

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Genetic Diversity of *Coffea arabica* L.: A Genomic Approach

José Luis Spinoso-Castillo, Tarsicio Corona-Torres,  
Esteban Escamilla-Prado, Victorino Morales-Ramos,  
Víctor Heber Aguilar-Rincón and Gabino García-de los Santos

## Abstract

*Coffea arabica* L. produces a high-quality beverage, with pleasant aroma and flavor, but diseases, pests and abiotic stresses often affect its yield. Therefore, improving important agronomic traits of this commercial specie remains a target for most coffee improvement programs. With advances in genomic and sequencing technology, it is feasible to understand the coffee genome and the molecular inheritance underlying coffee traits, thereby helping improve the efficiency of breeding programs. Thanks to the rapid development of genomic resources and the publication of the *C. canephora* reference genome, third-generation markers based on single-nucleotide polymorphisms (SNPs) have gradually been identified and assayed in *Coffea*, particularly in *C. arabica*. However, high-throughput genotyping assays are still needed in order to rapidly characterize the coffee genetic diversity and to evaluate the introgression of different cultivars in a cost-effective way. The DArTseq™ platform, developed by Diversity Arrays Technology, is one of these approaches that has experienced an increasing interest worldwide since it is able to generate thousands of high quality SNPs in a timely and cost-effective manner. These validated SNP markers will be useful to molecular genetics and for innovative approaches in coffee breeding.

**Keywords:** *Coffea* spp., high throughput genotyping, molecular markers, plant breeding, DArTseq

## 1. Introduction

Coffee is an important crop and the second most traded commodity in the world (after petroleum) providing a living to more than 125 million people. Commercial coffee production is controlled by only two species belonging to the *Coffea* genus: *Coffea arabica* L. (Arabica coffee) and *Coffea canephora* Pierre ex A. Froehner (Robusta coffee), which supplied 60 and 40% of the world coffee production in 2018/19, respectively [1]. Although *C. canephora* does not have the cup quality of the more popular *C. arabica*, it continues to be widely grown, especially in regions where farming is low intensive because of its tolerance to diseases and pests as well as abiotic stresses [2].

*C. arabica* produces a high-quality beverage, with pleasant aroma and flavor, but a range of biotic and abiotic stresses often affect its yield [3]. Therefore, improving important agronomic traits of both commercial species remains a target for most

coffee breeding programs. Advances in genomic and sequencing technology, make possible to understand the coffee genome and the molecular inheritance underlying coffee traits, thereby helping improve the efficiency of coffee breeding [3].

The development of new genomic tools can help us explore, more deeply and more precisely, the genomic diversity at intra and inter-specific levels [4]. Two examples of high-throughput platforms include next-generation sequencing (NGS) [5] and the development of DNA microarrays [6]. Compared to a whole-genome sequencing methodology, an SNP array approach provides time-effective, low-cost and more straightforward genotyping technology for germplasm screening [7, 8].

Thanks to the rapid development of genomic resources and the publication of the reference genome [9], third-generation markers based on single-nucleotide polymorphisms (SNPs) have gradually been identified and assayed in *Coffea*, particularly in *C. arabica* [10, 11].

## 2. Genetic diversity of *Coffea arabica* L.

The *Coffea* genus belongs to the Rubiaceae family that includes around 124 species, most of them are diploids ( $2n = 2x = 22$ ). The only allotetraploid is *C. arabica* L. with  $2n = 4x = 44$  [12], which was originated from the natural cross between *Coffea eugenoides* S. Moore and *C. canephora* Pierre ex A. Froehner [13], *C. arabica* is the only self-fertile among the other cultivated species. This specie is genetically less diverse when compared to the diploid species [14, 15], a situation that has been associated with its susceptibility to the common coffee diseases [16].

*C. arabica* is mainly native to the highlands of southwestern Ethiopia, South Sudan (Boma plateau), and north Kenya (Mount Marsabi). *C. arabica* cultivars grown all around the world are derived from either 'Typica' or 'Bourbon' genetic base [17]. Studies report wide agronomic diversity of Arabica coffee accessions collected in these regions of Ethiopia regarding leaf size, height, biotic and abiotic stresses tolerance and yield [18, 19]. In addition, studies using molecular markers indicated the presence of higher genetic variability of Ethiopian (ET) accessions compared with cultivars, demonstrating the potential of these accessions for breeding purposes [10, 20–23]. These accessions also showed a great variability of metabolite profiles contents of coffee beans for cup quality improvement [10, 24].

The assessment of population structure and genetic relationships of these ET accessions, among themselves and in relation to traditional cultivars is fundamental for efficient use of genetic diversity of these genotypes in Arabica coffee breeding programs [25]. However, selection of genetically diverse parental lines based on morphological and agronomic traits is often difficult because of a high degree of morphological similarities [26].

During the past 30 years, molecular markers have been increasingly used in germplasm diversity assessment of various crops [27, 28]. The molecular information allows gaining insight into the genetic structure of individual genotypes, and eventually helps in accurate selection of superior genotypes for maximizing selection gains [29].

