# Assessment of genetic quality of Cacao seed gardens output using Single Nucleotide Polymorphism (SNPs) markers.

**POKOU N. Desire[1]., Tahi G. Mathias[1]., Guiraud B. Brigitte[1]., Ranjana Bhattacharjee[2]., Motamayor Juan Carlos[3]**

1- Centre National de recherche Agronomique (CNRA), 01 BP 1740 Abidjan 01

2- International Institute of Tropical Agriculture (IITA) PMB 5320, Ibadan, Nigeria.

3- MARS-Inc, C/o United States Department of Agriculture /Agricultural Research Service (USDA-ARS).13601 Old Cutler Road, Miami, Florida, USA.

## Summary

## Abstract

The development and cultivation of improved planting material of cacao have greatly contributed to ensure efficiency of production and increasing farmer's income in producing countries. The use of improved planting materials is a major key towards ensuring sustainability of the cacao. In Côte d'Ivoire, improved plant material is released under hybrids seeds. Seed gardens are plots established using the parental genotypes of selected progenies in the aims of controlling the crosses to produced hybrids seeds for planting material. Genetic quality of seed gardens output have been the target of scientists for years through testing of different layouts. Now, seeds in Côte-d'Ivoire are produced in monoclonal seed gardens using hand pollination techniques to control the crosses between male and female parents on open flowers. However, little is known about the risk of non-recommended crosses in the seed gardens. In this study, 80 SNPs markers have been used to test 1288 parental trees in the mains seed gardens of Côte-d'Ivoire.

Cocoa, SNPs markers, unique genotypes.

## Introduction

*Theobroma cacao* L. also referred to as cocoa is a tree crop native to tropical forests of American continent that belongs to the Malvaceae (Alverson *et al,* 1999; Motamayor *et al.,* 2003). Cocoa is cultivated as the source of cocoa butter and its derived product, chocolate. All the world's cocoa is supply by developing countries with predominant of West African countries which supply 70% of raw material. In, the largest producing country of Côte d'Ivoire with 40% of the word cocoa, cultivation of cocoa is one of the predominant income producing enterprises in rural area. Significant effort were done by the research institute to improve the varieties yielding from 500kg to more than 2 tons per hectare (Besse, 1977). In Côte d'Ivoire, the cocoa breeding program has been based on the creation of hybrids between different genetic groups. All the varieties developed are released also through seeds gardens. Improving the purity of varieties from seed gardens have been the target of research work for long. The layout of the fisrt generation seed gardens were established following multiclonal design in which more than one males and females genotypes are establish in the same plot. This designed rise a concern on releasing non desirable hybrids and therefore abandoned. The second generation of seed gardens are biclonal seed gardens in which only one female genotype and one male genotypes are established in a single plot. The design have been adopted with the assumption that females are self-incompatible so only crosses between the female and the surrounding male can be make. When the first molecular makers were applied to cacao, it turn out that the open pollination seeds contains some selfing (Lanaud, 1987). Then, this pollination technique have been changed into hand pollination in bi-clonal gardens. The presence of male and female in the same plot raise the concern of human mistake on the pollen donor that might lead to selfing. Thus, from 2000, a new design of monoclonal seed garden have been adopted in which male and female are established in separate plots. This design are expected to improve the genetic quality of the seed garden output by eliminating the human mistake that may happen in the

previous design. Although the used of controlled pollination methods on open flowers in seed gardens, no study to date has reported the verification of the identity of genotypes in hybrids seeds. Furthermore, despite the continued development of SNPs markers on cacao for their used in breeding (Zhang *et al,* 2009; Donald *et al.,* 2011), no study reports describing the realized performance of such markers for genotype identification in cocoa seed gardens in Côte-d'Ivoire. In this work, we used SNPs markers to ensure the identity of genotypes in seed garden in order to monitor the success of hand pollination and improve the purity of seeds in cacao seed gardens in Côte-d'Ivoire.

## Material and methods

### Sampling of plant material

Seed gardens installed at the research station of Abengourou, Divo and and Soubré were used for these study. Among which, 04 seed gardens of G15-1, G15-2, G15-3 and G15-4 at Abengourou research station, 08 seed gardens of Ei-11, Ai4-1, Ai4-2, Ai4-3, Ai4-4, Di9, B13-1 and B13-2 at Divo station and 02 seed gardens named Firca and FDPCC at Soubré. A total of 16 genotypes of ICS1, IFC1, IFC15, IFC412, IMC67, NA32, PA150, POR, SCA6, T60/887, T79/501, T85/799, UPA409, UPA419, and UPA608 were analyzed in the 14 seed gardens. In each plot, sampling was done by systematic collect of one tree per row. However, sampling was done randomly on trees within the row. At least one tree is selected per row depending on the size of the plot. A total of 1288 trees were analyzed across all seed gardens (Table 1).

