Developing core collections to optimize the management and the exploitation of diversity of the coffee *Coffea canephora*

**Thierry Leroy · Fabien De Bellis · Hyacinthe Legnate ·
Pascal Musoli · Adrien Kalonji · Rey Gastón Loor Solórzano ·
Philippe Cubry**

Abstract The management of diversity for conservation and breeding is of great importance for all plant species and is particularly true in perennial species, such as the coffee *Coffea canephora*. This species exhibits a large genetic and phenotypic diversity with six different diversity groups. Large field collections are available in the Ivory Coast, Uganda and other Asian, American and African countries but are very expensive and time consuming to establish and maintain in large areas. We propose to improve coffee germplasm management through the construction of genetic core collections derived from a set of 565 accessions that are characterized with 13 microsatellite markers. Core collections of 12, 24 and 48 accessions were defined using two methods aimed to maximize the allelic diversity (Maximization strategy) or genetic distance (Maximum-Length Sub-Tree method). A composite core collection of 77 accessions is proposed for both objectives of an optimal management of diversity and breeding. This core collection presents a gene diversity value of 0.8 and exhibits the totality of the major alleles (i.e., 184) that are present in the initial set. The seven proposed core collections constitute a valuable tool for diversity management and a foundation for breeding programs. The use of these collections for collection management in research centers and breeding perspectives for coffee improvement are discussed.

T. Leroy (✉) · F. De Bellis · P. Cubry
CIRAD-UMR AGAP, 34398 Montpellier, France
e-mail: thierry.leroy@cirad.fr

F. De Bellis
e-mail: fabien.de_bellis@cirad.fr

H. Legnate
CNRA, BP 808, Divo, Côte d'Ivoire

P. Musoli
COREC, PO Box 185, Mukono, Uganda

A. Kalonji
University of Kinshasa, BP 866, Kinshasa, RDC

R. G. Loor Solórzano
INIAP, Estación Experimental Pichilingue, Los Rios, Ecuador

P. Cubry
INRA, UR 629 Ecologie des Forêts Méditerranéennes, URFM, Domaine Saint Paul, Site Agroparc, 84914 Avignon Cedex 9, France

Introduction

The genus *Coffea* (Rubiaceae) is mainly endemic to the tropical forests of Africa and Madagascar. The genus consists of approximately 124 species (Davis et al. 2011), two of which are of commercial importance: *Coffea arabica* L., which produces Arabica coffee and *Coffea canephora* Pierre ex A. Froehner, which produces Robusta coffee. Robusta accounts for one-third of the world coffee trade in volume and approximately US$ 5 billion (ICO 2013).

Coffea canephora originated from lowland tropical African forests that stretch from Guinea to Uganda. Since 1980, the genetic diversity of *C. canephora* has been described in numerous studies (Berthaud 1986; Montagnon et al. 1992; Dussert et al. 1999; Montagnon 2000; Gomez et al. 2009; Musoli et al. 2009; Cubry et al. 2013a). These studies revealed the presence of two main diversity groups:

the Congolese group and the Guinean group. The Congolese group is subdivided into five subgroups: SG1, SG2, B, C and UW. The SG1 genotypes originate from Coastal Atlantic Africa. The SG2 accessions originate from the Congo Basin, mainly from the Democratic Republic of the Congo (DRC). The B genotypes come from the Central African Republic (CAR) and the DRC. The C genotypes come from the CAR and Cameroon (Cubry et al. 2013a). The UW genotypes are found in the forests of eastern Uganda (Musoli et al. 2009). Only a small portion of this large diversity is used in past and current breeding programs, with the exception of the recurrent selection program in the Ivory Coast (IC) in the 1990s (Leroy et al. 1993; Montagnon et al. 1998; Montagnon 2000).

Davis et al. (2011) highlighted the serious threat of a loss of genetic diversity of *Coffea* species mainly due to a decrease in the quality and quantity of the species' natural habitat. Despite its economic importance, no effective in situ conservation strategy exists. *Coffea canephora* is an allogamous perennial species, and consequently, its genotypes must be reproduced by cloning (cuttings) and conserved *ex situ* in the form of living collections.

The *ex situ* gene banks of *C. canephora* were established from biological material that was gathered during collecting missions (Montagnon et al. 2012) or created while implementing breeding programs. The richest field collection in terms of the number of accessions and diversity is located at the Centre National de la Recherche Agronomique (CNRA) in the Ivory Coast, with more than 1,000 genotypes of *C. canephora* that were collected throughout Africa (Berthaud and Charrier 1988, Anthony 1992). Another large field collection is present at the Coffee Research Centre (COREC) in Uganda, with a mix of local genotypes and genotypes that were introduced from the former Belgian Congo (Thomas 1947; Musoli et al. 2009). Other collections of importance can be found in Brazil (genotypes from the SG1 group, as confirmed by Cubry (2008b)), Ecuador (introduced genotypes from the SG1 and SG2 groups, Leroy, pers. com.), the DRC, Guinea, Madagascar, Cameroon and the CAR.

For most crop species, the genetic diversity that is conserved in *ex situ* gene banks is underutilized and frequently redundant. A small proportion of this diversity is actually used in breeding programs (Van Hintum et al. 2000). The core collection concept was introduced in the 1980s to define subsets from the whole collection that capture, with minimum redundancy, most of the genetic diversity of the target species (Frankel and Brown 1984; Brown 1989). The first core collections in plants were based on passport and phenotypic quantitative data (Upadhyaya et al. 2001). DNA markers, such as microsatellites (SSR), have been further used to construct core collections (Volk et al. 2005; Ronfort et al. 2006; Balfourier et al. 2007) either alone or in association with phenotypic traits (Volk et al. 2005; Belaj et al. 2012).

Different approaches have been compared for the construction of core collections in annual (Franco et al. 2006) and perennial species (Escribano et al. 2008), leading to the conclusion that the maximization (M) method, which maximizes the number of alleles, is highly suitable for constructing core collections (Schoen and Brown 1993). Reports on the construction and quality of core collections for their effective use in diversity management have recently been published (Pessoa-Filho et al. 2010; El Bakkali et al. 2013; Odong et al. 2013).

For most perennial crops, numerous teams are working on different germplasm collections throughout the world. A common subset of genotypes representing the global diversity of the species is therefore of high interest for sharing data and research results. Such core collections have been constructed for grapes (Le Cunff et al. 2008; Laucou et al. 2011) and olives (Belaj et al. 2012; Haouane et al. 2011).

The first coffee core collection was proposed by Hamon et al. (1995) for the whole genus using the principal components score strategy with quantitative data, including isozymes. Until now, no other core collection has been proposed for the genus *Coffea* or for the species *C. canephora* in particular.

Considering the great genetic diversity of *C. canephora*, the geographic structure of its diversity, the high number of *ex situ* collections worldwide and the cost of maintenance of field collections, it is important to improve the management strategy of these collections. Among other applications, small core collections are particularly pertinent and could be used for gene discovery, Genotyping by Sequencing (GbS) or massive resequencing. The core collections could also be the first step in genetic association studies (Le Cunff et al. 2008; El Bakkali et al. 2013), complementing Linkage Disequilibrium (LD) studies (Barnaud et al. 2006; Cubry et al. 2013b).

The purpose of the present study was (1) to propose different core collections for the *C. canephora* species using genetic diversity analyses of genotypes that were collected and maintained in field collections in the IC, Uganda, the DRC and French Guyana and (2) to determine the consequences of medium- and long-term strategies for the conservation and improvement of this species.

