



**FERNANDA DE ARAÚJO CARNEIRO**

**VALIDAÇÃO DE UMA PLATAFORMA DE GENOTIPAGEM  
(COFFEE AXIOM CHIP – 26K) COM APLICAÇÃO EM  
SELEÇÃO GENÔMICA AMPLA E GWAS, EM *Coffea  
canephora***

**LAVRAS – MG  
2019**

**FERNANDA DE ARAÚJO CARNEIRO**

**VALIDAÇÃO DE UMA PLATAFORMA DE GENOTIPAGEM (COFFEE AXIOM  
CHIP – 26K) COM APLICAÇÃO EM SELEÇÃO GENÔMICA AMPLA E GWAS, EM  
*Coffea canephora***

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Biotecnologia Vegetal, área de concentração em Análise Genômica e Funcional, para a obtenção do título de Doutor.

Dr. Alan Carvalho Andrade  
Orientador

**LAVRAS - MG  
2019**

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca  
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Carneiro, Fernanda de Araújo.

Validação de uma plataforma de genotipagem (Coffee Axiom  
Chip-26K) com aplicação em seleção genômica ampla e GWAS,  
em *Coffea canephora* / Fernanda de Araújo Carneiro. - 2019.

172 p. : il.

Orientador(a): Alan Carvalho Andrade.

Tese (doutorado) - Universidade Federal de Lavras, 2019.  
Bibliografia.

1. *Coffea canephora*. 2. Genotipagem em larga escala-SNP. 3.  
GWAS-Seleção Genômica. I. Andrade, Alan Carvalho. II. Título.

**FERNANDA DE ARAÚJO CARNEIRO**

**VALIDAÇÃO DE UMA PLATAFORMA DE GENOTIPAGEM (COFFEE AXIOM  
CHIP – 26K) COM APLICAÇÃO EM SELEÇÃO GENÔMICA AMPLA E GWAS, EM  
*Coffea canephora***

**VALIDATION OF A GENOTYPING PLATFORM (COFFEE AXIOM CHIP – 26K)  
WITH APPLICATION IN GENOMIC SELECTION AND GWAS IN *Coffea canephora***

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Biotecnologia Vegetal, área de concentração em Análise Genômica e Funcional, para a obtenção do título de Doutor.

APROVADA em 28 de fevereiro de 2019.

Dr(a) Flávia Maria Avelar Gonçalves

UFLA

Dr. Márcio Balestre

UFLA

Dr. Pierre Roger Rene Marraccini

CIRAD/UMR IPME

Dr. Orzenil Bonfim da Silva Júnior

Embrapa Recursos Genéticos e Biotecnologia



Dr. Alan Carvalho Andrade  
**Orientador**

**LAVRAS – MG  
2019**

## AGRADECIMENTOS

À Universidade Federal de Lavras, em especial ao Programa de Biotecnologia Vegetal, pela oportunidade oferecida.

A CAPES, pela concessão da bolsa de doutorado, e juntamente com o CNPq, INCT-Café e ao Consórcio Pesquisa Café pelo apoio financeiro para que esse trabalho fosse realizado.

Agradeço ao meu orientador, Alan Carvalho Andrade, por seu apoio, paciência e incentivo ao longo desse tempo, que inclui mestrado e doutorado.

Não poderia deixar de agradecer ao Pierre Marraccini que me acompanhou ao longo desse trabalho, que muito me ensinou e auxiliou.

Esse trabalho começou há muito tempo, as avaliações em campo tiveram início em 2012, e de lá até aqui muitas pessoas me ajudaram. Sendo assim, para não correr o risco de esquecer alguém, quero agradecer de coração a todos da Embrapa Cerrados, aos amigos do antigo Laboratório de Genética Molecular, quando ainda era localizado na Embrapa Recursos Genéticos e Biotecnologia, em Brasília, e aos amigos do novo Laboratório de Genética Molecular, agora situado na InovaCafé/UFLA. Muito obrigada mesmo, a ajuda de cada um de vocês tornou esse caminhar mais leve e prazeroso.

Gostaria de agradecer ao Gustavo Costa Rodrigues e a Milene Alves de Figueiredo Carvalho pela disponibilidade, boa vontade e disposição em participarem das avaliações de potencial hídrico.

Outros que deram seu tempo, conselhos e apoio para me ajudar a concluir esta tese incluindo o prof. Márcio Balestre (DES/UFLA), Indalécio e Guilherme.

Por último, mas não menos importante, agradeço a toda a minha família, especialmente aos meus pais, Osvaldo e Luzia, as minhas irmãs, Juliana e Renata, aos meus sobrinhos (Marcos Neto, Mariana, Cadu e Maitê) e ao meu companheiro Alan, que sempre estiveram ao meu lado, incentivando, apoiando e escutando. Amo vocês!!!

## RESUMO

O cafeeiro é uma espécie perene, de ciclo longo e, portanto, os programas de melhoramento genético convencional dessa espécie são demorados (vários anos) e de custo elevado. A geração de cultivares superiores via melhoramento genético ainda lida com o desafio de agregar simultaneamente, diversas características quantitativas de relevância agrônômica e de qualidade de bebida. As plataformas de genotipagem em larga escala e, em especial, os chips de DNA, possibilitaram a identificação de uma alta densidade de SNPs cobrindo todo o genoma, e aliado a isso, os avanços recentes na genômica do cafeeiro, como o sequenciamento do genoma de referência de *C. canephora*, proporcionaram os recursos necessários para o avanço do melhoramento da espécie por seleção genômica e descoberta de genes por GWAS, embora os estudos no gênero *Coffea* utilizando as metodologias de genotipagem em larga escala ainda sejam poucos. O desenvolvimento e a validação de um chip de SNP 26K Axiom para *C. canephora* foi possível, cobrindo a maior parte da diversidade genética conhecida da espécie, incluindo também uma grande parte dos genótipos pertencentes aos principais programas de melhoramento do Brasil. De um total de 25.456 SNPs presentes no chip, 25.411 SNPs foram convertidos com sucesso e mais de 85% classificados como polimórficos. Unindo os dados de genotipagem em larga escala, utilizando o chip desenvolvido, e os dados fenotípicos gerados para 1.319 indivíduos de *C. canephora*, cultivados na Embrapa Cerrados (Planaltina-DF), realizou-se nesse trabalho: (i) estudo de associação genômica ampla (GWAS) e (ii) seleção genômica ampla para diversas características importantes no melhoramento do cafeeiro. A GWAS tem sido amplamente utilizada na análise genética de características complexas e a SG tem sido estudada para aumentar o ganho genético e reduzir a duração dos ciclos de reprodução, especialmente em se tratando de plantas perenes como é o cafeeiro. Uma grande quantidade de marcadores associados foi identificada na GWAS para diversas características, sugerindo a utilidade dos SNPs identificados no chip como um recurso potencial para a identificação de alvos genômicos e de genótipos superiores que seriam fontes de genes de interesse. As análises de SG foram realizadas com o método G-BLUP e estimaram-se as herdabilidades e as acurácias. As características avaliadas nesse estudo apresentaram alta complexidade, pois são controladas por diversos genes, e baixa herdabilidade. A SG é uma abordagem promissora e inovadora a ser aplicada para o melhoramento de *C. canephora*. Na prática, comparando com a avaliação fenotípica tradicional, espera-se que GWAS e SG acelerem o ciclo de reprodução, mantenham a diversidade genética e aumentem o ganho genético por unidade de tempo, uma vez que indivíduos elite seriam selecionados em fase de muda.

Palavras-chave: *Coffea canephora*. GWAS. Seleção Genômica. Chip de genotipagem de DNA. SNP.

## ABSTRACT

Coffee is a perennial, long-cycle species and, therefore, conventional breeding programs of this species are time-consuming (several years) and costly. The generation of superior cultivars via genetic improvement still deals with the challenge of simultaneously aggregating several quantitative characteristics of agronomic relevance and the generation of superior cultivars via genetic improvement still deals with the challenge of simultaneously aggregating several quantitative characteristics of agronomic relevance and beverage quality. Large-scale genotyping platforms, and in particular DNA arrays, have enabled the identification of a high density of SNPs covering the entire genome, and in addition, recent advances in coffee genomics, such as the sequencing of the *C. canephora* reference genome, provided the necessary resources for the advancement of breeding of the species by genomic selection and gene discovery by GWAS, although *Coffea* genus studies using high-throughput genotyping methodologies are still few. The development and validation of a *C. canephora* 26K Axiom SNP array was possible, covering most of the known genetic diversity of the species, including also a large part of the genotypes belonging to the main breeding programs in Brazil. Of a total of 25,456 SNPs present on the array, 25,411 SNPs were successfully converted and more than 85% classified as polymorphic. Joining the high-throughput genotyping data, using the developed array, and the phenotypic data generated for 1,319 *C. canephora* individuals, cultivated at Embrapa Cerrados (Planaltina-DF), was carried out in this work: (i) genomic wide association study (GWAS) and (ii) genomic selection for several important characteristics in coffee breeding. GWAS has been widely used in the genetic analysis of complex traits and SG has been studied to increase genetic gain and reduce the duration of breeding cycles, especially in the case of perennial plants such as coffee. A large number of associated markers were identified in the GWAS for several characteristics, suggesting the usefulness of SNPs identified on the array as a potential resource for the identification of genomic targets and of higher genotypes that would be sources of genes of interest. GS analyzes were performed by the G-BLUP method and heritabilities and accuracy were estimated. The characteristics evaluated in this study presented high complexity, since they are controlled by several genes, and low heritability. GS is a promising and innovative approach to be applied for *C. canephora* breeding. In practice, in comparison with traditional phenotypic evaluation, GWAS and SG might now be employed to foster and accelerate the breeding programmes of *C. canephora* with a very good forecast of success in addition might maintain genetic diversity and increase the gain genetic selection per unit of time.

Keywords: *Coffea canephora*. GWAS. Genomic Selection. DNA array. SNP.

## LISTA DE FIGURAS

Figura 1 - História evolutiva do alotetraplóide <i>C. arabica</i> . .....	20
Figura 2 – Evolução da produção brasileira de café beneficiado para o período de 2003 a 2019. A produção para 2019 é uma estimativa realizada em janeiro/2019. ....	21
Figura 3 – Produção de <i>C. canephora</i> Conilon no Brasil. A produção para 2019 é uma estimativa realizada em janeiro/2019. ....	22
Figura 4 – Origem geográfica dos principais grupos genéticos de <i>C. canephora</i> . ....	24
Figura 5 – Árvore filogenética ( <i>Neighbor-joining</i> ) com base em uma matriz de dissimilaridade entre indivíduos de <i>C. canephora</i> . ....	26
Figura 6 – Diagrama esquemático mostrando o mapeamento de associação de todo o genoma e o mapeamento em genes candidatos. ....	32
Figura 7 – Indivíduos de <i>C. canephora</i> Conilon estabelecidos em um campo experimental da Embrapa Cerrados, localizado em Planaltina/DF. ....	47
Figura 8 – Resultado da análise de tamanho e formato de frutos de café cereja. ....	49
Figura 9 – Peneira com crivos circulares, para separar os grãos tipo chato (Figura 9A) e peneira com crivos oblongos, para separar os grãos tipo moca (Figura 9B). ....	50
Figura 10 – Análise exploratória para a característica de produção. ....	52
Figura 11 – Análise exploratória para as características Boia e Peso de 100 grãos (P100). ....	53
Figura 12 – Análise exploratória para as características de Morfologia dos frutos (Área, Perímetro, Eixo Maior e Eixo Menor). ....	54
Figura 13 – Análise exploratória para a característica de tamanho e formato dos grãos. ....	55
Figura 14 – Análise exploratória para a característica de potencial hídrico foliar de antemanhã. ....	55
Figura 15 – Potencial hídrico foliar de antemanhã para as 366 plantas de <i>C. canephora</i> selecionadas. ....	57
Figura 16 – Fenótipos contrastantes pertencentes à população de melhoramento após um período de déficit hídrico (Embrapa Cerrados). ....	58
Figura 17 – Produção (medida em litros de cereja por planta), para os 20 indivíduos que apresentaram maior média, nos anos de 2012 (linha azul), 2013 (linha vermelha) e 2014 (linha verde). ....	59
Figura 18 – Produção (medida em litros de cereja por planta), para os 30 indivíduos que apresentaram maior média, nos anos de 2012 (barras azuis), 2013 (barras vermelhas) e 2014 (barras verdes). ....	60



Figura 19 – Produção, em litros (L), e potencial hídrico foliar de antemanhã ( $\Psi_{am}$ ) de plantas de <i>C. canephora</i> LxPy selecionadas. As barras azuis correspondem às medidas realizadas em 2012 e as barras vermelhas às medidas realizadas em 2013. ....	62
Figura 20 – Dendrograma UPGMA representando a similaridade genética entre os representantes dos diferentes grupos de diversidade de <i>C. canephora</i> e dos Conilons oriundos do BAG do Incaper.....	76
Figura 21 – <i>Design</i> e desenvolvimento da plataforma de genotipagem Coffee Axiom Chip 26K para <i>C. canephora</i> .....	77
Figura 22 – Classificação, quanto à qualidade dos SNPs, gerada pelo <i>Axiom Best Practices Genotyping Workflow</i> .....	79
Figura 23 – Desenho representativo das sondas de DNA contendo o SNP. ....	81
Figura 24 – Distribuição dos SNPs contidos nas sondas para cada cromossomo de <i>C. canephora</i> .....	82
Figura 25 - Distribuição de SNPs em diferentes regiões genômicas do genoma de <i>C. canephora</i> . ....	83
Figura 26 – Posição das sondas na região gênica e intergênica do genoma de <i>C. canephora</i> . ....	83
Figura 27 – Dendrograma de proximidade dos grupos com as principais variedades comerciais do Brasil (Conilon) e os genótipos representantes dos diferentes grupos de diversidade de <i>C. canephora</i> .....	87
Figura 28 – Árvore filogenética ( <i>Neighbor-joining</i> ) entre os 1.319 indivíduos, utilizando-se 16.688 marcadores SNPs.....	89
Figura 29 – Análise do decaimento do desequilíbrio de ligação ( $r^2$ ) em <i>C. canephora</i> . (Continua).....	90
Figura 30 – Gráfico Manhattan de associação do efeito dos SNPs testados para a característica tamanho e formato de grãos utilizando o rrBLUP.....	102
Figura 31 – Gráfico quantil-quantil (Q-Q) dos resultados de GWAS para tamanho e formato de grãos.....	104
Figura 32 – Gráfico Manhattan de associação do efeito dos SNPs testados para a característica tamanho e formato de grãos utilizando o TASSEL.....	105
Figura 33 – Diagrama de Venn comparando as análises de GWAS realizadas no rrBLUP vs. TASSEL. ....	109
Figura 34 – Gráfico Manhattan de associação do efeito dos SNPs testados para a característica tamanho e formato de grãos utilizando o rrBLUP utilizando os marcadores de duas técnicas de genotipagem (chip 26K Axiom + nextRAD).....	113

Figura 35 – Gráfico quantil-quantil (Q-Q) dos resultados de GWAS para tamanho e formato de grãos utilizando marcadores identificados por duas técnicas de genotipagem (chip 26K Axiom + nextRAD).....	114
Figura 36 – Gráfico Manhattan de associação do efeito dos SNPs testados para as características de produção, avaliada no ano de 2012, e boia, utilizando o rrBLUP utilizando os marcadores de duas técnicas de genotipagem (chip 26K Axiom + nextRAD).....	116
Figura 37 – Gráfico quantil-quantil (Q-Q) dos resultados de GWAS, sem P3D, para produção e boia associando duas técnicas de genotipagem (chip 26K Axiom + nextRAD). A linha central na diagonal do gráfico representa a distribuição esperada e os pontos representam a distribuição observada. ....	117
Figura 38 – Gráfico Manhattan de associação do efeito dos SNPs testados para a característica de tamanho e formato de grãos, utilizando o rrBLUP utilizando os marcadores de duas técnicas de genotipagem (chip 26K Axiom + nextRAD).....	117
Figura 39 – Gráfico Manhattan de associação do efeito dos SNPs testados para a característica tamanho e formato de grãos utilizando o TASSEL com o conjunto de marcadores provenientes de duas técnicas de genotipagem (chip 26K Axiom + nextRAD). ....	120
Figura 40 – Gráfico Manhattan de associação do efeito dos SNPs testados para as características de peso de 100 grãos e potencial hídrico foliar de antemanhã utilizando o TASSEL com o conjunto de marcadores provenientes de duas técnicas de genotipagem (chip 26K Axiom + nextRAD). ....	121
Figura 41 – Porcentagens de SNPs identificados nos locos nextRAD para cada cromossomo de <i>C. canephora</i> .....	123
Figura 42 – Gráfico Manhattan de associação do efeito dos SNPs testados com as características de tamanho e formato de grãos e peso de 100 grãos. ....	124
Figura 43 – Gráfico Manhattan de associação do efeito dos SNPs testados com a característica de tamanho e formato de grãos.....	126
Figura 44 – Gráfico Manhattan de associação do efeito dos SNPs testados com a característica de peso de 100 grãos.....	127
Figura 45 – Gráfico Manhattan de associação do efeito dos SNPs testados com a característica de produção 2013, boia, peso de 100 grãos e eixo menor. ....	129
Figura 46 – Gráfico Manhattan de associação do efeito dos SNPs testados com a característica de tamanho e formato de grãos.....	131
Figura 47 – Localização dos genes associados nos cromossomos de <i>C. canephora</i> .....	143

Figura 48 – Gráfico Manhattan de associação do efeito dos SNPs testados com a característica de produção, avaliada nos anos de 2012 e 2014, utilizando o modelo bimodal.....	144
Figura 49 – Gráfico quantil-quantil (Q-Q) dos resultados de GWAS, utilizando o modelo bimodal, para produção avaliada nos anos de 2012 e 2014.....	145
Figura 50 – Gráfico com o VGG e o erro de predição do modelo para as características (A) Área, (B) Perímetro, (C) Maior Eixo e (D) Boia. Somente os indivíduos que apresentaram médias acima de zero estão presentes nos gráficos. ....	159
Figura 51 – Gráfico com o VGG e o erro de predição do modelo para as características (A) Potencial hídrico foliar de antemanhã, (B) Ferrugem, (C) Produção (2013) e (D) Média de produção para os três anos avaliados. Somente os indivíduos que apresentaram médias acima de zero estão presentes nos gráficos. ....	160
Figura 52 – Gráfico com o VGG e o erro de predição do modelo para tamanho e formato de grãos. ....	161
Figura 53 – Gráfico com o VGG e o erro de predição do modelo para as características (A) Área, (B) Perímetro, (C) Maior Eixo e (D) Menor Eixo na análise com os 1.319 indivíduos. Somente os indivíduos que apresentaram médias acima de zero estão presentes nos gráficos. ....	164
Figura 54 – Gráfico com o VGG e o erro de predição do modelo para as características (A) Boia, (B) P100, (C) Prod12, (D) Prod13, (E) Prod14 e (F) Média para as três produções, na análise com os 1.319 indivíduos. Somente os indivíduos que apresentaram médias acima de zero estão presentes nos gráficos.....	165
Figura 55 – Gráfico com o VGG e o erro de predição do modelo para as características (A) Chato Graúdo, (B) Chato Médio, (C) Chato Miúdo, (D) Moca Graúdo, (E) Moca Médio e (F) Moca Miúdo, na análise com os 1.319 indivíduos. Somente os indivíduos que apresentaram médias acima de zero estão presentes nos gráficos. ....	166
Figura 56 – Correlação entre os dados preditos, em alaranjado, e os dados reais, em azul, para produção (2014, 2015 e 2016) e peso de 100 grãos, para os 13 indivíduos de <i>C. canephora</i> utilizados na validação real do modelo de predição genômica. ....	167
Figura 57 – Correlação entre os dados preditos, em alaranjado, e os dados reais, em azul, para as características de formato e tamanho do grão, para os 13 indivíduos de <i>C. canephora</i> utilizados na validação real do modelo de predição genômica. ....	168

## LISTA DE TABELAS

Tabela 1 – Cultivares comerciais registradas e/ou protegidas junto ao MAPA. ....	27
Tabela 3 – Valores máximo (Máx), mínimo (Mín) e média, com o desvio padrão, para o fenótipo de Morfologia dos frutos. ....	56
Tabela 4 – Valor máximo, mínimo e média para a produção avaliada nos três anos e para 366 indivíduos. ....	59
Tabela 5 – Painel de resequenciamento composto por 48 indivíduos de <i>C. canephora</i> Conilon dividido em 14 pools (C1-C14). O número de indivíduos, o número de <i>reads</i> produzidos e o número de <i>reads</i> mapeados são indicados para cada pool. ....	74
Tabela 6 – Painel de resequenciamento composto por 51 indivíduos de <i>C. canephora</i> representando os diferentes grupos de diversidade, divididos em 10 pools (D1-D10). O grupo, subgrupo, número de indivíduos, o número de <i>reads</i> produzidos e o número de <i>reads</i> mapeados são indicados para cada pool. ....	75
Tabela 7 – Conjunto de validação do chip 26K de <i>C. canephora</i> . ....	78
Tabela 8 – Número de SNPs mapeados em cada cromossomo de <i>C. canephora</i> . ....	83
Tabela 9 – Número e porcentagem de SNPs classificação em cada categoria. ....	85
Tabela 10 – Marcadores associados para as características de tamanho e formato de grãos na análise com o rrBLUP. ....	103
Tabela 11 – Marcadores associados para as características de tamanho e formato de grãos na análise com o TASSEL. (Continua) ....	110
Tabela 12 – Marcadores associados para as características de tamanho e formato de grãos na análise com o rrBLUP (com P3D) e utilizando os marcadores de duas técnicas de genotipagem (chip 26K Axiom + nextRAD). ....	115
Tabela 13 – Marcadores associados para as características de produção, boia e tamanho e formato de grãos, na análise com o rrBLUP (sem P3D) e utilizando os marcadores de duas técnicas de genotipagem (chip 26K Axiom + nextRAD). (Continua) ....	118
Tabela 14 – Marcadores SNPs adicionais identificados na análise de associação utilizando o programa TASSEL, com o conjunto de marcadores provenientes de duas técnicas de genotipagem (chip 26K Axiom + nextRAD). ....	122
Tabela 15 – Marcadores associados para as características de peso de 100 grãos e tamanho e formato de grãos, na análise com os 1.319 indivíduos e utilizando o rrBLUP (com P3D) ...	125
Tabela 16 – Marcadores associados para as características de peso de 100 grãos e tamanho e formato de grãos, na análise com os 1.319 indivíduos e utilizando o rrBLUP (sem P3D). ...	127

Tabela 17 – Marcadores associados para as características de produção, boia, peso de 100 grãos e tamanho de fruto, na análise com os 1.319 indivíduos e utilizando o programa TASSEL. ....	130
Tabela 18 – Marcadores associados para as características de tamanho e formato de grãos (chato graúdo e chato miúdo) na análise com os 1.319 indivíduos e utilizando o programa TASSEL. ....	132
Tabela 19 – Marcadores associados para as características de tamanho e formato de grãos, indicando os 50 marcadores mais significativos para moca graúdo, na análise com os 1.319 indivíduos e utilizando o programa TASSEL. ....	133
Tabela 20 – Marcadores SNPs identificados como associados para as diferentes características testadas, comparando-se os programas estatísticos (TASSEL vs. rrBLUP), os diferentes métodos (com p3D vs. Sem P3D), os número de indivíduos (366 vs. 1.319) e os número de marcadores (Affymetrix vs. nextRAD). (Continua...) .....	138
Tabela 21 – Marcadores SNPs identificados como associados para as diferentes características testadas, comparando-se os programas estatísticos (TASSEL vs. rrBLUP), os diferentes métodos (com p3D vs. Sem P3D), os número de indivíduos (366 vs. 1.319) e os número de marcadores (Affymetrix vs. nextRAD). (Continua...) .....	139
Tabela 22 – Marcadores SNPs identificados como associados para as diferentes características testadas, comparando-se os programas estatísticos (TASSEL vs. rrBLUP), os diferentes métodos (com p3D vs. Sem P3D), os número de indivíduos (366 vs. 1.319) e os número de marcadores (Affymetrix vs. nextRAD). (Continua...) .....	140
Tabela 23 – Marcadores SNPs identificados como associados para as diferentes características testadas, comparando-se os programas estatísticos (TASSEL vs. rrBLUP), os diferentes métodos (com p3D vs. Sem P3D), os número de indivíduos (366 vs. 1.319) e os número de marcadores (Affymetrix vs. nextRAD). (Continua...) .....	141
Tabela 24 – Marcadores SNPs identificados como associados para as diferentes características testadas, comparando-se os programas estatísticos (TASSEL vs. rrBLUP), os diferentes métodos (com p3D vs. Sem P3D), os número de indivíduos (366 vs. 1.319) e os número de marcadores (Affymetrix vs. nextRAD). (Conclusão) .....	142
Tabela 25 – Resumo dos resultados obtidos para os componentes de variância e herdabilidade na análise dos caracteres avaliados para os 366 indivíduos de <i>C. canephora</i> .....	156
Tabela 26 – Acurácia preditiva do modelo, ganho com a seleção (GS) e pressão de seleção (PS) para cada característica analisada nos 366 indivíduos de <i>C. canephora</i> .....	158

Tabela 27 – Resumo dos resultados obtidos para os componentes de variância e herdabilidade na análise dos caracteres avaliados para os 1.319 indivíduos de <i>C. canephora</i> .....	162
Tabela 28 – Acurácia preditiva do modelo, ganho com a seleção (GS) e pressão de seleção (PS) para cada característica analisada nos 1.319 indivíduos de <i>C. canephora</i> .....	163

## SUMÁRIO

<b>1</b>	<b>CAPÍTULO 1 INTRODUÇÃO GERAL.....</b>	<b>16</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO.....</b>	<b>19</b>
<b>2.1</b>	<b>O cafeeiro.....</b>	<b>19</b>
<b>2.2</b>	<b><i>Coffea canephora</i>.....</b>	<b>22</b>
<b>2.2.1</b>	<b>Diversidade genética em <i>C. canephora</i>.....</b>	<b>23</b>
<b>2.3</b>	<b>Marcadores moleculares.....</b>	<b>27</b>
<b>2.4</b>	<b>Plataformas de genotipagem em larga escala.....</b>	<b>29</b>
<b>2.5</b>	<b>Estudo de associação genômica ampla.....</b>	<b>30</b>
<b>2.6</b>	<b>Seleção Genômica Ampla.....</b>	<b>33</b>
	<b>REFERÊNCIAS.....</b>	<b>36</b>
<b>3</b>	<b>OBJETIVOS.....</b>	<b>45</b>
<b>3.1</b>	<b>Objetivo geral.....</b>	<b>45</b>
<b>3.2</b>	<b>Objetivos específicos.....</b>	<b>45</b>
	<b>CAPÍTULO 2 FENOTIPAGEM DE INDIVÍDUOS DE UMA POPULAÇÃO DE <i>Coffea canephora</i>.....</b>	<b>46</b>
<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>46</b>
<b>2</b>	<b>MATERIAL E MÉTODOS.....</b>	<b>47</b>
<b>2.1</b>	<b>Material vegetal.....</b>	<b>47</b>
<b>2.2</b>	<b>Fenotipagem da população de <i>C. canephora</i>.....</b>	<b>47</b>
<b>2.2.1</b>	<b>Produção.....</b>	<b>48</b>
<b>2.2.2</b>	<b>Morfologia dos frutos.....</b>	<b>48</b>
<b>2.2.3</b>	<b>Boia.....</b>	<b>49</b>
<b>2.2.4</b>	<b>Peso de 100 grãos.....</b>	<b>49</b>
<b>2.2.5</b>	<b>Formato e tamanho dos grãos.....</b>	<b>50</b>
<b>2.2.6</b>	<b>Análises fisiológicas.....</b>	<b>51</b>
<b>3</b>	<b>RESULTADOS E DISCUSSÃO.....</b>	<b>52</b>
<b>3.1</b>	<b>Análise exploratória das variáveis fenotípicas.....</b>	<b>52</b>
<b>3.2</b>	<b>Morfologia dos frutos.....</b>	<b>56</b>
<b>3.3</b>	<b>Peso de 100 grãos e boia.....</b>	<b>56</b>
<b>3.4</b>	<b>Formato e tamanho do grão.....</b>	<b>56</b>
<b>3.5</b>	<b>Tolerância à seca.....</b>	<b>57</b>
<b>3.6</b>	<b>Produção.....</b>	<b>58</b>

4	CONCLUSÃO.....	63
	REFERÊNCIAS.....	64
	APÊNDICE.....	67
	<b>CAPÍTULO 3 VALIDAÇÃO DE UM CHIP DE DNA (COFFEE AXIOM CHIP – 26K) PARA GENOTIPAGEM EM ESCALA GENÔMICA DE <i>Coffea canephora</i>.....</b>	<b>72</b>
1	INTRODUÇÃO.....	72
2	MATERIAL E MÉTODOS.....	74
2.1	Material vegetal e resequenciamento.....	74
2.2	Seleção e validação dos SNPs.....	77
2.3	Chamada de SNP e análise dos dados.....	78
2.4	Genotipagem em larga escala usando o chip de DNA 26k de <i>C. canephora</i> .....	80
2.4.1	Material vegetal e extração de DNA.....	80
2.4.2	Análise da diversidade e desequilíbrio de ligação.....	80
3	RESULTADOS E DISCUSSÃO.....	81
3.1	Alinhamento e detecção dos SNPs.....	81
3.2	Validação do chip de genotipagem.....	84
3.3	Análise filogenética do conjunto de validação.....	85
3.4	Desequilíbrio de ligação e estrutura da população da Embrapa Cerrados..	88
4	CONCLUSÃO.....	92
	REFERÊNCIAS.....	93
	<b>CAPÍTULO 4 ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA (GWAS) EM <i>C. canephora</i>.....</b>	<b>95</b>
1	INTRODUÇÃO.....	95
2	MATERIAL E MÉTODOS.....	97
2.1	Material vegetal.....	97
2.2	Genotipagem.....	97
2.3	Análise de associação genômica ampla (GWAS).....	98
3	RESULTADOS E DISCUSSÃO.....	99
3.1	Análise de 366 indivíduos.....	99
3.1.1	Dados genotípicos do chip 26K Axiom de <i>C. canephora</i> – rrBLUP.....	99
3.1.2	Dados genotípicos do chip 26K Axiom de <i>C. canephora</i> – TASSEL.....	104
3.1.3	Dados genotípicos do chip 26K Axiom de <i>C. canephora</i> e nextRAD –	



	rrBLUP.....	112
3.1.4	Dados genotípicos do chip 26K Axiom de <i>C. canephora</i> e nextRAD – TASSEL.....	120
3.2	Análise de 1.319 indivíduos.....	124
3.2.1	rrBLUP.....	124
3.2.2	TASSEL.....	128
3.2.3	Modelo Bimodal.....	144
4	CONCLUSÃO.....	147
	REFERÊNCIAS.....	148
	CAPÍTULO 5 SELEÇÃO GENÔMICA AMPLA EM <i>C. canephora</i> .....	152
1	INTRODUÇÃO.....	152
2	MATERIAL E MÉTODOS.....	154
2.1	Dados fenotípicos.....	154
2.2	Dados genotípicos.....	154
2.3	Modelo de predição.....	155
3	RESULTADOS E DISCUSSÃO.....	156
3.1	Análise de 366 indivíduos.....	156
3.1.1	Componentes de variância e herdabilidade.....	156
3.1.2	Acurácia e Erro de predição do modelo.....	157
3.2	Análise de 1.319 indivíduos.....	162
3.2.1	Componentes de variância e herdabilidade.....	162
3.2.2	Acurácia e Erro de predição do modelo.....	162
3.2.3	Validação Real.....	167
4	CONCLUSÃO.....	170
	REFERÊNCIAS	171

## CAPÍTULO 1

### INTRODUÇÃO GERAL

O café é uma das commodities agrícolas mais comercializadas globalmente. A produção mundial de café se baseia em duas espécies, *Coffea arabica* L. e *Coffea canephora* Pierre ex A. Froehner, respondendo respectivamente, por 2/3 e 1/3 da produção mundial. Esta produção deve-se principalmente a pequenos produtores que enfrentam inúmeros desafios, especialmente em relação às mudanças climáticas. Nessas circunstâncias, tanto o rendimento do café como a qualidade da bebida podem ser afetados.

Nos últimos anos, tem havido uma crescente preocupação sobre como as mudanças climáticas afetarão as áreas tradicionais de produção de café. Vários estudos de modelagem estimaram os efeitos sobre a cultura do café, incluindo perdas severas de rendimentos (GAY et al., 2006; SCHROTH et al., 2009), reduções extensivas de áreas adequadas para a produção de café no Brasil (ZULLO et al., 2011) e possível extinção de populações selvagens de café arábica na Etiópia (DAVIS et al., 2012). Além disso, sob cenários de mudanças climáticas, condições favoráveis para algumas pragas e doenças importantes, como o bicho mineiro, nematoides e ferrugem no Brasil (GHINI et al., 2012; GHINI; BETTIOL; HAMADA, 2011), broca de café no leste da África (JARAMILLO et al., 2011) e broca branca do tronco, no Zimbábue (KUTYWAYO et al., 2013), deverão aumentar. A maioria desses potenciais efeitos negativos para os produtores de café tem sido associada a previsões de aumento das temperaturas e padrões de precipitação alterados, nos cenários de mudanças climáticas presentes e futuras.

A concentração atmosférica de CO<sub>2</sub> ([CO<sub>2</sub>]) aumentou cerca de 40% dos níveis pré-industriais e a temperatura média global da superfície aumentou 0,85 °C durante o mesmo período. As projeções sugerem que a [CO<sub>2</sub>] atmosférico pode exceder 936 μmol mol<sup>-1</sup> (RCP 8.5) até o final deste século e a temperatura global aumentará 1,5-4,0 °C em relação à média de 1850 a 1900 (IPCC, 2013).

O aumento na [CO<sub>2</sub>] pode afetar processos essenciais para a planta, como fotossíntese e respiração (KIRSCHBAUM, 2011), e como consequência, alterar o crescimento, produção e qualidade (DAMATTA et al., 2010). Embora o aumento do CO<sub>2</sub> represente um aumento da taxa fotossintética, altas temperaturas prolongadas levam à morte das plantas e, em períodos moderados, afeta a floração, o desenvolvimento do grão e, conseqüentemente, a produtividade

(CAMARGO, 2010; MARRACCINI et al., 2011), representando, portanto, uma grande ameaça para as culturas agrícolas.

Com uma vida útil média de cerca de 30 a 50 anos, significa que as plantações de café existentes podem, desta forma, experimentar as mudanças climáticas previstas. As variedades comerciais atualmente em uso possuem uma base genética estreita (ANTHONY et al., 2001) e, portanto, uma faixa de adaptação climática estreita (DAMATTA, 2004). Diante deste contexto, uma alternativa para obter plantas que possam suportar essas alterações climáticas e, ao mesmo tempo, não ter a produção e qualidade de bebida afetada, seria com o melhoramento genético, entretanto, em espécies perenes como o café, aproximadamente 30 anos são necessários para se desenvolver uma nova cultivar com as características de interesse, usando os métodos convencionais. Além do longo tempo demandado, o alto custo, as diferenças de ploidia entre as espécies do gênero *Coffea* e a incompatibilidade genética também são fatores que dificultam o melhoramento convencional (MISHRA; SLATER, 2012).

Com os avanços recentes na genômica do cafeeiro, como o sequenciamento completo do genoma de referência de *C. canephora* (DENOEUDE et al., 2014), uma redução significativa em tempo e custo na seleção de plantas com características de interesse ao melhoramento pode ser alcançada. O genoma de referência juntamente com as tecnologias de sequenciamento de nova geração (Next Generation Sequencing – NGS) forneceram as ferramentas necessárias para a genotipagem de SNPs (Single Nucleotide Polymorphisms) em larga escala. Desta forma, os marcadores SNPs tem se mostrado uma opção bastante atrativa devido ao custo, relativamente baixo, a abundância no genoma, a especificidade, a codominância, ao potencial de análise e a baixa taxa de erro de genotipagem (RAFALSKI, 2002; SCHLOTTERER, 2004; THUDI et al., 2012).

As plataformas de genotipagem de SNPs em larga escala resultam em uma alta cobertura e alta precisão, características essenciais para realizar predição genômica e descoberta de genes em estudos de associação genômica (Genome Wide Association - GWAS). O desenvolvimento de *chips* de genotipagem de DNA, com milhares de SNPs, tem aumentado significativamente, não somente em organismos modelos, mas também em inúmeras espécies que ainda não possuem os dados genômicos disponíveis (GUPTA et al., 2013).

As plataformas da Illumina e Affymetrix tornaram-se amplamente utilizadas na construção destes chips de DNA, possibilitando a identificação de SNPs com uma alta densidade e alta cobertura do genoma (GANAL et al., 2014). Recentemente, chips de genotipagem Axiom® (Affymetrix) foram desenvolvidos para realizar estudos de associação

genômica em morango, algodão, rosa, soja e maçã (BASSIL et al., 2015; BIANCO et al., 2016; HULSE-KEMP et al., 2015; KONING-BOUCOIRAN et al., 2015; LEE et al., 2015;).

Com base nessas tecnologias disponíveis e considerando o cenário climático futuro, o presente trabalho tem por finalidade realizar estudos de associação genômica ampla (GWAS) e de desenvolver modelos preditivos de seleção genômica ampla (SGA) em *C. canephora*. A escolha de *C. canephora*, se baseia no fato de que, além da sua importância econômica, esta espécie é geneticamente diversa, diploide, pode suportar temperaturas mais altas do que *C. arabica*, exibe resistência a importantes patógenos como *M. paranaensis* (LIMA et al., 2015) e tolerância a estresses abióticos como o estresse hídrico (MARRACCINI et al., 2012; VIEIRA et al., 2013). Além disso, sua sequência de genoma já é conhecida (DENOEUDE et al., 2014) e, mais importante, é um dos subgenomas de *C. arabica* (LASHERMES et al., 1999) e, nesse sentido, o conhecimento molecular gerado para *C. canephora* certamente poderá ser traduzido também, para programas de melhoramento genético de *C. arabica*.

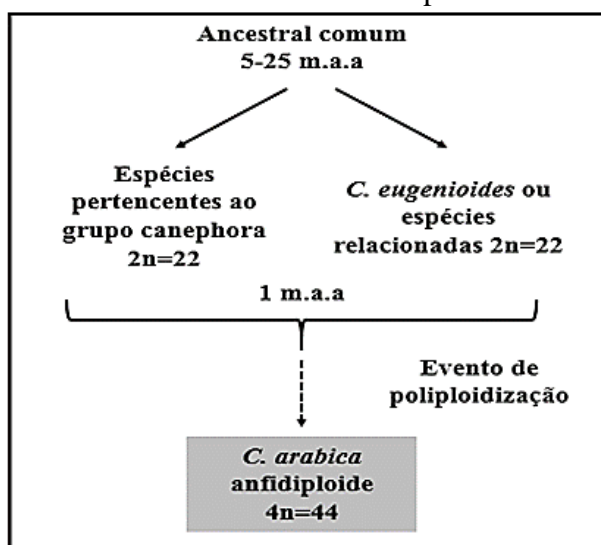
## 2 REFERENCIAL TEÓRICO

### 2.1 O cafeeiro

Nativo da África, Madagascar e Mascarenhas (MUSOLI et al., 2009; DE KOCHKO et al., 2010), o cafeeiro pertence à família *Rubiaceae*, subfamília *Ixoroideae*, tribo *Coffeae* e ao gênero *Coffea*, compreendendo mais de 120 espécies (DAVIS et al., 2006; HENDRE et al., 2008). Ao nível botânico, as plantas de *Coffea* são dicotiledôneas, de folhas persistentes e flores hermafroditas, de porte arbustivo ou arbóreo e caule lenhoso (FAZUOLI et al., 2000). Evolutivamente, o início da diversificação deste gênero ocorreu entre 5 e 25 milhões de anos atrás (CUBRY et al., 2008).

De todas as espécies pertencentes ao gênero, apenas duas se destacam para fins comerciais: *C. canephora* e *C. arabica*. As duas espécies são árvores lenhosas perenes e apresentam variações consideráveis quanto à morfologia, tamanho e adaptação ecológica (COMBES et al., 2015). *C. canephora* é uma espécie diploide ( $2n = 2X = 22$ ), cultivada em baixas e médias altitudes nas regiões intertropicais da África, América e Ásia, alógama apresentando autoincompatibilidade genética (CUBRY et al., 2012). Devido à alogamia, apresenta uma variabilidade genética maior quando comparado ao *C. arabica*, sendo mais resistente a doenças e pragas, e capaz de adaptar-se às diversas condições climáticas (BERTRAND et al., 2003).

*C. arabica* é autógama alotetraploide ( $2n = 4X = 44$ ) que se originou de um cruzamento entre *C. canephora* e *Coffea eugenioides* (FIGURA 1) ocorrido nos planaltos da Etiópia Central (LASHERMES et al., 1999). Por ser considerada superior em termos de qualidade de bebida, apresentando menor amargor e teor de cafeína, o *C. arabica* alcança um maior valor de mercado (ALONSO-SALCES et al., 2009; CUBRY et al., 2012; LEROY et al., 2011) e representa atualmente 70% da produção mundial.

Figura 1 - História evolutiva do alotetraplóide *C. arabica*.

Legenda: Origem do *C. arabica*: os genomas dos progenitores estão representados pelos diploides *C. eugenioides* e *C. canephora*. *C. arabica* surgiu cerca de 1 milhão de anos atrás (m.a.a) a partir do cruzamento de *C. canephora* (ou espécies relacionadas) e *C. eugenioides*.

Fonte: Vidal et al. (2010) adaptada.

Tradicionalmente, a produção de café no Brasil e no mundo concentrava-se apenas na espécie *C. arabica*. Entretanto, entre os anos de 1870 e 1900, diante de uma epidemia de ferrugem no sudeste asiático causada pelo fungo *Hemileia vastatrix* que devastou culturas em diversos países produtores de *C. arabica*, a espécie *C. canephora* passou a ser alvo de estudos científicos visando à sua exploração econômica, já que esta espécie apresentava resistência à doença.

O café destaca-se econômica e socialmente no Brasil, desde a chegada das primeiras mudas vindas da Guiana Francesa em meados do século XVIII e, diante de sua rápida adaptação ao solo e clima, o produto adquiriu importância no mercado, transformando-se em um dos principais itens de exportação, desde o Império até os dias atuais (Ministério da Agricultura Pecuária e Abastecimento-MAPA)<sup>1</sup>.

De acordo com os dados divulgados pela ICO (2019), a produção mundial de café no ano de 2018 foi de mais de 168 milhões de sacas (60 Kg), sendo que o Brasil contribuiu com mais de 1/3 deste total, produzindo pouco mais de 61 milhões de sacas, seguido pelo Vietnã, 29,55 milhões, e Colômbia, 15,55 milhões de sacas.

De acordo com a estimativa da Companhia Nacional de Abastecimento-CONAB (2019)<sup>2</sup> o Brasil deverá colher em 2019 mais de 50 milhões de sacas de café beneficiado

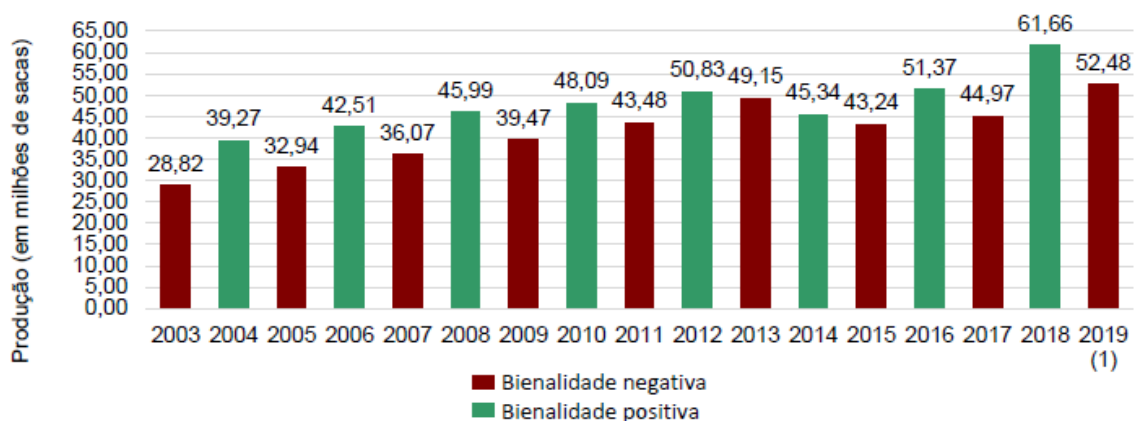
<sup>1</sup> Disponível em: <http://www.agricultura.gov.br/vegetal/culturas/cafe>

<sup>2</sup> Disponível em: <http://www.conab.gov.br>

(espécies *C. arabica* e *C. canephora*) (FIGURA 2), observando-se uma queda em relação ao ano anterior, o que já era esperado devido ao fato do ano de 2019 ser de bienalidade negativa.

A bienalidade é própria da natureza fisiológica de plantas perenes, como o cafeeiro, que necessita vegetar em um ano para produzir bem no ano seguinte. Nos anos de alta carga, a planta direciona a produção de fotossintetizados para a formação e crescimento dos frutos. Nos anos de baixa carga, estes são direcionados a formação de novas gemas vegetativas que gerarão novos ramos (PICINI, 1998, RENA; MAESTRI, 1985). É possível observar na Figura 2, que para o ano de 2014, considerado de alta carga, a redução na produção se deve principalmente a forte estiagem que ocorreu nos primeiros meses do ano, afetando diretamente o desenvolvimento do grão, favorecendo o aumento de frutos chochos e/ou mal granados (BELAN et al., 2011).

Figura 2 – Evolução da produção brasileira de café beneficiado para o período de 2003 a 2019. A produção para 2019 é uma estimativa realizada em janeiro/2019.



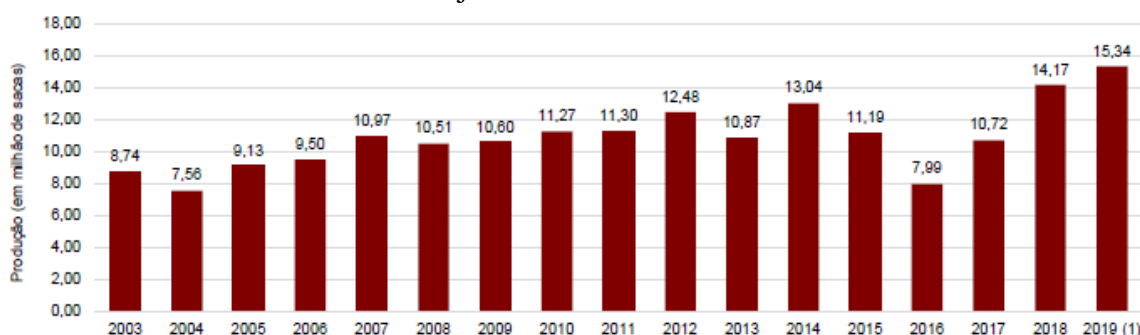
Fonte: CONAB (2019).

O café produzido por *C. canephora* é de grande importância econômica para um grande número de países da zona intertropical. Dentro de um século, inúmeros trabalhos em agronomia e melhoramento varietal aumentaram seu potencial de produção. Embora forneça uma bebida de qualidade inferior à bebida fornecida pelos grãos de *C. arabica*, os de *C. canephora* vem apresentando um crescimento constante no Brasil (CONAB, 2019). As indústrias de café solúvel optam pela sua utilização em misturas com o *C. arabica*, aumentando a lucratividade do produto, sem causar perdas excessivas na qualidade, uma vez que seu preço é inferior no mercado mundial.

No Brasil, estima-se que serão produzidas mais de 15 milhões de sacas de *C. canephora* no ano de 2019 (FIGURA 3), e o estado que se destaca quanto à produção é o Espírito Santo (66-69%), seguido de Rondônia e Bahia (CONAB, 2019). Observa-se que

nestes estados as lavouras têm sido renovadas com variedades superiores, por meio da multiplicação vegetativa de mudas, realizada por estacas (FERRÃO et al., 2007), e com outras tecnologias associadas, visando-se o aumento da produção e melhoria da qualidade final do produto. Como consequência, uma remuneração mais adequada para o café de qualidade superior servirá como incentivo para que cafeicultores aumentem a produtividade e se insiram intensamente em programas para a melhoria da qualidade do *C. canephora*.

Figura 3 – Produção de *C. canephora* Conilon no Brasil. A produção para 2019 é uma estimativa realizada em janeiro/2019.



Fonte: CONAB (2019).

## 2.2 *Coffea canephora*

*C. canephora* é uma rubiácea originária da África sub-equatorial, descrita por Louis Pierre, em 1895, e classificada em 1897 por Albrecht Froehner. As plantas selvagens desta espécie são naturalmente distribuídas na África intertropical, estendendo-se da Guiné para Uganda e da República Centro-Africana para Angola. As populações naturais são compostas de poucos indivíduos, submetidos a fluxo gênico de populações vizinhas até alguns quilômetros de distância (MONTAGNON et al., 1993).

Seu cultivo remonta ao século XIX na Uganda e na parte oriental da Bacia do Congo. Já no início do século XX, sementes originárias da África Central foram enviadas para Java, onde começaram os primeiros trabalhos de seleção da espécie. A variedade 'Robusta' de *C. canephora* é selecionada e, em seguida, substitui *C. arabica* que havia sofrido com a epidemia de ferrugem na Ásia. Por volta de 1910, sementes das variedades selecionadas em Java retornam à África e, posteriormente, os diferentes genótipos foram introduzidos na maioria dos países produtores, incluindo o Brasil.

A espécie possui caules lenhosos e múltiplos, folhas lanceoladas e grandes, com nervuras salientes e bordas onduladas (DE KOCHKO et al., 2010; MONTAGNON; LEROY; YAPO, 2012). Suas flores são hermafroditas e a reprodução ocorre de forma alógama devido



à incompatibilidade gametofítica (LASHERMES et al., 1996a). A espécie pode florir uma ou várias vezes por ano e, no caso específico do cafeeiro, existe um sincronismo do florescimento que está associado com ciclos de déficit hídrico nas plantas. Após longos períodos de estiagem, os quais quebrariam a dormência das gemas florais, o retorno das chuvas ou irrigação, resultaria em um florescimento uniforme (FERRÃO et al., 2007; SILVA et al., 2009).

Como mencionado, a espécie possui um mecanismo de autoincompatibilidade genética do tipo gametofítico, permitindo que as plantas reconheçam e rejeitem seu próprio pólen ou pólen com genótipo semelhante (NOWAK et al., 2011), favorecendo a alogamia e contribuindo para a manutenção da variabilidade genética. No caso de *C. canephora*, o sistema de autoincompatibilidade é controlado por um loco denominado S, no qual a formação do tubo polínico é impedida por S-RNases produzidas no pistilo, quando o alelo do loco-S do grão de pólen (haploide) é idêntico a um dos alelos do pistilo (diploide) (ASQUINI et al., 2011).

### **2.2.1 Diversidade genética em *C. canephora***

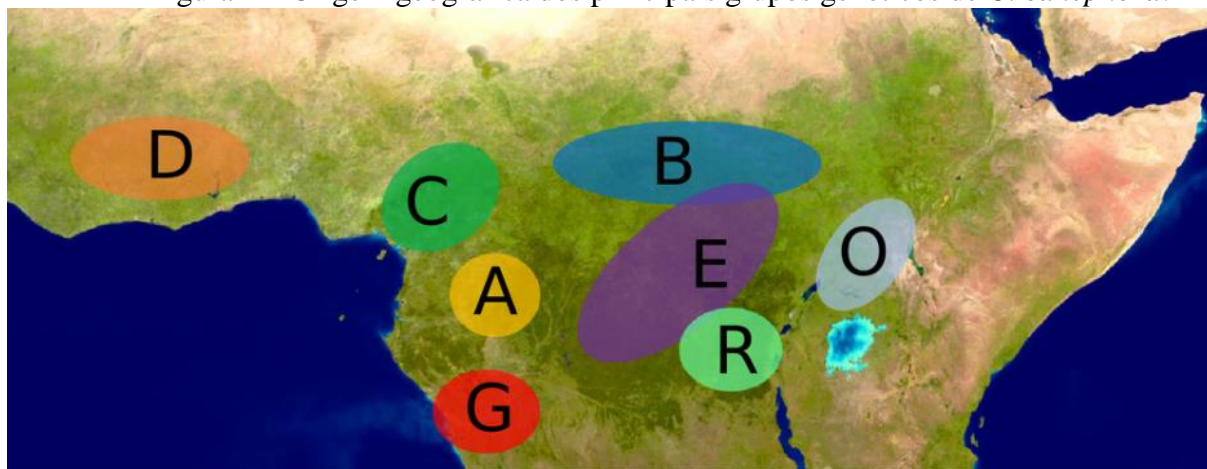
A primeira análise da diversidade genética em *C. canephora*, utilizando isoenzimas, foi realizado por Berthaud (1986), no qual quinze amostras foram classificadas de acordo com o cálculo da distância genética. Este estudo pioneiro demonstrou uma estruturação da espécie, em que dois grupos genéticos foram identificados, o Guineano e o Congolês. Posteriormente, Montagnon et al. (1992) incluíram dentro do grupo Congolês dois subgrupos, denominados SG1 e SG2. Musoli et al. (2009), trabalhando com microssatélites, acrescentaram ao grupo Congolês um novo grupo de diversidade, formado por indivíduos selvagens de Uganda. A análise realizada por Gomez et al. (2009) resultou em cinco grupos genéticos bem definidos (denominados A-E), definidos a priori por Dussert et al. (1999), causando uma confusão na comunidade científica. Esses cinco grupos genéticos corresponderam as diferentes origens geográficas dos indivíduos (GOMEZ et al., 2009) (FIGURA 4).

Enquanto essas análises anteriores resultaram em algum acordo de grupos genéticos e origem geográfica, cada trabalho independente deu nomes diferentes aos grupos genéticos, causando uma confusão para a comunidade cafeeira, sugerindo a importância de definir com precisão uma nomenclatura geral. Isso foi proposto por Garavito et al. (2016) que sugeriu o uso da seguinte nomenclatura para pelo menos cinco grupos genéticos anteriormente referenciados de *C. canephora*: (i) grupo guineano (às vezes chamado de grupo D) é o grupo

genético originário da área da Costa do Marfim-Guiné na África Ocidental; (ii) "Nana" (às vezes chamado de grupo C) grupo formado pelo café originário da periferia do Sudeste de Camarões, do sudoeste da África Central e do norte do Congo; (iii) grupo "Conilon" (denominado SG1 ou A), representado pelas populações domesticadas dos tipos Luki-, Niaouli e Kouilou, originárias do sul do Gabão; (iv) grupo "Robusta Congo-África Central" (por vezes denominado B) composto pelos cafés silvestres do norte da bacia central do Congo e do sul da África Central e (v) grupo "Robusta Congo Uganda" (por vezes denominado SG2) das populações selvagens ou variedades cultivadas nativas de Uganda e da bacia do Congo.

A caracterização dos grupos genéticos de *C. canephora* realizada por Merot-L'anthoene et al. (2019), utilizando SNPs, confirmou a existência de todos os grupos de diversidade anteriormente mencionados (A-E, e o grupo de Uganda, denominado de grupo O) e incluiu indivíduos de Angola e da República Democrática do Congo (RDC) com a identidade genética desconhecida, identificando mais dois novos grupos, denominados de G (Angola) e R (RDC), conforme Figura 4.

Figura 4 – Origem geográfica dos principais grupos genéticos de *C. canephora*.



Legenda: Distribuição global dos grupos genéticos de *C. canephora* com suas origens geográficas. O grupo A, corresponde ao SG1, os grupos B e E corresponde ao SG2, o grupo D que corresponde ao Guineano, o grupo O com indivíduos de Uganda, grupo C ou Nana e os grupos R e O, com indivíduos da RDC e Angola, respectivamente.

Fonte: MEROT-L'ANTHOENE et al. (2019).

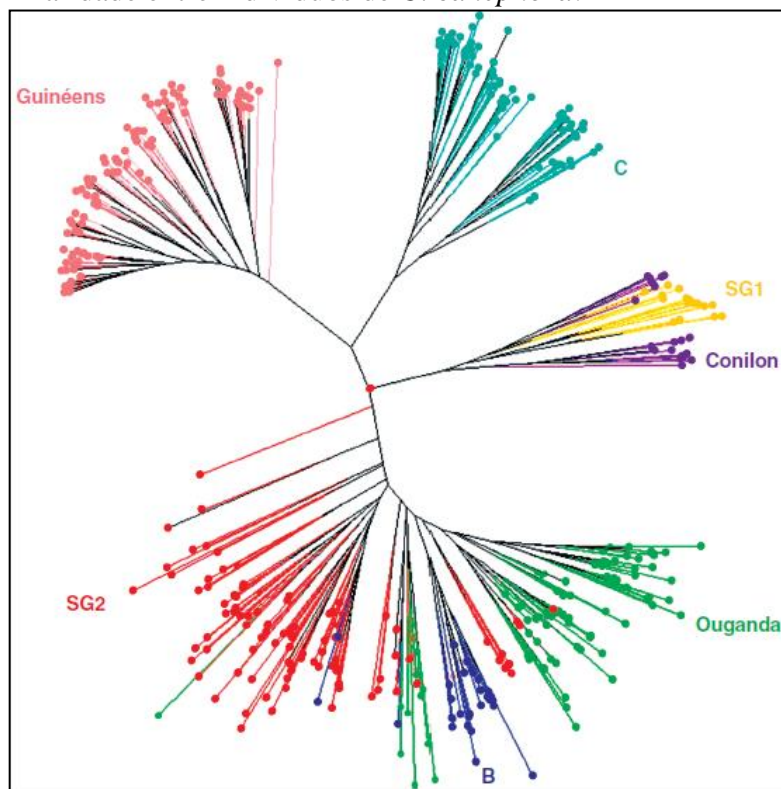
Outros trabalhos adicionais foram realizados caracterizando a diversidade genética dos materiais cultivados de *C. canephora*, atualmente em uso nos principais países produtores. O material genético cultivado no Vietnã, o primeiro país produtor de *C. canephora*, foi recentemente descrito utilizando o método DArTseq (GARAVITO et al., 2016). Materiais genéticos brasileiros de coleções de germoplasma e material cultivado também foram caracterizados (FERRÃO et al., 2013, 2014). Germoplasma de *C. canephora* da Indonésia foi

caracterizado por marcadores SSR (SUMIRAT et al, 2012). Foi também relatada uma análise da diversidade genética de robusta indiano (PRAKASH et al., 2005). Acessos genéticos de Uganda, compreendendo material cultivado e selvagem de *C. canephora* também foram estudados (MUSOLI et al., 2009).

Os países mencionados acima representam mais de 95% da produção mundial atual de café robusta (<http://www.ico.org/>), indicando que a maior parte da diversidade genética dos materiais genéticos cultivados de *C. canephora* em todo o mundo já é conhecida. O mais interessante nos resultados desses estudos, foi a proximidade entre os materiais cultivados analisados e os materiais da diversidade selvagem de *C. canephora*, com poucos sinais de domesticação, mostrando que ainda há um espaço enorme para a exploração da diversidade genética conhecida, pelos programas de melhoramento genético do café robusta. As ferramentas genômicas disponíveis agora podem ser empregadas para promover e acelerar os programas de melhoramento de *C. canephora* com uma boa perspectiva de sucesso (ANDRADE, 2018).

Durante a última década, vários programas de melhoramento buscam explorar essa diversidade para o desenvolvimento de novos clones. No Brasil, o programa de melhoramento realizado pela, até então, Empresa Capixaba de Pesquisa Agropecuária (EMCAPA), hoje Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (Incaper), explorou a variabilidade de genótipos SG1, caracterizando uma variedade clonal de *C. canephora* Conilon altamente produtiva em condições de seca, denominada EMCAPA 8141 – Robustão Capixaba (FERRÃO et al., 2000). Essa proximidade genética entre o Conilon, cultivado no Brasil, e os indivíduos do grupo SG1, originários do Sul do Gabão, pode ser confirmada ao se analisar a Figura 5. Isto corrobora com a hipótese de Montagnon (2000), com base no fenótipo destas plantas e do nome Conilon, derivado de Kouilou, que é o nome de um rio localizado no sul do Gabão. O genótipo Kouilou ou Conilon chegou ao Brasil por volta do ano 1912, nos estados do Rio de Janeiro e Espírito Santo (FERRÃO et al., 2007).

Figura 5 – Árvore filogenética (*Neighbor-joining*) com base em uma matriz de dissimilaridade entre indivíduos de *C. canephora*.



Legenda: Árvore filogenética (*Neighbor-joining*) com base em uma matriz de dissimilaridade entre 519 indivíduos, utilizando sete marcadores microsatélites. Esta árvore compreende todas as populações ou grupos de diversidade. Cada cor corresponde a um grupo genético: Guineano, em rosa, SG1, amarelo, SG2, em vermelho, B, azul escuro, C, azul claro, e Uganda, em verde. A cor violeta corresponde ao material representativo da variedade Conilon brasileira.

Fonte: MONTAGNON, CYBRY E LEROY (2012).

Além do Incaper, outra instituição que se tem destacado no desenvolvimento de novas variedades comerciais de *C. canephora* é a Embrapa Rondônia, que em 2012, em parceria com o Consórcio Pesquisa Café, lançaram a cultivar clonal BRS Ouro Preto. A Tabela 1 mostra as cultivares registradas e/ou protegidas junto ao MAPA<sup>3</sup> desenvolvidas pelo Incaper e Embrapa.

<sup>3</sup> Disponível em: <http://www.agricultura.gov.br/guia-de-servicos/registro-nacional-de-cultivares-rnc>

Tabela 1 – Cultivares comerciais registradas e/ou protegidas junto ao MAPA.

Nome da cultivar	Registrada	Protegida	Titular	Referência
BRS Ouro Preto	X	X	Embrapa	Ramalho et al., 2015
Centenária ES8132	X	X	Incaper	Ferrão et al., 2015
Diamante ES8112	X	X	Incaper	
ES8122	X	X	Incaper	
EMCAPA 8111 - Precoce	X		Incaper	Bragança et al., 2001
EMCAPA 8131	X		Incaper	
EMCAPA 8121	X		Incaper	
EMCAPA 8141 Robustão Capixaba	X		Incaper	Ferrão et al., 2000
Vitória Incaper 8142	X		Incaper	Fonseca et al., 2004

### 2.3 Marcadores moleculares

Marcador molecular é definido como todo e qualquer fenótipo molecular oriundo de um gene expresso ou de um segmento específico de DNA, que diferencie dois ou mais indivíduos e que sejam herdados geneticamente (FERREIRA; GRATTAPAGLIA, 1998; MILACH, 1998). Diversas técnicas de biologia molecular estão disponíveis atualmente para a detecção da variabilidade genética ao nível de DNA, ou seja, detecção de polimorfismo genético (FERREIRA; GRATTAPAGLIA, 1998). Esses polimorfismos estão no centro da genética moderna e podem ser medidos em uma ou mais populações, além de serem herdáveis, sendo comumente utilizados em estudos de genética de populações, genética ecológica e evolução (DAVEY et al., 2011).

