

UNIVERSIDADE FEDERAL DE ALAGOAS-UFAL
CAMPUS ARAPIRACA
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRICULTURA E AMBIENTE-PPGAA

GRAZIELE CLEMENTE COSTA

**ABUNDÂNCIA DIFERENCIAL DE FAMÍLIAS DE DNA SATÉLITE EM *Asclepias*
(APOCYNACEAE): UMA CARACTERIZAÇÃO DE SEQUÊNCIAS ESPECÍFICAS
DE ESPÉCIES**

ARAPIRACA
2019

GRAZIELE CLEMENTE COSTA

ABUNDÂNCIA DIFERENCIAL DE FAMÍLIAS DE DNA SATÉLITE EM *Asclepias*
(APOCYNACEAE): UMA CARACTERIZAÇÃO DE SEQUÊNCIAS ESPECÍFICAS DE
ESPÉCIES

Dissertação de mestrado apresentada ao Programa de pós-graduação em Agricultura e Ambiente da Universidade Federal de Alagoas – Campus Arapiraca, como requisito parcial para obtenção de título de Mestra em Agricultura e Ambiente.

Orientador: Prof. Dr. Cicero Carlos de Souza Almeida.

ARAPIRACA
2019

Universidade Federal de Alagoas – UFAL
Campus Arapiraca
Bibliotecário Responsável: Nestor Antônio Alves Junior
CRB - 4 / 1557

C837a Costa, Grazielle Clemente
Abundância diferencial de famílias de DNA satélite em *Asclepias*
(APOCYNACEAE): uma caracterização de seqüências específicas de espécies /
Grazielle Clemente Costa. – Arapiraca, 2019.

41 f.

Dissertação [Mestrado] – Programa de Pós-Graduação em Agricultura e
Ambiente - PPGAA, Universidade Federal de Alagoas – UFAL, Campus Arapiraca,
Arapiraca, 2019.

Orientador: Prof. Dr. Cícero Carlos de Souza Almeida .

Bibliografia: p. 36-41.

1. Sinal filogenético. 2. DNA satélite. 3. *Asclepias*. I. Almeida, Cícero Carlos de
Souza. II. Título.

CDU: 631

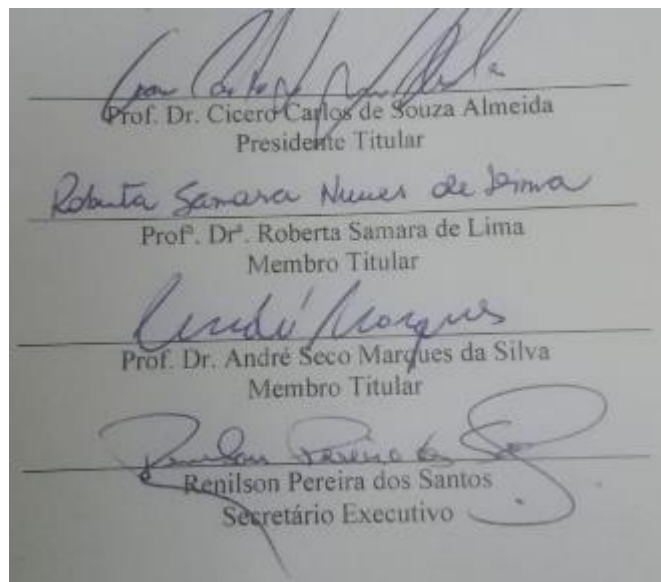
GRAZIELE CLEMENTE COSTA

**ABUNDÂNCIA DIFERENCIAL DE FAMÍLIAS DE DNA SATÉLITE EM *Asclepias*
(APOCYNACEAE): UMA CARACTERIZAÇÃO DE SEQUÊNCIAS ESPECÍFICAS
DE ESPÉCIES**

Dissertação de mestrado apresentada ao Programa de pós-graduação em Agricultura e Ambiente da Universidade Federal de Alagoas – Campus Arapiraca, como requisito parcial para obtenção de título de Mestre em Agricultura e Ambiente.

Data de Aprovação: 28/05/2019

Banca Examinadora



AGRADECIMENTOS

Agradeço a Deus acima de tudo, por toda proteção, amor, e por me proporcionar a coragem para enfrentar e superar os desafios a mim colocados.

Aos meus pais, em especial à minha mãe Ednalva, minha maior incentivadora e exemplo.

Ao meu orientador e grande pesquisador Cicero Carlos pela oportunidade, confiança e paciência.

Ao meu namorado Franklin Moismar por também ter sido muito paciente e me ouvido nos momentos de angústia onde tudo que eu precisava era ser entendida e apoiada, obrigada amor!

Aos amigos e colegas do Laboratório de Recursos Genéticos - LARGE, por todo aprendizado e companheirismo, em especial à minha amiga Gleica Martins, que com toda sua paciência e ciência não media esforços para me dar todo apoio.

À minha amiga/irmã Symonne Fernandes que sempre está do meu lado apesar de toda chatice do mundo.

Aos meus amigos/colegas de curso (sem citar nomes), por todos os laços ajustados e pelo apoio coletivo e mútuo que dávamos uns aos outros.

Ao corpo docente do Programa de Pós-Graduação, que com todo o compartilhamento de conhecimentos, foram de extrema importância para minha qualificação profissional e pessoal.

Ao secretário do Programa Renilson por toda ajuda burocrática durante o curso, pelo esclarecimento de dúvidas, pela torcida, paciência e amizade.

À Universidade Federal de Alagoas – *Campus* Arapiraca, por ser minha segunda casa durante o curso e à FAPEAL pela concessão da bolsa.

Enfim, obrigada a todos que contribuíram direta ou indiretamente para a concretização deste trabalho, desta espata, deste sonho.

RESUMO

Sequências de DNA satélite compõem juntamente com elementos móveis e DNA ribossomal a fração repetitiva dos genomas eucarióticos. Esta por sua vez, compreende valores que variam de acordo com a espécie, podendo ocupar até 90% do tamanho do genoma. As sequências de DNA satélites apresentam consideráveis taxas de evolução e diversificação intraespecífica e interespecífica, características que as tornam ferramentas importantes em estudos evolutivos ajudando a elucidar relações de parentesco entre as espécies bem como a entender aspectos relacionados à importância do DNA repetitivo. O gênero *Asclepias* é um dos mais importantes da família Apocynaceae e é usado mundialmente como modelo nas mais diversas áreas de pesquisa devido a grande quantidade de dados gerados a partir das tecnologias de Sequenciamento de Nova Geração (NGS). Partindo do pressuposto que a evolução do DNA satélite acompanha a evolução do gênero *Asclepias* e possibilita a reconstrução da filogenia do grupo. Neste sentido o objetivo deste trabalho foi analisar a fração repetitiva e a abundância de DNA satélite em 77 espécies de *Asclepias*. Para isso, foi construída uma filogenia de máxima verossimilhança de 77 espécies de *Asclepias* usando regiões gênicas concatenadas do DNA de cloroplastos a partir de *reads* obtidos do NCBI; em seguida foram identificadas as sequências de DNA satélite de *A. syriaca*, *A. curassavica*, *A. coulteri* e *A. fournieri* usando agrupamentos baseados em gráficos e a abundância dessas sequências foi analisada em 77 espécies de *Asclepias*; e por fim feita análise de correlação pela estatística lambda entre abundância de DNAs satélites e as topologias estimadas. As inferências filogenéticas e as topologias corroboraram com as inferências de estudos anteriores, indicando uma filogenia de alto suporte. Foram detectados onze satélites. O Sat1 apresentou seis famílias evolutivas distribuídas em todas as espécies, em que quatro destas (F1, F2, F5 e F6) se destacaram pela diferentes abundâncias entre os principais clados, com alta correlação com as topologias da árvore. Essa correlação foi evidenciada também nos Sat2 e Sat3, sendo estes grupos de sat uma característica evolutiva que pode ser utilizada para inferir sobre a diversificação do gênero. Aqui concluímos que o gênero *Asclepias* é caracterizado por muitos DNAs satélites e que a abundância dessas sequências de satélites está de acordo com a filogenia.

Palavras-chave: Sinal filogenético. DNA satellite. *Asclepias*.

ABSTRACT

Sequences of satellite DNA make up together with mobile elements and ribosomal DNA the repetitive fraction of eukaryotic genomes. This, in turn, comprises values that vary according to the species, and can occupy up to 90% of the size of the genome. Satellite DNA sequences present considerable intraspecific and interspecific evolution and diversification rates, characteristics that make them important tools in evolutionary studies, helping to elucidate kinship relationships between species as well as to understand aspects related to the importance of repetitive DNA. The genus *Asclepias* is one of the most important of the Apocynaceae family and is used worldwide as a model in the most diverse areas of research due to the large amount of data generated from the New Generation Sequencing (NGS) technologies. Based on the assumption that the evolution of the satellite DNA accompanies the evolution of the genus *Asclepias* and enables the reconstruction of the phylogeny of the group. In this sense, the objective of this work was to analyze the repetitive fraction and the abundance of satellite DNA in 77 species of *Asclepias*. For this, a maximum likelihood phylogeny of 77 species of *Asclepias* was constructed using concatenated DNA regions of chloroplasts from reads obtained from NCBI; then the satellite DNA sequences of *A. syriaca*, *A. curassavica*, *A. coulteri* and *A. fournieri* were identified using plot-based clusters and the abundance of these sequences was analyzed in 77 species of *Asclepias*; and finally performed correlation analysis by the lambda statistic between abundance of satellite DNAs and the estimated topologies. Phylogenetic inferences and topologies corroborate with the inferences of previous studies, indicating a high support phylogeny. Eleven satellites were detected. The Sat1 presented six evolutionary families distributed in all species, in which four of these (F1, F2, F5 and F6) were distinguished by the different abundances between the main clades, with a high correlation with the topologies of the tree. This correlation was also evidenced in Sat2 and Sat3, these sat groups being an evolutionary trait that can be used to infer gender diversification. Here we conclude that the genus *Asclepias* is characterized by many satellite DNAs and that the abundance of these satellite sequences is in agreement with the phylogeny.

Keywords: Phylogenetic sign. DNA satellite. *Asclepias*.

SUMÁRIO

1 INTRODUÇÃO GERAL	7
2 REVISÃO DE LITERATURA.....	9
2.1 GÊNERO <i>Asclepias</i>	9
2.2 DNA PLASTIDIAL	12
2.3 DNA NUCLEAR.....	13
2.3.1 DNA Repetitivo	13
2.3.1.1 DNA satélites	14
2.3.1.2 DNA ribossomal.....	16
2.3.1.3 Elementos móveis.....	17
2.3.1.3.1 Retrotransposons	17
2.3.1.3.2 Transposons	18
2.4 EVOLUÇÃO DE SATÉLITES E SINAL FILOGENÉTICO	19
3 OBJETIVOS	22
3.1 Geral	22
3.2 Específicos	22
4 MATERIAL E MÉTODOS	23
4.1 Dados de Sequenciamento de Rendimento.....	23
4.2 Análises filogenéticas.....	24
4.3 Agrupamento baseado em graficos das leituras de DNA.....	25
5 RESULTADOS	26
5.1 Árvore filogenética	26
5.2 Composição de Repetições	28
5.3 Sequências de satDNA.....	30
6 DISCUSSÃO	33
6.1 Árvore filogenética	33
6.2 Sequências de DNA repetitivo	33
6.3 Sequências de DNA de satélite.....	34
7 CONCLUSÕES.....	35
REFERÊNCIAS	36

1 INTRODUÇÃO GERAL

As sequências de DNA repetitivo estão presentes em todas as plantas superiores, podendo representar até 90% do seu genoma (MEHROTRA; GOYAL, 2014). Elas correspondem à fração mais importante do ponto de vista evolutivo (WICKER et al., 2007), podendo estar dispostos em conformações adjacentes denominadas em *tandem*, como é o caso de sequências de DNA satélite e DNA ribossomal, ou inseridas de forma dispersa e intercalada no genoma, como os elementos móveis (PISUPATI; VERGARA; KANE, 2018).

