



**UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE AGRONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E
MELHORAMENTO DE PLANTAS**

**VALIDAÇÃO DE MARCADORES SSR-Seq PARA
Eugenia klotzschiana O. BERG. (MYRTACEAE)**

Orientadora:
Prof.^a Dra. THANNYA NASCIMENTO SOARES

SYLLUANA RIBEIRO CORRÊA



UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE AGRONOMIA

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a [Lei 9.610/98](#), o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

Dissertação Tese Outro*: _____

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

SYLLUANA RIBEIRO CORRÊA

3. Título do trabalho

VALIDAÇÃO DE MARCADORES SSR-Seq PARA *Eugenia klotzschiana* O. BERG. (MYRTACEAE)

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento SIM NÃO¹

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:

a) consulta ao(à) autor(a) e ao(à) orientador(a);

b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.

O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.



Documento assinado eletronicamente por **Sylluana Ribeiro Corrêa, Discente**, em 25/07/2024, às 15:13, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Thannya Nascimento Soares, Professora do Magistério Superior**, em 14/03/2025, às 10:25, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **4643232** e o código CRC **A3B234A8**.

SYLLUANA RIBEIRO CORRÊA

**VALIDAÇÃO DE MARCADORES SSR-Seq
PARA *Eugenia klotzschiana* O. BERG.
(MYRTACEAE)**

Dissertação apresentada ao Programa de Pós-Graduação em Genética e Melhoramento de Plantas, da Escola de Agronomia, da Universidade Federal de Goiás, como requisito à obtenção do título de Mestra em Genética e Melhoramento de Plantas.

Área de Concentração: Genética e Melhoramento de Plantas

Linha de Pesquisa: Conservação e Melhoramento de Espécies Vegetais Nativas do Cerrado

Orientadora:

Prof.^a Dr.^a Thannya Nascimento Soares

Coorientadora:

Prof.^a Dr.^a Mariana Pires de Campos Telles

Goiânia, GO – Brasil

2024

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Corrêa, Sylluana Ribeiro
VALIDAÇÃO DE MARCADORES SSR-Seq PARA *Eugenia klotzschiana* O.BERG. (MYRTACEAE) [manuscrito] / Sylluana Ribeiro Corrêa. - 2024.
XLIV, 44 f.

Orientador: Prof. Thannya Nascimento Soares; co-orientador Mariana Pires de Campos Telles.
Dissertação (Mestrado) - Universidade Federal de Goiás, Escola de Agronomia (EA), Programa de Pós-graduação em Genética e Melhoramento de Plantas, Goiânia, 2024.

Apêndice.
Inclui gráfico, tabelas.

1. diversidade genética. 2. genotipagem por sequenciamento. 3. pera-do-cerrado. 4. recursos genéticos. I. Soares, Thannya Nascimento, orient. II. Título.

CDU 575



UNIVERSIDADE FEDERAL DE GOIÁS

ESCOLA DE AGRONOMIA

ATA DE DEFESA DE DISSERTAÇÃO

Ata nº 124 da sessão de Defesa de Dissertação de SYLLUANA RIBEIRO CORRÊA, que confere o título de Mestra em Genética e Melhoramento de Plantas, na área de concentração em Genética e Melhoramento de Plantas.

Aos dezoito dias do mês de julho de dois mil e vinte e quatro, a partir das 14h:00m, em formato híbrido, realizou-se a sessão pública de Defesa de Dissertação intitulada “VALIDAÇÃO DE MARCADORES SSR-Seq PARA *Eugenia klotzschiana* O. BERG. (MYRTACEAE)”. Os trabalhos foram instalados pela Orientadora, Professora Doutora Thannya Nascimento Soares (PPGGMP/UFG), com a participação dos demais membros da Banca Examinadora: Professora Doutora Renata de Oliveira Dias (ICB/UFG), membro titular externo; Professora Doutora Ramilla dos Santos Braga Ferreira (UFR), membro titular externo e a Co-orientadora Professora Doutora Mariana Pires de Campos Telles (ICB/UFG), membro titular interno. Durante a arguição os membros da banca não fizeram sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Dissertação, tendo sido a candidata **Aprovada** pelos seus membros. Proclamados os resultados pelo Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora.

TÍTULO SUGERIDO PELA BANCA



Documento assinado eletronicamente por **Thannya Nascimento Soares, Professora do Magistério Superior**, em 18/07/2024, às 16:56, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Ramilla Dos Santos Braga, Usuário Externo**, em 19/07/2024, às 09:03, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Renata De Oliveira Dias, Professora do Magistério Superior**, em 19/07/2024, às 17:12, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Mariana Pires De Campos Telles, Professor do Magistério Superior**, em 02/04/2025, às 13:25, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **4643233** e o código CRC **651F04FE**.

Agradecimentos

Gostaria de começar agradecendo as instituições de fomento que viabilizaram a realização desse trabalho. Primeiramente, a Fundação de Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), pela concessão da bolsa, permitindo a minha estadia no mestrado durante esses anos. Ao apoio financeiro e logístico do Instituto Nacional de Ciência e Tecnologia (INCT) em Ecologia, Evolução e Conservação da Biodiversidade financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (proc. 465610/2014-5) e pela Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG) (proc. 201810267000023).

Ao Programa de Pós Graduação em Genética e Melhoramento de Plantas (PPGGMP) da Universidade Federal de Goiás (UFG), especialmente as coordenadoras Dra. Patrícia Melo e Dra. Marcela Resende, aos professores por todos os ensinamentos e aprendizados durante esse período, aos secretários e técnicos por todo apoio e esclarecimentos.

Ao Núcleo de Estudos em Genética, Melhoramento e Conservação de Plantas (Grupo EuGeM), por realizar simpósios que proporcionaram grandes aprendizados e formação de uma rede de *networking*, aos integrantes do grupo pela parceria, suporte e dedicação na realização dos eventos, especialmente à Luciana Lopes e Alessandra Gonçalves.

Ao Laboratório de Genética e Biodiversidade (LGBio -UFG) pela disponibilidade do ambiente físico, equipamentos, materiais e insumos laboratoriais. Agradeço imensamente toda a equipe do laboratório, e também aqueles que se tornaram amigos no meio do processo, por toda a ajuda na bancada e análise de dados, especialmente à Cíntia Targueta, Amanda Melo, Leonardo Corvalán, Jocilaine Santos, Diego Ortiz e Marla Arianne. Agradeço também, o professor José Alexandre Diniz Filho por disponibilizar seus *scripts* e criar uma função no *software* R para ajudar nas análises deste trabalho, mesmo estando em viagem de coleta ao Rio Araguaia.

Aos membros da banca examinadora, Dra. Renata de Oliveira Dias, Dra. Cíntia Pelegrineti Targueta de Azevedo Brito, Dra. Ramilla dos Santos Braga Ferreira e Dra. Adriana Maria Antunes pela disponibilidade em compor a banca e por suas respectivas contribuições para o aperfeiçoamento deste trabalho.

Quero agradecer imensamente a minha orientadora Dra. Thannya Nascimento Soares, por toda a assistência, conduta e ensinamentos durante o mestrado, por nunca ter me deixado desistir e acreditar que eu era capaz de superar todas as barreiras e concluir mais essa

etapa da minha vida acadêmica. Agradeço por toda paciência, orientações e sugestões ao corrigir os trabalhos, mesmo com toda a correria e prazos acadêmicos, sou verdadeiramente grata!

Agradeço a minha coorientadora Dra. Mariana Pires de Campos Telles, por todo apoio e todas as sugestões e correções detalhadas na escrita da dissertação, de resumos científicos, projeto do mestrado e também durante, a qualificação por todas as observações acerca da minha vida acadêmica e profissional, por ter aceitado ser coorientadora e fazer parte deste projeto, estando disponível para esclarecimentos e dúvidas!

Agradeço a Deus por confortar meu coração nas horas difíceis, por minha família, minha mãe Mara Ribeiro por ter me concedido o dom da vida, pois sem esta nada é possível. Tenho admiração e sou profundamente grata, ao meu pai Erivan Corrêa por toda a assistência durante a graduação e mestrado, por ter estado ao meu lado nos momentos mais difíceis, por nunca ter largado a minha mão quando precisei. Sou grata a minha prima Suênya Ponciano, por todo auxílio, amparo, conselhos e por sempre ter acreditado em mim!

A todos obrigada!!

SUMÁRIO

RESUMO.....	8
ABSTRACT.....	10
1 INTRODUÇÃO.....	12
2 OBJETIVOS	15
2.1 OBJETIVO GERAL.....	15
2.2 OBJETIVOS ESPECÍFICOS.....	15
3 MATERIAL E MÉTODOS.....	16
4 RESULTADOS E DISCUSSÃO.....	22
5 CONCLUSÕES.....	36
6 REFERÊNCIAS.....	37
APÊNDICES.....	40

RESUMO

Eugenia klotzschiana O. Berg. (Myrtaceae) é conhecida popularmente como pera-do-cerrado. A espécie está listada no grupo de plantas prioritárias para a pesquisa na publicação “Plantas para o Futuro da região Centro-Oeste”. A pera-do-cerrado pode ser consumida *in natura* ou na produção de geleias e doces, apresentando alto valor nutritivo. Os marcadores moleculares são ferramentas úteis para a caracterização dos recursos genéticos vegetais. Dentre os marcadores moleculares disponíveis, os microssatélites ou Sequências Simples Repetidas (SSRs), são úteis na investigação da estrutura genética para estudos populacionais em plantas como a *E. klotzschiana*, auxiliando na conservação e no seu uso sustentável. O objetivo deste estudo foi estabelecer um protocolo base para a genotipagem por sequenciamento (GBS) de marcadores SSR e disponibilizar locos polimórficos e informativos para a espécie *E. klotzschiana*. Assim, 22 pares de *primers* previamente desenvolvidos para a espécie foram utilizados para testar a genotipagem em diferentes sistemas multiplex, com a finalidade de verificar a influência na quantidade de regiões amplificadas simultaneamente. O DNA foi extraído do tecido foliar de 37 indivíduos coletados no município de Senador Canedo-GO utilizando o protocolo CTAB 2%. Para verificar a quantidade e integridade da extração realizou-se uma eletroforese horizontal em gel de agarose 1%. Após a confirmação da amplificação, os *primers* foram testados em diferentes sistemas de PCR multiplex (mPCR), contendo entre 5 e 22 pares de *primers* por sistema. As multiplex foram montadas em três testes, com diferentes quantidades de *primers* nos sistemas mPCR. No teste 1 (T1) foram montadas quatro multiplex com o mínimo de cinco e o máximo de seis *primers* cada; no teste 2 (T2) foram montadas duas mPCRs com a metade da quantidade de *primers* cada; no teste 3 (T3) foi montada uma única multiplex com todas as regiões amplificadas simultaneamente. Visando obter um melhor padrão de amplificação, testou-se diferentes concentrações de *primers* para as mPCRs. A partir da validação do *pool* de bibliotecas a 4nM, preparou-se o cartucho *Miseq* v2 de 500 ciclos para sequenciamento na plataforma *MiSeq Illumina*®. A análise da qualidade das sequências foi realizada no *FastQC* e a chamada dos alelos utilizando o *pipeline SSR-GBS*. Mesmo com ajuste de concentração de *primers*, T1 e T2 apresentaram padrão de amplificação com bandas mais intensas, quando comparados ao teste T3 em que utilizou-se todos os 22 *primers* em única reação de PCR. As análises sugerem que houve grande sucesso na genotipagem de microssatélites por sequenciamento em *E. klotzschiana*. A chamada de alelos

durante o sequenciamento, obteve uma baixa taxa de erro de chamamento de bases, com um valor médio de *Phred* de 38. Conforme esperado houve diferença na chamada de alelos entre os testes, com um aumento de dados perdidos no teste 3. A atribuição de alelos que foram chamados iguais pros mesmos locos entre os testes T1 e T2, ficaram acima de 90%, com exceção de um único loco que obteve 70% de similaridade. Como no teste T3 teve uma maior taxa de erros de genotipagem que precisaram de checagem manual (MC), tiveram poucas *reads* (PR) e/ou ficaram sem amplificação (NA), este obteve uma menor similaridade entre os locos avaliados. Por fim, a população amostrada apresentou-se como sendo predominantemente de indivíduos clonais. Os resultados obtidos demonstram a eficácia da genotipagem por sequenciamento (GBS) para a espécie *E. klotzschiana*, fornecendo informações valiosas para sua caracterização genética. Esses dados podem contribuir para futuras estratégias de conservação e uso sustentável da pera-do-cerrado.

Palavras-chave: diversidade genética; genotipagem por sequenciamento; pera do cerrado; recursos genéticos.