A introdução do uso de marcadores moleculares em estudos genéticos ocorreu no início da década de 80 e desde então passou a ser utilizado para diversas finalidades, como por exemplo: estudos de diversidade genética, caracterização de bancos de germoplasma, genealogia, construção de mapas, mapeamento comparativo, mapeamento gênico, seleção de genitores, certificação de cruzamentos, predição de fenótipos, *fingerprinting*, análises de pureza genética de sementes, melhoramento assistido, proteção varietal, mapeamento físico de genomas, integração de mapas genéticos e físicos, clonagem posicional, estudos de

desequilíbrio de ligação, mapeamento de associação, filogenia, estudo de pedigree, isolamento de genes e diagnose (BORÉM; CAIXETA, 2009).

Diferentes classes de marcadores moleculares vêm sendo aplicados no melhoramento genético vegetal. Os marcadores de DNA são versáteis e possuem inúmeras vantagens quando comparados aos marcadores fenotípicos convencionais, pois estão presentes em maior número, são estáveis em todos os estádios de desenvolvimento da planta e são neutros, ou seja, não variam de acordo com o ambiente (RESENDE et al., 2012).

Em café, grande parte dos trabalhos utilizando marcadores moleculares, estão relacionados com o estudo da diversidade genética do gênero, geralmente utilizando algumas poucas dezenas de marcadores. Berthaud (1986) e Montagnon; Leroy e Yapo (1992) utilizaram isoenzimas, Dussert et al. (2003) utilizaram RFLP, e trabalhos mais recentes foram com microssatélites (CUBRY et al., 2008; 2012; 2013, LEROY et al., 2011; MUSOLI et al., 2009; PONCET et al., 2006; 2007; PRAKASH et al., 2005) e DArTSeq (GARAVITO et al., 2016).

Estudos mais avançados em genômica populacional e seleção genômica, entretanto, demandam uma cobertura genômica consideravelmente mais ampla. Para isso métodos de mais alto desempenho e elevada reprodutibilidade são necessários. Uma nova geração de metodologias para a genotipagem de marcadores moleculares surgiu nos últimos anos com o advento das novas tecnologias de sequenciamento. Com base em NGS, estão sendo desenvolvidos novos métodos para descobrir, sequenciar e genotipar uma grande quantidade de marcadores em praticamente qualquer genoma de interesse em uma única etapa, mesmo em populações com pouca ou nenhuma informação genética disponível (DAVEY et al., 2011).

O refinamento dessas metodologias está apenas no seu início, mas podemos citar alguns dos novos métodos para se identificar marcadores moleculares de forma ampla no genoma, tais como bibliotecas de representação reduzida-*RRLs*, sequências polimórficas com complexidade reduzida (redução de complexidade de sequências polimórficas-*CRoPS*), *DArTseq* (diversidade baseada em sequenciamento Tecnologia de matriz), sequenciamento em grande escala de fragmentos de DNA associados ao sítio de restrição (*RADseq*), genotipagem de “*shotgun*” multiplexada (MSG) ou genotipagem por sequenciamento (Genotyping By Sequencing-*GBS*), entre outros (DAVEY et al., 2011). Estas metodologias estão começando a revolucionar a genômica aplicada ao melhoramento e genética de populações (POLAND; RIFE, 2012).

Considerado o tipo mais comum de polimorfismo de DNA, os marcadores SNPs são hoje os mais versáteis e eficientes para o estudo de associações entre polimorfismos ao longo de todo o genoma e fenótipos quantitativos e qualitativos, construção de mapas genéticos e mapeamento de locos controladores de características quantitativas, diagnósticos genéticos, análise da estrutura genética de populações, análise filogenética, entre outras (RAFALSKI, 2002). Sua ocorrência e distribuição ao longo do genoma varia entre as espécies em função de diversos aspectos relacionados com o sistema preferencial de reprodução, história evolutiva da espécie e da população alvo do estudo (LANNES et al., 2007; LI et al., 2009; SOLEIMANI; BAUM; JOHNSON, 2003).

Alguns trabalhos identificando SNPs em cafeeiro utilizando bibliotecas ESTs (*Expressed Sequence Tag*) foram realizados no Brasil (MONDEGO et al., 2011; VIDAL et al., 2010; VIEIRA et al., 2006;). Em outros trabalhos, utilizando INDELS, Marraccini et al. (2011) estudaram a expressão dos diferentes alelos do gene que codifica a subunidade menor da ribulose 1,5-bis-fosfatocarboxilase (Rubisco), em folhas de *C. arabica* e *C. canephora*, e Cotta et al. (2014) analisaram a expressão dos genes homeólogos de nsLTP durante a maturação de frutos de *C. arabica*.

Marcadores RADseq também já foram produzidos e identificados em café, sendo utilizados para saturar o mapa genético de uma população de *C. canephora* da ICRI-Indonésia, que ajudou a montagem do genoma do café (DENOEUDE et al., 2014). Em outro estudo, marcadores DArTseq (CRUZ; KILIAN; DIERIG, 2013), foram utilizados para o estudo da diversidade genética de *C. canephora* (GARAVITO et al., 2016). Além disso, outro trabalho com o objetivo de avaliar e caracterizar a diversidade genética de 480 indivíduos de uma população de *C. canephora* no Brasil, por meio da técnica de genotipagem nextRAD, foi realizado. Neste, obteve-se uma lista de 5.412 SNPs em café, sendo que o cromossomo 2 obteve a maior densidade de sondas mapeadas e mais de 60% dos SNPs identificados ocorreram em regiões gênicas (CARNEIRO et al., 2014).

## **2.4 Plataformas de genotipagem em larga escala**

As plataformas de genotipagem de SNPs em larga escala resultam em uma alta cobertura e alta precisão, características essenciais para realizar predição genômica e descoberta de genes em estudos de associação genômica (GWAS). Os recentes avanços na genômica do café com o sequenciamento completo do genoma de referência de *C. canephora* forneceram à comunidade científica do café o recurso necessário para se desenvolver um

conjunto de ferramentas moleculares para genotipagem de SNPs em escala genômica (ANDRADE et al., 2017). *C. canephora*, uma espécie diploide alógama e um dos parentais do alotetraploide *C. arabica*, tem sido uma importante fonte de variabilidade genética para os programas de melhoramento genético de ambas as espécies cultivadas. Os genomas altamente heterozigóticos, como *C. canephora*, requerem uma profundidade de sequência muito maior para alcançar taxas de chamada de marcador aceitáveis e precisão do genótipo, quando se utilizam métodos de genotipagem baseados em sequência, de modo que a sua relação custo-benefício não seja realizada. Desta forma, o desenvolvimento de *chips* de genotipagem de DNA, com milhares de SNPs, tem aumentado significativamente, não somente em organismos modelos, mas também em inúmeras espécies que ainda não possuem os dados genômicos disponíveis (GUPTA et al., 2013).

## 2.5 Estudo de associação genômica ampla

A relação causal entre o polimorfismo genético dentro de uma espécie vegetal e as diferenças fenotípicas observadas entre indivíduos é de fundamental interesse biológico, para se prever os fatores genéticos associados com as características de interesse agrônomico e, portanto, requer uma compreensão tanto dos loci específicos que estão associados a um fenótipo, assim como a arquitetura genética das características de interesse (KORTE; FARLOW, 2013).

Um mapa genético ou de ligação pode ser visto como um "mapa rodoviário" dos cromossomos de uma população segregante. A aplicação mais importante de um mapa genético visa identificar QTLs (*Quantitative Trait Loci*) associados às características agrônomicas de interesse (LEONFORTE et al., 2013). O mapeamento de QTL baseia-se no princípio de que as características fenotípicas e os marcadores de DNA segregam concomitantemente durante a meiose, permitindo sua análise na progênie (PATERSON, 2002). As características e genes / marcadores que estão próximos ou geneticamente ligados serão transmitidos de genitores para progênies, mais frequentemente, do que os genes / marcadores que estão mais distantes (KEARSEY; FARQUHAR, 1998). Desta forma, a análise de ligação e o mapeamento de associação são ferramentas importantes e comumente utilizadas para a identificação de características complexas para a agricultura. A variação fenotípica dessas características é influenciada por QTLs, suas interações, além da influência do ambiente e da relação entre QTLs e ambiente (ZHU et al, 2008).



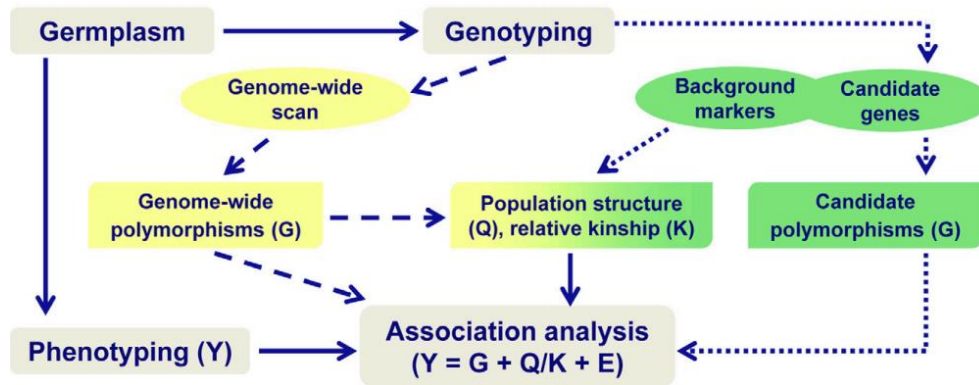
A genética direta (*Forward Genetics*), em que muitos indivíduos que diferem em genótipos são rastreados para fenótipos de interesse, tem sido uma poderosa ferramenta para abordar a relação entre genótipo e fenótipo. Em geral, as diferenças genéticas cruciais que estão sendo selecionadas são obtidas por mutagênese ou amostradas de uma população natural (KORTE; FARLOW, 2013). Todas as diferenças fenotípicas identificadas são conectadas aos loci causadores por meio de várias abordagens de mapeamento, incluindo o mapeamento de QTL. Hoje em dia, com os recentes avanços em genômica, *Genome-Wide Association Studies* (GWAS) ou estudo de associação genômica ampla, tornou-se possível e pode ser considerada uma ferramenta complementar e poderosa para conectar o mapa genótipo-fenótipo. No entanto, o mapeamento de QTL provou, e permanece, um método poderoso para se identificar regiões do genoma que co-segregam com uma determinada característica em uma determinada população.

Embora o GWAS tenha potencial para identificar polimorfismos associados a características importantes na agricultura, três fatores têm grande influência nas análises. O primeiro é o tamanho da amostra, que afeta o poder estatístico do estudo, especialmente na detecção de polimorfismos de pequeno efeito subjacentes às características mais complexas (KORTE; FARLOW, 2013). Os resultados apresentados por Goddard et al. (2016) enfatizam que, como a maioria dos polimorfismos explica uma pequena fração da variância (por exemplo, menos de 1%), para alcançar o nível de significância convencional ( $p < 5 \times 10^{-8}$ ), grandes tamanhos de amostra são necessários.

O segundo fator é a estrutura populacional e o parentesco entre os indivíduos, visto que estes podem levar a falsas descobertas devido a associações espúrias (ZHANG et al., 2010). O terceiro, e último fator, é o número adequado de marcadores distribuídos ao longo de todos os cromossomos. Recentemente, as plataformas de sequenciamento de próxima geração reduziram drasticamente o custo e o tempo na genotipagem de um grande número de marcadores cobrindo todo o genoma (DAVEY, et al., 2011).

O estudo de associação é particularmente importante em espécies perenes, como o cafeeiro, podendo ser feito em genes candidatos, relacionando polimorfismos nesses genes com o controle da variação fenotípica para características específicas, ou analisando-se a variação genética em todo o genoma para encontrar sinais de associação para várias características complexas (FIGURA 6). A escolha se o mapeamento de associação pode ser feito em todo o genoma ou somente em genes candidatos depende do foco e da escala do estudo realizado (CUBRY et al, 2013; RISCH; MERIKANGAS, 1996).

Figura 6 – Diagrama esquemático mostrando o mapeamento de associação de todo o genoma e o mapeamento em genes candidatos.



Fonte: Zhu et al, (2008).

Em *C. canephora* vários estudos de QTL foram realizados, como o mapeamento do locus de auto-incompatibilidade (COULIBALY et al., 2002; LASHERMES et al., 1996b), viabilidade do pólen (COULIBALY et al., 2003), características de floração (NUGROHO, 2013), produção e característica relacionadas com a qualidade do grão (LEROY et al., 2011; MÉROT-L'ANTHOËNE et al., 2014), assim como características relacionadas com o sistema radicular (ACHAR et al., 2015), entre outros.

Apesar desse sucesso, o mapeamento de QTLs sofre duas limitações fundamentais: (i) somente a diversidade alélica que segrega entre os pais do cruzamento específico utilizado nas análises pode ser testada (BOREVITZ; NORDBORG, 2003) e (ii) a taxa de recombinação que ocorre durante a criação da população coloca um limite na resolução do mapeamento (KORTE; FARLOW, 2013).

## 2.6 Seleção Genômica Ampla

O grande atrativo da genética molecular em benefício do melhoramento genético aplicado é a utilização direta das informações de DNA na seleção, de forma a permitir alta eficiência seletiva, grande rapidez na obtenção de ganhos genéticos com a seleção e baixo custo, em comparação com a tradicional seleção baseada em dados fenotípicos (RESENDE et al., 2008). Com o propósito de atender a esses objetivos, a seleção genômica (SG) ou seleção genômica ampla (*genome wide selection* – GWS) foi proposta por Meuwissen, Goddard e Hayes (2001). Este método pode ser aplicado em todas as populações em avaliação nos programas de melhoramento genético de espécies alógamas tal como *C. canephora*, apresenta alta acurácia seletiva para a seleção baseada exclusivamente em marcadores (após terem seus efeitos genéticos estimados a partir de dados fenotípicos em uma amostra da população de seleção) e não exige prévio conhecimento das posições (mapa) dos QTLs, não estando sujeita aos erros tipo II associados à seleção de marcadores ligados a QTLs (RESENDE et al., 2008). Esse método permaneceu discreto por alguns anos, devido à limitação dos marcadores moleculares disponíveis à época. Recentemente, com o desenvolvimento e baixo custo dos marcadores tipo SNP, o método tornou-se muito atrativo e, geneticistas e melhoristas renomados e adeptos de métodos tradicionais, têm demonstrado e confirmado a superioridade e exequibilidade prática do método em benefício do melhoramento animal (LEGARRA; MISZTAL, 2008; LONG et al., 2008) e vegetal (BERNARDO; YU, 2007). Esses trabalhos mostraram, definitivamente, que a seleção genômica terá grande utilidade no melhoramento genético, via métodos do tipo BLUP/GWS, que equivalem ao procedimento BLUP (melhor predição linear não viciada) aplicado sobre dados moleculares e permitem a predição de valores genéticos genômicos (RESENDE et al., 2008). A GWS é excelente para caracteres de baixa herdabilidade, ao contrário da Seleção Assistida por Marcadores (SAM), que não é útil para caracteres de baixa herdabilidade (MUIR, 2007).

A SAM tem por base, a detecção, mapeamento e uso de QTLs de grande efeito na seleção, focando na determinação do número, posição e efeitos dos QTLs marcados. Já a GWS é definida como a seleção simultânea para centenas ou milhares de marcadores, os quais cobrem o genoma de uma maneira densa, de forma que todos os genes de um caráter quantitativo estejam em desequilíbrio de ligação com pelo menos uma parte dos marcadores. Esses marcadores em desequilíbrio de ligação com os QTLs, tanto de grandes quanto de pequenos efeitos, explicarão quase a totalidade da variação genética de um caráter quantitativo (RESENDE et al., 2008). Devido ao número elevado de SNPs, a probabilidade de

se encontrar um QTL em desequilíbrio de ligação com pelo menos um marcador é muito alta. Desta forma, somente os marcadores em desequilíbrio de ligação com os QTLs serão úteis na determinação dos fenótipos e na explicação da variação genética. Os efeitos dos marcadores são estimados em uma amostra de indivíduos pertencentes a várias famílias e, sendo assim, o impacto de famílias específicas (com padrões específicos de desequilíbrio de ligação) nas estimativas dos efeitos dos marcadores será minimizado. É importante enfatizar que os marcadores terão seus efeitos genéticos estimados a partir de uma amostra de pelo menos 1.000 indivíduos genotipados e fenotipados, ou seja, com base em pelo menos 1.000 repetições experimentais (RESENDE et al., 2012). Assim, embora a herdabilidade de cada marcador efetivo (aquele que identifica um dos polígenes com precisão) seja muito baixa, com 1.000 repetições, essa herdabilidade torna-se alta. Em outras palavras, o efeito de ambiente será minimizado por meio do uso de um número de repetições muito alto. Essa é a mesma filosofia da avaliação e seleção de características quantitativas com base em fenótipos em experimentos de campo (RESENDE et al., 2008).

A SGA é ampla porque atua em todo o genoma, capturando todos os genes que afetam um caráter quantitativo. E isso sem a necessidade prévia de identificar os marcadores com efeitos significativos e de mapear QTLs, como no caso da SAM. Valores genéticos genômicos, associados a cada marcador ou alelo, são usados para prever o valor genético genômico (VGG) global de cada indivíduo. Há uma diferença básica na predição de valores genéticos tradicionais e na predição de valores genéticos genômicos. No primeiro caso, informações fenotípicas são utilizadas visando inferências sobre os efeitos dos genótipos dos indivíduos e, nos últimos, informações genotípicas (genótipos para os alelos marcadores) são usadas visando às inferências sobre os valores fenotípicos futuros (ou valores genéticos genômicos preditos) dos indivíduos. Em outras palavras, os métodos tradicionais usam o fenótipo para inferir sobre o efeito do genótipo e a SGA usa o genótipo, com efeito genético pré-estimado em uma amostra da população, para inferir sobre o fenótipo a ser expresso nos candidatos à seleção (RESENDE et al., 2008).

Segundo Crossa et al., (2010), os efeitos dos marcadores não serão necessariamente os mesmos em diferentes estudos e ambientes. Na SGA, os efeitos genéticos dos marcadores são estimados e usados na seleção para cada população de melhoramento e em um determinado ambiente. Modelos de predição, incluindo a interação genótipos x ambientes, podem também ser usados, visando verificar a possibilidade de se obter estimativas válidas para um conjunto de ambientes. Mas isso dependerá da magnitude da interação envolvendo os vários ambientes.

A SGA pode basear-se no uso de: (i) apenas dos marcadores; (ii) de haplótipos ou intervalos definidos por dois marcadores; (iii) haplótipos definidos por mais de dois marcadores, incluindo a covariância entre haplótipos devida à ligação. Segundo Calus et al., (2008), para caracteres de baixa herdabilidade (10%) não existem diferenças significativas entre essas três abordagens. Solberg et al., (2008) mostraram que é possível praticar a SGA eficientemente com o uso apenas dos marcadores, ou seja, com a predição direta dos efeitos dos marcadores. Relatam também que isso é vantajoso porque não há necessidade de estimar as fases de ligação entre os marcadores, as quais são estimadas com algum erro. Não apenas marcadores SNPs podem ser usados na SGA, marcadores microssatélites também se prestam a esse fim. Solberg et al., (2008) relatam que o uso de SNPs requer quatro a cinco vezes maior densidade de marcadores do que o uso de microssatélites. Isto se deve à natureza bi-alélica (binucleotídica) dos SNPs e multi-alélica dos microssatélites.

Os enormes avanços tecnológicos experimentados na genômica do gênero *Coffea* na última década proporcionaram aos pesquisadores do café as ferramentas moleculares necessárias para traduzir a genômica em aplicações práticas para auxílio aos programas de melhoramento genético. Estudos de diversidade em *C. canephora* da maioria dos materiais cultivados têm mostrado pequenos sinais de domesticação, o que significa que ainda há um espaço enorme para a exploração da conhecida diversidade genética natural aplicada ao melhoramento genético do café robusta. As ferramentas genômicas disponíveis agora podem ser empregadas para promover e acelerar os programas de melhoramento de *C. canephora* com uma ótima previsão de sucesso. À medida que os desafios das mudanças climáticas avançam, a aplicação dessas novas tecnologias visam garantir a sustentabilidade do café, proporcionando alternativas viáveis aos cafeicultores.

## REFERÊNCIAS

- ACHAR, D. et al. Identification of Putative Molecular Markers Associated with Root Traits in *Coffea canephora* Pierre ex Froehner. **Molecular Biology International**, v. 2015, p. 1-11, 2015.
- ALONSO-SALCES, R. M. et al. Botanical and geographical characterization of green coffee (*Coffea arabica* and *Coffea canephora*): chemometric evaluation of phenolic and methylxanthine contents. **Journal of Agricultural and Food Chemistry**, v. 57, n. 10, p. 4224-4235, 2009.
- ANDRADE, A. C. et al. 'Towards GWAS and Genomic Prediction in Coffee: Development and Validation of a 26K SNP Chip for *Coffea canephora*', In XX INTERNATIONAL PLANT AND ANIMAL GENOME CONFERENCE, p. W173. **Anais eletrônicos...** Disponível em: <<https://pag.confex.com/pag/xxv/webprogram/Paper23677.html>>
- ANDRADE, A. C. Developments in molecular breeding techniques in Robusta coffee In: LASHERMES, P. **Achieving sustainable cultivation of coffee: Breeding and quality traits**. Cambridge: Burleigh Dodds Science Publishing, 2018. p.
- ANTHONY, F. et al. Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers. **Euphytica**, v. 118, n. 1, p. 53-65, 2001.
- ASQUINI, E. et al. S-RNase-like sequences in styles of *Coffea* (Rubiaceae). Evidence for S-RNase based gametophytic self-incompatibility? **Tropical Plant Biology**, v. 4, p. 237-249, 2011.
- BASSIL, N. V. et al. Development and preliminary evaluation of a 90 K Axiom(R) SNP array for the allo-octoploid cultivated strawberry *Fragaria x ananassa*. **BMC Genomics**, v. 16, n. 1, p. 1-30, 2015.
- BELAN, L. L. et al. Aspectos fisiológicos do cafeeiro conilon: uma abordagem sistemática. **Nucleus**, v. 8, n. 1, p. 225-240, 2011.
- BERNARDO, R.; YU, J. Prospects for genomewide selection for quantitative traits in maize. **Crop Science**, v. 47, n. 3, p. 1082-1090, 2007.
- BERTHAUD, J. **Les ressources génétiques pour l'amélioration des caféiers africains diploïdes: évaluation de la richesse génétique des populations sylvestres et de ses mécanismes organisateurs. Conséquences pour l'application**. 1986. 379 p. Tese doutorado
- BERTRAND, B. et al. Impact of the *Coffea canephora* gene introgression on beverage quality of *C. arabica*. **Theoretical and Applied Genetics**, v. 107, n. 3, p. 387-94, 2003.
- BIANCO, L. et al. Development and validation of the Axiom® Apple480K SNP genotyping array. **The Plant Journal**, v. 86, n. 1, p. 62-74, 2016.
- BORÉM, A.; CAIXETA, E. T. **Marcadores moleculares**. Viçosa: UFV, 2009. 532 p.

BOREVITZ, J. O.; NORDBORG, M. The Impact of Genomics on the Study of Natural Variation in Arabidopsis. **Plant Physiology**, v. 132, n. 2, p. 718–725, 2003.

BRAGANÇA, S. M. et al. EMCAPA 8111, EMCAPA 8121, EMCAPA 8131: Variedades clonais de café Conilon para o Estado do Espírito Santo. **Pesquisa Agropecuária Brasileira**, v. 36, n. 05, p. 765-770, 2001.

CALUS, M. P. L. et al. Accuracy of genomic selection using different methods to define haplotypes. **Genetics**, v. 178, n. 1, p. 553–561, 2008.

CAMARGO, M. B. P. D. The impact of climatic variability and climate change on arabic coffee crop in Brazil. **Bragantia**, v. 69, p. 239-247, 2010.

CARNEIRO, F. A. et al. Phenotyping and genotyping a *Coffea canephora* population, cultivated at high altitude, aiming at a GWS program for coffee. In: THE 25TH INTERNATIONAL CONFERENCE ON COFFEE SCIENCE, 2014, Colômbia.

COMBES, M. C. et al. Regulatory divergence between parental alleles determines gene expression patterns in hybrids. **Genome Biology and Evolution**, v. 7, n. 4, p. 1110-1121, 2015.

CONAB - Companhia Nacional de Abastecimento. Disponível em: [www.conab.gov.br/](http://www.conab.gov.br/)> Acesso em 18 de jan. de 2019.

COTTA, M. G. et al. Lipid transfer proteins in coffee: isolation of *Coffea* orthologs, *Coffea arabica* homeologs, expression during coffee fruit development and promoter analysis in transgenic tobacco plants. **Plant Molecular Biology**, v. 85, n. 1-2, p. 11-31, 2014.

COULIBALY, I. et al. Introgression of self-compatibility from *Coffea heterocalyx* to the cultivated species *Coffea canephora*. **Theoretical and Applied Genetics**, v. 105, n. 6–7, p. 994–999, 2002.

COULIBALY, I. et al. Pollen viability restoration in a *Coffea canephora* P. and C-heterocalyx Stoffelen backcross. QTL identification for marker-assisted selection. **Theoretical and Applied Genetics**, v. 106, p. 311–316, 2003.

CROSSA, J. et al. Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers. **Genetics**, v. 186, n. 2, p. 713–724, 5 out. 2010.

CRUZ, V. M. V; KILIAN, A.; DIERIG, D. A. Development of DArT Marker Platforms and Genetic Diversity Assessment of the U.S. Collection of the New Oilseed Crop *Lesquerella* and Related Species. **PLoS ONE**, v. 8, n. 5, 2013.

CUBRY, P. et al. An initial assessment of linkage disequilibrium (LD) in coffee trees: LD patterns in groups of *Coffea canephora* Pierre using microsatellite analysis. **BMC Genomics**, v. 14, 2013.

CUBRY, P. et al. Global analysis of *Coffea canephora* Pierre ex Froehner (Rubiaceae) from the Guineo-Congolese region reveals impacts from climatic refuges and migration effects. **Genetic Resources and Crop Evolution**, v. 60, n. 2, p. 483-501, 2012.

- CUBRY, P. et al. Diversity in coffee assessed with SSR markers: structure of the genus *Coffea* and perspectives for breeding. **Genome**, v. 51, n. 1, p. 50-63, 2008.
- DAMATTA, F. M. et al. Impacts of climate changes on crop physiology and food quality. **Food Research International**, v. 43, n. 7, p. 1814-1823, 2010.
- DAMATTA, F. M. Ecophysiological constraints on the production of shaded and unshaded coffee: a review. **Field Crops Research**, v. 86, p. 99-114, 2004.
- DAVEY, J. W. et al. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. **Nature Reviews Genetics**, v. 12, n. 7, p. 499-510, 2011.
- DAVIS, A. P. et al. The Impact of Climate Change on Indigenous Arabica Coffee (*Coffea arabica*): Predicting Future Trends and Identifying Priorities. **PLoS ONE**, v. 7, n. 11, 2012.
- DAVIS, A. P. et al. An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). **Botanical Journal of the Linnean Society**, v. 152, n. 4, p. 465-512, 2006.
- DE KOCHKO, A. et al. Advances in *Coffea* Genomics. In: KADER, J. C.; DELSENY, M. (Ed.). **Advances in Botanical Research**, v. 53. Londres: Elsevier Science, 2010, p.23-63.
- DENOEUDE, F. et al. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. **Science**, v. 345, n. 6201, p. 1181-1184, 2014.
- DUSSERT, S. et al. 'Le caféier, *Coffea canephora*', In P. HAMON, P. et al.(Eds). Diversité génétique des plantes tropicales cultivées. CIRAD, Montpellier, pp. 175-794, 1999.
- FAZUOLI, L. C. et al. Seleção antecipada e sua eficiência no café Icatu. In: I SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL. 2000, Poços de Caldas. p.576-584.
- FERRÃO, R. G. et al. 'DIAMANTE ES 8112', 'ES 8122 – JEQUITIBÁ' e 'CENTENÁRIA ES 8132': novas cultivares clonais de café conilon com qualidade de bebida para o espírito santo. In: IX SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 2015, Curitiba. **Anais...** Curitiba: SPCB 9, 2015.
- FERRÃO, L. F. V. et al. Comparative study of different molecular markers for classifying and establishing genetic relationships in *Coffea canephora*. **Plant Systematics and Evolution**, v. 299, n. 1, p. 225-238, 3 jan. 2013.
- FERRÃO, L. F. V. et al. The effects of encoding data in diversity studies and the applicability of the weighting index approach for data analysis from different molecular markers, **Plant Systematics and Evolution**, v. 300, n. 7, p. 1649-61, 2014.
- FERRÃO, M. A. G. et al. Origem, dispersão geográfica, taxonomia e diversidade genética de *Coffea canephora*. In: FERRÃO, R. G. et al (Ed.). **Café Conilon**, 2007. cap. 3, p.65-91.
- FERRÃO, R. G. et al. Melhoramento genético de *Coffea canephora*. In: FERRÃO, R. G. et al (Ed.). **Café Conilon**, 2007. cap. 5, p.121-173.



FERRÃO, R. G. et al. EMCAPA 8141 - Robustão Capixaba, variedade clonal de café Conilon tolerante à seca, desenvolvida para o estado do Espírito Santo. **Ceres**, v. 47, p. 555-560, 2000.

FERREIRA, M. E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. Brasília: Embrapa Recursos Genéticos e Biotecnologia, 1998. 220 p.

FONSECA, A. F. A. et al. Conilon Vitória - Incaper 8142': improved *Coffea canephora* var. *kouillou* clone cultivar for the state of Espírito Santo. **Crop Breeding and applied Biotechnology**, v. 04, n. 4, 2004.

GANAL, M. W. et al. High-throughput SNP Profiling of Genetic Resources in Crop Plants Using Genotyping Arrays. In: TUBEROSA, R.; GRANER, A., et al (Ed.). **Genomics of Plant Genetic Resources: Volume 1. Managing, sequencing and mining genetic resources**. Springer, Dordrecht, 2014. p.113-130.

GARAVITO, A. et al. Identification by the DArTseq method of the genetic origin of the *Coffea canephora* cultivated in Vietnam and Mexico. **BMC Plant Biology**, v. 16, p. 242-254, 2016.

GAY, C. et al. Potential Impacts of Climate Change on Agriculture: A Case of Study of Coffee Production in Veracruz, Mexico. **Climatic Change**, v. 79, n. 3-4, p. 259–288, 2006.

GHINI, R. et al. Research approaches, adaptation strategies, and knowledge gaps concerning the impacts of climate change on plant diseases. **Tropical Plant Pathology**, v. 37, n. 1, p. 5-24, 2012.

GHINI, R.; BETTIOL, W.; HAMADA, E. Diseases in tropical and plantation crops as affected by climate changes: current knowledge and perspectives. **Plant Pathology**, v. 60, n. 1, p. 122-132, fev. 2011.

GODDARD, M. E. et al. Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture. **Proceedings. Biological Sciences**, v. 283, 2016.

GOMEZ, C. et al. Current genetic differentiation of *Coffea canephora* Pierre ex A. Froehn in the Guineo-Congolian African zone: cumulative impact of ancient climatic changes and recent human activities. **BMC Evolutionary Biology** v. 9, p. 1-19, 2009.

GUPTA, P. K.; RUSTGI, S.; MIR, R. R. Array-Based High-Throughput DNA Markers and Genotyping Platforms for Cereal Genetics and Genomics. In: GUPTA, P. K.; VARSHNEY, R. K. (Ed.). **Cereal Genomics II**. Springer, Dordrecht, 2013. p.11-55.

HENDRE, P. S. et al. Development of new genomic microsatellite markers from robusta coffee (*Coffea canephora* Pierre ex A. Froehner) showing broad cross-species transferability and utility in genetic studies. **BMC Plant Biology**, v. 8, 2008.

HULSE-KEMP, A. M. et al. Development of a 63K SNP Array for Cotton and High-Density Mapping of Intraspecific and Interspecific Populations of *Gossypium* spp. **G3: Genes, Genomes, Genetics**, v. 5, n. 6, p. 1187-1209, 2015.

ICO – International Coffee Organization. Disponível em: <<http://www.ico.org/>>. Acesso em: 15 fev. 2019.

IPCC, 2013: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp, doi:10.1017/CBO9781107415324.

JARAMILLO, J. et al. Some like it hot: the influence and implications of climate change on coffee berry borer (*Hypothenemus hampei*) and coffee production in East Africa. **PLoS One**, v. 6, n. 9, e24528, 2011.

KEARSEY, M. J.; FARQUHAR, A. G. QTL analysis in plants; where are we now? **Heredity**, v. 80, p. 137-142, 1998.

KIRSCHBAUM, M. U. F. Does Enhanced Photosynthesis Enhance Growth? Lessons Learned from CO<sub>2</sub> Enrichment Studies. **Plant Physiology**, v. 155, n. 1, p. 117-124, 2011.

KONING-BOUCOIRAN, C. F. et al. Using RNA-Seq to assemble a rose transcriptome with more than 13,000 full-length expressed genes and to develop the WagRhSNP 68k Axiom SNP array for rose (*Rosa L.*). **Frontiers in Plant Science**, v. 6, n. 249, p. 1-10, 2015.

KORTE, A.; FARLOW, A. The advantages and limitations of trait analysis with GWAS : a review. **Plant Methods**, v. 29, p. 1-9, 2013.

KUTYWAYO, D. et al. The Impact of Climate Change on the Potential Distribution of Agricultural Pests: The Case of the Coffee White Stem Borer (*Monochamus leuconotus P.*) in Zimbabwe. **PLoS ONE**, v. 8, n. 8, p. 1-11, 2013.

LANNES, S. D. et al. Polimorfismos nucleotídicos de genes envolvidos nas características químicas do grão de café. Complementaridade das estratégias in silico e in vivo. In: V SIMPÓSIO BRASILEIRO DE PESQUISA DOS CAFÉS DO BRASIL, 2007, Águas de Lindóia.

LASHERMES, P. et al. Molecular characterisation and origin of the *Coffea arabica L.* genome. **Molecular and General Genetics**, v. 261, n. 2, p. 259-266, 1999.

LEGARRA, A.; MISZTAL, I. Technical Note: Computing Strategies in Genome-Wide Selection. **Journal of Dairy Science**, v. 91, n. 1, p. 360–366, 2008.

LASHERMES, P. et al. Inheritance and genetic mapping of self-incompatibility in *Coffea canephora* Pierre. **Theoretical and Applied Genetics**, v. 93, n. 3, p. 458-462, 1996a.

\_\_\_\_\_. Inheritance and genetic mapping of self-incompatibility in *Coffea canephora* Pierre. **Theoretical and Applied Genetics**, 1996b.

LEE, Y. G. et al. Development, validation and genetic analysis of a large soybean SNP genotyping array. **The Plant Journal**, v. 81, n. 4, p. 625-636, 2015.

LEONFORTE, A. et al. SNP marker discovery, linkage map construction and identification of QTLs for enhanced salinity tolerance in field pea (*Pisum sativum* L.). **BMC Plant Biology**, v. 13, n. 1, p. 161, 2013.

LEROY, T. et al. Improving the quality of African robustas: QTLs for yield- and quality-related traits in *Coffea canephora*. **Tree Genetics & Genomes**, v. 7, n. 4, p. 781-798, 2011.

LI, F. et al. A *Brassica rapa* linkage map of EST-based SNP markers for identification of candidate genes controlling flowering time and leaf morphological traits. **DNA Research**, v. 16, n. 6, p. 311-323, 2009.

LIMA, E. A. et al. The Multi-Resistant Reaction of Drought-Tolerant Coffee “Conilon Clone 14” to *Meloidogyne* spp. and Late Hypersensitive-Like Response in *Coffea canephora*. **Phytopathology**, v. 105, n. 6, p. 805-814, 2015.

LONG, N. et al. Machine learning classification procedure for selecting SNPs in genomic selection: application to early mortality in broilers. **Developments in biologicals**, v. 132, p. 373–6, 2008.

MARRACCINI, P. et al. Differentially expressed genes and proteins upon drought acclimation in tolerant and sensitive genotypes of *Coffea canephora*. **Journal of Experimental Botany**, v. 63, n. 11, p. 4191-4212, 2012.

MARRACCINI, P. et al. *RBCS1* expression in coffee: *Coffea* orthologs, *Coffea arabica* homeologs, and expression variability between genotypes and under drought stress. **BMC Plant Biology**, v. 11, p. 1-23, 2011.

MÉROT-L'ANTHOËNE, V. et al. Comparison of three QTL detection models on biochemical, sensory, and yield characters in *Coffea canephora*. **Tree Genetics & Genomes**, v. 10, n. 6, p. 1541–1553, 25 dez. 2014.

MEROT-L'ANTHOENE, V. et al. Development and evaluation of a genome-wide Coffee 8.5K SNP array and its application for high-density genetic mapping and for investigating the origin of *Coffea arabica* L. **Plant Biotechnology Journal**, p. 1-13, 2019.

MEUWISSEN, T. H. E.; GODDARD, M. E.; HAYES, B. J. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, n. 4, p. 1819-1829, 2001.

MILACH, S. C. K. **Marcadores moleculares em plantas**. Porto Alegre: UFGRS, 1998. 140 p.

MISHRA, M. K.; SLATER, A. Recent advances in the genetic transformation of coffee. **Biotechnology Research International**, v. 2012, p. 1-17, 2012.

MONDEGO, J. M. C. et al. An EST-based analysis identifies new genes and reveals distinctive gene expression features of *Coffea arabica* and *Coffea canephora*. **BMC Plant Biology**, v. 11, 2011.

MONTAGNON, C. **Optimisation des gains génétiques dans le schéma de sélection récurrente réciproque de *Coffea canephora* Pierre**. 2000. 42 p. Tese doutorado Montpellier University

MONTAGNON, C.; CUBRY, P.; LEROY, T. Coffee *Coffea canephora* Pierre genetic improvement: acquired knowledge, strategies and perspectives. **Cahiers Agricultures**, v. 21, n. 2-3, p. 143-153, 2012.

MONTAGNON, C.; LEROY, T.; YAPO, A. Diversité génotypique et phénotypique de quelques groupes de caféiers (*Coffea canephora* Pierre) en collection. Conséquences sur leur utilisation en sélection. **Café, Cacao, Thé**, v. 36, n. 3, p. 187-198, 1992.

MUIR, W. M. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. **Journal of Animal Breeding and Genetics**, v. 124, n. 6, p. 342–355, 2007.

MUSOLI, P. et al. Genetic differentiation of wild and cultivated populations: diversity of *Coffea canephora* Pierre in Uganda. **Genome**, v. 52, n. 7, p. 634-646, 2009.

NOWAK, M. D. et al. Expression and trans-specific polymorphism of self-incompatibility RNases in *Coffea* (Rubiaceae). **Plos One**, v. 6, n. 6, 2011.

NUGROHO, D. Identification of Quantitative Trait Loci (QTLs) Determining Flowering in the Robusta Coffee. **Journal of Agricultural Science and Technology**, v. 3, p. 296-305, 2013.

PATERSON, A. H. What has QTL mapping taught us about plant domestication? **New Phytologist**, v. 154, n. 3, p. 591-608, 2002.

PICINI, A. G. **Desenvolvimento e teste de modelos agrometeorológicos para estimativa de produtividade do cafeeiro (*Coffea arabica* L.) a partir do monitoramento da disponibilidade hídrica do solo**. 1998. 132 p. Dissertação (mestrado)- Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1998.

POLAND, J. A.; RIFE, T. W. Genotyping-by-Sequencing for Plant Breeding and Genetics. **Plant Genome**, v. 5, n. 3, p. 92-102, 2012.

PONCET, V. et al. Development of genomic microsatellite markers in *Coffea canephora* and their transferability to other coffee species. **Genome**, v. 50, n. 12, p. 1156-61, 2007.

PONCET, V. et al. SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. **Molecular Genetics and Genomics**, v. 276, n. 5, p. 436-49, 2006.

PRAKASH, N. et al. Analysis of genetic diversity in Indian robusta coffee genepool (*Coffea canephora*) in comparison with a representative core collection using SSRs and AFLPs. **Genetic Resources and Crop Evolution**, v. 52, n. 3, p. 333-343, 2005.

RAFALSKI, J. A. Novel genetic mapping tools in plants: SNPs and LD-based approaches. **Plant Science**, v. 162, n. 3, p. 329-333, 2002.

RAMALHO, A. R. et al. Conilon ‘BRS Ouro Preto’: cultivar clonal de café para Amazônia Ocidental. In: IX SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 2015, Curitiba. **Anais...** Curitiba: SPCB 9, 2015.

RENA, A. B.; MAESTRI, M. Fisiologia do cafeeiro. **Informe agropecuário**, v. 11, p. 26-40, 1985.

RESENDE, M. D. V. et al. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal Brasileira**, n. 56, p. 63-77, 2008.

RESENDE, M. D. V. et al. Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees. **New Phytologist**, v. 194, n. 1, p. 116-128, 2012.

RESENDE JR., M. F. R. et al. Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. **New Phytologist**, v. 193, n. 3, p. 617–624, fev. 2012.

RISCH, N.; MERIKANGAS, K. The future of genetic studies of complex human diseases. **Science**, v. 273, p. 1516-1517, 1996.

SCHLOTTERER, C. The evolution of molecular markers - just a matter of fashion? **Nature Reviews Genetics**, v. 5, n. 1, p. 63-69, 2004.

SCHROTH, G. et al. Towards a climate change adaptation strategy for coffee communities and ecosystems in the Sierra Madre de Chiapas, Mexico. **Mitigation and Adaptation Strategies for Global Change**, v. 14, n. 7, p. 605–625, 2009.

SILVA, E. A. et al. Influência de déficits hídricos controlados na uniformização do florescimento e produção do cafeeiro em três diferentes condições edafoclimáticas do estado de São Paulo. **Bragantia**, v. 68, n. 2, p. 493-501, 2009.

SOLBERG, T. R. et al. Genomic selection using different marker types and densities. **Journal of Animal Science**, v. 86, n. 10, p. 2447–2454, 2008.

SOLEIMANI, V. D.; BAUM, B. R.; JOHNSON, D. A. Efficient validation of single nucleotide polymorphisms in plants by allele-specific PCR, with an example from barley. **Plant Molecular Biology Reporter**, v. 21, n. 3, p. 281-288, 2003.

SUMIRAT, U. et al. Genetic diversity assesment in Indonesian *Coffea canephora* collection using SSR Markers. In: 24TH INTERNATIONAL CONFERENCE ON COFFEE SCIENCE, 2012, Costa Rica. CD-ROM, PB215.

THUDI, M. et al. Current state-of-art of sequencing technologies for plant genomics research. **Briefings in Functional Genomics** v. 11, n. 1, p. 3-11, 2012.

VIDAL, R. O. et al. A high-throughput data mining of single nucleotide polymorphisms in *Coffea* species expressed sequence tags suggests differential homeologous gene expression in the allotetraploid *Coffea arabica*. **Plant Physiology**, v. 154, n. 3, p. 1053-1066, 2010.

VIEIRA, N. G. et al. Different Molecular Mechanisms Account for Drought Tolerance in *Coffea canephora* var. Conilon. **Tropical Plant Biology**, v. 6, n. 4, p. 181-190, 2013.

VIEIRA, L. G. E. et al. Brazilian coffee genome project: an EST-based genomic resource. **Brazilian Journal of Plant Physiology**, v. 18, n. 1, p. 95-108, 2006.

ZHANG, Z. et al. Mixed linear model approach adapted for genome-wide association studies. **Nature Genetics**. v. 42, p. 355-360, 2010.

ZHU, C. et al. Status and Prospects of Association Mapping in Plants. **The Plant Genome**, v. 1, n. 1, p. 5-20, 2008.

ZULLO, J. et al. Potential for growing Arabica coffee in the extreme south of Brazil in a warmer world. **Climatic Change**, v. 109, n. 3-4, p. 535–548, 2011.

### 3 OBJETIVOS

#### 3.1 Objetivo geral

Desenvolver e validar modelos preditivos de seleção genômica ampla (SGA) para produção, características físicas e bioquímicas dos grãos e tolerância à seca em *C. canephora*. Os modelos desenvolvidos no âmbito deste projeto poderão, na continuidade dos programas de melhoramento desta espécie, (1) acelerar significativamente os programas de seleção recorrente de *C. canephora*, dispensando o tempo necessário de testes de progênie e (2) permitir a clonagem imediata de indivíduos elite, selecionados ainda no estágio de mudas viabilizando a produção de clones elite de *C. canephora* em larga escala.

#### 3.2 Objetivos específicos

- ✓ Avaliar os dados fenotípicos agronômicos e fisiológicos (produção e tolerância à seca) e de propriedades físicas dos grãos em indivíduos selecionados em uma população de melhoramento de *C. canephora*, cultivada no campo experimental da Embrapa Cerrados (Planaltina-DF).
- ✓ Identificar SNPs em um grupo de genótipos representando a diversidade genética global de *C. canephora*.
- ✓ Construir de uma plataforma de genotipagem baseada no “Chip Axiom-Affymetrix” a partir dos SNPs identificados.
- ✓ Genotipar em larga escala dos 1.300 indivíduos em estudo, utilizando o *chip* desenvolvido.
- ✓ Realizar o estudo de associação genômica ampla para as características de interesse;
- ✓ Gerar e validar os modelos de predição de Seleção Genômica Ampla para as diversas características fenotípicas avaliadas na população.

## CAPÍTULO 2

### FENOTIPAGEM DE INDIVÍDUOS DE UMA POPULAÇÃO DE *Coffea canephora*

#### 1 INTRODUÇÃO

De acordo com dados das Nações Unidas, a população humana deverá crescer acima de 9 bilhões até 2050 e garantir que a produção de alimentos seja suficiente para suprir as necessidades dessa população é um grande desafio para a ciência. Além da pressão para o aumento da produção agrícola, existem ainda outras ameaças, como a escassez de água, restrições de fertilidade do solo e mudanças climáticas (COBB et al., 2013). Na tentativa de aumentar a produtividade mesmo diante de tantos obstáculos, os programas de melhoramento atuais baseiam-se na seleção de novas cultivares que possuam características promissoras, como alta produtividade, tolerância a estresses bióticos e abióticos, qualidade de bebida, explorando a relação entre genótipo e fenótipo desses indivíduos (JANNINK et al., 2010).

Com o progresso na área da genômica, exemplo disso são as novas tecnologias de sequenciamento de DNA, a caracterização de genomas tornou-se um processo mais eficiente e barato, gerando uma enorme quantidade de dados para a pesquisa. Porém, um avanço similar ainda não ocorreu na caracterização em larga escala de fenótipos, reconhecida como a parte mais laboriosa, cara e tecnicamente desafiadora desse processo, e isso retarda o avanço no entendimento das associações genótipo *versus* fenótipo. (FIORANI; SCHURR, 2013; MONTES et al., 2007).

Fenotipagem, ou a caracterização de recursos genéticos para atributos diversos, representa um componente essencial em programas de melhoramento e muito antes da descoberta de DNA e marcadores moleculares, produtores e fazendeiros já faziam a seleção de genótipos superiores com base somente em fenótipos (DURÃES et al., 2004; LI; ZHANG; HUANG, 2014). Os dados fenotípicos são utilizados na realização de diversos estudos, como: (1) na seleção recorrente assistida por marcadores (NEWELL; JANNINK, 2014); (2) na identificação de eventos positivos em estudos de transgenia (GAUDIN et al., 2013) e dentro deste trabalho, a fenotipagem foi realizada em uma população de *C. canephora* com o objetivo de que os dados gerados, juntamente com a genotipagem desta população, sejam utilizados no estudo de associação genômica ampla (GWAS) e de seleção genômica ampla (SGA) em *C. canephora*.



## 2 MATERIAL E MÉTODOS

### 2.1 Material vegetal

Os genótipos utilizados neste trabalho foram provenientes do campo experimental da Embrapa Cerrados (Planaltina, Distrito Federal, Brasil) (FIGURA 7), estabelecido em 2009/2010, com um total aproximado de 3.500 indivíduos de *C. canephora* Conilon. Essa população de melhoramento foi formada a partir de um pool de sementes coletadas em um campo de polinização aberta com 48 genitores provenientes do banco ativo de germoplasma (BAG) do Instituto de Pesquisa e Assistência Rural (INCAPER, Vitória, Espírito Santo, Brasil).

Em 2012, os 3.500 indivíduos da população de melhoramento foram avaliados visualmente, observando-se características fenotípicas relevantes como o vigor da planta, ramificação secundária, seca de ponteiros, suscetibilidade à ferrugem e precocidade do fruto, e, posteriormente, foram selecionados 1.319 indivíduos, objetos do presente estudo, nomeadas como LxPy (APÊNDICE A).

Figura 7 – Indivíduos de *C. canephora* Conilon estabelecidos em um campo experimental da Embrapa Cerrados, localizado em Planaltina/DF.



Fonte: Do autor (2013).

### 2.2 Fenotipagem da população de *C. canephora*

Os dados fenotípicos a serem utilizados nos experimentos de GWAS e de SGA foram levantados a partir de avaliações realizadas para os 1.319 indivíduos selecionados na

população de melhoramento. Foram obtidos dados de produção, boia, peso de 100 grãos, peneiras e morfologia de frutos, conforme metodologias já estabelecidas. Além disto, análises fisiológicas para avaliação da tolerância à seca, foram também realizadas.

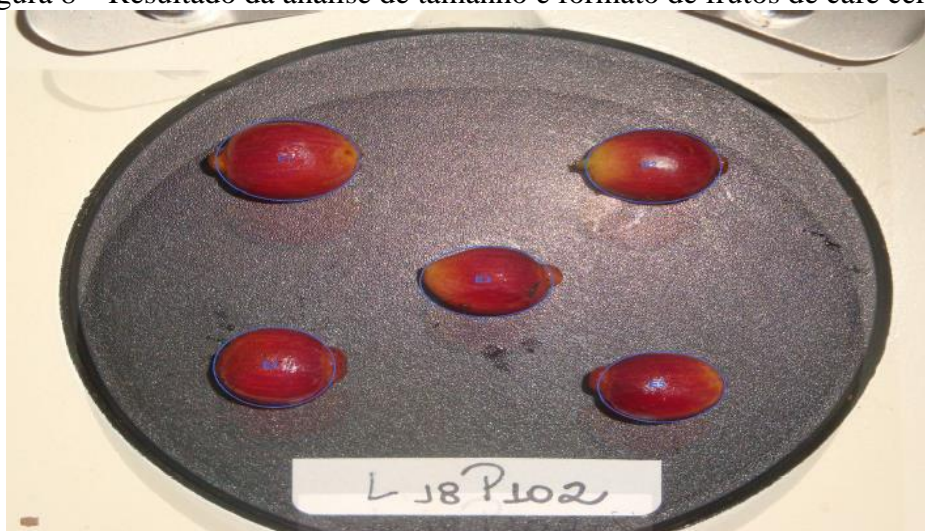
### **2.2.1 Produção**

Os dados de produção anual foram obtidos para cada indivíduo selecionado na população, por um período de 3 anos consecutivos (2011/12; 2012/13 e 2013/14), sendo que cada planta foi colhida individualmente e a produção do café cereja, por planta, medida em litros (L).

### **2.2.2 Morfologia dos frutos**

A partir da colheita, cerca de 50 frutos de café cereja, de cada planta, foram acondicionados em sacos de papel e mantidos à -20°C até a realização das análises morfológicas dos frutos. Os dados de tamanho e formato do grão foram obtidos com auxílio do software DIGI-Pro (Labomed), utilizando-se 5 frutos por amostra, conforme apresentado na Figura 8.

Figura 8 – Resultado da análise de tamanho e formato de frutos de café cereja.



Name	Type	Measurement Values
E1	Ellipse	Area : 2.6233 Centimeters Perimeter : 9.1447 Centimeters Minor Axis : 0.8548 Centimeters Major Axis : 0.9769 Centimeters
E2	Ellipse	Area : 2.2583 Centimeters Perimeter : 8.5064 Centimeters Minor Axis : 0.7758 Centimeters Major Axis : 0.9266 Centimeters
E3	Ellipse	Area : 2.1037 Centimeters Perimeter : 8.2238 Centimeters Minor Axis : 0.7399 Centimeters Major Axis : 0.9051 Centimeters
E4	Ellipse	Area : 2.1996 Centimeters Perimeter : 8.3468 Centimeters Minor Axis : 0.8261 Centimeters Major Axis : 0.8476 Centimeters
E5	Ellipse	Area : 2.0398 Centimeters Perimeter : 8.0714 Centimeters Minor Axis : 0.747 Centimeters Major Axis : 0.8691 Centimeters

Legenda: As medidas (em cm) de área, perímetro, menor eixo e maior eixo para cada planta com 5 repetições (E1 a E5).

Fonte: Do autor (2013).

### 2.2.3 Boia

Após a colheita, uma amostra de 100 frutos de café cereja, de cada planta, foi colocada em um recipiente com 1L de água para obter o valor, em porcentagem, de frutos boia.

### 2.2.4 Peso de 100 grãos

A partir da colheita, as amostras de cada planta foram processadas em despoldador elétrico (Palini & Alves, modelo: PA-DCC/E), seguindo o processo de cereja descascado,

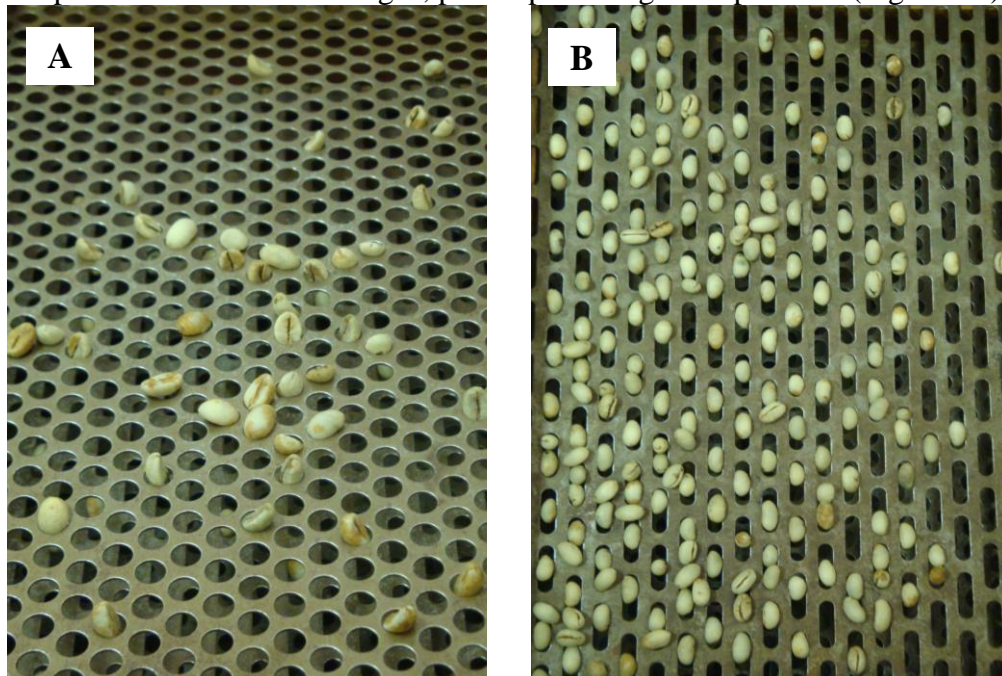
acondicionadas em sacos de filó e colocadas para secar em terreiro, até os grãos atingirem grau de umidade de aproximadamente 12%. Posteriormente, pesou-se uma amostra de 100 grãos secos de cada indivíduo em balança eletrônica modelo AL200C (Marte).

### 2.2.5 Formato e tamanho dos grãos

Amostras de grãos secos (100g) de cada indivíduo foram analisadas em um conjunto de peneiras (Pinhalense) para a classificação quanto ao formato e granulometria.

Conforme a Instrução Normativa do MAPA (BRASIL, 2003), os grãos podem ser enquadrados em duas categorias: (1) Chato – constituída de grãos com superfície dorsal convexa e a ventral plana ou ligeiramente côncava, com a ranhura central no sentido longitudinal e (2) Moca – constituída de grãos com formato ovoide, também com ranhura central no sentido longitudinal (FIGURA 9). Para cada categoria, os grãos são ainda classificados de acordo com a dimensão dos crivos das peneiras que os retêm, conforme Quadro 1.

Figura 9 – Peneira com crivos circulares, para separar os grãos tipo chato (Figura 9A) e peneira com crivos oblongos, para separar os grãos tipo moca (Figura 9B).



Fonte: Do autor (2013).

Quadro 1 – Classificação oficial de café por peneira.

Chato Graúdo	Peneira 17 e maiores
Chato Médio	Peneiras 15 e 16
Chato Miúdo	Peneiras 12, 13 e 14
Moca Graúdo	Peneiras 11 a 13
Moca Médio	Peneira 10
Moca Miúdo	Peneiras 8 e 9

Fonte: Brasil, 2003.

### 2.2.6 Análises fisiológicas

Medidas do potencial hídrico foliar de antemanhã ( $\Psi_{am}$ ) foram realizadas no final do mês de agosto e início de setembro/2013, após um período de estresse hídrico de aproximadamente 70 dias, conforme descrito em Marraccini et al. (2012). Para essas análises fenotípicas, selecionou-se uma amostra de 366 indivíduos, os quais estão entre os 1.319 indivíduos da população de melhoramento selecionados para as demais características.

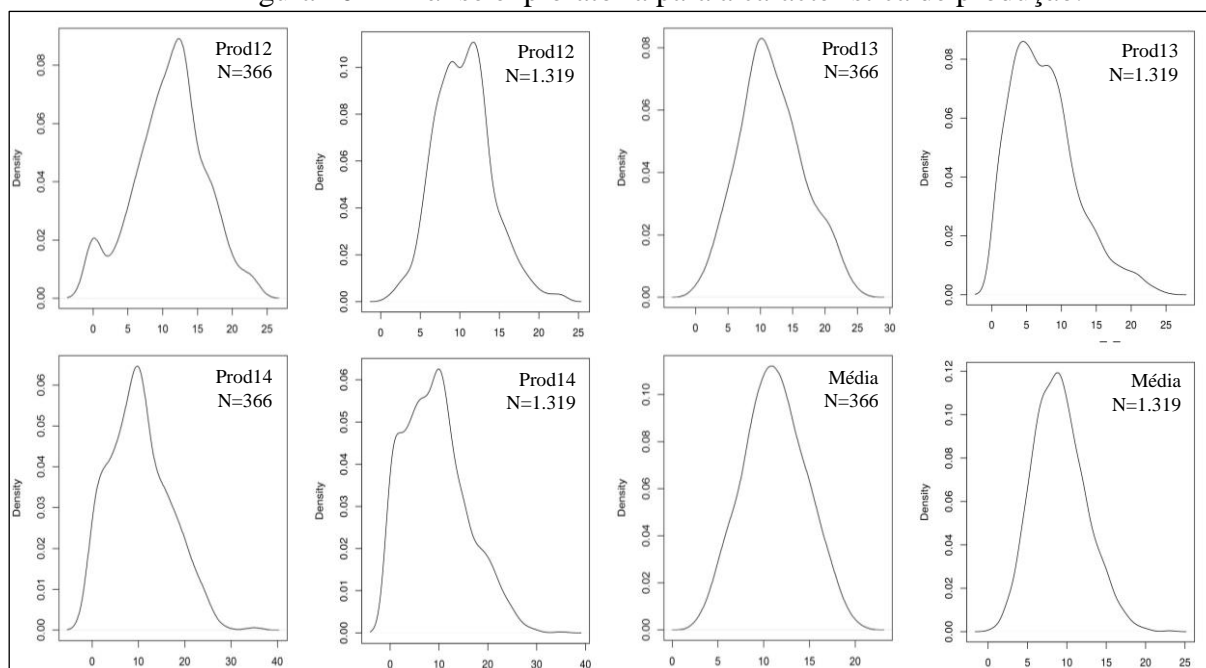
### 3 RESULTADOS E DISCUSSÃO

#### 3.1 Análise exploratória das variáveis fenotípicas

Primeiramente, realizou-se uma análise exploratória para as variáveis fenotípicas avaliadas na população de melhoramento utilizada neste estudo, dessa forma é possível observar o padrão de distribuição das mesmas para os 366 e 1.319 indivíduos selecionados.

Pode-se observar na Figura 10 que para a característica de Produção (Prod12; Prod13; Prod14) os dados referentes aos 1.319 indivíduos apresentaram uma distribuição bimodal. Entretanto, ao obter a média de produção para os três anos avaliados (Figura 10 – Média; N=1.319), os dados parecem se ajustar para uma distribuição normal. Para os 366 indivíduos, a análise exploratória apresentou um resultado mais próximo à distribuição normal, tanto para os anos separados de produção (Prod12, Prod13 e Prod14; N=366) como para a média dos três anos avaliados.

Figura 10 – Análise exploratória para a característica de produção.



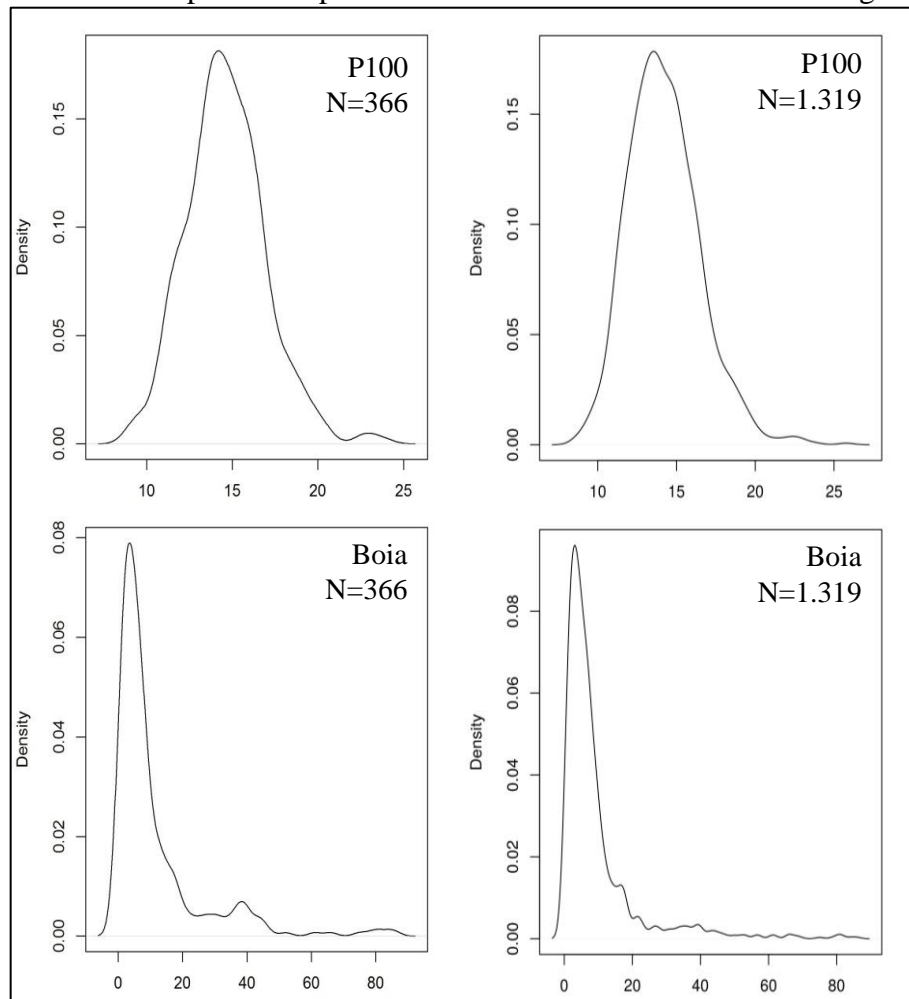
Legenda: Os dados de produção foram obtidos para os 366 (N=366) e 1.319 (N=1.319) indivíduos nos anos de 2011/2012 (Prod12), 2012/2013 (Prod13), 2013/2014 (Prod14) e média da produtividade para os 3 anos avaliados.

