Genetic resources of Coffea canephora have been introduced in several tropical countries with potential for crop development. In Ecuador, the species has been cultivated since the mid-20th century. However, little is known about the diversity and genetic structure of introduced germplasm. This paper provides an overview of the genetic and phenotypic diversity of C. canephora in Ecuador and some proposals for implementing a breeding program. Twelve SSR markers were used to analyze 1491 plants of C. canephora grown in different living collections in Ecuador, compared to 29 genotypes representing the main genetic and geographic diversity groups identified within the species. Results indicated that most of the genotypes introduced are of Congolese origin, with accessions from both main subgroups, SGI and SG2. Some genotypes were classed as hybrids between both subgroups. Substantial phenotypic diversity was also found, and correlations were observed with genetic diversity. Ecuadorian Robusta coffee displays wide genetic diversity and we propose some ways of improving Robusta in Ecuador. A breeding program could be based on three operations: the choice of elite clones, the introduction of new material from other countries (Ivory Coast, Uganda), and the creation of new hybrid material using genotypes from the different diversity groups.
1. Introduction

* Coffea canephora* is originated from the lowland tropical forests of Africa, which stretch from Guinea to Uganda, and its cultivation is recent (end of the 19th century). Robusta coffee fields are now widely found in all lowland intertropical regions of Africa, America, and Asia [1]. The genetic diversity of *C. canephora* was first described at molecular level in the 1980s [1–7]. Those studies revealed two main diversity groups, the Congolese and the Guinean groups (G). The Congolese group was subdivided into five subgroups (SG1, SG2, B, C, and UW). Only a small portion of this wide diversity (i.e., mainly SG1 and SG2) is used in current breeding programs, with the exception of the recurrent breeding program in Ivory Coast, which uses a larger share of this diversity, except that from Uganda [4, 8, 9]. A core collection encompassing a large share of known *C. canephora* diversity has been recently proposed [10]. This core collection contains genotypes from all the known diversity groups and is an interesting starting point from which to broaden genetic diversity in *C. canephora* breeding programs.

In Ecuador, *C. canephora* genetic resources were first introduced in the mid-20th century [11, 12]; the origin of this germplasm is diverse but little information is available on its true geographic origin or its diversity and genetic structure. This information is considered very important for future conservation and development conditions for a breeding program in the country.

Ecuadorian historical records show that the first introductions of *C. canephora* genetic material came from the “Tropical Agricultural Research and Higher Education Center (CATIE),” Costa Rica, in 1951, 1964, 1972, 1977, and 1986. They corresponded to the “Robusta” type (putative SG2) and all were planted at the Pichilingue Tropical Research Station (EETP) of the National Institute of Agricultural Research of Ecuador (INIAP), with the first *C. canephora* plantations appearing in 1952 in Los Ríos province, from where they were gradually extended to several coastal provinces and toward the north of Ecuadorian Amazonia [13].

Later, in 1987 and 2006, genetic material of the “Conilon” type (putative SG1) was imported from Brazil. Additionally, unofficial sources report introductions of genetic material imported as seeds from Vietnam and Indonesia (2009, 2010), as well as Robusta from Brazil (2010). Using seeds for genetic material transfers for a self-incompatible tree cannot ensure its genetic origin since crosses between genotypes from different genetic groups are likely to occur within germplasm collections [10].

During the second half of the 20th century, a nonspecific *C. canephora* breeding program was developed. However, a first group of elite material “clones” was identified by INIAP in 1998, based on yield and morphological traits. To date, these clonal *C. canephora* materials have been recommended for commercial planting under the conditions of northern Ecuadorian Amazonia. A recent study on the phenotypic characterization of *C. canephora* accessions planted in the living genebank collection, located in the EETP of the INIAP, showed a high variability between and within these accessions [13]. Consequently, knowledge on the genetic diversity of the material widely distributed in the Ecuadorian territory could help breeders and geneticists to understand the structure of introduced germplasm in order to design a *C. canephora* breeding program in Ecuador.

