



MARCELO ANTÔNIO DA TRINDADE

**ANÁLISE COMPARATIVA DE ELEMENTOS REPETITIVOS
EM GENOMAS DE ESPÉCIES DA SUBTRIBO CASSIINAE
(FABACEAE)**

**LAVRAS – MG
2021**

MARCELO ANTÔNIO DA TRINDADE

**ANÁLISE COMPARATIVA DE ELEMENTOS REPETITIVOS EM GENOMAS DE
ESPÉCIES DA SUBTRIBO CASSIINAE (FABCEAE)**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, para a obtenção do título de Mestre.

Profa. Dra. Giovana Augusta Torres
Orientadora

**LAVRAS – MG
2021**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Trindade, Marcelo Antônio da.

Análise comparativa de elementos repetitivos em genomas de espécies da subtribo Cassiinae (Fabaceae) / Marcelo Antônio da Trindade. - 2021.

52 p. : il.

Orientador(a): Giovana Augusta Torres.

Dissertação (mestrado acadêmico) - Universidade Federal de Lavras, 2021.

Bibliografia.

1. Fabaceae. 2. Cassiinae. 3. Elementos transponíveis. I. Torres, Giovana Augusta. II. Título.

MARCELO ANTÔNIO DA TRINDADE

**ANÁLISE COMPARATIVA DE ELEMENTOS REPETITIVOS EM GENOMAS DE
ESPÉCIES DA SUBTRIBO CASSIINAE (FABACEAE)**

**COMPARATIVE ANALYSIS OF REPETITIVE ELEMENTS IN GENOMES OF
SPECIES OF THE CASSIINAE SUBTRIBE (FABACEAE)**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, para a obtenção do título de Mestre.

APROVADA em 06 de agosto de 2021.

Profa. Dra. Magdalena Vaio Scvortzoff – Universidad de la Republica – Uruguay

Prof. Dr. Luiz Gustavo Rodrigues Souza – UFPE

Profa. Dra. Giovana Augusta Torres
Orientadora

**LAVRAS – MG
2021**

A Frederico (in memoriam), que me ensinou a amar e a me permitir ser amado. Pelos anos de companheirismo, risadas, tardes com muito café e conversas aleatórias acompanhadas de muito drama. Tenho fé que um dia ainda nos reencotraremos, até lá, guardo meu melhor abraço.

Dedico

AGRADECIMENTOS

À minha mãe, pelo carinho e suporte emocional.

Às minhas irmãs, pela presença e amizade.

À Janayne por ser o melhor presente que a UFU me deu, certamente levarei essa amizade sempre comigo.

Ao pessoal do laboratório de citogenética da UFLA, em especial ao Alex que foi fundamental para a execução deste trabalho.

À Ana Gabriela e Clara por todo o apoio e ajuda durante o mestrado.

Aos professores do Programa de Genética e melhoramento de plantas, em especial à professora Giovana, pela ajuda e paciência em me orientar durante este ciclo.

Aos técnicos do laboratório, responsáveis pela manutenção.

Aos amigos de mestrado Danyllo, Kátia, Carlos e Alessandra, pelo apoio e pelas longas conversas acompanhadas de uma boa cerveja.

À Universidade Federal de Lavras por toda estrutura e investimento.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

RESUMO

Fabaceae (Leguminosae) é considerada uma das três maiores famílias de Angiospermae, contendo 765 gêneros e 19.500 espécies de distribuição cosmopolita. No Brasil, são encontradas cerca de 2.756 espécies, sendo 1507 endêmicas. Diversos grupos de leguminosas têm tido sua classificação alterada ao longo do tempo, como é o caso da subtribo Cassiinae, que apresenta três gêneros resultantes da divisão de *Cassia* lato sensu: *Cassia* stricto sensu, *Chamaecrista* Moench e *Senna* Mill. Com base em análises filogenéticas, a tribo Cassiinae tem sido considerada complexa devido a hipóteses conflitantes em relação aos três gêneros que a constituem. Estudos moleculares e citogenéticos foram importantes para a contribuição da sistemática de Cassiinae, mas não suficientes para elucidar as controvérsias que envolvem a relação entre os três gêneros. A inclusão de outras informações genômicas são importantes para contribuir com o melhor entendimento das relações dentro do grupo. A partir do advento do sequenciamento de próxima geração (NGS) foi possível estudos que permitem a identificação e caracterização da fração repetitiva no genoma de uma espécie, mostrando ser uma interessante ferramenta para a compreensão da evolução genômica. Desta forma, nosso trabalho visou obter a fração repetitiva no genoma de espécies representantes dos três gêneros de Cassiinae. Sequenciamento de baixa cobertura (cerca de 0,4x) foi obtido na Plataforma Illumina HiSeq™ 4000 para uma espécie de cada gênero (*Cassia fistula*, *Chamaecrista ensiformes* e *Senna occidentalis*). As sequências obtidas foram analisadas na plataforma RepeatExplorer para caracterização e comparação da fração repetitiva das três espécies. O genoma de *C. fistula* (0,70 pg/1C) possui 41,74% de DNA repetitivo, *Ch. ensiformes* (0,85 pg/1C) apresenta 32,86% e *S. occidentalis* (0,70 pg/1C) 40,49%. Os elementos repetitivos mais abundantes no genoma das três espécies foram classificados como LTRs, onde a super -família Ty3/Gypsy foi a mais abundante, correspondente a 19,90% em *C. fistula*, 20,05% em *Ch. ensiformes* e 13,29% em *S. occidentalis*. Foram encontradas seis sequências satélites, três em *Cassia fistula*, duas em *Chamaecrista ensiformes* e uma em *Senna occidentalis*, todas são praticamente exclusivas de cada espécie. A variação na abundância e tipos de sequências repetitivas observada nas espécies avaliadas indica o potencial dessas informações para a melhor compreensão das relações entre os gêneros de Cassiinae, a partir da análise de um número maior de espécies.

Palavras-chave: Fabaceae. Cassiinae. Elementos Transponíveis.

ABSTRACT

Fabaceae (Leguminosae) is considered one of the three largest families of Angiospermae, containing 765 genera and 19,500 species with a cosmopolitan distribution. In Brazil, about 2,756 species are found, of which 1507 are endemic. Several groups of legumes have had their classification changed over time, such as the subtribe Cassiinae, which has three genera resulting from the division of *Cassia* lato sensu: *Cassia* stricto sensu, *Chamaecrista* Moench and *Senna* Mill. Based on phylogenetic analyses, the Cassiinae tribe has been considered complex due to conflicting hypotheses regarding the three genera that constitute it. Molecular and cytogenetic studies were important for Cassiinae systematics, but not enough to elucidate the controversies involving the relationship among the three genera. The inclusion of other genomic information is important to contribute to a better understanding of the relationships within the group. The advent of next-generation sequencing (NGS) allowed the characterization and comparison of the repetitive fraction in the genomes, proving to be an interesting tool for understanding genomic evolution. Thus, our work aimed to characterize the repetitive fraction in the genome of species representing the three genera of Cassiinae. We used genome skimming approach (about 0.5x coverage) with sequencing data obtained in Illumina platform HiSeq™ for one species of each genus (*Cassia fistula*, *Chamaecrista ensiformes* e *Senna occidentalis*). The data were analyzed in RepeatExplorer pipeline for characterization and comparison of repetitive fraction of the species. The genome of *C. fistula* (0.70 pg/1C) has 41.74% of repetitive DNA; *Ch. ensiformes* (0.85 pg/1C) has 32.86% and *S. occidentalis* (0.70 pg/1C) 40.49%. Retrotransposons LTRs are the most abundant elements, with the Ty3/Gypsy superfamily being the most abundant in all three species, corresponding to 19.90% in *C. fistula*, 20.05% in *Ch. ensiformes* and 13.29% in *S. occidentalis*. Six satellite sequences were identified, three in *C. fistula*, two in *Ch. ensiformes* and one in *S. occidentalis*. These sequences are almost exclusive of the species in which they were identified. The variation in abundance and types of repetitive sequences observed in the three species evaluated points to the potential of this information for understanding relationships among the genera of Cassiinae in studies involving more species representing genome diversity.

Keywords: Fabaceae. Cassiinae. Transposable Elements.

SUMÁRIO

1 INTRODUÇÃO	8
2 REFERENCIAL TEÓRICO	10
2.1 ASPECTOS SISTEMÁTICOS, BOTÂNICOS E EVOLUTIVOS DE CASSIINAE.	10
2.2 ASPECTOS CITOGENÉTICOS DE CASSIINAE.....	13
2.3 SEQUÊNCIAS DE DNA REPETITIVO.....	16
2.3.1 Sequências de DNA repetitivo dispersas - Elementos Transponíveis (ETs)	17
2.3.2 Elementos Transponíveis de Classe I – Transposição por RNA	18
2.3.3 Elementos Transponíveis de Classe II – Transposição por DNA.....	20
2.3.4 Sequências de DNA repetitivo em tandem – DNA Satélite.....	21
2.3.5 Caracterização da fração repetitiva.....	22
3 MATERIAL E MÉTODOS	26
3.1 MATERIAL VEGETAL.....	26
3.2 ANÁLISE IN SILICO DE DNA REPETITIVO.....	26
3.2.1 Extração de DNA genômico e sequenciamento.....	26
3.2.2 Pré-processamento de dados e análise de <i>clustering</i>.....	27
3.2.3 Identificação e quantificação dos elementos repetitivos	27
3.2.4 Repetições em tandem.....	28
4 RESULTADOS	30
5 DISCUSSÃO	38
6 CONCLUSÃO.....	43
REFERÊNCIAS	44

1 INTRODUÇÃO

Fabaceae Lindl. é a terceira maior família de plantas com flores depois de Orchidaceae e Asteraceae, formada por mais de 760 gêneros e mais de 19.500 espécies. A família possui representantes em todos os tipos de vegetação em torno do globo e é a segunda família mais importante depois das gramíneas em relação ao valor econômico, devido às suas espécies serem usadas como fonte de alimento, forragem, combustível, remédios, madeira e enriquecimento do solo por fixação de nitrogênio (LPWG, 2017).

Atualmente, Fabaceae encontra-se dividida em seis subfamílias: Caesalpinioideae, Cercidoideae, Detarioideae, Dialioideae, Duparquetioideae e Papilionoideae (LPWG, 2017). A subfamília Caesalpinioideae, polifilética e tida como grupo ancestral dentre as leguminosas, é formada por cinco tribos com várias linhagens agrupadas sem uma relação filogenética bem resolvida (DOYLE et al., 2003).

A subtribo Cassiinae, da tribo Cassieae e subfamília Caesalpinioideae, é composta por três gêneros *Cassia* L. sensu stricto, *Senna* Mill. e *Chamaecrista* Moench (SINGH, 2001). Cassiinae foi definida a partir da elevação do gênero *Cassia s.l.* a nível de subtribo, com base em características morfológicas (IRWIN; BARNEBY, 1981).

Esta segregação de *Cassia s.l.* em três gêneros distintos já havia sido proposta por Bentham (1871) e é atualmente reconhecida por meio de estudos com caracteres morfológicos (OWENS; LEWIS, 1989), moleculares (ACHARYA et al., 2011; HERENDEEN et al., 2003; TORRES et al., 2011) e citogenéticos (SOUZA; BENKO-ISEPPON, 2004; BIONDO et al., 2005b, BIONDO et al., 2006). Apesar disso, ainda há controvérsias quanto às relações de parentesco entre os gêneros dentro da subtribo.

Alguns estudos sugerem uma maior afinidade entre *Cassia* e *Senna* e *Chamaecrista* posicionada em um clado distinto (MARAZZI et al., 2006; TORRES et al., 2011). Entretanto, outros trabalhos sustentam *Senna* como grupo irmão de *Chamaecrista*, mas não relacionado a *Cassia* (KAJITA et al., 2001; HERENDEEN et al., 2003).

O tamanho do genoma, valor C, é uma característica importante no estudo de grupos de plantas. A enorme variação dentro do grupo das angiospermas se deve principalmente a eventos de poliploidia e amplificação de sequências repetitivas (PELLICER et al., 2018). Na ausência de poliploidia, as alterações na quantidade de DNA repetitivo são as principais causas pelas diferenças de tamanho do genoma entre as espécies (DU et al., 2017). Em espécies de Cassiinae existe ampla variação do conteúdo C de DNA, com relatos de (0,70 a 1,80 pg/1C), em espécies de *Cassia*, *Senna* e *Chamaecrista* (OHRI, 1986).

Sequências repetitivas de DNA representam uma grande porção do genoma em eucariotos superiores chegando, em alguns casos, até 90% do tamanho do genoma (MEHROTRA; GOYAL, 2014). Essas sequências são divididas em dois grupos, os DNAs satélites e os elementos transponíveis dispersos (BISCOTTI; OLMO; HESLOP-HARRISON, 2015). Essa fração repetitiva pode ser específica da espécie, do genoma ou até mesmo de cromossomos, pois algumas repetições são altamente conservadas enquanto outras evoluem mais rapidamente, apresentando diferenças entre espécies relacionadas (MEHROTRA; GOYAL, 2014; GARRIDO-RAMOS, 2015).

Uma forma de obter uma compreensão dos processos que interferem na abundância e distribuição da fração repetitiva nos genomas de plantas é através de uma caracterização detalhada dessas sequências entre espécies próximas (MACAS et al., 2011). O advento do sequenciamento de próxima geração (NGS) e ferramentas de bioinformática, como a abordagem de agrupamento baseado em gráficos, permitiram a caracterização abrangente de repetições de espécies de plantas não modelo a baixo custo e sem necessidade de um genoma montado (NOVÁK et al., 2020). Esta abordagem tem se provado eficiente para a identificação e caracterização das sequências repetidas, bem como para estudos comparativos de evolução do genoma de eucariotos (DODSWORTH et al., 2015; PAGAN et al., 2012; PELLICER et al., 2014; CAI et al., 2014).

Em *Senna tora*, a fração repetitiva foi estimada em 53,9% do genoma, com predominância de retrotransposons de repetição terminal longa (LTR), principalmente elementos Gypsy, que ocupam 15,6% do genoma (KANG et al., 2020). Waminal e colaboradores (2021) identificaram e localizaram, por hibridização *in situ* fluorescente, as principais repetições em tandem no genoma de *Senna tora*. Os resultados sugerem que as repetições em tandem desempenham papel importante na disploidia verificada em *S. tora* e que há o envolvimento de espaçadores intergênicos de rDNA 45S no "transporte" de repetições durante a remodelação do genoma.

Apesar do potencial do conhecimento sobre a fração repetitiva para o melhor entendimento dos mecanismos de evolução genômica e suas implicações para a taxonomia de grupos relacionados (MACAS et al., 2015), na subtribo Cassinae apenas o genoma de *Senna tora* foi estudada com essa abordagem. Diante disso, o presente trabalho tem como objetivo caracterizar a fração repetitiva em uma espécie de cada gênero de Cassiinae e compará-las para obter informações que possam auxiliar na melhor compreensão de padrões evolutivos do genoma de *Cassia fistula*, *Chamaecrista ensiformes* e *Senna occidentalis*.

2 REFERENCIAL TEÓRICO

2.1 ASPECTOS SISTEMÁTICOS, BOTÂNICOS E EVOLUTIVOS DE CASSIINAE

A família das leguminosas contém mais de 19.500 espécies distribuídas em 765 gêneros, 36 tribos e 6 subfamílias encontradas em todo o mundo, sendo a terceira maior família de angiospermas em diversidade de espécies. Possui alta variabilidade, desde pequenas ervas anuais até árvores de grande porte e são frequentes nos biomas tropicais e temperados (LPWG, 2017).

