DNA satélite pode ser consequência de restrições seletivas (Lima e cols. 2017; Kuhn e cols. 2021).

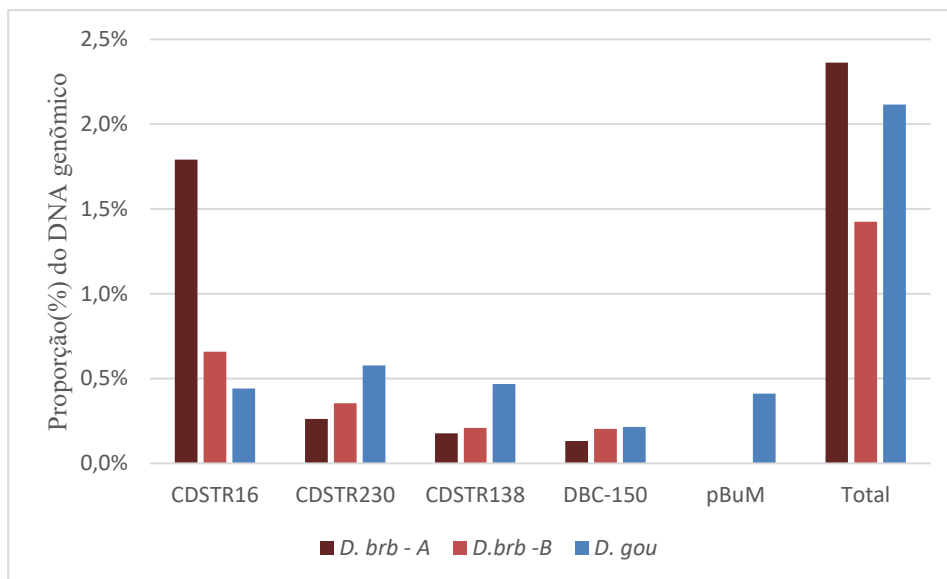


Figura 13. Proporção (%) dos DNAs satélites no DNA genômico de *D. borborema* (linhagens A e B) e *D. gouveai*.

4.3 LOCALIZAÇÃO CROMOSSÔMICA DOS satDNAs

Espécies do cluster *buzzatii* apresentam cariótipos com um par de cromossomos sexuais, quatro pares de autossomos telocêntricos de mesmo tamanho e um par de microcromossomos (Figura 14). Em algumas espécies, o cariótipo pode ser diferenciado pela quantidade de heterocromatina presente nos microcromossomos e nos cromossomos sexuais (Baimai e cols., 1983; Kuhn e cols., 1996; Kuhn e cols., 2007). Com o objetivo de determinar a localização dos satDNA encontrados em *D. gouveai*, realizamos experimentos de FISH em cromossomos metafásicos.

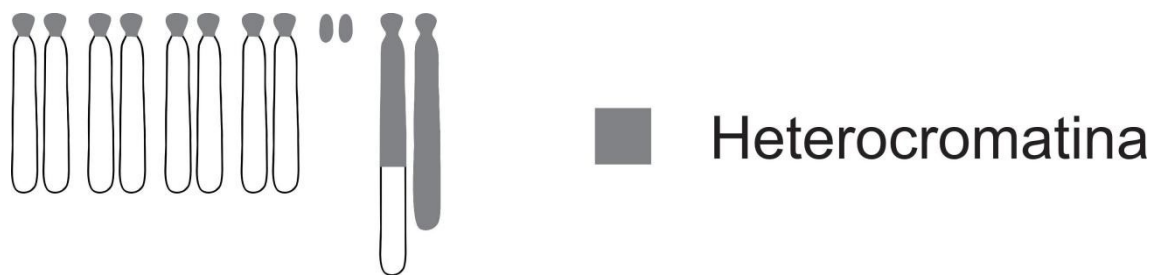


Figura 14. Idiograma representativo do cariótipo básico encontrado em espécies do cluster *buzzatii*

O satDNA pBuM2 foi localizado nos microcromossomos e na região centromérica de dois pares de cromossomos autossômicos e do cromossomo X, assim como previamente relatado por Kuhn e cols. (2008). No entanto, em nosso experimento de double-FISH contendo sondas para o pBuM2 e CDSTR16, percebemos que o pBuM2 está localizado em uma região mais distal do centrômero (Figura 15 e 18).

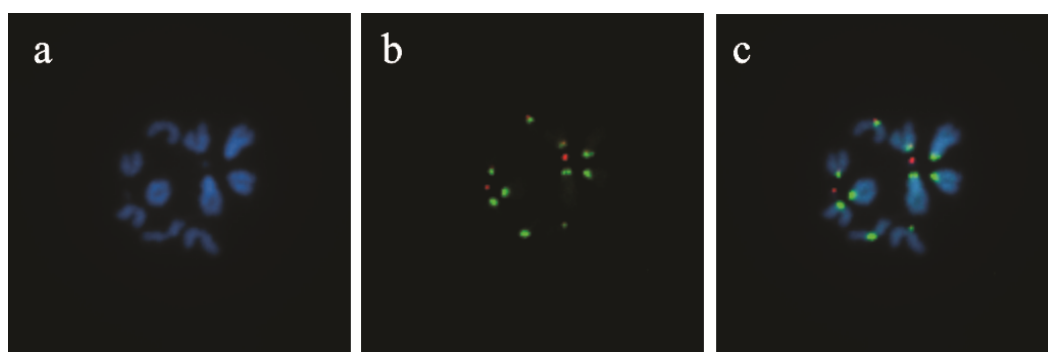


Figura 15. Hibridização *in situ* fluorescente (FISH) de pBuM2 (vermelho) e CDSTR16 (verde) em cromossomos metafásicos de *D. gouveai*. **a.** cromossomos metafásicos contraindo com DAPI **b.** sinal fluorescente da hibridização das sondas **c.** cromossomos contraindo com DAPI junto da hibridização

O novo DNA satélite identificado nesse trabalho, o CDSTR16, foi localizado nos centrômeros de todos os autossomos e no centrômero do cromossomo X. Já o satélite CDSTR230, satDNA mais abundante em *D. gouveai*, foi localizado nos microcromossomos e na região da beta-heterocromatina do cromossomo X e no braço longo do cromossomo Y.

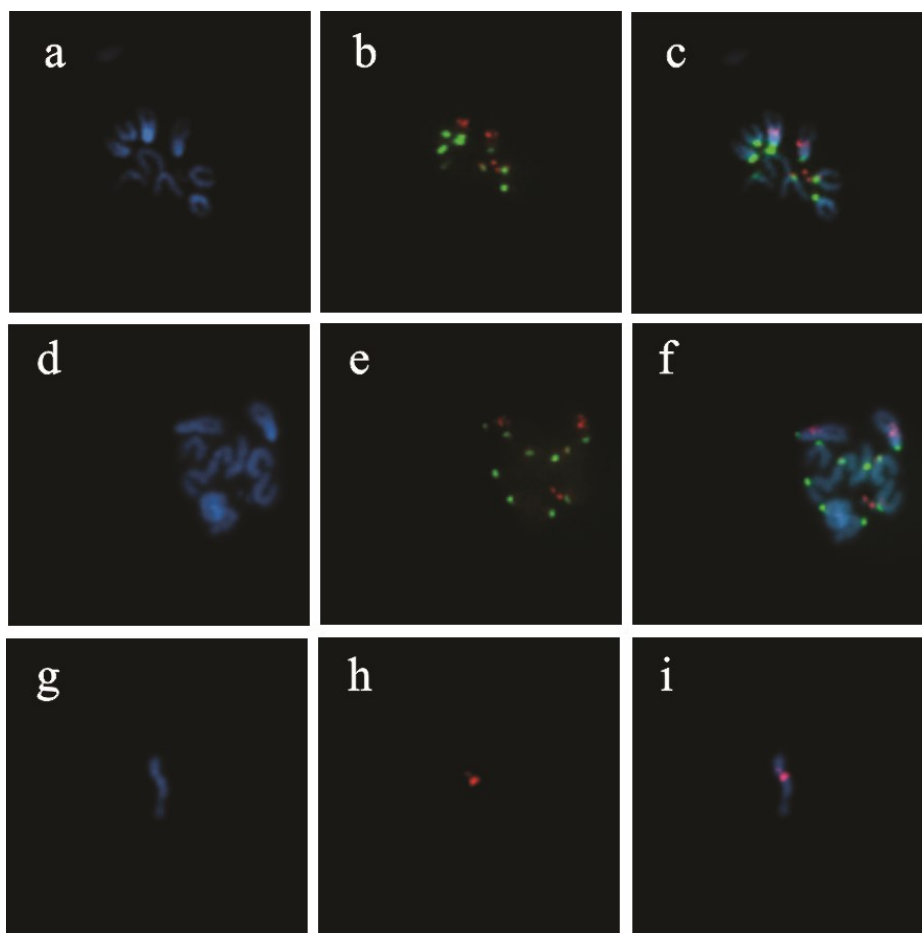


Figura 16. FISH de CDSTR230(vermelho) e CDSTR16 (verde) em cromossomos metafásicos de *D. gouveai*. **a. d. g.** cromossomos metafásicos contracolorados com DAPI **b. e. h.** sinal fluorescente da hibridização das sondas **c. f. i.** cromossomos contracolorados com DAPI junto da hibridização. **g. h. i.** sinal no cromossomo Y.

O SAT109-104 foi localizado nos centrômeros dos cromossomos sexuais, assim como observado em *D. serido* e *D. antonietae* por Laborne e cols (2024). Este resultado indica que, assim como nas últimas duas espécies, o SAT109-104 é um componente de espaçador intergênico (IGS), não sendo portanto, um satDNA.

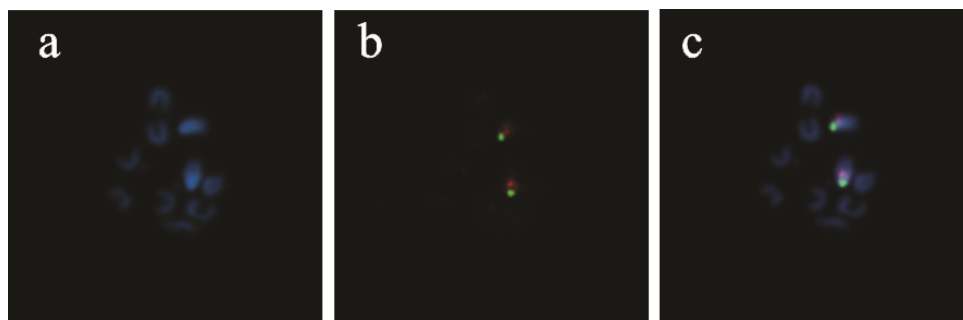


Figura 17. FISH de CDSTR230 (vermelho) e SAT109/104 (verde) em cromossomos metafásicos de *D. gouveai*. **a.** cromossomos metafásicos contracolorados com DAPI **b.** sinal fluorescente da hibridização das sondas **c.** cromossomos contracolorados com DAPI junto da hibridização

Em nosso experimento de double-FISH com os satélites pBuM2 e CDSTR138, o satDNA CDSTR138 foi localizado nos centrômeros de todos os autossomos, enquanto que o pBuM2 foi localizado em regiões mais distais.

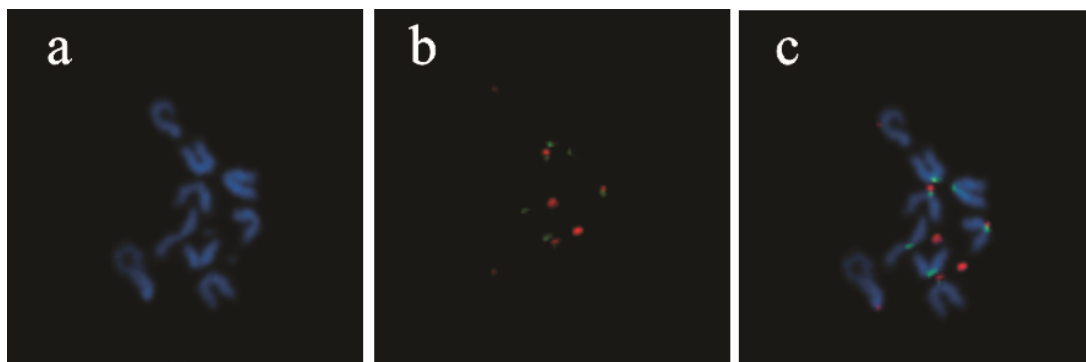


Figura 18. FISH de pBuM2 (vermelho) e CDSTR138 (verde) em cromossomos metafásicos de *D. gouveai*. **a.** cromossomos metafásicos contracolorados com DAPI **b.** sinal fluorescente da hibridização das sondas **c.** cromossomos contracolorados com DAPI junto da hibridização

Nos experimentos relatados acima, observamos que tanto o satélite CDSTR16 quanto o CDSTR138 estão localizados na região centromérica. Por este motivo, realizamos a double-FISH com sondas de ambos satDNAs. Com este experimento, foi possível identificar que ambos CDSTR16 e CDSTR138 co-

localizam na região centromérica. Resultados similares já foram identificados em outras espécies, por exemplo em *D. melanogaster*, onde mais de um satDNA se encontra no centrômero (Sun e cols., 2000).

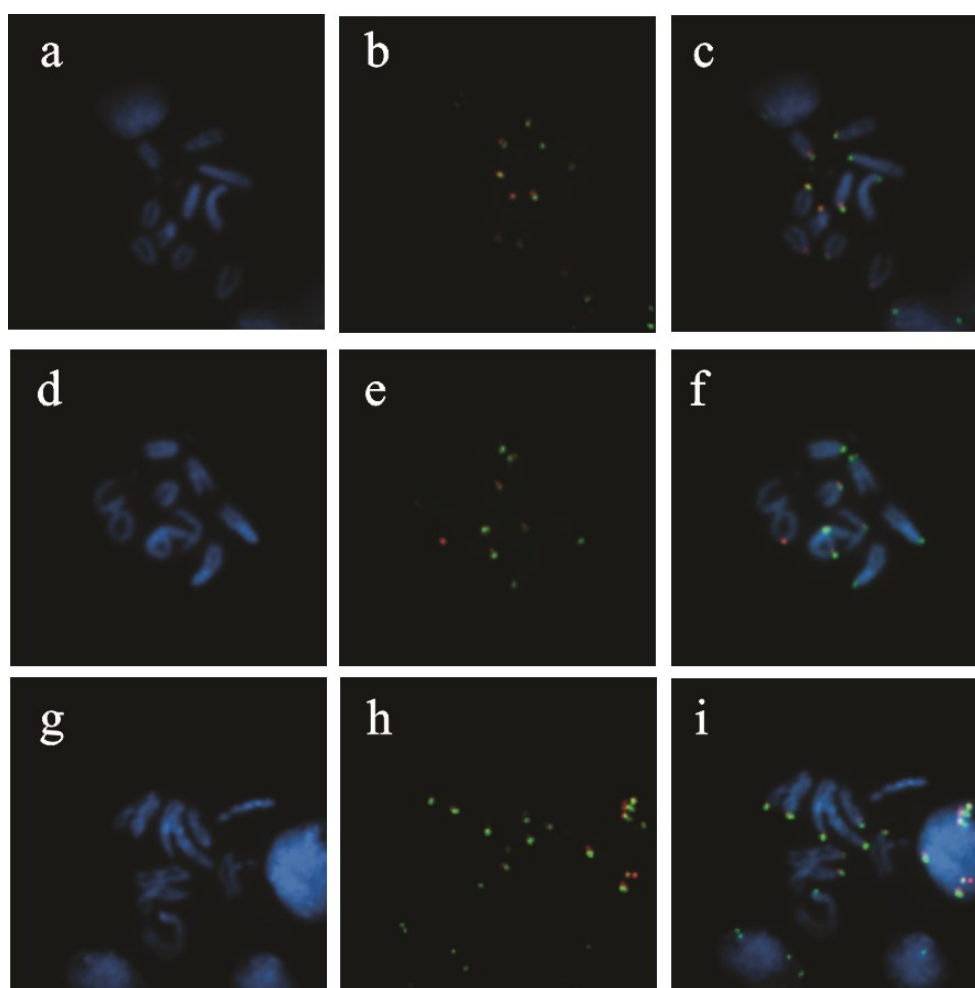


Figura 19. FISH de CDSTR138(vermelho) e CDSTR16 (verde) em cromossomos metafásicos de *D. gouveai*. **a-f** cromossomos metafásicos femininos **g-i**. cromossomos metafásicos masculino, nenhum sinal observado no cromossomo Y.

