

UNIVERSIDADE FEDERAL DE ALAGOAS - UFAL
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS GRADUAÇÃO EM AGRICULTURA E AMBIENTE

BRUNA MARIZA DA SILVA SANTOS

**GENOMA DE CLOROPLASTO DE *PISTACIA VERA* L. (ANACARDIACEAE):
ORGANIZAÇÃO ESTRUTURAL E SINAIS FILOGENETICOS EM SAPINDALES**

ARAPIRACA

2019

Bruna Mariza da Silva Santos

Genoma de cloroplasto de *Pistacia vera* L. (Anacardiaceae): organização estrutural e sinais filogenéticos em Sapindales

Dissertação apresentada ao Programa de Pós- Graduação em Agricultura e Ambiente da Universidade Federal de Alagoas, como requisito para obtenção do título de Mestre em Agricultura e Ambiente.

Orientador: Prof. Dr. Cícero Carlos de Souza Almeida

Arapiraca

2019

Catálogo na fonte
Universidade Federal de Alagoas – UFAL
Biblioteca *Campus* de Arapiraca - BCA

Divisão de Tratamento Técnico
Bibliotecária: Gerlane Costa Silva de Farias
CRB - 4/1802

S237

Santos, Bruna Mariza da Silva

Genoma de cloroplasto de *Pistacia vera* L. (Anacardiaceae): organização estrutural e sinais filogenéticos em Sapindales. / Bruna Mariza da Silva Santos. – 2019.

16 f.: il.

Orientador: Cícero Carlos de Souza Almeida

Dissertação Artigo científico (Mestrado) – Universidade Federal de Alagoas. Pós-Graduação em Agricultura e Ambiente, Arapiraca, 2019.

Bibliografia; f. 11-14

1. Plastoma. 2. Filogenia 3. Sapindaceae I. Título

CDU 575.113

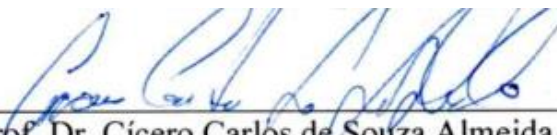
BRUNA MARIZA DA SILVA SANTOS

Genoma de cloroplasto de *Pistacia vera* L. (Anacardiaceae): Organização estrutural e sinais filogenéticos em Sapindales

Dissertação apresentada ao Programa de Pós-Graduação em Agricultura e Ambiente da Universidade Federal de Alagoas, como requisito para obtenção do título de Mestre em Agricultura e Ambiente

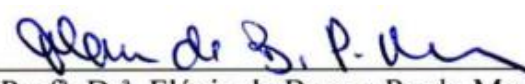
Data de aprovação: 26/02/2019

Banca Examinadora:



Prof. Dr. Cícero Carlos de Souza Almeida
Universidade Federal de Alagoas UFAL- *Campus* Arapiraca

Orientador



Prof.^ª. Dr.^ª. Flávia de Barros Prado Moura
Universidade Federal de Alagoas- UFAL

Examinador I



Prof. Dr. André Seco Marques da Silva
Universidade Federal de Alagoas UFAL- *Campus* Arapiraca

Examinador II

O Senhor é minha luz e minha salvação, a quem temerei?
O Senhor é o protetor de minha vida, de quem terei medo?

Salmos 27:1

À Deus, minha família e Luíz Neto.
Dedico

GENOMA DE CLOROPLASTO DE *Pistacia vera* L. (Anacardiaceae): ORGANIZAÇÃO ESTRUTURAL E SINAIS FILOGENETICOS EM SAPINDALES

THE CHLOROPLAST GENOME OF *Pistacia vera* L. (Anacardiaceae): STRUCTURAL ORGANIZATION AND PHYLOGENOMIC SIGNALS IN SAPINDALES

Bruna Mariza da Silva Santos¹, Cícero Carlos de Souza Almeida²

RESUMO: A ordem de Sapindales pertence ao clado Malvids e está próxima das ordens Huerteales, Malvales e Brassicales. A ordem é monofilética e possui nove famílias: Anacardiaceae, Burseraceae, Meliaceae, Rutaceae, Sapindaceae, Simaroubaceae, Biebersteiniaceae, Kirkiaceae e Nitrariaceae. Com exceção de Sapindaceae, Kirkiaceae, Biebersteiniaceae e Nitrariaceae, a filogenia da ordem é clara para a maioria das famílias. Desta forma, os genomas dos cloroplastos são altamente informativos e amplamente utilizados para estudos filogenéticos; no entanto, na ordem de Sapindales, a estrutura dos genomas dos cloroplastos não foi analisada. O objetivo deste estudo foi montar o genoma do cloroplasto de *Pistacia vera* L., construir a filogenia para a ordem de Sapindales usando genomas completos de cloroplasto e analisar os sinais filogenéticos com os conteúdos de LSC, SSC, IRA e CG. O genoma da *P. vera* foi montado usando leituras do NCBI e outros genomas completos foram baixados do NCBI no formato FASTA. O genoma do cloroplasto de *P. vera* mostrou 160.674 pb, dos quais 126 genes, oito rRNA e 37 tRNA foram identificados. A filogenia usando genomas completos apresentou as mesmas topologias com estudos prévios e a evolução do cloroplasto apresentou alta correlação com clados filogenéticos para o comprimento LSC e o conteúdo de CG, dos quais Sapindaceae é caracterizada por pequenos tamanhos de LSC e Rutaceae por alto CG. Conclui-se que Sapindales apresentou uma filogenia robusta e os genomas dos cloroplastos mostram sinais filogenéticos para o tamanho de LSC e o conteúdo de CG.