Muitas espécies de leguminosas são economicamente importantes, como fonte de proteínas vegetais altamente nutritivas para humanos e animais (DURANTI, 2006; VOISIN et al., 2014). Além disso, 88% das espécies de leguminosas têm a capacidade de estabelecer associações com bactérias fixadoras de nitrogênio por meio de nódulos radiculares e, portanto, são importantes para a agricultura sustentável e fundamentais para o funcionamento de alguns ecossistemas (LPWG, 2013; SPRENT, et al. 2017).

Atualmente, Fabaceae está dividida em seis subfamílias Caesalpinioideae, Cercidoideae, Detarioideae, Dialioideae, Duparquetioideae e Papilionoideae (LPWG, 2017). No Brasil é a família de plantas superiores com a maior diversidade na flora, com 2.756 espécies, das quais 1.507 são endêmicas (BFG, 2018).

A partir do aperfeiçoamento das técnicas moleculares, estudos confirmaram o monofiletismo de Fabaceae. No entanto, evidenciaram a importância de correções na classificação das subfamílias, com ênfase no grupo parafilético das Caesalpinioideae (BRUNEAU et al., 2013; LPWG, 2013).

Caesalpinioideae possui aproximadamente 4.400 espécies distribuídas em 148 gêneros e encontradas desde locais temperados a regiões secas e úmidas (AZANI et al., 2017). As espécies possuem folhagem comumente composta, bipinada ou pinada, inflorescências espiciformes ou capituliformes dispostas em apenas um eixo. Suas flores são frequentemente zigomorfas, de simetria radial. Os frutos são do tipo legume e as sementes sem a presença de ranhura hilar (BARROSO et al., 1999).

Caesalpinioideae, considerada a subfamília mais primitiva, da qual derivaram as subfamílias Papilionoideae e Mimosoideae (BURKANT, 1987) é a menos entendida e estudada entre as subfamílias de Fabaceae (HERENDEEN, 2000). Vários estudos biosistemáticos foram realizados com este grupo, nos quais foi confirmada a hipótese de que Caesalpinioideae é um grupo não natural (HERENDEEN, 2000; BRUNEAU et al., 2001; KAJITA et al., 2001).

A subfamília se encontra dividida em quatro tribos Cercideae, Caesalpinieae, Cassieae e Detarieae (LEWIS et al., 2005).

A tribo Cassieae é polifilética, com aproximadamente 720 espécies alocadas em 21 gêneros. As espécies possuem ampla distribuição e são facilmente reconhecidas pela combinação de caracteres morfológicos: altura variada desde grandes árvores a pequenas ervas, com folhas compostas, nectários extraflorais no pecíolo, corola amarela, comumente assimétrica, estames heterodínamos, anteras basifixas e fruto do tipo legume (LEWIS et al., 2005; QUEIROZ, 2009).

A partir de características morfológicas do androceu, arquitetura floral, presença ou ausência de bractéolas, Irwin e Barneby (1982) propuseram a elevação do gênero *Cassia* sensu lato a nível de subtribo. Atualmente essa subtribo, denominada Cassiinae, pertencente à tribo Cassieae subfamília Caesalpinioideae, é formada por três gêneros *Cassia* L., *Chamaecrista* Moench e *Senna* Mill (LEWIS, 2005).

Análises filogenéticas revelaram que Cassiinae constitui um grupo bastante complexo, havendo hipóteses conflitantes em relação aos três gêneros que a compõem. Os estudos conduzidos por Bruneau et al. (2008), Manzanilla e Bruneau (2012) sugerem uma maior afinidade entre *Cassia* e *Senna*, onde *Chamaecrista* faz parte de um clado distinto, o que corrobora com os estudos conduzidos por Marazzi et al. (2006). Já outros estudos apoiam *Senna* como grupo irmão de *Chamaecrista*, mas não relacionado a *Cassia* (HERENDEEN et al., 2003).

Apesar de Cassiinae ser reconhecida como monofilética (LEWIS, 2005), análises moleculares baseadas tanto em sequências de DNA cloroplastidial como em marcadores (RAPD e AFLP) obtiveram resultados controversos em relação à monofilia da subtribo e à relação entre os três gêneros que a constituem (KAJITA et al., 2001; ACHARYA; MUKHERJEE, 2011).

Cassia L. *sensu stricto* é composto por cerca de 30 espécies, das quais 11 estão amplamente distribuídas no Brasil (SOUZA; BORTOLUZZI, 2014). Seus frutos costumam ser indeiscentes, ao contrário do gênero *Senna*, que são tipicamente deiscentes e sem septos separando as sementes (LEWIS, 2005). As vagens indeiscentes e lenhosas de *Cassia* L. são bastante resistentes, assim como as sementes. Além disso, os frutos de *Cassia* geralmente não são carnudos, mas com resquícios de polpa.

Muitas espécies de *Cassia* têm valor medicinal e podem diferir entre si a partir da presença de constituintes fitoquímicos (LAVANYA et al., 2018). As folhas, caules, vagens e sementes de espécies de *Cassia* têm demonstrado possuir valores farmacológicos diferentes

(FEITOSA et al., 2011). Entre as aplicações etnofarmacológicas conhecidas estão o tratamento de doenças de pele como eczema, micose, sarna e hanseníase (SUNDARAMOORTHY et al., 2016). Estudos *in vitro* também mostraram seu possível uso como um tratamento para distúrbios do cérebro humano, como epilepsia e enxaqueca (DESHPANDE; BHALSING, 2013; BALKRISHNA; MISRA, 2017). *C. fistula*, por exemplo, também conhecida como chuva-de-ouro, é muito apreciada pela beleza de seus cachos de flores amarelas e pelas propriedades laxativas de suas vagens, folhas e flores, sendo também indicada em diversos países para tratamento de infecções bacterianas (TIWARI; MISRA, 1985), reumatismo e lesões cutâneas (BHAKTA et al., 1999).

Senna é um dos gêneros com maior diversidade dentro da família Fabaceae, com aproximadamente 350 espécies de árvores, arbustos e subarbustos distribuídas pelos continentes americano, africano e australiano, com ocorrências também na Ásia (MARAZZI et al., 2006). O gênero é dividido em seis seções: *Psilorhegma*, *Chamaefistula*, *Senna*, *Peiranisia*, *Paradyction* e *Astroites* (IRWIN; BARNEBY, 1982). *Senna* se difere de *Cassia* por apresentar filetes retos, mais curtos ou até duas vezes o comprimento das anteras, que são basifixas, e pela presença de glândulas extraflorais na maioria das espécies. Com relação a *Chamaecrista*, *Senna* se difere principalmente pela ausência de bractéolas, pelo androceu zigomorfo e pelos legumes que podem ser indeiscentes (IRWIN; BARNEBY, 1982).

Espécies de *Senna* são usadas como ervas medicinais e fonte de madeira para exportação, contribuindo significativamente como matéria prima para produtos comerciais (SEETHAPATHY et al., 2014). No Brasil, o gênero é representado por 82 espécies, sendo 27 delas endêmicas do país, distribuídas em todos os domínios fitogeográficos, com maior predominância na Mata Atlântica, Caatinga e Cerrado (BFG, 2015). Várias espécies do gênero são usadas na medicina popular como laxantes ou para tratar resfriados (SILVA et al., 2010).

Senna occidentalis é uma espécie usada para o controle de doenças tropicais. O extrato de suas folhas é usado como anti-inflamatório e bactericida. Infusões das folhas são utilizadas em alguns países africanos como purgativo, diurético e para o tratamento da febre. Existem relatos de que cada parte da planta possui atividade e aplicação distintas (VILA-NOVA et al., 2012).

O gênero *Chamaecrista*, anteriormente definido como *Lasiorhegma*, subgênero de *Cassia*, compreende aproximadamente 330 espécies, das quais 266 são nativas das Américas (LEWIS, 2005). No Brasil, ocorrem 257 espécies (BFG, 2015). *Chamaecrista* é o único gênero dentro de *Cassiinae* com nectários extraflorais côncavos e raízes com nódulos bacterianos (IRWIN; BARNEBY, 1982). Esse gênero pode ser reconhecido pelo pedicelo 2-bracteolado,

próximo ou acima da metade de seu comprimento total, androceu actinomorfo, com anteras pubescentes ao longo das suturas e nectário extrafloral em forma de disco ou taça (QUEIROZ, 2009; RANDO et al., 2013).

Chamaecrista ensiformes possui porte arbóreo, com uma altura que pode atingir até sete metros, racemos caulifloros e estípulas caducas. Esta espécie é endêmica do Brasil (FILARDI et al., 2007), distribuída nas regiões Nordeste, Centro Oeste e Sudeste e com ocorrência sempre associada às áreas mais abertas, como clareiras e bordas de mata (CESTARO, 2002).

2.2 ASPECTOS CITOGENÉTICOS DE CASSIINAE

O grupo das leguminosas constitui um modelo interessante para investigar os processos de diversificação e evolução, inclusive do cariótipo. O período entre a origem e diversificação do grupo foi estimado como curto, com a divergência das linhagens principais há cerca de 60 milhões de anos (LAVIN et al., 2005).

Dentro das leguminosas, as espécies da subfamília Caesalpinioideae apresentam ampla variabilidade intergenérica, interespecífica e intraespecífica em relação ao número de cromossomos, onde já foram encontrados $2n = 14, 16, 22, 26, 28, 32, 52$, sendo $2n = 28$ o mais comum. A poliploidia parece ser importante na diversificação inicial do grupo, com ocorrência de uma série de reduções dispolides no decorrer do processo evolutivo (BIONDO, 2005). A subfamília pode ser considerada um grupo ancestral dentro das leguminosas, pelo fato das suas espécies possuírem cromossomos pequenos e com cariótipos relativamente simétricos (SOUZA; BENKO-ISEPPON, 2004).

A tribo Cassieae possui espécies com uma ampla variação de conjuntos cromossômicos (GOLDBLATT, 1981). A maioria das espécies analisadas na tribo possuem cromossomos pequenos, com cerca de 2,5 μm e predomínio de cromossomos metacêntricos (BIONDO et al., 2005). Esta tribo apresenta espécies tetraploides, cujo número básico é $x = 14$, derivado por poliploidia de $x = 7$ (POLHILL et al., 1981).

Para as espécies de Cassiinae, as informações sobre o complemento cromossômico indicam que a subtribo possui espécies com número básico $x = 14$, além de $x = 11, 12$ e 13 , derivados por dispolidia (BIONDO et al., 2005b).

O número cromossômico de espécies de *Cassia* L é predominantemente $2n = 28$ cromossomos (IRWIN; TURNER, 1960; BIONDO et al., 2005). Porém também é citado $2n = 26$ para *C. fistula* L. (GOLDBLATT; JOHNSON, 1996) e para *C. fastuosa* (SOUZA; BENKO-

ISEPPON, 2004). As variações são provavelmente devidas ao processo de disploidia, comumente encontrado nas espécies da tribo Cassieae (GOLDBLATT, 1981a; SOUZA; BENKO-ISEPPON, 2004).

Os conjuntos cromossômicos registrados para *Chamaecrista* revelam uma ampla variação numérica ($2n=14, 16, 22, 24, 26, 28, 32, 42, 48, 52$ e 56), sendo $2n = 16$ e 28 os mais frequentes. É observado que tanto a poliploidia quanto a disploidia parecem ter atuado na evolução cariotípica do grupo, com intensidades diferentes dependendo do clado em questão. Os números $x = 7$ e 8 são propostos na literatura como números básicos para o gênero, o que poderia ser explicado pelos eventos de poliploidia e disploidia associados (CONCEIÇÃO, 2006). No gênero são observados números $x = 8$ nas seções *Chamaecrista* e *Caliciopsis*, $x = 7$ na seção *Xerocalyx* e $x = 14$ nas seções *Absus* e *Grimaldia* (BIONDO et al., 2006).

O caráter tamanho dos cromossomos sugere, juntamente com outros caracteres analisados, a segregação de *Chamaecrista* dos demais gêneros pertencentes à subtribo. Geralmente os cromossomos de *Cassia* L e *Senna* Mill são pequenos ($2,5 \mu\text{m}$) e predominantemente metacêntricos (BIONDO et al., 2005b; RESENDE et al., 2013). Em *Senna* e *Chamaecrista* foram observados cromossomos com variação de tamanho de $0,62$ até $2,5 \text{ mm}$ (SOUZA; BENKO-ISEPPON, 2004). Braz (2014) analisou o tamanho médio dos cromossomos em seções do gênero *Chamaecrista* onde foi encontrada uma variação entre $2,10 \mu\text{m}$ e $3,66 \mu\text{m}$. Em *Ch. cipoana* e *Ch. semaphora* foram observados cromossomos com $5,04 \mu\text{m}$ e $7,18 \mu\text{m}$, respectivamente, o que não confirma a tendência descrita na literatura de que os cromossomos desse gênero são pequenos.

Diferenças no tamanho dos cromossomos de duas populações de *Chamaecrista nictitans* parecem estar relacionadas com a quantidade de heterocromatina. Enquanto uma população possui cromossomos maiores, outra possui alguns pares menores. Após a coloração com fluorocromo, houve constatação de diferenças no número de bandas $\text{CMA}^+ / \text{DAPI}^-$ entre as populações em estudos. É possível que, como os grupos de populações estão isolados uns dos outros, tenha ocorrido divergência quanto à evolução da heterocromatina (SOUZA; BENKO, 2005).

Braz (2014) caracterizou o complemento cromossômico (número e morfometria), estimou o tamanho do genoma e localizou as regiões codificadoras dos RNAs ribossômicos 45S e 5S nos cromossomos de espécies representantes de diferentes seções e/ou séries de *Chamaecrista* coletadas na região de Diamantina – MG. A variabilidade no número e morfometria cromossômica, número e padrão de distribuição das sequências de DNA ribossomal e no tamanho do genoma revelaram que *Chamaecrista* é um grupo heterogêneo do

ponto de vista cariológico. Aparentemente a redução dispoloide está envolvida na evolução cariotípica do gênero. Rarranjos do tipo inversão, translocação e deleção também podem ser considerados como fatores que levaram a alterações estruturais nos cromossomos, diminuição na quantidade de DNA e diminuição no número de locos de rDNA.

Souza e Benko (2004) mostraram que o núcleo interfásico de três espécies de Cassiinae eram do tipo semireticulado, com diferenças no tamanho, número, forma e distribuição dos cromocentros. Em quatro populações de *Chamaecrista nictitans*, a cromatina difusa foi mais clara do que nos táxons restantes, frequentemente apresentando núcleos polarizados com cromocentros difusos. O gênero *Senna* apresentou núcleos polarizados com granulação fina e coloração mais profunda da cromatina difusa, embora a cromatina de *S. quinqueangulata* seja semelhante à de *Cassia fastuosa*, sendo mais granulada. O comportamento de condensação foi parecido para a maioria das espécies analisadas, com exceção das espécies de *Senna* que mostraram condensação homogênea com pequenas porções de condensação distal tardia. A espécie *Senna quinqueangulata* apresentou comportamento semelhante ao de *Cassia fastuosa*, com cromatina de condensação proximal.

O conjunto cromossômico de aproximadamente 20% das espécies de *Senna* já foram descritos, com predominância de $2n = 28$, embora também haja registros de $2n = 22$, 24 e 26 (RESENDE et al., 2014; CORDEIRO; FELIX, 2017). Também há registros de poliploidia no gênero, como $2n = 42$, 56 e 112 em *Senna rugosa* (RESENDE et al., 2014) e $2n = 52$ e 104 em *S. gardneri* (MATOS et al., 2011). A consistência de $2n = 28$ para a maioria das espécies demonstra que o número básico pode ser $x = 14$, enquanto os outros números ($x = 11$, 12 e 13) refletem eventos de disploidia (GOLDBLATT, 1981a).

