RESSALVA

Atendendo solicitação do(a) autor(a), o texto completo desta tese será disponibilizado somente a partir de 02/04/2020.

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Botucatu

Instituto de Biociências de Botucatu - Unesp

ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA APLICADA AO CONTEÚDO DE MACRONUTRIENTES EM GRÃOS DE *Coffea arabica* L.

MARIANE SILVA FELICIO

BOTUCATU – SP
2020

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Botucatu

UNIVERSIDADE ESTADUAL PAULISTA
"Júlio de Mesquita Filho"
INSTITUTO DE BIOCIÊNCIAS DE BOTUCATU

ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA APLICADA AO CONTEÚDO DE MACRONUTRIENTES EM GRÃOS DE *Coffea arabica* L.

ALUNA: MARIANE SILVA FELICIO
ORIENTADOR: PROF. DR. DOUGLAS SILVA DOMINGUES
COORIENTADOR: PROF. DR. LUIZ FILIPE PROTASIO PEREIRA

Tese apresentada ao Instituto de Biociências, Campus de Botucatu, UNESP, para obtenção do título de Doutor no Programa de Pós-Graduação em Ciências Biológicas (Genética).

BOTUCATU – SP
2020

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE-CRB 8/5651

Felicio, Mariane Silva.
Estudo de associação genômica ampla aplicada ao conteúdo de macronutrientes em grãos de Coffea arabica L. / Mariane Silva Felicio. - Botucatu, 2020

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu
Orientador: Douglas Silva Domingues
Coorientador: Luiz Filipe Protasio Pereira
Capes: 20203004

1. Café. 2. Genômica. 3. Estresse vegetal. 4. Critica de imputação de dados (Estatística). 5. Marcadores genéticos. 6. Ausência de dados (Estatística).

Palavras-chave: Dados ausentes; Espécies não-modelo; Estresses bióticos e abióticos; Imputação.

“Contudo, seja qual for o grau a que chegamos, o que importa é prosseguir decididamente. ”
(Fl 3, 16)

“Não vos conformeis com este mundo, mas transformai-vos pela renovação do vosso espírito, para que possais discernir qual é a vontade de Deus, o que é bom, o que lhe agrada e o que é perfeito. ”
(Rm 12, 2)

Dedico à Deus, minha fortaleza. Ao meu pai Benedito Felicio Filho (*in memoriam*) que sempre me incentivou a buscar meus sonhos.

Agradecimentos

À Deus, por todas as graças recebidas e por nunca me abandonar.

À minha família, meu pai Benedito (*in memoriam*) que sempre me incentivou a estudar e foi um grande exemplo para mim. À minha mãe Maria Izabel, pelo suporte para realizar esse curso. À minha irmã Marina, por todo amor e paciência que tem comigo e teve durante o desenvolvimento desse trabalho. Ao meu avô Benedito, um grande homem, meu melhor amigo e conselheiro.

Ao meu orientador Dr. Douglas Silva Domingues, pela amizade, por me confiar esse trabalho, me incentivar a buscar a capacitação em diferentes cursos, por todos os conselhos e pela orientação valiosa durante o mestrado e doutorado.

Ao meu coorientador Dr. Luiz Filipe Protasio Pereira, pela amizade, pelos ensinamentos, por todas as discussões que contribuíram para o andamento do trabalho e também por me coorientar desde a graduação.

Ao Dr. Gustavo Sant'Ana por todos os conhecimentos transmitidos e auxílio para a realização das análises de dados.

Aos amigos Lucinéia Maria da Silva e Manuel Luiz Martins, pela amizade, conselhos e dedicação às coletas que foram fundamentais para a realização desse trabalho.

Ao Dr. Gabriel Rodrigues Alves Margarido, e seus alunos Amanda Avelar, Lorena Guimarães Batista, Fernando Henrique Correr e Guilherme Kenichi Hosaka, por me receberem tão bem no Laboratório de Bioinformática Aplicada a Bioenergia e por todos os conhecimentos que me passaram que foram essenciais para o desenvolvimento desse trabalho.

Ao Leandro Carrijo Cintra, por me auxiliar com o necessário para a realização das análises de Bioinformática utilizando o servidor da EMBRAPA.

Aos funcionários e estagiários do laboratório de solos e tecido vegetal do IAPAR, especialmente à Rosineia Aparecida de Souza pelos ensinamentos transmitidos e dedicação para realizar as análises de nutrientes.

Aos Drs. Eduardo Fermino Carlos, Nelson da Silva Fonseca, Paula Cristina da Silva Ângelo, Leandro Simões Azeredo Gonçalves pela amizade e conhecimentos transmitidos que auxiliaram para o desenvolvimento desse trabalho.

Ao servidor do IAPAR Ovidio Mantoani, pela amizade, por todos os conhecimentos passados e auxílios para a moagem dos grãos de café.

A equipe do programa de melhoramento do cafeeiro do IAPAR pela manutenção do germoplasma utilizado nesse trabalho. Especialmente ao Eugênio Brandt e Fernando Carducci

pelo auxílio para a identificação das plantas no campo e com as coletas de frutos. À Luciana Harumi pelo auxílio com as informações sobre a manutenção do banco de germoplasma.

Aos Drs. Celso Luis Marino, Eveline Teixeira Caixeta, Roberto Fritsche Neto e João Ricardo Bachega Feijó Rosa, pela participação da banca de tese e por todas as sugestões para a melhoria do trabalho.

Às amigas Darley de Souza, Aline Silveira, Nathália Caroline Rodrigues e Nara Moreira, pelo suporte, amizade e orações durante essa caminhada.

Às amigas Jordana Oliveira e Najila Nolie, por me acolherem em casa em tantas viagens para Botucatu. E a todos os amigos que me acolheram na cidade, especialmente a Bruna Jerônimo, Vanessa Jacob, Camila Moreira, Marco Soares, Adauto Cardoso, Arno Butzge, Viviani Sene e Isabel Silverio.

À amiga Jessica Delfini, por todas as conversas, partilhas, desabafos e scripts compartilhados.

Aos amigos do Laboratório de Biotecnologia Vegetal (LBI) do IAPAR, pela amizade e companheirismo. Especialmente Bruna Silvestre Rodrigues da Silva, Caroline Ariyoshi, Rafaelle Vecchia Ferreira e Lívia Maria Nogueira Brito pela parceria e auxílios para o desenvolvimento desse trabalho.

Às amigas do Laboratório de Cultura de Tecidos do IAPAR, Cícera Martimiano e Suely Ario Kudo pela amizade e por me auxiliarem sempre que precisei.

A todos os professores do Programa de Pós-Graduação em Ciências Biológicas (Genética), especialmente ao coordenador Dr. Ivan de Godoy Maia, pela dedicação para proporcionar uma formação de qualidade para tantos alunos e por todos os eventos científicos produzidos para nosso desenvolvimento como pesquisadores.

Aos funcionários da Seção técnica de Pós-Graduação, pelo trabalho de excelência, por terem me assessorado e tirado minhas dúvidas sempre que preciso.

À Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP) e todos os seus funcionários, pela infraestrutura fornecida e a manutenção das atividades na universidade.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de estudos que possibilitou minha dedicação exclusiva ao trabalho.

Ao Instituto Agronômico do Paraná (IAPAR) pela infraestrutura e pelo acesso às plantas utilizadas nesse trabalho.

Muito obrigada!

