

UC Davis

UC Davis Previously Published Works

Title

Genetic diversity and re-classification of coffee (*Coffea canephora* Pierre ex A. Froehner) from South Western Nigeria through genotyping-by-sequencing-single nucleotide polymorphism analysis

Permalink

<https://escholarship.org/uc/item/30v0h34b>

Journal

Genetic resources and crop evolution, 66(3)

ISSN

0925-9864

Authors

Anagbogu, Chinyere F
Bhattacharjee, Ranjana
Ilori, Christopher
[et al.](#)

Publication Date

2019-03-01

DOI

10.1007/s10722-019-00744-2

Peer reviewed



Genetic diversity and re-classification of coffee (*Coffea canephora* Pierre ex A. Froehner) from South Western Nigeria through genotyping-by-sequencing-single nucleotide polymorphism analysis

Chinyere F. Anagbogu · Ranjana Bhattacharjee · Christopher Ilori · Pumipat Tongyoo · Keji E. Dada · Anna A. Muiyiwa · Paul Gepts · Diane M. Beckles

Received: 9 January 2018 / Accepted: 28 January 2019
© Springer Nature B.V. 2019

Abstract *Coffea canephora* is an important economic crop in Nigeria, however, little is known about the diversity inherent within, and the genetic relationship among coffee grown and conserved in the country. We examined the genetic diversity and relatedness among 48 *Coffea* genotypes which included: (a) *C. arabica*, *C. abeokutae*, *C. liberica*, and *C. stenophylla*, (b) 14 *C. canephora* accessions conserved in the germplasm of Cocoa Research Institute of Nigeria (CRIN), and (c) 30 farmer-cultivated genotypes collected from South-Western Nigeria. By analyzing 433048 single nucleotide polymorphisms (SNPs) identified through genotyping-by-sequencing we discovered that previous characterizations of *C. canephora* based on morphological

data were inconclusive. Here, we established the correct number of *C. canephora* varieties present in the CRIN genebank which was four and not six as previously described based on morphological characters. We found three distinct diversity structures within the *C. canephora* gene pool that were dominated by a single genetic group determined from passport descriptors to most likely be of Congolese (Democratic Republic of Congo) origin. High uniformity was also found among the farmer-cultivated accessions with 99% of them representing *C. canephora* var. Niaouli as their ancestral background. The analysis showed that the genetic base of coffee germplasm in Nigeria is narrow compared to the large genetic diversity of *C. canephora*. Therefore, broadening this genetic base through future acquisition and hybridization is imperative. However, the relatively high genetic differentiation (F_{ST} estimate = 0.3037)

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10722-019-00744-2>) contains supplementary material, which is available to authorized users.

C. F. Anagbogu (✉) · K. E. Dada · A. A. Muiyiwa
Crop Improvement Division, Cocoa Research Institute of Nigeria, Ibadan, Nigeria
e-mail: flora2na@yahoo.com

C. F. Anagbogu · C. Ilori
Crop Protection and Environmental Biology Department,
University of Ibadan, Ibadan, Nigeria

C. F. Anagbogu · P. Gepts · D. M. Beckles
Department of Plant Sciences, University of California,
Davis, CA, USA

R. Bhattacharjee
Bioscience Laboratory, International Institute of Tropical Agriculture, Ibadan, Nigeria

P. Tongyoo
Center of Excellence on Agricultural Biotechnology:
(AG-BIO/PERDO-CHE), Bangkok 10900, Thailand

identified between Java Robusta and Niaouli will be used as a starting point for our breeding program.

Keywords *Coffea canephora* · Genetic diversity · Genotyping-by-sequencing · Single nucleotide polymorphism

Introduction

Coffee (*Coffea* spp. Linnaeus) is a high-value crop globally. Over 2.25 billion cups are consumed daily (Dicum and Luttinger 2006) by more than a third of the world's population (Bolvenkel et al. 1993). It is also a source of income for millions of smallholder farmers who collectively produce over 70% of the world's coffee (Oxfam 2001).

The genus *Coffea*, has over 124 species. The two most cultivated are *C. arabica* Linnaeus and *C. canephora* Pierre ex A. Froehner (Zamir 2014), which make up 65% and 35% of the international coffee trade, respectively (Davis et al. 2006). *C. arabica* is prized for its cup quality attributes, while *C. canephora* is valued for its higher disease resistance and yield (Bertrand et al. 2003). *C. arabica* is an allotetraploid ($2n = 4x = 44$), while *C. canephora* is a self-incompatible diploid ($2n = 2x = 22$) and one of the progenitors of *C. arabica* (Pearl et al. 2004).

In Nigeria, *C. canephora* constitutes 95% of total coffee production while *C. arabica* contributes only 5%. Previous research activity has been limited to germplasm introductions, and agro-morphological characterizations of six *Coffea* spp. including *C. arabica*, *C. canephora*, *C. liberica* Bull. ex Hiern, *C. abeakutae* Cramer, *C. excelsa* Aug. Chevalier, and *C. stenophylla* G. Don (Omolaja et al. 1997). *C. canephora* has received the most attention, specifically, the six varieties: Gold Coast, Java Robusta, Uganda, Kouilou/Quillou, Java Robusta ex. Gamba and Niaouli. The true-to-type identity of genotypes/varieties cultivated by farmers in Nigeria is not known. The introduction of these genotypes was made in 1966 from Ghana, Indonesia, Zaire (Democratic Republic of Congo), Uganda and the Benin Republic (Williams 1989).

Information on the molecular characterization of *Coffea* in Nigeria is scarce. *Coffea* genotypes in this country have been traditionally distinguished using

morphological characteristics (Omolaja et al. 2000), which is insufficient because of environmental influences on phenotype (Souza et al. 2013). Molecular markers can partition environmental from genetic influences on phenotype, thus providing a higher level of accuracy on the genetic relatedness of different genotypes (Mishra et al. 2011). First- and second-generation molecular markers have already been adopted for coffee genotyping (Achar et al. 2015; Hendre and Aggarwal 2007; Hendre et al. 2008; Lashermes et al. 1999; Silvestrini et al. 2008; Garavito et al. 2016). “Next-Generation Sequencing” technologies such as genotyping-by-sequencing (GBS) however, are preferred for genome-wide diversity studies because of their high efficiency compared with other single nucleotide polymorphism (SNP) discovery techniques (He et al. 2014; Kwok 2001; Poland et al. 2012). Also in coffee, GBS and diversity array technology sequencing (DArTseq) were used to discover large number of SNPs which could be useful in subsequent coffee breeding programs and in understanding genetic background of varieties of coffee produced (Garavito et al. 2016; Hamon et al. 2017). The availability of the draft genome of *C. canephora* (Denoëud et al. 2014) has facilitated the utilization of GBS for such diversity studies.

