

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO – UFES
CENTRO DE CIÊNCIAS AGRÁRIAS E ENGENHARIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E MELHORAMENTO

LUARA LOPES SOUSA

**MARCADORES MOLECULARES PARA CARACTERIZAÇÃO GENÉTICA E
GENÔMICA EM ESPÉCIES DE *Psidium***

**ALEGRE – ES
2019**

LUARA LOPES SOUSA

**MARCADORES MOLECULARES PARA CARACTERIZAÇÃO GENÉTICA E
GENÔMICA EM ESPÉCIES DE *Psidium***

Dissertação apresentada ao Programa de Pós-graduação em Genética e Melhoramento da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do título de Mestre em Genética e Melhoramento, na linha de pesquisa Biotecnologia.

Orientadora: Prof.^a Dra. Marcia Flores da Silva Ferreira.

Coorientadores: Prof.^a Dra. Greiciane Gaburro Paneto e Prof. Dr. Adésio Ferreira.

**ALEGRE – ES
Fevereiro – 2019**

Ficha catalográfica disponibilizada pelo Sistema Integrado de Bibliotecas - SIBI/UFES e elaborada pelo autor

L864m Lopes Sousa, Luara, 1993-
Marcadores moleculares para caracterização genética e genômica em espécies de *Psidium* / Luara Lopes Sousa. - 2019. 114 f.

Orientadora: Marcia Flores da Silva Ferreira.
Coorientadores: Greiciane Gaburro Paneto, Adésio Ferreira.
Dissertação (Mestrado em Genética e Melhoramentos)
Universidade Federal do Espírito Santo, Centro de Ciências Agrárias e Engenharias.

1. Genética Vegetal. 2. Biotecnologia. 3. Marcadores Moleculares. 4. SNPs. 5. Microssatélites. 6. *Psidium*. I. Flores da Silva Ferreira, Marcia. II. Gaburro Paneto, Greiciane. III. Ferreira, Adésio. IV. Universidade Federal do Espírito Santo. Centro de Ciências Agrárias e Engenharias. V. Título.

CDU: 631.523

LUARA LOPES SOUSA

**MARCADORES MOLECULARES PARA CARACTERIZAÇÃO GENÉTICA E
GENÔMICA EM ESPÉCIES DE *Psidium***

Dissertação apresentada ao Programa de Pós-Graduação em Genética e Melhoramento do Centro de Ciências Agrárias e Engenharias da Universidade Federal do Espírito Santo, como requisito para a obtenção do título de Mestre em Genética e Melhoramento.

Alegre/ES, 28 de fevereiro de 2019.

Comissão Examinadora:



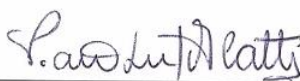
Prof.^a. Dr.^a. Marcia Flores da Silva Ferreira
Universidade Federal do Espírito Santo
(Orientadora)



Prof.^a. Dr.^a. Greiciane Gaburro Paneto
Universidade Federal do Espírito Santo
(Coorientadora)



Pesquisadora Amélia Carlos Tuler
Instituto de Pesquisa Jardim Botânico do Rio de Janeiro
(Examinadora Externa ao Programa)



Pesquisadora Paola de Avelar Carpinetti-Oliveira
Universidade Federal de Viçosa
(Examinadora Externa ao Programa)

Agradecimentos

À Deus por guiar meus caminhos e por me conceder forças para seguir em frente lutando e buscando o meu melhor.

Às pessoas que mais amo, Marta Julieta e Davi Grossi, minha mãe e meu companheiro, pela imensa paciência, carinho e apoio em todas as escolhas que fiz.

Ao Programa de Pós-Graduação em Genética e Melhoramento da Universidade Federal do Espírito Santo pela oportunidade de crescimento e aperfeiçoamento que me concedeu.

À Fundação de Amparo à Pesquisa e Inovação do Espírito Santo – à FAPES, à VALE, CAPES e CNPq, pela bolsa e apoio financeiro deste trabalho.

À Prof.^a Dr.^a. Marcia Flores da Silva Ferreira, pela orientação, ensinamentos e paciência com a minha falta de experiência na área. Agradeço por ter me aceitado como aluna.

Aos Prof.^a Dr.^a. Greiciane Gaburro Paneto e Prof. Dr. Adésio Ferreira pela coorientação, ensinamentos e sempre estarem disponíveis para auxiliar na construção deste trabalho.

Aos demais professores, técnicos e funcionários da Universidade Federal do Espírito Santo, que contribuíram direta ou indiretamente para a minha formação acadêmica, profissional e pessoal.

À Prof. Dr. Enrique Medina e a Dr.^a. Thais Louvain por terem me recebido tão bem no NUDIM/UENF e pela ajuda nos testes e coleta dados.

Aos amigos do laboratório pela disponibilidade em sempre ajudar, por contribuírem para meu crescimento pessoal e profissional e por todos os momentos de alegria que dividimos. Agradeço especialmente ao meu companheiro de Laboratório Matheus Alves pelo trabalho e paciência.

As queridas amigas Luziane e Aléxia pela amizade, palavra amiga e apoio.

Aos membros da banca que se disponibilizaram a ceder seu tempo para contribuir com o crescimento deste trabalho.

À todos que contribuíram para realização deste trabalho.

BIOGRAFIA

Luara Lopes Sousa, nascida em 10 de agosto de 1993 em Itanhém - Ba, filha única de Marta Julieta Ferreira Lopes e Joaquim de Sousa, casada com Davi Reale Grossi.

Em setembro de 2011, ingressou como aluna de graduação no curso Licenciatura em Ciências Biológicas, na Universidade Estadual da Bahia (UNEB), campus de Teixeira de Freitas-Ba. Concluiu o curso em junho de 2016, colando grau em julho de 2016 onde obteve o título de Licenciada em Ciências Biológicas. Na graduação atuou como bolsista de Extensão do Núcleo de Pesquisa e Extensão da UNEB/ Campus X, bolsista de Iniciação Científica (PIBIC) da FAPESB, e bolsista do Programa Institucional de Iniciação à Docência (PIBID).

Em abril de 2017, iniciou o curso de mestrado em Genética e Melhoramento na Universidade Federal do Espírito Santo (UFES), em Alegre-ES, na linha de pesquisa Biotecnologia, sob orientação da professora Marcia Flores da Silva Ferreira. Atuou em atividades de pesquisa com ênfase em marcadores moleculares para espécies vegetais, submetendo-se à defesa de dissertação em fevereiro de 2019.

RESUMO

Marcador molecular pode ser um caracter visível ou um fenótipo molecular proveniente de um segmento específico de DNA, que pode ser expresso ou não. Os marcadores de DNA microssatélites e SNPs são os mais utilizados atualmente, possibilitam diferentes estudos de espécies vegetais por apresentarem vantagens como polimorfismo, abundância e robustez. Muitos desses são conservados, podendo ser estudados em espécies afins àquela para o qual foi desenvolvido. Esses marcadores são importantes no estudo de grupos pouco conhecidos, como as espécies do gênero *Psidium* L., as quais possuem importante papel ecológico e são de grande interesse econômico, entre as quais está a goiabeira (*Psidium guajava*). Neste trabalho objetivou-se desenvolver dois painéis de marcadores moleculares, um de SSR e outro de SNPs, que sejam polimórficos e estejam presentes em regiões conservadas para espécies de *Psidium*. Um painel multiplex, para caracterização interespecífica de 26 espécies do gênero, foi montado por marcadores microssatélites conservados e polimórficos. Os 15 SSR selecionados foram marcados com fluoróforos, testados em combinações e genotipados através de eletroforese capilar, enquanto os marcadores SNPs foram sequenciados por meio do método DArTseq. Para ambos marcadores foram feitas análises descritivas e agrupamentos, sendo um *heatmap* e agrupamento para os microssatélites e uma análise de coordenadas principais (PCoA) para os SNPs. Os marcadores microssatélites e SNPs selecionados apresentaram índices de diversidade úteis para estudos em *P. guajava*. Foi estabelecido um painel multiplex contendo sete marcadores microssatélites para genotipagem eficiente em genótipos de goiabeiras. Um total de 5.951 SNPs polimórficos em sequências conservadas foram selecionados para compor o painel em chip para espécies de *Psidium*. No geral, as duas metodologias foram eficientes e ambos os marcadores informativos e úteis para estudos intra e interespecíficos.

ABSTRACT

Molecular marker can be a visible character or molecular phenotype from a specific segment of DNA, which may or may not be expressed. The DNA markers microsatellites and SNPs are the most used nowadays, allowing different studies of plant species due to their advantages such as polymorphism, abundance and robustness. Many of these are conserved and can be studied in species related to the one for which it was developed. These markers are important in the study of little known groups, such as the species of the genus *Psidium* L., which have an important ecological role and are of great economic interest, among them guava (*Psidium guajava*). In this work we aimed to develop two panels of molecular markers, one of SSR and another of SNPs, which are polymorphic and present in conserved regions for *Psidium* species. A multiplex panel, for interspecific characterization of 26 species of the genus, was assembled by conserved and polymorphic microsatellite markers. The 15 SSRs selected were labeled with fluorophores, tested in combinations and genotyped by capillary electrophoresis, while the SNPs markers were sequenced by the DArTseq method. For both markers, descriptive analyzes and clusters were performed, being a heatmap and dendrogram for the microsatellites and a main coordinate analysis (PCoA) for the SNPs. Selected microsatellite markers and SNPs presented diversity indexes useful for studies in *P. guajava*. A multiplex panel containing seven microsatellite markers was established for efficient genotyping in guava genotypes. A total of 5,951 polymorphic SNPs in conserved sequences were selected to compose the chip panel for *Psidium* species. In general, both methodologies were efficient and both markers informative and useful for intra and interspecific studies.

LISTA DE FIGURAS

REVISÃO BIBLIOGRÁFICA 18

Figura 1. Passo a passo da metodologia DArTseq adaptada para plataformas NGS. 1) Digestão do DNA com uma enzima de restrição de corte coesivo visando reduzir a complexidade do genoma. 2) Ligaç o, com aux lio de enzima de ligaç o, do adaptador *barcodes* a uma das extremidades dos fragmentos clivados e do adaptador comum a outra extremidade. 3) PCR com a adiç o de *primers* apropriados capazes de se ligar aos adaptadores ligados aos fragmentos clivados. 4) A enzima T4 ligase   inativada e uma al quota de cada amostra   reunida e aplicada a uma coluna de exclus o por tamanho para remoç o de adaptadores n o ligados ou falhas. 5) Purificaç o do produto da PCR e avaliaç o dos tamanhos dos fragmentos da biblioteca. 6-7-8) An lise de dados. Fonte: Sansaloni (2012).
..... 26

CAP TULO I – DESENVOLVIMENTO DE PAINEL MULTIPLEX DE MICROSSAT LITES PARA A CARACTERIZAÇ O GEN TICA DE ESP CIES DE *Psidium* 36

Figura 1. Figura 1. Heatmap para ampliaç o dos 15 marcadores SSRs selecionados, em 26 Oesp cies de *Psidium*. As bandas azuis representam a presenç a de ampliaç o, as vermelhas, n o ampliaç o e brancas marcador n o testado em dada esp cie. A linha vermelha representa o ponto de corte feito no agrupamento. Distancia coincid ncia simples. 47

Figura 2. Perfil na ampliaç o em gel de poliacrilamida dos *primers* **A)** mPgCIR 16, **B)** mPgCIR 206 e **C)** mPgCIR 378 para amostragem de quatro populaç es (separadas pela linha pontilhada) de *P. cattleyanum* com 45 indiv duos, sendo: 15 Guarapari; 02 Vale; 12 S o Floriano/Domingos Martins; 16 Santa Rosa de Lima – SC. 48

Figura 3. Gel de Poliacrilamida apresentado   ampliaç o de 12 marcadores selecionados (mPgCIR 02, 10, 16, 20, 29, 104, 187, 206, 208, 216, 285 e 439) para compor o painel multiplex. As ampliaç es est o representadas individualmente e agrupados em tr plex. (Marcador de 100 pb). 55

Figura 4. Perfil de ampliaç o do heptaplex na eletroforese capilar. Picos preenchidos representam marcadores mPgCIR 187 (PET), 439 (NED), 104 (FAM), 216 (VIC), 29 (PET), 206 (VIC) e 208 (NED) respectivamente em amostra de *Psidium guajava*. Setas de cor preta indicam a presenç a de ampliaç es/picos inespec ficos. 55

Figura 5. Perfil de ampliaç o do heptaplex na eletroforese capilar separados por fluoresc ncia em amostra de *P. guajava* Cortibel 21. Cada quadro representa uma fluoresc ncia, sendo elas 6-FAM, VIC, NED e PET respectivamente. Marcadores mPgCIR 187, 439, 104, 216, 29, 206 e 208 respectivamente. Na figura o eixo X (horizontal) indica o tamanho do fragmento, enquanto o eixo Y (vertical) absorv ncia do fluor foro. 56

Figura 6. Perfil de ampliaç o do heptaplex na eletroforese capilar em *P. appendiculatum*. Marcadores representados mPgCIR 206, 187 e 29 com tamanhos de 180/189, 103/105/112 e 165/169/171 respectivamente. A presenç a de mais de dois picos nos locus sugere que essa esp cie seja um poliploide. Seta preta indica que a fluoresc ncia NED se espelhou na

amplificação produzida pelo marcador com fluorescência VIC. Na figura o eixo X (horizontal) indica o tamanho do fragmento, enquanto o eixo Y (vertical) indica graus de absorvância do fluoróforo. 58

Figura 7. Agrupamento obtido através da ligação média entre grupos (UPGMA) para 56 genótipos de goiabeiras, através de marcadores microssatélites genotipados através de eletroforese capilar. Genótipos identificados em três grupos: Cultivares (verde), Cortibel (amarela) e Silvestres.62

CAPÍTULO II - CARACTERIZAÇÃO DE SNPs ORIUNDOS DE DArTseq VISANDO UTILIZAÇÃO EM MÚLTIPLAS ESPÉCIES DE *Psidium*. 76

Figura 1. Diagrama mostrando o panorama geral dos marcadores SNPs genotipados a através da metodologia DArTseq. Os marcadores foram divididos em dois grupos: Ancorados em *E. grandis* e Não Ancorados, dentro de cada grupo eles foram subdivididos em Funcionais e Aleatórios. Essas divisões auxiliaram no estudo e seleção dos marcadores 86

Figura 2. Distribuição de 124.069 marcadores SNPs identificados em nove espécies do gênero *Psidium* nas classes: Funcionais e Aleatórios; Ancorados no genoma de *E. grandis* e ancorados funcionais; Não ancorados e não ancorados funcionais. 90

Figura 3. Distribuição dos marcadores SNPs em três grupos: Para todos, ancorados no genoma de *Eucalyptus* e não ancorados. Em cada grupo os SNPs foram subdivididos para cada espécie em Monomórfico, Polimórficos, Polimórficos Funcionais, Polimórficos Aleatórios e Ausência do marcador (NA). 92

Figura 4. Relação dos marcadores SNPs Funcionais com sua ontologia gênica (GO) para genes relacionado a Processos Biológicos (bp). As informações de cada gráfico estão divididas entre marcadores ancorados e não ancorados ao genoma de *Eucalyptus*. 94

Figura 5. Relação dos marcadores SNPs Funcionais com sua anotação funcional (GO) para genes relacionado a Componentes celulares (cc). As informações de cada gráfico estão divididas entre marcadores ancorados e não ancorados ao genoma de *Eucalyptus*. 95

Figura 6. Relação dos marcadores SNPs Funcionais com sua anotação funcional (GO) para genes relacionado a Função molecular (mf). As informações de cada gráfico estão divididas entre marcadores ancorados e não ancorados ao genoma de *Eucalyptus*. 96

Figura 7. Análise de principais coordenada (PCoA) baseada em marcadores SNPs para 94 amostras de *Psidium*. Cada espécie está representada por cores diferentes para todos os marcadores, ancorados e não ancorados. 98

Figura 8. Apresentação dos marcadores SNPs que foram encontrados em pelo menos 99% das amostras. Cada conjunto de barras representam os marcadores totais, ancorados e não ancorados respectivamente. 99

Figura 9. Agrupamento obtido através dos valores de PIC dos 5.951 marcadores SNPs calculados para 94 amostras de *Psidium*. Espécies identificadas por cor: 63 *P. guajava* (Verde escuro), 05 *P. guineense* (Laranja), 01 Híbrido (Rosa), 04 *P. oblongatum* (Azul escuro), 07 *P. myrtoides* (Lilás), 01 *P. friedrichsthalianum*(Verde claro), 10 *P. cattleyanum* (Vermelho), 01

P. gaudichaudianum (Preto), 01 *P. acutangulum* (Amarelo) e 01 *Psidium sp.* (Azul claro).
..... 101

LISTA DE TABELAS

CAPÍTULO I – DESENVOLVIMENTO DE PAINEL MULTIPLEX DE MICROSSATÉLITES PARA A CARACTERIZAÇÃO GENÉTICA DE ESPÉCIES DE <i>Psidium</i>	36
Tabela 1. Populações naturais de diferentes espécies e genótipos de <i>P. guajava</i> utilizadas para estimar polimorfismo de 15 marcadores microssatélites conservados entre 26 espécies de <i>Psidium</i>	42
Tabela 2. Descrição de 15 marcadores microssatélites selecionados para o desenvolvimento de painel multiplex para múltiplas espécies de <i>Psidium</i> , marcados com fluorocromos e selecionados para compor um painel multiplex. Em negrito estão os marcadores presentes no heptaplex.	46
Tabela 3. Análise descritiva de 15 <i>loci</i> microssatélites utilizados em 92 amostras de <i>Psidium guajava</i> : 20 cultivares, 31 Cortibeis e 41 amostras de quatro populações coletadas no Espírito Santo. Número de alelos por loci (N); heterozigosidade esperada (He); heterozigosidade observada (Ho); Conteúdo de informação polimórfica (PIC).	51
Tabela 4. Análise descritiva de 15 <i>loci</i> microssatélites utilizados em 45 amostras de <i>P. cattleyanum</i> coletadas no Espírito Santo. Número de alelos por loci (N); heterozigosidade esperada (He); heterozigosidade observada (Ho); Conteúdo de informação polimórfica (PIC).	52
Tabela 5. Análise descritiva de 15 <i>loci</i> microssatélites utilizados em 48 amostras de <i>Psidium guineense</i> coletadas no Espírito Santo. Número de alelos por loci (N); heterozigosidade esperada (He); heterozigosidade observada (Ho); Conteúdo de informação polimórfica (PIC).	53
Tabela 6. Análise descritiva de 15 <i>loci</i> microssatélites utilizados em sete amostras de <i>P. grandifolium</i> e 14 amostras de <i>P. myrtoides</i> coletadas no Espírito Santo. Número de alelos por loci (N); heterozigosidade esperada (He); heterozigosidade observada (Ho); Conteúdo de informação polimórfica (PIC).	54
Tabela 7. Análise descritiva dos <i>loci</i> de microssatélites avaliados no painel multiplex. Amostras avaliadas consistem em 48 genótipos de <i>Psidium guajava</i> , sendo 16 cultivares, 30 Cortibel, nove genótipos de ocorrência natural e um híbrido de <i>P. guajava</i> X <i>P. guineense</i>	61
Tabela 8. Análise descritiva dos <i>loci</i> de microssatélites avaliados no painel multiplex. Amostras avaliadas consistem em 24 espécies diferentes de <i>Psidium</i> (<i>P. guajava</i> e <i>P. sartorianum</i> não estão incluídas).	63
CAPÍTULO II - CARACTERIZAÇÃO DE SNPs ORIUNDOS DE DArTseq VISANDO UTILIZAÇÃO EM MÚLTIPLAS ESPÉCIES DE <i>Psidium</i>	76

Tabela 1. Caracterização das espécies utilizadas quanto a Número de Amostras (NA), Ploidia, Número de Cromossomos, Origem/Endemismo e Domínios Fitogeográficos.	79
Tabela 2. Quadro definindo as principais métricas utilizadas para caracterização dos marcadores SNPs através do pacote DArTR.	82
Tabela 3. Descrição, por cromossomo, da distribuição dos 22050 SNPs presentes em espécies de <i>Psidium</i> que foram detectados como ancorados no genoma do <i>Eucalyptus grandis</i>	87
Tabela 4. Análise descritiva para os marcadores selecionados.	100

LISTA DE APÊNDICES

CAPÍTULO I – DESENVOLVIMENTO DE PAINEL MULTIPLEX DE MICROSSATÉLITES PARA A CARACTERIZAÇÃO GENÉTICA DE ESPÉCIES DE <i>Psidium</i>	36
Apêndice 1. Heatmap para amplificação dos 54 marcadores SSRs selecionados, em 22 espécies de <i>Psidium</i> . As bandas azuis representam a presença de amplificação e as vermelhas não amplificação. Distancia coincidência simples.	73
Apêndice 2. Genotipagem das amostras de <i>P. guajava</i>	73
Apêndice 3. Genotipagem das amostras de 24 espécies de <i>Psidium</i>	75
CAPÍTULO II - CARACTERIZAÇÃO DE SNPs ORIUNDOS DE DArTseq VISANDO UTILIZAÇÃO EM MÚLTIPLAS ESPÉCIES DE <i>Psidium</i>	76
Apêndice 1. Médias utilizadas para caracterização de Todos marcadores SNPs (124.069), juntos e separados por espécies.	109
Apêndice 2. Médias utilizadas para caracterização dos marcadores SNPs Ancorados ao genoma de <i>Eucalyptus</i> (22.050), juntos e separados por espécies.	110
Apêndice 3. Médias utilizadas para caracterização dos marcadores SNPs Não Ancorados ao genoma de <i>Eucalyptus</i> (102.019), juntos e separados por espécies.	111
Apêndice 4. Agrupamento obtido através dos marcadores SNPs com valores de PIC menores que 0.02, calculados para 94 amostras de <i>Psidium</i> . Espécies identificadas por cor: 63 <i>P. guajava</i> (Verde escuro), 05 <i>P. guineense</i> (Laranja), 01 Híbrido (Rosa), 04 <i>P. oblongatum</i> (Azul escuro), 07 <i>P. myrtoides</i> (Lilás), 01 <i>P. friedrichsthalianum</i> (Verde claro), 10 <i>P. cattleyanum</i> (Vermelho), 01 <i>P. gaudichaudianum</i> (Preto), 01 <i>P. acutangulum</i> (Amarelo) e 01 <i>Psidium sp.</i> (Azul claro).	112
Apêndice 5. Agrupamento obtido através dos marcadores SNPs com valores de PIC maiores que 0.2, calculados para 94 amostras de <i>Psidium</i> . Espécies identificadas por cor: 63 <i>P. guajava</i> (Verde escuro), 05 <i>P. guineense</i> (Laranja), 01 Híbrido (Rosa), 04 <i>P. oblongatum</i> (Azul escuro), 07 <i>P. myrtoides</i> (Lilás), 01 <i>P. friedrichsthalianum</i> (Verde claro), 10 <i>P. cattleyanum</i> (Vermelho), 01 <i>P. gaudichaudianum</i> (Preto), 01 <i>P. acutangulum</i> (Amarelo) e 01 <i>Psidium sp.</i> (Azul claro).	113
Apêndice 6. Agrupamento obtido através dos marcadores SNPs com valores de PIC intermediários (menores que 0,02 e maiores que 0,2), calculados para 94 amostras de <i>Psidium</i> . Espécies identificadas por cor: 63 <i>P. guajava</i> (Verde escuro), 05 <i>P. guineense</i> (Laranja), 01 Híbrido (Rosa), 04 <i>P. oblongatum</i> (Azul escuro), 07 <i>P. myrtoides</i> (Lilás), 01 <i>P. friedrichsthalianum</i> (Verde claro), 10 <i>P. cattleyanum</i> (Vermelho), 01 <i>P. gaudichaudianum</i> (Preto), 01 <i>P. acutangulum</i> (Amarelo) e 01 <i>Psidium sp.</i> (Azul claro).	114

SUMÁRIO

INTRODUÇÃO	16
2. REVISÃO BIBLIOGRÁFICA	18
2.1 Gênero <i>Psidium</i>	18
2.1.3 <i>Psidium guajava</i>	19
2.2 Marcadores moleculares microssatélites	20
2.2.1 Transferibilidade de microssatélites para identificação de espécies.....	21
2.2.2 Painel multiplex de microssatélites	22
2.3 Nucleotídeo de Polimorfismo Único (SNP).....	23
2.3.1 Diversity Array Technology (DArTseq).....	25
2.3.2 Desenvolvimento de uma coleção de SNPs representativa para <i>Psidium</i>	27
2.4. Referências.....	29
3. OBJETIVOS	35
3.1 Objetivo Geral	35
3.2 Objetivos Específicos.....	35
4. CAPÍTULO I:	36
DESENVOLVIMENTO DE PAINEL MULTIPLEX DE MICROSSATÉLITES PARA A CARACTERIZAÇÃO GENÉTICA DE ESPÉCIES DE <i>Psidium</i>	36
4.1 Introdução	36
4.2 Material e Métodos.....	39
4.2.1 Seleção de Marcadores Microssatélites	39
4.2.2 Caracterização dos marcadores selecionados	40
4.2.3 Desenvolvimento de painéis multiplex de <i>loci</i> microssatélites.....	43
4.2.4 Validação de painéis multiplex de marcadores microssatélites para uso em estudos genéticos.....	44
4.3 Resultados	45
4.3.1 Marcadores microssatélites selecionados	45
4.3.2 Caracterização dos marcadores selecionados em cinco espécies de <i>Psidium</i>	45
4.3.2 Desenvolvimento do painel multiplex	55
4.4 Discussão	64
4.5 Conclusão	68
4.6 Referências	69
4.7 Apêndices.....	73
5. CAPÍTULO II:.....	76
CARACTERIZAÇÃO DE SNPs ORIUNDOS DE DArTseq VISANDO UTILIZAÇÃO EM MÚLTIPLAS ESPÉCIES DE <i>Psidium</i>.....	76

5.1 Introdução	76
5.2 Materiais e Métodos	79
5.2.1 Material Vegetal	79
5.2.2 Extração de DNA, quantificação e incubação	80
5.2.3 Genotipagem por Sequenciamento em plataforma DArTseq	81
5.2.5 Análise de dados	81
5.3 Resultados	86
5.3.1 Caracterização dos marcadores SNPs	86
5.4 Discussão	102
5.5 Conclusão	105
5.6 Referências	106
5.7 Apêndices	109

INTRODUÇÃO

Myrtaceae é uma família composta por plantas com grande representatividade nos ecossistemas brasileiros, sendo destaque na Floresta Atlântica, além de apresentar conhecidas fruteiras nativas do Brasil (BFG 2015; FLORA DO BRASIL 2020 em construção). Esta família é composta por cerca de 144 gêneros e 5.500 espécies (WILSON et al., 2005; WILSON, 2011; WCSP, 2018), destacando-se não apenas pela grande diversidade de espécies, mas também pela complexidade destas (BARROSO et al., 1984). É uma família de fácil reconhecimento morfológico, caracterizando-se como árvores ou arbustos apresentando flores solitárias, agrupadas em racemos ou dicásios axilares, com folhas simples e glândulas oleíferas. Entretanto, a nível de gênero e espécie, é considerada complexa pela grande diversidade de espécies, sobreposição de caracteres e, principalmente pela precariedade de estudos taxonômicos (LANDRUM & KAWASAKI, 1997).

Dentro da família Myrtaceae, o gênero *Psidium* é um dos mais importantes, representativos e explorado (SOBRAL et al., 2014). Este gênero possui distribuição em toda região neotropical, apresentando espécies que ocorrem nos mais variados biomas e sujeitas às mais diferentes pressões ambientais, dificultando a identificação e delimitação de suas espécies (COSTA, 2009). Seus representantes desempenham papel ecológico importante, sendo o fruto fonte de alimento, também apresentam uso medicinal e recuperação de áreas degradadas, além do grande potencial de recursos genéticos a serem descobertos (BEZERRA et al., 2006; GRESSLER et al., 2006; NERI et al., 2011).

Dentro do gênero, o principal representante é a goiabeira (*Psidium guajava* L.), devido ao valor econômico para a indústria e para o comércio. O Brasil se destaca na produção de goiabas vermelhas, sendo um dos maiores produtores mundiais deste fruto. A goiaba é utilizada tanto *in natura* quanto pela indústria na fabricação de doces, sucos e biscoitos, além do seu alto teor de fibra e vitamina C (POMMER E MURAKAMI, 2009). Grande tem sido os avanços em estudos de goiabeiras nos últimos anos, graças ao desenvolvimento de cultivares e à associação com pesquisas moleculares.

Além da goiabeira outras espécies de *Psidium*, conhecidas popularmente como araçás, apresentam possibilidades para estudos devido às características de seu fruto e folhas que podem ser utilizados no comércio de frutas, indústria farmacêutica e estudos

de melhoramento, visando a resistência a fatores bióticos e abióticos. Entretanto, devido à falta de conhecimentos básicos sobre essas espécies poucos tem sido os avanços na utilização desses recursos. Além disso, estudos desta natureza dependem da correta identificação das espécies, procedimento considerado complexo em Myrtaceae, devido não só ao grande número de espécies, mas também a existência de muita semelhança morfológica entre elas e falta de estudos taxonômicos abrangentes (BARROSO et al., 1984; TULER, 2014).

Devido à importância ecológica, econômica e genética deste gênero, cada vez mais estudos moleculares tem se dedicado a produção de conhecimento visando informações genéticas novas que possam ser utilizadas em programas de melhoramento (COSER et al., 2012; TULER et al., 2015; SILVA et al., 2016). Assim sendo, ferramentas biotecnológicas como marcadores moleculares vem sendo amplamente utilizados para diversas finalidades, tais como estudos de diversidade, distância genética, genotipagem, mapeamento de regiões específicas, estudos de associação, seleção genética, etc.

Marcadores moleculares são úteis para estudos de organismos pouco conhecidos, pois permitem mensurar a diversidade genética e as relações entre organismos ao nível do DNA, excluindo influências ambientais (AGARWAL et al., 2008). Dentre os marcadores mais utilizados atualmente em estudos de plantas destacam-se os microssatélites e SNPs. Os microssatélites são amplamente utilizados devido a sua disponibilidade e abundância ao longo do genoma, por apresentarem natureza polimórfica, codominantes e detectados pelas condições de PCR. Os marcadores SNPs, consistem em uma variação na sequência de DNA que afeta somente uma base na sequência do genoma, além disso, podem também se apresentar como uma inserção ou deleção de base.

Em *Psidium*, o uso de marcadores moleculares pode revelar diferenças no genoma nem sempre expressas no fenótipo das espécies, fornecendo informações sobre o *pool* genético de espécies selvagens. Levando em consideração o genoma eucarioto, onde encontramos vários tipos de mutação em grande quantidade, a presença de um marcador pode nos mostrar regiões genômicas funcionais importantes. Desta forma, objetivou-se com a realização deste trabalho, desenvolver dois painéis de marcadores moleculares, um de microssatélites e outro de SNPs, que sejam polimórficos e estejam presentes em sequência conservadas para diferentes genótipos de *Psidium*, contribuindo com seu melhoramento e com o conhecimento do genoma de espécies próximas.

2. REVISÃO BIBLIOGRÁFICA

2.1 Gênero *Psidium*

A grande família Myrtaceae, inserida na ordem Myrtales, é composta de plantas lenhosas, eudicotiledôneas, e abrange quase 6.000 espécies organizadas em 132 gêneros (GOVAERTS et al., 2019; TULER, 2014). Seus representantes estão presentes principalmente nas regiões tropicais e subtropicais do mundo, com poucas espécies ocorrendo nas regiões temperadas (BARROSO, 1984). No Brasil, Myrtaceae representa cerca de 1000 espécies distribuídas em diferentes fitofisionomias (SOBRAL et al., 2014).

Dentro da família Myrtaceae, o gênero *Psidium* L. é um dos que abrange maior riqueza de espécies. Apresenta cerca de aproximadamente 100 espécies no mundo, das quais 59 ocorrem no Brasil (LANDRUM 2003; SOBRAL et al., 2014). Das 38 espécies que ocorrem na Mata Atlântica brasileira, 18 são endêmicas (BFG, 2015; WCSP, 2016), e 15 estão listadas no Espírito Santo, sendo elas: *Psidium brownianum* Mart. ex DC., *Psidium cattleyanum* Sabine., *Psidium cauliflorum* Landrum & Sobral, *Psidium gaudichaudianum* Proença & Faria, *Psidium grazielae* Tuler & Souza, *Psidium guajava* L., *Psidium guineense* Sw., *Psidium longipetiolatum* D.Legrand, *Psidium myrtoides* O.Berg., *Psidium oblongatum* O.Berg., *Psidium oligospermum* DC., *Psidium ovale* (Spreng.) Burret, *Psidium rhombeum* O.Berg, *Psidium rufum* Mart. ex DC., *Psidium sartorianum* (O.Berg) Nied., (TULER et al., 2017; Flora do Brasil 2020 em construção).

O gênero é amplamente distribuído em toda a América e região neotropical, constituído por espécies que ocorrem em diferentes biomas (caatinga, cerrado, florestas amazônica e atlântica, restingas, entre outros) e sujeitas às diferentes pressões ambientais, que ocasionam uma grande plasticidade fenotípica, dificultando a identificação e delimitação de suas espécies (COSTA, 2009). O gênero possui frutos com muitas sementes envolvidos por polpa carnosa, rica em água e carboidratos, e pobre em proteínas e lipídeos (LANDRUM & KAWASAKI, 1997).

Psidium também é constituído por muitas frutíferas nativas, conhecidas como araçás, estas vêm despertando grande interesse para o agronegócio e para indústria farmacêutica, devido aos compostos bioativos e aromáticos encontrados em abundância nos óleos essenciais (DIAS et al., 2015; MENDES et al., 2017). Além disso, representam fonte de variabilidade e resistência, e constituem um importante banco de germoplasma nativo, o qual pode ser usado no melhoramento (SILVA, 2017).

Com exceção de *P. guajava*, todas as espécies do gênero são silvestres e pouco exploradas economicamente. A goiabeira tem valor elevado na economia brasileira para a indústria e para o comércio de frutos *in natura*. Entretanto, algumas espécies nativas também representam fonte de variabilidade e resistência, bem como apresentam interesse para a indústria farmacêutica, pois as frutas são ricas em vitaminas e em substâncias antioxidantes, além dos compostos bioativos e aromáticos encontrados nos óleos essenciais (FRANZON et al., 2009).

Os frutos das espécies silvestres também são comestíveis e apresentam potencial para exploração econômica, tendo polpa suculenta e levemente ácida, com sabor exótico, alto teor de vitamina C, baixo teor de açúcar, elevado teor de compostos fenólicos e sais minerais. Essas espécies podem ser exploradas pela agroindústria para sucos e para uso na fabricação de sorvetes, geleias, doces, licores e outros produtos. A fabricação de doces e geleias, produzidos em pequenas unidades de base familiar, é a principal forma de aproveitamento dos araçazeiros nativos (FRANZON et al., 2009).

Atualmente, o estudo dessas espécies vem sendo obtido principalmente via marcadores moleculares, focando na identificação, no estudo da relação entre espécies, na filogenia e no melhoramento, propiciando modelos para estudos citogenéticos e evolutivos. A caracterização taxonômica, morfológica e genética para as espécies nativas é pouco estudada. Entretanto a partir de estudos com marcadores microsatélites, foi possível verificar proximidade genômica de espécies de *Psidium* dada a grande quantidade de marcadores conservados entre espécies (TULER et al., 2015).

2.1.3 *Psidium guajava*

A goiabeira é uma árvore perene comumente encontrada em quintais. Seu fruto possui de 3 a 6 cm de comprimento e apresenta cores variadas a depender do cultivar (vermelho, branco, roxo ou amarelo) quando maduro. Sendo nativa dos países sul-americanos, a goiabeira está adaptada ao clima subtropical (EMBRAPA, 2015).

A goiaba é considerada uma das frutas mais completas e equilibradas entre as frutas tropicais, por seus constituintes nutricionais e funcionais. O fruto é rico em zinco, fibras e apresenta uma excelente fonte de vitaminas A, B, C e E, além de concentrar altas quantidades de selênio, cobre, fósforo, magnésio, cálcio, ferro e ácido fólico (MANICA et al., 2000; SINGH, 2005). A goiaba contém quatro vezes mais vitamina C

que a laranja e todas as partes da goiabeira são empregadas em aplicações medicinais (GUTIÉRREZ et al., 2008). Por todos esses atributos, aliados aos baixos custos do cultivo, a safra de goiaba é economicamente importante em vários países tropicais e subtropicais (RODRÍGUEZ et al., 2010).

A goiaba apresenta destaque entre as frutas nativas, sendo uma das mais cultivadas no Brasil. Segundo dados da FAO (Organização das Nações Unidas para Alimentação e Agricultura), o Brasil está entre os dez maiores produtores de goiaba do mundo, sendo o maior produtor de goiabas vermelhas. A produção está concentrada nas regiões sudeste e nordeste, nos Estados de São Paulo, Pernambuco e Bahia, que juntos respondem por mais de 80% da produção. O Estado do Espírito Santo contribui com uma produção de 6.231 toneladas de goiaba e uma área plantada de 414 ha (IBGE, 2016).

2.2 Marcadores moleculares microssatélites

O termo marcador refere-se a uma função que identifica algo, é qualquer caráter visível ou um fenótipo molecular proveniente de um segmento específico de DNA. Dessa forma, os marcadores genéticos são utilizados para marcar alelos cuja expressão seja de difícil identificação (BERED et al., 1997). Considerando as alterações nas sequências moleculares, estes podem ser de evolução lenta (adequados para estudos interespecíficos) ou de evolução rápida (apropriados para estudo de indivíduos, famílias e populações) (TISHKOFF, et al., 2000).

