OXFORD

Journal of AOAC INTERNATIONAL, 103(2), 2020, 315–324

doi: 10.1093/jaocint/qsz002
Article

SPECIAL GUEST EDITOR SECTION

Accurate Differentiation of Green Beans of Arabica and Robusta Coffee Using Nanofluidic Array of Single Nucleotide Polymorphism (SNP) Markers

Dapeng Zhang,[1] Fernando E. Vega,[1] Francisco Infante,[2] William Solano,[3] Elizabeth S. Johnson,[4] and Lyndel W. Meinhardt[5]

[1]Sustainable Perennial Crops Laboratory, U.S. Department of Agriculture (USDA), Agricultural Research Service (ARS), Beltsville, MD 20705, USA, [2]El Colegio de la Frontera Sur (ECOSUR), Carretera Antiguo Aeropuerto Km. 2.5, Tapachula, 30700 Chiapas, Mexico, [3]Centro Agronómico Tropical de Investigación y Enseñanza, Turrialba, Costa Rica, [4]The Inter-American Institute for Cooperation on Agriculture (IICA), Hope Gardens, Kingston 6, Jamaica, [5]Sustainable Perennial Crops Laboratory, USDA, Agricultural Research Service, Beltsville, MD 20705, USA

Corresponding author's e-mail: dapeng.zhang@ars.usda

Abstract

Green (unroasted) coffee is one of the most traded agricultural commodities in the world. The Arabica (*Coffea arabica* L.) and Robusta (*Coffea canephora* Pierre ex A. Froehner) species are the two main types of coffees for commercial production. In general, Arabica coffee is known to have better quality in terms of sensory characteristics; thus, it has a higher market value than Robusta coffee. Accurate differentiation of green beans of the two species is, therefore, of commercial interest in the coffee industry. Using the newly developed single nucleotide polymorphism (SNP) markers, we analyzed a total of 80 single green bean samples, representing 20 Arabica cultivars and four Robusta accessions. Reliable SNP fingerprints were generated for all tested samples. Unambiguous differentiation between Robusta and Arabica coffees was achieved using multivariate analysis and assignment test. The SNP marker panel and the genotyping protocol are sufficiently robust to detect admixture of green coffee in a high-throughput fashion. Moreover, the multilocus SNP approach can differentiate every single bean within Robusta and 55% of Arabica samples. This advantage, together with the single-bean sensitivity, suggests a significant potential for practical application of this technology in the coffee industry.

Coffee is one of the most popular global beverages and green (unroasted) coffee is one of the most traded agricultural commodities in the world. The yearly value of the global coffee industry has been estimated at US$173 billion (1). The genus *Coffea* (Rubiaceae) comprises 124 species (2, 3) while commercially-traded coffee comprises two species: Arabica (*Coffea arabica* L.) and Robusta (*C. canephora* Pierre ex. A. Froehner). These two main types of coffee account for 99% of the commercial market (4). In general, Arabica coffee has more favorable sensory characteristics and commands a higher commercial value. In 2016, Robusta green beans sold for ca 54% of the price of Arabica green beans in the commodities market (5). This price differential creates commercial interest within the coffee industry for accurate differentiation and identification of green coffee beans.

Green beans of the two species have different morphological characteristics that can be visually distinguished by experienced people. Nonetheless, a robust procedure is still required to reliably differentiate green beans of Arabica from samples admixed with Robusta. Various analytical methods have been

Received: 22 October 2019; **Accepted:** 22 October 2019

successfully developed and most of them are based on chemical or biochemical compositions of the green beans, such as chlorogenic and fatty acids, amino acid, caffeine, trigonelline, 3,4-dimethoxycinnamic acid, theobromine and theophylline, tocopherols, triglycerides, sterolic, and protein profiles (6–15). Detailed reviews of these approaches were summarized by Toci et al. (16) and Burns et al. (17).

Deoxyribonucleic acid (DNA)-based methods offer a complementary advantage over approaches based on chemical analysis because they directly identify the genotype thus bypassing the influence of environment and post-harvest handling of the green coffee beans. Real-Time PCR was suggested as a viable alternative to chemistry-based methods (17). Trantakis et al. (18) developed a PCR-based coffee authenticity assay using a diagnostic single nucleotide polymorphism (SNP) marker, whereby species identification is accomplished by visual detection of the strongly colored nano-particles of Robusta and Arabica coffees. In this study, as low as 5% of Robusta coffee was detected in the presence of Arabica coffee. Spaniolas et al. (19) developed SNP-based analytical assays to differentiate Arabica and Robusta varieties for the authentication of green coffee beans. Their results show that both single-base primer extension (SNaPShot®) and pyrosequencing could be used in the detection of powder mixtures of Arabica and Robusta green beans. They also showed that the PCR-RFLP assay could be used to authenticate the mixture, thus suggesting that this assay could be a highly useful method for a laboratory that lacks sequencing facilities. Combes et al. (20) developed a protocol for authenticating green and roasted coffee products using quantitative high resolution melting analysis of SNP markers. They showed that *Coffea* species of origin can be identified in green and roasted coffee, with Robusta adulteration as low as 1% in Arabica products.

In the present study, we propose an alternate approach using a multi-locus SNP genotyping approach based on a nanofluidic array. Compared with the previously reported procedure based on DNA markers, this method can be established based on a set of universally consistent SNP markers and the genotyping can be performed on single green beans. The method can be easily implemented using high-throughput SNP arrays which typically do not need gel separation, thus making it more suitable for large-scale industrial applications. Moreover, this procedure is based on multilocus SNP genotyping, which can potentially be applied not only for the detection of *Coffea* species, but to the differentiation of different cultivars or individual genotypes. For these reasons, we conducted the study with the aim of evaluating the efficacy of differentiating green bean samples of the two main coffee species using a nanofluidic array with 96 SNP markers. Considering commercial practices in the green coffee bean trade, the evaluations were conducted based on single coffee beans so that admixtures of green beans from the two species could be readily detected.

Materials and Methods

Green Coffee Bean Samples and DNA Extraction

Green coffee beans of different Arabica cultivars or accessions were obtained from the International Coffee Germplasm Collection maintained in The Tropical Agricultural Research and Higher Education Center (CATIE), Costa Rica. Twenty Arabica cultivars, breeding lines and farmer selections (hereafter referred to as "cultivars") were sampled, and from each cultivar, three single green beans were used. Green beans of Robusta coffee were obtained from commercial vendors. Shade-grown, high-altitude peaberry robustabeans from Dalat (Annam) Highlands of Vietnam and shade-grown, high-altitude Philippine robusta beans were purchased from Heirloom Coffee LLC (Medford, MA) via Amazon.com. Organic green robusta beans from Madagascar were purchased from Dean's Beans (Orange, MA). The India Josuma Kaapi Royale Premium robusta green coffee beans were purchased from Josuma Coffee Company (Menlo Park, CA). *Coffea liberica* green beans from Indonesia were obtained from Joseph A. Rivera (coffeechemistry.com) and was used as an outgroup reference. The detailed list of green coffee beans is presented in Table 1.

For DNA extraction, the green beans were oven dried at 50°C for 24 h to remove additional moisture. A single bean was then placed between paper towels and crushed with a hammer to a mixture of fine/semi-fine pieces. The crushed bean was then used for DNA extraction with the DNeasy® Plant Mini kit (Qiagen Inc., Valencia, CA), which uses a silica-based affinity matrix. The crushed green bean was placed in a 2 mL microcentrifuge tube with one 1/4-in. ceramic sphere and 0.15 g garnet matrix (Lysing Matrix A; MP Biomedicals, Solon, OH). The samples were disrupted by high-speed shaking in a TissueLyser II (Qiagen Inc.) at 30 Hz for 1 min. Lysis solution (DNeasy® kit buffer AP1 containing 25 mg/mL polyvinylpolypyrrolidone), along with RNase A, was added to the powdered samples and the mixture was incubated at 65°C, as specified in the kit instructions. The remainder of the extraction method followed manufacturers' suggestions. DNA was eluted from the silica column with two washes of 50 μL Buffer AE, which were pooled, resulting in 100 μL DNA solution. Using a NanoDrop spectrophotometer (Thermo Scientific, Wilmington, DE), DNA concentration was determined by absorbance at 260 nm. DNA purity was estimated by the 260:280 ratio and the 260:230 ratio.