## 3. *C. arabica* diversity assessment by molecular markers

Several works on the assessment of Arabica genetic diversity have been carried out with different results. Generally, among different types of material (cultivars, accessions, hybrids, and spontaneous genotypes) practically all studies show a very

low genetic variation by using different marker systems [3]. Arabica's genetic diversity has been evaluated by a range of molecular markers, such as Random Amplified Polymorphic DNA (RAPD) [30, 31], Inter Simple Sequence Repeat (ISSR) [32], Simple Sequence Repeat (SSR) [23, 29, 33, 34], SSR and Amplified Fragment Length Polymorphism (AFLP) [35, 36].

In a recent study presented in the World Coffee Research annual report a genetic diversity assessment of 800 Arabica's accessions from the collection at CATIE, Costa Rica, shows the least genetic diversity of *C. arabica* compared to other major crops [37]. This study also found that coffee cultivars contain almost 45% of the genetic diversity found in the 800 above-mentioned accessions indicating the limitation of variability for breeding programs [3]. Therefore, it is crucial to assess the population structure and its genetic diversity in *Coffea* genus.

Of course, all *C. arabica* germplasm available in ex situ collections may represent only a fraction of the total genetic diversity of the remaining wild and semi-wild forest coffees in S.W. Ethiopia [38]. However, Arabica's breeders do have already an idea of the potential and limits of ET germplasm, in particular in regard to host resistances to diseases and pests. For example, none of the modern Arabica cultivars with host resistances to CLR derive from these ET germplasm [39]. Also cultivars resistant to CBD outside Ethiopia do not have ET germplasm as progenitors [40], while nematode resistance found in ET accessions provide only limited protection to the severe nematode problems in Central America [41].

In contrast, ET germplasm may be a good source for sensory quality traits in cup. The cup quality profile of the new Arabica's F1-hybrids developed for Central America is said to derive largely from one of the two progenitors, being a selected ET accession of the FAO-1964 pool [42]. Silvarolla et al. found three coffee plants in offspring of ET germplasm, which were nearly caffeine-free [43]. Male sterility has been detected in a few ET accessions, a character useful for F1-hybrid seed production [44].

#### 4. Next generation sequencing techniques in *C. arabica*

NGS incorporate technologies which, at low cost and in short time, produce millions of short DNA sequence. The most commonly used platforms for high-throughput, useful genomic research, especially in non-model plant species include second generation sequencing techniques (SGseqTs): Illumina/Solexa, 454/Roche, ABI/SOLiD, and Helicos (read mostly in the range of 25 and 700 bp in length) [45]. Results obtained from such research point to the fact that NGS techniques (NGSTs) should not be restricted to the genomes of model organisms only as non-model plants have provided useful resources for genomic studies [45].

In contrast to classical molecular markers, SNPs are the most abundant markers, particularly in the non-coding regions of the genome [46]. NGS used jointly with different complexity reduction methods, Genotyping by sequencing (GBS) and DArTseq™ (Sequencing-based diversity array technology) methods, enable a large-scale discovery of SNPs in a wide variety of non-model organisms [47–49]. These techniques provide measures of genetic divergence and diversity within the major genetic clusters that comprise crop germplasm [50].

The genotyping profiles of SNPs can be compared across laboratories and sequencing platforms. These benefits have resulted in the increasing use of SNPs as high-quality markers for genotype identification in a wide range of crops [51], as recently demonstrated in cacao (*Theobroma cacao*; [52]), pummelo (*Citrus maxima*; [53]), tea (*Camellia sinensis*; [54]), longan (*Dimocarpus longan*; [55]), and litchi (*Litchi chinensis*; [56]).



Although significant, the number of reports concerning genomic resources in *Coffea*, even for a specie of commercial importance, such as *C. arabica*, is still low. Already, genotyping profiles of SNPs were identified and tested in *C. arabica* by Moncada et al. [57], Sousa et al. [29], Sant'Ana et al. [10] and Merot-L'anthoene et al. [4]. High-throughput genotyping assays are still needed in order to rapidly characterize the coffee genetic diversity and to evaluate the introgression of different cultivars in a cost-effective way. Measures must be taken to construct high-density genetic maps in *Coffea* [57, 58]. However, the use of SNP markers to generate denser maps is still low.

## 5. DArTseq™: an effective tool for genome diversity in *C. arabica*

The DArTseq™ technology, developed by DArT company (<https://www.diversityarrays.com>), is one of those methods that have received increasing interest worldwide since it can generate thousands of high-quality SNPs in a timely and cost-effective manner [59, 60]. The DArTseq™ method, a variation of GBS, implements complexity reduction methods that effectively targets low-copy sequences of the genome [61]. Besides, this process is optimized for each organism and type of study, by using combinations of restriction enzymes (REs) and selecting the most effective in reducing genome complexity [59].

The DArTseq™ technology has been utilized in diploid but more often in polyploid plant species, such as rice (*Oryza sativa*; [62]), barley (*Hordeum vulgare*; [63]) and maize (*Zea mays*; [64]), because SNP detection is facilitated by high fidelity REs, rather than relying on the annealing of primers to genomic targets in the presence of homologous annealing sequences [65].

