



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
DEPARTAMENTO DE AGRONOMIA
PÓS-GRADUAÇÃO EM AGRONOMIA – MELHORAMENTO GENÉTICO DE
PLANTAS

FLÁVIA LAYSE BELÉM MEDEIROS

DESENVOLVIMENTO DE MAPAS GENÉTICOS E FÍSICOS
COM BASE EM SNPs EM *Psidium guajava* E TRIAGEM *in*
***silico* DE LOCI MICROSSATÉLITES DE GOIABEIRA PARA**
TRANSFERIBILIDADE EM *Eucalyptus globulus*

RECIFE – PE
2021

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO

**DESENVOLVIMENTO DE MAPAS GENÉTICOS E FÍSICOS
COM BASE EM SNPs EM *Psidium guajava* E TRIAGEM *in
silico* DE LOCI MICROSSATÉLITES DE GOIABEIRA PARA
TRANSFERIBILIDADE EM *Eucalyptus globulus***

DISSERTAÇÃO DE MESTRADO

FLÁVIA LAYSE BELÉM MEDEIROS

**RECIFE – PE
2021**

FLÁVIA LAYSE BELÉM MEDEIROS

**DESENVOLVIMENTO DE MAPAS GENÉTICOS E FÍSICOS COM BASE EM SNPs
EM *Psidium guajava* E TRIAGEM *in silico* DE LOCI MICROSSATÉLITES DE
GOIABEIRA PARA TRANSFERIBILIDADE EM *Eucalyptus globulus***

Dissertação apresentada ao Programa de Pós-Graduação em Agronomia – Melhoramento Genético de Plantas da Universidade Federal Rural de Pernambuco como requisito para a obtenção do grau de Mestre em Melhoramento Genético de Plantas.

ORIENTADOR: DR. CARLOS ANTÔNIO FERNANDES SANTOS

**RECIFE – PE
2021**

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

- M488d Medeiros, Flávia Layse Belém
Desenvolvimento de mapas genéticos e físicos com base em SNPs em *Psidium guajava* e triagem in silico de loci microssatélites de goiabeira para transferibilidade em *Eucalyptus globulus* / Flávia Layse Belém Medeiros. - 2021.
117 f. : il.
- Orientador: Carlos Antonio Fernandes Santos.
Inclui referências, apêndice(s) e anexo(s).
- Dissertação (Mestrado) - Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em Agronomia - Melhoramento Genético de Plantas, Recife, 2021.
1. Goiabeira. 2. Triagem in silico. 3. Mapa genético. 4. Mapa físico. 5. JoinMap e GACD. I. Santos, Carlos Antonio Fernandes, orient. II. Título

CDD 581.15

FLÁVIA LAYSE BELÉM MEDEIROS

**DESENVOLVIMENTO DE MAPAS GENÉTICOS E FÍSICOS COM BASE EM
SNPs EM *Psidium guajava* E TRIAGEM *in silico* DE LOCI MICROSSATÉLITES
DE GOIABEIRA PARA TRANSFERIBILIDADE EM *Eucalyptus globulus***

Dissertação defendida e aprovada pela banca examinadora em: 18/08/2021

ORIENTADOR:

Dr. Carlos Antônio Fernandes Santos - Embrapa Semiárido

EXAMINADORES:

Prof. Dr. Frederico Inácio da Silva

Dr. Rômulo Maciel de Moraes Filho

DEDICATÓRIA

Dedico este trabalho a todos que acreditam que a ciência é o caminho para o futuro e para um desenvolvimento sustentável, e que acreditaram em mim como parte desse processo.

AGRADECIMENTOS

Gostaria de agradecer primeiramente a Deus por proporcionar saúde a todos que eu amo em tempos de pandemia do COVID-19. Não faltaram obstáculos, mas a fé em dias melhores me moveu nos momentos de dificuldade.

À minha mãe Eliane Mara Belém Medeiros por me apoiar em todos os meus sonhos e acreditar na minha capacidade de realizá-los e ao meu irmão Fernando Antônio Medeiros da Silva Júnior pelo suporte, incentivo e exemplo de vida. À toda minha família: minha cunhada Marina Falcão Barbosa, meus pais de consideração Wellington César Gaia e Valdeliz Duete Gaia e todos que direta ou indiretamente me apoiaram na jornada acadêmica.

Ao meu orientador Dr. Carlos Antônio Fernandes Santos pelo grande exemplo de profissional. Agradeço por estar sempre presente e estimular os alunos a atingirem seu máximo potencial.

Aos colegas do programa de pós-graduação em Melhoramento Genético de Plantas e do Laboratório de Genética Vegetal (Embrapa Semiárido) Cinthia, Nina, Jamile, Deisy, Andressa, Antônio, e ao técnico Carlos Santos, por todas as histórias e conhecimento divididos. A jornada acadêmica é bem mais leve com vocês.

Aos amigos Wesley Albuquerque Maranhão, Jessyca Maria Gomes Florêncio, Ellienay Simony Wyatt pelo incentivo e ombro amigo nos momentos mais difíceis, especialmente Carla Morta Leal por ser uma irmã de alma e por todos os conselhos. Sem vocês esse trabalho não seria possível.

Aos professores da Universidade Federal Rural de Pernambuco, grandes mentores e exemplos de profissão. Espero um dia chegar a ponto de formar jovens profissionais da mesma forma que fui formada pela casa.

À Universidade Federal Rural de Pernambuco pela oportunidade de participar do programa de pós-graduação e proporcionar momentos valiosos que vão além do currículo.

À Embrapa Semiárido pela estrutura oferecida para o desenvolvimento da pesquisa.

Ao CNPq pela concessão da bolsa.

LISTA DE FIGURAS

CAPÍTULO I

- Figura 1** Produção de goiaba nos últimos 10 anos no Brasil: área plantada em hectares, quantidade produzida em toneladas e valor de produção. 19
- Figura 2** Popularidade de diferentes marcadores moleculares nos últimos 40 anos. Fonte: Grover A, Sharma PC (2016). 23
- Figura 3** Região microssatélite em indivíduo homozigoto (A) e heterozigoto (B). Fonte: Ferreira ME, Grattapaglia D (1998). 25
- Figura 4** Diagrama representando o polimorfismo de um único nucleotídeo (SNP). 26
- Figura 5** Grupos de ligação 1 a 4 em *P. guajava* provenientes do mapeamento de ligação utilizando marcadores SSR e AFLP. Fonte: Lepitre *et al.* (2010). 32
- Figura 6** Exemplo de alinhamento de SSR mPgCIR001 de *P. guajava* no genoma de *E. grandis*, cromossomo 6, obtido pelo BLAST (Altschul *et al.*, 1997): bit score – 164, e-value – 73 e-81 e identidade – 82% (303/368). 39

CAPÍTULO II

- Figura 1** Mapa de ligações genéticas do cruzamento Pedro Sato × Roxa (PSR) realizado pelo JoinMap 4.0 (Van Ooijen 2006). 60
- Figura 2** Mapa de ligações genéticas do cruzamento Pedro Sato × Roxa (PSR) realizado pelo GACD (Zhang *et al.* 2015). 61
- Figura 3** Mapa de físico dos cromossomos 1 ao 6 do cruzamento Pedro Sato × Roxa (PSR) realizado através do BLAST de SNPs de *Eucalyptus* contra o genoma da goiaba e alinhado no programa MapGene2Chrom (Jiangtao *et al.* 2015). 65
- Figura 4** Mapa de físico dos cromossomos 7 ao 11 do cruzamento Pedro Sato × Roxa (PSR) realizado através do BLAST de SNPs de *Eucalyptus* contra o genoma da goiaba e alinhado no programa MapGene2Chrom (Jiangtao *et al.* 2015). 66

CAPÍTULO III

- Figura 1.** Resolução de amplicons de locus SSR de *Psidium guajava* em quatro plantas de *Eucalyptus deglupta*. 87

LISTA DE TABELAS

CAPÍTULO II

Tabela 1 Comprimento dos grupos de ligação, número de marcadores/grupo de ligação, número de marcadores codominantes nos mapas de ligações do cruzamento Pedro Sato × Roxa (PSR) feitos pelos programas JoinMap 4.0 × GACD.....	62
Tabela 2 Comprimento total dos cromossomos de <i>Psidium guajava</i> reportado por Feng <i>et al.</i> (2020), cobertura do mapa físico e número de marcadores SNPs de <i>Eucalyptus</i> por cromossomo de <i>P. guajava</i>	64
Tabela 3 Número de SNPs por cromossomo/grupo de ligação nos diferentes mapas genéticos e físico de <i>P. guajava</i>	67
Tabela 4 Comparações do mapa de ligação JoinMap 4.0 × mapa físico de <i>P. guajava</i>	68
Tabela 5 Comparações do mapa de ligação GACD × mapa físico de <i>P. guajava</i>	69

CAPÍTULO III

Tabela 1 Locus SSR, número de acesso EMBL, tamanho do clone e tamanho dos alelos das 23 sequências SSR de <i>Psidium guajava</i> identificadas por Risterucci <i>et al.</i> (2005).	82
Tabela 2. Score, e-value e identidade de nove sequências de SSR (Risterucci <i>et al.</i> 2005) de <i>P. guajava</i> , em cromossomos de <i>Eucalyptus globulus</i>	85
Tabela 3 Score, e-value, identidade (ID) e distância entre primers (bp) de três sequências de SSR do projeto GUAVAMAP de <i>P. guajava</i> , em cromossomos de <i>Eucalyptus globulus</i> (Chr).	86

LISTA DE ANEXOS

ANEXO I - RESULTADOS DO BLAST DE SEQUÊNCIAS DE SNPs DE EUCALIPTO CONTRA O GENOMA DA GOIABEIRA (*P. guajava*).

Tabela 1 Resultados BLAST das sequências de SNPs utilizadas no mapa físico de *P. guajava*.
.....95

ANEXO II - BLAST DE SEQUÊNCIAS DE MICROSSATÉLITES DE *P. guajava* EM *Eucalyptus globulus*

Fig 1 Espécie utilizada inserida na busca em todos os bancos de dados do site.....	111
Fig 2 Seleção da espécie de interesse.....	111
Fig 3 Adição do “accession number” e utilização da opção MEGABLAST.....	112
Fig 4 Resumo das informações dos hits encontrados, como query ID, molecular type e length of the query.....	112
Fig 5 Resultados fornecidos pela BLAST Search.....	113
Fig 6 Alinhamento dos nucleotídeos do marcador mPgCIR001 contra genoma de <i>E. globulus</i>	113
Fig 7 Alinhamento dos nucleotídeos do marcador mPgCIR005 contra genoma de <i>E. globulus</i>	114
Fig 8 Alinhamento dos nucleotídeos do marcador mPgCIR007 contra genoma de <i>E. globulus</i>	114
Fig 9 Alinhamento dos nucleotídeos do marcador mPgCIR009 contra genoma de <i>E. globulus</i>	115
Fig 10 Alinhamento dos nucleotídeos do marcador mPgCIR011 contra genoma de <i>E. globulus</i>	115
Fig 11 Alinhamento dos nucleotídeos do marcador mPgCIR018 contra genoma de <i>E. globulus</i>	116
Fig 12 Alinhamento dos nucleotídeos do marcador mPgCIR020 contra genoma de <i>E. globulus</i>	116
Fig 13 Alinhamento dos nucleotídeos do marcador mPgCIR021 contra genoma de <i>E. globulus</i>	117
Fig 14 Alinhamento dos nucleotídeos do marcador mPgCIR026 contra genoma de <i>E. globulus</i>	117

SUMÁRIO

RESUMO.....	XII
ABSTRACT	XIII
CAPÍTULO I – Considerações gerais.....	14
1. INTRODUÇÃO GERAL	15
2. REVISÃO DE LITERATURA.....	16
2.1 Taxonomia, origens e aspectos botânicos da espécie <i>Psidium guajava</i> L.	16
2.2 Composição química e utilização da <i>P. guajava</i>	17
2.3 Importância econômica.....	18
2.4 Melhoramento genético da goiabeira	20
2.5 Marcadores moleculares	21
2.6 Marcadores microssatélites e SNPS e suas aplicações na goiabeira	24
2.7 Mapeamento genético aplicado no melhoramento genético de plantas	27
2.8 Análises de QTL e suas aplicações em <i>P. guajava</i> L.....	33
2.9 Transferibilidade de locus SSR	35
2.10. Genômica comparativa e Alinhamento local (BLAST)	37
3 REFERÊNCIAS BIBLIOGRÁFICAS	40
CAPÍTULO II – Desenvolvimento de mapas genéticos e físicos com base em SNPs em <i>Psidium guajava</i>	52
RESUMO.....	53
ABSTRACT	53
INTRODUÇÃO	54
MATERIAIS E MÉTODOS	56
RESULTADOS	59
DISCUSSÃO	69
CONCLUSÃO.....	72
REFERÊNCIAS BIBLIOGRÁFICAS	72
CAPÍTULO III – Triagem <i>in silico</i> de loci microssatélites de goiabeira para transferibilidade em <i>Eucalyptus globulus</i>.....	78
RESUMO.....	79
ABSTRACT	79
INTRODUÇÃO	80
MATERIAIS E MÉTODOS	81
RESULTADOS	85
DISCUSSÃO	88

CONCLUSÃO	90
REFERÊNCIAS BIBLIOGRÁFICAS	90
ANEXOS	94
ANEXO I – RESULTADOS DO BLAST DE SEQUÊNCIAS DE SNPs DE EUCALIPTO CONTRA O GENOMA DA GOIABEIRA (<i>P. guajava</i>)	95
ANEXO II - BLAST DE SEQUÊNCIAS DE MICROSSATÉLITES DE <i>P. guajava</i> EM <i>Eucalyptus globulus</i>	111

RESUMO

A goiabeira (*Psidium guajava*) é conhecida mundialmente pela facilidade de manejo, riqueza nutricional e versatilidade em consumo. Apesar de seu mercado em expansão, ainda é uma espécie pouco estudada do ponto de vista molecular, sendo importante o fomento de estudos que possibilitem a aplicação de seleção assistida por marcadores na espécie. Neste sentido, este trabalho teve com objetivos: 1) desenvolver e comparar mapas genéticos e físicos com base em SNPs de *Eucalyptus* para *P. guajava* e 2) propor a utilização da ferramenta BLAST para a triagem *in silico* de sequências microssatélites (SSR) de *P. guajava* para *E. globulus*, de forma a orientar estudos de transferibilidade. O material vegetal utilizado de goiabeira foi oriundo do cruzamento das cultivares Pedro Sato × Goiabeira Roxa (PSR), disponível na Embrapa Semiárido, Petrolina, PE. Folhas novas e sadias de ambas as espécies foram coletadas, devidamente identificadas, e acondicionadas em isopor para transporte ao laboratório, onde as extrações de DNA foram realizadas de acordo com o protocolo CTAB 2x modificado com uma pré-lavagem de sorbitol. A quantificação do DNA foi realizada por espectrofotometria, utilizando a placa Microdrop na faixa de absorbância a 260 nm. Após quantificação, as amostras foram diluídas para concentração de trabalho de 60 ng/μL (*P. guajava*) e 20 ng/μL (*E. deglupta*) e armazenadas a -20°C. A genotipagem de 112 progênies da população de PSR foi efetuada com o Euchip60K de *Eucalyptus*. A análise de ligação realizada no JoinMap 4.0 considerou um LOD score de 5,0 a 12, enquanto no programa GACD considerou LOD de 8,0. O mapa físico foi ordenado em cromossomos da goiabeira considerando o alinhamento de sequências dos SNPs de *Eucalyptus* (query) e o genoma da goiabeira (subject), estimando as posições dos marcadores através do Blastn, com e-value < E⁻¹⁰. O JoinMap 4.0 gerou um mapa de 1.405,2 cM e o mapa GACD 1.392,7 cM. Os mapas gerados apresentaram grupos com segmentos de vários cromossomos, quando comparados com o mapa físico, indicando limitações. O GACD apresentou maior limitação em relação ao JoinMap 4.0, ao separar os marcadores de acordo com sua origem parental. O mapa físico gerado com Blastn, e-values variando de 8xE⁻¹⁰ a 1,15xE⁻²⁶, cobriu 434,88 Mb, com distâncias médias de 0,62 Mb, sendo referência para estudos de mapeamento e estimativas de QTLs em goiabeira. Para a triagem *in silico* foram alinhadas sequências de 23 clones de SSR para *P. guajava* (query) contra o genoma completo de *E. globulus* (subject.) através do Blastn otimização MEGABLAST e em seguida selecionadas aquelas com e-value < 1e⁻²⁰. Foram avaliados primers de 140 loci disponibilizados pelo projeto GuavaMap, com base no e-value < 1.7 e proximidade entre as sequências forward e reverse (máximo de 300 nucleotídeos de distância). Das 23 sequências avaliadas 39% apresentaram alinhamentos significativos, com identidade média das sequências de 87%. Dos loci GuavaMap somente três, 2,1%, apresentaram alinhamentos significativos. Os loci SSR com resultados significativos no BLAST amplificaram em gel de poliacrilamida, enquanto os loci que não apresentaram “hits” ou e-value no Blastn também não produziram amplicons *in vitro*. Desta forma, a triagem *in silico* demonstrou ser efetiva para orientar estudos genômicos entre as duas espécies, reduzindo custos e tempo.

Palavras-chave: goiabeira, triagem *in silico*, mapa genético, mapa físico, JoinMap, GACD

ABSTRACT

Guava tree (*Psidium guajava*) is worldwide known for its easy cultivation, nutritional richness and consumption versatility. Although this crop is expanding commercially, it is still little studied from a molecular point of view. Thus, it is important to promote studies that enable marker assisted selection in guava. Given that, this study aimed to 1) develop and compare *P. guajava* genetic and physical maps based on SNPs from *Eucalyptus* and 2) propose BLAST for *in silico* screening of microsatellite sequences (SSR) from *P. guajava* to *E. globulus* in order to guide transferability studies. Guava's plant material was originated from the cross between Pedro Sato × Goiabeira Roxa (PSR), available at Embrapa Semiárido, Petrolina – PE, Brazil. Young and healthy leaves, from both species, were collected, properly identified, and put in styrofoam for transportation to the laboratory where samples were extracted with a modified CTAB 2x protocol, including a sorbitol pre-wash. DNA quantification was performed by spectrophotometry, using a Microdrop plate in the absorbance range of 260nm. After quantification, samples were diluted to work concentration of 60 ng/μL (*P. guajava*) e 20 ng/μL (*E. deglupta*) and stored at -20°C. Genotyping of the 112 PSR was performed with Euchip60K from *Eucalyptus*. Linkage analysis was performed using JoinMap 4.0, LOD score from 5 to 12 meanwhile GACD analysis a LOD score of 8. A physical map was ordered in guava chromosomes considering *Eucalyptus* SNPs (query) alignment to guava's genome (subject), estimating marker position via Blastn with a cut off e-value < E⁻¹⁰. JoinMap 4.0 generated a map of 1.405,2 cM and GACD a map of 1.392,7 cM. Both maps presented linkage groups with segments from several chromosomes when compared to the physical map, indicating limitations. GACD showed a greater limitation in comparison to JoinMap 4.0 because it divides markers according to their parental origin. The physical map generated with Blastn, e-values ranging from 8xE⁻¹⁰ to 1.15xE⁻²⁶ covered 434. 88Mb, with mean distances of 0.6Mb, being a reference for mapping and QTL estimations in guava. For the *in silico* screening 23 *P. guajava* SSR clones (query) were aligned against *E. globulus* genome (subject), using Blastn and e-value < 1E⁻²⁰. Also 140 primers made available by GuavaMap were analyzed based on e-value < 1.7 and the distance between forward and reverse sequences (a maximum of 300 nucleotides distance). 39% of the 23 SSR clones showed significant alignments with sequences mean identity of 87%. Regarding GuavaMap loci only three, 2.1%, presented significant alignments. SSR loci with significant BLAST results also amplified in polyacrylamide gel meanwhile the loci that did not show “hits” or e-value on Blastn did not produce amplicons *in vitro*. Therefore, *in silico* screening showed to be effective in reducing costs, time and in orienting transferability studies between these species.

Keywords: guava, *in silico* screening, genetic map, physical map, JoinMap, GACD.

CAPÍTULO I

CONSIDERAÇÕES GERAIS

1. INTRODUÇÃO GERAL

A espécie *Psidium guajava*, conhecida popularmente como goiabeira, é uma das culturas de destaque na fruticultura brasileira. Tem como centro de origem a América do Sul (Risterucci *et al.* 2005). Pertence à família Myrtaceae, a qual se destaca pela riqueza de biodiversidade e por abrigar outras culturas de importância social e econômica, como o eucalipto (*Eucalyptus spp.*). O gênero *Psidium* contém 64 espécies no Brasil, sendo 38 delas na Mata Atlântica (Flora do Brasil 2020), que podem ser importantes para introgressão de alelos de resistência a certas pragas e doenças na goiabeira (Grossi *et al.* 2021)

A goiaba se destaca pelo alto valor nutricional, a exemplo do teor de vitamina C, que chega ser 6x maior do que em frutos cítricos (Barbosa e Lima 2010), e altos níveis de compostos fenólicos e carotenoides, sendo popularmente referida como “Super fruit” (Chang *et al.* 2019). Do ponto de vista econômico *P. guajava* se sobressai pela sua disponibilidade o ano inteiro, preços acessíveis, aptidão para transporte, manuseio pós-colheita e o mais importante: a preferência do consumidor (Nimisha *et al.* 2013). No mundo Índia, China, Tailândia, Paquistão e México são produtores de goiaba, sendo o Brasil o décimo maior produtor mundial da cultura (Masud Parvez *et al.* 2018, FAOSTAT 2020). A produção brasileira atingiu aproximadamente 584 mil toneladas de frutos de acordo com o Censo Agropecuário (IBGE 2019), com maior produção em São Paulo, Pernambuco, Bahia e Ceará.

Apesar de sua importância econômica a goiabeira ainda carece de estudos para otimizar os programas de melhoramento genético, visto que o melhoramento de planta perene tende a ser mais complexo, devido ao longo período juvenil da planta. O desenvolvimento de ferramentas biotecnológicas, como o uso de marcadores moleculares, pode auxiliar programas de melhoramento genético da goiabeira. Na literatura há relatos do uso de marcadores moleculares na caracterização de germoplasma (Briceño *et al.* 2010, Sánchez-Teyer *et al.* 2010), divergência genética (Corrêa *et al.* 2011, Costa e Santos 2017), transferibilidade (Rai *et al.* 2013), mapeamento genético (Rodríguez *et al.* 2007, Lepitre *et al.* 2010, Padmakar *et al.* 2015) e análise de QTL (Ritter, Rodríguez-Medina *et al.* 2010).

Microssatélites (SSRs) são loci de herança codominante, abundantes e conservados no genoma e de alta taxa de transferibilidade (Valdés-Infante *et al.* 2010, Fagundes *et al.* 2016). Os microssatélites são extensivamente utilizados em estudos de transferibilidade a fim de elucidar as relações filogenéticas. O tempo necessário para a conclusão de um estudo de

transferibilidade pode ser longo, devido a dependência da validação via produtos de PCR que depende da especificidade de nucleotídeos e da temperatura de anelamento (Bustin *et al.* 2020).

Atualmente com o aperfeiçoamento das tecnologias de segunda geração de sequenciamento a utilização de SNPs tem crescido na pesquisa em plantas (McCallum *et al.* 2016). SNP, do inglês “single nucleotide polymorphism” são marcadores dominantes, de alta densidade no genoma porém de baixa capacidade de identificar polimorfismos devido a sua natureza bi-alélica (Ball *et al.* 2010). Sua abundância nos genomas o torna uma escolha interessante para o mapeamento genético de alta densidade (Akond *et al.* 2013) pois permite uma maior cobertura do genoma.

O mapeamento de ligação consiste na estimativa da posição e distância entre marcadores moleculares baseada no número de crossing-overs, já o mapeamento físico se baseia na distância real entre nucleotídeos no genoma. Os principais entraves na construção de um mapa genético são: a determinação dos grupos de ligação, a ordenação e a estimativa das distâncias entre marcadores (cM) (Cheema e Dicks 2009), que podem ser estimados por meio de diferentes programas e algoritmos (Van Ooijen 2006, Zhang *et al.* 2015, Meng *et al.* 2015).

A utilização de estratégias que minimizem tempo necessário e reduzam custos em estudos de transferibilidade de loci de SSR, assim como estratégias que reduzam as inconsistências entre mapas genéticos e físicos são de extrema importância pois contribuem para a aplicação da seleção assistida por marcadores em programas de melhoramento de goiabeira, que resultem na recomendação de novas cultivares.

2. REVISÃO DE LITERATURA

2.1 Taxonomia, origens e aspectos botânicos da espécie *Psidium guajava* L.

A goiabeira (*Psidium guajava* L.) é uma planta tropical pertencente à ordem Myrtales, família Myrtaceae. Esta família engloba mais de 5650 espécies divididas entre 130 a 150 gêneros (Govaerts *et al.* 2008) alguns deles de importância econômica e social, como o gênero *Eucalyptus*, com destaque para *Eucalyptus globulus*, utilizado na produção de lenha, carvão vegetal, celulose, papel, entre outros subprodutos; *Eugenia* no qual se destaca *Eugenia uniflora* L., amplamente consumida tanto *in natura* quanto industrializada; o gênero *Myrcia*, representado pela *Myrciaria cauliflora*, popular pelo seu uso no paisagismo, propriedades

medicinais dos frutos e consumo *in natura*. No gênero *Psidium* as diversas espécies de araçá (como *P. cattleyanumm*, *P. guineense*, *P. myrsinites*, entre outras) se destacam pelo sabor exótico, alto teor de vitamina C e potencial para exploração por pesquisadores (Franzon *et al.* 2009), sendo aplicados em estudos de caracterização de germoplasma e em hibridação com *P. guajava*.

Embora o centro de origem da *P. guajava* seja o norte da América do sul (Risterucci *et al.* 2005), atualmente está amplamente distribuída nos trópicos e produzida por diversos países como México, Tailândia, Índia, Cuba, China, Paquistão e Estados Unidos, no estado no Havaí. Esta cultura possui alta adaptabilidade edafoclimática, tolerância a estresses bióticos e abióticos, alta produtividade e propriedades medicinais (Padmakar *et al.* 2015), características as quais tornam a goiabeira interessante do ponto de vista produtivo e, em conjunto com suas características de sabor e aroma inconfundíveis, justificam sua popularidade a nível mundial.

A espécie *P. guajava* é uma planta perene de porte pequeno a médio (3 a 5m de altura) que apresenta flores hermafroditas em inflorescências do tipo racemo. As flores podem ocorrer em botões isolados ou em grupos de dois ou três, dependendo do cultivar, mas sempre na axila das folhas (Gonzaga Neto e Soares 1994). O cálice é completo, tem pelos e é persistente, com quatro a seis lobos. As pétalas, em número de quatro ou cinco, são brancas, obovadas e côncavas. Os estames são numerosos e dispostos em ruas no disco. O ovário apresenta quatro ou cinco lóculos (Barbosa e Lima 2010). Suas folhas adultas apresentam forma oval, oblonga ou elíptica com pecíolo arredondado de cor verde-amarelada. Os frutos da goiabeira são globosos, ovoides ou piriformes, com 4 a 10 cm de diâmetro e peso que varia entre 100 a 450 g. O fruto é uma baga, circundada pelos lobos do cálice, sendo globoso, ou ovoide, ou piriforme, com 4 cm a 10 cm de diâmetro. A coloração da casca varia de verde-clara a amarelo-brilhante (Barbosa e Lima 2010), tornando-se amarelada a medida que ocorre a maturação do fruto. A coloração de polpa pode ser branca, vermelha, amarela ou rosa, sendo sempre carnuda e com grande quantidade de sementes.

2.2 Composição química e utilização da *P. guajava*

A goiaba é um fruto versátil que pode ser consumido tanto *in natura* como também industrializado na forma de doces, geleias, néctar, xarope, sorvete ou suco (Padmakar *et al.* 2015). Do ponto de vista nutricional os frutos se destacam pelos altos teores de vitamina A, vitamina C, fibras, cálcio e licopeno (Bramley 2000). Pommer *et al.* (2013) discutem o licopeno como um diferencial na goiaba vermelha e destacam o alto valor nutricional da goiaba, que

chega a superar os frutos cítricos. Além disso, os frutos também apresentam um alto teor de pigmentos antioxidantes como carotenoides e polifenóis, assim como alto teor de ácido ascórbico, o que aumenta o valor nutricional da goiaba (Wilson *et al.* 1982). Além dos frutos, as folhas também são ricas em compostos fenólicos e óleos essenciais e são utilizadas na medicina popular (Campos 2010).

Sendo uma cultura originária das Américas, a goiabeira fez parte da alimentação de populações indígenas na América Central, como no México, e é culturalmente utilizada também no tratamento de doenças. As folhas, frutos, raízes e caule da goiabeira são utilizados no tratamento de gastroenterites, desinteria, dores de estômago, antibactericida para patógenos do intestino, anti-inflamatório, hipertensão, cáries, analgésico e na redução de febre (Gutiérrez *et al.* 2008). Holetz *et al.* (2002) testaram a capacidade antibacteriana de treze extratos vegetais utilizados comumente no Brasil, dentre eles de *E. uniflora* e *P. guajava*, no qual a goiabeira apresentou atividade antibacteriana moderada contra bactérias gram-positivas e também gram-negativas. Os mesmos autores também destacam o uso de frutos e folhas no tratamento de anorexia, diarreia, problemas intestinais, insuficiência gástrica, mucosas inflamadas, laringite, problemas de pele, corrimento vaginal e garganta inflamada.

2.3 Importância econômica

Segundo dados do IBGE (2019), a produção nacional atingiu 584 mil toneladas de frutos produzidos em 22.624 hectares de área plantada. O rendimento médio de produção estimado é de 26.402 kg/ha. Entre os estados de maior produção no país destacam-se São Paulo, Pernambuco, Bahia e Ceará. Esta produção é direcionada majoritariamente para o mercado nacional, pois a goiaba de polpa vermelha (mais produzida) não tem boa aceitação no mercado externo (Barbosa e Lima 2010). Em Pernambuco, o Vale do São Francisco vem se destacando na produção de goiaba irrigada, com destaque para Petrolina que no ano de 2019 obteve produção de mais de 94 mil toneladas. Nos últimos 10 anos tem se observado aumento na produtividade e no valor da produção, sem aumento acentuado na área plantada com goiabeira no Brasil (Figura 1).

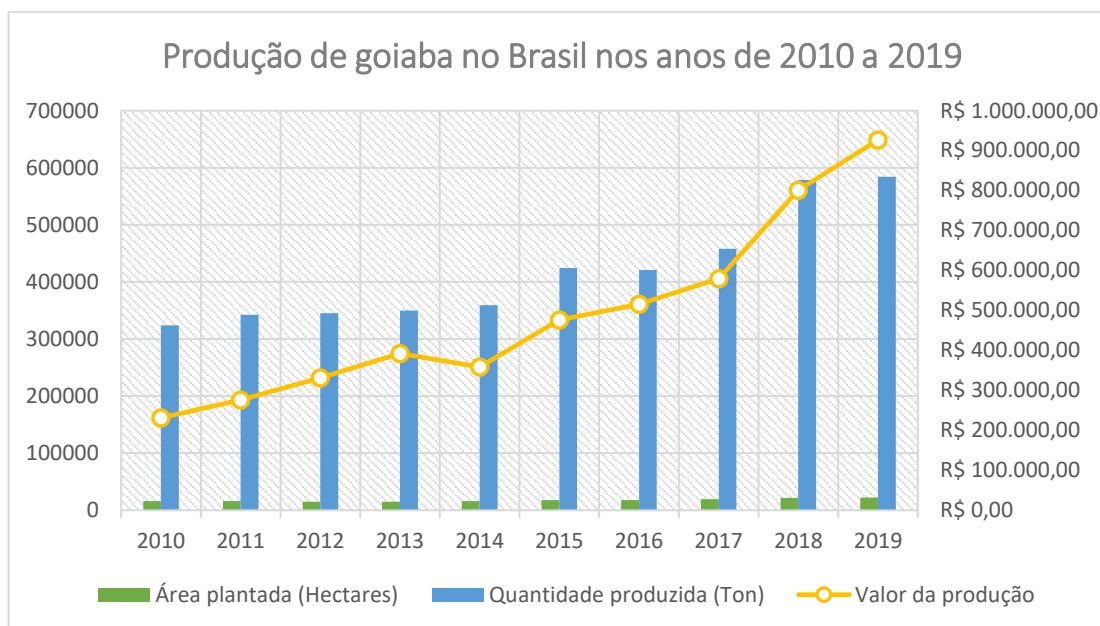


Figura 1 Produção de goiaba nos últimos 10 anos no Brasil: área plantada em hectares (valores de referência à esquerda), quantidade produzida em toneladas (valores de referência à esquerda) e valor de produção (valores de referência à direita).

O aumento da produção nacional é justificado não somente pelo consumo da fruta *in natura*, mas também pelo consumo das suas formas industrializadas como sucos, geleias, doces, entre outros (Quintal *et al.* 2017). O melhoramento genético da cultura, aliado as boas práticas culturais, auxilia no aumento da produtividade sem que haja aumento significativo na área plantada, minimizando os impactos ambientais da agricultura. O mercado da goiabicultura está em expansão, devido à facilidade de manejo e longo período de colheita, com os produtores investindo na cultura. Nos anos de 2017 e 2018 teve aumento significativo no valor de produção da goiabeira (Figura 1).

De acordo com Padmakar *et al.* (2015) em se tratando do consumo da fruta *in natura* a preferência é por frutos de polpa avermelhada, tamanho médio e com alto teor de sólidos solúveis, portanto para atender as demandas de mercado estas características se tornam objetivos dentro dos programas de melhoramento genético. Um dos principais desafios da cultura da goiaba é o manejo das nematoses, principalmente da espécie *Meloidogyne enterolobii*, conhecido como nematoide-das-galhas. Esta doença causa perda na produtividade e até mesmo perda total da área de produção. A Embrapa Semiárido (Petrolina – PE) têm realizado pesquisas para contornar esse problema, desenvolvendo a cultivar porta-enxerto BRS Guaraçá (Santos *et al.* 2017) resistente ao nematoide.

2.4 Melhoramento genético da goiabeira

Plantas alógamas são aquelas que apresentam 95% ou mais de sua reprodução por fecundação cruzada, como é o caso de *P. guajava*. A goiabeira é uma espécie altamente heterozigota devido a sua forma de reprodução e apresenta dois conjuntos básicos de cromossomos, com $2n = 22$. A grande variabilidade genética presente na goiabeira é justificada por Pommer (2012) pelo fato de a propagação ser sexuada (via semente) gerando elevada variabilidade entre suas cultivares, o que para os programas de melhoramento é muito importante pois representa um maior quantitativo de material genético a ser trabalhado.

Programas de melhoramento para a goiaba existem no Brasil há mais de 50 anos com a liberação das cultivares pelo Instituto Agrônomo de Campinas (IAC) e destaque para o programa realizado pelo Prof. F.M. Pereira (UNESP – Jaboticabal) que produziu boas cultivares, como Paluma, Rica e Século XXI (Pereira e Nachtigal 2002). Muitas outras cultivares importantes foram obtidas pelos próprios agricultores como a Kumagai, Ogawa, Pedro Sato e Sassaoka (Pommer *et al.* 2013). Na condução de programas de melhoramento genético Negi e Rajan (2007) descrevem alguns métodos, como introdução de genótipos e seleção, uso de hibridações intervarietais e/ou hibridação interespecífica controlada ou de polinização aberta. Segundo Pommer e Murakami (2009) a maioria dos programas de melhoramento genético de goiaba são oriundos de cruzamentos artificiais, tanto intraespecíficos como interespecíficos.

O primeiro passo de um programa bem-sucedido é conhecer o material que será utilizado por meio da caracterização de germoplasma. Esta caracterização pode ser feita por descritores morfo-agronômicos (fenotípicos) ou por marcadores moleculares. Valdés-infante *et al.* (2007) realizaram a caracterização de acessos cubanos de goiabeira utilizando marcadores microssatélites. Neste estudo, sete loci exibiram 34 alelos nos acessos, sendo 10 alelos raros. O conhecimento da variabilidade da espécie aumenta a quantidade de matéria-prima à disposição do melhorista e permite o planejamento dos cruzamentos que tenham maior probabilidade de gerar indivíduos com alta concentração de alelos favoráveis.

Para seleção de novas cultivares, Pereira e Nachtigal (2002), sugerem valores de peso médio do fruto superior a 100 g, polpa vermelha com rendimento de polpa superior a 70%, sólidos solúveis superior a 10 °Brix, relação açúcar/acidez superior a 11, teor de vitamina C em torno de 100 mg de ácido ascórbico por 100 g de polpa e produção mínima de frutos de 30 t/ha. Outros autores sugerem outros critérios para seleção: I. Frutos: a) tamanho grande (200-340 g),

com poucas sementes e polpa grossa; b) polpa branca para mesa e rosa escuro a vermelho, para a indústria; c) sabor e aroma característicos do fruto; d) conteúdo de sólidos solúveis totais superior a 10%; e) acidez de 1,25-1,50%, nos destinados a processamento, e de 0,2-0,6%, naqueles para mesa; f) conteúdo de vitamina C igual ou maior que 300g.kg⁻¹; g) um número mínimo de células pétreas, que dão o toque de dureza na polpa; h) boa qualidade pós-colheita; i) resistência a doenças e pragas que danificam os frutos; II. Plantas: a) árvores vigorosas, com copa bastante aberta e baixo desenvolvimento em altura; b) resistência a pragas e doenças; c) alta produção; d) porta-enxertos ananícantes.