SNP Genotyping

SNP loci were identified from expressed sequence tags from a wide range of coffee plant tissues that displayed a good representation within the coffee transcriptomes (Table 2). The detailed process of SNP screening and validation has been reported by Zhou et al. (21). The protocol for SNP genotyping of coffee used the Fluidigm 96.96 Dynamic Array™ Integrated Fluidic Circuit® (IFC) (Fluidigm, San Francisco, CA), which can run 96 samples against 96 SNP assays generating a total of 9216 genotypes. Specific target amplification (STA), was also used to enrich the template molecules for each individual IFC reaction, in order to facilitate the multiplexing during genotyping. An advantage of STA is that it allows the use of limited or low-quality DNA samples, reducing the bias that may occur when samples are loaded to the 96-well plate and reducing the effect of compounds that can potentially inhibit PCR amplification. The STA master mix was composed of 2.5 μL of TaqMan® Taq polymerase (Life Technologies, Carlsbad, CA), PreAmp Master Mix (2X), 1.25 μL of pooled assay mix (0.2X), and 1.25 μL of genomic DNA for a total reaction volume of 5.0 μL.

PCR was performed with an initial denaturation step of 95°C for 10 min, followed by 14 cycles of a two-step amplification profile consisting of 15 s at 95°C and 4 min at 60°C. The resulting amplified DNA was then diluted 1:5 in TE buffer to reduce the concentration of any remaining PCR by-products. Samples were then genotyped using the nanofluidic 96.96 Dynamic Array™ IFC for SNP genotyping, as described by Wang et al. (22).

Table 1. List of 20 Arabica cultivars, four Robusta accessions and one *C. liberica* coffee accession, represented by 81 single green beans used in the genotyping experiment

Sample name	Species	No. of single beans	Sample provider	Country of origin
Madagascar	*C. canephora*	5	Dean's Beans	Madagascar
India Kaapi	*C. canephora*	5	Josuma Coffee Company	India
Philippine	*C. canephora*	5	Heirloom Coffee	Philippines
Vietnam Peaberry	*C. canephora*	5	Heirloom Coffee	Vietnam
Anfilo	*C. arabica*	3	CATIE, Costa Rica	Kenya
BA 21	*C. arabica*	3	CATIE, Costa Rica	India
Barbuk	*C. arabica*	3	CATIE, Costa Rica	Sudan
Bronze	*C. arabica*	3	CATIE, Costa Rica	Congo
Cioiccie-S6	*C. arabica*	3	CATIE, Costa Rica	Ethiopia
CRI Arabica	*C. arabica*	3	CATIE, Costa Rica	Costa Rica
Dilla Alghe	*C. arabica*	3	CATIE, Costa Rica	Kenya
E 118	*C. arabica*	3	CATIE, Costa Rica	Costa Rica
Ennarea-S2	*C. arabica*	3	CATIE, Costa Rica	Ethiopia
Geisha	*C. arabica*	3	CATIE, Costa Rica	Tanzania
H 66	*C. arabica*	3	CATIE, Costa Rica	Puerto Rico
Harar-S10	*C. arabica*	3	CATIE, Costa Rica	Ethiopia
Jimma 6	*C. arabica*	3	CATIE, Costa Rica	Ethiopia
Laurina	*C. arabica*	3	CATIE, Costa Rica	Cameroon
Leekemti	*C. arabica*	3	CATIE, Costa Rica	Ethiopia
Lejeune 12	*C. arabica*	3	CATIE, Costa Rica	Ethiopia
Rume	*C. arabica*	3	CATIE, Costa Rica	Tanzania
SL 28	*C. arabica*	3	CATIE, Costa Rica	Kenya
Villalobos	*C. arabica*	3	CATIE, Costa Rica	Costa Rica
Zeghie-S13	*C. arabica*	3	CATIE, Costa Rica	Ethiopia
Liberica	*C. liberica* (bean)	1	Coffeechemistry.com	Indonesia
Total	26	81		

End-point fluorescent images of the 96.96 IFC were acquired on an EP1™ imager (Fluidigm) and the data was recorded with Fluidigm Genotyping Analysis Software.

Statistical Analysis

Raw data for each SNP locus and sample calls was organized in Microsoft Excel 2007. The genotype consistency among the three single beans representing each cultivar or accessions in the two coffee species was first examined using multilocus matching procedure implemented in GenAlEx 6.5 (23, 24). The three SNP profiles for each cultivar that were fully matched at all genotyped SNP loci were declared as inbreeding progenies from the same cultivar or accession. One accession from each duplicate group was retained for subsequent analysis.

After checking the homogeneity in each cultivar or accession, a multivariate analysis was used to assess the relationship between Arabica and Robusta coffees, as well as cultivars or accessions within each species. The Pairwise genetic distances were computed using the DISTANCE procedure implemented in GenAlEx 6.53 (24). The same program was then used to perform principal coordinates analysis (PCoA), based on the pairwise distance matrix and both distance and covariance were standardized.

In addition to the multivariate analysis, assignment tests were employed to check that all of the coffee beans from the two species were assigned to different "home species" with clearly defined statistical rigor. An exclusion test (25) as implemented in the program GENECLASS (26) was used to identify if there was a presence of the adulterant Robusta bean in the tested samples. Both the Bayesian assignment test and classical frequency-based assignment were applied. In both cases, the resampling algorithm of Cornuet et al. (27) was used and the minimum number of simulated individuals was set at 10 000. The "Type I error" was set at 0.01.

Results

The Polymorphism of SNP Markers

Out of the chosen 96 SNP markers, a total of 65 polymorphic markers were scored across the 80 green beans of Arabica and Robusta coffee. Among the remaining 31 SNPs, eight generated a no call in the Fluidic array and 12 generated monomorphic SNP profiles across all the samples. These SNPs were likely due to the sequence complexity or the presence of polymorphisms within the flanking sequences. In addition, there were 11 SNPs that produced profiles with missing data at least 10% of the time, and these were excluded from data analysis. Among the 65 successful SNPs with high call rate, 23 were polymorphic between the two species but monomorphic within each species, i.e. only one SNP variant was identified in all individuals (Table 3). These within-species monomorphic markers were not useful for differentiating cultivar or genotypes within Arabica or Robusta, but they are useful for differentiating the two species. The within-species polymorphism is much larger in Robusta, which had a total of 50 polymorphic SNPs whereas Arabica only had six polymorphic markers.

The result of multilocus matching showed that most of the tested Arabica cultivars had consistent SNP profiles, where the three single beans fully matched in their multilocus profiles (Table 3). The results showed that these Arabica beans were produced via self-fertilization. However different SNP profiles were detected in cultivars "Lejeune 12," "Zeghie-S-13" and "Jimma-6" (all originally from Ethiopia; Table 1), showing possible outcross fertilization origin of these green beans and/or

Table 2. Examples of inter- and intraspecific SNP genotype profiles based on single green bean fingerprinting (showing truncated profiles)