Table 1: Number of trees sampled per plot

| | Seed Garden | Number of trees |
|---|---|---|
| Abengourou | G15-1 | 100 |
| | G15-2 | 117 |
| | G15-3 | 116 |
| | G15-4 | 90 |
| Divo | Ei-11 | 185 |
| | Ai4-1 | 104 |
| | Ai4-2 | 49 |
| | Ai4-3 | 31 |
| | Ai4-4 | 41 |
| | | |
| | B13-1 | 91 |
| | B13-2 | 63 |
| Soubré | FDPCC | 100 |
| | Firca | 201 |
| **TOTAL** | | **1288** |

**DNA extraction and SNP genotyping**

DNA was isolated from 0.15 g of fresh leaves collected from trees growing in seed gardens and breeders working collection. Leaves were cleaned, frozen in liquid nitrogen and ground. DNA isolation was performed using MAATB protocol as described in pokou *et al*. (2009). The DNA

solution was then purified using the phenol–chloroform method (Karakousis and Langridge 2003). SNPs fingerprinting was performed at K-Bioscience using Kompetitive Allele-Specific PCR (KASP). KASP is a homogeneous, fluorescence-based genotyping technology, initially developed by KBioscience. The technology is based on allele-specific oligo extension and fluorescence resonance energy transfer (FRET) for signal generation (Semagn *et al.,* 2014). KAPS was performed on genotypes from the seed gardens trees and breeders working collection using 80 SNPs selected from markers set developed by CIRAD (Argout et al, 2008).

**SNP data analysis**

The first step of the analysis is to identify the reference genotype for each of the seed gardens clone. This is done by multiple matching analysis performed wih Flap Jack version 1.16.10.04 (The James Hutton Institute). The predominant genotypes across all seed gardens and breeders' working collection is considered as a reference genotype for seed production. Mislabeling among seed gardens. First hierarchical clustering was performed on all unique genotypes based using XL-Stat, then pairwise multilocus matching was performed among pair of trees including trees previous identified as reference. Tree with the same names as the reference but not matching the reference genotype were declare off-type (homonymous mislabeling). Trees with different names as the reference but matching in genotype with the reference or having single mismatching were declare synonymous. Unique genotypes are define as the genotype belonging to least 10 individuals within the same label in the samples. Unknown genotypes are genotype that not match any of the 16 genotypes analyzed.

**Results and discussion**

**Analysis of Homonymous**

Of the total of 80 markers analyzed, were removed because of the high missing data. The multiple matching analysis reveals different unique genotypes and some unknown genotypes considered as off-types. The presence of off-type in cacao germplasm were revealed several times based on morphological and molecular analysis (Padi, et al., 2015). In this study, the clone in seed garden that showed more than one unique genotypes are IFC5, NA32, UPA109, UPA419 and UPA 608. These unique genotypes were renamed as follow: two genotypes of IFC5 named IFC15-1 and IFC15-2, two genotypes of NA32 named NA32-1 and NA32-2, two genotypes of UPA409 named UPA 409-1 and UPA 409-2, three genotypes of UPA419 named UPA419-1, UPA419-2, UPA419-3, and two genotypes named UPA608 named UPA 608-1 and UPA608-2. This list involve only one amelonado. The others genotypes are belong to upper amazon. Indeed, this group were shown to be more divers in many previous studies (N'Goran et al., 2000; Zhang et al.2009). The cross check of unique genotype reveals that IFC15-2 found with 25 individuals match with the IFC5 while genotypeUPA419-1 match entirely with NA32, This genotypes match with the profile of NA 32. Therefore the unique genotype of NA32 is the NA32-1. Within the NA32 labeled, NA32-2 match with T60/887 and the NA32-3 match with the T79/501

The three UPA genotypes of UPA 409, UPA419 and UPA608 analyzed presents also unique genotypes each. UPA409 reveals three genotypes named hereafter UPA409-1, UPA409-2 and UPA409-3. The UPA409-1 match most of the trees in the breeders' working collection (budwood garden). The UPA409-2 was found within individual from the seed garden G15-2 only and doesn't match any of the 16 genotypes analyzed from the budwood gardens while the UPA409-3 match with T79/501.

The UPA 419 also reveals three unique genotypes of UPA419-1, UPA419-2, and UPA419-3. The first one, UPA419-1 match with NA32 and the UPA419-2 match T85/799. It turned out that true UPA419 is the UPA419-3. As regard the genotype UPA608, two unique genotypes

were found: UPA608-1and UPA608-2. The UPA608-2 match with the genotype T60/887.The genotype UPA608-1 involved individuals from seed gardens and budwood garden and is considered to be the true UPA608. The cluster dendrogram performed using reference genotypes and unique genotypes shows three main group: one composed with ICS1 and POR, the second involve all amelonado and the third group is composed with the UPA, the T clones, PA150 and NA32. SCA6 is isolated as a single group (Figure 1).