Os marcadores microssatélite ou SSR (*Simple Sequence Repeat*) consistem em trechos de DNA repetidas em tandem, abundantes e dispersos por todo o genoma e que possuem comportamento mendeliano (HAMADA et al., 1982). Esses marcadores revelam polimorfismo em um *locus* através da diferença no número de vezes que o motivo se repete, assim seu conteúdo informativo é alto, pois se trata de sequências muito variáveis no genoma. A sequência de DNA que flanqueia estas regiões contendo repetições curtas são conservadas dentro de uma espécie, permitindo a seleção de primers de PCR que podem ser utilizados para amplificar os SSR.

São marcadores codominantes, multialélicos e dependentes de pequena quantidade de DNA dos indivíduos analisados. A análise dos SSR é realizada por meio da técnica de PCR (*Polymerase Chain Reaction*), utilizando *primers* com 18 a 25 bases complementares às regiões que flanqueiam as repetições. Essas variações no número de

repetições são variações no comprimento do segmento detectado pela reação de polimerase em cadeia e na separação de fragmentos amplificados em gel de eletroforese ou na eletroforese capilar (POWELL et al., 1996).

Os microssatélites são classificados quanto ao tipo de unidade de repetição como: 1) simples perfeitos, onde os *loci* de microssatélites são formados por um único motivo (TATATATA); 2) compostos perfeitos, onde mais de um motivo compõe o microssatélite (CACACAGATGATGATGAT); 3) simples ou compostos imperfeitos, os motivos são intercalados por nucleotídeos que não fazem parte da unidade de repetição (WANG et al., 2009).

Em plantas, esses marcadores podem ser amplamente empregados em estudos de diversidade genética, filogenia, estrutura genética de populações naturais, conservação e uso de recursos genéticos, melhoramento vegetal, obtenção de mapas de ligação, identificação de *loci* associados a características importantes ao melhoramento, tanto de natureza qualitativa, quanto quantitativa (QTL). Entretanto, a grande vantagem dos SSRs é a facilidade de troca de informações entre laboratórios, visto que as sequências dos *primers* proporcionam uma linguagem comum e a identificação alélica é direta e objetiva (POWELL et al., 1996).

2.2.1 Transferibilidade de microssatélites para identificação de espécies

Os microssatélites têm sido considerados marcadores genéticos ideais para uma ampla variedade de aplicações. Estudos utilizando os polimorfismos nos tamanhos dos alelos tem sido utilizado na identificação de unidades taxonômicas em diversos grupos: Vitaceae (SEFC et al., 1999), Rosaceae (TESTOLIN et al., 2000), Cupressaceae (BERUBÉ et al., 2003) e Myrtaceae (KIRST et al., 2005). Entre as vantagens desses marcadores estão a obtenção de um grande número de polimorfismos genéticos, a identificação do genótipo sem influência do ambiente e a possibilidade de detecção de polimorfismos em qualquer estágio do desenvolvimento da planta (MIAH et al., 2013).

As maiores restrições quanto ao uso de SSR são o custo e tempo investidos para a construção de uma biblioteca de microssatélites, especialmente para espécies com pouca informação genômica ou de sequenciamento indisponível. Os marcadores microssatélites são normalmente pouco conservados ao longo de grandes distâncias evolutivas, entretanto alguns *loci* microssatélites podem ser conservados entre espécies

que divergiram 100 milhões de anos atrás ou mais (STOLLE, KIDNER & MORITZ, 2013). Levando em consideração o genoma eucarioto, no qual encontramos vários tipos de mutações em grande quantidade, a presença de marcadores conservados nos mostra que tais sequências estão localizadas em regiões genômicas funcionais importantes (GUO, 2013).

Graças à conservação das regiões que flanqueiam o SSR, estudos com transferibilidade de marcadores entre espécies ou gênero próximos, têm sido cada vez mais relatados, a fim de atenuar os problemas gerados com uma nova biblioteca (KALIA et al., 2011). Essas sequências flanqueadoras conservadas podem estar presentes em regiões gênicas, levando-as a acumular menor quantidade de mutações. Entretanto, estando em regiões conservadas, possuem maior tempo de existência e, portanto, maior possibilidade de acúmulo de mutações. Assim, para estudo intraespecíficos e interespecíficos é importante o conhecimento de um grupo de marcadores microssatélites que sejam tão conservados e polimórficos quanto um grupo de marcadores aleatórios são.

Estudos envolvendo a transferibilidade de marcadores entre espécies filogeneticamente próximas vem sendo desenvolvidos e aplicadas com diferentes propósitos em pesquisas com diversas espécies de plantas. O conhecimento de microssatélites altamente conservados permite abordar questões em diferentes linhas de evolução, filogenia e melhoramento com o mesmo conjunto de marcadores. Portanto, ao transferir *primers* de regiões conservadas que flanqueiam os microssatélites para as demais espécies, espera-se que os fenótipos oriundos da amplificação sejam capazes de separar as espécies (TULER, 2014).

2.2.2 Painel multiplex de microssatélites

Apesar das vantagens descritas para os marcadores microssatélites, estes apresentam desvantagens associadas principalmente ao custo e tempo consumido por ensaios individuais. Com a finalidade de aumentar a eficiência da análise de SSR, abordagens de PCR multiplex combinadas com análise baseada em fluorescência e chamada de alelo semiautomática foram desenvolvidas (ZAROURI et al., 2015).

Edwards et al. (1991) foi o primeiro a desenvolver um painel multiplex a partir de PCR com detecção simultâneas de *loci* tri e tetranucleotídeos em sistema fluorescente. Essa estratégia permite amplificar, simultaneamente, mais de um *locus* em

uma única reação, usando múltiplos pares de *primers* (ZAROURI et al., 2015). Os dados obtidos em trabalhos realizados demonstram que o emprego dessa metodologia é eficaz, rápido e altamente informativo, permitindo a genotipagem de um elevado número de acessos, favorecendo o gerenciamento e uso dos recursos genéticos nos programas de conservação e melhoramento.

Para o agrupamento do PCR multiplex é necessário conhecer previamente a variedade de tamanhos de alelos para cada marcador de modo a evitar a sobreposição dos fragmentos e otimizar as condições de PCR a fim de se obter uma boa amplificação de todos os fragmentos alvo (BUTLER, 2005; ZAROURI et al., 2015). Protocolos de PCR multiplex foram desenvolvidos e aplicados com sucesso para diferentes usos em muitas espécies, como genotipagem (ZAROURI et al., 2015), *fingerprinting* (SCHLAUTMAN et al., 2017), identidade e caracterização genética de diferentes espécies de plantas (POSTOLACHE et al., 2013; DRAŠNAROVÁ et al., 2014; MORAIS et al., 2016).

Outros estudos têm sido propostos para melhorar a análise das espécies e construção do painel (DIJK et al., 2014). Entretanto, para a genotipagem de um grande número de marcadores amplificados numa mesma PCR, somente a visualização em gel não é suficiente, pois este pode não separar e apresentar resolução satisfatória das moléculas, ainda mais quando estas possuem mesmo tamanho. Nestes casos, faz-se necessário a eletroforese capilar com fluorescência. Assim, o uso desses painéis, aliado à marcação dos *primers*, aumenta consideravelmente a quantidade de informação gerada e a capacidade de genotipagem de um grande número de acessos, linhagens ou indivíduos em cada reação de amplificação (PESSOA-FILHO, 2007; SCHMIDT, 2009).

Dessa forma, a importância da construção e otimização de um painel multiplex para as espécies de *Psidium* está relacionada à maximização da genotipagem em vários *loci* simultaneamente através de um único PCR. Além disso, quando bem caracterizado o conjunto de marcadores microssatélites deverá apresentar elevado conteúdo informativo para a formação de painéis e análise em várias espécies de *Psidium*.

2.3 Nucleotídeo de Polimorfismo Único (SNP)

O sequenciamento tem mostrado que, como esperado, as sequências genômicas dos indivíduos em uma espécie são muito parecidas e que quase todas as alterações

mais comuns do genoma são baseadas na variação genética em um único nucleotídeo. As substituições de pares de bases desse tipo produzem um grande número de Polimorfismo de Nucleotídeo Único (SNP), que além das substituições, também podem ocorrer como inserções ou deleções de uma única base (indels).

Os SNPs são originados por uma mutação pontual no genoma, que substitui um nucleotídeo por outro. Apesar de ser possível a ocorrência de qualquer um dos quatro nucleotídeos, comumente os SNPs são bialélicos devido à maneira que são originados e distribuídos entre os indivíduos da população (BROOKES, 1999). Tais mutações são do tipo transição, quando há uma substituição de uma purina por outra purina (A-G) ou de uma pirimidina por outra pirimidina (T-C) e do tipo transversão, quando a substituição é de uma purina por uma pirimidina e vice-versa. Quando tais mutações ocorrem em células germinativas, são herdadas pela progênie, transmitidas às gerações seguintes e após algumas gerações, o SNP pode se fixar na população em uma frequência mínima de 1%, quando deixam de ser mutação e passam a ser consideradas polimorfismos (BROOKES, 1999).

Os SNPs são herdados como variantes alélicas, gerando ou não diferenças fenotípicas, a depender da região do genoma em que estão presentes. Em regiões codificadoras, quando a mutação substitui o aminoácido, é denominada mutação não-sinônima. Essa substituição pode ser do tipo conservativa ou não-conservativa em função das características dos aminoácidos envolvidos, podendo ocorrer modificações estruturais e funcionais nas proteínas. Quando as substituições não alteram o aminoácido, elas são denominadas mutações sinônimas (KWOK & GU, 1999).

A diferença dos SNPs em relação aos demais marcadores é que constituem as formas mais frequentes de variações e tem grande distribuição no genoma de várias espécies. A identificação de SNPs mostrou-os como importantes marcadores moleculares para genotipagem, produzindo um conjunto útil de marcadores para mapeamento em escala fina, com grande aplicação para o melhoramento de plantas. O estudo de haplótipos definidos por SNPs em um cromossomo ou segmento, pode fornecer informações importantes sobre as relações entre diferentes grupos e sobre a evolução dos mesmos (VIGNAL et al., 2002; HAYASHI et al., 2004).

Os marcadores SNPs tem sido cada vez mais empregados em trabalhos com plantas, sendo desenvolvidos com sucesso em várias espécies, revelando como marcador molecular informações importantes acerca da diversidade e estrutura genética de uma população. A possibilidade de automação em plataformas de sequenciamento

possibilita, com essa grande quantidade de marcadores moleculares disponíveis, que seja possível realizar estudos de pedigree, distância genética, mapeamento fino de regiões específicas e estudos de associação com características fenotípicas (MAMMADOV et al., 2012; YANG et al., 2013; VIGNAL et al., 2002).

Bhatramaki et al. (2000) apresentaram um trabalho referente à frequência de SNPs em milho, onde foi constatado um polimorfismo a cada 70 bases em média. Evidentemente, a maioria desses polimorfismos ocorre em regiões que não codificam, uma vez que a parcela do genoma que codifica corresponde de 1% a 5%, e, portanto, não produz fenótipos mutantes. Já o trabalho de Setter et al. (2010) usou uma abordagem de associação para identificar SNPs que afetam a acumulação de metabolitos no desenvolvimento de tecidos florais de milho em resposta ao estresse hídrico.

Marcadores SNPs podem ser úteis ao fornecer uma fonte para saturação de mapas genéticos, aumentando sua precisão, ou serem utilizados na análise e caracterização de recursos genéticos, possibilitando um melhor gerenciamento dos mesmos para o melhoramento vegetal. Atuam como ferramentas para seleção assistida, onde os indivíduos de uma população que reunirem o maior número de marcadores mostram-se mais próximos do ideal e portanto são selecionados (GUIMARÃES & COSTA, 2002; RAFALSKI, 2002; BUNDOCK & HENRY, 2004; GUPTA et al., 2004; RAMALHO et al., 2012).

Finalmente, utilizando diferentes abordagens NGS (*Next Generation Sequencing*), em um único experimento, é possível a descoberta e genotipagem de centenas ou milhares de SNPs ao longo do genoma. A genotipagem de SNPs por NGS geralmente é realizada através de uma estratégia multiplex, que consiste no sequenciamento simultâneo de diferentes amostras. Para isto, é necessário adicionar um pequeno adaptador (*barcode*) que difere para cada uma das amostras a serem sequenciadas (SANSALONI, 2012).

2.3.1 Diversity Array Technology (DArTseq)

Diversas metodologias para a detecção de SNPs, através da genotipagem por sequenciamento, surgiram para explorar o potencial das tecnologias NGS, dentre elas está o DArTseq. Inicialmente esta técnica foi desenvolvida para detectar polimorfismos via hibridização de sondas e foi primeiro demonstrada no genoma do arroz por

Jaccound et al. 2001. Entretanto, a partir de 2009, a empresa DArT P/L em Canberra na Austrália (www.diversityarrays.com), otimizou e adaptou a metodologia para ser utilizada em plataforma NGS, sendo nomeada DArTseq (Figura 1).

Trata-se de uma abordagem robusta, tecnicamente simples, multiplex e de alto rendimento que é capaz de identificar centenas de milhares de polimorfismos do tipo SNPs e indels, de forma simultânea em todo genoma. Apresenta custo relativamente baixo, elevada reprodutibilidade e acurácia, capacidade de fornecer informações de todo o genoma e flexibilidade na utilização desta técnica. Além disso, por não depender do conhecimento prévio de sequência, pode ser utilizado em espécies que não possuem dados moleculares ou quando os recursos são limitados (EGEA et al., 2017).

A incorporação da plataforma de Sequenciamento de Nova Geração, aumentou substancialmente o número de fragmentos genômicos analisados e, conseqüentemente, o número de marcadores obtidos, criando assim uma tecnologia com melhor custo benefício comparado ao método DArT inicial (SANSALONI, 2012).

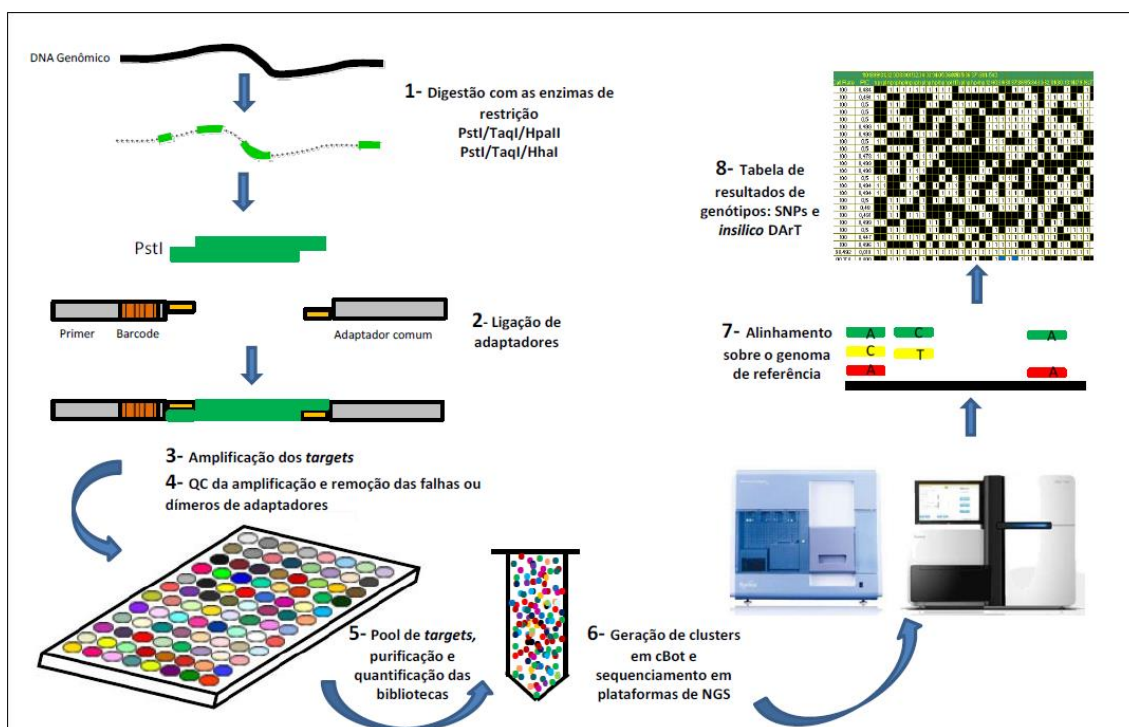


Figura 1. Passo a passo da metodologia DArTseq adaptada para plataformas NGS. 1) Digestão do DNA com uma enzima de restrição de corte coesivo visando reduzir a complexidade do genoma. 2) Ligação, com auxílio de enzima de ligação, do adaptador *barcodes* a uma das extremidades dos fragmentos clivados e do adaptador comum a outra extremidade. 3) PCR com a adição de *primers* apropriados capazes de se ligar aos adaptadores ligados aos fragmentos clivados. 4) A enzima T4 ligase é inativada e uma alíquota de cada amostra é reunida e aplicada a uma coluna de exclusão por tamanho para remoção de adaptadores não ligados ou falhas. 5) Purificação do produto da PCR e

avaliação dos tamanhos dos fragmentos da biblioteca. 6-7-8) Análise de dados. Fonte: Sansaloni (2012).

Tem sido crescente o número de estudos utilizando a abordagem DArTseq em plantas. Suas aplicações mais populares são a caracterização de amostras (VALDISSER *et al.*, 2017) e elucidação de relações filogenéticas (AL-BEYROUTIOVÁ *et al.*, 2016), uma vez que relações resolvidas entre espécies cultivadas e seus parentes silvestres fornecem um recurso importante para o desenvolvimento das culturas, além do conhecimento das espécies silvestres, que são recursos genéticos valiosos para a introdução de características desejáveis.

O estudo de Al-Beyroutiová *et al.* (2016), com acessos de *Secale*, utilizou mais de 13.000 marcadores moleculares obtidos a partir da tecnologia DArTseq e forneceu uma nova visão sobre a taxonomia desse gênero relativamente jovem, sugerindo que diferentes espécies e subespécies surgiram de diferentes acessos de *Secale* sp. SNPs oriundos da metodologia DArTseq, podem também ser aplicados na construção de mapas de ligação altamente saturados, importantes para a detecção de *loci* importantes para seleção assistida por marcadores (REN *et al.*, 2015).

Dados DArTseq também podem ser utilizados em estudos de associação de Genoma Amplo (Genome-Wide Association Studies - GWAS) e Desequilíbrio de Ligação (DL). Nimmakayala *et al.* (2016), obteve a partir de marcadores SNPs um painel GWAS com a melhor representação da diversidade de melões, demonstrando que marcadores SNPs de alta densidade desenvolvidos no estudo fornecem uma ferramenta eficaz para dissecar a arquitetura genética da firmeza dos frutos.

A metodologia DArT possui diversas aplicações, permitindo identificar indivíduos (GARAVITO *et al.*, 2016), caracterizar germoplasmas e analisar diversidade genética (EGEA *et al.*, 2017), mapeamento genético e marcação de genes (REN *et al.*, 2015), seleção assistida por marcadores moleculares (SANTOS *et al.*, 2016) e rastreamento de alterações de metilação do genoma. Além disso, de acordo com as informações disponibilizadas pela empresa Diversity Arrays Technology Pty. Ltd., a metodologia DArTseq pode ser aperfeiçoada visando obter uma densidade diferenciada de marcadores de acordo com os objetivos a serem cumpridos.

2.3.2 Desenvolvimento de uma coleção de SNPs representativa para *Psidium*

Avanços no campo da genética são altamente dependentes de tecnologias capacitadas para realizar análises genômicas de alta resolução. Essas tecnologias têm o potencial de revolucionar nossa capacidade de identificar características de interesse. Com o advento da tecnologia NGS, grandes quantidades de dados de sequência de DNA podem ser geradas em um curto período de tempo. A descoberta de SNP de alto rendimento é uma necessidade devido à disponibilidade de sistemas de análise como a plataforma Infinium da Illumina (<http://www.illumina.com/>) que permite o ensaio de grandes números de SNPs por amostra de DNA em paralelo em um único slide de silício (SONG et al., 2013).

O painel de genotipagem da Illumina, BeadChip é uma placa de vidro que contém milhares de sondas com material genético. O BeadChip, permite aos investigadores genotipar milhares ou milhões de SNPs ou analisar outras variantes genômicas, como o número de cópias, em um grande número de amostras de DNA (ADLER, WILEY & GAFFNEY, 2013). Essa ferramenta é capaz de cruzar informações entre marcadores de DNA e perfis economicamente interessantes, revelando associações importantes de genes e inferindo sobre diversas características de interesse.

Essas ofertas de múltiplas amostras são normalmente usadas para estudos de mapeamento fino por pesquisadores que já realizam Genotipagem de Genoma Inteiro (*Whole genome genotyping* - WGG) ou possuem regiões genômicas específicas com SNPs de interesse. É eficiente para caracterização da diversidade genética e do desequilíbrio de ligação, bem como para construir mapas de ligação de alta resolução e melhorar a montagem da sequência do organismo em estudo. Layouts simples permitem que diferentes amostras sejam aplicadas às diferentes faixas do BeadChip (STEEMERS & GUNDERSON, 2007).

Atualmente, muitos arranjos de SNPs para vários organismos vegetais foram projetados nesta plataforma, e provaram ser bastante eficientes, incluindo arroz (RiceSNP50, CHEN et al., 2014), soja (SoySNP6K, AKOND et al., 2013; SoySNP50K, SONG et al., 2013), milho (MaizeSNP50, UNTERSEER et al., 2014), girassol (BACHLAVA et al., 2012) e muitos outros.

Assim, os SNPs selecionados deverão ser polimórficos para serem úteis no estudo da variação dentro de *P. guajava*, avaliando o desempenho em diversos genótipos comerciais e amostras selvagens. Além disso, com o BeadChip será possível identificar regiões de SNPs comuns para as diferentes espécies de *Psidium* e caracterizá-las quanto a estrutura e função.

2.4. Referências

- Adler, A. J., Wiley, G. B., & Gaffney, P. M. (2013). Infinium Assay for Large-scale SNP Genotyping Applications. *Journal of Visualized Experiments*, 10(81), 1–10. <https://doi.org/10.3791/50683>
- Agarwal, M., Shrivastava, N. and Padh, H. (2008) Advances in Molecular marker techniques and their applications in plant science. *Plant cell rep* 27:617- 613
- Akond, M., Liu, S., Schoener, L., Anderson, J. A., Kantartzi, S. K., Meksem, K., ... Kassem, M. A. (2013). A SNP-Based Genetic Linkage Map of Soybean Using the SoyS - NP6K Illumina Infinium BeadChip Genotyping Array. *Plant Genetics, Genomics, and Biotechnology*, 1(3), 80–89. <https://doi.org/10.5147/jpgs.2013.0090>
- Al-Beyroutiová, M., Sabo, M., Slezniak, P., Dušinský, R., Birčák, E., Hauptvogel, P., ... Švec, M. (2016). Evolutionary relationships in the genus *Secale* revealed by DArTseq DNA polymorphism. *Plant Systematics and Evolution*, 302(8), 1083–1091. <https://doi.org/10.1007/s00606-016-1318-2>
- Bachlava, E., Taylor, C. A., Tang, S., Bowers, J. E., Mandel, J. R., Burke, J. M., & Knapp, S. J. (2012). Snp discovery and development of a high-density genotyping array for sunflower. *PLoS ONE*, 7(1). <https://doi.org/10.1371/journal.pone.0029814>
- Barroso, G.M., Peixoto, A.L., Ichaso, C.L.F., Costa, C.G., Guimarães, E.F. & Lima, H.C. (1984). Myrtaceae. In: Sistemática de Angiospermas do Brasil. Viçosa. v. 2, Ed. Univ. Fed. Viçosa. 114-126 p.
- Bered, F., Barbosa Neto, F. J., Carvalho, F. I. F. (1997). Marcadores moleculares e sua aplicação no melhoramento genético de plantas. *Ciência Rural*, v.27, p.513-520.
- Berubé, Y., Ritland, C., & Ritland, K. (2003). Isolation, characterization, and cross species utility of microsatellites in yellow cedar (*Chamaecyparis nootkatensis*). *Genome*, 361(2001), 353–361.
- Bezerra, J. E. F.; Lederman, I. E.; Silva Junior, J. F.; Proença, C. E. B. Araújo. in: Vieira, R. F.; Costa, T. S. A.; Silva, D. B.; Ferreira; Sano, F. R. *Frutas Nativas da Região Centro-Oeste do Brasil. Brasília: Embrapa Recursos Genéticos e Biotecnologia*, 2006, 320p.
- BFG. (2015). Growing knowledge: an overview of Seed Plant diversity in Brazil. *Rodriguésia*, v. 66, p.1085-1113, 2015. DOI: 10.1590/2175-7860201566411.
- Bhatramaki, D. et al. (2000). Conserved single nucleotide polymorphism (SNP) haplotypes in maize. In: *Plant and Animal Genome Conference*, p. 188, Proceedings, San Diego: Scherago International.
- Brookes, A. J. (1999). The essence of SNPs. *Gene*, 234(2), 177–186. [https://doi.org/10.1016/S0378-1119\(99\)00219-X](https://doi.org/10.1016/S0378-1119(99)00219-X)
- Bundock, P. C., Henry, R. J. (2004). Single nucleotide polymorphism, haplotype diversity and recombination in the *Isa* gene of barley. *TAG. Theoretical and applied genetics.*, v. 109, n. 3, p. 543-551.
- Butler, J. M. (2005) Forensic DNA typing protocols. *Methods Mol Biol* 297:53–65. doi:10.1385/1592598676
- Chen, H., Xie, W., He, H., Yu, H., Chen, W., Li, J., ... Zhang, Q. (2014). A high-density snp genotyping array for rice biology and molecular breeding. *Molecular Plant*, 7(3), 541–553. <https://doi.org/10.1093/mp/sst135>
- Costa, I. R. da (2009). *Estudos evolutivos em Myrtaceae: aspectos citotaxonômicos e filogenéticos em Myrteae, enfatizando Psidium e gêneros relacionados*. Tese (doutorado) – Universidade Estadual de Campinas/ Instituto de Biologia –

- Campinas, SP.
- Dias, C. N., Alves, L. P., Rodrigues, K. A., Brito, M. C., Rosa, C dos, Amaral, F. M. dos, Monteiro O. dos S., Andrade, E. H., Maia, J. G. & Moraes, D. F. (2015). Chemical composition and larvicidal activity of essential oils extracted from brazilian legal Amazon plants against *Aedes aegypti* L. (Diptera: Culicidae). *J. Evid. Based Complementary Altern. Med.*, p. 1–8.
- Drašnarová, A., Krak, K., Vít, P., Doudová, J., Douda, J., Hadincová, V., ... Mandák, B. (2014). Cross-amplification and multiplexing of SSR markers for *Alnus glutinosa* and *A. incana*. *Tree Genetics and Genomes*, 10(4), 865–873. <https://doi.org/10.1007/s11295-014-0727-z>
- Egea, L. A., Mérida-García, R., Kilian, A., Hernandez, P., & Dorado, G. (2017). Assessment of genetic diversity and structure of large garlic (*Allium sativum*) germplasm bank, by diversity arrays technology “genotyping-by-sequencing” platform (DArTseq). *Frontiers in Genetics*, 8(JUL), 1–9. <https://doi.org/10.3389/fgene.2017.00098>
- EMBRAPA (2015). *Costa Sistema de produção de goiabas para pequenos produtores do Rio Grande do Sul*. Editores técnicos: Jair Costa Nachtigal, Carlos Roberto Martins e Glaucia Figueiredo Nachtigal. – Pelotas: Embrapa Clima Temperado, 105 p. (ISSN 1676-7683; 22).
- FAO - CCP, Food and Agriculture Organization of the United Nations Committee on Commodity Problems, Tropical Fruits Compendium (2010).
- Franzon, R. C., Campos, L. Z. de O., Proença, C. E. B., & Sousa-Silva, J. C. (2009). *Araçás do Gênero Psidium: principais espécies, ocorrência, descrição e usos* (No. ISSN 1517-5111, ISSN online 2176-5081 ; 266). Planaltina, DF.
- Garavito, A., Montagnon, C., Guyot, R., & Bertrand, B. (2016). Identification by the DArTseq method of the genetic origin of the *Coffea canephora* cultivated in Vietnam and Mexico. *BMC Plant Biology*, 16(1), 1–12. <https://doi.org/10.1186/s12870-016-0933-y>
- Govaerts, R., Sobral, M., Ashton, P., Barrie, F., Holst, B., Landrum, L., Lucas, E., Matsumoto, K., Mazine, F., Proença, C., Soares-Silva, L., Wilson, P. & Niclughdha, E. (2019+). World Checklist of selected plant families – Myrtaceae. *The Board of Trustees of the Royal Botanic Gardens, Kew*. Disponível em <<http://apps.kew.org/wcsp/>>. Acesso em 02 de mai de 2019.
- Gressler, E., Pizo, M. A., & Morellato, L. P. C. (2006). Polinização e dispersão de sementes em Myrtaceae do Brasil. *Revista Brasileira de Botânica*, 29(4): 509-530.
- Guimarães, P. E. M., Costa, M. C. R. (2002). SNPs: Sutis diferenças em um código. *Biotecnologia Ciência e Desenvolvimento*, n.26, maio/junho 2002.
- Guo, Y. L. (2013). Gene family evolution in green plants with emphasis on the origination and evolution of *Arabidopsis thaliana* genes. *Plant Journal*, v. 73, n. 6, p. 941–951, jan.
- Gupta, V. et al. (2004). Molecular tagging of erucic acid trait in oilseed mustard (*Brassica juncea*) by QTL mapping and single nucleotide polymorphisms in FAE1 gene. *Theoretical and Applied Genetics*, v.108, p. 743 - 749.
- Gutiérrez, R. M. P., Mitchell, S., & Solis, R. V. (2008). *Psidium guajava*: A review of its traditional uses, phytochemistry and pharmacology. *Journal of Ethnopharmacology*, 117(1), 1–27. <https://doi.org/10.1016/j.jep.2008.01.025>
- Hamada, H., Petrino, M. G., Kakunaga, T. A. (1982). A novel repeat element with Z-DNA-forming potential is widely found in evolutionary diverse eucariotic genomes. *Proceedings National Academy Science United States America*, v. 79, p. 6465- 6469.

- Hayashi, K., Hashimoto, N., Daigen, M., & Ashikawa, I. (2004). Development of PCR-based SNP markers for rice blast resistance genes at the Piz locus. *Theoretical and Applied Genetics*, 108(7), 1212–1220. <https://doi.org/10.1007/s00122-003-1553-0>
- IBGE (2016). Disponível em <https://sidra.ibge.gov.br/Tabela/106#resultado>. Acesso em: 17 julho, 2019.
- Kalia, R. K., Rai, M. K., Kalia, S., Singh, R., Dhawan, A. K. (2011). Microsatellite markers: an overview of the recent progress in plants. *Euphytica*. 177: 09–334.
- Kirst, M., Cordeiro, C. M., Rezende, G. D. S. P., & Grattapaglia, D. (2005). Power of microsatellite markers for fingerprinting and parentage analysis in *Eucalyptus grandis* breeding populations. *The Journal of Heredity*, 96(2), 161–6.
- Kwok, P. Y.; Gu Z. (1999). Single nucleotide polymorphism libraries: why and how are we building them?. *Molecular Medicine Today*, v. 5, p. 538 - 543.
- Landrum, L.R. (2003). A revision of the *Psidium salutare* complex. *Sida* 20: 1449-1469.
- Landrum, L.R. & Kawasaki, M.L. (1997). The genera of Myrtaceae in Brasil: an illustrated synoptic treatment and identification keys. *Brittonia* 49: 508-536.
- Lima, L. S. (2007). *Identificação de polimorfismo em ESTs da interação cacau-Moniliophthora perniciosa*. Dissertação (Mestrado) – Universidade Estadual de Santa Cruz/ Faculdade de Ciências Biológicas e Ambientais – Ilhéus, Ba.
- Mammadov, J., Aggarwal, R., Buyyarapu, R., Kumpatla, S. (2012). Snp markers and their impact on plant breeding. *Int. J. Plant Genom.* 728398.
- Manica, I. *Frutas nativas, silvestres e exóticas I: técnicas de produção e mercado: abiu, amora-preta, araçá, bacuri, biriba, carambola, cereja-do-rio-grande, jabuticaba*. Porto Alegre: Cinco Continentes, 2000. 327 p.
- Mendes, L. A., Martins, G. F., Valbon, W. R., Souza, T. S. de, Menini, L., Ferreira, A., Ferreira, M. F. da S. (2017). Larvicidal effect of essential oils from Brazilian cultivars of guava on *Aedes aegypti* L. *Industrial Crops & Products*, v. 108, p. 684–689.
- Miah, G., Rafii, M. Y., Ismail, M. R., Puteh, A. B., Rahim, H. a, Islam, K. N., & Latif, M. A. (2013). A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance. *International Journal of Molecular Sciences*, 14(11): 499–528.
- Morais, S. R. P. de, Vieira, A. F., Almeida, L. C. da S., Rodrigues, L. A., Melo, P. G. S., Faria, L. C. de, ... Souza, T. L. P. O. de. (2016). Application of microsatellite markers to confirm controlled crosses and assess genetic identity in common bean. *Crop Breeding and Applied Biotechnology*, 16(3), 234–239. <https://doi.org/10.1590/1984-70332016v16n3n35>
- Neri, A. V.; Soares, M. P.; Neto, J. A. A.; Dias, L. E. (2011). Espécies de cerrado com potencial para recuperação de áreas degradadas por mineração de ouro, Paracatu-Mg. *Revista Árvore*, Viçosa-MG.35(4):907-918.
- Nimmakayala, P., Tomason, Y. R., Abburi, V. L., Alvarado, A., Saminathan, T., Vajja, V. G., ... Reddy, U. K. (2016). Genome-Wide Differentiation of Various Melon Horticultural Groups for Use in GWAS for Fruit Firmness and Construction of a High Resolution Genetic Map. *Frontiers in Plant Science*, 7(September), 1–15. <https://doi.org/10.3389/fpls.2016.01437>
- Pessoa-Filho, M., Belo, A., Alcochete, A., Hideo, P. N. R., Ferreira, M. E. (2007). A set of multiplex panel of microsatellite markers for rapid molecular characterization of rice accessions. *BMC Plant Biology*. 1-10 7:23doi:10.1186/1471-2229-7-23.
- Postolache, D., Leonarduzzi, C., Piotti, A., Spanu, I., Roig, A., Fady, B., ... Vendramin, G. G. (2013). Transcriptome versus Genomic Microsatellite Markers: Highly Informative Multiplexes for Genotyping *Abies alba* Mill. and Congeneric Species.

- Plant Molecular Biology Reporter*, 32(3), 750–760.
<https://doi.org/10.1007/s11105-013-0688-7>
- Powell, W.; Machray, G. C.; Provan, J. (1996). Polymorphism revealed by simple sequence repeats. *Trends in Plant Science*, Oxford, n. 7, v. 1, p. 215 – 222.
- Psidium* in Flora do Brasil 2020 em construção. Jardim Botânico do Rio de Janeiro. Disponível em:
 <<http://floradobrasil.jbrj.gov.br/reflora/floradobrasil/FB605023>>. Acesso em: 17 Ago, 2018.
- Ramalho, M. A. P. *et al.* (2012). *Genética na Agropecuária UFLA* (Vol. 5). Lavras.
- Rafalski, A. (2002). Applications of single nucleotide polymorphisms in crop genetics. *Current opinion in plant biology*, v. 5, n. 2, p. 94-100.
- Ren, R., Ray, R., Li, P., Xu, J., Zhang, M., Liu, G., ... Yang, X. (2015). Construction of a high-density DArTseq SNP-based genetic map and identification of genomic regions with segregation distortion in a genetic population derived from a cross between feral and cultivated-type watermelon. *Molecular Genetics and Genomics*, 290(4), 1457–1470. <https://doi.org/10.1007/s00438-015-0997-7>
- Rodríguez, N. N., Valdés, J., Rodríguez, J. A., Velásquez, J. B., Rivero, D., Martínez, F., ... Cañizares, J. (2010). Genetic resources and breeding of guava (*Psidium guajava* L.) in Cuba. *Journal of Genetics*, 2–5.
- Sansaloni, C. P. (2012). *Desenvolvimento e aplicações de DArT (Diversity Arrays Technology) e genotipagem por sequenciamento (Genotyping-by-Sequencing) para análise genética em Eucalyptus*. Tese (doutorado) – Universidade de Brasília / Departamento de Biologia Celular – Brasília, DF.
- Santos, J. P. R. dos, Pires, L. P. M., Vasconcellos, R. C. C., Pereira, G. S., Pinho, R. G. Von, & Balestre, M. (2016). Genomic selection to resistance to *Stenocarpella maydis* in maize lines using DArTseq markers. *BMC Genetics*, 17(1), 1–10. <https://doi.org/10.1186/s12863-016-0392-3>
- Schlautman, B., Bolivar-Medina, J., Hodapp, S., & Zalapa, J. (2017). Cranberry SSR multiplexing panels for DNA horticultural fingerprinting and genetic studies. *Scientia Horticulturae*, 219, 280–286. <https://doi.org/10.1016/j.scienta.2017.03.005>
- Schmidt, A. B. (2009). *Desenvolvimento de painéis multiplex de marcadores microsatélites e mapeamento de QTLs de tolerância à seca e ao frio em linhagens puras recombinantes de arroz (Oryza sativa L.)*. Tese (doutorado) – Universidade Federal de Santa Catarina / Centro de Ciências Agrárias – Florianópolis, SC.
- Sefc, K. M., Regner, F., Turetschek, E., Glössl, J., & Steinkellner, H. (1999). Identification of microsatellite sequences in *Vitis riparia* and their applicability for genotyping of different *Vitis* species. *Genome*, 42(3): 367–73.
- Setter, T. L., Yan, J., Warburton, M., Ribaut, J. M., Xu, Y., Sawkins, M., Buckler, E. S., Zhang, Z., Gore, M. A. (2010). Genetic association mapping identifies single nucleotide polymorphisms in genes that affect abscisic acid levels in maize floral tissues during drought. *J Exp Bot* 62:701–716
- Silva, S. N. (2017). *Diversidade de genótipos de goiabeiras por caracteres vegetativos, reprodutivos e viabilidade polínica*. Tese (Doutorado) – Universidade Federal do Espírito Santo/ Centro de Ciências Agrárias – Alegre, ES.
- Silva, E. F.; De Araújo, R. F.; Martins, C. S. R.; Martins, L. S. S.; Veasey, E. A. (2016). Diversity and genetic structure of natural populations of araçá (*Psidium guineense* Sw.). *Revista Caatinga*, v. 2125, p. 37-44.
- Singh, G. (2005). High density planting in guava application of canopy architecture. *ICAR News* 11, 9–10.

