

**UNIVERSIDADE FEDERAL DE PERNAMBUCO**

**CENTRO DE BIOCÊNCIAS**

**LUIZ GUSTAVO SANTANA LUNA**

**DECLÍNIO DOS DINOSSAUROS E SUCESSO DOS  
POLIPLOIDES: CONSEQUÊNCIAS GENÔMICAS DA  
PALEOPOLIPLOIDIA NA SUBFAMÍLIA *Bombacoideae*  
(*Malvaceae*) NO FINAL DO CRETÁCEO**

**RECIFE**

**2022**

**LUIZ GUSTAVO SANTANA LUNA**

**DECLÍNIO DOS DINOSSAUROS E SUCESSO DOS  
POLIPLOIDES: CONSEQUÊNCIAS GENÔMICAS DA  
PALEOPOLIPLOIDIA NA SUBFAMÍLIA *Bombacoideae*  
(*Malvaceae*) NO FINAL DO CRETÁCEO**

Trabalho de Conclusão de Curso de Graduação apresentado ao Bacharelado em Ciências Biológicas com ênfase em Ciências Ambientais da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de bacharel.

Orientador (a): Lucas Alexandre de Souza Costa

Coorientador (a): Luiz Gustavo Rodrigues Souza

RECIFE

2022

Ficha de identificação da obra elaborada pelo autor,  
através do programa de geração automática do SIB/UFPE

Luna, Luiz Gustavo Santana.

Declínio dos dinossauros e sucesso dos poliploides: consequências genômicas da paleopoliploidia na subfamília Bombacoideae (Malvaceae) no final do Cretáceo / Luiz Gustavo Santana Luna. - Recife, 2022.  
50 : il., tab.

Orientador(a): Lucas Alexandre de Souza Costa  
Coorientador(a): Luiz Gustavo Rodrigues Souza  
Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Biociências, Ciências Biológicas /Ciências Ambientais - Bacharelado, 2022.

1. Evolução. 2. DNA repetitivo. 3. Bombacoideae. 4. Paleopoliploidia. 5. Malvaceae. I. Costa, Lucas Alexandre de Souza. (Orientação). II. Souza, Luiz Gustavo Rodrigues. (Coorientação). III. Título.

580 CDD (22.ed.)

**LUIZ GUSTAVO SANTANA LUNA**

**DECLÍNIO DOS DINOSSAUROS E SUCESSO DOS  
POLIPLÓIDES: CONSEQUÊNCIAS GENÔMICAS DA  
PALEOPOLIPLÓIDIA NA SUBFAMÍLIA *Bombacoideae*  
(*Malvaceae*) NO FINAL DO CRETÁCIO**

Trabalho de Conclusão de Curso apresentado à Coordenação do Curso de Bacharelado em Ciências Biológicas com ênfase em Ciências Ambientais, da Universidade Federal de Pernambuco, como parte dos requisitos à obtenção do grau de Bacharel em Ciências Biológicas.

Data de Aprovação: 13/10/2022

Nota: \_\_\_\_\_

BANCA EXAMINADORA:

\_\_\_\_\_  
Dr. Lucas Alexandre de Souza Costa (Orientador)

Departamento de Botânica - UFPE

\_\_\_\_\_  
Dra. Yhanndra Karine Dias Silva (1º Titular)

Departamento de Botânica - UFPE

\_\_\_\_\_  
MSc. Claudio César Montenegro Júnior (2º Titular)

Departamento de Botânica – UFPE

\_\_\_\_\_  
MSc. Jéssica Nascimento de Aguiar (Suplente)

Departamento de Botânica – UFPE

RECIFE

2022

## AGRADECIMENTOS

Dentre todos esses 23 anos vividos, acredito que agradecer com palavras escritas em um papel não supre o mínimo que pode ser feito ao terminar mais um ciclo como esse, mas é o que me resta então vamos lá. Primeiramente queria agradecer a minha mãe que, como figura materna e presente, sempre me incentivou, assim como meu pai e meu irmão me mostrando os trilhos que, para eles, é o melhor a ser traçado, o da educação. Sem a presença da minha família muito provavelmente eu nunca iria experimentar os quatro anos mais importantes da minha vida até então.

Formação de um pensamento crítico, me entender como um ser político, ter experiências profissionais e ver o mundo com outros olhos foram, entre outras coisas, o que a UFPE estava pronta pra me oferecer mesmo com tanta dificuldade proporcionada pro um desgoverno que perdurou por toda a minha formação. Dessa forma eu agradeço a todos os professores, técnicos e servidores que resistem e continuam a fazer da UFPE o ser vivo e pensante que ela é. Mais especificamente, agradeço aos amigos que fiz na graduação, Gabriel, Vinícius, Giovanna, Larissa, Jeisa e Nathalia e aos Professores Luiz Gustavo e Andrea Pedrosa, aos alunos de iniciação científica, mestres, mestrandos, doutores e doutorandos que fazem e já fizeram parte do laboratório de Citogenética Vegetal do departamento de botânica da UFPE, como Acalene, Amália, Amanda, Ana, Bruna, Arthur, Brena, Cláudio, Erton, Géssica, Gustavo (GG), Jéssica, Mariela, Natália, Pablo, Paulo, Rayssa, Thiago, Yhanndra, Yennifer e Júlia. Além, é claro, de toda orientação e companheirismo de Lucas Costa, um amigo que a ciência me deu.

Não poderia de deixar de agradecer aos meus amigos de infância, o “Nós 5”, que por algum motivo ainda não conhecido, nunca afroxou os la. Desde 2010, Arthur, Flávio, Thomaz e Thiago tem feito parte da minha história e o meu desenvolvimento ao longo dos anos, sempre presentes, apoiando e de braços abertos para qualquer tipo de situação. Dizem que na vida amigo é algo raro, dessa forma talvez eu tenha gastado muito da minha sorte naquele ano.

Esse último parágrafo eu dedico para agradecer alguém mais que especial, a pessoa que me mostrou uma realidade totalmente fora da minha bolha, me deu amor, atenção e esteve ao meu lado nos momentos mais difíceis não largando

minha mão em nenhum segundo, me ajudando a enfrentar todos os meus medos e correr atrás dos meus objetivos, alguém que me fez persistir. Jéssica Nascimento é uma das cientistas que mais admiro, seu compromisso com a ciência é algo inspirador e tenho sorte de dividir a vida ao seu lado.

***“Um passo à frente e você não está no mesmo lugar”.*** (Chico Science, 1996)

## RESUMO

A poliploidia (também conhecida como *Whole Genome Duplication* - WGD) e a amplificação/desamplificação de DNAs repetitivos (*repeats*) são os principais mecanismos responsáveis pela evolução do genoma em plantas. Muitas linhagens de angiospermas têm evidências de antigas WGD, a maioria estimada em 66 milhões de anos, sugerindo um sucesso dos poliploides após as mudanças globais que levaram à extinção dos dinossauros no final do Cretáceo. Nesse sentido, pouco é conhecido sobre a evolução das frações repetitivas dos genomas em linhagens que passaram por eventos antigos de poliploidia (paleopoliploidia). O clado Malvatheca (Malvaceae) representa um modelo interessante para estudar tal relação, pois a subfamília Bombacoideae apresenta números cromossômicos altos ( $2n = 86$  a  $2n = 276$ ) e uma provável origem paleopoliploide. Por outro lado, a subfamília Malvoideae apresenta números cromossômicos mais baixos e eventos pontuais de neopoliploide. Nesse contexto, a análise dos elementos repetitivos nessas subfamílias ajuda a elucidar o impacto da paleopoliploidia na evolução dos repeats. Aqui, caracterizamos a fração repetitiva genômica de 14 espécies de Bombacoideae e 11 Malvoideae e inferimos relações filogenéticas com base em três conjuntos de dados: plastomas, abundância/similaridade dos repeats e domínios específicos de proteínas de elementos transponíveis. As filogenias baseadas na abundância e na similaridade de repeats não apresentaram topologias inteiramente congruentes com a filogenia de plastoma, sugerindo que esse método pode falhar com grupos antigos ou de origem paleopoliploide. Além disso, para a filogenia baseada em abundância de repeats, os tamanhos dos ramos foram mais longos em Malvoideae e mais curtos em Bombacoideae, indicando um possível efeito da diploidização sobre a fração repetitiva deste grupo. No geral, domínios proteicos de diferentes elementos transponíveis não mostraram um padrão claro, com alguns elementos evoluindo aleatoriamente, enquanto outros evoluíram de acordo com as relações filogenéticas, sendo possível separar bem elementos pertencentes a representantes das duas subfamílias, sem efeito claro da paleopoliploidia. Por fim, foi encontrada uma correlação significativa entre número cromossômico e diversidade de repeats, denotando um possível efeito da poliploidia sobre a variedade de linhagens de elementos repetitivos. Desta forma, nossos resultados mostram que a paleopoliploidia exerce um efeito mais forte sobre a diversidade de elementos repetitivos do que sobre a variação da abundância ou sobre as sequências que os compõem. Este estudo fornece insights sobre a evolução

da fração repetitiva em grupos paleopoliploides e destaca o potencial dos repeats para entender a sistemática e a evolução das plantas.

Palavras-chave: Bombacoideae, Malvoideae, Paleopoliploidia, DNA repetitivo, Malvatheca

## ABSTRACT

Polyploidy (also known as Whole Genome Duplication - WGD) and amplification/de-amplification of repetitive DNAs (repeats) are the main mechanisms responsible for genome evolution in plants. Many angiosperm lineages have evidence of ancient WGD, most estimated to be 66 million years old, suggesting a success of the polyploids after the global changes that led to the extinction of the dinosaurs at the end of the Cretaceous. On the other hand, it is still not very well known about the evolution of the repetitive fractions of genomes in lineages that underwent ancient events of polyploidy (paleopolyploidy). The Malvaceae clade (Malvaceae) represents an interesting model to study this relationship, as the Bombacoideae subfamily has high chromosome numbers ( $2n = 86$  to  $2n = 276$ ) and a probable paleopolyploid origin. On the other hand, the Malvoideae subfamily has lower chromosome numbers and punctual neopolyploid events. In this context, the analysis of the repetitive elements in these subfamilies can help to elucidate the impact of paleopolyploidy on the evolution of repeats. Here, we characterized the genomic repeat fraction of 14 Bombacoideae and 11 Malvoideae species and inferred phylogenetic relationships based on three datasets: plastomes, repeat abundance/similarity, and protein-specific domains of transposable elements. The phylogenies based on the abundance and similarity of repeats did not show topologies entirely congruent with the plastome phylogeny, suggesting that this method may fail with ancient groups or groups of paleopolyploid origin. Furthermore, for the phylogeny based on repeat abundance, branch sizes were longer in Malvoideae and shorter in Bombacoideae, indicating a possible effect of diploidization on the repetitive fraction of this group. In general, protein domains of different transposable elements did not show a clear pattern, with some elements evolving randomly, while others evolved according to phylogenetic relationships, being possible to separate well elements belonging to representatives of the two subfamilies, with no clear effect of paleopolyploidy. Finally, a significant correlation was found between chromosome number and repeat diversity, denoting a possible effect of polyploidy on the variety of lineages of repetitive elements. Thus, our results show that paleopolyploidy exerts a stronger effect on the diversity of repetitive elements than on the variation in abundance or on the sequences that compose them. This study provides insights into the evolution of the repeat fraction in paleopolyploid groups and highlights the potential of repeats to understand plant systematics and evolution.

Keywords: Bombacoideae, Malvoideae, Paleopolyploidy, repetitive DNA, Malvaceae

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>10</b>
1.1 OBJETIVOS.....	12
1.1.1 GERAL .....	12
1.1.2 ESPECÍFICOS.....	12
<b>2 REFERENCIAL TEÓRICO</b> .....	<b>12</b>
<b>2.1 GENOMA DAS PLANTAS</b> .....	<b>12</b>
2.1.1 DNA REPETITIVO .....	13
2.1.2 DNA REPETITIVO EM TANDEM .....	14
2.1.3 DNA REPETITIVO DISPERSO .....	15
<b>2.2 SEQUENCIAMENTO DE NOVA GERAÇÃO</b> .....	<b>16</b>
<b>2.3 FILOGENÔMICA</b> .....	<b>16</b>
2.4 POLIPLOIDIA.....	17
<b>2.5 MALVACEAE</b> .....	<b>18</b>
<b>3 MATERIAIS E MÉTODOS</b> .....	<b>19</b>
<b>3.1 AQUISIÇÃO DE DADOS NGS E CARACTERIZAÇÃO DA FRAÇÃO REPETITIVA</b> .....	<b>19</b>
<b>3.2 DIVERSIDADE DOS REPEATS E CORRELAÇÃO COM O NÚMERO CROMOSSÔMICO</b> .....	<b>20</b>
<b>3.3 RELAÇÕES FILOGENÉTICAS BASEADAS EM PLASTOMAS COMPLETOS</b> .....	<b>20</b>
<b>3.4 RELAÇÕES FILOGENÉTICAS BASEADAS EM DNA REPETITIVO</b> .....	<b>20</b>
<b>3.5 FILOGENIA DE DOMÍNIOS CONSERVADOS DE ELEMENTOS TRANSPONÍVEIS</b> .....	<b>21</b>
<b>4 RESULTADOS</b> .....	<b>21</b>
<b>4.1 ARTIGO: DECLÍNIO DOS DINOSSAUROS E SUCESSO DOS POLIPLOIDES: CONSEQUÊNCIAS GENÔMICAS DA PALEOPOLIPLOIDIA NA SUBFAMÍLIA BOMBACOIDEAE (MALVACEAE) NO FINAL DO CRETÁCEO</b> .....	<b>21</b>
<b>5 REFERÊNCIAS</b> .....	<b>40</b>

## 1 INTRODUÇÃO

O estudo de dados cariológicos, como os números cromossômicos, contribuem para uma maior compreensão acerca da evolução das espécies, bem como da evolução do genoma eucarioto (Guerra 2008). As angiospermas, particularmente, apresentam uma ampla variação de número cromossômico, com números entre  $2n = 4$  até  $2n = 640$  (Uhl, 1978; Roberto, 2005), e essa variação está principalmente relacionada com alguns eventos de rearranjos cromossômicos, tais como fissão, fusão e com a duplicação do complemento cromossômico total, também conhecida por poliploidia (Mayrose e Lysak 2021). A poliploidia é o fenômeno caracterizado pela presença de dois ou mais genomas completos, podendo ser de dois tipos distintos: a autopoliploidia, que se refere à multiplicação de conjuntos de cromossomos geneticamente idênticos dentro de uma única espécie, e a alopoliploidia, a qual envolve a hibridação entre duas espécies geneticamente diferentes gerando um genoma com dois subgenomas (Ramsey e Schemske, 1998, 2002, Husband et al., 2013). Esses eventos são muito importantes na evolução das plantas, não só por afetar a constituição genômica de uma espécie, mas também por refletir diretamente na diversificação dos organismos, gerando uma alta taxa de especiação (Weiss-Schneeweiss et al., 2013, 2015).

