APPLICATION OF SIMPLE SEQUENCE REPEAT (SSR) MARKERS FOR GENOTYPE DIFFERENTIATION OF 24 CASSAVA (*Manihot esculenta* Crantz) ACCESSIONS

by

Ana María Correa Morales, B. S.

A thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE
in
BIOLOGY

UNIVERSITY OF PUERTO RICO
MAYAGÜEZ CAMPUS
2008

Approved by:

_____	_____
Dimuth Siritunga, PhD President, Graduate Committee	Date
_____	_____
Duane A. Kolterman, PhD Member, Graduate Committee	Date
_____	_____
Juan Carlos Martínez Cruzado, PhD Member, Graduate Committee	Date
_____	_____
Brian M. Irish, PhD Representative of Graduate Studies	Date
_____	_____
Nanette Diffoot Carlo, PhD Chairperson of the Department	Date

ABSTRACT

The genetic diversity and differentiation of 24 cassava (*Manihot esculenta* Crantz) accessions from the Puerto Rican collection are assessed in this study. In addition, 63 cassava DNA samples from the CIAT (Centro Internacional de Agricultura Tropical) germplasm collection, representing Latin America (Colombia, Costa Rica, Guatemala, Mexico, and Brazil) and Africa (Benin, Cameroon, Ghana, Nigeria, and Sierra Leone) were included from previous cassava diversity studies. Using simple sequence repeat (SSR) markers, variation in allele frequency at 37 unlinked loci was used to estimate genetic diversity and differentiation and to find the relationships between these 24 accessions and their cultivated relatives from primary and secondary diversity centers. The SSR markers were chosen because they represent a broad coverage of the cassava genome with moderate to high polymorphism information content (PIC) and robust amplification. For all the tested SSR loci studied there were on average 96.85 % ± 6.86 polymorphic loci across all country samples with an average of 4.58 ± 1.83 alleles per locus, and a mean observed heterozygosity of 0.7 ± 0.055. Despite the low level of differentiation [F_{ST} (theta) = 0.047± 0.005] found between country samples, sufficient genetic distance (Pairwise Nei's and F_{ST} distances) existed between individual accessions to separate them according to their country of origin. UPGMA analysis of country samples revealed that Puerto Rican accessions clustered together with South American landraces. The SSR markers detected two genetically identical accessions in the Puerto Rican collection and led us identify the possibly closest cultivated relatives from Brazil and Colombia for some of the collection's accessions.

Finally, we evaluate the embryogenic capacity of the accessions present in the *in vitro* collection using the conventional solid induction medium system with 8 mg/l of 2,4-D. The explants used were young leaf lobes isolated from *in vitro* grown plants. Somatic embryogenesis was achieved in 58 % of the accessions, but only SG804, CM3064, Tremesiana, PI12900, PI12903 and Seda produced germinated somatic embryos. Differences found among the explants for their capacity to form embryogenic structures and germinate could be due to their intrinsic genetic characteristics. Our results support the thought that the

capacity to produce embryogenic tissue in cassava varies among different accessions. Furthermore, the solid system used could be forming gradients in the uptake of nutrients and hormones promoting variation and reducing the growth of the embryogenic structures.

RESUMEN

La diversidad genética y diferenciación de 24 accesiones de yuca (*Manihot esculenta* Crantz) de la colección puertorriqueña fueron estimados en este estudio. Adicionalmente, 63 accesiones de la colección de germoplasma del CIAT (Centro Internacional de Agricultura Tropical) que representan Latino América (Colombia, Costa Rica, Guatemala, México, Brasil) y África (Benin, Camerún, Ghana, Nigeria y Sierra León) fueron incluidos a partir de estudios de diversidad de yuca realizados previamente. La variación en la frecuencia alélica de 37 loci no ligados correspondiente a marcadores de Secuencias Sencillas Repetidas en tandem (SSR, acrónimo en inglés) fue usada para estimar la diversidad genética y diferenciación, así como para descifrar las relaciones entre las 24 accesiones con sus correspondientes parientes cultivados provenientes de los centros de diversidad primaria y secundaria. Los marcadores SSR fueron escogidos porque muestran una amplia cobertura del genoma de la yuca y un polimorfismo (PIC) moderado a alto. Para todos los loci SSR estudiados se observó un 96.85% ± 6.86 de loci polimórficos entre países, un promedio de 4.58 ± 1.83 alelos por locus y una heterozigosidad media observada de 0.7 ± 0.055. A pesar del bajo nivel de diferenciación [F_{ST} (theta) = 0.047± 0.005] que se encontró entre países, la distancia genética (Distancia de Nei y F_{ST}) existente entre las accesiones es suficiente para separarlas de acuerdo con su país de origen. Los análisis de UPGMA entre países revelaron que las accesiones puertorriqueñas se agrupan con las de Suramérica. Adicionalmente con los SSR se detectaron dos accesiones genéticamente iguales que exhibían nombres diferentes y se identificaron los posibles parientes cultivados de algunas de las accesiones de la colección puertorriqueña provenientes de Brasil y Colombia.

Finalmente se evaluó la capacidad embriogénica de los cultivares presentes en la colección *in vitro* mediante el método de inducción convencional en sólido usando 8 mg/l de 2,4-D. Para ello se usaron como explantes hojas jóvenes aisladas de plantas crecidas en *in vitro*. Se logró la embriogénesis somática en el 58 % de los genotipos, pero solo las accesiones SG804, CM3064, Tremesiana, PI12900, PI12903 y Seda produjeron embriones germinados. Las diferencias observadas entre los explantes para formar estructuras

embriogénicas y germinar, puede atribuirse a las características genéticas intrínsecas de cada individuo, y en este sentido nuestros resultados apoyan la idea de que la producción de estructuras embriogénicas en la yuca varía entre los diferentes accesiones. Además, el sistema en sólido que se usó podría formar gradientes en la toma de nutrientes y de hormonas que estarían promoviendo la variación y reduciendo el crecimiento de las estructuras embriogénicas.

TO MY FAMILY AND THOSE YET TO COME

"The Lord in His wisdom made the fly
And then forgot to tell us why"
Ogden Nash

ACKNOWLEDGMENTS

During my growth as a graduate student in the University of Puerto Rico several persons and institutions collaborated directly and indirectly with my work. Without their support it would have been impossible for me to finish this research. That is why I would like to dedicate this section to recognize their helpful support and love.

First, I would like to express a sincere acknowledgement to my advisor, Dr. Dimuth Siritunga because he gave me the opportunity to research under his guidance and because he followed my steps and encouraged me to continue growing as an independent researcher and supported all my ideas. His ability to handle many tasks and his dedication have been an inspiration for me and will guide me in the future work. I also want to thank the inspiration and support I received from Dr. Martin Fregene (CIAT). To these two persons, I am completely grateful. Special thanks I owe to my committee, Dr. Duane Kolterman and Dr. Juan Carlos Martínez Cruzado for critically evaluating my research through these three years. I also wish to thank to Dr. Fernando Bird Picó for his helpful comments, dedication and transmitted knowledge in population genetics.

I would also thank to Wilson Castelblanco (Swedish University of Agricultural Sciences) for his useful assistance with all the computer packages I used to analyze the data. Special thanks to Charles Buitrago, Paula Hurtado and Graciela Mafla (CIAT) for providing us the cassava DNA samples from the cassava diversity kit and their backing in the SSR analysis. Their help and transmitted of knowledge were an important component in the development of this work. I would like to thank Mr. Agenol González from Corozal Agricultural Experiment Substation for supplying us the plant material for this work.

I am also grateful with my lab partners Josean Marrero, Morgan Echeverry, María del Milagro Montero, Amaury Ortíz, Ruth Almodovar and José Mari for their suggestions, technical assistance and for their friendship.

Finally, but not least I thank to my family, especially to Ligia, Patricia, Mercedes and José Luis for their love and unconditional support. Without their constant encouragement I

would not have accomplished this goal. I would also like to extend my gratitude to my friends Erika Stechauner, Carlos Echeverri, and Carlos E. Cuadros for their friendship, camaraderie, love and support during all these years.

Page

Table of Contents

ABSTRACT ii
RESUMEN iv
DEDICATION..... vi
ACKNOWLEDGMENTS vii
LIST OF TABLES..... xi
LIST OF FIGURES..... xii
CHAPTER I..... 1
1. Introduction to cassava 1
1.1 Importance of an ancient crop..... 1
1.2 Biology, origin and genetic diversity of cassava..... 2
1.2.1 Biology 2
1.2.2 Origin 2
1.2.3 Genetic diversity 3
1.3 Introduction of cassava to Puerto Rico 4
1.4 Cassava SSR markers 5
1.5 Diversity and differentiation: measures of the genetic variation 7
1.6 Cassava breeding and transgenic technologies 9
1.7 Economic importance 10
1.8 Motivation of this study and objectives 10
CHAPTER II..... 12
2. SSR markers and genetic diversity assessment of cassava (*Manihot esculenta* Crantz) of the Puerto Rican germplasm collection 12
2.1 Introduction 12
2.2 Materials and methods 14
2.2.1 Plant materials 14
2.2.2 DNA sources, SSR amplification and analysis 15
2.2.3 Genetic and statistical analysis of SSR data 22
2.3 Results 24
2.3.1 Number of alleles in SSR loci 24
2.3.2 Genetic diversity 26
2.3.3 Genetic differentiation between Puerto Rican, Latin American and African cassava accessions 26
2.3.4 Genetic similarity among accessions and cluster analysis 32
2.4 Discussion 37
CHAPTER III 40
3. Tissue culture and somatic embryogenesis of cassava (*Manihot esculenta* Crantz, Euphorbiaceae) accessions of Puerto Rico..... 40
3.1 Introduction 40
3.2 Materials and methods 42

3.2.1 Plant material and micropropagation ……..……... 42
3.2.2 Induction of somatic embryogenesis ……..…… 43
3.3 Results ……. 44
3.3.1 Cassava micropropagation ……. 44
3.3.2 Induction of somatic embryogenesis and embryo germination ……. 45
3.4 Discussion ……..……. 50
4. **CONCLUSIONS**……..…….. 51
5. **SUGESTIONS AND PERSPECTIVES**…..……..……... 53
6. **BIBLIOGRAPHY**…….……. 54
APPENDIX A ……..…….. 63
APPENDIX B ……..…….. 65
APPENDIX C ……..…….. 66
APPENDIX D……..…….. 69
APPENDIX E ……..…….. 71
APPENDIX F ……..…….. 73
APPENDIX G……..…….. 80
APPENDIX H ……..…….. 82

List of tables

Page

Table 1. Cassava DNA obtained directly from CIAT collection..... 17-18

Table 2. Properties of cassava SSR loci and their primer pairs modified at the 5' end with the M13 tail sequence……..... 19-20

Table 3. Number of alleles per SSR locus for the cassava accessions from different countries……..... 25

Table 4. Genetic diversity of cassava accessions from different countries origins 27

Table 5. Average pairwise genetic distance between cassava populations using Nei's distance (Nei 1978) ……... 28

Table 6. Pairwise estimator of F_{ST} (theta) between pairs of country groups of cassava accessions……... 30

Table 7. Similarity coefficients for the cassava accessions found as closest relatives……..... 36

Table 8. Number of cassava nodes produced per shoot by each genotype after two months growing in 4E medium …….. 45

Table 9. Cassava embryogenic cultures growing as clumps or single embryos on young leaf explants……..……...……. 48

Table 10. Dellaporta et al. (1983) DNA extraction buffer recipe..... 64

Table 11 . PCR master mix used to amplify SSR loci…… 65

Table 12. Thermocycling program ……..…….. 65

Table 13. Standardized conditions to run cassava SSR loci…… 68

Table 14. Proportion of differentiation among populations (G_{ST}=coefficient of gene differentiation) computed according to Nei and Chesser (1983) and *F*-statistics computed according to Weir & Cockerham (1984) over loci ……. 72

Table 15. 4E medium for cassava micropropagation growth……. 80

Table 16. MS8 (8 mg/l 2,4-D) medium for embryogenesis induction in cassava. 80

Table 17. RM1 medium for the germination of somatic embryos……..…... 81

Table 18. Embryogenesis and germination results…….. 82

List of figures

Page

Figure 1. Genomic DNA of cassava *in vitro* plants isolated using the modified protocol of Dellaporta et al. (1983) ……..... 15

Figure 2. M13-tailed primer method used here is based on a forward primer for each microsatellite primer set designed with a 5'-tail that has the sequence 5'-CACGACGTTGTAAAACGAC-3'…… 21

Figure 3. Categorical (a) and binary (b) matrices made based on the allele sizes matrix……..... 21

Figure 4. Number of SSR alleles per locus and their frequency in cassava accessions from Puerto Rican collection, and some selected samples from the world's cassava collection from Latin America and Africa……..…..……. 24

Figure 5. Unweighted pair-group method with arithmetic averaging (UPGMA) dendrogram illustrating the genetic relationships among cassava accessions grouped by country of origin as revealed by pairwise genetic distance (Nei 1978) 29