- Sobral, M.; Proença, C.; Souza, M.; Mazine, F.; Lucas, E. (2014). 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: 16 Ago. 2018.
- Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., & Cregan, P. B. (2013). Development and Evaluation of SoySNP50K, a High-Density Genotyping Array for Soybean. *PLoS ONE*, 8(1), 1–12. <https://doi.org/10.1371/journal.pone.0054985>
- Stemmers, F. J., & Gunderson, K. L. (2007). Whole genome genotyping technologies on the BeadArray platform. *Biotechnology Journal*, 2(1), 41–49. <https://doi.org/10.1002/biot.200600213>
- Stolle, E., Kidner, J. H., & Moritz, R. F. A. (2013). Patterns of evolutionary conservation of microsatellites (SSRs) suggest a faster rate of genome evolution in hymenoptera than in Diptera. *Genome Biology and Evolution*, 5(1), 151–162. <https://doi.org/10.1093/gbe/evs133>
- Testolin, R., Marrazzo, M. T., Cipriani, G., Quarta, R., Verde, I., Dettori, M. T., Pancaldi, M., Sansavini, S. (2000). Microsatellite DNA in peach (*Prunus persica* L. Batsch) and its use in fingerprinting and testing the genetic origin of cultivars. *Genome* 43:512–520
- Tishkoff, S. A.; Pakstis, A. J.; Stoneking, M.; Kidd, J. R.; Destro-Bisol, G.; Sanjantila, A.; Lu, R. B.; Deinard, A. S.; Sirugo, G.; Jenkins, T.; Kidd, Kk.; Clark, A. G. (2000). Short tandem-repeat polymorphism/Alu haplotype variation at the PLAT locus: implications for modern human origins. *Am J Hum Genet.* 67(4):901-25.
- Tuler, A. C. (2014). *Diversidade de Psidium L. (Myrtaceae) no Estado do Espírito Santo, Brasil*. Dissertação (mestrado) – Instituto de Pesquisas Jardim Botânico do Rio de Janeiro / Escola Nacional de Botânica Tropical – Rio de Janeiro, RJ.
- Tuler, A. C., Carrijo, T. T., Ferreria, M. F. S., Peixoto, A. L., Santo, E., P, P., & Santo, E. (2017). Flora of Espírito Santo: *Psidium* (Myrtaceae), 68(5), 1791–1805. <https://doi.org/10.1590/2175-7860201768515>
- Tuler, A. C., Carrijo, T. T., Nóia, L. R., Ferreira, A., Peixoto, A. L., & Ferreira, M. F. da S. (2015). SHORT COMMUNICATION SSR markers: a tool for species identification in *Psidium* (Myrtaceae). *Molecular Biology Reports*. <https://doi.org/10.1007/s11033-015-3927-1>
- Unterseer, S., Bauer, E., Haberer, G., Seidel, M., Knaak, C., Ouzunova, M., ... Schön, C. C. (2014). A powerful tool for genome analysis in maize: Development and evaluation of the high density 600 k SNP genotyping array. *BMC Genomics*, 15(1), 1–15. <https://doi.org/10.1186/1471-2164-15-823>
- Valdisser, P. A. M. R., Pereira, W. J., Almeida Filho, J. E., Müller, B. S. F., Coelho, G. R. C., de Menezes, I. P. P., ... Vianello, R. P. (2017). In-depth genome characterization of a Brazilian common bean core collection using DArTseq high-density SNP genotyping. *BMC Genomics*, 18(1), 1–19. <https://doi.org/10.1186/s12864-017-3805-4>
- Vignal, A., Milan, D., SanCristobal, M., & Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics Alain. *Genetics Selection Evolution*, 34(March 2002), 275–305. <https://doi.org/10.1051/gse:2002009>
- Wang, Z., Gerstein, M., Snyder, M. (2009). RNA-seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, London. 10:57-63.
- WCSP. (2016). World Checklist of Selected Plant Families, apps.kew.org/wcsp/. Acesso em 6 de Junho de 2018.

- Yang, W., Kang, X., Yang, Q., Lin, Y., & Fang, M. (2013). Review on the development of genotyping methods for assessing farm animal diversity. *Journal of Animal Science and Biotechnology*, 4(1), 2–7. <https://doi.org/10.1186/2049-1891-4-2>
- Zarouri, B., Vargas, A. M., Gaforio, L., Aller, M., de Andrés, M. T., & Cabezas, J. A. (2015). Whole-genome genotyping of grape using a panel of microsatellite multiplex PCRs. *Tree Genetics and Genomes*, 11(2). <https://doi.org/10.1007/s11295-015-0843-4>

3. OBJETIVOS

3.1 Objetivo Geral

Desenvolver dois painéis de marcadores moleculares, um de SSR e outro de SNPs, conservados e polimórficos para espécies de *Psidium*.

3.2 Objetivos Específicos

1. Estudar e caracterizar geneticamente as espécies do gênero *Psidium*, por marcadores moleculares do tipo SNPs e Microssatélites (SSR).
2. Identificar marcadores microssatélites conservados desenvolvidos para *P. guajava* e transferidos para 26 espécies do gênero;
3. Desenvolver e validar um painel multiplex de marcadores microssatélites polimórficos útil para diferentes espécies de *Psidium*;
4. Estudar as relações genéticas entre 26 espécies de *Psidium* através do uso do marcador molecular microssatélites;
5. Detectar SNPs interespecíficos no genoma de espécies de *Psidium*;
6. Identificar e caracterizar um grupo de marcadores SNPs com múltiplas finalidades para espécies de *Psidium*;
7. Caracterizar marcadores conservados que estão ancorados no genoma de *Eucalyptus*, quanto a sua distribuição e polimorfismo;
8. Classificar os SNPs identificados como funcionais e aleatórios e conhecer as principais variantes entre as espécies;
9. Caracterizar relação intraespecífica e interespecífica entre espécies do gênero *Psidium* e cultivares de *P. guajava*;
10. Caracterizar o germoplasma selvagem e cultivado de espécies de *Psidium* spp.;
11. Selecionar SNPs para constituir um painel de SNPs de alta resolução para genotipagem de espécies do gênero *Psidium*.

4. CAPÍTULO I:

DESENVOLVIMENTO DE PAINEL MULTIPLEX DE MICROSSATÉLITES PARA A CARACTERIZAÇÃO GENÉTICA DE ESPÉCIES DE *Psidium*

4.1 Introdução

Psidium L. é um gênero com ampla distribuição por toda região neotropical, com cerca de 95 espécies vegetais, das quais 59 ocorrem no Brasil (LANDRUM 2003, SOBRAL *et al.*, 2014). Várias espécies apresentam importância econômica, ecológica e farmacêutica, com destaque para *Psidium guajava*, uma frutífera amplamente cultivada. Na família Myrtaceae o gênero *Psidium* está entre os de maior riqueza de espécies, seus integrantes representam grande potencial de recursos genéticos, fonte de variabilidade e resistência, constituindo um importante banco de germoplasma nativo a ser devidamente caracterizado (SILVA, 2017).

Os marcadores moleculares de DNA microsatélites, também conhecidos como SSR (simples sequências repetidas) são reconhecidamente importantes ferramentas moleculares em estudos genéticos de espécies vegetais, os quais são utilizados na caracterização das variações genéticas em populações naturais (ANDERSON *et al.*, 2018), em cultivares (HONIG *et al.*, 2018), em estudos de delimitação entre espécies (TULER *et al.*, 2015), em análises de *fingerprinting* (SCHLAUTMAN *et al.*, 2017), no mapeamento genético (DIÉGUEZ *et al.*, 2018) e na seleção de indivíduos com características relevantes (KHERWAR *et al.*, 2018).

Os SSR são amplamente usados por suas vantagens quanto ao polimorfismo, abundância e robustez. Estes marcadores são desenvolvidos para espécies individuais, por um procedimento que demanda custo e tempo, tanto pela construção de bibliotecas de DNA genômico ou enriquecidas, quanto pela busca de microsatélites *in silico* em bancos de sequências. Assim, para a maioria das espécies vegetais não existem estes marcadores desenvolvidos. Entretanto, a conservação de microsatélites entre espécies relacionadas que divergiram é um fato conhecido (STOLLE, KIDNER & MORITZ, 2013) e estudos de transferibilidade tem sido úteis para detecção de regiões microsatélites comuns em diferentes espécies de um grupo relacionado (NOGUEIRA *et al.*, 2015; TULER *et al.*, 2015; FAGUNDES *et al.*, 2016; BERNARDES *et al.*, 2018). Este tipo de estudo tem especial relevância para grupos de espécies nativas pouco

estudadas e sem ferramentas genômicas desenvolvidas, a exemplo das espécies do gênero *Psidium* (TULER, 2018).

Considerando estudos de diversidade genética em uma espécie, um bom marcador molecular deve ser eficiente para a identificação de polimorfismos entre indivíduos e populações. Entretanto, a presença de marcadores comuns entre espécies distintas, gera perspectivas de identificação de alelos específicos por espécies, bem como a possibilidade de estudar a diversidade genética de populações distintas em um mesmo ambiente com as mesmas regiões genômicas, podendo gerar um banco de dados genéticos de espécies relacionadas, que podem servir a múltiplos estudos. No gênero *Psidium*, alta frequência de SSR comuns entre espécies foram relatados em estudos de transferibilidade (TULER *et al.*, 2015; BERNARDES *et al.*, 2018), evidenciando um cenário favorável para o desenvolvimento de métodos de utilização destes marcadores para múltiplas espécies do gênero.

O potencial de identificação de polimorfismo de um marcador em populações naturais ou experimentais pode ser estimado considerando o número e a frequência dos alelos amostrados, visando a maior probabilidade de detecção de genótipos divergentes. Esta estimativa é o conteúdo de informação polimórfica (PIC) do marcador, o qual é utilizado para identificação de *locus* importantes visando diferentes propósitos (KANUPRIYA *et al.*, 2011; CHAITHANYA *et al.*, 2015; MEHMOOD *et al.*, 2015).

Visando reduzir custos, trabalho e tempo da genotipagem por microssatélites em diferentes populações, é possível otimizar reações de PCR para ampliações em multiplex, nas quais diferentes *loci* são amplificados em uma única reação de PCR. Esta reação é possível pelo uso de marcação dos *primers* com fluorocromos distintos para diferentes *loci*, possibilitando que mesmo amplicons de tamanhos semelhantes possam ser amplificados juntos, sendo visualizados por fluorescências distintas (SCHMIDT 2009). Amplificações em multiplex para genotipagem são casos de sucesso para microssatélites para humanos (KOHLRUSS *et al.*, 2018), animais (LI *et al.*, 2018; BEUGIN *et al.*, 2017; MUNYARD *et al.*, 2009;) e vegetais (SCHLAUTMAN *et al.*, 2017; MORAIS *et al.*, 2016; ASHKANI *et al.*, 2015; DIJK *et al.*, 2014; POSTOLACHE *et al.*, 2013).

A amplificação de diferentes *loci* em reação de PCR multiplex é influenciada por diferentes fatores, sendo necessário um conhecimento prévio dos marcadores candidatos observando sua faixa de amplificação, qualidade dos amplicons, conteúdo polimórfico e interação dos *primers*. Os marcadores devem ser testados em grupos de

dois ou três e visualizados em gel para a padronização do alelo de referência. Com os *primers* marcados, uma padronização das concentrações de cada *um* é necessária, pois estes apresentam comportamentos distintos com sutis mudanças na reação de PCR.

No gênero *Psidium*, a alta taxa de transferibilidade de marcadores microssatélites entre espécies (TULER *et al.*, 2015; BERNARDES *et al.*, 2018), evidencia a possibilidade de otimização e utilização de marcadores comuns que possam ser de múltiplas utilidades. Neste sentido, o desenvolvimento de painéis multiplex de *loci* microssatélites conservados em espécies de *Psidium* torna-se uma alternativa rápida prática e de baixo custo, viabilizando estudos nestas espécies. Este capítulo dedica-se à otimização e validação de um painel multiplex, em *P. guajava*, a goiabeira, a partir de marcadores SSR conservados em múltiplas espécies do gênero, visando utilizações intra e interespecíficas do mesmo grupo de marcador.

De forma complementar destaca-se que marcadores microssatélites são importantes ferramentas por serem considerados regiões neutras no genoma e, portanto, acumularem mutações aleatórias, constituindo uma poderosa ferramenta na explicação da variabilidade natural existente. Entretanto, no caso das regiões microssatélites conservadas entre espécies, devem-se considerar que estes pressupostos podem não ser satisfeitos e o nível de polimorfismo ser reduzido. Sendo necessário, portanto, a análise de polimorfismos dos *loci* em estudo de diferentes populações naturais, antes da otimização em painel multiplex, a fim da utilização apenas de *loci* com alta capacidade polimórfica. Dessa forma, neste capítulo objetivou-se desenvolver um painel multiplex com *loci* microssatélites conservados para 26 espécies de *Psidium*, sendo este painel desenvolvido e validado usando 69 genótipos da espécie referência *P. guajava*.

4.2 Material e Métodos

4.2.1 Seleção de Marcadores Microssatélites

A seleção dos marcadores microssatélites para o desenvolvimento do painel multiplex, foi realizada baseada em dados de amplificação do conjunto de marcadores de *P. guajava* desenvolvidos por Risterucci *et al.* (2005) e Guavamap (2008) em 26 espécies de *Psidium*. Dentre esse conjunto, foram selecionados 54 marcadores a partir de resultados de transferibilidade em 13 espécies do gênero *Psidium* (TULER *et al.*, 2015) e entre gêneros da família Myrtaceae (BERNARDES *et al.*, 2018). Nestes trabalhos foram observados o número de espécies amplificadas, número de alelo e qualidade de amplificação em gel. Os dados de transferibilidade obtidos por Tuler *et al.* (2015) foram para as espécies: *P. brownianum*, *P. cattleyanum*, *P. cauliflorum*, *P. guajava*, *Psidium grazielae*, *P. guineense*, *P. longipetiolatum*, *P. myrtoides*, *P. oblongatum*, *P. oligospermum*, *P. ovale*, *P. sartorianum* e *Psidium* spp.1 (caatinga).

Os 54 loci SSR selecionados foram avaliados, neste trabalho, para outras 13 espécies (Apêndice 1): *Psidium acutangulum* DC., *Psidium appendiculatum* Kiaersk., *Psidium bahianum* Landrum & Funch, *Psidium bergianum* (Nied.) Burret, *Psidium firmum* O. Berg, *Psidium friedrichsthalianum*, *Psidium ganevii* Landrum & Funch, *P. gaudichaudianum* Proença & Faria, *P. grandifolium* Mart. ex DC., *Psidium myrsinites* DC., *Psidium rhombeu* O. Berg, *Psidium robustum* O. Berg e *Psidium rufum* Mart. ex DC. Um indivíduo de cada espécie foi utilizado. Ao final do trabalho, indivíduos das 26 espécies foram utilizados para validação do painel multiplex a fim de uso interespecífico. A validação para a espécie *Psidium guajava* foi realizada em 63 indivíduos.

A partir dos resultados de transferibilidade dos 54 *loci* nas 26 espécies de *Psidium*, foram selecionados 15 marcadores para o desenvolvimento do painel multiplex. A seleção destes marcadores foi realizada visando o desenvolvimento de um painel útil à um maior número de espécies, utilizando os critérios abaixo descritos:

- 1- Conservação do microssatélite entre as 26 espécies de estudo. Para avaliação, o *locus* deve estar presente em cada espécie;
- 2- Polimorfismo – foram selecionados apenas marcadores que apresentaram no mínimo quatro alelos interespecíficos nas análises realizadas por Tuler *et al.* (2015). Esse parâmetro estabelece diferença entre o *locus* em diferentes espécies;

3- Ausência de interação entre *primers* na extremidade 3' (com a utilização do Programa PerlPrimer). Esse tipo de interação faz os *primers* anelarem entre si, o que atrapalha a amplificação das sequências durante a PCR;

4- Qualidade de amplificação em gel, de acordo com o descrito por (TULER *et al.*, 2015; BERNARDES *et al.*, 2018). Esse parâmetro auxilia a filtrar *primers* que anelam em vários locais e formam muitas bandas inespecíficas, o que pode atrapalhar durante a genotipagem;

5- Variação na faixa de tamanho dos fragmentos amplificados entre *loci* microssatélites. A variação de tamanho permite amplificar em uma mesma PCR vários *primers* que sejam marcados com mesma fluorescência, o que aumenta o poder de informatividade do multiplex.

A partir dos dados de transferibilidade foi gerada uma matriz de dados que se baseou na amplificação dos 15 marcadores selecionados em 25 espécies avaliadas juntamente com *P. guajava* como controle. Os dados foram representados pela presença (1) ou ausência (0) de amplificação, a dissimilaridade foi estimada entre as espécies e entre os marcadores usando o índice de coincidência simples. Os dados de amplificação foram apresentados no formato *heatmap* e analisados utilizando o software R versão 3.5.0 (R CORE TEAM, 2018).

4.2.2 Caracterização dos marcadores selecionados

Os 15 marcadores microssatélites selecionados (Tabela 2) foram otimizados para a reação em dois genótipos comerciais de *P. guajava*, as goiabeiras Paluma e Cortibel RM, visando delimitar os tamanhos de fragmentos amplificados e a qualidade de amplificação de cada marcador.

A reação de PCR para amplificação dos microssatélites foi de 14 μ L, composta por: 50 ng de DNA genômico; 1X de tampão comercial da Genbiotech; 0,21 μ M de cada *primer* (*forward* e *reverse*); 2,14 mM MgCl₂; 0,86 mM dNTPs; 0,17 μ g/ml de (BSA) e 1,2 U de Taq DNA polimerase. Foi utilizado o programa: 94°C por 4 minutos; 35 ciclos compostos por 94°C por 45 segundos; temperatura de anelamento do primer 55°C por um minuto; 72°C por um minuto; uma extensão final de 72°C por 8 minutos.

Os produtos de amplificação foram corados com gel red e aplicados para separação por eletroforese em géis de poliacrilamida 10% e tampão TBE 1 X a 100

volts com duração de aproximadamente 3 horas. Após a eletroforese os géis foram fotografados por sistema de fotodocumentação (ChemiDoc XRS+System - Bio-Rad™)

Posteriormente, estes 15 marcadores foram submetidos a análise de polimorfismos em diferentes espécies utilizando amostragens de populações naturais de *Psidium*, bem como genótipos de goiabeiras dos quais: 31 genótipos denominados Cortibel, (XVIII ao XLVIII) e 19 cultivares comerciais (Chinesa Branca, Indiana, Kuse, Maçã, Paluma, Pedro Sato, Roxa, Século XXI, Tailandesa, Tailandesa Branca, Petri, Cortibel LG, Cortibel LM, Cortibel Branca LG, Cortibel RM, Cortibel Branca RM, Cortibel RM2, Cortibel RG, Cortibel SLG) (Tabela 1).

Após as genotipagens dos 15 microssatélites em gel, os dados foram analisados e codificados quanto ao polimorfismo. As análises realizadas foram estatísticas descritivas através do programa GENES (CRUZ, 2013) para as espécies diploides e manualmente no programa Excel para os poliploides. Para cada marcador microssatélite foi contabilizado:

Número de alelos observados em cada *locus* (No) dentro de cada população;

$$No = \sum_{j=1}^L a_j$$

Heterozigosidade esperada (*He*), definida como uma fração estimada de todos os indivíduos que poderiam ser heterozigóticos de um *loco*. Se trata do complementar do somatório do quadrado das frequências alélicas (*Pi*)

$$He = 1 - \sum Pi^2$$

Heterozigosidade observada (*Ho*), refere-se ao número de indivíduos heterozigotos (*n*) dividido pelo número total de amostras (*N*) dentro de cada população;

$$Ho = \sum_{k < k'}^{a_j} \frac{n_{ijkk'}}{N_i}$$

Conteúdo informativo de polimorfismo (PIC), que indica a qualidade do marcador em estudos genéticos.

$$PIC = 1 - \sum_{k=1}^{a_j} \hat{p}_{ijk}^2 - \sum_{k < k'}^{a_j} 2\hat{p}_{ijk}^2 \hat{p}_{ijk'}$$

Tabela 1. Populações naturais de diferentes espécies e genótipos de *P. guajava* utilizadas para estimar polimorfismo de 15 marcadores microssatélites conservados entre 26 espécies de *Psidium*.

Populações			
Espécie	Local/Identificação		Tamanho amostral
<i>P. cattleyanum</i>	Guarapari		15
	Vale		02
	São Florianópolis/Domingos Martins		12
	Santa Rosa de Lima – SC		16
Total			45
<i>P. grandifolium</i>	Licínio de Almeida -		07
Total			07
<i>P. guineense</i>	Guarapari		08
	Águia Branca		08
	Itaúnas		08
	São Florianópolis/Domingos Martins		08
	Mimoso		08
	Venda Nova do Imigrante		08
Total			48
<i>P. myrtooides</i>	Guarapari		07
	Vale		07
Total			14
<i>P. guajava</i> Populações	Guarapari		08
	Itapemirim		15
	Itaúnas		05
	Águia Branca		13
Cultivares	Chinesa Branca	Petri	20
	Indiana	Cortibel LG	
	Kuse	Cortibel LM	
	Maçã	Cortibel Branca LG	
	Paluma	Cortibel RM	
	Pedro Sato	C. Branca RM	
	Roxa	Cortibel RM2	
	Século XXI	Cortibel RG	
	Tailandesa	Cortibel SLG	
	Tailandesa Branca	Goiaba amarela	
Cortibel	18 a 48		31
Total			92

Todos os DNAs foram extraídos de amostras de folhas, pertencentes ao Laboratório de Genética e Melhoramento Vegetal do Centro de Ciências Agrárias da Universidade Federal do Espírito Santo (CCA/UFES) de acordo com as modificações feitas pelo CTAB no método Doyle & Doyle (1990). A integridade e a concentração das amostras foram verificadas por espectrofotômetro *Nanodrop*TM 2000 (*Thermo Scientific*).

4.2.3 Desenvolvimento de painéis multiplex de *loci* microssatélites

A partir dos 15 marcadores selecionados, as combinações dos *primers* para amplificação em multiplex foram avaliadas usando o programa AutorDimer (VALLONE & BUTLER, 2004), visando identificar a interferência de anelamento entre *primers*. Esta análise minimiza o número de reações e possibilitou a avaliação de 12 microssatélites marcados com fluorocromos para uso em PCRs simultâneos.

Para amplificação multiplex os *primers* foram marcados com quatro fluorescências diferentes (NED, VIC, PET e 6-FAM) (Tabela 2). As reações de PCR para testar os painéis multiplex foram realizadas com o Kit QIAGEN® PCR Multiplex, com volume total de 12,5 µL composta por: 50 ng de DNA; 1X de Multiplexer PCR Master Mix; 1X Q-Solution e 0,21 µM de cada primer (*forward* e *reverse*). As reações foram amplificadas sob uma desnaturação inicial de 15 minutos a 95° C, para ativação da enzima, seguida de 35 ciclos compostos por 94°C por um minuto; temperatura de anelamento 55°C por um minuto; 72°C por um minuto; com uma extensão final de 10 minutos a 72°C.

Inicialmente foram organizados quatro grupos contendo três marcadores com tamanhos diferentes e marcados com mesma fluorescência (Tabela 2). Esses tríplex foram organizados de modo que marcadores com mesmo tamanho não possuíssem a mesma fluorescência. Testes foram feitos para amplificação individual de cada grupo e para amplificação de todos juntos. Posteriormente, um heptaplex foi padronizado com concentrações de 0,48 µM de cada primer para os marcadores mPgCIR 29 e 187; 0,24 µM para mPgCIR 104, 206, 216 e 439; e para o mPgCIR 208 foi padronizado uma concentração de 0,8 µM.

Para a eletroforese capilar, a 1µL dos produtos de amplificação foi adicionada 9,35 µL de formamida (Hi-Di Formamida, Applied Biosystems) e 0,4µL do padrão de peso molecular GeneScan LIZ 500, marcado com o fluorocromo LIZTM (fluorescência laranja) (Applied Biosystems). Os produtos de amplificação foram separados por eletroforese em capilar utilizando o equipamento ABI Prism™ 310 Genetic Analyzer com matriz para eletroforese capilar de fragmentos DS-33.

4.2.4 Validação de painéis multiplex de marcadores microssatélites para uso em estudos genéticos

Entre os 12 *primers* marcadores, apenas 7 amplificaram em conjunto com sucesso, estes marcadores são compostos por três motivos tetranucleotídeos, um di e três duplos, di e trinucleotídeos (Tabela 1). O painel heptaplex otimizado e padronizado foi submetido à validação nos genótipos de goiabeiras relatados na Tabela 2 (18 cultivares, 30 genótipos denominados Cortibel, 10 genótipos de ocorrência natural oriundos de diferentes locais e 1 híbrido de *P. guajava* com *P. guineense*). Também foram avaliados nas demais 24 espécies (a exceção de *P. sartorianum*) utilizadas neste estudo. Durante a padronização e validação do painel, foram observados parâmetros como número de alelos em cada marcador, interação entre marcadores, ausência de picos falsos nas corridas e não sobreposição de picos para minimizar as possibilidades de erros na genotipagem dos eletroferogramas.

As avaliações dos marcadores em goiabeiras foram feitas através de estatística descritiva através do programa GENES (CRUZ, 2013). Para cada marcador presente no painel foi estimado o número de alelos observados por *locus*, a heterozigosidade observada, esperada e o PIC. A partir da matriz de distância euclidiana entre os genótipos de goiabeira analisados pelo heptaplex, foi obtido o agrupamento através da ligação média entre grupos (UPGMA), utilizando o programa GENES (CRUZ, 2013).

4.3 Resultados

4.3.1 Marcadores microssatélites selecionados

Os 54 marcadores testados inicialmente, selecionados com base no trabalho de Tuler *et al.* (2015), e que apresentaram alta transferibilidade em 26 espécies de *Psidium*, provenientes de outros estados e fitofisionomias (Apêndice 1). Destes marcadores 85% foram transferidos para 25 espécies diferentes. Todos os marcadores foram amplificados em *P. guineense*, evidenciando a proximidade filogenética com *P. guajava*.

Considerando a alta taxa de transferibilidade nas 26 espécies (Figura 1), foi possível selecionar os marcadores com melhor qualidade de amplificação em gel e sem interação na extremidade 3' dos *primers*, visando evitar formação de bandas inespecíficas e interferências na amplificação entre *loci* (Tabela 2). Assim foram selecionados 15 marcadores para avaliação da amplificação em multiplex. Esses marcadores apresentaram taxa de 90% de transferibilidade, sendo que nove espécies amplificaram em todos os marcadores.

O heatmap obtido mostra dados de transferibilidade. Os marcadores mPgCIR 02, 187 e 439 amplificaram para todas as espécies, demonstrando a conservação destas regiões genômicas no gênero (Figura 1). O marcador com menor transferibilidade foi o mPgCIR 16, com 69% de amplificação.

Pelo agrupamento das espécies verificou-se a formação de três grupos. No grupo 1 nove espécies amplificaram todos os marcadores, neste grupo, com exceção de *P. guajava* todas as espécies são poliploides além de serem encontrados basicamente em Floresta Atlântica (FLORA DO BRASIL 2020, em construção). *P. cauliflorum*, espécie muito diferente morfológicamente, continua distante ao se agrupar. Próximo dela, as espécies *P. myrsinites* e *P. robustum* foram as que menos amplificaram, evidenciado que, quanto mais distantes, menor o compartilhamento de regiões. Dessa forma, observamos que a taxa de amplificação diminui à medida que aumenta a divergência genética entre as espécies, esse fato foi discutido por Kalia *et al.* (2010).

4.3.2 Caracterização dos marcadores selecionados em cinco espécies de *Psidium*

Os 15 *loci* microssatélites foram polimórficos, informativos E presentes em pelo menos 90% das espécies. Nas populações o número médio de alelos variou entre 1,7 a 3

Tabela 2. Descrição de 15 marcadores microssatélites selecionados para o desenvolvimento de painel multiplex para múltiplas espécies de *Psidium*, marcados com fluorocromos e selecionados para compor um painel multiplex. Em negrito estão os marcadores presentes no heptaplex.

<i>Loci</i>	Fluorescência	Fragmento (pb)*	Primer Foward	Primer Reverse	Motivo	**N. de espécies	***N. de espécies	**N° de alelos
mPgCIR02	6 - FAM	199	AGTGAACGACTGAAGACC	ATTACACATTCAGCCACTT	(GA)20	13	26	5
mPgCIR20	6 - FAM	268	TATACCACACGCTGAAAC	TTCCCCATAAACATCTCT	(CT)14(CA)17	09	21	6
mPgCIR104	6 - FAM	120	ATTCCCGTGGATTATGTATC	ACAACCATTTTCTCCTCATC	(GGTT)3	10	24	5
mPgCIR10	VIC	287	GTTGGCTCTTATTTTGGT	GCCCCATATCTAGGAAG	(CT)12	12	22	4
mPgCIR206	VIC	180	GGAAGTTTCAAAGTAACAGC AC	AGAATGAGTCCATGCTCAAA	(GA)8/(GT)11	12	24	5
mPgCIR216	VIC	131	GATGGAAGTCAATGTATGA	AGACCTGCTCTGACTTGTGA	(CGTA)3	12	25	4
mPgCIR16	NED	280	AATACCAGCAACACCAA	CATCCGTCTCTAAACCTC	(TC)25	13	18	5
mPgCIR208	NED	240	AGCACGATGTAGGAAGAGAC	CGGCAAGCTCAACAATATAA	(GA)29 / (GAG)5	10	24	10
mPgCIR439	NED	116	GCATCTTGCTTCTGTCACTT	GGAGATGTGCAACGTATTTT	(CA)8(ACAC)	11	26	4
mPgCIR29	PET	168	CTCGCTTCAATCTCCATCTA	AGCGACACAGACTCTTCATT	(GA)8/(GAA)5	09	23	4
mPgCIR187	PET	116	AAGCCTAATAATGCGGAAGT	TCTTTCCCAGGATGGAGTA	(GGAA)3	10	26	4
mPgCIR285	PET	220	CTTGGGTGTCTCTGTGTCT	CAGCCAGGGCATAAGTATAG	(GA)11(GA)18	12	17	4
mPgCIR231		177	CTCCAAGAAAATGGAAAGG	TGAAAACACCAAACAGCAC	(GT)9/(GA)12	10	23	5
mPgCIR378		162	GTCACTTGCCTGTTGATTTT	GCTTTCTCATTTCTTAACCAAG	(CT)10(CT)6	08	25	6
mPgCIR182		181	GAGGAAGAAACCCGAAGTTA	GGTAGAAAGATCGGAAAGAC		13	25	13

Dados dos marcadores presentes no heptaplex apresentados em negrito.

* Descrito para o genótipo que os marcadores foram desenvolvidos.

** Para as espécies estudadas por Tuler *et al.* (2015)

*** Amplificação para as espécies avaliadas neste estudo.

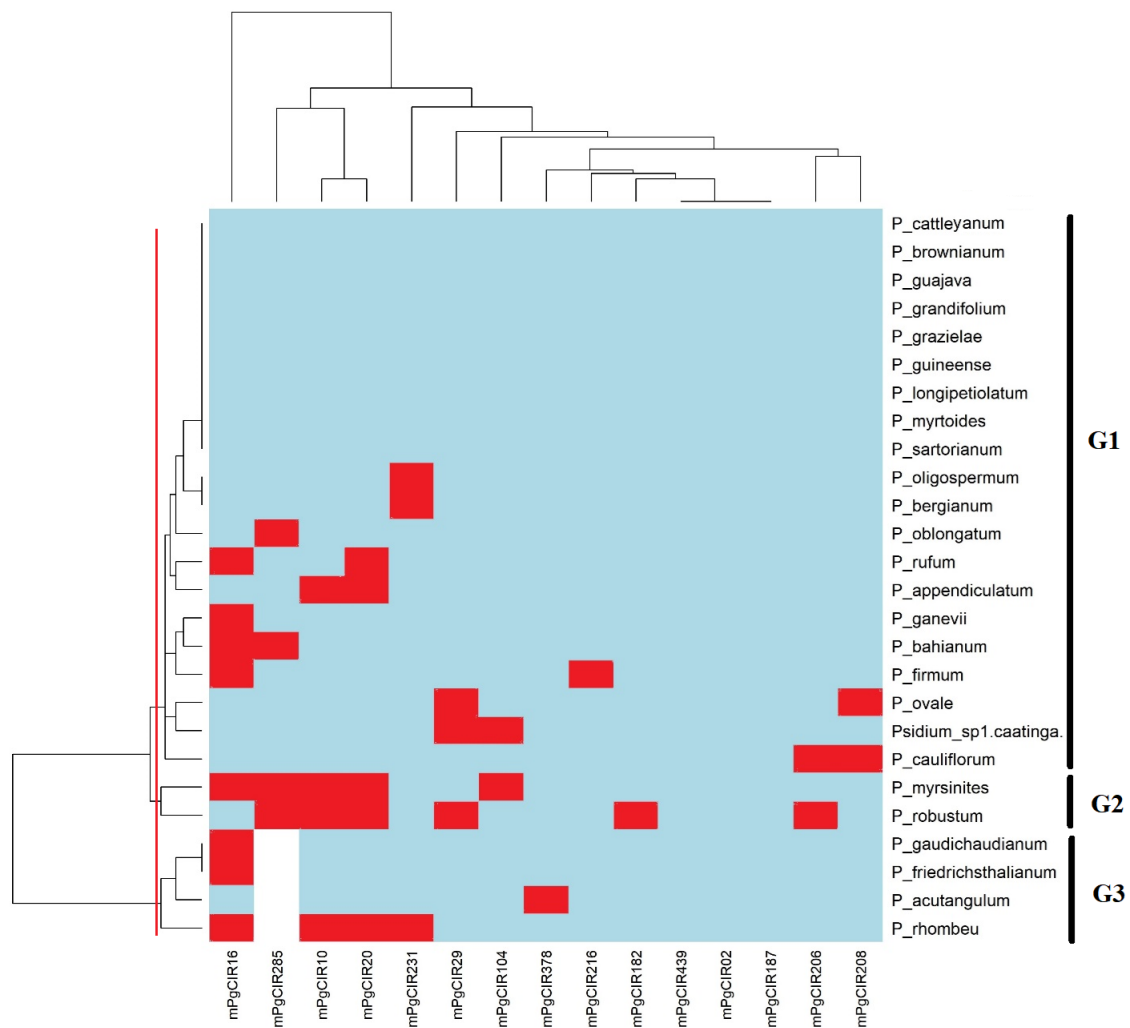


Figura 1. Heatmap para amplificação dos 15 marcadores SSRs selecionados, em 26 espécies de *Psidium*. As bandas azuis representam a presença de amplificação, as vermelhas, não amplificação e brancas marcador não testado em dada espécie. A linha vermelha representa o ponto de corte feito no agrupamento. Distância coincidência simples.

(Tabelas 3 a 6). Sendo que cada população foi analisada de forma independente das outras, o maior número de alelos foi visualizado nos géis das populações de *P. cattleyanum* provavelmente por ser um poliploide, apresentou até seis alelos em algumas amostras de população de Santa Catarina (Figura 2). A população de *P. myrtoides* apresentou o menor número de alelos (17), ficando abaixo de *P. grandifolium* (25 alelos) representado por apenas sete indivíduos (Tabela 6). Para *P. guineense* foram identificados 36 alelos nos 48 indivíduos (média de 3,2 alelos por *locus*) (Tabela 5). Para os três grupos de *P. guajava* (20 cultivares, 31 Cortibel e população) 37, 32, 36 alelos respectivamente foram identificados (Tabela 3).