Materials and methods

Plant material

An initial set of 565 genotypes was used for our study (Table 1). These genotypes were planted under different field collections in the IC (Cubry et al. 2013a), Uganda (Musoli et al. 2009), the DRC (Luki origin) and French Guyana. Some additional genotypes were collected from

Table 1 Description of the different collections that were studied: collection site and provider, number of genotypes for each diversity group and country of origin

Collection	No. of genotypes	Provider	Diversity groups	No. of accessions per Diversity Group (No of populations concerned)	Country of origin	Reference
Initial set						
The IC	299	CNRA, Ivory Coast (IC)	G	161 (8)	The IC, Guinea	Cubry et al. 2013a
			B	29 (1)	The CAR	
			C	83 (1)	The CAR	
			SG1	9 (1)	Gabon or Congo via Benin	
			SG2	16 (2)	The DRC	
Uganda	223	NARO-COREC, Uganda	SG2	56 (2)	Introduced from the DRC	Musoli et al. 2009
			UW	167 (2)	Uganda	
The DRC	30	INERA, Democratic Republic of Congo (DRC)	SG1	30 (2)	The DRC	Present study
Guyana	11	CIRAD, French Guyana	G	4 (1)	The IC	Present study
			SG1	3 (1)	Selected in plantation in the IC, origin unknown	
			SG2	3 (1)	Selected in plantation in the IC, origin unknown	
			Intergroup Hybrid[a]	1 (1)	Selected in plantation in the IC, origin unknown	
The IRD	2	IRD, France	G	2 (1)	Guinea	Present study
	Total 565					
Additional material						
Ecuador	87	INIAP, Ecuador	SG1	20 (1)	Unknown	Present study
			SG2	67 (1)	Unknown	
Uganda	27	NARO-COREC, Uganda	SG2	27 (1)	Introduced from the DRC	Musoli et al. 2009
Brazil	4	IAPAR, Brazil	SG2	4 (1)	Unknown	Cubry et al. 2013a
	Total 118					

[a] This accession is a hybrid between the existing diversity groups and shall not be considered as a diversity group *per se*

the collection of the Institut de Recherche Pour le Développement (IRD) in Montpellier (France). For a complete description of the material, see Supplementary material 1.

In addition, 118 genotypes from breeding collections outside of Africa (Brazil and Ecuador) and from feral populations (Uganda) were included for comparison purposes (Table 1).

DNA preparation and genotyping

Genomic DNA extractions and PCR reactions were performed according to Cubry et al. (2008). The PCR products were analyzed by electrophoresis on a 6.5 %

polyacrylamide gel using a LI-COR® 4300 automated sequencer (LI-COR Biosciences, Nebraska, USA). Size calling was automatic and manually checked using the manufacturer's program SAGA[GT].

Data analysis

Validation and relevance of the set of markers

Previous studies on the field collections of the IC (Cubry et al. 2013a, b) and Uganda (Musoli et al. 2009) led us to select a set of 13 SSR markers (Table 2). These markers were shared between these studies, easy to analyze and pertinent

Table 2 Detailed information on the 13 SSR markers that were used, their motif, GenBank number, reference and primers that were used for their amplification

Locus	Forward sequence	Reverse sequence	Motif	GenBank accession number	Reference
Me_BE_CWI_DL011	ATACATAAGCAAGCACTGA	CAGAACAAATGAAATGGA	(GCT)4(CAT)8	AJ871890	Leroy et al. 2005
Me_BE_CWI_DL013	AGAGGGATGTCAGCATAA	ATTTGTGTTTGGTAGATGTG	(CA)6(CT)8	AJ871892	Leroy et al. 2005
Mg_M257	GCATTTTGTTGCACACTGTA	GACCATTACATTTCACACAC	(CA)9	AJ250257	Combes et al. 2000
Mg_M358	CATGCACTATTATGTTTGTGTTTT	TCTCGTCATATTTACAGGTAGGTT	(CA)11	AM231554	Poncet et al. 2007
Mg_M368	AACCACATCTCCATCCATAACC	TCCTACCTACTTGCCTGTGCT	(TG)3(TT)1(TG)13	AM408761	Poncet et al. 2007
Mg_M445	CCACAGCTTGAATGACCAGA	AATTGACCAAGTAATCACCGACT	(AC)10	AM231567	Poncet et al. 2007
Mg_M461	CGGCTGTGACTGATGTG	AATTGCTAAGGGTCGAGAA	(AC)9	AM231570	Poncet et al. 2007
Mg_M477	CGAGGGTGGGAAAAGGT	ACCACCTGATGTTCCATTTGT	(AC)16	AM231574	Poncet et al. 2007
Mg_M501	CACCACCATCTAAATGCACCT	CTGCACCAGCTAATTCAAGC	(TG)8	AM231576	Poncet et al. 2007
Mg_M753	TCGAGAAGTCTTGGGGTGTT	GGAGACGCAGGTGGTAGAAG	(CA)15	AJ308753	Rovelli et al. 2000
Mg_M779	GGGAGTGTTTTGTGTGTGCTT	TCCCCCATCTTTTTCTTTCC	(TG)17	AJ308779	Rovelli et al. 2000
Mg_M782	TCCACATACATTTCCCAGCA	AAAGGAAAATTGTTGGCTCTGA	(GT)15	AJ308782	Rovelli et al. 2000
Mg_M837	CGGTATGTTCCTCGTTCCTC	CTCGCTTTCACGCTCTCTCT	(TG)16(GA)11	AJ308837	Rovelli et al. 2000

for diversity studies. These markers were located in eight of the 11 different linkage groups of a *C. canephora* genetic map (Leroy et al. 2011). We verified the relevance and the good discrimination of accessions with this set of 13 markers using a correlation test between the *Rst* diversity statistics (Slatkin 1995) for 356 accessions using a large set of 101 makers (data from Cubry et al. 2013b) compared with our 13 markers using Arlequin 3.5.1.2 software (Excoffier et al. 2005). This set of 356 accessions was used for linkage disequilibrium studies (Cubry et al. 2013b) and divided into seven groups of diversity. We generated two *Rst* matrices, one from our set of 13 markers and another one from the set of 101 SSR markers. A Mantel correlation test using Gen-AlEx software (Peakall and Smouse 2006) was performed to compare both of the *Rst* matrices after 999 permutations.

Genetic diversity representation

We computed a dissimilarity matrix between individuals using a simple matching index for the entire dataset (DARwin. v. 5.0.137, Perrier and Jacquemoud-Collet 2006). The level of dissimilarity between individuals can be used to assess the uniqueness of the genotypes. We inferred diversity trees from this matrix using the Neighbor-Joining (NJ) method (Saitou and Nei 1987). Two hundred bootstrap iterations were calculated to test the robustness of the nodes.

Construction of the core collections

We first defined a compulsory subset of five genotypes considering their specific interests:

- The accession "126" from the Ivory Coast, which is a natural hybrid between Congolese and Guinean, is considered as best *C. canephora* genotype regarding its agronomic characteristics and the quality of the coffee produced (Montagnon et al. 1992).
- The accession "200" has been used for genome sequencing.
- The accessions "410", "466" and "A03" were the best progenitors of the intergroup hybrids between Congolese and Guinean from the Reciprocal Recurrent Selection program on *C. canephora* breeding in the Ivory Coast (Leroy et al. 1997).