Programas de melhoramento genético no Brasil e no mundo tem se beneficiado dos avanços de metodologias como a seleção assistida por marcadores moleculares (SAM) a fim de reduzir custos dentro do programa, reduzir seu tempo de duração, ter maior precisão na escolha de indivíduos superiores e maior confiabilidade na recomendação dos genótipos parentais para cruzamentos posteriores. A identificação do polimorfismo a nível molecular, sem a necessidade de esperar a expressão fenotípica, é muito vantajosa, pois exclui a influência do ambiente e permite ao melhorista melhor entendimento da característica. A seleção assistida por marcadores é de suma importância em programas de melhoramento de plantas de ciclo longo, características de difícil avaliação ou de baixa herdabilidade (Souza 2001).

2.5 Marcadores moleculares

Na década de 70 os estudos da genética passaram por uma transformação com a descoberta das enzimas de restrição, com Meselson e Yuan (1968) relatando a primeira endonuclease isolada e purificada a partir do material genético da bactéria *Escherichia coli*. A importância das enzimas de restrição foi reconhecida pelo prêmio Nobel de Fisiologia e Medicina, o qual foi dividido entre as contribuições de três trabalhos (Arber 1964, Smith e Welcox 1970, Danna *et al.* 1971). As enzimas de restrição têm capacidade de clivar o material genético em locais específicos dos genes, dividindo a dupla hélice do DNA, estrutura proposta por Watson e Crick (1953), em pedaços menores. Isto permitiu o desenvolvimento de diversas áreas de conhecimento como a engenharia genética e a biotecnologia. Com o passar do tempo, técnicas como a tecnologia do DNA recombinante, produção de transgênicos, edição gênica, entre outros, ganharam destaque em diversas áreas de conhecimento. O uso de marcadores também ganhou destaque pelo seu papel fundamental em estudos de diversidade, variabilidade genética, filogenia e construção de mapas de ligação (Kalendar *et al.* 2011).

Marcadores moleculares são sequências de DNA que revelam polimorfismo. Em genética são definidos como um fragmento de DNA que está associado a parte de um genoma, onde está localizado um gene responsável por uma característica (Bagali *et al.* 2010). Marcadores moleculares podem ser divididos em marcadores enzimáticos e marcadores de DNA. Os marcadores enzimáticos (isoenzimas e aloenzimas) ganharam importância por um curto período de tempo antes da disponibilização de marcadores de DNA (Grover e Sharma 2016).

Os marcadores de DNA são classificados de acordo com a técnica aplicada para sua utilização, podendo ser baseados em hibridização de fragmentos, na técnica de PCR (polymerase chain reaction) e baseados em uma sequência de DNA alvo. Grover e Sharma (2016) descrevem os principais marcadores moleculares e alguns serão descritos a seguir:

a) RFLP (Restriction Fragment Length Polymorphism): baseado no tratamento com enzimas de restrição e posterior separação dos fragmentos de acordo com o tamanho durante a eletroforese;

b) RAPD (Random Amplified Polymorphic DNA): baseado na técnica de PCR utiliza um único primer que se anela aleatoriamente, o que ocasiona à amplificação ao acaso de segmentos de DNA no genoma;

c) ISSR (Inter-Simple Sequence Repeat): as reações de amplificação são realizadas com um único oligonucleotídeo que vai amplificar diversas regiões entre os microssatélites utilizados. Essencialmente combina os benefícios do RAPD com a reprodutibilidade e especificidade;

d) AFLP (Amplified fragment length polymorphism): também baseado na amplificação via PCR de fragmentos seletivos. Estes fragmentos são gerados por enzimas de restrição de corte raro e de corte frequente e são ligados à nucleotídeos adaptadores;

e) SCARS (Sequence-Characterized Amplified Region): envolve clonar e sequenciar fragmentos de DNA amplificados de interesse e desenhar primers complementares aos seus fins 3’;

f) RGAP (Resistance Gene Analogue Polymorphism): outra técnica baseada em PCR onde primers são usados para amplificar genes de resistência em plantas e seus análogos.

g) Rep-PCR (Repetitive sequence based): essa impressão digital do DNA é baseada no uso de primers que correspondem a elementos naturalmente repetitivos do genoma como a REP, ERIC e BOX;

h) SSR (Simple Sequence Repeat): também conhecido como microssatélites, são regiões curtas no genoma e repetidas em tandem;

i) SNP (Single nucleotide polymorphism): são polimorfismos específicos em um único nucleotídeo, ocorrem frequentemente no genoma e, embora a maioria dos SNPs residam em regiões que não codificarão para proteínas, uma grande parte corresponde a mutações em genes que estão associados com doenças ou outros fenótipos.

j) GBS (Genotyping By Sequencing): com o intuito de reduzir a complexidade de grandes genomas o GBS é atingido através do sequenciamento de regiões próximas a regiões de reconhecimento de enzimas de restrição.

Grover e Sharma (2016) também representaram graficamente (Figura 2) a evolução do uso destes marcadores nos últimos 40 anos, sendo a popularidade interpretada com base no número total de citações para cada marcador no Google Scholar, destacando o crescimento da popularidade de marcadores SSR e SNPs.

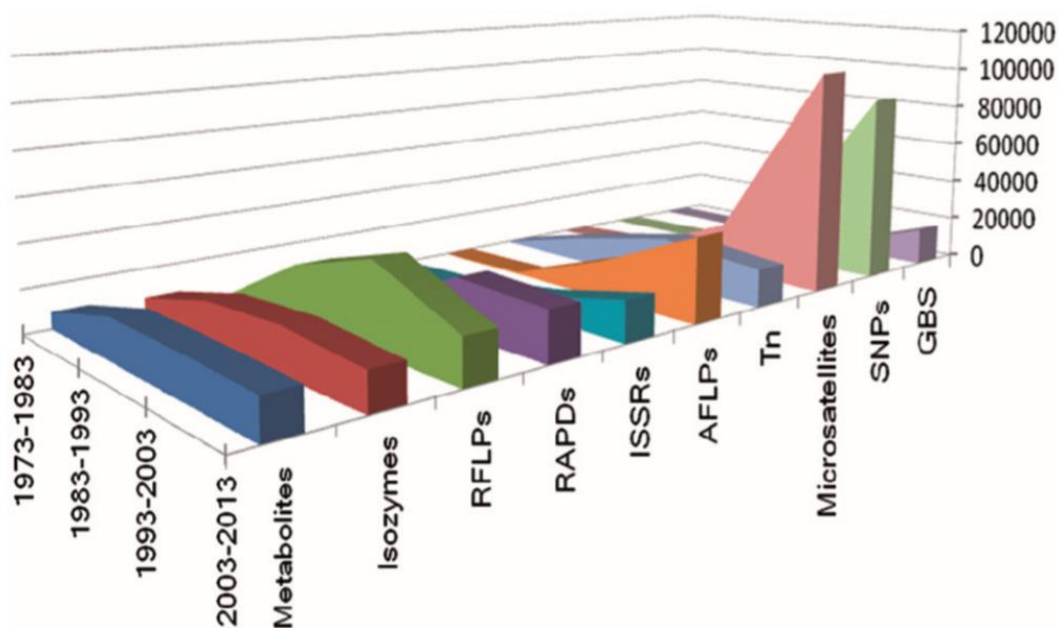


Figura 2 Popularidade de diferentes marcadores moleculares nos últimos 40 anos. Fonte: Grover A., Sharma, PC (2016).

A utilização de marcadores moleculares no melhoramento genético de plantas veio em contraposição ao uso exclusivo de marcadores fenotípicos, uma vez que eles geralmente são de difícil distinção entre indivíduos muito próximos geneticamente e sofrem influência ambiental. Já os marcadores moleculares são capazes de identificar praticamente um número ilimitado de polimorfismos genéticos, não sofrem influência ambiental e tem a capacidade de identificar esses polimorfismos em qualquer estágio de desenvolvimento da planta ou a partir de cultura de tecidos (Faleiro 2007). Bagali *et al.* (2010) descrevem as características ideais para marcadores de DNA: herança codominante, fácil acesso e troca de dados entre laboratórios, ensaios rápidos e fáceis, ocorrência frequente no genoma, alta reprodutibilidade e alta natureza polimórfica.

Rodríguez *et al.* (2007) descrevem a combinação de caracteres morfoagronômicos e marcadores de DNA como uma ferramenta para caracterizar o germoplasma da goiabeira, estimar níveis de diversidade e parentesco entre acessos e para recomendar genótipos para programas de melhoramento e conservação. Alguns trabalhos tem sido desenvolvidos em goiabeira com diversos marcadores como o RAPD para a avaliação de diversidade genética (Filho *et al.* 2010) e avaliação de quantidade quercetina nas folhas da goiabeira (Feria-Romero *et al.* 2009), uso de marcadores AFLP para avaliação da similaridade genética entre acessos de goiabeira e araçazeiro (Corrêa *et al.* 2011), marcadores ISSR na avaliação de fidelidade clonal de goiabeiras micropropagadas (Liu e Yang 2012), transferibilidade de marcadores SSR de goiabeira para quatro espécies de Myrtaceae (Rai *et al.* 2013), confirmação de híbridos interespecíficos no gênero *Psidium* através de marcadores SSR (Costa *et al.* 2016), entre outros.

2.6 Marcadores microssatélites e SNPS e suas aplicações na goiabeira

Marcadores microssatélites ou “single sequence repeat” (SSR) são marcadores codominantes amplamente utilizados na construção de mapas de ligações genéticas, análises de QTL (quantitative trait loci), diversidade genética, análise filogenética, análise de paternidade, estudos comparativos, “fingerprinting” e seleção assistida (Mnejja *et al.* 2010, Kalia *et al.* 2011). Estes marcadores também podem dar suporte à programas de melhoramento e gerar conhecimento sobre a variabilidade genética de populações naturais para adoção de estratégias eficientes de conservação (Oliveira *et al.* 2006).

Microssatélites são sequências repetidas em tandem, de 2 a 5 nucleotídeos, que apresentam alto polimorfismo e podem ser amplificados por reações em cadeia da DNA polimerase (PCR) com primers específicos ou iniciadores derivados de sequência flanqueando

as regiões repetidas no genoma (Risterucci *et al.* 2010). Estes marcadores são codominantes, ou seja, são capazes de identificar as duas cópias dos alelos em indivíduos diploides e assim diferenciar indivíduos homocigotos dos heterocigotos (Figura 3). Mason (2015) destaca que nos últimos 20 anos os SSR foram amplamente utilizados na genotipagem de plantas por serem multialélicos, informativos, codominantes e experimentalmente reproduzíveis entre plantas aparentadas.

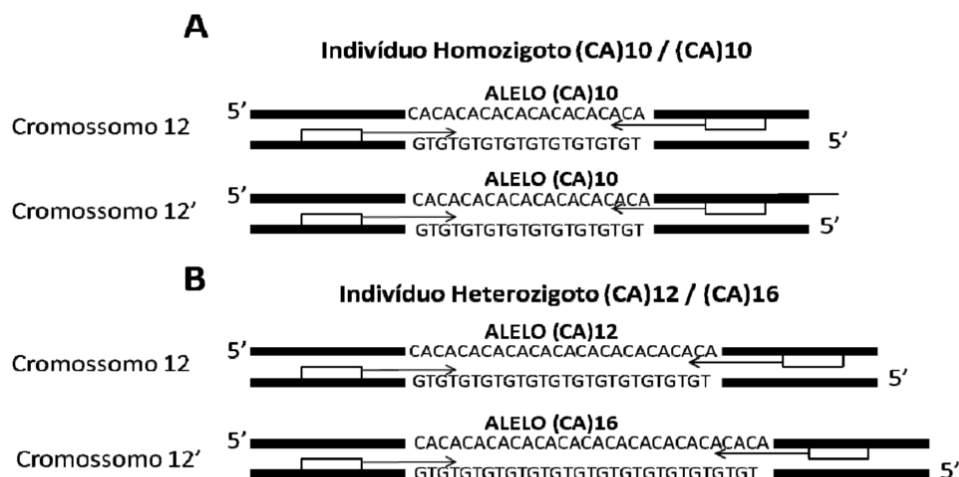


Figura 3 Região microssatélite em indivíduo homocigoto (A) e heterocigoto (B). Fonte: Ferreira ME, Grattapaglia D (1998).

As características positivas dos microssatélites os colocam em destaque na escolha para diferentes pesquisas, mas sua utilização depende de sequências de nucleotídeos pré disponibilizadas. Avanços nos estudos de *P. guajava* foram possíveis quando Risterucci *et al.* (2005) disponibilizaram uma biblioteca de referência de microssatélites para goiaba, para identificação de cultivares, análise de pedigree, assim como estudos de mapeamento genético. Com base nestas informações, diversos trabalhos com marcadores microssatélites foram utilizados em estudos de caracterização e diversidade genética em *P. guajava* (Valdés-infante *et al.* 2007, Aranguren *et al.* 2010, Sánchez-Teyer *et al.* 2010, Chaithanya *et al.* 2014, Sither *et al.* 2014). Costa (2013) utilizou marcadores microssatélite para genotipagem de híbridos interespecíficos de *P. guajava* x *P. guineense*. Lepitre *et al.* (2010) utilizaram marcadores SSR e AFLP para construção de um mapa de referência da cultura utilizando a F1 do cruzamento de acessos heterocigotos cubanos (Enana Roja x N6). A partir da biblioteca massiva de marcadores SSR, o GUAVAMAP, um consórcio internacional composto por instituições europeias e latino-americanas, expandiu o quantitativo de pesquisas realizadas em goiabeira, visando o

desenvolvimento de programas de melhoramento genético a nível mundial. O principal objetivo do projeto “Improvement of guava: Linkage mapping and QTL analysis as a basis for marker-assisted selection” foi disponibilizar informações moleculares sobre a cultura, conseguindo impulsionar a caracterização de germoplasma de goiabeira ao redor do mundo. Neste consórcio foram disponibilizadas as sequências de 428 marcadores microssatélites para a goiabeira.

Marcadores SNPs, ou polimorfismo de um único nucleotídeo, são baseados em pequenas variações na sequência de DNA, alterando somente uma das quatro bases: adenina, guanina, citosina e timina (Figura 4) em mais de 1% da população em questão. Podem ser aplicados de diversas formas em projetos de pesquisa, como a construção de mapas genéticos de alta resolução, mapeamento baseado no desequilíbrio de ligação, diagnósticos genéticos, análise de diversidade, identificação de cultivares, análise filogenética e caracterização de recursos genéticos (Rafalski 2002). De acordo com Van Ooijen e Jansen (2013) os SNPs representam uma novidade tecnológica de extrema importância para o mapeamento genético pois aumentam o número de marcadores identificados no genoma.

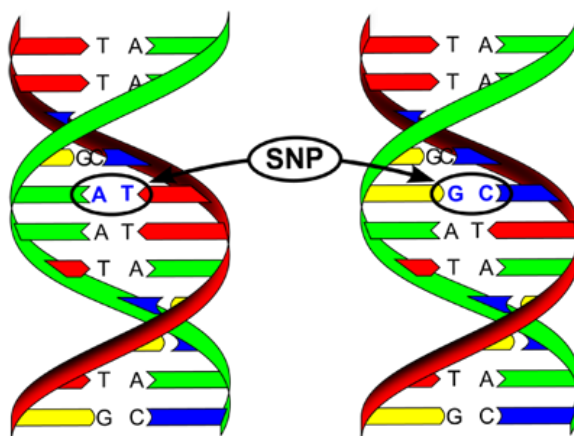


Figura 4 Diagrama representando o polimorfismo de um único nucleotídeo (SNP).

A grande desvantagem descrita no uso de marcadores SNPs é a sua natureza bi-alélica, o que significa que eles proporcionam menos informação por locus que um marcador multialélico. Em estudos de mapeamento genético a informação da variabilidade nos permite estimar a localização e distância dos marcadores através do número de recombinações ocorridas entre eles. Entretanto os SNPs são de grande valia para o mapeamento, de ligação e de associação, pois eles se encontram em todo o genoma e em grande quantidade. De acordo com Ball *et al.*

(2010) a combinação de microssatélites e SNPs são capazes de construir mapas de ligação de alta resolução, pois os microssatélites servem como pontos âncora para alocar os SNPs nos seus grupos de ligação correspondentes.

Silva-Junior et al. (2015) desenvolveram o EUChip60K, um chip multi-espécie de *Eucalyptus* com 59.222 SNPs, gerando dados de milhares de marcadores SNPs para diferentes espécies de *Eucalyptus* e alguns táxons aparentados, como a goiabeira. A transferibilidade destes marcadores SNPs foi testada por Costa e Santos (2017), onde 7% dos marcadores desenvolvidos foram eficientemente transferidos para a goiabeira. Com base nestes marcadores foi caracterizada a divergência genética ente 53 acessos de *Psidium*, sendo 47 de *P. guajava*, três de *P. guineense* e três de *P. cattleyanum*. Atualmente Grossi et al. (2021) reportaram o desenvolvimento de SNPs originados do sequenciamento DArT-seq específicos para espécies de *Psidium*, inclusive *P. guajava*.

2.7 Mapeamento genético aplicado no melhoramento genético de plantas

O mapeamento genético foi fundamentado por Thomas Hunt Morgan (1910), que propôs a teoria da herança cromossômica ligada ao sexo por meio de experimentos utilizando a mosca das frutas *Drosophila melanogaster*. Morgan propôs que fatores mendelianos (hoje são conhecidos como genes) estariam alocados nos cromossomos e que certos fatores só se expressarão conforme o sexo do indivíduo. Para Morgan (1910) as possíveis distorções da segregação mendeliana podem ser explicadas pela recombinação (crossing-over), evento que ocorre quando os fatores estão localizados no mesmo cromossomo. A posição do gene no cromossomo e as distâncias entre os genes são informações cruciais para construção de um mapa genético e estudo da hereditariedade. Sturtevant (1913) utilizou a mosca das frutas para testar se proporção de crossing-overs poderia ser usada como um índice da distância entre dois fatores no cromossomo, tornando possível estimar a estrutura linear do cromossomo e assim propondo o primeiro mapa genético da história. Creighton e McClintock (1931) demonstraram a associação definitiva dos mapas de ligação com o cromossomo em milho (*Zea mays*) demonstrando que o crossing-over é resultado de troca entre segmentos cromossômicos.

Um mapa genético, ou mapa de ligação, ou mapa de ligação genética é a representação das posições relativas dos genes e dos marcadores genéticos nos cromossomos de uma espécie. Estas posições são mensuradas em Morgans, em homenagem ao geneticista; na prática, centiMorgans são normalmente utilizados e abreviados para cM (Van Ooijen e Jansen 2013). A importância de um mapa de ligação genética de alta densidade está no fato dele proporcionar

uma grande quantidade de informação para clonagem, pesquisas de comparação genômica, sequenciamento de genoma completo, mapeamento de QTL e seleção assistida (Van Ooijen e Jansen 2013, Zhao *et al.* 2013).

Van Ooijen e Jansen (2013) descrevem que se pode estimar as distâncias entre dois loci através de suas taxas de recombinação, portanto se combinarmos as frequências de recombinação de múltiplos loci no mesmo cromossomo podemos determinar as posições relativas destes loci no cromossomo. A grande disponibilidade de marcadores neutros, altamente polimórficos, cuja herança não sofre influência ambiental, aliada a procedimentos estatísticos complexos, tem permitido a construção de mapas de ligação para a maioria das espécies vegetais de interesse agrônomo (Carneiro e Vieira 2002). Os mesmos autores também descrevem pontos essenciais a serem levados em consideração na construção de um mapa genético:

- a) Seleção apropriada da população - a população deve ser segregante como em F₂ de progenitores heterozigotos, retrocruzamento ou de auto-fecundação a partir de uma F₂ para a construção de uma linhagem recombinante consanguínea, a fim de que as diferenças entre os alelos sejam visualizadas no mapa. Fehr (1987) propôs a utilização de uma população F₁ como uma F₂ (pseudo-F₂) quando se tratar de um cruzamento controlado entre pais altamente heterozigotos.
- b) Tipo de marcador a ser utilizado - a escolha do marcador molecular baseia-se no número de locus que podem ser detectados, conteúdo informativo do marcador (dominante ou codominante), grau de polimorfismo, número de marcadores (que está diretamente ligado a resolução pré-estabelecida do mapa), custo, tempo necessário para realização das avaliações, praticidade, entre outros.
- c) Verificação da segregação mendeliana - deve ser realizada através do teste de qui-quadrado (χ^2) a fim de calcular se o resultado obtido segue a segregação esperada com base nas leis de Mendel. Os testes probabilísticos assumem segregação mendeliana, caso contrário ocorre a chamada distorção de segregação. Em locus onde haja distorção da segregação podem ocorrer ligações espúrias (ligações falsas) e estas deverão ser desconsideradas do mapa.
- d) Método estatístico a ser aplicado - os métodos estatísticos cruciais no mapeamento são direcionados para o cálculo da frequência de recombinação entre pares de marcadores e funções de mapeamento. A frequência de recombinação pode ser calculada pelo método dos quadrados mínimos, método da máxima verossimilhança, regressão linear, e método

dos momentos. As funções de mapeamento são utilizadas para estabelecer as distâncias genéticas (cM) e também estimar o número de recombinações, sendo as funções de Haldane (1919) e Kosambi (1943) as mais utilizadas no mapeamento genético de vegetais (Staub *et al.* 1996). A diferença entre as funções é que a função de Kosambi admite que haja interferência entre o crossing-over de segmentos adjacentes, enquanto a função de Haldane não assume que ocorra interferência.

- e) Determinação dos grupos de ligação – grupos de ligação são conjuntos de genes localizados no mesmo cromossomo. Com base nas razões de verossimilhança convencionou-se a determinação de um logaritmo na base 10, o LOD score, para determinação do grupo de ligação, sendo o LOD score de 3 a 6 considerados ideais para a formação dos grupos.
- f) Ordenação de marcadores no grupo de ligação - achar a melhor ordem de marcadores no cromossomo é comparável ao problema do caixeiro viajante, onde é necessário encontrar a melhor rota de viagem (de menor custo e tempo) para um vendedor visitar várias cidades diferentes. A ordenação de marcadores no mapeamento genético difere deste problema, pois se entende que o cromossomo é uma estrutura linear e, portanto, a distância entre marcadores é unidirecional. A melhor ordem dos marcadores não é conhecida antes da análise de ligação, para um grupo de ligação com “m” marcadores as ordens possíveis são $m!/2$ (Van Ooijen e Jansen 2013). Existem diversos algoritmos que podem ser utilizados para ordenação de marcadores, sendo alguns deles: delineação rápida em cadeia (“rapid chain delineation” – RCD) (Doerge e Weir 1994), seriação (“seriation” - SER), (Buetow e Chakravarti 1987), recombination counting and ordering algorithm (RECORD) (Van Os *et al.* 2005), nearest neighbour (vizinho mais próximo) (Stam 1993), simulated annealing (anelamento simulado) (Kirkpatrick *et al.* 1983), algoritmo genético, otimização da colônia de formigas, entre outros.

O RCD consiste inicialmente na identificação de qual par de marcadores possui a menor taxa de recombinação e posterior identificação de qual marcador possui a menor taxa de recombinação com os marcadores iniciais e assim sucessivamente até que todos os marcadores sejam adicionados à cadeia. O algoritmo SER difere do RCD, pois posiciona marcadores à esquerda e a direita do marcador inicial a fim de comparar essas distâncias. Se a distância (ou taxa de recombinação) for menor no marcador da direita (por exemplo), este marcador será posicionado à direita do marcador inicial, este processo se repete até que todos os marcadores sejam posicionados. O algoritmo RECORD consiste na contagem dos eventos de recombinação

e ordenação. Neste caso utiliza-se o número esperado de eventos de recombinação entre marcadores para criar uma matriz. O critério “COUNT” vai avaliar o menor valor que será correspondente a melhor ordem do grupo de ligação. O algoritmo do vizinho mais próximo, ou algoritmo ganancioso, constrói o mapa locus por locus, começando do par de loci no meio do mapa. O próximo locus é escolhido baseado na soma dos LODs de todos os outros loci. Esse critério implica que o locus escolhido tem ligação máxima com o loci avaliado no momento (Van Ooijen e Jansen 2013). O algoritmo “simulated annealing” examina todas as possíveis ordens de uma amostra e é capaz de encontrar o mínimo local através de métodos de probabilidade, mesmo na presença de vários mínimos locais (Nomelini *et al.* 2009).

Após a estimativa da ordem dos marcadores é necessária à avaliação da acurácia do algoritmo escolhido. Para tal vários critérios podem ser utilizados: soma mínima das frações de recombinação adjacentes (“sum of adjacent recombination fractions” – SARF) (Falk, 1989); produto mínimo das frações de recombinação adjacentes (“product of adjacent recombination fractions” – PARF (Wilson 1988); soma máxima dos LOD Scores adjacentes (“sum of adjacent LOD Scores” – SALOD) (Weeks e Lange 1987).

Alguns *softwares* são utilizados na construção de mapas de ligação e associação de QTL:

- 1) MapMaker (Lander *et al.* 1987) que calcula todas as distâncias de máxima verossimilhança e os seus LOD scores estabelecendo LOD score maior que 3 e a distâncias entre marcadores < 80 cM utilizando a função de Haldane (Boopathi 2020);
- 2) MapManager QTX (Manly *et al.* 2001) que oferece três funções de mapeamento (Haldane, Kosambi e Morgan), utiliza o método Markov-chain para reconstrução genética de marcadores perdidos e possui métodos opcionais para cálculo de ligação e distâncias que não são afetadas pela distorção de segregação;
- 3) JoinMap®4 (Van Ooijen 2006) um programa desenhado para gerar mapas de ligação, integrar mapas diferentes em um mapa consensus usando marcadores (âncoras); a partir de diferentes tipos de populações experimentais (inclusive famílias de irmãos completos F₁). O mapa pode ser complementado utilizando o pacote MapQTL para análise de ligação de caracteres quantitativos. Sua versão mais atual (JoinMap 5) substituiu o método de Gibbs para estimativa multiponto das frequências de recombinação pelo método Markov Chain, por se tratar de um método mais rápido e mais preciso;
- 4) OneMap (Margarido *et al.* 2007) foi desenvolvida em ambiente R para facilitar a análise de ligação em espécies de polinização cruzada utilizando a metodologia de Wu *et al.* (2002) que

permite estimar simultaneamente a ligação e fase de ligação de um conjunto de marcadores informativos e parcialmente informativos em uma família de irmãos completos;

5) GACD (Zhang *et al.* 2015) é um programa integrado para análise de ligação, construção de mapas e mapeamento de QTL em F₁ clonal e populações de polinização cruzada;

6) QTLiciMapping (Meng *et al.* 2015) um programa integrado utilizado para mapeamento de alta densidade e análise de QTL, o qual se baseia no mapeamento intervalado composto para detecção de QTL. Neste programa também pode-se encontrar os comandos BIN para exclusão de marcadores redundantes, MAP para a construção do mapa genético, CMP para a construção do mapa consensus através dos marcadores âncora, SDL mapeamento dos loci em distorção de segregação, BIP para mapeamento de genes aditivos, dominantes e epistáticos, MET análise de QTL pela interação com ambiente, CSL mapeamento de genes aditivos e epistáticos em segmentos cromossômicos em substituição e NAM: mapeamento de QTL em populações NAM (populações onde ocorre múltiplos cruzamentos com um parental comum).

O primeiro mapa de ligação em goiabeira foi realizado por Valdés-Infante *et al.* (2003) utilizando marcadores AFLP e posteriormente estendido (Rodríguez *et al.* 2007) totalizando 220 marcadores alocados em 11 grupos de ligação. (Ritter, Herran, *et al.* 2010) compararam os grupos de ligações de três populações diferentes, com base em marcadores SSR e AFLP, utilizando os SSR como pontos âncora. Lepitre *et al.* (2010) disponibilizaram um mapa contendo 850 marcadores com média de 77 marcadores por grupos de ligação (Figura 5). Padmakar *et al.* (2015) destacam a importância de gerar mapas genéticos saturados e marcadores funcionais a fim de avançar nos programas de melhoramento de goiabeira e reportaram o desenvolvimento de mapas de ligação intra específicos utilizando marcadores SSR juntamente com a genotipagem da população utilizando marcadores SRAP, com o intuito de enriquecer os mapas de ligação disponíveis para goiabeira.

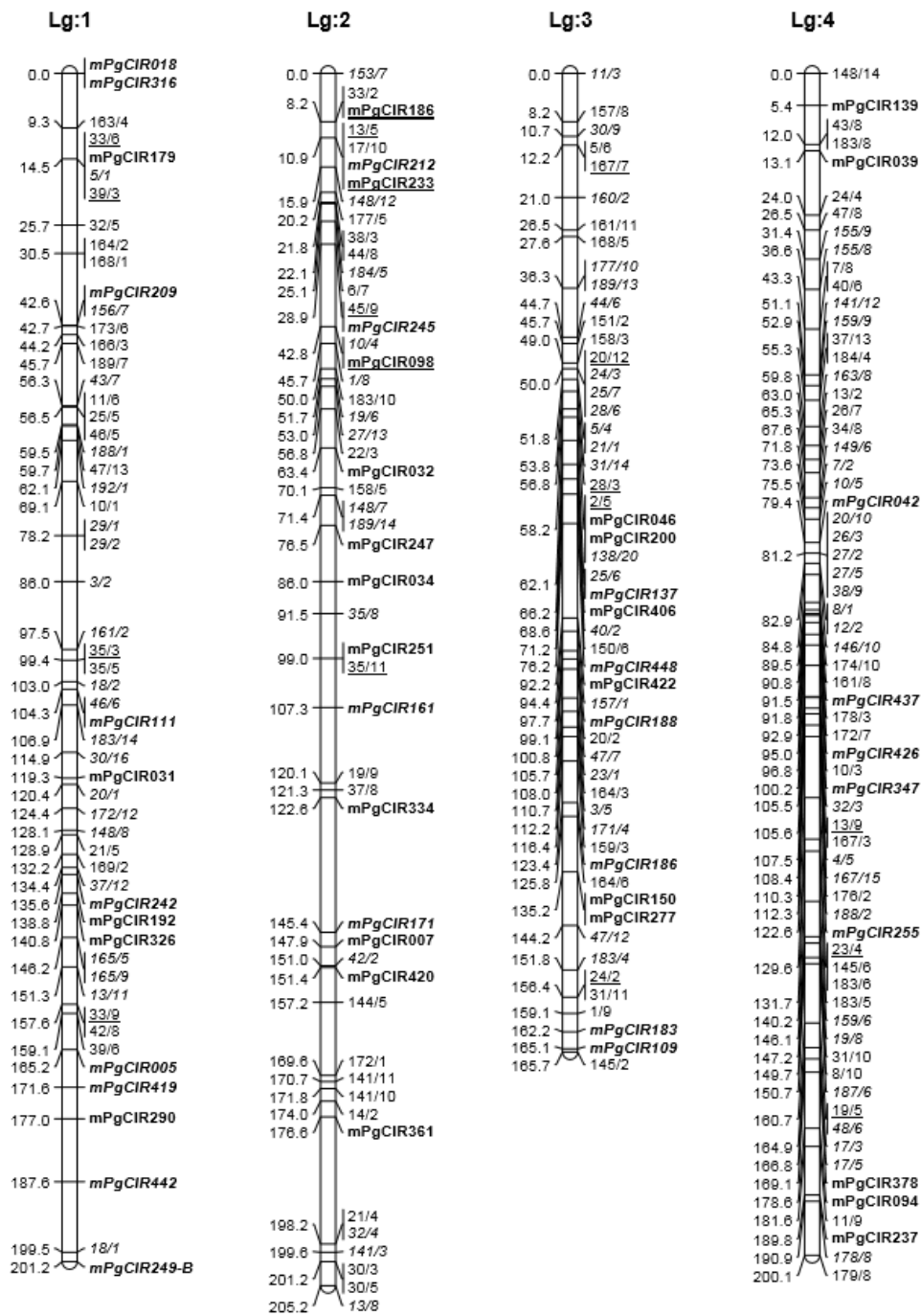


Figura 5 Grupos de ligação 1 a 4 em *P. guajava* provenientes do mapeamento de ligação utilizando marcadores SSR e AFLP. Fonte: Lepitre *et al.* (2010).

2.8 Análises de QTL e suas aplicações em *P. guajava* L.

Características de herança discreta são bons marcadores fenotípicos e permitem ao pesquisador estudar características como nanismo, albinismo e coloração de pétalas. Entretanto, quando se estudam características de distribuição contínua o nível de dificuldade é bem maior. Essas características são denominadas poligênicas, quantitativas ou de herança complexa, e os loci que as controlam são chamados de QTL's (Tanksley 1993). Este tipo de herança tem lugar de destaque no melhoramento vegetal pois diversas características de importância econômica são poligênicas, como produtividade e peso de frutos, Sakiyama *et al.* (2014) destaca que tais características geralmente apresentam herdabilidade baixa à moderada, o que é um problema quando se trabalha com métodos de melhoramento convencional.

Um estudo pioneiro realizado por Thoday (1961) propôs que os poligenes poderiam ser mapeados caso fossem localizados indiretamente, através de sua proximidade a genes marcadores genéticos. Esta metodologia embasa até hoje as associações de QTL's. Com a utilização de marcadores moleculares abundantes no genoma vegetal essa associação é cada vez mais precisa. Ritter *et al.* (2005) descrevem a análise de QTL como uma forma de dissecar a variação observada e entender o efeito do locus quantitativo na expressão de uma característica. Dessa forma marcadores moleculares que estejam próximos ou ligados a QTLs podem ser utilizados na seleção assistida por marcadores (SAM).

A seleção assistida utiliza a presença de marcadores moleculares para auxiliar a seleção fenotípica, de forma a torná-la mais eficiente, eficaz, confiável e econômica comparada ao melhoramento de plantas mais tradicional (Collard *et al.* 2005). É recomendada a seleção assistida quando a seleção fenotípica for onerosa, quando a seleção do gene depende de condições ambientais específicas ou estágios de desenvolvimento e na piramidização de vários QTL para uma característica (Schuster 2011, Sakiyama *et al.* 2014). Para realizar a seleção assistida para QTL necessita-se de um mapa genético de alta resolução, pois assim tem-se maior confiança de que os QTLs estão flanqueados por marcadores (pelo menos < 5 cM, para garantir que eles estão ligados e não ocorrerá recombinação entre o marcador e o gene). Um dos fatores limitantes para o uso da seleção assistida para QTL é a produção de um mapa genético de baixa acurácia e necessidade de validação em diferentes ambientes (Schuster 2011). Ferreira e Grattapaglia (1998) destacam fatores importantes a se considerar quando utilizar marcadores moleculares associados a QTLs: a repetibilidade da informação de ligação em diferentes populações, a interação QTLs x ambiente e a seleção simultânea de vários caracteres.

Existem metodologias amplamente utilizadas para a detecção de QTL como: análise de um marcador único, mapeamento de intervalo simples e mapeamento de intervalo composto (composite interval mapping - CIM).

1) A análise de um marcador único é um método simples, pode ser realizado sem o mapa de ligação completo e utiliza a análise de variância e regressão linear para detecção de QTL associados a marcadores simples (Collard *et al.* 2005);

2) O mapeamento de intervalo simples utiliza o mapa de ligação e a análise de intervalo entre pares adjacentes de marcadores ligados (Lander e Botstein 1989). Este método é considerado mais eficiente para a detecção de QTL pois leva em consideração a recombinação entre os marcadores, estimada pelo LOD score e pelo método de máxima verossimilhança, diferentemente da análise de um único marcador. A razão de verossimilhança é estimada pela proporção entre a probabilidade de o QTL estar ligado ao marcador e a probabilidade de eles não estarem ligados;

3) O método de intervalo composto é o mais utilizado e estima os QTL's fora do intervalo em questão por um modelo de regressão linear múltipla (Jansen 1993, Zeng 1993, 1994). A maior vantagem do método de intervalo composto é a sua precisão e efetividade no mapeamento de QTLs comparado à análise de um único marcador e o mapeamento de intervalo simples, especialmente quando QTLs ligados estão envolvidos (Collard *et al.* 2005).

Diferentes programas são utilizados na detecção e análise de QTLs, geralmente integrados com *softwares* de mapeamento de ligação como: o GACD (Zhang *et al.* 2015) e QTL Ici Mapping (Meng *et al.* 2015), descritos anteriormente. Outros programas populares são: MapManager QTX (Manly *et al.* 2001) que mapeia os QTL através de três funções de mapeamento (Haldane, Kosambi e Morgan), utiliza o método Markov-chain para reconstrução genética de marcadores perdidos e possui métodos opcionais para cálculo de ligação e distâncias que não são afetadas pela distorção de segregação; o QTL Cartographer (Basten *et al.*, 1999) usa regressão linear, mapeamento intervalado (Lander e Botstein 1989) ou método de intervalo composto (Zeng 1993, 1994) e possui um algoritmo que permite diversas análises estatísticas em várias ações gênicas, interações QTL x ambiente e ligações próximas de marcadores; a extensão R/qtl (Arends *et al.* 2010) que utiliza a linguagem R para mapear QTLs em populações experimentais utilizando mapeamento de QTL múltiplo (MQM) (Jansen 1993) definido como a combinação da força dos modelos lineares generalizados de regressão com os modelos de mapeamento intervalado. Arends *et al.* (2010) descreve algumas vantagens que o MQM traz para o programa R/qtl: mais potência quando QTL é responsável por boa parte da

variação e a prevenção de detecção de QTLs fantasmas (QTLs inexistentes que podem aparecer devido “acúmulo” de efeitos poligênicos uniformemente distribuídos no genoma); e por fim o MapQTL5.0 (Van Ooijen, 2004) que é utilizado para diversos tipos de populações experimentais diploides, é capaz de analisar experimentos através do mapeamento intervalado (Lander e Botstein 1989), do mapeamento múltiplo (Jansen 1993) equivalente ao mapeamento intervalado composto proposto por Zeng (1994). Outra adição a este programa foi o método de regressão de Haley e Knott (1992) que trouxe vantagens como maior velocidade de análise, redução do uso de memória RAM.