RESUMO

O café é uma das commodities agrícolas tropicais mais comercializadas no mundo. *Coffea arabica* é a principal espécie utilizada para a produção comercial de café. A espécie é originária da Etiópia. Ela é única espécie alotetraploide do gênero (2n = 4x = 44) e se reproduz predominantemente por autofecundação. As cultivares comerciais de *C. arabica* possuem baixa diversidade genética, o que indica a necessidade de introgressão de alelos de germoplasma para o melhoramento dessas cultivares. Acessos do centro de origem da espécie possuem maior diversidade que as cultivares comerciais e podem ser utilizados para a identificação de novos alelos. O conteúdo de macronutrientes em grãos do cafeeiro tem impacto direto na qualidade do produto. No entanto, a base molecular da composição mineral de grãos de cafeeiro ainda é pouco conhecida. Com isso, o objetivo desse trabalho foi identificar marcadores SNP possivelmente associados com a composição de macronutrientes em grãos de *C. arabica*. Para alcance deste objetivo, foram comparados três métodos de imputação de genótipos, bem como foi realizado o mapeamento associativo em estudo de associação genômica ampla (GWAS). Foi utilizado um painel de 110 genótipos de *C. arabica*, composto por genótipos elite do programa de melhoramento do Instituto Agronômico do Paraná (3), cultivares comerciais (11) e acessos selvagens (96). Foram realizadas análises da composição de cinco macronutrientes (N, P, K, Ca e Mg) em grãos de cafeeiro coletados de 70 e 105 genótipos de *C. arabica* nos anos de 2017 e 2018, respectivamente. Foram calculados valores de BLUP para 65 genótipos, para os quais foram realizadas coletas nos dois anos. Foi identificado que os acessos selvagens de *C. arabica* possuem maior variabilidade genética e de composição de macronutrientes nos grãos que as cultivares comerciais. Foram identificados mais marcadores SNPs quando o mapeamento dos dados de GBS foi realizado com o genoma de referência da cultivar Caturra (11.230 SNPs) do que com o diaploide Et39 (9.991 SNPs). O método Beagle apresentou o melhor desempenho para a imputação dos painéis de marcadores e a imputação contribuiu para reduzir a taxa de associações espúrias identificadas nas análises de GWAS. Foram identificados 5, 12, 13, 6, 6 e 1 marcadores em regiões codificantes associados a composição de N, P, K, Ca, Mg, Mg e K em grãos de *C. arabica*, respectivamente. As regiões genômicas identificadas foram relacionadas a diversas rotas metabólicas em plantas, incluindo o transporte de nutrientes, a germinação de sementes, o controle da época de florescimento e algumas vias de resposta a estresses bióticos e abióticos. As proteínas identificadas nesse trabalho podem ser utilizadas como alvos em futuros estudos para a caracterização de vias metabólicas que atuam no controle do acúmulo de macronutrientes em plantas da espécie *C. arabica*. Além disso, os genótipos

selvagens de *C. arabica* utilizados nesse trabalho podem ser utilizados para identificação de alelos favoráveis a serem introduzidos em cultivares comerciais.

Palavras-chave: imputação, dados ausentes, espécies não-modelo, estresses bióticos e abióticos.

ABSTRACT

Coffee is one of the most traded tropical commodities in the world. *Coffea arabica* is the main species used for commercial production. The species is originally from Ethiopia. In the *Coffea* genus, *C. arabica* is the only allotetraploid species (2n = 4x = 44) and it reproduces predominantly by self-fertilization. The commercial cultivars of *C. arabica* have a narrow genetic base that indicates the need for the introgression of new alleles from germplasm into coffee breeding programs. Wild accessions of *C. arabica*, from Ethiopia, have higher genetic diversity and can be used to identify new alleles. The macronutrient composition of the coffee grains has a direct impact on grain quality. However, the molecular basis for the mineral composition in coffee grains still poorly understood. Thus, the aim of this work was to perform mapping association analyses using the genome-wide association study (GWAS) technique to identify single nucleotide polymorphisms (SNPs) associated with macronutrient content in coffee grains from *C. arabica*. We also tested three imputation methods (haplotype missing allele imputation - Beagle, K-nearest neighbors, and Random Forest) in the genotypic data, and mapped it to two *C. arabica* reference genomes from the cultivar Caturra red and the spontaneous dihaploid Et39. We used a panel of 110 *C. arabica* genotypes, including elite landraces from the IAPAR coffee breeding program (3), commercial cultivars (11) and wild accessions (96). Analysis of the composition of five macronutrients (N, P, K, Ca and Mg) was carried out on coffee grains collected from 70 and 105 genotypes in the years 2017 and 2018, respectively. BLUP values were estimated for 65 genotypes in which the grains were collected in both years. Our results indicate that the *C. arabica* wild accessions have higher genetic variability and higher diversity for grains macronutrient content than the commercial cultivars. More SNPs markers were identified when the GBS data were mapped to the Caturra reference genome (11,230 SNPs) than to the Et39 reference genome (9,991 SNPs). Beagle presented the best performance in markers dataset imputation. The imputation reduced the false discovery rate in GWAS. We identified 5, 12, 13, 6, 6 and 1 markers in coding regions associated with the content of N, P, K, Ca, Mg, and Mg and K in coffee grains, respectively. The proteins identified participates in several metabolic pathways, including nutrient transport, seed germination, flowering time control and plant response to biotic and abiotic stresses. These proteins can be used as targets in further studies for the characterization of metabolic pathways controlling macronutrient accumulation in *C. arabica* plants. Also, the wild *C. arabica* genotypes used in this work can be used to identify favorable alleles to be introduced in

commercial cultivars, and for the selection of promising genotypes with altered levels of macronutrients in their grains than the observed among *C. arabica* commercial cultivars.

Keywords: imputation, missing genotypes, non-model species, biotic and abiotic stress.

Lista de figuras

Capítulo 1

Capítulo 2

Lista de tabelas

the GWAS models and the minimum and maximum values of LOD score, QTN effects, minor allele frequency (MAF), and correlation (r^2) to the associated trait estimated for each marker.

Sumário

1. INTRODUÇÃO

O café é uma das *commodities* agrícolas mais comercializadas no mundo. O produto possui grande importância para o desenvolvimento socioeconômico do Brasil, que é o principal país produtor e exportador de café no mundo. A principal espécie utilizada para a produção comercial de café é *Coffea arabica* L. (café arábica) (ICO 2020).

A espécie *C. arabica* pertence à família Rubiaceae, gênero *Coffea* (Davis et al. 2011). Estudos indicam que a espécie tenha uma origem relativamente recente, estimada entre 10 mil e 600 mil anos atrás (Yu et al. 2011; Scalabrin et al. 2020) e originada na região da Etiópia por meio da hibridação natural entre genótipos ancestrais das espécies *Coffea canephora* Pierre ex A. Froehner e *Coffea eugenioides* S. Moore (Lashermes et al. 1999). *Coffea arabica* é a única espécie alotetratetraploide natural (2n = 4x = 44) e com reprodução predominantemente autógama do gênero *Coffea*, enquanto o restante das espécies foi caracterizado como diploide e a maioria auto incompatível (Charrier e Berthaud 1985).

Embora o centro de origem do café arábica seja a região da Etiópia, o cultivo comercial dessa espécie teve início no Iêmen, região que foi classificada como o centro secundário de dispersão da espécie (Meyer et al. 1968). A partir de plantas cultivadas no Iêmen, surgiram as subpopulações Típica e Bourbon, que deram origem a maior parte das cultivares comerciais de *C. arabica* utilizadas no mundo (Anthony et al. 2002).