In this study we used GBS-SNPs to analyze a total of 48 coffee genotypes of importance to south-western Nigeria, the primary coffee-production area in the country. These comprised 30 accessions selected directly from farmers' fields and 18 accessions acquired from the Cocoa Research Institute of Nigeria (CRIN) germplasm repository. Our aim was to determine (i) the extent to which there was genetic uniformity among the farmer-cultivated coffee cultivars, (ii) the genetic diversity among the conserved germplasm, and (iii) the genetic background of the farmer-cultivated genotypes.

Materials and methods

Plant material

Forty eight coffee samples comprising of 18 accessions from the Cocoa Research Institute of Nigeria (CRIN) germplasm repository (Table 1) and 30 accessions from six farmers' plots with differing locations were used in this study (Table S1). The

CRIN accessions were obtained from five *Coffea* species: *C. arabica*, *C. abeokutae*, *C. liberica*, *C. stenophylla* and *C. canephora*. Each of the species was represented by one genotype except *C. canephora* species which was represented by six different varieties: Kouillou, Gold Coast, Java Robusta, Niaouli, Uganda and Java Robusta Ex Gamba (Table 1), where two to three genotypes were selected to represent each variety (Table 1). The farmers' fields were located in south-western Nigeria (Figure S1A), and the altitude and latitude of the locations were recorded using ArcGIS software (Redlands, California). This included three farms each in both Kogi (7°79'N 5°80'E) and Ekiti states (7°77'N 5°77'E), regions where coffee production is dominant within Nigeria (Figure S1B).

Genotyping-by-sequencing analysis of coffee genotypes

DNA extraction

Young leaves were harvested from all 48 genotypes and placed into zip-lock bags filled with silica gel (Rabbi et al. 2015). The material was lyophilized at the International Institute of Tropical Agriculture in Nigeria and DNA extraction performed using a cetyl trimethylammonium bromide (CTAB) method optimized for coffee (Santa Ram and Sreenath 2000) at the University of California Davis, USA.

DNA library preparation and sequencing

Genomic DNA was sent to the Cornell University Biotechnology Resource Center for GBS sequencing and analysis: <http://www.biotech.cornell.edu/brc/brc/services/terms-and-policies>. A GBS 96-plex protocol commonly used by the maize research community was applied in this study (Elshire et al. 2011). The restriction endonuclease, *ApeKI* (New England Biolabs, Ipswich, MA) that recognizes a degenerate 5 bp sequence GCWGC (where W is A or T) and leaves 2 to 3 bp (CWG) overhangs was chosen. Oligonucleotides (Table S2) comprising the top and bottom strands of each barcode adapter and a common adapter, were diluted and annealed in a thermocycler according to Elshire et al. (2011). Adapters quantification and dilution, DNA and adapter plating, and DNA digest were all performed following the protocol developed for maize GBS (Elshire et al. 2011).

The digested DNA samples, each with a different barcode adapter, were combined (5 µL each) and purified using a commercial kit (QIAquick PCR Purification Kit; Qiagen, Valencia, CA) according to the manufacturer's instructions. DNA samples were eluted into a final volume of 50 µL. Restriction fragments from each sample were pooled and amplified by PCR in 50 µL volumes containing 2 µL pooled DNA fragments, 16 µL *Taq* Master Mix (New England Biolabs), and 25 pmol each of the primers (Table S3). These PCR primers were complementary to the ligated adapters, allowing the amplified product to bind the oligonucleotides that coat the Illumina

Table 1 The known background i.e. genome size, variety, and origin of the 18 coffee genotypes acquired from the CRIN Germplasm repository prior to GBS analysis

Species	Genome size	Variety	Origin	Sample name
<i>C. arabica</i>	1174	–	Kenya (Williams 1989)	Ara_18
<i>C. abeokutae</i>	587	–	Nigeria (Omolaja et al. 2000)	Abe_02
<i>C. liberica</i>	636	–	Nigeria (Omolaja et al. 2000)	Lib_02
<i>C. stenophylla</i>	587	–	Ivory Coast (Razafinarivo et al. 2013)	Ste_02
<i>C. canephora</i>	807	Kouillou	Zaire, DRC (Montagnon et al. 1998)	C90, C111, C36
		Gold Coast	Ghana (Williams 1989)	A111, A81
		Java Robusta	Indonesia (Williams 1989)	E106, E77
		Niaouli	Republic of Benin (Montagnon et al. 1998; Williams 1989)	M10, M36
		Uganda	Uganda (Montagnon et al. 1998, Williams 1989)	G129, G37
		Java Robusta ex Gamba	Zaire, DRC (Williams 1989)	T1049, T921, T797

sequencing flow cell, and to prime subsequent DNA sequencing reactions (Bentley et al. 2008). PCR cycling parameters were: 72 °C for 5 min, and 98 °C for 30 s, followed by 18 cycles of 98 °C for 30 s, 65 °C for 30 s, and 72 °C for 30 s, with a final step at 72 °C for 5 min (Elshire et al. 2011).

The amplified library was purified using a commercial kit (QIAquick PCR Purification Kit; Qiagen, Valencia, CA) and the DNA was loaded onto a capillary sizing system to evaluate the fragment sizes contained within the library (Experion[®] automated electrophoresis station; BioRad Experion[®] Hercules, CA). DNA fragment sizes between 170–350 bp were used for, single-end sequencing on a flow cell channel of HiSeq 2500 Illumina (Illumina, Inc., San Diego, CA) (Bradbury et al. 2007; Glaubitz et al. 2014). The pre-processing of the sequenced read data was performed on a TASSEL (Trait Analysis by Association, Evolution and Linkage)-GBS Pipeline (Bradbury et al. 2007; Glaubitz et al. 2014). Of the 267 million reads generated, 9.2% of the reads were acceptable and retained for SNP calling.

