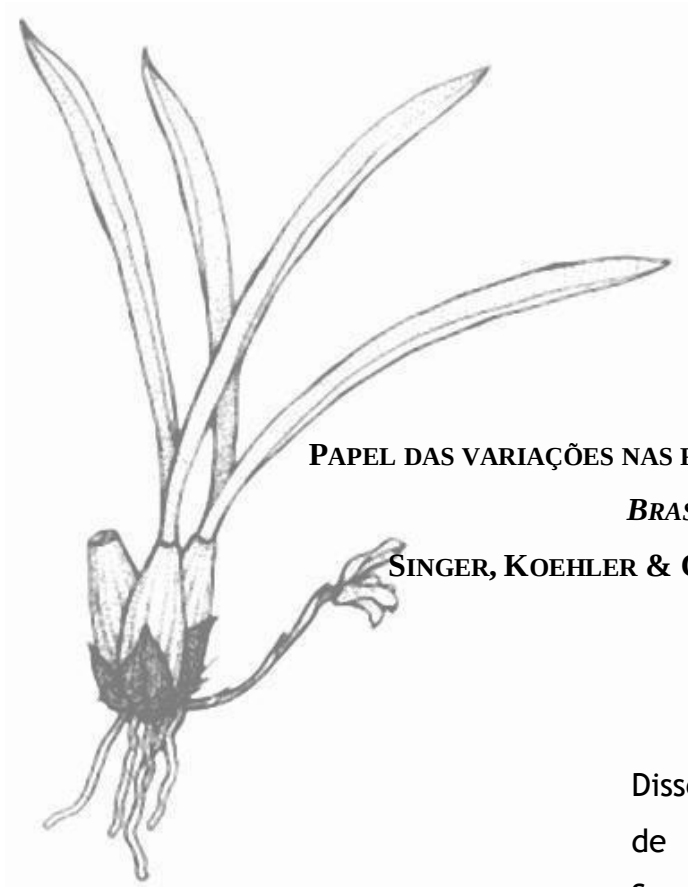


Thayná Juliane Guerra da Silva



**PAPEL DAS VARIAÇÕES NAS FAMÍLIAS DE DNA REPETITIVO EM
BRASILIORCHIS
SINGER, KOEHLER & CARNEVALI (ORCHIDACEAE)**

Dissertação apresentada ao Instituto de Pesquisas Ambientais, da Secretaria de Meio Ambiente, Infraestrutura e Logística, como parte dos requisitos exigidos para a obtenção do título de MESTRE em BIODIVERSIDADE VEGETAL E MEIO AMBIENTE, na Área de concentração de Plantas Vasculares em Análises Ambientais.

São Paulo - SP

2023

Thayná Juliane Guerra da Silva

PAPEL DAS VARIAÇÕES NAS FAMÍLIAS DE DNA REPETITIVO EM *BRASILIORCHIS*

R.B.SINGER, S.KOEHLER & CARNEVALI (ORCHIDACEAE)

Dissertação apresentada ao Instituto de Pesquisas Ambientais, da Secretaria de Meio Ambiente, Infraestrutura e Logística, como parte dos requisitos exigidos para a obtenção do título de MESTRE em BIODIVERSIDADE VEGETAL E MEIO AMBIENTE, na Área de concentração de Plantas Vasculares em Análises Ambientais.

ORIENTADORA: DRA. ANA PAULA DE MORAES

Ficha Catalográfica elaborada pelo **NÚCLEO DE BIBLIOTECAS E MAPOTECAS**

Silva, Thayná Juliane Guerra da
S586p Papel das variações nas famílias de DNA Repetitivo em *Brasiliorchis* Singer, Koehler & Carnevali (Orchidaceae) / Thayná Juliane Guerra da Silva - - São Paulo, 2023.
86p.; il.

Dissertação (Mestrado) -- Instituto de Pesquisas Ambientais da Secretaria de Meio Ambiente, Infraestrutura e Logística, 2023.

Bibliografia.

1. DNA satélite. 2. Elemento transponíveis. 3. NGS. I. Título.

CDU: 582.594.2

COMISSÃO JULGADORA DA DISSERTAÇÃO



Documento assinado digitalmente

ANA PAULA DE MORAES

Data: 12/06/2023 16:01:35-0300

Verifique em <https://validar.it.gov.br>

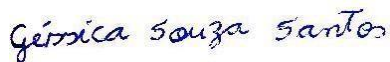
Profa. Dra. Ana Paula de Moraes

(Orientadora, Presidente)

A handwritten signature in grey ink that reads "Magdalena Vaio".

Profa. Dra. Magdalena Vaio

Universidad de la República Udelar (UDELAR)

A handwritten signature in blue ink that reads "Gessica Souza Ramos".

Dra. Gécica Souza Ramos

Universidade do Estado da Bahia - UNEB

Ao meu querido avô, Gerson Gonçalves Guerra,
que partiu antes do fim dessa aventura, sou grata
pelo amor e inspiração que trouxe á minha vida,
sempre lembrarei com carinho de sua presença e
ensinamentos.

In memoriam.

AGRADECIMENTOS

Dedico minha profunda gratidão a todas as pessoas que cruzaram meu caminho, pois cada uma delas trouxe valiosos ensinamentos. Agradeço especialmente àqueles que, de alguma forma, contribuíram para minha formação profissional como bióloga, pesquisadora e professora. Em especial:

Expresso minha sincera gratidão ao Instituto de Pesquisas Ambientais (IPA), especificamente ao programa de Pós-Graduação em Biodiversidade Vegetal e Meio Ambiente (PPG-IPA) por fornecer condições para realização deste trabalho,

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de mestrado,

À Universidade Federal do ABC pelo apoio no programa de Pesquisador Colaborador,

Ao Laboratório de Citogenômica e Evolução de plantas da UFABC, por todo suporte ao longo dessa caminhada. Em especial à Laura e Sylvia, sou imensamente grata a todo apoio e carinho que recebi de vocês, agradeço por todas as dicas de edição e todo ensinamento nos debates. Minha querida Laura, gostaria de expressar mais uma vez minha gratidão por sua ajuda e apoio mesmo quando estava ocupada e cheia de disciplinas, sua generosidade, compreensão e disposição em ajudar são verdadeiras e notáveis, obrigada. Agradeço também ao José Juliano Amorim, por sua ajuda e por ceder imagens que ajudaram na construção dessa etapa,

À Pesquisadora Dra. Magdalena Vaio, pela contribuição tão valiosa e fundamental para melhoria deste trabalho. Muito Obrigada!

À minha orientadora, Dra. Ana Paula Moraes, agradeço o seu valioso apoio, dedicação e orientação ao longo deste processo, por toda sua paciência para me ajudar nos momentos de dúvidas, seu carinho e principalmente os puxões de orelha que foram fundamentais para o meu desenvolvimento acadêmico, profissional e pessoal. Muito Obrigada, Ana Paula, sou extremamente grata por tê-la como minha orientadora!

À minha família, especialmente Tassia Guerra, Tatiane Guerra, Cecilia Mariz, Graziella Guerra, Gleyvison Guerra, por todo apoio nessa jornada, sem dúvida cada contribuição foi fundamental para minha formação, seja profissional ou como pessoa.

À minha querida vizinha Maria de Lourdes, sua presença amorosa, seu cuidado constante e seu carinho incondicional significam o mundo para mim.

Queridos Gleyce Guerra e Marivaldo Rodrigues, palavras não são suficientes para expressar minha gratidão por tudo que vocês fizeram por mim. Sou verdadeiramente abençoada por ter vocês como meus pais. Obrigada por seu amor incondicional, apoio constante ao longo dessa batalha e por serem os melhores pais do mundo. Amo vocês infinitamente.

A Miguel Fernandes, meu namorado e melhor amigo, agradeço por ter você em minha vida. Sou grata por cada momento que compartilhamos juntos, seu apoio, compreensão, incentivo. Obrigada por ser o companheiro perfeito em todas as aventuras da minha vida. Te amo!

Aos colegas e amigos que fiz ao longo dessa trajetória acadêmica, em especial Marcela Barbosa e Laís Marta, sou grata pela amizade de vocês, tantas risadas e

sonhos compartilhados. Obrigada pelo eterno incentivo de vocês, que a nossa amizade perdure muitos anos.

Aos que aqui não foram citados e que em algum momento me auxiliaram neste trabalho, muito obrigada!!

Obrigada a todos!

RESUMO

A espécie *Brasiliorchis schunkeana*, uma espécie de flores negras típicas, se destaca também por apresentar grandes blocos de heterocromatina ricos em AT em seis dos seus 40 cromossomos. A presença de blocos heterocromáticos é uma característica comum, porém estes são comumente formados por DNAsat rico em CG que formam faixas nos cromossomos. Blocos heterocromáticos ricos em AT e que ocupam quase a totalidade dos cromossomos são raros. Em associação a esses grandes blocos, essa espécie apresenta um aumento do GS ($2C=4,19pg$) em comparação com as demais espécies do gênero ($2C$ médio= $3,19pg$), mas com manutenção do número cromossômico ($2n=40$ em todas as espécies do gênero). Alterações no GS em plantas são frequentemente atribuídas à expansão/retração de TEs, sequências repetitivas dispersas. Dessa forma, esta dissertação visou identificar se a variação do GS em *B. schunkeana* deve-se à variação na fração de TEs ou DNAsat e qual(is) sequência(s) causaram a mudança no GS. Para realizar tal estudo, foi feita uma análise genômica comparativa entre *B. schunkeana* e a espécie tipo do gênero, *B. picta*. O principal componente do repetoma das duas espécies é constituído por retroelementos do tipo LTR, com a linhagem OGRE sendo a mais representativa. Porém a composição de TE é homogênea entre as duas espécies, diferente da fração de DNAsat: 6% do genoma (9% do repetoma) de *B. schunkeana* foi anotado como DNAsat, o dobro do anotado para *B. picta*. Dos três cluster mais representativos de DNAsat, dois apresentam composição rica em AT, o que é compatível com a presença de blocos heterocromáticos DAPI⁺ no genoma de *B. schunkeana*. Por fim, concluímos que a variação do tamanho de genoma em *B. schunkeana* deve-se a amplificação de DNAsat, sequências CL 32 e 51, e não a expansão de TE.

Palavra-chave: DNA satélite, Elemento transponíveis, NGS.

ABSTRACT

The species *Brasiliorchis schunkeana*, a species of typical black flowers, stands out also for presenting large blocks of AT-rich heterochromatin in six of its 40 chromosomes. The presence of heterochromatic blocks is a common characteristic, but these are commonly formed by CG-rich DNAsat that form bands on the chromosomes. AT-rich heterochromatic blocks that occupy almost all of the chromosomes are rare. In association with these large blocks, this species presents an increase in GS ($2C=4.19\text{pg}$) compared to other species in the genus (average $2C=3.19\text{pg}$), but with maintenance of the chromosome number ($2n=40$ in all species of the genus). Changes in GS in plants are often attributed to the expansion/retraction of TEs, dispersed repetitive sequences. Thus, this dissertation aimed to identify whether the variation in GS in *B. schunkeana* is due to variation in the fraction of TEs or DNAsat and which sequence(s) caused the change in GS. To perform this study, a comparative genomic analysis was carried out between *B. schunkeana* and the type species of the genus, *B. picta*. The main component of the repeatome of both species is constituted by LTR retroelements, with the OGRE lineage being the most representative. However, the TE composition is homogeneous between the two species, unlike the fraction of DNAsat: 6% of the genome (9% of the repeatome) of *B. schunkeana* was annotated as DNAsat, twice the amount annotated for *B. picta*. Of the three most representative DNAsat clusters, two have an AT-rich composition, which is compatible with the presence of DAPI+ heterochromatic blocks in the genome of *B. schunkeana*. Finally, we conclude that the variation in genome size in *B. schunkeana* is due to the amplification of DNAsat, CL 32 and 51 sequences, and not to the expansion of TEs.

Keyword: Satellite DNA, Transposable Elements, NGS.

LISTA DE FIGURAS

Figura 1. Representação esquemática da composição do genoma dos eucariotos.

Ilustração baseada no modelo de Wicker (2012).

PG. 02

Figura 2. Representação dos elementos de transposição encontrados nos seres vivos.

Os TE de Classe I e de Classe II diferem de acordo com o mecanismos de mobilização, sendo que os retrotransposon (classe I) envolvem uma cópia de RNA mensageiro e tradução de proteínas no citoplasma da célula, enquanto os transposons (classe II) realizam a transposição envolvendo apenas cópias de DNA e não saem do núcleo da célula. Imagem retirada de Varani *et al.* (2015). PG. 07

Figura 3. Representação do mecanismo de transposição. A) Retrotransposons, ou elementos classe I, realizam o processo conhecido como “copia e cola”. A barra verde escura representa a sequência original, enquanto a barra verde clara representa a copia nova do retrotransposon. B) Transposons, ou elementos classe II, que realizam o processo conhecido como “recorta e cola”. A barra vermelha representa o transposon que é exsistido de uma região e se insere em uma nova região. PG. 09

Figura 4. Estrutura e classificação de retrotransposons com e sem LTR. A) Os retroelementos com LTR (sequência de terminação longa repetida e em orientação direta nas extremidades de cada elemento) podem ser autônomos (i) e não-autônomos (ii). Nos autônomos (i) estão classificados os Ty3/Gypsy, Ty1/Copia e BEL-Pao. Dentre os não-autônomos (ii) temos os LARD e TRIM. B) Retroelementos sem LTR. O retroelemento LINE é autônomo, enquanto os SINE-Alu e SVA são não autônomos. C) Outros retroelementos mais recentes, não classificados em LTR e não-LTR, são os DIRS e Penelope. Em cada TE estão indicadas as regiões típicas de cada elemento: **região TSD:** sequência-alvo com 4 a 6 pb duplicada nas extremidades 5' e 3' de um elemento (é um sinal da inserção de um elemento de transposição), está

repetida e em orientação direta nas extremidades de cada elemento; **gag**: capsídeo, **AP**: protease aspártica, **INT**: integrase, **RT**: transcriptase reversa, **RH**: RNaseH; **PBS**: *Primer Binding Site*, **PPT**: *PolyPurine Tract*, **EN**: endonuclease. No elemento SINE, o monômero esquerdo representa a região derivada de tRNA com sítios A e B homólogos a promotores da RNA polimerase III, enquanto o monômero direito representa a região relacionada a elementos LINE. Tanto o monômeros direito e esquerdo finalizam com uma cauda poli-A. Elemento SVA, (CCCTCC)_n é o hexâmero repetido; o domínio Alu-like consiste de dois fragmentos Alu antissenso (a seta indica a orientação das sequências Alu); o domínio VNTR (*Variable Number Tandem Repeat*); o domínio SINE-R são sequências similares ao gene env, e regiões U3, R, sinal de poliadenilação (3' LTR) e cauda poli-A. Imagem retirada de Varani *et al.* (2015).

PG. 10

Figura 5. Estrutura e organização dos elementos transponíveis de classe II. Os transposons são divididos em duas subclasses, I e II. A) Subclasse I, transposas de corte de dupla fita. Aqui incluem-se o Mariner, hAT, Mutator e elemento P, dentre os autônomos, e o MITE, nos não-autônomos. B) Subclasse II, transposas de corte de fita simples. Aqui incluem-se os Helitrons e Maverik. Em cada TE estão indicadas as regiões típicas de cada elemento: a caixa cinza clara na posição central nos transposons refere-se à transposase. **DR**: repetições diretas (do inglês, *Direct Repeat*), **TIR**: repetições invertidas terminais (do inglês, *Terminal Inverted Repeat*), **Y2**: transposase-HUH, **HEL**: Helicase de DNA, **POL B**: DNA Polimerase, **INT**: transposase-HUH. Imagem editada de Varani *et al.* (2015).

PG. 12

Figura 6. Rearranjos cromossômicos numéricos. A) Evento da aneuploidia podendo ser ascendente ou descendente, B) Displóidia causada por fissão ou fusão

cêntrica, C) Evento de poliploidia causando a duplicação completa do genoma. Imagem elaborado por José Juliano Amorim. PG. 17

Figura 7. Rearranjos cromossômicos estruturais. A) Deleção, note a falta de um pedaço terminal de uma cromátide, B) Duplicação, note a duplicação terminal de uma cromátide, C) Inversão, note os bloco roxos grandes e pequenos invertidos no braço curto do cromossomo, D) Translocação recíproca balanceada entre dois cromossomos não homólogos. Imagem elaborado por José Juliano Amorim. PG. 19

Figura 8. Etapas da análise do DNA repetitivo via RepeatExplorer2. A) O sequenciamento NGS. Em zigzag, tem-se a sequência total do genoma do indivíduo, em verde, vermelho e azul as partes de DNA repetitivo e em preto, os reads gerados pelo NGS. B) A montagem dos cluster por agrupamento dos reads feita pelo Repeat Explorer. As linhas cinzas ao fundo mostram a conexão entre reads. C) Representação dos grafos gerados pela organização dos reads dentro dos clusters. Imagem baseada em ilustração fornecida no curso Workshop Repeat Explorer 2022. PG. 24

Figura 9. Sistemática de Orchidaceae. (A) A família Orchidaceae é dividida em cinco subfamílias (Apostasioideae, Vanilloideae, Cypripedioideae, Orchidoideae e Epidendroideae). (B) A subfamília Epidendroideae (ramos em rosa em A e cladograma em B) é a maior subfamília, contemplando 16 tribos (setas verdes em B) e 70% das espécies de Orchidaceae. Estrela verde indica a posição da subtribo Maxillariinae, na tribo Cymbidiae. Filogenia de Orchidaceae simplificada de Serna-Sánchez *et al.* 2021 e filogenia de Epidendroideae simplificada a partir de Chase *et al.* (2015). Ramos da filogenia em A estão coloridos de acordo com a subfamília segundo legenda. Fotos de flores retiradas de Serna-Sánchez *et al.* (2021) eOrchiBase. PG. 26

Figura 10. Representantes da subtribo Maxillariinae. A-D. *Brasiliorchis*: A: *B. schunkeana*, B: *B. marginata*, C: *B. picta*, D: *B. barbosea*. E-F: *Christensonella*. E: *C. neowiide*, F: *C. uncata*. G-H: *Sauvetea alpestris*. I: *Nitidobulbon*. I: *N. nasutum*, J: *N. proboscideum*. K-M: *Heterotaxis*. K: *H. brasiliensis*, L: *H. equitans*, M: *H. superflua*. N: *Ornithidium vestitum*. O: *Ornithidium fulgens*. P: *Inti bicallosa*. Q-R: *Mapinguari desvauxiniana*. S: *Maxillariella variabilis*. T: *Rhetinantha scorpioidea*. U-X: *Mormolyca*. U: *Mormolyca ringes*, V: *Mormolyca rufecens*, X: *Mormolyca richii*. A'-B': *Trigonidium*. A': *T. acuminatum*, B': *T. obtusum*. C': *Pityphyllum laricinum*. D'-F': *Cyrtidiorchis*. D', E': *C. alata*, F': *C. stumpfleii*. G'-I': *Cryptocentrum*. G' e H': *Cryp. latifolium*, I': *Cryp. roseans* (previamente considerado *Anthosiphon roseans* Schltr). Créditos: A-F, K-N, Q, S-V, A'-B': APMoraes; G, J, O-P, R, X-Z, C'F'' e H' - fotos publicadas em Whitten & Blanco (2011); H - desenho retirado de Pridgeon *et al.* (2009); I, H', I' - fotos retiradas de www.orchidsppecies.com. PG. 29

Figura 11. Morfologia de *Brasiliorchis* R. Singer, S. Koehler e Carnevali. (A) Indivíduo inteiro. Observe os pseudobulbos cônicos bifoliados, com folhas lanceoladas. (A') coluna. (A'') Capuz do polinário. (A''') Polinário. (B) *B. marginata*. (C) *B. picta*. (D) *B. barbosea*. (E) *B. porphyrosteles*. (F e G) *B. schunkeana*, flor e hábito respectivamente. Desenho apresentado em 1A foram modificados a partir de Singer *et al.* (2007). Fotos de B a G foram feitas por APMoraes. PG. 31

Figura 12. Filogenia da subtribo Maxillariinae (A) e do gênero *Brasiliorchis* (B). Suporte de Probabilidade Posterior é indicado pelas cores dos ramos de acordo com a escala. Em A, os insertos mostram as imagens de representantes dos gêneros de Maxillariinae: *Scuticaria steleei*, *Rudolfiella aurantiaca*, *Bifrenaria leucorhoda*, *Anguloa uniflora*, *Lycaste macrobulbon*, *Xylobium variegatum*, *Heterotaxis equitans*, *Mapinguari desvauxianus*, *Christensonella uncata*, *Mormolyca ringes*, *Rhetinantha*

cerifera, *Maxillariella variabilis*, *Trigonidium acuminatum*, *Maxillaria leucaimata*, *Camaridium ochroleucum*, *Brasiliorchis picta* e *Brasiliorchis barbosea*. Em B) inserto da filogenia de *Brasiliorchis* com imagens de *Brasiliorchis picta* e *Brasiliorchis barbosea*. Fotos de A. P. Moraes and J. S. Cabral. Imagem editada de Moraes *et al.* 2022. PG 33