We used this set as a kernel for the core collection construction. We removed the rare alleles with a frequency <0.05 % to limit their impact on further analyses. Then, we used two different sampling methods:

1. Using the M strategy (Schoen and Brown 1993) as implemented in the MStrat program (Gouesnard et al. 2001), we evaluated the efficiency of this approach and

Table 3 SSR diversity among the different defined core collections: the number of alleles, heterozygosity, gene diversity and repartition of the selected accessions through the diversity groups

Core set	Size	No. of common individuals MStrat/MLST	Number of alleles	% of total SSR diversity	Heterozygosity (Ho)	Gene diversity (GD)	B	C	G	Hybrids	SG1	SG2	UW
Available_565i	565	–	202	–	0.37	0.73	29	83	167	1	42	76	167
565i-RareAlleles	565	–	184	100	0.37	0.73	29	83	167	1	42	76	167
Core12_MStrat	12	7	126	68	0.51	0.81	0	2	3	1	3	2	1
Core12_MLST	12		109	59	0.43	0.76	1	1	2	1	3	4	0
Core24_MStrat	24	9	161	88	0.53	0.81	3	3	6	1	5	4	2
Core24_MLST	24		142	77	0.40	0.78	1	2	7	1	4	7	2
Core48_MStrat	48	19	184	100	0.44	0.79	3	8	11	1	8	9	8
Core48_MLST	48		160	87	0.40	0.80	4	7	12	1	9	10	5
Core77_composite	77	–	184	100	0.41	0.79	6	11	19	1	13	15	12

the optimal range size of the coffee core collection. For this purpose, we compared the number of alleles that were captured in core collections of increasing size using both the M strategy and a random sampling (200 replicates by core size and 15 iterations) and located the inflection point of the M-strategy redundancy curve. Using the same software settings, we successively built core collections from 12 accessions to the optimal size, selecting as kernels the accessions that were retained in the previous step. We named these core collections CoreXX_MStrat, with XX representing the number of accessions in the core.

2. The maximum length sub-tree method (MLST strategy) as implemented in the DARwin program was used to define core collections that were the same size as those that were generated with MStrat. This iterative method is based on resampling from an initial diversity tree that was constructed using the NJ method. From this tree, at each step, a maximum length sub-tree is constructed by the successive pruning of most of the redundant units. We named the resulting core collections CoreXX_MLST.

Genetic parameters

The number of alleles, the observed heterozygosity (H_o), the gene diversity (GD) and the number of accessions within each diversity group for the different core collections were calculated with PowerMarker v. 3.25 (Liu and Muse 2005).

Results

Validation of the marker set

The *Rst* matrices were compared for both sets of markers using the Mantel correlation test. The results of the significance of relationships between the matrices are presented in Fig. 1. With an R^2 value of 0.782 (R = 0.884, $p < 0.005$), we accept the strong relationship between both of the matrices, which validates our set of 13 markers for diversity analyses within our species. Thus, the proposed set of 13 markers is adequate for evaluating the global diversity within our species.

Genetic structure of *C. canephora* accessions

The first diversity tree was constructed with 565 accessions and 13 markers (Fig. 2). The high degree of stratification within the species is clearly illustrated by this tree. A total of 24 pairs of genotypes presented no significant dissimilarity.

The 118 additional genotypes from Ecuador and Brazil and from the feral populations of Uganda did not exhibit additional genetic diversity when included in the initial set (data not shown). Therefore, these genotypes were not used for the construction of the core collections as their diversity appeared to be redundant with the reference collections for the SG1 and SG2 groups.

Construction of the core collections

A total of 202 alleles were detected in the 565 genotypes, of which 184 alleles with a frequency > 0.05 % and that occurred at least two times were retained for the analysis. The efficiency of the M strategy versus the random sampling was visually checked by constructing random sampling (dotted line, Fig. 3) and M strategy (plane line, Fig. 3) redundancy curves. The optimal size for the core collection to avoid redundancy was between 24 and 48 accessions, leading us to construct core collections of up to 48 accessions.

The core collections were identified using MStrat software and the MLST stategy for nested core collections of

Correlation test between Rst values for sets of 101 markers vs 13 markers

R = 0.884
(p<0.005)
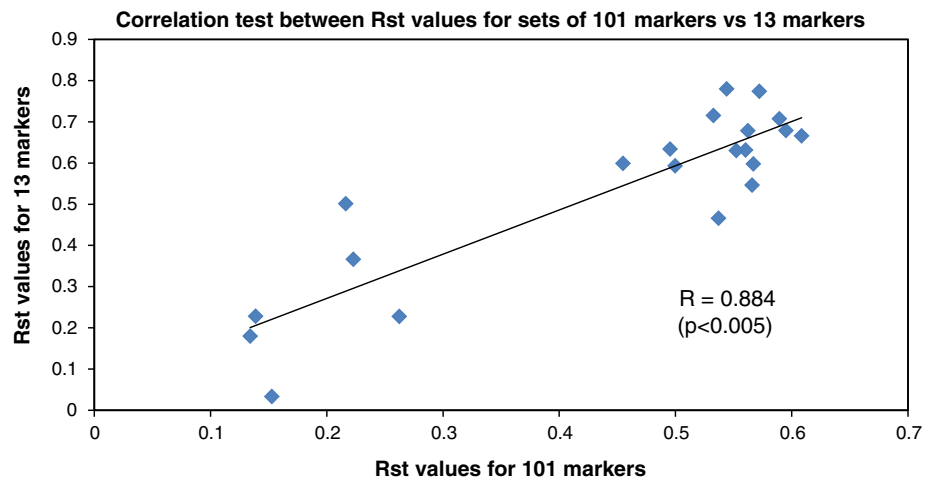
12, 24 and 48 accessions. The genotypes in Core12, Core24 and Core48 for both of the strategies are presented in (Tables 3) (M strategy) and 4 (MLST strategy). The number of alleles, the observed heterozygosity (H_o), the gene diversity (GD) and the number of accessions within each diversity group for the different core collections are presented in Table 5. The number of alleles was constantly higher in the core collections that were constructed with the M strategy than in those that were constructed with the MLST strategy for either size. In Core48_MStrat, all of the alleles were present, while only 87 % of the alleles were present using the MLST strategy. The H_o increased in Core12_MStrat (0.51) and Core24_MStrat (0.53) and decreased in Core48_MStrat (0.45). The H_o was always lower when using the MLST strategy, with the highest value of 0.43 in Core12_MLST. The GD increased in all the core collections using the MLST strategy, reaching a value of 0.80 in Core48_MLST, but decreased in Core48_MStrat, with a value of 0.79. The GD was slightly higher in Core48_MLST than in Core48_MStrat. The number of accessions per diversity group was different in both of the approaches, illustrating the differences in the strategies when constructing core collections. Regarding Core12, five out of the six diversity groups were retained by both of the methods; the B diversity group was missing in Core12_MStrat and the UW group was missing in Core12_MLST. For the Core48 collections, 19 accessions were common to both of the methods.

