

Lilian de Oliveira Machado

**ANÁLISE COMPARATIVA DO GENOMA PLASTIDIAL DE
Myrtaceae: *Acca sellowiana* (O. Berg) Burret, *Eugenia uniflora* L.,
Campomanesia xanthocarpa (Mart.) O.Berg, *Plinia cauliflora* (Mart.)
Kausel, *Plinia aureana* (Mattos) Mattos e *Plinia* sp.**

Tese submetida ao Programa de Pós
Graduação em Recursos Genéticos
Vegetais da Universidade Federal de
Santa Catarina para obtenção do grau
de Doutor (a) em Ciências.

Orientador: Prof. Dr. Rubens Onofre
Nodari

Coorientador: Prof. Dr. Valdir Marcos
Stefenon

Florianópolis

2017

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Machado, Lilian de Oliveira

ANÁLISE COMPARATIVA DO GENOMA PLASTIDIAL DE Myrtaceae :
Acca sellowiana (O. Berg) Burret, Eugenia uniflora L.,
Campomanesia xanthocarpa (Mart.) O.Berg, Plinia cauliflora
(Mart.) Kausel, Plinia aureana (Mattos) Mattos e Plinia
sp. / Lilian de Oliveira Machado ; orientador, Rubens
Onofre Nodari ; coorientador, Valdir Marcos Stefenon. -
Florianópolis, SC, 2017.
197 p.

Tese (doutorado) - Universidade Federal de Santa
Catarina, Centro de Ciências Agrárias. Programa de Pós
Graduação em Recursos Genéticos Vegetais.

Inclui referências

1. Recursos Genéticos Vegetais. 2. Evolução de plantas.
3. Sequenciamento de nova geração. 4. Genoma plastidial. 5.
Myrtaceae. I. Nodari, Rubens Onofre. II. Stefenon, Valdir
Marcos. III. Universidade Federal de Santa Catarina.
Programa de Pós-Graduação em Recursos Genéticos Vegetais.
IV. Título.

Banca Examinadora: Colocar após formatação

Dedico este trabalho aos meus amados pais, Saionara e Altivo e as minhas amadas avós Vera e Eva (*in memoriam*).

AGRADECIMENTOS

Gratidão à vida e a todas as oportunidades que o universo me proporcionou. Esses quatro anos foram muito intensos, de muito aprendizado, crescimento profissional e pessoal.

Agradeço aos meus amados pais, Altivo e Saionara, a minha irmã Daniele, a minha avó Vera e minhas primas Querolla e Ketelen pelo apoio, confiança e por acreditarem em mim e nos meus sonhos. Obrigada pelo apoio incondicional em todos os momentos.

Ao meu orientador, Rubens Onofre Nodari, primeiramente, por ter aceitado me orientar. Pela confiança em realizar este trabalho, muito obrigada. Sou muito grata pelos ensinamentos e oportunidade de crescimento profissional e pessoal que tive nesses quatro anos.

Ao meu coorientador, Valdir Marcos Stefenon, pois foi ele que me apresentou e iniciou nos trabalhos de biologia molecular e genética de plantas e foi por meio dele que conheci o programa de Recursos Genéticos Vegetais. Nada é por acaso.

Aos meus amigos queridos do Paradigma, Daniela, Ricardo e Vinícius pelos cafés no mercado São Jorge, cinema no paradigma e pelos “Le croque madame”. Sou grata a vocês pela amizade e bons momentos partilhados. Obrigada de todo meu coração.

Ao Gustavo e Denise, pois foram os primeiros a me receber no RGV e me acolheram com muito carinho. Obrigada pela amizade e ensinamentos.

Aos amigos Montagna, Daniel, Rafael, Dani, Vini, Denise, Gustavo, Sarah, Clarissa pelos momentos de descontração fora do laboratório.

Ao Vini e ao Daniboy pela ajuda com o inglês sempre que precisei.

Ao Vini, pela ajuda em um dos momentos que mais precisei. Obrigada pela ajuda no meu exame de qualificação.

Ao Montagna pela ajuda na coleta da *Plinia peruviana* que desde 2013 sempre foi uma espécie difícil de trabalhar hehehe. Montagna

obrigada pelos churrascos e pela administração do nosso lar (Casa na Montanha).

À Tassiane pela amizade, por partilhar comigo o quarto por um ano, pelos momentos bons que vivemos, pelos conselhos e carinho.

À Vanessinha pela amizade e companheirismo, por sempre estar ao meu lado nas alegrias e nas tristezas. Obrigada amora!

A minha amiga Grazi que mesmo distante sempre esteve presente em minha vida.

A todos os colegas do LFDGV.

À Leila, pela ajuda, que foi fundamental nesse último ano e principalmente nesse último semestre. Muito obrigada!

Ao Dr. Helisson Faoro, que desde o início tem contribuído com este trabalho.

Aos professores do RGV pelos ensinamentos e à Berna pelo auxílio sempre que precisei.

Ao professor Maurício Sedrez pela disciplina de Ecologia evolutiva, a melhor que já fiz na vida.

Aos professores Américo e Moeses e a Vanessa Sallas da UTFPR pelas espécies de Plinias.

Aos professores Emanuel Maltempi de Souza e Fábio de Oliveira Pedrosa e ao Valter Baura pelo sequenciamento dos genomas de cloroplasto.

A todos que contribuíram direta ou indiretamente para realização deste trabalho. Sou muito grata.

A CAPES e a FAPESC pelo suporte financeiro para realização desse trabalho.

Ao projeto amanhecer, que é um dos projetos mais lindos da UFSC e a Liliana que me apresentou esse projeto. Esse projeto cuidou da parte essencial desse trabalho, a mão de obra. Namastê!

“Há aqueles que se bastam com o grão de mostarda e aqueles que necessitam do Cosmos. Há aqueles que creem sem saber e aqueles que sentem a necessidade de saber para poder crer.”
(Bezerra de Menezes)

RESUMO

A família Myrtaceae pertence à ordem Myrtales e possui cerca de 130 gêneros e aproximadamente 6000 espécies. Essa família de frutíferas é considerada importante em várias formações vegetais, distribuindo-se em todos os continentes, exceto na região Antártica. No Brasil a distribuição ocorre em todos os estados. *Acca sellowiana* (goiabeira-serrana), *Campomanesia xanthocarpa* (guabiroba), *Eugenia uniflora* (pitanga) e *Plinia* spp. (jabuticabas) são espécies de Myrtaceae que têm mostrado boas perspectivas de aumento do uso comercial, tanto que estão no projeto plantas para o futuro da Região Sul e já são cultivadas em pequena escala, comparativamente a outras frutíferas exóticas. Estas espécies destacam-se pelo potencial organoléptico dos frutos, propriedades farmacológicas e valor ornamental, entre outros. Além disso, frutos destas espécies já são atualmente produzidos no Brasil, o que torna um excelente mercado a ser explorado no país. Desta forma, este trabalho tem por objetivo identificar e comparar os genomas plastidiais (cpDNA) de cinco espécies frutíferas de Myrtaceae: *Acca sellowiana*, *Campomanesia xanthocarpa*, *Eugenia uniflora*, *Plinia cauliflora*, *Plinia aureana* e uma espécie híbrida (*Plinia* sp.) visando avançar na compreensão de processos evolutivos e futuramente desenvolver marcadores microssatélites de cloroplasto (cpSSR) específicos visando estudos de diversidade genética dessas espécies em seus habitats. Para tanto, o genoma plastidial (cpDNA) das espécies foi extraído e purificado e seus plastomas foram sequenciados por meio de sequenciamento de nova geração (NGS), em plataforma Illumina MiSeq. Após o sequenciamento foi adotada a estratégia de montagem *de novo*, ou seja, sem utilização de outra espécie como referência. O plastoma completo de *A. sellowiana* possui 159.370 pb de comprimento, região longa de cópia única (LSC) de 88.028 pb, região curta de cópia única (SSC) de 18.598 pb e regiões repetidas invertidas (IRs) com 26.372 pb; *E. uniflora* possui um tamanho de plastoma completo de 158.680 pb, LSC de 87.495 pb, SSC de 18.537 pb e IRs com 26.324 pb; *C. xanthocarpa* possui um plastoma de 158.131 pb de comprimento,

LSC de 87.596 pb, SSC de 18.595 pb e IRs com 25.970 pb; *Plinia aureana* 158.918 pb de comprimento, LSC de 88.204 pb, SSC de 18.462 pb e IRs com 26.126 pb; *Plinia cauliflora* 159.095 pb de comprimento, LSC de 88.162 pb, SSC de 18.615 pb e IRs com 26.159 pb; *Plinia* sp. 158.912 pb de comprimento, LSC de 88.132 pb, SSC de 18.462 pb e IRs com 26.159 pb. Dados de sequenciamento de plastomas completos de espécies da mesma família possibilitarão a realização de estudos comparativos para a identificação de polimorfismos espécie-específicos, os quais poderão gerar marcadores moleculares para estudos de evolução, ecologia, filogenia, filogeografia e diversidade genética.

Palavras-chave: cpDNA, cpSSR, plastoma, feijoa, pitanga, jabuticabas, guabiroba

ABSTRACT

Myrtaceae belongs to the order Myrtales in which about 130 genera and 6000 species are assigned. This family is considered important in various vegetation types, occurring in every continent, except Antarctica region. In Brazil, the family's species are distributed in all states. *Acca sellowiana* (pineapple-guava or feijoa), *Plinia peruviana* (jaboticaba), *Campomanesia xanthocarpa* (guabiroba) and *Plinia* spp. (jaboticabas) are Myrtaceae species that have shown potential for commercial use and are included in the project "Plants to the Future". In addition, these four species stand out for their organoleptic fruit flavours, pharmacological properties and ornamental value, among others. Moreover, currently, fruits of these species are produced on a small scale in Brazil, comparatively to the exotic fruit species, making them an excellent market opportunity to be further exploited in the country. Thus, this study aimed at to identify and to compare the plastid genomes (cpDNA) of four fruit species of Myrtaceae: *Acca sellowiana*, *Campomanesia xanthocarpa*, *Eugenia uniflora*, *Plinia cauliflora*, *Plinia aureana* and a hybrid species (*Plinia* sp.), in order to deepen on their evolutionary process and to develop specific chloroplast microsatellite markers (cpSSR) to carry out further studies on the genetic diversity of these species in their habitats. Therefore, the chloroplast DNA (cpDNA) of the species was extracted and purified, and the plastomas were sequenced by Illumina MiSeq platform next-generation. After the sequencing, the de novo strategy was used, that this, without using another species as a reference. The complete plastid genome of *A. sellowiana* has 159,370 pb Kb in length, large single copy region (LSC) of 88,028 bp, small single copy region (SSC) of 18,598 bp and inverted repeat regions (IRs) with 26,372bp. *E. uniflora* has a complete plastid genome of 158,680 Kb, LSC of 87,495 bp, SSC of 18,537 bp and IRs with 26,324 bp. *C. xanthocarpa* has a complete plastid genome of 158,131 Kb in length, LSC of 87,596 pb bp, SSC of 18,595 bp and IRs with 25,970 bp; *P. aureana* has a complete plastid genome of 158,918

Kb in length, LSC of 88,204 pb bp, SSC of 18,462 bp and IRs with 26,126 bp; *P. cauliflora* has a complete plastid genome of 159,095 Kb in length, LSC of 88,162 pb bp, SSC of 18,615 bp and IRs with 26,159 bp and *Plinia* sp. has a complete plastid genome of 158,912 Kb in length, LSC of 88,132 pb bp, SSC of 18,462 bp and IRs with 26,159 bp. Complete plastomas of species of the same family will allow to perform comparative studies for the identification of species-specific polymorphisms, which can generate molecular markers for the study on evolution, ecology, phylogeny, phlogeography and genetic diversity.

Keywords: cpDNA, cpSSR, plastome, feijoa, pitanga, jaboticabas, guabiroba

LISTA DE FIGURAS

ESTADO DA ARTE

Figura 1. Distribuição geográfica da família Myrtaceae.....	33
Figura 2. Goiabeira-serrana: A- Flor e botão floral; B- Frutos.	35
Figura 3. Pitanga: A- Flor; B- Frutos.....	38
Figura 4. Guabiroba: A- Ramo; B- Frutos; C- Flores.	41
Figura 5. Jabuticabeira: A- Botão floral e flor; B- Flor com polinizado; C- fruto com coloração escura; D- Fruto com coloração verde-clara...	43
Figura 6. Mapa circular do genoma plastidial de <i>Eucalyptus globulus</i> , primeira espécie de Myrtaceae com plastoma publicado.	48

CAPÍTULO 1

Figure 1. Gene map of <i>Acca sellowiana</i> (feijoa) chloroplast genome.	80
Supplementary Fig. 1 Multiple genome alignment performed using Mauve software and the chloroplast genomes sequence of eight Myrtaceae family members.	81
Figure 2. Comparison of boundary positions between single copy (large, LSC, or small, SSC) and inverted repeat (IR) regions among nine plastomes of Myrtales.	82
Figure 3. Bayesian phylogeny based on the complete plastome sequence of eight Myrtoideae (Myrtaceae) species and the outgroup <i>Lagerstroemia fauriei</i> (Myrtales: Lythraceae; KT358807).	96

CAPÍTULO 2

Figure 1. Gene map of <i>Campomanesia xanthocarpa</i> chloroplast genome.	113
Figure 2. Multiple genome alignment performed using Mauve software and the chloroplast genome sequences of nine genera of Myrtaceae.	116
Figure 3. Comparison of final parts (C-terminal) of the amino acid sequences of ycf2 gene in 45 species of Myrtaceae.	117
Figure 4. Bayesian phylogeny based on the cp ycf2 sequence of 45 Myrtaceae species and the outgroup <i>Lagerstroemia fauriei</i> (Myrtales: Lythraceae; KT358807).	124

CAPÍTULO 3

Figura 1. Sequências de três regiões com 557 pb, 757 pb e 963 pb, respectivamente, resolvidas pelo método de sequenciamento Sanger no genoma plastidial de <i>E. uniflora</i>	144
Figura 2. Organização dos genes do genoma do cloroplasto de <i>E. uniflora</i>	146
Figura 3. Análises pontuais da sequência do genoma plastidial de <i>E. uniflora</i> KY392761 contra <i>E. uniflora</i> KR867678.	155
Figura 4. Alinhamento múltiplo do genoma plastidial de dois indivíduos de <i>E. uniflora</i> utilizando o programa MAUVE.	156

CAPÍTULO 4

Figura 1. Sequências de duas regiões com 791 pb e 1076 pb, respectivamente, resolvidas pelo método de sequenciamento Sanger no genoma plastidial de <i>P. aureana</i> .	177
Figura 2. Organização dos genes do genoma do cloroplasto de <i>Plinia aureana</i> .	180
Figura 3. Alinhamento múltiplo do genoma plastidial da tribo Myrteae utilizando o programa MAUVE.	182
Figura 4. Inferência Bayesiana baseada na sequência completa do genoma plastidial de seis diferentes espécies pertencentes à tribo Myrteae (Myrtaceae) e <i>Lagerstroemia fauriei</i> (Myrtales: Lythraceae; KT358807) como outgroup.	186

LISTA DE QUADROS

CAPÍTULO 1

Table 1. Summary of feijoa chloroplast genome characteristics.	77
Table 2. Comparison of chloroplast genomes of Myrtoideae (Myrtaceae) species and Lythraceae analyzed in this study.	83
Table 3. List of genes identified in feijoa chloroplast genome.	84
Table 4. Genes with introns in feijoa chloroplast genome and length of exons and introns.	85
Table 5. SNP and indel variation between feijoa and pitanga plastomes.	87
Table 6. Frequency of SNPs and frequency of transition and transversion substitutions in feijoa compared to pitanga cp genomes.	88
Table 7. Comparison of nonsynonymous (Ka) and synonymous (Ks) substitution rates and Ka/Ks ratio between feijoa and pitanga.	90

CAPÍTULO 2

Table 1. Comparison of chloroplast genomes of Myrtaceae species and outgroup analyzed in this study.....	108
Table 2. <i>Campomanesia xanthocarpa</i> chloroplast genome sequencing and assembly data.....	112
Table 3. Summary of <i>Campomanesia xanthocarpa</i> chloroplast genome characteristics.....	115

Table 4. List of genes identified in <i>Campomanesia xanthocarpa</i> chloroplast genome.....	118
Table 5. Genes with introns in <i>Campomanesia xanthocarpa</i> chloroplast genome and length of exons and introns.....	119
Table 6. Comparison of nonsynonymous (Ka) and synonymous (Ks) substitution rates and Ka/Ks ratio among <i>Campomanesia xanthocarpa</i> and 44 species of Myrtaceae available in GenBank.	120

CAPÍTULO 3

Tabela 1. Listagem de iniciadores de <i>E. uniflora</i> para sequenciamento de fragmentos do genoma plastidial de <i>E. uniflora</i>	138
Tabela 2. Resultado do sequenciamento e montagem de novo do cpDNA de <i>E. uniflora</i>	142
Tabela 3. Resumo das características do genoma plastidial de <i>E. uniflora</i>	145
Tabela 4. Lista de genes identificados no genoma plastidial de <i>E. uniflora</i>	147
Tabela 5. Lista de genes contendo íntrons no genoma plastidial de <i>E. uniflora</i>	148
Tabela 6. Variações intraespecíficas do genoma plastidial de <i>E. uniflora</i>	149
Tabela 7. Diferenças no número e frequência de SNPs e tipo de mutação intraespecífica no genoma plastidial de <i>E. uniflora</i>	151
Tabela 8. Localização de possíveis SNPs e indels entre os genomas plastidiais de <i>E. uniflora</i>	152
Tabela 9. Características do genoma plastidial de dois indivíduos de <i>E. uniflora</i>	154

CAPÍTULO 4

Tabela 1. Lista das coordenadas geográficas e número de tomo das exsicatas de <i>Plinia cauliflora</i> , <i>Plinia aureana</i> e <i>Plinia</i> sp.	168
Tabela 2. Lista de iniciadores para sequenciamento de fragmentos específicos do genoma plastidial de <i>P. aureana</i>	172
Tabela 3. Resultado do sequenciamento e montagem do cpDNA de três espécies de <i>Plinia</i>	176
Tabela 4. Resumo das características dos genomas plastidiais de <i>Plinia</i> spp.	179
Tabela 5. Lista de genes identificados no genoma plastidial de <i>Plinia</i> spp.	181

LISTA DE ABREVIATURAS

AT -- Porcentagem das nucleobases Adenosina e Timina em determinada sequência ou grupos de sequências de DNA.

°C - Graus Celsius

260/230 – Relação entre as absorbâncias 260 nm e 230 nm, que demonstra a pureza de RNA e DNA, indicando contaminação por compostos fenólicos, carboidratos e tiocianato de guanidina

260/280 – Relação entre as absorbâncias 260 nm e 280 nm, que demonstra a pureza de RNA e DNA, indicando contaminação por proteínas, compostos fenólicos e outras substâncias que absorvem fortemente a 280 nm

µg – Micrograma

µl – Microlitro

A – Adenina (nucleobase)

BAG – Banco Ativo de Germoplasma

BLAST – *Basic Local Alignment Search Tool* – Ferramenta para realização de alinhamento local

bp - *Base pairs* - Pares de base

C – Citosina (nucleobase)

cp - cloroplasto

cDNA - DNA complementar, obtido pela polimerização do mRNA pela enzima transcriptase reversa

contig - Sequência contígua, formada pela sobreposição de *reads* em determinado alinhamento

cpDNA - DNA plastidial

cpSSR - Marcador SSR localizado no genoma plastidial

de novo - Método de alinhamento e montagem de genoma ou transcriptoma sem a utilização de outro genoma como referência

DNA – Ácido desoxirribonucleico

DNase - Nuclease que catalisa a degradação do ácido desoxirribonucleico

G – Guanina (nucleobase)

g – grama

GC - Porcentagem das nucleobases Guanina e Citosina em determinada sequência ou grupos de sequências de DNA/RNA

indel - Inserções ou deleções de bases nitrogenadas em determinada sequência de DNA/RNA

intron - *INTRagenic regiON* - região intra-gênica, excisada durante a etapa de *splicing*

IR - regiões invertidas repetidas

kb – **kpb** – Kilo Bases = 1.000 pb

LSC – região longa de cópia simples

LFDGV – Laboratório de Fisiologia do Desenvolvimento e Genética Vegetal

mg – Miligrama

ml - Mililitro

mtDNA - DNA mitocondrial

mRNA - RNA mensageiro

Mya - *Million years ago* (Milhões de anos atrás)

NaCl – Cloreto de sódio

NCBI – *National Center for Biotechnology Information* - Centro Nacional para Informações Biotecnológicas, EUA

NGS - *Next Generation Sequencing* – Sequenciamento de DNA de Nova

Geração

ng/μl – Nanogramas/microlitro

nm - Nanômetros

nM - Nanomolar

ORF - *Open Reading Frame* - Fase aberta de leitura

LFDGV – Laboratório de Fisiologia do Desenvolvimento e Genética Vegetal

pb – Pares de base

PEG - Polietileno Glicol

pH - Potencial hidrogeniônico

reads - Leituras de sequenciamento, com tamanho entre 30 pb a 400 pb, oriundas de NGS

RNA – Ácido Ribonucleico

RNAse – Nucleases que catalisam a degradação do ácido ribonucleico

rRNA - RNA ribossômico ou ribossomal

RuBisCo - Enzima ribulose-1,5-bisfostato carboxilase oxigenasse

S - Sub-unidades do rRNA, composto pelo rRNA citoplasmático, mitocondrial e cloroplastidial

s – Segundos

SSC – região curta de cópia simples

S/I – Relação entre o número de substituições (SNPs) e *indels*

SNP – *Single Nucleotide Polymorphism* - Polimorfismo de base única

SSR - Marcador molecular baseado em sequências repetidas simples (Microsatélite)

T – Timina (nucleobase)

Ti/Tv – “*Transitions / Transversions*” - Relação entre o número de transições e transversões de bases nitrogenadas em SNPs

U – Uracil (nucleobase)

UFPR – Universidade Federal do Paraná

UTFPR – Universidade Tecnológica Federal do Paraná

X - Cobertura de sequenciamento, número de vezes em que a mesma base nitrogenada é sequenciada

xg – Força centrífuga relativa

SUMARIO

1.INTRODUÇÃO E JUSTIFICATIVA	29
2.REVISÃO BIBLIOGRÁFICA	32
2.1 Família Myrtaceae.....	32
2.2 <i>Acca sellowiana</i> (O. Berg.) Burret.....	34
2.3 <i>Eugenia uniflora</i> L.....	37
2.4 <i>Campomanesia xanthocarpa</i> (Mart.) O. Berg.....	40
2.5 <i>Plinia</i> spp.....	42
2.6 Genoma plastidial.....	44
2.7 Relevância genética nos avanços da sistemática de Myrtac.....	48
3.OBJETIVOS.....	52
3.1 Objetivo Geral.....	52
3.1.2 Objetivos Específicos para as seis espécies-alvo do estudo..	52
4.HIPÓTESES	52
REFERÊNCIAS BIBLIOGRÁFICAS	53

CAPÍTULO 1

Abstract.....	71
1.Introduction.....	72
2.Methods.....	74
2.1 Plant material and cpDNA purification.....	75
2.2. Chloroplast genome sequencing, assembling and annotating	74
2.3 Phylogenomic analysis	75
2.4 Synonymous (<i>Ks</i>) and nonsynonymous (<i>Ka</i>) substitution rates	75

2.5 Variation of nucleotide substitutions/indels (S/I) ratio and transition/transversion (Ti/Tv) ratio	76
3.Results and Discussion.....	76
3.1 Feijoa plastome content and organization.....	76
3.2 Comparative analyses with other Myrtaceae cp genomes.....	79
3.3 Evolutionary comparison between feijoa and pitanga	86
3.4 Phylogenomic analysis of Myrtaceae	94
4Conclusion	95
References.....	97

CAPÍTULO 2

Abstract.....	105
1Introduction.....	106
2Methods.....	107
2.1 Plant material and cpDNA purification.....	107
2.2 Chloroplast genome assembly and annotation.....	107
2.3 Synonymous (<i>Ks</i>) and nonsynonymous (<i>Ka</i>) substitution rates. 108	
2.4 Phylogenetic analysis	111
3.Results and Discussion.....	112
3.1 <i>Campomanesia xanthocarpa</i> plastome assembly and gene content.....	112
3.2 Diversity of <i>ycf2</i> protein-coding gene sequence in Myrtaceae	120
3.3 Phylogenetic Inference	123
4.Conclusion	125

References	126
------------------	-----

CAPÍTULO 3

1.Introdução.....	132
2.MATERIAL E MÉTODOS.....	134
2.1 Coleta de material vegetal	134
2.2 Isolamento do cloroplasto e do DNA de cloroplasto (cpDNA).....	134
2.3 Sequenciamento e anotação do genoma plastidial.....	136
2.3.1 Sequenciamento em sequenciador de nova geração.....	136
2.3.2 Desenho de iniciadores e padronização da amplificação via reação em cadeia de polimerase (PCR)	137
2.3.3 Purificação com Polietileno Glicol (PEG)	138
2.3.4 Reação de sequenciamento pelo Método Sanger e purificação pós-reação de sequenciamento.....	139
2.3.5 Anotação do genoma plastidial de <i>E. uniflora</i>	140
2.4 Estruturas Repetitivas e SNPs	141
3.RESULTADOS.....	142
3.1 Montagem do genoma plastidial e conteúdo gênico	142
3.2 Variação intraespecífica do genoma plastidial de <i>E. uniflora</i> ...	149
4.DISSCUSSÃO	154
4.1 Montagem do genoma plastidial e conteúdo gênico de <i>E. uniflora</i>	154
4.2 Variação intraespecífica do genoma plastidial de <i>E. uniflora</i> ...	157
REFERÊNCIAS BIBLIOGRÁFICAS	160

CAPÍTULO 4

1.Introdução.....	167
--------------------------	------------

2.MATERIAL E MÉTODOS.....	168
2.1 Coleta de material vegetal.....	168
2.2 Isolamento do cloroplasto e do DNA de cloroplasto (cpDNA) .	169
2.3 Sequenciamento e anotação dos genomas plastidiais	170
2.3.1 Sequenciamento em sequenciador de nova geração	170
2.3.2 Desenho de iniciadores e padronização da amplificação via reação em cadeia de polimerase (PCR)	171
2.3.3 Purificação com Polietileno Glicol (PEG)	172
2.3.4 Reação de sequenciamento pelo Método Sanger e purificação pós-reação de sequenciamento	173
2.3.5 Anotação dos genomas plastidiais de <i>Plinia cauliflora</i> , <i>Plinia aureana</i> e <i>Plinia</i> sp.	174
2.4 Análise filogenômica da tribo Myrteae.....	175
3.RESULTADOS E DISCUSSÃO	176
3.1 Genoma plastidial e conteúdo e arranjo dos genes	176
3.2 Análise filogenômica da tribo Myrteae.....	184
4.CONCLUSÃO	185
REFERÊNCIAS BIBLIOGRÁFICAS.....	187
CONSIDERAÇÕES FINAIS E PERSPECTIVAS FUTURAS	191
APÊNDICE A - Localização dos SNPs e <i>indels</i> entre os genomas plastidiais de Feijoa/Pitanga	193

1. INTRODUÇÃO E JUSTIFICATIVA

Este estudo envolve cinco espécies e um híbrido da família Myrtaceae, que além de produzirem frutos comestíveis, tem potencial de uso farmacológico. Myrtaceae é considerada uma família de plantas importante em várias formações vegetais. Essa família distribui-se em todos os continentes, exceto em ambientes com níveis de temperatura muito baixos, como na região Antártica (MARCHIORI; SOBRAL, 1997; JUDD et al. 2002; SOUSA; LORENZI, 2005; GOVAERTS et al. 2008). Predomina em regiões tropicais e subtropicais sendo encontrada em grande abundância na América do Sul e Oceania. No Brasil, as espécies distribuem-se principalmente nos Estados de Goiás, Bahia, Minas Gerais, Tocantins, Mato Grosso do Sul, São Paulo, Rio de Janeiro, Paraná, Santa Catarina e Rio Grande do Sul (SOBRAL et al. 2015).

Do ponto de vista econômico, essa família possui representantes como os Eucaliptos (*Eucalyptus* spp.), de origem australiana, cultivados para a obtenção de madeira, além de espécies de uso alimentício, como o cravo-da-índia (*Syzygium aromaticum*), as jabuticabeiras *Plinia* spp., a pitanga (*Eugenia uniflora*) (MARCHIORI; SOBRAL, 1997) e, mais recentemente, tanto a feijoa (*Acca sellowiana*) quanto guabiroba (*Campomanesia xanthocarpa*) vem apresentando aumento de uso conforme SANTOS, 2005 e LISBÔA et al. 2011, respectivamente.

Acca sellowiana, *C. xanthocarpa*, *E. uniflora*, *P. cauliflora*, *P. aureana* e *Plinia* sp. são as espécies frutíferas de Myrtaceae de interesse neste trabalho. Estas espécies destacam-se pelo potencial organoléptico dos frutos, propriedades farmacológicas e valor ornamental. Além disso, frutos de feijoa, guabiroba, pitanga e jabuticaba são atualmente produzidos em pequena escala no Brasil, evidenciando um excelente mercado a ser explorado no país. Portanto, devido às diversas características já mencionadas, essas espécies contribuem para o aumento da variabilidade genética em cultivo e conservação no manejo da biodiversidade.

O sequenciamento do genoma plastidial pode ser considerado uma importante fonte de dados que permite realizar diversas análises comparativas por meio de estudos baseados em filogenia e evolução de plantas. Além disso, as novas tecnologias de sequenciamento automatizado de DNA tornaram possível o acesso ao sequenciamento de genomas organelares completos de forma mais eficiente, custo reduzido e rapidez no processo. Portanto, o sequenciamento de genomas plastidiais pode ser realizado rapidamente, devido à sua pequena dimensão e simplicidade estrutural quando comparado aos genomas nucleares. Ainda, a disponibilidade de genomas de cpDNA completos de diferentes espécies da mesma família são importantes em estudos evolutivos pois poderão identificar polimorfismos espécie-específicos.

Além dos trabalhos já realizados com *A. sellowiana*, o Laboratório de Fisiologia do Desenvolvimento e Genética Vegetal (LFDGV) tem desenvolvido trabalhos em análise da diversidade, sequenciamento de DNA e isolamento e sequenciamento de cloroplasto e análise de transcriptoma. Esse grupo tem mostrado avanços expressivos por meio do projeto “Análises genômicas e transcriptômicas nas coníferas brasileiras *Araucaria angustifolia*, *Podocarpus sellowii*, *Podocarpus lambertii* e *Retrophyllum piresii* visando uso, conservação, estudos evolutivos, moleculares e biotecnológicos”, financiado pela FAPESC (Proc. 14.848/2011-2). O referido projeto apresenta como objetivos principais o sequenciamento e análises do genoma plastidial de coníferas nativas e a realização de análises transcriptômicas em *A. angustifolia* e *P. lambertii*. Para realização dos trabalhos envolvendo sequenciamento de genoma plastidial e transcriptoma, o LFDGV tem parceria com o laboratório do grupo de pesquisa do Dr. Fábio Pedrosa e do Dr. Emanuel Maltempi de Souza do Departamento de Bioquímica e Biologia Molecular da Universidade Federal do Paraná (UFPR).

A presente tese contempla o estudo de genomas plastidiais de *Acca sellowiana*, *Campomanesia xanthocarpa*, *Eugenia uniflora*, *Plinia cauliflora*, *Plinia aureana* e *Plinia* sp., constituindo-se em avanço no conhecimento científico sobre os padrões evolutivos das mesmas. Também se espera contribuir para o melhor conhecimento da biologia e aportar subsídios adicionais ao uso e aperfeiçoamento das estratégias de

conservação destas espécies. Neste contexto, as análises comparativas entre as sequências dos genomas plastidiais destas espécies são cruciais. Os avanços científicos e tecnológicos possibilitaram o desenvolvimento do sequenciamento de nova geração, o qual torna possível sequenciar genomas de diversas organelas de uma só vez, em curto prazo e com baixo custo. Assim, distintas questões relevantes acerca do genoma plastidial podem ser agora respondidas com maior facilidade que anteriormente. Dentre estas questões podem ser destacadas: i) espécies pertencentes a um mesmo grupo possuem genomas plastidiais muito semelhantes? ii) regiões intergênicas de cópia única são mais variáveis que regiões intergênicas repetidas invertidas? iii) espécies pertencentes à mesma subfamília possuem polimorfismos nas mesmas regiões do genoma? iv) os polimorfismos do genoma plastidial podem ser utilizados como marcadores?

Neste contexto, o presente trabalho objetivou a identificação e comparação dos genomas plastidiais (cpDNA) de seis espécies frutíferas de Myrtaceae visando a obtenção de informações genéticas que podem proporcionar um avanço no conhecimento da filogenia, e contribuir para estudos de evolução, ecologia, filogenia, filogeografia e diversidade genética.

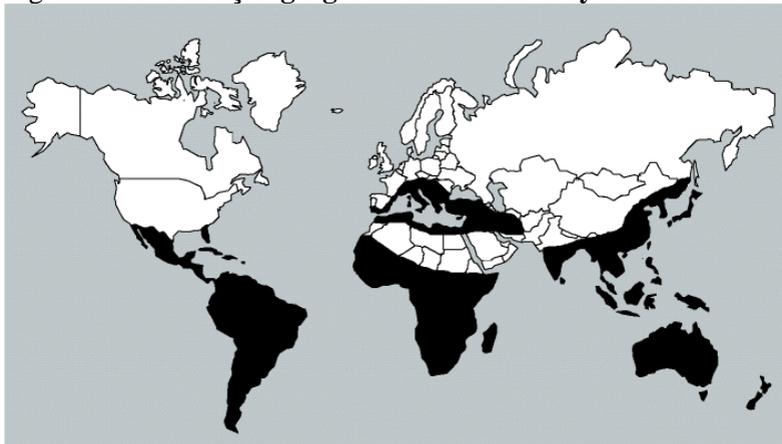
2. REVISÃO BIBLIOGRÁFICA

2.1 Família Myrtaceae

A família Myrtaceae pertence à ordem Myrtales (MARCHIORI, 1995) e é considerada a maior da ordem da família (JOHNSON; BRIGGS, 1984; WATSON; DALLWITZ, 2007), com 132 gêneros e 5.671 espécies (GOVAERTS et al. 2008). Esta família ocorre naturalmente em todos os continentes (Figura 1), exceto na região Antártica, com predominância nas regiões tropicais e subtropicais do globo terrestre, especialmente na América e na Oceania (PEREIRA; NACHTIGAL, 2002; MARCHIORI; SOBRAL, 1997). Myrtaceae é de ampla distribuição na América do Sul, sendo que no Brasil, elas estão distribuídas em todos os Estados (SOBRAL et al. 2015).

Esta família possui dois centros de diversidade no mundo, que podem ser facilmente associados a determinadas características morfológicas: um destes centros é a Oceania onde ocorrem os gêneros *Eucalyptus*, *Melaleuca* e *Callistemon*, com folhas alternas e fruto seco e o outro centro de diversidade é a região Neotropical, onde as plantas apresentam folhas opostas ou verticiladas e frutos carnosos (SOUZA; LORENZI, 2005).

Figura 1. Distribuição geográfica da família Myrtaceae.



Fonte: Grattapaglia et al. 2012 adaptado de Heywood, 1996

Myrtaceae era dividida em duas grandes subfamílias, Myrtoideae com frutos bacoides e Leptospermoideae com frutos capsulares (MCVAUGH, 1968); porém essa divisão já foi questionada por Johnson; Briggs (1984) e Wilson et al. (2001). Então, uma nova classificação foi proposta por Wilson et al. (2005), dividindo Myrtaceae em duas subfamílias, Psiloxylloideae e Myrtoideae, contendo duas e 15 tribos, respectivamente (WILSON et al. 2005).

Myrtaceae é uma das famílias de plantas mais importantes em várias formações vegetais brasileiras, especialmente nas florestas (PEIXOTO; GENTRY 1990; LANDRUM; KAWASAKI 1997; TABARELLI; MANTOVAN 1999; OLIVEIRA FILHO; FONTES 2000; GUILHERME et al. 2004). Todas as mirtáceas existentes no Brasil são da subfamília Myrtoideae, com 26 gêneros distribuídos na flora brasileira: *Acca* O. Berg, *Accara* Landrum, *Algrizea* Proença & NicLugh., *Blepharocalyx* O.Berg, *Calycolpus* O.Berg, *Calyptranthes* Sw., *Calyptrogenia* Burret, *Campomanesia* Ruiz et Pav., *Curitiba* Salywon & Landrum, *Eugenia* L., *Marlierea* Cambess., *Mitranthes*

O.Berg, *Myrceugenia* O.Berg, *Myrcia* DC., *Myrcianthes* O.Berg, *Myrciaria* O.Berg, *Myrrhinium* Schott, *Myrteola* O.Berg, *Myrtus* Tourn. ex L., *Neomitranthes* Kausel ex D.Legrand, *Pimenta* Lindl., *Plinia* L., *Psidium* L., *Siphoneugena* O.Berg, *Syzygium* Gaertn., *Ugni* Turcz. Myrtaceae in Flora do Brazil 2020 (<<http://floradobrasil.jbrj.gov.br/reflora/floradobrasil/FB171>>).

As espécies dessa família geralmente não produzem madeiras valiosas, limitando-se ao fornecimento de lenhas (MARCHIORI; SOBRAL, 1997). Entretanto, a família possui algumas espécies frutíferas de grande interesse comercial (FRANZON et al. 2009) como a pitangueira *Eugenia uniflora* L., a goiabeira *Psidium guajava* L., as jabuticabeiras e a cerejeira *Eugenia involucrata* DC., (REITZ, KLEIN; REIS, 1988) e, mais recentemente, a *Acca sellowiana* (O. Berg), com boas perspectivas de uso comercial (SANTOS, 2005).

Dentre as espécies de Myrtaceae que apresentam grande potencial de uso futuro destacam-se *A. sellowiana*, *E. uniflora*, *P. peruviana* e *C. xanthocarpa* que, devido a características fitodiversas, poderão contribuir para o enriquecimento da matriz agrícola, a conservação dos seus habitats, o uso e a conservação da agrobiodiversidade, entre outros (CORADIN et al. 2011).

2.2 *Acca sellowiana* (O. Berg) Burret

Acca sellowiana (O. Berg.) Burret tem como sinônimos *Feijoa sellowiana* (Berg), *Feijoa obovata* (Berg), *Feijoa schenkiana* Kiaerskou, *Orthostemon sellowianus* (Berg), *Orthostemon obovatus* Berg, e é também conhecida popularmente como goiabeira-serrana, goiabeira-domato, goiabeira da serra ou feijoa (Figura 2). A primeira coleta botânica da *Acca sellowiana* ocorreu em 1856, quando Fredrich Sellow obteve na região de Pelotas, no Estado do Rio Grande do Sul (RS), material que foi identificado pelo botânico Alemão Otto Berg. O nome *Feijoa sellowiana* dado por Berg manteve-se até 1941, quando Burret propôs a inclusão desse gênero em *Acca* criando assim um novo gênero (THORP; BIELESKI, 2002). Embora alguns botânicos insistam no enquadramento

de plantas da espécie no gênero *Acca*, ainda é possível encontrar na literatura científica ambos nomes *Acca sellowiana* e *Feijoa sellowiana*.

Essa espécie está distribuída no sul do Brasil, Uruguai e Argentina (MARCHIORI; SOBRAL, 1997; DUCROQUET et al. 2000; KELLER; TRESSSENS, 2007). No Cone Sul da América do Sul, há duas grandes populações nativas, ou tipos, uma que ocorre nas regiões da Serra Gaúcha e dos campos de altitude do Rio Grande do Sul, Santa Catarina (SC) e Paraná (PR), e a outra que está restrita ao Norte do Uruguai e Sul do Rio Grande do Sul (THORP; BIELESKI, 2002).

Figura 2. Goiabeira-serrana: A- Flor e botão floral; B- Frutos.



Fonte: A - Autor, 2015; B - Poliana Francescato, 2014

Esta espécie possui características importantes como resistência a seca, folhas persistentes e frutos organolépticos (DUARTE et al. 1992), com sabor único. Seus frutos podem ser consumidos *in natura* ou na forma de sucos, geleias e doces (SHARPE et al. 1993; DUCROQUET et al. 2000). Além do aproveitamento dos frutos, suas flores merecem destaque por serem comestíveis e atrativas a pássaros que a polinizam (SAZIMA; SAZIMA, 2007; FRANZON et al. 2004). *Acca sellowiana*, além de ser utilizada em jardins como planta ornamental, pode ter suas flores utilizadas em decorações de ambientes (MATTOS, 1986; FRANZON, et al. 2004). Pode ser utilizada para fins ornamentais e plantios mistos em áreas degradadas, e sua madeira é empregada para pequenas obras (FAIAD et al. 2003). Basile et al. (1997); Vuotto et al. (2000); Ielpo et al. (2000) e Bontempo et al. (2007) comprovaram que dentre as principais propriedades farmacológicas estão

as atividades bactericida e antioxidante, além da existência de flavonoides, cujas propriedades auxiliam na atividade imunológica e na apoptose de células tumorais em casos de leucemia. Além disso, o extrato do fruto da goiabeira-serrana utilizado no tratamento de artrite reumatoide e diabetes tipo-2 foi patenteado nos Estados Unidos da América (2010157678) e vários outros países.