A forma de dispersão do cultivo de *C. arabica*, baseado em poucos genótipos, aliada a origem relativamente recente da espécie e o modo de reprodução predominantemente autógamo, contribuíram para a base genética estreita observada atualmente entre as cultivares comerciais (Anthony et al. 2002; Silvestrini et al. 2007; Setotaw et al. 2013). A baixa variabilidade genética entre cultivares comerciais representa um problema para a cultura, pois a maioria das cultivares são suscetíveis à estresses bióticos e abióticos (van der Vossen et al. 2015).

Para ampliar os recursos genéticos a serem explorados pelos programas de melhoramento de *C. arabica*, foram realizadas coletas de genótipos na região do centro de origem da espécie, a Etiópia (Meyer et al. 1968). Análises realizadas em grãos de café revelaram que os acessos selvagens da Etiópia abrigam ampla variabilidade para aspectos morfológicos e de compostos que afetam a qualidade da bebida de café (Gaspari-Pezzopane 2014; Sant'Ana et al. 2018; dos Santos Scholz et al. 2016).

A composição mineral do grão do cafeeiro tem impacto direto na qualidade do produto. No entanto, até o momento, a base genética da composição de macronutrientes em grãos de *C. arabica* ainda é pouco conhecida.

A análise de mapeamento associativo, ou mapeamento por desequilíbrio de ligação (DL), é uma alternativa que pode ser utilizada para a identificação de regiões genômicas associadas a características fenotípicas que possuem controle genético complexo, como o conteúdo de macronutrientes em grãos (Ziegler et al. 2017; Ziegler et al. 2018). Análises de mapeamento associativo que utilizam a informação de marcadores moleculares distribuídos ao longo de regiões do genoma recebem o nome de associação genômica ampla (*genome wide association study* - GWAS).

A técnica de GWAS geralmente é aplicada a populações com genótipos diversos (Hayward et al. 2015), como por exemplo, populações que incluem acessos de coleções de germoplasma (Rafalski 2010). Nessa técnica são identificados marcadores de polimorfismos de nucleotídeo único (*single nucleotide polymorphisms* - SNP) que possuem associação significativa com características fenotípicas complexas (Hayward et al. 2015).

O desenvolvimento das tecnologias de sequenciamento de nova geração contribuiu para o avanço de técnicas destinadas a identificação de marcadores SNPs. Em plantas, a técnica de genotipagem por sequenciamento (*genotyping-by-sequencing* - GBS) tem sido amplamente utilizada para essa finalidade (Rasheed et al. 2017; Nadeem et al. 2018). Nessa técnica, enzimas de restrição são utilizadas para reduzir a complexidade dos genomas, limitando o sequenciamento às regiões que flanqueiam o sítio de corte das enzimas. A ligação de adaptadores específicos às extremidades dos insertos de DNA permite a multiplexação de amostras para o sequenciamento em uma mesma reação, reduzindo assim o custo da genotipagem (Elshire et al. 2011). No entanto, um limitante para o uso de dados de GBS está na proporção de dados ausentes por marcador, que tende a ser alta quando o sequenciamento é realizado com baixa cobertura (Fu 2014; Torkamaneh et al. 2018).

Os dados ausentes podem reduzir a acurácia de análises de associação, como GWAS, contribuindo para a identificação de falsos positivos (Rahimi et al. 2019). A imputação de genótipos é uma alternativa que pode reduzir os prejuízos causados pelos dados ausentes, e consiste na substituição desses dados por genótipos prováveis (Marchini e Howie 2010; Torkamaneh et al. 2018). A imputação pode ser realizada por métodos desenvolvidos especificamente para substituir dados ausentes em painéis de marcadores ou por métodos estatísticos gerais (He et al. 2015; Nazzicari et al. 2016). No primeiro grupo é possível citar o método de imputação de alelos em haplótipos localizados, implementado no *software* Beagle (Browning e Browning 2016). Dentre os métodos estatísticos gerais estão incluídos o método da média ponderada entre os marcadores mais próximos (*K-nearest neighbor* - KNN) e o

método não paramétrico de regressões de florestas aleatórias (*Random forest* - RF) (Rutkoski et al. 2013; Nazzicari et al. 2016).

A acurácia da imputação de marcadores pode ser alterada de acordo com o método de imputação adotado, o genoma de referência utilizado para o mapeamento dos dados de GBS, a proporção de dados ausentes permitida por marcador, o número de marcadores identificados e as características intrínsecas às espécies e populações genotipadas, como por exemplo a relação de parentesco entre os acessos da população (Rutkoski et al. 2013; Torkamaneh e Belzile. 2015; Nazzicari et al. 2016).

Estudos recentes indicam que a imputação de marcadores pode contribuir para o aumento da acurácia na identificação de associações significativas em GWAS, como observado em populações de soja (*Glycine max* L., Torkamaneh e Belzile 2015) e trigo (*Triticum aestivum* L., Rahimi et al. 2019).

No primeiro trabalho de GWAS realizado com a espécie *C. arabica*, em uma população com 107 acessos provenientes do centro de origem da espécie, foram identificados 21 marcadores possivelmente associados à composição de lipídeos nos grãos, compostos que interferem na qualidade da bebida de café (Sant'Ana et al. 2018). Em trabalhos com soja e milho (*Zea mays* L.) foram identificados marcadores associados a regiões genômicas que controlam o acúmulo de nutrientes nos grãos (Ziegler et al. 2017; Ziegler et al. 2018).

Até o momento não foi publicado nenhum estudo em que o desempenho de métodos de imputação tenha sido analisado em painéis de marcadores da espécie *C. arabica*. Além disso, para a mesma espécie também não se sabe qual é o efeito da imputação de marcadores em análises de GWAS. Com isso, os objetivos desse trabalho foram: i) mapear dados de genotipagem de *C. arabica* em genomas de referência da própria espécie; ii) identificar o melhor método de imputação em uma população de 110 genótipos de *C. arabica*; iii) aplicar a técnica de GWAS para identificar regiões genômicas em associação com a composição de nutrientes em grãos; iv) verificar o efeito da imputação de marcadores sobre os resultados de GWAS.

5. CONCLUSÕES GERAIS

Os resultados obtidos nesse trabalho indicam que os genótipos selvagens de *C. arabica* abrigam uma ampla variabilidade para o conteúdo de macronutrientes em grãos. Além disso, as análises genotípicas também confirmaram maior variabilidade genética entre genótipos selvagens quando comparados às cultivares comerciais. Com isso, esse trabalho pode contribuir para a identificação de genótipos com diferentes níveis de concentração de macronutrientes nos grãos a serem introduzidos aos programas de melhoramento do café, visando maior qualidade do grão e da bebida. Portanto, a preservação de genótipos selvagens de *C. arabica* é essencial para o melhoramento da espécie e sua sobrevivência em frente a diferentes fatores bióticos e abióticos.

O alinhamento dos dados de GBS de *C. arabica* à dois genomas de referência completos da própria espécie favoreceu a identificação de um grande número de polimorfismos na população (9000 - 11230 SNPs). O desempenho dos métodos de imputação de marcadores Beagle, KNN e RF foi similar, com aproximadamente 80% de acurácia para os dois alinhamentos. Como *C. arabica* é uma espécie não modelo e não possui painéis de referência para a imputação de marcadores, a acurácia obtida pode ser considerada elevada. Além disso, a imputação de marcadores com o software Beagle favoreceu o controle da identificação de associações espúrias, as quais foram identificadas nos painéis não imputados.