Palavras-chave: Plastomas. Filogenia. Sapindaceae

ABSTRACT: The Sapindales order belongs to Malvids clade and is close to orders Huerteales, Malvales, and Brassicales. The order is monophyletic and has nine families: Anacardiaceae, Burseraceae, Meliaceae, Rutaceae, Sapindaceae, Simaroubaceae, Biebersteiniaceae, Kirkiaceae, and Nitrariaceae. Except for Sapindaceae, Kirkiaceae, Biebersteiniaceae and Nitrariaceae, the phylogeny of the order is clear for most families. This way, the chloroplast genomes are highly informative and are widely utilized for phylogenetics studies; however, in the Sapindales order, the structure of chloroplast genomes has not been analyzed. The aim of this study was to assemble the chloroplast genome of *Pistacia vera* L., build the phylogeny for Sapindales order using complete chloroplast genomes, and to analyze the phylogenetics signals with LSC, SSC, IRA, and CG content. The *P. vera* genome was assembled using reads from NCBI and other complete genomes were downloaded from NCBI in FASTA format. The chloroplast genome of *P. vera* showed 160,674 bp, of which 126 genes, eight rRNA, and 37 tRNA were identified. The phylogeny using complete genomes displayed the same topologies with prior studies and the evolution of the chloroplast showed high correlation with phylogenetic clades for the LSC length and the CG content, of which Sapindaceae is characterized by small LSC sizes and Rutaceae by high CG content. We conclude that Sapindales displayed a robust phylogeny and the chloroplast genomes shows signals phylogenetic for the LSC size and the CG content.

¹ Mestranda em Agricultura e Ambiente, na Universidade Federal de Alagoas –UFAL, campus Arapiraca

² Professor adjunto na Universidade Federal de Alagoas – UFAL, campus Arapiraca

Keywords: Plastomes. Phylogeny. Sapindaceae

1 INTRODUÇÃO

O Sapindales é uma ordem monofilética, que compreende nove famílias com aproximadamente 5.800 espécies (ANGIOSPERM PHYLOGENY GROUP - APG 2016) e teve origem há cerca de 140 milhões de anos, no período cretáceo (MUELLNER-RIEHL et al. 2016). A ordem é composta por Anacardiaceae, Burseraceae, Meliaceae, Rutaceae, Sapindaceae (incluindo Aceraceae, Hippocastanaceae e Xanthoceraceae), Simaroubaceae, Biebersteiniaceae, Kirkiaceae e Nitrariaceae (JUDD et al. 2009), e as análises filogenéticas foram suportadas por três clados: 1) clade Burseraceae + Anacardiaceae, incluindo Kirkiaceae, (2) Clade Sapindaceae, e (3) Clade Meliaceae, Simaroubaceae e Rutaceae (MUELLNER et al. 2007; MUELLNER-RIEHL et al. 2016).

Os avanços no sequenciamento completo de genomas permitiram a obtenção dos genomas completos de cloroplasto para muitas espécies, que foram usados para construir inferências filogenéticas robustas (MOORE et al. 2010; BARRETT et al. 2016). Os genomas dos cloroplastos são conservados entre as espécies, que possuem alta conservação na ordem e conteúdo dos genes, sintenia e permitem a obtenção de alinhamento entre as sequências (RAVI et al. 2008), permitindo a obtenção de inferências filogenéticas e divergentes com alto suporte (SHI et al. 2012; TARKKA et al. 2013; XIE et al. 2016). Entretanto, os estudos utilizando inferências baseadas nas sequências alinhadas e a conservação das estruturas do genoma do cloroplasto pode ser altamente informativa em uma correlação com uma árvore filogenética.

O presente estudo analisou as seguintes questões: (1) Qual é a estrutura do genoma do cloroplasto de *Pistacia vera* L; (2) A árvore filogenética usando o genoma completo dos cloroplastos é mais robusta do que usando sequências parciais para Sapindales; e (3) existem correlações entre os clados filogenéticos com a evolução das estruturas dos IR's, LSC, SSC e conteúdo de CG nos Sapindales. Para responder a essas questões, obtivemos o genoma do cloroplasto de *P. vera*; O genoma completo do cloroplasto foi usado para construir uma árvore filogenômica para a ordem de Sapindales, e uma análise evolucionária foi construída para a estrutura do cloroplasto, os tamanhos LSC, SSC, IRs e o conteúdo GC.

2 MATERIAL E MÉTODOS

2.1 Genoma do cloroplasto de *Pistacia vera*

Para *P. vera* (SRR4453368) os *reads* (*short paired-end reads*) foram obtidas de dados publicados no NCBI, para os quais os arquivos SRA foram descompactados no FASTQ usando o executável FASTQ-DUMP do SRA Toolkit. Em seguida, os arquivos FASTQ foram filtrados com uma qualidade mínima de 10, foram convertidos em arquivos FASTA e foram utilizados para a montagem do genoma.

Quatro milhões de *reads* foram utilizadas para montagem *de novo* usando o *software* Ray (BOISVERT et al. 2012). O BLAST foi usado para identificar os *contigs de novo* que correspondiam aos genomas dos cloroplastos. O IRa e o IRb e a circularização do genoma foram realizados manualmente no *software* Geneious R9. A anotação do genoma foi realizada usando *Sapindus mukorossi* como referência pelo *software* Geneious e foi verificada com a anotação obtida usando Verdant (MCKAIN et al. 2016). Para a anotação usando *Geneious*, foi considerado um mínimo de 80% de corte de identidade entre os genomas. As anotações foram verificadas individualmente e foram corrigidas manualmente para os códons de início e parada, se necessário. *Geneious R9* foi usado para criar uma representação gráfica dos mitogenomas.