Definition of a composite core collection

For both of the Core48 collections, we constructed the diversity trees using DARwin software (Fig. 4). By comparison with the tree in Fig. 2 with 565 genotypes, the diversity of the species was well represented in both of the core collections (Fig. 4). The percentage of present alleles was slightly higher for the M strategy, but the
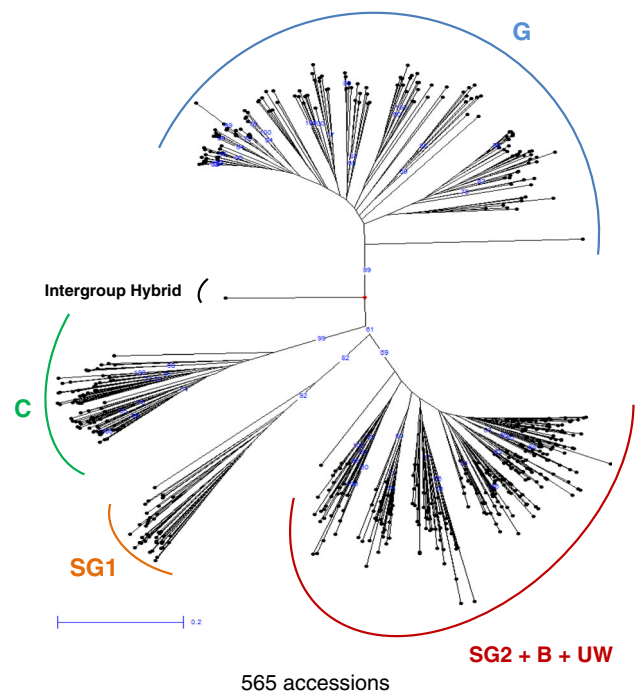

565 accessions

Fig. 2 Neighbor-joining diversity tree based on the dissimilarity matrices for the set of 565 accessions that were used for the construction of the core collections. The diversity groups are indicated on the tree

tree had longer branches for the MLST strategy because it was based on the pruning of the edges that presented the minimal length, thus maximizing the interleaf distances.

Because both of the strategies had different purposes and gave complementary results, we proposed to merge both of the Core48 collections, leading to a composite core collection of 77 accessions (Table 6). We observed values of 0.41 for the H_o and 0.79 for the GD, with the presence of all of the alleles from the original set of accessions.

Fig. 3 Redundancy curves that were obtained with MStrat software for the determination of the optimal size of our core collection. The plane line represents the curve using the Maximization strategy, and the dotted line represents the curve using random sampling. The box indicates the location around the inflection point, which determines the optimal size of the core collection
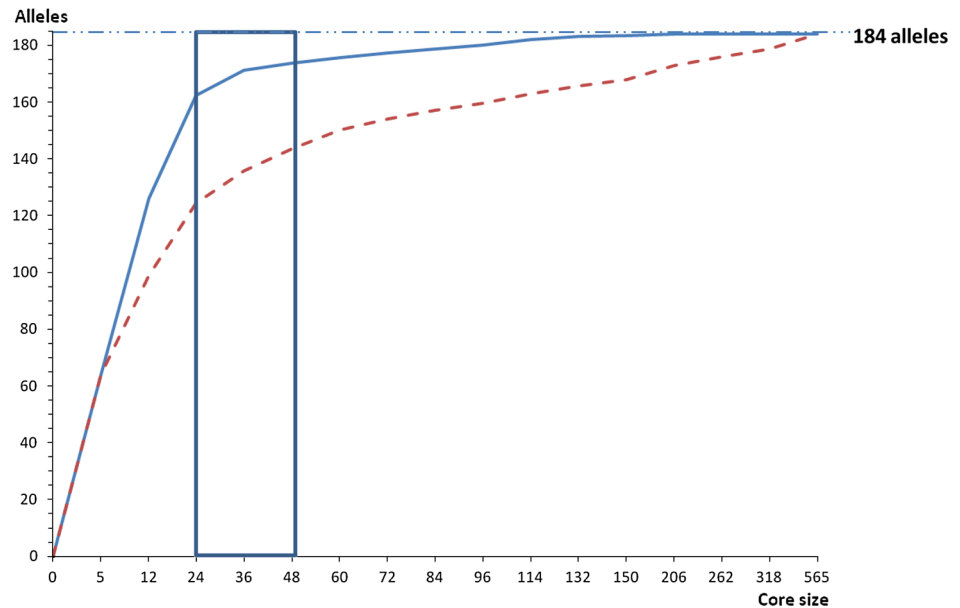
Discussion

General consideration for building nested core-collections for *Coffea canephora*

We constructed core collections for the *C. canephora* germplasm from the available genetic diversity within the species using a set of 13 SSR markers that were validated as relevant for our analysis. To address a wide variety of purposes and projects, we choose to construct collections of 12, 24 and 48 accessions (i.e., up to the optimal size as assessed with the M Strategy) using both the MStrat and MLST strategies. The construction of core collections in a highly structured species such as *C. canephora* is possible with a good representation of the diversity in terms of alleles using 48 genotypes.

For perennial tropical crops such as coffee, constructing these core collections is of great interest. Field collections cover large areas and are costly to manage for research centers worldwide. It will be easier to manage and characterize smaller core collections. These core collections could be associated with "reference" general collections, such as those from the Ivory Coast and Uganda, which cover a high variability within *C. canephora* and other coffee species. For coffee germplasm management, we should consider both germplasm conservation and breeding perspectives.

These core collections are useful for a variety of purposes: gene discovery, Genotyping by Sequencing and the massive sequencing of selected genotypes. The construction of these collections could also be used as the first step of association studies on the whole diversity or on a specific diversity group. These core collections can also be used for the detection of signature of adaptive selection and the detection of genomic regions that are submitted to selective pressure such as abiotic stress.

Different strategies for constructing core collections lead to different results

Both of the strategies for core collection construction were based on different finalities for germplasm management and valorization. With the M strategy, the number of alleles was maximized, meaning that the main objective of this strategy was to preserve the highest number of alleles, leading to a "conservation"-oriented strategy in core collection construction. The MLST strategy differs because the aim was to maintain most of the diverse combinations of alleles in specific genotypes. We could state that this strategy was more orientated toward "breeding" to maintain the most diverse combinations of alleles within selected genotypes. Comparing both of the Core48 collections, the difference in the sampling strategies is well illustrated by the higher conservation of diversity groups in Core48MStrat with higher bootstraps values.

For conservation purposes, the M strategy is the most adapted. Core12_MStrat will be a valuable and affordable tool for targeted resequencing. The Core24_MStrat collection could be a good equilibrium for a first core collection to be shared by most researchers because it gathers 88 % of the alleles with a gene diversity of 0.81 and a good representation of all of the diversity groups. Core48_MStrat is more complete and should be considered in a second step when the material exchange will be more comfortable for all researchers.

Table 4 List of the selected genotypes for the 12, 24 and 48 core collections using the MStrat strategy: name, population of origin, diversity group and country of collection