Figura 13. Gráfico de pizza mostrando as proporções das diferentes categorias anotadas na análise genômica individual por espécie. A) *Brasiliorchis schunkeana* e B) *Brasiliorchis picta*. Estrela indica a fração anotada como DNAsat nas duas espécies. PG. 39

Figura 14. Gráfico de pizza mostrando as proporções dos diferentes TEs anotadas na análise genômica individual por espécie. A) *Brasiliorchis schunkeana* e B) *Brasiliorchis picta*. PG. 40

Figura 15. Análise comparativa entre as espécies *Brasiliorchis schunkeana* e *Brasiliorchis picta*. A) Tamanho em pares de bases dos clusters anotados em *B. schunkeana* e *B. picta*. B) Frequência relativa dos clusters dentro de cada espécie, plotados de forma comparativa entre as espécies *B. schunkeana* (linha superior) e *B. picta* (linha inferior), a seta em vermelho indica o destaque nos três clusters que foram identificados. A anotação dos cluster está indicada na legenda de cores na porção superior direita da imagem. Retângulo vermelho sinaliza os cluster com maior diferença entre espécies, em roxo são os clusters anotado como DNA satélite. PG. 44

Figura 16. Análise *dot-plot* e grafo dos três clusters com distribuição diferencial entre as espécies. A) CL32, B) CL51, C) CL67. Na linha superior estão apresentados os gráficos gerados na análise de *dot-plot*, enquanto na linha inferior, para cada cluster, estão indicados os grafos gerados pela ferramenta Seqgraph no RE2. PG. 46

LISTA DE TABELA

Tabela 1. Percentagens das diferentes categorias anotadas nos genomas de *Brasiliorchis schunkeanae* e *Brasiliorchis picta*. Para cada espécie são apresentadas as percentagens de cada categoria balizadas pelo total do genoma anotado (coluna Genoma), total do repetoma, excluindo-se os rDNA (coluna Repetoma) e total de elementos transponíveis anotados (coluna TE). PG 41

SUMÁRIO

1	Introdução Geral	1
1.1	O que é genoma?	1
1.2	As sequências repetitivas de DNA	3
1.3	Tamanho de Genoma	13
1.4	O cariótipo	15
1.4.1	Alterações Cariotípicas	16
1.4.2	Cariotipagem via citogenética clássica e molecular	20
1.5	Como analisar as sequências genômicas?	21
1.6	<i>Brasiliorchis</i> como modelo de estudo	25
2	Objetivos	34
3	Material e Métodos	35
3.1	Material	35
3.2	Método: Análises no Repeat Explorer 2	35
4	Resultados	37
4.1	Análise individual	37
4.2	Análise Comparativa	43
5	Discussão	47
5.1	A composição de elementos transponíveis em plantas e no modelo de estudo	47
5.2	A composição de DNA satélite nos genomas vegetais	50
6	Conclusão	53
7	Referências	54
8	Material suplementar	67

1. Introdução Geral

1.1 O que é genoma?

O termo genoma, criado em meados de 1920 por Hans Winkler, frequentemente é referido, e entendido, como conjunto de genes de um organismo, porém o genoma é mais que um simples agrupamento de genes (Nick 2022). O genoma, no seu sentido mais amplo, refere-se a todo material genético presente no núcleo da célula, assim como nas organelas mitocôndria (genoma mitocondrial) e plastídeo (genoma plastídial) (Gregory 2005). O estudo da organização e composição do genoma permite visualizar facilmente a diversidade e complexidade dos diferentes tipos de sequências que compõem o núcleo (Fig. 1; Gregory 2005). Enquanto o conjunto de genes (*ie.*, sequências de nucleotídeos codificantes, que serão transcritas a RNA e, posteriormente, traduzidas à proteína; Krebs 2017, Nick 2022) responde por uma pequena fração do genoma (aproximadamente 1.5-2% do genoma), a maior parte do genoma é composto por sequências não-codificantes, especialmente as sequências repetitivas (Fig. 1).

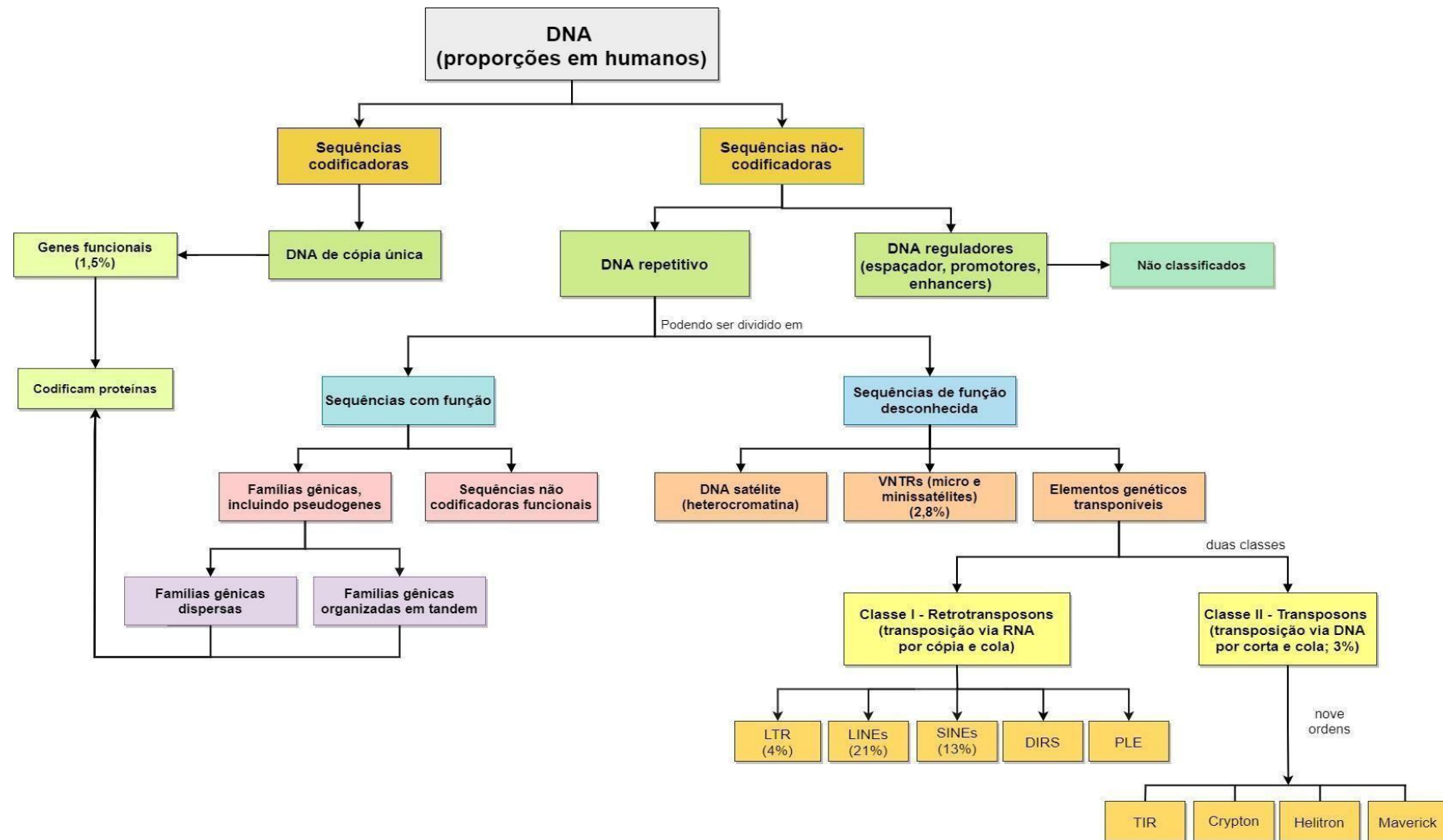


Figura 1. Representação esquemática da composição do genoma dos eucariotos. Ilustração baseada no modelo Wicker (2012).

1.2 As sequências repetitivas de DNA

As sequências não codificadoras abrangem as sequências reguladoras, os espaçadores gênicos e, principalmente, as sequências de DNA repetitivo que podem ser classificadas com base em algumas características como o tamanho da unidade de repetição, a organização e a frequência destas unidades no genoma. Por exemplo, os microssatélites e os minissatélites diferem em tamanho da unidade de repetição e frequência de repetição: enquanto os microssatélites são repetições extremamente curtas de até 10 pares de bases (pb) repetidas poucas vezes, os minissatélites podem ter até algumas dezenas de pares de bases (<100pb) repetidas dezenas de vezes (Biscotti *et al.* 2015, Krebs 2017). Entretanto, essas duas categorias se assemelham quanto a organização das repetições: ambas são organizadas em tandem (*ie*, uma unidade de repetição seguida da outra). A terceira categoria organizada em tandem é o DNA satélite (DNAsat), classe de DNA bastante frequente nas regiões de heterocromatina constitutiva, podendo ser classificado como altamente repetitivo (Fig. 1), representando um importante componente do genoma eucarioto. Paralelamente, outra classe muito importante de DNA repetitivo no genoma eucarioto são os elementos transponíveis (do inglês, *Transposable Elements* - TE; Wicker *et al.* 2007, 2012). Entretanto, estas sequências, diferentemente dos DNAsat, estão organizados de forma dispersas no genoma, sem formar blocos conspícuos nos cromossomos (Gregory 2005).

1.2.1 As sequências de DNA satélite

Os primeiros trabalhos sobre o DNAsat foram apresentados em 1960 por Kit *et al.* e por Seuoka *et al.* Os autores, ao estudarem a cinética de reassociação do DNA, observaram que após centrifugar o DNA nuclear em um gradiente de cloreto de cério

(CsCl) havia a formação de duas bandas de tamanhos distintos e, para cada uma das bandas, foi determinado um nome: banda maior foi chamada de adjacente e a banda menor, de satélite (Guerra 1988). Em um mesmo intervalo de tempo, esta última banda apresentava taxas mais elevadas de renaturação do DNA em comparação à banda maior, o que evidenciava a presença de sequências de nucleotídeos repetidas. A maioria das famílias de DNAsat são compostas por sequências não-codificantes com evolução rápida (Biscotti *et al.* 2015). Essas sequências têm um papel potencial importante na estabilidade do genoma e no reconhecimento de homólogas na meiose I, garantindo a segregação cromossômica e graças a evolução rápida, gera diferenças genômicas entre espécies/gêneros (Biscotti *et al.* 2015, 2017), o que atribui a essa fração do genoma um potencial para estudos taxonômicos e evolutivos (Sousa 2022, Ibiapino *et al.* 2022).

Enquanto os estudos de desnaturação e reassociação (renaturação) do genoma permitiram identificar a existência de regiões com elementos repetitivos, o aprimoramento de técnicas de biologia molecular como, por exemplo, o estabelecimento da técnica de reação em cadeia da polimerase (do inglês, *Polymerase Chain Reaction* - PCR) e, mais atualmente, o uso das ferramentas de bioinformática associada com a hibridização *in situ* têm permitido a identificação, a organização e a localização de tais elementos no genoma e nos cromossomos. O DNAsat normalmente constitui regiões de heterocromatina constitutiva de modo que as sequências não ficam dispersas no genoma, mas formam blocos bem conspícuos (Mehrotra *et al.* 2014, Garrido 2015, 2017).

O termo heterocromatina foi definido inicialmente por Emil Heitz em 1928, que a denominava como “cromatina que não condensa no final da telófase” (Heitz 1928). A cromatina pode ser dividida em duas categorias de acordo com sua

composição e organização: a eucromatina e a heterocromatina. A separação nestas duas categorias se dá com base na compactação da cromatina na interfase, quando os filamentos de cromatina estão menos condensados e disponíveis para transcrição. Se a cromatina estiver “frouxa” e acessível à transcrição, chamamos de eucromatina, porém, a fração que mesmo na intérfase está permanentemente condensada e indisponível para transcrição, chamamos heterocromatina (Guerra 1988).

Entretanto a heterocromatina ainda pode ser subdividida em duas categorias: facultativa ou constitutiva. Enquanto a primeira apresenta um comportamento alternado: em alguns momentos do desenvolvimento se comporta como heterocromatina e em outros, como eucromatina, a segunda permanece condensada ao longo de todo ciclo celular em todas as fases do desenvolvimento do organismo. Porém, quando se fala em heterocromatina, normalmente nos referimos à essa última categoria: heterocromatina constitutiva. Esta fração apresenta algumas características típicas, como: (1) encontrada principalmente nas regiões pericentromérica e telomérica do cromossomo, (2) composta por DNA repetitivo, especialmente DNAsat, e por isso, é pobre em genes e apresenta baixa atividade transcricional, (3) normalmente é rica em AT ou em CG e (4) apresenta replicação tardia. Esta fração costuma formar blocos conspícuos nos cromossomos, os chamados blocos heterocromáticos (Elgin & Shiv 2003). A baixa atividade transcricional observada na heterocromatina está associada à própria organização desta fração do genoma: por formar alças pequenas na cromatina, a heterocromatina produz a aparência distinta e típica da heterocromatina nos cromossomos mitóticos (Fitz-James *et al.* 2020). A partir da descoberta da heterocromatina não foi difícil compreender a relação positiva entre presença de blocos heterocromáticos e tamanho do genoma e as consequências evolutivas essa relação pode trazer para uma espécie.

1.2.2 Os elementos transponíveis

O segundo grupo de sequências altamente repetitivas são os TEs, que encontram-se dispersos pelo genoma, sem formar blocos conspícuos nos cromossomos (Gregory 2005). Os TEs foram descritos inicialmente por Barbara McClintock em seus estudos sobre genoma do milho, já na década de 40 (McClintock 1950). A pesquisadora sugeriu que as sequências eram capazes de se movimentar no genoma, saltando de uma região para outra, o que não foi prontamente aceito na comunidade científica, mas com o tempo comprovou-se verdadeiro. Os TEs têm capacidade de se transpor entre regiões não homólogas do genoma, fazendo, ou não, cópias adicionais de si mesmas (Biémont 2010). Tais sequências são denominadas “*Jumping genes*” (genes saltadores; McClintock 1950) ou “*Selfish DNA*” (DNA egoísta), pois o conhecimento inicial sobre TEs sugeria que tais sequências se alastravam pelo genoma sem contribuir efetivamente para a atividade gênica da espécie, sendo caracterizados como parasitas do genoma graças a sua atividade proliferativa e mutacional (Doolittle & Sapienza 1980, Orgel & Crick 1980). Contudo, atualmente sabe-se que os TEs constituem componentes importantes dos genomas eucarióticos, impactando a regulação da expressão gênica, afetando as funções enzimáticas, a plasticidade fenotípica dos organismos e suas respostas ao estresse ambiental biótico e abiótico (Kazazian 2004, Varani *et al.* 2015, Bourque *et al.* 2018). Esses fatores estão todos interligados, pois a expansão/retração dos TEs gerada como resposta ao estresse causado por fatores abióticos, como variação de altitude, hábitat e hábito pode afetar a regulação gênica, podendo levar a mudanças fenotípicas nos organismos (Hsu *et al.* 2020). Essa mudança de paradigma, faz que hoje os TEs sejam considerados *facilitadores da evolução* (Biémont 2010, Mhiri *et al.* 2022). Dada a grande diversidade dos TEs, estes são classificados em classes, ordens e famílias. No

primeiro nível, os TEs são organizados em duas grandes classes de acordo com o mecanismo de transposição (Gregory 2005, Wicker *et al.* 2007): os retrotransposons (classe I) e os transposons (classe II) (Fig. 2).

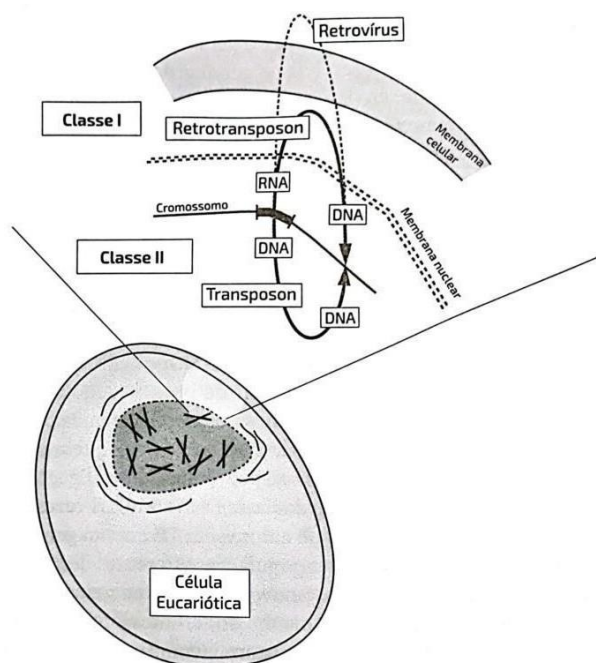


Figura 2. Representação dos elementos de transposição encontrados nos seres vivos. Os TE de Classe I e de Classe II diferem de acordo com o mecanismos de mobilização, sendo que os retrotransposon (classe I) envolvem uma cópia de RNA mensageiro e tradução de proteínas no citoplasma da célula, enquanto os transposons (classe II) realizam a transposição envolvendo apenas cópias de DNA e não saem do núcleo da célula. Imagem retirada de Varani *et al.* (2015).

Os retrotransposons, os elementos transponíveis de classe I, realizam a transposição autônoma e não autônoma, mas sempre envolvendo um RNA mensageiro (RNAm) intermediário para que essa sequência se locomova (Fig. 2). Durante o processo de transcrição do retrotransposon é gerado um RNAm que é direcionado ao citoplasma, onde é traduzido um conjunto de enzimas relacionadas com a transposição do TE e todas essas enzimas são codificadas pelo próprio retroelemento. Uma das enzimas codificadas é a transcriptase reversa (do inglês, *Reverse Transcriptase* - RT; Fig. 2 e 3a), responsável pela síntese da nova cópia de

DNA do retroelemento. Essa cópia de DNA, junto das enzimas traduzidas, como a integrase (ver INT na Fig. 4) que auxiliá-ra no processo bioquímico de mobilização com a inserção do retroelemento em uma nova posição do genoma, são deslocadas ao núcleo da célula para realizar o processo (Varani *et al.* 2015). Esse processo é conhecido como "copia e cola" ou "transposição replicativa". O retroelementos são encontrados em grande quantidade no genoma, sendo considerados os principais contribuintes da fração repetitiva dos genomas vegetais (Varani *et al.* 2015).

Por sua vez, os transposons, os elementos transponíveis de classe II, são TEs de DNA, que também podem ser autônomos ou não-autônomos. Entretanto, esses TEs, diferente dos retrotransposons, utilizam o processo de "recorta e cola" para sua movimentação (Fig. 2 e 3b). Neste caso, a sequência é excisada de um ponto e movida para outra região dentro do genoma. Para tal é necessário o auxílio da enzima transposase (ver Tpase na Fig. 4), codificada pela célula hospedeira, sendo que a inserção da sequência na nova região é feita também pela enzima integrase (ver INT na Fig. 4).

Dentro de cada uma das classes, os TE podem ser classificados em ordens, com cada ordem apresentando as suas famílias (Wicker *et al.* 2007, Varani *et al.* 2015; Fig. 1 e Fig. 4). Os retrotransposon podem ser classificados de acordo com a presença (Fig. 4A) ou ausência (Fig. 4B) de longas sequências terminais em suas extremidades, os LTR (do inglês, *Long Terminal Repeat*). Dentro dos LTRs, os Ty3/Gypsy e o Ty1/Copia, ambos autônomos, são os mais representativos nos genomas (Fig. 4A), enquanto entre os não-LTR temos os LINEs (do inglês, *Long Interspersed Nucleotide Elements*) e SINEs (do inglês, *Short Interspersed Nuclear Elements*) (Fig. 4B). Outros elementos de classe I recentemente descobertos são os PLE (do inglês,

Penelope-Like Elements) e DIRS (do inglês, *Dictyostelium Intermediate Repeat Sequence*) (Varani *et al.*, 2015; Fig. 4C).