In coffee, we have reported a genetic diversity study in 87 accessions of *Coffea* spp. These accessions were selected from the National Coffee Germplasm Bank located at 19°10' 27" N and 96° 57' 50" W and 1345 masl, in Huatusco, Veracruz, Mexico. Accessions were previously characterized by DArTseq™ method and SNP markers in Spinoso-Castillo et al. [66].

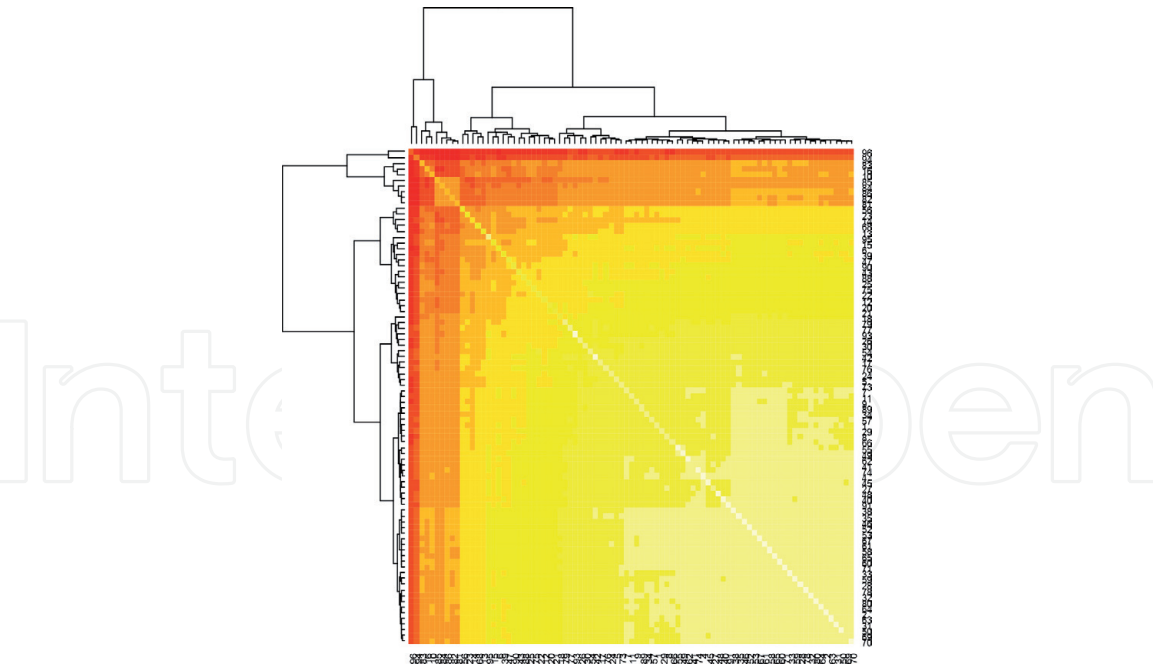
As a result, 16,995 SNP markers, derived from 34,000 unique sequences, were obtained by DArTseq™ from 87 accessions of different *Coffea* spp. After removing the markers with more than 10% of the missing data and MAF <5%, there were 1,739 polymorphic SNP markers for the analysis. After imputation and elimination of markers based on MAF, a heat map of the 87 accessions was obtained by using the genomic relations matrix *G* (Figure 1).

For the heat map, the genomic relations matrix *G* can be easily calculated using the following expression:

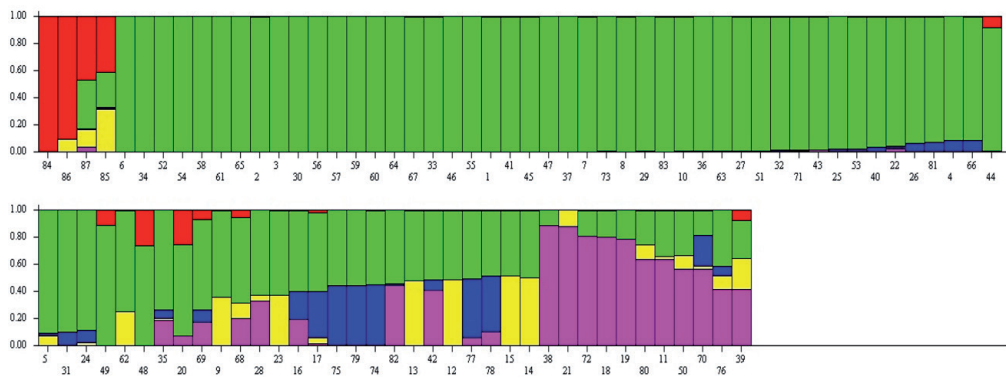
$$G = \frac{ZZ'}{p}, \quad (1)$$

where *Z* is the matrix of markers of dimension *n* = 87 rows (individuals) and *p* = 1,739 columns (markers), which is obtained by centering and standardizing the columns of the matrix of markers. The model-based Bayesian cluster analysis in STRUCTURE visualized the population structure under examination (Figure 2). Five distinct sub-populations were found across cultivars.