Genotype name	Population	Diversity group	Collection	Core
NA005	nana	C	The IC	12
NA087	nana	C	The IC	12
DJE6	guincultiv	G	The IC	12
02134	ira2	G	The IC	12
410_Guyane	IC_via_Guyana	G	Guyana	12
126_RCI	IC_via_Guyana	Intergroup Hybrid	Guyana	12
200_Guyane	IC_via_Guyana	SG2	Guyana	12
S_9/23-3	luki	SG1	The DRC	12
466_Guyane	IC_via_Guyana	SG1	Guyana	12
A03_Guyane	IC_via_Guyana	SG1	Guyana	12
3/20	erect	SG2	The IC	12
Mbale_Mahungu_T23	K_MM	UW	Uganda	12
02569	libengue	B	The IC	24
02527	libengue	B	The IC	24
02528	libengue	B	The IC	24
NA001	nana	C	The IC	24
02826	pine	G	The IC	24
150	guincultiv	G	The IC	24
02159	ira2	G	The IC	24
LAF_159/58-3	luki	SG1	The DRC	24
NIAOULI3	niaouli	SG1	The IC	24
024	ineac7	SG2	The IC	24
1/2	erect	SG2	The IC	24
Kyamuhoro_Cpt11_T8	I_11	UW	Uganda	24
NA012	nana	C	The IC	48
NA019	nana	C	The IC	48
NA023	nana	C	The IC	48
NA075	nana	C	The IC	48
NA101	nana	C	The IC	48
02363	pelezi	G	The IC	48
416	guincultiv	G	The IC	48
02044	ira1	G	The IC	48
02201	fourougbankoro	G	The IC	48
Maclaudi_S/BIL4	Maclaudi	G	The IRD	48
S_19/58-2	luki	SG1	The DRC	48
NIAOUL14	niaouli	SG1	The IC	48
042	ineac2	SG2	The IC	48
3/5	erect	SG2	Uganda	48
202/63	nganda	SG2	Uganda	48
257/18	nganda	SG2	Uganda	48
077_Guyane	IC_via_Guyana	SG2	Guyana	48
392_Guyane	IC_via_Guyana	SG2	Guyana	48
Kyamuhoro_Cpt11_T15	I_11	UW	Uganda	48
Ngogo_Kibuguta_A_T9	K_NK	UW	Uganda	48
Kabwegyemere_T6_S1	K_Kab	UW	Uganda	48
Ngogo_Kibuguta_A_T10_S2	K_NK	UW	Uganda	48
Ngogo_Kibuguta_T10_S3	K_NK	UW	Uganda	48
Ngogo_Kibuguta_B_T2_S3	K_NK	UW	Uganda	48

Compulsory accessions are shown in bold

Table 5 List of the selected genotypes for the 12, 24 and 48 core collections using the MLST strategy: name, population of origin, diversity group and country of collection

| Genotype name	Population	Diversity group	Collection	Core
02522	libengue	B	The IC	12
NA005	nana	C	The IC	12
Maclaudi_S/BIL4	maclaudi	G	The IRD	12
410_Guyane	IC_via_Guyana	G	Guyana	12
126_RCI	IC_via_Guyana	Intergroup Hybrid	Guyana	12
200_Guyane	IC_via_Guyana	SG2	Guyana	12
S_19/58-3	luki	SG1	The DRC	12
466_Guyane	IC_via_Guyana	SG1	Guyana	12
A03_Guyane	IC_via_Guyana	SG1	Guyana	12
3/20	erect	SG2	Uganda	12
14/19	erect	SG2	Uganda	12
202/63	nganda	SG2	Uganda	12
NA013	nana	C	The IC	24
02827	pine	G	The IC	24
150	guincultiv	G	The IC	24
418	guincultiv	G	The IC	24
02102	ira2	G	The IC	24
02962	mouniandougou	G	The IC	24
LK_35/2-3	luki	SG1	The DRC	24
066	ineac2	SG2	The IC	24
024	ineac7	SG2	The IC	24
203/74	nganda	SG2	Uganda	24
Mbale_Mahungu_T29	K_MM	UW	Uganda	24
Mbale_Mahungu_T9_S6	K_MM	UW	Uganda	24
02519	libengue	B	The IC	48
02569	libengue	B	The IC	48
02551	libengue	B	The IC	48
NA003	nana	C	The IC	48
NA001	nana	C	The IC	48
NA010	nana	C	The IC	48
NA019	nana	C	The IC	48
NA087	nana	C	The IC	48
02377	pelezi	G	The IC	48
02821	pine	G	The IC	48
416	guincultiv	G	The IC	48
GO2	guincultiv	G	The IC	48
02038	ira1	G	The IC	48
LAF_159/33-3	luki	SG1	The DRC	48
LK_35/2-1	luki	SG1	The DRC	48
S_9/23-1	luki	SG1	The DRC	48
S_9/23-3	luki	SG1	The DRC	48
NIAOUL14	niaouli	SG1	The IC	48
002	ineac7	SG2	The IC	48
3/66	erect	SG2	Uganda	48
247/62	nganda	SG2	Uganda	48
Kyamuhoro_Cpt11_T8	I_11	UW	Uganda	48
Ngogo_Kibuguta_B_T3_S2	K_NK	UW	Uganda	48
Kabwegyemere_T5_S1	K_Kab	UW	Uganda	48

Compulsory accessions are shown in bold
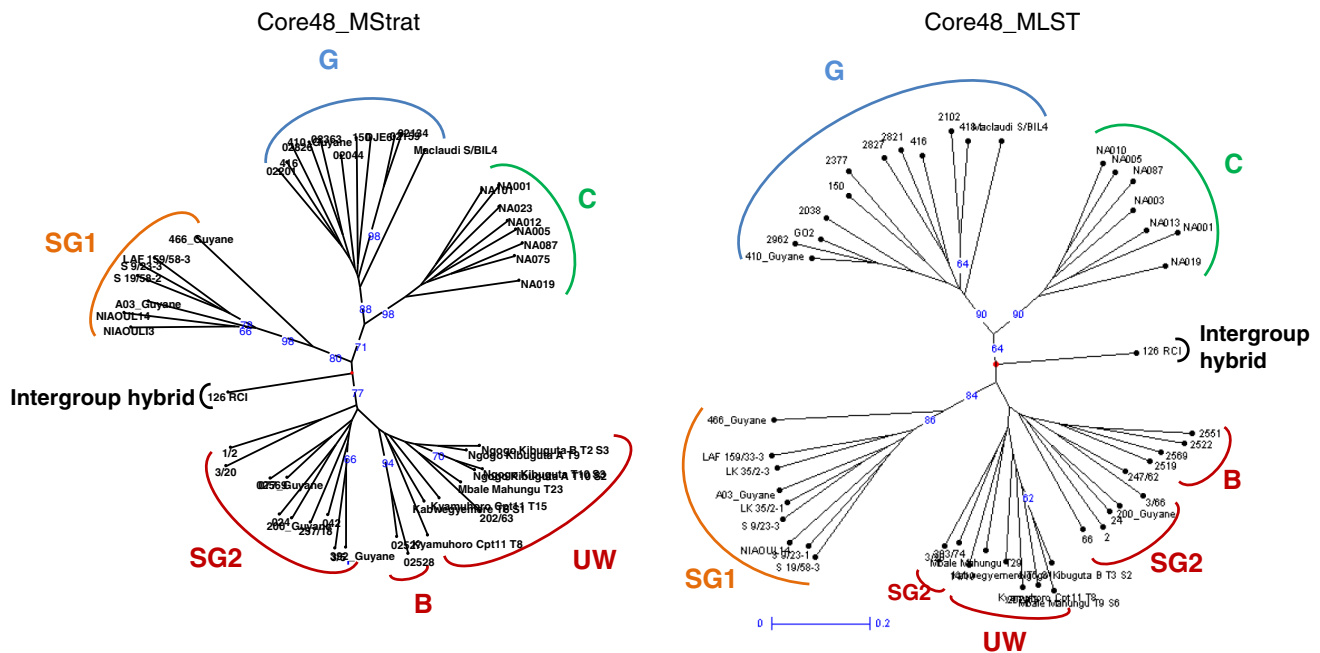
Fig. 4 Neighbor-joining diversity trees based on dissimilarity matrices for the Core48_MStrat and Core48_MLST accessions. Each accession is identified by its name on the tree, and the diversity groups are indicated

In a breeding strategy, MLST appears to be more adapted because this methodology maximizes the combination of allele diversity at the genome level. Breeders managing diversity in long-term breeding programs, such as those in the Ivory Coast and Uganda, could prefer to use and exchange accessions from the Core48_MLST collection, which gathers 87 % of the alleles for a gene diversity of 0.80 and displays a complete representation of the diversity groups.