Seu cultivo no Brasil ocorre apenas em escala local (SANTOS et al. 2009) devido aos poucos avanços em termos de pesquisa e desenvolvimento de um sistema de cultivo que transforme esta frutífera em mais uma opção de renda para os agricultores da região de origem. Entretanto, a espécie já despertou o interesse de outros países produtores de frutas e de estudiosos que visam estratégias de conservação e suporte ao cultivo comercial, (SANTOS et al. 2011a). A produção da espécie é mais amplamente feita em países como Estados Unidos da América (Califórnia e Flórida), Nova Zelândia e Colômbia (BARNI et al. 2004) e países bálticos, entre outros, sendo que a Nova Zelândia já possui 13 produtos feitos a partir do fruto (THORP; BIELESKI, 2002). Na Colômbia se pode destacar a produção de frutos desidratados (NAGLE, 2004), comercialização de polpa congelada para sucos, bebidas, bolos, biscoitos e exportação de frutos *in natura* para diversos países, inclusive para o Brasil.

No Brasil, já foram desenvolvidas quatro cultivares de goiabeira-serrana: Alcântara, Helena, Mattos e Nonante (DUCROQUET et al. 2007; DUCROQUET et al. 2008). Entretanto, ainda não foram disponibilizadas para os agricultores. Além disso, o Brasil possui alta diversidade genética e conhecimento local mantido por agricultores, quilombolas e povos indígenas. Além deste conhecimento local, as ferramentas biotecnológicas podem auxiliar no melhoramento genético dessa espécie.

Há diversos estudos de diversidade genética visando ações de conservação da *A. sellowiana*, utilizando marcadores moleculares isoenzimáticos, RAPDs (*Random Amplified Polymorphic DNA*) e microssatélites (SSRs - *Simple Sequence Repeats*) (SANTOS et al. 2011b; SANTOS et al. 2007; WELTER et al. 1999; NODARI et al. 1997). A caracterização parcial dos acessos do Banco Ativo de

Germoplasma (BAG) de São Joaquim revelou uma alta variabilidade genética entre os acessos, com 82% dos locos sendo polimórficos, através de marcadores isoenzimáticos (NODARI et al. 1997). Welter et al. (1999) também encontraram alta variabilidade genética utilizando marcadores RAPD. Além disso, a partir da transferibilidade de marcadores microssatélites do gênero *Eucalyptus* e do desenvolvimento dos próprios marcadores específicos para a espécie (SANTOS et al. 2008), foi possível estimar a diversidade genética de parte dos acessos do BAG de São Joaquim (SANTOS et al. 2007), assim como caracterizar 122 plantas oriundas de populações naturais (SANTOS et al. 2011b). Por meio desses estudos pode-se destacar a importância das áreas de ocorrência natural como fonte de diversidade, tanto para os agricultores quanto para o BAG.

2.3 *Eugenia uniflora* L.

Eugenia uniflora (Figura 3), sinonímia *Eugenia arechavaletae herter*, *Eugenia dasyblasta (o.berg) nied.*, *Eugenia decidua merr.*, *Eugenia fuscopunctata Kiaersk.*, *Eugenia gracilipes Kiaersk.* *Eugenia michelii lam.*, *Syzygium michelii (lam.) Duthie*, *Plinia pedunculata l.f.*, *Plinia petiolata l.*, *Eugenia strigosa (o.berg) arechav.*, *Plinia tetrapetalal.* (SOBRAL et al. 2010), é popularmente conhecida como pitangueira, pitangueira-vermelha, pitanga, pitanga-mulata. Essa espécie é nativa, mas não endêmica do Brasil, distribuindo-se desde o Nordeste até o Sul do País. Sua ocorrência se dá nos domínios fitogeográficos do Cerrado, Mata-Atlântica e Pampa e, segundo Sobral et. al. (2015), os tipos de vegetação de ocorrência da espécie são o Cerrado, Floresta Ciliar ou de Galeria, Floresta Estacional Perenifólia, Floresta Estacional Semidecidual, Floresta Ombrófila, Floresta Ombrófila Mista e Restinga.

Figura 3. Pitanga: A- Flor; B- Frutos.



Fonte: A - Óscar Chaves, 2012; B - Sérgio Bordignon, 2014

A espécie apresenta diversos usos, principalmente por suas propriedades farmacológicas conhecidas, podendo ser utilizada como fitoterápico. As folhas produzem grande quantidade de óleos essenciais utilizados na indústria de perfumes, cosméticos em geral e farmacêutica (CONSOLINI; SARUBBIO, 2002). Estudos farmacológicos comprovam a capacidade antibactericida, antifúngica, antioxidante, imunomodulatória, anti-inflamatória e de redução da pressão arterial (SCHAPOVAL, 1994; CONSOLINI et al. 1999; CONSOLINI; SARUBBIO, 2002; SANTOS et al. 2013; FIGUEIRÔA et al. 2013). Além disso, como planta pioneira, pode ser utilizada na recuperação de áreas degradadas. Produz grande quantidade de flores durante o período de florescimento, tornando-a útil na polinização por abelhas. Outro uso refere-se à excelente aceitação para utilização como planta ornamental de jardins (BOURSCHEID et al. 2011).

Conforme BOURSCHEID et al. (2011), em regiões tropicais e subtropicais a pitangueira apresenta bom crescimento e produtividade, embora ocorra também em regiões de clima temperado e altitude relativamente elevada. Suporta o frio, mesmo em temperaturas abaixo de 0°C e é resistente a geadas. O principal potencial de exploração agroindustrial da pitangueira é a produção de frutos para obtenção da polpa integral congelada e suco engarrafado, além da utilização da polpa

na fabricação de sorvetes, picolés, licores, geleias, vinhos e cosméticos (DONADIO, 1983; FERREIRA et al. 1987; LEDERMAN et al. 1992).

No que se refere à produção e comercialização da pitanga, acredita-se que o Brasil seja o maior produtor mundial da fruta (BOURSCHEID et al. 2011). De acordo com Franzão e Melo (2017), com a demanda crescente dos mercados interno e externo por produtos à base de frutas nativas e de sabor exótico, vislumbra-se a possibilidade de crescimento do mercado interno em pelo menos 100% sobre o volume atual. O mercado externo se constitui em ótima opção, já que é totalmente inexplorado. O cultivo dessa espécie no Nordeste vem crescendo devido à utilização do fruto no preparo de polpa e suco e na fabricação de sorvetes, refrescos, geleias, licores e vinhos (LEDERMAN et al. 1992; BEZERRA et al. 2000). Bezerra et al. (2004) comprovaram que a pitangueira conduzida sob irrigação apresenta boa adaptação às condições do Vale do Rio Moxotó, em Pernambuco, tanto para características de crescimento como de produção e qualidade do fruto. Além disso, os acessos avaliados naquele estudo mostraram grande variabilidade em relação às características físicas e químicas do fruto.

Há uma ampla diversidade genética manifestada na cor do fruto maduro, variando do vermelho-claro até o quase negro. Além disso, outras características variáveis são o tamanho do fruto, presença e ausência de sulcos, acidez, teor de sólidos solúveis totais, número de sementes e tolerância às geadas e à seca. O Brasil, além de conter o maior germoplasma de pitangueira conservado *ex situ*, também possui enorme variabilidade *in situ* ainda não coletada nos vários centros de diversidade e domesticação. Estudos sobre diversidade genética por meio de marcadores moleculares microssatélites (SSRs) e AFLP e análise filogeográfica utilizando espaçadores intergênicos de cloroplasto têm mostrado alta diversidade genética entre as populações estudadas (FERREIRA-RAMOS et al. 2008; FERREIRA-RAMOS et al. 2014; MARGIS et al. 2002; SALGUEIRO et al. 2004; TURCHETTO-ZOLET et al. 2011; TURCHETTO-ZOLET et al. 2016).

2.4 *Campomanesia xanthocarpa* (Mart.) O. Berg

Campomanesia xanthocarpa (Figura 4), sinonímia *Psidium punctulatum* Miq., *Psidium eugenioides* Miq., *Campomanesia malifolia* O. Berg., é popularmente conhecida como guabiroba, guaviroba, guabirova, guabirobeira e guavirobeira. Esta distribuída geograficamente de Minas Gerais e São Paulo até o Rio Grande do Sul, ocorrendo principalmente em solos úmidos e bem drenados das submatas de pinhais de capões e de galeria nos domínios fitogeográficos do Cerrado e Mata Atlântida (SOBRAL *et al.* 2015).

O fruto da guabiroba é suculento, doce e acidulado, podendo ser consumido *in natura* ou utilizado na fabricação de geleias, doces, sucos, licores e sorvetes. Porém, não há informação de plantios comerciais dessa espécie e sua conservação depende da manutenção dos habitats naturais (LISBÔA *et al.* 2011). Conforme Lisbôa *et al.* (2011), a guabiroba está na lista de espécies frutíferas nativas de viveiristas da Região Sul. Além disso, também é considerada uma espécie apícola, importante na produção de mel na região e possui grande potencial de uso futuro pelas características dos seus frutos (aroma e sabor) e propriedades farmacológicas. Estudos de farmacologia indicaram que a espécie auxilia na redução de peso corpóreo, reduz a glicemia, auxilia no tratamento de úlceras e tem potencial antioxidante (BIAVATTI *et al.* 2004; MARKMAN *et al.* 2004; PEREIRA *et al.* 2012).

Figura 4. Guabiroba: A- Ramo; B- Frutos; C- Flores.



Fonte: A - Daniela Werner, 2016; B - João Bagatini, 2014; C - Martin Molz, 2007

Os frutos da guabirobeira constituem um recurso nativo com potencial tecnológico e econômico para a região sul, principalmente para aplicações nos setores de nutrição, farmacêutico e cosmético (PEREIRA et al. 2012). No entanto, a produção e os investimentos no cultivo da espécie ainda são ínfimos e algumas ações são necessárias para a introdução desses frutos no sistema de produção. Assim, se fazem necessários estudos genéticos que possam auxiliar no melhoramento da espécie e o cultivo desta, permitindo a produção do fruto em escala comercial.

2.5 *Plinia* spp.

As espécies de jabuticabeiras (Figura 5) foram reclassificadas para o gênero *Plinia* por Mattos (1998). Contudo, o gênero *Myrciaria* é ainda amplamente utilizado no meio científico, na classificação botânica de jabuticabeiras e pode ser considerado como sinónímia.

Conforme Mattos, (1983) há nove espécies de jabuticabeiras, dentre elas a *Plinia aureana* (Mattos.) Mattos, popularmente conhecida como jabuticaba branca. Algumas espécies são consideradas em extinção, das quais apenas três tem distribuição natural e são cultivadas no Brasil: *Plinia peruviana* (Poir.) Govaerts, conhecida popularmente como jabuticaba de cabinho; *Plinia cauliflora* (DC.) Berg, conhecida popularmente como jabuticaba paulista, pohnema ou açu; e *Plinia jaboticaba* (Vell.) O. Berg, conhecida popularmente como sabará, sendo a mais cultivada e conhecida no Brasil, principalmente nos estados de Minas Gerais e São Paulo, que possuem alguns pomares comerciais.

A jabuticaba é uma das frutas nativas mais conhecidas e apreciadas no Brasil. Os frutos da jabuticabeira podem ser consumidos *in natura*, como também podem ser utilizados nas indústrias de sucos, vinhos, sorvetes, geleias, doces, vinagres, xaropes, licores e jeropigas e tem alto potencial gastronômico, conforme Kinupp et al. (2011). Além disso, alguns estudos evidenciam as propriedades farmacológicas antibactericida e antifúngica que podem ser exploradas (LAGO et al. 2011, ALEZANDRO et al. 2013).

Dentre as principais ameaças as espécies destacam-se a erosão genética causada pela destruição de habitats e das áreas de jabuticabais silvestres. Além disso, outros fatores também ameaçam a espécie, como a redução do cultivo e a perda de possíveis variedades cultivadas há centenas de anos em fazendas do interior do país, sem que estes germoplasmas sejam resgatados, cultivados e difundidos no mercado pela carência de estudos aprofundados e de longo prazo (KINUPP et al. 2011).

Figura 5. Jabuticabeira: A- Botão floral e flor; B- Flor com polinizador; C- fruto com coloração escura; D- Fruto com coloração verde-clara.



Fonte: A;B;C- Moeses Danner, 2015; D- Paulo Aguiar de Oliveira, 2005

Assim, estudos genéticos, farmacológicos, fitossociológicos e ecológicos sobre estas espécies são necessários para que se possa investir no cultivo em larga escala. Segundo Kinupp et al. (2011), a jabuticaba é uma espécie com mercados nacional e internacional garantidos, desde que com produção em quantidades e qualidades satisfatórias.

2.6 Genoma plastidial

As plantas contêm três genomas, o nuclear, o mitocondrial e o plastidial. Neste estudo, o objeto é o genoma do cloroplasto. Por ser menor, haploide e, na maioria das espécies, de herança uniparental, o genoma do cloroplasto se torna atrativo e pertinente para estudos filogenéticos e filográficos.

O cloroplasto é uma organela intracelular da classe dos plastóides presente em algas e plantas e tem função fotossintética. Um componente importante na estrutura do cloroplasto é o extenso sistema de membranas do tilacoide, as quais se estendem por todo o cloroplasto dominando sua estrutura interna. Os tilacoides são os locais de transporte de elétrons e da síntese de adenosina trifosfato (ATP) (BOCK, 2007). Essa organela é responsável desde a captação da energia luminosa até a formação dos compostos orgânicos.

Na década de 1980, Lynn Margulis propôs a teoria endossimbionte, que sugere que a origem do cloroplasto em células vegetais foi o segundo evento endossimbionte e ocorreu pela interação mutuamente vantajosa entre uma célula eucarionte, que já possuía mitocôndria, e uma cianobactéria fotossintética. Conforme Bock; Khan (2004), esse acontecimento teria marcado o surgimento da célula vegetal e fotossinteticamente ativa. Inicialmente, muitos genes podem ter sido transferidos para o núcleo, entretanto, percebe-se que agora existem taxas diferentes na transferência de genes e na perda de linhagens eucarióticas (MARTIN; HERRMANN, 1998; GRAY et al. 1998). Além disso, as organelas são capazes de emitir diferentes sinais para o núcleo, que por sua vez pode interpretar e responder a esses sinais alterando a expressão gênica (WOODSON; CHORY, 2008).

Os plastídeos formam um grupo distinto de organelas em plantas superiores e inferiores e são considerados uma das características que distinguem as plantas dos animais (BOCK, 2007). Acreditava-se que os plastídeos eram exclusivos de plantas, porém descobriu-se que existem organelas semelhantes a plastídeos

denominados apicoplastos em alguns protozoários (ZHANG et al. 1999; GREEN, 2004; HOWE et al. 2008), aumentando assim a dificuldade em determinar características que possam diferenciar um conjunto de plantas e de animais unicelulares. Portanto, o cloroplasto, por ser portador de importante estrutura fotossintética, é a classe de plastídeos mais estudada.

Segundo Bock; Khan (2004), no decorrer da evolução a maioria dos 3000 genes de bactérias endossimbiontes foram perdidos ou transferidos para o genoma nuclear. No entanto, o sequenciamento de genomas inteiros de cloroplastos revelou, pelo menos para angiospermas, que os seus plastomas são consideravelmente conservados, enquanto que os genomas mitocondriais (mtDNA) exibem pouca variação interespecífica. Em plantas, o mtDNA apresenta uma taxa de substituição muito baixa que, combinada com as frequentes recombinações intramoleculares, reduzem a utilidade do genoma mitocondrial como fonte de marcadores, sendo utilizado mais frequentemente o genoma do cloroplasto (AVISE, 1998; PROVAN, 1999; PETIT; VENDRAMIN, 2007). Conforme Bock (2007), as características desejáveis dos marcadores organelares são a ausência de recombinação gênica, rearranjos genéticos, herança materna ou paterna, e a facilidade de acesso à informação contida em seus pequenos genomas circulares.

O DNA do cloroplasto (cpDNA) é o ácido nucleico mais amplamente usado na filogenia de plantas, pois oferece diversas vantagens para a sistemática. O cpDNA é composto por uma cápsula extranuclear e contém informações genéticas que são codificadas nas proteínas enzimáticas e estruturais. Essa organela sofreu especiação e as sequências nucleotídicas de seu DNA podem ser consideradas conservadas e importantes fontes de evidência sobre as relações evolutivas entre taxas fotossintéticas (DHINGRA; FOLTA, 2005). O plastoma da maioria das plantas terrestres possui entre 120 a 160 Kb e estrutura-se em um cromossomo que em parte encontra-se na forma circular e parte na forma linear (BOCK, 2007). Nas plantas terrestres, a maior parte dos genomas plastidiais é dividida em quatro regiões: região longa de cópia única (LSC), região curta de cópia única (SSC) e duas

regiões repetidas invertidas (IR) (BOCK, 2007; YUKAWA et al. 2005; SUGIURA, 1989).

As duas IRs são idênticas na composição, com duas cópias por genoma, porém em sentido de leitura contrário. A função dessas regiões IRs ainda não foi totalmente desvendada, sendo que teorias como o aumento na dosagem de genes altamente expressos, como genes que codificam para rRNA e a estabilização do genoma foram propostas (PALMER; THOMPSON, 1982). Uma evidência circunstancial para a ação da conversão gênica nas IRs vem da observação de que a frequência de mutação de genes das regiões IR é significativamente menor do que para genes localizados em ambas as regiões de cópia única do plastoma (WOLFE et al. 1987; MAIER et al. 1995).

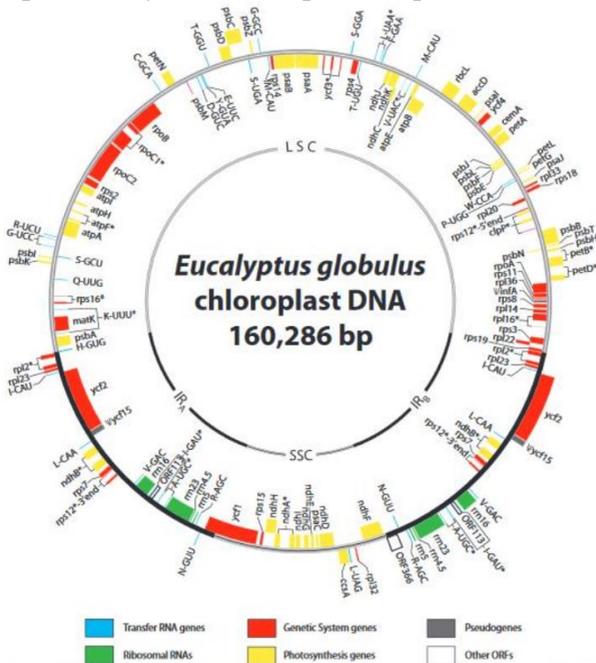
Devido à herança uniparental e ao fato de não ocorrer recombinação nestes genomas, linhagens podem ser identificadas e inferências filogenéticas podem ser feitas ao nível intraespecífico (WALKER; AVISE, 1998; OUBORG et al. 1999, TANGPHATSORNRUANG et al. 2010). A herança materna é relatada na maioria das angiospermas, porém existem exceções significativas como demonstrado por Harris; Ingram (1991) quando revisaram dados que mostraram que 21% dos 233 gêneros de angiospermas e 27% de 398 espécies estudadas apresentam potencial para a transmissão biparental de plastomas. Em estudos realizados com *Eucalyptus*, algumas espécies apresentaram herança materna (BYRNE; MORAN, 1994; MCKINNON et al. 2001).

As plataformas de sequenciamento de DNA de nova geração (*next generation sequencing* - NGS) têm permitido muitos avanços nos estudos de genômica estrutural e funcional. Além disso, são capazes de gerar milhões de pares de base em uma única corrida gerando bilhões de leitura de DNA com mais de 100 nucleotídeos cada (BAKER 2012). Assim, essas plataformas são consideradas uma ótima ferramenta para facilitar os estudos de genômica de plantas por meio do sequenciamento de transcritos, ressequenciamento, que permitem a obtenção da sequência completa dos genomas de organelas (PEARSON et al. 2007; BARBAZUK et al. 2007; HUBER et al. 2007; BOYLE et al. 2008).

Desde os primeiros plastomas sequenciados (SHINOZAKI et al. 1986; OHYAMA et al. 1986), cerca de nove mil organelas foram sequenciadas, dentre estas cerca de 1300 genomas de cloroplastos de diferentes espécies foram sequenciados até agora e encontram-se disponíveis no banco de dados do NCBI (<<http://www.ncbi.nlm.nih.gov/>>). Desses, 43 são plastomas pertencentes à família Myrtaceae. O primeiro genoma plastidial de Myrtaceae publicado foi o de *Eucalyptus globulus* (Steane, 2006) (Figura 6).

O sequenciamento do genoma plastidial pode ser considerado uma importante fonte de dados que permite realizar diversas análises comparativas relacionadas com estudos baseados em filogenia e evolução de plantas. As sequências de genomas plastidiais também servem como excelentes fontes de informação para estudar eventos de transferência horizontal de genes e rearranjos ocorridos nestes compartimentos genéticos durante o processo de evolução da célula vegetal (CHUMLEY et al. 2006; WU et al. 2007; GOREMYKIN et al. 2009; LIN et al. 2010).

Figura 6. Mapa circular do genoma plastidial de *Eucalyptus globulus*, primeira espécie de Myrtaceae com plastoma publicado.



Fonte: Dorothy Steane, 2006

2.7 Relevância genética nos avanços da sistemática de Myrtaceae

A similaridade morfológica entre as espécies da família Myrtaceae tem gerado diversos problemas taxonômicos. Entretanto, estudos filogenéticos moleculares envolvendo sequências do gene plastidial *matK* associadas a dados morfológicos (GADEK et al. 1996, WILSON et al. 2001, WILSON et al. 2005; BIFFIN et al. 2010; THORNHILL et al. 2012; THORNHILL et al. 2015), sequências *ITS*-1, *ITS*-2 e 5.8S rRNA do genoma nuclear (LUCAS et al. 2005, 2007;

BIFFIN, et al. 2010, THORNHILL et al. 2012; THORNHILL et al. 2015), e sequências de rRNA 18S-26S e sequências do gene plastidial *ndhF* (WILSON et al. 2005; BIFFIN et al. 2010; THORNHILL et al. 2012; THORNHILL et al. 2015), têm contribuído para esclarecer as relações entre os gêneros de Myrteae (Myrtaceae). Esses estudos têm indicado que essa tribo é monofilética, com exceção dos gêneros *Syzygium* e *Acmena*, aliança *Acmena* (BRIGGS; JOHNSON, 1979).

Myrteae era dividida em três subtribos, com base na morfologia dos embriões: Eugeniinae, Myrciinae e Myrtinae (MCVAUGH, 1956, LANDRUM; KAWASAKI, 1997). Entre as subtribos, Myrciinae é exclusivamente neotropical, sendo que Myrtinae e Eugeniinae têm distribuição pantropical, alcançando o Mediterrâneo (LANDRUM, 1981). Conforme Lucas et al. (2005), as análises filogenéticas, baseadas nas sequências dos genes ITS-1, ITS-2 e 5.8S rRNA do genoma nuclear, indicam Myrciinae como monofilética, ao contrário das Myrtinae e Eugeniinae, que são parafiléticas. Atualmente, a tribo Myrteae está classificada em grupos informais ao redor dos gêneros *Plinia* L., *Myrcia* DC., *Myrceugenia* O. Berg, *Myrteola* O. Berg, *Pimenta* Lindl. e *Eugenia* L. (LUCAS et al. 2007).

Sytsma et al. (2004) identificaram duas grandes linhagens de cloroplasto dentro dos Myrtales, uma das quais incluiu a linhagem Myrtaceae, e argumentaram que o ancestral da ordem evoluiu em meados do Cretáceo (cerca de 100 Mya) no Sudeste de África. A linhagem Myrtaceae, definida por meio de sequências dos genes cloroplastidiais (*matK*, *ndhF* e *rbcL*), foi datada de 70-80 Mya. Os autores sugeriram que a família diversificou na Australásia, com mudanças mais recentes para as Américas, África, Mediterrâneo e possível dispersão posterior de volta à Austrália. A maioria dos gêneros amostrados parece ter origem no Oligoceno há 30 Mya.

Análises filogenéticas baseadas em dados morfológicos (JOHNSON; BRIGGS, 1984) e posteriores sequências de DNA (WILSON et al, 2001, 2005; SYTSMA et al. 2004) sugerem a origem da família Myrtaceae na Gondwana, com origem e diversificação de Myrteae na Australásia entre 77-56 Mya, quando a Austrália era conectada com a América do Sul através da Antártida. Foram sugeridos

ainda pelos autores pelo menos dois eventos de dispersão distintos trazendo *Eugenia* da América do Sul para a África, *Syzygium* e *Metrosideros* da Australásia para a África, *Tepualia* da Australásia para a América do Sul, e *Myrteae* a partir de uma localidade desconhecida para o Mediterrâneo. *Myrteae* inicialmente se estabeleceu no sul da América do Sul e, posteriormente, se espalhou por todo o continente. Conforme Lucas et al. (2007), a área específica de origem *Myrteae* na América do Sul não está confirmada. Características semelhantes em espécies andinas são encontradas em gêneros que ocorrem na Nova Zelândia. Além disso, características intermediárias são encontradas em gêneros da Austrália e Nova Caledônia, sugerindo um cenário de especiação de *Myrtaceae* nos Andes com posterior migração para Australasia (LUCAS et al. 2007).

Wilson et al. (2001) sugeriram que o frutos deiscentes são plesiomórficos e que os frutos indeiscentes surgiram em quatro ocasiões independentes, conforme análises combinadas de dados morfológicos e moleculares. Além disso, todos os fatores apresentados levam a apoiar a hipótese de que *Myrteae* originou-se na Gondwana. Conforme Thornhill et al. 2015, a estimativa média de idade de *Myrteae* e *Myrtaceae* baseada em taxas de substituições (nucleares e plastidiais) é de 50,7 Ma e 85 Ma, respectivamente. Todos esses estudos evidenciam a importância da genética e biologia molecular pela contribuição a sistemática e de *Myrtaceae*.

3. OBJETIVOS

3.1 Objetivo Geral

Identificar e comparar os genomas plastidiais de seis espécies frutíferas de Myrtaceae: *Acca sellowiana*; *Campomanesia xanthocarpa*; *Eugenia uniflora*; *Plinia cauliflora*; *Plinia aureana* e *Plinia* sp. visando a obtenção de informações genéticas que podem proporcionar um avanço no conhecimento da filogenia, e contribuir para estudos de evolução, ecologia, filogenia, filogeografia e diversidade genética.

3.1.2 Objetivos Específicos para as seis espécies alvo do estudo

- 1) isolar o cloroplasto e posteriormente o cpDNA;
- 2) Sequenciar, montar e anotar os genomas plastidiais;
- 3) comparar os genomas sequenciados com os genomas de outras Myrtaceae já disponíveis no GenBank;
- 4) comparar as regiões codificantes e não codificantes entre as espécies;
- 5) analisar filogeneticamente a relação entre as espécies.

4. HIPÓTESES

- 1) Os genomas plastidiais das espécies alvo deste estudo (*Acca sellowiana*; *Campomanesia xanthocarpa*; *Eugenia uniflora*; *Plinia cauliflora*; *Plinia aureana* e *Plinia* sp.), pertencentes ao grupo *Myrtoidea* são semelhantes aos genomas plastidiais de outras Myrtaceae disponíveis no GenBank tais como,

Eucalyptus globulus (STEANE, 2005); *Eucalyptus grandis* (PAIVA et al. 2011) pertencentes ao grupo *Eucalyptus*; e *Syzygium cumini* (ASIF et al. 2013) pertencente ao grupo *Syzygium*.

Esta hipótese foi testada através da comparação interespecífica dos cpDNAs completos sequenciados.

- 2) As regiões intergênicas possuem taxa evolutiva (frequência de mutação em um sítio nucleotídeo ou aminoácido/ ou região gênica que ocorrem ao longo do tempo) (mutação/sítio/ano) mais elevada que as regiões gênicas;

Esta hipótese foi testada por meio da comparação entre a frequência de mutações em regiões gênicas e intergênicas do genoma plastidial;

- 3) Há polimorfismos nas regiões intergênicas (maior variabilidade), regiões de cópia única, regiões de regiões repetidas invertidas;
- 4) As espécies alvo deste estudo apresentam polimorfismo nas mesmas regiões do genoma plastidial, pois, pela análise filogenética fazem parte da mesma subfamília *Myrtoideae*;
As hipóteses três e quatro foram testadas por meio da comparação da estrutura e organização dos genomas para fornecer elementos para identificação dos polimorfismos.

REFERÊNCIAS BIBLIOGRÁFICAS

ALEZANDRO, M.R.; DUBÉ, P.; DESJARDINS, Y.; LAJOLO, F.M.; GENOVESE, M.I. Comparative study of chemical and phenolic compositions of two species of jaboticaba: *Myrciaria jaboticaba* (Vell.) Berg and *Myrciaria cauliflora* (Mart.) O. Berg. **Food Research International**. 54, p. 468–477. 2013.

AVISE, J.C. History and pureview of phylogeography. **Molecular Ecology**, 7: 371-379. 1998.

BAKER, M. *De novo genome assembly: what every biologist should know*. **Nat Meth**. 9(4): p. 333-337. 2012.

BARBAZUK, W.B., et al. *SNP* discovery via 454 transcriptome sequencing. **Plant J**. 51(5): p. 910-8. 2007.

BARNI, E.J.; DUCROQUET, J.P.; SILVA, M.C.; NETO, R. B.; PRESSER, R. F. Potencial de Mercado para goiabeira-serrana catarinense. Florianópolis: EPAGRI, 2004. 48p. (Documento n. 212).

BASILE, A.; VUOTTO, M.L.; VIOLANTE, U.; SORBO, S.; MARTONE, CASTALDO-COBIANCHI, R. Antibacterial activity in *Actinidia chinensis*, *Feijoa sellowiana* and *Aberia caffra*. *International Journal of Antimicrobial Agents*, v. 8, n. 3, p.199-203, 1997.

BEZERRA, J.E.F.; SILVA JÚNIOR, J.F.; da LEDERMAN, I.E. Pitanga (*Eugenia uniflora* L.). Jaboticabal: FUNEP, 30p. 2000. (Série Frutas Nativas, 1).

BEZERRA, J.E.F.; LEDERMAN, I.E.; SILVA JÚNIOR, J.F.; ALVES, M.A.; Comportamento da pitangueira (*Eugenia uniflora* L) sob irrigação na Região Do Vale Do Rio Moxotó, Pernambuco. **Rev. Bras. Frutic**. BEZERRA Jaboticabal - SP, v. 26, n. 1, p. 177-179, 2004.

BIAVATTI, M.W.; FARIAS, C.; CURTIUS, F.; BRASIL, L.M.; HORT, S.; SCHUSTER, L.; LEITE, S.N.; PRADO, S.R.T. Preliminary studies on *Campomanesia xanthocarpa* (Berg.) and *Cuphea carthagenensis* (Jacq.) J.F. Macbr. aqueous extract: weight control and biochemical parameters. **Journal of Ethnopharmacology** 93. p.385–389. 2004.

BIFFIN, E.; LUCAS, E.J.; CRAVEN, L.A. et al. Evolution of exceptional species richness among lineages of fleshy-fruited Myrtaceae. **Ann Bot.** 106:79–93. doi: 10.1093/aob/mcq088. 2010.

BOCK, R.; KHAN, M.S. Taming plastids for a green future. **Trends Biotechnol.** 22:311-318. 2004.

BOCK, R. Structure, function, and inheritance of plastid genomes. In: Cell and Molecular Biology of Plastids, R. Bock, ed. (Göteborg: Göteborg University), pp. 29-64. 2007.

BONTEMPO, P. MITA, L. MICELI, M. DOTO, A. et al. Feijoa sellowiana derived natural flavone exerts anti-cancer action displaying hDac inhibitory activities. **The International Journal of Biochemistry ; Cell Biology.** v. 39, p.1902-1914, 2007.

BOURSCHEID K, VIEIRA, N. K. LISBÔA G N., KINUPP V F. BARROS, I.B.I. O. Grupos de Uso e as Espécies Prioritárias: *Eugenia uniflora*. In: CORADIN, L.; SIMINSKI, A.; REIS, A. **Espécies nativas da Flora Brasileira de Valor Econômico Atual ou Potencial: Plantas para o Futuro- Região sul.** Brasília: MMA, 2011.170-177p.

BOYLE, A.P., et al. High-resolution mapping and characterization of open chromatin across the genome. **Cell.**132(2): p. 311-22. 2008.

BYRNE, M., MORAN, G.F., Population divergence in the chloroplast genome of *Eucalyptus nitens*. **Heredity** 73, 18–28. 1994.

BRIGGS, B.G.L.; JOHNSON, A.S. Evolution in the Myrtaceae: evidence from inflorescence structure. **Proceedings of the Linnean Society of New South Wales**.102: 158-256. 1979.

CONSOLINI, A.E.; BALDINI, O.A.N.; AMAT, A.G. Pharmacological basis for the empirical use of *Eugenia uniflora* L. (Myrtaceae) as antihypertensive. **Journal of Ethnopharmacology**. 66, p. 33–39. 1999.

CONSOLINI, A.E.; SARUBBIO, M.G. Pharmacological effects of *Eugenia uniflora* (myrtaceae) aqueous crude extract on rats heart. **Journal of Ethnopharmacology**, v. 81, p. 57-63, 2002.

CORADIN, L.; SIMINSKI, A.; REIS, A. *Espécies nativas da Flora Brasileira de Valor Econômico Atual ou Potencial: Plantas para o Futuro-Região sul*. Brasília: MMA, 2011. 934p.

CHUMLEY, T.W.; PALMER, J.D.; MOWER, J.P. The complete chloroplast genome sequence of *Pelargonium x hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. **Mol. Biol. Evol.** 23:2175-2190. 2006.

DHINGRA, A.; FOLTA, K.M. ASAP: Amplification, sequencing ; annotation of plastomes. **BMC Genomics**. 6:176.2005.

DONADIO, L.C. *Fruticultura para pomares domésticos*. Jaboticabal, SP: UNESP-FCAV, 1983. 126p

DUCROQUET, J.P.H.J.; RIBEIRO, P. A goiabeira-serrana: velha conhecida, nova alternativa. *Agropecuária Catarinense*, v. 4, n. 3, p. 27-29, 1991.

DUCROQUET, J.P.H.J.; HICKEL, E.R.; NODARI, R.O. *Goiabeira serrana (Feijoa sellowiana)*. Jaboticabal: FUNEP, 2000. 66p. (Série Frutas Nativas, 5).

DUCROQUET, J.P.H.J.; SANTOS, K.L.; ANDRADE, E.R.; BONETI, J.I.S.; BONIN, V.; NODARI, R. O. As primeiras cultivares brasileiras de

goiabeira serrana: SCS 411 Alcântara e SCS 412 Helena. Agropecuária Catarinense, v. 20, p. 77-80, 2007.

DUCROQUET, J.P.H.J.; NUNES, E.C.; GUERRA, M.P.; NODARI, R. O. Novas cultivares brasileiras de goiabeira serrana: SCS 414- Mattos scs 415- Nonante. Revista Agropecuária Catarinense, v. 21, n. 2, p. 77- 80, 2008.

FAIAD, M.G.R.; SALOMÃO, A.N.; PADILHA, L.S.; MUNDIN, R.C. **Sobrevivência de Colletotrichum gloeosporioides (Penz.) Sacc. em sementes de feijoa (Acca Sellowiana Burr.) durante o armazenamento.** Embrapa. Brasília-DF, 2003.

FERREIRA, F.R.; FERREIRA, S.A.N.; CARVALHO, J.E.U. Espécies frutíferas pouco exploradas, com potencial econômico e social para o Brasil. Revista Brasileira de Fruticultura, v.9, n. extra, p.11-22, 1987

FERREIRA-RAMOS, R.; LABORDA, P.R.; SANTOS, M.O.; MAYOR, M.S.; MESTRINER, M.A.; DE SOUZA, A.P.; ALZATE-MARIN, A.L. Genetic analysis of forest species *Eugenia uniflora* L. through of newly developed SSR markers. **Conserv Genet.** 9:1281–1285. 2008.

FERREIRA-RAMOS, R.; ACCORONI, K.A.G.; ROSSI, A.; GUIDUGLI, M.C.; MESTRINER, M.A.; MARTINEZ, C.A.; ALZATE-MARIN, A.L. Genetic diversity assessment for *Eugenia uniflora* L., *E. pyriformis* Cambess., *E. brasiliensis* Lam. and *E. francavilleana* O. Berg neotropical tree species (Myrtaceae) with heterologous SSR markers. **Genetic Resources and Crop Evolution.** v. 61. 267-272. 2014.

FRANZÃO, A.A.; MELO, B. Cultura da Pitangueira. núcleo de estudos em fruticultura no cerrado, Uberlândia, s. d. Disponível em: <<http://www.fruticultura.iciag.ufu.br/pitangueira.html>>. Acesso em 28 Dez. 2017.

FRANZON, R.C.; CORRÊA, E.R.; RASEIRA, M.C.B. Potencialidades de Produção de Mirtáceas Frutíferas Nativas do Sul do Brasil. In. Resumos do

II Simpósio Nacional do Morango e 1º Encontro de Pequenas Frutas Nativas do Mercosul, Pelotas, RS 2004.

FRANZON, R. C. et al. **Araçás do Gênero *Psidium***: principais espécies, ocorrência, descrição e usos. Planaltina, DF: Embrapa Cerrados, 2009. p. 10-12. 48p.

FIGUEIRÔA, E.O.; SILVA, L.C.N.; MELO, C.M.L.; NEVES, J.K.A.; SILVA, L.N.H.; PEREIRA, V.R.A.; CORREIA, M.T.S. Evaluation of Antioxidant, Immunomodulatory, and Cytotoxic Action of Fractions from *Eugenia uniflora* L. and *Eugenia malaccensis* L.: Correlation with Polyphenol and Flavanoid Content. **The ScientificWorld Journal**. v. 2013. 2013.

FINATTO, T.; SANTOS, K.L.; STEINER, N.; BIZZOCCHI, L.; HOLDERBAUM, D. F. DUCROQUET, J.P.H.J. Guerra, M. P. Nodari, R. O. Late-acting self-incompatibility in *Acca sellowiana* (Myrtaceae). **Australian Journal of Botany**, 59, p. 53–60, 2011.

GADEK, P.A.; WILSON, P.G.; QUINN, C.J. Phylogenetic reconstruction in Myrtaceae using matK, with particular reference to the position of *Psiloxylon* and *Heteropyxis*. **Australian Systematic Botany**. 9:283–290. 1996.

GOREMYKIN, V.V.; SALAMINI, F.; VELASCO, R. Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. **Mol. Biol. Evol.** 26:99-110. 2009.

GRATTAPAGLIA, D.; VAILLANCOURT, R.E.; SHEPHERD, M.; THUMMA, B.R.; FOLEY, W.; KÜLHEIM, C.; POTTS, B.M.; MYBURG, A.A. Progress in Myrtaceae genetics and genomics: *Eucalyptus* as the pivotal genus. **Tree Genetics ; Genomes**. 8:463–508. 2012.

GOVAERTS, R.; SOBRAL, M.; ASHTON, P.; BARRIE, F.; HOLST, B.K.; LANDRUM, L.R.; MATSUMOTO, K.; MAZINE, F.F.; NIC

LUGHADHA, E.; PROENÇA, C.; SOARES-SILVA, L.H.; WILSON, P.G.; LUCAS, E.J. World checklist of Myrtaceae. **Kew Publishing**, Royal Botanic Gardens, Kew. 2008.

GRAY, M.W. et al. Genome structure and gene content in protist mitochondrial DNAs. *Nucleic Acids Res.* 26, 865–878.1998.

GREEN, B.R. The chloroplast genome of dinoflagellates: a reduced instruction set? **Protist.** 155:23-31. 2004.

GUILHERME, F.A.G.; MORELLATO, L.P.C.; ASSIS, M.A. Horizontal and vertical tree community structure in a lowland Atlantic rain forest, Southeastern Brazil. **Revista Brasileira de Botânica.** 27:725-737. 2004.

