

**VANESSA EMANUELLE PEREIRA SANTOS**

**CARACTERIZAÇÃO DA FRAÇÃO REPETITIVA DO GENOMA DE *Hancornia speciosa*  
GOMES**

RECIFE - PE

2018

**CARACTERIZAÇÃO DA FRAÇÃO REPETITIVA DO GENOMA DE *Hancornia speciosa*  
GOMES**

Dissertação apresentada ao Programa de Pós-Graduação em Agronomia – Melhoramento Genético de Plantas da Universidade Federal Rural de Pernambuco, como parte dos requisitos exigidos para obtenção do título de Mestre em Melhoramento Genético de Plantas.

**COMITÊ DE ORIENTAÇÃO:**

Dr. Edson Ferreira Silva - Orientador - DB/UFRPE

Dr. Cícero Carlos de Souza Almeida – Coorientador -UFAL

RECIFE – PE

2018

**CARACTERIZAÇÃO DA FRAÇÃO REPETITIVA DO GENOMA DE *Hancornia speciosa*  
GOMES**

**VANESSA EMANUELLE PEREIRA SANTOS**

Data de Aprovação:

**Banca Examinadora**

---

Dr. Edson Ferreira da Silva  
Universidade Federal Rural de Pernambuco – UFRPE  
Orientador

---

Dr. Reginaldo de Carvalho  
Universidade Federal Rural de Pernambuco – UFRPE  
Examinador

---

Dr. Amaro de Castro Lira Neto  
Instituto Agronômico de Pernambuco – IPA  
Examinador

*A Deus, à minha família, aos meus amigos, professores e a todos que direta ou indiretamente contribuíram na conclusão desta etapa de minha vida.*

## AGRADECIMENTOS

Agradeço primeiramente a Deus pelo dom da vida e possibilidade de realização deste trabalho, sem Ele nada seria possível.

Aos meus pais Verônica e Josevaldo pelo incentivo, paciência, amor e por não medirem esforços para que eu chegasse aonde cheguei.

Ao meu irmão João Victor e minha bebê Lilica pelo apoio emocional ao longo dos anos.

Ao meu orientador Edson Ferreira da Silva, por toda orientação, aprendizado e auxílio que me foram dados.

Ao meu coorientador Cícero Carlos de Souza Almeida que contribui com meu sucesso acadêmico desde a graduação, proporcionando conhecimentos que levarei para toda a vida.

Aos amigos do Laboratório de Recursos Genéticos da Universidade Federal de Alagoas (UFAL) – *Campus Arapiraca*, em especial Daniel, Eliane, Jackeline, Merieli, Élvia e Gleica por toda ajuda que me foi dada na realização deste trabalho.

A minha amiga Leylanne por toda ajuda com as correções do trabalho, por todo incentivo, paciência e torcida por todos os meus objetivos de vida.

Aos meus amigos de infância Ricardo, Luzia, Flávio, Gabriella, Tayslanne, Fabiana e Johanderson pelo incentivo e torcida nesta caminhada acadêmica.

Aos amigos de graduação Nathaly, Bruna, Karla, Daniela, Suzilaine, Maria Cristina, Rodrigo e Ana Paula, que se fizeram presentes nestes dois anos de mestrado.

As amigas que Recife me deu, Bruna, Sammara e Josinete pela paciência, correções e conselhos, os quais foram muito importantes para que eu não desistisse no meio do caminho.

Ao meu amigo Joelson Crispim, minha dupla de mestrado e de vida, por toda ajuda que me foi dada na execução e conclusão do trabalho, desde os conselhos até as ajudas com formatações.

Ao Laboratório de Citogenética Vegetal da Universidade Federal de Pernambuco (UFPE) pela ajuda em parte das análises.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudos que permitiu minha moradia em Recife, manutenção e execução do projeto.

Ao Conselho Nacional de Pesquisa e Aperfeiçoamento (CNPq) pela concessão dos recursos para realização deste trabalho.

Ao programa de Pós-graduação em Agronomia – Melhoramento genético de plantas (PPAMGP) da Universidade Federal Rural de Pernambuco (UFRPE).

## RESUMO GERAL

A mangabeira é uma espécie nativa do Brasil, amplamente distribuída no Brasil, e que apresenta importância socioeconômica, principalmente nas regiões Nordeste e Centro-Oeste do país. *Hancornia speciosa* ocorre na Caatinga, no Cerrado, nos Tabuleiros Costeiros e nas Restingas. Seus frutos contêm vitamina C e apresentam os maiores teores de proteína e ferro e suas folhas possuem propriedades medicinais sendo utilizadas no controle da pressão arterial e no tratamento contra diabetes. Devido à expansão imobiliária, utilização de áreas nativas para cultivo de grandes culturas e fragmentação das florestas, a espécie vem sofrendo significativa erosão genética. Apesar de sua importância e vulnerabilidade, estudos genômicos em *H. speciosa* são negligenciados, sendo assim, o conhecimento de suas características genômicas, os quais permitirão a aplicação de técnicas que contribuam com programas de conservação e melhoramento genético da espécie, são de suma importância. As plantas, geralmente, contêm elevadas proporções de DNA repetitivo, os quais são importantes elementos de regulação gênica e que surgem ao longo da evolução, ocasionando mutações que podem alterar fenotipicamente o organismo em questão. Neste sentido, o objetivo deste trabalho foi caracterizar a fração repetitiva de *H. speciosa* fornecendo o primeiro estudo genômico na espécie o qual contribuirá futuramente para técnicas aplicadas ao melhoramento genético envolvendo associação genômica. Indivíduos de mangaba, provenientes do Estado de Pernambuco foram coletados e tiveram seu DNA extraído e posteriormente fragmentado para construção da biblioteca *paired-end-reads* de sequenciamento de nova geração (NGS), usando a plataforma Illumina HiSeq2500. O tamanho do genoma foi estimado usando citometria de fluxo e a fração repetitiva foi analisada utilizando o software RepeatExplorer e TAREAN. Três espécies relacionadas geneticamente com *H. speciosa* foram utilizadas para comparar os resultados e avaliar se a quantidade de DNA repetitivo tem relação com a proximidade genética das espécies. O genoma de *H. speciosa* possui tamanho pequeno (430 Mb) e pouca quantidade de DNA repetitivo (15%) a qual é proveniente principalmente de elementos LTR/Gypsy e de sequências de DNAr. A espécie não apresenta sequências de DNA satélite e sugerimos que o papel deste elemento no genoma é assumido por sequências LTR/Gypsy chromovirus, os quais possuem ampla proporção em genomas, principalmente em regiões centroméricas. As informações obtidas no presente estudo fornecem informações iniciais aos estudos genômicos em *H. speciosa* os quais contribuirão para o melhoramento genético da mangaba.

**Palavras-chave:** mangaba, NGS, LTR/Gypsy, genoma, DNAr.

## GENERAL ABSTRACT

The mangabeira is a native species of Brazil, widely distributed and that presents socioeconomic importance mainly in the Northeast and Center-West regions of the country. In Brazil, *H. speciosa* occurs in the Caatinga, in the Cerrado, in the Coastal Tracks and in the Restingas. Its fruits contain vitamin C and have the highest levels of protein and iron and its leaves have medicinal properties being used to control blood pressure and to treat diabetes. Due to the real estate expansion, use of native areas for cultivation of large crops and fragmentation of the forests, the species has suffered significant genetic erosion. Despite its importance and vulnerability genetic and genomic studies in *H. speciosa* and therefore projects with the purpose of knowing and studying the genomic and genetic characteristics of the species are important in order to contribute to the conservation and genetic improvement of the species. Plants generally contain high proportions of repetitive DNA, which are important elements of gene regulation and that appear along the evolution causing mutations that can change phenotypically the organism in question. In this sense, the objective of this work was to characterize the repetitive fraction of *H. speciosa* providing the first genomic study in the species which will contribute in future to techniques applied to genetic improvement involving genomic association. Mangrove individuals from the State of Pernambuco were collected and their DNA extracted and later fragmented to construct the new generation sequencing *paired-end-reads* (NGS) library, using the Illumina HiSeq2500 platform. Genome size was estimated using flow cytometry and the repetitive fraction was analyzed using the RepeatExplorer and TAREAN software. Three species genetically related to *H. speciosa* were used to compare the results and to evaluate if the amount of repetitive DNA is related to the genetic proximity of the species. The genome of *H. speciosa* has a small size (430 Mb) and a small amount of repetitive DNA (15%) which comes mainly from LTR / Gypsy elements and from rDNA sequences. The species does not present satellite DNA sequences and we suggest that the role of this element in the genome is assumed by LTR / Gypsy chromovirus sequences which have a large proportion in genomes, especially in centromeric regions. The information obtained in the present study provides initial information to the genomic studies in *H. speciosa* which will contribute to the genetic improvement of the mangaba

**Keywords:** mangaba, NGS, LTR / Gypsy, genome, rDNA.

## LISTA DE ABREVIATURAS

**DNAsat** – DNA satélite

**DNAr** – DNA ribossomal

**IGS** – espaçadores intergênicos

**ITS** – espaçadores internos

**LTR** - *Long Terminal Repeat*

**NGS** - Sequenciamento de Nova Geração

**TE's** - Elementos transponíveis

**TIR** – *Terminal inverted repeat*

**LINE** – *Long Interspersed Nucleotide Elements*

**SINE** – *Short Interspersed Nuclear Elements*

**NTS** – Espaçador não transcrito

**ETS** – *External Transcribed Spacer*

**RT** – *Transcriptase reversa*

**INT** – *Integrase*

**MITE** – Elementos transposíveis de repetição invertida em miniatura

**QTL** – *Quantitative trait loci*



## LISTA DE FIGURAS

**Figura 1.** Tamanho do genoma de duas variedades de *Hancornia speciosa* (var. *gardneri* e var. *speciosa*) à teste de médias usando 1C value, pelo teste de Tukey ao nível de 5% de significância.

**Figura 2.** Árvore filogenética da ordem Gentianales baseada em genomas plastidiais inteiros, composta pelas subfamílias Ravoldioideae, Apocynoideae, Periplocoideae, Secamonoideae e Asclepiadeae e Gentianaceae (grupo externo).

**Figura 3.** (A) Proporção de elementos repetitivo nos genomas de *H. speciosa*, *C. roseus*, *R. stricta* e *A. syriaca* (B) comparação da fração repetitiva nos genomas das espécies identificando: sequências de DNA satélite; sequências repetitivas não classificadas (repetições simples); regiões repetitivas do DNAr e; elementos transponíveis classes I e II retrotransposons (LTRs e N-LTRs) e transposons, respectivamente.

**Figura 4.** Caracterização da fração repetitiva dos genomas de (A) *Asclepias syriaca* (B) *Catharanthus roseus* (C) *Hancornia speciosa* e (D) *Rhazya stricta*, quanto às famílias de elementos transponíveis e subunidades do DNAr.

**Figura 5.** DNA ribossomal 45S de *Hancornia speciosa* composto pelas subunidades 5.8s, 18s e 25.8s e regiões repetitivas ricas em sequências de LTR Gypsy Chromovirus dos tipos Galadriel e Reina dispostas em *tandem*.

## LISTA DE TABELAS

**Tabela 1.** Referências dos genomas plastidiais utilizados na análise filogenética e disponíveis no GenBank.

**Tabela 2 -** Tamanho dos genomas e resultados do agrupamento de sequências usando RepeatExplorer para as espécies *H. speciosa*, *C. roseus*, *R. stricta* e *A. syriaca*.

**Tabela 3 –** Distribuição de elementos repetitivos com mais de 0,01% de representação nos genomas de *H. speciosa*, *C. roseus*, *R. stricta* e *A. syriaca*, classificada de acordo com Wicker *et al.* (2007) em Classe, Subclasse, Superfamília e Família.

## Sumário

1. INTRODUÇÃO GERAL .....	13
2. REVISÃO DE LITERATURA .....	15
2.1 Origem, distribuição e características botânicas de <i>Hancornia speciosa</i> Gomes .....	15
2.2 Importância e vulnerabilidade da espécie.....	16
2.3 Recursos genéticos e melhoramento da espécie.....	17
2.4 Evolução do genoma vegetal e DNA repetitivo .....	18
2.4.1 DNA satélite .....	19
2.4.2 DNA ribossomal.....	20
2.4.3 Elementos transponíveis.....	20
2.4.3.1 Retrotransposons .....	22
2.4.3.2 Transposons .....	23
2.3 Aplicações do estudo de DNA repetitivo no melhoramento de plantas .....	24
2.4 Next-Generation Sequencing (NGS).....	25
3. REFERÊNCIAS .....	27
CAPÍTULO II – Tamanho genômico e identificação de sequências repetitivas no genoma de <i>Hancornia speciosa</i> Gomes (Apocynaceae; Gentianales) .....	33

---

## **CAPÍTULO I - REVISÃO DE LITERATURA**

## 1. INTRODUÇÃO GERAL

A mangabeira (*Hancornia speciosa* Gomes) é uma espécie arbórea neotropical, nativa do Brasil, classificada na ordem Gentianales e família Apocynaceae. É uma frutífera de ampla distribuição e que ocorre em locais de vegetação aberta, especialmente na Caatinga, no Cerrado, nos Tabuleiros Costeiros e nas Restingas (LIMA & SCARIOT, 2010). Além do Brasil (do Nordeste para o Centro-Oeste) esta espécie também foi relatada no Paraguai, Bolívia e Peru (COLLEVATTI *et al.*, 2016). A espécie engloba seis variedades: *H. speciosa* var. *speciosa*, *H. speciosa* var. *maximiliani*, *H. speciosa* var. *cuyabensis*, *H. speciosa* var. *lundii*, *H. speciosa* var. *gardneri* e *H. speciosa* var. *pubescens* (MONACHINO, 1945). Dentre estas, *H. speciosa* var. *speciosa* é a mais representativa no Brasil e ocorre desde a região Sudeste até a região Norte do país (LIMA & SCARIOT, 2010).

Os frutos da mangabeira são consumidos *in natura* e utilizados principalmente em indústrias de processamento de alimentos para fabricação de polpas, geleias e doces. A comercialização dos frutos é a atividade econômica mais expressiva para a espécie, porém, suas folhas, casca, látex e raízes são usados para finalidades medicinais (LIMA & SCARIOT, 2010). É uma planta alógama, que possui flores hermafroditas e é auto-incompatível carecendo, portanto, de agentes polinizadores e diferentes genótipos para que ocorra a fecundação cruzada e a produção dos frutos (COLLEVATTI *et al.*, 2016). A polinização é entomófila, geralmente realizada por mariposas e abelhas (LIMA & SCARIOT, 2010).

A coleta dos frutos de mangaba é feita principalmente por agroextrativistas, visto que não existem grandes áreas de cultivo de mangabeira. Em Sergipe 90% dos frutos comercializados provém de áreas nativas onde o extrativismo existe há séculos. Estima-se que neste Estado 2.500 famílias praticam esta atividade, sendo ela responsável por 60% da renda de cada família (LIMA & SCARIOT, 2010). O rápido avanço das fronteiras agrícolas para a implantação de grandes culturas, bem como a expansão imobiliária e a fragmentação das florestas, vem contribuindo significativamente para a redução de populações naturais de *H. speciosa*, sendo assim, estudos genéticos são de suma importância para auxiliar em programas de melhoramento e conservação nesta espécie (AMORIM *et al.*, 2015).

Ainda não existem estudos genômicos em *H. speciosa*, no entanto, os mesmos são de suma importância tanto para o melhoramento genético quanto para conservação da espécie.

Sequências de DNA repetitivo são contribuintes importantes no aumento do tamanho genômico de espécies vegetais e se fazem presente em elevadas proporções em organismos de genoma grande (KAZAZIAN JÚNIOR, 2004). Durante muito tempo essas sequências foram conhecidas como “DNA lixo”, no entanto, estudos genômicos realizados na última década constataram sua importância na organização cromossômica, controle da expressão gênica e na rápida adaptação de organismos as condições adversas a partir da interrupção dos mecanismos de silenciamento destes elementos no genoma, principalmente sob condições de estresse (SEIDL & THOMMA, 2017). Considerados como estruturas importantes para estudos de evolução de espécies, esses elementos estão cada vez mais envolvidos em pesquisas para as mais diversas aplicações e atualmente utilizados como marcadores eficazes em estudos de associação genômica, técnica importante para o melhoramento genético de plantas (ZERJAL *et al.*, 2012).