Em *Senna obtusifolia*, uma população apresentou duas bandas terminais $CMA^+/DAPI^-$, e outra população, duas bandas terminais $DAPI^+/CMA^-$. As diferenças observadas nos padrões de bandas CMA/DAPI em *Senna* mostram a eficácia dessa técnica em facilitar a diferenciação cariotípica de suas espécies (SOUZA; BENKO, 2004).

Cordeiro (2018) relatou variabilidade para distribuição da heterocromatina em dez espécies de *Senna*, tanto com relação ao número quanto localização das bandas. Apenas *S. georgica* e *S. rizzinii* apresentaram o mesmo padrão de bandas (quatro bandas $CMA^+/DAPI^-$ terminais e vinte duas $CMA^+/DAPI^-$ proximais). *S. martiana* e *S. aversiflora* apresentaram o mesmo número de bandas (quatro bandas $CMA^+/DAPI^-$ terminais), embora esta última tenha demonstrado um cariótipo poliplóide ($2n = 52$).

Rezende (2012) caracterizou o núcleo interfásico, o número e a morfometria dos cromossomos de 11 espécies do gênero *Senna*. Houve variabilidade interespecífica com relação

a todas as características avaliadas: tipo de núcleo interfásico, sendo encontrados núcleos arreticulados e semi-reticulados; número cromossômico ($2n = 24, 26, 28$ e 56); morfologia cromossômica (meta e submetacêntricos). Os cariótipos foram predominantemente simétricos. Observou-se variação cromossômica numérica entre células meristemáticas de um mesmo indivíduo para as espécies *S. macranthera*, *S. rugosa* e *S. splendida*. Os resultados sugeriram que mecanismos de poliploidia e disploidia estão envolvidos na evolução cariotípica do gênero *Senna*, mas não forneceram subsídios conclusivos para inferências citotaxonômicas.

Em *Senna occidentalis*, o núcleo interfásico é arreticulado e a condensação cromossômica é homogênea, com porções de condensação distal atrasada na prometáfase. O cariótipo da espécie foi descrito com $2n = 28$ cromossomos, com fórmula $9 m + 5 sm$ e constrição secundária presente em um par de cromossomos. A partir da coloração com nitrato de prata foram detectados quatro nucléolos, o que pode indicar a presença de dois pares cromossômicos com Região Organizadora de Nucléolo (RON). O comportamento meiótico para a espécie foi regular, com alto índice meiótico (FERREIRA et al., 2010).

Em relação ao tamanho do genoma (GS), quantidade de DNA contido no núcleo da célula, também conhecido como valor C, existe grande diversidade entre as angiospermas. Os principais mecanismos que contribuem para a expansão do genoma são a poliploidia e o acúmulo de DNA repetitivo. Normalmente, esses processos são contrabalançados de forma eficiente para evitar a expansão descontrolada dos genomas (LEITCH et al., 2019). Em espécies de Cassiinae os valores C disponíveis em (<https://cvalues.science.kew.org/>) variam de 0,70 a 1,80 pg/1C. Espécies como *Cássia fistula* e *Senna occidentalis* possuem (0,70 pg/1C). Rezende (2012) quantificou o tamanho de genoma de espécies pertencentes ao gênero *Senna*, identificando variação de conteúdo genômico entre 0,95 a 2,63pg/2C. Braz (2014), ao analisar a quantidade C de DNA em espécies do gênero *Chamaecrista*, encontrou variação de 0,44 a 1,08 pg/1C, sendo que *Chamaecrista ensiformes* apresentou 0,85 pg/1C.

2.3 SEQUÊNCIAS DE DNA REPETITIVO

Apesar da grande variação no tamanho do genoma entre os eucariotos, via de regra os genes representam apenas uma pequena fração do material genético nuclear. Essas regiões codificadoras são a parte mais estável do genoma de um organismo, em grande parte devido à seleção que age para preservar a função do gene. Para muitos eucariotos, o espaço genômico remanescente é extremamente dinâmico e predominantemente composto por uma mistura variada de sequências repetitivas (CRAIG et al., 2002).

A fração repetitiva representa sequências DNA que variam amplamente em tamanho e dispersão. Estão presentes em todas as plantas superiores e podem representar até 90% do genoma em algumas espécies (MEHROTRA; GOYAL, 2014). Estas sequências estão intimamente relacionadas à morfologia e dinamismo dos cromossomos, onde participam de importantes mecanismos evolutivos e de diferenciação cariotípica (SILVA et al., 2013). Além de desempenhar papel estrutural, a região formada por sequências repetitivas também desempenha função central na estabilidade do cromossomo, no ciclo celular, na regulação da expressão gênica e é um fator importante para a evolução do genoma (LIU et al., 2019).

Em relação ao conjunto cromossômico, estas sequências estão envolvidas direta ou indiretamente nos eventos de rearranjo estrutural (deleções, duplicações, inversões e translocações) que são responsáveis por uma parte da variação cariotípica observada no processo evolutivo de muitos grupos de organismos (PRAKHONGCHEEP et al., 2017).

Os mecanismos envolvidos na evolução das sequências de DNA repetitivo operam em nível intragenômico e estão diretamente relacionados à organização dessas sequências nos cromossomos. Sequências de DNA repetitivo podem ser divididas em dois grupos, um de sequências em tandem e outro de sequências dispersas. O primeiro é constituído por sequências satélites, cujo os monômeros estão repetidos em tandem, ou seja, as cópias estão dispostas adjacentes umas às outras, comumente nas regiões centroméricas e teloméricas ou próximas a elas. O segundo grupo que consiste em elementos que estão dispersos por todo o genoma, os elementos móveis como transposons de DNA e retrotransposons (BISCOTTI; OLMO; HESLOP-HARRISON, 2015; MLINAREC et al., 2019).

Enquanto algumas classes de sequências repetitivas são conservadas, outras são extremamente dinâmicas durante a evolução de espécies intimamente relacionadas. Assim, o estudo dessa fração repetitiva pode fornecer importantes informações para compreensão do cromossomo e evolução de grupos taxonômicos próximos (RUIZ-RUANO et al., 2016).

2.3.1 Sequências de DNA repetitivo dispersas - Elementos Transponíveis (ETs)

Os elementos transponíveis (ETs) são segmentos de DNA que foram descobertos no final da década de 40, em estudos citogenéticos de milho (MCCLINTOCK, 1950). Após sua descoberta, os ETs eram considerados apenas "DNA lixo" e "parasitas de DNA", sequências simples que se multiplicavam dentro do genoma, mas não forneciam contribuição para seu hospedeiro (DOOLITTLE; SAPIENZA, 1980).

ETs são sequências de DNA de tamanho variável que podem se mover e replicar dentro de um genoma (CASACUBERTA; GONZÁLEZ, 2013). Com isso, tornam-se fontes importantes de mutações genômicas (RHO et al, 2007) e agentes da evolução do genoma e dos genes (ZHAO; FERGUSON; JIANG, 2016; VICIENT; CASACUBERTA, 2017). Constituem matéria-prima para o surgimento de novas sequências regulatórias, como promotores e intensificadores (SPRINGER; LISCH; LI, 2016), alterando a estrutura e o ambiente regulatório e/ou epigenético, gerando uma expressão gênica modificada (MITA; BOEKE, 2016).

Um número significativo de genes em plantas é oriundo de ETs devido a um processo conhecido como “exaptação”. ETs também contribuíram para a evolução de íntrons, exons e promotores (ZHAO; FERGUSON; JIANG, 2016). Os mecanismos pelos quais os ETs podem modificar genes são diversos, sendo mais óbvia a inativação das regiões codificantes ou reguladoras do gene. No entanto, a inserção de um ET dentro de um gene também pode gerar mutações mais sutis, como mudanças na sequência de aminoácidos da proteína codificada, mudanças no padrão de expressão ou novas formas de *splicing* (HUANG et al., 2015). Os elementos transponíveis determinaram e ainda estão determinando variações na estrutura do genoma dos eucariotos, mesmo dentro de uma mesma espécie (NISHIHARA, 2019).

Foi proposto por Wicker et al. (2007) um sistema de classificação e nomenclatura para os ETs, dividindo-os em duas classes principais, retrotransposons (Classe I) e transposons de DNA (Classe II), que diferem entre si em relação ao seu mecanismo de transposição. Os retrotransposons são subdivididos a partir de suas características de transposição, organização e filogenia enzimática. Os transposons de DNA são divididos em duas subclasses com base no número de fitas de DNA que são cortadas durante a transposição. ETs também podem ser separados em dois grupos principais: os autônomos, que possuem região codificadora para a transcriptase ou para a transposase e são capazes de deslocar no genoma por si próprios, e os não-autônomos, que possuem alguma mutação na região codificante da transcriptase ou transposase e, portanto, incapazes de se deslocar sem o auxílio de ETs autônomos (WERREN, 2011).

2.3.2 Elementos Transponíveis de Classe I – Transposição por RNA

Os elementos transponíveis pertencentes à classe I se propagam através da ação da enzima transcriptase reversa que tem a capacidade de sintetizar uma molécula de DNA a partir de RNA, gerando novas cópias de elementos que, após a integração, aumentam o tamanho do genoma do hospedeiro. Nenhum segmento é cortado e transferido para outra região do genoma.

A classe I foi subdividida em duas grandes categorias que se distinguem pela presença ou ausência de longas repetições terminais: LTR (*Long Terminal Repeat*). Porém existem outras categorias de retrotransposons eucarióticos que são distinguíveis pelo mecanismo de transposição ou características filogenéticas da enzima transcriptase reversa, como os DIRS (*Dictyostelium Intermediate Repeat Sequence*) e *Penelope-Like Elements* (PLEs) (JURKA et al., 2007).

A sequência entre as extremidades dos RT-LTRs representa a fração codificadora do elemento e inclui os genes *Gag* e *Pol*. *Gag* codifica uma cadeia polipeptídica semelhante ao capsídeo viral, enquanto *Pol* codifica as enzimas necessárias para o processamento do RNA do retrotransposon, que inclui: uma transcriptase reversa, uma integrase, uma RNase H e uma protease (KAZAZIAN, 2011). Outras características estruturais dos RT - LTRs são um sítio de ligação do primer e um fragmento de poli-A, ambos envolvidos na replicação do elemento (WICKER et al., 2007).

Christoph et al. (2020) caracterizaram e investigaram a composição de famílias de RT - LTRs presentes no genoma de *Brachypodium distachyon*. Elementos da família *Copia* foram os mais dinâmicos do genoma da espécie e uma fonte importante de polimorfismos dentro da mesma. O estudo mostrou como algumas linhagens de retrotransposons mudaram a constituição genética e epigenética da espécie, além de apenas influenciar no tamanho do genoma.

Retrotransposons *non - LTR* são elementos antigos que evoluíram e permaneceram em genomas eucariotos por milhões de anos (CRAIG et al., 2002). Como a própria nomenclatura pressupõe, são elementos que não possuem longas repetições terminais. São autônomos, com uma ou duas estruturas de leitura aberta – ORFs e comumente chamados de elementos nucleares intercalados longos (LINES) (WICKER et al., 2007). Os elementos intercalados curtos (SINEs) são elementos móveis que podem se mover com o auxílio de elementos intercalados longos (LINES). SINEs são compostos de várias regiões de origem independente, chamadas cabeça, corpo e cauda. Essas três regiões nem sempre estão presentes em todos os SINEs. As cabeças dos SINEs normalmente se originam de RNAs não codificantes, como 7 SL RNA, tRNA, 5S rRNA ou pequenos RNA nucleares, onde servem principalmente como regiões promotoras para o início da transcrição (KAPITONOV; JURKA, 2003). SINEs com cabeças originadas de tRNA, chamados SINE2, são os mais comuns e amplamente distribuídos entre os eucariotos (KOJIMA, 2015).

As características estruturais que distinguem os elementos DIRS dos demais são a presença de uma região codificadora para tirosina recombinase, mas não para integrase ou protease. Diferente dos LTRs, os DIRS possuem suas repetições terminais invertidas. Essa

estrutura incomum das repetições terminais dos elementos do tipo DIRS parece estar relacionada à sua replicação por meio de intermediários circulares livres (POULTER; GOODWIN, 2005).

Penelope foi descrito pela primeira vez em *D. virilis* (EVGEN'EV et al., 1997). Por causa de suas repetições terminais longas, foi considerado primeiramente como membro dos RT-LTR, embora suas características sejam diferentes de outros RT-LTR. A presença de endonuclease do tipo GIY-YIG a jusante do domínio da transcriptase reversa levou a classificar *Penelope* e seus parentes como um novo grupo de retrotransposons (LYOZIN et al., 2001).

2.3.3 Elementos Transponíveis de Classe II – Transposição por DNA

Os transposons de classe II codificam uma transposase, enzima que permite a transposição destes elementos, tanto em eucariotos quanto em procariotos. As transposases possuem domínios conservados que basicamente atuam cortando o elemento transponível de um determinado sítio para posterior inserção em outra região do genoma do hospedeiro, sendo o reconhecimento mediado por regiões específicas das duas extremidades do elemento (VARANI et al., 2015). Esse mecanismo de transposição é conhecido como conservativo, por basicamente cortar e colar o elemento em outra região do genoma. As transposases podem reconhecer sequências das extremidades de transposons diferentes e conseqüentemente imobilizá-los, o que é comum em elementos MITES (WICKER et al., 2007).

Os MITES são ETs de sequências curtas, originadas a partir de elementos autônomos que ao sofrerem deleções perderam a região codificante da transposase, porém permaneceram com suas extremidades conservadas. Essa característica estrutural permite o reconhecimento dos MITES pela transposase de um ET autônomo com características relacionadas (WICKER et al., 2007; VARANI et al., 2015).

Os transposons de DNA são fragmentos grandes com tamanhos de centenas a milhares de pares de bases. Elementos simples de procariotos, chamados de sequências de inserção (IS), são representantes dos transposons de DNA. Os IS são separados em duas subclasses pelo número de fitas de DNA que são cortadas durante o mecanismo de transposição.

A subclasse I utiliza o mecanismo de clivar e colar as duas fitas de DNA e as duas ordens de maior representação são a TIR e *Crypton* (WICKER et al., 2007). A subclasse II corta e cola apenas uma das fitas da molécula de DNA e suas ordens de maior representação são *Maverick* e *Helitron*. A ordem *Helitron* possui elementos nos quais a transposição ocorre através de um

mecanismo chamado “círculo rolante” onde apenas uma das fitas de DNA é cortada (KAPITONOV; JURKA, 2003).

2.3.4 Sequências de DNA repetitivo em tandem – DNA Satélite

A existência de DNAs repetitivos em genoma de eucariotos foi descrita pela primeira vez na década de 60. A partir do DNA genômico de camundongo ultra centrifugado em meio a cloreto de cério foram visualizadas duas bandas com densidades diferentes, a banda de menor peso representava cerca de 10% do genoma e foi chamada de DNA satélite. O DNA desta banda sofreu taxas de renaturação mais altas do que o restante do DNA nuclear ou do que genomas mais simples compostos principalmente de sequências únicas, indicando que o genoma de camundongos possuía repetições pelo genoma (KIT et al., 1961).