HARRIS, S.A.; INGRAM R. Chloroplast DNA and biosystematics: the effects of intraspecific diversity and plastid transmission. *Taxon* 40, 393-412. 1991.

HEYWOOD, V.H. **Flowering plants of the world.** B.T. Batsford Ltd., London. 1996.

HOWE, C.J.; NISBET, R.R.E.; BARBROOK, A.C. The remarkable chloroplast genome of dinoflagellates. *Journal of Experimental Botany.* vol. 59, No. 5, pp. 1035–1045, 2008.

HUBER, J.A., et al. Microbial population structures in the deep marine biosphere. **Science.** 318(5847): p. 97-100. 2007.

IELPO, M.T.L.; BASILE, A.; MIRANDA, R.; MOSCATELLO, V.; NAPPO, C.; SORBO, S.; LAGHI, E.; RICCIARDI, M.M.; RICCIARDI, L.; VUOTTO, M. L. Immunopharmacological properties of flavonoids. *Fitoterapia*, v. 71, p.101-109, 2000.

JOHNSON, L.A.S.; BRIGGS, B.G. Myrtales and Myrtaceae a phylogenetic analysis. *Annals of the Missouri Botanical Garden* 71: 700-756, 1984.

JUDD, W. S. **Plant systematics: A phylogenetic approach**. 2 ed. Sunderland, Massachusetts: Sinauer Associates, 2002. 576 p.

KELLER, H.A.; TRESSENS, S.G. Presencia en argentina de especies de uso múltiple: *Acacia sellowiana* (myrtaceae) y *Casearia lasiophylla* (flacourtiaceae). *Darwiniana*, v. 45, n. 2, p. 204-212, 2007.

KINUPP, V.F.; LISBÔA, G.N.; BARROS, I.B.I.O. Grupos de Uso e as Espécies Prioritárias: *Plinia peruviana*. In: CORADIN, L.; SIMINSKI, A.; REIS, A. *Espécies nativas da Flora Brasileira de Valor Econômico Atual ou Potencial: Plantas para o Futuro- Região sul*. Brasília: MMA, 2011. 198-204p.

KINUPP, V.F. Espécies Alimentícias Nativas da Região Sul do Brasil. In: CORADIN, L.; SIMINSKI, A.; REIS, A. *Espécies nativas da Flora Brasileira de Valor Econômico Atual ou Potencial: Plantas para o Futuro- Região sul*. Brasília: MMA, 2011. 107-110p.

LAGO, J.H.G.; SOUZA, E.D.; MARIANE, B.; PASCON, R.; VALLIM, M.A.; MARTINS, R.C.C.; BAROLI, A.A.; CARVALHO, B.A.; SOARES, M.G.; SANTOS, R.T.; SARTORELLI, P. Chemical and Biological Evaluation of Essential Oils from Two Species of Myrtaceae — *Eugenia uniflora* L. and *Plinia trunciflora* (O. Berg) Kausel. **Molecules**. 16, p. 9827-9837. 2011.

LANDRUM, L.R.A monograph of the genus *Myrceugenia* (Myrtaceae). *Flora Neotropica*, Monograf. 29: 1-137. 1981.

LANDRUM, L.R.; KAWASAKI, M.L. The genera of Myrtaceae in Brazil: an illustrated synoptic treatment and identification keys. **Brittonia**, v.49, p.508-536, 1997.

LEDERMAN, I.E.; BEZERRA, J.E.F.; CALADO, G. **A pitangueira em Pernambuco**. Recife: IPA, 1992. 20p. (Documentos, 19).

LIN, C.P.; HUANG, J.P.; WU, C.S.; HSU, C.Y.; CHAW, S.M. 2010. Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. **Genome Biol. Evol.** 2:504-517. 2010.

LISBÔA, G.N.; KINUPP, V.F.; BARROS, I.B.I.O. Grupos de Uso e as Espécies Prioritárias: *Campomanesia xanthocarpa*. In: CORADIN, L.; SIMINSKI, A.; REIS, A. *Espécies nativas da Flora Brasileira de Valor Econômico Atual ou Potencial: Plantas para o Futuro- Região sul*. Brasília: MMA, 2011. 159-162 p.

LUCAS, E.J.; BELSHAM, S.A.; NICLUGHADA, E.M.; ORLOVICH, D.; SAKURAGUI, C.; CHASE, M.; WILSON, P.G. Phylogenetics patterns in the fleshy-fruited Myrtaceae - preliminary molecular evidence. **Plant Systematics and Evolution**, v.251, p. 35-51, 2005.

LUCAS, E.J, HARRIS, S.A.; MAZINE, F.F.; BELSHAM, S.R.; NICLUGHADA, E.M.; TELFORD, A., GASSON, P.E.; MARK W.; CHASE, M.W. Suprageneric phylogenetics of Myrteae, the generically richest tribe in Myrtaceae (Myrtales). **Taxon**,v.56, n.4, p.1105–1128, 2007.

MAIER, R.M.; NECKERMANN, K.; IGLOI, G.L.; KÖSSEL, H. Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. **Journal Molecular Biology**, 251:614-628. 1995.

MARCHIORI, J. N. C. **Elementos de dendrologia**. Santa Maria: Ed. UFSM, 1995. p.35. 163p.

MARCHIORI, J. N. C. SOBRAL, M. **Dendrologia das angiospermas – Myrtales**. Santa Maria: Ed. da UFSM, 1997. 304 p.

MARKMAN, B.E.O.; BACCHI, E.M.; KATO, E.T.M. Antiulcerogenic effects of *Campomanesia xanthocarpa*. **Journal of Ethnopharmacology**. 94, p. 55–57. 2004.

MARGIS, R.; FELIX, D.; CALDAS, J.F.; SALGUEIRO, F.; DE ARAUJO D.S.D.; BREYNE, P.; VAN MONTAGU, M.; DE OLIVEIRA, D.; MARGIS-PINHEIRO, M. Genetic differentiation among three neighboring Brazil-cherry (*Eugenia uniflora* L.) populations within the Brazilian Atlantic rain forest. **Biodiversity and Conservation**. 11: 149–163, 2002.

MARTIN, W.; HERRMANN, R.G. Gene transfer from organelles to the nucleus: how much, what happens and why? *Plant Physiol.* 118, 9–17. 1998.

MATTOS, J. R. **Myrtaceae do Rio Grande do Sul**. Roesslária, Porto Alegre, v. 5, n. 2, p. 315-317, 1983.

MATTOS, J. R. A Goiabeira-Serrana. Porto Alegre: Instituto de Pesquisas de Recursos Naturais Renováveis, (Publicação IPRNR, 19), 1986. 84p.

MATTOS, J.R. Novidades taxonômicas em Myrtaceae – XV. **Loefgrenia**: comunicações avulsas de Botânica, Florianópolis, n.112. 1998. 9p.

MCKINNON, G.E.; VAILLANCOURT, R.E.; JACKSON, H.D.; POTTS, B.M. Chloroplast sharing in the Tasmanian Eucalypts. *Evolution* 55 (4), 703–711. 2001.

MCVAUGH, R. Tropical American Myrtaceae. Notes on generic concepts and descriptions of previously unrecognized species. **Fieldiana, Botany** 29(3): 145-228. 1956.

MCVAUGH R. The genera of American Myrtaceae – an interim report. **Taxon** 17: 354– 418. 1968

NAGLE, A.R. El cultivo de la feijoa (*Feijoa sellowiana* Berg). Universidad De Caldas, Facultad de Ciências Agropecuarias, Programa de Agronomia, 2004. 48p.

NODARI, R.O.; DUCROQUET, J.P.H.J. GUERRA, M. P.; MELER, K. Genetic variability of Feijoa sellowiana germoplasm. *Acta Horticulturae*, v. 452, p. 41-46. 1997.

OHYAMA, K.; FUKUZAWA, H.; KOHCHI, T.; et al. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* 322:572-574. 1986.

OLIVEIRA FILHO, A.T.; FONTES, M.A. Patterns of floristic differentiation among Atlantic forests in southeastern Brazil, and the influence of climate. *Biotropica* 32:793-810. 2000.

OUBORG, N.J.; PIQUOT, Y.; VAN GROENENDAEL, J. M. Population genetics. molecular markers and the study of dispersal in plants. *Journal of Ecology*, 87: 551-568. 1999.

PALMER, J.D.; THOMPSON, W.F. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell*, 29:537-550. 1982.

PEARSON, B.M.; et al. The complete genome sequence of *Campylobacter jejuni* strain 81116 (*NCTC11828*). *J Bacteriol.* 189 (22): p. 8402-3. 2007.

PEIXOTO, A.L.; GENTRY, A.H. Diversidade e composição florística da mata de tabuleiro na Reserva Florestal de Linhares (Espírito Santo, Brasil). *Revista Brasileira de Botânica*. 13:19-25. 1990.

PEREIRA, F.M.; NACHTIGAL, J.C. Goiabeira. In: BRUCKNER. C.H. *Melhoramento de Fruteiras Tropicais*. Viçosa: UFV, 2002. P. 267-289.

PEREIRA, M.C.; STEFFENS, R.S.; JABLONSKI, A.; HERTZ, P.F.; RIOS, A.O.; VIZZOTTO, M.; FLÔRES, S.H. Characterization and Antioxidant Potential of Brazilian Fruits from the Myrtaceae Family. *J. Agric. Food Chem.* 60, p. 3061–3067. 2012.

PETIT, R.J.; VENDRAMIN, G.G. Plant phylogeography based on organelle genes: na introduction. *In*: Weiss, S.; Ferrand, N. (Eds). **Phylogeography of Southern European Refugia**. Springer. p. 23-97. 2007.

PROVAN, J.; POWELL, W.; HOLLINGSWORTH, P.M. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. **Trends in Ecology; Evolution**, 16: 142-147. 2001.

Plinia in **Flora do Brasil 2020 em construção**. Jardim Botânico do Rio de Janeiro. Disponível em: <<http://floradobrasil.jbrj.gov.br/reflora/floradobrasil/FB10824>>. Acesso em: 30 Dez. 2016

SALGUEIRO, F.; FELIX, D.; CALDAS, J.F.; MARGIS-PINHEIRO, M.; MARGIS, R. Even population differentiation for maternal and biparental gene markers in *Eugenia uniflora*, a widely distributed species from the Brazilian coastal Atlantic rain Forest. **Diversity and Distributions**. 10:201-210. 2004.

SANTOS, K. I. Bases genéticas de características de importância agrônômica em goiabeira-serrana (*Acca sellowiana*) 2005. 125 f. Dissertação (mestrado em ciências, área de concentração em Recursos Genéticos Vegetais) - faculdade de agronomia, universidade federal de santa catarina, Florianópolis, SC.

SANTOS, K.L.; WELTER, L.J.; DANTAS, A.C.M.; GUERRA, M.P.; DUCROQUET, J.P.H.J.; NODARI, R. O. Transference of microsatellite markers from *Eucalyptus* spp. to *Acca sellowiana* and the successful use of this technique in genetic characterization. *Genetics and Molecular Biology*, v. 30, n. 1, p. 73-79, 2007.

SANTOS, K.L.; SANTOS, M.O.; LABORDA, P.R.; SOUZA, A.P.; PERONI, N.; NODARI, R.O. Isolation and characterization of microsatellite markers in *Acca sellowiana* (berg) burret. *Molecular Ecology Resources*, v. 8, p. 998-1000, 2008.

SANTOS, K.L.; SIMINSKI, A.; DUCROQUET, J.P.H.J.; GUERRA, M.P.; PERONI, N.; NODARI, R. O. Grupos de Uso e as Espécies Prioritárias: *Acca sellowiana*. In: CORADIN, L.; SIMINSKI, A.; REIS, A. *Espécies nativas da Flora Brasileira de Valor Econômico Atual ou Potencial: Plantas para o Futuro- Região sul*. Brasília: MMA, 2011a. 111-129p.

SANTOS, K.L.; DUCROQUET, J.P.H.J.; NODARI, R.O. Caracterização genética de populações naturais de goiabeira serrana (*Acca sellowiana*) com marcadores microssatélites heterólogos. **Biotemas**, v.4, p. 75-83, 2011b.

SANTOS, K.L.; PERONI, N.; GURIES, R.P.; NODARI, R.O. Traditional Knowledge and Management of Feijoa (*Acca sellowiana*) in Southern Brazil. **Economic Botany**, v.63, p.204-214, 2009.

SANTOS, K.L. **Diversidade cultural, genética e fenotípica da goiabeira-serrana (*Acca sellowiana*)**: implicações para a domesticação da espécie. Florianópolis, 2009b. Tese (Doutorado) Programa de Pós-Graduação em Recursos Genéticos Vegetais, Centro de Ciências Agrárias, Universidade Federal de Santa Catarina.

SANTOS, K.L.; WELTER, L.J.; DANTAS, A.C.M.; GUERRA, M.P.; DUCROQUET, J.P.H.J.; NODARI, R.O. Transference of microsatellite markers from *Eucalyptus* spp to *Acca sellowiana* and the successful use of this technique in genetic characterization. **Genetics and Molecular Biology**, v. 1, p. 73-79, 2007.

SANTOS, K.K.A.; MATIAS, E.F.F.; TINTINO, S.R.; SOUZA, C.E.S.; BRAGA, M.F.B.M.; GUEDES, G. M.M.; COSTA, J.G.M.; MENEZES, I.R.A.; COUTINHO, H.D.M. Enhancement of the Antifungal Activity of Antimicrobial Drugs by *Eugenia uniflora* L. **J Med Food**.16 (7), p.669–671. 2013.

SANTOS, C.M.R.; FERREIRA, A.G.; ÁQUILA, M.E.A. Características de frutos e germinação de sementes de seis espécies de Myrtaceae nativas do Rio Grande Do Sul. **Ciência Florestal**. Santa Maria. v.14 n. 2 p. 13-20, 2004.

SAZIMA, I.; SAZIMA, M. Petiscos florais: pétalas de *Acca sellowiana* (Myrtaceae) como fonte alimentar para aves em área urbana no sul do Brasil. *Biota Neotropica*, v. 7, n. 2, p. 307-312, 2007.

SCHAPOVAL, E.E.S.; SILVEIRA, S.M.; MIRANDA, M.L.; ALICE, C.B.; HENRIQUES, A.T. Evaluation of some pharmacological activities of *Eugeniuni unijlora* L. *Journal of Ethnopharmacology*. 44, p. 137-142. 1994.

SHARPE, R.H.; SHERMAN, W.B.; MILLER, E.P. Feijoa history and improvement. *Proceedings of the Florida State Horticultural Society*, Winter Haven, v.106, p.134-139, 1993.

SHINOZAKI, K.; OHME, M.; TANAKA, M. et al. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J.* 5:2043-2049. 1986.

SOBRAL, M. A Família das Myrtaceae no Rio Grande do Sul. São Leopoldo: Ed. Unisinos. 2003. 215p.

SOBRAL, M.; JARENKOW, J. A. **Flora Arbórea e Arborescente do Rio Grande do Sul, Brasil**. São Carlos: RiMa: Novo Ambiente, 2006. p. 130. 350 p.

SOBRAL, M.; PROENÇA, C.; SOUZA, M.; MAZINE, F.; LUCAS, E. 2015 Myrtaceae in **Lista de Espécies da Flora do Brasil**. Jardim Botânico do Rio de Janeiro. Disponível em: <<http://floradobrasil.jbrj.gov.br/jabot/floradobrasil/FB171>>. Acesso em: 29 Dez. 2016

SOUZA, V.C.; LORENZI, H. Botânica sistemática: Guia ilustrado para identificação das famílias de Angiospermas da flora brasileira, baseado em APG II. Plantarum, Nova Odessa. 2005.

STEANE, D.A. Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum *Eucalyptus globulus* (Myrtaceae). **DNA Res.**12(3):215–220. 2005.

SUGIURA, M. The chloroplast chromosomes in land plants. **Annu. Rev. Cell. Biol.** 5:51-70. 1989.

SYTSMA, K.J.; LITT, A.; ZJHRA, M.L.; PIRES, J.C.; NEPOKROEFF, M.; CONTI, E.; WALKER, J.; WILSON PG. Clades, clocks, and continents: historical and biogeographical analysis of Myrtaceae, Vochysiaceae, and relatives in the southern hemisphere. **Int J Plant Sci** 165: p.85–105, 2004.

TABARELLI, M.; MANTOVANI, W. A riqueza de espécies arbóreas na floresta atlântica de encosta no estado de São Paulo (Brasil). **Revista Brasileira de Botânica.** 22:217-223. 1999.

TANGPHATSORNRUANG, S.; SANGSRAKRU, D.; CHANPRASERT, J.; UTHAIPAISANWONG, P. YOOCHA, T.; JOMCHAI, N.; TRAGOONRUNG, S. The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships. **DNA Res** 17:11–22. 2010.

THORP, G.; BIELESKI, R. **Feijoas**: origins, cultivation and uses. Auckland: David Bateman, 2002. 87p.

TURCHETTO-ZOLET, A.C.; SALGUEIRO, F.; CRUZ, F.; VETO, N.; MARGIS, R. Chloroplast DNA variation and phylogeography of *Eugenia uniflora* L. (Myrtaceae) in the Brazilian Atlantic forest. **BMC Proceedings**, v.5 (Suppl 7), p.19, 2011.

TURCHETTO-ZOLET, A.C.; SALGUEIRO, F.; TURCHETTO, C.; CRUZ, F.; VETO, N.M.; BARROS, M. J. F.; SEGATTO, A.L.A. FREITAS, L.B.; MARGIS, R. Phylogeography and ecological niche modelling in *Eugenia uniflora* (Myrtaceae) suggest distinct vegetational

responses to climate change between the southern and the northern Atlantic Forest. **Botanical Journal of the Linnean Society**, 2016.

WALKER, D.; AVISE, J.C. Principles of phylogeography as illustrated by freshwater and terrestrial turtles in the southeastern United States. **Annual Review of Ecology and Systematics**. 29: 23-58. 1998.

WATSON, L.; DALLWITZ, M.J. **The families of flowering plants: descriptions, illustrations, identification, and information retrieval.** <http://delta-intkey.com> Version: 1st June. 2007.

WELTER, L.J.; BELÓ, A.; DUCROQUET, J.P.; GUERRA, M.P.; NODARI, R.O. Genetic characterization of the goiabeira-serrana (*Feijoa sellowiana* Berg) germoplasm. in: Congresso Brasileiro de Genética, 45, Gramado (RS). Anais Ribeirão Preto, SP, Anais, Revista Brasileira de Genética, v. 22, n. 3, p. 301, 1999.

WILSON P.G.; O'BRIEN, M.M.; GADEK, P.A.; QUINN; C.J. Myrtaceae revisited: a reassessment of infrafamilial groups. *Amer. J. Bot.* 88: 2013–2025, 2001.

WILSON, P.G.; O'BRIEN, M.M.; HELSEWOOD, M.M.; QUINN, C.J. Relationships within Myrtaceae sensu lato based on *matK* phylogeny. **Plant Systematics and Evolution**, v.251, p.3-19, 2005.

WOODSON, J.D.; CHORY, J. Coordination of gene expression between organellar and nuclear genomes. **Nature Reviews**. Vol.9. 2008.

WOLFE, K.H.; LI, W.H.; SHARP, P.M. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear DNAs. **Proc. Natl. Acad. Sci. USA** 84:9054-9058. 1987.

WU, C.S.; WANG, Y.N.; LIU, S.M. Chloroplast Genome (*cpDNA*) of *Cycas taitungensis* and 56 cpProtein-Coding Genes of *Gnetum parvifolium*: Insights into cp DNA Evolution and Phylogeny of Extant Seed Plants. **Mol. Biol. Evol.** 24:1366-1379. 2007.

THORNHILL, A.H.; HO, S.Y.W.; KÜLHEIM, C.; CRISP, M.D. Interpreting the modern distribution of Myrtaceae using a dated molecular phylogeny. **Mol Phylogenet Evol.** 93:29–43. doi: 10.1016/j.ympev.2015.07.007. 2015.

THORNHILL, A.H.; POPPLE, L.W.; CARTER, R.J.; et al. Are pollen fossils useful for calibrating relaxed molecular clock dating of phylogenies? A comparative study using Myrtaceae. **Mol Phylogenet Evol.**63:15–27. doi: 10.1016/j.ympev.2011.12.003. 2012.

VUOTTO, M.L.; BASILE, A.; MOSCATIELLO, V.; DE SOLE, P.; CASTALDO-COBIANCHI, R.; LAGHI, E.; IELPO, M.T.L. Antimicrobial and antioxidant activities of Feijoa sellowiana fruit. **International Journal of Antimicrobial Agents**, v. 13, p. 197-201, 2000.

YUKAWA, M.; TSUDZUKI, T.; SUGIURA, M. The 2005 Version of the Chloroplast DNA Sequence from Tobacco (*Nicotiana tabacum*). *Plant Mol. Biol. Rep.* 23:359–365. 2005.

ZHANG, Z.; GREEN, B.R.; CAVALIER-SMITH, T. Single gene circles in dinoflagellate chloroplast genomes. **Nature.** 400:155–159. 1999.

CAPÍTULO 1

Manuscrito aceito pelo periódico *Genetica*

**Phylogenomic relationship of feijoa (*Acca sellowiana* (O.Berg)
Burret) with other Myrtaceae based on complete chloroplast
genome sequences**

Abstract

Given their distribution, importance, and richness, Myrtaceae species comprise a model system for studying the evolution of tropical plant diversity. In addition, chloroplast (cp) genome sequencing is an efficient tool for phylogenetic relationship studies. Feijoa (*Acca sellowiana* (O. Berg) Burret; CN: pineapple-guava) is a Myrtaceae species that occurs naturally in southern Brazil and northern Uruguay. Feijoa is known for its exquisite perfume and flavorful fruits, pharmacological properties, ornamental value and increasing economic relevance. In the present work, we reported the complete cp genome of feijoa. The feijoa cp genome is a circular molecule of 159,370 bp with a quadripartite structure containing two single copy regions, a Large Single Copy region (LSC, 88,028 bp) and a Small Single Copy region (SSC, 18,598 bp) separated by Inverted Repeat regions (IRs, 26,372 bp). The genome structure, gene order, GC content and codon usage are similar to those of typical angiosperm cp genomes. When compared to other cp genome sequences of Myrtaceae, feijoa showed closest relationship with pitanga (*Eugenia uniflora* L.). Furthermore, a comparison of pitanga synonymous (Ks) and nonsynonymous (Ka) substitution rates revealed extremely low values. Maximum Likelihood and Bayesian Inference analyses produced phylogenomic trees identical in topology. These trees supported monophyly of three Myrtoideae clades.

Keywords: Myrtaceae, feijoa, plastid, next-generation sequencing, phylogenomics

1. Introduction

Acca sellowiana (O. Berg) Burret (syn. *Feijoa sellowiana*) is a Myrtaceae species native to the Neotropical region of South America. This species is commonly known as feijoa, or pineapple-guava, in English, and Goiabeira-serrana, Goiabeira-do-mato, Goiabeira-da-serra or Feijoa in Brazil, its main center of diversity (Nodari et al. 2008). Both fruit and leaves have pharmacological potential for antibacterial activity (Basile et al. 1997), effects of acetonic fruit extracts on human phagocyte functions (Ielpo et al. 2000), antimicrobial and antioxidant activities of fruits (Vuotto et al. 2000), anti-cancer activities (Bontempo et al. 2007; Abe et al. 2000; Motohashi et al. 2000), and antioxidant activity of leaves, as well as dry and fresh fruit (Beyhan et al. 2010). In addition, the use of feijoa fruit extract in the treatment and prevention of rheumatoid arthritis and Type-2 diabetes was patented in the United States (2010157678) and several other countries.

Ecologically, Myrtaceae is the dominant plant family among several vegetation types in South America through a variety of ecotypes (Mori and Boom 1983; Pennington et al. 2009). However, little is known about the evolutionary dynamics of Myrtaceae species across geographical space and time. Given their distribution and species richness, Myrtaceae comprises a model system for studying the evolution of tropical plant diversity.

Phylogenetic analyses are central to systematics and evolution, providing insights into diversification patterns, as well as the origin of desirable species characteristics. A number of studies have provided insights into Myrtaceae phylogeny using nuclear ribosomal DNA and chloroplast (cp) markers (Wilson et al. 2001; Wilson et al. 2005; Lucas et al. 2007; Biffin et al. 2010; Thornhill et al. 2015; Berger et al. 2016), and phylogenetic analyses based on complete cp genome sequences have been performed mainly on *Eucalyptus* and related genera (Steane 2005; Bayly et al. 2013). The newly sequenced cp genome of feijoa improves the Myrtaceae taxonomic sampling. Feijoa belongs to the tribe Myrteae, which is the most species-rich tribe in the Myrtaceae family, containing 2,500 species (Thornhill et al. 2015). Myrteae is a reasonably young clade, with an onset of diversification beginning in the late Oligocene, about 30 million years ago. At the Oligocene-Miocene boundary, about 23 million years ago, this tribe colonized in South

America, which sparked lineage diversification at high rates (Biffin et al. 2010; Berger et al. 2016). Therefore, a better understanding of the tribe Myrteae is pivotal for unravelling the evolution of this important tropical group in South America. The tribe Myrteae and *Eucalyptus* (tribe Eucalypteae) represent two very divergent lineages in the Myrtaceae, possibly sharing a recent common ancestor as early as the Late Cretaceous (Biffin et al. 2010).

In this study, we sequenced and analyzed the complete cp sequence of feijoa. We then performed a cp phylogenomic analysis using the newly sequenced feijoa and seven previously sequenced Myrtaceae species. This is the first cp phylogenomic analysis that includes three tribes of Myrtaceae. Furthermore, we compared gene order, structure and evolutionary forces within Myrteae.

2. Methods

2.1 Plant material and cpDNA purification

For cpDNA isolation, fresh leaves obtained from a single feijoa tree (Type Brazil) were collected from among those cultivated in the Agricultural Science Center at the Federal University of Santa Catarina (UFSC), Brazil (27°34'95,0"S, 48°30'31,9"W). The related voucher specimen, namely Machado, O.L. 01 - FLOR0059178, is kept at the Herbarium, Department of Botany, UFSC. Chloroplasts and plastid DNA from young leaves were obtained according to Vieira et al. (2014) with the following modifications during cpDNA isolation: cp lyse was obtained by incubating the chloroplast pellet with 8 ml of DNA isolation buffer, 1.5 ml 20% SDS, 450 µl 2-mercaptoethanol and 50 µl proteinase K (10 mg/ml) in a centrifuge tube at 55°C for 4 h.

2.2. Chloroplast genome sequencing, assembling and annotating

A total of 1 ng of cpDNA was used to prepare sequencing libraries with the Nextera XT DNA Sample Prep Kit (Illumina Inc., San Diego, CA), according to the manufacturer's instructions. Libraries were sequenced using MiSeq Reagent Kit v3 (600 cycles) on the Illumina MiSeq Sequencer (Illumina Inc., San Diego, California, USA). The obtained paired-end reads (2 x 300 bp) were used for *de novo* assembly performed with CLC Genomics Workbench v8.0.1. The same software was used to estimate cp genome coverage. Initial annotation of the feijoa plastome was performed using Dual Organellar GenoMe Annotator (DOGMA) (Wyman et al. 2004). From this initial annotation, putative starts, stops, and intron positions were determined based on comparisons to homologous genes in other cp genomes. The tRNA genes were further verified by using tRNAscan-SE (Schattner et al. 2005). The circular cp genome map was drawn using OrganellarGenomeDRAW (OGDRAW) (Lohse et al. 2007). REPuter (Kurtz and Schleiermacher 1999) was used to identify and locate the Inverted Repeat regions (IRs) in feijoa plastome sequenced by forward versus reverse complement (palindromic) alignment. The minimal repeat size was set to 30 bp, and identity of repeats was $\geq 90\%$. The complete feijoa cp genome sequence was deposited in the GenBank database under accession number KX289887. We also used REPuter

(Kurtz and Schleiermacher 1999) to identify and locate LSC/IRb/SSC/IRa sizes and boundaries in eight other previously published cp genomes. The *rpl2* gene of *Eucalyptus grandis* (HM347959) was reannotated using DOGMA (Wyman et al. 2004). The Mauve version 2.4.0 (Darling et al. 2004) was used for genome comparisons because it identified matches into locally collinear blocks (LCBs) and allow to evaluate rearrangements, large insertions or deletions, and substantial sequence divergence.

2.3 Phylogenomic analysis

Phylogenomic analysis was performed with the cp genomes of the newly sequenced feijoa and eight previously published species of Myrtaceae. As an ingroup species, we arbitrarily chose one species from each Myrtaceae genus. The *Lagerstroemia fauriei* (Myrtales: Lythraceae; KT358807) cp genome was used as the outgroup. The IRa was omitted to prevent overrepresentation of the IR sequences. The cp genomes were aligned using MAFFT (Multiple Alignment using Fast Fourier Transform) (Kato and Standley 2013). Nucleotide positions that contained one or more gaps introduced by the alignments were omitted from the matrix. The substitution model was selected using jModelTest (Darriba et al. 2012) and 5 was used as the number of substitution schemes. The General Time Reversible model of substitution, incorporating invariant sites and a gamma distribution (GTR + I + G), was used in subsequent analyses. Maximum likelihood (ML) analysis was performed using RAxML v 7.2.8 (Stamatakis 2006) with 1000 non-parametric bootstrap replicates. MrBayes 3.2.2 (Ronquist and Huelsenbeck 2003) was used to perform a Bayesian inference (BI) analysis. The Markov chain Monte Carlo (MCMC) analysis was run for 2,000,000 generations, and the first 25% of trees was discarded as burn-in. Resulting trees were represented and edited using FigTree v1.4.1.

2.4 Synonymous (K_s) and nonsynonymous (K_a) substitution rates

The complete feijoa cp genome sequence was compared with the sequence of pitanga (*Eugenia uniflora*) (GenBank accession KR867678). We analyzed synonymous (K_s) and nonsynonymous (K_a) substitution rates and their ratio, K_a/K_s , using Model Averaging in the KaKs_Calculator program (Zhang et al. 2006). The protein-coding

genes were aligned using MUSCLE (MUltiple Sequence Comparison by Log-Expectation) (Edgar 2004) to identify synonymous and nonsynonymous substitution. The correlated genes of pitanga were used as reference in the alignment.

2.5 Variation of nucleotide substitutions/indels (S/I) ratio and transition/transversion (Ti/Tv) ratio

The ratio of nucleotide substitutions to indels (S/I) and patterns of transition (Ti) to transversion (Tv) ratio (Ti/Tv) were calculated between feijoa and pitanga cp genomes. The single nucleotide variations (SNPs) and insertions/deletions were calculated using CLC Genomics Workbench 8.0.1v with the following parameters: a) minimum coverage of 30x and b) variant frequency of 35% for SNPs and 20% for indels.

3. Results and Discussion

3.1 Feijoa plastome content and organization

The length of the feijoa cp genome is 159,370 bp, about 915 bp larger than the size of the pitanga cp genome. Previous phylogenetic analyses, based on 18S-26S rRNA internal transcribed spacers (*ITS*) and plastid regions *matK* and *ndhF*, revealed that pitanga is the closest related species to feijoa with a complete cp genome sequence available in GenBank (Wilson et al. 2005; Biffin et al. 2010; Thornhill et al. 2012; Thornhill et al. 2015). Analogous to most other angiosperms, the feijoa cp genome is circular with a quadripartite structure: a pair of Inverted Repeat regions (IRs) (26,272 bp) separated by the Small Single Copy region (SSC) (18,598 bp) and Large Single Copy region (LSC) (88,228 bp) (Figure 1). SSC and LSC length correspond to 11.7% and 55.2% of the cp genome, respectively (Table1). The Feijoa cp genome size is similar to the genome sizes of other Myrtaceae species (Table 1 and 2). Its overall GC content is 37.0%. This value is slightly higher in protein coding genes (37.6%) than in intergenic spacers (IGS) (34.4%), introns (35.4%), the LSC region (34.9%), and the SSC region (30.6%). In contrast, rRNA, tRNA and the IRs all showed high GC content values with 55.3%, 53.2% and 42.8%, respectively. The GC content of the Feijoa cp genome is similar to that of other species of Myrtales, whose

plastomes range from 37-39% overall GC content, with 35-37% in the LSC region, 30-35% in the SSC region, and 43% previously reported for the IR regions(Gu et al. 2016).

A total of 135 functional genes were predicted, including 115 distinct genes comprising 78 protein-coding genes, 30 transfer RNA (tRNA) genes and 4 ribosomal RNA (rRNA) genes (Table 3). All four rRNA genes, seven tRNA genes, seven protein-coding genes, and two pseudogenes are repeated in the IR. The identified genes are summarized in Figure 1 and Table 3.

Protein-coding genes, tRNAs and rRNAs represent, respectively, 50.6%, 1.7% and 5.7% of the feijoa cp genome (Table 1), and non-coding DNA, including intergenic spacers (IGSs) and introns, represent 42.0% of the cp genome (Table 1). These values are consistent with other Myrtales and comparable to other cp genomes of Myrtaceae (Gu et al. 2016).

Among the 135 genes present in the feijoa cp genome, 18 were intron-containing genes, including nine protein-coding genes with a single intron, two protein-coding with a double intron, six tRNA genes with a single intron and one trans-splicing gene (*rps12*) (Table 4). The *rps12* is a trans-spliced gene with the exon 1 located in the LSC region and the C-terminal exons 2 and 3 duplicated in the IR region. Of the intron-containing genes, 12 are located in the LSC region (*rps16*, *rpoC1*, *atpF*, *petB*, *petD*, *rpl16*, *ycf3*, *clpP*, *trnK*^{-UUU}, *trnG*^{-UCC}, *trnV*^{-UAC}, *trnL*^{-UAA}), one in the SSC (*ndhA*), and four in the IR region (*rpl2*, *ndhB*, *trnI*^{-GAU}, *trnA*^{-UGC}). Similar to other species of the Myrtaceae, the *rpl2* gene has two highly conserved exons in feijoa. However, feijoa is slightly different from *L. fauriei* (outgroup), which lost the intron of the *rpl2* gene (Gu et al. 2016).

Table 1. Summary of feijoa chloroplast genome characteristics

Characteristics of cp genome	<i>A. sellowiana</i>
cp Genome Size (bp)	159,370
LSC size in bp (%)	88,028 (55.2)
SSC size in bp (%)	18,598 (11.7)
IR length in bp	26,372
Size in bp (%) of PCG	80,715 (50.6)

Size in bp (%) of introns	16,387 (10.3)
Size in bp (%) of rRNA	9,042 (5.7)
Size in bp (%) of tRNA	2,735 (1.7)
Size in bp (%) of IGS	50,491 (31.7)
Different genes	115
Different PCG	78
Different tRNA genes	30
Different rRNA genes	4
Different genes duplicated by IR	20
Different genes with introns	18
Overall % GC content	37.0
% GC content in LSC	34.9
% GC content in SSC	30.6
% GC content in IR	42.8
% GC content in PCG	37.6
% GC content in introns	35.4
% GC content in IGS	34.4
% GC content in rRNA	55.3
% GC content in tRNA	53.2

cp = chloroplast. PCG=Protein-coding gene

Three supposed pseudogenes, highly conserved in flowering plants, were identified: *ψ-infA*, *ψ-ycf15* and *ψ-ycf68* (Raubeson et al. 2007). The first two pseudogenes were also previously reported in Myrtaceae species of *Eucalyptus sp.* and *Syzygium cumini* (Steane 2005; Paiva et al. 2011; Asif et al. 2013; Bayly et al. 2013) and are assumed to be nonfunctional on the basis of premature stop codons.

3.2 Comparative analyses with other Myrtaceae cp genomes

We compared the feijoa cp genome with the cp genome of eight other Myrtales (seven Myrtaceae and one Lythraceae species) (Table 2). The eight cp genomes vary in length from 152,440 bp (*L. fauriei*) to 160,373 bp (*S. cumini*). The cp genome of *L. fauriei* is the shortest because of a reduction in intergenic regions (Gu et al. 2016). The overall GC content among plastid genomes of the Myrtaceae species ranged from 36.7% to 37.0%. These percentages are consistent with most land plants (Ravi et al. 2008). The GC content in Myrtaceae and Lythraceae LSC ranged from 35% to 36%, respectively, while the SSC and IR showed equal values, ranging from 31% in Myrtaceae and 43% in Lythraceae.

The feijoa cp genome is similar in gene order to the eight other Myrtales species analyzed (Fig. 1, Supplementary Fig 1). Thus, Myrtaceae is highly conserved in cp genome content, gene order and genomic structure (Supplementary Fig 1), as proposed by Gu et al. (2016), who also analyzed species of Myrtaceae.

The cp genomes were rather conserved within Myrtales, and neither inversions nor translocations were detected among the analyzed species. The LSC/IRBb/SSC/IRa boundary regions were examined to compare four junctions (JLA, JLB, JSA, and JSB) among nine species of Myrtales and to evaluate the potential impact of evolutionary changes in the cp genome of feijoa. The IR lengths ranged from 25,792 bp to 26,409 bp, and the position of IRs boundaries varied for each species. In three out of nine analyzed plastid genomes (*A. ternata*, *S. quadrifida*, and *L. fauriei*), the boundary between the LSC and IR regions was located within the *rps19* gene, resulting in the formation of an *rps19* pseudogene. The longest *rps19* pseudogene in Myrtales (75 bp) was observed in *L. fauriei* (outgroup) (Fig. 2). In the other six species, the LSC comprises an intact *rps19* gene together with 2 bp (*E. grandis*), 3 bp (feijoa, pitanga), 6 bp (*S. cumini*), and 8 bp (*A. costata*, *C. maculata*) of non-coding region beyond the LSC-IRb border.

The IRa-LSC border in these six species is located in the intergenic spacer (IGS) between *rpl2* and *trnH*. The *trnH* gene in *A. ternata*, *S. quadrifida* and *L. fauriei* extends to the IRa by 5bp and 3bp, respectively, whereas the same gene for feijoa, pitanga, *A. costata* and *C. maculata*, *E. grandis*, and *S. cumini* is 53 bp, 44 bp, 9 bp, 9bp, 2 bp, and 56 bp, respectively, away from the IRa-LSC border. The boundary

of the SSC-IRb junction in Myrtales plastomes was located within the *ycf1* gene, also resulting in the formation of a *ψycf1* pseudogene, which varied in length between 1,046 bp and 2,251 bp. The *ndhF* gene at the IRb-SSC border extends 35 bp and 38 bp into the IRb region in *A. ternata* and *L. fauriei*, respectively.

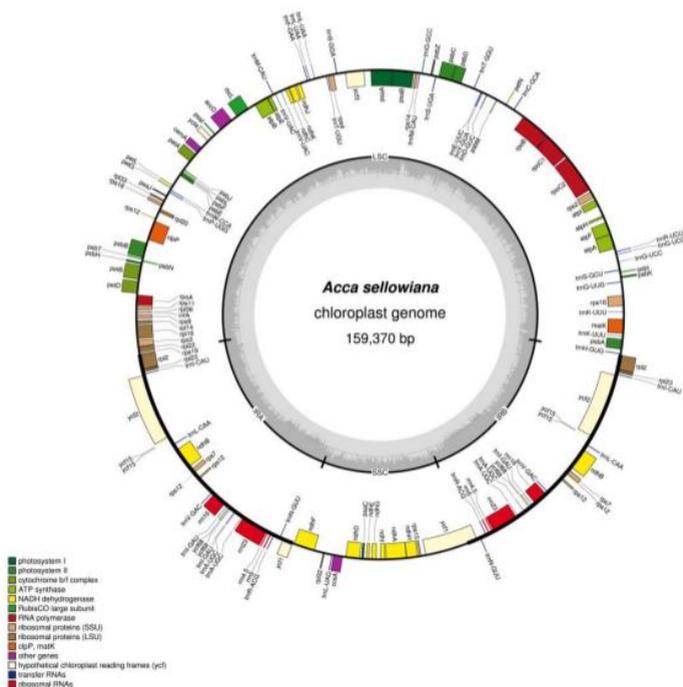


Figure 1. Gene map of *Acca sellowiana* (feijoa) chloroplast genome. Genes drawn inside the circle are transcribed clockwise, and genes drawn outside are transcribed counterclockwise. Genes belonging to different functional groups are color-coded. The darker gray in the inner circle corresponds to GC content, and the lighter gray to AT content.



Supplementary Fig. 1 Multiple genome alignment performed using Mauve software and the chloroplast genomes sequence of eight Myrtaceae family members. The presence of only one local colinear block (LCB), here represented on pink colour, indicates homologous DNA region without sequence rearrangements.

This gene is located in the SSC region in the seven other species and is separated from the IRb-SSC border by 10 to 225 bp. While other studies involving Myrtales species used the IRb-LSC boundary region to resolve phylogenetic relationships (Asif et al. 2013; Bayly et al. 2013; Gu et al. 2016), we observed low variation of IRb-SSC border within Myrtaceae species.