O sequenciamento de nova geração (NGS) é um processo de sequenciamento de DNA desenvolvido com o objetivo de acelerar e baixar o custo do processo de sequenciamento (VARUZZA, 2013). Essa técnica permite obter informações genômicas, incluindo os genes e a fração repetitiva do genoma, como os elementos de transposição, sequências de DNA satélites e microssatélites (METZKER, 2010). Diante do exposto, neste trabalho teve por objetivo caracterizar a fração repetitiva do genoma de *H. speciosa*, no intuito de compreender o processo evolutivo da espécie e inferir sobre a expansão no tamanho do genoma e dinâmica evolutiva de DNA repetitivo, oferecendo contribuições genômicas para as mais diversas aplicações no melhoramento vegetal.

## 2. REVISÃO DE LITERATURA

### 2.1 Origem, distribuição e características botânicas de *Hancornia speciosa* Gomes

*Hancornia speciosa* Gomes pertence à família Apocynaceae a qual encontra-se distribuída principalmente em regiões tropicais e subtropicais do mundo (MONACHINO, 1945). No Brasil, a família é composta por cerca de 90 gêneros e 850 espécies, habitando vários tipos de vegetação. Dentro desta grande família botânica, o gênero *Hancornia* é considerado monotípico, composto apenas pela espécie *H. speciosa* Gomes (BASTOS, *et al.*, 2017).

A espécie *H. speciosa*, conhecida popularmente como mangabeira, é uma árvore neotropical, nativa do Brasil e de ampla distribuição, tem ocorrência em locais de vegetação aberta, especialmente na Caatinga, no Cerrado, nos Tabuleiros Costeiros e nas Restingas, havendo relatos de sua presença na região Amazônica (LIMA & SCARIOT, 2010). Conforme Collevatti *et al.* (2016) essa espécie ocorre do Nordeste ao Centro-Oeste do Brasil e também no Paraguai, Bolívia e Peru.

A mangabeira é uma árvore perene, com porte variando de 4 a 7 m de altura, podendo chegar até 15 m. A raiz é pivotante e profunda, toda a planta exsuda látex de cor branca ou róseo-pálida. Possui folhas simples, opostas, ovaladas ou lanceoladas e pecioladas. Suas flores são hipocrateriformes, possui tubo floral relativamente longo e estreito, de coloração branca, campanuladas e aromáticas. O fruto é uma baga elipsóide e carnosa com número de sementes variável, coloração amarela ou esverdeada com manchas avermelhadas. As sementes são recalcitrantes, discóides, achatadas, rugosas e de coloração castanho-clara com hilo central (MONACHINO, 1945).

Geralmente suas flores surgem de agosto a novembro e os frutos de outubro a maio, porém, as épocas de floração e frutificação são irregulares e a variação pode acontecer conforme a época do ano, em árvores de lugares diferentes e ainda em árvores presentes em um mesmo local (LIMA & SCARIOT, 2010). A época de colheita dos frutos varia de acordo com a região, no norte de Minas Gerais, a atividade é realizada nos meses de outubro a janeiro. Em Sergipe há duas safras para colheita, a qual é realizada de dezembro a abril para a safra de verão e de maio a julho para safra de inverno, na Bahia esta atividade é feita de novembro a abril (LIMA & SCARIOT, 2010).

O tamanho e o peso dos frutos bem como a produção de frutos por planta são bastante variados, podendo chegar, neste último caso, a 800 frutos por planta ao ano (LIMA & SCARIOT, 2010). A árvore de mangaba desenvolve-se bem em solos ácidos e pobres, apresentando boa tolerância aos períodos de seca (LIMA & SCARIOT, 2010) e

possui desenvolvimento ótimo sob temperaturas variando de 24 a 26 °C e índice pluviométrico anual de 750 a 1600 mm (SOARES *et al.*, 2006).

De acordo com as diferenças morfológicas de suas folhas e flores, foram descritas seis variedades botânicas da espécie, sendo estas: *H. speciosa* var. *speciosa*, *H. speciosa* var. *maximiliani*, *H. speciosa* var. *cuyabensis*, *H. speciosa* var. *lundii*, *H. speciosa* var. *gardneri* e *H. speciosa* var. *pubescens* (MONACHINO, 1945). A variedade *H. speciosa* var. *speciosa* é a mais abrangente no Brasil e ocorre desde a região Sudeste até a região Norte do país. As demais variedades encontram-se restritas em algumas regiões do Cerrado. A variedade *lundii* ocorre nos Estados de Minas Gerais, Pernambuco, Bahia e Goiás, a *pubescens* em Goiás e Minas Gerais, a *gardneri* no Brasil Central, a *cuyabensis* em Mato Grosso, mais especificamente na Chapada dos Guimarães e a *maximiliani* em Minas Gerais (MONACHINO, 1945). A diferenciação morfológica na espécie sugere restrição de longo prazo no fluxo gênico e adaptação ecológica ou mutação de caracteres decorrentes de um pequeno número de genes (COLLEVATTI *et al.*, 2016).

Quanto ao sistema reprodutivo, *H. speciosa* é uma planta alógama, que possui flores hermafroditas e é auto-incompatível necessitando, portanto, de agentes polinizadores e diferentes genótipos para que ocorra a fecundação cruzada e a produção dos frutos (COLLEVATTI *et al.*, 2016). A polinização é entomófila, geralmente realizada por mariposas e abelhas (LIMA & SCARIOT, 2010).

Estudos recentes envolvendo hibridação entre variedades de *H. speciosa* demonstraram a capacidade de troca de pólen entre as diferentes variedades, indicando que a diferenciação morfológica intervartietal não ocorreu devido a fatores genéticos (COLLEVATTI *et al.*, 2016). Além disto, os autores observaram que o cruzamento entre populações parentais não é aleatório e que apesar da divergência existente entre variedades e híbridos, os mesmos não apresentaram fenômenos de heterose ou depressão endogâmica, sugerindo assim que as diferenças morfológicas presentes são resultado de adaptação ecológica e deriva genética na espécie (COLLEVATTI *et al.*, 2016).

## **2. 2 Importância e vulnerabilidade da espécie**

Os frutos da mangabeira são consumidos *in natura* e principalmente utilizados em indústrias de processamento de alimentos para fabricação de polpas, geleias, doces, entre outros (LIMA & SCARIOT, 2010). A fruta é tipicamente tropical e altamente apreciada pelos consumidores devido às suas características organolépticas e ao valor nutricional (Silva *et al.*, 2016). Ressalta-se que os frutos possuem elevados teores de ferro



e proteína além de ser rico em vitamina C e ácido ascórbico (SANTOS *et al.*, 2017).

As folhas de *H. speciosa* são tradicionalmente usadas para tratar diabetes no Brasil, as quais apresenta potencial anti-diabético devido à redução na concentração de glicose no sangue por meio da inibição da  $\alpha$ -glucosidase intestinal e estimulação da absorção de glicose por adipócitos (PEREIRA *et al.*, 2014). Além disto, suas folhas também são utilizadas no controle da pressão arterial (SILVA *et al.*, 2016). O látex exsudado da árvore de mangaba é amplamente utilizado na medicina tradicional para tratar uma variedade de doenças, incluindo diarreia, úlcera, gastrite, tuberculose, acne e verrugas (RIBEIRO, *et al.*, 2016).

A coleta dos frutos é feita por agroextrativismo, atividade esta que tem uma forte influência econômica e social para muitas famílias na região Nordeste, que fazem do lucro obtido, componente indispensável de sua renda mensal (SCHMITZ *et al.*, 2009). Mais de 5000 famílias praticam a coleta extrativista dos frutos da mangabeira no Estado de Sergipe durante a temporada de frutos. A produção brasileira de mangaba ultrapassa 800 toneladas por ano (IBGE, 2016).

Devido à redução na área dos ecossistemas onde ocorre, a mangabeira vem sofrendo erosão genética no Brasil (ALMEIDA *et al.*, 2017). Fatores exógenos, como o intenso mercado de terras, a produção de monoculturas, a construção de infraestruturas turísticas e a agricultura contribuíram para a redução de áreas de ocorrência natural de mangabeira (ALMEIDA *et al.*, 2017).

### **2. 3 Recursos genéticos e melhoramento da espécie**

*Hancornia speciosa* ainda está em processo de domesticação, devido a isto, os objetivos do melhoramento genético da espécie ainda não são bem definidos, no entanto, como a principal finalidade econômica da mangabeira é a produção de frutos, a seleção de genótipos superiores às plantas naturais é uma finalidade já conhecida no melhoramento da espécie, buscando genótipos mais produtivos de boa qualidade e estabilidade na produção de frutos (GANGA *et al.*, 2010).

O estabelecimento de bancos de germoplasma é o primeiro passo para a implementação de um programa de melhoramento e um dos principais requisitos é a capacidade de identificar genótipos superiores e populações segregantes (SANTOS *et al.*, 2017). Tradicionalmente, a diversidade genética é medida pela análise de uma combinação de características morfológicas e agrônômicas, no entanto, *H. speciosa* é uma planta heterogênea com muitos atributos morfológicos sobrepostos. Além disso, muitas

características vegetativas são influenciadas por fatores ambientais e mostram variação contínua e alta plasticidade, o que dificulta a identificação de grupos taxonômicos discretos (JIMENEZ *et al.*, 2015).

Outro problema enfrentado quanto à conservação do germoplasma de mangaba é a baixa porcentagem de germinação e recalcitrância de suas sementes, fator limitante da produtividade e que influencia na conservação convencional e disponibilidade de material vegetal (PRUDENTE *et al.*, 2015). Estudos acerca da obtenção de explantes *in vitro* de *H. speciosa*, visando alcançar produção de mudas em larga escala, proporcionam importante suprimento para a micropropagação e conservação de germoplasma e vem sendo realizados com sucesso (PRUDENTE *et al.*, 2015)

As técnicas de biologia molecular constituem um conjunto de ferramentas importantes que permite identificar, caracterizar e mapeamento genético. Essas tarefas estão sendo realizadas com maior segurança e mais rapidamente devido ao avanço das tecnologias aplicadas à biologia (SANTOS *et al.*, 2017).

#### **2.4 Evolução do genoma vegetal e DNA repetitivo**

A análise genômica permite a quantificação de genomas (Valor 1C, medido em picograma- pg ou megabases – Mb, sendo 1 pg igual a 978 Mb) que constitui característica importante a ser levada em consideração em estudos de evolução. Sequências de DNA repetitivo são contribuintes importantes no aumento do tamanho genômico, estes elementos dispõem-se de forma dispersa ou em *tandem* (KAZAZIAN JÚNIOR, 2004).

A evolução na fração repetitiva dos genomas ocorre mais rapidamente do que nas regiões gênicas e a velocidade com que isto ocorre é proporcional a quão repetitiva determinada região está, ou seja, quanto maior a porcentagem de repetição de determinada sequência mais rapidamente ela evolui e maior é a heterogeneidade entre as sequências devido à mutação (MEHROTRA & GOYAL, 2014).

O reconhecimento das propriedades funcionais das sequências de DNA repetitivo é de interesse biológico significativo, pois essas sequências podem desempenhar papel importante na organização estrutural dos cromossomos, na regulação gênica e na dinâmica evolutiva do genoma (FESCHOTTE, 2008).

A fração repetitiva do genoma é distribuída em dois grupos: os elementos de transposição (TEs) e as repetições em tandem (*tandem repeats*), ambos se encontram

repetidos diversas vezes no genoma de forma dispersa e sequenciada, respectivamente (Mehrotra & Goyal, 2014). As repetições em *tandem* podem ser classificadas como microssatélites (tamanho de 2-5pb); minissatélites (tamanho de 6-100 pb) e os satélites (tamanho de 150-400pb) (BABENKO *et al.*, 2012; MEHROTRA & GOYAL, 2014).

Conhecer a origem evolutiva, a localização cromossômica e a organização genômica das sequências de DNA repetitivo, é importante para elucidar sua função e entender o envolvimento destas nos processos evolutivos dos genomas de eucariotos, sabendo que ambos estão diretamente relacionados com a estrutura e função do centrômero (PLOHL, 2010). Sendo que o centrômero é o local onde o cinetócoro se organiza para que as fibras do fuso conectem-se a ele promovendo assim a segregação durante a divisão e multiplicação celular (JIANG *et al.*, 2003).

#### **2.4.1 DNA satélite**

DNAs satélite (DNAsat) são sequências relacionadas, encontradas repetidas em *tandem* (em sequência) que possuem unidades de repetição (monômeros) (JURKA *et al.*, 2007). Apresentam tamanho variando de 150 a 400 pb, sendo mais comuns tamanhos de aproximadamente 170 pb (BABENKO *et al.*, 2012; MEHROTRA & GOYAL, 2014). Estes elementos encontram-se geralmente em regiões específicas que desempenham função nos cromossomos, como nos genes de DNA ribossomal, nos telômeros, subtelômeros e nos centrômeros (JIANG *et al.*, 2003). Segundo Khost *et al.* (2017) estas sequências apresentam papel potencial na estabilidade do genoma e segregação cromossômica.

O DNAsat é importante componente das regiões de heterocromatina dos genomas eucariotos que, geralmente, possuem dinâmica evolutiva elevada (SAMOLUK *et al.*, 2017). Sequências de DNA satélite apresentam variação na composição de nucleotídeos, porcentagem de representação no genoma e na distribuição cromossômica, mesmo em espécies relacionadas, sendo assim contribuem significativamente em estudos taxonômicos (PLOHL, 2010).

As sequências de DNA satélite divergem de forma gradual, principalmente devido ao acúmulo de substituições de nucleotídeos. Este tipo de sequência contribui para estudos taxonômicos, pois quando se trata de espécies relacionadas, o número de repetições entre indivíduos do mesmo grupo é diferente ainda que contenham o mesmo *repeat* (MEHROTRA & GOYAL, 2014).

## 2.4.2 DNA ribossomal

O esclarecimento de eventos evolutivos relativamente recentes bem como a diferenciação genética entre espécies fenotipicamente semelhantes recebe fortes contribuições do estudo do RNA ribossomal e suas regiões separadoras chamadas de DNA ribossomal (DNAr) (HILLIS & DIXON 1991). Em espécies vegetais, o DNA ribossomal (DNAr) é composto por regiões denominadas 5S e 45S que são separadas por espaçadores internos e intergênicos e transcritas pela RNA-polimerase III e RNA-polimerase I, respectivamente (ROA & GUERRA, 2015).

O DNAr 45S é formado por três genes (18S, 5.8S e 25S), dois espaçadores internos (ITS-1 e ITS-2) e um espaçador intergênico (IGS). Cada unidade de repetição do RNA ribossomal possui regiões transcritas, que compreendem o chamado espaçador transcrito externo ETS (*External Transcribed Spacer*), as regiões codificadoras dos três RNAs e os dois espaçadores internos ITS 1 (entre os genes 18S e 5.8S) e ITS 2 (entre o 5.8S e o 25S) e uma região não transcrita (região intergênica – IGS), localizada entre o terminal 3' do ETS próximo ao gene 25S e o terminal 5' do ETS ao lado do gene 18S (LONG & DAWID, 1980). Os sítios de DNAr são encontrados em posições terminais dos cromossomos e são bastante utilizados para estudos de reconstrução filogenética (ALMEIDA *et al.*, 2012; BAO *et al.*, 2010). As regiões gênicas 18S, 5.8S e 25S são altamente conservadas em todos os organismos, tanto em seu comprimento quanto em sua sequência nucleotídica (ALMEIDA *et al.*, 2012).