		Inter-specific polymorphic only				Inter- and intra-specific polymorphic							
Bean samples	Species	Cac19	Cac73	Cac20	Cac178	Cac12	Cac32	Cac230	Ca396	Cac1	Cac206	Cac55	Cac101
Lejeune12-1	Arabica	C T	C T	C G	G G	C C	C T	A T	C T	A A	G G	G G	C C
Lejeune12-2	Arabica	C T	C T	C G	G G	C C	C T	A T	C T	A A	G G	G G	C C
Lejeune12-3	Arabica	C T	C T	C G	G G	C C	C T	A T	C T	A A	G G	G G	C C
H66-1	Arabica	C T	C T	C G	G G	A C	C C	T T	C C	A A	G G	G G	C C
H66-2	Arabica	C T	C T	C G	G G	A C	C C	T T	C C	A A	G G	G G	C C
H66-3	Arabica	C T	C T	C G	G G	A C	C C	T T	C C	A A	G G	G G	C C
Jimma6-1	Arabica	C T	C T	C G	G G	C C	T T	A A	C T	A A	G G	G G	C C
Jimma6-2	Arabica	C T	C T	C G	G G	C C	C T	A A	C T	A A	G G	G G	C C
Jimma6-3	Arabica	C T	C T	C G	G G	C C	C T	A A	C T	A A	G G	G G	C C
BA21-1	Arabica	C T	C T	C G	G G	A C	C C	T T	C C	A A	G G	G G	C C
BA21-2	Arabica	C T	C T	C G	G G	A C	C C	T T	C C	A A	G G	G G	C C
BA21-3	Arabica	C T	C T	C G	G G	A C	C C	T T	C C	A A	G G	G G	C C
Villalobos-1	Arabica	C T	C T	C G	G G	A C	C C	T T	C C	A A	G G	G G	C C
Villalobos-2	Arabica	C T	C T	C G	G G	A C	C C	T T	C C	A A	G G	G G	C C
Villalobos-3	Arabica	C T	C T	C G	G G	A C	C C	T T	C C	A A	G G	G G	C C
Zeghie-S13-1	Arabica	C T	C T	C G	G G	C C	C T	A T	C C	A A	G G	G G	C C
Zeghie-S13-2	Arabica	C T	C T	C G	G G	C C	C T	A A	C C	A A	G G	G G	C C
Zeghie-S13-3	Arabica	C T	C T	C G	G G	C C	C T	A T	C C	A A	G G	G G	C C
Madagascar1	Robusta	T T	C C	G G	A G	C C	T T	A T	C C	A C	G T	G G	C C
Madagascar2	Robusta	T T	C C	G G	A G	C C	T T	A A	C C	A A	T T	G G	A C
Madagascar3	Robusta	T T	C C	G G	A G	C C	T T	A A	C T	A C	G T	A A	A C
Madagascar4	Robusta	T T	C C	G G	A G	C C	T T	A T	C C	A C	G T	A A	C C
Madagascar5	Robusta	T T	C C	G G	A G	C C	T T	A A	C C	A C	G T	A A	A C
India Kaapi1	Robusta	T T	C C	G G	A G	C C	T T	A A	C C	A C	G T	G G	C C
India Kaapi2	Robusta	T T	C C	G G	A G	C C	T T	A A	C C	A C	G T	A A	A C
India Kaapi3	Robusta	T T	C C	G G	A G	C C	T T	A A	C T	A C	G T	G G	A C
India Kaapi4	Robusta	T T	C C	G G	A G	C C	T T	A A	C C	A C	G T	G G	C C
India Kaapi5	Robusta	T T	C C	G G	A G	C C	T T	A T	C C	A C	T T	A A	A C
Philippines1	Robusta	T T	C C	G G	A G	C C	T T	A A	C C	A C	T T	A A	C C
Philippines2	Robusta	T T	C C	G G	A G	C C	T T	A A	C C	A A	G T	A G	C C
Philippines3	Robusta	T T	C C	G G	A G	C C	T T	A A	C T	A A	G T	A A	C C
Philippines4	Robusta	T T	C C	G G	A G	C C	T T	A T	C C	A A	G T	A A	C C
Philippines5	Robusta	T T	C C	G G	A G	C C	T T	A A	C C	A C	T T	A A	C C
Vietnam Peaberry1	Robusta	T T	C C	G G	A G	C C	T T	A A	C C	A C	T T	A A	C C
Vietnam Peaberry2	Robusta	T T	C C	G G	A G	C C	T T	A A	C C	A A	G T	A A	C C
Vietnam Peaberry3	Robusta	T T	C C	G G	A G	C C	T T	A A	C C	A C	T T	G G	A C
Vietnam Peaberry4	Robusta	T T	C C	G G	A G	C C	T T	A A	C C	A C	T T	G G	A C
Vietnam Peaberry5	Robusta	T T	C C	G G	A G	C C	T T	A T	C C	A C	G T	G G	C C

residual heterozygosity of the parental trees. In the Robusta coffee, every individual single bean is different within each accession, demonstrating that all the beans were produced by outcrossing fertilization. After eliminating duplicates, a total of 43 distinctive samples were used for subsequent analysis.

Differentiation between Arabica and Robusta Coffee

The genetic relationships among the 43 distinctive single beans (out of the 80 single bean samples), representing 20 Arabica and four Robusta accessions, together with two samples of *C. liberica*, were presented in the PCoA plot, (Figure 1A). The three PCoA axes, which accounted for 93.9% of the total variation (77.2%, 14.7%, and 2.0% for axis 1, 2, and 3 respectively). The 43 single beans were clearly separated into three species without any ambiguity, demonstrating the differentiation power of this SNP panel. Within the Robusta samples, each of the tested single beans were well separated from one another, but all the Arabica samples formed a tight cluster in the PCoA, showing a much smaller intraspecific variation. The average pairwise genetic distances among the Robusta samples is 21.8, whereas the average pairwise genetic distance among the Arabica bean samples was 3.95. The difference between *C. liberica* and these two species was also clearly reflected in the PCoA plot, which suggested that this panel of SNPs can be used to differentiate green beans of *C. liberica* as well (Figure 1A).

The PCoA plot exclusively showing relationships within the Robusta samples is presented in Figure 1B. Each single bean of Robusta coffee was different. The three PCoA axes explained a total of 64.1% of the total variation (33.9%, 18.9%, and 11.3% for axis 1, 2, and 3 respectively). It also showed that the five single beans from the same accession may not share the same parentage, in spite of their shared origin. For example, bean #2 of "Madagascar" was not grouped with the other four beans of "Madagascar" and beans #3 and #4 of "Philippines" were distinctly separated from the other three beans of "Philippines" (Figure 1B).

Within the Arabica samples, only 11 out of the 20 cultivars can be differentiated even though the samples presumably included a diverse set of Arabica cultivars from African countries such as Ethiopia, Sudan, Congo, Cameroon, Malawi, Kenya, and