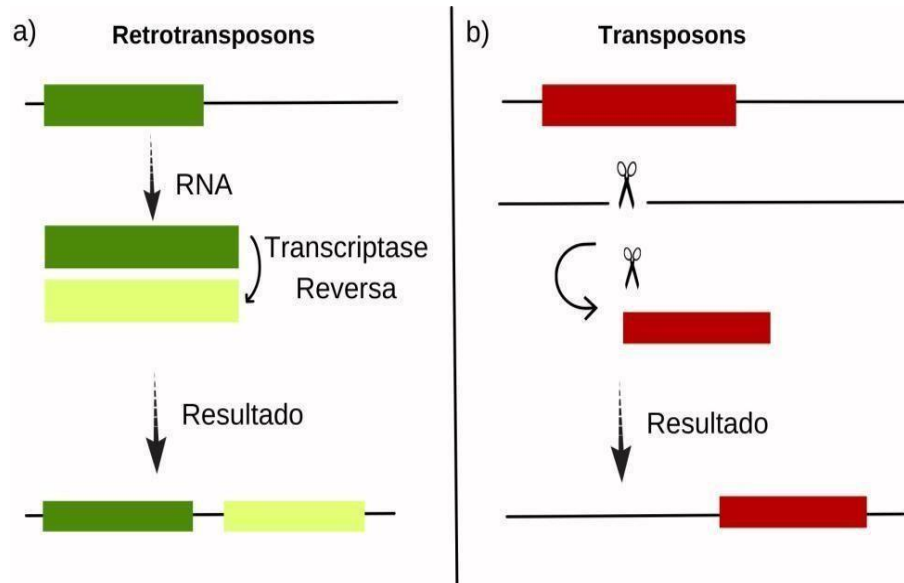


Figura 3. Representação do mecanismo de transposição. A) Retrotransposons, ou elementos classe I, realizam o processo conhecido como “copia e cola”. A barra verde escura representa a sequência original, enquanto a barra verde clara representa a cópia nova do retrotransposon. B) Transposons, ou elementos classe II, que realizam o processo conhecido como “recorta e cola”. A barra vermelha representa o transposon que é exsistado de uma região e se insere em uma nova região.

Classe I - Retrotransposons

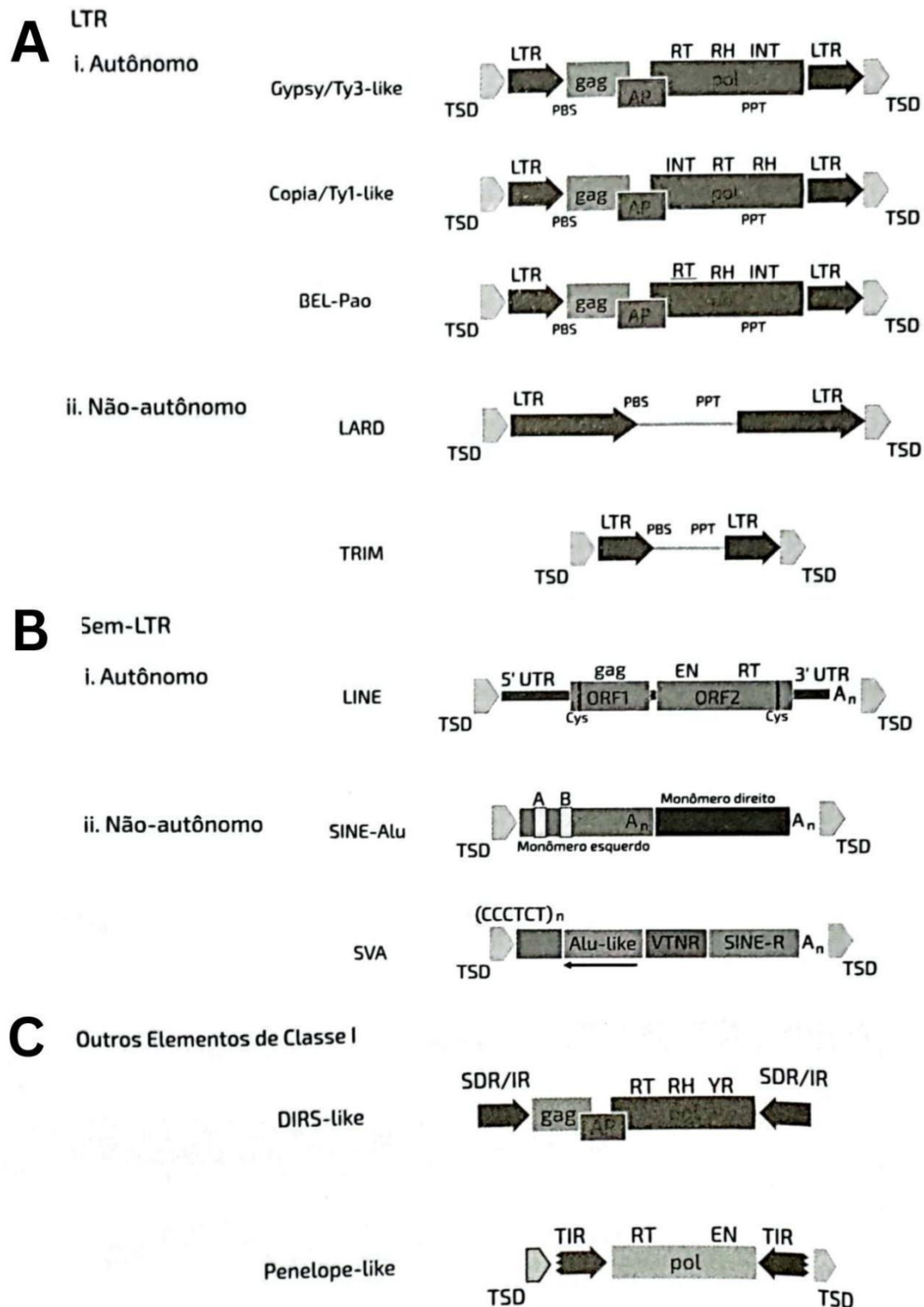


Figura 4. Estrutura e classificação de retrotransposons com e sem LTR. (continua na próxima página)

A) Os retroelementos com LTR (sequência de terminação longa repetida e em orientação direta nas extremidades de cada elemento) podem ser autônomos (i) e não-autônomos (ii). Nos autônomos (i) estão classificados os Ty3/Gypsy, Ty1/Copia e BEL-Pao. Dentre os não-autônomos (ii) temos os LARD e TRIM. B) Retroelementos sem LTR. O retroelemento LINE é autônomo, enquanto os SINE-Alu e SVA são não autônomos. C) Outros retroelementos mais recentes, não classificados em LTR e não-LTR, são os DIRS e Penelope. Em cada TE estão indicadas as regiões típicas de cada elemento: **região TSD**: sequência-alvo com 4 a 6 pb duplicada nas extremidades 5' e 3' de um elemento (é um sinal da inserção de um elemento de transposição), está repetida e em orientação direta nas extremidades de cada elemento; **gag**: capsídeo, **AP**: protease aspártica, **INT**: integrase, **RT**: transcriptase reversa, **RH**: RNaseH; **PBS**: *Primer Binding Site*, **PPT**: *PolyPurine Tract*, **EN**: endonuclease. No elemento SINE, o monômero esquerdo representa a região derivada de tRNA com sítios A e B homólogos a promotores da RNA polimerase III, enquanto o monômero direito representa a região relacionada a elementos LINE. Tanto o monômeros direito e esquerdo finalizam com uma cauda poli-A. Elemento SVA, (CCCTCC)_n é o hexâmero repetido; o domínio Alu-like consiste de dois fragmentos Alu antissenso (a seta indica a orientação das sequências Alu); o domínio VNTR (*Variable Number Tandem Repeat*); o domínio SINE-R são sequências similares ao gene env, e regiões U3, R, sinal de poliadenilação (3' LTR) e cauda poli-A. Imagem retirada de Varani *et al.* (2015).

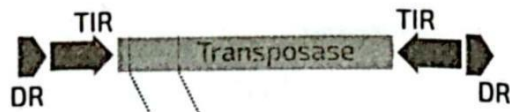
Por sua vez, os transposons são classificados em duas subclasses de acordo de como a transposase corta o DNA no momento da inserção do transposon (Fig. 5). Enquanto que nos transposons da subclasses I, a transposase realiza corte nas duas fitas de DNA no momento da inserção, nos transposons da subclasse II, a transposase realiza um único corte na fita de DNA no momento da inserção do transposon (Varani *et al.* 2015). Exemplo da subclasse I é o transposons TIR (do inglês, *Terminal Inverted Repeats*) e MITE, e da subclasse II, os transposons Helitron, muito presentes em plantas, e Maverick, ausentes em plantas (Kazazian 2004, Wicker *et al.* 2007, Varani *et al.* 2015; Fig. 1 e Fig. 5).

Classe II - Transposons

Subclasse I - corta-e-cola

i. Autônomo

Tc1-Mariner
hAT
Mutator
Elemento P



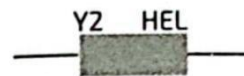
ii. Não-autônomo

MITE



Subclasse II - copia-e-cola

Helitron



Maverick

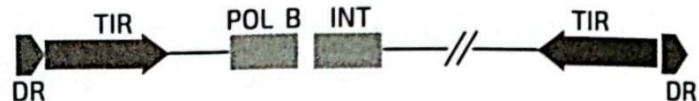


Figura 5. Estrutura e organização dos elementos transponíveis de classe II. Os transposons são divididos em duas subclasses, I e II. A) Subclasse I, transposas de corte de dupla fita. Aqui incluem-se o Mariner, hAT, Mutator e elemento P, dentre os autônomos, e o MITE, nos não-autônomos. B) Subclasse II, transposas de corte de fita simples. Aqui incluem-se os Helitrons e Maverik. Em cada TE estão indicadas as regiões típicas de cada elemento: a caixa cinza clara na posição central nos transposons refere-se à transposase. **DR:** repetições diretas (do inglês, *Direct Repeat*), **TIR:** repetições invertidas terminais (do inglês, *Terminal Inverted Repeat*), **Y2:** transposase-HUH, **HEL:** Helicase de DNA, **POL B:** DNA Polimerase, **INT:** transposase-HUH. Imagem editada de Varani *et al.* (2015).

1.3 Tamanho de Genoma

No final da década de 1940, com o trabalho de Boivin, R. Vendrely e C. Vendrely em tecidos bovinos, surgiram os primeiros estudos de quantificação do DNA, pois mesmo antes da conclusão dos estudos sobre estrutura do DNA, os pesquisadores já se interessavam em saber a quantidade de material genético presente nos genomas das espécies (Ogur *et al.* 1951). A quantidade total de DNA do complemento cromossômico gamético não replicado de uma célula é dito o tamanho de genoma (do inglês, *Genome Size* - GS; Levin 2002), o qual pode ser medido em picogramas (pg, g) ou em megapares de base (1 pg = 978 Mpb). O tamanho de genoma pode ser referido como o valor-C, termo determinado por Swift em 1950. Desde os primeiros estudos em animais e plantas (Ris & Mirsky 1949, Swift 1950), o conteúdo de DNA entre diferentes indivíduos de uma mesma espécie sempre foi constante, de forma que Swift relata que a escolha pela letra 'C' é fazer menção a esta constância observada do tamanho de genoma dentro da espécie. Tal constância teve grande importância no início das hipóteses sobre o papel hereditário do DNA observada por Vendrely e Vendrely (1948).

No início dos estudos sobre tamanho de genoma, a falta de correlação do valor-C com a complexidade biológica e/ou com o número de genes do organismo intrigou os pesquisadores, gerando dúvidas que compunham o “paradoxo do valor-C”: porquê nem sempre as espécies mais complexas apresentam os maiores genomas? E porquê espécies intimamente relacionadas, sem grandes diferenças de comportamento e desenvolvimento, podem apresentar ampla variação do valor-C? (Thomas 1993, MacGregor 2013). Entretanto, ao se descobrir que o genoma é composto em grande parte por DNA repetitivo não codificante, inclusive com correlação positiva entre tamanho de genoma e tamanho de íntrons (Vinogradov 1999, Levin 2002), foi

possível esclarecer que a complexidade das espécies não tem relação com o tamanho de genoma, solucionando o paradoxo do valor-C (MacGregor 2013). Por exemplo, o compacto genoma de *Arabidopsis thaliana* (L.) Heynh. (1C = 0,16pg) contém 27.000 genes e 20 a 25 Mb de retrotransposons (Bennet & Leitch 2005, The Arabidopsis Genome Initiative 2000), enquanto que o genoma de milho (*Zea mays* L.), 17 vezes maior do que o de *A. thaliana* (1C = 2,73pg), apresenta apenas o dobro de genes (40.000 genes) e 90 vezes mais retrotransposons (1800 Mb; Wicker 2012).

As variações do tamanho de genoma influenciam em diferentes características morfológicas, fisiológicas e ecológicas das plantas, tais como o tamanho da célula, do núcleo e no tempo do ciclo celular (Levin 2002) e, dessa forma, tais variações impactam o crescimento e o desenvolvimento desses organismos (Beaulieu *et al.* 2008, Knight & Beaulieu 2008, MacGregor 2013). Desta forma, o GS se destaca como uma característica importante para compreensão da dinâmica evolutiva das plantas, podendo até mesmo prever como diferentes espécies podem responder aos efeitos das mudanças climáticas e outros fatores ambientais (Beaulieu *et al.* 2008, Knight & Beaulieu 2008).

As mudanças no GS são causadas por dois mecanismos: (1) a poliploidia e (2) a expansão/retração das sequências de DNA repetitivo (Gregory 2005). A poliploidia, a duplicação do genoma completo levando à duplicação do número de cromossomos, é considerado o principal mecanismo na evolução das plantas (Jiao 2011, Jiao *et al.* 2018, Magallón *et al.* 2019). Tal evento de duplicação pode fornecer novas oportunidades para neofuncionalização e subfuncionalização dos genes, aumentando a diversidade genética, além de gerar novos padrões de expressão genética (Chen 2007). Além disso, a poliploidia também pode estar associada ao controle epigenético, que é responsável por regular vários processos que estão envolvidos na regulação do

crescimento e desenvolvimento das plantas (Van der Peer *et al.* 2021). Por fim, a sugere-se que a poliploidia possa proporcionar uma vantagem adaptativa a ambientes adversos (Pellicer *et al.* 2018, Moraes *et al.* 2022).

O segundo mecanismo é a amplificação/retração das sequências repetitiva de DNA, isso pode afetar tanto o DNAsat, como os elementos transponíveis, sendo que na maioria das plantas são os TEs que afetam mais frequentemente a variação de GS (por exemplo, Ribeiro *et al.* 2021, 2022, Ibiapino *et al.* 2022, Samoluk *et al.* 2022). Sendo assim, a amplificação ou retração das sequências repetitivas de DNA pode levar a alterações no tamanho dos cromossomos, o que pode ser identificado pela análise do cariótipo (Guerra 2008).

1.4 O cariótipo

Como já revisado, o genoma nuclear completo de uma célula é composto por todo DNA armazenado no núcleo. Este DNA é organizado nos cromossomos sendo que cada cromossomo é uma dupla-fita de DNA. Assim, pode-se dizer que os cromossomos são moléculas de DNA altamente condensadas e associadas a moléculas proteicas e de RNA. A quantidade de cromossomos varia de espécie para espécie e pode, em alguns casos, ter até mesmo importância taxonômica (Guerra 2012).

O número de cromossomos é o carácter citogenético mais simples e econômico de ser obtido e, por representar o conjunto completo de pares de cromossomos, o cariótipo tem papel fundamental em estudos evolutivos. Por meio de rearranjos cromossômicos, também conhecidos como mutações cromossômicas, é possível a obtenção de caracteres específicos, sejam eles morfológicos, comportamentais ou bioquímicos. Estas alterações têm seu papel fundamental na

diversificação de espécies e na evolução de linhagens, caracterizando-se como elementos essenciais para a compreensão da biologia evolutiva (Mayrose *et al.* 2021).

1.4.1 Alterações Cariotípicas

As alterações cromossômicas podem ser numéricas, quando ocorre alteração no número de cromossomos estruturalmente normais (Fig. 6) ou podem ser estruturais, quando afeta a estrutura do cromossomo (Fig. 7) (Guerra 2008). Em suma, as alterações numéricas podem ser euploides ou aneuploides, dependendo se as células têm conjuntos completos ou incompletos de cromossomos. Assim, são ditos euploides quando as células apresentam conjuntos completos de cromossomos e, desta forma, são considerados euploides desde as células com número haplóide (células gametofíticas, n), as diplóides (esporofíticas, $2n$), até os múltiplos de n , e nesse caso as células são ditas poliplóides (Fig. 6A). Dentro de poliplóides há dois tipos básicos: (1) autopoliploides, quando há duplicação do genoma de uma mesma espécie, e (2) alopoliplóides, aqueles provenientes de hibridização interespecífica. Aneuploidia é a mudança no número de cromossomos, mas sem envolver o complemento completo, ou seja, ganho ou perda de um, ou poucos cromossomos (Fig. 6B). Dentre as aneuploidias podemos citar as monossomia quando perde-se apenas um cromossomo ($2n-1$); trissomia, quando ganha-se um cromossomo ($2n+1$); e nulissomia, quando perde-se os dois cromossomos de um par cromossômico ($2n-2$). As possíveis origens dos aneuploides são a não-disjunção dos cromossomos durante a meiose ou mitose, tendo em conta que, no caso da mitose, devido a um erro na divisão mitótica do zigoto, são geradas duas ou mais linhagem celular de nome mosaicismo cromossômico (Mayrose *et al.* 2021).

Por outro lado, os rearranjos estruturais afetam a estrutura do cromossomo sem, na maioria das vezes, afetar o número cromossômico. Porém é importante realçar algumas peculiaridades aqui: exceção deve ser feita às fusões e fissões cromossômicas causadas pela translocação Robertsoniana envolvendo cromossomos acrocêntricos e também às fissões e fusões cêntricas (Fig. 6C), que apesar de comporem rearranjos estruturais levam à mudança de número cromossômico classificada como disploidia (Fig. 6C).

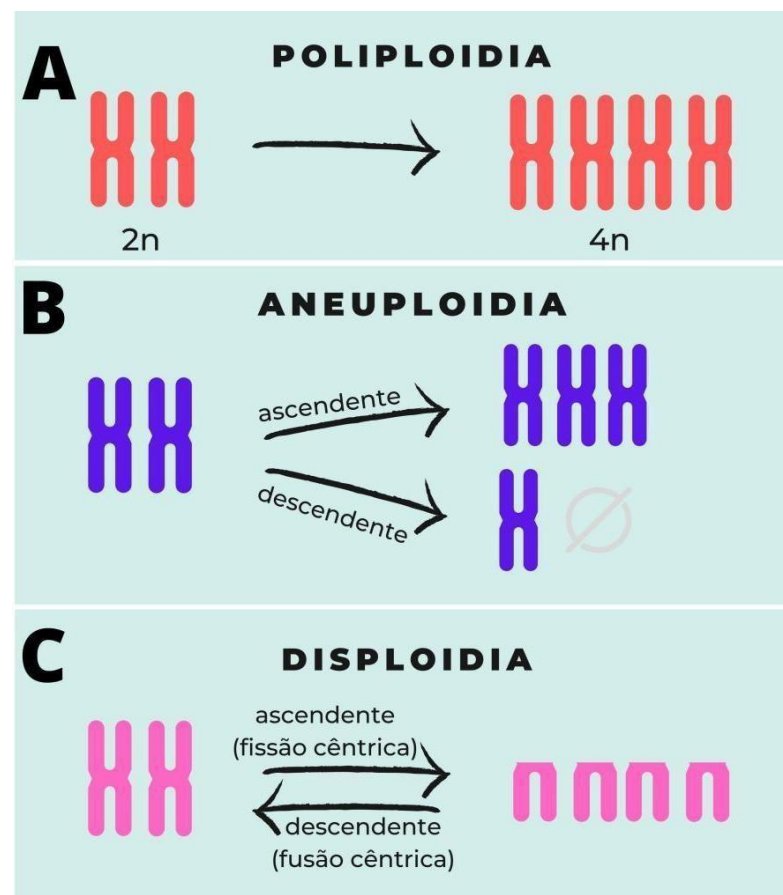


Figura 6. Rearranjos cromossômicos numéricos. A) Evento de poliploidia causando a duplicação completa do genoma B) Evento da aneuploidia podendo ser ascendente ou descendente, C) Disploidia causada por fissão ou fusão cêntrica. Por José Juliano Amorim.

Os rearranjos estruturais podem ser classificados como rearranjos balanceados e não balanceados: enquanto os rearranjos balanceados não afetam a quantidade de material genético, os rearranjos não balanceados produzem uma mudança na quantidade de material genético e geralmente tem consequências fenotípicas mais graves.

Nos rearranjos cromossômicos temos os quatro tipos básicos (Fig. 7): (1) deleção, resultando na perda de um segmento cromossômico (Fig. 7A), (2) duplicação, ocorre quando temos uma repetição do material genético, normalmente após uma deleção já que o fragmento que foi perdido é conectado a cromátide irmã (Fig. 7B), (3) inversão, quando o cromossomo é quebrado em duas partes e o segmento intermediário faz um giro em um ângulo de 180° e volta a ligar-se ao restante do cromossomo, e as inversões podem envolver o centrômero (chamadas de inversões pericêntricas) ou não (inversões paracêntricas; Fig. 7C), e por fim temos a (4) translocação, quando fragmento de um cromossomo é transferido para outro cromossomos não homólogo (Fig. 7D) (Schubert & Vu 2016). As translocações podem ser subdivididas em três tipos: translocação simples, translocação dupla e translocação complexa. Na translocação simples, um fragmento de um cromossomo é transferido para outro cromossomo não homólogo (Fig. 7D). Na translocação dupla, dois fragmentos de cromossomos diferentes são trocados. Na translocação complexa, mais de dois fragmentos de cromossomos são implicados. Entretanto, um último tipo pode ser elencado: a translocação Robertsoniana que resulta da junção de dois cromossomos acrocêntricos, ou seja, cromossomos que possuem seu centrômero bem próximo do seu extremo. Esta translocação é considerada um tipo de disploidia, pois causa alteração de número de cromossomos mas sem alterar o patrimônio genético (Guerra 2012).