Nesse trabalho também foram identificadas diversas regiões genômicas associadas à composição de cinco macronutrientes (N, P, K, Ca e Mg) em grãos de *C. arabica*. Portanto, a estratégia de utilizar dois genomas para o alinhamento dos dados de GBS se mostrou eficaz, considerando que diferentes regiões foram identificadas através das análises de GWAS a partir de cada genoma de referência.

As proteínas codificadas pelos genes candidatos anotados nesse trabalho desempenham diferentes funções no metabolismo da planta, por isso, sugerimos que esses genes sejam

utilizados como alvos de estudos futuros para a melhor caracterização da base genética da composição de macronutrientes em grãos de *C. arabica*. O presente trabalho contribuiu para a identificação de diversos marcadores associados com a composição de macronutrientes nos grãos que após validados poderão ser utilizados para a seleção assistida por marcadores de genótipos de *C. arabica*.

6. REFERÊNCIAS

Anthony F, Astorga C, Berthaud J (1999) Los recursos genéticos: las bases de una solución genética a los problemas de la caficultura latinoamericana. In: Desafíos de la caficultura centroamericana. IICA-PROMECAFE-CIRAD, San José, pp 369-406.

Anthony F, Bertrand B, Quiros O, et al (2001) Genetic diversity of wild coffee (Coffea arabica L.) using molecular markers. Euphytica 118:53–65. https://doi.org/10.1023/A:1004013815166

Anthony F, Combes MC, Astorga C, et al (2002) The origin of cultivated Coffea arabica L. varieties revealed by AFLP and SSR markers. Theor Appl Genet 104:894–900. https://doi.org/10.1007/s00122-001-0798-8

Ashburner M, Ball CA, Blake JA, et al (2000) Gene Ontology : tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25:25–29. https://doi.org/10.1038/75556

Associação Brasileira da Indústria de Café (ABIC) (2015) Café é a segunda bebida mais consumida no Brasil. http://www.consorciopesquisacafe.com.br/index.php/imprensa/noticias/580-cafe-e-a-segunda-bebida-mais-consumida-no-brasil. Accessed 24 January 2018

Barman AR, Banerjee J (2015) Versatility of germin-like proteins in their sequences, expressions, and functions. Funct Integr Genomics 15:533–548. https://doi.org/10.1007/s10142-015-0454-z

Bazzo BR, Eiras A de L, DeLaat DM, et al (2013) Gene Expression Analysis Suggests Temporal Differential Response to Aluminum in Coffea arabica Cultivars. Trop Plant Biol 6:191–198. https://doi.org/10.1007/s12042-013-9120-6

Berthaud J (1976) Etude cytogénétique d'un haploïde de Coffea arabica L. Café, Cacao, Thé

(Francia) v. 20 (2):91-96.

Bradbury PJ, Zhang Z, Kroon DE, et al (2007) TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Branca A, Paape TD, Zhou P, et al (2011) Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume Medicago truncatula. Proc Natl Acad Sci U S A 108:864–870. https://doi.org/10.1073/pnas.1104032108

Breiman L (2001) Random forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324

Browning BL, Browning SR (2016) Genotype Imputation with Millions of Reference Samples. Am J Hum Genet 98:116–126. https://doi.org/10.1016/j.ajhg.2015.11.020

Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet 81:1084–1097. https://doi.org/10.1086/521987

Caixeta ET, Oliveira AD, Brito GD, Sakiyama N S (2016). Tipos de marcadores moleculares. In: Marcadores moleculares, Editora UFV, Viçosa, pp 9-93.

Catani RA, Pellegrino D, Alcarde, JC, Graner CAF (1967). Variação na concentração e na quantidade de macro e micronutrientes no fruto do cafeeiro, durante o seu desenvolvimento. Anais da escola superior de agricultura Luiz de Queiroz 24:249-263. https://doi.org/10.1590/S0071-12761967000100024

Catchen J, Hohenlohe PA, Bassham S, et al (2013) Stacks: An analysis tool set for population genomics. Mol Ecol 22:3124–3140. https://doi.org/10.1111/mec.12354

Cenci A, Combes MC, Lashermes P (2012) Genome evolution in diploid and tetraploid Coffea species as revealed by comparative analysis of orthologous genome segments. Plant Mol Biol 78:135–145. https://doi.org/10.1007/s11103-011-9852-3

Chan AW, Hamblin MT, Jannink JL (2016) Evaluating imputation algorithms for low-depth genotyping-by-sequencing (GBS) data. PLoS One 11:1–17. https://doi.org/10.1371/journal.pone.0160733

Chardin C, Girin T, Roudier F, et al (2014) The plant RWP-RK transcription factors: Key regulators of nitrogen responses and of gametophyte development. J Exp Bot 65:5577–5587. https://doi.org/10.1093/jxb/eru261

Charrier A, Berthaud J (1985) Botanical classification of coffee. In: Coffee, Springer, Boston, pp 13-47.

Cheek S, Zhang H, Grishin N V. (2002) Sequence and structure classification of kinases. J Mol Biol 320:855–881. https://doi.org/10.1016/S0022-2836(02)00538-7

Cheng Q, Sun MF, Kravtsov DV, Aktimur LA, Gailani D (2003) Factor XI apple domains and protein dimerization. Journal of Thrombosis and Haemostasis, 1(11): 2340-2347. https://doi.org/10.1046/j.1538-7836.2003.00418.x

Ching ADA, Caldwell KS, Jung M, Dolan M, Smith OSH, Tingey S et al. (2002). SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC genetics, 3(1), 19.

Clarindo WR, Carvalho CR (2008). Comparison of the Coffea canephora and C. arabica karyotype based on chromosomal DNA content. Plant Cell Rep, 28:73-81. https://doi.org/10.1007/s00299-008-0621-y

Clifford MN (1985) Chemical and physical aspects of green coffee and coffee products. In: Coffee Botany, Biochemistry and Production of Beans and Beverage. Springer, Boston, pp 305-374.

Companhia Nacional de Abastecimento (CONAB) (2018) Acompanhamento da Safra Brasileira: Café, Safra 2018 - Quarto Levantamento Dezembro 2018. Brasília, pp 1-84.

Companhia Nacional de Abastecimento (CONAB) (2019) Acompanhamento da Safra

Brasileira: Café, Safra 2019 - Terceiro Levantamento Setembro 2019. Brasília, pp 1-48.

da Silva BSR, Sant'Ana GC, Chaves CL, et al (2019) Population structure and genetic relationships between Ethiopian and Brazilian Coffea arabica genotypes revealed by SSR markers. Genetica 147:205–216. https://doi.org/10.1007/s10709-019-00064-4

Danecek P et al. (2011). The variant call format and VCFtools. Bioinformatics, 27(15):2156-2158. https://doi.org/10.1093/bioinformatics/btr330

Das S, Abecasis GR, Browning BL (2018) Genotype Imputation from Large Reference Panels. Annu Rev Genomics Hum Genet 19:73–96. https://doi.org/10.1146/annurev-genom-083117-021602

Davey JW, Hohenlohe PA, Etter PD, et al (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12:499–510. https://doi.org/10.1038/nrg3012

Davis AP, Chadburn H, Moat J, O'Sullivan R, Hargreaves S, Lughadha EN (2019) High extinction risk for wild coffee species and implications for coffee sector sustainability. Science advances, 5(1), eaav3473. https://doi.org/10.1126/sciadv.aav3473

Davis AP, Gole TW, Baena S, Moat J (2012). The impact of climate change on indigenous arabica coffee (Coffea arabica): predicting future trends and identifying priorities. PloS one, 7(11). https://doi.org/10.1371/journal.pone.0047981

Davis AP, Tosh J, Ruch N, Fay MF (2011) Growing coffee: Psilanthus (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of Coffea. Bot J Linn Soc 167:357–377. https://doi.org/10.1111/j.1095-8339.2011.01177.x

de Kochko A. and ACGC (2018) Deciphering the Allotetraploid Genome of Coffea arabica L. In Plant and Animal Genome Conference XXVI (Januray 13- 17, 2018). San Diego, CA, USA.

