GENÉTICA FORENSE APLICADA A IDENTIFICAÇÃO GENÉTICA DE CLONES DE *Coffea canephora*

Aline dos Santos Bergamin, Viviane Alexia Correia Silva, Jônatas Gomes Santos, Anamaria Rissi Vargas, Pedro Silva de Araujo, Adésio Ferreira, Marcia Flores da Silva Ferreira.

Universidade Federal do Espírito Santo, Alto Universitário, S/N Guararema, 29500-000 - Alegre-ES, Brasil, aline.bergamin@edu.ufes.br, vivianeac.correia@gmail.com, gomesjonatas21@gmail.com, anamaria.vargas@edu.ufes.br, pedro.s.araujo@edu.ufes.br, adesio.ferreira@ufes.br, marcia.ferreira@ufes.br

Resumo
Este estudo utilizou 11 marcadores microssatélites (SSRs) para avaliar a identidade genética de 15 clones de *Coffea canephora* comercializados por um viveiro de Minas Gerais, comparando-os a um banco de dados de referência com 703 amostras. Foram identificados 68 alelos (média de 6,18 por locus), com alta informatividade (PIC médio 0,66) e probabilidade de identidade combinada muito baixa ($4{,}94 \times 10^{-11}$), confirmando o potencial discriminatório do painel. Apesar disso, apenas 13% dos clones, apenas as amostras Z13 e Z13 "sombrite" apresentaram correspondência esperada com o genótipo de referência, enquanto 87% não coincidiram. Os resultados reforçam a eficácia técnica do painel de SSRs para identificação e certificação clonal, mas apontam a necessidade de obtenção de maior análise de dados e, também, conhecimento de taxa de mutações somáticas na espécie, que possam auxiliar na padronização e critérios para aplicação prática em larga escala na cafeicultura.

Palavras-chave: Banco de dados. Cafeicultura. Microssatélites.
Área do Conhecimento: Ciências Agrárias – Agronomia

Introdução

A identificação genética correta de clones de *Coffea canephora* é de extrema importância para a cafeicultura, especialmente diante da crescente diversificação de materiais genéticos comercializados por viveiros em todo o país. A confusão entre clones, seja por erros de rotulagem, mistura de mudas ou comercialização irregular, pode comprometer diretamente a produtividade e a qualidade das lavouras, além de gerar prejuízos econômicos para os produtores que investem em materiais supostamente selecionados para características agronômicas específicas. Sendo assim, a construção de bancos de dados de perfis genéticos multillocus, surge como uma ferramenta estratégica, permitindo o armazenamento sistemático de informações moleculares que possibilitam a comparação, o rastreamento e a certificação da identidade genética de diferentes clones ao longo do tempo (Ferreira, & Grattapaglia 1998; Li et al 2025).

Dentre as ferramentas moleculares disponíveis para a caracterização genética de plantas, os marcadores microssatélites (SSRs - *Simple Sequence Repeats*), se destacam devido a sua elevada capacidade discriminatória, codominância, reprodutibilidade e ampla distribuição ao longo do genoma (Geethanjali et al. 2024). Essas características tornam esses marcadores adequados para a construção de perfis genéticos multilocus, nos quais a combinação de alelos detectados em múltiplos *loci* gera um padrão único capaz de individualizar cada genótipo analisado (Faria et al. 2011; Queiroz et al. 2023).

Em espécies de propagação vegetativa, como é o caso de *C. canephora*, essa individualização é ainda mais relevante, uma vez que a identidade clonal deve ser mantida ao longo de sucessivas multiplicações, tornando os SSRs uma ferramenta robusta para a verificação da autenticidade genética dos materiais propagados. A partir desses perfis, torna-se possível não apenas confirmar ou refutar a identidade varietal declarada pelos viveiros, mas também estabelecer uma base de dados de referência que sirva como ferramenta contínua de controle de qualidade genética na cadeia produtiva da espécie *C. canephora*. A comparação entre os perfis multilocus obtidos e aqueles previamente catalogados permite identificar duplicações, sinonímias ou possíveis trocas de material entre clones, contribuindo para a rastreabilidade e a confiabilidade dos materiais propagados.

Diante deste contexto, o presente trabalho teve como objetivo utilizar um banco de dados de perfis multilocus de *C. canephora*, já desenvolvido pelo grupo de pesquisa, avaliando seu potencial para ser destinado à identificação e certificação genética desses materiais.