Os DNAs satélites (geralmente a segunda classe mais abundante) são caracterizados por matrizes denominadas monômeros que, a depender do seu tamanho, podem ser classificados em microssatélites, com repetições de 2 a 6 pares de bases (pb), minissatélites, de 6 a 100 pb, e os satélites propriamente ditos, com monômeros que variam de 150 a 400 pb. Eles estão geralmente localizados nas regiões terminais e centroméricas dos cromossomos, sendo os principais componentes da heterocromatina (MEHROTRA; GOYAL, 2014).

As sequências de satélites apresentam altas taxas de diversificação (MACAS; MESZÁROS; NOUZOVÁ, 2002) que associadas a eventos de evolução combinada resultam em sequências homogêneas (MEHROTRA; GOYAL, 2014). Essa homogeneidade agregada à abundância dessas frações dentro do genoma confere o suporte para o uso de sequências de satélites em estudos que visam estimar melhor as relações de evolução entre espécies e dessa forma esclarecer e/ou corroborar com os resultados obtidos em estudos filogenéticos feitos com árvores gênicas (DODSWORTH et al., 2015).

O uso de sequências de satélites como ferramenta para a compreensão de relações evolutivas em nível de gênero e em nível de família de plantas vem se tornando comum nos últimos anos (MACAS et al., 2015; DODSWORTH et al., 2015). No entanto, ainda são necessários estudos sobre a abundância e a evolução de sequências de DNA satélite envolvendo um grande número de espécies dentro do mesmo gênero, podendo esta abordagem contribuir ainda mais para a compreensão da evolução das frações repetitivas em genomas eucarióticos.

Nesse sentido, usamos a grande quantidade de dados de Sequenciamento de Nova Geração (NGS) disponíveis no banco de dados do NCBI (National Center for Biotechnology Information) sobre o gênero *Asclepias* para analisar a evolução de sequências de DNA satélite em 77 espécies do grupo. Este gênero é um dos mais importantes da família Apocynaceae com suas 490 espécies distribuídas de forma pantropical (FONTELLA-PEREIRA; KONNO, 2005). As análises filogenéticas de 103 espécies e 5 subespécies do gênero sustentaram quatro

clados bem suportados, descritos como (1) clado Sonoran Desert; (2) clado Incarnatae, incluindo as espécies sul-americanas; (3) clado Temperate North American, que consiste nas espécies endêmicas dos Estados Unidos e Canadá e algumas espécies mexicanas e (4) clado Mexican Highland com a maioria das espécies endêmicas do México, América Central e algumas dos Estados Unidos da América (Fishbein et al., 2018). Partindo do pressuposto que a evolução do DNA satélite acompanha a evolução do gênero *Asclepias* e possibilita a reconstrução da filogenia do grupo, este trabalho objetivou analisar a fração repetitiva e a abundância de DNA satélite em 77 espécies de *Asclepias*.

2 REVISÃO DE LITERATURA

2.1 GÊNERO *Asclepias*

O gênero *Asclepias*, também conhecido como *milkweeds* (FISHBEIN et al., 2011) é um dos mais importantes da família Apocynaceae, com cerca de 490 espécies distribuídas de forma pantropical, ou seja, estão distribuídas nas regiões tropicais dos continentes (FONTELLA-PEREIRA; KONNO, 2005).

De acordo com Fishbein et al. (2011) existe uma maior incidência de *Asclepias* na América do Norte se comparada à América do Sul. Nesse sentido, o site Flora do Brasil (2019) descreve como espécies de *Asclepias* nativas do país, *A. candida*, *A. mellodora* e *A. curassavica*. *A. aequicorну*, *A. bracteolata* e *A. langsdorffii*. Dentre as espécies mencionadas, *A. curassavica* é a única com ocorrência no estado de Alagoas.

Plantas do gênero *Asclepias* são conhecidas por sua seiva leitosa, que consiste em um látex que nada mais é que um mecanismo de defesa contra possíveis predadores (PELLISSIER et al., 2016). Nele estão presentes os terpenóides, seu principal componente (FISHBEIN et al., 2018), e os cardenolídeos (presentes em quantidades variáveis em cada espécie), estes últimos são esteroides de sabor amargo que possuem valor terapêutico para humanos agindo principalmente no miocárdio através do rompimento do fluxo de sódio e potássio, além de ser extremamente venenoso para vertebrados em doses elevadas (MALCOLM, 1991). Além de tornar as plantas impalatáveis, os cardenolídeos presentes contribuem para a redução do consumo de tecidos das plantas por herbívoros (PELLISSIER et al., 2016).

Algumas espécies de *Asclepias* são consideradas tóxicas, é o caso de *A. curassavica* que vem sendo constantemente associada à morte de bovinos no Brasil (TOKARNIA; BRITO; CUNHA, 2001), e, portanto sendo responsável por perdas econômicas diretas e indiretas. As primeiras estão relacionadas à própria morte do animal, diminuição das taxas de reprodução, diminuição das taxas de produção de leite e carne, e aumento de susceptibilidade a doenças. Já as perdas indiretas estão relacionadas ao manejo dessas plantas tóxicas, envolvendo para isso a produção de cercas, gastos associados aos diagnósticos das intoxicações, além do tratamento do animal afetado ou reposição de bovinos mortos (TOKARNIA et al., 2000).

Nesse contexto, espécies do gênero são utilizadas mundialmente como modelos em estudos relacionados à defesa de plantas (PELLISSIER et al., 2016), ecologia, reprodução (WYATT; BROYLES, 1994), coevolução de plantas e herbívoros (relação existente entre *A.*

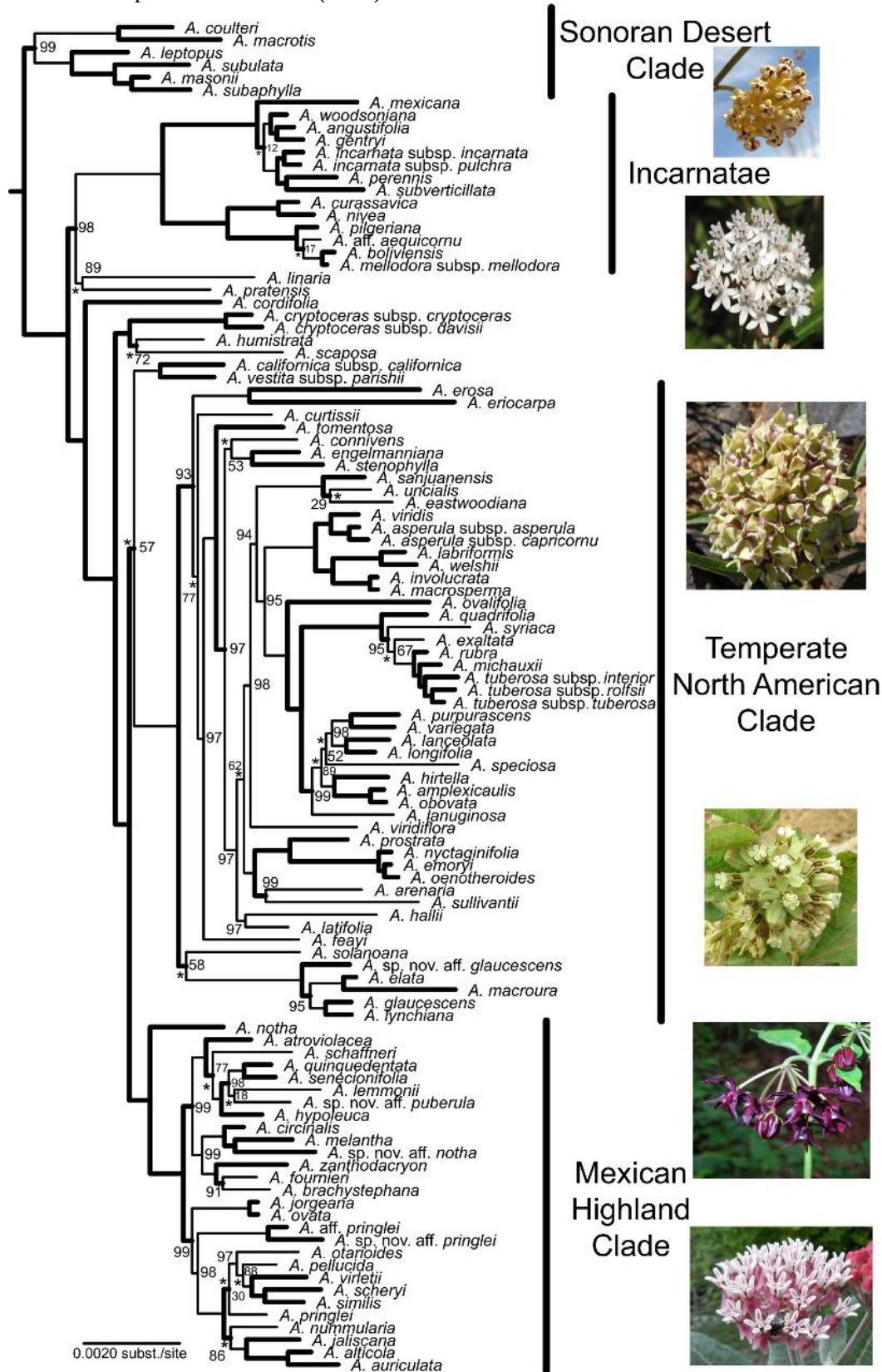
curassavica e a borboleta monarca - *Dannaus plexippus*) (MALCOLM, 1991), estudos baseados em DNA plastidial não codificante (FISHBEIN et al., 2011), DNA ribossômico nuclear, DNA mitocondrial (STRAUB et al., 2012), e mais recentemente estudo baseado em sequências plastômicas contribuindo bastante para o esclarecimento de muitas relações entre as espécies do gênero (FISHBEIN et al., 2018) (Figura 1).

No que tange a aspectos evolutivos, Woodson (1954) foi o primeiro a estudar e classificar espécies norte americanas do gênero usando para isso a comparação taxonômica entre as espécies, tendo como resultado nove subgêneros, apresentando o maior deles (subgênero *Asclepias*) oito séries incluídas.

Apesar de os resultados obtidos por Woodson terem sido considerados a espinha dorsal da árvore filogenética de *Asclepias* norte americanas, estudos posteriores realizados por Fishbein et al. (2011) utilizando três loci do gene plastidial rpl16 intron, (trnCGCA-rpoB espaçador, seu adjacente trnSGCU-trnG UUC espaçador e trnGUUC intron) identificaram treze clados para o gênero os quais apresentaram relações internas pouco resolvidas, e principalmente, demonstram que a classificação de Woodson em subgêneros e séries não refletia a filogenia do grupo, deixando clara a necessidade de estudos mais aprofundados sobre a temática.