Combining the best of two worlds, proposing a composite core collection that is built with different approaches and objectives

Regarding the specificity and different objectives of the two methods that were used, we propose a global core collection of 77 accessions that should be adapted for both germplasm conservation and breeding perspectives.

The composite Core77 collection gathered genotypes representing both the maximum number of alleles and the most diverse allelic combinations. This "optimal" core collection could be considered as the basic *C. canephora* collection that allows for all possible uses, from the simple choice of genotypes for specific aims (tolerance to a specific pest or specific characteristics of the coffee bean) to long-term breeding strategies using a large diversity panel to improve populations and select specifically adapted high-yielding genotypes that are tolerant to biotic and abiotic stress and that produce a high-quality coffee.

Limitations of phenotypic information

Our core collections were constructed considering the marker diversity without reference to the phenotypic diversity as reliable phenotypic data are not available for all of the accessions. We only have partial data from the field collections, mainly in the Ivory Coast and Uganda, under different edaphoclimatic conditions. These data provide some indications of the yield components, quality traits and tolerance to biotic and abiotic stresses. The choice of the five mandatory genotypes was based on the phenotypic or genotypic values of the genotypes. For example, the accession "126" is presents the best organoleptic quality associated with a high agronomic value. This accession has been widely distributed in a high number of African countries, including Togo, Guinea and Cameroon, as a reference for the yield and quality. The accessions "410", "A03" and "466" have specific genotypic values by crossing with genotypes from other diversity groups in the Reciprocal Recurrent Selection program in the Ivory Coast. These accessions are progenitors of the best hybrids in the Ivory Coast and have been distributed all over the world through improved seeds. The use of our proposed core collections will facilitate the construction of a multilocal phenotypic evaluation of the proposed accessions under the different edaphoclimatic conditions from the Ivory Coast to Brazil and from Uganda to Ecuador with different levels of biotic and abiotic stresses.

Table 6 List of the 77 accessions for the composite core collection: name, population of origin, diversity group, country of collection and core origin

Genotype name	Population	Diversity group	Collection	Core
02519	libengue	B	The IC	MLST
02522	libengue	B	The IC	MLST
02551	libengue	B	The IC	MLST
02527	libengue	B	The IC	MStrat
02528	libengue	B	The IC	MStrat
02569	libengue	B	The IC	MStrat/MLST
NA003	nana	C	The IC	MLST
NA010	nana	C	The IC	MLST
NA013	nana	C	The IC	MLST
NA012	nana	C	The IC	MStrat
NA023	nana	C	The IC	MStrat
NA075	nana	C	The IC	MStrat
NA101	nana	C	The IC	MStrat
NA001	nana	C	The IC	MStrat/MLST
NA005	nana	C	The IC	MStrat/MLST
NA019	nana	C	The IC	MStrat/MLST
NA087	nana	C	The IC	MStrat/MLST
418	guincultiv	G	The IC	MLST
02038	ira1	G	The IC	MLST
02102	ira2	G	The IC	MLST
02377	pelezi	G	The IC	MLST
02821	pine	G	The IC	MLST
02827	pine	G	The IC	MLST
02962	mouniandougou	G	The IC	MLST
GO2	guincultiv	G	The IC	MLST
02044	ira1	G	The IC	MStrat
02134	ira2	G	The IC	MStrat
02159	ira2	G	The IC	MStrat
02201	fourougbankoro	G	The IC	MStrat
02363	pelezi	G	The IC	MStrat
02826	pine	G	The IC	MStrat
DJE6	guincultiv	G	The IC	MStrat
150	guincultiv	G	The IC	MStrat/MLST
416	guincultiv	G	The IC	MStrat/MLST
410_Guyane	IC_via_Guyana	G	Guyana	MStrat/MLST
Maclaudi_S/BIL4	Maclaudi	G	The IRD	MStrat/MLST
126_RCI	IC_via_Guyana	Intergroup Hybrid	Guyana	MStrat/MLST
200_Guyane	IC_via_Guyana	SG2	Guyana	MStrat/MLST
LAF_159/33-3	luki	SG1	The DRC	MLST
LK_35/2-1	luki	SG1	The DRC	MLST
LK_35/2-3	luki	SG1	The DRC	MLST
S_19/58-3	luki	SG1	The DRC	MLST
S_9/23-1	luki	SG1	The DRC	MLST
LAF_159/58-3	luki	SG1	The DRC	MStrat
NIAOULI3	niaouli	SG1	The IC	MStrat
S_19/58-2	luki	SG1	The DRC	MStrat
466_Guyane	IC_via_Guyana	SG1	Guyana	MStrat/MLST
A03_Guyane	IC_via_Guyana	SG1	Guyana	MStrat/MLST
NIAOUL14	niaouli	SG1	The IC	MStrat/MLST
S_9/23-3	luki	SG1	The DRC	MStrat/MLST

Table 6 continued

Genotype name	Population	Diversity group	Collection	Core
002	ineac7	SG2	The IC	MLST
066	ineac2	SG2	The IC	MLST
3/66	erect	SG2	Uganda	MLST
14/19	erect	SG2	Uganda	MLST
203/74	nganda	SG2	Uganda	MLST
247/62	nganda	SG2	Uganda	MLST
042	ineac2	SG2	The IC	MStrat
1/2	erect	SG2	The IC	MStrat
3/5	erect	SG2	Uganda	MStrat
077_Guyane	IC_via_Guyana	SG2	Guyana	MStrat
257/18	nganda	SG2	Uganda	MStrat
392_Guyane	IC_via_Guyana	SG2	Guyana	MStrat
024	ineac7	SG2	The IC	MStrat/MLST
3/20	erect	SG2	The IC	MStrat/MLST
202/63	nganda	SG2	Uganda	MStrat/MLST
Kabwegyemere_T5_S1	K_Kab	UW	Uganda	MLST
Mbale_Mahungu_T29	K_MM	UW	Uganda	MLST
Mbale_Mahungu_T9_S6	K_MM	UW	Uganda	MLST
Ngogo_Kibuguta_B_T3_S2	K_NK	UW	Uganda	MLST
Kabwegyemere_T6_S1	K_Kab	UW	Uganda	MStrat
Kyamuhoro_Cpt11_T15	I_11	UW	Uganda	MStrat
Mbale_Mahungu_T23	K_MM	UW	Uganda	MStrat
Ngogo_Kibuguta_A_T10_S2	K_NK	UW	Uganda	MStrat
Ngogo_Kibuguta_A_T9	K_NK	UW	Uganda	MStrat
Ngogo_Kibuguta_B_T2_S3	K_NK	UW	Uganda	MStrat
Ngogo_Kibuguta_T10_S3	K_NK	UW	Uganda	MStrat
Kyamuhoro_Cpt11_T8	I_11	UW	Uganda	MStrat/MLST

Potential for enhancing *Coffea canephora* breeding programs and research networks

Beyond the limited number of genotypes, core collections are of great interest for facilitating and promoting material exchange between research centers. The exchange of genotypes for strictly allogamous perennial species such as *C. canephora* should be based on cuttings. As far as we know, all of the genotypes that are described in the 565 set are available from each field collection manager. This exchange will then allow breeders and collection managers to build an international network of coffee conservation and breeding. The diverse climatic conditions, the diversity of stresses and the specific selection criteria will allow breeders from all countries to use this core collection for their specific interest and to share their results with other research teams. These exchanges are currently quite infrequent due mainly to the lack of knowledge about the collections that are available elsewhere and their diversity. Our work will contribute to a better mutual knowledge of the available diversity throughout the "*canephora* world" and of the potential to construct a common set of diversity analyses through producing countries.