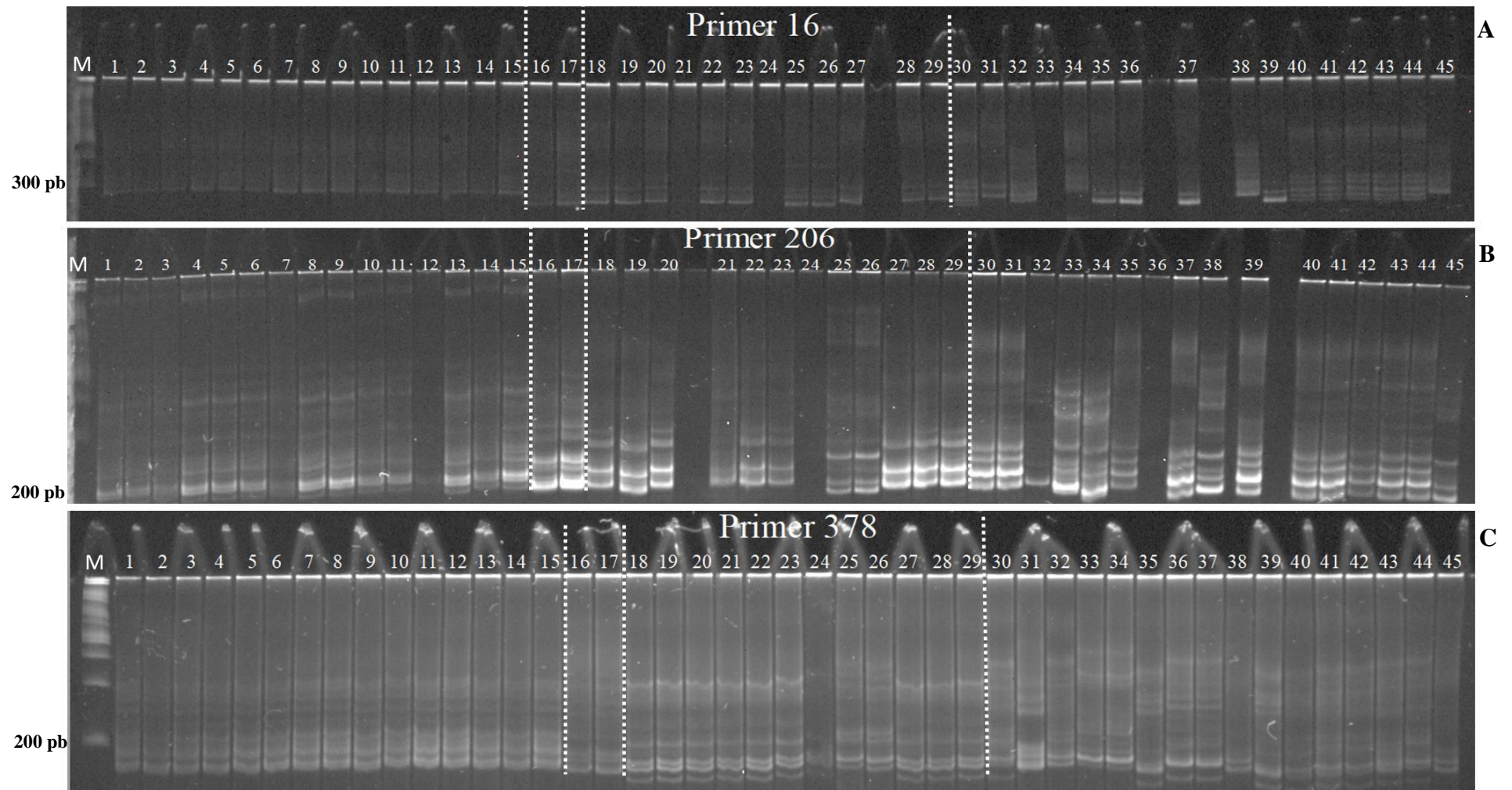


Figura 2. Perfil na amplificação em gel de poliacrilamida dos *primers* A) mPgCIR 16, B) mPgCIR 206 e C) mPgCIR 378 para amostragem de quatro populações (separadas pela linha pontilhada) de *P. cattleyanum* com 45 indivíduos, sendo: 15 Guarapari; 02 Vale; 12 São Floriano/Domingos Martins; 16 Santa Rosa de Lima – SC.

A heterozigosidade esperada (*He*), apresentou valores entre 0,2 (mPgCIR 208 e 439) à 0,8 (mPgCIR 104, 206, 285 e 378) nas avaliações em *P. guajava* e *P. cattleyanum* (Tabela 3 e 4) e com a maior média geral (0,63) entre populações sendo representada por *P. guineense* (Tabela 5). A *He* média dos *loci* em *P. guajava* foi de 0,41, sendo que o menor valor foi atribuído ao grupo dos 31 genótipos Cortibel, que também apresentou menores valores de alelos, *Ho* e PIC, este grupo é oriundo de uma mesma população de polinização aberta, justificando sua semelhança (Tabela 3). A média da *He* foi superior para o marcador mPgCIR 285 (0,66), esse *loci* apresentou até seis alelos nas populações de *P. cattleyanum* e *P. guineense*, o que explica seus valores se apresentarem sempre altos. Em contrapartida o marcador mPgCIR 231, apresentou um valor médio (0,32) abaixo do esperado para estes marcadores previamente selecionados (Tabelas 7).

A heterozigosidade observada (*Ho*) apresentou valores entre 0,15 (mPgCIR 216 para as cultivares (Tabela 3) a 1 em pelo menos duas populações para cada marcador testado, com exceção do mPgCIR 439, demonstrando mesmo assim a alta heterozigosidade desses *loci*. *Psidium cattleyanum* e *P. guineense* apresentaram frequentemente mais de dois alelos por *locus* e, portanto, foram analisados como poliploides e apresentaram os maiores valores para *Ho*, com uma média variando de 0,76 a 0,89 (Tabelas 4 e 5).

Em todas as populações a média de *Ho* foram maiores que *He*. O marcador mPgCIR 20 apresentou 100% de heterozigosidade em todas as amostras que foi observada, entretanto houve muito dado perdido para este *locus*. A *Ho* dos *loci* em *P. guajava* também foi alta, com uma média que variou de 0,54 a 0,78. Em relação ao marcador, esse grupo apresentou altos valores para a *Ho*, com exceção para o mPgCIR 216 (Tabela 3).

O conteúdo de informação polimórfica (PIC) apresentou uma média que varia de 0,27 (mPgCIR 439) a 0,60 (mPgCIR 285) por marcador, com média geral de 0,43. Para os marcadores selecionados, nenhum foi considerado pouco informativo, mPgCIR 104, 285 e 378 foram considerados muito informativos, apresentando médias acima de 0,5. Os outros 12 marcadores foram considerados medianos para informatividade. A média de maiores PICs foram representados pelas populações de *P. cattleyanum* (0,54) e *P. guineense* (0,55) (Tabelas 4 e 5). Para *P. guajava* a média do PIC variou de 0,23 a 0,37 (Tabelas 3). A populações de *P. grandifolium* e *P. myrtoides* apresentaram os menores valores médios (0,29 e 0,28 respectivamente) (Tabelas 6).

Considerando apenas os valores de todas as populações para os marcadores que compõe o heptaplex (mPgCIR 29, 104, 182, 206, 208, 216 e 439), marcados em negritos nas tabelas, a *Ho* média foi de 0,71 para a amostragem geral, indicando um favorecimento aos alelos heterozigotos nestes *loci*. O PIC foi representado por uma média de 0,43 indicativo de que esses *loci* são considerados medianamente informativos, entretanto, considerando a característica conservadas desses marcadores pode se inferir que eles possuem um alto grau de polimorfismo, atributo que torna eles desejáveis para compor um primeiro painel multiplex de marcadores microssatélites para genótipos de *P. guajava*.

Tabela 3. Análise descritiva de 15 *loci* microssatélites utilizados em 92 amostras de *Psidium guajava*: 20 cultivares, 31 genótipos Cortibel e 41 amostras de quatro populações coletadas no Espírito Santo. Número de alelos por *loci* (*N*); heterozigosidade esperada (*He*); heterozigosidade observada (*Ho*); Conteúdo de informação polimórfica (PIC).

<i>Locus</i>	Cultivares (2n=22)				Genótipos Cortibel (2n=22)				<i>P. guajava</i> (2n=22)				Média			
	N	He	Ho	PIC	N	He	Ho	PIC	N	He	Ho	PIC	N	He	Ho	PIC
mPgCIR 02	3	0,54	1,0	0,44	2	0,5	1,0	0,37	2	0,5	1,0	0,37	2,3	0,51	1,00	0,39
mPgCIR 10	4	0,68	0,94	0,63	4	0,61	1,0	0,53	4	0,55	0,94	0,46	4,0	0,61	0,96	0,54
mPgCIR 16	---	---	---	---	2	0,5	1,0	0,37	---	---	---	---	2,0	0,50	1,00	0,37
mPgCIR 20	2	0,5	1,0	0,37	2	0,5	1,0	0,37	4	0,54	1,0	0,44	2,6	0,51	1,00	0,39
mPgCIR 29	4	0,54	1,0	0,44	2	0,5	1,0	0,37	2	0,5	1,0	0,37	2,6	0,51	1,00	0,39
mPgCIR 104	3	0,33	0,1	0,30	2	0,03	0,03	0,03	2	0,49	0,97	0,37	2,3	0,28	0,37	0,23
mPgCIR 182	3	0,57	1,0	0,48	3	0,53	1,0	0,41	4	0,68	1,0	0,62	3,3	0,59	1,00	0,50
mPgCIR 187	1	0,0	0,0	0,0	2	0,06	0,0	0,06	2	0,04	0,05	0,04	1,6	0,03	0,02	0,03
mPgCIR 206	4	0,64	1,0	0,58	2	0,06	0,0	0,06	2	0,47	0,77	0,36	2,6	0,39	0,59	0,33
mPgCIR 208	2	0,5	1,0	0,37	2	0,06	0,06	0,06	---	---	---	---	2,0	0,28	0,53	0,22
mPgCIR 216	3	0,56	0,15	0,47	2	0,03	0,03	0,03	2	0,34	0,45	0,28	2,3	0,31	0,21	0,26
mPgCIR 231	2	0,5	1,0	0,37	1	0,0	0,0	0,0	2	0,5	1,0	0,37	1,6	0,33	0,67	0,25
mPgCIR 285	3	0,54	1,0	0,44	3	0,59	1,0	0,51	4	0,55	0,94	0,45	3,3	0,56	0,98	0,47
mPgCIR 378	2	0,5	1,0	0,37	2	0,5	1,0	0,37	2	0,5	1,0	0,37	2,0	0,50	1,00	0,37
mPgCIR 439	1	0,0	0,0	0,0	1	0,0	0,0	0,0	4	0,39	0,12	0,34	2,0	0,13	0,04	0,11
Média	2,6	0,46	0,72	0,37	2,1	0,29	0,54	0,23	2,7	0,47	0,78	0,37	2,4	0,41	0,68	0,32

Dados dos marcadores presentes no heptaplex apresentados em negrito em todas as tabelas.

Tabela 4. Análise descritiva de 10 *loci* microssatélites utilizados em 45 amostras de *P. cattleyanum* coletadas no Espírito Santo. Número de alelos por *loci* (*N*); heteroziguidade esperada (*He*); heteroziguidade observada (*Ho*); Conteúdo de informação polimórfica (PIC).

<i>P. cattleyanum</i> (2n=66)*																				
<i>Locus</i>	<i>Guarapari</i>				<i>Vale</i>				<i>São Floriano/DM</i>				<i>S. Rosa de Lima - SC</i>				<i>Total</i>			
	<i>N</i>	<i>He</i>	<i>Ho</i>	PIC	<i>N</i>	<i>He</i>	<i>Ho</i>	PIC	<i>N</i>	<i>He</i>	<i>Ho</i>	PIC	<i>N</i>	<i>He</i>	<i>Ho</i>	PIC	<i>N</i>	<i>He</i>	<i>Ho</i>	PIC
mPgCIR 02	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
mPgCIR 10	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
mPgCIR 16	1	0,0	0,0	0,0	1	0,0	0,0	0,0	3	0,6	0,8	0,56	5	0,7	0,8	0,74	2,5	0,3	0,4	0,33
mPgCIR 20	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
mPgCIR 29	3	0,6	0,9	0,57	3	0,6	1,0	0,56	3	0,6	0,8	0,56	3	0,5	0,4	0,48	3	0,6	0,8	0,54
mPgCIR 104	4	0,6	1,0	0,63	3	0,7	1,0	0,59	4	0,7	0,9	0,67	4	0,7	1,0	0,68	3,8	0,7	1,0	0,64
mPgCIR 182	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
mPgCIR 187	4	0,7	0,9	0,66	4	0,7	1,0	0,68	4	0,7	1,0	0,7	4	0,7	1,0	0,69	4	0,7	1,0	0,68
mPgCIR 206	4	0,7	0,9	0,7	4	0,8	1,0	0,7	4	0,7	0,9	0,66	4	0,7	0,9	0,7	4	0,7	0,9	0,69
mPgCIR 208	4	0,4	0,3	0,41	1	0,0	0,0	0,0	2	0,2	0,1	0,14	5	0,6	0,4	0,58	3	0,3	0,2	0,28
mPgCIR 216	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
mPgCIR 231	4	0,6	1,0	0,60	2	0,5	1,0	0,38	3	0,6	1,0	0,57	3	0,	0,2	0,39	3	0,4	0,8	0,49
mPgCIR 285	4	0,6	1,0	0,61	3	0,7	1,0	0,59	4	0,7	0,9	0,63	6	0,7	1,0	0,7	4,3	0,7	1,0	0,63
mPgCIR 378	4	0,7	1,0	0,7	4	0,8	1,0	0,7	4	0,7	1,0	0,7	5	0,7	1,0	0,72	4,3	0,7	1,0	0,71
mPgCIR 439	2	0,5	0,9	0,37	2	0,5	1,0	0,38	3	0,6	0,9	0,54	3	0,6	0,9	0,52	2,5	0,6	0,9	0,45
Média	3,4	0,5	0,8	0,53	2,7	0,5	0,8	0,46	3,4	0,6	0,8	0,57	4,2	0,6	0,8	0,62	3,4	0,6	0,8	0,55

* Ploidia referente ao trabalho de Tuler et al., (2019). Espécie apresenta diferentes ploidias.

Tabela 5. Análise descritiva de 15 *loci* microssatélites utilizados em 48 amostras de *Psidium guineense* coletadas no Espírito Santo. Número de alelos por *loci* (*N*); heterozigiosidade esperada (*He*); heterozigiosidade observada (*Ho*); Conteúdo de informação polimórfica (PIC).

<i>P. guineense</i> (2n=44)*																												
<i>Locus</i>	<i>Guarapari</i>				<i>Águia Branca</i>				<i>Itaúnas</i>				<i>São Floriano/DM</i>				<i>Mimoso</i>				<i>VNM</i>				<i>Total</i>			
	<i>N</i>	<i>He</i>	<i>Ho</i>	<i>PIC</i>	<i>N</i>	<i>He</i>	<i>Ho</i>	<i>PIC</i>	<i>N</i>	<i>He</i>	<i>Ho</i>	<i>PIC</i>	<i>N</i>	<i>He</i>	<i>Ho</i>	<i>PIC</i>	<i>N</i>	<i>He</i>	<i>Ho</i>	<i>PIC</i>	<i>N</i>	<i>He</i>	<i>Ho</i>	<i>PIC</i>	<i>N</i>	<i>He</i>	<i>Ho</i>	<i>PIC</i>
mPgCIR 02	4	0,5	0,4	0,48	4	0,7	1,0	0,69	4	0,7	0,9	0,65	4	0,7	0,9	0,7	2	0,5	1,0	0,38	4	0,7	0,6	0,69	3	0,6	0,8	0,6
mPgCIR 10	2	0,4	0,4	0,33	2	0,5	1,0	0,38	2	0,5	0,8	0,38	2	0,5	0,8	0,38	2	0,5	0,8	0,38	2	0,5	0,8	0,38	2	0,5	0,8	0,37
mPgCIR 16	-	---	---	---	-	---	---	---	-	---	---	---	-	---	---	---	-	---	---	---	-	---	---	---	-	---	---	---
mPgCIR 20	-	---	---	---	-	---	---	---	-	---	---	---	-	---	---	---	-	---	---	---	-	---	---	---	-	---	---	---
mPgCIR 29	-	---	---	---	-	---	---	---	-	---	---	---	-	---	---	---	-	---	---	---	-	---	---	---	-	---	---	---
mPgCIR 104	3	0,7	1,0	0,59	5	0,8	0,9	0,76	5	0,8	0,8	0,72	4	0,8	1,0	0,72	5	0,8	1,0	0,74	4	0,7	1,0	0,7	4	0,8	1,0	0,71
mPgCIR 182	-	---	---	---	-	---	---	---	-	---	---	---	-	---	---	---	-	---	---	---	-	---	---	---	-	---	---	---
mPgCIR 187	3	0,6	1,0	0,5	3	0,7	1,0	0,59	3	0,7	1,0	0,59	3	0,7	1,0	0,59	4	0,7	1,0	0,63	4	0,7	1,0	0,69	3	0,7	1,0	0,6
mPgCIR 206	3	0,6	1,0	0,5	3	0,6	1,0	0,50	3	0,6	1,0	0,51	4	0,7	0,8	0,67	4	0,7	1,0	0,68	3	0,5	1,0	0,5	3	0,6	1,0	0,56
mPgCIR 208	3	0,7	0,8	0,59	3	0,7	0,6	0,59	3	0,7	0,5	0,59	3	0,6	0,5	0,56	3	0,7	0,9	0,59	3	0,6	0,9	0,59	3	0,7	0,7	0,59
mPgCIR 216	3	0,7	0,6	0,59	3	0,7	1,0	0,59	3	0,7	0,8	0,59	3	0,7	0,8	0,59	3	0,7	1,0	0,59	3	0,6	1,0	0,56	3	0,7	0,9	0,59
mPgCIR 231	-	---	---	---	-	---	---	---	-	---	---	---	-	---	---	---	-	---	---	---	-	---	---	---	-	---	---	---
mPgCIR 285	3	0,7	0,9	0,58	4	0,7	1,0	0,67	6	0,8	0,8	0,73	4	0,7	0,8	0,67	6	0,8	1,0	0,76	4	0,7	0,9	0,7	4	0,7	0,9	0,69
mPgCIR 378	3	0,7	0,6	0,59	3	0,7	1,0	0,59	3	0,7	1,0	0,59	3	0,7	0,9	0,59	3	0,7	0,8	0,59	3	0,6	0,8	0,59	3	0,7	0,9	0,59
mPgCIR 439	2	0,5	1,0	0,38	2	0,2	0,1	0,18	2	0,5	1,0	0,38	2	0,4	0,3	0,3	2	0,4	0,4	0,33	1	0	0	0	1	0,3	0,5	0,26
Média	2	0,59	0,76	0,51	3	0,62	0,86	0,55	3	0,65	0,84	0,57	3	0,65	0,75	0,58	3	0,64	0,88	0,57	3	0,60	0,7	0,54	3	0,63	0,81	0,56
	,				,				,				,				,				,	9			,			
	9				2				4				2				4				1				2			

* Ploidia referente ao trabalho de Tuler et al., (2019). Espécie apresenta diferentes ploidias.

Tabela 6. Análise descritiva de 15 *loci* microssatélites utilizados em sete amostras de *P. grandifolium* e 14 amostras de *P. myrtooides* coletadas no Espírito Santo. Número de alelos por *loci* (*N*); heterozigosidade esperada (*He*); heterozigosidade observada (*Ho*); Conteúdo de informação polimórfica (PIC).

<i>Locus</i>	<i>P. grandifolium</i>				<i>P. myrtooides</i> (2n=66)*				Total							
	<i>Lícínio Almeida</i>				<i>Guarapari</i>				<i>Vale</i>							
	<i>N</i>	<i>He</i>	<i>Ho</i>	<i>PIC</i>	<i>N</i>	<i>He</i>	<i>Ho</i>	<i>PIC</i>	<i>N</i>	<i>He</i>	<i>Ho</i>	<i>PIC</i>	<i>N</i>	<i>He</i>	<i>Ho</i>	<i>PIC</i>
mPgCIR 02	3	0,62	1,0	0,55	1	0,0	0,0	0,0	1	0,0	0,0	0,0	1	0,0	0,0	0,0
mPgCIR 10	1	0,0	0,0	0,0	---	---	---	---	---	---	---	---	---	---	---	---
mPgCIR 16	3	0,62	1,0	0,54	---	---	---	---	---	---	---	---	---	---	---	---
mPgCIR 20	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
mPgCIR 29	2	0,5	1,0	0,37	2	0,5	1,0	0,37	2	0,5	1,0	0,37	2,0	0,5	1,0	0,37
mPgCIR 104	2	0,24	0,0	0,21	2	0,45	0,71	0,35	2	0,5	1,0	0,37	2,0	0,48	0,86	0,36
mPgCIR 182	3	0,57	0,8	0,50	2	0,5	1,0	0,37	2	0,5	1,0	0,37	2,0	0,5	1,0	0,37
mPgCIR 187	1	0,0	0,0	0,0	2	0,5	1,0	0,37	2	0,5	1,0	0,37	2,0	0,5	1,0	0,37
mPgCIR 206	1	0,0	0,0	0,0	3	0,56	1,0	0,46	2	0,5	1,0	0,37	2,5	0,53	1,0	0,42
mPgCIR 208	2	0,5	1,0	0,37	2	0,5	1,0	0,37	2	0,5	1,0	0,37	2,0	0,5	1,0	0,37
mPgCIR 216	1	0,0	0,0	0,0	2	0,5	1,0	0,37	2	0,5	1,0	0,37	2,0	0,5	1,0	0,37
mPgCIR 231	2	0,5	1,0	0,37	1	0,0	0,0	0,0	1	0,0	0,0	0,0	1,0	0,0	0,0	0,0
mPgCIR 285	2	0,5	1,0	0,37	---	---	---	---	---	---	---	---	---	---	---	---
mPgCIR 378	1	0,0	0,0	0,0	---	---	---	---	---	---	---	---	---	---	---	---
mPgCIR 439	1	0,0	0,0	0,0	---	---	---	---	---	---	---	---	---	---	---	---
Média	1,7	0,28	0,4	0,23	1,8	0,39	0,75	0,30	1,7	0,39	0,78	0,29	1,8	0,39	0,76	0,29

* Ploidia referente ao trabalho de Tuler et al., (2019).

4.3.2 Desenvolvimento do painel multiplex

Dente os 15 marcadores microssatélites selecionados, 12 foram marcados para compor o painel multiplex, estes marcadores apresentam amplificação consistente e para a construção e validação para emprego em genotipagem de goiabeiras. A escolha dos 12 pares de *primers* marcados levou em consideração o tamanho dos fragmentos observados e o fluorocromo associado, pois estes deveriam formar grupos que não tivessem tamanhos com mesma fluorescência, uma vez que poderia interferir durante a identificação do alelo correto. Assim foram organizados em grupos de três marcadores com fragmentos não sobrepostos para serem marcados com o mesmo corante fluorescente: FAM, NED, PET ou VIC (Figura 3). O tamanho e o perfil de amplificação desses tríplex foram observados em gel de poliacrilamida.

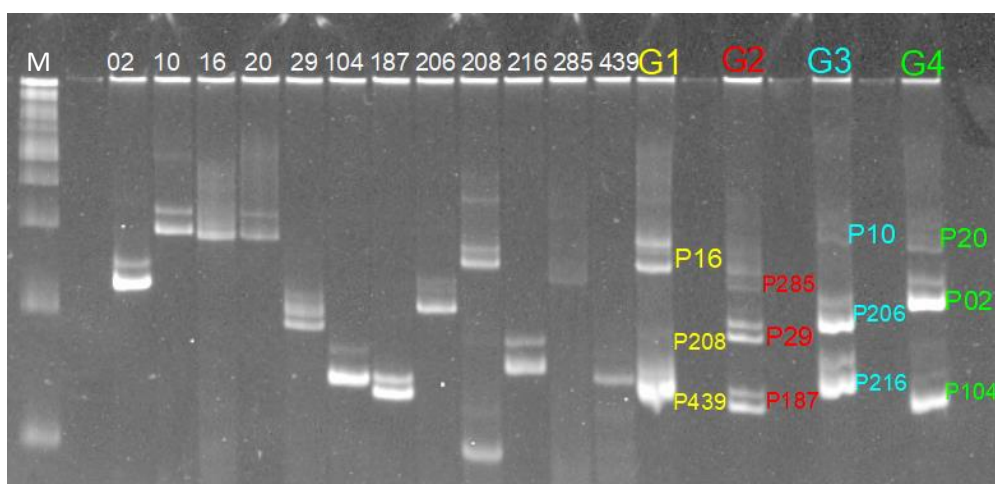


Figura 3. Gel de Poliacrilamida apresentado à amplificação de 12 marcadores selecionados (mPgCIR 02, 10, 16, 20, 29, 104, 187, 206, 208, 216, 285 e 439) para compor o painel multiplex. As amplificações estão representadas individualmente e agrupados em tríplex. (Marcador de 100 pb).

Os testes de construção de painéis multiplex procuraram maximizar a quantidade de marcadores por painel e manter pelo menos um marcador em comum entre diferentes painéis. Foram testadas várias combinações de marcadores por painel. Os testes foram realizados utilizando o kit QIAGEN® Multiplex PCR. Teste com reagentes comuns foram feitos para os tríplex formados inicialmente, porém sem sucesso na amplificação.

Diversas combinações de tríplex e duplex foram realizadas e testadas na eletroforese capilar para a formação do painel, entretanto apenas um grupo com sete marcadores, formado pelos marcadores mPgCIR 187, 439, 104, 216, 29, 206 e 208,

apresentaram consistência adequada para repetição da metodologia em outros genótipos (Figura 4).

O grupo de marcadores formando um heptaplex (Figura 4) apresentou um perfil de eletroferograma satisfatório, pois os picos exibiram absorbância aceitável, próximo a 1000, para a maioria dos *loci* amplificados, sendo que apenas o mPgCIR 208 demonstrou problemas na amplificação do heptaplex e também quando testado sozinho em uma reação isolada, por isso a concentração deste *primer* na reação é muito maior em relação aos outros.

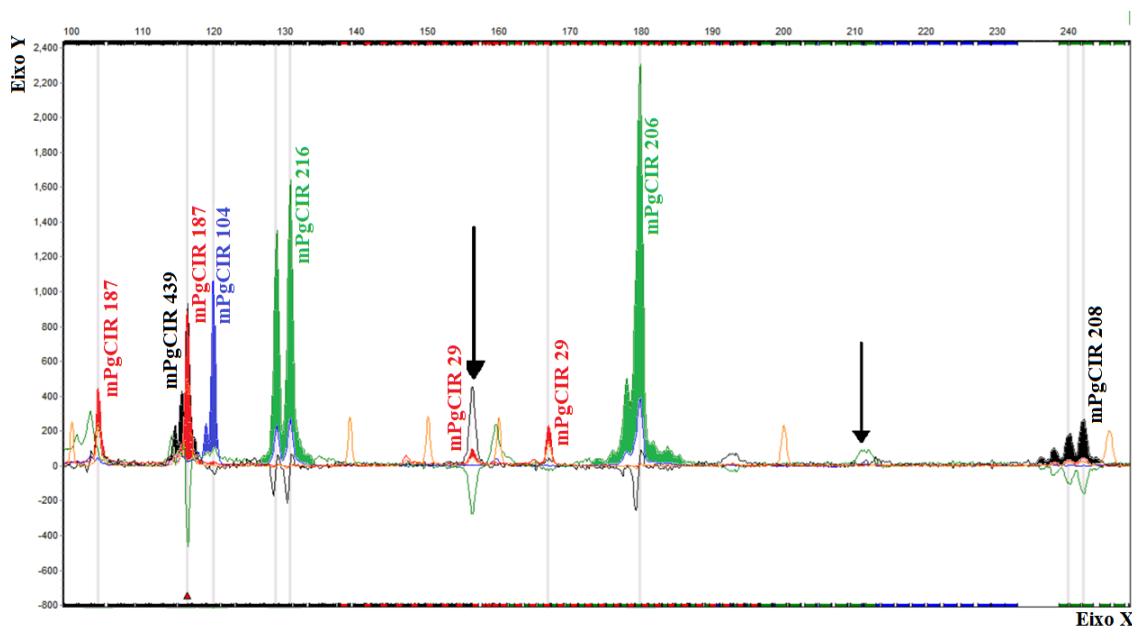


Figura 4. Perfil de amplificação do heptaplex na eletroforese capilar. Picos preenchidos representam marcadores mPgCIR 187 (PET), 439 (NED), 104 (FAM), 216 (VIC), 29 (PET), 206 (VIC) e 208 (NED) respectivamente em amostra de *Psidium guajava*. Setas de cor preta indicam a presença de ampliações/picos inespecíficos.

Após várias tentativas de combinações, o heptaplex foi formado a partir dos marcadores que amplificaram fragmento menores (entre 100 e 240), com exceção do mPgCIR 208, de cerca de 240 pares de base, mesmo assim, como já mencionado, este marcador é o que possui pior qualidade de amplificação, exibindo sempre um baixo valor de absorbância (Figura 5).

O heptaplex formado foi validado através da amplificação em 63 genótipos de *P. guajava*, todos apresentaram padrão de amplificação exemplificado na Figura 5. Além disso, com o intuito de desenvolver uma ferramenta útil para múltiplas espécies este grupo de marcadores foi testado em outras 24 espécies de *Psidium* diferentes (Figura 6).

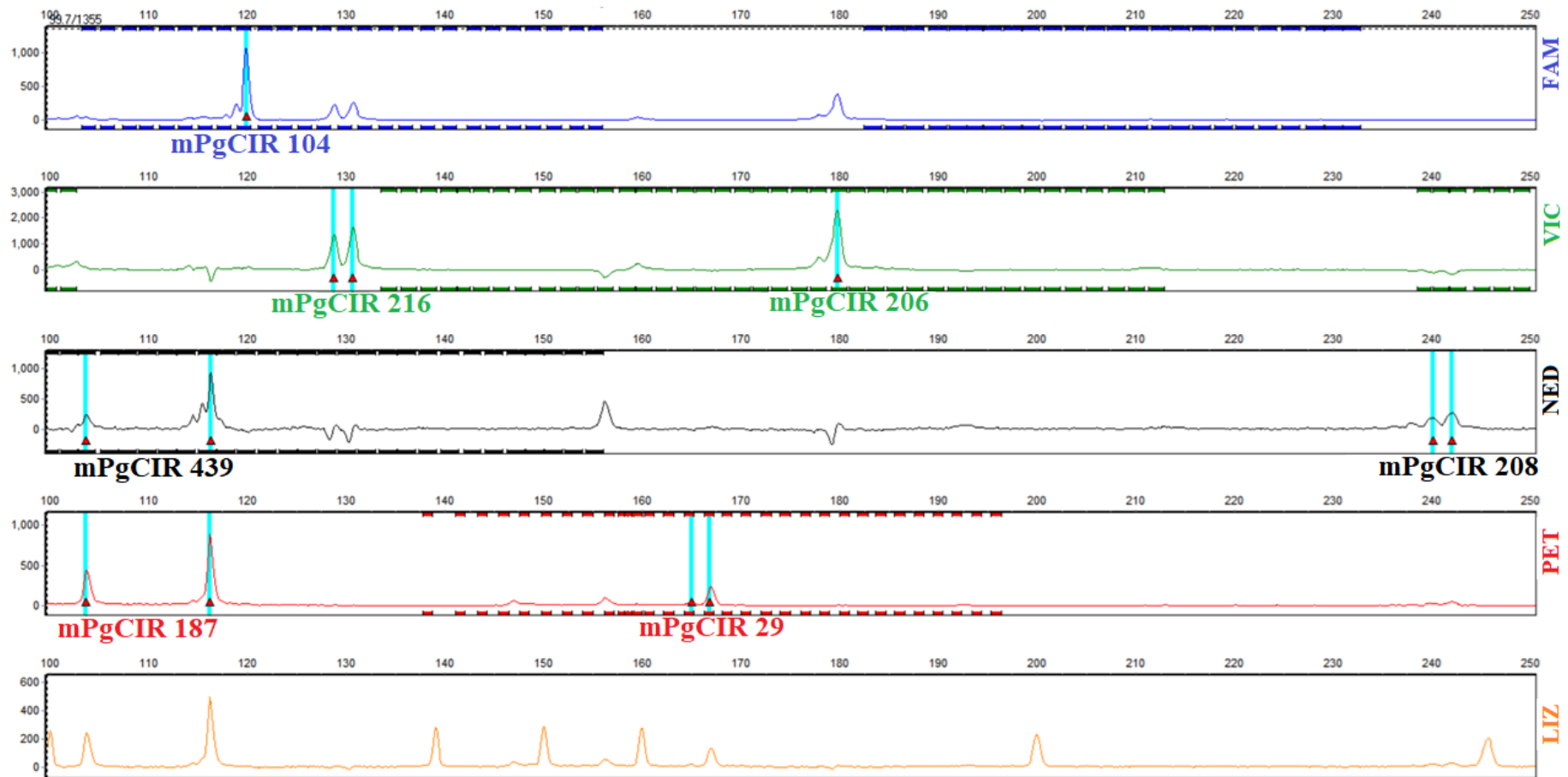


Figura 5. Perfil de amplificação do heptaplex na eletroforese capilar separados por fluorescência em amostra de *P. guajava* Cortibel 21. Cada quadro representa uma fluorescência, sendo elas 6-FAM, VIC, NED e PET respectivamente. Marcadores mPgcIR 187, 439, 104, 216, 29, 206 e 208 respectivamente. Na figura o eixo X (horizontal) indica o tamanho do fragmento, enquanto o eixo Y (vertical) absorvância do fluoróforo.

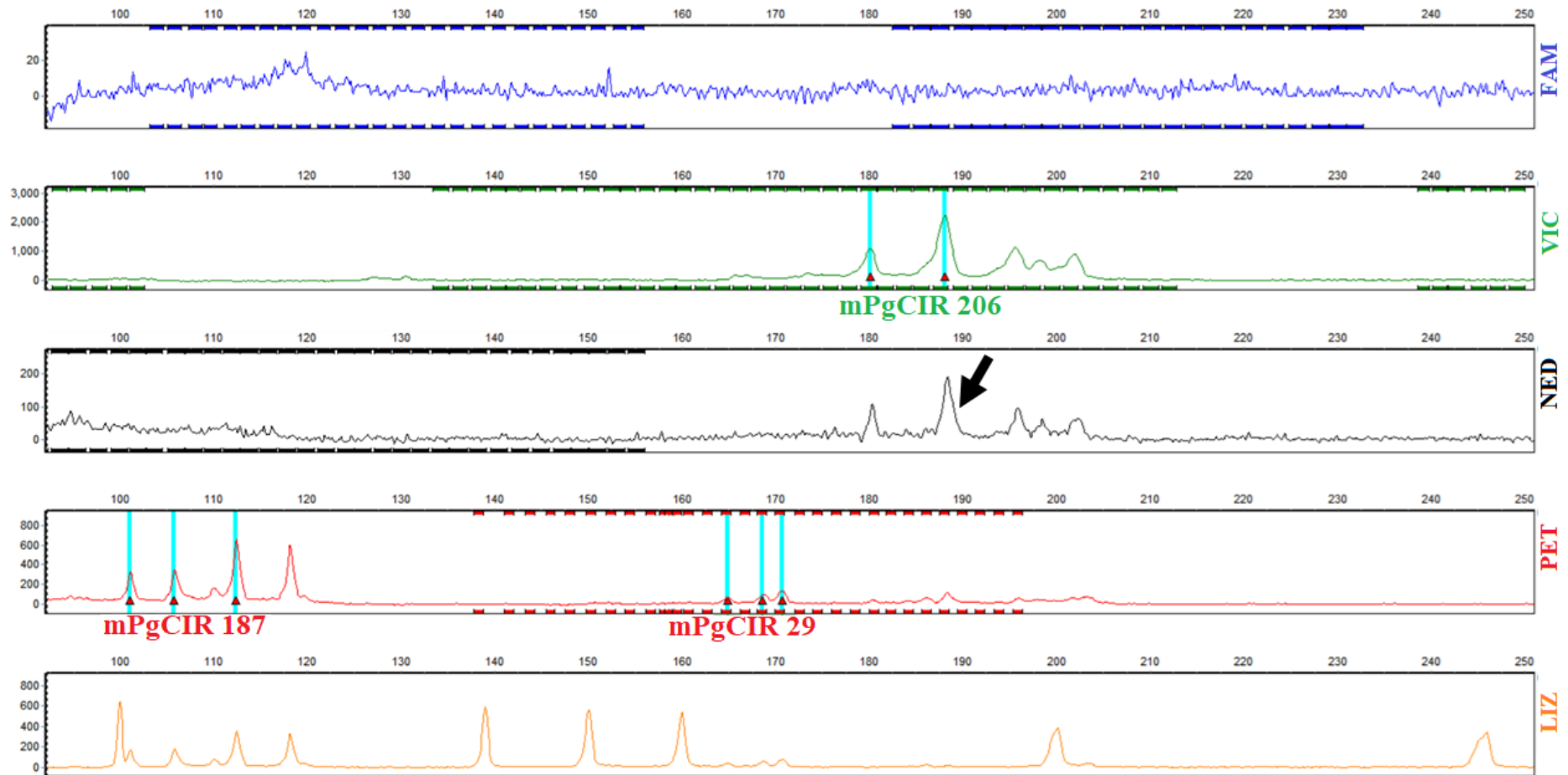


Figura 6. Perfil de amplificação do heptaplex na eletroforese capilar em *P. appendiculatum*. Marcadores representados mPgCIR 206, 187 e 29 com tamanhos de 180/189, 103/105/112 e 165/169/171 respectivamente. A presença de mais de dois picos nos *locus* sugere que essa espécie seja um poliploide. Seta preta indica que a fluorescência NED se espelhou na amplificação produzida pelo marcador com fluorescência VIC. Na figura o eixo X (horizontal) indica o tamanho do fragmento, enquanto o eixo Y (vertical) indica graus de absorvância do fluoróforo.