Diante disso, através do estudo com plastomas parciais de 103 espécies de *Asclepias* e 5 subespécies Fishbein et al. (2018) propuseram a existência de quatro clados bem definidos para o gênero, são eles: Sonoran Desert (espécies sem folhas, arbustivas), Incarnatae (com todas as espécies sul-americanas), Temperate North American (maioria das espécies endêmicas dos Estados Unidos da América, Canadá e algumas espécies do México) e Mexican Highland (maioria das espécies endêmicas do México, América Central e algumas dos Estados Unidos da América), dos quais surpreendentemente (devido a grandes diferenças morfológicas e adaptativas) Sonoran Desert e Incarnatae foram colocados como irmãos. Aqui, diferente da abordagem apresentada por Fishbein et al. (2018), inferimos as relações filogenéticas de 77 espécies de *Asclepias* através da evolução de sequências de satélites dentro da filogenia com a finalidade de explicar melhor e/ou corroborar com os dados já existentes.

Figura 1 – Árvore de Máxima Verossimilhança (ML) baseada em seqüências de plastomas proposta por Fishbein et al. (2018).



Fonte: Fishbein et al. (2018).

2.2 DNA PLASTIDIAL

Os cloroplastos são organelas presentes em plantas e que estão diretamente associadas à fotossíntese e a fixação de carbono (KOLODNER e TEWARI, 1979; 1972). Acredita-se que a origem de cloroplastos em células vegetais surgiu há mais de um bilhão de anos (EMBLEY; MARTIN, 2006) através de uma associação endossimbiótica, onde cianobactérias foram fagocitadas por células eucariontes primitivas, formando assim, os cloroplastos e consequentemente, as células fotossintetizantes (ARCHIBALD, 2009). Essas organelas possuem material genético próprio, o DNA plastidial, também chamado de cpDNA ou plastoma (KOLODNER; TEWARI, 1979;1972), presente de 1.000 a 10.000 cópias por célula (RAVI et al., 2008). Além disso, é a estrutura mais estudada dos genomas de plantas, e tem como característica um formato circular (KOLODNER; TEWARI, 1972).

Estudos indicam habilidade de exportação de genes dos cloroplastos para o núcleo (GREEN, 2011). Essa transferência resultou na redução do tamanho do cpDNA e em um sistema genético eucariótico compartimentado (TIMMIS et al., 2004) sugerindo assim, que esses genomas não são totalmente independentes. Sabe-se hoje que a maioria das proteínas produzidas nos cloroplastos são codificadas pelo genoma nuclear e depois importadas para essas organelas (RAVI et al., 2008; LEISTER, 2003).

Os genomas plastidiais apresentam herança uniparental. Em angiospermas, essa herança é predominantemente materna, tendo raros casos de herança paterna ou biparental (ZHANG; SODMERGEN, 2010). No que diz respeito às gimnospermas, possuem em sua maior parte origem paterna (JANSEN e RUHLMAN, 2012).

Quanto a organização, o tamanho do cpDNA varia podendo chegar a 230 kilobases (kb) (FIGUEROA-MARTINEZ et al., 2017), numa estrutura “quadripartida”, com uma região longa de cópia única (LSC) e uma região pequena de cópia única (SSC) separadas por dois grandes *repeats* invertidos (IRA e IRB) (GREEN, 2011). As regiões IR são idênticas, porém possuem sentido de leitura contrário (MAIER et al., 1995).

De forma geral, os genes encontrados no cpDNA são conservados no que diz respeito ao conteúdo e à ordem, principalmente em grupos específicos de organismos e podem ser organizados em três grandes categorias. A primeira compreende os genes do aparato fotossintético, incluindo os fotossistemas I (psaA, psaB, por exemplo) e II (psbA e psbB), citocromo b6f (petA, petB, etc), ATP sintase (atpA, atpB, etc) RuBisCo, (rbcL), e genes de NAD (P) H desidrogenase (ndhA, ndhB, etc). A segunda categoria engloba genes de RNA e genes do aparato genético, como RNA ribossômico (rrn16, rrn5, etc), RNA de transferência

(trnH, trnK), RNA polimerase (rpoA, rpoB) e genes de subunidade ribossômica (rps2, rps3, rpl2, rpl16, etc). Já a terceira envolve os ORFs (quadros de leitura aberta) conservados chamados ycfs além dos potenciais genes codificadores de proteínas matK e cemA (RAVI et al., 2008).

Além disso, sabe-se que a frequência de mutação das regiões IR é significativamente menor que nas regiões de cópia única (MAIER et al., 1995) e a variação de tamanho dessas regiões é a principal, mas não a única, causa de variação do tamanho do cpDNA entre os gêneros, sendo a perda de genes, diferenças nas regiões intergênicas e rearranjos outros motivos importantes de variação desse tamanho (RAVI, et al., 2008).

Ravi et al., (2008) ainda afirmam que por serem as frações mais conservadas do plastoma, os IRs atuam como regiões estabilizadoras, evoluindo de forma duas ou três vezes mais lenta que as regiões LSC e SSC. Dentre as regiões de cópia única, a região longa é a menos conservada.

Conhecer as regiões dos plastomas tem se mostrado de extrema importância em estudos filogenéticos e filogeográficos, permitindo assim, inferir sobre os relacionamentos intra e interespecificamente, além de inferências das relações demográficas da biodiversidade nos mais diversos biomas (SCARCELLI et al., 2011).

2.3 DNA NUCLEAR

A maior parte do DNA de plantas está presente no núcleo celular, se comparado às mitocôndrias e aos cloroplastos, organelas que também apresentam material genético próprio (SCHIFINO-WITTMANN, 2001). O tamanho dos genomas nucleares das espécies varia muito em decorrência da quantidade de sequências repetitivas existentes, no entanto, não há evidência de relação entre tamanho dos genomas eucarióticos e a complexidade dos mesmos. É o famoso paradoxo do valor constante, ou valor C (quantidade de DNA haploide em um núcleo) (MEHROTRA et al., 2014). As frações repetitivas do genoma incluem famílias de elementos móveis, sequências de satélites e DNA ribossomal (MACAS et al., 2015).

2.3.1 DNA Repetitivo

A maior parte dos genomas eucarióticos típicos é constituída de frações repetitivas (HAWBOLD; WIEHE, 2006). Sua presença, juntamente com eventos de poliploidização (MACAS et al., 2015) estão diretamente relacionados ao tamanho desses genomas, que

podem variar muito entre as espécies, como é o caso de *Paris japonica* com mais de 152 Gb (PELLICER; FAY; LEITCH, 2010) e *Arabidopsis thaliana* com 150 Mpb (PISUPATI; VERGARA; KANE, 2018).

As proporções de sequências repetitivas também variam de acordo com a espécie, podendo chegar a 90% do tamanho do genoma (MEHROTRA; GOYAL, 2014). Em *A. thaliana* 20% do seu genoma é constituído de frações repetitivas. *Helianthus annuus*, o girassol, apresenta 80% do seu material genético constituído por frações repetitivas (GILL et al., 2014), e em *Cannabis sativa* as taxas variam de 64 a 65% (PISUPATI, VERGARA e KANE, 2018). Em *Vicia sylvatica* os retrotransposons de LTR representam 81% do seu genoma nuclear, e em *Vicia faba* os elementos Ogre (elementos LTR pertencentes à superfamília Ty3Gypsy) representam sozinhos 54% do genoma (MACAS et al., 2015).

Essas frações provavelmente possuem efeitos negativos e positivos dentro do genoma, agindo como matéria prima para a produção de novos genes (MARTIN et al., 2009), alterando a função de outros (ZHOU et al., 2004), atuando na evolução de genomas (MACAS et al., 2015), na estruturação e reestruturação física de genomas maiores, incluindo rearranjos cromossômicos que podem contribuir de forma significativa para o fenômeno da especiação (FARIA; NAVARRO, 2010), podendo ainda estar relacionadas à ligação do cromossomo ao fuso mitótico, ao arranjo cromossômico, a interação de proteínas da cromatina, a ligação de histonas, a determinação da estrutura de cromossomos e a evolução cariotípica (MEHROTRA; GOYAL, 2014).

As frações repetitivas dos genomas eucarióticos são muito complexas (MACAS et al., 2015) e podem ser compreendidas em dois grandes grupos de sequências, as que se repetem de forma adjacente, chamadas de repetições em *tandem*, como é o caso dos DNA satélites e rDNA (45S e 5S), e as que estão presentes de forma intercalada no genoma, como os elementos transponíveis (PISUPATI; VERGARA; KANE, 2018).

2.3.1.1 DNA satélites

Os DNA satélites (satDNA) são caracterizados por monômeros organizados lado a lado dentro do genoma. Essas unidades podem variar de tamanho, como é o caso do microsatélites também conhecidas como repetições simples repetidas (SSR) (ELLEGREN, 2004) com motivos de 2 a 5 pb, os minissatélites, de 6 a 100 pb, e os satélites propriamente ditos, com monômeros que geralmente variam de 150 a 400 pb (MEHROTRA; GOYAL, 2014). Essas sequências vêm sendo associadas a papéis estruturais no genoma, afetando a

expressão de genes por modificações epigenéticas (FELICIELLO; AKRAP; UGARKOVIC, 2015) induzidas por mudanças específicas no ambiente (ROBLEDILLO et al., 2018).

Os satélites estão geralmente localizados nas regiões terminais e centroméricas dos cromossomos, sendo os principais componentes da heterocromatina (MEHROTRA; GOYAL, 2014). As sequências de satélites localizadas nas regiões centroméricas são comumente encontradas em plantas e animais, já as encontradas nas regiões subteloméricas são específicas de gêneros (MEHROTRA et al., 2014).

Os satDNA estão entre os componentes mais dinâmicos dos genomas eucarióticos, e sua alta taxa evolutiva resulta em considerável diversificação de sequências (MACAS; MESZÁROS; NOUZOVÁ, 2002). Essa evolução por sua vez é coesa e combinada, o que culmina em altas taxas de homogeneidade na sequência (MEHROTRA; GOYAL, 2014).

Tanto a homogeneidade quanto a evolução das sequências de satélites estão relacionadas ao número de cópias, de forma que as sequências que possuem baixo número de cópias evoluem mais lentamente e de forma homogênea, diferentemente das sequências com alto número de cópias, que são mais heterogêneas e possui elevada taxa de evolução (MEHROTRA; GOYAL, 2014; STADLER et al., 1995).

Uma dada sequência de satélite pode ser particular de uma espécie, isso permite a identificação rápida e confiável da espécie (MEHROTRA; GOYAL, 2014), ou estar presente em muitas espécies ou famílias, sugerindo que algumas sequências repetitivas evoluem rapidamente, diferentemente de outras, que podem ser bem conservadas (ANAMTHAWAT-JÓNSSON; HESLOP-HARRISON, 1992). Um exemplo disso, é o satélite pCtKpn I-I isolado inicialmente em *Carthamus tinctorius* e outras espécies do mesmo gênero, e descoberto posteriormente em diversas famílias de angiospermas (MEHROTRA et al., 2014).