Trabalhos com *P. guajava* ainda são escassos, mas grandes contribuições no mapeamento genético e análise de QTL foram realizadas por Valdés-Infante *et al.* (2003) e posteriormente Rodríguez *et al.* (2007), que integraram ao mapa genético 15 loci QTLs para comprimento de folha, largura de folha, comprimento do pecíolo, altura de planta, taxas de crescimento para diâmetro e altura de planta. Ritter, Rodríguez-Medina *et al.* (2010) realizaram análise de QTL pelo método do mapeamento de intervalo simples em três populações de goiabeiras cubanas. As características avaliadas foram: altura de planta, comprimento de pecíolo, comprimento e largura de folha, número de frutos e média de peso de fruto, comprimento e largura de fruto, firmeza interna e externa de polpa, número de sementes, peso de sementes, conteúdo de vitamina C acidez e total de sólidos solúveis. Mais de 100 QTLs foram detectados para diferentes características nas três populações e variaram entre 2 e 13 QTLs por característica. Alguns desse QTLs estão localizados ou ligados a marcadores SSR utilizados no estudo, o que permitirá uma seleção assistida eficiente (Ritter 2012).

2.9 Transferibilidade de locus SSR

A transferibilidade de marcadores microssatélite é uma técnica bastante utilizada quando não há disponibilização de uma biblioteca genômica da espécie. As regiões que flanqueiam os microssatélites são extremamente conservadas entre espécies correlatas, o que permite a utilização dos primers SSR em espécies não alvo, gerando uma diminuição de custos e trabalhos necessários para a construção de bibliotecas e desenvolvimento de primers (Ferreira e Grattapaglia 1998, Wang *et al.* 2009). Estudos de transferibilidade de locus SSR têm sido desenvolvidos a partir de diversas culturas, como a oliveira (Rallo *et al.* 2003), trigo, centeio e triticale (Kuleung *et al.* 2004), côco (Mistura *et al.* 2012), pêra (Fan *et al.* 2013), mangueira (Ribeiro *et al.* 2008), cebolinha (Santos *et al.* 2009), palmeiras (Fortes *et al.* 2016), entre outros. As análises de transferibilidade são realizadas através da técnica de PCR. Os dados obtidos são

avaliados em função da amplificação ou não do fragmento de DNA, avaliando-se também a qualidade da banda produzida. Este processo demanda a confecção dos primers específicos e ajustes o protocolo até que se tenha certeza da qualidade da reação, o que torna o estudo laborioso e caro.

De modo geral, pode-se afirmar que a ocorrência de SSRs é menor em regiões codificantes do genoma pois os microssatélites têm uma alta taxa de mutação ocasionadas por erros nos processos de replicação, o que pode comprometer a expressão gênica (Vieira *et al.* 2016). Porém, devido a sua natureza altamente informativa e abundância no genoma os marcadores SSR podem ser utilizados âncoras para identificação de genes de interesse. Bouvier *et al.* (2012) trabalharam com a cultura da pera e descobriram um gene resistente a sarna da pera, *Rvp1* localizado próximo ao microssatélite da cultivar Nayara. Este marcador está localizado numa região sintênica a um gene resistente a sarna na maçã, no grupo de ligação 2, indicando a primeira sentença funcional para resistência a sarna em maçã e pera.

Rai *et al.* (2013) destacam a transferibilidade como ferramenta útil em estudos de genômica comparativa e filogenética. A genômica comparativa é uma área de conhecimento que surgiu na década de 90 com os avanços no sequenciamento de genomas completos e baseia-se no princípio que características comuns entre dois organismos, frequentemente codificadas no DNA, se localizam em porções conservadas entre espécies próximas filogeneticamente. Segundo Wei *et al.* (2002) a genômica comparativa pode comparar a estrutura genômica, as regiões codificantes e as regiões não codificantes, incluindo a predição de elementos regulatórios.

A transferibilidade dos 23 primers desenvolvidos para goiabeira (Risterucci *et al.* 2005) foi testada por Rai *et al.* (2013) para 4 espécies da família Myrtaceae: *Eucalyptus citriodora*, *Eucalyptus camaldulensis*, *Callistemon lanceolatus* e *Syzygium aromaticum*. A média de amplificação foi de 71,2% e apenas 8 pares de primers produziram produtos de amplificação em todas as quatro espécies. Nogueira *et al.* (2016) utilizaram 158 SSR de *P. guajava* para avaliar a transferibilidade em 18 espécies arbóreas dos gêneros: *Eugenia*, *Campomanesia*, *Myrciaria* e *Syzygium*, dos quais 152 SSR transferiram para pelo menos 1 espécie, 47 amplificaram para 10 ou mais espécies e o SSR mais conservado (mPgCIR162) amplificou em 16 espécies. Fagundes *et al.* (2016) avaliaram a transferibilidade de 21 microssatélites de *Eucalyptus sp.*, *Eugenia uniflora* e *Melaleuca alternifolia* Cheel para 8 espécies da floresta Atlântica, incluindo *P. guajava* e *Psidium spp.*, onde a taxa de transferibilidade dos pares de

primers foi alta (>52%) em 7 das 8 espécies estudadas. Apesar da disponibilidade de mapas de ligações genéticas (Lepitre *et al.* 2010, Padmakar *et al.* 2015) para a goiaba, esta espécie ainda carece de estudos que sejam capazes de fomentar a utilização da seleção assistida por marcadores com o objetivo de dar suporte aos programas de melhoramento de forma precisa e que reduza o tempo de seleção.

2.10. Genômica comparativa e Alinhamento local (BLAST)

A genômica comparativa é um campo de estudo que possibilita inferências sobre evolução e distância genética entre indivíduos, utilizando ferramentas de alinhamento de sequências ou comparação de genomas completos. Um dos aspectos mais importantes deste campo é a identificação de similaridades, a fim de entender a estrutura genômica de indivíduos próximos filogeneticamente e possivelmente permitir a transferência de informações genéticas entre eles (Butler *et al.* 2017). Alguns conceitos causam confusão dentro da genômica comparativa, sendo eles: sintenia, colinearidade, homologia e similaridade.

Sintenia refere-se à ordem dos genes no cromossomo (e se houve alteração ao longo dos anos), de acordo com Paterson *et al.* (2000), na maioria das plantas, os processos evolutivos ocorreram de forma lenta e não modificaram a porção do genoma que realmente codifica um gene, assim como os arranjos de genes ao longo dos cromossomos, o que caracteriza blocos sintênicos em espécies taxonomicamente próximas. Regiões sintênicas ou blocos sintênicos entre espécies divergentes são de grande valia para a localização de genes ortólogos e predição da posição de marcadores em cromossomos de diferentes espécies, conforme encontra-se na literatura científica (Mateos-Hernandez *et al.* 2006, Wu *et al.* 2009, Hudson, Kullán, *et al.* 2012).

A colinearidade é a correspondência linear entre as sequências de nucleotídeos, o que pode ser usado para identificar blocos sintênicos no genoma. Paterson *et al.* (2000) afirma que as alterações na colinearidade são oriundas de eventos de recombinação nos cromossomos, mutações, deleção de genes e translocações ao longo do processo evolutivo. A sintenia e colinearidade foram avaliadas em três espécies de eucalipto (*E. grandis*, *E. urophylla* e *E. globulus*) através da comparação de mapas de ligação densos por Hudson *et al.* (2012) onde os mesmos relataram alta conservação entre os genomas (altos níveis de sintenia e colinearidade), que sugerem grande possibilidade de transferibilidade de marcadores entre as espécies em questão. Resultados semelhantes foram encontrados na literatura (Grattapaglia *et al.* 2015, Butler *et al.* 2017, Sumathi *et al.* 2018).

Wei *et al.* (2002) definem homologia como a identificação de um ancestral comum entre dois genes ou produtos desses genes, já a similaridade é a medida de semelhança entre alinhamentos sendo, portanto, útil como evidência de homologia. Kuleung *et al.* (2004) afirmam que sequências obtidas de diferentes espécies de plantas demonstram altos níveis de homologia nas regiões que flanqueiam SSR, o que torna essa informação relevante para inferências na posição e função dos genes a partir da identificação dessas homologias. Para identificação de homologias entre sequências de diferentes espécies o BLAST tem sido amplamente utilizado (Varshney *et al.* 2002, 2005, Acuña *et al.* 2012).

A ferramenta BLAST ou “Basic Local Alignment Search Tool” (Altschul *et al.* 1990, 1997) se baseia em algoritmos heurísticos para localizar sequências estimadamente similares em diferentes indivíduos. Esta análise permite o alinhamento de uma dada sequência (query) com outra sequência ou genoma completo (subject). Esta ferramenta encontra-se disponível no site do NCBI (National Center for Biotechnology Information) de forma gratuita para uso online ou utilização como executável (BLAST+) atualmente na versão 2.11.

Pertsemlidis e Fondon (2001) revisaram o funcionamento do BLAST e dividiram em 3 etapas: na primeira o BLAST identifica regiões de baixa complexidade, à exemplo de sequências curtas, e as retira da query, uma vez que essas sequências podem gerar alto número de resultados que não são biologicamente significantes. A partir disto é gerada uma lista que será avaliada por um “scoring matrix” e determinará quais são os “high scoring matching words” e com esses resultados gerará uma nova lista. Na etapa subsequente ocorre a filtragem de erros e “gaps” onde o algoritmo procura no “subject” as sequências exatamente iguais às sequências presentes na lista, e na terceira e última etapa o BLAST se baseia nas “matching words” para estender os alinhamentos em ambas as direções, contanto que o “score” continue crescente, o que gerará “high segments pairs” (HSPs). Os HSPs são identificados por meio da sua comparação a um valor de corte S (determinado empiricamente pelo próprio algoritmo). Se o valor estimado for maior que S, então o alinhamento é considerado relevante e será designado com sua nota (score) e poderá ser avaliado quanto a sua significância estatística.

As sequências utilizadas no BLAST podem ser de nucleotídeos (blastn e tblastx) ou proteínas (blastx e tblastn). Os parâmetros utilizados pela ferramenta BLAST são: bit score, expected value (e-value) e a identidade. O Bit score é um valor normalizado que pode ser comparado entre alinhamentos de diferentes buscas: quanto maior o bit score melhor o alinhamento em questão. O e-value é utilizado como índice de significância do alinhamento, pois quanto menor o e-value menor é a probabilidade de um dado alinhamento ter ocorrido por

acaso: e-value $1e^{-50}$ = indica significância de alta qualidade, e-value de 0,01 pode ser considerada com boa qualidade e e-value=10 indica que não é significativa. A identidade é expressa em porcentagem e diz respeito à coincidência de um mesmo nucleotídeo (ou aminoácido) na mesma posição nas sequências comparadas (Pertsemlidis e Fondon 2001).

Os três parâmetros avaliados em conjunto fornecem uma perspectiva melhor da qualidade de um alinhamento. Além de prever onde o primer em questão possivelmente se anelará no subject, este tipo de informação ainda é útil pois permite uma exploração eficiente dos atuais recursos de goiabeira através da análise de co-locação entre QTLs mapeados e potenciais genes candidatos (Ritter 2012) (Figura 6).

```

Query= P.guajava microsatellite, mPgCIR001
Length=367
Sequences producing significant alignments:
Chr06                                     Score      E
                                         (Bits)     Value
                                         303       7e-81

> Chr06
Length=57472304

Score = 303 bits (164), Expect = 7e-81
Identities = 303/368 (82%), Gaps = 17/368 (5%)
Strand=Plus/Minus

Query  2      ACTCTGCACACCATATTTGGCATTGAGCTCTCCCTAGTGCTTTGGTTGCTTGGAAAAGACC 61
      |||
Sbjct  9586660  ACTCTGCACACCATATTTGGCATTGAGCTCTCCCTGTGCTTTGGTGCTTGGAAAAGACC 9586601

Query  62      ATCGGCATATCCTGAGAAGTTGAGCATTGGGTTTAAGAGTAAC TTATGCTGCGCCAGTT 121
      |||
Sbjct  9586600  GTCGGCATATCTTGAGAAATTGAGCATTGGGTTTAAGACTAAC TTATGCGCCACAGTT 9586541

Query  122     GCAGCACAGAGTGAGACCATTGGCATAAGAAGATGAGTATTCGGGTTTTGCGGgagagag 181
      |||
Sbjct  9586540  GCAGCA--GAGGAAGACCATCGGCATATCTTGA-GAA-ATT-GAGCAGAA-GGG-GTTTG 9586488

Query  182     agagagagagagagagagagagagagagagTTG--CATGTAAGGCGGTTCCGTTTCATTTTC 239
      |||
Sbjct  9586487  CCAGAGAGAGAGAGAGAGATA-A-AGAGTTGTCATGTAAGGC-GTTCTGTTTCATTTTC 9586431

Query  240     TTTTACTTATTTCCGACCT-T-ATATCCACCTGCGTAGGACATGATCATCTGATTACAT 297
      ||
Sbjct  9586430  TTCTTCTTATTTCCGACCTGTAATATCCACCTATGTAGAATATGATCATCTGATCACAT 9586371

Query  298     TCTTTCTGAAATTCCTATTGGGCTTCTGGTTAC---GTGAGTTGAGCTCAATTGGGTGAT 354
      ||
Sbjct  9586370  TCATTTGTAAATCCATATTGGGTTCTGATTACTGCGTGAGTTCAAGCTCAATTGGGTGAC 9586311

Query  355     GATTGTCC 362

```

Figura 6 Exemplo de alinhamento de SSR mPgCIR001 de *P. guajava* no genoma de *E. grandis*, cromossomo 6, obtido pelo BLAST (Altschul *et al.*, 1997): bit score – 164, e-value – 73×10^{-81} e identidade – 82% (303/368).

MEDEIROS, F. L. B. Desenvolvimento de mapas genéticos e físicos com base em SNPs em *Psidium guajava* e triagem *in silico* de loci microssatélites de goiabeira para transferibilidade em *Eucalyptus globulus*

3 REFERÊNCIAS BIBLIOGRÁFICAS

Acuña CV, Fernandez P, Villalba PV, García MN, Hopp HE and Poltri MSN (2012) Discovery, validation, and *in silico* functional characterization of EST-SSR markers in *Eucalyptus globulus*. **Tree Genetics and Genomes** **8**: 289–301.

Akond M, Liu S, Schoener L, Anderson JA, Kantartxi, SK, Meksem K, Song Q, Wang D, Wen Z, Lightfoot DA eKassem MA (2013) A SNP-Based Genetic Linkage Map of Soybean Using the SoyS- NP6K Illumina Infinium BeadChip Genotyping Array. **Journal of Plant Genome Sciences** **3**: 80–89.

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic Local Alignment Search Tool. **Journal of Molecular Biology** **215**: 403–410.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Research** **25**: 3389–3402.

Aranguren Y, Briceño A and Fermin G (2010) Assessment of the variability of Venezuelan guava landraces by microsatellites. **Acta Horticulturae** **849**: 147–154.

Arber W (1964) Host-controlled modification of Bacteriophage. **Annual Review of Microbiology** **19**: 365–378.

Arends D, Prins P, Jansen RC and Broman KW (2010) R/qtl: High-throughput multiple QTL mapping. **Bioinformatics** **26**: 2990–2992.

Bagali Prashanth G, Prabhu AH, Raghavendra K, Bagali Pramod G, Hittalmani S and Vadivelu JS (2010) Application of molecular markers in plant tissue culture. **Asia-Pacific Journal of Molecular Biology and Biotechnology** **18**: 83–85.

Ball AD, Stapley J, Dawson DA, Birkhead TR, Burke T and Slate J (2010) A comparison of SNPs and microsatellites as linkage mapping markers: Lessons from the zebra finch (*Taeniopygia guttata*). **BMC Genomics** **11**: 218 - 233.

Barbosa FR and Lima MF (2010) **A cultura da goiaba**. Embrapa, 180p.

Boopathi NM (2020) **Genetic Mapping and Marker Assisted Selection**. Springer US, New York, 293p.

Bouvier L, Bourcy M, Boulay M, Tellier M, Guérif P, Denancé C, Durel CE and Lespinasse Y

MEDEIROS, F. L. B. Desenvolvimento de mapas genéticos e físicos com base em SNPs em *Psidium guajava* e triagem *in silico* de loci microssatélites de goiabeira para transferibilidade em *Eucalyptus globulus*

(2012) A new pear scab resistance gene Rvp1 from the European pear cultivar “Navara” maps in a genomic region syntenic to an apple scab resistance gene cluster on linkage group 2. **Tree Genetics and Genomes** **8**: 53–60.

Bramley PM (2000) Is lycopene beneficial to human health? **Phytochemistry** **54**: 233–236.

Briceño A, Aranguren Y and Fermin G (2010) Assessment of Guava-Derived SSR Markers for the Molecular Characterization of Myrtaceae from Different Ecosystems in Venezuela. **Acta Horticulturae** **849**: 139-142.

Buetow KH and Chakravarti A (1987) Multipoint Gene mapping using Seriation. **American Journal of Human Genetics** **41**: 180–188.

Bustin SA, Mueller R, Nolan T (2020) Parameters for Successful PCR Primer design. In Biassoni R, Raso A (eds) **Quantitative Real-time PCR. Methods in Molecular Biology**. Humana Press, New York, p.5-22.

Butler JB, Vaillancourt RE, Potts BM, Lee DJ, King GJ, Baten A, Shepherd M and Freeman JS (2017) Comparative genomics of Eucalyptus and Corymbia reveals low rates of genome structural rearrangement. **BMC Genomics** **18**: 1–13.

Campos LZ de O (2010) Etnobotânica do gênero *Psidium* L. (Myrtaceae) no Cerrado Brasileiro. **Dissertação de Mestrado**, 71p.

Carneiro MS and Vieira MLC (2002) Mapas genéticos em plantas. **Bragantia** **61**: 89–100.

Chaithanya MVN, Dinesh MR, Vasugi C, Reddy DCL, Sailaja D and Aswath C (2014) Assessment of genetic diversity in guava (*Psidium guajava*) germplasm using microsatellites. **Journal of Horticultural Science** **9**: 117–125.

Chang SK, Alasalvar C and Shahidi F (2019) Superfruits: Phytochemicals, antioxidant efficacies, and health effects—A comprehensive review. **Critical Reviews in Food Science and Nutrition** **59**: 1580–1604.

Cheema J and Dicks J (2009) Computational approaches and software tools for genetic linkage map estimation in plants. **Briefings in Bioinformatics** **10**: 595–608.

Collard BCY, Jahufer MZZ, Brouwer JB and Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. **Euphytica** **142**: 169–196.

MEDEIROS, F. L. B. Desenvolvimento de mapas genéticos e físicos com base em SNPs em *Psidium guajava* e triagem *in silico* de loci microssatélites de goiabeira para transferibilidade em *Eucalyptus globulus*

Corrêa LC, Santos CAF, Lima GPP, Rodrigues MA and Costa TPP (2011) Similaridade genética entre acessos de goiabeiras e araçazeiros baseada em marcadores moleculares AFLP. **Revista Brasileira de Fruticultura** **33**: 859–867.

Costa SR (2013) Divergência genética em acessos de psidium guajava e avaliação da resistência de híbrido interespecífico de psidium ao nematóide *Meloidogyne enterolobbi*. **Dissertação de Mestrado**, 62p.

Costa SR and Santos CAF (2017) Genetic divergence among *Psidium* accessions based on single nucleotide polymorphisms developed for *Eucalyptus*. **Genetics and Molecular Research** **16**: 1–9.

Costa SR, Santos CAF and da Silva DOM (2016) Confirmação de híbridos interespecíficos no gênero *Psidium* por meio de marcadores microssatélites (SSR) In **IV Congresso Brasileiro de Recursos Genéticos**, Curitiba – Paraná, p. 315.

Creighton HB and McClintock B (1931) The Correlation of Cytological and Genetical Crossing-Over in *Zea Mays*. A Corroboration. **Proceedings of the National Academy of Sciences** **21**: 148–150.

Danna K, Nathans D and Roberts RJ (1971) Specific cleavage of simian virus 40 DNA by restriction endonuclease of hemophilus influenzae. **Reviews in Medical Virology** **9**: 75–81.

Doerge RW and Weir BS (1994) Constructing Genetic Maps By Rapid Chain Delineation. **Journal of Quantitative Trait Loci** **2**: 1–14.

Fagundes BS, da Silva LF, Giacomini RM, Secco D, Díaz-Cruz JA and Da-Silva PR (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** **9**: 408–422.

Faleiro FG (2007) Marcadores Genético-moleculares aplicados a programas de conservação e uso de recursos genéticos. **Embrapa Informações Tecnológicas** 99p.

Falk CT (1989) A simple scheme for preliminary ordering of multiple loci: application to 45 CF families. In: Elston RC, Spence MA, Rode SE, Mcclue JW (eds) **Multipoint mapping and linkage based upon affected pedigree members**. Liss, New York, p.17-22.

Fan L, Zhang MY, Liu QZ, Li LT, Song Y, Wang LF, Zhang SL and Wu J (2013)

MEDEIROS, F. L. B. Desenvolvimento de mapas genéticos e físicos com base em SNPs em *Psidium guajava* e triagem *in silico* de loci microssatélites de goiabeira para transferibilidade em *Eucalyptus globulus*

Transferability of Newly Developed Pear SSR Markers to Other Rosaceae Species. **Plant Molecular Biology Reporter** **31**: 1271–1282.

FAOSTAT 2020. Food and Agriculture Organization of the United Nations. Disponível em: <http://www.fao.org/faostat/en/#data>. Acesso em 20 mai 2021.

Flora do Brasil 2020. Jardim Botânico do Rio de Janeiro. Disponível em: < <http://floradobrasil.jbrj.gov.br/> >. Acesso em: 08 mai. 2021

Fehr WR (1987) **Principles of Cultivar Development: theory and technique** Macmillan Publishing Company, New York, 536p.

Feria-Romero IA, Astudillo-De La Vega H, Chavez-Soto MA, Rivera-Arce E, López M, Serrano H and Lozoya X (2009) RAPD markers associated with quercetin accumulation in *Psidium guajava*. **Biologia Plantarum** **53**: 125–128.

Ferreira ME and Grattapaglia D (1998) **Introdução ao uso de marcadores moleculares em análise genética**. Embrapa CENARGEN, Brasília, 220p.

Filho AG, de Oliveira JG, Viana AP, de Oliveira Siqueira AP, Oliveira MG and Pereira MG (2010) Marcadores moleculares RAPD e descritores morfológicos na avaliação da diversidade genética de goiabeiras (*Psidium guajava* L.). **Acta Scientiarum - Agronomy** **32**: 627–633.

Fortes ACR, Oliveira M do SP de, Oliveira NP de, Sanches E de NM and Cunha EFM (2016) Transferibilidade de locos microssatélites desenvolvidos em outras espécies de palmeiras para *Astrocaryum vulgare* Mart. **Revista de Ciências Agrárias** **59**: 80–86.

Franzon RC, Campos LZO, Proença CEB, Sousa-Silva JC (2009) **Araçás do gênero *Psidium*: principais espécies, ocorrência, descrição e uso**. Embrapa Cerrados, Planaltina, 48p.

Gonzaga Neto L and Soares JM (1994) **Goiaba para exportação: aspectos técnicos da produção**. EMBRAPA-SPI, Brasília, 49p.

Govaerts R, Sobral M, Ashton P, Barrie F, Holst BK, Landrum LL, Matsumoto K, Mazine FF, Lughadha EN, Proença C, Soares-Silva LH, Wilson PG and Lucas E (2008) **World Checklist of Myrtaceae**. Royal Botanic Gardens, Kew, 455p.

Grattapaglia D, Mamani EMC, Silva-Junior OB and Faria DA (2015) A novel genome-wide microsatellite resource for species of *Eucalyptus* with linkage-to-physical correspondence on the reference genome sequence. **Molecular Ecology Resources** **15**: 437–448.

MEDEIROS, F. L. B. Desenvolvimento de mapas genéticos e físicos com base em SNPs em *Psidium guajava* e triagem *in silico* de loci microssatélites de goiabeira para transferibilidade em *Eucalyptus globulus*

Grossi LL, Fernandes M, Silva MA, de Oliveira Bernardes C, Tuler AC, dos Santos PHD, Ferreira A and da Silva Ferreira MF (2021) DArTseq-derived SNPs for the genus *Psidium* reveal the high diversity of native species. **Tree Genetics and Genomes** **17**

Grover A and Sharma PC (2016) Development and use of molecular markers: Past and present. **Critical Reviews in Biotechnology** **36**: 290–302.

Gutiérrez RMP, Mitchell S and Solis RV (2008) *Psidium guajava*: A review of its traditional uses, phytochemistry and pharmacology. **Journal of Ethnopharmacology** **117**: 1–27.

Haldane JBS (1919) The combination of linkage values and the calculation of distances between linked factors. **Journal of Genetics** **8**: 299–309.

Haley CS and Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. **Heredity** **69**: 315–324.

Holetz FB, Pessini GL, Sanches NR, Cortez AG, Nakamura CV, Prado B and Filho D (2002) Screening Pl Medicinais 2.Pdf. **Mem Inst Oswaldo Cruz** **97**: 1027–1031.

Hudson CJ, Kullam ARK, Freeman JS, Faria DA, Grattapaglia D, Kilian A, Myburg AA, Potts BM and Vaillancourt RE (2012) High synteny and colinearity among *Eucalyptus* genomes revealed by high-density comparative genetic mapping. **Tree Genetics and Genomes** **8**: 339–352.

IBGE - Instituto Brasileiro de Estatística e Geografia (2019). Área destinada à colheita, área colhida e rendimento médio da produção da lavoura permanente. Disponível em: <<https://sidra.ibge.gov.br/tabela/5457#resultado>>. Acesso em: 19 abr. 2020.

Jansen RC (1993) Interval mapping of multiple quantitative trait loci. **Genetics** **135**: 205–211.

Kalendar R, Flavell AJ, Ellis THN, Sjakste T, Moisy C and Schulman AH (2011) Analysis of plant diversity with retrotransposon-based molecular markers. **Heredity** **106**: 520–530.

Kalia RK, Rai MK, Kalia S, Singh R and Dhawan AK (2011) Microsatellite markers: An overview of the recent progress in plants. **Euphytica** **177**: 309–334.

Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. **Science** **220**: 671–680.

Kosambi DD (1943) The estimation of map distances from recombination values. **Annals of**

MEDEIROS, F. L. B. Desenvolvimento de mapas genéticos e físicos com base em SNPs em *Psidium guajava* e triagem *in silico* de loci microssatélites de goiabeira para transferibilidade em *Eucalyptus globulus*

Eugenics **12**: 172–175.

Kuleung C, Baenziger PS and Dweikat I (2004) Transferability of SSR markers among wheat, rye, and triticale. **Theoretical and Applied Genetics** **108**: 1147–1150.

Lander ES and Botstein S (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics** **121**: 185.

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE and Newburg L (1987) MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. **Genomics** **1**: 174–181.

Lepitre V, Nansot G, Grangeon R, Pomies V, Rivallan R, Risterucci AM, Valdés-Infante J, Rodríguez-Medina NN, Muth J, Boike J, Prüfer D, Becker D, Rohde W, Ritter E and Billotte N (2010) The microsatellite (SSR)/AFLP reference linkage map of guava. **Acta Horticulturae** **849**: 183–192.

Liu X and Yang G (2012) Assessment of clonal fidelity of micro-propagated guava (*Psidium guajava*) plants by ISSR markers. **Australian Journal of Crop Science** **6**: 291–295.

Manly KF, Cudmore RH and Meer JM (2001) Map Manager QTX, cross-platform software for genetic mapping. **Mammalian Genome** **12**: 930–932.

Margarido GRA, Souza AP and Garcia AAF (2007) OneMap: Software for genetic mapping in outcrossing species. **Hereditas** **144**: 78–79.

Mason AS (2015) SSR Genotyping. In: Batley J (eds) **Plant Genotyping methods in Molecular Biology (Methods and Protocols)**. Humana Press, New York, p.77-89.

Masud Parvez GM, Uzzaman S, Miah Akanda K and Mehjabin S (2018) A short review on a Nutritional Fruit : Guava. **Open Access: Toxicology & Research** **1**: 1–8

Mateos-Hernandez M, Singh RP, Hulbert SH, Bowden RL, Huerta-Espino J, Gill BS and Brown-Guedira G (2006) Targeted mapping of ESTs linked to the adult plant resistance gene Lr46 in wheat using synteny with rice. **Functional and Integrative Genomics** **6**: 122–131.

McCallum S, Graham J, Jorgensen L, Rowland LJ, Bassil N V., Hancock JF, Wheeler EJ, Vining K, Poland JA, Olmstead JW, Buck E, Wiedow C, Jackson E, Brown A and Hackett CA (2016) Construction of a SNP and SSR linkage map in autotetraploid blueberry using genotyping by sequencing. **Molecular Breeding** **36**: 1–24.

MEDEIROS, F. L. B. Desenvolvimento de mapas genéticos e físicos com base em SNPs em *Psidium guajava* e triagem *in silico* de loci microssatélites de goiabeira para transferibilidade em *Eucalyptus globulus*

Meng L, Li H, Zhang L and Wang J (2015) QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. **Crop Journal** **3**: 269–283.

Meselson M and Yuan R (1968) DNA restriction enzyme from *E. coli*. **Nature** **217**: 1110–1114.

Mistura CC, Barbieri RL, Castro CM, Priori D, Castelo J and Villela B (2012) Transferibilidade de marcadores microssatélites de coco (*Cocos nucifera*) para butiá (*Butia odorata*) **Magistra** **24**: 360-369.

Mnejja M, Garcia-Mas J, Audergon JM and Arús P (2010) Prunus microsatellite marker transferability across rosaceous crops. **Tree Genetics and Genomes** **6**: 689–700.

Morgan TH (1910) Sex Limited Inheritance in *Drosophila*. **Science** **32**: 120–122.

Negi SS and Rajan S (2007) Improvement of guava through breeding. **Acta Horticulturae** **735**: 31–37.

Nimisha S, Kherwar D, Ajay KM, Singh B and Usha K (2013) Molecular breeding to improve guava (*Psidium guajava* L.): Current status and future prospective. **Scientia Horticulturae** **164**: 578–588.

Nogueira AM, Ferreira A and da Silva Ferreira MF (2016) Transferability of Microsatellites from *Psidium guajava* to *Eugenia*, *Myrciaria*, *Campomanesia*, and *Syzygium* Species (*Myrtaceae*). **Plant Molecular Biology Reporter** **34**: 249–256.

Nomelini QSS, Silva HD and Faria TC (2009) Eficiência dos métodos de otimização Simulated annealing, Delineação rápida em cadeia e Ramos e conexões para construção de mapas genéticos. **Ciênc. agrotec., Lavras** **3**: 1534–1537.

Oliveira EJ, Pádua JG, Zucchi MI, Vencovsky R and Vieira MLC (2006) Origin, evolution and genome distribution of microsatellites. **Genetics and Molecular Biology** **29**: 294–307.

Padmakar B, Kanupriya C, Latha PM, Prashant KS, Dinesh MR, Sailaja D and Aswath C (2015) Development of SRAP and SSR marker-based genetic linkage maps of guava (*Psidium guajava* L.). **Scientia Horticulturae** **192**: 158–165.

Paterson AH, Bowers JE, Burow MD, Draye X, Elsiek CG, Jiang CX, Katsar CS, Lan TH, Lin YR, Ming R and Wright RJ (2000) Comparative genomics of plant chromosomes. **Plant Cell**

MEDEIROS, F. L. B. Desenvolvimento de mapas genéticos e físicos com base em SNPs em *Psidium guajava* e triagem *in silico* de loci microssatélites de goiabeira para transferibilidade em *Eucalyptus globulus*

12: 1523–1539.

Pereira FM and Nachtigal JC (2002) **Melhoramento da goiabeira** Bruckner CH (ed). UFV, 53–78p.

Pertsemlidis A and Fondon JW (2001) Having a BLAST with bioinformatics (and avoiding BLAST phemy). **Genome Biology 2:** 1–10.

Pommer C (2012) Guava world-wide breeding: major techniques and cultivars and future challenges. **Acta Horticulturae 959:** 81–88.

Pommer C, Murakami KRN (2009) Breeding Guava (*Psidium guajava* L.). In Jain SM, Priyadarshan (eds) **Breeding Plantation Tree Crops: Tropical Species**. Springer, New York, p.83-120.

Pommer C, Odaci O and Santos CAF (2013) **Goiaba: Recursos genéticos e melhoramento**. EdUfersa, Mossoró, 126p.

Quintal SSR, Viana AP, Campos BM, Vivas M and Amaral Júnior AT (2017) Seleção via modelos mistos em famílias segregantes de goiabeira baseada em características de produtividade e qualidade. **Revista Brasileira de Fruticultura 39.**

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. **Current Opinion in Plant Biology 5:** 94–100.

Rai MK, Phulwaria M and Shekhawat NS (2013) Transferability of simple sequence repeat (SSR) markers developed in guava (*Psidium guajava* L.) to four Myrtaceae species. **Molecular Biology Reports 40:** 5067–5071.

Rallo P, Tenzer I, Gessler C, Baldoni L, Dorado G and Martín A (2003) Transferability of olive microsatellite loci across the genus *Olea*. **Theoretical and Applied Genetics 107:** 940–946.

Ribeiro HLC, Santos CAF, Rodrigues MA, Araújo JS, Costa TPP, Oliveirai MM de, Oliveira VR de (2008) Transferência de primers SSRs da mangueira (*Mangifera indica*) para o umbuzeiro (*Spondias tuberosa* Arr.). In: **Jornada de Iniciação Científica Embrapa Semi-árido**.

Risterucci AM, Duval MF, Rohde W and Billotte N (2005) Isolation and characterization of microsatellite loci from *Psidium guajava* L. **Molecular Ecology Notes 5:** 745–748.

MEDEIROS, F. L. B. Desenvolvimento de mapas genéticos e físicos com base em SNPs em *Psidium guajava* e triagem *in silico* de loci microssatélites de goiabeira para transferibilidade em *Eucalyptus globulus*

Risterucci AM, Nansot G, Grangeon R, Lepitre V, De Reeper A, Argout X, Ruiz M and Billotte N (2010) Development of guava microsatellite (SSR) markers using the SAT software. **Acta Horticulturae 849**: 113–120.

Ritter E (2012) Guava biotechnologies, genomic achievements and future needs. **Acta Horticulturae 959**: 131–140.

Ritter E, Lucca F, Sánchez I, Ruiz de Galarreta JI, Aragonés A, Castañón S, Bryan G, Waugh R, Lefebvre V, Rouselle-Bourgoise F, Gebhardt C, van Eck H, van Os H, Taco J, Baker K (2005) Genomic resources in potato and possibilities for exploitation. In: Haverkort AJ, Struik PC (eds) **Potato in progress, science meets practice**. Wageningen Acad. Pub., pp.55-65.

Ritter E, Rodríguez-Medina NN, Velásquez B, Rivero D, Rodríguez JA, Martínez F and Valdés-Infante J (2010) QTL (quantitative trait loci) analysis in guava. **Acta Horticulturae 849**: 193–202.

Ritter E, Herran A, Valdés-Infante J, Rodríguez-Medina NN, Briceño A, Fermin G, Sanchez-Teyer F, O’connor-Sanchez A, Muth J, Boike J, Prüfer D, Santos CA, Dos Santos ICN, Rodrigues MA, Risterucci AM, Billotte N, Becker D and Rohde W (2010) Comparative linkage mapping in three guava mapping populations and construction of an integrated reference map in guava. **Acta Horticulturae 849**: 175–182.

Rodríguez N, Valdés-Infante J, Becker D, Velázquez B, González G, Sourd D, Rodríguez J, Billotte N, Risterucci AM, Ritter E and Rohde W (2007) Characterization of guava accessions by ssr markers, extension of the molecular linkage map, and mapping of qtls for vegetative and reproductive characters. **Acta Horticulturae 735**: 201–215.

Sakiyama NS, Ramos HCC, Caixeta ET and Pereira MG (2014) Plant breeding with marker-assisted selection in Brazil. **Crop Breeding and Applied Biotechnology 14**: 54–60.

Sánchez-Teyer LF, Barraza-Morales A, Keb L, Barredo F, Quiroz-Moreno A, O’Connor-Sánchez A and Padilla-Ramírez JS (2010) Assessment of genetic diversity of Mexican guava germplasm using DNA molecular markers. **Acta Horticulturae 849**: 133–138.

Santos CAF, Correa LC, Leonardo H, Ribeiro C, Oliveira MM De and Araujo JS (2009) Transferência de Primers SSRs de *Allium Fistulosum* Para A. Ceba. **Horticultura Brasileira 27**: 3419–3422.

Santos CAF, Costa SR and Souza RRC (2017) **Brs Guaraçá**.