Deschamps S, Llaca V, May GD (2012) Genotyping-by-sequencing in plants. Biology (Basel) 1:460–483. https://doi.org/10.3390/biology1030460

Dias Chaves J, Sarruge J (1984) Alteraçoes nas concentraçoes de macronutrientes nos frutos e folhas do cafeeito durante um ciclo productivo. Pesqui agropecu bras 19:427–432

Diniz I et al. (2015) Unveilling the involvement of oxidases in the resistance of Coffea sp. to Colletotrichum kahawae. In: Proceedings of 25th International Conference on Coffee Science (ASIC). Armenia, Colombia.

Disch S, Anastasiou E, Sharma VK, et al (2006) The E3 Ubiquitin Ligase BIG BROTHER Controls Arabidopsis Organ Size in a Dosage-Dependent Manner. Curr Biol 272–279. https://doi.org/10.1016/j.cub.2005.12.026

Dominski Z (2007) Nucleases of the metallo-β-lactamase family and their role in DNA and RNA metabolism. Crit Rev Biochem Mol Biol 42:67–93. https://doi.org/10.1080/10409230701279118

dos Santos TB, Meda AR, Sitta RB, et al (2015) Nutritional characterization of Arabica coffee accession from Ethiopia. Coffee Sci 10:10–19. https://doi.org/10.25186/cs.v10i1.716

dos Santos Scholz MB, Kitzberger CSG, Pagiatto NF, et al (2016) Chemical composition in wild ethiopian Arabica coffee accessions. Euphytica 209:429–438. https://doi.org/10.1007/s10681-016-1653-y

Dunwell JM, Purvis A, Khuri S (2004) Cupins: The most functionally diverse protein superfamily? Phytochemistry 65:7–17. https://doi.org/10.1016/j.phytochem.2003.08.016

Eklund A (2016). beeswarm: The Bee Swarm Plot, an Alternative to Stripchart. R package version 0.2.3. https://CRAN.R-project.org/package=beeswarm

Elshire RJ, Glaubitz JC, Sun Q, et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:1–10. https://doi.org/10.1371/journal.pone.0019379

Food and Agriculture Organization of the United Nations (FAO) (2015) FAO Coffee Pocketbook 2015, Rome.

Frichot E, Mathieu F, Trouillon T, Bouchard G, François O (2014). Fast and efficient estimation of individual ancestry coefficients. Genetics, 196(4), 973-983. https://doi.org/10.1534/genetics.113.160572

Frichot E, François O (2015) LEA: An R package for landscape and ecological association studies. Methods Ecol Evol 6:925–929. https://doi.org/10.1111/2041-210X.12382

Fu YB (2014) Genetic diversity analysis of highly incomplete snp genotype data with imputations: An empirical assessment. G3 Genes, Genomes, Genet 4:891–900. https://doi.org/10.1534/g3.114.010942

Garcia ALA, de CARVALHO CHS, Garcia AWR (2009) Extração de nutrientes em cafeeiros da espécie Coffea arabica. In Embrapa Café-Artigo em anais de congresso (ALICE). In: Congresso Brasileiro de Pesquisa Cafeeiras, 34., 2008, Caxambú. Anais. Brasília, DF: Embrapa Café, 2009.

Gaspari-Pezzopane CD, Medina Filho HP, Bordignon R (2004). Variabilidade genética do rendimento intrínseco de grãos em germoplasma de Coffea. Bragantia, 63(1), 39-54. https://doi.org/10.1590/S0006-87052004000100005

Glaubitz JC, Casstevens TM, Lu F, et al (2014) TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS One 9(2): e90346. https://doi.org/10.1371/journal.pone.0090346

Gole TW (2003) Vegetation of the Yayu Forest in SW Ethiopia: impacts of human use and implications for in situ conservation of wild Coffea arabica L. populations. Ph.D. Thesis. University of Bonn, Germany.

Gong X, Shen L, Peng YZ, et al (2017) DNA Topoisomerase I a Affects the Floral Transition. Plant physiol, 173:642–654. https://doi.org/10.1104/pp.16.01603

Gore M, Bradbury P, Hogers R, et al (2007) Evaluation of target preparation methods for single-feature polymorphism detection in large complex plant genomes. Crop Sci 47:135–148. https://doi.org/10.2135/cropsci2007.02.0085tpg

Guerra-Guimarães L, Tenente R, Pinheiro C, et al (2015) Proteomic analysis of apoplastic fluid of Coffea arabica leaves highlights novel biomarkers for resistance against Hemileia vastatrix. Front Plant Sci 6:1–16. https://doi.org/10.3389/fpls.2015.00478

Guerreiro-Filho O et al. (2008) Origem e classificação botânica do cafeeiro. In: Cultivares de café: origem, características e recomendações, Embrapa, Brasília, pp 27-33.

Guillaumet JL, Hallé F (1978) Echantillonnage du matériel Coffea arabica récolté en Ethiopie. Bulletin IFCC, 14:13-18.

Guimarães PT, Reis THP (2010) Nutrição e adubação do cafeeiro. In: Café arábica do plantio à colheita. Epamig, Lavras, pp 343-414.

Harrel Jr F E et al. (2019) Hmisc: Harrel Miscellaneous. R package version 4.2-0. https://CRAN.R-project.org/package=Hmisc

Hayward A C, Tollenaere R, Dalton-Morgan J, Batley J (2015) Molecular marker application in plants. In: Plant Genotyping Methods and Protocols, Humana Press, New York, pp 13-27.

He J, Zhao X, Laroche A, et al (2014) Genotyping-by-sequencing (GBS), An ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front Plant Sci 5:1–8. https://doi.org/10.3389/fpls.2014.00484

He S, Zhao Y, Mette MF, et al (2015) Prospects and limits of marker imputation in quantitative genetic studies in European elite wheat (Triticum aestivum L.). BMC Genomics 16:1–12. https://doi.org/10.1186/s12864-015-1366-y

Healey A, Furtado A, Cooper T, Henry RJ (2014) Protocol : a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. Plant

Methods, 1–8. https://doi.org/10.1186/1746-4811-10-21

Hu X, Song F, Zheng Z (2006) Molecular cloning and expression analysis of rice OsTVLP1, encoding a protein with similarity to TGF-β receptor interacting proteins and vacuolar assembly Vam6p/Vps39p proteins. DNA Seq 17:152–158. https://doi.org/10.1080/10425170600700212

Hu Z, Olatoye MO, Marla S, Morris GP (2019) An integrated genotyping-by-sequencing polymorphism map for over 10,000 sorghum genotypes. The Plant Genome, 12(1). https://doi.org/10.3835/plantgenome2018.06.0044

Huang X, Han B (2014) Natural Variations and Genome-Wide Association Studies in Crop Plants. Annu Rev Plant Biol 65:531–551. https://doi.org/10.1146/annurev-arplant-050213-035715

International Coffee Organization (ICO) (2019a) Coffee market report October 2019. http://www.ico.org/documents/cy2019-20/cmr-1019-e.pdf. Accessed 27 November 2019