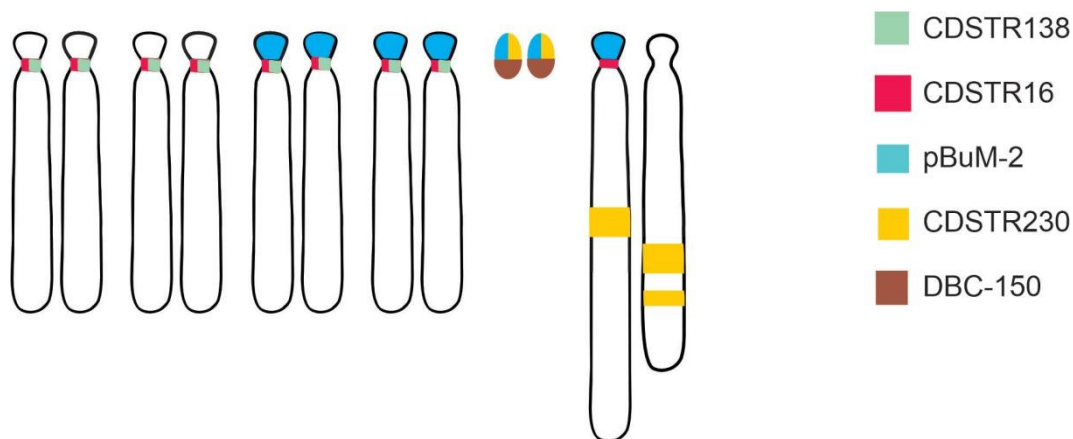


Figura 20. Idiograma com a localização cromossômica dos satDNAs em *D. gouveai*

Os resultados de FISH mostraram que em *D. gouveai*, os principais satDNAs da região centromérica são o CDSTR138 e CDSTR16. O satDNA CDSTR230 é no entanto o mais abundante de *D. gouveai*. Este resultado mostra que nem sempre o satDNA mais abundante de uma espécie corresponde ao centromérico, como proposto por Melters e cols. (2013).

4.4 EVOLUÇÃO DOS satDNAs EM *D. borborema* e *D. gouveai*

Os DNAs satélites são componentes que evoluem rapidamente nos genomas eucarióticos. Em contraste, as repetições dentro da mesma família de satDNA costumam exibir alta homogeneidade dentro da espécie. Esse processo é conhecido como evolução combinada (Zimmer e cols. 1980; Dover 1982).

Os dados obtidos em nosso trabalho mostraram que o satDNA pBuM apresentou o padrão de evolução combinada em *D. borborema* (A) e *D. gouveai* (Figura 21), consistente com pesquisas anteriores sobre este satDNAs em espécies do grupo *repleta*. Nesses estudos, a família de satDNA pBuM, abrangendo as variantes pBuM-1 e pBuM-2, seguiu um padrão de evolução combinada (Kuhn e cols., 2007; Lima e cols., 2017; Laborne e cols., 2024). Para a construção da árvore

filogenética, não foram utilizadas cópias de pBuM-2 de *D. borborema* (B), pois sequencias desse satDNA so foram identificadas na linhagem (A), no genoma montado após sequenciamento pela plataforma PACBIO.

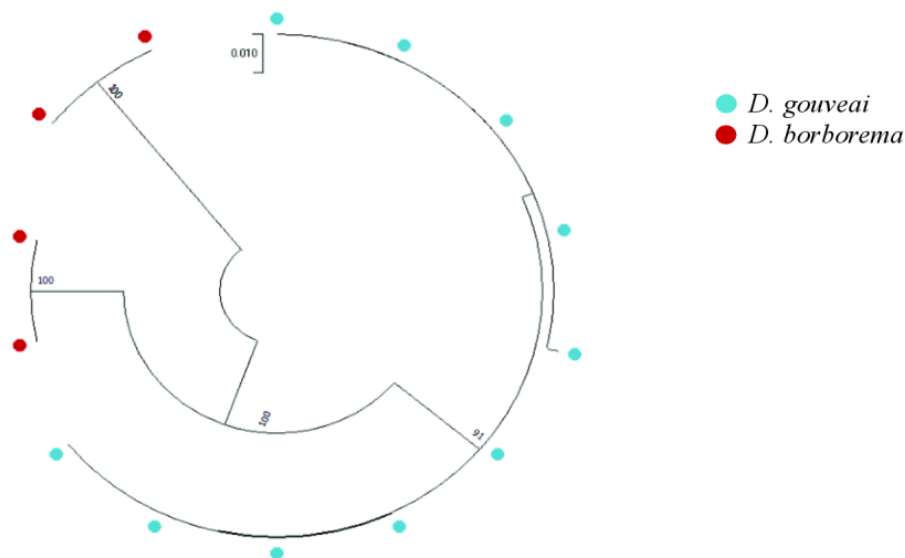


Figura 21. – Árvore de máxima verossimilhança (ML) contendo cópias de pBuM-2 de *D. borborema* A(vermelho) e *D. gouveai* (azul). A árvore foi inferida pelo modelo Tamura de 3 parâmetros com 1000 réplicas de bootstrap. O suporte mínimo de bootstrap é de 50%.

As repetições do satDNA DBC-150 analisadas neste estudo (Figuras 22-23) seguiram um padrão de evolução combinada, com cópias de *D. borborema* e *D. gouveai* agrupadas em ramos espécie-específicos na filogenia, o que contrasta com resultados anteriores obtidos por Kuhn e cols. (2007). Esta discrepância pode ser explicada pelo fato das cópias do DBC-150 analisadas em Kuhn e cols (2007) terem sido obtidas através de PCR com apenas um par de oligos, o que pode ter trazido viés na amplificação das sequencias.

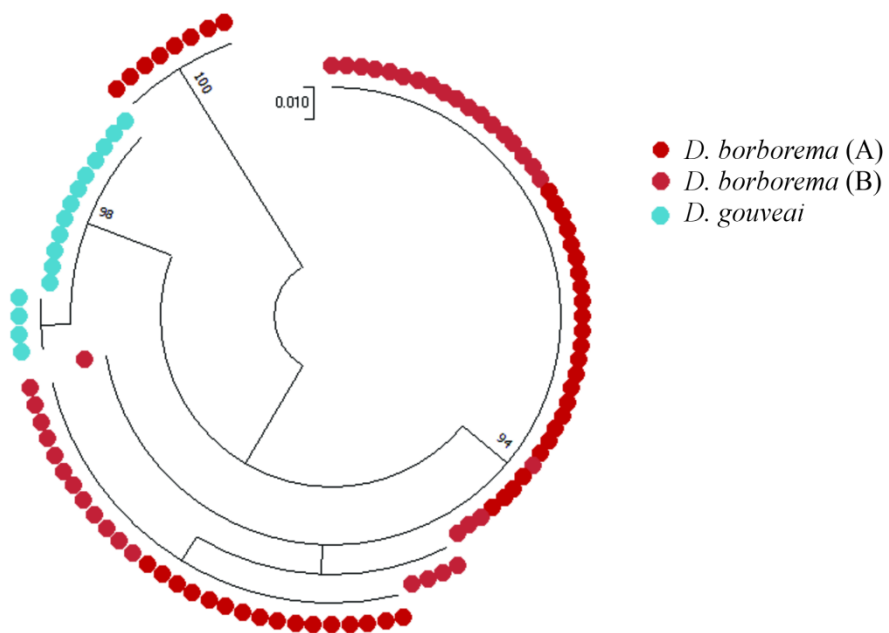


Figura 22 – Árvore de máxima verossimilhança (ML) contendo cópias de DBC-112 de *D. borborema* A (vermelho), *D. borborema* B (rosa) e *D. gouveai* (azul). A árvore foi inferida pelo modelo Jukes-Cantor com 1000 réplicas de bootstrap. O suporte mínimo de bootstrap é de 50%.

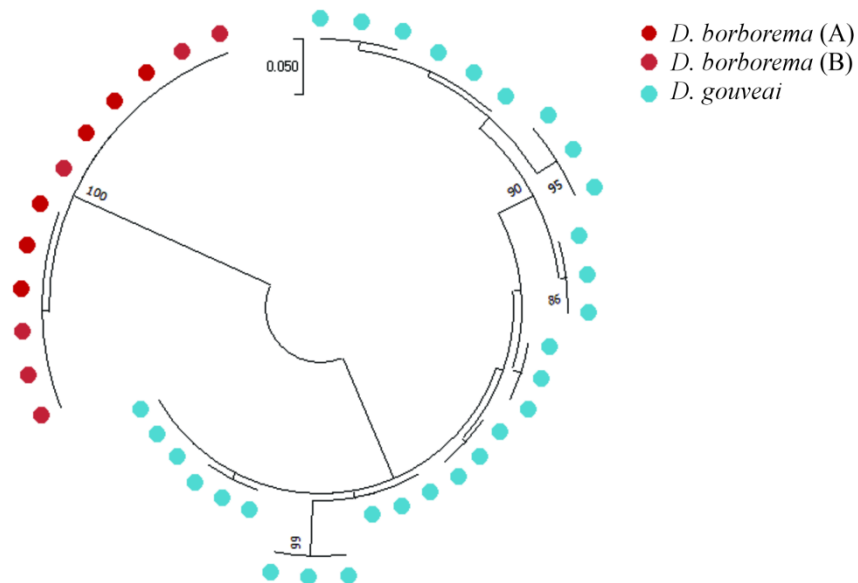


Figura 23 - Árvore de máxima verossimilhança (ML) contendo cópias de DBC-150 de *D. borborema* A (vermelho), *D. borborema* B (rosa) e *D. gouveai* (azul). A árvore foi inferida pelo modelo Hasegawa-Kishino-Yano, utilizando uma distribuição Gamma discreta (+G) com 5 categorias, e 1000 réplicas de bootstrap. O suporte mínimo de bootstrap é de 50%.

Com os satélites CDSTR138 e CDSTR230, nós também observamos o padrão de evolução combinada, pois cópias destes satDNAs também formam ramos espécies-específicos nas filogenias (Figuras 24 e 25)

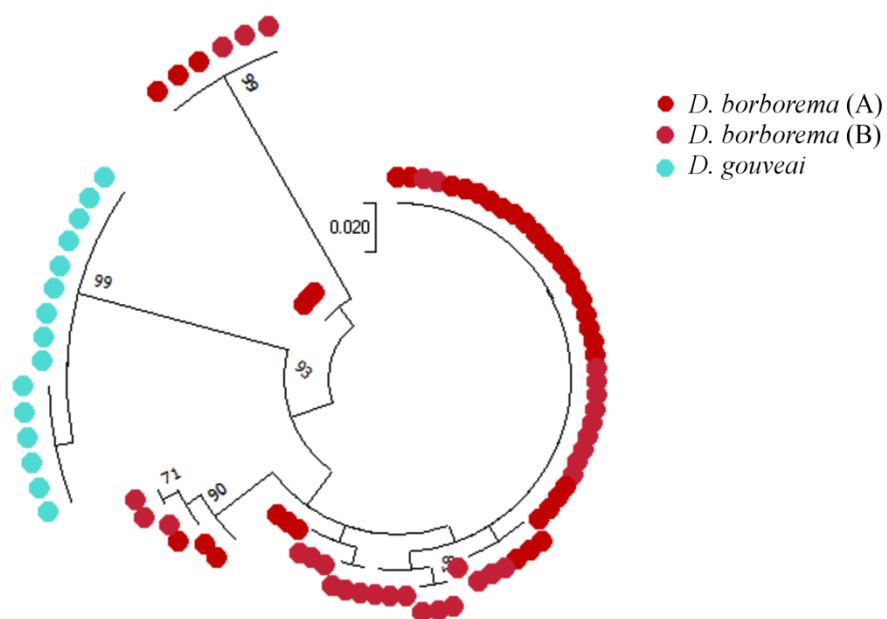


Figura 24 - Árvore de máxima verossimilhança (ML) contendo cópias de CDSTR138 de *D. borborema* A(vermelho), *D. borborema* B (rosa), e *D. gouveai* (azul). A árvore foi inferida pelo modelo Tamura de 3 parâmetros com 1000 réplicas de bootstrap. O suporte mínimo de bootstrap é de 50%.

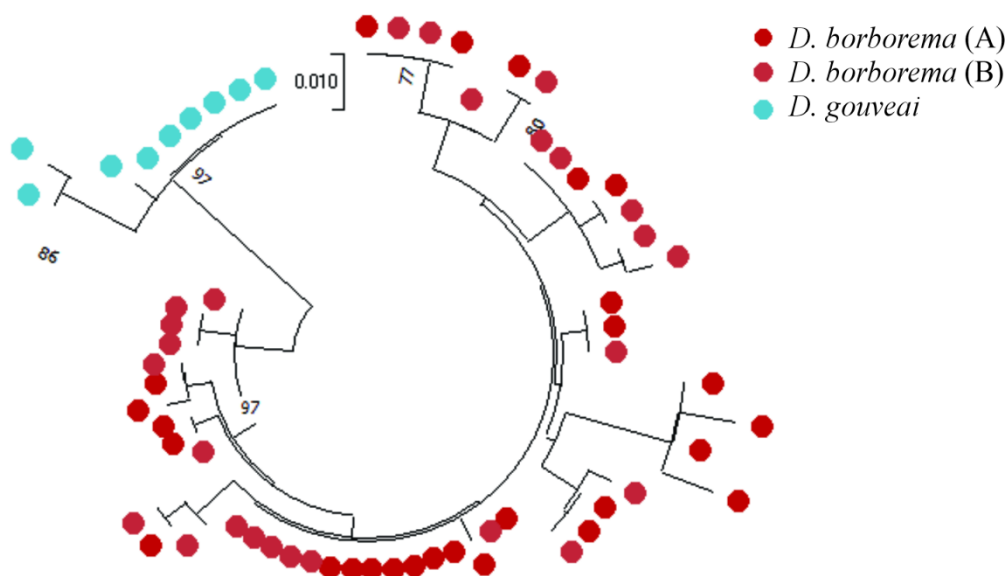


Figura 25 - Árvore de máxima verossimilhança (ML) contendo cópias de CDSTR230 de *D. borborema* A (vermelho), *D. borborema* B (rosa) e *D. gouveai* (azul). A árvore foi inferida pelo modelo Hasegawa-Kishino-Yano, utilizando uma distribuição Gamma discreta (+G) com 5 categorias, e 1000 réplicas de bootstrap. O suporte mínimo de bootstrap é de 50%.

Através da análise das filogenias acima, podemos também concluir que não houve evolução combinada em *D. borborema* no nível inter-populacional. Em outras palavras, cópias dos satDNAs das linhagens (A) e (B) não formaram ramos exclusivos de cada população. Entre as hipóteses, esta ausência de divergência pode ter sido causada por fatores evolutivos, como fluxo gênico, seleção, ou tempo recente de separação entre as populações. Ausência de diferenciação inter-populacional já foi relatada em espécies de *Drosophila* do cluster *buzzatii* (Franco e cols., 2006).

4.5 DNAs SATÉLITES COMO MARCADORES FILOGENÉTICOS

Atualmente, a relação filogenética entre as espécies *D. borborema*, *D. gouveai*, *D. seriema* e *D. serido* não está bem definida. Para investigar se os satDNAs podem auxiliar no esclarecimento dessa filogenia, realizamos análises de árvores de máxima verossimilhança utilizando as sequências de cada satDNA. Para satDNAs das espécies *D. serido* e *D. seriema*, as cópias utilizadas foram obtidas de estudos anteriores do grupo (Lima e cols., 2017; Laborne e cols., 2025).

As árvores resultantes (Figuras 26 a 29) revelaram vários obstáculos que impediram a elucidação das relações filogenéticas entre as espécies, como cópias da mesma espécie distribuídas em ramos contendo cópias de outras espécies e topologias diferentes entre as filogenias. Estes resultados provavelmente refletem a divergência recente entre estas espécies (<1My) (Hurtado e cols., 2019).

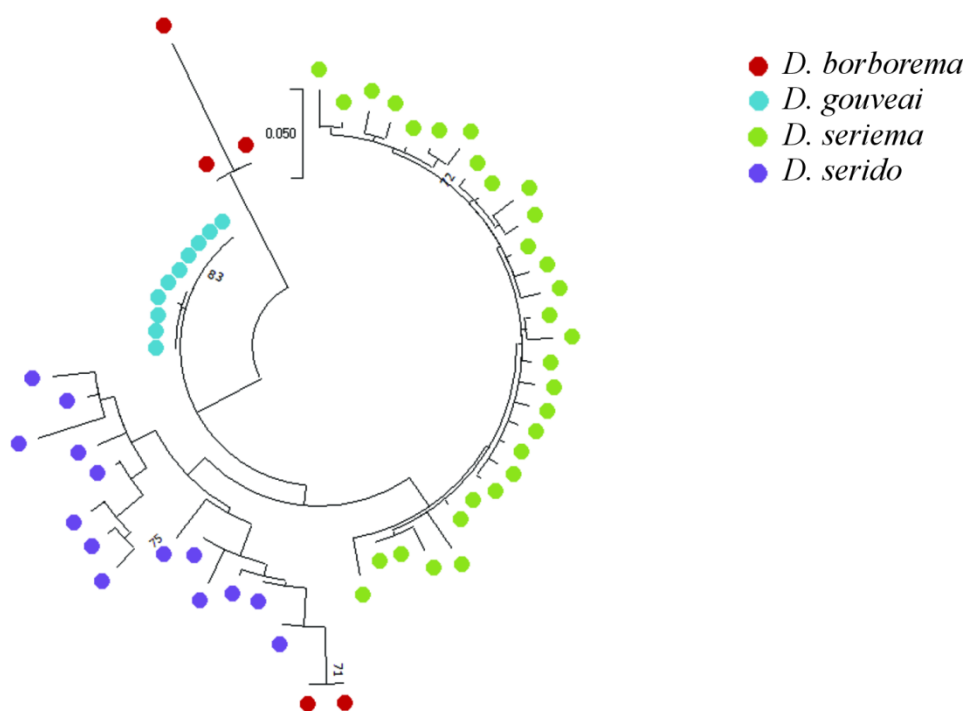


Figura 26 - Árvore de máxima verossimilhança (ML) contendo cópias de pBuM-2 de *D. borborema* (vermelho), *D. gouveai* (azul), *D. serido* (roxo), e *D. seriema* (verde). A árvore foi inferida pelo

modelo Tamura de 3 parâmetros, utilizando uma distribuição Gamma discreta (+G) com 5 categorias, com 1000 réplicas de bootstrap. O suporte mínimo de bootstrap é de 50%.

