JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

ARTICLES

DNA Extraction and Analysis from Processed Coffee Beans

CHIARA MARTELLOSSI,†,‡ EMILY J. TAYLOR,† DAVID LEE,†
GIORGIO GRAZIOSI,*,‡ AND PAOLO DONINI†

NIAB, Huntingdon Road, CB3 0LE Cambridge, United Kingdom, and Università degli Studi di Trieste, piazzale Europa 1, 34127 Trieste, Italy

The authenticity of coffee is an important issue for both producers and consumers. Premium Arabica material is especially prone to being adulterated, and a number of different techniques have been employed to determine the quality of both roasted and instant coffee. Currently, assessment of coffee authenticity relies on chemical methods which can discriminate between coffee species, but not varieties. Several genetic markers are available for assessing coffee origin, but their suitability to testing commercial coffee is limited by the ability to extract DNA from highly processed beans. In this paper, we demonstrate that PCR-grade DNA may be obtained from roasted beans and even instant coffee. This would allow analysis of commercial samples, provided that suitable markers for species/variety identification are found.

KEYWORDS: Coffee; DNA extraction; PCR; authenticity

INTRODUCTION

World coffee production relies on just two species, *Coffea canephora* (commercially known as "Robusta") and *C. arabica* ("Arabica"). The latter alone accounts for 70% of the market: the former is usually employed in blends, for espresso or instant coffee. Arabica coffee is associated with better quality products and is sold at a higher price; even within this species some varieties, such as Kona or Jamaican Blue Mountain, are specially prized and command a premium.

"Green", unroasted coffee may be adulterated by producers with the addition of husks, low-rate varieties, and Robusta beans. Also retailers may add inferior grade or even extraneous material such as chicory root, acorns, or barley to create sophisticated products, especially when roasted beans are sold loose after grinding; nowadays, the subject is aggravated by the possibility that the material may derive from genetically modified plants. It has also been reported that fraudulent retailers have traded green Arabica beans from South America labeled as premium Hawaii Kona coffee (*1*). The presence of husks or alien species in coffee may be revealed through chromatographic (*2*) or spectroscopic techniques (*3*), and D-5-avenasterol is key evidence for the presence of Robusta material (*4*), but adulteration from lower grade Arabica varieties is more difficult to identify.

The use of DNA molecular markers to identify species or varieties within a blend is a way to guarantee a product of constant quality and prevent adulteration with extraneous material or low-grade varieties. Molecular markers such as microsatellites have been employed with good results for the characterization of *Coffea* species, enabling the discrimination between Robusta and Arabica and even among different Arabica varieties (*5*).

The main problem with using diagnostic DNA sequences for commercial coffee testing lies in the extent of its degradation during processing. Nucleic acids decompose rapidly at high temperatures such as those reached during roasting (up to 245 °C); therefore, the DNA extracted from roasted beans is fragmented and denatured, making further analysis difficult. Other chemical alterations, such as depurination and nicking, make the DNA unsuitable for analysis, even though it may appear intact when visualized in a nondenaturing agarose gel. Furthermore, several compounds that are present in the seed or originate during roasting are copurified together with the DNA in the extraction process and may hinder the PCR assay; in particular, polyphenols and acidic polysaccharides are known to inhibit the Taq polymerase (*6*). During roasting, sugars in the bean break down to more soluble compounds (*7*), and these are then found in the final extract.

We focused on PCR-amplification techniques because these methods are generally sensitive and reproducible, capable of providing results even when testing scarce amounts of degraded DNA. To minimize the deleterious effects of DNA degradation, we analyzed chloroplast sequences, which are present in high copy number, other highly repetitive genomic sequences, and

* To whom correspondence should be addressed. E-mail: graziosi@units.it. Phone: +39 040 811836. Fax: +39 040 810860.
† NIAB.
‡ Università degli Studi di Trieste.