Com o aperfeiçoamento das técnicas moleculares foi possível ampliar o conhecimento sobre os DNAs satélites, mostrando sua função na estruturação do centrômero e telômeros, montagem do cinetócoro, segregação cromossômica e reconhecimento dos cromossomos homólogos durante a meiose. Além desses aspectos funcionais, eventos de amplificação ou deleção podem causar mutação em grupos de DNA satélite, ocasionando alterações evolutivas entre espécies (PLOHL et al., 2014). Transcritos de DNA satélite também desempenham outras funções, tais como regulação genômica e manutenção da heterocromatina (KOO et al., 2016).

De acordo com a hipótese desenvolvida por Ruiz-Ruano et al. (2016), o surgimento de uma sequência satélite ou de uma nova classe pode ocorrer pela duplicação de um ponto único no genoma, onde dois ou mais pares de bases originam uma curta repetição em tandem. Esse evento pode ocorrer por vários mecanismos, como, por exemplo, o deslizamento de uma fita da molécula de DNA durante a replicação ou a inserção de cópias replicadas de DNA por intermediários extracromossômicos.

Sequências de DNA satélite participam de um processo evolutivo constante que ocorre em nível molecular e que acarreta duas mudanças principais no genoma: o número de cópias e a composição nucleotídica. Tais mudanças tendem a afetar sequências únicas ou conjuntos de sequências e podem levar à formação de perfis de DNA satélite específicos de cada espécie. Em alguns casos, essas mudanças podem ser correlacionadas com a evolução cariotípica e influenciar no processo evolutivo das espécies (UGARKOVIC; PLOHL, 2002).

As sequências repetitivas em tandem foram agrupadas em três classes de acordo com o tamanho do segmento repetido: Microssatélites, minissatélites e satélites (BISCOTTI; OLMO; HESLOP-HARRISON, 2015).

Os satélites são as repetições em tandem de sequências de tamanho maior ou igual a 100 nucleotídeos (GRIFFITHS et al., 1996). Normalmente, estas sequências encontram-se organizadas em unidades muito extensas, presente em maior concentração em regiões heterocromáticas pericentromérica e subtelomérica. São encontradas também em regiões intersticiais dos cromossomos em posições específicas dos braços e também entre as regiões subtelomérica e pericentromérica, especialmente em invertebrados e plantas (LÓPEZ; GARRIDO, 2012).

Microsatélites, ou repetições de sequência simples (SSRs), são repetições formadas por arranjos de até 6 pares de bases de comprimento variável e de distribuição por todo o genoma (CUADRADO et al., 2008). Dado seu alto grau de polimorfismo, os SSRs têm sido amplamente usados como marcadores moleculares em mapeamento, impressão digital de DNA e estudos de evolução genética. A maioria dos estudos tem se baseado na análise de SSRs localizados em regiões codificantes da eucromatina nas quais o tamanho das repetições sequenciadas é em torno de 100 pares de bases. No entanto, experimentos de hibridização fluorescente *in situ* (FISH) mostraram que algumas regiões formadas por SSRs em genomas de plantas são organizadas em grupos, geralmente associados com a heterocromatina constitutiva (CUADRADO; SCHWARZACHER, 1998). O mapeamento cromossômico de microsatélites possui enorme valor para a elucidação da diferenciação cariotípica de espécies evolutivamente próximas (TRALDI et al., 2016).

Em eucariotos, os genes 18S, 5.8S e 26S (referidos como rDNA 35S em plantas) são transcritos como um único operon e são separados por espaçadores transcritos internos (ITS) e espaçadores intergênicos (IGS). Citogeneticamente, os locos de rDNA 35S são agrupados formando domínios cromossômicos distintos chamados de regiões organizadoras de nucléolo (NORs) (NEVES et al. 2005). Apesar do alto número de cópias (centenas a milhares de unidades nas plantas), os genes são quase idênticos, o que é explicado por sua evolução combinada. Sob tal processo evolutivo, se toda a unidade 35S rDNA for igualmente afetada pela evolução em concerto, então é esperado observar pouca ou nenhuma diferença no grau de homogeneidade intragenômica entre regiões codificantes e não codificantes (EICKBUSH, 2007).

2.3.5 Caracterização da fração repetitiva

A evolução do genoma envolve vários processos, como por exemplo, poliploidia, deleções ou duplicações de segmentos cromossômicos, reestruturação cromossômica e

abundância de sequências repetitivas (RUPRECHT et al., 2017). O DNA repetitivo constitui uma fração que corresponde entre 25 e 90% do genoma das plantas, onde varia tanto em padrões de estrutura quanto na dispersão no genoma (BENNETZEN; WANG, 2014).

Uma maneira de obter uma compreensão abrangente dos processos que moldam a composição repetitiva dos genomas das plantas é conduzir uma caracterização detalhada do DNA repetitivo em várias espécies, seguida pela análise dentro de uma estrutura filogenética. Esses estudos eram escassos devido às grandes quantidades e considerável complexidade das sequências repetitivas que precisariam ser analisadas. No entanto, muito progresso foi feito nos últimos anos devido à introdução de tecnologias de sequenciamento de próxima geração (NGS) e abordagens de bioinformática correspondentes (WEISS-SCHNEEWEISS et al., 2015).

Ferramentas de bioinformática baseadas no agrupamento de sequências em gráficos é um método de identificação de elementos repetitivos a partir do sequenciamento de próxima geração (NGS) (MACAS et al., 2015). O RepeatExplorer é uma plataforma desenvolvida com o objeto de identificar elementos de repetição *de novo* através de um algoritmo de agrupamento de sequências similares em clusters. Esses clusters compreendem principalmente sequências repetidas, pois apenas as leituras derivadas de sequências presentes no genoma várias vezes podem produzir um número suficiente de ocorrências de similaridade em dados de sequenciamento genômico de baixa cobertura. O número de leituras em cada cluster é proporcional à abundância de repetições, assim, possibilita sua quantificação (NOVAK et al., 2013).

Ao estudar o impacto do DNA repetitivo na evolução do genoma, as angiospermas estão entre os melhores modelos de plantas pois o grupo exibe uma variação extraordinária na abundância dessa fração genômica (FLEISCHMANN et al., 2014).

Neste contexto, Dodsworth et al. (2015) testaram a utilidade da abundância de diferentes elementos repetitivos, classificados por meio de análises bioinformáticas, com o intuito de propor uma nova abordagem filogenética. Os genomas de angiospermas e de insetos foram utilizados e as árvores filogenéticas inferidas com base na abundância da fração repetitiva dos genomas onde foram tratados como caracteres variáveis. Esta metodologia forneceu relações bem suportadas em níveis interespecíficos e intergenéricos que concordam com os resultados de análises filogenéticas mais comumente usadas. Essa abordagem pode ser especialmente útil em grupos onde há pouca diferenciação genética, onde pode fornecer dados adicionais para inferência filogenética, e assim, produzir dados para estudos comparativos da evolução do genoma.

Macas et al. (2015) utilizaram o sequenciamento de baixa cobertura para identificar e quantificar a fração repetitiva do DNA de vinte e três espécies de quatro gêneros pertencentes a tribo Fabeae usando a plataforma RepeatExplorer. O DNA repetitivo foi responsável por 85% da diferença encontrada no tamanho do genoma das espécies, e 57% desta variação foi determinada por uma única linhagem de retrotransposon. Alguns grupos de retrotransposons foram encontrados apenas entre espécies filogeneticamente próximas. A presença desses grupos possui um certo grau de correspondência dentro de espécies mais estreitamente relacionadas do que com todas as Fabeae.

Em um contexto citogenético, o mapeamento de sequências repetitivas por hibridização *in situ* fluorescente (FISH) foi aplicado com sucesso em muitos estudos, especialmente quando se trata de espécies relacionadas. Informações genômicas em conjunto com a citogenética é uma abordagem eficiente para compreender o mais alto nível de organização estrutural e funcional dos cromossomos (LIANG; CHEN, 2015).

Sucre et al. (2020) realizaram a caracterização da fração repetitiva no genoma de *Erythrostemon hughesii*, espécie pertencente ao grupo Caesalpinia (Fabaceae), seguida da análise da distribuição dos elementos mais abundantes nos cromossomos. A fração repetitiva de *E. hughesii* representa 28,73% do genoma. A redução da abundância dos elementos LTR *Ty3/Gypsy* da linhagem *Tekay* no genoma de *E. hughesii* está correlacionada com a perda de bandas heterocromáticas CMA+ proximal e com a invasão de várias outras repetições em regiões subtelo méricas nos cromossomos. A variabilidade de sequências repetitivas em espécies do grupo Caesalpinia pode estar relacionado com a heterogeneidade da heterocromatina, um fator de divergência genômica, que é o principal pré-requisito para o processo de especiação.

No genoma de *Senna tora* foi realizado a identificação *in silico* das repetições em tandem mais abundantes, seguida da visualização da distribuição cromossômica por meio de FISH para entender a composição e o papel das repetições em tandem na evolução do cariótipo da espécie. A distribuição cromossômica das repetições em tandem de *S. tora*, especialmente quando comparada com sua espécie intimamente relacionada, *Senna occidentalis*, forneceu evidências citogenéticas de que rearranjos cromossômicos podem ter moldado o genoma atual de *S. tora* (WAMINAL et al., 2021).

A fração repetitiva no genoma de *S. tora* inicialmente foi prevista usando o método *de novo* e classificadas em subclasses de repetição. A identificação de repetições foi conduzida usando RepeatModeler por Kang et al. (2020), que estimou em 53,9% do genoma total a fração

repetitiva da espécie. Retrotransposons de repetição longa terminal (LTRs), principalmente LTRs do tipo *Gypsy*, foram os mais abundantes, ocupando 15,6% do genoma.

3 MATERIAL E MÉTODOS

3.1 MATERIAL VEGETAL

As sementes de *Senna occidentalis*, *Cassia fistula* e *Chamaecrista ensiformes* foram coletadas na região sul do estado de Minas Gerais e se encontram depositadas no laboratório de Citogenética da Universidade Federal de Lavras, Lavras-Minas Gerais.

Tabela 1 - Espécies da subtribo Cassiinae com seus respectivos locais de coleta, número cromossômico e tamanho do genoma.

Espécies	Número de cromossomos	Genoma 1C/Mpb	Local de coleta
<i>Cassia fistula</i>	28	700*	Lavras – MG
<i>Chamaecrista ensiformes</i>	28	850**	Serra do Cipó – MG
<i>Senna occidentalis</i>	28	700*	Lavras – MG

Fonte: OHRI (1986); BRAZ (2014).

3.2 ANÁLISE IN SILICO DE DNA REPETITIVO

3.2.1 Extração de DNA genômico e sequenciamento

As análises foram feitas no Laboratório de Citogenética Vegetal, Departamento de Biologia da Universidade Federal de Lavras. Folhas jovens de *Senna occidentalis*, *Cassia fistula* e *Chamaecrista ensiformes* foram obtidas de plantas cultivadas em casa de vegetação para a extração de DNA genômico com o Kit de Extração DNeasy® Plant Mini Kit (Qiagen® Inc., Venlo, Netherlands). A qualidade e a concentração do DNA foram avaliadas por meio de eletroforese em gel de agarose 1% e em espectrofotômetro NanoDrop™.

O sequenciamento do DNA genômico foi realizado em parceria com o Instituto René Rachou – Fiocruz Minas, localizado em Belo Horizonte – MG. O sequenciamento de baixa cobertura foi realizado na plataforma HiSeq™ 4000 da Illumina, onde, a biblioteca foi preparada com o kit Nextera™ DNA Flex Library Prep e foram gerados *reads paired-ends* de 2x300pb (Illumina™).

3.2.2 Pré-processamento de dados e análise de *clustering*

As sequências (reads) foram analisadas na plataforma RepeatExplorer (RE), onde foram realizadas as etapas de pré-processamento dos *reads* e as análises de *clustering* (NOVÁK et al., 2013).

Inicialmente, foi utilizada a ferramenta FastQC Quality Reports, para identificar eventuais problemas e filtrar sequências de baixa qualidade para etapas posteriores (BABRAHAM BIOINFORMATICS, 2016).

Após a análise de qualidade, foi utilizada a ferramenta Trim sequences com o propósito de manter apenas as regiões de alta confiança dos *reads*. Após o corte, o comprimento dos reads foram de 207 pb para a espécie *Cassia fistula*, 134 pb para *Chamaecrista ensiformes* e 205 pb para *Senna occidentalis*. Posteriormente, foi realizada a filtragem de sequências através da ferramenta Filter by quality, foram eliminados reads que não apresentassem 90% das suas bases com qualidade igual ou superior a 20 (Phred quality score).

Os *reads Forward* e *Reverse* de cada espécie foram concatenados em um só conjunto de dados através da ferramenta Interlacer Paired Reads.

Na análise comparativa, para resolver incongruências entre o comprimento das sequências foi feita a homogeneização do tamanho dos *reads*, com 134 pb para as três espécies. Para a identificação de cada espécie foi adicionado um prefixo (“CaF” “ChE” e “SeO”), indicando o genoma de origem de *Cassia fistula*, *Chamaecrista ensiformes* e *Senna occidentalis* respectivamente. Finalmente os *reads* das três espécies foram agrupados em um único arquivo. Este, por fim, foi utilizado como entrada para o processo de clusterização da análise comparativa.

3.2.3 Identificação e quantificação dos elementos repetitivos

A análise de *clustering* foi feita por meio da ferramenta mgblast tool, onde todos os reads foram comparados um contra todos e agrupados em clusters por similaridade, de acordo com definições padrão da plataforma RE (<https://repeatexplorer-elixir.cerit-sc.cz>). Para que um read seja incorporado ao um cluster, é necessário que ele possua 55% de cobertura em relação ao comprimento e 90% de identidade em relação a outro read (NOVÁK; NEUMANN; MACAS, 2020).

A caracterização e identificação foram realizadas seguindo os mesmos parâmetros descritos para as configurações em RepeatExplorer 2, isso tanto para a análise individual como

para a análise conjunta. Na análise individual os *reads* foram agrupados em clusters com base em parâmetros de similaridade, sendo representados por gráficos com características distintas. (BELYAYEV et al., 2018). A anotação automática é realizada com base no banco de dados REXdb (NEUMAN et al., 2019). Na análise comparativa foi seguido o mesmo procedimento, porém, com análise simultânea dos *reads* das três espécies mesclados em um único conjunto de dados.

A partir do agrupamento por abundância e similaridade, os elementos transponíveis presentes nas espécies foram classificados e anotados de forma automática, dentro do pipeline Repeat Explorer 2 Clustering. Para os clusters formados e não identificados na análise comparativa, foi feita a identificação manual através de comparação contra o banco de dados Giri Repbase (BAO; KOJIMA; KOHANY, 2015) e classificados conforme similaridade com as sequências disponíveis.

Dessa forma, os resultados da abundância e caracterização dos elementos transponíveis foram baseados na análise comparativa.

3.2.4 Repetições em tandem

As sequências classificadas como satélites na análise individual foram identificadas pelo algoritmo TAREAN (NOVÁK et al., 2017). A partir dos clusters caracterizados como satélites, foi realizada a análise para confirmar se as sequências repetitivas são de arranjos em tandem. Essa análise consistiu em alinhar pelo menos cinco contigs contra si mesmos através da plataforma Dotmatcher – EMBOSS e verificar o padrão do gráfico Dotplot produzido.

A relação de identidade e similaridade entre os satélites oriundos das três espécies foi analisada na plataforma Repeat Masker (SMIT et al., 2015). Já a ferramenta Map to Reference foi utilizada para mapeamento e quantificação de sequências satélite contra o genoma pré-processado de *Cassia fistula*, *Chamaecrista ensiformes* e *Senna occidentalis*, através do algoritmo MAFFT dentro do software Geneious Prime 2019 (KEARSE et al., 2012).