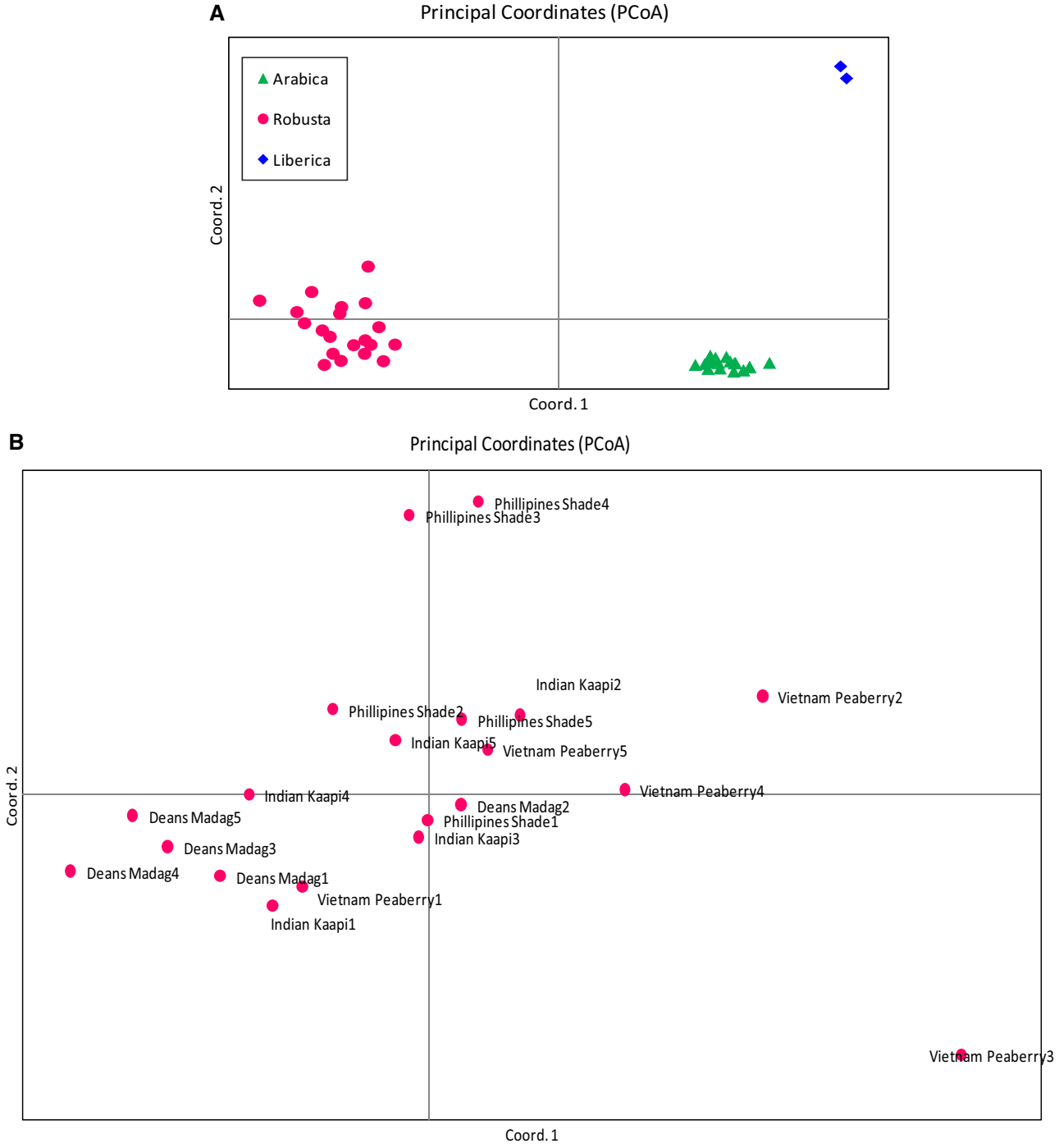

Figure 1. (A) Principle Coordinate Analysis (PCoA) plot of 44 coffee samples representing four Robusta accessions, 20 Arabica cultivars and one accession of *Coffea liberica*. The plane of the first three main PCO axes accounted for 93.9% of the total variation (First axis = 77.2%, of total information, the second = 14.7%, and the third =2.0%). (B) PCoA plot of 20 single green beans, representing four Robusta accessions. The plane of the first three main PCoA axes explained a total of 64.1% of the total variation (First axis = 33.9% of total information, the second = 18.9%, and the third =11.3%). (C) PCoA plot of 23 single green beans, representing 20 Arabica coffee cultivars. The plane of the first three main PCoA axes explained a total of 86.2% of the total variation. (First axis = 46.9% of total information, the second = 24.1%, and the third =16.2).

Tanzania. There were nine cultivars that formed three duplicate clusters. The first duplicate cluster included cultivars "Dila Alghe", "CRI Arabica", and E-118. The second cluster included "Rume" and "Anfilo" and the third one includes "Villalobos", H-66, "Ennarea-S-2", and BA 21 (Figure 1C).

The result of the assignment by the exclusion-simulation method showed that all of the Arabica bean samples could be clearly assigned into one genetic cluster at the threshold probability of >0.90. All the Robusta beans had an assignment probability smaller than 0.001, and therefore could be clearly categorized as admixtures (Table 4). When the algorithm was changed from Bayesian clustering analysis to frequency-based assignment, the overall results remained unchanged (Table 4).

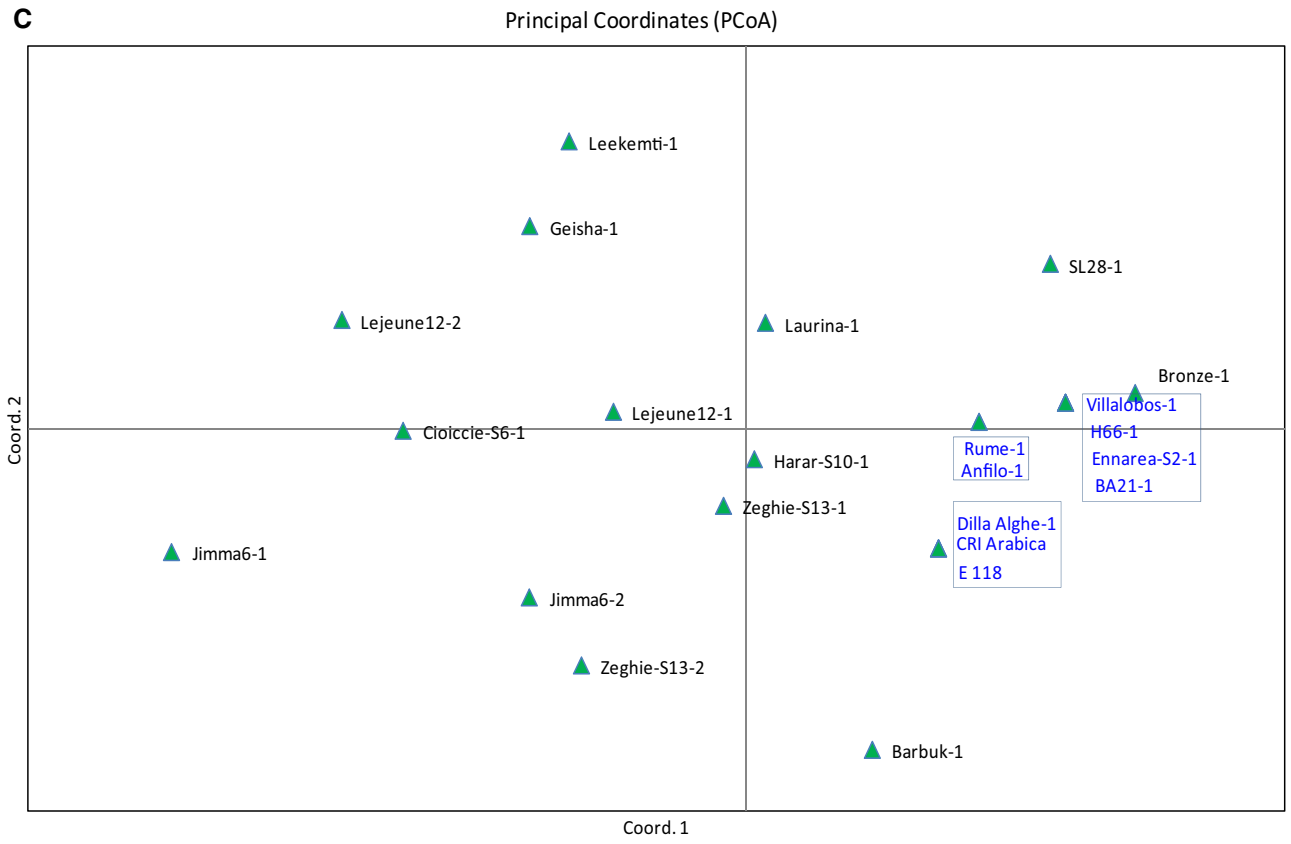

Figure 1. Continued

Discussion

Admixture Detection of Arabica and Robusta Green Beans

An accurate identification of Arabica and Robusta green coffee beans is important for the coffee industry. Various approaches have been reported (see Introduction), however, a high-throughput system is still needed for large scale industry applications and, ideally, such an application would combine species differentiation with cultivar identification in the same process. The present study, therefore, tested the efficacy of using multilocus SNP genotyping on green coffee beans. Reliable SNP fingerprints were generated for all tested samples using a system of nanofluidic arrays. Unambiguous differentiation between Robusta and Arabica coffee was achieved using multivariant analysis and assignment tests with high statistical rigor. This SNP method can handle a large number of samples in a short period and the results are robust and consistent. Our results also showed that the nanofluidic array, along with specific targeted amplification, efficiently dealt with potential problems of quality or quantity of DNA and are particularly suitable for SNP marker analysis in coffees. The STA protocol, which is performed before genotyping, multiplexes primers for all loci of interest without targeting the specific alleles, resolving the problem of low single coffee bean DNA concentrations. The results from the cultivar-specific, three green beans analysis (three independent DNA extractions performed from the same tree) showed 98.5% concordance, thus proving the reliability of the nanofluidic platform for generating highly accurate DNA fingerprints for green coffee beans.

Coffee Cultivar/Genotype Identification within Arabica and Robusta

In addition to the high-throughput and single bean sensitivity features of this platform, another advantage offered by the present method is the potential capacity to simultaneously identify cultivars or genotypes within each species, as suggested in Figure 1B and C. The SNP panel used in the present study successfully differentiated every Robusta bean sample and 55% of the Arabica bean samples. The results of within-species cultivar/genotype differentiation were robust and highly repeatable, as examined in the present experiment by genotyping three single beans from independent DNA extractions. The results showed that this method can be applied to green beans for genotype identification, in addition to species differentiation, in spite of the high concentration of polyphenolic and polysaccharide compounds found in the green coffee bean samples.