Metodologia

Material vegetal e extração de DNA genômico

Foram utilizadas 15 amostras de diferentes clones de *Coffea canephora*: A1, Pirata, Z13, JC, 17, R22, R8, LB88, R25, R07, LB15, Z13 "sombrite", R04, LB10, AS2 e R156 oriundas para análise de um viveiro do estado de Minas Gerais. Todos os materiais foram identificados e armazenadas em sacos de papel no freezer a -80 ºC por um período de 24 horas. Posteriormente, as amostras foram liofilizadas por aproximadamente 72 horas e maceradas utilizando o MagnaLyser.

A extração foi realizada utilizando folhas jovens e saudáveis, seguindo o protocolo CTAB, descrito por Doyle e Doyle (1990), adaptado por Zaidan et al. (2023) para *C. canephora*. O DNA isolado foi quantificado e avaliado quanto a pureza por espectrofotometria utilizando NanoDrop One UV-Vis, considerando ideal a razão absorbância de 260 a 280 nm (A260/A280) e pureza entre 1.8 e 2.0, e a integridade do DNA avaliada por eletroforese em gel de agarose a 0.8%. As amostras foram diluídas para 20 ng.μL^{-1} e armazenadas a -20 °C. O experimento foi realizado no Laboratório de Genética e Melhoramento e Laboratório de Biometria da Universidade Federal do Espírito Santo, no campus Alegre.

Análise de fragmentos por eletroforese capilar

Para a análise de similaridade e obtenção dos perfis genéticos multilocus por amostra, foram utilizados 11 microssatélites (SSRs), marcados com fluoróforos na extremidade 5' do marcador forward e selecionados com base na ótima amplificação e genotipagem consistente, picos definidos, alto sinal fluorescente, distinção clara entre alelos e baixo stutter, a saber: CAM 16 (TC)11, CAM 22 (AT)15, CaEST-002 (-), CCRM17 (CT)17, SSR34 (AC)10 (CA)6, CAM18 (TC)9, CAM32 (TA)12, Mg_257 (CA)9, CC2P4 (AT)22, CCRM16 (TC) 26, CCRM33 (CT)12 (Silva 2026).

Para as PCRs utilizou-se o protocolo da Sinapse Biotecnologia com volume final de 20 µL, contendo por reação: 20 ng.µL-1 de DNA, 1X PCR Buffer Mg2+ plus, 1.25 U de Taq DNA Polimerase (Sinapse Biotecnologia), 0.25 µM dos primers forward/reverse e 0.2 mM de dNTP. As amplificações foram feitas no termociclador VeritiTM 96-Well Thermal Cycler (Thermo Fisher Scientific) sob as seguintes condições: um ciclo de três minutos a 94°C; 35 ciclos de 40 segundos a 94 °C, um minuto a temperatura de anelamento entre 51 e 61°C (de acordo com a TA do primer), um minuto a 72 °C; seguido de extensão final de 72 °C por 30 minutos (Silva 2026). O anelamento dos primers, tamanho e amplificação dos fragmentos, foram confirmados utilizando os produtos das reações de PCR em gel de agarose a 1.2%, corado com GelRed, a 80V por 40 minutos.

Os produtos da PCR foram organizados em bulks de acordo com a fluorescência e tamanho de fragmentos dos marcadores, e submetidos à análise de fragmentos por eletroforese capilar, conduzida no sequenciador automático SeqStudio Genetic Analyzer (Thermo Fisher Scientific). Foram transferidos para a placa do sequenciador: 1 uL do bulk, juntamente com 0.25 µL do size standart GeneScan 600 LIZ (Thermo Fisher Scientific) e 8.75 µL de formamida Hi-Di (Thermo Fisher Scientific). O mix, no bulk foi desnaturado a 95 °C por 3 minutos no termociclador Veriti 96-Well Thermal Cycler, e após a desnaturação, as amostras foram resfriadas no gelo por cerca de 5 minutos e, em seguida, submetidas à eletroforese capilar.