Os efeitos da poliploidia tais como a duplicação de genes e as mudanças ocorridas no espaço gênico podem ocasionar o surgimento de novidades evolutivas. Um exemplo é o surgimento das flores, o qual é atribuído a uma duplicação do genoma total muito antigo, ocorrido na base de diversificação das angiospermas (Dodsworth; Chase; Leitch, 2016). Sabendo-se da existência de eventos antigos e recentes de poliploidia, é possível também caracterizar as espécies como paleo- e neopoliploides, de acordo com a origem (antiga ou recente, respectivamente) do evento de poliploidização (Guerra 2008). Apesar dessa caracterização, por vezes, se torna difícil reconhecer neo e paleopoliploides, uma vez que, as espécies, após a divergência do genoma poliploide, passam por vários processos de diploidização, tais como a dispoliploidia descendente, fazendo com que o organismo volte a ter um comportamento diploide (Soltis et al. 2015), bem como mudanças em sua composição e organização genômica (Weiss-Schneeweiss et al, 2015).

Além da presença de sequências codificantes, os genomas das espécies apresentam também uma grande porcentagem composta por DNA não-codificantes altamente repetitivas, (Elliott; Gregory, 2015). Essas sequências repetitivas podem ser

divididas em dois grupos: as sequências de DNA repetitivo de forma dispersa e a sequências de DNA repetitivos em tandem (Weiss-Schneeweiss et al., 2015). Os DNA repetitivos em tandem são elementos repetitivos um seguido do outro formando blocos, se destacando entre eles os satélites (Garrido-Ramos, 2017). Já o DNA disperso, é encontrado disposto de forma espalhada ao longo de todo o conjunto cromossômico, se destacando os elementos transponíveis, que por sua vez tem a capacidade de se deslocar dentro do genoma (Arkhipova, 2017). Esses elementos repetitivos podem ser classificados em Classe I e Classe II dependendo do seu mecanismo de transposição (Finnegan, 1989). Dessa forma, os elementos de Classe II, representados principalmente por retrotransposons, também podem ser subdivididos em LTR (*Long Terminal Repeats*), o qual, também são subdivididos em Ty1/copia e Ty3/gypsy, elementos encontrados principalmente na heterocromatina das angiospermas (Macas et al, 2015).

Atualmente, diversos estudos vêm sendo realizados com foco na análise e identificação da fração repetitiva dos genomas, sendo estes realizados a partir de sequenciamento do genoma como os de baixa cobertura (Dodsworth et al. 2015). A análise da fração repetitiva pode ser analisada e caracterizada com a utilização de plataformas como o RepeatExplorer e o TAREAN (Novak et al, 2013). Estudo atuais, revelaram a possibilidade do uso de sequenciamentos de DNA de baixa cobertura, como por exemplo o “Target-capture”, que atua sequenciando regiões alvos, para estudos que visam a análise do DNA repetitivo podendo assim reutilizar dados utilizados em estudos anteriores. (Costa et al. 2021). Outro importante avanço no estudo do DNA repetitivo foi a criação de métodos de inferência filogenômica baseados tanto na abundância (Dodsworth 2015) como na similaridade (Fan et al 2015) de sequências repetitivas. Além do potencial de informação taxonômica, filogenias baseadas em DNA repetitivo podem ser alvo interessante de estudo para grupos poliploides de origem antiga.

A família Malvaceae, pertencente a ordem Malvales, possui aproximadamente 4.225 espécies aceitas, divididas em 9 subfamílias, as quais se distribuem em 244 gêneros com traços morfológicos de difícil interpretação (Stevens 2001; Bayer e Kubitzki 2003; Christenhusz e Byng 2016). Dessa forma, atrelado ao grande número de representantes, as relações filogenéticas dentro dessa família são bastante controversas. As subfamílias Malvoideae (~1800 spp) e Bombacoideae (160 spp) estão distribuídas por todo o globo e são muito importantes economicamente na

agricultura, silvicultura e horticultura, como o algodão (*Gossypium* spp) e, por serem próximas filogeneticamente, são agrupadas em um clado específico denominado Malvatheca. Do ponto de vista citogenético, a subfamília Bombacoideae, possui uma provável origem paleopoliploide, se diferenciando das outras subfamílias por ter números de cromossomos significativamente maiores (Marinho et al. 2014). No presente trabalho foram caracterizadas as frações repetitivas de 13 espécies de Bombacoideae e 11 de Malvoideae, bem como a inferência filogenética da abundância e similaridade dos *repeats* e uma topologia de domínios proteicos específicos. Além disso, foi construída uma filogenia de plastomas, a fim de testar o impacto da paleopoliploidia na fração repetitiva dos genomas de representantes da subfamília Bombacoideae.

## 1.1 OBJETIVOS

### 1.1.1 Geral

Avaliar o impacto da paleopoliploidia na fração repetitiva dos genomas de espécies da subfamília Bombacoideae.

### 1.1.2 Específicos

- Caracterizar a fração de DNA repetitivo de 24 espécies da família Malvaceae, sendo 13 pertencentes a subfamília Bombacoideae e 11 a subfamília Malvoideae.
- Inferir relações filogenéticas com base na abundância de *repeats*, similaridade de *repeats* e domínios proteicos de elementos transponíveis
- Realizar uma filogenia a partir de plastomas completos para as espécies analisadas no estudo.
- Comparar as diferentes árvores filogenéticas geradas no presente estudo, verificando as diferenças entre as topologias de cada uma.

## 2 REFERENCIAL TEÓRICO

### 2.1 Genoma das plantas

O genoma é a sequência completa de DNA de um organismo. Além do genoma nuclear, nas plantas, encontram-se também o genoma de cloroplasto (plastoma) e o genoma das mitocôndrias (mitogenoma). O genoma nuclear das plantas é composto por uma gama

de sequências, contudo, as mais estudadas são as de DNA de cópia única e o DNA repetitivo (Heslop-Harrison; Schwarzacher, 2011). A complexidade e diversificação das plantas é

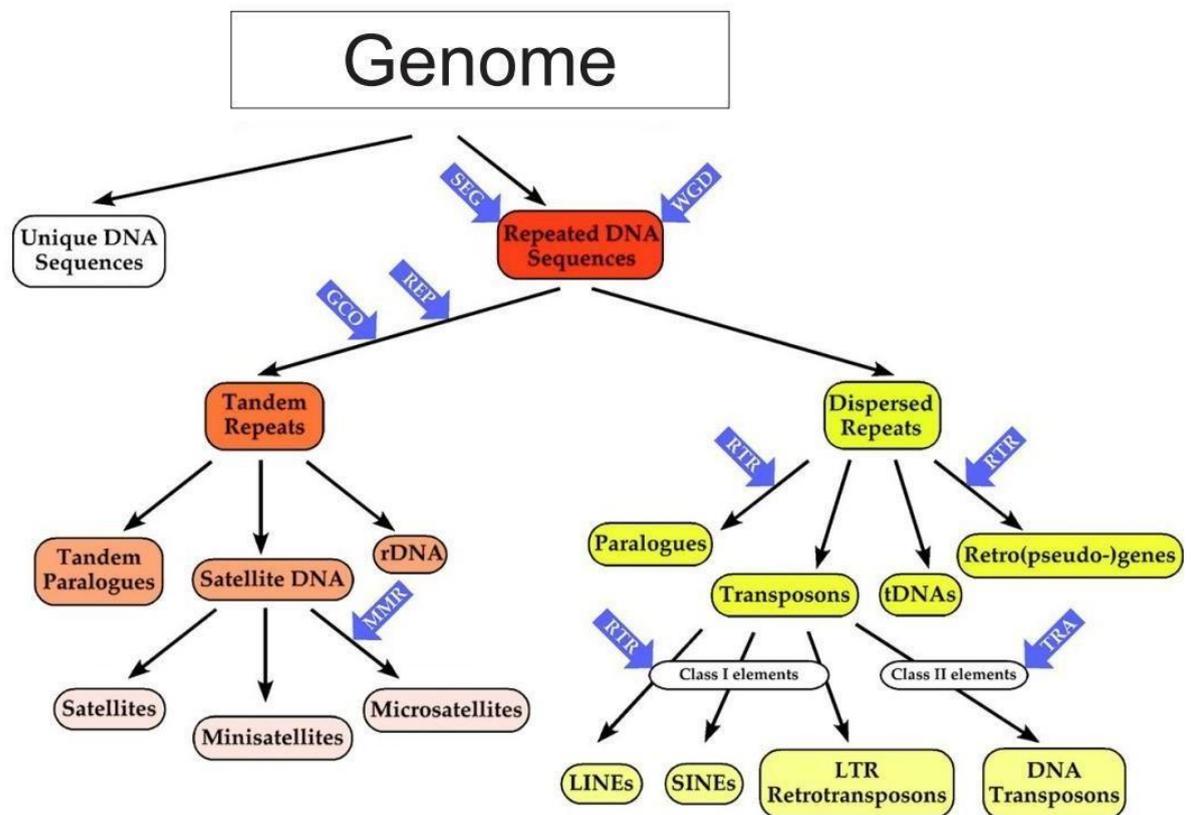
refletida também em seus genomas, onde as variações encontradas nos genomas das espécies são geralmente ocasionadas, principalmente, por processos evolutivos como a poliploidia e a amplificação e/ou deleção de DNA repetitivo, na forma de satélites e elementos de transposição (Mehrotra; Goyal, 2014; Kelly et al., 2015).

### 2.1.1 DNA repetitivo

O DNA repetitivo é o componente principal de grande parte dos genomas das plantas e foi por muito tempo considerado como DNA lixo, uma vez que, pouco se sabia sobre suas funções e importância nos genomas das espécies hospedeiras (Biscotti; Olmo; Heslop-Harrison, 2015). Nesse âmbito, os estudos de Barbara McClintock com o milho na década de 40 com a descoberta de um grupo de DNA repetitivo em particular, denominados “Elementos transponíveis”, reforçam a necessidade de se estudar mais a fundo alguns desses DNAs, visando compreender o papel desses elementos na evolução dos genomas das espécies (McClintock, 1948 e 1984; Biémont, 2010).

Com o avanço das técnicas moleculares e genômicas é possível inferir sobre o papel do DNA repetitivo dentro dos genomas, seu comportamento conforme as variáveis ambientais, o aumento ou diminuição dos tamanhos dos genomas através de deleções e/ou ampliações (Lyu et al., 2018; Schley et al., 2021), ou até mesmo ativando ou silenciando genes por meio de inserções em regiões gênicas (Achrem; Szućko; Kalinka, 2020).

Além dos aspectos que dizem respeito às funcionalidades e papéis que o DNA repetitivo assume nos genomas, hoje muito se sabe também acerca de sua diversidade, classes e tipos, sendo ele normalmente dividido, para fins de estudo, em duas grandes categorias, os DNAs repetitivos em tandem e os DNAs dispersos pelo genoma (Rao et al., 2010; Šatovic-Vukšić & Plohl, 2021).



Fonte: Modificado de Richard et al. 2008.

### 2.1.2 DNA repetitivo em tandem

O DNA repetitivo em tandem, representado pelos satélites, minissatélites, microssatélites e o DNA ribossomal formam longos arranjos de unidades de repetições chamados monômeros, sendo os principais componentes da heterocromatina e considerados como as sequências mais dinâmicas dos genomas, devido à sua grande variabilidade estrutural e de abundância, refletindo assim na sua alta taxa de evolução (Garrido-Ramos, 2017). Esses DNAs podem formar grandes blocos nos genomas e são representados principalmente por DNA ribossomal, DNA centromérico, DNA telomérico, famílias de genes codificadores de proteínas e pelo DNA satélite (Biscotti; Olmo; Heslop-Harrison, 2015; Garrido-Ramos, 2017).

Dentre os diversos tipos de DNAs repetitivos em tandem, destaca-se o DNA satélite, o qual tem sido considerado como um dos principais componentes da heterocromatina, podendo ser encontrado em diferentes partes dos cromossomos das espécies eucariotas, desde o centrômero até principalmente as regiões subtelo méricas e pericentroméricas (Guerra, 2000; Dos Santos et al., 2021). Essas sequências, além de serem um dos tipos mais dinâmicos nos genomas, devido à sua

grande variabilidade estrutural e de abundância (o que reflete em sua alta taxa de evolução) (Garrido-Ramos, 2017); podem formar famílias distintas e serem espécie-específicas, ou até mesmo compartilhadas entre espécies distintas. Desse modo, mesmo que em menor proporção que os elementos repetitivos dispersos, os satélites podem vir a contribuir para o aumento e/ou diminuição do tamanho dos genomas das espécies (Garrido-Ramos, 2017; Kelly et al., 2015).

### 2.1.3 DNA repetitivo disperso

O DNA repetitivo disperso, representado principalmente pelos Elementos de transposição, trata-se da fração repetitiva mais abundante dentro dos genomas das plantas (Bourque 2009). Esses elementos transponíveis são divididos em duas grandes classes, os elementos de classe I, representados principalmente pelos retrotransposons; e os elementos de classe II, que tem como principal representante os transposons (BOURQUE et al., 2018). Esses elementos, que se encontram em abundância nos genomas devido à sua notável capacidade de se replicar e inserir em novos locais do genoma independente da homologia de sequências, diferenciam-se entre si pelos seus intermediários de transposição, tendo como intermediários de RNA os de classe I e intermediários de DNA os de classe II (Feschotte; Jiang; Wessler 2002).

Os elementos de classe I são divididos em elementos não LTRs (do inglês *Long Terminal Repeats*), representados pelos LINEs (do inglês *Long Interspersed Elements*) que são elementos autônomos longos intercalados, os SINEs (do inglês *Short Interspersed Nuclear Elements*), elementos curtos intercalados não-autônomo, os quais dependem dos LINEs para sua amplificação, ambos presentes em alta abundância nos genomas de mamíferos. E os retrotransposons LTR (do inglês *Long Terminal Repeats*), os quais possuem uma repetição terminal longa flanqueando o próprio elemento e representam uma fração considerável dos genomas das plantas (Feschotte; Jiang; Wessler, 2002).

