PROJECT: Studies on the genetic bases of flavour in Nacional varieties in order to produce more productive and full arriba flavour varieties

Final Scientific Report of the Project

Claire Lanaud
Senior Scientist
Cocoa Department-CIRAD

2010

Final Scientific Report of the project:

STUDIES ON THE GENETIC BASES OF FLAVOUR IN NACIONAL VARIETIES IN ORDER TO PRODUCE MORE PRODUCTIVE AND FULL ARRIBA FLAVOUR VARIETIES

Our objectives were to understand the structure, diversity and origin of the Nacional cacao varieties from Ecuador and to characterise the Nacional collection for sensorial and biochemical traits.
One of the objective of the present study is also to evaluate the relationships between genetic, biochemical and sensorial traits in order to identify markers linked to cocoa aromatic flavor of Nacional varieties.

Here we present:
• The results of diversity and domestication analyses described in 2 papers:

I. Tracing the native ancestors of the modern *Theobroma cacao* L. population in Ecuador (p1)
II. Insight about the wild origin, migration and domestication events of the fine flavour Nacional *Theobroma cacao* L. variety from Ecuador (p 20)

• The first results of association studies related to biochemical and sensoriel traits of the Nacional population studied (p 42)

I. Diversité and domestication of the Ecuadorian Nacional variety

1) Tracing the native ancestors of the modern *Theobroma cacao* L. population in Ecuador

Abstract:

The native ***Theobroma cacao*** **L** variety from Ecuador, known as Nacional, is famous by its particular floral and fine cocoa flavour called "ARRIBA". It has been subjected, however, to genetic erosion, due principally to successive introductions of foreign germplasm, which have hybridised with the native variety. This hybrid population, which now constitute the modern Nacional variety, gradually replaced the native plantations from the beginning of 20th century. In order to increase our knowledge on

the pattern of diversity among modern Nacional cacao populations, forty SSR markers were selected to analyse three hundred and twenty-two samples collected within germplasm banks of CCA Tenguel and EETPichilingue. These 332 samples originated from different geographical origins along of the coast region from Ecuador. A total of 169 alleles were observed in this Ecuadorian cacao population. Overall, the genetic parameter evaluated showed a high heterozygosity level. The F-statistics showed negative values for F_{IS} and F_{IT}, reflecting the high level of heterozygosity observed. A low level of differentiation was observed between the two studied Ecuadorian collections and a factorial analysis of correspondence (FAC) showed a continuous variation in the pool of these hybrid individuals. Some cacao trees showed high levels of homozygosity. Based on the dissimilarity analysis, this homozygous group was located at an extreme side of the hybrid Ecuadorian population. The origin of this hybrid population was assessed using paternity analysis. The highly homozygous genotypes appeared to be at the origin of more than 90% of the modern Nacional population; a small part of this population could correspond to direct crosses between the highly homozygous Nacional genotypes and Trinitario or Lower Amazon Forastero types probably introduced at the beginning of 20 century; however, the larger part of the population probably represent more advanced generation of recombination involving crosses between hybrid plants. Our analyses suggest that the highly homozygous individuals identified in this study could represent the native Nacional variety present in Ecuador before the foreign introductions.

Introduction:

***Theobroma cacao* L.** is a perennial species member of the Malvaceae family. This species has its putative centre of origin and diversity in the upper basin of the Amazon river tributaries (Napo, Putumayo and Caqueta) at the foot of the Ecuadorian Andes (Cheesman, 1944). The greatest wild cacao diversity is located in the Amazonian region. During the course of cacao domestication a small part of this primary cacao diversity has been exported to various destinations to be cultivated, and three main varieties were spread and cultivated around the world before 1950. They have been traditionally described as: Criollo, Forastero and Trinitario.

Criollo was originally cultivated by the Mayas in Central America and represents the first domesticated cacao. Its cacao beans have been used in food for more than 2,600 years as shown by archaeological evidence in cacao residues found in Mayan cooking vessels which date back to approximately 600 years BC (Hurts *et al*, 2002).

Forastero was domesticated later than Criollo during the 17th century. One variety called "Amelonado", originated in the Lower Amazon region of Brazil and was spread around the world; especially in West Africa, which presently accounts for over 80% of the world's production of cocoa.

Trinitario is an hybrid population developed from natural hybridization between Criollo and Lower Amazon Forastero (Pound 1938, 1945) that took place in the island of Trinidad, when the Criollo plantations were partly destroyed by a natural disaster in the late eighteenth century (1727). Subsequently, Forastero plants were imported to Trinidad (Chessman 1944) and hybridised with the few surviving Criollo trees giving rise to hybrid forms called Trinitario.

In addition to these three main cacao groups, a primitive cultivated population of ***T. cacao*** known as "***Nacional***" variety existed in the Pacific Coast Region of Ecuador. Although its exact origin is unknown, it is considered as native of this country. Through the years its classification has varied : it was classified as Forastero by Cheesman (1944) and Soria (1970), placed among Criollo by Enriquez (1992) and more recently considered as a group different from Criollo or Forastero, but genetically closer to the later (Lerceteau *et al* (1997a). According to Allen *et al* (1983), the Nacional variety might have been derived from a local wild population which could now have completely disappeared together with the original forest cover of the coast region. Soria (1970) and Vera (1987, 1993) suggested as its center of origin the oriental declivities of the Andes mountain in the Amazonian area of Ecuador.

According to historical archives of Ecuador, in early 1600s, small plantations of Nacional cacao variety existed along the Guayas river shores that spread in its tributaries, Daule and Babahoyo (upward rivers), which are located in a region called "ARRIBA" (Figure 1). This variety develops a strong floral aroma known as "ARRIBA" flavour in international markets. This quality is highly appreciated to make fine chocolates. The Nacional variety is exclusively produced in Ecuador and is considered as a strategic crop for its economic development. With exception of the province of Esmeraldas, the Nacional cacao was the only one grown in the western

coastal plain of Ecuador until 1890 (Van Hall, 1932) when, for the first time, , pods of cultivars called "Venezuela" characterised by a high productivity, were introduced from Trinidad. This Trinidad accession probably corresponded to Criollo/Trinitario types. With the appearance in 1916 of ***Monilia roreri*** (Rorer, 1926); and, in 1919 of ***Crinipellis perniciosa*** (Pound, 1938), fungi that respectively cause the frosty pod and witches' broom diseases, foreign materials were introduced in larger quantity in Ecuador, and particularly genotypes of the Scavina family from Upper Amazon region. Scavina was used from this time onwards in the breeding programs, because of its known high level of resistance to witches' broom (Ampuero, 1960). Presently, more than 95% of the original area planted with Nacional cacao has been replaced with hybrid material involving foreign clones, especially Trinitario type (Loor, 2002). This genetic mixing has lead to a dilution of the "ARRIBA" flavour of the traditional Nacional variety.

In an effort to preserve the Nacional cacao genetic resources and to broaden its genetic basis, several expeditions have been made since the 1950's, and a substantial amount of germplasm has been collected along the Pacific Coast Region of Ecuador. Today, an important fraction of these accessions are maintained as ex-situ collections in the two main cacao germplasm banks of Ecuador (Figure 1): Estación Experimental Tropical Pichilingue (EET-P) and Centro de Cacao Aroma Tenguel (CCAT). Although the majority of the passport information was lost in the case of CCAT, these gene banks appear attractive to cacao breeders. Very little is known however, about the genetic structure of these collections. The understanding of their patterns of genetic diversity will help to optimize new breeding strategies to produce modern Nacional varieties with a conserved "ARRIBA" flavour.

Recent developments in molecular techniques provide a larger array of genetic tools for studying genetic diversity of populations. Microsatellite markers or simple sequence repeat (SSR) is one of the most powerful molecular-marker types to reveal the genetic diversity of natural populations. For cacao, a large number of microsatellite (SSR) markers was developed by Lanaud *et al* (1999); Risterucci *et al* (2000); Lanaud *et al* (2004) and Pugh *et al.* (2004). However, few genetic studies have been carried out on the Nacional cacao of Ecuador (Lerceteau *et al*, 1997b; Quiroz, 2002; and Loor, 2002); and, unfortunately a small number of samples have been used. In the present study, SSR

genotyping technology was applied to study the diversity of a large number of samples (a) to assess the genetic structure and diversity present in modern Nacional cacao, (b) to determine the genetic relationships between the different cacao accessions, and (c) to identify individuals that could represent the native Nacional cacao; and, which could be used as source of genes for a new cacao breeding program of aromatic varieties.

Materials and methods

Plant material

Leaves were collected from 322 cacao trees conserved at CCAT and EETP germplasm banks (Figure 1). The CCAT is presently managed by the UTEQ (Universidad Técnica Estatal Quevedo) and is located in the south coast region of Ecuador. The EET-P is located in the central coast region of Ecuador and it is managed by the Ecuadorian Agricultural Research Center (INIAP). One old cacao tree selected in a traditional farm (B-240) was used as reference for Nacional (Figure1). In addition, Moreover, four genotypes representing a typical Trinitario (UF-676) and its two probable ancestors: Lan-28 (Criollo) and Mat-1-6 (Lower Amazon) (Motamayor *et al* 2003); and, the resistant clone Scavina 6 (SCA6), a Forastero genotype originated from the Upper Amazon region of Peru, were included in this analysis.

DNA isolation

Adult fresh leaves were collected and stored at -20°C until DNA extraction was performed. Genomic DNA was isolated following the method described by Risterucci et al (2000) with some slight modifications. The DNA was purified by a silica-based anion-exchange resin as recommended by the supplier (NucleoBond® AX, Macherey-Nagel), and the concentration was estimated using a fluorometer (Fluoroskan Ascent 2.5) according to the manufacturer's instructions. The proper DNA work dilution was established at 0.2ng/µl.

PCR amplification and SSR analysis

PCR amplification and SSR analyses were carried out following the methods described by Risterucci *et al* (2000) and Pugh *et al* (2004). A total of forty SSR primers pairs (Table 1) were chosen to analyse the genetic diversity, based upon their position in the most recent cacao genetic map (Pugh *et al* 2004). Four loci per chromosome were selected. The PCR products were denatured and separated by electrophoresis in 0.5% TBE buffer at 60W for approximately 2h. The gel was dried for 25 min at 80°C and the results were revealed by autoradiography. SSR loci were scored individually and alleles were recorded by the presence of polymorphic DNA fragments (alleles) among the individuals of each population. Only those alleles that showed consistent amplification were considered. Smeared and weak alleles were ignored.

Data analysis

The SSR bands were scored as alleles. The genetic diversity was represented using multivariate analysis. A factorial analysis of correspondence (FAC) (Benzecri 1973) was carried out using GENETIX V.4.05 software (Belkhir *et al,* 2004) and a graphic representation of the variation among individuals was obtained (Dervin 1992; Phillips 1995). POWER MARKER V.4.03 (Liu and Muse 2005) was used to calculate the allelic and genotypic frequencies. The following parameters were also computed: 1) Nei's genetic parameters (Nei 1972; Nei 1978): genetic distance, observed heterozygosity (*Ho*) and gene diversity, often referred to as expected heterozygosity (*He*); 2) Genetic polymorphisms (Hart & Clark, 1997) was calculated by the effective number of alleles/locus (N_e); 3) F-statistics were measured at different hierarchical levels, according to Weir & Cockerhan (1984): Wright's F_{IS} (correlation of alleles within individuals of one population), F_{ST} (correlation of alleles between individuals of a population compared to the whole populations) and F_{IT} (correlation of alleles within individuals "inbreeding"). In addition, to identify the two most likely parents of each sample, paternity inference was conducted using CERVUS V.2.0 software (Marshall *et al.*, 1998), a Windows-based maximum likelihood program, designed for use with co-dominant markers. Paternity assignment using likelihood techniques was determined at, 95% statistical confidence level and with 10 000 simulated offspring as suggested by Marshall *et al* (1998).

Results

Level of polymorphism revealed in "modern" Nacional genotypes

Forty SSRs primers were used to reveal polymorphism among a subset of 332 "modern" Nacional cacao accessions. As shown in Table 1, genetic variability was detected for all loci in the natural cacao populations, covering most of the Pacific Coast Province in Ecuador. A total of 169 alleles were scored, ranging from 2 alleles at mtcCIR84 or mtcCIR258 to 7 alleles at mtcCIR136, mtcCIR290 and mtcCIR258. The frequency of the most common allele was also determined per locus and ranged from 0.36 (mtcCIR73) to 0.95 (mtcCIR242). High levels of gene diversity (He >0.50) were observed for 28 SSR markers among the 40 SSR studied (Table 1). Indeed, for most SSR loci, gene diversity (Nei's He) was apparently lower than the observed heterozygosity (Ho). As showed in Table 1, the polymorphism information content (PIC) ranged from 0.09 (mtcCIR242) to 0.68 (mtcCIR136); an example of this polymorphism profile can be observed in Figure 2.

Genetic structure and diversity in modern Ecuadorian cacao population

Based on polymorphic SSR, eight accessions showed a higher level of homozygosity (> 75%). Among these, four accessions showed 90% of homozygosity. These individuals, characterised by a higher level of homozygosity, will be called HoN in the further results or discussions reported in this paper.

No significant difference was observed (0.031) between the two collections from EETP and CCAT on the basis of Nei's genetic distance (1972). The genetic structure of the studied population was also analyzed in term of F statistics. A summary of genetic statistics values measured for each population and the overall sample is shown in Table 2. In each population, Wright's F_{IS} measured the heterozygous deficiency within each sample, and negative values were obtained in each group collected. Overall negative values were obtained for F_{IS} (-0.1393) and F_{IT} (-0.1169), in both cases, reflecting high values of heterozygosity level. The result of F_{ST} showed low level of differentiation among individuals of each population (F_{ST}= 0.0369).

A Factorial Analysis of Correspondence (FAC) gives a visual representation of the relationships and variation among individuals (Figure 3).

As showed in Figure 3, the two first axes of FAC explained 15% of the total variation. The modern Ecuadorian cacao accessions present a continuous variation and were widespread in an area delimited by the old Nacional genotype B-240, the Trinitario type UF-676 and the Lower Amazon ancestor MAT-1-6. This pattern of diversity may indicate the gene flow occurred between cultivated Nacional and foreign genotypes introduced at the beginning of the 20^{th} century.

The highly homozygous genotypes (HoN) previously identified were found at the extreme side of the hybrid population, and close to the Nacional reference B-240. These genotypes were circled at the left-hand end on axis 1 of FAC (Figure 3). The diversity between HoN individuals was very low, despite the fact that the trees have been collected from diverse geographic zones along the coast provinces of Ecuador. These HoN genotypes shared the lowest percentage of common alleles with the Trinitario type represented by UF676.

Nearly all HoN genotypes clustered together and with the Nacional reference genotype B240, showing a high genetic similarity among them, and the higher genetic divergence with the three reference genotypes representing the first potential foreign introductions: the Trinitario (UF676) and their two potential ancestors LAN28 and MAT1-6.

From our results, most of the cacao trees in the germplasm banks of EET-Pichilingue and CCA-Tenguel appeared as hybrids that mainly shared alleles with the typical Trinitario type UF-676 and with the HoN individuals. These results suggested that the HoN individuals identified in this study could represent the few remaining samples from the native Nacional variety cultivated in Ecuador before the foreign cacao introductions and that they could be at the origin of the modern and hybrid Nacional population. To check this hypothesis, a paternity analysis was carried out on all samples studied.

Identification of potential parents of modern hybrid Nacional genotypes

According to the history of foreign cacao introduction in Ecuador, approximately one century separates the first hybridization events of foreign cacao with the native Nacional variety from the modern Nacional population. Thus, probably no more than 4 generations of recombination occurred during this time. Using paternity inferences from Cervus software, we attempted to identify the potential parents of the genotypes that form the actual Nacional cacao population in Ecuador.

We carried out several preliminary paternity analyses to finally include a total of 77 potential progenitors of the modern Nacional cacao population analysed in this study, represented by:

- the three reference genotypes (Criollo, Lower Amazon Forastero and Trinitario);
- the HoN Nacional genotypes;
- SCA6 and the modern Nacional genotypes close to it, and, as distant from Trinitario as the HoN individuals;
- different hybrid genotypes involving crosses between HoN genotypes only, or crosses between HoN genotypes and Criollo, Trinitario or Lower Amazon Forastero.

Thus, the two most likely parents were identified for each individual. The first most likely parent was identified at 95% of confidence in the 78% of individuals and only for 2 individuals; however, paternity was unresolved.

The identification of two most likely parents allowed to clarify the genetic structure of the cacao population analyzed in this study. Several classes of crosses, probably occurred during the last century of Nacional cocoa history in Ecuador, could be identified according to the two most likely parents assigned for each individual from our paternity analysis:

- crosses between 2 HoN genotypes (22,91%)
- crosses between HoN x Foreign genotypes (3,94%)
- back crosses (HoN x Foreign genotypes) x HoN (11,48%)
- back crosses (HoN x Foreign genotypes) x Foreign genotypes (40,01%)
- crosses (HoN x Foreign genotypes) x (HoN x Foreign genotypes) (13,45%)
- crosses between only foreign genotypes (8,21%)

About 12% of the individuals, widespread in the several classes, involve SCA6 as one of their probable parent.

Our results showed clear evidences of high introgression of foreign alleles, principally from Trinitario type, in the modern Nacional population; and, confirm the hybrid nature of modern Nacional variety. Our results indicated also that the HoN cacao samples could be at the origin of more than 67% of the modern Nacional varieties, associated to various levels of introgression of foreign genotypes. They are also at the origin of 23% of the population which result only from crosses between HoN.

These results showed that different hierarchic levels of crosses have occurred during the last century in the Ecuadorian cacao plantation, from direct crosses between foreign genotypes and the native variety, to more advanced generation.

Discussion - conclusion

In this work, we used 40 SSR markers to estimate the genetic structure of modern Nacional cacao, represented by 322 cacao accessions. These cacao trees were originally sampled across their distribution in the Ecuadorian Coast Region and are maintained as living collections at EETP and CCAT germplasm banks. No significant level of genetic differentiation between these two collections was evidenced. As also demonstrated by F-Statistics, the genetic structure of modern Nacional cacao population is characterized by a high level of heterozygosity in agreement with the historical events of introduction of foreign germplasm, since the end of 1890's, and its subsequent gene flow with the native Nacional variety (Van Hall 1932; Fowler *et al,* 1949).

The modern Nacional genotypes presented a wide range of variations. Among them, a few highly homozygous genotypes (HoN) were also identified, in contrast with the hybrid nature of most of them. These HoN genotypes showed the lowest level of introgression of Trinitario alleles, but shared the same alleles, at most loci, between them; and, with the reference genotype of Nacional (B-240). Such highly homozygous individuals have been also identified by Lerceteau *et al,* (1997b) and Loor (2002) using RFLP and SSR marker, respectively. Using paternity analyses, these genotypes

appeared to be at the origin of more than 90% of the population studied. These HoN genotypes, among which some correspond to trees of more than 80 years old, could represent the native Nacional variety cultivated in Ecuador before foreign introductions, and now nearly lost among a wide hybrid cultivated pool.

Despite to the wide geographical area where the HoN genotypes were collected, they showed a high genetic similarity. This result could indicate a pronounced bottleneck effect of the original Nacional cacao plantations along of the Coast Provinces of Ecuador. In this way, we can suppose that at the origin of Nacional variety the size of the ancestral population was very small. The differentiation of this homozygous group may have occurred through allele fixation (genetic drift). Subsequently, intervention of man, through cultivation may have fixed and maintained some phenotypes, as hypothesized by Motamayor *et al*, (2002) for the Criollo group.

The higher level of homozygosity of these genotypes could be explained by an incompatibility system present in T. cacao but variable according to the different genetic origins. The Trinitario are generally self incompatible, but the oldest Nacional genotypes from Ecuador are self-compatible (Quiroz, 1990; Pastorelli, 1992; Enriquez, 1992; Loor, 1998). Hybrid forms appear, most likely, to be variable, according to the introgressed part of the Trinitario genome.

Paternity analyses allowed us to refine the genetic structure of modern Nacional varieties. From our results, we estimated that more than 65% of the cacao population analyzed in this study correspond to crosses of a second generation. It is important to indicate that almost the totality of cacao accessions from CCATenguel were collected in 1950's, by its similarity with the Nacional variety. Currently, the majority of traditional cacao plantations are most likely to be established with crosses of fourth or fifth generation.

The course of continuous germplasm introduction has caused considerable genetic erosion within original Ecuadorian Nacional cacao, occasioning changes in its allele frequencies and in its fine cocoa flavour, which has traditionally characterized the Ecuadorian cocoa beans. Nowadays, the number of native trees of Nacional variety has decreased alarmingly and this variety will likely to disappear in the near future (Loor, 2002). Subsequently, the "fine flavour" quality of Ecuadorian cocoa, economically

important for this country, has also degraded. Until now Nacional cocoa has been classified as "fine flavour" and benefited from higher, premium prices in the international market.