Análises estatísticas

A identificação dos alelos e a visualização dos eletroferogramas foi realizada no software GeneMapper V5.0 (SoftGenetics LLC, State College, PA, EUA). O perfil genético multilocus das amostras foi obtido por genotipagem dos loci SSRs e todos os alelos registrados automaticamente pelo software foram verificados manualmente para corrigir possíveis erros de processamento. Os dados resultantes foram organizados em uma tabela, diferenciando cada viveiro ao qual as amostras pertenciam. A identificação dos alelos e a visualização dos eletroferogramas foi realizada no software GeneMapper V5.0 (SoftGenetics LLC, State College, PA, EUA). Foram estimados os parâmetros para cada SSR no software GenAlEx v. 6.503 (Peakall; Smouse, 2012): número de alelos (Na), Ho, He e a probabilidade de identidade (PI). O conteúdo de informação polimórfica (PIC) foi calculado no Cervus (Kalinowski et al., 2007). O software GMDA (Stefenon et al. 2026) foi utilizado para avaliar a similaridade

entre os genótipos oriundos do viveiro em relação a um banco de dados composto por 703 amostras de *C. canephora*. Ainda no mesmo software, foi estimado parâmetros genéticos básicos multilocus para cada genótipo, sendo: A: número total de alelos e a diversidade genética, medida como a heterozigosidade observada (Ho). A nível populacional, foi estimado A, Ho, número efetivo de alelos (Ae) e o índice de fixação (F).

Resultados

Para o conjunto de 11 SSRs, foram identificados 68 alelos, com média de 6,18 por *locus* (Tabela 1). O número de alelos variou de três (Mg257) a oito (CaEST002). A Ho apresentou valor máximo (Ho = 1,00) para os SSRs CAM18, CAM22, CaEST002, CC2P4, CAM16 e CCRM33. O menor valor foi observado no SSR Mg257 (0,29) e a média entre os loci foi de 0,82. Por sua vez, a variou de 0,36 (Mg257) a 0,84 (CaEST002), com média de 0,69. O F em geral apresentou valores negativos para maioria dos *loci*, o que concorda com os altos valores de Ho. O PIC variou de 0,32 (Mg257) a 0,82 (CaEST002), com média de 0,66. Dos 11 SSRs, 10 apresentaram PIC superior a 0,5. A PI esteve entre 0,05 (CCRM17) e 0,44 (Mg257), com média de 0,15 e PI combinada entre os 11 SSRs de $4,94 \times 10^{-11}$.

Tabela 1 – Caracterização dos 11 SSRs avaliados em 15 clones de *Coffea canephora*. Na: número de alelos; Ho: heterozigosidade observada; He: heterozigosidade esperada; PIC: conteúdo de informação polimórfica.

Locus	**Na**	**Ho**	**He**	**F**	**PIC**	**PI**
CAM18	4	1,00	0,70	-0,42	0,65	0,14
CAM22	6	1,00	0,80	-0,25	0,77	0,07
SSR34	6	0,75	0,57	-0,31	0,54	0,21
CCRM17	9	0,92	0,83	-0,10	0,81	0,05
CC2P4	7	0,54	0,78	0,31	0,75	0,08
CaEST002	8	1,00	0,84	-0,20	0,82	0,05
CAM32	6	0,83	0,75	-0,12	0,71	0,10
CAM16	5	1,00	0,67	-0,50	0,62	0,16
CCRM16	7	0,64	0,52	-0,24	0,50	0,25
CCRM33	7	1,00	0,78	-0,29	0,74	0,08
Mg257	3	0,29	0,36	0,20	0,32	0,44
Média	6,18	0,82	0,69	-0,17	0,66	0,15

Fonte: elaborado pelo próprio autor.

A nível populacional (Tabela 2), os parâmetros genéticos para os 15 clones avaliados revelaram número total de alelos entre 11 e 21 alelos, com média de 17,27, enquanto o número médio de alelos efetivos foi de 16,11. A Ho variou de 0,71 a 1,00, com média de 0,81. A He apresentou média de 0,93, resultando em um índice de fixação médio de 0,131.

A comparação dos clones oriundos do viveiro, utilizados como amostras de consulta, com o banco de dados de referência permitiu estimar a similaridade genética entre os materiais avaliados. Os valores de similaridade entre as amostras de consulta e suas referências mais similares variaram de 80,54%, observado para LB10 e B30/Z13, e 100% observado para Z13 “sombrite” e B30/Z13 (amostra de muda pertencente a um viveiro do município Águia Branca do estado do Espírito Santo). Apenas 13% dos clones (dois de 15), sendo eles Z13 e Z13 “sombrite”, apresentaram o genótipo de referência esperado entre as três maiores similaridades. Já os 87% restante dos clones não apresentaram correspondência com o genótipo de referência esperado.

Tabela 2 – Três maiores similaridades genéticas entre os 15 clones do viveiro e amostras do banco de dados de referência. A: número de alelos e heterozigosidade observada (Ho).