Os LTR, por sua vez, agrupam-se em duas grandes superfamílias, que são a Ty1/copia e a Ty3/gypsy, sendo estes os principais componentes dos genomas das angiospermas, sendo presente em mais de 70% do genoma do milho por exemplo (Kumar & Bennetzen, 1999; Meyers et al., 2001, Macas et al., 2015, Lee et al, 2017). Os elementos das linhagens Ty1/copia e Ty3/gypsy são encontrados principalmente nas regiões heterocromáticas dos cromossomos, e muitos estudos vem mostrando

que tais elementos são sensíveis a fatores ambientais, podendo assim, levar a suas ampliações e/ou deleções, impactando diretamente na evolução e diversificação genômica de espécies intrinsecamente relacionadas (Cao et al. 2014; Matsunaga et al. 2015).

Os dados genômicos obtidos através de sequenciamentos de baixa cobertura, vem permitindo uma análise acurada dos elementos repetitivos que compõem os genomas das espécies, podendo serem utilizados para reconstruções filogenômicas e análises comparativas entre as espécies (Macas et al., 2015; Novák et al., 2020; Costa et al., 2021).

## 2.2 Sequenciamento genômico de Nova geração

Opções cada vez mais rápidas e menos onerosas de sequenciamentos estão se fazendo presentes em contraste as maneiras mais clássicas, como o sequenciamento Sanger por exemplo. Os sequenciamentos de Nova Geração ou também conhecido por NGS (*Next Generation Sequence*) tem a vantagem de conseguir realizar um número maior de leituras em um curto espaço de tempo, indo desde a obtenção de dados mais simples a mais complexos. Técnicas como *Genome skimming*, por exemplo, é uma das metodologias mais simples que consiste em um sequenciamento superficial do DNA genômico, na qual se obtêm uma considerável fração de alta cópia do genoma (desde plastoma, mitogenoma e DNA repetitivo). Esses dados podem ser utilizados em análises filogenômicas e também para a caracterização da fração repetitiva de DNA das espécies (Dodsworth 2015).

Por outro lado, sequenciamentos do tipo target-capture, visam uma subamostragem diferente dos genomas sequenciados, onde o foco são sequências específicas de baixa cópia do genoma nuclear (Andermann et al. 2020). Contudo, a fração off-target do sequenciamento pode ser reutilizada, onde pode ser aproveitada a fração repetitiva e utilizada para fins tais quais os de genome skimming (Costa et al. 2021).

## 2.3 Filogenômica

Os dados de sequenciamento NGS utilizados para análises *in silico* das frações repetitivas do genoma podem ser utilizados para inferências filogenômicas, sendo muito encontradas em trabalhos mais recentes que visam elucidar questões filogenéticas entre as espécies (Costa et al. 2020). Utilizando informações como a

abundância dos elementos repetitivos para a construção de árvores, foi possível perceber sinais filogenéticos significativos, tal como os gerados a partir de dados de regiões nucleares, plastidiais e ribossomais (Dodsworth et al. 2016). Análogo a isso, a similaridade dos *repeats* também pode apresentar sinais filogenômicos presente em alguns grupos de espécies quando se utilizam, por exemplo, abordagens como de inferências filogenéticas como o AAF (do inglês *Assembly and Alignment-Free*) que gera árvores a partir de dados de sequenciamentos de genomas não montados (Fan et al 2015). Atrelada a esse tipo de análise, dados de números cromossômicos se apresentam de forma complementar, uma vez que ajudam a endossar aspectos ligados a variação cromossômica e relações de parentesco.

#### 2.4 Poliploidia

O conjunto cromossômico de uma espécie é denominado cariótipo. Um dos principais caracteres utilizados para estudos relacionados a filogenia e a evolução cromossômica das espécies são a displóidia, que consiste em rearranjos cromossômicos estruturais gerando o aumento ou a perda do número cromossômico, a qual pode ser ascendente e descendente respectivamente que, geralmente não resulta em perda ou ganho do material genético; e a poliploidia que consiste na duplicação ou multiplicação de um conjunto cromossômico inteiro (Guerra 2008). Além disso, outras variações podem afetar nos números cromossômicos das espécies, como por exemplo, aneuploidias e os cromossomos B sendo assim rearranjos numérico gerando a perda ou ganho do material genético. O nível de ploidia varia geralmente durante a fase ontogênica das plantas, é utilizado a letra (n) para expressar a fase gamética e o (2n) para a fase esporofítica não especificando o nível de ploidia do organismo (Guerra 2008, 2012).

A duplicação inteira do genoma das plantas WGD (Whole Genome Duplication) chamada também de poliploidia, foi inicialmente dividida em quatro tipos pelos estudos do botânico Stabbins no final da década de 1940, considerando aspectos genéticos e citogenéticos. Atualmente, consideram-se apenas dois: os alopoliploides, que é fruto da hibridação entre duas espécies diferentes e os autoploiploides, evento que ocorre com a duplicação do conjunto genético dentro de uma só espécie (Ramsey e Schemske, 1998). Além disso, é considerada também um dos eventos mais importantes para a evolução e diversificação, principalmente em plantas com flores, uma vez que, seu surgimento provavelmente se deu após um evento antigo de WGD

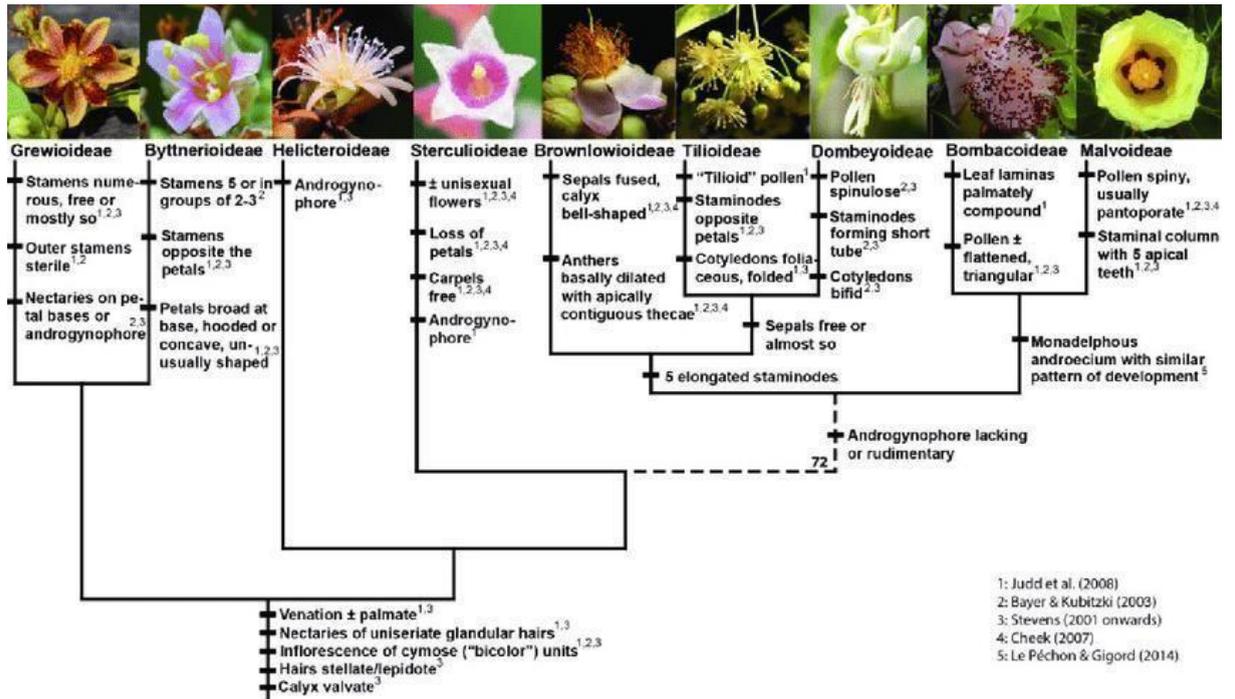
além de outras novidades evolutivas reveladas por dados de sequenciamento (Soltis et al., 2009; Jiao et al., 2011).

Conceitos introduzidos a esse tema, também se referem a origem ancestral recente ou antiga da poliploidia, a neopoliploidia e a paleopoliploidia respectivamente (Guerra 2008), que por sua vez se mostram importantes, uma vez que o cariótipo dos poliploides antigos passam por processos chamados de diploidização que, com a perda de cópias de genes duplicados, entre outras coisas, consiste na volta do comportamento diploide em organismos poliploides (Soltis et al; 2015).

## 2.5 Família Malvaceae

Malvaceae é uma família de angiospermas que agrupa cerca de 4,225 espécies, distribuídas entre 224 gêneros e nove subfamílias, sendo elas Bombacoideae, Brownlowioideae, Byttnerioideae, Dombeyoideae, Grewioideae, Helicteroideae, Malvoideae, Sterculioideae e Tilioideae (Bayer e Kubitzki). Esse grupo é muito bem distribuído sendo Pantropical e possuindo plantas economicamente importantes como o Cacao (*Theobroma cacao*), o Quiabo (*Abelmoschus esculentus*), Durian (*Durio zibethinus*), Cola (*Cola* spp.) e o Baoba (*Adansonia* spp.). As espécies dessa família possuem diferentes tipos de hábitos como árvores, arbustos e ervas, podendo ter folhas simples, lobadas palmadas, digitatas ou palmatinervias, por vezes apresentando nectário extrafloral com frutos deiscentes ou indeiscentes e flores unissexuais ou hermafroditas (Bayer e Kubitzki 2003).

Do ponto de vista citogenético, a subfamília Bombacoideae se destaca diferenciando-se das outras subfamílias por ter uma provável origem paleopoliploide (Marinho et al 2014). Trabalhos recentes trazem informações de novas contagens cromossômicas demonstrando um número alto de cromossomos na maioria das espécies do grupo, evidenciando uma suposta poliploidia nesse grupo. Além disso, as espécies do grupo também demonstraram comportamento diploide apresentando pares de sítios de DNAr e pares de bandas heterocromáticas sugerindo a diploidização em Bombacoideae (Figueredo et al 2016; Costa et al 2017).



Fonte: Strijk et al. 2021.

### 3. MATERIAIS E MÉTODOS

#### 3.1 Aquisição de dados ngs e caracterização da fração repetitiva

Os dados foram aproveitados de sequenciamentos de genome skimming e target-capture encontradas disponíveis no banco de dados NCBI e também cedida por nossos colaboradores.<sup>24</sup> espécies do clado Malvatheca, das quais 11 representaram a subfamília Malvoideae e 13 a subfamília Bombacoideae foram analisadas, filtradas e tratadas por qualidade através da plataforma RepeatExplorer2, (Novak et al., 2013) e foi realizada uma análise de agrupamento de reads baseada em similaridade de sequência. Foram feitas análises individuais para cada espécie e uma comparativa utilizando todo o conjunto de dados. A análise comparativa foi realizada com um total de 1.091.952 reads e a cobertura genômica de cada uma das espécies foi calculada pela fórmula  $(r \times l)/g$ , onde  $r$  foi o número de leituras utilizadas na análise;  $l$  corresponde ao comprimento das leituras e  $g$  é o tamanho do genoma haploide em pb. Foi feito um cálculo da cobertura considerando um tamanho aproximado de acordo com a espécie filogeneticamente mais próxima e com número cromossômico mais próximo já que não tínhamos o tamanho do genoma para todas as espécies.

### 3.2 Diversidade dos repeats e correlação com o número cromossômico

Calculamos a diversidade de linhagens de repeats nos genomas utilizando o Índice de diversidade de Shannon (H, Shannon, 1948) e o Índice de diversidade de Simpson (D, Simpson, 1949). Desta forma mostrando a probabilidade de que dois indivíduos (neste caso cópias de repeats) retirados aleatoriamente de uma comunidade (neste caso um genoma) pertençam a mesma espécie (neste caso linhagem de repeat). Para isso utilizamos as abundâncias de cada linhagem de repeats encontradas pela análise comparativa do RepeatExplorer.

Para investigar se a relação entre a diversidade de repeats em um dado genoma estava relacionada com a poliploidia, realizamos uma análise de correlação de Pearson entre os índices de Shannon e de Simpson com o número cromossômico das espécies de *Malvatheca* utilizando o pacote 'stats' implementado no RStudio (R Core Team, 2019).

### 3.3 Relações filogenéticas baseadas em plastomas completos

As montagens dos plastomas das espécies analisadas foram montadas contrarreferência da espécie *Ceiba speciosa*, obtida através do banco de dados do NCBI (acesso NC\_057077). Para isso foi mapeados os reads de cada espécie contra o genoma referência utilizando a função "Map to reference" do software Geneious v.6.0.3 (Kearse et al. 2012). Os alinhamentos foram realizados utilizando o MAFFT (Kato and Standley, 2013). Para simplificação, foi utilizado o modelo mais geral de substituição de DNA GTR + I + G (Abadi et al., 2019). Foram inferidas as relações filogenéticas utilizando Máxima Verossimilhança (ML) pelo FastTree (Price et al. 2010). A visualização e a edição das árvores foram realizadas pelo FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>).

### 3.4 Relações filogenéticas baseadas em DNA repetitivo

O método de inferência filogenética empregado foi o baseado em abundância de repeats que permite fazer a comparação entre a abundância de diferentes classes de elementos repetitivos entre diferentes espécies que, possuem sinal filogenéticos quando tratadas como caracteres contínuos. (Dodsworth et al., 2015). Primeiramente, foi concatenada as leituras para as 24 espécies com cobertura de aprox. 0,065 x, com códigos de quatro caracteres para cada espécie. Foi executada uma análise de agrupamento comparativa (agrupamento simultâneo de todas as espécies no conjunto

de dados) no Repeat Explorer com configurações padrão sobre os dados (Novak et al. 2013).

Já para a análise filogenética baseada em similaridade de sequência, foi usada a abordagem de alinhamento e montagem livre (AAF) (Fan et al., 2015) usando todas as leituras identificadas como repetições pelo RepeatExplorer. Essa abordagem gera filogenias a partir de dados de sequência de genoma não montados, independente da montagem ou do alinhamento do genoma. Dessa forma é calculada as propriedades estatísticas das distâncias pareadas entre genomas, otimizando a seleção de parâmetros e realizar bootstrapping.