International Coffee Organization (ICO) (2019b) Coffee Development Report 2019. Growing for prosperity. United Nations, New York Geneva 1–84

International Coffee Organization (ICO) (2020) Trade Statistics. Coffee production by exporting countries. http://www.ico.org/trade_statistics.asp?section=Statistics. Accessed 01 February 2020

Ilyas M, Rasheed A, Mahmood T (2016) Functional characterization of germin and germin-like protein genes in various plant species using transgenic approaches. Biotechnol Lett 38:1405–1421. https://doi.org/10.1007/s10529-016-2129-9

Imai A, Nonaka K, Kuniga T, et al (2018) Genome-wide association mapping of fruit-quality traits using genotyping-by-sequencing approach in citrus landraces, modern cultivars, and breeding lines in Japan. Tree Genet Genomes 14(2):24. https://doi.org/10.1007/s11295-018-1238-0

Jones P, Binns D, Chang H, et al (2014) InterProScan 5 : genome-scale protein function classification. Bioinformatics, 30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Katuuramu DN, Hart JP, Porch TG, et al (2018) Genome-wide association analysis of nutritional composition-related traits and iron bioavailability in cooked dry beans (Phaseolus vulgaris L.). Mol Breed 38(4):44. https://doi.org/10.1007/s11032-018-0798-x

Kaye Y, Golani Y, Singer Y, et al (2011) Inositol polyphosphate 5-phosphatase7 regulates the production of reactive oxygen species and salt tolerance in arabidopsis. Plant Physiol 157:229–241. https://doi.org/10.1104/pp.111.176883

Khan MA, Korban SS (2012) Association mapping in forest trees and fruit crops. J Exp Bot 63:4045–4060. https://doi.org/10.1093/jxb/ers105

Kim C, Guo H, Kong W, et al (2016) Application of genotyping by sequencing technology to a variety of crop breeding programs. Plant Sci 242:14–22. https://doi.org/10.1016/j.plantsci.2015.04.016

Kishor PBK, Hima Kumari P, Sunita MSL, Sreenivasulu N (2015) Role of proline in cell wall synthesis and plant development and its implications in plant ontogeny. Front Plant Sci 6:1–17. https://doi.org/10.3389/fpls.2015.00544

Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: A review. Plant Methods 9:1–9. https://doi.org/10.1186/1746-4811-9-29

Kumar A, Batra R, Gahlaut V, et al (2018) Genome-wide identification and characterization of gene family for RWP-RK transcription factors in wheat (Triticum aestivum L .). PloS One 13(12) https://doi.org/10.1371/journal.pone.0208409

Kumar S, Banks TW, Cloutier S (2012) SNP discovery through next-generation sequencing and its applications. Int J Plant Genomics 2012. https://doi.org/10.1155/2012/831460

Kupriyanova E V., Albert E V., Bliznina AI, et al (2017) Arabidopsis DNA topoisomerase I

alpha is required for adaptive response to light and flower development. Biol Open 6:832–843. https://doi.org/10.1242/bio.024422

Kusano M, Fukushima A, Redestig H, Saito K (2011) Metabolomic approaches toward understanding nitrogen metabolism in plants. J Exp Bot 62:1439–1453. https://doi.org/10.1093/jxb/erq417

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923

Langmead B, Wilks C, Antonescu V, Charles R (2019) Scaling read aligners to hundreds of threads on general-purpose processors. Bioinformatics 35:421–432. https://doi.org/10.1093/bioinformatics/bty648

Lashermes P, Combes M, Robert J, et al (1999) Molecular characterisation and origin of the Coffea arabica L . genome. Mol Gen Genet MGG, 259–266. https://doi.org/10.1007/s004380050965

Laviola BG, Martinez HEP, De Souza RB, Víctor Hugo Alvarez V (2007) Dinâmica de cálcio e magnésio em folhas e frutos de coffea arabica. Rev Bras Cienc do Solo 31:319–329. https://doi.org/10.1590/s0100-06832007000200014

Li L, Xu X, Chen, C., & Shen, Z. (2016) Genome-wide characterization and expression analysis of the germin-like protein family in rice and Arabidopsis. International journal of molecular sciences, 17(10): 1622. https://doi.org/10.3390/ijms17101622

Lin H, Goodenough UW (2007) Gametogenesis in the Chlamydomonas reinhardtii minus Mating Type Is Controlled by Two Genes, MID and MTD1. Genetics, 1:913–925. https://doi.org/10.1534/genetics.106.066167

Lipka AE, Kandianis CB, Hudson ME, et al (2015) From association to prediction: Statistical methods for the dissection and selection of complex traits in plants. Curr Opin Plant Biol 24:110–118. https://doi.org/10.1016/j.pbi.2015.02.010

Lu F, Lipka AE, Glaubitz J, et al (2013) Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol. PLoS Genet 9(1). https://doi.org/10.1371/journal.pgen.1003215

Lunetti P, Cappello AR, Marsano RM, et al (2013) Mitochondrial glutamate carriers from Drosophila melanogaster: Biochemical, evolutionary and modeling studies. Biochim Biophys Acta - Bioenerg 1827:1245–1255. https://doi.org/10.1016/j.bbabio.2013.07.002

Malavolta E (1986) Nutrição mineral e adubação do cafeeiro – passado, presente e perspectiva. In: Nutrição e adubação do cafeeiro. Instituto de Potassa & Fosfato (EUA), Piracicaba, pp 138-178.

Marchini J, Howie B (2010) Genotype imputation for genome-wide association studies. Nat Rev Genet 11:499–511. https://doi.org/10.1038/nrg2796

Matiello JB et al. (2015) Cultura do café no Brasil: manual de recomendações, edição 2015. MAPA, Rio de Janeiro.

McClure KA, Gardner KM, Douglas GM, et al (2018) A Genome-Wide Association Study of Apple Quality and Scab Resistance. Plant Genome 11:1–14. https://doi.org/10.3835/plantgenome2017.08.0075

Medina Filho H, Bordingnon R, Carvalho CHS (2008) Desenvolvimento de novas cultivares de café arábica. In: Cultivares de café: origem, características e recomendações, Embrapa, Brasília, pp 79-101.

Mendes ANG et al. (2008) História das primeiras cultivares de café plantadas no Brasil. In: Cultivares de café: origem, características e recomendações, Embrapa, Brasília, pp 69-78.

Merot-L'anthoene V, Tournebize R, Darracq O, et al (2019) 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 Biotechnol J 17:1418–1430. https://doi.org/10.1111/pbi.13066

Meyer F et al. (1968) FAO Coffee mission to Ethiopia, 1964:1965. Food and agriculture organization of the United Nations.

Minamikawa MF, Nonaka K, Kaminuma E, et al (2017) Genome-wide association study and genomic prediction in citrus: Potential of genomics-assisted breeding for fruit quality traits. Sci Rep 7:1–2. https://doi.org/10.1038/s41598-017-05100-x

Mindrebo JT, Nartey CM, Seto Y, et al (2016) Unveiling the functional diversity of the alpha/beta hydrolase superfamily in the plant kingdom. Curr Opin Struct Biol 41:233–246. https://doi.org/10.1016/j.sbi.2016.08.005

Miyazawa M et al. (1999) Análise química de tecido vegetal. In: Manual de análises químicas de solos, plantas e fertilizantes, Embrapa, Brasília, pp 171-223.