Considering the accessions that are currently present under the different field collections, we can already propose some exchanges between research centers based on our results. For example, it should certainly be of interest for the breeding program of recurrent selection in the Ivory Coast to integrate certain UW genotypes from the Ugandan group as a specific population. Guinean genotypes, mainly in field collections in the Ivory Coast, should be of interest for use as genitors in breeding strategies in Uganda, DRC, Brazil and Ecuador. The SG1 genotypes from the DRC seem to be highly representative from this group, and their integration in programs in other countries, in Uganda as well as the Ivory Coast, will be of interest.

The proposed CC48 and CC77 core collections should provide these research centers access to a representation of the whole genetic diversity that is currently available for *C. canephora*.

Paving the way for association mapping

The definition of core collections could also be the first step for further analyses such as association studies. In *C. canephora*, diversity studies have been complemented with genetic mapping (Leroy et al. 2011; Crouzillat et al. 2013) and Linkage Disequilibrium (LD) studies (Cubry et al. 2013b). These two types of analyses are preliminary steps in setting up association studies.

Association studies are performed on non-structured populations to associate markers with phenotypic diversity. As has already been proposed in grapes (Barnaud et al. 2006) and olives (El Bakkali et al. 2013), core collections can help to manage association studies for breeding complex traits within species. As a first step, LD studies in coffee populations can define the density of markers that are necessary for efficient association studies (Rafalski 2009). In our species, the density of markers should be very different depending on the breeding populations from the six diversity groups. For example, the marker density should be far lower for the Guinean genotypes than for the SG2 genotypes (Cubry et al. 2013a). Ongoing work using GbS will allow in the short-term for a high number of markers that are well dispersed throughout the genome.

In a second step, association studies could be performed in the target populations by identifying marker alleles that are associated with favorable phenotypic traits. For example, in Guinean populations where LD is rather persistent, association studies could be initiated for certain agronomical such as bean size and certain quality traits such as caffeine content with a medium marker density (Cubry et al. 2013b). In the whole core collection that was defined in the present study (77 accessions), the structure of the genetic diversity has been erased. Therefore, association studies could be performed on this core collection for traits of high heritability, such as bean size, vigor or tolerance to biotic stresses, when a high density of markers throughout the genome will become available.

Large perspectives for purpose-oriented core collections

This study provides a foundation for building different core collections depending on the objectives of the different research programs. We provide here only a global diversity-oriented construction of a core collection as this construction appears the most appropriate for immediate material exchange. However, one should be interested in studying adaptive evolution and identifying selective marks at the diversity group level. We provide here useful information for the construction of specific, small core collections in every known diversity group. These "group-nested" core collections might be useful for specific purposes and research objectives. We suggest using the MStrat strategy when building these nested core-collections because this approach will better capture the whole genetic diversity of the considered genetic group. We propose such core collections in the supplementary material (Supplementary 2). In the near future, these sets will be useful for gene discovery, the study of adaptive selection, massive genome resequencing and GbS studies. From a medium- and long-term perspective, these core collections could be used as kernels for the implementation of association studies in *C. canephora* supplemented with genotypes regarding specific objectives.

Conclusion

In the present study, we define core collections within our working species, *C. canephora*, that are well-suited for different objectives and projects. Given its economic importance and new challenges to coffee growers, such as pests and the adaptation to climate changes, it appears highly important to develop such resources. These core collections are valuable tools for all researchers working on coffee diversity and breeding and should facilitate the construction of an international network for the management of coffee diversity, conservation and breeding with an active mutual exchange of material and information. These collections will enhance breeding programs worldwide and pave the way for association studies and for the global management of coffee diversity in breeding programs.

References

Anthony F (1992) Les ressources génétiques des caféiers : collecte, gestion d'un conservatoire et évaluation de la diversité génétique. Collection Travaux et Documents Microfichés n°81, ORSTOM (now IRD), Paris

Balfourier F, Roussel V, Strelchenko P, Exbrayat-Vinson F, Sourdille P, Boutet G, Koenig J, Ravel C, Mitrofanova O, Beckert M, Charmet G (2007) A worldwide bread wheat core collection arrayed in a 384-well plate. Theor Appl Genet 114:1265–1275

Barnaud A, Lacombe T, Doligez A (2006) Linkage disequilibrium in cultivated grapevine, *Vitis vinifera* L. TheorAppl Genet 112:708–716

Belaj A, Dominguez-GarcíaMdC AS, Urdíroz NM, De la Rosa R, Satovic Z, Martín A, Kilian A, Trujillo I, Valpuesta V, Del Río C (2012) Developing a core collection of olive (*Olea europaea* L.) based on molecular markers (DArTs, SSRs, SNPs) and agronomic traits. Tree GenetGenomes 8:365–378

Berthaud J (1986) Les ressources génétiques pour l'amélioration des caféiers africains diploïdes. Evaluation de la richesse génétique des populations sylvestres et de ses mécanismes organisateurs. Conséquences pour l'application, Paris (FRA), ORSTOM, 379 pp

Berthaud J, Charrier A (1988) Genetic resources of Coffea. In: Clarke RJ and Macrae R (eds) Coffee, vol. 4 Agronomy, London: Elsevier Applied Science, pp. 1–42

Brown AHD (1989) Core collections: a practical approach to genetic resources management. Genome 31:818–824

Combes MC, Andrzejewski S, Anthony F, Bertrand B, Rovelli P, Graziosi G, Lashermes P (2000) Characterization of microsatellite loci in *Coffea arabica* and related coffee species. Mol Ecol 9:1178–1180

Crouzillat D, Rigoreau M, Lefebvre-Pautigny F, Priyono, Broun P, Lambot C (2013) A coffee high density genetic map for quantitative trait loci analysis on agronomical, technological and biochemical characteristics in robusta and arabica. In: ASIC 24th International Conference on Coffee Science (ASIC Costa Rica 2012), 11–16 Nov 2012, San José, Costa Rica, 6 p

Cubry P (2008b) Structuration de la diversité génétique et analyse des patrons de déséquilibre de liaison de l'espèce Coffeacanephora Pierre ex Froehner. Thèse de doctorat de l'Université Montpellier II, Montpellier. http://tel.archivesouvertes.fr/tel-00365078/fr/

Cubry P, Musoli P, Legnaté H, Pot D, De Bellis F, Poncet V, Anthony F, Dufour M, Leroy T (2008) Diversity in coffee assessed with SSR markers: structure of the genus *Coffea* and perspectives for breeding. Genome 51:50–63

Cubry P, De Bellis F, Pot D, Musoli P, Leroy T (2013a) Global analysis of *Coffea canephora* Pierre ex Froehner (Rubiaceae) from the Guineo-Congolese region reveals impacts from climatic refuges and migration effects. Genet Resour Crop Evol 60(2):483–501. doi:10.1007/s10722-012-9851-5

Cubry P, De Bellis F, Avia K, Bouchet S, Pot D, Dufour M, Legnate H, Leroy T (2013b) An initial assessment of linkage disequilibrium (LD) in coffee trees: LD patterns in groups of *Coffea canephora* Pierre using microsatellite analysis. BMC Genom 14:10. doi:10.1186/1471-2164-14-10