Fonte: Do autor (2018).

Para as características de Boia e peso de 100 grãos (P100), é possível observar na Figura 11 que os resultados foram semelhantes em sua distribuição independente do número de indivíduos. Os dados obtidos para a característica de Boia apresentou assimetria, contudo para P100 nota-se uma distribuição normal, tanto para os 366 como para os 1.319 indivíduos.



Figura 11 – Análise exploratória para as características Boia e Peso de 100 grãos (P100).

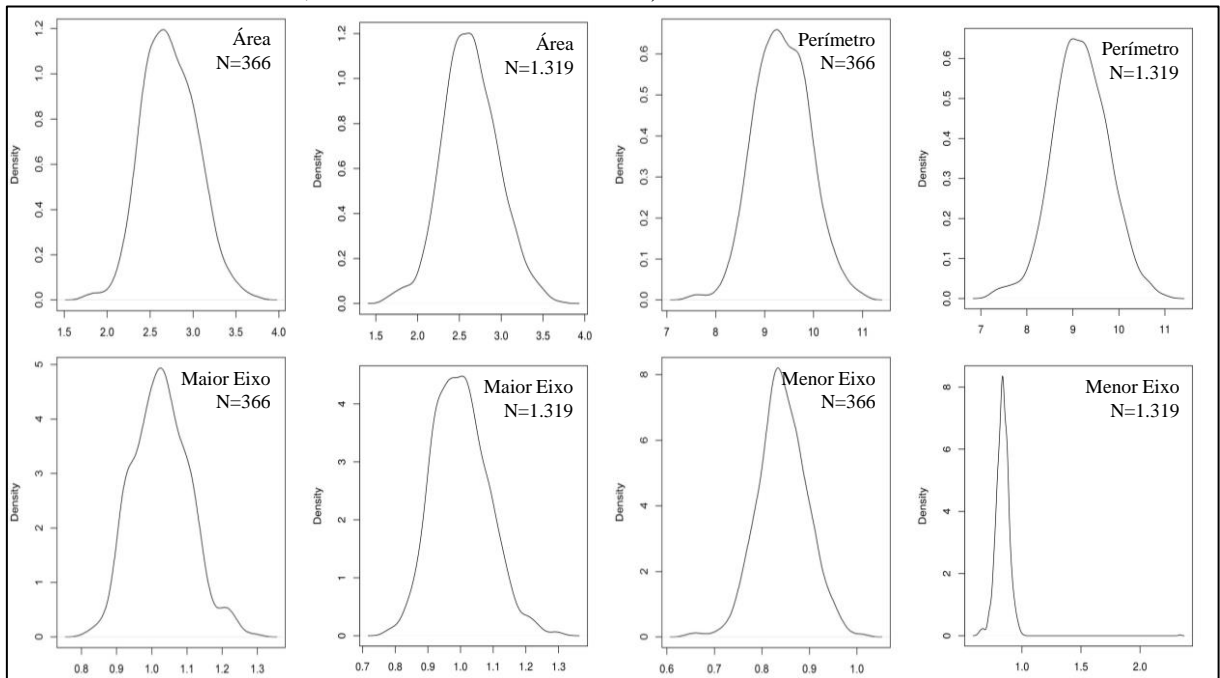


Legenda: Os dados de Boia e peso de 100 grãos foram obtidos para os 366 (N=366) e 1.319 (N=1.319) indivíduos selecionados.

Fonte: Do autor (2018).

Para a maioria das características avaliadas referentes à morfologia dos frutos, como área, perímetro, eixo maior e eixo menor, observa-se (FIGURA 12) uma tendência à normalidade dos dados, exceto a característica de eixo menor avaliada para os 1.319 indivíduos que apresentou uma assimetria para os dados. Já para os dados obtidos a partir da análise de tamanho e formato dos grãos (FIGURA 13), a maioria demonstrou uma distribuição assimétrica, com exceção para Moca Graúdo, tanto para os 366 como para os 1.319 indivíduos e Chato Médio (N=366), que apresentaram distribuição normal. Outro dado interessante pode ser observado para os grãos classificados como médio, tanto chato como moca, que apresentaram uma distribuição bimodal para os 1.319 indivíduos.

Figura 12 – Análise exploratória para as características de Morfologia dos frutos (Área, Perímetro, Eixo Maior e Eixo Menor).

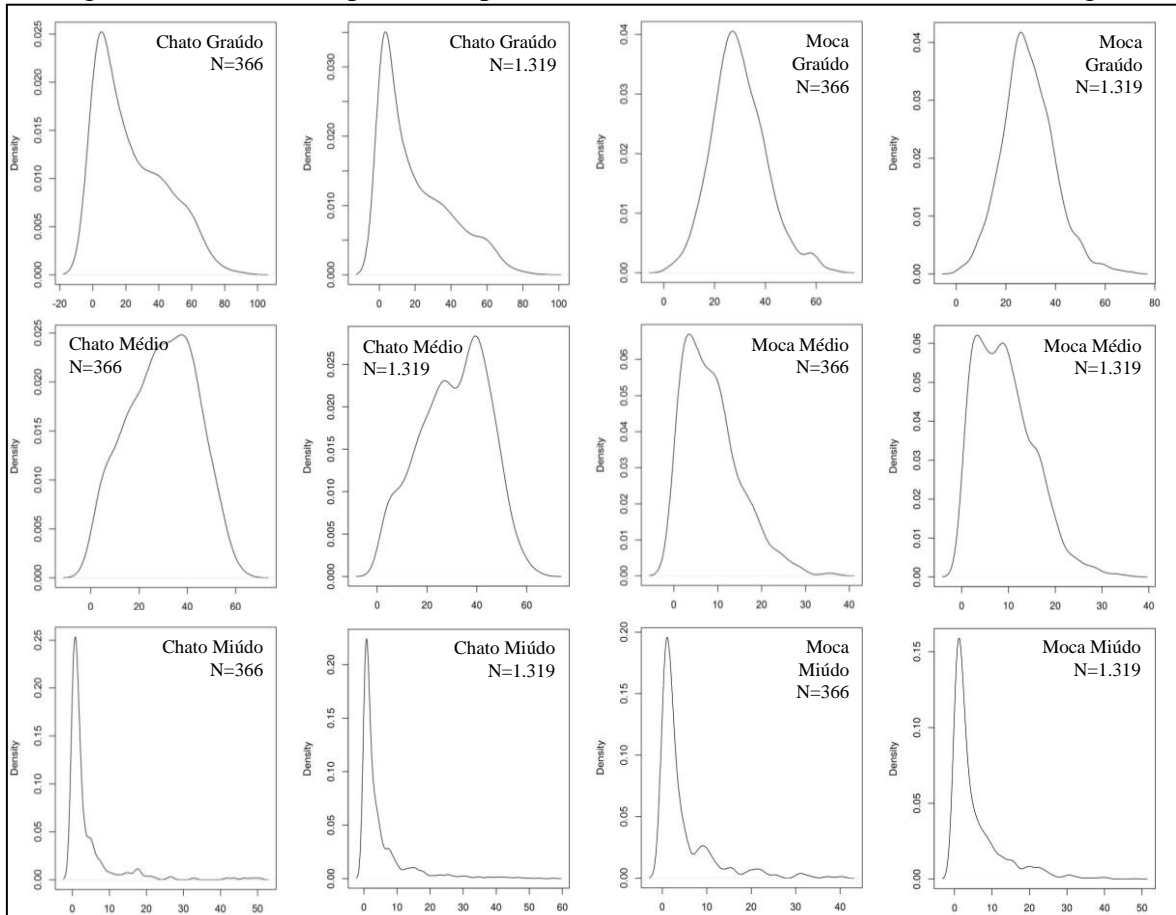


Legenda: Para a característica de Morfologia, medidas de área, perímetro, eixos (maior e menor) foram obtidas a partir de amostra de frutos cereja para os 366 (N=366) e 1.319 (N=1.319) indivíduos selecionados.

Fonte: Do autor (2018).



Figura 13 – Análise exploratória para a característica de tamanho e formato dos grãos.

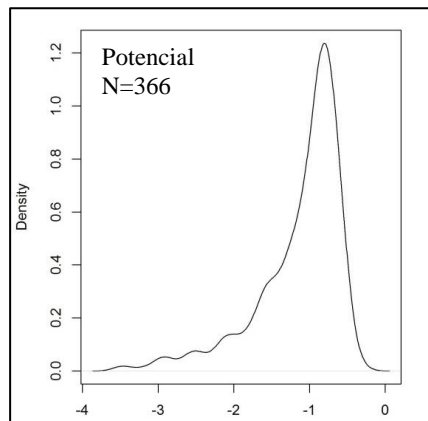


Legenda: Análise exploratória dos resultados obtidos para tamanho e formato de grãos, por meio da técnica de peneiras, para os 366 (N=366) e 1.319 (N=1.319) indivíduos selecionados.

Fonte: Do autor (2018).

Na Figura 14 verifica-se a distribuição dos resultados obtidos para a característica de potencial hídrico foliar de antemanhã, fenótipo medido somente para os 366 indivíduos da população da Embrapa Cerrados. Para essa característica nota-se uma distribuição assimétrica, em que a maioria dos indivíduos apresentou potencial em torno de -1,0 MPa.

Figura 14 – Análise exploratória para a característica de potencial hídrico foliar de antemanhã.



Fonte: Do autor (2018).

### 3.2 Morfologia dos frutos

Conforme apresentado na Tabela 3, na avaliação para a morfologia de frutos, os valores para área variaram entre 1,62 a 3,72 cm, com média de 2,63 cm. Os eixos maior e menor apresentaram médias de 1,0 e 0,83 cm respectivamente. Para perímetro, a média foi de 9,17 cm, variando de 0,62 a 2,34 cm.

Tabela 2 – Valores máximo (Máx), mínimo (Mín) e média, com o desvio padrão, para o fenótipo de Morfologia dos frutos.

	Área	Perímetro	Eixo Menor	Eixo Maior
<b>Máx</b>	3,72	11,03	1,01	2,34
<b>Mín</b>	1,62	7,22	0,62	0,77
<b>Média</b>	2,63 ( $\pm 0,26$ )	9,17 ( $\pm 0,48$ )	0,83 ( $\pm 0,04$ )	1,00 ( $\pm 0,07$ )

Fonte: Do autor (2018).

Brum (2007) em trabalho com a variedade clonal EMCAPA 8131 na Bacia do Itaperimim a 146 m de altitude, encontrou valores médios de 0,91 cm para a medida de diâmetro menor, correspondente à característica eixo menor do presente trabalho, e de 1,15 cm para à medida que corresponde ao eixo maior.

### 3.3 Peso de 100 grãos e boia

Na análise de percentagem de frutos chochos (boia), a média encontrada foi de 9,27%. Este valor é muito próximo do relatado por Silva et al. (2012), em seu trabalho com *C. arabica*, indicando baixa ocorrência de “lojas vazias”.

A variação encontrada para peso de 100 grãos foi de 8,27 a 25,77 g, com média em torno de 14,2 g.

### 3.4 Formato e tamanho do grão

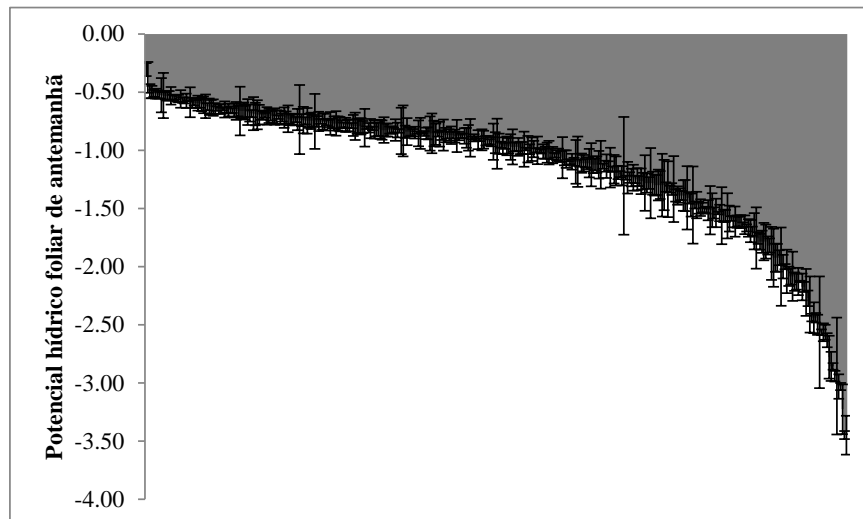
Por meio das análises de formato e tamanho dos grãos, utilizando as peneiras, verificou-se que mais aproximadamente 50% dos indivíduos selecionados na população de melhoramento apresentaram peneira maior que 13. Este resultado é inferior aos 65% relatado por Brum (2007) e aos 94,5 e 96,13% identificados por Oliveira et al. (2009) aos 600 m e 630

m de altitude, respectivamente, porém indica uma boa diversidade da população, para essa característica.

### 3.5 Tolerância à seca

Pode-se observar na Figura 15 que o potencial hídrico foliar de antemanhã ( $\Psi_{am}$ ) apresentou uma grande variação entre os 366 indivíduos avaliados, com o indivíduo mais tolerante ao déficit hídrico apresentando potencial de -0,31MPa e mais da metade dos indivíduos com  $\Psi_{am}$  de até -1,0 MPa, mesmo após um período prolongado de seca, cerca de 70 dias.

Figura 15 – Potencial hídrico foliar de antemanhã para as 366 plantas de *C. canephora* selecionadas.



Fonte: Do autor (2015).

Na Figura 16 nota-se o contraste fenotípico entre os indivíduos da população de melhoramento após o período de déficit hídrico. A planta tolerante ( $L_5P_{47}$ ) apresentou um potencial de -0,46Mpa (Figura 16A) enquanto que a planta  $L_{12}P_{100}$ , sensível ao déficit hídrico, apresentou -2,47MPa (Figura 16B). Além das diferenças entre os valores para o  $\Psi_{am}$ , nota-se visivelmente na planta sensível, um número de folhas reduzido, indicando maior sensibilidade ao estresse hídrico.

Figura 16 – Fenótipos contrastantes pertencentes à população de melhoramento após um período de déficit hídrico (Embrapa Cerrados).



Fonte: Do autor (2013).

Alguns trabalhos que avaliaram o  $\Psi_{am}$  entre diferentes clones de *C. canephora* Conilon, tolerantes (14, 73 e 120) e sensível (22) ao déficit hídrico, identificaram também uma variabilidade entre os diferentes genótipos. Esses estudos, realizados em casa de vegetação, analisaram o número de dias em que cada clone apresentou  $\Psi_{am}$  de -3,0 MPa. O clone 22 chegou no potencial estabelecido com seis dias, seguido pelos clones 14 e 120, ambos com 12 dias e aos 15 dias, o clone 73 (MARRACINNI et al., 2012; VIEIRA et al., 2013).

Assim como observado por Marraccini et al. (2012) e Vieira et al. (2013), as plantas utilizadas nesse presente estudo também apresentaram uma diversidade fenotípica em relação a medida de potencial hídrico, com valores entre -0,31 a -3,45 MPa, sendo que aproximadamente 60% das plantas apresentaram potencial com valor até -1,0 MPa, indicando uma tolerância aos 70 dias de déficit hídrico, por que passaram.

### 3.6 Produção

Para os 366 indivíduos que possuem os dados de  $\Psi_{am}$ , a variação fenotípica encontrada para a produção pode ser observada na Tabela 4. Como se pode observar, todas as características apresentaram uma ampla variação fenotípica, ou seja, um bom indicativo da existência de variabilidade genética e possibilidade de ganho com a seleção.

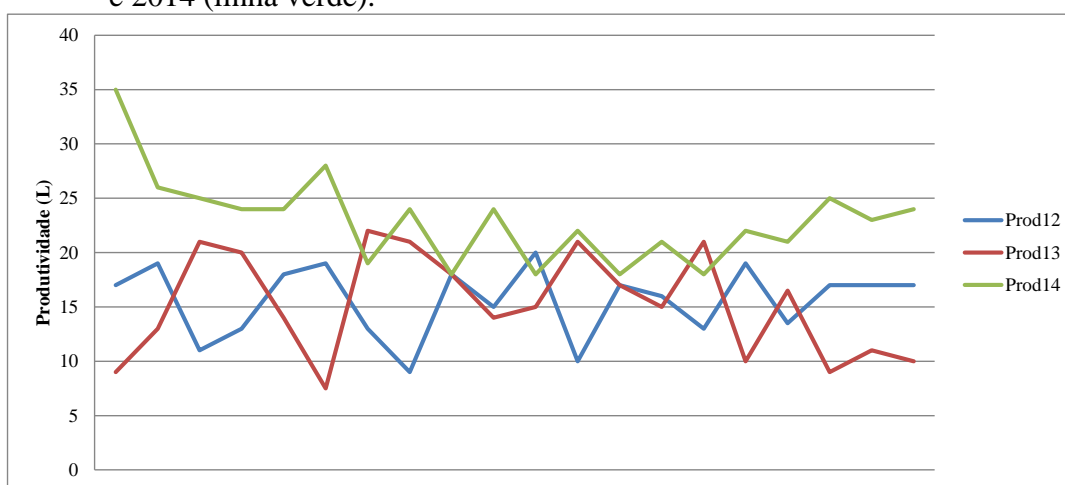
Tabela 3 – Valor máximo, mínimo e média para a produção avaliada nos três anos e para 366 indivíduos.

	<b>Prod12</b>	<b>Prod13</b>	<b>Prod14</b>
<b>Máximo</b>	23,00	25,00	35,00
<b>Mínimo</b>	0,00	0,50	0,00
<b>Média</b>	10,98	11,74	10,30

Fonte: Do autor (2018).

A média alcançada para os 3 anos avaliados foi de 11L, os valores máximos foram de 23, 25 e 35 litros, para os anos de 2012, 2013 e 2014, respectivamente. Classificando esses indivíduos pela média da produção e selecionando-se os 20 mais produtivos, visualiza-se na Figura 17 o padrão da produção para cada ano avaliado. Nota-se na Figura 17 que as plantas apresentam um ano de alta carga e outro de baixa carga, conhecido como bienalidade do cafeeiro.

Figura 17 – Produção (medida em litros de cereja por planta), para os 20 indivíduos que apresentaram maior média, nos anos de 2012 (linha azul), 2013 (linha vermelha) e 2014 (linha verde).

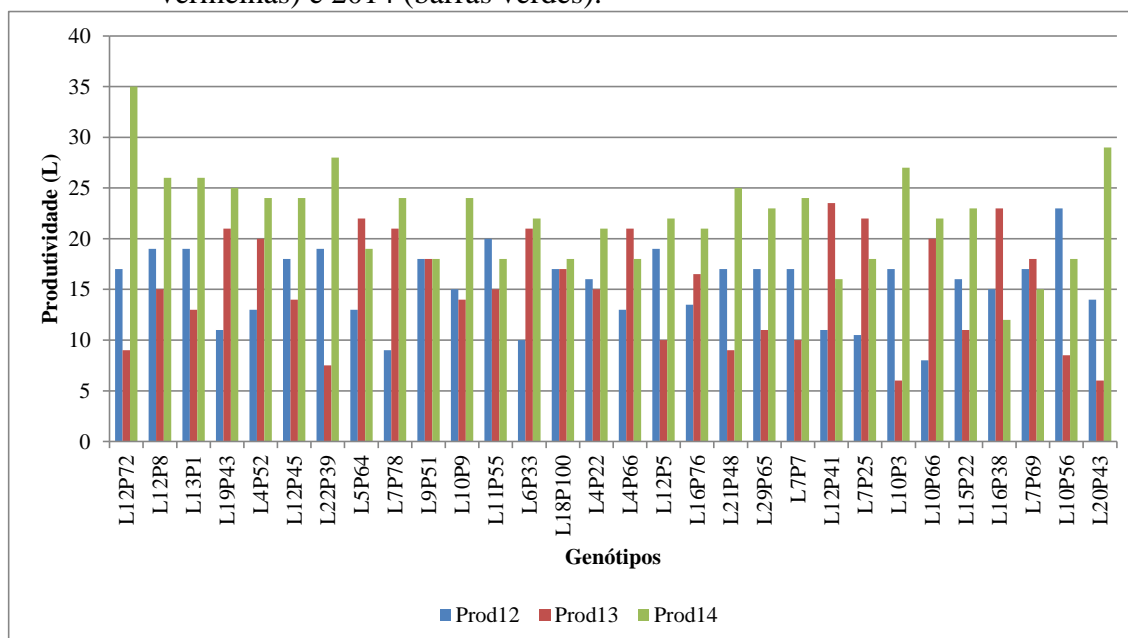


Fonte: Do autor (2018).

Já para os 1.319 indivíduos, a produção média foi de pouco mais de 9L e o ano de 2012 representou o de maior produção, em torno de 10,5 L/planta. A Figura 18 mostra, para as 30 plantas com maior média de produção, a variação já esperada em cafeeiro, com um ano de alta e o outro de baixa carga. Este efeito da bienalidade fica claro ao se observar os dados de produção das plantas L<sub>12</sub>P<sub>72</sub>, L<sub>12</sub>P<sub>8</sub> e L<sub>12</sub>P<sub>1</sub>, entre outras (FIGURA 18). Entretanto, pode-se também observar nos dados apresentados na Figura 19, que existem plantas, que sofrem menos o efeito da bienalidade, como por exemplo, a L<sub>9</sub>P<sub>51</sub> e L<sub>18</sub>P<sub>100</sub>, que produziram aproximadamente a mesma quantidade, acima de 15L, para os diferentes anos avaliados. Mais

ainda, existem também plantas que aumentaram a produção, ano após ano, tais como as plantas L<sub>19</sub>P<sub>43</sub> e L<sub>4</sub>P<sub>52</sub>, por exemplo (FIGURA 18).

Figura 18 – Produção (medida em litros de cereja por planta), para os 30 indivíduos que apresentaram maior média, nos anos de 2012 (barras azuis), 2013 (barras vermelhas) e 2014 (barras verdes).



Fonte: Do autor (2018).

Os dados de produção apresentaram uma tendência de serem bimodais, como apresentado na Figura 10. Esses resultados estão de acordo com o esperado para a espécie estudada visto que a bienalidade é própria da natureza fisiológica do cafeeiro, que necessita vegetar em um ano para produzir bem no ano seguinte. Em anos de alta carga, a planta direciona a produção de fotossintetizados para a formação de novas gemas e crescimento dos frutos; diferentemente dos anos de baixa carga, em que os fotossintetizados são direcionados para a formação de gemas vegetativas que gerarão novos ramos (RENA; MAESTRI, 1985).

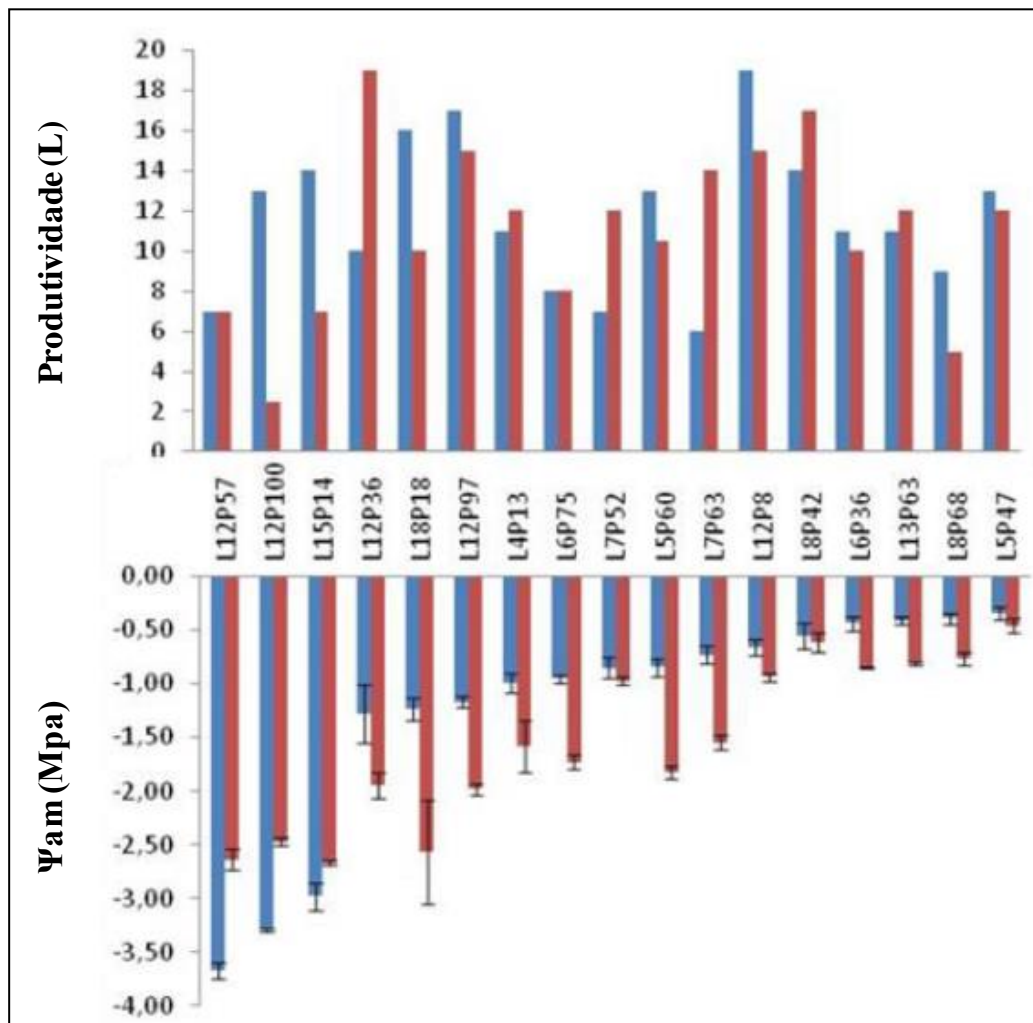
Outro dado observado para o fenótipo de produção é a presença, no mesmo campo experimental, de plantas altamente produtivas ao lado de plantas com uma produção muito baixa ou, até mesmo, que não produziram. Essa variabilidade espacial e temporal da produtividade já havia sido relatada por Carvalho et al. (2004) em regiões produtoras de café no sul de Minas Gerais.

Um fato importante de ser mencionado é que o *C. canephora* é cultivado em regiões com menor altitude, normalmente abaixo de 500 m, e temperatura mais elevada, com média anual entre 22°C e 26°C, conforme descrito por Matiello (1991). Em um trabalho realizado por Oliveira et al. (2009), os autores analisaram a produção de café Conilon cultivado em altitude acima do recomendado e obtiveram uma produção por planta de 6,75L e 5,25L em

ambientes com altitudes de 600 m e 630 m, respectivamente. Resultado semelhante foi descrito por Brum (2007), porém a aproximadamente 150 m de altitude, em que a produção média por planta foi de 6,0 L. O presente trabalho, realizado no campo experimental da Embrapa Cerrados e que possui altitude acima de 1.000 m, apresentou média de produção por planta de 9L para os 1.319 indivíduos avaliados e de 11,2L considerando somente os 366 indivíduos que possuem as medidas de potencial hídrico foliar de antemanhã.

Na relação tolerância *versus* produtividade, pode-se identificar na Figura 19 que algumas plantas são produtivas e tolerantes à seca (por exemplo, L<sub>13</sub>P<sub>63</sub>, L<sub>8</sub>P<sub>68</sub> e L<sub>5</sub>P<sub>47</sub>) e outras apresentam o fenótipo contrastante, sendo sensíveis ao estresse hídrico e pouco produtivas (L<sub>12</sub>P<sub>57</sub>, L<sub>12</sub>P<sub>100</sub> e L<sub>15</sub>P<sub>14</sub>).

Figura 19 – Produção, em litros (L), e potencial hídrico foliar de antemanhã ( $\Psi_{am}$ ) de plantas de *C. canephora* LxPy selecionadas. As barras azuis correspondem às medidas realizadas em 2012 e as barras vermelhas às medidas realizadas em 2013.



Fonte: Do autor (2014).

Trabalhos como o de Ferrão et al. (2000) e Fonseca et al. (2004) caracterizaram plantas de *C. canephora* Conilon como tolerantes ao estresse hídrico e produtivas. Essa característica de tolerância poderia ser justificada pela maior proximidade genética dos materiais brasileiros com o grupo SG1 de *C. canephora* (MONTAGNON; CUBRY; LEROY, 2012). O grupo SG1 apresenta-se mais tolerante ao déficit hídrico comparado ao grupo SG2 (Robusta) (MONTAGNON; LEROY, 1993).



#### 4 CONCLUSÃO

Os resultados obtidos a partir das avaliações fenotípicas realizadas neste trabalho nos permitem concluir que existe variabilidade fenotípica razoável para as diversas características avaliadas na população de melhoramento utilizada, e que existe potencial para a realização dos estudos de GWAS e de SGA, utilizando-se esses dados. Entretanto, é importante salientar que os modelos disponíveis para a realização desses estudos de GWAS e SGA, estão estabelecidos para dados que apresentam uma distribuição normal e, portanto, para algumas características ajustes podem ser necessários. Além disto, nota-se também, que existe potencial de cultivo, sob condições irrigadas, de *C. canephora* em altitudes elevadas (acima de 1000 m), como é o caso do campo experimental da Embrapa Cerrados em que se localiza a população de melhoramento, utilizada neste estudo.

## REFERÊNCIAS

- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Instrução Normativa** n. 8. de 11 de junho de 2003.
- BRUM, V. J. **Café conilon em sombreamento com pupunheira**. 2007. 149 p. Dissertação (Mestrado em Produção Vegetal) – Universidade Federal do Espírito Santo, Vitória, 2007.
- BUFFO, R. A.; CARDELLI-FREIRE, C. Coffee flavour: An overview. **Flavour and Fragrance Journal**, v. 19, n. 2, p. 99-104, 2004.
- CARVALHO, L. G. et al. A regression model to predict coffee productivity in Southern Minas Gerais, Brazil. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 8, n. 2/3, p. 204-211, 2004.
- COBB, J. N. et al. Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. **Theoretical and Applied Genetics**, v. 126, n. 4, p. 867-887, 2013.
- DURÃES, F.O.M. et al. **Fenotipagem associada à tolerância a seca em milho para uso em melhoramento, estudos genômicos e seleção assistida por marcadores**. Circular Técnica n. 39, Embrapa Milho e Sorgo, 2004.
- FARAH, A. Coffee as a speciality and functional beverage. In PAQUIN, P. **Functional and Speciality Beverage Technology**. Oxford, Woodhead Publishing, 2009, p. 370–395.
- FERRÃO, R. G. et al. EMCAPA 8141 - Robustão Capixaba, variedade clonal de café Conilon tolerante à seca, desenvolvida para o estado do Espírito Santo. **Revista Ceres**, v. 47, n. 273, p. 555-560, 2000.
- FIORANI, F.; SCHURR, U. Future scenarios for plant phenotyping. **Annual Review of Plant Biology**, v. 64, p. 267-291, 2013.
- FONSECA, A. F. A., et al. Conilon Vitória—Incapar 8142: improved *Coffea canephora* var. Kouillou clone cultivar for the State Espírito Santo. **Crop Breeding and Applied Biotechnology**, v. 4, p. 503-505, 2004.
- GAUDIN, A. C. M. et al. Taking transgenic rice drought screening to the field. **Journal of Experimental Botany**, v. 64, n. 1, p. 109-117, 2013.
- JANNINK, J. L.; LORENZ, A. J.; IWATA, H. Genomic selection in plant breeding: from theory to practice. **Briefings in Functional Genomics**, v. 9, n. 2, p. 166-177, 2010.
- JOËT, T. et al. Metabolic pathways in tropical dicotyledonous albuminous seeds: *Coffea arabica* as a case study. **New Phytologist**, v. 182, n. 1, p. 146–162, 2009.
- LI, L.; ZHANG, Q.; HUANG, D. A. Review of Imaging Techniques for Plant Phenotyping. **Sensors**, v. 14, n. 11, p. 20078-20111, 2014.

- MARRACCINI, P. et al. Differentially expressed genes and proteins upon drought acclimation in tolerant and sensitive genotypes of *Coffea canephora*. **Journal of Experimental Botany**, v. 63, n. 2, p. 695–709, 2012.
- MATIELLO, J. B. Clima e solos para o cafeeiro. In: **O café do cultivo ao consumo**. São Paulo: Ed. Globo, 1991, cap. 2, p. 26-39.
- MONTAGNON, C.; CUBRY, P.; LEROY, T. Amélioration génétique du caféier *Coffea canephora* Pierre: connaissances acquises, stratégies et perspectives. **Cahiers Agricultures**, v. 21, n. 2-3, p. 143-153, 2012.
- MONTAGNON, C.; LEROY, T. Réaction à la sécheresse de jeunes caféiers *Coffea canephora* de Côte-d'Ivoire appartenant à différents groupes génétiques. **Café, Cacao, Thé**, v. 37, n. 3, p. 179-190, 1993.
- MONTES, J. M.; MELCHINGER, A. E.; REIF, J. C. Novel throughput phenotyping platforms in plant genetic studies. **Trends in Plant Science**, v. 12, n. 10, p. 433-436, 2007.
- NEWELL, M. A.; JANNINK, J. L. Genomic selection in plant breeding. In: FLEURY, D.; WHITFORD, R. **Methods in Molecular Biology (Methods and Protocols)**, v. 1145, New York, Humana Press, 2014, 117-130.
- OESTREICH-JANZEN, S. Chemistry of coffee. **Comprehensive Natural Products II**, v. 3, p. 1085-1117, 2010.
- OLIVEIRA, C. M. et al. Maturação e produção do café conilon cultivado em altitude acima do recomendado. In: XIII ENCONTRO LATINO AMERICANO DE INICIAÇÃO CIENTÍFICA E IX ENCONTRO LATINO AMERICANO DE PÓS-GRADUAÇÃO, 2009, Vale do Paraíba. **Anais eletrônicos...** Vale do Paraíba. Disponível em: [http://www.inicepg.univap.br/cd/INIC\\_2009/anais/arquivos/0720\\_0713\\_01.pdf](http://www.inicepg.univap.br/cd/INIC_2009/anais/arquivos/0720_0713_01.pdf)
- RENA, A. B.; MAESTRI, M. Fisiologia do cafeeiro. **Informe Agropecuário**, v. 11, n. 126, p. 26-40, 1985.
- SILVA, V. A. et al. Sistemas intercalares com abacaxizeiro como alternativa de renda durante a formação de cafezais irrigados. **Pesquisa Agropecuária Brasileira**, v. 47, n. 10, p. 1471-1479, 2012.
- SUNARHARUM, W. B.; WILLIAMS, D. J.; SMYTH, H. E. Complexity of coffee flavor: A compositional and sensory perspective. **Food Research International**, v. 62, p. 315-325, 2014.
- TRAN, H. T. M. et al. Advances in genomics for the improvement of quality in coffee. **Journal of the Science of Food and Agriculture**, v. 96, n. 10, p. 3300-3312, 2016.
- VIEIRA, N. G. et al. Different molecular mechanisms account for drought tolerance in *Coffea canephora* var. Conilon. **Tropical Plant Biology**, v. 6, n. 4, p. 181-190, 2013.

VINECKY, F. et al. Controlled irrigation and nitrogen, phosphorous and potassium fertilization affect the biochemical composition and quality of Arabica coffee beans. **The Journal of Agricultural Science**, v. 155, n. 6, p. 902-918, 2016.

**APÊNDICE A** – Lista com os 1.319 indivíduos de *C. canephora* selecionados para o estudo.

L01P1	L02P51	L04P13	L05P39	L06P6	L07P63
L01P10	L02P52	L04P14	L05P40	L06P51	L07P66
L01P11	L02P53	L04P16	L05P42	L06P52	L07P67
L01P13	L02P55	L04P19	L05P45	L06P55	L07P68
L01P15	L02P57	L04P22	L05P46	L06P59	L07P69
L01P16	L02P59	L04P24	L05P47	L06P62	L07P7
L01P19	L02P62	L04P26	L05P48	L06P64	L07P71
L01P2	L02P65	L04P27	L05P50	L06P66	L07P72
L01P20	L02P66	L04P28	L05P51	L06P68	L07P80
L01P24	L02P7	L04P30	L05P57	L06P69	L07P84
L01P26	L02P70	L04P37	L05P58	L06P70	L07P85
L01P27	L02P71	L04P38	L05P59	L06P71	L07P90
L01P3	L02P8	L04P39	L05P6	L06P73	L07P92
L01P32	L02P9	L04P4	L05P60	L06P74	L07P96
L01P34	L03P1	L04P42	L05P64	L06P75	L08P102
L01P35	L03P10	L04P46	L05P66	L06P79	L08P11
L01P37	L03P15	L04P49	L05P67	L06P9	L08P13
L01P47	L03P16	L04P50	L05P77	L06P92	L08P15
L01P49	L03P17	L04P51	L05P81	L07P1	L08P17
L01P50	L03P18	L04P52	L05P82	L07P103	L08P19
L01P60	L03P19	L04P53	L06P10	L07P105	L08P2
L01P62	L03P21	L04P55	L06P11	L07P13	L08P21
L01P65	L03P22	L04P6	L06P12	L07P15	L08P22
L02P10	L03P24	L04P62	L06P13	L07P16	L08P24
L02P13	L03P25	L04P66	L06P2	L07P17	L08P25
L02P14	L03P27	L04P67	L06P22	L07P19	L08P26
L02P15	L03P32	L04P7	L06P24	L07P22	L08P28
L02P17	L03P33	L04P70	L06P26	L07P24	L08P29
L02P19	L03P35	L04P72	L06P3	L07P25	L08P3
L02P20	L03P37	L04P8	L06P30	L07P26	L08P30
L02P21	L03P38	L04P9	L06P31	L07P28	L08P32
L02P24	L03P4	L05P1	L06P33	L07P29	L08P33
L02P27	L03P40	L05P11	L06P34	L07P35	L08P34
L02P28	L03P41	L05P16	L06P35	L07P36	L08P35
L02P30	L03P42	L05P17	L06P36	L07P45	L08P36
L02P31	L03P43	L05P18	L06P37	L07P46	L08P37
L02P34	L03P49	L05P2	L06P38	L07P47	L08P38
L02P4	L03P55	L05P20	L06P40	L07P48	L08P39
L02P40	L03P56	L05P21	L06P41	L07P50	L08P4
L02P42	L03P57	L05P25	L06P42	L07P51	L08P40
L02P44	L03P59	L05P28	L06P43	L07P52	L08P41
L02P46	L03P6	L05P32	L06P47	L07P55	L08P42
L02P49	L03P7	L05P33	L06P48	L07P57	L08P44
L02P5	L04P1	L05P37	L06P5	L07P60	L08P45

L08P46	L09P113	L10P108	L10P84	L11P64	L12P38
L08P47	L09P17	L10P11	L10P86	L11P66	L12P4
L08P48	L09P19	L10P116	L10P9	L11P68	L12P40
L08P49	L09P2	L10P12	L10P95	L11P69	L12P41
L08P5	L09P20	L10P13	L11P10	L11P74	L12P42
L08P50	L09P23	L10P14	L11P101	L11P76	L12P44
L08P51	L09P24	L10P16	L11P102	L11P8	L12P45
L08P53	L09P28	L10P18	L11P107	L11P83	L12P46
L08P54	L09P29	L10P19	L11P109	L11P84	L12P48
L08P56	L09P3	L10P20	L11P11	L11P86	L12P49
L08P57	L09P30	L10P22	L11P110	L11P9	L12P5
L08P58	L09P31	L10P23	L11P112	L11P91	L12P50
L08P59	L09P33	L10P24	L11P14	L11P92	L12P53
L08P60	L09P34	L10P25	L11P15	L11P93	L12P55
L08P62	L09P36	L10P26	L11P2	L11P97	L12P56
L08P64	L09P37	L10P28	L11P20	L11P99	L12P57
L08P65	L09P39	L10P3	L11P23	L12P10	L12P59
L08P66	L09P40	L10P30	L11P24	L12P100	L12P6
L08P67	L09P43	L10P33	L11P3	L12P101	L12P60
L08P68	L09P44	L10P35	L11P31	L12P102	L12P61
L08P69	L09P45	L10P36	L11P37	L12P103	L12P63
L08P7	L09P48	L10P39	L11P39	L12P106	L12P65
L08P70	L09P49	L10P40	L11P4	L12P107	L12P67
L08P71	L09P5	L10P43	L11P40	L12P108	L12P68
L08P73	L09P50	L10P44	L11P43	L12P11	L12P70
L08P76	L09P51	L10P46	L11P44	L12P113	L12P71
L08P78	L09P53	L10P47	L11P45	L12P12	L12P72
L08P79	L09P55	L10P49	L11P46	L12P13	L12P73
L08P80	L09P56	L10P5	L11P47	L12P14	L12P74
L08P81	L09P63	L10P51	L11P48	L12P17	L12P77
L08P82	L09P66	L10P56	L11P49	L12P18	L12P78
L08P88	L09P7	L10P57	L11P5	L12P2	L12P8
L08P89	L09P70	L10P58	L11P50	L12P21	L12P80
L08P9	L09P73	L10P6	L11P51	L12P23	L12P86
L08P91	L09P8	L10P63	L11P52	L12P24	L12P87
L08P92	L09P83	L10P64	L11P53	L12P25	L12P9
L08P96	L09P91	L10P65	L11P55	L12P26	L12P90
L08P98	L09P92	L10P66	L11P56	L12P27	L12P91
L09P1	L09P95	L10P67	L11P57	L12P28	L12P92
L09P10	L10P1	L10P68	L11P58	L12P3	L12P94
L09P106	L10P100	L10P7	L11P59	L12P30	L12P95
L09P11	L10P102	L10P78	L11P60	L12P33	L12P96
L09P111	L10P103	L10P8	L11P61	L12P34	L12P97
L09P112	L10P105	L10P83	L11P62	L12P36	L12P99

L13P03	L13P66	L14P5	L15P19	L15P97	L16P76
L13P1	L13P67	L14P51	L15P2	L15P98	L16P78
L13P100	L13P68	L14P52	L15P21	L15P99	L16P79
L13P104	L13P7	L14P53	L15P22	L16P10	L16P8
L13P105	L13P76	L14P54	L15P23	L16P100	L16P80
L13P110	L13P77	L14P6	L15P24	L16P101	L16P81
L13P13	L13P79	L14P60	L15P26	L16P102	L16P82
L13P15	L13P81	L14P61	L15P27	L16P105	L16P84
L13P17	L13P82	L14P62	L15P28	L16P111	L16P88
L13P2	L13P83	L14P63	L15P3	L16P112	L16P9
L13P20	L13P9	L14P64	L15P30	L16P114	L16P99
L13P21	L13P90	L14P65	L15P32	L16P117	L17P10
L13P23	L13P92	L14P68	L15P33	L16P121	L17P100
L13P24	L14P1	L14P69	L15P35	L16P122	L17P105
L13P26	L14P10	L14P7	L15P37	L16P13	L17P106
L13P27	L14P101	L14P70	L15P39	L16P19	L17P107
L13P28	L14P12	L14P73	L15P42	L16P2	L17P108
L13P29	L14P13	L14P74	L15P43	L16P20	L17P109
L13P34	L14P14	L14P75	L15P45	L16P22	L17P11
L13P35	L14P15	L14P77	L15P47	L16P25	L17P110
L13P36	L14P16	L14P8	L15P48	L16P26	L17P112
L13P39	L14P18	L14P80	L15P49	L16P29	L17P113
L13P4	L14P19	L14P81	L15P50	L16P3	L17P122
L13P40	L14P2	L14P82	L15P53	L16P30	L17P133
L13P41	L14P20	L14P84	L15P55	L16P32	L17P14
L13P42	L14P21	L14P86	L15P56	L16P35	L17P16
L13P44	L14P23	L14P9	L15P57	L16P38	L17P19
L13P45	L14P24	L14P92	L15P6	L16P4	L17P20
L13P46	L14P26	L14P94	L15P60	L16P43	L17P21
L13P48	L14P27	L14P96	L15P61A	L16P44	L17P22
L13P49	L14P3	L14P99	L15P63	L16P47	L17P27
L13P50	L14P30	L15P1	L15P64	L16P49	L17P29
L13P53	L14P31	L15P101	L15P66	L16P51	L17P3
L13P54	L14P32	L15P105	L15P68	L16P53	L17P31
L13P55	L14P33	L15P106	L15P7	L16P56	L17P35
L13P57	L14P34	L15P109	L15P71	L16P57	L17P4
L13P58	L14P35	L15P11	L15P72	L16P58	L17P40
L13P6	L14P38	L15P113	L15P73	L16P6	L17P46
L13P60	L14P39	L15P115	L15P76	L16P60	L17P48
L13P61	L14P40	L15P119	L15P77	L16P61	L17P5
L13P62	L14P43	L15P14	L15P80	L16P62	L17P51
L13P63	L14P45	L15P15	L15P97	L16P66	L17P52
L13P64	L14P46	L15P17	L15P98	L16P67	L17P54
L13P65	L14P49	L15P18	L15P99	L16P68	L17P55

L17P56	L18P53	L19P42	L20P28	L21P120	L22P127
L17P58	L18P55	L19P43	L20P29	L21P126	L22P130
L17P59	L18P57	L19P44	L20P32	L21P127	L22P2
L17P6	L18P6	L19P46	L20P35	L21P132	L22P26
L17P63	L18P60	L19P50	L20P37	L21P14	L22P29
L17P64	L18P63	L19P52	L20P38	L21P17	L22P39
L17P66	L18P64	L19P55	L20P4	L21P19	L22P42
L17P7	L18P65	L19P56	L20P40	L21P2	L22P44
L17P70	L18P68	L19P59	L20P41	L21P20	L22P46
L17P72	L18P69	L19P6	L20P42	L21P22	L22P49
L17P73	L18P7	L19P63	L20P43	L21P24	L22P50
L17P81	L18P73	L19P64	L20P44	L21P28	L22P54
L17P82	L18P76	L19P65	L20P45	L21P29	L22P57
L17P84	L18P77	L19P69	L20P46	L21P3	L22P6
L17P85	L18P79	L19P7	L20P49	L21P30	L22P62
L17P87	L18P80	L19P72	L20P5	L21P32	L22P63
L17P89	L18P83	L19P74	L20P50	L21P33	L22P68
L18P100	L18P84	L19P76	L20P55	L21P34	L22P69
L18P102	L18P86	L19P77	L20P56	L21P36	L22P7
L18P104	L18P88	L19P78	L20P6	L21P4	L22P71
L18P106	L18P89	L19P79	L20P61	L21P48	L22P74
L18P108	L18P95	L19P8	L20P63	L21P5	L22P81
L18P11	L19P100	L19P82	L20P64	L21P52	L22P82
L18P122	L19P102	L19P87	L20P68	L21P53	L22P83
L18P123	L19P108	L19P9	L20P70	L21P57	L22P85
L18P14	L19P109	L20P10	L20P72	L21P6	L22P86
L18P15	L19P11	L20P102	L20P73	L21P60	L22P87
L18P18	L19P110	L20P103	L20P74	L21P63	L22P90
L18P19	L19P111	L20P104	L20P79	L21P64	L22P94
L18P2	L19P117	L20P106	L20P80	L21P66	L23P100
L18P22	L19P126	L20P107	L20P82	L21P69	L23P109
L18P25	L19P128	L20P118	L20P83	L21P7	L23P115
L18P26	L19P130	L20P12	L20P86	L21P72	L23P123
L18P27	L19P131	L20P120	L20P87	L21P8	L23P127
L18P28	L19P14	L20P121	L20P88	L21P80	L23P129
L18P3	L19P16	L20P123	L20P9	L21P85	L23P139
L18P32	L19P18	L20P124	L20P92	L21P88	L23P14
L18P4	L19P2	L20P125	L20P93	L21P95	L23P34
L18P41	L19P20	L20P13	L20P99	L21P97	L23P4
L18P42	L19P23	L20P130	L21P1	L22P100	L23P48
L18P43	L19P25	L20P14	L21P10	L22P101	L23P5
L18P45	L19P28	L20P15	L21P103	L22P104	L23P60
L18P46	L19P29	L20P19	L21P112	L22P110	L23P61
L18P47	L19P36	L20P21	L21P116	L22P117	L23P7
L18P48	L19P37	L20P22	L21P117	L22P118	L24P10
L18P5	L19P4	L20P24	L21P118	L22P121	L24P106
L18P51	L19P41	L20P27	L21P12	L22P125	L24P112



L24P124	L25P58	L26P68	L27P65	L29P111	L30P50
L24P125	L25P6	L26P7	L27P67	L29P114	L30P52
L24P127	L25P61	L26P71	L27P68	L29P116	L30P55
L24P132	L25P65	L26P8	L27P69	L29P123	L30P56
L24P135	L25P66	L26P94	L27P77	L29P124	L30P69
L24P139	L25P69	L26P95	L27P84	L29P133	L30P99
L24P142	L25P7	L26P96	L27P85	L29P135	
L24P2	L25P77	L27P103	L27P9	L29P137	
L24P21	L25P78	L27P107	L27P93	L29P141	
L24P24	L25P87	L27P109	L27P95	L29P145	
L24P26	L25P88	L27P11	L28P100	L29P147	
L24P30	L25P90	L27P110	L28P107	L29P150	
L24P31	L25P93	L27P113	L28P109	L29P27	
L24P35	L25P94	L27P120	L28P110	L29P32	
L24P4	L26P1	L27P125	L28P111	L29P39	
L24P41	L26P103	L27P126	L28P112	L29P40	
L24P48	L26P108	L27P13	L28P117	L29P51	
L24P49	L26P110	L27P130	L28P122	L29P6	
L24P50	L26P113	L27P131	L28P126	L29P62	
L24P58	L26P114	L27P139	L28P131	L29P65	
L24P59	L26P117	L27P143	L28P133	L29P76	
L24P64	L26P120	L27P15	L28P134	L29P83	
L24P72	L26P121	L27P16	L28P139	L30P100	
L24P73	L26P123	L27P17	L28P141	L30P101	
L24P74	L26P126	L27P19	L28P143	L30P102	
L24P81	L26P128	L27P20	L28P147	L30P106	
L24P83	L26P130	L27P25	L28P32	L30P11	
L24P86	L26P132	L27P29	L28P34	L30P120	
L24P95	L26P137	L27P3	L28P35	L30P122	
L25P104	L26P16	L27P30	L28P4	L30P123	
L25P106	L26P23	L27P34	L28P40	L30P16	
L25P112	L26P25	L27P35	L28P41	L30P18	
L25P114	L26P26	L27P36	L28P51	L30P2	
L25P117	L26P29	L27P37	L28P53	L30P20	
L25P119	L26P32	L27P39	L28P55	L30P21	
L25P121	L26P35	L27P4	L28P57	L30P24	
L25P123	L26P36	L27P41	L28P6	L30P26	
L25P13	L26P4	L27P43	L28P60	L30P29	
L25P14	L26P42	L27P44	L28P62	L30P35	
L25P2	L26P45	L27P45	L28P63	L30P36	
L25P23	L26P46	L27P46	L28P7	L30P37	
L25P24	L26P50	L27P47	L28P74	L30P38	
L25P25	L26P58	L27P50	L28P77	L30P39	
L25P36	L26P6	L27P51	L28P78	L30P43	
L25P37	L26P63	L27P53	L28P84	L30P45	
L25P41	L26P65	L27P54	L29P102	L30P47	
L25P44	L26P66	L27P55	L29P108	L30P48	

### CAPÍTULO 3

## VALIDAÇÃO DE UM CHIP DE DNA (COFFEE AXIOM CHIP – 26K) PARA GENOTIPAGEM EM ESCALA GENÔMICA DE *Coffea canephora*

### 1 INTRODUÇÃO

O café é uma das *commodities* agrícolas mais importantes mundialmente e é o principal meio de subsistência para mais de 125 milhões de pessoas, sendo produzido em mais de 60 países. As duas espécies *Coffea arabica* e *Coffea canephora* predominam na produção mundial representando aproximadamente 60% e 40% do mercado cafeeiro, respectivamente (ICO, 2018).

Como para qualquer cultura, o estresse ambiental é um grande desafio para a produção de alimentos, uma vez que reduz os rendimentos em até 70% (AGARWAL et al., 2006). Em virtude do aquecimento global, as intempéries climáticas que ocorrem no país estão mais frequentes, provocando mudanças com um aumento previsto em radiação solar, temperatura, umidade, dentre outros, comprometendo o desenvolvimento e a produtividade das plantas (ASSAD et al., 2004).

Obter plantas que suportem estresses abióticos e que mantenham a produção e a qualidade se tornou o foco em programas de melhoramento, porém, por se tratar de uma espécie perene, o melhoramento convencional do cafeeiro demanda muito tempo e um custo elevado (MISHRA; SLATER, 2012). Uma alternativa para reduzir esse tempo são as novas abordagens que vem sendo utilizadas em estudos de genômica populacional e seleção genômica, entretanto, um grande número de marcadores distribuídos por todo o genoma é requerido, de forma a capturar a maioria dos efeitos (DAVEY et al., 2011; RESENDE et al., 2008). Estas metodologias estão começando a revolucionar a genômica aplicada ao melhoramento e genética de populações (POLAND; RIFE, 2012), exemplo disso são as tecnologias de *chips* de genotipagem de DNA, capazes de identificar uma enorme quantidade de SNPs por locus em plataformas como Illumina Infinium<sup>4</sup> ou Axiom Affymetrix<sup>5</sup> (CHEN et al., 2013).

---

<sup>4</sup> Disponível em: <http://www.illumina.com>

<sup>5</sup> Disponível em: <http://www.affymetrix.com>

Neste capítulo descrevemos o desenvolvimento e a validação de um chip de genotipagem de DNA (26K Axiom SNP array), desenvolvido para *C. canephora*.

## 2 MATERIAL E MÉTODOS

### 2.1 Material vegetal e resequenciamento

Primeiramente foi realizado o resequenciamento genômico de 14 pools formados por indivíduos de *C. canephora* Conilon (Tabela 5) e 10 pools abrangendo indivíduos dos diferentes grupos de diversidade de *C. canephora* (Tabela 6). Estes pools foram formados a partir de um estudo anterior realizado por Alekcevetch (2013), utilizando 11 microssatélites, e com base nos perfis multiloco, os indivíduos geneticamente mais próximos pertencem ao mesmo pool genético (Figura 20). Este painel de resequenciamento foi escolhido para cobrir a maior parte da diversidade genética presente na espécie estudada.

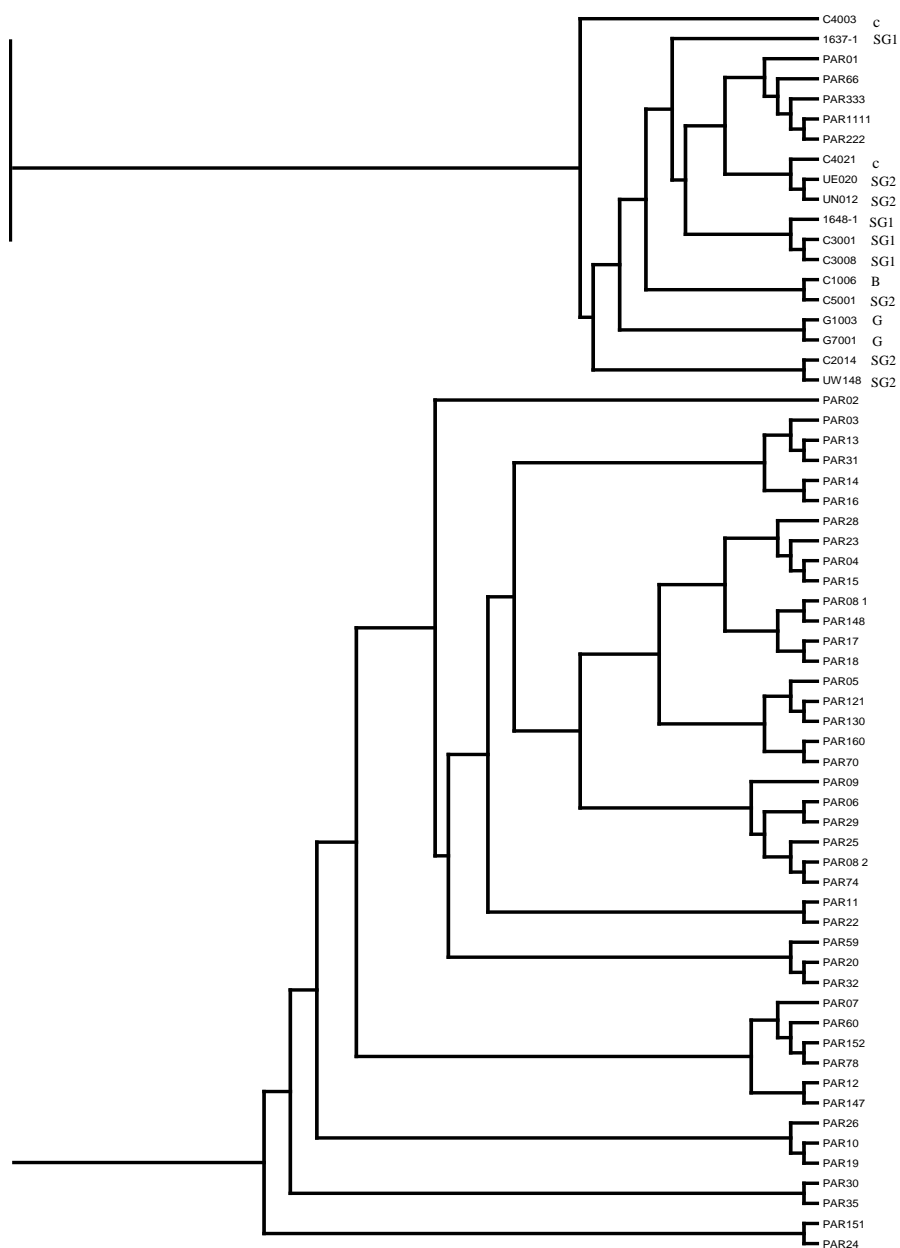
Tabela 4 – Painel de resequenciamento composto por 48 indivíduos de *C. canephora* Conilon dividido em 14 pools (C1-C14). O número de indivíduos, o número de *reads* produzidos e o número de *reads* mapeados são indicados para cada pool.

<b>Pool</b>	<b>Grupo</b>	<b>Nº de indivíduos</b>	<b>Read pairs</b>	<b>Reads mapeados</b>
C1	Conilon	5	51 935 630	49 014 161
C2	Conilon	1	52 852 588	48 707 510
C3	Conilon	5	59 725 554	54 196 408
C4	Conilon	4	52 128 826	48 008 860
C5	Conilon	4	60 632 930	55 733 196
C6	Conilon	5	63 529 706	59 253 749
C7	Conilon	6	40 549 350	36 535 906
C8	Conilon	2	47 569 112	41 653 248
C9	Conilon	3	53 429 582	48 412 359
C10	Conilon	4	40 807 252	36 557 443
C11	Conilon	2	42 186 500	37 608 626
C12	Conilon	3	58 682 334	52 871 878
C13	Conilon	2	47 381 834	44 309 089
C14	Conilon	2	65 571 840	60 637 139
<b>TOTAL</b>		<b>48</b>	<b>736 983 038</b>	<b>673 499 572</b>

Tabela 5 – Painel de resequenciamento composto por 51 indivíduos de *C. canephora* representando os diferentes grupos de diversidade, divididos em 10 pools (D1-D10). O grupo, subgrupo, número de indivíduos, o número de *reads* produzidos e o número de *reads* mapeados são indicados para cada pool.

<b>Pool</b>	<b>Grupo</b>	<b>Subgrupo</b>	<b>Nº de indivíduos</b>	<b><i>Read pairs</i></b>	<b><i>Reads mapeados</i></b>
D1	SG1	Luki	10	64 857 944	60 869 442
D2	SG1	Niaouli	9	69 425 608	58 272 686
D3	C	Nana	10	43 455 882	35 453 285
D4	B	Libengue	9	56 216 552	51 298 018
D5	SG2	INEAC	4	59 473 836	49 373 546
D6	SG2	Erect	2	67 321 990	62 140 815
D7	SG2	Nganda	2	61 836 730	58 363 232
D8	SG2	Wild	3	70 825 792	65 238 622
D9	G	Pelezi	1	50 711 840	39 259 360
D10	G	Mouniandougou	1	46 877 658	41 498 798
<b>TOTAL</b>			<b>51</b>	<b>591 003 832</b>	<b>521 767 804</b>

Figura 20 – Dendrograma UPGMA representando a similaridade genética entre os representantes dos diferentes grupos de diversidade de *C. canephora* e dos Conilons oriundos do BAG do Incaper.



Fonte: (Alekcevetch, 2013).

Extraíu-se o DNA desses indivíduos utilizando-se o protocolo CTAB (DOYLE; DOYLE, 1990). A quantificação e a integridade do DNA foram analisadas usando espectrofotômetro NanoVue (GE Healthcare Life Sciences) e gel de agarose 0,8% corado com brometo de etídio ( $1.0 \text{ mg L}^{-1}$ ), respectivamente.

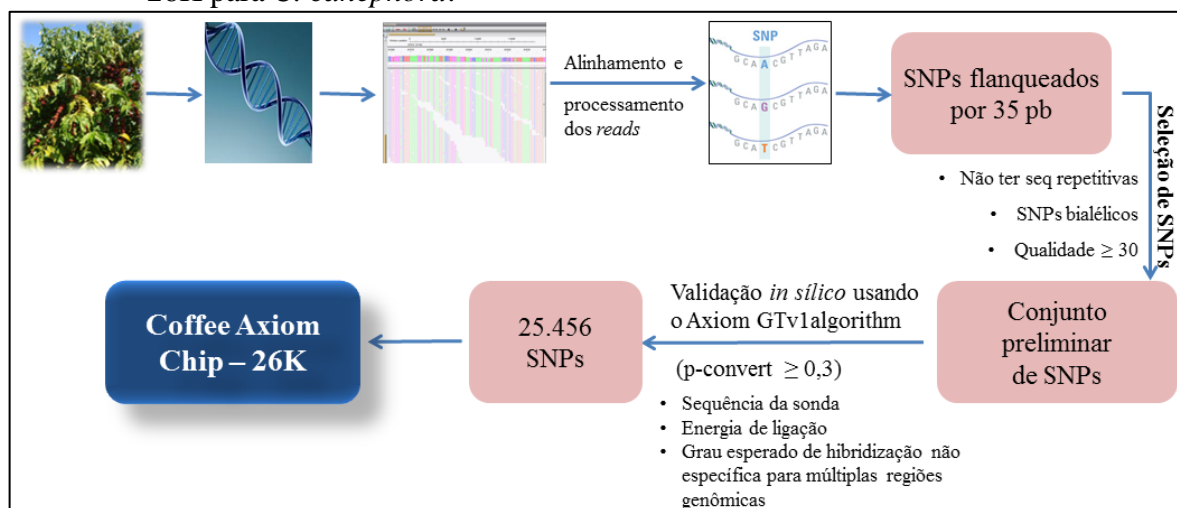
As bibliotecas de sequenciamento foram preparadas de acordo com o kit TruSeq DNA SamplePrep (Illumina) e o DNA genômico foi fragmentado entre 500 a 900 pares de base (pb), com tamanho médio de 700 pb. As bibliotecas foram quantificadas por qPCR e

sequenciadas por HiSeq2000, usando kit de sequenciamento SBS versão 3. Os arquivos Fastq foram gerados com Casava 1.8.2.