10.1021/jf050776p CCC: $30.25 © 2005 American Chemical Society
Published on Web 10/04/2005

Table 1. Extraction of DNA from Green and Processed Coffee with Different Procedures

kit/protocol	starting material mass	yield (green beans) (ng)	PCR	yield (roasted beans) (ng)	PCR	yield (instant coffee) (ng)	PCR
ClonTech NucleoSpin Plant Kit	50 mg	250	yes	~80	yes		
Macherey-Nagel NucleoSpin Tissue Kit	50 mg	120	no	none	no		
Promega Wizard Magnetic DNA Purification Kit for Food (plus GeneClean II purification)	1 g	40–100	yes	~100 (purified)	yes	~100	yes
Qiagen DNeasy	100 mg	250	yes	none			
Qiagen Mixer Mill	50 mg	250	yes	none			
Qiagen Stool Kit	220 mg	250	no	none			
ClonTech NucleoSpin Food Kit	50 mg	NA		~80	no		
CTAB	1 g	200-400	yes	~100	no		

short microsatellite markers, therefore increasing the chance of obtaining amplicons from intact, full-length copies of the target fragments.

EXPERIMENTAL PROCEDURES AND RESULTS

All the samples used for DNA extractions were kindly supplied by Kraft Foods. These consisted of 25 samples of green coffee beans, 3 samples of Arabica beans roasted to 8.4, 10.3, and 18.8 *L*, 3 samples of Robusta roasted to 10.2, 11, and 19.1 *L*, (a higher *L* value corresponds to a lower degree of roasting), and 6 samples of instant coffee from different stages of the production process.

Green Coffee. Beans were ground to a fine powder in a fixed-speed Moulinex Super Junior coffee grinder. To minimize binding of phenolic inhibitors to DNA, 4% of either polyvinylpolypyrrolidone (PVPP) or activated charcoal was added to the beans before grinding (*8*, *9*).

Several commercially available kits provided suitable yields of high-molecular-weight genomic DNA, although not always amplifiable (see **Table 1** for an overview of the results). Because of the rapidity of the procedure, after the preliminary assays, the ClonTech NucleoSpin Plant Kit protocol was routinely used following the manufacturer's instructions with the following modification: after the initial incubation step at 65 °C, 5 μL of 1 U/μL RNase A was added to each tube, and the samples were incubated for 15 min at room temperature (rt).

Where a higher yield was required, a CTAB extraction was performed as follows: 1 g of ground beans was incubated at 65 °C in a sterile Falcon 15 mL tube with 3 mL of freshly stirred CTAB extraction buffer (1.5 M NaCl, 0.12 M sorbitol, 1% PVPP, 0.08% CTAB, 1% *N*-lauroylsarcosine, 5 mM EDTA, pH 8.0) and then allowed to cool to rt; after addition of 1 mL of 5 M potassium acetate, the tubes were incubated overnight at −20 °C, thawed under ice, and spun at 4000*g* for 15 min. A 1 mL aliquot of supernatant was transferred to a sterile 2 mL microcentrifuge tube before addition of 5 μL of 1 U/μL RNase A. Samples were incubated at 37 °C for 15 min and extracted with 600 μL of chloroform/isoamylic alcohol (24:1), followed by centrifugation at 4300*g* for 15 min. The aqueous phase was transferred to a 1.5 mL microcentrifuge tube, and 800 μL of 100% 2-propanol was added. Samples were incubated at room temperature for 30 min and spun at 4300*g* for 15 min. Supernatant was discarded and the pellet washed once with 70% ethanol and resuspended in 50 μL of TE buffer (10 mM Tris−HCl, 1 mM EDTA, pH 8). At this stage, samples showing residual traces of polysaccharides were subjected to an additional centrifugation at 13200*g* for 10 min on a benchtop centrifuge, and the supernatant was transferred to a fresh tube. All samples were stored at −20 °C.