Figure 6. UPGMA dendrogram of the pairwise fixation index F_{ST} between cassava accessions grouped by country of origin……..……. 31

Figure 7. Scatter plot of SSR marker diversity in cassava accessions was derived from plotting the first two principal coordinates, which were computed from Jaccard's similarity matrix.… ……..... 33

Figure 8. Cluster analysis for SSR assessment of genetic distance among 86 cassava accessions using Euclidean distance and UPGMA grouping method…..... 35

Figure 9. Cassava plants at the Corozal Experimental Station field (a) and grown in pots at the Department of Biology of the University of Puerto Rico (b)……. 43

Figure 10. *In vitro* cassava explants growing in tubes and magenta boxes…… 44

Figure 11. Cassava friable calli and somatic embryos in mature torpedo shape developing on the embryogenic tissue (a) and somatic embryos in globular shape (b) (200X) …….. 46

Figure 12. Percentage of cassava somatic embryos developed on the surface calli by the 23 accessions from Puerto Rican collection and one control (MCol2215) under the effect of 8 mg/l of 2,4-D treatment…..... 47

Figure 13. Germinating cassava somatic embryos of SG804 growing on RM1…. 49

Figure 14. Percentage of cassava embryo germination on RM1 medium for the cassava 23 accessions and the control 49

CHAPTER I

1 Introduction to Cassava

1.1 Importance of an ancient crop

Historically, cassava, *Manihot esculenta* Crantz, has played an important role on the world's agricultural stage. This starchy root is grown especially in the lowlands of developing countries, where it is a staple food of some of the most hard-pressed populations in the world. According to the Consultative Group on International Agricultural Research (CGIAR) system, cassava is the most important food crop in sub-Saharan Africa, where the average consumption exceeds 300 kg per capita per year, and also feeds more than half a billion people in the world (Mba et al. 2001, Fregene et al. 2001, 2003). By the year 2050, 90% of humankind will live in developing countries, where agriculture remains the most important economic activity (Raven et al. 2006). In that way, cassava offers an opportunity to solve the problem of the calorie requirements and food security in those areas, and also will be one of the most promising crops in terms of growth and new market opportunities (Tonukari 2004). Scott et al. (2000) cited by Siritunga (2002) estimated that the value of cassava grown in the developing countries was US $8.8 billion per year (1995-1997) and will increase to US $14 billion by the year 2020. In spite of its importance as a major food crop, it is the least genetically understood of any of the major staple crops (including rice, maize, wheat, and potatoes) that feed mankind (Fregene et al. 1997, 2001). Developments towards understanding the genetics of important agronomic traits in cassava have been made with the creation of a molecular genetic map (Fregene et al. 1997), the isolation and characterization of microsatellites (SSR) and other molecular markers (Chavarriaga-Aguirre et al. 1998; Mba et al. 2001) and the genetic transformation and tissue culture propagation (Siritunga and Sayre 2003; Siritunga et al. 2004). These tools are currently being used to achieve the same goal as through breeding but in a faster and cost-effective way.

1.2 Biology, origin and genetic diversity of cassava

1.2.1 Biology

Cassava (*Manihot esculenta* Crantz) is one of the few cultigens in the family Euphorbiaceae. The genus *Manihot* comprises 98 species spread throughout the Neotropics, 17 of which are native to North America and the others to South America (Rogers and Appan 1973). Each of the more than 10,000 varieties of cassava has its own distinctive plant form, genetic structure, and adaptability to different environments. It is a perennial, woody shrub that grows from about 1 m to about 3 m (3 to 9 feet) tall. The woody stems are topped by hand-shaped, dark green leaves (sometimes purplish when they are young). Flowering varieties show hypogynous, actinomorphic and unisexual flowers with a trilocular ovary of three carpels. Male and female flowers grow on the same plant (monoecious plant) and are cross-pollinated by insects. The female flowers mature earlier than the male flowers (protogyny). Fruits are dehiscent and seed production is low and erratic. Seeds are oval, about 10 mm long, and mottled brown and gray. The cone-shaped roots vary in size and number, according to the variety and environmental conditions. Normally, they have a dark or light brown bark epidermis, which often peels off, leaving a pink or cream cortex. The starch extracted from its roots is used for human and animal feeding and is also used to make a wide range of sweet and savory foods (Best and Hargrove 1994; CIAT 2002; Judd et al. 1999; Perry 1943).

1.2.2 Origin

Cassava has been cultivated in tropical America for over 10,000 years (Rogers 1963 cited by Fregene et al. 2003). Before 1600, during the European exploration of America, Portuguese traders took it to Congo and later to eastern Africa and Asia. By the end of the 19th century, it was growing throughout the tropical world (Best and Hargrove 1994) ranking fourth in importance as a dietary staple and sixth on an overall agricultural basis (Cock 1985).

According to Rogers and Appan (1973), the genus *Manihot* originates in South America and Mesoamerica, where there are two centers of diversity, in Brazil and Mexico. Apart from the *Manihot esculenta* species there is little geographical overlap between these two groups. Olsen and Schaal (1999) have demonstrated that cassava probably originated in wild *M. esculenta* populations growing along the southern rim of the Amazon Basin in Brazil. They demonstrated that cassava haplotypes of the gene glyceraldehyde-3-phosphate dehydrogenase (*G3pdh*) are present in natural populations from other regions of Brazil. Their findings suggested that cassava is closely related to populations of *M. esculenta* subsp. *flabellifolia*. Like most phylogeographic analysis, the study by Olsen and Schaal (1999) was based on a single locus, providing a single estimate of phylogenetic relationships in the study; therefore, microsatellite studies were utilized in 2001 by Olsen and Schaal in order to address the question about cassava's origin. Those data corroborated the findings of the phylogenetic study, thus, the southern ecotone was identified as the crop's likely geographical origin of domestication (Olsen and Schaal 2001). On the other hand, the savannas of Colombia, Venezuela, Guatemala and southern Mexico that had previously been proposed as likely places of origin for cassava due to the large number of landraces present there, now are considered as secondary centers of domestication of the crop (Fregene et al. 2003).

1.2.3 Genetic diversity

The vegetative propagation and the spread of cassava in the world should have shown many founder events with the concomitant effect of reduced genetic diversity over time, because of the accumulation of pathogens and the spread of a few, vigorous, well adapted landraces. Nonetheless, levels of molecular marker diversity observed for cassava accessions from Africa and several Neotropical countries are comparable with those from Brazil (Beeching et al. 1993, Fregene et al. 2003). Using SSR markers, Fregene et al. (2003) suggested that spontaneous recombination and farmer selection from volunteer seedlings of new varieties maintains the high levels of genetic diversity in the African field collections. These findings were unexpected because it is known that cassava was probably domesticated from populations of *M. esculenta* subsp. *flabellifolia* along the southern rim of the Amazon

Basin within the last 10,000 years (Olsen and Schaal 2001). However, the agricultural practices and the allogamous nature of cassava have been demonstrated to produce a large pool of volunteer seedlings that natural and human selection can act on to produce new varieties. The production of new varieties not only maintains a high level of genetic diversity, but also serves as insurance against crop failure due to biotic and abiotic stresses. Fregene et al. (2003) findings also demonstrated the separation between Neotropical and African accessions and a more pronounced substructure in the African ones. This divergence of the African accessions seems to be related with the presence of the cassava mosaic disease (CMD), which does not occur in the Neotropics. This event may have served as a selection pressure toward resistant genotypes and therefore new variability (Fregene et al. 2003).

1.3 Introduction of cassava to Puerto Rico

The cassava collection present at the Corozal Experimental Station in Puerto Rico consisted of 23 accessions, seven of them were introduced to Puerto Rico by the Tropical Agriculture Research Station (TARS) in 1994 from the International Center for Tropical Agriculture (CIAT, from its acronyms in Spanish). Those clones are: CM523, CM 3064, CM 3311, CM 3380, CM4484, SG804, SM494. Other accessions have been in the cassava collections of the Agriculture Experimental Stations of Puerto Rico (EEA, from its acronym in Spanish) for approximately 26 years (A. González, personal communication). Among them are: Abuelo, Brava, Chilena, Cubana, Forastera, Jamaica 18, Llanera, PI 12900, PI 12902, PI 12903, Seda, Senon, Serrallés, Tremesiana, Trinidad 14-56, and Valencia. Additional to these 23 accessions, a new accession called Amarillo was added from Gurabo EEA, since 2008 (A. González, personal communication). The identification of desirable accessions for diverse end uses is the target of many cassava breeding programs. Therefore, identification of accessions within the Corozal collection is a necessity in order to maintain a cassava core collection for the development of an effective germplasm conservation program.

1.4 Cassava SSR markers

In a crop where cultivars have been selected and distributed by the farmers themselves and where different accessions can often possess the same name (in addition to the same accession bearing more than one name), an objective method of identification, classification and measurement of genetic diversity is required (Beeching et al. 1993; Chavarriaga-Aguirre et al. 1998, 1999; Haysom et al. 1994). To achieve this goal, the assessment of the genetic variability of germplasm collections requires highly polymorphic markers as well as high-throughput genotyping systems. Simple sequence repeat (SSR) markers are particularly attractive to study the genetic variability because they are abundant and widely distributed in plant genomes. They also show high levels of polymorphism based on the number of repeat units found for each locus in any given population, and are adaptable to automation (Morgante & Olivieri 1993; Fregene et al. 2003). Cassava SSR markers have been useful in searching for duplicates in the CIAT's core collection (Chavarriaga-Aguirre et al. 1998) and in the analysis of variation in natural populations of putative progenitors of cassava (Olsen and Schaal 2001). These studies have contributed to the establishment of relations between accessions or cultivated genepools, thus facilitating the development of cassava cultivars to satisfy the market standards (Chavarriaga-Aguirre et al. 1998) and also have been improved to respond to some biotic and abiotic stresses (CIAT 2002). All eukaryote genomes are densely interspersed with simple sequences which consist of stretches of tandemly repeated nucleotide motifs that can be as short as 1-5 nucleotides. These elements can also be found to a lesser extent in prokaryotic genomes. Litt and Lutty (1989) cited by Mba et al. (2001) were the first to call those sequences microsatellites, but later, Jacob et al. (1991) cited by Mba et al. (2001) called them SSRs (Simple Sequence Repeats). When such regions are individually amplified by means of the Polymerase Chain Reaction (PCR), using a pair of flanking unique oligonucleotides as primers, they almost invariably show extensive polymorphism due to site-specific length variation, which occurs as a consequence of different numbers of repeat units (Morgante and Olivieri 1993). In eukaryotes one can expect to encounter at least one simple sequence stretch every 10 kb of DNA sequence. Also, SSRs have been correlated with regions of high divergence on an evolutionary time scale. On the basis of these facts,

scientists have proposed that simple sequences do not generally provide a defined function for the genome, but simply reflect internal genomic mechanisms, which have the tendency to dynamically produce and delete these sequences. This assumption predicts that simple sequence stretches should be generally hypervariable in length (Tautz 1989). SSR loci are highly variable on account of the number of repeat units found for each locus in any given population (Morgante and Olivieri 1993). The high levels of heterozygosity, the PCR-based nature of these repeat loci and their codominant nature have made SSRs the molecular markers of choice for genetic mapping and diversity studies (Tautz 1989; Mba et al. 2001). Currently, microsatellites have been evaluated for many crop plants, including, sugarbeet, Brassicas, melon, tomato, pepper, cotton, sorghum, maize, rice, soybean, apple, among others (*Plant & Animal Genome IX Conference* http://www.intl-pag.org/9/abstracts/P3b_13.html> checked: April 19, 2006). SSRs also have been produced for other angiosperm plants, such as *Eucalyptus pilularis, E. globulus, E. grandis, E. cloeziana, Grevillea robusta, Corymbia variegata, Toona ciliata*, and for gymnosperms, such as *Araucaria cunninghamii, Pinus elliottii, P. caribaea* (Scott et al. 1999). Also, SSR markers are now being applied to plant mapping projects in many species to saturate the existing genetic maps. For example, in wheat (University of Missouri http://www.agron.missouri.edu/mnl/76/77coe.html> checked: October 2, 2008), oat (University of Saskatchewan, Crop Molecular Genetics Laboratory <http://www.usask.ca/agriculture/plantsci/cmg/oatMMP.htm> checked: October 2, 2008), bean (CIAT <http://www.ciat.cgiar.org> checked: October 2, 2008), rice, barley, rye, pearl millet, maize, sorghum, sugarcane, tomato, potato, wheat, soybean, mungbean, cowpea, pea, chickpea, lentil, lettuce, cotton, coffee, papaya, sugar beet, among others (Kumar 1999).