## 2.2 Seleção e validação dos SNPs

A partir dos dados de resequenciamento uma lista com possíveis SNPs para serem incluídos no *chip* foi obtida. Posteriormente, essa lista foi enviada a empresa desenvolvedora do *chip* e esses SNPs foram submetidos a uma validação *in silico*. Essa validação envolveu uma triagem preliminar em que ambas as sondas (*forward* e *reverse*) de cada SNP receberam um valor de *p-convert* gerados pelo *Affymetrix power tool (APT) AxiomGTv1algorithm*<sup>6</sup>, para prever a probabilidade de conversão do SNP e garantir um *chip* final de alta qualidade. O modelo considera fatores que incluem a sequência da sonda, a energia de ligação e o grau esperado de hibridização não específica para múltiplas regiões genômicas. Finalmente, as sondas desenhadas para os SNPs selecionados com altos valores de *p-convert* foram projetadas e sintetizadas com sucesso no *chip* (FIGURA 21).

Figura 21 – *Design* e desenvolvimento da plataforma de genotipagem Coffee Axiom Chip 26K para *C. canephora*.



Legenda: O *design* e o desenvolvimento do *chip* incluindo as quatro etapas principais: detecção dos SNPs, a partir de dados de resequenciamento de *C. canephora*, seleção dos SNPs que incluía filtragem de qualidade, validação *in silico* e colocação final na matriz.

Para validar os SNPs incluídos no *chip*, um total de 296 indivíduos (TABELA 7) de *C. canephora* foram genotipados. Vale ressaltar que esses materiais utilizados como “conjunto de validação” incluem os diferentes grupos de diversidade de *C. canephora*, as principais

<sup>6</sup> [http://www.affymetrix.com/estore/partners\\_programs/programs/developer/tools/powertools.affx](http://www.affymetrix.com/estore/partners_programs/programs/developer/tools/powertools.affx)

variedades clonais resultantes dos programas de melhoramento do INCAPER e Embrapa Rondônia, bem como representantes da população proveniente da Embrapa Cerrados.

Tabela 6 – Conjunto de validação do chip 26K de *C. canephora*.

<b>Genótipos</b>	<b>Grupo</b>	<b>Instituição</b>
BRS Ouro Preto	Conilon	Embrapa Rondônia
Centenária ES8132	Conilon	INCAPER
Diamante ES8112	Conilon	INCAPER
Jequitibá	Conilon	INCAPER
EMCAPA 8111	Conilon	INCAPER
EMCAPA 8121	Conilon	INCAPER
EMCAPA 8131	Conilon	INCAPER
EMCAPA 8141-Robustão Capixaba	Conilon	INCAPER
Vitória Incaper 8142	Conilon	INCAPER
CCE	Conilon	Embrapa Cerrados
B	Libengue	
C	Nana	
Hyb	Cong x Gui	
SG1	Luki	
SG1	Niaouli	
SG2		
Guineano		

Fonte: Do autor (2018).

A plataforma Affymetrix Gene Titan® foi usada para genotipar o “conjunto de validação” com o *chip* desenvolvido para *C. canephora*. Primeiramente, as sondas alvo foram preparadas usando cada amostra de DNA, extraído conforme o protocolo descrito por Russel et al. (2010), com uma quantidade mínima de 20 µL de DNA de boa qualidade e concentração de 10 ng/ µL. Estas amostras foram então amplificadas, fragmentadas aleatoriamente (25-125 pb), purificadas e hibridizadas no *chip*. As ligações aleatórias não específicas ligadas a alvos foram lavadas e para a identificação do SNP presente, o *chip* foi corado, fotografado e processado utilizando o instrumento Gene Titan® Multi-Channel (MC) para gerar os dados.

Para verificar a reprodutibilidade dos resultados obtidos, algumas amostras foram incluídas como replicatas técnicas dentro de uma mesma placa e entre placas.

### 2.3 Chamada de SNP e análise dos dados

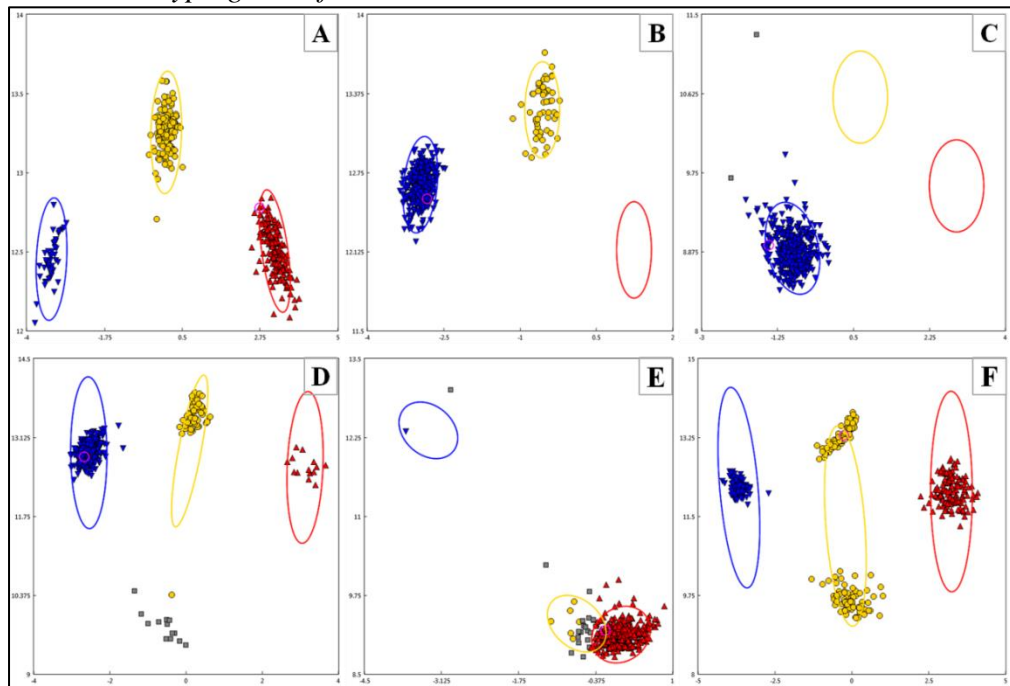
O *software* Axiom™ Analysis Suite (v. 1.1.1.66) foi usado para processamento dos dados de intensidade de hibridização, agrupamento e chamada de genótipos. Amostras que



apresentaram valor de *Dish QC* (DQC)  $< 0,82$  e *call rate*  $< 0,97$  foram removidas da análise, pois sua inclusão pode reduzir a qualidade dos resultados analisados.

Executando o *Axiom Best Practices Genotyping Workflow*, os SNPs identificados foram classificados dentro de seis tipos, de acordo com critérios de qualidade (FIGURA 22). Essas classes são: “*Poly High Resolution*” (PHR) (FIGURA 22A), que apresenta SNPs polimórficos e que passaram por todos os controles de qualidade (CQ); “*No MinorHomozygote*” (NMH), os SNPs passaram pelo CQ, porém somente dois *clusters* são observados (FIGURA 22B); “*Mono High Resolution*” (MHR), os SNPs passaram em relação à qualidade, mas são monomórficos (FIGURA 22C); “*Call Rate Below Threshold*” (CRBT), apresentam *call rate* abaixo de 97% (FIGURA 22D); “*Other*”, em que o padrão resultante de *cluster* de SNP não se encontra em nenhuma das classes anteriores (FIGURA 22E) e “*Off-Target Variant*” (OTV), apresentam um *cluster* adicional, de baixa intensidade, resultante de desajustes entre a sonda e as sequências para esse grupo de indivíduos (FIGURA 22F).

Figura 22 – Classificação, quanto à qualidade dos SNPs, gerada pelo *Axiom Best Practices Genotyping Workflow*.



Legenda: Classificação dos SNPs identificados em diferentes categorias. (A) *Poly High Resolution* (PHR); (B) *No MinorHomozygote* (NMH); (C) *Mono High Resolution* (MHR); (D) *Call Rate Below Threshold* (CRBT); (E) *Other*; (F) *Off-Target Variant* (OTV).

Fonte: Do autor (2018).

## **2.4 Genotipagem em larga escala usando o chip de dna 26k de *C. canephora***

### **2.4.1 Material vegetal e extração de DNA**

Os 1.319 indivíduos de *C. canephora* selecionados na população da Embrapa Cerrados, a partir das avaliações fenotípicas, como descrito no Capítulo 2, foram utilizados para realizar uma genotipagem em larga escala usando o *chip* de DNA 26K desenvolvido e validado para *C. canephora*. O DNA genômico foi extraído, conforme descrito por Russel et al., (2010). A quantificação e a integridade do DNA foram analisadas usando espectrofotômetro NanoVue (GE Healthcare Life Sciences) e gel de agarose 0,8% corado com brometo de etídio (1.0 mg L<sup>-1</sup>), respectivamente.

### **2.4.2 Análise da diversidade e desequilíbrio de ligação**

Os dados de genotipagem gerados foram usados nas análises de diversidade genética e para o cálculo do desequilíbrio de ligação (DL) da população. Informações como MAF, número de heterozigotos, dados faltantes, entre outras, foram obtidas usando o software TASSEL (v. 5.2.43).

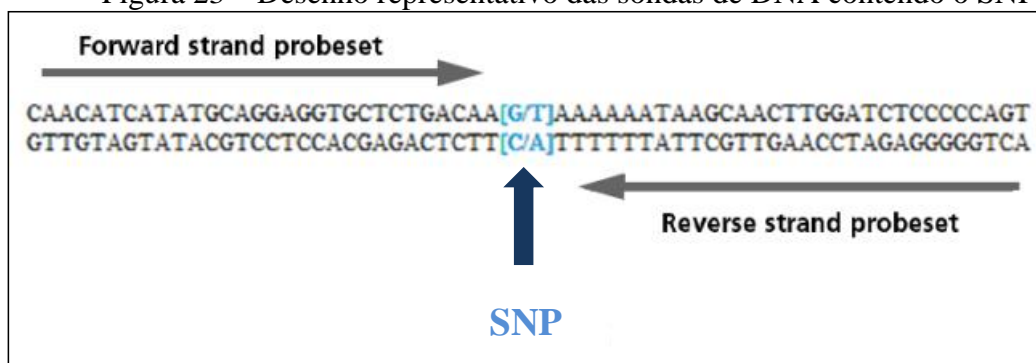
O desequilíbrio de ligação (DL) foi calculado para cada cromossomo em função da distância física dos marcadores. O valor de  $r^2$  foi estimado usando o pacote LDCORSV e a curva de decaimento foi ajustada por uma regressão não linear (HILL; WEIR, 1988, REMINGTON et al., 2001) usando o *script* R de Marroni et al. (2011).

### 3 RESULTADO E DISCUSSÃO

#### 3.1 Alinhamento e detecção dos SNPs

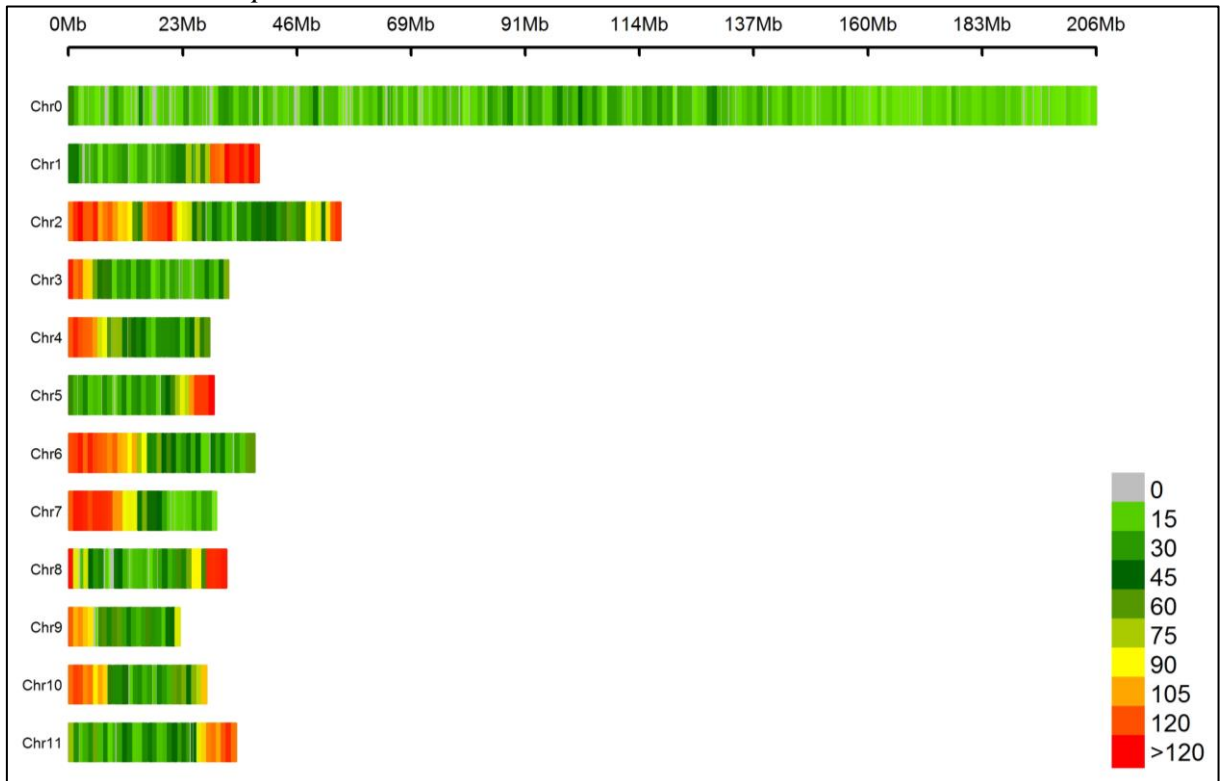
Conforme descrito na seção “Material e Métodos”, a partir dos dados de mapeamento dos *reads* obtidos no resequenciamento dos 99 indivíduos (TABELAS 5 e 6) contra o genoma de referência de *C. canephora* (DENOEUDE et al., 2014), assim como das análises de bioinformática posteriores para a filtragem dos SNPs, um total de 25.456 sondas de DNA, contendo os SNPs de alta qualidade (FIGURA 23), foram selecionadas e incluídas na construção do *chip*. A seleção dos SNPs teve por base os critérios estabelecidos pela Affymetrix, para a construção das sondas. Do conjunto de sondas incluídas no *chip*, 25.411 cobriram todos os 11 cromossomos de *C. canephora* e o “cromossomo 0”, que consiste em um conjunto de *scaffolds* não ordenados, fornecendo uma boa representação do genoma total da espécie (FIGURA 24), com uma média de 2.117 sondas por cromossomo e 45 sondas foram excluídas por mapearem em múltiplos locais.

Figura 23 – Desenho representativo das sondas de DNA contendo o SNP.



Legenda: SNP, em azul, flanqueado pelo conjunto de sondas *forward* e *reverse*. As sondas apresentam 71 pb, sendo 35 pb anteriores e 35 pb posteriores ao SNP.

Figura 24 – Distribuição dos SNPs contidos nas sondas para cada cromossomo de *C. canephora*.



Legenda: Número de SNPs dentro de uma janela de 1Mb. A quantidade de SNPs varia de 0 (em cinza claro) até >120 (em vermelho) SNPs por Mb para os 11 cromossomos de *C. canephora*, incluindo o cromossomo 0.

Fonte: Do autor (2019).

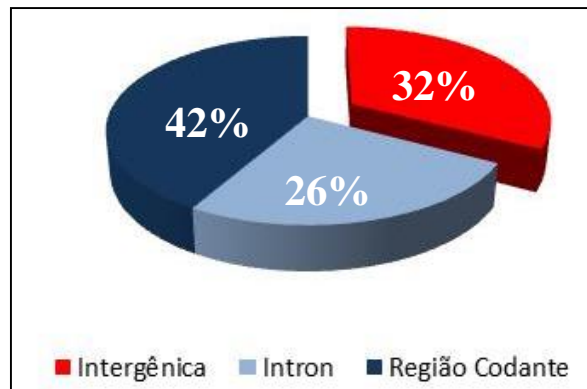
Verifica-se na Tabela 8 que o maior número de sondas localiza-se no cromossomo 2 (4.209), sendo este o maior cromossomo da espécie, com aproximadamente 54 Mb, enquanto que o menor número de sondas foram mapeadas no cromossomo 9 (1.257).

Com relação às posições genômicas dos SNPs fixados no *chip* 26K de *C. canephora*, 68% estavam presentes em região gênica, sendo que desses, 42% estavam localizados em éxons (região codante – CDS) e 26% em íntrons. Os demais SNPs identificados no *chip* (8.348) estavam localizados em regiões intergênicas (FIGURAS 25 e 26).

Tabela 7 – Número de SNPs mapeados em cada cromossomo de *C. canephora*.

	Tam. cromossomo (Mb)	Nº de sondas mapeadas	Região Gênica		Região intergênica
			CDS	Íntron	
Chr 0	205.6	3.381	830	563	1.988
Chr 1	38.2	2.121	928	576	617
Chr 2	54.5	4.209	1.915	1.248	1.046
Chr 3	32.0	1.400	599	372	429
Chr 4	28.2	1.789	841	436	512
Chr 5	29.1	1.478	651	332	495
Chr 6	37.3	2.553	1.102	715	736
Chr 7	29.8	2.047	984	575	488
Chr 8	31.6	1.688	685	490	513
Chr 9	22.3	1.257	503	331	423
Chr 10	27.6	1.762	777	444	541
Chr 11	33.5	1.726	764	402	560
	<b>570</b>	<b>25.411</b>	<b>10.579</b>	<b>6.484</b>	<b>8.348</b>

Fonte: Do autor (2018).

Figura 25 - Distribuição de SNPs em diferentes regiões genômicas do genoma de *C. canephora*.

Legenda: Distribuição das sondas identificadas no *chip* na região intergênica (em vermelho), região intrônica (em azul escuro) e região codante (em azul claro).

Fonte: Do autor (2019).

Figura 26 – Posição das sondas na região gênica e intergênica do genoma de *C. canephora*.

Legenda: As sondas, representadas pelo triângulo cinza, localizadas na região codante (CDS), em amarelo, ou em região de íntron, do gene *Cc10\_g11070* de *C. canephora*, indicado em verde. Outra sonda é representada fora do gene, em região intergênica.

Fonte: Do autor (2019).

Os marcadores moleculares têm sido amplamente utilizados para avaliação da diversidade genética, estudos evolutivos e de mapeamento. A aplicabilidade desses no melhoramento depende do custo, facilidade e precisão, tornando os SNPs uma escolha indispensável. O rápido progresso nas tecnologias de NGS durante a última década permitiu a produção de dados de sequência a baixo custo e em muito menos tempo (THUDI et al., 2012).

As tecnologias baseadas em NGS foram efetivamente utilizadas para sequenciamento e resequenciamento de genomas, possibilitando a identificação de um número muito grande de marcadores SNPs. No cafeeiro, diferentes plataformas de genotipagem de SNPs foram desenvolvidas para várias aplicações (CARNEIRO et al., 2014, FERRÃO et al., 2017; SANT'ANA et al., 2018; SOUSA et al., 2017). Atualmente são os *chips* de DNA, com uma alta densidade de SNPs, que tem se mostrado uma ferramenta poderosa para o melhoramento molecular e a investigação genética de populações dentro de espécies de café (ANDRADE et al., 2017; MEROT-L'ANTHOENE et al., 2018), pois são de fácil utilização e econômicos para a geração e análise de dados de genotipagem. Além disso, a plataforma apresenta alta precisão e reprodutibilidade.

### 3.2 Validação do chip de genotipagem

O conjunto de validação do *chip* de alta densidade com 25.456 SNPs foi formado por 296 indivíduos de *C. canephora*, produzindo mais de 97% de taxa de sucesso de amostragem, sendo que as oito amostras excluídas não passaram devido ao QC *call rate*, que variou de 92,03 a 96,94 para essas amostras. As amostras incluídas como replicatas, dentro e entre placas, apresentaram uma concordância > 99,82 entre os resultados.

A validação da genotipagem resultou em 22.679 SNPs (89%) de excelente qualidade (TABELA 9). Destes, cerca de 90% eram polimórficos (PHR), os outros 10% eram marcadores que apresentavam somente dois *clusters* (NMH) e marcadores monomórficos (MHR). Os marcadores remanescentes (CRBT+OTV+Other) e que apresentam menor acurácia, representaram menos de 11% (2.777) do total de SNPs.

Tabela 8 – Número e porcentagem de SNPs classificação em cada categoria.

<b>Categoria do SNP</b>	<b>Número de SNPs</b>	<b>% dos SNPs</b>
PHR	20.920	82,18%
NMH	1.569	6,16%
MHR	190	0,74%
CRBT	571	2,24%
OTV	583	2,29%
Other	1.623	6,38%
<b>TOTAL</b>	<b>25.456</b>	

Legenda: Os SNPs são classificados como PHR – Poly High Resolution, NMH – No Minor Homozygote, MHR – Mono High Resolution, CRBT – Call Rate Below Threshold, OTV – Off-Target Variant e Other.

Fonte: Do autor (2018).

A grande maioria dos SNPs pertence à classe mais bem definida de polimorfismos PHR (20.920, 82%). Este resultado está de acordo com outros chips publicados, melhorando as 60% e 74% das variantes PHR obtidas no *chip* 180K SoyaSNP (LEE et al., 2015) e Apple 480K (BIANCO et al., 2016), desenvolvidos para soja e maçã, respectivamente, mas ficando aquém dos 92% de variantes PHR apresentadas pela matriz Maize 600K (UNTERSEER et al., 2014), desenvolvida para milho. No entanto, vale ressaltar que, neste último caso, uma abordagem mais onerosa foi seguida com a produção de dois *arrays* de triagem e a incorporação final das 50% variantes mais confiáveis.

Diversos *chips* de genotipagem de SNP, incluindo de média a alta densidade, foram desenvolvidos e usados para GWAS e outras aplicações no melhoramento de amendoim (PANDEY et al., 2017), arroz (CHEN et al., 2014; McCOUCH et al., 2010; SINGH et al., 2015; YU et al., 2014), trigo (MACCAFERRI et al., 2015), dentre outras culturas. Em café não é diferente, sendo que essa grande quantidade obtida de SNPs de qualidade, fornece uma ferramenta poderosa e essencial para desenvolver GWAS, SGA, pesquisa de genética populacional, e outros vários trabalhos que auxiliarão no melhoramento genético do cafeeiro.

### 3.3 Análise filogenética do conjunto de validação

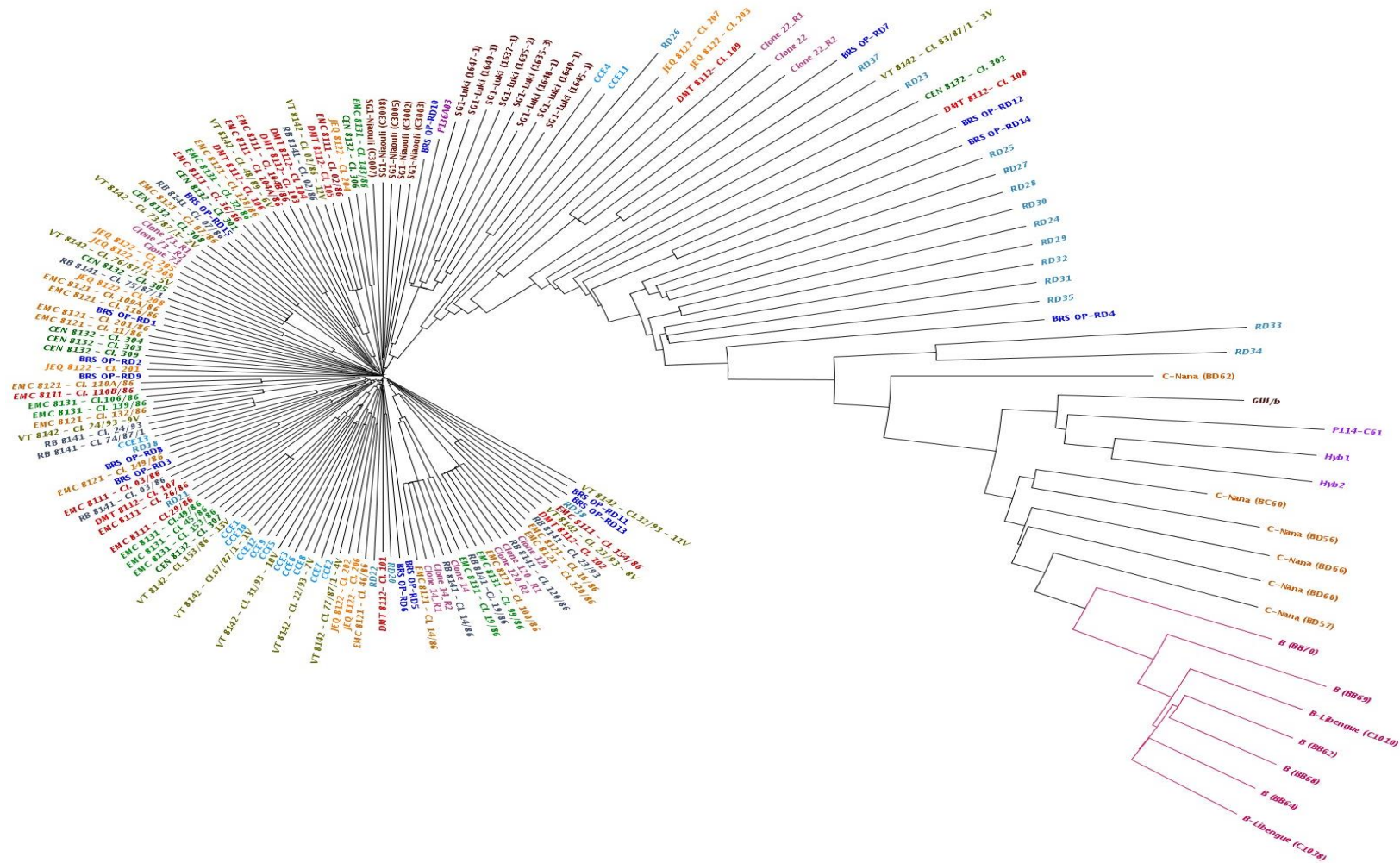
Utilizando somente os SNPs classificados como polimórficos (PHR), realizou-se uma análise filogenética para os indivíduos utilizados na validação do *chip*. Nesta análise estão presentes os indivíduos de *C. canephora* Conilon pertencentes às principais variedades comerciais lançadas pelo Incaper e pela Embrapa Rondônia, os genótipos representantes dos grupos de diversidade da espécie e representantes da população proveniente da Embrapa Cerrados.

Observando a Figura 27, nota-se que a maior parte das variedades comerciais do Brasil estão próximas dos indivíduos que representam o grupo SG1 de *C. canephora*, corroborando com os dados apresentados por Montagnon, Cubry e Leroy. (2012). Observa-se também que diferentes variedades registradas/ protegidas possuem materiais muito semelhantes geneticamente, é o caso, por exemplo, das variedades RB8141-Cl. 02/86 (Robustão Capixaba), VT 8142-Cl. 02/86-12V (Vitória-Incaper 8142), DMT 8112-Cl. 105 (Diamante ES8112) e EMC8111-Cl. 02/86 (Emcapa 8111).

Característica interessante é em relação à variedade BRS Ouro Preto (BRS OP-RD), percebe-se que os 15 clones que a compõem estão amplamente distribuídos na árvore, abrangendo genótipos próximos ao grupo SG1, mas também apresentando alguns materiais que se aproximam dos demais grupos de diversidade, como por exemplo, o clone BRS OP-RD4 que está mais próximo dos indivíduos do grupo C – Nana.



Figura 27 – Dendograma de proximidade dos grupos com as principais variedades comerciais do Brasil (Conilon) e os genótipos representantes dos diferentes grupos de diversidade de *C. canephora*

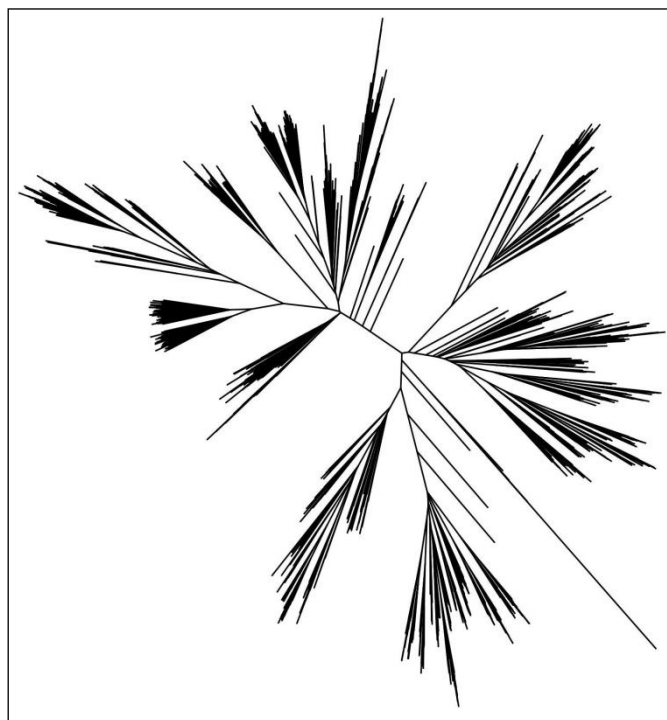


A partir da validação do chip 26K Axiom de *C. canephora* pode-se dizer que a genotipagem se mostrou coerente e confiável, gerando uma grande quantidade de marcadores SNPs polimórficos de qualidade. Ao realizar a genotipagem, todas as replicatas incluídas foram identificadas, como os clones 14, 22, 73 e 120 de *C. canephora*, presentes na Figura 27. O dendograma nos permitiu concluir que o agrupamento para a maioria dos indivíduos está de acordo com o esperado, exemplo disso são os indivíduos que pertencem a um mesmo grupo de diversidade estar em um mesmo ramo da árvore, bem como alguns clones das variedades comerciais estarem próximos.

### **3.4 Desequilíbrio de ligação e estrutura da população da Embrapa Cerrados**

Com o intuito de obter uma ideia geral a respeito da estrutura da população localizada na Embrapa Cerrados, visto que a estrutura afeta diretamente as análises de GWAS e de SGA, a árvore filogenética com base na análise do vizinho mais próximo (*Neighbor-joining*) foi gerada para os 1.319 indivíduos selecionados na população de melhoramento, utilizando 16.688 marcadores SNPs (FIGURA 28). Visualiza-se na Figura 28 a possível existência de 5 a 6 grupos nessa população.

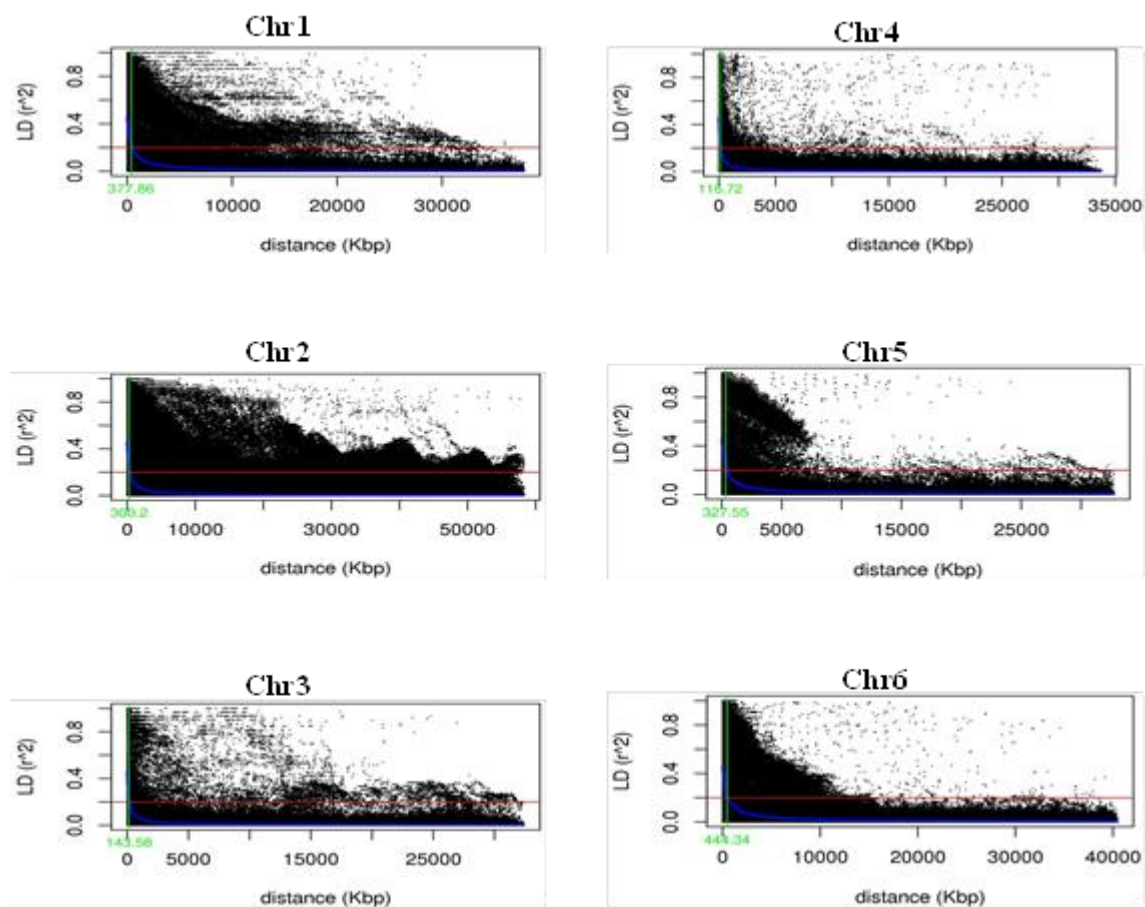
Figura 28 – Árvore filogenética (*Neighbor-joining*) entre os 1.319 indivíduos, utilizando-se 16.688 marcadores SNPs.



Fonte: Do autor (2018).

A Figura 27 apresenta os parâmetros de  $r^2$  estimados em função da distância física entre loci para cada cromossomo de *C. canephora*. Verifica-se que o decaimento do desequilíbrio de ligação variou de 444,34 Kbp para o cromossomo 6 até 116,72 Kbp, para o cromossomo 4.

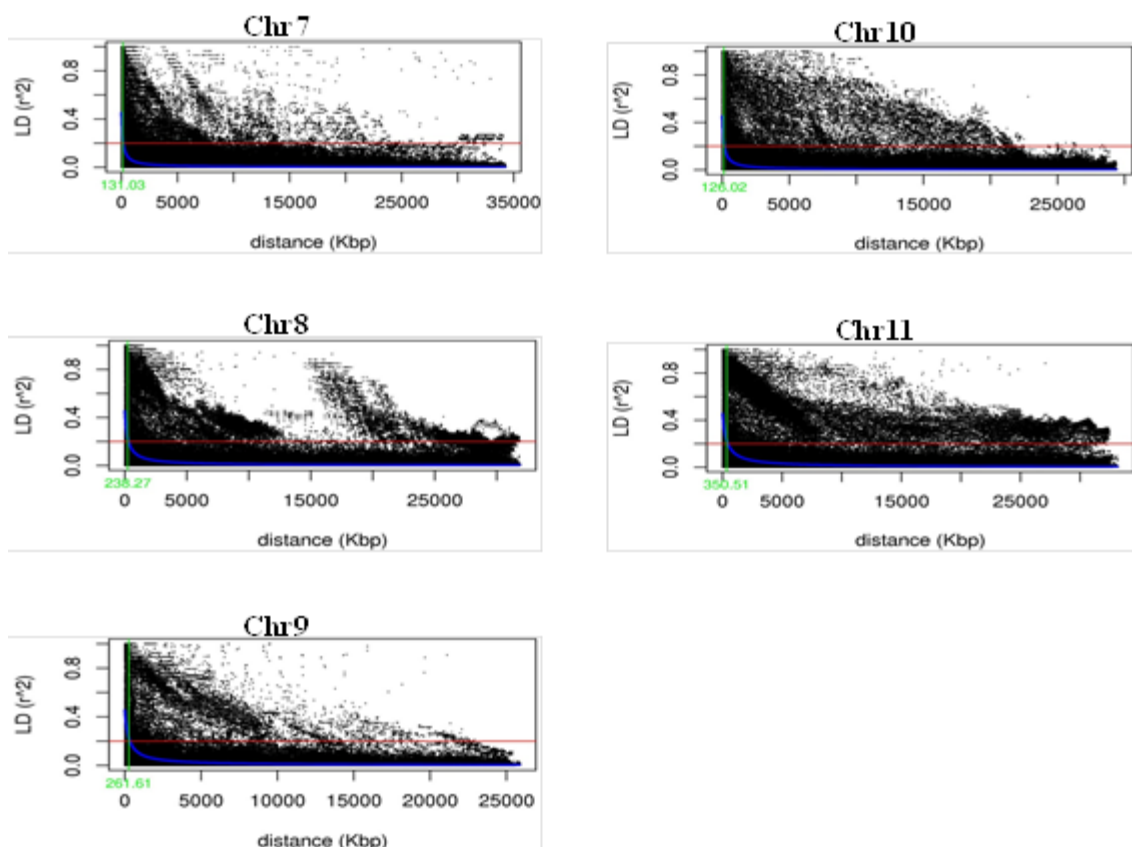
Figura 29 – Análise do decaimento do desequilíbrio de ligação ( $r^2$ ) em *C. canephora*.  
(Continua)



Legenda: O desequilíbrio de ligação, em função da distância física (Kpb) entre os marcadores de cada cromossomo (Chr 1 – Chr 11) de *C. canephora* para os 1.319 indivíduos provenientes da população da Embrapa Cerrados. Em verde encontram-se os valores (em Kbp) que o valor de  $r^2$  atingiu 0,2.

Fonte: Do autor (2018).

Figura 30 – Análise do decaimento do desequilíbrio de ligação ( $r^2$ ) em *C. canephora*. (Conclusão)



Legenda: O desequilíbrio de ligação, em função da distância física (Kbp) entre os marcadores de cada cromossomo (Chr 1 – Chr 11) de *C. canephora* para os 1.319 indivíduos provenientes da população da Embrapa Cerrados. Em verde encontram-se os valores (em Kbp) que o valor de  $r^2$  atingiu 0,2.

Fonte: Do autor (2018).

O DL foi calculado para cada cromossomo separadamente, pois para GWAS e SGA esse conhecimento é fundamental, uma vez que fornece uma indicação da forma como os marcadores e os traços fenotípicos a serem estudados serão correlacionados. Além disso, estabelece a quantidade de marcas necessárias para a realização de diversos estudos, incluindo os estudos de associação.

Merot-L'Anthoene et al. (2018) relataram como muito rápido o decaimento do DL em *C. canephora* calculado a partir de um conjunto de resequenciamento formado por 12 genótipos de diferentes grupos genéticos, chegando a um  $r^2 = 0,2$  a uma distância de 8,1Kb. Ao realizar esse cálculo dentro de um único grupo genético, o decaimento relatado foi de 17,5Kb, em média. Nesse trabalho, o DL foi calculado para os 1.319 indivíduos dentro de uma mesma população, não abrangendo diferentes grupos genéticos, isso poderia explicar o decaimento ser mais lento (variando de 116 a 444Kb entre os cromossomos de *C. canephora*).

## 4 CONCLUSÃO

O desenvolvimento da plataforma de genotipagem 26K Axiom® SNP array para *C. canephora*, com um grande número de SNPs validados e, acima de tudo, com a alta qualidade e reprodutibilidade destes marcadores, representam uma inovação na espécie estudada e disponibiliza uma ferramenta robusta para posteriores estudos de GWAS e SGA, genética populacional, podendo contribuir para o aumento na eficiência dos programas de melhoramento genético do cafeeiro.

## REFERÊNCIAS

- AGARWAL, P. K. et al. Role of DREB transcription factors in abiotic and biotic stress tolerance in plants. **Plant Cell Reports**, v. 25, n. 12, p. 1263-1274, 2006.
- ALEKCEVETCH, J. C. **Estudo da diversidade genética, por meio de marcadores moleculares de uma população de *Coffea canephora* var. Conilon**. 2013. 92 p. Dissertação (Mestrado em Biotecnologia Vegetal) – Universidade Federal de Lavras, Lavras, 2013.
- ASSAD, E. D. et al. Impacto das mudanças climáticas no zoneamento agroclimático do café no Brasil. **Pesquisa Agropecuária Brasileira**, v. 39, n. 11, p. 1057-1064, 2004.
- BIANCO, L. et al. Development and validation of the Axiom® Apple480K SNP genotyping array. **The Plant Journal**, v. 86, n. 1, p. 62-74, 2016.
- CHEN, H. et al. A High-Density SNP Genotyping Array for Rice Biology and Molecular Breeding. **Molecular Plant**, v. 7, n. 3, p. 541-553, 2014.
- CHEN, H. et al. Development of genomics-based genotyping platforms and their applications in rice breeding. **Current Opinion in Plant Biology**, v. 16, n. 2, p. 247-254, 2013.
- DAVEY, J. W. et al. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. **Nature Reviews Genetics**, v. 12, n. 7, p. 499-510, 2011.
- DENOEUDE, F. et al. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. **Science**, v. 345, n. 6201, p. 1181-1184, 2014.
- DOYLE, J. J.; DOYLE, J. L. Isolation of plant DNA from fresh tissue. **Focus**, v. 12, p. 13-15, 1990.
- HILL, W. G.; WEIR, B. S. Variances and covariances of squared linkage disequilibria in finite populations. **Theoretical Population Biology**, v. 33, n. 1, p. 54-78, 1988.
- ICO – International Coffee Organization. Disponível em: <<http://www.ico.org/>>. Acesso em: 18 jan. 2019.
- LEE, Y. G. et al. Development, validation and genetic analysis of a large soybean SNP genotyping array. **The Plant Journal**, v. 81, n. 4, p. 625-636, 2015.
- MACCAFERRI, M. et al. A high-density, SNP-based consensus map of tetraploid wheat as a bridge to integrate durum and bread wheat genomics and breeding. **Plant Biotechnology Journal**, v. 13, n. 5, p. 648-663, 2015.
- MARRONI, F. et al. Nucleotide diversity and linkage disequilibrium in *Populus nigra* cinnamyl alcohol dehydrogenase (CAD4) gene. **Tree Genetics & Genomes**, v. 7, n. 5, p. 1011-1023, 2011.
- MCCOUCH, S. R. et al. Development of genome-wide SNP assays for rice. **Breeding Science**, v. 60, p. 524-535, 2010.

MISHRA, M. K.; SLATER, A. Recent advances in the genetic transformation of coffee. **Biotechnology Research International**, p. 1-17, 2012.

MONTAGNON, C.; CUBRY, P.; LEROY, T. Coffee *Coffea canephora* Pierre genetic improvement: acquired knowledge, strategies and perspectives. **Cahiers Agricultures**, v. 21, n. 2-3, p. 143-153, 2012.

PANDEY, M. K. et al. Development and Evaluation of a High Density Genotyping 'Axiom\_Arachis' Array with 58 K SNPs for Accelerating Genetics and Breeding in Groundnut. **Scientific Reports**, v. 7, n. 40577, 2017.

POLAND, J. A.; RIFE, T. W. Genotyping-by-Sequencing for Plant Breeding and Genetics. **Plant Genome**, v. 5, n. 3, p. 92-102, 2012.

REMYINGTON, D. L. et al. Structure of linkage disequilibrium and phenotypic associations in the maize genome. **Proceedings of the National Academy of Sciences**, v. 98, n. 20, p. 11479-84, 2001.

RESENDE, M. D. V. et al. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal Brasileira**, n. 56, p. 63-77, 2008.

RUSSELL, A. et al. Phylogenetics and cytology of a pantropical orchid genus *Polystachya* (Polystachyinae, Vandaeae, Orchidaceae): evidence from plastid DNA sequence data. **Taxon**, v. 59, n. 2, p. 389-404, 2010.

SINGH, N. et al. Single-copy gene based 50 K SNP chip for genetic studies and molecular breeding in rice. **Scientific Reports**, v. 5, n. 11600, 2015.

UNTERSEER, S. et al. A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. **BMC Genomics**, v. 15, n. 1, p. 1-15. 2014.

YU, H. et al. A whole-genome SNP array (RICE6K) for genomic breeding in rice. **Plant Biotechnology Journal**, v. 12, n. 1, p. 28-37, 2014.



## CAPÍTULO 4

### ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA (GWAS) EM *C. canephora*

#### 1 INTRODUÇÃO

A relação causal entre o polimorfismo genético dentro de uma espécie e as diferenças fenotípicas observadas entre indivíduos é de interesse biológico fundamental. Essa relação entre fenótipo e genótipo tem sido de grande interesse desde que Mendel postulou a existência de "fatores internos" que são transmitidos para a próxima geração (KORTE; FARLOW, 2013).

Um dos principais desafios que a comunidade agrícola enfrenta, na medida em que busca integrar a informação genética no melhoramento de plantas, é que, até recentemente, a aquisição de dados genotípicos era caro e laborioso. Como resultado, as aplicações genômicas ficavam restritas a marcadores moleculares que poderiam ser testados ao longo das gerações, dessa forma, a maioria dos genes caracterizados nas espécies eram aqueles associados a traços mendelianos, ou seja, traços controlados por alguns genes de grande efeito. Sabe-se, no entanto, que a maioria das características agrônômicas de interesse no melhoramento é quantitativa e poligênica, controlada por muitos genes de pequeno efeito (SPINDEL et al., 2016).

O Estudo de Associação Genômica Ampla (*Genome-Wide Association Studies* – GWAS) surgiu com o intuito de capturar o efeito desses genes na determinação da característica. Para isso, os indivíduos selecionados são genotipados e fenotipados para as características de interesse e então, a partir de análises estatísticas entre os polimorfismos de DNA e a variação no fenótipo, os genes que controlam essas características são identificados (VARSHNEY, TERAUCHI, MCCOUCH, 2014).

A cafeicultura sempre teve grande importância para a economia brasileira, contribuindo de maneira decisiva para a industrialização do país. Atualmente, o Brasil se destaca na produção e exportação dessa *commodity* (ICO, 2018), sendo a cultura conduzida, em sua maioria, por pequenos produtores e proporcionando melhorias na qualidade de vida.

A produção cafeeira pode ser afetada por diversos fatores, como a bienalidade, característica intrínseca da espécie, e as alterações climáticas, que podem prejudicar o desenvolvimento da planta, a formação de grãos e, conseqüentemente, a produção e a

qualidade final do produto (DAMATTA et al., 2010). Os fatores bióticos também podem causar grandes prejuízos, dentre eles estão o bicho mineiro (*Leucoptera coffeella* Guérin-Mèneville & Perrottet), a cercosporiose (*Cercospora coffeicola* Berk et Cook e a ferrugem (*Hemileia vastratrix* Berk et Br.).

Por se tratar de uma espécie perene e de ciclo longo, os programas de melhoramento genético convencional dessa espécie são demorados (vários anos) e de custo elevado. A geração de cultivares superiores via melhoramento genético ainda lida com o desafio de agregar, simultaneamente, diversas características quantitativas de relevância agrônômica e qualidade de bebida.

Neste sentido, a GWAS representa um grande avanço no melhoramento da espécie, auxiliando a identificação de genótipos superiores que seriam fontes de genes de interesse. Desta forma, o presente trabalho tem como objetivo conduzir um estudo de associação genômica ampla (GWAS) para algumas características de interesse no melhoramento de *C. canephora*.

## 2 MATERIAL E MÉTODOS

### 2.1 Material vegetal

Na condução de GWAS, duas análises foram realizadas separadamente. A primeira, com os dados dos 366 indivíduos que contém as informações de produção, peso de 100 grãos, boia, tamanho e formato dos grãos (peneiras), morfologia dos frutos e potencial hídrico foliar de antemanhã. Na segunda, todos os 1.319 indivíduos de *C. canephora* que possuem todas as informações fenotípicas citadas acima, exceto de potencial hídrico foliar de antemanhã. Todos os dados relacionados ao fenótipo dos indivíduos foram apresentados no Capítulo 2.

### 2.2 Genotipagem

Os dados genotípicos utilizados para a realização de GWAS foram os marcadores SNPs identificados no chip de genotipagem 26K Axiom de *C. canephora*. Após um processamento mínimo, utilizando a função *A.mat* do pacote rrBLUP (ENDELMAN, 2011), no software R (R Core Team, 2017), obteve-se um conjunto de 18.942 e 16.688 marcadores, considerando os 366 e 1.319 indivíduos, respectivamente. Portanto, após essa filtragem, somente os marcadores polimórficos, com *call rate* acima de 90% e com MAF (*minimum allele frequency*) superior a 1% foram utilizados nas análises.

Além dos SNPs identificados por meio do *chip* de genotipagem, outros marcadores SNPs foram também incorporados nas análises com os 366 indivíduos (CARNEIRO, 2014), já que estes haviam sido genotipados anteriormente utilizando outra técnica denominada nextRAD (*Nextera-tagmented reductively-amplified DNA*), assim pode-se verificar se esse acréscimo de marcadores aumentaria o poder das análises de associação.

De um total de 11.230 SNPs identificados pela genotipagem nextRAD, utilizando os mesmos critérios de filtragem acima citados, 5.982 SNPs polimórficos, com *call rate* acima de 90% e com MAF superior a 1% foram adicionados as análises para os 366 indivíduos.

### 2.3 Análise de associação genômica ampla (GWAS)

Para as análises de associação genômica ampla (GWAS) entre os marcadores SNP e os dados fenotípicos, dois programas diferentes foram utilizados com a finalidade de evidenciar a capacidade computacional dessas plataformas. O primeiro foi utilizando o pacote rrBLUP (ENDELMAN, 2011) implementado no *software* R, função GWAS, e o segundo foi o *software* TASSEL (BRADBURY et al., 2007), ambos tem como princípio o uso da equação de modelos mistos proposto por Yu et al. (2006).

Para confirmar a significância das associações entre os SNPs e os fenótipos de interesse, dois métodos foram utilizados: (i) correção de Bonferroni e (ii) FDR (*False Discovery Rate*) (BENJAMINI; HOCHBERG, 1995), adotando um nível de significância de 5% (p-valor < 0,05). A correção de Bonferroni simplesmente divide o nível de significância em cada locus pelo número de testes. Em outras palavras, ela ajusta o valor alfa ( $\alpha$ ) de 0,05 para  $\alpha = (0,05 / k)$ , em que, nesse estudo, k é o número de marcadores utilizados na análise. Para o cálculo do FDR, utilizou-se o pacote *qvalue* (STOREY; TIBSHIRANI, 2003; STOREY, 2002), disponível em R, que tem por finalidade reduzir o número de falsos positivos que podem decorrer de associações por vínculo genético. Os marcadores foram definidos como estando significativamente associados com base no p-valor ajustado pelo FDR de 0,05, ou seja, todos SNPs associados com um p-valor (FDR) < 0,05 foram considerados significativos durante a análise. A partir dos dados de p-valores de significância para cada método, o pacote CMplot, implementado no *software* R, gerou os gráficos do tipo *Manhattan Plot*, para observar a dispersão das associações entre os marcadores SNP e a característica de interesse. Para gerar os gráficos Manhattan, o cromossomo 0 foi denominado de cromossomo 12 pois o pacote utilizado não aceita a informação “Cromossomo 0”. Os gráficos do tipo QQ plot, que tem como objetivo relacionar as associações encontradas em relação às associações esperadas, foram gerados pelo rrBLUP.

### 3 RESULTADOS E DISCUSSÃO

#### 3.1 Análise de 366 indivíduos

##### 3.1.1 Dados genotípicos do chip 26K Axiom de *C. canephora* – rrBLUP

Ao analisar os 366 indivíduos com os 18.942 marcadores presentes no *chip* e mantidos após os filtros aplicados, como descrito em Material e Métodos, observa-se que as características de formato e tamanho de grãos apresentaram associação para Chato Médio, Moca Graúdo e Moca Miúdo, conforme apresentado na Figura 30A, 30B e 30C, respectivamente. Considerando o FDR, como limite de significância, três SNPs foram identificados para a característica Chato Médio, todos localizados no cromossomo 6 (FIGURA 30A). Um desses SNPs está localizado numa região intergênica (AX-168345461) e os outros dois em regiões gênicas (AX-168307159 e AX-168298103). O marcador AX-168307159, se localiza no gene Cc06\_g14970, que codifica uma proteína *leucina-aminopeptidase-like*, pertencente à superfamília peptidase M-17. Já o marcador AX-168298103, se localiza no gene Cc06\_g14920, que codifica uma *flavonol synthase/flavanone 3-hydroxylase-like*, pertencente à superfamília PcbC. É importante ressaltar que flavonoides já foram descritos como importantes reguladores do tamanho de sementes em *Arabidopsis* (DOUGHTY; ALJABRI; SCOTT, 2004).

A característica de Moca Graúdo apresentou 50 SNPs associados, sendo que 28 marcadores encontram-se no cromossomo 2, 15 marcadores no cromossomo 6 e os 7 restantes, nos cromossomos 0 (5 SNPs) e 11 (2 SNPs) (FIGURA 30B). Dos 50 SNPs que apresentaram associação significativa com a característica de Moca Graúdo, 16 SNPs se localizam em regiões intergênicas, podendo estar na região promotora ou terminadora de genes (TABELA 10). Os outros 34 SNPs se encontram localizados em regiões gênicas. Em alguns casos, mais de um SNP associado estava localizado no mesmo gene. Como exemplo, podem-se citar os marcadores AX-168300389 e AX-168343107, que estão ambos, localizados no gene Cc02\_g34530, que codifica uma proteína “*Topless-related protein 3-like isoform XI*”. Essas proteínas, pertencentes à família de fatores de transcrição (TPL ou TPRs) são co-supressores (reguladores negativos) que interagem com proteínas adaptadoras do tipo NINJA, os quais por sua vez, interagem diretamente com proteínas do grupo 2 da família proteica TIFY, como o BS1, proteína diretamente envolvida na regulação do tamanho do grão, em soja (GE et

al., 2016). Outros genes com mais de um SNP associado para a característica Moca Graúdo foram: Cc02\_g34510 (AX-168350209 e AX-168349364) que codifica uma “*Phospholipase D zeta 1-like isoform XI*”, Cc02\_g33440 (AX-168354236 e AX-168307761), que codifica em proteína não caracterizada (“*Uncharacterized protein LOC113727614 isoform XI*”), Cc06\_g05430 (AX-168302790 e AX-168301070) que codifica uma proteína transportadora do tipo ABC (“*ABC transporter G family member 39-like isoform X2*”). Proteínas do tipo ABC já foram caracterizadas como transportadores de flavonoides (HWANG et al, 2016), metabólitos importantes na regulação do tamanho de sementes em *Arabidopsis*, conforme já mencionado acima (DOUGHTY; ALJABRI; SCOTT, 2004). Por fim, a outra proteína com mais de um marcador associado, Cc06\_g05330 (AX-168357947 e AX-168305381), codifica um outro fator de transcrição (“*Trihelix transcription factor PTL-like*”). Membros dessa família de fatores de transcrição (*Trihelix*), também já foram funcionalmente caracterizados como importantes proteínas reguladoras da maturação e desenvolvimento de sementes em *Arabidopsis* (GAO et al, 2009).

Marcadores SNPs associados significativamente com a característica de tamanho e formato de grãos de café (Moca Graúdo, neste caso), localizados em regiões gênicas como o AX-168308075, que está posicionado no gene Cc02\_g33260, que codifica uma proteína “*cyclin-H1-1-like*”, o AX-168303111, localizado no gene Cc02\_g34000, que codifica um fator de transcrição “*transcription factor bHLH18-like*”, assim como o marcador AX-168345105, posicionado no gene Cc06\_g05340 que codifica uma proteína “*ethylene-responsive transcription factor ERF024-like*”, merecem destaque.

Em um trabalho recente, Ren et al (2018), em estudo similar de GWAS em *Arabidopsis*, identificou associação de CYCB1-4, um gene que codifica uma proteína “*cyclin*” envolvida no ciclo celular, com o tamanho de sementes. Plantas transgênicas com aumento de expressão de CYCB1-4, apresentaram sementes maiores devido a uma progressão acelerada do ciclo celular, tendendo a apresentar maiores produções (REN et al, 2018). Função similar pode ser postulada para o gene Cc02\_g33260, que também codifica uma proteína “*cyclin*”, encontrada neste estudo associada com a característica de Moca Graúdo.

Em outro trabalho, a função de um membro da família de fatores de transcrição bHLH, também foi relacionada com o envolvimento dessas proteínas no desenvolvimento do carpelo e do fruto em *Arabidopsis* (GROSZMANN et al, 2008), corroborando com o possível envolvimento do gene Cc02\_g34000, detectado neste

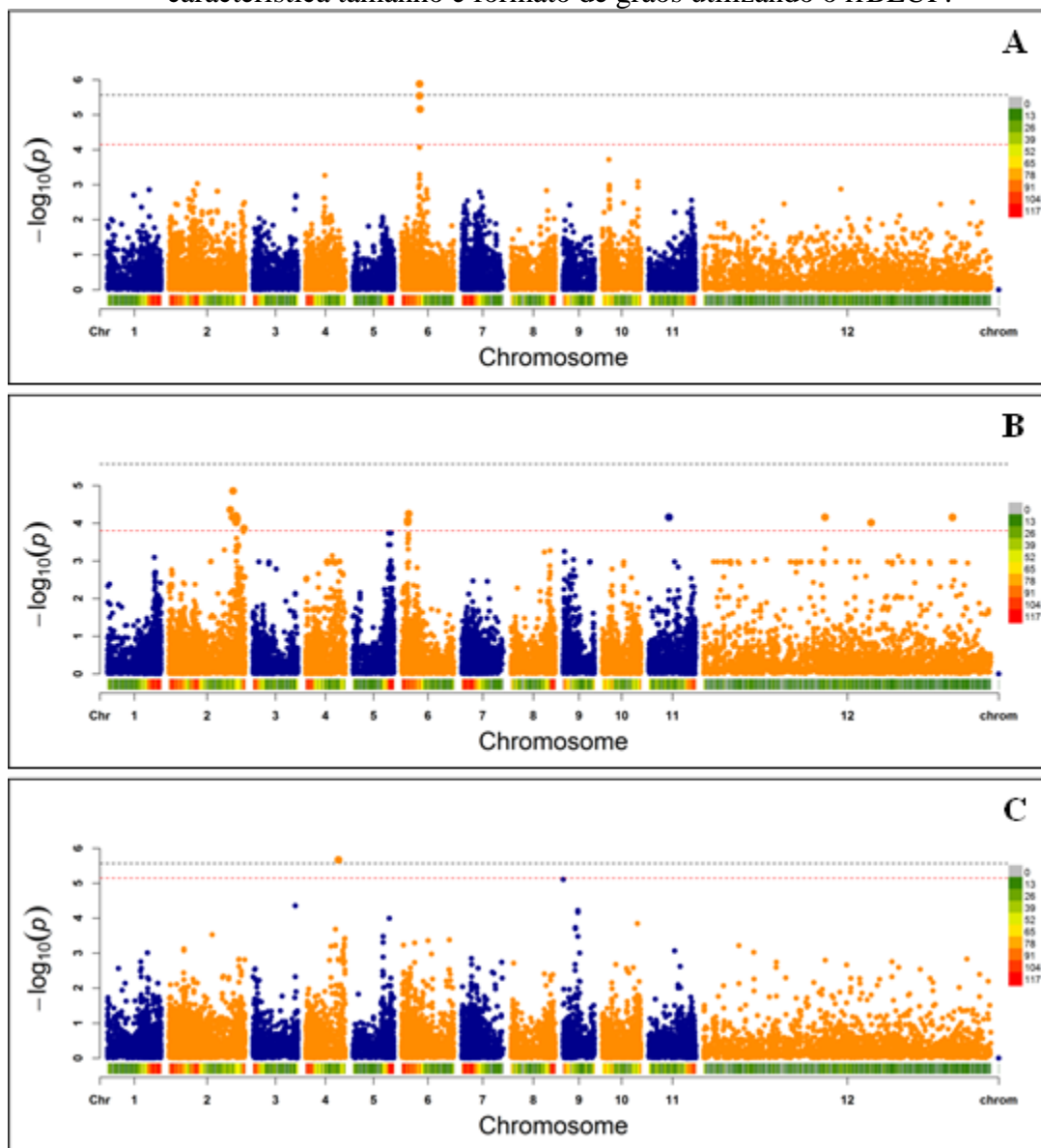
estudo como importante na determinação do tamanho e formato dos grãos de café. Por fim, vale destacar também a detecção de um marcador associado com a característica de Moca Graúdo, localizado no gene Cc06\_g05340, que conforme o resultado de Blast, o classificou como codificando uma proteína “*ethylene-responsive transcription factor ERF024-like*”. Entretanto, vale a pena mencionar que esse gene, codifica uma proteína que contém o domínio AP2 (APETALA), que pertence a uma proteína de *Arabidopsis* diretamente envolvida no desenvolvimento do embrião e endosperma, determinando o tamanho das sementes nesta espécie vegetal (OHTO et al, 2009). Interessante notar, que mutantes de APETALA em *Arabidopsis*, produzem sementes maiores que as plantas controle, condizente com um possível envolvimento do gene Cc06\_g05340, identificado nesse estudo com o fenótipo de grãos maiores, do tipo Moca Graúdo.

Como pode-se observar, vários dos marcadores SNPs encontrados em associação com as características de tamanho e formato dos grãos de café, neste estudo, estão localizados em regiões gênicas que codificam proteínas já caracterizadas em outras espécies vegetais, como importantes reguladores ou proteínas envolvidas no desenvolvimento ou tamanho de sementes. Na Tabela 10 estão apresentados todos os marcadores que apresentaram associação significativa com as características em estudo, assim como sua localização.

O marcador AX-168299085, localizado no cromossomo 4, mostrou-se associado a característica Moca Miúdo (FIGURA 30C). Para as demais características analisadas, não foi possível identificar marcadores associados. Os gráficos QQ plot para todos os formatos (chato ou moca) e tamanhos de grãos (graúdo, médio e miúdo) estão apresentados na Figura 31.

Considerando a correção de Bonferroni, nota-se que o número de SNPs associados para as características apresentadas na Figura 30 reduzem drasticamente, especialmente para Moca Graúdo, em que nenhum marcador apresentou associação.

Figura 30 – Gráfico Manhattan de associação do efeito dos SNPs testados para a característica tamanho e formato de grãos utilizando o rrBLUP.



Legenda: *Manhattan plot* para as características de tamanho e formato do grão. SNPs associados para os grãos classificados como Chato Médio (A), Moca Graúdo (B) e Moca Miúdo (C). As linhas tracejadas representam o valor obtido para  $-\log_{10}(p - \text{valor})$  considerando a correção de Bonferroni (linha tracejada preta) e FDR (linha tracejada vermelha). O cromossomo 0 de *C. canephora* está representado como cromossomo 12.