ABSTRACT

Eugenia klotzschiana O. Berg. (Myrtaceae) is popularly known as the Cerrado pear. The species is listed in the group of priority plants for research in the publication “Plants for the Future of the Central-West Region”. The Cerrado pear can be consumed in natura or in the production of jellies and sweets, presenting high nutritional value. Molecular markers are useful tools for the characterization of plant genetic resources. Among the available molecular markers, microsatellites or Simple Sequence Repeats (SSRs) are useful in the investigation of the genetic structure for population studies in plants such as *E. klotzschiana*, aiding in their conservation and sustainable use. The objective of this study was to establish a base protocol for genotyping by sequencing (GBS) of SSR markers and to make available polymorphic and informative loci for the species *E. klotzschiana*. Thus, 22 pairs of *primers* previously developed for the species were used to test genotyping in different multiplex systems, in order to verify the influence of the quantity of *primers* amplified simultaneously in the detection of alleles. DNA was extracted from the leaf tissue of 37 individuals collected in the municipality of Senador Canedo-GO using the CTAB 2% protocol. To verify the quantity and integrity of the extraction, horizontal electrophoresis was performed in 1% agarose gel. After confirmation of amplification, the *primers* were tested in different multiplex PCR (mPCR) systems, containing between 5 and 22 pairs of *primers* per system. The multiplexes were assembled in three tests, with different quantities of *primers* in the mPCR systems. In test 1 (T1), four multiplexes were assembled with a minimum of five and a maximum of six *primers* each; in test 2 (T2), two mPCRs were assembled with half the quantity of *primers* each; in test 3 (T3), a single multiplex was assembled with all *primers* amplified simultaneously. In order to obtain a better amplification pattern, different *primer* concentrations were tested for mPCRs. From the validation of the library pool at 4nM, the 500-cycle Miseq v2 cartridge was prepared for sequencing on the *MiSeq Illumina*® platform. Sequence quality analysis was performed in *FastQC* and allele calling using the *SSR-GBS pipeline*. Even with *primer* concentration adjustment, T1 and T2 presented amplification patterns with more intense bands, when compared to the T3 test in which all 22 *primers* were used in a single PCR reaction. The analyses suggest that there was great success in microsatellite genotyping by sequencing in *E. klotzschiana*. Allele calling during sequencing obtained a low base calling error rate, with an average *Phred* value of 38. As expected, there was a difference in allele calling between tests,

with an increase in lost data in test 3. The assignment of alleles that were called the same for the same loci between the T1 and T2 tests was above 90%, with the exception of a single locus that obtained 70% similarity. As the T3 test had a higher rate of genotyping errors that required manual checking (MC), had few reads (PR) and/or were left without amplification (NA), this obtained a lower similarity between the loci evaluated. Finally, the sampled population was predominantly made up of clonal individuals. The results obtained demonstrate the effectiveness of genotyping by sequencing (GBS) for the species *E. klotzschiana*, providing valuable information for its genetic characterization. These data may contribute to future conservation strategies and sustainable use of the Cerrado pear.

Keywords: genetic diversity; genotyping by sequencing; cerrado pear; genetic resources.

1. INTRODUÇÃO

Em meio as espécies que compõem o conjunto de recursos genéticos vegetais do Cerrado, destaca-se as frutíferas, que produzem frutos comestíveis, com formas variadas, cores atrativas e sabores característicos (Ribeiro & Rodrigues, 2006). Dentre as espécies vegetais frutíferas do Cerrado destaca-se a *Eugenia klotzschiana* O. Berg. (Myrtaceae), popularmente conhecida como pera-do-cerrado. Essa espécie distingue-se por seu potencial de uso, especialmente como uma fonte promissora de antioxidantes naturais com interesse industrial no setor alimentício, farmacêutico e cosmético (Carneiro et al., 2019; Sardi et al., 2017).

Eugenia klotzschiana está listada no grupo de plantas prioritárias para a pesquisa na publicação “Plantas para o futuro na região centro-oeste”, devido a sua precocidade de produção em larga escala, por ainda não existir pomares comerciais, além da importância na diversificação e enriquecimento da dieta para a população rural. Ademais, como a área de distribuição da espécie, a densidade de sua ocorrência no ambiente e a frequência de adultos produtivos são baixas, se torna limitado a disponibilidade de frutos para o consumo (Ministério do Meio Ambiente, 2016).

Os marcadores moleculares são ferramentas úteis para a caracterização dos recursos genéticos vegetais. Dentre os marcadores moleculares disponíveis, os microssatélites ou Sequências Simples Repetidas (SSRs), são úteis na investigação da estrutura genética para estudos populacionais em plantas. Os SSRs são importantes ferramentas na identificação, caracterização e avaliação dos recursos genéticos vegetais em diversas espécies, incluindo as frutíferas da família Myrtaceae (Portela et al., 2024). Esses marcadores são codominantes e altamente polimórficos, apresentando uma grande variedade de alelos em uma população, o que possibilita a identificação de indivíduos e a caracterização da diversidade genética de uma espécie (Faleiro, 2007).

Existem várias estratégias de genotipagem que foram desenvolvidas ao longo do tempo, como o RFLP (Polimorfismo de Comprimento de Fragmentos de Restrição) e AFLP (Polimorfismo de Comprimento de Fragmentos Amplificados) que são baseados na digestão de DNA com enzimas; RAPD (Polimorfismo de DNA Amplificado ao Acaso), SSRs e SNPs (Polimorfismo de Nucleotídeo Único) que utilizam *primers* para amplificar regiões do genoma e também, o NGS (Sequenciamento de Nova Geração) que permite a leitura de milhões de

fragmentos de DNA simultaneamente (Hodel et al., 2016). Atualmente, existe o SSR-Seq (Sequenciamento de Microssatélites) que faz uso da combinação de SSRs com NGS, sendo uma das mais avançadas técnicas de genotipagem para se obter dados robustos e confiáveis em sistemas multiplex para análise de múltiplos locos simultaneamente (Lepais et al., 2020). Essa evolução das estratégias de genotipagem, reflete um avanço significativo na capacidade de analisar a diversidade genética de uma espécie não-modelo.

Embora exista um conjunto crescente de outros marcadores moleculares e técnicas de sequenciamento, os SSRs ainda são uma das principais opções de genotipagem em espécies vegetais (Neophytou et al., 2018). Além disso, o uso desses marcadores moleculares para realizar a caracterização genética é ainda mais acessível, após o surgimento das tecnologias de sequenciamento de nova geração (NGS) e a identificação de numerosos locos em espécies não-modelo (Curto et al., 2019; Guichoux et al., 2011).

O uso de plataformas NGS na genotipagem por SSR-Seq tem impulsionado avanços significativos na caracterização genética de plantas nativas do cerrado. Tradicionalmente, a genotipagem de microssatélites era realizada por eletroforese capilar em sequenciadores Sanger, um método de alto custo, baixa reprodutibilidade e limitado pelo número reduzido de locos analisáveis por experimento (Sarhanová et al., 2018). Com a transição para plataformas NGS, como *MiSeq* da *Illumina*, a genotipagem por SSR-Seq tornou-se mais eficiente e acessível devido ao aumento da resolução e da profundidade dos dados obtidos, como também possibilita a análise de múltiplos indivíduos em um único ensaio, reduzindo custos e tempo experimental (Taheri et al., 2018). Além disso, a maior sensibilidade das plataformas NGS facilita a identificação de variantes alélicas raras e a detecção de polimorfismos em espécies de plantas nativas do Cerrado, contribuindo para a conservação e melhoramento genético desses recursos vegetais (Gonçalves et al., 2020).

A genotipagem por SSR-Seq é realizada usando a chamada de alelos com base em informações do fragmento amplificado e sequenciado em plataformas de NGS (Darby et al., 2016; Vartia et al., 2016). O SSR-Seq permite uma maior eficiência que pode ser alcançada sequenciando produtos de PCR multiplexados com um número muito maior de locos, além da análise de vários indivíduos simultaneamente (Sarhanová et al., 2018). O sequenciamento de várias amostras em um mesmo ensaio (multiplex) pode ser implementado marcando os fragmentos de DNA alvo dessas amostras com diferentes combinações de ‘identificadores’ por amostra de DNA (Chung et al., 2017). Na plataforma de sequenciamento *Illumina*®, essa etapa é realizada principalmente através da incorporação de indexes que são adicionados durante a

preparação das bibliotecas de sequenciamento por meio de reações de PCR e ligação de adaptadores (De Barba et al., 2017; Lepais et al., 2020).

Apesar do potencial de uso da espécie, até o presente momento, foram desenvolvidos poucos estudos genéticos para *E. klotzschiana*. Em 2014, foi realizada a transferibilidade de marcadores microssatélites gênicos, desenvolvidos para *Eucalyptus*, para caracterizar geneticamente indivíduos de *E. klotzschiana* (Siqueira, 2014). Em 2020, testou-se a transferibilidade de *primers* desenvolvidos para *Eugenia dysenterica* (cagaita) em *E. klotzschiana* para genes relacionados à tolerância hídrica (Araújo, 2020). Em 2023, foi publicado o primeiro rascunho do genoma nuclear parcialmente montado para *E. klotzschiana* (Corvalán et al., 2023). Neste artigo também foi disponibilizado pares de *primers* para amplificação de regiões microssatélites, desenvolvidos para serem utilizados no contexto de genotipagem por sequenciamento de Nova Geração (SSR-Seq), que necessitam de validação para ser utilizado. Neste contexto, o presente estudo teve como objetivo avaliar estratégias de amplificação em multiplex e validação de um protocolo para a genotipagem por sequenciamento de marcadores SSRs para *E. klotzschiana*, utilizando a plataforma *Illumina-MiSeq*®, para aplicação em estudos genético-populacionais.

2. OBJETIVOS

2.1 OBJETIVO GERAL

Estabelecer um protocolo base para a genotipagem por sequenciamento (GBS) de marcadores SSR e disponibilizar locos polimórficos e informativos para a espécie *Eugenia klotzschiana*.

2.2 OBJETIVOS ESPECÍFICOS

- i. Avaliar a eficiência de amplificação via PCR em multiplex dos *primers* em diferentes combinações;
- ii. Verificar se a chamada de genótipos/alelos é influenciada pela quantidade de *primers* do sistema multiplex;
- iii. Definir o número ideal de *primers* do sistema multiplex para uma genotipagem eficiente;
- iv. Disponibilizar a lista com os melhores locos, baseada nas estimativas de diversidade genética;
- v. Verificar a ocorrência de indivíduos clonais na população avaliada.

3. MATERIAL E MÉTODOS

Todas as etapas laboratoriais e de análise de dados deste estudo foram realizadas no Laboratório de Genética & Biodiversidade (LGBio), localizado no Instituto de Ciências Biológicas (ICB I) da Universidade Federal de Goiás (UFG) Campus Samambaia, Goiânia – GO.

Os 37 indivíduos utilizados no presente estudo são provenientes de uma coleta realizada em outubro de 2010 no município de Senador Canedo – GO de latitude 16°37 '32,197 " S e longitude 49°4 '22,696" W. As folhas coletadas foram armazenadas em sílica em gel, em um saco plástico e identificadas com o número e o código de cada indivíduo. As amostras foram transportadas até o laboratório da universidade e armazenadas em *freezer* – 80°C.

O DNA foi extraído a partir do tecido foliar de *Eugenia klotzschiana*, utilizando o protocolo de extração vegetal com tampão CTAB 2% adaptado de Doyle & Doyle (1987). A quantidade e a integridade do DNA extraído de cada amostra foi avaliada em eletroforese horizontal, com a aplicação de 3µL de DNA extraído em gel de agarose 1% corado com brometo de etídeo e tampão para eletroforese TBE 1X (Tris Borato EDTA) e com auxílio de marcadores de DNA (DNA Lâmbda - λ) nas concentrações de 10, 20, 50, 100 e 200 ng/µL. O DNA foi diluído para a concentração de estoque com aproximadamente 2,5 ng/µL.

Foram testados 25 pares de *primers* desenhados para amplificação de regiões microssatélites da espécie *E. klotzschiana* (Corvalán et al., 2023) (Tabela 1). Inicialmente os 25 pares de *primers* foram testados em um único indivíduo (EKLSCGO-25) de *E. klotzschiana*, para analisar a capacidade de amplificação de cada um dos *primers* individualmente, pois somente os que obtiveram amplificação nessa primeira avaliação, fizeram parte da etapa de multiplexação. A reação de PCR foi montada com volume final de 10 µL, utilizando o kit PCR multiplex da *Qiagen*®. Nesta reação os pares de *primers* foram testados a uma concentração final de 0,9 µM, conforme frequentemente se utiliza para a amplificação de pares de *primers* individualmente (Neophytou et al., 2018). A amplificação dos fragmentos de DNA foi realizada no equipamento *Veriti*® 96-Well Fast Thermal Cycler (*Applied Biosystems*®) e o programa de termociclagem continha as seguintes etapas: I) Desnaturação inicial do DNA a 95 °C por 15 minutos; II) 35 ciclos contendo os passos de desnaturação do DNA a 94 °C por 30 segundos, anelamento inicial de 52 °C e anelamento final de 54 °C por 90 segundos e extensão da molécula de DNA a 72 °C por 60 segundos; III) Extensão final da molécula de DNA a 60 °C por 30 minutos. O ajuste da temperatura de anelamento dos pares de *primers* de 54 °C, foi atingido

testando-se o seu aumento ou diminuição da temperatura, até que a visualização das bandas no gel indicasse a melhor resolução possível.