É inegável que a maior parte das frações repetitivas que aparecem nos genomas eucariontes são os elementos móveis, no entanto, os satélites são responsáveis por uma considerável proporção. Em *Vicia faba*, por exemplo, 937 Mpb/1C são constituídos por satélites, isso corresponde a 6,97% do genoma, em *Latirus sativus* o índice sobe para 10,7% (699 Mpb/1C) (MACAS et al., 2015). Ainda segundo Macas et al., (2015), os satélites também são frações significativas em genomas pequenos como *Vicia villosa*, com 12,3% (250Mpb/1C) e *Vicia sativa* com 10,4% (184 Mpb/1C) .

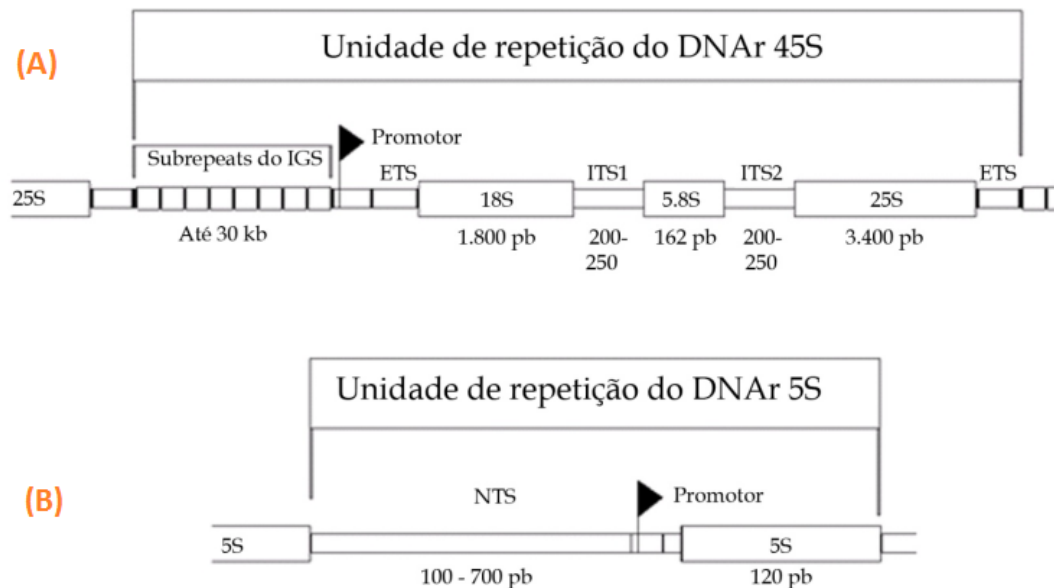
2.3.1.2 DNA ribossomal

Normalmente em proporção menor, mas altamente importante, o DNA ribossomal (rDNA) é outro tipo de sequência arranjada em sequências longas em *tandem* (PISUPATI; VERGARA; KANE, 2018) que vem contribuindo para o esclarecimento de eventos evolutivos (HILLIS; DIXON, 1991). De acordo com Pisupati et al. (2018) existem normalmente de 500 a 4000 cópias de rDNA em células diplóides de plantas a depender da espécie. Em *Cannabis sativa*, por exemplo, cerca de 2% do tamanho do genoma é composto por genes de rDNA. Em contrapartida, *Humulus lupulus* apresentou 0,1% de presença desses genes, apesar de ter o genoma maior que *C. sativa* (PISUPATI; VERGARA; KANE, 2018).

Existem duas subunidades de rDNA, a 45S e a 5S. A primeira (também conhecida com rDNA 35S), é constituída pelos genes 18S, 5.8S e 25S, separados por espaçadores transcritos internos não codificantes denominados ITS1 e ITS2 (*Internal transcribed spacer*), além de uma sequência intergênica conhecida como IGS (*Intergenic spacer*) entre as repetições adjacentes. Os espaçadores ITS e IGS são transcritos pela RNA polimerase-I e RNA polimerase-III respectivamente. Todas as regiões mencionadas, excetuando os espaçadores IGS, constituem juntamente com o espaçador transcrito externo ETS (*External Transcribed Spacer*) a região transcrita. A região intergência IGS, que não é transcrita, é localizada entre o terminal 3' do ETS próximo ao rRNA 25S e o terminal 5' do ETS, ao lado do rDNA 18S (LONG; DAWID, 1980).

A segunda subunidade possui apenas uma região codificante altamente conservada constituída pelo gene 5S e um espaçador intergênico. Cada unidade de repetição desse gene corresponde a 120 pb, que ao contrário dos demais RNA ribossomais, é transcrita fora do nucléolo. Tanto no rRNA 5S quanto no 45S a região codificante é altamente conservada e a espaçadora é variável entre as espécies (EICKBUSH; EICKBUSH, 2007) (figura 2).

Figura 2 – Organização do RNA ribossomal (rRNA) e suas regiões separadoras. Unidade de repetição do DNA ribossomal 45S (A) e 5S (B) e tamanho aproximado em pares de bases.



Fonte: OVALLE, (2011).

Ainda segundo Eickbush e Eickbush (2007), as sequências dos genes de rRNA são altamente conservadas em cada espécie, isso sugere o crossing - over desigual como a principal força motriz que impulsiona essa evolução, com trocas entre cromátides irmãs ocorrendo mais frequentemente do que a trocas entre homólogos. Essa evolução combinada é tão estável que os elementos transponíveis inseridos em grandes frações dos genes rRNA provavelmente têm pouco efeito sobre esse processo.

2.3.1.3 Elementos móveis

Os elementos transponíveis, também conhecidos como elementos móveis, são sequências que podem mover-se e duplicar-se dentro do genoma de acordo com a necessidade de presença ou ausência da enzima de transcrição reversa. No primeiro caso, são chamados de retrotransposons, ou classe I, e no segundo, de transposons, ou classe II (SCHIETGAT et al., 2018).

2.3.1.3.1 Retrotransposons

Os elementos de classe I, também conhecidos como retrotransposons compreendem o grupo de elementos móveis mais representativo (KEJNOVSKY, HAWKINS e FESCHOTTE 2012). Eles se movem pelo mecanismo de “copiar e colar” que consiste na transcrição

completa do elemento, dando origem a uma cópia de RNA que codifica a enzima transcriptase reversa (KIDWELL, 2002). Essa molécula pode dar origem a uma fita de DNA, resultando na formação de um intermediário híbrido DNA/RNA. Por fim, a fita de DNA utiliza a enzima integrase para se inserir em lugares aleatórios nos cromossomos (KAZAZIAN Jr., 2004).

São considerados os principais responsáveis pelo aumento ou diminuição do tamanho do genoma, sendo, geralmente, as diferenças na quantidade de retrotransposons uma das razões da alta variabilidade do tamanho de genomas de plantas, como na variação do tamanho do genoma do milho e do sorgo (CASACUBERTA, 2003).

Com base na presença ou ausência de repetições terminais longas (*Long terminal repeats* - LTR), necessárias para a transcrição e incorporação do elemento em outro local (KAPITONOV; JURKA, 2008), os retrotransposons podem ser classificados em retrotransposons de LTR e retrotransposons não-LTR (*Non long terminal repeat*) (GOGVADZE; BUZDIN, 2009). O primeiro grupo compreende as ordens LTR (superfamílias TY1-copia, BEL-PAO e TY-3 gypsy) e DIRS (superfamílias DIRS, Ngaro, VIPER). O segundo grupo (retrotransposons não-LTR) é representado pela ordem de elementos intercalados longos, LINE (superfamílias R2, RTE, Jockey, L1) (SHIETGAT et al., 2018), que codifica as proteínas necessárias para a transcrição reversa (não-LTR autônomos) (BOWEN; JORDAN, 2002), e a ordem dos elementos intercalados curtos, SINE (superfamílias tRNA, 7SL, 5S) (SHIETGAT et al., 2018), que não codificam as proteínas necessárias para a sua replicação e inserção (não-LTR não autônomos) (BOWEN; JORDAN 2002).

2.3.1.3.2 Transposons

Os elementos de classe II, também conhecidos como transposons, correspondem a 3% do genoma e possuem o mecanismo de “cortar e colar”, não necessitando de transcrição reversa (GOGVADZE; BUZDIN, 2009; GOODIER; KAZAZIAN JR., 2008). Para isso, a sequência inicial se mantém no local original, e a transposição é acompanhada por um aumento no número de cópias do transposon, através do mecanismo de “recorta e cola”, sendo excisado de um ponto e alocado em outro (KAZAZIAN Jr., 2004; WICKER, et al., 2007). Embora a maioria das transposições desses elementos ocorra em sítios genômicos menores do que 100 kb de seu sítio original, alguns são capazes de realizar pulos distantes (KAZAZIAN Jr, 2004).

Essa classe de elementos transponíveis é encontrada em quase todos os eucariontes e em procariontes, nestes últimos em formas mais simples denominadas sequências de inserção (IS) (WICKER et al., 2007).

Os transposons apresentam duas subclasses, que são distinguidas pelo número de filamentos de DNA cortados durante a transposição (; SHIETGAT et al., 2018; KAZAZIAN Jr, 2004). A subclasse I, compreende a ordem TIR (superfamílias TC1-Mariner, hAT, Mutator, Merlin, PiggyBac e CACTA), que tem como característica repetições terminais invertidas (TIR), e subclasse II apresentando as ordens Helitron (superfamília Helitron) e Maverick (superfamília Maverick) (SHIETGAT et al., 2018).

Os elementos da subclasse I (TIR) aumentam seus números de cópias transpondo de uma posição que já foi replicada dentro do cromossomo, para outra posição que a forquilha de replicação ainda não passou. Esse mecanismo é mediado pela enzima transposase, que reconhece as TIRs e corta ambos os filamentos em cada extremidade. Já os elementos da subclasse II são copiados, colados e se transpõem através de um mecanismo de replicação envolvendo o deslocamento de apenas uma cadeia de DNA (WICKER et al., 2007).

Durante a transposição, os elementos de classe II podem causar mutações afetando a expressão ou alterando a estrutura dos genes (KIDWELL, 2002) e embora a maioria dessas mutações seja prejudicial, algumas vezes os transposons podem contribuir para originar diversidade alélica para a criação de novos genes (GOGVADZE; BUZDIN, 2009).

2.4 EVOLUÇÃO DE SATÉLITES E SINAL FILOGENÉTICO

Como relatado anteriormente, os monômeros de satélites possuem uma alta taxa de diversificação de sequências (MACAS; MESZÁROS; NEUZOVÁ, 2002) associada a eventos de evolução combinada, fatos que resultam em sequências altamente homogêneas (MEHROTRA; GOYAL, 2014).

Essa homogeneidade como característica, associada à abundância dessas frações repetitivas, dão forte respaldo para a utilização de DNA satélites em estudos que visam estimar melhor as relações filogenéticas entre espécies e dessa forma corroborar ou explicar de forma mais congruente os resultados de estudos filogenéticos feitos com árvores gênicas, visto que essas árvores podem ter topologias conflitantes devido ao tempo de divergência entre as espécies, diferenças das histórias evolutivas dos alelos de diferentes espécies, entre outros fatores (DODSWORTH et al., 2015).

Edwards et al., (2002) descobriram que as repetições de sequências curtas fornecem assinaturas genômicas que podem ser usadas para inferir relações filogenéticas. Posteriormente, Dodsworth et al., (2015) usaram a abundância de diferentes frações repetitivas para resolver as relações evolutivas entre cinco grupos de angiospermas e tiveram resultados semelhantes aos encontrados usando cpDNA e regiões ribossômicas nucleares.