2.2 Análises filogenéticas

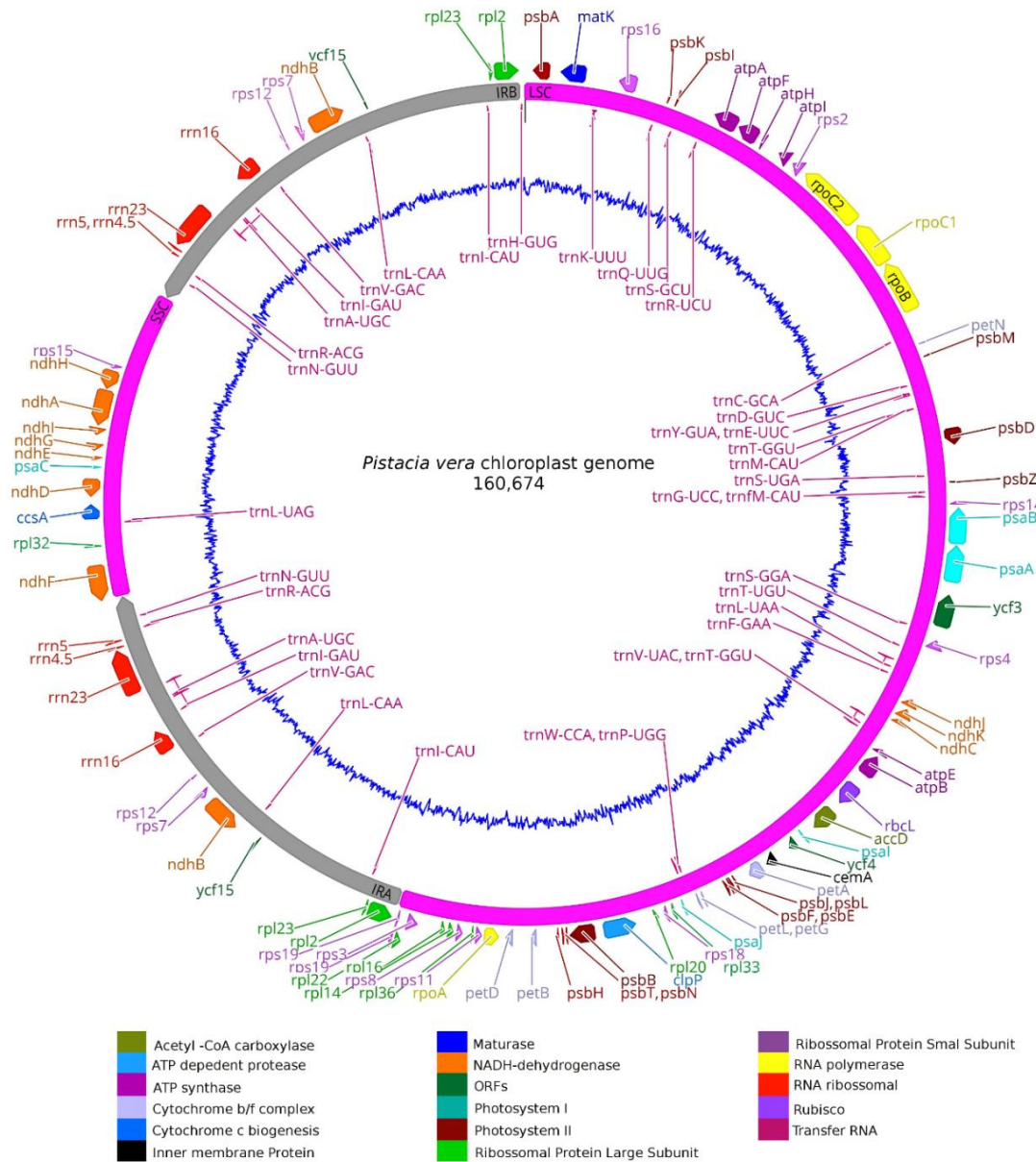
Os genomas de cloroplasto disponíveis no NCBI foram baixados e utilizados para a reconstrução filogenômica. As sequências completas do genoma do cloroplasto foram alinhadas usando o programa MAFFT v7.017 (KATO; STANDLEY 2013) implementado como a ferramenta “*Multiple align*” no *Geneious R9*. O modelo GTR foi determinado utilizando Evolução do Critério de Informação Bayesiana implementada em *jmoldetest* (POUSADA, 2008). A história evolutiva foi inferida utilizando-se a Máxima Verossimilhança e o suporte ao ramo foi avaliado com 1.000 réplicas (*bootstrap*) realizadas no *software* MEGA7 (KUMAR et al. 2016). As relações genéticas entre espécies também foram investigadas através de uma análise de componentes principais (PCA) usando a função *glPca*, pacote implementado no R (JOMBART; AHMED, 2011). A correlação entre a estrutura do genoma do cloroplasto e a árvore filogenética foi realizada com o tamanho de LSC, SSC e IRs e o conteúdo de CG, utilizando a função com Map implementada em *phytools R packager* (REVELL, 2012).

3 RESULTADOS

3.1 Genoma do cloroplasto de *P. vera*

A sequência do genoma do cloroplasto de *P. vera* (KY549635) revelou 160.674 pb, mostrando uma estrutura quadripartidária típica e conservada, com um par de repetições invertidas (IRa e IRb), separadas por regiões LSC e SSC (Figura 1). O IRa e o IRb apresentaram 27.075 pb e 27.156 pb, respectivamente, separados por CSL com 89.606 pb e SSC com 18.381 pb (Figura 1). O genoma compreendia 128 genes, incluindo 82 codificadores de proteínas, 37 RNAt e oito RNAr (Figura 1, Tabela S1), e todos os genes codificadores de proteínas apresentaram AUG como o codão de início.