O DNA 5S é formado por uma unidade altamente repetitiva e conservada que contém centenas de unidades 5S arranjadas em *tandem*, sendo assim, mais repetitivo que o DNAr 45S. Esta sequência pode fornecer bons modelos para o entendimento da natureza evolutiva em *loci* compostos por unidades repetitivas arranjadas em *tandem* (SASTRI *et al.*, 1992). Cada unidade de repetição do DNAr 5S consiste de uma região codificante de 120 pb, que ao contrário dos outros RNA ribossomais é transcrita fora do nucléolo, sendo flanqueada por um espaçador não transcrito (NTS) de 100 a 700 pb. Assim como no DNAr 45S, a região codificante do DNAr 5S é altamente conservada e a espaçadora é variável entre espécies (EICKBUSH & EICKBUSH, 2007).

## 2.4.3 Elementos transponíveis

Elementos transponíveis (TEs) são sequências de DNA móveis que desempenham papéis na regulação de genes e tem potencial para influenciar a expressão de genes próximos,

funcionando como sequências reguladoras (KARAKULAH & SUNER, 2017). Desde a sua descoberta em 1956, por Bárbara McClintock em estudos no milho, os elementos móveis têm sido encontrados em quase todos os organismos e atualmente é sabido que TEs são os principais representantes do genoma de eucariotos (GOGVADZE & BUZDIN, 2009).

Durante muito tempo os elementos transponíveis (TEs), assim como todo DNA repetitivo, foram considerados “DNA lixo” entretanto, trabalhos realizados na última década identificaram papéis positivos para TEs na geração de variabilidade genômica e diversidade em plantas. Em particular, estudos recentes sugeriram que as alterações epigenéticas induzidas por TE e a modificação da expressão gênica contribuem para a variação fenotípica e adaptação a ambientes desfavoráveis e/ou aos estresses. Essas descobertas levaram TEs (antigo DNA lixo) a serem conhecidos como fontes de controle e diversidade genômica (SONG & CAO, 2017).

As mutações geradas por TEs, como qualquer outro tipo de mutação que ocorrem no genoma, podem ser prejudiciais, neutras ou benéficas. Os TEs demonstraram causar instabilidade genômica e, de fato, sua disseminação descontrolada pode levar à morte do indivíduo. Por esse motivo, existem vários mecanismos para controlar a expressão e transposição da TEs no genoma, estes incluem a colocação de modificações de histonas repressivas, remodelação de cromatina e metilação do DNA, que ocasionam silenciamento da expressão de TE. No entanto, esses mecanismos de repressão não são 100% eficientes e alguns TEs permanecem ativos no genoma (HORTVÁTH *et al.*, 2017)

O grande potencial evolutivo e adaptativo das angiospermas é devido à ação dos elementos transponíveis que resulta em uma ampla variedade de mudanças genéticas incluindo modificações de genes, padrões de herança alterados e duplicações (OLIVER *et al.*, 2013). Os TEs constituem frações de DNA que podem se mover no genoma usando um intermediário de DNA (transposons) ou um intermediário de RNA (retrotransposons) (GALINDO-GONZÁLEZ *et al.*, 2017).

Em escala crescente de especialização, estes elementos são classificados em Classe, Subclasse, Ordem, Subordem, Superfamília, Família e Subfamília. O nível mais elevado (Classe) separa os elementos em relação à presença ou ausência de um RNA intermediário no processo de transposição, as classes subdividem elementos que se autocopiam daqueles que deixam o local de origem e se integram em outra região (autônomos e não autônomos). A classificação quanto à ordem os diferencia de acordo com o mecanismo de inserção, já as superfamílias são divididas por características como estrutura das proteínas e presença e

tamanho de TSD (Target Site Duplication), que consiste na duplicação de uma sequência genômica curta no sítio de ligação do elemento. As famílias são classificadas quanto a conservação das sequências de DNA e as subfamílias com base em dados filogenéticos podendo ainda distinguir elementos autônomos de não autônomos (WICKER *et al.*, 2007).

Os TEs são atualmente conhecidos como potentes agentes de adaptação ofertando suas contribuições através de um processo conhecido por exaptação em que, os elementos que se replicam no genoma, transformam-se em novos genes hospedeiros conferindo ao organismo benefícios fenotípicos (JOLY-LOPEZ *et al.*, 2016). O processo de exaptação (também conhecido como co-opção ou domesticação molecular) consiste na utilização de uma sequência, ou característica estrutural, para uma função diferente da qual foi desenvolvida através do processo de seleção natural, em outras palavras, a exaptação ocorre quando um TE originalmente conservado através da seleção auto-replicativa transita para aumentar a aptidão do organismo e se conserva através da seleção fenotípica (FESCHOTTE & PRITHAM, 2007). Os TEs contribuem para a evolução genômica através de diversos mecanismos principalmente em relação a inativação e/ou alteração da expressão gênica (SEIDL & THOMMA, 2017).

#### **2.4.3.1 Retrotransposons**

Retrotransposon, ou elementos classe I, são sequências móveis que se locomovem a partir de um RNA intermediário. A transposição do elemento ocorre inicialmente com a transcrição completa do elemento, originando um RNA, que por ação da enzima *transcriptase reversa* (RT) produz uma cópia de DNA genômico que é integrada ao genoma por ação da enzima *integrasse* (INT), desta forma ao final de cada ciclo de transcrição uma nova cópia é produzida (KAZAIAN JÚNIOR, 2004). Essa característica faz com que os retrotransposons sejam os componentes de elevada importância para o aumento do tamanho dos genomas. Estes elementos podem ser classificados de acordo com a presença de sequências longas terminais presentes em suas extremidades (*Long Terminal Repeat* -LTR), ou pela ausência dessa sequência (*Non-Long Terminal Repeat* – Non LTR) (KAZAIAN JÚNIOR, 2004; WICKER *et al.*, 2007).

Os retrotransposons da ordem LTR são os principais responsáveis pelo paradoxo do valor C, no qual o tamanho do genoma não está correlacionado com a complexidade fisiológica do organismo (FESCHOTTE, 2008). Estes elementos estão intimamente

relacionados com os retrovírus, acreditando-se que estes organismos podem ter evoluído a partir de retrotransposons LTR *Gypsy* que adotaram um estilo de vida viral através da aquisição de uma proteína de envelope (ENV) e um conjunto de proteínas adicionais e sequências reguladoras. Retrovírus também podem ser transformados em retrotransposons a partir da inativação ou deleção de domínios que permitem sua mobilidade celular (WICKER *et al.*, 2007).

Na estrutura dos retrotransposons LTR, existe um complexo de genes responsáveis por sua replicação e inserção (genes *gag* e *pol* que codificam um revestimento de partículas virais (*gag*) e uma transcriptase inversa (RT), a ribonuclease H (RH), e a integrase (INT) para fornecer atividades enzimáticas e produzir o cDNA a partir do RNA e inserindo-o no genoma) (KAZAZIAN JÚNIOR, 2004). Os elementos Ty1-*Copia* e os elementos Ty3-*Gypsy* são as duas principais superfamílias de retrotransposons LTR (WICKER *et al.*, 2007). Estas superfamílias são os componentes TE mais abundantes dos genomas de plantas (GOGVADZE & BUZDIN, 2009). Eles podem ser distinguidos uns dos outros pela organização de seus domínios internos, que codificam proteínas necessárias para a retrotransposição (GALINDO-GONZÁLEZ *et al.*, 2017). Uma vez que as LTRs são recriadas durante o ciclo de retrotransposição, as LTRs em cada extremidade de uma cópia recém-inserida são idênticas, assim as variações entre LTRs podem ser usadas como um relógio molecular para determinar o tempo desde a inserção (assumindo uma taxa de mutação constante) (GALINDO-GONZÁLEZ *et al.*, 2017).

Os retrotransposons Non-LTR, podem ser divididos em duas subclasses: Non-LTR autônomos (*Long Interspersed Nucleotide Elements* – LINES), que codificam proteínas para sua transposição e Non-LTR não autônomos (*Short Interspersed Nuclear Elements* – SINES) que não codificam as proteínas necessárias à sua transposição e são então dependentes de Non-LTR autônomos para se transpor (KAZAZIAN JÚNIOR, 2004).

#### **2.4.3.2 Transposons**

Os transposons, ou elementos classe II, são sequências repetitivas móveis que se locomovem de forma dispersa por mecanismos variados. O mecanismo mais comum de transposição destes elementos é o chamado “recorta e cola”, em que a sequência é excisada de um local para o outro no genoma, desta forma o processo, que é não replicativo, consiste na codificação de uma proteína (*transposase*) que identifica a sequência da região terminal

invertida (TIRs) do TE para recortá-la e inserir em um novo local do genoma (LEVIM & MORAM, 2011). Além deste, outros mecanismos de transposição são conhecidos como, por exemplo, o de “copia e cola” em que a maquinaria celular promove a duplicação da sequência e as cópias geradas são inseridas em novos locais do genoma (KAZAZIAN JÚNIOR, 2004; WICKER *et al.*, 2007).

A determinação do local que o transposon será inserido é realizada pela enzima *integrase* (INT) e a integração de novas sequências é limitada a um pequeno número de nucleotídeos e dinucleotídeos que ocorre na proximidade de inserções parentais, processo este denominado “salto local”. Embora a maioria das transposições ocorra em sítios genômicos menores do que 100 kb de seu sítio original, alguns transposons são capazes de realizar saltos mais distantes (KAZAZIAN JÚNIOR, 2004).

As atividades robustas de pequenos transposons na expressão gênica provavelmente proporcionam resposta rápida a condições ambientais desvantajosas e, portanto, podem representar um modo regulatório avançado na evolução e adaptação de espécies (SONG & CAO, 2017). Isto é evidenciado pela observação de que plantas superiores acumularam mais MITEs (Elementos transposíveis de repetição invertida em miniatura) do que plantas inferiores (CHEN *et al.*, 2014). Esta característica indica que estes elementos constituem uma forma rápida de evolução genômica. Comparado com transposons de DNA autônomas, MITEs são geralmente sequências curtas (<800 pb) e não codificam *transposases* (CHEN *et al.*, 2014).

### **2.3 Aplicações do estudo de DNA repetitivo no melhoramento de plantas**

Em plantas, os elementos transponíveis são contribuintes importantes para diversidade e evolução, e existem muitos casos onde se constata que TEs atuam na alteração de genes ou geração de novos, causando conseqüentemente alteração genotípica e fenotípica ampliando assim a variabilidade genética do organismo, sob a qual melhoristas podem atuar selecionando genótipos de interesse (OLIVER *et al.*, 2013).

Elementos transponíveis, geralmente, permanecem silenciados no genoma por mecanismos epigenéticos como a metilação de DNA, no entanto, estes elementos podem ativados sob condições de estresse (SEIDL & THOMMA, 2017). Exemplo disto é o transposon *mPING* no arroz é ativado em resposta a estresses salinos (YASUDA *et al.*, 2013). O raciocínio por trás dessa hipótese é que a ativação de TEs levaria a um aumento na taxa de



mutação gerando variabilidade sobre a qual a seleção pode atuar. Além disso, sequências de TE reguladoras de resposta ao estresse podem ser distribuídas ao longo do genoma, ajudando na reprogramação gênica em resposta ao estresse (COWLEY & OAKEY, 2013).

Em casos particulares, os elementos transponíveis podem contribuir na adaptação e evolução do genoma como a inserção de um transposon no primeiro íntron do *Flowering Locus C* (FLC) em *Arabidopsis thaliana* (L.) Heynh, ocasionando florescimento precoce e dispensando a necessidade de vernalização (STRANGE *et al.*, 2011). Em soja, interrupções causadas por TEs no gene *GmphyA2* que codifica o fitocromo A está associada a adaptação da espécie a altitudes elevadas (KANAZAWA *et al.*, 2009)

Os TEs podem ser utilizados como marcadores moleculares em estudos de mapeamento e associação no genoma (LAI *et al.*, 2017). Polimorfismos de inserção de elementos transponíveis tem sido utilizado com êxito em estudos de estrutura populacional e análises de associação de características de interesse no milho (LAI *et al.*, 2017). Estudos mostram uma associação altamente significativa entre um marcador de transposon MITE e o tempo de floração em uma região nunca mapeada antes com uso de QTL - *Quantitative trait loci* (ZERJAL *et al.*, 2012).

Pesquisas demonstram que os polimorfismos ocasionados por inserções de TE em milho afetam diretamente a expressão fenotípica (LAI *et al.*, 2017). Uma inserção de TE classe II de tipo CACTA mostrou silenciar a expressão do gene *ZmCCT*, ocasionando redução da sensibilidade ao fotoperíodo e contribuindo para a adaptação do milho aos ambientes de dias longos (YANG *et al.*, 2013).

A relação entre TEs e regulação da expressão gênica é complexa devido ao grande número e tipos de TE's em todo o genoma e a sua capacidade de regular não só sua expressão, mas também a de genes próximos (HORTVÁTH *et al.*, 2017). O elevado número de dados disponibilizados pelas técnicas de sequenciamento de nova geração (NGS) permite analisar e compreender melhor a atividade dos elementos transponíveis e suas contribuições passivas dentro do genoma (SEIDL & THOMMA, 2017).

## **2. 4 Next-Generation Sequencing (NGS)**

As plataformas NGS produzem fragmentos mais curtos que o sequenciamento Sanger, porém com número de leituras muito maior podendo chegar a seis bilhões em um único ciclo

de sequenciamento. As quatro plataformas comercialmente disponíveis de sequenciamento de nova geração são Roche/454, Illumina, Applied Biosystems Solid, e a recentemente lançada Helicos HeliScope, que produzem enorme quantidade de sequências curtas a uma taxa de transferência bem mais elevada do que o antigo método de Sanger (MOROZOVA *et al.*, 2009).

O sequenciamento na plataforma Solexa (Illumina), é realizado por síntese usando DNA polimerase e nucleotídeos terminadores marcados com diferentes fluoróforos, seguindo o método de Sanger, seu diferencial consiste na clonagem *in vitro* dos fragmentos em uma plataforma sólida de vidro. A superfície de clonagem (*flow cells*) é dividida em linhas, em cada linha, adaptadores são fixados à superfície pela extremidade 5', deixando a extremidade 3' livre para servir na iniciação da reação de sequenciamento dos fragmentos imobilizados no suporte por hibridização (CARVALHO & SILVA, 2010).

Dados oriundos de NGS vem sendo utilizados em trabalhos que visam compreender melhor os efeitos do DNA repetitivo nos genomas (VITTE *et al.*, 2014; KHOST *et al.*, 2017). A elevada quantidade de dados disponibilizados pelas técnicas de sequenciamento de nova geração necessita do auxílio de tecnologias de bioinformática para que os resultados sejam melhor analisados e interpretados. O software Repeat Explorer é amplamente utilizado em estudos de DNA repetitivo atuando no agrupamento das sequências disponíveis (dados brutos) de acordo com parâmetros definidos, formando *clusters* e permitindo uma melhor análise da fração repetitiva dos genomas (NOVÁK *et al.*, 2010).