Tabela 1. Conjunto de 25 *primers* (Ekl_SSR-Seq) com os respectivos adaptadores (*Illumina*®) para genotipagem por sequenciamento de alto rendimento de marcadores microssatélites em *Eugenia klotzschiana*.

Nome do Primer	Motivo de repetição SSR (Número de repetições)	Primer Foward	Primer Reverse
Ekl_NGS2	AG ₁₂	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGTTGTACCCAAAGAAGAGCT	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGACTCTCCTAAACGCCTTA
Ekl_NGS3	AG ₁₂	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGGGAACGAAGCCTATCTCAA	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGAATGGAGCAAATCGAAACAC
Ekl_NGS5	AG ₁₃	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGCACCTATTGCCATTGGATCA	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGTGACCTCATGTTGCGATTTA
Ekl_NGS6	AG ₁₃	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGAGTTTGAGAACCGAATCGA	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGGCTCTCTCAAACCATCTGA
Ekl_NGS7	AG ₁₃	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGTCACTCAAGTCAGCCTAAA	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGGGTCTTCAGAAGTTGGGAT
Ekl_NGS9	AG ₁₄	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGAGGGCTTATATGGTAATCTCAC	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGATGAGTCGTGTAAGCAACA
Ekl_NGS10	AG ₁₂	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGTTTCTTTATCGGGTCACCTC	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGCTGATGCACCATTCTCT
Ekl_NGS12	AACG ₆	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGCCTCCTAACAAGACTTGCAT	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGCCTTACCTTATGACGTCGTT
Ekl_NGS13	AG ₁₄	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGAGAAGGGACTAGTGTAGCTT	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGCTGAATAGACAACCGAGGTT
Ekl_NGS16	AG ₁₂	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGCGGGACATATAACAATCCGT	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGCCGTTGTTAATAGGGTCACA
Ekl_NGS17	AG ₁₃	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGAATTCCAAAGTTCGTCTCA	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGATATGACCGATGTATGCTCC
Ekl_NGS18	AG ₁₃	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGCAGTGATGTCTATCAGCTCC	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGCGCTAGTTTGAAGAGCAAA
Ekl_NGS19	AG ₁₃	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGACTACACGGTCAAGAAAGTC	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGAGCTTGGACCAGAACTAT
Ekl_NGS27	AG ₁₂	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGCAGTTCACCAGATTTAGCCT	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGCTCTCCCTGTTCAATCAGT
Ekl_NGS29	AG ₁₂	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGATCGAGTCTTTTCAGTGCT	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGATCAACGACTTCCAATGTGA
Ekl_NGS30	AGGG ₆	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGGCAGGTAATCCAGTTTCAGA	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGAAATCTTCTGTTACGTCGA
Ekl_NGS31	AG ₁₄	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGCTGTGTTGCGTTAAACCA	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGGATTCGCGGTTATCGAAAT
Ekl_NGS32	AG ₁₃	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGTCCGCATTTGGTACTACTT	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGAAGTCTCTAGTCTGACC
Ekl_NGS33	AG ₁₂	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGTCAAGAATCATTCCAAGC	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGTTCACCTGTCTCTCCCAT
Ekl_NGS34	AG ₁₂	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGAATCGCACTGATACCAAAGT	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGCAGTTTAGCTGTCTCATCGA
Ekl_NGS35	ACATG ₆	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGACTGAGCAATTGAAGTTGGA	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGCCCTGTTCTTTCTAGAC
Ekl_NGS38	AG ₁₄	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGATATTCTCTTTGCGGACC	GTCTCGTGGGCTCGGAGGTGTATAAGAG ACAGCGAAACAAGTTGCATAGCTT
Ekl_NGS39	AG ₁₃	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGTGAGACGTTGGAGTACTT	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGTTCCTAATGTTCCGCCATA
Ekl_NGS41	AG ₁₂	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGTCAGATCTCACAACATTGCA	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGAACCATTGCACTCTCTT
Ekl_NGS42	AG ₁₂	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGGCCATCCTTGTAACCCTA	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGCTCATGTGCGAGTCTTACA

Em negrito são as sequências dos adaptadores. Fonte: Corvalán et al. (2023) (com adaptações).

Para verificar a qualidade de amplificação de cada par de *primer*, os produtos de PCR foram submetidos a uma eletroforese horizontal em gel de agarose 3%, corado com brometo de etídeo, solução tampão TBE 1x e marcador 1 Kb plus (4 µL), com uma corrente elétrica de 70V, por aproximadamente 1 hora. Para verificar os locos utilizou-se produto da reação de PCR (2 µL/amostra) e tampão de corrida (4 µL/amostra). Os fragmentos amplificados foram visualizados por meio de fotodocumentador de imagem com exposição à luz ultravioleta.

Os pares de *primers* que apresentaram bom perfil de amplificação individual foram submetidos à três testes de amplificação em sistemas de PCR multiplex (mPCRs). O teste 1 (T1) foi composto por quatro sistemas multiplex, distribuindo o total de pares de *primers* em quatro conjuntos; o teste 2 (T2) foi composto por duas multiplex, sendo cada uma com a metade da quantidade de *primers*; e o teste 3 (T3) consistiu de uma única multiplex com todos pares de *primers* (Tabela 2). Esses testes foram realizados com 37 indivíduos nas mesmas condições de amplificação individual, conforme apresentado anteriormente. Visando obter um melhor padrão de amplificação em sistemas multiplex, testou-se diferentes concentrações de *primers*. As condições de amplificação no termociclador e a visualização das bandas no gel seguiram as mesmas condições citadas anteriormente para a amplificação individual dos pares de *primers*.

Tabela 2. Distribuição dos 22 pares de *primers* SSR-Seq de *Eugenia klotzschiana* nos sistemas de PCR multiplex.

Teste 1			Teste 2			Teste 3		
Multiplex	Primers (3 µM)	Qtd	Multiplex	Primers (4 µM)	Qtd	Multiplex	Primers (5 µM)	Qtd
M1	Ek1_NGS2	5	M1	Ek1_NGS2	11	M1	Ek1_NGS2	22
	Ek1_NGS3			Ek1_NGS3			Ek1_NGS3	
	Ek1_NGS5			Ek1_NGS5			Ek1_NGS5	
	Ek1_NGS6			Ek1_NGS6			Ek1_NGS6	
	Ek1_NGS9			Ek1_NGS9			Ek1_NGS9	
M2	Ek1_NGS10	5	M1	Ek1_NGS10	11	M1	Ek1_NGS10	22
	Ek1_NGS12			Ek1_NGS12			Ek1_NGS12	
	Ek1_NGS13			Ek1_NGS13			Ek1_NGS13	
	Ek1_NGS16			Ek1_NGS16			Ek1_NGS16	
	Ek1_NGS17			Ek1_NGS17			Ek1_NGS17	
M3	Ek1_NGS18	6	M2	Ek1_NGS18	11	M1	Ek1_NGS18	22
	Ek1_NGS27			Ek1_NGS27			Ek1_NGS27	
	Ek1_NGS29			Ek1_NGS29			Ek1_NGS29	
	Ek1_NGS30			Ek1_NGS30			Ek1_NGS30	
	Ek1_NGS31			Ek1_NGS31			Ek1_NGS31	
	Ek1_NGS32			Ek1_NGS32			Ek1_NGS32	
M4	Ek1_NGS33	6	M2	Ek1_NGS33	11	M1	Ek1_NGS33	22
	Ek1_NGS34			Ek1_NGS34			Ek1_NGS34	
	Ek1_NGS35			Ek1_NGS35			Ek1_NGS35	
	Ek1_NGS38			Ek1_NGS38			Ek1_NGS38	
	Ek1_NGS41			Ek1_NGS41			Ek1_NGS41	
	Ek1_NGS42			Ek1_NGS42			Ek1_NGS42	

Qtd = quantidade; as diferentes escalas de cinza indicam o conjunto de *primers* para um sistema multiplex; Temperatura de anelamento em cada multiplex de 54 °C; tamanho esperado do fragmento amplificado 250 pb.

O preparo da biblioteca genômica dos indivíduos de *E. klotzschiana* seguiu o protocolo de 16S da *Illumina*® com algumas modificações. Os produtos amplificados foram purificados com 12 µL de *beads* magnéticas *AMPure*. Em seguida, as *beads* foram eluídas em 22 µL de Tris-HCl pH 8,0, transferindo-se para uma placa nova 20 uL do preparo de biblioteca. Posteriormente, foi realizada uma PCR para a inserção dos *indexes* de identificação de cada um dos indivíduos em cada uma das multiplex. Após a PCR de inserção de *index*, foi feita a limpeza com as *beads*. Para a quantificação de todas as amostras amplificadas foi utilizado o Kit Qubit *High sensitivity* e para a verificação das bibliotecas usou-se o Kit DNA1000 do *Bioanalyzer*.

Depois da quantificação, realizou-se a diluição das bibliotecas para 4 nM, utilizando Tris-HCL a 10 mM. Foi montado o *pool* das bibliotecas para sequenciamento na plataforma *MiSeq Illumina*®. Pipetou-se 5 µL de cada biblioteca em um tubo de 1,5 mL. Quantificou-se o *pool* no Qubit para checar se a concentração permaneceu a 4 nM. Para o sequenciamento, foi utilizado o cartucho *MiSeq v2* de 500 ciclos.

As análises dos dados brutos foram realizadas usando os servidores do Laboratório LGBio do ICB I – UFG. Para a análise e controle de qualidade dos dados do sequenciamento, foi utilizado o *software FastQC* (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), que permitiu detectar possíveis problemas no conjunto de dados do sequenciamento. O *software* PEAR (Zhang et al., 2014) foi utilizado para fazer a montagem das sequências R1 (*Foward*) e R2 (*Reverse*), que são as sequências resultantes do sequenciamento *paired-end*, que podem ser unidas em uma sequência única (consenso), assim esta etapa sucedeu-se para todas as bibliotecas que foram sequenciadas.

Para a obtenção da tabela de genótipos foi utilizado o *pipeline SSR-GBS* (<https://github.com/mcurto/SSR-GBS-pipeline>), descrito por Tibihika et al. (2019) e Curto et al. (2019). Para tanto, foi preparado um arquivo de *input* das análises, com as seguintes informações: coluna A nome no *primer* separado por hífen e motivo de repetição separado por *underline* (ex.: EKL-NGS2_AG), seguido da coluna B contendo a sequência flanqueadora do *primer* 5' *forward* e coluna C com a sequência flanqueadora do *primer* 3' *reverse*, conforme descrito na Tabela 1.

O *pipeline SSR-GBS* (Tibihika et al., 2019) é dividido em duas partes, em que na primeira é produzida uma matriz codominante com genótipos baseados no comprimento do alelo e na segunda parte os genótipos são baseados na composição de bases das sequências. Nesta primeira etapa, o *script primer_demultiplex.py* demultiplexa arquivos *fastq* mesclados de acordo com o conteúdo do *primer*, as saídas são um arquivo *fastq* por amostra e loco. Em

seguida, o *script CountLengths.sh* conta o número de ocorrências (*reads*) de cada comprimento de sequência por alelo, gerando um arquivo em que a primeira coluna é o comprimento do alelo e a segunda coluna o número de ocorrências deste. O *script Rscript_Markerlength_develop_Color.R* usa as contagens de comprimento de sequência para chamar alelos e definir genótipos. Um genótipo é considerado homocigoto se a contagem relativa de leitura do alelo mais abundante for maior que o valor alfa definido (0,7 ou 70%). Um genótipo foi considerado heterocigoto quando a soma da abundância relativa de dois comprimentos de leitura mais abundantes excedeu o valor alfa (0,7 ou 70%), considerando ainda: (i) se a diferença no comprimento dos alelos potenciais foi maior que uma vez o comprimento do motivo de repetição; (ii) se os dois alelos potenciais mais abundantes diferirem apenas no comprimento de uma repetição, o alelo menos abundante foi considerado verdadeiro se for mais longo que o alelo mais abundante; (iii) se o segundo maior alelo potencial for apenas uma vez menor que o comprimento do motivo repetido, ele só foi considerado se sua abundância relativa for igual a 0,75 do alelo mais abundante; (iv) e ainda se, foi considerado mutações pontuais que levam a múltiplos não inteiros do comprimento da repetição, se a sua abundância relativa fosse de pelo menos 0,6 do alelo mais abundante. Caso esses critérios não fossem atendidos a amostra era considerada para controle manual (*man_check*). Ao final desta primeira etapa é produzida uma planilha de genótipos e histogramas de comprimento destacando os alelos chamados com cores diferentes, sendo azul para o alelo homocigoto, verde para os alelos heterocigotos e preto para os demais comprimentos de sequência. Quando não havia leituras para um dado indivíduo em um dado loco, o pipeline insere “NA” no local do genótipo. Caso a contagem de leituras para um dado indivíduo em um dado loco fosse muito baixa, o *pipeline* insere a mensagem “Poucas reads” no local do genótipo.