The sub-populations were denoted as Pop1, Pop2, Pop3, Pop4 and Pop5. The first group clustered *C. liberica* (84) and *C. canephora* (85, 86 and 87) species, the second group clustered mostly *C. arabica* accessions of the central collection, which evidenced the greater dissimilarity of these accessions with *C. liberica* and *C. canephora* species; the third group clustered CIRAD's F1-hybrids (74–79). Fourth and fifth clusters compiled different *C. arabica* accessions among them. Also, it was shown by



**Figure 1.**  
Heat map for the 87 accessions of *Coffea* spp. from the National Bank of Coffee Germplasm in Mexico using DArTseq Technology. Red small squares indicate an individual's genetic relatedness to itself, dark orange color represents high kinship relations while lighter colors (yellow) represent weaker relations.



**Figure 2.**  
Bar graphic of the STRUCTURE software used to study the diversity of the 87 coffee accessions using SNP marker data. The 87 genotypes are represented below the graphic, and were divided into five ( $K = 5$ ) groups.

Steiger et al. [67], using AFLP markers, that *C. canephora* and *C. arabica* were more genetically similar, revealing inter-species diversity even though *C. arabica* resulted from a recent hybridization between *C. canephora* and *C. eugenioides* [13].

The results obtained from this *Coffea* spp. central collection are similar to those reported in the study of Sant'Ana et al. [10] who found in the population structure analyses the presence of two to three groups ( $K = 2$  and  $K = 3$ ), corresponding to the east and west sides of the Great Rift Valley and an additional group formed by wild *C. arabica* accessions collected in the western forests. Sousa et al. [29] analyzed the population structure of coffee genotypes of interest for breeding studies, they used 11,187 SNP markers from which two groups ( $K = 2$ ) were obtained.

## 6. Advantages and disadvantages of NGS techniques in *C. arabica* genomics

High quality reference genome assemblies accelerate plant breeding by selecting desirable genes with improved agronomic traits, including high yield, tolerance to

various abiotic and biotic stresses, and resistance to pathogens [68]. However, draft genomes are suffering from unknown sequences and ambiguous assembly due to homologous sequences, while high-quality genomes are required for comparative genomics and functional annotation to crop improvement [68, 69].

These NGSTs are classified as second and third generation. The success of these NGSTs is mainly due to advancement in nanofluidics and automated single molecule imaging [69]. SGseqTs refer to those methods which require a PCR step for signal intensification prior to sequencing and third generation sequencing techniques (TGSeqTs) are those which can perform single molecule sequencing (SMS) [70].

As an advantage, in SGseqT the variation is different in their sequencing chemistry, cost, accuracy, speed and read length; SGseqTs produce thousands to billions of nucleotide long reads (25–800 nucleotides) as compared to first generation sequencing method [69, 70]. However, as a disadvantage, the accuracy of SGseqTs differs due to dependence on several multiplication steps during library preparation, each manipulation causes various artifacts in DNA measurements; additionally, the small reads produced by these procedures are not suitable for de novo genome assembly [69, 70].

Therefore, novel technologies are being designed in such a way that involve a minimum or no manipulation of the natural DNA molecule; TGSTs are able to analyze natural DNA/RNA molecules without any manipulation and without amplification [70] TGSTs have average read length longer to 10 kb, the availability of long reads constitutes a great advantage.

The first SMS technology, was developed by Quake and commercialized in 2009 by Helicos BioSciences; it worked similar to Illumina sequencers, but without any bridge amplification [70, 71]. However, it was slow, expensive and produced relatively short reads, around 35 bp long; therefore, two single-molecule approaches were technologically advanced to overcome these disadvantages [72].

The first approach, Single Molecule Real-Time (SMRT) sequencing was developed by Craighead, Korlach, Turner and Webb and was further refined and commercialized by Pacific Biosciences (PacBio) since 2011 [73]. The second approach, Nanopore sequencing, was first hypothesized in the 1990s and further developed and commercialized by Oxford Nanopore Technologies (ONT) since 2005; the advantages of SMRT sequencing over NGS have come at the price of higher per base sequencing costs [70].

Finally, DArTseq™ technique is based on genomic complexity reduction. This technique benefitted from the development in NGSTs and now DArTseq™ markers are replaced by NGS-DArT markers. Sansaloni et al. [60] found that the combined use of DArTseq™ with NGS make available more quantity of markers than conventional DArT method. DArTseq™ markers in combination with other molecular techniques have been used to create deeper genetic maps in *C. arabica* to perform association studies [4, 74, 75].

## 7. A future in genomic resources of *C. arabica*

Arabica's cultivars and landraces are generally propagated by seed. The mating system is primarily based on self-fertilization. Thereby, autogamy leads to high levels of inbreeding. Besides, an effective clonal propagation system is being adopted but limited for F1 Arabica hybrids. It is evident that molecular analyses of genetic diversity are needed to support this scenario [74, 75].

The development of a new coffee variety takes about 25 years. An efficient selection can be addressed when sequencing approaches are adopted in the variety

development process [66, 76]. In the 1990s, Marker-Assisted Selection (MAS) was proposed, which enabled selecting individuals with specific alleles. However, MAS has shown to be inefficient in polygenic and/or low heritability traits [77]. Due to its potential and importance, genome-wide selection (GS) was developed by Meuwissen et al. [78].