### 3. REFERÊNCIAS

- ALMEIDA, C.; FONSECA, A.; SANTOS, K. G. B.; MOSIOLEK, M.; PEDROSA-HARAND, A. Contrasting evolution of a satellite DNA and its ancestral IGS DNAr in *Phaseolus* (Fabaceae). **Genome**, Ottawa, v. 55, p.1-7, 2012
- ALMEIDA, V. M.; LUZ, G. A.; MARTINS, P. P.; GOMES, M. F. C.; COSTA, M.F.; LIMA, P. S. C.; VALENTE, S. E. S. Comparasion of eight methods to isolate genomic DNA from *Hancornia speciosa*. **Genetics and Molecular Research**, Ribeirão Preto, v. 16, n.3, p. 1-7, 2017.
- AMORIM, J. A. E., MATA, L. R.; LÉDO, A. S.; AZEVEDO, V. C. R.; SILVA, A. V. C. Diversity and genetic structure of mangaba remnants in states of northeastern Brazil. **Genetics and Molecular Research**, Ribeirão Preto, v. 14, p. 823-833, 2015.
- BABENKO, V. N.; KUTASHEV, K. O.; MATVIENKO, V. F.; Nucleosome organization in plant DNA satellite sequences. **Russian Journal of Genetics: Applied Research**, v. 2, p. 271-278, 2012.
- BAO, Y.; WENDEL, J. F.; GE, S. Multiple patterns of rDNA evolution following polyploidy in *Oryza*. **Molecular Phylogenetic and Evolution**, Orlando, v. 55, p. 136-142, 2010.
- BASTOS, K. X.; DIAS, C. N.; NASCIMENTO, Y. M.; SILVA, M. S.; LANGASSNER, S. M. Z.; WESSJOHANN, L. A.; TAVARES, J. F. Identification of phenolic compounds from *Hancornia speciosa* (Apocynaceae) leaves by UHPLC orbitrap-HRMS. **Molecules**, Basel, v. 22, p. 143, 2017.
- CARVALHO, M. C. G; SILVA, D. C. G. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. **Ciência Rural**, Santa Maria, v.40, n.3, p.735-744, 2010.
- CHEN, J.; HU Q.; ZHANG, H.; LU, C.; KUANG, H. P-MITE: A database for plant miniature inverted-repeat transposable elements. **Nucleic Acids Research**, Oxford v. 42, p. 176-181, 2014.
- COLLEVATTI, R. G.; OLIVATTI A. M.; TELLES, M. P. C.; CHAVES, L. J. Gene flow among *Hancornia speciosa* (Apocynaceae) varieties and hybrid fitness. **Tree Genetics & Genomes**, Berlin, v. 12 n. 74, p. 1-12, 2016.
- COWLEY, M.; OAKLEY, R. J. Transposable elements re-wireand find-tune the transcriptome. **PLOS Genetics**, San Francisco, v. 9, n. 1, p. e1003234, 2013.

- EICKBUSH, T. H.; EICKBUSH, D. G. Finely orchestrated movements: evolution of the ribosomal RNA genes. **Genetics**, Austin, v. 175, n. 2, p. 477-485, 2007.
- FESCHOTTE, C.; PRITHAN, E. J. DNA Transposon and evolution of eukaryotic genomes. **Annual Review of Genetic**, Palo Alto, v. 41, p. 331-368, 2007.
- FESCHOTTE, C. The contribution of transposable elements to the evolution of regulatory networks. **Nature Reviews Genetics**, London, v. 9, n. 5, p. 397-405, 2008.
- GALINDO-GONZÁLEZ, L.; MHIRI C.; DEYHOLOS, M. K.; GRANDBASTIEN, M. A. LTR retrotransposons in plants: Engines of evolution. **Gene**, Amsterdam, p. 14-25, 2017.
- GANGA, R. M. D.; FERREIRA, G. A.; CHAVES, L. J.; NAVES, R. V.; NASCIMENTO, J. L. Caracterização de frutos e árvores de populações naturais de *Hancornia speciosa* Gomes do Cerrado. **Revista Brasileira de Fruticultura**. Jaboticabal, v. 32, n.1, p.101-113, 2010.
- GOGVADZE E.; BUZDIN, E. A. Retroelements and their impact on genome evolution and functioning. **Cellular and Molecular Life Sciences**, Basel, v.66, p.3727–3742, 2009.
- HILLIS, D. M.; DIXON, M. T. Ribosomal DNA: Molecular evolution and phylogenetic inference. **The Quarterly Review of Biology**, New York, v.66, p.411-453, 1991.
- HORVÁTH, V.; MERENCIANO, M.; GONZÁLEZ, J. Revisiting the relationship between transposable elements and the eukaryotic stress response. **Trends in Genetics**, Amsterdam, v. 33, n. 11, p. 832-841, 2017.
- IBGE, 2016. Instituto Brasileiro De Geografia E Estatística. <<https://censos.ibge.gov.br/agro/2017/en/agencia-press-room/2185-news-agency/releases-en/17008-pevs-2016-silviculture-and-wild-crop-harvesting-reach-r-18-5-billion.html>> (Acesso em maio de 2017).
- JANGAM, D.; FESCHOTTE, C.; BETRÁN, E. Transposable elements domestication as an adaptation to evolutionary conflicts. **Trends in Genetics**, Amsterdam, v. 33, n. 11, p. 817-831, 2017.
- JIANG, J.; BIRCHLER, J. A.; PARROT, W. A.; DAWE, R. K. A molecular view of plants centromeres. **Trends in Plant Science**, Oxford v.8, n. 12, p. 570-575, 2003.
- JIMENEZ, H. J.; MARTINS, L. S. S.; MONTARROYOS A. V. V.; SILVA-JÚNIOR, J. F.; ALZATE-MARIN, A. L.; MORAES-FILHO, R.M. Genetic diversity of the neotropical tree *Hancornia speciosa* Gomes in natural populations in Northeastern Brazil. **Genetic and Molecular Research**, Ribeirão Preto, v. 14, n. 4, p. 17749-17757, 2015.
- JOLY-LOPEZ, Z.; HOEN, D. R.; BLANCHETTE, M.; BUREAU, T. Phylogenetic and

genomic analyses resolve the origin of important plant genes derived transposable elements. **Molecular Biology and Evolution**, Chicago, v. 33, n. 8, p. 1937-1956, 2016.

JURKA, J.; KAPITONOV, V. V.; KOHANY, O.; JURKA, M. V. Repetitive sequences in complex genomes: structure and evolution. **Annual Review of Genomics and Human Genetics**, Palo Alto, v. 8, p. 241-259, 2007.

KANAZAWA, A.; LIU, B.; KONG, F.; ARASE, S.; ABE, J. Adaptative evolution involving gene duplication and insertion of a novel Ty1/copia retrotransposon in soybean. **Journal of Molecular Evolution**, New York, v. 69, n. 2, p. 164-175, 2009.

KARAKULAH, G.; SUNER, A. PlanTEnrichment: a tool for enrichment analysis of transposable elements in plants, **Genomics**, San Diego, v. 109, p. 336-340, 2017.

KAZAZIAN JÚNIOR, H. H. Mobile elements: drivers of genome evolution. **Science**, Washington, v. 303, p. 1626-1632, 2004.

KHOST, D. E.; EICKBUSH, D. G.; LARRACUENTE, A. M. Single-molecule sequencing resolves the detailed structure of complex satellite DNA loci in *Drosophila melanogaster*. **Genome Research**, New York, v. 21, p. 709-721, 2017.

LAI, X.; SCHANBLE, J. C.; LIAO, Z.; XU, J.; ZHANG, G.; CHUAN, LI.; HU, E.; RONG, T.; XU, Y.; LU, Y. Genome-wide characterization of non-reference transposable element insertion polymorphisms reveals genetic diversity in tropical and temperate maize. **BMC Genomics**, London, v. 18, n. 1, p. 702, 2017.

LEVIN, H.; MORAN, J. V. Dynamic interaction between transposable elements and their hosts, **Nature Reviews Genetics**, London, v. 12, n. 9, p. 615-627, 2011

LIMA, I. L. P.; SCARIOT A. Boas práticas de manejo para o extrativismo sustentável da mangaba. Brasília –DF: **Embrapa recursos genéticos e biotecnologia**, p. 1-68, 2010.

METZKER, M. L. Sequencing technologies – the next generation. **Nature Reviews Genetics**. Houston, v. 11, p. 31-46, 2010.

MONACHINO, J. A revision of *Hancornia* (Apocynaceae). **Lilloa**, Tucumán, v.11, p.19-48, 1945.

MOROZOVA, O.; HIRST, M.; MARRA, M. A. Applications of new sequencing technologies for transcriptome analysis. **Annual Review of Genomics and Human Genetics**, v. 10, p. 135-1351, 2009.

MOURA, N. F.; CHAVES, L. J.; VENKOVSKY, R.; NAVES, R. V. N.; AGUIAR, A. V.;

- MOURA, M. F. Genetic structure of manbaga (*Hancornia speciosa* Gomes) populations in the cerrado region of central Brazil. **Bioscience Journal**, Uberlândia, v. 27, n.3, p.473-481, 2011.
- NABOUT, J. C.; MAGALHÃES, M. R.; GOMES, M. A. A.; CUNHA, H. F. The impact global climate change on the geographic distribution and sustainable harvest of *Hancornia speciosa* Gomes (Apocynaceae) in Brazil. **Environmental Management**, New York, v. 57, p. 814-821, 2016.
- NOVÁK, P.; NEUMANN, P.; MACAS, J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. **BMC Bioinformatics**, London, v. 11, p. 1-12, 2010.
- OLIVER, K. R.; MCCOMB, J. A.; GREENE, W. K. Transposable elements: powerful contributors to angiosperm evolution and diversity. **Genome Biology and Evolution**, Oxford, v. 5, n. 10, p. 1889-1901, 2013.
- PEREIRA, A. C.; PEREIRA, A. B.; MOREIRA, C. C. L.; BOTION, L. M.; LEMOS, V. S.; BRAGA, F. C.; CORTES, S.F.; *Hancornia speciosa* Gomes (Apocynaceae) as a potential anti-diabetic drug. **Journal of Ethnopharmacology**, Lausanne, v. 162, p. 30-36, 2015.
- PLOHL, M. Those mysterious sequences of satellite DNAs. **Periodicum Biologorum**, Zagreb, v. 112, p. 403-140, 2010.
- PRUDENTE, D. O.; PAIVA, R.; NERY, F. C.; REIS, M. V.; PAIVA, P. D. O.; SILVA, L. C. Encapsulation of lateral buds of *Hancornia speciosa*. **Acta Horticulturae**, The Hague, v. 1155, p. 59-64, 2015.
- RIBEIRO, T. P.; SOUSA, T. R.; ARRUDA, A. S.; PEIXOTO, N.; GONÇALVES, P. J.; ALMEIDA, L. M. Evaluation of citotoxicity and genotoxicity of *Hancornia speciosa* látex in *Allium cepa* root model. **Brazilian Journal of Biology**, São Carlos, v. 76, n. 1, p. 245-249, 2016.
- ROA, F.; GUERRA, M. Non-random distribution of 5S rDNA sites and its association with 45S rDNA in plant chromosomes. **Cytogenetic and Genome Research**, Basel, v.146, p.243–249. 2015.
- RUIZ-RUANO, F. J.; LÓPEZ-LEÓN, M. D.; CABRERO, J.; CAMACHO, J. P. M. High-throughput analysis of the satellitome of illuminates satellite DNA evolution. **Scientific Reports**, Tokyo, v.6, p. 28333, 2016.
- SAMOLUK, S. S.; ROBLEDO, G.; BERTIOLI, D.; SEIJO, J. G. Evolutionary dynamics of an at-rich satellite DNA and its contribution of karyotype differentiation in wild diploid *Arachis*

- species. **Molecular Genetic and Genomics**, Berlin v. 292, n. 2, p. 283-296, 2017.
- SANTOS, P. S.; FREITAS, L. S.; SANTANA, J. G. S.; MUNIZ, E. N.; RABBANI, A. R. C.; SILVA, A. V. C. Genetic diversity and quality of mangabeira tree fruits (*Hancornia speciosa* Gomes – Apocynaceae) a native species from Brazil. **Scientia Horticulturae**, Amsterdam, v. 226, p. 372-378, 2017.
- SASTRI, D. C.; HILU, K.; APPELS, R.; LAGUDAH S.; PLAYFORD, J.; BAUM, B. R. An overview of evolution in plant 5S DNA. **Plant Systematics and Evolution**, New York, v. 183, p. 169-181, 1992.
- SEIDL, M. F.; THOMMA, B. P. H. J. Transposable elements direct the coevolution between plants and microbes. **Trends in Genetics**, Amsterdam, v. 33, n. 11, p. 842-851, 2017.
- SILVA, G. C.; BRAGA, F. C.; LEMOS, V. S.; CORTES, S. F. Potent antihypertensive effect of *Hancornia speciosa* leaves extract, **Phytomedicine**, Jena, v. 23, n. 2, p. 214-219, 2016.
- SOARES, F. P.; PAIVA, R.; CRAVO, N. R.; OLIVEIRA, L. M.; SILVA, D. R. G.; PAIVA, P. D. O. Cultura da mangaba (*Hancornia speciosa* Gomes). **Boletim Agropecuário**, Lavras, n. 67, p. 1-12, 2006.
- SONG, X.; CAO, X. Transposon-mediated epigenetic regulation contributes to phenotypic diversity and environmental adaptation in rice. **Current Opinion in Plant Biology**, London, v. 36, p. 111-118, 2017.
- SOTERO-CAIO, C. G.; CABRAL-DE-MELLO, D. C.; CALIXTO, M. S.; VALENTE, G. T.; MARTINS, C.; LORETO, V.; SOUZA, M. J.; SANTOS, N. Centromeric enrichment LINE-L1 retrotransposons and its significance for the chromossome evolution of phyllostomid bats. **Chromossome Research**, Oxford, v. 25, p. 313-325, 2017.
- STRANGE, A.; PEIJIN, L.; LISTER, C.; ANDERSON, J.; WARTHMAN, N.; SHINDO, C.; IRWIN, J.; NORDBORG, M.; DEAN, C. Major-effects alleles at relatively few loci underlie distinct vernalization and flowering variation in *Arabidopsis* accessions. **PLoS ONE**, San Francisco, v. 6, n. 5, p. e19949, 2011.
- VARUZZA, L. Introdução à análise de dados de sequenciadores de nova geração. v.2.0.1, 2013.
- VITTE, C.; FUSTIER, M. A.; ALIX, K.; TENAILLON M. I. The bright side of transposons in crop evolution. **Briefings in Functional Genomics**. Oxford, v. 13, n. 4, 276-295, 2014.
- WICKER, T.; SABOT F.; HUA-VAN, A.; BENNETZEN J. L.; COPY, O.; CHALHOUB B.; FLAVELL A.; LEROY, P.; MORGANTE, M.; PANAUD, O.; PAUX, E.; SANMIGUEL, P.;

SCHULMAN, A. H. A unified classification system for eukaryotic transposable elements. **Nature Reviews Genetics**, v. 8, p. 973-982, 2007.

YANG, Q.; LI, Z.; LI, W. KU, L.; WANG, C.; YE, J.; LI, K.; YANG, N.; LI, Y.; ZHONG, T.; LI, J.; CHEN, Y.; YAN, J.; YANG, X.; XU, M. CACTA-like transposable element in *ZmCCT* attenuated photoperiod sensitivity and accelerated the post domestication spread of maize. **Proceedings of the National Academy of Sciences**, Washington, v. 110, n. 42, p. 16969-16974, 2013.

YASUDA K.; ITO, M.; SUGITA, T.; TSUKIYAMA, T.; SAITO, H.; NAITO, K.; TERAISHI, M.; TANISAKA, T.; OKUMOTO, Y. Utilization of transposable element mPing as a novel genetic tool for modification of the stress response in rice. **Molecular Breeding**, Dordrecht, v. 32, n. 3, p. 505-516, 2013.

ZERJAL, T.; ROUSSELET, A.; MHIRI, C.; COMBES, V.; MADUR, D.; GRANDBASTIEN, M. A.; CHARCOSSET, A.; TENAILLON, M. I. Maize genetic diversity and association mapping using transposable element insertion polymorphisms. **Theoretical and Applied Genetics**, Berlin, v. 124, n. 8, p. 1521-1537, 2012.