Na segunda etapa do *pipeline*, o *script Sequence_Allele_Call.py* extraiu as leituras correspondentes aos alelos chamados na etapa anterior, produzindo uma sequência consenso para cada alelo baseado em comprimento, destacando possíveis variações de SNPs.

Os genótipos provenientes de cada um dos três testes foram comparados entre si visando avaliar diferenças nas taxas de erros e na chamada de alelos. Os erros de genotipagem foram classificados como *man_check* (MC), que indica a necessidade de conferência manual do alelo atribuído pelo *pipeline*; ausência de amplificação (NA), que indica que não houve produto amplificado para aquele alelo/indivíduo; e poucas *reads* (PR), que significa que não foi possível atribuir o alelo porque a contagem de *reads* foi muito baixa (normalmente abaixo de 50 *reads*) para aquele genótipo (Apêndice A).

Para comparar se havia diferença significativa na quantidade de erros de genotipagem entre os três testes, foi conduzida uma Análise de Variância de Medidas Repetidas (<https://www.r-bloggers.com/2021/04/repeated-measures-of-anova-in-r-complete-tutorial/>), para cada um dos tipos de erros (MC, NA e PR). A comparação da quantidade de erros entre os pares de testes foi realizada por um Teste de T par a par.

A comparação da chamada de alelos entre os testes, ou seja, para verificar se o alelo atribuído para cada indivíduo foi igual entre os testes, foi realizada uma medida de similaridade de alelos atribuídos, atribuindo-se 0 para os locos que tinham NA e 1 para os locos contendo PR. Para tanto, foi utilizado um *script* em linguagem R (cedido gentilmente por Diniz-Filho, (Apêndice B) que inicialmente “contou” o número de alelos diferentes atribuídos para os mesmos indivíduos, entre os pares de testes diferentes. Em seguida, foi criada uma matriz de distância par a par entre os testes, baseada no número de alelos diferentemente atribuídos. A similaridade de atribuição dos alelos foi obtida a partir do cálculo do inverso da matriz de distância. O resultado foi apresentado como a proporção de alelos atribuídos de forma similar entre os pares de testes. Essa análise também foi realizada por loco entre os pares de testes.

A partir dos resultados das análises de erros e atribuição de alelos, foi definido o melhor teste/esquema de PCR multiplex para a genotipagem dos SSR-Seq para os locos avaliados. A planilha de genótipos do teste escolhido foi utilizada para avaliar a qualidade dos locos, quanto aos seguintes parâmetros de diversidade genética: número médio de alelos por loco (A); diversidade genética de Nei (h); heterozigosidade observada (H_o); riqueza alélica (Ar) e o índice de fixação intrapopulacional (f). Também foi realizada a identificação dos possíveis clones, ou seja, indivíduos com genótipos idênticos a similaridade na chamada de alelos entre os indivíduos, usando a mesma função do *software* R descrita anteriormente (cedido gentilmente por Diniz-Filho, (Apêndice B).

Por fim, o editor *BioEdit* foi utilizado para fazer o alinhamento global das sequências para cada indivíduo e cada loco analisado, utilizando a ferramenta *ClustalW*, para confirmar a região microssatélite do loco, produzir uma “escada alélica” e confirmar a chamada dos alelos, além de detectar a presença de SNPs dentro e fora (regiões flanqueadoras) dessas regiões microssatélites.

4. RESULTADOS E DISCUSSÃO

No que se refere a integridade do DNA genômico de *Eugenia klotzschiana*, os resultados das extrações dos 37 indivíduos utilizados neste experimento apresentaram bandas de DNA bastante intensas com uma quantificação de DNA em torno de 60 ng/ μ L na maioria das amostras, com exceção das amostras dos indivíduos 3, 6, 17 e 44 que apresentaram bandas mais fracas, e por isso precisaram ser submetidos a uma nova extração (Figura 1).

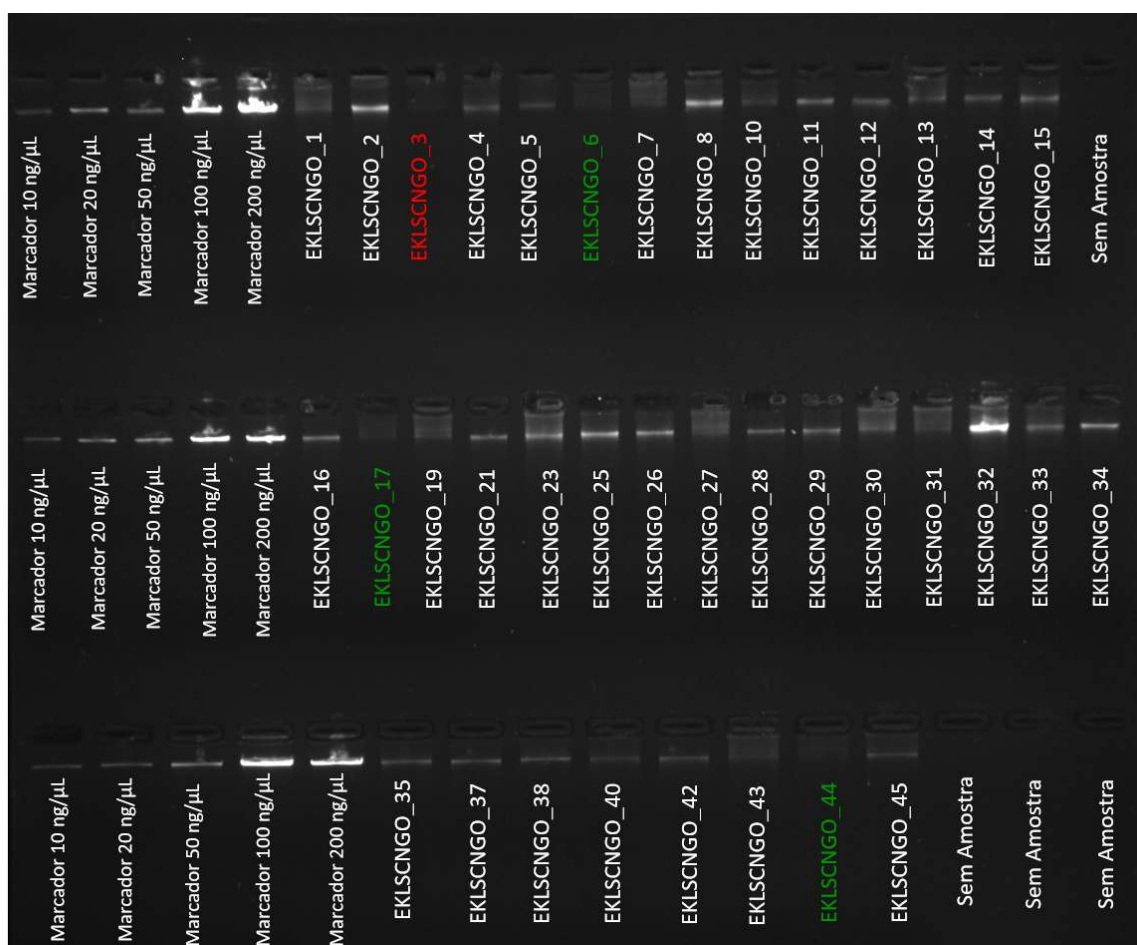


Figura 1. Imagem do gel de agarose 1% da extração de DNA vegetal dos 37 indivíduos de *Eugenia klotzschiana* de Senador Canedo utilizando protocolo CTAB 2%.

Um total de 22 (88%) dos 25 pares de *primers* de *E. klotzschiana* testados durante a realização da PCR singleplex amplificaram de forma satisfatória, assim os *primers* EKL_NGS7, EKL_NGS19 e EKL_NGS39 não obtiveram êxito e, portanto, foram excluídos na realização das PCRs seguintes (Figura 2).

Para se obter um produto de PCR que seja satisfatório, é necessário que se tenha um par de *primers* específico capaz de se ligar exclusivamente na região genômica de interesse a ser amplificada. Portanto, os *primers* EKL_NGS7, EKL_NGS19 e EKL_NGS39, indicaram uma baixa especificidade para o genoma de *E. klotzschiana*. Contudo, a quantidade de *primers* amplificados (88%), está dentro do esperado, por se tratar de *primers* que só haviam sido testados *in silico* para a espécie, até o presente estudo.

O número de *primers* amplificados foi superior aos encontrados em outros trabalhos, já desenvolvidos com *E. klotzschiana* que realizaram a transferibilidade de *primers* provenientes de outras espécies do Cerrado, com o objetivo de também montar sistemas de PCRs multiplex, para a pera-do-cerrado, como no estudo de Siqueira (2014), por isso os pares de *primers* utilizados neste estudo foram considerados amplificados com sucesso .

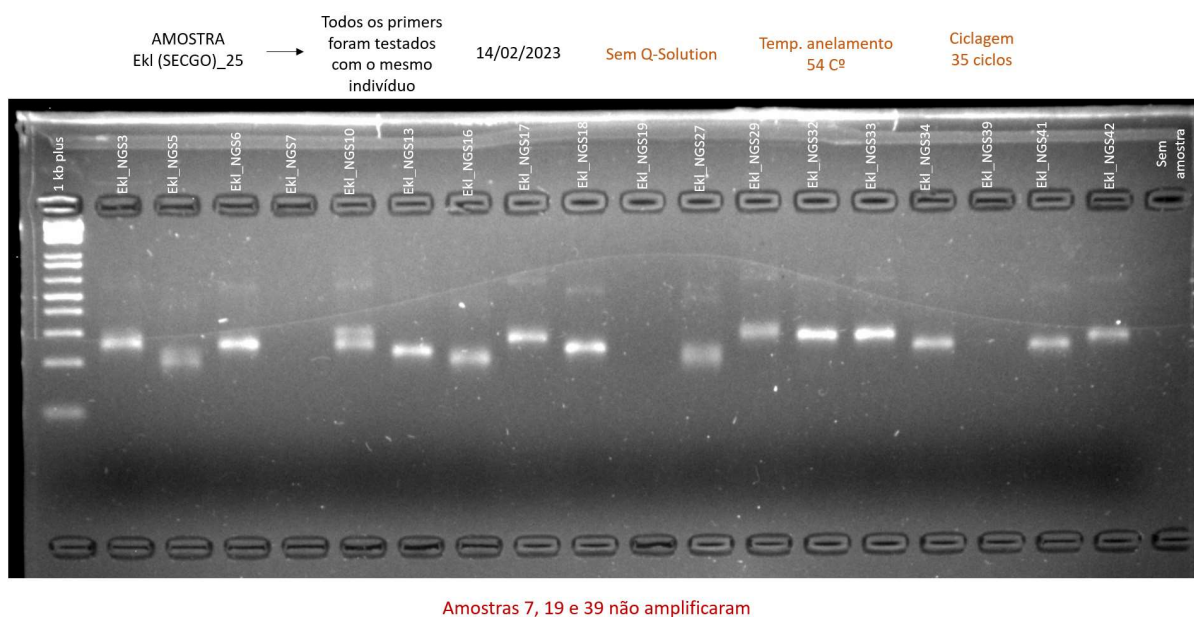


Figura 2. Padrão de amplificação individual dos locos microsatélites de *Eugenia klotzschiana* em gel de agarose 3% corado com brometo de etídeo e marcador molecular 1 kb plus.

Verificou-se que as concentrações de 3, 4 e 5 μ M, respectivamente, foram melhores para a amplificação dos testes T1, T2 e T3. Nas PCRs multiplex mesmo com o ajuste de concentração de *primers*, os testes T1 (Figuras 3 a 6) e T2 (Figuras 7 e 8) apresentaram padrão de amplificação com bandas mais intensas, quando comparados ao teste T3 (Figura 8) em que utilizou-se todos os 22 *primers* em uma única reação de PCR. No entanto, foi possível verificar a presença de amplificação nos três esquemas de multiplex.