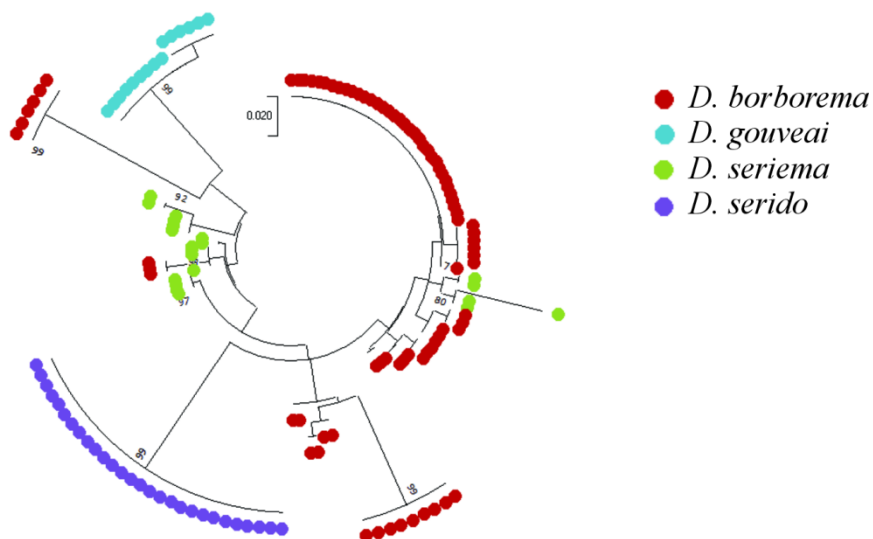


Figura 27 - Árvore de máxima verossimilhança (ML) contendo cópias de CDSTR138 de *D. borborema* (vermelho), *D. gouveai* (azul), *D. serido* (roxo), e *D. seriema* (verde). A árvore foi inferida pelo modelo Tamura de 3 parâmetros, com 1000 réplicas de bootstrap. O suporte mínimo de bootstrap é de 50%.

Kimura de 2 parâmetros, utilizando uma distribuição Gamma discreta (+G) com 5 categorias, com 1000 réplicas de bootstrap. O suporte mínimo de bootstrap é de 50%.

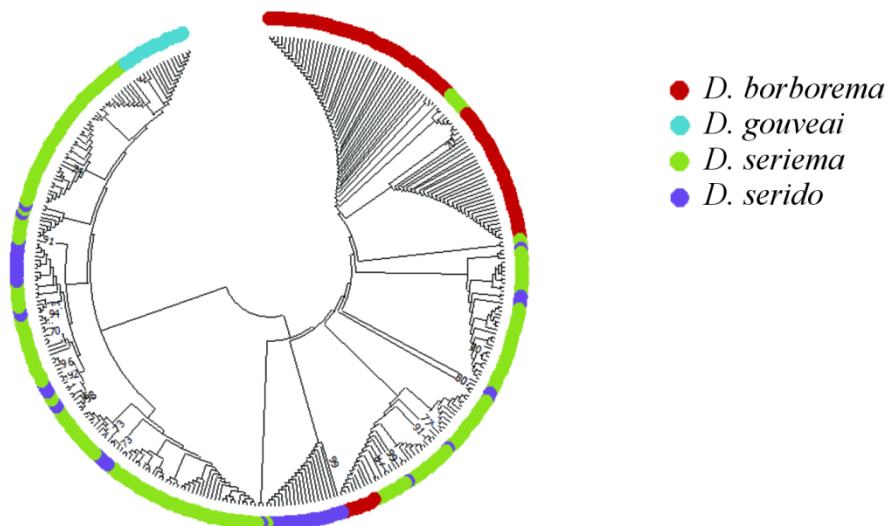


Figura 30 – Árvore de máxima verossimilhança (ML) contendo cópias de DBC-112 de *D. borborema* (vermelho), *D. gouveai* (azul), *D. serido* (roxo), e *D. seriema* (verde). A árvore foi inferida pelo modelo Kimura de 2 parâmetros, utilizando uma distribuição Gamma discreta (+G) com 5 categorias, com 1000 réplicas de bootstrap. O suporte mínimo de bootstrap é de 50%.

5 CONCLUSÃO

Em relação aos nossos objetivos:

- Identificar e caracterizar todos os satDNAs de *D. borborema* e *D. gouveai*

Com a pipeline RepeatExplorer identificamos quatro DNAs satélites em *D. borborema* e cinco em *D. gouveai*. Conseguimos identificar nestas espécies os satDNAs previamente descritos no cluster *buzzatii*: CDSTR138 e CDSTR230, além de identificarmos um novo satDNA. Observamos que a proporção genômica ocupada pelos satDNAs é baixa (<5%) assim como descrita anteriormente em

outros estudos do grupo *repleta*. Identificamos a presença do minissatélite CDSTR198 em ambas as espécies, o qual também foi descrito anteriormente em outras espécies do *cluster buzzatii*. Nenhum satDNA apresentou homologia com TEs.

Utilizando sequenciamento PacBio da população (A) de *D. borborema*, nós identificamos algumas cópias do satDNA pBuM2, o qual não foi identificado pela pipeline RepeatExplorer, provavelmente pelo baixo número de cópias.

- Determinar a localização cromossômica dos satDNAs identificados;

Como esperado, os satDNAs identificados foram localizados nas regiões heterocromáticas dos cromossomos de *D. gouveai* (Figura 20). O satDNA mais abundante desta espécie, o CDSTR230, foi localizado exclusivamente nos cromossomos sexuais da espécie, na região de beta-heterocromatina do X e em duas bandas no braço longo do Y. Ao contrário do que se esperava, ele não é o principal componente do DNA centromérico da espécie. Até o momento, considera-se que tanto o CDSTR138 quanto o novo satDNA descrito, CDSTR16, possam ser os principais componentes para o centrômero de *D. gouveai*.

- Analisar o padrão de evolução dos DNAs satélites encontrados e sua utilização como marcadores filogenéticos

Ao analisar o padrão de evolução dos satDNAs de *D. borborema* e *D. gouveai*, constatamos que em ambas espécies houve o padrão de evolução combinada. No entanto, ao comparar as quatro espécies *D. borborema*, *D. gouveai*, *D. seriema* e *D. serido* esse padrão não se manteve, provavelmente consequência da recente divergência entre estas espécies.

- Estudar a variação inter-populacional dos DNAs satélites em *D. borborema*;

Ao analisar as duas populações de *D. borborema*, verificamos que ambas possuem os mesmos satélites, com poucas diferenças na **proporção e abundância** de

cada um. Observamos também que as cópias dos diferentes satDNAs não apresentaram divergência inter-populacional.

REFERENCIAS

ADAMS, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. **Science**, v. 287, n. 5461, p. 2185-2195, 2000.

ALTEMOSE, N.; MIGA, K. H.; MAGGIONI, M.; WILLARD, H. F. Genomic characterization of large heterochromatic gaps in the human genome assembly. **PLoS computational biology**, v. 10, n. 5, p. e1003628, 2014.

ALTEMOSE, N. A classical revival: Human satellite DNAs enter the genomics era. In: Seminars in cell & developmental biology. **Academic Press**, p. 2-14. 2022.

ASHBURNER, M. *Drosophila*. A laboratory handbook. **Cold Spring Harbor Laboratory Press**, 1989.

BAIMAI, V. Chromosomal polymorphisms of constitutive heterochromatin and inversions in *Drosophila*. **Genetics**, 1977, v. 85, n. 1, p. 85-93.

BELLEN, H. J.; TONG, C.; TSUDA, H. 100 years of *Drosophila* research and its impact on vertebrate neuroscience: a history lesson for the future. **Nature Reviews Neuroscience**, v. 11, n. 7, p. 514-522, 2010.

BELYAYEV, A.; JOSEFIOVÁ, J.; JANDOVÁ, M.; KALENDAR, R.; KRAK, K.; MANDÁK, B. Natural History of a Satellite DNA Family: From the Ancestral Genome Component to Species-Specific Sequences, Concerted and Non-Concerted Evolution. **Int. J. Mol. Sci.**, 2019, 20, 1201.

BELYAYEV, A., JANDOVÁ, M., JOSEFIOVÁ, 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**, 2020, 15(10), e0241206.

BOSCO, G.; CAMPBELL, P.; LEIVA-NETO, J. T.; MARKOW, T. A. Analysis of *Drosophila* species genome size and satellite DNA content reveals significant differences among strains as well as between species. **Genetics**, v. 177, n. 3, p. 1277-1290. 2007.

BRACEWELL, R.; CHATLA, K.; NALLEY, M. J.; BACHTROG, D. Dynamic turnover of centromeres drives karyotype evolution in *Drosophila*. **Elife**, v. 8, p. e49002, 2019.

BROWN, J. D.; O'NEILL, R. J. The Evolution of Centromeric DNA Sequences. **eLS**, p. 1–11, 2014.

BUNTJER, J. B.; NIJMAN, I. J.; ZIJLSTRA, C.; LENSTRA, J. A. A satellite DNA element specific for roe deer (*Capreolus capreolus*). **Chromosoma**, v. 107, p. 1-5, 1998.

CAMACHO, J. P. M., CABRERO, J., LÓPEZ-LEÓN, M. D.; MARTÍN-PECIÑA, M., PERFECTI, F., GARRIDO-RAMOS, M. A., RUIZ-RUANO, F. J. Satellitome comparison of two *oedipodine grasshoppers* highlights the contingent nature of satellite DNA evolution. **BMC biology**, 20(1), 1-24. 2022.

CHARLESWORTH, B.; SNIEGOWSKI, P.; STEPHAN, W. The evolutionary dynamics of repetitive DNA in eukaryotes. **Nature**, 1994, v. 371, n. 6494, p. 215-220.

COBB, M. A gene mutation which changed animal behaviour: Margaret Bastock and the yellow fly. **Animal behaviour**, v. 74, n. 2, p. 163-169, 2007.

COMPEAU P E, PEVZNER P A, TESLER G. How to apply de Bruijn graphs to genome assembly. **Nat Biotechnol**. 2011 8;29(11):987-91

CRADDOCK, E. M.; GALL, J. G.; JONAS, M. Hawaiian *Drosophila* genomes: size variation and evolutionary expansions. **Genetica**, 2016, v. 144, p. 107-124.

DALAL, Y.; FURUYAMA, T.; VERMAAK, D.; HENIKOFF, S. Structure, dynamics, and evolution of centromeric nucleosomes. **Proceedings of the National Academy of Sciences**, v. 104, n. 41, p. 15974-15981, 2007.

DALTON, R. A fly by any other name: *Drosophila* experts argue over reclassification proposal. **Nature**, v. 457, n. 7228, p. 368-369, 2009.

DAWE, R. K.; HENIKOFF, S. Centromeres put epigenetics in the driver's seat. **Trends in Biochemical Sciences**, 2006, v. 31, n. 12, p. 662–669.

DIAS G.B., SVARTMAN M., DELPRAT A., RUIZ A., KUHN G.C. Tetris is a foldback transposon that provided the building blocks for an emerging satellite DNA of *Drosophila virilis*. **Genome Biol Evol**. 2014, May 24;6(6):1302-13.

DIAS, G. B., HERINGER, P., SVARTMAN, M., KUHN, G. C. Helitrons shaping the genomic architecture of *Drosophila*: enrichment of DINE-TR1 in α - and β -

heterochromatin, satellite DNA emergence, and piRNA expression. **Chromosome Research**, 2015, 23, 597-613.

EDGAR, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic acids research**, 2004, v. 32, n. 5, p. 1792-1797.

ELDER JR, J. F.; TURNER, B. J. Concerted evolution of repetitive DNA sequences in eukaryotes. **The Quarterly review of biology**, 1995, v. 70, n. 3, p. 297-320.

FERREIRA, R. F. **Filogenia do Complexo *Drosophila Buzzatii* (Grupo Repleta): Inferências de Análises Multilocus Mitocondriais e Nucleares**. 2011. Dissertação (Mestrado em Genética) - Faculdade de Medicina de Ribeirão Preto, University of São Paulo, Ribeirão Preto, 2011.

FERREE P.M, BARBASH D.A. Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. **PLoS Biol.** 2009 Oct;7(10):e1000234. 2009.

FINET, C.; KASSNER, V. A.; CARVALHO, A. B.; CHUNG, H.; DAY, J. P.; DAY, S.; DELANEY, E. K.; DE RÉ, F. C.; DUFOUR, H. D.; DUPIM, E.; IZUMITANI, H. F.; GAUTÉRIO, T. B.; JUSTEN, J.; KATOH, T.; KOPP, A.; KOSHIKAWA, S.; LONGDON, B.; LORETO, E. L.; NUNES, M. D. S.; RAJA, K. K. B.; REBEIZ, M.; RITCHIE, M. G.; SAAKYAN, G.; SNEDDON, T.; TERAMOTO, M.; TYUKMAEVA, V.; VANDERLINDE, T.; WEY, E. E.; WERNER, T.; WILLIAMS, T. M.; ROBE, L. J.; TODA, M. J.; MARLÉTAZ, F. Drosophyla: Resources for Drosophilid Phylogeny and Systematics. **Genome Biology and Evolution**, v. 13, n. 8, p. evab179, 2021

FRANCO, F. F.; KUHN, G. C. S.; DE SENE, F. M.; MANFRIN, M. H. Conservation of pBuM-2 satellite DNA sequences among geographically isolated *Drosophila gouveai* populations from Brazil. **Genetica**, v. 128, p. 287-295, 2006.

FRANCO, F. F.; MANFRIN, M. H. Recent demographic history of cactophilic *Drosophila* species can be related to Quaternary palaeoclimatic changes in South America. **Journal of Biogeography**, v. 40, n. 1, p. 142-154, 2013.

FU J, ZHANG H, GUO F, MA L, WU J, YUE M, ZHENG X, QIU Z, LI L. Identification and characterization of abundant repetitive sequences in *Allium cepa*. **Sci Rep.** 2019

GALL, J. G.; ATHERTON, D. D. Satellite DNA sequences in *Drosophila virilis*. **Journal of molecular biology**, v. 85, n. 4, p. 633-664, 1974.

GAO, J.; HU, Y.; TODA, M. J.; KATOH, T.; TAMURA, K. Phylogenetic relationships between *Sophophora* and *Lordiphosa*, with proposition of a hypothesis on the vicariant divergences of tropical lineages between the Old and New Worlds

in the family Drosophilidae. **Molecular Phylogenetics and Evolution**, v. 60, n. 1, p. 98-107, 2011.

GARRIDO-RAMOS, M.A. Satellite DNA: An Evolving Topic. **Genes**, 2017, 8, 230.

GENT, J. I.; SCHNEIDER, K. L.; TOPP, C. N.; RODRIGUEZ, C.; PRESTING, G. G.; DAWE, R. K. Distinct influences of tandem repeats and retrotransposons on CENH3 nucleosome positioning. *Epigenetics & chromatin*, v. 4, p. 1-12, 2011.

GREGORY, T. R.; JOHNSTON, J. S. Genome size diversity in the family Drosophilidae. **Heredity**, v. 101, n. 3, p. 228-238, 2008.

GUILLÉN Y, RIUS N, DELPRAT A, WILLIFORD A, MUYAS F, PUIG M, CASILLAS S, RÀMIA M, EGEA R, NEGRE B, MIR G, CAMPS J, MONCUNILL V, RUIZ-RUANO FJ, CABRERO J, DE LIMA LG, DIAS GB, RUIZ JC, KAPUSTA A, GARCIA-MAS J, GUT M, GUT IG, TORRENTS D, CAMACHO JP, KUHN GC, FESCHOTTE C, CLARK AG, BETRÁN E, BARBADILLA A, RUIZ A. Genomics of ecological adaptation in cactophilic *Drosophila*. **Genome Biol Evol**. 2014. Dec 31;7(1):349-66.