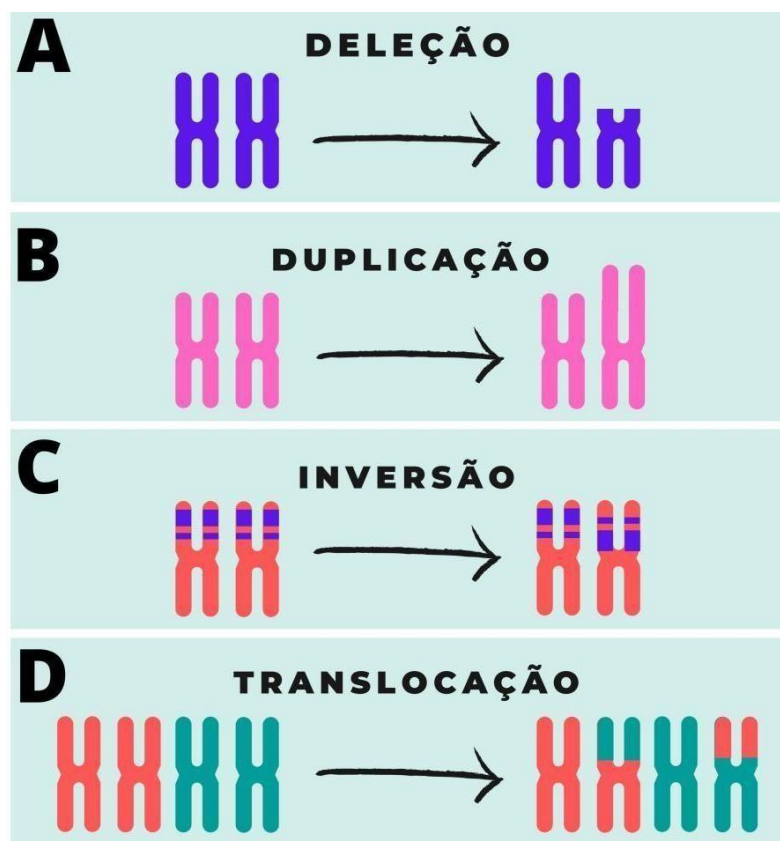


Figura 7. Rearranjos cromossômicos estruturais. A) Deleção, note a falta de um pedaço terminal de uma cromátide, B) Duplicação, note a duplicação terminal de uma cromátide, C) Inversão, note os bloco roxos grandes e pequenos invertidos no braço curto do cromossomo, D) Translocação recíproca balanceada entre dois cromossomos não homólogos. Imagem elaborado por José Juliano Amorim.

Os TE, vistos no tópico anterior, também podem facilitar, e até promover a ocorrência de rearranjos cromossômicos estruturais, em especial translocações. Como revisado por Garcia & Gaiesky (2015), há dois possíveis mecanismos pelos quais rearranjos cromossômicos associados a TE podem originar-se. O primeiro mecanismo é a recombinação ectópica. O TE apresenta múltiplas cópias muito semelhantes, se não idênticas, ao longo do genoma do hospedeiro e tais cópias podem promover recombinação ectópica entre essas diferentes cópias de TEs não-homólogas. Já o segundo mecanismo origina rearranjos cromossômicos por ação direta do TE, em decorrência do próprio evento de transposição. Neste segundo mecanismo as regiões

terminais dos TEs envolvidos são provenientes de elementos distintos (dando origem a um elemento híbrido), ao invés de pertencerem à mesma cópia de um elemento, como na transposição tradicional. Esse mecanismo também é conhecido como transposição aberrante, não convencional ou alternativa, e até agora tem sido descrito apenas para os elementos de Classe II (os quais utilizam um mecanismo de corte em fita dupla para se transpor). Quando os elementos envolvidos se encontram em cromátides irmãs ou em cromossomos homólogos, o mecanismo denomina-se Transposição entre Cromátides-Irmãs; já quando o mecanismo envolve elementos que se encontram na mesma cromátide, denomina-se Transposição dos Terminais Invertidos (Garcia & Gaiesky 2015).

1.4.2 Cariotipagem via citogenética clássica e molecular

Tais rearranjos estruturais podem ser identificados via marcas nos cromossomos, como as bandas cromossômicas (Guerra 2008, Moraes *et al.* 2016). As bandas cromossômicas são formadas por regiões onde o material genético de cada cromátide está altamente condensado afetando o padrão de coloração da região, que pode ser parcialmente ou totalmente visíveis ao microscópio. E essas bandas são úteis na identificação dos cromossomos homólogos, pois estes possuem as mesmas características morfológicas (Guerra 2008). A identificação é feita via técnicas de coloração diferencial, como as técnicas de bandejamento. Dentre tais técnicas, o bandejamento CMA/DAPI (cromomicina A3 / 4',6-diamidino-2-fenilindole) revela-se muito informativo: além de revelar o padrão de distribuição de heterocromatina, permite também classificá-la como rica em bases C e G (citosina e guanina – bandas CMA⁺) ou rico em bases A e T (adenina e timina – bandas DAPI⁺) (Schweizer 1976). Por exemplo, em orquídeas este bandejamento tem revelado um grande polimorfismo

de bandas (de Oliveira *et al.* 2015, Moraes *et al.* 2012, 2013, 2016, 2017), permitindo até mesmo indicar pontos de quebras e/ou fusões cromossômicas (Cabral *et al.* 2006, Moraes *et al.* 2016).

Porém, a hibridização *in situ* fluorescente (do inglês, *Flourescent In Situ Hybridization* - FISH; Jiang 2019) permite localizar sequências de DNA diretamente nos cromossomos, inserido mais um nível na caracterização das sequências e permitindo um refinamento final da análise (por exemplo, veja Ribeiro *et al.* 2022, Ibiapino *et al.* 2022). Da mesma forma, é possível realizar o reconhecimento de homólogos com alto refinamento da montagem do cariótipo (p.ex., Moraes *et al.* 2008, Mendes *et al.* 2011).

1.5 Como analisar as sequências genômicas?

Os genomas vegetais são muito diversos e variam de tamanho desde ~10 Mb em algas até entre ~60 Mb e ~150 Gb em plantas terrestres (revisado por Kersey 2019). Cada dia mais estudos mostram que os DNAs repetitivos são importantes para processos evolutivos, sendo que tais sequências podem ser usadas como marcadores genômicos e mapeadas diretamente nos cromossomos (Ibiapino *et al.* 2022). Entretanto, o primeiro passo é a caracterização dessas sequências no genoma da espécie foco de estudo. Este passo tem sido facilitado pelo sequenciamento genômico massivo (ie., sequenciar o genoma completo), também chamado de sequenciamento de nova geração (do inglês, *Next Generation Sequence* - NGS; Fig. 8A) (Young & Gillung 2020). Enquanto o sequenciamento do tipo Sanger “lê” a sequência de um fragmento de DNA específico, o NGS “lê” inúmeros fragmentos diferentes do genoma. Além disso, existem também outras plataformas popularmente disponíveis para sequenciamento de nova geração, no caso da Solexa (Illumina). Na plataforma

Ilumina o processo de análise é feito por síntese que usa DNA polimerase junto dos nucleotídeos, uma técnica de leitura que conecta adaptadores em *chips* de leitura para amplificar a leitura das sequências, fazendo um esquema de “pontes” que no final irá emitir feixes de luzes nas regiões de interesse.

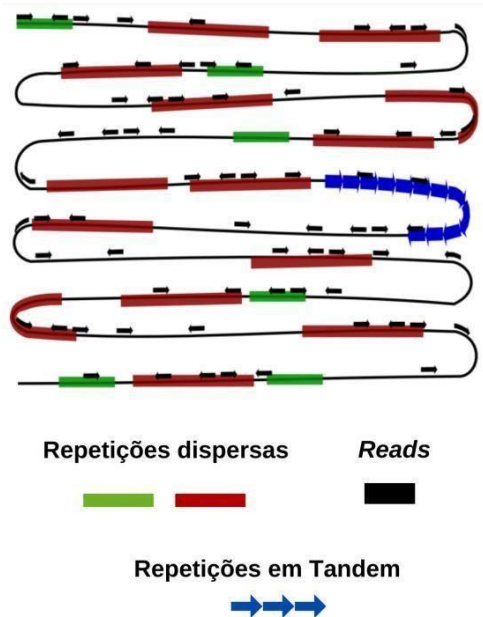
A bioinformática vem para ajudar a “montar o quebra cabeça” destas inúmeras sequências geradas. É graças à bioinformática que os pequenos fragmentos sequenciados (os *reads*) são agrupados entre si por similaridade, formando os *clusters* (Fig. 8B). Os *clusters*, ao serem comparados com bancos de dados de sequências genômicas conservadas entre plantas, são identificados (ao que chamamos de anotação). Por não necessitarem de comparação com genoma de referência, essas abordagens recentes permitem analisar o genoma até mesmo de organismos não-modelos (Egan *et al.* 2012). Para anotar as sequências repetitivas presentes no genoma a partir dos dados de sequenciamento, lança-se mão das análises bioinformáticas *in silico*, o que pode ser feito por meio de diferentes plataformas digitais, como REViewer (You *et al.* 2013), RepEx (Gurusaran *et al.* 2013), RepeatProfile (Negm *et al.* 2021) ou o RepeatExplorer2 (RE2; Novák *et al.* 2013, 2017, 2020).

O RE2 é uma das ferramentas mais amigáveis para este tipo de análise e, possivelmente por isso, tal *pipeline* tem sido amplamente utilizado para identificar sequências repetitivas no genoma de diferentes espécies de plantas e realizar análises comparativas entre grupos (Lee *et al.* 2018, Hsu *et al.* 2020, Eriksson *et al.* 2022, Wang *et al.* 2022). O RE2 é um *pipeline* que agrupa ferramentas acessíveis via servidor Galaxy via web (<https://galaxy-elixir.cerit-sc.cz/>), permitindo que os *clusters* montados durante a análise *in silico* (Fig. 8A e B) sejam anotados por comparação com um banco de dados de Viridiplantae, mas também sejam interpretados segundo a topologias dos grafos gerados pela organização dos *reads* dentro do *cluster* (Fig. 8C).

Desta forma, o RE2 permite anotar sequências repetitivas, em especial DNAsat, que não estão no banco de dados.

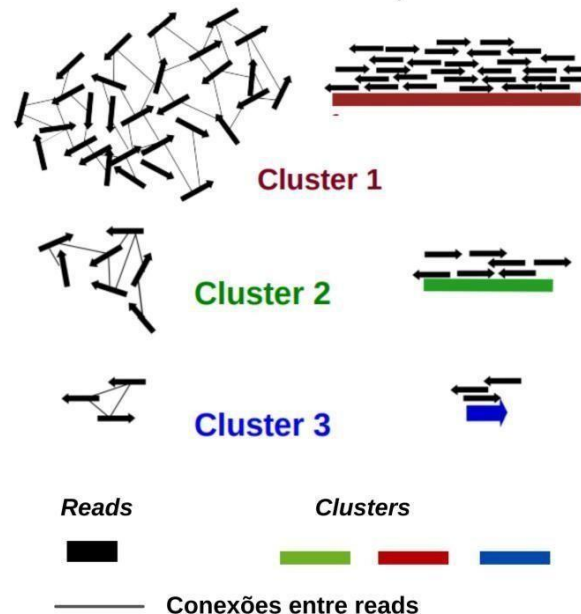
Os grafos são gerados a partir do número de conexões entre os *reads* de um *cluster* (Fig. 8C). Os *reads* que apresentam muitos pontos de similaridade dentro da própria sequência irão gerar *clusters* com grafos de formato circulares e densos (Fig. 8C'), enquanto os *reads* que apresentam um, ou poucos, pontos de similaridade entre si tendem a gerar *clusters* com grafos mais lineares (Fig. 8C''). Enquanto o primeiro exemplo sugere uma sequência de DNAsat, o segundo exemplo refere-se a um TE. O RE2, uma plataforma amigável e intuitiva, pode auxiliar nos estudos que busquem revelar mudanças genômicas nas espécies da nossa biodiversidade, sendo especial valorosa em estudo que não envolvam plantas modelos. Famílias vegetais grandes, com diversidade de hábitos, morfologias, distribuição geográfica ampla e bom conhecimento da citogenética básica podem se beneficiar de forma especial - como é o caso das orquídeas.

A) Sequenciamento do genoma por Next Generation Sequence



Análise do Repeat Explorer

B) Montagem dos clusters



C) Análise de grafos

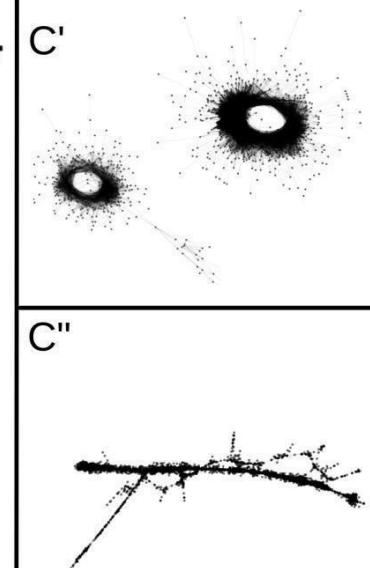


Figura 8. Etapas da análise do DNA repetitivo via RepeatExplorer2. A) O sequenciamento NGS. Em zigzag, tem-se a sequência total do genoma do indivíduo, em verde, vermelho e azul as partes de DNA repetitivo e em preto, os *reads* gerados pelo NGS. B) A montagem dos cluster por agrupamento dos *reads* feita pelo *Repeat Explorer*. As linhas cinzas ao fundo mostram a conexão entre *reads*. C) Representação dos grafos gerados pela organização dos *reads* dentro dos clusters. Imagem baseada em ilustração fornecida no curso Workshop *Repeat Explorer* 2022.

1.6 *Brasiliorchis* como modelo de estudo

Orchidaceae Juss. é a maior família de angiospermas (Christenhusz & Byng 2016, Ulloa *et al.* 2017), apresentando grande diversidade de espécies e ampla distribuição geográfica, em especial nos Neotrópicos (Givnish *et al.* 2015, 2016). Originada aproximadamente há 88-112 milhões de anos (Ma) na Austrália (Givnish *et al.* 2016, Serna-Sánchez *et al.* 2021), essa família conta atualmente com cerca de 28 mil espécies distribuídas em 880 gêneros organizados em cinco subfamílias (Freudenstein & Chase 2015, Chase *et al.* 2015; Fig. 9): Apostasioideae (40-90 Ma), Vanilloideae (80-84 Ma), Cypripedioideae (31-76 Ma), Orchidoideae (57-64 Ma) e Epidendroideae (44-64 Ma) (idades de acordo com Givnish *et al.* 2016 e Serna-Sánchez *et al.* 2021, respectivamente).

Representantes de Orchidaceae podem ser encontrados em praticamente todos os continentes, com exceção da Antártica, e seu habitat inclui desde florestas tropicais úmidas até desertos quentes. As orquídeas são conhecidas principalmente por suas flores exuberantes (veja Fig. 9 e 10), mas também possuem uma variedade de usos, uma vez que Orchidaceae agrupa espécies ornamentais, medicinais e alimentícias.

Dentre os representantes da tribo Epidendroideae (Fig. 9B), alguns gêneros são bastante grandes como *Epidendrum* L., com ca. de 1.500 espécies, e *Maxillaria* Ruiz & Pav. sensu lato (s.l.), com ca. de 580 espécies (Govaerts 2017). Enquanto o primeiro gênero continua com sua ampla circunscrição original, *Maxillaria* (subtribo Maxillarinae, veja estrela em Fig. 9B) foi foco de vários rearranjos taxonômico nas últimas décadas (Whitten *et al.* 2007, Blanco *et al.* 2007, Whitten & Blanco 2011).

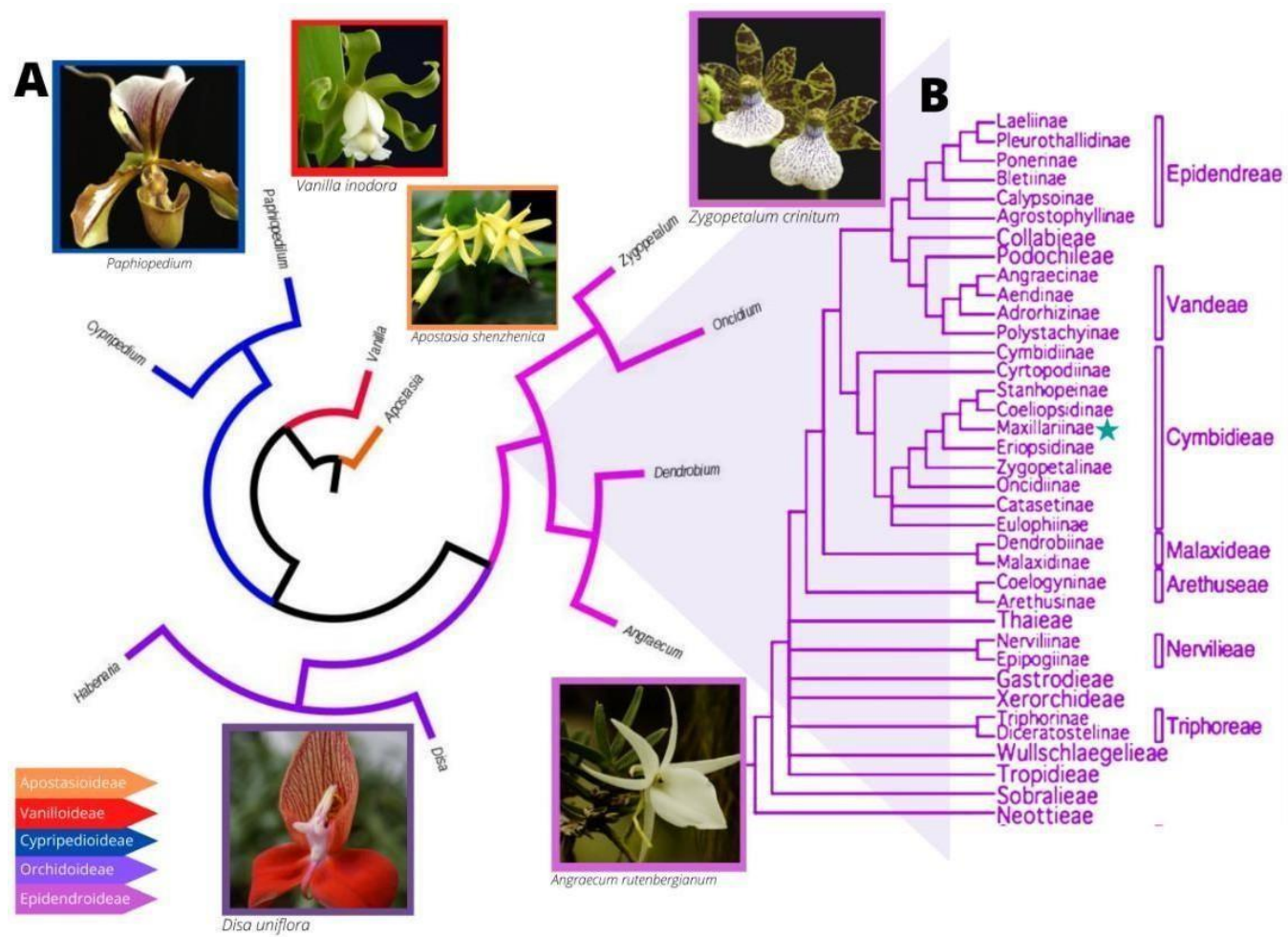


Figura 9. Sistemática de Orchidaceae.(continua na próxima página)

<< **Figura 9. Sistemática de Orchidaceae.** (A) A família Orchidaceae é dividida em cinco subfamílias (Apostasioideae, Vanilloideae, Cypridioideae, Orchidoideae e Epidendroideae). (B) A subfamília Epidendroideae (ramos em rosa em A e cladograma em B) é a maior subfamília, contemplando 16 tribos (setas verdes em B) e 70% das espécies de Orchidaceae. Estrela verde indica a posição da subtribo Maxillariinae, na tribo Cymbidiae. Filogenia de Orchidaceae simplificada de Serna-Sánchez *et al.* 2021 e filogenia de Epidendroideae simplificada a partir de Chase *et al.* (2015). Ramos da filogenia em A estão coloridos de acordo com a subfamília segundo legenda. Fotos de flores retiradas de Serna-Sánchez *et al.* (2021) e OrchiBase.