As matrizes de identidade foram avaliadas conforme proposto por Ruiz-Ruano et al. (2016), onde similaridade maior que 95% indica a mesma variante de satélite, similaridade entre 80% e 95% indica diferentes variantes de uma mesma família e abaixo de 80% indicam sequências distintas com proximidade ao nível de superfamília.

A nomenclatura adotada para identificação de cada satélite se deu pelo código de três letras para representar a espécie, seguidas do sufixo “Sat” e posterior número para indicar a família e/ou superfamília, por ordem decrescente de abundância.

4 RESULTADOS

O sequenciamento produziu sequências com cobertura genômica de 0,4 vezes para *Cassia fistula* e *Senna occidentalis* e de 0,5 vezes para *Chamaecrista ensiformes*. Após o pré-processamento e retirada dos reads de baixa qualidade, o número de sequências submetidas para análise individual foi de 1.229.946 para *Cassia fistula*, 1.165.326 para *Chamaecrista ensiformes* e de 939.212 para *Senna occidentalis*, o que corresponde a aproximadamente 0,1x o genoma de cada espécie.

Do total de 3.334.484 reads submetidos para análise comparativa, 2.520.089 reads foram organizados em 329.236 clusters, enquanto 814.395 reads foram classificados como sequências não repetitivas.

Na análise comparativa onde os genomas das três espécies foram analisados simultaneamente, 419 clusters foram os mais representativos, ou seja, com mais de 0,01% de abundância. Os clusters classificados como genoma extranuclear foram removidos da análise. Os resultados da análise comparativa descritos a seguir estão apresentados na Tabela 2.

Chamaecrista ensiformes possui a menor proporção de elementos repetitivos 32,86%, enquanto *Cassia fistula* e *Senna occidentalis* tem proporções maiores e similares entre si, 41,74% e 40,49%, respectivamente. A maior contribuição para o tamanho do genoma das espécies é dada pelos retrotransposons, ou elementos Classe I, os quais representam 31,67% para *Cassia fistula*, 26,42% para *Chamaecrista ensiformes* e 20,37% para *Senna occidentalis* (Figura 1).

Os retrotransposons Ty3/Gypsy, pertencentes à ordem LTR (*Long Terminal Repeat*), foram os mais abundantes nas três espécies, porém em menor proporção no genoma de *Senna occidentalis*, onde corresponde a 13,29% da fração repetitiva. No genoma de *Cassia fistula* e *Chamaecrista ensiformes*, Ty3/Gypsy apresenta proporção similar, 19,90% e 20,05% respectivamente.

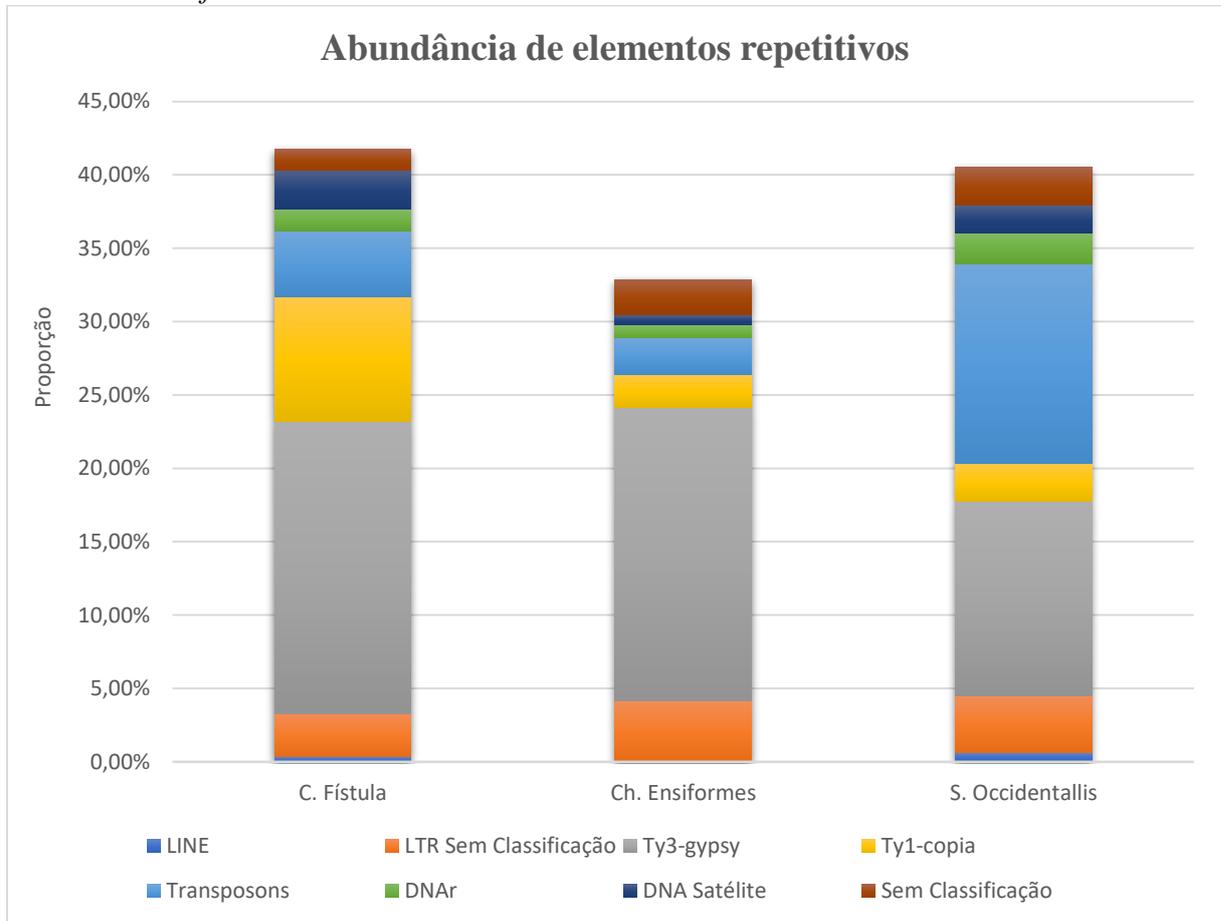
Já a abundância dos retrotransposons da superfamília Ty1/Copia, também pertencentes à ordem LTR, é bem menor quando comparada com a proporção de Ty3/Gypsy e mais divergente entre as espécies. No genoma de *Cassia*, Ty1/Copia (8,47%), é aproximadamente três vezes mais abundante que em *Chamaecrista ensiformes* (2,24%) e em *Senna occidentalis* (2,61%).

Tabela 2 - Fração repetitiva detalhada no genoma de espécies pertencentes à subtribo Cassiinae. *Cassia fistula* (Caf), *Chamaecrista ensiformes* (Che) e *Senna occidentalis* (Seo).

Ordem	Super-família	Família	Abundância (%)		
			Caf	Che	Seo
		Classe I – Retrotransposons			
		LTR Sem Classificação	2,93	4,13	3,82
		<i>Tat/Retand</i>	4,24	0,34	5,26
		Chromovirus (<i>Tekay</i> + CRM)	5,43	13,23	4,06
	<i>Ty3-Gypsy</i>	<i>Athila</i>	9,45	2,45	2,79
		<i>Ogre</i>	0,55	0,74	0,83
		<i>Gypsy</i> Sem Classificação	0,23	3,29	0,35
		Total <i>Gypsy</i>	19,90	20,05	13,29
LTR		<i>Ale</i>	0,04	0,37	0,33
		<i>Alesia</i>	0,06	0,00	0,00
		<i>TAR</i>	0,12	0,01	0,46
		<i>SIRE</i>	6,01	0,58	0,28
	<i>Ty1-Copia</i>	<i>Ivana</i>	0,14	0,05	0,00
		<i>Bianca</i>	0,50	0,13	0,59
		<i>Ikeros</i>	0,75	0,69	0,70
		<i>Tork</i>	0,15	0,31	0,17
		<i>Angela</i>	0,60	0,09	0,00
		<i>Copia</i> Sem Classificação	0,10	0,00	0,07
		Total <i>Copia</i>	8,47	2,24	2,61
	LINE		0,37	0,00	0,65
		Total Retrotransposons	31,67	26,42	20,37
		Classe II – Transposons			
		<i>PIF_Harbinger</i>	0,00	0,55	2,42
Subclasse I – TIR		<i>Tc1-Mariner</i>	0,50	0,00	0,00
		<i>EnSpm – CACTA</i>	1,03	0,13	1,45
		<i>MuDR – Mutator</i>	1,93	0,47	1,34
		<i>hAT</i>	0,84	0,55	0,53
Subclasse II		<i>Helitron</i>	0,24	0,77	7,82
		Total Transposons	4,54	2,47	13,56
		Total Elementos Transponíveis	36,21	28,89	33,93
		DNA Satélite	2,66	0,70	1,86
		rDNA 5S	0,21	0,16	0,01
		rDNA 45S	1,21	0,71	2,12
		rDNA Total	1,42	0,87	2,13
Sem Classificação			1,45	2,41	2,58
		Total DNA Repetitivo	41,74	32,86	40,49

Fonte: Do autor (2021).

Figura 1 - Composição da fração repetitiva no genoma de *Cassia fistula*, *Chamaecrista ensiformes* e *Senna occidentalis*.



Fonte: Do autor (2021).

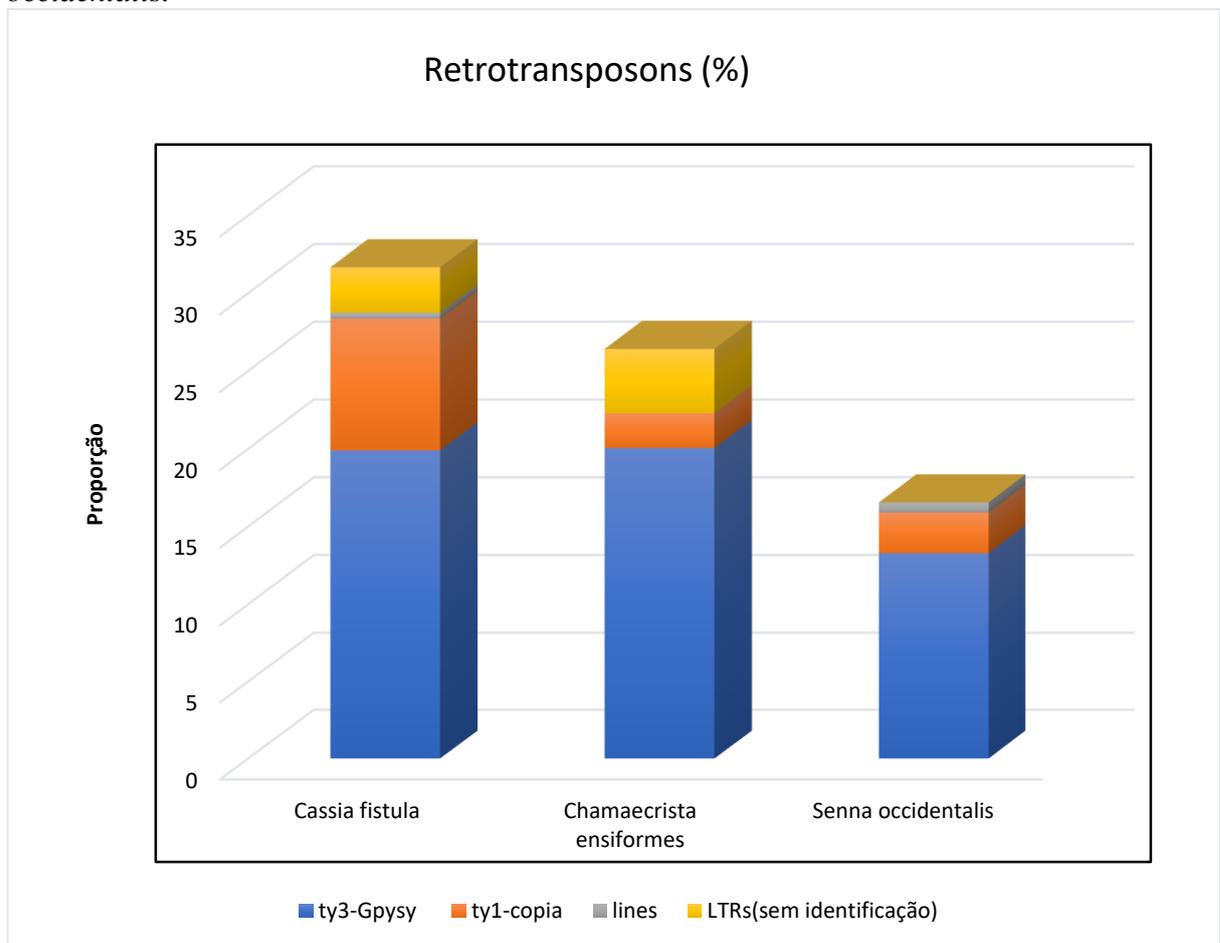
Os retrotransposons ou elementos de classe I possuem fração divergente entre as três espécies, embora a super-família de Ty3/Gypsy seja similar entre *C. fistula* e *Ch. ensiformes*. A proporção da super-família *Ty1/copia* é mais similar entre *Ch. ensiformes* e *S. occidentalis* (Figura 2).

Em relação às famílias de elementos LTR Ty3/Gypsy, a família *Chromovirus* foi mais abundante em *Ch. ensiformes* com 13,23%. Para *C. fistula* a abundância foi de 5,43% e para *S. occidentalis* foi de 4,06%. As famílias *Athila* e *Tat/Retand* também foram bem representadas. A proporção de *Athila* foi mais abundante no genoma de *Cassia* (9,45%) e semelhante entre as outras duas espécies (2,45% e 2,79%). Já *Tat/Retand* teve uma variação de 1% entre *C. fistula* (4,24%) e *S. occidentalis* (5,26%), enquanto em *Ch. ensiformes* apresentou apenas 0,34% do genoma constituído por essa família.

Os elementos sem classificação foram mais presentes em *Ch. ensiformes* (2,41%) e em *S. occidentalis* (2,58%), enquanto que em *C. fistula* foi de 1,45%. Os LTRs sem classificação

foram mais abundantes em *Ch. ensiformes* (4,13%), *S. occindetalis* (3,82%) e em *C. fistula* a abundância desses elementos foi de 2,93%. As superfamílias *Ty3/Gypsy* e *T71/Copia* também tiveram reads não identificados. Em *Ch. ensiformes* elementos da super-família *Ty3/Gypsy* não identificados correspondem a 3,29%, enquanto que em *C. fistula* e *S. occindentalis* foi bem menor, 0,23 e 0,35%, respectivamente. Já os elementos não classificados de *Ty1/Copia* tiveram uma proporção bem inferior, variando de 0,00 a 0,10% entre as espécies.

Figura 2 - Proporção de retrotransposons em *Cassia fistula*, *Chamaecrista ensiformes* e *Senna occidentalis*.



Fonte: Do autor (2021).