A much larger SNP variation was found in the Robusta bean samples, in which 42 SNP markers (out of the total of 65 for both species) were found to be polymorphic. Individual genotype matching (pairwise comparisons) showed that each of the 20 Robusta beans had a unique SNP profile, with a minimum difference of four SNPs between the most similar pair of single bean samples (i.e. Philippines Shade3 vs. Philippines Shade4). This level of polymorphism suggested that this set of SNPs can provide sufficient differentiation power for Robusta genotype identifications. Although, in general, Robusta coffee has a lower commercial value relative to Arabica, there are large within-species variations in Robusta in terms of agronomic traits and bean quality. Moreover, Robusta coffee is easier to cultivate and can produce higher yields than Arabica, due to its broader

Table 3. Sixty-five SNPs and flanking sequences showing inter- and/or intra-specific polymorphism in Arabica and Robusta

Code	Sixth-five polymorphic coffee SNPs and flanking sequences
Cac001	CCGCGAATCCTGTAAACCCAAAGCCACCCAATTCTCTCATGCACACTCACATTCCTTAC[A/C]ACTCTTTACCAGAATCAA CAAAATCTCCAGGCTGCCATATTGTTTATGTTTATCGAGACCC
Cac006	CCCATTTTGGTCCCAGCTGTATTTACCCTAGGGCTGGCCCTGGCCGGGTTCTTGACCTC[C/T]GGTGCTTTCGGGATCACTGCA CTTGCTTCATTGTCGTGGATGCTGAACTACATCCGACTCA
Cac007	TATTGAAGTGATTAATCGTGGAAGAGGAATGTCTGTTTTTGTTTTCTTGCAAATAAGCT[T/C]AATGTTGTGAGGTTTGCT GCAGATGCAGTTGGTTACGCAGTTGTGTTTGCCTGTTCATAGC
Cac008	CCCGTCTTCCACAAAAAGGCCAAAACTACAAAGAAGATTGTGCTAAGGCTGCAATGCCA[A/G]GGTTGTAAGCACGTTTCA CAGCACCCTATCAAGAGGTGCAAGCACTTTGAAATTGGTGGTG
Cac010	ATCTTATTCATAGTCTTCATTATTCTCCTGATTGTCACTGCATTCATCACTGTGGCTCT[C/T]ACATATTTTCAACTTGCTGCT GAAGACCATGAATGGTGGTGGAGATCTTTTCTTTGCGGTG
Cac012	CTAAAAATGCGAGGTGCTGGATCTGGAGCTGGAAGCTTACTTAAAGTTTTGGCCAACAA[C/A]TTTGATGTTCTTGCTTGG CCTGTGGTTAGTTTGGTGTATCCTCTTTATGCTTCCATCAGGG
Cac013	GTGGGTACTTTATTAGGAGCTATGACTGGAGCCTTGATAGGCCAAGAGACTGAAAGTGG[T/A]TTTGTTAGAGGGGCTGCA GTTGGTGCAATATCTGGAGCTGTTTTCTCCATTGAGGTCTTCG
Cac017	GAGTTGATGGAGCAGTTTCAGTTTTTGTCAGAAGAAGCTCTTCAACAGGAGTTCAACGA[A/T]CATGACTTTGTCTTGTAA TTTACCACTTCTGTTCAACACAATTCTTGATCATTAATTGTAC
Cac019	AGAGGCCTTGGCAAGCTCAAGAGGGCTATCATGGGCAGCGTTAGCAACTACGTGGTAAA[C/T]AACGCTTCTTGCCCTGTG ACAGTTGTGAAGAATGTTGAACATGATTGATCTTACCTCACAT
Cac020	GAGCACTTGGCAAGTTTTCGGGCAATGCCTAATATTTTGATGTTGCGTCCAGCTGATGG[C/G]AATGAGACAGCTGGTTCT TACAAGGTAGCTGTCCTCAATAGGAAGAGACCATCAATCCTTG
Cac030	GGAGTGTTGGAGGAAGTAATGAAGGGACTGTACTACGGAACTAAGGAGACCGTGGGTTG[C/T]GCTGCTGAGATGGTGAAG AGGAATGCTGTTGAGATCGGGGACTTCAGATTCTTTGATGGAT
Cac032	GATGCTCAGCGAAGGCTGAGTAAATCCCATATACTTGTCAGTGGACTTACAGGCACTGT[C/T]GTTGAGTTCTGCAAGAAC ATTGTCCTTGCTGGAGTTGGTAGTTTGACATTGAATGATGATC
Cac036	TGGTACGGCAAGTTTACGCCCACCCAGCGCTCCATTATTGTCGATTTTCTTCAATCCCT[T/A]AACTCCCCCAGGGCGGCTTCT CCCTCCGCCGCCTCCTGGTGGATGACGACCGAGAAGTACA
Cac040	CTCGTGCTATCTATTGCTGATGTTTATTGCCCTCCTAGAAAAACGCTCACGTGTTAGCGC[A/G]CCATATGCTGTTGACAGTCGT TTGTTTAACAAAGAGCGGAACCCTTCCATTGAAACTCTTC
Cac048	TTTGAAATCAAATGTATGAATGACCCAAAAGCTTGCCTTCCTGGTTCCATTATTGTCAC[G/A]GCTACCAACTTTTGCCCTCCT AACAATGCACTCCCAAACAATGATGGAGGCTGGTGCAATC
Cac050	TCTCTATATGGCCTACCAAATTCCAAAACAAGACTAATGGGATTACTCCTCGCCGGTGG[A/C]TTCGGTTTTGTAGTCCTG AGCTTAGTCAAATAATAACCAAATGGTTAAAAACTGATAAATG
Cac054	GAAGAAGGATTCATGGTCGACCCATCTGTCTGGCGTTAGAGCCACTTTAACATTTGATC[A/C]CCCTACAACCAATAAGGA GAACTCCAAGCAAAGGAAGCACACTGTGGATCCCAATGCTCCT
Cac055	CACACAACGTGTCCACAAAGAAGAGAAAAGAGATAGTGGAGCGTGCAGCTCAGCTAGAT[G/A]TTGTTGTTACTAACAAGC TTGCTAGGCTGCGGAGCCAGGAGGATGAATGAGCTTTTTGGCT
Cac056	GCGGAGCCAGGAGGATGAATGAGCTTTTTGGCTTTCATGCGCTGAACTTGGTTCATTTT[A/G]ATTATCATGCCTTTTTGTTTA CGCTATCCCTTGTTAGTAACATTGTAATTTTGACAGATGA
Cac061	GACAAGGTTGCATTTGATGAACAAACCGGCACTTGGAAGCGTCGCCATGGTTATGATCG[T/C]GTCAATGATGACAAAGAT GTACCAATCATTGAGGCCAAGATGAATGATGAGCCAGGGGAAG
Cac063	TGGTGACATGTTTTGGTGTCTTTCTGAGGATTATTTTGAGAACAAAAGGTGACTGCTAC[T/C]AAAACCATGTGTTGCGCT CTATCTACTTCCAGAAGGTGAGTTTTCACGGAAACCCTGTTGT
Cac065	CCTGTTGTAACTGTTCATAGGGATATAATTATCTGTAGCAAAATGTCAGGAAGTGATTG[C/T]TGTTGCTTTGTTTTCATTGTG GAACATCAATTGCGGTCTACTGCTGTATATTGGCTTGTAT
Cac072	TTTTTTCCTTTTCTCGGTTCTGTGAATTATTTTTATCCAATCCTCTTTTGCGGTTAATA[T/G]CTGAAAGCCTATCAAGATGGT GAAGTTCACAGCAGAAGAGCTTCGAAGGATTATGGACTAC
Cac073	CGTAATATGTCTGTTATTGCACATGTTGATCATGGGAAGTCCACTCTTACTGATTCTCT[T/C]GTGGCTGCTGCTGGTATCATT GCTCAAGAAGTTGCTGGAGATGTTCGAATGACGGATACAA
Cac083	ATTCTTTTAGACTCGGAATTCAATGCAAAGTTGTCCGACTTTGGCTTGGCCAAGGCAGG[G/T]CCAACTGGTGATAGGACT CATGTATCCACCCAAGTTATGGGAACACACGGCTATGCTGCAC
Cac087	AATATCGCCACGCCTCATCACCATGAGGTTGGCTATCAGGGCTATGGGCAGCAGCACAG[C/A]ATTAATGGTGATGGGTAT GGGAATCACCACAAGTACAATGACTACAACAGCCATGGCTATG
Cac092	TCCACCCTCTACGCCGTAGGCAGCCGTTCGGTGGAAAAAGCCTCAAACTTTGCGAAAGA[T/G]AATGGCTTTCCGGCTTCA GCAAAGGTATACGGCAGTTATGACGCCCTTCTAGATGACCCAG
Cac093	ATGGTTTTCATGGGTCCGACATTTTATCAACGCCTTACTCATATGGCTGAAGATAAAGT[C/A]AAATTTCGGAACACGGGA CCAGTCCATCCACTCACTCGTCAGCCAGTGGCAGACAGGAAGC
Cac094	CACAAGTCCAAGAAAGATCCTGAGCATGCCCACAAGCACAAGATAGAAGAAGAGATTGC[A/T]GCAGTAGCTGCAGTGGGT GCCGGTGGATTTGCGTTCCATGAGCATCACGAGAAGAAGGATG
Cac101	ACCACGTGCTGTCTTGTCCAGTCCAGATAATGATCAAATGATTGGAAGCAAAAACAAG[A/C]CAAAAGGTGATATACTTG CCAGTATGAAAAGGCAAAGTCTGTTTGAGAATAGACACGCCCGG
Cac102	GTTTGAGAATAGACACGCCCGGTGTAAGGTTACTCCGAGGCCTGTTGCTGCTGATGGCT[C/G]TATAAGCACAAGAACATC GCTTAAGGAAGTACCTGACGGTAAAGGTGATCTTCGAACTAGA