SNP Calling

The read or raw data (C81ECANXX_8_fastq.gz) was aligned to the reference genome, i.e. the *C. canephora*, pseudomolecules.fa.gz (<http://coffee-genome.org>), using the Burrows-Wheeler Alignment (BWA) (Li and Durbin 2009) within the TASSEL commands (net-maizegenetics.pipeline.TasselPipeline—Tassel Version: 3.0.173). After alignment, SNP calling was initiated with Sequence Alignment Map (SAM) (Li and Durbin 2009) on aligned tags, using a default settings to generate a Hapmap genotype (Danecek et al. 2011; Etter et al. 2011). Genotype quality (GQ) score was calculated to the GATK version (<http://gatkforums.broadinstitute.org/discussion/1268/how-should-i-interpret-vcf-files-produced-by-the-gatk>).

SNPs were filtered with VCFtools. Those with a minor allele frequency > 1%, and, with less than 10% missing data per site across taxa, were retained. During this process *Can_47* and *C. abeokutae* genotypes were removed because of their low SNP coverage (as a result of high missing data at > 10% missing sites) and the occurrence of SNPs at a low allele frequency of < 0.01. This reduced the number of genotypes from 48 to 46. All raw GBS sequencing data was submitted to the National Center for Biotechnology Information

(NCBI), Sequence Read Archive (study accession number SRP096172: <https://www.ncbi.nlm.nih.gov/sra/?term=SRP096172>). The filtered SNPs were used for downstream statistical analyses.

Multivariate analyses

The genetic diversity and relatedness of 46 coffee genotypes, both conserved and cultivated, were assessed using TASSEL software version 5.0 (Bradbury et al. 2007) to generate a tree based on the Neighbour-Joining dendrogram approach; PCA and hierarchical clustering on 43,3048 SNPs were carried out using SNPRelate software (<http://www.Rproject.org>) (Zheng et al. 2012) to estimate the genetic relatedness among 46 genotypes. Pairwise analysis of Identity by State (IBS) distance matrices was used to relate genetic distance to genetic diversity among all genotypes (Purcell et al. 2007). F-statistic (F_{ST}) was calculated using VCFtools (Danecek et al. 2011) based on Weir and Cockerham estimation, to identify the genetic differentiation between and among populations (Balloux and Lugon-Moulin 2002). Admixture analysis was carried out using STRUCTURE (structureHarvester.py v0.6.94) (Dent and Bridgett 2012).

Results and discussion

To our knowledge, this is the first GBS analysis of *C. canephora* genotypes in Nigeria to assess the extent of genetic diversity and genetic uniformity among conserved genotypes and farmer-cultivated coffee cultivars, and to determine the genetic background of the farmer-cultivated genotypes.

SNP distribution/characterization and heterozygosity of alleles

The number of raw SNPs detected was 440,481 while 433,048 SNPs were obtained after filtering for low (< 10%) minor allelic frequency (MAF; 0.01), across 46 genotypes. Of these SNPs, 329,577 were distributed across the 11 coffee chromosomes. The remaining 103,471 SNPs were anchored on “chromosome 0”, because they could not be assigned to any of the known 11 chromosomes. Most SNPs were recorded on chromosome 2, which had three times

more SNPs compared to Chromosome 9, which had the least number (Table S4).

The highest number of heterozygous alleles (106,365) was observed in *C. arabica* (Ara_18). The genotype classified as *C. liberica* (Lib_02), the farmer-cultivated accession Can_19, and the *C. canephora* var Kouillou accession C90, also had high number of heterozygous alleles (more than 60,000), but the *C. canephora* genotypes var. Java Robusta, (E106) had the highest number of heterozygous alleles (between 60,000 and 80,000). In contrast, the farmers' accessions Can_20, Can_38, and Can_44 had the fewest alleles i.e. 6697, 4462 and 4721 respectively (Figure S3), almost 10-fold fewer compared with the genotypes with the highest number of heterozygous alleles.

Twice as many heterozygous alleles were found in *C. arabica* compared to *C. canephora*. This is likely due to the polyploid nature of *C. arabica*, and the presence of polymorphisms existing between the loci of the two homeologous genomes of the tetraploid *C. arabica*. Lashermes et al. (1999) also found a high level of fixed heterozygosity in the *C. arabica* genome and regarded the level of its internal genetic variability to be twice that present within its diploid relatives.

Relatedness of the 46 genotypes (conserved and cultivated)

There was a high degree of similarity in SNP polymorphisms between M10 and each of the farmer-cultivated accessions, irrespective of where they were grown in Nigeria (Fig. 2 and Figure S4). The genetic uniformity observed among the cultivated accessions is an indication that they may have been propagated vegetatively. This suggests a production system characterized by a high level of clonal multiplication and cooperation among Nigeria coffee farmers. The lack of genetic diversity in these accessions contradicts previous report that there were as many as 26 varieties of *C. canephora* in the distribution zone that encompasses Nigeria (Montagnon et al. 1998; Gomez et al. 2009).

Of the 30 farmer-cultivated accessions analysed, only one (Can_19), was genetically distinct. This exceptional genotype (Can_19) resembled Lib_02 and C90, both of which were conserved in the CRIN coffee germplasm. The Lib_02 genotype labelled as *C. liberica*, was introduced from farmers' field to the CRIN coffee germplasm (Omolaja et al. 1997). This

genotype seemed to be divergent from the other farmers' accessions (Figure S4).

Previous morphological characterizations of the CRIN coffee germplasm classified C90 and Lib_02 as *C. canephora* and *C. liberica* respectively (Omolaja et al. 1997). The floral morphology of C90 and Lib_02 appeared to be different from *C. canephora* and even *C. liberica* (Fig. 2b). However, the GBS-SNP analysis in this study showed a high level of divergence in the SNPs detected among C90/Lib_02/Can_19 (Figures S4 and 3A) and *C. canephora*. Our analysis found C90 and Lib_02 (G1; Tables 2, 3) to be closer to *C. canephora* than to *C. arabica*. First, the average genetic distance was 0.2014 with *C. canephora*, compared with the higher value of 0.3346 found with *C. arabica* (Table S4). Second, the mean F_{ST} estimate of 0.50006 with *C. arabica* indicates higher genetic differentiation, while the lower values (0.1321–0.2501) found with *C. canephora* indicate greater similarity (Tables 2, 3). Finally, genetic structure analysis (Fig. 3) confirmed C90 and Lib_02 to be *C. canephora*.