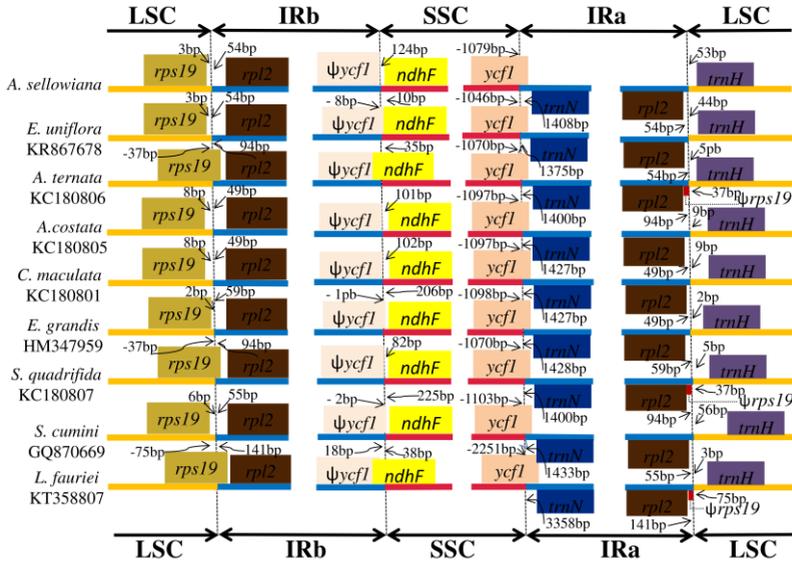


Figure 2. Comparison of boundary positions between single copy (large, LSC, or small, SSC) and inverted repeat (IR) regions among nine plastomes of Myrtales. Genes above lines are transcribed clockwise, while genes below the lines are transcribed counterclockwise. Ψ indicates a pseudogene.

Table 2. Comparison of chloroplast genomes of Myrtoideae (Myrtaceae) species and Lythraceae analyzed in this study

Family	Myrtoideae (Myrtaceae)										Lythraceae
Tribe	Myrtaceae	Myrtaceae	Eucalyptaceae	Eucalyptaceae	Eucalyptaceae	Eucalyptaceae	Eucalyptaceae	Eucalyptaceae	Syzygiaceae	Outgroup	
Species	<i>A.ellowianza</i> ^a	<i>E. uniflora</i>	<i>A. tenuia</i>	<i>A. costata</i>	<i>C. maculata</i>	<i>E. grandis</i>	<i>S. quadrifida</i>	<i>S. cuminii</i>	<i>L. furetei</i>		
GenBank accession	KX289887	KR867678	KC180806	KC180805	KC180801	HMA47959	KC180807	Q0870669	KI358807		
Plastome (bp)	159,370	158,445	159,593	160,326	160,045	160,137	159,561	160,373	152,440		
Length (%)	37.0	37.0	36.8	36.7	36.8	36.9	36.9	36.8	37.6		
LSC^b Region											
Length (bp)	88,028	87,459	88,218	88,769	88,557	88,872	88,247	89,081	83,923		
GC content (%)	34.9	34.8	34.6	34.5	34.6	34.8	34.7	34.7	35.9		
Length percentage (%) ^c	55.2	55.2	55.3	55.4	55.3	55.5	55.3	55.5	55.1		
SSC^c Region											
Length (bp)	18,598	18,318	18,571	18,773	18,670	18,475	18,544	18,508	16,933		
GC content (%)	30.6	30.8	30.5	30.5	30.5	30.5	30.7	30.6	30.8		
Length percentage (%) ^c	11.7	11.6	11.6	11.7	11.7	11.5	11.6	11.5	25.1		
IR^d Region											
Length (bp)	26,372	26,334	26,402	26,392	26,409	26,395	26,385	26,392	25,792		
GC content (%)	42.8	42.8	42.7	42.7	42.7	42.7	42.7	42.7	42.1		
Length percentage (%) ^e	33.1	33.2	33.1	32.9	33.0	33.0	33.1	33.0	19.8		

^a Species with chloroplast genome sequenced in this study; ^b Large Single Copy Region; ^c Short Single Copy Region; ^d Inverted Repeat Region; ^e Percent of entire chromosome

Table 3. List of genes identified in feijoa chloroplast genome

Category of genes	Group of genes		Name of genes							
Self - replication	Large subunit of ribosome	<i>rpl14</i>	<i>rpl16</i>	<i>rpl2</i> ^{ab}	<i>rpl20</i>	<i>rpl22</i>	<i>rpl23</i> ^b	<i>rpl32</i>	<i>rpl33</i>	<i>rpl36</i>
		DNA-dependent RNA polymerase	<i>rpoA</i>	<i>rpoB</i>	<i>rpoC1</i> ^a	<i>rpoC2</i>				
	Small subunit of ribosome	<i>rps11</i>	<i>rps12</i> ^{b,*}	<i>rps14</i>	<i>rps15</i>	<i>rps16</i> ^a	<i>rps18</i>	<i>rps19</i>	<i>rps2</i>	<i>rps3</i>
			<i>rps4</i>	<i>rps7</i> ^b	<i>rps8</i>					
	Ribosomal RNA Genes	<i>rnl6</i> ^b	<i>rnl23</i> ^b	<i>rnl4</i> ^b	<i>rnl5</i> ^b					
		<i>trnA-UGC</i> ^{ab}	<i>trnC-GCA</i>	<i>trnD-GUC</i>	<i>trnE-UUC</i>	<i>trnF-GAA</i>	<i>trnM-CAU</i>	<i>trnG-GCC</i>	<i>trnG-UCC</i> ^a	<i>trnH-GUG</i>
	Transfer RNA Genes	<i>trnL-CAU</i> ^b	<i>trnL-GAU</i> ^{ab}	<i>trnK-UUU</i> ^b	<i>trnL-CAA</i> ^a	<i>trnL-UAA</i> ^a	<i>trnL-UAG</i>	<i>trnM-CAU</i>	<i>trnV-GUU</i> ^b	<i>trnP-UGG</i>
		<i>trnQ-UUG</i>	<i>trnR-ACG</i> ^b	<i>trnR-UUCU</i>	<i>trnS-GCU</i>	<i>trnS-GGA</i>	<i>trnS-UGA</i>	<i>trnT-GGU</i>	<i>trnT-UGU</i>	<i>trnI-GAC</i> ^c
		<i>trnV-UAC</i> ^a	<i>trnW-CCA</i>	<i>trnI-GUA</i>						
	Translational initiation factor	<i>infA</i> ^c								
Genes for photosynthesis	Subunits of photosystem I	<i>psaA</i>	<i>psaB</i>	<i>psaC</i>	<i>psaI</i>	<i>psaJ</i>	<i>ycf3</i> ^a	<i>ycf4</i>		
	Subunits of photosystem II	<i>psbA</i>	<i>psbB</i>	<i>psbC</i>	<i>psbD</i>	<i>psbE</i>	<i>psbF</i>	<i>psbH</i>	<i>psbI</i>	<i>psbJ</i>
			<i>psbK</i>	<i>psbL</i>	<i>psbM</i>	<i>psbN</i>	<i>psbT</i>	<i>psbZ</i>		
	Subunits of cytochrome	<i>petA</i>	<i>petB</i> ^a	<i>petD</i> ^b	<i>petG</i>	<i>petL</i>	<i>petN</i>			
	Subunits of ATP synthase	<i>atpA</i>	<i>atpB</i>	<i>atpE</i>	<i>atpF</i> ^a	<i>atpH</i>	<i>atpI</i>			
	Large subunit of rubisco	<i>rbcL</i>								
	Subunits of NADH-dehydrogenase	<i>ndhA</i> ^a	<i>ndhB</i> ^b	<i>ndhC</i>	<i>ndhD</i>	<i>ndhE</i>	<i>ndhF</i>	<i>ndhG</i>	<i>ndhH</i>	<i>ndhI</i>
			<i>ndhJ</i>	<i>ndhK</i>						
	Other genes	Subunit of Acetyl-CoA-carboxylase	<i>accD</i>							
		c-type cytochrome synthesis gene	<i>ccsA</i>							
Envelope membrane protein		<i>cemA</i>								
Protease		<i>clpP</i> ^a								
Maturase		<i>matK</i>								
Component of TIC complex		<i>ycf1</i> ^{b,c*}								
Genes unknown function	Unknown function	<i>ycf2</i> ^b	<i>ycf15</i> ^{b,c}	<i>ycf68</i> ^{b,c}						

^a Intron-containing gene; ^b Duplicated genes; ^c Pseudogene; ^{c*} *ycf1* pseudogene at the boundary between IRB and SSC regions; ^{*} *rps12* is trans-spliced with the 5'-end located in the LSC region and the duplicated 3'-end in the IR region.

Table 4. Genes with introns in feijoa chloroplast genome and length of exons and introns

Region	Gene	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II(bp)	Exon III (bp)
LSC	<i>rps16</i>	206	875	38		
LSC	<i>rpoC1</i>	1616	728	431		
LSC	<i>atpF</i>	410	750	143		
LSC	<i>petB</i>	5	780	641		
LSC	<i>petD</i>	8	752	473		
LSC	<i>rpl16</i>	398	1006	8		
LSC	<i>ycf3</i>	152	724	227	767	125
LSC	<i>clpP</i>	227	617	290	861	68
LSC	<i>trnK-UUU</i>	34	2527	36		
LSC	<i>trnG-UCC</i>	22	739	48		
LSC	<i>trnV-UAC</i>	36	596	38		
LSC	<i>trnL-UAA</i>	36	503	49		
SSC	<i>ndhA</i>	539	1062	551		
LSC/I Rs	<i>rps12</i> *	113	–	209	–	26
IR	<i>rpl2</i>	434	664	389		

IR	<i>ndhB</i>	755	681	776
	<i>trnI-</i>			
IR	<i>GAU</i>	41	952	34
	<i>trnA-</i>			
IR	<i>UGC</i>	37	803	34

**rps12* is trans-spliced with the 5'-end located in the LSC region and the duplicated 3'-end in the IR region.

3.3 Evolutionary comparison between feijoa and pitanga

Two close species of Myrtaceae [feijoa and pitanga (KR867678)] were compared by phylogenetic analysis. The respective lengths of feijoa and pitanga were 159,370 bp and 158,445 bp. Both species contained 78 protein-coding genes, four rRNAs and 30 tRNAs. Comparison between these species was based on single nucleotide polymorphisms (SNPs) and insertion/deletion (indel). We detected 95 variations, including 85 putative SNPs and 10 indels (Table 5). In the coding region, a slightly higher degree of variation was found, with 41 putative SNPs and three putative indels, whereas in noncoding regions, 44 putative SNPs and six putative indels were found. These values are consistent with those of previous studies (Chen et al. 2009; Doorduyn et al. 2011; Bayly et al. 2013), who found fewer indels in coding regions compared to noncoding regions, in addition to high variation in nucleotide polymorphisms (Bayly et al. 2013). The S/I ratio is 13.7 in coding regions and 5.6 in noncoding regions. These results agree with those of Chen et al. (2009), who found the S/I ratio in coding regions to be higher than the S/I ratio in noncoding regions, indicating a universal constant by the relative occurrence of the two types of mutations. The frequency of SNPs/Kb in feijoa and pitanga cp genomes was 0.53 SNP/Kb (Table 6), corroborating the results of Tang et al. (2004), who detected 0.5 SNP/Kb in the cp genomes.

Table 5. SNP and indel variation between feijoa and pitanga plastomes

Variation type	Region of genome	feijoa/pitanga
Single Nucleotide Polymorphism (SNP)	Protein-coding gene	41
	Intron	7
	tRNA gene	5
	rRNA gene	0
	Intergenic region (IGS)	32
Total of SNPs		85
Insertion/Deletion	Protein-coding gene	3
	Intron	1
	tRNA gene	0
	rRNA gene	0
	Intergenic region (IGS)	6
Total of indels		10

Table 6. Frequency of SNPs and frequency of transition and transversion substitutions in feijoa compared to pitanga cp genomes

feijoa/pitanga	
Total of SNPs	85
Frequency of SNPs/Kb	0.53
Transition substitutions	
A <-> G (purine <-> purine)	23 (27.0%)
C <-> T (pyrimidine <-> pyrimidine)	26 (30.6%)
Total transitions (Ti)	49 (57.6%)
Frequency of total Ti/Kb	0.30
Transversion substitutions	
A <-> T	0 (0%)
A <-> C	13 (15.3%)
G <-> C	4 (4.7%)
G <-> T	19 (22.4%)
Total transversions (Tv)	36 (42.4%)
Frequency of total Tv/Kb	0.23
Ti/Tv	1.36

Compared to feijoa, the pitanga plastid genome showed a total of 49 transition substitutions (57.6%), corresponding to 0.30 Ti/Kb, and a total of 36 transversions (42.4%), corresponding to 0.23 Tv/Kb. The typical ti/tv ratio shows transition substitutions ranging from 2 to 10 (Purvis and Bromham 1997; Bakker et al. 1998; Yang and Yoder 1999;

Chen et al. 2009), whereas higher rates of transversions result in ti/tv ratios of about 1 (Hillis et al. 1993). In feijoa compared to pitanga cp genome, the ti/tv ratio was 1.36 (Table 6). The divergence between the two analyzed species can be explained by the decreased number of transitions relative to transversions (Yang 2006). Thus, this relatively high transversion rate suggests a divergence between feijoa and pitanga over evolutionary time.

Synonymous substitution does not change the amino acid within a peptide chain, whereas nonsynonymous substitution does (Page and Holmes 1998). Comparison of the substitution rates of each gene of the two closely related species was used to determine evolutionary forces. The Ka/Ks ratio was calculated for 78 protein-coding genes in common between feijoa and pitanga cp genomes (Table 7). Genes with Ka/Ks values of 50 and NA (not applicable or infinite) were changed to 0 in Table 7. According to Redwan et al. (2015) and Wang et al. (2011) this happens when the Ks value is extremely low (ranging from 0 to 0.0004) or when there were no substitution in the alignment, respectively. In this study, the *petL* gene had the highest synonymous rate, 0.04284, while the *rpl22* gene associated with large subunits of ribosome had the highest nonsynonymous rate, 0.02438. No changes were observed in the synonymous or nonsynonymous rate in *rps7*, *rpl23*, *rpl33*, *rpl36*, *rpoC1*, *atpF*, *atpH*, *atpI*, *ndhB*, *ndhE*, *petG*, *petN*, *psaJ*, *psbE*, *psbF*, *psbH*, *psbJ*, *psbL*, *psbM*, *psbN*, *psbT* and *psbZ* genes. The genes related to cytochrome b/f complex had the highest synonymous rate, with an average of 0.01444, while those related to the ATP synthase gene showed the lowest average rate of 0.00003. Genes with unknown functions (*ycf1* and *ycf2*) showed the highest nonsynonymous average rate of 0.00815, while those associated with the cytochrome b/f complex contained the lowest average value of 0.00044.

The Ka/Ks ratio indicates positive selection for eight genes analyzed, one gene associated with large subunit of ribosome (*rpl122*), one gene of RNA polymerase subunit (*rpoC2*), three genes associated with NADH dehydrogenase (*ndhD*, *ndhF*, *ndhH*), one gene associated with photosystem I (*psaA*), one gene with unknown functions (*ycf1*) and one other gene (*matK*). The Ka/Ks ratio indicated purifying selection for 27 genes. These genes are associated with a small subunit of ribosome

(*rps12*, *rps16*, *rps18*, *rps3*, *rps8*), large subunit of ribosome (*rpl2*, *rpl32*), RNA polymerase subunits (*rpoA*, *rpoB*), NADH dehydrogenase (*ndhA*, *ndhC*, *ndhG*, *ndhK*), photosystem I (*psaB*, *psaI*), photosystem II (*psbB*, *psbC*), unknown and other functions (*ycf2*, *ycf4*, *accD*, *ccsA*, *cemA*), cytochrome b/f complex (*petA*, *petB*, *petD*, *petL*), and large chain of rubisco (*rbcL*). Furthermore, the Ka values and Ka/Ks ratios in genes for photosynthesis are low, suggesting that plastid genes of photosynthetic plants are under strong purifying selection (Guisinger et al. 2008; Wang et al. 2008). However, the weighted average values of Ka/Ks ratios were lower in the IR (0.074) region than in the SSC and LSC regions (1.379 and 0.190, respectively), corroborating the findings of Perry and Wolfe (2002).

Table 7. Comparison of nonsynonymous (Ka) and synonymous (Ks) substitution rates and Ka/Ks ratio between feijoa and pitanga

Gene	feijoa/pitanga		
	Ka	Ks	Ka/Ks
Small subunit of ribosome			
<i>rps2</i>	0.01088	0.00022	0
<i>rps3</i>	0.00402	0.02491	0.161
<i>rps4</i>	0.00276	0.00006	0
<i>rps7</i>	NA	NA	0
<i>rps8</i>	0.00001	0.01403	0.001
<i>rps11</i>	0.01651	0.00033	0
<i>rps12</i>	0.00001	0.01502	0.001
<i>rps14</i>	0.01028	0.00021	0
<i>rps15</i>	0.00547	0.00011	0
<i>rps16</i>	0.00002	0.01819	0.001
<i>rps18</i>	0.00002	0.01709	0.001
<i>rps19</i>	0.00459	0.00009	0

Average	0.00455	0.00752	0
Large subunit of ribosome			
<i>rpl2</i>	0.00163	0.01154	0.141
<i>rpl14</i>	0.00649	0.00013	0
<i>rpl16</i>	0.00647	0.00013	0
<i>rpl20</i>	0.00423	0.00008	0
<i>rpl22</i>	0.02438	0.01711	1.425
<i>rpl23</i>	NA	NA	0
<i>rpl32</i>	0.00003	0.03397	0.001
<i>rpl33</i>	NA	NA	0
<i>rpl36</i>	0	0	0
Average	0.00480	0.00699	0
RNA polymerase subunits			
<i>rpoA</i>	0.00525	0.00553	0.950
<i>rpoB</i>	0.00547	0.00677	0.808
<i>rpoC1</i>	NA	NA	0
<i>rpoC2</i>	0.00536	0.00502	1.067
Average	0.00402	0.00433	0.707
Total average	0.00446	0.006284	0.235
ATP synthase gene			
<i>atpA</i>	0.00246	0.00005	0
<i>atpB</i>	0.00169	0.00003	0
<i>atpE</i>	0.00430	0.00008	0
<i>atpF</i>	NA	NA	0
<i>atpH</i>	NA	NA	0
<i>atpI</i>	NA	NA	0

Average	0.00141	0.00003	0
NADH dehydrogenase			
<i>ndhA</i>	0.00259	0.00541	0.480
<i>ndhB</i>	NA	NA	0
<i>ndhC</i>	0.00003	0.03187	0.001
<i>ndhD</i>	0.00785	0.00252	3.118
<i>ndhE</i>	NA	NA	0
<i>ndhF</i>	0.00933	0.00201	4.655
<i>ndhG</i>	0.00527	0.02271	0.232
<i>ndhH</i>	0.00757	0.00260	2.908
<i>ndhI</i>	0.01035	0.00021	0
<i>ndhJ</i>	0.00341	0.00007	0
<i>ndhK</i>	0.00365	0.03198	0.114
Average	0.00455	0.00903	1.046
Cytochrome b/f complex			
<i>petA</i>	0.00256	0.01271	0.201
<i>petB</i>	0.00001	0.01449	0.001
<i>petD</i>	0.00002	0.01659	0.001
<i>petG</i>	NA	NA	0
<i>petL</i>	0.00004	0.04284	0.001
<i>petN</i>	NA	NA	0
Average	0.00044	0.01444	0.034
Photosystem I			
<i>psaA</i>	0.00368	0.00257	1.429
<i>psaB</i>	0.00289	0.00323	0.892
<i>psaC</i>	0.02036	0.00041	0

<i>psaI</i>	0.00003	0.03468	0.001
<i>psaJ</i>	NA	NA	0
<i>ycf3</i>	0.00540	0.00011	0
<i>ycf4</i>	0.01007	0.01558	0.649
Average	0.00606	0.00808	0.424

Photosystem II

<i>psbA</i>	0.00301	0.00006	0
<i>psbB</i>	0.00089	0.01122	0.080
<i>psbC</i>	0.00003	0.02506	0.001
<i>psbD</i>	0.00121	0.00002	0
<i>psbE</i>	NA	NA	0
<i>psbF</i>	NA	NA	0
<i>psbH</i>	NA	NA	0
<i>psbI</i>	0.00852	0.00017	0
<i>psbJ</i>	NA	NA	0
<i>psbK</i>	0.00852	0.00017	0
<i>psbL</i>	NA	NA	0
<i>psbM</i>	NA	NA	0
<i>psbN</i>	NA	NA	0
<i>psbT</i>	NA	NA	0
<i>psbZ</i>	NA	NA	0
Average	0.00148	0.00245	0.005

Large chain of rubisco

<i>rbcL</i>	0.00529	0.00735	0.720
Total average	0.00321	0.00690	0.372

Other genes

<i>clpP</i>	0.00275	0.00005	0
<i>ccsA</i>	0.00858	0.03160	0.271
<i>accD</i>	0.00529	0.00735	0.720
<i>cemA</i>	0.00362	0.00809	0.447
<i>matK</i>	0.01771	0.01011	1.752
Total average	0.00759	0.01144	0.638
Unknown functions			
<i>ycf1</i>	0.01523	0.00719	2.120
<i>ycf2</i>	0.00107	0.00350	0.304
Total average	0.00815	0.00534	1.212

3.4 Phylogenomic analysis of Myrtaceae

The alignment matrix of cp genomes of the nine Myrtales species, after exclusion of gaps, was 118,672 sites, whereas the original alignment matrix including gaps was 141,515 nucleotide positions in length Bayesian analysis ($\ln L = -237\ 706.9$) of the full dataset recovered the same tree topology as that produced by ML analysis ($\ln L = -237\ 698.3$) (Fig. 3). The ML bootstrap supports (BS) and the BI posterior probabilities (PP) showed maximum support for all nodes, values were 100% and 1.0, respectively (Fig. 3).

In this work, we present, for the first time, a phylogenetic tree based on complete cp genome sequences involving eight species of Myrtaceae. When substitution rates were compared between plastid genomes, these rates were found to be much lower than those of nuclear genomes, suggesting that plastid genomes are evolutionarily conserved (Palmer 1985). In addition, different chloroplast regions had contrasting substitution rates, providing a useful tool for the understanding of evolutionary relationships (Palmer 1985; Shaw et al. 2005; Shaw et al.

2007). The phylogenetic tree clearly supports the monophyly of the three Myrtoideae tribes: Myrteae, Eucalypteae, Syzygieae.

This result corroborates previous studies based on plastidial genes such as *rbcL*, *ndhF* and *matK*, and nuclear sequences such as 18S–26S rRNA and *ITS* (Sytsma et al. 2004; Wilson et al. 2005; Biffin et al. 2010; Thornhill et al. 2012; Thornhill et al. 2015).

As an outgroup, we used one Australasian taxon, *L. fauriei*. The Myrteae clade of feijoa (*A. sellowiana*) + pitanga (*E. unflora*) is strongly supported as monophyletic (BS = 100%, PP = 1.0; node 2; Fig. 3) and sister of the equally strongly supported clade of Syzygieae + Eucalypteae (BS = 100%, PP = 1.0; node 3). These results are in agreement with Biffin et al. (2010), who concluded that Syzygieae and Myrteae show highly significant positive variation in diversification rates associated with both of these lineages relative to the overall evolutionary radiation of Myrtaceae. The Eucalypteae tribe was divided into two monophyletic clades with maximum support: *E. grandis*+ *A. costata* + *C. maculata* (node 6) and *A. ternata* + *S. quadrifida* (node 7). Previous analyses based on cpDNA also reported this relationship (Bayly et al. 2013).

4. Conclusion

In this work, we provide the complete cp genome of feijoa. Genome gene contents and arrangement are similar to those found in the cp genome of other Myrtoideae (Myrtaceae) species. The IRb-LSC boundary region is similar among the nine species analyzed. Between feijoa and pitanga, two closely related species of the tribe Myrteae, a difference of 85 SNPs and 10 indels was recorded. For both of these species, the S/I ratio was higher in coding regions compared to noncoding regions. In addition, the ti/tv ratio was slightly lower than that of other land plant plastomes, suggesting a divergence between feijoa and pitanga over time. Moreover, the Ka/Ks ratio was lower in the IR compared to single copy regions. For the first time, we also developed a phylogenetic tree based on the complete plastid sequences of eight Myrtaceae species. The ML and BI trees strongly supported the monophyly of three tribes of Myrtoideae.

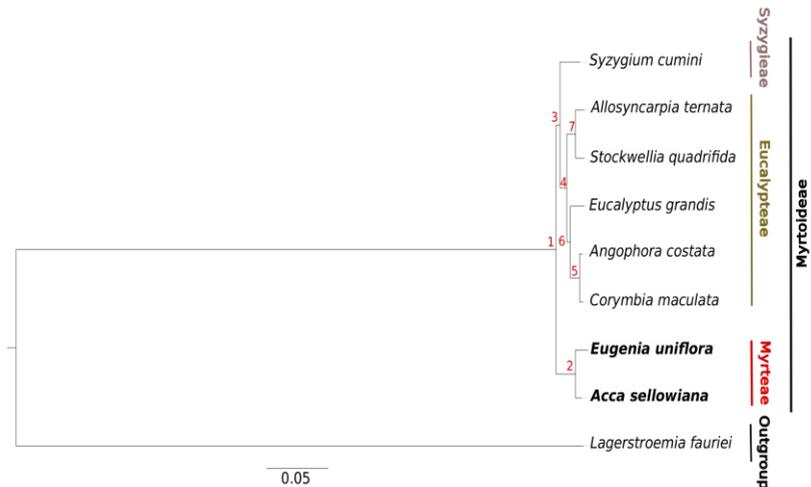


Figure 3. Bayesian phylogeny based on the complete plastome sequence of eight Myrtoideae (Myrtaceae) species and the outgroup *Lagerstroemia fauriei* (Myrtales: Lythraceae; KT358807). All nodes were supported with 100% bootstrap supports (BS)/1.0 posterior probabilities (PP). The branch length is proportional to the inferred divergence level. The scale bar indicates the number of inferred nucleic acid substitutions per site. *Acca sellowiana* and *Eugenia uniflora* corresponds to feijoa and pitanga cp genome, respectively.

References

- Abe I, Seki T, Umehara K, et al (2000) Green tea polyphenols: novel and potent inhibitors of squalene epoxidase. *Biochem Biophys Res Commun* 268:767–71. doi: 10.1006/bbrc.2000.2217
- Asif H, Khan A, Iqbal A, et al (2013) The chloroplast genome sequence of *Syzygium cumini* (L.) and its relationship with other angiosperms. *Tree Genet Genomes* 9:867–877. doi: 10.1007/s11295-013-0604-1
- Bakker FT, Culham A, Gomez-martinez R, et al (1998) Patterns of Nucleotide Substitution in Angiosperm cpDNA trnL (UAA)– trnF (GAA) Regions. 1146–1155.
- Basile A, Vuotto ML, Violante U, et al (1997) Antibacterial activity in *Actinidia chinensis*, *Feijoa sellowiana* and *Aberia caffra*. *Int J Antimicrob Agents* 8:199–203. doi: 10.1016/S0924-8579(97)00376-2
- Bayly MJ, Rigault P, Spokevicius A, et al (2013) Chloroplast genome analysis of Australian eucalypts - *Eucalyptus*, *Corymbia*, *Angophora*, *Allosyncarpia* and *Stockwellia* (Myrtaceae). *Mol Phylogenet Evol* 69:704–716. doi: 10.1016/j.ympev.2013.07.006
- Berger BA, Kriebel R, Spalink D, Sytsma KJ (2016) Divergence times, historical biogeography, and shifts in speciation rates of Myrtales. *Mol Phylogenet Evol* 95:116–136. doi: 10.1016/j.ympev.2015.10.001
- Beyhan Ö, Elmastaş M, Gedikli F (2010) Total phenolic compounds and antioxidant capacity of leaf, dry fruit and fresh fruit of feijoa (*Acca sellowiana*, Myrtaceae). *J Med Plants Res* 4:1065–1072.
- Biffin E, Lucas EJ, Craven LA, et al (2010) Evolution of exceptional

- species richness among lineages of fleshy-fruited Myrtaceae. *Ann Bot* 106:79–93. doi: 10.1093/aob/mcq088
- Bontempo P, Mita L, Miceli M, et al (2007) Feijoa sellowiana derived natural Flavone exerts anti-cancer action displaying HDAC inhibitory activities. *Int J Biochem Cell Biol* 39:1902–1914. doi: 10.1016/j.biocel.2007.05.010
- Chen JQ, Wu Y, Yang H, et al (2009) Variation in the ratio of nucleotide substitution and indel rates across genomes in mammals and bacteria. *Mol Biol Evol* 26:1523–1531. doi: 10.1093/molbev/msp063
- Darling ACE, Mau B, Blattner FR, Perna NT (2004) Mauve : Multiple Alignment of Conserved Genomic Sequence With Rearrangements
Mauve : Multiple Alignment of Conserved Genomic Sequence With Rearrangements. 1394–1403. doi: 10.1101/gr.2289704
- Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9:772–772. doi: 10.1038/nmeth.2109
- Doorduyn L, Gravendeel B, Lammers Y, et al (2011) The complete chloroplast genome of 17 individuals of pest species *Jacobaea vulgaris*: SNPs, microsatellites and barcoding markers for population and phylogenetic studies. *DNA Res* 18:93–105. doi: 10.1093/dnares/dsr002
- Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. doi: 10.1186/1471-2105-5-113
- Gu C, Tembrock LR, Johnson NG, et al (2016) The Complete plastid genome of *Lagerstroemia fauriei* and loss of *rpl2* intron from

- lagerstroemia (Lythraceae). PLoS One 11:1–18. doi: 10.1371/journal.pone.0150752
- Guisinger MM, Kuehl J V, Boore JL, Jansen RK (2008) Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. Proc Natl Acad Sci U S A 105:18424–18429. doi: 10.1073/pnas.0806759105
- Hillis DM, MW A, MM M (1993) Analysis of DNA Sequence Inference Data: Phylogenetic. Analysis 224:456–490.
- Ielpo MTL, Basile A, Miranda R, et al (2000) Immunopharmacological properties of flavonoids. Fitoterapia 71:101–109. doi: 10.1016/S0367-326X(00)00184-2
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol 30:772–780. doi: 10.1093/molbev/mst010
- Kurtz S, Schleiermacher C (1999) REPuter: Fast computation of maximal repeats in complete genomes. Bioinformatics 15:426–427. doi: 10.1093/bioinformatics/15.5.426
- Lohse M, Drechsel O, Bock R (2007) OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Curr Genet 52:267–274. doi: 10.1007/s00294-007-0161-y
- Lucas EJ, Harris SA, Mazine FF, et al (2007) Suprageneric phylogenetics of Myrteae, the generically richest tribe in Myrtaceae (Myrtales). Taxon 56:1105–1128. doi: 10.2307/25065906
- Mori SA, Boom BM (1983) Ecological importance of Myrtaceae in a Eastern Brazilian Wet Forest. 68–70.

- Motohashi N, Kawase M, Shirataki Y, et al (2000) Biological activity of Feijoa peel extracts. *Anticancer Res* 20:4323–4329.
- Nodari RO, Santos KL dos, Ducroquet JP, Guerra MP (2008) Goiabeira-serrana: domesticação. In: Barbieri RL, Stumpf ERT (eds) *Origem e evolução de plantas cultivadas*. p 914
- Page RDM, Holmes EC (1998) *Molecular Evolution: A Phylogenetic Approach*.
- Paiva J a P, Prat E, Vautrin S, et al (2011) Advancing Eucalyptus genomics: identification and sequencing of lignin biosynthesis genes from deep-coverage BAC libraries. *BMC Genomics* 12:137. doi: 10.1186/1471-2164-12-137
- Palmer JD (1985) Chloroplast DNA and molecular phylogeny. *BioEssays* 2:263–267. doi: 10.1002/bies.950020607
- Pennington RT, Lavin M, Oliveira A (2009) Woody Plant Diversity, Evolution, and Ecology in the Tropics: Perspectives from Seasonally Dry Tropical Forests. *Annu Rev Ecol Evol Syst* 40:437–457. doi: 10.1146/annurev.ecolsys.110308.120327
- Perry AS, Wolfe KH (2002) Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. *J Mol Evol* 55:501–508. doi: 10.1007/s00239-002-2333-y
- Purvis A, Bromham L (1997) Estimating the transition/transversion ratio from independent pairwise comparisons with an assumed phylogeny. *J Mol Evol* 44:112–119. doi: 10.1007/PL00006117
- Raubeson LA, Peery R, Chumley TW, et al (2007) Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics* 8:174.

doi: 10.1186/1471-2164-8-174

- Ravi V, Khurana JP, Tyagi AK, Khurana P (2008) An update on chloroplast genomes. *Plant Syst Evol* 271:101–122. doi: 10.1007/s00606-007-0608-0
- Redwan RM, Saidin a, Kumar S V (2015) Complete chloroplast genome sequence of MD-2 pineapple and its comparative analysis among nine other plants from the subclass Commelinidae. *BMC Plant Biol* 15:196. doi: 10.1186/s12870-015-0587-1
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574. doi: 10.1093/bioinformatics/btg180
- Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* 33:686–689. doi: 10.1093/nar/gki366
- Shaw J, Lickey EB, Beck JT, et al (2005) The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am J Bot* 92:142–166. doi: 10.3732/ajb.92.1.142
- Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The Tortoise and the hare III. *Am J Bot* 94:275–288. doi: 10.3732/ajb.94.3.275
- Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690. doi: 10.1093/bioinformatics/btl446
- Steane DA (2005) Complete nucleotide sequence of the chloroplast genome

- from the Tasmanian blue gum, *Eucalyptus globulus* (Myrtaceae).
DNA Res 12:215–220. doi: 10.1093/dnares/dsi006
- Sytsma KJ, Litt A, Zjhra ML, et al (2004) Clades , Clocks , and Continents :
Historical and Biogeographical Analysis of Myrtaceae , Vochysiaceae
, and Relatives in the Southern Hemisphere Source : International
Journal of Plant Sciences , Vol . 165 , No . S4 , Tropical
Intercontinental Disjunctio. Int J Plant Sci 165:S85–S105.
- Tang J, Xia H, Cao M, et al (2004) A Comparison of Rice Chloroplast
Genomes. Plant Physiol 135:412–20. doi: 10.1104/pp.103.031245
- Thornhill AH, Ho SYW, Külheim C, Crisp MD (2015) Interpreting the
modern distribution of Myrtaceae using a dated molecular phylogeny.
Mol Phylogenet Evol 93:29–43. doi: 10.1016/j.ympcv.2015.07.007
- Thornhill AH, Popple LW, Carter RJ, et al (2012) Are pollen fossils useful
for calibrating relaxed molecular clock dating of phylogenies? A
comparative study using Myrtaceae. Mol Phylogenet Evol 63:15–27.
doi: 10.1016/j.ympcv.2011.12.003
- Vieira LDN, Faoro H, De Freitas Fraga HP, et al (2014) An improved
protocol for intact chloroplasts and cpDNA isolation in conifers.
PLoS One 9:1–8. doi: 10.1371/journal.pone.0084792
- Vuotto ML, Basile A, Moscatiello V, et al (2000) Antimicrobial and
antioxidant activities of *Feijoa sellowiana* fruit. Int J Antimicrob
Agents 13:197–201. doi: 10.1016/S0924-8579(99)00122-3
- Wang D, Liu F, Wang L, et al (2011) Nonsynonymous substitution rate
(Ka) is a relatively consistent parameter for defining fast-evolving and
slow-evolving protein-coding genes. Biol Direct 6:13. doi:
10.1186/1745-6150-6-13

- Wang R-J, Cheng C-L, Chang C-C, et al (2008) Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol Biol* 8:36. doi: 10.1186/1471-2148-8-36
- Wilson PG, O'Brien MM, Gadek PA, Quinn CJ (2001) Myrtaceae revisited: A reassessment of infrafamilial groups. *Am J Bot* 88:2013–2025. doi: 10.2307/3558428
- Wilson PG, O'Brien MM, Heslewood MM, Quinn CJ (2005) Relationships within Myrtaceae sensu lato based on a matK phylogeny. *Plant Syst Evol* 251:3–19. doi: 10.1007/s00606-004-0162-y
- Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* (Oxford, England) 20:3252–3255.
- Yang Z (2006) *Computational molecular evolution*. Oxford Ser Ecol Evol xvi, 357 p. doi: 10.1093/acprof:oso/9780198567028.001.0001
- Yang Z, Yoder AD (1999) Estimation of the transition/transversion rate bias and species sampling. *J Mol Evol* 48:274–283. doi: 10.1007/PL00006470
- Zhang Z, Li J, Zhao XQ, et al (2006) KaKs_Calculator: Calculating Ka and Ks Through Model Selection and Model Averaging. *Genomics, Proteomics Bioinforma* 4:259–263. doi: 10.1016/S1672-0229(07)60007-2

CAPÍTULO 2

Manuscrito formatado de acordo com as normas do periódico *Plant Molecular Biology*

Phylogenetic inference of *Campomanesia xanthocarpa* (Mart.) O. Berg with eight genera of Myrtaceae based on giant *ycf2*

Abstract

Campomanesia xanthocarpa (Mart.) O. Berg is a representative of the Myrteae tribe for which the smallest chloroplast genome within Myrtaceae has been reported. This genome is 158,131 bp in length with 36.9% GC content and consists of two single-copy regions separated by a pair of 25,970 bp inverted repeats. The large single copy and the small single copy regions span 87,596 bp and 18,595 bp, respectively. The genome contains 135 genes, including 20 located in each inverted repeat. The large chloroplast gene *ycf2* is essential for cell survival. In *C. xanthocarpa*, *ycf2* encodes the protein product of 2294 amino acids. Phylogenetic analysis among *Campomanesia* and eight other genera sampled from Myrtaceae based on the large chloroplast *ycf2* gene showed it to be useful for evolutionary studies. In addition, the synonymous (*Ks*) and nonsynonymous (*Ka*) substitution rates, as well as *Ka/Ks* ratio, revealed extremely low values for the *ycf2* gene among *C. xanthocarpa* and another 44 chloroplast genomes of Myrtaceae species. Bayesian inference analyses produced phylogenetic trees, which supported monophyly of three Myrtaceae tribes: Myrteae, Syzygieae and Eucalypteae.

Keywords: Myrtaceae, *Campomanesia*, plastid, *ycf2*, phylogeny, evolution

1. Introduction

Chloroplast (cp) genomes are a useful tool to perform comparative analyses associated with phylogenetic and evolutionary studies. In land plants, cp genome size ranges from 130 to 160 kb. The genome contains two inverted repeat regions (IRs) divided between large (LSC) and small (SSC) single-copy regions (Bock 2007).

In addition, the cp genome contains essential genes in conserved ORFs. In higher plant chloroplast, some *orfs* have unknown function, and these are called *ycfs*, or hypothetical open reading frame, such as the giant open reading frame *ycf2*. The large *ycf2* cp gene specifies an expressed protein (Glick and Sears 1993) whose function remains unknown. However, the function of *ycf2* has been hypothesized to exhibit similarities with *fstH*, such as ATPase-related activities, chaperone function and activity associated with cell division (Wolfe 1994).

The elevated substitution rates of the *ycf2* gene led to a pseudogenization process (Downie et al. 1994; Oliver et al. 2010; Wolf et al. 2011). This gene is commonly absent in some cp genomes (Downie et al. 1994; Millen et al. 2001), especially monocot grasses, such as maize, rice, and sugarcane (Maier et al. 1995; Matsuoka et al. 2002; Asano et al. 2004). Thus, the high variability of *ycf2* makes it a potential candidate for species-level DNA barcoding (Kumar et al. 2009).

Studies involving natural biodiversity or species with significant economic potential, such as *Campomanesia xanthocarpa* (Mart.) O. Berg, are tied together by the framework of phylogenetics. Phylogeny inferences may shed light on diversification patterns and speciation, as well as the evolution of useful and desirable features. Complete cp genome sequences, as well as phylogenetic analyses based on these data, have been conducted on *Eucalyptus* and related genera (Steane 2005; Bayly et al. 2013; Bayly 2016). More recently, data including species of different tribes of Myrtaceae have been reported (MACHADO et al. 2017). In addition, sequencing the cp genome of *C.*

xanthocarpa, the newest member of Myrtaceae, will contribute to a better understanding of the evolutionary patterns of Myrtaceae.

Thus, we herein sequenced and analysed the complete cp genome of *C. xanthocarpa*. In addition, we performed a cp phylogenetic analysis based on the large cp *ycf2* gene using *C. xanthocarpa* and 44 previously sequenced Myrtaceae species in order to test the utility of this gene for phylogenetic inferences. Finally, we compared gene order, structure of cp genome and evolutionary characteristics of *ycf2* in Myrtaceae.