Partindo das premissas que espécies originadas de um ancestral comum compartilham boa parte das informações genéticas (HARVEY; PAGEL, 1991) e que espécies ecologicamente relacionadas tendem a ser mais semelhantes (LOSOS, 2008), inferências com conjuntos de dados de sequências múltiplas (DODSWORTH et al., 2015), também chamadas de atributos moleculares, morfológicos, fisiológicos e comportamentais vêm sendo realizadas (HARVEY; PAGEL, 1991).

Uma forma de fazer essas inferências evolutivas é através da medição do sinal filogenético, que é a tendência de as espécies relacionadas se assemelharem mais do que se assemelham as espécies desenhadas aleatoriamente na árvore filogenética (BLOMBERG; GARLAND, 2002).

Em resumo, se as duas espécies compartilham uma proporção maior de história comum, então, sob muitas circunstâncias, elas devem ser mais semelhantes que a média. Em contraste, se duas espécies compartilham pouca ou nenhuma história comum, então elas provavelmente serão menos similares que a média (REVELL; HARMON; COLLAR, 2008).

Asseverando essa proposição, Dodsworth et al. (2014) afirmam que quanto à estrutura e posição cromossômica, repetições em espécies intimamente relacionadas são quase idênticas, enquanto espécies mais distantes divergem em estrutura e localização de repetições à medida que a similaridade genética diminui.

Existem vários índices de medição de sinal filogenético disponíveis (MUNKEMULLER et al., 2012), como o K de Blomberg (BLOMBERG; GARLAND; IVES, 2003) e o λ (lambda) de Pagel (PAGEL, 1999), ambos quantitativos, dependentes de filogenia prévia e assumindo um modelo de Movimento Browniano (BM) de evolução de traços (MUNKEMULLER et al., 2012) sendo esta última estatística usada para inferir processos evolutivos neste estudo. Para o índice lambda, um valor próximo de zero (baixo sinal filogenético) indica que a característica evoluiu independentemente da filogenia (KAMILAR; COOPER, 2013), e um valor igual a um, ou próximo a ele (alto sinal filogenético), indica que as características das espécies estão distribuídas como esperado no BM, ou seja, com mudança de característica gradual, aleatória e não direcional ao longo do tempo (MUNKEMULLER et al., 2012).

O sinal filogenético tem implicações em uma ampla gama de áreas de pesquisas, como biogeográficas, estudos em biologia e conservação, filogenéticos, etc., contribuindo de forma considerável em inferências de processos evolutivos (DODSWORTH et al., 2015).

3 OBJETIVOS

3.1 Geral

Analisar a fração repetitiva e a abundância de DNA satélite em 77 espécies de *Asclepias*.

3.2 Específicos

- a) Construir uma filogenia de Máxima Verossimilhança de 77 espécies do gênero *Asclepias*;
- b) Identificar e quantificar as sequências de DNA satélite de quatro espécies dos principais clados de *Asclepias* (*A. syriaca*, *A. curassavica*, *A. coulteri* e *A. Fournieri*).
- c) Correlacionar a abundância de sequências de satélite e a topologia da árvore.

4 MATERIAL E MÉTODOS

4.1 Dados de Sequenciamento de Rendimento

As leituras únicas (*single-end-reads*) para 76 espécies de *Asclepias* e as leituras pareadas (*paired-end reads*), apenas para *Asclepias syriaca*, foram obtidas no banco de dados do NCBI (<https://www.ncbi.nlm.nih.gov/>) (Tabela 1). Os arquivos SRA foram descompactados no FASTQ usando o executável FASTQ-DUMP do SRA Toolkit. As leituras foram filtradas usando o BBDuk (implementado no *software* Geneious R9) para remover quaisquer artefatos e ajustar sua qualidade.

Tabela 1 - Dados de sequenciamento de rendimento das *Asclepias* utilizadas no estudo.

	Species	NCBI number		Species	NCBI number
1	<i>A. subulata</i> x <i>A. albicans</i>	SRR1612212	40	<i>A. solanoana</i>	SRR1612205
2	<i>A. albicans</i>	SRR1612109	41	<i>A. feayi</i>	SRR1612142
3	<i>A. meadii</i>	SRR1612167	42	<i>A. cinerea</i>	SRR1612126
4	<i>A. subulata</i>	SRR1612211	43	<i>A. glaucescens</i>	SRR1612228
5	<i>A. subaphylla</i>	SRR1612208	44	<i>A. hallii</i>	SRR1612107
6	<i>A. longifolia</i>	SRR1612160	45	<i>A. engelmanniana</i>	SRR1612140
7	<i>A. cutleri</i>	SRR1612135	46	<i>A. tomentosa</i>	SRR1612215
8	<i>A. coulteri</i>	SRR1612131	47	<i>A. connivens</i>	SRR1612128
9	<i>A. masoni</i>	SRR1612166	48	<i>A. circinalis</i>	SRR1612127
10	<i>A. macrotis</i>	SRR1612165	49	<i>A. eriocarpa</i>	SRR1612139
11	<i>A. subverticillata</i>	SRR1612231	50	<i>A. viridula</i>	SRR1612224
12	<i>A. woodsoniana</i>	SRR1612226	51	<i>A. humistrata</i>	SRR1612147
13	<i>A. angustolia</i>	SRR1612113	52	<i>A. oenotheroides</i>	SRR1612176
14	<i>A. curassavica</i>	SRR1612134	53	<i>A. nyctaginifolia</i>	SRR1612174
15	<i>A. aff.candida</i>	SRR1612106	54	<i>A. emoryi</i>	SRR1612137
16	<i>A. mexicana</i>	SRR1612169	55	<i>A. arenaria</i>	SRR1612115
17	<i>A. boliviensis</i>	SRR1612123	56	<i>A. próstata</i>	SRR1612189
18	<i>A. cordifolia</i>	SRR1612130	57	<i>A. welshii</i>	SRR1612225
19	<i>A. cryptoceras</i>	SRR1612133	58	<i>A. macrosperma</i>	SRR1612163
20	<i>A. scaposa</i>	SRR1612199	59	<i>A. involucrata</i>	SRR1612151
21	<i>A. vestita</i>	SRR1612222	60	<i>A. viridis</i>	SRR1612223
22	<i>A. californica</i>	SRR1612125	61	<i>A. asperula</i>	SRR1612116

23	<i>A. alticola</i>	SRR1612111	62	<i>A. sanjuanensis</i>	SRR1612197
24	<i>A. auriculata</i>	SRR1612121	63	<i>A. ruthiae-</i>	SRR1612195
25	<i>A. ovalifolia</i>	SRR1612181	64	<i>A. syriaca</i> x <i>A. speciosa</i>	SRR1612214
26	<i>A. otarioides</i>	SRR1612180	65	<i>A. syriaca</i>	SRR098429
27	<i>A. jorgeana</i>	SRR1612154	66	<i>A. tuberosa</i>	SRR1612217
28	<i>A. perennis</i>	SRR1612184	67	<i>A. quadrifolia</i>	SRR1612193
29	<i>A. similis</i>	SRR1612203	68	<i>A. verticillata</i>	SRR1612221
30	<i>A. Scheryi</i>	SRR1612201	69	<i>A. speciosa</i>	SRR1612207
31	<i>A. sp nov cf notha</i>	SRR1612173	70	<i>A. amplexicaulis</i>	SRR1612112
32	<i>A. notha</i>	SRR1612229	71	<i>A. latifolia</i>	SRR1612157
33	<i>A. pelúcida</i>	SRR1612183	72	<i>A. lynchiana</i>	SRR1612161
34	<i>A. lanceolata</i>	SRR1612155	73	<i>A. lanuginosa</i>	SRR1612156
35	<i>A. fournieri</i>	SRR1612143	74	<i>A. purpurascens</i>	SRR1612191
36	<i>A. brachystephana</i>	SRR1612124	75	<i>A. uncialis</i>	SRR1612219
37	<i>A. standleyi</i>	SRR1612206	76	<i>A. variegata</i>	SRR1612220
38	<i>A. atrovioleacea</i>	SRR1612120	77	<i>A. schaffner</i>	SRR1612200
39	<i>A. rosea</i>	SRR1612194			

Fonte: A autora (2019).

4.2 Análises filogenéticas

As leituras para todas as espécies foram mapeadas no *software* Bowtie (LANGMEAD et al., 2009) usando o genoma do cloroplasto de *Asclepias syriaca* como referência e os genomas de esboço (Draft Genome) foram obtidos usando Geneious R9 9.1 (KEARSE et al., 2012). A anotação do genoma foi realizada usando *A. syriaca* como referência pelo *software* Geneious R9. Para a anotação foi considerado um mínimo de 80% de identidade entre os genomas. As anotações foram verificadas individualmente e foram corrigidas manualmente para os códons de início e parada, se necessário. As anotações genéticas foram extraídas e as seqüências foram concatenadas para inferências filogenéticas.

As seqüências concatenadas foram alinhadas usando MAFFT v.7.017 (KATO; STANDLEY, 2013) implementado como a ferramenta “*Multiple align*” no *software* Geneious. A história evolutiva foi inferida utilizando o método da Máxima Verossimilhança (ML) baseado no modelo de substituição de nucleotídeos GTR + I + G (NEI; KUMAR, 2000), e a sustentação do ramo foi avaliada com 1.000 réplicas *bootstrap*. A determinação do modelo GTR e a análise ML foram realizadas no IQ-TREE (NGUYEN et al., 2015)

implementado no Geneious. A espécie *Vincetoxicum rossicum* (Asclepiadaceae) foi utilizada como grupo externo na árvore filogenética (MATTILA et al., 2003).

4.3 Agrupamento baseado em graficos das leituras de DNA

As leituras dos clados principais observados na árvore filogenética (ver Figura 1) foram ajustadas usando o BBDuk e foram usadas como entrada para agrupamentos baseados em gráficos comparativos com o RepeatExplorer (NOVÁK et al., 2010) e com o Tandem Repeat Analyzer (TAREAN) *software* (NOVÁK et al., 2017) para identificação de sequências repetitivas, com ambos os pacotes de software implementados no ambiente Galaxy (<http://repeatexplorer.umbr.cas.cz/>). A análise do RepeatExplorer permitiu detectar a proporção genômica do DNA repetitivo, enquanto o TAREAN é um canal computacional para a identificação de repetições a partir das leituras de sequência desmontadas. As espécies utilizadas no agrupamento gráfico foram: *A. coulteri*, *A. founieri*, *A. curassavica* e *A. syriaca* (Tabela 1). As seqüências de satDNA identificadas foram utilizadas para a detecção da abundância destas sequências nas 77 espécies usando uma abordagem de leitura de mapeamento no *software* Geneious. A correlação entre a proporção de famílias de satDNA e a filogenia foi analisada pela estatística lambda implementada na função comMap no pacote R de phytools (REVELL, 2012).