GUILLÉN, Y.; RIUS, N.; DELPRAT, A.; WILLIFORD, A.; MUYAS, F.; PUIG, M.; CASILLAS, S.; RÀMIA, M.; EGEA, R.; NEGRE, B.; MIR, G.; CAMPS, J.; MONCUNILL, V.; RUIZ-RUANO, F. J.; CABRERO, J.; DE LIMA, L. G.; DIAS, G. B.; RUIZ, J. C.; KAPUSTA, A.; GARCIA-MAS, J.; GUT, M.; GUT, I. G.; TORRENTS, D.; CAMACHO, J. P.; KUHN, G. C. S.; FESCHOTTE, C.; CLARK, A. G.; BETRÁN, E.; BARBADILLA, A.; RUIZ, A. Genomics of ecological adaptation in cactophilic *Drosophila*. **Genome Biology and Evolution**, v. 7, n. 1, p. 349-366, 2015.

HENIKOFF, S.; AHMAD, K.; MALIK, H. S. The centromere paradox: stable inheritance with rapidly evolving DNA. **Science**, 2001, v. 293, n. 5532, p. 1098-1102.

HURTADO, J.; ALMEIDA, F.; REVALE, S.; HASSON, E. Hurtado, J., Almeida, F., Revale, S., & Hasson, E. Revised phylogenetic relationships within the *Drosophila buzzatii* species cluster (Diptera: Drosophilidae: *Drosophila repleta* group) using genomic. **Systematics & Phylogeny**, 2019, v. 77, p.239-250.

JENNINGS, B. H. *Drosophila*—a versatile model in biology & medicine. **Materials today**, v. 14, n. 5, p. 190-195, 2011.

JOHN, B.; MIKLOS, G. L. G. Functional aspects of satellite DNA and heterochromatin. **International review of cytology**, 1979, v. 58, p. 1-114.

KIT, S Equilibrium sedimentation in density gradients of DNA preparations from animal tissues. **Journal of molecular biology**, 1961, v. 3, n. 6, p. 711-IN2.

KIM, B. Y.; WANG, J. R.; MILLER, D. E.; BARMINA, O.; DELANEY, E.; THOMPSON, A.; COMEAULT, A. A.; PEDE, D.; D'AGOSTINO, E. R. R.; PELAEZ, J.; AGUILAR, J. M.; HAJI, D.; MATSUNAGA, T.; ARMSTRONG, E. E.; ZYCH, M.; OGAWA, Y.; STAMENKOVIĆ-RADAK, M.; JELIĆ, M.; SAVIĆ VESSELINOVIĆ, M.; TANASKOVIĆ, M.; ERIĆ, P.; GAO, J. J.; KATOH, T. K.; TODA, M. J.; WATABE, H.; WATADA, M.; DAVIS, J. S.; MOYLE, L. C.; MANOLI, G.; BERTOLINI, E.; KOŠŤÁL, V.; HAWLEY, R. S.; TAKAHASHI, A.; JONES, C. D.; PRICE, D. K.; WHITEMAN, N.; KOPP, A.; MATUTE, D. R.; PETROV, D. A. Highly contiguous assemblies of 101 drosophilid genomes. **Elife**, v. 10, p. e66405, 2021.

KIM, B. Y.; GELLERT, H. R.; CHURCH, S. H.; SUVOROV, A.; ANDERSON, S. S.; BARMINA, O.; BESKID, S. G.; COMEAULT, A. A.; CROWN, K. N.; DIAMOND, S. E.; DORUS, S.; FUJICHICA, T.; HEMKER, J. A.; HRCEK, J.; KANKARE, M.; KATOH, T.; MAGNACCA, K. N.; MARTIN, R. A.; MATSUNAGA, T.; MEDEIROS, M. J.; MILLER, D. E.; PITNICK, S.; SCHIFFER, M.; SIMONI, S.; STEENWINKEL, T. E.; SYED, Z. A.; TAKAHASHI, A.; WEI, K. H. C.; YOKOYAMA, T.; EISEN, M. B.; KOPP, A.; MATUTE, D.; OBBARD, D. J.; O'GRADY, P. M.; PRICE, D. K.; TODA, M. J.; WERNER, T.; PETROV, D. A. Single-fly genome assemblies fill major phylogenomic gaps across the Drosophilidae Tree of Life. **PLoS Biology**, v. 22, n. 7, p. e3002697, 2024.

KROUPIN, P.; KUZNETSOVA, V.; ROMANOV, D.; KOCHESHKOVA, A.; KARLOV, G.; DANG, T. X.; KHUAT, T. M. L.; KIROV, I.; ALEXANDROV, O.; POLKHOVSKIY, A.; RAZUMOVA, O.; DIVASHUK, M. Pipeline for the rapid development of cytogenetic markers using genomic data of related species. **Genes**, 2019, v. 10, n. 2, p. 113.

KUHN G C S, RUIZ A, ALVES M A R, SENE F M The metaphase and polytene chromosomes of *Drosophila seriema* (*repleta* group; *mulleri* subgroup). *Brazil J Genet* 19:209–216, 1996.

KUHN, G. C. S., S. BOLLGÖNN, D. SPERLICH, AND L. BACHMANN. Characterization of a species-specific satellite DNA of *Drosophila buzzatii*. **J. Zoological Syst. Evol. Res.** 1999, 37(2): 109–112.

KUHN G.C, SENE F.M. Evolutionary turnover of two pBuM satellite DNA subfamilies in the *Drosophila buzzatii* species cluster (*repleta* group): from alpha to alpha/beta arrays. **Gene**. 2005, Apr 11;349:77-85.

KUHN G.C., FRANCO F.F., MANFRIN M.H., MOREIRA-FILHO O., SENE F.M. Low rates of homogenization of the DBC-150 satellite DNA family restricted to a single pair of microchromosomes in species from the *Drosophila buzzatii* cluster. **Chromosome Res.** 2007, 15(4):457-69.

KUHN, G. C., SENE, F. M., MOREIRA-FILHO, O., SCHWARZACHER, T., HESLOP-HARRISON, J. S. Sequence analysis, chromosomal distribution and long-range organization show that rapid turnover of new and old pBuM satellite DNA repeats leads to different patterns of variation in seven species of the *Drosophila buzzatii* cluster. **Chromosome Research**, 2008, 16, 307-324.

KUHN, G.C.S.; KÜTTLER, 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. **Mol. Biol. Evol.** 2012, 29, 7–11.

KUHN G C S, HERINGER P, DIAS G B Structure, Organization, and Evolution of Satellite DNAs: Insights from the *Drosophila repleta* and *D. virilis* Species Groups. **Prog Mol Subcell Biol.** 60:27-56. (2021).

LABORNE, A. M., BARRIOS-LEAL, D. Y., HESLOP-HARRISON, J. S., MANFRIN, M. H., & KUHN, G. Genome location, evolution and centromeric contribution of satellite DNAs shared between the two closely related species *Drosophila serido* and *D. antonietae* (*repleta* group, *buzzatii* cluster). **Chromosoma**, 134(1), 1-15. 2025.

LANDER, E S. *et al.* Initial sequencing and analysis of the human genome. 2001.

LEVY, S.; SUTTON, G.; NG, P. C.; FEUK, L.; HALPERN, A. L.; WALENZ, B. P.; AXELROD, N.; HUANG, J.; KIRKNESS, E. F.; DENISOV, G.; LIN, Y.; MACDONALD, J. R.; PANG, A. W. C.; SHAGO, M.; STOCKWELL, T. B.; TSIAMOURI, A.; BAFNA, V.; BANSAL, V.; KRAVITZ, S. A.; BUSAM, D. A.; BEESON, K. Y.; MCINTOSH, T. C.; REMINGTON, K. A.; ABRIL, J. F.; GILL, J.; BORMAN, J.; ROGERS, Y. H.; FRAZIER, M. E.; SCHERER, S. W.; STRAUSBERG, R. L.; VENTER, J. C. The diploid genome sequence of an individual human. **PLoS biology**, v. 5, n. 10, p. e254, 2007.

DE LIMA, L. G.; SVARTMAN, M.; KUHN, G.C.S. Dissecting the satellite DNA landscape in three cactophilic *Drosophila* sequenced genomes. **G3: Genes, Genomes, Genetics**, 2017. v. 7, n. 8, p. 2831-2843.

LIN, C. C.; LI, Y. C. Chromosomal distribution and organization of three cervid satellite DNAs in Chinese water deer (*Hydropotes inermis*). **Cytogenetic and genome research**, v. 114, n. 2, p. 147-154, 2006.

LOHE, A. R.; BRUTLAG, D. L. Multiplicity of satellite DNA sequences in *Drosophila melanogaster*. **Proceedings of the National Academy of Sciences**, v. 83, n. 3, p. 696-700, 1986.

MANFRIN M H, BRITO R O A, SENE F M. Systematics and evolution of the *Drosophila buzzatii* (Diptera; Drosophilidae) cluster using mtDNA. *Ann Entomol Soc Am* 94: 333–346. 2001.

MANFRIN M.H; SENE F.M. Cactophilic *Drosophila* in South America: a model for evolutionary studies. **Genetica**. 2006 Jan;126(1-2):57-75.

MARKOW, T A.; O'GRADY, P. *Drosophila*: a guide to species identification and use. **Elsevier**, 2005.

MARKOW, T. A., O'GRADY, P. M. Evolutionary genetics of reproductive behavior in *Drosophila*: connecting the dots. *Annu. Rev. Genet.*, v. 39, n. 1, p. 263-291, 2005.

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.; GARCIA, J. F.; DERISI, J. L.; SMITH, T.; TOBIAS, C.; ROSS-IBARRA, J.; KORF, I.; CHAN, S. W. L. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. **Genome biology**, v. 14, p. 1-20, 2013.

MESTROVIC, N., PLOHL, M., MRAVINAC, B., UGARKOVIC, D. Evolution of satellite DNAs from the genus *Palorus*--experimental evidence for the "library" hypothesis. **Molecular biology and evolution**, 1998, 15(8), 1062-1068.

MIGA, K. H. Completing the human genome: the progress and challenge of satellite DNA assembly. **Chromosome Research**, v. 23, p. 421-426, 2015.

MERALDI, P., McAINSH, A. D., RHEINBAY, E., SORGER, P. K. Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins. **Genome biology**, v. 7, p. 1-21, 2006.

MOREYRA NN, ALMEIDA FC, ALLAN C, FRANKEL N, MATZKIN LM, HASSON E. Phylogenomics provides insights into the evolution of cactophily and host plant shifts in *Drosophila*. **Mol Phylogenet Evol**. 2023.

MORGAN, T. H. Sex limited inheritance in *Drosophila*. **Science**, v. 32, n. 812, p. 120-122, 1910.

NASCIMENTO, Ana Mattioli Laborne do. **Caracterização e evolução de DNAs satélites em *Drosophila antonietae* e *D. serido* (grupo *repleta*, *cluster buzzatii*) e em genomas da zona híbrida**. 2023. Dissertação (Mestrado em Genética) - Programa de Pós-Graduação em Genética, Universidade Federal de Minas Gerais, Belo Horizonte, 2023.

NOVÁK P, ÁVILA ROBLLEDILLO L, KOBLÍŽKOVÁ A, VRBOVÁ I, NEUMANN P, MACAS J. TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. **Nucleic Acids Res.** 2017.

NOVÁK P, NEUMANN P, MACAS J. Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. **Nat Protoc.** 2020.

O'GRADY, P. M.; DESALLE, R. Phylogeny of the genus *Drosophila*. **Genetics**, v. 209, n. 1, p. 1-25, 2018.

PATTERSON, J. T. The Drosophilidae of the southwest. **The University of Texas Publication**, v. 4313, p. 7, 1943.

PEREIRA, M A Q R; VILELA, C R; SENE, F M. Notes on breeding and feeding sites of some species of the *Repleta* group of the genus *Drosophila* (Diptera, *Drosophilidae*). **Ciência e Cultura**, 1983, São Paulo, v. 35, n. 9. p. 1313-1319.

PERVEEN, F. K. Introduction to *Drosophila*. ***Drosophila melanogaster: model for recent advances in genetics and therapeutics***, p. 3, 2018.

PLOHL, M.; MEŠTROVIĆ, N.; MRAVINAC, B. Satellite DNA evolution. **Repetitive DNA**, 2012, v. 7, p. 126-152.

PRYTKOVA, T. R., ZHU, X., WIDOM, J., SCHATZ, G. Modeling DNA-bending in the nucleosome: role of AA periodicity. **The Journal of Physical Chemistry B**, v. 115, n. 26, p. 8638-8644, 2011.

PODGORNAYA, O. I. Nuclear organization by satellite DNA, SAF-A/hnRNPU and matrix attachment regions. In: *Seminars in Cell & Developmental Biology*. **Academic Press**, 2022. p. 61-68.

PONS, J., BUCUR, R., VOGLER, A. P. Higher-order repeats in the satellite DNA of the cave beetle *Pholeuon proserpinae glaciale* (Coleoptera: Cholevidae). **Hereditas**, v. 139, n. 1, p. 28-34, 2003.

POWELL, J. R., DeSalle, R. *Drosophila* molecular phylogenies and their uses. **Evolutionary biology**, 28, 87-87, (1995)

POWELL, J. R. Progress and prospects in evolutionary biology: the *Drosophila* model. 1997.

RUIZ, A., HEED, W.B. Host plant specificity in the cactophilic *Drosophila mulleri* species complex. **J. Anim. Ecol.** 57, 237–249. 1988

RUIZ-RUANO, F. J.; NAVARRO-DOMÍNGUEZ, B.; CAMACHO, J. P. M.; GARRIDO-RAMOS, M. A. Characterization of the satellitome in lower vascular

plants: The case of the endangered fern *Vandenboschia speciosa*. **Annals of botany**, 2019, 123(4), 587-599.

RUDD, M. K., WRAY, G. A., WILLARD, H. F. The evolutionary dynamics of α -satellite. **Genome research**, v. 16, n. 1, p. 88-96, 2006.

ŠATOVIĆ-VUKŠIĆ, E., PLOHL, M. Satellite DNAs—From localized to highly dispersed genome components. **Genes**, v. 14, n. 3, p. 742, 2023.

SILVA BSML, HERINGER P, DIAS GB, SVARTMAN M, KUHN GCS. *De novo* identification of satellite DNAs in the sequenced genomes of *Drosophila virilis* and *D. americana* using the RepeatExplorer and TAREAN pipelines. **PLoS One**. 19;14(12). (2019).

SILVA, Bráulio Soares Macedo Leão. **Identificação de novo e citogenômica comparativa de DNAs satélites em espécies de *Drosophila* dos grupos *virilis* e *montium***. 2021. Dissertação (Mestrado em Genética) - Programa de Pós-Graduação em Genética, Universidade Federal de Minas Gerais, Belo Horizonte, 2021.

SILVA, BSML; PICORELLI, ACR; KUHN, GCS. In Silico Identification and Characterization of Satellite DNAs in 23 *Drosophila* Species from the *Montium* Group. **Genes**, v. 14, n. 2, p. 300, 2023.

SINGH, N. D., LARRACUENTE, A. M., SACKTON, T. B., CLARK, A. G. Comparative genomics on the *Drosophila* phylogenetic tree. **Annual Review of Ecology, Evolution, and Systematics**, v. 40, n. 1, p. 459-480, 2009.

SHANG, W., HORI, T., TOYODA, A., KATO, J., POPENDORF, K., SAKAKIBARA, Y., FUJIYAMA, A., FUKAGAWA, T. Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. **Genome research**, v. 20, n. 9, p. 1219-1228, 2010.

SPROUL, J. S.; KHOST, D. E.; EICKBUSH, D. G.; NEGM, S.; WEI, X.; WONG, I.; LARRACUENTE, A. M. Dynamic evolution of euchromatic satellites on the X chromosome in *Drosophila melanogaster* and the *simulans* clade. **Molecular biology and evolution**, 2020, v. 37, n. 8, p. 2241-2256.

STÖVER BC, MÜLLER KF. 2010. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. **BMC Bioinf.** 11:7

STRUHL, K., SEGAL, E. Determinants of nucleosome positioning. **Nature structural & molecular biology**, v. 20, n. 3, p. 267-273, 2013.

SUN, F. L.; CUAYCONG, M. H.; CRAIG, C. A.; WALLRATH, L. L.; LOCKE, J.; ELGIN, S. C. The fourth chromosome of *Drosophila melanogaster*: interspersed

euchromatic and heterochromatic domains. **Proceedings of the National Academy of Sciences**, v. 97, n. 10, p. 5340-5345, 2000.