---

**CAPÍTULO II – Tamanho genômico e identificação de sequências repetitivas no genoma de *Hancornia speciosa* Gomes (Apocynaceae: Gentianales)**

1 Tamanho genômico e identificação de sequências repetitivas no genoma de *Hancornia speciosa*  
2 Gomes (Apocynaceae: Gentianales)

3 **Short running title:** DNA repetitivo em *H. speciosa*

4 **Resumo**

5 A mangabeira é uma arbórea neotropical, nativa do Brasil, classificada na ordem Gentianales.  
6 Neste trabalho analisou-se a fração repetitiva do genoma de algumas espécies desta ordem, no  
7 intuito de estudar comparativamente, elementos repetitivos em *Hancornia speciosa*. Para tanto,  
8 foram usados *reads* de *H. speciosa*, provenientes de sequenciamento de nova geração (NGS) os  
9 quais foram obtidos a partir da biblioteca genômica *paired-end reads* e sequenciados usando a  
10 plataforma Illumina HiSeq2500. Os *reads* utilizados na análise das demais espécies foram  
11 oriundos do banco de dados online NCBI e a relação entre as espécies demonstrada por análise  
12 PCA, utilizando genomas plastidiais inteiros. As frações repetitivas dos genomas foram analisadas  
13 em *clusters* no software RepeatExplorer. O genoma de duas variedades de *H. speciosa*,  
14 quantificados usando citômetro de fluxo, apresentaram 0,88 e 0,87 pg (Valor 2C). Os padrões de  
15 repetição foram semelhantes entre *H. speciosa*, *C. roseus* e *R. stricta*. A análise de DNA repetitivo  
16 mostrou que os retrotransposons foram os mais representativos e os Transposons apresentaram  
17 menor proporção nas espécies analisadas. As espécies *H. speciosa* e *R. stricta* não apresentaram  
18 sequências de DNA satélite, sugerimos, portanto, que seus genomas, enriquecidos de LTR  
19 Ty3/Gypsy/Chromovírus, podem ter estes elementos atuando como DNA satélite.

20

21 **Palavras-chave:** mangabeira, NGS, retrotransposons, chromovirus, DNA satélite.

22

## Abstract

23 The mangabeira is a neotropical tree, native to Brazil, classified in the order Gentianales. In this  
24 work we have analyzed the repetitive genome of some species of this order, in order to study  
25 comparatively, repetitive elements in *Hancornia speciosa*. For this, reads of *H. speciosa*, from  
26 new generation sequencing (NGS) were used which were obtained from the paired-end reads  
27 genomic library and sequenced using the Illumina HiSeq2500 platform. The reads used in the  
28 analysis of the other species were from the NCBI online database and the relationship between  
29 the species demonstrated by PCA analysis using whole plastid genomes. Repetitive fractions of  
30 the genomes were analyzed in clusters in the RepeatExplorer software. The genome of two  
31 varieties of *H. speciosa*, quantified using flow cytometer, showed 0.88 and 0.87 pg (Value 2C).  
32 Repeat patterns were similar between *H. speciosa*, *C. roseus* and *R. stricta*. The repetitive DNA  
33 analysis showed that the retrotransposons were the most representative and the Transposons  
34 showed a lower proportion in the analyzed species. The species *H. speciosa* and *R. stricta* did not  
35 present satellite DNA sequences, therefore, we suggest that their genomes, enriched with LTR  
36 Ty3 / Gypsy / Chromovirus, may have these elements acting as satellite DNA.

37

38 **Keywords:** mangabeira, NGS, retrotransposons, chromovirus, satellite DNA.

## 39 **Introdução**

40 A mangabeira (*Hancornia speciosa* Gomes), classificada na família Apocynaceae e ordem  
41 Gentianales, é uma frutífera típica de regiões tropicais e subtropicais. No Brasil, essa espécie  
42 apresenta ampla distribuição, ocorrendo em locais de vegetação aberta, especialmente na  
43 Caatinga, no Cerrado, nos Tabuleiros Costeiros e nas Restingas (Lima and Scariot, 2010). Além  
44 do Brasil, Collevatti *et al.* (2016) reporta a ocorrência de *H. speciosa* no Paraguai, Bolívia e Peru.

45 Os frutos da árvore de mangaba apresentam os maiores teores de ferro e proteína além de  
46 serem ricos em vitamina C e ácido ascórbico (Santos *et al.*, 2017). São consumidos *in natura* e  
47 amplamente utilizados em indústrias de processamento de alimentos para fabricação de polpas,  
48 geleias, doces, entre outros (Lima and Scariot, 2010), As folhas da mangabeira apresentam  
49 propriedades medicinais tendo potencial anti-diabético (Pereira *et al.*, 2015) e utilização no  
50 controle da pressão arterial conforme Silva *et al.* (2016). O látex exsudado do caule é amplamente  
51 utilizado na medicina tradicional para tratar doenças como diarreia, úlcera, gastrite, tuberculose,  
52 acne e verrugas (Ribeiro *et al.*, 2016).

53 Práticas de monocultivo agrícola e construções de infraestruturas turísticas tem  
54 contribuído para a redução de áreas de ocorrência natural da mangabeira, ocasionando erosão  
55 genética na espécie (Almeida *et al.*, 2017). *Hancornia speciosa* ainda está em processo de  
56 domesticação, devido a isto, os objetivos do melhoramento genético na espécie não estão bem  
57 definidos. Entretanto, a principal finalidade econômica dessa espécie é a produção de frutos e a  
58 seleção de genótipos superiores; mais produtivos, com boa qualidade de fruto e estabilidade de  
59 produção é almejada em programas de melhoramento (Ganga *et al.*, 2010).

60 A genômica em plantas é uma área que permite conhecer as funções estruturais e  
61 funcionais do genoma por meio de técnicas como como DNA recombinante, sequenciamento de  
62 DNA e ferramentas computacionais essenciais para o melhoramento genético de plantas (Unamba  
63 *et al.*, 2015). Estas informações podem ser utilizadas no melhoramento genético de plantas por  
64 diversas finalidades tais como seleção assistida e associação genômica. Apesar de sua importância

65 e vulnerabilidade ainda não existem estudos genômicos em *H. speciosa*, conhecimentos neste  
66 âmbito apresentariam contribuições informativas que poderão ser utilizadas na aplicabilidade de  
67 técnicas tanto para o melhoramento genético da espécie, quanto para conservação de seus recursos  
68 genéticos.

69 O sequenciamento de nova geração (NGS) foi desenvolvido com o objetivo de acelerar e  
70 baixar o custo deste processo (Metzker, 2010). É uma técnica que permite obter informações  
71 genômicas, incluindo os genes e a fração repetitiva do genoma, elementos de transposição,  
72 sequências de DNA satélites e microssatélites (Metzker, 2010). Para organismos não modelos ou  
73 pouco estudados no âmbito científico, NGS permite obter dados para exploração de seus genomas.

74 A análise citometria de fluxo permite quantificar genomas (Valor 1C, medido em  
75 picograma- pg ou megabases – Mb, sendo 1 pg igual a 978 Mb) sendo esta, característica  
76 importante a ser levada em consideração em estudos que visam conhecer a composição genômica  
77 dos organismos. Dentre os diversos tipos de sequências presentes nos genomas, as referentes ao  
78 chamado DNA repetitivo são contribuintes importantes no aumento do tamanho genômico  
79 (Kazazian Júnior, 2004).

80 A fração repetitiva do genoma é distribuída em dois grupos: os elementos de transposição  
81 (TEs) e as repetições em tandem (*tandem repeats*), ambos se encontram repetidos diversas vezes  
82 no genoma de forma dispersa e sequenciada, respectivamente (Mehrotra and Goyal, 2014). Em  
83 escala crescente de especialização, elementos transponíveis são classificados em Classe,  
84 Subclasse, Ordem, Subordem, Superfamília, Família e Subfamília. O nível mais elevado (Classe)  
85 separa os elementos em relação a presença (retrotransposons) ou ausência (transposons) de um  
86 RNA intermediário no processo de transposição (Wicker *et al.*, 2007).

87 O reconhecimento das propriedades funcionais das sequências de DNA repetitivo é de  
88 interesse biológico, pois essas sequências podem desempenhar papel importante na organização  
89 estrutural dos cromossomos, na regulação gênica e na dinâmica evolutiva do genoma (Feschotte,  
90 2008). Para tanto são considerados como estruturas importantes para estudos de evolução de

91 espécies e estão cada vez mais envolvidos em pesquisas para diferentes aplicações, sendo  
92 atualmente utilizados como marcadores eficazes em estudos de associação genômica, técnica  
93 importante para o melhoramento genético de plantas (Zerjal *et al.* 2012). Neste trabalho foi  
94 quantificado e sequenciado parte do genoma de *H. speciosa* para analisar sua fração repetitiva e  
95 compará-la com espécies relacionadas, pertencentes a família Apocinácea, fornecendo as  
96 primeiras informações a nível de genoma em mangaba.

97

## 98 **Material e métodos**

### 99 **Citometria de fluxo**

100 Os genomas de duas variedades de *H. speciosa* (var. *speciosa* e var. *gardneri*),  
101 provenientes do Estado de Goiás (14° 3' 29" S; 47° 59' 19" W) e da Paraíba (7° 30' 53" S; 34° 53'  
102 07" W), foram medidos e quantificados por citometria de fluxo, utilizando suspensão de núcleos  
103 a partir de folhas jovens, preparada seguindo a descrição de Loureiro *et al.* (2007), usando tampão  
104 WPB. Um fluxo CyFlow SL cytometer (Partec, Görlitz, Alemanha) foi utilizado para estimar os  
105 tamanhos dos genomas. O conteúdo de DNA final para cada variedade foi calculado com base em  
106 três repetições para cada planta. Como controle, foram utilizadas folhas jovens de *Solanum*  
107 *lycopersium* (L.) (1C = 1,96 pg DNA) conforme Doležel and GreilHuber (2010). Os dados de  
108 tamanho dos genomas foram analisados em teste de média, usando 1C value, pelo teste de Tukey  
109 ao nível de 5% de significância. A análise estatística foi realizada na plataforma R (Team, 2016).  
110 A quantificação do genoma por citometria fluxo foi realizada no Laboratório de Citogenética da  
111 Universidade Federal de Pernambuco (UFPE).

112

### 113 **Material vegetal, isolamento de DNA e sequenciamento de DNA**

114 Foram coletados indivíduos de mangaba provenientes do município de Caaporã, Estado da  
115 Paraíba (7° 30' 53" S; 34° 53' 07" W), para posterior extração de DNA, a qual foi realizada na  
116 Universidade Federal de Alagoas (UFAL) – *Campus* Arapiraca, no Laboratório de Recursos

117 Genéticos. O procedimento foi efetuado utilizando aproximadamente 2 cm<sup>2</sup> (1g) de folhas  
118 (material vegetal vivo), pelo protocolo de Doyle and Doyle (1987), com modificações. O DNA  
119 extraído foi analisado por visualização em gel de agarose 1% e quantificado em  
120 espectrofotômetro. As amostras de DNA foram fragmentadas para obter sequências de  
121 aproximadamente 500-600 pares de base (pb) e a partir destas foi construída uma biblioteca para  
122 sequenciamento de nova geração (NGS). Os fragmentos foram ligados a adaptadores usando  
123 Illumina Nextera kit e sequenciados *paired-end reads* com tamanho de 100nt usando a plataforma  
124 Illumina HiSeq2500. O sequenciamento foi realizado no Laboratório Central de Tecnologias de  
125 Alto Desempenho em Ciências da Vida (LacTad) na Universidade Estadual de Campinas  
126 (Unicamp), Brasil.

127

## 128 **Análise filogenética**

129 Para que os resultados obtidos com *H. speciosa* fossem passíveis de comparações, foi  
130 construída uma árvore filogenética com base nos genomas plastidiais inteiros de espécies da  
131 ordem Gentianales, disponíveis no banco de dados online *National Center for Biotechnology*  
132 *Information – NCBI* (Tabela 1), a qual nos permitiu selecionar para este estudo as espécies:  
133 *Catharanthus roseus* (L.) G. Don, *Rhazya stricta* Decne (espécies classificadas subfamília  
134 Rauvolfioideae) e *Asclepias syriaca* L. (espécie mais divergente, classificada na subfamília  
135 Asclepiadoideae). A análise foi realizada utilizando a ferramenta Adegnet do software R (Team,  
136 2016).

137

## 138 **Análise da fração repetitiva dos genomas de espécies da ordem Gentianales usando os** 139 **softwares RepeatExplorer e TAREAN**

140 Os *reads* de *H. speciosa*, resultantes do sequenciamento foram usados para montar *clusters*  
141 utilizando um software livre chamado RepeatExplorer (Novák *et al.*, 2010), que engloba  
142 ferramentas que agrupam e caracterizam sequências repetitivas provenientes de dados de

143 sequenciamento de nova geração, gerando uma análise de *cluster*. Os *reads* das espécies *C. roseus*,  
144 *R. stricta* e *A. syriaca* utilizados neste trabalho, foram oriundos do NCBI e utilizados também na  
145 montagem de clusters no RepeatExplorer. Os dados de tamanho do genoma destas espécies foram  
146 oriundos do banco de dados online *Kew – Royal Botanic Gardens* (disponível em  
147 [data.kew.org/cvalues](http://data.kew.org/cvalues)).

148 Foram utilizados parâmetros padrões do software, sendo estes: sobreposição mínima (K-  
149 mer) de 40 pb entre os *reads*; e comprimento mínimo de 55 nt, para ser considerado significativo  
150 para formação do *cluster*. O detalhamento dos *clusters* foi realizado para os que apresentarem no  
151 mínimo 0,01% de repetição no genoma. As análises de bioinformática foram realizadas na  
152 Universidade Federal de Alagoas (UFAL) – *Campus Arapiraca*. O software Tarean identificou  
153 sequências de DNA satélite, permitindo obter monômeros e a proporção de repetição no genoma.

154

## 155 **Resultados**

156

### 157 **Tamanho do genoma**

158 Os resultados da quantificação do genoma de *H. speciosa* var. *speciosa* e *H. speciosa* var.  
159 *gardneri*, por meio da citometria de fluxo, mostraram valor 2C de  $0,87 \pm 0,02$  e  $0,88 \pm 0,01$  pg,  
160 respectivamente (aproximadamente 860 Mb). Os valores foram submetidos a teste de Tukey 5%  
161 sendo constatado que não diferem estatisticamente entre si (Figura 1).

162

### 163 **Análise de componentes principais**

164 A análise de componentes principais mostra a separação da subfamília Asclepiadoideae  
165 das demais pelo componente 1 e a divisão de Asclepiadoideae, Secamonoideae e Gentinaceae de  
166 Pericocloideae, Apocynoideae, Rauvolfioideae e um grupo de Asclepiadoideae pelo componente  
167 2. Foi possível obter a distância entre Asclepiadoideae e Rauvolfioideae, as quais englobam  
168 espécies deste estudo (Figura 2).



169

## 170 **Identificação da fração repetitiva na subfamília Rauvolfioideae**

171 A análise da fração repetitiva a partir de *reads* oriundos de NGS formou 162 clusters para  
172 *H. speciosa*, 316 para *Catharanthus roseus*, 225 para *Rhazya stricta* e 239 para *Asclepias syriaca*,  
173 correspondendo a 15,07%, 35,62%, 21,90% e 49,45%, do genoma, respectivamente (Tabela 2).  
174 Os resultados permitiram observar que *H. speciosa* possui a menor proporção de elementos  
175 repetitivos em relação às espécies avaliadas. A maior contribuição na fração repetitiva é dada  
176 pelos Retrotransposons, ou Elementos Classe I, os quais representam 6,215% do genoma de *H.*  
177 *speciosa* e 25,535% do genoma de *A. syriaca*, sendo estas a menor e a maior proporção de  
178 Retrotransposons dentre as espécies aqui analisadas, respectivamente (Tabela 3 e Figura 3B).

179 Os retroelementos LTR Ty3/*Gypsy* foram os mais abundantes em todas as espécies, porém  
180 a porcentagem de retroelementos da superfamília LTR Ty1/*Copia* é aproximada variando de 0,5  
181 a 1 % , em relação aos da superfamília LTR Ty3/*Gypsy*, com exceção da espécie *A. syriaca* a qual  
182 apresenta cerca de quatro vezes mais elementos LTR Ty3/*Gypsy* do que LTR Ty1/*Copia*. Os  
183 retrotransposons Non-LTR foram ausentes em *A. syriaca* e encontrados em pequenas proporções  
184 em *H. speciosa*, *C. roseus* e *R. stricta*, com representação de 0,22 % , 0,019% e 0,22%,  
185 respectivamente (Tabela 3 e Figura 3B).