Clone	**Top 3 similaridade (%)**	**A**	**Ho**
A1	B30/Z13 (84,33%) A36/R22 (83,06%) A36/143 (80,70%)	15	0,88
Pirata	A6/R8 (90,07%) A21/AS2 (90,02%) B2/AS2 (88,95%)	20	0,82
Z13	B30/Z13 (99,31%) A12/LB15 (95,91%)	20	0,82

	B38/A1 (91,94%)		
17	B31/R25 (97,95%) B35/R25 (97,95%) C4/R25 (97,81%)	11	0,83
R22	41/Z37 (99,20%) A8/Z13 (98,00%) A38/R8 (97,92%)	12	0,71
R8	B30/Z13 (96,78%) A3/CM1 (85,04%) B37/Z13 (84,54%)	19	0,90
LB88	A12/LB15 (88,43%) A3/Z13 (88,26%) A15/Z13 (88,26%)	17	0,70
R25	A36/R22 (88,13%) B36/R22 (86,38%) C1/R156 (84,42%)	20	0,82
R07	A36/R22 (97,01%) B36/R22 (97,01%) A6/R8 (96,30%)	14	1,00
LB15	B30/Z13 (82,59%) A36/R22 (80,30%) C1/R156 (77,31%)	20	0,82
Z13 “sombrite”	B30/Z13 (100,00%) B37/Z13 (100,00%) A20/R22 (97,92%)	12	0,71
R04	B30/Z13 (99,74%) B37/Z13 (99,56%) 25/Imbigudinho (96,25%)	19	0,73
LB10	B30/Z13 (80,54%) A36/143 (79,10%) A21/AS2 (78,95%)	19	0,73
AS2	B30/Z13 (94,01%) 25/Imbigudinho (89,00%) A12/LB15 (86,91%)	21	0,91
R156	B30/Z13 (96,47%) A12/LB15 (96,26%) A23/LB80 (93,02%)	20	0,82
Média	-	17,27	0,81

Fonte: elaborado pelo próprio autor.

Discussão

O conjunto de SSRs foi altamente poplimórfico, evidenciada pela variabilidade genética detectada nos materiais avaliados. A Ho foi em média, maior que a He, o que demonstra uma predominância de heterozigotos no conjunto avaliado e concorda com os valores de F negativos obtidos para maior parte dos SSRs. A alta diversidade alélica identificada nos clones contribuiu para a discriminação de clones geneticamente próximos, como evidenciado pelo PIC, sendo que dez dos 11 SSRs foram altamente informativo (PIC >0,5), segundo critério definido por Botstein et al. 1980. A elevada informatividade dos SSRs reflete diretamente na PI, que apresentou um valor combinado extremamente baixo (PIcomb = $4,94 \times 10^{-11}$), o que indica uma probabilidade baixa de que dois genótipos não relacionados apresentarem o mesmo perfil multilocus ao acaso.

A baixa taxa de correspondência das amostras (87% sem match esperado), pode ser devido a erros durante a propagação das mudas dentro do viveiro, seja pela mistura de material clonal ou etiquetagem incorreta, que se trata de um problema comum e já relatado entre viveiristas e produtores de café. Além disso, uma das limitações do trabalho até então, é ausência de uma taxa de variação ou porcentagem

aceitável devido al alta taxa de propagação vegetativa e possibilidade de ocorrência de mutação somática, ainda não mensurada.

Existem na literatura, trabalhos sobre o uso de microssatélites na identificação genética de espécies como *Tectona grandis*, em que análises de parentesco e distância genética revelaram identidades inesperadas ou ancestralidade compartilhada recente significativa entre clones considerados únicos (Queiroz et al. 2023). Já em *Eucalyptus*, observaram que os marcadores SSR, demonstraram boa resolução para identificação individual, distinção de espécies e testes de atribuição individual para algumas das principais espécies plantadas em todo o mundo (Faria et al. 2011). Todos esses resultados reforçam a importância da inclusão de marcadores de DNA para certificação varietal em café e outras culturas perenes.

A espécie *C. canephora* é uma cultura de grande importância para o Espírito Santo, o que reforça a necessidade de controle quanto a identificação de clones durante a produção de mudas, e o quanto isso pode impactar a produtividade e sustentabilidade da cadeia produtiva da espécie. Ademais, a ampliação do painel de marcadores para aumentar a PI combinada e a resolução entre clones proximamente relacionados, assim como a expansão do banco de dados de referência dos materiais, podem auxiliar na aplicação do protocolo em maior escala como ferramenta de certificação oficial de viveiros.