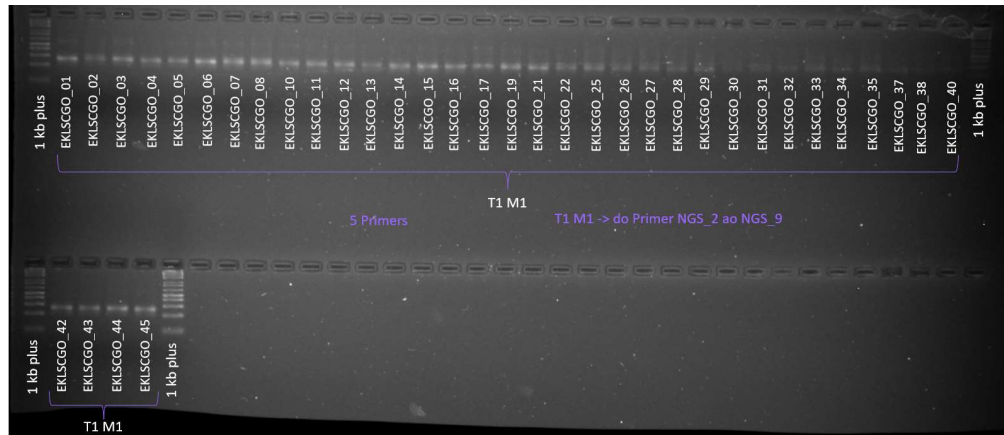


Figura 3. Amplificação dos locos microssatélites de *Eugenia klotzschiana* em gel de agarose 3% corado com brometo de etídeo e marcador molecular 1 kb plus. Teste 1 multiplex 1.

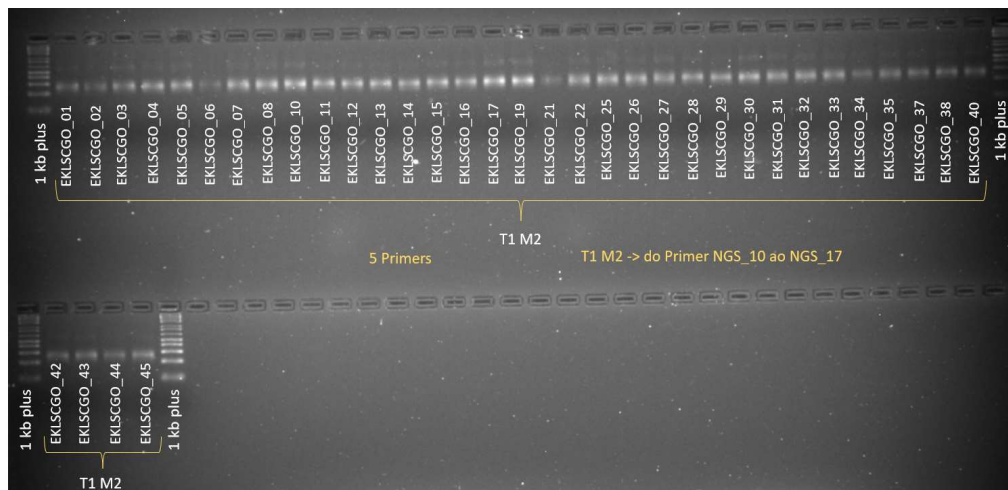


Figura 4. Amplificação dos locos microssatélites de *Eugenia klotzschiana* em gel de agarose 3% corado com brometo de etídeo e marcador molecular 1 kb plus. Teste 1 multiplex 2.

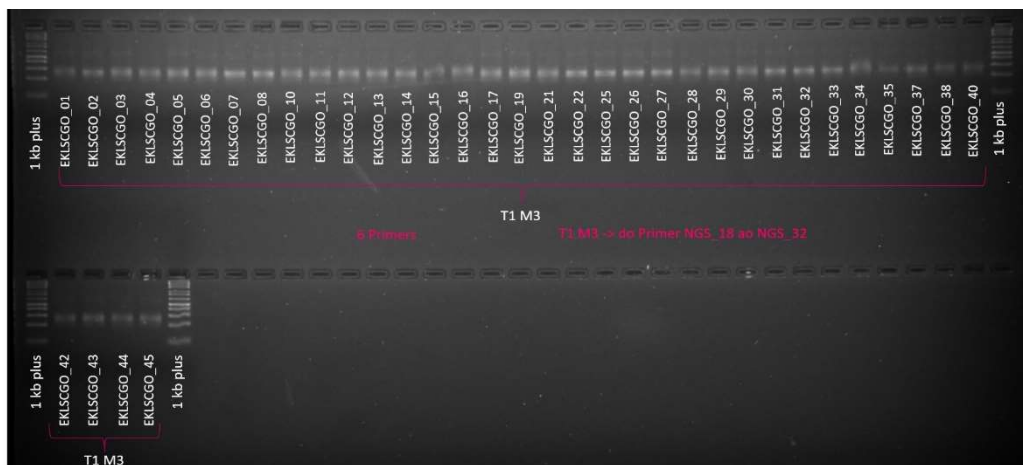


Figura 5. Amplificação dos locos microssatélites de *Eugenia klotzschiana* em gel de agarose 3% corado com brometo de etídeo e marcador molecular 1 kb plus. Teste 1 multiplex 3.

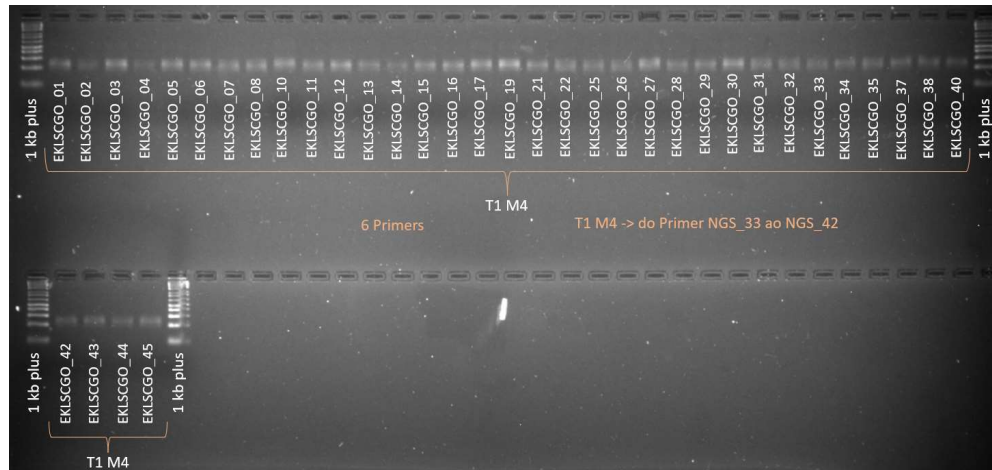


Figura 6. Amplificação dos locos microssatélites de *Eugenia klotzschiana* em gel de agarose 3% corado com brometo de etídeo e marcador molecular 1 kb plus. Teste 1 multiplex 4.

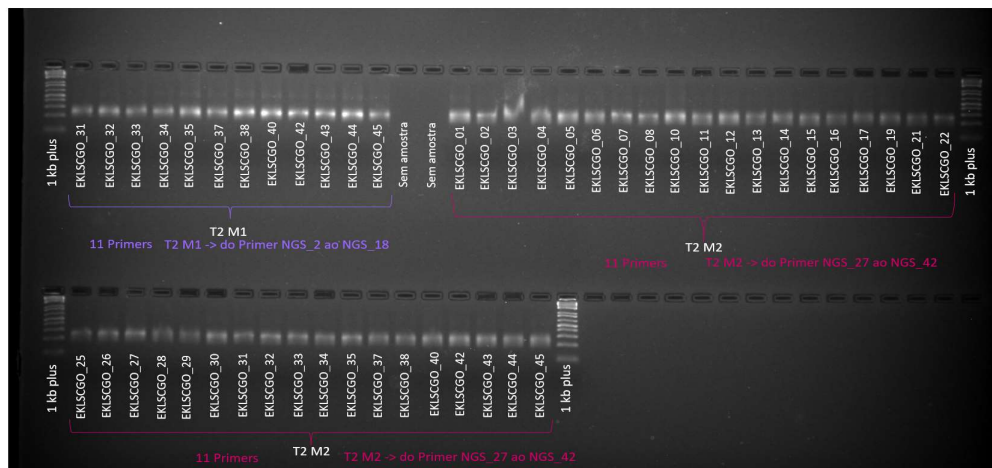


Figura 7. Amplificação dos locos microssatélites de *Eugenia klotzschiana* em gel de agarose 3% corado com brometo de etídeo e marcador molecular 1 kb plus. Teste 2 multiplex 1 e 2.

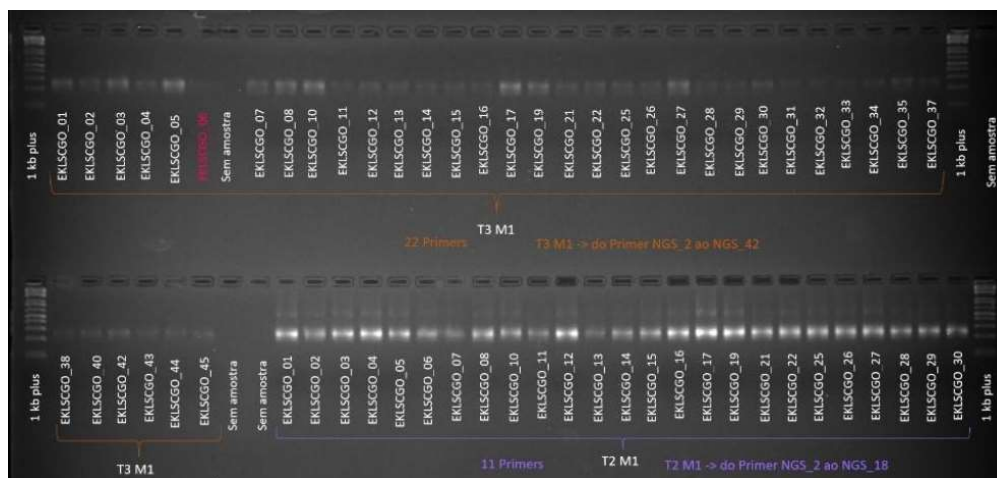


Figura 8. Amplificação dos locos microssatélites de *Eugenia klotzschiana* em gel de agarose 3% corado com brometo de etídeo e marcador molecular 1 kb plus. Teste 3 multiplex 1 e Teste 2 multiplex 1

A quantificação da diluição do pool das bibliotecas para sequenciamento na plataforma *Illumina® Miseq* obteve o resultado ideal esperado de 4 nM, valor considerado adequado para realizar a genotipagem por sequenciamento de alto rendimento. Ademais, o tamanho médio da biblioteca ficou em torno de 300 (pb), enquanto a quantificação da biblioteca no Qubit foi de 0,792 ng/ μ L.

Os arquivos resultantes do sequenciamento *paired-end*, para o teste singleplex do indivíduo EKLSCGO-25, tanto para o sequenciamento dos testes multiplex obtiveram um total de 250 pb em relação ao comprimento da sequência mais longa do conjunto, pois devido ao sequenciamento ser *paired-end* e ao modelo do cartucho utilizado, era esperado que este fosse o maior comprimento de sequência que seria fornecida com o sequenciamento de alto rendimento.

Na avaliação dos 37 indivíduos a partir dos gráficos obtidos pelo *software FastQC*, a qualidade média da chamada de bases durante o sequenciamento foi considerada muito boa, com um valor médio de *Phred* de 38, que indica uma alta precisão na chamada de bases durante o sequenciamento SSR-Seq, com uma probabilidade de erro na chamada de bases de apenas 0,02%(EWING; GREEN, 1998). Esse aspecto, reflete uma maior confiança na qualidade dos dados do sequenciamento de *Eugenia klotzschiana*, o que é essencial para a obtenção de parâmetros de diversidade genética robustos e confiáveis. Além disso, a queda da qualidade da chamada de bases no sequenciamento só começou a aparecer à medida que a execução avançou, o que é normal na maioria das plataformas de sequenciamento, por isso foi comum perceber a chamada de bases caindo ao final de cada leitura.