The discovery, inheritance and variability of hundreds of SSR markers have been described in cassava (Chavarriaga-Aguirre et al. 1998; Mba et al. 2001). In 1998, Chavarriaga-Aguirre and collaborators isolated fourteen SSRs containing GA repeats. A subset of four of those SSR markers was used to evaluate the genetic diversity of the core collection, about 600 accessions in the cassava world germplasm bank at CIAT. The results obtained by Chavarriaga et al. (1998) showed high levels of heterozygosity (up to 0.88) with the markers and also showed the existence of duplicates in the cassava core collection. Then

in 2001, Mba et al. (2001) isolated and characterized 182 SSR markers in cassava in order to saturate the existing genetic map. Since 2001, new SSR molecular markers have been developed and introduced to saturate the existing map (Zárate 2002). The new SSR markers have been used to construct other linkage genetic maps using F2 progenies enabling the search for Quantitative Trait Loci (QTL) (Marín 2002) and to assess the genetic diversity of the crop (Castelblanco et al. 2004; Fregene et al. 2003).

1.5 Diversity and differentiation: measures of the genetic variation

Diversity and differentiation are estimators of the genetic variation broadly used in population genetics studies. They differ in that diversity is a measure of individual variation within a population and is widely agreed to reflect the number of different types in the population, taking into account their frequencies. In contrast, differentiation measures variation between two or more populations, demes or subpopulations (Gregorius 1987). There are many diversity estimators which are based on specific models and probabilistic concepts. Among them, heterozygosity is the most widespread measure of genetic variation within a population. Expected heterozygosity (*He*), also known as gene diversity, is a concept introduced by Nei (1978) to explain the probability by which two alleles that were arbitrarily selected are different, and is particularly useful because it is applicable for genes of different ploidy levels and in organisms of different reproductive systems (Hedrick 2005). Natural selection and genetic drift favor the genetic differentiation between populations which is associated to high levels of endogamy and low genetic diversity. On the other hand, gene flux enhances the homogenization of the genetic diversity between populations, which is associated to low differentiation and high genetic diversity. Several approaches have been used to estimate the amount of differentiation in the subdivisions of a population. Wright (1951, 1965) cited by Hedrick (2005) developed an approach to partition the genetic variation in a subdivided population that is commonly used and provides an obvious description of differentiation. This consist of three different F coefficients (correlation coefficients that are different from the F statistics used in the analysis of variance) used to allocate the genetic variability to the total population level (T), subpopulations (S), and individuals (I). These three coefficients, F_{ST}, F_{IT} and F_{IS} are interrelated so that

$$F_{ST} = F_{IT}-F_{IS}/(1-F_{IS})$$

F_{ST} is a measure of the genetic differentiation over subpopulations and is always positive. F_{IS} and F_{IT} are measures of the deviation from Hardy-Weinberg equilibrium within subpopulations and in the total population.

To interpret the F_{ST} values, Wright (1978) suggested a qualitative guide in which F_{ST} values show a fluctuation:

- 0 to 0.05-low genetic differentiation
- 0.05 to 0.15-moderate genetic differentiation
- 0.15 to 0.25-high genetic differentiation
- 0.25 or more-highest levels of genetic differentiation.

Nei (1977) has shown how these coefficients can be expressed in terms of allele frequencies and observed and expected genotype frequencies, and also he extended the analysis of these coefficients to multiple loci as an estimate of F_{ST} while assuming Hardy-Weinberg proportions. Nei (1973) defined the new coefficient of gene differentiation as

$$G_{ST} = D_{ST}/ H_T = F_{ST}$$

Where $D_{ST} = H_T - H_S$ is the average genetic diversity within subpopulations, H_T is an average of the expected heterozygosity in the total population over loci and G_{ST} is the measure of the relative differentiation among subpopulations. The dependence of G_{ST} on the amount of genetic variation is true for highly variable loci such as SSR loci where H_S and H_T can approach unity (Nei 1987; Hedrick 2005).

These statistical estimators make a partition of the genetic variability, measured by levels of heterozygosity, into components of variability among and within populations. Furthermore they provide a way to estimate gene flux patterns among populations.

1.6 Cassava breeding and transgenic technologies

Plant breeding can be defined as the science of changing heredity of plants to improve their economic utility to man. Then, plant breeding is credited with all the useful changes in crop plants starting from domestication to the current modification at the molecular level. Cassava breeding programs focus on multiples traits that involve increased yield, control of pests and diseases through the development of resistant plants, increase of desirable traits according to consumer preferences such us early bulking of storage roots, combined with high dry matter content and low cyanide content (Chahal and Gosal 2002 cited by Kizito 2006). Ceballos et al. (2004) emphasized that given the recent start in cassava improvement, local accessions still play a more important role in cassava breeding than in other major crop breeding schemes. Today, molecular techniques have become important tools for basic and applied research in cassava. Genetic markers, are an example of how molecular tools have been used to understand the inheritance and diversity of natural variation in the crop (Beeching et al. 1993; Fregene et al. 1994, 2003, Roa et al. 1997; Olsen and Schaal 1999).

During the last few years the transformation of cassava plants having value-added traits has become a reality. Currently, two different transformation systems have been employed to engineer cassava. The first one uses *Agrobacterium*-mediated transformation of somatic embryos or friable embryogenic callus (FEC), while the second system uses biolistics (particle gun-mediated bombardment) to transform the same tissues. Cassava transformation experiments have been done mainly on four accessions: TMS 60444, SM 1219-9, CM 2306-4, and MCOL 2215 (CIAT <http://www.ciat.cgiar.org/biotechnology/crops_cassava.htm#cassava> checked: October 2, 2008). To date, *Agrobacterium*-mediated transformation of cassava has proven to be more successful than particle gun-mediated transformation (Siritunga and Sayre 2003; Siritunga et al. 2004). The genetic transformation of cassava is being used as a tool to support conventional breeding programs. This action is due to conventional breeding approaches being hampered by a variety of factors, including the heterozygous nature of the crop, low production of flowers, apomixis, its long growing cycle of 9-18 months, its low seed yield per pollination, and inbreeding depression caused by cassava's strong outcrossing monoecious habit (mediated by protogyny) (Fregene et al. 1997). Furthermore, cassava is commonly propagated by clonal stem cuttings. Therefore, all

these factors make cassava a suitable target for genetic improvement via transgenic approaches.

1.7 Economic importance

The starchy roots of cassava are mainly used for human consumption, but they are also used in many processed forms and products, such as starch and flour that has industrial and animal feed applications. In Africa, most of the cassava is used for food consumption with 50% in processed form and 38 % in the boiled form. The other 12 % is used for animal feed. In some sub-Saharan countries, the average consumption exceeds 300 kg per person per year. Africa is also the world's largest producer of cassava with a production of 90 million metric tons in 1999. Nigeria is the top cassava producing country, with an estimated total production capacity of 34 million metric tons (Phillips et al. 2004). In Latin America between 35 % and 40 % of the cassava production is mainly used for food consumption and a third of the cassava produced is for animal feed; the remaining percentage is used in industry. The starch has a significant use in Asia where cassava chips are exported to Europe for animal feed purposes (Westby 2002 cited by Kizito 2006). Because of its excellent adaptability to erratic rainfall and low-fertility soils, it has become a major dietary staple, a famine-reserve crop and a source of income (Fregene et al. 2003).

1.8 Motivation of this study and objectives

Despite the increasing demand and its economic importance in developing countries, cassava remains as an "orphan crop". This is the primary reason for continuing to research this crop with a goal of improving it.

In 2000, cassava production in Puerto Rico was almost 750,000 kg with an annual farm value of US $386,000. During the same year more than 6,800,000 kg of cassava tubers were imported, showing the need to increase production, reduce cost and meet local demand (Goenaga et al. 2002). At present, twenty-four cassava accessions are maintained at the University of Puerto Rico Agricultural Experimental Stations (UPR-EEA) of Corozal and

Isabela, but little is known about their genetic diversity and origin. With respect to this, the objectives of this thesis were:

- To introduce 23 accessions from UPR-EEA Corozal fields to sterile *in vitro* culture conditions, in order to maintain the cassava germplasm for a prolonged time.
- To access the genetic diversity and differentiation of cassava accessions present at Corozal collection based on the allele frequency at 37 SSR markers.
- To establish the relationships between Puerto Rican accessions and their cultivated relatives from primary and secondary centers of diversity.
- To measure the levels of embryogenesis of the Puerto Rican acessions compared with those of the model cultivar MCol2215.

CHAPTER II

2. SSR markers and genetic diversity assessment of cassava (*Manihot esculenta* Crantz) of the Puerto Rican germplasm collection

2.1 Introduction

Historically, cassava (*Manihot esculenta* Crantz) has played an important role on the world's agricultural stage. This starchy root is especially grown in the lowlands tropics, where it is a staple food of some of the most hard-pressed populations in the world (Mba et al. 2001). While being a staple food, cassava occupies fourth place after rice, sugar cane and maize as a source of calories in the human diet (Cock 1985). According to the Consultative Group on International Agricultural Research (CGIAR) system, cassava is the most important food crop in sub-Saharan Africa, where the average annual consumption exceeds 300 kg per capita per year, and also feeds more than half a billion people in the world (Mba et al. 2001, Fregene et al. 2001, 2003). With 90 % of humankind living in developing countries by the year 2050, where agriculture remains the most important economic activity (Raven et al. 2006), cassava offers an opportunity of fulfilling human calorie requirements as well as providing food security to people in those areas. In spite of its importance as a major food crop, it is the least genetically understood of any of the major staple crops that feed mankind (Fregene et al. 1997, 2001) and remains as an "orphan crop" despite its increasing demand and economic importance.

The origins of cassava have long been obscure. Until 1982 it was assumed that cassava has no known wild ancestry (Allem 1994; Heysom et al. 1994). The savannas of northeastern South America and Mesoamerica had been previously proposed as likely places of origin of cassava based on the relative abundance of wild species (Renvoize 1972; Rogers 1963). These evidences led to the development of two hypotheses about the cassava origin; the first one proposed by Rogers (1963) considered that cassava is a "compilospecies", in other words, it is the product of the introgression and hybridization process of many native species

of the *Manihot* genus. The second hypothesis proposed by Allem (1994) supports that wild populations of *M. esculenta* in Brazil are the origins of the crop. More recently, the phylogeographic studies made by Olsen and Schaal (1999) demonstrated that haplotypes of the gene G3*pdh* are present in natural populations of *M. esculenta* subspecies, giving credibility to the hypothesis proposed by Allem (1994), who showed that wild *M. esculenta* populations growing along the southern border of the Amazon basin are the early ancestors.

In 2000, cassava production in Puerto Rico was almost 750,000 kg with an annual farm value of US $386,000. During the same year more than 6,800,000 kg of cassava tubers were imported. This illustrates the need to increase the local cassava production, reduce cost and meet the local demand (Goenaga et al. 2002). Even though there are twenty-four cassava accessions in Puerto Rico that have been maintained at the UPR-EEA of Corozal and Isabela, little is known about their genetic diversity and origin. Recently, SSR markers have been used to study genetic diversity of crops because of they are particularly useful to study the variation in allele frequency at many unlinked loci, which is the preferred way to assess genetic diversity and differentiation. Additionally, SSR markers exhibit high levels of polymorphism, and seem to be somatically stable, inherited in a co-dominant Mendelian manner and their automation (Morgante and Olivieri 1993; Fregene et al. 2003). The identification of desirable genotypes for diverse end uses is the target of many cassava breeding programs. Therefore, identification of genotypes within Puerto Rican collection is a necessity in order to develop an effective management and conservation program for the cassava genetic resources.

The objectives of this study are to use the SSR information obtained from screening a collection of 24 cassava accessions present in Puerto Rico and 63 others representing Latin America and Africa diversity centers: (1) to determine genetic identities in order to characterize the germplasm resources present in Puerto Rico, necessary to identify duplication in the collection. (2) Examine the genetic diversity and differentiation derived from SSR data to recognize genetic relationships of Puerto Rican accessions with their cultivated relatives from primary and secondary diversity centers.

2.2 Materials and methods

2.2.1 Plant materials

Cassava landraces were collected from the Corozal Experimental Station of the University of Puerto Rico (latitude 18˚20'N, longitude 66˚31'W. altitude 185 m), located in the highland agricultural zone of Puerto Rico. Woody stakes (~ 30 cm) containing two to three nodes were planted in pots with sterile soil-sand mix and grown under screen-house conditions. Cuttings of three months old plants were then introduced to *in vitro* conditions by sterilizing the nodal segments with 20 % (v/v) household bleach (sodium hypochlorite solution) and 0.05 % (v/v) Tween 20. After washing with sterile deionized water 4 to 5 times, the explant tissues were cultured on semi-solid Murashige and Skoog (MS)-4E micropropagation medium (Murashige and Skoog 1962) [1X MS salts basal medium (4.3 g) supplemented with 2 % (w/v) sucrose, 0.04 mg/l of N6-benzylaminopurine (BAP), 0.05 mg/l of GA3, 0.02 mg/l of NAA, myo-inositol (100 mg/l), thiamine-HCl (1 mg/l), and 0.7 % of Agar (Fisher Scientific, Pittsburg, PA) at pH 5.7] for the maintenance of the plants for an extended period of time. Plants were partially shadded from light for two weeks to induce etiolation and then were fully exposed to light under a 12 h/day photoperiod (5 µmol photons m-2 s-1) at 27 ± 2˚C according to protocols described by Escobar (1991) and Siritunga and Sayre (2003) to induce the sprouting of the buds.