Baseado nos marcadores moleculares Whitten *et al.* (2007) propuseram o realinhamento do gênero *Maxillaria*, até então polifilético. A divisão de *Maxillaria* em 13 gêneros menores e um core de *Maxillaria* deixou a subtribo Maxillariinae com 17 gêneros monofiléticos (Whitten *et al.* 2007; Fig. 10). Segundo os autores, *Maxillaria* s.s. se justificaria pois é um agrupamento monofilético caracterizado pela sinapomorfia morfológica da linha de abscisão entre a folha e o bulbo. Entretanto, o novo realinhamento não foi aceito por toda comunidade científica.

Divergindo da proposta de Whitten *et al.* (2007), Chase *et al.* (2015) e Schuiteman & Chase (2015) sugeriram que seria melhor agrupar todas as espécies em um mega-gênero *Maxillaria*, ainda maior do que previamente aceito. Tal proposta se baseia na justificativa de que, para os especialistas, esses gêneros recentemente reconhecidos/descritos/ressuscitados fazem sentido, mas para a grande maioria dos usuários esse sistema taxonômico é extremamente desafiador (Chase *et al.* 2015). O site Re flora, ao classificar as espécies da subtribo Maxillariinae, segue a sugestão de Chase *et al.* (2015) e Schuiteman & Chase (2015). Porém, a reorganização realizada

pelos autores do Re flora, e por Schuiteman & Chase (2015), não foi cuidadosa o suficiente, fazendo com que duas espécies claramente diferentes agora tenham o mesmo nome.

As mudanças taxonômicas necessitam ser feitas com responsabilidade, pois tais mudanças podem levar a mudanças nos status de conservação das espécies afetando os planos de conservação (Garnet & Christidis 2017). Por exemplo, a mudança taxonômica proposta (Chase *et al.* 2005, Schuiteman & Chase 2015) fez com que centenas de espécies passassem a ser consideradas sinônimos, por mais diferente que sejam morfologicamente (veja Fig. 10). Os nomes agora unem espécies muito distantes em uma única espécie como, por exemplo *Trigonidium ringes* Lindl., que segundo a Flora e Funga do Brasil (2022), agora é considerado sinônimo de *Mormolyca ringes* (Lindl.) Gentil (Fig. 10U), ambos sob o nome de *Maxillaria lineolata* (Frenzl) Molinari, três espécies claramente diferentes. Para se ter uma ideia, compare as fotos de flores do gênero *Trigonidium* (Fig. 10A' e B') com *Mormolyca ringer* (Fig. 10U) para observar a diferença clara entre os táxons. Devido a essas inconsistências, neste trabalho seguiremos com a classificação proposta por Whiten *et al.* (2007) e Blanco *et al.* (2007).

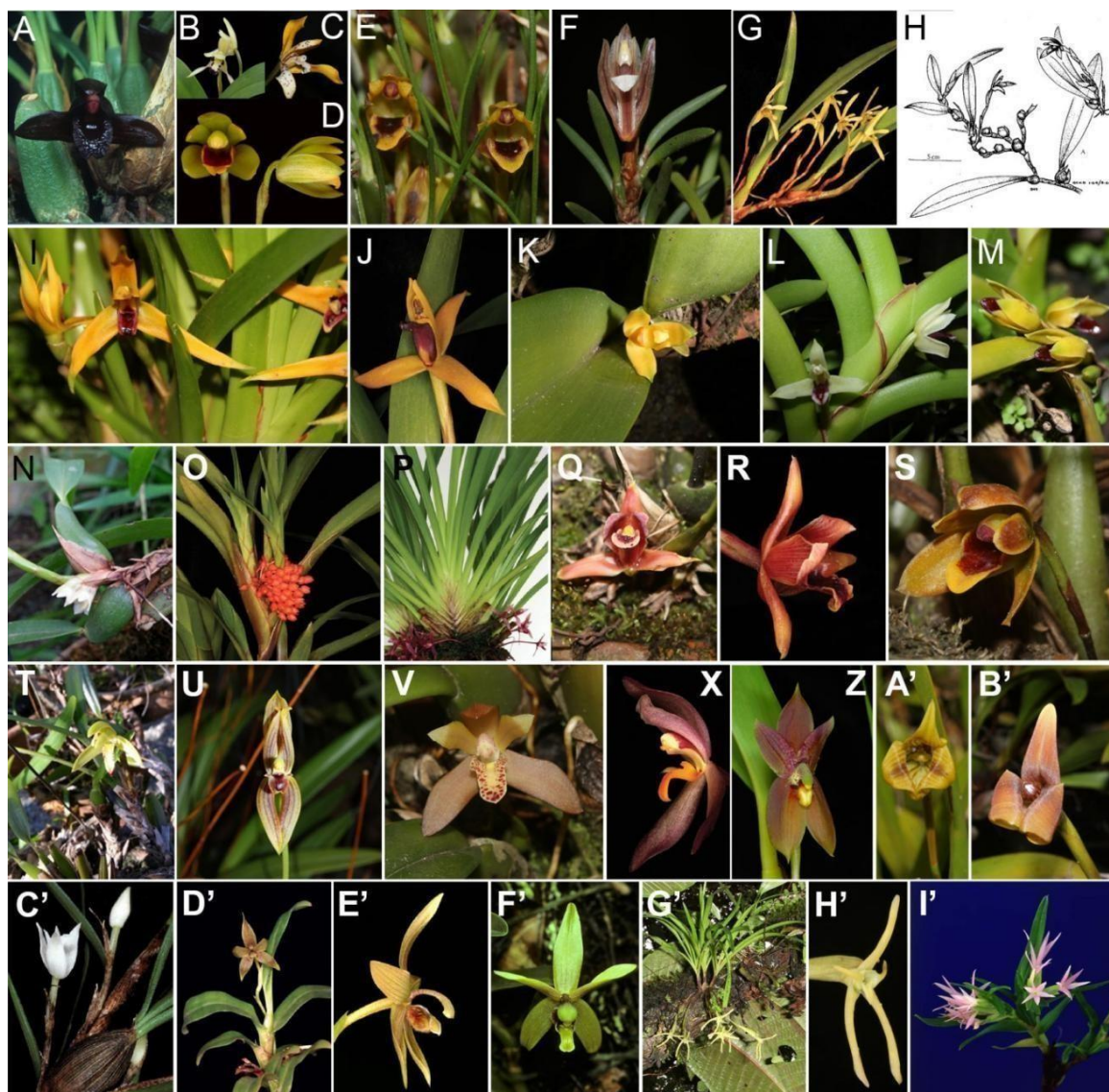


Figura 10. Representantes da subtribo Maxillariinae. A-D. *Brasiliorchis*: A: *B. schunkeana*, B: *B. marginata*, C: *B. picta*, D: *B. barbosea*. E-F: *Christensonella*. E: *C. neowiide*, F: *C. uncata*. G-H: *Sauvetrea alpestris*. I: *Nitidobulbon*. I: *N. nasutum*, J: *N. proboscideum*. K-M: *Heterotaxis*. K: *H. brasiliensis*, L: *H. equitans*, M: *H. superflua*. N: *Ornithidium vestitum*. O: *Ornithidium fulgens*. P: *Inti bicallousa*. Q-R: *Mappinguaria desvauxiniana*. S: *Maxillariella variabilis*. T: *Rhetinantha scorpioidea*. U-X: *Mormolyca*. U: *Mormolyca ringes*, V: *Mormolyca rufecens*, X: *Mormolyca richii*. A'-B': *Trigonidium*. A': *T. acuminatum*, B': *T. obtusum*. C': *Pityphyllum laricinum*. D'-F': *Cyrtidiorchis*. D', E': *C. alata*, F': *C. stumpfleii*. G'-I': *Cryptocentrum*. G' e H': *Cryp. latifolium*, I': *Cryp. roseans* (previamente considerado *Anthosiphon roseans* Schltr). Créditos: A-F, K-N, Q, S-V, A'-B': APMoraes; G, J, O-P, R, X-Z, C'-F' e H' - fotos publicadas em Whitten & Blanco (2011); H - desenho retirado de Pridgeon *et al.* (2009); I, H', I' - fotos retiradas de www.orchidspecies.com.

Dentre os gêneros novos propostos por Whiten *et al.* (2007) e Blanco *et al.* (2007) está *Brasiliorchis* R. Singer, S. Koehler e Carnevali (Singer *et al.* 2007; Fig. 10A-D), gênero monofilético que agrupa as espécies antes consideradas da aliança picta. As espécies deste gênero são endêmicas da Mata Atlântica do Brasil, ocorrendo em altitudes entre 500 e 800 m (Singer *et al.* 2007). As espécies se distribuem de forma homogênea, exceção feita à *Brasiliorchis schunkeana* (Campacci & Kautsky) R.B.Singer, S.Koehler & Carnevali, endêmica de regiões de Mata Atlântica do estado do Espírito Santo, região esta sugerida como refúgio no Quaternário e à *Brasiliorchis barbozae* (Loefgr.) R.B.Singer, S.Koehler & Carnevali, endêmica da Serra do Espinhaço. Morfologicamente, *Brasiliorchis* pode ser caracterizada por suas flores desprovidas de secreções ou tricomas pluricelulares, não oferecendo nenhum recurso aos seus polinizadores (polinização por engodo). O pedicelo floral é normalmente bem desenvolvido (veja seta em Fig. 11A); a coluna é quase reta e não tem apêndices laterais (asas da coluna) (Fig. 11A'), o pé da coluna não é tão bem desenvolvido como em outros membros da subtribo *Maxillariinae*, o polinário é recoberto por um capuz (Fig. 11A''), incumbente e composto por dois pares de polínias dorsalmente comprimidas de coloração amarela gema (Fig. 11A'''). As flores destas espécies apresentam com formato campanulado (Fig. 11 e 12), têm longa duração (10 dias ou mais), são perfumadas e de coloração geralmente creme ou amarelo com pintas roxas (Fig. 11B, C e E), exceção feita novamente às espécies *B. schunkeana*, que apresenta flores totalmente negras (Fig. 11F e G), e *B. barbozae*, que exibe labelo vermelho escuro e restante das flores amarelo-creme (Fig. 11D e Fig. 12B).

Dentre as características vegetativas das espécies de *Brasiliorchis* podemos destacar: **(1)** pseudobulbos sulcados e bifoliados (Fig. 11A), normalmente fortemente agregados (Fig. 11G), com exceção de *Brasiliorchis marginata* (Lindl.) R.B.Singer,

S.Koehler & Carnevali e *Brasiliorchis chrysantha* (Barb.Rodr.) R.B.Singer, S.Koehler & Carnevali, nas quais os pseudobulbos são separados por um rizoma conspícuo recoberto por brácteas secas; (2) folhas conduplicadas e lanceoladas (Fig. 11A e G) (exceção feita à *Brasiliorchis kautskyi* (Pabst) R.B.Singer, S.Koehler & Carnevali que exibe folhas aciculares), as folhas são comumente mais longas em indivíduos que crescem em lugares sombrios e mais duros, curtos e de coloração verde-amarelado em plantas que crescem em condições ensolaradas; (3) raízes lisas com a ponta avermelhada ou acastanhada, quando expostas.

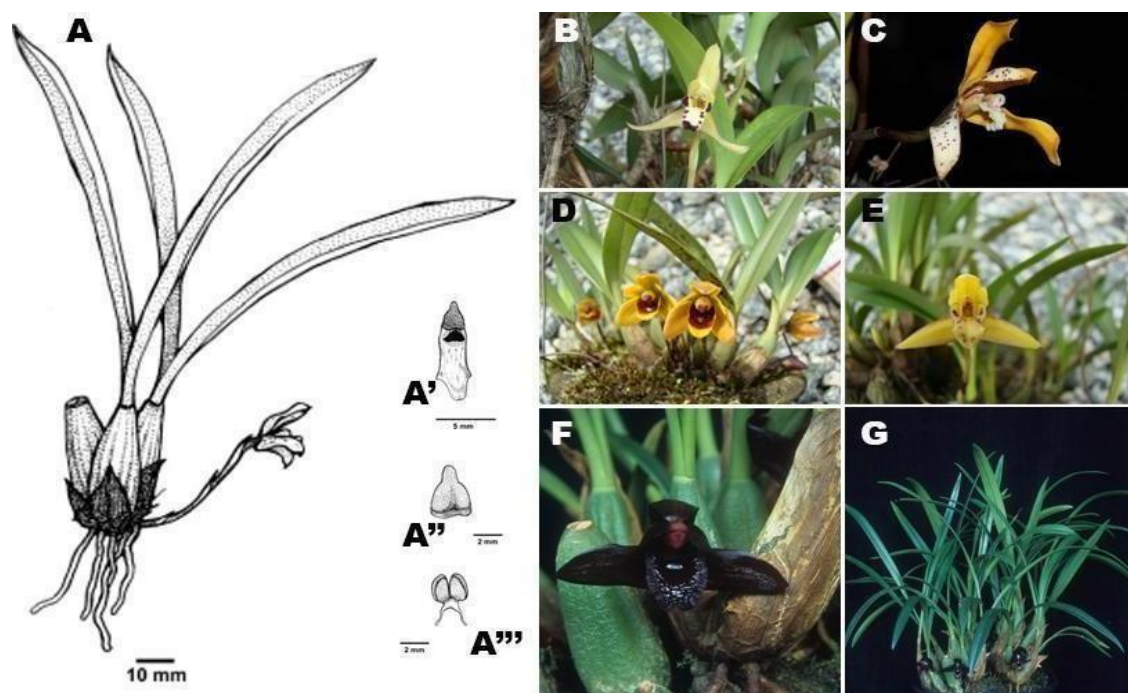


Figura 11. Morfologia de *Brasiliorchis* R. Singer, S. Koehler e Carnevali. (A) Indivíduo inteiro. Observe os pseudobulbos cônicos bifoliados, com folhas lanceoladas. (A') coluna. (A'') Capuz do polinário. (A''') Polinário. (B) *B. marginata*. (C) *B. picta*. (D) *B. barbosea*. (E) *B. porphyrosteles*. (F e G) *B. schunkeana*, flor e hábito respectivamente. Desenho apresentado em 1A foram modificados a partir de Singer *et al.* (2007). Fotos de B a G foram feitas por APMoraes.

Assim como a morfologia do gênero, a análise cromossômica já explorou boa parte das espécies do gênero sendo conhecidos dados de número cromossômico para nove das 13 espécies (Moraes *et al.* 2012, 2022), sendo que todas apresentam $2n=40$. Mesmo com a constância no número cromossômico, duas espécies apresentam aumento do GS: *B. barbozae* ($1C=5,54pg$; Moraes *et al.* 2022) e *B. schunkeana* ($1C=4,19$; Moraes *et al.* 2012), mas o cariótipo de *B. schunkeana* traz uma curiosidade, não observada em *B. barbozae*: grandes blocos heterocromáticos ricos em AT (brilhantes com o corante DAPI⁺) que cobrem os dois braços de dois pares cromossômicos, além dos braços longos de um terceiro par cromossômico (os sítios de DNA ribossomal 45S localizam-se nos braços curtos desse terceiro par) (Moraes *et al.* 2012). Dada todas as peculiaridades observadas em *B. schunkeana* e *B. schunkeana*, não é de se estranhar que essas duas espécies sejam espécies irmãs das demais espécies do gênero (Whitten *et al.* 2007, Moraes *et al.* 2022; Fig. 12).

Considerando todas as características destoantes de *B. schunkeana*, esta espécie destaca-se como um excelente modelo para compreender o papel de sequências repetitivas na mudança do tamanho de genoma graças a amplificação de DNAsat, potencialmente não devido a expansão de TE, como comumente visto em plantas. Assim, neste projeto visa-se identificar se há sequências de DNAsat com distribuição diferencial em *B. schunkeana* que possa explicar o aumento de tamanho de genoma nesta espécie. Entretanto, tal estudo depende de análises comparativas com outra espécie do gênero. Considerando-se a constância dos caracteres citogenéticos das espécies de *Brasiliorchis*, qualquer uma das espécies do gênero poderia servir de comparação. Neste sentido, escolhemos a espécie-tipo do gênero, *Brasiliorchis picta* (Hook.) R.B.Singer, S.Koehler & Carnevali (Singer *et al.* 2007, Pinheiro *et al.* 2009).

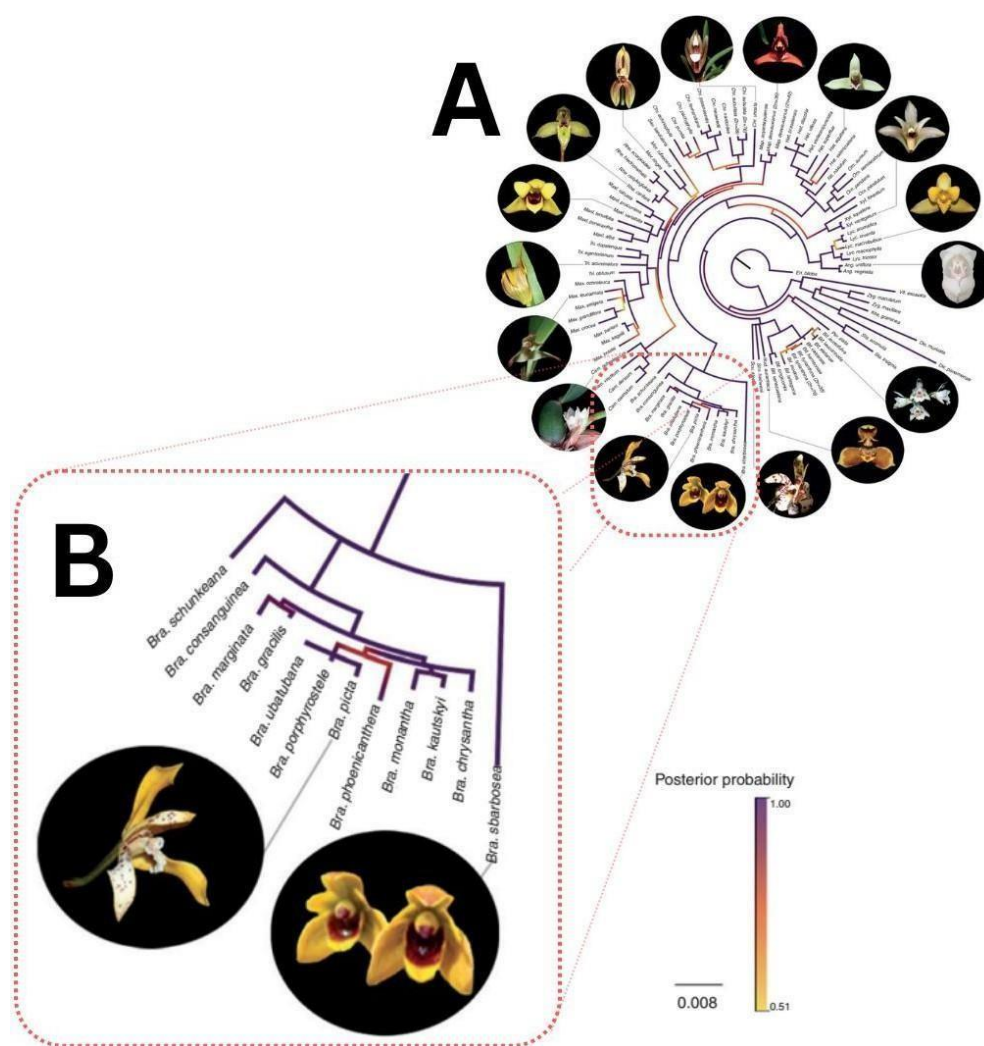


Figura 12. Filogenia da subtribo Maxillariinae (A) e do gênero *Brasiliorchis* (B).

Suporte de Probabilidade Posterior é indicado pelas cores dos ramos de acordo com a escala. Em A, os inserts mostram as imagens de representantes dos gêneros de Maxillariinae: *Scuticaria steleei*, *Rudolfiella aurantiaca*, *Bifrenaria leucorhoda*, *Anguloa uniflora*, *Lycaste macrobulbon*, *Xylobium variegatum*, *Heterotaxis equitans*, *Mapinguari desvauxianus*, *Christensonella uncata*, *Mormolyca ringes*, *Rhetinantha cerifera*, *Maxillariella variabilis*, *Trigonidium acuminatum*, *Maxillaria leucaimata*, *Camaridium ochroleucum*, *Brasiliorchis picta* e *Brasiliorchis barbosea*. Em B) inserto da filogenia de *Brasiliorchis* com imagens de *Brasiliorchis picta* e *Brasiliorchis barbosea*. Fotos de A. P. Moraes and J. S. Cabral. Imagem editada de Moraes *et al.* 2022.

2. Objetivos

Neste projeto objetivamos determinar qual(is) sequência(s) repetitiva(s) é responsável pelo aumento do tamanho de genoma em *B. shunkeana*, determinando se a mudança do tamanho de genoma é graças a amplificação de sequências de TE, como comumente visto em plantas, ou de DNAsat associado aos blocos heterocromáticos DAPI⁺ previamente identificados em *B. shunkeana*. Para tal será realizada análise *in silico* comparativa entre *B. shunkeana* e *B. picta* buscando identificar *in silico* a sequência repetitiva mais representativa de *B. schunkeana*.