A comparação visual em *boxplot* da quantidade de “erros de genotipagem” entre os testes evidenciou que T1 e T2 apresentaram os maiores números de alelos que necessitaram de checagem manual (MC) em comparação ao T3. Entretanto, vale ressaltar que a necessidade de checagem manual por si só não indica um problema, pois os testes que têm mais chamada de alelos também estão mais propensos a terem a necessidade de checagem manual. Isso pode ser corroborado pelo fato de que o teste T3 apresentou muito mais falhas de amplificação (NA) de alelos do que os testes T1 e T2, que apresentaram médias de 0,541 e 0,919 para esse erro de genotipagem. No entanto, quando se compara o número de alelos não chamados em função da baixa quantidade de *reads* (PR), fica evidente que os três testes se diferenciam entre si, com o teste T1 apresentando o menor valor para esse parâmetro (Figura 9).

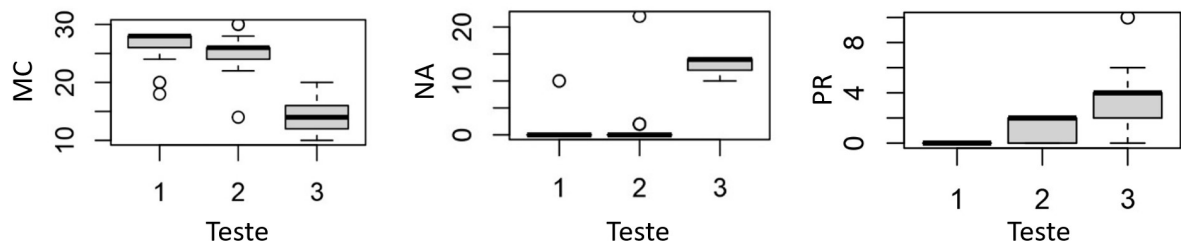


Figura 9. *Boxplot* dos “erros de genotipagem” do tipo MC (Checagem manual), NA (Sem amplificação) e PR (Poucas *reads*), para os testes T1, T2 e T3, mostrando a média para cada um dos erros em cada teste e os pontos *outliers*, indicando valores que foram discrepantes do conjunto de dados.

A análise de variância, seguida do teste de T pareado, corroborou com a análise visual dos *boxplots*, apresentando que os testes T1 e T2 não apresentaram diferenças significativas para os erros MC e NA, enquanto para o erro do tipo PR a diferença foi significativa entre os três testes (Tabela 3). Esse resultado indica que o uso de sistemas de PCR multiplex com menor número de pares de primers diminui a chance de ter erros de genotipagem devido ao baixo número de *reads*, provavelmente por ter maior quantidade de produto amplificado.

Tabela 3. Análise de variância de medidas repetidas comparando o número de alelos com erros de genotipagem por sequenciamento SSR-Seq. Testes a posteriori (Teste de T par a par) entre os diferentes testes de esquemas de amplificação em sistemas de PCR multiplex de locos microssatélites para *Eugenia klotzschiana*.

Tipo de Erro	F	p	Valor de p dos Testes a Posteriori		
			T1-T2	T1-T3	T2-T3
MC	290,21	$3,47 \cdot 10^{-35}$	$2,12 \cdot 10^{-1}$ ns	$4,65 \cdot 10^{-22}$	$3,90 \cdot 10^{-20}$
NA	255,47	$2 \cdot 10^{-33}$	1 ns	$2,49 \cdot 10^{-23}$	$3,96 \cdot 10^{-20}$
PR	58,81	$7,25 \cdot 10^{-16}$	$4,38 \cdot 10^{-7}$	$4,35 \cdot 10^{-10}$	$7,38 \cdot 10^{-7}$

MC=checagem manual; NA=não amplificado; PR=poucas *reads*

F= indica a variabilidade do erro entre os testes

p= significância estatística de F

ns=não significativo

Os testes T1 e T2 apresentaram alta similaridade na chamada de alelos (92,75%), enquanto o T3 apresentou baixa similaridade na chamada de alelos com T1 (61,7%) e T2 (56,5%), quando se considera o total de indivíduos e locos analisados (Tabela 4). Quando se compara a chamada de alelos entre os testes, considerando cada loco, fica mais evidente a semelhança entre T1 e T2, com altos valores de similaridade, exceto para o loco Ekl_NGS30 (27,02%) (Tabela 5). A análise na tabela de genótipos revelou que o T2 apresentou muitos erros

do tipo PR e NA (que foram considerados como informação nesta análise) para esse loco, o que justifica a baixa similaridade na chamada de alelos. Analisando a chamada de alelos de forma isolada, é possível afirmar que os diferentes esquemas de PCR multiplex dos testes T1 e T2 não apresentaram diferenças relevantes na chamada de alelos.

Embora o teste T3 tenha apresentado menor similaridade de chamada de alelos com os demais testes (Tabela 4), a análise por loco evidencia que um pouco mais da metade (12 no total) apresenta alta similaridade (maior que 85%) com T1 (Tabela 5). Isso demonstra que a chamada de alelos é influenciada pelo par de primers que amplifica a região alvo.

Tabela 4. Similaridade de atribuição de alelos entre os testes considerando todos os indivíduos.

	T1(%)	T2(%)	T3(%)
T1(%)	-	-	-
T2(%)	0,9275184	-	-
T3(%)	0,6167076	0,5651106	-

Tabela 5. Similaridade de atribuição de alelos entre os testes por loco considerando todos os indivíduos.

Loco	Similaridade por loco entre testes (%)		
	T1 e T2	T1 e T3	T2 e T3
Ekl_NGS2	0,9459459	0,05405405	0
Ekl_NGS3	0,9189189	0,05405405	0
Ekl_NGS5	0,9324324	0,05405405	0
Ekl_NGS6	0,9459459	0,05405405	0
Ekl_NGS9	0,9324324	0,05405405	0
Ekl_NGS10	0,972973	0	0
Ekl_NGS12	1	1	1
Ekl_NGS13	1	0,29729730	0,2972973
Ekl_NGS16	1	0,97297297	0,97297297
Ekl_NGS17	1	1	1
Ekl_NGS18	1	0,93243243	0,93243243
Ekl_NGS27	0,9189189	0,85135135	0,85135135
Ekl_NGS29	0,972973	0,82432432	0,81081081
Ekl_NGS30	0,2702703	1	0,27027027
Ekl_NGS31	0,972973	1	0,97297297
Ekl_NGS32	0,972973	0,95945946	0,94594595
Ekl_NGS33	0,972973	0,79729730	0,7972973
Ekl_NGS34	0,972973	0	0,02702703
Ekl_NGS35	0,972973	0,95945946	0,93243243
Ekl_NGS38	0,972973	1	0,97297297
Ekl_NGS41	0,7837838	0,70270270	0,67567568
Ekl_NGS42	0,972973	1	0,97297297

A partir da análise global de comparação entre os testes, o esquema de PCR multiplex do teste T1, com os menores números de pares de *primers* por sistema apresenta uma qualidade global maior, sugerindo que este deve ser o adotado para as análises de SSR-Seq com a espécie *E. klotzschiana*. Nesse sentido, as análises de diversidade genética dos locos foram realizadas a partir da planilha de genótipos do teste T1.

Dos 22 locos analisados na população foram verificados 21 polimórficos e apenas 1 monomórfico, com uma média de 2,9 alelos, variando de 1 a 5 alelos nos locos (Tabela 6). Embora o número de locos polimórficos tenha sido alto, o número de alelos por loco foi baixo, levando em consideração que locos microssatélites são potencialmente multialélicos (Pedra et al., 2023). No entanto, como foi analisada apenas uma população e, considerando que a espécie pode apresentar clones, o número de alelos por locos pode ser maior, quando se analisar novas populações naturais.

A heterozigosidade esperada (H_e) média da população foi igual a 0,55 (Tabela 6), esse valor está dentro do esperado para o médio número de alelos detectados (2,9). A heterozigosidade observada (H_o) média da população foi muito alta (0,87), com muitos indivíduos heterozigotos na população (Tabela 6). Em consonância a esse resultado, os valores do índice de fixação intrapopulacional f foram negativos, evidenciando um excesso de heterozigotos em relação ao esperado pelo equilíbrio de *Hardy-Weinberg*. Novamente observa-se uma limitação de interpretação deste resultado, em função da amostragem de uma única população com indivíduos potencialmente clones.

Tabela 6. Análise de diversidade genética com marcadores microssatélites (SSR-Seq) da população de *Eugenia klotzschiana* do município de Senador Canedo – GO.

Loco	Parâmetros de Diversidade Genética					
	<i>n</i>	<i>A</i>	<i>Ar</i>	<i>He</i>	<i>Ho</i>	<i>f</i>
Ekl_NGS2	35	3	3,000	0,617	1,000	-0,637
Ekl_NGS3	35	2	2,000	0,507	1,000	-1,000
Ekl_NGS5	35	4	4,000	0,631	1,000	-0,599
Ekl_NGS6	35	3	3,000	0,617	1,000	-0,637
Ekl_NGS9	35	4	4,000	0,726	1,000	-0,385
Ekl_NGS10	37	5	4,946	0,728	1,000	-0,381
Ekl_NGS12	37	3	2,946	0,471	0,703	-0,501
Ekl_NGS13	37	2	2,000	0,462	0,703	-0,532
Ekl_NGS16	37	4	3,946	0,622	1,000	-0,621
Ekl_NGS17	37	4	4,000	0,719	1,000	-0,399
Ekl_NGS18	37	3	3,000	0,613	1,000	-0,646
Ekl_NGS27	37	4	4,000	0,715	0,919	-0,290
Ekl_NGS29	37	3	3,000	0,613	1,000	-0,646

Ekl_NGS30	37	1	1,000	0,000	0,000	NA
Ekl_NGS31	37	3	3,000	0,613	1,000	-0,646
Ekl_NGS32	37	3	3,000	0,613	1,000	-0,646
Ekl_NGS33	37	2	2,000	0,507	1,000	-1,000
Ekl_NGS34	37	2	2,000	0,507	1,000	-1,000
Ekl_NGS35	37	3	3,000	0,468	0,297	0,368*
Ekl_NGS38	37	2	2,000	0,507	1,000	-1,000
Ekl_NGS41	37	2	2,000	0,501	0,892	-0,800
Ekl_NGS42	37	2	2,000	0,462	0,703	-0,532
Média Geral	36,5	2,9	2,902	0,555	0,873	-0,586

n = número de indivíduos analisados; A = número de alelos; Ar = riqueza alélica; He = diversidade genética de Nei; Ho = heterozigosidade observada; f = coeficiente de endogamia total.

*Valor de f significativo

A qualidade dos locos e confirmação dos alelos atribuídos na planilha do teste T1 foi verificada a partir do alinhamento das sequências obtidas no editor *BioEdit*. A figura 10 exemplifica a análise do loco EKL_NGS10 com todos os alelos confirmados. As regiões com *gaps*, ilustram a formação de uma “escada alélica”, representando todos os alelos identificados na população analisada. As escadas alélicas forneceram uma referência adicional de tamanho para os alelos de cada loco, além daquela já determinada pelo *pipeline* utilizado. Além disso, com o alinhamento pode-se confirmar a região amplificada, por exemplo no caso do loco EKL_NGS30 que mostrou ser o único monomórfico, pois não apresentou alelos microssatélites de diferentes tamanhos e sequencias de bases. Dessa forma, a imagem obtida com o *BioEdit* confirmou que o loco é microssatélite, mas não possui polimorfismo.

A partir do alinhamento das sequências observou-se que o loco EKL_NGS5, possui duas regiões microssatélites amplificadas, apesar de o pipeline ter atribuído alelos referentes apenas uma dessas regiões (Figura 11, altura de alelos de aproximadamente 130 pb), foi identificada uma outra região de microssatélites com alelos claramente definidos (Figura 12, altura de alelos de aproximadamente 190 pb). Isso pode ser explicado pelo fato de que algumas vezes os pares de *primers* podem anelar ou se complementar a mais de uma região do genoma.

O *pipeline SSR-GBS* atribuiu alelos normalmente para o loco EKL_NGS35 e o sinalizou como portador de mutações pontuais. Porém a análise no *BioEdit* evidenciou que o par de *primers* não amplificou uma região microssatélite (Figura 13) e os genótipos deste loco devem ser desconsiderados.