Figura 1 - Mapa do genoma do cloroplasto de *P. vera*. Os genes anotados são coloridos de acordo com as categorias funcionais, conforme descrito na legenda. Os genes são transcritos no sentido horário e no sentido anti-horário. LSC, região de cópia única grande; SSC, região única de cópia única; repetições invertidas (IRA e IRB). O círculo azul interno representa o conteúdo C + G.

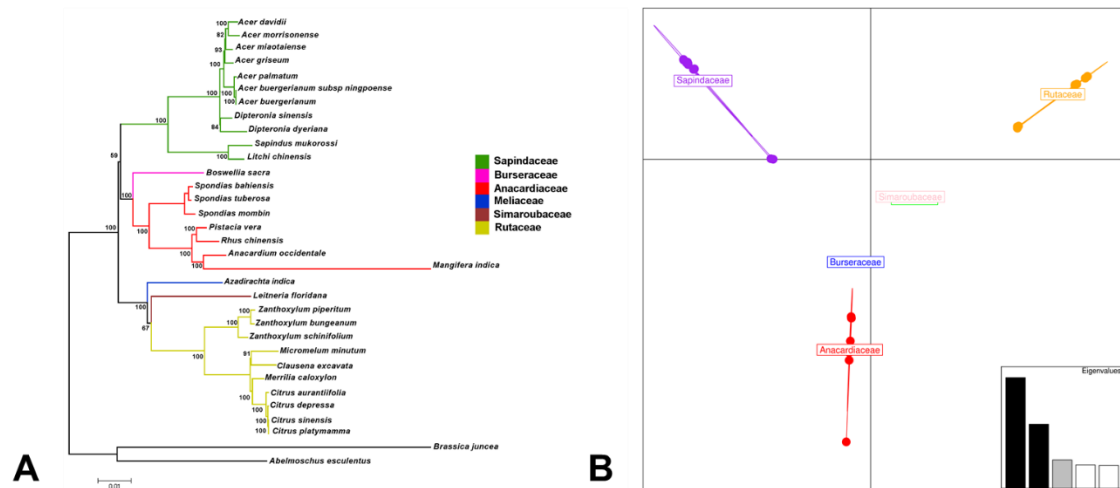


Fonte: A autora (2019)

3.2 Análise comparativa do genoma de cloroplasto em Sapindales

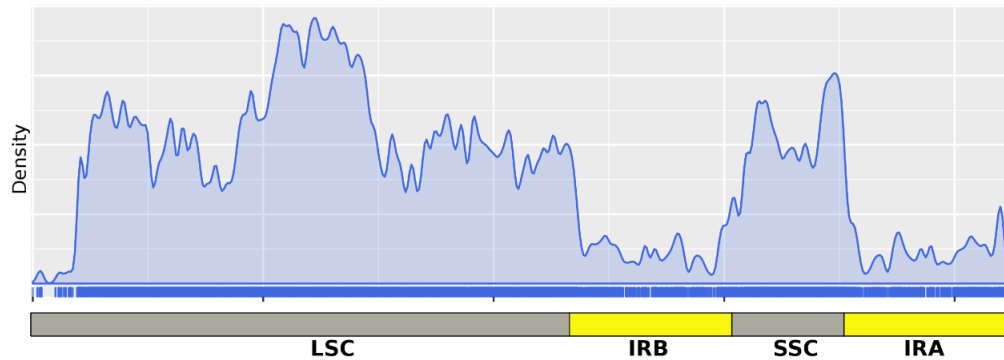
Um total de 31 genomas completos de cloroplasto de Sapindales foi utilizado para a construção de uma árvore filogenética e os resultados mostraram altos valores de sustentação para Anacardiaceae, Rutaceae, Sapindaceae e Burseraceae, com altos valores de *bootstrap* (Figura 2A). Os resultados do PCA para os genomas completos mostraram um claro apoio para as famílias Aracardiaceae, Sapindaceae, Rutaceae e Burseraceae, enquanto para Simaroubaceae e Meliaceae os resultados mostraram em um grupo (Figura 2B). A espécie *Mangifera indica* L. apresentou ramo longo, sugerindo alta evolução; no entanto, uma região invertida de 16.972 pb na região LSC (33,119–50,090 pb) foi observada, incluindo os genes *rps4*, *ycf3*, *psaA*, *psaB*, *rps14*, *psbZ*, *psbC* e *psbD*. Para todas as outras espécies, a ordem e a orientação dos genes foram conservadas. Quando comparadas com as regiões IRA e IRB, as variações entre os genomas completos de cloroplasto para Sapindales, incluindo SNP e *indels*, foram mais abundantes nas regiões LSC e SSC (Figura 3).

Figura 2 - Relacionamentos genéticos em Sapindales. (A) Análise Filogenética Molecular pelo método de Máxima Verossimilhança para genomas completos de cloroplasto; os valores de suporte de ramificação foram estimados por *bootstrap*. (B) PCA usando os genomas completos dos cloroplastos. As cores denotam as famílias Sapindales; os eixos horizontal e vertical representam os dois primeiros componentes principais.