5 RESULTADOS

5.1 Árvore filogenética

A árvore filogenética foi inferida usando regiões anotadas concatenadas do genoma do cloroplasto, incluindo os genes codificadores de proteínas, RNAt e rRNAs. A árvore filogenética mostrou quatro clados principais com alto suporte (Figura 3). O primeiro clado (A) corresponde à primeira linhagem divergente, correspondente ao clado do Deserto de Sonora, enquanto o segundo clado (B) foi formado pela espécie *Incarinatae*; o clado (C) era formado pelas espécies norte-americanas temperadas e o clado (D) era composto por espécies do grupo de espécies das terras altas mexicanas. A árvore filogenética mostrou alto suporte de ramos, sugerindo forte suporte e uma história mais clara das *Asclepias*.

Figura 3 - Filogenia da Máxima Verossimilhança (ML) utilizando genes de cloroplastos concatenados e suporte estimado por bootstraps (em porcentagens). As cores e letras correspondem aos principais clados.



Fonte: A autora (2019).

5.2 Composição de Repetições

As análises foram realizadas usando *single-end-reads* com comprimentos de 76 nt disponíveis no banco de dados do NCBI. O seqüenciamento de DNA de alto rendimento foi realizado em quatro espécies que pertencem aos principais clados de *Asclepias* usando *clustering* baseado em gráfico. A fração repetitiva correspondeu a 38,758% em *A. coulteri*, 40,66% em *A. founieri*, 31,871% em *A. curassavica* e 45,596% em *A. syriaca* (Tabela 2). A repetição da composição das espécies mostrou que os elementos transportáveis (TE) foram mais abundantes em todos os genomas, dentre os quais os retrotransposons de LTR, incluindo ty1 / copia e Ty3 / gypsy, representaram a maior fração entre todos os genomas analisados. Entre as famílias de retrotransposons de LTR, as Bianca, Athila e Ogre / Tat foram as mais abundantes no gênero. Para os transposons, as espécies *A. coulteri*, *A. founieri* e *A. curassavica* apresentaram baixa abundância e não foram observados transposons em *A. syriaca* (Tabela 2).

Tabela 2 – Proporções da fração repetitiva em quatro genomas de *Asclepias*.

Class	Order	Superfamily	Family	Genome [%]				
				<i>A. coulteri</i>	<i>A. fournieri</i>	<i>A. curassavica</i>	<i>A. syriaca</i>	
Classified TEs- Retrotransposons	LTR	Ty1/copia	Ale	1.276	0.283	0.952	0.069	
			Angela	-	0.727	-	-	
			keros	0.246	0.41	0.262	-	
			Bianca	2.935	2.044	1.182	2.409	
			Ivana/Oryc o	-	0.305	1.409	0.222	
			SIRE	0.185	0.1698	-	-	
			Tork	1.389	0.916	0,589	0.508	
			Alesia	-	0.038	-	-	
			Total	6.031	4.8929	4.394	3.208	
		Ty3/gypsy	Athila	5.17	7.328	2.618	7.762	
			Chromoviru s	1.824	0.45	1.31	0.586	
			Ogre/Tat	1.57	1.084	1.765	0.489	
		Total	8.564	8.862	5.693	8.837		
		Non-LTR (LINE)			1.024	-	0.51	-
		Pararetrovirus			0.674	0.067	0.474	-
Classified TEs-Transposons	TIR	hAT	-	-	-	-		
		Mutator	0.027	0.011	0.034	-		
		MULE	-	-	-	-		
		CACTA	-	0.16	-	-		
Total for TEs			16.32	13.83	11.11	12,05		
rDNA			0.688	1.992	2.68	0.695		
Classified repeats			Satellites	0.985	.462	7.732	3.38	
Unidentified			20.765	20.2132	10.354	29.476		
Total Repeat			38.758	40.6601	31.871	45.596		

Fonte: A autora (2019).

5.3 Sequências de satDNA

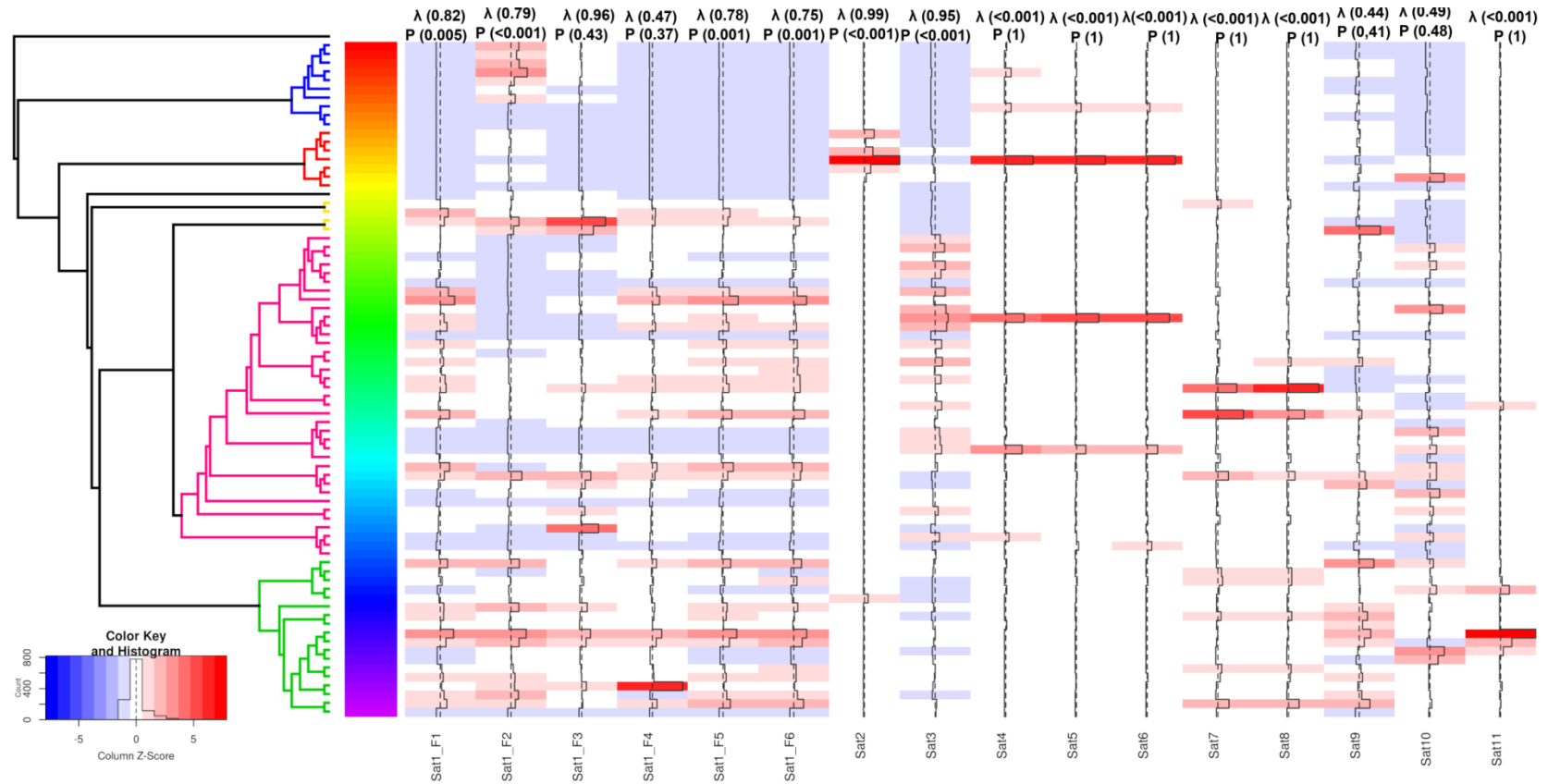
A fração repetitiva que correspondeu ao DNA satélite variou de 0,985% (*A. coulteri*) a 7,732% (*A. curassavica*), indicando uma proporção considerável de satDNA. O rDNA se mostrou moderadamente repetitivo, variando de 0,688% (*A. coulteri*) a 2,68% (*A. curassavica*) (Tabela 2). Notavelmente, houve muitas sequências de satélites observadas, entre as quais 11 satélites foram encontrados nas quatro espécies analisadas. O comprimento dos monômeros apresentou uma variação moderada, como Sat1 com 166-170 pb (detectado como variação entre as espécies), Sat2 com 92 pb, Sat3 com 73 pb, satélites Sat4, Sat7 e Sat8 com 42 pb, Sat5 com 171 pb, Sat6 com 26 pb, Sat9 com 56 pb, Sat10 com 166 pb e Sat11 com 91 pb (Tabela 3). O conteúdo de CG variou de 19 a 45,1% e não foram detectados satélites ricos em CG. A análise das famílias de satDNA mostrou que Sat1 tinha seis famílias em *A. coulteri*, *A. curassavica*, *A. coulteri* e *A. curassavica* (Figura 4). Entre os satélites, Sat1 foi observado em todas as espécies e foi o mais abundante em todas as espécies em relação aos outros satélites. A correlação entre a abundância de satélites de DNA e a filogenia de *Asclepias* foi significativa para quatro famílias Sat1 ($\lambda = 0,75-0,82$, $p < 0,005$), Sat2 ($\lambda = 0,99$, $p < 0,001$) e Sat3 ($\lambda = 0,95$, $p < 0,001$) (Figura 2). Notavelmente, Sat2 foi mais abundante no clado B (espécies *Incarinatae*), enquanto Sat3 foi mais abundante no Clade C (espécies temperadas da América do Norte) (Figura 4).

Tabela 3 – Sequências de monômeros de DNA satélite em *Asclepias*.

Names	Sequence	Size (pb)	CG (%)
SAT1	TCAACATGTAAGGCACAAAAGTAGTTAAATCACATTACAATGAA TCTAATATAACGAATGTAACGCTTCAAAGTTCATCAAATTTAGGT CAAATTTGGTTTGAATTTCAACGCAAGTAATGCACTTGAATCTTG AAATAAGAAGTTATTTGGTTACTTTT	166-170	27.6
SAT2	TTCATGATAGAAAGTATAATATAGAGCCACAACAATGGTAACTCT ACTCGGTATTTTGAAAGAGTTGGTAAACAACTCCTAAATATTTT TT	92	29.3
SAT3	ATCACGTTTGAAAATTATGTGAAACATTAGGACTCATCTTCTCAA TGAGTCCTATAATTCACTACTGGAATTT	73	31.5
SAT4	ATAATATATATTTGGTGTACTTTGATATTTTATGTACAGTAT	42	19.0
SAT5	TTTTGTAGAATCTGCAAGTGGATATTTGGACCGCTTTGAGGCCTT CGTTGGAAACGGGAATATCTTCCATAAAAACTAGACAGAAGCA TTCTCAGAACTTCTTTGTGATGTGTGCATTCAACTCACAGAGTT GAACCTTTCTTTTCATAGAGCAGTTTTGAAACACTCT	171	39.2
SAT6	TCCATTCGATTCCATTTCGATGATGAT	26	38.5
SAT7	CTTCTGAGGGTTTCTGAACTTTCCTTCAACAGCTATGAATTT	42	38.1
SAT8	TGTTGAAGGAAAGTCCAGAAACCCTCACAAGAAATTCATAGC	42	40.5
SAT9	ATAATTCCAACATACATGGAGTATCCAATCCAACAAAAACACA ATGTGATAAAAA	56	28.6
SAT10	TTTAAATTACTAAAAAAGTGTAATTCATCAGTAACAAATTTATA TCAAGTGTGAAAAGTCTTCAATTCAAATATTCCTAAATTTTAAAG TGGAAGTTATAATTCTGTGTTTGTAATAAGCAAAAAAACACTT CTTCCCCGCCCTTATATAAAGGTAAACATT	166	24.7
SAT11	GTCATCTCCATAGGTCCGTATCATCGTTACAGATACGATCGGCAT CAAGATAACAAAATGAAAGCTATGATGGGAGAGAACGGAGCCA GAG	91	45.1

Fonte: A autora (2019)

Figura 4 - Mapa de calor mostrando a abundância de seqüências de satDNA nas 77 espécies de *Asclepias*. As cores correspondem à abundância onde o vermelho representa o mais abundante. O topo mostra a estatística lambda e o p valor .