With the development of NGSTs, GS has become a reality for several economically important species. However, the procedure requires precaution for polyploid species, which have subgenomes with duplicate regions or with high similarity, such as *C. arabica* [77]. Despite the economic importance of *C. arabica*, GS works in Arabica coffee are scarce. Coffee trees have been selected based on biometric analyses using phenotypic data of yield and resistance to biotic and abiotic stresses. However, due to the complexity and number of genes that control most of the agronomic traits of this *Coffea* spp., GS studies are promising for they allow estimating the effects of all loci that explain the genetic variation and the genomic estimated breeding value (GEBV) [74, 75, 77].

Genome sequencing initiatives of Arabica accessions have been launched by several research groups (<https://coffeegenome.ucdavis.edu/>, among others) but an open-access genome assembly, with a reliable sorting of homologous sequences, is not yet available [77, 79]. Decoding the allotetraploid genome of *C. arabica* is therefore required to have accurate GS studies in this species.

## 8. Conclusions

DARtseq™ technology identifies thousands of high quality SNP polymorphic markers in a timely and cost-effective manner. Our study confirmed that the genotyping method by DARtseq™ can be successfully used in studies of genetic diversity specially in coffee. In addition, trait-associated-SNPs identified by GWAS may be helpful to develop strategies aiming to improve the biochemical quality of coffee or another important trait. These SNPs markers may be useful for marker-assisted selection (MAS) in Arabica coffee breeding programs and genomic selection.

## Acknowledgements

This work was supported by the FONDO SECTORIAL SAGARPA-CONACYT [2016-2101-277838].

## Conflict of interest

The authors have no conflicting interests, and all authors have approved the manuscript and agree with its submission to IntechOpen.



IntechOpen

## Author details

José Luis Spinoso-Castillo<sup>1</sup>, Tarsicio Corona-Torres<sup>1\*</sup>, Esteban Escamilla-Prado<sup>3</sup>, Victorino Morales-Ramos<sup>2</sup>, Víctor Heber Aguilar-Rincón<sup>1</sup> and Gabino García-de los Santos<sup>1</sup>

<sup>1</sup> Postgraduate College, Campus Montecillo, Texcoco, State of Mexico, Mexico

<sup>2</sup> Postgraduate College, Campus Córdoba, Amatlán de los Reyes, Veracruz, Mexico

<sup>3</sup> Autonomous University of Chapingo, Eastern Regional University Center, Veracruz, Mexico

\*Address all correspondence to: tcoronat@gmail.com

## IntechOpen

© 2021 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

## References

- [1] ICO. International coffee organization. 2020. Available from: <https://www.ico.org/prices/po-production.pdf>. [Accessed: 2020-11-05]
- [2] Oxfam. The Coffee Market: A Background Study; Oxfam: London, UK; 2001
- [3] Tran HT, Lee LS, Furtado A, Smyth H, Henry RJ. Advances in genomics for the improvement of quality in coffee. *J Sci Food Agric*. 2016;96:3300-3312
- [4] Merot-L'anthoene V, Tournebize R, Darracq O, Rattina V, Lepelley M, Bellanger L, et al. Development and evaluation of a genome-wide Coffee 8.5 K SNP array and its application for high density genetic mapping and for investigating the origin of *Coffea arabica* L. *Plant Biotechnol J*. 2019;1-13
- [5] Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet*. 2011;12:499-510
- [6] Gupta PK, Rustgi S, Mir RR. Array-based high-throughput DNA markers for crop improvement. *Heredity*. 2008;101:5-18
- [7] You Q, Yang X, Peng Z, Xu L, Wang J. Development and applications of a high throughput genotyping tool for polyploidy crops: single nucleotide polymorphism (SNP) array. *Front. Plant Sci*. 2018;9:104
- [8] Yu H, Xie W, Li J, Zhou F, Zhang, Q. A whole-genome SNP array (RICE6K) for genomic breeding in rice. *Plant Biotechnol. J*. 2014;12:28-37
- [9] Denoeud F, Carretero-Paulet L, Dereeper A, Droc G, Guyot R, Pietrella M, Zheng C, et al. The coffee genome provides insight into the convergent evolution of Caffeine biosynthesis. *Science*. 2014;345:1181-1184
- [10] Sant'Ana GC, Pereira LFP, Pot D, Ivamoto ST, Domingues DS, Ferreira RV, Pagiatto NF, et al. Genome-wide association study reveals candidate genes influencing lipids and diterpenes contents in *Coffea arabica* L. *Sci. Rep*. 2018;8:465.
- [11] Tran HTM, Ramaraj T, Furtado A, Lee LS, Henry RJ. Use of a draft genome of coffee (*Coffea arabica*) to identify SNPs associated with caffeine content. *Plant Biotechnol. J*. 2018;16:1756-1766
- [12] Davis AP, Tosh J, Ruch N, Fay M. 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*. 2011;167:357-377
- [13] Lashermes P, Combes MC, Robert J, et al. Molecular characterization and origin of the *Coffea arabica* L. genome. *Mol Gene Genet*. 1999;261:259-266
- [14] Pruvot-Woehl S, Krishnan S, Solano W, Schilling T, Toniutti L, Bertrand B, Montagnon C. Authentication of *Coffea arabica* varieties through DNA fingerprinting and its significance for the coffee sector. *Journal of AOAC International*. 2020;103(2):325-334
- [15] Baruah A, Naik V, Hendre PS, Rajkumar P, Aggarwal RK. Isolation and characterization of nine microsatellite markers from *Coffea arabica* L. showing wide cross-species amplifications. *Mol Ecol Notes*. 2003;3:647-650
- [16] Prakash NS, Combes MC, Somanna N, Lashermes P. AFLP analysis

of introgression in coffee cultivars (*Coffea arabica* L.) derived from a natural interspecific hybrid. *Euphytica*. 2002;124:265-271