Fonte: A autora (2019)

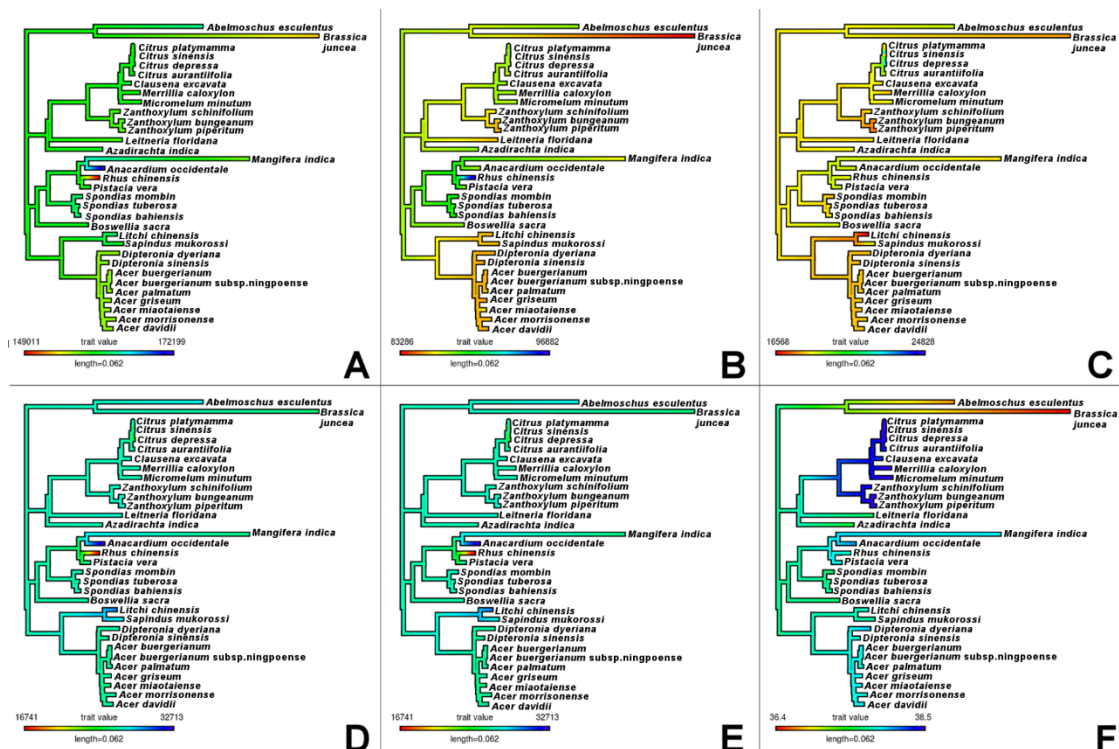
Figura 3 - Distribuição de SNPs entre os genomas de cloroplastos de Sapindales. LSC - cópia única grande, SSC - cópia única pequena, IRA e IRB - regiões repetidas invertidas.



Fonte: A autora (2019)

Para a estrutura quadripartite, a compaixão entre os genomas do cloroplasto mostrou que o LSC variou de 83.286 pb a 96.882 pb, SSC de 16.568 pb a 24.828 pb e os IRs de 16.741 a 32.713 pb (Tabela S1). Entre os genomas, o grande genoma foi *Anacardium occidentale*, com 172.199 pb e o menor genoma foi *Rhus chinensis* 149.011 pb (ambas espécies pertencem à família Anacardiaceae). A correlação entre os tamanhos do genoma completo, LSC, SSC, IRs e o conteúdo de CG, analisados por sinais filogenéticos, mostrou correlação significativa apenas para o tamanho LSC e o conteúdo de CG (Tabela S2 e Figura 4), que a família Sapindaceae apresentou tamanhos menores de LSC (Figura 4B) e a família Rutaceae apresentaram maior conteúdo de CG (Figura 4F).

Figure 4 - Evolução da estrutura do cloroplasto reconstruída utilizando a função “contMap” no pacote R dos *phytools*. (A) tamanho completo do genoma, (B) LSC - tamanho de cópia única grande, (C) SSC - tamanho de cópia única pequena, (D) IRA- tamanho de repetição invertida A, (E) - tamanho de repetição invertida IRB e (F) conteúdo de CG.



Fonte: A autora (2019)

4 DISCUSSÃO

4.1 Relacionamentos genéticos em Sapindales

O genoma do cloroplasto de *P. vera* foi semelhante a outros genomas da família Anacardiaceae, que exibiram o conteúdo similar de genes, ordem e orientação. A árvore filogenômica usando genomas de cloroplastos completos apresentou topologias semelhantes em comparação com APG IV (APG IV 2016), com altos valores de sustentação de ramos para três clados principais: Sapindaceae, Burseraceae + Anacardiaceae e Meliaceae + Simaroubaceae e Rutaceae. Na família Sapindaceae, as topologias corroboram com estudos prévios (MUELLNER et al. 2007; MUELLNER-RIEHL et al. 2016) e na largura de Sapindaceae um clado que corresponde a Aceraceae (antiga família) foi observado, sugerindo que a filogenia molecular sustenta a separação entre Sapindaceae e Aceraceae.

4.2 Evolução do cloroplasto em Sapindales

Os genomas dos cloroplastos em Sapindales, em sua maioria, mostraram uma alta conversação na estrutura. As principais diferenças no genoma do cloroplasto ocorrem na perda de genes, *indels* e rearranjos genômicos (JANSEN et al. 2007; DUCHENE; BROMHAM 2013; SMITH 2015; AMIRYOUSEFI et al. 2018) e as regiões não codificadoras mostram mais variação do que as regiões codificantes (HU et al. 2016; ZHOU et al. 2016; Ren et al. 2018).

Em Sapindales, com exceção de *Mangifera indica*, todos os genomas que foram analisados mostraram similar ordem genética, alta sentença e tamanhos semelhantes das regiões LSC, SSC e IRs. Para a distribuição dos SNPs, as regiões LSC e SSC apresentaram maior variação do que as regiões IRs, sugerindo que as IRs são mais conservadas, corroborando com as análises anteriores (ZHU et al. 2016; REN et al. 2018). O baixo número de SNPs e *InDels* nas regiões de IR mostrou uma correlação com os tamanhos dos IRs, cujos sinais filogenéticos não mostraram significantes entre os clados, corroborando a alta conservação dos IRs. A alta conservação das regiões de IRs pode ser devida aos genes dos ribossomos, o que corresponde a 1/3 dos IRs e é altamente conservada (PALMER, 1985).