3. Material e Métodos

3.1 Material

Duas espécies de *Brasiliorchis* foram sequenciadas previamente a realização deste projeto. O material testemunho dos espécimes utilizados no sequenciamento foram depositadas no Herbário Sinningia da Universidade Federal do ABC (*B. picta* - HUFABC 001181, *B. schunkeana* - HUFABC 000019).

3.2 Método: Análises no Repeat Explorer 2

Os *reads* brutos do sequenciamento *Illumina* obtidos previamente a realização deste estudo foram analisados, de forma individual (*ie.*, por espécie) e de forma comparativa entre espécies, através do *pipeline* RE2 (Novak *et al.* 2013, 2017, 2020). Os *reads* passaram por uma triagem de qualidade e agrupados, com mais de 10% de sobreposição e sequência superior a 95% de similaridade. Os *clusters* resultantes foram alinhados de acordo com sua semelhança no *pipeline* usando o banco de dados de domínio conservado de proteína chamado *Viridiplantae* para anotar as sequências classificadas como repetidas.

Neste *pipeline*, os *clusters* além de serem anotados por comparação com o banco de dados *Viridiplantae* implementado no RE2, também são classificados por meio da topologia dos grafos. Nesta etapa, o suporte para essa identificação vem do TAREAN que é uma ferramenta implementada no próprio RE2. Além disso, a confirmação da anotação de cluster pobremente anotados pelo RE2 foi realizado pelo *Basic Local Alignment Search Tool* (BLAST) (Altschul *et al.* 1990) e RepBase. Para *clusters* potencialmente de DNAsat, o software Geneious Prime 2022.1.1 foi usado para gerar o gráfico de *dot-plot*, ajudaram a refinar e finalizar tal anotação. A

comparação de cluster anotados nas duas diferentes espécies foi feita via alinhamento de sequências no MAFFT (Kato *et al.* 2013, 2018).

A análise comparativa também foi realizada via RE2, tomando por base a anotação de sequências da rodada anterior. Os *cluster* foram anotados como citados anteriormente. No final da rodada foi possível identificar os clusters que estão ausente, ou em baixa frequência, em *B. picta* e presentes em alta frequência em *B. schunkeana*.

4. Resultados

4.1 Análise individual

As duas espécies foram analisadas individualmente, visando anotar as sequências do repetoma e avaliar a composição da fração repetitiva de cada espécie. As percentagens de representação das diferentes frações anotadas do genoma foram bastante consistentes entre *B. schunkeana* e *B. picta* (Fig. 12, Tabela 1). Sequências de cloroplasto e mitocôndria representam 1-2% do genoma anotado, enquanto sequências de DNA ribossomal, 0,34-0,52% (*B. schunkeana* e *B. picta*, respectivamente). Aproximadamente 21-22% das sequências (*B. schunkeana* e *B. picta*, respectivamente) não puderam ser agrupadas em *clusters* (*singlets* representado em verde no gráfico de pizza na Figura 12) e 8%, em ambas as espécies, formaram *clusters* de baixa repetição. Além dessas, uma pequena fração de cluster grandes (os Top-clusters) foram montados mas não puderam ser anotados, nem mesmo empregando os métodos adicionais de anotação realizados após o RE2. Ao total, c. 33-36% dos cluster não foram anotados em *B. schunkeana* e *B. picta*, respectivamente (Tabela 1). Removendo estes grupos de contaminantes e de sequências com baixa qualidade de anotação, observa-se que o repetoma é composto de 90-95% de TE (representando 60-59% do total do genoma) em *B. schunkeana* e *B. picta*, respectivamente (Tabela 1, veja inserto em Fig. 13).

As composições dentro dos elementos móveis também são similares entre as duas espécies (Fig. 13, Tabela 1). Entre os elementos transponíveis identificados, os TE de classe I são os mais representativos nas duas espécies, representando c. 96% dos TE em ambas as espécies, sendo que TE de Classe II compreendem apenas 3,94-3,51% dos TE anotados *B. schunkeana* e *B. picta*, respectivamente (veja fatia em tons de azul na Fig. 13, Tabela 1). Dentro de TE de classe I, nota-se a prevalência da ocorrência

de TE com LTR (96,06 e 94,07% em *B. schunkeana* e *B. picta*, respectivamente; veja porção em vermelho escuro indicada por LTR na Fig. 13), mas uma distribuição equivalente entre as categorias Ty1/Copia e Ty3/Gypsy (Fig. 13). Dentro de Ty1/Copia, o elemento transponível mais representativo, nas duas espécies, foi SIRE, enquanto que entre os Ty3/Gypsy foi o elemento OTA/Tat/Ogre, representando 27,88 e 29,32% do genoma anotado de *B. schunkeana* e *B. picta*, respectivamente (52,30 e 49,64% dos TEs nestas mesmas duas espécies; Tabela 1).

Entretanto, apesar da constância dos resultados de TE apresentados, nota-se uma discrepância quanto à representatividade de DNAsat nas duas espécies: enquanto em *B. schunkeana* 6% do genoma foi anotado como DNAsat, apenas 2,73% do genoma de *B. picta* 3% foi anotado nesta mesma categoria (veja estrela em Figura 12, Tabela 1). Desta forma, nossos resultados indicam que comparando o repetoma das duas espécies, os resultados foram similares, exceto pela presença muito mais marcante de DNAsat em *B. schunkeana*, sendo que provavelmente é essa fração do genoma a responsável pelo aumento do tamanho de genoma em *B. schunkeana*.

Tabela 1. Percentagens das diferentes categorias anotadas nos genomas de *Brasiliorchis schunkeanae* *Brasiliorchis picta*. Para cada espécie são apresentadas as percentagens de cada categoria balizadas pelo total do genoma anotado (coluna Genoma), total do repetoma, excluindo-se os rDNA (coluna Repetoma) e total de elementos transponíveis anotados (coluna TE).

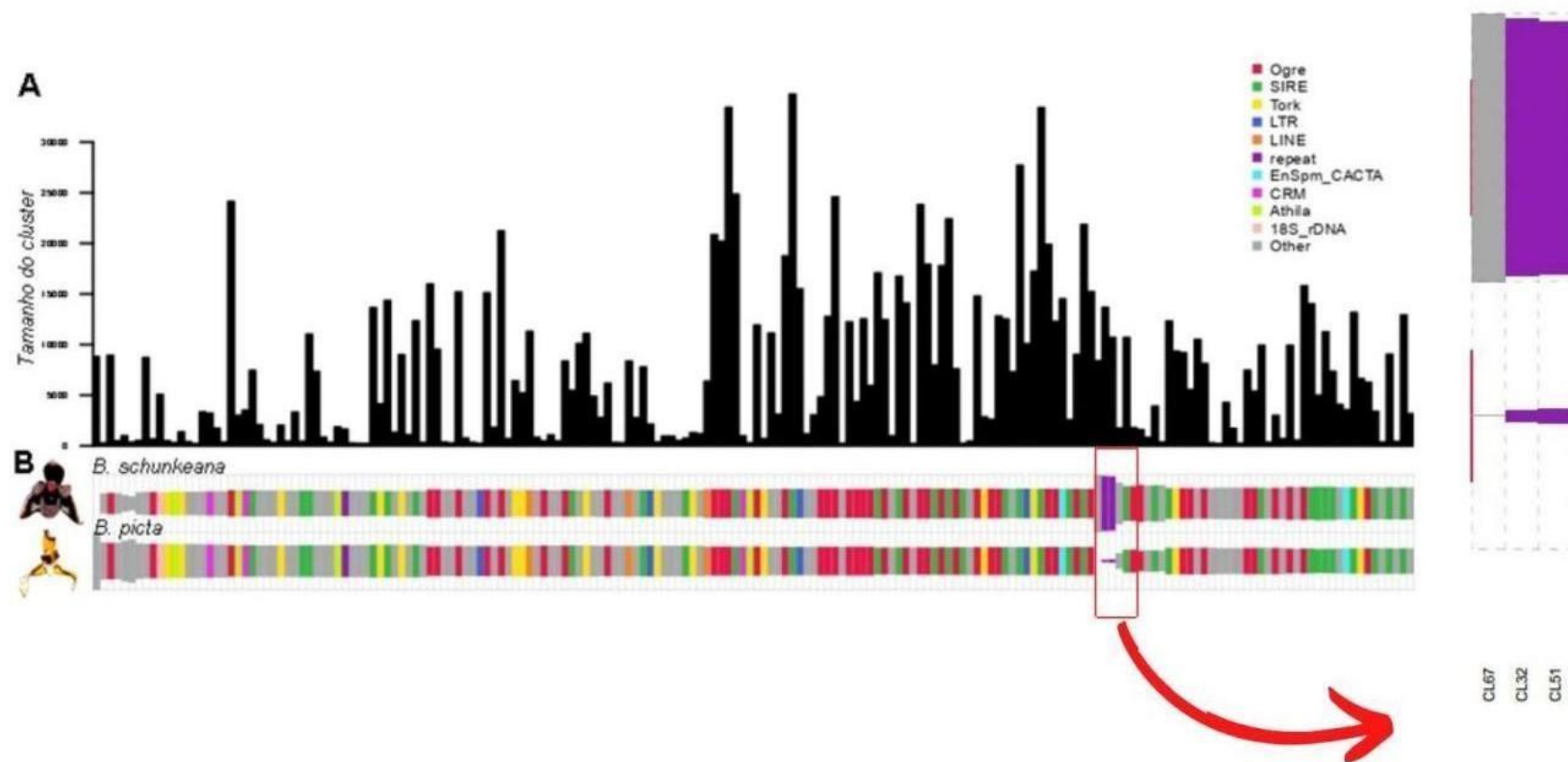
Categorias anotadas	Genoma	Repetoma	TE	Genoma	Repetoma	TE
	<i>B. schunkeana</i>			<i>B. picta</i>		
Singlets, cluster pequenos e Top Cluster não anotados	32,96%			35,96%		
Organelas	1% do genoma			1,73% do genoma		
plastid	0,991278934			1,677499304		
mitochondria	0			0,04850053		
DNA ribossomal	0,34% do genoma			0,52% do genoma		
45S_rDNA	0,34			0,50		
5S_rDNA	0,00			0,03		
Repetoma (removendo rDNA)	61,79% do genoma			61,79% do genoma		
DNA satélite	6,04	9,19		2,73	4,42	
TEs	59,67% do genoma			59% do genoma		
Classe I	96,06% dos TEs			96,49% dos TEs		
LINE	0	0	2,55	1,43	2,31	2,42

Classe I - LTR				96,06% dos TEs			94,07% dos TEs		
LTR				2,91	4,42	0,20	0,11	0,18	0,19
Classe I - LTR - Ty1/Copia				42,74% dos TEs			41,92% dos TEs		
SIRE				21,21	32,28	36,08	20,22	32,73	34,25
Tork				4,11	6,26	7,69	4,31	6,98	7,30
Ale				0,03	0,05	0,11	0,06	0,10	0,10
Ikeros				0,15	0,22	0,29	0,16	0,26	0,27
Classe I - LTR - Ty3/Gypsy				48,45% dos TEs			51,96% dos TEs		
chromovirus	CRM			0,45	0,68	1,15	0,64	1,04	1,09
chromovirus	Tekay			0,18	0,27	0,39	0,22	0,36	0,37
non-chromovirus	OTA	Athila		0,40	0,61	0,91	0,51	0,82	0,86
non-chromovirus	OTA	Tat	Ogre	27,88	42,43	52,30	29,32	47,45	49,64
Classe II				3,94% dos TEs			3,51% dos TEs		
Subclass_1	TIR	EnSpm_CACTA		1,10	1,68	1,68	0,94	1,52	1,59
Subclass_1	TIR	hAT		0,24	0,36	0,53	0,30	0,48	0,50
Subclass_1	TIR	MuDR_Mutator		0,34	0,51	0,61	0,34	0,55	0,58
Subclass_1	TIR	PIF_Harbinger		0,67	1,02	0,88	0,49	0,80	0,84

4.2 Análise Comparativa

Depois dos dados individuais, realizamos uma análise comparativa entre *B. schunkeana* e *B. picta* visando verificar quais *clusters* apresentam frequência diferencial entre as duas espécies. Essa análise permite indicar qual(is) *cluster(s)* dentro da fração de DNAsat já identificada como contraste são diferente entre as duas espécies. A análise comparativa realizada no RE2 indicou três *clusters* mais representativos *B. schunkeana* e menos representativos em *B. picta*. Importante observar que dois dos três *clusters* foram identificados como DNAsat já na anotação automática do RE2 (veja coloração roxa da barra em Fig. 14B), enquanto último não foi anotado (veja coloração cinza da barra em Fig. 14B). Esses *clusters* foram identificados como CL32, CL51 e CL67 sendo este último o cluster não anotado pelo RE2 e sendo ausente em *B. picta*. Enquanto as sequências consenso dos CL32 e 51 evidenciam sequências ricas em AT, a sequência consenso do CL67 apresenta mais CG do que AT (Mat. Sup. 1).

>>Figura 15. Análise comparativa entre as espécies *Brasiliorchis schunkeana* e *Brasiliorchis picta*. A) Tamanho em pares de bases dos clusters anotados em *B. schunkeana* e *B. picta*. B) Frequência relativa dos clusters dentro de cada espécie, plotados de forma comparativa entre as espécies *B. schunkeana* (linha superior) e *B. picta* (linha inferior), a seta em vermelho indica o destaque nos três clusters que foram identificados. A anotação dos cluster está indicada na legenda de cores na porção superior direita da imagem. Retângulo vermelho sinaliza os cluster com maior diferença entre espécies, em roxo são os clusters anotado como DNA satélite.



Visando confirmar e refinar as anotações destes três *clusters* foi realizada busca no banco de dados do GenBank por sequências similares. A busca, empregando o BLASTn não retornou nenhum resultado. O alinhamento entre as sequências consenso destes três *clusters* também não identificou similaridade entre os *clusters*, confirmando que deva se tratar de três DNAsat diferentes. Por fim, realizamos a análise de *dot-plot* dos três *clusters* buscando identificar se o CL67 é um DNAsat, e ratificar a anotação de DNAsat para os outros dois clusters (Fig. 15). A análise do *dot-plot* dos três *clusters* com distribuição diferencial entre *B. schunkeana* e *B. picta* reforça a anotação do CL32 e 51 como DNAsat (Fig. 15A-B), além de indicar que CL67 seja um DNAsat, uma vez que o gráfico de *dot-plot* é bastante denso (Fig. 15C). Adicionalmente, o padrão repetitivo destes três *clusters*, em especial CL32 e 51 (Fig. 15A-B) é evidenciado pela topologia dos grafos que apresentam muitas conexões entre os contigs que formam os *clusters* (Fig. 15).

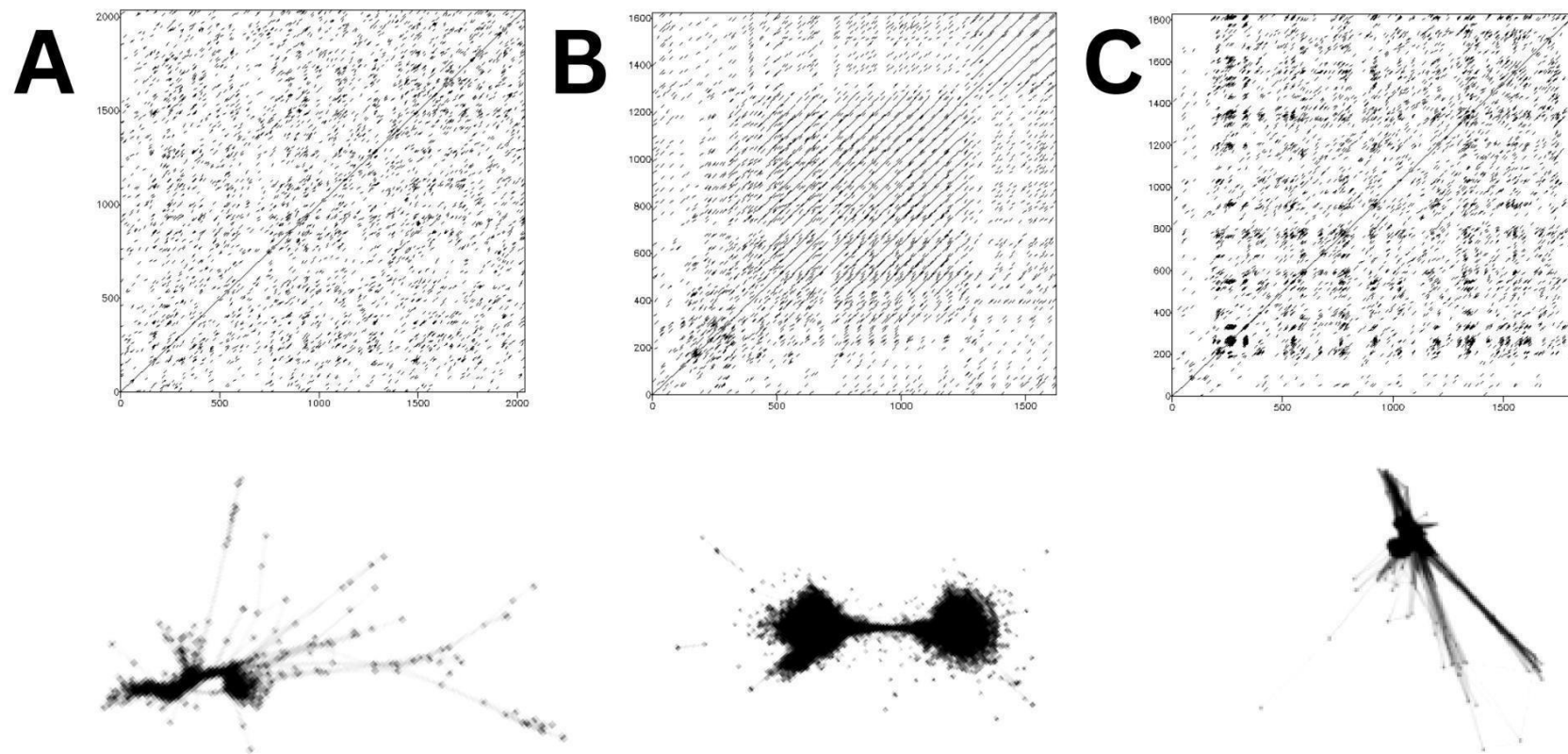


Figura 16. Análise *dot-plot* e grafo dos três *clusters* com distribuição diferencial entre as espécies. A) CL32, B) CL51, C) CL67. Na linha superior estão apresentados os gráficos gerados na análise de *dot-plot*, enquanto na linha inferior, para cada *cluster*, estão indicados os grafos gerados pela ferramenta Seqgraph no RE2.

5. Discussão

5.1 A composição de elementos transponíveis em plantas e no modelo de estudo

Os trabalhos dedicados ao estudo da composição dos genomas vegetais têm indicado, desde muito tempo, o papel importante da fração repetitiva (Varani *et al.* 2015). As sequências altamente repetitivas compreendem a maior fração do genoma (Wicker *et al.* 2007), desempenhando um papel fundamental no aumento do tamanho de genoma (Kazazian 2004, Biscotti *et al.* 2015). Comumente, com relação a distribuição dos TEs ao longo de todo o genoma, a maior parte destas sequências em vegetais costumam ser da ordem retrotransposons com LTR, mais especificamente da superfamília Gypsy, seguida pela superfamília Copia, transposons de DNA e finalmente de retrotransposons sem LTR (Bennetzen *et al.* 2000, Biémont 2010, Varani *et al.* 2015, Bourque *et al.* 2018). Devido à sua abundância no genoma e ao potencial de causar mutações, os elementos Ty3/Gypsy tendem a ser frequentemente suprimidos pela maquinaria molecular celular (Qiu & Ungerer 2018).

Apesar dos mecanismos de remoção de TE, os TE seguem respondendo pela maior parte dos genomas vegetais. E dentre os TEs, a literatura mostra de forma consistente que as duas maiores superfamílias de retrotransposons LTR, Ty1/Copia e Ty3/Gypsy, costumam responder pela maior parte da fração repetitiva do genoma em plantas (veja por exemplo Sader *et al.* 2021, Ribeiro *et al.* 2022, Ibiapinho *et al.* 2022, Samoluk *et al.* 2022, Wang *et al.* 2022). Esse resultado foi visto também em *Arachis*, com Ty3/Gypsy, com suas cinco linhagens, sendo o mais representativos no genoma, apesar de Ty1-Copia ser mais diversa, com sete linhagens identificadas (Samoluk *et al.* 2015). Resultados similar, de maior diversidade de Ty1/Copia, mas maior representatividade de Ty3/Gypsy são comuns em plantas como em legumes da tribo

Fabae com genomas grandes, como *Vicia*, *Lens*, *Pisum* e *Lathyrus* (Macas *et al.* 2015).