Moat J, Gole TW, Davis AP (2019) Least concern to endangered: Applying climate change projections profoundly influences the extinction risk assessment for wild Arabica coffee. Glob Chang Biol 25:390–403. https://doi.org/10.1111/gcb.14341

Moncada MDP, Tovar E, Montoya JC, et al (2016) A genetic linkage map of coffee (Coffea arabica L.) and QTL for yield, plant height, and bean size. Tree Genet Genomes 12:1–17. https://doi.org/10.1007/s11295-015-0927-1

Müller R, Morant M, Jarmer H, Nilsson L, Nielsen T H (2007) Genome-wide analysis of the Arabidopsis leaf transcriptome reveals interaction of phosphate and sugar metabolism. Plant Physiology, 143(1):156-171. https://doi.org/10.1104/pp.106.090167

Murachelli AG, Ebert J, Basquin C, et al (2012) The structure of the ASAP core complex reveals the existence of a Pinin-containing PSAP complex. Nat Struct Mol Biol 19:378–386. https://doi.org/10.1038/nsmb.2242

Nadeem MA, Nawaz MA, Shahid MQ, et al (2018) DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnol Biotechnol Equip 32:261–285.

https://doi.org/10.1080/13102818.2017.1400401

Nazzicari N, Biscarini F, Cozzi P, et al (2016) Marker imputation efficiency for genotyping-by-sequencing data in rice (Oryza sativa) and alfalfa (Medicago sativa). Mol Breed 36:1–16. https://doi.org/10.1007/s11032-016-0490-y

Ogura T, Busch W (2015) From phenotypes to causal sequences: Using genome wide association studies to dissect the sequence basis for variation of plant development. Curr Opin Plant Biol 23:98–108. https://doi.org/10.1016/j.pbi.2014.11.008

Patel DA, Zander M, Dalton-Morgan J, Batley J (2015). Advances in plant genotyping: where the future will take us. In: Plant genotyping. Humana Press, New York, pp 1-11.

Pereira AA, Carvalho GR, Moura WM, Botelho CE, Rezende JC, Oliveira ACB, Silva FL (2010). Cultivares: origem e suas características. Café arábica do plantio à colheita. Epamig, Lavras, pp167-221.

Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7(2). https://doi.org/10.1371/journal.pone.0032253

Poland JA, Rife TW (2012) Genotyping-by-sequencing for plant breeding and genetics. Plant Genome 5:92–102. https://doi.org/10.3835/plantgenome2012.05.0005

R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rafalski JA (2010) Association genetics in crop improvement. Curr Opin Plant Biol 13:174–180. https://doi.org/10.1016/j.pbi.2009.12.004

Rahimi Y, Bihamta MR, Taleei A, et al (2019) Genome-wide association study of agronomic traits in bread wheat reveals novel putative alleles for future breeding programs. BMC Plant Biol 19:1–19. https://doi.org/10.1186/s12870-019-2165-4

Rasheed A, Hao Y, Xia X, et al (2017) Crop Breeding Chips and Genotyping Platforms:

Progress, Challenges, and Perspectives. Mol Plant 10:1047–1064. https://doi.org/10.1016/j.molp.2017.06.008

Resende MDV (2016) Software Selegen-REML/BLUP: a useful tool for plant breeding. Crop Breed Appl Biotechnol 16:330–339. https://doi.org/10.1590/1984

Rutkoski JE, Poland J, Jannink JL, Sorrells ME (2013) Imputation of unordered markers and the impact on genomic selection accuracy. G3 Genes, Genomes, Genet 3:427–439. https://doi.org/10.1534/g3.112.005363

Salmona J, Dussert S, Descroix F, De Kochko A, Bertrand B, Joët T (2008). Deciphering transcriptional networks that govern Coffea arabica seed development using combined cDNA array and real-time RT-PCR approaches. Plant Mol Biol, 66:105-124. https://doi.org/10.1007/s11103-007-9256-6

Sanchez R, Zhou MM (2011) The PHD finger: A versatile epigenome reader. Trends Biochem Sci 36:364–372. https://doi.org/10.1016/j.tibs.2011.03.005

Sant'Ana GC, Pereira LFP, Pot D, et al (2018) Genome-wide association study reveals candidate genes influencing lipids and diterpenes contents in Coffea arabica L. Sci Rep 8:1–12. https://doi.org/10.1038/s41598-017-18800-1

Scalabrin S, Toniutti L, Di Gaspero G et al (2020). A single polyploidization event at the origin of the tetraploid genome of Coffea arabica is responsible for the extremely low genetic variation in wild and cultivated germplasm. Sci Rep, 10(1):1-13. https://doi.org/10.1038/s41598-020-61216-7

Schneider C, Anderson JT, Tollervey D (2007) The Exosome Subunit Rrp44 Plays a Direct Role in RNA Substrate Recognition. Mol Cell 27:324–331. https://doi.org/10.1016/j.molcel.2007.06.006

Schwender H (2007) Statistical Analysis of Genotype and Gene Expression Data. Dissertation, University of Dortmund.

Schwender H, Fritsch A (2013) scrime: Analysis of High-Dimensional Categorical Data such as SNP Data. R package version 1.3.3. https://CRAN.R-project.org/package=scrime

Sengupta D, Naik D, Reddy AR (2015) Plant aldo-keto reductases (AKRs) as multi-tasking soldiers involved in diverse plant metabolic processes and stress defense: A structure-function update. J Plant Physiol 179:40–55. https://doi.org/10.1016/j.jplph.2015.03.004

Sera T (2001) Coffee Genetic Breeding at IAPAR. Crop Breed Appl Biotechnol 1:179–199. https://doi.org/10.13082/1984-7033.v01n02a08

Setotaw TA, Caixeta ET, Pereira AA, et al (2013) Coefficient of parentage in Coffea arabica L. cultivars grown in Brazil. Crop Sci 53:1237–1247. https://doi.org/10.2135/cropsci2012.09.0541

Silvarolla MB, Mazzafera P, Fazuoli LC (2004) A naturally decaffeinated arabica coffee. Nature 429(6994):826. https://doi.org/10.1038/429826a

Silvestrini M, Junqueira MG, Favarin AC, et al (2007) Genetic diversity and structure of Ethiopian, Yemen and Brazilian Coffea arabica L. accessions using microsatellites markers. Genet Resour Crop Evol 54:1367–1379. https://doi.org/10.1007/s10722-006-9122-4

Slatkin M (2008) Linkage disequilibrium - Understanding the evolutionary past and mapping the medical future. Nat Rev Genet 9:477–485. https://doi.org/10.1038/nrg2361

Stekhoven DJ, Bühlmann P (2012) Missforest-Non-parametric missing value imputation for mixed-type data. Bioinformatics 28:112–118. https://doi.org/10.1093/bioinformatics/btr597

Swarts K, Li H, Alberto Romero Navarro J, et al (2014) Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. Plant Genome 7:1–12. https://doi.org/10.3835/plantgenome2014.05.0023

Tamba CL, Ni YL, Zhang YM (2017) Iterative sure independence screening EM-Bayesian

LASSO algorithm for multi-locus genome-wide association studies. PLoS Comput Biol 13:1–20. https://doi.org/10.1371/journal.pcbi.1005357

Tamba CL, Zhang Y (2018) A fast mrMLM algorithm for multi-locus genome-wide association studies. bioRxiv, 341784 https://doi.org/10.1101/341784

Tedeschi F, Rizzo P, Rutten T, et al (2017) RWP-RK domain-containing transcription factors control cell differentiation during female gametophyte development in Arabidopsis. New Phytol 213:1909–1924. https://doi.org/10.1111/nph.14293