As consequence of the introduction of foreign germplasm, presently 25% of the Ecuadorian cacao production has been disqualified as fine flavour and considered as "bulk cocoa" with a lower price in the international market. Additionally, to compound these problems, a very limited number of Nacional accessions have been used in traditional cacao breeding programs in Ecuador, due to the utilisation of foreign varieties, providing disease resistance and higher productivity.

The present study provides important information for long-term conservation and management of *T. cacao* in Ecuador with emphasis on the Nacional variety. If the proportion of specific alleles from Nacional variety continues decreasing, the consequences will be very negative for the "fine flavour" quality of Ecuadorian cocoa. The identification of non introgressed Nacional genotypes will facilitate the improvement of "fine flavour" in new varieties.

The large number of diverse hybrid forms between Nacional and Trinitario genotypes currently present in the Ecuadorian germplasm banks, offer a segregation population with continuous variations between them, adapted to carry out association mapping studies to identify the genetic bases of "fine cocoa flavour" of Nacional varieties. This knowledge will facilitate the improvement of new varieties associating Nacional quality traits and disease resistance, provided by some foreign cacao genotypes using a marker assisted strategy.

The origin of Nacional variety cultivated in Ecuador from 1600 is currently unknown. The identification of wild Nacional cacao populations at the origin of still existing Nacional varieties, could allow for the enlargement of the narrow genetic base of non introgressed Nacional genotypes, identified in this study, favouring the breeding strategies for the modern Nacional varieties.

Acknowledgments

We are grateful to the United States Department of State (U.S. Foreign Ministry), the U.S. Embassy, Quito; and the U.S. Department of Agriculture (USDA-ARS), for the financial support of this research. Thanks also should go to the Centre de Coopération International en recherche agronomique pour le développement (CIRAD), France; and, the Agricultural Research Center of Ecuador (INIAP) for their financial and technical support.

References

Allen BJ, Lass AR (1983). London cocoa trade Amazon project. Final report phase 1. Cocoa Growers' Bulletin 34:1-72.

Ampuero E (1960). Progresos alcanzados en el Ecuador en el estudio de selección para resistencia de la Escoba de Bruja. *In* Inter American Cacao Conference (Trinidad and Tobago). Proceeding. Trinidad, Bowen Press. P166-173.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations, Laboratoire Génome, Populations, Interactions; CNRS UMR 5000; Université Montpellier II, Montpellier (France).

Benzecri JP (1973) L'analyse des données. Tome 2 : L'analyse des correspondances. Dunod (eds), Paris, France.

Braudeau J (1970) El cacao. Técnicas agrícolas y producciones tropicales. Blume. Barcelona, España. 207p.

Cheesman EE (1944) Notes on the nomenclature, classification and possible relationships of cacao populations. Trop Agric 21:144-159

Dias LAS (2001) Melhoramento genético do cacaueiro. Viçosa:FUNAPE.

Dias LAS, Barriga PJ, Kageyama YP, Almeida CV (2003) Variation and its distribution in wild cacao populations from the Brazilian Amazon. Brazilian Archive of biology and technology.46:507-514

Dervin C (1992) Analyses des Correspondances. Comment interpréter les résultats?. Collection STAT-ITCF. Paris, France.

Enriquez GA (1992) Characteristics of cacao "Nacional" of Ecuador. International workshop on conservation, characterisation and utilisation of cocoa genetic resources in the 21st century. Port of Spain, Trinidad 13-17th September. The Cocoa Research Unit, The University of the West Indies, pp 269-278.

Fowler RL and Lopez GH (1949). *The Cacao Industry of Ecuador*. Foreign Agriculture Report N°34, US Dept of Agriculture, Washington.

Harland, S. C. 1925. Studies in Cacao. The method of pollination. Ninth West Indian Agr. Conf. Proc. Kingston, Jamaica, 1924: 61 - 69.

Hartl DL, Clark AG (1997) Principles of Population Genetics, 3rd edn. Sinauer Associates, Sunderland, MA

Hasenstein KH and Zavada MS (2001). Auxin modification of the incompatibility response in *Theobroma cacao*. *Physiologia Plantarum* **112**:113-118

Hurst WJ, Tarka SM, Powis TG, Valdez F, Hester RT (2002) "Archaeology: cacao usage by the earliest Maya civilization." Nature 418, 289-290.

Juan A, Crespo BM, Cowan SR, Lexer C, Fay FM (2004) Patterns of variability and gene flow in *Medicago citrine*, an endangered endemic of islands in the western Mediterranean, as revealed by amplified fragment length polymorphism (AFLP). Molecular Ecology 13:2679-2690.

Lanaud, C., Risterucci, A.M., Pieretti, I., Falque, M., Bouet, A., Lagoda, PJL (1999) Isolation and characterisation of microsatellites in *Theobroma cacao* L. *Molecular Ecology* 8: pp 2141 – 2152

Lanaud, C., Risterucci, A.M., Pieretti, I., N'Goran JAK, Fargeas D (2004) Characterization and genetic mapping of resistance and defence gene analogs in cocoa (*Theobroma cacao* L.). Mol Breeding 13: 211-227

Lerceteau E, Robert T, Pétiard V, Crouzillat D (1997a) Evaluation of the extent of genetic variability among *Theobroma cacao* L accessions using RAPD and RFLP markers. *Theoretical and Applied Genetics* 95:10-19

Lerceteau E, Flipo S, Pétiard V, Crouzillat D (1997b) Genetic differentiation among Ecuadorian *Theobroma cacao* L. accessions using molecular and morphological analyses. Euphytica 95: 77-87.

Liu K, Muse VS (2005) Power Marker: Integrated analysis environment for genetic marker data. Bioinformatics 21(9): 2128-2129Lockwood G (1976) A comparison of the growth and yield during a 20 year period of Amelonado and Upper Amazon hybrid cocoa (*Theobroma cacao*) in Ghana. Euphytica 24(3):647-658.

Loor RG (1998) Obtención de híbridos de cacao tipo Nacional provenientes de materiales de alta productividad y resistentes a enfermedades. Tesis de Ingeniero Agrónomo. Universidad Técnica de Manabí, Portoviejo, Ecuador. 67p

Loor RG (2002) Caracterización morfológica y molecular de 37 clones de cacao (*Theobroma cacao* L) Nacional de Ecuador. Tesis de Maestro en Ciencias. Colegio de Postgraduados, Montecillo, texcoco, Edo de México. 96p

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, **7**, 639-655.

Moreno M (1970). Determinación de la autocompatibidad y la compatibilidad cruzada de sesenta clones de cacao. Tesis In. Agr. Universidad de Guayaquil, Ecuador. 23p.

Motamayor JC, Risterucci AM, Lopez PA, Ortiz CF, Moreno A, Lanaud C (2002). Cacao domestication I: the origin of the cacao cultivated by the Mayas. Heredity **89**: 380-386.

Motamayor JC, Risterucci AM, Heath M, Lanaud C (2003). Cacao domestication II: progenitor germplasm of Trinitario cacao cultivar. Heredity **91**: 322-330.

Nei, M. (1972). Genetic distance between populations. *Am. Nat. 106*: 283-292.

Nei M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583–590.

N'Goran JAK, Laurent V, Risterucci AM, Lanaud C (2000). The genetic structure of cocoa populations (*Theobroma cacao* L.) revealed by RFLP analysis. *Euphytica* **115**:83-90

Pastorelly RD (1992) Evaluación de algunas características del cacao tipo Nacional, en la zona de Tenguel. Tesis Ing. Agr. Universidad Agraria del Ecuador. Guayaquil, Ecuador. P7.

Phillips D (1995). Correspondence analysis. Social Research Update. Department of Sociology, University of Surrey, Guildford GU2 7XH, England.

Pound FJ. (1938). Cacao and Witches' broom disease (*Marasmius perniciosus*) of South America, with notes on other species of *Theobroma*. Yuille's Printerie, Port of Spain, Trinidad and Tobago, 9–49. Reprinted 1982, Archives of Cocoa Research 1:21–64.

Pound FJ (1945). A note about the cacao populations of South America. *Report and Proceedings Cocoa Research Conference, London*. Colonial 192:95-7. (reprinted 1982 in *Arch. Cocoa Res* **1**:93-97).

Pugh T, Fouet O, Risterucci AM, Brottier P, Abouladze M, Deletrez C, Courtois B, Clement D, Larmande P, N'Goran JAK, Lanaud C (2004) A new cacao linkage map based on codominant markers: development and integration of 201 new microsatellite markers. Theor Appl Genet 108:1151-1161

Quiroz J (1990) Estudio de la compatibilidad en algunos cultivares de cacao (Theobroma cacao L.). Tesis de Ing. Agr. Universidad Técnica de Babahoyo. Babahoyo, Ecuador. 30p

Quiroz J (2002) Caracterización molecular y morfológica de genotipo superiores de cacao nacional (Theobroma cacao L.) de Ecuador. Tesis Ms. Sc. Turrialba, CR, Centro Agronómico Tropical de Investigación y Enseñanza. 111 p.

Risterucci AM, Grivet L, N'Goran JAK, Pieretti I, Flament MH and Lanaud C. (2000) A high-density linkage map of *Theobroma cacao* L. *Theoretical and Applied Genetics* 101: pp 948 – 955

Rorer JB (1926). Ecuador cacao. Tropical Agriculture (Trinidad) 3:69-69.

Soria JV (1970). Principal varieties of cocoa cultivated in tropical America. Cocoa Growers' Bulletin 19:12-21.

Van Hall CJ (1932). Cocoa. Second edition. MacMillan & Co., Ldt., London. pp.304-320.

Vera BJ (1969). Estudio de la compatibilidad en híbridos interclonales de cacao (*Theobroma cacao* L.). Tesis de Ingeniero Agrónomo. Universidad de Guayaquil, 30p.

Vera BJ (1987) Antecedentes históricos. Manual del cultivo de cacao. INIAP, EET-Pichilingue. Quevedo, Ecuador, pp: 6-9.

Vera BJ (1993) Botánica del cacao. Manual del cultivo de cacao 2da edición. INIAP, EET-Pichilingue. Quevedo, Ecuador, pp: 10-15.

Weir BS, Cockerhan CC (1984). Estimating F-statistics for the analysis of population structure. Evolution, 38, 1358-1370

FIGURES

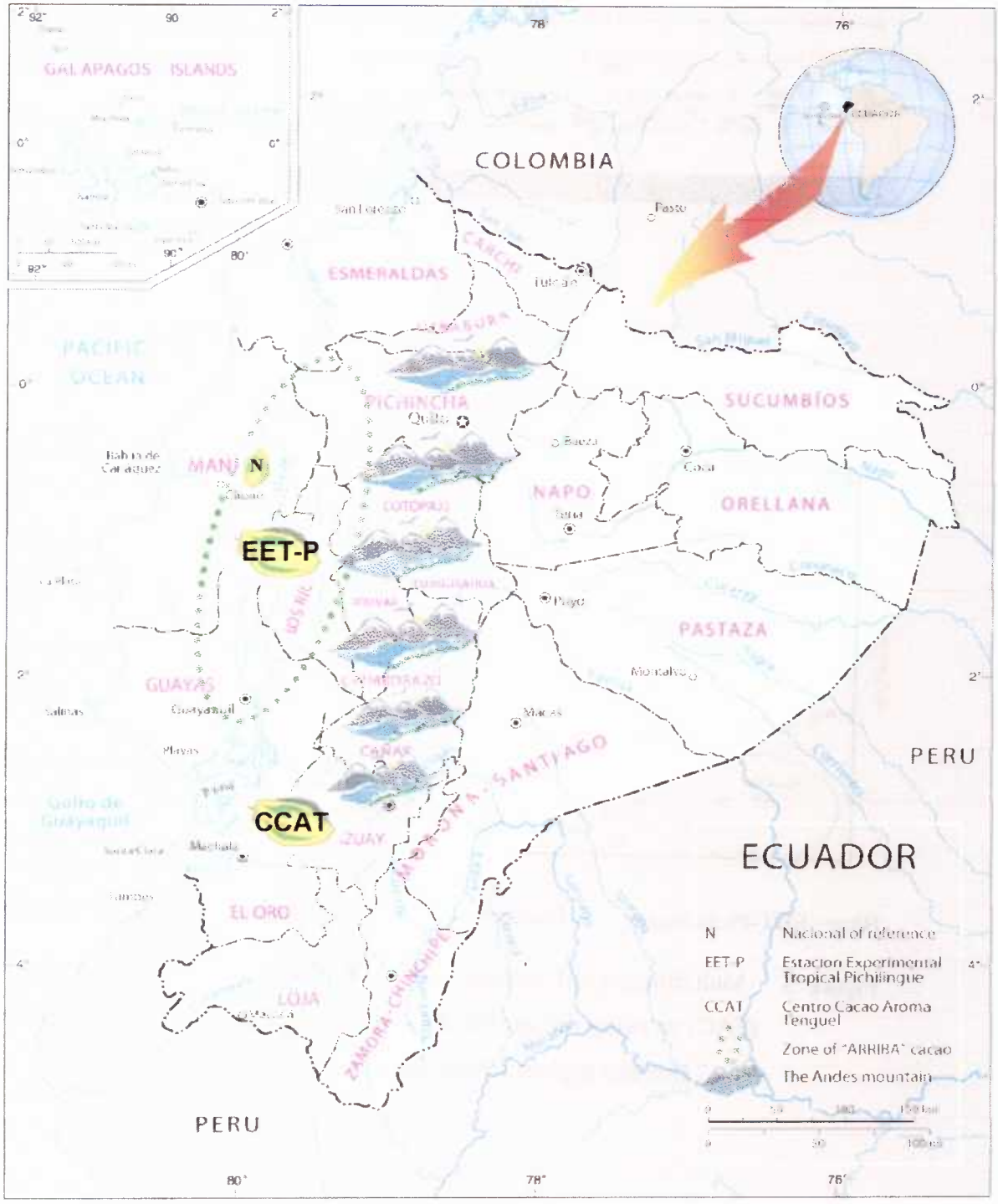


Figure1. Map of Ecuador showing the different ecological zones, the geographic locations of Nacional germplasm banks and the zone called ARRIBA where was cultivated the first *Nacional* cacao population in Ecuador.

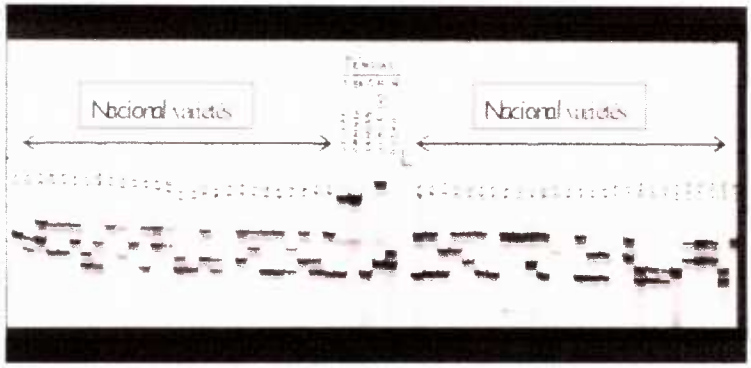
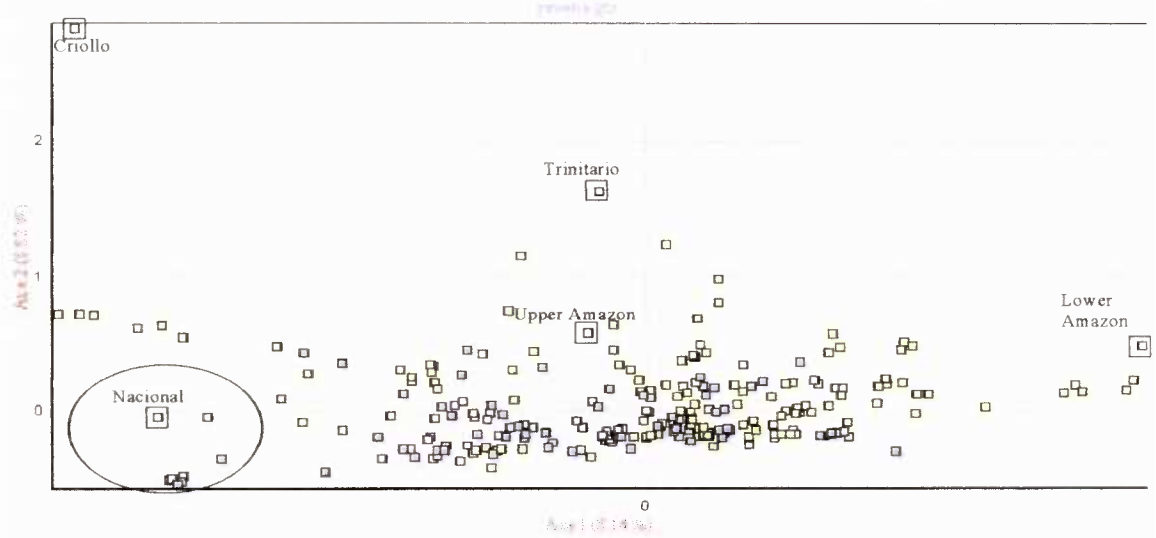


Figure 2. Polymorphism profile of some Ecuadorian samples on chromosome 6 using mtcCIR 136.



Blue= EET-Pichilingue Boxed= Reference Genotype yellow= CCATenguel

Figure 3. Multidimensional scaling plot of Factorial Analysis of Correspondence (FAC) showing the genetic diversity among the Ecuadorian cacao samples. The 2 first axe explain 15% of the total variation.

Table 1. Summary of genetic parameters values for the 40 microsatellite loci evaluated on the cacao populations collected in the coast region from Ecuador. Allele parameters, Ne: number of allele per locus; Fa: frequency of most common allele. Genotype parameters, Ng: number of effective genotype; Fg: frequency of most common genotype). Gene diversity, He: Nei's expected heterozygosity; Ho: observed heterozygosity, and PIC: polymorphism information content.

SSRs*	Allele		Genotype		Gene diversity	Ho	PIC
Locus	Ne	Fa	Ng	Fg	(He)		
mtcCIR15	6	0.5729	11	0.5186	0.5492	0.6475	0.4678
mtcCIR84	2	0.9476	3	0.9021	0.0994	0.0909	0.0945
mtcCIR275	3	0.5859	6	0.5219	0.5408	0.6768	0.4597
mtcCIR94	5	0.5530	12	0.5166	0.5780	0.7119	0.5052
mtcCIR230	4	0.5932	7	0.6122	0.4960	0.6274	0.3889
mtcCIR268	5	0.3848	12	0.2695	0.6839	0.6680	0.6245
mtcCIR100	5	0.4860	8	0.5979	0.5635	0.7343	0.4671
mtcCIR73	4	0.3639	10	0.3836	0.6887	0.7410	0.6282
mtcCIR82	4	0.9400	6	0.8867	0.1143	0.1067	0.1107
mtcCIR167	4	0.5214	7	0.6335	0.5108	0.6512	0.3925
mtcCIR175	5	0.4966	12	0.5593	0.5613	0.6678	0.4657
mtcCIR204	4	0.8704	7	0.7807	0.2287	0.1827	0.2081
mtcCIR57	3	0.5419	6	0.5940	0.5289	0.6644	0.4241
mtcCIR107	3	0.5201	6	0.4818	0.6040	0.7445	0.5306
mtcCIR32	3	0.4850	6	0.6113	0.5360	0.6545	0.4277
mtcCIR242	3	0.9521	5	0.9175	0.0922	0.0693	0.0901
mtcCIR109	5	0.4946	8	0.6304	0.5342	0.6957	0.4255
mtcCIR267	3	0.8490	4	0.7083	0.2570	0.2813	0.2248
mtcCIR265	6	0.4443	11	0.3716	0.6473	0.7500	0.5766
mtcCIR80	3	0.5261	6	0.5765	0.5450	0.6678	0.4453
mtcCIR238	4	0.5811	9	0.4570	0.5526	0.5894	0.4767
mtcCIR136	7	0.3656	17	0.2143	0.7309	0.7755	0.6839
mtcCIR291	6	0.5286	13	0.5724	0.5398	0.6397	0.4385
mtcCIR290	7	0.4785	12	0.5033	0.6192	0.6490	0.5443
mtcCIR56	5	0.7374	9	0.5863	0.3996	0.3094	0.3381
mtcCIR93	5	0.4710	12	0.4539	0.6391	0.7235	0.5716
mtcCIR110	3	0.5446	6	0.3927	0.5997	0.8152	0.5328
mtcCIR186	3	0.5524	5	0.5804	0.5037	0.5874	0.3870
mtcCIR163	3	0.5409	4	0.6548	0.4983	0.6584	0.3759
mtcCIR189	7	0.4589	15	0.3980	0.6071	0.6020	0.5280
mtcCIR225	4	0.4908	8	0.4890	0.5954	0.7096	0.5142
mtcCIR258	2	0.8221	3	0.6577	0.2924	0.3289	0.2497
mtcCIR30	3	0.5964	6	0.5179	0.5189	0.5964	0.4290
mtcCIR79	4	0.5212	9	0.3497	0.6177	0.6928	0.5519
mtcCIR157	4	0.5470	9	0.6342	0.5160	0.6678	0.4054
mtcCIR58	5	0.5033	13	0.3679	0.6435	0.7124	0.5867
mtcCIR31	3	0.8364	4	0.6909	0.2743	0.2909	0.2377
mtcCIR91	3	0.6071	4	0.5170	0.4861	0.5408	0.3790
mtcCIR220	5	0.8163	9	0.7789	0.3178	0.1020	0.2965
mtcCIR229	6	0.5447	10	0.5704	0.5416	0.6735	0.4452
Σ	169			330			
Mean	4.2256	0.5918		8.25	0.4963	0.5675	0.4232

* More information on the mapped SSRs markers used in this study can be found in Lanaud *et al* (1999), Risterucci *et al* (2000) and Pugh *et al* (2004).