The names of the accessions collected at the UPR-EEA of Corozal appears as follow: PI12902, Jamaica 18, Cubana, PI12903, PI12900, Forastera, Llanera, Trinidad 14-56, Brava, Tremesiana, Seda, Serrallés, Senon, CM3064, CM3311, CM3380, CM4484, SG804, SM494, SM523, Abuelo, Chilena, Valencia and Amarillo.

2.2.2 DNA sources, SSR amplification and analysis

DNA isolation from Puerto Rican accessions was carried out according to Dellaporta et al. (1983) modified protocol, using 0.25 g of *in vitro* cassava leaves. Tissue was ground with sterile sea sand and a pestle in a 1.5 ml Eppendorf tube containing 600 µl of 0.1 M of Tris-HCl pH 8.0, 0.05M of EDTA, 0.5 M of NaCl, 1% of PVP, 1.6% of SDS, 0.07% of β-mercaptoethanol extraction buffer. The mix was incubated at 65 °C for 15 min with agitation every five min. Then the samples were placed at room temperature for two min followed by incubation with 250 µl of ice-cold 5M potassium acetate in ice for 20 min. After incubation the samples were centrifuged for 15 min at 13,000 rpm at room temperature, and the supernatant was transferred to another tube where one volume of ice-cold isopropanol was added and mixed gently. The mix was kept at -20 °C for one hour and then centrifuged for 10 min at 13,000 rpm. Supernatant was removed and the pellet was washed with 70 % of ethanol, followed by a brief centrifugation of five minutes at 13,000 rpm. DNA was air dried and later resuspended in 50 µl of TE 10:1 (10 mM: 1 mM) buffer and 1 µl of RNAse A 10mg/ml (Appendix A). Between 0.5 µg and 1 µg of high quality DNA was obtained from each extraction and quantified by spectrophotometry (Figure 1). Samples were diluted to 0.01 µg/µl with deionized distilled water for the PCR amplification.

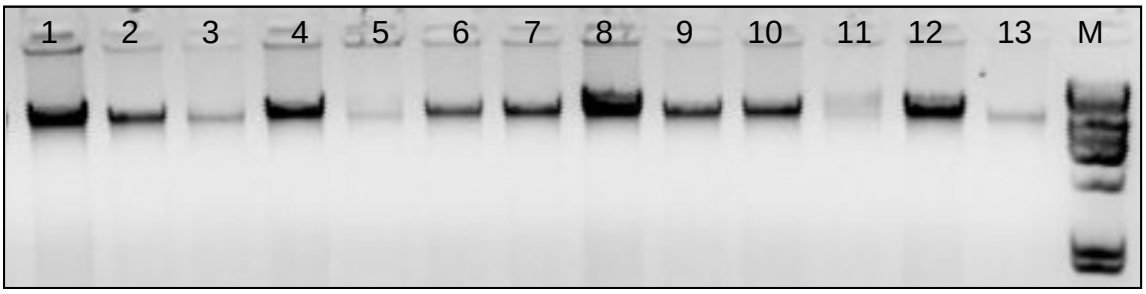

Figure 1. Genomic DNA of cassava *in vitro* plants isolated using the modified protocol of Dellaporta *et al*. (1983). The numbers correspond with the serial codes given to Corozal plants: PI1903, Jamaica18, Cubana, PI12904, PI12900, Forastera, Llanera, Trinidad 14-56, Brava, Tremesiana, Seda, Serrallés and Senon; M = 1 Kb molecular size marker (Invitrogen Co.)

A subset of 63 DNA samples from accessions from Africa and Neotropics was obtained directly from CIAT. These DNA samples consisted of 9 from five different African countries, 26 from Brazil, 16 from Colombia, 4 from Costa Rica, 4 from Guatemala, 3 from Mexico and 1 from Ecuador. Forty-five accessions from the 63 were selected by CIAT because they represent the allelic diversity of the crop (Table 1) and have been designated as part of the reference microsatellite kit developed to analyze the genetic diversity in cassava (Hurtado et al. 2008).

A subset of 37 SSR markers was selected from a data base of SSRs developed at CIAT (Chavarriaga et al. 1998; Mba et al. 2001) because of their high polymorphic content and broad coverage of the cassava genome.

Table 1. Cassava DNA obtained directly from CIAT collection.

Accession*	Accession Holder	Accession Origin
tme204	IITA	Benin
tme200	IITA	Benin
bra158	CIAT	Brasil
bra376	CIAT	Brasil
bra32	CIAT	Brasil
bra126	CIAT	Brasil
bra150	CIAT	Brasil
bra156	CIAT	Brasil
bra164	CIAT	Brasil
bra210	CIAT	Brasil
bra216	CIAT	Brasil
bra354	CIAT	Brasil
bra370	CIAT	Brasil
bra376	CIAT	Brasil
bra426	CIAT	Brasil
bra432	CIAT	Brasil
bra501	CIAT	Brasil
bra508	CIAT	Brasil
bra739	CIAT	Brasil
bra849	CIAT	Brasil
bra906	CIAT	Brasil
bra969	CIAT	Brasil
bra1062	CIAT	Brasil
bra1150	CIAT	Brasil
bra1197	CIAT	Brasil
bra1275	CIAT	Brasil
bra1392	CIAT	Brasil
bra1396	CIAT	Brasil
tme145	IITA	Cameroon
tme190	IITA	Cameroon
mcol2215	CIAT	Colombia
col1905	CIAT	Colombia
col306	CIAT	Colombia
col1853	CIAT	Colombia
col2269	CIAT	Colombia
sm734-5	CIAT	Colombian-breeding
cm523-7	CIAT	Colombian-breeding
cm2177-2	CIAT	Colombian-breeding
cm3064-4	CIAT	Colombian-breeding
cm3171-8	CIAT	Colombian-breeding
cm3311-3	CIAT	Colombian-breeding
cm3380-7	CIAT	Colombian-breeding
cm3401-2	CIAT	Colombian-breeding
cm4484-2	CIAT	Colombian-breeding

*The accessions listed here were selected from previous cassava diversity studies because they represent the allelic diversity of the crop (Hurtado et al. 2008).

Table 1. (Cont.)

Accession	Accession Holder	Accession Origin
sg804-5	CIAT	Colombian-breeding
sm494-2	CIAT	Colombian-breeding
cr54	CIAT	Costa Rica
cr136	CIAT	Costa Rica
cr142	CIAT	Costa Rica
cr30	CIAT	Costa Rica
ecu47	CIAT	Ecuador
tme635	IITA	Ghana
gua11	CIAT	Guatemala
gua15	CIAT	Guatemala
gua18	CIAT	Guatemala
gua6	CIAT	Guatemala
mex41	CIAT	Mexico
mex20	CIAT	Mexico
mex58	CIAT	Mexico
tms60444	IITA	Nigeria
tme113	IITA	Nigeria
nga2	IITA	Nigeria
tme857	IITA	Sierra Leone

Table 2 shows the sequences of the primers, the modifications made to the forward primer adding the M-13 sequence at the 5'end (Oetting et al. 1995; Zho et al. 2002), and the band size ranges observed for each primer set. The PCR reactions were performed using a Mastercycler (Eppendorf) in a total volume of 12.5 µl containing 1X of Buffer B (500 mM KCl, 100 mM Tris-HCl pH 9.0), 2.0 mM of $MgCl_2$, 200 µM of dNTP mix, 0.1 µM each forward and reverse primers, 0.5 µM of M-13 fluorescence primer, 1 U of Taq DNA Polymerase (Fisher), 20 ng of template DNA, and 6.3 µl of ddH_2O. PCR conditions used were the following: 95 °C for 5 min, followed by 34 cycles of 30 s at 94 °C, 1 min at 45 °C or 55 °C, 1 min at 72 °C, and a final extension step of 5 min at 72 °C was added (Appendix B). The PCR product was diluted 1 to 10 with blue stop solution (LI-COR) and then denatured at 95 °C for 3 min before loading on 6.5 % polyacrylamide gels using an automated 4300 DNA analyzer (LI-COR) for high-throughput SSR data collection (Appendix C). Allele sizes were scored using the program Saga GT and a 50-350 bp molecular size ladder (LI-COR) (Figure 2). Two control DNA samples (Nigeria 2 and CM2177-2) were included for the scoring procedure because their amplification pattern is known for the SSR markers used here and because they have been used as controls in previous cassava diversity studies.

Table 2. Properties of cassava SSR loci and their primer pairs modified at the 5' end with the M13 tail sequence.

Name	Motif	M13 5'-Tail	Forward primer	Reverse primer	Annealing	Allele Sizes
SSRY4	GA(16)TA GA(3)	CACGACGTTGTAAAACGAC	ATAGAGCAGAAGTGCAGGCG	CTAACGCACACGACTACGGA	55	278-320
SSRY9	GT(15)	CACGACGTTGTAAAACGAC	ACAATTCATCATGAGTCATCAACT	CCGTTATTGTTCCTGGTCCT	55	267-293
SSRY12	CA(19)	CACGACGTTGTAAAACGAC	AACTGTCAAACCATTCTACTTGC	GCCAGCAAGGTTTGCTACAT	55	264-284
SSRY19	CT(8)CA(18)	CACGACGTTGTAAAACGAC	TGTAAGGCATTCCAAGAATTATCA	TCTCCTGTGAAAAGTGCATGA	55	203-237
SSRY20	GT(14)	CACGACGTTGTAAAACGAC	CATTGGACTTCCTACAAATATGAAT	TGATGGAAAGTGGTTATGTCCTT	55	146-188
SSRY21	GA(26)	CACGACGTTGTAAAACGAC	CCTGCCACAATATTGAAATGG	CAACAATTGGACTAAGCAGCA	55	180-214
SSRY34	GGC(5)GGT GGC GGT(2)	CACGACGTTGTAAAACGAC	TTCCAGACCTGTTCCACCAT	ATTGCAGGGATTATTGCTCG	55	288-306
SSRY47	CA(17)	CACGACGTTGTAAAACGAC	GGAGCACCTTTTGCTGAGTT	TTGGAACAAAGCAGCATCAC	55	200-290
SSRY51	CT(11)CG CT(11) CA(18)	CACGACGTTGTAAAACGAC	AGGTTGGATGCTTGAAGGAA	GGATGCAGGAGTGCTCAACT	55	277-315
SSRY52	GT(19)	CACGACGTTGTAAAACGAC	GCCAGCAAGGTTTGCTACAT	AACTGTCAAACCATTCTACTTGC	55	266-284
SSRY59	CA(20)	CACGACGTTGTAAAACGAC	GCAATGCAGTGAACCATCTTT	CGTTTGTCCTTTCTGATGTTC	55	146-202
SSRY63	GA(16)	CACGACGTTGTAAAACGAC	TCAGAATCATCTACCTTGGCA	AAGACAATCATTTTGTGCTCCA	55	303-315
SSRY64	CT(13)CG CT(6)	CACGACGTTGTAAAACGAC	CGACAAGTCGTATATGTAGTATTCACG	GCAGAGGTGGCTAACGAGAC	55	192-220
SSRY69	CT(18)ATT AT(2) CTTTCTT CTTT(2)CCTTCT	CACGACGTTGTAAAACGAC	CGATCTCAGTCGATACCCAAG	CACTCCGTTGCAGGCATTA	55	213-267
SSRY82	GA(24)	CACGACGTTGTAAAACGAC	TGTGACAATTTTCAGATAGCTTCA	CACCATCGGCATTAAACTTTG	55	166-208
SSRY100	CT(17)TT CT(7) CCCT	CACGACGTTGTAAAACGAC	ATCCTTGCCTGACATTTTGC	TTCGCAGAGTCCAATTGTTG	55	209-273
SSRY102	(GT)11	CACGACGTTGTAAAACGAC	TTGGCTGCTTTCACTAATGC	TTGAACACGTTGAACAACCA	55	198-200
SSRY103	GA(22)	CACGACGTTGTAAAACGAC	TGAGAAGGAAACTGCTTGCAC	CAGCAAGACCATCACCAGTTT	55	274-308
SSRY106	CT(24)	CACGACGTTGTAAAACGAC	GGAAACTGCTTGCACAAAGA	CAGCAAGACCATCACCAGTTT	55	268-302
SSRY108	CT(24)CCT	CACGACGTTGTAAAACGAC	ACGCTATGATGTCCAAAGGC	CATGCCACATAGTTCGTGCT	55	192-228
SSRY120	CA(7)AA GA(8)GGA	CACGACGTTGTAAAACGAC	TCACCGTTAATTGTAGTCTGCG	GCGAGGTTCAAATATGCGAT	55	150-174
SSRY135	(CT)16	CACGACGTTGTAAAACGAC	CCAGAAACTGAAATGCATCG	AACATGTGCGACAGTGATTG	45	183-277
SSRY147		CACGACGTTGTAAAACGAC	GTACATCACCACCAACGGGC	AGAGCGGTGGGGCGAAGAGC	45	118-136
SSRY148		CACGACGTTGTAAAACGAC	GGCTTCATCATGGAAAAACC	CAATGCTTTACGGAAGAGCC	45	128-138
SSRY151		CACGACGTTGTAAAACGAC	AGTGGAAATAAGCCATGTGATG	CCCATAATTGATGCCAGGTT	45	194-234
SSRY155		CACGACGTTGTAAAACGAC	CGTTGATAAAGTGGAAAGAGCA	ACTCCACTCCCGATGCTCGC	55	163-175
SSRY161	CT(11)TT CT(21)CA(19)	CACGACGTTGTAAAACGAC	AAGGAACACCTCTCCTAGAATCA	CCAGCTGTATGTTGAGTGAGC	55	188-256

Table 2. (Cont.)