MEDEIROS, F. L. B. Desenvolvimento de mapas genéticos e físicos com base em SNPs em *Psidium guajava* e triagem *in silico* de loci microssatélites de goiabeira para transferibilidade em *Eucalyptus globulus*

Schuster I (2011) Marker-assisted selection for quantitative traits. **Crop Breeding and Applied Biotechnology 11**: 50–55.

Silva-Junior OB, Faria DA and 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.

Sitther V, Zhang D, Harris DL, Yadav AK, Zee FT, Meinhardt LW and Dhekney SA (2014) Genetic characterization of guava (*Psidium guajava* L.) germplasm in the United States using microsatellite markers. **Genetic Resources and Crop Evolution 61**: 829–839.

Smith HO and Welcox KW (1970) A Restriction enzyme from *Hemophilus influenzae*. I. Purification and general properties. **Journal of Molecular Biology 51**: 379–391.

Souza AP (2001) Biologia molecular aplicada ao melhoramento. In: Nass LL, Valois ACC, Melo IS, Valadares-inglis MC (eds) **Recursos Genéticos e melhoramento**. Fundação MT, Rondonópolis, p. 939-965.

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JOINMAP. **Plant Journal 3**: 739–744.

Staub JE, Serquen FC and Gupta M (1996) Genetic markers, map construction, and their application in plant breeding. **HortScience 31**: 729–740.

Sturtevant A (1913) The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. **Journal of Experimental Zoology 14**: 43–59.

Sumathi M, Bachpai VKW, Mayavel A, Dasgupta MG, Nagarajan B, Rajasugunasekar D, Sivakumar V and Yasodha R (2018) Genetic linkage map and QTL identification for adventitious rooting traits in red gum eucalypts. **3 Biotech 8**: 242.

Tanksley SD (1993) Mapping Polygenes. **Annual Review of Genetics 27**: 205–233.

Thoday J (1961) Location of Polygenes. **Nature 191**: 368–370.

Valdés-infante J, Becker D, Rodriguez N, Velasquez B, Gonzalez G, Sourd D, Rodriguez J, Ritter E, Rohde W (2003) Molecular characterization of Cuban accessions of guava (*Psidium guajava* L) establishment of a first molecular linkage map and mapping of QTLs for vegetative characters. **Journal of Genetics of Breeding 57**: 349–358.

Valdés-infante J, Rodríguez NN, Becker D, Velázquez B, Sourd D, Espinosa G and Rohde W

MEDEIROS, F. L. B. Desenvolvimento de mapas genéticos e físicos com base em SNPs em *Psidium guajava* e triagem *in silico* de loci microssatélites de goiabeira para transferibilidade em *Eucalyptus globulus*

(2007) Microsatellite characterization of guava (*Psidium guajava* L.) germplasm collection in Cuba. **Cultivos Tropicales** **28**: 61–67.

Valdés-Infante J, Rodríguez NN, Velásquez B, Rivero D, Martínez F, Espinosa G, Risterucci AM, Billotte N, Becker D and Rohde W (2010) Simple sequence repeats (SSRs) for diversity characterization of guava (*Psidium guajava* L.). **Acta Horticulturae** **849**: 155–162.

Van Ooijen JW (2006) JoinMap®4, Software for the Calculation of Genetic Linkage Maps in Experimental Populations. Kyazma BV, Wageningen.

Van Ooijen JW (2004) MapQTL® 5. Software for the mapping of quantitative trait loci in experimental populations. Kyazma BV, Wageningen.

Van Ooijen JW, Jansen J (2013) **Genetic mapping in experimental populations**. Cambridge University Press, Cambridge, 139p.

Van Os H, Stam P, Visser RGF and Van Eck HJ (2005) RECORD: A novel method for ordering loci on a genetic linkage map. **Theoretical and Applied Genetics** **112**: 30–40.

Varshney RK, Sigmund R, Börner A, Korzun V, Stein N, Sorrells ME, Langridge P and Graner A (2005) Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. **Plant Science** **168**: 195–202.

Varshney RK, Thiel T, Stein N, Langridge P and Graner A (2002) *In Silico* Analysis on Frequency and Distribution of microsatellites in ESTs of some cereal species. **Cellular & Molecular Biology letters** **7**: 537–546.

Vieira MLC, Santini L, Diniz AL and Munhoz C de F (2016) Microsatellite markers: What they mean and why they are so useful. **Genetics and Molecular Biology** **39**: 312–328.

Wang M, Barkley N and Jenkins T (2009) Microsatellite markers in plants and insects. Part I: Applications of biotechnology. **Genes, Genomes and Genomics** **3**: 54–67.

Watson JD and Crick F (1953) Molecular Structure of Nucleic Acids. **Nature** **171**: 737–738.

Weeks DE and Lange K (1987) Preliminary ranking procedures for multilocus ordering. **Genomics** **1**: 236–242.

Wei L, Liu Y, Dubchak I, Shon J and Park J (2002) Comparative genomics approaches to study organism similarities and differences. **Journal of Biomedical Informatics** **35**: 142–150.

MEDEIROS, F. L. B. Desenvolvimento de mapas genéticos e físicos com base em SNPs em *Psidium guajava* e triagem *in silico* de loci microssatélites de goiabeira para transferibilidade em *Eucalyptus globulus*

Wilson CW, Shaw PE and Campbell CW (1982) Determination of organic acids and sugars in guava (*Psidium guajava* L.) cultivars by high-performance liquid chromatography. **Journal of the Science of Food and Agriculture** **33**: 777–780.

Wilson SR (1988) A major simplification in the preliminary ordering of linked loci. **Genetic Epidemiology** **5**: 75–80.

Wu F, Eannetta NT, Xu Y, Durrett R, Mazourek M, Jahn MM and Tanksley SD (2009) A COSII genetic map of the pepper genome provides a detailed picture of synteny with tomato and new insights into recent chromosome evolution in the genus *Capsicum*. **Theoretical and Applied Genetics** **118**: 1279–1293.

Wu R, Ma CX, Painter I and Zeng ZB (2002) Simultaneous Maximum Likelihood Estimation of Linkage and Linkage Phases in Outcrossing Species. **Theoretical Population Biology** **61**: 349–363.

Zeng Z-B (1994) Precision Mapping of Quantitative Trait Loci. **Genetics** **136**: 492–496.

Zeng Z-B (1993) Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. **Proceedings of the National Academy of Sciences of the United States of America** **90**: 10972–10976.

Zhang L, Meng L, Wu W and Wang J (2015) GACD: Integrated software for genetic analysis in clonal fl and double cross populations. **Journal of Heredity** **106**: 741–744.

Zhao N, Yu X, Jie Q, Li Hui, Li Hua, Hu J, Zhai H, He S and Liu Q (2013) A genetic linkage map based on AFLP and SSR markers and mapping of QTL for dry-matter content in sweetpotato. **Molecular Breeding** **32**: 807–820.

CAPÍTULO II

DESENVOLVIMENTO DE MAPAS DE LIGAÇÕES GENÉTICAS E FÍSICOS COM BASE EM SNPs DE EUCALYPTUS EM *Psidium guajava*

RESUMO

O mapeamento de um genoma é uma representação gráfica simplificada de marcadores ou sequências de nucleotídeos nos cromossomos. Existem dois tipos de mapas: os físicos e os mapas de ligações genéticas. O desenvolvimento de mapas densos e precisos é necessário para a aplicação da seleção assistida em espécies vegetais, como a goiaba. Na literatura científica não há relatos de mapas físicos para *P. guajava* nem do desenvolvimento de mapas genéticos com base em SNPs. Desta forma, o objetivo deste estudo foi desenvolver e comparar mapas de ligações genéticas, com base nos programas JoinMap 4.0 e GACD, bem como mapas físicos, resultantes da ordenação de SNPs de *Eucalyptus*, via BLAST, no genoma da goiabeira, para servirem de referências para o mapeamento de caracteres quantitativos e qualitativos em populações de goiabeira. Foram genotipados 112 indivíduos do cruzamento Pedro Sato × Goiabeira Roxa (PSR), utilizando o Euchip60K, com 72.202 SNPS, dos quais 79% foram monomórficos. Após filtragem foram utilizados 1120 marcadores para análises de mapeamento. O mapa JoinMap 4.0 apresentou 203 marcadores nos onze grupos de ligação de *P. guajava* (n=11), cobrindo 1.405,2 cM com intervalo médio de 7,7 cM. O mapa GACD apresentou 186 marcadores cobrindo 1.392,7 cM com intervalo médio de 8,8 cM. Os programas JoinMap e GACD não apresentaram concordância nas distâncias estimadas e também na ordenação dos SNPs nos grupos de ligações. Os mapas gerados pelo JoinMap 4.0 e pelo GACD apresentaram grupos com segmentos de vários cromossomos, quando comparados com o mapa físico, indicando limitações. O GACD apresentou maior limitação em relação ao JoinMap 4.0, ao separar os marcadores de acordo com sua origem parental. O mapa físico gerado com BLASTn, e-values variando de 8×10^{-10} a $1,15 \times 10^{-26}$, consistiu de 694 SNPs, cobrindo 434,88 Mb, com distâncias médias de 0,62 Mb, sendo referência para estudos de mapeamento e estimativas de QTLs em goiabeira.

Palavras-chave: mapa de ligação, mapa físico, BLAST, *P. guajava*.

ABSTRACT

Mapping of a genome is a simplified representation of molecular markers or nucleotide sequences in the chromosome. There are two types of maps: physical and linkage maps. The development of dense and precise maps is important to molecular-assisted selection in plants, such as guava. There are no reports on physical mapping in guava or linkage mapping based in SNP markers. Thus, the goal of this study was to develop and compare linkage maps, based on softwares JoinMap 4.0 and GACD, as well as physical maps from BLAST and ordering *Eucalyptus* SNPs against guava genome, to serve as reference for qualitative and quantitative traits mapping in guava populations. Genotyping was carried out with 112 individuals from a Pedro Sato x Goiabeira Roxa (PSR) cross and using Euchip60K with 72.202 SNPS, from which 79% were monomorphic. After filtering, 1120 markers were applied for mapping analysis. JoinMap 4.0 map contained 203 markers along the eleven linkage groups of *P. guajava* (n=11) spanning a total distance of 1.405,2 cM with a mean distance of 7.7 cM. GACD map presented 186 markers spanning a distance of 1392.7 cM with mean interval of 8.8 cM. JoinMap and GACD programs did not show agreement in the estimated distances neither in SNP ordering. Both maps presented linkage groups with segments of several chromosomes when compared to the physical map, indicating linkage mapping limitations. GACD demonstrated greater limitation in comparison to JoinMap 4.0 because it ordered SNPs according to their parental origin. The physical map generated through BLASTn, e-values ranging from 8×10^{-10} to 1.15×10^{-26} , consisted in 694 SNPs covering a total of 434.88 Mb with a mean interval of 0.62 Mb, being a reference for mapping and QTL estimates in guava.

Keywords: linkage map, physical map, BLAST, *P. guajava*.

INTRODUÇÃO

A goiabeira (*Psidium guajava* - Myrtaceae) é uma das culturas de destaque na fruticultura brasileira, tendo como centro de origem a América do Sul (Risterucci *et al.* 2005). A versatilidade de consumo, tanto *in natura* como industrializado, (Padmakar *et al.* 2015, Costa e Santos 2017) é um atrativo para a produção e comercialização dessa espécie, além de um alto valor nutricional (Pommer *et al.* 2013, Masud Parvez *et al.* 2018). O Brasil é um dos maiores produtores mundiais de goiaba, com as cultivares Paluma e Pedro Sato como as mais importantes. A importância econômica da goiaba tem crescido e, de maneira análoga, os estudos com enfoque na cultura. Recentemente Feng *et al.* (2020) sequenciaram o genoma de *P. guajava* (cultivar New Age) e montaram 443,8 Mb em 11 cromossomos (n=11), representando 95,7% do genoma, o que a coloca no mesmo patamar de grandes culturas como o eucalipto (Myburg *et al.* 2014).

O mapeamento de um genoma é uma representação gráfica dos marcadores ou sequências de nucleotídeos nos cromossomos (Mozo *et al.* 1999, Lander *et al.* 2001). Para tal, utilizam-se dois tipos de mapas: os físicos e os mapas de ligações genéticas (ou mapas genéticos). O mapa físico difere do genético por refletir a distância real entre nucleotídeos, enquanto o mapa genético baseia-se nas taxas de recombinações que ocorrem nos cromossomos (Bueno 2009).

O mapeamento genético, em cM, é um método tradicional e permite a clonagem de regiões genômicas, junção de mapas físicos, identificação de “hotspots” de recombinação, genômica comparativa, entre outros (Sumathi *et al.* 2018). Na literatura há relatos da aplicação de mapeamento de ligação em diferentes espécies de Myrtaceae, como eucalipto (*Eucalyptus spp.*) (Brondani *et al.* 2006, Lima *et al.* 2011, Hudson, Freeman, *et al.* 2012, Sumathi *et al.* 2018), érica-japonesa (*Leptospermum scoparium*) (Chagné *et al.* 2019). Na cultura da goiabeira (*P. guajava*) o primeiro mapa de ligação foi reportado por Valdés-Infante *et al.* (2003), posteriormente estendido por Rodríguez *et al.* (2007). Mapas de ligações baseados em SSR, AFLPs e SRAP também foram disponibilizados para *P. guajava* (Lepitre *et al.* 2010, Ritter, Herran, *et al.* 2010, Padmakar *et al.* 2015).

O mapeamento físico consiste na sobreposição de grandes insertos de clones de DNA gerando uma representação das distâncias entre bases no genoma (bp) (Soto *et al.* 2015), que podem ser mensuradas em megabases (Mb), o que corresponde a um milhão de bases (ISOGG 2017). Estes mapas podem ser ancorados a mapas genéticos através de marcadores âncoras

(Grattapaglia 2004), fornecendo uma estrutura para análise de QTL e clonagem de genes. Paiva *et al.* (2011) relataram a construção de bibliotecas baseados em “bacterial artificial chromosome” (BAC) de *E. grandis* para mapas físicos (clone BRASUZ1), montando genoma de referência para a espécie. Na literatura não foram encontrados relatos de mapas físicos ou bibliotecas BAC para a espécie *P. guajava*.

O sequenciamento de genomas tem se tornado uma prática mais acessível em razão das tecnologias de sequenciamento de nova geração (NGS), tornando a genotipagem de alto rendimento de milhares de marcadores moleculares uma realidade (Nielsen *et al.* 2011). Os “Single Nucleotide Polymorphism” (SNP) são marcadores abundantes, amplamente distribuídos no genoma e importantes para o mapeamento de ligação e seleção assistida (Bertioli *et al.* 2014). Costa e Santos (2017) avaliaram a divergência genética em *Psidium* utilizando SNPs desenvolvidos para *Eucalyptus* (Silva-Junior *et al.* 2015). Grossi *et al.* (2021) reportaram o desenvolvimento dos primeiros SNPs específicos de *P. guajava*, derivados de genotipagem DArTseq, aplicando-os na identificação e caracterização das espécies *P. guajava*, *P. cattleianum*, *P. guineense*, *P. myrtoide*s e *P. oblongatum*. Mapas físicos com base em SNPs já foram relatados por Kong *et al.* (2004) em humanos e por Blair *et al.* (2018), em feijão. Também há relatos da utilização de SNPs na integração de mapas físicos e genéticos (Han *et al.* 2009, González *et al.* 2010, Soto *et al.* 2015), mas não existem relatos de mapas físicos para goiabeira.

Segundo Hohmann *et al.* (1994) a comparação entre mapas físicos e genéticos é uma forma de combinar as informações genéticas, de mapeamento e moleculares em diferentes espécies. Esses autores compararam mapas genéticos e físicos do cromossomo 7 de trigo e cevada e observaram que a maioria das posições eram conservadas e colineares em loci homólogos. DeWan *et al.* (2002) reportaram inconsistências entre mapas físicos e mapas genéticos, nos painéis 9 e 10 do genoma humano, tendo como referência o sequenciamento de dois genomas: o Human Genome Project–Santa Cruz e o Celera. Comparações entre o mapa genético com base em SNPs de *Capsicum baccatum* e o mapa físico de *C. annuum* indicaram inconsistências, translocações ou inversões, assim como do grupo de ligações de *C. baccatum* com segmentos cromossômicos dos Chr.3, Chr.9, e Chr.5 de *C. annuum*.

A ferramenta Basic Local Alignment Search Tool (BLAST) (Altschul *et al.* 1990) é comumente utilizada na busca por homologias entre espécies, mas pode ser utilizada para ancorar marcadores em um mapa tal qual reportado por Xu *et al.* (2008) que utilizaram o

BLASTn para ancorar as sequências disponíveis de unigenes de *Gossypium hirsutum* ao mapa físico/genético dos cromossomos 12 e 26 da espécie, estratégia previamente utilizada por Cheng *et al.* (2005) em arroz. O BLAST pode ser uma alternativa para a montagem de um mapa físico parcial ou alocação de marcadores aos cromossomos, de acordo com a similaridade (hits) encontrada no genoma, conforme demonstrou Gao *et al.* (2018) ao utilizar o BLAST para localizar fisicamente scaffolds e SNPs de *Pleurotus tuoliensis*, adotando e-value de 1×10^{-3} . Blair *et al.* (2018) reportaram múltiplos BLAST “hits” para 66 do total 768 SNPs indicando sequências parálogas que podem provocar inconsistências na formação e ordenação dos grupos em mapas genéticos. Na literatura não foi reportado o desenvolvimento de mapas físicos com base em SNPs para *P. guajava*, impossibilitando comparações com mapas genéticos.

O objetivo do presente estudo foi desenvolver e comparar mapas de ligações genéticas, com base nos programas JoinMap 4.0 e GACD, bem como mapas físicos, resultantes da ordenação de SNPs de *Eucalyptus*, via BLAST, no genoma da goiabeira, para servirem de referências ao mapeamento de caracteres quantitativos e qualitativos em populações de goiabeira.

MATERIAIS E MÉTODOS

Obtenção e coleta do material vegetal

Foram genotipados indivíduos oriundos do cruzamento manual das cultivares Pedro Sato × Goiabeira Roxa (PSR), no Banco de Germoplasma da Embrapa Semiárido, localizado no Campo experimental de Bebedouro, Petrolina, PE. A `Pedro Sato` apresenta frutos de polpa avermelhada, enquanto a goiabeira roxa apresenta folhas frutos, caules e polpa arroxeados.

As sementes das gerações F₁, resultante deste cruzamento, foram semeadas em potes de 2 L contendo solo autoclavado e o substrato Plantmax®, na proporção de 1:1. Quando as plantas atingiram 6 a 8 cm de altura foram repicadas para sacos de polietileno de 22 x 30 cm, contendo a mesma mistura utilizada no substrato da germinação, e colocadas em viveiro telado. Após atingirem em torno de 50 cm de altura as plantas foram transplantadas, sendo plantadas lado a lado com mudas do porta-enxerto BRS Guaraçá, resistente ao nematoide, todos no espaçamento de 4 m x 4 m. Enxertia de garfagem, tipo fenda completa, foi efetuada no campo, tendo início após o quarto mês do transplante.

Folhas novas e sadias de 112 indivíduos da população de irmãos completos foram coletadas e colocadas em saco de papel, devidamente identificados, e acondicionadas em isopor

para transporte ao laboratório. Posteriormente os sacos foram armazenados em freezer -80°C até o momento da extração de DNA.

Extração e Quantificação do DNA

O DNA foi extraído de acordo com protocolo de Inglis *et al.* (2018). A maceração mecânica foi realizada na presença de nitrogênio líquido para o rompimento da parede vegetal, até a obtenção de um pó fino. O vegetal macerado de cada amostra foi transferido para tubo de eppendorf de 2 mL, contendo 900 µL solução de sorbitol [100 mM Tris-HCl pH 8.0, 0.35 M Sorbitol, 5mM EDTA pH 8.0, 1% polivinilpirrolidona (PVP- 40)] para a pré-lavagem e adição de 1% (v/v) de 2-mercaptoetanol antes da extração. As amostras foram centrifugadas a 7500 rpm durante 5 min, seguido da retirada do sobrenadante e adição de 700 µL do tampão de extração CTAB (3%) (20 mM EDTA, 100 mM Tris-HCl pH 8.0, 3 M NaCl, 3% CTAB). As amostras foram colocadas em banho Maria a 60°C durante 30 min, realizando-se a inversão dos tubos de eppendorf a cada 10 min. Em seguida foi adicionado 700 µL de clorofórmio e álcool isoamílico (24:1) às amostras e as mesmas centrifugadas a 10.500 rpm, por 10 min. Foram transferidos 600 µL do sobrenadante para um novo tubo de eppendorf de 1,5 mL, adicionando-se 396 µL de álcool isopropílico gelado e 60 µL de acetato de sódio, seguido da inversão suave dos tubos e a disposição em um isopor com gelo para repouso por 60 min. Após 1 h as amostras foram centrifugadas a 10.500 rpm para formação do “pellet” no fundo do tubo, descartando-se o sobrenadante em béquer, dentro de capela de exaustão de gases. O “pellet” foi ressuspenso em 30 µL Tris-EDTA, permanecendo por 24 h na geladeira para completa dissolução do “pellet”.

A quantificação do DNA foi realizada por espectrofotometria, utilizando-se a placa Microdrop na faixa de absorbância a 260 nm. Uma curva de regressão foi ajustada aos valores de absorbância de amostras de DNA de concentrações conhecidas. A determinação do grau de pureza foi obtida pela razão entre as leituras A260/A280, sendo utilizadas as amostras que apresentaram valor próximo ou acima de 1,8, seguido da nova extração para amostras abaixo desse valor. Após a quantificação as amostras foram diluídas para concentração de trabalho de 20 ng/µL e armazenadas a -20°C.

Genotipagem com marcadores SNPs e desenvolvimento de mapas de ligações genéticas

As amostras de DNA da população PSR foram genotipadas com o chip EUChip60K (Silva-Junior *et al.* 2015). SNPs no EUChip60K são codificados com o número do cromossomo,

seguido da localização física em pares de base na versão 1.1 de *E. grandis* e atualizado na versão 2.0 para 72.202 SNPs. A genotipagem foi realizada na GeneSeek (Lincoln, NE, USA). Marcadores monomórficos, com segregação inesperada ou com dados ausentes >20%, foram eliminados em planilha Excel. Os mapas genéticos foram identificados com as letras “A” e “C” ao final do marcador a fim de representar a origem parental no cruzamento, e em caso de codominância com “H”.

A análise de ligação realizada no JoinMap 4.0 (Van Ooijen 2006) considerou o seguinte `script`: 1) marcadores com qui-quadrado >20, para as segregações ll, lm, nn e np, e >70, para a segregação hh, hk e hk, forma eliminados das análises; 2) os dados foram codificados para opção cross-pollination (CP), hh, hk, kk, ll, lm, np e nn, como definidos no manual do programa; 3) a formação de grupos foi efetuada com base no “LOD score”, com valores variando de 5,0 a 12; 4) a função do maximum likelihood foi usada na ordenação dos marcadores de um grupo; 5) marcadores com N.N. stress (cM) > 4.0 foram eliminados de um dado grupo; 6) os valores obtidos para as frequências de recombinação foram convertidos em distância de mapa genético (centiMorgan, cM), usando a função de Kosambi (1943).

No programa GACD (Zhang *et al.* 2015) a formação dos grupos de ligação foi efetuada na janela “Grouping”, com LOD de 8,0. Na janela “Ordering” foram selecionados “k-optimality”, por LOD score, 3-OptMAP, NN Initials, 10. Na janela “Rippling” foram selecionados “Windows size 8”, por LOD. Os valores obtidos para as frequências de recombinação foram convertidos em distância de mapa genético (cM), usando a função de Kosambi (1943).

Análise *in silico* de marcadores SNPs e desenvolvimento de mapa físico em *P. guajava*

O mapa físico foi ordenado em cromossomos da goiabeira considerando o alinhamento de sequências dos SNPs de *Eucalyptus* (query) e o genoma completo da goiabeira (subject) (Feng *et al.* 2020), disponibilizado no National Center for Biotechnology Information (NCBI). Cada SNP tem, aproximadamente, 70 nucleotídeos de extensão.

Para a estimativa das posições dos marcadores SNPs utilizou-se a ferramenta BLAST, com otimização do algoritmo blastn e alteração no parâmetro “expected threshold” para 1. Os hits encontrados entre os SNPs de *Eucalyptus* e o genoma da goiabeira foram analisados em planilha Excel. Os resultados foram ordenados por cromossomo da goiabeira, do primeiro ao décimo primeiro, e na ordem crescente da coluna “subject start”, nos respectivos cromossomos. Utilizou-se o ponto de corte de e-value < E⁻¹⁰, desconsiderando-se alinhamentos de baixa ou

duvidosa significância biológica. Alinhamentos repetidos e com e-values próximos foram mantidos no mapa e identificados por “*” no mesmo cromossomo e “#” em diferentes cromossomos. Os mapas físicos foram construídos com programa MapGene2Chrom. (Jiangtao *et al.* 2015).

RESULTADOS

Dos 72.202 SNPs do chip EUChip60K (Silva-Junior *et al.* 2015), utilizados para a genotipagem da população Pedro Sato × Goiaba Roxa, 79% foram monomórficos e consequentemente não informativos para mapeamento. Após filtragem dos dados 1120 SNPs (1,6% do total inicial) foram utilizados para confecção dos mapas de ligações e do mapa físico.

Mapas de ligações genéticas

O mapa de ligação gerado pelo programa JoinMap 4.0 apresentou 203 marcadores SNPs distribuídos nos onze grupos de ligações da goiabeira, cobrindo uma distância de 1.405,2 cM. O intervalo médio entre marcadores foi de 7,7 cM. O comprimento dos grupos de ligações variou entre 49,3 (GL9) e 196,8 cM (GL7) e de 11 a 26 marcadores por grupo. A distância entre os marcadores variou de 0,1 a 28,4 cM (Figura 1).

O mapa de ligação gerado pelo programa GACD resultou em 1.392,7 cM, com 186 marcadores nos onze grupos de ligações da goiabeira. O intervalo médio entre os marcadores foi de 8,8 cM. Os grupos de ligações 5 e 2 foram os mais densos e o grupo 11 o menos denso. Em relação ao comprimento, o grupo de ligações 1 foi o menor com 49,86 cM e o grupo 5 foi o maior, 180,80 cM. A distância entre os marcadores variou de 0,94 a 37,93 cM (Figura 2).

No mapeamento com o JoinMap 4.0 foram observadas misturas de fases de atração e repulsão, com os marcadores provenientes dos diferentes pais (identificados com “C” e “A”), enquanto no mapeamento com o GACD essa mistura de fases não foi observada, tendo os marcadores de um pai ordenados de forma sequencial (Figuras 1 e 2). observou-se a presença de 91 marcadores parentais dominantes tipo “A”, 75 do tipo “C” e 37 co-dominantes (H) no mapa JoinMap 4.0. Já no mapa GACD foram observados 88 marcadores tipo “A”, 81 “C” e 17 “H” (Figuras 1 e 2). Dos 37 marcadores “H” (JoinMap 4.0), 25 não coincidiram com o mapa GACD, estando os marcadores tipo “H” não coincidentes nos GL6, G7 e G10 (Figuras 1 e 2, Tabela 1).

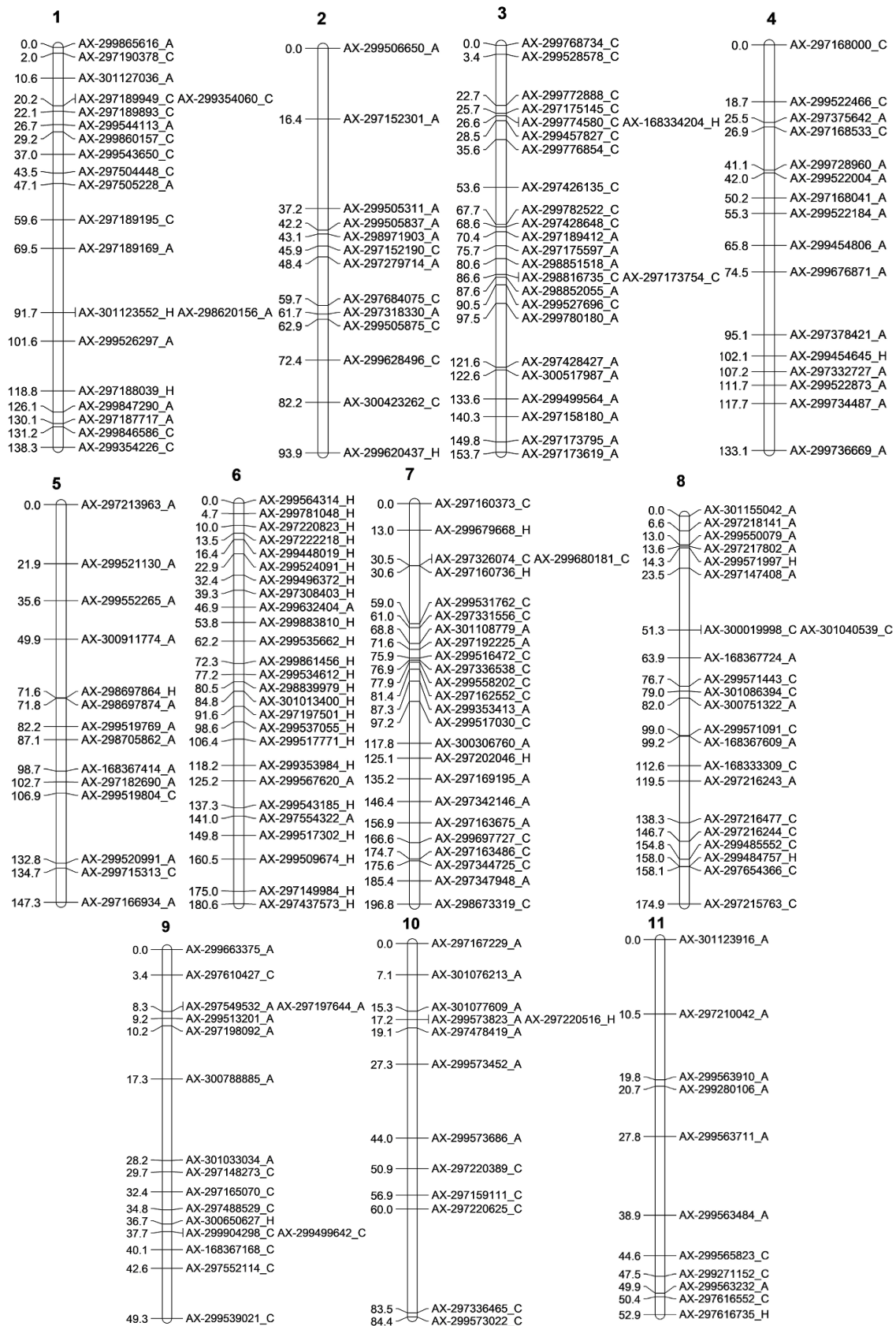


Figura 1 Mapa de ligações genéticas do cruzamento Pedro Sato × Roxa (PSR) realizado pelo JoinMap 4.0 (Van Ooijen 2006)

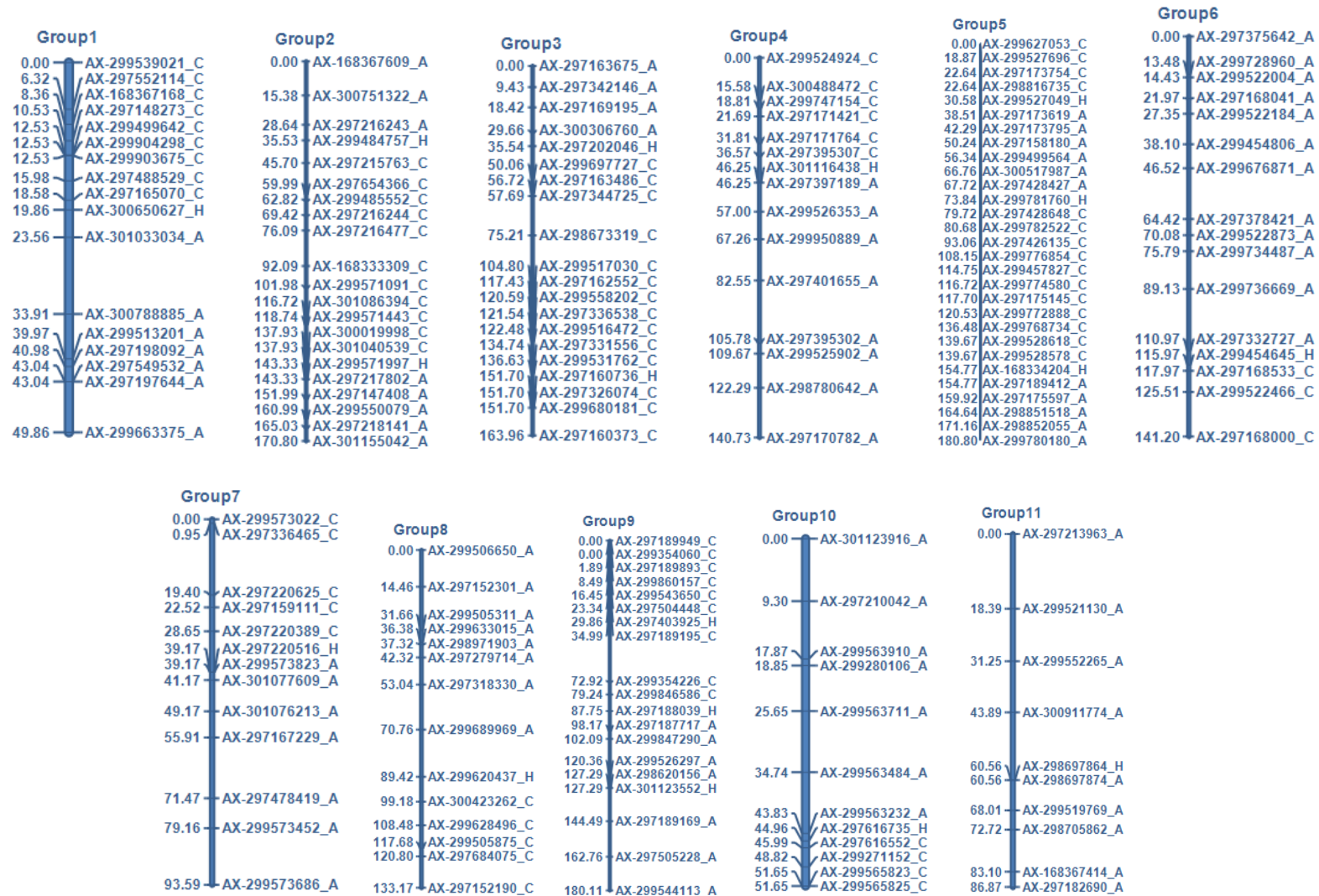


Figura 2 Mapa de ligações genéticas do cruzamento Pedro Sato × Roxa (PSR) realizado pelo GACD (Zhang *et al.* 2015).

Os GLs apresentaram marcadores comuns para todos os grupos mapeados pelos dois programas, exceto o GL6 do JoinMap que não teve correspondente no GACD (Tabela 1, Figuras 1 e 2). Dentre estes 85,7% dos marcadores mapeados no GL1 (JoinMap 4.0) foram mapeados no GL9 (GACD), 92,3% dos marcadores do GL2 (JoinMap 4.0) foram mapeados no GL8 (GACD) (Figuras 1 e 2). Os grupos 3, 4 e 11 (JoinMap 4.0) tiveram 100% dos marcadores mapeados nos grupos 5, 6 e 10 (GACD), respectivamente. O GL5 (JoinMap 4.0) apresentou 71,4% dos seus marcadores mapeados no GL11 do mapa 2. O GL7 (JoinMap 4.0) apresentou 80% dos marcadores em comum com o GL3 (GACD) (Figuras 1 e 2). No GL8 (JoinMap 4.0) 95,5% dos marcadores foram iguais ao GL2 (GACD), assim como 94,1% do GL9 correspondeu ao GL1 (GACD). 92,3% dos marcadores do GL10 (JoinMap 4.0) foram iguais ao GL7 (GACD) (Figuras 1 e 2).

Tabela 1 Comprimento dos grupos de ligação, número de marcadores/grupo de ligação, número de marcadores codominantes (H) nos mapas de ligações do cruzamento Pedro Sato × Roxa (PSR) feitos pelos programas JoinMap 4.0 × GACD.

GL	JoinMap 4.0			GACD			GL
	Comp. GL (cM)	Nº de marcadores	Marcadores (H)	Comp. GL (cM)	Nº de marcadores	Marcadores (H)	
GL 1	138,3	21	2	49,86	17	1	9
GL 2	93,9	13	1	170,8	21	2	8
GL 3	153,7	25	1	163,9	20	2	5
GL 4	133,1	16	1	140,73	15	1	6
GL 5	147,3	14	1	180,8	29	3	11
GL 6	180,6	26	23	141,2	16	1	Não há
GL 7	196,8	25	3	93,9	13	1	3
GL 8	174,9	22	2	133,17	14	1	2
GL 9	49,3	17	1	180,11	19	2	1
GL 10	84,4	13	1	51,65	12	1	7
GL 11	52,9	11	1	86,6	10	1	10
Total	1405,2	203	37	1392,7	186	17	

Foram observadas ordenações diferentes nos grupos: GL1 x GL9, GL2 x GL8, GL3 x GL5, GL4 x GL6, GL7 x GL3, GL8 x GL2, GL9 x GL1, GL10 x GL7 e GL11 x GL10 (Figuras 1 e 2), à exemplo dos últimos cinco marcadores do GL10 (JoinMap 4.0) que foram mapeados no início do GL7 (GACD). Essas mudanças de orientações não são importantes, pois os marcadores podem ser ordenados de diferentes formas, no início ou no final do grupo de ligação.