(continued)

Table 3. (continued)

Code	Sixth-five polymorphic coffee SNPs and flanking sequences
Cac103	GTGGAAGGCTACAGCCAGCATGACAATGAAGTCCAATATAGTCTCCTATTAGTTACTTA[A/T]AGGAATAAAGAGACTACT CATTTGAACTTCACAAATATGAACTTTATGATGTATTTTCTGT
Cac107	CCAACAAAAGCAAGAAAGCATGAACGCCGAGGTTATTTCCAGCGGTCACCTGCAACTTT[A/T]CCCAACAAGTTCAAATAC AAGACATTGTGGCTCCCAGCTTCCATTTAGCGGCAGGGCCAGT
Cac108	AGAGCAAAGACACCCTTCTCAAGCCAAGAGCCCCTGCTTCCATTCCTCTCTTGTCCCTC[A/C]ACGATGGACAAACTGAAG CTGTTTTCTACCGGTGCGGCTTTGGTTACAATTGTAACTATGT
Cac111	CATCTTCGTCAACTTCACATGACCAGTCACGACCCCGTAATGCAGGGTCAACTGGAAGA[G/A]CATCTGGTGCATCTACAA CACAAACTCCTAGTGCAACATCTCTGCGATGGGATCGGCAAAC
Cac148	AAGTTCTTGAAGAATGGTGATGCTGGTTTTGTCAAGATGATTCCCACCAAGCCCATGGT[G/T]GTTGAAACTTTCTCAGAG TACCCTCCCCTTGGTCGTTTTGCTGTTAGGGACATGCGACAAA
Cac149	AAGTCCTGAAGGTAAAGTCCCTGTGTTAAAGCTGGAAGATAAGTGGATTCCAGATTCAG[G/A]CGTTATTACACAGGCAAT AGAAGAAAAGTTTCCTGAACCGCCATTGGCAACGCCACCTGAG
Cac150	GCCCCAAAGATGCCTTAGTCGGCGGTTGGAGTAAGGCTGACCCCAAGGACCCAGAGGTG[C/G]TAGAGAACGGAAAATTTG CCATAGATGAGCACAACAAGGAGGCCGGTACCAAGTTGGAGTT
Cac152	TATGGAATGACAAGCGTGAAAGGGAGATGTATGACAACTTTGCTGAGCTTTTTGCAATT[A/G]TAAAGGCCACCGAGAAGC TTGAAAAAGCTTATGTTCGTGACATCATATCACCAATCGAGTA
Cac155	CACTCAAAGGTGGGTGCTGATGAAGAAGAGGAGCCTGAGATAATCGAATCTGATGTTGA[C/G]CTTGATGACACTGAAGTT GTGGAGCCTGATAATGATCCTCCGCAACAGATGGGAGACCCTT
Cac158	TGCAACACCGTTTTCCTATTCTCTATTTTTCTTGCCCTAATTTTCCTTCAGAGCACTGC[G/A]GATGATGTTGGGAGTGTTGTC GATGATTTCTCCAAGGATGGCTCTGATTTGTCAGCTGAAT
Cac160	AGTTGAAAGCGAAGGATGATGCAATTTCTGAGCTGGAGAAGAAGATTAAAGAGAAATCT[A/G]ATAGCATTTCTTCATTGC AGAGTGAGATAGTGTCTCTGCAGAAAAAAGGAACTCTAGATGC
Cac169	TGGGACATGATGCTTCCAAACGAAGACCCTTTTAGAATCTTGGAGCACAGCCCTTTAAC[T/G]GTCCCCAAAGGGGTGGAG ACGCTGGCCTTGGCACGCGCTGACTGGAAGGAGACGGCGAAGG
Cac171	CATTTGACATCACTTGATGGAGCTAAGGAAAGGCTTCAGTTGTACTCGGCAAACTTACT[G/A]GAAGAGGGATCGTTTGAT GCAATAGTCGAGGGATGTGAAGGGGTTTTCCATACTGCATCTC
Cac175	TTTAACGATGTAATCGAGAAAATCTGTTGTGTCATCAAATTTGAACCCTCTGCTGATGG[G/A]GGTTCAATCTGCAAAACC ACTAATACATACTACCCCAAAGGTGGTGCTCAGATCAGTGAGG
Cac176	GCCAGGGCCAGCCCTGCTCAAGCTAGCATGGTTGCACCCTTCACCGGCCTCAAAGCTGC[A/T]TCTTCTTTCCCCATTTCCAAG AAGTCCGTCGACATTACTTCCCTTGCCACCAACGGTGGAA
Cac178	TCAACCGCGTGAATGGCGGCCTGCAGTGGAAAATTGTTATTGGCACTCTCTATATCCTT[A/G]TCCTTGCAACTCAGGATT CTAAGGGCACATATACCGATTATGCAGTGGTTTTTGAGACCTT
Cac184	GCCTGAGGCAGTCCTTCAGACTGTTTCAAAGACCGGGAAGAAGACTTCTTTCTGGGAAG[G/A]AGGAGCATCAGCTGCACC TGAATCGAAGCCCGCAGAAACTGTTGCAGCTGCATAATTTGGG
Cac191	AAGTGCTTTCATTTTTGTGTCACCTCATGACTATCGTTTGGAATGGTGTTTTACACCTT[A/T]TGTGCGGAAAGTTGCATATCT TTGGTTACTCAGATAACGGATGAGGATGTTCAACAGCTAG
Cac194	AGTCGAAACTGATGCGGCGGTGACCAAATTCAAGAAAGTCATCTCTCTTCTAGGCCGAA[G/C]CAGAACTGGCCATGCTCG TTTTAGAAGAGGCCCCGTCCCCGTGGCTACAAATCCGGTTCCT
Cac196	AAATGGGGTAATAAAAATCCCTCCTTTTCCTGTTGGAGATGTTACAGAAACTTTCTTCA[G/A]AAATTTAGTGGCATACAA AAAATACATTCTTACCGGCTATGAGGATCGCAGATGTGTAACT
Cac197	GGACTTCACAATGGCAGGAATTTGATTGGAGGCATCAATAGCAAAAGGGCTTCAACATG[T/G]AAAGCAAATGCCTTCCCA GATTGGCCATTGATGGCAGTACTGGTTGAGCATGCTGAAGGAC
Cac206	GAGTTTGCTACTCGTCTGGGTAATGTCTTCATCATTGGAAAAGGTGCAAAACCCTGGGT[T/G]TCTCTTCCAAAGGGCAAA GGTATCAAGTTGTCAGTTATAGAGGAACAAAGGAAGAGGATTG
Cac208	GATCAATATCAACAATTGCATCAACAGCAGCAATTTGTTCCTACCAGTAATCAATACAT[T/G]CACCATCATCCTGGGGCA GTGCCAATGTATTATTACTCCATTTACCCTCCTCAGCAGTTAG
Cac214	TCTCTGTTTAACTTCATAAGCTGTCAAATATTTAACGTCACCGTCACCGTCAGTTCTGC[A/T]GAAAATGTCGCTGATTCCAAG CGTCTTCGGTGGCCGAAGAAGCAACGTTTTCGACCCATTT
Cac222	ATATATTGCTTCTACTCTCTTCCCTTTTCCTTTCTCCCTTCTCCCCTTTTCCAAATTAA[C/G]TCCCGCTGATCATTTCTCTTT TCCAAATTCTCTCCTTCCTTTCTTAAATCCACCGCCCCCT
Cac225	AGGCACAAGATAGGAGAAGAGATTGGAGCGGTAGCCGCACTTGGAGCTGGTGGATTTGC[A/T]TTCCATGAGCATCACGAG AAGAAGGAAGCTAAAGAAGAAGAGGAAGAGGCCGAGGGGAAAA
Cac230	AGGGATAAGGTAGTTCTTGATGTGGGTGCTGGGACTGGAATCCTCTCGCTGTTTTGTGC[A/T]AAAGTTGGAGCAAAACAT GTCTATGCCATTGAATGCTCCAGTATGGCTGACATGGCGCGAG
Cac320	AGATCTGAGGAGTATGCTTGAAGAAGGAGATTTTGATGGAGATGGGGCTCTCAATCAAA[T/C]GGAGTTCTGCGTGCTCAT GTTCAGATTGAGTCCTGAGTTGATGGAGCAGTTTCAGTTTTTG
Cac331	CGTTTCGGATGAAGCAAGGCTTTGCTGGGATTGTGATGAAAAAGTGCACAGTGCAAATT[C/T]TCTGGTCGCCAAACATTC AAGAAATCTTCTTTGCCATGTTTGTCAATCTCCGACCCCATGG
Cac346	ATTTCCAAGGCCATTGCTGAACATAAGCTCAGAATCATTTGCAACTGTAATTAACAGTG[A/G]AATCGGGAGAACCCTGGA GCCACCTTTTGATCCTTATGCTAATGACATCAACTATCTCATT
Cac359	ATAAAAGAAAATGAGAAATTCCAGAAGAATGGTGAAAGTCAGGTTGGCCTGAAGCAGCT[G/A]GTTCAAGACACCAAGCAG GGCCATGATGTCAAGTACGTCTATGTGTGGCATGCTCTAGCAG