Para os 63 genótipos de goiabeira, o heptaplex exibiu perfil de eletroferograma satisfatório, com boa absorbância para a maioria dos marcadores amplificados, assim como na amostra padrão, sendo indicado para genotipagem em qualquer amostra de *P. guajava* (Apêndice 2). Os sete *loci* microssatélites analisados apresentaram uma média de 5,1 alelos para *P. guajava* e nove alelos para as 24 espécies, número muito superior aos detectados com genotipagem em gel de poliacrilamida.

O número de alelos variou entre três (mPgCIR 208) a nove alelos (mPgCIR 439). O conteúdo polimórfico superior ao apresentado nos testes, sendo que cinco marcadores tiveram valores médios de PIC próximos de 0,5, apenas o mPgCIR 104 e 208 ficaram distantes desse valor, apresentando valores de 0,30 e 0,38 respectivamente (Tabela 7). Assim, o heptaplex apresentou boa amplificação com níveis de polimorfismos úteis para estudos populacionais na espécie.

Entretanto, na transferência do heptaplex para 24 espécies de *Psidium* é necessária uma otimização que seja compatível com as particularidades de cada genoma, pois é possível observar na tabela de genotipagem (Apêndice 3) a amplificação de diferentes marcadores conforme varia a espécie.

Para as espécies houve a transferência de apenas 52% dos marcadores no heptaplex, sendo que o marcador mPgCIR 208 amplificou somente em *P. grandifolium*. O número médio de alelos foi nove (Tabela 8), número superior a todas as amostras genotipadas em gel e capilar, inclusive a *P. catleyanum*. O maior número de ampliações foi para o marcador mPgCIR 187, que amplificou em 21 espécies. Para os outros resultados presentes na Tabela 8 faz-se necessário uma complementação dos dados, visto que apenas o mPgCIR 187 apresentou pouca falha na amplificação cruzada.

Para as amostras de *Psidium* distintas os picos no eletroferograma apresentaram valores de absorbância da fluorescência baixos. Vários fatores podem ter contribuído para isso, desde mutação na região flanqueadoras das amostras de DNA utilizadas, a presença de muitos compostos fenólicos nessas espécies que inibem a reação, além do perfil de amplificação conjunta que pode variar conforme o genoma.

A partir dos dados de diversidade dos marcadores no painel foi gerado um agrupamento com a distribuição dos genótipos de *P. guajava*. Com base no ponto de corte indicado pelo programa foi observado a formação de seis grupos. O primeiro, segundo e terceiro grupo foram representados apenas por um genótipo, sendo eles as amostras de Pernambuco, Pedro Sato e Alagoas, respectivamente. O quarto

agrupamento foi formado por duas amostras silvestres e duas amostras de Cortibel. O quinto agrupamento foi representado por Cortibel 34 e o sexto incluiu todas as demais espécies em um grande grupo. Apenas os genótipos Cortibel 40 e Cortibel 42 não foram discriminados (Figura 7).

Para a maior parte dos materiais nomeados Cortibel foram observados proximidade genética. Esses materiais assemelharam-se tanto aos cultivares comerciais quanto a genótipos espontâneos. Não foi possível acessar diferenças entre os materiais Cortibel 40 e Cortibel 42. Para a cultivar Pedro Sato uma maior dissimilaridade em relação aos demais genótipos foi notada. Situação semelhante para as goiabeiras oriunda dos Estados de Pernambuco e Alagoas e para Cortibel 34. Já os materiais Cortibel 7 e 48 juntamente com as goiabeiras da Paraíba e Bahia alocaram-se em um mesmo grupo (Figura 7).

Tabela 7. Análise descritiva dos *loci* de microsatélites avaliados no painel multiplex. Amostras avaliadas consistem em 48 genótipos de *Psidium guajava*: 16 cultivares, 30 Cortibel, nove genótipos de ocorrência natural e um híbrido de *P. guajava* x *P. guineense*.

Primer	Cultivares (2n=22)				Cortibel (2n=22)				Naturais (2n=22)			
	N	He	Ho	PIC	N	He	Ho	PIC	N	He	Ho	PIC
mPgCIR 29	3	0,61	0,80	0,54	4	0,62	0,90	0,55	5	0,71	0,80	0,66
mPgCIR 104	2	0,18	0,07	0,16	4	0,30	0,27	0,28	3	0,56	0,57	0,47
mPgCIR 187	3	0,53	0,94	0,42	4	0,55	0,83	0,46	2	0,44	0,67	0,35
mPgCIR 206	3	0,65	0,19	0,57	4	0,24	0,13	0,23	3	0,59	0,10	0,51
mPgCIR 208	3	0,37	0,45	0,33	3	0,54	0,93	0,44	2	0,50	1,00	0,38
mPgCIR 216	4	0,60	0,44	0,54	4	0,64	0,57	0,58	4	0,62	0,22	0,57
mPgCIR 439	7	0,71	0,88	0,68	6	0,72	0,85	0,68	2	0,28	0,00	0,24
Média	3,57	0,52	0,54	0,46	4,14	0,52	0,64	0,46	3,00	0,53	0,48	0,45

N= Número de alelos.

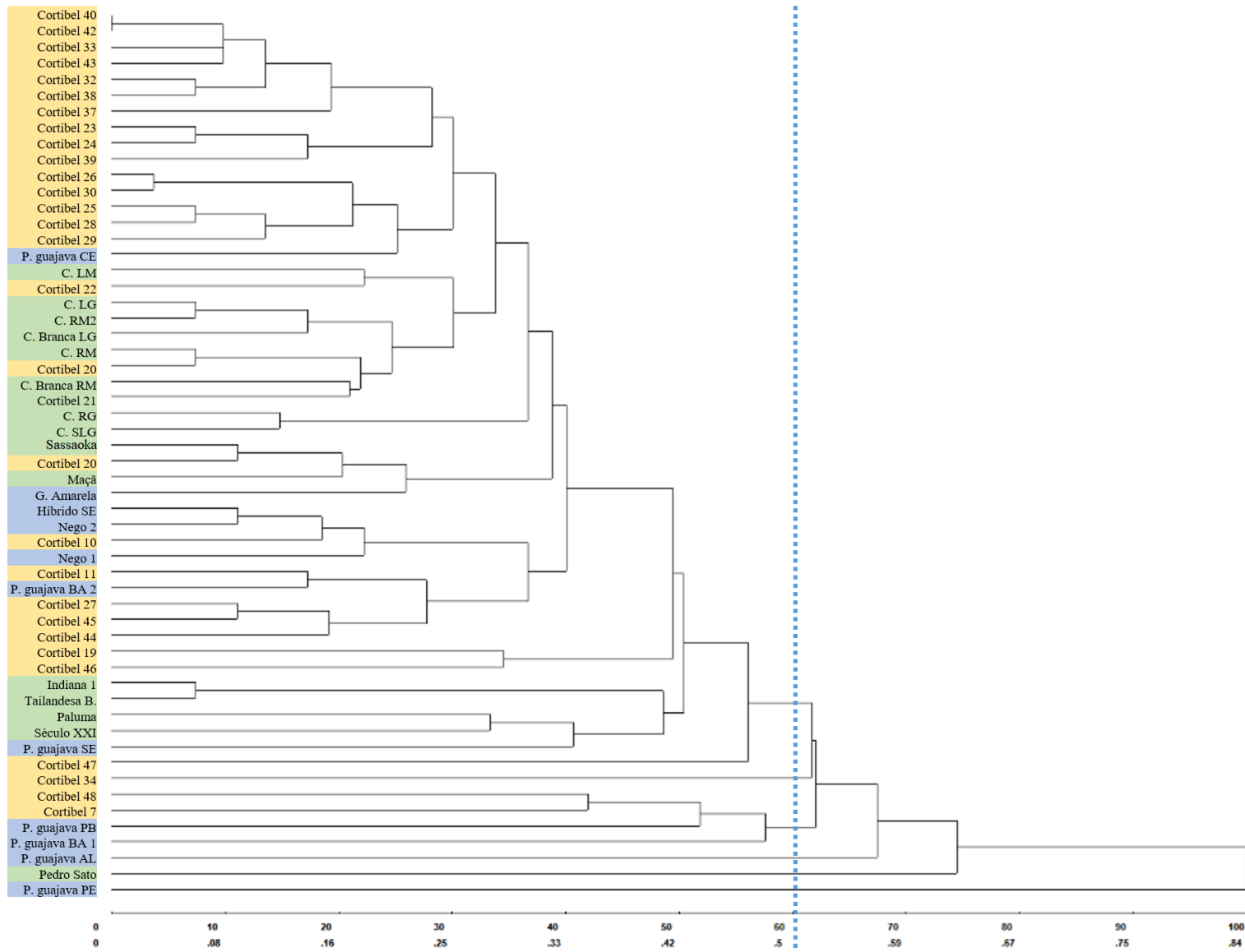


Figura 7. Agrupamento obtido através da ligação média entre grupos (UPGMA) para 56 genótipos de goiabeiras, através de marcadores microssatélites genotipados através de eletroforese capilar. Genótipos identificados em três grupos: Cultivares (verde), Cortibel (amarela) e Silvestres (azul).

Tabela 8. Análise descritiva dos *loci* de microssatélites avaliados no painel multiplex. Amostras avaliadas consistem em 24 espécies diferentes de *Psidium* (*P. guajava* e *P. sartorianum* não estão incluídas).

24 Espécies de <i>Psidium</i>		
Primer	NA	NE
mPgCIR 29	17	19
mPgCIR 104	6	08
mPgCIR 187	12	21
mPgCIR 206	12	13
mPgCIR 208	2	1
mPgCIR 216	8	12
mPgCIR 439	3	7
Média	9	11,5

NA = Número de alelos por *locus*; NE = Número de espécies onde houve amplificação;

4.4 Discussão

Diferentes estudos têm tido sucesso ao realizar transferibilidade de marcadores desenvolvidos para *P. guajava* em outras espécies e até em outros gêneros (NOGUEIRA *et al.*, 2015; TULER *et al.*, 2015; BERNARDES *et al.*, 2018). Evidenciando uma relação de proximidade genética entre a espécie, uma vez que mutações nas regiões flanqueadoras impediriam o anelamento do *primer*, sugerindo que estas sequências estão em regiões importantes do genoma sendo conservadas e também que estas espécies divergiram recentemente. Além disso, segundo Tuler *et al.* (2015) esta alta taxa de transferibilidade pode estar relacionado à diversificação do gênero em um passado recente.

As regiões com microssatélites apresentam uma taxa de mutação mais elevada do que outras regiões do genoma. Por isso o seu elevado grau de multialelismo e a elevada geração de alelos exclusivos ou com baixa frequência em uma determinada população (SCHMIDT, 2009). Entretanto, os 15 marcadores selecionados formaram um grupo com nove espécies morfologicamente diferentes entre si, porém que amplificaram todos os *loci*, demonstrando a origem conservada de um grupo de marcadores, que podem ser tidos como referências para estudos inter e intraespecíficos de diferentes propósitos.

Após a seleção e caracterização houve a construção de um sistema multiplex para a caracterização de espécies de *Psidium* informativo, polimórfico e com elevado poder discriminatório dos indivíduos de uma população. Observou-se também que ampliações de fragmentos de tamanhos menores e similares pelas diferentes combinações de *primers*, foram de maior sucesso, pois os marcadores SSR devem apresentar alelos que possuam intervalos de tamanho que permitem a combinação entre si, sem prejudicar ou confundir interpretações de resultados (BUTLER, 2005).

A abordagem multiplex tem sido cada vez mais utilizada por autores para a caracterização de *loci* microssatélites em diversas espécies (POSTOLACHE *et al.*, 2013; DRAŠNAROVÁ *et al.*, 2014; ZAROURI *et al.*, 2015), dado a importância deste marcador. Em espécies anuais e cultivadas esta utilização é mais comum, mas para a espécie de interesse, como exemplo os trabalhos de Morais *et al.* (2016), no qual foi desenvolvido um painel multiplex para 24 *loci* microssatélites divididos em quatro grupos, em feijão, e o trabalho de Ashkani *et al.* (2015), no qual desenvolveram, a partir

de 63 *loci* microssatélites, 13 sistemas multiplex para identificação de genes de resistência em arroz.

Para validação do sistema multiplex heptaplex, os resultados obtidos na genotipagem de 63 genótipos de *P. guajava* (Tabela 7), demonstraram maior eficiência quando comparados ao PCR comum, repetibilidade das amplificações, bem como aumento do número de alelos observados com a eletroforese capilar, ao se comparar os mesmos *loci* avaliados em poliacrilamida, com conseqüente aumento dos valores de PIC, H_e e H_o , nesta análise. Estes resultados esperados, pela conhecida eficiência de discriminação de alelos do sistema de eletroforese capilar, demonstraram que o sistema heptaplex otimizado apresenta poder de resolução maior que o observado nos 15 *loci* marcadores no gel de poliacrilamida. Esse resultado pode ser notado ao se comparar os resultados das tabelas três e oito, visto que os genótipos de cultivares e de genótipos Cortibel foram analisadas das duas formas.

Para todos os parâmetros da análise descritiva foram encontrados valores superiores ao compararmos os dois resultados, com exceção da H_o para as cultivares. Este efeito pode ser conseqüência da formação de bandas inespecíficas que são erroneamente registradas no gel, sendo que ao visualizar o eletroferograma a chances de identificação incorretamente diminui devido ao esperado para o padrão de pico formado pelo alelo.

Como já mencionado, o PIC também apresentou maiores resultados, segundo Ellegren (2004) o polimorfismo de determinado *locus* pode estar positivamente correlacionado à natureza gênica ou genômica, com os tipos de motivos de repetição destas regiões (mono, di, tri, tetra, penta ou hexa), e o número de repetições dos motivos e interrupções nas repetições.

Neste estudo o loco mPgCIR 104 apresentou um baixo PIC, este marcador apresenta um motivo tetranucleotídeo. Além de menor quantidade no genoma, os microssatélites formados por tetra ou hexanucleotídeos apresentam menor polimorfismo, em contrapartida para os *loci* mono e dinucleotídeos, espera-se maior polimorfismo, pois estes tipos de motivos, por serem menores, estão mais sujeitos ao *slippage* e com isso a mutações mais frequentes (ELLEGREN, 2004).

Trabalhos descrevendo os PICs obtidos a partir de marcadores desenvolvidos para *P. guajava* (CHAITHANYA *et al.*, 2015; MEHMOOD *et al.*, 2015; KANUPRIYA *et al.*, 2011) apresentam valores médios de PIC superiores, 0,74 para 13 *loci*, 0,61 para 18 *loci* e 0,74 para 23 *loci*, entretanto, esses valores são para grupos de marcadores

aleatórios, que não possuem seleção prévia, como os deste estudo. Assim, é possível notar que os valores médios de PIC não estão destoando fortemente dos valores médios apresentados na tabela 7, visto que os locus estudados aqui foram selecionado e estão presentes em regiões conservadas do genoma, sendo, portanto, considerados informativos e úteis para estudos intra e interespecíficos.

Sequências de DNA das regiões não traduzidas ou de íntrons, normalmente contém mais repetições de di e tetranucleotídeos que regiões traduzidas, onde os trinucleotídeos encontram-se em maior quantidade (CARDLE *et al.*, 2000). Para *P. guajva*, a maioria das sequências de *loci* microssatélites conhecidas são de dinucleotídeos, como pode se verificar no trabalho de Tuler *et al.* (2015) mostrando que a maioria dos marcadores são aleatórios no genoma.

A variabilidade dos materiais utilizados nesse estudo foi observada pela análise de marcadores microssatélites, na qual grande parte dos genótipos Cortibel apresentou proximidade genética. Esses dados evidenciam o potencial de genótipos Cortibel, que apresentaram características semelhantes a genótipos amplamente cultivados, sendo importante ao melhoramento por aumentar a variabilidade de genótipos comerciais disponíveis.

Século XXI e Paluma ficaram bem próximas, visto que Século XXI foi obtido através do cruzamento Supreme 2 x Paluma (POMMER; OLIVEIRA; SANTOS, 2013), é justificável essa proximidade, embora Coser *et al.* (2012) em estudo com os esses materiais verificou que os mesmos alocaram-se em grupos divergentes, possivelmente pelo fato de terem sido utilizados outros marcadores, e pelo número de materiais avaliados, que foi menor comparado ao presente estudo. Um ponto que se assemelha ao estudo de Coser *et al.* (2012) é que o material Pedro Sato se mostrou divergente em relação a todos os outros materiais, evidenciando que esse comportamento se manteve em diferentes estudos e também com SNPs.

Com o bom desempenho do heptaplex nas amostras de goiabeira, o painel foi testado em 24 espécies de *Psidium*, porém os resultados indicam a necessidade de maior padronização para as diferentes espécies avaliadas. Segundo Chapuis e Estoup (2007) quando se realiza amplificação cruzada é muito provável encontrar na espécie alvo, altas incidências de alelos nulos ou frequências genotípicas diferentes das esperada, portanto estes resultados podem ser devido ao comportamento diferente dos *primers* quando amplificados juntos em cada genoma.

Entretanto, mesmo que o heptaplex não tenha funcionado devidamente para as espécies de *Psidium*, esse conjunto de *loci* se mostrou eficiente para genotipagem em amostra de goiabeira comerciais ou de ocorrência natural, permitindo a detecção de variabilidade genética e discriminação individual por apresentarem polimorfismo considerável. Para a amplificação em outras espécies, faz-se necessários novos estudos da história evolutiva de cada espécie, presença de hibridação e testes com condições diferentes dos marcadores visto que o trabalho da dissertação de Miranda (2014) mostrou ser possível a formação de um painel que contemple mais de uma espécie.

4.5 Conclusão

1. Houve sucesso na transferibilidade de *primers* de regiões microssatélites para diferentes espécies de *Psidium*;
2. Foi possível estabelecer um sistema multiplex a partir de sete marcadores microssatélites polimórficos e conservados entre espécies para genotipagem de *P. guajava*;
3. As amostras de goiabeira genotipadas na eletroforese capilar por meio de marcadores microssatélites, apresentaram elevada diversidade genética;
4. O heptaplex deve ser otimizado para a amplificação em outras espécies do gênero, visto que, o comportamento dos marcadores muda conforme as especificidades de cada genoma.
5. Os marcadores microssatélites aqui estudados conservados em múltiplas espécies de *Psidium* apresentam índice de diversidade por *locus* similar à microssatélites aleatórios no genoma, servindo para estudos tanto intra como interespecíficos.

4.6 Referências

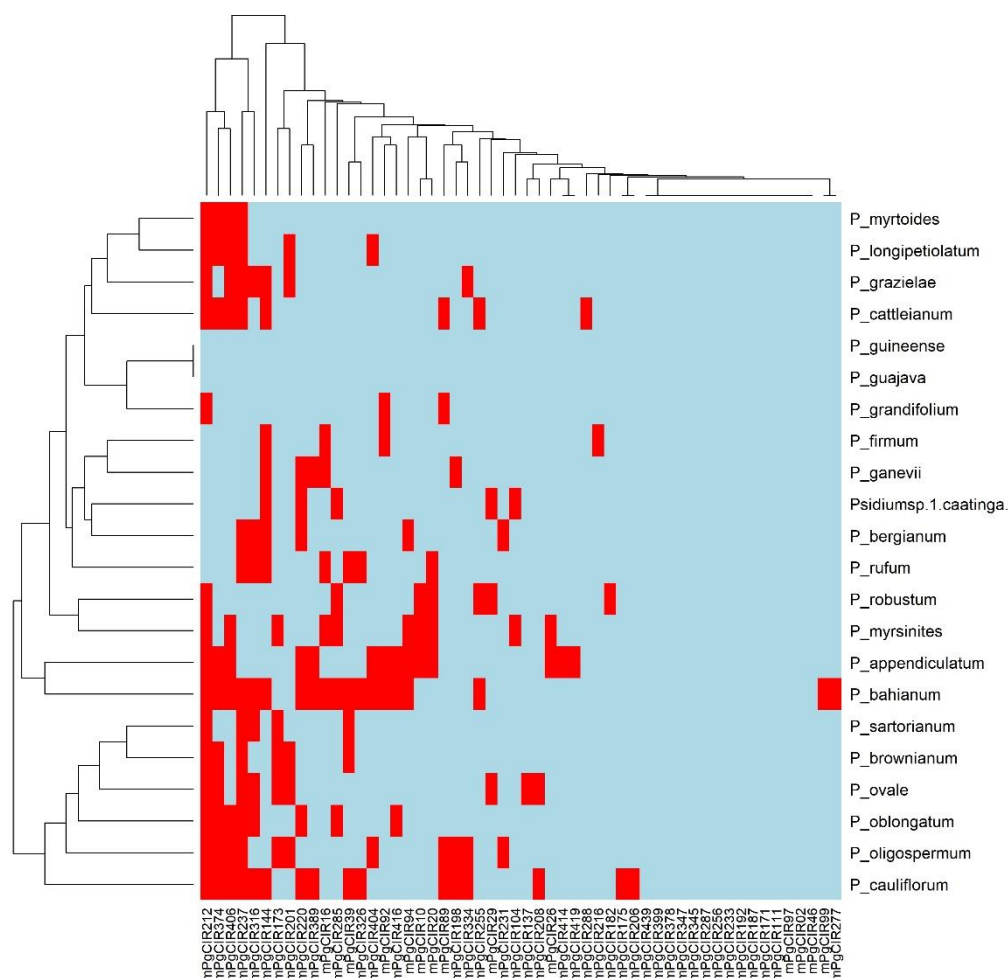
- Anderson, N. O., Jocienė, L., Krokaitė, E., Rekašius, T., Paulauskas A., Kupčinskienė, E. (2018). Genetic diversity of phalaris arundinacea populations in relation to river regulation in the Merkys basin, Lithuania. *River Research and Applications*. 1–10 p. DOI: 10.1002/rra.3259
- Ashkani, S., Rafii, M. Y., Shabanmofrad, M., Foroughi, M., Azizia, P., Akhtar, M. S., Sahebi, M., Harun, A. R., Nasehi, A. (2015). Multiplex SSR–PCR approaches for semi-automated genotyping and characterization of loci linked to blast disease resistance genes in rice, C. R. *Biologies*. <http://dx.doi.org/10.1016/j.crv.2015.07.007>
- Bernardes, C. de O., Tuler, A. C., Ferreira, A., Carvalho, M. S., Nogueira, A. M., & Ferreira, M. F. da S. (2018). Transferability of Psidium microsatellite loci in Myrteae (Myrtaceae): a phylogenetic signal. *Euphytica*. <https://doi.org/10.1007/s10681-018-2195-2>
- Beugin, M. P., Letty, J., Kaerle, C., Guitton, J. S., Muselet, L., Queney, G., Pontier, D. (2017). A single multiplex of twelve microsatellite markers for the simultaneous study of the brown hare (*Lepus europaeus*) and the mountain hare (*Lepus timidus*). *Ecology and Evolution*. 7:3931–3939. DOI: 10.1002/ece3.2943
- Butler, J. M. (2005) Forensic DNA typing protocols. *Methods Mol Biol* 297:53–65. doi:10.1385/1592598676
- Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics* 156:847–854
- Chaithanya, M. V. N., Sailaja, D., Dinesh, M. R., Vasugi, C., Reddy, D. C. L., Aswath, C. (2015). Microsatellite-Based DNA Fingerprinting of Guava (*Psidium guajava*) Genotypes. *Proc. Natl. Acad. Sci., India, Sect. B Biol. Sci.* DOI 10.1007/s40011-015-0660-4
- Chapuis, M. P. & Estoup, A. (2007). Microsatellite Null Alleles and Estimation of Population. *Molecular Biology and Evolution* 24 (3): 621-631, 2007.
- Coser, S. M.; Ferreira, M. F. S.; Ferreira, A.; Mitre, L. K.; Carvalho, C. R.; Clarindo, W. R. (2012). Assessment of genetic diversity in *Psidium guajava* L. using different approaches. *Scientia Horticulturae*, v. 148, p. 223-229.
- Guavamap (2008). Improvement of Guava: Linkage mapping and QTL analysis as a basis for marker-assisted selection. Disponível em: <<http://www.neiker.net/neiker/guavamap/>> Acesso em 04 Abr. 2013.
- Cruz, C. D. (2013). GENES - a software package for analysis in experimental statistics and quantitative genetics. *Acta Scientiarum*, 271–276. <https://doi.org/10.4025/actasciagron.v35i3.21251>
- Diéguez, M. J., Petignat, C., Ferella, L., Fiorentino, G., Silva, M., Dabove, M. A., Yañez, G. I. R., López, M., Pergolesi, M. F., Ingala, L., Cuyeu, A. R., Sacco, F. (2018). Mapping a gene on wheat chromosome 4BL involved in a complementary interaction with adult plant leaf rust resistance gene LrSV2. *Theoretical and Applied Genetics*. <https://doi.org/10.1007/s00122-018-3155-x>
- Dijk, K.-J. V., Mellors, J., Waycott, M. (2014). Development of multiplex microsatellite PCR panels for the seagrass *thalassia hemprichii* (Hydrocharitaceae). Applications in *Plant Sciences*. 2 (11): 1400078. doi:10.3732/apps.1400078
- Doyle, J.J. & Doyle, J.L. (1990). Isolation of plant DNA from fresh tissue. *Focus* 1213-1215. Erdtman, G. 1952. Pollen Morphology and Plant Taxonomy – Angiosperms.

- Stockolm, Almqvist & Wiksel.
- Drašnarová, A., Krak, K., Vít, P., Doudová, J., Douda, J., Hadincová, V., ... Mandák, B. (2014). Cross-amplification and multiplexing of SSR markers for *Alnus glutinosa* and *A. incana*. *Tree Genetics and Genomes*, 10(4), 865–873. <https://doi.org/10.1007/s11295-014-0727-z>
- Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, 5(6), 435-445.
- Fagundes, B. S., da Silva, L. F., Giacomini, R. M., Secco, D., Díaz-Cruz, J. A., & Da-Silva, P. R. (2016). Transferability of Microsatellite Markers Among Myrtaceae Species and Their use to Obtain Population Genetics Data to Help the Conservation of the Brazilian Atlantic Forest. *Tropical Conservation Science*, 408–422. <https://doi.org/10.1177/194008291600900122>
- Honig, J. A., Averello, V., Kubik, C., Vaiciunas, J., Bushman, B. S., Bonos, S. A., Meyer, W. A. (2018). An Update on the Classification of Kentucky Bluegrass Cultivars and Accessions Based on Microsatellite (SSR) Markers. *Crop Science*, vol. 58:1776–1787. Doi: 10.2135/cropsci2017.11.0689.
- Kalia, R. K., Rai, M. K., Kalia, S., Singh, R., Dhawan, A. K. (2011). Microsatellite markers: an overview of the recent progress in plants. *Euphytica*. 177: 09–334.
- Kanupriya, Latha P. M., Aswath, C., Reddy, L., Padmakar, B., Vasugi, C., Dinesh, M. R. (2011). Cultivar Identification and Genetic Fingerprinting of Guava (*Psidium guajava*) Using Microsatellite Markers, *International Journal of Fruit Science*, 11:2, 184-196, DOI:10.1080/15538362.2011.578521
- Kherwar, D., Usha, K., Mithra, S. V. A., Singh, B. (2018). Microsatellite (SSR) marker assisted assessment of population structure and genetic diversity for morpho-physiological traits in guava (*Psidium guajava* L.). *Journal of Plant Biochemistry and Biotechnology*. Vol. 27, Issue 3, pp 284–292. <https://doi.org/10.1007/s13562-017-0438-2>
- Kohlross, M., Reiche, M., Jesinghaus, M., Grosser, B., Slotta-Huspenina, J., Hapfelmeier, A., Bauer, L., Novotny, A., Weichert, W., Keller, G. (2018). A microsatellite based multiplex PCR method for the detection of chromosomal instability in gastric cancer. *Scientific Reports*, 8:12551, DOI:10.1038/s41598-018-30971-z
- Landrum, L.R. (2003). A revision of the *Psidium salutare* complex. *Sida* 20: 1449-1469.
- Li, D., Wang, S., Shen, Y., Meng, X., Xu, X., Wang, R., & Li, J. (2018). A multiplex microsatellite PCR method for evaluating genetic diversity in grass carp (*Ctenopharyngodon idellus*). *Aquaculture and Fisheries*, 3(6), 238–245. doi:10.1016/j.aaf.2018.09.001
- Miranda, E. A. G. C. (2014). *Transferibilidade e validação de marcadores microssatélites derivados de EST para duas espécies de Campomanesia (Myrtaceae) do Cerrado*. Dissertação (mestrado) – Universidade Federal de Goiás / Instituto de Ciências Biológicas – Goiânia, GO.
- Morais, S. R. P. de, Vieira, A. F., Almeida, L. C. da S., Rodrigues, L. A., Melo, P. G. S., Faria, L. C. de, ... Souza, T. L. P. O. de. (2016). Application of microsatellite markers to confirm controlled crosses and assess genetic identity in common bean. *Crop Breeding and Applied Biotechnology*, 16(3), 234–239. <https://doi.org/10.1590/1984-70332016v16n3n35>
- Munyard, K. A., Ledger, J. M., Lee, C. Y., Babra, C., & Groth, D. M. (2009). Characterization and multiplex genotyping of alpaca tetranucleotide microsatellite markers. *Small Ruminant Research*, 85(2-3), 153–156. doi:10.1016/j.smallrumres.2009.07.012

- Nogueira, A. M., Ferreira, A., & Ferreira, M. F. da S. (2015). Transferability of Microsatellites from *Psidium guajava* to *Eugenia*, *Myrciaria*, *Campomanesia*, and *Syzygium* Species (Myrtaceae). *Plant Molecular Biology Reports*. <https://doi.org/10.1007/s11105-015-0921-7>
- Pommer, C. V.; Oliveira, O. F.; Santos, C. A. F. (2013). Goiaba: recursos genéticos e melhoramento. Edufersa: Mossoró - RN. 126 p
- Postolache, D., Leonarduzzi, C., Piotti, A., Spanu, I., Roig, A., Fady, B., ... Vendramin, G. G. (2013). Transcriptome versus Genomic Microsatellite Markers: Highly Informative Multiplexes for Genotyping *Abies alba* Mill. and Congeneric Species. *Plant Molecular Biology Reporter*, 32(3), 750–760. <https://doi.org/10.1007/s11105-013-0688-7>
- Psidium* in Flora do Brasil 2020 em construção. Jardim Botânico do Rio de Janeiro. Disponível em: <<http://floradobrasil.jbrj.gov.br/reflora/floradobrasil/FB605023>>. Acesso em: 17 Ago, 2018.
- R Core Team (2018) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available on: <http://www.R-project.org>
- Risterucci, A. M.; Duval, M. F.; Rohde, W.; Billotte, N. (2005). Isolation and characterization of microsatellite loci from *Psidium guajava* L. *Molecular Ecology Notes*.5: 745–748.
- Schlautman, B., Bolivar-Medina, J., Hodapp, S., & Zalapa, J. (2017). Cranberry SSR multiplexing panels for DNA horticultural fingerprinting and genetic studies. *Scientia Horticulturae*, 219, 280–286. <https://doi.org/10.1016/j.scienta.2017.03.005>
- Schmidt, A. B. (2009). *Desenvolvimento de painéis multiplex de marcadores microssatélites e mapeamento de QTLs de tolerância à seca e ao frio em linhagens puras recombinantes de arroz (Oryza sativa L.)*. Tese (doutorado) – Universidade Federal de Santa Catarina / Centro de Ciências Agrárias – Florianópolis, SC.
- Sobral, M.; Proença, C.; Souza, M.; Mazine, F.; Lucas, E. (2014). 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: 16 Ago. 2018.
- Stolle, E., Kidner, J. H., & Moritz, R. F. A. (2013). Patterns of evolutionary conservation of microsatellites (SSRs) suggest a faster rate of genome evolution in hymenoptera than in Diptera. *Genome Biology and Evolution*, 5(1), 151–162. <https://doi.org/10.1093/gbe/evs133>
- Silva, S. N. (2017). *Diversidade de genótipos de goiabeiras por caracteres vegetativos, reprodutivos e viabilidade polínica*. Tese (Doutorado) – Universidade Federal do Espírito Santo/ Centro de Ciências Agrárias – Alegre, ES.
- Tuler, A. C., Carrijo, T. T., Nóia, L. R., Ferreira, A., Peixoto, A. L., & Ferreira, M. F. da S. (2015). SHORT COMMUNICATION SSR markers: a tool for species identification in *Psidium* (Myrtaceae). *Molecular Biology Reports*. <https://doi.org/10.1007/s11033-015-3927-1>
- Tuler, A. C.; Carrijo, T. T.; Peixoto, A. L.; Garbin, M. L.; Ferreira, M. F. da S.; Carvalho, C. R.; Spadeto, M. S.; Clarindo, W. R. (2019). Diversification and geographical distribution of *Psidium* (Myrtaceae) species with distinct ploidy levels. *Trees*. <https://doi.org/10.1007/s00468-019-01845-2>
- Vallone, P. M., Butler, J. M. (2004). AutoDimer: a screening tool for primer-dimer and hairpin structures. *Biotechniques*. 37(2):226–231.

Zarouri, B., Vargas, A. M., Gaforio, L., Aller, M., de Andrés, M. T., & Cabezas, J. A. (2015). Whole-genome genotyping of grape using a panel of microsatellite multiplex PCRs. *Tree Genetics and Genomes*, *11*(2). <https://doi.org/10.1007/s11295-015-0843-4>

4.7 Apêndices



Apêndice 1. Heatmap para amplificação dos 54 marcadores SSRs selecionados, em 22 espécies de *Psidium*. As bandas azuis representam a presença de amplificação e as vermelhas não amplificação. Distancia coincidência simples.

Apêndice 2. Genotipagem das amostras de *P. guajava*.

IND.	P 29	P 104	P 187	P 206	P 208	P 216	P 439							
Paluma	168	168	120	120	116	118	180	180	240	240	130	132	114	116
Pedro Sato	-	-	-	-	116	116	180	180	240	240	132	132	98	108
Século XXI	168	168	120	120	104	116	180	180	240	242	128	130	110	116
Indiana	156	166	118	120	104	116	180	180	238	240	116	130	110	124
Tailandesa	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Tailandesa B.	156	166	118	118	104	116	180	180	238	240	116	130	110	124
Chinesa	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C. LG 1	166	168	120	120	104	116	184	184	-	-	116	130	104	116
C. LM 1	168	168	120	120	104	116	186	186	240	240	132	132	104	116
C. Branca LG	166	168	120	120	104	116	184	184	-	-	132	132	104	116

C. RM	166	168	120	120	104	116	184	186	240	240	116	132	104	116
C. Branca RM	166	168	120	120	104	116	186	186	240	240	130	130	104	116
C. RM2	166	168	120	120	104	116	184	184	-	-	116	130	104	116
C. RG	166	168	120	120	104	116	184	186	-	-	130	130	116	116
C. SLG	166	168	120	120	104	116	184	184	-	-	130	130	116	116
Kuse	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Maçã	156	166	120	120	104	116	180	180	238	240	130	130	108	116
Petri	156	168	120	120	104	116	184	184	240	240	130	130	108	116
G. Roxa	-	-	-	-	-	-	-	-	-	-	-	-	-	-
G. amarela	156	168	120	120	104	116	184	186	238	240	130	130	104	116
Cortibel 18	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cortibel 19	156	166	114	118	100	114	184	184	240	240	114	130	114	114
Cortibel 20	166	168	120	120	104	116	184	186	-	-	116	132	104	116
Cortibel 21	166	168	120	120	104	116	180	180	240	242	130	132	104	116
Cortibel 22	168	170	120	120	104	116	184	184	238	240	132	132	104	116
Cortibel 23	166	168	120	120	104	116	182	184	238	240	132	132	110	116
Cortibel 24	166	168	120	120	104	116	184	186	238	240	132	132	110	116
Cortibel 25	166	168	120	120	104	116	184	184	238	240	116	132	104	116
Cortibel 26	166	168	116	120	104	116	184	184	238	240	116	132	110	116
Cortibel 27	156	166	120	120	104	116	184	184	238	240	116	132	110	116
Cortibel 28	166	168	116	120	104	116	184	184	238	240	116	132	104	116
Cortibel 29	166	168	116	120	104	116	184	184	238	240	130	130	104	116
Cortibel 30	166	168	116	120	104	116	184	184	240	242	116	132	110	116
Cortibel 31	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cortibel 32	166	168	118	120	104	116	184	184	238	240	130	130	108	116
Cortibel 33	166	168	120	120	104	116	184	184	238	240	132	132	108	116
Cortibel 34	-	-	120	120	116	116	184	184	-	-	132	132	116	116
Cortibel 35	156	168	120	120	104	116	180	184	238	240	130	130	108	116
Cortibel 36	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cortibel 37	166	168	116	120	104	116	184	184	240	242	132	132	108	116
Cortibel 38	166	168	120	120	104	116	184	184	238	240	130	130	108	116
Cortibel 39	166	168	120	120	104	116	184	184	238	240	116	130	110	116
Cortibel 40	166	168	120	120	104	116	184	184	238	240	116	130	108	116
Cortibel 42	166	168	120	120	104	116	184	184	238	240	116	130	108	116
Cortibel 43	166	168	120	120	104	116	184	184	238	240	116	132	108	116
Cortibel 44	166	166	120	120	104	116	184	184	238	240	116	132	108	116
Cortibel 45	166	166	120	120	104	116	184	184	238	240	116	132	110	116
Cortibel 46	166	168	114	118	114	114	184	184	238	240	114	130	105	114
Cortibel 47	156	168	120	120	104	116	186	186	238	240	116	132	-	-
Cortibel 48	166	168	120	120	116	116	184	184	240	240	132	132	-	-
Cortibel 7	156	168	120	120	116	116	184	184	238	240	132	132	-	-
Cortibel 10	156	166	120	120	104	116	184	184	238	240	132	132	110	110
Cortibel 11	156	156	120	120	116	116	184	184	238	240	116	132	110	110
P. sp (Edson)	-	-	-	-	-	-	-	-	-	-	-	-	-	-
P. guajava BA	156	168	-	-	116	116	180	180	238	240	132	132	-	-

P. guajava BA	156	168	120	120	104	116	184	184	238	240	116	132	110	110
P. guajava SE 1	168	168	116	116	116	116	180	180	238	240	128	128	112	112
P. guajava AL	156	166	-	-	104	116	184	184	238	240	128	128	-	-
P. guajava PE 1	174	174	-	-	-	-	186	186	-	-	-	-	-	-
P. guajava PB	156	168	118	120	116	116	184	184	238	240	130	130	-	-
P. guajava CE	166	168	116	120	104	116	180	180	238	240	116	132	110	110
Híbrido SE 1	156	168	116	120	104	116	184	184	238	240	132	132	110	110
Nego 1	154	168	116	116	104	116	184	186	238	240	132	132	110	110
Nego 2	154	168	116	120	104	116	184	184	238	240	132	132	110	110

Apêndice 3. Genotipagem das amostras de 24 espécies de *Psidium*.