Fonte: A autora (2019).

6 DISCUSSÃO

6.1 Árvore filogenética

No presente estudo, construímos uma árvore filogenética para auxiliar na compreensão da evolução de sequências de DNA satélite no gênero *Asclepias*. Nossa árvore filogenética mostrou que os clados eram fortemente apoiados e suas topologias eram semelhantes aos resultados de estudos anteriores (FISHBEIN et al., 2011; FISHBEIN et al., 2018). Estas diferentes inferências filogenéticas são importantes para esclarecer as relações entre os táxons. Dessa forma, o presente trabalho contribui para o entendimento da evolução da *Asclepias* e fornece forte suporte para os principais clados descritos por Fishbein et al. (2018)

6.2 Sequências de DNA repetitivo

As sequências de DNA repetitivo constituem uma fração importante dos genomas de plantas e são ferramentas importantes para entender sua dinâmica evolutiva, organização, distribuição e as relações entre espécies congêneres (MEHROTRA; GOYAL, 2014). As sequências de DNA repetitivas mostraram alta variação de tamanho entre as plantas, como *Arabidopsis thaliana*, que possui 20% de seu genoma composto por sequências repetitivas de DNA (INITIATIVE et al., 2000), enquanto em *Helianthus annuus*, sequências repetitivas de DNA respondem por 80% de seu genoma (GILL et al., 2014). No entanto, a abundância e diversidade de sequências de DNA repetitivas em relações próximas entre espécies não foram relatadas e esta abordagem pode contribuir para a nossa compreensão da evolução destes elementos repetitivos em plantas. Entre as classes de sequências repetitivas de DNA, os estudos de DNA satélites explodiram após o desenvolvimento de novas abordagens de sequenciamento, porque a abordagem de Sequenciamento de Nova Geração (NGS) pode ser usada na identificação de sequências repetidas baseada em agrupamento de sequências moderadas a altamente repetidas. Desta forma, o gênero *Asclepias* é adequado para entender a evolução dinâmica do DNA satélite, pois existem dados NGS disponíveis no NCBI para muitas espécies.

As estimativas do tamanho do genoma nuclear das espécies variam entre 366 a 548 Mb nas quatro espécies (BAI et al., 2012), os quais podem ser considerados pequenos; no entanto, a fração repetitiva apresentou abundância moderada de sequências repetitivas de DNA, dentre as quais os elementos do transposon foram os mais abundantes em todas as

quatro espécies. Transposons foram relatados como as sequências repetitivas mais abundantes na maioria dos genomas de plantas (RABINOWICZ; BENNETZEN, 2006). As sequências de DNA satélite mostraram considerável abundância entre as espécies, sendo uma parte importante da evolução do genoma de *Asclepias*. Notavelmente, o gênero *Asclepias* é caracterizado por suas muitas sequências de DNA satélite, entre as quais pelo menos 11 sequências de DNA satélite foram observadas, sugerindo que esta classe de sequências repetitivas de DNA mostra uma alta diversidade. Em outras espécies, como *Vicia faba* (Fabaceae), que possui uma estimativa do tamanho do genoma de 13,41 Gb, 31 sequências de DNA satélite foram descritas (MACAS et al., 2015), correspondendo a 7% do tamanho do genoma, enquanto no gênero *Spondias* (Anacardiaceae), que tem uma estimativa do tamanho do genoma de 0,5 Gb, existem apenas quatro sequências de DNA satélite (MARTINS et al., 2019 em publicação).

6.3 Sequências de DNA de satélite

As sequências de DNA satélites mostraram uma alta diversidade no número de satDNAs e no comprimento dos monômeros, sugerindo que essa classe de DNA repetitivo tem uma evolução e diversificação altamente dinâmica em *Asclepias*. Algumas sequências de DNA satélite mostraram evolução de acordo com a filogenia, como Sat2, que é característico do clado Incarnatae, que inclui espécies da América do Sul, sugerindo que a evolução do satélite está correlacionada com a distribuição geográfica das espécies. Por outro lado, algumas sequências de DNA satélite eram estáveis no gênero e estavam distribuídas na maioria dos clados. Notavelmente, o clado Incarnatae, que mostrou uma pequena abundância de Sat1, sugeriu que houve uma substituição de Sat1 por Sat2. A evolução diferencial das sequências de DNA satélite foi observada em outros táxons, entre os quais as sequências podem ser espécies específicas ou específicas do gênero e podem ser altamente conservadas dentro de gêneros e famílias (ANAMTHAWAT-JÓNSSON; HESLOP-HARRISON, 1992).

7 CONCLUSÕES

Nós concluimos: (1) *Asclepias* é caracterizada por muitos satDNAs, sugerindo que o satDNA contribui para a alta abundância de DNA repetitivo no genoma de *Asclepias*; (2) a abundância dos mesmos satDNAs está correlacionada com sua filogenia; e (3) as sequências de satDNAs podem ser usadas como características adicionais para o estudo da evolução de *Asclepias*.

REFERÊNCIAS

- ANAMTHAWAT-JONSSON, K.; HESLOP-HARRISON, J. S. Species specific DNA sequences in the Triticeae. **Hereditas**, v. 116, n.1, p.49-54, 1992. DOI: 10.1111/j.1601-5223.1992.tb00204.x.
- ARCHIBALD, J. M. The puzzle of plastid evolution. **Curr. Biol**, v.19, n.2, p.81-88, 2009. DOI: 10.1016/j.cub.2008.11.067.
- Asclepias* in **Flora do Brasil 2020 em construção**. Jardim Botânico do Rio de Janeiro. Disponível em: <http://reflora.jbrj.gov.br/reflora/floradobrasil/FB4519>. Acesso em: 14 fev. 2019.
- BAI, C. et al. New reports of nuclear DNA content for 407 vascular plant taxa from the United States. **Ann. Bot.**, v.110, n.8, p. 1623-1629, 2012. DOI:10.1093/aob/mcs222.
- BLOMBERG, S. P.; GARLAND, T.; IVES, A. R. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. **Evolution**, v.57, n.4, p.717-745, 2003. DOI: 10.1111/j.0014-3820.2003.tb00285.x.
- BOWEN, N. J.; JORDAN, I. K. Transposable elements and the evolution of eukaryotic complexity. **Current Issues in Molecular Biology**, v.4, p.65-76, 2002. DOI:: 10.1023/A:1016072014259.
- CASACUBERTA, J. M.; SANTIAGO, N. Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. **Gene**, v.311, p.1-11, 2003. DOI: 10.1016/S0378-1119(03)00557-2.
- DODSWORTH, S. et al. Genomic repeat abundances contain phylogenetic signal. **Syst. Biol.**, v.64, n.1, p.112-126, 2015. DOI:10.1093/sysbio/syu080.
- EDWARDS, S. V. et al. A genomic schism in birds revealed by phylogenetic analysis of DNA strings. **Syst. Biol.** v. 51, n. p.599-613, 2002. DOI: 10.1080/10635150290102285.
- EICKBUSH, T.H.; EICKBUSH, D.G. Finely orchestrated movements: Evolution of the ribosomal RNA genes. **Genetics**, v.175, p.477-485, 2007. DOI: 10.1534/genetics.107.071399.
- ELLEGREN, H. Microsatellites: simple sequence with complex evolution. **Nature**, v.5, n.1, p.435-445, 2004. DOI: 10.1038/nrg1348.
- EMBLEY, T. M.; MARTIN, W. Eukaryotic evolution changes and challenges. **Nature**, v.440, n.7084, p.623-630, 2006. DOI: 10.1038/nature04546.
- FARIA, R.; NAVARRO, A. Chromosomal speciation revisited: rearranging theory with pieces of evidence. **Trends in Ecology & Evolution**, v.25, n.11, p.660-669, 2010. DOI:10.1016/j.tree.2010.07.008
- FELICIELLO, I.; AKRAP, I.; UGARKOVIC, D. Satellite DNA modulates gene expression in the beetle *Tribolium castaneum* after heat stress. **PLOS Genetics**, v.11, n.9, p.1-18, 2015. DOI: doi.org/10.1371/journal.pgen.1005466.

- FIGUEROA-MATINEZ, F. et al. The plastid genome of *Polytoma uvella* is the largest known among colorless algae and plants and reflects contrasting evolutionary paths to nonphotosynthetic lifestyles. **Plant physiology**, v.173, n.2, p.932-943. 2017. DOI: 10.1104 / pp.16.01628.
- FISHBEIN, M. et al. Evolution at the tips: *Asclepias* phylogenomics and new perspectives on leaf surfaces. **American Journal of Botany**, v.105, n.3, p.514-524, 2018. DOI: 10.1002/ajb2.1062.
- FISHBEIN, M. et al. Phylogenetic relationships of *Asclepias* (Apocinaceae) inferred from non-coding chloroplast DNA sequences. **Systematic Botany**, v. 36, n.4, p.1008-1023, 2011.
- FONTELLA-PEREIRA, J.; KONNO, T. U. P. *Asclepias* (Asclepiadaceae). In: WANDERLEY, M.G.L.; SHEPHERD, G. J.; MELHEM, T.S.; GIULIETTI, E A.M. (Eds.). **Flora Fanerogâmica do Estado de São Paulo**, v. 4. São Paulo: FAPESP & Rima, 2005. p. 98-100.
- GILL, N. et al. Sequence-based analysis of structural organization and composition of the cultivated sunflower (*Helianthus annuus* L.) genome. **Biology** v.3, n. 2, p. 295-319, 2014. DOI: 10.3390/biology3020295.
- GOGVADZE E.; BUZDIN, E, A. Retroelements and their impact on genome evolution and functioning. **Cellular and Molecular Life Sciences**, v.66, p.3727–3742, 2009. DOI: 10.1007/s00018-009-0107-2.
- GOODIER, J. L.; KAZAZIAN, H. H. JR. Retrotransposons revisited: the restraint and rehabilitation of parasites. **Elsevier Inc**, v. 135, n.3, p.23–35, 2008. DOI: 10.1016/j.cell.2008.09.022.
- GREEN, B. R. The plant genome: an evolutionary view on structure and function: chloroplast genomes of photosynthetic eukaryotes. **The Plant Journal**, v.66, n.1, p. 34–44, 2011. DOI: 10.1111/j.1365-313X.2011.04541.x.
- HARVEY, P. H; PAGEL, M. D. **The comparative method in evolutionary biology**. Oxford, UK: Oxford University Press, 1991.
- HAWBOLD, B; WIEHE, T. How repetitive are genomes? **BMC Bioinformatics**. v.7 n. 1, p. 1-10 541, 2006. DOI: 10.1186/s13015-017-0118-8.
- HILLIS, D. M.; DIXON, M. T. Ribosomal DNA: Molecular evolution and phylogenetic inference. **The Quarterly Review of Biology**, v.66, p.411-453, 1991. DOI: 10.1086/417338.
- INITIATIVE, A. G. et al. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. **Nature**, v.408, n.1, p796-815, 2000.
- JANSEN, R. K.; RUHLMAN, T. A. Plastid genomes of seed plants. **Genomics of Chloroplast and Mitochondria**, v.35, n.1. p. 103-126, 2012. DOI: 10.1007/978-94-007-2920-9_5.