Conclusão

Portanto, esses resultados demonstram a eficiência do painel para identificação e diferenciação de clones *C. canephora*, como visto para os genótipos de Z13, e mesmo com a baixa taxa de correspondência genotípica exata, o estudo demonstrou que o banco de dados e o painel de SSRs têm potencial técnico (alta PIC, baixa PI combinada) para uso em identificação e certificação, mas sua efetividade prática depende de ampliação da base de referência.

Referências

BOTSTEIN, D.; WHITE, R. L.; SKOLNICK, M.; DAVIS, R. W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **American Journal of Human Genetics**, v. 32, n. 3, p. 314–331, 1980.

Doyle, J. J. & Doyle, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. **Focus**, v. 12, p. 13-15, 1990.

Faria, D. A., Mamani, E. M. C., Pappas Jr, G. J. & Grattapaglia, D. Genotyping systems for Eucalyptus based on tetra-, penta-, and hexanucleotide repeat EST microsatellites and their use for individual fingerprinting and assignment tests. **Tree Genetics & Genomes**, v. 7, n. 1, p. 63-77, 2011. https://doi.org/10.1007/s11295-010-0315-9

Ferreira, M. E. & Grattapaglia, D. Introdução ao uso de marcadores moleculares em análise genética. 3ª ed. Brasília: EMBRAPA-CENARGEN, 1998.

Geethanjali, S., Kadirvel, P., Anumalla, M., Hemanth Sadhana, N., Annamalai, A. & Ali, J. Streamlining of simple sequence repeat data mining methodologies and pipelines for crop scanning. **Plants**, v. 13, n. 18, p. 2619, 2024. https://doi.org/10.3390/plants13182619

Kalinowski, S. T.; Taper, M. L. & Marshall, T. C. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. **Molecular Ecology**, v. 16, n. 5, p. 1099-1106, 2007.

Li X., Hu Z., Yu W., Xie H., Wang X., Huang P., Zhang X., Yang J., Li Y., Zhao W., Yang X., Zhang Z., Pu W. & Li Z. Advances and challenges in plant molecular marker technologies and their applications in the artificial intelligence empowered era. **Frontiers in Plant Science**, v. 16, p. 1757949, 2025. https://doi.org/10.3389/fpls.2025.1757949

Peakall, R. & Smouse, P. E. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. **Bioinformatics**, v. 28, n. 19, p. 2537–2539, 2012.

Queiroz, M. D. M., Arriel, D. A., Caldeira, S. F., Higa, A. R., Araújo, S. P., Sebbenn, A. M. & Grattapaglia, D. Microssatellite markers for plant variety protection of clonally propagated forest trees: a case study in teak (*Tectona grandis* Lf). **Silvae Genetica**, v. 72, n. 1, p. 189-199, 2023. https://doi.org/10.2478/sg-2023-0019

Silva, V. A. C. *Fingerprinting* em *Coffea canephora* Pierre ex Froehner: seleção e validação de microssatélites e obtenção de banco de perfis de DNA para identificação clonal. 2026. 126p. Dissertação (Mestrado em Genética e Melhoramento) - Universidade Federal do Espírito Santo, Alegre - ES, 2026.

Stefenon, V. M., Poletto, T., Girardello, G. M., & Thalmayr, P. N. Selection of pecan (*Carya illinoinensis*) genotypes using GMDA: Software for genetic and morphometric data analyses. Crop Breeding and Applied Biotechnology, v. 26, n. 2, p. e54342625, 2026.

Zaidan, I. R., Ferreira, A., Noia, L. R., Santos, J. G., de Arruda, V. C., do Couto, D. P. & da Silva Ferreira, M. F. Diversity and structure of *Coffea canephora* from old seminal crops in Espírito Santo, Brazil: genetic resources for coffee breeding. **Tree Genetics & Genomes**, v. 19, n. 2, p. 19, 2023. https://doi.org/10.1007/s11295-023-01594-x

Agradecimentos

O presente trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa e Inovação do Espírito Santo (Fapes)/SEAG – Processo: 281/2024 - P: 2024-88LGZ. Os autores agradecem ao Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (Incaper), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), à Secretaria de Estado da Agricultura, Abastecimento, Aquicultura e Pesca (SEAG), à Universidade Federal do Espírito Santo (UFES) e aos Laboratórios de Biometria e de Genética da UFES por todo o apoio.