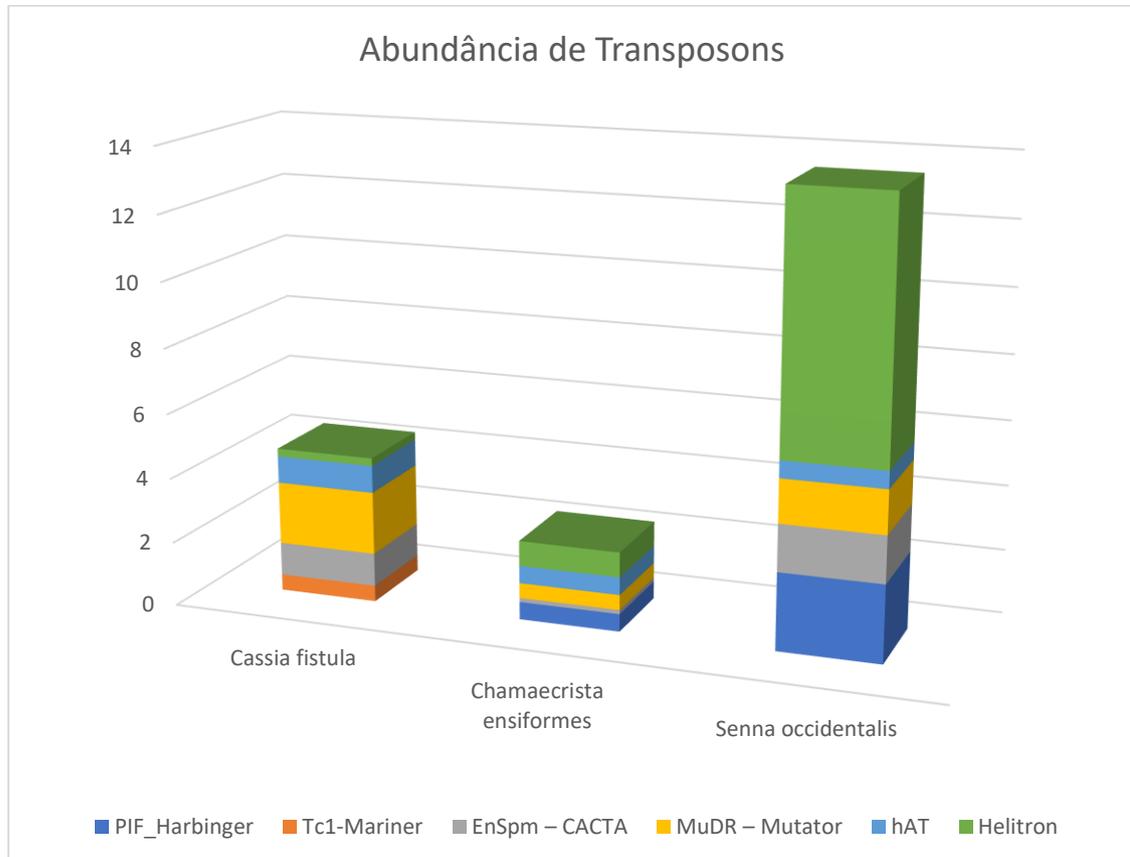
De forma geral, as famílias de *Ty1/Copia* no genoma das três espécies estão presentes em uma proporção relativamente baixa. A família mais abundante é a SIRE, com 6,01% do genoma de *C. fistula*, 0,58% de *Ch. ensiformes* e 0,28% de *S. occidentalis*. A família *Alesia* está presente apenas no genoma de *C. fistula*, na proporção de 0,06%. As famílias *Ivana* e *Angela* não estão presentes no genoma de *S. occidentalis*, enquanto que em *Ch. Ensiformes* representam 0,14% e 0,05% e em *C. fistula* a abundância é superior, com 0,60% e 0,14%

respectivamente. A família *Ikeros* possui proporção similar entre as três espécies, variando de 0,69% a 0,75%.

Os retrotransposons Non-LTRs não foram amostrados no genoma de *Ch. ensiformes*, enquanto que a família de elementos intercalados longos (LINEs) correspondem a uma proporção de 0,37% da fração repetitiva de *C. fistula* e de 0,65% em *S. occidentalis*.

Os elementos classe II ou transposons foram mais abundantes na espécie *S. occidentalis*, onde correspondem a 13,56% da fração repetitiva. Em *C. fistula* e *Ch. ensiformes* essa classe é bem menos abundante, com 4,54% e 2,47%, respectivamente (Figura 3). Em relação à subclasse I, foram identificados apenas elementos da Ordem TIR (*Terminal Inverted Repeat*), em baixas proporções (0,5 a 2,42%). Os elementos identificados pertencem às famílias CACTA, hAT e *Mutator* no genoma das três espécies, à família *Harbinger* no genoma de *Ch. ensiformes* e de *S. occidentalis* e à família *Mariner*, apenas em *C. fistula*. A subclasse II foi representada apenas pela família *Helitron*, nas três espécies, com uma abundância superior em *S. occidentalis* (7,82%) e numa proporção inferior a 1% nas outras duas espécies.

Figura 3 - Famílias de transposons de DNA presentes no genoma de *C. fistula*, *Ch. ensiformes* e *S. occidentalis*.



Fonte: Do autor (2021).

O DNA ribossomal apresentou menor abundância no genoma de *Ch. Ensiformes* (0,87%), enquanto que *S. occidentalis* foi a espécie com maior proporção (2,58%) e *C. fistula* apresentou abundância intermediária (1,42%). A proporção de DNA satélites em *Chamaecrista* também foi menor (0,70%) quando comparada as outras duas espécies, que apresentaram 1,86% (*S. occidentalis*) e 2,66% (*C. fistula*).

A análise fornecida pela ferramenta TAREAN identificou, inicialmente, 25 possíveis satélites, sendo 15 pertencentes a *Cassia fistula*, quatro a *Chamaecrista ensiformes* e seis a *Senna occidentalis*. Destes, seis satélites foram confirmados através da plataforma Dotmatcher – EMBOSS, sendo três para *C. fistula*, um para *Ch. ensiformes*, dois para *S. occidentalis*. Os outros 19 satélites identificados previamente pelo TAREAN não foram confirmados como sequências repetitivas em tandem pelo padrão do gráfico gerado pela plataforma Dotmatcher – EMBOSS.

O alinhamento via algoritmo MAFFT revelou uma identidade inferior a 80% para quatro satélites (ChESat1W, SeOSat2W, CaFSat3W, SeOSat4W), os quais foram organizados ao nível de superfamília. CaFSat1A e CaFSat2A tiveram uma similaridade de 82% e foram agrupados ao nível de família. Na família 1 (F1; Sufixo A) foram agrupados dois DNA satélites (CaFSat1A e CaFSat2A) pertencentes ao genoma de *Cassia fistula*. Os dois satélites da família F1 possuem comprimento de 111 pb. Na superfamília 1 (SF1; Sufixo W) foram agrupados quatro DNA satélites com tamanhos que variam de 154 a 181 pb.

Em *S. occidentalis* os dois satélites encontrados SeoSat2W e SeoSat4W apresentaram similaridade de 70% e comprimento bem similar, 180 pb e 181bp. De acordo com o sistema de classificação de Ruano e colaboradores (2016) os dois satélites dessa espécie pertencem a uma mesma super-família. Em *Ch. ensiformes* foi encontrado apenas um satélite com 154 pb.

A partir do Mapeamento por Referência, foi possível quantificar a abundância dos reads e identificar a presença ou ausência nas espécies. De forma geral, todos os satélites encontrados foram praticamente específicos de cada espécie, ocorrendo a presença de alguns com abundância bem pequena em outras espécies (Tabela 3, figura 4).

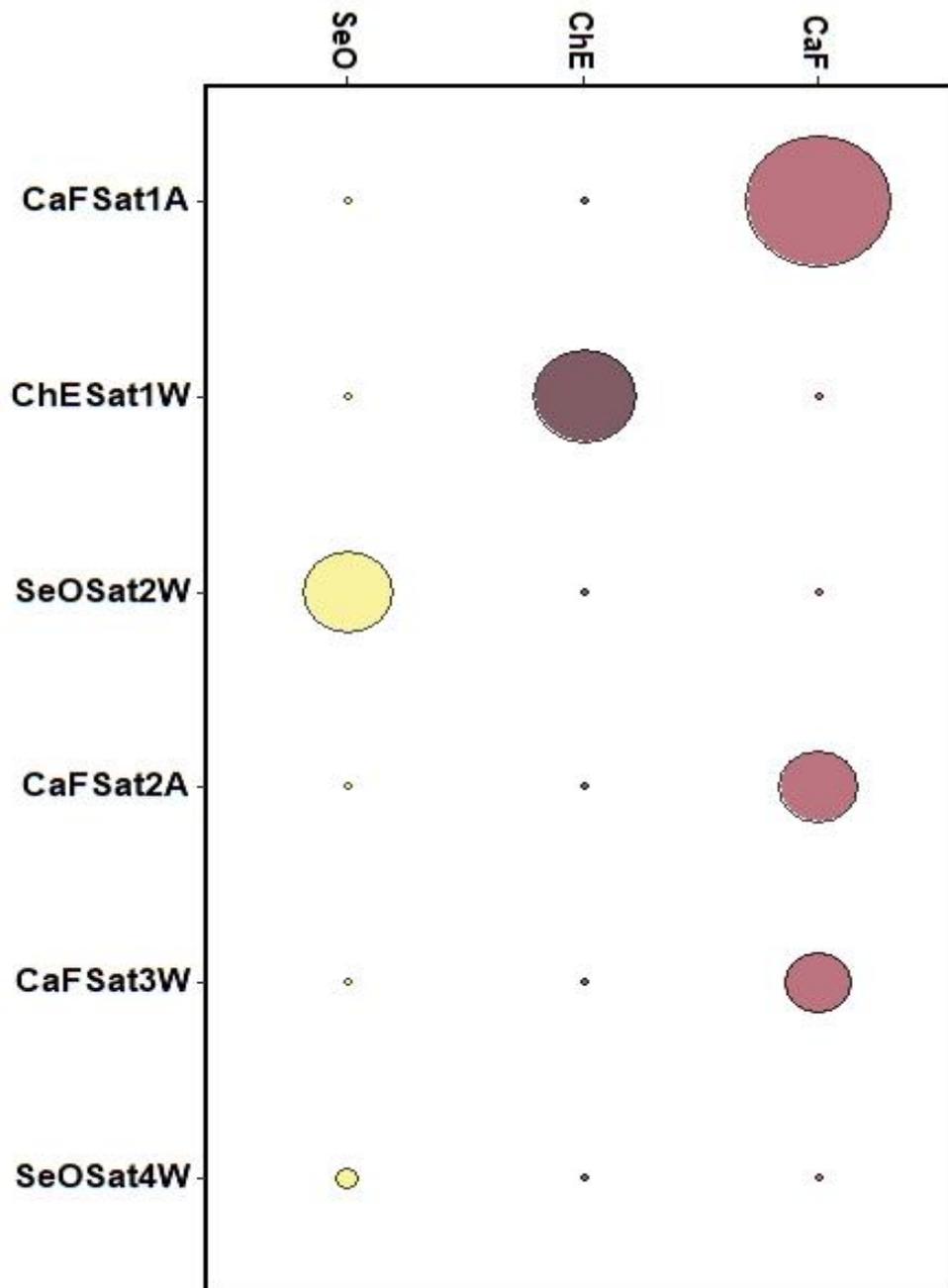
Os satélites (CaFSat1A, CaFSat2A) identificados no genoma de *C. fistula* foram também mapeados no genoma de *Ch. ensiformes* numa abundância pequena, 4 e 2 reads respectivamente. Já o satélite (ChESat1W) identificado no genoma de *Ch. ensiformes* foi mapeado no genoma de *C. fistula* numa abundância de oito reads. Os satélites identificados em *S. occidentalis* (SeOSat2W e SeOSat4W) são específicos para a espécie.

Tabela 3 – Abundância e comprimento das sequências satélite encontradas em *Cassia fistula* (caf), *Chamaecrista ensiformes* (Che) e *Senna occidentalis*, organizados em família (F), super-família (SF).

F/SF	DNASat	Comprimento (pb)	Abundância		
			Caf	Che	Seo
F1	CaFSat1A	111	64.435	4	0
SF1	ChESat1W	154	8	31.425	0
SF1	SeOSat2W	180	0	0	23.323
F1	CaFSat2A	111	17.884	2	0
SF1	CaFSat3W	169	12.135	0	0
SF1	SeOSat4W	181	0	0	842
Total de reads de cada espécie			1.229.946	1.165.326	939.212

Fonte: Do autor (2021).

Figura 4 - Abundância de seqüências satélite no genoma de *Cassia fistula* (1.229.946 reads), *Chamaecrista ensiformes* (1.165.326 reads) e *Senna occidentalis* (939.212 reads).



Fonte: Do autor (2021).

5 DISCUSSÃO

Sequências de DNA repetitivo estão presentes em todas as plantas superiores e podem representar até 90% do tamanho do genoma em algumas espécies. Essas sequências são consideradas um dos principais fatores responsáveis por gerar diferenças no tamanho do genoma e, assim, refletir distâncias evolutivas entre as espécies (MEHROTRA; GOYAL, 2014). Essa fração repetitiva, onipresente e dinâmica nos genomas, ocupa no genoma total as proporções de 41,74% em *Cassia fistula* (0,70 pg/1C), 32,86% em *Chamaecrista ensiformes* (0,85 pg/1C) e 40,49% para *Senna occidentalis* (0,70 pg/1C).

Embora a poliploidização resulte na multiplicação instantânea de todo o genoma nuclear, a evolução da composição repetitiva do genoma explica a maior parte da variação observada no tamanho do genoma em angiospermas (BENNETZEN, 2005). Das espécies analisadas neste estudo, *Ch. ensiformes* é a que possui maior conteúdo genômico (0,80 Gb) e a menor quantidade de DNA repetitivo. Embora haja, geralmente, correlação entre tamanho e quantidade de DNA repetitivo, essas repetições também podem ser removidas do genoma pela ação de vários mecanismos baseados em recombinação cromossômica (DEVOS; BROWN; BENNETZEN, 2002).

Os retrotransposons de repetição terminal longa (LTR) são os elementos que ocorrem com maior frequência na maioria dos genomas de plantas superiores e são normalmente distribuídos por todos os cromossomos. Essas repetições abrangem várias superfamílias, das quais *Ty3-Gypsy* e *Ty1-copia* são as mais comuns em plantas (WICKER et al., 2007). Nossos resultados corroboram essa tendência, onde os retrotransposons foram os elementos de maior proporção na fração repetitiva das três espécies em estudo, sendo a super-família *Ty3-Gypsy* mais representativa do que *Ty1-Copia*. No genoma de *Ch. ensiformes* por exemplo, a proporção de *Ty3-Gypsy* chega a ser quase 10 vezes maior que *Ty1-Copia*. Em alguns estudos já foi demonstrado que os retrotransposons LTR representam a principal classe de DNA repetitivo na maioria dos genomas de plantas e seu acúmulo é, junto com várias rodadas de poliploidização, uma força chave que governa a expansão do tamanho do genoma (NEUMANN et al., 2006).

Em relação às famílias de *Ty1-Copia*, apenas a família SIRE foi identificada em *C. fistula*, correspondendo a 6,01%, enquanto nas outras duas espécies não alcança 1% das repetições.

Esses resultados corroboram os estudos feitos até o momento em espécies de leguminosas, onde parece não haver um padrão simples na evolução dos retrotransposons LTR que explique a diversidade de tamanho observada no genoma de plantas. Em parte, isso pode

ser devido ao conhecimento limitado de dados de poucas espécies ou pequenos grupos de táxons relacionados, sendo necessário uma amostragem mais ampla para preencher essas lacunas (AGREN et al., 2015). Nossos resultados indicam que os retrotransposons têm papel relevante na evolução do genoma de espécies de Cassiinae, uma vez que uma família particular de retrotransposon pode ser específica da espécie ou estar presentes em espécies relacionadas em proporções diferentes, indicando que algumas sequências repetitivas evoluem rapidamente, enquanto outras podem ser conservadas. Embora mais estudos, com um maior número espécies seja necessário para se constatar o papel dos elementos repetitivos na evolução do genoma de espécies de Cassiinae, n

Em *Senna tora* a fração repetitiva representa cerca de 53,9% de todo o genoma, onde os retrotransposons de repetição terminal longa (LTRs), principalmente LTRs do tipo *Gypsy*, são os mais abundantes, ocupando 15,6% do genoma (KANG, 2020). Em nosso estudo, *S. occidentalis*, espécie mais proximamente relacionada a *S. tora*, possui fração repetitiva de 40,49%, com LTRs do tipo *Ty3-Gypsy* em maior abundância.