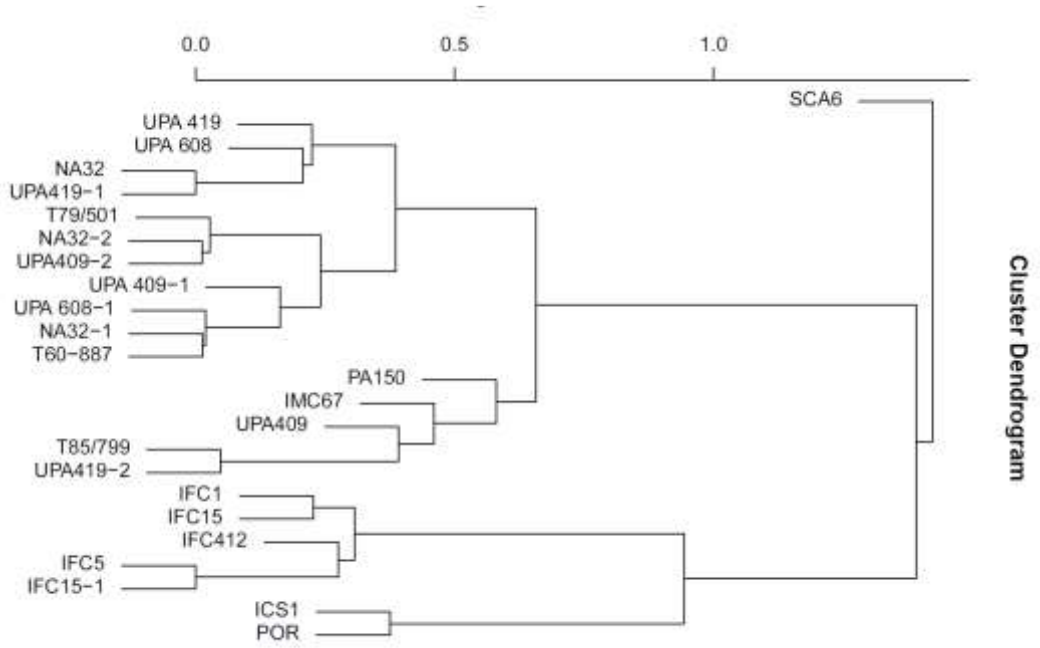


Figure1: cluster dendrogram of unique genotypes in seed gardens

**Conclusion**

In this work, we analyzed 1288 trees from 16 genotypes in 14 seed gardens. We found that the following three genotypes were generated the most unique genotypes: NA32, T60/87 and T79/501. The involvement of these genotypes in many of released varieties across seed gardens

might explain why there are within the most propagated genotypes in the different nurseries. However, only one genotype UPA409-2 was not found in the recommended list of genotypes used for establishing seed gardens.

**Reference**.


Alverson S, William Whitlock BA, Nyffeler R, Bayer C, Baum AD. (1999). Phylogeny of the core malvales: evidence from *NDHF* sequence data. Ame. J. Botany. 86(10): 1474–1486.

Argout X, Fouet O, Wincker P, Gramacho K, Legavre T, Sabau X, Risterucci A, Da Silva C, Cascardo J, Allegre M, Kuhn D, Verica J, Courtois B, Loor G, Babin R, Sounigo O, Ducamp M, Guiltinan M, Ruiz M, Alemanno L, Machado R, Phillips W, Schnell R, Gilmour M, Rosenquist E, Butler D, Maximova S, Lanaud C (2008) Towards the understanding of the cocoa transcriptome: production and analysis of an exhaustive dataset of ESTs of Theobroma cacao L. generated from various tissues and under various conditions. BMC Genomics 9:512

Besse J., (1977). La sélection générative du cacaoyer en Côte-d'Ivoire : bilan et orientation des recherches. 5ème Conférence International sur la Recherche Cacaoyère. *Ibadan, Nigeria. 95-103*

Lanaud C., 1987. Nouvelles données sur la biologie du cacaoyer ( *Theobroma cacao* L.) – Diversité des populations, système d'incompatibilité, haploïde spontanés, leurs conséquences pour l'amélioration génétique de cette espèce. Thèse de doctorat d'état. Université de Paris-Sud. 106p

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

Karakousis, A., and P. Langridge, (2003): A high-throughput plant DNA extraction method for marker analysis. Plant Mol. Biol. 21,
95a—95f.

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

Francis K. Padi, Atta Ofori , Jemmy Takrama1, Esther Djan1, Stephen Y. Opoku1, Abu M. Dadzie, Ranjana Bhattacharjee2, Juan Carlos Motamayor, Dapeng Zhang.(2015). The impact of SNP fingerprinting and parentage analysis on the effectiveness of variety recommendations in cacao. Tree Genetics & Genomes 11: 44

Semagn Kassa, Raman Badu, Sarah Hearne, Michael Olsen. (2014). Single Nucleotiode Polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of technology and its application in crop improvement. Molecular Breeding (33) 1-14.

Zhang D, Mischke S, Johnson E, Phillips-Mora W, Meinhardt L. (2009) Molecular characterization of an international cacao collection using microsatellite markers. Tree Genet Genom 5:1–10