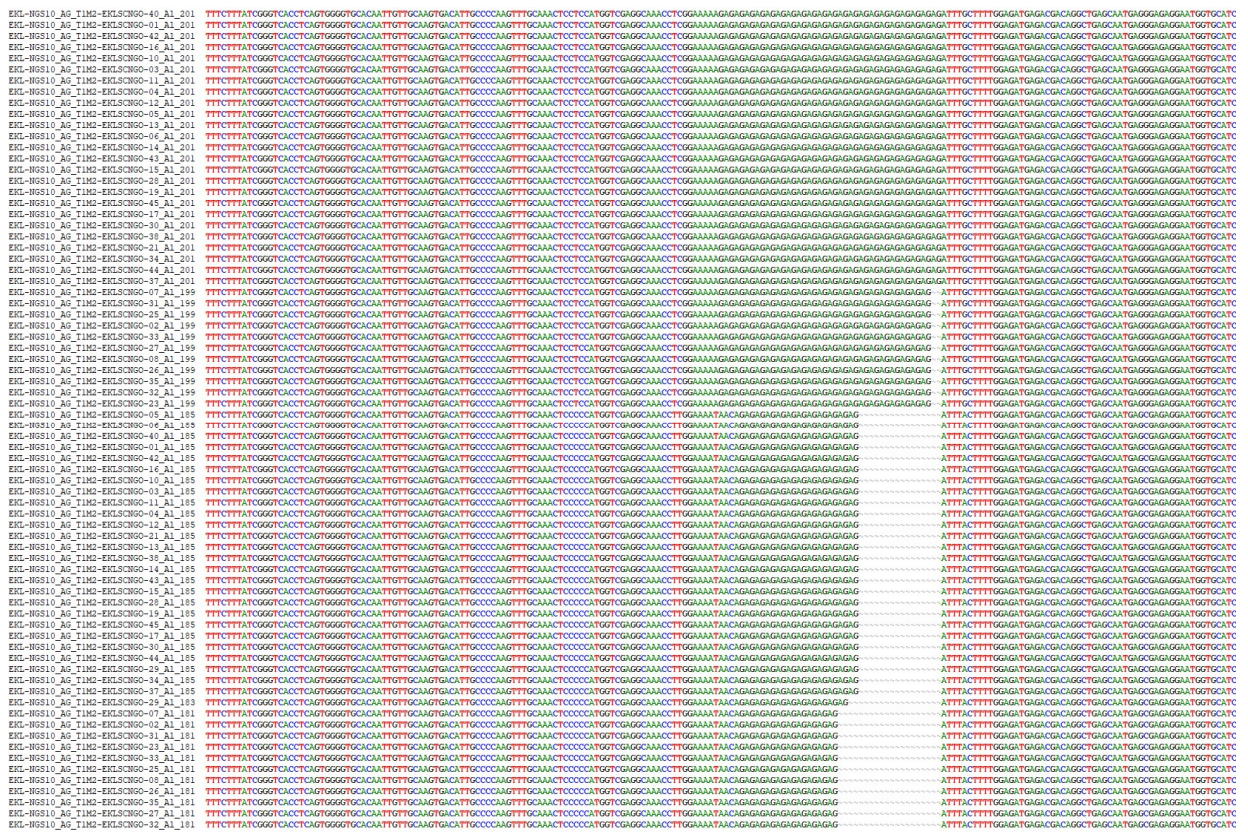


Figura 10. Alinhamento das sequências do loco de SSR-Seq, EKL_NGS10, para 37 indivíduos de *Eugenia klotzschiana*, por ordem de tamanho, formando uma “escada alélica” com as chamadas dos alelos e com a detecção dos SNPs dentro das regiões microssatélites (regiões com *gaps*).

A partir desse cenário e, considerando as análises de diversidade genética, os locos EKL_NGS30 e EKL_NGS35 devem ser descartados para compor a lista de melhores locos, para genotipagem por sequenciamento de *E. klotzschiana* (Tabela 7). Quanto às condições de amplificação, o Teste 1 (quatro multiplex com cinco primers cada, utilizando primers na concentração de 3 μ M e 54°C de temperatura de anelamento) apresenta as melhores condições para a amplificação e genotipagem satisfatória de pares de primers SSR-Seq em *E. klotzschiana*.

Um número maior de marcadores microssatélites fornecem resultados robustos de genética populacional (Capote et al., 2012; Rymanet al., 2006). Geralmente o número de locos microssatélites entre 8 e 20 são considerados adequadamente informativos para determinar a estrutura genética entre os indivíduos de uma espécie de acordo com os estudos desenvolvidos nos últimos anos (Arthofer et al., 2018; Vartia et al., 2014). Neste estudo são apresentados 20 pares de primers polimórficos e informativos para os estudos genético populacionais com *E. klotzschiana*.

Tabela 7. Melhores pares de primers para a genotipagem de SSR-Seq em *Eugenia klotzschiana*, com as respectivas propostas de sistemas de PCR multiplex, motivos de repetição, número e amplitude de tamanho de alelos obtidos.

Multiplex	Nome do <i>Primer</i> (3 μ M)	Motivo de repetição (Número de repetições)	<i>A</i>	Amplitude de tamanho de alelos
M1	Ekl_NGS2	AG ₁₂	3	187-197
	Ekl_NGS3	AG ₁₂	2	182-184
	Ekl_NGS5	AG ₁₃	4	129-137 e 187-201
	Ekl_NGS6	AG ₁₃	3	181-185
	Ekl_NGS9	AG ₁₄	4	187-201
M2	Ekl_NGS10	AG ₁₂	5	181-201
	Ekl_NGS12	AACG ₆	3	178-186
	Ekl_NGS13	AG ₁₄	2	160-172
	Ekl_NGS16	AG ₁₂	4	133-141
	Ekl_NGS17	AG ₁₃	4	193-205
M3	Ekl_NGS18	AG ₁₃	3	166-170
	Ekl_NGS27	AG ₁₂	4	132-144
	Ekl_NGS29	AG ₁₂	3	199-209
	Ekl_NGS31	AG ₁₄	3	191-200

	Ekl_NGS32	AG ₁₃	3	195-205
M4	Ekl_NGS33	AG ₁₂	2	198-200
	Ekl_NGS34	AG ₁₂	2	164-166
	Ekl_NGS38	AG ₁₄	2	124-126
	Ekl_NGS41	AG ₁₂	2	158-160
	Ekl_NGS42	AG ₁₂	2	191-197

A = número de alelos

A população de *E. klotzschiana* do município de Senador Canedo – GO, apresentou 6 grupos de indivíduos provavelmente clonais (Figura 14). Esse aspecto reforça alguns estudos anteriores que indicam que *E. klotzschiana* pode ser de reprodução assexuada (vegetativa ou apomítica) (Oliveira et al., 1999). Outros estudos, como os de Balloux et al. (2003), indicam que populações de plantas clonais podem manter ou até aumentar a diversidade genética, especialmente se os clones forem heterozigotos, como é o caso do presente estudo com a espécie em que evidenciou-se uma população majoritariamente heterozigota.

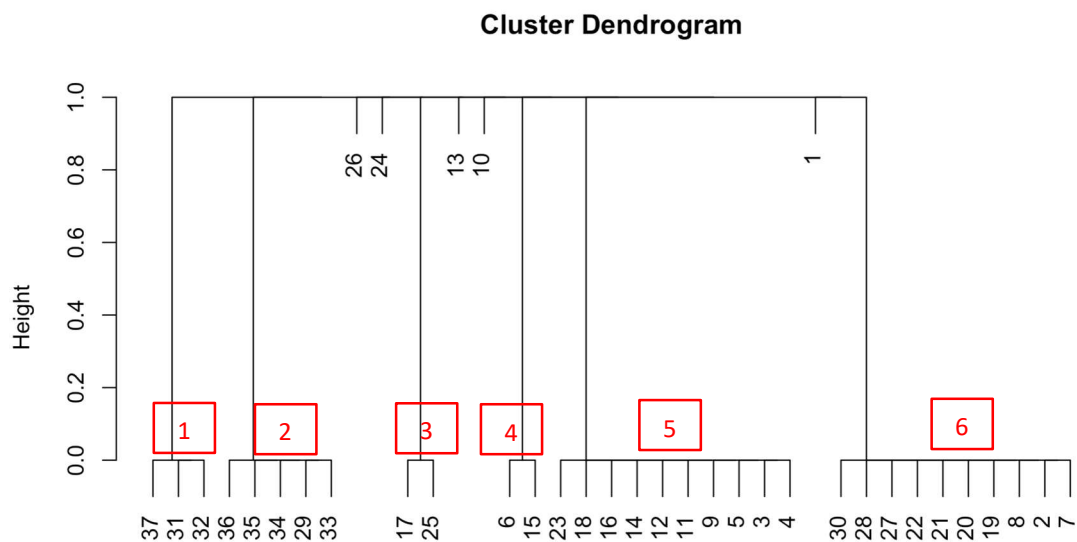


Figura 14. Dendrograma de similaridade clonal entre os 37 indivíduos de *Eugenia klotzschiana* encontrados no município de Senador Canedo – GO.

5. CONCLUSÕES

- As condições de amplificação do teste T1 (quatro multiplex com cinco primers cada, utilizando primers na concentração de 3 μ M e 54°C de temperatura de anelamento) são as mais adequadas para garantir a qualidade da genotipagem dos locos SSR-Seq desenvolvidos para *E. klotzschiana*;
- A partir dos resultados de diversidade genética e características das sequências microssatélites dos locos, foram descartados dois locos, ficando com uma bateria de locos com níveis adequados de polimorfismo e qualidade de sequências;
- Foram identificados seis grupos de indivíduos provavelmente clonais na população de *E. klotzschiana* avaliada.

6. REFERÊNCIAS

- ARAÚJO, A. M. DE. **Transferibilidade de Primers e Validação de SNPs em Genes Relacionados a Tolerância à Estresses Bióticos e Abióticos**. Universidade Federal de Goiás - UFG, 2020.
- ARTHOFER, W. et al. Identifying the minimum number of microsatellite loci needed to assess population genetic structure: A case study in fly culturing. **Fly**, v. 12, n. 1, p. 13–22, 2018.
- BALLOUX, F.; LEHMANN, L.; DE MEEÛS, T. The population genetics of clonal and partially clonal diploids. **Genetics**, v. 164, n. 4, p. 1635–1644, 2003.
- CAPOTE, N. et al. Molecular Tools for Detection of Plant Pathogenic Fungi and Fungicide Resistance. **Plant Pathology**, 2012.
- CARNEIRO, N. S. et al. Eugenia Klotzschiana O. Berg fruits as new sources of nutrients: Determination of their bioactive compounds, antioxidant activity and chemical composition. **Brazilian Archives of Biology and Technology**, v. 62, p. 1–11, 2019.
- CHUNG, Y. S. et al. Genotyping-by-sequencing: a promising tool for plant genetics research and breeding. **Horticulture Environment and Biotechnology**, v. 58, n. 5, p. 425–431, 2017.
- CORVALÁN, L. et al. Dados de primers SSRs para genotipagem por sequenciamento de alto rendimento (SSR-Seq) com base na montagem parcial do genoma de Eugenia klotzschiana (Myrtaceae). **Data in brief**, v. 47, 2023a.
- CORVALÁN, L. C. J. et al. Data of SSRs primers for high-throughput genotyping-by-sequencing (SSR-Seq) based on the partial genome assembly of Eugenia klotzschiana (Myrtaceae). **Data in Brief**, v. 47, 2023b.
- CURTO, M. et al. Application of a SSR-GBS marker system on investigation of European Hedgehog species and their hybrid zone dynamics. **Ecology and Evolution**, v. 9, n. 5, p. 2814–2832, 2019.
- DARBY, B. J. et al. Digital fragment analysis of short tandem repeats by high-throughput amplicon sequencing. **Ecology and Evolution**, v. 6, n. 13, p. 4502–4512, 2016.
- DE BARBA, M. et al. High-throughput microsatellite genotyping in ecology: improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. **Molecular Ecology Resources**, v. 17, n. 3, p. 492–507, 2017.
- EWING, B.; GREEN, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. **Genome Research**, v. 8, n. 3, p. 186–194, 1998.