Os mapas também apresentaram algumas ordenações semelhantes, à exemplo do grupo de ligação 4 (JoinMap 4.0) no qual seis dos marcadores estavam ordenados da mesma forma que no grupo 6 (GACD) (Figuras 1 e 2). Esta situação foi observada em todos os outros grupos, em proporções diferentes, como nos grupos GL3 x GL5 e GL8 x GL2 (Figuras 1 e 2) que apresentaram 3 e 2 marcadores, respectivamente, ordenados de maneira semelhante. O GL 9 x GL1 (Figuras 1 e 2) não apresentaram coincidências na ordenação dos marcadores. Os grupos de ligações 5 x 11 e 11 x 10 (Figuras 1 e 2) apresentaram a maior coincidência na ordenação dos SNPs entre os mapas, sendo o grupo 5 (JoinMap 4.0) quase idêntico ao GL11 (GACD), com exceção de 4 marcadores; e o grupo de ligação 11, com 54% de seus marcadores ordenados de maneira semelhante ao GL 10 (Figuras 1 e 2).

Nos GL4 x GL6 (Figuras 1 e 2) as posições dos marcadores diferiram quase a totalidade, com exceção dos marcadores AX-297332727, posicionado à 107,2 cM (GL4) e 110,97 cM (GL6) e AX-299454645 posicionado à 102,2 cM (GL4) e 115,97 cM (GL 6) (Figuras 1 e 2). No GL 5 x GL11 o marcador AX-297213963 foi mapeado na mesma posição em ambos os mapas (Figuras 1 e 2). A maior diferença entre as posições estimadas para estes grupos foi observada no AX-297182690 que foi posicionado a 102,7 cM (GL5) e 86,87 cM (GL11), enquanto os demais marcadores apresentaram posições relativas próximas, variando entre 3,51 cM (AX-299521130) de diferença entre os mapas, até 15 cM (AX-168367414 e AX-297182690).

Entre os grupos GL11 x GL10 (Figuras 1 e 2) observou-se posicionamento quase idêntico dos marcadores. O marcador AX-301123916 foi o primeiro em ambos os grupos de ligação, a maior distância observada foi entre o marcador AX-299565823 o qual foi mapeado a 52,9 cM (GL11) e 44,96 cM (GL10) (Figuras 1 e 2), os demais marcadores apresentaram distâncias relativas estimadas ainda mais próximas, variando de 1,15 cM (AX-299563711) a 7,94 cM (AX-297616735) entre os mapas.

Mapa físico de SNPs de *Eucalyptus* em *P. guajava*

A análise *in silico* dos SNPs de *Eucalyptus* gerou 694 hits dos 1120 marcadores filtrados inicialmente para as análises de ligações genéticas, resultando na ordenação de um mapa físico da goiabeira (Figuras 3 e 4). Os e-values dos alinhamentos no mapa físico variaram de 8×10^{-10} a $1,15 \times 10^{-26}$.

O cromossomo 3 apresentou o maior comprimento, 50,58 Mb, enquanto o cromossomo 9 apresentou o menor comprimento, 32,33 Mb (Tabela 2). Os SNPs alinhados no genoma da

goiabeira cobriram quase a totalidade (mais de 90%) dos comprimentos de cromossomos reportados, variando de 31,4 Mb (cromossomo 9) a 49,27 Mb (cromossomo 3) e cobrindo em média 39,5 Mb/cromossomo (Tabela 2). O número de marcadores por grupo de ligação variou de 40 (cromossomo 8) a 88 (cromossomos 5 e 7), apresentando em média 63 SNPs por cromossomo (Tabela 2). A distância média entre marcadores SNPs foi de 0,62 Mb, variando entre 0,01 Mb a 7,36 Mb.

Tabela 2 Comprimento total dos cromossomos de *Psidium guajava* reportado por Feng et al. (2020), cobertura do mapa físico e número de marcadores SNPs de *Eucalyptus* por cromossomo de *P. guajava*.

Cromossomo	Comprimento total cromossomo (Mb)	Cobertura dos marcadores (Mb)	Número de marcadores/cromossomo
1	40,37	38,21	45
2	38,47	38,13	62
3	50,58	49,27	65
4	48,29	48,03	69
5	44,77	44,46	88
6	42,82	42,80	63
7	35,36	35,59	88
8	33,38	32,96	40
9	32,33	31,4	50
10	37,88	37,78	71
11	37,02	36,25	53
Média	40,12	39,53	63
Total	441,27	434,88	694

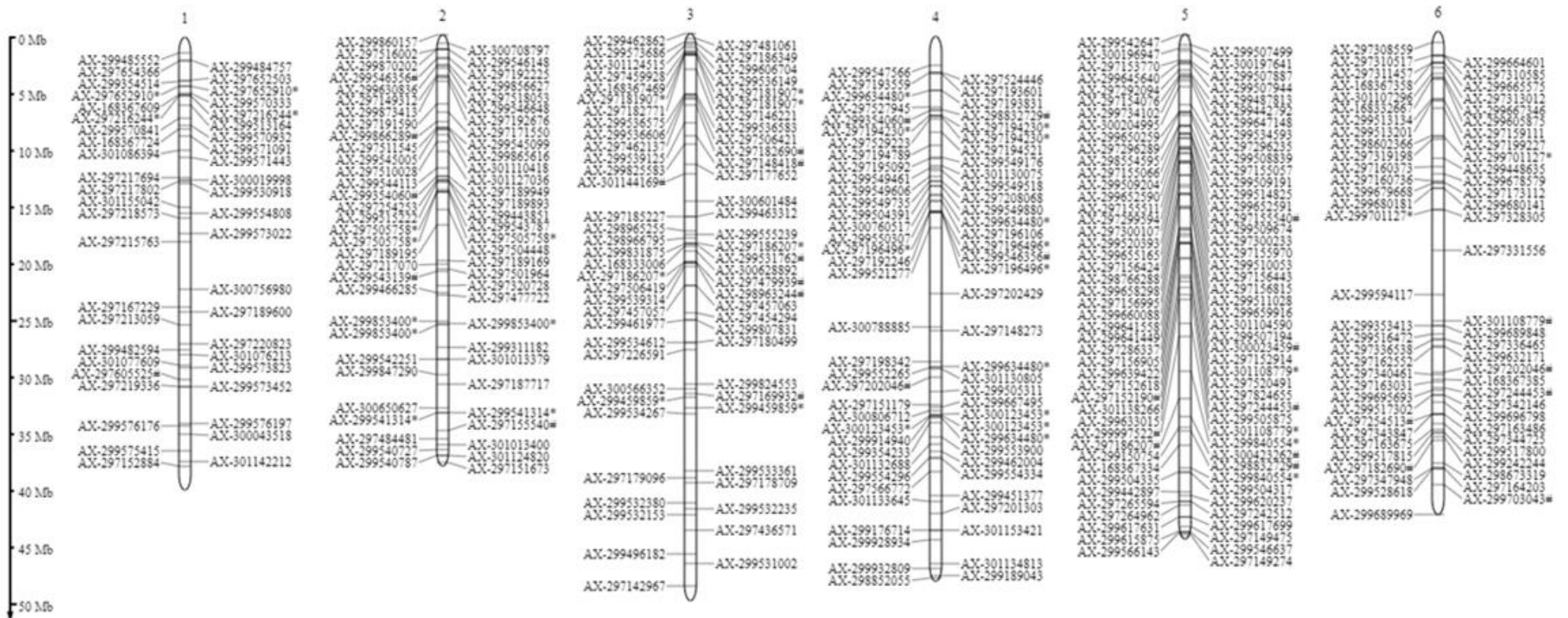


Figura 3 Mapa de físico dos cromossomos 1 ao 6 do cruzamento Pedro Sato × Roxa (PSR) realizado através do BLAST de SNPs de *Eucalyptus* contra o genoma da goiaba e alinhado no programa MapGene2Chrom (Jiangtao *et al.* 2015). Marcadores com “*” identificam marcadores repetidos no mesmo cromossomo. Marcadores com “#” identificam marcadores repetidos em mais de um cromossomo.

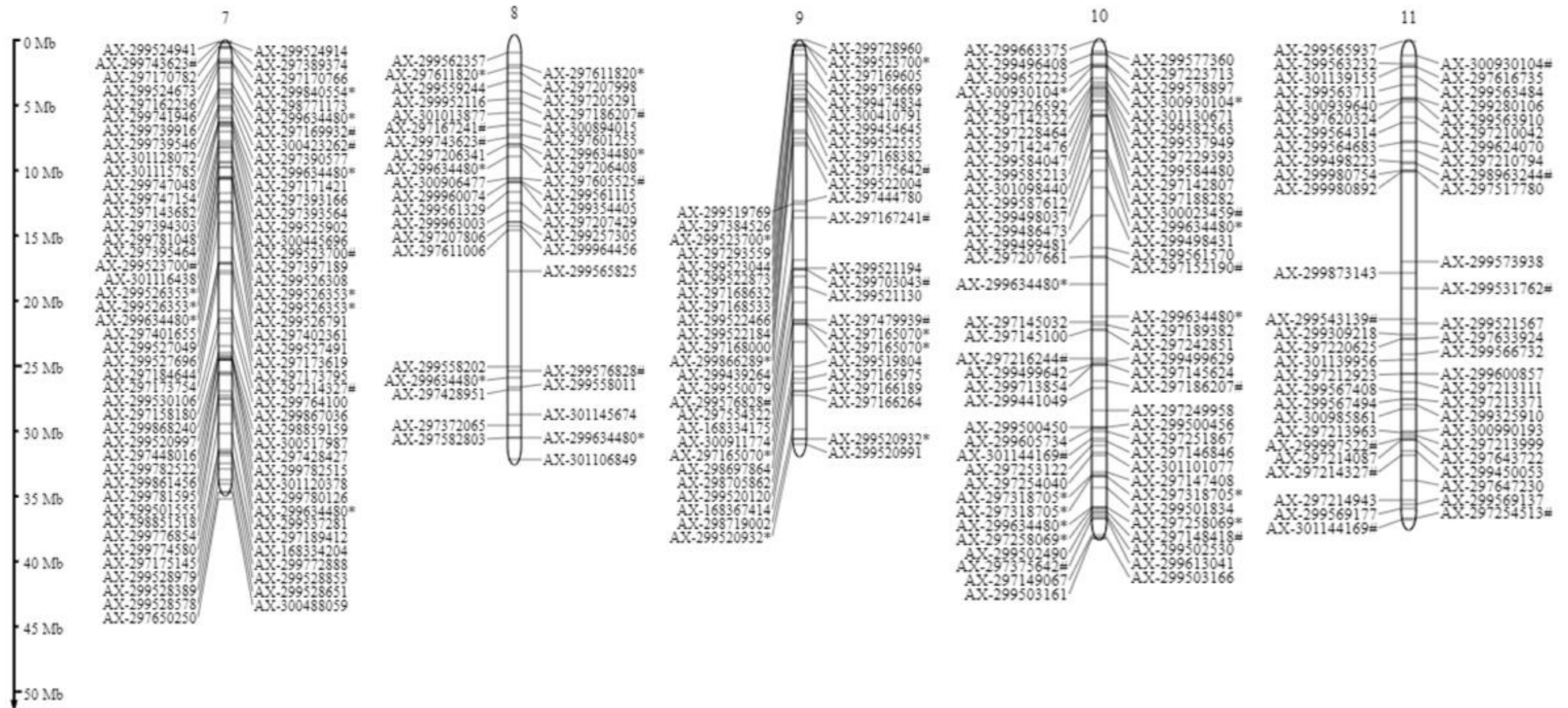


Figura 4 Mapa de físico dos cromossomos 7 ao 11 do cruzamento Pedro Sato x Roxa (PSR) realizado através do BLAST de SNPs de *Eucalyptus* contra o genoma da goiaba e alinhado no programa MapGene2Chrom (Jiangtao *et al.* 2015). Marcadores com “*” identificam marcadores repetidos no mesmo cromossomo. Marcadores com “#” identificam marcadores repetidos em mais de um cromossomo.

Comparações entre mapas de ligações genéticas e mapa físico

O mapa físico foi o mais denso quando comparado com os dois mapas de ligações genéticas (Tabela 3). O mapa Joinmap 4.0 apresentou média de 18 SNPs por grupo de ligação, sendo o GL6 o mais denso, seguido pelos GL3 e GL7, e o GL11 o menos denso (Tabela 3). O mapa GACD apresentou média 16 SNPs/grupo de ligação, e variou entre 10 marcadores (GL11) a 29 marcadores (GL5), enquanto o mapa físico apresentou média 63 SNPs/cromossomo, variando entre 40 marcadores (cromossomo 8) e 88 marcadores (cromossomos 5 e 7) (Tabela 3).

Tabela 3 Número de SNPs por cromossomo/grupo de ligação nos diferentes mapas genéticos e físico de *P. guajava*.

Cromossomo/grupo de ligação	Quantidade de SNPs por cromossomo/grupo de ligação		
	JoinMap 4.0	GACD	Físico
1	21	17	45
2	13	21	62
3	25	20	65
4	16	15	69
5	14	29	88
6	26	16	63
7	25	13	88
8	22	14	40
9	17	19	50
10	13	12	71
11	11	10	53
Média de SNPS	18	16	63
Comprimento total do mapa	1.405,2 cM	1392,7 cM	434,88 Mb

O mapa JoinMap 4.0 apresentou 122 marcadores comuns ao mapa físico. Alguns destes grupos de ligações tiveram sequências de SNPs em diferentes cromossomos da goiabeira, como no GL6 que teve sequências de SNPs em oito cromossomos: 1, 2, 3, 5, 6, 7, 9 e 11 (Tabela 4). O mapa GACD apresentou 117 marcadores comuns ao mapa físico, onde situação semelhante foi observada no GL1 com as sequências de SNPs distribuídas nos cromossomos 2, 4, 6, 9 e 10 cromossomos (Tabela 5).

Em três grupos de ligação do mapa JoinMap 4.0 (GL1, GL4 e GL11) houve alinhamento com somente um cromossomo do mapa físico (Tabela 4). Ao comparar o mapa JoinMap 4.0 e o mapa físico observamos que houve uma correspondência 57% entre o GL1 e o cromossomo 2, seguindo a ordem linear estimada pela recombinação dos marcadores, com exceção dos

marcadores AX-299544113 (12,5 Mb) e AX-299860157 (0,74 Mb). Nos GL 4 x cromossomo9 houve correspondência de 62,5% dos marcadores e não se observou ordenação semelhante entre os mapas, já no GL11 x cromossomo11 houve 72,7% de correspondência dos marcadores, porém com a ordenação alterada (Figuras 1 e 3).

Tabela 4 Comparações do mapa de ligação JoinMap 4.0 × mapa físico de *P. guajava*.

Grupo de ligação	Comprimento do grupo de ligação (cM)	Cromossomos correspondentes no mapa físico
1	138,3	Chr. 2
2	93,9	Chr. 4, Chr. 5
3	153,7	Chr. 4, Chr. 7
4	133,1	Chr. 9
5	147,3	Chr. 3, Chr. 4, Chr. 9, Chr. 11
6	180,6	Chr. 1, Chr. 2, Chr. 3, Chr. 5, Chr. 6, Chr. 7, Chr. 9, Chr. 11
7	196,8	Chr. 2, Chr. 3, Chr. 4, Chr. 5, Chr. 6, Chr. 8
8	174,9	Chr. 1, Chr. 9, Chr. 10
9	49,3	Chr. 2, Chr. 4, Chr. 6, Chr. 9, Chr. 10
10	84,4	Chr.1 Chr. 3, Chr. 6, Chr. 11
11	52,9	Chr. 11

Ao comparar o mapa GACD ao mapa físico observou-se que todos os grupos de ligação apresentaram alinhamentos com mais de um cromossomo, excetuando os GL4, GL6 e G9 (Tabela 5). Não foi mantida a mesma ordenação entre os GL4 x cromossomo7 (53,3% de marcadores correspondentes), bem como entre o GL 6 x cromossomo9 (62,5%), com exceção dos marcadores AX-299454645, AX-297168533, AX-299522466 e AX-29716800, ordenados em sequência em ambos os mapas (Figuras 2 e 4). A distância do marcador AX-299454645 foi de 115,97 cM (GL6), enquanto no mapa 3 a distância foi de 25,37 Mb. No GL 9 x cromossomo2 observou-se uma taxa de correspondência de 52,6% e alterações na ordem dos marcadores.

Tabela 5 Comparações do mapa de ligação GACD × mapa físico de *P. guajava*.

Grupo de ligação	Comprimento do grupo de ligação (cM)	Cromossomos correspondentes no mapa físico
1	49,86	Chr. 2, Chr.4, Chr.6, Chr.9, Chr. 10
2	170,80	Chr. 1, Chr. 9, Chr. 10
3	163,9	Chr. 3, Chr 4, Chr. 6, Chr. 8
4	140,73	Chr. 7
5	180,80	Chr. 4, Chr. 6, Chr. 7
6	141,20	Chr. 9
7	93,59	Chr. 1, Chr. 3, Chr. 6, Chr. 11
8	133,17	Chr. 4, Chr. 5, Chr. 6
9	180,11	Chr. 2
10	51,65	Chr. 8, Chr. 11
11	86,87	Chr. 3, Chr. 4, Chr. 9, Chr. 11

DISCUSSÃO

Mapas de ligações genéticas são representações simplificadas dos cromossomos de uma espécie e contribuem para o estudo de características de herança complexa. Este é o primeiro relato da construção de mapas de ligações genéticas e de mapa físico para *P. guajava* baseados em SNPs, associando distâncias de recombinação, cM, com a posição real no genoma da goiabeira, Mb, e comparando os resultados obtidos por dois diferentes programas de construção de mapas genéticos. Os SNPs são marcadores abundantes e amplamente distribuídos no genoma (Grossi *et al.* 2021) e sua utilização tem crescido em projetos de mapeamento de ligações (Ball *et al.* 2010, Lima *et al.* 2011, Soto *et al.* 2015, Lee *et al.* 2016, Gao *et al.* 2018).

Os grupos de ligações foram estabelecidos pelo LOD score (logarithm of odds), um indicador da razão de máxima verossimilhança e da probabilidade de ligação ou segregação independente destes marcadores. Nos mapas JoinMap 4.0 e GACD o LOD score adotado foi de 5 a 12 e 8, respectivamente, ambos iguais ou mais altos do que os reportados para os mapas da espécie (Lepitre *et al.* 2010, Padmakar *et al.* 2015).

Ao comparar os mapas de ligações JoinMap 4.0 e GACD aqui reportados observou-se que a quantidade de SNPs ligados nos mapas foi próxima, tanto no total dos SNPs (203 e 186, respectivamente), quanto em marcadores/grupo de ligação (18 SNPs/GL e 16 SNPs/GL, respectivamente). Entretanto a ordenação e distância entre os marcadores foram divergentes, salvo poucas exceções como o demonstrado no GL 5 x GL11 (Figuras 1 e 2). A ordenação de

marcadores em um mapa de ligações é semelhante ao problema do caixeiro viajante, onde deve-se escolher a melhor rota em um conjunto de $m!/2$ possibilidades (Cheema e Dicks 2009, Van Ooijen e Jansen 2013). O programa JoinMap 4.0 utilizando o algoritmo “maximum likelihood” demonstrou ser mais eficiente na ordenação dos marcadores SNPs em *P. guajava* que o mapa GACD, pois estes marcadores foram ordenados em sequência, de acordo com a sua origem parental (identificados pelas letras A e C), o que não tem fundamentação biológica, considerando-se que as recombinações resultantes da meiose são processos aleatórios, que dificilmente apresentam sequência ordenada de origem de um dos pais. A detecção de eventos de recombinação e integração de mapas é difícil pois um número diferente de alelos pode segregar em cada locus F1 e as fases de ligação não são conhecidas previamente (Maliepaard *et al.* 1997, Wu *et al.* 2002). Além disso, o mapa GACD apresentou menor quantidade de marcadores codominantes (17) enquanto o JoinMap 4.0 apresentou 37 marcadores codominantes, os quais são mais informativos para o mapeamento (Wu *et al.* 2002), sendo que a integração dos mapas parentais depende da utilização de marcadores codominantes, juntando fases de atração e repulsão.

O mapas genéticos gerados pelo JoinMap 4.0 (Van Ooijen 2006) e pelo GACD (Zhang *et al.* 2015) apresentaram 11 grupos de ligações cada, conforme esperado para a espécie ($n=11$). Os mapas previamente reportados para goiaba por Lepitre *et al.* (2010) e Padmakar *et al.* (2015), com base em SSR e SRAP, foram estimados em 2.179 cM e 2.551,3 cM, respectivamente, sendo maiores que os aqui reportados (1.405,2 cM e 1.392,7 cM). O mapa de ligação com base em SNPs de *L. scoparium*, espécie da família Myrtaceae, resultou em dois mapas parentais de 1.242,8 e 1.616,2 cM (Chagné *et al.* (2019), valores semelhantes aos mapas do presente estudo. Grossi *et al.* (2021) também reportaram um número pequeno de marcadores SNPs de goiabeira que foram ancorados em *E. grandis*, que pode ter influenciado nos comprimentos dos grupos de ligações do presente estudo.

A construção de mapas físicos através do alinhamento de sequências de SNPs pode ser uma alternativa para estimar as posições reais de marcadores em espécies em que o genoma sequenciado esteja disponível, orientando a construção de bibliotecas BAC e a montagem destes genomas. A ferramenta BLAST foi eficiente para esse objetivo no presente estudo, identificando sequências homólogas com significância biológica, variando entre 8×10^{-10} a $1,15 \times 10^{-26}$. Gao *et al.* (2018) trabalharam com um valor de corte de E^{-3} ao localizar fisicamente SNPs de aproximadamente 20 nucleotídeos em “scaffold” de *P. tuolensis*. No presente estudo,

utilizou-se sequências de aproximadamente 70 nucleotídeos de extensão e um valor de corte de E^{-10} , critérios que demonstraram ser eficientes na identificação de alinhamentos de significância biológica para a construção do mapa físico. Com o aumento das sequências (query) espera-se maior precisão nos alinhamentos e na identificação de regiões parálogas entre os genomas uma vez que o comprimento da sequências de SNPs influencia diretamente nos resultados dos alinhamentos, como bit score e e-value (Pertsemlidis e Fondon 2001, Blair *et al.* 2018).

Ao comparar os mapas de ligações do JoinMap 4.0 ao mapa físico observou-se que 60,1% (122 SNPs) foram comuns aos dois mapas, bem como a grande correspondência do GL11 com o cromossomo 11 e do GL1 com cromossomo 2. No mapa GACD observou-se 117 (62,9%) marcadores comuns ao mapa físico e ausência de correspondências exata entre os GLs e os cromossomos. Grattapaglia *et al.* (2015) relataram o uso do BLAST para identificar as consistências entre o mapa de ligação e o mapa físico de *E. grandis* (disponível no Phytozome) e puderam observar que 94,7% dos EST-SSR utilizados foram associados ao mapa físico. Estes valores são maiores do que os aqui reportados para os mapas genéticos, em função da especificidade dos marcadores para a espécie estudada. Entretanto o JoinMap 4.0 apresentou maior número de marcadores comuns e maior conservação da ordenação que o GACD, estando mais próximo do cenário real retratado pelo mapa físico e, portanto, foi mais eficiente na construção dos grupos de ligações e ordenação de marcadores em *P. guajava*.

A disponibilização de mapa genéticos, físicos e informações moleculares para *P. guajava* auxilia no melhoramento da cultura e na identificação de QTL ligados a marcadores. Alguns QTLs relativos ao fruto (morfologia, peso e qualidade) e à altura de planta já foram relatados e podem ser utilizados na seleção assistida quando ligados à marcadores moleculares (Rodríguez 2007, Ritter, Rodríguez-Medina *et al.* 2010). Para tal, necessita-se de mapas genéticos densos (Nimisha *et al.* 2013) e cada vez mais precisos. Apesar dos mapas genéticos serem extensivamente utilizados na literatura científica e cumprirem papel importante ao oferecer uma base para estudos de análise de QTL e seleção assistida por marcadores, trata-se de uma ferramenta propensa a erros em várias de suas etapas, e onerosa. Para minimizar os erros ou inconsistências advindas dos algoritmos utilizados (especialmente na identificação de marcadores ligados ou ordenação) a análise dos dados com mapas físicos baseados no alinhamento de sequências é necessária.

Desta forma, os mapas físicos aqui propostos podem ser utilizados como referência no mapeamento genético e de QTL para *P. guajava* pela facilidade de aplicação, baixo custo e

potencialidade de gerar informação de forma precisa a nível de nucleotídeo no genoma. A associação de marcadores a sua posição real no genoma e a ancoragem de novos marcadores pode auxiliar na identificação de genes candidatos em QTLs para a seleção assistida de caracteres por marcadores.

CONCLUSÃO

Foram desenvolvidos mapas de ligações genéticas pioneiros, com base em SNPs, para *P. guajava* através dos programas JoinMap 4.0, cobrindo 1.405,2 cM e com intervalo médio entre marcadores de 7,7 cM, e o programa GACD, cobrindo 1.392,7 cM e intervalo de 8,8 cM. Os programas JoinMap e GACD não apresentaram concordância nas distâncias estimadas e também na ordenação dos SNPs nos grupos de ligações. Ambos os mapas apresentaram grupos de ligações com segmentos de vários cromossomos, quando comparados ao mapa físico, indicando limitações. O GACD apresentou maior limitação em relação ao JoinMap 4.0, ao separar os marcadores de acordo com sua origem parental. O mapa físico gerado com BLASTn consistiu de 694 SNPs, cobrindo 434,88 Mb, com distâncias médias de 0,62 Mb. Devido à proximidade filogenética, os SNPs desenvolvidos para *Eucalyptus* foram úteis na genotipagem de *P. guajava* e montagem do mapa físico através do BLAST, sendo uma alternativa, até que um número massivo de marcadores específicos seja disponibilizado. Os mapas aqui reportados, especialmente o mapa físico, se mostram úteis ferramentas para o melhoramento genético da cultura, servindo de referência para futuros estudos de mapeamento e estimativas de QTLs em goiabeira.

REFERÊNCIAS BIBLIOGRÁFICAS

- Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic Local Alignment Search Tool. **Journal of Molecular Biology** 215: 403–410.
- Ball AD, Stapley J, Dawson DA, Birkhead TR, Burke T and Slate J (2010) A comparison of SNPs and microsatellites as linkage mapping markers: Lessons from the zebra finch (*Taeniopygia guttata*). **BMC Genomics** 11.
- Bertioli DJ, Ozias-Akins P, Chu Y, Dantas KM, Santos SP, Gouvea E, Guimarães PM, Leal-Bertioli SCM, Knapp SJ and Moretzsohn MC (2014) The use of SNP markers for linkage mapping in diploid and tetraploid peanuts. **G3: Genes, Genomes, Genetics** 4: 89–96.

MEDEIROS, F. L. B. Desenvolvimento de mapas genéticos e físicos com base em SNPs em *Psidium guajava* e triagem *in silico* de loci microssatélites de goiabeira para transferibilidade em *Eucalyptus globulus*

Blair MW, Cortés AJ, Farmer AD, Huang W, Ambachew D, Varma Penmetsa R, Carrasquilla-Garcia N, Assefa T and Cannon SB (2018) Uneven recombination rate and linkage disequilibrium across a reference SNP map for common bean (*Phaseolus vulgaris* L.). **PLoS ONE** **13**: 1–21.

Brondani RPV, Williams ER, Brondani C and Grattapaglia D (2006) A microsatellite-based consensus linkage map for species of *Eucalyptus* and a novel set of 230 microsatellite markers for the genus. **BMC Plant Biology** **6**: 20–35.

Bueno MRP (2009) O Projeto Genoma Humano. **Revista Bioética** **5**: 1–10.

Chagné D, Ryan J, Saeed M, Van Stijn T, Brauning R, Clarke S, Jacobs J, Wilcox P, Boursault E, Jaksons P, Jones D, Thrimawithana A, Schwinn K and Lewis D (2019) A high density linkage map and quantitative trait loci for tree growth for New Zealand mānuka (*Leptospermum scoparium*). **New Zealand Journal of Crop and Horticultural Science** **47**: 261–272.

Cheema J and Dicks J (2009) Computational approaches and software tools for genetic linkage map estimation in plants. **Briefings in Bioinformatics** **10**: 595–608.

Cheng CH, Chung MC, Liu SM, Chen SK, Kao FY, Lin SJ, Hsiao SH, Tseng IC, Hsing YIC, Wu HP, Chen CS, Shaw JF, Wu J, Matsumoto T, Sasaki T, Chen HH and Chow TY (2005) A fine physical map of the rice chromosome 5. **Molecular Genetics and Genomics** **274**: 337–345.

Costa SR and Santos CAF (2017) Genetic divergence among *Psidium* accessions based on single nucleotide polymorphisms developed for *Eucalyptus*. **Genetics and Molecular Research** **16**: 1–9.

DeWan AT, Parrado AR, Matise TC and Leal SM (2002) The map problem: a comparison of genetic and sequence-based physical maps. **Am. J. Hum. Genet.** **70**: 101–107.

Feng Chen, Feng Chao, Lin X, Liu S, Li Y and Kang M (2020) A chromosome-level genome assembly provides insights into ascorbic acid accumulation and fruit softening in guava (*Psidium guajava*). **Plant Biotechnology Journal** 717–730.

Gao W, Qu J, Zhang J, Sonnenberg A, Chen Q, Zhang Y and Huang C (2018) A genetic linkage map of *Pleurotus tuoliensis* integrated with physical mapping of the de novo sequenced genome and the mating type loci. **BMC genomics** **19**: 18.

MEDEIROS, F. L. B. Desenvolvimento de mapas genéticos e físicos com base em SNPs em *Psidium guajava* e triagem *in silico* de loci microssatélites de goiabeira para transferibilidade em *Eucalyptus globulus*

González VM, Garcia-Mas J, Arús P and Puigdomènech P (2010) Generation of a BAC-based physical map of the melon genome. **BMC Genomics 11**.

Grattapaglia D (2004) Integrating genomics into Eucalyptus breeding. **Genetics and molecular research : GMR 3**: 369–379.

Grattapaglia D, Mamani EMC, Silva-Junior OB and Faria DA (2015) A novel genome-wide microsatellite resource for species of Eucalyptus with linkage-to-physical correspondence on the reference genome sequence. **Molecular Ecology Resources 15**: 437–448.

Grossi LL, Fernandes M, Silva MA, de Oliveira Bernardes C, Tuler AC, dos Santos PHD, Ferreira A and da Silva Ferreira MF (2021) DArTseq-derived SNPs for the genus *Psidium* reveal the high diversity of native species. **Tree Genetics and Genomes 17**.

Han Y, Chagné D, Gasic K, Rikkerink EHA, Beever JE, Gardiner SE and Korban SS (2009) BAC-end sequence-based SNPs and Bin mapping for rapid integration of physical and genetic maps in apple. **Genomics 93**: 282–288.

Hohmann U, Endo TR, Gill KS and Gill BS (1994) Comparison of genetic and physical maps of group 7 chromosomes from *Triticum aestivum* L. **MGG Molecular & General Genetics 245**: 644–653.

Hudson CJ, Freeman JS, Kullán ARK, Petroli CD, Sansaloni CP, Kilian A, Detering F, Grattapaglia D, Potts BM, Myburg AA and Vaillancourt RE (2012) A reference linkage map for Eucalyptus. **BMC Genomics 13**.

Inglis PW, Marília de Castro RP, Resende L V. and Grattapaglia D (2018) Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. **PLoS ONE 13**: 1–14.

ISOGG - International Society of Genetic Genealogy (2017). Megabase. Disponível em: <<https://isogg.org/wiki/Megabase>>. Acesso em: 03 abr. 2021.

Jiangtao C, Yingzhen K, Qian W, Yuhe S, Daping G, Jing L and Guanshan L (2015) MapGene2Chrom, a tool to draw gene physical map based on Perl and SVG languages. **Hereditas 37**: 91–97.

Kong X, Murphy K, Raj T, He C, White PS and Matisse TC (2004) A combined linkage-physical

MEDEIROS, F. L. B. Desenvolvimento de mapas genéticos e físicos com base em SNPs em *Psidium guajava* e triagem *in silico* de loci microssatélites de goiabeira para transferibilidade em *Eucalyptus globulus*

map of the human genome. **American Journal of Human Genetics** **75**: 1143–1148.

Kosambi DD (1943) The estimation of map distances from recombination values. **Annals of Eugenics** **12**: 172–175.

Lander ES, Linton LM, Birren B, *et al* (2001) Initial sequencing and analysis of the human genome. **Nature** **409**: 860–921.

Lee YR, Yoon JB and Lee J (2016) A SNP-based genetic linkage map of *Capsicum baccatum* and its comparison to the *Capsicum annuum* reference physical map. **Molecular Breeding** **36**: 1–11.

Lepitre V, Nansot G, Grangeon R, Pomies V, Rivallan R, Risterucci AM, Valdés-Infante J, Rodríguez-Medina NN, Muth J, Boike J, Prüfer D, Becker D, Rohde W, Ritter E and Billotte N (2010) The microsatellite (SSR)/AFLP reference linkage map of guava. **Acta Horticulturae** **849**: 183–192.

Lima BM, Silva-Junior OB, Faria DA, Mamani EM, Pappas GJ and Grattapaglia D (2011) Assessment of SNPs for linkage mapping in *Eucalyptus*: construction of a consensus SNP/microsatellite map from two unrelated pedigrees. **BMC Proceedings** **5**: 4–5.

Maliepaard C, Jansen J and Van Ooijen JW (1997) Linkage analysis in a full-sib family of an outbreeding plant species: Overview and consequences for applications. **Genetical Research** **70**: 237–250.

Masud Parvez GM, Uzzaman S, Miah Akanda K and Mehjabin S (2018) A short review on a Nutritional Fruit : Guava. **Open Access: Toxicology & Research** **1**: 1–8.

Mozo T, Dewar K, Dunn P, Ecker JR, Fischer S, Kloska S, Lehrach H, Marra M, Martienssen R, Meier-Ewert S and Altmann T (1999) A complete BAC-based physical map of the *Arabidopsis thaliana* genome. **Nature Genetics** **22**: 271–275.

Nielsen R, Paul JS, Albrechtsen A and Song YS (2011) Genotype and SNP calling from next-generation sequencing data. **Nature Reviews Genetics** **12**: 443–451.

Nimisha S, Kherwar D, Ajay KM, Singh B and Usha K (2013) Molecular breeding to improve guava (*Psidium guajava* L.): Current status and future prospective. **Scientia Horticulturae** **164**: 578–588.

Padmakar B, Kanupriya C, Latha PM, Prashant KS, Dinesh MR, Sailaja D and Aswath C (2015)

MEDEIROS, F. L. B. Desenvolvimento de mapas genéticos e físicos com base em SNPs em *Psidium guajava* e triagem *in silico* de loci microssatélites de goiabeira para transferibilidade em *Eucalyptus globulus*

Development of SRAP and SSR marker-based genetic linkage maps of guava (*Psidium guajava* L.). **Scientia Horticulturae** **192**: 158–165.

Paiva JAP, Prat E, Vautrin S, Santos MD, San-Clemente H, Brommonschenkel S, Fonseca PGS, Grattapaglia D, Song X, Ammiraju JSS, Kudrna D, Wing RA, Freitas AT, Bergès H and Grima-Pettenati J (2011) Advancing Eucalyptus genomics: Identification and sequencing of lignin biosynthesis genes from deep-coverage BAC libraries. **BMC Genomics** **12**.

Pertsemlidis A and Fondon JW (2001) Having a BLAST with bioinformatics (and avoiding BLAST phemy). **Genome Biology** **2**: 1–10.

Pommer C, Odaci O and Santos CAF (2013) **Goiaba: Recursos genéticos e melhoramento**. EdUfersa, Mossoró, 126p.

Risterucci AM, Duval MF, Rohde W and Billotte N (2005) Isolation and characterization of microsatellite loci from *Psidium guajava* L. **Molecular Ecology Notes** **5**: 745–748

Ritter E, Herran A, Valdés-Infante J, Rodríguez-Medina NN, Briceño A, Fermin G, Sanchez-Teyer F, O’connor-Sanchez A, Muth J, Boike J, Prüfer D, Santos CA, Dos Santos ICN, Rodrigues MA, Risterucci AM, Billotte N, Becker D and Rohde W (2010) Comparative linkage mapping in three guava mapping populations and construction of an integrated reference map in guava. **Acta Horticulturae** **849**: 175–182.

Ritter E, Rodríguez-Medina NN, Velásquez B, Rivero D, Rodríguez JA, Martínez F and Valdés-Infante J (2010) QTL (quantitative trait loci) analysis in guava. **Acta Horticulturae** **849**: 193–202.

Rodríguez N, Valdés-Infante J, Becker D, Velásquez B, González G, Sourd D, Rodríguez J, Billotte N, Risterucci AM, Ritter E and Rohde W (2007) Characterization of guava accessions by ssr markers, extension of the molecular linkage map, and mapping of qtls for vegetative and reproductive characters. **Acta Horticulturae** **735**: 201–215.

Silva-Junior OB, Faria DA and 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.