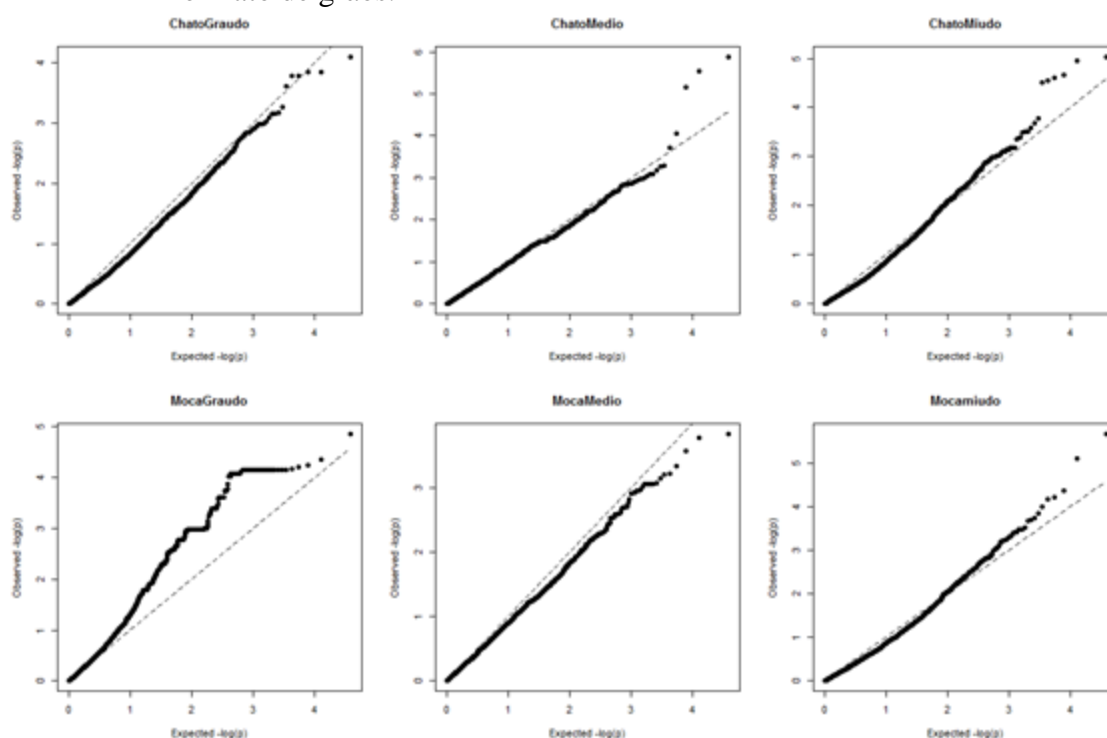


Tabela 9 – Marcadores associados para as características de tamanho e formato de grãos na análise com o rrBLUP.

Marcador	Chr	Posição	Fenótipo	Gene	Proteína	Log10 (pvalor)
AX-168298103	chr6	12791230	chM	Cc06_g14920	flavonol synthase/flavanone 3-hydroxylase-like	5,54
AX-168307159	chr6	12838501	chM	Cc06_g14970	leucine aminopeptidase 1-like	5,88
AX-168345461	chr6	13075988	chM	intergênica		5,16
AX-168351582	chr0	86833249	mcG	Cc00_g10480	autophagy-related protein 18a-like isoform X1	4,15
AX-168358000	chr0	86842748	mcG	Cc00_g10490	protein SRC2 homolog	4,15
AX-168350302	chr0	119869013	mcG	intergênica		4,01
AX-168358401	chr0	178020530	mcG	Cc00_g28760	probable pectinesterase/pectinesterase inhibitor 58	4,15
AX-168349493	chr0	178021959	mcG	intergênica		4,15
AX-168307233	chr2	43845078	mcG	Cc02_g32570	piezo-type mechanosensitive ion channel homolog	4,35
AX-168345527	chr2	45306957	mcG	intergênica		4,15
AX-168308075	chr2	45584774	mcG	Cc02_g33260	cyclin-H1-1-like	4,15
AX-168307761	chr2	45870306	mcG	<b>Cc02_g33440</b>	<b>uncharacterized protein LOC113727614 isoform X1</b>	4,85
AX-168354236	chr2	45873076	mcG	intergênica		4,15
AX-168344465	chr2	45965255	mcG	intergênica		4,15
AX-168350938	chr2	46680708	mcG	Cc02_g33780	probable linoleate 9S-lipoxygenase 5	4,15
AX-168306591	chr2	46700921	mcG	Cc02_g33800	probable linoleate 9S-lipoxygenase 5	4,15
AX-168303111	chr2	47249367	mcG	Cc02_g34000	transcription factor bHLH18-like	4,15
AX-168359111	chr2	47305915	mcG	Cc02_g34030	stress-related protein-like	4,15
AX-168352813	chr2	47499197	mcG	Cc02_g34140	uncharacterized protein LOC113727677 /Taffazin	4,15
AX-168314271	chr2	47659533	mcG	intergênica		4,15
AX-168350690	chr2	47691038	mcG	Cc02_g34320	serine/threonine-protein phosphatase PP1 isozyme 9	4,15
AX-168345203	chr2	47801512	mcG	intergênica		4,15
AX-168349364	chr2	47922088	mcG	<b>Cc02_g34510</b>	<b>phospholipase D zeta 1-like isoform X1</b>	4,02
AX-168350209	chr2	47923944	mcG	intergênica		4,19
AX-168343107	chr2	47984747	mcG	<b>Cc02_g34530</b>	<b>topless-related protein 3-like isoform X1</b>	4,15
AX-168300389	chr2	47987950	mcG	intergênica		4,15
AX-168344012	chr2	48014671	mcG	intergênica		4,16
AX-168350907	chr2	48024076	mcG	intergênica		4,07
AX-168356150	chr2	48051681	mcG	Cc02_g34560	probable protein phosphatase 2C 5	4,13
AX-168352613	chr2	48102576	mcG	Cc02_g34580	uncharacterized protein LOC113727756 isoform X1	4,02
AX-168299845	chr2	48202193	mcG	Cc02_g34690	calmodulin-binding protein 60 D-like	4,15
AX-168304747	chr2	48319349	mcG	intergênica		4,15
AX-168351290	chr2	48391519	mcG	Cc02_g34840	pentatricopeptide repeat-containing protein At1g09190-like	4,14
AX-168353968	chr2	48527652	mcG	Cc02_g35000	lon protease homolog 1, mitochondrial-like	4,15
AX-168306176	chr2	48714647	mcG	intergênica		4,10
AX-168305094	chr2	53747500	mcG	intergênica		3,86
AX-168300636	chr6	4165722	mcG	Cc06_g05270	polyadenylation and cleavage factor homolog 4-like isoform X1	4,07
AX-168305381	chr6	4213885	mcG	<b>Cc06_g05330</b>	<b>trihelix transcription factor PTL-like</b>	4,07
AX-168357947	chr6	4214603	mcG	intergênica		4,07
AX-168303541	chr6	4224625	mcG	intergênica		4,07
AX-168344569	chr6	4232666	mcG	intergênica		4,07
AX-168345105	chr6	4245430	mcG	Cc06_g05340	ethylene-responsive transcription factor ERF024-like	4,07
AX-168300402	chr6	4246582	mcG	intergênica		4,07
AX-168300079	chr6	4247807	mcG	intergênica		4,07
AX-168357236	chr6	4256203	mcG	Cc06_g05350	UPF0481 protein At3g47200-like	4,07
AX-168358192	chr6	4258781	mcG	Cc02_g05380	tyrosine N-monoxygenase-like	4,07
AX-168357178	chr6	4314349	mcG	Cc06_g05400	1,4-alpha-glucan-branching enzyme 2-2, chloroplastic/amyloplastic-like isoform X1	4,07
AX-168301070	chr6	4334744	mcG	intergênica		4,02
AX-168302790	chr6	4338286	mcG	<b>Cc06_g05430</b>	<b>ABC transporter G family member 39-like isoform X2</b>	4,07
AX-168342602	chr6	4375659	mcG	Cc06_g05480	serine/threonine-protein kinase D6PKL2-like	4,07
AX-168303120	chr6	4922381	mcG	Cc06_g06250	uncharacterized protein LOC113691887	4,25
AX-168357486	chr11	14383294	mcG	Cc11_g03570	glutamate receptor 1.3-like	4,15
AX-168305338	chr11	14385622	mcG	intergênica		4,15
AX-168299085	chr4	23325422	mcP	intergênica		5,66

Legenda: Identificação dos marcadores associados, utilizando o rrBLUP, para as características de tamanho e formato de grãos (chM – chato médio, mcG – moca graúdo e mcP – moca miúdo), indicando o cromossomo (Chr), posição, gene, proteína e log10 (pvalor).

Figura 31 – Gráfico quantil-quantil (Q-Q) dos resultados de GWAS para tamanho e formato de grãos.



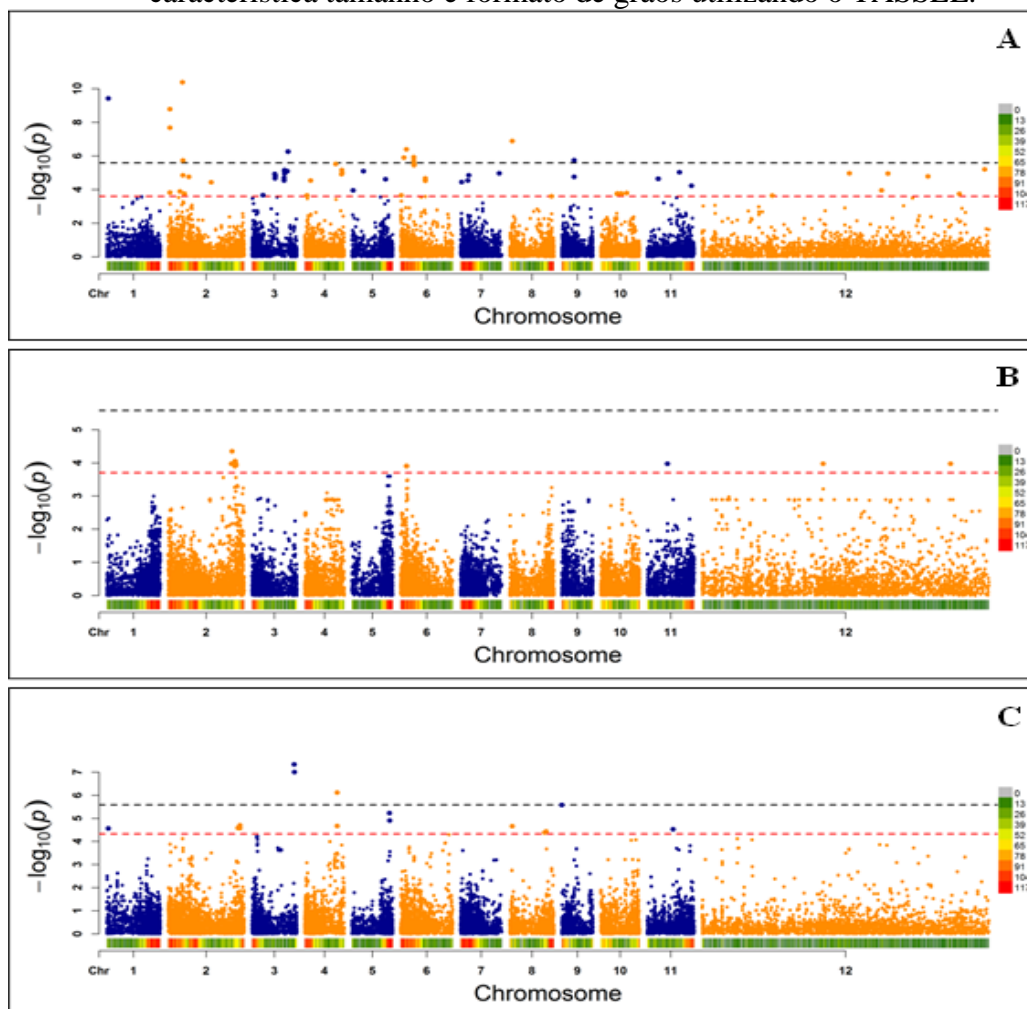
Legenda: A linha central na diagonal do gráfico representa a distribuição esperada e os pontos representam a distribuição observada.

Fonte: Do autor (2019).

### 3.1.2 Dados genotípicos do chip 26K Axiom de *C. canephora* – TASSEL

As análises realizadas no software TASSEL, para o mesmo conjunto de dados, apresentou um resultado diferente comparado ao rrBLUP. Somente com o TASSEL pode-se verificar associação para a característica Chato Miúdo, como se observa na Figura 32A, detectando um total de 82 marcadores associados. Para Moca Graúdo e Moca Miúdo, 47 e 17 SNPs estavam associados, respectivamente (FIGURA 32B e C), totalizando 146 marcadores com associação significativa para as características em estudo. A análise não identificou SNPs relacionados a Chato Médio, como obtido na análise realizada com o pacote rrBLUP.

Figura 32 – Gráfico Manhattan de associação do efeito dos SNPs testados para a característica tamanho e formato de grãos utilizando o TASSEL.



Legenda: *Manhattan plot* para as características de tamanho e formato do grão. SNPs associados para os grãos classificados como Chato Miúdo (A), Moca Graúdo (B) e Moca Miúdo (C). As linhas tracejadas representam o valor obtido para  $-\log_{10}(p - \text{valor})$  considerando a correção de Bonferroni (linha tracejada preta) e FDR (linha tracejada vermelha). O cromossomo 0 de *C. canephora* está representado como cromossomo 12.

Comparando-se as duas metodologias utilizadas para os 366 indivíduos de *C. canephora* e o conjunto de dados genotípico e fenotípico sendo o mesmo, notam-se diferenças entre as duas análises (rrBLUP vs. TASSEL). Inicialmente, em termos de quantidade de marcadores associados, o rrBLUP identificou 57 SNPs, enquanto que com o TASSEL esse número foi bem mais elevado, com um total de 146 SNPs. Outra diferença é em relação às características que apresentaram associação, verifica-se que somente o rrBLUP detectou três SNPs relacionados com Chato Médio, em contrapartida, somente com a análise utilizando-se o TASSEL, a característica Chato Miúdo apresentou associação, com 82 marcadores identificados na GWAS. Para Moca Graúdo, 50 SNPs foram detectados com rrBLUP e 47 com o TASSEL, sendo que 46 desses SNPs foram coincidentes nas duas análises. Para Moca Miúdo, o único SNP associado na GWAS realizada no rrBLUP foi também identificado com as análises no TASSEL, que resultou na identificação de 16 marcadores adicionais para essa mesma característica. A Figura 33 ilustra o resultado comparativo das análises de GWAS realizadas utilizando-se os diferentes programas estatísticos (rrBLUP vs TASSEL).

Com relação à localização desses marcadores com associação significativa para a característica Chato Miúdo dos grãos de café, especificamente detectados com as análises realizadas com o programa TASSEL, pode-se observar, conforme os dados apresentados na Tabela 11 e Figura 32A, que a maior parte dos SNPs associados estão localizados nos cromossomos 3 (19 SNPs), 6 (13 SNPs) e 2 (11 SNPs). Alguns genes tiveram mais de um marcador SNP associado, sendo que o gene Cc03\_g12920, que codifica uma “*serine/threonine-protein kinase-like protein ACR4*” apresentou três SNPs associados. Essa proteína pertence à família de receptores com atividade quinase, que estão envolvidos nas respostas imunes de plantas ao ataque de pragas e patógenos (AFZAL; WOOD; LIGHTFOOT, 2008), mas também na morfogênese e diferenciação de tecidos vegetais, determinando o tamanho e formato de órgãos vegetais (DE SMET et al, 2009). O gene CRINKLY 4 (ACR4) de *Arabidopsis* foi caracterizado funcionalmente como importante regulador da diferenciação de células da epiderme de raízes, restringindo a divisão celular na formação das raízes (DE SMET et al, 2008). Além disto, o envolvimento deste gene na organização das camadas celulares no desenvolvimento do integumento do óvulo, assim como na definição das margens das sépalas, também foi descrito (GIFFORD; DEAN; INGRAM, 2003).

Com dois marcadores SNPs associados, pode-se citar também, os genes Cc02\_g12670 (“*glycine-rich cell wall structural protein-like*”), Cc04\_g16330

(“*thioredoxin-like 1-2, chloroplastic*”), Cc06\_g11920 (“*mechanosensitive ion channel protein 10-like*”), Cc06\_g12050 (“*uncharacterized protein LOC113696977 isoform XI*”), Cc06\_g17650 (“*ABC transporter A family member 1*”) e Cc10\_g10700 (“*uncharacterized protein LOC113714577 isoform XI*”).

A morfogênese de tecidos vegetais depende de uma intensa orquestra de sinais intracelulares. Em plantas, a situação é mais complexa pelo fato das células vegetais estarem envolvidas pela parede celular e não se moverem, uma em relação à outra. Desta forma, alterações no tamanho e formato das células, precisam ser coordenadas entre as células de um tecido vegetal, conjuntamente com a dinâmica da parede celular (VADEPALLI et al, 2011). Desta forma, é importante salientar que a detecção de genes codificando componentes da parede celular, assim como enzimas que remodelam a mesma, associados com as características de tamanho e formato de grãos de café é extremamente justificada e compreensível. Este é o caso, por exemplo do gene Cc02\_g12670, que codifica uma “*glycine-rich cell wall structural protein-like*”, proteína importante na composição da parede celular em plantas (RINGLI; KELLER; RYSER, 2001). Da mesma forma, a associação com o gene Cc02\_g16230, que codifica uma “*beta-galactosidase 13-like*”, também pode ser incluída nesse grupo.

Conforme descrito anteriormente, genes que codificam proteínas de receptores com atividade quinase (“*Receptor-Like Kinases*”), associados com o desenvolvimento e diferenciação celular, como os genes detectados nesse estudo, Cc10\_g10760 (“*receptor kinase-like protein Xa21*”), assim como as proteínas do ciclo celular codificada pelo gene Cc11\_g06580 (“*cyclin-SDS*”), já foram descritos como potenciais candidatos a estarem relacionados ao fenótipo de tamanho e formato de grãos (DE SMET et al, 2009; REN et al, 2018).

O envolvimento de proteínas com similaridade às que são codificadas pelos genes Cc04\_g02150 (“*dof zinc finger protein DOF3.1-like*”), Cc06\_g11810 (“*protein EARLY FLOWERING 5-like*”), Cc06\_g11820 (“*WD repeat-containing protein VIP3-like*”) e Cc07\_g07970 (“*MADS-box transcription factor 23-like isoform X2*”) em processos de desenvolvimento e diferenciação de tecidos vegetais, também já foram descritos na literatura.

Proteínas DOF3 são fatores de transcrição que se ligam especificamente nas sequências consenso de DNA 5'-AAAG-3', também conhecido como bloco “*prolamim*” (YAMAMOTO et al, 2006). Podem ativar a expressão de genes que codificam proteínas de armazenamento como glutelina, prolamina e globulina. Essas proteínas atuam

sinergicamente com RISBZ/BZIP58 para regular positivamente a expressão quantitativa de muitas proteínas de armazenamento das sementes (KAWAKATSU et al, 2009). Atuam também, na regulação positiva de genes chave no desenvolvimento da camada de aleurona em sementes (YAMAMOTO et al, 2006; KAWAKATSU et al, 2009). Função similar pode ser desenvolvida pela proteína Cc04\_g02150, detectada neste estudo como envolvida no tamanho e forma de grãos de café.

Já as proteínas ELF5, similar à proteína codificada pelo gene Cc06\_g11810, foram caracterizadas como importantes reguladoras do tempo de florescimento em *Arabidopsis*, sendo responsáveis pelo florescimento precoce, independente do fotoperíodo (NOH et al, 2004).

O domínio WD (também conhecido como Trp-Asp or “WD40 motif”) é encontrado em uma enorme quantidade de proteínas em eucariotos, envolvidas numa variedade de processos celulares. Na planta modelo *Arabidopsis thaliana*, membros dessa superfamília de proteínas estão sendo reconhecidos como reguladores chave de processos de desenvolvimento de tecidos vegetais (van NOCKER, LUDWIG, 2003). Função similar, no desenvolvimento de grãos de café, pode ser desempenhada pela proteína codificada pelo gene Cc06\_g11820, detectado com associação significativa na determinação do formato e tamanho de grãos de café. De maneira similar, pode-se também destacar o gene Cc07\_g07970 (“*MADS-box transcription factor 23-like isoform X2*”), uma vez que proteínas dessa família (MADS) já foram descritos como importantes reguladores de processos de desenvolvimento vegetal como identidade de meristema, tempo de florescimento, assim como desenvolvimento de fruto e sementes (MASIERO et al, 2011).

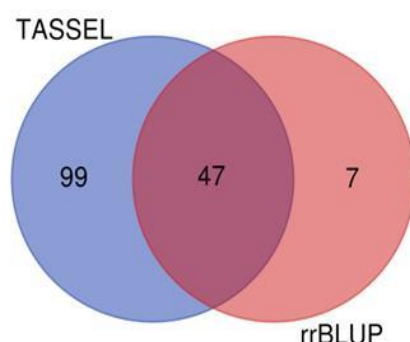
Com relação aos dados de associação, detectados significativamente com a característica Moca Graúdo (TABELA 11), somente 1 marcador adicional aos previamente detectados com o programa rrBLUP, foi encontrado. O marcador AX-168352019, localizado no gene Cc02\_g34790 que codifica uma proteína com função desconhecida (“*uncharacterized protein LOC113727770 isoform XI*”).

Já para Moca Miúdo (TABELA 11), 16 marcadores adicionais foram detectados com o TASSEL, em comparação com as análises realizadas com o rrBLUP (TABELA 10). Observa-se nos dados apresentados que dois marcadores SNPs (AX-168299145 e AX-168350848) foram também encontrados com associação significativa com a característica Chato Miúdo, ambas as características relacionadas ao tamanho dos grãos. O marcador AX-168299145 está posicionado no gene Cc01\_g00640, que codifica uma



proteína “*IQ domain-containing protein IQM2-like*”. Já o marcador AX-168350848, está localizado numa região intergênica do cromossomo 8. Relatos na literatura com relação a membros da família IQM de proteínas foram encontrados, porém somente com dados de expressão, indicando que genes que codificam essas proteínas são expressos em diversos tecidos de *Arabidopsis*, como folhas, caule, flores e siliques (ZHOU et al, 2010).

Figura 33 – Diagrama de Venn comparando as análises de GWAS realizadas no rrBLUP vs. TASSEL.



Análises	Total	Marcadores
TASSEL e rrBLUP (comuns)	47	AX-168345105/mcG AX-168345203/mcG AX-168358401/mcG AX-168308075/mcG AX-168352813/mcG AX-168305338/mcG AX-168359111/mcG AX-168345527/mcG AX-168350690/mcG AX-168357486/mcG AX-168342602/mcG AX-168349364/mcG AX-168306591/mcG AX-168303541/mcG AX-168353968/mcG AX-168300402/mcG AX-168300636/mcG AX-168303111/mcG AX-168349493/mcG AX-168314271/mcG AX-168300389/mcG AX-168357947/mcG AX-168305381/mcG AX-168350938/mcG AX-168352613/mcG AX-168357236/mcG AX-168351290/mcG AX-168350907/mcG AX-168299085/mcP AX-168351582/mcG AX-168357178/mcG AX-168302790/mcG AX-168350209/mcG AX-168299845/mcG AX-168306176/mcG AX-168344465/mcG AX-168304747/mcG AX-168344012/mcG AX-168358192/mcG AX-168358000/mcG AX-168307761/mcG AX-168343107/mcG AX-168300079/mcG AX-168354236/mcG AX-168344569/mcG AX-168356150/mcG AX-168301070/mcG
TASSEL (exclusivos)	99	AX-168343647/chP AX-168309767/mcP AX-168303383/chP AX-168305683/chP AX-168354238/chP AX-168354968/chP AX-168358222/chP AX-168343864/chP AX-168354529/chP AX-168351372/mcP AX-168346334/chP AX-168353870/chP AX-168345166/chP AX-168310468/chP AX-168353697/chP AX-168305324/chP AX-168357997/chP AX-168300621/chP AX-168310228/chP AX-168344168/chP AX-168302268/mcP AX-168344507/chP AX-168356073/chP AX-168307006/chP AX-168372147/chP AX-168311657/chP AX-168350156/chP AX-168359459/chP AX-168355206/chP AX-168346346/chP AX-168298549/chP AX-168345100/chP AX-168311555/chP AX-168356219/mcP AX-168351301/chP AX-168357173/chP AX-168357314/chP AX-168311674/chP AX-168371779/chP AX-168343656/chP AX-168356774/chP AX-168298126/chP AX-168348260/mcP AX-168348505/chP AX-168372309/chP AX-168350848/mcP AX-168354923/chP AX-168350759/mcP AX-168347111/chP AX-168297908/chP AX-168309899/chP AX-168298456/mcP AX-168350848/chP AX-168357775/chP AX-168351351/chP AX-168349013/chP AX-168300945/chP AX-168297140/chP AX-168297476/chP AX-168354492/chP AX-168371724/chP AX-168310426/chP AX-168308959/chP AX-168352736/chP AX-168351369/chP AX-168344856/chP AX-168344460/chP AX-168348359/chP AX-168302930/chP AX-168346035/chP AX-168372190/mcP AX-168299145/mcP AX-168345765/chP AX-168354676/chP AX-168307893/chP AX-168310634/chP AX-168315406/chP AX-168308219/chP AX-168344607/chP AX-168341050/chP AX-168341207/chP AX-168355753/mcP AX-168311865/chP AX-168297274/chP AX-168345653/chP AX-168350139/mcP AX-168297497/chP AX-168351061/mcP AX-168300300/chP AX-168300213/chP AX-168346721/chP AX-168343673/chP AX-168299145/chP AX-168302843/mcP AX-168312699/mcP AX-168349331/chP AX-168352019/mcG AX-168342868/chP AX-168312875/mcP
rrBLUP (exclusivos)	7	AX-168298103/chM AX-168303120/mcG AX-168307233/mcG AX-168305094/mcG AX-168307159/chM AX-168345461/chM AX-168350302/mcG

Legenda: Diagrama de Venn indicando a quantidade de marcadores comuns e distintos nas análises realizadas no rrBLUP e TASSEL, bem como a identificação desses marcadores.

Tabela 10 – Marcadores associados para as características de tamanho e formato de grãos na análise com o TASSEL. (Continua)

Marcador	Chr	Posição	Fenótipo	Gene	Proteína	Log10 (pvalor)
AX-168351301	chr0	50327373	chP	intergênica		3,63
AX-168311657	chr0	105528281	chP	Cc00_g14940	disease resistance protein RPM1-like	4,96
AX-168351351	chr0	128705129	chP	Cc00_g20120	AAA-ATPase ASD, mitochondrial-like	3,95
AX-168308959	chr0	133279405	chP	intergênica		4,95
AX-168345100	chr0	161912726	chP	intergênica		4,78
AX-168372309	chr0	184367809	chP	intergênica		3,74
AX-168353870	chr0	202536015	chP	Cc00_g35180	mannosyl-oligosaccharide 1,2-alpha-mannosidase MNS1-like	5,19
<b>AX-168299145</b>	<b>chr1</b>	<b>1126012</b>	<b>chP</b>	<b>Cc01_g00640</b>	<b>IQ domain-containing protein IQM2-like</b>	<b>9,42</b>
AX-168305324	chr2	1408748	chP	Cc02_g01760	cyclic nucleotide-gated ion channel 1-like	8,78
AX-168371724	chr2	1411894	chP	intergênica		7,68
AX-168371779	chr2	1464001	chP	Cc02_g01810	ATP-dependent zinc metalloprotease FTSH 11, chloroplastic/mitochondrial-like	3,81
AX-168343673	chr2	8528963	chP	Cc02_g10460	cytochrome P450 703A2-like	3,89
AX-168307006	chr2	10336701	chP	Cc02_g12290	pre-mRNA-splicing factor SYF1-like isoform X1	10,37
AX-168346346	chr2	10767477	chP	<b>Cc02_g12670</b>	<b>glycine-rich cell wall structural protein-like</b>	5,72
AX-168347111	chr2	10767519	chP			5,72
AX-168297476	chr2	10787245	chP	Cc02_g12690	uncharacterized protein LOC113730922	4,84
AX-168346721	chr2	11752422	chP	Cc02_g13600	squamosa promoter-binding protein 1	3,75
AX-168359459	chr2	15011257	chP	Cc02_g16230	beta-galactosidase 13-like	4,74
AX-168352736	chr2	31001072	chP	intergênica		4,43
AX-168345166	chr3	7872046	chP	intergênica		3,66
AX-168344168	chr3	16412991	chP	intergênica		4,92
AX-168300945	chr3	16547070	chP	intergênica		4,66
AX-168348359	chr3	16590846	chP	intergênica		4,81
AX-168345653	chr3	22992321	chP	intergênica		4,52
AX-168305683	chr3	23071449	chP	intergênica		4,69
AX-168303383	chr3	23072264	chP	intergênica		4,69
AX-168343864	chr3	23073093	chP			4,69
AX-168344607	chr3	23073508	chP	<b>Cc03_g12920</b>	<b>serine/threonine-protein kinase-like protein ACR4 isoform X1</b>	4,69
AX-168351369	chr3	23074459	chP			4,69
AX-168297140	chr3	23226970	chP	intergênica		5,12
AX-168349013	chr3	23231432	chP	Cc03_g12950	eIF-2-alpha kinase GCN2 isoform X1	5,11
AX-168297497	chr3	23238786	chP	intergênica		5,11
AX-168297908	chr3	23240174	chP	intergênica		5,11
AX-168297274	chr3	23410454	chP	intergênica		5,16
AX-168308219	chr3	23440945	chP	intergênica		4,95
AX-168300300	chr3	25246170	chP	intergênica		5,09
AX-168346334	chr3	25809464	chP	intergênica		6,25
AX-168310426	chr3	25819170	chP	Cc03_g13390	pentatricopeptide repeat-containing protein At1g06710, mitochondrial-like	6,25
AX-168344856	chr4	1676585	chP	Cc04_g02150	dof zinc finger protein DOF3.1-like	3,65
AX-168310634	chr4	1708338	chP	Cc04_g02180	dentin sialophosphoprotein	3,65
AX-168298549	chr4	4380942	chP	Cc04_g05880	uncharacterized protein LOC113738625 isoform X1	4,53
AX-168343656	chr4	22257865	chP	intergênica		5,52
AX-168311865	chr4	26419140	chP			4,90
AX-168346035	chr4	26419235	chP	<b>Cc04_g16330</b>	<b>thioredoxin-like 1-2, chloroplastic</b>	4,90
AX-168300621	chr4	26619424	chP	intergênica		5,15
AX-168354968	chr4	26779757	chP	intergênica		4,97
AX-168341050	chr5	790300	chP	intergênica		3,95
AX-168357314	chr5	8224703	chP	intergênica		5,09
AX-168302930	chr5	24054345	chP	Cc05_g09570	two-pore potassium channel 3-like	4,61
AX-168344507	chr6	550354	chP	Cc06_g00670	short-chain dehydrogenase reductase 2a-like	3,65
AX-168348505	chr6	2538355	chP	Cc06_g03280	uncharacterized protein LOC113694498	5,90
AX-168344460	chr6	4164904	chP	Cc06_g05270	polyadenylation and cleavage factor homolog 4-like isoform X1	6,39
AX-168357173	chr6	9545143	chP	Cc06_g11810	protein EARLY FLOWERING 5-like	5,92
AX-168315406	chr6	9552364	chP	Cc06_g11820	WD repeat-containing protein VIP3-like	5,71
AX-168298126	chr6	9677559	chP	<b>Cc06_g11920</b>	<b>mechanosensitive ion channel protein 10-like</b>	5,52
AX-168354529	chr6	9680854	chP			5,50
AX-168354238	chr6	9760152	chP	<b>Cc06_g12050</b>	<b>uncharacterized protein LOC113696977 isoform X1</b>	5,44
AX-168341207	chr6	9761259	chP			5,51
AX-168356774	chr6	9770901	chP	intergênica		5,52
AX-168350156	chr6	17701432	chP	Cc06_g17600	uncharacterized protein LOC113773371	4,65
AX-168357997	chr6	17773743	chP			4,52
AX-168353697	chr6	17785532	chP	<b>Cc06_g17650</b>	<b>ABC transporter A family member 1</b>	4,65
AX-168343647	chr7	873756	chP	Cc07_g01330	protein COFACTOR ASSEMBLY OF COMPLEX C SUBUNIT B CCB1, chloroplastic	4,43
AX-168307893	chr7	5156376	chP	intergênica		4,53
AX-168300213	chr7	5864059	chP	Cc07_g07970	MADS-box transcription factor 23-like isoform X2	4,84
AX-168354923	chr7	27723856	chP	intergênica		4,96
<b>AX-168350848</b>	<b>chr8</b>	<b>1803097</b>	<b>chP</b>	<b>intergênica</b>		<b>6,88</b>
AX-168357775	chr8	29947471	chP	Cc08_g15090	uncharacterized protein LOC113703701	3,60
AX-168349331	chr9	8852357	chP	intergênica		5,74
AX-168342868	chr9	8880269	chP	intergênica		4,75
AX-168345765	chr10	11726758	chP	intergênica		3,74
AX-168372147	chr10	11748552	chP	intergênica		3,74
AX-168309899	chr10	11777153	chP	intergênica		3,76
AX-168310488	chr10	14573473	chP			3,77
AX-168355206	chr10	14573896	chP	<b>Cc10_g10700</b>	<b>uncharacterized protein LOC113714577 isoform X1</b>	3,74
AX-168310228	chr10	14742543	chP	intergênica		3,74
AX-168356073	chr10	14780965	chP	Cc10_g10760	receptor kinase-like protein Xa21	3,74
AX-168354676	chr10	18454699	chP	Cc10_g11460	uncharacterized protein LOC113714478	3,79
AX-168311555	chr11	7777719	chP	intergênica		4,63
AX-168311674	chr11	7913567	chP	Cc11_g02420	uncharacterized protein LOC113716855 isoform X1	4,63
AX-168354492	chr11	22999877	chP	Cc11_g06580	cyclin-SDS	5,03
AX-168358222	chr11	31714783	chP	intergênica		4,21

Legenda: Identificação dos marcadores associados, utilizando o TASSEL, para as características de tamanho e formato de grãos (chP – chato miúdo, mcG – moca grão e mcP – moca miúdo), indicando o cromossomo (Chr), posição, gene, proteína e log10 (pvalor).



Tabela 11 – Marcadores associados para as características de tamanho e formato de grãos na análise com o TASSEL. (Conclusão)

Marcador	Chr	Posição	Fenótipo	Gene	Proteína	Log10 (pvalor)
AX-168351582	chr0	86833249	mcG	Cc00_g10480	autophagy-related protein 18a-like isoform X1	3,97
AX-168358000	chr0	86842748	mcG	Cc00_g10490	protein SRC2 homolog	3,97
AX-168358401	chr0	178020530	mcG	Cc00_g28760	probable pectinesterase/pectinesterase inhibitor 58	3,97
AX-168349493	chr0	178021959	mcG	intergênica		3,97
AX-168345527	chr2	45306957	mcG	intergênica		3,97
AX-168308075	chr2	45584774	mcG	Cc02_g33260	cyclin-H1-1-like	3,97
AX-168307761	chr2	45870306	mcG	Cc02_g33440	uncharacterized protein LOC113727614 isoform X1	4,34
AX-168354236	chr2	45873076	mcG	intergênica		3,97
AX-168344465	chr2	45965255	mcG	intergênica		3,97
AX-168350938	chr2	46680708	mcG	Cc02_g33780	probable linoleate 9S-lipoxygenase 5	3,97
AX-168306591	chr2	46700921	mcG	Cc02_g33800	probable linoleate 9S-lipoxygenase 5	3,97
AX-168303111	chr2	47249367	mcG	Cc02_g34000	transcription factor bHLH18-like	3,97
AX-168359111	chr2	47305915	mcG	Cc02_g34030	stress-related protein-like	3,97
AX-168352813	chr2	47499197	mcG	Cc02_g34140	uncharacterized protein LOC113727677 /Taffazzin	3,97
AX-168314271	chr2	47659533	mcG	intergênica		3,97
AX-168350690	chr2	47691038	mcG	Cc02_g34320	serine/threonine-protein phosphatase PP1 isozyme 9	3,97
AX-168345203	chr2	47801512	mcG	intergênica		3,97
AX-168349364	chr2	47922088	mcG	Cc02_g34510	phospholipase D zeta 1-like isoform X1	4,05
AX-168350209	chr2	47923944	mcG	intergênica		3,99
AX-168343107	chr2	47984747	mcG	Cc02_g34530	topless-related protein 3-like isoform X1	3,97
AX-168300389	chr2	47987950	mcG	intergênica		3,97
AX-168344012	chr2	48014671	mcG	intergênica		3,97
AX-168350907	chr2	48024076	mcG	intergênica		3,90
AX-168356150	chr2	48051681	mcG	Cc02_g34560	probable protein phosphatase 2C 5	3,98
AX-168352613	chr2	48102576	mcG	Cc02_g34580	uncharacterized protein LOC113727756 isoform X1	4,05
AX-168299845	chr2	48202193	mcG	Cc02_g34690	calmodulin-binding protein 60 D-like	3,97
AX-168304747	chr2	48319349	mcG	intergênica		3,97
AX-168352019	chr2	48356484	mcG	Cc02_g34790	uncharacterized protein LOC113727770 isoform X1	3,95
AX-168351290	chr2	48391519	mcG	Cc02_g34840	pentatricopeptide repeat-containing protein At1g09190-like	3,97
AX-168353968	chr2	48527652	mcG	Cc02_g35000	lon protease homolog 1, mitochondrial-like	3,97
AX-168306176	chr2	48714647	mcG	intergênica		3,93
AX-168300636	chr6	41657222	mcG	Cc06_g05270	polyadenylation and cleavage factor homolog 4-like isoform X1	3,90
AX-168305381	chr6	4213885	mcG	Cc06_g05330	trihelix transcription factor PTL-like	3,90
AX-168357947	chr6	4214603	mcG	intergênica		3,90
AX-168303541	chr6	4224625	mcG	intergênica		3,90
AX-168344569	chr6	4232666	mcG	intergênica		3,90
AX-168345105	chr6	4245430	mcG	Cc06_g05340	ethylene-responsive transcription factor ERF024-like	3,90
AX-168300402	chr6	4246582	mcG	intergênica		3,90
AX-168300079	chr6	4247807	mcG	intergênica		3,90
AX-168357236	chr6	4256203	mcG	Cc06_g05350	UPF0481 protein At3g47200-like	3,90
AX-168358192	chr6	4258781	mcG	Cc02_g05380	tyrosine N-monoxygenase-like	3,90
AX-168357178	chr6	4314349	mcG	Cc06_g05400	1,4-alpha-glucan-branching enzyme 2-2, chloroplastic/amyloplastic-like isoform X1	3,90
AX-168301070	chr6	4334744	mcG	Cc06_g05430	ABC transporter G family member 39-like isoform X2	3,90
AX-168302790	chr6	4338286	mcG	intergênica		3,90
AX-168342602	chr6	4375659	mcG	Cc06_g05480	serine/threonine-protein kinase D6PKL2-like	3,90
AX-168357486	chr11	14383294	mcG	Cc11_g03570	glutamate receptor 1.3-like	3,97
AX-168305338	chr11	14385622	mcG	intergênica		3,97
AX-168299145	chr1	1126012	mcP	Cc01_g00640	IQ domain-containing protein IQM2-like	4,56
AX-168312699	chr2	50146016	mcP	intergênica		4,58
AX-168350759	chr2	51481156	mcP	Cc02_g36670	protein cornichon homolog 4-like	4,57
AX-168302268	chr2	51499977	mcP	Cc02_g36700	ATP-dependent DNA helicase 2 subunit KU80-like isoform X1	4,69
AX-168312875	chr3	30300282	mcP	Cc03_g15070	BTB/POZ and TAZ domain-containing protein 3-like	7,33
AX-168351061	chr3	30317284	mcP	Cc03_g15080	altered inheritance rate of mitochondria protein 25-like isoform X2	7,00
AX-168350139	chr3	30317666	mcP	intergênica		7,00
AX-168298456	chr3	30319343	mcP	intergênica		7,00
AX-168299085	chr4	23325422	mcP	intergênica		6,11
AX-168348260	chr4	23388423	mcP	intergênica		4,67
AX-168356219	chr5	26887049	mcP	Cc05_g13220	uncharacterized protein LOC113690300	5,23
AX-168309767	chr5	27043049	mcP	intergênica		4,90
AX-168350848	chr8	1803097	mcP	intergênica		4,66
AX-168372190	chr8	24994732	mcP	Cc08_g1013	uncharacterized protein LOC113706144	4,39
AX-168302843	chr8	26017160	mcP	Cc08_g11270	beta-glucosidase 24-like	4,43
AX-168355753	chr9	83842	mcP	Cc09_g00110	E3 ubiquitin-protein ligase WAV3-like isoform X1	5,57
AX-168351372	chr11	18451769	mcP	intergênica		4,52

Legenda: Identificação dos marcadores associados, utilizando o TASSEL, para as características de tamanho e formato de grãos (chP – chato miúdo, mcG – moca graúdo e mcP – moca miúdo), indicando o cromossomo (Chr), posição, gene, proteína e log10 (pvalor).

### 3.1.3 Dados genotípicos do chip 26K Axiom de *C. canephora* e nextRAD – rrBLUP

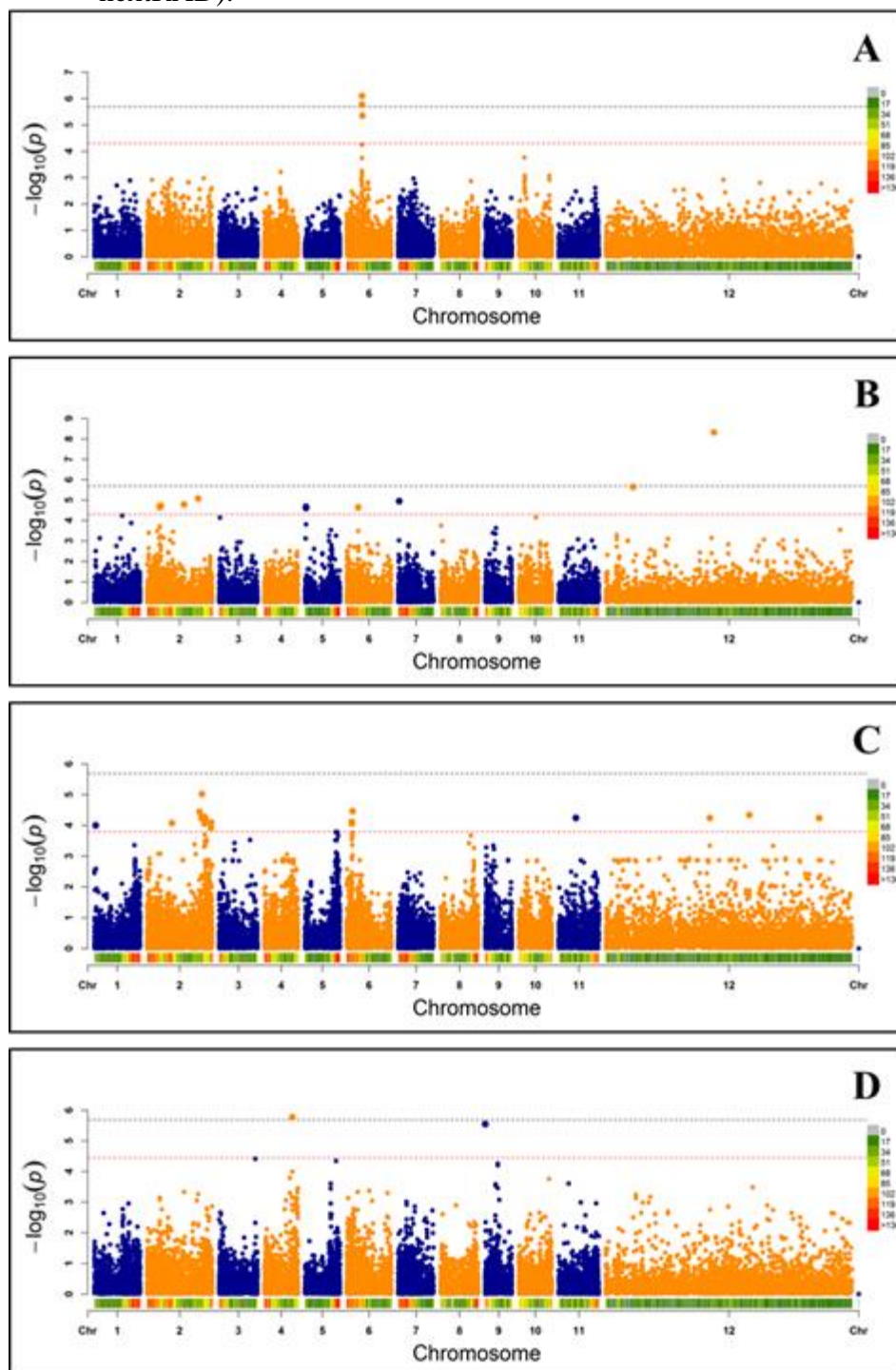
O acréscimo de mais de 5.000 SNPs, identificados pela técnica de genotipagem nextRAD, aos 18.942 do *chip* 26K Axiom, teve o intuito de verificar se o poder na análise de associação entre marcador e fenótipo de interesse seria mais elevado.

Nota-se nos gráficos Manhattan e QQ Plot, respectivamente Figuras 34B e 35, que a característica Chato Miúdo apresentou alguns marcadores associados e, na análise anterior, somente com os marcadores identificados no *chip*, nada havia sido detectado. Para essa característica (TABELA 12), 13 SNPs mostraram-se associados, 7 destes provenientes da técnica de genotipagem nextRAD e os demais identificados no *chip* de *C. canephora*. Pode-se observar, analisando-se os dados apresentados na Tabela 12, que todos os sete marcadores em associação significativa com Chato Miúdo, provenientes de nextRAD, se localizaram em regiões intergênicas. Por outro lado, todos SNPs da Affymetrix identificados com associação significativa para Chato Miúdo, na presente análise, já haviam sido detectados na análise com o TASSEL, sem a adição dos SNPs de genotipagem nextRAD (TABELA 11). Comparando-se com os resultados obtidos para Moca Graúdo (FIGURA 34C, TABELA 12), quatro SNPs adicionais foram identificados como associados para a característica, um marcador presente no *chip* de DNA (AX-168343237), cromossomo 2, e outros três identificados pela técnica de genotipagem nextRAD, localizados nos cromossomos 1 e 2. Destes SNPs adicionais, somente o marcador S2\_20759823, está localizado numa região gênica (Cc02\_g23620-“*cullin-associated NEDD8-dissociated protein 1*”). Em *Arabidopsis*, uma proteína similar, pertencente à família das “*cullin-RING E3 ligases (CRLs)*” interage com a proteína ABERRANT LATERAL ROOT FORMATION 4 (ALF4), sendo que o mutante de ALF4, exibe um fenótipo que sugere deficiência nas respostas hormonal (BAGCHI et al, 2017).

Para a característica Moca Miúdo (FIGURA 34D, TABELA 12), outro marcador presente no *chip* também se mostrou associado, sendo este (AX-168355753), localizado no gene Cc09\_g00110 (“*E3 ubiquitin-protein ligase WAV3-like*”).

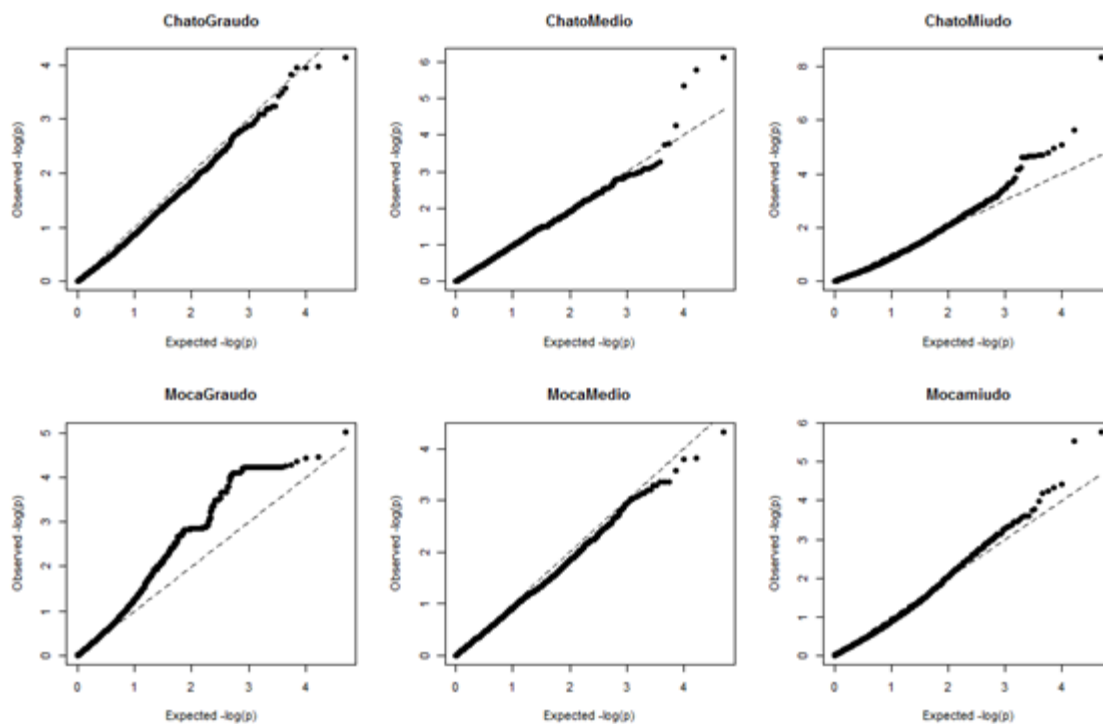
Para Chato Médio, Figura 34A, o resultado obtido nas análises com o diferente conjunto de marcadores não se modificou, sendo que os mesmos três SNPs identificados anteriormente foram também identificados na presente análise.

Figura 34 – Gráfico Manhattan de associação do efeito dos SNPs testados para a característica tamanho e formato de grãos utilizando o rrBLUP utilizando os marcadores de duas técnicas de genotipagem (chip 26K Axiom + nextRAD).



Legenda: *Manhattan plot* para as características de tamanho e formato do grão. SNPs associados para os grãos classificados como Chato Médio (A), Chato Miúdo (B), Moca Graúdo (C) e Moca Miúdo (D). As linhas tracejadas representam o valor obtido para  $-\log_{10}(p)$  considerando a correção de Bonferroni (linha tracejada preta) e FDR (linha tracejada vermelha). O cromossomo 0 de *C. canephora* está representado como cromossomo 12.

Figura 35 – Gráfico quantil-quantil (Q-Q) dos resultados de GWAS para tamanho e formato de grãos utilizando marcadores identificados por duas técnicas de genotipagem (chip 26K Axiom + nextRAD).



Legenda: A linha central na diagonal do gráfico representa a distribuição esperada e os pontos representam a distribuição observada.

Fonte: Do autor (2019).

Tabela 11 – Marcadores associados para as características de tamanho e formato de grãos na análise com o rrBLUP (com P3D) e utilizando os marcadores de duas técnicas de genotipagem (chip 26K Axiom + nextRAD).

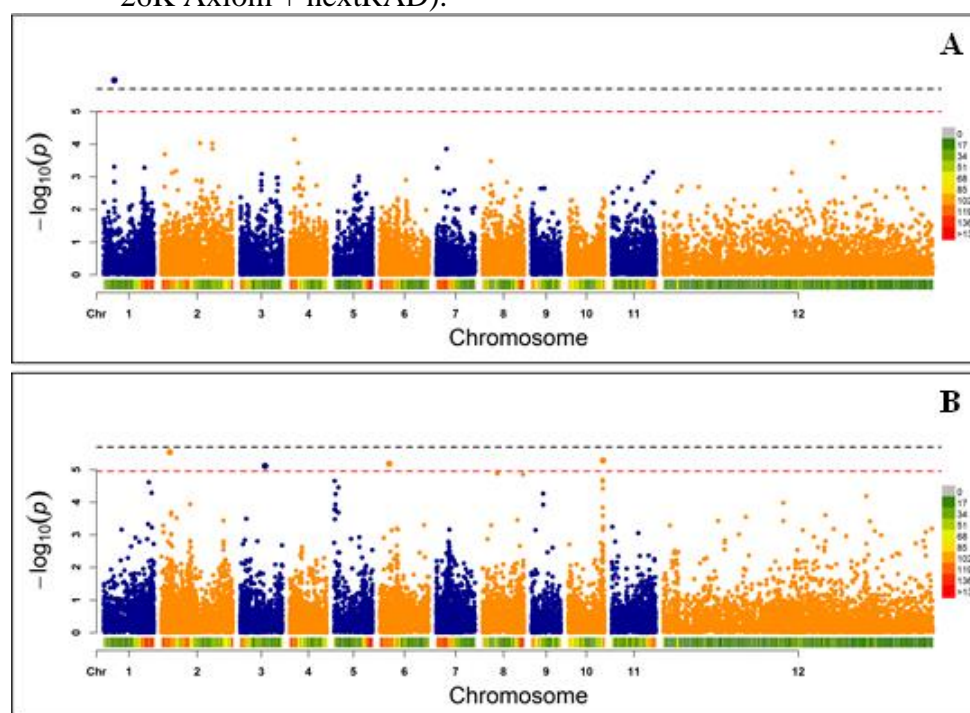
Marcador	Chr	Posição	Fenótipo	Gene	Proteína	Log10 (pvalor)
AX-168298103	chr6	12791230	chM	Cc06_g14920	flavonol synthase/flavanone 3-hydroxylase-like	5,77
AX-168307159	chr6	12838501	chM	Cc06_g14970	leucine aminopeptidase 1-like	6,10
AX-168345461	chr6	13075988	chM	intergênica		5,35
S0_22957200	chr0	22957199	chP	intergênica		5,64
S0_90384806	chr0	90384805	chP	intergênica		8,32
AX-168346346	chr2	10767477	chP	<b>Cc02_g12670</b>	<b>glycine-rich cell wall structural protein-like</b>	4,71
AX-168347111	chr2	10767519	chP			4,66
AX-168346721	chr2	11752422	chP	Cc02_g13600	squamosa promoter-binding protein 1	4,72
AX-168352736	chr2	31001072	chP	intergênica		4,79
S2_42640423	chr2	42640422	chP	intergênica		5,08
S5_962780	chr5	962779	chP	intergênica		4,61
S5_962789	chr5	962788	chP	intergênica		4,61
S5_962794	chr5	962793	chP	intergênica		4,68
S5_962799	chr5	962798	chP	intergênica		4,68
AX-168357173	chr6	9545143	chP	Cc06_g11810	protein EARLY FLOWERING 5-like	4,64
AX-168343647	chr7	873756	chP	Cc07_g01330	protein COFACTOR ASSEMBLY OF COMPLEX C SUBUNIT B CCB1, chloroplastic	4,95
AX-168351582	chr0	86833249	mcG	Cc00_g10480	autophagy-related protein 18a-like isoform X1	4,24
AX-168358000	chr0	86842748	mcG	Cc00_g10490	protein SRC2 homolog	4,24
AX-168350302	chr0	119869013	mcG	intergênica		4,34
AX-168358401	chr0	178020530	mcG	Cc00_g28760	probable pectinesterase/pectinesterase inhibitor 58	4,24
AX-168349493	chr0	178021959	mcG	intergênica		4,24
S1_1107611	chr1	1107610	mcG	intergênica		4,00
S1_1107596	chr1	1107595	mcG	intergênica		4,00
S2_20759823	chr2	20759822	mcG	Cc02_g23620	cullin-associated NEDD8-dissociated protein 1	4,08
AX-168307233	chr2	43845078	mcG	Cc02_g32570	piezo-type mechanosensitive ion channel homolog	4,44
AX-168345527	chr2	45306957	mcG	intergênica		4,24
AX-168308075	chr2	45584774	mcG	Cc02_g33260	cyclin-H1-1-like	4,24
AX-168307761	chr2	45870306	mcG	<b>Cc02_g33440</b>	<b>uncharacterized protein LOC113727614 isoform X1</b>	5,02
AX-168354236	chr2	45873076	mcG			4,24
AX-168344465	chr2	45965255	mcG	intergênica		4,24
AX-168350938	chr2	46680708	mcG	Cc02_g33780	probable linoleate 9S-lipoxygenase 5	4,24
AX-168306591	chr2	46700921	mcG	Cc02_g33800	probable linoleate 9S-lipoxygenase 5	4,24
AX-168303111	chr2	47249367	mcG	Cc02_g34000	transcription factor bHLH18-like	4,24
AX-168359111	chr2	47305915	mcG	Cc02_g34030	stress-related protein-like	4,24
AX-168352813	chr2	47499197	mcG	Cc02_g34140	uncharacterized protein LOC113727677 /Taffazzin	4,24
AX-168314271	chr2	47659533	mcG	intergênica		4,24
AX-168350690	chr2	47691038	mcG	Cc02_g34320	serine/threonine-protein phosphatase PP1 isozyme 9	4,24
AX-168345203	chr2	47801512	mcG	intergênica		4,24
AX-168349364	chr2	47922088	mcG	<b>Cc02_g34510</b>	<b>phospholipase D zeta 1-like isoform X1</b>	4,06
AX-168350209	chr2	47923944	mcG			4,28
AX-168343107	chr2	47984747	mcG	<b>Cc02_g34530</b>	<b>topless-related protein 3-like isoform X1</b>	4,24
AX-168300389	chr2	47987950	mcG			4,24
AX-168344012	chr2	48014671	mcG	intergênica		4,25
AX-168350907	chr2	48024076	mcG	intergênica		4,15
AX-168356150	chr2	48051681	mcG	Cc02_g34560	probable protein phosphatase 2C 5	4,21
AX-168352613	chr2	48102576	mcG	Cc02_g34580	uncharacterized protein LOC113727756 isoform X1	4,06
AX-168299845	chr2	48202193	mcG	Cc02_g34690	calmodulin-binding protein 60 D-like	4,24
AX-168304747	chr2	48319349	mcG	intergênica		4,24
AX-168351290	chr2	48391519	mcG	Cc02_g34840	pentatricopeptide repeat-containing protein At1g09190-like	4,22
AX-168353968	chr2	48527652	mcG	Cc02_g35000	lon protease homolog 1, mitochondrial-like	4,24
AX-168306176	chr2	48714647	mcG	intergênica		4,20
AX-168343237	chr2	53259323	mcG	intergênica		3,93
AX-168305094	chr2	53747500	mcG	intergênica		4,10
AX-168300636	chr6	4165722	mcG	Cc06_g05270	polyadenylation and cleavage factor homolog 4-like isoform X1	4,11
AX-168305381	chr6	4213885	mcG	<b>Cc06_g05330</b>	<b>trihelix transcription factor PTL-like</b>	4,11
AX-168357947	chr6	4214603	mcG			4,11
AX-168303541	chr6	4224625	mcG	intergênica		4,11
AX-168344569	chr6	4232666	mcG	intergênica		4,11
AX-168345105	chr6	4245430	mcG	Cc06_g05340	ethylene-responsive transcription factor ERF024-like	4,11
AX-168300402	chr6	4246582	mcG	intergênica		4,11
AX-168300079	chr6	4247807	mcG	intergênica		4,11
AX-168357236	chr6	4256203	mcG	Cc06_g05350	UPF0481 protein At3g47200-like	4,11
AX-168358192	chr6	4258781	mcG	Cc02_g05380	tyrosine N-monooxygenase-like	4,11
AX-168357178	chr6	4314349	mcG	Cc06_g05400	1,4-alpha-glucan-branching enzyme 2-2, chloroplastic/amyloplastic-like isoform X1	4,12
AX-168301070	chr6	4334744	mcG	<b>Cc06_g05430</b>	<b>ABC transporter G family member 39-like isoform X2</b>	4,06
AX-168302790	chr6	4338286	mcG			4,11
AX-168342602	chr6	4375659	mcG	Cc06_g05480	serine/threonine-protein kinase D6PKL2-like	4,11
AX-168303120	chr6	4922381	mcG	Cc06_g06250	uncharacterized protein LOC113691887	4,46
AX-168357486	chr11	14383294	mcG	Cc11_g03570	glutamate receptor 1.3-like	4,24
AX-168305338	chr11	14385622	mcG	intergênica		4,24
AX-168299085	chr4	23325422	mcP	intergênica		5,77
AX-168355753	chr9	83842	mcP	Cc09_g00110	E3 ubiquitin-protein ligase WAV3-like isoform X1	5,55

Legenda: Identificação dos marcadores associados, utilizando o rrBLUP, para as características de tamanho e formato de grãos (chM – chato médio, chP – chato miúdo, mcG – moca gráudo e mcP – moca miúdo), indicando o cromossomo (Chr), posição, gene, proteína e log10 (pvalor).



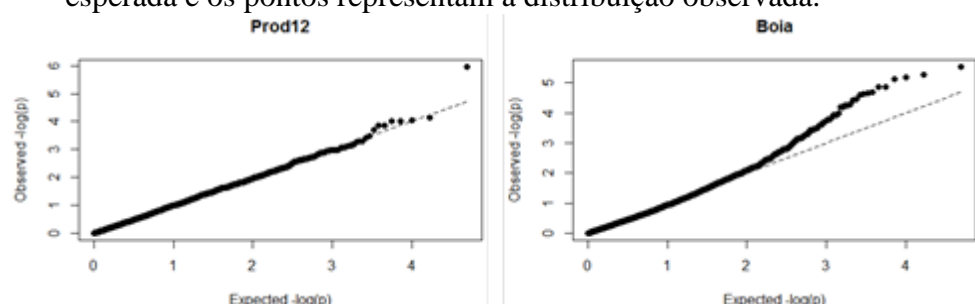
Na tentativa de aumentar ainda mais o poder de análise de GWAS e identificar outros marcadores ou até mesmo outras características que apresentassem associação, uma análise com os mesmos dados (366 indivíduos e marcadores do *chip* + nextRAD) foi realizada no rrBLUP, entretanto, modificou-se a opção para rodar sem P3D. Nestas condições, um SNP associado à produção foi encontrado no cromossomo 1 (S1\_8001113), e outros quatro SNPs (S2\_6272091, S3\_18778694, AX-168299277 e S6\_7033632) associados a variável fenotípica Boia, como apresentado nas Figuras 36 e 37 e Tabela 13. Conforme mostram os dados da TABELA 13, o marcador S1\_8001113 que apresentou associação com a produção no ano de 2012, está localizado em uma região intergênica. Já no caso da característica Boia, 3 dos quatro SNPs em associação significativa estão localizados em regiões gênicas. O marcador S3\_18778694, localizado no gene *Cc03\_g12090* (“*flowering time control protein FPA-like*”), AX-168299277 posicionado no gene *Cc06\_g08760* (“*zinc finger CCCH domain-containing protein 32-like*”) e o S6\_7033632 (“*dihydrofolate reductase isoform XI*”).