Assessment of genetic diversity and reclassification of coffee germplasm

The multidimensional scaling (MDS) using SNP-related analysis is generally used in assessing diversity and relatedness of genotypes. This analysis grouped the 46 genotypes into four diverse clusters on a principal component plot, showing the first (PC1) and the second principal components (PC2) which explained 36.2% of the variation (Fig. 1a, b). The genotypes belonging to a cluster are more genetically similar compared to those in other clusters. This clustering differentiated *C. canephora* into 3 sub-groups (II, III and IV). Also, an admixture analysis discovered three populations among the *C. canephora* genotypes, (Q1, Q2 and Q3), of which two (Q2 and Q3) were regarded as being sub-populations by STRUCTURE (Fig. 3 and Table S7).

Identity By State (IBS) distance matrix and principal component analysis (PCA) clustering were used to identify both between and within-species diversity. The analysis revealed that the average genetic distances of *C. canephora* from (i) *C. arabica*, (ii) the *C. canephora* variety once labeled as *C. liberica*, and (iii) the farmers' accessions were 0.3346, 0.2014, and 0.1867, respectively (Table S5). These data may be

Table 2 The grouping of genotypes based on hierarchical clustering

G1	Lib_02	Can_19	C90					
G2	Ara_18							
G3	T979	T921	T1049	E106				
G4	C36	C111	G37	G129	E77	A81	A111	Ste_02
G5	M36	M10	Can_46	Can_45	Can_44	Can_43	Can_42	Can_41
	Can_40	Can_39	Can_38	Can_37	Can_36	Can_35	Can_34	Can_33
	Can_32	Can_31	Can_30	Can_29	Can_28	Can_27	Can_26	Can_25
	Can_24	Can_23	Can_22	Can_21	Can_20	Can_18		

Table 3 Pairwise F_{ST} estimation of the five groups identified from HC (Hierarchical Clustering) analysis

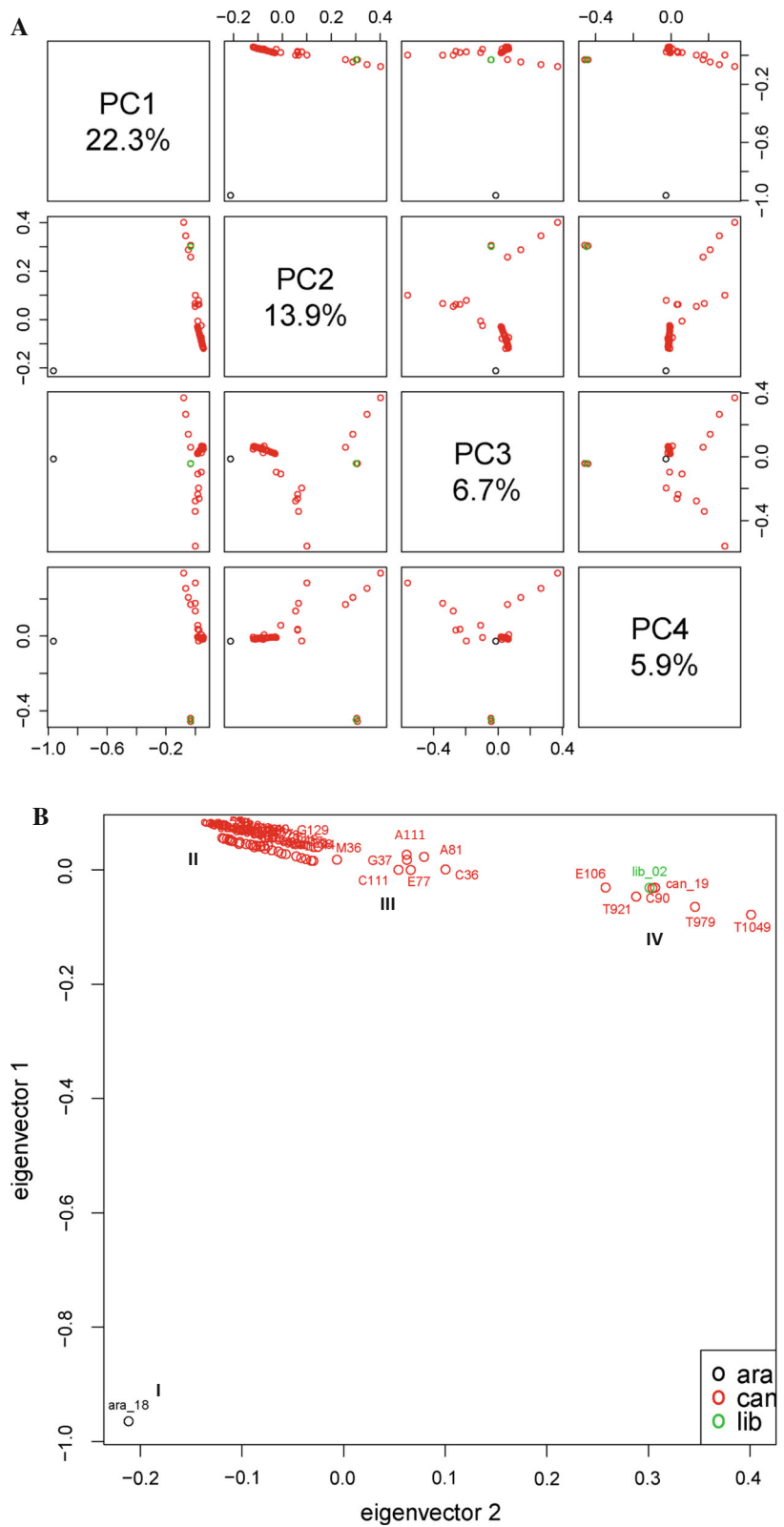
Comparing group	Total sample	Mean F_{ST} estimate	Weighted F_{ST} estimate
All groups	46	0.3648	0.35337
G1_vs_G2	4	0.50006	0.59913
G1_vs_G3	7	0.1321	0.22404
G1_vs_G4	11	0.17983	0.3079
G1_vs_G5	33	0.25018	0.38704
G2_vs_G3	5	0.25351	0.43656
G2_vs_G4	9	0.29954	0.51683
G2_vs_G5	31	0.3037	0.6111
G3_vs_G4	12	0.14999	0.27821
G3_vs_G5	34	0.24287	0.39527
G4_vs_G5	38	0.13977	0.26971