5 CONSIDERAÇÕES FINAIS

Em resumo, demonstramos a sequência completa do genoma do cloroplasto para *P. vera*, a alta conservação das regiões de IRs em Sapindales, e que Sapindales mostrou uma filogenia robusta corroborada com APG IV. Para a estrutura do cloroplasto, mostramos que o tamanho LSC e o conteúdo de CG apresentaram correlação com as famílias Sapindales.

REFERÊNCIAS

- AMIRYOUSEFI A; HYVONEN J; POZAI P. The chloroplast genome sequence of bittersweet (*Solanum dulcamara*): Plastid genome structure evolution in Solanaceae. **PLOS ONE**, 2018. DOI: <https://doi.org/10.1371/journal.pone.0196069>. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/29694416/>. Acesso em: 20 fev. 2018.
- APG IV (2016)- Angiosperm Phylogeny Group IV. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. **Bot. J. Linn. Soc.**, v.181, p. 1–20, 2016. DOI: <http://dx.doi.org/10.1111/boj.12385>. Disponível em: <https://academic.oup.com/botlinnean/article/181/1/1/2416499#>. Acesso em: 20 fev. 2018.

BARRETT CF , BAKER WJ , COMER JR , CONAN JG , LAHMEYER SC , LEEBENS-MACK JH , LI J , LIM. GS , MAYFIELD-JONES DR , PÉREZ L , MEDINA J , PIRES JC , SANTOS C , WM STEVENSON D , ZOMLEFER WB , DAVIS JI . Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. **New Phytologist**, v. 209, n. 2, p. 855–870. 2016. DOI: <http://doi: 10.1111 / nph.13617>. Epub 2015 09 de setembro. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/?term=Plastid+genomes+reveal+support+for+deep+phylogenetic+relationships+and+extensive+rate+variation+among+palms+and+other+commelinid+monocots>. Acesso em: 20 mar. 2018.

BOISVERT S, RAYMOND F, GODZARIDIS E, LAVIOLETTE F, CORBEIL J Ray Meta: montagem escalável de metagenoma de *nov*o e perfilamento. **Genome Biology**, 2012. DOI: 13: R122. 2012. <http://doi.org/10.1186/gb-2012-13-12-r122>. Disponível em: <https://genomebiology.biomedcentral.com/articles/10.1186/gb-2012-13-12-r122>. Acesso em: 20 fev. 2018.

DUCHENE D, E BROMHAM L. Rates of molecular evolution and diversification in plants: chloroplast substitution rates correlate with species-richness in the Proteaceae. **BMC Evol. Biol.** 13, 65, 2013. DOI: <http://doi:10.1186/1471-2148-13-65>. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/?term=Rates+of+molecular+evolution+and+diversification+in+plants%3A+chloroplast+substitution+rates+correlate+with+species-richness+in+the+Proteaceae>. Acesso em: 20 mar. 2018.

HU Y; WOESTE KE; ZHAO P. Completion of the chloroplast genomes of five Chinese *Juglans* and their contribution to chloroplast phylogeny. **Front. Plant Sci.** 7, 1955. 2016. DOI: <https://doi: 10.3389 / fpls.2016.01955>. eCollection 2016. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/?term=Completion+of+the+chloroplast+genomes+of+five+Chinese+Juglans+and+their+contribution+to+chloroplast+phylogeny>. Acesso em: 20 fev. 2018.

JANSEN RK, CAI Z, RAUBESON LA, DANIELL H, DEPAMPHILIS CW, LEEBENS-MACK J, KF MÜLLER , GUISINGER-BELLIAN M , RC HABERLE , HANSEN AK , TW CHUMLEY , LEE SB , PEERY R , MCNEAL JR , KUEHL JV , BOORE JL. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. **Proc Natl Acad Sci USA**. 104:19369–19374. 2007. DOI: <https://doi.org/10.1073/pnas.0709121104> PMID: 18048330. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/?term=Analysis+of+81+genes+from+64+plastid+genomes+resolves+relationships+in+angiosperms+and+identifies+genome-scale+evolutionary+patterns>. Acesso em: 20 fev. 2018.

JUDD WS; CAMPBELL CS, KELLOGG EA, STEVENS PF, DONOGHUE, MJ. **Sistemática vegetal: um enfoque filogenético**. 3. ed. Artmed, Porto Alegre. 2009. 612p. Disponível em: <https://www.amazon.com.br/Plant-Systematics-Phylogenetic-Walter-Judd/dp/1605353892>. Acesso em: 20 fev. 2018.

KATO H K E STANDLEY DM. MAFFT Multiple Sequence Alignment Software version 7: performance improvements and usability. **Mol Biol Evol.** 2013 abr; 30 (4): 772-80. 2013. DOI: <http:// dx.doi: 10.1093 / molbev / mst010>. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/?term=MAFFT+Multiple+Sequence+Alignment+Software+version+7%3A+performance+improvements+and+usability>. Acesso em: 20 fev. 2018.

MOORE MJ, SOLTIS PS, BELL CD, BURLEIGH JG, SOLTIS DE. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. **Proceedings of the National Academy of Sciences**. 107(10):4623±8. 2010. [http:// DOI: 10.1073 / pnas.0907801107](http://doi.org/10.1073/pnas.0907801107). Disponível em:

<https://www.ncbi.nlm.nih.gov/pubmed/?term=Phylogenetic+analysis+of+83+plastid+genes+further+resolves+the+early+diversification+of+eudicots>. Acesso em: 20 fev. 2018.