186 Dentre as famílias de elementos LTR Ty3/*Gypsy* encontradas, a família Chromovirus foi  
187 a que apresentou maior proporção repetitiva dos genomas de *H. speciosa*, *C. roseus* e *R. stricta*,  
188 compondo 2,662 % , 5,103 % , 2,941%, respectivamente. *Asclepias syriaca* por sua vez, possui  
189 maior proporção da família Athila, a qual representa 15,58% de seu genoma (Tabela 2 e Figura  
190 2B). A família LTR Retrovirus não foi encontrada em *A. syriaca* e presente em pequenas  
191 proporções nas espécies *H. speciosa*, *C. roseus* e *R. stricta*, compondo apenas 0,38%, 0,68%,  
192 0,735% de seus genomas. Quanto aos elementos LTR Ty1/*Copia*, as famílias mais abundantes  
193 foram variadas em relação as espécies. No genoma de *H. speciosa* destaca-se a família Ikeros,  
194 com 1,699% de repetição, já para as espécies *C. roseus* e *R. stricta*, a família Maximus/SIRE foi

195 a mais representativa com 1,628% e 1,049% de repetição, respectivamente. Em *A. syriaca* a maior  
196 proporção de repetição no genoma, nesta superfamília, foi da família Bianca (2,078%) (Tabela 3  
197 e Figura 4).

198 Os transposons, ou elementos Classe II, apresentaram maiores proporções nas espécies *H.*  
199 *speciosa* e *C. roseus* (1,241% e 3,160%, respectivamente) enquanto que para *R. stricta* e *A.*  
200 *syriaca*, estes elementos não chegaram nem a 1% de representatividade em seus genomas (Tabela  
201 3 e Figura 3B). Em relação a subclasse I foram encontrados apenas elementos da Ordem TIR  
202 (*Terminal Inverted Repeat*) com as famílias CACTA e hAT presentes em quase todas as espécies  
203 estudadas, com exceção de *A. syriaca* na qual apenas as famílias CACTA e Mutator forem  
204 encontradas em proporções inferiores a 1%. *Rhazya stricta* apresentou a maior variabilidade de  
205 transposons subclasse I, tendo em seu genoma a presença das superfamílias CACTA, hAT, PIF-  
206 Harbinger e Mutator, todas em pequenas proporções (menos que 1%). *Catharanthus roseus*  
207 apresentou famílias CACTA, hAT e Mutator, destacando-se a família CACTA com 1,179% de  
208 repetição, seguida da família Mutator com 1,439%. *Hancornia speciosa* contém em seu genoma  
209 apenas as famílias CACTA e hAT em proporções menores que 1%, no entanto, a espécie foi a  
210 única, dentre as estudadas, que apresentou a presença de transposons subclasse II, da superfamília  
211 Helitron com proporção de 0,011%. (Tabela 3 e Figura 4).

212 A porcentagem de sequências de DNA satélite foi de 5,332% para *C. roseus* e 0,274% para  
213 *A. syriaca*, no entanto, para as espécies *H. speciosa* e *R. stricta* não foram encontradas tais  
214 sequências, nem mesmo em pequena quantidade (Tabela 2 e Figura 3). O DNA ribossomal  
215 apresenta alta representatividade em *H. speciosa*, totalizando 4,818% de seu genoma. Para *C.*  
216 *roseus*, *R. stricta* e *A. syriaca*, as porcentagens de DNAr foram de 1,008%, 1,134% e 2,22%,  
217 respectivamente (Tabela 3 e Figura 3B). *Hancornia speciosa* apresenta um subrepeat que engloba  
218 13 clusters na região intergênica (IGS) do DNAr 45S (Figura 5).

219 Os resultados da análise de componentes principais relacionam-se com as porcentagens e  
220 distribuição de elementos repetitivos no genoma, as quais apresentam-se mais semelhantes entre

221 as espécies *H. speciosa*, *C. roseus* e *R. stricta* (subfamília Rauvolfioideae) e mais divergentes com  
222 as características da fração repetitiva do genoma de *A. syriaca* (subfamília Asclepiadoideae)  
223 (Tabela 3, Figuras 2 e 3A).

224

## 225 **Discussão**

### 226 **Tamanho do genoma**

227 Elementos repetitivos são sequências que aparecem diversas vezes no genoma e são os  
228 importantes contribuintes para aumento de seu tamanho (Kazazian Júnior, 2004). Anteriormente  
229 achava-se que o tamanho do genoma estava correlacionado com a especificidade do organismo,  
230 no entanto, hoje acredita-se que os retrotransposons da ordem LTR são os principais responsáveis  
231 para o aumento do tamanho genômico (Feschotte, 2008). O presente estudo revelou conteúdos  
232 variados de DNA repetitivo em espécies da ordem Gentianales, as quais possuem genomas  
233 classificados como pequenos (< 3,5 pg).

234 Ainda que correlacionados com o aumento do tamanho genômico de espécies vegetais,  
235 genomas pequenos também podem conter elevados conteúdos de elementos repetitivos. Aqui  
236 mostramos que *A. syriaca* (Valor 1C 0,42 pg) possui aproximadamente 50% de DNA repetitivo  
237 em seu conteúdo genômico, 25% correspondente a retrotransposons LTR, enquanto *H. speciosa*  
238 (Valor 1C 0,44 pg) possui, em média, apenas 15%, de conteúdo repetitivo total, destes 6%  
239 retrotransposons. Desta forma, ainda que responsáveis pelo aumento no tamanho do genoma, estes  
240 elementos encontram-se presentes em espécies de genoma pequeno em proporções variadas,  
241 podendo chegar a 50%.

242 Estudo recente mostrou a comparação de conteúdos de elementos repetitivos entre  
243 *Vallisneria spinulosa* S. Z. Yan e outras espécies monocotiledôneas, em que apesar de *V.*  
244 *spinulosa* possuir genoma cerca de 5 vezes maior que *Sorghum bicolor* L. este apresenta maior  
245 quantidade de retrotransposons (Feng *et al.*, 2017). No presente estudo observamos que embora  
246 *H. speciosa* apresente tamanho genômico próximo ao de *A. syriaca* e duas vezes maior que *R.*

247 *stricta* seu conteúdo de repetição é o menor dentre as espécies analisadas.

248 Os elementos Ty1-*copia* e Ty3-*gypsy* são as duas principais superfamílias de  
249 retrotransposons LTR (Wicker *et al.*, 2007). Os resultados comprovam sua abundância nas  
250 espécies da família Apocynaceae, correspondendo a aproximadamente 50% de toda a fração  
251 repetitiva com exceção apenas de *C. roseus* que apresenta, além destes elementos, transposons,  
252 em especial a superfamília Mutator.

253 Estudos acerca da atividade transcricional de elementos das superfamílias Ty3-*Gypsy* e  
254 Ty1-*Copia* mostraram que elementos *Copia* possuem transcrição ativada, sendo assim, expressos  
255 em níveis mais elevados, isto porque elementos *Gypsy* (mais abundantes) tendem a ser  
256 frequentemente silenciados pela maquinaria molecular celular, justamente por serem  
257 demasiadamente presentes no genoma e ocasionando mutações. As regiões de elementos *Copia*,  
258 que estão presentes em menores quantidades, são mais propensos a escapar dos mecanismos de  
259 silenciamento (Qiu and Ungerer, 2018). Neste estudo observamos que a superfamília *Gypsy* foi  
260 encontrada em maior proporção que *Copia* em Apocynaceae e, dentre as linhagens de Ty3-*Gypsy*  
261 a cromossoviral foi a mais abundante em *H. speciosa*, *C. roseus* e *R. stricta* (subfamília  
262 Rauvolfioideae) resultados semelhantes ao encontrado em estudos recentes de DNA repetitivo do  
263 tipo LTR retrotransposons no gênero *Helianthus* (Mascagni *et al.*, 2017).

264 O terminal C de integrase é altamente divergente entre os retrotransposons Ty3-*Gypsy*, em  
265 sequência e tamanho, exceto para uma linhagem específica referida como cromovírus. Esta  
266 linhagem cromoviral, ou LTR/*Gypsy* Chromovirus foi a mais representativa nas espécies *H.*  
267 *speciosa*, *R. stricta* e *C. roseus*, diferentemente de *A. syriaca* a qual teve o grupo Athila em maior  
268 proporção. Este grupo cromoviral é classificado pela presença de um cromodomínio localizado  
269 no C-terminal da integrase, sugerindo um papel potencial para a especificidade do alvo de  
270 retrotransposons LTR no genoma hospedeiro (Malik and Eickbush, 1999).

271 Sequências de DNAsat encontram-se geralmente em regiões específicas e que  
272 desempenham função nos cromossomos, como nos genes de DNA ribossomal, nos telômeros,

273 subtelômeros e nos centrômeros apresentando papel potencial na estabilidade do genoma e  
274 segregação cromossômica (Jiang *et al.*, 2003; Khost *et al.*, 2017). A ausência de DNAsat em *H.*  
275 *speciosa* e *R. stricta* sugere que algum elemento transponível pode ter assumido o papel das  
276 sequências de DNAsat ao longo da evolução das espécies.

277 Os retrotransposons LTR são frequentemente localizados em centrômeros de plantas  
278 (Weber and Schmidt, 2009). Os LTRs *Gypsy/Chromovirus* contêm um domínio presente em várias  
279 proteínas envolvidas na remodelação da cromatina e na regulação da expressão gênica em  
280 eucariotas durante o desenvolvimento do organismo. Os cromodomínios são capazes de interagir  
281 com diversos componentes da cromatina, incluindo proteínas, RNA e DNA permitindo a interação  
282 destes elementos com proteínas centroméricas (Nagaki *et al.*, 2003). Estas características  
283 associadas à distribuição destes elementos em *tandem* no DNAr (Figura 5) e a sua elevada  
284 proporção em três das quatro espécies analisadas permitem sugerir que os retrotransposons LTR  
285 *Gypsy/Chromovirus* atuam como DNA satélite no genoma de *H. speciosa*.

286 Características como a ausência de DNA satélite, em duas das três espécies de  
287 *Rauvolfioideae* analisadas, e a presença de elementos LTR/Retrovirus Pararetovirus parecem ser  
288 características desta superfamília, no entanto, de forma geral, mesmo dentro de uma mesma  
289 subfamília características divergentes acerca do componente repetitivo dos genomas podem ser  
290 encontrados como: a presença de DNA satélite em *C. roseus*, a presença de transposons subclasse  
291 II em *H. speciosa*, assim como a ausência nesta espécie de famílias LTR/*Gypsy* *Ivana/Oryco* e  
292 *Maximus/SIRE*. Sequências repetitivas abundantes apresentam assinaturas genômicas assinaturas  
293 filogenéticas (Dodsworth *et al.*, 2014). Sendo assim, a dinâmica dos retrotransposons LTR e suas  
294 contribuições para a evolução do genoma podem ser usadas para analisar a variabilidade que  
295 ocasionou a subclassificação de *H. speciosa* em suas seis variedades.

296 Diferentemente da maioria das sequências repetitivas, cujas funções em sua maioria não  
297 são claras, as atividades dos genes de DNAr 5S e 45S são vitais para os organismos, fornecendo  
298 RNAs ribossomais (RNAr) necessários para a montagem de ribossomos funcionais, que

299 representam mais de 90% do RNA celular total (Volkov *et al.*, 2017). *Hancornia speciosa*  
300 apresenta DNAr em elevada proporção (aproximadamente 5%), sendo este um indicativo na  
301 eficiência do mecanismo molecular de tradução de proteínas ou amplificação de sítios ao longo  
302 dos cromossomos.

303 Embora numerosas cópias de repetições de DNAr coexistam no mesmo genoma, elas  
304 tendem a ser quase idênticas em muitas espécies diploides devido ao processo de homogeneização  
305 de sequência, sendo assim, cópias individuais de elementos repetidos não evoluem de forma  
306 independente, mas de forma concentrada (Volkov *et al.*, 2003). Desta forma, o acúmulo de DNAr  
307 pode estar ligado à evolução da espécie ou apenas estar relacionado com o funcionamento eficaz  
308 de sua maquinaria de transcrição. Trabalhos posteriores envolvendo técnicas de citogenômica  
309 como a hibridização fluorescente in situ (FISH) podem esclarecer o papel do DNAr em sua  
310 evolução em *H. speciosa*.

311

## 312 **Conclusões**

313 Apesar da forte contribuição de elementos transponíveis (TE's) para o aumento no  
314 tamanho genômico, espécies de genomas pequenos como o de *H. speciosa* podem apresentar  
315 elevada proporção de TE's.

316 A ausência de DNAsat em *H. speciosa* aliada a elevada proporção de LTR Ty3-  
317 Gypsy/Chromovirus e sua conformação em *tandem* na região do DNAr sugere que ao longo da  
318 evolução estes elementos assumiram o papel do DNAsat nesta espécie .,

319

320

321

322

323

324

325 **Referências**

- 326 Almeida VM, Luz GA, Martins PP, Gomes MFC, Costa MF, Lima PSC and Valente SES (2017)  
327 Comparation of eight methods to isolate genomic DNA from *Hancornia speciosa*. Genet  
328 Mol Res 16: 1-7.
- 329 Collevatti RG, Olivatti AM, Telles MPC and Chaves IJ (2016) Gene flow among *Hancornia*  
330 *speciosa* (Apocynaceae) varieties and hybrid fitness. Tree Genet Genomes 12: 1-12.
- 331 Dodsworth S, Chase MW, Kelly LJ, Leitch IJ, Macas J, Novák P, Piednoel M, Weiss-Schneeweiss  
332 and Leitch AR (2014). Genomic repeat abundances contain phylogenetic signal. Syst. Biol  
333 64: 112-126.
- 334 Doležel J, Greilhuber J (2010) Nuclear genome size: are we getting closer?. Cytometry 77A: 635-  
335 642.
- 336 Doyle JJ and Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf  
337 tissue. Phytochemical Bulletin 19:11-15.
- 338 Feng R, Wang X, Tao M, Du G and Wang K (2017) Genome size and identification of abundant  
339 repetitive sequences in *Vallisneria spinulosa*. PeerJ 5:e3982;DOI10.7717/peerj.3982.
- 340 Fechotte C (2008) The contribution of transposable elements to the evolution of regulatory  
341 networks. Nat Ver Genet 9: 397-405.
- 342 Ganga RMD, Ferreira GA, Chaves LJ, Naves RV and Nascimento JL (2010) Caracterização de  
343 frutos e árvores de populações naturais de *Hancornia speciosa* Gomes do Cerrado. Rev Bras  
344 Frut 32: 101-113.
- 345 Jiang J, Birchler JA, Parrot WA and Dawe RK (2003). A molecular view of plants centromeres.  
346 Trend Plant Sci 8: 570-575.
- 347 Kazazian Júnior HH (2004) Mobile elements: drivers of genome evolution. Science 303:1626-  
348 1632.
- 349 Khost DE, Eickbush DG and Larracuente AM (2017) Single-molecule sequencing resolves the  
350 detailed structure of complex satellite DNA loci in *Drosophila melanogaster*. Genome Res

351 21: 709-721.

352 Lima LL and Scariot A (2010) Boas práticas de manejo para o extrativismo sustentável da  
353 mangaba. Brasília –DF: Embrapa recursos genéticos e biotecnologia 1-68.

354 Loureiro J, Rodriguez e, Doležel J and Santos C (2007) Two new nuclear isolation buffers for  
355 plant DNA flow cytometry: a test with 37 species. *Ann Bot* 100: 875-888.

356 Malik HS and Eickbush TH (1999) Modular evolution of the integrase domain in the Ty3/Gypsy  
357 class of LTR retrotransposons. *J Virol.* 73: 5186-5190.

358 Mascagni F, Giordani T, Ceccarelli M, Cavalline A and Natali L (2017) Genome wide – analysis  
359 of LTR-retrotransposon diversity and its impact on the evolution of the genus *Helianthus*.  
360 *BMC Genomics* 18:1-16.