Key is as follows: 0–0.05 indicates little genetic differentiation; 0.05 and 0.15, moderate differentiation; 0.15 and 0.25, great differentiation and above 0.25, very great genetic differentiation. For the interpretation of F_{ST} estimate see Balloux and Lugon-Moulin (2002)

interpreted as follows: firstly, there is high genetic distance between *C. canephora* and *C. arabica*, revealing inter-species diversity even though *C. arabica* resulted from a recent hybridization between *C. canephora* and *C. eugenioides* (Lashermes et al. 1999). Secondly, there is low genomic variation between *C. canephora* and the genotype formerly classified as *C. liberica*, revealing intraspecific diversity. This is evidence that the formally classified *C. liberica* is actually a *C. canephora* genotype because it shares more common alleles with *C. canephora* (Fig. 1b) and belong to the same sub-population with Java Robusta (*C. canephora* variety). If Lib_02 was a *C. liberica* genotype it would have a higher genetic distance compared with *C. canephora*. This was shown by Steiger et al. (2002) using AFLP markers. They reported that *C. canephora* and *C. arabica* were more genetically similar, while *C. canephora* and *C. liberica* were more genetically distinct. The low value

(0.17983) detected with F_{ST} estimation between Lib_02 (G1) and *C. canephora* (G3 and G4) confirmed Lib_02 to be a *C. canephora* genotype. Thirdly, there is low genetic distance between *C. canephora* and the farmers' accessions (0.1867). From both hierarchical analysis and IBS genetic distance values, it is possible to assume that coffee farmers in south-western Nigeria are cultivating one variety of *C. canephora*, specifically the Niaouli variety. This was illustrated by the low genetic distance (IBS value) of 0.1194 between *C. canephora* var. Niaouli (M10) and the farmers' accessions (Table S5), and the formation of a cluster (II) between the farmers' accessions and M10 (Fig. 1b). Also the ancestral inference detected with Structure analysis grouped farmers accessions into the same population structure as M10 (Fig. 3). The reason for this widespread adoption of a single genotype among farmers is not known. It is possible that there was an exchange of coffee seedlings from the

Fig. 1 a Principal component analysis of 46 coffee genotypes categorized by SNPRelate software, **b** The multi dimensional scale (MDS) plot of PC1 (22.3%) and PC2 (13.9%) of 46 genotypes (cluster I: *C. arabica*, II-IV: *C. canephora*)



neighboring country of Benin (Gomez et al. 2009) the center of origin of *C. canephora* var. Niaouli (Montagnon et al. 1998), leading to its introduction into Nigeria, and its subsequent cultivation was due to farmer preferences based on its unique characters or due to its availability.

While comparing *C. arabica* with the four *C. canephora* varieties (Table 2, 3), it was observed that var. Java Robusta grouped closer to *C. arabica* (G2) having a mean F_{ST} estimate of 0.25351 (Tables 2, 3), while *C. canephora* var. Niaouli was more distant with mean F_{ST} estimate of 0.3037 (Table 3 and S5). *C. canephora* var. Java Robusta is known to have good cup quality and leaf rust resistance, and has the shortest genetic distance from *C. arabica* (high quality coffee) and should be utilized for for intraspecific hybridization, and interspecific hybridization after confirming their reproductive compatibility.

The intra-species diversity among different genotypes of the six *C. canephora* varieties used in the study was compared in a pairwise manner, and the average genetic distance was 0.1867 (Table S6). The highest IBS value of 0.2552 (Table S6), pairwise F_{ST} estimate of 0.24287 (Table 3) and unrelated population structures of Q2 and Q1 (Fig. 3) were found between var. Java Robusta ex Gamba (T1049) and var. Niaouli (M10), suggesting that these two varieties are the most diverse. The heterosis within these *C. canephora* varieties may therefore be utilized in intra-specific crossing, to develop improved varieties of coffee. In addition, hierarchical clustering (Fig. 2) disclosed four inter-genetic diversity among *C. canephora* constituting four varietal levels: (1) Java Robusta and Java Robusta ex Gamba (Group III), (2) Kouillou/Quillou and Java Robusta (Group IV), (3) Uganda and Gold Coast (Group V) and (4) Niaouli (Group VI). Interestingly, the results from the population structure (Fig. 3) and PCA analysis (Fig. 1b) are in conformity, since both revealed three genetic units in the *C. canephora* germplasm from the CRIN. The varieties of *C. canephora* identified here are of Congolese' origin (Montagnon et al. 1998; Dussert et al. 1999; Gomez et al. 2009; Musoli et al. 2009; Razafinarivo et al. 2013; Leroy et al. 2014), and thus represent a very narrow genetic pool. There has been little acquisition of new coffee genetic resources at the CRIN. The need to broaden this genetic base for continued coffee improvement in Nigeria becomes imperative. Low genetic diversity has also been

observed by Omolaja and Fawole (2004) using morphological characteristics.

The low number of heterozygous alleles found in different *C. canephora* genotypes was surprising, since it is a self-incompatible species. A study by Souza et al. (2013) on *C. canephora* cultivated in Brazil detected high genetic diversity. Low genetic diversity in self-incompatible *C. canephora* genotypes maintained and cultivated in Nigeria indicated little or no major breeding efforts targeting these genotypes, which may be a reason for their current low yield and quality. To ensure proper utilization of genetic resources, a full knowledge of the inherent genetic diversity and relationship within the genepool of interest is necessary (Li and Durbin 2009). All analytical approaches used in the current study revealed the existence of misclassified genotypes conserved in the CRIN germplasm repository. It is recommended that the two varieties currently described as 'Uganda' and 'Gold Coast' be merged into a single variety (Uganda/Gold Coast) since no significant genetic difference was detected between them (Figs. 1b, 2a, b). Although the selected 14 *C. canephora* conserved genotypes from the CRIN germplasm repository were previously categorized into six varietal groups based on morphological characteristics (Table 4), our genomic analyses using next-generation sequencing method (GBS) confirmed that these actually belonged to four varietal groups (Table 4). The comprehensive nature of the SNPs-GBS genomic analysis instilled a high degree of confidence with respect to the classification of these varieties. A similar result was achieved with EST-SSR markers for proper grouping of different populations and varietal groups in coffee from Brazil (Souza et al. 2013).