[17] Benti T, Gebre E, Tesfaye K, Berecha G, Lashermes P, Kyallo M, Yao NK. Genetic diversity among commercial arabica coffee (*Coffea arabica* L.) varieties in Ethiopia using simple sequence repeat markers. *Journal of Crop Improvement*. 2020

[18] Bertrand B, Etienne H, Cilas C, Charrier A, Baradat P. *Coffea arabica* hybrid performance for yield, fertility and bean weight. *Euphytica*. 2005;141(3):255-262

[19] Pot D, Scholz MBS, Lannes SD, Del Grossi L, Pereira LFP, Vieira LG, Sera T. Phenotypic analysis of *Coffea arabica* accessions from Ethiopia: contribution to the understanding of *Coffea arabica* diversity. In: 22nd International Conference on Coffee Science, Campinas. Anais. Campinas; 2008. p 165

[20] Silvestrini M, et al. Genetic diversity and structure of Ethiopian, Yemen and Brazilian *Coffea arabica* L. accessions using microsatellites markers. *Genet. Resour. Crop Ev*. 2007;54:1367-1379

[21] López-Gartner G, Cortina H, Couch MC, Susan R, Moncada MDP. Analysis of genetic structure in a sample of coffee (*Coffea arabica* L.) using fluorescent SSR markers. *Tree Genet Genomes*. 2009;5:435-446

[22] Teresa A, Crouzillat D, Petiard V, Brouhan P. Genetic diversity of Arabica coffee (*Coffea arabica* L.) collections. *EJAST*. 2010;1:63-79

[23] Aerts R, Berecha G, Gijbels P, Hundera K, Glabeke S, Vandepitte K, Muys B, Roldán RI, Honnay O. Genetic Variation and Risks of Introgression in the Wild *Coffea Arabica* Gene Pool in South-Western Ethiopian Montane

Rainforests. *Evolutionary Applications*. 2013;6(2):243-252

[24] Scholz MBS, Kitzberger CSG, Pagiatto NF, Pereira LFP, Davrieux F, Pot D, et al. Chemical composition in wild Ethiopian Arabica coffee accessions. *Euphytica*. 2016;209(2):429-438

[25] da Silva, BSR, Sant'Ana GC, Chaves CL, et al. Population structure and genetic relationships between Ethiopian and Brazilian *Coffea arabica* genotypes revealed by SSR markers. *Genetica*. 2019;147:205-216

[26] Jingade P, Huded AK, Kosaraju B, Mishra MK. Diversity Genotyping of Indian Coffee (*Coffea Arabica* L.) Germplasm Accessions by Using SRAP Markers. *Journal of Crop Improvement*. 2019;33:327-345

[27] Kesawat MS, Das Kumar B. Molecular Markers: It's Application in Crop Improvement. *Journal of Crop Science and Biotechnology*. 2009;12:169181

[28] Garrido-Cardenas JA, Mesa-Valle C, Manzano-Agugliaro F. Trends in Plant Research Using Molecular Markers. *Planta*. 2018;247:543

[29] Sousa TV, Caixeta ET, Alkimim ER, de Oliveira ACB, Pereira AA, Zambolim L, Sakiyama NS. Molecular Markers Useful to Discriminate *Coffea Arabica* Cultivars with High Genetic Similarity. *Euphytica*. 2017;213

[30] Lashermes P, Trouslot P, Anthony F, Combes MC, Charrie A. Genetic Diversity for RAPD Markers between Cultivated and Wild Accessions of *Coffea Arabica*. *Euphytica*. 1996;87:59-64

[31] Anthony F, Bertrand B, Quiros O, Lashermes P, Berthaud J, Charrier A. Genetic Diversity of Wild Coffee (*Coffea arabica* L.) Using Molecular Markers. *Euphytica*. 2001;118:53-65