### 3.5 Filogenia de domínios conservados de elementos transponíveis

Foram extraídos o domínio proteico da transcriptase reversa dos três elementos mais abundantes da superfamília Ty1/copia (ALE, Ikeros e SIRE) e dos três elementos mais abundantes da superfamília Ty3/gypsy (CRM, Ogre e Tekay) encontrados nas espécies utilizando a ferramenta DANTE (Domain-based Annotation of Transposable Elements) com configurações padrões (Novák et al. 2013), com o intuito de investigar a homologia e a evolução dinâmica dos retrotransposons LTR. O DANTE realiza anotações e a classificação dos domínios protéicos homólogos aos dados disponíveis no banco de dados de domínios proteicos Viridiplantae (Neumann et al. 2019). Os domínios encontrados foram alinhados usando MAFFT (Katoh and Standley, 2013). Foram criadas filogenias para cada um dos elementos com base nos alinhamentos realizados por meio do plugin FastTree implementado no Geneious v.6.0.3 (Kearse et al. 2012)

## 4 RESULTADO

### 4.1 Artigo: Declínio dos dinossauros e sucesso dos poliploides: consequências genômicas da paleopoliploidia na subfamília *Bombacoideae* (*Malvaceae*) no final do Cretáceo

Gustavo Luna<sup>1</sup>, Flavia Fonseca Pezzini<sup>2</sup>, Nisa Karimi<sup>3</sup>, Yennifer Mata-Sucre<sup>1</sup>, Joeri Sergej Strijk<sup>4</sup>, Jefferson Guedes Carvalho-Sobrinho<sup>5</sup>, Gustavo Souza<sup>1#</sup>, Lucas Costa<sup>1</sup>

<sup>1</sup> Laboratory of Plant Cytogenetics and Evolution, Department of Botany, Federal University of Pernambuco, Recife, Pernambuco, Brazil;

<sup>2</sup> Royal Botanic Garden Edinburgh, Edinburgh, United Kingdom;

<sup>3</sup> Department of Botany, University of Wisconsin – Madison, 430 Lincoln Drive, Madison, WI 53706, USA;

<sup>4</sup> Institute for Biodiversity and Environmental Research, Universiti Brunei Darussalam;

<sup>5</sup> Universidade Federal do Vale do São Francisco, Petrolina, Brazil, Universidade Federal Rural de Pernambuco, Recife, Brazil.

#Corresponding authors: luiz.rodriguessouza@ufpe.br;

## Resumo

A poliploidia (também conhecida como Whole Genome Duplication - WGD) e a amplificação/ desamplificação de DNAs repetitivos (repeats) são os principais mecanismos responsáveis pela evolução do genoma em plantas. Muitas linhagens de angiospermas têm evidências de antigas WGD, na maioria estimadas em 66 milhões de anos, sugerindo um sucesso dos poliploides após as mudanças globais que levaram à extinção dos dinossauros no final do Cretáceo. Nesse sentido, pouco é conhecido sobre os mecanismos de evolução das frações repetitivas dos genomas em linhagens que passaram por eventos antigos de poliploidia (paleopoliploidia). A subfamília Bombacoideae (Malvaceae) representa um modelo interessante para estudar tal relação, pois apresenta números cromossômicos altos ( $2n = 86$  a  $2n = 276$ ) e uma provável origem paleopoliploide. Por outro lado, a subfamília Malvoideae apresenta números cromossômicos mais baixos e eventos pontuais de neopoliploide. Aqui, caracterizamos a fração repetitiva genômica de 14 espécies de Bombacoideae e 11 Malvoideae e inferimos relações filogenéticas com base em três conjuntos de dados: plastomas, abundância/ similaridade dos repeats e domínios específicos de proteínas de elementos transponíveis. A caracterização das frações repetitivas dos representantes do clado Malvatheca revelaram uma alta diversidade de repeats com predomínio de LTRs Ty3/gypsy, sem tendências diferenciais entre Bombacoideae e Malvoideae. No geral, domínios proteicos de diferentes elementos transponíveis não revelaram evidências de impacto da paleopoliploidia na sua diversificação. As filogenias baseadas na abundância e similaridade de repeats não apresentaram topologias congruentes com a filogenia de plastoma, sugerindo que esse método pode falhar com grupos antigos. Por outro lado, para a filogenia baseada em abundância de repeats, os tamanhos dos ramos foram mais longos em Malvoideae e mais curtos em Bombacoideae, evidenciando diferentes taxas evolução dos repeats nessas subfamílias. Por fim, foi encontrada uma correlação significativa entre número cromossômico e diversidade de repeats, denotando um possível efeito da poliploidia sobre a variedade de linhagens de elementos repetitivos. Desta forma, nossos

resultados mostram que a paleopoliploidia provavelmente exerce um efeito mais forte sobre a diversidade de elementos repetitivos no genoma do que sobre a variação quantitativa/ qualitativa dos repeats. Este estudo fornece insights sobre a evolução da fração repetitiva do genoma em grupos paleopoliploides, o que é desafiador dada a idade antiga dessas WGD.

**Palavras-Chave:** Bombacoideae, Malvoideae, Paleopoliploidia, DNA repetitivo

## INTRODUÇÃO

Um dos eventos mais importantes relacionado a evolução das plantas é a poliploidia (também conhecida como Whole Genome Duplication - WGD), que se trata da presença de dois ou mais genomas completos em uma espécie, o qual afeta não só o genoma, mas também interfere na especiação e diversificação dos organismos (Soltis et al., 2009; Weiss-Schneeweiss; Schneeweiss, 2013). Especialmente nas angiospermas, há evidências de WGDs recorrentes, e o ancestral de todas as plantas com flores foi provavelmente originado de uma linhagem poliploide (Jiao et al. 2011). Muitos WGDs de plantas parecem associados a períodos de aumento do estresse ambiental e/ou flutuações, uma tendência que é evidente tanto para os poliploides atuais (neopoliploides) quanto para poliploides antigos (paleopoliploides), formados independentemente em diversas famílias de angiospermas no período da extinção do Cretáceo-Paleogeno (aprox. 66 Ma) (Vanneste et al. 2014; Marinho et al. 2014). Curiosamente, o evento que extinguiu os dinossauros parece ter atuado como um filtro ambiental que selecionou positivamente os poliploides em plantas (Vanneste et al. 2014).

Existem basicamente dois tipos de poliploidia: (i) autopoliploidia, que se refere à multiplicação de conjuntos de cromossomos geneticamente idênticos dentro de uma única espécie e, (ii) alopoliploidia, que envolve a hibridização entre duas espécies geneticamente diferentes (Ramsey e Schemske, 1998, 2002, Husband et al., 2013). A poliploidia tem sido considerada um dos processos mais importantes para o sucesso da evolução e diversidade das espécies vegetais por estar associada a uma alta taxa de especiação (Wood et al, 2009). Por conseguinte, a duplicação de genes e as mudanças no espaço gênico consequentes da poliploidia podem ocasionar o surgimento de novidades evolutivas, como por exemplo, a aparição das flores, a qual é atribuída a uma duplicação total do genoma na base de diversificação das angiospermas (Soltis et al 2015; Dodsworth; Chase; Leitch, 2016). Após a divergência

do genoma poliploide, acontece um processo conhecido como diploidização, que inclui a perda de genes duplicados, diminuição do número cromossômico (displóidia descendente) e diminuição do tamanho do genoma, fazendo com que o organismo volte a um estado diploide (Soltis et al. 2015).

O genoma dos eucariotos, além da presença de sequências codificantes de cópia única (genes), é também composto, em sua maior parte, por sequências não-codificantes altamente repetitivas especialmente em plantas (Elliott; Gregory, 2015). Este DNA repetitivo é dividido em duas grandes categorias: os DNAs repetitivos em tandem e os DNAs repetitivos dispersos (Mehrotra e Goyal, 2014). Os DNA repetitivos em tandem são elementos repetitivos em grandes blocos adjacentes, destacando-se os chamados DNA satélite (Garrido-Ramos, 2017). Os elementos dispersos estão dispostos de forma dispersa ao longo de todo o genoma (Weiss-Schneeweiss; Schneeweiss, 2013). Dentre os DNAs repetitivos dispersos, os mais estudados são os elementos transponíveis, que possuem a capacidade de se “mover” ao longo do genoma hospedeiro (Arkhipova, 2017). Estes são classificados de acordo com o mecanismo utilizado para sua transposição, sendo divididos em elementos Classe I retrotransposons e os de Classe II, transposons de DNA (Finnegan, 1989). Dentre os retrotransposons, os elementos que mais se destacam são os do tipo LTR (Long Terminal Repeat) que podem ser divididos em duas superfamílias, Ty1/copia e Ty3/gypsy, encontrados principalmente na heterocromatina das angiospermas e responsáveis por compor elementos como o centrômero, e contribuir significativamente para a variação no tamanho do genoma (Marques et al. 2015; Macas et al. 2015).

A identificação da fração repetitiva pode ser feita a partir de um sequenciamento do genoma de baixa cobertura, também conhecido como genome skimming (Dodsworth et al. 2015). Atualmente, a plataforma mais utilizada para a análise do DNA repetitivo é o RepeatExplorer, que utiliza um algoritmo de agrupamento baseado em gráficos para agrupar sequências de alta cópia com base na similaridade, com posterior identificação de domínios codificadores de proteínas por referência cruzada com um extenso conjunto de bases de dados atualizadas (Novak et al., 2013). Recentemente, foi visto que, além do tradicional uso de dados de baixa cobertura, dados obtidos por sequenciamentos como target-capture (no qual várias sondas genômicas são projetadas para “capturar” e enriquecer regiões codificantes específicas de baixa cópia do genoma nuclear) podem ser utilizadas para

esse tipo de estudo, por meio da reutilização da fração off-target do sequenciamento (Costa et al. 2021). Outro importante avanço no estudo do DNA repetitivo foi a criação de métodos de inferência filogenômica baseados tanto na abundância (Fan et al. 2015) como na similaridade (Dodsworth et al. 2015) de sequências repetitivas (abordagens repeat-based). Além do potencial informação taxonômica, árvores baseadas em DNA repetitivo podem trazer insights sobre a evolução das diferentes linhagens de DNA repetitivo, o que pode ser potencialmente informativo em linhagens paleopoliploides.

Na família Malvaceae, a subfamília Bombacoideae se destaca por ter uma origem paleopoliplóide entre 62–56 Ma, (Marinho et al. 2014; Costa et al. 2017). Essa família pertence a ordem Malvales e possui cerca de mais de 4.225 espécies aceitas distribuídas em 244 gêneros (Stevens 2001; Bayer e Kubitzki 2003; Christenhusz e Byng 2016). As relações filogenéticas dentro dessa família são bastante controversas por ter uma grande quantidade de espécies e pela dificuldade em interpretar traços morfológicos que, de acordo com dados moleculares parecem ser homoplásicos (Alverson et al. 1999; Le Péchon and Gigord 2014). Malvaceae é dividida em 9 subfamílias (Bombacoideae, Brownlowioideae, /Byttnerioideae, Dombeyoideae, Grewioideae, Helicteroideae e Malvoideae) (Bayer et al. 1999). As subfamílias Malvoideae (~1.800 spp) e Bombacoideae (160 spp) são comumente agrupadas em clado informal conhecido como Malvatheca. Essas subfamílias apresentam distribuição pantropical e são muito importantes economicamente na agricultura, silvicultura e horticultura, tendo como alguns dos principais representantes o algodão (*Gossypium* spp.), o quiabo (*Abelmoschus esculentus*), o baobá (*Adansonia* spp.) e Hibiscus (Bayer e Kubitzki 2003).

Do ponto de vista citogenético, a subfamília Bombacoideae se diferencia das outras subfamiliais por apresentar números cromossômicos altos ( $2n= 86$  a  $2n = 276$ ) (Baum e Oginuma 1994; Costa et al 2017; Figueredo et al 2016; Marinho et al. 2014). Revisões citogenéticas recentes de gêneros da subfamília Bombacoideae, indicam estabilidade numérica em nível genérico sugerindo que a variabilidade de números cromossômicos é menor do que reportada na literatura (Baum e Oginuma 1994; Costa et al 2017; Figueiredo et al 2016). Recentes estudos mostram que apesar do elevado número cromossômico, as espécies possuem poucos pares de bandas heterocromáticas, assim como poucos sítios de DNA ribossomal, sugerindo que sejam genomas altamente diploidizados (Figueredo et al 2016; Costa et al 2017).

O objetivo do estudo foi testar o impacto da paleopoliploidia na fração repetitiva dos genomas de representantes da subfamília Bombacoideae. Para isso, foi realizada a caracterização da fração repetitiva de 13 espécies de Bombacoideae e 11 de Malvoideae e inferidas topologias filogenéticas com base em três conjuntos de dados: abundância de repeats, similaridade de repeats e domínios proteicos específicos de elementos transponíveis. Além disso, foi realizada uma filogenia a partir dos plastomas completos a fim de realizar a comparação das diferentes topologias das árvores no estudo. A partir desse conjunto de dados, visamos responder às seguintes perguntas: 1) A origem paleopoliploide gerou diferentes tendências evolutivas dos DNAs repetitivos no clado Bombacoideae? 2) É possível inferir filogenias repeat-based (Dodsworth et al. 2015) para o clado Malvatheca?

## MATERIAIS E MÉTODOS

### *Aquisição de dados NGS e caracterização da fração repetitiva*

No presente estudo, foram utilizados dados aproveitados de sequenciamento target-capture e genome skimming disponíveis no banco de dados NCBI e cedida por nossos colaboradores. Foram analisadas 24 espécies pertencentes ao clado Malvatheca, das quais 11 representaram a subfamília Malvoideae e 13 a subfamília Bombacoideae. A lista de todas as espécies utilizadas, bem como os códigos de acesso ao NCBI se encontram listados na Tabela 1.

Tabela 1. Lista de espécies e dados citológicos utilizados nesse estudo.