Name	Motif	M13 5'-Tail	Forward primer	Reverse primer	Annealing	Allele Sizes
SSRY164	GA(29)	CACGACGTTGTAAAACGAC	TCAAACAAGAATTAGCAGAACTGG	TGAGATTTCGTAATATTCATTTCACTT	45	156-204
SSRY175	GA(38)	CACGACGTTGTAAAACGAC	TGACTAGCAGACACGGTTTCA	GCTAACAGTCCAATAACGATAAGG	55	100-156
SSRY177	CCT(6)CTN(65)CT(4)AT CT(18)	CACGACGTTGTAAAACGAC	ACCACAAACATAGGCACGAG	CACCCAATTCACCAATTACCA	45	244-286
SSRY179	GA(28)	CACGACGTTGTAAAACGAC	CAGGCTCAGGTGAAGTAAAGG	GCGAAAGTAAGTCTACAACTTTTCTAA	55	195-253
SSRY180	GA(16)G(4) GA(5)	CACGACGTTGTAAAACGAC	CCTTGGCAGAGATGAATTAGAG	GGGGCATTCTACATGATCAATAA	55	131-145
SSRY181	GA(22)G(3)C GA(3) GGAA GA(4)	CACGACGTTGTAAAACGAC	GGTAGATCTGGATCGAGGAGG	CAATCGAAACCGACGATACA	55	192-216
SSRY182	CA(17)N(31)GAGG GA(8)	CACGACGTTGTAAAACGAC	GGAATTCTTTGCTTATGATGCC	TTCCTTTACAATTCTGGACGC	55	238-258
NS189		CACGACGTTGTAAAACGAC	TGGGCTGTTCGTGATCCTTA	CATGAGTTTAAAAATTATCACATCCG	55	106-124
NS376		CACGACGTTGTAAAACGAC	TCAAGACCCTTGCTTTGGTT	GGACTATCAAGGCGCAAAAG	55	213-233
NS911		CACGACGTTGTAAAACGAC	TGTTGTTCAGACGATGTCCAA	TTGAAGCAGTTATGAACCGT	55	135-149

A di-allelic model of inheritance was adopted and markers showing three or more alleles per individual were eliminated. The band sizes were exported directly into Microsoft Excel (Microsoft Inc) for further formatting as input files for statistical analysis (Figure 3).

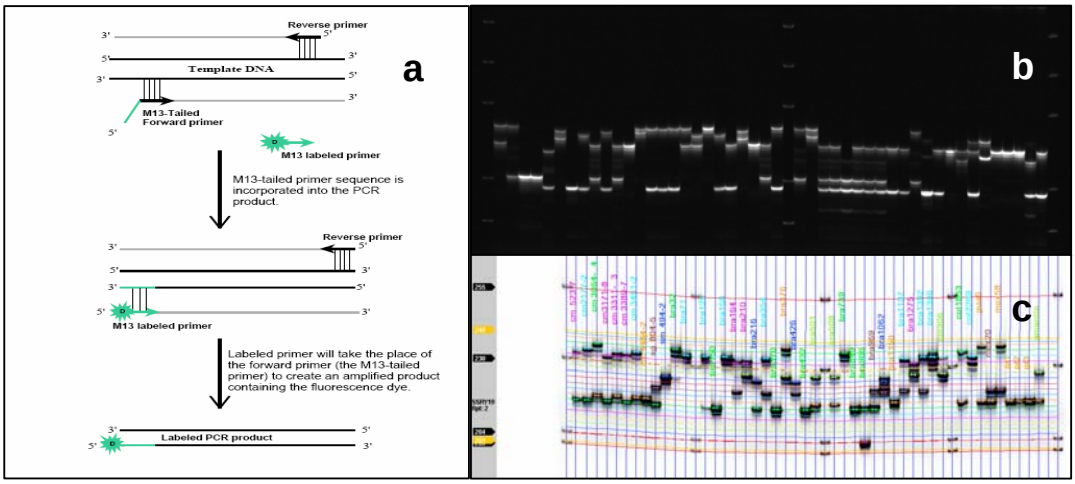

Figure 2. M13-tailed primer method used here is based on a forward primer for each microsatellite primer set designed with a 5'-tail that has the sequence 5'-CACGACGTTGTAAAACGAC-3' (a). SSR gel image depicting the reaction products from PCR amplifications made from genomic DNA of 47 cassava accessions using the SSRY180 primer pair (b). Genotyping samples using Saga GT software (c).

a b

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		SampleID	SSRY4	SSRY4	SSRY9	SSRY9	SSRY12	SSRY12	SSRY19	SSRY19	SSRY20	SSRY20	SSRY21	SSRY21
1		SampleID	SSRY4	SSRY4	SSRY9	SSRY9	SSRY12	SSRY12	SSRY19	SSRY19	SSRY20	SSRY20	SSRY21	SSRY21
2	1	pi12902	11	4	3	1	3	1	14	11	13	2	9	5
3	2	jamaica18	11	11	4	4	4	2	9	3	13	12	8	3
4	3	cubana	12	12	4	1	3	1	6	6	12	6	10	9
5	4	pi12903	12	12	4	1	3	1	7	6	12	6	10	9
6	5	pi12900	7	3	4	4	3	1	11	3	7	4	10	2
7	6	forastera	12	12	4	2	3	1	8	4	16	12	5	2
8	7	llanera	2	1	2	1	4	1	8	2	12	12	10	2
9	8	trinidad1456	11	8	2	1	6	6	3	1	6	2	10	5
10	9	brava	12	8	2	1	3	1	11	6	12	6	10	9
11	10	tremesiana	11	8	4	2	6	3	7	6	12	11	5	5
12	11	seda	12	8	2	1	3	1	11	6	12	6	10	9
13	12	serralles	5	4	2	1	2	1	11	4	12	11	5	2
14	13	senon	11	11	6	2	6	3	7	6	6	2	7	5

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	1	298	87	9										
2		pi12902*	jamaica18*	cubana*	pi12903*	pi12900*	forastera*	llanera*	trinidad1456*	brava*	tremesiana*	seda*	serralles*	senon*
3	1	0	0	0	0	0	0	1	0	0	0	0	0	0
4	2	0	0	0	0	0	0	1	0	0	0	0	0	0
5	3	0	0	0	0	1	0	0	0	0	0	0	0	0
6	4	1	0	0	0	0	0	0	0	0	0	0	1	0
7	5	0	0	0	0	0	0	0	0	0	0	0	1	0
8	6	0	0	0	0	0	0	0	0	0	0	0	0	0
9	7	0	0	0	0	1	0	0	0	0	0	0	0	0
10	8	0	0	0	0	0	0	0	1	1	1	1	0	0
11	9	0	0	0	0	0	0	0	0	0	0	0	0	0
12	10	0	0	0	0	0	0	0	0	0	0	0	0	0
13	11	1	1	0	0	0	0	0	1	0	1	0	0	1
14	12	0	0	1	1	0	1	0	0	1	0	1	0	0

Figure 3. Categorical (a) and binary (b) matrices made based on the allele sizes matrix.

2.2.3 Genetic and statistical analysis of SSR data

Gene diversity parameters within and among country groups were estimated utilizing the 34 SSR markers data, chosen for their diallelic nature. The software package GEN-SURVEY (Vekemans and Lefebvre 1997) was used for the following statistics analysis: percentage of polymorphic loci, mean number of alleles per polymorphic locus, average observed heterozygosity (Ho), and average gene diversity (He) (Nei 1978). For all loci and accessions, the total heterozygosity (H_T=total heterozygosity in the entire data set) was partitioned into within population diversity (H_S=heterozygosity within populations averaged over the entire data set) and between population diversity (D_{ST}) estimates, where $H_T = H_S + D_{ST}$. The proportion of among-accession differentiation (G_{ST}) was estimated according to Nei (1978). Standard deviations for the above parameters were estimated over loci and sampled by jackknifing (200 replications) and 95 % confidence intervals sampled by bootstrapping (1000 bootstraps over loci).

Pairwise genetic distances between pairs of populations were calculated with the software GEN-SURVEY using the following statistics: original Nei's distance (Nei 1972) and corrected Nei's distance (Nei 1978) for small sample sizes.

Genetic differentiation between pairs of country groups was estimated over all loci using F-statistics (Wright 1965) and pairwise calculations of F_{ST} (Weir and Cockerham 1984) using the software FSTAT 2.9.3.2 (Goudet 1995) and GEN-SURVEY. The pairwise F_{ST} values were used to construct a dendrogram based on the Unweighted Paired Group Method with Arithmetic mean (UPGMA) using the NTSYS-PC computer program (Rohlf 1993). Additionally, genetic disimilarity among country groups was evaluated using the UPGMA cluster method of Nei's genetic distances (Sneath and Sokal 1973) leading to the generation of a second dendrogram using the program NTSYS-PC.

Genetic distance among accessions was estimated by Principal Coordinate Analysis and clustering analysis based on UPGMA. For this purpose, DNA bands were scored as 1 for the presence and 0 for the absence of a particular DNA fragment to generate a binary matrix data which was analyzed using the program NTSYS-PC by selecting the computational module SIMQUAL which computes the association coefficient of Jaccard to construct a similarity matrix of the cassava individuals. The similarity matrix was then double-centered and factored to compute eigenvector and eigenvalue matrices that were

subjected to a principal coordinate analysis (PCoA) to evaluate the relationships among the cassava accessions. Further cluster analyses were performed to measure the degree of relatedness between accessions. For this purpose, a similarity matrix was performed based on the eigenvectors matrix and computing the Euclidian coefficient. The results were represented in a dendrogram based on the UPGMA method by selecting the Sequential and Hierarchical Numeric (SAHN) option. A brief description about each program used here is detailed in Appendix D.

2.3 Results

2.3.1 Number of alleles in SSR loci

A total of 34 of 37 SSR loci (three were eliminated for showing a polyploidy pattern) were analyzed in 86 cassava accessions representing Puerto Rico, Latin America and Africa and distributed among eight groups according to their country or continent of origin. The observed alleles at each locus in the whole data set were high and ranged from 2 at the SSRY 102 to 17 at the SSRY 20 (Table 3). The distribution of the number of alleles at each locus over all 34 loci is shown in Figure 4.

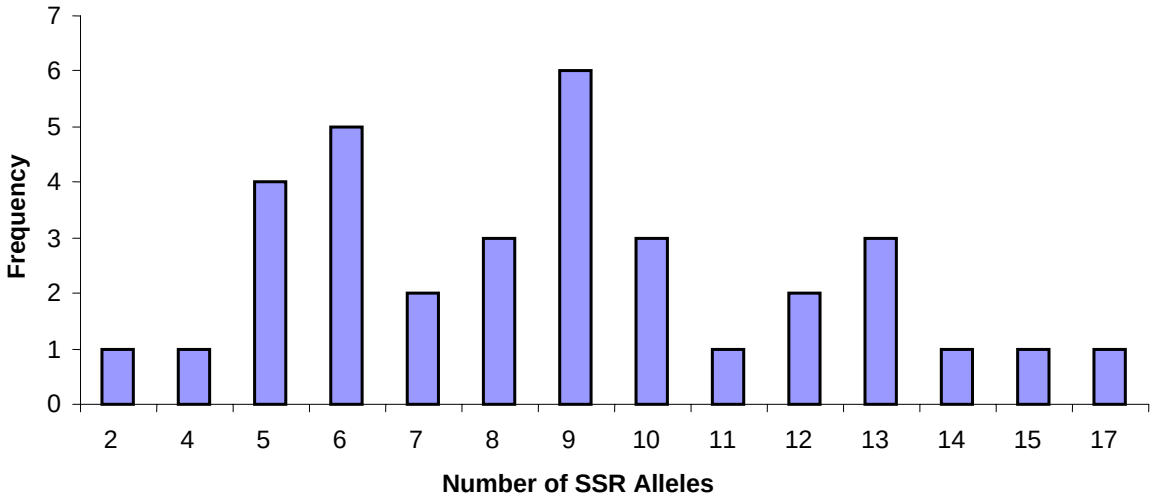

Figure 4. Number of SSR alleles per locus and their frequency in cassava accessions from Puerto Rican collection, and some selected samples from the world's cassava collection from Latin America and Africa.