KAMILAR, J. M.; COOPER, N. Phylogenetic signal in primate behavior ecology and life history. **Philosophical Transactions of the Royal Society B**, v.368, 2012341, 2013. DOI: 10.1098/rstb.2012.0341.

KAPITONOV, V. V.; JURKA, J. A universal classification of eukaryotic transposable elements implemented in Repbase. **Nature**, v.9, n.5, p.411-412, 2008. DOI: 10.1038/nrg2165-c1.

KATOH, K.; STANDLEY, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. **Mol. Biol. Evol.**, v.30, n.4, p. 772-780, 2013. DOI: 10.1093/molbev/mst010.

KAZAZIAN Jr., H. H. Mobile elements: drivers of genome evolution. **Science**, v.303, p.1-8, 2004. DOI: 10.1126/science.1089670.

KEARSE, M. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. **Bioinformatics**, v.28, n.12, p. 1648-1649, 2012. DOI: 10.1093/bioinformatics/bts199.

KEJNOVSKY, E.; HAWKINS, J. S.; FESCHOTTE, C. Plant Transposable Elements: Biology and Evolution. **Plant Genome Diversity**, v.1, p.17-34, 2012. DOI: 10.1007/978-3-7091-1130-7_2.

KIDWELL, M G. Transposable elements and the evolution of genome size in eukaryotes. **Genetica**, Arizona, USA. v.115, n.1, p.49–63, 2002.

KOLODNER, R.; TEWARI, K. Inverted repeats in chloroplast DNA from higher plants. **Proc. Natl. Acad. Sci.**, USA, v. 76, p.41-45, 1979.

KOLODNER, R.; TEWARI, K. Molecular size and conformation of chloroplast deoxyribonucleic acid from pea leaves. **J. Biol. Chem.** v.247,n.19, p.6355- 6364, 1972.

LANGMEAD, B. et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. **Genome Biology**, v.10, n.3, p.1-10, 2009. DOI: 10.1186/gb-2009-10-3-r25.

LONG, E. O.; DAWID, I. B. Repeated genes in eukaryotes. **Annu Rev Biochem**, v. 49, p. 727 – 764, 1980. DOI: 10.1146/annurev.bi.49.070180.003455.

LOSOS, J. B. Phylogenetic niche conservatism, phylogenetic signal and the relationships between phylogenetic relatedness and ecological similarity among species. **Ecology Letters**, v. 11, n. 10, p. 995-1003, 2008. DOI: 0.1111/j.1461-0248.2008.01229.x.

LEISTER, D. Chloroplast research in the genomic age, **Trends Genet**, v.19, n.1, p.47-56, 2003. DOI: 10.1016/S0168-9525(02)00003-3.

MACAS, J. et al. In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe Fabaeae. **Plos one**, v.10, n.1, p. 1-23, 2015. DOI: 10.1371/journal.pone.0143424.

MACAS, J.; MÉSZÁROS, T.; NOUZOVÁ, M. Plant-sat: a specialized database for plant satellite repeats. **Bioinformatics**, v.18, n.1, p.28-35, 2002. DOI: 10.1093/bioinformatics/18.1.28.

MAIER, R. M. et al. Complete sequence of the maize chloroplast genome: gene content , hotspots of divergence and fine tuning of genetic information by transcript editing. **J. Mol. Biol.**, v. 251, n.5, p.614-628, 1995. DOI: 10.1006/jmbi.1995.0460.

MALCOLM, S. B. **Herbivore:** theyer interactions with secondary plant metabolites. Cardenolide-mediated interactions between plants and herbivores. 2. ed. California: Academic Press Inc, 1991.

MARTIN, A. et al. A transposon-induced epigenetic change leads to sex determination in melon. **Nature**, v.461, n.7267, p. 1135-1138, 2009. DOI: 0.1038/nature08498.

MARTINS, G. S. M. et al. Genome size and cytomolecular analysis revealed a high similarity of satellite DNA among Spondias species (Anacardiaceae), 2019. (Não publicado).

MATTILA, H. R.; OTIS, G. W. A comparison of the host preference of monarch butterflies (*Danaus plexippus*) for milkweed (*Asclepias syriaca*) over dog-strangler vine (*Vincetoxicum rossicum*). **Entomoliga Experimentalis at applicate**, v.107, n.3. p.193-199. DOI: 10.1046/j.1570-7458.2003.00049.x

MEHROTRA, S. et al. Significance of satellite DNA revealed by conservation of a widespread repeat DNA sequence among angiosperms. **Applied Biochemistry and Biotechnology**, v.173, n.7, p.1790-1801, 2014. DOI: 10.1007/s12010-014-0966-3.

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. DOI: 10.1016/j.gpb.2014.07.003.

MUNKEMULLER, T. et al. How to measure and test phylogenetic signal. **Methods in Ecology and Evolution**, v. 2, n.4, p.743-756, 2012. DOI: 10.1111/j.2041-210X.2012.00196.x.

NEI, M. KUMAR, S. **Molecular Evolution and Phylogenetics**. New York: Oxford University Press, 2000.

NGUYEN et al. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. **Molecular Biology and Evolution**, v.32, n. 1, p.268-274, 2015. DOI: 10.1093/molbev/msu300.

NOVÁK et al. TAREAN: a computacional toll for identification and characterization of satellite DNA from unassembled short reads. **Nucleic Acids Research**, v.45, n.12, p.1-10, 2017. DOI: 10.1093/nar/gkx257.

NOVÁK et al. Graph-based clustering and characterization repetitive sequences in next-generation sequencing data. **BMC Bioinformatics**, v.11, n.378, p.1-12, 2010. DOI: 10.1186/1471-2105-11-378.

OVALLE, F. R. **Análise da distribuição dos sítios de DNA ribossomal 5S e 45S em cariótipos de espécies vegetais.** Tese (Doutorado em Biologia Vegetal) - Universidade Federal de Pernambuco, Centro de Ciências Biológicas, Recife, 2011.

PAGEL, M. Inferring the historical patterns of biological evolution. **Nature**, v. 401, n.6756, p.877-884, 1999. DOI: 10.1038/44766.

PELLICER, J.; FAY, M. F.; LEITCH, I. F. The largest eukaryotic genome of them all? **Botanical journal of the Linnean society**, v.164, n.1 p. 10-15, 2010. DOI: [org/10.1111/j.1095-8339.2010.01072.x](https://doi.org/10.1111/j.1095-8339.2010.01072.x).

PELLISSIER, L. et al. Different rates of defense evolution and niche preferences in clonal and nonclonal milkweeds (*Asclepias* spp.). **New Phytologist**, v. 209, n.3, p.1230-1239, 2016. DOI: 10.1111/nph.13649.

PISUPATI, R.; VERGARA, D.; KANE, N. C. Diversity and evolution of the repetitive genomic content in *Cannabis sativa*. **BMC Genomics**, v.19, n.1, p. 1-9, 2018. DOI: 10.1186/s12864-018-4494-3.

RABINOWICZ, P. D.; BENNETZEN, J. L. The maize genome as a model for efficient sequence analysis of large plant genomes. **Curr. Opin. Plant. Biol.**, v.9, n.2, p.149-156, 2006. DOI: 10.1016/j.pbi.2006.01.015.

RAVI, V. et al. An update on chloroplast genomes. **Pl Syst Evol**, v.271, n.1-2, p.101-122, 2008. DOI: 10.1007/s00606-007-0608-0.

REVELL, L. J. Phytools: an R package for phylogenetic comparative biology (and other things). **Ecol. Evol.**, v.6, n.9, p.2774-2780, 2016. DOI: 10.1002/ece3.2051.

REVELL, L. J.; HARMON, L.J.; COLLAR, D. C. Phylogenetic signal, evolutionary, and rate. **Syst. Biol.**, v.57, n.4, p.591-601, 2008. DOI: 10.1080/10635150802302427.

ROBLEDILLO, L. A. et al. Satellite DNA in *Vicia faba* is characterized by remarkable diversity in its sequence composition, association with centromeres, and replication timing. **Nature**, v. 8, 5838, p.1-11, 2018. DOI: 10.1038/s41598-018-24196-3.

SCARCELLI, N. et al. A Set of 100 Chloroplast DNA Primer Pairs to Study Population Genetics and Phylogeny in Monocotyledons. **PLoS ONE**, v.6, n.5, p.e19954, 2011. DOI: 10.1371/journal.pone.0019954.

SCHIETGAT, L. et al. A machine learning based framework to identify and classify long terminal repeat retrotransposons. **PLOS Comput Biol**, v.14, n.4, p.1-21, 2018. DOI: 10.1371/journal.pcbi.1006097.

SCHIFINO-WITTMANN, M. T. Nuclear DNA content determination in plants. **Ciência Rural**, v.31, n.5, p.89-902, 2001. DOI: 10.1590/S0103-84782001000500028.

STADLER, M. et al. Distribution of novel and known repeated elements of *Solanum* and application for the identification of somatic hybrids among *Solanum* species. **Theor Appl Genet**, v. 91, n.8, p.1271-1278, 1995. DOI: 10.1007/BF00220940.

- STRAUB, S. C. K. et al. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. **American Journal of Botany**, v.99, n.2, p.349-364, 2012. DOI: 10.3732/ajb.1100335
- TIMMIS, J. N. et al. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. **Nat Rev Genet**, v.5, n.1, p.123–135, 2004. DOI: 10.1038/nrg1271.
- TOKARNIA, C. H. BRITO, M. F.; CUNHA, B. R. M. Intoxicação experimental por *Asclepias curassavica* (Asclepiadaceae) em bovinos. Dados complementares. **Pesq. Vet. Bras**, v.21, n. 1, p.1-4, 2001. DOI: 10.1590/S0100-736X2001000100002 .
- TOKARNIA, C. H; DOBEREINER, J; PEIXOTO, P. V. **Plantas Tóxicas do Brasil** , Rio de Janeiro: Helianthus, 2000.
- WICKER, T. et al. A unified classification system for eukaryotic transposable elements. **Nature Reviews Genetics**, v.8, p.973-982, 2007. DOI: 10.1126/science.1089670.
- WOODSON, R. E. Jr. The North American species of *Asclepias* L. **Missouri Botanical Garden**, v.41, n.1, p. 1-211, 1954. DOI: 10.2307/2394652
- WYATT, R. BROYLES, S. B. Ecology and evolution of reproduction in milkweeds. **Annu. Rev. Ecol. Syst.**, v.25, n.1 p. 423-441, 1994. DOI: 10.1146/annurev.es.25.110194.002231.
- ZHANG, Q.; SODMERGEN. Why does biparental plastid inheritance revive in angiosperms? **J Plant Res**, v.123, n.2, p.201-206, 2010. DOI: 10.1007/s10265-009-0291-z.
- ZHOU, L. et al. Transposition of hAT elements links transposable elements and V(D)J recombination. **Nature**, v.432, n.7020, p.995-1001, 2004. DOI: 10.1038/nature03157.