Figura 36 – Gráfico Manhattan de associação do efeito dos SNPs testados para as características de produção, avaliada no ano de 2012, e boia, utilizando o rrBLUP utilizando os marcadores de duas técnicas de genotipagem (*chip* 26K Axiom + nextRAD).



Legenda: *Manhattan plot* com os SNPs associados para as características de produção 2012 (A) e boia (B). As linhas tracejadas representam o valor obtido para  $-\log_{10}(p - \text{valor})$  considerando a correção de Bonferroni (linha tracejada preta) e FDR (linha tracejada vermelha). O cromossomo 0 de *C. canephora* está representado como cromossomo 12.

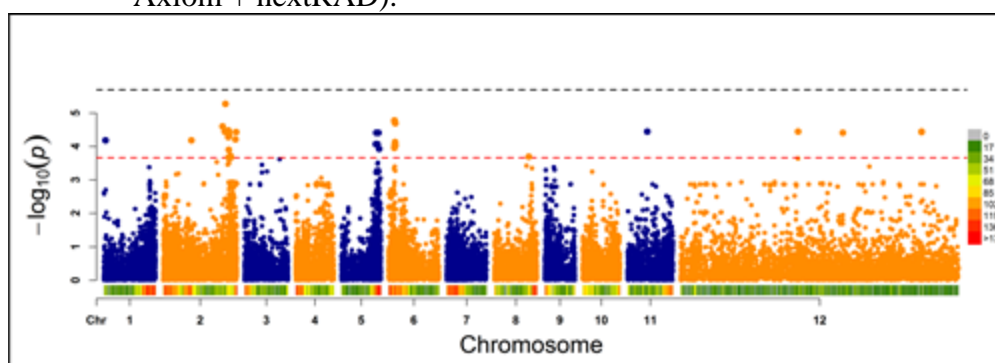
Figura 37 – Gráfico quantil-quantil (Q-Q) dos resultados de GWAS, sem P3D, para produção e boia associando duas técnicas de genotipagem (chip 26K Axiom + nextRAD). A linha central na diagonal do gráfico representa a distribuição esperada e os pontos representam a distribuição observada.



Para as características de grãos Chato Médio e Chato Miúdo, nenhuma diferença nos resultados para essas características entre as análises utilizando-se o rrBLUP com ou sem P3D, foi identificada. Ou seja, os mesmos SNPs foram identificados em associação nas duas análises.

Já para a característica Moca Graúdo, além de todos os 54 SNPs detectados na análise anterior (TABELA 12), 41 marcadores SNPs adicionais foram encontrados em associação com essa característica (FIGURA 38). Esses SNPs estão destacados e apresentados na Tabela 13. Já para a característica Moca Miúdo, os mesmos dois marcadores encontrados anteriormente, foram confirmados.

Figura 38 – Gráfico Manhattan de associação do efeito dos SNPs testados para a característica de tamanho e formato de grãos, utilizando o rrBLUP utilizando os marcadores de duas técnicas de genotipagem (chip 26K Axiom + nextRAD).



Legenda: *Manhattan plot* com os SNPs associados para a característica de tamanho e formato dos grãos, peneira Moca Graúdo. As linhas tracejadas representam o valor obtido para  $-\log_{10}(p - \text{valor})$  considerando a correção de Bonferroni (linha tracejada preta) e FDR (linha tracejada vermelha). O cromossomo 0 de *C. canephora* está representado como cromossomo 12.

Ao modificar a opção de análise para sem o P3D, os componentes de variância são estimados para cada marcador separadamente, resultando em um aumento do poder da análise e consequentemente o aumento do tempo da mesma. Quando o P3D é

utilizado, os componentes de variância são estimados apenas uma vez, reduzindo o tempo computacional e o poder da análise de GWAS.

Nota-se que os resultados obtidos corroboram com a questão do aumento de poder da análise ao estimar a variância para cada marcador separadamente. Além de capturar um número maior de marcadores associados, ao modificar o tipo da análise, novas variáveis fenotípicas foram identificadas na GWAS.

Tabela 12 – Marcadores associados para as características de produção, boia e tamanho e formato de grãos, na análise com o rrBLUP (sem P3D) e utilizando os marcadores de duas técnicas de genotipagem (chip 26K Axiom + nextRAD). (Continua)

Marcador	Chr	Posição	Fenótipo	Gene	Proteína	Log10 (pvalor)
S1_8001113	chr1	8001112	p12	intergênica		5,97
S2_6272091	chr2	6272090	boia	intergênica		5,54
S3_18778694	chr3	18778693	boia	Cc03_g12090	flowering time control protein FPA-like	5,28
S6_7033632	chr6	7033631	boia	Cc06_g08760	zinc finger CCCH domain-containing protein 32-like	5,19
AX-16829927	chr10	26426206	boia	Cc10_g15510	dihydrofolate reductase isoform X1	5,12
AX-16829810	chr6	12791230	chM	Cc06_g14920	flavonol synthase/flavanone 3-hydroxylase-like	5,85
AX-16830715	chr6	12838501	chM	Cc06_g14970	leucine aminopeptidase 1-like	6,25
AX-16834546	chr6	13075988	chM	intergênica		5,52
S0_22957200	chr0	22957199	chP	intergênica		5,71
S0_90384806	chr0	90384805	chP	intergênica		9,66
AX-16834634	chr2	10767477	chP	<b>Cc02_g12670</b>	<b>glycine-rich cell wall structural protein-like</b>	4,78
AX-16834711	chr2	10767519	chP			4,73
AX-16834672	chr2	11752422	chP	Cc02_g13600	squamosa promoter-binding protein 1	4,79
AX-16835273	chr2	31001072	chP	intergênica		5,35
S2_42640423	chr2	42640422	chP	intergênica		5,57
S5_962780	chr5	962779	chP	intergênica		4,81
S5_962789	chr5	962788	chP	intergênica		4,81
S5_962794	chr5	962793	chP	intergênica		4,88
S5_962799	chr5	962798	chP	intergênica		4,88
AX-16835717	chr6	9545143	chP	Cc06_g11810	protein EARLY FLOWERING 5-like	4,72
AX-16834364	chr7	873756	chP	Cc07_g01330	protein COFACTOR ASSEMBLY OF COMPLEX C SUBUNIT B CCB1	5,18

Legenda: Identificação dos marcadores associados, utilizando o rrBLUP (sem P3D), para as características de produção (p12), boia e tamanho e formato de grãos (chM – chato médio, chP – chato miúdo, mcG – moca graúdo e mcP – moca miúdo), indicando o cromossomo (Chr), posição, gene, proteína e log10 (pvalor). Destacados em azul estão os marcadores exclusivos identificados somente nessa análise (sem P3D) para o fenótipo de Moca Graúdo.



Tabela 13 – Marcadores associados para as características de produção, boia e tamanho e formato de grãos, na análise com o rrBLUP (sem P3D) e utilizando os marcadores de duas técnicas de genotipagem (chip 26K Axiom + nextRAD). (Conclusão)

Marcador	Chr	Posição	Fenótipo	Gene	Proteína
AX-168351582	chr0	86833249	mcG	Cc00_g10480	autophagy-related protein 18a-like isoform X1
AX-168358000	chr0	86842748	mcG	Cc00_g10490	protein SRC2 homolog
AX-168350302	chr0	119869013	mcG	intergênica	
AX-168358401	chr0	178020530	mcG	Cc00_g28760	probable pectinesterase/pectinesterase inhibitor 58
AX-168349493	chr0	178021959	mcG	intergênica	
S1_1107595	chr1	1107595	mcG	intergênica	
S1_1107611	chr1	1107610	mcG	intergênica	
S2_20759823	chr2	20759822	mcG	Cc02_g23620	cullin-associated NEDD8-dissociated protein 1
AX-168307233	chr2	43845078	mcG	Cc02_g32570	piezo-type mechanosensitive ion channel homolog
AX-168345527	chr2	45306957	mcG	intergênica	
AX-168308075	chr2	45584774	mcG	Cc02_g33260	cyclin-H1-1-like
AX-168307761	chr2	45870306	mcG		
AX-168354236	chr2	45873076	mcG	<b>Cc02_g33440</b>	<b>uncharacterized protein LOC113727614 isoform X1</b>
AX-168344465	chr2	45965255	mcG	intergênica	
AX-168350938	chr2	46680708	mcG	Cc02_g33780	probable linoleate 9S-lipoxygenase 5
AX-168306591	chr2	46700921	mcG	Cc02_g33800	probable linoleate 9S-lipoxygenase 5
AX-168303111	chr2	47249367	mcG	Cc02_g34000	transcription factor bHLH18-like
AX-168359111	chr2	47305915	mcG	Cc02_g34030	stress-related protein-like
AX-168352813	chr2	47499197	mcG	Cc02_g34140	uncharacterized protein LOC113727677 /Taffazzin
AX-168314271	chr2	47659533	mcG	intergênica	
AX-168350290	chr2	47691038	mcG	Cc02_g34320	serine/threonine-protein phosphatase PP1 isozyme 9
AX-168345203	chr2	47801512	mcG	intergênica	
AX-168349364	chr2	47922088	mcG		
AX-168350209	chr2	47923944	mcG	<b>Cc02_g34510</b>	<b>phospholipase D zeta 1-like isoform X1</b>
AX-168343107	chr2	47984747	mcG		
AX-168300389	chr2	47987950	mcG	<b>Cc02_g34530</b>	<b>topless-related protein 3-like isoform X1</b>
AX-168344012	chr2	48014671	mcG	intergênica	
AX-168350907	chr2	48024076	mcG	intergênica	
AX-168356150	chr2	48051681	mcG	Cc02_g34560	probable protein phosphatase 2C 5
AX-168352613	chr2	48102576	mcG	Cc02_g34580	uncharacterized protein LOC113727756 isoform X1
AX-168299845	chr2	48202193	mcG	Cc02_g34690	calmodulin-binding protein 60 D-like
AX-168342448	chr2	48217852	mcG	intergênica	
AX-168304747	chr2	48319349	mcG	intergênica	
AX-168352019	chr2	48356484	mcG	Cc02_g34790	uncharacterized protein LOC113727770 isoform X1
AX-168351290	chr2	48391519	mcG	Cc02_g34840	pentatricopeptide repeat-containing protein Atg09190-like
AX-168353968	chr2	48527652	mcG	Cc02_g35000	ion protease homolog 1, mitochondrial-like
AX-168351650	chr2	48592726	mcG	Cc02_g35040	nicotianamine synthase-like
AX-168300212	chr2	48609755	mcG	intergênica	
AX-168353529	chr2	48642133	mcG	Cc02_g35070	RNA polymerase II-associated protein 3-like isoform X1
AX-168355301	chr2	48655516	mcG	intergênica	
AX-168306176	chr2	48714647	mcG	intergênica	
AX-168353797	chr2	48728000	mcG	intergênica	
AX-168308758	chr2	48794085	mcG	intergênica	
AX-168352993	chr2	49435102	mcG	intergênica	
AX-168357531	chr2	49495819	mcG	intergênica	
AX-168348341	chr2	49497297	mcG	intergênica	
AX-168357029	chr2	49582983	mcG	intergênica	
AX-168352190	chr2	49635690	mcG	Cc02_g35680	pleiotropic drug resistance protein 3-like
AX-168352761	chr2	49685638	mcG	Cc02_g35700	pleiotropic drug resistance protein 3-like
AX-168359149	chr2	49899023	mcG	intergênica	
AX-168343237	chr2	53259323	mcG	intergênica	
AX-168305094	chr2	53747500	mcG	intergênica	
AX-168308957	chr5	25357499	mcG	intergênica	
AX-168345962	chr5	25599111	mcG	Cc05_g11480	glutamate receptor 2.7-like
AX-168357936	chr5	25664762	mcG	Cc05_g11570	reticulon-4-interacting protein 1, mitochondrial-like
AX-168310386	chr5	25969191	mcG	Cc05_g12010	protein SPA1-RELATED 3 isoform X2
AX-168306839	chr5	26207976	mcG	intergênica	
AX-168346202	chr5	26581667	mcG	intergênica	
AX-168354263	chr5	27091604	mcG	Cc05_g13540	serine/threonine-protein kinase PBL23
AX-168354235	chr5	27259421	mcG	Cc05_g13780	purple acid phosphatase 18
S5_27752781	chr5	27752780	mcG	intergênica	
AX-168300636	chr6	4165722	mcG	Cc06_g05270	polyadenylation and cleavage factor homolog 4-like isoform X1
AX-168305381	chr6	4213885	mcG		
AX-168357947	chr6	4214603	mcG	<b>Cc06_g05330</b>	<b>trihelix transcription factor PTL-like</b>
AX-168300541	chr6	4224625	mcG	intergênica	
AX-168344569	chr6	4232666	mcG	intergênica	
AX-168345105	chr6	4245430	mcG	Cc06_g05340	ethylene-responsive transcription factor ERF024-like
AX-168300402	chr6	4246582	mcG	intergênica	
AX-168300079	chr6	4247807	mcG	intergênica	
AX-168357236	chr6	4256203	mcG	Cc06_g05350	UPF0481 protein At3g47200-like
AX-168358192	chr6	4258781	mcG	Cc02_g05380	tyrosine N-monooxygenase-like
AX-168303067	chr6	4305097	mcG		
AX-168357178	chr6	4314349	mcG	<b>Cc06_g05400</b>	<b>1,4-alpha-glucan-branching enzyme 2-2</b>
AX-168301070	chr6	4334744	mcG		
AX-168302790	chr6	4338286	mcG	<b>Cc06_g05430</b>	<b>ABC transporter G family member 39-like isoform X2</b>
AX-168342602	chr6	4375659	mcG	Cc06_g05480	serine/threonine-protein kinase D6PKL2-like
AX-168354345	chr6	4447830	mcG	intergênica	
AX-168351134	chr6	4454535	mcG	intergênica	
AX-168371811	chr6	4467903	mcG	Cc06_g05550	GDSL esterase/lipase At5g03610-like
AX-168299990	chr6	4531144	mcG		
AX-168349250	chr6	4531827	mcG	<b>Cc06_g05640</b>	<b>CAAX prenyl protease 2</b>
AX-168350483	chr6	4531992	mcG		
AX-168304804	chr6	4569946	mcG	intergênica	
AX-168351225	chr6	4572473	mcG	intergênica	
AX-168310946	chr6	4590621	mcG	Cc06_g05710	protein BIC1-like
AX-168354124	chr6	4696954	mcG	Cc06_g05910	F-box/kelch-repeat protein At1g57790-like
AX-168351028	chr6	4720389	mcG	Cc06_g05960	uncharacterized protein LOC113692255
AX-168303538	chr6	4738239	mcG	Cc06_g05990	uncharacterized protein LOC113691749
AX-168352371	chr6	4790340	mcG	Cc06_g06080	beta-fructofuranosidase, insoluble isoenzyme 1-like isoform X1
AX-168306989	chr6	4867384	mcG	Cc06_g06170	peptidyl-prolyl cis-trans isomerase CYP18-2-like
AX-168343151	chr6	4890314	mcG	Cc06_g06220	probable F-box protein At2g36090
AX-168303120	chr6	4922381	mcG	Cc06_g06250	uncharacterized protein LOC113691887
S8_25372478	chr8	25372477	mcG	intergênica	
AX-168357486	chr11	14363294	mcG	Cc11_g03570	glutamate receptor 1.3-like
AX-168305338	chr11	14385622	mcG	intergênica	
AX-168299085	chr4	23325422	mcP	intergênica	
AX-168355753	chr9	83842	mcP	Cc09_g00110	E3 ubiquitin-protein ligase WAV3-like isoform X1

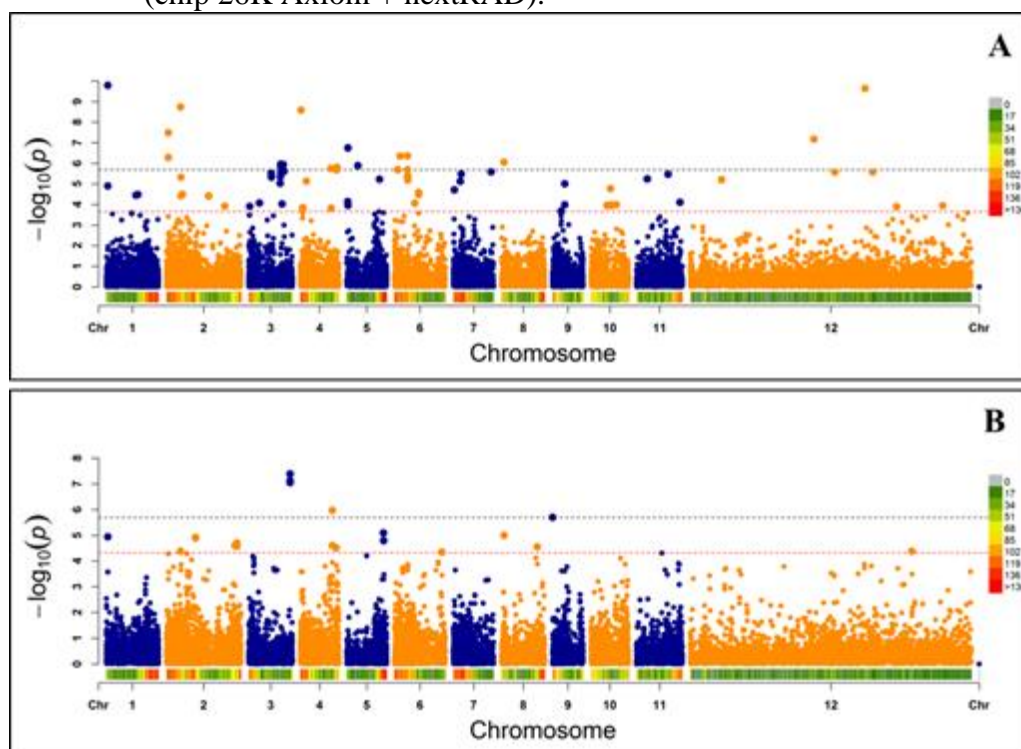
Legenda: Identificação dos marcadores associados, utilizando o rrBLUP (sem P3D), para as características de produção (p12), boia e tamanho e formato de grãos (chM – chato médio, chP – chato miúdo, mcG – moca graúdo e mcP – moca miúdo), indicando o cromossomo (Chr), posição, gene, proteína e log10 (pvalor). Destacados em azul estão os marcadores exclusivos identificados somente nessa análise (sem P3D) para o fenótipo de Moca Graúdo.

### 3.1.4 Dados genotípicos do chip 26K Axiom de *C. canephora* e nextRAD – TASSEL

Na GWAS realizada com o *software* TASSEL, associações significativas de marcadores SNPs foram encontradas para quatro características. Analisando-se o tamanho e formato de grãos, 21 SNPs adicionais foram detectados associados aos grãos do tipo Chato Miúdo e oito SNPs ao tipo Moca Miúdo, conforme apresentado na Figura 39 e Tabela 14.

Além disto, o resultado do TASSEL identificou ainda marcadores para duas novas características; para peso de 100 grãos o marcador do *chip* de DNA AX-168309741, localizado no cromossomo 10, estava associado (FIGURA 40A e TABELA 14) e para potencial hídrico foliar de antemanhã quatro marcadores nextRAD (S0\_111609941, S0\_150055733, S0\_198072686 e S10\_10265721) estavam associados (FIGURA 40B e TABELA 14).

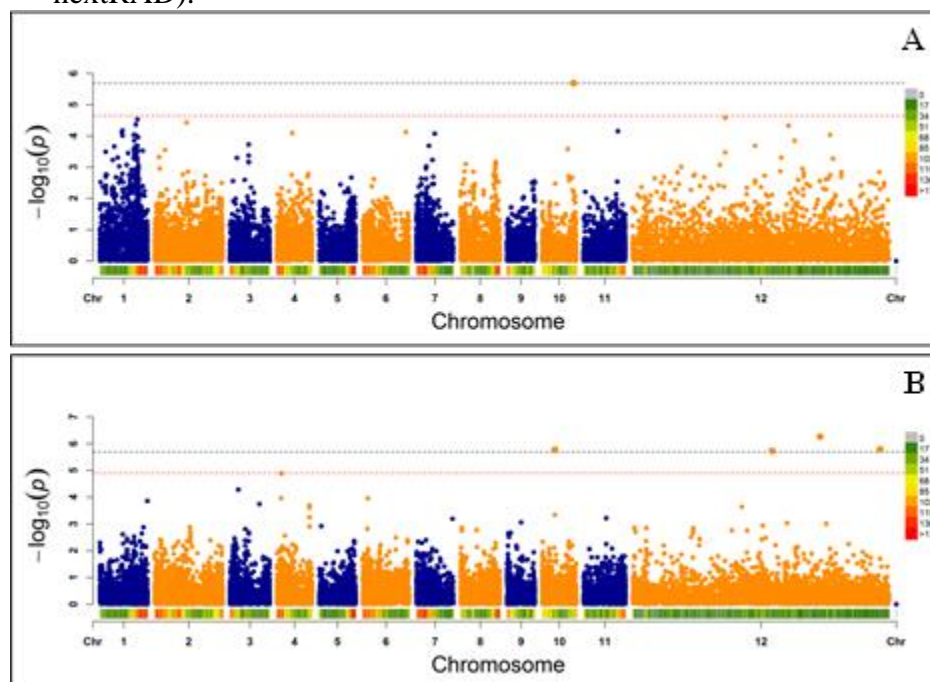
Figura 39 – Gráfico Manhattan de associação do efeito dos SNPs testados para a característica tamanho e formato de grãos utilizando o TASSEL com o conjunto de marcadores provenientes de duas técnicas de genotipagem (chip 26K Axiom + nextRAD).



Legenda: *Manhattan plot* para as características de tamanho e formato do grão. SNPs associados para os grãos classificados como Chato Miúdo (A) e Moca Miúdo (B). As linhas tracejadas representam o valor obtido para  $-\log_{10}(p)$  – valor considerando a correção de

Bonferroni (linha tracejada preta) e FDR (linha tracejada vermelha). O cromossomo 0 de *C. canephora* está representado como cromossomo 12.

Figura 40 – Gráfico Manhattan de associação do efeito dos SNPs testados para as características de peso de 100 grãos e potencial hídrico foliar de antemanhã utilizando o TASSEL com o conjunto de marcadores provenientes de duas técnicas de genotipagem (chip 26K Axiom + nextRAD).



Legenda: *Manhattan plot* para os SNPs associados as características de peso de 100 grãos (A) e potencial hídrico foliar de antemanhã (B). As linhas tracejadas representam o valor obtido para  $-\log_{10}(p)$  considerando a correção de Bonferroni (linha tracejada preta) e FDR (linha tracejada vermelha). O cromossomo 0 de *C. canephora* está representado como cromossomo 12.

Tabela 13 – Marcadores SNPs adicionais identificados na análise de associação utilizando o programa TASSEL, com o conjunto de marcadores provenientes de duas técnicas de genotipagem (chip 26K Axiom + nextRAD).

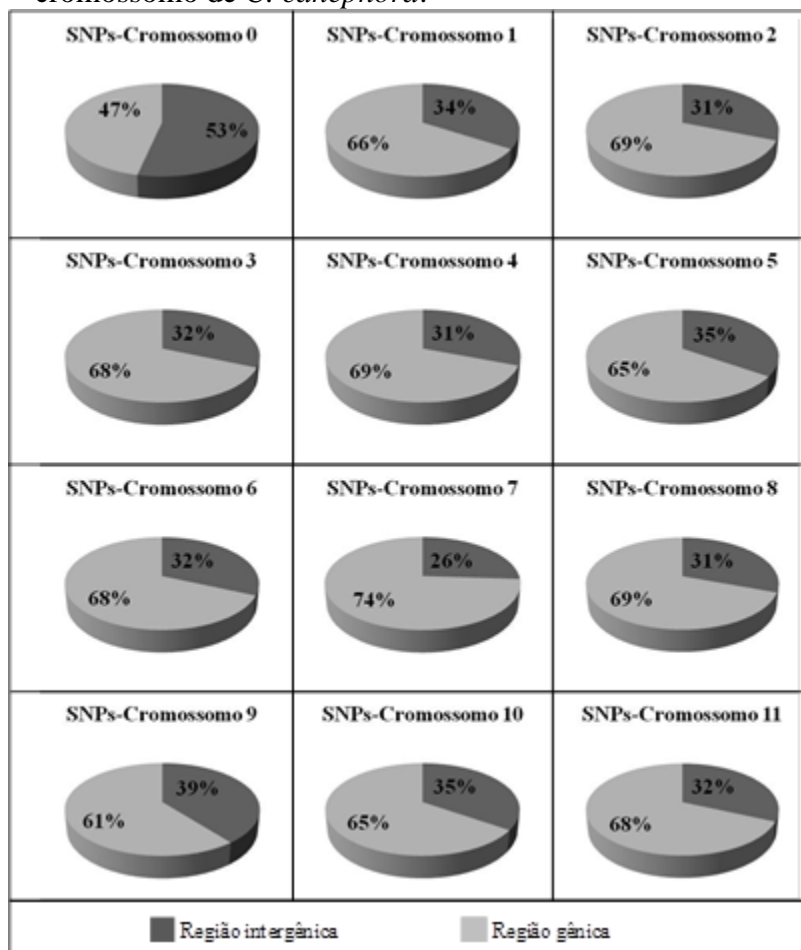
Marcador	Chr	Posição	Fenótipo	Gene	Proteína	Log10 (pvalor)
AX-168309741	chr10	25244253	p100	intergênica		5,69
S0_22957200	chr0	22957199	chP	intergênica		5,21
S0_90384806	chr0	90384805	chP	intergênica		7,18
S0_127662678	chr0	127662677	chP	intergênica		9,63
AX-168345855	chr0	150787969	chP	intergênica		3,89
S1_1190308	chr1	1190307	chP	intergênica		4,90
AX-168343633	chr1	21662434	chP	intergênica		4,45
S1_23113352	chr1	23113351	chP	Cc01_g06230	uncharacterized protein LOC113720213 isoform X1	4,49
S2_42640423	chr2	42640422	chP	intergênica		3,92
S3_838902	chr3	838901	chP	Cc03_g01110	uncharacterized protein LOC113733901	3,90
S3_24286751	chr3	24286750	chP	intergênica		4,02
S4_508477	chr4	508476	chP	Cc04_g00690		8,58
AX-168313313	chr4	1881580	chP	Cc04_g02490		3,80
AX-168343159	chr4	22461703	chP	Cc04_g14730	putative disease resistance RPP13-like protein 1	3,83
S5_962753	chr5	962752	chP	intergênica		3,95
S5_962780	chr5	962779	chP	intergênica		6,75
S5_962789	chr5	962788	chP	intergênica		6,75
S5_962794	chr5	962793	chP	intergênica		6,74
S5_962799	chr5	962798	chP	intergênica		6,74
AX-168358748	chr6	14822146	chP	intergênica		4,06
AX-168371703	chr9	6259614	chP	Cc09_g05840	potassium transporter 11-like isoform X1	3,68
S10_14346652	chr10	14346651	chP	intergênica		4,77
AX-168345100	chr0	161912726	mcP	intergênica		4,38
AX-168307006	chr2	10336701	mcP	Cc02_g12290	pre-mRNA-splicing factor SYF1-like isoform X1	4,38
S2_21294997	chr2	21294996	mcP	intergênica		4,89
S2_21295008	chr2	21295007	mcP	intergênica		4,89
S2_21295021	chr2	21295020	mcP	intergênica		4,92
S3_30315983	chr3	30315982	mcP	<b>Cc03_g15080</b>	<b>altered inheritance rate of mitochondria protein 2</b>	7,05
S4_26107964	chr4	26107963	mcP	Cc04_g16160	TORTIFOLIA1-like protein 3	4,51
AX-168350047	chr6	34672019	mcP	intergênica		4,34
S0_111609941	chr0	111609940	pot	intergênica		5,73
S0_150055733	chr0	150055732	pot	intergênica		6,26
S0_198072686	chr0	198072685	pot	intergênica		5,79
S10_10265721	chr10	10265720	pot	intergênica		5,77

Legenda: Identificação dos marcadores adicionais associados nas análises realizadas no TASSEL com as duas metodologias de genotipagem (chip 26K Axiom + nextRAD). Para cada marcador está indicado o cromossomo (Chr), a posição, a característica associada (p100 – peso de 100 grãos, chP – chato miúdo, mcP – moca miúdo e pot – potencial hídrico foliar de antemanhã), a proteína e log10 (pvalor).

Sem dúvidas ao incluir um maior número de marcas distribuídas ao longo do genoma as chances de identificar associação para as características de interesse aumentam, ao acrescentar os mais de 5.000 marcadores nextRAD para os 366 indivíduos, notou-se que outras características foram detectadas na GWAS. Outro fator que poderia influenciar as análises de associação seria a localização desses marcadores, se em região gênica ou intergênica. Se estiverem localizados dentro ou muito próximos ao gene de interesse, existiria uma maior probabilidade em encontrar associações. As duas técnicas de genotipagem utilizadas nesse estudo identificaram uma maior

porcentagem de marcadores em regiões gênicas, pouco mais de 65,5% dos marcadores nextRAD (FIGURA 41) e cerca de 68% das sondas contidas no *chip* estavam localizadas em genes.

Figura 41 – Percentagens de SNPs identificados nos locos nextRAD para cada cromossomo de *C. canephora*.



Legenda: Os SNPs localizados em regiões gênicas estão representados em cinza claro e em regiões intergênicas de cinza escuro em cada cromossomo de *C. canephora*.

Fonte: Do autor (2018).

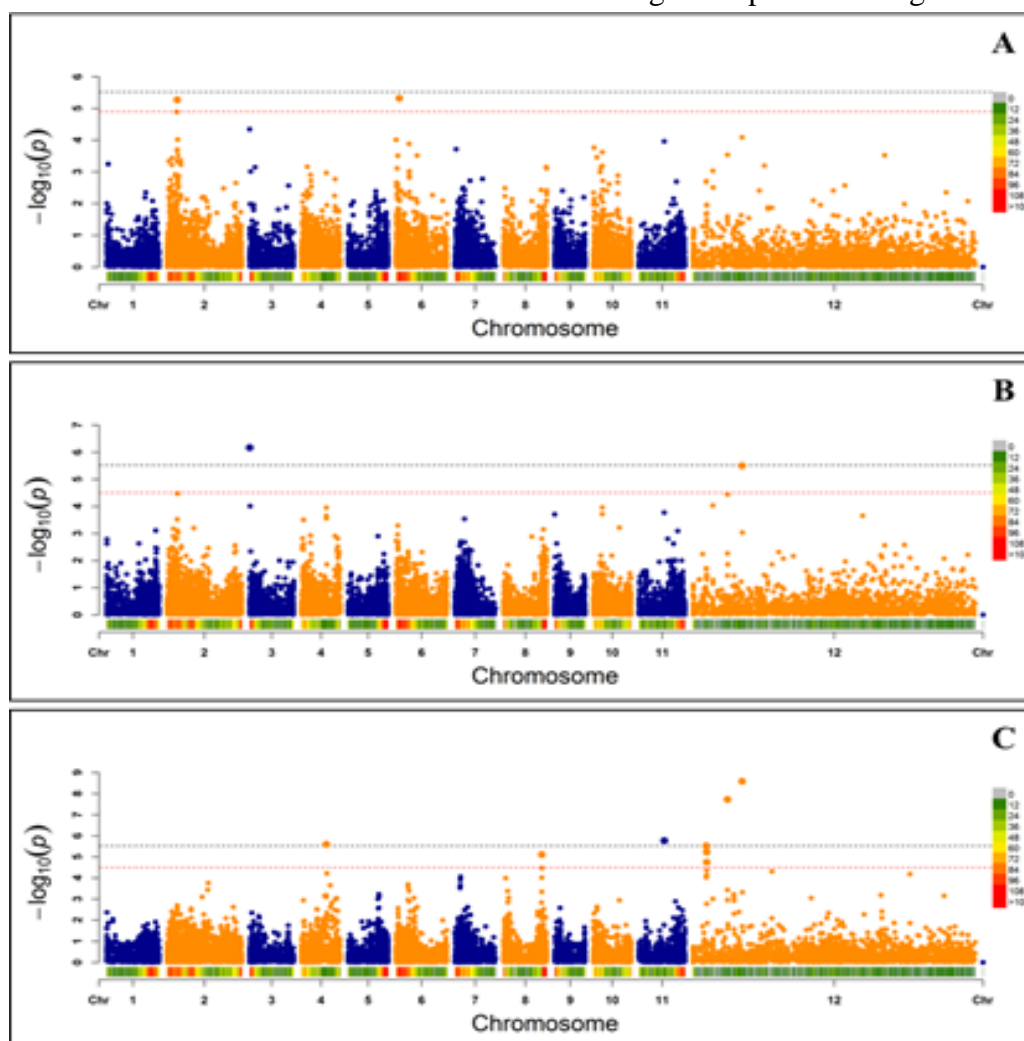


## 3.2 Análise de 1.319 indivíduos

### 3.2.1 rrBLUP

A análise de GWAS realizada para os 1.319 indivíduos selecionados na Embrapa Cerrados, utilizando 16.688 SNPs identificados no *chip* de *C. canephora*, identificou um total de 13 SNPs associados (FIGURA 42 e TABELA 15).

Figura 42 – Gráfico Manhattan de associação do efeito dos SNPs testados com as características de tamanho e formato de grãos e peso de 100 grãos.



Legenda: *Manhattan plot* para as características de tamanho e formato do grão e peso de 100 grãos. SNPs associados para os grãos classificados como Chato Miúdo (A), Moca Miúdo (B) e peso de 100 grãos (C). As linhas tracejadas representam o valor obtido para  $-\log_{10}(p\text{-valor})$  considerando a correção de Bonferroni (linha tracejada preta) e FDR (linha tracejada vermelha). O cromossomo 0 de *C. canephora* está representado como cromossomo 12.

Para as características de tamanho e formato de grãos, os SNPs associados para Chato Miúdo e Moca Miúdo estão apresentados nas Figuras 42A e 42B, respectivamente, identificando dois marcadores para cada característica. Na Figura 42C

e Tabela 15 visualizam-se os nove marcadores identificados na GWAS para peso de 100 grãos, sendo que seis localizam-se no cromossomo 0 e os demais nos cromossomos 4, 8 e 11.

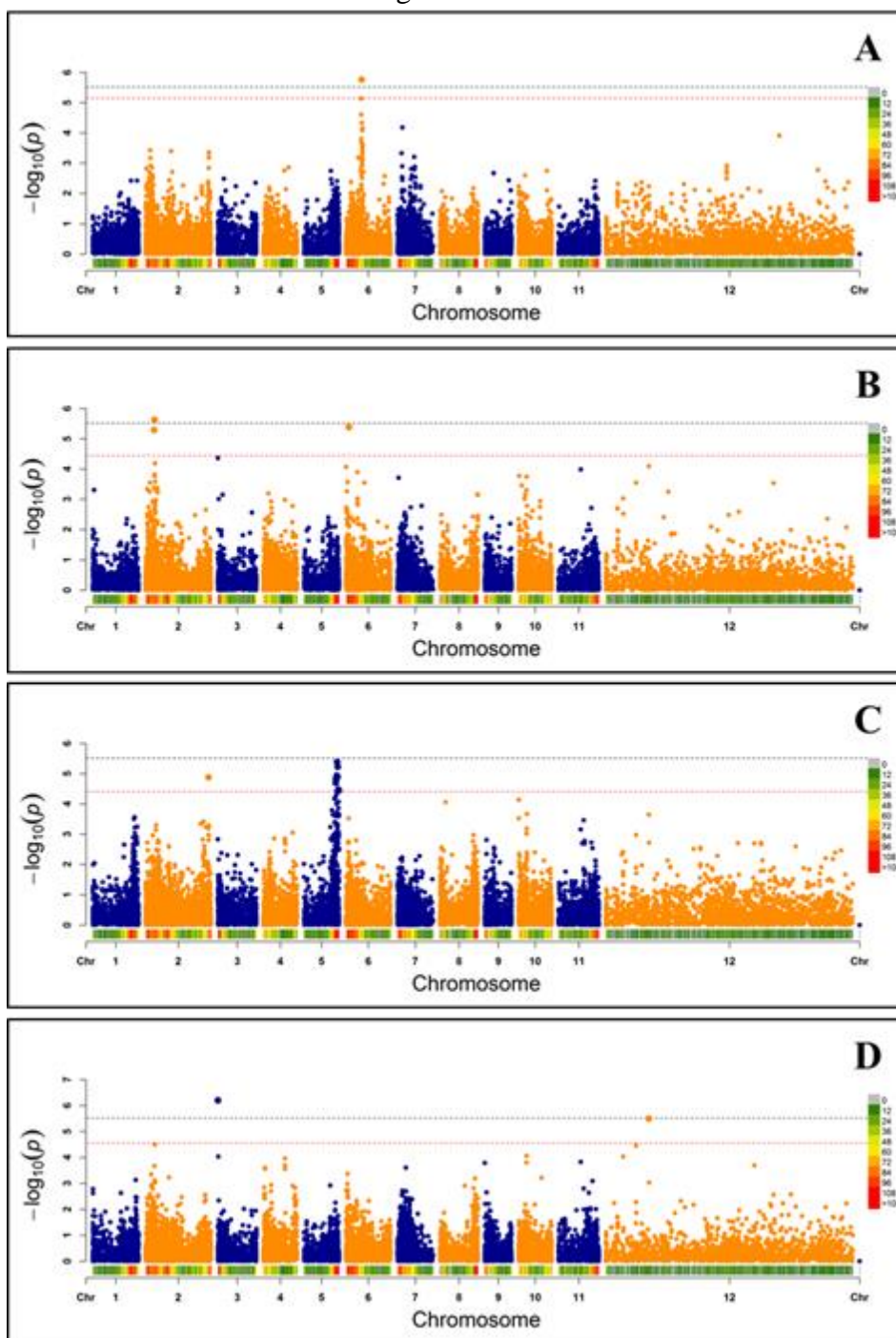
Tabela 14 – Marcadores associados para as características de peso de 100 grãos e tamanho e formato de grãos, na análise com os 1.319 indivíduos e utilizando o rrBLUP (com P3D)

Marcador	Chr	Posição	Fenótipo	Gene	Proteína	Log10 (pvalor)
AX-168308807	chr0	9786333	p100	intergênica		5,52
AX-168355686	chr0	9842073	p100	Cc00_g01520	uncharacterized protein LOC113743019	5,51
AX-168348258	chr0	10030733	p100	intergênica		5,23
AX-168342756	chr0	10065719	p100	intergênica		4,75
AX-168358131	chr0	25149927	p100	Cc00_g03450	ABC transporter B family member 25-like isoform X2	7,72
AX-168342522	chr0	35764294	p100	intergênica		8,59
AX-168348970	chr4	18081017	p100	Cc04_g13590	pyridoxine/pyridoxamine 5'-phosphate oxidase 1, chloroplastic-like	5,60
AX-168357068	chr8	27955474	p100	intergênica		5,11
AX-168351372	chr11	18451769	p100	intergênica		5,76
AX-168355324	chr2	7588600	chP	Cc02_g09340	Flavonol 4'-sulfotransferase, putative	5,27
AX-168348505	chr6	2538355	chP	Cc06_g03280	uncharacterized protein LOC113694498	5,32
AX-168342522	chr0	35764294	mcP	intergênica		5,49
AX-168355983	chr3	74856	mcP	intergênica		6,16

Legenda: Identificação dos marcadores associados, utilizando o rrBLUP (com P3D), para as características de peso de grãos (p100) e tamanho e formato de grãos (chP – chato miúdo e mcP – moca miúdo), indicando o cromossomo (Chr), posição, gene, proteína e log10 (pvalor).

Para esse mesmo conjunto de dados foi realizada a análise no rrBLUP sem P3D, estimando os componentes de variância para cada marcador separadamente, identificando no total, para tamanho e formato de grãos, 19 marcadores associados (FIGURA 43 e TABELA 16); sendo que um marcador presente no cromossomo 6 estava associado para a característica Chato Médio (FIGURA 43A), três tiveram associação para a característica Chato Miúdo (FIGURA 43B), posicionados nos cromossomos 2 (AX-168348734, AX-168355324) e 6 (AX-168348505), 13 SNPs estavam associados para Moca Graúdo, podendo ser observado na Figura 43C que somente um marcador está localizado no cromossomo 2 e os demais encontram-se localizados em regiões próximas (26.401.552 – 28.859.558) do cromossomo 5, indicando a possibilidade de ser um QTL. Dois outros marcadores, associados à característica de Moca Miúdo (FIGURA 43D), estavam localizados nos cromossomo 3 (AX-168355983) e o no cromossomo 0 (AX-168342522).

Figura 43 – Gráfico Manhattan de associação do efeito dos SNPs testados com a característica de tamanho e formato de grãos.



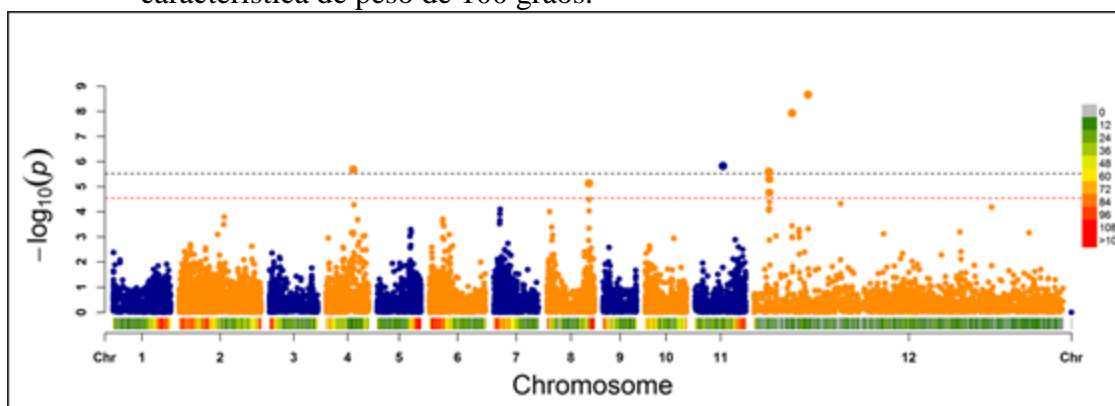
Legenda: *Manhattan plot* para as características de tamanho e formato do grão. SNPs associados para os grãos classificados como Chato Médio (A), Chato Miúdo (B), Moca Graúdo (C) e Moca Miúdo (D). As linhas tracejadas representam o valor obtido para  $-\log_{10}(p)$  considerando a correção de Bonferroni (linha tracejada preta) e FDR (linha tracejada vermelha). O cromossomo 0 de *C. canephora* está representado como cromossomo 12.

Na Figura 44 visualizam-se os marcadores associados à característica de peso de 100 grãos, em que nove marcadores estavam associados, sendo que seis encontram-se



localizados no cromossomo 0 e um destes, AX-168342522, também foi associado à característica de Moca Miúdo.

Figura 44 – Gráfico Manhattan de associação do efeito dos SNPs testados com a característica de peso de 100 grãos.



Legenda: *Manhattan plot* para a característica peso de 100 grãos. As linhas tracejadas representam o valor obtido para  $-\log_{10}(p)$  considerando a correção de Bonferroni (linha tracejada preta) e FDR (linha tracejada vermelha). O cromossomo 0 de *C. canephora* está representado como cromossomo 12.

Tabela 15 – Marcadores associados para as características de peso de 100 grãos e tamanho e formato de grãos, na análise com os 1.319 indivíduos e utilizando o rrBLUP (sem P3D).

Marcador	Chr	Posição	Fenótipo	Gene	Proteína	Log10 (pvalor)
AX-168308807	chr0	9786333	p100	intergênica		5,60
AX-168355686	chr0	9842073	p100	Cc00_g01520	uncharacterized protein LOC113743019	5,59
AX-168348258	chr0	10030733	p100	intergênica		5,30
AX-168342756	chr0	10065719	p100	intergênica		4,76
AX-168358131	chr0	25149927	p100	Cc00_g03450	ABC transporter B family member 25-like isoform X2	7,93
AX-168342522	chr0	35764294	p100	intergênica		8,66
AX-168348970	chr4	18081017	p100	Cc04_g13590	pyridoxine/pyridoxamine 5'-phosphate oxidase 1, chloroplastic-like	5,68
AX-168357068	chr8	27955474	p100	intergênica		5,13
AX-168351372	chr11	18451769	p100	intergênica		5,83
AX-168345461	chr6	13075988	chM	intergênica		5,76
AX-168348734	chr2	7138427	chP	Cc02_g08840	cysteine desulfurase 1, chloroplastic-like isoform X1	5,29
AX-168355324	chr2	7588600	chP	Cc02_g09340	Flavonol 4'-sulfotransferase, putative	5,63
AX-168348505	chr6	2538355	chP	Cc06_g03280	uncharacterized protein LOC113694498	5,39
AX-168297987	chr2	52458925	mcG	Cc02_g37620	T-complex protein 1 subunit theta-like	4,88
AX-168344359	chr5	26401552	mcG	intergênica		4,70
AX-168346202	chr5	26581667	mcG	intergênica		4,85
AX-168346321	chr5	26690962	mcG	Cc05_g12960	G-type lectin S-receptor-like serine/threonine-protein kinase At4g27290	4,78
AX-168371903	chr5	27299271	mcG	intergênica		4,94
AX-168357408	chr5	27563621	mcG	Cc05_g14220	ethylene-responsive transcription factor 3-like	5,38
AX-168371643	chr5	27839867	mcG	intergênica		4,90
AX-168342644	chr5	28021680	mcG	Cc05_g14940	fasciclin-like arabinogalactan protein 1	5,23
AX-168350481	chr5	28114364	mcG	Cc05_g15070	Thioredoxin-like fold protein	4,42
AX-168343448	chr5	28347586	mcG	Cc05_g15500	enhancer of mRNA-decapping protein 4-like isoform X1	4,44
AX-168301499	chr5	28428018	mcG	intergênica		4,41
AX-168353973	chr5	28563082	mcG	Cc05_g15780	chloride channel protein CLC-f isoform X1	4,47
AX-168358243	chr5	28859558	mcG	Cc05_g16200	molybdenum cofactor sulfuryase-like isoform X1	4,46
AX-168342522	chr0	35764294	mcP	intergênica		5,50
AX-168355983	chr3	74856	mcP	intergênica		6,20

Legenda: Identificação dos marcadores associados, utilizando o rrBLUP (sem P3D), para as características de peso de grãos (p100) e tamanho e formato de grãos (chM – chato médio, chP – chato miúdo, mcG – moca grão e mcP – moca miúdo), indicando o cromossomo (Chr), posição, gene, proteína e log10 (pvalor).

Nota-se que o número de marcas e de características associadas resultantes das análises no rrBLUP sem a opção P3D aumentou consideravelmente para os 1.319 indivíduos. Duas novas características apresentaram associação e o número de marcadores associados saltou de 13 para 28.

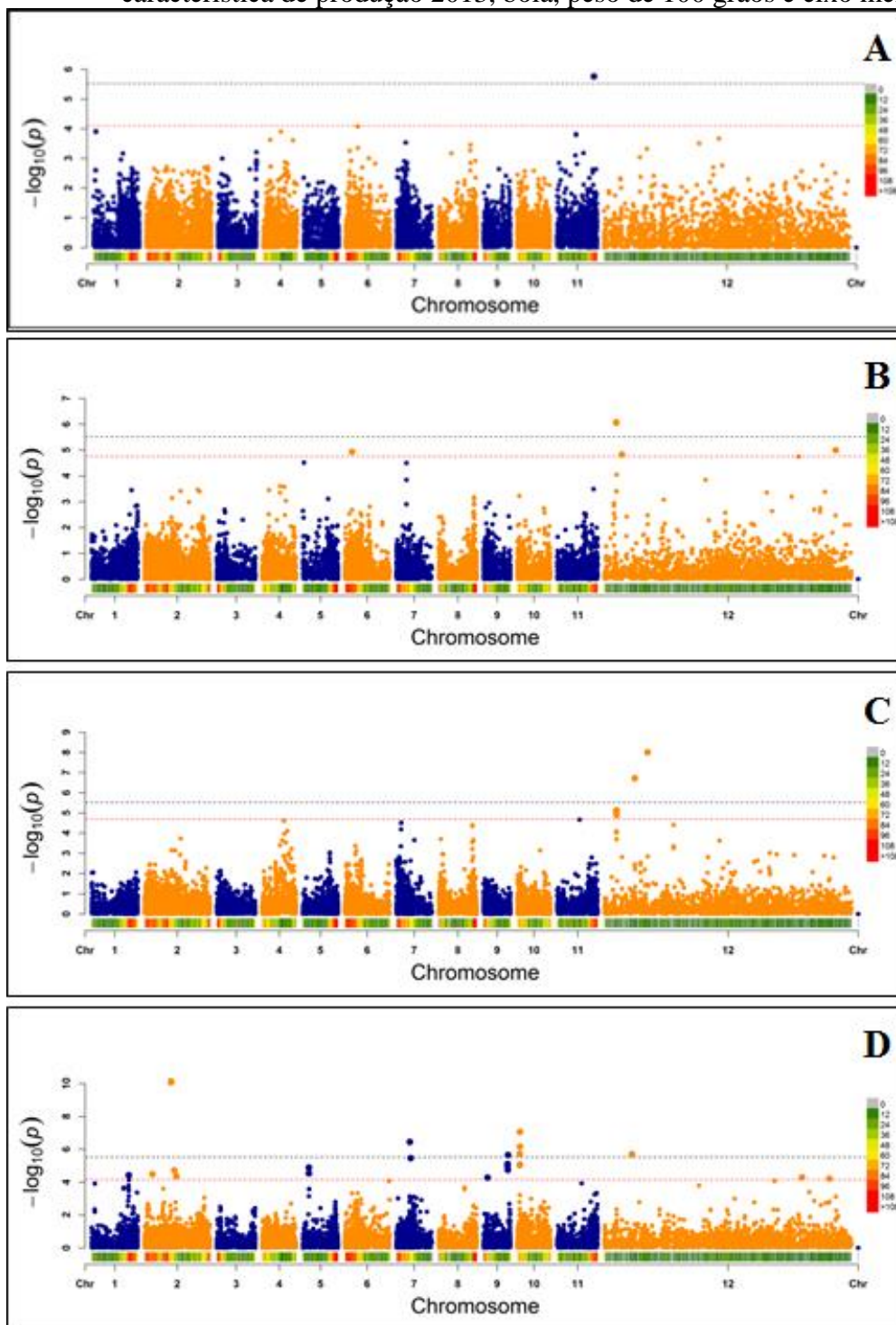
### 3.2.2 TASSEL

Para as análises realizadas no software TASSEL, utilizando os mesmo conjunto de dados utilizado no rrBLUP, um marcador presente no cromossomo 11 foi associado para a característica de produção, avaliada no ano de 2013 (FIGURA 45A e TABELA 17).

Duas outras características que não haviam sido detectadas pelo rrBLUP foram identificadas pela análise com o TASSEL, 6 SNPs apresentaram associação para o fenótipo Boia (FIGURA 45B), 5 presentes no cromossomo 0 e um no cromossomo 6, e 23 marcadores estavam associados ao tamanho do fruto, relacionados à característica de eixo menor (FIGURA 45D). A Figura 45C apresenta os seis marcadores associados à peso de 100 grãos, todos eles localizados no cromossomo 0.

Observando-se os dados apresentados na Tabela 17, pode-se notar que 3 SNPs associados à característica boia, também foram detectados em associação com a característica de p100 (AX-168308807, AX-168355686 e AX-168358258). Todos esses, localizados no cromossomo 0. Já para a característica do fruto, eixo menor, a qual, somente nesta análise, marcadores associados, foram detectados, pode-se observar pelos dados apresentados na Tabela 17, que mais de um SNP em associação significativa foi encontrado em um mesmo gene. Cc02\_g24790 (AX-168303576 e AX-168343505), que codifica uma “*GPI-anchored protein like*”, sendo que “*Glycosylphosphatidylinositol-Anchored Proteins*” GPI proteínas são importantes na formação da parede celular e também na morfogênese (GILMOR et al, 2005) e, o gene Cc10\_g02630 que codifica um frutokinase (“*putative pfkB family carbohydrate kinase*”, proteínas envolvidas no metabolismo de frutoses (PEGO; SMEEKENS, 2001).

Figura 45 – Gráfico Manhattan de associação do efeito dos SNPs testados com a característica de produção 2013, boia, peso de 100 grãos e eixo menor.



Legenda: *Manhattan plot* para as características de Produção 2013 (A), boia (B), peso de 100 grãos (C) e eixo menor (D). As linhas tracejadas representam o valor obtido para  $-\log_{10}(p)$  considerando a correção de Bonferroni (linha tracejada preta) e FDR (linha tracejada vermelha). O cromossomo 0 de *C. canephora* está representado como cromossomo 12.

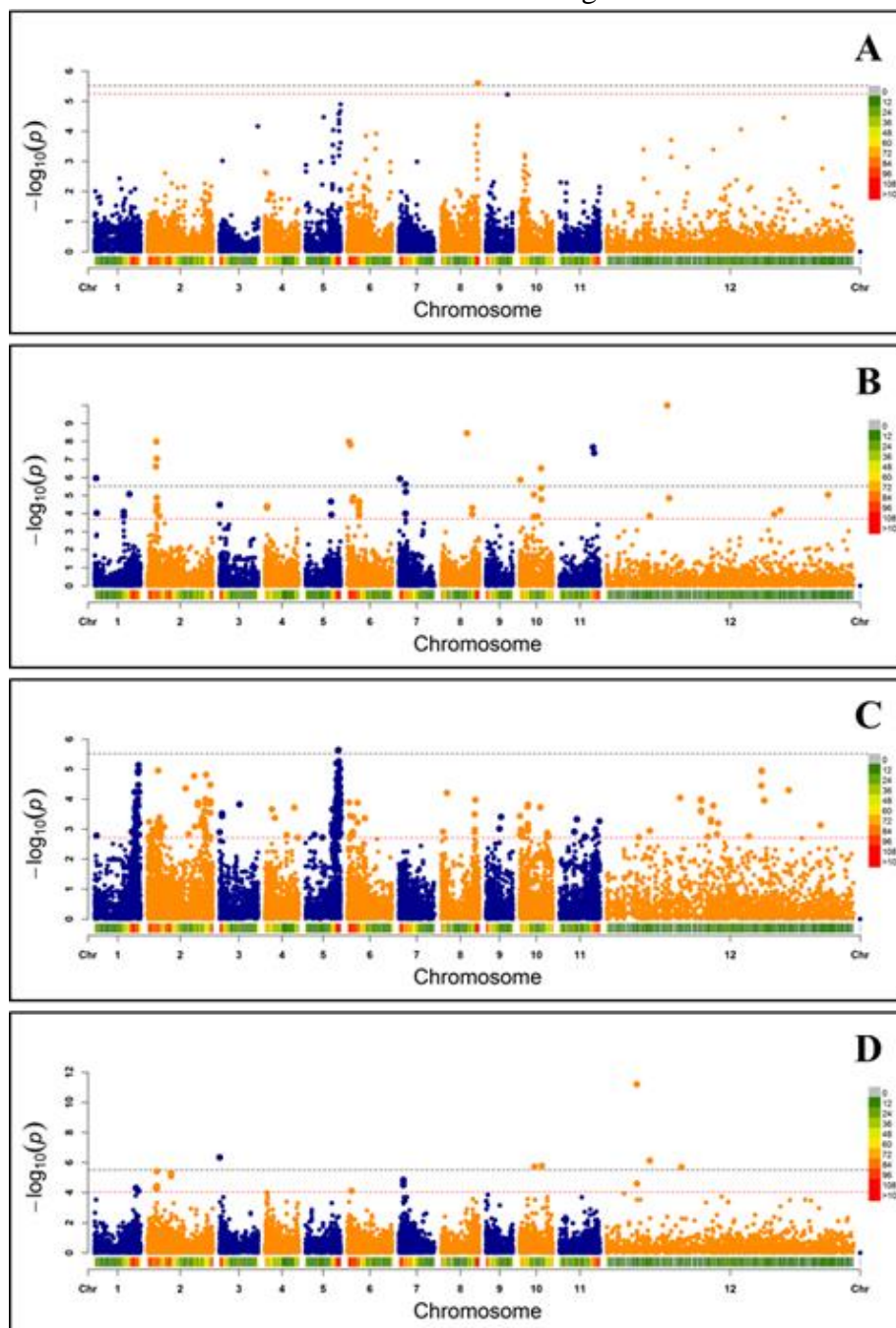
Tabela 16 – Marcadores associados para as características de produção, boia, peso de 100 grãos e tamanho de fruto, na análise com os 1.319 indivíduos e utilizando o programa TASSEL.

Marcador	Chr	Posição	Fenótipo	Gene	Proteína	Log10 (pvalor)
AX-168306468	chr11	30890580	p13	intergênica		5,76
AX-168308807	chr0	9786333	Boia	intergênica		6,06
AX-168355686	chr0	9842073	Boia	Cc00_g01520	uncharacterized protein LOC113740446 isoform X1	6,06
AX-168313352	chr0	14547561	Boia	intergênica		4,82
AX-168358258	chr0	161279672	Boia	intergênica		4,74
AX-168350461	chr0	192412669	Boia	intergênica		4,99
AX-168341377	chr6	5939217	Boia	intergênica		4,94
AX-168308807	chr0	9786333	p100	intergênica		5,10
AX-168355686	chr0	9842073	p100	Cc00_g01520	uncharacterized protein LOC113740446 isoform X1	5,07
AX-168348258	chr0	10030733	p100	intergênica		5,13
AX-168342756	chr0	10065719	p100	intergênica		4,89
AX-168358131	chr0	25149927	p100	Cc00_g03450	ABC transporter B family member 25-like isoform X2	6,72
AX-168342522	chr0	35764294	p100	intergênica		8,00
AX-168309730	chr0	22915872	emenor	Cc00_g03210	uncharacterized protein LOC113708797	5,70
AX-168307262	chr0	164357922	emenor	intergênica		4,29
AX-168358902	chr0	187027050	emenor	intergênica		4,21
AX-168299745	chr1	30731985	emenor	intergênica		4,41
AX-168343597	chr2	6668157	emenor	Cc02_g08330	zinc finger protein ZAT11-like	4,48
AX-168303576	chr2	22040497	emenor	<b>Cc02_g24790</b>	<b>GPI-anchored protein like</b>	10,13
AX-168343505	chr2	22041597	emenor			10,07
AX-168355466	chr2	24758035	emenor	Cc02_g27080	protein N-terminal asparagine amidohydrolase	4,71
AX-168308319	chr2	26577824	emenor	intergênica		4,36
AX-168354342	chr5	5013195	emenor	Cc05_g01700	putative methyltransferase NSUN6	4,53
AX-168353013	chr5	5057074	emenor	intergênica		4,88
AX-168346600	chr7	11180970	emenor	intergênica		6,45
AX-168309934	chr7	11879643	emenor	Cc07_g15170	ras-related protein RABH1e	5,45
AX-168305682	chr9	3268799	emenor	intergênica		4,27
AX-168308409	chr9	19961980	emenor	Cc09_g09610	Zinc transport protein like	5,10
AX-168353406	chr9	19983202	emenor	intergênica		4,97
AX-168355113	chr9	20133026	emenor	intergênica		4,74
AX-168297043	chr9	20245702	emenor	intergênica		5,85
AX-168357124	chr10	2011076	emenor	<b>Cc10_g02630</b>	<b>putative pfkB family carbohydrate kinase</b>	5,72
AX-168351596	chr10	2011145	emenor			5,03
AX-168354675	chr10	2048242	emenor	intergênica		5,07
AX-168309693	chr10	2227748	emenor	intergênica		6,15
AX-168297375	chr10	2249135	emenor	Cc10_g03010	uncharacterized protein LOC113713027 isoform X2	7,05

Legenda: Identificação dos marcadores associados, utilizando o TASSEL, para as características de produção (p13), boia, peso de grãos (p100) e morfologia de fruto (emenor-eixo menor), indicando o cromossomo (Chr), posição, gene, proteína e log10 (pvalor).

Para tamanho e formato de grãos, um marcador estava associado para grãos do tipo Chato Graúdo (FIGURA 46A), 57 SNPs para Chato Miúdo (FIGURA 46B), 485 SNPs para Moca Graúdo (FIGURA 46C) e para Moca Miúdo, 24 marcas associadas (FIGURA 46D). Observando-se a localização dos SNPs associados para Moca Graúdo (FIGURA 46C), aproximadamente 84% dessas marcas foram identificadas nos cromossomos 1 (113 SNPs), cromossomo 2 (117 SNPs) e cromossomo 5 (176 SNPs) de *C. canephora*. Nas Tabelas 18 e 19 estão apresentados os marcadores em associação para as características de tamanho e formato do fruto, encontradas na presente análise. No entanto, para moca graúdo, somente estão relacionados os 50 marcadores com menores p-valores (TABELA 19).

Figura 46 – Gráfico Manhattan de associação do efeito dos SNPs testados com a característica de tamanho e formato de grãos.



Legenda: *Manhattan plot* para as características de tamanho e formato do grão. SNPs associados para os grãos classificados como Chato Graúdo (A), Chato Miúdo (B), Moca Graúdo (C) e Moca Miúdo (D). As linhas tracejadas representam o valor obtido para  $-\log_{10}(p\text{-valor})$  considerando a correção de Bonferroni (linha tracejada preta) e FDR (linha tracejada vermelha). O cromossomo 0 de *C. canephora* está representado como cromossomo 12.



Tabela 17 – Marcadores associados para as características de tamanho e formato de grãos (chato grúdo e chato miúdo) na análise com os 1.319 indivíduos e utilizando o programa TASSEL.