2. Methods

2.1 Plant material and cpDNA purification

For cpDNA isolation, fresh leaves obtained from a single *C. xanthocarpa* tree were collected in the Department of Botany, University of Santa Catarina (UFSC), Brazil (27° 36. 094” S, 48° 31. 310” W). The plastid DNA was obtained according to Vieira et al. (2014), with the following modifications during cpDNA isolation: cp lyse was obtained by incubating the chloroplast pellet with 8 ml of DNA isolation buffer, 1.5 ml 20% SDS, 450 µl 2-mercaptoethanol and 50 µl proteinase K (10 mg/ml) in a centrifuge tube at 55°C for 4 h or overnight.

2.2 Chloroplast genome assembly and annotation

The sequencing libraries were prepared using 1 ng of cpDNA with the NexteraXT DNA Sample Prep Kit (Illumina Inc., San Diego, CA), according to the manufacturer’s instructions. Libraries were sequenced using MiSeq Reagent Kit v3 (600 cycles) on the Illumina MiSeq Sequencer (Illumina Inc., San Diego, California, USA). The obtained paired-end reads (2 x 300 bp) were used for *de novo* assembly performed with CLC Genomics Workbench v8.0.1. The same software was used to estimate cp genome coverage. Initial annotation of the *C. xanthocarpa* plastome was performed using Dual Organellar GenoMe

Annotator (DOGMA) (Wyman et al. 2004). From this initial annotation, putative starts, stops, and intron positions were determined based on comparisons to homologous genes in other cp genomes. The tRNA genes were further verified by using tRNAscan-SE (Schattner et al. 2005). The circular cp genome map was drawn using OrganellarGenomeDRAW (OGDRAW) (Lohse et al. 2007). REPuter (Kurtz and Schleiermacher 1999) was used to identify and locate the IRs in the *C. xanthocarpa* plastome sequenced by forward versus reverse complement (palindromic) alignment. The minimal repeat size was set as 30 bp, and identity of repeats was $\geq 90\%$. The complete *C. xanthocarpa* cp genome sequence was deposited in the GenBank database under accession number KY392760. Mauve version 2.4.0 (Darling et al. 2004) was used for genome comparisons among nine genera of Myrtaceae to identify matches into locally collinear blocks (LCBs). In addition, this approach allows evaluating rearrangements, large insertions or deletions, and substantial sequence divergence.

2.3 Synonymous (K_s) and nonsynonymous (K_a) substitution rates

The sequence of the *C. xanthocarpa* *ycf2* gene was compared to that of 44 other Myrtaceae species (Table 1). The evolutionary characteristics, nonsynonymous (K_a) and synonymous substitution rates (K_s), as well as K_a/K_s ratio, were calculated using Model Averaging in the KaKs_Calculator program (Zhang et al. 2006). The genes were aligned in pairs using MULTiple Sequence Comparison by Log-Expectation (MUSCLE) (Edgar 2004) to identify synonymous and nonsynonymous substitution.

Table 1. Comparison of chloroplast genomes of Myrtaceae species and outgroup analyzed in this study.

Species	Accession	Size	LSC ^b	SSC ^c	IR ^d
<i>Acca sellowiana</i>	KX289887	159,370	88,028	18,598	26,372
<i>Allosyncarpia ternata</i>	KC180806	159,593	88,218	18,571	26,402
<i>Angophora costata</i>	KC180805	160,326	88,769	18,773	26,392

<i>Angophora floribunda</i>	KC180804	160,245	88,715	18,746	26,392
<i>Campomanesia xanthocarpa</i> ^a	KY392760	158,131	87,596	18,595	25,970
<i>Corymbia eximia</i>	KC180802	160,012	88,522	18,672	26,409
<i>Corymbia gummifera</i>	KC180800	160,713	88,310	17,197	27,603
<i>Corymbia henryi</i>	KP015032	160,095	88,589	18,688	26,409
<i>Corymbia maculata</i>	KC180801	160,045	88,557	18,670	26,409
<i>Corymbia tessellaris</i>	KC180803	160,127	88,617	18,692	26,409
<i>Corymbia torelliana</i>	KP015033	159,994	88,494	18,682	26,409
<i>Eucalyptus aromaphloia</i>	KC180789	160,149	88,925	18,468	26,378
<i>Eucalyptus baxteri</i>	KC180773	160,032	88,926	18,368	26,369
<i>Eucalyptus camaldulensis</i>	KC180791	160,164	88,874	18,492	26,399
<i>Eucalyptus cladocalyx</i>	KC180786	160,213	89,045	18,376	26,396
<i>Eucalyptus cloeziana</i>	KC180779	160,015	88,867	18,446	26,351
<i>Eucalyptus curtisii</i>	KC180782	160,038	88,828	18,448	26,381
<i>Eucalyptus deglupta</i>	KC180792	160,177	88,936	18,425	26,408
<i>Eucalyptus delegatensis</i>	KC180771	159,724	88,490	18,498	26,368
<i>Eucalyptus diversicolor</i>	KC180795	160,214	88,994	18,416	26,402
<i>Eucalyptus diversifolia</i>	KC180774	159,954	88,901	18,315	26,369
<i>Eucalyptus elata</i>	KC180776	159,899	88,762	18,401	26,368

<i>Eucalyptus erythrocorys</i>	KC180799	159,742	88,691	18,287	26,382
<i>Eucalyptus globulus</i>	AY780259	160,286	89,012	18,488	26,393
<i>Eucalyptus grandis</i>	HM347959	160,137	88,872	18,475	26,395
<i>Eucalyptus guilfoylei</i>	KC180798	160,520	89,054	18,096	26,685
<i>Eucalyptus marginata</i>	KC180781	160,076	88,828	18,476	26,386
<i>Eucalyptus melliodora</i>	KC180784	160,386	89,073	18,557	26,378
<i>Eucalyptus microcorys</i>	KC180797	160,225	89,051	18,410	26,382
<i>Eucalyptus nitens</i>	KC180788	160,271	89,005	18,468	26,399
<i>Eucalyptus obliqua</i>	KC180769	159,527	88,293	18,498	26,368
<i>Eucalyptus patens</i>	KC180780	160,187	88,902	18,543	26,371
<i>Eucalyptus polybractea</i>	KC180785	160,268	88,944	18,530	26,397
<i>Eucalyptus radiata</i>	KC180770	159,529	88,295	18,498	26,368
<i>Eucalyptus regnans</i>	KC180777	160,031	88,860	18,447	26,362
<i>Eucalyptus saligna</i>	KC180790	160,015	89,041	18,426	26,274
<i>Eucalyptus salmonophloia</i>	KC180796	160,413	89,173	18,466	26,387
<i>Eucalyptus sieberi</i>	KC180775	159,985	88,848	18,401	26,368
<i>Eucalyptus spathulata</i>	KC180793	161,071	88,729	17,116	27,613
<i>Eucalyptus torquata</i>	KC180794	160,223	89,018	18,439	26,383
<i>Eucalyptus umbra</i>	KC180778	159,576	88,864	18,658	26,027

<i>Eucalyptus verrucata</i>	KC180772	160,109	88,890	18,481	26,369
<i>Eugenia uniflora</i>	KR867678	158,445	87,459	18,318	26,334
<i>Stockwellia quadrifida</i>	KC180807	159,561	88,247	18,544	26,385
<i>Syzygium cumini</i>	GQ870669	160,373	89,081	18,508	26,392
<i>Lagerstroemia fauriei</i> ^c	KT358807	152,440	83,923	16,933	25,792

^a Species with plastid genomes sequenced in this study; ^b Large Single Copy Region; ^c Short Single Copy Region; ^d Inverted Repeat Region; ^e Outgroup

2.4 Phylogenetic analysis

Phylogenetic analysis was performed with the *ycf2* gene of the newly sequenced cpDNA of *C. xanthocarpa* and 44 previously published species of Myrtaceae (Table 1). The cp genomes of *Lagerstroemia fauriei* (Myrtales: Lythraceae; KT358807) and *Pelargonium x hortorum* (Geraniales: Geraniaceae; DQ897681) were tested as possible outgroups. The *ycf2* genes were aligned using MAFFT (Multiple Alignment using Fast Fourier Transform) (Katoh and Standley 2013). Nucleotide positions that contained one or more gaps introduced by the alignments were omitted from the matrix. The substitution model was selected using jModelTest (Darriba et al. 2012), and 7 was used the number of substitution schemes. The General Time Reversible model of substitution, incorporating invariant sites (GTR + I), was used in subsequent analyses. MrBayes 3.2.2 (Ronquist and Huelsenbeck 2003) was used to perform a Bayesian inference (BI) analysis. Markov Chain Monte Carlo (MCMC) analysis was run for 3,000,000 generations until standard deviation of split frequencies fell below 0.01, and the first 25% of trees was discarded as burn-in. Remaining trees were represented and edited using FigTree v1.4.1.

3. Results and Discussion

3.1 *Campomanesia xanthocarpa* plastome assembly and gene content

The Illumina Miseq reads obtained and submitted to *de novo* assembly resulted in high average genome coverage (~370x), as shown in Table 2. Thus, we consider the assembled genome to be of high quality. *C. xanthocarpa* plastome size was determined to be 158,131 bp, which is the smallest plastome size within the Myrtaceae family (Table 1). The *C. xanthocarpa* plastome presents the general quadripartite structure typical of angiosperms, consisting of a pair of IRs (25,970 bp) separated by the LSC (87,596 bp) and SSC (18,595 bp) regions (Fig. 1; Table 3).

Table 2. *Campomanesia xanthocarpa* chloroplast genome sequencing and assembly data.

Species	<i>C. xanthocarpa</i>
Plastome size (bp)	158,131
GC content in %	36.98
Total plastid read count	398,70
Mean read length (bp)^a	147.52
Reads mapped in aligned pairs	333,21
Average coverage	369.78

^a Mean read length (bp) after trimming using CLC Genomics Workbench with quality score limit of 0.05

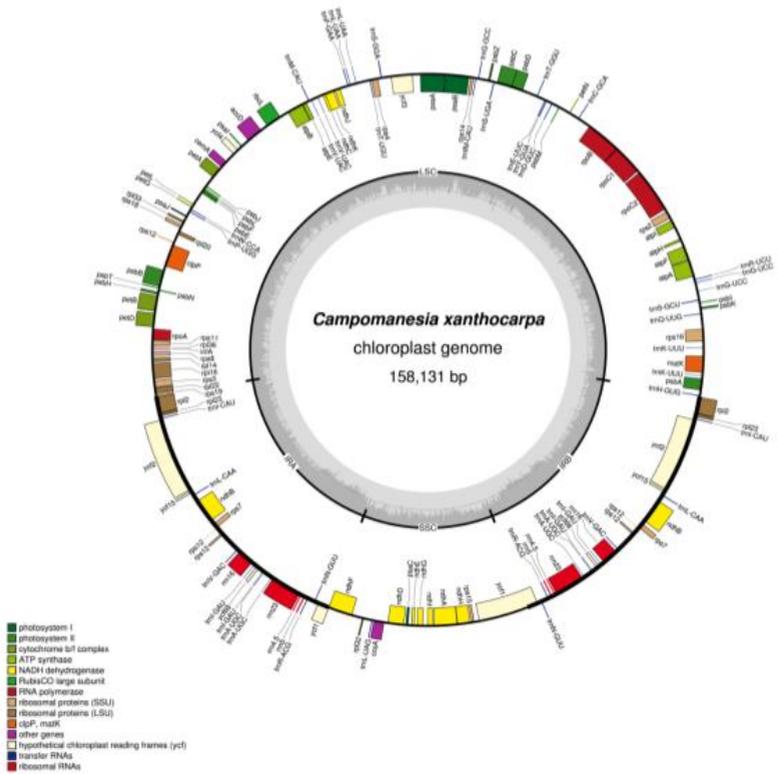


Figure 1. Gene map of *Campomanesia xanthocarpa* chloroplast genome. Genes drawn inside the circle are transcribed clockwise, and genes drawn outside are transcribed counterclockwise. Genes belonging to different functional groups are color-coded. The darker gray in the inner circle corresponds to GC content, and the lighter gray corresponds to AT content.

Table 3. Summary of *Campomanesia xanthocarpa* chloroplast genome characteristics.

Characteristics of cp genome	<i>C. xanthocarpa</i>
cp Genome Size (bp)	158,131
LSC size in bp (%)	87,596 (55.39)
SSC size in bp (%)	18,595 (11.76)
IR length in bp	25,970
Different genes	115
Different PCG	78
Different tRNA genes	30
Different rRNA genes	4
Different pseudogenes	4
Different genes duplicated by IR	20
Different genes with introns	18
Overall % GC content	36.98
% GC content in LSC	34.8
% GC content in SSC	30.6
% GC content in IR	42.9

cp = chloroplast; PCG=Protein-coding gene

The GC content determined for *C. xanthocarpa* is 36.98%, which is very similar to other Myrtales species (Gu et al. 2016; Machado et al. 2017). The cp genome of *C. xanthocarpa* contains 112 genes and 4 pseudogenes with the same gene order and gene clusters as other Myrtaceae (Fig. 2). Of the 112 genes, 90 were single copy and 19 were duplicated (Fig. 1; Table 4). In addition, 18 were intron-containing genes (Table 5), including nine protein-coding genes with a single

intron, two protein-coding genes with a double intron, six tRNA genes with a single intron and one trans-splicing gene (*rps12*). Among intron-containing genes, 12 are located in the LSC region, one in the SSC region, and four in the IR region. These results corroborate the findings of Machado et al. (2017) for feijoa (*Acca sellowiana*).

The protein *ycf2* has a different C-terminal region in *C. xanthocarpa* when compared with the other 44 species of Myrtaceae (Fig. 3). Previous studies have shown that this gene presented high variability in Myrtales (Gu et al. 2016).

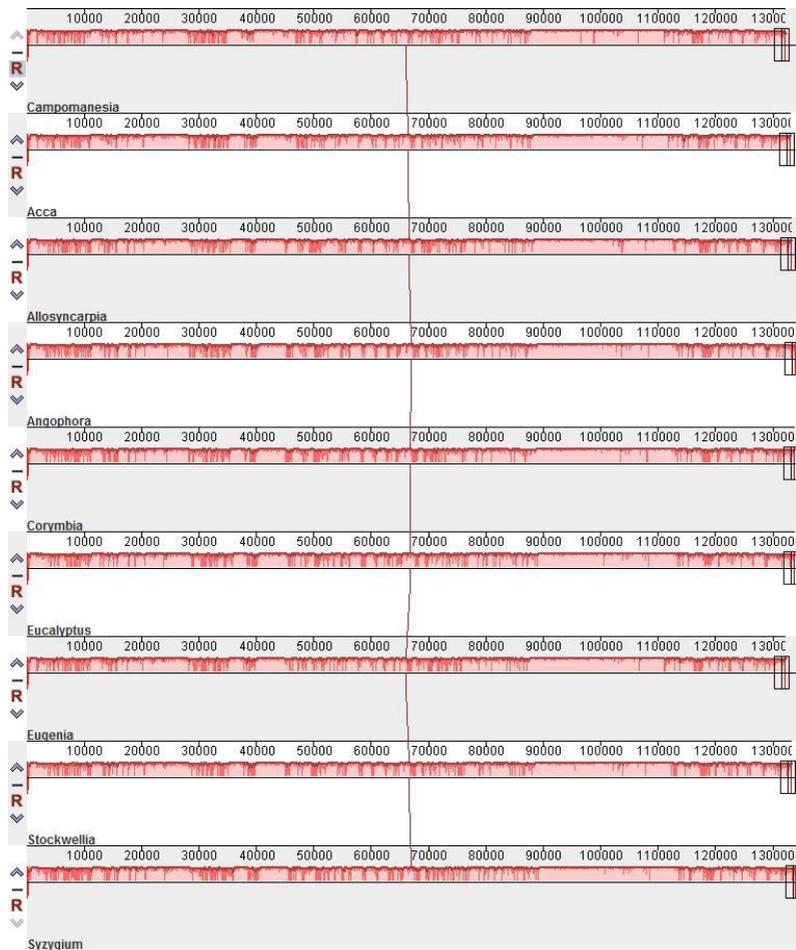


Figure 2. Multiple genome alignment performed using Mauve software and the chloroplast genome sequences of nine genera of Myrtaceae. The presence of only one local collinear block (LCB), as herein represented in pink color, indicates homologous DNA region without sequence rearrangements.

<i>C. xanthocarpa</i>	M	V	V	A	I	C	S	N	N	E	S	L	V	*
<i>A. sellowiana</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>A. ternata</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>A. costata</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>A. floribunda</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>C. eximia</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>C. gummifera</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>C. henryi</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>C. maculata</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>C. tessellaris</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>C. torelliana</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. aromaphloia</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. baxteri</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. camaldulensis</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. cladocalyx</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. cloeziana</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. curtisii</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. deglupta</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. delegatensis</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. diversicolor</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. diversifolia</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. elata</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. erythrocorys</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. globulus</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. grandis</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. guilfoylei</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. marginata</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. melliodora</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. microcorys</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. nitens</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. obliqua</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. patens</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. polybractea</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. radiata</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. regnans</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. saligna</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. salmonophloia</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. sieberi</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. spatulata</i>	M	K	I	G	F	I	V	E	R	F	P	I	T	*
<i>E. torquata</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. umbra</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. verrucata</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>E. uniflora</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>S. quadrifida</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-
<i>S. cumini</i>	M	K	I	G	F	M	*	-	-	-	-	-	-	-

Figure 3. Comparison of final parts (C-terminal) of the amino acid sequences of *ycf2* gene in 45 species of Myrtaceae.

Table 4. List of genes identified in *Campomanesia xanthocarpa* chloroplast genome.

Category of genes	Group of genes	Name of genes										
Self - replication	Large subunit of ribosome	<i>rpl14</i>	<i>rpl16</i>	<i>rpl2^{ab}</i>	<i>rpl20</i>	<i>rpl22</i>	<i>rpl23^b</i>	<i>rpl32</i>	<i>rpl33</i>	<i>rpl36</i>		
	DNA-dependent RNA polymerase	<i>rpoA</i>	<i>rpoB</i>	<i>rpoC1^a</i>	<i>rpoC2</i>							
	Small subunit of ribosome	<i>rps11</i>	<i>rps12^{a*}</i>	<i>rps14</i>	<i>rps15</i>	<i>rps16^a</i>	<i>rps18</i>	<i>rps19</i>	<i>rps2</i>	<i>rps3</i>		
	Ribosomal RNA Genes	<i>rml6</i>	<i>rml2^a</i>	<i>rml4^a</i>	<i>rml5^a</i>							
	Transfer RNA Genes		<i>trnA-UGC^{a*}</i>	<i>trnC-GCA</i>	<i>trnD-GUC</i>	<i>trnE-UUC</i>	<i>trnF-GAA</i>	<i>trnG4-CAU</i>	<i>trnG-GCC</i>	<i>trnG-UCC^a</i>	<i>trnH-GUG</i>	
			<i>trnI-CAU^b</i>	<i>trnI-GAU^{ab}</i>	<i>trnK-UUU^b</i>	<i>trnL-CAA^b</i>	<i>trnL-UAA^a</i>	<i>trnL-UAG</i>	<i>trnM-CAU</i>	<i>trnN-GUU^b</i>	<i>trnP-UGG</i>	
			<i>trnQ-UUG</i>	<i>trnR-ACG^a</i>	<i>trnR-UUU</i>	<i>trnS-GCU</i>	<i>trnS-GGA</i>	<i>trnS-UGA</i>	<i>trnT-GGU</i>	<i>trnT-UGU</i>	<i>trnV-GAC^c</i>	
	Translational initiation factor	<i>infA^c</i>										
	Genes for photosynthesis	Subunits of photosystem I	<i>psaA</i>	<i>psaB</i>	<i>psaC</i>	<i>psaI</i>	<i>psaJ</i>	<i>ycf3^a</i>	<i>ycf4</i>			
		Subunits of photosystem II	<i>psbA</i>	<i>psbB</i>	<i>psbC</i>	<i>psbD</i>	<i>psbE</i>	<i>psbF</i>	<i>psbH</i>	<i>psbI</i>	<i>psbJ</i>	
<i>psbK</i>			<i>psbL</i>	<i>psbM</i>	<i>psbN</i>	<i>psbT</i>	<i>psbZ</i>					
Subunits of cytochrome		<i>petA</i>	<i>petB^a</i>	<i>petD^a</i>	<i>petG</i>	<i>petL</i>	<i>petN</i>					
Subunits of ATP synthase		<i>atpA</i>	<i>atpB</i>	<i>atpE</i>	<i>atpF^a</i>	<i>atpH</i>	<i>atpI</i>					
Large subunit of rubisco		<i>rbcL</i>										
Subunits of NADH-dehydrogenase		<i>ndhA^a</i>	<i>ndhB^{ab}</i>	<i>ndhC</i>	<i>ndhD</i>	<i>ndhE</i>	<i>ndhF</i>	<i>ndhG</i>	<i>ndhH</i>	<i>ndhI</i>		
	<i>ndhJ</i>	<i>ndhK</i>										
Other genes	Subunit of Acetyl-CoA-carboxylase	<i>accD</i>										
	c-type cytochrome synthesis gene	<i>ccsA</i>										
	Envelope membrane protein	<i>cemA</i>										
	Protease	<i>clpP^a</i>										
	Maturase	<i>matK</i>										
	Component of TIC complex	<i>ycf1^{bc*}</i>										
Genes unknown function	Unknown function	<i>ycf2^b</i>	<i>ycf13^{bc}</i>	<i>ycf68^{bc}</i>								

^a Intron-containing gene; ^b Duplicated genes; ^c Pseudogene; ^{c*} *ycf1* is pseudogene at the boundary between IRB and SSC regions; ^{*} *rps12* is trans-spliced with the 5'-end located in the LSC region and the duplicated 3'-end in the IR region.

Table 5. Genes with introns in *Campomanesia xanthocarpa* chloroplast genome and length of exons and introns.

Region	Gene	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
LSC	rps16	206	889	38		
LSC	rpoC1	1616	730	452		
LSC	atpF	410	747	146		
LSC	petB	5	772	647		
LSC	petD	8	753	473		
LSC	rpl16	398	1005	8		
LSC	ycf3	152	725	227	760	125
LSC	clpP	227	620	290	871	68
LSC	trnK-UUU	34	2530	36		
LSC	trnG-UCC	22	750	48		
LSC	trnV-UAC	36	595	38		
LSC	trnL-UAA	36	504	49		
SSC	ndhA	539	1061	551		
LSC/IRs	rps12 *	113	–	209	–	26
IR	rpl2	434	662	392		
IR	ndhB	755	695	776		
IR	trnI-GAU	36	956	33		
IR	trnA-UGC	37	804	34		

*rps12 is trans-spliced with the 5'-end located in the LSC region and the duplicated 3'-end in the IR regions.

3.2 Diversity of *ycf2* protein-coding gene sequence in Myrtaceae

The cp *ycf2* gene has a vital, but as yet unknown, function in higher plants, essentially because silencing of, or reduction in, mRNA synthesis of this gene induces cell apoptosis (Drescher et al. 2000). In order to characterize evolutionary patterns of the cp *ycf2* gene in the Myrtaceae family, the *Ka/Ks* ratio (Table 6) was calculated for the *ycf2* protein-coding gene for *C. xanthocarpa* and 44 species of Myrtaceae, which included 31 species of the *Eucalyptus* genus, six of the *Corymbia* genus, two of the *Angophora* genus, and one each of *Allosyncarpia*, *Eugenia*, *Stockwellia*, *Syzygium*, and *Acca*.

Ka ranged from 0.0021 to 0.0054, and *Ks* ranged from 0.0010 to 0.0138, while the average *Ka/Ks* ratio was 0.46, values similar to those of Machado et al. (2017), who detected a *Ka/Ks* ratio of 0.30 when comparing *Acca sellowiana* and *Eugenia uniflora* cp genomes. Thus, these *ycf2* genes gave evidence of strong purifying selection, as indicated by their low *Ka/Ks* ratios. This result stands in contrast to previous studies with *Campanulastrum americanum* for which the *ycf2* gene exhibited evidence of relaxed purifying selection, or positive selection (Barnard-Kubow et al. 2014). In addition, polymorphisms data (Table 6) indicated the closest relationship to be among *C. xanthocarpa*, *A. sellowiana* and *E. uniflora* members of Myrteae (Myrtaceae), and this was confirmed by phylogenetic inference.

Table 6. Comparison of nonsynonymous (*Ka*) and synonymous (*Ks*) substitution rates and *Ka/Ks* ratio among *Campomanesia xanthocarpa* and 44 species of Myrtaceae available in GenBank.

Species	Gene	<i>C. xanthocarpa</i> /Species			
		Polymorphisms	<i>Ka</i>	<i>Ks</i>	<i>Ka/Ks</i>
<i>Acca sellowiana</i>	<i>ycf2</i>	13	0.0021	0.0010	2.0114
<i>Allosyncarpia ternata</i>	<i>ycf2</i>	40	0.0046	0.0122	0.3778
<i>Angophora costata</i>	<i>ycf2</i>	41	0.0049	0.0113	0.4391

<i>Angophora floribunda</i>	ycf2	41	0.0049	0.0113	0.4391
<i>Corymbia eximia</i>	ycf2	43	0.0051	0.0120	0.4269
<i>Corymbia gummifera</i>	ycf2	42	0.0054	0.0101	0.5314
<i>Corymbia henryi</i>	ycf2	42	0.0051	0.0110	0.4677
<i>Corymbia maculata</i>	ycf2	42	0.0051	0.0110	0.4677
<i>Corymbia tessellaris</i>	ycf2	43	0.0052	0.0117	0.4420
<i>Corymbia torelliana</i>	ycf2	44	0.0052	0.0127	0.4053
<i>Eucalyptus aromaphloia</i>	ycf2	42	0.0050	0.0116	0.4309
<i>Eucalyptus baxteri</i>	ycf2	40	0.0046	0.0118	0.3921
<i>Eucalyptus camaldulensis</i>	ycf2	43	0.0052	0.0117	0.4433
<i>Eucalyptus cladocalyx</i>	ycf2	44	0.0052	0.0127	0.4062
<i>Eucalyptus cloeziana</i>	ycf2	40	0.0046	0.0118	0.3921
<i>Eucalyptus curtisii</i>	ycf2	44	0.0053	0.0121	0.4403
<i>Eucalyptus deglupta</i>	ycf2	44	0.0054	0.0115	0.4704
<i>Eucalyptus delegatensis</i>	ycf2	41	0.0048	0.0119	0.4043
<i>Eucalyptus diversicolor</i>	ycf2	44	0.0052	0.0124	0.4199
<i>Eucalyptus diversifolia</i>	ycf2	41	0.0048	0.0116	0.4189
<i>Eucalyptus elata</i>	ycf2	40	0.0046	0.0118	0.3921
<i>Eucalyptus erythrocorys</i>	ycf2	46	0.0053	0.0138	0.3811
<i>Eucalyptus globulus</i>	ycf2	43	0.0052	0.0114	0.4584
<i>Eucalyptus</i>	ycf2	45	0.0054	0.0125	0.4318

grandis

<i>Eucalyptus guilfoylei</i>	ycf2	41	0.0048	0.0119	0.4038
<i>Eucalyptus marginata</i>	ycf2	42	0.0050	0.0117	0.4302
<i>Eucalyptus melliodora</i>	ycf2	45	0.0053	0.0128	0.4173
<i>Eucalyptus microcorys</i>	ycf2	39	0.0044	0.0121	0.3655
<i>Eucalyptus nitens</i>	ycf2	42	0.0050	0.0116	0.4309
<i>Eucalyptus obliqua</i>	ycf2	41	0.0048	0.0119	0.4043
<i>Eucalyptus patens</i>	ycf2	39	0.0046	0.0108	0.4304
<i>Eucalyptus polybractea</i>	ycf2	44	0.0052	0.0127	0.4062
<i>Eucalyptus radiata</i>	ycf2	41	0.0048	0.0119	0.4043
<i>Eucalyptus regnans</i>	ycf2	41	0.0046	0.0128	0.3594
<i>Eucalyptus saligna</i>	ycf2	42	0.0051	0.0119	0.4283
<i>Eucalyptus salmonophloia</i>	ycf2	44	0.0050	0.0136	0.3641
<i>Eucalyptus sieberi</i>	ycf2	41	0.0046	0.0128	0.3591
<i>Eucalyptus spathulata</i>	Ycf2	44	0.0053	0.0121	0.4362
<i>Eucalyptus torquata</i>	ycf2	43	0.0052	0.0117	0.4429
<i>Eucalyptus umbra</i>	ycf2	42	0.0050	0.0120	0.4162
<i>Eucalyptus verrucata</i>	ycf2	41	0.0048	0.0116	0.4186
<i>Eugenia uniflora</i>	ycf2	15	0.0021	0.0028	0.7413
<i>Stockwellia quadrifida</i>	ycf2	34	0.0037	0.0115	0.3207
<i>Syzygium</i>	ycf2	35	0.0043	0.0093	0.4587

Total average	0.0048	0.0114	0.4620
----------------------	--------	--------	--------

3.3 Phylogenetic Inference

Campomanesia xanthocarpa belongs to the tribe Myrteae, which is the most species-rich in Myrtaceae (Thornhill et al. 2015). The complete cp genome aligned matrix using the *ycf2* gene of 45 species of Myrtaceae (Table 1) was 6,963 nucleotide positions in length. After removal of gaps, the matrix length was 6,702 sites. We compared the phylogenetic trees generated using *L. fauriei* (Lythraceae) or *Pelargonium x hortorum* (Geraniaceae) as outgroup. The log-likelihoods (lnL) obtained using this species were determined through BI (lnL = -10,967 and -11,813, respectively). Thus, we used *Lagerstroemia fauriei* as the outgroup because it showed better results than *Pelargonium x hortorum*. Support of BI posterior probabilities (PP) is shown in Figure 4.

Phylogenetic analysis of cp *ycf2* gene sequences using the BI approach (Fig. 4) resulted in three monophyletic tribes of Myrtoideae. Phylogenetic inferences were based on plastidial genes, such as *rbcL*, *ndhF* and *matK*, and nuclear genes, such as 18S–26S rRNA and *ITS* (Sytsma et al. 2004; Wilson et al. 2005; Biffin et al. 2010; Thornhill et al. 2012; Thornhill et al. 2015), and complete plastome sequences (Machado et al. 2017) reported monophyly of Myrteae, Eucalypteae e Syzygieae tribes. BI posterior probability values were between 0.7 and 0.9 for 11 nodes and 1.0 for the rest (Fig. 4).

In Myrteae, the *E. uniflora* + *A. sellowiana* + *C. xanthocarpa* clade showed low support as monophyletic (PP = 0.75; node 1). However, the sister clade Syzygieae + Eucalypteae was a well-supported clade (PP = 0.99; node 3). These results were previously reported in phylogenetic studies using combined analysis of plastid DNA regions (*matK* and *ndhF*) and nuclear regions (18S–26S rRNA and *ITS*) in Biffin et al. (2010) and complete cp sequences (Machado et al. 2017). Relative to the overall evolutionary radiation of Myrtaceae, as happened in the Oligocene–Miocene, the Syzygieae and Myrteae lineages show elevated evolutionary diversification rates as a

consequence of the highly significant positive shift in those rates (Biffin et al. 2010).

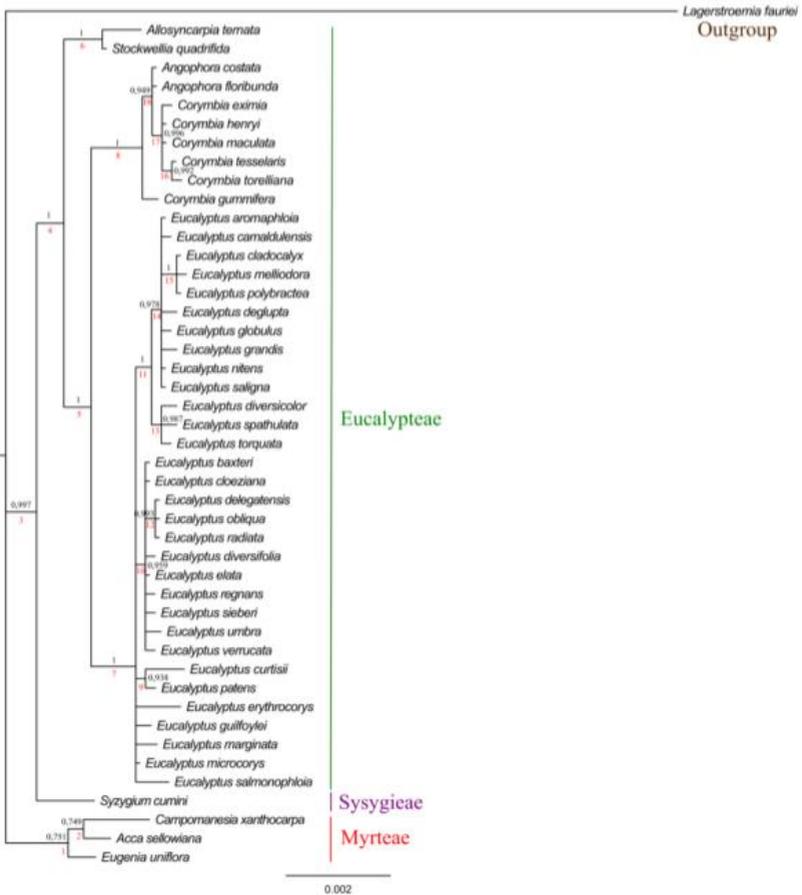


Figure 4. Bayesian phylogeny based on the cp *ycf2* sequence of 45 Myrtaceae species and the outgroup *Lagerstroemia fauriei* (Myrtales: Lythraceae; KT358807). Branch length is proportional to the inferred divergence level. The scale bar indicates the number of inferred nucleic acid substitutions per site.

The Eucalypteae tribe segregated into two well-supported monophyletic clades: *Eucalyptus* + *Angophora* + *Corymbia* (PP = 1.0; node 5) and *A. ternata* + *S. quadrifida* (PP = 1.0; node 6). Our results using the cp *ycf2* gene showed that the tribe Eucalypteae (node 5) is topologically similar to that proposed by Bayly et al. (2013) and Bayly (2016) based on plastid genome. In addition, the well-supported monophyly, as exhibited in the *Angophora* + *Corymbia* sister relationship (PP = 1.0; node 8) has already been reported in studies based on nuclear ribosomal ETS sequences (Parra-O et al. 2006). Taken together, these results suggest that the sequence of the *ycf2* gene is a useful tool for understanding evolutionary relationships.

4. Conclusion

In conclusion, *ycf2* gene analysis revealed purifying evolution, and the polymorphisms indicated sister relationships among species of the tribe Myrteae, which is equivalent to that observed in phylogenetic analysis at genus level in BI analyses. In addition, the BI tree supports monophyly of Myrteae, Syzygieae and Eucalypteae. Moreover, these characteristics of the *ycf2* gene make it a potential marker that corroborates, as well as complements and resolves, the phylogenetic relationships at species level within one to several genera. Furthermore, the complete cp genome of *C. xanthocarpa* will allow the development of molecular markers, such as short sequence repeats (SSRs), to contribute in future phylogeographic and population genetics studies for this species.

References

- Asano T, Tsudzuki T, Takahashi S, et al (2004) Complete Nucleotide Sequence of the Sugarcane (*Saccharum*. 99:93–99.
- Barnard-Kubow KB, Sloan DB, Galloway LF (2014) Correlation between sequence divergence and polymorphism reveals similar evolutionary mechanisms acting across multiple timescales in a rapidly evolving plastid genome. *BMC Evol Biol* 14:1. doi: 10.1186/s12862-014-0268-y
- Bayly MJ (2016) Phylogenetic studies of eucalypts: Fossils, morphology and genomes. *Proc R Soc Victoria* 128:12–24. doi: 10.1071/RS16002
- Bayly MJ, Rigault P, Spokevicius A, et al (2013) Chloroplast genome analysis of Australian eucalypts - *Eucalyptus*, *Corymbia*, *Angophora*, *Allosyncarpia* and *Stockwellia* (Myrtaceae). *Mol Phylogenet Evol* 69:704–716. doi: 10.1016/j.ympev.2013.07.006
- Biffin E, Lucas EJ, Craven LA, et al (2010) Evolution of exceptional species richness among lineages of fleshy-fruited Myrtaceae. *Ann Bot* 106:79–93. doi: 10.1093/aob/mcq088
- Bock R (2007) Structure, function, and inheritance of plastid genomes. *Cell Mol Biol Plast* 19:29–63. doi: 10.1007/4735
- Darling ACE, Mau B, Blattner FR, Perna NT (2004) Mauve : Multiple Alignment of Conserved Genomic Sequence With Rearrangements
Mauve : Multiple Alignment of Conserved Genomic Sequence With Rearrangements. 1394–1403. doi: 10.1101/gr.2289704
- Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9:772–

772. doi: 10.1038/nmeth.2109

Downie SR, Katz-Downie DS, Wolfe KH, et al (1994) Structure and evolution of the largest chloroplast gene (ORF2280): internal plasticity and multiple gene loss during angiosperm evolution. *Curr Genet* 25:367–378.

Drescher A, Ruf S, Jr TC, et al (2000) The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes.

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. doi: 10.1186/1471-2105-5-113

Glick RE, Sears BB (1993) Large unidentified open reading frame in plastid DNA (ORF2280) is expressed in chloroplasts. *Plant Mol Biol* 21:99–108. doi: 10.1007/BF00039621

Gu C, Tembrock LR, Johnson NG, et al (2016) The Complete plastid genome of *lagerstroemia fauriei* and loss of rpl2 intron from *lagerstroemia* (Lythraceae). *PLoS One* 11:1–18. doi: 10.1371/journal.pone.0150752

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 30:772–780. doi: 10.1093/molbev/mst010

Kumar S, Hahn FM, McMahan CM, et al (2009) Comparative analysis of the complete sequence of the plastid genome of *Parthenium argentatum* and identification of DNA barcodes to differentiate *Parthenium* species and lines. *BMC Plant Biol* 9:131. doi: 10.1186/1471-2229-9-131

Kurtz S, Schleiermacher C (1999) REPuter: Fast computation of maximal repeats in complete genomes. *Bioinformatics* 15:426–427. doi:

10.1093/bioinformatics/15.5.426

- Lohse M, Drechsel O, Bock R (2007) OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet* 52:267–274. doi: 10.1007/s00294-007-0161-y
- Machado LO, Vieira LN, Stefenon, VM, Pedrosa FO, Souza EM, Guerra MP, Nodari RO (2017) Phylogenomic relationship of feijoa (*Acca sellowiana* (O.Berg) Burret) with other Myrtaceae based on complete chloroplast genome sequences. *Genetica*. doi: 10.1007/s10709-017-9954-1
- Maier RM, Neckermann K, Igloi GL, Ko H (1995) Complete Sequence of the Maize Chloroplast Genome : Gene Content , Hotspots of Divergence and Fine Tuning of Genetic Information by Transcript Editing. 614–628.
- Matsuoka Y, Yamazaki Y, Ogihara Y, Tsunewaki K (2002) Whole Chloroplast Genome Comparison of Rice, Maize, and Wheat: Implications for Chloroplast Gene Diversification and Phylogeny of Cereals. *Mol Biol Evol* 19:2084–2091. doi: 10.1093/oxfordjournals.molbev.a004033
- Millen RS, Olmstead RG, Adams KL, et al (2001) Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell* 13:645–658. doi: 10.1105/tpc.13.3.645
- Oliver MJ, Murdock AG, Mishler BD, et al (2010) Chloroplast genome sequence of the moss *Tortula ruralis*: gene content, polymorphism, and structural arrangement relative to other green plant chloroplast genomes. *BMC Genomics* 11:143. doi: 10.1186/1471-2164-11-143

- Parra-O C, Bayly M, Udovicic F, Ladiges P (2006) ETS sequences support the monophyly of the eucalypt genus *Corymbia* (Myrtaceae). *Taxon* 55:653–663. doi: 10.2307/25065641
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574. doi: 10.1093/bioinformatics/btg180
- Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* 33:686–689. doi: 10.1093/nar/gki366
- Steane DA (2005) Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum, *Eucalyptus globulus* (Myrtaceae). *DNA Res* 12:215–220. doi: 10.1093/dnares/dsi006
- Sytsma KJ, Litt A, Zjhra ML, et al (2004) Clades , Clocks , and Continents : Historical and Biogeographical Analysis of Myrtaceae , Vochysiaceae , and Relatives in the Southern Hemisphere Source : International Journal of Plant Sciences , Vol . 165 , No . S4 , Tropical Intercontinental Disjunctio. *Int J Plant Sci* 165:S85–S105.
- Thornhill AH, Ho SYW, Külheim C, Crisp MD (2015) Interpreting the modern distribution of Myrtaceae using a dated molecular phylogeny. *Mol Phylogenet Evol* 93:29–43. doi: 10.1016/j.ympev.2015.07.007
- Thornhill AH, Popple LW, Carter RJ, et al (2012) Are pollen fossils useful for calibrating relaxed molecular clock dating of phylogenies? A comparative study using Myrtaceae. *Mol Phylogenet Evol* 63:15–27. doi: 10.1016/j.ympev.2011.12.003
- Vieira LDN, Faoro H, De Freitas Fraga HP, et al (2014) An improved protocol for intact chloroplasts and cpDNA isolation in conifers. *PLoS One* 9:1–8. doi: 10.1371/journal.pone.0084792

- Wilson PG, O'Brien MM, Heslewood MM, Quinn CJ (2005) Relationships within Myrtaceae sensu lato based on a matK phylogeny. *Plant Syst Evol* 251:3–19. doi: 10.1007/s00606-004-0162-y
- Wolf PG, Der JP, Duffy AM, et al (2011) The evolution of chloroplast genes and genomes in ferns. *Plant Mol Biol* 76:251–261. doi: 10.1007/s11103-010-9706-4
- Wolfe KH (1994) Similarity between putative ATP-binding sites in land plant plastid ORF2280 proteins and the FtsH/CDC48 family of ATPases. *Curr Genet* 25:379–383.
- Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* (Oxford, England) 20:3252–3255.
- Zhang Z, Li J, Zhao XQ, et al (2006) KaKs_Calculator: Calculating Ka and Ks Through Model Selection and Model Averaging. *Genomics, Proteomics Bioinforma* 4:259–263. doi: 10.1016/S1672-0229(07)60007-2

CAPÍTULO 3

**Análise comparativa intraespecífica do genoma plastidial de
Eugenia uniflora L.**

1. Introdução

Eugenia uniflora pertence à tribo Myrteae, a mais rica em espécies da família Myrtaceae. Essa espécie é conhecida popularmente no Brasil como pitanga ou Brazilian cherry em países de língua inglesa. No Brasil, a pitangueira apresenta diferentes formas de mudas, tamanho, cor e sabor dos frutos; porém não são conhecidas variedades perfeitamente definidas (LEDERMAN et al. 1992; FRANZÃO; MELO 2017). Essa espécie apresenta um incipiente interesse econômico não só pelo fruto, que é muito utilizado na confecção de sucos, geleias, caipiras, sorvetes e licores, já que, a comercialização *in natura* é difícil pela pericibilidade dos frutos (BOURSCHEID et al. 2011). Além disso, o extrato das folhas de *E. uniflora* são de interesse farmacológico, pois as folhas da pitangueira contêm propriedades diuréticas, antipirética, antirreumática, antidiarreica, antidiabetes (ARAI et al. 1999; SANTOS et al. 2012, VICTORIA et al. 2012, DENARDIN et al. 2014) e podem ajudar no tratamento de fibrose hepática (DENARDIN et al. 2016).