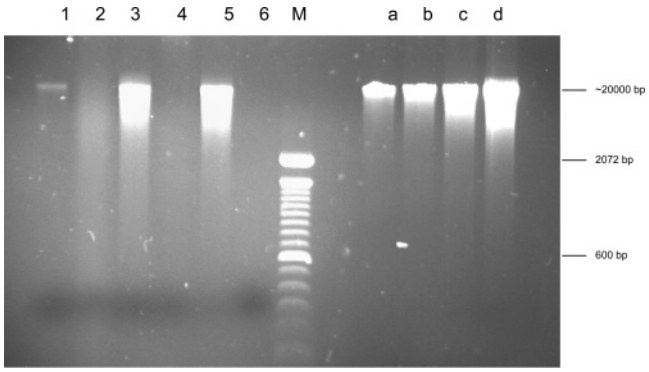

Figure 1. Extraction of DNA from green and roasted coffee beans. The final eluate/resuspended pellet was extracted with different protocols (10 μL each). Key: lanes 1 and 2, CTAB method on green and roasted beans; lanes 3 and 4, Nucleon Phytopure Plant DNA Extraction Kit on green and roasted beans; lanes 5 and 6, ClonTech NucleoSpin Plant Kit on green and roasted beans; lane M, 100 bp marker; lanes a–d, 6.25, 12.5, 25, and 50 ng of calf thymus DNA, respectively.

Roasted Coffee. Beans were ground as previously described in the presence of either 4% PVPP or activated charcoal. Several kits for DNA extraction were tested, most employing DNA-binding columns, the exception being Promega Wizard Magnetic DNA Purification Kits, which employ a magnetic bead capture system. The CTAB extraction protocol described above was also tested.

To determine DNA concentration and quality, 10 μL of the final eluate/resuspended pellets was electrophoresed on an 0.8% agarose gel containing 0.3 mg/mL ethidium bromide (EtBr). DNA from roasted beans was visible as a faint smear of ca. 200 bp; a blue smear appeared in the 20−2 kb range when the CTAB protocol was used (see **Figure 1**).

Sterical interaction of degraded DNA with EtBr and other dyes, such as Hoechst 33258, may be suboptimal and result in lesser fluorescence, affecting both visual and spectrophotometric quantitation. Values thus obtained, although approximate, were consistent and were used as references to compare the performance of different extraction methods and to assess the quantity of template in PCR reactions.

After the initial evaluation phase, DNA extraction for analytical purposes was routinely performed on 0.3 g of material with ClonTech NucleoSpin columns, following the manufacturer's protocol. DNA was eluted in 50 μL of the kit's CE buffer. This method took less than 3 h of hands-on time, and yielded ~0.1 μg of total DNA.

For applications requiring DNA of higher purity, such as amplification of low-copy-number sequences, extractions were performed from 1 g of starting material with the Wizard magnetic food kit, eluting the magnetic particles in 400 μL of

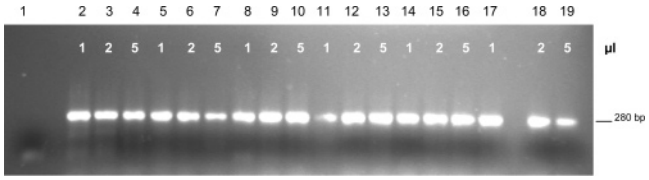

Figure 2. PCR on DNA extracted from roasted coffee with the Promega Wizard Magnetic DNA Purification Kit for Food followed by purification with Bio101 GeneClean II. Key: lane 1, negative control; lanes 2–4, Arabica roasted at 8.4 *L*; lanes 5–7, Arabica roasted at 10.3 *L*; lanes 8–10, Arabica roasted at 18.8 *L*; lanes 11–13, Robusta roasted at 10.2 *L*; lanes 14–16, Robusta roasted at 11 *L*; lanes 17–19, Robusta roasted at 19.1 *L*.