Table 3. Summary of genetic parameters

	CCAT	EETP	Overall
He	0.4853	0.4805	0.4829
Ho	0.5922	0.5255	0.5588
P(0.95)	0.9500	0.9750	0.9625
P(0.99)	1.0000	1.0000	1.0000
Alleles/locus	3.6500	4.2000	3.9250
Fis	-0.2181	-0.0604	-0.1393
Fit	x	x	-0.1169
Fst	x	x	0.0369

This work has been published in the following paper:

Loor Solorzano R. G., Risterucci A.M., Courtois B., Fouet O., Jeanneau M., Rosenquist E., Amores F., Vasco A., Medina M., Lanaud C. (2009). Tracing the native ancestors of modern *Theobroma cacao* L. population in Ecuador. *Tree genetics and genomes*, 5 (*3*) : 421-433. [20100330]. http://dx.doi.org/10.1007/s11295-008-0196-3.

2) Insight about the wild origin, migration and domestication events of the fine flavour Nacional *Theobroma cacao* L. variety from Ecuador

(paper in preparation)

Loor R.G., Lemainque, A., Fouet O., Pavek S., Argout X., Wincker P., Risterucci A.M., Boccara M., Amores F., Courtois B., Lanaud C.

ABSTRACT

The Ecuador's economic history has been strongly linked to the ***Theobroma cacao*** **L** culture, specifically to a native variety known as "Nacional". Due to foreign germplasm introductions the original Nacional cacao trees are presently considered in danger of extinction. In a previous work, a few non-introgressed "Nacional" types were identified as potential founders of the modern Ecuadorian cacao population, but up to now, their origin could not be formally identified. We used 80 microsatellite markers to analyse the relationships between these potential Nacional founders and 13 cocoa groups representing a wide range of wild and cultivated genotypes from South and Central America. The main objectives were to identify the putative centre of origin of Nacional and its wild potential ancestors; and, to trace the domestication history of Nacional cocoa variety.

A high genetic similarity was observed between the Nacional pool and some wild genotypes from the south Amazonian region of Ecuador; this result was confirmed by the proportion of shared alleles and a parentage analysis. The wild genotypes that showed the highest parentage relationships with the Nacional pool were sampled along the Yacuambi, Nangaritza and Zamora in the Province of Zamora Chinchipe.

Based on our results, and on data about pre-Columbian civilization and Spanish colonization history of Ecuador, we traced for the first time the possible centre of origin and migration events of Nacional variety, since the Amazonian area until its arrival to the Coast Provinces.

The results of this paper could guide us towards precious genetic resources suitable to enlarge the genetic diversity necessary to improve both the flavour and disease resistance in modern Ecuadorian cacao varieties. Large unexplored forest areas still exist at south of Ecuadorian Amazonian region, and must be considered for new cacao collections.

Introduction:

***Theobroma cacao* L.** is the most important economic species of the *Theobroma* genus, with a particularly long history in Ecuador. The cacao cultivation in this country is very ancient and dates back to the pre-Columbian age, before the colonization of its territory by the Spanish. This was evidenced by Pizarro during his first travel in 1526 on the South American coasts that nowadays are Ecuador, finding small plantations of an apparently native cacao (Steveson 1829; Prescott 1847). Admittedly, the population of Ecuadorian cacao named Nacional must have existed for several centuries prior to the arrival of the Europeans (Bartley, 2005), but its origin in the Ecuadorian littoral has never been known. Two hypotheses have been advanced: according to Allen and Lass (1983) the Nacional cacao variety could have originated from a local wild population, that has nowadays completely disappeared together with the original forest cover of the region; according to Pound (1938), Soria (1970) and Vera (1987) the Nacional could have been introduced to the coast from the Amazonian area of Ecuador where wild cacao is common. There is no evidence that *T. cacao* and its product played any part in the lives of the inhabitants of the Ecuadorian coast during the Chorrera phase of the pre-Columbian Valdivia culture (2000BC), in contrast to extensive information of such use for the Meso American people, most notably the Mayas and Aztecas. In Ecuador, the available information, related to the history of Nacional cacao cultivation, is associated to the Spanish colonization history in this country.

Since the moment when the first Spanish colonists initiated the deforestation of the Ecuadorian coast region, a large number of native cocoa trees have been reported (Crawfort 1980, Allen and Lass 1983) principally along the Guayas basin. Apparently, these colonists began to sow seeds of these native trees approximately 100 years after the discovery of America, and when the native meso-American populations started to decline. The cacao cultivated areas expanded, and rapidly the native Nacional variety became worldwide known due to a strong floral aroma called "ARRIBA", exclusively produced by its cocoa beans. From the beginning, the quality produced by the Nacional cocoa beans has been highly appreciated by chocolate manufacturers to make fine flavour chocolates.

The native "Nacional" variety was the only one planted in Ecuador until the early 1890s, when foreign germplasm began to be introduced in this country. In 1890, due to the quality traits of the Nacional cocoa beans, Ecuador had a privileged position in the markets of Hamburg and London (Preuss, 1901). From 1910's, foreign germplasm introductions progressively increased due to the appearance of two fungus diseases known as witches' broom (***Moniliophthora perniciosa***) and frosty pod (***Moniliophthora roreri***) that together devastated the native plantations.

Presently, a large genetic admixture between the native variety and foreign germplasm can be observed within the modern Ecuadorian cacao plantations (Loor *et al*, 2009). The fine flavour cocoa aroma has decreased in this hybrid complex. However, reports of the International Cocoa Organization (ICCO) indicate, that until now, Ecuador continues being the main supplier of the fine flavour cocoa in the world. Its production in the year 2006 reached 60% of the offer in this special quality market (ICCO, 2006). This situation brings a competitive advantage for Ecuador in the international market of the fine flavour cocoa, where the prices for quality are superior to those from the current cocoa. There is nowadays an increasing demand of fine flavour cocoa, which presently, does not exceed 5 % of the world cocoa production.

In a previous work (Loor *et al*,) a few non-introgressed "Nacional" types were identified within this hybrid population as potential representatives of the native Nacional variety. Until now, however, the origin of this Nacional variety could not be formally identified. In this paper, we analyze the relationships between the non introgressed Nacional genotypes identified in Loor *et al* (2009) and a wide range of wild cocoa genotypes covering an extensive geographical spectrum from upper and lower Amazonian regions (Pound 1938, 1945; Allen and Lass 1983; Lachenaud and Salle 1993; Lachenaud et al 1997): a) to identify Nacional putative centre of origin and wild potential ancestors, and b) to trace the domestication history of Nacional cacao variety. Representative individuals of the native Criollo (Motamayor, 2001), first domesticated variety in Central America by meso Amerindian population, were also included in this work to analyse their relationship with the wild Amazonian cocoa population.

Materials and methods

Plant material

A total of 176 individuals representing 14 cocoa groups were used for SSR analyses in this work (Table 1):

- Seven cacao individuals identified as putative ancient Nacional cacao variety cultivated along the Ecuadorian coast region (Loor *et al,* 2009).
- Sixty-five wild genotypes collected by Allen and Lass (1983) at north, centre and south from the upper Amazonian region of Ecuador (Figure 1)
- Sixty-six wild genotypes collected by Pound (1938, 1945) in the upper Amazonian region of Peru (Figure 1)
- Eleven wild cacao samples collected by Lachenaud and Sallée (1993) in the lower Amazonian area from Guyana.
- Three cultivated cacao genotypes from lower Amazonian region of Brazil.
- Twenty-four Criollo variety samples collected by Motamayor (2001) in Central America from Venezuela to Mexico were also used in this study.

The non Ecuadorian cacao samples were selected from genebanks located at Cacao Research Unit (CRU) in Trinidad; and, Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD) in Montpellier, France. The wild cacao accessions from Ecuador were selected from the CRU living collection on the basis of the geographical area sampled by Allen and Lass (1983) along of the Amazonian Provinces (Figure 1). The Nacional genotypes were selected from the living collection bank located at EET-Pichilingue (Ecuador).

Molecular markers

Eighty SSR markers isolated from EST and recently mapped (Fouet et al., 2011) were chosen for these analyses. They are widespread in all the 10 cocoa chromosomes (Table 2)

DNA isolation and PCR amplification

DNA samples were isolated from each individual as described by Risterucci *et al* (2000). Eighty microsatellite markers were used for the analysis (Table 2). PCR amplifications were performed as reported by Risterucci *et al* (2000) and fragments were subsequently detected using a "Megabase 1000" DNA analysis system (Molecular Dynamics/Amersham Life Science).

Statistical analysis

Amplified SSR fragments were scored as alleles. The following genetic parameters were calculated using Power Marker (Liu and Muse 2005) and Genetix (Belkhir et al. 2004) software: allele frequencies, expected and observed heterozygosity; and, proportion of polymorphic loci.

In a first approach, a Principal Coordinates Analysis (PCoA) was carried out to assess the pattern of the genetic diversity among groups. The first two axes of the PCoA plot allowed to visualise the relationships among cacao samples. In a subsequent analysis, a dendrogram was constructed using a dissimilarity index (simple matching) and the neighbour-joining (NJ) method. Both, principal coordinate's analysis (PCoA) and neighbour-joining (NJ) cluster analysis were carried out using Darwin sofware-5.0 (Perrier *et al*, 2003).

In addition, analysis of genetic distance among pairs of populations (Nei's 1972) and proportion of shared alleles were carried out to determine the genetic similarity between the Nacional pool and the wild and cultivated genotypes. Finally, a paternity analysis was performed using CERVUS software, version 2.0 (Marshall *et al*, 1998) to identify the most likely ancestral population of the Nacional variety. An 80% confidence threshold was used for paternity assignment.

Results

Patterns of genetic diversity among cocoa accessions.

A total of 463 alleles were detected at the 80 analysed SSR loci across all 176 individuals sampled (Table 2). A wide range of allele number was generally observed

for each locus (from 2 to 15), with an average of 5.78 alleles per locus. In each group, genetic diversity parameters were evaluated. As shown in Table 3, the mean number of observed alleles per locus and per genetic group corresponding to wild accessions ranged from 1.33 (Guy) to 4.45 (LCT-EENa). The most polymorphic wild cocoa accessions were generally found in the north Amazonian region of Ecuador (data not shown).

The Principal Coordinates Analysis (PCoA) produced a two dimensional representation of the genetic relationships between the cacao samples. The first two axes of the PCoA, based on SSR data, accounted for 42.7% of the total variation (Figure 2). As shown in Figure 2, the cacao samples could be separated in three different groups: (i) the representatives of Criollo forming a very compact and isolated group, placed in an extreme position in the upper left corner side of the PCoA plot; (ii) a second cluster formed with the majority of Peruvian accessions on the right quadrants. This second group includes also some accessions from the Upper Amazonian region of Ecuador (LCT-EEN) and all accessions from Lower Amazonian and Orinoco regions (Venezuela, Brazil and Guyana). A last cluster (iii) placed in the left side of the first axis contained the majority of the wild accessions from Ecuador and some wild accessions from Peru (SCAs and Morona river). The representative individuals of the native Nacional variety were also found within this last cluster. Some accessions from the South Amazonian region of Ecuador (LCT-EENc) were found very close to the Nacional pool.

In a subsequent analysis, a neighbour-joining tree (Figure 3a) calculated from simple matching allelic dissimilarity index, clustered the majority of cacao samples according to their geographical origins. Three main clusters could be observed in this analysis (Figure 3a):

- Cluster n°1 is divided into 2 subgroups : one sub group (1a) includes most of the accessions collected in the North and Centre of Amazonian regions from Ecuador; the other subgroup (1b) includes all Criollo accessions.
- Cluster n°2 included most of Peruvian accessions. It is also divided into 2 sub groups: One subgroup (2b) pooled the Scavina accessions; the other group (2a) gathered the remaining Peruvian accessions with wild accessions from French Guyana and Venezuela, and cultivated Lower Amazon Forastero samples from Brazil.

- Cluster n°3 in Figure 3a (amplified in Figure 3b), included the Nacional cocoa accessions, the wild cocoa genotypes from the southern part of Ecuadorian Amazonia and the wild cocoa accessions from Peru collected along the Morona river.

The NJ tree (Figure 3a) confirmed the pattern of relationship previously observed in Figure 2 between the Nacional pool and the wild cocoa genotypes (LCT-EEN) sampled in the Amazonian region C from Ecuador. The accessions sampled in Peru, along the Morona river were also genetically close to the Nacional. The Morona river region is geographically close to region C of Ecuadorian Amazonia. Compared to the genetic diversity revealed by Peruvian populations, a widest genetic diversity was found in the LCT-EEN accessions from Ecuador (Figure 2 and 3).

Genetic distance between the native Nacional genotypes and wild populations present in the Amazonian regions

To refine the previous analysis, coefficients of genetic distance were calculated for pair-wise comparisons of the 14 cocoa groups (Nei 1972). A matrix of genetic distance values for all groups is presented in Table 4. The highest genetic distance (2.445) was obtained between the Criollo and Lower Amazon group from Brazil (BA), while the most similar populations were LCT-EENa and LCT-EENb with a genetic distance of 0.039. In the case of the Nacional pool, the lowest genetic distance was evidenced with the groups from LCT-EENc region (0.250) and Morona (0.307).

Subsequent analyses of genetic distance was carried out to identify the genetic distances between the Nacional pool and wild cacaos collected along the neighbouring rivers within each Amazonian region of Ecuador (a, b, c) including also all accessions collected in the upper Amazonian region from Peru (data not shown). This new approach allowed to identify a few LCT-EENc samples collected along the adjacent rivers (Nangaritza, Yacuambi and Zamora) as the wild cocoa trees genetically closer to the Nacional pool cultivated in the coast region. An important characteristic of the wild cocoa trees from these adjacent rivers is their low level of heterozygosity. Amazonian region C is geographically close to Guayas basin (Figure 1) where the first Nacional cocoa plantations were established.

Potential wild ancestors of the Nacional variety

The genetic similarity between Nacional individuals and the wild genotypes from Amazonian region was analysed on the basis of proportion of shared alleles and parentage relationship (Table 4). Our results indicated that the wild genotypes coded as LCT-EEN85, LCT-EEN86 and LCT-EEN91 had the strongest parentage relationship with the Nacional variety pool from coast region, suggesting that the individuals representing the native Nacional pool most probably descended from wild cacaos adjacent to rivers Nangaritza, Yacuambi at south of Amazonian region "C". Thus a clear consistency with our previous results has been obtained. These three LCT-EENc samples could be representatives of the putative wild cacao population at the origin of the Nacional variety pool.

Discussion-Conclusion

In the course of cacao cultivation history in Ecuador, substantial genetic changes, including reproductive behaviour, have occurred as a consequence of foreign germplasm introduction and genetic admixture, associated to natural and human selection. The first study of genetic diversity of modern Ecuadorian cacao accessions was undertaken by Lerceteau *et al* (1997); however, studies aiming at identifying the origin of Nacional cocoa variety and assessment its genetic relationships with wild cocoa trees from Amazonian regions have not been conducted up to now. Patterns of domestication, colonization or geographical distribution, have been reported in many other crop species (White et al. 1999; Gaudeul et al. 2000; Vijverberg et al. 2000; Juan et al. 2004). In *T. cacao* Motamayor et al. (2002) studied the domestication of Criollo, a cocoa variety originally cultivated by the Mayas in Central America, and concluded at a very narrow genetic basis of the 'ancient" Criollo cultivated in Central America before cocoa foreign introductions during the 18th century.

In this study, we used molecular markers to try to identify the wild genetic origin and elucidate the domestication events at the origin of the first Nacional plantations established in Ecuador in the Guayas Basin. The Amazonian area from South America was considered as a potential source of wild material from which the Nacional cocoa population cultivated in the Ecuadorian littoral could have derived.

In this study, the diversity of wild accessions is divided into 3 main groups structured according to their geographic origins.

• One of the groups includes accessions from the north and the central region of Ecuadorian Amazonia. The cultivated Criollo accessions are also included in this group in spite of a high level of differentiation with the other members of this group. All Criollo accessions collected from Mexico to Venezuela clustered together in a very compact group, reflecting its narrow genetic base. In a previous study, Motamayor et al (2002) identified some accessions from Colombia (EBC) as the closest accessions to Criollo. Colombia has been suggested as the centre of origin of Criollo. In our study, we did not have access to Colombian accessions, nevertheless, the genetic similarity observed between Criollo and accessions from the north Amazonian region of Ecuador show the same geographic trend than the results obtained by Motamayor et al. (2002).

• Another group includes some accessions from Ecuador and most of wild Peruvian accessions, mainly those collected in the northern region of Peru, associated to wild accessions collected in French Guyana and in the Orinoco region of Venezuela. The cultivated genotypes from Lower Amazonia in Brazil are also included in the group.

The Upper Amazon region has always been considered as the primary centre of origin and diversity of *T. cacao* L. species. The high genetic similarity observed between some wild cocoa accessions from upper Amazonian region (Parinari or Pound) and the group from Guyana, suggest that during *T. cacao* evolution, the extension of cocoa populations towards the eastern part of South America could have started from these upper Amazonian regions.

• Another group includes most of the accessions of the south Amazonian region of Ecuador, the accessions collected in Peru along the Morona river, close to this south Ecuadorian region, and the Nacional accessions. In this group, a high genetic relationship was found among some wild accessions from Region C and the Nacional pool from the coast region. These results are supported by the evaluation of genetic distance, the proportion of shared alleles, as well as, by a parentage analysis. From our results, we hypothesize that the Amazonian region located in the Province of Zamora Chinchipe could be the centre of origin of the first Nacional cacao plantation observed on the Guayas river banks. Later, from this last place, the expansion of the cacao cultivation along Guayas tributary rivers (upper waters), could have been initiated,

materializing nowadays the producing zone known as "ARRIBA" among the Province of Guayas, Los Ríos and Manabí (Figure 1).

This hypothesis of migration and domestication event is supported by the geographical proximity between the first cocoa plantation established in the Guayas river banks (Preuss, 1901) and the wild cacao collected along of the Nangaritza, Yacuambi and Zamora rivers banks, in the Province of Zamora Chinchipe (Figure 1, region C). The morphological similarity that exists between the Nacional variety and the native populations of the Ecuadorian jungle (fruit and seed) close to the Amazonian cities of Archidona and Macas in the region C, and already observed by Vera (1993) reinforces also this hypothesis.

Most of the representatives of the native Nacional variety have a high level of homozygosity probably associated to their self compatibility (Quiroz, 1990, Pastorelly, 1992; Loor, 1998). The question raised is whether the domestication process has fixed and selected by selfing the highly homozygous genotypes, or whether a natural selection, leading to an increased homozygosity level, was already effective in wild populations before this domestication event. A common feature of domestication is the reduction of genetic diversity in crops relative to wild progenitors. The severity of genetic loss ascribed to bottleneck effects varies greatly among crop species. Tanksley et al (1997) indicated that this reduction resulted from two major forces. First, most domestication events are thought to have involved initial populations of small size (relative to wild ancestors) that contained a narrow levels of genetic diversity. The second factor to have an impact on crop genomes is the selection for the agronomic traits that distinguish crops from their ancestors. A selective pressure on the ancestral gene pool is generally excited by men to select for favourable traits, increasing or fixing the favourable alleles at genes controlling these traits.

In this work, highly homozygous wild genotypes were identified in each region sampled. This last result suggests that self compatible alleles already existed in these wild populations and that natural selection could have already eliminated a part of the genetic burden due to inbreeding, facilitating, by this way, the domestication process from some wild populations.

The particular characteristics of some cacao varieties could have helped the selection and migration of specific cacao materials. From this point of view, the specific aromatic flavour of cocoa beans from the Nacional, detectable on bean pulp, even without a fermentation step, could have been one of the criteria used by this primitive population to choose the cocoa mother tree for further seed sowing. Indeed, it is possible that the ambulatory merchants transporting cacao pods along the routes, used the fresh pulp only for their own refreshment and alimentation without consuming the cacao beans, and thus, introducing the cacao tree into a new environment (Stone, 1984). An extensive study of flavour aroma of wild cacao genotypes associated to an evolutionary analysis of this trait variation could help to identify candidate genes responsible of flavour in the aromatic Nacional variety.