The two admixture genotypes, M36 and G129 detected with STRUCTURE (Fig. 3) contain almost equal proportion of Niaouli and Kouillou. These genotypes will be incorporated into var. Java Robusta genome through conventional hybridization thereby generating F_1 with the combined genomic composition of all the three genetic structures (Niaouli, Kouillou and Java Robusta). From this study it can be deduced that the analysed genotypes of germplasm representative of the CRIN collection which was started in 1966, comprised of three genetic structures. Some of the nomenclature of the accessions were assigned based to the donor countries' name, and

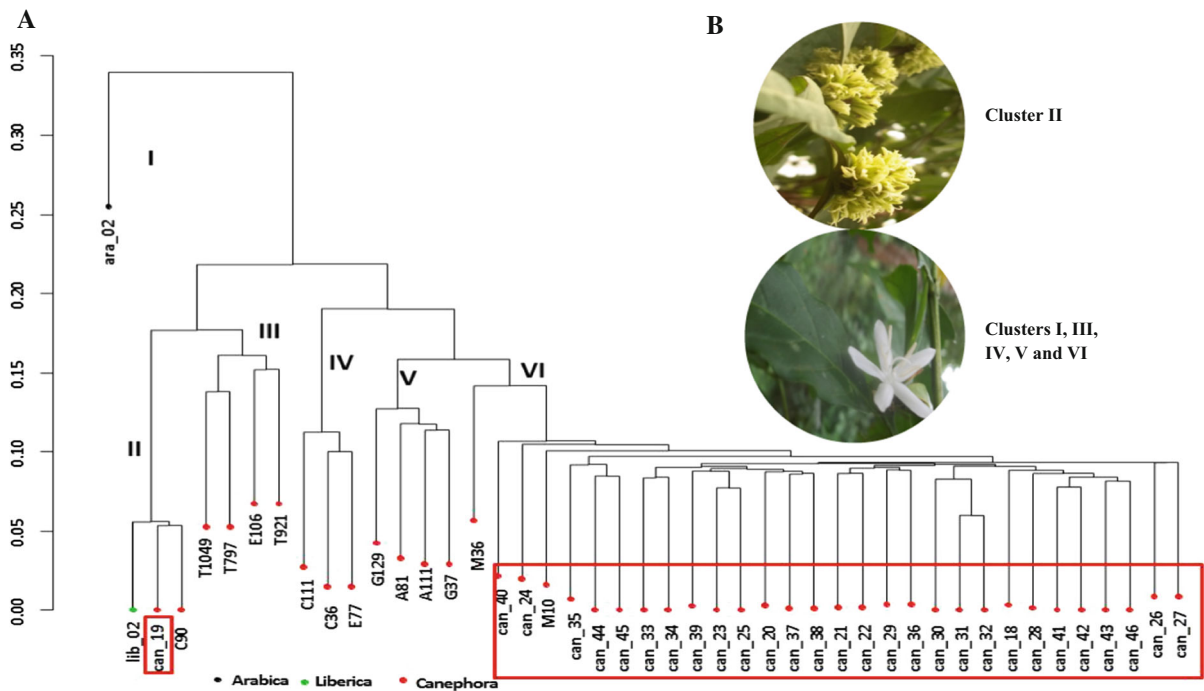


Fig. 2 a Hierarchical clustering of the genotypes. Cluster I: A *C. arabica* genotype; II: Unknown *C. canephora* genotypes, III: Java Robusta ex Gamba and Java Robusta (E106); IV: Kouillou and Java Robusta (E77); V: Uganda and Gold Coast; VI: Niaouli

and farmers' accessions. The samples in red rectangle are the accessions from farmers' field except M10. **b** Morphological differences in the flowers produced by genotypes in Cluster II and I, III, IV, V and VI. (Color figure online)

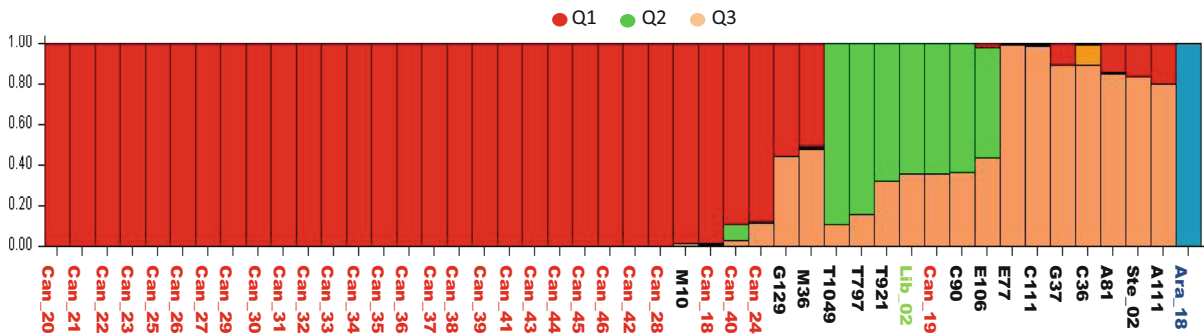


Fig. 3 The Genetic Structure analysis of the 46 genotypes between K = 1 and K = 10 with 3 populations Q1, Q2 and Q3. Q2 and Q3 are regarded as sub-populations. The Q1 population represents the farmers' accessions, while Q2 and Q3 are

from the CRIN conserved genotypes (Black are conserved genotypes of *C. canephora*, green is formerly *C. liberica*, blue is a genotype of *C. arabica*, while red are cultivated accessions). (Color figure online)

collection was by seeds. These three genetic structures were grouped based on the clustering of the genotypes into Niaouli, Gold Coast/Kouillou and Java Robusta.

Improvement in the CRIN germplasm collection should be channelled towards acquiring more genetic material for *C. canephora* genotypes from Ivory Coast and Uganda as they have high yield components, quality traits and tolerance to biotic and abiotic

stresses. In order to improve the organoleptic quality, the introduction of var. '126' should be prioritized, as it is the best accession for quality and yield traits (Leroy et al. 2014). This accession has been widely distributed in Togo, Guinea and Cameroon. Also, accessions '410', 'A03 and '466' all have specific genotypic values between Guinean and Congolese groups (Leroy et al. 2014) and should be targeted for

Table 4 Grouping of the *C. canephora* genotypes based on SNP-GBS genotyping. The number decreased from six varieties to four. (a) Hierarchical clustering formed the following varietal groups: Java Robusta ex Gamba and Java Robusta

(E106), Kioullou and Java Robusta (E77), Uganda and Gold Coast, and Niaouli alone (Fig. 2). (b) Similar classification was confirmed using PCA and IBS (Fig. 1 and Table S5)