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

Dussert D, Lashermes P, Anthony F, Montagnon C, Trouslot P, Combes MC, Berthaud J, Noirot M, Hamon S (1999) Le caféier, *Coffea canephora*. In: Hamon P, Seguin M, Perrier X, Glaszmann JC (eds) Diversité génétique des plantes tropicales cultivées. CIRAD, Montpellier, pp 175–194

El Bakkali A, Haouane H, Moukhli A, Costes E, Van Damme P, Khadari B (2013) Construction of core collections suitable for association mapping to optimize use of mediterranean olive (*Olea europaea* L.) genetic resources. PLoS ONE 8(5):e61265

Escribano P, Viruel MA, Hormaza JI (2008) Comparison of different methods to sequence repeat markers. A case study in cherimoya (*Annona cherimola*, Annonaceae), an underutilised subtropical fruit tree species. Ann Appl Biol 153:25–32

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evolut Bioinform Online 1:47–50

Franco J, Crossa J, Warburton ML, Taba S (2006) Sampling strategies for conserving maize diversity when forming core subsets using genetic markers. Crop Sci 46:854–864

Frankel OH, Brown AHD (1984) Plant genetic resources today: a critical appraisal. Crop genetic resources. In: Holden JHW, Williams JT (eds) Conservation and evaluation. Georges Allen and Unwin Ltd, London, pp 249–257

Gomez C, Dussert S, Hamon P, Hamon S, 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

Gouesnard B, Bataillon TM, Decoux G, Rozale C, Schoen DJ, David JL (2001) MSTRAT: an algorithm for building germplasm core collections by maximizing allelic or phenotypic richness. J Hered 92:93–94

Hamon S, Noirot M, Anthony F (1995) Developing a coffee core collection using the principal components score strategy with quantitative data. In: Brown AHD, van Hintum TJL, Morales EAV (eds) Hodgkin T. IPGRI Wiley-Sayce publication, Core collections of Plant Genetic resources, pp 117–126

Haouane H, El Bakkali A, Moukhli A, Tollon C, Santoni S, Oukabli A, El Modafar C, Khadari B (2011) Genetic structure and core collection of the world olive germplasm bank of Marrakech: towards the optimized management and use of Mediterranean olive genetic resources. Genetica 139:1083–1094

ICO (2013) International Coffee Organization.ICO Annual Review 2012/13.http://www.ico.org/news/annual-review-2012-13-e.pdf

Laucou V, Lacombe T, Dechesne F, Siret R, Bruno JP, Dessup M, Dessup T, Ortigosa P, Parra P, Roux C, Santoni S, Varès D, Péros JP, Boursiquot JM, This P (2011) High throughput analysis of grape genetic diversity as a tool for germplasm collection management. Theor Appl Genet 122:1233–1245

Le Cunff L, Fournier-Level A, Laucou V, Vezzulli S, Lacombe T, Adam-Blondon AF, Boursiquot JM, This P (2008) Construction of nested genetic core collections to optimize the exploitation of natural diversity in *Vitis vinifera* L. subsp. sativa. BMC Plant Biol 8:31

Leroy T, Montagnon C, Charrier A, Eskes AB (1993) Reciprocal recurrent selection applied to Coffeacanephora Pierre. I. Characterization and evaluation of breeding populations and value of intergroup hybrids. Euphytica 67:113–125

Leroy T, Montagnon C, Cilas C, Yapo AB, Charmetant P, Eskes AB (1997) Reciprocal recurrent selection applied to *Coffea canephora* Pierre. III.Genetic gains and results of first intergroup crosses. Euphytica 95:347–354

Leroy T, Marraccini P, Dufour M, Montagnon C, Lashermes P, Sabau X, Ferreira LP, Jourdan I, Pot D, Andrade AC, Glaszmann JC, Vieira LGE, Piffanelli P (2005) Construction and characterization of a *Coffea canephora* BAC library to study the organization of sucrose biosynthesis genes. Theor Appl Genet 111:1032–1041

Leroy T, De Bellis F, Legnate H, Kanamura E, Gonzales G, Pereira LFP, Andrade AC, Charmetant P, Montagnon C, Cubry P, Marraccini P, Pot D, de Kochko A (2011) Improving the quality of African robustas: QTLs for yield- and quality-related traits in *Coffea canephora*. Tree Genet Genomes 7:781–798. doi:10.1007/s11295-011-0374-6

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129. doi:10.1093/bioinformatics/bti282

Montagnon C (2000) Optimisation des gains génétiques dans le schéma de sélection récurrente réciproque de Coffea canephora Pierre. ENSA Montpellier, France, PhD thesis

Montagnon C, Leroy T, Yapo A (1992) Diversité génotypique et phénotypique de quelques groupes de caféiers (*Coffea canephora* Pierre) en collection. Conséquences sur leur utilisation en sélection. Café Cacao Thé 36:187–198

Montagnon C, Leroy T, Eskes AB (1998) Amélioration variétale de Coffea canephora. II. Les programmes de sélection et leurs résultats. Plantations, recherche, développement 5(2): 18–31

Montagnon C, Cubry P, Leroy T (2012) Amélioration génétique du caféier *Coffea canephora* Pierre :connaissances acquises, stratégies et perspectives. Cahiers de l'Agriculture 21:143–153. doi:10.1684/agr.2012.0556

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 52:634–646. doi:10.1139/G09-037

Odong TJ, Jansen J, van Eeuwijk FA, van Hintum TJL (2013) Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. Theor Appl Genet 126: 289–305

Peakall R, Smouse P (2006) GENALEX6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295

Perrier X, Jacquemoud-Collet JP (2006) DARwin software. http://darwin.cirad.fr/darwin

Pessoa-Filho M, Rangel PHN, Ferreira ME (2010) Extracting samples of high diversity from thematic collections of large gene banks using a genetic-distance based approach. BMC Plant Biol 10:127

Poncet V, Dufour M, Hamon P, Hamon S, de Kochko A, Leroy T (2007) Development of genomic microsatellite markers in *Coffea canephora* and their transferability to other coffee species. Genome 50:1156–1161

Rafalski JA (2009) Association genetics in crop improvement. Curr Opin Plant Biol 13:174–180

Ronfort J, Bataillon T, Santoni S, Delalande M, David JL, Prosperi JM (2006) Microsatellite diversity and broad scale geographic structure in a model legume: building a set of nested core collection for studying naturally occurring variation in *Medicago truncatula*. BMC Plant Biol 6:28

Rovelli P, Mettulio R, Anthony F, Anzueto F, Lashermes P (2000) Microsatellites in Coffea arabica L. In: Sera T, Soccol CR, Pandey A, Roussos S (eds) Coffee biotechnology and quality, Kluwer Academic Publishers, The Netherlands, pp 123–133

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Schoen DJ, Brown AHD (1993) Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. Proc Natl Acad Sci USA 90:10623–10627

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462

Thomas AS (1947) The cultivation and selection of Robusta coffee in Uganda. Emp J Exp Agric 15:66–81

Upadhyaya HD, Bramiel PJ, Sube S (2001) Development of a chickpea core subset using geographic distribution and qualitative traits. Crop Sci 41:206–210

Van Hintum TJL, Brown AHD, Spillane C, Hodgkin T (2000) Core collections of plant genetic resources. IPGRI Technical Bulletin 3, International Plant Genetic Resource Institute, Rome

Volk GM, Richards CM, Reilley AD, Henk AD, Forsline PL, Aldwinckle HS (2005) *Ex situ* conservation of vegetatively propagated species: development of a seed-based core collection for *Malus sieversii*. J Am Soc Hort Sci 130:203–210