Many crop species have been domesticated for thousands of years. In cocoa, only one variety (Criollo) is known to have been domesticated for a long time (more than 2000 years) by Maya populations in Central America (Hurt *et al* 2002). Cocoa domestication seems much more recent in most of the other cultivating countries, even if wild cocoa populations exist in their areas. In the case of Ecuador, information about the domestication of its cocoa native plantations dates back for only four centuries.

Before to first cacao plantations planted by the Spanish in the Ecuadorian coast, no evidence existed of human dispersal of this species; therefore, the knowledge of the dispersal mechanism implicated in the long-distance migration process is important to explain our results. There are several hypotheses trying to explain the cocoa migration such as: the transport of fruits or seeds by birds, animals or men, but until now none of them has been formally demonstrated. Vera (1987) suggested that the Nacional cocoa variety was introduced by the old Incas ways and planted by the native people of this time in the coast regions. However, archaeological evidences suggest that these old routes were not built by the Incas but existed a long time before. These ways were already established by the pre-Columbian native people, who inhabited the Coast, Andes and Amazonian regions of Ecuador (Salazar, 1995) thousands of years before the arrival of the Incas to Ecuadorian territories. In the pre-Columbian time, trade, swap or exchange of plant material between adjacent cultures could occur using these rudimentary ways.

Unfortunately, contrary to Maya population who used hieroglyphs and represented cacao, the more primitive people that inhabited Ecuador, did not use writing symbols, limiting by this way, the available proofs about cacao domestication and use before Spanish arrival. However, archaeological evidences about contacts and exchanges of products between pre-Columbian peoples from coast, sierra and Amazonia of Ecuador, dating back 2000 years B.C. (Ayala, 1983) could provide new explanations on the origin of cocoa trees in the Ecuadorian Littoral. The archaeological evidences showed that these exchanges went as far as a place presently known as "La Cueva de los Tayos" located in the province of Morona Santiago, region C, where the ancestors of the Nacional variety were identified.

From our results, future cacao expeditions should be supported to investigate the south Amazonian area from Ecuador as the putative centre of origin of the Nacional variety and to collect new genetic resources with the same Nacional flavour specificities, suitable for fine flavour cocoa improvement in Ecuador.

References

Allen BJ, Lass AR (1983). London cocoa trade Amazon project. Final report phase 1. Cocoa Growers' Bulletin 34:1-72.

Ayala ME (1983) Nueva historia del Ecuador. Editorial Grijalbo Ecuatoriana.

Bartley BGD (2005) The genetic diversity of cacao and its utilization. CABI Publishing, London, UK.

Belkhir K, Borsa P, Chikhi L, Raufaste N and Bonhomme F (2004). GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations, Laboratoire Génome, Populations, Interactions; CNRS UMR 5000; Université Montpellier II, Montpellier (France).

Crawford LR (1980) El Ecuador en la época cacaotera. Respuestas locales al auge y colapso en el ciclo monoexportador. Editorial universitaria. Universidad Central del Ecuador.

Gaudeul M., Taberlet P. & Till-Bottraud I. (2000) Genetic diversity in an endangered alpine plant, *Eryngium alpinum* L.. (Apiaceae), inferred from amplified fragment length polymorphism markers. *Molecular Ecology* 9, 1625-1637

Hurst WJ, Tarka SM, Powis TG, Valdez F and Hester RT (2002). Archaeology: cacao usage by the earliest Maya civilization. *Nature* 418 : 289-290.

International Cocoa Organization (2006) ICCO Quarterly Bulletin of Cocoa Statistics, Vol. XXXIII, No.1, Cocoa year 2006/07

Juan A, Crespo BM, Cowan SR, Lexer C, Fay FM (2004) Patterns of variability and gene flow in *Medicago citrine*, an endangered endemic of islands in the western Mediterranean, as revealed by amplified fragment length polymorphism (AFLP). Molecular Ecology 13:2679-2690.

Lachenaud P, & Sallée B. (1993). Les cacaoyers spontanés de Guyane. Localisation, écologie et morphologie. *Café cacao thé* (Paris), **37**, n°2: 101-114.

Lachenaud P, Mooleedhar V, Caouturier C. (1997). Les cacaoyers spontanés de Guyane : Nouvelles prospections. Plantations, Recherche, Développement 4 : 25-32.

Lerceteau E, Flipo S, Pétiard V, Crouzillat D (1997) Genetic differentiation among Ecuadorian *Theobroma cacao* L. accessions using molecular and morphological analyses. Euphytica 95: 77-87.

Liu K and Muse VS (2005). Power Marker: Integrated analysis environment for genetic marker data. Bioinformatics 21 : 2128-2129.

Loor RG (1998) Obtención de híbridos de cacao tipo Nacional provenientes de materiales de alta productividad y resistentes a enfermedades. Tesis de Ingeniero Agrónomo. Universidad Técnica de Manabí, Portoviejo, Ecuador. 67p

Loor Solorzano R. G., Risterucci A.M., Courtois B., Fouet O., Jeanneau M., Rosenquist E., Amores F., Vasco A., Medina M., Lanaud C. (2009). Tracing the native ancestors of modern *Theobroma cacao* L. population in Ecuador. *Tree genetics and genomes*, 5 (*3*) : 421-433. [20100330]. http://dx.doi.org/10.1007/s11295-008-0196-3

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, 7, 639-655.

Motamayor JC (2001). Etude de la diversité génétique et de la domestication des cacaoyers du groupe Criollo (Theobroma cacao L) à l'aide de marqueurs moléculaires. Thèse de doctorat, Université Paris XI, Orsay, France, 176p.

Motamayor JC, Risterucci AM, Lopez PA, Ortiz CF, Moreno A, Lanaud C (2002). Cacao domestication I: the origin of the cacao cultivated by the Mayas. Heredity **89**: 380-386.

Nei, M. (1972). Genetic distance between populations. *Am. Nat. 106*: 283-292.

Pastorelly RD (1992) Evaluación de algunas características del cacao tipo Nacional, en la zona de Tenguel. Tesis Ing. Agr. Universidad Agraria del Ecuador. Guayaquil, Ecuador. P7.

Perrier X, Flori A, Bonnot F (2003). Data analysis methods. In: Hamon P, Seguin M, Perrier X, Glaszmann JC. Ed., Genetic diversity of cultivated tropical plants. Enfield, Science Publishers. Montpellier, France pp 43 - 76.

Pound FJ (1938). Cacao and witchbroom disease (***Marasmius perniciosus***) of South America. In H Toxopeus (ed.) Reprint Archives of cocoa research Vol 1 Amer Cacao Res. Institute.

Pound FJ (1945). A note about the cacao populations of South America. Report and Proceedings Cocoa Research Conference, London. Colonial 192 : 95-7. (Reprinted 1982 in *Arch. Cocoa Res* 1: 93-97).

Prescott WH (1847). History of the conquest of Peru with a preliminary view of the civilization of Incas. London.

Preuss P (1901) Expedition nach Central- und Sudamerika 1899/1900. Kolonial-Wirtschaftlichen Komitees, Berlin, 374pp.

Quiroz J (1990) Estudio de la compatibilidad en algunos cultivares de cacao (Theobroma cacao L.). Tesis de Ing. Agr. Universidad Técnica de Babahoyo. Babahoyo, Ecuador. 30p

Risterucci AM, Grivet L, N'Goran JAK, Pieretti I, Flament MH and Lanaud C. (2000) A high-density linkage map of *Theobroma cacao* L. *Theoretical and Applied Genetics* 101: pp 948 – 955

Salazar E (1995). El Ecuador aborigen, Quito, Corporación Editora Nacional.

Soria JV (1970). Principal varieties of cocoa cultivated in tropical America. Cocoa Growers' Bulletin 19:12-21.

Stevenson WB (1829). Historical and descriptive narrative of twenty year's residence in South America. Vol II Londres.

Tanksley SD and McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* **277**: 1063-1066

Vera BJ (1987) Antecedentes históricos. Manual del cultivo de cacao. INIAP, EET-Pichilingue. Quevedo, Ecuador, pp: 6-9.

Vera BJ (1993) Botánica del cacao. Manual del cultivo de cacao 2da edición. INIAP, EET-Pichilingue. Quevedo, Ecuador, pp: 10-15.

Vijverberg K, Kuperus P, Breeuwer JAJ, Bachmann K. (2000) Incipient adaptive radiation of New Zealand and Australian Microseris (Asteraceae), an amplified fragment length polymorphism (AFLP) study. *Journal of Evolutionary Biology* **13**:997-1008.

White GM, Boshier DH, Powell W (1999) Genetic variation within a fragmented population of *Swietenia humilis* Zucc. *Molecular Ecology* **8**: 1899-1909.

FIGURES

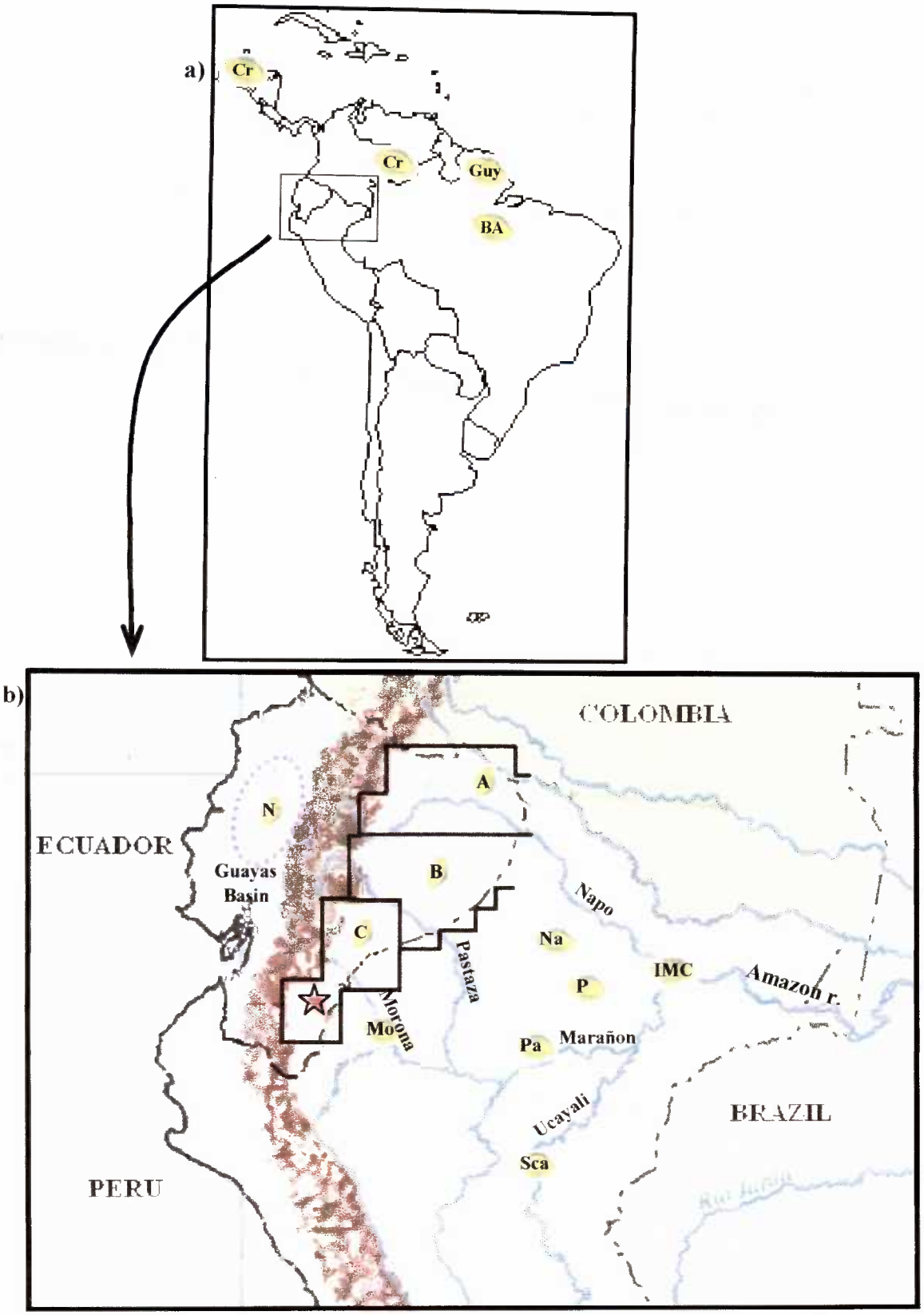


Figure 1. Geographic distribution and localization of the 14 cocoa groups used in this study. (a) Central and South America; and (b) Ecuadorian coast and upper Amazonian region. = Zone of "ARRIBA" cocoa; and, = localization of Nangaritza and Yacuambi rivers. See Table 1 for explanations of name codes.

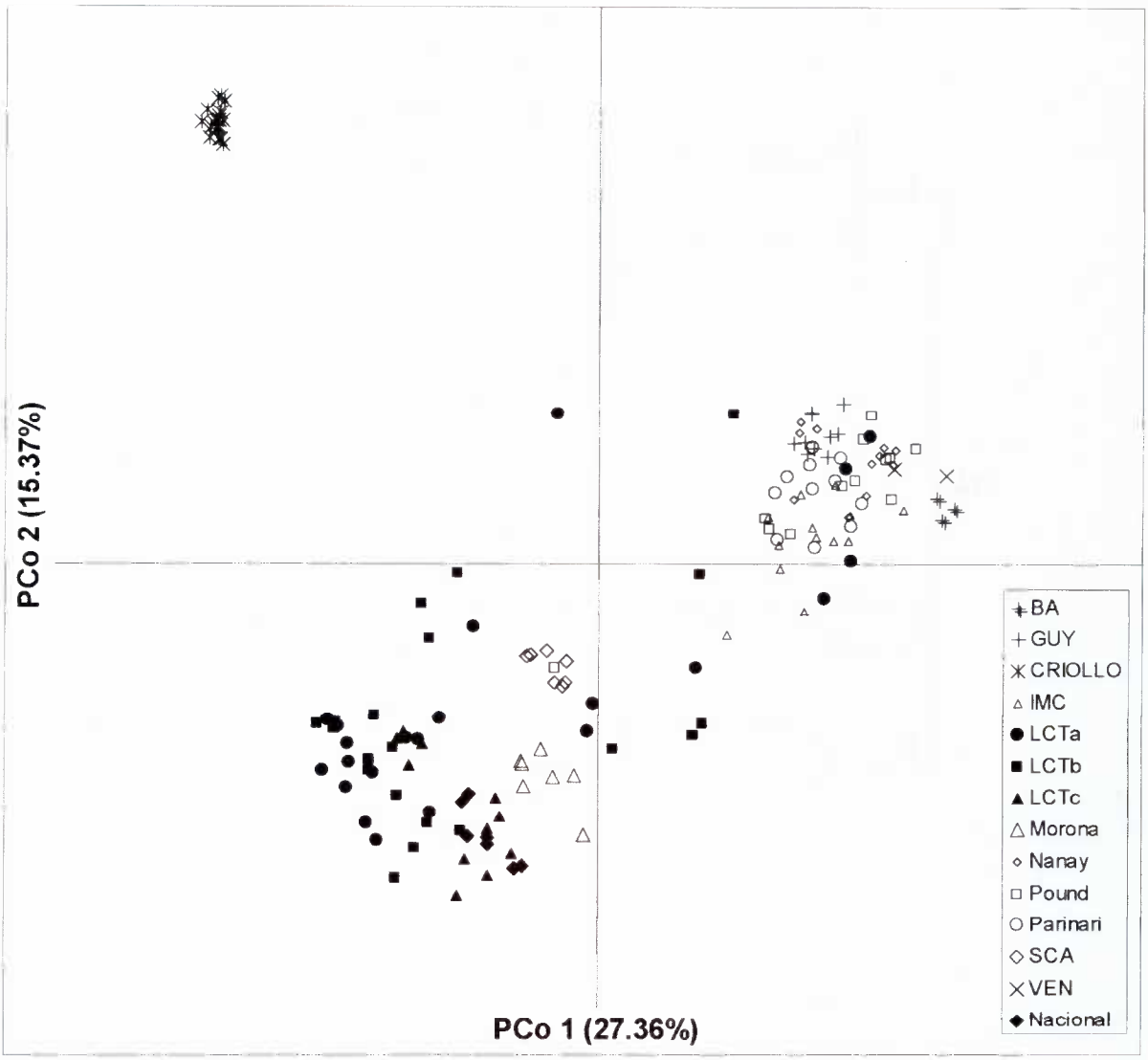


Figure 2.PCoA plot of the first and second axes based on SSRs fingerprinting of ***Theobroma cacao*** **L**. populations. The identification for each population code is shown in Table 1.

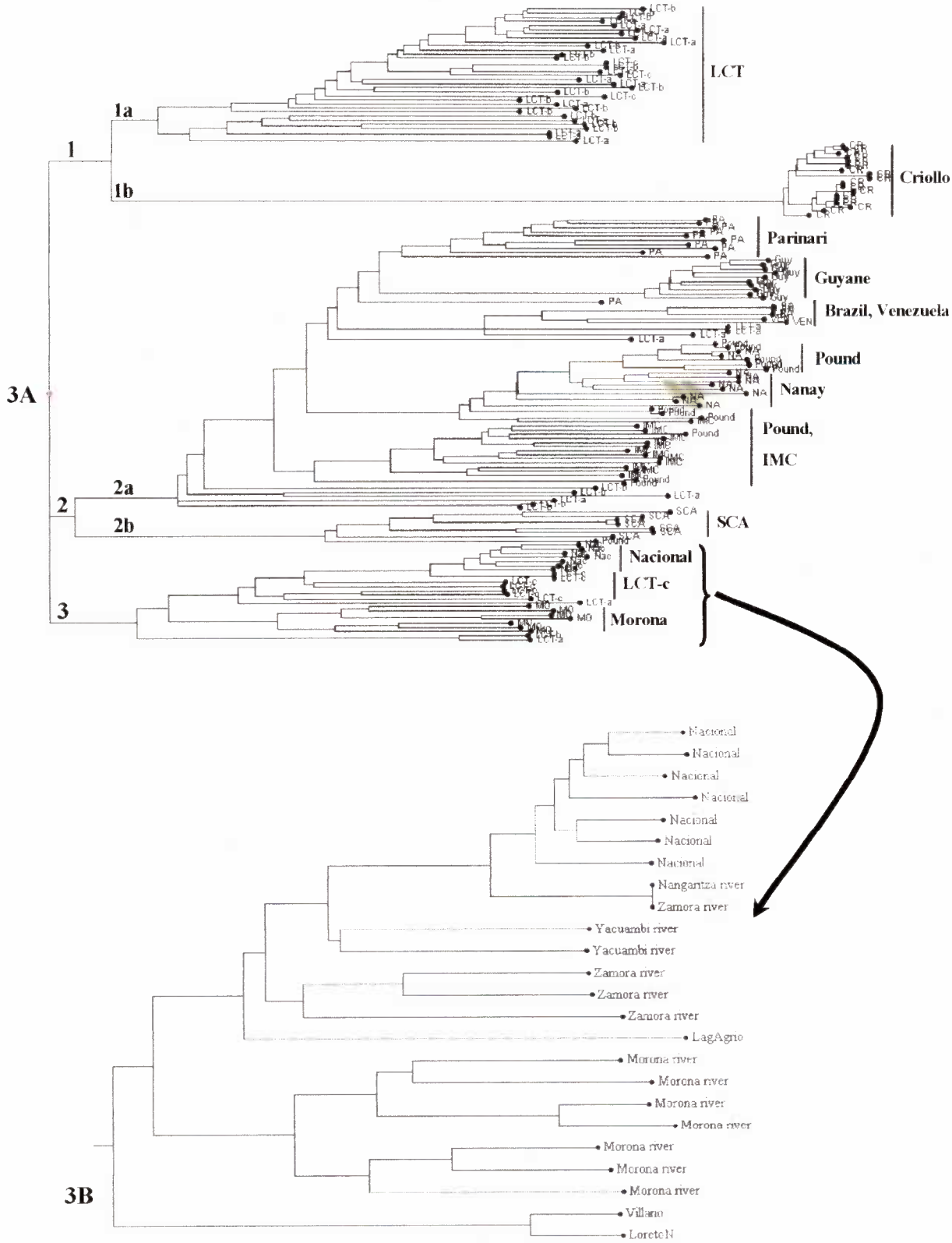


Figure 3. 3A: Neighbour-joining tree from dissimilarity matrix, representing genetic relationship among the wild and cultivated cacao accessions analysed in this study. In Fig 3B: the cluster is amplified and represented with the information about the rivers where were collected the wild genotypes closer to the Nacional. See Table 1 for code identification.