IND.	P29	P 104	P 187	P 206	P 208	P 216	P 439								
1.P. acutangulum	154	171	-	-	-	113	113	-	-	-	-	-	-	-	
2.P. appendiculatum	165	169	171	-	-	101	106	112	180	188	-	-	-	-	
3.P. bahianum	161	161	-	-	-	118	118	-	-	-	-	-	-	-	
4.P. bergianum	-	-	-	-	-	117	119	-	183	192	-	-	127	131	-
5.P. brownianum	167	170	-	123	123	118	118	-	184	186	-	-	129	131	-
6.P. sp.1 (caatinga)	167	169	-	-	-	106	118	-	182	191	-	-	131	131	-
7.P. cattleyanum	171	171	-	-	-	-	-	-	-	-	-	-	130	130	-
8.P. cauliflorum	156	156	-	123	123	106	117	-	190	198	-	-	-	-	111
9.Psidium firmum	167	167	-	-	-	106	118	-	182	192	-	-	-	-	-
10.P. friedrichsthalianum	176	178	-	-	-	106	118	-	183	192	-	-	-	-	-
11.P. ganevii	180	180	-	-	-	109	118	-	-	-	-	-	119	134	107
12.P. gaudichaudianum	150	171	-	-	-	105	117	-	-	-	-	-	-	-	-
13.P. grandifolium	155	167	-	114	118	103	115	-	184	184	244	246	128	128	-
14.P. grazielae	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
15.P. guineense	156	165	-	119	119	105	117	-	190	200	-	-	-	-	-
16.P. longipetiolatum	-	-	-	120	120	106	118	-	-	-	-	-	127	127	-
17.P. myrsinites	175	179	-	120	120	105	117	-	192	192	-	-	127	127	-
18.P. myrtoides	181	185	-	121	121	118	120	-	-	-	-	-	-	-	-
19.P. oblongatum	155	155	-	-	-	120	120	-	-	-	-	-	131	131	110
20.P. oligospermum	-	-	-	-	-	-	-	-	-	-	-	-	131	131	111
21.P. ovale	185	185	-	-	-	106	118	-	-	-	-	-	127	127	-
22.P. rombeu	-	-	-	118	118	116	116	-	-	-	-	-	-	-	110
23.P. robustum	165	165	-	-	-	118	118	-	192	192	-	-	126	126	110
24.Psidium rufum	165	165	-	-	-	105	117	-	189	189	-	-	-	-	-

5. CAPÍTULO II:

CARACTERIZAÇÃO DE SNPs ORIUNDOS DE DArTseq VISANDO UTILIZAÇÃO EM MÚLTIPLAS ESPÉCIES DE *Psidium*

5.1 Introdução

Marcadores moleculares são polimorfismos de DNA que podem ocorrer em regiões codificadoras ou não codificadoras no genoma e são de ampla utilização em espécies vegetais (SANSALONI, 2012). Dentre os vários marcadores moleculares, os SNPs (*Single Nucleotide Polymorphism*) vem sendo cada vez mais utilizados, por apresentarem precisão, alto rendimento, resolução e reprodutibilidade (ILLUMINA, 2017). Esses marcadores são os mais abundantes, amplamente distribuídos e de alta densidade no genoma, favorecendo a detecção de múltiplos haplótipos pela combinação de alelos de SNPs fortemente ligados. São comumente bi-alélicos e podem ser identificados através da troca, inserção ou deleção de nucleotídeos. A utilização destes marcadores tem sido amplamente realizada devido a possibilidade de sua identificação automatizada em plataformas de Sequenciamento de Nova Geração (NGS - *Next Generation Sequencing*) (DESCHAMPS *et al.*, 2012).

Marcadores moleculares genômicos não afetados pela seleção são denominados neutros (ANDERSEN & LUBBERSTEDT, 2003). Estes marcadores são importantes para avaliar parâmetros como a diversidade dentro de populações, diferenciação genética entre populações, endogamia, melhoramento e eventos demográficos. Por outro lado, marcadores não neutros marcam regiões genéticas que são diretamente influenciadas pela seleção natural, permitindo uma série de estudos sobre seleção natural e adaptação local de populações naturais para uma ampla quantidade de grupos taxonômicos (KIRK & FREELAND, 2011).

Os marcadores moleculares podem também ser classificados como funcionais (*Functional markers*), quando são derivados de sítios polimórficos dentro de genes, ou aleatórios (*Randon DNA markers*), quando se encontram em sítios polimórficos e aleatórios no genoma (ANDERSEN & LUBBERSTEDT, 2003). A presença de SNPs em regiões codificadoras podem ter uma significância funcional, se resultar em troca de aminoácidos e alterar o fenótipo (GAITÁN-SOLÍS *et al.*, 2008), o que é de particular

interesse ao melhoramento. Entretanto, é relatado que o acúmulo de variações é maior em regiões não codificadoras do genoma (GAITÁN-SOLÍS *et al.*, 2008)

A escolha do marcador mais adequado a ser utilizado depende de suas propriedades físicas e localização genômica, do custo envolvido, da facilidade de uso e informatividade. Milhares de marcadores moleculares neutros, aleatórios ou funcionais podem ser gerados para qualquer espécie e têm sido utilizados com sucesso em muitos estudos para representar genomas de vegetais. Em geral, o conteúdo genômico entre indivíduos de uma espécie, ou até mesmo entre espécies diferentes (em táxons próximos) é altamente conservado, sendo as variações genéticas mais comuns aquelas caracterizadas por diferenças em um único nucleotídeo quando comparada com regiões equivalentes de genomas distintos (BROOKES, 1999).

Algumas metodologias baseadas em NGS combinam a descoberta de polimorfismo e genotipagem em um único passo (ELSHIRE *et al.*, 2011), dessa forma, os SNPs são detectados e genotipados nos indivíduos, e depois é feita uma seleção de quais serão usados com base em um conjunto de critérios de filtragem. Neste sentido, a metodologia DArTseq, foi adaptada para utilização em plataforma NGS, possibilita a genotipagem e sequenciamento de centenas ou milhares de SNPs ao longo do genoma (A. KIllian com. pessoal).

O sequenciamento realizado por DArTseq permite detecção de SNPs baseando-se na redução da complexidade do genoma por meio do uso de enzimas de restrição especificamente otimizadas para cada espécie. Tem como principal vantagem a seleção inteligente das frações do genoma a ser estudado, visto que a enzima atua principalmente próximo a regiões metiladas, o que corresponde principalmente a genes ativos. A metodologia DArTseq é realizada através de uma estratégia multiplex, que consiste no sequenciamento simultâneo de diferentes amostras. Para isto, é adicionado um pequeno adaptador (*barcode*) que difere para cada uma das amostras a ser sequenciada (JACCOUD *et al.*, 2001).

Marcadores SNPs sequenciados a partir da estratégia DArTseq vêm sendo muito utilizados em estudos com diferentes finalidades para várias espécies vegetais, tais como diversidade genética e estrutura de populações (EGEA *et al.*, 2017; NDJIONDJOP *et al.*, 2017), identificação e origem genética (GARAVITO *et al.*, 2016), filogenia (AL-BEYROUTIOVÁ *et al.*, 2016), caracterização genômica (VALDISSER *et al.*, 2017), seleção genômica para resistência (SANTOS *et al.*, 2016), estudo de

associação genômica ampla - GWAS (QIU *et al.*, 2015; SUKUMARAN REYNOLDS & SANSALONI, 2018) e mapa genético saturado (NIMMAKAYALA *et al.*, 2016; REN *et al.*, 2015).

Com base nessas características, marcadores SNPs são ideais para estudos no gênero *Psidium*, representado por um grande número de espécies presentes nos mais diversificados domínios fitogeográficos. Dentre as inúmeras espécies frutíferas do gênero, apenas a goiabeira é amplamente cultivada e estudada, principalmente com a finalidade de caracterizar seu germoplasma, e desenvolver novas cultivares (POMMER & MURAKAMI, 2009). Entretanto, às espécies não domesticadas ou cultivadas também exibem fonte de variabilidade e resistência, bem como apresentam interesse para a indústria farmacêutica, sendo fonte de recursos genéticos desconhecidos (DIAS *et al.*, 2015; MENDES *et al.*, 2017).

Para otimizar a eficiência desses estudos, as tecnologias de nova geração têm permitido experimentos com genotipagem de vários marcadores SNPs por amostras simultaneamente. Uma dessas tecnologias é denominada painel de genotipagem ou BeadChip, que consiste em uma placa de vidro com milhares de marcadores genéticos funcionando como sondas. O BeadChip permite aos investigadores genotipar milhares ou milhões de SNPs ou analisar outras variantes genômicas, como o número de cópias, em um grande número de amostras de DNA (ADLER, WILEY & GAFFNEY, 2013).

Com o BeadChip é possível identificar regiões de SNPs comuns para as diferentes espécies de *Psidium* e caracterizá-las quanto a estrutura e função. Para isso, é necessário que os marcadores selecionados sejam conservados, com alta cobertura interespecífica, sendo selecionados preferencialmente SNPs presentes em regiões gênicas. Além disso, eles devem ser polimórficos e úteis no estudo da variação intra e interespecíficas. Dessa forma, este trabalho objetiva identificar e caracterizar um grupo de marcadores SNPs, úteis em múltiplas finalidades, para o gênero *Psidium*.

5.2 Materiais e Métodos

5.2.1 Material Vegetal

Um total de 94 amostras, oriundas de nove espécies de *Psidium* foram utilizadas neste estudo (Tabela 1). As amostras de *Psidium guajava* foram as mais numerosas, totalizando 64 amostras, das quais: 19 são cultivares que foram obtidas de viveiro comercial (Chinesa Branca, Indiana, Kuse, Maçã, Paluma, Pedro Sato, Roxa, Século XXI, Tailandesa, Tailandesa Branca, Petri, Cortibel LG, Cortibel LM, Cortibel Branca LG, Cortibel RM, Cortibel Branca RM, Cortibel RM2, Cortibel RG, Cortibel SLG); 33 são genótipos de provenientes de seleção massal de material oriundo se polinização aberta denominadas Cortibel (XVIII ao XLVIII); 12 amostras são goiabeiras de ocorrência natural, coletadas em Alagoas, Bahia, Ceará, Espírito Santo, Pernambuco, Paraíba, Sergipe, Paraná e São Paulo; Um híbrido natural de *P. guajava* com *P. guineense* coletado em Sergipe; Finalmente mais oito espécies de *Psidium* com quantidade de amostras variáveis, sendo elas: um *P. gaudichaudianum*, um *P. acutangulum*, dez *P. cattleyanum*, um *P. friedrichsthalianum*, cinco *P. guineense*, sete *P. myrtoides* e quatro *P. oblongatum*.

Tabela 1. Caracterização das espécies utilizadas quanto ao Número de Amostras (NA), Ploidia, Número de Cromossomos, Origem/Endemismo e Domínios Fitogeográficos.

Espécies	NI	Ploidia*	N. de* cromossomos	Origem*/ Endemismo	Domínios* Fitogeográficos
<i>Psidium acutangulum</i>	01	4x	44	Nativa/ Não endêmica	Amazônia
<i>Psidium cattleyanum</i>	10	4x	44	Nativa/ Endêmica	Caatinga, Cerrado, Mata Atlântica
		7x	77		
		8x	88		
<i>Psidium friedrichsthalianum</i>	01	2x	22	-	-
		4x	44		
<i>Psidium gaudichaudianum</i>	01	-	-	Nativa/ Endêmica	Mata Atlântica

Continuação

<i>Psidium guajava</i>	64	2x	22	Naturalizada/ Não endêmica	Amazônia, Caatinga, Cerrado, Mata Atlântica
<i>Psidium guineense</i>	05	4x	44	Nativa/ Não endêmica	Amazônia, Caatinga, Cerrado, Mata Atlântica
		5x	55		
<i>Psidium myrtoides</i>	07	8x	88	Nativa/ Endêmica	Caatinga, Cerrado, Mata Atlântica
<i>Psidium oblongatum</i>	04	2x	22	Nativa/ Endêmica	Mata Atlântica
Híbrido	01	-	-	-	-

Número de indivíduos amostrados (NI);

*Informações retiradas da literatura (BERNARDES, 2017; FLORA DO BRASIL 2020, em construção).

5.2.2 Extração de DNA, quantificação e incubação

O DNA de cada amostra foi extraído a partir de folhas maceradas com nitrogênio líquido. A extração de DNA foi realizada no Laboratório de Genética e Melhoramento Vegetal da Universidade Federal do Espírito Santo, Campus Alegre. Para extração do tecido foliar macerado, foi utilizado o método CTAB (Doyle e Doyle, 1990) com modificações do IAC, para amostras que contém muito açúcar. A caracterização inicial da quantidade e qualidade das amostras de DNA foram verificadas em NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA, EUA) e visualizada em gel de agarose a 0,8%.

Para atender à exigência da metodologia de genotipagem DArTseq, 2 µL de cada amostra do DNA genômico purificado foi submetido a uma incubação, com tampão da enzima HINFI e corante loading buffer 10x, totalizando um volume final de 12 µL, por 2 horas a 37°C. As amostras submetidas à eletroforese, em gel de agarose a 0,8% com tampão TAE 1X. Apenas as amostras que permaneceram integras foram aceitas para o sequenciamento. Um volume de 60 µL cada amostra contendo entre 60 a 100 ng de

DNA foi pipetado em placas de PCR e enviados para a empresa *Diversity Arrays Technology Pty. Ltd*, de acordo com as recomendações da empresa (<https://www.diversityarrays.com/faq/>) em Canberra – Austrália para produção da biblioteca e da genotipagem por sequenciamento.

5.2.3 Genotipagem por Sequenciamento em plataforma DArTseq

A tecnologia foi otimizada para *Psidium* considerando tanto o tamanho da representação quanto a fração do genoma selecionada para análise. A representação do genoma foi obtida a partir de biblioteca genômica adquirida a partir de digestão com as enzimas de restrição PstI (corte raro) e MseI (corte comum). Às extremidades dos fragmentos clivados foram ligados a um adaptador com *barcode* e um adaptador comum. Os fragmentos misturados foram amplificados utilizando dois iniciadores com sequências complementares aos adaptadores ligados e aos oligonucleotídeos da plataforma de sequenciamento Illumina HiSeq2000.

Todas as amplificações que tiveram sucesso foram agrupadas e levadas a uma *flowcell* para amplificação, onde foram gerados vários moldes de moléculas de DNA. Posteriormente os clusters foram sequenciados em plataforma NGS (KILIAN *et al.*, 2012).

5.2.5 Análise de dados

Os dados DArTseq foram baixados como arquivos .csv e processados para controle de qualidade, análises e interpretação. Considerando que o estudo envolve nove espécies, os marcadores avaliados não necessariamente podem estar presentes em todas as espécies, bem como poderia ocorrer a presença de marcadores espécie específicos. Assim a partir das 94 análises não foram aplicados nos dados obtidos filtragem dos marcadores baseados em MAF e callRate mínimos embora todos os parâmetros relativos a cada *locus* tenha sido fornecido pela empresa para os parâmetros listados na Tabela 2.

Tabela 2. Quadro definindo as principais métricas utilizadas para caracterização dos marcadores SNPs através do pacote DArTR.

Metadados	
AvgCountRef	A soma das contagens de leitura de tags para todas as amostras, dividida pelo número de amostras com contagens de leitura de tags diferentes de zero, para a linha de alelos de referência
AvgCountSnp	A soma das contagens de leitura de tags para todas as amostras, dividida pelo número de amostras com contagens de leitura de tag diferente de zero, para a linha de alelos SNP
AvgPIC	A média do conteúdo de informação de polimorfismo (PIC) das linhas de alelos de Referência e SNP
CallRate	A proporção de amostras para as quais a chamada de genótipo é "1" ou "0", em vez de "-"
FreqHets	A proporção de amostras que pontuam como heterozigotas.
FreqHomRef	A proporção de amostras com pontuação homozigótica para o alelo de referência
FreqHomSnp	A proporção de amostras que pontuam como homozigotas para o alelo SNP
OneRatioRef	A proporção de amostras para as quais a pontuação genotípica é "1", na linha de alelos de referência
OneRatioSnp	A proporção de amostras para as quais a pontuação genotípica é "1", na linha alélica do SNP
PICRef	O conteúdo de informação de polimorfismo (PIC) para a linha de alelos de referência
PICSnp	O conteúdo de informação de polimorfismo (PIC) para a linha de alelos do SNP
RepAvg	A proporção de pares de ensaios técnicos replicados para os quais a pontuação do marcador é consistente

Para as análises por espécie, foram obtidas novas planilhas apenas com amostras de *P. guajava*, *P. cattleyanum*, *P. guineense*, *P. myrtoides* e *P. oblongatum* e todos os parâmetros da Tabela 2 foram recalculados para cada grupo de espécie. Adicionalmente também foram calculados: o número de marcadores SNPs polimórficos, número de SNPs monomórficos, número de SNPs totais por espécie, número de marcadores não amplificados do total amplificado nas nove espécies.

Nos dados dos marcadores SNPs nas 94 amostras fornecidos pela empresa DArT Pty Ltd, também foram apresentadas as informações sobre a presença das sequências que continham os SNPs no genoma de *Eucalyptus grandis*, a espécie referência para estudos genômicos em Myrtaceae. Os dados de presença ou ausência no genoma de *E. grandis* foram utilizados para classificar os marcadores como ancorados, se presentes no genoma de *E. grandis*, ou não ancorados, caso contrário.

Identificação e caracterização de marcadores presentes em regiões codificadoras

Para o estudo dos polimorfismos e as relações filogenéticas entre espécies utilizando marcadores presentes em regiões codificadoras e não-codificadoras, as sequências consenso trimadas, nas quais estavam os SNPs, e que foram fornecidas pela empresa de genotipagem, foram utilizadas para realização de buscas de bancos de dados de sequências públicas e em sequências genômicas de goiabeiras do grupo de pesquisa.

Dados de anotação e sequências genômicas de *Eucalyptus grandis* foram utilizadas para obter a anotação funcional dos marcadores. No intuito de identificar os marcadores mais conservados obtidos neste estudo, as sequências únicas nas quais os SNPs estão contidos foram divididas em dois grupos: sequências presentes no genoma de *Eucalyptus grandis* (Ancorados) e sequências não encontradas neste genoma (Não Ancorados). Para determinar a presença foi realizado um *Blast* das sequências contra o genoma de *E. grandis* obtido no banco de dados *Phytozome*. Para determinar a localização gênica, o grupo Ancorados foi submetido a um *Blast* na sequências gênicas de *E. grandis*, obtidas no banco de dados *Phytozome*.

Para a obtenção das sequências gênicas dos marcadores Não Ancorados em eucalipto foi realizado *Blast* contra sequências gênicas de *P. guajava* fornecidas pelo grupo de pesquisa do PPGM UFES (sequências ainda não publicadas). O *Blast* foi feito através do software *Blastn* (NCBI), com parâmetros *default* com formato de saída alterado para *csv* (valores separados por vírgulas) e *e-value* 1e-7.

Após determinadas as sequências gênicas sua anotação funcional foi obtida através de estratégias diferentes para cada grupo. Para o grupo Ancorados, a anotação foi obtida com auxílio da tabela de anotação de *E. grandis*, disponibilizada no banco de dados *Phytozome*, cruzando as tabelas com programa *Join* disponível nos sistemas

operacionais Linux. No caso dos Não Ancorados, foi necessário obter a sequência de aminoácido traduzido por cada gene disponível em dados da anotação estrutural de *P. guajava*, para então realizar *Blastp* no banco de dados NR e obter uma tabela com associação de gene predito em *P. guajava* X identificador do gene (*gi*). A anotação funcional de cada *gi* foi obtida com auxílio do banco de dados *Uniprot* (www.uniprot.org/uploadlists/) ao submeter os identificadores na página.

Finalmente, os dados de três tabelas diferentes foram cruzados utilizando o banco de dados *MySQL* (Tabela 3: relaciona sequência gênica do grupo Não Ancorados com genes predito em *P. guajava*; Tabela 2: relaciona gene predito em *P. guajava* com *gi*; Tabela 3 relaciona *gi* a sua anotação funcional). Os arquivos oriundos dos processamentos foram analisados para obter estatísticas e gerar gráficos de abundância para GOs (*Gene Ontology*) em cada grupo de sequência citados.

Uma ontologia é uma representação formal e consistem em um conjunto de classes com relações que operam entre elas. O vocabulário GO foi projetado para ser independente de espécie e inclui termos aplicáveis a procariotas e eucariotas, bem como a organismos únicos e multicelulares. A *GO* descreve o domínio biológico com relação a três aspectos:

1. Função Molecular: atividades de nível molecular realizadas por produtos gênicos. Termos de função molecular descrevem atividades que ocorrem no nível molecular.
2. Componente Celular: locais relativos a estruturas celulares nas quais um produto genético desempenha uma função. Não se refere a processos, mas sim a anatomia celular.
3. Processo Biológico: Os processos maiores realizados por múltiplas atividades moleculares (Disponível em <http://geneontology.org/docs/ontology-documentation/>).

Seleção de marcadores

A análise entre espécies foi realizada usando distância euclidiana para análise de coordenadas principais (*PCoA*) e um dendograma com agrupamento através da ligação média entre grupos (*UPGMA*), ambos realizados pelo software R.

Para seleção dos melhores marcadores SNPs visando compor um painel para múltiplas espécies de *Psidium*, foram analisados os parâmetros descritos na Tabela 2.

Esses parâmetros constituem uma análise descritiva que foi realizada para cada espécie separadamente utilizando o pacote DArTR através do programa R. Para contemplar o maior número de espécies, foram filtrados marcadores que estão presentes nas 94 amostras, Call Rate =1,00) e MAF > 0,01.

5.3 Resultados

5.3.1 Caracterização dos marcadores SNPs

Os resultados da genotipagem realizada pelo método DArTseq com base em 94 amostras (com nove espécies de *Psidium*, Tabela 1) resultou em um total de 124.069 marcadores SNPs, os quais representaram 64.672 sequencias distintas. Para a caracterização destes marcadores foram realizadas análises em grupos subdivididos. Primeiramente, os SNPs foram separados em ancorados ou não ancorados no genoma do *Eucalyptus*, a fim de observar os marcadores mais conservados, uma vez que são compartilhados por dois gêneros distantes na família *Myrtaceae*. Alternativamente foram avaliados os que ocorreram apenas nas espécies de *Psidium*. Um total de 22.050 SNPs foram encontrados ancorados e distribuídos nos 11 cromossomos de *E. grandis* (Figura 1, Tabela 3). O baixo valor de SNPs ancorados em Eucalipto, com relação ao número total, representa a distância filogenética entre os dois gêneros.

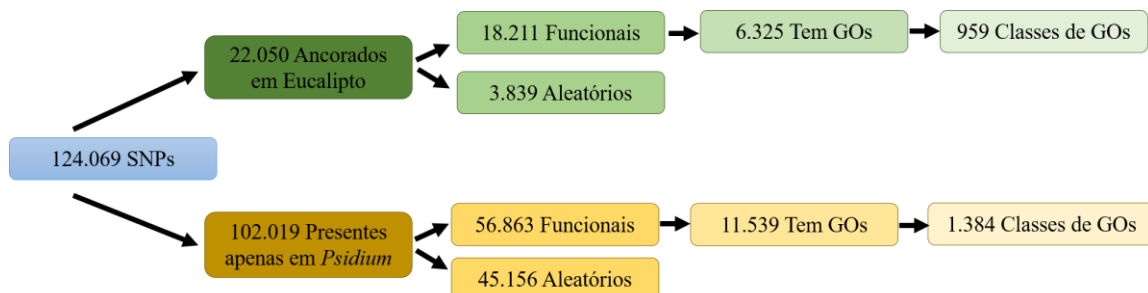


Figura 1. Diagrama mostrando o panorama geral dos marcadores SNPs genotipados a através da metodologia DArTseq. Os marcadores foram divididos em dois grupos: Ancorados em *E. grandis* e Não Ancorados, dentro de cada grupo eles foram subdivididos em Funcionais e Aleatórios. Essas divisões auxiliaram no estudo e seleção dos marcadores.

Todos os dados da análise descritiva para todos os marcadores (Apêndice 1), para os ancorados no genoma de *Eucalyptus* (Apêndice 2) e para os marcadores não ancorados (Apêndice 3) podem ser consultados no material suplementar. As métricas para cada grupo estão subdivididas por espécies, assim é possível visualizar as variações intra e interespecíficas.

Tabela 3. Descrição, por cromossomo, da distribuição dos 22050 SNPs presentes em espécies de *Psidium*, que foram ancorados no genoma do *Eucalyptus grandis*.

Cromossomo de <i>E. grandis</i>	Número de SNPs	Média dos PICs*	Amplitude dos PICs	Média de Frequência de Homozigotos (R)	Média de Frequência de Homozigotos SNP	Média de Frequência de heterozigotos	Média Call Rate**	Amplitude do Call Rate
1	1733	0.1962	0.4992 – 0.0105	0.7629	0.2095	0.0275	0.7893	1 – 0.2021
2	2241	0.1909	0.4965 – 0.0105	0.7773	0.1974	0.0252	0.8003	1 – 0.2021
3	1891	0.1898	0.4986 – 0.0105	0.7944	0.1809	0.0245	0.7829	1 – 0.2021
4	1139	0.1931	0.4950 – 0.0105	0.7778	0.1941	0.0279	0.7856	1 – 0.2021
5	1581	0.1890	0.4992 – 0.0105	0.7759	0.1976	0.0264	0.7838	1 – 0.2021
6	2690	0.1951	0.5 – 0.0105	0.7758	0.1973	0.0267	0.7846	1 – 0.2021
7	1698	0.1944	0.4992 – 0.0105	0.7823	0.1914	0.0261	0.7981	1 – 0.2021
8	2594	0.1944	0.5 – 0.0105	0.7708	0.2030	0.0261	0.7816	1 – 0.2021
9	1560	0.1940	0.4980 – 0.0105	0.7701	0.2053	0.0245	0.7997	1 – 0.2021
10	1983	0.2005	0.5 – 0.0105	0.7822	0.1930	0.0246	0.7871	1 – 0.2021
11	2022	0.2039	0.5 – 0.0105	0.7586	0.2130	0.0282	0.7819	1 – 0.2021
Scaffolds	918	0.1887	0.4986 – 0.0108	0.7787	0.1925	0.0287	0.7833	1 – 0.2021
Não Ancorados	102.019	0.1968	0.5 – 0.0105	0.7537	0.2225	0.0238	0.7182	1 – 0.2021
Total	124.069	0.1943	0.4988 – 0,0105	0.7739	0.1998	0.0262	0.7828	1 – 0.2021

*Média do PIC dos alelos referências e SNPs

** Proporção de amostras para a qual a chamada de genótipo é “1” ou “0” em vez de “–“

Os marcadores SNPs ancorados foram separados por presença em cada um dos 11 cromossomos de eucalipto ou *Scaffolds* (Tabela 3). O número de marcadores variou entre 1139 (cromossomo 4) à 2640 (cromossomo 6), sendo o cromossomos 6, 8 e 11 com maior quantidade de SNPs em sequências conservadas com *Psidium*. Os parâmetros de PIC, frequência do homozigoto referência ou SNPs, frequência do heterozigoto foram similares por cromossomo. A heterozigosidade dos marcadores foi baixa (0.02) nos cromossomos e a cobertura de dados foi alta, (Call Rate acima de 70%).

A maior variabilidade foi detectada nos marcadores presentes nos cromossomos 10 e 11 (PIC acima de 0,2), valor mínimo considerado em alguns trabalhos com SNP, sugerindo conservação dos *loci* e alto número de genótipos em homozigose, tanto do alelo referência, quanto do SNP. Entretanto, quando observamos a faixa de amplitude (Tabela 3), vemos que os cromossomos 06, 08, 10 e 11 possuem marcadores com PIC de 0,5, valor máximo para marcadores SNPs que são binários, e para os demais cromossomos o valor máximo chega bem próximo de 0,5. Estes resultados mostraram ser possível selecionar marcadores polimórficos, mesmo em *loci* conservados na família.

Os marcadores SNPs da classe não ancorados foram mais numerosos totalizando 102.019 SNPs, presentes apenas no genoma das amostras de *Psidium*. Estes marcadores Não Ancorados, apresentaram baixa média para o PIC (0,19), porém uma faixa de amplitude grande (0.5 – 0.0105) (Tabela 3). As espécies *P. guajava* e *P. guineense* apresentam números de SNPs parecidos entre si, com diferença de aproximadamente 5.000 SNPs nos valores gerais, os quais indicam marcadores específicos de cada espécie.

O número total de SNPs ancorados também foi similar entre *P. guajava* e *P. guineense*, reduzindo para as demais espécies. *Psidium oblongatum* e *Psidium myrtoides* apresentaram o menor número de marcadores ancorados em *Eucalyptus*. Além disso, *P. myrtoides* foi a espécie que menos apresentou marcadores comuns para o gênero (Figura 2). Após visualizarmos os marcadores ancorados e não ancorados, esses dois grupos foram subdivididos em marcadores presentes em regiões gênicas e regiões não gênicas (Figura 2). No geral foram encontrados 75.074 marcadores (60%) presentes em regiões codificadoras, mostrando a efetividade da metodologia DArTseq em reduzir a complexidade e selecionar a fração do genoma que corresponde a genes.

As estimativas para presença de SNPs em regiões codificadoras foram maiores ainda na classe de marcadores ancorados no genoma de eucalipto, para estes 82% dos SNPs (Figura 2) e menores na classe dos marcadores não ancorados, 55%.

No geral, a Figura 2 mostra padrão similar para os parâmetros avaliados em *P. guajava* e *P. guineense*. Bem como, valores similares entre *P. acutangulum* e *P. friedrichsthalianum*. De forma peculiar *P. oblongatum* exibiu valores próximos de marcadores funcionais e aleatórios muito próximos, diferindo das demais espécies, que apresentaram número de marcadores funcionais superior aos aleatórios (Figura 2).

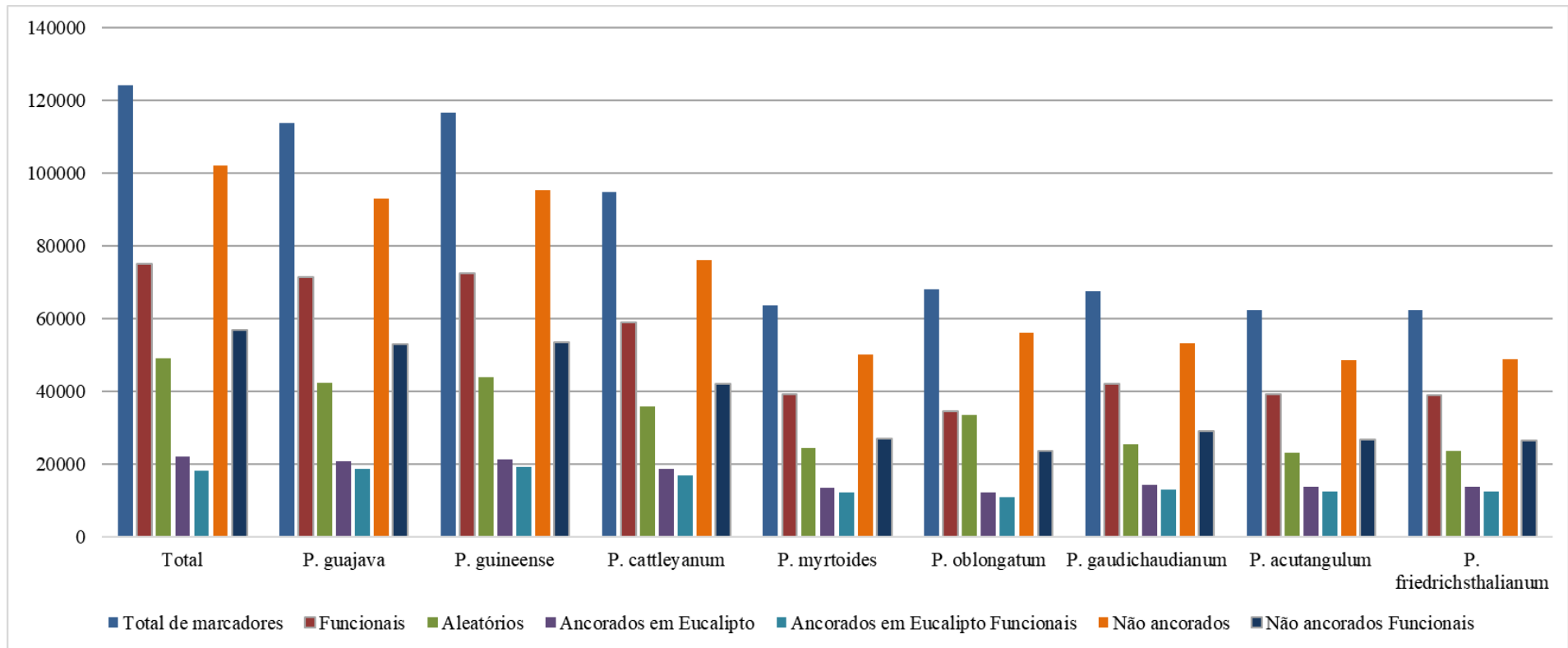


Figura 2. Distribuição de 124.069 marcadores SNPs identificados em nove espécies do gênero *Psidium* nas classes: Funcionais e Aleatórios; Ancorados no genoma de *E. grandis* e ancorados funcionais; Não ancorados e não ancorados funcionais.

A análise dos SNPs por espécies, pelos critérios descritivos de cada marcador, permitiu identificar a presença de marcadores monomórficos, ou seja, quando a sequência do marcador foi amostrada em uma dada espécie, mas sem variação intraespecífica (Figura 3). O número de marcadores monomórficos, detectados em todas as espécies foi expressivo, indicando o compartilhamento de regiões genômicas, mas não a presença de polimorfismos nas amostras avaliadas. O número de marcadores monomórficos foi maior nas espécies *P. guajava*, *P. guineense* e *P. cattleyanum*. Isto revela um maior potencial destes marcadores para revelar diferenças interespecíficas entre as espécies. As três espécies anteriormente citadas apresentaram também uma quantidade muito superior de marcadores polimórficos, em relação às espécies *P. myrtoides* e *P. oblongatum*. Estas por sua vez, foram as que menos compartilharam regiões genômicas com as demais espécies, mostrado pela maior quantidade de marcadores ausentes (NA). Para compor o painel de SNPs é preciso selecionar marcadores que sejam polimórficos intra e interespecíficos.

Diferente do esperado, ao avaliar os marcadores polimórficos quanto a presença em regiões gênicas e não gênicas, verificou-se que para todas as espécies, o número de marcadores polimórficos funcionais foi superior ou similar aos marcadores polimórficos aleatórios. Esses valores ficam mais evidenciados na análise dos marcadores ancorados ao genoma de eucalipto. Entretanto, quando observamos os não ancorados vemos que esses valores diminuem drasticamente. Nesta comparação é a primeira vez que observamos diferença entre as amostras de *P. guajava* e *P. guineense*, visto que a primeira espécie possui 106 marcadores polimórficos aleatórios a mais que os polimórficos funcionais. Este resultado para *P. guajava* contraria o observado nas demais espécies (Figura 3).

Na Figura 3 também é possível observar o número de marcadores que não apareceram nas espécies (identificado como NA). Assim como os marcadores monomórficos, esses dados mostram as regiões não compartilhadas entre as espécies, dessa forma, percebemos que quanto maior for o número de NAs maior será a diferença da espécie em relação às demais. Nos três grupos avaliados *P. oblongatum* é a espécie que possui maior número de NAs, seguida de *P. myrtoides*, essas duas espécies são nativas, endêmicas (Tabela 1), com ocorrência no Espírito Santo e no trabalho de Tuler *et al.* (2015) elas foram agrupadas juntamente com *P. sartorianum* e *P. brownianum*.

Em *P. cattleyanum* é possível notar os valores mais diferenciados. O número de regiões não compartilhadas (NAs) por essa espécie é maior que as observadas em *P. guajava* e *P. guineense*, e é menor que as compartilhadas por *P. myrtoides* e *P. oblongatum*, indicando que esta espécie possui maior variabilidade (Figura 3).