De fato, alguns estudos tem demonstrado a importância ecológica da espécie *E. uniflora* revelando informações acerca da análise da diversidade genética em populações dessa espécie por meio da utilização de marcadores moleculares, como marcadores microssatélites (SSRs) e polimorfismo no comprimento de fragmentos amplificados (AFLPs) (FERREIRA-RAMOS et al. 2008; FERREIRA-RAMOS et al. 2014; MARGIS et al. 2002; SALGUEIRO et al. 2004). Além disso, a análise filogeográfica da espécie foi realizada utilizando dois marcadores intergênicos plastidiais (*psbA/trnH* and *trnC/ycF6*) (TURCHETTO-ZOLET et al. 2011) e, posteriormente, utilizando três marcadores intergênicos plastidiais (*psbA-trnH*, *trnS-trnG* e *trnC-ycf6*) em 46 populações de *E. uniflora* (TURCHETTO-ZOLET et al. 2016). Contudo, mesmo com estes estudos, há pouca disponibilidade de resultados com o uso de ferramentas genômicas, tais como polimorfismos de nucleotídeo único (SNPs), para serem utilizados em apoio aos programas de melhoramento da espécie.

Neste estudo, o genoma plastidial de *E. uniflora* foi sequenciado e analisado quanto a estrutura e ordem gênica. Além disso, foram realizadas análises comparativas intraespecíficas entre os

genomas plastidiais de dois indivíduos de *E. uniflora* visando identificar polimorfismos, tais como SNPs e de inserções e deleções (*indels*), que poderão ser eficientes ferramentas genômicas para estudos de genética de populações e seleção assistida dessa espécie.

2. MATERIAL E MÉTODOS

2.1 Coleta de material vegetal

Foram coletadas folhas novas e sadias de um indivíduo de *E. uniflora*, cuja exsicata foi depositada no herbário do Departamento de Botânica da UFSC com o seguinte número de tomo: FLOR0059177. As coordenadas geográficas do indivíduo *E. uniflora* são (27° 34. 911' S/48° 30. 275' W). Após a coleta, as amostras foliares foram armazenadas no escuro, por dois dias a temperatura de 4°C, em geladeira, no Laboratório de Fisiologia e Desenvolvimento Vegetal (LFDGV), procedimento este que tem por função o decréscimo dos níveis de amido e resina das folhas.

2.2 Isolamento do cloroplasto e do DNA de cloroplasto (cpDNA)

O isolamento dos cloroplastos foi realizado por meio da metodologia descrita por VIEIRA et al. 2014. O protocolo de isolamento de cpDNA foi otimizado, a partir dos cloroplastos isolados, para as espécies de Myrtaceae. Para o isolamento dos cloroplastos, foram coletadas aproximadamente 100 g de folhas, as quais foram mantidas no escuro a 4°C por dois dias antes da realização do procedimento. As folhas foram homogeneizadas em 400 ml de tampão de extração A pH 3,8 (1,25 M NaCl, 0,25 M ácido ascórbico, 10 mM metabissulfito de sódio, 0,0125 M Borax, 50 mM Tris-HCl (pH 8.0), 7 mM EDTA, 1% PVP-40 (m/v), 0,1% BSA (m/v)) em liquidificador por 30 s para a quebra da parede celular. Em seguida, foram realizadas duas filtrações, sendo a primeira em atadura de crepom Cremer com densidade de 18 fios/cm², e a segunda em Miracloth[®]. Essas duas filtrações são importantes para eliminar grande quantidade de resíduo do tampão contendo os cloroplastos. O material filtrado foi centrifugado a 200 g por 15 min a 4°C. Após esta etapa os sedimentos nucleares e de parede celular foram descartados. O sobrenadante, contendo os cloroplastos suspensos, foi centrifugado a 3000 g por 25 min a 4°C. O *pellet* formado nesta etapa contém os cloroplastos e resquícios de DNA nuclear. Posteriormente, o novo sobrenadante foi descartado e o *pellet* resultante ressuspensionado em 10 ml de tampão de extração B pH 8,0 (1,25 M NaCl, 0,0125 M Borax, 1% PVP-40 (m/v), 50 mM Tris-HCl (pH 8.0), 25 mM EDTA, 0,1% BSA (m/v)), com o auxílio de um pincel,

com a função de lavar os resquícios de DNA nuclear. Os cloroplastos ressuspensos foram centrifugados em gradiente de Percoll® (70%/30%) a 5.000 g durante 25 min a 4°C. A interface contendo os cloroplastos, entre as soluções 30% e 70% foi coletada e foram realizadas duas lavagens em tampão de extração B pH 8,0 para a retirada do resíduo de Percoll®. Para tanto, foram adicionados 100 ml do tampão de extração B, para cada lavagem, e foi realizada a centrifugação a 3000 g por 20 min a 4°C. Entre as lavagens, o *pellet* de cloroplastos resultante foi ressuspensionado com pincel. Por fim, foi obtido o *pellet* de cloroplastos que foi armazenado em freezer a -80°C até o isolamento do cpDNA.

Para isolamento do cpDNA foi utilizado o tampão de isolamento de DNA (100 mM NaCl, 100 mM Tris - HCl (pH 8,0), 50 mM EDTA). A lise dos cloroplastos foi obtida por incubação do *pellet* de cloroplastos em 8 ml de tampão de isolamento de DNA, 1,5 ml de SDS 20 %, 450 µl de 2-mercaptoetanol e 50 µl de Proteinase K (10 mg / ml) em thermo-shaker a 55°C por 4h ou *overnight* para diminuir níveis de polifenóis das amostras. Após, os tubos foram incubados em gelo durante 5 min e, posteriormente, foi adicionado uma solução contendo 1,5 ml de KAc 5M (pH 6,4) a qual foi refrigerada por cerca de 30 min a -20°C. O KAc tem a função de precipitar impurezas como proteínas e polissacarídeos separando-os do DNA. Depois disso, os tubos foram centrifugados a 10.000 g durante 15 min a 4°C, e os sedimentos foram descartados. O sobrenadante foi extraído com um volume igual de fenol saturado e CIA, na proporção (24:1), fenol e CIA são eficientes na desnaturação de proteínas, o material foi centrifugado duas vezes a 10.000 g durante 20 min. Após, foi adicionado ao sobrenadante um volume igual de álcool isopropílico (cerca de 10 ml) e os tubos foram incubados em freezer -20°C *overnight*, para precipitar o DNA plastidial. Por fim, para obtenção do *pellet* de cpDNA, os tubos foram centrifugados a 10.000 g por 20 min a 4°C. O sedimento contendo os cpDNAs foi lavado com etanol 70 % e 96 % e em temperatura ambiente. O *pellet* de cpDNA foi dissolvido em 50µl de água destilada UltraPure™ DNase/RNase-Free Invitrogen™ e posteriormente tratado com RNase (10µg /ml). A qualidade das amostras foi verificada por meio da visualização em gel de agarose a 0,8 %. A pureza e concentração de DNA foram avaliadas por: Nanodrop H, baseado nas taxas 260/280 e 260/230 e Qubit® 2.0 fluorímetro Invitrogen™.

2.3 Sequenciamento e anotação do genoma plastidial

2.3.1 Sequenciamento em sequenciador de nova geração

Foi utilizado 1 ng do DNA plastidial extraído a partir dos cloroplastos isolados para o sequenciamento. A construção das bibliotecas de DNA plastidial foi realizada utilizando o kit Nextera® XT DNA Sample Prep v3 (Illumina, San Diego, USA) de acordo com as recomendações do fabricante. A qualidade dos fragmentos gerados pela biblioteca foram checadas em Bioanalyzer 2100 (Agilent Technologies) e, após apresentarem qualidade satisfatória, foram sequenciados em Illumina MiSeq (San Diego, USA). O sequenciamento foi realizado no Departamento de Bioquímica e Biologia Molecular da Universidade Federal do Paraná. Após o sequenciamento foi adotada a estratégia de montagem *de novo* utilizando o programa CLC workbench versão 8.0.1.

2.3.2 Desenho de iniciadores e padronização da amplificação via reação em cadeia de polimerase (PCR)

Foram desenhados três pares de iniciadores (Tabela 1) para verificar três regiões do cpDNA de *E. uniflora*, com cobertura baixa, por meio do método Sanger de sequenciamento no sequenciador automático ABI 3500XL (Applied Biosystems™). Para o desenho dos iniciadores foi utilizado o programa *on line* Primer3Plus (UNTERGASSER et al. 2007). Para amplificação dos fragmentos foi utilizado 5 ng de amostra de DNA, 0,2mg/ml de albumina de soro bovino (BSA), 0,2 mM de cada dNTP, 1U de Taq DNA Polymerase (Quatro G), 1x Taq buffer, 2 mM MgCl₂ e 0,3 μM de cada primer em um volume total de 20 μL. Termociclador Veriti (Life Technologies, Carlsbad, CA) foi utilizado para amplificação dos fragmentos seguindo o protocolo: 95 °C por 2 min, com 35 ciclos de 95 °C por 30 s, temperatura de anelamento por 1 min, 72 °C por 1 min, e uma extensão final de 7 min a 72 °C. Para determinar a temperatura de anelamento de cada primer foi testada inicialmente gradiente de 52 °C a 60 °C. Os produtos da amplificação foram quantificados em gel de agarose 1,5 % corado com GelRed (*Biotium*®) por comparação com o padrão de peso molecular 1Kb DNA ladder (Invitrogen™).

Tabela 1. Listagem de iniciadores de *E. uniflora* para sequenciamento de fragmentos do genoma plastidial de *E. uniflora*.

Iniciadores	Amplificon (pb)	Sequência	Tm °C	Comprimento (pb)
Pitanga_01	614	Forward TGTCTAATGGATAGGAC AGAGGTC	58,7	24
		Reverse CGGGAACGAATGGTTA TCTT	58,9	20
Pitanga_02	794	Forward AATTTAGTAAACTCTTG GGATCTTTTC	58,1	27
		Reverse TGGTTTGGAAATGCATTT ATCTT	58,5	22
Pitanga_03	740	Forward TCGGGTTGTGAGACAC ATTC	59,5	20
		Reverse TCTGTTCAGGGCGATTC C	59,7	18

Fonte: Autor

2.3.3 Purificação com Polietileno Glicol (PEG)

Após a reação de amplificação, os produtos foram purificados, com a função de remover os iniciadores direto, reverso e dNTPs não incorporados durante a reação de amplificação. Esta purificação é essencial, pois a reação subsequente (sequenciamento de DNA – terminação de cadeia) acontece na presença de apenas um iniciador. A cada reação, foram adicionados 20,0 µL de PEG 800 20% contendo cloreto de sódio 2,5 M. As reações foram levemente agitadas (em Vortex®) e incubadas por 30 min em incubadora a 37 °C. Após, as reações foram centrifugadas em temperatura ambiente durante 15 min a

15.115 g. O sobrenadante foi removido e foi adicionado 125,0 µL de etanol 80% (-20°C) para a lavagem do pellet. As reações foram centrifugadas em temperatura ambiente durante 8 min a 15.115 g. O sobrenadante foi novamente removido e em seguida foram adicionados 125,0 µL de etanol 80% (-20°C) para a lavagem do pellet. As reações foram novamente centrifugadas em temperatura ambiente durante 8 min a 15.115 g. O sobrenadante foi removido e o pellet formado secou em incubadora a 37 °C. Após esta etapa, o pellet foi ressuscitado (durante 30 min a 37 °C) em 10,0 µL de água UltraPure™ DNase/RNase-Free Invitrogen™. O produto de PCR purificado foi posteriormente quantificado em gel de agarose 0,8%, corado com GelRed (Biotium®), por comparação com o padrão de peso molecular 1 kb DNA ladder (Invitrogen™)

2.3.4 Reação de sequenciamento pelo Método Sanger e purificação pós-reação de sequenciamento

Após a otimização, a amplificação e a purificação, os produtos de amplificação dos três pares iniciadores plastidiais foram sequenciados, pelo método de terminação de cadeia (SANGER et al. 1977) utilizando o Kit BigDye® Terminator v3.1 (Applied Biosystems™) em um sequenciador de DNA automático ABI 3500XL (Applied Biosystems™). As reações de sequenciamento de DNA foram realizadas em um volume total de 10,0 µL, contendo: 1,0 µL (estocados a 10,0 ng) de produto de amplificação purificado da primeira reação de amplificação; 0,3 µL de iniciador (estocado a 10,0 µM); 4,0 µL do mix BigDye® Terminator v3.1 (Applied Biosystems™), e 5,35 µL de água UltraPure™ DNase/RNase-Free Invitrogen™. Termociclador Veriti (Life Technologies, Carlsbad, CA) foi utilizado para amplificação da reação de sequenciamento, seguindo o protocolo: 96 °C por 2 min, com 35 ciclos de 96 °C por 10 s, temperatura de anelamento por 15 min, 60 °C por 4 min.

Após a reação de sequenciamento foi realizada uma nova purificação para remover iniciadores, dNTPs e ddNTPs não incorporados nos amplicons. A cada reação de sequenciamento, foi adicionado 2,0 µL de acetato de amônio 7,5 M, 2,0 µL de EDTA 125 mM e 50,0 µL de etanol absoluto (100%) (-20 °C). Esse procedimento foi realizado em placa de PCR (Applied Biosystems™). Cada placa foi

agitada levemente para homogeneizar a solução. As placas foram mantidas por 15 min, no escuro em geladeira (4,0 °C) para permitir a precipitação dos ácidos nucleicos. Após isso, foram centrifugadas a 4,0 °C durante 30 min a 3.000 g. O excesso de isopropanol foi descartado e as placas foram invertidas em papel toalha. Após, foi adicionado 70 µL de etanol 70% (-20 °C) para a lavagem do pellet. As reações foram novamente centrifugadas a 4,0 °C durante 15 min a 1.650 g. O sobrenadante foi novamente removido por inversão e logo após breve centrifugação durante 1 min a 185 g, com a placa invertida em papel toalha. O pellet foi seco a temperatura ambiente e no escuro. Após esta etapa, o pellet foi ressuscitado por 60 min em 10,0 µL de formamida HiDi™ (Applied Biosystems™). Após, as placas foram desnaturadas por 5min a 95°C em termociclador e logo após incubada em gelo por 3 min. Após esta etapa as placas foram colocadas em sequenciador automático ABI 3500XL (Applied Biosystems™).

2.3.5 Anotação do genoma plastidial de *E. uniflora*

O genoma plastidial de *E. uniflora* foi anotado inicialmente utilizando o programa Dual Organellar GenoMe Annotator DOGMA (WYMAN et al. 2004). Os genes previamente anotados foram confirmados utilizando o BLAST (ALTSCHUL et al. 1990). Os genes de tRNAs foram preditos utilizando o programa tRNAscan-SE (LOWE & EDDY, 1997). O mapa circular do cpDNA foi desenhado com o auxílio do programa OrganellarGenomeDRAW (LOHSE et al. 2013). A ordem dos genes foi verificada por meio do alinhamento dos genomas plastidiais de *E. uniflora* (KY392761 e KR867678) pelo programa Mauve versão 2.4.0 (DARLING et al. 2004), este programa foi utilizado para identificar combinações no genoma em blocos localmente colineares (LCBs) que permite avaliar rearranjos, grandes inserções ou deleções e divergência entre sequências. O programa PROtein MUMmer (PROmer) Perl script em MUMmer 3.0 (KURTZ et al. 2004), disponível em <http://Mummer.sourceforge.net>, foi utilizado para visualizar a sintenia entre *E. uniflora* (KY392761 e KR867678). A região SSC, da espécie sequenciada neste estudo, está com a orientação invertida, isto é considerado usual, já que, as duas orientações da região SSC ocorrem constantemente durante a replicação do cpDNA Walker et al. 2015. Porém para a realização dessas análises a região SSC foi reorientada.

2.4 Estruturas Repetitivas e SNPs

Para a identificação e localização das sequências diretas (*Forward*) e invertidas repetidas (palindrômicas) no genoma do cloroplasto foi utilizado o programa REPuter (KURTZ & SCHLEIERMACHER, 1999) considerando um tamanho mínimo ≥ 30 pb e uma distância Hamming de 3, ou seja, uma identidade de sequência de 90% ou superior.

A identificação dos polimorfismos de nucleotídeo único (SNP) putativos e de inserções e deleções (*indels*) foi realizada com o auxílio do programa CLCbios Genomics Workbench (versão 8.0.1). Foram utilizados, para identificação das marcas, os seguintes parâmetros: frequência mínima de alelos (MAF) de 60%, cobertura mínima de 35X e composição mínima de *contigs* de 10 *reads*.

3. RESULTADOS

3.1 Montagem do genoma plastidial e conteúdo gênico

O sequenciamento do genoma plastidial de *E. uniflora* foi realizado utilizando *Next-Generation Sequencing* (NGS). A montagem dos *reads* das sequências de nucleotídeos foi realizada por meio do método de montagem *de novo*, sem necessidade de espécie como referência. A cobertura média do genoma plastidial foi cerca de 160X (Tabela 2), que consideramos uma alta cobertura. As três regiões do genoma plastidial que apresentavam baixa cobertura foram resolvidas pelo método de sequenciamento Sanger como mostra a Figura 1. As sequências das três regiões apresentam 557 pb, 757 pb e 963 pb, respectivamente. A primeira sequência corresponde aos genes *trnR-UCU* e *atpA*, enquanto a segunda e a terceira sequências correspondem a regiões intergênicas (IGS).

Tabela 2. Resultado do sequenciamento e montagem de novo do cpDNA de *E. uniflora*.

Espécie	<i>E. uniflora</i>
Tamanho do cpDNA (pb)	158,68
Conteúdo GC %	36,95
Total de <i>reads</i> contados	173,31
Média do comprimento dos <i>reads</i> (pb)	148,23
<i>Reads</i> mapeados alinhados em pares	137,60
Cobertura do cpDNA *	160,31

*Cobertura: nº de bases sequenciadas/tamanho estimado do cpDNA.

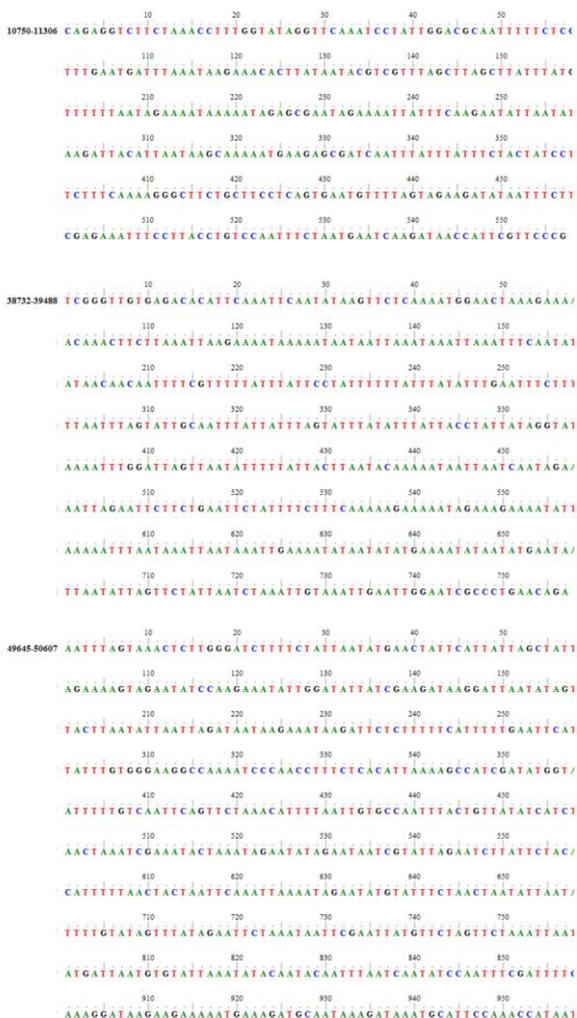
Fonte: Autor

O genoma plastidial de *E. uniflora* apresentou um tamanho de 158,680 pb de comprimento (Tabela 3), incluindo: uma região longa de

cópia única (LSC) de 87,495 pb (que corresponde a 55,1% do genoma plastidial); uma região curta de cópia única (SSC) de 18,537 pb (11,7% do genoma plastidial); e um par de sequências invertidas repetidas (IRa e IRb) com 26,324 pb cada (Figura 2; Tabela 4). As regiões LSC e SSC correspondem a 55,1% e 11,7% do genoma plastidial, respectivamente (Tabela 2). O conteúdo GC total do genoma foi de 37,0 %. As regiões LSC, SSC e IR apresentaram 34,8 %; 30,7 % e 42,8 % do conteúdo GC das respectivas regiões do genoma plastidial (Tabela 3).

Um total de 115 diferentes genes foi identificado, sendo que 78 codificam para proteínas, 30 genes de tRNA e quatro genes de rRNA (Tabela 3). Um total de 20 genes encontra-se duplicado nas IR, dentre esses, sete genes de tRNA (*trnN-GUU*; *trnR-ACG*; *trnA-UGC*; *trnI-GAU*; *trnV-GAC*; *trnL-CAA*; *trnI-CAU*), seis genes codificadores de proteínas (*rps12*; *rps7*; *ndhB*; *ycf2*; *rpl23*; *rpl2*), três *pseudogenes* (*ycf68*; *ycf1*; *ycf15*) e quatro genes de rRNA (*rrn16*; *rrn23*; *rrn4.5*; *rrn5*). Além disso, no genoma plastidial de *E. uniflora* 18 genes contêm ítrons (Tabela 5).

Figura 1. Sequências de três regiões com 557 pb, 757 pb e 963 pb, respectivamente, resolvidas pelo método de sequenciamento Sanger no genoma plastidial de *E. uniflora*.



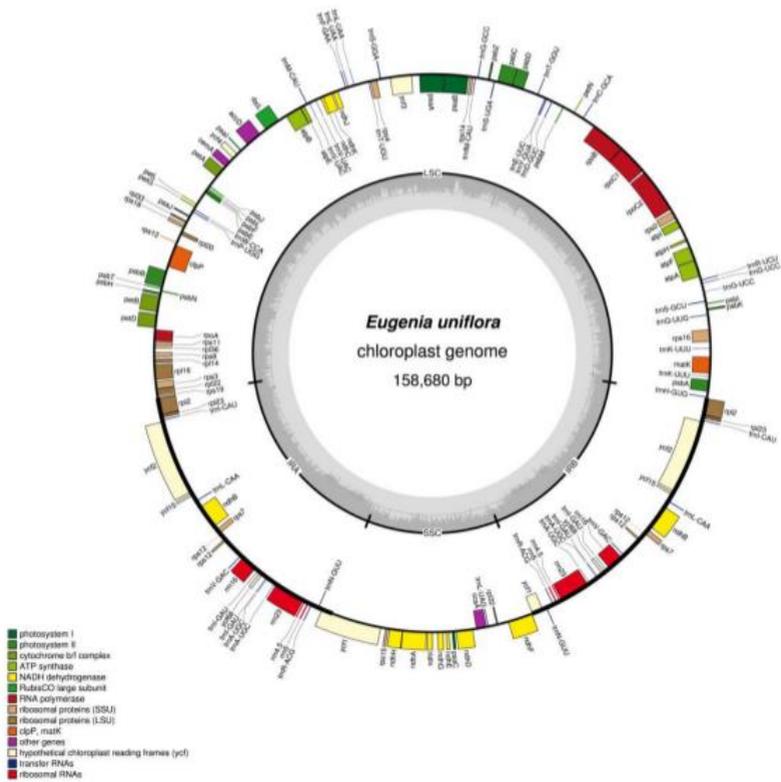
Fonte: Autor

Tabela 3. Resumo das características do genoma plastidial de *E. uniflora*.

Parâmetros	<i>E. uniflora</i>
Tamanho do cpDNA (Kb)	158680
LSC tamanho em pb (%)	87.495 (55,1%)
SSC tamanho em pb (%)	18.537 (11,7%)
IR tamanho em pb	26.324
Genes diferentes	115
Diferentes GCP	78
Diferentes genes de tRNA	30
Diferentes genes de rRNA	4
Diferentes genes duplicados em IR	20
Diferentes genes com introns	18
Conteúdo total % GC	37,0
% GC LSC	34,8
% GC SSC	30,7
% GC IR	42,8

Fonte: Autor

Figura 2. Organização dos genes do genoma do cloroplasto de *E. uniflora*. Genes fora do círculo são transcritos no sentido horário e genes no interior o círculo são transcritos no sentido anti-horário. O círculo central representa as diferentes regiões (LSC, SSC e duas IR) do genoma plastidial.



Fonte: Autor

Tabela 4. Lista de genes identificados no genoma plastidial de *E. uniflora*.

Categoria funcional	Grupo de genes	Nome dos genes													
Auto - replicação	Subunidade ribossomal grande	<i>rpl14</i>	<i>rpl16</i>	<i>rpl2^{ab}</i>	<i>rpl20</i>	<i>rpl22</i>	<i>rpl23^b</i>	<i>rpl32</i>	<i>rpl33</i>	<i>rpl36</i>					
	DNA polimerase dependente de RNA	<i>rpoA</i>	<i>rpoB</i>	<i>rpoC1^a</i>	<i>rpoC2</i>										
	Subunidade ribossomal pequena	<i>rps11</i>	<i>rps12^{a*}</i>	<i>rps14</i>	<i>rps15</i>	<i>rps16^a</i>	<i>rps18</i>	<i>rps19</i>	<i>rps2</i>	<i>rps3</i>					
		<i>rps4</i>	<i>rps7^b</i>	<i>rps8</i>											
	Genes do RNA ribossomal	<i>rrm1^b</i>	<i>rrm2^{3b}</i>	<i>rrm4.5^b</i>	<i>rrm5^b</i>										
		<i>trnA-UGC</i>	<i>trnC-</i>	<i>trnD-</i>	<i>trnE-</i>	<i>trnF-</i>	<i>trnG4-</i>	<i>trnG-</i>	<i>trnG-</i>	<i>trnH-</i>					
		^a	<i>GCA</i>	<i>GUC</i>	<i>UUC</i>	<i>GAA</i>	<i>CAU</i>	<i>GCC</i>	<i>UCC^a</i>	<i>GUG</i>					
	Genes para RNA transportadores	<i>trnI-GAU</i>	<i>trnK-</i>	<i>trnL-</i>	<i>trnL-</i>	<i>trnL-</i>	<i>trnL-</i>	<i>trnM-</i>	<i>trnN-</i>	<i>trnP-</i>					
		<i>trnI-CAU^b</i>	^{ab}	<i>UUU^a</i>	<i>CAA^a</i>	<i>UAA^a</i>	<i>UAG</i>	<i>CAU</i>	<i>GUU^b</i>	<i>UGG</i>					
		<i>trnR-</i>	<i>trnR-</i>	<i>trnS-</i>	<i>trnS-</i>	<i>trnS-</i>	<i>trnT-</i>	<i>trnT-</i>	<i>trnV-</i>	<i>trnV-</i>					
		<i>trnQ-UUG</i>	<i>ACG^b</i>	<i>UCU</i>	<i>GCU</i>	<i>GGA</i>	<i>UGA</i>	<i>GGU</i>	<i>UGU</i>	<i>GAC^b</i>					
		<i>trnW-</i>	<i>trnW-</i>	<i>trnY-</i>	<i>GUA</i>										
	Fator de iniciação de tradução	<i>infA^c</i>													
	Genes relacionados a fotossíntese	Subunidade do fotossistema I	<i>psaA</i>	<i>psaB</i>	<i>psaC</i>	<i>psaI</i>	<i>psaJ</i>	<i>ycf3^a</i>	<i>ycf4</i>						
		Subunidade do fotossistema II	<i>psbA</i>	<i>psbB</i>	<i>psbC</i>	<i>psbD</i>	<i>psbE</i>	<i>psbF</i>	<i>psbH</i>	<i>psbI</i>	<i>psbJ</i>				
<i>psbK</i>			<i>psbL</i>	<i>psbM</i>	<i>psbN</i>	<i>psbT</i>	<i>psbZ</i>								
Subunidade do citocromo b 6/f		<i>petA</i>	<i>petB^a</i>	<i>petD^b</i>	<i>petG</i>	<i>petL</i>	<i>petN</i>								
Subunits of ATP synthase		<i>atpA</i>	<i>atpB</i>	<i>atpE</i>	<i>atpF^a</i>	<i>atpH</i>	<i>atpI</i>								
Grande subunidade do Rubisco		<i>rbcL</i>													
Subunidades da NADH desidrogenase		<i>ndhA^a</i>	<i>ndhB^{a,b}</i>	<i>ndhC</i>	<i>ndhD</i>	<i>ndhE</i>	<i>ndhF</i>	<i>ndhG</i>	<i>ndhH</i>	<i>ndhI</i>					
		<i>ndhJ</i>	<i>ndhK</i>												
Outros genes		Subunidade de acetil-CoA-carboxilase	<i>accD</i>												
		Gene que sintetiza o citocromo c	<i>ccsA</i>												
	Envelope de proteína de membrana	<i>cemA</i>													
	Protease	<i>clpP^a</i>													
	Maturase	<i>matK</i>													
Componente do complex TIC	<i>ycf7^{b,c*}</i>														
Genes com funções desconhecidas	Função desconhecida	<i>ycf2^b</i>	<i>ycf15^{b,c}</i>	<i>ycf68^{b,c}</i>											

^a Gene contendo íntron; ^b Genes duplicados; ^c Pseudogene; ^{c*} *ycf1* é pseudogene na fronteira entre as regiões IRb e SSC; ^{*}*rps12* é trans-splicing com a extremidade 5' localizada na região LSC e a extremidade 3' duplicada na região IR.

Fonte: Autor

Tabela 5. Lista de genes contendo íntrons no genoma plastidial de *E. uniflora*.

Região	Gene	Exon I (pb)	Intron I (pb)	Exon II (pb)	Intron II(pb)	Exon III (pb)
LSC	rps16	206	869	38		
LSC	rpoC1	1718	613	446		
LSC	atpF	410	743	146		
LSC	petB	5	766	647		
LSC	petD	8	702	524		
LSC	rp116	401	998	8		
LSC	ycf3	152	727	227	759	125
LSC	clpP	227	620	290	871	68
LSC	trnK-UUU	34	2536	36		
LSC	trnG-UCC	22	756	48		
LSC	trnV-UAC	36	596	38		
LSC	trnL-UAA	36	507	49		
SSC	ndhA	563	1046	539		
LSC/IRs	rps12 *	113	–	209	–	26
IR	rp12	434	682	392		
IR	ndhB	755	695	776		
IR	trnI-GAU	36	958	34		
IR	trnA-UGC	37	804	34		

*rps12 é trans-splicing da extremidade 5' na região LSC com a extremidade 3' duplicada nas regiões IR

Fonte: Autor

3.2 Variação intraespecífica do genoma plastidial de *E. uniflora*

Na comparação do genoma plastidial de *E. uniflora* sequenciado neste trabalho com o genoma plastidial da espécie de *E. uniflora* (KR867678), disponível no GenBank, foram encontrados 29 possíveis polimorfismos de nucleotídeo único (SNPs) em genes codificadores de proteínas, sete SNPs em íntrons e dois em regiões intergênicas (IGS) (Tabela 6). O total de 38 possíveis SNPs no do genoma plastidial de *E. uniflora* representa uma frequência de 1 a cada 4,2 kb (Tabela 7). Além disso, foram encontrados 11 possíveis inserções/deleções (*indels*) em IGSs, três em genes codificadores de proteínas, seis em íntrons e dois *indel* em genes de RNA transportadores (Tabela 7). A posição dos possíveis SNPs e *indels* encontrados por meio da comparação intraespecífica do genoma plastidial de *E. uniflora* foi listada na Tabela 8.

Tabela 6. Variações intraespecíficas do genoma plastidial de *E. uniflora*.

Tipo de variação	Regiões do genoma plastidial	<i>E. uniflora</i>
Nucleotídeo de polimorfismo único (SNP)	Genes codificadores de proteínas	29
	Íntron	7
	Genes tRNA	0
	Genes rRNA	0
	Regiões intergênicas (IGS)	2
Total de SNPs		38
Inserção/Deleção	Genes codificadores de proteínas	3
	Íntron	6
	Genes tRNA	0
	Genes rRNA	0
	Regiões intergênicas (IGS)	2
Total de <i>indels</i>		11
S/I regiões codificantes		9,7

A relação entre o número de substituições (SNP) e entre o número de *indels* (S/I) foi de 9,7 em regiões codificantes e de 1,1 em regiões não codificantes (Tabela 6). Além disso, a quantidade de mutações encontradas nos possíveis SNPs do tipo transição (substituição purina <-> purina ou pirimidina <-> pirimidina) foi mais abundante que a do tipo transversão (substituição, purina <-> pirimidina) (Tabela 7). A frequência de SNP/Kb no genoma plastidial de *E. uniflora* de substituições do tipo transição foi de 0,15 SNPs/Kb, enquanto que as substituições do tipo transversão ocorrem com a frequência de 0,09 por 1 SNPs/Kb. A relação Ti/Tv, caracterizada pela divisão do número de transições (Ti) dividido pelo número de transversões (Tv) de bases nitrogenadas, foi de 1,71 no genoma plastidial de *E. uniflora* (Tabela 7).

Tabela 7. Diferenças no número e frequência de SNPs e tipo de mutação intraespecífica no genoma plastidial de *E. uniflora*.

<i>E. uniflora</i>	
Número total de SNPs	38
Frequência de SNP/Kb	0,23
Transições	
A <-> G (purina <-> purina)	10 (26,3%)
C <-> T (pirimidina <-> pirimidina)	14 (36,9%)
Total de transições	24 (63,2%)
Frequência de Ti/Kb	0,15
Transversões	
A <-> T	3 (7,9%)
A <-> C	4 (10,5%)
G <-> C	3 (7,9%)
G <-> T	4 (10,5%)
Total de transversões	14 (36,8%)
Frequência de Tv/Kb	0,09
Ti/Tv	1,71

SNP: polimorfismos de nucleotídeo único; Transição (Ti): base púrica x base púrica; base pirimídica x base pirimídica; Transversão (Tv): base púrica x base pirimídica.

Fonte: Autor

Tabela 8. Localização dos SNPs e *indels* entre os genomas plastidiais de *E. uniflora*.

Posição no genoma (pb)	Substituição	<i>E. uniflora</i> (KR867678)	<i>E. uniflora</i> (KY392761) ^a	Região no genoma
2908		C	T	<i>matK</i>
5325		G	A	Íntron <i>rps16</i>
5905		T	C	Íntron <i>rps16</i>
13613		G	A	Íntron <i>atpF</i>
13955		A	G	<i>atpF</i>
19628		C	G	<i>rpoC2</i>
22467		C	G	<i>rpoC1</i>
23760		A	T	<i>rpoC1</i>
37763		C	T	<i>psbC</i>
43435		C	A	<i>psaA</i>
46773		G	T	Íntron II <i>ycf3</i>
53541		G	A	<i>ndhK</i>
54028	SNP	T	C	<i>ndhC</i>
56352		G	C	<i>atpB</i>
59679		A	T	<i>rbcL</i>
60657		C	A	<i>accD</i>
60791		C	T	<i>accD</i>
74091		A	T	Íntron I <i>clpP</i>
82113		T	C	<i>rpoA</i>
86042		A	C	<i>rps3</i>
114737		G	A	<i>ycf1</i>
117981		G	A	<i>ycf1</i>
118737		C	T	<i>ycf1 - rps15</i>
119212		G	T	<i>ndhH</i>
119272		T	C	<i>ndhH</i>

119662		A	G	<i>ndhH</i>
119800		C	T	<i>ndhH</i>
120053		A	G	<i>ndhH</i>
120973		C	T	Ínton <i>ndhA</i>
121368		T	G	Ínton <i>ndhA</i>
124527		A	C	<i>ndhE</i>
126375		G	T	<i>ndhD</i>
126987		T	C	<i>ndhD</i> - <i>ccsA</i>
130435		G	A	<i>ndhF</i>
130843		C	T	<i>ndhF</i>
131156		G	A	<i>ndhF</i>
131203		C	T	<i>ndhF</i>
131977		C	T	<i>ndhF</i>
<hr/>				
45944		T	-	Ínton I <i>ycf3</i>
47420		A	-	Ínton II <i>ycf3</i>
74079		A	-	Ínton I <i>clpP</i>
124796		T	-	<i>ndhE-psaC</i>
124891		TAGGTC	-	<i>psaC</i>
132100	<i>Indel</i>	TTTAAATA	-	<i>ndhF</i>
5678		T	-	<i>trnK-UUU</i> - <i>rps16</i>
47321		-	A	Ínton II <i>ycf3</i>
53692		-	T	<i>ndhK</i>
73896		-	TTT	Ínton I <i>clpP</i>
74602		-	TT	Ínton II <i>clpP</i>

^a Espécie sequenciada neste estudo

Fonte: Autor

4. DISCUSSÃO

4.1 Montagem do genoma plastidial e conteúdo gênico de *E. uniflora*

O genoma plastidial de *E. uniflora* é similar ao de outras Myrtaceae como mostrou o estudo realizado por Machado et al. (2017). A comparação intraespecífica do genoma plastidial de *E. uniflora* apresentou diferenças tais como, *E. uniflora* (KY392761) apresentou o tamanho do genoma plastidial de 158.680 pb enquanto o indivíduo *E. uniflora* (KR867678) tem o tamanho do genoma de 158.445 pb (EGUILUZ et al. 2015) (Tabela 9), uma diferença de 235 pb entre os dois genomas plastidiais. A LSC, SSC e IRs dos dois indivíduos apresentaram diferenças de 36 pb e 219 pb e 10 pb, respectivamente (Tabela 9). Além disso, as diferenças entre os dois genomas pode ser visualizada na Figura 3. Comparativamente, o genoma plastidial sequenciado no presente estudo é muito similar aos já sequenciados de outras Myrtaceae (STEANE, 2005; PAIVA et al. 2011; ASIF et al. 2013; BAYLY et al. 2013; MACHADO et al. 2017).