TABLES

Table 1. Passport information available for the group accessions used in this study

Group Name	Country Origin	Geographical localities (within country)	Sample Size
Nacional	Ecuador	Coast Provinces	7
LCT-EENa	Ecuador	Orient (North region)	30
LCT-EENb	Ecuador	Orient (Center region)	23
LCT-EENc	Ecuador	Orient (South region)	12
IMC	Peru	Iquitos	13
Nanay	Peru	Nanay river	12
Parinari	Peru	Tributaries of Marañon river	13
Morona	Peru	Morona river	8
Guy	Guyana		11
VEN	Venezuela		2
BA	Brazil		3
SCA	Peru	Loreto, Ucayali river	8
Pound	Peru	Nanay river	12
Criollo	Mexico and Venezuela		22
Total samples			176

Table 2. List of microsatellite loci and number of alleles per locus identified in this study.

mTcCIR	Nº alleles per locus
1	7
6	11
15	10
26	10
37	11
60	8
292	4
293	5
294	8
297	2
298	6
299	2
302	13
304	6
305	3
308	5
309	4
312	3
316	7
322	15
323	9
324	2
325	7
326	7
327	3
329	10
331	10
333	3
336	7
337	3
339	3
341	2
342	4
344	10
348	3
350	4
352	13
356	4
359	5
361	6
364	2

mTcCIR	Nº alleles per locus
366	8
369	8
373	5
374	6
375	4
376	6
378	6
379	5
380	4
382	11
383	6
387	4
391	7
392	8
393	8
394	12
396	3
397	2
398	2
400	6
402	3
404	2
405	3
411	7
413	4
417	4
418	2
419	2
420	4
422	13
426	4
427	4
429	8
432	5
435	5
436	3
438	2
446	6
447	4
80	463
	Main 5.78

Table 3. Summary of genetic diversity revealed by microsatellite markers within each group analyzed. Hexp.: Expected heterozygosity ; Hn.b.: Unbiased gene diversity (Nei, 1978); Hobs.: Observed heterozygosity; Proportion of polymorphic loci when most frequent allele does not exceed 95% (0.95) and 99% (0.99); and, mean number of alleles per locus (NMA/L).

	Hexp.	H n.b.	Hobs.	P(0.95)	P(0.99)	NMA/locus
BA	0.056	0.068	0.000	0.125	0.125	1.125
Guy	0.108	0.114	0.064	0.287	0.325	1.337
CR	0.049	0.050	0.006	0.150	0.175	1.250
IMC	0.376	0.391	0.479	0.800	0.862	2.700
LCTa	0.523	0.533	0.317	0.950	0.975	4.450
LCTb	0.473	0.485	0.358	0.950	0.962	4.150
LCTc	0.411	0.432	0.271	0.787	0.812	2.637
MO	0.297	0.321	0.314	0.700	0.700	2.000
NA	0.223	0.234	0.242	0.650	0.750	2.150
Pound	0.381	0.399	0.292	0.850	0.900	3.200
PA	0.299	0.312	0.304	0.637	0.750	2.337
SCA	0.319	0.343	0.359	0.750	0.750	2.125
VEN	0.108	0.152	0.056	0.225	0.225	1.237
Nac	0.106	0.114	0.094	0.312	0.312	1.375

Table 4. Coefficients of genetic distance (Nei, 1972) calculated for pair-wise comparisons of the 14 groups of *T. cacao* studied.

BA	0.000													
Guy	0.557	0.000												
CR	2.445	1.611	0.000											
IMC	0.416	0.473	1.435	0.000										
LCT-a	0.724	0.689	0.910	0.468	0.000									
LCT-b	0.861	0.789	0.929	0.534	0.039	0.000								
LCT-c	1.110	0.930	1.079	0.632	0.143	0.144	0.000							
MO	0.928	0.991	1.293	0.571	0.446	0.493	0.431	0.000						
NA	0.412	0.586	1.574	0.269	0.619	0.715	0.792	0.906	0.000					
Pound	0.347	0.475	1.461	0.148	0.493	0.588	0.673	0.718	0.051	0.000				
PA	0.329	0.304	1.468	0.312	0.517	0.618	0.718	0.669	0.422	0.303	0.000			
SCA	1.218	0.955	1.352	0.765	0.631	0.652	0.673	0.615	0.967	0.742	0.775	0.000		
VEN	0.233	0.535	2.141	0.385	0.735	0.892	1.098	0.977	0.380	0.322	0.402	1.172	0.000	
Nac	1.300	1.226	1.395	0.748	0.492	0.582	0.250	0.307	1.017	0.855	0.900	0.779	1.276	0.000
	BA	Guy	CR	IMC	LCT-a	LCT-b	LCT-c	MO	NA	Pound	PA	SCA	VEN	Nac

Table 5. The two most likely parents of Nacional cacao individuals. According to Cervus 2.0 software parentage analysis. All putative parents are from the south Amazonian region (c) of Ecuador.

Individual	Prob. Non-exclusion	Parent	
		First	Second
Sa16	6.937×10^{-7}	LCT-EEN91	LCT-EEN86
SNA409a	4.072×10^{-6}	LCT-EEN86	LCT-EEN91
SNA503	5.139×10^{-6}	LCT-EEN91	LCT-EEN85
SNA604	4.001×10^{-6}	LCT-EEN91	LCT-EEN85
SNA1001	1.065×10^{-6}	LCT-EEN86	LCT-EEN85
SNA1003	6.188×10^{-7}	LCT-EEN86	LCT-EEN85
B240	3.893×10^{-6}	LCT-EEN91	LCT-EEN86

II. First results of association studies related to biochemical and sensoriel traits of the Nacional population

1) Studied population

A population of 316 individuals corresponding to hybrid forms of Nacional has been initially focused for this study. These trees are part of 2 germplasm collections: the "finca la Buseta" in Tenguel and the collection of EET Pichilingue

2) Molecular analyses

The 319 individuals were genotyped with 143 microsatellite markers widespread in all chromosomes (Fouet et al., 2011).

Identification of marker/trait associations: In order to minimize the risk of false positives, we used TASSEL (Trait Analysis by ASSociation, Evolution and Linkage) software (Remington *et al.*, 2001), available at http://www.maizegenetics.net, which integrates population structure as a cofactor in the association analyses (Yu, *et al.*, 2006).

The model-based software Structure (Pritchard *et al.*, 2000) was used to infer population structure using a burn-in of 100,000, run length of 1,000,000, and 10 independent runs. A model with admixture and correlated allele frequencies was chosen. The tested K values (equivalent to the number of sub-populations) ranged from 1 to 9. SSR genotype data for 21 independent microsatellite loci located in all cocoa chromosomes were used for these analyses.

The least square solution to the fixed effects general linear model (GLM) (Searle, 1987) implemented in TASSEL 2.1 and taking into account the admixture percentages estimated by STRUCTURE, was used to identify associations. The significance of the association between a marker and a trait was determined by the *p values*.

3) Sensorial characterisation

The sensorial characterisation was made on a part of this population (about 200 individuals) by 2 panels:

- Ed Seguine, in GUITTARD lab
- Alejandra Santos and Sophie Assemat in CIRAD lab

Tasting
The liquors of cocoa samples were prepared: Beans were roasted (120°C, 25 min), shelled, coarsely then finely ground (<20 µm) and stored.
Tasting sessions were realized at Cirad and at Guittard Chocolate Company. Liquors are tasted at 50°C. Each sample was tasted at least 3 times.

Respectively 157 (at CIRAD) and 201 (at GUITTARD) individuals were evaluated for their sensorial traits.

The traits subjected to sensorial evaluation and their statistical characterisation are the following:

Sensorial evaluations made at GUITTARD

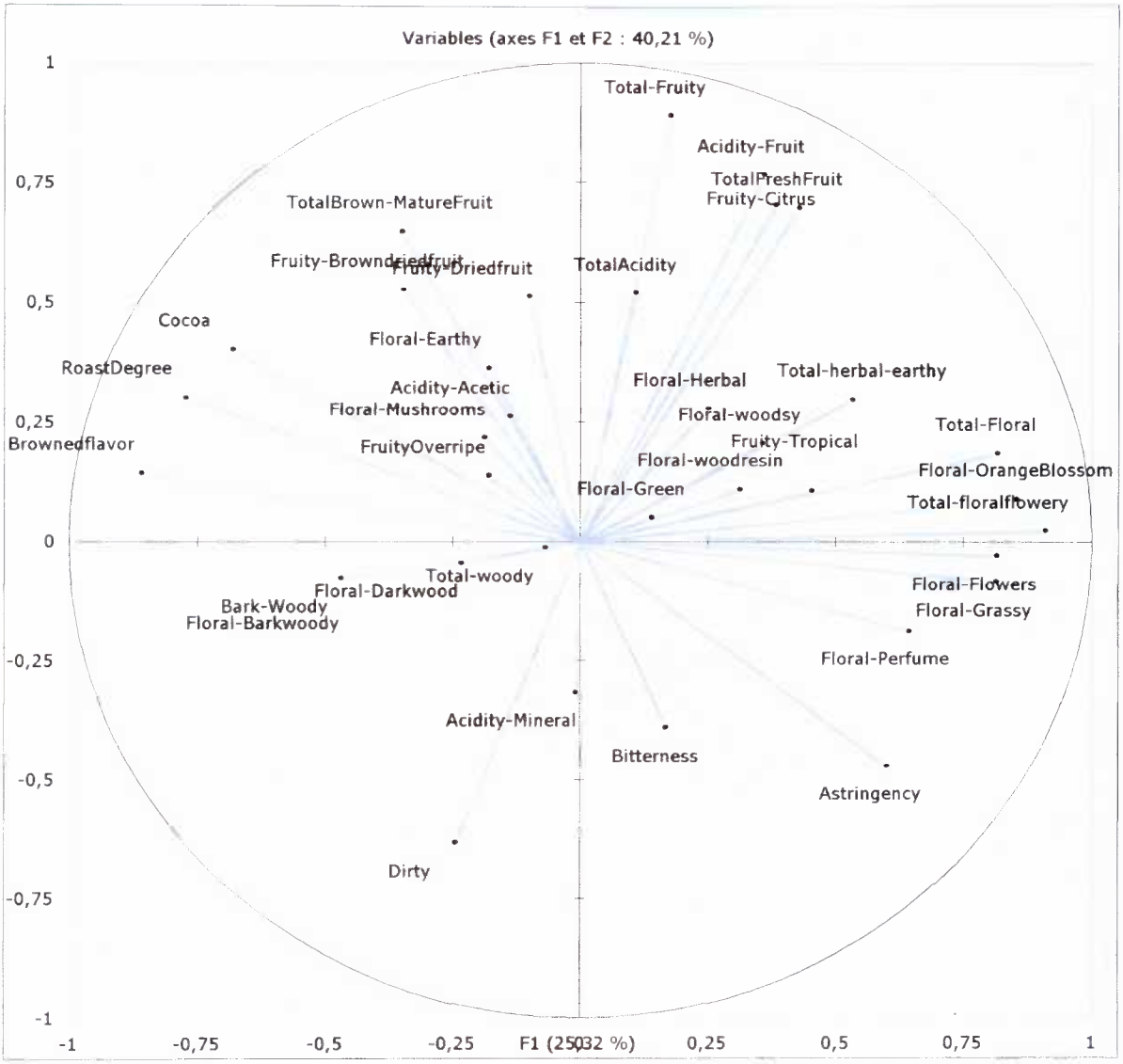
trait	nb ind.	Mini	Max	Mean	Sdt deviat.
Cocoa	201	1,20	5,80	3,88	0,81
TotalAcidity	201	0,00	4,80	1,95	0,95
Acidity-Fruit	201	0,00	2,80	0,72	0,57
Acidity-Acetic	201	0,00	3,00	0,93	0,57
Acidity-Mineral	201	0,00	1,50	0,26	0,31
Bitterness	201	2,50	5,00	3,73	0,49
Astringency	201	1,25	9,00	4,99	1,69
Fruity-Citrus	201	0,00	2,50	0,62	0,54
Fruity-Tropical	201	0,00	1,25	0,08	0,20
Fruity-Driedfruit	201	0,00	1,50	0,16	0,30
Fruity-Browndriedfruit	201	0,00	1,50	0,20	0,33
FruityOverripe	201	0,00	1,75	0,14	0,27
Floral-woodsy	201	0,00	1,00	0,11	0,24
Floral-Grassy	201	0,00	3,00	0,55	0,69
Floral-Green	201	0,00	2,00	0,36	0,44
Floral-Earthy	201	0,00	2,50	0,60	0,55
Floral-Herbal	201	0,00	1,80	0,44	0,47
Floral-Mushrooms	201	0,00	1,75	0,17	0,33
Floral-OrangeBlossom	201	0,00	3,80	0,86	0,80
Floral-Flowers	201	0,00	4,75	0,59	0,81
Floral-Perfume	201	0,00	3,00	0,19	0,45
Floral-Barkwoody	201	0,00	3,75	1,84	0,78
Floral-Darkwood	201	0,00	3,20	0,73	0,69
Floral-woodresin	201	0,00	2,00	0,12	0,30
Brownedflavor	201	2,50	7,20	4,50	0,75
RoastDegree	201	3,00	6,60	4,60	0,61
Raw	201	0,00	2,20	0,19	0,40
Total-Fruity	201	0,00	2,83	0,95	0,63
TotalFreshFruit	201	0,00	2,58	0,67	0,57
TotalBrown-MatureFruit	201	0,00	2,13	0,41	0,44
Total-Floral	201	0,00	6,81	2,20	1,31
Total-herbal-earthy	201	0,00	3,58	1,38	0,80
Total-floralflowery	201	0,00	5,85	1,19	1,15
Total-woody	201	0,00	3,20	0,86	0,69
Bark-Woody	201	0,00	3,75	1,84	0,78
Brownedflavor	201	2,50	7,20	4,50	0,75
Dirty	201	0,00	4,00	0,69	0,84

Sensorial evaluations made at CIRAD

trait	nb ind.	Mini	Max	Mean	Sdt deviat.
Aroma intensity (nose)	157	1,50	5,00	3,24	0,56
Aroma cacao (nose)	157	0,00	4,00	1,96	0,78

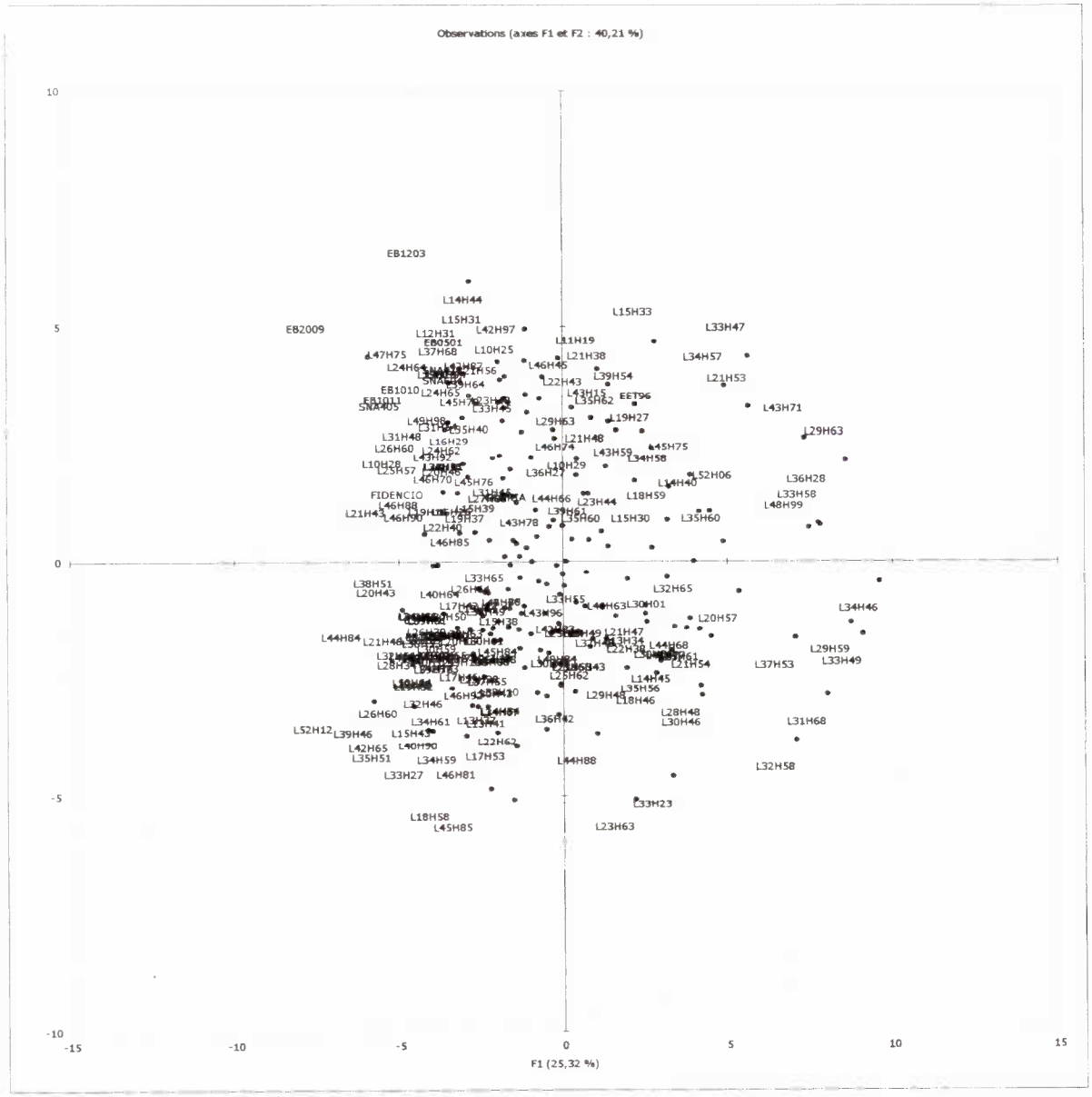
Flavour intensity	157	2,00	4,50	3,33	0,50
Acidity	157	0,00	3,50	1,82	0,78
Bitterness	157	1,00	5,00	2,75	0,78
Astringency	157	0,50	4,00	2,14	0,83
Cacao	157	0,50	4,00	2,21	0,74
Fresh fruit	157	0,00	4,00	1,23	0,98
Dry fruit	157	0,00	3,50	1,77	0,71
Floral	157	0,00	5,00	1,32	1,21
Nutty	157	0,00	3,50	1,37	0,73
Herbal	157	0,00	3,50	0,94	0,80
Mouthfeel	157	2,50	5,00	3,63	0,48
Overall evaluation	157	0,00	5,00	2,59	0,87

A principal component analysis was made with the Guittard and sensorial data to have a synthetic view of the several sensorial floral and fruity traits of the Nacional population. The 2 first axis of the ACP made on Guittard data represent 40% of the variation of the population, those of CIRAD data give a similar representation of the traits variation (42%). The contribution of the traits and the distribution of the genotypes studied are represented for each set of data in the following figures:

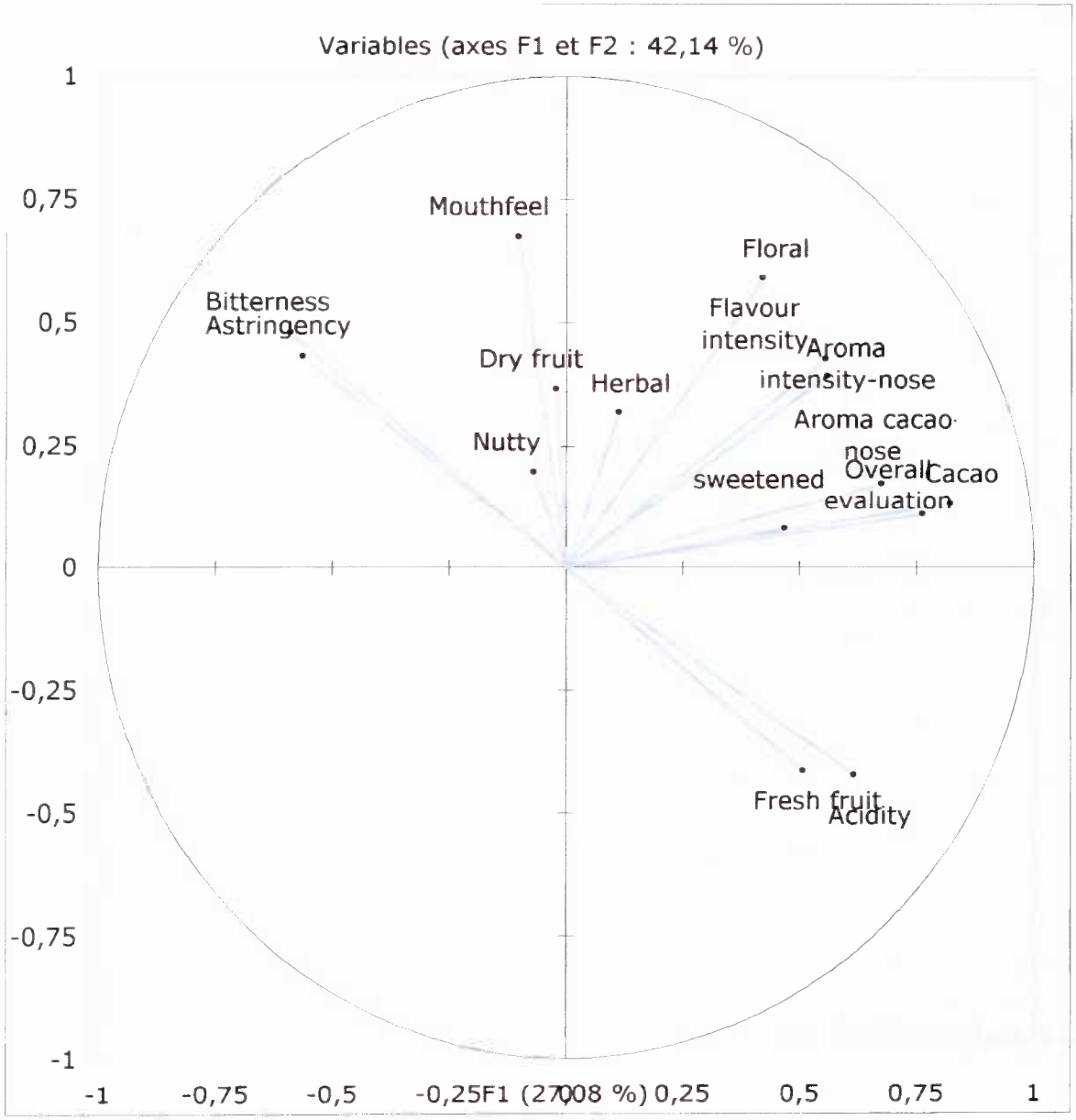


ACP made on GUITTARD data, representation of traits

The higher values of the floral and fruity traits are gathered in the right side of the Axis 1 of the figure, and in opposition with the cocoa, browned and woody flavours.
The Axix 2 discriminate the individuals with a higher acidity and fruity flavor at the top of the axis from the individuals with a higher bitterness and astringency towards the bottom of this axis.