Em espécies do gênero *Caesalpinia*, grupo irmão de Cassiinae (Caesalpinoideae), a fração repetitiva dos genomas de *Cenostigma microphyllum*, *Libidibia ferrea* e *Pau-brasilia echinata* foram estimados em 41,70%, 38,44% e 72,51%, respectivamente. Os retransposons da família *Ty3/Gypsy*, especificamente a linhagem *Tekay*, foram as repetições mais abundantes em cada um dos três genomas (VAN-LUME et al., 2020). Nas espécies *C. fistula*, *Ch. ensiformes* e *S. occidentalis* a linhagem *Tekay* corresponde a 5,35%, 4,13% e 3,96% respectivamente.

No genoma de *Erythrostemon hughesii* ($2n = 24$), espécie de Caesalpineae, a proporção de DNA repetitivo foi de 28,73%, onde os retrotransposons-LTRs e satDNAs foram as repetições mais abundantes, representando 9,76% e 7,83%. (SUCRE et al., 2020). A proporção de DNA repetitivo de *E. hughesii* é similar à de *Ch. ensiformes*.

Na nossa análise, os elementos repetitivos de classe II, ou transposons, possuem uma porção inferior quando comparada com a quantidade de retrotransposons. Entre as três espécies em nosso estudo, *S. occidentalis* é a que possui maior quantidade dessa classe de elementos, correspondendo a 13,56% da fração repetitiva. *C. fistula* e *Ch. ensiformes* possuem 4,54% e 2,47% respectivamente. Não houve uma relação entre as espécies quanto à presença das famílias nem à proporção das mesmas, ou seja, cada família contribuiu com uma quantidade diferente, sendo que algumas estão presentes enquanto outras ausentes entre as três espécies de Cassiinae. As diferenças mais marcantes foram observadas na família *Helitron*, que contribuiu com mais de 7% da fração repetitiva nessa espécie, não chegando a 1% nas outras duas espécies.

Evidências sugerem que a reestruturação cromossômica está frequentemente associada a regiões heterocromáticas compostas de sequências de DNA repetitivo. Sequências repetitivas, especialmente ETs, estão envolvidas em vários rearranjos cromossômicos (YAKOVLEV et al., 2017). Evidências citogenéticas mostraram que os transposons *EnSpm* estavam envolvidos no rearranjo cromossômico em andamento, levando ao aumento de uma nova população de plantas férteis de *Aegilops speltoides* (RASKINA et al., 2004). Em *C. fistula* a família de elementos *EnSpm* é representada por 1,03% enquanto que em *S. occidentalis* e em *Ch. ensiformes* essa família contribui com 1,45% e 0,13%, respectivamente.

Com relação ao DNA ribossomal em *S. tora* a proporção de rDNA 5S não foi detectada pelo TAREAN, o que implica numa fração consideravelmente pequena no genoma desta espécie. Posteriormente a fração de 5S foi identificada com a ferramenta dnaLCW, correspondendo a 0,01% do genoma, enquanto a porção 45 S correspondeu a 1,04%. Em *S. occidentalis*, a fração 5S também foi de 0,01%, mas a sequência 45S foi praticamente o dobro (2,12%).

Distribuição diferenciada de sequências repetitivas nos cromossomos fornece dados comparativos úteis para estudar a dinâmica do genoma e história evolutiva de espécies relacionadas (AMMIRAJU et al. 2008). Dois tipos de elementos repetitivos frequentemente usados nestes estudos são as sequências 45S e 5S dos genes ribossomais, esses dois genes possuem regiões altamente repetitivas e conservadas, tornando-os úteis para pesquisas citogenéticas (WAMINAL; KIM, 2015).

Em *S. tora*, sequências 45S e 5S do rDNA foram localizadas em apenas um loco cada, nos cromossomos 2S e 12S, respectivamente. A localização dessas regiões nos cromossomos de *Senna* possui padrões distintos de distribuição. Em um estudo com 10 espécies do gênero apenas um par de rDNA 5S foi detectado em todas as espécies, a maioria frequentemente no penúltimo cromossomo. Os sinais foram geralmente localizados em regiões pericentroméricas, exceto em *Senna viarum*, onde estavam presentes nas regiões intersticiais dos braços longos. *S. occidentalis* possui $2n = 28$ cromossomos, com um loco 5S localizado no cromossomo 13 e um loco de rDNA 45S localizado no cromossomo 2 (PELLERIN; WAMINAL; KIM, 2019).

Ch. ensiformes possui $2n = 28$ cromossomos e três locos de rDNA 45S e 5S, localizados de forma contígua nas regiões terminais do braço curto do cromossomo 2 e do braço longo dos cromossomos 1 e 7. (BRAZ, 2014). Apesar do maior número de locos, em nosso estudo, *Ch. ensiformes* possui a menor proporção de rDNA entre as três espécies, com 0,87% do total da fração repetitiva, o que pode ser reflexo da eliminação ou menor amplificação destas sequências nos diferentes locos.

Na nossa análise as três espécies apresentaram uma proporção de DNAs satélite relativamente baixa, sendo que *Ch. ensiformes* possui menos da metade dessa fração repetitiva quando comparada às outras duas espécies. As repetições organizadas em tandem foram os elementos mais divergentes entre as espécies de Cassiinae estudadas, onde os satélites encontrados são quase que exclusivos de cada espécie. Foi verificado, pelo mapeamento, que em *C. fistula* e *Ch. ensiformes* há um compartilhamento de DNA satélite entre as espécies, porém numa quantidade extremamente baixa. Esse resultado corrobora a hipótese de que o conjunto de satélites de uma espécie pode ou não ser parcial ou totalmente compartilhado com outras espécies e populações proximamente relacionadas (RUIZ-RUANO et al., 2016). O fato de os satélites serem quase que específicos de cada espécie pode ser explicado pelo fato dessas sequências evoluírem rapidamente. Este fenômeno pode ser explicado pela hipótese da biblioteca, onde as espécies de um grupo relacionado contêm diferentes quantidades de um conjunto comum de repetições de satélite que são amplificados diferencialmente e podem evoluir independentemente após a especiação (UGARKOVIĆ, 2008).

Há divergências em relação à abundância de repetições em tandem entre *C. fistula*, *Ch. ensiformes* e *S. occidentalis*, o que corrobora hipóteses de que espécies de Cassiinae passaram por sucessivos processos de rearranjos cromossômicos durante a sua evolução (BIONDO et al., 2005), o que pode ter contribuído com o surgimento e eliminação de famílias de DNA satélites. É importante salientar que essas sequências repetitivas mudam rapidamente ao longo da evolução dos organismos, onde se tornam excelentes marcadores de segmentos que permitem rastrear os rearranjos cromossômicos (MOSCONI et al., 2007).

Todos os seis satélites encontrados em nossa análise possuem um comprimento similar, variando de 111 a 181 pb. O comprimento de monômeros de ocorrência mais frequentes encontrado entre os DNAs de satélite está em torno de 180 bases. Isso se encaixa em um padrão comum encontrado em DNAs satélites em que monômeros em torno de 165 nucleotídeos parecem ser favorecidos, o que pode ser consequência de restrições estruturais impostas pelo nucleossomo (GARRIDO-RAMOS, 2015).

Wanimal et al (2021), realizaram uma análise comparativa da localização de sequências satélite por meio de hibridização *in situ* fluorescente (FISH) entre nove espécies de *Senna* intimamente relacionadas dentro de um contexto citotaxonômico. Das nove repetições em tandem encontradas em *S. tora*, duas não mostraram nenhum sinal de FISH, enquanto sete mostraram tanto padrões semelhantes quanto contrastantes com outras espécies de *Senna*. A sequência StoTR01_86, que estava localizada nas regiões pericentroméricas em todos os cromossomos de *S. tora*, mas não no sítio da região organizadora de nucléolo, foi

co-localizado no sítio NOR em todas as demais espécies, exceto em *S. siamea*. Já StoTR02_7_tel foi localizado principalmente em regiões terminais, algumas espécies tiveram uma repetição telomérica intersticial em alguns cromossomos. StoTR05_180 foi distribuído na região subtelomérica na maioria das espécies e foi altamente amplificado na região pericentromérica em algumas espécies. StoTR06_159 estava ausente ou co-localizado no sítio NOR em algumas espécies, e StoIGS_463, que estava localizado no sítio NOR em *S. tora*, estava ausente ou localizado nas regiões subteloméricas ou pericentroméricas em outras espécies. A partir deste estudo foi possível compreender a dinâmica dessas repetições em tandem no gênero *Senna*, que apresentou evolução compartilhada e independente entre as espécies. O estudo também sugere que sequências de rDNA 45S possam desempenhar um papel transportador durante os rearranjos cromossômicos. Esses dados também possibilitam a visualização da expansão, contração e reorganização de famílias de sequências repetidas no gênero *Senna*.

O DNA repetitivo pode ser uma ferramenta importante para o melhor entendimento da evolução genômica de Cassiinae, sendo necessário mais estudos com um maior número de taxons dos três gêneros que constituem a subtribo. Os resultados evidenciaram a grande diversidade e abundância dos elementos repetitivos em *Cassia fistula*, *Chamaecrista ensiformes* e *Senna occidentalis*, o que pode vir a contribuir para a construção de um panorama da fração repetitiva em espécies da subtribo.

6 CONCLUSÃO

A fração repetitiva do genoma de *Cassia fistula*, *Chamaecrista ensiformes* e *Senna occidentalis* corresponde, respectivamente, 41,74%, 32,86% e 40,49%.

Os genomas das três espécies são compostos por diferentes super-famílias de elementos transponíveis. Os elementos repetitivos mais abundantes foram classificados como LTRs, onde a super -família Ty3/Gypsy foi a mais abundante, correspondendo a 19,90% em *C. fistula*, 20,05% em *Ch. ensiformes* e 13,29% em *S. occidentalis*.

Foram identificadas 6 DNA satélite, 3 para *C. fistula*, 2 para *Ch. ensiformes* e 1 para *S. occidentalis* sendo todas praticamente exclusivas de cada espécie.

Os diferentes tipos de elementos repetitivos estiveram ou ainda estão sujeitos a diferentes tipos de mecanismos evolutivos, que modelam o genoma das três espécies de forma independente entre elas.

REFERÊNCIAS

- ACHARYA, L.; MUKHERJEE, A. K.; PANDA, P. C. Separation of the genera in the subtribe Cassiinae (Leguminosae: Caesalpinioideae) using molecular markers. **Acta Botanica Brasilica**, Feira de Santana, v. 25, n. 1, p. 223-233, 2011.
- ÅGREN, J. A.; GREINER, S.; JOHNSON, M. T. J.; WRIGHT, S. I. No evidence that sex and transposable elements drive genome size variation in evening primroses. **Evolution**, v. 69, p. 1053-1062, 2015.
- ARTONI, R. F. *et al.* Inferring Diversity and Evolution in Fish by Means of Integrative Molecular Cytogenetics. **The Scientific World Journal**, v. 2015, 2015.
- AZANI, N. *et al.* A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. **Taxon**, v. 66, n. 1, p. 44-77, 2017.
- BABRAHAM BIOINFORMATICS. **A quality control tool for high throughput sequence data**. [S. l.], 2016. Disponível em: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> Acesso em: 5 out. 2021.
- BALTISBERGER, M.; HORANDL, E. Karyotype evolution supports the molecular phylogeny in the genus *Ranunculus* (Ranunculaceae). **Perspectives in Plant Ecology, Evolution and Systematics**, v. 18, p. 1-14, 2016.
- BARROSO, G. M.; MORIM, M. P.; PEIXOTO, A. L.; ICHASO, C. L. F. **Frutos e sementes: morfologia aplicada a sistemática de dicotiledôneas**. Viçosa, MG: Editora UFV, 1999.
- BELYAYEV, A. Bursts of transposable elements as an evolutionary driving force. **Journal of Evolutionary Biology**, v. 27, n. 12, p. 2573-2584, 2014.
- BFG (The Brazil Flora Group). Brazilian Flora 2020: Innovation and collaboration to meet Target 1 of the Global Strategy for Plant Conservation (GSPC). **Rodriguésia**, v. 69, n. 4, p. 1513-1527, 2018.
- BIONDO, E.; MIOTTO, S. T. S.; SCHIFINO-WITTMANN, M. T. Números cromossômicos e implicações sistemáticas em espécies da subfamília Caesalpinioideae (Leguminosae) ocorrentes na região sul do Brasil. **Revista Brasileira de Botânica**, São Paulo, v. 28, n. 4, p. 797-808, 2005.
- BISCOTTI, M. A.; OLMO, E.; HESLOP-HARRISON, P. Repetitive DNA in eukaryotic genomes. **Chromosome Research**, v. 23, n. 3, p. 415-420, 2015.
- BRUNEAU, A. *et al.* Phylogenetic patterns and diversification in the caesalpinoid legumes. **Botany**, v. 86, n. 7, p. 697-718, 2008.
- BRUNEAU, A. *et al.* Legume phylogeny and classification in the 21st century: Progress, prospects and lessons for other species-rich clades. **Taxon**, v. 62, n. 2, p. 217-248, 2013.

- CAI, Z. *et al.* Differential genome evolution and speciation of *Coix lacryma-jobi* L. and *Coix aquatica* Roxb. hybrid guangxi revealed by repetitive sequence analysis and fine karyotyping. **BMC Genomics**, v. 15, n. 1025, p. 1-16, 2014.
- CARDOSO, D. *et al.* A molecular phylogeny of the vataireoid legumes underscores floral evolvability that is general to many early-branching papilionoid lineages. **American Journal of Botany**, v. 100, n. 2, p. 403-421, 2013.
- CASACUBERTA, E.; GONZÁLEZ, J. The impact of transposable elements in environmental adaptation. **Molecular Ecology**, v. 22, n. 6, p. 1503-1517, 2013.
- CHARLESWORTH, B.; SNIEGOWSKI, P.; STEPHAN, W. The evolutionary dynamics of repetitive DNA in eukaryotes. **Nature**, n. 371, p. 215-220, 1994.
- CONCEIÇÃO, A. S. **Filogenia do gênero Chamaecrista (Leguminosae-Caesalpinioideae) e Taxonomia do grupo Baseofhyllum**. 2006. Tese (Doutorado em Botânica) - Universidade Estadual de Feira de Santana, Feira de Santana, 2006.
- CONCEIÇÃO, A. S. *et al.* Phylogeny of Chamaecrista (leguminosae - Caesalpinioideae) based on nuclear and chloroplast DNA regions. **Taxon**, v. 58, n. 4, p. 1168-1180, 2009.
- CORDEIRO, J. M. P.; FELIX, L. P. Intra and interspecific karyotypic variations of the genus *Senna* Mill. (Fabaceae, Caesalpinioideae). **Acta Botanica Brasilica**, Belo Horizonte, v. 32, n. 1, p. 128-134, 2018.
- CRAIG, N. L. *et al.* **Mobile DNA II**. Washington, DC: American Society for Microbiology Press, 2002.
- CUADRADO, A.; SCHWARZACHER, T. The chromosomal organization of simple sequence repeats in wheat and rye genomes. **Chromosoma**, v. 107, p. 587-594, 1998.
- CUADRADO, A. *et al.* Physical organization of simple sequence repeats (SSRs) in Triticeae: structural, functional and evolutionary implications. **Cytogenet Genome**, v. 120, p. 210-219, 2008.
- DENG, H. *et al.* Molecular cytogenetic analysis of genome-specific repetitive elements in *Citrus clementina* Hort. Ex Tan. and its taxonomic implications. **BMC Plant Biology**, v. 19, 2019.
- DEVOS, K. M.; BROWN, J. K.; BENNETZEN, J. L. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. **Genome Research**, v. 12, n. 7, p. 1075-1079, 2002.
- DODSWORTH, S. *et al.* Genomic repeat abundances contain phylogenetic signal. **Systematic biology**, v. 64, p. 112-126, 2015.
- EVGEN'EV, M. B. *et al.* Penelope, a new family of transposable elements and its possible role in hybrid dysgenesis in *Drosophila virilis*. **Proceedings of the National Academy of Sciences**, v. 94, n. 1, p. 196-201, 1997.