Table 3. Number of alleles per SSR locus for the cassava accessions from different countries

	Population								
Locus	AFRICA (9)[c]	BRASIL (25)[c]	COLOMBIA (16)[c]	COSTA RICA (4)[c]	ECUADOR (1)[c]	GUATEMALA (4)[c]	MEXICO (3)[c]	PUERTO RICO (24)[c]	Number of alleles
SSRY4	5	8	6	5	2	3	4	9	13
SSRY9	4	6	5	3	2	2[a]	3	6	7
SSRY12	5	5	5	5	2	3	2	5	6
SSRY19	7	11	8	4	2	4	5	11	14
SSRY20	9	12	8	4	2	3	5	10	17
SSRY21	4	9	8	4	1	4	6	7	10
SSRY34	2	5	5	3	1	3	3[a]	5	6
SSRY51	4	7	7	4	2	1[a]	3	8	9
SSRY52	5	7	6	5	2	3	2	5	8
SSRY59	4	10	8	5	2	4	5	9	15
SSRY63	3	4	4	3	[b]	3	2[a]	3	4
SSRY64	5	8	7	5	2	3	5	8	11
SSRY69	5	7	6	4	2	4	3	8	8
SSRY100	6	11	9	5	1	3	3	8	13
SSRY102	2	2	2	2	1	2	2	2	2
SSRY103	8	8	8	3	2	3	4	8	10
SSRY106	6	7	8	4	2	4	4	9	9
SSRY108	3	6	5	3	2	3	3	5	8
SSRY120	6	7	6	4	2	2[a]	2	5	9
SSRY135	6	7	6	3	[b]	4	4	5	9
SSRY147	3	4	2	2	2	1	2	3	5
SSRY148	3	5	3	2	2	2	4	4	5
SSRY151	7	7	8	5	1	2	2	7	8
SSRY155	4	5	5	4	2	3	2	5	6
SSRY161	6	9	5	4	2	4	3	7	12
SSRY164	7	6	8	5	2	5	3	7	10
SSRY175	6	11	9	4	2	4	3	9	12
SSRY177	5	6	5	3	2	2[a]	3	7	9
SSRY179	6	11	10	4	2	[b]	5	9	13
SSRY181	4	7	4	4	2	3[a]	2	5	7
SSRY182	4	6	5	4	2	2	3	5	6
NS189	3	4	4	2	1	2	2	3	5
NS376	7	7	7	3	2	3	2	9	9
NS911	3	3	5	3	[b]	1	2	5	5
Total Alleles	167	238	207	127	56	95	108	221	

[a] Some individuals showed three or more alleles at a given locus

[b] Missing value

[c] Number in parenthesis represents the number of samples per country group

2.3.2 Genetic diversity

Genetic diversity parameters were calculated from the SSR data within and between country samples. For all the SSR loci studied there were on an average 96.85 % ± 6.86 polymorphic loci across all countries samples, showing a high genetic polymorphisms, using the criterion that the frequency of the most common allele is less than 0.95 (Table 4). A high number of alleles were found per locus, with an average of 4.58 ± 1.83, with Brazilian accessions having the highest number at an average of 7.0 alleles per locus, followed by accessions from Puerto Rico (6.5) and Colombia (6.1). The observed heterozygosity (H_o) was high in all country groups and ranged from 0.6121 to 0.8065 at an average 0.7 ± 0.0552 (Table 4), which confirms the highly heterozygous nature of cassava. The probability that two randomly selected alleles in a given accession are different was estimated by H_e. The expected heterozygosity corrected for small samples sizes (H_{ec}; Nei 1978) ranged from 0.6567 to 0.8065 and was on an average more than half (0.7217 ± 0.0476, Table 4), with the highest H_{ec} being found for Ecuador, followed by Costa Rican, Brazilian, Colombian, Puerto Rican, Mexican, African and Guatemalan accessions. This implies that the probability that two randomly selected alleles in Puerto Rico are different is greater than half. The total heterozygosity (H_T) over all loci was high in all country groups (H_T = 0.7581 ± 0.1442, Table 4), but only 4.68 % (G_{ST} = 0.0468 ± 0.0534, Table 4) was due to differentiation among country groups, while most of the diversity was found within country groups (H_S = 0.7194 ± 0.1341).

2.3.3 Genetic differentiation between Puerto Rican, Latin American and African cassava accessions

The F_{ST} (theta) estimator of genetic differentiation, averaged over all loci with mean and standard deviation obtained thorough jackknifing were 0.047 ± 0.005, and 0.033 ± 0.061 as calculated by bootstrapping, with a confidence interval of 99 % (Appendix E). These results reveal again a low level of differentiation between country samples as was observed with estimates of G_{ST} (Table 4).

Table 4. Genetic diversity of cassava accessions from different country origins.

Population	Sample size	No. of Loci	No. of pol Loci	Percent of pol. Loci	Mean No. of alleles/locus	Mean No. of alleles/pol. locus	Ho[1]	He[2]	Hec[3]	Fis_p[4]
AFR	8	34	34	100.0	4.8	4.8	0.7033	0.6391	0.6834	-0.031
BRA	24	34	34	100.0	7.0	7.0	0.6911	0.7246	0.7399	0.0674
COL	15	34	34	100.0	6.1	6.1	0.6707	0.7040	0.7280	0.0773
CR	4	34	34	100.0	3.7	3.7	0.7353	0.6416	0.7569	0.0053
ECU	1	31	25	80.6	1.8	2.0	0.8065	0.4032	0.8065	-0.3333
GUA	4	34	32	94.1	3.6	3.7	0.6912	0.5743	0.6567	-0.0766
MEX	3	34	34	100.0	3.2	3.2	0.6127	0.5658	0.6853	0.0708
PR collection	24	34	34	100.0	6.5	6.5	0.6890	0.7020	0.7170	0.0384
PR unique*	17	34	34	100.0	6.0	6.0	0.6950	0.6937	0.7149	0.0269
Mean				96.85	4.58	4.62	0.7000	0.6193	0.7217	-0.0227
SD				6.86	1.83	1.77	0.0552	0.1053	0.0476	0.1367

	HT[5]	HS[6]	DST[7]	GST[8]
Mean	0.7581	0.7194	0.0387	0.0468
SD[9]	0.1442	0.1341	0.0389	0.0534
95% CI[10]	0.7092	0.6726	0.0256	0.0290
95% CI[10]	0.8010	0.7614	0.0516	0.0646

[1]Mean observed heterozygosity. [2] Mean expected heterozygosity. [3]Mean expected heterozygosity corrected for small sample sizes according to Nei 1978. [4]Average inbreeding coefficient with correction for small samples sizes. [5]Gene diversity in the total set of populations. [6]Average gene diversity within populations. [7] Average gene diversity between populations. [8]Coefficient of gene differentiation. [9]Standard deviations were estimated by Jackknifing over loci (200 replications). [10] 95 % Confidence intervals were obtained through 1000 bootstraps over loci. H_T, H_S, D_{ST} and G_{ST} are given over loci and over groups. *Puerto Rican accessions avoiding those that exhibit similar names or codes to those from CIAT collection.

Pairwise genetic distances between populations using Nei's distance (Nei 1978) provide an estimate of the germplasm exchange between countries. The genetic distances ranged widely, from 0.0138 between accessions from Costa Rica and Ecuador, to a maximum of 0.3598 (Table 5) between accessions of Guatemala and Puerto Rico. The very low genetic differentiation between Costa Rican and Ecuador accessions could be due to the effect of the small sample size from Ecuador group. On the other hand, the high differentiation between Guatemala and Puerto Rico suggest a limited germplasm exchange between these two countries.

Table 5. Average pairwise genetic distance between cassava populations using Nei's distance (Nei 1978)

Population	Africa	Brasil	Colombia	Costa Rica	Ecuador	Guatemala	Mexico	Puerto Rico
Africa	0							
Brasil	0.1975	0						
Colombia	0.2165	0.0367	0					
Costa Rica	0.0866	0.1712	0.1454	0				
Ecuador	0.1174	0.1515	0	0.0138	0			
Guatemala	0.3090	0.2262	0.2818	0	0.1498	0		
Mexico	0.3193	0.1669	0.1710	0.1845	0.0791	0.3076	0	
Puerto Rico	0.2596	0.0789	0	0.2298	0.0662	0.3598	0.1978	0

An UPGMA dendrogram of the cassava landraces grouped by country using Nei's distance estimates separates the African from Neotropical accessions. Similar dendrograms have been reported in previous cassava diversity studies (Fregene et al. 2003; Kizito et al. 2005). Figure 5 show the differentiation found between the accessions from Costa Rica and Guatemala, and between Colombia and Ecuador. The accessions of Colombia, Ecuador and Brazil representing South America, and Puerto Rico from the Caribbean, clustered closely together. Interestingly, the Guatemalan and Costa Rican accessions grouped distantly to the Latin American accessions compared to the African ones (Figure 5).

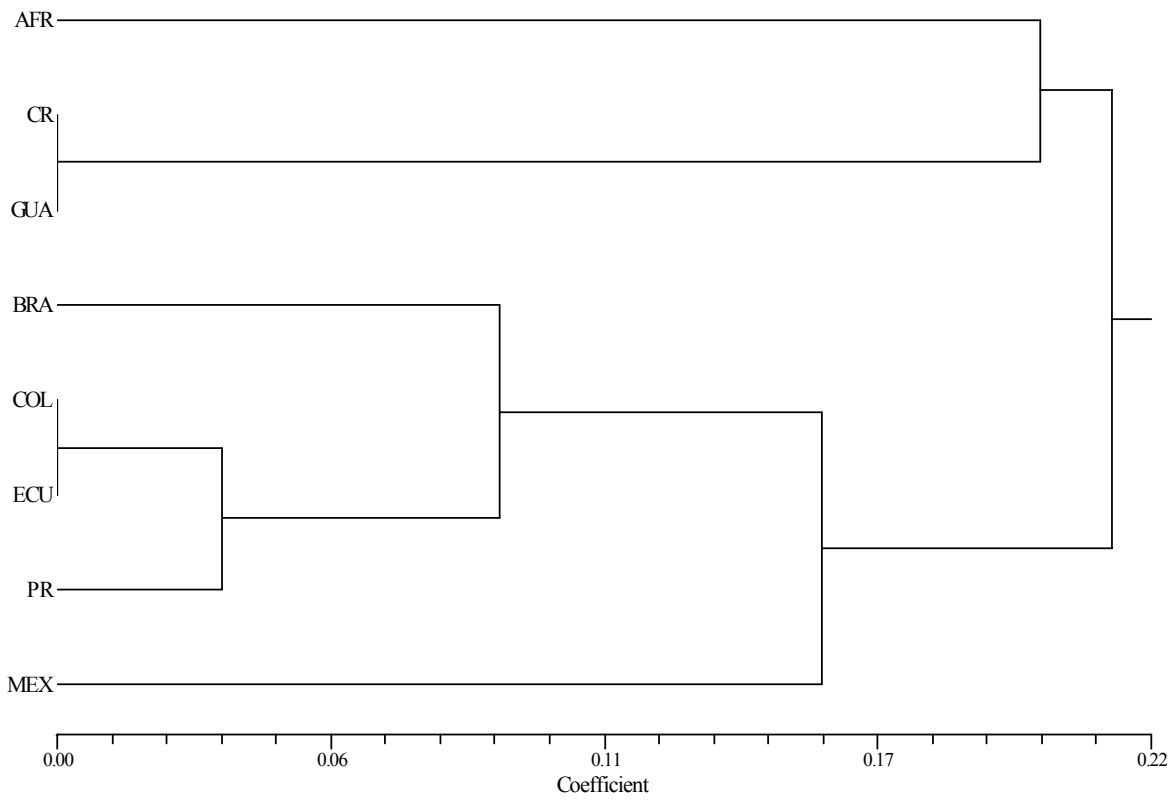

Figure 5. Unweighted pair-group method with arithmetic averaging (UPGMA) dendrogram illustrating the genetic relationships among cassava accessions grouped by country of origin as revealed by pairwise genetic distance (Nei 1978).

Pairwise calculations of F_{ST} (theta) over all loci between pairs of country groups were obtained to corroborate the results obtained using Nei's genetic distance (Table 6). F_{ST} values ranged widely, from -0.0365 between accessions from Costa Rica and Guatemala, to a maximum of 0.163 between accessions from Guatemala and Africa. The very low genetic differentiation between Guatemalan and Costa Rican accessions reveals an exchange of germplasm among them. On the other hand, medium differentiation among Guatemalan and African accessions suggest limited germplasm exchange between these two groups. The UPGMA of the pairwise F_{ST} values resulted in a dendrogram (Figure 6) that indicates a moderate differentiation of the African accessions from Neotropical ones. Additionally, it reveals the presence of a high differentiation between Guatemalan and African accessions and a moderate differentiation between Guatemalan and the other Neotropical groups.