<b>Bombacoideae</b>	<b>n</b>	<b>1C</b>	<b>Acesso</b>	<b>Sequenciamento</b>
<i>Adansonia digitata</i> L.	80	1.67	<a href="#">SRX7064907</a>	HybSeq
<i>Adansonia grandidieri</i> Baill.	44	-	<a href="#">SRX7064909</a>	HybSeq
<i>Adansonia gregorii</i> F.Muell.	44	-	<a href="#">SRX7064910</a>	HybSeq
<i>Adansonia madagascariensis</i> Baill.	44	-	<a href="#">SRX7064914</a>	HybSeq
<i>Adansonia perrieri</i> Capuron	44	-	<a href="#">SRX7064913</a>	HybSeq
<i>Adansonia rubrostipa</i> Jum. & H.Perrier	44	-	<a href="#">SRX7064901</a>	HybSeq
<i>Adansonia suarezensis</i> H.Perrier	44	-	<a href="#">SRX7064911</a>	HybSeq
<i>Adansonia za</i> Baill.	44	-	<a href="#">SRX7064903</a>	HybSeq
<i>Bombax ceiba</i> L.	46	1.62	<a href="#">SRX7064904</a>	HybSeq
<i>Bombax ceiba</i> L.	46	1.62	Nisa	HybSeq
<i>Pachira aquatica</i> Aubl.	44	2.28	Nisa	HybSeq
<i>Pseudobombax croizatii</i> A.Robyns	-	-	<a href="#">SRX7064905</a>	HybSeq
<i>Pseudobombax croizatii</i> A.Robyns	-	-	Nisa	HybSeq

<i>Scleronema micranthum</i> (Ducke) Ducke	-	-	<a href="#">SRX7064906</a>	HybSeq
<i>Scleronema micranthum</i> (Ducke) Ducke	-	-	Nisa	HybSeq
<i>Ceiba pentandra</i> (L.) Gaertn.	46	1.75	Flávia	HybSeq
<b>Malvoideae</b>	<b>n</b>	<b>1C</b>	<b>Acesso</b>	<b>Sequenciamento</b>
<i>Abelmoschus esculentus</i> (L.) Moench	60	1.70	SRX2990989	HiSeq
<i>Abutilon theophrasti</i> Medik.	21	1.40	SRX9130922	HiSeq X Ten
<i>Althaea officinalis</i> L.	21	-	SRX2855182	MiSeq
<i>Gossypium arboreum</i> L.	13	1.72	SRX5759203	HiSeq X Ten
<i>Gossypium raimondii</i> Ulbr.	13	0.90	SRX341462	HiSeq
<i>Hibiscus syriacus</i> L.	40	-	SRX529352	HiSeq
<i>Kokia drynarioides</i> Lewton	12	0.60	SRX3304973	HiSeq
<i>Malva sylvestris</i> L.	21	1.50	ERX3149988	HiSeq
<i>Malva moschata</i> L.	21	1.36	ERX3149982	HiSeq
<i>Malva pusilla</i> Sm.	21	-	SRX7432590	HiSeq
<i>Thespesia populneoides</i> (Roxb.) Kostel.	26	4.10	SRX706235	HiSeq

As sequências de todas as espécies analisadas foram filtradas e tratadas por qualidade. Através do pipeline RepeatExplorer2, implementado no servidor Elixir-Cerit (<https://repeatexplorer-elixir.cerit-sc.cz>) (Novak et al., 2013), foi realizada uma análise de agrupamento de reads baseada em similaridade de sequência. Dois tipos de análises foram realizadas, uma individual (onde cada espécie foi analisada separadamente a fim de caracterizar a fração repetitiva do genoma de forma mais detalhada) e uma análise comparativa (utilizando os dados concatenados das sequências das espécies, com o intuito de comparar a porção repetitiva de todas elas). A análise comparativa foi realizada com um total de 1.091.952 reads e a cobertura genômica de cada uma das espécies foi calculada pela fórmula  $(r \times l)/g$ , onde r foi o número de leituras utilizadas na análise; l corresponde ao comprimento das leituras e g é o tamanho do genoma haploide em pb. Como não possuíamos informações de tamanho do genoma para todas as espécies, para o cálculo da cobertura foi considerado um tamanho aproximado de acordo com a espécie filogeneticamente mais próxima e com número cromossômico mais próximo. O total de leituras analisadas por espécie tanto para a análise individual quanto para a comparativa para cada uma delas se encontra detalhado na Tabela 1.

### *Diversidade dos repeats e correlação com o número cromossômico*

Tem sido proposto que métricas de diversidade utilizadas na ecologia de comunidades podem ser aplicadas de forma análoga em análises genômicas, onde o genoma representa uma comunidade e as diferentes linhagens de elementos repetitivos representam diferentes espécies (Schley et al. 2022). Desta forma, buscamos calcular a diversidade de linhagens de repeats nos genomas analisados utilizando duas métricas: o Índice de diversidade de Shannon (H, Shannon, 1948) e o Índice de diversidade de Simpson (D, Simpson, 1949). Estas métricas informam a probabilidade de que dois indivíduos (neste caso cópias de repeats) retirados aleatoriamente de uma comunidade (neste caso um genoma) pertençam a mesma espécie (neste caso linhagem de repeat). Quanto maior o valor dos índices, menor essa probabilidade, o que indica uma alta diversidade para a comunidade (genoma) analisado. Para esta análise, utilizamos as abundâncias de cada linhagem de repeats encontradas na análise comparativa do RepeatExplorer.

Para investigar se a diversidade de repeats em um dado genoma estava relacionada com a poliploidia, fizemos também uma análise de correlação de Pearson entre os índices de Shannon e de Simpson com o número cromossômico das espécies de *Malvatheca*. Para isso, utilizamos o pacote 'stats' implementado no RStudio (R Core Team, 2019).

### *Relações filogenéticas baseadas em plastomas completos*

Os plastomas referentes às espécies analisadas no trabalho foram montadas contra referência de *Ceiba speciosa*, obtida no banco de dados do NCBI (acesso NC\_057077). Para isso mapeamos os reads de cada espécie contra o genoma referência usando a função "Map to reference" do software Geneious v.6.0.3 (Kearse et al. 2012). Os alinhamentos dos plastomas foram feitos usando MAFFT (Kato and Standley, 2013). Para simplificação, usamos o modelo mais geral de substituição de DNA GTR + I + G (Abadi et al., 2019). As relações filogenéticas foram inferidas usando Máxima Verossimilhança (ML) usando FastTree (Price et al. 2010). As árvores foram visualizadas e editadas no FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>).

### *Relações filogenéticas baseadas em DNA repetitivo*

Foi empregado um método de inferência filogenética baseado em abundância de repeats que permite comparar a abundância de diferentes classes de elementos

repetitivos entre espécies diferentes que, tratadas como caracteres contínuos, possuem sinal filogenético (Dodsworth et al., 2015). Primeiro, concatenamos as leituras para nossas 24 espécies com cobertura de aprox. 0,065 x, com códigos de quatro caracteres para cada espécie. Sobre esses dados, foi executada uma análise de agrupamento comparativa (agrupamento simultâneo de todas as espécies no conjunto de dados) no Repeat Explorer com configurações padrão (Novak et al. 2013). Como as sequências foram codificadas com os nomes das espécies, pudemos identificar o número de leituras que cada espécie contribuiu para cada cluster gerado, que é proporcional à abundância de cada repetição no genoma de cada espécie.

Para a análise filogenética baseada na similaridade de sequência, foi usada a abordagem de alinhamento e montagem livre (AAF) (Fan et al., 2015) usando todas as leituras identificadas como repetições pelo RepeatExplorer. A AAF gera filogenias a partir de dados de sequência de genoma não montados, independente tanto da montagem quanto do alinhamento do genoma. Assim, ele calcula as propriedades estatísticas das distâncias pareadas entre genomas, permitindo otimizar a seleção de parâmetros e realizar bootstrapping.

#### *Filogenia de domínios conservados de elementos transponíveis*

O domínio proteico da Transcriptase Reversa dos três elementos mais abundantes da superfamília Ty1/copia (ALE, Ikeros e SIRE) e dos três elementos mais abundantes da superfamília Ty3/gypsy (CRM, Ogre e Tekay) encontrados nas espécies analisadas foram extraídos usando a ferramenta DANTE (Domain-based Annotation of Transposable Elements) com configurações padrões (Novák et al. 2013), a fim de explorar a homologia e entender a evolução dinâmica dos retrotransposons LTR. Esta ferramenta faz a anotação e a classificação dos domínios protéicos homólogos aos dados disponíveis no banco de dados de domínios proteicos Viridiplantae (Neumann et al. 2019). Todos os domínios encontrados para cada elemento foram alinhados usando MAFFT (Kato and Standley, 2013). Com base nesses alinhamentos, filogenias para cada um dos elementos foram criadas por meio do plugin FastTree implementado no Geneious v.6.0.3 (Kearse et al. 2012).

## **RESULTADOS**

### *Caracterização da fração repetitiva*

Para a análise individual do RepeatExplorer, o número de reads analisados variou de 84.423 para a espécie *Abutilon theophrasti* até 1.794.593 para a espécie *Hibiscus syriacus*. Foi observada uma variação na abundância total de DNA repetitivo de 19,9% (*Adansonia perrieri*) a 68,7% (*Ceiba pentandra*) entre as espécies da subfamília Bombacoideae e de 21,4% (*Malva pusilla*) a 59,4% (*Thespesia populneoides*) na subfamília Malvoideae (Figura 1).

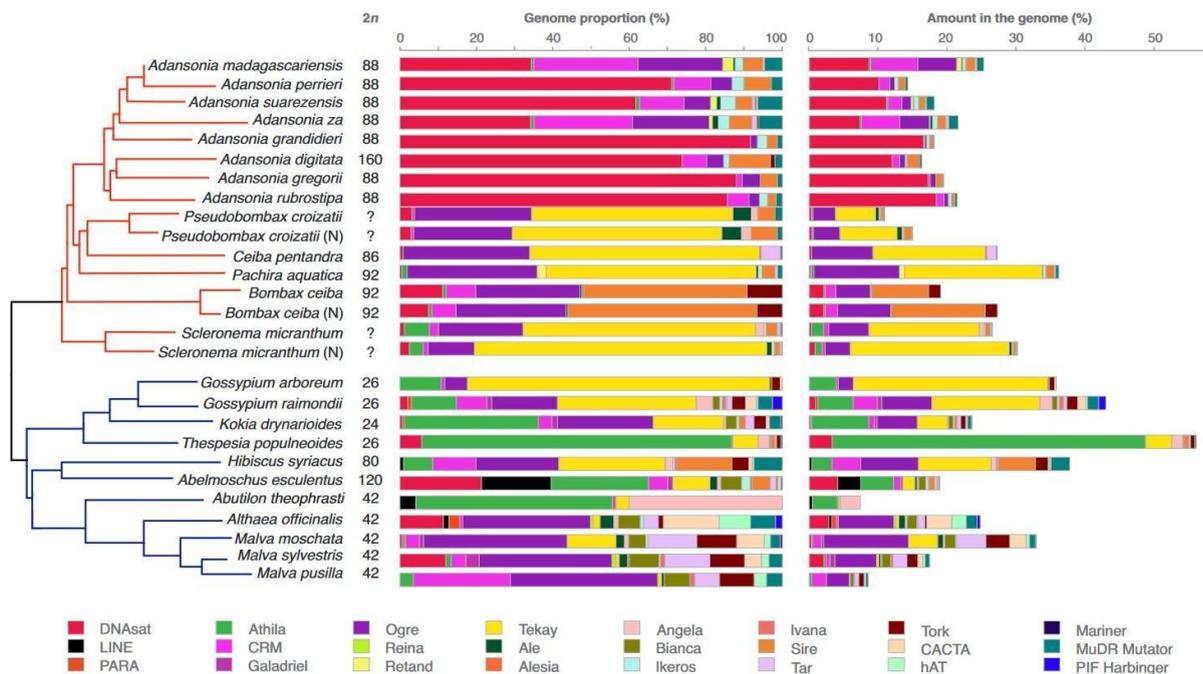


Figura 1. Composição da fração repetitiva de espécies das subfamílias Bombacoideae (clado vermelho) e Malvoideae (clado azul).

Com relação aos elementos transponíveis, as duas subfamílias apresentaram uma maior porcentagem de elementos de Classe I do que Classe II, com o destaque para a superfamília de retrotransposons LTR Ty3/gypsy que variou de 1,1% em *Adansonia gregorii* a 48,5% em *Thespesia populneoides*, com predominância das famílias de elementos Ogre, Tekay e Athila para ambas as subfamílias. Já em relação aos retrotransposons LTR do tipo Ty1/copia, as proporções variaram de 0,8% em *Adansonia gregorii* a 15% em um dos acessos de *Bombax ceiba*. Os elementos mais predominantes dentre os Ty1/copia foram SIRE, Ikeros, e Tork para ambas as subfamílias (Figura 1; Tabela 1). O DNA satélite variou de 0,02% em *Gossypium arboreum* a 18% em *Adansonia rubrostipa*. De forma geral, a subfamília Bombacoideae apresentou uma maior abundância de DNA satélite, especialmente dentre as espécies do gênero *Adansonia* (7,2% em *Adansonia za* até 18% em

*Adansonia rubrostipa*). Não foi observado nenhuma sinapomorfia genômica (presença/ ausência ou abundância de repeats) que distinga claramente as subfamílias Bombacoideae e Malvoideae (Figura 1).

### Análise comparativa das frações repetitivas

Para investigar o compartilhamento de clusters de sequências repetitivas, foi também realizada uma análise comparativa pelo RepeatExplorer. Na subfamília Bombacoideae pode-se observar um alto compartilhamento de diferentes clusters entre as espécies do gênero *Adansonia*, que apresentaram composição bastante similar. Interessantemente, *Pseudobombax croizatii* e *Pachira aquatica*, espécies de gêneros distintos dentro de Bombacoideae, apresentaram alto compartilhamento de clusters identificados como pertencentes às linhagens OGRE e CRM (Figura 2). No geral foi observado um grau mais alto de compartilhamento de repeats na subfamília Bombacoideae no que na subfamília Malvoideae (Figura 2).

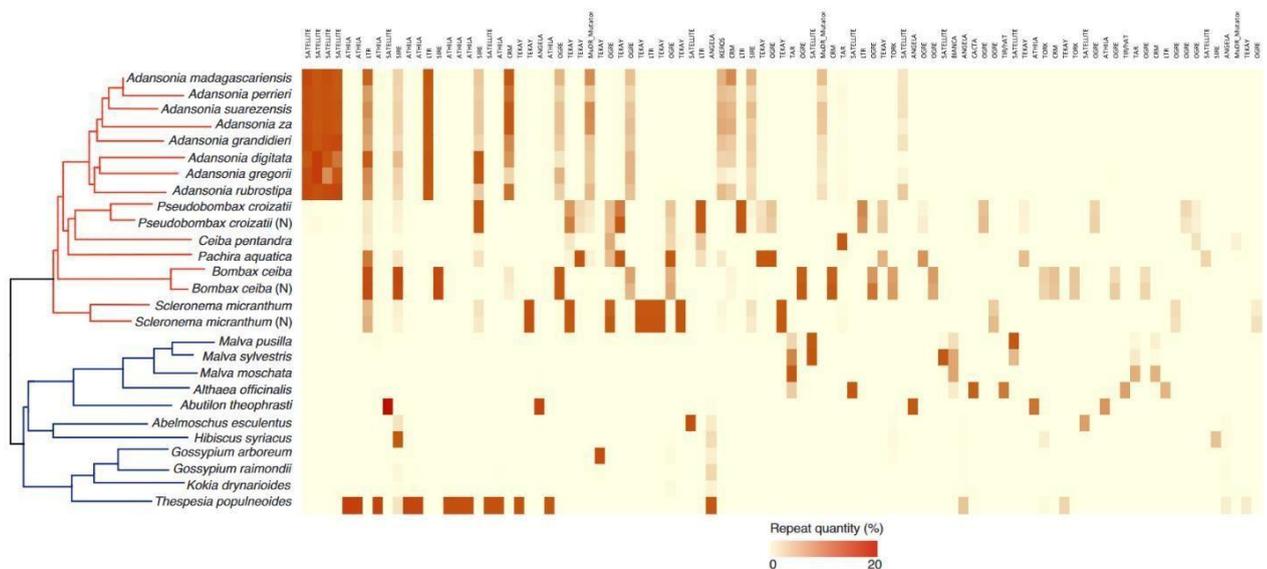


Figura 2. Análise comparativa dos elementos repetitivos de espécies das subfamílias Bombacoideae (vermelho) e Malvoideae (azul).