FLORA DO BRASIL 2020. Jardim Botânico do Rio de Janeiro. [S. l.], 2020. Disponível em: <http://floradobrasil.jbrj.gov.br/> Acesso em: 13 set. 2021.

GARRIDO-RAMOS, M. A. Satellite DNA in plants: more than just rubbish. **Cytogenet and Genome Research**, v. 146, n. 2, p. 153-170, 2015.

GLADYSHEV, E. A.; ARKHIPOVA, I. R. Telomere-associated endonuclease-deficient *Penelope*-like retroelements in diversos eukaryotes. **Proceedings of the National Academy of Sciences**, v. 104, n. 22, p. 9352-9357, 2007.

GOLDBLATT, P. Cytology and the phylogeny of Leguminosae. *In.*: POLHILL, R. M.; RAVEN, P. H. (eds.). **Advances in Legume Systematic**. Kew: Royal Botanic Gardens, 1981. p. 427-463.

GOLDBLATT, P.; JOHNSON, D. E. **Index to plant chromosome numbers 1992-1993**. Saint Louis: Missouri Botanical Garden, 1996.

GOLDBLATT, P.; JOHNSON, D. E. **Index to plant chromosome numbers 1996-1997**. Saint Louis: Missouri Botanical Garden, 2000.

GUERRA, M. Cytotaxonomy: the end of childhood. **Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology**, v. 146, n. 3, p. 703-710, 2012.

HERENDEEN, P. S. **Phylogenetic relationships in caesalpinoid legumes: a preliminary analysis based on morphological and molecular data**. Kew: Royal Botanic Gardens, 2003. p. 37-62.

HUANG, J. *et al.* Comparative transcriptomics uncovers alternative splicing changes and signatures of selection from maize improvement. **BMC Genomics**, v. 16, n. 363, 2015.

IRWIN, H. S.; BARNEBY, R. C. **The American Cassiinae, a synoptical revision of Leguminosae tribe Cassiae subtribe Cassiinae in the New World**. Bronx: New York Botanical Garden, 1982.

IRWIN, H. S.; BARNEBY, R. C. Tribe Cassieae Bronn. *In.*: POLHILL, R. M.; RAVEN, P. H. (eds.). **Recent Advances in Legume Systematics**. London: Royal Botanic Gardens, 1981. p. 97-106.

IRWIN, H. S.; TURNER, B. L. Chromosome relationships and taxonomic considerations in the genus *Cassia*. **American Journal of Botany**, v. 47, n. 4, p. 309-318, 1960.

JURKA, J. *et al.* Repetitive Sequences in Complex Genomes: Structure and Evolution. **Annual review of genomics and human genetics**, v. 8, p. 241-259, 2007.

KAJITA, T. *et al.* rbcL and legume phylogeny, with particular reference to Phaseoleae, Millettieae, and Allies. **Systematic Botany**, Kent, v. 26, n. 3, p. 515-536, 2001.

KAPITONOV, V. V.; JURKA, J. Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. **Proceedings of the National Academy of Sciences**, v. 100, n. 11, p. 6569-6574, 2003.

- KAZAZIAN, H. H. Mobile DNA transposition in somatic cells. **BMC Biology**, v. 9, n. 62, 2011.
- KIT, S. Equilibrium sedimentation in density gradients of DNA preparations from animal tissues. **Journal of molecular biology**, v. 3, n. 6, p. 711-716, 1961.
- KOJIMA, K. K. A new class of SINEs with snRNA gene-derived heads. **Genome biology and Evolution**, v. 7, n. 6, p. 1702-1712, 2015.
- KOO, D. H. *et al.* Molecular cytogenetic mapping of satellite DNA sequences in *Aegilops geniculata* and wheat. **Cytogenet and Genome Research**, v. 148, p. 314-321, 2016.
- LEVINSON, G.; GUTMAN, G. A. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. **Molecular Biology and Evolution**, v. 2, p. 203-221, 1987.
- LEWIS, G. P. Tribe Cassieae. *In.*: LEWIS, G. P. *et al.* (eds.). **Legumes of the World**. Kew: Royal Botanic Gardens, 2005. p. 111-126.
- LEWIS, G. P. *et al.* **Legumes of the world**. Kew: Royal Botanic Gardens, 2005.
- LIANG, G.; CHEN, H. Scaling chromosomes for an evolutionary karyotype: a chromosomal tradeoff between size and number across woody species. **PLoS One**, v. 10, n. 12, 2015.
- LIU, Q. *et al.* The repetitive DNA landscape in *Avena* (Poaceae): chromosome and genome evolution defined by major repeat classes in whole-genome sequence reads. **BMC Plant Biology**, v. 19, n. 226, 2019.
- LÓPEZ-FLORES, I.; GARRIDO-RAMOS, M. A. The Repetitive DNA content of eukaryotic genomes. **Genome Dyn. Basel**. v. 7, p. 1-28, 2012.
- LYOZIN, G. T. *et al.* The structure and evolution of Penelope in the virilis species group of *Drosophila*: an ancient lineage of retroelements. **Journal of Molecular Evolution**, v. 52, n. 5, p. 445-456, 2001.
- MACAS, J. *et al.* In Depth Characterization of Repetitive DNA in 23 Plant Genomes Reveals Sources of Genome Size Variation in the Legume Tribe *Fabeae*. **PloS one**, v. 10, n. 11, 2015.
- MACAS, J. *et al.* Next generation sequencing-based analysis of repetitive DNA in the model dioecious plant *Silene latifolia*. **PloS one**, v. 6, n. 11, 2011.
- MANZANILLA, V.; BRUNEAU, A. Phylogeny reconstruction in the Caesalpinieae grade (Leguminosae) based on duplicated copies of the sucrose synthase gene and plastid markers. **Molecular Phylogenetics and Evolution**, v. 65, n. 1, p. 149-162, 2012.
- MARAZZI, B. *et al.* Phylogenetic relationships within *Senna* (Leguminosae, Cassiinae) based on three chloroplast DNA regions: Patterns in the evolution of floral symmetry and extrafloral nectaries. **American journal of botany**, v. 93, n. 2. p. 288-303, 2006.

MATOS, L. P. *et al.* Análise citogenética em 16 espécies dos gêneros *Senna* Mill. e *Cassia* L. (Leguminosae), com ênfase nas espécies ocorrentes na Bahia. *In.*: Seminário De Iniciação Científica, 15., 2011, Feira de Santana. **Anais [...]**. Feira de Santana: UEFS, 2011. p. 114-117.

MCCLINTOCK, B. The origin and behaviour of mutable loci in maize. **Proceedings of the National Academy of Sciences**, v. 36, n. 6, p. 344–355, 1950.

MEHROTRA, S.; GOYAL, V. Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. **Genomics Proteomics Bioinformatics**, v. 12, n. 4, p. 164-171, 2014.

MISHRA, P. *et al.*, DNA barcoding: an efficient tool to overcome authentication challenges in the herbal market. **Plant Biotechnology Journal**, v. 14, p. 8-21, 2016.

MITA, P.; BOEKE, J. D. How retrotransposons shape genome regulation. **Current Opinion in Genetics and Development**, v. 37, p. 90-100, 2016.

MLINAREC, J. *et al.* The repetitive DNA composition in the natural pesticide producer *Tanacetum cinerariifolium*: interindividual variation of subtelomeric tandem repeats. **Frontiers in Plant Science**, v. 10, n. 613, p. 1-14, 2019.

NEUMANN, P. *et al.* Significant expansion of *Vicia pannonica* genome size mediated by amplification of a single type of giant retroelement. **Genetics**, v. 173, n. 2, p 1047-1056, 2005.

NISHIHARA, H. Retrotransposons spread potential *cis*-regulatory elements during mammary gland evolution. **Nucleic Acids Research**, v. 47, n. 22, p. 11551-11562, 2019.

NOVÁK, P. *et al.* Repeat Explorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. **Bioinformatics**, v. 29, n. 6, p. 792-793, 2013.

OHRI, D.; KUMAR, A.; PAL, M. Correlations between 2C DNA values and habit in *Cassia* (Leguminosae: Caesalpinioideae). **Plant Systematics and Evolution**, v. 153, p. 223-227, 1986.

PAGAN, H. J. T. *et al.* Survey sequencing reveals elevated DNA transposon activity, novel elements, and variation in repetitive landscapes among vesper bats. **Genome Biology and Evolution**, v. 4, n. 4, p. 575-585, 2012.

PELLERIN, R. J.; WAMINAL, N. E.; KIM, H. H. Fish mapping of rDNA and telomeric repeats in 10 *Senna* species. **Horticulture, Environment, and Biotechnology**, v. 60, n. 2, p. 253-260, 2019.

PELLICER, J. *et al.* A universe of dwarfs and giants: genome size and chromosome evolution in the monocot family Melanthiaceae. **New Phytologist**, v. 201, p. 1484-1497, 2014.

PLOHL, M.; MESTROVIC, N.; MRAVINAC, B. Centromere identity from the DNA point of view. **Chromosoma**, v. 123, n. 4, p. 313-325, 2014.

POULTER, R. T. M.; GOODWIN, T. J. D. DIRS-1 and the other tyrosine recombinase retrotransposons. **Cytogenetic and Genome Research**, v. 110, n. 1-4, p. 575-588, 2005.

PRAKHONGCHEEP, O. *et al.* Lack of satellite DNA species-specific homogenization and relationship to chromosomal rearrangements in monitor lizards (Varanidae, Squamata). **BMC Evolutionary Biology**, v. 17, n. 193, 2017.

RESENDE, K. *et al.* Polyploidy and apomixis in accessions of *Senna rugosa* (G.Don) H.S.Irwin & Barneby. **Turkish Journal of Biology**, v. 38, n. 4, p. 510-515, 2014.

RESENDE, K. F. M.; DAVIDE, L. C.; TORRES, G. A. Chromosome number and meiosis in populations of *Senna* species (Caesalpinioideae - Fabaceae) from Southeast Brazil. **Caryologia**, v. 66, n. 1, p. 1-5, 2013.

RHO, M. *et al.* De novo identification of LTR retrotransposons in eukaryotic genomes. **BMC genomics**, v. 8, n. 90, 2007.

RUIZ-RUANO, F. J. *et al.* High-throughput analysis of the satellitome illuminates satellite DNA evolution. **Scientific Reports**, v. 6, n. 28333, 2016.

SEETHAPATHY, G. S. *et al.* Assessing product adulteration in natural health products for laxative yielding plants, *Cassia*, *Senna*, and *Chamaecrista*, in southern India using DNA barcoding. **International Journal of Legal Medicine**, v. 129, p. 693-700, 2014.

SILVA, D. M. Z. A. *et al.* Chromosomal organization of repetitive DNA sequences in *Astyanax bockmanni* (Teleostei, Characiformes): dispersive location, association and co-localization in the genome. **Genetica**, v. 141, p. 329-336, 2013.

SINGH, V. **Monograph of Indian subtribe Cassiinae (Caesalpinaceae)**. Jodhpur: Scientific Publishers, 2001.

SOUZA, M. G. C.; BENKO-ISEPPON, A. M. Cytogenetics and chromosome banding patterns in Caesalpinioideae and Papilionoideae species of Pará, Amazonas, Brazil. **Botanical Journal of the Linnean Society**, v. 144, p. 181-191, 2004.

SOUZA, V. C.; LORENZI, H. **Botânica sistemática: Guia ilustrado para identificação das famílias de fanerógamas nativas e exóticas no Brasil, baseado em APG II**. 2. ed. Nova Odessa: Instituto Plantarum, 2008.

SPRINGER, N. M.; LISCH, D.; LI, Q. Creating order from chaos: epigenome dynamics in plants with complex genomes. **The Plant Cell**, v. 28, p. 314-325, 2016.

SUCRE, Y. C. M. **Análise da diversificação da heterocromatina do grupo Caesalpinia (Leguminosae) baseado em dados citomoleculares e genômicos**. 2020. Dissertação (Mestrado em Biologia Vegetal) - Universidade Federal de Pernambuco, Recife, 2020.

TRALDI, J. B. *et al.* Chromosome Analyses of *Apareiodon argenteus* and *Apareiodon davisi* (Characiformes, Parodontidae): An Extensive Chromosomal Polymorphism of 45S and 5S Ribosomal DNAs. **Zebrafish**, v. 13, n. 1, p. 19-25, 2016.

UGARKOVIĆ, D.; PLOHL, M. Variation in satellite DNA profiles - causes and effects. **EMBO J.** v. 21, n. 22, p. 5955-5959, 2002.

VARANI, A. M. *et al.* Elementos de transposição: classificação e mecanismos de mobilização. *In.*: CARARETO, C. M. A.; VITORELLO, C. B. M.; SLUYS, M-A. V. **Elementos de transposição: diversidade, evolução, aplicações e impacto nos genomas dos seres vivos**. 22. ed. Rio de Janeiro: Fiocruz, 2015. p. 11-42.

VICIENT, C. M.; CASACUBERTA, J. M. Impact of transposable elements on polyploid plant genomes. **Annals of botany**, Oxford, v. 120, n. 2, p. 195-207, 2017.

WAMINAL, N. E. *et al.* Chromosomal Mapping of Tandem Repeats Revealed Massive Chromosomal Rearrangements and Insights Into *Senna tora* Dysploidy. **Frontiers in plant science**, v. 12, 2021.

WEISS-SCHNEEWEISS, H. *et al.* Employing next generation sequencing to explore the repeat landscape of the plant genome. *In.*: HORÄNDL, E.; APPELHANS, M. S. (eds.). **Next generation sequencing in plants systematics**. [S. l.]: International Association for Plant Taxonomy, 2015. Capítulo 5, p. 1-25.

WERREN, J. H. Selfish Genetic Elements, Genetic Conflict, and Evolutionary Innovation. **Proceedings of the National Academy of Sciences**, v. 108, n. 2, p. 10863-10870, 2011.

WICKER, T. *et al.* A unified classification system for eukaryotic transposable elements. **Nature Reviews Genetics**, v. 8, p. 973-982, 2007.

WYLER, M. *et al.* Impact of transposable elements on methylation and gene expression across natural accessions of *Brachypodium distachyon*. **Genome Biology and Evolution**, v. 180, 2020.

ZHAO, D.; FERGUSON, A. A.; JIANG, N. What makes up plant genomes: the vanishing line between transposable elements and genes. **Biochimica et Biophysica Acta**, v. 1859, n. 2, p. 366-380, 2016.

ZHU, Q. *et al.* Repetitive sequence analysis and karyotyping reveal different genome evolution and speciation of diploid and tetraploid *Tripsacum dactyloides*. **The Crop Journal**, v. 4, n. 4, p. 247-255, 2016.