TAMURA, K., SUBRAMANIAN, S., KUMAR, S. Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. **Mol. Biol. Evol.** 21, 36–44. 2004

TAMURA, K.; STECHER, G.; KUMAR, S. MEGA11: molecular evolutionary genetics analysis version 11. **Molecular biology and evolution**, 2021, v. 38, n. 7, p. 3022-3027.

TAUTZ, D. Notes on the definition and nomenclature of tandemly repetitive DNA sequences. **DNA fingerprinting: State of the science**, 1993, p. 21-28.

TEIXEIRA JR, DIAS GB, SVARTMAN M, RUIZ A, KUHN GCS. Concurrent Duplication of *Drosophila* Cid and Cenp-C Genes Resulted in Accelerated Evolution and Male Germline-Biased Expression of the New Copies. **J Mol Evol.** 2018, Jul;86(6):353-364.

THAKUR, J.; PACKIARAJ, J.; HENIKOFF, S. Sequence, chromatin and evolution of satellite DNA. **International journal of molecular sciences**, 2021, v. 22, n. 9, p. 4309.

TRINDER, M., DAISLEY, B. A., DUBE, J. S., REID, G. *Drosophila melanogaster* as a high-throughput model for host–microbiota interactions. **Frontiers in microbiology**, v. 8, p. 751, 2017.

TREANGEN, T. J.; SALZBERG, S. L. Repetitive DNA and next-generation sequencing: computational challenges and solutions. **Nature Reviews Genetics**, v. 13, n. 1, p. 36-46, 2012.

UGUR, B., CHEN, K., BELLEN, H. J. *Drosophila* tools and assays for the study of human diseases. **Disease models & mechanisms**, v. 9, n. 3, p. 235-244, 2016.

VENTER, J. C *et al.* The sequence of the human genome. **Science**, v. 291, n. 5507, p. 1304-1351, 2001

WASSERMAN, M. Cytological studies of the repleta group. **Univ. Texas Publ**, v. 5422, p. 130-152, 1954.

WHARTON, L. T. Interspecific hybridization in the repleta group. **Univ. Texas Publ.**, v. 4445, p. 175-193, 1944.

WILLARD, H. F.; WAYE, J. S. Hierarchical order in chromosome-specific human alpha satellite DNA. **Trends in Genetics**, v. 3, p. 192-198, 1987.

ZIMMER E.A.; MARTIN S.L.; BEVERLEY S.M.; KAN Y.W.; WILSON A.C. Rapid duplication and loss of genes coding for the alpha chains of hemoglobin. **Proceedings of the National Academy of Sciences**, 1980, v. 77, n. 4, p. 2158-2162.

ZHANG, P.; WU, W.; CHEN, Q.; CHEN, M. Non-Coding RNAs and their Integrated Networks. **Journal of Integrative Bioinformatics**, v. 16, n. 3, 13 jul. 2019.

ANEXOS

D. borborema A

sequencias com 1 monomero:

CATCAATA ou **CATCGATA**

Sequencias com 2 monomeros:

CATCGATA e **CATCAATA** ou **CATCAATA** e **CACCAATA**

Sequencias com 3 monomeros:

CATCGATA **CATCAATA** **CACCAATA**

CATCGATA **CATCAATA** **ATTGATA**

Figura Suplementar 1 - Variações das sequências dos quatro monômeros encontrados no satélite CDSTR16 nas linhagens de *D. borborema A*.

D. borborema B

1 monomero:

CATCGATA ou CATCAATA

2 monomeros:

CATCAATA e CATCGATA

CATCAATA e CACCAATA

CATCAATA e CATCTATA

3 monomeros:

CATCAATA CATCGATA e CTTCGATA

4 monomeros:

CATCAATA CATCGATA CTTCGATA e CCTCGATA

5 monomeros:

CATCAATA CATCGATA CTTCGATA CCTCGATA e CATCTATA

Figura Suplementar 2 - Variações das sequências dos seis monômeros encontrados no satélite CDSTR16 nas linhagens de *D. borborema* B .

Table S1. Comparação dos clusters recuperados a partir de três runs de *D. borborema*(A) pelo RepeatExplorer2. As famílias de satDNA analisadas neste estudo estão destacadas em roxo., *minisatélite

Run1

Cluster	Prop. (%)	Proportion adjusted (%)	N reads	Satellite Probability	Consensus length (pb)	C index	P index	Similarity hits
1	1.800	1,792	72635	0,979	16	0,999	0,992	CDSTR16
17	0,27	0,269	10925	0,815	208	0,944	0,907	CDSTR198*
20	0,26	0,262	10621	0,994	245	0,991	0,969	CDSTR230
32	0,18	0,177	7189	0,994	138	0,967	0,574	CDSTR138
40	0,13	0,133	5387	0,985	112	0,977	0,558	DBC-150

Run2

1	1.800	1,803	73539	0,979	16	0,999	0,992	CDSTR16
---	-------	-------	-------	-------	----	-------	-------	---------

2	0.650	0,667	37599	0.985	24	0.997	0.983	CDSTR16
8	0.340	0,349	19702	0.986	245	0.996	0.975	CDSTR230
24	0.210	0,214	12085	0.994	138	0.995	0.967	CDSTR138
26	0.210	0,211	11918	0.915	86	0.998	0.920	SAT109
29	0.200	0,206	11640	0.994	112	0.994	0.967	DBC150
38	0.180	0,182	10262	0.941	178	0.961	0.907	CDSTR198*

Table S3. Comparação dos clusters recuperados a partir de três runs de *D. gouveai*, pelo RepeatExplorer2. As famílias de satDNA analisadas neste estudo estão destacadas em roxo., *minisatélite.

Run1

Cluster	Prop. (%)	Proportion adjusted (%)	N reads	Satellite Probability	Consensus length (pb)	C index	P index	Similarity hits
1	0,57	0,578	23859	0,992	230	0,986	0,649	CDSTR230
5	0,47	0,469	19357	9,55E-01	86	0,937	0,609	SAT109
6	0,47	0,468	19341	9,81E-01	136	0,973	0,619	CDSTR138
8	0,44	0,442	18251	9,33E-01	24	0,971	0,717	CDSTR16
9	0,41	0,412	16992	9,92E-01	370	0,985	0,835	pBuM2
30	0,21	0,216	8923	0,933	149	0,973	0,553	DBC-150
100	0,063	0,063	2617	6,61E-02	188	0,821	0,857	CDSTR198*

Run2

2	0.590	0,604	25505	0.992	230	0.988	0.982	CDSTR230
3	0.540	0,556	23457	0.992	136	0.988	0.977	CDSTR138
6	0.410	0,422	17821	0.955	86	0.977	0.935	SAT109
8	0.400	0,407	17171	0.986	370	0.990	0.985	pBuM2
9	0.400	0,406	17141	0.974	24	0.979	0.977	CDSTR16
29	0.200	0,210	8847	0.981	149	0.977	0.973	DBC-150
82	0.077	0,079	3341	0.0701	188	0.853	0.864	CDSTR198*

Run3

2	0.550	0,567	24242	0.976	230	0.987	0.942	CDSTR230
3	0.540	0,556	23764	0.992	136	0.987	0.977	CDSTR138
5	0.420	0,425	18172	0.819	86	0.969	0.932	SAT109

7	0.400	0,406	17379	0.974	24	0.980	0.977	CDSTR16
8	0.400	0,406	17359	0.986	370	0.991	0.985	pBuM2
28	0.200	0,209	8953	0.981	149	0.977	0.973	DBC-150
82	0.077	0,079	3381	0.06800	188	0.830	0.864	CDSTR198*

Table S4. Description of all clusters retrieved from reads of *D. borborema*(A) by RepeatExplorer2. The satDNA families analyzed in this study are highlighted in purple.

N input reads	111976370	N nuclear reads	4054375
N analyzed reads	4074818	Nn = Na - Nc - No	
N contamination reads	0		
N organelle reads	20443		

Cluster	Proportion (%)	Proportion adjusted (%)	N reads	Satellite Probability	Consensus length (pb)	C index	P index	Similarity hits
1	1.800	1,792	72635	0,979	16	0,999	0,992	CDSTR16
2	0,64	0,641	26006	4,14e-07		0,504	0,079	unknow
3	0,46	0,462	18723	1,60E-08		0,392	0,15	unknow
4	0,43	0,428	17352	3,16E-13		0,184	0,227	unknow
5	0,43	0,428	17334	1,72E-06		0,904	0,114	unknow
6	0,42	0,426	17270	1,43E-13		0,193	0,661	unknow
7	0,34	0,337	13671	2,35E-03		0,653	0,409	unknow
8	0,33	0,334	13539	1,79E-06		0,787	0,0699	unknow
9	0,32	0,320	12956	1,01E-16		0,115	0,701	27.90% Class_I/LI NE:LINE- RT
10	0,29	0,296	12008	2,13E-11		0,257	0,755	4.11% Class_I/LI NE:LINE- RT
11	0,29	0,291	11793	8,35E-20		0,0552	0,858	16.44% Class_I/LT R/Bel- Pao:BEL-

								GAG
12	0,29	0,287	11645	3,08E-05		0,618	0,147	unknow
13	0,29	0,286	11615	8,65E-04		0,933	0,387	unknow
14	0,28	0,286	11588	7,54E-08		0,362	0,479	unknow
15	0,28	0,282	11442	7,61E-24		0,000265	0,865	27.12% Class_I/LT R/Bel- Pao:BEL- RT
16	0,27	0,276	11177	3,18E-13		0,174	0,48	unknow
17	0,27	0,269	10925	0,815	208	0,944	0,907	CDSTR198
18	0,27	0,269	10902	0,0443	4126	0,77	0,852	13.46% Class_I/LI NE:LINE- RT
19	0,26	0,266	10792	9,26E-14		0,152	0,361	unknow
20	0,26	0,262	10621	0,994	245	0,991	0,969	CDSTR230
21	0,26	0,261	10574	1,30E-22		0,00681	0,866	16.64% Class_I/LT R/Bel- Pao:BEL- RT
22	0,24	0,246	9973	7,58E-09		0,34	0,302	unknow
23	0,21	0,211	8568	0,0588	6918	0,899	0,766	5.84% Class_I/LT R/Ty3_gyp sy:Ty3-RT
24	0,2	0,202	8197	1,88E-13		0,158	0,36	unknow
25	0,2	0,202	8188	2,29E-22		0,000977	0,676	19.72% Class_I/LI NE:LINE- ENDO
26	0,2	0,201	8132	1,17E-22		0,00221	0,718	6.78% Class_I/LT R/Ty3_gyp sy:Ty3-INT
27	0,2	0,200	8092	9,66E-25		0	0,947	IGS
28	0,19	0,196	7940	0,0285	909	0,78	0,6	unknow
29	0,19	0,192	7778	5,69E-23		0,00489	0,758	7.75% Class_I/LT R/Bel- Pao:BEL- GAG
30	0,19	0,190	7699	0,059	6337	0,88	0,79	6.77% Class_I/LT R/Ty3_gyp

								sy:Ty3-RT
31	0,18	0,182	7369	2,14E-18		0,0863	0,845	10.58% Class_I/LT R/Ty3_gyp sy:Ty3-INT
32	0,18	0,177	7189	0,994	138	0,967	0,574	CDSTR138
33	0,17	0,169	6840	1,76E-23		0,00132	0,821	46.37% Class_I/LI NE:LINE- RT
34	0,16	0,160	6501	0,0666	2614	0,872	0,826	unknow
35	0,16	0,158	6406	3,18E-21		0,00516	0,618	6.26% Class_I/LT R/Bel- Pao:BEL- INT
36	0,15	0,154	6264	1,97E-24		0,00064	0,916	20.96% organelle/m itochondria
37	0,15	0,154	6239	0,624	6166	0,966	0,827	6.81% Class_I/LT R/Ty3_gyp sy:Ty3-RT
38	0,14	0,144	5831	1,51E-21		0,159	0,855	8.20% Class_I/LT R/Ty3_gyp sy:Ty3-RT
39	0,14	0,142	5744	2,14E-03		0,631	0,702	unknow
40	0,13	0,133	5387	0,985	112	0,977	0,558	DBC-150
41	0,13	0,129	5219	0,0453	3319	0,912	0,739	5.42% Class_I/LT R/Ty1_copi a:Ty1-GAG
42	0,12	0,125	5075	1,76E-23		0,00138	0,82	41.83% Class_I/LI NE:LINE- ENDO
43	0,12	0,124	5009	4,91E-24		0,0002	0,881	41.11% 45S_rDNA/ 25S_rDNA
44	0,12	0,123	4993	0,78	5974	0,957	0,867	10.92% Class_I/LT R/Bel- Pao:BEL- INT
45	0,12	0,119	4810	1,11E-06		0,439	0,738	10.23% Class_I/LT R/Ty3_gyp sy:Ty3-RT

46	0,12	0,116	4689	1,44E-23		0,000854	0,826	47.34% organelle/pl astid
47	0,11	0,114	4640	3,18E-21		0,0104	0,615	26.44% Class_I/LT R/Bel- Pao:BEL- INT
48	0,11	0,113	4595	0,0678	226	0,809	0,883	IGS
49	0,11	0,111	4520	8,21E-23		0,000664	0,74	20.15% Class_II/Su bclass_2/He litron:Helitr on-HEL2
50	0,11	0,111	4514	1,64E-12		0,232	0,0799	unknow
51	0,11	0,110	4476	1,93E-08		0,374	0,806	1.01% Class_I/LI NE:LINE- RT
52	0,11	0,110	4474	7,09E-21		0,0438	0,87	16.45% Class_II/Su bclass_2/He litron:Helitr on-HEL1
53	0,11	0,106	4298	7,61E-24		0,000237	0,863	5.14% organelle/m itochondria
54	0,1	0,103	4162	5,24E-17		0,101	0,689	unknow
55	0,1	0,101	4109	0,014	6393	0,708	0,764	6.94% Class_I/LT R/Bel- Pao:BEL- RT
56	0,094	0,094	3825	5,30E-07		0,415	0,641	3.79% Class_I/LT R/Ty3_gyp sy:Ty3-RT
57	0,094	0,094	3823	0,975	3629	0,947	0,865	unknow
58	0,09	0,090	3661	7,61E-24		0,000273	0,858	38.00% 45S_rDNA/ 18S_rDNA
59	0,088	0,088	3573	7,33E-11		0,281	0,794	unknow
60	0,087	0,088	3562	0,0546	7088	0,832	0,789	7.19% Class_I/LT R/Ty3_gyp sy:Ty3-RT
61	0,085	0,085	3456	9,83E-03		0,681	0,629	4.31% Class_I/LT R/Ty3_gyp

								sy:Ty3-RH
62	0,085	0,085	3455	0,372	4145	0,921	0,829	16.67% Class_I/LT R/Ty1_copi a:Ty1-RT
63	0,083	0,083	3364	4,49E-04		0,565	0,538	unknow
64	0,082	0,082	3342	3,80E-22		0,00898	0,804	5.69% Class_I/LT R/Ty3_gyp sy:Ty3-RT
65	0,082	0,082	3340	5,34E-04		0,586	0,741	unknow
66	0,082	0,082	3324	4,69E-21		0,00361	0,436	unknow
67	0,08	0,080	3248	7,67E-18		0,105	0,818	3.11% Class_I/LT R/Ty3_gyp sy:Ty3-RT
68	0,075	0,075	3042	1,19E-20		0,0253	0,768	14.30% Class_I/LT R/Bel- Pao:BEL- GAG
69	0,074	0,074	3006	5,69E-23		0,000333	0,759	11.31% organelle/m itochondria
70	0,072	0,073	2947	1,05E-08		0,33	0,595	unknow
71	0,072	0,072	2924	2,58E-05		0,504	0,667	unknow
72	0,067	0,067	2730	2,68E-17		0,0875	0,678	unknow
73	0,065	0,066	2658	6,18E-05		0,528	0,729	10.16% Class_I/LT R/Ty1_copi a:Ty1-RT
74	0,064	0,065	2616	6,64E-22		0,00191	0,609	1.91% Class_I/LT R/Ty1_copi a:Ty1-INT
75	0,064	0,064	2603	0,708	6249	0,964	0,838	7.30% Class_I/LT R/Bel- Pao:BEL- INT
76	0,061	0,061	2476	1,08E-20		0,00808	0,252	1.45% Class_I/LT R/Ty3_gyp sy:Ty3- GAG
77	0,06	0,060	2452	0,992	255	0,981	0,807	unknow
78	0,06	0,060	2427	2,51E-03		0,658	0,4	unknow