Teo YY (2008) Common statistical issues in genome-wide association studies: A review on power, data quality control, genotype calling and population structure. Curr Opin Lipidol 19:133–143. https://doi.org/10.1097/MOL.0b013e3282f5dd77

Tesseema A, Alamerew S, Kufa T, Garedew W (2011) Variability and association of quality and biochemical attributes in some promising Coffea arabica germplam collections in Southwestern Ethiopia. Int J Plant Breed and Genetics, 4:302-316. https://doi.org/10.3923/ijpbg.2011.302.316

The Gene Ontology Consortium (2019) The Gene Ontology Resource : 20 years and still GOing strong. Nucleic Acids Res, 47:330–338. https://doi.org/10.1093/nar/gky1055

Torkamaneh D, Belzile F (2015) Scanning and filling: Ultra-dense SNP genotyping combining genotyping-by-sequencing, SNP array and whole-genome resequencing data. PLoS One 10:1–16. https://doi.org/10.1371/journal.pone.0131533

Torkamaneh D, Boyle B, Belzile F (2018) Efficient genome-wide genotyping strategies and data integration in crop plants. Theor Appl Genet 131:499–511. https://doi.org/10.1007/s00122-018-3056-z

Tran HTM, Vargas CAC, Slade Lee L, et al (2017) Variation in bean morphology and biochemical composition measured in different genetic groups of arabica coffee (Coffea arabica L.). Tree Genet Genomes 13(3):54. https://doi.org/10.1007/s11295-017-1138-8

Troyanskaya O, Cantor M, Sherlock G, et al (2001) Missing value estimation methods for DNA microarrays. Bioinformatics 17:520–525. https://doi.org/10.1093/bioinformatics/17.6.520

Usman MG, Rafii MY, Martini MY, et al (2017) Molecular analysis of Hsp70 mechanisms in plants and their function in response to stress. Biotechnol Genet Eng Rev 33:26–39. https://doi.org/10.1080/02648725.2017.1340546

van der Vossen H, Bertrand B, Charrier A (2015) Next generation variety development for sustainable production of arabica coffee (Coffea arabica L.): a review. Euphytica 204:243–256. https://doi.org/10.1007/s10681-015-1398-z

Veerappan V, Wang J, Kang M, et al (2012) A novel HSI2 mutation in Arabidopsis affects the PHD-like domain and leads to derepression of seed-specific gene expression. Planta 236:1–17. https://doi.org/10.1007/s00425-012-1630-1

Vidal RO, Mondego JMC, Pot D, et al (2010) 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 Physiol 154:1053–1066. https://doi.org/10.1104/pp.110.162438

Vieira LGE et al. (2006). Brazilian coffee genome project: an EST-based genomic resource. Braz. J. of Plant Physiol., 18(1): 95-108. https://doi.org/10.1590/S1677-04202006000100008

Vilhjálmsson BJ, Nordborg M (2013) The nature of confounding in genome-wide association studies. Nat Rev Genet 14:1–2. https://doi.org/10.1038/nrg3382

Voiniciuc C, Dean GH, Jonathan S. Griffiths, et al (2013) FLYING SAUCER1 Is a Transmembrane RING E3 Ubiquitin Ligase That Regulates the Degree of Pectin Methylesteri fi cation in Arabidopsis Seed Mucilage. The Plant Cell, 25:944–959. https://doi.org/10.1105/tpc.112.107888

Vos SM, Tretter EM, Schmidt BH, Berger JM (2011) All tangled up: how cells direct, manage

and exploit topoisomerase function. Nature reviews Molecular cell biology, 12(12), 827-841. https://doi.org/ 10.1038/nrm3228

Waki T, Hiki T, Watanabe R (2011) Report The Arabidopsis RWP-RK Protein RKD4 Triggers Gene Expression and Pattern Formation in Early Embryogenesis. Curr Biol, 4:1277–1281. https://doi.org/10.1016/j.cub.2011.07.001

Wang SB, Feng JY, Ren WL, et al (2016) Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. Sci Rep 6:1–10. https://doi.org/10.1038/srep19444

Wen YJ, Zhang H, Ni YL, et al (2018) Methodological implementation of mixed linear models in multi-locus genome-wide association studies. Brief Bioinform 19:700–712. https://doi.org/10.1093/bib/bbw145

Wickham H (2016) Elegant Graphics for Data Analysis, Springer-Verlag, New York.

Wu W, Cheng Z, Liu M, et al (2014) C3HC4-type RING finger protein NbZFP1 is involved in growth and fruit development in Nicotiana benthamiana. PLoS One 9(6). https://doi.org/10.1371/journal.pone.0099352

Xu, S (2010). An expectation–maximization algorithm for the Lasso estimation of quantitative trait locus effects. Heredity, 105(5):483-494. https://doi.org/ 10.1038/hdy.2009.180

Yokoyama Y, Kobayashi S, Kidou S ichiro (2019) PHD type zinc finger protein PFP represses flowering by modulating FLC expression in Arabidopsis thaliana. Plant Growth Regul 88:49–59. https://doi.org/10.1007/s10725-019-00487-1

Yu J, Pressoir G, Briggs WH, et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208. https://doi.org/10.1038/ng1702

Yu Q, Guyot R, Kochko A De, et al (2011) Micro-collinearity and genome evolution in the vicinity of an ethylene receptor gene of cultivated diploid and allotetraploid coffee species

(Coffea). 305–317. https://doi.org/10.1111/j.1365-313X.2011.04590.x

Zhang C, Dong SS, Xu JY, et al (2019) PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics 35:1786–1788. https://doi.org/10.1093/bioinformatics/bty875

Zhang J, Feng JY, Ni YL, et al (2017) PLARmEB: Integration of least angle regression with empirical Bayes for multilocus genome-wide association studies. Heredity 118:517–524. https://doi.org/10.1038/hdy.2017.8

Zhang N, Yao LL, Li X dong (2018b) Regulation of class V myosin. Cell Mol Life Sci 75:261–273. https://doi.org/10.1007/s00018-017-2599-5

Zhang Y, Zheng Q, Sun C, et al (2016) Palmitoylation of the Cysteine Residue in the DHHC Motif of a Palmitoyl Transferase Mediates Ca2+ Homeostasis in Aspergillus. PLoS Genet 12:1–30. https://doi.org/10.1371/journal.pgen.1005977

Zhang Y, Li P, Ren W, Ni Y, Zhang Y (2018a) mrMLM: Multi-Locus Random-SNP-Effect Mixed Linear Model Tools for Genome-Wide Association Study. R package version 3.1. https://CRAN.R-project.org/package=mrMLM

Zhao Y, Jian Y, Liu Z, et al (2017) Network analysis reveals the recognition mechanism for dimer formation of bulb-type lectins. Sci Rep 7:1–9. https://doi.org/10.1038/s41598-017-03003-5

Zhong P, Li J, Luo L, et al (2019) TOP1 α regulates FLOWERING LOCUS C expression by coupling histone modification and transcription machinery. Development 146(4):dev167841. https://doi.org/10.1242/dev.167841

Ziegler G, Kear PJ, Wu D, et al (2017) Elemental accumulation in kernels of the maize nested association mapping panel reveals signals of gene by environment interactions. bioRxivd. https://doi.org/10.1101/164962

Ziegler G, Nelson R, Granada S, et al (2018) Genomewide association study of ionomic traits

on diverse soybean populations from germplasm collections. Plant Direct 2:e00033. https://doi.org/10.1002/pld3.33