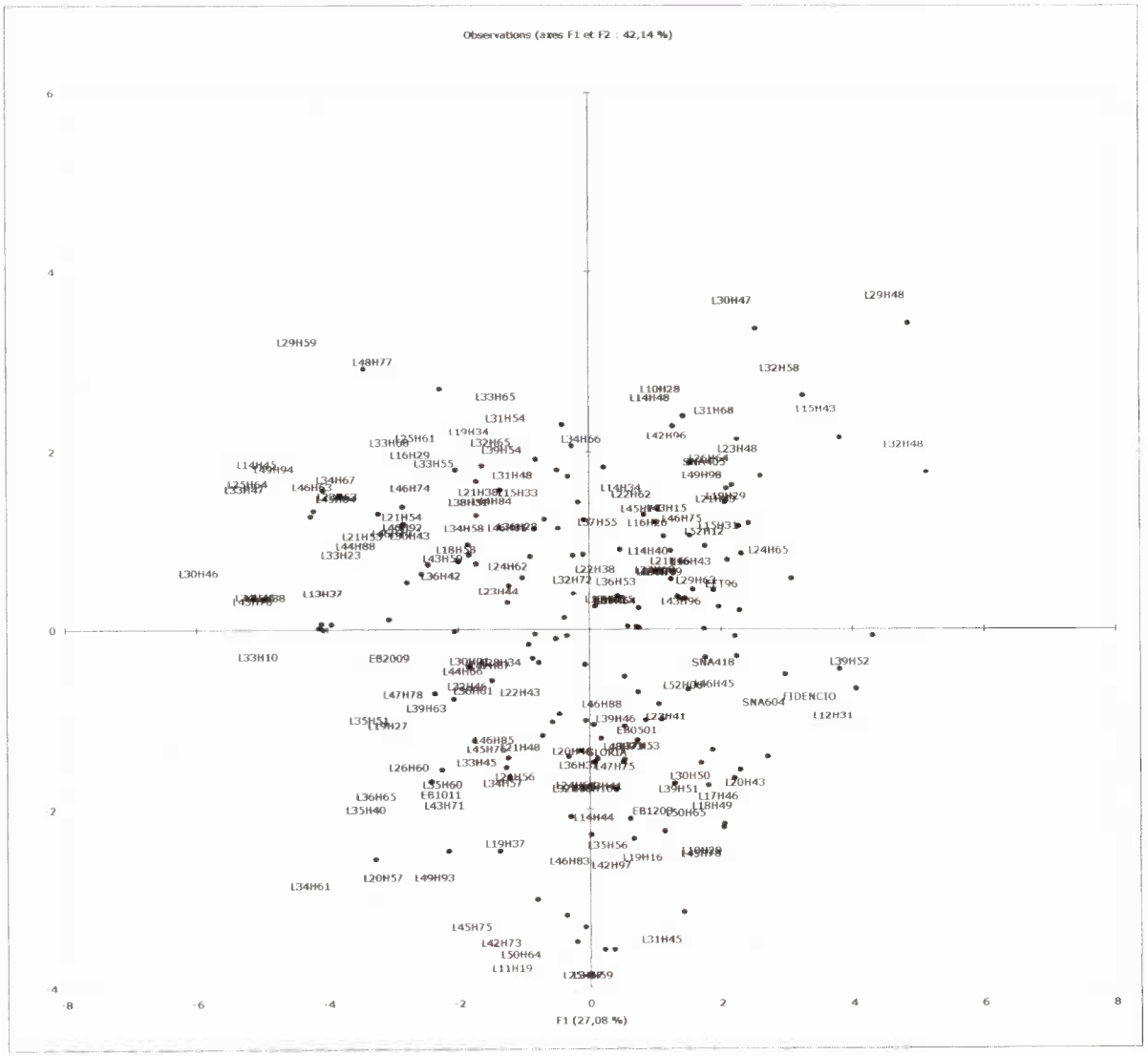


ACP made on GUITTARD data, representation of individuals



ACP made on CIRAD data- repartition of traits

In the axis 1 bitterness and astringency are associated, and in opposition with the cocoa and overall estimation of quality. Cocoa aroma evaluated by the nose or mouth are completely associated. The floral and flavour intensity are more intense for the individuals gathered in the upper right side of this ACP. The corresponding individuals can be identified in the following figure.



ACP made on CIRAD data- repartition of individuals

3) Biochemical characterisation

The evaluation of the relationships between genetic, biochemical and sensorial traits in order to identify markers linked to cocoa aromatic flavour of Nacional varieties implies the biochemical characterization of each tree of the selected population of Nacional cocoa type.

A - Methods

Sample preparation
Cocoa beans came from the pods of 319 individual trees of the "finca la Buseta" in Tenguel and of the collection of EET Pichilingue. Nineteen defective pods were discared.
The beans corresponding to the pods of each tree were respectively placed in nylon net and fermented according to the micro fermentation technique. Fermentations were carried out in cubic wooden boxes (60 cm), and the cocoa layer was 50 cm. The duration of fermentation was 5 days; the beans were mixed up respectively at the 2^{nd} and 4^{th} day.
Each cocoa sample was respectively sun dried before (UFSD: unfermented and sun dried) and after (FSD: fermented and sun dried cocoa beans) fermentation.
To carry out all the chemical and sensory analyses, 100 g of UFSD and 750 g of FSD are necessary.
Thirty two trials (harvesting, fermentation and drying) were carried out to obtain 271 samples which were clearly identified, 29 which were doubtful (problems of labeling) and 2 were defective..

The number of defective beans of healthy pods, the weight of 200 fresh beans, the weight of 100 fermented and dried beans, the ratio weight of fresh beans / weight of fermented and dried beans, were the factors used in Ecuador to characterize the potential production and the quality of the production of each tree.

Chemical analyses: volatile compounds
The volatile compounds were analysed by SPME-GC-MS. A Carboxen/DVB/PDMS SPME fibre (Supelco Co., Bellefonte, PA, Ca.) was used to extract volatile constituents from the cocoa headspace.

Preliminary study
In order to improve the extraction method reproductibility, cocoa sample number 187 (tree L46 H88 of the EET collection) was analysed as nibs and as powder.
Cocoa powder was obtained by cooling cocoa nibs with liquid nitrogen then grinding them in a coffee grinder. All the cococa powder particules were smaller than 0.5mm.
The extraction-analysis of (2 g of) the cocoa powder and nibs were replicate 5 times.

Extraction conditions used in the study
50 g of cocoa beans were shelled, and then crushed in a mortar to obtain small particles of cocoa (nibs). 2.85 g of these nibs were poured in a glass vial, and wet with 3 ml of the internal standard solution. The vial was then sealed (septum) and put in an oven at 50°C for 1 h.

Internal standard solution preparation
Solution I : 5 ml of 1-butanol are dissolved in hexane (completed to 100 ml).
Internal standard solution: 5 ml of solution I are homogenized in 49.5 ml of distilled water.

Volatile compound extraction
After 15 min in the oven, the SPME device allows the fiber to be placed in the vial. The duration of adsorbtion is 45 min.

Volatile compounds analysis
The fiber is desorbed for 4 min in the GC injector.

GC analysis
HP 5890
Column : DB-WAX (60 m*0,25 mm i.d., with 0.25 µm phase coating)
Column temperature : 5 min isothermal at 40 °C, increased from 40 to 180 °C at 2 °C.min^{-1} and maintained for 15 min.
Temperature injector :250°C detector FID : 250°C
Gaz vector He : 2.8 ml.min^{-1},
Split/Splitless with Split ratio: 1/10
Results are expressed by the ratio pic area / internal standard area (undimentional value)
GC-MS analysis
GC : Agilent 6890, SM : Agilent 5973N
Column : DB-WAX (60 m*0,25 mm i.d., with 0.25 µm phase coating)
Column temperature : the same program that GC analysis was used.
Temperature injector : 250 °C
He 1.5 ml.min^{-1},
Split/Splitless with ratio1/20
M/z range : 40 to 350 uma
Librairy : Wiley Mass Spectral Data

Chemical analyses: non volatile compounds

Wet chemistry
40 cocoa samples representative of genetic diversity of the population were analyzed according to the standard methods. All the samples were shelled, finely ground (<0.5 mm) and stored in a deep freeze.
Purines: cocoa beans were deshelled and ground (<0.5 mm).
0.3 g of ground cocoa was extracted by boiling water during 30 minutes. After filtration, the extract was analyzed by HPLC
HPLC HP serie Agilent 1100
Column : Lichrospher RP C18, 5 µm, 25 cm x 4.6 mm)
Solvant : MeOH / H_2O (30/70 V/V), 1ml.min^{-1}
Detection UV (λ_{max}) 278 nm
Injection : 40 µL

Cocoa butter: the ground cocoa samples were refined (<20 µm) and extracted by petroleum ether in a Soxhlet apparatus.
Procyanidins: ground cocoa samples were extracted with acetone /water (75/25). The extract was saturated with NaCl and the organic phase was evaporated. The residue was dissolved in water, washed with $CHCl_3$, and then extracted with ethyl acetate. The organic phase was dried on anhydrous sodium sulphate and evaporated. The final residue was dissolved in methanol and analyzed by HPLC.

NIRS analyses

Absorption spectra (diffuse reflexion) were obtained with a Perstorp 6500 Analyzer. Spectra were collected and analyzed using the software NIRS 2 version 4.11 (Infra Soft International). The near infrared spectra of the 40 cocoa samples were registered and added to the existing NIR data base. This allows one to check if the samples fit with the existing calibration curve and to develop a more robust calibration.

The spectra of all the unfermented and dried cocoa samples were also registered.

B - Results

Chemical analyses: volatile compounds, unroasted cocoa beans

Preliminary study

The same number of compounds were characterized (tab. 1) in cocoa nibs and cocoa powder. It is logical that the mean values (5 reps) are different. All in all, the coefficients of variation are smaller in the case of cocoa nibs.

Table 1 - Volatile compounds identified in cocoa nibs and cocoa powder

Ref	Volatile compounds	Nibs		Powder	
		Mean	C. Var (%)	Mean	C. Var (%)
1	Pentanal	14	9	901	162
2	Methyl sulfide	7	17	36	7
3	2-Methyl-propanal + propan-2-one	5	26	44	19
4	Methyl acetate	26	21	69	19
5	Ethyl acetate	11	11	115	19
6	Butan-2-one	756	21	214	23
7	2-Methyl-butanal	34	14	498	30
8	3-Methyl-butanal	90	17	17	21
9	Ethanol	134	25	227	18
10	Ethyl propanate	7	29	53	13
11	Ethyl 2-Methylpropanate	8	26	44	27
12	Penta-2 one	9	18	26	25
13	2-méthylpropyl acetate	6	21	64	20
14	2-Ethyl-5-methyl furane	57	26	81	23
15	2-Methyl-3-butan-2-ol	3	14	15	10
16	(1)* + (2)*	52	17	802	17
17	1-Méthylbutyl acétate	5	5	35	31
18	Hexanal (or hexan-2-one)	386	21	14	11
19	2- and 3-méthyl-butyl acétate	675	17	1444	17
20	2-Pentanol	226	16	305	26
21	Heptan -2-one	908	24	1228	18
22	2-Methyl -1-butanol	19	21	37	21
23	3-Methyl-1-butanol	81	21	153	21
24	2 Hexanol + 2-pentyl furan	4	38	69	25
25	Cis ocimene + Ethyl hexanoate	36	20	54	16
26	Styrene	23	35	22	32
27	3-Méthyl-but-2-èn-1-yl acetate	10	76	79	19
28	1-Methylhexyl acetate	115	30	76	16
29	Hexyl acetate	15	32	374	11
30	3-Hydroxy-butan-2-one	87	36	41	10
31	Heptan-2-ol	840	14	1614	19

32	Nonan-2-ol	62	32	29	20
33	Nonanal	35	30	248	11
34	Ethyl octanoate	11	28	18	8
35	Acetic acid	2435	31	89	24
36	Furfural	3	13	2806	26
37	Trans linalol oxyde	92	22	51	23
38	2,3,4,5 Tetramethyl-pyrazine	731	33	263	21
39	2,3-Butanediyl diacetate	3	1	1703	21
40	Benzaldehyde	172	24	230	22
41	2,3-Butanediyl diacetate (Meso)	22	25	699	13
42	Linalol	182	22	15	18
43	γ – butyrolactone	5	48		
44	Phenylethanal	81	50	73	34
45	Acetophenone	21	70	325	80
46	Ethyl benzoate	11	43	416	13
47	Pentanoic acid	5	45	35	13
48	Benzyl acetate	7	23	18	13
49	Cis Epoxylinalol	4	29	28	34
50	Trans Epoxylinalol	15	31	23	65
51	Ethylphenyl acetate	28	25	37	30
52	4-Hydroxy-acetophenone	5	26	91	13
53	Phenylethyl acetate	166	19	60	20
54	Butyl benzoate	5	15	591	20
55	Guaiacol	14	27	28	17
56	Benzylic alcohol	9	23	25	26
57	2-Phenylethanol	9	70	19	24
58	2-Acetyl pyrrole	260	22	715	23
59	Octanoïque acid	9	21	23	26
60	Nonanoïque Acid	4	45	30	87
61	Ethyl hexadecanoate	6	12	35	81

(1)* = Diméthyl disulfide ; (2)* = Ethyl 3-méthyl-butanoate

Analyses of the population

A study of repeatability of the extraction method (used for the whole population) was carried out with the sample (tree) n° 58. Analyses were replicated 12 times. The coefficients of variation are acceptable (Annex 1).

Two hundred eighteen samples were analyzed. Sixty five volatile compounds were characterized of which 62 were identified (tab. 2).

Table 2 - Identification of the volatile compounds of "Nacional" cocoa

Volatile compounds	min	max	average
Ethanal	0	307	36
Methyl sulfide	0	389	53
2-Methyl-propanal+propan-2-one	0	362	120
Methyl acetate	40	1424	351
Ethyl acetate	11	37093	3949
Butan-2-one	0	7964	1509
2-Methyl-butanal	0	2500	385
3-Methyl-butanal	26	2528	732
Ethanol	19	2252	216
Ethyl propanoate	0	333	39
Ethyl-(2-Methyl)-propanate	0	501	53
Pentan-2-one	42	10503	1080
2-Methylpropyl acetate	23	884	202
2-Ethyl-5-methyl furane	0	351	79
Toluene	0	128	28
2-Methyl-butan-2-ol	0	419	95
Dimethyl disulfide + ethyl 3-methyl-butanoate	0	1041	23
1-Methyl-butyl acetate	21	1859	587
Hexanal (or hexan-2-one)	16	629	41
2 and 3-Methyl-butyl acetate	333	4171	1290
Pentan-2-ol	14	1595	425
Hepta-2-one	278	9249	1558
2-Methyl-1-butanol	0	390	44
3-Methyl-1-butanol	12	889	234
2 Hexanol + 2-pentylfuran	0	969	140
Cis ocimene + ethyl hexanoate	0	227	58
Styrene	13	282	76
3 Methyl-but-2-en-1-yl acetate	9	176	65
1-Methylhexyl acetate	14	2545	348
Hexyl acetate	0	217	20
3-Hydroxy-butan-2-one	27	1044	275
Heptan-2-ol	189	9114	1633
Nonan-2-ol	0	2725	76
Nonanal	0	3284	295
Unidentified	18	360	103
Ethyl octanoate	0	7670	89
Acetic acid	265	17706	5486
Furfural	0	229	74
Linalol trans furanic oxyde	72	527	222
2,3,4,5 Tetramethylpyrazine	34	6859	1715
2,3-butan-di-yl diacetate	10	1030	163
Benzaldehyde	26	3074	765
Meso 2,3 butan-di-yl diacetate	15	929	171

Table 2 - End

Volatile compounds	min	max	average
Linalol	73	4067	625
Unidentified 2	144	2443	678
Unidentified 3	75	2397	551
g – butyrolactone	15	809	168
Phenylethanal	25	1208	225
Acetophenone	138	2272	617
Ethyl benzoate	12	457	57
pentanoic acid	0	1308	480
benzyl acetate	0	112	19
Linalol cis pyranic oxide	0	93	39
Linalol trans pyranic oxide	0	293	63
Ethyl phenylacetate	0	116	16
4-Hydroxy-acetophenone	0	83	4
2-phenylethyl acetate	87	1623	384
Butyl benzoate	0	337	32
Guaiacol	13	115	41
Benzylic alcohol	9	155	33
2-Phenylethanol	212	3384	867
2-Acetylpyrrole	0	163	30
Octanoic acid	0	86	20
Nonanoic acid	0	240	9
Ethyl hexadecanoate	0	60	5

min (smallest area), max (greater area) and average area of the 218 samples

Chemical analyses: non volatile compounds

Equations of calibration

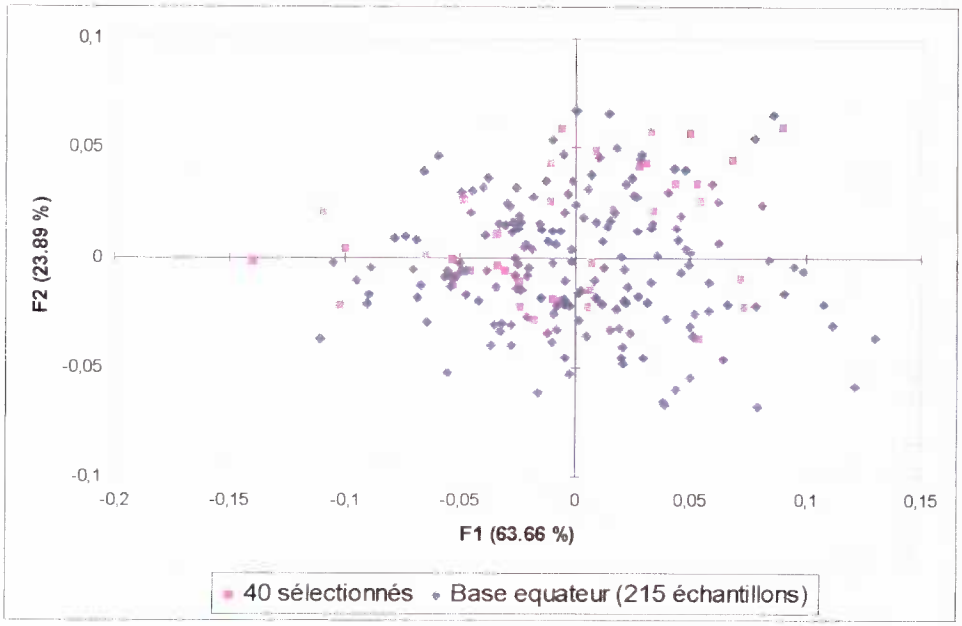


Figure 1 – Principal component analysis of the population based on NIR spectra of the 215 samples of the population

Fourty cocoa samples coming from 40 trees were selected in the population (fig. 1) in order to be representative of the diversity.
Using the results obtained by wet chemistry, new equations of calibration were calculated (tab. 3). All the graphs showing the relation between wet chemistry data and NIR prediction were determined. An example is given for theobromine (fig.2).
The contents of fat, purines and flavan-3-ols were predicted using the new equations of calibration. The distributions (occurrence frequency %) of the contents within the population are similar to that of the ratio theobromine / caffeine (fig 3).