(continued)

Table 3. (continued)

Code	Sixth-five polymorphic coffee SNPs and flanking sequences
Cac392	GCTGGGGCTCAGAACGGTGTCGTTCTTGTCCAGAAGGAGAAGGATAAGGAGACGGCTGC[T/G]GCTCCTGCCGCTGCTTATTTGTCAATGGTGGACCCTTTTTTGGTTGAGGCCCTGCAAAATC
Cac396	CGCAACAGTTGAAAAATCCCAAGTTGAGGATGAAGGTGTCAATTTCTTATGATTTAGAT[T/C]ACCCTGATACTGAGAAGGAAGGGAAGAGTGATAAACAGGTTAAGAAGACCAAGAGGAAGCA
Cac398	GGGACTCGTTAAGAGGGAGGAGCTTTTCATTACTACCAAGCTGTGGAATTCAGACCATG[T/C]CCACGTTCTCGAGGCTTGCAAAGACAGCCTGAAAAAGCTTCGTCTTGATTATCTTGACCTG

Table 4. Self-assignment test of single green beans of Arabica and Robusta coffees for differentiating resident samples from non-resident ones, based on their multilocus SNP profiles

		Probability of self-assignment	
Sample name	Species	Bayesian method	Allele freq.
Villalobos-1	*C. Arabica*	0.984	0.995
Laurina-1	*C. Arabica*	0.975	0.962
Rume-1	*C. Arabica*	0.952	0.966
Anfilo-1	*C. Arabica*	0.998	0.966
Barbuk-1	*C. Arabica*	0.986	0.991
Della Alghe-1	*C. Arabica*	0.951	0.962
BA21-1	*C. Arabica*	0.961	0.951
Jimma6-1	*C. Arabica*	0.991	0.983
Jimma6-2	*C. Arabica*	0.978	0.963
Ennarea-S2-1	*C. Arabica*	0.934	0.986
Harar-S10-1	*C. Arabica*	0.986	0.998
SL28-1	*C. Arabica*	0.947	0.965
Cioiccie-S6-1	*C. Arabica*	0.986	0.991
Geisha-1	*C. Arabica*	0.965	0.934
Leekemti-1	*C. Arabica*	0.934	0.945
Zeghie-S13-1	*C. Arabica*	0.955	0.969
Zeghie-S13-2	*C. Arabica*	0.982	0.985
Bronze-1	*C. Arabica*	0.972	0.984
E 118	*C. Arabica*	0.922	0.914
Lejeune12-1	*C. Arabica*	0.995	0.991
Lejeune12-2	*C. Arabica*	0.975	0.986
H66-1	*C. Arabica*	0.985	0.991
CRI_Arabica_1	*C. Arabica*	0.992	0.996
Deans Madag1	*C. canephor*	<0.001	<0.001
Deans Madag2	*C. canephor*	<0.001	<0.001
Deans Madag3	*C. canephor*	<0.001	<0.001
Deans Madag4	*C. canephor*	<0.001	<0.001
Deans Madag5	*C. canephor*	<0.001	<0.001
Indian Kaapi1	*C. canephor*	<0.001	<0.001
Indian Kaapi2	*C. canephor*	<0.001	<0.001
Indian Kaapi3	*C. canephor*	<0.001	<0.001
Indian Kaapi4	*C. canephor*	<0.001	<0.001
Indian Kaapi5	*C. canephor*	<0.001	<0.001
Philippines Shade1	*C. canephor*	<0.001	<0.001
Philippines Shade2	*C. canephor*	<0.001	<0.001
Philippines Shade3	*C. canephor*	<0.001	<0.001
Philippines Shade4	*C. canephor*	<0.001	<0.001
Philippines Shade5	*C. canephor*	<0.001	<0.001
Vietnam Peaberry1	*C. canephor*	<0.001	<0.001
Vietnam Peaberry2	*C. canephor*	<0.001	<0.001
Vietnam Peaberry3	*C. canephor*	<0.001	<0.001
Vietnam Peaberry4	*C. canephor*	<0.001	<0.001
Vietnam Peaberry5	*C. canephor*	<0.001	<0.001

adaptability and better resistance to diseases and pests (28). Recently, fine flavored Robusta coffees have been branded and new standards for fine flavored beans have been developed (29), piloted by a grant from the United States Agency for International Development. Therefore, this SNP panel can be used to assist traceability and authentication of fine-flavored Robusta coffee, thus supporting the market differentiation for Robusta coffee.

However, the current SNP panel has not reached the same level of functionality for differentiating every cultivar/accession in Arabica, because of the much smaller intra-specific variation in *C. arabica*. Nine cultivars were found to fall into three synonymous (duplicate) groups (Figure 1C), although they are from diverse geographical origins. These results suggest that cultivars from each duplicate cluster shared similar origin, and were possibly mutants or offspring derived from a single original cultivar. Moreover, the SNPs used in the present study were originally selected based on the result of genotyping good DNA samples extracted from freeze-dried coffee leaves. When these SNPs were applied to coffee beans, some of them had high rates of no calls (>10%), likely due to the lower quality of DNA extracted from green coffee beans (30). Therefore, more SNP markers will be needed to genotype coffee bean samples and assess genetic relationships among the different Arabica cultivars. Now that the draft genome of *C. canephora* and *C. arabica* have been sequenced (31–33) and large amounts of genomic resources have been deposited in the public domain, mutation localization and identification of causal sequence variants are possible. We are currently developing more SNP markers through whole genome resequencing of 40 of the most commonly cultivated Arabica cultivars. It will soon be possible to develop a more powerful SNP panel that can differentiate all Arabica cultivars. Both the Arabica and Robusta green beans tested in the current study were retained for future validation.