Marcador	Chr	Posição	Fenótipo	Gene	Proteína	Log10 (pvalor)
AX-168342455	chr8	30285270	chG	Cc08_g15550	cation/H(+) antiporter 18-like isoform X1	5,60
AX-168342522	chr0	35764294	chP	intergênica		3,87
AX-168351301	chr0	50327373	chP	intergênica		9,99
AX-168342372	chr0	52042011	chP	intergênica		4,85
AX-168371943	chr0	139468519	chP	Cc00_g21970	MDIS1-interacting receptor like kinase 2-like	3,96
AX-168312126	chr0	144649094	chP	intergênica		4,19
AX-168372309	chr0	184367809	chP	intergênica		5,04
AX-168299145	chr1	1126012	chP	Cc01_g00640	IQ domain-containing protein IQM2-like	5,96
AX-168296877	chr1	1634705	chP	intergênica		4,02
AX-168353642	chr1	24284629	chP		<b>La-related protein 1A</b>	3,86
AX-168302945	chr1	24284681	chP	<b>Cc01_g06920</b>		4,09
AX-168353469	chr1	28904683	chP	Cc01_g10220	pheophorbide a oxygenase, chloroplastic-like	5,08
AX-168348734	chr2	7138427	chP	Cc02_g08840	cysteine desulfurase 1, chloroplastic-like isoform X1	6,60
AX-168305336	chr2	7551514	chP		<b>7-methylguanosine phosphate-specific 5'-nucleotidase A-like</b>	4,24
AX-168349189	chr2	7562684	chP	<b>Cc02_g09290</b>		4,12
AX-168355324	chr2	7588600	chP	Cc02_g09340	Flavonol 4'-sulfotransferase, putative	7,98
AX-168309352	chr2	7802958	chP	Cc02_g09610	ribonuclease H2 subunit A-like	7,04
AX-168348474	chr2	7806809	chP	intergênica		4,87
AX-168344634	chr2	7956595	chP	intergênica		4,24
AX-168344537	chr2	7957761	chP	Cc02_g09820	R3H domain-containing protein 2-like	4,24
AX-168312035	chr2	8063368	chP	intergênica		4,47
AX-168359112	chr2	9959463	chP	intergênica		3,85
AX-168355983	chr3	74856	chP	intergênica		4,48
AX-168358718	chr4	1742432	chP	Cc04_g02250	60S ribosomal protein L28-1-like	4,40
AX-168358482	chr4	1823968	chP	intergênica		4,32
AX-168302488	chr5	21080225	chP	intergênica		4,66
AX-168352281	chr5	21454337	chP	intergênica		3,92
AX-168357727	chr6	1333284	chP	Cc06_g01620	laccase-4-like	7,98
AX-168348505	chr6	2538355	chP	Cc06_g03280	uncharacterized protein LOC113694498	7,80
AX-168344460	chr6	4164904	chP	Cc06_g05270	polyadenylation and cleavage factor homolog 4-like isoform X1	4,70
AX-168343288	chr6	5023959	chP	intergênica		4,90
AX-168357173	chr6	9545143	chP	Cc06_g11810	protein EARLY FLOWERING 5-like	4,34
AX-168315406	chr6	9552364	chP	Cc06_g11820	WD repeat-containing protein VIP3-like	4,28
AX-168354529	chr6	9680854	chP	intergênica		4,00
AX-168354238	chr6	9760152	chP	Cc06_g12050	uncharacterized protein LOC113696977 isoform X1	4,67
AX-168341207	chr6	9761259	chP	intergênica		4,59
AX-168356774	chr6	9770901	chP	intergênica		3,87
AX-168358996	chr6	9982684	chP	Cc06_g12320	ATP-dependent DNA helicase Q-like 3 isoform X2	4,67
AX-168343647	chr7	873756	chP	Cc07_g01330	protein COFACTOR ASSEMBLY OF COMPLEX C SUBUNIT B CCB1	5,92
AX-168356907	chr7	5407344	chP	Cc07_g07350	growth-regulating factor 1-like isoform X1	5,62
AX-168308739	chr7	5505044	chP	intergênica		4,00
AX-168349545	chr7	5702520	chP	Cc07_g07740	tRNA-dihydrouridine(20) synthase [NAD(P)+]-like isoform X2	5,21
AX-168354649	chr8	21391981	chP	Cc08_g07910	chorismate mutase 2-like isoform X1	8,45
AX-168342859	chr8	25748304	chP	Cc08_g10930	Golgi transport complex subunit	3,95
AX-168350329	chr8	25801832	chP	Cc08_g10990	serine/threonine-protein kinase tricorner	4,30
AX-168349506	chr10	574254	chP	Cc10_g00650	uncharacterized protein LOC113712807	5,87
AX-168345765	chr10	11726758	chP	intergênica		5,04
AX-168372147	chr10	11748552	chP	intergênica		3,83
AX-168309899	chr10	11777153	chP	intergênica		3,82
AX-168310468	chr10	14573473	chP		<b>uncharacterized protein LOC113714577 isoform X1</b>	3,84
AX-168355206	chr10	14573896	chP	<b>Cc10_g10700</b>		3,83
AX-168310228	chr10	14742543	chP	intergênica		3,84
AX-168356073	chr10	14780965	chP	Cc10_g10760	receptor kinase-like protein Xa21	3,83
AX-168351366	chr10	17902648	chP		<b>protein TIC 40, chloroplastic-like isoform X2</b>	5,40
AX-168351350	chr10	17906346	chP	<b>Cc10_g11320</b>		6,51
AX-168348689	chr10	18050185	chP	intergênica		4,78
AX-168372376	chr11	27910613	chP	Cc11_g10490	mediator of RNA polymerase II transcription subunit 25	7,68
AX-168354354	chr11	28763501	chP	Cc11_g11480	phosphatidylinositol 4-kinase gamma 7-like isoform X1	7,36

Legenda: Identificação dos marcadores associados, utilizando o TASSEL, para as características de tamanho e formato de grãos (chG – chato grúdo e chP – chato miúdo) indicando o cromossomo (Chr), posição, gene, proteína e log10 (pvalor).

Tabela 18 – Marcadores associados para as características de tamanho e formato de grãos, indicando os 50 marcadores mais significativos para moca graúdo, na análise com os 1.319 indivíduos e utilizando o programa TASSEL.

Marcador	Chr	Posição	Fenótipo	Gene	Proteína	Log10 (pvalor)
AX-168352499	chr0	128965321	mcG	intergênica		4,95
AX-168351974	chr0	128971355	mcG	intergênica		4,93
AX-168303844	chr0	128973523	mcG	intergênica		4,95
AX-168304865	chr0	128974922	mcG	intergênica		4,95
AX-168344491	chr0	128977308	mcG	intergênica		4,94
AX-168350099	chr1	35916313	mcG	Cc01_g18940	beta-carotene isomerase D27, chloroplastic-like	4,91
AX-168351864	chr1	35954271	mcG	intergênica		4,91
AX-168350800	chr1	35965818	mcG	Cc01_g18990	protein SPIRRIG-like isoform X1	4,91
AX-168350278	chr1	36047250	mcG	intergênica		4,91
AX-168344238	chr1	36117292	mcG	<b>Cc01_g19150</b>	<b>homocysteine S-methyltransferase 2-like</b>	4,93
AX-168304760	chr1	36117547	mcG			4,93
AX-168349254	chr1	36136097	mcG	intergênica		4,88
AX-168371733	chr1	36154034	mcG	<b>Cc01_g19250</b>	<b>non-specific lipid-transfer protein-like protein At2g13820</b>	4,92
AX-168344419	chr1	36154295	mcG			4,91
AX-168307890	chr1	36163621	mcG	Cc01_g19260	F-box/kelch-repeat protein SKIP30-like	4,90
AX-168350197	chr1	36330604	mcG	intergênica		4,91
AX-168371659	chr1	36398376	mcG	intergênica		5,14
AX-168358897	chr1	36423435	mcG	Cc01_g19640	basic 7S globulin-like	4,95
AX-168350068	chr1	36485520	mcG	Cc01_g19730	rab11 family-interacting protein 3 isoform X3	4,95
AX-168358332	chr1	36643242	mcG	intergênica		4,96
AX-168350567	chr1	36685502	mcG	intergênica		4,95
AX-168349565	chr2	9004542	mcG	Cc02_g11040	amino acid transporter AVT1C-like	4,95
AX-168344672	chr2	38979591	mcG	intergênica		4,76
AX-168356859	chr2	48929062	mcG	Cc02_g35240	Histone-lysine N-methyltransferase	4,82
AX-168357936	chr5	25664762	mcG	Cc05_g11570	reticulon-4-interacting protein 1, mitochondrial-like	5,19
AX-168344320	chr5	26775484	mcG	Cc05_g13080	scarecrow-like protein 21	4,89
AX-168354263	chr5	27091604	mcG	Cc05_g13540	serine/threonine-protein kinase PBL23	5,63
AX-168304428	chr5	27249415	mcG	Cc05_g13760	polyadenylate-binding protein RBP45-like isoform X1	4,74
AX-168354235	chr5	27259421	mcG	Cc05_g13780	purple acid phosphatase 18	5,64
AX-168349220	chr5	27265924	mcG	intergênica		4,74
AX-168357884	chr5	27274821	mcG	Cc05_g13830	protein AMN1 homolog	4,71
AX-168350422	chr5	27288885	mcG	intergênica		4,71
AX-168352527	chr5	27320714	mcG	Cc05_g13910	autophagy-related protein 8C-like	4,70
AX-168354224	chr5	27359501	mcG			4,71
AX-168354156	chr5	27361703	mcG	<b>Cc05_g13960</b>	<b>DNA mismatch repair protein MSH5-like isoform X4</b>	4,70
AX-168307263	chr5	27425644	mcG	Cc05_g14000	splicing factor, suppressor of white-apricot homolog isoform X2	4,72
AX-168357204	chr5	27488185	mcG	intergênica		4,71
AX-168344132	chr5	27514424	mcG	Cc05_g14150	uncharacterized protein LOC113691025	4,83
AX-168350174	chr5	27600665	mcG	Cc05_g14300	trigger factor-like protein TIG, Chloroplastic isoform X1	4,71
AX-168314683	chr5	27620400	mcG	intergênica		4,69
AX-168300545	chr5	27649534	mcG	Cc05_g14360	putative BPI/LBP family protein At1g04970 isoform X1	4,71
AX-168301998	chr5	27657009	mcG	Cc05_g14380	serine/threonine-protein phosphatase PP-X isozyme 2	4,70
AX-168302797	chr5	27682900	mcG	Cc05_g14410	cell division protein FtsZ homolog 1, chloroplastic-like	4,71
AX-168302552	chr5	27718368	mcG	Cc05_g14470	protein ENHANCED DOWNY MILDEW 2-like	5,17
AX-168344030	chr5	27743513	mcG	Cc05_g14490	protein TRI1	4,71
AX-168304702	chr5	27749889	mcG	Cc05_g14500	sodium/hydrogen exchanger 4-like	5,26
AX-168344218	chr5	27816294	mcG	Cc05_g14590	uncharacterized protein LOC113689014	4,71
AX-168343893	chr5	27898571	mcG	Cc05_g14760	putative DUF21 domain-containing protein At3g13070	4,71
AX-168344096	chr5	28596778	mcG	Cc05_g15820	F-box protein At-B	4,69
AX-168359296	chr5	28705373	mcG	Cc05_g15970	protein RRNAD1-like isoform X3	5,01
AX-168301111	chr0	25089844	mcP	intergênica		11,21
AX-168358131	chr0	25149927	mcP	Cc00_g03450	ABC transporter B family member 25-like isoform X2	4,60
AX-168342522	chr0	35764294	mcP	intergênica		6,14
AX-168313058	chr0	62217631	mcP	Cc00_g07670	probable ubiquitin-like-specific protease 2B isoform X2	5,70
AX-168353798	chr1	34267496	mcP	Cc01_g16730	membrane protein of ER body 2-like	4,31
AX-168350995	chr1	34299159	mcP	Cc01_g16760	oleoyl-acyl carrier protein thioesterase, chloroplastic-like	4,32
AX-168349367	chr1	36278881	mcP	intergênica		4,12
AX-168348734	chr2	7138427	mcP	Cc02_g08840	cysteine desulfurase 1, chloroplastic-like isoform X1	4,34
AX-168355324	chr2	7588600	mcP	Cc02_g09340	Flavonol 4'-sulfotransferase, putative	4,32
AX-168311082	chr2	7694431	mcP	intergênica		4,22
AX-168309352	chr2	7802958	mcP	Cc02_g09610	ribonuclease H2 subunit A-like	5,41
AX-168344634	chr2	7956595	mcP	intergênica		4,44
AX-168344537	chr2	7957761	mcP	Cc02_g09820	R3H domain-containing protein 2-like	4,44
AX-168303613	chr2	19739507	mcP	Cc02_g22310	U3 small nucleolar RNA-associated protein 18 homolog	5,10
AX-168300036	chr2	19795491	mcP			5,12
AX-168349520	chr2	19796373	mcP	<b>Cc02_g22420</b>	<b>protein RETICULATA-RELATED 1, chloroplastic</b>	5,31
AX-168355983	chr3	74856	mcP	intergênica		6,34
AX-168311212	chr6	3280700	mcP	Cc06_g04140	cell division cycle protein 48 homolog	4,11
AX-168343126	chr7	3646570	mcP	Cc07_g05180	lysosomal amino acid transporter 1 homolog	4,50
AX-168341219	chr7	3651082	mcP			4,78
AX-168341437	chr7	3651136	mcP	<b>Cc07_g05190</b>	<b>protein YeeZ</b>	4,72
AX-168349133	chr7	3651853	mcP			4,85
AX-168342562	chr10	12414950	mcP	intergênica		5,72
AX-168355517	chr10	18290603	mcP	intergênica		5,78

Legenda: Identificação dos marcadores associados, utilizando o TASSEL, para as características de tamanho e formato de grãos, especificamente os 50 marcadores mais significativos, de um total de 485, para moca graúdo, indicando o cromossomo (Chr), posição, gene, proteína e log10 (pvalor).

Comparando as análises de GWAS realizadas no *software* TASSEL e no rrBLUP, com os mesmos dados utilizados em ambas as análises, diferentes resultados foram obtidos e o TASSEL se mostrou mais eficiente, identificando um número muito maior de marcadores associados (603 SNPs) quando se compara as duas análises realizadas pelo rrBLUP, com P3D e sem P3D, que somente identificou 13 e 28 SNPs, respectivamente. As análises com o TASSEL também identificaram associação para outras características diferentes daquelas obtidas pelo rrBLUP, como por exemplo produção, boia, eixo menor e chato graúdo.

A partir dos resultados de GWAS obtidos notou-se que diversas marcas estavam associadas a diferentes fenótipos, como por exemplo, os marcadores AX-168308807 e AX-168355686 que no rrBLUP associou-se a peso de 100 grãos e no TASSEL associou-se para peso de 100 como também para Boia. O mesmo ocorreu com o marcador AX-168342522 associado para peso de 100 e moca miúdo (rrBLUP e TASSEL) e para chato miúdo e moca graúdo (TASSEL). Nas análises do rrBLUP o marcador AX-168355983 associou-se a moca miúdo e para as análises do TASSEL, além de moca miúdo, ele se associou as características de chato miúdo e moca graúdo.

Ao comparar os resultados obtidos com a inclusão de indivíduos, houve um aumento de marcas associadas na GWAS realizada no *software* TASSEL, 147 associações para os 366 indivíduos e 603 para os 1.319 indivíduos de *C. canephora*. No rrBLUP, entretanto, notou-se uma redução de marcas associadas ao aumentar o número de indivíduos, sendo 54 SNPs nas análises realizadas com os 366 indivíduos e 13 SNPs para os 1.319 indivíduos. Este fato, pode estar relacionado à diferente distribuição dos dados fenotípicos, ao se variar o número de indivíduos, proporcionando uma maior detecção de marcadores associados.

Em GWAS diversas correções de teste para o valor limite do teste de significância podem ser aplicadas, os resultados de associação apresentados nesse trabalho levaram em consideração a correção de FDR, contudo, em todos os gráficos Manhattan apresentados observa-se a correção de Bonferroni, e nota-se que esta é muito mais conservadora, portanto muitos loci importantes podem não passar no critério rigoroso do teste de significância (WANG et al., 2016).



Apesar de alguns resultados diferentes, no exercício de se comparar diferentes programas estatísticos (TASSEL *vs.* rrBLUP), diferente número de indivíduos (366 *vs.* 1.319), assim como número de SNPs (Affymetrix *vs.* nextRAD) e, também, opções metodológicas (sem p3D *vs.* p3D+) os resultados obtidos neste trabalho, várias vezes confirmados por uma ou mais, das análises realizadas, corroboram com sua robustez e significância, considerando-se ou não a estrita correção de Bonferroni, em muitos casos, neste trabalho, com seu limite superado. Neste sentido, na Tabela 20, estão apresentados os marcadores associados com as respectivas características estudadas, encontrados neste trabalho. A Tabela 20, apresenta uma lista sem redundância de todos os marcadores encontrados em associação, com uma classificação de reprodutibilidade das diversas análises realizadas neste estudo.

Alguns destaques devem ser mencionados, como por exemplo, o marcador AX-168342522 localizado numa região intergênica do cromossomo 0, se apresentou em associação com 3 características nas análises realizadas com as 1.319 plantas. Nas análises com rrBLUP (sem p3D e p3D+), para as características de Moca Miúdo e p100. Já com as análises utilizando-se o TASSEL, foi também encontrada associação com Chato Miúdo, além de Moca Miúdo e p100. Outro destaque é o marcador AX-168343647, localizado no gene Cc07\_g01330 (“*protein COFACTOR ASSEMBLY OF COMPLEX C SUBUNIT B CCBI*”), no cromossomo 7, que só não esteve associado com a característica de Chato miúdo, em duas das análises realizadas (rrBLUP/366 e rrBLUP/1.319). Em todas as outras análises, associação significativa foi detectada. O mesmo aconteceu para outros dois marcadores AX-168348505 (Cc06\_g03280-“*uncharacterized protein LOC113694498*”) e AX-168357173 (Cc06\_g11810-“*protein EARLY FLOWERING 5-like*”), ambos também associados significativamente com a característica Chato miúdo. Outro destaque a ser mencionado, diz respeito ao gene Cc03\_g12920 (“*serine/threonine-protein kinase-like protein ACR4*”), pois foi o único gene a ter 3 SNPs associados, com a característica de Chato miúdo, detectado em mais de uma das análises realizadas. Também tiveram 3 SNPs associados, porém, neste caso em somente uma das análises realizadas, os genes Cc03\_g15080 (“*altered inheritance rate of mitochondria protein 25*”), o gene Cc07\_g05190 (“*protein YeeZ*”) e o Cc06\_g05640 (“*CAAX prenyl protease 2*”), sendo os dois primeiros em associação com Moca miúdo e o último, com Moca gráudo.

Outros 13 genes, tiveram associação com 2 marcadores SNPs, consistentes com mais de uma das diferentes análises realizadas. Associados com a característica Chato

miúdo, Cc02\_g12670-“*glycine-rich cell wall structural protein-like*” (AX-168346346 e AX-168347111), Cc04\_g16330-“*thioredoxin-like 1-2, chloroplastic*” (AX-168311865 e AX-168346035), Cc06\_g11920-“*mechanosensitive ion channel protein 10-like*” (AX-168298126 e AX-168354529), Cc06\_g12050-“*uncharacterized protein LOC113696977 isoform XI*” (AX-168354238 e AX-168341207), Cc06\_g17650-“*ABC transporter A family member 1*” (AX-168357997 e AX-168353697), Cc10\_g10700-“*uncharacterized protein LOC113714577 isoform XI*” (AX-168310468 e AX-168355206). Já para a característica Moca Graúdo, associação significativa com mais de um SNP localizado no mesmo gene, foram detectadas para os genes Cc02\_g33440-“*uncharacterized protein LOC113727614 isoform XI*” (AX-168307761 e AX-168354236), Cc02\_g34510-“*phospholipase D zeta 1-like isoform XI*” (AX-168349364 e AX-168350209), Cc02\_g34530-“*topless-related protein 3-like isoform XI*” (AX-168343107 e AX-168300389), Cc06\_g05330-“*trihelix transcription factor PTL-like*” (AX-168300636 e AX-168305381), Cc06\_g05430-“*ABC transporter G family member 39-like isoform X2*” (AX-168357947 e AX-168301070) e o gene Cc06\_g05400-“*1,4-alpha-glucan-branching enzyme 2-2*” (AX-168357178 e AX-168371811). Por fim, o gene Cc06\_g05270-“*polyadenylation and cleavage factor homolog 4-like*”, teve um SNP associado com a característica Chato miúdo (AX-168344460) e outro marcador, AX-168300636 associado com a característica Moca graúdo. Outros 09 genes também tiveram mais de uma SNP associado significativamente, porém em somente uma das análises realizadas. Todos esses resultados estão sumarizados na Tabela 20.

No total, foram 759 marcadores SNPs encontrados significativamente em associação com uma das características estudadas. Certamente, a característica Moca graúdo foi a que apresentou o maior número de SNPs associados (502 SNPs), pois somente na análise realizada com o TASSEL/1.319, foram detectados 485 SNPs em associação com essa característica. Destes 485 SNPs, somente os 50 com menores p-valores estão apresentados na Tabela 20, totalizando 143 SNPs não redundantes. Em seguida, a característica Chato miúdo apresentou 127 SNPs associados, Moca miúdo (31 SNPs), eixo menor do fruto (21 SNPs), p100 (10 SNPs) e boia (8 SNPs). Por fim, a característica potencial hídrico de antemanhã (4 SNPs), Chato médio (3 SNPs) e Chato graúdo, produções dos anos de 2012 (p12) e 2013 (p13), com um SNP cada (Tabela 20).

Do total de 350 SNPs apresentados na Tabela 20, 180 SNPs estão localizados em regiões gênicas e 170 SNPs em regiões intergênicas. Entretanto, vale ressaltar que a maioria desses 170 SNPs (intergênicos) estão localizados nas proximidades dos genes

encontrados, muito provavelmente na região promotora ou terminadora desses mesmos genes, podendo estar diretamente ligados à função dos mesmos.

Tabela 19 – Marcadores SNPs identificados como associados para as diferentes características testadas, comparando-se os programas estatísticos (TASSEL vs. rrBLUP), os diferentes métodos (com p3D vs. Sem P3D), os número de indivíduos (366 vs. 1.319) e os número de marcadores (Affymetrix vs. nextRAD). (Continua...)

Marcador	Análise	SNP	Chr	Posição	Fenótipo	Gene	Proteína
AX-168351582	***		chr0	86833249	mcG	Cc00_g10480	autophagy-related protein 18a-like isoform X1
AX-168358000	***		chr0	86842748	mcG	Cc00_g10490	protein SRC2 homolog
AX-168350302	**		chr0	119869013	mcG	intergênica	
AX-168351974			chr0	128965321	mcG	intergênica	
AX-168303844			chr0	128971355	mcG	intergênica	
AX-168304865			chr0	128973523	mcG	intergênica	
AX-168344491			chr0	128974922	mcG	intergênica	
AX-168312126			chr0	128977308	mcG	intergênica	
AX-168358401	***		chr0	178020530	mcG	Cc00_g28760	probable pectinesterase/pectinesterase inhibitor 58
AX-168349493	***		chr0	178021959	mcG	intergênica	
S1_1107596	*		chr1	1107595	mcG	intergênica	
S1_1107611	*		chr1	1107610	mcG	intergênica	
AX-168350800			chr1	35916313	mcG	Cc01_g18940	beta-carotene isomerase D27, chloroplastic-like
AX-168350278			chr1	35954271	mcG	intergênica	
AX-168344238			chr1	35965818	mcG	Cc01_g18990	protein SPIRRIG-like isoform X1
AX-168349254			chr1	36047250	mcG	intergênica	
AX-168304760	B		chr1	36117292	mcG	<b>Cc01_g19150</b>	<b>homocysteine S-methyltransferase 2-like</b>
AX-168349367			chr1	36136097	mcG	intergênica	
AX-168344419	B		chr1	36154034	mcG	<b>Cc01_g19250</b>	<b>non-specific lipid-transfer protein-like protein At2g13820</b>
AX-168358897			chr1	36163621	mcG	Cc01_g19260	F-box/kelch-repeat protein SKIP30-like
AX-168371659			chr1	36330604	mcG	intergênica	
AX-168358332			chr1	36398376	mcG	intergênica	
AX-168350068			chr1	36423435	mcG	Cc01_g19640	basic 7S globulin-like
AX-168371779			chr1	36485520	mcG	Cc01_g19730	rab11 family-interacting protein 3 isoform X3
AX-168350567			chr1	36643242	mcG	intergênica	
AX-168354675			chr1	36685502	mcG	intergênica	
AX-168359459			chr2	9004542	mcG	Cc02_g11040	amino acid transporter AVT1C-like
S2_20759823	*		chr2	20759822	mcG	Cc02_g23620	cullin-associated NEDD8-dissociated protein 1
AX-168342448			chr2	38979591	mcG	intergênica	
AX-168307233	**		chr2	43845078	mcG	Cc02_g32570	piezo-type mechanosensitive ion channel homolog
AX-168345527	***		chr2	45306957	mcG	intergênica	
AX-168308075	***		chr2	45584774	mcG	Cc02_g33260	cyclin-H1-1-like
AX-168307761	***	B	chr2	45870306	mcG	<b>Cc02_g33440</b>	<b>uncharacterized protein LOC113727614 isoform X1</b>
AX-168344465	***		chr2	45965255	mcG	intergênica	
AX-168350938	***		chr2	46680708	mcG	Cc02_g33780	probable linoleate 9S-lipoxygenase 5
AX-168306591	***		chr2	46700921	mcG	Cc02_g33800	probable linoleate 9S-lipoxygenase 5
AX-168303111	***		chr2	47249367	mcG	Cc02_g34000	transcription factor bHLH18-like
AX-168359111	***		chr2	47305915	mcG	Cc02_g34030	stress-related protein-like
AX-168352813	***		chr2	47499197	mcG	Cc02_g34140	uncharacterized protein LOC113727677 /Taffazzin
AX-168314271	***		chr2	47659533	mcG	intergênica	
AX-168350690	***		chr2	47691038	mcG	Cc02_g34320	serine/threonine-protein phosphatase PP1 isozyme 9
AX-168345203	***		chr2	47801512	mcG	intergênica	
AX-168349364	***	B	chr2	47922088	mcG	<b>Cc02_g34510</b>	<b>phospholipase D zeta 1-like isoform X1</b>
AX-168343107	***	B	chr2	47984747	mcG	<b>Cc02_g34530</b>	<b>topless-related protein 3-like isoform X1</b>
AX-168344012	***		chr2	48014671	mcG	intergênica	
AX-168350907	***		chr2	48024076	mcG	intergênica	
AX-168356150	***		chr2	48051681	mcG	Cc02_g34560	probable protein phosphatase 2C 5
AX-168352613	***		chr2	48102576	mcG	Cc02_g34580	uncharacterized protein LOC11372756 isoform X1
AX-168299845	***		chr2	48202193	mcG	Cc02_g34690	calmodulin-binding protein 60 D-like
AX-168300212	***		chr2	48217852	mcG	intergênica	
AX-168304747	***		chr2	48319349	mcG	intergênica	
AX-168352019	*		chr2	48356484	mcG	Cc02_g34790	uncharacterized protein LOC113727770 isoform X1
AX-168351290	**		chr2	48391519	mcG	Cc02_g34840	pentatricopeptide repeat-containing protein At1g09190-like
AX-168353968	***		chr2	48527652	mcG	Cc02_g35000	lon protease homolog 1, mitochondrial-like
AX-168353529			chr2	48592726	mcG	Cc02_g35040	nicotianamine synthase-like
AX-168355301			chr2	48609755	mcG	intergênica	
AX-168356859			chr2	48642133	mcG	Cc02_g35070	RNA polymerase II-associated protein 3-like isoform X1
AX-168353797			chr2	48655516	mcG	intergênica	
AX-168306176	***		chr2	48714647	mcG	intergênica	
AX-168308758			chr2	48728000	mcG	intergênica	
AX-168352993			chr2	48794085	mcG	intergênica	
AX-168352190			chr2	48929062	mcG	Cc02_g35240	Histone-lysine N-methyltransferase
AX-168357531			chr2	49435102	mcG	intergênica	
AX-168348341			chr2	49495819	mcG	intergênica	
AX-168357029			chr2	49497297	mcG	intergênica	
AX-168359149			chr2	49582983	mcG	intergênica	
AX-168352761			chr2	49635690	mcG	Cc02_g35680	pleiotropic drug resistance protein 3-like
AX-168350759			chr2	49685638	mcG	Cc02_g35700	pleiotropic drug resistance protein 3-like
AX-168312699			chr2	49869023	mcG	intergênica	
S3_838902			chr2	52458925	mcG	Cc02_g37620	T-complex protein 1 subunit theta-like
AX-168343237	*		chr2	53259323	mcG	intergênica	
AX-168305094	**		chr2	53747500	mcG	intergênica	

Tabela 20 – Marcadores SNPs identificados como associados para as diferentes características testadas, comparando-se os programas estatísticos (TASSEL vs. rrBLUP), os diferentes métodos (com p3D vs. Sem P3D), os número de indivíduos (366 vs. 1.319) e os número de marcadores (Affymetrix vs. nextRAD). (Continua...)

Marcador	Análise	SNP	Chr	Posição	Fenótipo	Gene	Proteína
AX-168306839			chr5	25357499	mcG	intergênica	
AX-168310386			chr5	25599111	mcG	Cc05_g11480	glutamate receptor 2.7-like
AX-168357936	*		chr5	25664762	mcG	Cc05_g11570	reticulon-4-interacting protein 1, mitochondrial-like
AX-168346321			chr5	25969191	mcG	Cc05_g12010	protein SPA1-RELATED 3 isoform X2
AX-168344359			chr5	26207976	mcG	intergênica	
AX-168309767			chr5	26401552	mcG	intergênica	
AX-168346202	*		chr5	26581667	mcG	intergênica	
AX-168344320			chr5	26690962	mcG	Cc05_g12960	G-type lectin S-receptor-like serine/threonine-protein kinase At4g27290
AX-168356219			chr5	26775484	mcG	Cc05_g13080	scarecrow-like protein 21
AX-168354263	*		chr5	27091604	mcG	Cc05_g13540	serine/threonine-protein kinase PBL23
AX-168357884			chr5	27249415	mcG	Cc05_g13760	polyadenylate-binding protein RBP45-like isoform X1
AX-168354235	*		chr5	27259421	mcG	Cc05_g13780	purple acid phosphatase 18
AX-168350422			chr5	27265924	mcG	intergênica	
AX-168352527			chr5	27274821	mcG	Cc05_g13830	protein AMN1 homolog
AX-168371903			chr5	27288885	mcG	intergênica	
AX-168357204			chr5	27299271	mcG	intergênica	
AX-168354224			chr5	27320714	mcG	Cc05_g13910	autophagy-related protein 8C-like
AX-168354156		B	chr5	27359501	mcG	<b>Cc05_g13960</b>	<b>DNA mismatch repair protein MSH5-like isoform X4</b>
AX-168344132			chr5	27425644	mcG	Cc05_g14000	splicing factor, suppressor of white-apricot homolog isoform X2
AX-168314683			chr5	27488185	mcG	intergênica	
AX-168357408			chr5	27514424	mcG	Cc05_g14150	uncharacterized protein LOC113691025
AX-168350174			chr5	27563621	mcG	Cc05_g14220	ethylene-responsive transcription factor 3-like
AX-168300545			chr5	27600665	mcG	Cc05_g14300	trigger factor-like protein TIG, Chloroplastic isoform X1
S5_27752781			chr5	27620400	mcG	intergênica	
AX-168301998			chr5	27649534	mcG	Cc05_g14360	putative BPI/LBP family protein At1g04970 isoform X1
AX-168302797			chr5	27657009	mcG	Cc05_g14380	serine/threonine-protein phosphatase PP-X isozyme 2
AX-168302552			chr5	27682900	mcG	Cc05_g14410	cell division protein FtsZ homolog 1, chloroplastic-like
AX-168344030			chr5	27718368	mcG	Cc05_g14470	protein ENHANCED DOWNY MILDEW 2-like
AX-168304702			chr5	27743513	mcG	Cc05_g14490	protein TRI1
AX-168344218			chr5	27749869	mcG	Cc05_g14500	sodium/hydrogen exchanger 4-like
AX-168371643			chr5	27752780	mcG	intergênica	
AX-168343893			chr5	27816294	mcG	Cc05_g14590	uncharacterized protein LOC113689014
AX-168301499			chr5	27839867	mcG	intergênica	
AX-168342644			chr5	27898571	mcG	Cc05_g14760	putative DUF21 domain-containing protein At3g13070, chloroplastic isoform X1
AX-168350481			chr5	28021680	mcG	Cc05_g14940	fasciclin-like arabinogalactan protein 1
AX-168343448			chr5	28114364	mcG	Cc05_g15070	Thioredoxin-like fold protein
AX-168353973			chr5	28347586	mcG	Cc05_g15500	enhancer of mRNA-decapping protein 4-like isoform X1
AX-168354345			chr5	28428018	mcG	intergênica	
AX-168344096			chr5	28563082	mcG	Cc05_g15780	chloride channel protein CLC-f isoform X1
AX-168359296			chr5	28596778	mcG	Cc05_g05270	F-box protein At-B
AX-168358243			chr5	28705373	mcG	Cc05_g15970	protein RRNAD1-like isoform X3
AX-168344507			chr5	28859558	mcG	Cc05_g16200	molybdenum cofactor sulfurase-like isoform X1
AX-168300636	***	B	chr6	4165722	mcG	Cc06_g05330	polyadenylation and cleavage factor homolog 4-like isoform X1
AX-168305381	***	B	chr6	4213885	mcG	<b>Cc06_g05330</b>	<b>trihelix transcription factor PTL-like</b>
AX-168303541	***		chr6	4224625	mcG	intergênica	
AX-168344569	***		chr6	4232666	mcG	intergênica	
AX-168345105	***		chr6	4245430	mcG	Cc06_g05340	ethylene-responsive transcription factor ERF024-like
AX-168300402	***		chr6	4246582	mcG	intergênica	
AX-168300079	***		chr6	4247807	mcG	intergênica	
AX-168357236	***		chr6	4256203	mcG	Cc06_g05350	UPF0481 protein At3g47200-like
AX-168358192	***		chr6	4258781	mcG	Cc02_g05380	tyrosine N-monoxygenase-like
AX-168357178	***	B	chr6	4314349	mcG	<b>Cc06_g05400</b>	<b>1,4-alpha-glucan-branching enzyme 2-2, chloroplastic/amyloplastic-like isoform X1</b>
AX-168301070	***	B	chr6	4334744	mcG	<b>Cc06_g05430</b>	<b>ABC transporter G family member 39-like isoform X2</b>
AX-168342602	***		chr6	4375659	mcG	Cc06_g05480	serine/threonine-protein kinase D6PKL2-like
AX-168351134			chr6	4447830	mcG	intergênica	
AX-168304804			chr6	4454535	mcG	intergênica	
AX-168299990			chr6	4467803	mcG	Cc06_g05550	GDSL esterase/lipase At5g03610-like
AX-168349250		A	chr6	4531144	mcG	<b>Cc06_g05640</b>	<b>CAAX prenyl protease 2</b>
AX-168351225			chr6	4569946	mcG	intergênica	
AX-168343288			chr6	4572473	mcG	intergênica	
AX-168354124			chr6	4590621	mcG	Cc06_g05710	protein BIC1-like
AX-168351028			chr6	4696954	mcG	Cc06_g05910	F-box/kelch-repeat protein At1g57790-like
AX-168303538			chr6	4720389	mcG	Cc06_g05960	uncharacterized protein LOC113692255
AX-168352371			chr6	4738239	mcG	Cc06_g05990	uncharacterized protein LOC113691749
AX-168306989			chr6	4790340	mcG	Cc06_g06080	beta-fructofuranosidase, insoluble isoenzyme 1-like isoform X1
AX-168343151			chr6	4867384	mcG	Cc06_g06170	peptidyl-prolyl cis-trans isomerase CYP18-2-like
S6_7033632			chr6	4890314	mcG	Cc06_g06220	probable F-box protein At2g36090
AX-168303120	**		chr6	4922381	mcG	Cc06_g06250	uncharacterized protein LOC113691887
AX-168305682			chr8	25372477	mcG	intergênica	
AX-168357486	***		chr11	14383294	mcG	Cc11_g03570	glutamate receptor 1.3-like
AX-168305338	***		chr11	14385622	mcG	intergênica	

Tabela 21 – Marcadores SNPs identificados como associados para as diferentes características testadas, comparando-se os programas estatísticos (TASSEL vs. rrBLUP), os diferentes métodos (com p3D vs. Sem P3D), os número de indivíduos (366 vs. 1.319) e os número de marcadores (Affymetrix vs. nextRAD). (Continua...)

Marcador	Análise	SNP	Chr	Posição	Fenótipo	Gene	Proteína
S0_22957200	**		chr0	22957199	chP	intergênica	
AX-168351301	*		chr0	50327373	chP	intergênica	
S0_111609941			chr0	52042011	chP	intergênica	
S0_90384806	**		chr0	90384805	chP	intergênica	
AX-168311657	*		chr0	1,06E+08	chP	Cc00_g14940	disease resistance protein RPM1-like
AX-168352499			chr0	1,28E+08	chP	intergênica	
AX-168351351			chr0	1,29E+08	chP	Cc00_g20120	AAA-ATPase ASD, mitochondrial-like
AX-168308959	*		chr0	1,33E+08	chP	intergênica	
AX-168371943			chr0	1,39E+08	chP	Cc00_g21970	MDIS1-interacting receptor like kinase 2-like
S0_150055733			chr0	1,45E+08	chP	intergênica	
AX-168358258			chr0	1,51E+08	chP	intergênica	
AX-168345100	*		chr0	1,62E+08	chP	intergênica	
AX-168372309	**		chr0	1,84E+08	chP	intergênica	
AX-168353870			chr0	2,03E+08	chP	Cc00_g35180	mannosyl-oligosaccharide 1,2-alpha-mannosidase MNS1-like
AX-168299146	***		chr1	1126012	chP	Cc01_g00640	IQ domain-containing protein IQM2-like
AX-168296877			chr1	1190307	chP	intergênica	
S1_8001113			chr1	1634705	chP	intergênica	
AX-168299745			chr1	21662434	chP	intergênica	
S1_23113352			chr1	23113351	chP	Cc01_g06230	uncharacterized protein LOC113720213 isoform X1
AX-168353642		B	chr1	24284629	chP	<b>Cc01_g06920</b>	<b>La-related protein 1A</b>
AX-168353798			chr1	28904683	chP	Cc01_g10220	pheophorbide a oxygenase, chloroplastic-like
AX-168305324	*		chr2	1408748	chP	Cc02_g01760	cyclic nucleotide-gated ion channel 1-like
AX-168371724	*		chr2	1411894	chP	intergênica	
AX-168343597			chr2	1464001	chP	Cc02_g01810	ATP-dependent zinc metalloprotease FTSH 11, chloroplastic/mitochondrial-like
AX-168348734	***		chr2	7138427	chP	Cc02_g08840	cysteine desulfurase 1, chloroplastic-like isoform X1
AX-168349189		B	chr2	7551514	chP	<b>Cc02_g09290</b>	<b>7-methylguanosine phosphate-specific 5'-nucleotidase A-like</b>
AX-168355324	***		chr2	7588600	chP	Cc02_g09340	Flavonol 4'-sulfotransferase, putative
AX-168309352	*		chr2	7802958	chP	Cc02_g09610	ribonuclease H2 subunit A-like
AX-168312035			chr2	7806809	chP	intergênica	
AX-168344634	*		chr2	7956595	chP	intergênica	
AX-168344537	*		chr2	7957761	chP	Cc02_g09820	R3H domain-containing protein 2-like
AX-168359112			chr2	8063368	chP	intergênica	
AX-168349565			chr2	8528963	chP	Cc02_g10460	cytochrome P450 703A2-like
S2_21294997			chr2	9959463	chP	intergênica	
AX-168307006	**		chr2	10336701	chP	Cc02_g12290	pre-mRNA-splicing factor SYF1-like isoform X1
AX-168346346	***	B	chr2	10767477	chP	<b>Cc02_g12670</b>	<b>glycine-rich cell wall structural protein-like</b>
AX-168297476	*		chr2	10787245	chP	Cc02_g12690	uncharacterized protein LOC113730922
AX-168346721	***		chr2	11752422	chP	Cc02_g13600	squamosa promoter-binding protein 1
AX-168303613			chr2	15011257	chP	Cc02_g16230	beta-galactosidase 13-like
AX-168352736	***		chr2	31001072	chP	intergênica	
S2_42640423	**		chr2	42640422	chP	intergênica	
S3_18778694			chr3	838901	chP	Cc03_g01110	uncharacterized protein LOC113733901
AX-168345166	*		chr3	7872046	chP	intergênica	
AX-168344168	*		chr3	16412991	chP	intergênica	
AX-168300945	*		chr3	16547070	chP	intergênica	
AX-168348359	*		chr3	16590846	chP	intergênica	
AX-168345653	*		chr3	22992321	chP	intergênica	
AX-168305683	*		chr3	23071449	chP	intergênica	
AX-168303383	*		chr3	23072264	chP	intergênica	
AX-168343864	*	A	chr3	23073093	chP	<b>Cc03_g12920</b>	<b>serine/threonine-protein kinase-like protein ACR4 isoform X1</b>
AX-168297140	*		chr3	23226970	chP	intergênica	
AX-168349013	*		chr3	23231432	chP	Cc03_g12950	eIF-2-alpha kinase GCN2 isoform X1
AX-168297497	*		chr3	23238786	chP	intergênica	
AX-168297908	*		chr3	23240174	chP	intergênica	
AX-168297274	*		chr3	23410454	chP	intergênica	
AX-168308219	*		chr3	23440945	chP	intergênica	
AX-168298456	*		chr3	24286750	chP	intergênica	
AX-168300300	*		chr3	25246170	chP	intergênica	
AX-168346334	*		chr3	25809464	chP	intergênica	
AX-168310426	*		chr3	25819170	chP	Cc03_g13390	pentatricopeptide repeat-containing protein At1g06710, mitochondrial-like



Tabela 22 – Marcadores SNPs identificados como associados para as diferentes características testadas, comparando-se os programas estatísticos (TASSEL vs. rrBLUP), os diferentes métodos (com p3D vs. Sem P3D), os número de indivíduos (366 vs. 1.319) e os número de marcadores (Affymetrix vs. nextRAD). (Continua...)

Marcador	Análise	SNP	Chr	Posição	Fenótipo	Gene	Proteína
AX-168358718			chr4	508476	chP	Cc04_g00690	
AX-168344856	*		chr4	1676585	chP	Cc04_g02150	dof zinc finger protein DOF3.1-like
AX-168310634	*		chr4	1708338	chP	Cc04_g02180	dentin sialophosphoprotein
AX-168313313			chr4	1742432	chP	Cc04_g02250	60S ribosomal protein L28-1-like
AX-168348260			chr4	1823968	chP	intergênica	
AX-168343159			chr4	1881580	chP	Cc04_g02490	
AX-168298549	*		chr4	4380942	chP	Cc04_g05880	uncharacterized protein LOC113738625 isoform X1
AX-168343656	*		chr4	22257865	chP	intergênica	
S4_26107964			chr4	22461703	chP	Cc04_g14730	putative disease resistance RPP13-like protein 1
AX-168311865	*	B	chr4	26419140	chP	<b>Cc04_g16330</b>	<b>thioredoxin-like 1-2, chloroplastic</b>
AX-168300621	*		chr4	26619424	chP	intergênica	
AX-168354968	*		chr4	26779757	chP	intergênica	
AX-168341050	*		chr5	790300	chP	intergênica	
AX-168353013			chr5	962752	chP	intergênica	
S5_962780	**		chr5	962779	chP	intergênica	
S5_962789	**		chr5	962788	chP	intergênica	
S5_962794	**		chr5	962793	chP	intergênica	
S5_962799	**		chr5	962798	chP	intergênica	
AX-168357314	*		chr5	8224703	chP	intergênica	
AX-168352281			chr5	21080225	chP	intergênica	
AX-168308957			chr5	21454337	chP	intergênica	
AX-168302930	*		chr5	24054345	chP	Cc05_g09570	two-pore potassium channel 3-like
AX-168357727			chr6	550354	chP	Cc06_g00670	short-chain dehydrogenase reductase 2a-like
AX-168311212			chr6	1333284	chP	Cc06_g01620	laccase-4-like
AX-168348505	****		chr6	2538355	chP	Cc06_g03280	uncharacterized protein LOC113694498
AX-168341377			chr6	5023959	chP	intergênica	
AX-168357173	****		chr6	9545143	chP	Cc06_g11810	protein EARLY FLOWERING 5-like
AX-168315406	**		chr6	9552364	chP	Cc06_g11820	WD repeat-containing protein VIP3-like
AX-168298126	*	B	chr6	9677559	chP	<b>Cc06_g11920</b>	<b>mechanosensitive ion channel protein 10-like</b>
AX-168354238	**	B	chr6	9760152	chP	<b>Cc06_g12050</b>	<b>uncharacterized protein LOC113696977 isoform X1</b>
AX-168356774	**		chr6	9770901	chP	intergênica	
AX-168343126			chr6	9982684	chP	Cc06_g12320	ATP-dependent DNA helicase Q-like 3 isoform X2
AX-168350047			chr6	14822146	chP	intergênica	
AX-168350156	*		chr6	17701432	chP	Cc06_g17600	uncharacterized protein LOC113773371
AX-168357997	**	B	chr6	17773743	chP	<b>Cc06_g17650</b>	<b>ABC transporter A family member 1</b>
AX-168343647	****		chr7	873756	chP	Cc07_g01330	rotein COFACTOR ASSEMBLY OF COMPLEX C SUBUNIT B CCB1, chloroplastic
AX-168307893	*		chr7	5156376	chP	intergênica	
AX-168349545			chr7	5407344	chP	Cc07_g07350	growth-regulating factor 1-like isoform X1
AX-168346600			chr7	5505044	chP	intergênica	
AX-168309934			chr7	5702520	chP	Cc07_g07740	tRNA-dihydrouridine(20) synthase [NAD(P)+]-like isoform X2
AX-168300213	*		chr7	5864059	chP	Cc07_g07970	MADS-box transcription factor 23-like isoform X2
AX-168354923	*		chr7	27723856	chP	intergênica	
AX-168350848	**		chr8	1803097	chP	intergênica	
AX-168342859			chr8	21391981	chP	Cc08_g07910	chorismate mutase 2-like isoform X1
AX-168350329			chr8	25748304	chP	Cc08_g10930	Golgi transport complex subunit
AX-168302843			chr8	25801832	chP	Cc08_g10990	serine/threonine-protein kinase tricorner
AX-168342455			chr8	29947471	chP	Cc08_g15090	uncharacterized protein LOC113703701
AX-168308409			chr9	6259614	chP	Cc09_g05840	potassium transporter 11-like isoform X1
AX-168349331	*		chr9	8852357	chP	intergênica	
AX-168342868	*		chr9	8880269	chP	intergênica	
AX-168357124			chr10	574254	chP	Cc10_g00650	uncharacterized protein LOC113712807
AX-168345765	**		chr10	11726758	chP	intergênica	
AX-168372147	**		chr10	11748552	chP	intergênica	
AX-168309899	**		chr10	11777153	chP	intergênica	
AX-168348689			chr10	14346651	chP	intergênica	
AX-168310468	**	B	chr10	14573473	chP	<b>Cc10_g10700</b>	<b>uncharacterized protein LOC113714577 isoform X1</b>
AX-168310228	**		chr10	14742543	chP	intergênica	
AX-168356073	**		chr10	14780965	chP	Cc10_g10760	receptor kinase-like protein Xa21
AX-168351350		B	chr10	17902648	chP	<b>Cc10_g11320</b>	<b>protein TIC 40, chloroplastic-like isoform X2</b>
AX-168355517			chr10	18050185	chP	intergênica	
AX-168354676	*		chr10	18454699	chP	Cc10_g11460	uncharacterized protein LOC113714478
AX-168311555	*		chr11	7777719	chP	intergênica	
AX-168311674	*		chr11	7913567	chP	Cc11_g02420	uncharacterized protein LOC113716855 isoform X1
AX-168354492	*		chr11	22999877	chP	Cc11_g06580	cyclin-SDS
AX-168354354			chr11	27910613	chP	Cc11_g10490	mediator of RNA polymerase II transcription subunit 25
AX-168313352			chr11	28763501	chP	Cc11_g11480	phosphatidylinositol 4-kinase gamma 7-like isoform X1
AX-168358222	*		chr11	31714783	chP	intergênica	

Tabela 23 – Marcadores SNPs identificados como associados para as diferentes características testadas, comparando-se os programas estatísticos (TASSEL vs. rrBLUP), os diferentes métodos (com p3D vs. Sem P3D), os número de indivíduos (366 vs. 1.319) e os número de marcadores (Affymetrix vs. nextRAD). (Conclusão)

Marcador	Análise	SNP	Chr	Posição	Fenótipo	Gene	Proteína
AX-168301111			chr0	14547561	Boia	intergênica	
AX-168307262			chr0	161279672	Boia	intergênica	
S0_198072686			chr0	192412669	Boia	intergênica	
AX-168311082			chr2	6272090	boia	intergênica	
AX-168312875			chr3	18778693	boia	Cc03_g12090	flowering time control protein FPA-like
AX-168358748			chr6	5939217	boia	intergênica	
AX-168358996			chr6	7033631	boia	Cc06_g08760	zinc finger CCH domain-containing protein 32-like
AX-168372376			chr10	26426206	boia	Cc10_g15510	dihydrofolate reductase isoform X1
AX-168371703			chr8	30285270	chG	Cc08_g15550	cation/H(+) antiporter 18-like isoform X1
AX-168298103	**		chr6	12791230	chM	Cc06_g14920	flavonol synthase/flavanone 3-hydroxylase-like
AX-168307159	**		chr6	12838501	chM	Cc06_g14970	leucine aminopeptidase 1-like
AX-168345461	***		chr6	13075988	chM	intergênica	
AX-168309730			chr0	22915872	emenor	Cc00_g03210	uncharacterized protein LOC113708797
AX-168358902			chr0	164357922	emenor	intergênica	
AX-168350461			chr0	187027050	emenor	intergênica	
AX-168351864			chr1	30731985	emenor	intergênica	
AX-168305336			chr2	6668157	emenor	Cc02_g08330	zinc finger protein ZAT11-like
AX-168343505		B	chr2	22040497	emenor	<b>Cc02_g24790</b>	<b>GPI-anchored protein like</b>
AX-168351650			chr2	24758035	emenor	Cc02_g27080	protein N-terminal asparagine amidohydrolase
AX-168344672			chr2	26577824	emenor	intergênica	
AX-168345962			chr5	5013195	emenor	Cc05_g01700	putative methyltransferase NSUN6
AX-168302488			chr5	5057074	emenor	intergênica	
S8_25372478			chr7	11180970	emenor	intergênica	
AX-168354649			chr7	11879643	emenor	Cc07_g15170	ras-related protein RABH1e
AX-168353406			chr9	3268799	emenor	intergênica	
AX-168349506			chr9	19961980	emenor	Cc09_g09610	Zinc transport protein like
AX-168355113			chr9	19983202	emenor	intergênica	
AX-168297043			chr9	20133028	emenor	intergênica	
AX-168302945			chr9	20245702	emenor	intergênica	
AX-168351596		B	chr10	2011076	emenor	<b>Cc10_g02630</b>	<b>putative pfkB family carbohydrate kinase</b>
AX-168309693			chr10	2048242	emenor	intergênica	
S10_10265721			chr10	2227748	emenor	intergênica	
AX-168351366			chr10	2249135	emenor	Cc10_g03010	uncharacterized protein LOC113713027 isoform X2
AX-168342372			chr0	25089844	mcP	intergênica	
AX-168313058			chr0	62217631	mcP	Cc00_g07670	probable ubiquitin-like-specific protease 2B isoform X2
AX-168350995			chr1	34267496	mcP	Cc01_g16730	membrane protein of ER body 2-like
AX-168350099			chr1	34299159	mcP	Cc01_g16760	oleoyl-acyl carrier protein thioesterase, chloroplastic-like
AX-168350197			chr1	36278881	mcP	intergênica	
AX-168348474			chr2	7694431	mcP	intergênica	
AX-168300036			chr2	19739507	mcP	Cc02_g22310	U3 small nucleolar RNA-associated protein 18 homolog
AX-168349520		B	chr2	19795491	mcP	<b>Cc02_g22420</b>	<b>protein RETICULATA-RELATED 1, chloroplastic</b>
S2_21295008			chr2	21294996	mcP	intergênica	
S2_21295021			chr2	21295007	mcP	intergênica	
AX-168308319			chr2	21295020	mcP	intergênica	
S3_24286751			chr2	50146016	mcP	intergênica	
AX-168302268			chr2	51481156	mcP	Cc02_g36670	protein cornichon homolog 4-like
AX-168297987			chr2	51499977	mcP	Cc02_g36700	ATP-dependent DNA helicase 2 subunit KU80-like isoform X1
AX-168355983	***		chr3	74856	mcP	intergênica	
S3_30315983			chr3	30300282	mcP	Cc03_g15070	BTB/POZ and TAZ domain-containing protein 3-like
AX-168351061		A	chr3	30315982	mcP	<b>Cc03_g15080</b>	<b>altered inheritance rate of mitochondria protein 25-like isoform X2</b>
AX-168358482			chr3	30319343	mcP	intergênica	
AX-168290085	***		chr4	23325422	mcP	intergênica	
S5_962753			chr4	23388423	mcP	intergênica	
AX-168354342			chr4	26107963	mcP	Cc04_g16160	TORTIFOLIA1-like protein 3
AX-168304428			chr5	26887049	mcP	Cc05_g13220	uncharacterized protein LOC113690300
AX-168349220			chr5	27043049	mcP	intergênica	
AX-168303067			chr6	3280700	mcP	Cc06_g04140	cell division cycle protein 48 homolog
AX-168308739			chr6	34672019	mcP	intergênica	
AX-168341219			chr7	3646570	mcP	Cc07_g05180	lysosomal amino acid transporter 1 homolog
AX-168341437		A	chr7	3651082	mcP	<b>Cc07_g05190</b>	<b>protein YeeZ</b>
AX-168357775			chr8	26017160	mcP	Cc08_g11270	beta-glucosidase 24-like
AX-168355753	**		chr9	83842	mcP	Cc09_g00110	E3 ubiquitin-protein ligase WAV3-like isoform X1
S10_14346652			chr10	12414950	mcP	intergênica	
AX-168309741			chr10	18290603	mcP	intergênica	
AX-168308807	***		chr0	9786333	p100	intergênica	
AX-168355686	***		chr0	9842073	p100	Cc00_g01520	uncharacterized protein LOC113743019
AX-168348258	**		chr0	10030733	p100	intergênica	
AX-168342756	**		chr0	10065719	p100	intergênica	
AX-168358131	***		chr0	25149927	p100	Cc00_g03450	ABC transporter B family member 25-like isoform X2
AX-168342522	*****		chr0	35764294	p100	intergênica	
AX-168306468			chr10	25244253	p100	intergênica	
AX-168351372	*		chr11	18451769	p100	intergênica	
AX-168348970	*		chr4	18081017	p100	Cc04_g13590	pyridoxine/pyridoxamine 5'-phosphate oxidase 1, chloroplastic-like
AX-168357068	*		chr8	27955474	p100	intergênica	
AX-168343633			chr1	8001112	p12	intergênica	
S2_6272091			chr11	30890580	p13	intergênica	
S0_127662678			chr0	111609940	pot	intergênica	
AX-168345855			chr0	150055732	pot	intergênica	
S1_1190308			chr0	198072685	pot	intergênica	
AX-168342562			chr10	10265720	pot	intergênica	

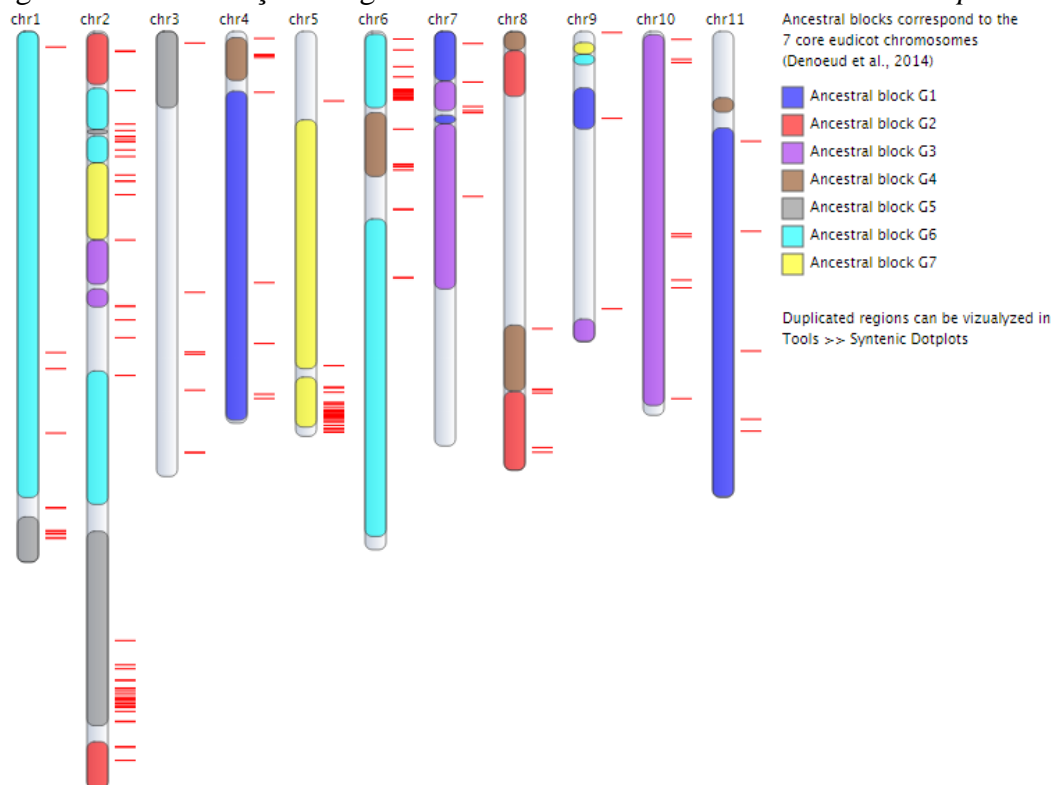
Legenda: Marcadores SNPs associados às respectivas características. Na coluna “Análise”, os marcadores associados em: (\*) duas análises, (\*\*) três análises, (\*\*\*) quatro análises, (\*\*\*\*) cinco análises. O marcador (AX-168342522), em azul, foi associado para três diferentes



características. Na coluna “SNP”, a letra (A) indica que 3 SNPs foram identificados no gene e (B) indica que 2 SNPs foram identificados no gene.

Na Figura 47, estão plotados os genes encontrados em associação neste trabalho, nas respectivas posições dos cromossomos. Observa-se claramente algumas regiões com maior intensidade de marcadores associados, principalmente nos cromossomos 2, 5 e 6, onde um maior número de traços vermelhos significa maior número de genes.

Figura 47 – Localização dos genes associados nos cromossomos de *C. canephora*.



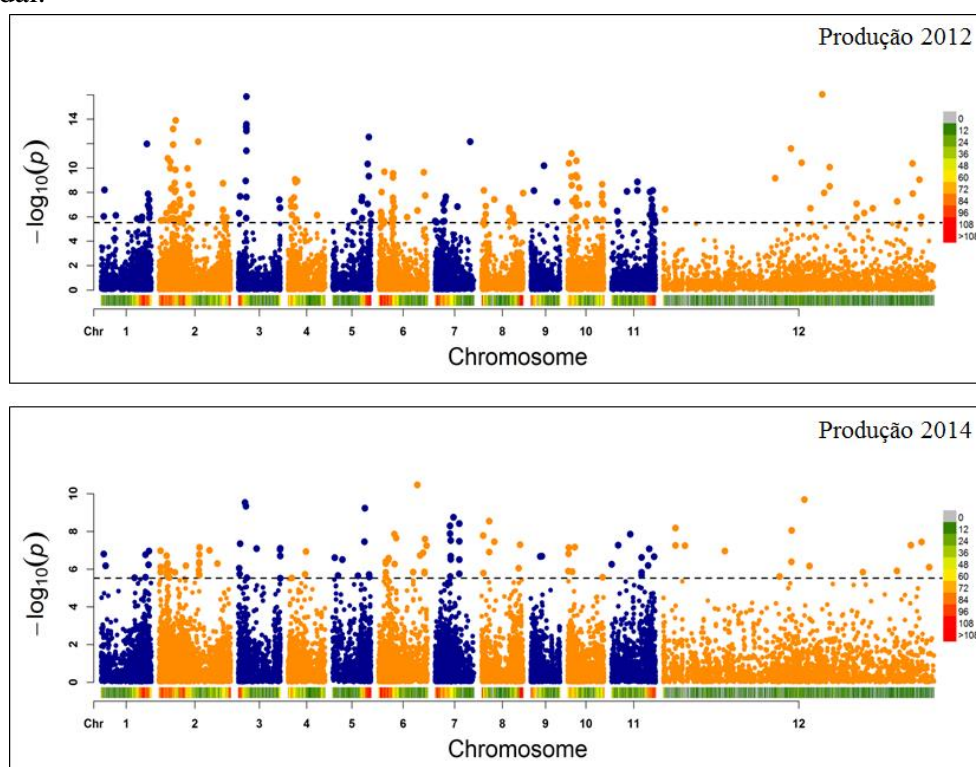
Legenda: Localização dos genes identificados como associados (traços vermelhos) nos 11 cromossomos de *C. canephora* (chr1 – chr11). Os genes localizados no cromossomo 0 não estão representados.

### 3.2.3 Modelo Bimodal

Apesar de muitas associações significativas terem sido encontradas, neste trabalho, principalmente para as características de tamanho e formato dos grãos de café, para várias outras características estudadas poucas marcas ou nenhuma associação foi encontrada com as metodologias existentes. Como por exemplo, com os dados de produção, obtidos neste trabalho. Ocorre que as análises de GWAS, pressupõem uma distribuição normal dos dados e ajustes estatísticos devem ser perseguidos para uma melhor exploração e obtenção dos resultados. No presente trabalho, tendo em vista a característica de bienalidade da produção em café, ou seja, com altas produções em um ano e baixas produções no ano seguinte, resultando numa distribuição bimodal dos dados, foi tentada uma análise de GWAS com um ajuste estatístico para tal distribuição (VIEIRA JUNIOR, dados não publicados).

Ao realizar a análise de GWAS com o modelo bimodal para a característica de produção, inúmeros marcadores foram associados com a característica para os anos de 2012 e 2014, como pode-se observar nas Figura 48 e 49.

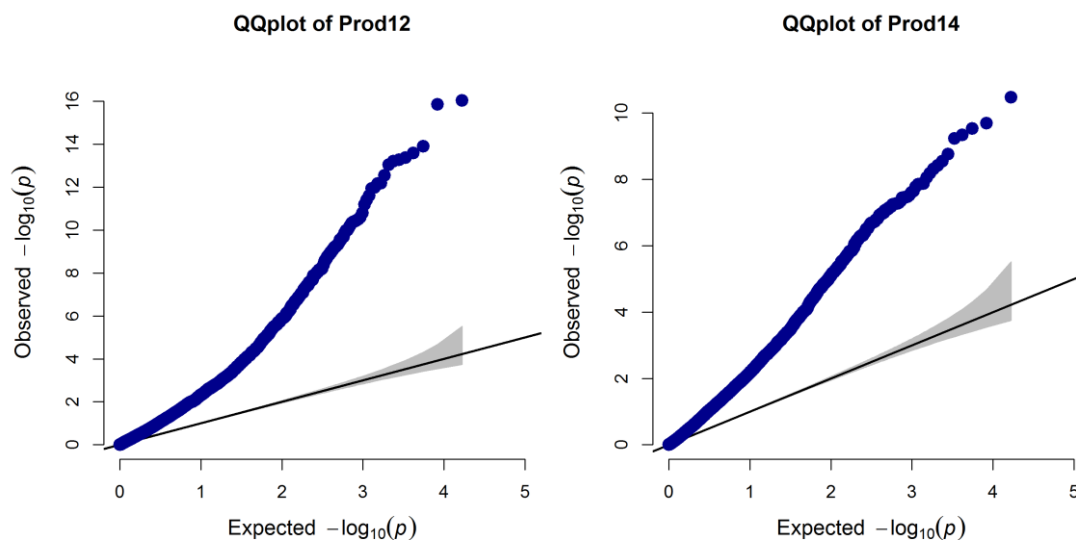
Figura 48 – Gráfico Manhattan de associação do efeito dos SNPs testados com a característica de produção, avaliada nos anos de 2012 e 2014, utilizando o modelo bimodal.