Table 3 – Parameters of the new equations of calibration.

Compound	N	Average	SD	SEC	R^2	SECV	RPD
Proc C1	194	0,50	0,68	0,14	0,96	0,18	3,80
epicatequin	220	0,83	1,14	0,18	0,98	0,27	4,15
Proc B5	224	0,10	0,14	0,02	0,97	0,03	4,11
Proc B2	224	0,28	0,41	0,07	0,97	0,10	4,22
caffein	349	0,28	0,13	0,04	0,91	0,05	2,77
theobromine	341	1,02	0,28	0,08	0,93	0,09	3,20
Fat	479	55,09	3,06	0,96	0,90	1,15	2,66
Dry matter	816	94,08	0,93	0,15	0,97	0,17	5,51

N: number of samples of the population
SD: standard deviation

SEC: standard error of calibration
R^2: Coefficient of multiple determinations
SECV: standard error of cross validation
RPD = SD/SECV

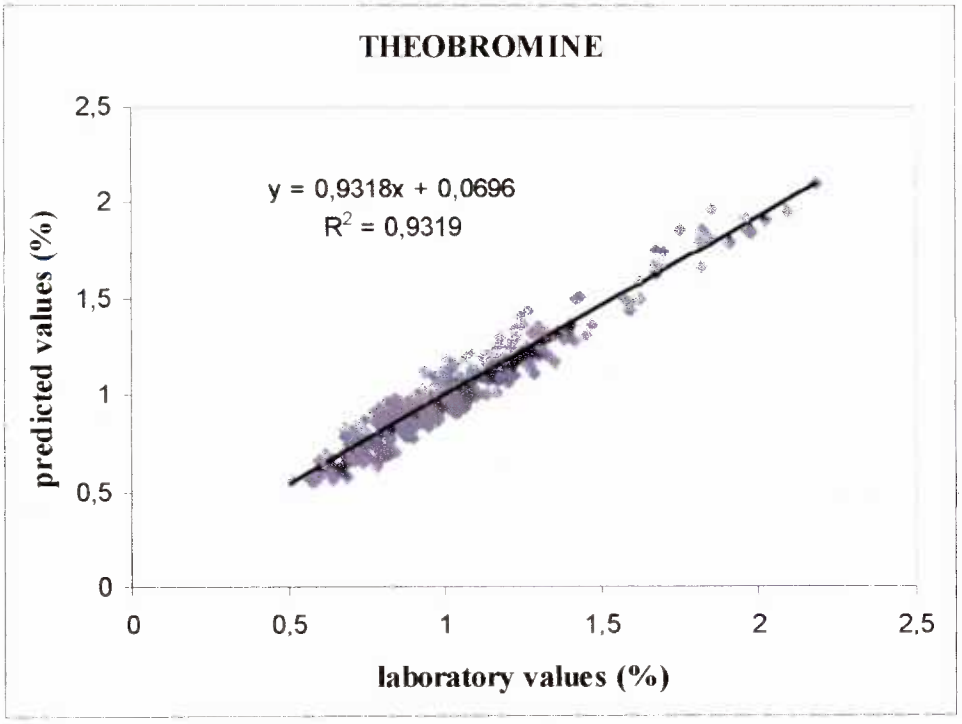


Figure 2 – Correlation between laboratory and predicted values for theobromine

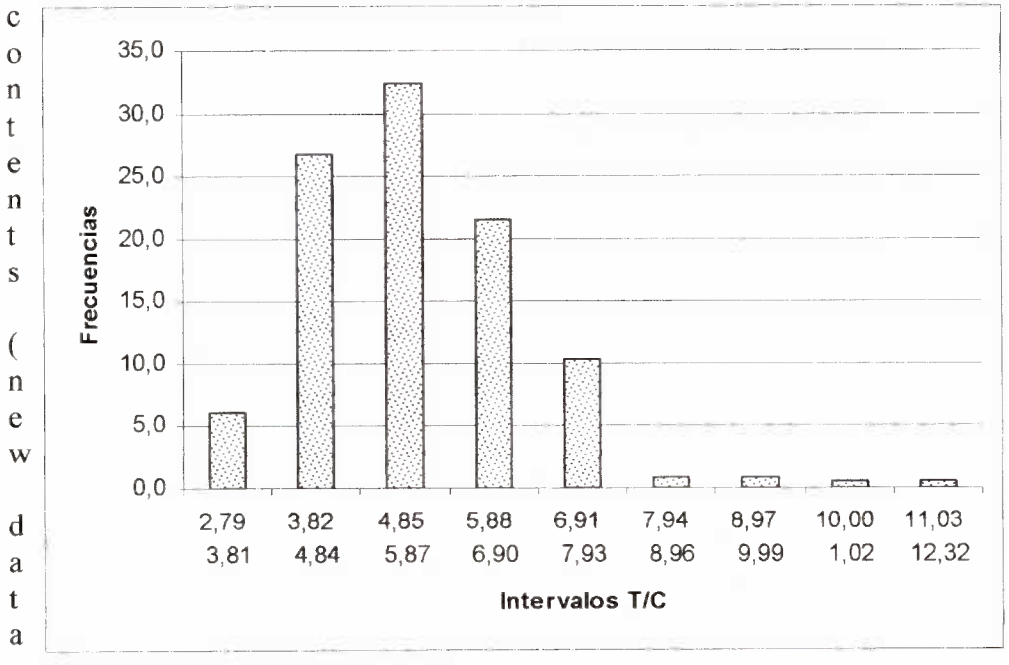


Figure 3. Distribution of the ratio theobromine/caffeine (unfermented and dried beans)

The statistics of non volatils compounds observed in the population are the following:

trait	nb ind.	Mini	Max	Mean	Sdt deviat.
C1NIRBS	169	0,933	2,873	1,952	0,355
EpiNIRBS	169	1,727	5,671	3,916	0,725
B5NIRBS	169	0,248	0,675	0,487	0,079
B2NIRBS	169	0,697	1,889	1,386	0,223
Caféine-NIRBS	169	0,168	0,597	0,330	0,080
Théobromine-NIRBS	169	1,160	2,175	1,718	0,203
FAT-NIRBS	169	45,575	56,768	51,059	1,851

EPI : epicathechin, B2, B5 : procyanidin dimer, C1 : procyanidin trimer

Roasted cocoa beans

The roasted cocoa bean volatile compounds analysed are reported in table 4

One hundred eight compounds were identified, 43 were present in both fermented (and dried) and roasted cocoa beans. 63 compounds were characterized in fermented and dried beans, 87 in roasted beans.

The data statistical analyses allowed one to characterize the great variability between the analyzed cocoa samples. They showed some intra-group (non volatile compounds, volatile compounds, sensory profile) correlations. Nevertheless, a single inter-group correlation was found between the non volatile fraction of the fermented and dried cocoa beans and the volatile fraction of the roasted cocoa beans.

Calendar of activities

Nov. 2003: selection of the plant material, starting the micro fermentation of the cocoa beans from each selected tree

2004 and 2005: preparation of the cocoa samples

20 June 2005 – 15 June 2006

Analyses of the volatile compounds of the cocoa samples by Juan Carlos Jimenez

November 2006 – Nov 2008

Analyses of the non-volatile compounds (unfermented and dried cocoa samples), of the volatile compounds of roasted (125°C / 25 min) cocoa beans and participation to the tasting sessions by Alejandra Saltos.

Cirad's people involved in the Biochemical part of the project: Fabrice Davrieux (NIRS), Renaud Boulanger (volatile compounds), Sophie Assemat (sensory profiles), Emile Cros (coordination).

An annual mission in Ecuador was made every year (selection of the trees, training in micro fermentation technique, follow up of the project, participation in various workshop).

Other activities

INIAP has organized a big workshop untitled "Calidad integral de cacao - Teoría y Práctica" in November 2004 (15^{th} to 17^{th}), taking the opportunity of our presence in Pinchilingue. So, M. Jacquet and E. Cros gave respectively the lectures "Métodos de campo para estimar la

calidad organoléptica del cacao" and "Factores que afectan el desarrollo del sabor a cacao - Bases bioquímicas del perfil aromático". Furthermore they were involved in the practical sessions.

Table 4 - Volatile compounds identified before and after roasting of fermented and dried cocoa beans (to be continued)

		Roasted	Fermented
1	Ethanal		X
2	Methanethiol	X	X
3	Acetaldehyde	X	
4	Dim˘thyl sulfide	X	
5	2-methylpropanal	X	X
6	Acetone	X	
7	Methyl acetate	X	X
8	Ethyl acetate	X	X
9	Butan-2-one		X
10	2-methylbutanal	X	X
11	3-methylbutanal	X	X
12	Ethanol	X	X
13	Propanoate d'˘thyle	X	X
14	Ethyl 2-m˘thylpropanate	X	X
15	Ac˘tate de propyle	X	
16	Pentan-2-one	X	X
17	Ac˘tate de 3-m˘thylbutyle	X	X
18	2-˘thyl-5-m˘thylfuranne	X	X
19	Butanoate d'˘thyle	X	
20	Toluene		X
21	2-m˘thyl -but-3-¸n-2-ol	X	X
22	2-m˘thylbutanoate d'˘thyle	X	
23	Pentan-2,3-dione	X	
24	Disulfure de dim˘thyle	X	X
25	Ac˘tate de 1-m˘thylbutyle	X	X
26	und˘cane	X	
27	Hexanal (or hexan-2-one)		X
28	Ac˘tate de 3-m˘thylbutyle	X	
29	Pentan-2-ol	X	
30	Heptan-2-one	X	X
31	2-pentylfuranne + ocim¸ne	X	
32	hexanoate d'˘thyle	X	
33	2-Methyl-1-butanol		X
34	2 ou 3-m˘thylbutan-1-ol	X	X
35	Hexan-2-ol	X	X
36	Cis ocimene + ethyl hexanoate		X
37	Styr¸ne	X	X
38	Ac˘tate de 3-m˘thyl-2-butan-1-yl	X	X
39	Ac˘tate 1-m˘thylhexyl	X	X
40	Hexyl acetate		X
41	1,2,5-trim˘thylbenz¸ne	X	
42	Pentan-1-ol	X	

Table 4 - Volatile compounds identified before and after roasting of fermented and dried cocoa beans (to be continued)

		Roasted	Fermented
43	méthylpyrazine	X	
44	Octan-2-one	X	X
45	3-hydroxy-butan-2-one	X	X
46	Heptan-2-ol	X	X
47	Nonan-2-ol		X
48	Nonanal		X
49	Unidentified		X
50	Ethyl octanoate		X
51	2,6-diméthylpyrazine	X	
52	Cis hept-4-èn-1-ol	X	
53	Trisulfure de diméthyle	X	
54	2,3-diméthylpyrazine	X	
55	Octanoate de méthyle	X	
56	Nonan-2-one	X	
57	2-éthyl-6-méthylpyrazine	X	
58	2-éthyl-5-méthylpyrazine	X	
59	2,4,6-triméthylpyrazine	X	
60	2,3,5-triméthylpyrazine	X	
61	Triméthylpyridine	X	
62	Octatnoate d'éthyle	X	
63	Acide acétique	X	X
64	Furfural	X	X
65	2,3-diméthyl-5-éthylpyrazine	X	
66	linalol trans furanic oxide	X	X
67	2,3,5,6-tétraméthylpyrazine	X	X
68	2,3-butan-di-yl diacetate		X
69	2-éthylhexan-1-ol	X	
70	2-(2-furanyl)-ethanone	X	
71	Benzaldéhyde	X	X
72	Butan-2,3-diol	X	X
73	2-hydroxyhéxanoate d'éthyle	X	
74	Linalol	X	X
75	Butane-1,3-diol	X	
76	Unident.1		X
77	(43,45,88) Unident.1		X
78	g-butyrolactone	X	X
79	Phenylethanal		X
80	Acétophénone	X	X
81	Ethyl benzoate		X
82	Acétate de 1-phényléthyle	X	
83	Benzyl acetate	X	

Table 4 - Volatile compounds identified before and after roasting of fermented and dried cocoa beans (to be continued)

		Roasted	Fermented
84	5-méthyl-2furanméthanol	X	
85	Acide pentanoïque	X	X
86	benzyl acetate		X
87	Epoxylinalol	X	X
88	Epoxylinalol	X	X
89	Acétate d'éthylphényle	X	
90	4-Hydroxy-acetophenone		X
91	Acétate de phényléthyle + 2-éthylphénol	X	X
92	Butyl benzoate		X
93	Guaiacol		X
94	Cinnamaldéhyde	X	
95	Dodécanoate d'éthyle	X	
96	Acide hexanoïque	X	
97	2-méthoxyphénol	X	
98	Alcool benzylique	X	X
99	2-phényléthanol	X	X
100	2-phénylbut-2-ènal	X	
101	2-acétylpyrrole	X	X
102	Phénol	X	
103	2-formylpyrrole (1H-pyrrole-2-carboxaldéhyde)	X	
104	1 ou 3 phenoxypropan-(1 ou 2)-ol	X	
105	Acide octanoïque	X	X
106	5-méthyl-2-phénylhéx-2-ènal	X	
107	Acide nonanoïque	X	X
108	Ethyl hexadecanoate		X

ANNEX 1

Repeatability of the method of extraction (to be continued)

		Average	S.D.	CV%
n°	**Aldehydes**			
1	Ethanal	27	4	15
2	2-Methyl-butanal	587	244	42
3	3-Methyl-butanal	635	147	23
4	Nonanal	125	22	17
5	Benzaldehyde	71	19	26
6	Phenylethanal	513	134	26
	Sulfides			
7	Methyl sulfide	39	12	31
	Esters			
8	Methyl acetate	440	116	26
9	Ethyl acetate	15332	3360	22
10	Ethyl propanoate	590	96	16
11	Ethyl-(2-Methyl)-propanate	1797	283	16
12	2-Methylpropyl acetate	56	10	18
13	1-Methyl-butyl acetate	119	25	21
14	2 and 3-Methyl-butyl acetate	15	2	11
15	3 Methyl-but-2-en-1-yl acetate	86	22	26
16	1-Methylhexyl acetate	18	3	16
17	Hexyl acetate	362	88	24
18	Ethyl octanoate	15	4	25
19	2,3-butan-di-yl diacetate	551	117	21
20	Meso 2,3 butan-di-yl diacetate	143	34	24
21	Ethyl benzoate	268	45	17
22	benzyl acetate	48	8	18
23	Ethyl phenylacetate	77	21	27
24	2-phenylethyl acetate	9	1	14
25	Butyl benzoate	40	11	27
26	Ethyl hexadecanoate	16	3	17
	Ketones			
27	Butan-2-one	2386	375	16
28	Pentan-2-one	1282	175	14
29	Hexan-2-one	79	15	20
30	Hepta-2-one	605	180	30
31	3-Hydroxy-butan-2-one	904	175	19
32	Acetophenone	192	56	29
33	4-Hydroxy-acetophenone	129	25	19

Repeatability of the method of extraction (end)

		Average	S.D.	CV%
n°	**Alcohols**			
34	Ethanol	1270	573	45
35	2-Methyl-butan-2-ol	191	34	18
36	Pentan-2-ol	687	122	18
37	2-Methyl-butan-1-ol	1014	149	15
38	3-Methyl-butan-1-ol	53	13	25
39	2 Hexanol (+ trace of 2-pentylfuran)	21	5	25
40	Heptan-2-ol	24	6	25
41	Nonan-2-ol	166	29	17
42	Benzylic alcohol	1132	170	15
43	2-Phenylethanol	1060	199	19
	Furanes			
44	2-Ethyl-5-methyl furane	143	26	18
45	Furfural	34	9	28
	Aromatic hydrocarbons			
46	Toluene	33	5	15
47	Styrene	100	29	28
48	Guaiacol	35	7	21
	Terpenes			
49	Linalol trans furanic oxyde	152	24	16
50	Linalol	154	28	18
51	Linalol cis pyranic oxide	25	5	22
52	Linalol trans pyranic oxide	75	26	34
	Acids			
53	Acetic acid	5387	1495	28
54	pentanoic acid	205	47	23
55	Octanoic acid	24	6	23
56	Nonanoic acid	19	6	31
	Co-eluted compounds			
57	Cis ocimene + ethyl hexanoate	14	7	50
58	2-Methyl-propanal+propan-2-one	136	24	17
59	Dimethyl disulfide + ethyl 3-methyl-butanoate	95	22	23
	Pyrazines			
60	2,3,4,5 Tetramethylpyrazine	604	184	30
	Unidentified			
61	A	758	212	28
62	B	51	15	28
63	C	379	108	29
	Lactones			
64	g – butyrolactone	176	41	23
	Pyrroles			
65	2-Acetylpyrrole	21	5	25

Results of association studies

Association studies were conducted on sensorial traits and on some classes of biochemical compounds : the terpene compounds particularly important for the floral Nacional specific trait, the pyrazine related compounds potentially involved in nutty and roast taste, and the non volatil compounds (flavonoids, purines, fat content).

The following associations were identified, and the corresponding pvalue reported ; only one association per fragment of 15 cM was reported for a same trait.

Sensorial traits

The results of associations are reported separately for each taster :

Trait - E.S. eval. (Guittard)	Locus	Chr	Chr_pos	p_Marker	Rsq_Marker
Astringency	mTcCIR015	1	18,2	0,00003	0,15
Astringency	mTcCIR001	8	0,0	0,00019	0,07
Astringency	mTcCIR409	9	5,2	0,00037	0,09
Bitterness	mTcCIR423	2	17,6	0,00026	0,56
Bitterness	mTcCIR408	5	42,2	0,00005	0,14
Bark-Woody	mTcCIR391	8	48,9	0,00075	0,12
Floral-Barkwoody	mTcCIR391	8	48,9	0,00075	0,12
Floral-Darkwood	mTcCIR026	8	35,4	0,00096	0,15
Floral-woodresin	mTcCIR342	1	77,0	0,00012	0,15
Floral-woodresin	mTcCIR411	2	33,9	0,00001	0,18
Floral-woodresin	mTcCIR439	6	7,4	0,00066	0,10
Floral-woodresin	mTcCIR427	9	3,5	0,00015	0,12
Total-woody	mTcCIR433	10	10,2	0,00028	0,15
Fruity-Browndriedfruit	mTcCIR419	1	18,7	0,00000	0,13
Fruity-Browndriedfruit	mTcCIR438	3	68,2	0,00029	0,09
Fruity-Browndriedfruit	mTcCIR382	8	38,1	0,00057	0,15
Fruity-Browndriedfruit	mTcCIR320	9	46,9	0,00006	0,12
Brownedflavor	mTcCIR015	1	18,2	0,00014	0,14
Brownedflavor	mTcCIR368	5	68,5	0,00064	0,15
Brownedflavor	mTcCIR320	9	46,9	0,00100	0,09
TotalBrown-MatureFruit	mTcCIR419	1	18,7	0,00040	0,08
TotalBrown-MatureFruit	mTcCIR320	9	46,9	0,00036	0,10
Raw	mTcCIR363	9	56,9	0,00001	0,18
Total Acidity	mTcCIR318	2	83,8	0,00049	0,08
Dirty	mTcCIR444	8	20,5	0,00041	0,12

Trait -S.A. eval. (CIRAD)	Locus	Chr	Chr_pos	p_Marker	Rsq_Marker
Floral	mTcCIR319	9	97,8	0,00200	0,11
Floral	mTcCIR410	3	68,6	0,00240	0,10
dry fruit	mTcCIR292	2	26,0	0,00008	0,16
Herbal	mTcCIR327	7	26,6	0,00110	0,09

Trait - A.S. eval.(CIRAD)	Locus	Chr	Chr_pos	p_Marker	Rsq_Marker
Astringency	mTcCIR301	6	37,4	0,00037	0,19
Int-arom. (nose)	mTcCIR413	6	25,2	0,00047	0,15
Floral	mTcCIR422	1	76,2	0,00190	0,17

Biochemical traits

Terpen compounds

Linalol and other terpenic compounds are known to be involved in many floral and aromatic traits in other species. In cocoa, Linalol has been observed at higher level in Nacional varieties, and such terpen compounds could be involved in the floral taste of Nacional.
A specific focus has been made on these terpene compounds evaluated in both non roasted and roasted beans. The following associations have been identified in the Nacional population :