Species	Genotypes	Varietal Grouping based on Morphological Markers	Varietal Grouping based on SNP-GBS
<i>C. canephora</i>	T1049	Java Robusta Ex Gamba	Java Robusta Ex Gamba
<i>C. canephora</i>	T797	Java Robusta Ex Gamba	Java Robusta Ex Gamba
<i>C. canephora</i>	E106	Java Robusta	Java Robusta Ex Gamba
<i>C. canephora</i>	T921	Java Robusta Ex Gamba	Java Robusta Ex Gamba
<i>C. canephora</i>	C111	Kouillou	Kouillou
<i>C. canephora</i>	C36	Kouillou	Kouillou
<i>C. canephora</i>	E77	Java Robusta	Kouillou
<i>C. canephora</i>	G129	Uganda	Uganda/Gold Coast
<i>C. canephora</i>	G37	Uganda	Uganda/Gold Coast
<i>C. canephora</i>	A81	Gold Coast	Uganda/Gold Coast
<i>C. canephora</i>	A111	Gold Coast	Uganda/Gold Coast
<i>C. canephora</i>	M10	Niaouli	Niaouli
<i>C. canephora</i>	M36	Niaouli	Niaouli

possible incorporation into the farmer's germplasm. The high variability found within the Java Robusta and Niaouli should also be utilized in crosses to obtain hybrids.

Conclusion

The ability to capture and efficiently use abundant genetic resources are considered essential for sustainable coffee production in Nigeria. The utilization of information on the diversity, relatedness, and consistency of the coffee genetic resources found in this study will help in planning a worthwhile coffee improvement program in Nigeria. Despite the limited number of genotypes used, we were able to determine that the previous characterization performed only with morphological characters was inconclusive and that the efficient utilization of genetic resources has been lacking. There is a need to broaden the genetic base of *C. canephora* and, generally, the *Coffea* species in Nigeria. Collaboration with other scientists from the Ivory Coast and Uganda, countries that harbor large collections of *C. canephora* genotypes, becomes imperative, in order to acquire new genetic material, and to ascertain their true genetic identity.

Acknowledgements Support for this research was provided in part by the Norman E. Borlaug Leadership Enhancement in Agriculture Program (Borlaug LEAP) through a grant to the University of California-Davis by the United States Agency for International Development. The opinions expressed herein are those of the authors and do not necessarily reflect the views of USAID. We are grateful to the International Institute of Tropical Agriculture and the University of California Davis, USA for providing free access to their facilities. Funding was also provided by the Federal Ministry of Agriculture in Nigeria through a grant to the Cocoa Research Institute of Nigeria. We are also grateful to Dr. May Thitisaksakul, Ms. Karin Alborno, Dr. Shaoyun Dong, Ms. Tamara Miller, Dr. Sarah Dohle, and Dr. Jorge Dubcovsky at UC Davis, USA for their advice and technical support.

Funding United States Agency for International Development (USAID) and Federal Ministry of Agriculture, Nigeria.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Achar D, Awati MG, Udayakumar M, Prasad TG (2015) Identification of putative molecular markers associated with

- root traits in *Coffea canephora* Pierre ex Froehner. *Mol Biol Int* 2015;11. <https://doi.org/10.1155/2015/532386>
- Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. *Mol Ecol* 11:155–165. <https://doi.org/10.1046/j.0962-1083.2001.01436.x>
- Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Cheetham RK, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IMJ, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DMD, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Catenazzi MCE, Chang S, Cooley RN, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang G-D, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ling Ng B, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Pike AC, Chris Pinkard D, Pliskin DP, Podhasky J, Quijano VJ, Raczky C, Rae VH, Rawlings SR, Rodriguez AC, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Ernest Sohna Sohna J, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, Vandevondele S, Verhovskiy Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Rogers J, Mullikin JC, Hurler ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R, Smith AJ (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456(7218):53–59
- Bertrand B, Guyot B, Anthony F, Lashermes P (2003) Impact of the *Coffea canephora* gene introgression on beverage quality of *C. arabica*. *Theor Appl Genet* 107(3):387–394. <https://doi.org/10.1007/s00122-003-1203-6>
- Bolvenkel E, Buckley T, Eijgendaal C (1993) Report on a mission for the International Trade Centre. Dissemination activities for coffee—an exporters guide
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23(19):2633–2635. <https://doi.org/10.1093/bioinformatics/btm308>
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G et al (2011) The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Davis AP, Govaerts R, Bridson DM, Stoffelen P (2006) An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). *Bot J Linn Soc* 152(4):465–512. <https://doi.org/10.1111/j.1095-8339.2006.00584.x>
- Denoeud F, Carretero-Paulet L, Dereeper A, Droc G, Guyot R, Pietrella M, Zheng CF, Alberti A et al (2014) The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* 345(6201):1181–1184. <https://doi.org/10.1126/science.1255274>
- Dent EA, Bridgett MH (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4(2):359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- Dicum G, Luttinger N (2006) The coffee book: anatomy of an industry from the crop to the last drop. The New Press, New York, NY
- Dussert S, Lashermes P, Anthony F, Montagnon C, Trouslot P, Combes M-C, Berthaud JMN, Hamon S (1999) Le caféier, *Coffea canephora*. Diversité génétique des plantes tropicales cultivées. CIRAD, Montpellier
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6(5):e19379. <https://doi.org/10.1371/journal.pone.0019379>
- Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA (2011) SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *Mol Methods Evolut Genet* 772:157–178. https://doi.org/10.1007/978-1-61779-228-1_9
- Garavito A, Montagnon C, Guyot R, Bertrand B (2016) Identification by the DArTseq method of the genetic origin of the *Coffea canephora* cultivated in Vietnam and Mexico. *BMC Plant Biol* 16:242. <https://doi.org/10.1186/s12870-016-0933-y>
- Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES (2014) TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS ONE* 9(2):e90346. <https://doi.org/10.1371/journal.pone.0090346>
- Gomez C, Dussert S, Hamon P, Hamon S, de Kochko A, Poncet V (2009) Current genetic differentiation of *Coffea canephora* Pierre ex A. Froehn in the Guineo-Congolian African zone: cumulative impact of ancient climatic changes and recent human activities. *BMC Evol Biol* 9:167. <https://doi.org/10.1186/1471-2148-9-167>
- Hamon P, Grover CE, Davis AP, Rakotomalala JJ, Raharimalala NE, Albert VA, Sreenath HL, Stoffelen P et al (2017) Genotyping-by-sequencing provides the first well-resolved phylogeny for coffee (*Coffea*) and insights into the evolution of caffeine content in its species GBS coffee phylogeny and the evolution of caffeine content. *Mol Phylogenet Evol* 109:351–361. <https://doi.org/10.1016/j.ympev.2017.02.009>