79	0,057	0,057	2303	6,95E-21	0,00782	0,14	unknow
80	0,055	0,055	2250	1,47E-21	0,000449	0,554	12.76% Class_I/LI NE:LINE- RT
81	0,055	0,055	2227	2,31E-11	0,27	0,883	unknow
82	0,053	0,053	2169	2,52E-20	0,0171	0,658	11.53% Class_I/LI NE:LINE- RT
83	0,051	0,051	2078	3,16E-22	0,00311	0,661	unknow
84	0,05	0,050	2043	5,12E-06	0,475	0,852	7.88% Class_I/LT R/Ty1_copi a:Ty1-RT
85	0,049	0,050	2010	3,21E-23	0,00149	0,791	unknow
86	0,049	0,049	1980	1,96E-18	0,0519	0,479	unknow
87	0,046	0,047	1890	1,49E-03	0,625	0,724	unknow
88	0,045	0,045	1823	5,55E-19	0,038	0,505	unknow
89	0,045	0,045	1817	8,45E-20	0,0282	0,624	unknow
90	0,043	0,043	1751	2,70E-22	0,00171	0,668	unknow
91	0,042	0,042	1700	1,40E-19	0,0341	0,57	unknow
92	0,041	0,041	1665	4,82E-18	0,0611	0,545	unknow
93	0,04	0,040	1626	4,67E-19	0,059	0,735	5.41% organelle/m itochondria
94	0,038	0,039	1568	1,38E-19	0,0172	0,49	unknow
95	0,038	0,038	1557	2,38E-14	0,314	0,431	unknow
96	0,038	0,038	1538	5,86E-11	0,271	0,722	2.67% Class_I/LT R/Ty3_gyp sy:Ty3- PROT
97	0,037	0,037	1503	1,26E-06	0,44	0,718	9.31% Class_I/LT R/Ty3_gyp sy:Ty3-INT
98	0,037	0,037	1491	3,91E-21	0,000674	0,465	3.35% Class_I/LI NE:LINE- RT
99	0,037	0,037	1489	1,86E-04	0,558	0,744	8.60% Class_I/LI NE:LINE-

							RT	
100	0,035	0,036	1446	7,88E-21		0,0124	0,494	unknow
101	0,035	0,036	1446	1,00E-03		0,608	0,88	unknow
102	0,035	0,036	1442	7,64E-22		0,00208	0,595	14.15% Class_I/LI NE:LINE- RT
103	0,035	0,035	1412	1,23E-24		0,000708	0,937	9.99% Class_I/LI NE:LINE- ENDO
104	0,034	0,034	1392	1,94E-20		0,02	0,679	9.34% Class_I/LT R/Ty3_gyp sy:Ty3-RT
105	0,034	0,034	1386	6,94E-20		0,0153	0,566	unknow
106	0,034	0,034	1379	0,0723	69	0,893	0,805	unknow
107	0,033	0,033	1346	6,64E-18		0,0609	0,494	unknow
108	0,033	0,033	1335	5,60E-18		0,0728	0,654	unknow
109	0,032	0,032	1314	3,89E-19		0,0327	0,0674	unknow
110	0,032	0,032	1307	1,54E-20		0,0023	0,219	unknow
111	0,031	0,031	1268	5,17E-19		0,0386	0,517	unknow
112	0,031	0,031	1255	1,33E-09		0,331	0,907	unknow
113	0,031	0,031	1244	1,39E-16		0,109	0,681	unknow
114	0,03	0,030	1233	4,80E-21		0,00579	0,573	unknow
115	0,03	0,030	1215	1,41E-12		0,231	0,782	unknow
116	0,028	0,028	1149	1,18E-06		0,415	0,605	unknow
117	0,028	0,028	1131	9,43E-24		0,000884	0,848	16.36% Class_I/LT R/Ty3_gyp sy:Ty3-INT
118	0,027	0,027	1111	2,70E-22		0,00466	0,673	unknow
119	0,027	0,027	1094	4,98E-22		0,000941	0,633	unknow
120	0,026	0,026	1046	6,84E-23		0,000956	0,749	unknow
121	0,025	0,025	1024	1,40E-21		0,00745	0,701	unknow
122	0,025	0,025	1011	2,90E-21		0,00396	0,491	unknow
123	0,024	0,024	982	9,81E-23		0,00311	0,732	unknow
124	0,022	0,022	911	5,69E-23		0,0011	0,755	5.93% Class_I/LT R/Ty3_gyp

								sy:Ty3-RH
125	0,021	0,021	863	1,62E-13		0,176	0,586	unknow
126	0,021	0,021	844	2,70E-22		0,00118	0,671	unknow
127	0,021	0,021	844	0,0225	2233	0,851	0,602	2.73% Class_I/LT R/Bel- Pao:BEL- GAG
128	0,02	0,020	816	1,09E-20		0,0049	0,323	2.70% Class_I/LI NE:LINE- RT
129	0,02	0,020	796	8,75E-22		0,00383	0,586	unknow
130	0,02	0,020	796	1,14E-21		0,00126	0,57	unknow
131	0,02	0,020	795	1,65E-22		0,00126	0,699	14.84% Class_I/LI NE:LINE- RT
132	0,019	0,019	787	6,70E-21		0,00508	0,516	unknow
133	0,019	0,019	768	2,61E-21		0,0013	0,503	23.96% Class_I/LI NE:LINE- RT
134	0,018	0,018	743	9,38E-21		0,00679	0,454	unknow
135	0,018	0,018	727	6,64E-22		0,00138	0,608	unknow
136	0,017	0,017	707	3,93E-24		0,00141	0,89	18.53% Class_I/LI NE:LINE- RT
137	0,017	0,017	697	1,06E-20		0,00739	0,405	unknow
138	0,016	0,016	657	4,71E-23		0,00152	0,771	12.02% Class_I/LI NE:LINE- RT
139	0,016	0,016	654	1,09E-19		0,0489	0,758	unknow
140	0,016	0,016	650	7,49E-21		0,00929	0,501	unknow
141	0,016	0,016	637	1,29E-21		0,00159	0,561	unknow
142	0,015	0,015	624	2,29E-22		0,00481	0,682	1.60% Class_I/LI NE:LINE- ENDO
143	0,015	0,015	624	3,91E-21		0,00481	0,461	unknow
144	0,015	0,015	596	6,30E-21		0,00671	0,532	unknow

145	0,015	0,015	591	1,23E-24		0,00169	0,938	unknow	Run 1 - <i>D. borborema</i> (A)
146	0,014	0,014	570	0,955	436	0,939	0,962	5S rRNA	
147	0,014	0,014	560	1,03E-20		0,00179	0,327	17.68% organelle/m itochondria	
148	0,014	0,014	555	3,21E-23		0,0018	0,79	10.99% Class_I/LI NE:LINE- ENDO	Run 1 - <i>D. borborema</i> (B)
149	0,014	0,014	552	9,38E-21		0,00543	0,453	13.22% Class_I/LI NE:LINE- RT	
150	0,013	0,013	538	8,09E-21		0,00186	0,372	unknow	
151	0,013	0,013	537	0,0575	1799	0,806	0,833	unknow	
152	0,013	0,013	515	4,07E-16		0,151	0,88	unknow	
153	0,012	0,013	509	2,70E-22		0,00196	0,669	19.84% Class_I/LT R/Ty1_copi a:Ty1-RT	
154	0,012	0,012	506	2,29E-22		0,00198	0,681	unknow	
155	0,012	0,012	498	1,91E-20		0,0462	0,879	unknow	
156	0,012	0,012	484	1,92E-13		0,215	0,813	unknow	
157	0,012	0,012	482	2,47E-19		0,0249	0,393	unknow	
158	0,011	0,012	467	4,98E-22		0,00214	0,633	unknow	
159	0,011	0,011	466	4,29E-22		0,00215	0,635	unknow	
160	0,011	0,011	464	5,53E-21		0,00875	0,552	unknow	
161	0,011	0,011	443	4,91E-24		0,00226	0,877	unknow	
162	0,011	0,011	436	5,69E-23		0,00229	0,758	4.36% Class_I/LT R/Ty3_gyp sy:Ty3- GAG	

Table S5. Description of all clusters retrieved from reads of *D. borborema*(B) by RepeatExplorer2. The satDNA families analyzed in this study are highlighted in purple.

N input reads	41590124	N nuclear reads	5647839
N analyzed reads	5717574	$N_n = N_a - N_c$ - N_o	
N contamination reads	0		

N organelle reads 69735

Cluster	Proportion (%)	Proportion adjusted (%)	N reads	Satellite Probability	Consensus length (pb)	C index	P index	Similarity hits
1	0.650	0,660	37280	2.05e-06		0.850	0.103	unknow
2	0.650	0,659	37203	0.985	32	0.996	0.981	CDSTR16
3	0.540	0,544	30747	6.54e-16		0.102	0.168	1.19% organelle/plastid
4	0.520	0,530	29933	8.66e-06		0.836	0.151	unknow
5	0.370	0,372	21014	8.61e-18		0.0585	0.4280	unknow
6	0.350	0,357	20189	1.12e-03		0.6480	0.3490	unknow
7	0.350	0,354	19996	0.986	245	0.996	0.972	CDSTR230
8	0.330	0,331	18688	8.77e-14		0.166	0.155	unknow
9	0.330	0,329	18595	6.18e-15		0.130	0.171	unknow
10	0.310	0,314	17712	4.24e-09		0.312	0.541	unknow
11	0.300	0,304	17143	5.48e-11		0.295	0.0916	unknow
12	0.280	0,279	15735	1.56e-21		0.00852	0.6850	10.82% Class_I/LINE:LINE-ENDO
13	0.270	0,277	15640	1.34e-03		0.622	0.758	8.31% Class_I/LTR/Bel-Pao:BEL-GAG
14	0.260	0,262	14806	3.13e-06		0.754	0.0683	unknow
15	0.250	0,256	14459	7.70e-12		0.247	0.7760	10.17% Class_I/LTR/Ty3_gypsy:Ty3-INT
16	0.250	0,252	14256	2.08e-21		0.0187	0.8270	7.13% Class_I/LTR/Ty3_gypsy:Ty3-RT
17	0.250	0,252	14247	2.02e-11		0.239	0.5940	20.52% Class_I/LINE:LINE-RT
18	0.240	0,238	13439	0.01630	1337	0.926	0.599	unknow
19	0.230	0,230	12997	6.68e-12		0.253	0.0959	unknow
20	0.220	0,228	12860	7.94e-06		0.582	0.120	unknow
21	0.220	0,228	12850	2.16e-23		0.004120	0.8150	14.60% Class_I/LTR/Bel-Pao:BEL-GAG
22	0.220	0,226	12786	1.16e-06		0.775	0.053	unknow
23	0.210	0,216	12222	2.85e-20		0.0193	0.650	4.53% Class_I/LINE:LINE-RT
24	0.210	0,209	11826	0.853	86	0.998	0.927	IGS

25	0.210	0,209	11791	0.986	138	0.995	0.966	CDSTR138
26	0.200	0,207	11686	1.69e-16		0.0768	0.370	unknow
27	0.200	0,206	11635	4.15e-20		0.293	0.676	unknow
28	0.200	0,205	11603	1.39e-08		0.348	0.315	unknow
29	0.200	0,205	11572	5.65e-16		0.0923	0.3070	unknow
30	0.200	0,204	11509	0.994	112	0.994	0.967	DBC-150
31	0.200	0,203	11449	1.36e-13		0.164	0.304	unknow
32	0.200	0,202	11395	1.60e-03		0.621	0.6990	unknow
33	0.190	0,194	10932	1.43e-10		0.258	0.3670	unknow
34	0.190	0,191	10785	3.64e-05		0.783	0.1810	unknow
35	0.180	0,185	10470	5.74e-10		0.294	0.6270	27.65% Class_I/LTR/Bel-Pao:BEL-INT
36	0.180	0,184	10375	8.21e-23		0.0027	0.7450	13.65% Class_I/LINE:LINE-RT
37	0.180	0,179	10090	5.04e-22		0.006740	0.7790	37.58% Class_I/LINE:LINE-ENDO
38	0.180	0,178	10077	0.941	178	0.965	0.905	CDSTR198
39	0.170	0,173	9783	7.55e-22		0.116	0.7450	50.72% Class_I/LINE:LINE-RT
40	0.170	0,172	9709	0.4410	5974	0.921	0.855	9.17% Class_I/LTR/Bel-Pao:BEL-INT
41	0.170	0,170	9603	1.40e-21		0.0119	0.702	30.99% Class_I/LTR/Bel-Pao:BEL-GAG
42	0.170	0,169	9548	0.04060	5057	0.798	0.711	14.87% Class_II/Subclass_2/Helitron:Helitron-HEL2
43	0.170	0,169	9536	1.40e-21		0.005770	0.6960	54.18% Class_I/LTR/Bel-Pao:BEL-RT
44	0.170	0,167	9443	1.17e-23		0.000318	0.8430	22.33% organelle/mitochondria
45	0.160	0,167	9429	1.10e-21		0.0122	0.7160	unknow
46	0.160	0,166	9397	6.70e-21		0.0121	0.5240	unknow
47	0.160	0,165	9316	1.14e-21		0.004830	0.568	28.71% Class_I/LINE:LINE-RT
48	0.160	0,165	9304	1.27e-12		0.189	0.3340	unknow
49	0.160	0,161	9096	0.03930	3325	0.886	0.709	5.21% Class_I/LTR/Ty1_copia:Ty1-GAG

50	0.150	0,152	8562	1.56e-21		0.005 260	0.685	59.53% Class_I/LTR/Bel-Pao:BEL-INT
51	0.150	0,150	8493	1.69e-05		0.490	0.670	unknow
52	0.150	0,150	8489	2.16e-23		0.000 4	0.8140	4.09% organelle/mitochondria
53	0.150	0,150	8489	4.01e-05		0.682	0.140	unknow
54	0.150	0,149	8420	3.31e-11		0.267	0.145	unknow
55	0.150	0,148	8375	0.0564	3171	0.864	0.793	12.26% Class_I/LTR/Ty3_gypsy:Ty3-RT
56	0.140	0,144	8161	5.27e-14		0.197	0.873	17.18% Class_II/Subclass_2/Helitron:Helitron-HEL1
57	0.140	0,144	8149	2.71e-10		0.298	0.179	unknow
58	0.140	0,138	7787	2.16e-23		0.002 31	0.807	4.97% organelle/mitochondria
59	0.130	0,137	7716	0.1460	7087	0.794	0.617	5.56% Class_I/LTR/Ty3_gypsy:Ty3-RT
60	0.130	0,134	7571	3.90e-23		0.002 25	0.777	15.00% Class_I/LTR/Ty3_gypsy:Ty3-RT
61	0.130	0,131	7375	0.0308	4843	0.759	0.690	6.06% Class_I/LTR/Ty3_gypsy:Ty3-RT
62	0.130	0,130	7349	1.84e-20		0.031 6	0.740	9.93% Class_I/LTR/Bel-Pao:BEL-GAG
63	0.120	0,124	6981	3.21e-23		0.004 01	0.7870	unknow
64	0.120	0,123	6939	0.2700	4145	0.915	0.836	15.68% Class_I/LTR/Ty1_copia:Ty1-RT
65	0.120	0,122	6866	4.27e-13		0.198	0.642	unknow
66	0.120	0,120	6794	1.93e-12		0.227	0.737	unknow
67	0.120	0,120	6794	3.50e-19		0.056 7	0.761	18.94% Class_I/LTR/Ty3_gypsy:Ty3-RT
68	0.120	0,120	6769	8.27e-20		0.035 8	0.704	48.65% Class_I/LTR/Bel-Pao:BEL-RT
69	0.120	0,119	6749	9.43e-24		0.002 07	0.847	28.77% 45S_rDNA/18S_rDNA
70	0.120	0,117	6591	3.77e-17		0.101	0.7110	unknow
71	0.110	0,116	6531	8.21e-23		0.003 22	0.7390	unknow
72	0.110	0,116	6526	0.0317	3491	0.752	0.834	9.56% Class_I/LTR/Ty3_gypsy:Ty3-RT
73	0.110	0,115	6519	3.74e-03		0.645	0.704	unknow