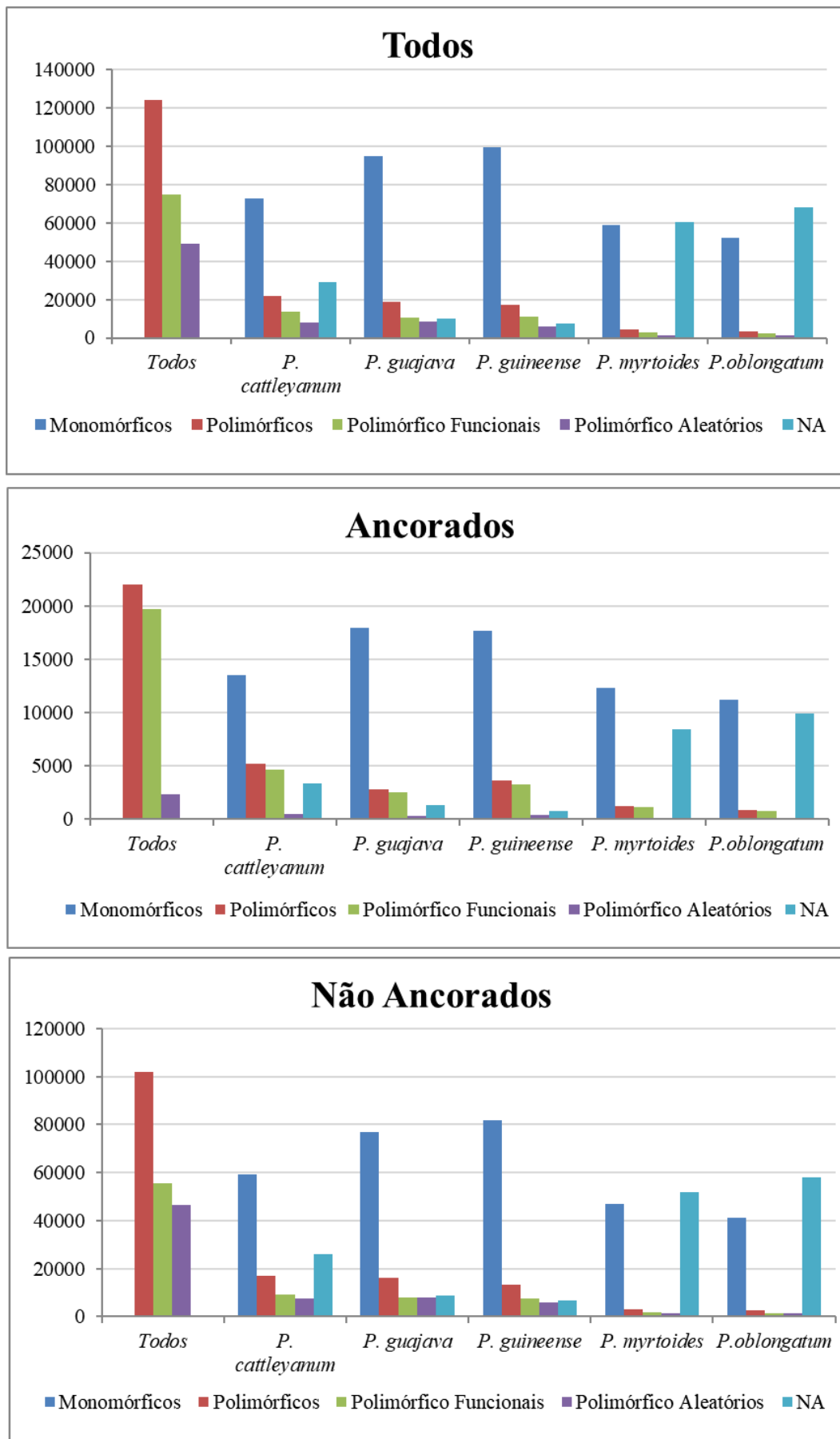


Figura 3. Distribuição dos marcadores SNPs em três grupos: Para todos, ancorados no genoma de *Eucalyptus* e não ancorados. Em cada grupo os SNPs foram subdivididos para cada espécie em Monomórfico, Polimórficos, Polimórficos Funcionais, Polimórficos Aleatórios e Ausência do marcador (NA).

Uma vez separados os marcadores funcionais ancorados (18.211) e não ancorados (75.074), a identificação da anotação funcional foi obtida, mostrando uma lista de potenciais ontologias gênicas (GO). Um gene pode ser encontrado em uma ou mais categorias que estão divididas em três estruturas ontológicas principais: Processo biológico (bp), Componente celular (cc) e Função molecular (mf). Para os marcadores ancorados em eucalipto foram detectados 6.325 SNPs, presentes em regiões codificadoras com anotação funcional, representados por 959 GOs diferentes. Para os SNPs não ancorados foram encontradas 11.539 anotações que identifica 1.384 GOs diferentes (Figura 1).

Visando caracterizar grupos funcionais de genes polimórficos nas espécies de *Psidium* estudadas, na Figura 4 estão representadas as 20 classes gênicas mais abundantes dentro da ontologia processos biológicos, que diz respeito a atividade biológica pertencente ao gene. Apesar do número superior de GOs associados a marcadores não ancorados, para esta categoria ontológica o número de marcadores ancorados é muito maior, sendo 4.810 GOs associados a marcadores ancorados e 2.499 associados à não ancorados. É importante considerar que um gene pode ser encontrado em uma ou mais categorias.

O gráfico relacionado a Componente Celulares (Figura 5), indica os GOs que representam o local na célula onde o gene ou seu produto é ativo. Para este grupo os marcadores não ancorados apresentaram um número muito maior de GOs (6.343). Também é possível observar nos gráficos que tanto os marcadores ancorados e não ancorados apresentaram como mais abundante a classe que representa os Componentes integrais da membrana. O terceiro gráfico (Figura 6), representa os genes ligados a Função Molecular, que diz respeito a atividade bioquímica do gene ou produto. Neste grupo, a classe de genes relacionada a Ligação de ATP foi a mais abundante com 1.518 GOs nos marcadores ancorados e 2.276 nos marcadores não ancorados.

A classe ontológica mais representada pelos marcadores funcionais deste estudo foi a Função molecular, exibindo 48% dos GOs, seguida de Componentes Celulares com 28,5% e Processos Biológicos com 23,5%. Quando discriminado em marcadores ancorados e não ancorados, apenas os Componentes Celulares apresentaram valores superior para não ancorados (72%).

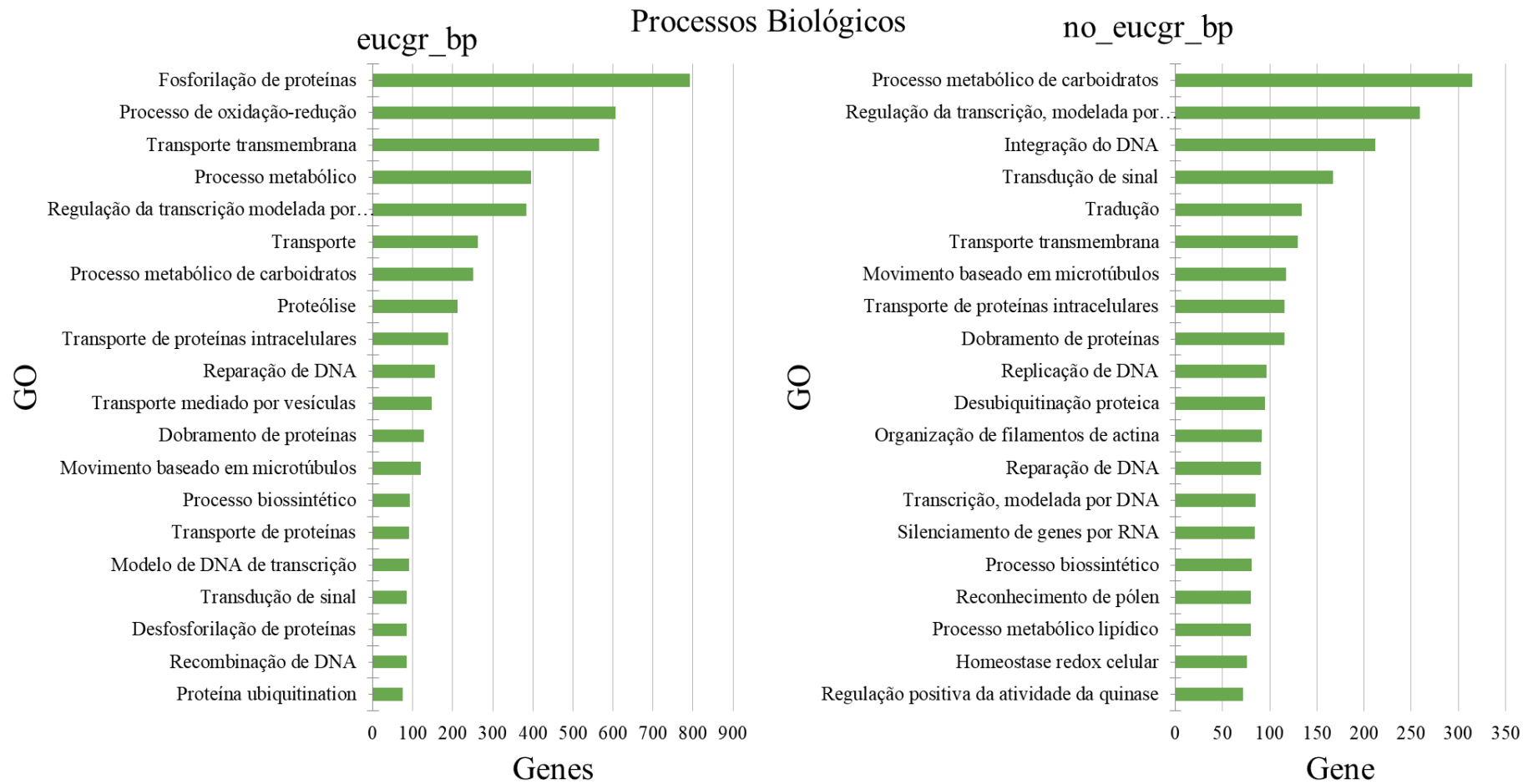


Figura 4. Relação dos marcadores SNPs funcionais com sua ontologia gênica (GO) para genes relacionado a Processos Biológicos (bp). As informações de cada gráfico estão divididas entre marcadores ancorados e não ancorados ao genoma de *Eucalyptus*.

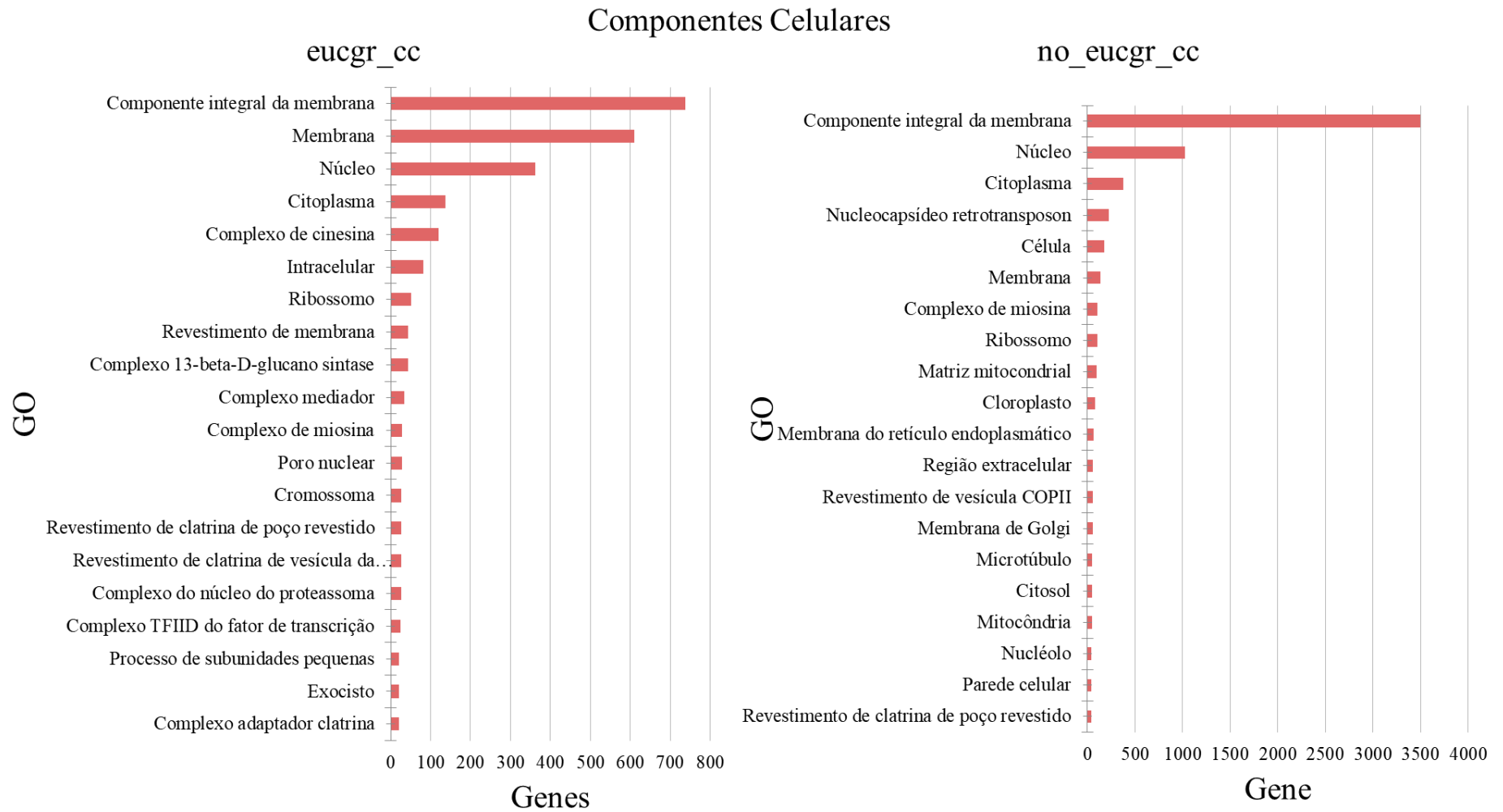


Figura 5. Relação dos marcadores SNPs funcionais com sua anotação funcional (GO) para genes relacionado a Componentes celulares (cc). As informações de cada gráfico estão divididas entre marcadores ancorados e não ancorados ao genoma de *Eucalyptus*.

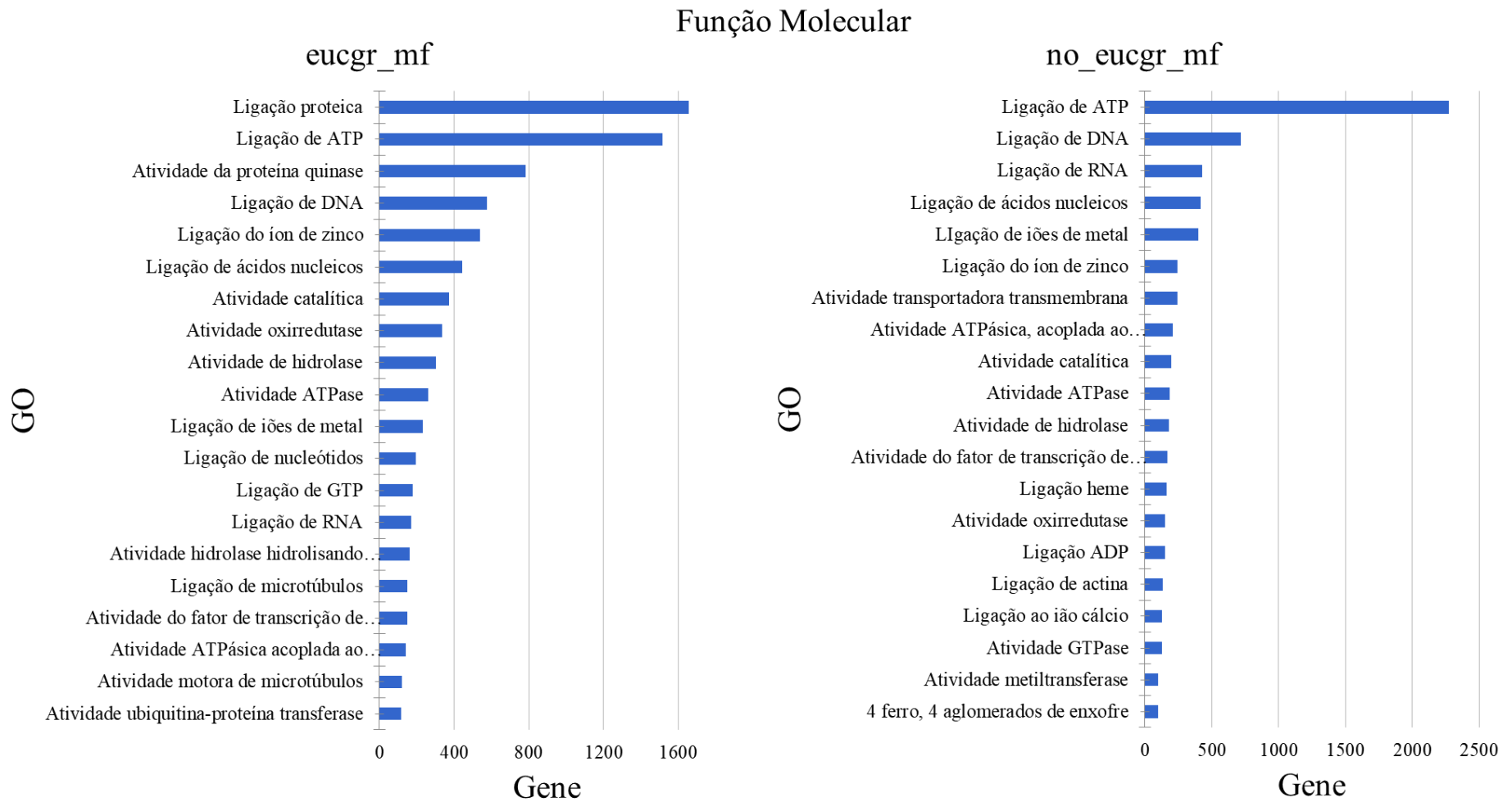


Figura 6. Relação dos marcadores SNPs funcionais com sua anotação funcional (GO) para genes relacionado a Função molecular (mf). As informações de cada gráfico estão divididas entre marcadores ancorados e não ancorados ao genoma de *Eucalyptus*.

A relação genética das nove espécies estudadas foi mostrada nos resultados da análise de coordenadas principais, realizada com os três grupos de dados: todos os marcadores, marcadores ancorados e não ancorados (Figura 7). Os resultados desta análise forneceram uma representação espacial das distâncias genéticas relativas entre os indivíduos, com pouca diferença na proporção da variação explicada na análise com todos os marcadores (59,6%) e com os marcadores não ancorados (58,1%). A maior representatividade dos dados, foi observada na análise com os marcadores ancorados (variação explicada = 66,1%), justificando a seleção destes marcadores como prioridade, para avaliações interespecíficas.

Para as três representações do *PCoA*, a distribuição resultou em três grupos distintos. As amostras de *P. guajava* e *P. guineense* ficaram agrupadas no primeiro eixo juntamente com o Híbrido, que apresentou maior semelhança com *P. guajava*. Pouca variação foi detectada entre os cinco indivíduos de *P. guineense* e as 63 amostras de *P. guajava*.

As espécies *P. myrtoides* e *P. oblongatum* ficaram próximas, confirmando o comportamento analisado na Figura 3, principalmente para o número de marcadores não compartilhados. Entretanto, as quatro amostras de *P. oblongatum* ficaram bem próximas, enquanto, as amostras de *P. myrtoides* exibiram afastamento espacial expressivo.

O eixo três, formado por *P. acutangulum*, *P. cattleyanum*, *P. friedrichsthalianum*, *P. gaudichaudianum* e *Psidium* sp1 (classificada anteriormente como uma goiabeira), foi um grupo de espécies que chamou a atenção, sendo seus integrantes formados por árvores altas. Para as dez amostras de *P. cattleyanum* foi formada uma subdivisão com seis amostras próximas a *P. acutangulum*, *P. friedrichsthalianum* e *P. gaudichaudianum*, que apresentaram muitos marcadores em comum e só possuem uma representação nas amostras. As outras quatro amostras de *P. cattleyanum* se agruparam com uma amostra coletada no Espírito Santo e inicialmente identificada como *P. guajava*, porém essa amostra se comportou de forma muito divergência das goiabeiras, assim ela foi separada e identificada com *Psidium* sp.1 para avaliações futuras.

Após discriminação e avaliação, foram selecionados para compor o painel de marcadores para *Psidium* um grupo que consiste em 5.951 SNPs, com essa seleção foi retirado do painel 95% dos SNPs, restando apenas os conservados e polimórficos entre e dentro das espécies. Entretanto, os demais marcadores ainda devem ser estudados pois servem para possuírem informações intraespecíficas e separar as espécies através da presença ou ausência.

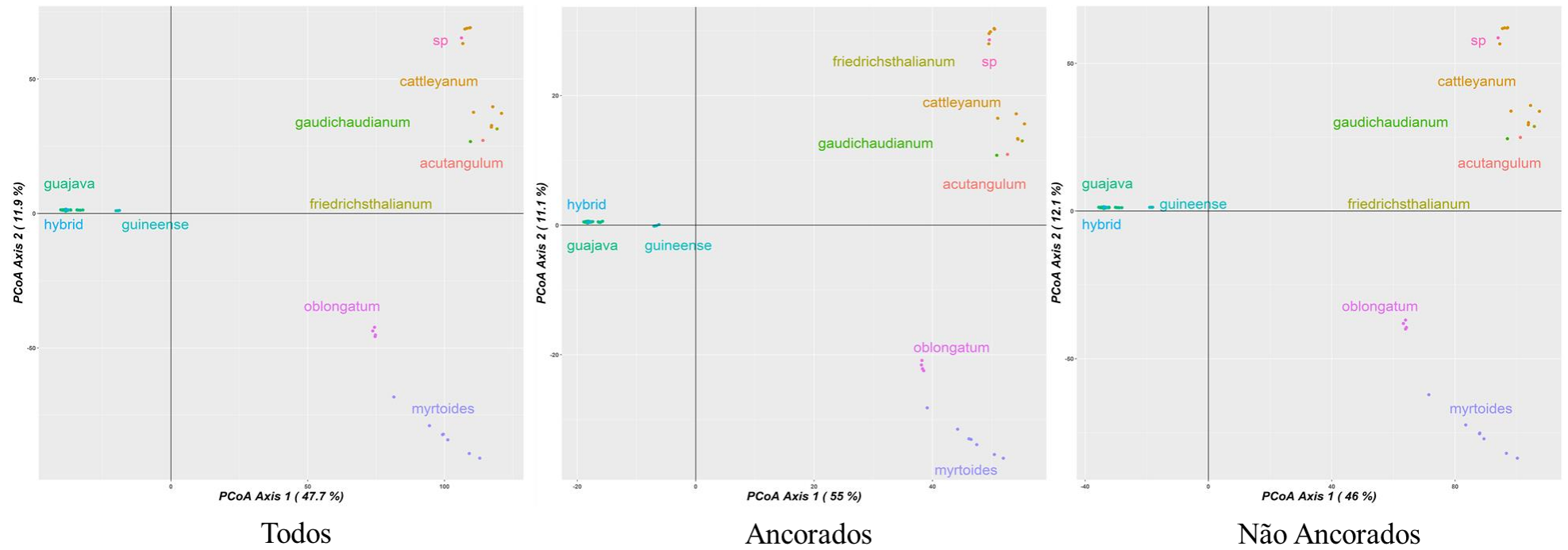


Figura 7. Análise de principais coordenada (*PCoA*) baseada em marcadores SNPs para 94 amostras de *Psidium*. Cada espécie está representada por cores diferentes para todos os marcadores, ancorados e não ancorados.

Dentre os marcadores selecionados 78% são marcadores funcionais (Figura 8). Esses marcadores estão presentes em pelo menos 99% das amostras, apresentando cobertura máxima (Tabela 4), a intenção ao escolher este grupo de SNPs, foi obter marcadores que representassem o maior número de espécies, visto que o número de marcadores ancorados em eucalipto foi alto (37%) e alguns estudos recentes tem mostrado a transferibilidade de marcadores entre gêneros de Myrtaceae.

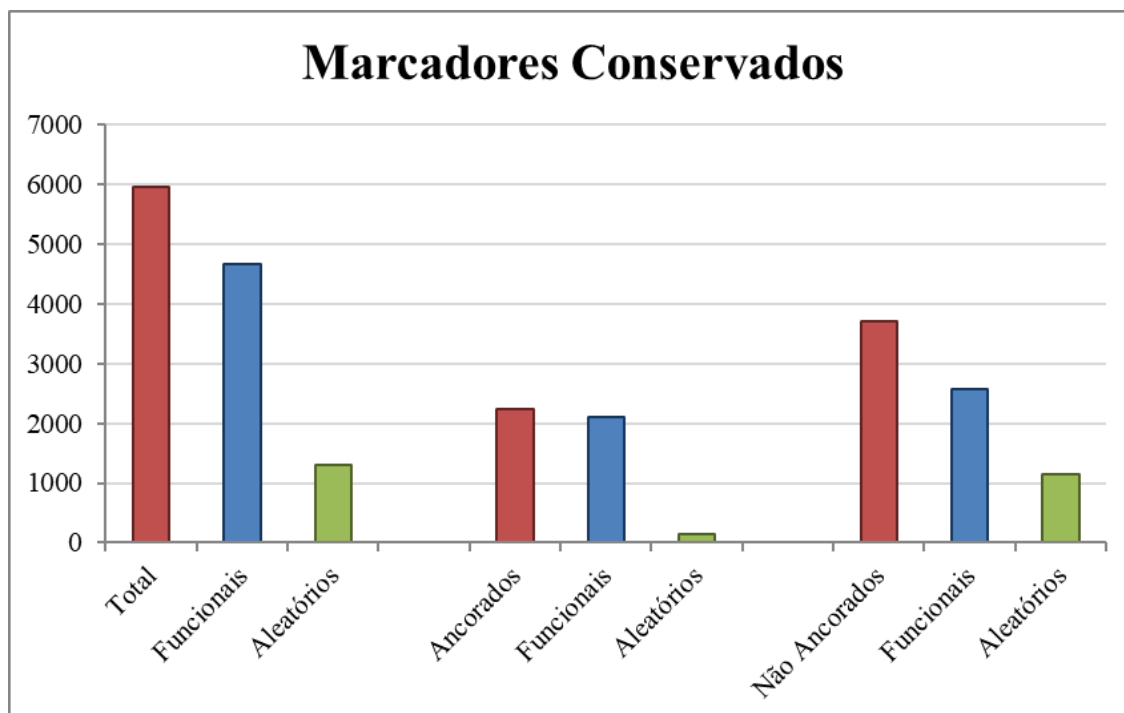


Figura 8. Apresentação dos marcadores SNPs que foram encontrados em pelo menos 99% das amostras. Cada conjunto de barras representam os marcadores totais, ancorados e não ancorados respectivamente.

O número de marcadores ancorados funcionais foi muito alto (93%), este resultado já era esperado, visto que a metodologia prioriza regiões gênicas e para comparação o genoma de *E. grandis* e sua anotação funcional estão disponíveis para consulta.

Finalmente, análise descritiva foi realizada para os marcadores conservados ancorados e não ancorados (Tabela 4). Com esses dados é possível observar valores semelhantes para os três grupos de marcadores além da cobertura máxima. O valor médio da frequência de homozigoto referência são os mais alto (0,74) em contrapartida o frequência média de heterozigoto é bem baixa (0,06), juntando essas informações com o fato de serem marcadores conservados, esperava-se uma média de PIC bem baixa, entretanto para marcadores bialélicos, cujo PIC máximo é 0,5, os valores médios

apresentados são satisfatórios pois variam de 0,22 a 0,25 permitindo um bom desempenho para compor o painel e auxiliar na identificação e seleção de genótipos para o melhoramento.

Tabela 4. Análise descritiva para os 5.951 marcadores selecionados.

	Número de Marcadores	CallRate	FreqHomRef	FreqHomSnp	FreqHets	AvgPIC
Total	5.951	1	0,75	0,19	0,06	0,24
Funcionais	4.658	1	0,76	0,18	0,06	0,24
Aleatórios	1.293	1	0,71	0,23	0,06	0,25
Ancorados	2.237	1	0,76	0,18	0,06	0,22
Funcionais	2.091	1	0,76	0,18	0,06	0,22
Aleatórios	146	1	0,76	0,17	0,07	0,22
Não Ancorados	3.714	1	0,74	0,20	0,06	0,25
Funcionais	2.567	1	0,76	0,19	0,06	0,25
Aleatórios	1.147	1	0,71	0,24	0,06	0,25
Média	----	1	0,74	0,20	0,06	0,24

Com base nos valores de PIC foi realizado um agrupamento para os 5.951 marcadores selecionados, resultando na formação de cinco grupos principais que revela novamente a relação entre as espécies (Figura 9). No primeiro agrupamento é possível observar a maior parte das amostras, sendo composto por *P. gujava* e *P. guineense* juntamente com o híbrido formado a partir dessas duas espécies. Neste agrupamento, uma amostra de *P. guineense* se agrupo distante das demais, este resultado também foi observado nos agrupamentos realizados para PICs maiores que 0,2 e intermediários (Apêndice 5 e 6). Entretanto, as amostras de *P. guineense* usadas neste estudo são de ocorrência natural, assim existe a possibilidade que uma delas seja um híbrido, que devido as semelhanças morfológicas entre as espécies foi identificado de forma incorreta.

O segundo e terceiro agrupamento são formados apenas por amostras de *P. oblongatum* e *P. myrtoides* respectivamente. Quando observado o segundo grupo no dendrograma para PICs menores que 0,02 (Apêndice 5), as amostras de *P. oblongatum* ficam separadas, além disso, esses marcadores formam um grande grupo e não são suficientes para separar as espécies de forma satisfatória.

No quarto grupo foi possível observar a maior diversidade de espécies, sendo composto por, *P. friedrichsthalianum*, duas amostras de *P. cattleyanum*, *P.*

gaudichaudianum e *P. acutangulum*. Essas amostras precisam ser melhor estudadas, pois *P. cattleyanum* é uma espécie que apresenta vários tipos de ploidia e as demais espécies são pouco conhecidas. Finalmente o quinto agrupamento foi formado pelas demais amostras de *P. cattleyanum* juntamente com *Psidium* sp. (identificado anteriormente como goiabeira). Para essa amostra sem identificação, mais estudos também serão realizados.



Figura 9. Agrupamento obtido através dos valores de PIC dos 5.951 marcadores SNPs calculados para 94 amostras de *Psidium*. Espécies identificadas por cor: 63 *P. guajava* (Verde escuro), 05 *P. guineense* (Laranja), 01 Híbrido (Rosa), 04 *P. oblongatum* (Azul escuro), 07 *P. myrtoides* (Lilás), 01 *P. friedrichsthalianum* (Verde claro), 10 *P. cattleyanum* (Vermelho), 01 *P. gaudichaudianum* (Preto), 01 *P. acutangulum* (Amarelo) e 01 *Psidium* sp. (Azul claro).

5.4 Discussão

O número de marcadores genotipados permitiram discriminação interespecífica das espécies avaliadas, o que suplementa a avaliação morfológica das espécies e permite resolver numerosas questões operacionais relacionados à seleção, melhoramento e conservação. Neste estudo, os SNPs oriundos da metodologia DArTseq forneceram informações robustas sobre a variabilidade genética das amostras. Essa ferramenta de sequenciamento apresenta alta produtividade e fornece confiabilidade ao analisar espécies diferentes.

Muitos estudos têm confirmado que os marcadores SNPs derivados da metodologia DArTseq demonstraram que esta tecnologia é um método eficiente de genotipagem com ampla cobertura genômica e pode ser útil para análises da diversidade genética em um *pool* de genótipos composto por amostras silvestre e cultivares (VALDISSE *et al.*, 2017). Considerando os dados de presença do homozigoto referência (0), SNP (1), heterozigoto (2) e ausência (-) usados neste estudo, foram aplicado vários métodos de análise que incluem localização no genoma, polimorfismo e distância genética na análise de coordenadas principais.

A disponibilidade de um genoma de referência dentro da família com alta qualidade (MYBURG *et al.*, 2014) contribuiu positivamente para as análises métricas de genotipagem de SNP. Os altos números de marcadores ancorados ao genoma de *Eucalyptus grandis* aumentam as aplicações desse estudo e os SNPs compartilhados entre esses gêneros serão valiosos para a análise de evolução, filogenia, identidade e parentesco, sintenia e em programas de melhoramento. Os SNPs compartilhados entre *Psidium* e *Eucalyptus*, a maioria presentes em regiões codificadoras, provavelmente representam variantes antigas que surgiram antes da divisão destes táxons e persistiram em linhagens separadas (SILVA-JUNIOR *et al.*, 2015).

Os dados também apresentaram um número muito alto de marcadores funcionais, ou seja, que estão presentes em regiões gênicas. Esses valores eram previstos, pela metodologia DArTseq apresentar a vantagem de redução da complexidade do genoma através de enzimas que atuam principalmente em regiões de metilação intensa, o que corresponde predominantemente a genes ativos (SOUSA *et al.*, 2018). Segundo Andersen e Lubberstedt (2003), os marcadores funcionais são superiores em diversas aplicações devido à sua ligação completa aos motivos

funcionais. Assim, permitem a aplicação confiável de marcadores em populações sem mapeamento devido à fixação alélica mais eficiente e melhor representação da variação genética em populações naturais ou selecionadas.

A análise de coordenadas principais (PCoA) mostrou a representação espacial da variabilidade entre os genótipos. O gráfico mostrou a dispersão diferenciando em três grupos os quais ficaram bem separados nos agrupamentos. O grupo mais homogêneo formado por *P. guajava* X *P. guineense* era esperado, visto que evidências sugerem que a origem do poliploide *P. guineense* se deu a partir de *P. guajava* (MARQUES *et al.*, 2016), além da proximidade genética observada em estudos com microssatélites (TULER *et al.*, 2015). Entretanto, a baixa variabilidade intraespecífica de *P. guajava* e *P. guineense*, em relação às demais espécies estudadas é um fato inédito observado neste estudo.

Um outro agrupamento foi composto por *P. oblongatum* e *P. myrtoides*. Essas duas espécies são nativas, com ocorrência no Espírito Santo e no trabalho de Tuler *et al.* (2015) com microssatélites elas foram agrupadas juntamente com *P. sartorianum* e *P. brownianum*. Três das espécies mencionadas são poliploides, entretanto *P. oblongatum* não (TULER *et al.*, 2019).

O estudo de Costa e Santos (2017) foi o único encontrado utilizando SNPs para avaliar amostras de *Psidium*, entretanto para este trabalho os autores não desenvolveram marcadores próprios para o gênero, fazendo deste o único estudo com marcadores próprios do gênero, até o momento. Para a genotipagem, eles utilizaram EUChip60Kchip, desenvolvido por Silva-Junior *et al.* (2015) para espécies de eucalipto. Utilizando essa metodologia, foi transferido 3.523 SNPs para as amostras de *Psidium*, no presente estudo tivemos um número muito maior ao procurar marcadores de *Psidium* em *Eucalyptus*. Os autores concluíram que a transferência de SNPs entre gêneros foi muito confiável concordando com estudos anteriores de divergência genética através de microssatélites.

Diversas abordagens têm sido empregadas na seleção de SNPs para compor um painel de genotipagem. Silva-Junior *et al.* (2015), para amostras de eucalipto, utilizaram marcadores que exibiram (1) cobertura acima de 80%, (2) GenCall acima de 0.15, (3) agrupamento de genótipos, (4) intensidade média normalizada acima de 0.2, (5) concordância de herança alélica Mendeliana acima de 95% e (6) 100% de

reprodutibilidade em 4 amostras replicadas. Já no trabalho de Unterseer *et al.* (2014), com amostras de milho, os marcadores passaram independentemente por dois softwares, SAMtools e GATK e foram escolhidos baseados em elevadas pontuações nestes programas, além da necessidade de existência de um alelo referência específico para o trabalho (B73).

Para amostras de arroz, Chen *et al.* (2014) seguiram uma lista com diferentes critérios para seleção, sendo eles (1) SNPs heterozigotos, (2) o segundo alelo de cada SNP deve ser encontrado em pelo menos três variantes, (3) não existem outros SNPs ou InDels em 55 pb em pelo menos um dos lados do SNP, (4) as sequências de 55 pb em qualquer um dos lados não tem outra combinação genômica acima de 85%. No trabalho de Bachlava *et al.* (2012), com amostras de girassol, foram selecionado marcadore que apresentavam (1) mínimo de dois ESTs por genótipo na posição do SNP, (2) mínimo de dois genótipos distintos na posição do SNP (que também satisfaçam o critério 1), (3) cobertura acima de 90%.

Como observados, diferentes parâmetros são utilizados para seleção dos marcadores, dependendo das amostras e dos objetivos a serem cumpridos. Essa ferramenta de genotipagem de alta produtividade fornece flexibilidade sem precedentes para genotipar várias espécies do mesmo gênero e demonstra a viabilidade técnica e as vantagens do desenvolvimento deliberado de um chip de genotipagem de SNP entre espécies (SILVA-JUNIOR *et al.*, 2015).

Neste estudo, foram selecionados os marcadores com cobertura total de amostras, seguido de uma filtragem que excluiu os menos informativos (menores PIC). Dessa forma, obtemos um conjunto de marcadores que representam nove espécies diferentes de *Psidium*, esses marcadores apresentam variação interespecíficas, muitos são compartilhados com *Eucaliptus* e estão presentes em regiões funcionais. Espera-se que este estudo, detalhando todas as estratégias adotadas para descoberta, caracterização e seleção de SNPs, possa fornecer dados valiosos para desenvolvimentos de um painel para genotipagem SNPs de larga escala e flexíveis para diferentes espécies de *Psidium*.