This paper will

(i) provide an overview of the genetic and phenotypic diversity and conservation of *C. canephora* in Ecuador,

(ii) address the bases for implementing a *C. canephora* breeding program.

2. Material and Methods

2.1. Plant Material. In all, 1491 leaf samples were taken from different *C. canephora* clonal plants collected during 2011, 2012, and 2013. In 2011, one hundred thirty-seven leaf samples were collected from two fields. The first field is the Dublins-Denisse Farm, nearby Isidro Ayora in Guayas province that has a living collection of *C. canephora* collected from different locations at the Amazonian region of Ecuador. The second field is located at the EETP research station, nearby Quevedo city, Los Ríos province. In 2012, two hundred leaf samples from 12 clonal accessions and 154 leaf samples from seven clonal accessions and a polyclonal mix were collected at the Dublins-Denisse Farm and the EETP, respectively. In 2013, one thousand leaf samples from 7 clones were sampled at the INIAP’s Amazonian Central Station (EECA).

The summary of the plant material used for genotypic analysis across the different years is shown in Tables 1(a) and 1(b). Among these, 48 samples were considered as duplicates and used to check the experimental reproducibility of the data and accordingly homogenize the data whenever needed.

2.2. DNA Preparation and Genotyping. Genomic DNA extractions were performed according to Cubry et al. [14]. The 1491 accessions were genotyped with the 12 SSR markers used by Leroy et al. [10]. Two different methods were used for genotyping and allele calling in 2011 and 2012/2013, respectively.

In 2011, PCR reactions ran as described in Cubry et al. [14]. 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+.

In 2012 and 2013, PCR reactions were performed according to De Bellis et al. [15]. In a solution A (25 μL total volume) containing 2.5 μL of PCR buffer (10 mM Tris-HCl, 50 mM KCl, 2 mM MgCl2, and 0.001% glycerol), 2.5 μL of dNTP (Jena Bioscience GmbH, Jena, Germany), 0.25 μL of MgCl2, 0.2 μL of 10 μM forward primer with an M13 tail at the 5’-end (5’-CACGACGTTGAAAAACGAC-3’), 0.25 μL of 10 μM reverse primer, 0.25 μL of fluorescently labeled M13-tail (6-FAM, NED, VIC, or PET from Applied Biosystems, Foster City, California, USA), 0.1 μL of Taq DNA polymerase (Sigma, St. Louis, Missouri, USA), 5 μL of template DNA (5 ng/μL), and 14 μL of H2O. The PCR conditions were as follows: an initial denaturation at 94°C for 5 min; 30 cycles...
Table 1: Genetic material analyzed and controls used in the genotypic analyses.

(a) Analyzed plants

<table>
<thead>
<tr>
<th>Year</th>
<th>Locations</th>
<th>Number of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>Dublinsa-Denisse Farm</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>EETP</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Total 2011</td>
<td>137</td>
</tr>
<tr>
<td>2012</td>
<td>Dublinsa-Denisse Farm</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>EETP</td>
<td>154</td>
</tr>
<tr>
<td></td>
<td>Total 2012</td>
<td>354</td>
</tr>
<tr>
<td>2013</td>
<td>EECA</td>
<td>1000</td>
</tr>
</tbody>
</table>

(b) Controls

<table>
<thead>
<tr>
<th>Diversity group</th>
<th>Diversity subgroup</th>
<th>Number of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guinean</td>
<td>G</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>3</td>
</tr>
<tr>
<td>Congolese</td>
<td>SG1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>SG2</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>UW</td>
<td>5</td>
</tr>
<tr>
<td>Outgroup</td>
<td>Coffea eugenioides</td>
<td>1</td>
</tr>
</tbody>
</table>

at 94°C for 45 s, 55°C for 45 s, and 72°C for 1 min; and a final extension at 72°C for 10 min. PCR products were pooled in a solution B containing 2 μL of 6-FAM, 2 μL of VIC, 2.5 μL of NED, and 3.5 μL of PET of A’s solutions. From this solution B, a volume of 4 μL was taken and added to 10 μL of Hi-Di formamide and 0.12 μL of GeneScan 600 LIZ size standard and analyzed on an ABI 3500xL Genetic Analyzer (Life Technologies, Carlsbad, California, USA). Alleles were scored using GeneMapper v4.1 software (Applied Biosystems, Foster City, California, USA).