74	0.110	0,113	6371	9.34e-06		0.714	0.0941	unknow
75	0.110	0,112	6310	0.755	3524	0.953	0.926	unknow
76	0.110	0,111	6293	0.7800	6107	0.963	0.866	9.12% Class_I/LTR/Bel-Pao:BEL-INT
77	0.110	0,111	6275	0.04150	1873	0.798	0.716	unknow
78	0.110	0,110	6199	1.82e-10		0.293	0.757	6.86% Class_I/LTR/Ty3_gypsy:Ty3-INT
79	0.110	0,109	6146	0.08060	211	0.880	0.875	unknow
80	0.100	0,103	5834	1.57e-15		0.111	0.182	unknow
81	0.100	0,103	5824	0.2340	6392	0.957	0.793	8.57% Class_I/LTR/Bel-Pao:BEL-RT
82	0.098	0,099	5575	1.80e-19		0.019 8	0.450	unknow
83	0.095	0,096	5405	1.23e-18		0.048 1	0.546	unknow
84	0.093	0,094	5295	0.0319	6577	0.756	0.762	4.74% Class_I/LTR/Ty3_gypsy:Ty3-RT
85	0.090	0,091	5141	8.21e-23		0.004 47	0.7350	unknow
86	0.088	0,089	5049	3.60e-18		0.086	0.807	3.03% Class_I/LTR/Ty3_gypsy:Ty3-RT
87	0.087	0,088	4961	7.49e-21		0.008 09	0.5040	unknow
88	0.086	0,087	4927	1.39e-22		0.001 22	0.7080	11.97% organelle/mitochondria
89	0.084	0,085	4821	8.58e-22		0.013 1	0.7370	unknow
90	0.082	0,084	4716	3.31e-19		0.022 1	0.1790	unknow
91	0.082	0,083	4701	2.47e-15		0.149	0.7130	2.55% Class_I/LTR/Ty3_gypsy:Ty3-GAG
92	0.081	0,082	4630	8.21e-23		0.003 46	0.7390	unknow
93	0.080	0,081	4579	9.81e-23		0.003 2	0.7330	5.94% Class_I/LTR/Ty3_gypsy:Ty3-PROT
94	0.079	0,080	4524	1.00e-20		0.007 74	0.427	17.84% Class_I/LTR/Bel-Pao:BEL-PROT
95	0.077	0,078	4389	7.88e-21		0.009 58	0.490	unknow
96	0.077	0,077	4376	0.03430	3045	0.816	0.646	2.33% Class_I/LTR/Ty3_gypsy:Ty3-RH

97	0.076	0,077	4329	1.72e-19		0.038 8	0.6370	unknow
98	0.075	0,075	4260	1.40e-21		0.006 57	0.6970	1.31% Class_I/LINE:LINE-RT
99	0.073	0,074	4196	2.16e-23		0.000 238	0.8060	28.46% 45S_rDNA/25S_rDNA
100	0.072	0,073	4135	0.00754	693	0.760	0.448	unknow
101	0.072	0,073	4104	3.65e-19		0.051 8	0.670	unknow
102	0.071	0,072	4044	2.16e-23		0.001 48	0.810	unknow
103	0.070	0,071	4011	1.76e-23		0.000 249	0.817	32.04% 45S_rDNA/25S_rDNA
104	0.068	0,069	3874	4.24e-05		0.503	0.474	unknow
105	0.067	0,068	3818	5.34e-04		0.594 0	0.7410	unknow
106	0.067	0,067	3811	3.61e-18		0.070 10	0.6820	3.88% Class_I/LTR/Ty3_gypsy:Ty3-RT
107	0.066	0,067	3777	3.34e-13		0.203 0	0.6650	unknow
108	0.065	0,066	3727	2.95e-17		0.070 0	0.4840	unknow
109	0.065	0,066	3708	7.57e-20		0.026 2	0.626	unknow
110	0.065	0,066	3708	0.979	255	0.999	0.990	unknow
111	0.063	0,064	3588	6.53e-21		0.003 34	0.4050	58.58% Class_I/LTR/Bel-Pao:BEL-RH
112	0.062	0,063	3566	3.16e-22		0.000 28	0.6650	1.35% organelle/mitochondria
113	0.062	0,063	3533	1.18e-20		0.010 8	0.3260	unknow
114	0.060	0,061	3459	4.29e-22		0.000 289	0.6430	4.34% organelle/mitochondria
115	0.060	0,061	3441	1.08e-18		0.055 6	0.6810	unknow
116	0.060	0,061	3422	1.49e-14		0.154	0.595	unknow
117	0.058	0,059	3310	7.64e-22		0.003 02	0.5980	13.14% Class_I/LINE:LINE-RT
118	0.056	0,057	3218	0.00924	18	0.702	0.473	unknow
119	0.056	0,057	3198	1.00e-21		0.002 51	0.582	4.44% Class_I/LTR/Ty1_copia:Ty1-RT
120	0.054	0,054	3070	1.27e-19		0.027	0.582	6.58% Class_I/LINE:LINE-RT

						4		
121	0.052	0,052	2949	3.96e-17		0.071 2	0.3490	unknow
122	0.051	0,051	2902	5.16e-21		0.006 55	0.555	2.79% Class_I/LINE:LINE-RT
123	0.050	0,051	2874	5.53e-21		0.008 76	0.5460	unknow
124	0.050	0,051	2874	6.70e-21		0.010 8	0.5171	unknow
125	0.050	0,051	2868	1.13e-09		0.300 0	0.3270	unknow
126	0.050	0,050	2846	2.03e-11		0.230 0	0.5150	unknow
127	0.049	0,050	2803	6.64e-22		0.002 510	0.6120	unknow
128	0.049	0,050	2800	0.7270	4217	0.947	0.486	15.04% Class_I/LTR/Ty1_copia:Ty1-RT
129	0.049	0,049	2784	6.64e-22		0.002 51	0.6130	unknow
130	0.049	0,049	2776	1.95e-22		0.002 91	0.6930	2.34% Class_I/LTR/Ty3_gypsy:Ty3-RH
131	0.047	0,048	2692	1.28e-19		0.015 6	0.5030	unknow
132	0.046	0,047	2655	2.49e-24		0.000 377	0.9060	8.74% Class_I/LINE:LINE-RT
133	0.046	0,047	2649	1.00e-20		0.005 680	0.4340	unknow
134	0.044	0,044	2500	1.00e-20		0.013 6	0.4330	2.16% Class_I/LINE:LINE-ENDO
135	0.043	0,043	2441	1.80e-07		0.415	0.796	unknow
136	0.042	0,043	2409	1.18e-19		0.019 9	0.5130	unknow
137	0.042	0,042	2394	3.63e-20		0.017 6	0.6300	unknow
138	0.042	0,042	2393	0.02450	4155	0.738	0.778	3.18% Class_I/LTR/Ty3_gypsy:Ty3-RT
139	0.042	0,042	2386	2.18e-19		0.054 7	0.7080	unknow
140	0.042	0,042	2378	3.12e-18		0.063 1	0.5930	unknow
141	0.041	0,041	2343	1.56e-21		0.008 020	0.6880	unknow

142	0.040	0,041	2307	4.80e-21		0.007 81	0.5750	12.01% Class_I/LINE:LINE-RT
143	0.040	0,041	2294	5.16e-21		0.012 7	0.5640	unknow
144	0.039	0,039	2221	6.70e-21		0.009 92	0.5230	unknow
145	0.036	0,037	2062	4.45e-21		0.005 83	0.5790	unknow
146	0.035	0,036	2007	0.7610	4349	0.995	0.902	14.25% Class_I/LTR/Ty1_copia:Ty1-RT
147	0.035	0,036	2005	0.02750	2233	0.746	0.616	unknow
148	0.034	0,035	1995	6.28e-20		0.023 7	0.5800	unknow
149	0.034	0,035	1967	5.74e-10		0.288 0	0.6340	7.22% Class_I/LINE:LINE-RT
150	0.034	0,035	1963	1.50e-20		0.003 060	0.0863	unknow
151	0.033	0,034	1899	1.17e-22		0.001 58	0.7150	unknow
152	0.033	0,033	1881	1.17e-20		0.006 3	0.313	unknow
153	0.033	0,033	1866	3.18e-21		0.008 34	0.6160	unknow
154	0.032	0,033	1839	1.65e-22		0.001 630	0.7000	10.44% Class_I/LINE:LINE-ENDO
155	0.032	0,032	1807	0.02620	2023	0.851	0.622	unknow
156	0.029	0,030	1684	4.32e-04		0.571 0	0.6430	unknow
157	0.029	0,029	1656	0.06510	3433	0.845	0.842	14.07% Class_I/LINE:LINE-RT
158	0.029	0,029	1641	0.53100	1226	0.965	0.819	unknow
159	0.029	0,029	1639	1.13e-20		0.012 2	0.3820	unknow
160	0.028	0,028	1599	5.06e-19		0.026 9	0.3610	unknow
161	0.028	0,028	1584	0.0514	4291	0.916	0.748	13.19% Class_I/LTR/Ty1_copia:Ty1-RT
162	0.027	0,027	1546	4.71e-23		0.003 88	0.7730	7.37% Class_I/LTR/Ty3_gypsy:Ty3-RT
163	0.026	0,026	1474	3.18e-21		0.010 5	0.6160	unknow
164	0.025	0,026	1457	5.53e-21		0.011 9	0.553	8.72% Class_I/LTR/Ty3_gypsy:Ty3-RT

165	0.025	0,025	1410	8.43e-18		0.066	0.6240	4.47% Class_I/LTR/Ty3_gypsy:Ty3-RH
166	0.025	0,025	1408	1.92e-19		0.043 3	0.6280	unknow
167	0.023	0,024	1340	9.02e-21		0.008 3	0.4610	unknow
168	0.023	0,023	1317	1.40e-20		0.000 759	0.2600	13.74% organelle/mitochondria
169	0.023	0,023	1289	3.31e-19		0.019 4	0.1780	unknow
170	0.022	0,023	1280	3.47e-21		0.007 0	0.6120	2.58% Class_I/LTR/Ty3_gypsy:Ty3-RT
171	0.022	0,023	1277	4.80e-21		0.007 83	0.575	unknow
172	0.022	0,022	1268	2.36e-20		0.041 4	0.788	unknow
173	0.022	0,022	1259	0.01480	1387	0.772	0.513	2.07% Class_I/LTR/Ty3_gypsy:Ty3-RT
174	0.022	0,022	1256	8.21e-23		0.000 796	0.7420	5.73% Class_I/LTR/Ty3_gypsy:Ty3-RT
175	0.022	0,022	1242	1.66e-21		0.004 03	0.545	unknow
176	0.022	0,022	1237	8.75e-22		0.004 85	0.5900	unknow
177	0.021	0,022	1219	1.27e-18		0.038 6	0.2650	unknow
178	0.021	0,021	1182	3.17e-19		0.037 6	0.5800	unknow
179	0.021	0,021	1175	6.30e-21		0.008 67	0.5260	unknow
180	0.021	0,021	1175	1.54e-20		0.002 55	0.2210	unknow
181	0.020	0,020	1144	7.94e-19		0.054 2	0.5960	unknow
182	0.020	0,020	1143	1.27e-19		0.035 00	0.5830	9.97% Class_I/LTR/Ty1_copia:Ty1-INT
183	0.020	0,020	1123	9.72e-21		0.008 010	0.4430	20.93% Class_I/LINE:LINE-RT
184	0.020	0,020	1119	0.25400	1679	0.885	0.939	unknow
185	0.019	0,019	1095	5.12e-21		0.002 74	0.4260	9.04% Class_I/LINE:LINE-RT
186	0.019	0,019	1091	2.86e-21		0.016 5	0.8120	5.77% Class_I/LINE:LINE-ENDO

187	0.018	0,018	1037	0.883	436	0.928	0.946	unknow
188	0.018	0,018	1028	4.45e-21		0.010 0	0.5820	unknow
189	0.018	0,018	1021	6.92e-19		0.070 5	0.7880	unknow
190	0.018	0,018	1008	5.76e-22		0.004 960	0.6230	unknow
191	0.017	0,017	962	1.07e-19		0.041 6	0.6760	unknow
192	0.016	0,016	900	3.18e-21		0.011 1	0.6160	unknow
193	0.015	0,015	864	0.992	456	0.988	0.977	unknow
194	0.014	0,015	823	2.96e-15		0.122	0.547	unknow
195	0.014	0,014	788	3.21e-23		0.001 27	0.7870	29.31% Class_I/LTR/Ty3_gypsy:Ty3-INT
196	0.013	0,013	732	2.47e-19		0.024 6	0.389	unknow
197	0.013	0,013	729	1.03e-17		0.061 7	0.2770	unknow
198	0.012	0,012	697	4.91e-24		0.001 430	0.8840	17.79% Class_I/LTR/Ty1_copia:Ty1-RT
199	0.012	0,012	679	1.00e-19		0.019 3	0.7860	10.77% Class_I/LTR/Ty3_gypsy:Ty3-RT
200	0.011	0,012	650	3.92e-21		0.001 55	0.6910	unknow
201	0.011	0,011	646	1.95e-22		0.001 55	0.6910	unknow
202	0.011	0,011	629	3.07e-19		0.015 9	0.330	1.91% Class_I/LINE:LINE-RT
203	0.011	0,011	627	1.37e-20		0.001 59	0.0627	unknow
204	0.011	0,011	619	7.81e-11		0.262	0.6120	unknow
205	0.011	0,011	608	1.63e-20		0.001 64	0.1710	unknow
206	0.010	0,011	600	3.90e-23		0.001 67	0.7800	unknow
207	0.010	0,010	585	3.21e-18		0.051 3	0.327	unknow

Run 1 - *D. gouveai***Table S6.** Description of all clusters retrieved from reads of *D. gouveai* by RepeatExplorer2. The satDNA families analyzed in this study are highlighted in purple.

N input reads	55525038		N nuclear reads	4129185				
N analyzed reads	4156234		Nn = Na - Nc - No					
N contaminat ion reads	0							
N organelle reads	27049							

Cluster	Proportion (%)	Proportion adjusted (%)	N reads	Satellite Probability	Consensus length (pb)	C index	P index	Similarity hits
1	0,57	0,578	23859	0,992	230	0,986	0,649	CDSTR230
2	0,52	0,523	21590	2,22e-03	0	0,658	0,389	unknow
3	0,5	0,505	20845	5,70E-11	0	0,263	0,246	unknow
4	0,47	0,474	19556	3,49E-11	0	0,272	0,788	35.91% Class_I/LIN E:LINE-RT
5	0,47	0,469	19357	9,55E-01	86	0,937	0,609	IGS
6	0,47	0,468	19341	9,81E-01	136	0,973	0,619	CDSTR138
7	0,45	0,453	18719	1,34E-15	0	0,16	0,846	7.37% Class_I/LT R/Bel- Pao:BEL- GAG
8	0,44	0,442	18251	9,33E-01	24	0,971	0,717	CDSTR16
9	0,41	0,412	16992	9,92E-01	370	0,985	0,835	pBuM2
10	0,41	0,408	16858	6,83E-11	0	0,25	0,351	unknow
11	0,4	0,399	16479	4,18E-13	0	0,169	0,394	unknow
12	0,39	0,395	16291	1,33E-02	535	0,856	0,555	unknow
13	0,37	0,371	15324	7,85E-02	4999	0,881	0,848	7.33% Class_I/LIN

								E:LINE-RT
14	0,37	0,371	15306	2,86E-21	0	0,0172	0,814	27.62% Class_I/LIN E:LINE- ENDO
15	0,36	0,361	14895	1,38E-05	0	0,51	0,314	unknown
16	0,33	0,335	13834	4,14E-05	0	0,512	0,658	unknown
17	0,31	0,313	12937	0,017	511	0,722	0,75	TRANSPO SON
18	0,31	0,311	12833	9,25E-04	0	0,615	0,795	unknown
19	0,3	0,303	12528	4,76E-14	0	0,141	0,436	unknown
20	0,29	0,292	12057	6,76E-15	0	0,148	0,643	unknown
21	0,28	0,282	11662	2,33E-17	0	0,071	0,516	unknown
22	0,26	0,267	11005	7,41E-02	2324	0,867	0,866	unknown
23	0,26	0,261	10792	4,66E-04	0	0,587	0,834	9.22% Class_I/LT R/Ty3_gyps y:Ty3-RT
24	0,26	0,259	10700	5,31E-01	6315	0,929	0,892	8.89% Class_I/LT R/Ty3_gyps y:Ty3-RT
25	0,25	0,252	10403	2,73E-01	5046	0,906	0,869	17.73% Class_I/LIN E:LINE-RT
26	0,25	0,248	10259	1,17E-23	0	0,00487	0,844	13.98% Class_I/LT R/Ty3_gyps y:Ty3-INT
27	0,24	0,245	10100	7,30E-15	0	0,178	0,883	8.79% Class_I/LT R/Ty3_gyps y:Ty3-RT
28	0,24	0,240	9914	9,32E-23	0	0,00514	0,893	57.94% 45S_rDNA/ 25S_rDNA
29	0,29	0,227	9386	6,44E-12	0	0,213	0,463	unknown
30	0,21	0,216	8923	0,933	149	0,973	0,553	DBC-150
31	0,2	0,206	8512	7,10E-16	0	0,125	0,683	unknown
32	0,2	0,204	8429	0,757	6215	0,941	0,895	7.98% Class_I/LT R/Ty3_gyps y:Ty3-RT
33	0,2	0,204	8405	9,32E-23	0	0,00595	0,889	36.50% 45S_rDNA/