- [32] Aga E, Bekele E, Bryngelsson T. Inter-Simple Sequences Repeat Variation in Forest Coffee Trees (*Coffea arabica* L.) Populations from Ethiopia. *Genetica*. 2005;124:213-221
- [33] Tornincasa P, Dreos R, Nardi D, Barbara E, Asquini J, Devasia M, Mishra K, et al. Genetic Diversity of Commercial Coffee (*C. Arabica* L.) From America, India and Africa Assessed by Simple Sequence Repeats (SSR). *Proceedings of 21st International Association for Coffee Science (ASIC)*, Montpellier, France. 2007; pp. 778-785
- [34] Geleta M, Herrera I, Monzón A, Bryngelsson T. Genetic Diversity of Arabica Coffee (*Coffea arabica* L.) In Nicaragua as Estimated by Simple Sequence Repeat Markers. *The Scientific World Journal*. 2012;1-11
- [35] Dessalegn Y, Herselman L, Labuschagne M. Comparison of SSR and AFLP Analysis for Genetic Diversity Assessment of Ethiopian Arabica Coffee Genotypes. *South African Journal of Plant and Soil*. 2009;26(2):119-125
- [36] Maluf MP, Silvestrini M, Ruggiero LM, de Guerreiro-Filho CO, Colombo CA. Genetic Diversity of Cultivated *Coffea Arabica* Inbred Lines Assessed by RAPD, AFLP and SSR Marker Systems. *Scientia Agricola*. 2005;62:366-373
- [37] World Coffee Research (WCR), Assessment of genetic diversity in *Coffea arabica*. *World Coffee Research 2014 Annual Report*, Texas, USA.;2014
- [38] Labouisse JP, Bellachew B, Kotecha S, Bertrand B. Current status of coffee (*Coffea arabica*) genetic resources in Ethiopia: implications for conservation. *Genet Res Crop Evol*. 2008;55:1079-1093
- [39] Silva MC, Várzea V, Guerra-Guimaraes L, Gil Azinheira H, Fernandez D, Petitot AS, Bertrand B, Lashermes Ph, Nicole M. Coffee resistance to the main diseases: leaf rust and coffee berry disease. *Braz J Plant Physiol*. 2006;18:119-147
- [40] Van der Vossen HAM. The cup quality of disease-resistant cultivars of arabica coffee (*Coffea arabica*). *Expl Agric*. 2009;45:323-332
- [41] Anzueto F, Bertrand B, Sarah JL, Eskes AB, Decazy B. Resistance to *Meloidogyne incognita* in Ethiopian *Coffea arabica* origins: detection and study of resistance transmission. *Euphytica*. 2001;118:1-8
- [42] Bertrand B, Vaast Ph, Alpizar E, Etienne H, Davrieux F, Charmetant P. Comparison of bean biochemical composition and beverage quality of arabica hybrids involving Sudanese-Ethiopian origins with traditional varieties at various elevations in Central America. *Tree Physiol*. 2006;26:1239-1248
- [43] Silvarolla MB, Mazzafera P, Fazuoli LC. A naturally decaffeinated arabica coffee. *Nature*. 2004;429:826
- [44] Georget F, Alpizar E, Courtel P, Hidalgo JM, Dechamp E, Poncon C, Etienne H, Bertrand B. *Arabica* F1 hybrid seed production based on genetic male sterility. *Proceedings of the 25th ASIC international conference on coffee science*, Armenia, Colombia. 2015
- [45] Unamba CIN, Nag A, Sharma RK. Next Generation Sequencing Technologies: The Doorway to the Unexplored Genomics of Non-Model Plants. *Front. Plant Sci*. 2015;6:1074
- [46] Mondini L, Noorani A, Pagnotta MA. Assessing Plant Genetic Diversity by Molecular Tools. *Diversity*. 2009;1(1):19-35.
- [47] Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA.



Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLoS ONE. 2008;3(10):e3376

[48] Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS ONE. 2011;6(5):e19379

[49] Cruz VMV, Kilian A, Dierig DA. 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. 2013;8(5):e64062

[50] Kumar S, Banks TW, Cloutier S. SNP Discovery through Next-Generation Sequencing and Its Applications. Int J Plant Genomics. 2012;2012:15

[51] Zhou L, Vega FE, Tan H, Ramírez Lluch AE, Meinhardt LW, Fang W, Mischke S, Irish B, Zhang D. Developing Single Nucleotide Polymorphism (SNP) Markers for the Identification of Coffee Germplasm. Tropical Plant Biol. 2016;9:82-95

[52] Ji K, Zhang D, Motilal L, et al. Genetic diversity and parentage in farmer varieties of cacao (*Theobroma cacao* L.) from Honduras and Nicaragua as revealed by single nucleotide polymorphism (SNP) markers. Genet Resour Crop Evol. 2013;60:441-453

[53] Wu GA, Prochnik S, Jenkins J, et al. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. Nat Biotechnol. 2014;32:656-662

[54] Fang W, Meinhardt LW, Tan H, et al. Varietal identification of tea (*Camellia sinensis*) using nanofluidic array of single nucleotide polymorphism (SNP) markers. Hortic Res. 2014;1:14035

[55] Wang B, Tan H, Fang W, et al. Developing single nucleotide polymorphism (SNP) markers from transcriptome sequences for identification of longan (*Dimocarpus longan*) germplasm. Hortic Res. 2015;2:14065

[56] Liu W, Xiao ZD, Bao XL, et al. Identifying litchi (*Litchi chinensis* Sonn.) cultivars and their genetic relationships using single nucleotide polymorphism (SNP) markers. PLoS ONE. 2015;10(8):e0135390

[57] Moncada MD, Tovar E, Montoya JC, Gonzalez A, Spindel J, McCouch S. A genetic linkage map of coffee (*Coffea arabica* L.) and QTL for yield, plant height, and bean size. Tree Genet. Genomes. 2016;12:5