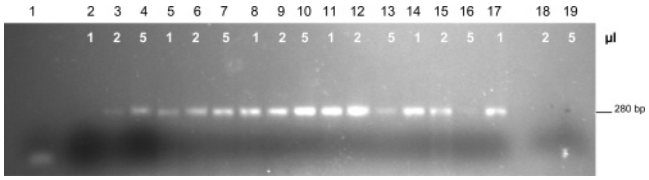

Figure 3. PCR on DNA extracted from roasted coffee with ClonTech Nucleospin columns. Key: lane 1, negative control; lanes 2–4, Arabica roasted at 8.4 *L*; lanes 5–7, Arabica roasted at 10.3 *L*; lanes 8–10, Arabica roasted at 18.8 *L*; lanes 11–13, Robusta roasted at 10.2 *L*; lanes 14–16, Robusta roasted at 11 *L*; lanes 17–19, Robusta roasted at 19.1 *L*.

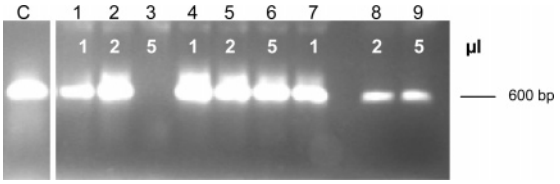

Figure 4. Inhibition test on a control PCR (ClonTech Nucleospin columns). Key: lane C, positive control (600 bp PCR product from a cloned insert); lanes 1–3, Robusta roasted at 10.2 *L*; lanes 4–6, Robusta roasted at 11 *L*; lanes 7–9, Robusta roasted at 19.1 *L*.

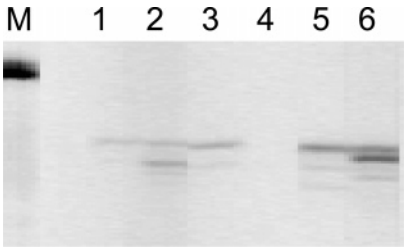

Figure 5. Microsatellite amplification from roasted coffee. Key: lane M, 106 bp PCR product (molecular weight marker); lane 1, Robusta roasted at 10.2 *L*; lane 2, Robusta roasted at 11 *L*; lane 3, Robusta roasted at 19.1 *L*; lane 4, Arabica roasted at 8.4 *L*; lane 5, Arabica roasted at 10.3 *L*; lane 6, Arabica roasted at 18.8 *L*.

H_2O in the final step. The eluate was further purified with Bio101 GeneClean II following the manufacturer's protocol, and the DNA was eluted in 20 μL of H_2O. Sterile PCR-grade water (Sigma) was used for all applications.

To assess PCR performance, we amplified a 210 bp fragment from a coffee chloroplast sequence for the *trnL* (UAA)−*trnF* (GAA) intergenic spacer, obtained from the NCBI sequence database (www.ncbi.nlm.nih.gov, accession no. U93387). Higher plants have a large number of chloroplasts per cell, and therefore, multiple copies of the chloroplast genome are present for each nuclear genome; hence, although DNA is degraded during roasting, enough target copies may remain intact for the PCR to succeed.

Volumes of 1, 2, and 5 μL of the final eluate (corresponding to ~5, 10, and 25 ng of DNA, respectively) were amplified in 25 μL of 10 mM Tris−HCl, 1.5 mM $MgCl_2$, 50 mM KCl, pH 8.3 at 20 °C, 2.0 μM dNTPs, 2.5% DMSO, 0.025 U/μL Taq polymerase (Roche), and a 1.0 μM concentration of each primer. The forward primer was 5′-AAT CGA TCT GGA CGG AAA AG-3′ and the reverse primer 5′-AAA GAA GGA AAG GGG ATT ACA A-3′. The PCR program was 94 °C for 3 min followed by 30 cycles of 94 °C for 30 s−55 °C for 30 s−72 °C for 1 min and by 72 °C for 5 min. The same conditions, except for the annealing temperature of 45 °C, were used to amplify a 280 bp fragment from a conserved domain of the reverse transcriptase, representative of a high-copy-number genomic sequence (*10*). The forward primer was 5′-GGA ATT CGA YGT NAA RAC NGC NTT YYT-3′ and the reverse primer 5′-GGG ATC CAY RTC RTC NAC RTA NAR NA-3′.