2.3. Data Analysis

2.3.1. Whole of Ecuador Genetic Analyses. Data were obtained separately in 2011, 2012 and 2013 for each sampling year, from which merged three matrices. Standard controls of known genotypes and duplicates samples together with genotyping database (SagaCityWeb, CIRAD, unpublished) were used to help us to pool/shift alleles to ensure good coherence of the genotyping data. From the raw data, were removed the 48 duplicates and the samples with more than 17% of missing data (i.e., two missing markers) for further analysis. The number of alleles detected per marker was recorded and compared to those detected in Leroy et al. [10]. We computed dissimilarity matrices between individuals using a simple matching index with Darwin 6.0 software [16].

A first diversity tree was drawn up using data from the 2011 sampling operation in Ecuador. Diversity from Ecuador was evaluated in relation to the global diversity of the species.

A global analysis, using data from the three sampling years, was performed to identify unique genotypes for future conservation and for breeding programs. In all, 1168 samples were kept for analysis, based on missing values (<8% not available, i.e., one marker missing). A Neighbor-Joining (NJ) tree was built on the whole data removing pairs of data with more than 70% of missing values. To take into account the noise due to highly repeated genotypes, the max-length subtree procedure was used to eliminate redundancy and to identify the number of unique individuals without loss in the number of alleles. A Principal Coordinates Analysis (PCoA) was used to construct a good image of the diversity between genotypes.

Specific analyses were performed at the DUBLINSA and EECA stations to check the homogeneity and the diversity of the materials introduced.

2.3.2. The EETP Collection. Phenotypic analysis was performed at the EETP collection, using data from 2010 to 2012. The evaluation was carried out in 256 plants, corresponding to 16 accessions. Each accession contained different number of individual (between 12 and 20). A phenotypic dendrogram was performed by UPGMA clustering method using the Euclidean distance.

Genotyping data were available for 146 plants. Tree diversity was determined for molecular data, and their relationship with phenotypic data generated by Plaza et al. [13] was studied.

Phenotypic data were measured between 2010 and 2012 for all the trees planted in 2007. During this time, the traits observed were the following: plant height (PH), stem diameter (SD), number of branches (NB), number of nodes per branch (NN), and internode length (IL). Outlier data were removed whenever found and replaced the missing values. PCA analyses were performed using the dudi.pca function from the ade4 R package [17]. Five principal components were chosen after observing the screen plot of
Two types of analyses were performed on these data. A clustering tree was first built on 249 plants using the dissimilarity Euclidian distance evaluated by average linkage for the 17 traits observed across the three years. Another analysis was performed using a Principal Components Analysis (PCA) on five vigor traits (yield of cherry beans, number of cushion flowers, total of productive branches, total of branches per tree, and tree height) across three years observed on the 146 trees analyzed for genetic diversity.

3. Results

3.1. Whole of Ecuador Genetic Analyses. The results on the clonal trials at the EETP and Dublinsa revealed great diversity between “clones,” and also a high level of diversity within “clones,” which was more surprising. The global genetic analysis indicated the Congolese origin for all the genotypes. They were classed as SGI or SG2 genotypes, with some hybrids between these two Congolese subgroups. This result can be explained by considering the origin of plants introduced in Ecuador: the SGI genotypes could have come from Brazil (Conilon genotypes are known to be from the SGI group) and the SG2 genotypes from CATIE (from the Congo basin). Figure 1 presents the diversity tree identifying the genotypes from Ecuador within the global diversity of C. canephora.