								18S_rDNA
34	0,19	0,192	7928	4,20E-03	0	0,664	0,808	3.30% Class_I/LT R/Ty3_gyps y:Ty3-RT
35	0,18	0,182	7526	1,12E-11	0	0,217	0,495	unknow
36	0,18	0,181	7475	4,29E-01	6308	0,919	0,875	7.28% Class_I/LT R/Ty3_gyps y:Ty3-RT
37	0,18	0,180	7423	0,334	6114	0,897	0,92	4.41% Class_I/LT R/Ty3_gyps y:Ty3-RT
38	0,17	0,172	7097	1,46E-15	0	0,0984	0,286	unknow
39	0,17	0,169	6979	5,71E-10	0	0,31	0,789	40.30% Class_I/LIN E:LINE- ENDO
40	0,17	0,169	6972	4,16E-03	0	0,646	0,674	unknow
41	0,17	0,167	6892	9,32E-04	0	0,606	0,792	3.53% Class_I/LT R/Bel- Pao:BEL- GAG
42	0,15	0,154	6378	5,58E-02	6920	0,814	0,815	6.10% Class_I/LT R/Ty3_gyps y:Ty3-RT
43	0,15	0,154	6339	2,87E-12	0	0,206	0,263	unknow
44	0,15	0,150	6202	1,76E-23	0	0,0032	0,822	48.37% Class_I/LIN E:LINE-RT
45	0,14	0,144	5965	6,80E-02	226	0,826	0,856	unknow
46	0,14	0,144	5953	7,56E-25	0	0,00134	0,957	14.78% organelle/m itochondria
47	0,14	0,143	5923	2,57E-21	0	0,0279	0,869	11.78% Class_I/LIN E:LINE-RT
48	0,14	0,141	5804	2,83E-22	0	0,0112	0,817	36.27% 45S_rDNA/ 25S_rDNA
49	0,14	0,138	5689	5,15E-09	0	0,339	0,255	unknow
50	0,14	0,137	5650	8,61E-02	3278	0,843	0,958	unknow
51	0,14	0,136	5617	1,59E-08	0	0,372	0,866	4.38% Class_I/LT

								R/Bel- Pao:BEL- INT
52	0,13	0,136	5600	6,60E-01	7086	0,939	0,889	7.38% Class_I/LT R/Ty3_gyps y:Ty3-RT
53	0,13	0,135	5577	2,83E-22	0	0,00628	0,824	42.39% 45S_rDNA/ 18S_rDNA
54	0,12	0,125	5166	1,08E-18	0	0,0553	0,681	unknow
55	0,12	0,125	5142	5,53E-20	0	0,0362	0,728	unknow
56	0,12	0,123	5087	1,40E-21	0	0,0106	0,705	3.11% organelle/m itochondria
57	0,12	0,119	4909	0,0501	3905	0,759	0,91	5.24% Class_I/LIN E:LINE-RT
58	0,12	0,118	4871	7,27E-02	7650	0,824	0,887	4.99% Class_I/LT R/Ty3_gyps y:Ty3-INT
59	0,11	0,114	4713	1,36E-13	0	0,207	0,872	9.70% Class_I/LT R/Bel- Pao:BEL- INT
60	0,11	0,113	4657	1,44E-23	0	0,00236	0,832	13.66% Class_I/LT R/Ty3_gyps y:Ty3-RT
61	0,11	0,112	4613	1,56E-24	0	0,00217	0,929	40.65% organelle/m itochondria
62	0,11	0,109	4494	0,022	3199	0,73	0,822	4.43% Class_I/LT R/Ty3_gyps y:Ty3-INT
63	0,11	0,109	4482	2,49E-24	0	0,00134	0,906	13.70% Class_I/LT R/Ty3_gyps y:Ty3-INT
64	0,11	0,108	4439	6,44E-01	6838	0,929	0,909	5.52% Class_I/LT R/Ty3_gyps y:Ty3-INT
65	0,11	0,106	4394	1,09E-18	0	0,0428	0,371	unknow
66	0,1	0,105	4328	6,79E-04	0	0,598	0,776	20.22% Class_II/Su bclass_2/He

								litron:Helitron-HEL2
67	0,1	0,104	4287	1,56E-24	0	0,000233	0,925	23.63% organelle/mitochondria
68	0,099	0,100	4123	6,60E-01	6391	0,939	0,892	10.72% Class_I/LTR/Bel-Pao:BEL-INT
69	0,098	0,098	4062	2,39E-21	0	0,0116	0,647	2.39% Class_I/LTR/Ty3_gypsy:Ty3-RT
70	0,09	0,091	3748	1,75E-21	0	0,00587	0,684	1.23% Class_I/LINE:LINE-ENDO
71	0,088	0,089	3663	1,95E-13	0	0,159	0,366	unknow
72	0,088	0,089	3658	7,25E-02	5787	0,875	0,86	6.83% Class_I/LTR/Ty3_gypsy:Ty3-RT
73	0,088	0,088	3640	8,27E-02	3733	0,881	0,886	2.61% Class_I/LTR/Ty3_gypsy:Ty3-RT
74	0,087	0,088	3625	2,39E-21	0	0,0127	0,654	13.32% Class_I/LINE:LINE-RT
75	0,087	0,088	3620	0,979	1997	0,997	0,993	unknow
76	0,087	0,087	3577	4,11E-21	0	0,0143	0,595	unknow
77	0,087	0,086	3570	4,71E-23	0	0,0014	0,772	11.88% Class_I/LINE:LINE-RT
78	0,085	0,086	3539	1,18E-11	0	0,261	0,846	2.66% Class_I/LTR/Ty3_gypsy:Ty3-INT
79	0,085	0,086	3531	3,90E-23	0	0,00172	0,782	12.29% Class_I/LINE:LINE-RT
80	0,085	0,085	3513	3,78E-21	0	0,00626	0,598	6.18% Class_I/LTR/Bel-Pao:BEL-PROT
81	0,084	0,084	3487	2,03E-02	4145	0,715	0,857	12.27% Class_I/LTR/Ty1_copi

								a:Ty1-RT
82	0,083	0,083	3436	1,09E-19	0	0,0463	0,76	unknow
83	0,082	0,082	3398	2,18E-02	4278	0,732	0,669	5.41% Class_I/LT R/Ty3_gyps y:Ty3-RT
84	0,081	0,082	3385	2,11E-13	0	0,203	0,702	1.18% Class_I/LT R/Ty3_gyps y:Ty3-INT
85	0,08	0,080	3313	5,78E-22	0	0,00755	0,769	unknow
86	0,08	0,080	3309	6,59E-02	5919	0,849	0,838	10.00% Class_I/LT R/Bel- Pao:BEL- RT
87	0,076	0,076	3148	6,31E-02	3265	0,842	0,834	6.35% Class_I/LT R/Ty3_gyps y:Ty3-GAG
88	0,076	0,076	3138	5,50E-20	0	0,0561	0,894	22.05% Class_II/Su bclass_2/He litron:Helitr on-HEL1
89	0,075	0,075	3115	2,12E-02	3310	0,733	0,803	7.58% Class_I/LT R/Bel- Pao:BEL- GAG
90	0,074	0,074	3069	7,01E-02	6806	0,808	0,885	6.65% Class_I/LT R/Ty3_gyps y:Ty3-RT
91	0,072	0,072	2987	1,24E-21	0	0,00814	0,707	unknow
92	0,071	0,072	2965	1,17E-17	0	0,0941	0,729	unknow
93	0,07	0,070	2909	6,75E-20	0	0,0251	0,635	unknow
94	0,069	0,070	2885	7,78E-02	5126	0,844	0,902	unknow
95	0,067	0,067	2772	5,10E-17	0	0,117	0,814	unknow
96	0,066	0,066	2734	8,75E-22	0	0,00476	0,593	unknow
97	0,065	0,065	2700	4,60E-25	0	0,00148	0,979	8.59% Class_I/LIN E:LINE-RT
98	0,063	0,063	2622	1,70E-19	0	0,0575	0,811	unknow
99	0,063	0,063	2620	2,48E-03	0	0,64	0,749	4.43% Class_I/LT R/Ty3_gyps

								y:Ty3-INT
100	0,063	0,063	2617	6,61E-02	188	0,821	0,857	CDSTR198
101	0,062	0,063	2593	9,43E-24	0	0,00309	0,848	unknow
102	0,061	0,061	2537	6,12E-24	0	0,000394	0,871	36.07% organelle/pl astid
103	0,057	0,057	2368	2,88E-04	0	0,565	0,72	7.52% Class_I/LT R/Bel- Pao:BEL- RT
104	0,057	0,057	2349	2,57E-02	3233	0,738	0,81	3.79% Class_I/LT R/Ty3_gyps y:Ty3-INT
105	0,056	0,056	2320	8,46E-01	4229	0,948	0,922	16.47% Class_I/LT R/Ty1_copi a:Ty1-RT
106	0,056	0,056	2310	7,63E-21	0	0,0329	0,796	8.96% Class_I/LIN E:LINE-RT
107	0,055	0,055	2282	2,90E-21	0	0,0123	0,631	unknow
108	0,054	0,054	2243	8,19E-01	4217	0,974	0,927	16.94% Class_I/LT R/Ty1_copi a:Ty1-RT
109	0,054	0,054	2231	2,58E-04	0	0,566	0,748	5.65% Class_I/LT R/Ty1_copi a:Ty1-GAG
110	0,053	0,054	2216	4,98E-22	0	0,00226	0,632	unknow
111	0,053	0,054	2213	5,01E-13	0	0,211	0,696	4.20% Class_I/LT R/Ty3_gyps y:Ty3-RT
112	0,053	0,053	2196	5,25E-03	0	0,663	0,901	15.80% Class_I/LT R/Ty1_copi a:Ty1-RT
113	0,048	0,048	1981	1,10E-21	0	0,0145	0,721	unknow
114	0,044	0,044	1833	9,23E-01	2743	0,974	0,982	unknow
115	0,043	0,043	1773	1,63E-04	0	0,563	0,89	12.13% Class_I/LIN E:LINE-RT
116	0,041	0,041	1707	4,71E-23	0	0,0041	0,765	unknow

117	0,04	0,040	1671	2,29E-22	0	0,0039	0,683	unknow
118	0,04	0,040	1670	1,40E-21	0	0,0102	0,697	unknow
119	0,04	0,040	1667	1,65E-22	0	0,00373	0,699	1.32% Class_I/LT R/Ty3_gyps y:Ty3-INT
120	0,039	0,039	1616	9,81E-23	0	0,00495	0,734	unknow
121	0,038	0,038	1588	2,76E-25	0	0,000646	0,997	unknow
122	0,037	0,037	1548	6,55E-01	3922	0,928	0,995	unknow
123	0,036	0,036	1502	3,22E-19	0	0,0519	0,682	unknow
124	0,036	0,036	1502	8,58E-22	0	0,00652	0,74	unknow
125	0,035	0,035	1448	5,06E-02	1461	0,787	0,828	8.63% Class_I/LT R/Ty3_gyps y:Ty3-RT
126	0,034	0,034	1412	2,39E-17	0	0,113	0,803	10.48% Class_I/LT R/Ty3_gyps y:Ty3-RT
127	0,033	0,033	1358	0,0421	3672	0,757	0,868	3.24% Class_I/LIN E:LINE- ENDO
128	0,033	0,033	1351	1,14E-21	0	0,0348	0,916	unknow
129	0,032	0,033	1346	4,91E-24	0	0,000743	0,883	unknow
130	0,032	0,033	1345	1,24E-21	0	0,00595	0,713	unknow
131	0,031	0,031	1287	4,98E-22	0	0,000777	0,631	unknow
132	0,03	0,030	1248	7,42E-01	6484	0,937	0,932	7.05% Class_I/LT R/Ty3_gyps y:Ty3-RT
133	0,029	0,030	1220	9,79E-01	4951	0,999	0,993	unknow
134	0,029	0,029	1188	1,20E-01	1587	0,869	0,964	unknow
135	0,028	0,029	1184	9,81E-23	0	0,000845	0,726	1.10% Class_I/LT R/Ty3_gyps y:Ty3-INT
136	0,027	0,027	1112	6,84E-23	0	0,0045	0,746	unknow
137	0,026	0,026	1068	1,17E-23	0	0,000936	0,835	16.39% Class_I/LIN E:LINE-RT
138	0,025	0,026	1057	8,19E-01	4349	0,969	0,925	16.93% Class_I/LT R/Ty1_copi

								a:Ty1-RT
139	0,025	0,025	1038	5,05E-02	2477	0,843	0,756	1.35% Class_I/LT R/Ty3_gyps y:Ty3-INT
140	0,025	0,025	1028	2,49E-24	0	0,00292	0,911	11.19% 45S_rDNA/ 25S_rDNA
141	0,024	0,024	1004	1,40E-21	0	0,00697	0,702	unknow
142	0,023	0,024	971	5,76E-22	0	0,00412	0,624	6.90% Class_I/LIN E:LINE-RT
143	0,023	0,023	970	9,75E-01	6794	0,982	0,952	9.07% Class_I/LT R/Ty3_gyps y:Ty3-RT
144	0,023	0,023	968	2,29E-22	0	0,0031	0,683	2.38% organelle/pl astid
145	0,023	0,023	957	2,70E-22	0	0,00104	0,667	unknow
146	0,023	0,023	948	1,23E-18	0	0,591	0,672	unknow
147	0,022	0,023	934	8,02E-02	5338	0,876	0,868	11.24% Class_I/LT R/Ty3_gyps y:Ty3-RT
148	0,022	0,023	932	1,14E-21	0	0,00107	0,572	13.73% Class_I/LIN E:LINE-RT
149	0,022	0,022	897	2,70E-22	0	0,00447	0,67	unknow
150	0,022	0,022	895	2,16E-21	0	0,00687	0,66	unknow
151	0,021	0,021	877	1,65E-22	0	0,00466	0,696	unknow
152	0,021	0,021	865	9,59E-01	4424	0,998	0,998	unknow
153	0,02	0,021	852	1,56E-21	0	0,0117	0,687	unknow
154	0,02	0,020	832	6,72E-02	423	0,839	0,849	1.08% organelle/pl astid
155	0,019	0,020	807	4,91E-24	0	0,0124	0,877	12.02% Class_I/LT R/Bel- Pao:BEL- INT
156	0,019	0,019	794	8,33E-01	69	0,933	0,941	unknow
157	0,019	0,019	769	1,30E-22	0	0,013	0,867	17.17% Class_I/LT R/Ty1_copi

								a:Ty1-INT
158	0,017	0,017	716	1,65E-22	0	0,0014	0,705	unknow
159	0,017	0,017	690	6,84E-23	0	0,00435	0,751	unknow
160	0,017	0,017	688	7,88E-21	0	0,00581	0,489	1.16% organelle/pl astid
161	0,016	0,016	680	2,49E-24	0	0,00147	0,91	unknow
162	0,016	0,016	660	1,97E-24	0	0,00152	0,919	15.45% organelle/m itochondria
163	0,016	0,016	657	1,12E-20	0	0,0198	0,72	unknow
164	0,016	0,016	649	6,61E-22	0	0,00633	0,764	unknow
165	0,016	0,016	647	4,11E-21	0	0,00927	0,586	unknow
166	0,015	0,015	624	5,16E-21	0	0,00643	0,556	unknow
167	0,015	0,015	622	2,16E-21	0	0,00838	0,663	unknow
168	0,015	0,015	611	0,991	255	0,987	0,984	unknow
169	0,014	0,014	598	9,43E-24	0	0,00168	0,846	unknow
170	0,014	0,014	597	7,80E-15	0	0,173	0,756	4.69% Class_I/LT R/Ty1_copi a:Ty1-RT
171	0,014	0,014	596	9,66E-25	0	0,173	0,756	unknow
172	0,014	0,014	591	0,0715	1209	0,802	0,913	unknow
173	0,014	0,014	576	6,61E-22	0	0,00168	0,948	unknow
174	0,014	0,014	575	1,23E-24	0	0,0107	0,756	18.09% Class_I/LIN E:LINE-RT
175	0,014	0,014	567	1,97E-24	0	0,00174	0,943	8.11% organelle/m itochondria
176	0,012	0,012	515	8,21E-23	0	0,00176	0,916	unknow
177	0,012	0,012	503	1,76E-23	0	0,00194	0,74	9.74% Class_I/LT R/Ty3_gyps y:Ty3-INT
178	0,012	0,012	496	1,40E-21	0	0,00199	0,816	unknow
179	0,012	0,012	485	2,09E-22	0	0,00619	0,837	18.56% 45S_rDNA/ 5.8S_rDNA
180	0,012	0,012	484	0,974	456	0,975	0,984	unknow
181	0,011	0,011	456	1,75E-21	0	0,0111	0,683	25.44% Class_I/LIN

								E:LINE-RT
182	0,011	0,011	446	2,70E-22	0	0,00224	0,67	unknow
183	0,011	0,011	442	7,92E-19	0	0,0928	0,939	unknow
184	0,01	0,010	431	2,69E-19	0	0,0232	0,368	unknow
185	0,01	0,010	430	9,43E-24	0	0,00233	0,853	13.72% organelle/m itochondria
186	0,01	0,010	427	2,49E-24	0	0,00234	0,915	2.81% organelle/m itochondria
187	0,01	0,010	419	8,43E-18	0	0,0692	0,618	15.27% Class_1/LT R/Ty3_gyps y:Ty3-INT