### Diversidade dos repeats e correlação com o número cromossômico

Os índices de diversidade de Shannon e Simpson para linhagens de repeats em *Malvaceae* variaram de  $H = 0,21$  e  $D = 0,08$  em *G. arboreum* até  $H = 1,59$  e  $D = 0,78$  em *A. rubrostipa*. No geral, espécies de Malvoideae apresentaram valores menores para ambos os índices ( $H = 0.21$  e  $D = 0.08$  em *G. arboreum* até  $H = 1.12$  e  $D = 0.63$  em *M. pusilla*), enquanto representantes de Bombacoideae apresentaram

valores maiores ( $H = 0.87$  e  $D = 0.51$  em *Pa. aquatica* até  $H = 1,59$  e  $D = 0,78$  em *A. rubrostipa*). Também encontramos correlações significativas entre os valores de ambos os índices de diversidade e número cromossômico em Malvatheca ( $R = 0.55$ ,  $p = 0.0036$  para Shannon e  $R = 0.55$  e  $p = 0.0034$  para Simpson), sugerindo uma possível relação entre poliploidia e incremento de diversidade dos repeats no clado Bombacoideae (Figura 3).

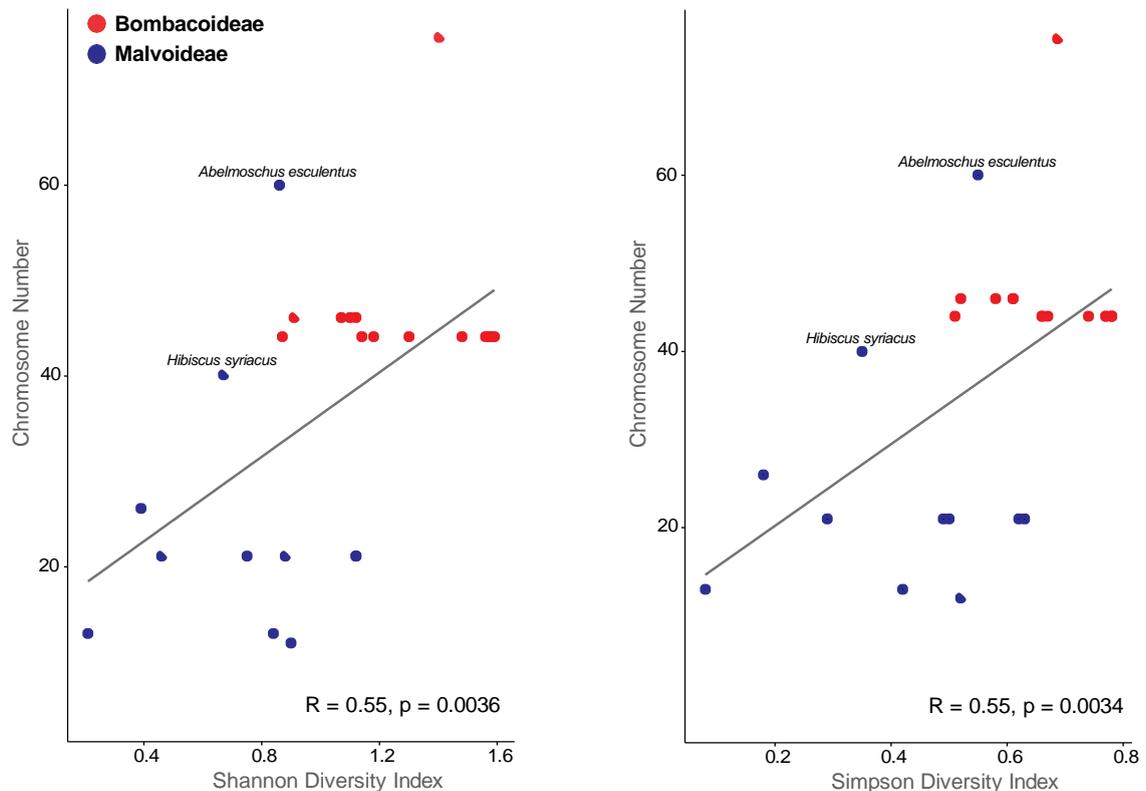


Figura 3. Índice de diversidade dos *repeats* (Shannon e Simpson) em representantes do clado Malvatheca.

#### *Filogenia de domínios proteicos dos elementos filogeneticamente abrangentes no clado Malvatheca*

Foram identificados seis elementos, três Ty3/gypsy (Ogre, CRM e Tekay) e três Ty3/copia (Sire, Ikeros e ALE) como abundância relativamente alta e compartilhada em todas as espécies analisadas. Assim, nós investigamos a homologia entre esses elementos transponíveis, construindo topologias baseadas nas sequências de domínios proteicos conservados identificados pela ferramenta DANTE do RepeatExplorer. Foi possível observar uma alta estruturação separando os domínios da transcriptase reversa das subfamílias Bombacoideae e Malvoideae. A exceção foi

a topologia dos elementos LTR Ty3/gypsy/CRM que apresentaram clados agrupando juntos elementos das duas subfamílias. No geral foi observado um número relativamente equilibrado de elementos para cada uma das subfamílias, exceto para a árvore ALE e Ikeros (com maior diversidade de elementos da subfamília Malvoideae) e SIRE (com maior diversidade de elementos da subfamília Bombacoideae).

Essa estruturação pode ser vista principalmente nas árvores baseadas nos domínios de elementos do tipo Tekay, SIRE, Ikeros e ALE. Em contraste, apesar de ainda apresentarem um certo grau de estruturação entre as subfamílias, as árvores dos elementos Ogre e CRM possuem um número maior de clados onde se observam domínios referente a espécies das duas subfamílias aninhados (Figura 4).

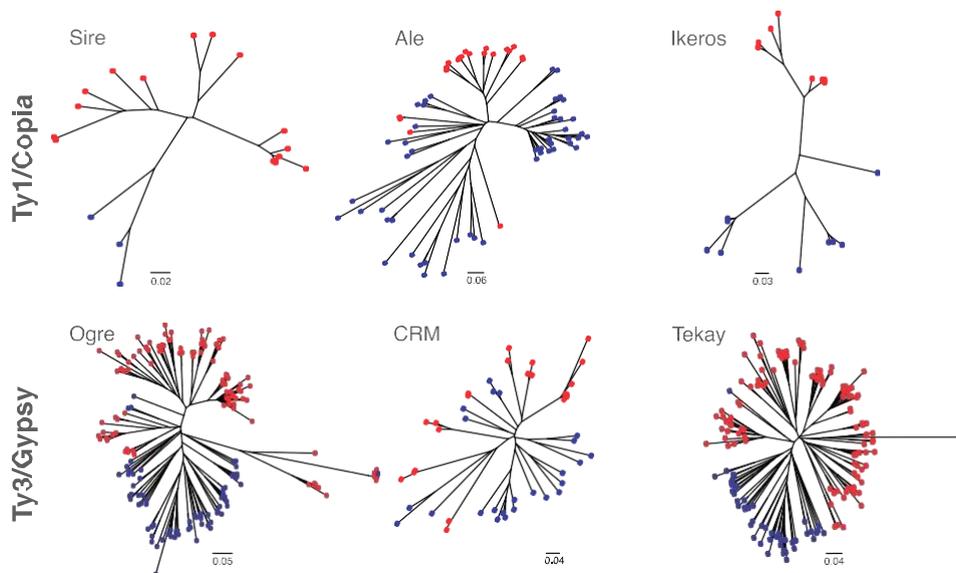


Figura 4. Árvores geradas a partir dos domínios proteicos dos elementos repetitivos encontrados em representantes da subfamília Bombacoideae (vermelho) e da subfamília Malvoideae (azul).

#### *Relações filogenéticas baseadas em DNA repetitivo*

As relações filogenéticas utilizando os plastomas completos das espécies analisadas apresentou nós em geral bem suportados, entre 0.98 e 1 pp (probabilidade posterior), com a exceção do nó que originou *Adansonia rubrostipa* (Figura 5a e 5b). Em ambas as subfamílias, espécies e/ou acessos pertencentes ao mesmo gênero formaram clados monofiléticos, como visto em *Adansonia*, *Pseudobombax*, *Bombax*, *Malva* e *Gossypium*. Na subfamília Bombacoideae, os gêneros *Pseudobombax*, *Ceiba* e *Pachira* formaram uma linhagem monofilética com suporte máximo diretamente irmã do clado formado pelas espécies analisadas do gênero *Adansonia*. Estes quatro

gêneros, por sua vez, tiveram como irmão o gênero *Bombax*, com o gênero *Scleronema* sendo a primeira linhagem a divergir dentro da subfamília. Na subfamília Malvatheca foi possível identificar dois clados com alto suporte. O primeiro clado foi formado por dois subclados distintos, sendo um formado por espécies do gênero *Malva*, *Althaea officinalis*, *Abutilon theophrasti*, e um subclado irmão composto pelas espécies *Abelmoschus esculentus*, *Hibiscus syriacus*. O segundo clado foi formado pelas espécies de *Gossypium*, *Kokia drynarioides* e *Thespesia populneoides*.

Em relação à árvore de similaridade de repeats baseada na abordagem AAF, a grande maioria dos nós apresentaram suporte máximo com exceção do nó ancestral do primeiro clado principal de Malvoideae (approx. 57%) (Figura 5b). As topologias não se mostraram inteiramente congruentes com a árvore baseada em plastomas, com poucas diferenças em relação às duas árvores. Dentre estas diferenças, destacam-se várias trocas de posições dentro do gênero *Adansonia* na subfamília Bombacoideae e a troca de posição entre os gêneros *Thespesia* e *Kokia* no segundo clado da subfamília Malvoideae. Outra diferença significativa na subfamília Bombacoideae foi a mudança de posição das linhagens do gênero *Scleronema* e *Bombax*. Enquanto na topologia dos plastomas essas linhagens foram a primeira e segunda a divergir respectivamente, na árvore AAF elas foram reconstruídas como irmãs de *Adansonia*. Consequentemente, o clado *Pseudobombax*+*Ceiba*+*Pachira* apareceu na topologia AAF como irmão do resto da subfamília.

Em contraste, a filogenia baseada na abundância dos elementos repetitivos não se mostrou congruente à baseada em plastomas, apresentando maiores diferenças (Figura 5a). Na subfamília Bombacoideae, observou-se a troca de posição de todas as espécies do gênero *Adansonia*, além desse ser reconstruído como linhagem irmã de *Bombax*, enquanto *Scleronema* apareceu como irmã de *Pseudobombax*+*Ceiba*+*Pachira*. Para a subfamília Malvoideae, *Hibiscus syriacus*, *Abelmoschus esculentus* e *Abutilon theophrasti* apareceram aninhadas no segundo clado, junto com as espécies de *Gossypium*, *Kokia drynarioides* e *Thespesia populneoides*.

Além das diferenças posicionais, foi também observada uma diferença em relação ao tamanho dos ramos entre as duas árvores baseadas em repeats. Na árvore AAF foram observados ramos relativamente grandes e uniformes para a maioria das espécies. Já a topologia de abundância dos elementos repetitivos mostrou um contraste de variação nos tamanhos dos ramos entre as duas subfamílias, com

espécies de Malvoideae apresentando ramos claramente maiores em comparação com Bombacoideae. Em Bombacoideae, o tamanho dos ramos variou de 0 nas espécies do gênero *Adansonia* a 0,2 em *Pachira aquatica*, enquanto em Malvoideae a variação foi de 0,1 em *Malva sylvestris* a 0,38 em *Abutilon theophrasti*.

## DISCUSSÃO

### *A origem paleopoliploide no clado Bombacoideae relacionada a incremento de diversidade dos repeats*

Nós apresentamos aqui a primeira análise genômica comparativa do clado Malvatheca, com foco em representantes da subfamília Bombacoideae. Essa subfamília é monofilética e se destaca por apresentar números cromossômicos altos (Baum e Oginuma 1994; Figueredo et al. 2016; Costa et al. 2017) associado a um tempo de origem antigo (62–56 milhões de anos) o que sugere uma origem paleopoliploide (Marinho et al. 2014). A estabilidade citomolecular, evidenciada pelo número/ posição dos sítios de DNAr, sugere que esses genomas já estejam altamente diploidizados e corrobora a hipótese de origem paleopoliploide (Figueredo et al. 2016; Costa et al. 2017). As análises das frações repetitivas dos genomas revelaram uma alta diversidade dos repeats no clado Malvatheca. Esses resultados contrastam com uma maior similaridade na composição de repeats em análises genômicas macroevolutivas (Beric et al. 2021; Zuo et al. 2022). Essa alta diversidade em Malvatheca pode estar associada a idade antiga do grupo e/ ou diversidade ecológica dos seus representantes (Zizka et al. 2019). É interessante que o destino da maioria dos poliploides recém-formados parece um beco sem saída evolutivo ao longo da competição por seus progenitores especializados diploides (Stebbins 1950, Wagner 1970, Mayrose et al. 2011) devido a uma gama de efeitos negativos associados, como exclusão de citótipos minoritários, anormalidades meióticas e mitóticas graves e instabilidade genômica associada à ploidia (Levin 1975, Madlung et al. 2005, Wang et al. 2010). No entanto, parece que existe uma forte ligação entre estresse ambiental e/ou flutuação e duplicação do genoma, como suportado pelo estabelecimento de linhagens paleopoliploides a partir do Cretáceo–Palaeogeno (Vannest et al. 2014; Marinho et al. 2014). Surpreendentemente, antigas WGD no Cretáceo–Palaeogeno (66 Ma) ocorreram independentemente em várias famílias de angiospermas (ex: Rosaceae, Fabaceae, Salicaceae, Brasicaceae, Malvaceae, Asteraceae, Ranunculaceae, Poaceae e Musaceae). Isso sugere que essa origem poliploide de

cada linhagem pode ter impulsionado a evolução de inovações biológicas e especiação no contexto de condições encontradas após fim do Cretáceo, por meio de extensa modificação de rede e diversificação funcional de genes reguladores e de desenvolvimento duplicados (Vanestrelle et al. 2014). Parece, assim, que o papel do ambiente tanto no estabelecimento dos poliploides quanto no seu sucesso evolutivo constitui um aspecto importante que merece uma investigação mais aprofundada em Bombacoideae.