Legenda: *Manhattan plot* com os SNPs associados para a característica de produção, avaliadas nos anos de 2012 e 2014, para os 1.319 indivíduos. A linha tracejada em preto representa o

valor obtido para  $-\log_{10}(p\text{-valor})$  considerando a correção de Bonferroni (nível de significância de 95%). O cromossomo 0 de *C. canephora* está representado como cromossomo 12.

Figura 49 – Gráfico quantil-quantil (Q-Q) dos resultados de GWAS, utilizando o modelo bimodal, para produção avaliada nos anos de 2012 e 2014.



Os resultados obtidos foram significativos, com um aumento enorme do número de marcadores associados. Para os dados de produção em 2012 (p12), 2.681 marcadores SNPs foram encontrados em associação significativa. Já com os dados de produção de 2014 (p14), o número de marcadores associados foi menor, com 1.843 marcadores SNPs com associação significativa. Uma comparação dos resultados para p12 e p14, indica que 325 desses marcadores foram comuns nas duas análises realizadas.

A distribuição dos marcadores nos cromossomos, com associação significativa com a produção de 2012, mostram que a menor ocorrência de marcadores ocorreu no cromossomo 9, com 83 SNPs associados e o de maior ocorrência no cromossomo 2 (447 SNPs). Em um estudo com produção em *C. canephora* (Leroy et al, 2011) foram capazes de identificar alguns QTLs para essa característica. Após verificarmos as correspondências dos grupos de ligação obtidos naquele trabalho com o genoma de referência de *C. canephora*, pudemos concluir que os QTLs identificados para produção estão localizados nos cromossomos 11 (Grupo de ligação K), 4 (grupo de ligação D) e 9 (grupo de ligação I). Uma análise mais detalhada está sendo realizada para verificar a co-localização dos SNPs associados encontrados neste caso, com os QTLs previamente identificados, para a importante, porém complexa característica de produção em café.

Entretanto, apesar do aumento na detecção de marcadores associados, com a utilização das correções para dados com distribuição bimodal, o número exagerado encontrado, de marcadores associados, ainda sugere ajustes estatísticos. Estudos posteriores ainda são necessários para uma exploração ainda maior dos dados de genotipagem e fenotipagem, obtidos neste trabalho. Neste sentido, os avanços nos métodos estatísticos são prioritários e estão em curso.

## 4 CONCLUSÃO

Apesar de alguns resultados diferentes, no exercício de se comparar diferentes programas estatísticos (TASSEL *vs.* rrBLUP), diferente número de indivíduos (366 *vs.* 1319), assim como número de SNPs (Affymetrix *vs.* nextRAD) e, também, opções metodológicas (sem p3D *vs.* p3D+) os resultados obtidos neste trabalho, várias vezes confirmados por uma ou mais, das análises realizadas, corroboram com sua robustez e significância, considerando-se ou não a estrita correção de Bonferroni, em muitos casos, neste trabalho, com seu limite superado. Além disto, com o suporte da caracterização funcional de vários dos genes detectados como importantes fatores na determinação fenotípica das características estudadas, encontrado na literatura, principalmente com os trabalhos pioneiros na planta modelo *Arabidopsis thaliana*, nos permitem concluir que vários dos genes encontrados em associação neste estudo, serão objeto de trabalhos futuros.

Por fim, apesar do aumento na detecção de marcadores associados, com a utilização das correções para dados com distribuição bi-modal, o número exagerado encontrado, de marcadores associados, ainda sugere ajustes estatísticos. Estudos posteriores ainda são necessários para uma exploração ainda maior dos dados de genotipagem e fenotipagem, obtidos neste trabalho. Neste sentido, os avanços nos métodos estatísticos são prioritários e, certamente, estão em curso.

## REFERÊNCIAS

- AFZAL, A.J.; WOOD, A. J.; LIGHTFOOT, D. A. Plant receptor-like serine threonine kinases: roles in signaling and plant defense. **Mol Plant Microbe Interact**, v. 21, p. 507-517, 2008.
- BAGCHI, R. et al. The *Arabidopsis* ALF4 protein is a regulator of SCF E3 ligases. **The EMBO Journal**, v. 37, n. 2, p. 255-268, 2018.
- BENJAMINI, Y.; HOCHBERG, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. **Journal of the Royal Statistical Society**, v. 57, n. 1, p. 289-300, 1995.
- BRADBURY, P. J. et al. TASSEL: software for association mapping of complex traits in diverse samples. **Bioinformatics**, v. 23, n. 19, p. 2633-2635, 2007,
- CARNEIRO, F. A. et al. Phenotyping and genotyping a *Coffea canephora* population, cultivated at high altitude, aiming at a GWS program for coffee. In: THE 25TH INTERNATIONAL CONFERENCE ON COFFEE SCIENCE, 2014, Colômbia.
- DAELE, I. V. et al. A comparative study of seed yield parameters in *Arabidopsis thaliana* mutants and transgenics. **Plant Biotechnology Journal**, v. 10, p. 488-500, 2012.
- DAMATTA, F. M. et al. Impacts of climate changes on crop physiology and food quality. **Food Research International**, v. 43, n.7, p. 1814-1823, 2010.
- DE SMET, I. et al. Receptor-like kinases shape the plant. **Nature Cell Biology**, v. 11, p. 1166-1173, 2009.
- DE SMET, I. et al. Receptor-like kinase ACR4 restricts formative cell divisions in the *Arabidopsis* root. **Science**, v. 322, p. 594-597, 2008.
- DOUGHTY, J.; ALJABRI, M.; SCOTT, R. J. Flavonoids and the regulation of seed size in *Arabidopsis*. **Biochemical Society Transactions**, v. 42, 2014.
- ENDELMAN, J. B. Ridge regression and other kernels for genomic selection with R package rrBLUP. **Plant Genome**, v. 4, p. 250-255, 2011.
- GAO, M-J. et al. Repression of Seed Maturation Genes by a Trihelix Transcriptional Repressor in *Arabidopsis* Seedlings. **The Plant Cell**, v. 21, p. 54-71, 2009.
- GARCIA, D. et al. *Arabidopsis haiku* Mutants Reveal New Controls of Seed Size by Endosperm. **Plant Physiology**, v. 131, p. 1661-1670, 2003.
- GE, L. et al. Increasing seed size and quality by manipulating BIG SEEDS1 in legume species. **Proceedings of the National Academy of Sciences**, v. 113, n. 44, 2016.

- GIFFORD, M. L.; DEAN, S.; INGRAM, G. C. The Arabidopsis ACR4 gene plays a role in cell layer organisation during ovule integument and sepal margin development. **Development**, v. 130, p. 4249-4258, 2003.
- GILMOR, C. S. et al. Glycosylphosphatidylinositol-Anchored Proteins Are Required for Cell Wall Synthesis and Morphogenesis in Arabidopsis. **The Plant Cell**, v. 17, p. 1128-1140, 2005.
- GROSZMANN, M.; PAICU, T; SMYTH, D. R. Functional domains of SPATULA, a bHLH transcription factor involved in carpel and fruit development in Arabidopsis. **The Plant Journal**, v. 55, n. 1, p. 40-52, 2008.
- HU, Z. et al. Determination of the genetic architecture of seed size and shape via linkage and association analysis in soybean (*Glycine max* L. Merr.). **Genetica**, v. 141, p. 247-254, 2013.
- HWANG, J-U. et al. Plant ABC Transporters Enable Many Unique Aspects of a Terrestrial Plant's Lifestyle. **Molecular Plant**, v. 9, p. 338-355, 2016.
- ICO – International Coffee Organization. Disponível em: <<http://www.ico.org/>>. Acesso em: 18 jan. 2019.
- KAWAKATSU, T. et al. Compensation and interaction between RISBZ1 and RPBF during grain filling in rice. **The Plant Journal**, v. 59, n. 6, p. 908-920, 2009.
- KORTE, A.; FARLOW, A. The advantages and limitations of trait analysis with GWAS: a review. **Plant Methods**, v. 29, p. 1-9, 2013.
- LAFON-PLACETTE, C.; KÖHLER, C. Embryo and endosperm, partners in seed development. **Current Opinion in Plant Biology**, v. 17, p. 64-69, 2014.
- LEROY, T. et al. Improving the quality of African robustas: QTLs for yield- and quality-related traits in *Coffea canephora*. **Tree Genetics & Genomes**, v. 7, n. 4, p. 781-798, 2011.
- MASIERO, S. et al. The Emerging Importance of Type I MADS Box Transcription Factors for Plant Reproduction. **The Plant Cell**, v. 23, n. 3, p. 865–872, 2011.
- NOH, Y. S. et al. EARLY FLOWERING 5 acts as a floral repressor in Arabidopsis. **The Plant Journal**, v. 38, n. 4, p. 664-72, 2004.
- NIU, Y. et al. Association mapping for seed size and shape traits in soybean cultivars. **Molecular Breeding**, v. 31, n. 4, p. 785-794, 2013.
- OHTO, M. et al. Effects of APETALA2 on embryo, endosperm, and seed coat development determine seed size in *Arabidopsis*. **Sexual Plant Reproduction**, v. 22, p. 277-289, 2009.

- PATIL, G. et al. Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. **Theoretical and Applied Genetics**, v. 130, p. 1975–1991, 2017.
- PEGO, J. V.; SMEEKENS, S. Plant fructokinases: A sweet family get-together. **Trends in Plant Science**, v. 5, n. 12, p. 531-536, 2001.
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2017 .URL <http://www.R-project.org/>.
- RAJKUMAR, M. S.; GARG, R.; JAIN, M. Genome-wide discovery of DNA polymorphisms among chickpea cultivars with contrasting seed size/weight and their functional relevance. **Scientific Reports**, v. 8, 2018.
- REN, D. et al. A new regulator of seed size control in *Arabidopsis* identified by a genome-wide association study. **New Phytologist**, 2018.
- RINGLI, C.; KELLER, B.; RYSER, U. Glycine-rich proteins as structural components of plant cell walls. **Cellular and Molecular Life Sciences**, v. 58, p. 1430-1441, 2001.
- ROBERT, H. S. et al. BTB and TAZ domain scaffold proteins perform a crucial function in *Arabidopsis* development. **The Plant Journal**, v. 58, p. 109–121, 2009.
- SCHRUFF, M. C. et al. The AUXIN RESPONSE FACTOR 2 gene of *Arabidopsis* links auxin signalling, cell division, and the size of seeds and other organs. **Development**, v. 133, n. 2, p. 251-261, 2005.
- SPINDEL et al. Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. **Heredity**, v. 116, p. 395-408, 2016.
- STOREY, J. D. A direct approach to false discovery rates. **Journal of the Royal Statistical Society**, v. 64, n. 3, p. 479-498, 2002.
- STOREY, J. D.; TIBSHIRANI, R. Statistical significance for genomewide studies. **Proceedings of the National Academy of Sciences**, v. 100, n. 16, p. 9440-9445, 2003.
- SUN, X. et al. Transcriptional and hormonal signaling control of *Arabidopsis* seed development. **Current Opinion in Plant Biology**, v. 13, p. 611-620, 2010.
- VADDEPALLI, P. et al. Structure-Function Analysis of STRUBBELIG, an *Arabidopsis* Atypical Receptor-Like Kinase Involved in Tissue Morphogenesis, **PLoS ONE**, v. 6, e19730, 2011.
- van NOCKER, S.; LUDWIG, P. The WD-repeat protein superfamily in *Arabidopsis*: conservation and divergence in structure and function. **BMC Genomics**, v. 12;4(1):50, 2003.
- VARSHNEY, R. K., TERAUCHI, R, MCCOUCH, S. R. Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. **PLOS Biology**, v. 12, n. 6, 2014.



WANG, S. B. et al. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. **Scientific Reports**, v. 6, 19444, 2016.

YAMAMOTO, M. P. et al. Synergism between RPBFD1 and RISBZ1 bZIP Activators in the Regulation of Rice Seed Expression Genes. **Plant Physiology**, v. 141, p. 1694-1707, 2006.

YU, J. et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. **Nature Genetics**, v. 38, p. 203-208, 2006.

ZHANG, D. et al. Comparative genetics of seed size traits in divergent cereal lineages represented by sorghum (Panicoidae) and rice (Oryzoidae). **G3: GENES, GENOMES, GENETICS**, v. 5, n. 6, p. 1117-1128, 2015.

ZHOU, Y. et al. Sequence and expression analysis of the Arabidopsis IQM family. **Acta Physiologiae Plantarum**, v. 32, n. 1, p. 191-198, 2010.

## CAPÍTULO 5

### SELEÇÃO GENÔMICA AMPLA EM *C. canephora*

#### 1 INTRODUÇÃO

O grande atrativo da genética molecular em benefício do melhoramento de plantas é a utilização direta das informações do DNA na seleção, permitindo assim, alta eficiência seletiva, maior rapidez na obtenção de ganhos genéticos e baixo custo, isto em comparação com a seleção fenotípica tradicional (RESENDE et al., 2008). Nesse contexto, a seleção genômica ampla (SGA), proposta por Meuwissen, Goddard, Hayes (2001), usa uma alta cobertura de marcadores no genoma e captura seus efeitos simultaneamente. Hoje, com os avanços nas plataformas de genotipagem de alto rendimento, a GS foi implementada com sucesso, primeiramente no melhoramento de gado, e posteriormente adaptada para outros programas de melhoramento, tanto animal (HAYES et al., 2009; HAYES; GODDARD, 2010) como vegetal (BERNARDO; YU, 2007, HEFFNER, SORRELS, JANNINK, 2009; JANNINK; LORENZ; IWATA, 2010,).

Esse sucesso foi atribuído principalmente às técnicas atuais de genotipagem de DNA, que reduziram os custos e permitiram que um grande número de genótipos pudessem ser genotipados a um preço acessível. A SGA assume que, com uma alta densidade de marcadores, os genes de interesse provavelmente apresentem DL com parte importante desses marcadores (GODDARD; HAYES, 2007) e, como somente os marcadores em DL são úteis para explicar a variação genética, é esperado que grande parte ou quase a totalidade desta variação seja explicada pelos marcadores.

O *C. canephora* é um bom ponto de partida para desenvolver estudos de SGA dentro do gênero *Coffea*, por razões econômicas e genéticas, incluindo a ploidia e a maior variabilidade genética, quando comparado ao *C. arabica* (TRAN et al., 2016). A motivação econômica baseia-se no fato de que a produção de *C. canephora* representa 40% da produção mundial de café, sendo que o Brasil ocupa a primeira posição na produção e exportação desses grãos (ICO, 2018).

O melhoramento genético do cafeeiro via SGA representa uma mudança de paradigma, pois permite a seleção precoce de plantas elite ainda no estágio de mudas, principalmente para características de expressão tardia tais como, produção, qualidade

da bebida e tolerância a estresses bióticos e abióticos. Avanços pioneiros na área de genômica do cafeeiro foram alcançados nos últimos anos como, por exemplo, o genoma de referência de *C. canephora* (DENOEUDE et al., 2014), gerando as ferramentas básicas para o estabelecimento de um programa de melhoramento com base na SGA.

Diante do exposto, este estudo objetivou (i) estimar os parâmetros genéticos e (ii) avaliar a aplicabilidade e acurácia da SGA na predição dos valores genéticos genômicos (VGGs) para *C. canephora*.

## 2 MATERIAL E MÉTODOS

### 2.1 Dados fenotípicos

Da mesma forma como descrito para GWAS (Capítulo 3), duas análises foram realizadas separadamente. A primeira, com os dados dos 366 indivíduos que contém as informações de produção, avaliadas em três safras consecutivas (2012, 2013 e 2014), peso de 100 grãos, boia, tamanho e formato dos grãos (peneiras), morfologia dos frutos e potencial hídrico foliar de antemanhã. Na segunda análise, todos os 1.319 indivíduos de *C. canephora* que possuem todas as informações fenotípicas citadas acima, exceto de potencial hídrico foliar de antemanhã. Todos os dados relacionados ao fenótipo dos indivíduos encontram-se descritos no Capítulo 2.

### 2.2 Dados genotípicos

Os dados genotípicos utilizados para a realização de SGA foram os marcadores SNPs identificados no chip de genotipagem 26K Axiom de *C. canephora*. Após um processamento mínimo, utilizando a função *A.mat* do pacote rrBLUP (ENDELMAN, 2011), no *software* R versão 3.4.3 (R Core Team, 2017), obteve-se um conjunto de 18.980 e 16.688 marcadores, considerando os 366 e 1319 indivíduos, respectivamente. Portanto, após essa filtragem, somente os marcadores polimórficos, com *call rate* acima de 90% e com MAF (*minimum allele frequency*) superior a 1% foram utilizados nas análises.

Além dos SNPs identificados por meio do chip de genotipagem, outros marcadores SNPs foram também incorporados nas análises com os 366 indivíduos (CARNEIRO, 2014), já que estes haviam sido genotipados anteriormente utilizando outra técnica denominada nextRAD (*Nextera-tagmentedreductively-amplified* DNA). De um total de 11.230 SNPs identificados pela genotipagem nextRAD, utilizando os mesmos critérios de filtragem acima citados, 5.982 SNPs polimórficos, com *call rate* acima de 90% e com MAF superior a 1% foram adicionados as análises para os 366 indivíduos.

### 2.3 Modelo de predição

As análises de SGA foram realizadas utilizando o método G-BLUP, com o uso do pacote rrBLUP, desenvolvido por Endelman (2011), implementado na plataforma R (R Core Team, 2017), com o uso do comando *mixed.solve*. Uma vez estimado o efeito de cada marca para os diferentes caracteres, foi possível estimar o valor genético genômico (VGG) dos indivíduos aplicando a seguinte fórmula:

$$VGG = \sum_i^n Z_i \hat{g}_i$$

em que  $n$  é o número de SNPs,  $Z_i$  é uma matriz de genótipos, codificado como -1, 0 ou 1 para os genótipos homocigoto recessivo, heterocigoto ou homocigoto dominante, respectivamente, e  $\hat{g}_i$  é uma matriz de efeitos dos SNPs obtidos na população de treinamento.

A validação cruzada é um método usado para avaliar a capacidade de generalização de um modelo preditivo para um conjunto de dados. Ao aplicar esse método, o conjunto de dados é dividido em subconjuntos. Nesse trabalho, utilizou-se a validação *10-fold*, realizando a predição dos VGGs para um dos grupos a partir do modelo de predição gerado com o uso dos outros 9 grupos, portanto para a população de 1.319 indivíduos, 1.189 foram utilizados para treinamento ou estimação e 130 indivíduos na validação. Da mesma forma, para os 366 indivíduos, 330 foram utilizados como conjunto de treinamento ou estimação e 36 para validação.

### 3 RESULTADOS E DISCUSSÃO

#### 3.1 Análise de 366 indivíduos

##### 3.1.1 Componentes de variância e herdabilidade

A análise dos componentes de variância e de herdabilidade foram utilizadas para retratar os valores obtidos para cada um dos caracteres avaliados nos 366 indivíduos de *C. canephora*. Os maiores valores obtidos para os componentes de variância genética foram 278,413 para Chato Graúdo e 146,060 para a característica Chato Médio. As medidas de Eixo foram as características que apresentaram os menores componentes de variância genético, 0,003 e 0,00034, para eixo maior e eixo menor, respectivamente. Os maiores componentes de variância ambientais foram observados para boia (200,433) e para Chato Graúdo (180,158) e o menor foi observado para Eixo Menor (0,002) (TABELA 25).

A partir dos componentes de variância genético e de ambiente, foram obtidas as estimativas para a herdabilidade. Os resultados para a herdabilidade estão apresentados na Tabela 25, e variaram de 0,092 (boia) a 0,726 (Moca Miúdo).

Tabela 24 – Resumo dos resultados obtidos para os componentes de variância e herdabilidade na análise dos caracteres avaliados para os 366 indivíduos de *C. canephora*.

Caracteres	Componentes da variância		Herdabilidade
	Genético	Ambiental	
Potencial	0.041	0.130	0.239
Prod12	3.943	22.470	0.149
Prod13	8.905	16.461	0.351
Prod14	4.692	37.376	0.112
Mean1_2_3	3.187	8.507	0.273
P100	14.621	27.866	0.344
Boia	20.411	200.433	0.092
Ferrugem	0.122	0.289	0.297
Área	0.028	0.070	0.289
Perímetro	0.100	0.207	0.327
Eixo Maior	0.003	0.003	0.560
Eixo Menor	0.000	0.002	0.129
Chato Graúdo	278.413	180.158	0.607
Chato Médio	146.060	84.034	0.635
Chato Miúdo	47.936	18.906	0.717
Moca Graúdo	87.253	36.863	0.703
Moca Médio	23.620	21.297	0.526
Moca Miúdo	39.821	15.049	0.726

### 3.1.2 Acurácia e Erro de predição do modelo

Os valores de acurácia de predição variaram de 0,15, para a característica de Eixo Menor, até 0,53, para Moca Graúdo, conforme Tabela 26. A pressão de seleção (PS) de cada característica foi calculada de acordo com o número de indivíduos que se apresentavam acima da média, portanto para cada característica uma PS foi calculada. O caráter Chato Graúdo apresentou a maior PS, de 11%, seguido pela característica de Eixo Maior (10%) e com PS de 9% estão as características de Moca Graúdo e Miúdo. Para as características de Prod12, Prod14, P100 e Eixo Menor, não se obtiveram PS, pois nenhum indivíduo apresentou médias acima de zero.

Os ganhos obtidos com a seleção (GS) variaram de 0,05 para Perímetro até 0,84 para Moca Médio (Tabela 26). Nota-se que as características de Chato Graúdo, Chato Miúdo e Moca Miúdo tiveram um GS acima de 1,0, o que não seria possível. Possivelmente, esses resultados obtidos devem-se ao fato de que o modelo de SGA utilizado está ajustado para dados gaussianos e essas características, conforme apresentado na Figura 13 do Capítulo 2, apresentam uma distribuição assimétrica. Provavelmente esses resultados poderiam ser melhor ajustados utilizando um modelo corrigido para assimetria.

Tabela 25 – Acurácia preditiva do modelo, ganho com a seleção (GS) e pressão de seleção (PS) para cada característica analisada nos 366 indivíduos de *C. canephora*.

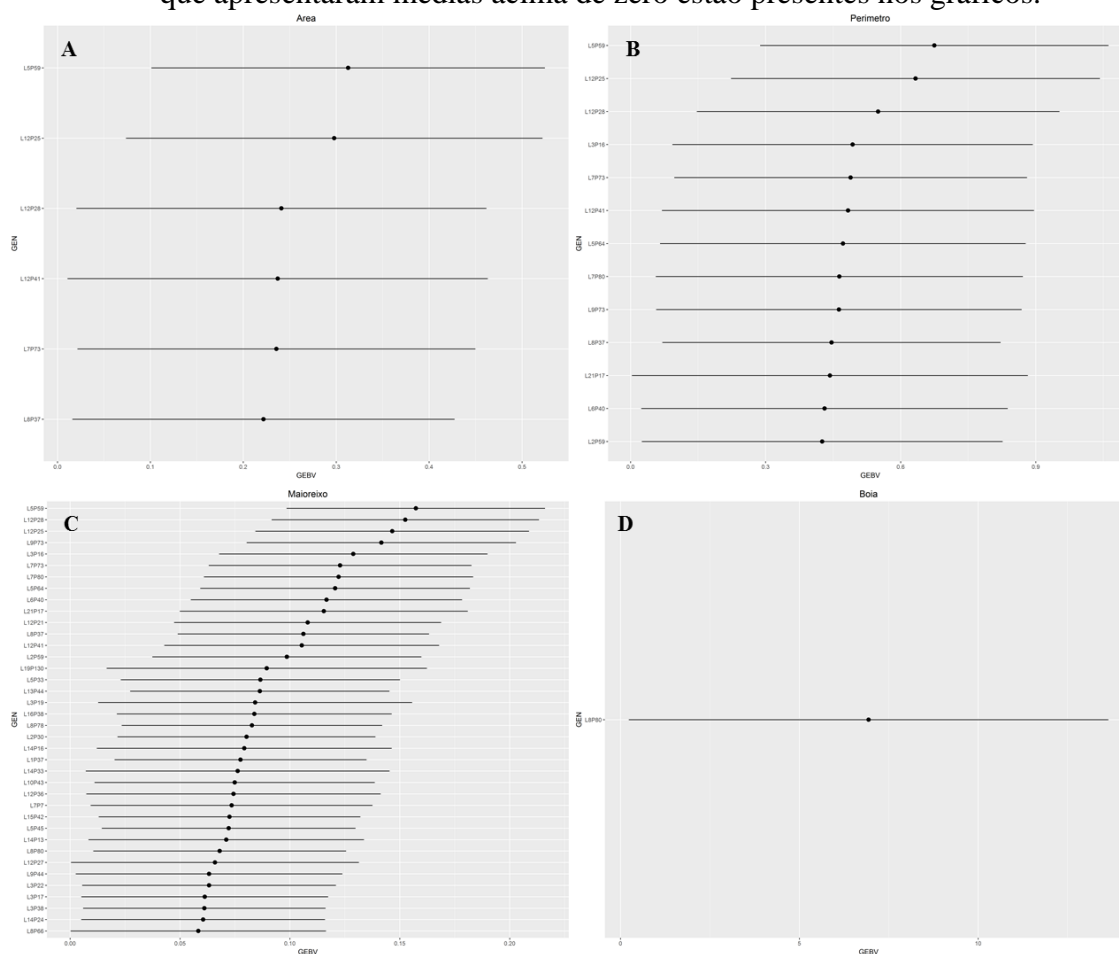
<b>Caracteres</b>	<b>Acurácia</b>	<b>GS</b>	<b>PS</b>
Potencial	0.32	0.31	0.03
Prod12	0.17	0.00	NaN
Prod13	0.32	0.38	0.03
Prod14	0.16	0.00	NaN
Mean1_2_3	0.30	0.24	0.03
P100	0.27	0.00	NaN
Boia	0.22	0.61	NaN
Ferrugem	0.18	0.57	0.01
Área	0.30	0.09	0.02
Perímetro	0.34	0.05	0.04
Eixo Maior	0.51	0.09	0.10
Eixo Menor	0.15	0.00	NaN
Chato Graúdo	0.49	1.03	0.11
Chato Médio	0.42	0.49	0.04
Chato Miúdo	0.48	3.21	0.07
Moca Graúdo	0.53	0.47	0.09
Moca Médio	0.36	0.84	0.05
Moca Miúdo	0.47	2.29	0.09

A partir dos VGGs e do erros de predição para todos os indivíduos, gráficos de lagarta com o erro de predição do modelo para cada característica foram gerados e somente aqueles indivíduos que apresentavam médias acima de zero foram considerados.

Na Figura 50 encontram-se os gráficos de lagarta para Área, Perímetro, Maior Eixo e Boia. Um maior número de indivíduos acima da média foi observado para a característica de Maior Eixo (FIGURA 50C), contendo 38 indivíduos, e somente um genótipo acima da média foi identificado para boia (FIGURA 50 D). Área e perímetro tiveram 6 e 13 plantas acima da média, respectivamente (FIGURAS 50A e B).

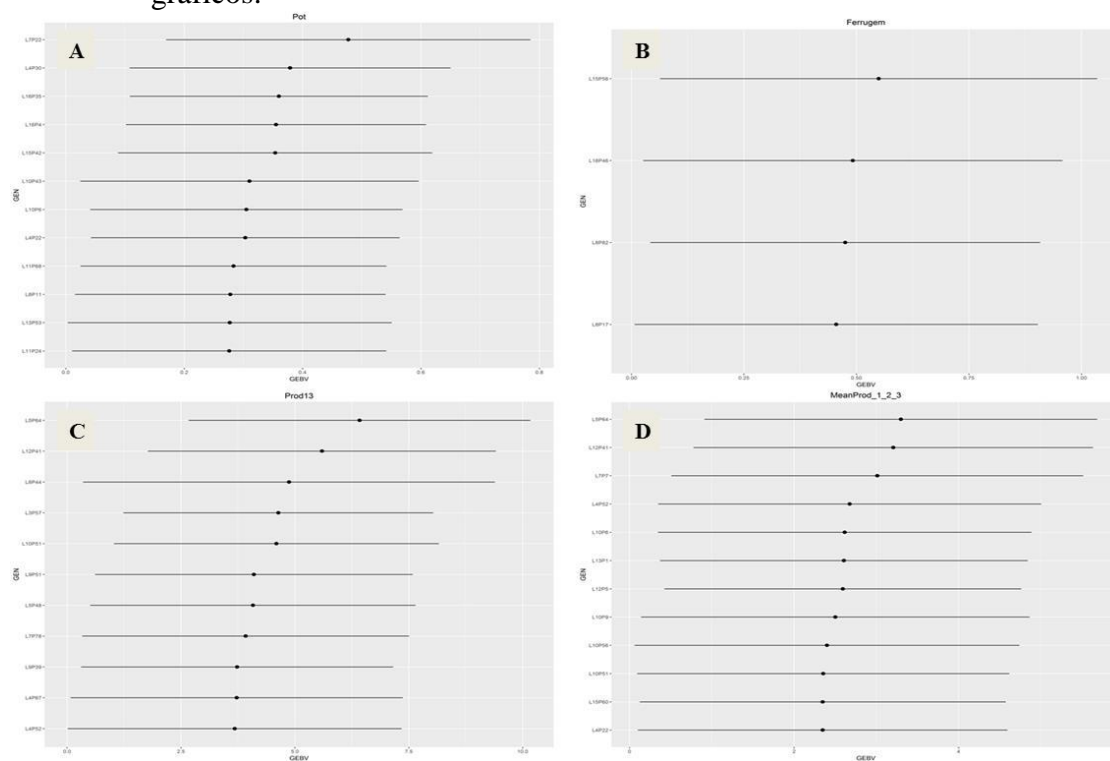


Figura 50 – Gráfico com o VGG e o erro de predição do modelo para as características (A) Área, (B) Perímetro, (C) Maior Eixo e (D) Boia. Somente os indivíduos que apresentaram médias acima de zero estão presentes nos gráficos.



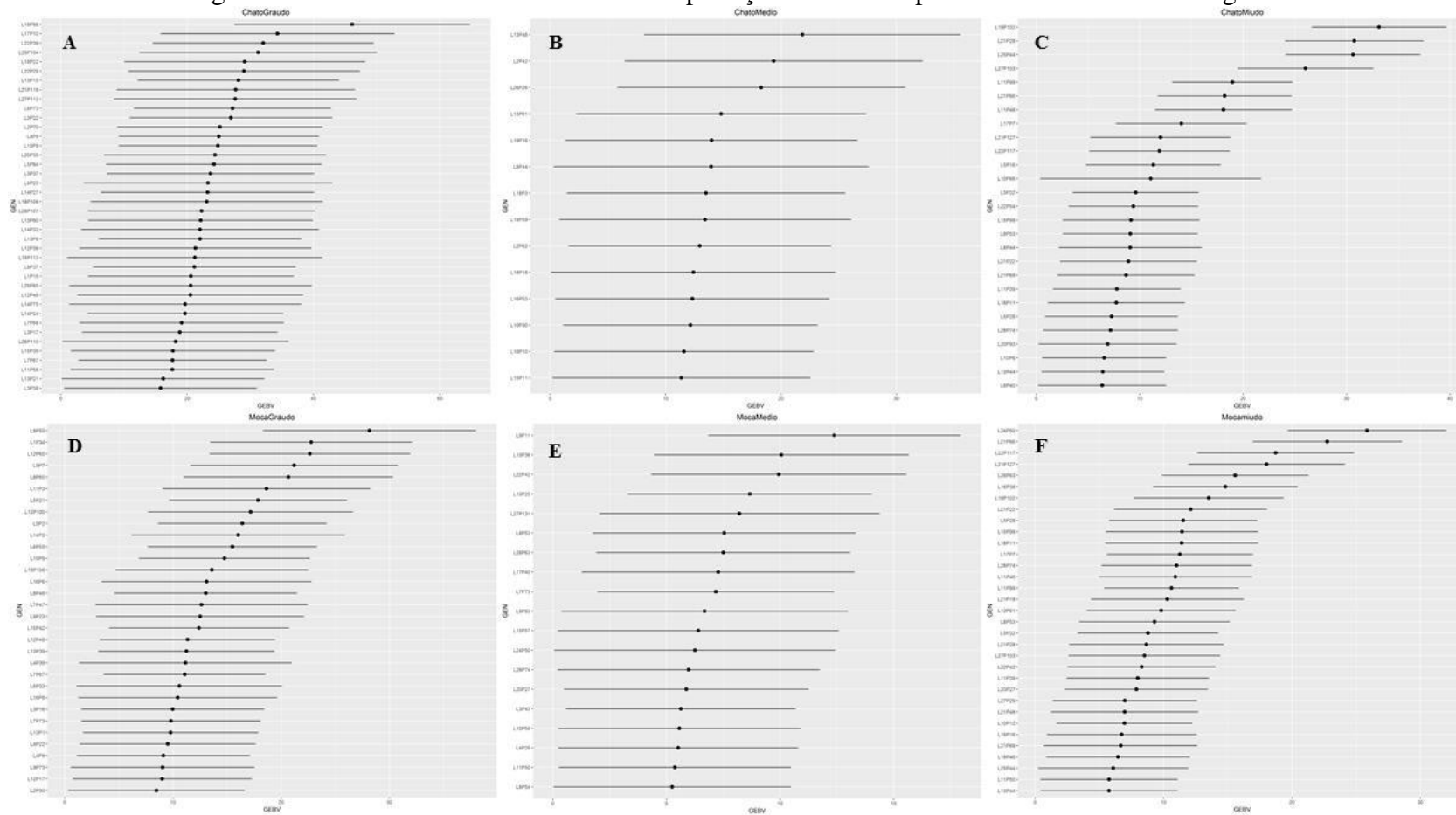
As características de potencial hídrico foliar de antemanhã e a média de produção para os três anos avaliados apresentaram 12 indivíduos acima da média (FIGURA 51A e D), seguida por produção (2013), com 11 plantas (FIGURA 51C), e ferrugem, com 4 indivíduos (FIGURA 51B). Somente a produção avaliada em 2013 e a média de produção para os três anos avaliados (2012, 2013 e 2014) apresentaram plantas que estavam acima da média.

Figura 51 – Gráfico com o VGG e o erro de predição do modelo para as características (A) Potencial hídrico foliar de antemanhã, (B) Ferrugem, (C) Produção (2013) e (D) Média de produção para os três anos avaliados. Somente os indivíduos que apresentaram médias acima de zero estão presentes nos gráficos.



Para tamanho e formato de grãos, uma maior quantidade de indivíduos que apresentaram VGG maiores que 0 foram identificados para os formatos chato e moca de grãos graúdos e miúdos, conforme Figura 50. Os grãos chato e moca médios apresentaram 14 e 19 plantas acima da média (FIGURAS 52B e E).

Figura 52 – Gráfico com o VGG e o erro de predição do modelo para tamanho e formato de grãos.



Legenda: VGG e erro de predição para: (A) Chato Graúdo, (B) Chato Médio, (C) Chato Miúdo, (D) Moca Graúdo, (E) Moca Médio e (F) Moca Miúdo. Somente os indivíduos que apresentaram médias acima de zero estão presentes nos gráficos.

## 3.2 Análise de 1.319 indivíduos

### 3.2.1 Componentes de variância e herdabilidade

Na análise dos componentes de variância e de herdabilidade das marcas para os 1.319 indivíduos, nota-se que as maiores herdabilidades foram encontradas para Moca Miúdo (56%), seguidas por Chato Graúdo (55%) e Moca Médio (50%). Mais uma vez, assim como ocorreu com os 366 indivíduos, o valor mais baixo foi encontrado para Boia (8%) (TABELA 27). As maiores variâncias genética e ambiental foram obtidas para a característica Chato Graúdo e os caracteres de Eixo Maior e Menor apresentaram os menores valores (TABELA 27).

Tabela 26 – Resumo dos resultados obtidos para os componentes de variância e herdabilidade na análise dos caracteres avaliados para os 1.319 indivíduos de *C. canephora*.

Caracteres	Componentes da variância		Herdabilidade
	Genético	Ambiental	
Prod12	1.47	11.73	0.11
Prod13	5.17	16.77	0.24
Prod14	8.84	32.59	0.21
Mean1_2_3	3.49	7.76	0.31
P100	2.52	2.78	0.48
Boia	11.44	130.45	0.08
Área	0.04	0.07	0.35
Perímetro	0.12	0.21	0.37
Eixo Maior	0.00286	0.00437	0.40
Eixo Menor	0.00078	0.00224	0.26
Chato Graúdo	210.71	168.97	0.55
Chato Médio	45.56	152.08	0.23
Chato Miúdo	30.33	34.96	0.46
Moca Graúdo	57.30	63.76	0.47
Moca Médio	20.09	19.95	0.50
Moca Miúdo	29.76	23.17	0.56

### 3.2.2 Acurácia e Erro de predição do modelo

Como apresentado na Tabela 28, as acurácias do modelo proposto para os 1.319 indivíduos foi de 61% para a característica Eixo Maior, seguido pela característica de Chato Graúdo, explicando 60% da variação fenotípica. As acurácias obtidas para a característica de Produção variaram de 0,18, para Produção de 2012, a 0,4, para a média entre os três anos.

Tabela 27 – Acurácia preditiva do modelo, ganho com a seleção (GS) e pressão de seleção (PS) para cada característica analisada nos 1.319 indivíduos de *C. canephora*.

<b>Caracteres</b>	<b>Acurácia</b>	<b>GS</b>	<b>PS</b>
Prod12	0.18	0.17	0.02
Prod13	0.39	0.46	0.10
Prod14	0.29	0.48	0.05
Mean1_2_3	0.40	0.30	0.09
P100	0.56	0.17	0.13
Boia	0.25	0.59	0.03
Área	0.50	0.11	0.15
Perímetro	0.53	0.06	0.16
Eixo Maior	0.61	0.09	0.17
Eixo Menor	0.30	0.05	0.04
Chato Graúdo	0.60	1.03	0.16
Chato Médio	0.26	0.31	0.02
Chato Miúdo	0.50	1.94	0.10
Moca Graúdo	0.45	0.38	0.08
Moca Médio	0.55	0.69	0.12
Moca Miúdo	0.57	1.58	0.12

Os gráficos de lagarta com o VGG e o erro de predição para as características que apresentaram indivíduos acima da média estão apresentados nas Figuras 53-55. Nota-se que foram identificados indivíduos acima da média para todas as características avaliadas para os 1.319 indivíduos de *C. canephora*. Considerando as medidas de frutos, área, perímetro e eixo maior (FIGURAS 53A, B e C, respectivamente) apresentaram um elevado número de indivíduos acima da média, comparado ao resultado obtido para eixo menor (FIGURA 53 D).

Para peso de 100 grãos, as plantas identificadas acima da média apresentaram VGGs variando de 1 a 5 (FIGURA 54B). A produção avaliada em 2012 apresentou 26 plantas acima da média com o VGG variando de 1,5 a 2,5 (FIGURA 54C). Já para a produção (2013), o número de plantas que ficaram acima da média foi bem mais elevado, assim como o VGG, variando de 2,5 até 7,5 (FIGURA 54D). A produção de 2014 apresentou 56 indivíduos acima da média, com o VGG variando de 3-7 (FIGURA 54E).

Figura 53 – Gráfico com o VGG e o erro de predição do modelo para as características (A) Área, (B) Perímetro, (C) Maior Eixo e (D) Menor Eixo na análise com os 1.319 indivíduos. Somente os indivíduos que apresentaram médias acima de zero estão presentes nos gráficos.

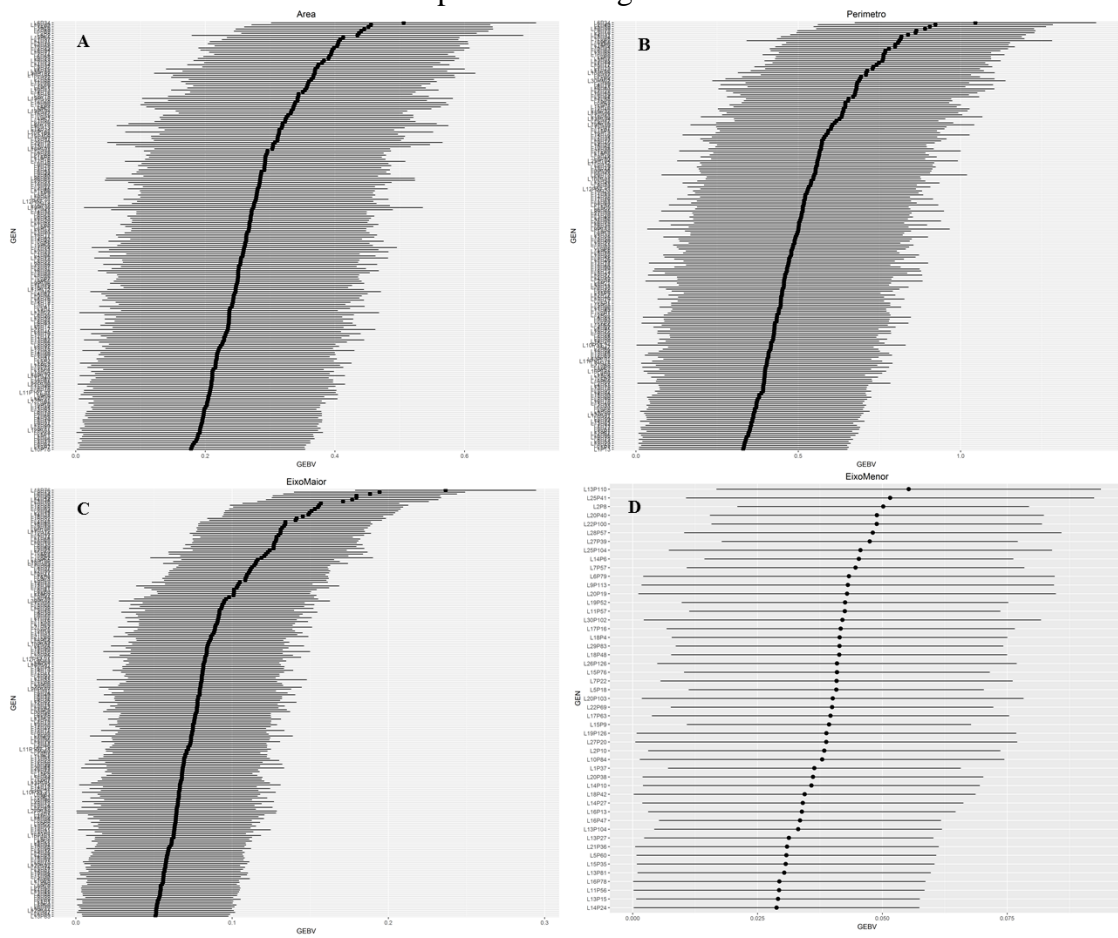


Figura 54 – Gráfico com o VGG e o erro de predição do modelo para as características (A) Boia, (B) P100, (C) Prod12, (D) Prod13, (E) Prod14 e (F) Média para as três produções, na análise com os 1.319 indivíduos. Somente os indivíduos que apresentaram médias acima de zero estão presentes nos gráficos.

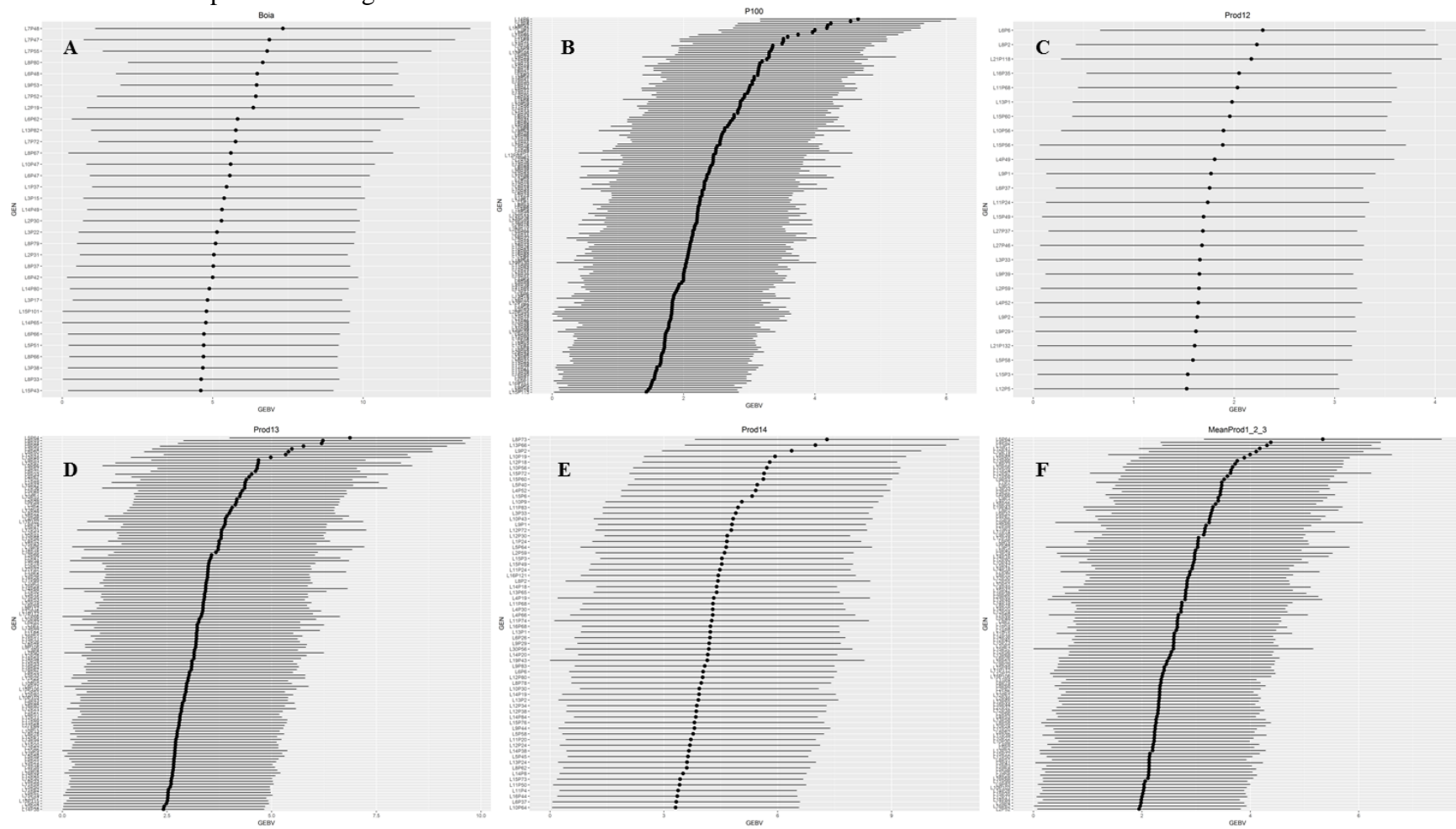
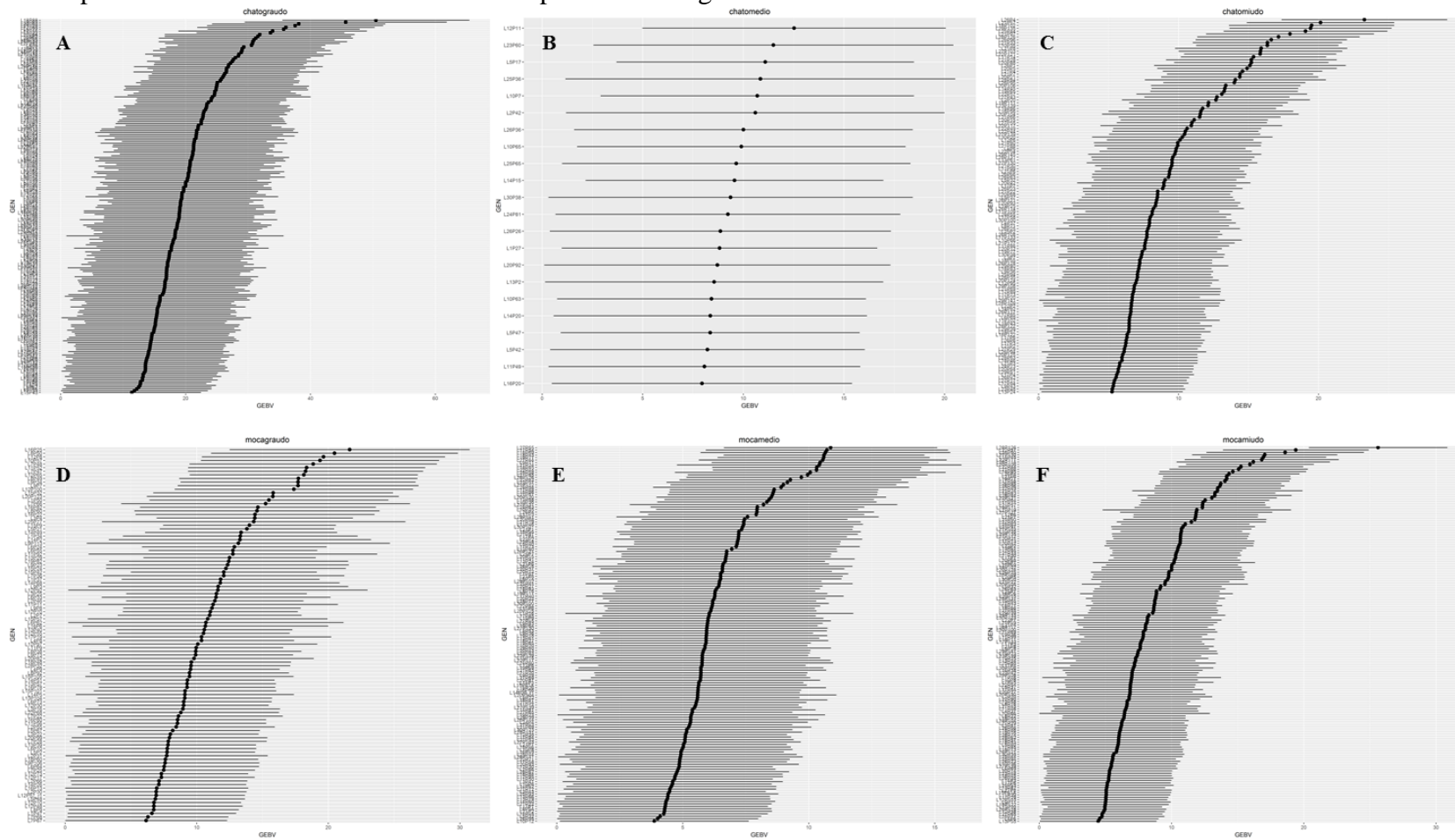


Figura 55 – Gráfico com o VGG e o erro de predição do modelo para as características (A) Chato Graúdo, (B) Chato Médio, (C) Chato Miúdo, (D) Moca Graúdo, (E) Moca Médio e (F) Moca Miúdo, na análise com os 1.319 indivíduos. Somente os indivíduos que apresentaram médias acima de zero estão presentes nos gráficos.





### 3.2.3 Validação Real

Na tentativa de verificar a acurácia real do modelo preditivo obtido, 13 plantas relacionadas aos 1.319 indivíduos, genotipadas com o *chip* de *C. canephora* e estabelecidas na Embrapa Cerrados, porém em outro campo experimental, foram avaliadas quanto à produtividade para as três primeiras produções (anos de 2014, 2015 e 2016), classificação quanto ao formato e granulometria (peneiras) e peso de 100 grãos. As correlações obtidas entre os dados reais e os dados preditos, de acordo com o modelo, encontram-se apresentadas nas Figuras 56 e 57.

Para as três safras avaliadas, a safra de 2015 apresentou uma correlação de 25 entre o dado real e o que foi predito pelo modelo (FIGURA 56). Já para as demais características, as correlações mais altas foram para peso de 100 grãos e moca gráudo, apresentando correlações de 40 (FIGURA 56) e 15% (FIGURA 57), respectivamente.

Figura 56 – Correlação entre os dados preditos, em alaranjado, e os dados reais, em azul, para produção (2014, 2015 e 2016) e peso de 100 grãos, para os 13 indivíduos de *C. canephora* utilizados na validação real do modelo de predição genômica.

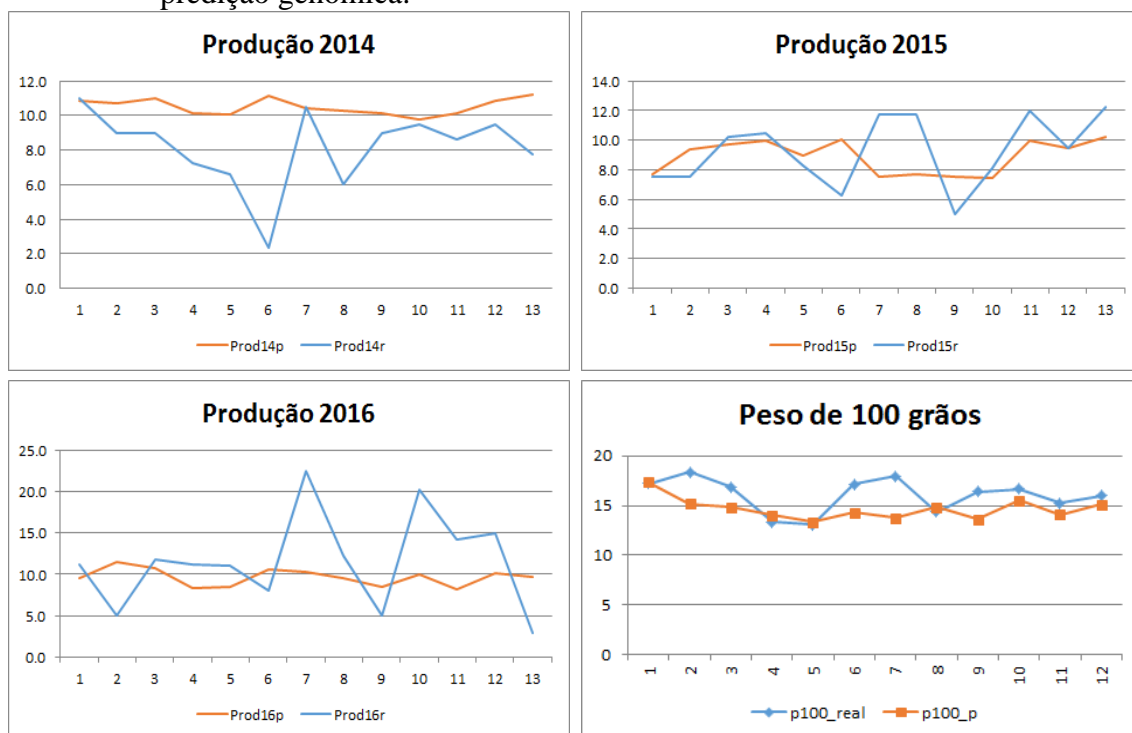
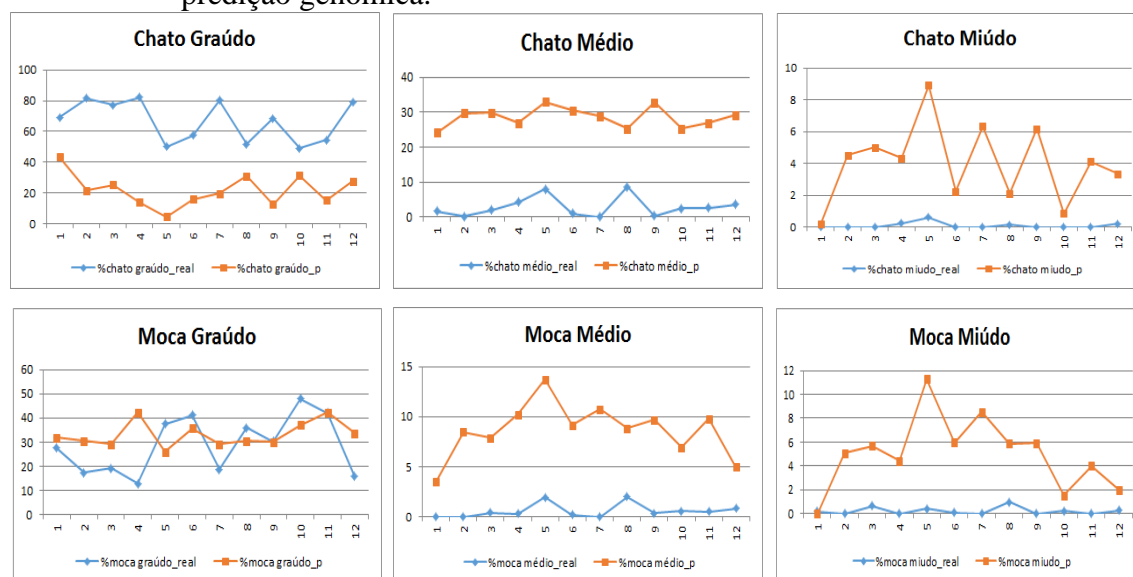


Figura 57 – Correlação entre os dados preditos, em alaranjado, e os dados reais, em azul, para as características de formato e tamanho do grão, para os 13 indivíduos de *C. canephora* utilizados na validação real do modelo de predição genômica.



Dentre as diversas metodologias envolvendo marcadores moleculares no melhoramento de plantas perenes, a SGA provavelmente tem gerado a maior expectativa, em termos de possibilidade de seleção para caracteres quantitativos (RESENDE et al., 2008; RESENDE, M. D. et al., 2012; RESENDE, M. F. et al., 2012), uma vez que utiliza uma grande quantidade de marcas, o que aumenta a probabilidade de capturar uma maior proporção da variação genética (MEUWISSEN; GODDARD; HAYES., 2001; WIENTJES; VEERKAMP; CALUS, 2013), facilitando a seleção de genótipos superiores.

A partir das informações das equações preditivas da SGA, estimou-se a herdabilidade genômica, apresentando magnitudes baixas ou moderadas. De certo modo, a herdabilidade permite prever o sucesso a ser alcançado com a seleção, uma vez que características com baixa herdabilidade devem apresentar menor capacidade preditiva (LEGARRA et al., 2008). Em estudo realizado por Grattapaglia e de Resende (2011), algumas simulações foram realizadas e verificou-se que o aumento na herdabilidade levava a um aumento na acurácia da SG. Dentre os parâmetros genéticos avaliados, o componente que tem maior interesse por parte dos melhoristas é a herdabilidade, por refletir todos os outros componentes de variância e indicar a proporção da variação fenotípica que pode ser herdada (FALCONER; MACKAY, 1996).

Os resultados mostraram uma alta correlação entre as acurácias da validação e as herdabilidades, de mais de 90% para as duas análises realizadas com os diferentes números de indivíduos e marcadores. Uma elevada correlação (82%) entre acurácia e herdabilidade também foi reportada por Sousa et al. (2019). Considerando especificamente o resultado obtido para a média de produção, verifica-se que os resultados de herdabilidade e acurácia do modelo estão um pouco mais elevados comparado ao demonstrado por Sousa et al. (2019). Herdabilidade e acurácia foram, respectivamente, de 0,27 e 0,3 (366 indivíduos) e de 0,31 e 0,4 (1.319 indivíduos).

As acurácias obtidas entre as análises realizadas com o diferente número de indivíduos mostrou claramente que o tamanho da população interfere na predição do modelo. Somente para as características de chato médio e moca graúdo que não ocorreu um aumento da acurácia, para as demais características pode-se observar um aumento de até 29%, como foi o caso de peso de 100 grãos.

O número de marcadores utilizados nesse estudo mostra-se suficiente, uma vez que não necessariamente o aumento progressivo de marcadores resulte em aumento da acurácia, ao contrário, Sousa et al. (2019) demonstraram que o máximo de acurácia obtida foi utilizando 16.000 marcadores, aumentando esse número para 20.000 a acurácia reduzia. O sucesso da SGA pode ser influenciado por inúmeros fatores, que consequentemente interferem na acurácia do modelo, como tamanho da população de treinamento, tamanho real da população, densidade de marcadores, herdabilidade e número de QTLs controlando as características (GRATTAPAGLIA; RESENDE, 2011; DESTA; ORTIZ, 2014).

#### **4 CONCLUSÃO**

O uso da SGA no gênero *Coffea* é uma abordagem promissora, uma vez que a maioria dos estudos moleculares no gênero ainda contempla a abordagem molecular tradicional. Considerando que a espécie é perene, de ciclo longo e, em comparação com a avaliação fenotípica tradicional, espera-se que com a incorporação da SGA acelere o ciclo de reprodução, mantenha a diversidade genética e aumente o ganho genético por unidade de tempo.

**REFERÊNCIAS**

- BERNARDO, R.; YU, J. Prospects for genome wide selection for quantitative traits in maize. **Crop Science**, v. 47, p. 1082-1090, 2007.
- CARNEIRO, F. A. et al. Phenotyping and genotyping a *Coffea canephora* population, cultivated at high altitude, aiming at a GWS program for coffee. In: THE 25TH INTERNATIONAL CONFERENCE ON COFFEE SCIENCE, 2014, Colômbia.
- DENOEUDE, F. et al. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. **Science**, v. 345, n. 6201, p. 1181-1184, 2014.
- DESTA, Z. A.; ORTIZ, R. Genomic selection: genome-wide prediction in plant improvement. **Trends in Plant Science**, v. 19, p. 592-601, 2014.
- ENDELMAN, J. B. Ridge regression and other kernels for genomic selection with R package rrBLUP. **Plant Genome**, v. 4, p. 250-255, 2011.
- GODDARD, M. E.; HAYES, B. J. Genomic selection. **Journal of Animal Breeding and Genetics**, v. 124, p. 323-330, 2007.
- GRATTAPAGLIA, D.; de RESENDE, M. D. V. Genomic selection in forest tree breeding. **Tree Genetics & Genomes**, v. 7, n. 2, p. 241-255, 2011.
- HAYES, B. J. et al. Invited review: genomic selection in dairy cattle: progress and challenges. **Journal of Dairy Science**, v. 92, p. 433-443, 2009.
- HAYES, B. J.; GODDARD, M. E. Genome-wide association and genomic selection in animal breeding. **Genome**, v. 53, p. 876-883. 2010.
- HEFFNER, E. L.; SORRELLS, M. E.; JANNINK, J. L. Genomic selection for crop improvement. **Crop Science**, v. 49, p. 1-12, 2009.
- ICO – International Coffee Organization. Disponível em: <<http://www.ico.org/>>. Acesso em: 18 jan. 2019.
- JANNINK, J. L.; LORENZ, A. J.; IWATA, H. Genomic selection in plant breeding: from theory to practice. **Brief Funct Genomics**, v. 9, p. 166-177, 2010.
- LEGARRA, A. et al. Performance of genomic selection in mice. **Genetics**, v. 180, p. 611-618, 2008.
- MEUWISSEN, T. H. E.; GODDARD, M. E.; HAYES, B. J. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, n. 4, p. 1819-1829, 2001.
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2017 .URL <http://www.R-project.org/>.

RESENDE, M. D. V. de et al. Seleção Genômica Ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal Brasileira**, n. 56, p. 63-77, 2008.

SOUSA, T. V. et al. Early Selection Enabled by the Implementation of Genomic Selection in *Coffea arabica* Breeding. **Frontiers in Plant Science**, v. 9, p. 1-12, 2019.

TRAN, H. T. M. et al. Advances in genomics for the improvement of quality in coffee. **Journal of the Science of Food and Agriculture**, v. 96, n. 10, p. 3300-3312, 2016.

WIJNTJES, Y. C. J.; VEERKAMP, R. F.; CALUS, M. P. L. The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. **Genetics**, v. 193, n. 2, p. 621-631, 2013.