361 Mehrotra S and Goyal V (2014) Repetitive sequences in plant nuclear DNA: types, distribution,  
362 evolution and function. *Genomics Proteomics Bioinformatics* 12:164-171.

363 Metzker ML (2010) Sequencing technologies – the next generation. *Nat Rev Genet.* 11:31-46.

364 Nagaki K, Song J, Stupar RM, Parokony AS, Yuan Q, Ouyang S, Liu J, Hsiao J, Jones KM,  
365 Dawe RK and Buell CR (2003). Molecular and cytological analyses of large tracks of  
366 centromeric DNA reveal the structure and evolutionary dynamics of maize centromeres.  
367 *Genetics*,163:759-770.

368 Novák P, Neumann P and Macas J (2010) Graph-based clustering and characterization of  
369 repetitive sequences in next-generation sequencing data. *BMC Bioinformatics* 11:1-12.

370 Pereira AC, Pereira AB, Moreira CC, Borion LM, Lemos VS, Braga FC and Cores SF (2015)  
371 *Hancornia speciosa* Gomes (Apocynaceae) as a potential anti-diabetic drug. *J*  
372 *Ethnopharmacol*, 30-36.

373 Qiu F and Ungeres MC (2018) Genomic abundance and transcriptional activity of diverse gypsy  
374 and copia long terminal repeat retrotransposons in three wild sunflower species. *BMC Plant*  
375 *Biol* 18:1-8.

376 Ribeiro TP, Sousa TR, Arruda AS, Peixoto N, Gonçalves PJ and Almeida LM (2016) Evaluation



377 of citotoxicity and genotoxicity of *Hancornia speciosa* látex in *Allium cepa* root model. Braz J  
378 Biol 76: 245-249.

379 Santos PS, Freitas LS, Santana JGS, Muniz EN, Rabbani ARC and Silva AVC (2017) Genetic  
380 diversity and quality of mangabeira tree fruits (*Hancornia speciosa* Gomes – Apocynaceae)  
381 a native species from Brazil. Sci Hortic 226: 372-378.

382 Silva GC, Braga FC, Lemos VS and Cortes SF (2016) Potent antihypertensive effect of *Hancornia*  
383 *speciosa* leaves extract. Phytomedicine 23: 214-219.

384 Song X and Cao X (2017) Transposon-mediated epigenetic regulation contributes to phenotypic  
385 diversity and environmental adaptation in rice. Curr Opin Plant Biol 36:111-118.

386 Team RCR (2016) A language and environment for statistical computing. R Foundation for  
387 Statistical Computing, Vienna, Austria, 2016. URL <https://www.R-project.org/>

388 Unamba CIN, Nag A, Ram KS (2015) Next Generation Sequencing Technologies: The Doorway  
389 to the Unexplored Genomics of Non-Model Plants. Front Plant Sci 6: 1-16.

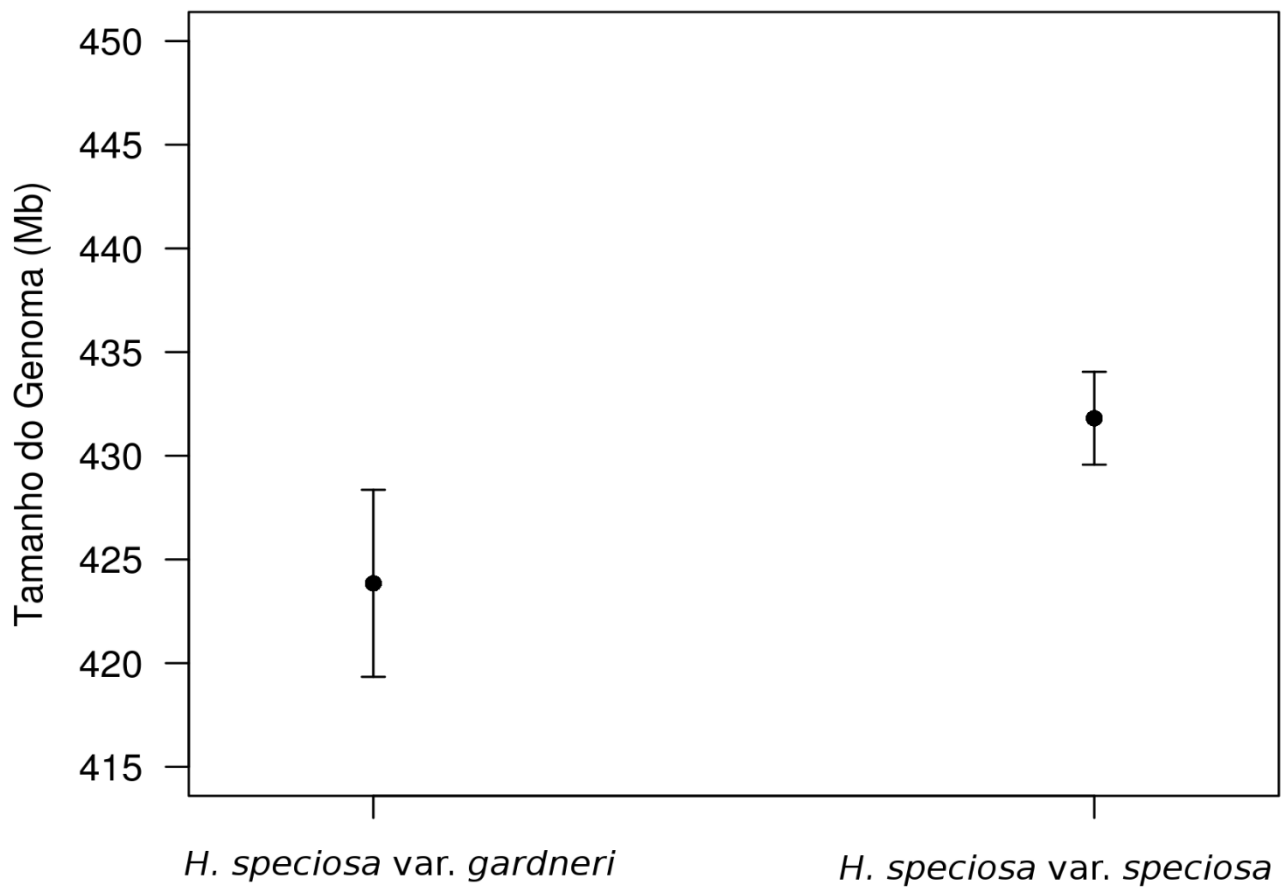
390 Volkov RA, Panchuk II, Borisjuk NV, Hosiawa-Baranska MH, Maluszynksa J and Hemleben V  
391 (2017) Evolutional dynamics of 45S and 5S ribosomal DNA in ancient allohexaploid.  
392 *Atropa belladonna*. BMC Plant Biol 17: 1-15.

393 Weber B and Schmidt T (2009) Nested Ty3-gypsy retrotransposons of a single Beta procumbens  
394 centromere contain a putative chromodomain. Chromosome Res. 17:379-396.

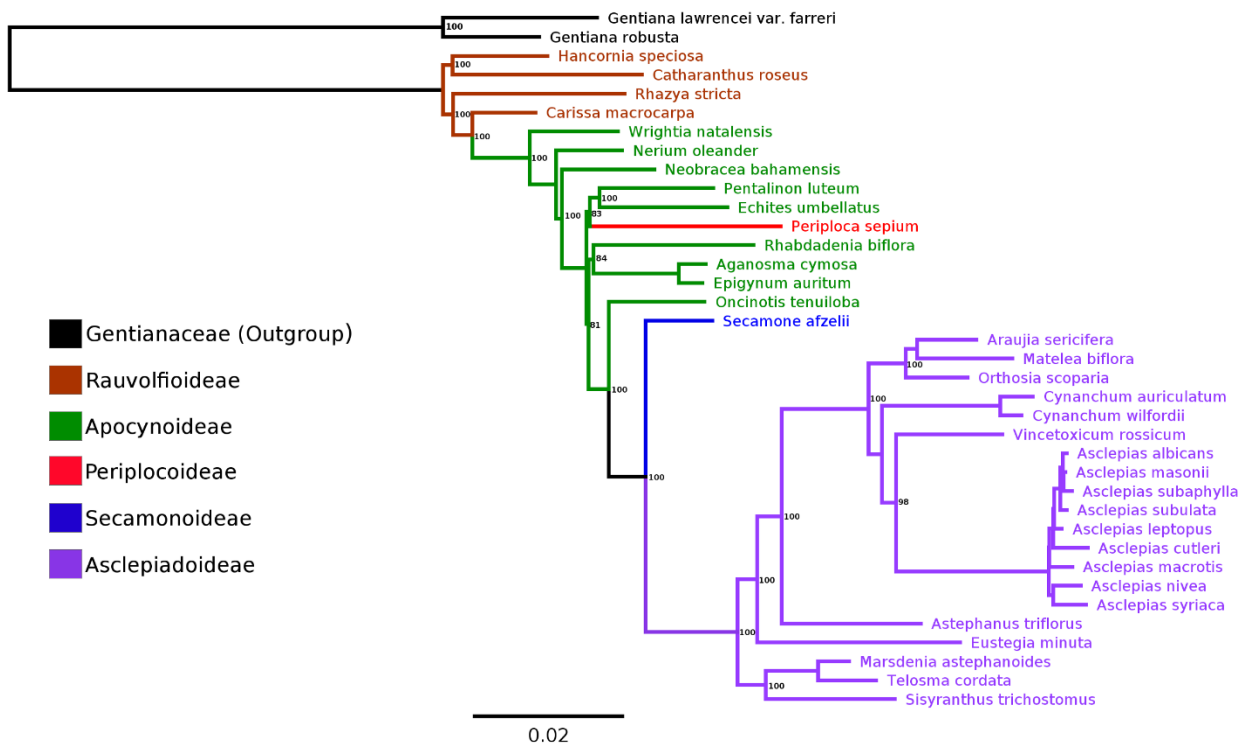
395 Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Copy P, Chalhou B, Flavell A, Leroy P, Morgante  
396 M, Panaud O, Paux E, SanMiguel F and Schulman AH (2007) A unified classification  
397 system for eukaryotic transposable elements. Nat Rev Genet 8:973-982.

398 Yang J, Li D and Li H (2014) Highly effective sequencing whole chloroplast genomes of  
399 angiosperms by nine novel universal primer pairs. Mol Ecol Resour, 14:1024-1031.

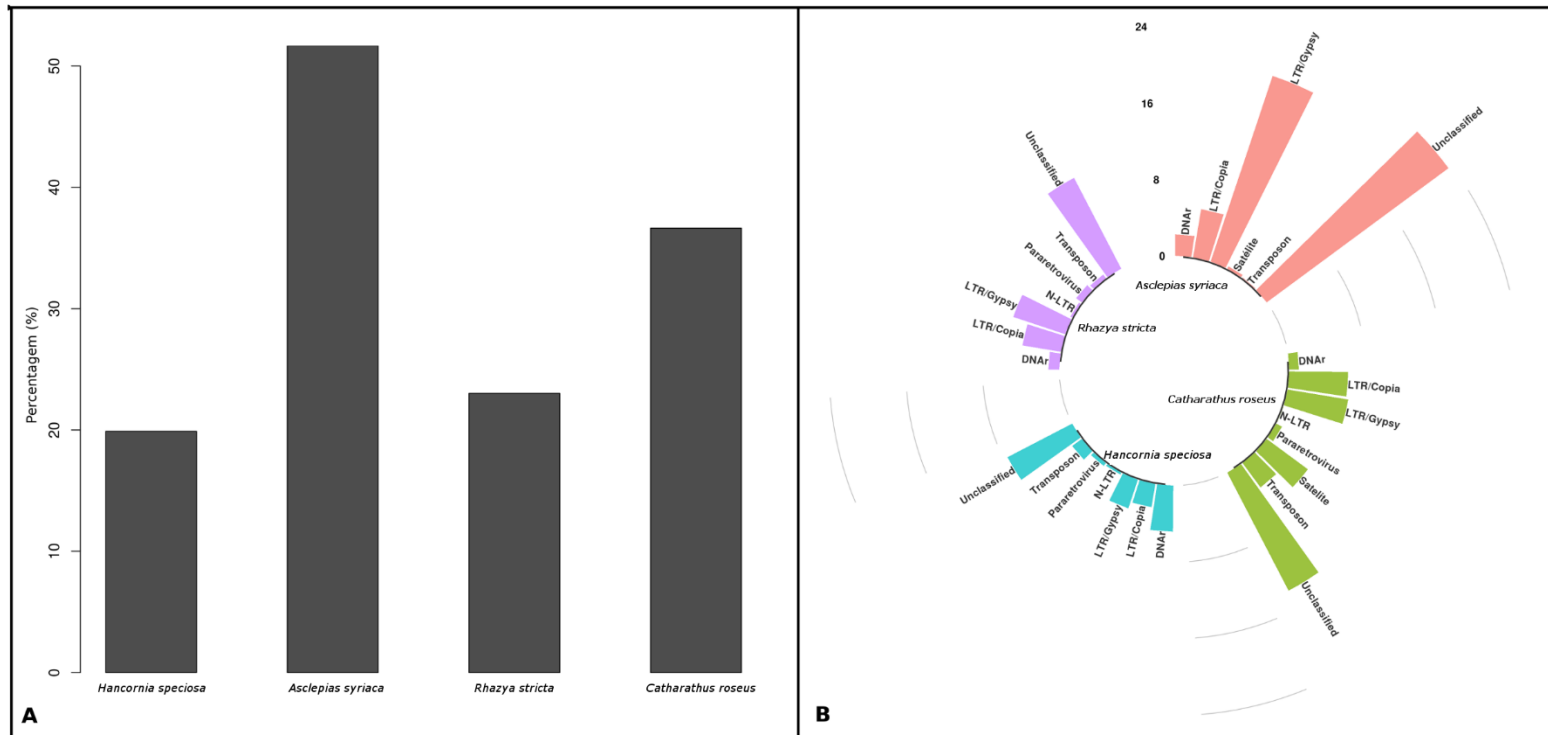
400 Zerjal T, Rousselet A, Mhiri C, Combes V, Madur D, Grandbastien MA, Charcosset A and  
401 Tenailon MT (2012) Maize genetic diversity and association mapping using transposable  
402 element insertion polymorphisms. Theor Appl Genet 124: 1521-1537.