In our experience, PCR performance of DNA extracted with the Wizard magnetic food kit followed by purification with GeneClean II showed little or no variation with regard to the quantity of template (see **Figure 2**). PCR amplification of DNA extracted with ClonTech NucleoSpin columns, instead, depended on the quantity of the template to a greater extent. Higher *L* values and greater quantity of template resulted in stronger bands: this was interpreted as a consequence of DNA degradation at high temperature. Amplification from Robusta samples was more efficient for smaller quantities of template and failed occasionally with larger volumes of eluate; this behavior was more evident in mildly roasted samples (**Figure 3**). To assess whether poor PCR performance was due to the residual presence of Taq inhibitors, we added 1, 2, and 5 μL of eluate to a control reaction, using M13 forward and reverse universal primers to amplify a 600 bp insert cloned in pUC19. PCR inhibition alone did not explain the behavior of the 10.2 Robusta sample, and DNA degradation may have occurred as well (**Figure 4**).

Although chloroplast DNA is present in high copy number in the total DNA extracted from beans, it is highly conserved and unlikely to show intraspecific polymorphism within *C. arabica*. Hypervariable sequences, such as microsatellites, would be better suited for variety discrimination purposes; however, single-locus amplification strategies find a limitation in the quantity of available target DNA. To assess PCR performance of low-copy-number sequences, a short coffee genomic DNA fragment (~94 bp) containing a microsatellite motif was also amplified using 5 μL of eluate in 25 μL of 10 mM Tris−HCl, 1.5 mM $MgCl_2$, 50 mM KCl, pH 8.3 at 20 °C, 2.0 μM dNTPs, 2.5% DMSO, 0.025 U/μL Taq Gold polymerase (Roche), 0.1 μM tailed forward primer, 0.5 μM labeled tail, and 0.5 μM reverse primer. The primer set used was CMA068, obtained from the site www.coffeedna.net (sequences are available upon subscription). The PCR program consisted of 45 cycles of 94 °C for 30 s−55 °C for 30 s−72 °C for 1 min and a 72 °C for 5 min final elongation step. The primers used were designed for use in conjunction with a labeled universal tail that allowed visualization of the PCR products on a fluorescence or infrared genotyper, but did not otherwise affect amplification. Only DNA extracted with the Wizard magnetic food kit followed by purification with GeneClean II was used for the experiment. PCR products were electrophoresed in a 6% polyacrylamide gel on an LI-COR infrared automated genotyper (**Figure 5**); PCR from one highly roasted Arabica sample failed, possibly because of extreme template degradation.

Instant Coffee. The material did not require previous grinding; 1 g of instant coffee was weighed into a sterile tube and DNA extracted using the Wizard Magnetic DNA Purifica-

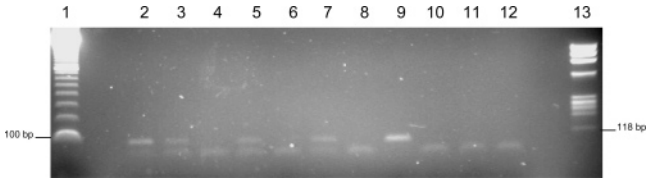

Figure 6. PCR on DNA extracted from instant coffee (PTC-100 thermal cycler). Key: lane 1, 100 bp marker; lanes 2–6, five different samples of instant coffee (a–e); lane 7, Kenco instant coffee; lane 8, negative control (PCR water); lane 9, positive control (genomic DNA from green coffee beans); lane 10, negative control (extraction elution buffer); lane 11, negative control (purification elution buffer); lane 12, aspecific amplification control (genomic DNA from oilseed rape); lane 12, ΦX174/Hae III marker.