Table 6. Pairwise estimator of F_{ST} (theta) between pairs of country groups of cassava accessions.

Population	Africa	Brasil	Colombia	Costa Rica	Guatemala	Mexico	Puerto Rico
Africa	0	0.0627	0.0729	0.0396	0.1630	0.1110	0.0891
Brasil	0.0627	0	0.0109	0.0503	0.0934	0.0468	0.0267
Colombia	0.0729	0.0109	0	0.0437	0.1063	0.0481	-0.0024
Costa Rica	0.0396	0.0503	0.0437	0	-0.0365	0.0528	0.0737
Guatemala	0.1630	0.0934	0.1063	-0.0365	0	0.1342	0.1315
Mexico	0.1110	0.0468	0.0481	0.0528	0.1342	0	0.0632
Puerto Rico	0.0891	0.0267	-0.0024	0.0737	0.1315	0.0632	0

No F_{ST} values were obtained for Ecuador, this could be explained as a result of the small sample size for that particular group. This also explains why the dendrograms show differences in the arrangement of the country groups.

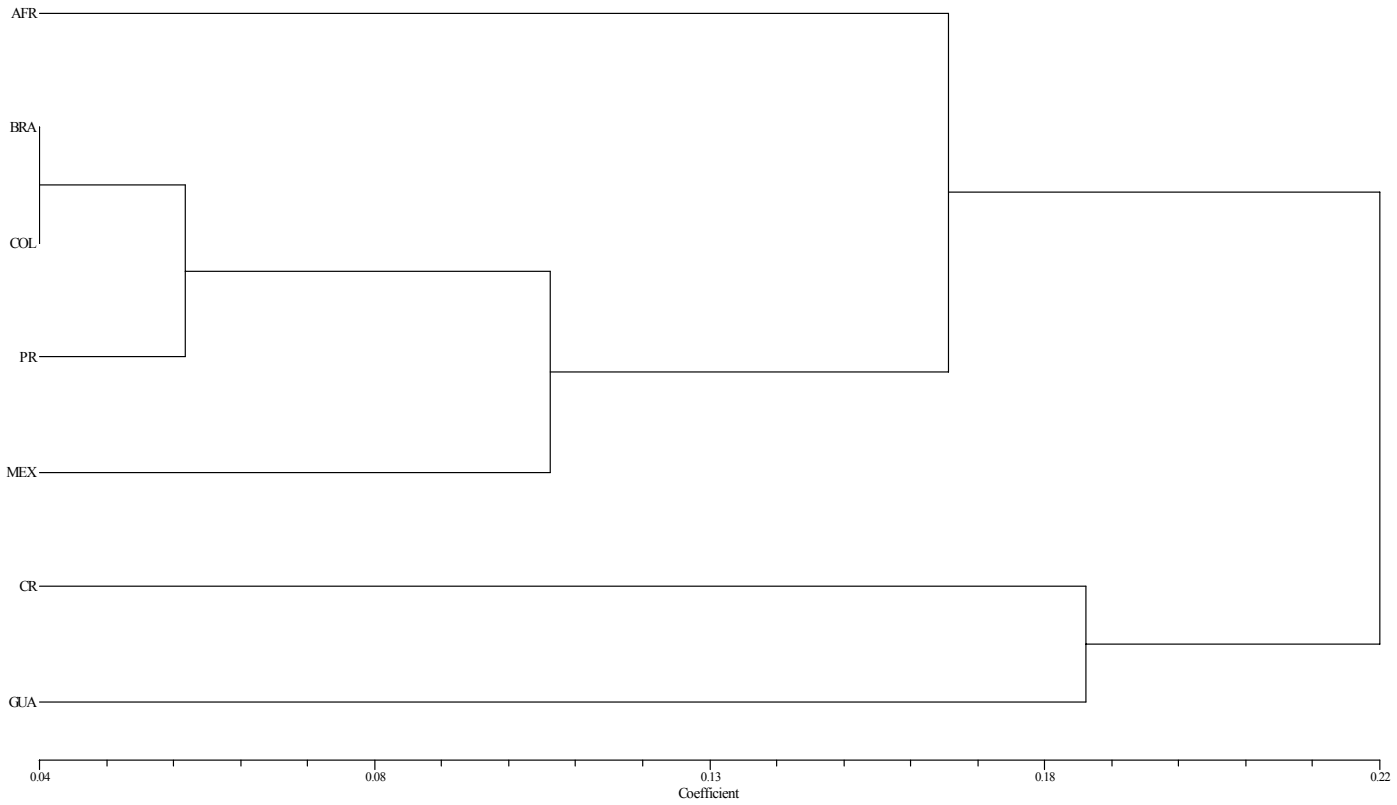

Figure 6. UPGMA dendrogram of the pairwise fixation index F_{ST} between cassava accessions grouped by country of origin.

2.3.4 Genetic similarity among accessions and cluster analysis

To get a graphical representation of the relationships between individual cassava accessions, a principal coordinate analysis (PCoA) was performed based on the genetic similarity matrix calculated using the Jaccard coefficient. The three main eigenvalues of the PCoA on the similarity matrix explained 6.85 %, 5.28 %, 3.87% of the total variation, respectively (Appendix F). However, there is a total of 19 eigenvalues that explain 52.04% of the total variation, which means that all the characteristics exhibit a similar weight. Figure 7 shows the scatter plot that splits the accessions in five main groups defined by the first two coordinates. The first group is the largest one and contains accessions from all the countries represented in this study. The second group is constituted by Puerto Rican accessions Seda, Brava, PI12903 and Cubana. The third group is constituted by accessions from Colombia which are also present in Puerto Rico. The fourth group is constituted by two accessions from the Puerto Rican collection, one from Africa (TME 857) and one from Costa Rica (CR 30). Finally, the fifth group containing accessions from Guatemala and Costa Rica appears to be separated from the Neotropical and African accessions as was previously observed in the UPGMA dendrogram based on Nei's distance for the cassava accessions grouped by country of origin (Figure 7).

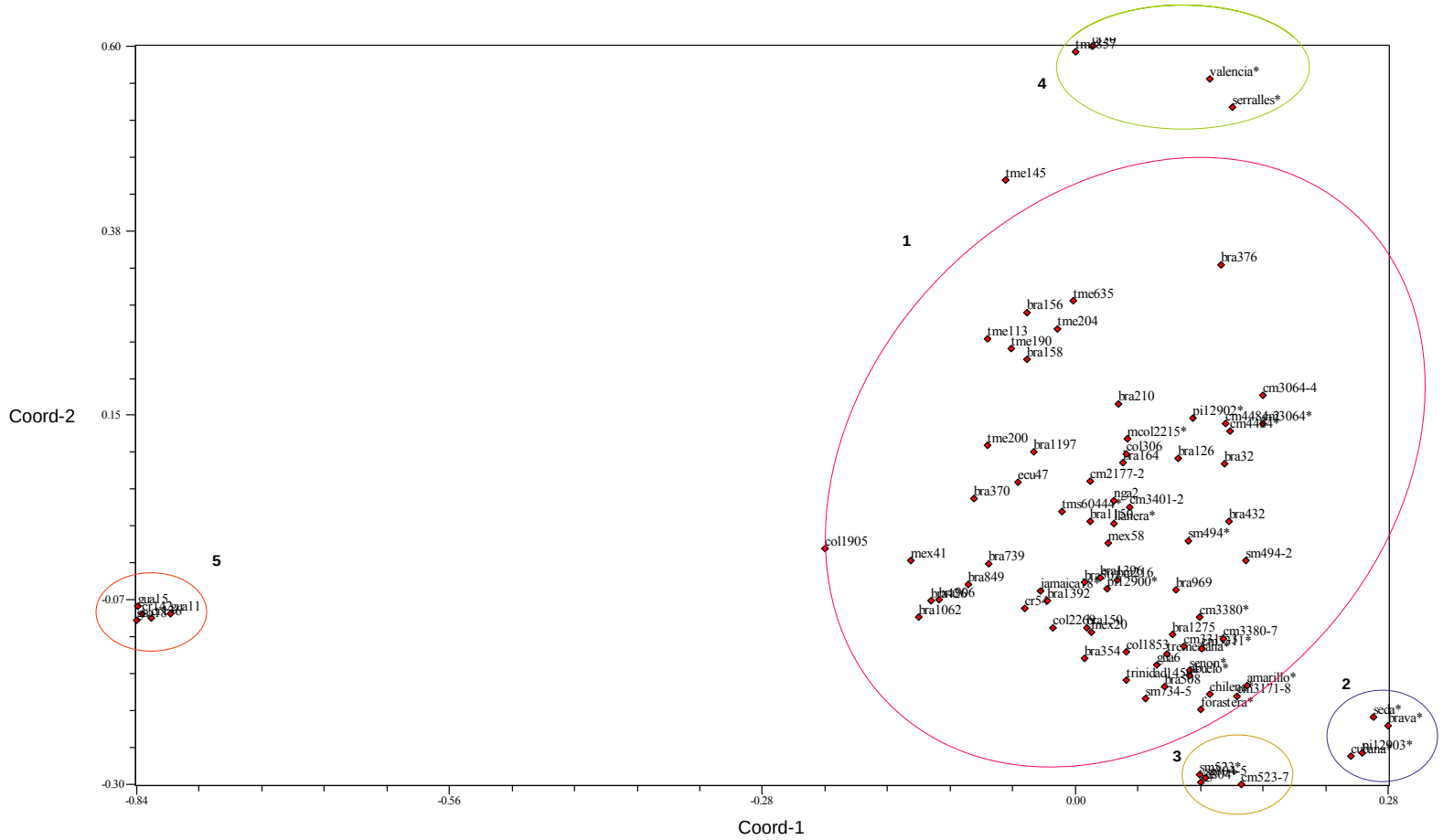

Figure 7. Scatter plot of SSR marker diversity in cassava accessions was derived from plotting the first two principal coordinates, which were computed from Jaccard's similarity matrix. Coordinates one and two explain 6.85 and 5.28 % of the variance observed, respectively.

Similarly, a clustering analysis for SSR assessment of genetic distance among the 86 cassava accessions using Euclidean distance and UPGMA grouping method again revealed five cluster groups at the 0.53 similarity coefficient. The fifth group constituted by Costa Rican and Guatemalan accessions is distantly related (0.07 similarity coefficient) to the other four clusters as was previously observed in the PCoA results, but the similarity among the accessions inside the cluster is more than 0.95. Group number four shares a similarity of 0.346 with other cluster groups, and shows a within-group similarity higher than 0.852. Group number three is 0.438 similar with respect to the other four groups, and the similarity among its accessions is more than 0.76. Finally, groups number two and one show a similarity of 0.53 compared to the other main clusters. There is a similarity of 0.852 among the accessions which are part of the cluster number 2, and 0.668 among the accessions that are part of the cluster number 1 (Figure 8).