Soto JC, Ortiz JF, Perlaza-Jiménez L, Vásquez AX, Lopez-Lavalle LAB, Mathew B, León J, Bernal AJ, Ballvora A and López CE (2015) A genetic map of cassava (*Manihot esculenta* Crantz) with integrated physical mapping of immunity-related genes. **BMC Genomics** **16**: 1–

MEDEIROS, F. L. B. Desenvolvimento de mapas genéticos e físicos com base em SNPs em *Psidium guajava* e triagem *in silico* de loci microssatélites de goiabeira para transferibilidade em *Eucalyptus globulus*

16.

Sumathi M, Bachpai VKW, Mayavel A, Dasgupta MG, Nagarajan B, Rajasugunasekar D, Sivakumar V and Yasodha R (2018) Genetic linkage map and QTL identification for adventitious rooting traits in red gum eucalypts. **3 Biotech 8**: 242.

Valdés-infante J, Becker D, Rodriguez N, Velasquez B, Gonzalez G, Sourd D, Rodriguez J, Ritter E, Rohde W (2003) Molecular characterization of Cuban accessions of guava (*Psidium guajava* L) establishment of a first molecular linkage map and mapping of QTLs for vegetative characters. **Journal of Genetics of Breeding 57**: 349 358

Van Ooijen JW (2006) JoinMap®4, Software for the Calculation of Genetic Linkage Maps in Experimental Populations. Kyazma BV, Wageningen.

Van Ooijen JW, Jansen J (2013) **Genetic mapping in experimental populations**. Cambridge University Press, Cambridge, 139p.

Wu J, Jenkins JN, McCarty JC and Lou XY (2011) Comparisons of four approximation algorithms for large-scale linkage map construction. **Theoretical and Applied Genetics 123**: 649–655.

Wu R, Ma CX, Painter I and Zeng ZB (2002) Simultaneous Maximum Likelihood Estimation of Linkage and Linkage Phases in Outcrossing Species. **Theoretical Population Biology 61**: 349–363.

Xu Z, Kohel RJ, Song G, Cho J, Yu J, Yu S, Tomkins J and Yu JZ (2008) An integrated genetic and physical map of homoeologous chromosomes 12 and 26 in Upland cotton (*G. hirsutum* L.). **BMC Genomics 9**: 1–11.

Zhang L, Meng L, Wu W and Wang J (2015) GACD: Integrated software for genetic analysis in clonal fl and double cross populations. **Journal of Heredity 106**: 741–744.

CAPÍTULO III

TRIAGEM *IN SILICO* DE LOCI SSR DE *Psidium guajava* DE GOIABEIRA PARA TRANSFERIBILIDADE EM *Eucalyptus globulus*

RESUMO

BLAST é uma ferramenta de busca por alinhamentos locais em genomas, utilizado na identificação de homologias entre genótipos e localização de possíveis genes ortólogos. O objetivo do presente trabalho foi utilizar a ferramenta BLAST para uma triagem *in silico* de sequências SSR de *Psidium guajava* em cromossomos de *Eucalyptus globulus*, de forma a orientar estudos de transferibilidade entre as espécies. Foram utilizados 23 clones de SSR de *P. guajava* (query), depositados no site NCBI, contra o genoma completo de *E. globulus* (subject), e selecionados os alinhamentos com valor de corte de $e\text{-value} < 1e^{-20}$. Também foram utilizados 140 loci de SSR disponibilizados pelo consórcio GuavaMap adotando os parâmetros: $e\text{-value} < 1,7$ e proximidade entre as sequências forward e reverse (máximo de 300 nucleotídeos). Para ambas análises utilizou o Blastn com otimização MEGABLAST. A extração de DNA de quatro indivíduos de eucalipto foi realizada de acordo com o protocolo CTAB 2x, modificado com uma pré-lavagem de sorbitol. A validação da triagem *in silico* foi efetuada com PCR, com resolução em gel de poliacrilamida. Das 23 sequências avaliadas 39% apresentaram alinhamentos significativos, com identidade média das sequências de 87%. Dos 140 locos de SSR GuavaMap somente três apresentaram alinhamentos significativos, representando 2,1% do total. A identidade média dos primers forward e reverse, para os loci SSR GuavaMap, foi de 100% e 95% respectivamente. Os sete SSR que apresentaram resultados significativos na triagem *in silico* (mPgCIR001, mPgCIR005, mPgCIR007, mPgCIR009, mPgCIR018, mPgCIR020 e mPgCIR026) apresentaram amplicons de fácil visualização em géis de poliacrilamida, enquanto outros cinco, sem alinhamentos significativos, não apresentaram amplicons. Esses resultados indicam a efetividade da triagem *in silico*, aqui relatada pela primeira vez, na previsão do anelamento de sete sequências SSR em *E. globulus*. Desta forma, a triagem *in silico* demonstra ser uma alternativa para orientar estudos de transferibilidade entre as duas espécies, reduzindo custos e tempo.

Palavras-chave: SSR, BLAST, triagem *in silico*, *P. guajava*.

ABSTRACT

BLAST is a genome local alignment search tool used on homology identification between genotypes and identifying possible orthologue genes. This study aimed to use BLAST for an *in silico* screening of SSR sequences from *Psidium guajava* to *Eucalyptus globulus* to guide transferability studies between these species. Twenty-three SSR clone sequences from *P. guajava* (query), deposited on NCBI's website, were aligned against the whole genome of *E. globulus* (subject), using a cut off $e\text{-value} < 1e^{-20}$. Another 140 loci made available by GuavaMap were analyzed with the following parameters: $e\text{-value} < 1.7$, maximum distance of 300 nucleotides between forward and reverse sequences. All loci were analyzed using Blastn with MEGABLAST optimization. DNA extraction of four eucalypt samples was performed with CTAB 2x protocol modified with a sorbitol pre-wash. Validation of *in silico* screening was performed through PCR technique with visualization in polyacrylamide gel. 39% of the 23 SSR clones showed significant alignments, with sequences mean identity of 87%. Only three GuavaMap loci (2.1%) presented significant alignments with mean identity of 100% and 95% in primers forward and reverse, respectively. Seven SSR clones which presented significant alignments *in silico* (mPgCIR001, mPgCIR005, mPgCIR007, mPgCIR009, mPgCIR018, mPgCIR020 and mPgCIR026) presented easy-to-score amplicons in the polyacrylamide gels, meanwhile the other five, without significant alignments, did not amplified. These results indicate the effectiveness of *in silico* screening, here reported for the first time, in predicting annealing of seven SSR sequences in *E. globulus*, and therefore being an alternative for cost and time reduction in transferability studies between these species.

Keywords: SSR, BLAST, *in silico* screening, *P. guajava*.

INTRODUÇÃO

Marcadores microssatélites ou “simple sequence repeat” (SSR) são amplamente utilizados em diferentes áreas da biologia e genética molecular pela sua abundância no genoma, codominância e elevada taxa de informação (Carrer *et al.* 2010). Rafalski *et al.* (1996) foi um dos primeiros a propor um método de identificação de microssatélites e posterior desenho de primers, que inclui: construção de uma biblioteca genômica, “screening” por hibridização, determinação das sequências de DNA de clones positivos, design dos primers, análise da PCR e identificação de polimorfismos. Trata-se de uma técnica laboriosa e onerosa do ponto de vista financeiro (Varshney *et al.* 2002, 2005). Com os avanços nas tecnologias de sequenciamento de próxima geração (NGS) já se pode identificar um alto número de microssatélites com menor custo e maior agilidade (Taheri *et al.* 2018).

Na ausência de uma biblioteca genômica para a espécie alvo, estudos de variabilidade genética, mapeamento, “fingerprinting” e genômica comparativa, têm sido viabilizados pela transferibilidade de marcadores SSR, de uma espécie para outra espécie (Singleton *et al.* 2020). Estes marcadores são regiões conservadas entre espécies próximas taxonomicamente e, portanto, úteis na identificação de sintenia e colinearidade (Shepherd *et al.* 2006). Fagundes *et al.* (2016) utilizaram a transferibilidade para auxiliar na caracterização e estudo de biodiversidade de diferentes espécies da Mata Atlântica, pertencentes à família Myrtaceae. Na família Myrtaceae a transferibilidade tem sido empregada com ênfase no gênero *Eucalyptus* (dos Santos *et al.* 2007, Grattapaglia *et al.* 2011, Hudson *et al.* 2012, Butler *et al.* 2017) devido a sua importância econômica e grande investimento em recursos moleculares. Nessa mesma família, trabalhos de transferibilidade foram realizados com sucesso por Rai *et al.* (2013) e Sittther *et al.* (2014).

Zou *et al.* (2020) relatam as dificuldades da transferibilidade de marcadores moleculares, como: a presença de alelos nulos devido ao polimorfismo local, falta de polimorfismo e a variação genômica mesmo entre indivíduos aparentados; além da necessidade de validação dos dados através da amplificação dos produtos da PCR. Com o advento de técnicas da bioinformática, como o alinhamento local de sequências, a genômica comparativa ganhou aliados na otimização de projetos e redução de custos.

A ferramenta BLAST vem sendo utilizado na busca por homologias entre espécies e também na localização de possíveis genes ortólogos. Varshney *et al.* (2005) utilizaram o BLAST para estimar a transferibilidade teórica de marcadores SSR-ESTs (expressed sequence tags) mapeados em cevada contra SSR-ESTs de trigo, centeio, arroz, milho, sorgo, *Arabidopsis*

e *Medicago*, e para tal consideraram alinhamentos significativos com e-value $< 1,00 \times 10^{-10}$. Os autores foram capazes de encontrar homologias significativas entre trigo, centeio e arroz de 93,5, 37,3 e 57,3%, respectivamente, e estimar a localização de regiões putativamente ortólogas, comparando SSR-EST mapeados fisicamente em cevada contra as outras espécies de cereais.

Trabalhos similares foram desenvolvidos na família Myrtaceae por Acuña *et al.* (2012) no qual os pesquisadores utilizaram a ferramenta BLASTX na tentativa de prever a função de proteínas das regiões SSR-EST descobertas para *Eucalyptus* e estimar identidades destas regiões a partir de um valor de corte de e-value $\leq 1 \times 10^{-5}$. Sumathi *et al.* (2018) utilizaram o BLAST para ancorar marcadores microssatélites recém-mapeados do híbrido de *E. tereticornis* \times *E. camaldulensis* no genoma de *E. grandis* e observaram altos níveis de sintonia, além de uma colinearidade de 64% entre os genomas comparados.

O objetivo do presente trabalho foi utilizar a ferramenta BLAST para uma triagem *in silico* de sequências de SSR de *Psidium guajava* em cromossomos de *Eucalyptus globulus*, de forma a orientar estudos de transferibilidade entre as espécies.

MATERIAIS E MÉTODOS

Triagem *in silico* de loci microssatélites de *P. guajava*

Foram utilizados sequências de 23 clones de SSR desenvolvidos por Risterucci *et al.* (2005) para *P. guajava* (query), depositadas no site NCBI (National Center for Biotechnology Information). O genoma completo de *Eucalyptus globulus* (subject), disponível por cromossomos no NCBI, foi utilizado na busca de sequências idênticas às sequências dos SSR de *P. guajava*. Pela plataforma do NCBI utilizou-se a ferramenta Blastn (Zhang *et al.* 2000) contra o genoma de *E. globulus* (Anexo I). Utilizou-se a otimização “MEGABLAST” para localizar sequências altamente similares e posteriormente selecionadas aquelas com e-value $< 1 \times 10^{-20}$.

Também foram utilizados 140 loci disponibilizados pelo consórcio GuavaMap dos quais as sequências curtas destes primers foram utilizadas. As duas regiões genômicas (sequence forward e sequence reverse) correspondentes foram utilizadas em formato FASTA, aplicando a ferramenta Blastn (Zhang *et al.* 2000), com a otimização “MEGABLAST”, para localizar sequências altamente similares. Os primers foram selecionados com base no score, e-value < 1.7 , proximidade entre as sequências forward e reverse (máximo de 300 nucleotídeos de

distância considerando o Subject Start e Subject end) e a presença das sequências identificadas no mesmo cromossomo.

Tabela 1 Locus SSR, número de acesso EMBL, tamanho do clone e tamanho dos alelos das 23 sequências SSR de *Psidium guajava* identificadas por (Risterucci *et al.* 2005).

Locus SSR	Número de acesso EMBL	Tamanho do clone	Faixa de tamanho dos alelos
mPgCIR01	AJ639775	237	236–250
mPgCIR02	AJ639753	224	202–230
mPgCIR03	AJ639754	158	118–164
mPgCIR04	AJ639755	148	126–150
mPgCIR05	AJ639756	252	224–280
mPgCIR07	AJ639757	149	148–160
mPgCIR08	AJ639758	214	210–224
mPgCIR09	AJ639759	173	156–176
mPgCIR10	AJ639760	261	262–320
mPgCIR11	AJ639761	298	298–314
mPgCIR13	AJ639762	245	240–260
mPgCIR14	AJ639763	185	184–186
mPgCIR15	AJ639764	147	144–172
mPgCIR16	AJ639765	292	268–296
mPgCIR17	AJ639766	231	230–240
mPgCIR18	AJ639767	195	192–204
mPgCIR19	AJ639768	274	258–280
mPgCIR20	AJ639769	266	270–298
mPgCIR21	AJ639770	154	150–164
mPgCIR22	AJ639771	235	236–252
mPgCIR23	AJ639772	185	184–198
mPgCIR25	AJ639773	124	104–130
mPgCIR26	AJ639774	185	180–198

Extração e Quantificação do DNA

Foram utilizados quatro genótipos de *Eucalyptus deglupta*, coletados na sede da Embrapa Semiárido em Petrolina – PE e extraídos de acordo com protocolo de Inglis *et al.* (2018), o qual consiste na adição de uma pré-lavagem da amostra com sorbitol. A maceração mecânica foi realizada na presença de nitrogênio líquido para o rompimento da parede vegetal, até a obtenção de um pó fino. O vegetal macerado de cada amostra foi transferido para tubos de eppendorf de 2 mL, contendo cada um 900 µL solução de sorbitol [100 mM Tris-HCl pH 8,0, 0,35 M Sorbitol, 5mM EDTA pH 8,0, 1% polivinilpirrolidona (PVP- 40)] para a pré-

lavagem e adição de 1% (v/v) de 2-mercaptoetanol antes da extração. As amostras foram centrifugadas a 7500 rpm durante 5 min, posteriormente foi retirado o sobrenadante e adicionado à amostra 700 µL do tampão de extração CTAB (3%) (20 mM EDTA, 100 mM Tris-HCl pH 8,0, 3 M NaCl, 3% CTAB). As amostras foram colocadas em banho Maria a 60°C durante 30 min, realizando a inversão dos tubos de eppendorf a cada 10 min. Em seguida foi adicionado 700 µL de clorofórmio e álcool isoamílico (24:1) às amostras e as mesmas centrifugadas a 10.500 rpm, por 10 min. Foram transferidos 600 µL do sobrenadante para um novo tubo de eppendorf de 1,5 ml e neles adicionados 396 µL de álcool isopropílico gelado e 60 µL de acetato de sódio, seguido da inversão suave dos tubos e a disposição em um isopor com gelo para repouso por 60 min. Após 1 hora as amostras foram centrifugadas a 10.500 rpm para formação do “pellet” no fundo do tubo, descartando-se o sobrenadante em béquer em capela de exaustão de gases. O “pellet” foi ressuscitado em 30 µL Tris-EDTA, permanecendo por 24 h a 6°C para completa dissolução do “pellet”.

A quantificação do DNA foi realizada por espectrofotometria, utilizando a placa Microdrop na faixa de absorbância a 260 nm. Uma curva de regressão foi ajustada aos valores de absorbância de amostras de DNA de concentrações conhecidas. A determinação do grau de pureza foi obtida pela razão entre as leituras A260/A280, sendo utilizadas as amostras que apresentaram valor próximo ou acima de 1,8, caso contrário as amostras foram reextraídas. Após quantificação as amostras foram diluídas para concentração de trabalho de 60 ng/µL e armazenadas a -20°C.

Validação da triagem *in silico*

Foram selecionados sete loci com “hits” no BLAST e outros cinco loci que não apresentaram “hits”, entre os loci publicados por Risterucci *et al.* (2005) para comparar aos resultados observados na triagem *in silico*, por meio da amplificação via PCR e resolução em gel de poliacrilamida. As reações de amplificação de PCR foram realizadas para um volume final de 20 µL de acordo com o descrito por Rai *et al.* (2013) com algumas modificações: 60ng de DNA, 2x de tampão para TaqDNA polimerase, 2,5 mM MgCl₂, 2,5 mM de dNTP's, 0,8 µM de cada “prime” e 1 unidade de enzima Taq DNA Polimerase. A programação do termociclador consistiu na desnaturação do ciclo inicial 94°C por 4 min; 30 ciclos a 94°C por 45 s, 52°C por 60 s e 72°C por 60 s e uma etapa de extensão final a 72°C, por 8 min.

Foram adicionados 5µL do tampão desnaturado de formamida 98% (EDTA pH 8,0 10mM, 1mg/ml de XileneCyanol e 1 mg/mL de Bromofhenol blue), seguido da completa

desnaturação a 94°C por 5 min em termociclador, anteriormente à aplicação do gel de poliacrilamida. As amostras foram mantidas em gelo até a aplicação. Os produtos de amplificação foram separados em gel de poliacrilamida 6%, preparado em placas de vidro tipo sanduíche com capacidade de 60 poços.

As placas foram limpas utilizando lenços de papel embebidos de etanol. Para limpeza da placa maior (de adesão do gel) utilizou-se 1,0 mL de solução contendo 95% de Etanol + 0,5 % de ácido Acético e 1 µL de Bindsilane. Essa solução foi espalhada em toda a superfície da placa com um auxílio de um lenço de papel, sendo o excesso retirado com um lenço umedecido com álcool etílico. Para a limpeza da placa menor, substituiu-se a solução contendo Bindsilane por um produto repelente a água, usado para limpar vidros de automóveis, com nome comercial de Glass Shield Autoshine. Uma pré-corrida de 30 min a 45 W foi realizada antes da aplicação das amostras de PCR. Foram aplicados 2,5 µL da reação de PCR desnaturada no gel de poliacrilamida 6%, sendo a corrida de eletroforese realizada por um período de aproximadamente 3h, com potência constante de 45 W.

As placas contendo o gel foram corados com nitrato de prata conforme proposto por Creste *et al.* (2001), com as seguintes modificações: A) Imersão em solução de fixação contendo etanol absoluto 10% e ácido acético 1% por 20 min; B) Lavagem com água destilada sob agitação por 1 min; C) Imersão em solução de ácido nítrico 0,2 mol.L por 3 min; D) Nova lavagem do gel com água destilada por 1 min; E) A impregnação do gel foi realizada com solução de nitrato de prata 0,2% por 20 min, seguida de duas lavagens de 30 s com água destilada; E) O gel foi imerso em solução de revelação contendo carbonato de sódio (60g) e formaldeído (1350 µL), utilizando-se metade da solução reveladora na primeira lavagem, até o início do aparecimento das bandas e posteriormente transferido para uma segunda bandeja contendo o restante da solução até o aparecimento do padrão desejado; F) Uma nova lavagem do gel foi realizada com ácido acético 5% por a 1 min, seguida de uma lavagem final com água por 1 min. Ao final do processo de revelação do gel o mesmo foi colocado em posição vertical para secagem e posterior análise dos fragmentos e digitalização.

RESULTADOS

Triagem *in silico* de loci SSR de *P. guajava* para os cromossomos de *Eucalyptus globulus*

Dentre os 23 primers disponibilizados por Risterucci *et al.* (2005), nove apresentaram alinhamentos significativos, ou seja 39% das sequências SSR (Tabela 2). Observaram-se altos scores variando entre 141 (mPgCIR021) a 435 (mPgCIR007). Todos os primers tiveram baixos valores do e-value variando entre $3,00 \times 10^{-32}$ (mPgCIR0021) a $2,00 \times 10^{-120}$ (mPgCIR007). O primer mPgCIR018 apresentou maior valor de identidade (92,91%). Todos os loci apresentaram identidades maior do que 80%, exceto o locus mPgCIR021 (79,9%), sendo a identidade média dos loci de 87%.

Tabela 2. Score, e-value e identidade de nove sequências de SSR (Risterucci *et al.* 2005) de *P. guajava*, em cromossomos de *Eucalyptus globulus*.

Locus de <i>P. guajava</i>	Cromossomo de <i>E. globulus</i>	Score	e-value	Identidade (%)
mPgCIR001	6	339	$2,00 \times 10^{-91}$	83,64
mPgCIR005	1	259	$2,00 \times 10^{-67}$	84,59
mpPCIR007	10	435	$2,00 \times 10^{-120}$	91,35
mPgCIR009	4	274	$6,00 \times 10^{-72}$	83,08
mPgCIR011	8	372	$3,00 \times 10^{-101}$	90,71
mPgCIR018	3	396	$9,00 \times 10^{-109}$	92,91
mPgCIR020	11	289	$2,00 \times 10^{-76}$	90,83
mPgCIR021	8	141	$3,00 \times 10^{-32}$	79,91
mPgCIR026	4	326	$1,00 \times 10^{-87}$	83,83

Dos 140 loci de SSR desenvolvidos no projeto GuavaMap somente três (mPgCIR261, mPgCIR313, mPgCIR378) apresentaram alinhamentos significativos, representando 2,1% do total de marcadores testados. A identidade média forward foi de 100% e nas sequências reverse a identidade média foi de 95,08% (Tabela 3). Os marcadores mPgCIR261, mPgCIR313, mPgCIR378 apresentarem scores de: 44,1, 40,1, 40,1 nas sequências forward e 32,2, 34,2 e 32,2 nas sequências reverse, respectivamente. Os alinhamentos apresentaram e-values variando de 0,19 a 0,0004 (Tabela 3). O BLAST foi capaz de posicionar as sequências do mPgCIR261, mPgCIR313, mPgCIR378 à 274, 62 e 86 nucleotídeos de distância, respectivamente (Tabela 3), que pode resultar em transferibilidade efetiva por PCR e resolução em gel de poliacrilamida, devido a pequena distância entre eles no genoma do *E. globulus*.

Tabela 3 Score, e-value, identidade (ID) e distância entre primers (bp) de três sequências de SSR do projeto GUAVAMAP de *P. guajava*, em cromossomos de *E. globulus* (Chr).

Marcador	Número de acesso	Chr	Score	e-value	ID (%)	Distância entre primers (pb)
mPgCIR261	Não publicado	8	44,1	2,00E ⁻⁰⁴	100	274
mPgCIR313	Não publicado	1	40,1	2,00E ⁻⁰³	100	62
mPgCIR378	Não publicado	2	40,1	2,00E ⁻⁰³	100	86
mPgCIR261	Não publicado	8	32,2	4,40E ⁻⁰¹	95	274
mPgCIR313	Não publicado	1	34,2	1,90E ⁻⁰¹	95	62
mPgCIR378	Não publicado	2	32,2	6,60E ⁻⁰¹	95	86

Validação da triagem *in silico* de loci SSR de *P. guajava* através de amplificação em gel de poliacrilamida

Todos os sete loci com e-value < 3,00xE⁻³² apresentaram amplificação de PCR, com boa resolução em gel de poliacrilamida, enquanto os cinco loci que não apresentaram “hits” na análise *in silico* não produziram amplicons (Figura 1).

O locus mPgCIR001 (e-value = 2,00xE⁻⁹¹) apresentou amplificação das quatro amostras de DNA com ótima qualidade e definição de bandas, o que pode estar associado a alta qualidade do alinhamento. Este locus foi monomórfico para as amostras utilizadas (Figura 1). Para o locus mPgCIR005 (e-value = 2,00xE⁻⁶⁷) três das quatro amostras amplificaram com boa resolução, sendo também monomórficas (Figura 1). O locus de menor e-value (2,00xE⁻¹²⁰), mPgCIR007, apresentou a melhor resolução no gel de poliacrilamida, amplificação de todas as amostras e monomorfismo (Figura 1). O locus mPgCIR009 (e-value = 6,00xE⁻⁷²) também amplificou em todas as amostras, apesar de apresentar resolução inferior quando comparados aos outros e com aproximadamente 250 pb (Figura 1).

O locus mPgCIR018 (e-value = 9,00xE⁻¹⁰⁹) apresentou amplificação de todas as amostras com boa qualidade e resolução, monomorfismo e aproximadamente >400 pb (Figura 1). Já no locus mPgCIR020 (e-value = 2,00xE⁻⁷⁶) todas as amostras amplificaram e este marcador foi polimórfico, sendo assim considerado o melhor devido ao maior nível de informação disponibilizada (Figura 1). O locus mPgCIR026 (e-value = 1,00xE⁻⁸⁷) apresentou amplificação de todas as amostras e também foi monomórfico, embora a resolução em gel tenha sido inferior quando comparada aos outros loci (Figura 1). Os demais loci utilizados (mPgCIR003, mPgCIR008, mPgCIR014, mPgCIR017, mPgCIR023) não apresentaram padrão de amplificação ou a amplificação foi inespecífica.

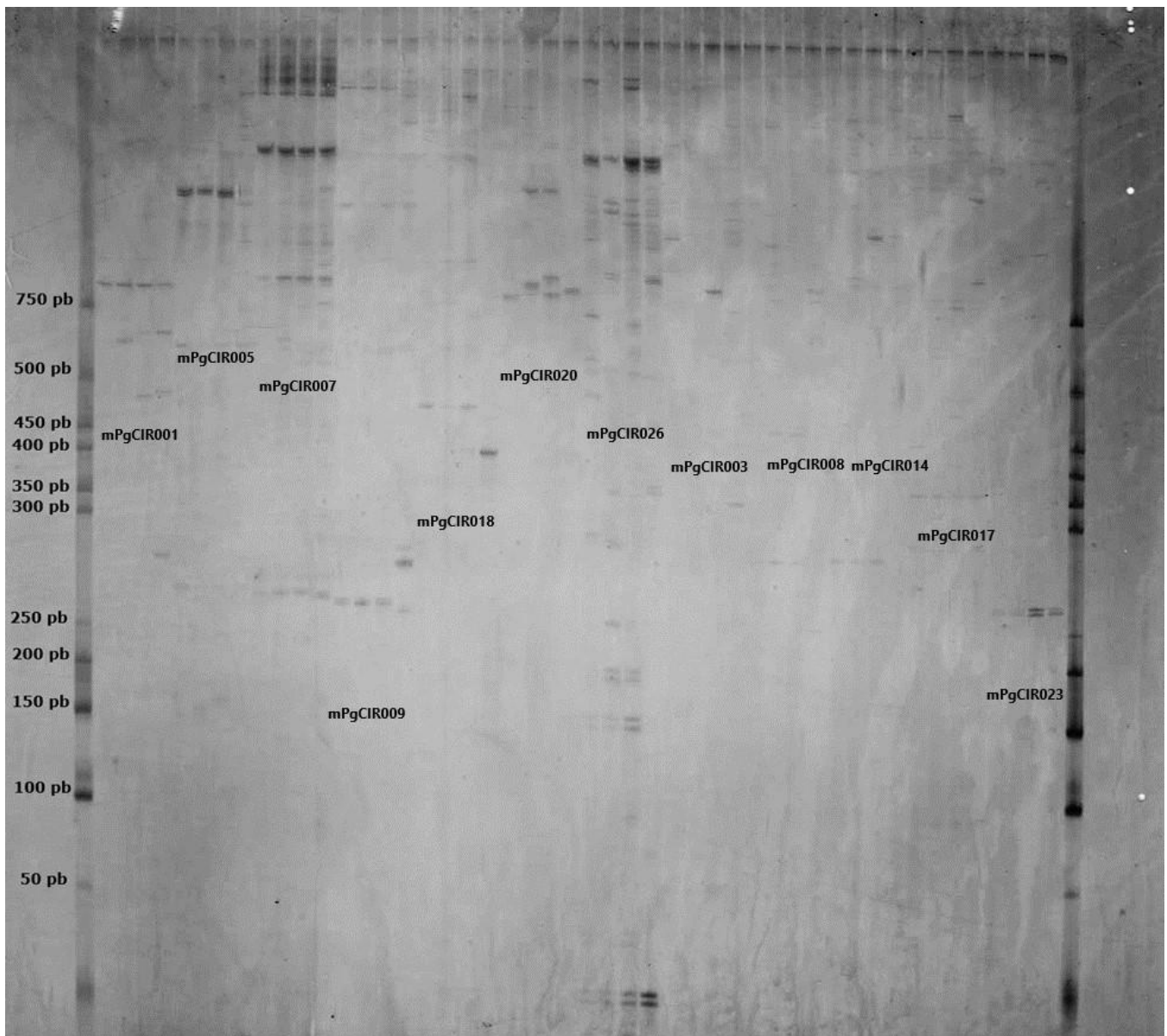


Figura 1. Resolução de amplicons de locus SSR de *Psidium guajava* em quatro plantas de *Eucalyptus deglupta*. 1 a 4 – mPgCIR001, mPgCIR005, mPgCIR007, mPgCIR009, 5 a 8: mPgCIR018, mPgCIR020, mPgCIR026, mPgCIR003, 9 a 12: mPgCIR008, mPgCIR014, mPgCIR017, mPgCIR023. A primeira e última coluna correspondem ao ladder 50pb Ludwig.

De acordo com o mapa genético de Lepitre *et al.* (2010) o locus mPgCIR005 se encontra no grupo de 1 de *P. guajava* e de *E. globulus*. O locus mPgCIR007 no grupo 2 de *P. guajava* e no grupo 10 de *E. globulus*. O mPgCIR018 no grupo de ligação 1 de *P. guajava* e no grupo 3 de *E. globulus*. O locus mPgCIR020 no grupo 7 de *P. guajava* e no 11 de *E. globulus*. Por fim o locus mPgCIR378 encontra-se no grupo 4 de *P. guajava* e no grupo 2 de *E. globulus*. Já de acordo com Padmakar *et al.* (2015) o locus mPgCIR005 se encontra no grupo de ligação 2 e 6

de *P. guajava*, o locus mPgCIR011 no grupo 2 e no *E. globulus* no grupo 11 e no mPgCIR018 está no grupo de ligação 3 de ambas as espécies.

DISCUSSÃO

Este primeiro estudo a propor o uso da ferramenta BLAST como forma de filtrar sequências SSR de *P. guajava* e prever *in silico* o anelamento das sequências em *E. globulus*, podendo assim estimar taxas de transferibilidade e orientar análises *in vitro*. Estudos que envolvem análises *in silico* necessitam levar em consideração não só a avaliação estatística mas também score dos alinhamentos, pois são informações importantes que trazem objetividade à comparação das sequências e avaliar a qualidade do alinhamento (Lobo 2008). De acordo com Pertsemlidis e Fondon (2001) e-values menores que 0,1 ou 0,05 são geralmente considerados de significância biológica. Ainda segundo os autores, a alta significância do e-value demonstra a conservação da sequência entre as duas espécies, o que é relevante tanto para comparação como para postular funções de proteínas.

Os marcadores disponibilizados por Risterucci *et al.* (2005) apresentaram elevada identidade e um valor teórico de transferibilidade de 39%, superior ao relatado por Siqueira (2014) que testou a transferibilidade de marcadores SSR gênicos do genoma de espécies de *Eucalyptus* para o genoma de *Eugenia klotzschiana* e obteve 9,2% de taxa de transferibilidade. Miranda (2014) obteve taxa de transferibilidade de 10% ao transferir EST-SSR de *Eucalyptus sp* para as espécies *Campomanesia adamantium* e *C. pubescens*. Ao utilizar a triagem *in silico* de marcadores microssatélites obtivemos resultados quatro vezes superiores aos encontrados na literatura em menor tempo e com menos recursos utilizados durante o processo.

Cabe e Marshall (2001) relatam que a chance de amplificação de qualquer sequência de DNA entre espécies diferentes através de PCR é inversamente proporcional à distância evolutiva entre elas, o que pode justificar a baixa transferibilidade entre os gêneros *Psidium* e *Eucalyptus*, também observada no presente estudo. Thornhill *et al.* (2015) estimam que houve uma separação entre as tribos que originaram estes gêneros entre 60 e 65 milhões de anos atrás, aumentando a distâncias evolutiva entre eles e possivelmente reduzindo as taxas de transferibilidade.

As altas identidades e e-values nos loci Guavamap, reportados no presente estudo, são atribuídos ao uso das sequências curtas (em torno de 20 nucleotídeos) dos marcadores analisados. Pertsemlidis e Fondon (2001) relatam que uma baixa complexidade na composição

de nucleotídeos de uma sequência ou sequências de curta periodicidade podem gerar altos valores de identidade, porém que não são confiáveis do ponto de vista biológico. A utilização dos parâmetros adotados no presente estudo, como alinhamento das sequências próximas no mesmo cromossomo, a menos de 300 nucleotídeos de distância e com baixo valor de e-value mostraram-se úteis para aproveitar as informações disponíveis na busca de homologias entre o genoma de *P. guajava* e *E. globulus*, podendo ser utilizada como indicativo da posição dos marcadores no genoma de *P. guajava*.

Todo dado gerado computacionalmente deve ser comparado à dados experimentais para validação (Verli 2011). Todos os loci de *Psidium* previstos de serem transferíveis para *Eucalyptus* na triagem *in silico* foram confirmados através de reação de PCR resultando em produtos de amplificação de boa qualidade. Rai *et al.* (2013) utilizando a metodologia proposta por Kuleung *et al.* (2004) classificou os produtos da PCR em quatro classes para confirmar a transferibilidade de marcadores, no presente estudo os locus mPgCIR001, mPgCIR005, mPgCIR007, mPgCIR009, mPgCIR018, mPgCIR020 e mPgCIR026 estão de acordo com as classes 1 e 2 (bandas de forte intensidade e fácil identificação e bandas de intensidade mais fraca mas ainda possível identificação), enquanto os demais locus utilizados (mPgCIR003, mPgCIR008, mPgCIR014, mPgCIR017 e mPgCIR023) estão de acordo com as classes 3 e 4 (bandas de intensidade muito fraca e difícil de identificação ou sem sinal).

A triagem *in silico* aqui descrita tem grande potencial para auxiliar estudos de transferibilidade, além de diversos trabalhos de mapeamento e genômica comparativa, visto que pode “filtrar” maior fluxo de informações moleculares antes das etapas experimentais. Jiang *et al.* (2020) relatam que o aumento de espécies a serem sequenciadas por tecnologias NGS aumentará a informação disponível em plataformas como GenBank, o que fornecerá muito material a ser explorado em estudos com as ferramentas do BLAST. Esta técnica pode ser aplicada em qualquer tipo de marcador com destaque para os EST-SSR, pois possuem maiores taxas de transferibilidade que SSR genômicos (Wu *et al.* 2014).

A partir do mapa de ligação de Lepitre *et al.* (2010) pôde-se inferir que o marcador mpPgCIR005 encontra-se em uma região homóloga nas duas espécies, assim como o locus mPgCIR018 de acordo com Padmakar *et al.* (2015). Ambos loci estão em grupos de ligação correspondentes nas duas espécies e a identificação de regiões como essas pode auxiliar na localização de genes de interesse para o melhoramento genético pois em regiões altamente conservadas existe grande chance de encontrar genes que evoluíram verticalmente do mesmo

MEDEIROS, F. L. B. Desenvolvimento de mapas genéticos e físicos com base em SNPs em *Psidium guajava* e triagem *in silico* de loci microssatélites de goiabeira para transferibilidade em *Eucalyptus globulus*

ancestral (Wei *et al.* 2002). Portanto o BLAST pode ser utilizado como indicativo da posição de locus não mapeados em goiabeira a partir de espécies geneticamente próximas como *Eucalyptus sp.*

CONCLUSÃO

A proposta pioneira de utilizar a ferramenta BLAST para triagem *in silico* de loci SSR de *P. guajava* foi realizada com sucesso ao prever o anelamento de sete loci microssatélites em *E. globulus*, os quais foram confirmados através da amplificação em gel de poliacrilamida. Os loci que não apresentaram “hits” ou e-value no Blastn também não produziram amplicons *in vitro*. Desta forma, a triagem *in silico* demonstra ser uma alternativa para orientar estudos genômicos entre as duas espécies, reduzindo custos e tempo.

REFERÊNCIAS BIBLIOGRÁFICAS

Acuña CV, Fernandez P, Villalba P V, García MN, Hopp HE and Poltri MSN (2012) Discovery, validation, and *in silico* functional characterization of EST-SSR markers in *Eucalyptus globulus*. **Tree Genetics and Genomes** 8: 289–301.

Butler JB, Vaillancourt RE, Potts BM, Lee DJ, King GJ, Baten A, Shepherd M and Freeman JS (2017) Comparative genomics of *Eucalyptus* and *Corymbia* reveals low rates of genome structural rearrangement. **BMC Genomics** 18: 1–13.

Cabe PR and Marshall KE (2001) Microsatellite loci from the house wren (*Troglodytes aedon*). **Molecular Ecology Notes** 1: 155–156.

Carrer H, Barbosa AL and Ramiro DA (2010) Biotecnologia na agricultura. **Estudos avançados** 24: 149–164.

Creste S, Tulmann Neto A and Figueira A (2001) Detection of Single Sequence Repeat Polymorphisms in Denaturing Polyacrylamide Sequencing Gels by Silver Staining. **Plant Molecular Biology Reporter** 19: 299–306.

dos Santos KL, Welter LJ, Dantas AC de M, Guerra MP, Ducroquet JPHJ and Nodari RO (2007) Transference of microsatellite markers from *Eucalyptus spp* to *Acca sellowiana* and the successful use of this technique in genetic characterization. **Genetics and Molecular Biology** 30: 73–79.