Trait - roasted beans (RO-)	Locus	Chr	Chr_pos	p_Marker	Rsq_Marker
undécane	mTcCIR350	1	53,2	0,000931	0,12
undécane	mTcCIR379	2	39,3	0,000412	0,13
undécane	mTcCIR405	3	21,7	0,000839	0,09
undécane	mTcCIR352	3	54,3	0,000673	0,18
undécane	mTcCIR298	5	26,3	0,000007	0,19
undécane	mTcCIR026	8	35,4	0,000673	0,18
undécane	mTcCIR305	9	30,0	0,000804	0,09
undécane	mTcCIR445	9	73,7	0,000006	0,19
cis linalol oxide	mTcCIR303	8	31,5	0,001100	0,12
Linalol	mTcCIR352	3	54,3	0,001800	0,17
Linalol	mTcCIR018	4	19,2	0,001200	0,14
Linalol	mTcCIR345	5	38,7	0,000436	0,15
Linalol	mTcCIR409	9	5,2	0,000596	0,10
Linalol	mTcCIR445	9	73,7	0,001400	0,13
Linalol	mTcCIR388	10	48,3	0,001700	0,17
Epoxylinalol	mTcCIR331	1	26,8	0,000000	0,22
Epoxylinalol	mTcCIR060	2	51,4	0,000132	0,18
Epoxylinalol	mTcCIR012	4	35,5	0,000773	0,17
Epoxylinalol	mTcCIR378	9	10,1	0,000326	0,12
Epoxylinalol	mTcCIR008	9	48,7	0,000019	0,23
Epoxylinalol	mTcCIR376	1	29,3	0,000022	0,17

Trait - non roasted beans (NRO)	Locus	Chr	Chr_pos	p_Marker	Rsq_Marker
Linalol trans furanic oxyde	mTcCIR434	2	5,4	0,000033	0,23
Linalol trans furanic oxyde	mTcCIR414	2	48,8	0,000000	0,35
Linalol trans furanic oxyde	mTcCIR316	3	62,0	0,000186	0,16
Linalol trans furanic oxyde	mTcCIR298	5	26,3	0,000321	0,14
Linalol trans furanic oxyde	mTcCIR408	5	42,2	0,000907	0,13
Linalol trans furanic oxyde	mTcCIR404	5	67,4	0,000014	0,15
Linalol trans furanic oxyde	mTcCIR006	6	0,0	0,000000	0,23
Linalol trans furanic oxyde	mTcCIR306	9	18,1	0,000155	0,11
Linalol trans furanic oxyde	mTcCIR024	9	30,4	0,000547	0,15
Linalol	mTcCIR414	2	48,8	0,000006	0,31
Linalol	mTcCIR438	3	68,2	0,000700	0,09
Linalol	mTcCIR408	5	42,2	0,000093	0,15
Linalol	mTcCIR404	5	67,4	0,000006	0,16
Linalol	mTcCIR006	6	0,0	0,000650	0,14
Linalol	mTcCIR329	8	4,1	0,000373	0,17
Linalol	mTcCIR306	9	18,1	0,001700	0,08
Linalol cis pyranic oxide	mTcCIR406	1	53,3	0,000251	0,16
Linalol cis pyranic oxide	mTcCIR379	2	39,3	0,000140	0,15
Linalol cis pyranic oxide	mTcCIR040	3	16,3	0,000065	0,18
Linalol cis pyranic oxide	mTcCIR394	4	19,7	0,000104	0,16
Linalol cis pyranic oxide	mTcCIR006	6	0,0	0,000273	0,15
Linalol cis pyranic oxide	mTcCIR327	7	26,6	0,000006	0,14
Linalol cis pyranic oxide	mTcCIR339	9	41,7	0,000129	0,11
Linalol cis pyranic oxide	mTcCIR388	10	48,3	0,000020	0,22
Linalol trans pyranic oxide	mTcCIR431	3	46,7	0,000418	0,11
Linalol trans pyr oxide	mTcCIR438	3	68,2	0,000019	0,13
Linalol trans pyranic oxide	mTcCIR344	4	64,3	0,000343	0,12
Linalol trans pyranic oxide	mTcCIR408	5	42,2	0,000816	0,13
Linalol trans pyranic oxide	mTcCIR006	6	0,0	0,000788	0,13
Linalol trans pyranic oxide	mTcCIR329	8	4,1	0,000098	0,19
Linalol trans pyranic oxide	mTcCIR306	9	18,1	0,000518	0,09

Pyrazine compounds

Pyrazine compounds are important flavour compounds of many fruits and vegetables, giving nutty, caramellic or roast aromas as well as herbaceous and earthy aromas. Several positive associations have been identified in these analyses :

	Locus	Chr	Chr_pos	p_Marker	Rsq_Marker
2,3,4,5 Tetramethylpyrazine (NRO)	mTcCIR329	8	4,1	5,66E-04	0,17
2,3,5,6-tétraméthylpyrazine	mTcCIR376	1	29,3	5,11E-06	0,19
2,3,5,6-tétraméthylpyrazine	mTcCIR375	1	47,4	1,86E-04	0,16
2,3,5,6-tétraméthylpyrazine	mTcCIR304	1	60,9	5,42E-05	0,17
2,3,5,6-tétraméthylpyrazine	mTcCIR438	3	68,2	3,74E-04	0,09
2,3,5,6-tétraméthylpyrazine	mTcCIR408	5	42,2	1,02E-04	0,15

2,3,5,6-tétraméthylpyrazine	mTcCIR327	7	26,6	1,63E-05	0,13
2,3,5,6-tétraméthylpyrazine	mTcCIR424	9	9,1	4,44E-05	0,14
2,3,5,6-tétraméthylpyrazine	mTcCIR008	9	48,7	4,61E-05	0,22
2,3,5-triméthylpyrazine	mTcCIR376	1	29,3	8,85E-06	0,18
2,3,5-triméthylpyrazine	mTcCIR408	5	42,2	6,09E-04	0,13
2,3,5-triméthylpyrazine	mTcCIR327	7	26,6	8,66E-04	0,09
2,3,5-triméthylpyrazine	mTcCIR424	9	9,1	1,84E-04	0,13
2,3,5-triméthylpyrazine	mTcCIR008	9	48,7	5,04E-04	0,18
2,3-diméthyl-5-éthylpyrazine	mTcCIR015	1	18,2	6,47E-07	0,46
2,3-diméthyl-5-éthylpyrazine	mTcCIR366	2	20,7	7,90E-07	0,54
2,3-diméthyl-5-éthylpyrazine	mTcCIR343	4	63,3	1,37E-04	0,31
2,3-diméthyl-5-éthylpyrazine	mTcCIR337	6	53,6	5,45E-05	0,39
2,3-diméthyl-5-éthylpyrazine	mTcCIR373	8	38,0	4,79E-05	0,44
2,3-diméthyl-5-éthylpyrazine	mTcCIR363	9	56,9	4,22E-05	0,39
2,3-diméthylpyrazine	mTcCIR314	1	84,2	3,71E-05	0,26
2,3-diméthylpyrazine	mTcCIR430	2	22,6	6,05E-05	0,17
2,3-diméthylpyrazine	mTcCIR379	2	39,3	1,09E-10	0,30
2,3-diméthylpyrazine	mTcCIR394	4	19,7	5,41E-06	0,20
2,3-diméthylpyrazine	mTcCIR322	5	9,1	4,52E-10	0,33
2,3-diméthylpyrazine	mTcCIR368	5	68,5	1,46E-07	0,28
2,3-diméthylpyrazine	mTcCIR413	6	25,2	1,11E-07	0,23
2,3-diméthylpyrazine	mTcCIR444	8	20,5	2,23E-05	0,18
2,3-diméthylpyrazine	mTcCIR391	8	48,9	1,41E-05	0,19
2,3-diméthylpyrazine	mTcCIR409	9	5,2	5,22E-05	0,13
2,3-diméthylpyrazine	mTcCIR008	9	48,7	6,95E-06	0,24
2,3-diméthylpyrazine	mTcCIR417	9	90,5	7,82E-05	0,16
2,6-diméthylpyrazine	mTcCIR380	4	43,9	1,41E-05	0,21
2-éthyl-5-méthylpyrazine	mTcCIR422	1	76,2	1,53E-05	0,21
2-éthyl-5-méthylpyrazine	mTcCIR414	2	48,8	1,65E-18	0,53
2-éthyl-5-méthylpyrazine	mTcCIR392	3	3,2	6,95E-06	0,19
2-éthyl-5-méthylpyrazine	mTcCIR405	3	21,7	2,59E-06	0,14
2-éthyl-5-méthylpyrazine	mTcCIR006	6	0,0	9,52E-04	0,12
2-éthyl-5-méthylpyrazine	mTcCIR337	6	53,6	8,80E-06	0,17
2-éthyl-5-méthylpyrazine	mTcCIR001	8	0,0	5,24E-06	0,11
2-éthyl-5-méthylpyrazine	mTcCIR310	8	20,9	4,97E-04	0,13
2-éthyl-5-méthylpyrazine	mTcCIR382	8	38,1	3,07E-08	0,26
2-éthyl-5-méthylpyrazine	mTcCIR409	9	5,2	1,30E-04	0,11
2-éthyl-6-méthylpyrazine	mTcCIR380	4	43,9	8,63E-07	0,22
méthylpyrazine	mTcCIR331	1	26,8	2,71E-06	0,20
méthylpyrazine	mTcCIR380	4	43,9	7,95E-04	0,14
méthylpyrazine	mTcCIR378	9	10,1	2,57E-05	0,16
méthylpyrazine	mTcCIR339	9	41,7	2,59E-04	0,10
Triméthylpyridine	mTcCIR008	9	48,7	2,21E-04	0,20

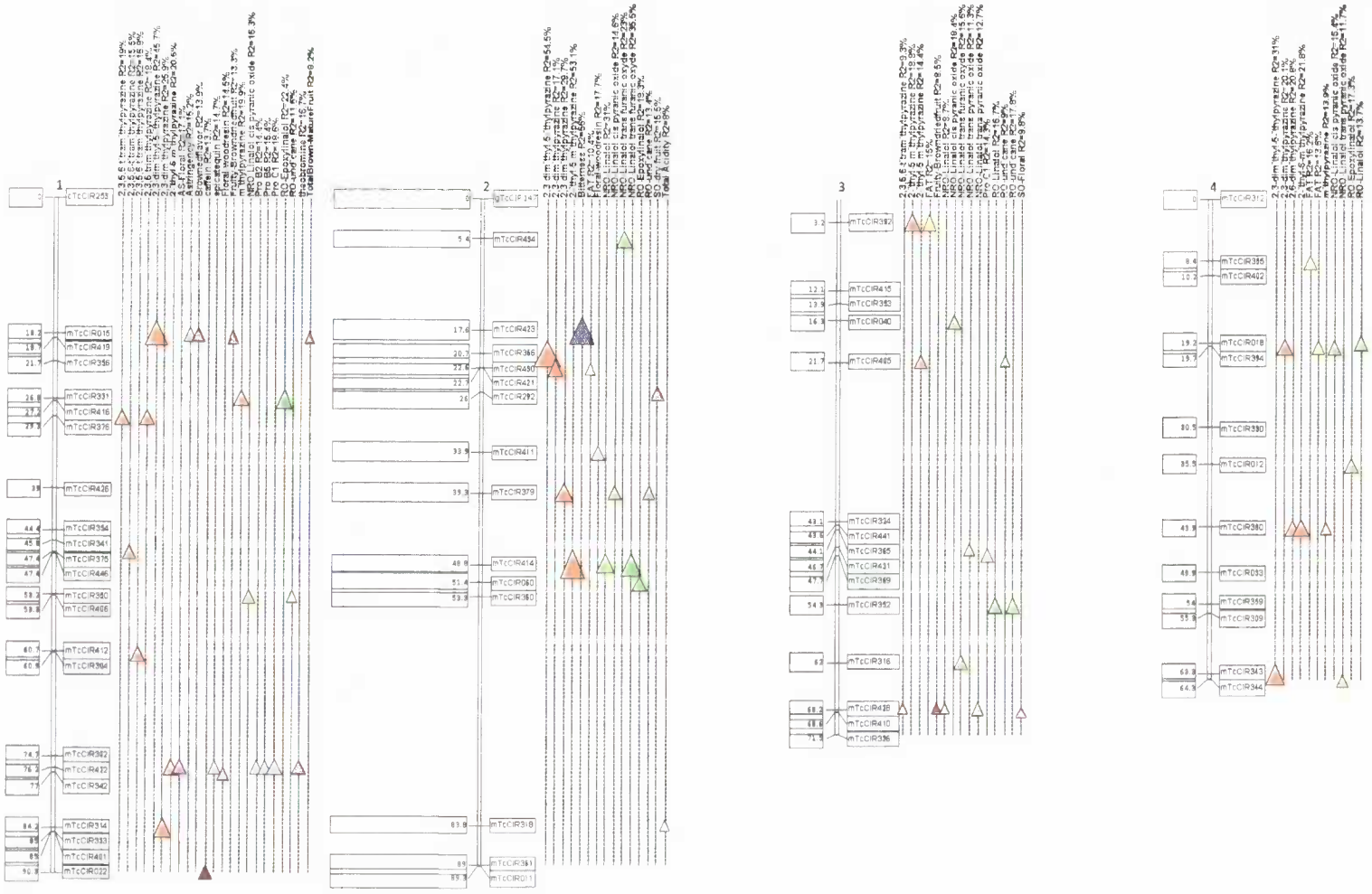
Non volatil compounds

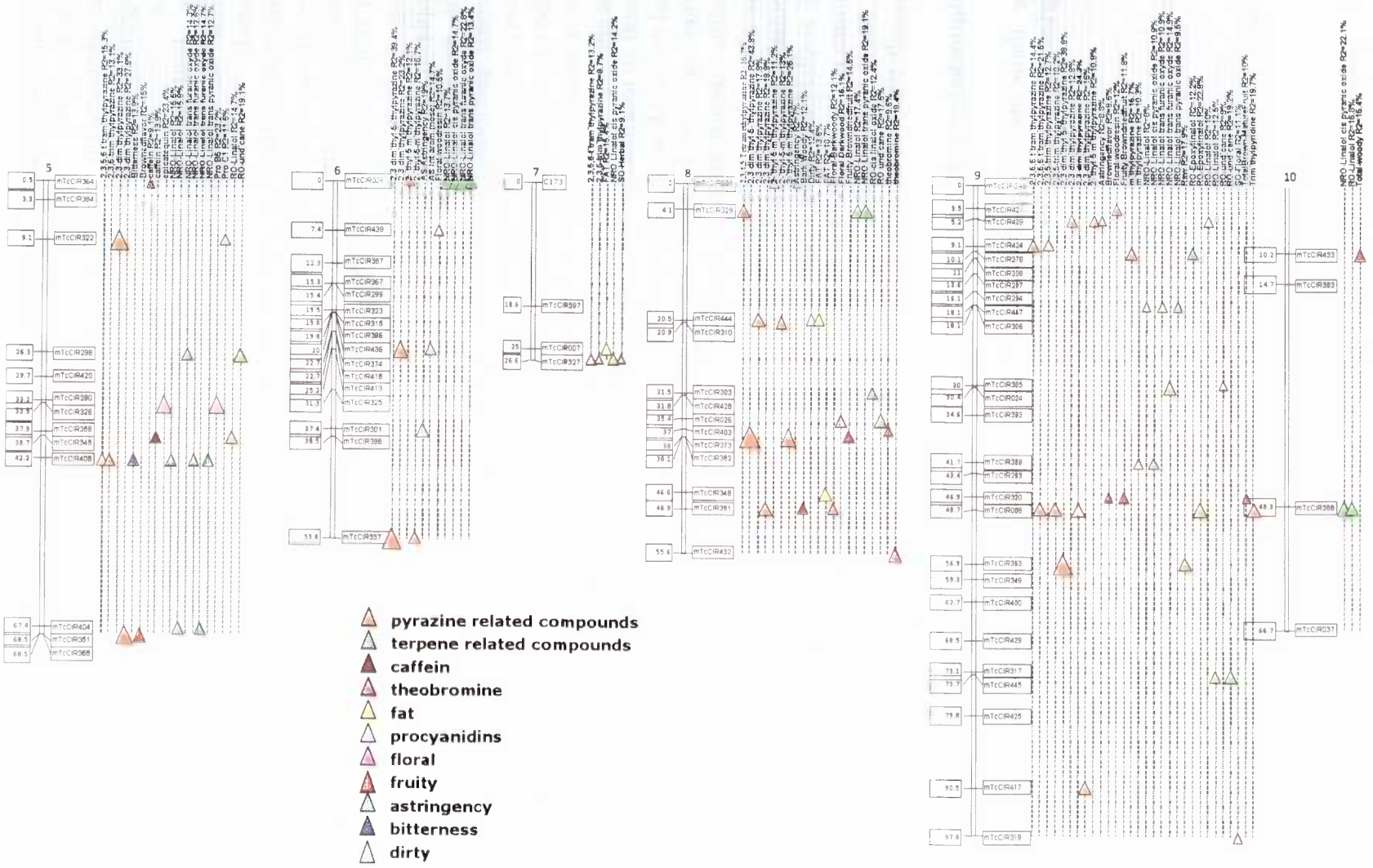
Procyanidins, purines and fat content were evaluated by NIRS on roasted beans. The following associations were identified :

Trait	Locus	Chr	Chr_pos	p_Marker	Rsq_Marker
epicatequin	mTcCIR422	1	76,2	5,44E-04	0,15
epicatequin	mTcCIR326	5	33,9	7,21E-04	0,23
Pro B2	mTcCIR422	1	76,2	5,99E-04	0,14
Pro B5	mTcCIR422	1	76,2	2,72E-04	0,15
Pro B5	mTcCIR326	5	33,9	7,07E-04	0,23
Pro C1	mTcCIR422	1	76,2	2,18E-05	0,19
Pro C1	mTcCIR369	3	47,7	0,0017	0,14
Pro C1	mTcCIR322	5	9,1	0,0016	0,12
Trait	**Locus**	**Chr**	**Chr_pos**	**p_Marker**	**Rsq_Marker**
caffein	mTcCIR022	1	90,3	5,43E-05	0,14
caffein	mTcCIR364	5	0,5	4,61E-04	0,09
caffein	mTcCIR345	5	38,7	3,00E-04	0,14
theobromine	mTcCIR422	1	76,2	4,68E-05	0,17
theobromine	mTcCIR403	8	37,0	6,60E-06	0,10
theobromine	mTcCIR432	8	55,6	3,74E-05	0,18

Trait	Locus	Chr	Chr_pos	p_Marker	Rsq_Marker
FAT	mTcCIR421	2	22,7	8,88E-05	0,10
FAT	mTcCIR392	3	3,2	2,82E-04	0,15
FAT	mTcCIR355	4	8,4	1,86E-04	0,16
FAT	mTcCIR394	4	19,7	2,25E-04	0,14
FAT	mTcCIR007	7	25,0	2,77E-04	0,15
FAT	mTcCIR444	8	20,5	1,15E-04	0,14
FAT	mTcCIR348	8	46,6	1,15E-06	0,19

All the associations identified for sensorial traits, for terpene, pyrazine and non volatil compounds are also reported in the following figures




pyrazine related compounds
terpene related compounds
caffein
theobromine
fat
procyanidins
floral
fruity
astringency
bitterness
dirty

CONCLUSIONS - PERSPECTIVES

The results obtained during this project have openned new ways for the improvement of Nacional varieties :

• The identification of representants of the pure native Nacional varieties allows a better management of the germplasm collection of Nacional clones, generally constituted by hybrid plants.

• The identification of the Amazon region probably at the origin of the domestication of Nacional has also open interesting perspectives to collect new genetic resources related to Nacional.

On the basis of these results a prospection of native cocoa trees has been undertaken in this region in august 2010 (Agropolis funded project). About 80 very old trees have been collected and multiplicated by grafting and seedlings. Until now, all the materials have been saved in the INIAP collections. The relationships of these trees with the native Nacional variety is beeing studied. This material, which presents variability represents a new source of germplasm to improve the Nacional variety

• This project has allowed to characterise at genetic, biochemical and sensorial levels the Nacional germplasm collections hold at INIAP and Tenguel stations. These analyses have shown a large variability displayed by this populaation for all studied traits, leading to a better choise of aromatic genotypes to be used in the breeding programs.

• The associations studies conducted on sensorial traits and on some biochemical compounds have allowed to identify several chromosome regions involved in these traits variations.

Several co localisations have observed between linalool and pyrazine derivatives and fruity or floral traits. However, the localisations of these associations have to be refined with new markers recently produced and with a higher density in the genetic map to maximise the possible identification of significative marker/trait associations. Complementary analyses have also to be conducted on the others biochemical volatil compounds characterised in the population and potentially involved in aroma variations.

The availability of the cocoa genome sequence and the identification already made of gene families related to cocoa qualities in the sequence (like terpens, flavonoids and fat) (Argout et al., 2011), will now facilitate the identification of candidate genes potentially involved in the aromatic traits variations.

Acknowledgements

We thank the U.S. Foreign Ministry, the U.S. Embassy, Quito, and the U.S. Department of Agriculture (USDA-ARS), for the financial support of this research.