No clado Malvatheca como um todo, assim como individualmente nas subfamílias Bombacoideae e Malvoideae, houve uma maior abundância genômica de elementos LTRs da superfamília Ty3/gypsy do que de Ty1/copia. Esse padrão é amplamente reportado em outros grupos plantas (Macas et al 2015; Neumann et al. 2020; Mata-Sucre et al. 2020). É interessante notar que o evento antigo de WGD na origem de Bombacoideae aparentemente não causou uma alteração nessa tendência. Esse resultado corrobora análises prévias que sugerem que não existe mudança na abundância de TE associada a WGD antigas (Beric et al. 2021). Por outro lado, nas análises genômicas comparativas foi observada uma pequena quantidade de repeats (TEs e DNAsat) com forte sinal filogenético, sendo espécie- ou gênero-específicos. Isso sugere uma heterogeneidade na idade de origem/ amplificação desses elementos repetitivos, que podem ter um rápido turnover (Dodsworth et al. 2020). Nesse sentido, os DNAsats parecem apresentar uma evolução mais acelerada, apresentando um forte significado filogenético sendo inclusive utilizados para remontar relações de parentesco e/ou origem de paleopoliploides (Oliveira et al. 2021).

Domínios proteicos de elementos transponíveis tendem a acumular mutações ao acaso após sua inserção, dando origem a novas cópias que podem ser amplificadas diferencialmente no genoma (Neumann et al. 2019). Dessa forma, ao longo do tempo, é possível que o repeatoma de diferentes espécies apresenta diferenças significativas na composição de sequências de uma mesma linhagem de repeats (Mascagni et al. 2020). Dada a idade de separação entre as duas subfamílias do clado Malvatheca, o esperado seria uma clara separação entre domínios proteicos de ambas as subfamílias. No entanto, nossos resultados corroboraram a heterogeneidade no modo de evolução dos diferentes elementos transponíveis em Malvatheca. Enquanto alguns elementos mostraram uma separação clara entre domínios das duas subfamílias, outros mostraram relações menos resolvidas, com menor estruturação. Estes resultados apontam não só para uma baixa resolução

filogenética promovida por *repeats* em *Malvatheca*, como também para uma baixa influência da poliploidia na diversificação dos domínios proteicos. Interessantemente, os dois elementos que mostram o maior grau de separação entre as subfamílias são representantes da superfamília Ty1/copia (SIRE e IKEROS). Tem sido reportado que, enquanto elementos Ty3/gypsy tendem a ter localização preferencial em regiões heterocromáticas (tais como centrômeros e peri-centrômeros), elementos Ty1/copia apresentam uma distribuição mais dispersa/aleatória, muitas vezes sendo encontrados próximos a regiões ricas em genes (Heckmann et al., 2013; De Souza et al., 2018; Van-Lume et al., 2019; Neumann et al., 2021). Sendo assim, a separação filogenética observada entre elementos Ty1/copia das duas subfamílias pode estar relacionada com uma distribuição mais conservada destes elementos no genoma.

A sugestão de que as diferentes classes de *repeats* que compõem um genoma operam de forma análoga a diferentes espécies em uma comunidade é a base da chamada “ecologia genômica” (Brookfield, 2005). Desta forma, métricas comumente aplicadas à ecologia de comunidades podem ser utilizadas para entender a dinâmica da evolução de *repeats* (Schley et al. 2022). Apesar de nossos resultados sugerirem que a paleopoliploidia de *Bombacoideae* não trouxe grandes alterações em relação à abundância e sequência dos *repeats* quando comparado à *Malvoideae*, métricas de diversidade dos *repeats* (índices de Shannon e Simpson) mostraram alta correlação com o número cromossômico. Valores baixos para os índices de Shannon e Simpson indicam alta probabilidade de que dois *repeats* escolhidos aleatoriamente em um genoma pertençam à mesma classe, denotando alta dominância de uma ou poucas classes específicas (Schley et al. 2022). Por conseguinte, este resultado mostra uma tendência clara de que os genomas paleopoliploides analisados aqui apresentarem uma maior variedade de diferentes classes de elementos repetitivos. Eventos de poliploidia geralmente ocasionam rápida mudança na fração repetitiva, seja por amplificação ou perda de diferentes elementos, como pode ser observado em neopoliploides (Vicent e Casacuberta 2017). Neste caso, é possível que poucas classes específicas de *repeats* sejam mais abundantes, dominando a fração repetitiva. Podemos observar este cenário em espécies diploides de *Malvoideae*, tais como *Gossypium arboreum*, *Hybiscus syriacus* e *Malva pusilla*, que apresentam poucas clusters de *repeats* com uma alta abundância em nossa análise comparativa. Por outro lado, espécies paleopoliploides de *Bombacoideae*, como os indivíduos

analisados de *Pseudobombax croizatii* e *Bombax ceiba*, apresentam vários clusters com abundâncias similares.

*Os repeats falham em determinar relações filogenéticas para o clado Malvatheca, mas apontam diferentes tendências entre Bombacoideae e Malvoideae*

Interpretando as frações repetitivas dos genomas do ponto de vista de inferência filogenética, tanto o método de abundância quanto o método de similaridade de repeats falham em mostrar relações verdadeiras entre as linhagens de Malvatheca. As topologias baseadas em abundância e similaridade de repeats revelaram tendências diferentes entre as duas subfamílias, onde Bombacoideae apresenta um mais alto grau de incongruência em relação à topologia plastidial. Esse resultado é interessante pois sugere que a paleopoliploidia em Bombacoideae pode estar associada a uma maior aleatoriedade na amplificação/ desamplificação dos repeats num padrão aparentemente randômico. Isso pode estar associado a mecanismos de diploidização, que podem ter levado de forma aleatória que alguns repeats apresentassem composições/ abundâncias similares em grupos filogeneticamente não relacionados (Renny-Byfield et al. 2013). Além disso, no(s) evento(s) de poliploidia que originou o clado Bombacoideae deve ter ocorrido “choque genômico” levando à diferentes tendências evolutivas em cada linhagem de Bombacoideae ao passar do tempo (Schrantz et al. 2012). É sabido que o estabelecimento de árvores filogenéticas utilizando repeats é mais complexa em amostragens incluindo poliploides, uma vez que podem gerar incongruências/ baixos suportes (Oliveira et al. 2021). Alternativamente o insucesso de estabelecimento de relações filogenéticas robustas em Malvatheca pode estar associada à idade antiga do grupo, uma vez que os repeats apresentam evolução rápida e podem perder o sinal filogenético depois de um longo tempo de divergência (Van-Lume et al. Em preparação).

A nossa análise de similaridade, realizada pelo método Assembly and Alignment-Free (AAF) (Fan et al. 2015), gerou topologias com melhores suportes, mas também revelou incongruências em relação à árvore plastidial. Esse método constrói filogenias diretamente a partir de dados de sequências genômicas não montados (reads brutos), ignorando as etapas de montagem e o alinhamento do genoma (Fan et al. 2015). Apesar do alto potencial do método AAF para resolver relações filogenéticas, inclusive entre organismos não-modelo (Fan et al. 2015; Oliveira et al. 2021; Van-Lume et al. Em preparação) nossos dados sugerem que essa abordagem

pode falhar em grupos com divergência muito antiga. Apesar da baixa resolução filogenética, a diferenciação dos clados Bombacoideae e Malvoideae é notória no que diz respeito aos tamanhos dos ramos na árvore das abundâncias dos repeats (Dodsworth et al. 2015). Nesse caso, se observam ramos mais longos no clado Malvoideae o que pode sugerir que o evento de paleopoliploidia no clado Bombacoideae resultou em uma menor variação nas abundâncias dos repeats. A hipótese de maior estabilidade genômica em representantes do clado Bombacoideae é corroborada por análises cito-moleculares, que revelaram baixa divergência cariotípica em nível genérico (Figueredo et al. 2016; Costa et al. 2017). Já Malvoideae, que é composta por espécies diploides e neopoliploides (ex. *Hybiscus syriacus* e *Abelmoschus esculentus*) apresentou uma maior heterogeneidade dos valores de abundância dos repeats. Neste caso, essa alta variação nas abundâncias dos repeats ocasionou ramos maiores para as espécies desta subfamília (Dodsworth et al. 2015). Essa tendência para Malvoideae é corroborada por trabalhos anteriores que indicam que a abundância de TE é impactada por eventos de neopoliploidia (Parisod et al. 2010; A'g ren et al. 2016; Vicent e Casacuberta 2017). Dessa forma, apesar de não resolver precisamente as relações filogenéticas, a árvore de abundância dos repeats apresentada aqui parece esclarecer tendências genômicas diferenciais entre os clados Bombacoideae e Malvoideae.

## CONCLUSÕES

A partir do presente estudo pode-se concluir que o uso de *repeats* para construções filogenéticas de famílias muito antigas como o caso da subfamília paleopoliploide Bombacoideae não é bem suportado, porém o resultado dessas análises aponta diferentes tendências entre as subfamílias Bombacoideae e Malvoideae. Em contrapartida a paleopoliploidia em Bombacoideae pode estar associada a uma taxa maior de aleatoriedade em relação à amplificação e ou desamplificação dos repeats num padrão que aparenta ser aleatório. Assim como também as análises das frações repetitivas dos genomas mostraram uma alta diversidade dos elementos repetitivos no clado Malvatheca. Dessa forma reforçando a ideia da idade antiga do grupo, e ou a diversidade ecológica dos seus representantes. Portanto, o papel do ambiente tanto no estabelecimento dos poliploides quanto no seu sucesso evolutivo, merece uma investigação mais aprofundada em Bombacoideae.

## REFERÊNCIAS

1. Å ÅGREN, J. A.; HUANG, H. R.; WRIGHT, S. I. Transposable element evolution in the allotetraploid *Capsella bursa-pastoris*. *American Journal of Botany*, v. 103, n. 7, p. 1197-1202, 2016.
2. ARKHIPOVA, I. R. Using bioinformatic and phylogenetic approaches to classify transposable elements and understand their complex evolutionary histories. **Mobile DNA**, v. 8, n. 1, p. 1-14, 2017.
3. BAUM, D. A.; OGINUMA, K. A review of chromosome numbers in Bombacaceae with new counts for *Adansonia*. **Taxon**, v. 43, n. 1, p. 11-20, 1994.
4. BAYER, C.; KUBITZKI, K. Malvaceae. In: **Flowering plants· dicotyledons**. Berlin, Heidelberg: Springer, 2003. p. 225-311.
5. BERIC, A. et al. Comparative phylogenetics of repetitive elements in a diverse order of flowering plants (Brassicales). **G3**, v. 11, n. 7, p. jkab140, 2021.
6. ALVERSON, W. S. et al. Phylogeny of the core Malvales: evidence from ndhF sequence data. **American Journal of Botany**, v. 86, n. 10, p. 1474-1486, 1999.
7. BROOKFIELD, J. F. The ecology of the genome—mobile DNA elements and their hosts. **Nature Reviews Genetics**, v. 6, n. 2, p. 128-136, 2005.
8. CHRISTENHUSZ, M. J.; BYNG, J. W. The number of known plants species in the world and its annual increase. **Phytotaxa**, v. 261, n. 3, p. 201-217, 2016.
9. R CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2019.
10. COSTA, L. et al. Comparative cytomelecular analyses reveal karyotype variability related to biogeographic and species richness patterns in Bombacoideae (Malvaceae). **Plant Systematics and Evolution**, v. 303, n. 9, p. 1131-1144, 2017.
11. COSTA, L. et al. Aiming off the target: recycling target capture sequencing reads for investigating repetitive DNA. **Annals of Botany**, v. 128, n. 7, p. 835-848, 2021.
12. DE SOUZA, T. B. et al. Analysis of retrotransposon abundance, diversity and distribution in holocentric *Eleocharis* (Cyperaceae) genomes. **Annals of Botany**, v. 122, n. 2, p. 279-290, 2018.
13. DODSWORTH, S.; CHASE, M. W.; LEITCH, A. R. Is post-polyploidization diploidization the key to the evolutionary success of angiosperms?. **Botanical Journal of the Linnean Society**, v. 180, n. 1, p. 1-5, 2016.
14. DODSWORTH, S. Genome skimming for next-generation biodiversity analysis. **Trends in Plant Science**, v. 20, n. 9, p. 525-527, 2015.
15. DODSWORTH, S. et al. Genomic repeat abundances contain phylogenetic signal. **Systematic Biology**, v. 64, n. 1, p. 112-126, 2015.
16. DODSWORTH, S. et al. Repetitive DNA restructuring across multiple *Nicotiana* allopolyploidisation events shows a lack of strong cytoplasmic bias in influencing repeat turnover. **Genes**, v. 11, n. 2, p. 216, 2020.
17. ELLIOTT, T. A.; GREGORY, T. R. What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 370, n. 1678, p. 20140331, 2015.
18. FAN, H. et al. An assembly and alignment-free method of phylogeny