- He JF, Zhao XQ, Laroche A, Lu ZX, Liu HK, Li ZQ (2014) Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Front Plant Sci* 5:484. <https://doi.org/10.3389/fpls.2014.00484>
- Hendre PS, Aggarwal RK (2007) DNA markers: development and application for genetic improvement of coffee. In: Varshney RK, Tuberosa R (eds) *Genomic Assisted Crop Improvement: Genomics Applications in Crops*. Springer, Berlin, pp 399–434
- Hendre PS, Phanindranath R, Annapurna V, Lalremruata A, Aggarwal RK (2008) Development of new genomic microsatellite markers from robusta coffee (*Coffea canephora* Pierre ex A. Froehner) showing broad cross-species transferability and utility in genetic studies. *BMC Plant Biol* 8:51. <https://doi.org/10.1186/1471-2229-8-51>
- Kwok PY (2001) Methods for genotyping single nucleotide polymorphisms. *Annu Rev Genom Hum Genet* 2:235–258
- Lashermes P, Combes MC, Robert J, Trouslot P, D'Hont A, Anthony F, Charrier A (1999) Molecular characterisation and origin of the *Coffea arabica* L. genome. *Mol Gen Genet* 261(2):259–266
- Leroy T, De Bellis F, Legnate H, Musoli P, Kalonji A, Solorzano RGL, Cubry P (2014) Developing core collections to optimize the management and the exploitation of diversity of the coffee *canephora*. *Genetica* 142(3):185–199. <https://doi.org/10.1007/s10709-014-9766-5>
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Mishra MK, Nishani S, Jayarama (2011) Genetic relationship among indigenous coffee species from India using RAPD, ISSR and SRAP markers. *Biharean Biol* 5:17–24
- Montagnon C, Leroy T, Eskes AB (1998) *Coffea canephora* L. Amélioration variétale de Critères et méthodes de sélection. *Plant Rech Dév* 5:18–33
- Musoli P, Cubry P, Aluka P, Billot C, Dufour M, De Bellis F, Pot D, Bieysse D, Charrier A, Leroy T (2009) Genetic differentiation of wild and cultivated populations: diversity of *Coffea canephora* Pierre in Uganda. *Genome* 52(7):634–646. <https://doi.org/10.1139/G09-037>
- Omolaja SS, Fawole I (2004) Characterization of Nigerian robusta coffee (*Coffea Canephora* Pierre ex. Froehner) germplasm. In: *Proceedings of 20th International Conference on coffee science*, Bangalore, India. 11–15 October, 2004.
- Omolaja SS, Williams JA, Obatolu CR (1997) Germplasm collection of *Coffea abekutae* and *C. liberica*. CRIN Annual Report
- Omolaja SS, Obatolu CR, Williams JA (2000) Collection of *Coffea abekutae* Cramer and *Coffea liberica*. *Bull in South Western Nigeria. Plant Genet Resour* 122:26–31
- Oxfam (2001) *The coffee market: a background study*. Oxfam, London
- Pearl HM, Nagai C, Moore PH, Steiger DL, Osgood RV, Ming R (2004) Construction of a genetic map for arabica coffee. *Theor Appl Genet* 108(5):829–835. <https://doi.org/10.1007/s00122-003-1498-3>
- 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):e32253. <https://doi.org/10.1371/journal.pone.0032253>
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, Bakker PIW, Daly MJ, Sham PCS (2007) PLINK: a tool set for whole-genome association and population-based linkage analysis. *Am J Hum Genet* 81:559–575
- Rabbi IY, Kulakow PA, Manu-Aduening JA, Dankyi AA, Asibuo JY, Parkes EY, Abdoulaye T, Girma G et al (2015) Tracking crop varieties using genotyping-by-sequencing markers: a case study using cassava (*Manihot esculenta* Crantz). *BMC Genet* 16:115. <https://doi.org/10.1186/s12863-015-0273-1>
- Razafinarivo NJ, Guyot R, Davis AP, Couturon E, Hamon S, Cruzillat D, Rigoreau M, Dubreuil-Tranchant C et al (2013) Genetic structure and diversity of coffee (*Coffea*) across Africa and the Indian Ocean islands revealed using microsatellites. *Ann Bot London* 111(2):229–248. <https://doi.org/10.1093/aob/mcs283>
- Santa Ram A, Sreenath HL (2000) Genetic fingerprinting of coffee leaf rust differentials with RAPD markers. In: Sera T, Soccol CR, Pandey A, Roussos S (eds) *Proceedings of the 3rd international seminar on biotechnology in the coffee agro-industry*. Springer Science, Dordrecht. <https://doi.org/10.1007/978-94-017-1068-8>
- Silvestrini M, Maluf MP, Silvarolla MB, Guerreiro O, Medina HP, Vanini MMT, Oliveira AS, de Gaspari-Pezzopane C, Fazuoli LC (2008) Genetic diversity of a *Coffea* Germplasm Collection assessed by RAPD markers. *Genet Resour Crop Evol* 55(6):901–910. <https://doi.org/10.1007/s10722-007-9295-5>
- Souza FD, Caixeta ET, Ferrao LFV, Pena GF, Sakiyama NS, Zambolim EM, Zambolim L, Cruz CD (2013) Molecular diversity in *Coffea canephora* germplasm conserved and cultivated in Brazil. *Crop Breed Appl Biotechnol* 13(4):221–227
- Steiger DL, Nagai C, Moore PH, Morden CW, Osgood RV, Ming R (2002) AFLP analysis of genetic diversity within and among *Coffea arabica* cultivars. *Theor Appl Genet* 105(2–3):209–215. <https://doi.org/10.1007/s00122-002-0939-8>
- Williams JA (1989) Coffee breeding in Nigeria. In: *Progress in tree crop research*, 2nd edn. Cocoa Research Institute of Nigeria (CRIN) Ibadan, pp 127–140
- Zamir D (2014) A wake-up call with coffee. *Science* 345(6201):1124–1124. <https://doi.org/10.1126/science.1258941>
- Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS (2012) A high-performance computing toolset for relatedness and principal component analysis on SNP data. *Bioinformatics* 28:3326–3328

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.