Fagundes BS, da Silva LF, Giacomini RM, Secco D, Díaz-Cruz JA and Da-Silva PR (2016)

MEDEIROS, F. L. B. Desenvolvimento de mapas genéticos e físicos com base em SNPs em *Psidium guajava* e triagem *in silico* de loci microssatélites de goiabeira para transferibilidade em *Eucalyptus globulus*

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** **9**: 408–422.

Grattapaglia D, Silva-Junior OB, Kirst M, de Lima BM, Faria DA and Pappas GJ (2011) High-throughput SNP genotyping in the highly heterozygous genome of Eucalyptus: Assay success, polymorphism and transferability across species. **BMC Plant Biology** **11**: 1–18.

Hudson CJ, Kullan ARK, Freeman JS, Faria DA, Grattapaglia D, Kilian A, Myburg AA, Potts BM and Vaillancourt RE (2012) High synteny and colinearity among Eucalyptus genomes revealed by high-density comparative genetic mapping. **Tree Genetics and Genomes** **8**: 339–352.

Inglis PW, Marilia de Castro RP, Resende LV. and Grattapaglia D (2018) Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. **PLoS ONE** **13**: 1–14.

Jiang Y, Xu S, Wang R, Zhou J, Dou J, Yin Q and Wang R (2020) Characterization, validation, and cross-species transferability of EST-SSR markers developed from *Lycoris aurea* and their application in genetic evaluation of *Lycoris* species. **BMC Plant Biology** **20**: 522–536.

Kuleung C, Baenziger PS and Dweikat I (2004) Transferability of SSR markers among wheat, rye, and triticale. **Theoretical and Applied Genetics** **108**: 1147–1150.

Lepitre V, Nansot G, Grangeon R, Pomies V, Rivallan R, Risterucci AM, Valdés-Infante J, Rodríguez-Medina NN, Muth J, Boike J, Prüfer D, Becker D, Rohde W, Ritter E and Billotte N (2010) The microsatellite (SSR)/AFLP reference linkage map of guava. **Acta Horticulturae** **849**: 183–192.

Lobo I (2008) Basic local alignment search tool (BLAST). **Biology and fertility of soils** **1**: 215–223.

Padmakar B, Kanupriya C, Latha PM, Prashant KS, Dinesh MR, Sailaja D and Aswath C (2015) Development of SRAP and SSR marker-based genetic linkage maps of guava (*Psidium guajava* L.). **Scientia Horticulturae** **192**: 158–165.

Pertsemlidis A and Fondon JW (2001) Having a BLAST with bioinformatics (and avoiding BLAST phemy). **Genome Biology** **2**: 1–10.

MEDEIROS, F. L. B. Desenvolvimento de mapas genéticos e físicos com base em SNPs em *Psidium guajava* e triagem *in silico* de loci microssatélites de goiabeira para transferibilidade em *Eucalyptus globulus*

Rafalski JA, Vogel JM, Morgante M, Powell IW, Andre C and Tingey S V (1996) Generating and Using DNA markers in Plants. In: Birren B and Lai E (eds) **Analysis of Nonmammalian Genomes: A Practical Guide**, Academic Press, Boca Raton, p. 75–134.

Rai MK, Phulwaria M and Shekhawat NS (2013) Transferability of simple sequence repeat (SSR) markers developed in guava (*Psidium guajava* L.) to four Myrtaceae species. **Molecular Biology Reports** **40**: 5067–5071.

Risterucci AM, Duval MF, Rohde W and Billotte N (2005) Isolation and characterization of microsatellite loci from *Psidium guajava* L. **Molecular Ecology Notes** **5**: 745–748.

Shepherd M, Kasem S, Lee D, Henry R (2006) Construction of micro-satellite linkage maps for *Corymbia*. **Silvae Genet** **55**: 228–238.

Singleton JJ, Mangat PK, Shim J, Vavra C, Coldren C and Angeles-Shim RB (2020) Cross-species transferability of *Solanum* spp. DNA markers and their application in assessing genetic variation in silverleaf nightshade (*Solanum elaeagnifolium*) populations from Texas, USA. **Weed Science** **68**: 396–404.

Siqueira MN de (2014) **Transferibilidade e Variabilidade genética nos microssatélites gênicos em *Eugenia klotzschiana* BERG (Myrtaceae)**. Dissertação de Mestrado (Universidade Federal de Goiás), 88p.

Sitther V, Zhang D, Harris DL, Yadav AK, Zee FT, Meinhardt LW and Dhekney SA (2014) Genetic characterization of guava (*Psidium guajava* L.) germplasm in the United States using microsatellite markers. **Genetic Resources and Crop Evolution** **61**: 829–839.

Sumathi M, Bachpai VKW, Mayavel A, Dasgupta MG, Nagarajan B, Rajasugunasekar D, Sivakumar V and Yasodha R (2018) Genetic linkage map and QTL identification for adventitious rooting traits in red gum eucalypts. **3 Biotech** **8**: 242.

Taheri S, Abdullah TL, Yusop MR, Hanafi MM, Sahebi M, Azizi P and Shamschiri RR (2018) Mining and development of novel SSR markers using Next Generation Sequencing (NGS) data in plants. **Molecules** **23**: 399–417.

Thornhill AH, Ho SYW, Külheim C and Crisp MD (2015) Interpreting the modern distribution of Myrtaceae using a dated molecular phylogeny. **Molecular Phylogenetics and Evolution** **93**: 29–43.

MEDEIROS, F. L. B. Desenvolvimento de mapas genéticos e físicos com base em SNPs em *Psidium guajava* e triagem *in silico* de loci microssatélites de goiabeira para transferibilidade em *Eucalyptus globulus*

Varshney RK, Sigmund R, Börner A, Korzun V, Stein N, Sorrells ME, Langridge P and Graner A (2005) Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. **Plant Science** **168**: 195–202.

Varshney RK, Thiel T, Stein N, Langridge P and Graner A (2002) *In Silico* Analysis on Frequency and Distribution of microsatellites in ESTs of some cereal species. **Cellular & Molecular Biology letters** **7**: 537–546.

Verli H (2011) **Bioinformática da Biologia à Flexibilidade Molecular**. Sociedade Brasileira de Bioquímica e Biologia Molecular, São Paulo, 282p.

Wei L, Liu Y, Dubchak I, Shon J and Park J (2002) Comparative genomics approaches to study organism similarities and differences. **Journal of Biomedical Informatics** **35**: 142–150.

Wu J, Cai C, Cheng F, Cui H and Zhou H (2014) Characterisation and development of EST-SSR markers in tree peony using transcriptome sequences. **Molecular Breeding** **34**: 1853–1866.

Zhang Z, Schwartz S, Wagner L and Miller W (2000) A greedy algorithm for aligning DNA sequences. **Journal of Computational Biology** **7**: 203–214.

Zou C, Karn A, Reisch B, Nguyen A, Sun Y, Bao Y, Campbell MS, Church D, Williams S, Xu X, Ledbetter CA, Patel S, Fennell A, Glaubitz JC, Clark M, Ware D, Londo JP, Sun Q and Cadle-Davidson L (2020) Haplotyping the *Vitis* collinear core genome with rhAmpSeq improves marker transferability in a diverse genus. **Nature Communications** **11**: 1–11.

MEDEIROS, F. L. B. Desenvolvimento de mapas genéticos e físicos com base em SNPs em *Psidium guajava* e triagem *in silico* de loci microssatélites de goiabeira para transferibilidade em *Eucalyptus globulus*

ANEXOS

ANEXO I – RESULTADOS DO BLAST DE SEQUÊNCIAS DE SNPs DE EUCALIPTO CONTRA O GENOMA DA GOIABEIRA (*P. guajava*).

Tabela 1 Resultados BLAST das sequências de 694 SNPS (query) de *Eucalyptus* utilizadas no mapa físico de *Psidium guajava*, contendo o cromossomo (Chr.), posição no genoma (Subject start e subject end), e-value e bit score.

Query	Chr. Goiaba	Subject start	Subject end	e-value	bit score
AX-299485552	1	1376372	1376440	1,95E-17	89,7
AX-299484757	1	1963535	1963466	2,39E-16	86,0
AX-297654366	1	2106025	2105955	1,40E-25	116,0
AX-297652503	1	3780033	3780103	7,31E-23	107,0
AX-299354514	1	3902661	3902608	5,27E-12	72,5
AX-297652910	1	4298830	4298898	3,79E-20	99,6
AX-297652910	1	4323791	4323861	3,79E-20	98,7
AX-299570333	1	5012208	5012278	7,31E-23	107,0
AX-168367609	1	5125397	5125466	1,09E-20	101
AX-297216244	1	5156108	5156178	3,09E-21	103,0
AX-297216244	1	5201485	5201555	3,09E-21	103,0
AX-299513164	1	6062361	6062431	1,61E-18	94,2
AX-299570841	1	7190006	7190076	1,02E-14	80,6
AX-299570932	1	7878114	7878184	7,31E-23	107,0
AX-168367724	1	8206596	8206666	1,61E-18	94,2
AX-299571091	1	8823475	8823541	1,60E-18	94,2
AX-301086394	1	10038975	10038909	6,38E-11	60,2
AX-299571443	1	10670471	10670539	1,01E-14	81,5
AX-297217694	1	12517767	12517831	2,91E-15	8,0
AX-300019998	1	12691154	12691224	3,09E-21	103,0
AX-297217802	1	12862150	12862214	8,35E-16	85,1
AX-299530918	1	13033706	13033775	4,93E-25	114,0
AX-301155042	1	15148641	15148572	1,51E-12	74,3
AX-299554808	1	15681282	15681220	2,24E-10	66,2
AX-297218573	1	16119397	16119467	1,61E-18	94,2
AX-299573022	1	17536138	17536208	3,09E-21	103,0
AX-297215763	1	18218738	18218797	1,51E-12	74,3
AX-300756980	1	22474807	22474877	8,35E-16	85,1
AX-297167229	1	24109126	24109191	2,24E-10	67,1
AX-297189600	1	24483483	24483413	3,11E-21	103,0
AX-297213059	1	25677825	25677884	7,82E-10	65,3
AX-297220823	1	27397777	27397852	4,32E-13	76,1
AX-299482594	1	27926876	27926945	1,96E-17	90,6
AX-301076213	1	28350334	28350404	7,31E-23	107
AX-301077609	1	29302834	29302904	5,97E-24	112,0
AX-299573823	1	29495636	29495706	3,79E-20	98,7
AX-297605525	1	30501182	30501241	1,47E-12	74,3

Query	Chr. Goiaba	Subject start	Subject end	e-value	bit score
AX-299573452	1	31122628	31122698	6,00E-24	112
AX-297219336	1	31131067	31131136	4,62E-19	95,1
AX-299576197	1	34500826	34500762	2,23E-10	66,2
AX-299576176	1	34644685	34644614	6,85E-17	88,7
AX-300043518	1	35418796	35418741	5,27E-12	71,6
AX-299575415	1	36937565	36937495	4,32E-13	76,1
AX-301142212	1	37892880	37892813	4,32E-13	75,2
AX-297152884	1	38214824	38214752	7,82E-10	65,3
AX-299860157	2	36234	36303	5,62E-18	92,4
AX-300708797	2	747791	747852	1,84E-11	70,7
AX-297516002	2	1273334	1273400	4,62E-19	96,0
AX-299546148	2	1284938	1285008	7,27E-23	107,0
AX-299870202	2	1315971	1316037	1,00E-13	76,1
AX-297192225	2	1996346	1996416	5,27E-12	71,6
AX-299546356	2	2085119	2085189	1,15E-26	121,0
AX-299856627	2	2643974	2644043	2,55E-22	105,0
AX-299630836	2	2774786	2774843	4,32E-13	75,2
AX-297518053	2	2933179	2933249	1,96E-17	89,7
AX-297149312	2	3598336	3598403	8,30E-16	84,2
AX-299546648	2	3676661	3676731	7,31E-23	107
AX-299873413	2	3802094	3802164	1,61E-18	94,2
AX-297192676	2	4132064	4132136	3,79E-20	99,6
AX-297191590	2	6097609	6097539	3,11E-21	103,0
AX-297171550	2	6849742	6849672	1,96E-17	89,7
AX-299866289	2	7756166	7756105	2,90E-15	82,4
AX-299545099	2	8286221	8286151	3,11E-21	103,0
AX-297511545	2	8319935	8320004	3,00E-15	82,4
AX-299865616	2	8352698	8352633	1,61E-18	94,2
AX-299545005	2	8458575	8458507	1,96E-17	90,6
AX-301110418	2	8934140	8934210	6,00E-24	112,0
AX-297510028	2	9584942	9584872	3,79E-20	98,7
AX-301127036	2	10373474	10373544	6,00E-24	112
AX-299544113	2	11895745	11895675	1,60E-18	94,2
AX-297189949	2	12564138	12564068	7,31E-23	107,0
AX-299354060	2	12564633	12564563	3,77E-20	98,7
AX-297189893	2	12882051	12881984	3,55E-14	79,7
AX-297254253	2	12974388	12974318	1,96E-17	89,7
AX-299443851	2	13079808	13079742	2,39E-16	86,9
AX-299515222	2	13138911	13138845	2,39E-16	86,9
AX-299543787	2	13796635	13796710	8,00E-16	85,1
AX-297505758	2	13993862	13993792	3,09E-21	103,0
AX-297505758	2	14016085	14016019	5,59E-18	91,5
AX-297505758	2	14026333	14026267	2,38E-16	86,9
AX-297504448	2	15551737	15551667	3,11E-21	103,0
AX-297189195	2	16913891	16913821	8,35E-16	85,1

Query	Chr. Goiaba	Subject start	Subject end	e-value	bit score
AX-297189169	2	20110950	20110880	6,00E-24	112
AX-297217070	2	20500877	20500807	3,09E-21	103,0
AX-297501964	2	20932158	20932220	2,24E-10	66,2
AX-299543139	2	21127806	21127736	7,31E-23	107,0
AX-297320728	2	22333503	22333570	1,24E-13	77,9
AX-299466285	2	23003011	23002941	1,96E-17	89,7
AX-297477722	2	23261553	23261483	3,77E-20	98,7
AX-299853400	2	25553473	25553412	7,77E-10	64,4
AX-299853400	2	25784053	25783984	1,83E-11	69,8
AX-299853400	2	25820916	25820847	3,53E-14	78,8
AX-299311182	2	27922091	27922162	6,85E-17	87,8
AX-299542251	2	28899787	28899717	3,79E-20	98,7
AX-301013379	2	28904912	28904844	2,39E-16	86,0
AX-299847290	2	30295752	30295689	6,38E-11	68
AX-297187717	2	31203189	31203119	1,61E-18	94,2
AX-300650627	2	33266519	33266449	3,79E-20	98,7
AX-299541314	2	33699949	33700007	2,00E-11	70,7
AX-299541314	2	33734507	33734450	2,24E-10	67,1
AX-297155540	2	35263138	35263069	3,79E-20	99,6
AX-301013400	2	36558163	36558093	3,79E-20	98,7
AX-299540727	2	37037394	37037464	1,61E-18	94,2
AX-301124820	2	37478961	37478902	6,42E-11	68
AX-299540787	2	37500892	37500960	8,91E-22	104
AX-297151673	2	38131970	38131900	1,96E-17	89,7
AX-297484481	2	36080494	36080563	1,00E-19	96,6
AX-299462862	3	10976	11045	1,51E-12	74,3
AX-297481061	3	472169	472239	3,00E-21	103,0
AX-299573686	3	746041	746111	3,11E-21	103
AX-297186349	3	817670	817595	7,82E-10	64,4
AX-301124515	3	1008587	1008520	3,11E-21	102,0
AX-299606704	3	1315971	1316037	4,32E-13	76,1
AX-297459928	3	1566501	1566570	6,81E-17	87,8
AX-299536149	3	1631462	1631530	5,59E-18	91,5
AX-168367469	3	1805033	1805103	3,11E-21	103,0
AX-297181907	3	1822610	1822679	6,81E-17	87,8
AX-297181907	3	1840766	1840835	3,53E-14	78,8
AX-297181907	3	1869367	1869436	3,53E-14	78,8
AX-297182771	3	1874389	1874459	1,96E-17	89,7
AX-297146221	3	3199582	3199652	1,61E-18	94,2
AX-299536575	3	5354779	5354848	1,32E-19	96,9
AX-299536583	3	5407416	5407486	7,31E-23	107
AX-299536606	3	5473542	5473609	1,02E-14	80,6
AX-297506421	3	5704147	5704217	7,31E-23	107,0
AX-297462137	3	5777373	5777443	1,01E-14	80,6
AX-297182690	3	6378469	6378539	1,61E-18	94,2

Query	Chr. Goiaba	Subject start	Subject end	e-value	bit score
AX-299539125	3	7953269	7953199	1,61E-18	94,2
AX-297148418	3	9094864	9094934	1,02E-14	80,6
AX-299825583	3	9827430	9827362	3,79E-20	99,6
AX-297177652	3	11630617	11630687	3,79E-20	98,7
AX-301144169	3	12463085	12463015	3,11E-21	103,0
AX-300601484	3	14975698	14975628	3,79E-20	98,7
AX-297185227	3	16272191	16272257	1,01E-14	80,6
AX-299463312	3	16285121	16285189	3,79E-20	99,6
AX-298965255	3	17551176	17551246	3,11E-21	103,0
AX-299555239	3	17927679	17927616	3,55E-14	79,7
AX-298966795	3	18223382	18223446	7,82E-10	65,3
AX-297186207	3	18692751	18692683	2,91E-15	82,4
AX-299831875	3	18757792	18757854	4,32E-13	75,2
AX-299531762	3	18780005	18779941	1,32E-19	97,8
AX-168333006	3	19014188	19014258	1,40E-25	116,0
AX-300628892	3	19345081	19345151	3,00E-12	77,0
AX-297186207	3	20314978	20315047	8,35E-16	84,2
AX-297479939	3	20366078	20366148	2,00E-18	94,2
AX-297506419	3	20415279	20415349	2,24E-10	67,1
AX-298963244	3	20450638	20450567	5,62E-18	92,4
AX-299539314	3	20764335	20764270	1,24E-13	77,0
AX-297457063	3	22426989	22426919	5,24E-12	71,6
AX-297457057	3	22433809	22433738	1,32E-19	96,9
AX-297454294	3	24955306	24955238	3,79E-20	99,6
AX-299461977	3	25561165	25561236	6,85E-17	87,8
AX-299807831	3	25563622	25563552	1,02E-14	80,6
AX-299534612	3	27549752	27549817	4,00E-13	76,1
AX-297180499	3	27554710	27554643	1,32E-19	97,8
AX-297226591	3	28180609	28180677	5,24E-12	72,5
AX-299824553	3	31243402	31243328	6,42E-11	68,9
AX-300566352	3	31656625	31656555	7,31E-23	107
AX-297169932	3	32128922	32128850	8,35E-16	85,1
AX-299459859	3	32431625	32431552	6,85E-17	88,7
AX-299459859	3	33404050	33403980	1,61E-18	94,2
AX-299534267	3	33878736	33878666	6,00E-24	112
AX-299533361	3	39027914	39027844	1,96E-17	89,7
AX-297179096	3	39609773	39609703	7,31E-23	107,0
AX-297178709	3	40069625	40069558	3,11E-21	102
AX-299532380	3	41864408	41864353	7,82E-10	65,3
AX-299532235	3	42405775	42405705	4,32E-13	76,1
AX-299532153	3	42969446	42969380	6,42E-11	68,9
AX-297436571	3	44273539	44273473	4,62E-19	96,0
AX-299496182	3	46474592	46474524	5,49E-18	91,5
AX-299531002	3	47299752	47299683	1,96E-17	90,6
AX-297142967	3	49271743	49271811	5,27E-12	72,5

Query	Chr. Goiaba	Subject start	Subject end	e-value	bit score
AX-299547566	4	2510506	2510576	3,11E-21	103,0
AX-297524446	4	3179996	3180066	8,35E-16	85,1
AX-297193559	4	3228519	3228589	7,31E-23	107
AX-297193601	4	3332112	3332182	3,11E-21	103,0
AX-299634480	4	4786198	4786268	1,00E-14	80,6
AX-297193831	4	4798362	4798432	3,11E-21	103,0
AX-297527945	4	6243410	6243476	2,91E-15	82,4
AX-298832729	4	6482290	6482355	8,00E-16	85,1
AX-299354060	4	6539552	6539491	1,50E-12	73,4
AX-297194230	4	6953284	6953354	7,00E-23	107,0
AX-297194230	4	6983189	6983259	3,00E-21	103,0
AX-297194230	4	7052671	7052741	3,00E-21	103,0
AX-297529223	4	7304355	7304420	4,00E-13	76,1
AX-297194521	4	8459309	8459377	6,42E-11	68,0
AX-297194789	4	9663926	9663997	6,85E-17	87,8
AX-299549176	4	10699140	10699208	1,96E-17	90,6
AX-297195092	4	10808688	10808759	2,91E-15	83,3
AX-301130075	4	11450247	11450178	2,10E-23	110,0
AX-299549461	4	11630617	11630687	3,79E-20	98,7
AX-299549518	4	11848497	11848567	6,00E-24	112
AX-299549606	4	12445756	12445826	1,61E-18	94,2
AX-297208068	4	12802182	12802115	2,24E-10	67,1
AX-299549735	4	12814239	12814309	1,61E-18	94,2
AX-299549880	4	13286608	13286676	6,42E-11	68,9
AX-299504391	4	13292876	13292946	8,35E-16	85,1
AX-299634480	4	13982969	13983035	2,00E-12	73,4
AX-300760517	4	14087408	14087477	2,55E-22	105,0
AX-297196106	4	14470222	14470294	1,51E-12	73,4
AX-299550307	4	14656420	14656490	1,60E-18	94,2
AX-297196496	4	15488156	15488222	1,51E-12	73,4
AX-297196496	4	15526284	15526351	2,24E-10	66,2
AX-299546356	4	15559841	15559911	7,27E-23	107,0
AX-297192246	4	15559982	15560043	2,39E-16	86,9
AX-297196496	4	15571719	15571789	7,31E-23	107
AX-299521277	4	16936485	16936549	1,84E-11	70,7
AX-297202429	4	22824624	22824689	2,00E-10	67,1
AX-300788885	4	25710523	25710458	5,27E-12	71,6
AX-297148273	4	26088336	26088273	1,96E-17	90,6
AX-297198342	4	28828709	28828777	2,39E-16	86,0
AX-299634480	4	29231413	29231479	2,00E-12	73,4
AX-299552265	4	29392798	29392860	6,85E-17	87,8
AX-301130805	4	29462795	29462866	6,85E-17	87,8
AX-297202046	4	30260663	30260594	3,55E-14	78,8
AX-299505311	4	31526713	31526651	4,32E-13	75,2
AX-297151179	4	32676676	32676606	2,23E-10	67,1

Query	Chr. Goiaba	Subject start	Subject end	e-value	bit score
AX-299667495	4	32831915	32831983	6,42E-11	68,0
AX-300806712	4	33169079	33169151	1,24E-13	77
AX-300123453	4	33503428	33503497	1,08E-20	101
AX-300123453	4	33583472	33583538	2,90E-15	82,4
AX-300123453	4	33597175	33597244	2,90E-15	83,3
AX-299914940	4	33715099	33715166	6,85E-17	88,7
AX-299634480	4	33779558	33779623	4,00E-13	76,1
AX-299354233	4	34253910	34253989	1,24E-13	77,9
AX-299553900	4	35533164	35533234	3,11E-21	103
AX-301132688	4	36129698	36129762	3,55E-14	78,8
AX-299462004	4	36796878	36796947	3,55E-14	79,7
AX-299554296	4	37311649	37311714	1,02E-14	80,6
AX-299554334	4	37377597	37377667	3,11E-21	103,0
AX-297566772	4	38622197	38622267	7,31E-23	107,0
AX-299451377	4	40698918	40698848	1,96E-17	89,7
AX-301133645	4	41210315	41210249	3,53E-14	78,8
AX-297201303	4	42299943	42300011	8,91E-22	104,0
AX-299176714	4	43740770	43740840	7,31E-23	107,0
AX-301153421	4	43820875	43820938	2,00E-16	86,0
AX-299928934	4	44667296	44667366	3,11E-21	103
AX-301134813	4	46793912	46793983	6,85E-17	87,8
AX-299932809	4	47214796	47214863	1,51E-12	73,4
AX-299189043	4	47796191	47796261	3,11E-21	103,0
AX-298852055	4	48028611	48028541	1,41E-25	116
AX-299546637	5	44238314	44238384	5,00E-21	98,7
AX-299542647	5	81186	81256	8,35E-16	85,1
AX-299507499	5	926305	926374	1,32E-19	96,9
AX-300196947	5	1320288	1320355	1,61E-18	93,3
AX-300197641	5	1525667	1525726	5,27E-12	71,6
AX-297153770	5	2288133	2288201	1,96E-17	90,6
AX-299507887	5	2438827	2438894	6,85E-17	88,7
AX-299645640	5	2556588	2556653	1,01E-14	81,5
AX-299507944	5	2651290	2651360	3,09E-21	103,0
AX-297292094	5	3138266	3138318	1,84E-11	70,7
AX-299487813	5	3445730	3445800	6,00E-24	112
AX-297154076	5	3671221	3671291	7,31E-23	107,0
AX-299442792	5	3683314	3683383	6,42E-11	68,0
AX-299734102	5	3971827	3971897	3,11E-21	103,0
AX-299647148	5	4108661	4108727	6,85E-17	87,8
AX-300204995	5	4682391	4682461	7,31E-23	107
AX-299534593	5	6218730	6218800	7,31E-23	107,0
AX-299650259	5	6893567	6893631	4,32E-13	75,2
AX-297296235	5	6932064	6932134	3,79E-20	98,7
AX-297296289	5	6998915	6998984	4,30E-13	75,2
AX-299508839	5	7233344	7233396	6,42E-11	68,9

Query	Chr. Goiaba	Subject start	Subject end	e-value	bit score
AX-298554595	5	8074285	8074355	1,41E-25	116,0
AX-297155057	5	8140483	8140555	2,39E-16	86
AX-297155066	5	8174508	8174578	1,96E-17	89,7
AX-299509191	5	8809592	8809662	3,11E-21	103
AX-299509204	5	8832339	8832409	5,97E-24	112,0
AX-299514825	5	8874698	8874768	3,11E-21	103,0
AX-299652590	5	9211777	9211847	3,11E-21	103,0
AX-299652591	5	9212038	9212108	7,31E-23	107
AX-297155511	5	9917869	9917934	1,96E-17	89,7
AX-297155540	5	10027145	10027214	2,00E-16	86,0
AX-297299594	5	10191018	10191088	6,42E-11	68,9
AX-299509674	5	10214833	10214900	2,00E-11	70,7
AX-297300107	5	10479248	10479317	1,51E-12	74,3
AX-297300233	5	10599851	10599920	4,93E-25	114
AX-299520393	5	11285359	11285429	8,35E-16	85,1
AX-297155970	5	11378856	11378926	3,11E-21	103,0
AX-299655165	5	11387991	11388059	2,39E-16	86,0
AX-299510053	5	11449924	11449993	1,08E-20	101,0
AX-297156424	5	12509841	12509911	6,00E-24	112,0
AX-297156443	5	12541226	12541296	3,09E-21	103,0
AX-298766288	5	13322010	13321940	7,31E-23	107,0
AX-297156815	5	13547169	13547234	1,24E-13	77,9
AX-299658298	5	14112849	14112919	1,96E-17	89,7
AX-299511028	5	14210717	14210787	8,35E-16	85,1
AX-297156995	5	14234789	14234858	1,32E-19	96,9
AX-299659916	5	15294438	15294501	6,42E-11	68,9
AX-299660088	5	15401014	15401084	6,00E-24	112
AX-301104590	5	15439427	15439357	7,27E-23	107,0
AX-299641558	5	17217787	17217718	2,55E-22	105
AX-299507194	5	17329502	17329435	2,55E-22	106,0
AX-299641449	5	17336924	17336868	7,82E-10	64,4
AX-300023459	5	17660652	17660582	4,32E-13	76,1
AX-297286337	5	17959363	17959298	1,02E-14	80,6
AX-297152914	5	18432885	18432817	5,27E-12	72,5
AX-297156905	5	18529966	18529892	1,09E-20	100,0
AX-301108779	5	18540508	18540442	2,38E-16	86,9
AX-299639422	5	18579119	18579064	5,27E-12	71,6
AX-297520491	5	19853554	19853486	2,39E-16	86,0
AX-297152618	5	19959797	19959727	3,11E-21	103,0
AX-297824655	5	21453064	21452994	3,11E-21	103,0
AX-297152190	5	21575755	21575824	2,54E-22	105,0
AX-297244453	5	21894394	21894464	2,00E-10	67,1
AX-301138266	5	22512042	22511972	2,24E-10	67,1
AX-299505875	5	23145086	23145016	4,30E-13	76,1
AX-299633015	5	23548577	23548509	2,24E-10	66,2

Query	Chr. Goiaba	Subject start	Subject end	e-value	bit score
AX-301108779	5	25633754	25633694	2,24E-10	67,1
AX-299997522	5	26781530	26781462	6,42E-11	68,0
AX-299840554	5	29892096	29892166	2,24E-10	67,1
AX-297186207	5	32374898	32374967	8,35E-16	84,2
AX-300423262	5	33924404	33924339	2,38E-16	86,0
AX-299150754	5	34918996	34918927	2,91E-15	83,3
AX-298832729	5	35176286	35176350	2,00E-12	74,3
AX-168367334	5	36437996	36438065	1,02E-14	81,5
AX-299840554	5	38525269	38525200	2,55E-22	105,0
AX-299504335	5	38781093	38781024	2,39E-16	86,0
AX-299504317	5	38888204	38888134	3,11E-21	103,0
AX-299442897	5	40615111	40615181	3,11E-21	103,0
AX-299620237	5	40928507	40928436	2,91E-15	83,3
AX-297265594	5	41420406	41420340	4,62E-19	96,0
AX-297242512	5	41421031	41420961	1,96E-17	89,7
AX-297264962	5	41713321	41713251	1,96E-17	89,7
AX-299617699	5	42798032	42797961	6,85E-17	87,8
AX-299617631	5	42842049	42841979	3,79E-20	98,7
AX-297149475	5	43695319	43695249	3,11E-21	103,0
AX-299615875	5	44121723	44121653	3,11E-21	103,0
AX-299566143	5	44238384	44238314	3,79E-20	98,7
AX-297149274	5	44460019	44459949	1,96E-17	89,7
AX-297308559	6	932048	932115	3,11E-21	102,0
AX-299664601	6	2102741	2102807	2,91E-15	82,4
AX-297310517	6	2128087	2128157	1,61E-18	94,2
AX-297310585	6	2160097	2160167	7,31E-23	107,0
AX-297311457	6	2755291	2755353	2,23E-10	66,2
AX-299665575	6	2757220	2757288	8,91E-22	104,0
AX-168367358	6	2757883	2757953	3,79E-20	98,7
AX-297313012	6	3224600	3224668	2,24E-10	67,1
AX-301107296	6	3757366	3757302	1,84E-11	69,8
AX-299667346	6	4009799	4009859	5,00E-12	71,6
AX-168332667	6	4125074	4125138	1,51E-12	74,3
AX-299605875	6	4231954	4231896	5,27E-12	72,5
AX-299513134	6	5908929	5908999	1,02E-14	80,6
AX-297159111	6	6027384	6027454	6,00E-24	112,0
AX-299513201	6	6181824	6181891	1,83E-11	69,8
AX-297199227	6	6803042	6803112	1,96E-17	89,7
AX-298602366	6	9197511	9197582	1,09E-20	101
AX-299701127	6	9295461	9295525	1,51E-12	74,3
AX-297319198	6	9511131	9511201	1,95E-17	89,7
AX-299448635	6	11214577	11214511	1,51E-12	74,3
AX-297160373	6	12028579	12028647	8,91E-22	104,0
AX-299678579	6	12615455	12615525	4,32E-13	76,1
AX-297160736	6	13202525	13202595	3,11E-21	103,0

Query	Chr. Goiaba	Subject start	Subject end	e-value	bit score
AX-297173112	6	13381092	13381151	7,82E-10	65,3
AX-299679668	6	13410013	13410081	2,39E-16	86,0
AX-299680141	6	13848028	13848084	1,83E-11	70,7
AX-299680181	6	13875781	13875851	3,79E-20	98,7
AX-297328305	6	15746725	15746795	1,96E-17	90,6
AX-299701127	6	15775986	15775916	3,79E-20	98,7
AX-297331556	6	19353188	19353258	1,02E-14	80,6
AX-299594117	6	23315314	23315384	1,61E-18	94,2
AX-301108779	6	25633754	25633694	2,23E-10	67,1
AX-299353413	6	26085428	26085498	3,11E-21	103
AX-299689848	6	26088457	26088516	2,91E-15	83,3
AX-299516472	6	26794867	26794937	6,00E-24	112,0
AX-297336465	6	27234735	27234803	1,83E-11	70,7
AX-297336538	6	27384034	27384102	1,41E-25	117,0
AX-299632171	6	27924933	27924996	8,00E-10	65,3
AX-297162552	6	28035890	28035959	2,55E-22	105,0
AX-297202046	6	30260663	30260663	4,00E-14	78,8
AX-297340461	6	30468599	30468669	1,96E-17	89,7
AX-168367385	6	30851443	30851513	7,31E-23	107,0
AX-297163031	6	31105088	31105157	7,82E-10	65,3
AX-297244453	6	31584749	31584813	1,10E-11	68,0
AX-299695693	6	31767124	31767190	1,00E-13	77,9
AX-297342146	6	32179888	32179958	1,96E-17	89,7
AX-299517302	6	32237999	32238067	2,39E-16	86,0
AX-299696798	6	32856773	32856841	1,23E-13	77,9
AX-297254513	6	33870962	33871027	5,27E-12	71,6
AX-297163486	6	33927904	33927973	2,10E-23	110,0
AX-297343847	6	33947767	33947837	1,01E-14	80,6
AX-297344725	6	34759587	34759651	2,91E-15	83,3
AX-297163675	6	35246204	35246267	2,24E-10	67,1
AX-299517800	6	35507384	35507455	1,32E-19	96,9
AX-299517815	6	35580174	35580244	1,60E-18	94,2
AX-299242244	6	35919673	35919743	8,35E-16	85,1
AX-297182690	6	36226050	36225980	5,27E-12	71,6
AX-298673319	6	38266231	38266302	2,55E-22	105,0
AX-297347948	6	38659444	38659508	1,09E-20	101
AX-297164203	6	38668095	38668161	1,84E-11	70,7
AX-299528618	6	38835384	38835314	3,11E-21	10,0
AX-299703043	6	40208366	40208436	3,00E-21	103,0
AX-299689969	6	42799373	42799441	4,62E-19	95,1
AX-299524941	7	145869	145799	3,77E-20	98,7
AX-299524914	7	210776	210842	2,00E-11	70,7
AX-299743623	7	221163	221233	1,00E-14	80,6
AX-297389374	7	488871	488801	1,01E-14	80,6
AX-297170782	7	509476	509539	9,00E-22	104,0

Query	Chr. Goiaba	Subject start	Subject end	e-value	bit score
AX-297170766	7	615474	615404	1,96E-17	89,7
AX-299524673	7	1516578	1516646	5,00E-19	95,1
AX-299840554	7	1674553	1674483	2,24E-10	67,1
AX-297162236	7	1690695	1690625	2,24E-10	67,1
AX-298771173	7	1847725	1847658	8,35E-16	84,2
AX-299741946	7	2093126	2093056	7,31E-23	107,0
AX-299634480	7	3385340	3385405	5,00E-12	71,6
AX-299739916	7	3844105	3844041	7,82E-10	64,4
AX-297169932	7	3877540	3877476	2,39E-16	86
AX-299739546	7	4184107	4184176	4,00E-14	78,8
AX-300423262	7	4502826	4502756	1,01E-14	80,6
AX-301128072	7	5115350	5115427	6,42E-11	68,9
AX-297390577	7	5187159	5187223	2,24E-10	66,2
AX-301115785	7	5412652	5412719	4,00E-14	78,8
AX-299634480	7	6047360	6047426	6,00E-11	68,9
AX-299747048	7	6381940	6382000	4,32E-13	76,1
AX-297171421	7	6383316	6383386	3,09E-21	103,0
AX-299747154	7	6455960	6456030	1,16E-26	121,0
AX-297393166	7	6606718	6606788	3,77E-20	98,7
AX-297143682	7	6641878	6641948	6,00E-24	112,0
AX-297393564	7	7064074	7064146	2,38E-16	86
AX-297394303	7	7852535	7852605	1,02E-14	80,6
AX-299525902	7	8103031	8103097	1,24E-13	77,9
AX-299781048	7	8278821	8278891	5,00E-12	71,6
AX-300445696	7	8411536	8411599	4,62E-19	95,1
AX-297395464	7	8669685	8669754	1,08E-20	101,0
AX-299523700	7	9397790	9397856	6,38E-11	68,9
AX-299523700	7	9519963	9520029	7,77E-10	64,4
AX-297397189	7	9826610	9826680	3,79E-20	98,7
AX-301116438	7	9828576	9828507	3,79E-20	99,6
AX-299526308	7	10567620	10567690	7,27E-23	107,0
AX-299526353	7	10688342	10688405	1,23E-13	77
AX-299526353	7	10703785	10703854	2,90E-15	83,3
AX-299526353	7	10721868	10721938	2,23E-10	67,1
AX-299526353	7	10764941	10765011	5,24E-12	71,6
AX-299634480	7	11846358	11846424	6,00E-11	68,9
AX-299526791	7	12535401	12535465	1,50E-12	74,3
AX-297401655	7	12600838	12600909	1,32E-19	96,9
AX-297402361	7	13368596	13368666	1,96E-17	89,7
AX-299527049	7	14242486	14242554	2,39E-16	86,9
AX-299527491	7	16084155	16084219	3,55E-14	78,8
AX-299527696	7	17245466	17245528	1,60E-18	93,3
AX-297173619	7	17252273	17252345	1,02E-14	81,5
AX-297184644	7	17422423	17422493	3,79E-20	98,7
AX-297173795	7	17935029	17934959	1,61E-18	94,2