A global analysis was performed in 2015, based on the results obtained for the 1168 accessions, by considering all the different genotypic classes revealed for each year in each plot. A global diversity tree was established using these 1168 accessions (figure not shown). After eliminating redundancy, the 138 unique genotypes (including 29 controls) were observed in the diversity tree (figure not shown). The PCoA analysis performed 138 unique genotypes, most of these genotypes were included in the SGI and SG2 diversity.
groups, and six genotypes could be considered as hybrids between the groups (Figure 2).

Figure 3 shows the results for the genotypic analysis, considering the “clones” analyzed in 2013 at the EECA station. For each accession/clone, 2 to 8 genotypic classes were identified. Accessions NP-3013 and NP-2044 only had one genotype. Some genotype classes were similar for certain accessions: genotype 1 from NP-2024 was similar to genotype 2 for NP-3056; genotype 1 from NP-3013 was similar to genotype 2 for NP-3018 and also to genotype 3 for NP-3056; genotype 2 from NP-3072 was similar to genotype 3 for NP-4024; genotype 4 from NP-2024 was similar to genotype 6 for NP-4024. The dissimilarity tree at the EECA living collection (Figure 4) yielded 50 genotyping classes. Several classes grouped individuals belonged to different accessions.

3.2. EETP Collection. A diversity tree was constructed considering the 146 plants from the 16 different origins with data available for molecular diversity. Twelve diverse groups were identified within the 146 (from 154) plants analyzed in 2012 (Figure 5).

On the other hand, a phenotypic characterization of this collection was reported by Plaza et al. [13] showing a wide range of variation in most of the agro-morphological traits evaluated for each plant. The existence at this level of phenotypic variation was found between and also within accessions. This phenotypic result was the first to open up the possibility of off-types in this collection; despite this, 25 plants were selected as “elite” material, but the most important variations considered when selecting those elite plants were plant yield and plant height [13]. These 25 plants (24 analyzed) were identified by their genetic diversity based on two analyses, per origin and individually:

(i) Per origin analysis: COF1 p2 was different from COF1 p10; COF3 p2, 5, 8, 18, and 19 were genetically similar; p7 was different; COF4 p7 and 9 were similar; p15 was different. COF 5 p6, 15, 16, 17, and 19 were similar; NP2024 p7, 10, 15, and 17 were similar; NP2044 p6, 16,
and 17 were similar; NP3018 p 8 and 19 were probably different, since the origin was very variable.

(ii) Global genetic diversity of the 24 “elite” clones: COF1 p10, COF3 p2, 5, 8, 18, and 19, and COF4 p7 and 9 were similar; COF5 p6, 15, 16, 17, and 19 and NP2024 p7, 10, 15, and 17 were similar; NP2044 p6, 16, and 17 were similar; COF4 p15 and COF3 p7 were similar; COF1 p2 and NP3018 p19 were almost identical; their similarity has to be verified.

The genetic diversity of the selected clones was therefore very low, since only 5 different groups of genotypes (six if including NP3018) were present in the selected clones. As regards the global diversity of the Pichilingue collection, almost half of this diversity was not present in the selected clones. The diversity of the Conilon/SG1 group was not present in the selected clones from this collection.

3.3. Phenotypic Diversity. The phenotypic dendrogram at the EETP C. canephora collection (Figure 6) indicated 3 clusters. One cluster (group I) included 14 accessions with a high diversity within the group. The second and third group consisted of one isolated accession, respectively.

Phenotypic results were highly variable in most plants belonging to the same accession. To determine the level of this variation, all individuals from NP2024 accession were used as sample. The dendrogram indicated 3 clusters, confirming the high level of variability among the individuals of this accession (Figure 7).

A complemented phenotypic PCA using the vigor traits was carried out with the 12 genetic groups previously identified at the EETP collection (i.e., Figure 5) showing a significant phenotypic diversity within of the groups (Figure 8). Group 7 was considered as presenting short internodes, and group 9 exhibited low vigor and short internodes, while group 5 had rather long internodes and high vigor, and group II was characterized by low vigor and long internodes.