- FALEIRO, F. G. **MARCADORES GENÉTICO-MOLECULARES**. Planaltina-DF: Embrapa, 2007.
- GONÇALVES, A. R. et al. Development and characterization of microsatellite markers in *Stryphnodendron adstringens* (Leguminosae). **Physiology and Molecular Biology of Plants**, v. 26, n. October, p. 2095–2101, 2020.
- GUICHOUX, E. et al. Tendências atuais em genotipagem de microssatélites. p. 591–611, 2011.
- HODEL, R. G. J. et al. The Report of My Death was an Exaggeration: A Review for Researchers Using Microsatellites in the 21st Century. **Applications in Plant Sciences**, v. 4, n. 6, p. 1600025, 2016.
- LEPAIS, O. et al. Fast sequence-based microsatellite genotyping development workflow. **PeerJ**, v. 2020, n. 3, p. 1–28, 2020.
- MMA, M. DO M. A. **Espécies Nativas da Flora Brasileira de Valor Econômico Atual ou Potencial: Plantas para o Futuro - Região Centro Oeste**. 1ª ed. ed. Brasília: [s.n.].
- NEOPHYTOU, C. et al. Analysis of microsatellite loci in tree of heaven (*Ailanthus altissima* (Mill.) Swingle) using SSR-GBS. **Tree Genetics and Genomes**, v. 14, n. 6, 2018.
- OLIVEIRA, G. C. DE et al. Caracterização de plantas de *Eugenia klotzschiana* Berg (pêra-do-Cerrado) e do ambiente de sua ocorrência na região fisiográfica dos campos das vertentes em Minas Gerais. **R. Un. Alfenas**, v. 5, n. October, p. 9–13, 1999.
- PEDRA, T. DA R. et al. Transferibilidade de primers microssatélites SSR para *Copernicia alba* Morong Ex & Britton (Carandá). **Peer Review**, v. 5, n. 16, p. 41–52, 2023.
- PORTELA, R. M. et al. **MARCADORES SSR: Aplicações em genética e melhoramento florestal**. 1ª ed. ed. Belém - PA: Home Editora, 2024.
- RIBEIRO, R.; RODRIGUES, F. Genética da conservação em espécies vegetais do cerrado. **Revista de Ciências Médicas e Biológicas**, v. 5, n. 3, p. 253–260, 2006.
- RYMAN, N. et al. Power for detecting genetic divergence: Differences between statistical methods and marker loci. **Molecular Ecology**, v. 15, n. 8, p. 2031–2045, 2006.
- SARDI, J. DE C. O. et al. Unexplored endemic fruit species from Brazil: Antibiofilm properties, insights into mode of action, and systemic toxicity of four *Eugenia* spp. **Microbial Pathogenesis**, v. 105, p. 280–287, 2017.
- ŠARHANOVÁ, P. et al. SSR-seq: Genotyping of microsatellites using next-generation sequencing reveals higher level of polymorphism as compared to traditional fragment size scoring. **Ecology and Evolution**, v. 8, n. 22, p. 10817–10833, 2018.

SIQUEIRA, M. N. DE. **TRANSFERIBILIDADE VARIABILIDADE GENÉTICA DE MARCADORES MICROSSATÉLITESGÊNICOS EM *Eugenia klotzschiana* BERG** UNIVERSIDADE FEDERAL DE GOIÁS INSTITUTO DE CIÊNCIAS BIOLÓGICAS PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR. [s.l.] Universidade Federal de Goiás - UFG, 2014.

TAHERI, S. et al. Mining and Development of Novel SSR Markers Using Next Generation Sequencing (NGS). **Molecules**, v. 23, p. 20p., 2018.

TIBIHIKA, P. D. et al. Application of microsatellite genotyping by sequencing (SSR-GBS) to measure genetic diversity of the East African *Oreochromis niloticus*. **Conservation Genetics**, v. 20, n. 2, p. 357–372, 2019.

VARTIA, S. et al. Multiplexing with three-primer PCR for rapid and economical microsatellite validation. **Hereditas**, v. 151, n. 2–3, p. 43–54, 2014.

VARTIA, S. et al. A novel method of microsatellite genotyping-by-sequencing using individual combinatorial barcoding. **Royal Society Open Science**, v. 3, n. 1, 2016.

ZHANG, J. et al. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. **Bioinformatics**, v. 30, n. 5, p. 614–620, 2014.

APÊNDICES

Apêndice A. Tabela da contagem dos erros de genotipagem do sequenciamento SSR-Seq de *Eugenia klotzschiana* com os 37 indivíduos para os 22 pares de *primers* nos testes multiplex T1, T2 e T3.

Quantificação de erros da genotipagem SSR-Seq				
Teste	Indivíduo	Tipo de Erro		
		MC	NA	PR
T1	EKLSCNGO-01	26	0	0
T2	EKLSCNGO-01	26	0	2
T3	EKLSCNGO-01	20	12	4
T1	EKLSCNGO-02	26	0	0
T2	EKLSCNGO-02	26	0	2
T3	EKLSCNGO-02	12	14	4
T1	EKLSCNGO-03	26	0	0
T2	EKLSCNGO-03	26	0	2
T3	EKLSCNGO-03	12	12	6
T1	EKLSCNGO-04	28	0	0
T2	EKLSCNGO-04	24	0	2
T3	EKLSCNGO-04	16	10	10
T1	EKLSCNGO-05	18	0	0
T2	EKLSCNGO-05	24	0	2
T3	EKLSCNGO-05	14	10	6
T1	EKLSCNGO-06	20	10	0
T2	EKLSCNGO-06	26	2	0
T3	EKLSCNGO-06	12	10	6
T1	EKLSCNGO-07	26	0	0
T2	EKLSCNGO-07	26	2	0
T3	EKLSCNGO-07	14	14	0
T1	EKLSCNGO-08	26	0	0
T2	EKLSCNGO-08	24	0	2
T3	EKLSCNGO-08	14	14	2
T1	EKLSCNGO-10	28	0	0
T2	EKLSCNGO-10	28	0	2
T3	EKLSCNGO-10	12	12	4
T1	EKLSCNGO-11	26	0	0
T2	EKLSCNGO-11	28	0	2
T3	EKLSCNGO-11	14	14	2
T1	EKLSCNGO-12	28	0	0
T2	EKLSCNGO-12	26	0	2
T3	EKLSCNGO-12	18	12	2
T1	EKLSCNGO-13	28	0	0
T2	EKLSCNGO-13	26	2	0
T3	EKLSCNGO-13	18	12	2
T1	EKLSCNGO-14	26	0	0
T2	EKLSCNGO-14	26	0	2
T3	EKLSCNGO-14	12	10	4
T1	EKLSCNGO-15	28	0	0
T2	EKLSCNGO-15	24	0	2
T3	EKLSCNGO-15	16	12	4

T1	EKLSCNGO-16	20	10	0
T2	EKLSCNGO-16	26	0	2
T3	EKLSCNGO-16	16	12	4
T1	EKLSCNGO-17	28	0	0
T2	EKLSCNGO-17	24	0	2
T3	EKLSCNGO-17	12	14	2
T1	EKLSCNGO-19	24	0	0
T2	EKLSCNGO-19	24	2	0
T3	EKLSCNGO-19	16	14	2
T1	EKLSCNGO-21	28	0	0
T2	EKLSCNGO-21	30	0	2
T3	EKLSCNGO-21	16	14	2
T1	EKLSCNGO-23	28	0	0
T2	EKLSCNGO-23	24	0	2
T3	EKLSCNGO-23	12	14	4
T1	EKLSCNGO-25	26	0	0
T2	EKLSCNGO-25	14	22	0
T3	EKLSCNGO-25	12	14	4
T1	EKLSCNGO-26	28	0	0
T2	EKLSCNGO-26	24	0	2
T3	EKLSCNGO-26	18	12	4
T1	EKLSCNGO-27	26	0	0
T2	EKLSCNGO-27	22	0	2
T3	EKLSCNGO-27	14	10	6
T1	EKLSCNGO-28	28	0	0
T2	EKLSCNGO-28	28	0	2
T3	EKLSCNGO-28	14	14	6
T1	EKLSCNGO-29	28	0	0
T2	EKLSCNGO-29	28	0	0
T3	EKLSCNGO-29	18	14	2
T1	EKLSCNGO-30	26	0	0
T2	EKLSCNGO-30	30	0	0
T3	EKLSCNGO-30	10	12	6
T1	EKLSCNGO-31	28	0	0
T2	EKLSCNGO-31	26	0	0
T3	EKLSCNGO-31	14	10	4
T1	EKLSCNGO-32	26	0	0
T2	EKLSCNGO-32	24	0	0
T3	EKLSCNGO-32	12	14	0
T1	EKLSCNGO-33	28	0	0
T2	EKLSCNGO-33	24	0	0
T3	EKLSCNGO-33	14	14	0
T1	EKLSCNGO-34	28	0	0
T2	EKLSCNGO-34	28	0	0
T3	EKLSCNGO-34	14	12	4
T1	EKLSCNGO-35	26	0	0
T2	EKLSCNGO-35	24	0	0
T3	EKLSCNGO-35	14	12	2
T1	EKLSCNGO-37	28	0	0
T2	EKLSCNGO-37	28	0	0

T3	EKLSCNGO-37	14	14	0
T1	EKLSCNGO-38	26	0	0
T2	EKLSCNGO-38	26	0	0
T3	EKLSCNGO-38	16	14	2
T1	EKLSCNGO-40	28	0	0
T2	EKLSCNGO-40	26	2	0
T3	EKLSCNGO-40	18	14	2
T1	EKLSCNGO-42	28	0	0
T2	EKLSCNGO-42	26	0	2
T3	EKLSCNGO-42	10	12	4
T1	EKLSCNGO-43	28	0	0
T2	EKLSCNGO-43	24	0	2
T3	EKLSCNGO-43	10	14	0
T1	EKLSCNGO-44	28	0	0
T2	EKLSCNGO-44	26	2	0
T3	EKLSCNGO-44	18	14	4
T1	EKLSCNGO-45	26	0	0
T2	EKLSCNGO-45	24	0	0
T3	EKLSCNGO-45	14	14	0

MC=checagem manual; NA=não amplificado; PR= poucas *reads*

Apêndice B. Função do *software* R para o cálculo da similaridade da chamada de alelos entre indivíduos de *Eugenia klotzschiana* para os testes T1, T2 e T3.

```
#COMPARAÇÃO NA CHAMADA DOS ALELOS ENTRE OS TESTES, GERAL E POR LOCUS
#Ler primeira a função perc_dist no final do script

#INPUT PLANILHAS GENOTIPOS, Ordenada por teste (dados perinha)
#library(vegan)

perinha <-as.matrix(read.table("Dados_perinha.txt",h=T))
dim(perinha)
loci <-perinha[,3:46]

#Variação individual total
#Teste
DIST <-as.matrix(perc_dist(loci))
indiv1 <-which(perinha[,2]==1)
DIST[indiv1,indiv1]

M.reduced <-as.dist(DIST[indiv1,indiv1],up=F,diag=F)
mean(M.reduced)
sd(M.reduced)

#Loop para calcular percentagem de similaridade total entre testes para todos os
indivíduos
D1 <-numeric()
D2 <-numeric()
D3 <-numeric()

for(i in 1:max(perinha[,2])){

  indiv <-which(perinha[,2]==i)
  DIST[indiv,indiv]

  M.reduced <-as.matrix(as.dist(DIST[indiv,indiv],up=T,diag=T))

  D1[i] <-M.reduced[1,2]
  D2[i] <-M.reduced[1,3]
  D3[i] <-M.reduced[2,3]

}

1 - D1 # % similaridade media entre T1 e T2 para todos os individuos
1 - D2 # % similaridade media entre T1 e T3 para todos os individuos
1 - D3 # % similaridade media entre T2 e T3 para todos os individuos

#medias
1 - mean(D1) # % similaridade media entre individuos entre T1 e T2, para todos os
loci
1 - mean(D2) # % similaridade media entre individuos entre T1 e T3, para todos os
loci
1 - mean(D3) # % similaridade media entre individuos entre T2 e T3, para todos os loci

#Por locus
DL1 <-numeric()
DL2 <-numeric()
DL3 <-numeric()

for(k in 1:22){

  #k <-10
  index <-c((k*2)-1,k*2)
  per.loci <-loci[,index]
```

```

DIST <-perc_dist(per.loci)
#DIST <-as.matrix(vegdist(per.loci,method="canberra",up=T,diag=T)) # dist
manhattan do vegan

D1 <-numeric()
D2 <-numeric()
D3 <-numeric()

for(i in 1:max(perinha[,2])){

  indiv <-which(perinha[,2]==i)
  DIST[indiv,indiv]

  M.reduced <-as.matrix(as.dist(DIST[indiv,indiv],up=T,diag=T)) #dist. euclidiana

  D1[i] <-M.reduced[1,2]
  D2[i] <-M.reduced[1,3]
  D3[i] <-M.reduced[2,3]

}
DL1[k] <-mean(D1,na.rm=T)
DL2[k] <-mean(D2,na.rm=T)
DL3[k] <-mean(D3,na.rm=T)
}

1 - DL1 # % similaridade por locus entre T1 e T2, para todos os individuos
1 - DL2 # % similaridade locus entre T1 e T3, para todos os individuos
1 - DL3 # % similaridade locus entre T2 e T3, para todos os individuos
mean(1 - DL1)
mean(1 - DL2)
mean(1 - DL3)

#FUNCAO PERC_DIST
#Function para calcular % similarity
perc_dist <-function(mat){

  n <-nrow(mat)
  p <-ncol(mat)
  pc <-matrix(0,n,n)

  dif <-numeric()

  for(i in 1:n){
    for(j in 1:n){
      for(k in 1:p){

        dif[k] <-ifelse(mat[i, k]==mat[j, k],0,1)

      }
      pc[i,j] <-sum(dif)/p
    }
  }

  return(as.matrix(pc))

} #end function

```