Query	Chr. Goiaba	Subject start	Subject end	e-value	bit score
AX-297173754	7	18155056	18154986	5,97E-24	112,0
AX-297214327	7	21013560	21013629	2,90E-15	83,3
AX-299485227	7	21612395	21612325	4,30E-13	76,1
AX-299530106	7	21612395	21612325	4,32E-13	76,1
AX-299764100	7	21941578	21941648	8,35E-16	85,1
AX-297158180	7	22799768	22799698	1,02E-14	80,6
AX-299867036	7	23756417	23756483	1,08E-20	100,0
AX-299868240	7	24295644	24295709	8,30E-16	85,1
AX-298859159	7	24520171	24520101	7,27E-23	107,0
AX-299520997	7	24637137	24637072	1,02E-14	80,6
AX-300517987	7	24658759	24658692	1,84E-11	69,8
AX-297448016	7	24804035	24803967	4,32E-13	75,2
AX-297428427	7	24826121	24826048	6,85E-17	88,7
AX-299782522	7	24833230	24833161	1,50E-12	74,3
AX-299782515	7	24836655	24836587	1,24E-13	77,9
AX-299781760	7	25742462	25742394	1,24E-13	77,9
AX-299861456	7	25742462	25742394	1,24E-13	77,9
AX-301120378	7	25876753	25876823	7,27E-23	107,0
AX-299781595	7	25924034	25923964	3,79E-20	98,7
AX-299780126	7	27264645	27264715	6,00E-24	112,0
AX-299501555	7	27563599	27563669	7,27E-23	107,0
AX-299634480	7	27747914	27747980	3,00E-15	82,4
AX-298851518	7	27850062	27849992	1,96E-17	89,7
AX-299537281	7	28279901	28279971	3,79E-20	98,7
AX-299776854	7	29819123	29819058	2,23E-10	67,1
AX-297189412	7	30509740	30509806	2,91E-15	82,4
AX-299774580	7	31841129	31841063	4,30E-13	75,2
AX-168334204	7	31982813	31982743	1,95E-17	89,7
AX-297175145	7	32030654	32030584	1,61E-18	94,2
AX-299772888	7	32207332	32207403	1,32E-19	96,9
AX-299528979	7	32855313	32855378	1,00E-14	80,6
AX-297182544	7	33290443	33290373	4,30E-13	76,1
AX-299528853	7	33290443	33290373	4,30E-13	76,1
AX-299528389	7	34179101	34179032	1,54E-12	74,3
AX-299528651	7	34483270	34483340	1,96E-17	89,7
AX-299528578	7	35090752	35090685	1,83E-11	70,7
AX-300488059	7	35094572	35094503	7,31E-23	108
AX-297650250	7	35592560	35592630	3,79E-20	98,7
AX-299562357	8	1422796	1422865	6,00E-18	92,4
AX-297611820	8	2383543	2383474	1,02E-14	81,5
AX-297611820	8	2493709	2493640	1,02E-14	81,5
AX-297207998	8	2982121	2982055	2,39E-16	86,9
AX-299559244	8	3591852	3591922	3,11E-21	103,0
AX-297205291	8	4324621	4324691	3,79E-20	98,7
AX-299952116	8	4932592	4932657	5,27E-12	71,6

Query	Chr. Goiaba	Subject start	Subject end	e-value	bit score
AX-297186207	8	5265594	5265525	6,81E-17	88,7
AX-301013877	8	5990860	5990927	3,55E-14	79,7
AX-300894015	8	6550701	6550771	7,31E-23	107,0
AX-297167241	8	6979599	6979524	6,38E-11	68
AX-297601255	8	7705968	7706033	8,35E-16	85,1
AX-299743623	8	7907299	7907230	4,62E-19	96,0
AX-299634480	8	8442782	8442848	1,00E-13	77,9
AX-297206341	8	8480299	8480368	7,82E-10	65,3
AX-297206408	8	8669235	8669305	3,77E-20	98,7
AX-299634480	8	9428872	9428938	2,00E-12	73,4
AX-297605525	8	11151452	11151522	1,61E-18	94,2
AX-300906477	8	11178926	11178991	2,23E-10	67,1
AX-299561115	8	11293612	11293670	1,02E-14	81,5
AX-299960074	8	11478257	11478327	8,35E-16	85,1
AX-299354405	8	11479530	11479600	6,00E-24	112
AX-299561329	8	12189004	12189074	7,00E-23	107,0
AX-297207429	8	13012992	13013062	7,31E-23	107
AX-299963003	8	13624994	13625064	1,00E-13	77,0
AX-299257305	8	14478942	14479012	3,11E-21	103
AX-297207806	8	14484654	14484723	6,85E-17	87,8
AX-299964456	8	14799054	14799121	7,77E-10	64,4
AX-297611006	8	15179947	15180017	1,61E-18	94,2
AX-299565825	8	18374057	18373990	2,24E-10	66,2
AX-299558202	8	25738957	25738887	3,79E-20	98,7
AX-299576828	8	26049858	26049925	7,82E-10	65,3
AX-299634480	8	26599290	26599356	1,00E-13	77,9
AX-299558011	8	27329642	27329712	3,00E-21	103,0
AX-297428951	8	27563599	27563669	7,27E-23	107,0
AX-301145674	8	29461411	29461481	3,79E-20	98,7
AX-297372065	8	30311424	30311489	8,00E-10	65,3
AX-299634480	8	31219541	31219607	3,00E-15	82,4
AX-297582803	8	31243402	31243328	6,42E-11	68,9
AX-301106849	8	32955981	32956050	2,10E-23	110
AX-299728960	9	10017	9947	1,96E-17	89,7
AX-297384526	9	313209	313140	3,79E-20	99,6
AX-299523700	9	390915	390847	2,38E-16	86
AX-299523700	9	408344	408276	8,86E-22	104
AX-297169605	9	468421	468351	3,79E-20	98,7
AX-297293559	9	784344	784274	1,61E-18	94,2
AX-299736669	9	1202676	1202606	1,96E-17	89,7
AX-299523044	9	2670910	2670990	1,00E-14	81,5
AX-299474834	9	3130634	3130568	1,44E-12	74,3
AX-299522873	9	3541158	3541094	1,32E-19	96,9
AX-300410791	9	3771987	3772055	3,00E-15	82,4
AX-297168632	9	4259112	4259042	1,61E-18	94,2

Query	Chr. Goiaba	Subject start	Subject end	e-value	bit score
AX-299454645	9	4577551	4577481	3,11E-21	103,0
AX-297168533	9	4579104	4579034	8,35E-16	85,1
AX-299522555	9	4734101	4734035	1,24E-13	77,9
AX-299522466	9	5165665	5165595	5,27E-12	72,5
AX-297168382	9	5460345	5460279	5,62E-18	91,5
AX-299522184	9	7039137	7039068	2,55E-22	105
AX-297375642	9	7202086	7202016	3,09E-21	103
AX-297168000	9	7657189	7657119	3,11E-21	103,0
AX-299522004	9	7897652	7897588	6,85E-17	87
AX-299866289	9	8206538	8206477	6,38E-11	68,9
AX-297444780	9	12511004	12511066	8,35E-16	84,2
AX-299439264	9	13257686	13257630	1,24E-13	77,9
AX-297167241	9	13794400	13794330	4,59E-19	95,1
AX-299550079	9	17106482	17106549	7,82E-10	65,3
AX-299521194	9	17659257	17659191	4,59E-19	96,0
AX-299576828	9	17770086	17770016	6,00E-24	112,0
AX-299703043	9	17790238	17790168	5,27E-12	71,6
AX-297554322	9	18345223	18345166	2,23E-10	66,2
AX-299521130	9	19170774	19170706	6,85E-17	87,8
AX-168334175	9	20318136	20318206	2,00E-10	67,1
AX-297479939	9	21776651	21776593	1,02E-14	81,5
AX-300911774	9	21779512	21779452	5,24E-12	71,6
AX-297165070	9	21930041	21930111	1,61E-18	94,2
AX-297165070	9	22021547	22021617	3,79E-20	98,7
AX-297165070	9	22057327	22057397	3,79E-20	98,7
AX-298697864	9	23439241	23439305	1,83E-11	69,8
AX-299519769	9	25337760	25337825	2,00E-11	69,8
AX-299519804	9	25370971	25371041	4,32E-13	76,1
AX-298705862	9	25765509	25765579	8,35E-16	85,1
AX-297165975	9	26290598	26290668	3,77E-20	98,7
AX-299520120	9	26622196	26622266	7,00E-23	107,0
AX-297166189	9	27228904	27228970	4,62E-19	96,0
AX-168367414	9	27235328	27235396	5,27E-12	72,5
AX-297166264	9	27537020	27537090	1,96E-17	89,7
AX-298719002	9	30249624	30249684	2,23E-10	67,1
AX-299520932	9	30963732	30963799	2,55E-22	106,0
AX-299520932	9	30977763	30977830	1,61E-18	93,3
AX-299520991	9	31397196	31397257	2,39E-16	86,9
AX-299663375	10	235506	235573	1,84E-11	70,7
AX-299577360	10	938604	938672	4,32E-13	75,2
AX-299496408	10	1179220	1179288	4,62E-19	95,1
AX-297223713	10	1204924	1204994	7,31E-23	107,0
AX-299652225	10	1907058	1907127	1,01E-14	80,6
AX-299578897	10	1965416	1965486	4,32E-13	76,1
AX-300930104	10	2145457	2145521	1,83E-11	70,7

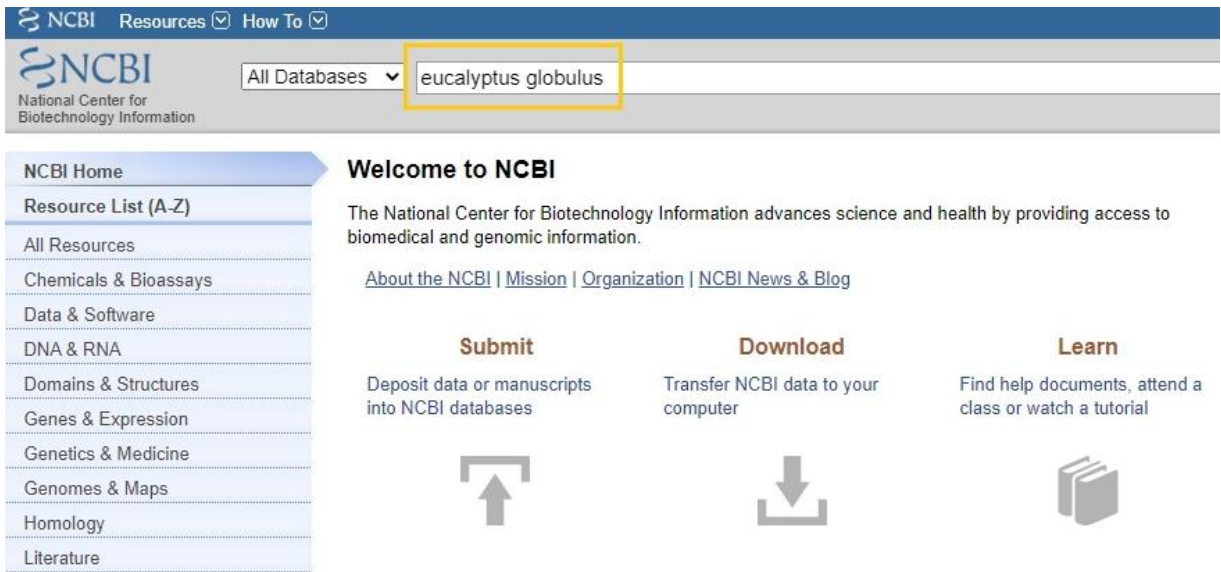
Query	Chr. Goiaba	Subject start	Subject end	e-value	bit score
AX-300930104	10	2156505	2156571	1,50E-12	73,4
AX-297226592	10	2918183	2918251	5,24E-12	72,5
AX-301130671	10	3336142	3336074	3,77E-20	99,6
AX-297142322	10	3541098	3541168	3,79E-20	98,7
AX-299582563	10	3700047	3700117	3,11E-21	103,0
AX-297228464	10	3710324	3710393	4,00E-14	78,8
AX-299537949	10	3887635	3887705	3,11E-21	103,0
AX-297142476	10	4128912	4128982	1,96E-17	89,7
AX-297229393	10	4324236	4324306	1,61E-18	94,2
AX-299584047	10	4644072	4644133	5,62E-18	91,5
AX-299584480	10	4798864	4798931	3,09E-21	102,0
AX-299585213	10	5441670	5441731	4,00E-13	75,2
AX-297142807	10	5764602	5764670	3,79E-20	99,6
AX-301098440	10	5844184	5844115	3,79E-20	99,6
AX-297188282	10	5873157	5873226	2,38E-16	86,9
AX-299587612	10	7178040	7178109	2,10E-23	110
AX-300023459	10	8452405	8452475	2,24E-10	67,1
AX-299498037	10	8465410	8465478	4,62E-19	95,1
AX-299634480	10	8976586	8976656	4,00E-13	76,1
AX-299486473	10	9978735	9978797	1,61E-18	93,3
AX-299498431	10	11230968	11231038	8,35E-16	85,1
AX-299499481	10	13321940	13322010	7,00E-23	107,0
AX-299561570	10	15791045	15791115	3,00E-21	103,0
AX-297207661	10	16448223	16448153	1,96E-17	89,7
AX-297152190	10	16586902	16586963	1,50E-12	73,4
AX-299634480	10	18526910	18526980	1,00E-14	80,6
AX-299634480	10	20959259	20959325	1,00E-13	77,9
AX-297145032	10	21432785	21432855	3,11E-21	103,0
AX-297189382	10	21596860	21596927	1,29E-19	97,8
AX-297145100	10	21896352	21896422	5,00E-12	71,6
AX-297242851	10	22032584	22032656	1,24E-13	77,9
AX-297216244	10	24181012	24181082	2,23E-10	67,1
AX-299499629	10	24390025	24390095	3,11E-21	103,0
AX-299499642	10	24609908	24609979	6,85E-17	87,8
AX-297145624	10	24620870	24620940	7,31E-23	107,0
AX-299713854	10	24741679	24741748	1,32E-19	96,9
AX-297186207	10	25840055	25839986	8,35E-16	84,2
AX-299441049	10	26438659	26438729	1,95E-17	89,7
AX-297249958	10	28142693	28142763	4,32E-13	76,1
AX-299500450	10	29339418	29339488	8,35E-16	85,1
AX-299500456	10	29357472	29357542	1,60E-18	94,2
AX-299605734	10	29507740	29507805	1,96E-17	90,6
AX-297251867	10	29638018	29638081	4,62E-19	95,1
AX-301144169	10	30190402	30190469	1,32E-19	97,8
AX-297146846	10	30429062	30429126	2,90E-15	83,3

Query	Chr. Goiaba	Subject start	Subject end	e-value	bit score
AX-297253122	10	30709462	30709528	2,23E-10	66,2
AX-301101077	10	31291723	31291780	3,55E-14	79,7
AX-297254040	10	31438577	31438642	4,32E-13	76,1
AX-297147408	10	32762118	32762048	3,79E-20	98,7
AX-297318705	10	33001267	33001198	2,90E-15	83,3
AX-297318705	10	33097474	33097544	2,38E-16	86,0
AX-297318705	10	33103786	33103856	4,59E-19	95,1
AX-299501834	10	33892726	33892796	3,00E-21	103,0
AX-299634480	10	35355504	35355570	4,00E-13	76,1
AX-297258069	10	35452705	35452774	5,62E-18	92,4
AX-297258069	10	35486276	35486344	3,79E-20	99,6
AX-297148418	10	35720210	35720280	7,31E-23	107,0
AX-299502490	10	35830287	35830357	8,30E-16	85,1
AX-299502530	10	36028669	36028739	1,00E-14	80,6
AX-297375642	10	36201226	36201295	7,77E-10	65,3
AX-299613041	10	36225065	36225135	3,79E-20	98,7
AX-297149067	10	37555967	37556036	5,62E-18	92,4
AX-299503161	10	37781897	37781961	2,00E-11	69,8
AX-299503166	10	37775919	37775989	3,00E-21	103,0
AX-299565937	11	72925	72856	1,32E-19	96,9
AX-300930104	11	1165377	1165447	3,11E-21	103
AX-299563232	11	1817237	1817303	1,24E-13	77,9
AX-297616735	11	2001662	2001732	7,31E-23	107,0
AX-301139155	11	2006716	2006770	4,00E-13	75,2
AX-299563484	11	2762856	2762926	3,79E-20	98,7
AX-299563711	11	3388286	3388355	2,10E-23	110
AX-299280106	11	4391622	4391692	3,09E-21	103
AX-300939640	11	4462150	4462218	5,27E-12	72,5
AX-299563910	11	4469904	4469974	7,27E-23	107
AX-297620324	11	4700693	4700761	5,00E-12	72,5
AX-297210042	11	5787483	5787546	2,38E-16	86
AX-299564314	11	6219215	6219285	2,24E-10	67,1
AX-299624070	11	7647460	7647528	4,00E-13	76,1
AX-299564683	11	7702707	7702777	1,02E-14	80,6
AX-297210794	11	8377402	8377471	2,08E-23	110,0
AX-299498223	11	9259142	9259074	4,62E-19	95,1
AX-298963244	11	9291737	9291808	1,32E-19	96,9
AX-299980754	11	9805780	9805850	1,41E-25	116,0
AX-297517780	11	9892954	9893018	1,02E-14	81,5
AX-299980892	11	9924102	9924172	1,61E-18	94,2
AX-299573938	11	16765475	16765410	1,02E-14	81,5
AX-299873143	11	17579712	17579774	1,61E-18	93,3
AX-299531762	11	18780005	18779941	1,32E-19	97,8
AX-299543139	11	21127806	21127736	7,31E-23	107,0
AX-299521567	11	21407001	21407069	8,00E-10	65,3

Query	Chr. Goiaba	Subject start	Subject end	e-value	bit score
AX-299309218	11	22251596	22251524	7,82E-10	64,4
AX-297633924	11	22536337	22536407	1,61E-18	94,2
AX-297220625	11	22602992	22602923	1,32E-19	97,8
AX-299566732	11	23692604	23692669	1,96E-17	89,7
AX-301139956	11	24152000	24152069	1,32E-19	96,9
AX-299600857	11	25217376	25217445	6,85E-17	87,8
AX-297212923	11	25257531	25257600	3,55E-14	79,7
AX-297213111	11	25863250	25863312	4,32E-13	75,2
AX-299567408	11	26578658	26578727	2,91E-15	83,3
AX-297213371	11	27126750	27126820	3,09E-21	103
AX-299567494	11	27181824	27181894	1,96E-17	89,7
AX-299325910	11	27595740	27595810	1,61E-18	94,2
AX-300985861	11	27863105	27863170	8,35E-16	85,1
AX-300990193	11	29579471	29579540	2,55E-22	105,0
AX-297213963	11	29707349	29707419	6,00E-24	112,0
AX-297213999	11	29928889	29928959	1,96E-17	89,7
AX-299997522	11	30086667	30086737	1,40E-25	116,0
AX-297643722	11	30193976	30194046	7,31E-23	107
AX-297214087	11	30197212	30197282	7,31E-23	107,0
AX-299450053	11	31060470	31060540	1,61E-18	94,2
AX-297214327	11	31330677	31330744	1,32E-19	96,9
AX-297647230	11	33254732	33254802	1,96E-17	89,7
AX-297214943	11	34764283	34764353	3,11E-21	103
AX-299569137	11	35138100	35138170	8,35E-16	85,1
AX-299569177	11	35420443	35420512	2,54E-22	105,0
AX-297254513	11	36094375	36094445	6,00E-24	112,0
AX-301144169	11	36252753	36252818	1,61E-18	94,2

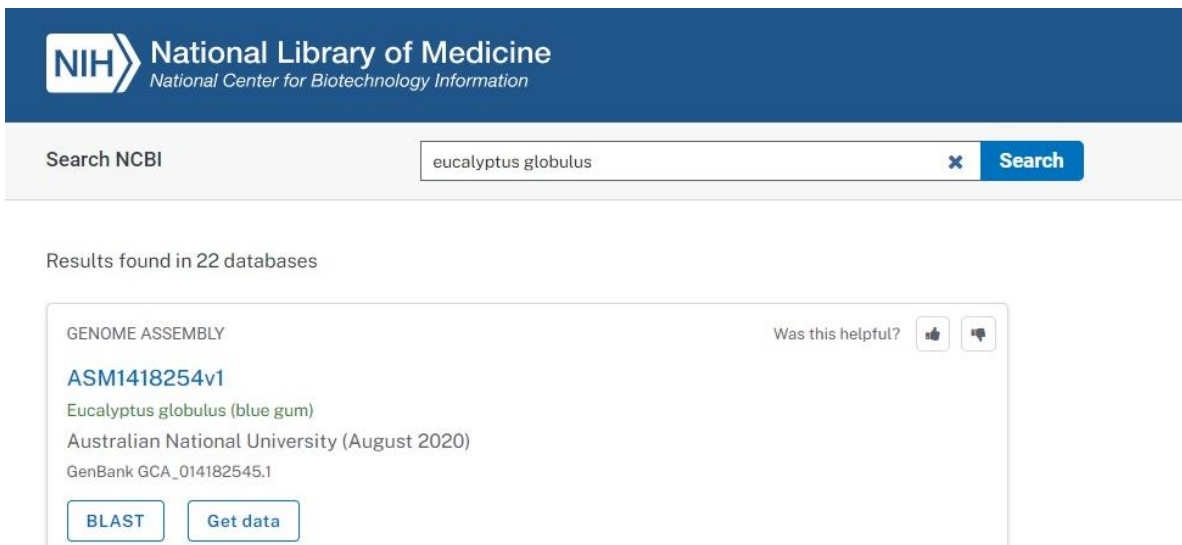
ANEXO II - BLAST DE SEQUÊNCIAS DE MICROSSATÉLITES DE *P. guajava* EM *Eucalyptus globulus*

Fig 1 Passo 1: Espécie utilizada inserida na busca em todos os bancos de dados do site.



The screenshot shows the NCBI homepage. At the top, there is a navigation bar with 'NCBI Resources' and 'How To'. Below this is the NCBI logo and a search bar. The search bar contains the text 'eucalyptus globulus' and is highlighted with a yellow box. To the left of the search bar is a dropdown menu labeled 'All Databases'. Below the search bar is a 'Welcome to NCBI' section with a navigation menu on the left and three main action buttons: 'Submit', 'Download', and 'Learn'. The 'Submit' button is highlighted with a yellow box.

Fig 2 Passo 2: Seleção da espécie de interesse.



The screenshot shows the NCBI search results page. At the top, there is the NIH logo and the text 'National Library of Medicine National Center for Biotechnology Information'. Below this is a search bar with the text 'eucalyptus globulus' and a 'Search' button. Below the search bar is a section titled 'Results found in 22 databases'. The first result is highlighted with a yellow box and contains the following information: 'GENOME ASSEMBLY', 'ASM1418254v1', 'Eucalyptus globulus (blue gum)', 'Australian National University (August 2020)', and 'GenBank GCA_014182545.1'. There are two buttons below the result: 'BLAST' and 'Get data'. To the right of the result is a 'Was this helpful?' section with a thumbs up and thumbs down icon.

Fig 3 Passo 3: Adição do “accession number” e utilização da opção MEGABLAST.

BLAST >> blastn suite

Eucalyptus globulus GenBank assembly GCA_014182545.1 Nucleotide BLAST

blastn | **tblastn** | tblastx

BLASTN programs search GenBank assembly GCA_014182545.1 databases using a nucleotide query. [more](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) Clear Query subrange [?](#)

AJ639775 From To

Or, upload file Escolher arquivo Nenhum arquivo selecionado [?](#)

Job Title
 Enter a descriptive title for your BLAST search [?](#)

Database ASM1418254v1 GenBank assembly [GCA_014182545.1] [?](#)

Program Selection

Optimize for

- Highly similar sequences (megablast)
- More dissimilar sequences (discontiguous megablast)
- Somewhat similar sequences (blastn)

Choose a BLAST algorithm [?](#)

Fig 4 Passo 4: Resumo das informações dos hits encontrados, como query ID, molecular type e length of the query

BLAST

[< Edit Search](#) [Save Search](#) [Search Summary](#) ▼

Job Title	AJ639775:Psidium guajava microsatellite DNA,...
RID	ZHBK1YW401R <i>Search expires on 01-10 02:11 am</i> Download All ▼
Program	BLASTN ? Citation ▼
Database	genomic/34317/GCA_014182545.1 See details ▼
Query ID	AJ639775.1
Description	Psidium guajava microsatellite DNA, clone mPgCIR1
Molecule type	nucleic acid
Query Length	367
Other reports	Distance tree of results MSA viewer ?

Fig 7 Resultado do alinhamento dos nucleotídeos do marcador mPgCIR005 contra genoma de *E. globulus*.

Eucalyptus globulus isolate CBG9404798.5 chromosome 1, whole genome shotgun sequence

Sequence ID: [CM024717.1](#) Length: 43159921 Number of Matches: 9

Range 1: 27813063 to 27813327 [GenBank](#) [Graphics](#)

[▼ Next Match](#) [▲ Previous Match](#)

Score	Expect	Identities	Gaps	Strand
259 bits(140)	2e-67	225/266(85%)	5/266(1%)	Plus/Minus
Query 4	CCATGATGTCTAAGTAGCCGGAATCGAAGGGA-GGTGCCATAGAGATCATCCGCAAGCTG			62
Sbjct 27813327	CCATGTTGTCTGAGTAACCGGAATCCAAGGAAACTG-CATGGAGATCGTTCAAAAGCTG			27813269
Query 63	AGCATCTCCGGTGTGT--TA-ATGGAAGTGAAGCATTATTTTATCTCCGTCTCAAAATG			119
Sbjct 27813268	AGCATATCCAGTGTGTGATAGATGGAACAAAGCATTATTTTGTCTCCATCTCAAAATG			27813209
Query 120	GTATCCTAGCCCTAGCCTTTGAACCACATCAATGAACTCAAGCTTGGCCACCTGGTTCAT			179
Sbjct 27813208	GTATCCTAGTCCGAGTCTTTGAACCGTGTCAATGAACTCGAGCTTGGCCACCTGGTTCAT			27813149
Query 180	CTCCTTGCCAAATAGACCCCTGACTTCTTCTTCAACCTTTGGACTTGTCTGGTATAGTC			239
Sbjct 27813148	CTCCCTGTTGAATAGACCCCTGATTTCTTCTTCAACCTTTGGACTTGTCTGGCATATTT			27813089
Query 240	CTCCTCCTATGCACGTGAAAAATTCA		265	
Sbjct 27813088	ATCCTCCTAGGCACGTAAAAACATTCA		27813063	

Fig 8 Resultado do alinhamento dos nucleotídeos do marcador mPgCIR007 contra genoma de *E. globulus*.

Eucalyptus globulus isolate CBG9404798.5 chromosome 10, whole genome shotgun sequence

Sequence ID: [CM024726.1](#) Length: 37343655 Number of Matches: 1

Range 1: 32043161 to 32043472 [GenBank](#) [Graphics](#)

[▼ Next Match](#) [▲ Previous Match](#)

Score	Expect	Identities	Gaps	Strand
435 bits(235)	2e-120	285/312(91%)	0/312(0%)	Plus/Minus
Query 6	GATTCTCACCGTATCTGAGGGGGTGATNCGTTCTCTGNCTGAAATTGCTGTTGAAGTCGN			65
Sbjct 32043472	GATTCTCACCGTATTTGTGGGGGTGATTCATTCTCTGCTGAGGTTGCTGTTGAAGTCAT			32043413
Query 66	AGAGGGAATATTTGAAACCCGNGAGCTGATGCTCAAGTTTCACCAGGAGTTTAGAGAGAG			125
Sbjct 32043412	AGAGGGAATATTTGAAACCTGTGAGCTGATGCTCAAGTTTGTCTAGGAGTTTAGAGAGAG			32043353
Query 126	CATGGTTGTGCAACTTTGCTAGTGTGAAGCTTCTTCCAAGCACCTCCCTCTCTTCTCTG			185
Sbjct 32043352	CATGATTGTGCAAAATTTGCTAGTGTGAAGCTTCTTCCAAGCACCTCCCTCCCTGCTCTG			32043293
Query 186	CTTTGATTATCCTCATTCTGGAAGACACCCCAATGGAGGTAGGTTGATGAACCCAAATT			245
Sbjct 32043292	CTTTGATTATCCTGATTCTGGAAGACACCCCAATGGAGGTAGGTTGATGAACCCGAATT			32043233
Query 246	TCCTTCCCCCATTCCATATATCTCTGCCAAATTTCAATAAAACTGCTCTCTAGTCC			305
Sbjct 32043232	TTCTTCCCCCTATCCATATATCTCTGCCAAATTTCAACAAACTGTTCTCTAGTCT			32043173
Query 306	CACcAcAcAcAc	317		
Sbjct 32043172	CACCACACACAC	32043161		

Fig 11 Resultado do alinhamento dos nucleotídeos do marcador mPgCIR018 contra genoma de *E. globulus*.

Eucalyptus globulus isolate CBG9404798.5 chromosome 3, whole genome shotgun sequence
 Sequence ID: [CM024719.1](#) Length: 67249429 Number of Matches: 1

Range 1: 1035784 to 1036051 [GenBank](#) [Graphics](#) [▼ Next Match](#) [▲ Previous Match](#)

Score	Expect	Identities	Gaps	Strand
396 bits(214)	9e-109	249/268(93%)	0/268(0%)	Plus/Plus
Query 89	gagagagagagagagaACCTTAATATTCTGTGAGTCTGTCAAAGCATGGCAAGCCTATC			148
Sbjct 1035784	GAGAGAGAGAGAGAGAACCTTAATATTCTGCGAATCGGTCAAAGCATGGTAAGCCTATC			1035843
Query 149	AATACCCATCCCCAGCCTCCAGTTGGAGGCAAACCATACTCAAGTGCAGAACAGAACGT			208
Sbjct 1035844	AATACCCATCCCCAGCCACCAGTTGGAGGCAAACCATACTCAAGTGCAGAACAGAACGT			1035903
Query 209	TTCATCCAAAGCCATGGCTTCATCATCTCCCGATTGCCTATCCTGAATACGATGCAGATT			268
Sbjct 1035904	TTCATCCAAAGCCATGGCTTCATCATCACCCGATTGCCTGTCTGAATACCATGCAGATT			1035963
Query 269	CAATTCAGAATTGCTTCCCGAGAGTATTACNAATATTGCACCCNGCAACGATNATACCT			328
Sbjct 1035964	CAATTCAGAATTGCTTCCCAAGAGTATAACAAAAAATTGCACCCAGCAATGATAATACCT			1036023
Query 329	TGAGTTGTTTCAGCAAATCGCTCCCGCTG			356
Sbjct 1036024	TGAGTTGTTTCGGCAAATCGCTCCCGCTG			1036051

Fig 12 Resultado do alinhamento dos nucleotídeos do marcador mPgCIR020 contra genoma de *E. globulus*.

Eucalyptus globulus isolate CBG9404798.5 chromosome 11, whole genome shotgun sequence
 Sequence ID: [CM024727.1](#) Length: 42041974 Number of Matches: 1

Range 1: 22321013 to 22321226 [GenBank](#) [Graphics](#) [▼ Next Match](#) [▲ Previous Match](#)

Score	Expect	Identities	Gaps	Strand
289 bits(156)	2e-76	198/218(91%)	4/218(1%)	Plus/Plus
Query 37	CCCTTATCGTGGTAGAGGTTCTATATGGTATATAGTGTTTAAGTATACCACACGCTGA			96
Sbjct 22321013	CCATTATTGTGGTAGAGGTTCTATA-GG-GTATAGTG-TTTATGCATACCAATGCTGA			22321069
Query 97	AACAAAATGTGAAAATGCTTAAACAGTGAAGTTTGCATATCTTTCATTGGGTCGCCAG			156
Sbjct 22321070	AACAAAATGTG-AAATGCTTAAACAGTGAAGTTTGCATTTCTTTCGTTGGGTCGCCAG			22321128
Query 157	AGTATCTTGAGGATTCTGAAATTTTATCAAATCGTTTTTCAGGTGCATctctctctctctc			216
Sbjct 22321129	AGTATCTTGAGGATTCTGAAATTTTATCAAACCGTTTTTCAGGTGCATCTCTCTCTCTC			22321188
Query 217	tctctctctctctctcacacacacacacacacacacac			254
Sbjct 22321189	TCTCTCTCTCTCTCTCTCTCTCACACGCACACGCAC			22321226

Fig 13 Resultado do alinhamento dos nucleotídeos do marcador mPgCIR021 contra genoma de *E. globulus*.

Eucalyptus globulus isolate CBG9404798.5 chromosome 8, whole genome shotgun sequence

Sequence ID: [CM024724.1](#) Length: 64768990 Number of Matches: 1

Range 1: 4350645 to 4350865 [GenBank](#) [Graphics](#)

[▼ Next Match](#) [▲ Previous Match](#)

Score	Expect	Identities	Gaps	Strand
141 bits(76)	3e-32	183/229(80%)	29/229(12%)	Plus/Plus
Query 1	ACCCACAAAAGTGGTGGAGGAATGAGATTTTTTCTTTCTTCAA-TTGCCCTTCTAAGTAT	59		
Sbjct 4350645	ACCCAGAAAAGTGCTAGGGATAAGA-TTCTCTTTCCTTCAAATG-CC--CTATGT-C	4350699		
Query 60	AACAG-AAGGAAAAGaaaaaaaaTTTCCTCCAAAATGGAAGCCAAACACAAATgaga--ga	116		
Sbjct 4350700	AGCAGAAAAGAAAA-AAAAAAA-TTCTTTCAAATGGAAGCCAAACACGAATGAGACAGA	4350757		
Query 117	gagag--agtttag-ga-ag--a-agag--agagagag---a-g--agagagagagagag	161		
Sbjct 4350758	GAGAGACAG-ACAGAGAGAGACAGAGAGACAGAGAGAGTTTAGGAAAAGAGAGAGAGAGAG	4350816		
Query 162	ag-g-gagagagagagagagTTTAGGAAGGTTTGTAGCTCAAAGAGGA	208		
Sbjct 4350817	AGAGAGAGAGAGAGAGAGAGTTTAGGAAGGTTTGTAGCTCAAAGAGGA	4350865		

Fig 14 Resultado do alinhamento dos nucleotídeos do marcador mPgCIR026 contra genoma de *E. globulus*.

Eucalyptus globulus isolate CBG9404798.5 chromosome 4, whole genome shotgun sequence

Sequence ID: [CM024720.1](#) Length: 36876699 Number of Matches: 1

Range 1: 36157376 to 36157715 [GenBank](#) [Graphics](#)

[▼ Next Match](#) [▲ Previous Match](#)

Score	Expect	Identities	Gaps	Strand
326 bits(176)	1e-87	312/372(84%)	32/372(8%)	Plus/Plus
Query 7	AGTAAAAATTCAGACAATTTCTTATACCATCAAGGACATCCCAAGCAGGGAAGCACTCTC	66		
Sbjct 36157376	AGTAAAAATTCAGACAATTCCTTATACCATCAATGACATTTCAAGCAGGGAAGCACTCTT	36157435		
Query 67	TGCAGGAAAGGAGATATACAGAAAGAGCATGCCTACCAAGGAGATAGCAAGTGCATATGT	126		
Sbjct 36157436	CACAAGAAAGGAGATATACAGAAAGAGCATGCCAAGCAAGGAGATAGCA--TG---A---	36157487		
Query 127	GCACGAATTGATTTGTGACTGGCGTGAGCAAgtgcgtatgtgcacaggcatgtatgtgca	186		
Sbjct 36157488	GCACAACCTGATTTGTGA-T---T--GC---TGC--A----CACAGGCACATATGTGAA	36157531		
Query 187	tgc atgtaatggtgtgcatatgtgtgcgtatgtgcgtgtgagagagagagagagagagag	246		
Sbjct 36157532	TGCATGTAAGGGGGTGCATGTGTGTCATATGTG--TGTGAGAGAGA-AGA-AGATAGAG	36157587		
Query 247	agagagagagagagTATACCTCCAAAGTCTCCATTTTCGTCCTTGGTAGGACAGAATTGA	306		
Sbjct 36157588	AGAGA-A-A-ATA-CATACCTTCAAAGTCTCCATTTCTTCTTGGTGGGACAGAATTTGA	36157643		
Query 307	TGAGATTTTCAACCTGATCAATATCCAAAGCTGAAGAATCCAAAGCCAGAACTGCATTCT	366		
Sbjct 36157644	TGAGATTTTCAACCTGATCAATGTCCAAAGCTGAAGAATCCAAAGCCAGAACTGTATTCT	36157703		
Query 367	GGGAAAAAATGG	378		
Sbjct 36157704	GGGAAAAAATGG	36157715		