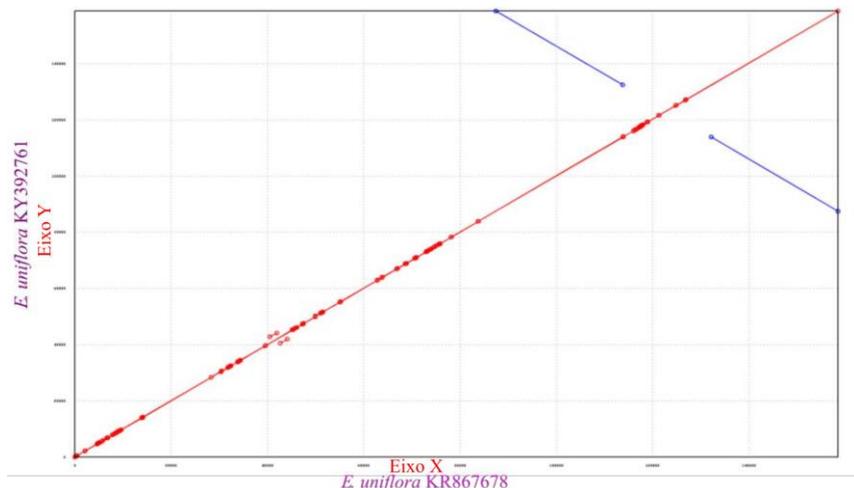
Tabela 9. Características do genoma plastidial de dois indivíduos de *E. uniflora*.

Espécie	Acesso	Comprimento (pb)	LSC em pb (%)	SSC em pb (%)	IR em pb
<i>Eugenia uniflora</i>	KY392		87.495	18.537	
	761 ^a	158.680	(55,1)	(11,7)	26.324
	KR8676		87.459	18.318	
	78	158.445	(55,2)	(11,6)	26.334

^a Espécie com o genoma plastidial sequenciado neste trabalho

Fonte: Autor

Figura 3. Análises pontuais da sequência do genoma plastidial de *E. uniflora* KY392761 contra *E. uniflora* KR867678. A inclinação positiva indica que as duas sequências estão alinhadas na mesma orientação. O gráfico representa a comparação entre *E. uniflora* KR867678 (eixo X) e *E. uniflora* KY392761 (eixo Y).



Fonte: Autor

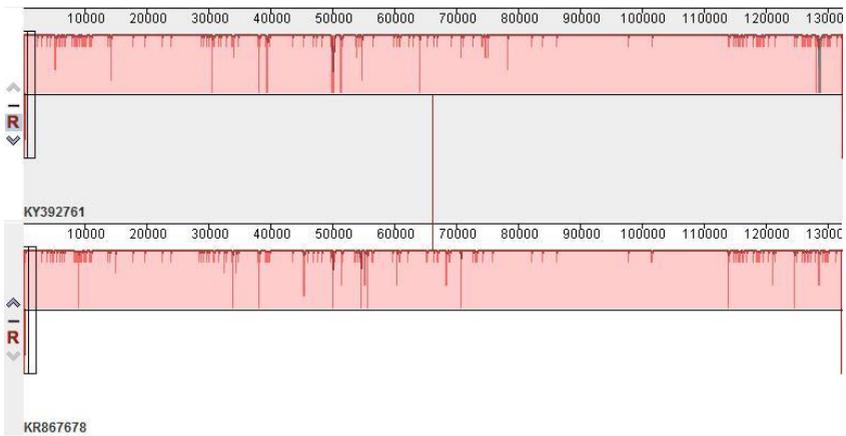
E. uniflora apresenta a mesma ordem gênica que outras Myrtaceae, como *S. cumini* (ASIF et al. 2013), *E. grandis* (PAIVA et al. 2011), *E. globulus* (STEANE, 2005) e *A. sellowiana*, como mostra no estudo realizado por Machado et al. (2017).

A análise do conteúdo gênico dos genomas plastidiais dos dois indivíduos de *E. uniflora* é apresentada na Figura 4. Os genes *infA* e *yef15* (pseudogene) são considerados altamente conservados em plantas com flores (RAUBESON et al. 2007; SHI et al. 2013); porém, para muitas espécies, incluindo *E. uniflora*, esses dois genes são considerados não-funcionais, pois possuem *stop* códons prematuros.

Além desses dois genes mencionados anteriormente, o gene *yef68* presente no indivíduo de *E. uniflora* sequenciado neste estudo, também foi considerado como gene não-funcional, assim como, em

outras angiospermas, tais como a maioria das Rosídeas (RAUBESON et al. 2007) e *Vigna radiata* (TANGPHATSORNRUANG et al. 2010). Em *A. sellowiana* (MACHADO et al. 2017) o *ycf68* também foi considerado pseudogene.

Figura 4. Alinhamento múltiplo do genoma plastidial de dois indivíduos de *E. uniflora* utilizando o programa MAUVE. A presença de apenas um bloco localmente colinear, representado na cor rosa, indica a região de DNA homóloga, sem rearranjos de sequência.



Fonte: Autor

O conteúdo de bases AT foi de aproximadamente 63% para os genomas plastidiais dos dois indivíduos de *E. uniflora*, valor similar aos encontrados em outros genomas plastidiais de Myrtaceae, como *S. Cumini* (63,2%) (ASIF et al. 2013) e *Eucalyptus corymbia* (63,0%) (BAYLY et al. 2013), *A. sellowiana* (63,0) (MACHADO et al. 2017). O conteúdo de AT% varia de 61,3% nas gramíneas a 62,9% nas dicotiledôneas sendo que a frequência de AT é menor nas IR que nas regiões de cópia única (SSC e LSC), este elevado conteúdo de AT tem sido considerado um dos fatores que facilitam a transferência de genes do genoma cloroplastidial para o genoma nuclear (HOWE et al. 2003). Neste sentido, o padrão de evolução dos genomas plastidiais de

Myrtaceae, em termos de transferência de genes do cloroplasto para o núcleo, foi similar as demais angiospermas.

4.2 Variação intraespecífica do genoma plastidial de *E. uniflora*

Quando compara-se os dois genomas plastidiais de *E. uniflora* observa-se que a quantidade de SNPs em regiões IGSs foi menor que em regiões codificadoras, discordando do trabalho de Doorduyn et al. (2011), o qual afirma que SNPs são encontrados 1,8 vezes mais frequentes em IGS e íntrons do que em regiões codificadoras; porém foi similar aos resultados de Machado et al. (2017), pois em *A. sellowiana* a quantidade de SNPs em IGS foi menor que em regiões codificantes de proteínas.

A frequência de SNPs/Kb em *E. uniflora* foi de 0,23, enquanto que em *A. sellowiana* esse valor foi de 0,53 (MACHADO et al. 2017). As frequências de SNPs/Kb de *E. uniflora* e *A. sellowiana* são consideradas mais baixas que outras angiospermas, tais como em *Glycine max*, onde foi observada uma frequência de 3 SNPs/Kb, e em *Populus tremula*, onde foi observada uma frequência de 60 SNPs/Kb.(SEBASTIANI et al. 2007). Assim, podemos inferir que quanto mais proximamente relacionadas às espécies ou quando são efetuadas comparações intraespecíficas, a frequência SNP/Kb tende a ser mais baixa.

O conhecimento do processo padrão de substituições de bases nitrogenadas é importante, pois, por meio do número de substituições entre as sequências de DNA pode-se inferir relações filogenéticas que incorporam modelos evolutivos de sequência de DNA. A relação Ti/Tv foi de 1,71 para *E. uniflora*, valor mais elevado do que foi constatado em *A. sellowiana* (MACHADO et al. 2017), de 1,36. Altas taxas de substituições do tipo transição apresentam valores de relação Ti/Tv entre dois e 10 (BAKKER et al. 2000, PURVIS et al. 1997), enquanto altas taxas de transversões apresentam valores próximos de um. No estudo envolvendo DNA organelar em *Cycas revoluta* e *Cycas taitungensis*, os valores da relação Ti/Tv foram de 1,808 e 1,707 para mtDNA e cpDNA, respectivamente (CHIANG et al, 2009), o que sugere baixa probabilidade de saturação. Neste cenário, os sítios observados mudaram tanto que retornaram ao estado ancestral (PAGE & HOLMES, 1998). As frequências de transições/Kb (Ti/Kb) e transversões/Kb (Tv/Kb) foram de 0,15 e 0,09, respectivamente, nos genomas plastidiais

de *E. uniflora*, valores bem menores que 0,30 e 0,23 constatados em *A. sellowiana* (MACHADO et al. 2017). Esses resultados mostram que a diferença entre essa relação considerando uma análise intraespecífica é similar aos resultados de uma análise interespecífica ou intragenérica, já que Machado et al. (2017) compararam os genoma plastidiais de *A. sellowiana* e *E. uniflora* (KR867678).

A quantidade de *indels* em genes de regiões codificadoras de proteínas foi relativamente menor do que em regiões não-codificadores, enquanto a quantidade de SNPs foi relativamente mais alta (Tabela 6). Este resultado corrobora com o proposto por Bayly et al. (2013), que menciona que *indels* em genes de regiões codificadoras de proteínas são relativamente baixos, sendo a maior parte da variação constituída pelos SNPs.

A relação S/I foi de 9,7 em regiões codificantes e de 1,1 em regiões não-codificantes de proteínas. Esses resultados corroboram com Chen et al. (2009), que afirmam que a relação S/I em seqüências de regiões codificadoras é maior do que em regiões não-codificantes, o que indica uma constância universal na ocorrência relativa dos dois tipos de mutações.

Neste estudo, foi sequenciado o genoma plastidial de *E. uniflora* que foi comparado com o genoma plastidial de um outro indivíduo da mesma espécie quanto a estrutura do genoma, incluindo tamanho, conteúdo gênico, polimorfismos (SNP e *indels*) e variações de substituição de nucleotídeos. Os resultados destes dois sequenciamentos mostraram que os genomas plastidiais dos dois indivíduos analisados contêm diferenças, que permite inferir que os indivíduos analisados são geneticamente distintos, mas da mesma espécie.

Além disso, constatamos que a diferença entre as relações Ti/Kb e Tv/Kb em análises intraespecífica foi similar aos resultados de análises interespecíficas ou intragenéricas. Assim, pode-se também inferir que quanto mais proximamente relacionadas são as espécies ou quando são realizadas comparações intraespecíficas, a frequência SNP/Kb tende a decrescer. Entretanto, estudos filogenéticos utilizando seqüências do genoma plastidial geralmente baseiam-se em regiões intergênicas, onde a variabilidade é maior. Neste sentido, quando comparamos as variações entre as regiões intergênicas dos genomas plastidiais de *E. uniflora* verificamos que essa variabilidade foi baixa tanto para SNP quanto para *indel*. Portanto, pode ser sugerido que análises utilizando o genoma completo são mais eficientes que àquelas

que utilizam parte do genoma em estudos evolutivos. Além disso, esses polimorfismos (possíveis SNPs) identificados neste estudo poderão ser validados e avaliados em futuros estudos de genética de populações envolvendo essa importante espécie frutífera.

REFERÊNCIAS BIBLIOGRÁFICAS

ARAI, I.; AMAGAYA, S.; KOMATSU, Y.; OKADA, M.; HAYASHI, T.; et al. Improving effects of the extracts from *Eugenia uniflora* on hyperglycemia and hypertriglyceridemia in mice. *Journal of Ethnopharmacology* 68: 307–314. 1999.

ASIF, H.; KHAN, A.; IQBAL, A.; KHAN, I.A.; HEINZE, B.; AZIM, M.K.; The chloroplast genome sequence of *Syzygium cumini* (L.) and its relationship with other angiosperms. ***Tree Genetics & Genomes***. 9:867–877. 2013.

BAYLY, M.J.; RIGAULT, P.; SPOKEVICIUS, A.; LADIGES, P.Y.; ADES, P.K.; ANDERSON, C.; BOSSINGER, G.; MERCHANT, A.; UDOVICIC, F.; WOODROW, I.E.; TIBBITS, J. Chloroplast genome analysis of Australian eucalypts – *Eucalyptus*, *Corymbia*, *Angophora*, *Allosyncarpia* and *Stockwellia* (Myrtaceae). ***Molecular Phylogenetics and Evolution***. 69,704–716. 2013.

BOURSCHEID K, VIEIRA, N. K. LISBÔA G N., KINUPP V F. BARROS, I.B.I. O. Grupos de Uso e as Espécies Prioritárias: *Eugenia uniflora*. In: CORADIN, L.; SIMINSKI, A.; REIS, A. ***Espécies nativas da Flora Brasileira de Valor Econômico Atual ou Potencial: Plantas para o Futuro- Região sul***. Brasília: MMA, 2011.170-177p.

CHEN, J.Q.; YING, W.; HAIWANG, Y.; JOY, B.; MARTIN, K.; DACHENG, T. Variation in the ratio of nucleotide substitution and indel rates across genomes in mammals and bacteria. ***Mol Biol Evol***. 26:1523–1531. 2009.

CHIANG, Y.C.; HUNG, K.H.; MOORE, S.J.; GE, X.J.; HUANG, S.; HSU, T.W.; SCHAAL, B.; CHIANG, T.Y. 2009. Paraphyly of organelle DNAs in *Cycas* sect. *Asiorientales* due to ancient ancestral polymorphisms. ***BMC Evolutionary Biology***. 9: 161.

DARLING, A.C.E.; MAU, B.; BLATTNER, F.R.; PERNA, N.T. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. ***Genome Research***. 1394–1403. 2004. doi:

10.1101/gr.2289704.

DENARDIN, C.C.; PARISI, M.M.; MARTINS, L.A.; et al. Antiproliferative and cytotoxic effects of purple pitanga (*Eugenia uniflora* L.) extract on activated hepatic stellate cells. **Cell Biochem Funct.** v.32:16–23. 2014.

DENARDIN, C.C.; MARTINS, L.A.M.; PARISI, M.M.; VIEIRA, M.Q.; TERRA, S.R.; BARBÉ-TUANA, F.M.; BOROJEVIC, R.; VIZZOTTO, M.; EMANUELLI, T.; GUMA, F.C.R. Autophagy induced by purple pitanga (*Eugenia uniflora* L.) extract triggered a cooperative effect on inducing the hepatic stellate cell death. **Cell Biol Toxicol.** DOI 10.1007/s10565-016-9366-5. 2016.

DOORDUIN, L.; GRAVENDEEL, B.; LAMMERS, Y.; ARIYUREK, Y.; CHIN-A-WOENG, T.; VRIELING, K. The Complete Chloroplast Genome of 17 Individuals of Pest Species *Jacobaea vulgaris*: SNPs, Microsatellites and Barcoding Markers for Population and Phylogenetic Studies. **DNA RES.** pp. 1–13. 2011.

EGUILUZ, M. M.; RODRIGUES, F. N.; GUZMAN, F.; YUYAMA, P.; MARGIS, R. The complete chloroplast genome sequence of neotropical Myrtaceae *Eugenia uniflora*: organization and phylogenetic relationships. In: 11th International Congress of Plant Molecular Biology. Foz do Iguacu, Brasil. 2015 p. 0487. 11th International Congress of Plant Molecular Biology, 2015.

FERREIRA-RAMOS, R.; LABORDA, P.R.; SANTOS, M.O.; MAYOR, M.S.; MESTRINER, M.A.; DE SOUZA, A.P.; ALZATE-MARIN, A.L. Genetic analysis of forest species *Eugenia uniflora* L. through of newly developed SSR markers. **Conserv Genet.** 9:1281–1285. 2008.

FERREIRA-RAMOS, ACCORONI, K.A.G.; ROSSI, A.; GUIDUGLI, M.C.; MESTRINER, M.A.; MARTINEZ, C.A.; ALZATE-MARIN, A.L. Genetic diversity assessment for *Eugenia uniflora* L., *E. pyriformis* Cambess., *E. brasiliensis* Lam. and *E. francavilleana* O. Berg neotropical tree species (Myrtaceae) with heterologous SSR markers. **Genetic Resources and Crop Evolution.** v. 61. 267-272. 2014.

FRANZÃO, A.A.; MELO, B.; Cultura da Pitangueira. núcleo de estudos em fruticultura no cerrado, Uberlândia, s. d. Disponível em: <<http://www.fruticultura.iciag.ufu.br/pitangueira.html>>. Acesso em 06 jan. 2017.

GOVAERTS, R.; SOBRAL, M.; ASHTON, P.; BARRIE, F.; HOLST, B.K.; LANDRUM, L.R.; MATSUMOTO, K.; MAZINE, F.F.; NIC LUGHADHA, E.; PROENÇA, C.; SOARES-SILVA, L.H.; WILSON, P.G.; LUCAS, E.J. World checklist of Myrtaceae. **Kew Publishing**, Royal Botanic Gardens, Kew. 2008.

LEDERMAN, I. E.; BEZERRA, J. E.F.; CALADO, G. **A pitangueira em Pernambuco**. Recife: Empresa Pernambucana de Pesquisa Agropecuária - IPA, 1992. 20p.

LOHSE, M.; DRECHSEL, O.; KAHLAU, S.; BOCK, R. OrganellarGenomeDRAW: a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. **Nucleic Acids Research**, v. 41, n. W575-81, 2013.

MACHADO, L.O.; VIEIRA, L.N.; STEFENON, V.M.; PREDROSA, F.O.; SOUZA, E.M.; GUERRA, M.P.; NODARI, R.O. Phylogenomic relationship of feijoa (*Acca sellowiana* (O.Berg) Burret) with other Myrtaceae based on complete chloroplast genome sequences. *Genetica*. doi: 10.1007/s10709-017-9954-1. 2017.

SANGER, F.; NICKLEN, S.; COULSON R. 1977. DNA sequencing with chain-terminating inhibitors. **Proceedings of the National Academy of Sciences of the United States of America** 74: 5463-5467.

MARGIS, R.; FELIX, D.; CALDAS, J.F.; SALGUEIRO, F.; ARAUJO D.S.D.; BREYNE, P.; VAN MONTAGU, M.; OLIVEIRA, D.; MARGIS-PINHEIRO, M. Genetic differentiation among three neighboring Brazil-cherry (*Eugenia uniflora* L.) populations within the Brazilian Atlantic rain forest. **Biodiversity and Conservation**. 11:149-163. 2002.

PAIVA, J.A.; PRAT, E.; VAUTRIN, S.; SANTOS, M.D.; SANCLEMENTE, H.; BROMMONSCHENKEL, S.; FONSECA, PG.; GRATTAPAGLIA, D.; SONG, X.; AMMIRAJU, J.S.; KUDRNA, D.; WING, R.A.; FREITAS, A.T.; BERGÈS, H.; GRIMA-PETTENATI, J. Advancing Eucalyptus genomics: identification and sequencing of lignin

biosynthesis genes from deep-coverage. BAC libraries. **BMC Genom.** 4(12):137. 2011.

UNTERGASSER, A.; NIJVEEN, H.; RAO, X.; BISSELING, T.; GEURTS, R.; LEUNISSEN, J.A.M. Primer3Plus, an enhanced web interface to Primer3. **Nucleic Acids Research.** 2007. 35: W71-W74; doi:10.1093/nar/gkm306.

RAUBESON, L.A.; PEERY, R.; CHUMLEY, T.W.; DZIUBEK, C.; FOURCADE, H.M.; et al. Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. **BMC Genomics** 8: 174. 2007.

SEBASTIANINI, F.; GONZÁLEZ-MARTÍNEZ, S.C. e VENDRAMIN, G.G. **Review on single nucleotide polymorphisms (SNPs) and population genetic studies in conifer species.** In: KOSKELA, J.; SAMUEL, C.J.A.; MÁTYÁS, Cs. e FADY, B. (Org.). *Conifer Network*. Bioersivity International. Roma, Itália. 2007.

STEANE, D.A. Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum *Eucalyptus globulus* (Myrtaceae). **DNA Res.**12(3):215–220. 2005.

SHI, C., LIU, Y., HUANG, H., XIA, E. H., ZHANG, H. B., AND GAO, L. Z. Contradiction between plastid gene transcription and function due to complex posttranscriptional splicing: an exemplary study of *ycf15* function and evolution in angiosperms. **PLoS ONE.** 8:e59620. doi: 10.1371/journal.pone.0059620. 2013.

TANGPHATSORNRUANG, S.; SANGSRAKRU, D.; CHANPRASERT, J.; UTHAIPAIWANWONG, P.; YOOCHA, T.; JOMCHAI, N.; et al. The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships. **DNA Res.** 17, 11–22. doi: 10.1093/dnares/dsp025. 2010.

SALGUEIRO, F.; FELIX, D.; CALDAS, J.F.; MARGIS-PINHEIRO, M.; MARGIS, R. Even population differentiation for maternal and biparental gene markers in *Eugenia uniflora*, a widely distributed species from the Brazilian coastal Atlantic rain Forest. **Diversity and Distributions.** 10:201-210. 2004.

SANTOS, K.K.A.; MATIAS, E.F.F.; TINTINO, S.R.; SOUZA, C.E.S.; BRAGA, M.F.B.M.; et al. Anti-Trypanosoma cruzi and cytotoxic activities of *Eugenia uniflora* L. *Experimental Parasitology*. 2012.

TURCHETTO-ZOLET, A.C.; SALGUEIRO, F.; CRUZ, F.; VETO, N.; MARGIS, R. Chloroplast DNA variation and phylogeography of *Eugenia uniflora* L. (Myrtaceae) in the Brazilian Atlantic forest. **BMC Proceedings**, v.5 (Suppl 7), p.19, 2011.

TURCHETTO-ZOLET, A.C.; SALGUEIRO, F.; TURCHETTO, C.; CRUZ, F.; VETO, N.M.; BARROS, M. J. F.; SEGATTO, A.L.A. FREITAS, L.B.; MARGIS, R. Phylogeography and ecological niche modelling in *Eugenia uniflora* (Myrtaceae) suggest distinct vegetational responses to climate change between the southern and the northern Atlantic Forest. **Botanical Journal of the Linnean Society**, 2016.

WALKER, J.F.; JANSEN, R.K.; ZANIS, M.J.; EMERY, N.C. Sources of inversion variation in the small single copy (SSC) region of chloroplast genomes. Letter to The Editor. *American Journal of Botany*. 102 (11): 1751 – 1752, 2015.

VICTORIA, F.N.; LENARDÃO, E.J.; SAVEGNAGO, L.; PERIN, G.; JACOB, R.G.; ALVES, D.; DA SILVA, W. P.; DA MOTTA, A.S.; NASCENTE, P.S. Essential oil of the leaves of *Eugenia uniflora* L.: Antioxidant and antimicrobial properties. **Food and Chemical Toxicology**. v. 50. 2668-2674. 2012.

VIEIRA, L.D.N.; FAORO, H.; DE FREITAS FRAGA, H.P.; et al. An improved protocol for intact chloroplasts and cpDNA isolation in conifers. **PLoS One**. 9:1–8. doi: 10.1371/journal.pone.0084792. 2014.

Capítulo 4

Sequencia do genoma plastidial completo de *Plinia cauliflora* (Mart.) Kausel, *Plinia aureana* (Mattos) Mattos e *Plinia* sp.: Análise filogenômica baseada no genoma plastidial completo de quatro gêneros da tribo Myrteae (Myrtaceae)

1. Introdução

As jabuticabeiras pertencem à família Myrtaceae e são nativas do Centro-Sul/Sudeste do Brasil. Dentre as principais ameaças as espécies destacam-se a erosão genética causada pela destruição de habitats e das áreas de jabuticabais silvestres. Além disso, outros fatores que também ameaçam as jabuticabeiras são a redução do cultivo e a perda de possíveis variedades cultivadas há centenas de anos em fazendas do interior do país (KINUPP et al. 2011).

A similaridade morfológica entre as espécies da família Myrtaceae tem gerado diversos problemas taxonômicos. Entretanto, estudos filogenéticos moleculares envolvendo sequências do gene plastidial *matK* associadas a dados morfológicos (GADEK et al. 1996, WILSON et al. 2001), sequências *ITS-1*, *ITS-2* e 5.8S rRNA do genoma nuclear (LUCAS et al. 2005, 2007; BIFFIN, et al. 2010, THORNHILL et al. 2012; THORNHILL et al. 2015) e sequências de rRNA 18S-26S e de dois genes plastidiais *matK* e *ndhF* (WILSON et al. 2005; BIFFIN et al. 2010; THORNHILL et al. 2012; THORNHILL et al. 2015), têm contribuído para esclarecer as relações filogenéticas entre os gêneros de Myrteae.

Neste sentido, no presente estudo o genoma plastidial de *Plinia cauliflora* (Mart.) Kausel, *Plinia aureana* (Mattos) Mattos e *Plinia* sp. foi sequenciado e analisado. Além disso, realizou-se uma análise filogenética utilizando sequências completas do genoma plastidial de seis diferentes espécies de Myrtaceae pertencentes à tribo Myrteae além de comparar a ordem gênica entre as espécies dessa tribo.

2. MATERIAL E MÉTODOS

2.1 Coleta de material vegetal

Foram coletadas folhas sadias de uma planta de cada espécie de *Plinia cauliflora* (Mart.) Kausel, *Plinia aureana* (Mattos) Mattos e *Plinia* sp., cujas exsicatas foram depositadas no Herbário da Universidade Tecnológica Federal do Paraná (UTFPR) com os números de tomo e as coordenadas geográficas listados na Tabela 1. *Plinia* sp. é considerada uma espécie híbrida pelos pesquisadores da UTFPR, cuja origem é comercial selecionada por um viveirista de São Paulo. Essa planta é considerada como híbrida pelos pesquisadores da UTFPR devido a precocidade de produção e pelas características da planta ela parece ter as folhas de *Plinia cauliflora* e os frutos parecem ser de *Plinia jaboticaba*.

Após a coleta, as amostras foliares foram armazenadas no escuro, por dois dias a temperatura de 4°C, em geladeira, no Laboratório de Fisiologia e Desenvolvimento Vegetal (LFDGV), procedimento este que tem por função o decréscimo dos níveis de amido e resina das folhas.

Tabela 1. Lista das coordenadas geográficas e número de tomo das exsicatas de *Plinia cauliflora*, *Plinia aureana* e *Plinia* sp.

Espécies	Coordenadas geográficas	Nº de tomo
<i>Plinia aureana</i>	25° 41' 44,8" S 53° 05' 36,0" W	DVPR3001
<i>Plinia cauliflora</i>	25° 41' 46,9" S 53° 05' 37,5" W	DVPR3002
<i>Plinia</i> sp.	25° 42' 13,5" S 53° 05' 52,7" W	DVPR3003

Fonte: Autor

2.2 Isolamento do cloroplasto e do DNA de cloroplasto (cpDNA)

O isolamento dos cloroplastos foi realizado por meio da metodologia descrita por VIEIRA et al. 2014. O protocolo de isolamento de cpDNA foi otimizado, a partir dos cloroplastos isolados, para as espécies de Myrtaceae. Para o isolamento dos cloroplastos, foram coletadas aproximadamente 100 g de folhas, as quais foram mantidas no escuro a 4°C por dois dias antes da realização do procedimento. As folhas foram homogeneizadas em 400 ml de tampão de extração A pH 3,8 (1,25 M NaCl, 0,25 M ácido ascórbico, 10 mM metabissulfito de sódio, 0,0125 M Borax, 50 mM Tris-HCl (pH 8,0), 7 mM EDTA, 1% PVP-40 (m/v), 0,1% BSA (m/v)) em liquidificador por 30 s para a quebra da parede celular. Em seguida, foram realizadas duas filtrações, sendo a primeira em atadura de crepom Cremer com densidade de 18 fios/cm², e a segunda em Miracloth[®]. Essas duas filtrações são importantes para eliminar grande quantidade de resíduo do tampão contendo os cloroplastos. O material filtrado foi centrifugado a 200 g por 15 min a 4°C. Após esta etapa os sedimentos nucleares e de parede celular foram descartados. O sobrenadante, contendo os cloroplastos suspensos, foi centrifugado a 3000 g por 25 min a 4°C. O *pellet* formado nesta etapa contém os cloroplastos e resquícios de DNA nuclear. Posteriormente, o novo sobrenadante foi descartado e o *pellet* resultante ressuspensionado em 10 ml de tampão de extração B pH 8,0 (1,25 M NaCl, 0,0125 M Borax, 1% PVP-40 (m/v), 50 mM Tris-HCl (pH 8,0), 25 mM EDTA, 0,1% BSA (m/v)), com o auxílio de um pincel, com a função de lavar os resquícios de DNA nuclear. Os cloroplastos ressuspensionados foram centrifugados em gradiente de Percoll[®] (70%/30%) a 5.000 g durante 25 min a 4°C. A interface contendo os cloroplastos, entre as soluções 30% e 70% foi coletada e foram realizadas duas lavagens em tampão de extração B pH 8,0 para a retirada do resíduo de Percoll[®]. Para tanto, foram adicionados 100 ml do tampão de extração B, para cada lavagem, e foi realizada a centrifugação a 3000 g por 20 min a 4°C. Entre as lavagens, o *pellet* de cloroplastos resultante foi ressuspensionado com pincel. Por fim, foi obtido o *pellet* de cloroplastos que foi armazenado em freezer a -80°C até o isolamento do cpDNA.

Para isolamento do cpDNA foi utilizado o tampão de isolamento de DNA (100 mM NaCl, 100 mM Tris - HCl (pH 8,0), 50 mM EDTA). A lise dos cloroplastos foi obtida por incubação do *pellet* de cloroplastos em 8 ml de tampão de isolamento de DNA, 1,5 ml de SDS 20 %, 450 µl de 2-mercaptoetanol e 50 µl de Proteinase K (10 mg / ml) em thermo-shaker a 55°C por 4h ou *overnight* para diminuir níveis de polifenóis das amostras. Após, os tubos foram incubados em gelo durante 5 min e, posteriormente, foi adicionado uma solução contendo 1,5 ml de KAc 5M (pH 6,4) a qual foi refrigerada por cerca de 30 min a -20°C. Depois disso, os tubos foram centrifugados a 10.000 g durante 15 min a 4°C, e os sedimentos foram descartados. O sobrenadante foi extraído com um volume igual de fenol saturado e CIA, na proporção (24:1), fenol e CIA são eficientes na desnaturação de proteínas, o material foi centrifugado duas vezes a 10.000 g durante 20 min. Após, foi adicionado ao sobrenadante um volume igual de álcool isopropílico (cerca de 10 ml) e os tubos foram incubados em freezer -20°C *overnight*, para precipitar o DNA plastidial. Por fim, para obtenção do *pellet* de cpDNA, os tubos foram centrifugados a 10.000 g por 20 min a 4°C. O sedimento contendo os cpDNAs foi lavado com etanol 70 % e 96 % e em temperatura ambiente. O *pellet* de cpDNA foi dissolvido em 50µl de água destilada UltraPure™ DNase/RNase-Free Invitrogen™ e posteriormente tratado com RNase (10µg /ml). A qualidade das amostras foi verificada por meio da visualização em gel de agarose a 0,8 %. A pureza e concentração de DNA foram avaliadas por: Nanodrop H, baseado nas taxas 260/280 e 260/230 e Qubit® 2.0 fluorímetro Invitrogen™.

2.3 Sequenciamento e anotação dos genomas plastidiais

2.3.1 Sequenciamento em sequenciador de nova geração

Foi utilizado 1 ng de cada DNA plastidial extraído a partir dos cloroplastos isolados para o sequenciamento. A construção das bibliotecas de DNA plastidial foi realizada utilizando o kit Nextera® XT DNA Sample Prep v3 (Illumina, San Diego, USA) de acordo com as recomendações do fabricante. A qualidade dos fragmentos gerados pela biblioteca foram checadas em Bioanalyzer 2100 (Agilent Technologies)

e, após apresentarem qualidade satisfatória, foram sequenciados em Illumina MiSeq (San Diego, USA). O sequenciamento foi realizado no Departamento de Bioquímica e Biologia Molecular da Universidade Federal do Paraná. Após o sequenciamento foi adotada a estratégia de montagem *de novo* utilizando o programa CLC workbench versão 8.0.1.

2.3.2 Desenho de iniciadores e padronização da amplificação via reação em cadeia de polimerase (PCR)

Foram desenhados dois pares de iniciadores (Tabela 2) para verificar duas regiões do cpDNA de *P.aureana*, com cobertura baixa, por meio do método Sanger de sequenciamento no sequenciador automático ABI 3500XL (Applied Biosystems™). No desenho dos iniciadores foi utilizado o programa *on line* Primer3Plus (UNTERGASSER et al. 2007). Visando a amplificação dos fragmentos foi utilizado 5 ng de amostra de DNA, 0,2mg/ml de albumina de soro bovino (BSA), 0,2 mM de cada dNTP, 1U de Taq DNA Polymerase (Quatro G), 1x Taq buffer, 2 mM MgCl₂ e 0,3 μM de cada primer em um volume total de 20 μL. O termociclador Veriti (Life Technologies, Carlsbad, CA) foi utilizado para amplificação dos fragmentos seguindo o protocolo: 95 °C por 2 min, com 35 ciclos de 95 °C por 30 s, temperatura de anelamento por 1 min, 72 °C por 1 min, e uma extensão final de 7 min a 72 °C. Para determinar a temperatura de anelamento de cada primer foi testada inicialmente gradiente de 52 °C a 60 °C. Os produtos da amplificação foram quantificados em gel de agarose 1,5 % corado com GelRed (Biotium®) por comparação com o padrão de peso molecular 1Kb DNA ladder (Invitrogen™).

Tabela 2. Lista de iniciadores para sequenciamento de fragmentos específicos do genoma plastidial de *P. aureana*.

Iniciadores	Sequência	Tm °C	Comprimento (pb)
Aureana_01	Forward TGTCTAATGGATAGGACAGAGGT	60,3	20
	C Reverse CGGGAACGAATGGTTATCTT		
Aureana_02	Forward AATTTAGTAAACTCTTGGGATCTT	60,3	20
	TTC Reverse TGGTTTGGGAATGCATTTATCTT		

Fonte: Autor

2.3.3 Purificação com Polietileno Glicol (PEG)

Após a reação de amplificação, os produtos foram purificados, com a função de remover os iniciadores direto, reverso e dNTPs não incorporados durante a reação de amplificação. Esta purificação é essencial, pois a reação subsequente (sequenciamento de DNA – terminação de cadeia) acontece na presença de apenas um iniciador. A cada reação, foram adicionados 20,0 µL de PEG 800 20% contendo cloreto de sódio 2,5 M. As reações foram levemente agitadas (em Vortex®) e incubadas por 30 min em incubadora a 37 °C. Após, as reações foram centrifugadas em temperatura ambiente durante 15 min a 15.115 g. O sobrenadante foi removido e foi adicionado 125,0 µL de etanol 80% (-20°C) para a lavagem do pellet. As reações foram centrifugadas em temperatura ambiente durante 8 min a 15.115 g. O sobrenadante foi novamente removido e em seguida foram adicionados 125,0 µL de etanol 80% (-20°C) para a lavagem do pellet. As reações foram novamente centrifugadas em temperatura ambiente durante 8 min

a 15.115 g. O sobrenadante foi removido e o pellet formado secou em incubadora a 37 °C. Após esta etapa, o pellet foi ressuspensionado (durante 30 min a 37 °C) em 10,0 µL de água UltraPure™ DNase/RNase-Free Invitrogen™. O produto de PCR purificado foi, posteriormente, quantificado em gel de agarose 0,8%, corado com GelRed (Biotium®), por comparação com o padrão de peso molecular 1 kb DNA ladder (Invitrogen™).

2.3.4 Reação de sequenciamento pelo Método Sanger e purificação pós-reação de sequenciamento

Após a otimização, a amplificação e a purificação, os produtos de amplificação dos dois pares de iniciadores plastidiais foram sequenciados, pelo método de terminação de cadeia (SANGER et al. 1977) utilizando o Kit BigDye® Terminator v3.1 (Applied Biosystems™) em um sequenciador de DNA automático ABI 3500XL (Applied Biosystems™). As reações de sequenciamento de DNA foram realizadas em um volume total de 10,0 µL, contendo: 1,0 µL (10,0 ng) de produto de amplificação purificado da primeira reação de amplificação; 0,3 µL de iniciador (estocado a 10,0 µM); 4,0 µL do mix BigDye® Terminator v3.1 (Applied Biosystems™), e 5,35 µL de água UltraPure™ DNase/RNase-Free Invitrogen™. Termociclador Veriti (Life Technologies, Carlsbad, CA) foi utilizado para amplificação da reação de sequenciamento, seguindo o protocolo: 96 °C por 2 min, com 35 ciclos de 96 °C por 10 s, temperatura de anelamento por 15 min, 60 °C por 4 min.

Após a amplificação dos produtos via reação de sequenciamento foi realizada uma nova purificação para remover iniciadores, dNTPs e ddNTPs não incorporados nos amplicons. A cada reação de sequenciamento, foi adicionado 2,0 µL de acetato de amônio 7,5 M, 2,0 µL de EDTA 125 mM e 50,0 µL de etanol absoluto (100%) (-20 °C). O procedimento foi realizado em placa de PCR (Applied Biosystems™). Cada placa foi agitada levemente para homogeneizar a solução. As placas foram mantidas por 15 min, no escuro em geladeira (4,0 °C) para permitir a precipitação dos ácidos nucleicos. Após isso, foram centrifugadas a 4,0 °C durante 30 min a 3.000 g. O excesso de isopropanol foi descartado e as placas foram invertidas em papel toalha. Após, foi adicionado 70 µL de etanol 70% (-20 °C) para a lavagem do

pellet. As reações foram novamente centrifugadas a 4,0 °C durante 15 min a 1.650 g. O sobrenadante foi novamente removido por inversão e logo após breve centrifugação durante 1 min a 185 g, com a placa invertida em papel toalha. O pellet foi seco a temperatura ambiente e no escuro. Após esta etapa, o pellet foi ressuscitado por 60 min em 10,0 µL de formamida HiDi™ (Applied Biosystems™). Após, as placas foram desnaturadas por 5 min a 95°C em termociclador e logo após incubada em gelo por 3 min. Após esta etapa as placas foram colocadas em sequenciador automático ABI 3500XL (Applied Biosystems™).

2.4.5 Anotação dos genomas plastidiais de *Plinia cauliflora*, *Plinia aureana* e *Plinia* sp.

Os genomas plastidiais das três espécies alvo deste estudo foram anotados inicialmente utilizando o programa Dual Organellar GenoMeAnnotator DOGMA (WYMAN et al. 2004). Os genes previamente anotados foram confirmados utilizando o BLAST (ALTSCHUL et al. 1990). Os genes de tRNAs foram preditos utilizando o programa tRNAscan-SE (LOWE & EDDY, 1997). O mapa circular do cpDNA foi desenhado com o auxílio do programa OrganellarGenomeDRAW (LOHSE et al. 2013). A ordem dos genes entre os quatro gêneros de interesse neste estudo foi verificada por meio do alinhamento dos genomas plastidiais pelo programa Mauve versão 2.4.0 (DARLING et al. 2004), este programa foi utilizado para identificar combinações no genoma em blocos localmente colineares (LCBs) que permite avaliar rearranjos, grandes inserções ou deleções e divergência entre sequências.

A região SSC do genoma plastidial de *Plinia aureana*, está com a orientação invertida, isto é considerado usual, já que as duas orientações da região SSC ocorrem constantemente durante a replicação do cpDNA (WALKER et al. 2015). Porém, para a realização das análises a região SSC foi reorientada.

2.4 Análise filogenômica da tribo Myrteae

A análise filogenômica foi realizada com os genomas plastidiais de seis diferentes espécies de Myrtaceae pertencentes à tribo Myrteae incluindo *P. aureana*, *P. cauliflora* e *Plinia* sp., sequenciadas neste estudo. Além disso, incluiu-se uma espécie de *Eugenia uniflora* cuja sequência está disponível no GenBank pelo número de acesso KR867678. *Lagerstroemia fauriei* (Myrtales: Lythraceae; KT358807) foi utilizada como *outgroup* na análise. A IRa foi omitida da análise para evitar a representação excessiva das sequências da região IR. Os genomas plastidiais foram alinhados utilizando o programa *Multiple Alignment using Fast Fourier Transform* (MAFFT) (KATO; STANLEY 2013). As posições dos nucleotídeos que continham um ou mais intervalos introduzidos pelos alinhamentos foram omitidas da matriz de distância. Após, realizou-se o teste para selecionar o modelo de evolução de sequências para todo o conjunto de genomas plastidiais completos de acordo com o critério de informação Akaike [*General Time Reversible + Invariant sites + Gamma distribution* (GTR + I + G)] utilizando o programa jModelTest (DARRIBA et al. 2012) e 7 foi utilizado como o número de substituições de Shemes. A análise Bayesiana foi estimada com base no modelo evolutivo que mais se ajustou aos dados (GTR + I + G), implementado no programa MrBayes 3.2.2 (RONQUIST; HUELSENBECK 2003). A análise de *Markov chain Monte Carlo* (MCMC) foi realizada até 2.000.000 de gerações, e todos os pontos amostrais anteriores a estabilidade, ou seja, (25% das primeiras árvores) foram descartados como amostras *burn in*. Os dados remanescentes posteriores ao descarte de amostras *burn in* foram utilizados para gerar a árvore consenso, sendo que, a porcentagem de amostras recobrindo qualquer clado em particular representa a probabilidade posterior (PP) do clado (RONQUIST; HUELSENBECK 2003). A árvore resultante foi representada e editada utilizando o programa FigTree v1.4.1.