Conclusions

We conducted a pilot study on genotyping green coffee beans using SNP markers. We genotyped 80 single bean samples, representing 20 different Arabica cultivars and four Robusta accessions, using a nanofluidic array. This technology enabled us to generate high-quality SNP profiles based on DNA extracted from both Arabica and Robusta coffees. Together with forensic statistical tools, these SNP-based DNA fingerprints allowed unambiguous differentiation of green coffee beans from the two species. The results showed that the detection of admixture of green coffee beans can be done in a high-throughput fashion. Moreover, the multilocus SNP approach can differentiate all Robusta single beans and 55% of Arabica beans, which can be further improved and applied for differentiating cultivars or genotypes within Arabica and Robusta coffees. This advantage, together with the single-bean sensitivity, suggests significant potential reliability for practical application in the coffee industry.

Acknowledgments

Guest edited as a special report on "Green and Roasted Coffee Authentication: Species, Origin and Diluent Methods of Analysis" by Brian T. Schaneberg.

The authors would like to give special thanks to Stephen Pinney and Stephen Emche (USDA, ARS) for assistance with SNP genotyping using the nanofluidic array system and Sue Mischke for the editing of this manuscript. The authors also thank the editor and two anonymous reviewers for their insightful comments and suggestions. Mention of trade names or commercial products in this publication is solely for the purposes of information and does not imply an official endorsement or recommendation by the United States Department of Agriculture or the Agricultural Research Service of the product to the exclusion of others that may also be suitable.

JAOAC does *not* publish color figures in the print version. Color images are published online only.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this article.

References

1. The Open University, The Value of Coffee, http://www.open.edu/openlearn/people-politics-law/the-value-coffee/content-section-0?intro=1 (accessed May 1, 2019)
2. Davis, A.P., Govaerts, R., Bridson, D.M., & Stoffelen, P. (2006) *Bot. J. Linn. Soc.* **152**, 465–512
3. Davis, A.P., Tosh, J., Ruch, N., & Fay, M.F. (2011) *Bot. J. Linn. Soc.*, **167**, 357–377
4. International Coffee Organization, http://www.ico.org/ (accessed May 1, 2019)
5. World Bank, Commodity Markets, http://www.worldbank.org/en/research/commodity-markets (accessed May 1, 2019)
6. Bertrand, B., Vaast, P., Alpizar, E., Etienne, H., Davrieux, F., & Charmetant, P. (2006) *Tree Physiol.* **26**, 1239–1248
7. Bertrand, B., Villarreal, D., Laffargue, A., Posada, H., Lashermes, P., & Dussert, S. (2008) *J. Agric. Food Chem.* **56**, 2273–2280
8. Martin, M.J., Pablos, F., Gonzalez, A.G., Valdenebro, M.S., & Leon-Camacho, M. (2011) *Talanta* **54**, 291–297
9. González, A.G., Pablos, F., Martín, M.J., León-Camacho, M., & Valdenebro, M.S. (2001) *Food Chem.* **73**, 93–101. doi:10.1016/S0308-8146(00)00282-X
10. Procida, G., Campisi, B., Seraglia, R., & Traldi, P. (2003) *Rapid Commun. Mass Spectrom.* **17**, 140–148
11. Gil-Agusti, M.T., Campostrini, N., Zolla, L., Ciambella, C., Invernizzi, C., & Righetti, P.G. (2005) *Proteomics* **5**, 710–718
12. Casal, S., Oliveira, M.B., & Ferreira, M.A. (1998) *J. Liq. Chromatogr. Relat. Technol.* **21**, 3187–3195
13. Casal, S., Alves, M.R., Mendes, E., Oliveira, M.B.P.P., & Ferreira, M.A. (2003). *J. Agric. Food Chem.* **51**, 6495–6501
14. Ky, C.L., Louarn, J., Dussert, S., Guyot, B., Hamon, S., & Noirot, M. (2001) *Food Chem.* **75**, 223–230. doi:10.1016/S0308-8146(01)00204-7
15. Andrade, P.B., Leitao, R., Seabra, R.M., Oliveira, M.B., & Ferreira, M. (1998) *Food Chem.* **61**, 511–514
16. Toci, A.T., Farah, A., Pezza, H.R., & Pezza L. (2016) *Crit. Rev. Anal. Chem.* **46**, 83–92
17. Burns, T.D., Tweed, L., & Walker, M.J. (2017) *Food Anal. Methods*. **10**, 2302–2310. doi:10.1007/s12161-016-0756-3
18. Trantakis, J.A, Spaniolas, S., Kalaitzis, P, Ioannou, P.C., Tucker, G.A., & Christopoulos, T.K. (2012) *J. Agric. Food Chem.* **60**, 713–717
19. Spaniolas, S., Bazakos, C., Tucker, G.A., & Bennett, M.J. (2014) *J. AOAC Int.* **94**, 1114–1120
20. Combes, M-C., Joët, T., & Lashermes, P. (2018) *Food Control* **88**, 198–206
21. Zhou, L., Vega, F.E., Tan, H., Ramirez Lluch, A.E., Meinhardt, L.W., Fang, W., Mischke, S., Irish, B.M., & Zhang, D. (2016) *Trop. Plant Biol.* **9**, 82–95
22. Wang, J., Lin, M., Crenshaw, A., Hutchinson, A., Hicks, B., Yeager, M., Berndt, S., Huang, W. Y., Hayes, R. B., Chanock, S. J., Jones, R. C., & Ramakrishnan, R. (2009) *BMC Genomics* **10**, 561. doi:10.1186/1471-2164-10-561
23. Peakall, R., & Smouse, P.E. (2006) *Mol. Ecol. Resour.* **6**, 288–295
24. Peakall, R., & Smouse, P.E. (2012) *Bioinformatics* **28**, 2537–2539
25. Rannala, B., & Mountain, J.L. (1997) *Proc. Natil. Acad. Sci. USA* **94**, 9197–9201
26. Piry, S., Alapetite, A., Cornuet, J.M., Paetkau, D., Baudouin, L., & Estoup, A. (2004) *J. Hered.* **95**, 536–569
27. Cornuet, J.M., Piry, S., Luikart, G., Estoup, A., & Solignac, M. (1999) *Genetics* **153**, 1989–2000
28. Vega, F.E. (2008) *Am. Sci.* **96**, 138–145
29. Coffee Quality Institute & Uganda Coffee Authority, Fine Robusta Standards and Protocols, https://coffeestrategies.com/wp-content/uploads/2015/04/compiled-standards-distribute1.1.pdf (accessed May 1, 2019)
30. Mishra, M.K., Sandhya Rani, N., Ram, A.S., Sreenath, H.L., & Jayarama, N. (2008) *African. J. Biotechnol.* **7**, 409–413
31. Denoeud, F., Carretero-Paulet, L., Dereeper, A., Gaëtan, D., Guyot, R., Pietrella, M., Zheng, C., Alberti, A., Anthony, F., Aprea, G., Aury, J.-M., Bento, P., Bernard, M., Bocs, S., Campa, C., Cenci, A., Combes, M.-C., Crouzillat, D., Da Silva, C., Daddiego, L., De Bellis, F., Dussert, S., Garsmeur, O., Gayraud, T., Guignon, V., Jahn, K., Jamilloux, V., Joët, T., Labadie, K., Lan, T., Leclercq, J., Lepelley, M., Leroy, T., Li, L.-T., Librado, P., Lopez, L., Muñoz, A., Noel, B., Pallavicini, A., Perrotta, G., Poncet, V., Pot, D., Priyono, Rigoreau, M., Rouard, M., Rozas, J., Tranchant-Dubreuil, C., VanBuren, R., Zhang, Q., Andrade, A.C., Argout, X., Bertrand, B., de Kochko, A., Graziosi, G., Henry, R.J., Jayarama, Ming, R., Nagai, C., Rounsley, S., Sankoff, D., Giuliano, G., Albert, V.A., Wincker, P., & Lashermes, P. *et al.* (2014) *Science* **345**, 1181–1184. doi:10.1126/science.1255274
32. U. C. Davis Coffee Genome Project, The University of California, Davis California. http://coffeegenome.ucdavis.edu/ (accessed May 1, 2019)
33. Tran, H.T.M., Ramaraj, T., Furtado, A., Lee, L.S., & Henry, R.J. (2018) *Plant Biotechnol. J.* **16**, 1756–1766. doi:10.1111/pbi.12912