Nossos resultados concordam com o já disposto na literatura, sendo que nas duas espécies os TEs de LTR/Ty3/Gypsy formam os mais frequentes, respondendo por quase metade dos TEs. Dentre os TE da superfamília Ty3/Gypsy, o não-cromovírus OTA/Tat/Ogre foi o mais representativo em *B. schunkeana* e *B. picta*, o que difere de outras espécies já analisadas. Por exemplo, esta linhagem foi pouco representativa em *Cuscuta nitida* E. Mey. Ex Choisy (Ibiapino *et al.* 2022), respondendo por apenas 0,14% do genoma anotado. A fração repetitiva do genoma de *C. nitida* corresponde a 42,83% do total sequenciado, sendo que a maior parte desta fração repetitiva anotada é composta por elementos transponíveis (22,01% do genoma ou metade da fração anotada). A classe mais representativa de TE em *C. nitida* também foi o retratransposon do tipo LTR (aprox. 38% do genoma anotado), sendo o SIRE o mais comum entre as linhagens Ty1/Copia e o Tekay, entre Ty3/Gypsy (Ibiapino *et al.* 2022). Sugere-se que os cariótipos bimodais, típicos do subgênero *Pachystigma* (Engelm.) Baker & C.H. Wright, poderiam ser causado por rearranjos cromossômicos, como fusões e fissões cromossômicas, mas as análises atuais têm sugerido que tal cariótipo tem origem no acúmulo de sequências repetitivas dispersas, os TEs (Ibiapino *et al.* 2022).

Em *Arachis* o principal elemento identificado foi Athila, outra linhagem dentro dos LTR não-cromovírus/OTA, assim como Ogre (Samoluk *et al.* 2022). Em *Agave* L., Ogre (3,36%) e Athila (2,37%) também compoem a fração principal anotada de TE (Ramos *et al.* 2023), mesmo com valores muito inferiores ao observado em *Brasiliorchis* (c. 28% do genoma). Uma diferença clara entre os dados de Ramos *et al.* (2023) é que a maioria dos TEs estão concentrados na categoria

“Ty3/Gypsy e Ty1/Copia não anotados”, enquanto que no nosso estudo, a maioria dos TE puderam ser anotados. Contudo, a importância do TE Athila já havia sido realçada, sendo que sugere-se que esta linhagem compreende uma porção significativa dos TEs em plantas (Wang *et al.* 2021). Em *Arachis*, os autores puderam também estimar a data dos *burst* de inserção que devem ter ocorrido há 2-5 milhões de anos (Samoluk *et al.* 2022), sendo que tais *burst* de atividade de TE-LTR é o principal mecanismo de aumento do genoma vegetal (Bennetzen & Wang 2014, Zhang & Gao 2017).

Concordando com o já descrito aqui, o gênero pantropical da família Leguminoae, *Tephrosia* Pers. (Souza 2022) também apresentou maior representatividade de TEs do tipo retrotransposons LTR, em especial as superfamílias Ty1/Copia e Ty3/Gypsy, porém aqui a linhagem SIRE, da superfamília Ty1/Copia foi a mais representativa. Contrastando com nossos resultados, a linhagem Ogre não foi anotada em 11 das 14 espécies analisadas e nas três espécies onde se detectou a linhagem Ogre, esta estava em baixa frequência (entre 0,07-0,94% dos TEs). Souza (2022) sugere que a expansão do TE nas 14 espécies de *Tephrosia* pode ter sido resultado de influências ambientais, incluindo tanto fatores bióticos como abióticos, sendo que a possível causa para essa relação entre o estresse ambiental e o aumento das repetições de TE pode ser devido ao fato desses elementos estarem frequentemente ligados a áreas de heterocromatina, o que os torna mais propensos a sofrer alterações. Contudo, apesar da importância dos TE, especialmente TE-LTR, na composição do repetoma das duas espécies analisadas aqui, reforçando a importância dos TEs com LTR da superfamília Ty3/Gypsy nos genomas vegetais, a expansão dos TE parece não explicar o aumento do tamanho de genoma em *B. schunkeana*.

5.2 A composição de DNA satélite nos genomas vegetais

As sequências de DNAsat formam blocos conspícuos nos cromossomos, estando localizadas geralmente na constrição secundária (formada pelos genes de DNA ribossomal), nos telômeros e nos centrômeros. Mas também podem formar blocos conspícuos nos braços do cromossomos, como é o caso de *Eleutherine* (Báez *et al.* 2019), *Cuscuta* (Ibiapinho *et al.* 2022), *Agave* L. (Ramos *et al.* 2023) e na espécie analisada aqui, *B. schunkeana* (Moraes *et al.* 2012). Graças a taxa evolução rápida, típica dessa fração do genoma, o DNAsat proporciona diferenciações genômicas entre taxa (Biscotti *et al.* 2015), podendo até mesmo gerar diferenças entre cromossomos, como observado em *Eleutherine*, que apresenta sequências específicas do par cromossômico maior (Báez *et al.* 2019).

Apesar de tradicionalmente ser aceito que a expansão de DNAsat gera um aumento do GS, essa relação não é tão clara como quando se trata da relação expansão de TE/aumento do GS. Inclusive, Samoluk *et al.* (2022) sugeriram que as espécies com menor GS tenderiam a acumular DNAsat em detrimento de TE-LTR, pois o acúmulo de DNAsat não geraria aumentos tão expressivos no tamanho de genoma. Isso pode ser observado em *Passiflora*, pois o satelitoma (conjunto de DNAsat) foi mais diverso e abundante especialmente nas espécies com GS menor (Sader *et al.* 2021). De forma similar, em *Agave* (Ramos *et al.* 2023), a única família de DNAsat identificada e isolada mostrou uma pequena banda *dot-like* no centrômero e, possivelmente, não deva responder pelas variações de GS no gênero. Em *Altromeria longistaminea*, os 11 DNAsat identificados representavam apenas 1,9% do genoma da espécie (Ribeiro *et al.* 2022). Contudo, a contribuição do conteúdo de DNAsat é variável entre os grupos vegetais, sendo que a proporção de DNAsat pode

ser diretamente, inversamente ou não relacionada com o GS (Biscotti *et al.* 2015, Samoluk *et al.* 2022).

Em *C. nitida*, as sequências de DNAsat respondem por aprox. 6,40% do genoma sequenciado (Ibiapino *et al.* 2022), valor equivalente ao encontrado para *B. schunkeana*. Outra similaridade entre *C. nitida* e *B. schunkeana* é a presença de blocos heterocromáticos bastante conspícuos e a expansão destes blocos parece estar associada à variação do GS nas duas espécies - em contraste com o comumente observado em plantas. Porém, a conclusão é embasa por motivos diferentes: enquanto em *C. nitida* os autores supõem que o aumento de GS deva-se a DNAsat pela representatividade desta fração em comparação aos dois TEs mais representativos (3,5% de SIRE e 4,19% de Tekay), nossos dados suportam que a variação de GS deva-se a expansão de DNAsat graças à análise comparativa com *B. picta*. Ainda em *Cuscuta*, *C. european* acumulou uma quantidade expressiva de DNAsat, 18%, e um único DNAsat isolado compõem todas as bandas heterocromáticas observadas nos cromossomos desta espécie (Vondrak *et al.* 2021). Assim, de maneira similar em *Cuscuta* e *Brasiliorchis*, DNAsat parece ser o principal responsável pela expansão do GS.

Além disso, da mesma forma que a anotação para um os cluster de *B. schunkeana* foi feita via *dot-plot*, e não pelo TAREAN, em *C. nitida* as autoras puderam anotar seis famílias de DNAsat não identificados pelo TAREAN (Ibiapino *et al.* 2022) e em *A. longistaminea*, oito dos 11 DNAsat identificados pelos autores, não foi identificado pelo TAREAN (Ribeiro *et al.* 2022). Isso reforça a necessidade de analisar todos os cluster bem representados, mesmo que a ferramenta automática do RE2 não indique como potencial DNAsat, especialmente levando em conta que o CL67 de *B. schunkeana* apresenta a distribuição diferencial mais impactante em

comparação à *B. picta* e não apresenta um grafo circular, mas apresenta um *dot-plot* bastante denso. O grafo do CL67 não remete a um DNAsat, mas apresenta traços que poderiam remeter a um TE. Ibiapino *et al.* (2022) mostraram que dois DNAsat mostraram similaridade com TE e sabe-se que TE e genes repetitivos podem contribuir para a formação e disseminação de DNAsat (Varani *et al.* 2015, Biscotti *et al.* 2015). Contudo, o cluster CL67 não mostrou similaridade com nenhuma sequência, nem mesmo TEs.

6. Conclusão

A despeito da grande importância dos TE na composição do repetoma das duas espécies de *Brasiliorchis* analisadas aqui, a expansão dos Ty3/Gypsy parece não explicar o aumento do tamanho de genoma em *B. schunkeana*, uma vez que a composição do repetoma entre *B. schunkeana* e *B. picta* é similar no que diz respeito aos TEs. Entretanto, diferenças claras e importantes foram detectadas na análise comparativa da representatividade dos DNAsat, suportando que, diferente da maioria das espécies vegetais, a variação do tamanho de genoma em *B. schunkeana* se deve a amplificação de DNAsat, em especial dos clusters CL 32 e 51.

Desta forma, *B. schunkeana* se destaca de diferentes formas: além das diferenças marcantes na morfologia, com suas flores negras, na distribuição geográfica, sendo a única espécie endêmica do Espírito Santo em uma área de refúgio do Quaternário, na citogenética, com aumento do GS e presença de grandes blocos heterocromáticos ricos em AT, *B. schunkeana* também difere na causa do aumento do GS, que nesta espécie foi graças a amplificação de DNAsat e não via expansão de TE, a causa mais frequente em plantas.

7. Referências

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215, 403–410. doi:10.1016/S0022-2836(05)80360-2
- Ambrožová, K., Mandáková, T., Bureš, P., Neumann, P., Leitch, I. J., Koblížková, A., ... Lysak, M. A. (2011). Diverse retrotransposon families and an AT-rich satellite DNA revealed in giant genomes of *Fritillaria* lilies. *Ann Bot*, Vol. 107, pp. 255–268. doi:10.1093/aob/mcq235
- Beaulieu, J., Leitch, I. J., Patel, S., Pendharkar, A., & Knight, C. A. (2008). Genome size is a strong predictor of cell size and stomatal density in angiosperms. *New Phytologist*, 975–986. doi:10.1111/j.1469-8137.2008.02528.x
- Bennett, M. D., & Leitch, I. J. (2005). Nuclear DNA amounts in angiosperms: progress, problems and prospects. *Annals of Botany*, 95, 45–90. doi:10.1093/aob/mci003
- Bennetzen, J. L. (2000). *Transposable element contributions to plant gene and genome evolution* (Vol. 42, pp. 251–269).
- Bennetzen, J. L., & Wang, H. (2014). The Contributions of Transposable Elements to the Structure, Function, and Evolution of Plant Genomes. *Https://Doi. Org/10.1146/Annurev-Arplant-050213-035811*, 65, 505–530. doi:10.1146/ANNUREV-ARPLANT-050213-035811
- Biscotti, M. A., Olmo, E., & Heslop-Harrison, J. S. (2015). Repetitive DNA in eukaryotic genomes. *Chromosome Research*, 23, 415–420. doi:10.1007/s10577-015-9499-z

- Blanco, M. A., Carnevali, G., Whitten, W. M., Singer, R. B., Koehler, S., Williams, N. H., ... Endara, L. (2007). Generic realignments in Maxillariinae (Orchidaceae). *Lankesteriana*, 7, 515–537.
- Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., ... Feschotte, C. (11 2018). Ten things you should know about transposable elements. *Genome Biology* 2018 19:1, 19, 1–12. doi:10.1186/S13059-018-1577-Z
- Cabral, J. S., Felix, L. P., & Guerra, M. (2006). Heterochromatin diversity and its co-localization with 5S and 45S rDNA sites in chromosomes of four Maxillaria species (Orchidaceae). *Genetics and Molecular Biology*, 664, 659–664.
- Chase, M. W., Cameron, K. M., Freudenstein, J. V., Pridgeon, A. M., Salazar, G., van den Berg, C., & Schuiteman, A. (2015). An updated classification of Orchidaceae. *Botanical Journal of the Linnean Society*, 177, 151–174. doi:10.1111/boj.12234
- Christenhusz, M. J. M., & Byng, J. W. (2016). The number of known plants species in the world and its annual increase. *Phytotaxa*, 261, 201–217. doi:10.11646/phytotaxa.261.3.1
- Doolittle, W. F., & Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. *Nature*, 284, 601–603. doi:10.1038/284601a0
- Egan, A. N., Schlueter, J., & Spooner, D. M. (2012). Applications of next-generation sequencing in plant biology. *American Journal of Botany*, 99, 175–185. doi:10.3732/ajb.1200020

- Elgin, S. C. R., & Grewal, S. I. S. (2003). Heterochromatin: silence is golden. *Current Biology* : CB, 13, R895-8. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/14654010>
- Eriksson, M. C., Mandáková, T., Mccann, J., Temsch, E. M., Chase, M. W., Hedrén, M., ... Wright, S. (2022). Repeat Dynamics across Timescales: A Perspective from Sibling Allotetraploid Marsh Orchids (*Dactylorhiza majalis* s.l.). *Molecular Biology and Evolution*, 39. doi:10.1093/molbev/msac167
- Flora e Funga do Brasil 2023. Jardim Botânico do Rio de Janeiro. Disponível em: <<https://floradobrasil.jbrj.gov.br/FB179>>. Acesso em: 15 out. 2022.
- Garcia, C. F., Gaiesky, V. L. S. V. (2015) Evolução cromossômica em eucariotos - rearranjos cromossômicos e elementos de transposição. In: Carareto, C. M. A., Monteiro-Vitorello, C. B., Van Sluys, M-A. (eds) Elementos de transposição - Diversidade, evolução, aplicação e impacto nos genomas dos seres vivos. Editora SBC, Ribeirão Preto & Editora FioCruz, Rio de Janeiro. 196p.
- Garnett, S. T., & Christidis, L. (2017). Taxonomy anarchy hampers conservation. *Nature*, 546, 25–27. doi:10.1038/546025a
- Givnish, T. J., Spalink, D., Ames, M., Lyon, S. P., Hunter, S. J., Zuluaga, A., ... Cameron, K. M. (2016). Orchid historical biogeography, diversification, Antarctica and the paradox of orchid dispersal. *Journal of Biogeography*, 43, 1905–1916. doi:10.1111/jbi.12854
- Givnish, T. J., Spalink, D., Ames, M., Lyon, S. P., Hunter, S. J., Zuluaga, A., ... Cameron, K. M. (2015). Orchid phylogenomics and multiple drivers of their

- extraordinary diversification. *Proceedings of the Royal Society B: Biological Sciences*, 282, 20151553. doi:10.1098/rspb.2015.1553
- Govaerts, R., Dransfield, J., Zona, S., Hodel, D. R., & Henderson, A. (2017). WCSP (2017). World Checklist of Selected Plant Families. *Royal Botanic Gardens*, 2016. Retrieved from <http://wcsp.science.kew.org>
- Gregory, R. (01 2005). The Evolution of the Genome. Elsevier Academic. 740p. doi:10.1016/B978-012301463-4/50003-6
- Guerra, M. (1988). Introdução à Citogenética Geral. Editora Guanabara.
- Guerra, M. (2012). Cytotaxonomy: The end of childhood. *Plant Biosystems*, 146, 703–710. doi:10.1080/11263504.2012.717973
- Guerra, M. (2008). Chromosome numbers in plant cytotaxonomy: Concepts and implications. *Cytogenetic and Genome Research*, 120, 339–350. doi:10.1159/000121083
- Gurusaran, M., Ravella, D., & Sekar, K. (10 2013). RepEx: Repeat extractor for biological sequences. *Genomics*, 102, 403–408. doi:10.1016/j.ygeno.2013.07.005
- Heitz, E. (1935). Chromosomenstruktur und Gene. *Zeitschrift Für Induktive Abstammungs- Und Vererbungslehre*, 70(1), 402–447. doi:10.1007/BF01741657
- Hsu, C. C., Chen, S. Y., Lai, P. H., Hsiao, Y. Y., Tsai, W. C., Liu, Z. J., ... Chen, H. H. (2020). Identification of high-copy number long terminal repeat retrotransposons and their expansion in *Phalaenopsis* orchids. *BMC Genomics*, 21. doi:10.1186/s12864-020-07221-6

- Ibiapino, A., Báez, M., García, M. A., Costea, M., Stefanović, S., & Pedrosa-Harand, A. (2022). Karyotype asymmetry in *Cuscuta* L. subgenus *Pachystigma* reflects its repeat DNA composition. *Chromosome Research*, 30, 91–107. doi:10.1007/s10577-021-09683-0
- Jiao, Y., Chanderbali, S., Landherr, L., Ralph, P. E., Wickett, N. J., & Ayyampalayam, S. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature*, 473, 97–102. doi:10.1038/nature09916
- Jiao, Y. (2018). Double the Genome, Double the Fun: Genome Duplications in Angiosperms. *Molecular Plant*, 11, 357–358. doi:10.1016/j.molp.2018.02.009
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30, 772–780. doi:10.1093/molbev/mst010
- Katoh, K., Rozewicki, J., & Yamada, K. D. (2018). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, 20, 1160–1166. doi:10.1093/bib/bbx108
- Kazazian, H. H. (2004). Mobile Elements: Drivers of Genome Evolution. *Science*, 303, 1626–1632. doi:10.1126/SCIENCE.1089670
- Kersey, P. J. (2019). Plant genome sequences: past, present, future. *Current Opinion in Plant Biology*, 48, 1–8. doi:10.1016/J.PBI.2018.11.001
- Kit S. (1961) Equilibrium sedimentation in density gradients of DNA preparations from animal tissues. *Journal of Molecular Biology* 3:711–716. doi: 10.1016/S0022-2836(61)80075-2.

- Knight, C. A., & Beaulieu, J. M. (2008). *Genome size scaling through phenotype space*. 101, 759–766. doi:10.1093/aob/mcm321
- Krebs, J. E., Goldstein, E. S., & Kilpatrick, S. T. (2017). *Lewin's genes XII*. Jones & Bartlett Learning.
- Levin, D.A. (2002) The role of chromosomal change in plant evolution. Oxford University Press, USA, 244p.
- Macas, J., Novak, P., Pellicer, J., Cizkova, J., Koblizkova, A., Neumann, P., ... Leitch, I. J. (2015). In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe fabaeae. *PLoS ONE*, 10. doi:10.1371/journal.pone.0143424
- MacGregor, H. C. (2013). C-Value Paradox. Brenner's Encyclopedia of Genetics: Second Edition, 258. <https://doi.org/10.1016/B978-0-12-374984-0.00368>
- Mayrose, I., & Lysak, M. A. (2021). The Evolution of Chromosome Numbers: Mechanistic Models and Experimental Approaches. *Genome Biology and Evolution*, 13, 0–34. doi:10.1093/gbe/evaa220
- Mehrotra, S., & Goyal, V. (2014). Repetitive Sequences in Plant Nuclear DNA: Types, Distribution, Evolution and Function. *Genomics, Proteomics & Bioinformatics*, 12, 164–171. doi:10.1016/J.GPB.2014.07.003
- Mendes, S., Moraes, A. P., Mirkov, T. E., & Pedrosa-Harand, A. (2011). Chromosome homeologies and high variation in heterochromatin distribution between Citrus L. and Poncirus Raf. as evidenced by comparative cytogenetic mapping. *Chromosome Research*, 19. doi:10.1007/s10577-011-9203-x

- Mhiri, C., Borges, F., & Grandbastien, M. A. (2022). Specificities and Dynamics of Transposable Elements in Land Plants. *Biology* 2022, Vol. 11, Page 488, 11, 488. doi:10.3390/BIOLOGY11040488
- Moraes, A. P., Leitch, I. J., & Leitch, A. R. (2012). Chromosome studies in Orchidaceae: Karyotype divergence in Neotropical genera in subtribe Maxillariinae. *Botanical Journal of the Linnean Society*, 170. doi:10.1111/j.1095-8339.2012.01266.x
- Moraes, A. P., Chinaglia, M., Palma-Silva, C., & Pinheiro, F. (2013). Interploidy hybridization in sympatric zones: The formation of *Epidendrum fulgens* × *E. puniceoluteum* hybrids (Epidendroideae, Orchidaceae). *Ecology and Evolution*, 3, 3824–3837. doi:10.1002/ece3.752
- Moraes, A. P., Engel, T. B. J., Forni-Martins, E. R., de Barros, F., Felix, L. P., & Cabral, J. S. (2022). Are chromosome number and genome size associated with habit and environmental niche variables? Insights from the Neotropical orchids. *Annals of Botany*, 130, 11–25. doi:10.1093/aob/mcac021
- Moraes, A. P., Simões, A. O., Alayon, D. I. O., Barros, F. D., & Forni-Martins, E. R. (2016). Detecting mechanisms of karyotype evolution in *Heterotaxis* (Orchidaceae). *PLoS ONE*, 11, 1–18. doi:10.1371/journal.pone.0165960
- Moraes, A. P., Koehler, S., Cabral, J. S., Gomes, S. S. L., Viccini, L. F., Barros, F., ... Forni-Martins, E. R. (2017). Karyotype diversity and genome size variation in Neotropical Maxillariinae orchids. *Plant Biology*, 19, 298–308. doi:10.1111/plb.12527