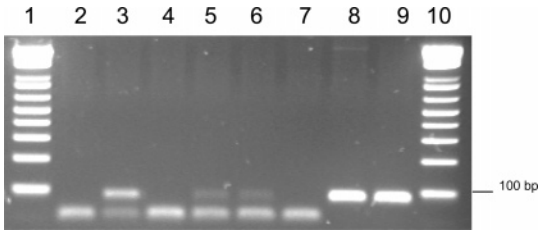

Figure 7. PCR on DNA extracted from instant coffee (Roche LightCycler). Key: lanes 1 and 10, 100 bp marker; lane 2, negative control (PCR water); lanes 3–6, four different samples of instant coffee (a, c–e); lane 7, Kenco instant coffee; lanes 8 and 9, positive control (*Coffea* genomic DNA from the leaf).

tion System for Food followed by purification with GeneClean II, as described for roasted coffee. Although 10 μL of DNA was not visible on a 2% agarose gel, measurements on a fluorimeter gave readings between 4 and 11 ng/μL. These readings are to be considered as indicative, for the reasons discussed. A PCR product of expected size was obtained from 5 μL of eluate, amplified on a PTC-100 thermal cycler in 25 μL of 10 mM Tris−HCl, 1.5 mM $MgCl_2$, 50 mM KCl, pH 8.3 at 20 °C, 2.0 μM dNTPs, 2.5% DMSO, 0.025 U/μL Taq Gold polymerase (Roche), and a 0.5 μM concentration of each primer. Primers were for an 86 bp fragment from the same *trnL*−*trnF* intergenic chloroplastic spacer described earlier. The forward primer was 5′-TTA TCC TAT CCC CCT TTC GTT A-3′ and the reverse primer 5′-GGG CTT TTC CGT CCA GAT-3′; the PCR program was 40 cycles of 95 °C for 30 s−55 °C for 30 s−72 °C for 1 min (**Figure 6**).

Real-time PCR was also carried out on a Roche LightCycler in 20 μL of 1× BioGene PCR master mix−3 mM $MgCl_2$ (cat. no. PCRM012), a 0.25 μM concentration of each primer, and 5% SYBR Green I solution (1‰ in dimethyl sulfoxide, DMSO); the PCR program was 40 cycles of 95 °C for 5 s−55 °C for 5 s−72 °C for 5 s. Electrophoresis of PCR products is shown in **Figure 7**. Results are comparable, although sample e was amplified on the LightCycler but not on the PTC-100; the opposite occurred for Kenco instant coffee. A description of the samples was not available to help us relate PCR performance with the type of processing to which the material tested was subjected during its formulation.

Conclusion. It is demonstrated that enough DNA survives during roasting and freeze-drying processes to enable successful extraction and subsequent amplification, provided that a suitable protocol is applied. In our experience, procedures requiring a final precipitation step allowed for a higher yield of DNA from both green and roasted beans; however, DNA thus obtained was not always amplifiable, presumably because of inhibitors coprecipitating with the DNA. Resin column methods provided DNA of a higher grade, although the yield was generally lower. Both activated charcoal and PVPP were shown to be efficient at preventing the binding of inhibitors to DNA during the grinding step; although the yield was lower, PCR performance improved.

Among the kits tested on roasted beans, the ClonTech NucleoSpin Plant Kit yielded readily amplifiable DNA from most samples; the Promega Wizard Magnetic DNA Purification Kit for Food required further purification with a DNA-binding resin to allow PCR amplification. However, this additional step resulted in a better PCR performance than was obtained using the NucleoSpin procedure. The Nucleon Phytopure plant DNA extraction kit and the CTAB protocol yielded about 250 ng of DNA, but none of the samples could be amplified, possibly due to the high amount of sugars contaminating the final extract. In our experience, the other kits tested did not yield detectable amounts of DNA from processed beans.

Suitability for amplification was shown to depend essentially on the nature of the starting material. Both degradation of DNA and likely the presence of PCR-inhibiting contaminants were observed, although it was possible to obtain successful amplification from all roasted samples simply by adjusting the quantity of template. It is however recommended that DNA extracted from roasted beans is thoroughly purified before being used for PCR, especially in applications that require a large quantity of template, such as low-copy-number sequences.