- reconstruction from next-generation sequencing data. **BMC Genomics**, v. 16, n. 1, p. 1-18, 2015.
19. FIGUEREDO, A. et al. Karyotypic stability in the paleopolyploid genus *Ceiba* Mill. (Bombacoideae, Malvaceae). **Brazilian Journal of Botany**, v. 39, n. 4, p. 1087-1093, 2016.
  20. FINNEGAN, D. J. Eukaryotic transposable elements and genome evolution. **Trends in Genetics**, v. 5, p. 103-107, 1989.
  21. GARRIDO-RAMOS, M. A. Satellite DNA: an evolving topic. **Genes**, v. 8, n. 9, p. 230, 2017.
  22. HECKMANN, S. et al. The holocentric species *Luzula elegans* shows interplay between centromere and large-scale genome organization. **The Plant Journal**, v. 73, n. 4, p. 555-565, 2013.
  23. HUSBAND, B. C.; BALDWIN, S. J.; SUDA, J. The incidence of polyploidy in natural plant populations: major patterns and evolutionary processes. In: **Plant Genome Diversity Volume 2**. Vienna: Springer, 2013. p. 255-276.
  24. JIAO, Y. et al. Ancestral polyploidy in seed plants and angiosperms. **Nature**, v. 473, n. 7345, p. 97-100, 2011.
  25. KATOH, K.; STANDLEY, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. **Molecular Biology and Evolution**, v. 30, n. 4, p. 772-780, 2013.
  26. KEARSE, M. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. **Bioinformatics**, v. 28, n. 12, p. 1647-1649, 2012.
  27. LEVIN, D. A. Minority cytotype exclusion in local plant populations. **Taxon**, v. 24, n. 1, p. 35-43, 1975.
  28. MACAS, J. et al. In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe Fabaeae. **PLoS One**, v. 10, n. 11, p. e0143424, 2015.
  29. MADLUNG, A. et al. Genomic changes in synthetic *Arabidopsis* polyploids. **The Plant Journal**, v. 41, n. 2, p. 221-230, 2005.
  30. MARINHO, R. C. et al. Do chromosome numbers reflect phylogeny? New counts for Bombacoideae and a review of Malvaceae sl. **American Journal of Botany**, v. 101, n. 9, p. 1456-1465, 2014.
  31. MARQUES, A. et al. Holocentromeres in *Rhynchospora* are associated with genome-wide centromere-specific repeat arrays interspersed among euchromatin. **Proceedings of the National Academy of Sciences**, v. 112, n. 44, p. 13633-13638, 2015.
  32. MASCAGNI, F. et al. A computational comparative study of the repetitive DNA in the genus *Quercus* L. **Tree Genetics & Genomes**, v. 16, n. 1, p. 1-11, 2020.
  33. MAYROSE, I. et al. Recently formed polyploid plants diversify at lower rates. **Science**, v. 333, n. 6047, p. 1257, 2011.
  34. MEHROTRA, S.; GOYAL, V. Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. **Genomics Proteomics Bioinformatics**, v. 12, p. 164-171, 2014.
  35. NEUMANN, P. et al. Systematic survey of plant LTR-retrotransposons elucidates

- phylogenetic relationships of their polyprotein domains and provides a reference for element classification. **Mobile DNA**, v. 10, n. 1, p. 1-17, 2019.
36. NEUMANN, P. et al. Impact of parasitic lifestyle and different types of centromere organization on chromosome and genome evolution in the plant genus *Cuscuta*. **New Phytologist**, v. 229, n. 4, p. 2365-2377, 2021.
  37. NOVÁK, P. et al. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. **Bioinformatics**, v. 29, n. 6, p. 792-793, 2013.
  38. OLIVEIRA, M. A. et al. High-Throughput Genomic Data Reveal Complex Phylogenetic Relationships in *Stylosanthes* Sw (Leguminosae). **Frontiers in Genetics**, v. 12, p. 1846, 2021.
  39. PARISOD, C. et al. Impact of transposable elements on the organization and function of allopolyploid genomes. **New Phytologist**, v. 186, n. 1, p. 37-45, 2010.
  40. PÉCHON, T. L.; GIGORD, L. D. On the relevance of molecular tools for taxonomic revision in Malvales, Malvaceae s.l., and Dombeyoideae. In: **Molecular Plant Taxonomy**. Humana Press, Totowa, NJ, 2014. p. 337-363.
  41. RAMSEY, J.; SCHEMSKE, D. W. Neopolyploidy in flowering plants. **Annual Review of Ecology and Systematics**, v. 33, p. 589-639, 2002.
  42. RAMSEY, J.; SCHEMSKE, D. W. Pathways, mechanisms, and rates of polyploid formation in flowering plants. **Annual Review of Ecology and Systematics**, v. 29, p. 467-501, 1998.
  43. RENNY-BYFIELD, S. et al. Diploidization and genome size change in allopolyploids is associated with differential dynamics of low-and high-copy sequences. **The Plant Journal**, v. 74, n. 5, p. 829-839, 2013.
  44. SCHLEY, R. J. et al. The ecology of palm genomes: repeat-associated genome size expansion is constrained by aridity. **New Phytologist**, 2022.
  45. SHANNON, C. E. A mathematical theory of communication. **The Bell System Technical Journal**, v. 27, n. 3, p. 379-423, 1948.
  46. SIMPSON, E. H. Measurement of diversity. **Nature**, v. 163, n. 4148, p. 688, 1949.
  47. SOLTIS, D. E. et al. Polyploidy and angiosperm diversification. **American Journal of Botany**, v. 96, n. 1, p. 336-348, 2009.
  48. SOLTIS, P. S. et al. Polyploidy and genome evolution in plants. **Current Opinion in Genetics & Development**, v. 35, p. 119-125, 2015.
  49. STEBBINS, G. L. **Variation and evolution in plants**. Columbia biological series, 1950.
  50. SCHRANZ, M. E.; MOHAMMADIN, S.; EDGER, P. P. Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. **Current Opinion in Plant Biology**, v. 15, n. 2, p. 147-153, 2012.
  51. WEISS-SCHNEEWEISS, H.; SCHNEEWEISS, G. M. Karyotype diversity and evolutionary trends in angiosperms. In: **Plant Genome Diversity Volume 2**. Springer, Vienna, 2013. p. 209-230.
  52. VANNESTE, K.; MAERE, S.; VAN DE PEER, Y. Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 369, n. 1648, p. 20130353, 2014.

53. VAN-LUME, B. et al. Evolutionary convergence or homology? Comparative cytogenomics of Caesalpinia group species (Leguminosae) reveals diversification in the pericentromeric heterochromatic composition. **Planta**, v. 250, n. 6, p. 2173-2186, 2019.
54. ICIENT, C. M.; CASACUBERTA, J. M. Impact of transposable elements on polyploid plant genomes. **Annals of Botany**, 2017.
55. WANG, Y. et al. Polyploidy-associated genomic instability in *Arabidopsis thaliana*. **Genesis**, v. 48, n. 4, p. 254-263, 2010.
56. WAGNER JR, W. H. Biosystematics and evolutionary noise. **Taxon**, v. 19, n. 2, p. 146-151, 1970.
57. WOOD, T. E. et al. The frequency of polyploid speciation in vascular plants. **Proceedings of the National Academy of Sciences**, v. 106, n. 33, p. 13875-13879, 2009.
58. ZIZKA, A. et al. Transitions between biomes are common and directional in Bombacoideae (Malvaceae). **Journal of Biogeography**, v. 47, n. 6, p. 1310-1321, 2020.
59. ZUO, S. et al. Genomes, repeatomes and interphase chromosome organization in the meadowfoam family (Limnanthaceae, Brassicales). **The Plant Journal**, 2022.

## 5 REFERÊNCIAS

1. ACHREM, M.; SZUĆKO, I.; KALINKA, A. The epigenetic regulation of centromeres and telomeres in plants and animals. **Comparative Cytogenetics**, v. 14, n. 2, p. 265, 2020.
2. ANDERMANN, T. et al. A guide to carrying out a phylogenomic target sequence capture project. **Frontiers in Genetics**, v. 11, p. 1407, 2020.
3. ARKHIPOVA, I. R. Using bioinformatic and phylogenetic approaches to classify transposable elements and understand their complex evolutionary histories. **Mobile DNA**, v. 8, n. 1, p. 1-14, 2017.
4. BAYER, C.; KUBITZKI, K. Malvaceae. In: **Flowering plants- dicotyledons**. Berlin, Heidelberg: Springer, 2003. p. 225-311.
5. BIÉMONT, C. A brief history of the status of transposable elements: from junk DNA to major players in evolution. **Genetics**, v. 186, n. 4, p. 1085-1093, 2010.
6. BISCOTTI, M. A.; OLMO, E.; HESLOP-HARRISON, J. S. Repetitive DNA in eukaryotic genomes. **Chromosome Research**, v. 23, n. 3, p. 415-420, 2015.
7. CAO, Y. et al. Molecular characterization of a transcriptionally active Ty1/copia-like retrotransposon in *Gossypium*. **Plant Cell Reports**, v. 34, n. 6, p. 1037-1047, 2015.
8. COSTA, L. et al. Aiming off the target: recycling target capture sequencing reads for investigating repetitive DNA. **Annals of Botany**, v. 128, n. 7, p. 835-848, 2021.
9. DODSWORTH, S. et al. Genomic repeat abundances contain phylogenetic signal. **Systematic Biology**, v. 64, n. 1, p. 112-126, 2015.
10. DODSWORTH, S.; CHASE, M. W.; LEITCH, A. R. Is post-polyploidization diploidization the key to the evolutionary success of angiosperms?. **Botanical Journal of the Linnean Society**, v. 180, n. 1, p. 1-5, 2016.

11. DODSWORTH, S. Genome skimming for next-generation biodiversity analysis. **Trends in Plant Science**, v. 20, n. 9, p. 525-527, 2015.
12. DOS SANTOS, R. Z. et al. A long-term conserved satellite DNA that remains unexpanded in several genomes of Characiformes fish is actively transcribed. **Genome Biology and Evolution**, v. 13, n. 2, p. evab002, 2021.
13. ELLIOTT, T. A.; GREGORY, T. R. What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 370, n. 1678, p. 20140331, 2015.
14. FAN, H. et al. An assembly and alignment-free method of phylogeny reconstruction from next-generation sequencing data. **BMC Genomics**, v. 16, n. 1, p. 1-18, 2015.
15. FESCHOTTE, C.; JIANG, N.; WESSLER, S. R. Plant transposable elements: where genetics meets genomics. **Nature Reviews Genetics**, v. 3, n. 5, p. 329-341, 2002.
16. FIGUEREDO, A. et al. Karyotypic stability in the paleopolyploid genus *Ceiba* Mill. (Bombacoideae, Malvaceae). **Brazilian Journal of Botany**, v. 39, n. 4, p. 1087-1093, 2016.
17. FINNEGAN, D. J. Eukaryotic transposable elements and genome evolution. **Trends in Genetics**, v. 5, p. 103-107, 1989.
18. GARRIDO-RAMOS, M. A. Satellite DNA: an evolving topic. **Genes**, v. 8, n. 9, p. 230, 2017.
19. HESLOP-HARRISON, J. P.; SCHWARZACHER, T. Genetics and genomics of crop domestication. In: **Plant Biotechnology and Agriculture**. 2012. p. 3-18.
20. HUSBAND, B. C.; BALDWIN, S. J.; SUDA, J. The incidence of polyploidy in natural plant populations: major patterns and evolutionary processes. In: **Plant Genome Diversity Volume 2**. Vienna: Springer, 2013. p. 255-276.
21. JIAO, Y. et al. Ancestral polyploidy in seed plants and angiosperms. **Nature**, v. 473, n. 7345, p. 97-100, 2011.
22. KELLY, L. J. et al. Analysis of the giant genomes of *Fritillaria* (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size. **New Phytologist**, v. 208, n. 2, p. 596-607, 2015.
23. KUMAR, A.; BENNETZEN, J. L. Plant retrotransposons. **Annual Review of Genetics**, v. 33, n. 1, p. 479-532, 1999.
24. LEE, J. et al. Rapid amplification of four retrotransposon families promoted speciation and genome size expansion in the genus *Panax*. **Scientific Reports**, v. 7, n. 1, p. 1-9, 2017.
25. LÓPEZ-FLORES, I.; GARRIDO-RAMOS, M. A. The repetitive DNA content of eukaryotic genomes. **Repetitive DNA**, v. 7, p. 1-28, 2012.
26. LYU, H. et al. Convergent adaptive evolution in marginal environments: unloading transposable elements as a common strategy among mangrove genomes. **New Phytologist**, v. 217, n. 1, p. 428-438, 2018.
27. MACAS, J. et al. In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe Fabeae. **PLoS One**, v. 10, n. 11, p. e0143424, 2015.
28. MEYERS, B. C.; TINGEY, S. V.; MORGANTE, M. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. **Genome**

- Research**, v. 11, n. 10, p. 1660-1676, 2001.
29. MAYROSE, I.; LYSAK, M. A. The evolution of chromosome numbers: mechanistic models and experimental approaches. **Genome Biology and Evolution**, v. 13, n. 2, p. evaa220, 2021.
  30. MARINHO, R. C. et al. Do chromosome numbers reflect phylogeny? New counts for Bombacoideae and a review of Malvaceae sl. **American Journal of Botany**, v. 101, n. 9, p. 1456-1465, 2014.
  31. MATSUNAGA, S. et al. LTR retrotransposons in the dioecious plant *Silene latifolia*. **Genome**, v. 45, n. 4, p. 745-751, 2002.
  32. MCCLINTOCK, B. Carnegie Inst. Wash. Year Book, v. 47, p. 155-169, 1948.
  33. MCCLINTOCK, B. The significance of responses of the genome to challenge. **Science**, v. 226, n. 4676, p. 792-801, 1984.
  34. MEHROTRA, S.; GOYAL, V. Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. **Genomics Proteomics Bioinformatics**, v. 12, p. 164-171, 2014.
  35. NOVÁK, P. et al. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. **Bioinformatics**, v. 29, n. 6, p. 792-793, 2013.
  36. RAMSEY, J.; SCHEMSKE, D. W. Pathways, mechanisms, and rates of polyploid formation in flowering plants. **Annual Review of Ecology and Systematics**, v. 29, p. 467-501, 1998.
  37. RAMSEY, J.; SCHEMSKE, D. W. Neopolyploidy in flowering plants. **Annual Review of Ecology and Systematics**, v. 33, p. 589-639, 2002.
  38. ROBERTO, C. Low chromosome number angiosperms. **Caryologia**, v. 58, n. 4, p. 403-409, 2005.
  39. ŠATOVIĆ-VUKŠIĆ, E.; PLOHL, M. Classification Problems of Repetitive DNA Sequences. **DNA**, v. 1, n. 2, p. 84-90, 2021.
  40. SCHLEY, R. J. et al. The ecology of palm genomes: repeat-associated genome size expansion is constrained by aridity. **New Phytologist**, 2022.
  41. SOLTIS, D. E. et al. Polyploidy and angiosperm diversification. **American Journal of Botany**, v. 96, n. 1, p. 336-348, 2009.
  42. SOLTIS, P. S. et al. Polyploidy and genome evolution in plants. **Current Opinion in Genetics & Development**, v. 35, p. 119-125, 2015.
  43. WEISS-SCHNEEWEISS, H. et al. Evolutionary consequences, constraints and potential of polyploidy in plants. **Cytogenetic and Genome Research**, v. 140, n. 2-4, p. 137-150, 2013.
  44. WEISS-SCHNEEWEISS, H. et al. Employing next generation sequencing to explore the repeat landscape of the plant genome. **Next Generation Sequencing in Plant Systematics**. *Regnum Vegetabile*, v. 157, p. 155-179, 2015.