- Moraes, A.P., T. G., Mirkov, Guerra, M. (2008). Mapping the chromosomes of *Poncirus trifoliata* Raf. by BAC-FISH. *Cytogenetic and Genome Research*, 281, 277–281. doi:10.1159/000138897
- Negm, S., Greenberg, A., Larracuenta, A. M., & Sproul, J. S. (2021). RepeatProfiler: A pipeline for visualization and comparative analysis of repetitive DNA profiles. *Molecular Ecology Resources*, 21, 969–981. doi:10.1111/1755-0998.13305
- Nick, P. (2022). Is a genome more than its genes? *Protoplasma*, 259, 233–235. doi:10.1007/s00709-022-01744-3
- Ogur, M., Erickson, R. O., Rosen, G. U., Sax, K. B., & Holden, C. (1951). Nucleic acids in relation to cell division in *Lilium longiflorum*. *Experimental Cell Research*, 2, 73–89. doi:10.1016/0014-4827(51)90007-9
- Orgel, L. E., & Crick, F. H. C. (1980). Selfish DNA: the ultimate parasite. *Nature*, 284, 604–607. doi:10.1038/284604a0
- Pellicer, J., Hidalgo, O., Dodsworth, S., & Leitch, I. J. (2018). Genome size diversity and its impact on the evolution of land plants. *Genes*, 9. doi:10.3390/genes9020088
- Pinheiro, F., Barros, F. D. E., Singer, R. R. B., Koehler, S., Singer, B. L. R. B., Lindl, B., ... Singer, B. H. R. B. (2009). Morphometric analysis of the *Brasiliorchis picta* complex (Orchidaceae). *Revista Brasileira de Botânica*, 2, 11–21.
- Ramos, L. C., Báez, M., Fuchs, J., Houben, A., Carvalho, R., & Pedrosa-Harand, A. (2023). Differential Repeat Accumulation in the Bimodal Karyotype of *Agave L.* *Genes*, 14, 491. doi:10.3390/GENES14020491/S1

- Ribeiro, T., Vaio, M., Félix, L. P., & Guerra, M. (2022). Satellite DNA probes of *Alstroemeria longistaminea* (Alstroemeriaceae) paint the heterochromatin and the B chromosome, reveal a G-like banding pattern, and point to a strong structural karyotype conservation. *Protoplasma*, 259, 413–426. doi:10.1007/s00709-021-01681-7
- Ribeiro, T., Nascimento, J., Santos, A., Félix, L. P., & Guerra, M. (2021). Origin and evolution of highly polymorphic rDNA sites in *Alstroemeria longistaminea* (Alstroemeriaceae) and related species. *Genome*, 64, 833–845. doi:10.1139/GEN-2020-0159
- Ris, H., & Mirsky, A. E. (1949). Quantitative cytochemical determination of desoxyribonucleic acid with the Feugen nuclear reaction. *The Journal of General Physiology*, 33, 125. doi:10.1085/JGP.33.2.125
- Sader, M., Vaio, M., Cauz-Santos, L. A., Dornelas, M. C., Vieira, M. L. C., Melo, N., & Pedrosa-Harand, A. (2021). Large vs small genomes in *Passiflora*: the influence of the mobilome and the satellitome. *Planta*, 253, 1–18. doi:10.1007/S00425-021-03598-0/METRICS
- Samoluk, Sergio S., Vaio, M., Ortiz, A. M., Chalup, L. M. I., Robledo, G., Bertoli, D. J., & Seijo, G. (2022). Comparative repeatome analysis reveals new evidence on genome evolution in wild diploid *Arachis* (Fabaceae) species. *Planta*, 256. doi:10.1007/s00425-022-03961-9
- Samoluk, Sergio Sebastián, Robledo, G., Podio, M., Chalup, L., Ortiz, J. P. A., Pessino, S. C., & Seijo, J. G. (2015). First insight into divergence, representation and chromosome distribution of reverse transcriptase fragments

- from L1 retrotransposons in peanut and wild relative species. *Genetica*, 143, 113–125. doi:10.1007/S10709-015-9820-Y/
- Schubert, I., & Vu, G. T. H. (2016). Genome Stability and Evolution: Attempting a Holistic View. *Trends in Plant Science*, 21, 749–757. doi:10.1016/j.tplants.2016.06.003
- Schuiteman, A., & Chase, M. (2015). A reappraisal of *Maxillaria* (Orchidaceae). *Phytotaxa*, 225, 1–78. doi:10.11646/phytotaxa.225.1.1
- Schweizer, D. (1976). Reverse fluorescent chromosome banding with chromomycin and DAPI. *Chromosoma*, 58, 307–324. doi:10.1007/BF00292840
- Singer, R. B., Koehler, S., Carnevali, G. N., & Carnevali, K. (2007). *Brasiliorchis*: A New Genus for the *Maxillaria picta* Alliance (Orchidaceae, Maxillariinae). *Novon*, 17, 91–99. Retrieved from <http://128.227.186.212/herbarium/>
- Souza, P. C. S. S. (2022). Análise in silico da fração repetitiva do genoma de espécies de *Tephrosia* Pers. (Leguminosae Juss.). Dissertação de Mestrado. Programa de Pós Graduação em Biodiversidade Vegetal - PPGBVeg. Universidade do Estado da Bahia – UNEB
- Sueoka N. (1961) Variation and heterogeneity of base composition of deoxyribonucleic acids: A compilation of old and new data. *J. Mol. Biol.* 1961: 31–40. doi: 10.1016/S0022-2836(61)80005-3.
- Swift, H. (1950). The Constancy of Desoxyribose Nucleic Acid in Plant Nuclei. *Proceedings of the National Academy of Sciences of the United States of America*, 36, 643. doi:10.1073/PNAS.36.11.643

- Van de Peer, Y., Ashman, T. L., Soltis, P. S., & Soltis, D. E. (2021). Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell*, 33, 11–26. doi:10.1093/plcell/koaa015
- Varani, A. M., Carvalho, L. C. B., Zerillo, M. M., Monteiro-Vitorello, C. B. (2015) Elementos de transposição: classificação e mecanismos de mobilização. In: Carareto, C. M. A., Monteiro-Vitorello, C. B., Van Sluys, M-A. (eds) Elementos de transposição - Diversidade, evolução, aplicação e impacto nos genomas dos seres vivos. Editora SBC, Ribeirão Preto & Editora FioCruz, Rio de Janeiro. 196p.
- Vinogradov, A. E. (2003). *Selfish DNA is maladaptive : evidence from the plant Red List*. 19, 609–614. doi:10.1016/j.tig.2003.09.010
- Vondrak, T., Oliveira, L., Novák, P., Koblížková, A., Neumann, P., & Macas, J. (2021). Complex sequence organization of heterochromatin in the holocentric plant *Cuscuta europaea* elucidated by the computational analysis of nanopore reads. *Computational and Structural Biotechnology Journal*, 19, 2179–2189. doi:10.1016/j.csbj.2021.04.011
- Wang, Y., Chen, Y., Wei, Q., Chen, X., Wan, H., & Sun, C. (2022). Characterization of repetitive sequences in *Dendrobium officinale* and comparative chromosomal structures in *Dendrobium* species using FISH. *Gene*, 846, 146869. doi:10.1016/J.GENE.2022.146869
- Whitten, W., Blanco, M., Williams, N., Koehler, S., Carnevali, G., Singer, R., ... Neubig, K. (2007). Molecular phylogenetics of *Maxillaria* and related genera

- (Orchidaceae: Cymbidieae) based on combined molecular data sets. *American Journal of Botany*, 94, 1860–1889. doi:10.3732/ajb.94.11.1860
- Whitten, W. M., & Blanco, M. A. (2011). Defining generic limits in *Maxillaria*. *Orchids*, 104–113. Retrieved from www.AOS.Org
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., ... Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. Retrieved from www.nature.com/reviews/genetics
- Wicker, T. (2012). So Many Repeats and So Little Time: How to Classify Transposable Elements in Grandbastien, M. A., & Casacuberta J. M. (eds.), Plant Transposable Elements, Topics in Current Genetics 24, <https://doi.org/10.1007/978-3-642-31842-9>.
- Winkler, H. (1916). Über Die Experimentelle Erzeugung. *Journal of Botany*, 8, 417–531.
- You, R. N., Kim, W. C., Lee, K. H., Lee, Y. K., Shin, K. S., Cho, K., & Cho, D. H. (2013). REViewer: A tool for linear visualization of repetitive elements within a sequence query. *Genomics*, 102, 209–214. doi:10.1016/j.ygeno.2013.07.008
- Young, A. D., & Gillung, J. P. (2020). Phylogenomics — principles, opportunities and pitfalls of big-data phylogenetics. *Systematic Entomology*, 45, 225–247. doi:10.1111/syen.12406
- Zhang, Q. J., & Gao, L. Z. (2017). Rapid and recent evolution of LTR retrotransposons drives rice genome evolution during the speciation of AA-genome *Oryza* Species. *G3: Genes, Genomes, Genetics*, 7, 1875–1885. doi:10.1534/G3.116.037572/-/DC1

8. Material suplementar

Material Suplementar 1. Sequências consenso dos cluster identificados como DNA satélite com distribuição diferencial entre *B. schunkeana* (bastante representativos) e *B. picta* (poucos representativos ou ausentes).

>CL67_CONSENSUS

ACGATCCACTGAGATCCTGCCCTACGTCGCGCATTTTTCCNTGCCCGTCTN
ATAGAGGGAAGCTCGAGGAGGGATCAAAATTGCCAAAATTTGTTCTCTNGT
TGACAGGCATGCTAGCCATGCCTTGCTCATGGTCATCCCATGTGCATGTNG
CCTTCGCGACATCTCTATTTTCGAGGCAAAAATTTAAGAGCGTAATGCCTCA
CGGGCACGAGGGGCGGGGAGCGAGGCACCCAGGTGCACAGGTGCGNGSA
CCTGGGAGGCACCGNGTGCGCAAATGCGTGCACATGAAAAAGNAANCCTT
TTTTGGKKGTGC

>CL51_CONSENSUS

TAACAAAAATATAATTATTAATAATAAAGCTATTCTAAAATATAATAATA
ATATATTTAANTNWATWTGAWTTTTYNAATATAAATATTAATAATAAACW
TATTTKNANAYATAATMATAATATANTTTAATAAATNTAAATTTTAAATA
TAAATATTAATAATAAATCTTATTTTCAAATATAATATTAATATATTTAAT
TAAATCTAAATTTTAANTATTAATATTAATAATAAACTTATCTTCAAACAT
WATAATAAAWTATAAAATTTAAANTATAAATATTAATAATAAAACTTATTT
TCAAACATAATAATAATATATTTAATTAAATCTAAATTTTAAATATAAATA
TTAATAATAAACTATANNTTTTAAAAAYATAATAATAATATATTTAATTAA
ATTYTAAATTTTAAATATAAATATTAATAATAAAMTTATTTTCAAACATA
TATAATAATATATTTAATTAAATACTAAATTTTAAATATAAATATTAATA
ATAAACTTATTTTCATAAACATAATAAATAATATATTTAATTAAATCTTAA
TTTTAAATATAAATTATTAATAATAAACTTATTTTCAAACATAATAAAT
ATATTNAAATAAATCTAACTTTAAANATAAATATTTAATNTAATAAACTT
ATTTTCAAANATAATAATANTNATATATTTAATTAAATCTGAATTTTAAAT
ATAATTATTAATAATAAANTTATTTTCAAACATAATAAANATNATTAAT
TAAAATCTWAATYTWAWWWNTAAAWATTAATAATAAANCATATTTTCN
NWWATWAWAATWAWATAWTTAATTNAAATCTNAAWTTTAANWWWTA
AKNMTTAANAATAAATNWAWTTYWWANTTAANAATWAMATWATTAA
TTAAAATNNTAATTTAWTTATWAAATATTAATAATAAATCNWATNTNTA
ATTAAAATWNAATAATTAATTTAAATCTAAATTTAATTATTAATATTA
TAATAAATCATATTTTCTAATTTAAATTAANNAANTAATTNNAATNTRAA
NMTANWWWTAAWTATTAANTGTTAATAATCAACTTATTTTCWAAAAW
ANAAATAAATAAATAATTANTATTAATCTTGKAATTATAAATATTAATTAT
AATAATAAATTNATTTTAAATATAATAATAAWATATTTAATTTAAATCTT
ATNAATTTTAWAAATAAATATTAATAATAAATATATTTTCNAAWAAAAA
TAATATATATTNAATTAAANTTTAAATTATAAGATAATAAATATTAATAAT

AAATTATATTTTAACTAATAWAATAATAATAATAATTAATTAATTAATAAT
CTTAAATTTTATAATATANTTAAATATTAATAATAAANTNATATTTTCTAAT
TAAAATATTAATAATTAATTAANATTTNAATTTAAATTTAATTAATANT
ATTAATAATATTAATAATCATATTTTNTNAATTAANAATTAATAATTA
ATTTAAATCNTTTTGAATTTAATTATTAATATTATATNATAAAGTTCAT
ATTTTNTNGAAAATTAATAATAATTAATTAATAACTTTCTTNAATT
TAATTATTAATATTTAATAATAATCATTATTTTCTAATTAATAATTAATA
TAAATTAATTAATAATCTAAATTTAATTATTAAKTATTAATAATAAGTTTAA
TTTTCTAATTAATAATNTAATTAATAATTTAATTAATAATTTTGAATTTAATTA
TTAAATATTAATATATAYWAATATTAAAWTATNATTNANTANTAAAAT
TTWATAATTAATAWTGAAATTTANTTAATTAANTNATWWNAAWWAGNW
NGTATNWTNTAKTWWWAWWWNTAATNMWTNANNNNAAATTTTNA
ANTTAATTATTATWNWTTAANTAKAWNMWATNATTTTTATTAATATTGA
TTAATTAATAATTTCAATTTTAATTAATTATTTAATTTTATTTAAAAAATATGA
TTTATTATTAATGCTTAATAATTAATTCAGATTTTAATTAATTATTTAATT
TTAATAAAAAATATGATTTATCATTAAATATTTAATAATTAGATTCAAATTT
TAATTGGTCATTTAATTTTAATTAATAAAGTATGATTTTCTATTAATATTTAA
TAATTAATTCAGATTTTAATTAATTATTATATTTTAATTAGAAGATAWN
WTTTNTTATTAATATTTAATAATTAATTYAGATTTTAATTAATTATTTAA
TTTTAATAAAAAATATRATTTATNATTAATRTTTAATAATTANATTCAAAT
TTAATTANTTATTTAATTTTAATTAATAAARTATNNTTTWTTATTAATGNT
WAATAATTAATAATTNANATTTTAATTAATNWTNTWATTWTNATWWNNGA
AAATAWNWWNWWTGTTTATTATTGCTAAATAGTTAAATTCAGATTTTAAT
TAATATTATTATTATTATTTGAAAATATTTAGATTTACTTAAATATATTA
TTATTAYNTTTGAAAATAAGNTTATYATTAAWTTTATATTTAARGTTCAA
ATTTAATTAATATATTATTATTATGTTTGAAAATATGATTTATTATTAATA
TATNTAATAAAAAANATATANNTTTAATTAATAATTTATTTTTATTATATTG
AAAANTAATGTTTATTATTAATATTTAATCATTAAATTCAAGTTTAAATTA
ATTATTATTTTTTAATKWKWRMAWATRNNTTTATWATTAWTWTTTANW
WNTTAAAWTYWWRWTTTAATTANNTWTTTAATTTATWWTWTTTAAAAA
ANANGTTTATTATTAATATTTATATTTAAATTTAGANTTTAATTAATAT
WTTATTATTATTTTTGNAAAAAATAAGTTTATTATTAATATTTATATTTA
AAATTATAAATTTAATTNAATATATTATTATTATRTTTGAAAATAAGT
TTATTATTAATTTATATTTAAATTCAGATTTAATTAATATTAATTAT
TATTATNTANTTTAAAAATANGTTTATTATTAATATTTATATTTAAATTTA
GATTTAATTAATATATTATTATTATGTTTAAAAANAWNNTTATTATTAAT
ATTTATATATTAATAATTCAGATTTAATTAATATATTATTATTATATTTGAA
AATAAGTTTAAANTTATTAATATTTATAATTTAAATTYANATTTAATTAA
ATATATTATTATTATGTTTAAAAATATGTTTATTATTAATATTTATATTTAA
AATTCAGATTTAATTAATATATTATTATTATATTTAAAAATAAGTTTATT
ATTNATATTWAWWTTMARATTNNNANNTNNTTAAATATATTATNATTA
TRTTTAAAAATAAGTTTANTATTAATAWTWWTATTTAAATNYATATTTA
ATTAANATANTATTATTATTTTWTAAATAWTGTTTATTTATTAANTTT
ATTATTTAAATTTWAATTTAATTATATATATTATTATNTATNAWTWAAA

ATTAANWTTTATTATTAATAATTATATTTTAAAATTTATATTTAATTAAATA
TATTATTATTATTTTTAAAAAATAAGTTTATTATTAATATTAATATTTAAAA
TTTAGATTTAATTAAATATATTATTATTATWWTTKTTNAAAATAANTTTGA
TTTATTATTAATATTTATATTTAAAATTTAGATTTAATTTAATNTATTANTA
TTWTAWWTAWAAWNNANNNTTTATTATTAATATTTAATATATTAAATTY
AAAATTTTAATTAATTATTTAATTTTAATTAGAAAATAAAACTTATTATTA
ATACTTAATAATTAATTTAGATTTTAATTAATTATTTTAATTTTAATTAGA
AAATATGATTTATTATTAATATTTAATAATTAAATTTAGATTTTAATTAATT
ATTTAATTTTAATNAATGAAAATATGATTTATNATTAATATTTWAAAATTA
GATTTAGNATTAATTAACCTTANTATWWNTAWWNWTGAAAATAAGTTTA
TTATTAATAKTWATATTNAAARTTTAGANTTAATTAANATATTAWTATT
ATNTTTGAAAATAWKTTTATWATTAATATTTATANTTCATACAAANTTTA
AATTTATTTAAATATATTATTATTTTNNNTTTGAAAATAAATTTATTATTAAT
ATTTATATTTAWAATWTANATTAAAWTAAATATTATTANTATATTTATAT
TTAAAATTTAGATYWAATTAATATTATTAWTATTATATTTNAAAATAAG
TTTATTATTAANTNTTNATATTTAAAATTCAAATTTAATTAATTATATTATT
ATTATGTTTGAAAATAAGTTTATTATTAATAATTATATTTAAAATTTAGAT
TTAATTAAATATATTATTATTATTATGTTTGAAAATAAGTTTATTATTAATA
TTTATATTTAAAATTTAGATTTAATTAATATATTATTATTATATTTAAAAA
TAAGTTTATTATTAATATTNATATTTAAAATTTAMATTTAATTAATATNA
AYTTTAAATATWAATWWTAATAATAAACTTATTTTCAAANATAATAATA
ATATATTTAATTAATMTTRAATTTTAAATATAATNATWAAYAANAANAANTT
ATTTTCAAAYATAATAWTAATAWWTTAATTAAAWNNWNAATTTWAW
WWWTNAAATATWAAWAWNAAWTNNTATTTTCWAAYTAAAATTAATAA
CTATTTAAAATTTGAATTTAATTATTAAGCATTAAATAATATATCATATTTT
TATTAATAATTAATAATTAATTAATAATATAAATTTAATTATTAATAATAA
TAATAAATCAGGACTTCTAATTAATAATAATAATTAATAATTAATAATTTCAAT
TTGATTATTAATATTAACAATAAATCATACTTTCTAATTATAATTAAATA
ATTAATTAATAATTTGAATATAATTATTAATATTAATAATAAATCAAATTT
TTTAAATTCATATTTGTTGCC

>CL32_CONSENSUS

TATANNTTTAACTTAANNNTAATTANATTAANCTTANATNANNACNTTAA
NTTAAAANTNNANTNATNANAANTTNNAATTTATTTANNTNANTANANAA
NATNAACATNTNTNATNTTANANNNTANNTTRAANNNAATTAANNNNNTA
AAATAACNTATANNAANTTNAAANATANNANANCNANAACNAAACNTTAA
TCNTNATNAANNCATAANTNATNAAANNNTTTTAANAANATAAAMAAAAN
ATAANNTAANAATAACTYTATNTTTNANAAANNCTAATNATAANCTCAAN
TNAATCANANTNNTNNNANNNNCNNNNANTTNNTNAAANTNANNNNNNAN
NTNNANNNNANNNNAANTNANAANNNAANNNNANANNTNANNNNNANNNN
NNANNNNNNTNNNNNTNNNNNTNANNNANNANNNNNANNNNATNNATAA
NNNNNNNTANTTTNNAANTNTAATNATAANNTANNNNNNNNCNNTNTNNAT