[58] Pearl HM, Nagai C, Moore PH, Steiger DL, Osgood RV, Ming R. Construction of a genetic map for Arabica coffee. Theor. Appl. Genet. 2004;108: 829-835

[59] Nguyen NH, Premachandra HKA, Kilian A, Knibb W. Genomic prediction using DArT-Seq technology for yellowtail kingfish *Seriola lalandi*. BMC Genomics. 2018;19(107):1-9

[60] Sansaloni C, Petroli C, Jaccoud D, Carling J, Detering F, Grattapaglia D, Kilian A. Diversity arrays technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. BMC Proc. 2011;5(S7):54

[61] Melville J, Haines ML, Boysen K, Hodkinson L, Kilian A, Smith Date KL, Potvin DA, Parris KM. Identifying hybridization and admixture using SNPs: application of the DArTseq platform in phylogeographic research on vertebrates. R. Soc. Open Sci. 2017;4(7):161061

- [62] Courtois B, Audebert A, Dardou A, Roques S, Ghneim-Herrera T, Droc G, Frouin J, Rouan L, Gozé E, Kilian A, Ahmadi N, Dingkuhn M. Genome-wide association mapping of root traits in a japonica rice panel. *PLoS One*. 2013;8(11):e78037
- [63] Dracatos PM, Khatkar MS, Singh D, Park RF. Genetic mapping of a new race specific resistance allele effective to *Puccinia hordei* at the Rph9/Rph12 locus on chromosome 5HL in barley. *BMC Plant Biol*. 2014;14(1598):1-12
- [64] Dos Santos JPR, Pires LPM, De Castro Vasconcellos RC, Pereira GS, Von Pinho RG, Balestre M. Genomic selection to resistance to *Stenocarpella maydis* in maize lines using DArTseq markers. *BMC Genet*. 2016;17(86):1-10
- [65] Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang S, Uszynski G, Mohler V, Lehmensiek A, Kuchel H, Hayden MJ, Howes N, Sharp P, Vaughan P, Rathmell B, Huttner E, Kilian A. Diversity Arrays Technology (DArT) for highthroughput profiling of the hexaploid wheat genome. *Theor. Appl. Genet*. 2006;113(8):1409-1420
- [66] Spinoso-Castillo JL, Escamilla-Prado E, Aguilar-Rincón VH, Morales Ramos V, García de los Santos G, Pérez-Rodríguez P, Corona-Torres T. Genetic diversity of coffee (*Coffea* spp.) in Mexico evaluated by using DArTseq and SNP markers. *Genet Resour Crop Evol*. 2020. DOI: [doi.org/10.1007/s10722-020-00940-5](https://doi.org/10.1007/s10722-020-00940-5)
- [67] Steiger DL, Nagai C, Moore PH, Morden CW, Osgood RV, Ming R. AFLP analysis of genetic diversity within and among *Coffea arabica* cultivars. *Theor Appl Genet*. 2002;105(2-3):209-215
- [68] Li C, Lin F, An D, Wang W, Huang R. Genome Sequencing and Assembly by Long Reads in Plants. *Genes*. 2018;9(1):6
- [69] Yadav P, Vaidya E, Rani R, et al. Recent Perspective of Next Generation Sequencing: Applications in Molecular Plant Biology and Crop Improvement. *Proc. Natl. Acad. Sci., India, Sect. B Biol. Sci*. 2018;88:435-449
- [70] Giani AM, Gallo GR, Gianfranceschi L, Formenti G. Long walk to genomics: History and current approaches to genome sequencing and assembly. *Computational and Structural Biotechnology Journal*. 2020;18:9-19
- [71] Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, et al. Single-molecule DNA sequencing of a viral genome. *Science*. 2008;320:106-109
- [72] Salazar AN, Nobrega FL, Anyansi C, Aparicio-Maldonado C, Costa AR, et al. An educational guide for nanopore sequencing in the classroom. *PLOS Computational Biology*. 2020;16(1):e1007314
- [73] Deamer D, Akeson M, Branton D. Three decades of nanopore sequencing. *Nat Biotechnol*. 2016;34:518-524
- [74] Scalabrin S, Toniutti L, Di Gaspero G, et al. 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*. 2020;10:4642
- [75] Alkimim ER, Caixeta ET, Sousa TV, et al. Selective efficiency of genome-wide selection in *Coffea canephora* breeding. *Tree Genetics & Genomes*. 2020;16:41
- [76] Gimase JM, Thagana WM, Omondi CO, et al. Genome-Wide Association Study identify the genetic loci conferring resistance to Coffee Berry Disease (*Colletotrichum kahawae*) in *Coffea arabica* var. Rume Sudan. *Euphytica*. 2020;216:86

[77] Sousa TV, Caixeta ET, Alkimim ER, Oliveira ACB, Pereira AA, Sakiyama NS, Zambolim L, Resende MDV. Early Selection Enabled by the Implementation of Genomic Selection in *Coffea arabica* Breeding. *Front. Plant Sci.* 2019;9:1934

[78] Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 2001;157:1819-1829

[79] Huang L, Wang X, Dong Y, et al. Resequencing 93 accessions of coffee unveils independent and parallel selection during *Coffea* species divergence. *Plant Mol Biol.* 2020; 103:51-61 (2020)