MUELLNER AN, VASSILIADES DD, RENNER SS. Placing Biebersteiniaceae, a herbaceous clade of Sapindales, in a temporal and geographic context. **Pl. Syst. Evol.** 266: 233–252. 2007. DOI: <http://dx.doi.org/10.1007/s00606-007-0546-x>. Disponível em: <https://page-one.springer.com/pdf/preview/10.1007/s00606-007-0546-x>. Acesso em: 20 fev. 2018.

MUELLNER-RIEHL AN, WEEKS A, CLAYTON JW, BUERKI S, NAUHEIMER L, CHIANG Y-C, CODY S, PELL SK. Molecular phylogenetics and molecular clock dating of Sapindales based on plastid rbcL, atpB and trnL-trnF DNA sequences. **TAXON**. 2016. DOI: <http://dx.doi.org/10.12705/655.5>. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.12705/655.5>. Acesso em: 20 mar. 2018.

PALMER JD. Comparative organization of chloroplast genomes. **Annu. Rev. Genet.** 19, 325–354. 1985. [http:// DOI: 10.1146 / annurev.ge.19.120185.001545](http://doi.org/10.1146/annurev.ge.19.120185.001545). Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/3936406>. Acesso em: 20 fev. 2018.

POUSADA D. jModelTest: Mean of the phylogenetic model. **Mol Biol Evol.** 2008 Jul; 25 (7): 1253-6. 2008. DOI: 10.1093 / molbev / msn083. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/18397919>. Acesso em: 20 fev. 2018.

RAVI V, KHURANA JP, TYAGY AK, KHURANA P. An update on chloroplast genomes. **Pl Syst Evol** , v. 271, p. 101–122, 2008. DOI: [http:// dx.doi 10.1007/s00606-007-0608-0](http://dx.doi.org/10.1007/s00606-007-0608-0). Disponível em: https://www.researchgate.net/profile/Paramjit_Khurana/publication/226182666_An_update_on_chloroplast_genome/links/0fcfd50c8242fa8cd9000000/An-update-on-chloroplast-genome.pdf. Acesso em: 25 fev. 2018.

REN T; YANG Y; ZHOU T; LIU ZL. Comparative Plastid Genomes of *Primula* Species: Sequence Divergence and Phylogenetic Relationships. **Int. J. Mol. Sci.** 19, 1050. 2018. DOI: [doi:10.3390/ijms19041050](https://doi.org/10.3390/ijms19041050). Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/?term=Comparative+Plastid+Genomes+of+Primula+Species%3A+Sequence+Divergence+and+Phylogenetic+Relationships>. Acesso em: 25 fev. 2018.

REVELL, L. J. phytools: An R package for phylogenetic comparative biology (and other things). **Methods Ecol. Evol.**, v. 3, p. 217-223, 2012. DOI: [http://dx.doi: 10.1111/j.2041-210X.2011.00169.x](http://dx.doi.org/10.1111/j.2041-210X.2011.00169.x). Disponível em: <https://besjournals.onlinelibrary.wiley.com/doi/10.1111/j.2041-210X.2011.00169.x>. Acesso em: 25 mar. 2018.

SHI C, HU N, HUANG H, GAO J, ZHAO YJ, GAO LZ. An Improved Chloroplast DNA Extraction Procedure for Whole Plastid Genome Sequencing. **Plos One**. 7(2), 2012. DOI: 10.1371 / journal.pone.0031468. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/?term=An+Improved+Chloroplast+DNA+Extraction+Procedure+for+Whole+Plastid+Genome+Sequencing>. Acesso em: 25 fev. 2018.

SMITH DR. Mutation Rates in Plastid Genomes: They Are Lower than You Might Think. **Genome Biol. Evol.** v. 7, p. 1227–1234, 2015. DOI:10.1093/gbe/evv069. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/?term=.+Mutation+Rates+in+Plastid+Genomes%3A+They+Are+Lower+than+You+Might+Think>. Acesso em: 25 fev. 2018.

TARKKA MT, HERRMANN S, WUBET T, FELDHAHN L, RECHT S, KURTH F, ET AL. OakContigDF159.1, a reference library for studying differential gene expression in *Quercus robur* during controlled biotic interactions: use for quantitative transcriptomic profiling of oak roots in ectomycorrhizal symbiosis. **New Phytol.** 199(2):529±40. DOI: Epub 2013/05/16. 2013. doi: 10.1111/nph.12317 PMID: 23672230. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/?term=OakContigDF159.1%2C+a+reference+library+for+studying+differential+gene+expression+in+Quercus+robur+during+controlled+biotic+interactions%3A+use+for+quantitative+transcriptomic+profiling+of+oak+roots+in+ectomycorrhizal+symbiosis>. Acesso em: 25 fev. 2018.

XIE Z, YU C, GUO L, LI M, YONG Z, LIU X, LIN H. Ion Torrent next-generation sequencing reveals the complete mitochondrial genome of black and reddish morphs of the Coral Trout *Plectropomus leopardus*. **Mitochondr DNA.** 27(1):609±12. 2016. DOI: 10.3109 / 19401736.2014.908369. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/?term=Ion+Torrent+next-generation+sequencing+reveals+the+complete+mitochondrial+genome+of+black+and+reddish+morphs+of+the+Coral+Trout+Plectropomus+leopardus>. Acesso em: 25 mar. 2018.

ZHOU T; CHEN C; WEI Y; CHANG Y; BAI G; LI Z; KANWAL N; ZHAO G. Comparative transcriptome and chloroplast genome analyses of two related *Dipteronia* species. **Front. Plant Sci.** 7, 1512. 2016. DOI: 10.3389/fpls.2016.01512. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/?term=Comparative+transcriptome+and+chloroplast+genome+analyses+of+two+related+Dipteronia+species>. Acesso em: 25 mar. 2018.