On the other hand, using PCA, the correlation circle between observed traits (Figure 9) indicated that the internode length (IL) for the three years was related to axis 2, since the other vigor traits, particularly stem diameter (SD), plant height (PH), and the total number of nodes per tree (NNT) were related to axis 1.

4. Discussion

Our analyses

(i) confirmed the genetic diversity of accessions from Ecuador, covering the SG1 and SG2 subgroups of Congolese diversity, in accordance with the history of introductions,

(ii) revealed great genetic and phenotypic diversity between clones, but also a large number of genotype classes within most of the “clonal” accessions,

(iii) suggested some ways of implementing a breeding strategy for C. canephora using the available diversity.

The first point concerns the reliability of our work for the three-year experiment. The first-year analyses were
Figure 4: Dissimilarity analysis of 7 "clonal accessions" from the EECA station in 2013. NJ-Tree on the 63 different unique genotypes using 12 SSR markers. Branch support is expressed in percent of presence after 10000 bootstraps (values above 0.5 are displayed).

performed using LI-COR 4300 technology. The following analyses were performed using an ABI sequencer, with different control plants. Misidentification of alleles during calling and binning processes are known caveats of SSR studies [18]. To solve this issue, laboratory good practices were implemented by including controls of known genotypes and by repeating some samples from one study to another. The overall results are in accordance with what was expected and thus could be considered of good quality.

It was a challenge to analyze concomitantly all the data for the final evaluation. The results could thus be considered as a compromise between all the different data. However, the low quality of some leaf material was also an element that we took into account, due to difficulties for DNA extraction and analysis with microsatellites. Some leaves samples were not correctly analyzed and were removed from the final analyses.

In 2012, at DUBLINSA collection, we detected genetic diversity within "clones" in our analyses, meaning that the "clones" were not genetically homogeneous, as they ought to have been (data not shown). We also observed mixtures between genotypes in both clonal trials. The CONERBO and POLICLON genotypes which belonged to Conilon type can be considered as plants from a mix of seeds, introduced from Brazil in the 80s; these origins present high genetic diversity.

Regarding the EETP collection, it should be noted that many plants, with different labels, were quite similar from a genetic point of view. We also had to consider that the Conilon exhibited wide diversity, possibly due to its environmental share of diversity or due to their seed origin.

In 2014, a global analysis of C. canephora diversity was carried out, using a core collection approach [10]. The comparison with the present study confirmed that the diversity observed within Ecuadorian accessions accounts for about 57% of core collection diversity, considering the different alleles. Therefore, the information provided by this study will help breeders choose the most appropriate plant(s) or accession(s) to be incorporated into their breeding programs.

Another finding was the small number of intergroup hybrids between the SG1 and SG2 diversity groups. This low level of hybridization can be explained by the history of the introduced material. Both introductions were composed of
genotypes from a single subgroup (SG2 from CATIE and SG1 from Brazil). The accessions were planted in one location (EETP) and then transferred to Manabí, Santo Domingo de los Tsáchilas, and Morona Santiago provinces, mainly by cuttings. Thus, few hybrids can be found between groups in the Ecuadorian collections.

Nevertheless, high diversity was observed within each diversity group (SG1 and SG2). For SG2, this diversity was related to the large number of accessions that were introduced, and previous studies [7] confirmed the high diversity within this group. The SG1 genotypes were mainly introduced by seeds, for which diversity is always greater than the clones in this allogamous species. High genetic and phenotypic diversity within this group has been recently established. In the case of the Pichilingue germplasm, it should be noted that high levels of phenotypic variation were previously reported by Plaza et al. [13] and could be related to the continuous pollen interchange and/or possible mix of seeds from different segregated populations.