Successful amplification of DNA extracted from instant coffee was obtained for some of the tested samples, provided that short, multiple-copy sequences are chosen as a template. One sample out of six never yielded amplifiable DNA.

Given the low amount of DNA extracted from processed coffee, contaminant DNA is likely to outcompete the proper template in the PCR reaction, especially when degenerated or "universal" primers are employed. It is strongly recommended that special care is taken during DNA extraction and PCR setup and that negative controls are always included to test for contamination. Several precautions such as the use of filter pipet tips, a "closed cap" reaction like that employed on the LightCycler platform, or the uracyl *N*-glycosilase (UNG) method (*11*) can be implemented to prevent the possibility of contamination and carryover from earlier PCR experiments.

These extraction methods will enable researchers to assess the authenticity of coffee varieties once suitable markers are found, allowing determination of the percentage of Robusta in blends, or reveal frauds such as mentioned earlier. The same considerations and procedures may apply, with due modifications, to other foods and products that undergo a roasting process, such as, although not limited to, peanuts, cocoa beans, tobacco, and black tea.

ACKNOWLEDGMENT

Dr. Simon Penson, Kraft Foods, is thanked for supplying the coffee material.

LITERATURE CITED

(1) Meadows, M. California man convicted of brewing a coffee scam. *FDA Consum.* **2001**, *35*, 6.

(2) Prodolliet, J.; Bruelhart, M.; Blanc, M. B.; Leloup, V.; Cherix, G.; Donnelly, C. M.; Viani, R. Adulteration of soluble coffee with coffee husks and parchments. *J. AOAC Int.* **1995**, *78*, 761−7.

(3) Fontes A. S.; Bento, A. C.; Miranda, L. C. M.; Baesso, M. L. Thermal lens evaluation of the presence of adulterants in brewed coffee. *Anal. Sci.* **2001**, *17*, s526−9.

(4) Valdenebro, M. S.; León-Camacho, M.; Pablos, F.; González, A. G.; Martín, M. J. Determination of the arabica/robusta composition of roasted coffee according to their sterolic content. *Analyst* **1999**, *124*, 999−1002.
(5) Anthony, F.; Combes, M. C.; Astorga, C.; Bertrand, B.; Graziosi, G.; Lashermes, P. The origin of cultivated Coffea arabica L. varieties revealed by AFLP and SSR markers. *Theor. Appl. Genet.* **2001**, *104*, 894−900.
(6) Wilson, I. G. Inhibition and facilitation of nucleic acid amplification. *Appl. Environ. Microbiol.* **1997**, *63*, 3741−51.
(7) Redgwell, R. J.; Trovato, V.; Curti, D.; Fischer, M. Effect of roasting on degradation and structural features of polysaccharides in Arabica coffee beans. *Carbohydr. Res.* **2002**, *337*, 421−31
(8) Brunner, I.; Brodbeck, S.; Buchler, U.; Sperisen, C. Molecular identification of fine roots of trees from the Alps: reliable and fast DNA extraction and PCR-RFLP analyses of plastid DNA. *Mol. Ecol.* **2001**, *10*, 2079−87.
(9) Bi, I. V.; Harvengt, L.; Chandeler, A.; Mergeai, G.; du Jardin, P. Improved RAPD amplification of recalcitrant plant DNA by the use of activated charcoal during DNA extraction. *Plant Breed.* **1996**, *115*, 205−206.
(10) Flavell, A. J.; Smith, D. B.; Kumar, A. Extreme heterogeneity of Ty1-copia group retrotransposons in plants. *Mol. Gen. Genet.* **1992**, *231*, 233−242.
(11) Longo, M. C.; Berninger, M. S.; Hartley, J. L. Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. *Gene* **1990**, *93* (1), 125−8.

Received for review April 6, 2005. Revised manuscript received July 22, 2005. Accepted August 2, 2005. This research was partially funded through a European Social Fund grant (SE 200115359003) with the assistance of the ENAIP Italian institute for professional instruction.

JF050776P