ZHU AD; GUO WH.; GUPTA S; FAN WS; MOWER, JP. Evolutionary dynamics of the plastid inverted repeat: The effects of expansion, contraction, and loss on substitution rates. **New Phytol.** 209, 1747–1756. 2016. DOI: 10.1111/nph.13743. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/?term=Evolutionary+dynamics+of+the+plastid+inverted+repeat%3A+The+effects+of+expansion%2C+contraction%2C+and+loss+on+substitution+rates>. Acesso em: 25 fev. 2018.

APÊNDICE A - MATEIRAL SUPLEMENTAR

Tabela S1- Genomas de cloroplasto de Sapindales e outgroup usados no estudo

Família	Espécie	NCBI	Tamanho	LSC	IRA	IRB	SSC	% CG
Anacardiaceae	<i>Spondias bahiensis</i>	KU756561	162,218	89,606	27,075	27156	18,381	37.7
Anacardiaceae	<i>Spondias tuberosa</i>	KU756562	162,039	89,453	27,078	27,139	18,369	37.7
Anacardiaceae	<i>Rhus chinensis</i>	KX447140	149,011	96,882	16,741	16,741	18,647	37.8
Anacardiaceae	<i>Pistacia vera</i>	KY549635	160,674	88,236	26,615	26,597	19,086	37.9
Anacardiaceae	<i>Anacardium occidentale</i>	KY635877	172,199	87,727	32,713	32,713	19,046	38.1
Anacardiaceae	<i>Mangifera indica</i>	KY635882	157,780	86,673	26,379	26,379	18,349	37.9
Anacardiaceae	<i>Spondias mombin</i>	KY828469	162,302	89,841	27,176	27,273	18,012	37.6
Brassicaceae	<i>Brassica juncea</i>	KT581449	153,483	83,286	26,211	26,211	17,775	36.4
Burseraceae	<i>Boswellia sacra</i>	KT934315	160,543	88,054	26,763	26,763	18,962	37.6
Malvaceae	<i>Abelmoschus esculentus</i>	KY635876	163,121	88,071	28,009	28,009	19,032	36.7
Meliaceae	<i>Azadirachta indica</i>	KF986530	160,737	88,137	26,982	26,982	18,636	37.5
Rutaceae	<i>Citrus sinensis</i>	DQ864733	160,129	88,744	26,996	26,996	18,393	38.5
Rutaceae	<i>Citrus aurantiifolia</i>	KJ865401	159,893	87,141	26,991	26,991	18,763	38.4
Rutaceae	<i>Citrus platymamma</i>	KR259987	160,121	87,732	26,998	26,998	18,393	38.5
Rutaceae	<i>Zanthoxylum piperitum</i>	KT153018	158,154	85,340	27,644	27,644	17,526	38.5
Rutaceae	<i>Zanthoxylum schinifolium</i>	KT321318	158,963	86,528	27,085	27,085	18,265	38.4
Rutaceae	<i>Clausena excavata</i>	KU949006	161,172	88,056	27,410	27,410	18,296	38.3
Rutaceae	<i>Merrillia caloxylon</i>	KU949003	159,969	87,912	27,013	27,013	18,031	38.5
Rutaceae	<i>Micromelum minutum</i>	KU949007	160,416	87,367	27,213	27,213	18,623	38.5
Rutaceae	<i>Zanthoxylum bungeanum</i>	KX497031	158,401	85,898	27,446	27,446	17,611	38.5
Rutaceae	<i>Citrus depressa</i>	LC147381	160,120	87,752	23,770	23,770	24,828	38.5
Sapindaceae	<i>Sapindus mukorossi</i>	KM454982	160,481	85,649	27,980	27,979	18,873	37.7
Sapindaceae	<i>Litchi chinensis</i>	KY635881	162,524	85,750	30,103	30,103	16,568	37.8
Sapindaceae-Aceraceae	<i>Acer buergerianum subsp. Ningpoense</i>	KF753631	156,911	85,314	26,752	26,752	18,093	37.9
Sapindaceae-Aceraceae	<i>Dipteronia sinensis</i>	KT878501	157,080	>85455	26,766	26,766	18,093	37.8
Sapindaceae-Aceraceae	<i>Acer morrisonense</i>	KT970611	157,197	85,655	26,728	26,728	18,086	37.8
Sapindaceae-Aceraceae	<i>Dipteronia dyeriana</i>	KT985457	157,071	>85529	26,730	26,730	18,082	38.0
Sapindaceae-Aceraceae	<i>Acer davidii</i>	KU977442	157,044	85,410	26,761	26,761	18,112	37.9
Sapindaceae-Aceraceae	<i>Acer miaotaiense</i>	KX098452	156,595	86,327	26,100	26,100	18,068	37.9
Sapindaceae-Aceraceae	<i>Acer buergerianum</i>	KY419137	156,477	86,245	26,090	26,082	18,060	37.9
Sapindaceae-Aceraceae	<i>Acer palmatum</i>	KY457568	157,367	85,829	26,689	26,689	18,160	37.8

Sapindaceae- Aceraceae	<i>Acer griseum</i>	KY511609	156,857	85,227	26,742	26,742	18,146	37.9
Simaroubaceae	<i>Leitneria floridana</i>	KT692940	158,763	85,689	27,443	27,443	18,188	37.6

Fonte: A autora (2019)

Tabela S2- Quantificação dos sinais filogenéticos

	Genome traits					
	Size	LSC	IRA	IRB	SSC	CG content
Lambda	0.16	0.97*	6.61 ⁻⁵	6.61 ⁻⁵	6.61 ⁻⁵	0.99*
p valor	0.35	< 0.001	>0.99	>0.99	>0.99	<0.001

* significant at p <0.005

Fonte: A autora (2019)