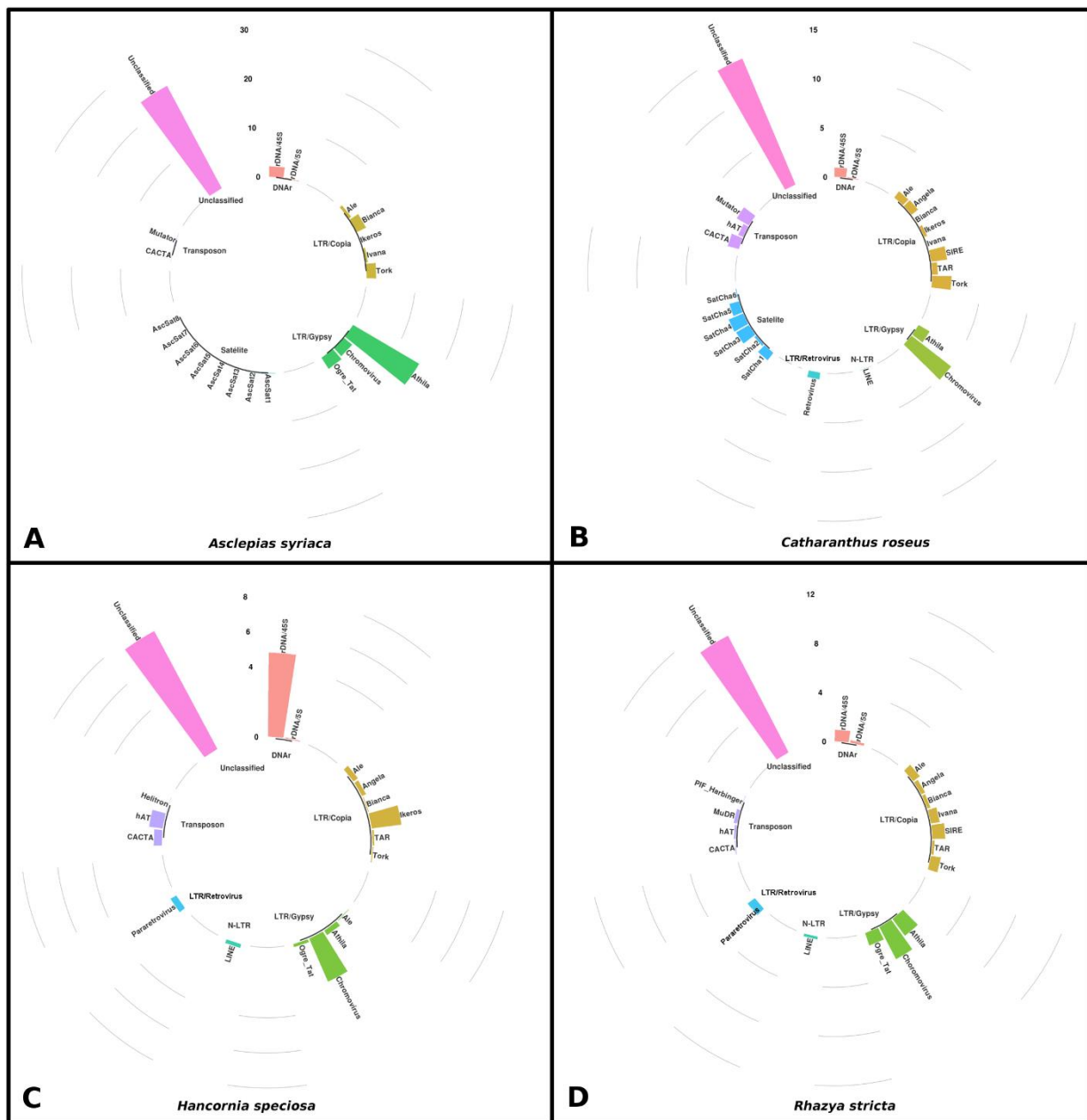
**Figura 1.** Tamanho do genoma de duas variedades de *Hancornia speciosa* (var. *gardneri* e var. *speciosa*) à teste de médias usando 1C value, pelo teste de Tukey ao nível de 5% de significância.



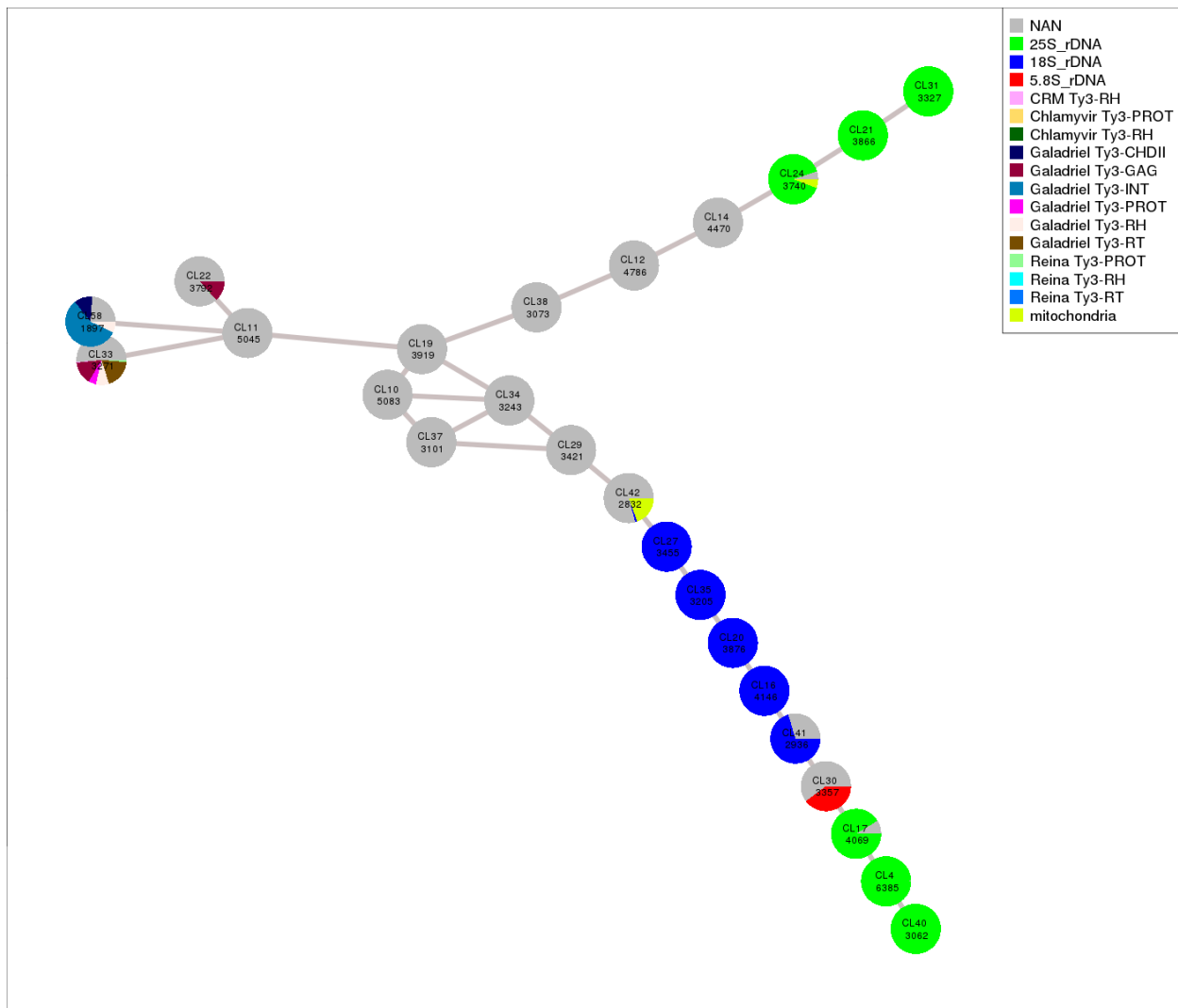
**Figura 2.** Árvore filogenética da ordem Gentianales baseada em genomas plastidiais inteiros, composta pelas subfamílias Ravoldioideae, Apocynoideae, Periplocoideae, Secamonoideae e Asclepiadeae e Gentianaceae (grupo externo).



**Figura 3.** (A) Proporção de elementos repetitivo nos genomas de *H. speciosa*, *C. roseus*, *R. stricta* e *A. syriaca* (B) comparação da fração repetitiva nos genomas das espécies identificando: sequências de DNA satélite; sequências repetitivas não classificadas (repetições simples); regiões repetitivas do DNAr e; elementos transponíveis classes I e II retrotransposons (LTRs e N-LTRs) e transposons, respectivamente.



**Figura 4.** Caracterização da fração repetitiva dos genomas de (A) *Asclepias syriaca* (B) *Catharanthus roseus* (C) *Hancornia speciosa* e (D) *Rhazya stricta*, quanto às famílias de elementos transponíveis e subunidades do DNAr.



**Figura 5.** DNA ribossomal 45S de *Hancornia speciosa* composto pelas subunidades 5.8s, 18s e 25.8s e regiões repetitivas ricas em seqüências de LTR Gypsy Chromovirus dos tipos Galadriel e Reina dispostas em *tandem*.

**Tabela 1.** Referências dos genomas plastidiais utilizados na análise filogenética e disponíveis no GenBank.

<b>Espécies</b>	<b>Subfamília</b>	<b>GenBank n°</b>	<b>Referências</b>
<i>Aganosma cymosa</i> (Roxb.) G.Don	Apocynoideae	KJ953903	Straub,S.C. et al. 2016
<i>Echites umbellatus</i> Jacq.	Apocynoideae	KJ953904	Straub,S.C. et al 2016
<i>Epigynum auritum</i> (C.K.Schneid.) Tsiang & P.T.Li	Apocynoideae	KJ953905	Straub,S.C. et al 2016
<i>Neobracea bahamensis</i> (Britton) Britton	Apocynoideae	KJ953906	Straub,S.C. et al 2016
<i>Nerium oleander</i> L.	Apocynoideae	KJ953907	Straub,S.C. et al 2016
<i>Oncinotis tenuiloba</i> Stapf	Apocynoideae	KJ953908	Straub,S.C. et al 2016
<i>Pentalinon luteum</i> (L.) B.F.Hansen & Wunderlin	Apocynoideae	KJ953909	Straub,S.C. et al 2016
<i>Rhabdadenia biflora</i> (Jacq.) Müll.Arg	Apocynoideae	KJ953911	Straub,S.C. et al 2016
<i>Wrightia natalensis</i> Stapf	Apocynoideae	KJ953913	Straub,S.C. et al 2016
<i>Asclepias syriaca</i> L.	Asclepiadoideae	JF433943	Straub,S.C. et al 2011
<i>Asclepias albicans</i> S.Watson	Asclepiadoideae	JN710458	Straub,S.C. et al 2012
<i>Asclepias cultriformis</i> Harv. ex Schltr.	Asclepiadoideae	JN710461	Straub,S.C. et al 2011
<i>Asclepias leptopus</i> I.M.Johnst.	Asclepiadoideae	JN710462	Straub,S.C. et al 2011
<i>Asclepias macrotis</i> Torr.	Asclepiadoideae	JN710463	Straub,S.C. et al 2011
<i>Asclepias masonii</i> Woodson	Asclepiadoideae	JN710465	Straub,S.C. et al 2011
<i>Asclepias subaphylla</i> Woodson	Asclepiadoideae	JN710467	Straub,S.C. et al 2011
<i>Asclepias subulata</i> Decne	Asclepiadoideae	JN710469	Straub,S.C. et al 2011
<i>Asclepias nivea</i> L.	Asclepiadoideae	KF539844	Straub,S.C.K. et al 2013
<i>Araujia sericifera</i> Brot.	Asclepiadoideae	KF539846	Straub,S.C.K. et al 2013
<i>Astephanus triflorus</i> (L.f.) R.Br.	Asclepiadoideae	KF539847	Straub,S.C.K. et al 2013
<i>Eustegia minuta</i> (L.f.) N.E.Br.	Asclepiadoideae	KF539848	Straub,S.C.K. et al 2013
<i>Marsdenia astephanoides</i> (A. Gray) Woodson	Asclepiadoideae	KF539849	Straub,S.C.K. et al 2013
<i>Mateleia biflora</i> (Raf.) Woodson	Asclepiadoideae	KF539850	Straub,S.C.K. et al 2013
<i>Orthosia scoparia</i> (Nutt.) Liede & Meve	Asclepiadoideae	KF539851	Straub,S.C.K. et al 2013
<i>Sisyranthus trichostomus</i> K. Schum.	Asclepiadoideae	KF539852	Straub,S.C.K. et al 2013
<i>Telosma cordata</i> (Burm. f.) Merr.	Asclepiadoideae	KF539853	Straub,S.C.K. et al 2013
<i>Vincetoxicum rossicum</i> (Kleopov) Barbar.	Asclepiadoideae	KF539854	Straub,S.C.K. et al 2013

<i>Cynanchum wilfordii</i> (Maxim.) Hemsl.	Asclepiadoideae	NC_029459	Jang, W. et al 2015
<i>Cynanchum auriculatum</i> Royle ex Wight	Asclepiadoideae	NC_029460	Jang, W. et al 2015
<i>Gentiana robusta</i> King ex Hook.f.	Gentianeae	KT159969	Ni,L.H. and Zhao,Z.L. 2016
<i>Gentiana lawrencei</i> var. <i>farreri</i> (Balf.f.) T.N.Ho	Gentianeae	KX096882	Fu,P.-C., Sun,S.-S. and Chen,S.-L. 2016
<i>Periploca sepium</i> Bunge	Periplocoideae	KJ953910	Straub,S.C.K. et al 2014
<i>Hancornia speciosa</i> Gomes	Rauvolfioideae	-	
<i>Catharanthus roseus</i> (L.) G. Don	Rauvolfioideae	KC561139	Ku,C. 2013
<i>Rhazya stricta</i> Decne	Rauvolfioideae	KJ123753	Yates,S.A. 2015
<i>Carissa macrocarpa</i> (Eckl.) A.DC.	Rauvolfioideae	KX364402	Jo,S. et al 2017
<i>Secamone afzelii</i> (Schult.) K.Schum.	Secamonoideae	KF539845	Straub, S.C.K. et al 2013

---



**Tabela 2** - Tamanho dos genomas e resultados do agrupamento de sequências usando RepeatExplorer para as espécies *H. speciosa*, *C. roseus*, *R. stricta* e *A. syriaca*.

<b>Espécies</b>	<b>Tamanho do genoma [pg - Mb]</b>	<b>Reads analisados [pb]</b>	<b>Cobertura</b>	<b>Reads in clusters [pb]</b>	<b>Proporção repetitiva [%]</b>
<i>H. speciosa</i>	0,44 – 430,32	660744	0,07	252652	15,07
<i>C. roseus</i>	0,76 – 743,28	661788	0,08	325788	35,62
<i>R. stricta</i>	0,28 – 274,00	660754	0,24	256683	21,89
<i>A. syriaca</i>	0,42 – 410,76	661946	0,16	491062	49,45

**Tabela 3** – Distribuição de elementos repetitivos com mais de 0,01% de representação nos genomas de *H. speciosa*, *C. roseus*, *R. stricta* e *A. syriaca*, classificada de acordo com Wicker *et al.* (2007) em Classe, Subclasse, Superfamília e Família.

Classe	Subclasse	Ordem	Superfamília	Família	% Genoma				
					<i>H. speciosa</i>	<i>C. roseus</i>	<i>R. stricta</i>	<i>A. syriaca</i>	
Retrotransposons	LTR			Ale	0,342	0,801	0,678	0,606	
				Angela	0,250	0,850	0,355	0	
				Bianca	0,042	0,015	0,273	2,078	
				Ikeros	1,699	0,300	0	0,040	
				Ty1-Copia	Ivana/Oryco	0	0,0099	0,721	0,427
				SIRE	0	1,628	1,049	0	
				TAR	0,100	0,571	0,180	0	
				Tork	0,052	1,970	0,808	1,947	
				<b>Total</b>	<b>2,485</b>	<b>6,1449</b>	<b>4,064</b>	<b>5,098</b>	
				Ty3- Gypsy	Athila	0,300	1,302	1,702	15,580
					Chromovirus	2,662	5,103	2,941	2,180
					Ogre/Tat	0,171	0	1,066	2,677
				<b>Total</b>	<b>3,133</b>	<b>6,405</b>	<b>5,709</b>	<b>20,437</b>	
				Retrovirus	Pararetrovirus	0,380	0,680	0,735	0
<b>Total</b>	Non-LTR (LINE)	L1		0,220	0,019	0,220	0		
			<b>6,218</b>	<b>13,249</b>	<b>10,728</b>	<b>25,535</b>			
Transposons	Suclasse I	TIR	CACTA	0,459	1,179	0,050	0,021		
			hAT	0,771	0,542	0,140	0		
			PIF-Harbinger	0	0	0,026	0		
			Mutator	0	1,439	0,265	0,042		
			<b>Total</b>	<b>1,230</b>	<b>3,160</b>	<b>0,481</b>	<b>0,063</b>		
Subclasse II	Helitron	Helitron	0,011	0	0	0			
<b>Total</b>			<b>1,241</b>	<b>3,160</b>	<b>0,481</b>	<b>0,063</b>			
<b>Total (TE's)</b>			<b>7,459</b>	<b>16,409</b>	<b>11,209</b>	<b>25,598</b>			
DNA ribossomal		rDNA		4,818	1,008	1,134	2,220		
DNA satélite		satDNA		0	5,332	0	0,274		
<b>Total não classificado</b>				7,609	13,882	10,682	23,582		
<b>Total de repetição</b>				<b>15,068</b>	<b>35,623</b>	<b>21,891</b>	<b>49,454</b>		