3. RESULTADOS E DISCUSSÃO

3.1 Genoma plastidial e conteúdo e arranjo dos genes

Os *reads* obtidos pelo sequenciamento no Illumina MiSeq e submetidos a estratégia de montagem *de novo* resultaram em alta cobertura dos genomas plastidiais de *P. aureana* (~130x), *P. cauliflora* (~450x) e *Plinia* sp. (~250x) (Tabela 3). As duas regiões do genoma plastidial de *P. aureana* que apresentava baixa cobertura foram resolvidas pelo método de sequenciamento Sanger como na mostra a Figura 1. As sequências apresentam 791 pb e 1076 pb, respectivamente. A primeira sequência corresponde ao gene *rrn16* pertencente a primeira IR, enquanto a segunda sequência corresponde aos genes *trnI-GAU*, *rrn16* (pertencente a segunda IR) e o pseudogene *ycf68*.

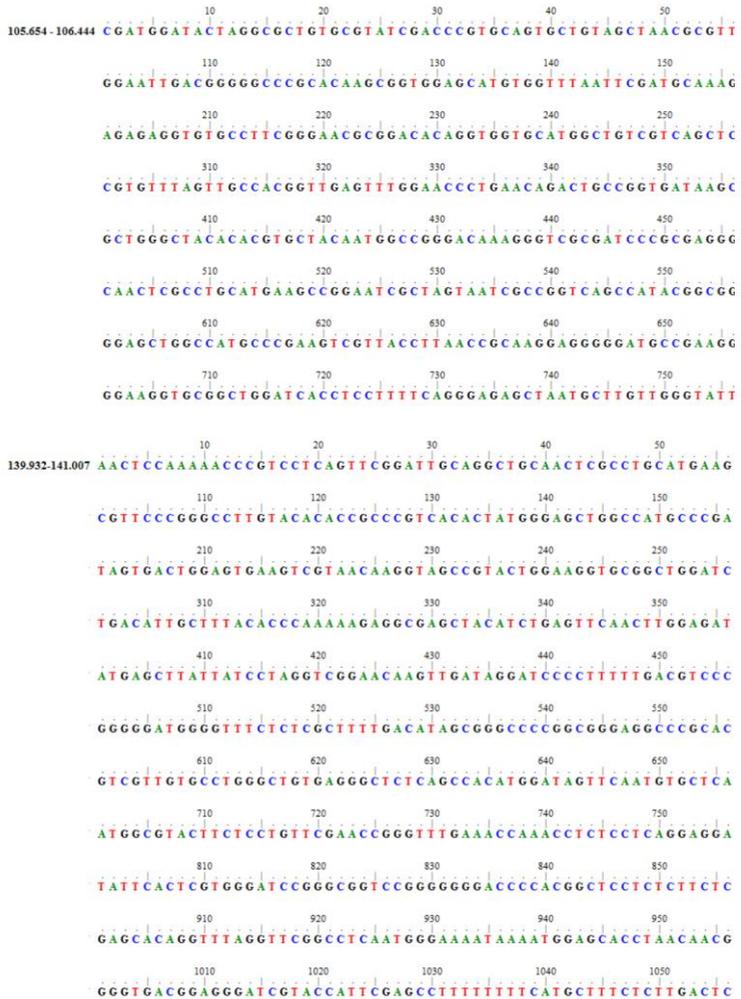
Tabela 3. Resultado do sequenciamento e montagem do cpDNA de três espécies de *Plinia*.

Espécies	<i>P. aureana</i>	<i>P. cauliflora</i>	<i>Plinia sp.</i>
Tamanho do cpDNA (pb)	158.918	159.095	158.912
Conteúdo GC %	36,96	36,97	36,97
Total de <i>reads</i> contados	74.73	414.98	215.01
Média do comprimento dos <i>reads</i> (pb)	280.90	172.97	186.74
<i>Reads</i> mapeados alinhados em pares	9.95	411.43	202.46
Cobertura do cpDNA*	129.73	450.02	251.36

*Cobertura: n° de bases sequenciadas/tamanho estimado do cpDNA.

Fonte: Autor

Figura 1. Sequências de duas regiões com 791 pb e 1076 pb, respectivamente, resolvidas pelo método de sequenciamento Sanger no genoma plastidial de *P. aureana*



Fonte: Autor

Os genomas plastidiais de *P. aureana*, *P. cauliflora* e *Plinia* sp. apresentaram os seguintes tamanhos 158.918 pb, 159.095 pb e 158.912 pb, respectivamente (Tabela 4), incluindo: uma região longa de cópia única (LSC) de 88.204 pb, 88.162 pb e 88.132 pb, respectivamente (que corresponde a aproximadamente 55,5% de cada genoma plastidial); uma região curta de cópia única (SSC) de 18.462 pb, 18.615 pb e 18.462 pb, respectivamente (aproximadamente 11,7% de cada genoma plastidial); e um par de sequências invertidas repetidas (IRa e IRb) com 26.126 pb, 26.159 e 26.159 cada, respectivamente (Figura 2; Tabela 4). O conteúdo GC total do genoma foi de 36,96 % para *P. aureana* e de 36,97 % para *P. cauliflora* e *Plinia* sp. As regiões LSC, SSC e IR apresentaram 34,8 %; 30,8 % e 42,7 % do conteúdo GC das respectivas regiões do genoma plastidial para as três espécies sequenciadas neste estudo (Tabela 4).

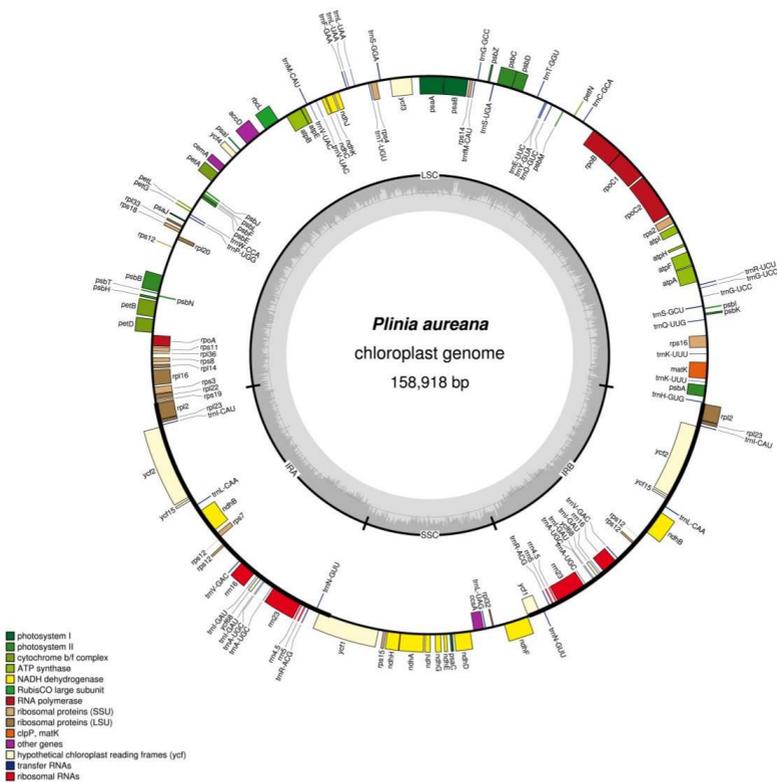
Nos três genomas plastidiais um total de 115 diferentes genes foram identificados, sendo que 78 codificam para proteínas, 30 genes de tRNA e quatro genes de rRNA (Figura 2; Tabela 5). Um total de 20 genes encontra-se duplicado nas IR, dentre esses, sete genes de tRNA (*trnN-GUU*; *trnR-ACG*; *trnA-UGC*; *trnI-GAU*; *trnV-GAC*; *trnL-CAA*; *trnI-CAU*), seis genes codificadores de proteínas (*rps12*; *rps7*; *ndhB*; *ycf2*; *rpl23*; *rpl2*), três *pseudogenes* (*ycf68*; *ycf1*; *ycf15*) e quatro genes de rRNA (*rrn16*; *rrn23*; *rrn4.5*; *rrn5*). Os genomas plastidiais de *P. aureana*, *P. cauliflora* e *Plinia* sp. são similares aos já sequenciados de outras Myrtaceae (STEANE, 2005; PAIVA et al. 2011; ASIF et al. 2013; BAYLY et al. 2013; MACHADO et al. 2017).

Tabela 4. Resumo das características dos genomas plastidiais de *Plinia* spp.

Parâmetros	<i>Plinia aureana</i>	<i>Plinia cauliflora</i>	<i>Plinia</i> sp.
Acesso GenBank	KY392759	KX527622	KX668206
Tamanho do cpDNA (Kb)	158918	159095	158912
LSC tamanho em bp (%)	88.204 (55,5%)	88.162 (55,4%)	88.132 (55,5%)
SSC tamanho em bp (%)	18.462 (11,6%)	18.615 (11,7%)	18.462 (11,6%)
IR tamanho em bp	26.126	26.159	26.159
Genes diferentes	115	115	115
Diferentes GCP	78	78	78
Diferentes genes de tRNA	30	30	30
Diferentes genes de rRNA	4	4	4
Diferentes genes duplicados em IR	20	20	20
Diferentes genes com introns	18	18	18
Conteúdo total % GC	36,96	36,97	36,97
% GC LSC	34,8	34,8	34,8
% GC SSC	30,8	30,8	30,8
% GC IR	42,7	42,7	42,7

Fonte: Autor

Figura 2. Organização dos genes do genoma do cloroplasto de *Plinia aureana*. Genes fora do círculo são transcritos no sentido horário e genes no interior o círculo são transcritos no sentido anti-horário. O círculo central representa as diferentes regiões (LSC, SSC e duas IR) do genoma plastidial. A organização dos genes de *Plinia* spp. é igual ao de *P. aureana* em ordem e conteúdo de genes.



Fonte: Autor

Tabela 5. Lista de genes identificados no genoma plastidial de *Plinia* spp.

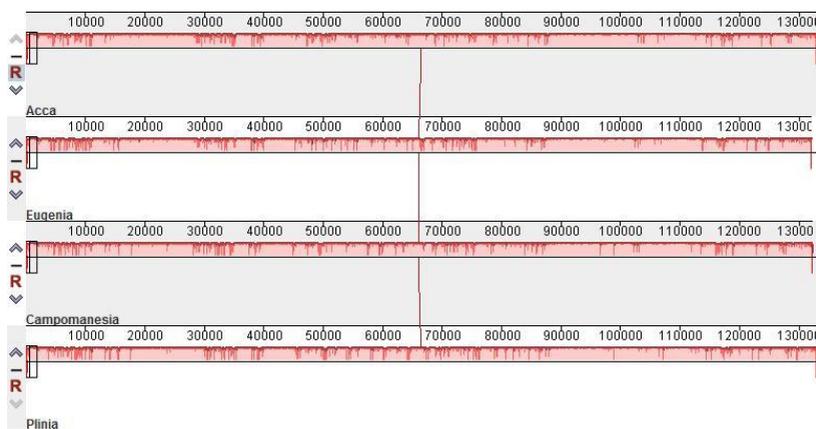
Categoria funcional	Grupo de genes	Nome dos genes									
Auto - replicação	Subunidade de ribossomal grande	<i>rpl14</i>	<i>rpl16</i>	<i>rpl2</i> ^{ab}	<i>rpl20</i>	<i>rpl22</i>	<i>rpl23</i> ^b	<i>rpl32</i>	<i>rpl33</i>	<i>rpl36</i>	
	DNA polimerase dependente de RNA	<i>rpoA</i>	<i>rpoB</i>	<i>rpoC1</i> ^a	<i>rpoC2</i>						
	Subunidade de ribossomal pequena	<i>rps11</i>	<i>rps12</i> ^{bc*}	<i>rps14</i>	<i>rps15</i>	<i>rps16</i> ^a	<i>rps18</i>	<i>rps19</i>	<i>rps2</i>	<i>rps3</i>	
		<i>rps4</i>	<i>rps7</i> ^a	<i>rps8</i>							
	Genes do RNA ribossomal	<i>rrn1</i> ^b	<i>rrn23</i> ^c	<i>rrn4.5</i> ^b	<i>rrn5</i> ^a	<i>trnE-</i>	<i>trnF-</i>	<i>trnG-M-</i>	<i>trnG-</i>	<i>trnG-</i>	<i>trnH-</i>
		<i>trnA-UGC</i> ^{ab}	<i>trnC-GCA</i>	<i>trnD-GUC</i>	<i>UUC</i>	<i>GAA</i>	<i>CAU</i>	<i>GCC</i>	<i>UCC</i> ^a	<i>GUG</i>	
	Genes para RNA transportadores	<i>trnI-CAU</i> ^b	<i>trnI-GAU</i> ^{ab}	<i>trnK-UUU</i> ^a	<i>CAA</i> ^b	<i>UAA</i> ^a	<i>UAG</i>	<i>CAU</i>	<i>GUU</i> ^a	<i>UGG</i>	
		<i>trnQ-UUG</i>	<i>trnR-ACG</i> ^a	<i>trnR-UUCU</i>	<i>GCU</i>	<i>GGA</i>	<i>UGA</i>	<i>GGU</i>	<i>UGU</i>	<i>GAC</i> ^b	
		<i>trnI-UAC</i> ^a	<i>trnW-CCA</i>	<i>trnY-GUA</i>							
		Fator de iniciação de tradução	<i>infA</i> ^c								
Genes relacionados a fotossíntese	Subunidade do fotossistema I	<i>psaA</i>	<i>psaB</i>	<i>psaC</i>	<i>psaI</i>	<i>psaJ</i>	<i>ycf3</i> ^a	<i>ycf4</i>			
	Subunidade do fotossistema II	<i>psbA</i>	<i>psbB</i>	<i>psbC</i>	<i>psbD</i>	<i>psbE</i>	<i>psbF</i>	<i>psbH</i>	<i>psbI</i>	<i>psbJ</i>	
		<i>psbK</i>	<i>psbL</i>	<i>psbM</i>	<i>psbN</i>	<i>psbT</i>	<i>psbZ</i>				
	Subunidade do citocromo b 6/f	<i>petA</i>	<i>petB</i> ^a	<i>petD</i> ^a	<i>petG</i>	<i>petL</i>	<i>petN</i>				
	Subunits of ATP synthase	<i>atpA</i>	<i>atpB</i>	<i>atpE</i>	<i>atpF</i> ^a	<i>atpH</i>	<i>atpI</i>				
	Grande subunidade do Rubisco	<i>rbcL</i>									
		Subunidades da NADH desidrogenase	<i>ndhA</i> ^a	<i>ndhB</i> ^{ab}	<i>ndhC</i>	<i>ndhD</i>	<i>ndhE</i>	<i>ndhF</i>	<i>ndhG</i>	<i>ndhH</i>	<i>ndhI</i>
		<i>ndhJ</i>	<i>ndhK</i>								
	Outros genes	Subunidade de acetil-CoA-carboxilase	<i>accD</i>								
		Gene que sintetiza o citocromo c	<i>ccsA</i>								
Envelope de proteína de membrana		<i>cemA</i>									
		Protease	<i>clpP</i> ^a								
Maturase		<i>matK</i>									
Componente do complex TIC		<i>ycf1</i> ^{bc*}									
Genes com funções desconhecidas	Função desconhecida	<i>ycf2</i> ^b	<i>ycf13</i> ^{bc}	<i>ycf68</i> ^{bc}							

^a Gene contendo íntron; ^b Genes duplicados; ^c Pseudogene; ^{c*} *ycf1* é pseudogene na fronteira entre as regiões IRb e SSC; ^{*}*rps12* é trans-splicing com a extremidade 5' localizada na região LSC e a extremidade 3' duplicada na região IR.

Fonte: Autor

A análise do conteúdo gênico dos genomas plastidiais das três espécies sequenciadas neste estudo foi identificado e está explicitado na Figura 2 e Tabela 5. Além disso, a similaridade entre o conteúdo gênico de espécies de quatro gêneros (*Acca*, *Campomanesia*, *Eugenia* e *Plinia*) pertencentes à tribo Myrteae é mostrado na Figura 3. A similaridade entre o conteúdo gênico dos genomas plastidiais de Myrtaceae já havia sido reportada por Machado et al. (2017).

Figura 3. Alinhamento múltiplo do genoma plastidial da tribo Myrteae utilizando o programa MAUVE. A presença de apenas um bloco localmente colinear, representado na cor rosa, indica a região de DNA homóloga, sem rearranjos de sequência.



Fonte: Autor

Conforme Howe et al. (2003) o conteúdo da porcentagem de AT varia de 61,3% nas gramíneas a 62,9% nas dicotiledôneas, sendo que a frequência de AT é menor nas IR que nas regiões de cópia única (SSC e LSC). Além disso, este elevado conteúdo de AT tem sido considerado um dos fatores que facilitam a transferência de genes do

genoma plastidial para o genoma nuclear. Assim, o conteúdo de bases AT das espécies *P. aureana*, *P. cauliflora* e *Plinia* sp. foi de aproximadamente 63%, valor similar aos valores do conteúdo de bases AT dos genomas plastidiais de outras Myrtaceae, como *S. Cumini* (63,2%) (ASIF et al. 2013) e *Eucalyptus corymbia* (63,0%) (BAYLY et al. 2013) e *A. sellowiana* (63,0%) (MACHADO et al. 2017).

3.2 Análise filogenômica da tribo Myrteae

A matriz de distância dos sete genomas plastidiais de Myrteae (Myrtaceae) completamente alinhados foi de 148.939 posições de nucleotídeos em comprimento; no entanto, após a exclusão dos *gaps* o referido valor diminuiu para 120.847 sítios. A análise Bayesiana BI (-lnl 21.792.256.407) produziu a árvore filogenômica representada pela Figura 4. As probabilidades posteriores da análise de Inferência Bayesiana mostraram suporte máximo para todos os nodos, com o valor de 1.0.

A sequência do genoma do cloroplasto completo é uma ferramenta bioinformática útil para o entendimento das relações evolutivas e, quando comparadas com o genoma nuclear, as sequências do genoma plastidial são mais conservadas evolutivamente (PALMER, 1985). Neste sentido, no presente estudo foi realizada uma análise filogenômica baseada na sequência do genoma plastidial completo envolvendo seis diferentes espécies pertencentes à tribo Myrteae. A tribo Myrteae era dividida em três subtribos, com base na morfologia dos embriões: Eugeniinae, Myrciinae e Myrtinae (MCVAUGH, 1956, LANDRUM; KAWASAKI, 1997, WILSON et al. 2001; 2005). Entretanto, Lucas et al. (2007) propôs uma nova classificação para tribo Myrteae. Atualmente, Myrteae está classificada em grupos ao redor dos gêneros *Plinia* L., *Myrcia* DC., *Myrceugenia* O. Berg, *Myrteola* O. Berg, *Pimenta* Lindl. e *Eugenia* L.

A árvore filogenética suporta claramente a monofilia da subtribo Myrtinae; no entanto, Eugeniinae não foi bem resolvida nesta filogenia, pois as análises indicaram ser parafilética. Esse resultado corrobora com o estudo realizado por Lucas et al. (2005), no qual Eugeniinae é considerada parafilética por meio de análise filogenética baseada nas sequências dos genes *ITS-1*, *ITS-2* e 5.8S rRNA do genoma nuclear e *psbA-trnH* espaçador intergênico de cloroplasto.

O clado do grupo *Pimenta* (anteriormente Myrtinae) formado por *A. sellowiana* + *C. xanthocarpa* é fortemente suportado como monofilético (PP = 1,0; nodo 5; Figura 4). Esse resultado está de acordo com estudo realizado por Lucas et al. (2007), onde os gêneros *Campomanesia*, *Psidium*, *Pimenta* e *Acca* aparecem como monofiléticos, quando por Inferência Bayesiana, no qual foram utilizadas as sequências do genes *ITS*, *ETS* do genoma nuclear e *psbA-trnA* e *matK* do genoma plastidial. O clado do grupo *Pimenta* sugerido

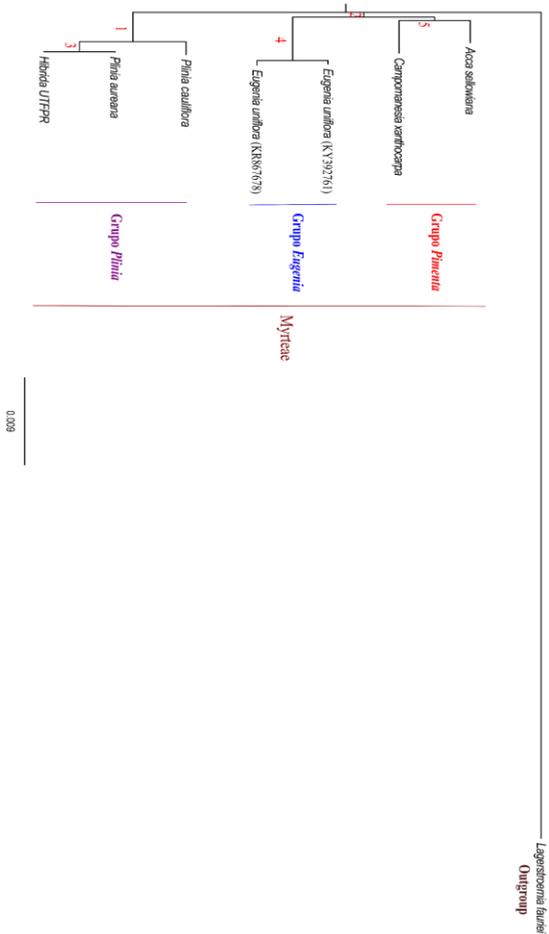
como monofilético igualmente concorda com o estudo realizado anteriormente por Murilo-A et al. (2013), que utilizaram as sequências *ITS* e *ETS* do genoma nuclear.

O grupo *Eugenia* é representado na análise do presente estudo pelos gêneros *Eugenia* formando pelo clado *E. uniflora* (KY392761) + *E. uniflora* (KR867678) (PP =1,0; nodo 4; Figura 4). Enquanto que, o grupo *Plinia* é formado pelo clado *P. cauliflora* + *P. aureana* + *Plinia* sp. (PP = 1,0; nodo 1; Figura 4). Os grupos *Eugenia* e *Plinia* eram gêneros classificados em Eugeniinae e eram considerados parafiléticos (LUCAS et al. 2005). Neste estudo os grupos *Eugenia* e *Plinia* formam dois cladogramas monofiléticos corroborando com o estudo realizado por LUCAS et al. (2007). O clado formado pelo grupo *Plinia* (Eugeniinae) mostrou maior similaridade entre *P. aureana* e *Plinia* sp. (PP =1,0; nodo 3; Figura 4).

4. CONCLUSÃO

Neste estudo, foi realizado o sequenciamento do genoma plastidial de três espécies do gênero *Plinia*: *P. aureana*, *P. cauliflora* e *Plinia* sp. O conteúdo e arranjos dos genes são similares ao que foi encontrado para outras espécies de Myrtaceae evidenciando a conservação do genoma plastidial dessa família. Além disso, realizou-se uma análise filogenômica baseada no genoma plastidial completo de seis diferentes espécies de Myrtaceae, todas pertencentes à tribo Myrteae. A tribo Myrteae é classificada em seis diferentes grupos: *Plinia* L., *Myrcia* DC., *Myrceugenia* O. Berg, *Myrteola* O. Berg, *Pimenta* Lindl. e *Eugenia* L., este estudo apresentou representantes do grupo *Pimenta* (*Acca sellowiana* e *Campomanesia xanthocarpa*), grupo *Eugenia* (*Eugenia uniflora*) e grupo *Plinia* (*Plinia cauliflora*, *Plinia aureana* e *Plinia* sp.). A análise de inferência Bayesiana indicou monofilia, para os três grupos de Myrteae com suporte máximo de probabilidade posterior.

Figura 4. Inferência Bayesiana baseada na sequência completa do genoma plastidial de seis diferentes espécies pertencentes à tribo Myrteae (Myrtaceae) e *Lagerstroemia fauriei* (Myrtales: Lythraceae; KT358807) como outgroup. Todos os nodos apresentaram probabilidade posterior (PP) igual a 1.0. O comprimento dos ramos é proporcional ao nível de divergência inferida. A barra de escala indica o número substituições de ácido nucléico inferidas por sítio.



REFERÊNCIAS BIBLIOGRÁFICAS

- ASIF, H.; KHAN, A.; IQBAL, A.; KHAN, I.A.; HEINZE, B.; AZIM, M.K.; The chloroplast genome sequence of *Syzygium cumini* (L.) and its relationship with other angiosperms. **Tree Genetics & Genomes**. 9:867–877. 2013.
- BAYLY, M.J.; RIGAUT, P.; SPOKEVICIUS, A.; LADIGES, P.Y.; ADES, P.K.; ANDERSON, C.; BOSSINGER, G.; MERCHANT, A.; UDOVICIC, F.; WOODROW, I.E.; TIBBITS, J. Chloroplast genome analysis of Australian eucalypts – *Eucalyptus*, *Corymbia*, *Angophora*, *Allosyncarpia* and *Stockwellia* (Myrtaceae). **Molecular Phylogenetics and Evolution**. 69,704–716. 2013.
- BIFFIN, E.; LUCAS, E.J.; CRAVEN, L.A. et al. Evolution of exceptional species richness among lineages of fleshy-fruited Myrtaceae. **Ann Bot**. 106:79–93. doi: 10.1093/aob/mcq088. 2010.
- DARLING, A.C.E.; MAU, B.; BLATTNER, F.R.; PERNA, N.T. Mauve : Multiple Alignment of Conserved Genomic Sequence With Rearrangements. **Genome Research**. 1394–1403. 2004. doi: 10.1101/gr.2289704.
- DARRIBA, D.; TABOADA, G.L.; DOALLO, R.; POSADA, D. jModelTest 2: more models, new heuristics and parallel computing. **Nat Methods**. 9:772–772. doi: 10.1038/nmeth.2109. 2012.
- GADEK, P. A. WILSON, P.G. QUINN, C.J. Phylogenetic reconstruction in Myrtaceae using matK, with particular reference to the position of *Psiloxylon* and *Heteropyxis*. **Australian Systematic Botany**. 9:283–290. 1996.
- KATOH, K.; STANDLEY, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. **Mol Biol Evol** 30:772–780. doi: 10.1093/molbev/mst010. 2013.

KINUPP, V. F. Espécies Alimentícias Nativas da Região Sul do Brasil. In: CORADIN, L.; SIMINSKI, A.; REIS, A. *Espécies nativas da Flora Brasileira de Valor Econômico Atual ou Potencial: Plantas para o Futuro-Região sul*. Brasília: MMA, 2011. 107-110p.

LOHSE, M.; DRECHSEL, O.; KAHLAU, S.; BOCK, R. OrganellarGenomeDRAW: a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. **Nucleic Acids Research**, v. 41, n. W575-81, 2013.

LANDRUM, L. R. A monograph of the genus *Myrceugenia* (Myrtaceae). *Flora Neotropica*, Monograf. 29: 1-137. 1981.

LANDRUM, L.R.; KAWASAKI, M.L. The genera of Myrtaceae in Brazil: an illustrated synoptic treatment and identification keys. **Brittonia**, v.49, p.508-536, 1997.

LUCAS, E.J.; BELSHAM, S.A.; NICLUGHADA, E.M.; ORLOVICH, D.; SAKURAGUI, C.; CHASE, M.; WILSON, P.G. Phylogenetics patterns in the fleshy-fruited Myrtaceae - preliminary molecular evidence. **Plant Systematics and Evolution**, v.251, p. 35-51, 2005.

LUCAS, E.J.; HARRIS, S.A.; MAZINE, F.F.; BELSHAM, S.R.; NICLUGHADA, E.M.; TELFORD, A., GASSON, P.E.; MARK W.; CHASE, M.W. Suprageneric phylogenetics of Myrteae, the generically richest tribe in Myrtaceae (Myrtales). **Taxon**,v.56, n.4, p.1105–1128, 2007.

MACHADO, L.O.; VIEIRA, L.N.; STEFENON, V.M.; PREDROSA, F.O.; SOUZA, E.M.; GUERRA, M.P.; NODARI, R.O. Phylogenomic relationship of feijoa (*Acca sellowiana* (O.Berg) Burret) with other Myrtaceae based on complete chloroplast genome sequences. *Genetica*. doi: 10.1007/s10709-017-9954-1. 2017.

MCVAUGH, R. Tropical American Myrtaceae. Notes on generic concepts and descriptions of previously unrecognized species. **Fieldiana, Botany** 29(3): 145228. 1956.

MATTOS, J.R. Novidades taxonômicas em Myrtaceae – XV. **Loefgrenia**: comunicações avulsas de Botânica, Florianópolis, n.112. 1998. 9p.

MURILLO-A, J.; STUESSY, T.; RUIZ, E. Phylogenetic relationships among *Myrceugenia*, *Blepharocalyx* and *Luma* (Myrtaceae) based on paired-sites models and the secondary structures of ITS and ETS sequences. **Plant Systematics and Evolution**. 299: 713-729. 2013.

PAIVA, J.A.; PRAT, E.; VAUTRIN, S.; SANTOS, M.D.; SANCLEMENTE, H.; BROMMONSCHENKEL, S.; FONSECA, PG.; GRATTAPAGLIA, D.; SONG, X.; AMMIRAJU, J.S.; KUDRNA, D.; WING, R.A.; FREITAS, A.T.; BERGÈS, H.; GRIMA-PETTENATI, J. Advancing Eucalyptus genomics: identification and sequencing of lignin biosynthesis genes from deep-coverage. BAC libraries. **BMC Genom.** 4(12):137. 2011.

PALMER, J.D. Chloroplast DNA and molecular phylogeny. *BioEssays* 2:263–267. doi: 10.1002/bies.950020607. 1985.

RONQUIST, F.; HUELSENBECK, J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574. doi: 10.1093/bioinformatics/btg180. 2003.

SANGER, F.; NICKLEN, S.; COULSON R. 1977. DNA sequencing with chain-terminating inhibitors. **Proceedings of the National Academy of Sciences of the United States of America** 74: 5463-5467.

STEANE, D.A. Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum *Eucalyptus globulus* (Myrtaceae). **DNA Res.**12(3):215–220. 2005.

THORNHILL, A.H.; HO, S.Y.W.; KÜLHEIM, C.; CRISP, M.D. Interpreting the modern distribution of Myrtaceae using a dated molecular phylogeny. **Mol Phylogenet Evol.** 93:29–43. doi: 10.1016/j.ympev.2015.07.007. 2015.

THORNHILL, A.H.; POPPLE, L.W.; CARTER, R.J.; et al. Are pollen fossils useful for calibrating relaxed molecular clock dating of phylogenies?

A comparative study using Myrtaceae. **Mol Phylogenet Evol.**63:15–27. doi: 10.1016/j.ympev.2011.12.003. 2012.

UNTERGASSER, A.; NIJVEEN, H.; RAO, X.; BISSELING, T.; GEURTS, R.; LEUNISSEN, J.A.M. Primer3Plus, an enhanced web interface to Primer3. **Nucleic Acids Research.** 2007. 35: W71-W74; doi:10.1093/nar/gkm306.

WALKER, J.F.; JANSEN, R.K.; ZANIS, M.J.; EMERY, N.C. Sources of inversion variation in the small single copy (SSC) region of chloroplast genomes. Letter to The Editor. *American Journal of Botany.* 102 (11): 1751 – 1752, 2015.

WILSON P.G., O'BRIEN M.M., GADEK P.A., QUINN C.J. Myrtaceae revisited: a reassessment of infrafamilial groups. *Amer. J. Bot.* 88: 2013–2025, 2001.

WILSON, P.G.; O'BRIEN, M.M; HELSEWOOD, M.M; QUINN, C.J. Relationships within Myrtaceae sensu lato based on *matK* phylogeny. **Plant Systematics and Evolution**, v.251, p.3-19, 2005.

VIEIRA, L.D.N.; FAORO, H.; DE FREITAS FRAGA, H.P.; et al. An improved protocol for intact chloroplasts and cpDNA isolation in conifers. **PLoS One.** 9:1–8. doi: 10.1371/journal.pone.0084792. 2014.

3. CONSIDERAÇÕES FINAIS E PERSPECTIVAS FUTURAS

A realização desta tese de doutorado apresentou resultados relevantes no conhecimento do genoma plastidial de espécies da família Myrtaceae. Os cpDNAs de seis espécies pertencentes a tribo Myrteae (Myrtaceae) foram sequenciados e tiveram seus genomas plastidiais montados a partir estratégia *de novo*. A partir dos dados obtidos do sequenciamento dos genomas plastidiais de *A. sellowiana*, *C. xanthocarpa*, *E. uniflora*, *P. aureana*, *P. cauliflora* e *Plinia* sp. foram realizadas análises comparativas, filogenéticas e evolutivas que apontaram diferenças intra e interespecíficas. Os resultados foram organizados em quatro seções, o que correspondem a quatro artigos, brevemente sumarizados abaixo, visando a melhor compreensão pelos leitores.

Na primeira seção, o genoma plastidial completo de *A. sellowiana* nos permitiu desenvolver pela primeira vez a árvore filogenética utilizando oito genomas plastidiais completos pertencentes à família Myrtaceae. Além disso, adicionamos ao estudo análises comparativas (S/I, Ti/Tv e Ka/Ks) entre os cpDNAs de *A. sellowiana* e *E. uniflora* que proporcionaram avaliar a relação de proximidade evolutiva entre os genomas plastidiais dessas duas espécies. Tais resultados permitiram a elaboração e submissão de um artigo científico, já aceito para ser publicado.

Adicionalmente, foi realizada uma análise filogenética baseada no gene *ycf2* utilizando 43 espécies de Myrtaceae já disponíveis no GenBank e dois genoma plastidial sequenciados neste trabalho *C. xanthocarpa* e *A. sellowiana*. As análises baseadas no gene *ycf2* revelaram que seleção purificadora para este gene. Além disso, o gene *ycf2* foi considerado como um potencial barcode para Myrtaceae. Um segundo artigo foi elaborado e submetido para publicação contendo os resultados destas análises acima referidas.

O genoma plastidial de *E. uniflora*, objeto da terceira seção, foi comparado com o genoma plastidial de um outro indivíduo da mesma espécie quanto a estrutura do genoma, incluindo tamanho, conteúdo

gênico, polimorfismos (SNP e *indels*) e variações de substituição de nucleotídeos. Esses resultados sugerem que os indivíduos analisados são geneticamente diferentes, mas da mesma espécie. Além disso, quando comparadas as variações entre as regiões intergênicas dos genomas plastidiais de *E. uniflora* verificou-se que essa variabilidade foi baixa tanto para SNP quanto para *indel*, assim, sugere-se análises utilizando o genoma completo são mais eficientes que àquelas utilizando parte do genoma em estudos evolutivos. Além disso, esses polimorfismos (possíveis SNPs) identificados neste estudo poderão ser validados e avaliados em futuros estudos de genética de populações envolvendo essa importante espécie frutífera de Myrtaceae. Estes avanços estão contidos no terceiro artigo elaborado, mas ainda não submetido.

Na última seção, o genoma plastidial de três espécies do gênero *Plinia* (*P. aureana*, *P. cauliflora* e *Plinia* sp.) foram sequenciados e tiveram o conteúdo e arranjo dos genes comparados com outros gêneros de Myrtaceae. Além disso, foi realizada a análise filogenômica baseada no genoma plastidial de seis diferentes espécies de Myrtaceae, pertencentes à tribo Myrteae incluindo *P. aureana*, *P. cauliflora* e *Plinia* sp. A árvore de inferência Bayesiana mostrou máximo suporte monofilético para os três grupos de Myrteae: *Pimenta*, *Plinia* e *Eugenia*. Além disso, o grupo *Plinia* revelou maior similaridade entre *P. aureana* e *Plinia* sp., do que com *P. cauliflora*. Estes avanços estão contidos no quarto artigo elaborado neste estudo, mas ainda não submetido.

Adicionalmente aos resultados organizados em quatro artigos, acima referidos, com os dados dos genomas plastidiais de *A. sellowiana*, *C. xanthocarpa*, *E. uniflora*, *P. aureana*, *P. cauliflora* e *Plinia* sp. poderão ser identificadas regiões repetidas do genoma plastidial de cada espécie alvo desse estudo visando a verificação dos possíveis polimorfismos. A partir disso, poderá ser realizado o desenho de iniciadores espécie-específicos para estudos de genética de populações e filogeografia dessas espécies. Além disso, essas informações servem de base para futuros trabalhos a serem realizados no Laboratório de Fisiologia do Desenvolvimento e Genética Vegetal (LFDGV) e em outros laboratórios que se dedicam aos Recursos Genéticos Vegetais.

Os polimorfismos de base única (SNPs), inserções e deleções (*indels*) e marcadores microssatélites identificados nesse estudo poderão

ser testados e validados em trabalhos de genética de populações envolvendo as espécies *A. sellowiana*, *C. xanthocarpa*, *E. uniflora*, *P. aureana*, *P. cauliflora* e *Plinia* sp.

Tomados em conjunto, os resultados obtidos no presente trabalho ampliam o conhecimento sobre a evolução do genoma plastidial das espécies alvo desse estudo, além de, poder elucidar lacunas que envolvem a família Myrtaceae. Esses resultados contribuirão para o conhecimento científico sobre essas frutíferas que demonstram grande potencial para futuros estudos e iniciativas visando o melhoramento e a conservação destas espécies nativas do sul do Brasil.

APÊNDICE

APÊNDICE A - Localização dos SNPs e *indels* entre os genomas plastidiais de Feijoa/Pitanga.

Position	Type	Pitanga	Feijoa	Region
719		G	A	psbA
1448		A	G	psbA
6049		A	C	Intron rps16
6105		C	T	Intron rps16
7622		G	T	IGS
7681		A	G	IGS
11741		T	C	atpA
12401		T	C	atpA
15840		G	T	IGS
17137		T	C	rps2
20524		T	C	rpoC2
20534		G	A	rpoC2
21365	SNP	T	C	rpoC2
23466		T	G	rpoC1
23473		G	T	rpoC1
26239		T	C	rpoB
26245		C	T	rpoB
26340		T	C	rpoB
26363		G	A	rpoB
26551		A	G	rpoB
26581		T	C	rpoB
26821		T	G	IGS
29535		A	C	IGS
29558		T	G	IGS
29610		C	A	IGS

29613	G	T	IGS
29632	G	C	IGS
29667	T	G	IGS
29701	C	T	IGS
30025	G	A	IGS
30067	G	T	IGS
30075	A	G	IGS
30093	C	A	IGS
30676	G	A	IGS
30706	G	T	IGS
30814	G	A	IGS
31712	G	A	IGS
31774	A	C	IGS
31780	G	T	IGS
31789	G	A	IGS
31797	T	C	IGS
36737	C	A	psbC
36860	A	G	psbC
36929	T	C	psbC
36971	C	T	psbC
38119	C	T	trnS-UGA
38152	G	T	trnS-UGA
38179	T	G	trnS-UGA
38188	G	T	trnS-UGA
41604	T	C	psaB
43409	C	G	psaA
43435	A	C	psaA
43534	G	T	psaA
43717	C	G	psaA
46946	T	C	Intron ycf3
52317	C	A	IGS
54808	G	T	Intron trnV-UAC

55439	C	A	IGS
58544	C	A	rbcL
58731	G	C	rbcL
59524	A	C	rbcL
60657	A	C	rbcL
61322	T	C	rbcL
71159	C	A	IGS
73106	T	C	IGS
73446	G	A	IGS
75887	A	G	Intron clpP
79526	G	A	IGS
80809	G	A	Intron petD
80872	G	A	Intron petD
86138	C	T	IGS
86156	C	T	IGS
86867	C	T	rps3
86874	C	T	rps3
86878	G	T	rps3
86920	T	C	rps3
87045	G	A	rps3
115007	C	T	ndhF
118678	A	G	trnL-UAG
119821	A	G	IGS
120868	A	G	ndhD
120973	T	C	ndhD
120982	G	A	ndhD
126429	T	G	ndhA
126645	G	T	ndhH
<hr/>			
30740	-	AC	IGS
55144	<i>Indel</i>	TTACG	IGS
59142	-	T	rbcL
62923	-	T	psaI

67935	-	G	IGS
73341	-	T	IGS
6073	G	-	Intron rps16
7675	T	-	IGS
29516	TAGGA	-	IGS
36133	T	-	psbD