The phenotypic characterization carried out by Plaza et al. [13] and our study enabled us to identify a wide range of variation in most of the agro-morphological traits evaluated per plant, with significant phenotypic diversity within genotypes. In this respect, it is important to keep in mind that a phenotype is a product of genotype × environment interaction, which we found in our results. Therefore, plants may be morphologically similar, but this does not necessarily imply genetic similarity, since different genetic bases can result in similar phenotypic expression [18] and, as observed in our results, the same genotype can lead to substantial differences in phenotypic expression.

In our view, a combined analysis of phenotypic and molecular marker results is crucial for a better understanding of evolutionary changes in this introduced species; this would allow a better analysis of variation patterns within *C. canephora* for evaluating their future adaptive potential in different geographical regions of Ecuador. Differences between phenotypic and genetic information have also been found in other crops [19–22].

Lastly, this study was intended to identify diversity and ways of using it to increase the production of Robusta coffee in Ecuador. In Ivory Coast, considering the high genetic diversity found in *C. canephora*, a program of reciprocal recurrent selection was conducted using the hybrid vigor observed between genotypes of different origins [8, 23, 24]. Based on the Ivorian experience, we might propose some steps for breeding *C. canephora* in Ecuador, with optimum use of the existing diversity and improved management of the existing material, based on the results presented. This strategy for the implementation of a *C. canephora* breeding program should also be of interest for other countries where coffee genotypes have been introduced in recent decades.
As a first step, we propose the following actions, using the diversity existing in Ecuador:

(i) To complete genotypic analyses with phenotypic data to increase knowledge on the accessions in the field for their vigor, productivity, and stress and disease tolerance

(ii) To reorganize the collections based on the genetic diversity observed in our studies to avoid duplications and identifying unique genotypes

(iii) To implement a breeding program by selecting the best genotypes for traits of interest (yield, biotic and abiotic stress tolerance, and adaptation to different edaphoclimatic conditions) (these genotypes will be planted as clones in a multisite trial design. Attention will be paid to the high diversity of plants from the Conilon/SG1 group. These clonal trials will enable the selection of a set of improved genotypes for farmers)

As a second step, we propose the following actions, by increasing genetic diversity:

(i) To introduce new genetic material from diverse groups (i.e., Guinean, Ugandan, and hybrids between them) that are not present in Ecuador and test them under Ecuadorian conditions

(ii) To establish a breeding program based on new hybrids obtained from crosses between genotypes from different diversity groups, adapted to the edaphoclimatic conditions in Ecuador (these new hybrids will use accessions from Ecuador and introduced accessions to create new hybrids. This program will use the diversity of both the SG1 and SG2 groups, as hybrids between these groups display hybrid vigor and good drought tolerance, as already observed in Ivory Coast. All the new hybrids will be tested under all the conditions in Ecuador, and this will lead to a new selection of elite clones or hybrids.

Figure 6: Phenotypic dendrogram for 16 accessions at EETP *C. canephora* collection. The data were collected between 2010 and 2012 using vigor traits (productivity, stem diameter, plant height, number of branches, number of nodes, and internode length). Branch support is expressed in percent of presence after 10000 bootstraps (values above 0.2 are displayed).
This improved material will be distributed to farmers through cutting gardens (for clones) or seed gardens (for hybrids)

Over the long term, hybrid selection might be the optimum breeding method, as seeds are more suitable for distribution to growers, and the nurseries are easier for farmers to manage.

5. Conclusions

The present research concluded that Ecuadorian Robusta coffee displays a wide genetic diversity between clones and also a high level of diversity within clones. This research confirms that most of the *C. canephora* genotypes introduced in Ecuador are of Congolese origin, containing accessions from both subgroups, SG1 and SG2.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors’ Contributions

Rey Gastón Loor Solórzano, Fabien De Bellis, and Thierry Leroy contributed equally to this work.

Acknowledgments

The authors thank DUBLINSA SA, the National Institute of Agricultural Research (INIAP), and the Secretary of High Education, Science, Technology and Innovation (SEnesCYT) for financial support and Solubles Instantaneos CA for its technical assistance in the postharvest analysis. The
authors also thank Mr. Peter Biggins for revising the English version.

References