5.5 Conclusão

1. Este estudo é o primeiro a utilizar tecnologia DArT para o gênero *Psidium*. Os resultados mostraram que a abordagem DArTseq gerou um grande conjunto de SNPs úteis para *Psidium* com ampla cobertura genômica que permitiram uma caracterização dos marcadores
2. O número de marcadores compartilhados entre *Psidium* e *Eucalyptus* é grande e informativo, sendo eucalipto considerado um gênero chave para pesquisas, esses resultados contribuem com avanços nos estudos de *Psidium*.
3. Verificou-se que o padrão de distribuição dos SNPs nas regiões do genoma difere, com marcadores representativos de regiões codificantes e não codificantes, obtendo-se um número maior de marcadores funcionais quando comparados com os aleatórios. Além disso, foi verificado a existência de marcadores específicos de cada espécie baseado em sua ausência (NA).
4. A partir dos valores de PIC de cada marcador foi possível gerar agrupamentos que mostraram as relações intra e interespecíficas. Estes agrupamentos revelaram que *P. guajava* foi amplamente selecionada e compartilha muitas semelhanças com a espécie poliploide *P. guineense*.
5. Os dados provenientes do *PCoA* revelaram a diversidade das amostras estudadas, permitindo o desenvolvimento de planos para futuras ações na seleção e melhoramento de genótipos com base na associação com características agrônômicas de interesse.
6. Um grande número de marcadores foi encontrado em todas as amostras, sendo considerados conservados entre espécies, e foram selecionados para compor o painel de SNPs interespecíficos para *Psidium*.

5.6 Referências

- Adler, A. J., Wiley, G. B., & Gaffney, P. M. (2013). Infinium Assay for Large-scale SNP Genotyping Applications. *Journal of Visualized Experiments*, 10(81), 1–10. <https://doi.org/10.3791/50683>
- Al-Beyroutiová, M., Sabo, M., Sleziak, P., Dušínský, R., Birčák, E., Hauptvogel, P., Kilian, A. & Švec, M. (2016). Evolutionary relationships in the genus *Secale* revealed by DArTseq DNA polymorphism. *Plant Systematics and Evolution*, 302(8), 1083–1091. <https://doi.org/10.1007/s00606-016-1318-2>
- Andersen, J. R. & Lubberstedt, T. (2003). Functional markers in plants. *Plant Science*. Vol.8. doi:10.1016/j.tplants.2003.09.010
- Bachlava, E., Taylor, C. A., Tang, S., Bowers, J. E., Mandel, J. R., Burke, J. M., & Knapp, S. J. (2012). Snp discovery and development of a high-density genotyping array for sunflower. *PLoS ONE*, 7(1). <https://doi.org/10.1371/journal.pone.0029814>
- Bernardes, C. de O. (2017). *Diversidade genética, caracterização e atividade de óleos essenciais em Psidium spp.* Tese (Doutorado) – Universidade Federal do Espírito Santo / Centro de Ciências Agrárias e Engenharias – Alegre, ES.
- Brookes, A. J. (1999). The essence of SNPs. *Gene*, 234(2), 177–186. [https://doi.org/10.1016/S0378-1119\(99\)00219-X](https://doi.org/10.1016/S0378-1119(99)00219-X)
- Chen, H., Xie, W., He, H., Yu, H., Chen, W., Li, J., ... Zhang, Q. (2014). A high-density snp genotyping array for rice biology and molecular breeding. *Molecular Plant*, 7(3), 541–553. <https://doi.org/10.1093/mp/sst135>
- Costa, S. R. & Santos, C. A. F. (2017). Genetic divergence among *Psidium* accessions based on single nucleotide polymorphisms developed for *Eucalyptus*. *Genetics and Molecular Research* 16 (2): gmr16029566. DOI <http://dx.doi.org/10.4238/gmr16029566>
- Deschamps, S., Llaca, V., & May, G. D. (2012). Genotyping-by-Sequencing in Plants. *Biology*, 1(3), 460–483. <https://doi.org/10.3390/biology1030460>
- Dias, C. N., Alves, L. P., Rodrigues, K. A., Brito, M. C., Rosa, C dos, Amaral, F. M. dos, Monteiro O. dos S., Andrade, E. H., Maia, J. G. & Moraes, D. F. (2015). Chemical composition and larvicidal activity of essential oils extracted from brazilian legal Amazon plants against *Aedes aegypti* L. (Diptera: Culicidae). *J. Evid. Based Complementary Altern. Med.*, p. 1–8
- Doyle, J. J. & Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus* 1213-1215. Erdtman, G. 1952. *Pollen Morphology and Plant Taxonomy – Angiosperms*. Stockolm, Almqvist & Wiksel.
- Egea, L. A., Mérida-García, R., Kilian, A., Hernandez, P., & Dorado, G. (2017). Assessment of genetic diversity and structure of large garlic (*Allium sativum*) germplasm bank, by diversity arrays technology “genotyping-by-sequencing” platform (DArTseq). *Frontiers in Genetics*, 8(JUL), 1–9. <https://doi.org/10.3389/fgene.2017.00098>
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, 6(5), 1–10. <https://doi.org/10.1371/journal.pone.0019379>
- Gaitán-Solís, E., Choi, I. Y., Quigley, C., Cregan, P., Tohme, J. (2008). Single Nucleotide Polymorphisms in Common Bean: Their Discovery and Genotyping Using a Multiplex Detection System. *Plant Genome* 1:125-134.

- doi:10.3835/plantgenome2008.04.0204
- Garavito, A., Montagnon, C., Guyot, R., & Bertrand, B. (2016). Identification by the DArTseq method of the genetic origin of the *Coffea canephora* cultivated in Vietnam and México. *BMC Plant Biology*, 16(1), 1–12. <https://doi.org/10.1186/s12870-016-0933-y>
- Illumina (2017). *Plant Bead Chip*. Maize LD BeadChip Kit: Cost-effective and scalable content for EDV assessment and other maize breeding applications. Disponível em: <https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/maize-ld-data-sheet-1370-2015-007.pdf> Acesso em: 15 de agosto de 2018
- Jaccoud, D., Peng, K., Feinberg, D., & Kilian, A. (2001). Diversity Arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Research*, 29(4), 25e–25. <https://doi.org/10.1093/nar/29.4.e25>
- Kilian A, Wenzl P, Huttner E, *et al* (2012). Diversity arrays technology: a generic genome profiling technology on open platforms. *Methods Mol Biol.* 888:67–89
- Marques, A. M.; Tuler, A. C.; Carvalho, C. R.; Carrijo, T. T.; Ferreira, M. F. S.; Clarindo, W. R. (2016). Refinement of the karyological aspects of *Psidium guineense* (Swartz, 1788): a comparison with *Psidium guajava* (Linnaeus, 1753). *Comparative Cytogenetics*, v. 10, n. 1, p. 117-128.
- Mendes, L. A., Martins, G. F., Valbon, W. R., Souza, T. S. de, Menini, L., Ferreira, A., Ferreira, M. F. da S. (2017). Larvicidal effect of essential oils from Brazilian cultivars of guava on *Aedes aegypti* L. *Industrial Crops & Products*, v. 108, p. 684–689.
- Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D *et al.* (2014). The genome of *Eucalyptus grandis*. *Nature* 510: 356–362.
- NCBI. NCBI BLAST. Disponível em: <https://www.ncbi.nlm.nih.gov/> Acesso em: jan. 2019.
- Ndjiondjop, M.-N., Semagn, K., Gouda, A. C., Kpeki, S. B., Dro Tia, D., Sow, M., Goungoulou, A., Sie, M., Perrier, X., Ghesquiere, A. & Warburton, M. L. (2017). Genetic Variation and Population Structure of *Oryza glaberrima* and Development of a Mini-Core Collection Using DArTseq. *Frontiers in Plant Science*, 8(October), 1–19. <https://doi.org/10.3389/fpls.2017.01748>
- Nimmakayala, P., Tomason, Y. R., Abburi, V. L., Alvarado, A., Saminathan, T., Vajja, V. G., Salazar, G., Panicker, G. K., Levi, A., Wechter, W. P., Mccreight, J. D., Korol, A. B., Ronin, Y., Garcia-Mas, J. & Reddy, U. K. (2016). Genome-Wide Differentiation of Various Melon Horticultural Groups for Use in GWAS for Fruit Firmness and Construction of a High Resolution Genetic Map. *Frontiers in Plant Science*, 7(September), 1–15. <https://doi.org/10.3389/fpls.2016.01437>
- Pommer, C. V.; Oliveira, O. F.; Santos, C. A. F. (2013). Goiaba: recursos genéticos e melhoramento. *Edufersa: Mossoró - RN*. 126 p
- Psidium* in Flora do Brasil 2020 em construção. Jardim Botânico do Rio de Janeiro. Disponível em: <http://floradobrasil.jbrj.gov.br/reflora/floradobrasil/FB605023>>. Acesso em: 17 Ago, 2018.
- Qiu, X., Pang, Y., Yuan, Z., Xing, D., Xu, J., Dingkuhn, M., Li, Z. & Ye, G. (2015). Genome-wide association study of grain appearance and milling quality in a worldwide collection of Indica rice germplasm. *PLoS ONE*, 10(12), 1–25. <https://doi.org/10.1371/journal.pone.0145577>
- Ren, R., Ray, R., Li, P., Xu, J., Zhang, M., Liu, G., Yao, X., Kilian, A. & Yang, X.

- (2015). Construction of a high-density DArTseq SNP-based genetic map and identification of genomic regions with segregation distortion in a genetic population derived from a cross between feral and cultivated-type watermelon. *Molecular Genetics and Genomics*, 290(4), 1457–1470. <https://doi.org/10.1007/s00438-015-0997-7>
- Sansaloni, C. P. (2012). *Desenvolvimento e aplicações de DArT (Diversity Arrays Technology) e genotipagem por sequenciamento (Genotyping-by-Sequencing) para análise genética em Eucalyptus*. Tese (doutorado) – Universidade de Brasília / Departamento de Biologia Celular – Brasília, DF.
- Santos, J. P. R. Dos, Pires, L. P. M., De Castro Vasconcellos, R. C., Pereira, G. S., Von Pinho, R. G., & Balestre, M. (2016). Genomic selection to resistance to *Stenocarpella maydis* in maize lines using DArTseq markers. *BMC Genetics*, 17(1), 1–10. <https://doi.org/10.1186/s12863-016-0392-3>
- Silva-Junior OB, Faria DA, Grattapaglia D. (2015). A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 Eucalyptus tree genomes across 12 species. *New Phytologist* 206: 1527–1540.
- Sousa, L., Ferreira, M. F. da S., MARCIEL, T. E. F. Marcadores Moleculares Baseados em Sequenciamento de Nova Geração. In: MIRANDA, F. D. de (Org.). *Tópicos Especiais em Genética e Melhoramento II*. Alegre - ES: CAUFES 2018. p. 144-173.
- Sukumaran, S., Reynolds, M. P., & Sansaloni, C. (2018). Genome-Wide Association Analyses Identify QTL Hotspots for Yield and Component Traits in Durum Wheat Grown under Yield Potential, Drought, and Heat Stress Environments. *Frontiers in Plant Science*, 9 (February), 1–16. <https://doi.org/10.3389/fpls.2018.00081>
- Tuler, A. C., Carrijo, T. T., Nóia, L. R., Ferreira, A., Peixoto, A. L., & Ferreira, M. F. da S. (2015). SHORT COMMUNICATION SSR markers: a tool for species identification in *Psidium* (Myrtaceae). *Molecular Biology Reports*. <https://doi.org/10.1007/s11033-015-3927-1>
- Tuler, A. C.; Carrijo, T. T.; Peixoto, A. L.; Garbin, M. L.; Ferreira, M. F. da S.; Carvalho, C. R.; Spadeto, M. S.; Clarindo, W. R. (2019). Diversification and geographical distribution of *Psidium* (Myrtaceae) species with distinct ploidy levels. *Trees*. <https://doi.org/10.1007/s00468-019-01845-2>
- Unterseer, S., Bauer, E., Haberer, G., Seidel, M., Knaak, C., Ouzunova, M., ... Schön, C. C. (2014). A powerful tool for genome analysis in maize: Development and evaluation of the high density 600 k SNP genotyping array. *BMC Genomics*, 15(1), 1–15. <https://doi.org/10.1186/1471-2164-15-823>
- Valdisser, P. A. M. R., Pereira, W. J., Almeida Filho, J. E., Müller, B. S. F., Coelho, G. R. C., De Menezes, I. P. P., Vianna, J. P. G., Zucchi, M. I., Lanna, A. C., Coelho, A. S. G., De Oliveira, J. P., Moraes, A. C., Brondani, C. & Vianello, R. P. (2017). In-depth genome characterization of a Brazilian common bean core collection using DArTseq high-density SNP genotyping. *BMC Genomics*, 18(1), 1–19. <https://doi.org/10.1186/s12864-017-3805-4>

5.7 Apêndices

Apêndice 1: Média utilizadas para caracterização de Todos marcadores SNPs (124.069), juntos e separados por espécies.

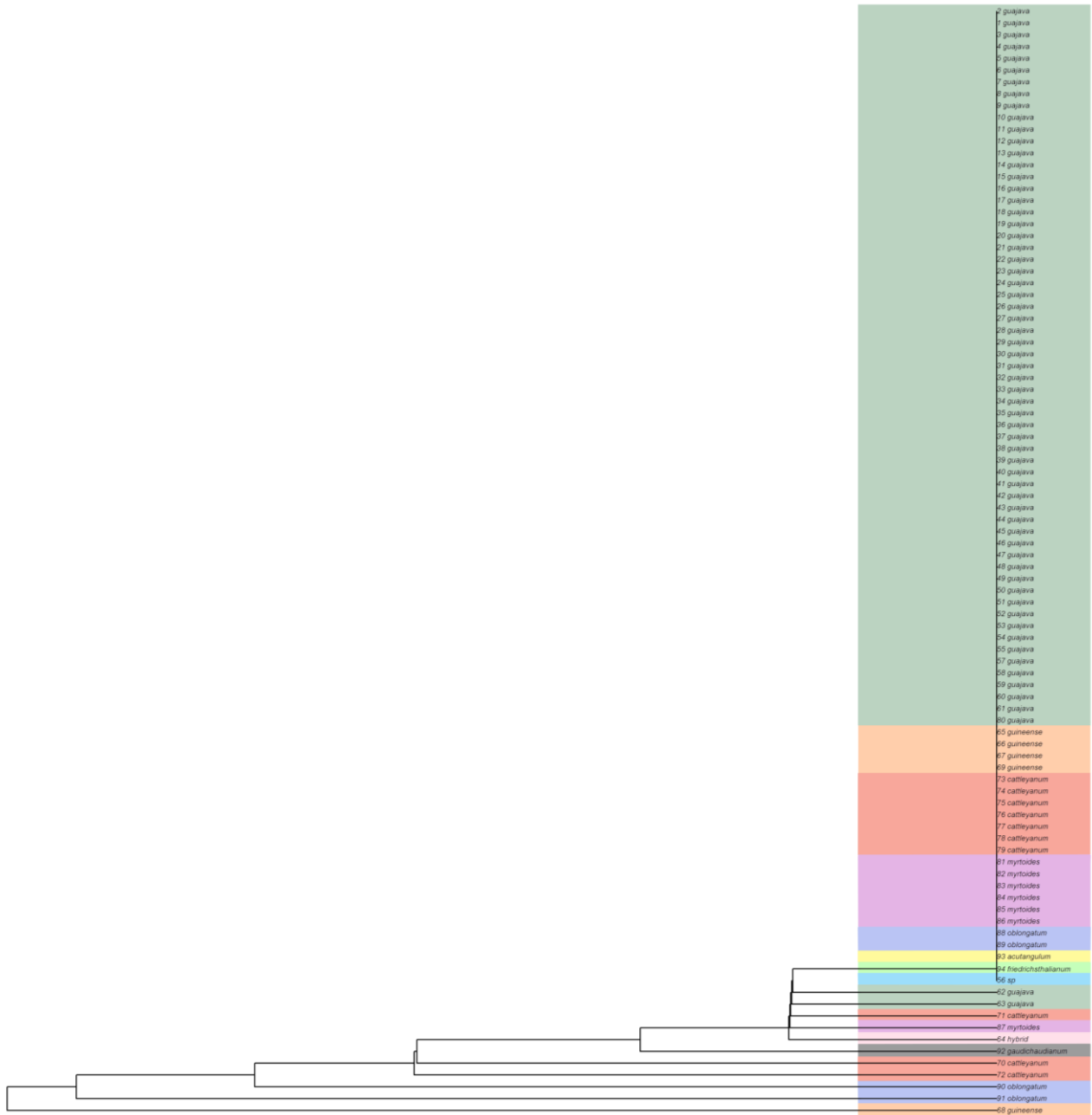
	Nº de Marcadores	Call Rate	OneRatio Ref	OneRatio SNP	FreqHom Ref	FreqHom SNP	Freq Hets	Freq_SNP	Frerq_Ref	PIC Ref	PIC SNP	Avg PIC	AvgCount Ref	AvgCount Snp	Rep Avg
Todos	124,069	0,73	0,78	0,24	0,76	0,22	0,02	0,23	0,77	0,19	0,21	0,20	1673374,27	641007,47	1,00
Todos Funcionais	75,074	0,77	0,82	0,21	0,79	0,18	0,02	0,19	0,81	0,18	0,20	0,19	1694498,28	608884,10	1,00
Todos Aleatórios	48,995	0,67	0,73	0,30	0,70	0,27	0,03	0,29	0,71	0,20	0,21	0,21	1641006,38	690229,42	1,00
<i>P. cattleyanum</i>	94,8099	0,76	0,59	0,61	0,77	0,18	0,05	0,20	0,80	0,28	0,29	0,28	1896109,80	742310,85	0,99
<i>P. cattleyanum</i> Funcionais	58,866	0,83	0,60	0,61	0,81	0,14	0,05	0,17	0,83	0,27	0,28	0,28	1896909,99	701939,41	0,99
<i>P. cattleyanum</i> Aleatórios	35,943	0,64	0,57	0,61	0,71	0,24	0,06	0,27	0,73	0,29	0,29	0,29	1894761,61	810330,72	0,99
<i>P. guajava</i>	113,832	0,75	0,79	0,29	0,68	0,24	0,08	0,28	0,72	0,13	0,19	0,16	1643408,52	925953,39	0,99
<i>P. guajava</i> Funcionais	71,555	0,78	0,82	0,26	0,70	0,22	0,08	0,26	0,74	0,12	0,17	0,15	1612786,08	890916,45	0,99
<i>P. guajava</i> Aleatórios	42,277	0,71	0,75	0,33	0,65	0,27	0,08	0,31	0,69	0,15	0,20	0,18	1681527,88	969567,99	0,99
<i>P. guineense</i>	116,540	0,81	0,73	0,65	0,71	0,23	0,06	0,26	0,74	0,19	0,32	0,25	1868245,98	918632,69	0,99
<i>P. guineense</i> Funcionais	72,590	0,84	0,76	0,67	0,73	0,21	0,06	0,24	0,76	0,17	0,32	0,24	1806986,46	876349,78	0,99
<i>P. guineense</i> Aleatórios	43,950	0,77	0,67	0,63	0,69	0,25	0,07	0,28	0,72	0,23	0,32	0,28	1976775,05	993542,28	0,99
<i>P. myrtoides</i>	63,611	0,82	0,71	0,77	0,75	0,17	0,08	0,21	0,79	0,23	0,21	0,22	1769553,70	733236,11	0,99
<i>P. myrtoides</i> Funcionais	39,311	0,87	0,71	0,78	0,78	0,14	0,08	0,18	0,82	0,23	0,20	0,22	1749300,66	687117,57	0,99
<i>P. myrtoides</i> Aleatórios	24,300	0,71	0,70	0,75	0,68	0,22	0,10	0,27	0,73	0,23	0,21	0,22	1810610,50	826727,26	0,99
<i>P. oblongatum</i>	68,129	0,78	0,61	0,65	0,71	0,24	0,05	0,27	0,73	0,30	0,29	0,30	1887187,51	879865,97	0,99
<i>P. oblogatum</i> Funcionais	34,598	0,85	0,60	0,67	0,77	0,18	0,05	0,21	0,79	0,30	0,27	0,28	1910834,47	801959,21	0,99
<i>P. oblongatum</i> Aleatórios	33,531	0,65	0,61	0,60	0,61	0,34	0,04	0,36	0,64	0,31	0,32	0,31	1847028,90	1012171,60	0,99

Apêndice 2: Médias utilizadas para caracterização dos marcadores SNPs Ancorados ao genoma de *Eucalyptus* (22.050), juntos e separados por espécies.

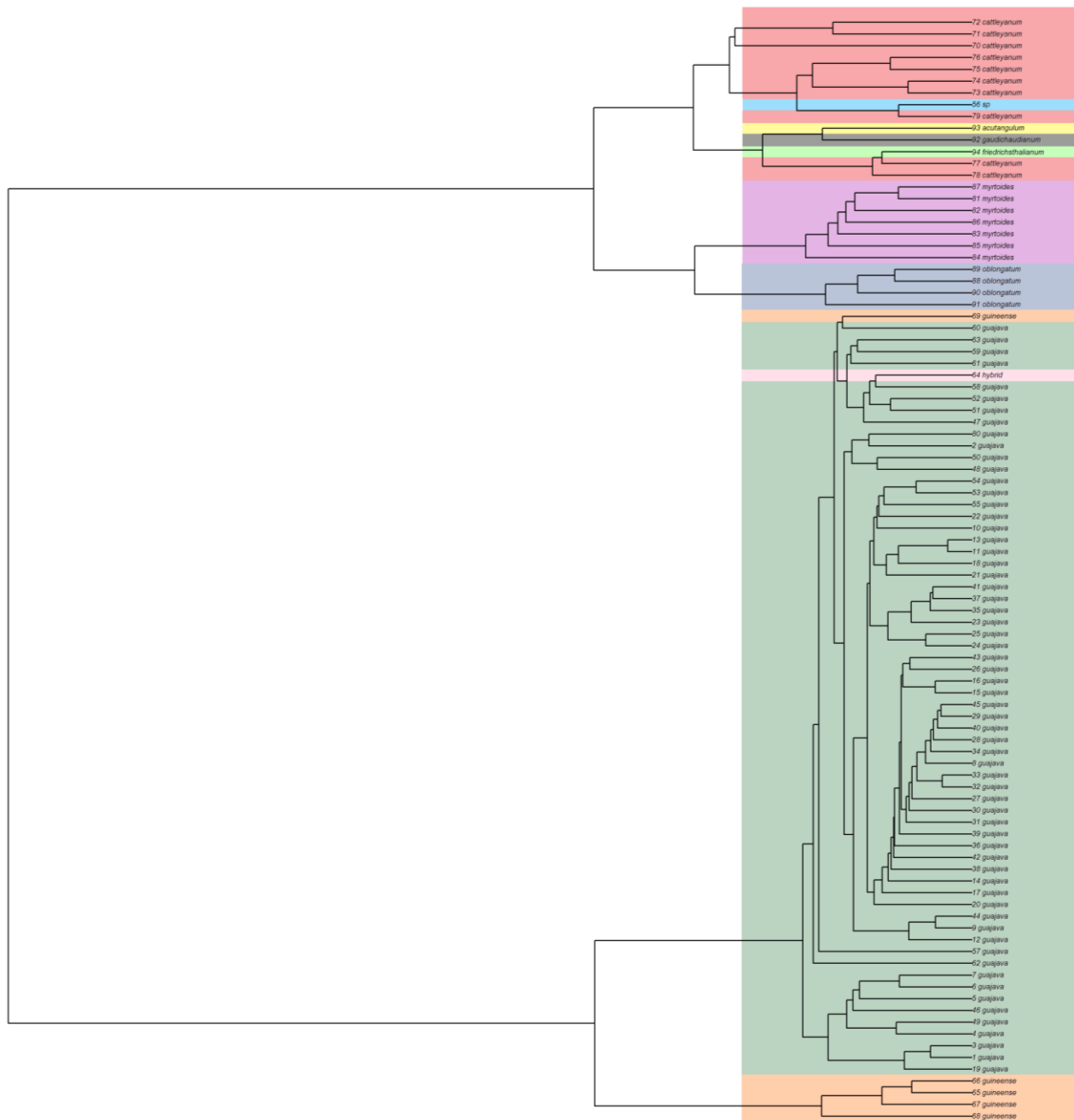
	Nº de Marcadores	Call Rate	OneRatio Ref	OneRatio Snp	FreqHom Ref	FreqHom SNP	Freq Hets	Freq_SNP	Freq_ Ref	PIC Ref	PIC SNP	Avg PIC	AvgCount Ref	AvgCount SNP	Rep Avg
Todos	22,050	0,79	0,80	0,22	0,78	0,20	0,03	0,21	0,79	0,18	0,21	0,19	1545437,68	590820,22	1,00
Todos Funcionais	18,211	0,79	0,80	0,22	0,78	0,20	0,03	0,21	0,79	0,18	0,21	0,19	1543466,40	590432,21	1,00
Todos Aleatórios	3,839	0,74	0,79	0,23	0,77	0,21	0,02	0,22	0,78	0,18	0,21	0,20	1562202,05	594119,96	1,00
<i>P. cattleyanum</i>	18,709	0,85	0,65	0,59	0,80	0,14	0,06	0,17	0,83	0,25	0,29	0,27	1719980,69	663031,75	0,99
<i>P. cattleyanum</i> Funcionais	16,817	0,86	0,65	0,59	0,80	0,14	0,06	0,17	0,83	0,25	0,28	0,26	1697745,85	660764,91	0,99
<i>P. cattleyanum</i> Aleatórios	1,892	0,77	0,62	0,59	0,80	0,15	0,05	0,18	0,82	0,29	0,30	0,29	1930156,04	684459,05	0,99
<i>P. guajava</i>	20,744	0,84	0,83	0,24	0,72	0,21	0,08	0,25	0,75	0,10	0,15	0,12	1504985,98	823286,47	0,99
<i>P. guajava</i> Funcionais	18,604	0,84	0,83	0,24	0,71	0,21	0,08	0,25	0,75	0,10	0,14	0,12	1463704,15	801498,09	0,99
<i>P. guajava</i> Aleatórios	2,140	0,80	0,83	0,24	0,72	0,20	0,08	0,24	0,76	0,11	0,17	0,14	1818249,66	988625,69	0,99
<i>P. guineense</i>	21,287	0,87	0,82	0,68	0,71	0,23	0,06	0,26	0,74	0,14	0,30	0,22	1706848,48	862057,30	0,99
<i>P. guineense</i> Funcionais	19,112	0,87	0,84	0,69	0,71	0,23	0,06	0,26	0,74	0,13	0,30	0,22	1667957,36	841482,69	0,99
<i>P. guineense</i> Aleatórios	2,175	0,84	0,72	0,65	0,71	0,24	0,05	0,26	0,74	0,20	0,31	0,25	2045036,62	1040969,25	0,99
<i>P. myrtoides</i>	13,575	0,89	0,75	0,77	0,77	0,14	0,09	0,18	0,82	0,22	0,21	0,21	1606831,99	682102,62	0,99
<i>P. myrtoides</i> Funcionais	12,286	0,89	0,74	0,77	0,77	0,14	0,09	0,18	0,82	0,22	0,21	0,21	1576655,96	653068,85	0,99
<i>P. myrtoides</i> Aleatórios	1,289	0,84	0,78	0,75	0,77	0,13	0,09	0,18	0,82	0,22	0,22	0,22	1937906,07	1000644,61	0,99
<i>P. oblongatum</i>	12,110	0,88	0,65	0,66	0,78	0,17	0,06	0,20	0,80	0,27	0,27	0,27	1795163,02	758010,96	0,99
<i>P. oblogatum</i> Funcionais	10,971	0,89	0,66	0,66	0,78	0,16	0,06	0,19	0,81	0,26	0,27	0,27	1680605,24	674954,84	0,99
<i>P. oblongatum</i> Aleatórios	1,139	0,82	0,65	0,64	0,73	0,22	0,05	0,25	0,75	0,30	0,29	0,30	3098862,87	1703213,03	0,99

Apêndice 3: Médias utilizadas para caracterização dos marcadores SNPs Não Ancorados ao genoma de *Eucalyptus* (102.019), juntos e separados por espécies.

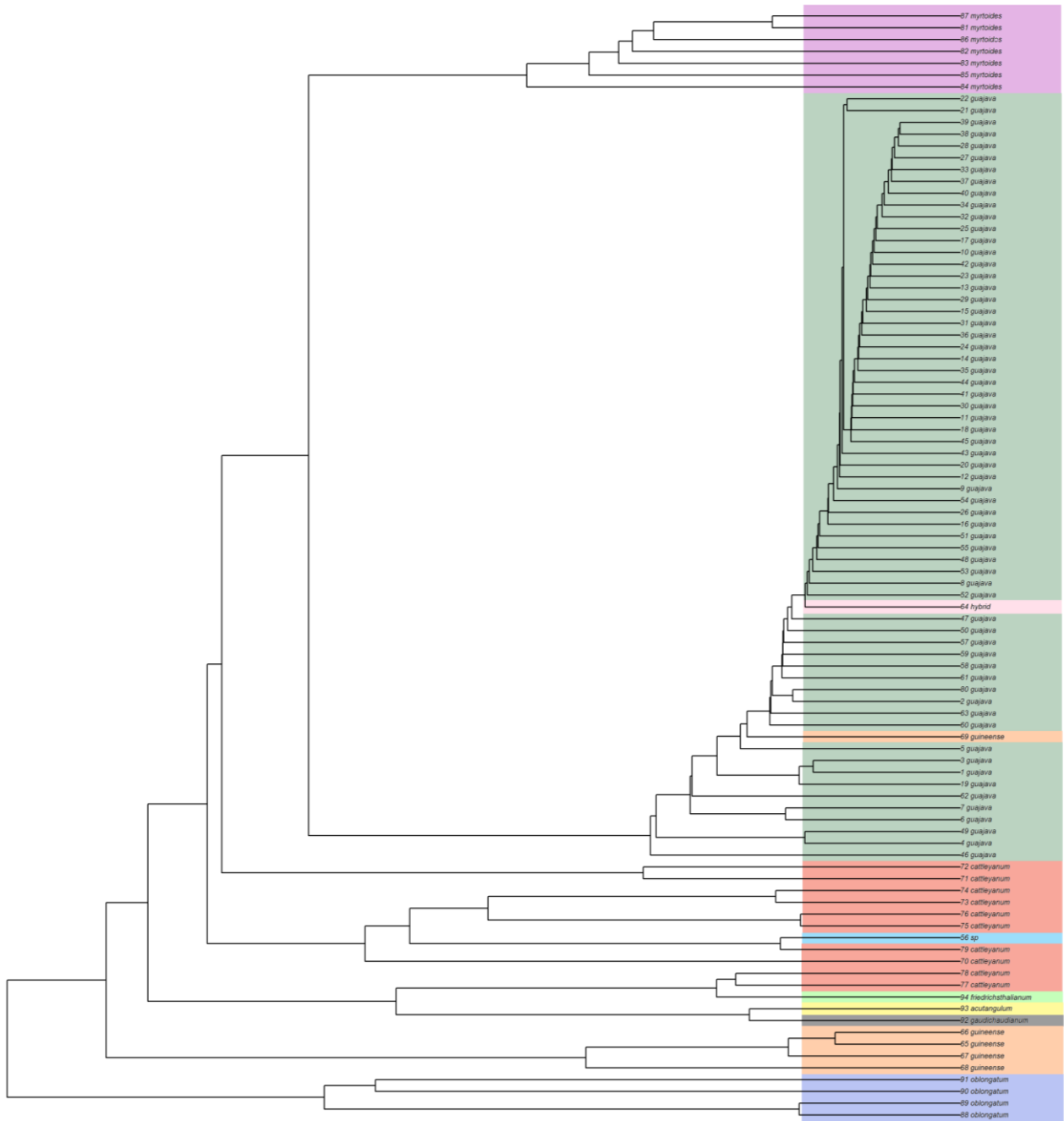
	Nº de Marcadores	Call Rate	OneRatio Ref	OneRatio SNP	FreqHom Ref	FreqHom SNP	Freq Hets	Freq_SNP	Freq_Rref	PIC Ref	PIC SNP	Avg PIC	AvgCount Ref	AvgCount SNP	Rep Avg
Todos	102,019	0,72	0,78	0,25	0,75	0,22	0,02	0,23	0,77	0,19	0,21	0,20	1701025,99	651854,75	1,00
Todos Funcionais	56,863	0,76	0,82	0,20	0,80	0,18	0,02	0,19	0,81	0,18	0,20	0,19	1748340,78	615462,15	1,00
Todos Aleatórios	45,156	0,67	0,72	0,30	0,70	0,28	0,03	0,29	0,71	0,20	0,21	0,21	1644923,38	695006,58	1,00
<i>P. cattleyanum</i>	76,100	0,73	0,57	0,62	0,76	0,19	0,05	0,21	0,79	0,29	0,29	0,29	1950115,50	766619,85	0,99
<i>P. cattleyanum</i> Funcionais	42,049	0,81	0,57	0,62	0,81	0,14	0,04	0,17	0,83	0,28	0,28	0,28	1998470,53	722945,57	0,99
<i>P. cattleyanum</i> Aleatórios	34,051	0,63	0,57	0,61	0,70	0,24	0,06	0,27	0,73	0,29	0,29	0,29	1892723,66	818456,17	0,99
<i>P. guajava</i>	93,088	0,73	0,78	0,30	0,67	0,25	0,08	0,29	0,71	0,14	0,19	0,17	1667740,00	943999,86	0,99
<i>P. guajava</i> Funcionais	52,951	0,76	0,82	0,27	0,70	0,22	0,08	0,26	0,74	0,13	0,18	0,16	1659520,81	918872,97	0,99
<i>P. guajava</i> Aleatórios	40,137	0,70	0,75	0,34	0,65	0,27	0,08	0,31	0,69	0,15	0,20	0,18	1675853,20	968802,73	0,99
<i>P. guineense</i>	95,253	0,79	0,70	0,65	0,71	0,22	0,06	0,26	0,74	0,21	0,32	0,26	1912353,19	934093,79	0,99
<i>P. guineense</i> Funcionais	53,478	0,82	0,73	0,66	0,74	0,21	0,06	0,24	0,76	0,19	0,32	0,26	1866895,83	891374,44	0,99
<i>P. guineense</i> Aleatórios	41,775	0,76	0,66	0,63	0,68	0,25	0,07	0,28	0,72	0,23	0,32	0,28	1972313,20	990442,26	0,99
<i>P. myrtoides</i>	50,036	0,79	0,69	0,77	0,73	0,18	0,08	0,22	0,78	0,23	0,21	0,22	1833552,82	753347,12	0,99
<i>P. myrtoides</i> Funcionais	27,025	0,86	0,69	0,78	0,78	0,15	0,07	0,19	0,81	0,24	0,20	0,22	1857981,92	708551,52	0,99
<i>P. myrtoides</i> Aleatórios	23,011	0,70	0,70	0,75	0,68	0,23	0,10	0,27	0,73	0,23	0,21	0,22	1800825,70	813358,79	0,99
<i>P. oblongatum</i>	56,019	0,74	0,59	0,64	0,69	0,27	0,04	0,29	0,71	0,31	0,30	0,30	1917370,16	919832,59	0,99
<i>P. oblogatum</i> Funcionais	23,627	0,83	0,57	0,68	0,76	0,19	0,04	0,22	0,78	0,32	0,27	0,29	2041672,37	874231,39	0,99
<i>P. oblongatum</i> Aleatórios	32,392	0,64	0,61	0,60	0,61	0,35	0,04	0,37	0,63	0,31	0,32	0,32	1774955,06	972078,63	0,99



Apêndice 4. Agrupamento obtido através dos marcadores SNPs com valores de PIC menores que 0.02, calculados para 94 amostras de *Psidium*. Espécies identificadas por cor: 63 *P. guajava* (Verde escuro), 05 *P. guineense* (Laranja), 01 Híbrido (Rosa), 04 *P. oblongatum* (Azul escuro), 07 *P. myrtoides* (Lilás), 01 *P. friedrichsthalianum* (Verde claro), 10 *P. cattleyanum* (Vermelho), 01 *P. gaudichaudianum* (Preto), 01 *P. acutangulum* (Amarelo) e 01 *Psidium sp.* (Azul claro).



Apêndice 5. Agrupamento obtido através dos marcadores SNPs com valores de PIC maiores que 0.2, calculados para 94 amostras de *Psidium*. Espécies identificadas por cor: 63 *P. guajava* (Verde escuro), 05 *P. guineense* (Laranja), 01 Híbrido (Rosa), 04 *P. oblongatum* (Azul escuro), 07 *P. myrtoides* (Lilás), 01 *P. friedrichsthalianum* (Verde claro), 10 *P. cattleyanum* (Vermelho), 01 *P. gaudichaudianum* (Preto), 01 *P. acutangulum* (Amarelo) e 01 *Psidium* sp. (Azul claro).



Apêndice 6. Agrupamento obtido através dos marcadores SNPs com valores de PIC intermediários (menores que 0,02 e maiores que 0,2), calculados para 94 amostras de *Psidium*. Espécies identificadas por cor: 63 *P. guajava* (Verde escuro), 05 *P. guineense* (Laranja), 01 Híbrido (Rosa), 04 *P. oblongatum* (Azul escuro), 07 *P. myrtoides* (Lilás), 01 *P. friedrichsthalianum* (Verde claro), 10 *P. cattleyanum* (Vermelho), 01 *P. gaudichaudianum* (Preto), 01 *P. acutangulum* (Amarelo) e 01 *Psidium* sp. (Azul claro).