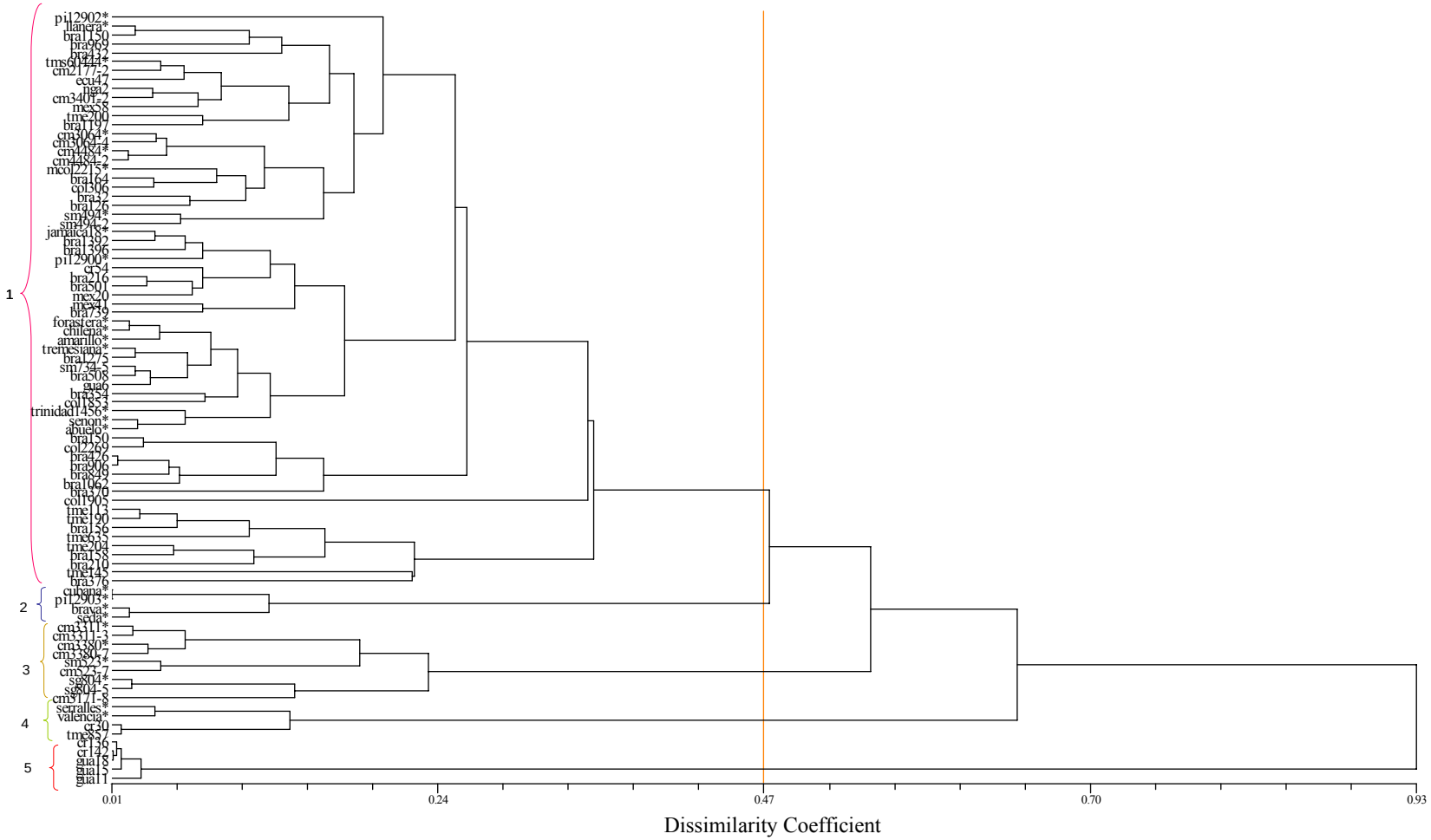

Figure 8. Cluster analysis for SSR assessment of genetic distance among 86 cassava accessions using Euclidean distance and UPGMA grouping method.

Table 7 shows the closest relatives found in this study and their associated similarity coefficients.

Table 7. Similarity coefficients for the cassava accessions found as closest relatives.

Closest Relatives	Cluster Location	Similarity Coefficient
Llanera* and BRA1150	1	0.97
TMS60444* and CM2177-2	1	0.96
NGA2 and CM3401-2	1	0.97
TME200 and BRA1197	1	0.93
CM3064* and CM3064-4	1	0.95
CM4484* and CM4484-2	1	>0.98
BRA164 and COL306	1	0.96
BRA32 and BRA126	1	0.94
SM494* and SM494-2	1	0.94
Jamaica18* and BRA1392	1	0.96
BRA216 and BRA501	1	0.97
Forastera* and Chilena*	1	0.98
Tremesiana* and BRA1275	1	0.97
SM734-5 and BRA508	1	0.97
BRA354 and COL1853	1	0.92
Senon* and Abuelo*	1	0.97
BRA150 and COL2269	1	>0.96
BRA436 and BRA906	1	>0.98
TME113 and TME190	1	0.97
TME204 and BRA158	1	>0.94
Cubana* and PI12903*	2	0.99
Brava* and Seda*	2	>0.97
CM3311* and CM3311-3	3	>0.97
CM3380* and CM3380-7	3	0.96
SM523* and CM523-7	3	0.95
SG804* and SG804-5	3	0.97
Serralles* and Valencia*	4	0.95
CR30 and TME857	4	>0.98
CR142 and GUA18	5	0.99

* Cassava accessions of the Puerto Rican collection

According to the results showed in Table 7, Cubana and PI12903 from the Puerto Rican collection and CR142 and GUA18 from CIAT collection seem to be the same genotypes.

2.4 Discussion

The number of unlinked SSR loci employed in this study enabled a good estimation of the diversity and genetic differentiation parameters for the cassava accessions included in this study, as was previously reported by Fregene et al. (2003). According to their studies, little or no additional increase in the polymorphism information content (PIC) is obtainable using more than 30 SSR markers. Additionally, Baverstock and Moritz (1996) affirmed that the reliability of estimates for genetic variation, such as H_e, H_o, F_{ST} and genetic distances, depends more on the number of loci used than on the number of individuals sampled. In this study we used the information of 34 (37 were analyzed but three of them showed a polyploidy pattern that was not compatible with the diallelic model of inheritance adopted and thus were voided). The mean number of alleles detected per locus in our study ranged from 2 to 17, with a mean value of 4.58 alleles per locus. These values are similar to those obtained in previous cassava diversity studies (Roa et al. 1997; Chavarriaga et al. 1998; Fregene et al. 2003; Kizito et al. 2005) and are on the low side of values reported for outcrossed woody perennial species, such as citrus (5.5 alleles per locus; Kijas et al. 1995), apple cultivars (12.1; Hokanson et al. 1998) and avocado, (9.5; Lavi et al. 1994). Genetic diversity in cassava has been previously studied using different types of DNA molecular markers, such as isozymes (Sarria et al. 1992), RFLP (Beeching et al. 1993; Haysom et al. 1994); AFLP (Roa et al. 1997) and SSR (Chavarriaga et al. 1998; Roa et al. 2000; Olsen and Schaal 2001; Fregene et al. 2003; Kizito et al. 2005; Moyib et al. 2007; Rocha et al. 2008) showing a low or medium genetic diversity. In the present study high values of observed (H_o) and expected (H_e) heterozygosity, the closest values of the inbreeding coefficient (F_{IS}) to zero meaning neither inbreeding nor outbreeding, and in some cases negative F_{IS} values indicate there are higher amounts of heterozygotes in some country groups. The number of alleles and the diversity estimators observed in the studies of Fregene et al. (2003) and Kizito et al. (2005) are comparable with the ones reported here and suggest that SSRs are an excellent tool for cassava diversity assessment. Genetic variation at SSR loci has been used to investigate the evolutionary and geographical origins of cassava and its wild relatives (Olsen

and Schaal 2001). The findings of that study suggested that cassava is derived from populations of *M. esculenta* subsp. *flabellifolia* (Pohl) growing on the southern border of the Amazon Basin, indicating that region as the likely site of domestication for the crop. According to Fregene et al. (2003) the expansion of cassava into other regions of Neotropics, Africa and Asia could have led to a founder effect of reduced diversity and an increase in genetic differentiation. Nevertheless, high genetic diversity of cassava landraces in all the countries studied has been found, and this is unexpected. Pujol et al. (2005) demonstrated that Amerindian farmers favor heterozygosity in cassava by selecting the large volunteer seedlings that appear spontaneously in fields, during the process of weeding. The high genetic diversity found in this study then can be explained as the result of the high diversity present in the original populations, the high individual heterozygosity and the allogamous nature of cassava.

The genetic relationships observed among cassava accessions here using diversity estimators, genetic distances, PCoA and cluster analysis affirmed the differentiation of the accessions according to their country of origin. Even though there is a low value of differentiation (F_{ST} (theta): 0.047 ± 0.005) among country groups, there is enough genetic distance among the accessions to separate them (as was observed for Nei's and F_{ST} pairwise distances). The separation between Latin America and Africa and the highly differentiated accessions from Guatemala observed here are comparable to the results reported by Fregene et al. (2003) and support to the idea that Guatemalan germplasm represents an independent event of domestication or that an introgression with wild genotypes has occurred (Rogers 1963; Renvoize 1972; Fregene et al. 2003). Additionally, the UPGMA cluster analysis of the 86 accessions in the present study produced meaningful groupings that permit the closer examination of the Puerto Rican collection. We were able to detect two misnamed accessions (Cubana and PI12903) by using these markers. The results also revealed the genetic similarity between pairs of accessions from Puerto Rican collection with some of their cultivated relatives from Colombia and Brazil. Such comparisons will direct future collection efforts on the island and guide the conservation strategies to preserve the maximal amounts of genetic diversity found for this species.

In conclusion, one of the important findings of the study is that Puerto Rican accessions are highly related to South American, especially to those from Colombia and Brazil. Other principal findings are the separation observed between Guatemalan-Costa Rican and the Latin American accessions, and the close relation between African and Brazilian accessions.

The introduction of disease resistant, abiotic stress resistant and productive accessions from other diversity centers such as Africa, Central America, and Caribbean would help increase the diversity in the collection, expand production, reduce cost and meet local demand in Puerto Rico.

CHAPTER III

3. Tissue culture and somatic embryogenesis of cassava (*Manihot esculenta* Crantz, Euphorbiaceae) accessions of Puerto Rico

3.1 Introduction

Cassava (*Manihot esculenta* Crantz) is a tropical crop, also known as yuca, manioc or tapioca, which is a perennial shrub grown in the lowland tropics of Latin America, Africa and Asia for its starchy roots. It constitutes a source of calories for more than 600 million people and is mostly produced by poor small-scale farmers for food, feed and cash (Best and Hargrove 1994). Currently, cassava is the most important source of dietary calories in the tropics after rice and maize because it is efficient in carbohydrate production, is well known for its ability to tolerate drought, its excellent adaptability to a wide range of environments and because it grows well in low fertility and acidic soils (Fregene et al. 2003; Kawano 2003; Raven et al. 2006). Additionally, several products can be obtained from it such as starch, flour, ethanol, glucose, mannitol, and sorbitol, among others, which have human nutrition, animal feed and industrial uses. It is expected that by the year 2050, 90% of the human population will live in developing countries where agricultural practices are an important component of their economies (Raven et al. 2006). Thus, cassava offers a good opportunity to solve the calorie requirements of people living in countries considered poor.

Problems in the cultivation of cassava include virus diseases, pests, low protein content and in some cultivars high levels of cyanogenic glucosides (Cock 1985). Classical breeding has been tried to overcome these problems, but several biological constraints such as the plant's long growth cycle, high degree of heterozygosity, allotetraploidy, allogamy and inbreeding depression on selfing have limited cassava breeding for those important traits (Fregene et al. 2001). Genetic engineering promises to complement traditional cassava breeding goals, but methods for efficient cassava transformation, including plant *in vitro*

generation, gene delivery, selection of transformed tissues and recovery of transgenic plants, still need to be improved (Zhang 2000). Furthermore, the optimization of embryogenic induction methods in different elite cassava accessions is a necessity to improve transformation systems.

As starting material, young leaves have been used for the induction of somatic embryos (Stamp and Henshaw 1987a, Szabados et al. 1987; Raemakers et al. 1993; Siritunga and Sayre 2003; Hankoua et al. 2005). Nevertheless, the formation of direct organized embryogenic structures has been obtained from other tissues such as zygotic embryo, meristem, and flower inflorescence explants (Hankoua et al. 2005). Siritunga and Sayre (2003) and Siritunga et al. (2004) provide the first evidences that genetic engineering can be used to modify cassava plants to alter useful agronomic traits. Two approaches were employed to reduce cyanogen toxicity in cassava: the first one included the reduction of cyanogenic glucosides using an antisense strategy to silence two target genes (cyp79d1 and cyp79d2) involved in the pathway, and the second took advantage of the over-expression of the HNL (hydroxyl nitrile lyase) enzyme activity to accelerate the detoxification process of cassava foods.

To date, none of the 23 accessions present in the Puerto Rican collection have been tested for their ability to grow under tissue culture conditions and for their capacity to form somatic embryos. In order to compare their embryogenic capacity, we used MS8 (8 mg/l) as induction medium in leaf explants of the 23 accessions. Accession MCol2215 was used as control since it has been the model cultivar in previous embryogenesis studies. Herein, we describe the process of adapting these accessions to the existing tissue culture systems of micropropagation and embryogenesis induction.

3.2 Materials and methods

3.2.1 Plant material and micropropagation

Twenty-three cassava accessions were collected in January of 2006 from the Corozal Experimental Station field of the University of Puerto Rico (latitude 18˚20'N, longitude 66˚31'W, altitude 185 m) (Figure 9a), located in the highland agricultural zone of Puerto Rico. Thirty cm long woody stakes, containing two to three nodes were brought to the University of Puerto Rico at Mayagüez Campus where they were planted in 3.78 L pots with sterile soil-sand mix in a proportion of 1:3 (Figure 9b) and grown under screenhouse conditions. Once the plants sprouted the stems having axillary and apical buds were introduced to *in vitro* tissue culture by sterilizing them with 20% (v/v) household bleach (sodium hypochlorite solution) and 0.05% (v/v) Tween 20. After washing the cuttings with sterile deionized water four to five times, they were cultured on semi-solid Murashige and Skoog (MS)-4E micropropagation medium (Murashige and Skoog 1962) containing 1X MS salts basal medium (4.3 g) supplemented with 2 % (w/v) sucrose, 0.04 mg/L N6-benzylaminopurine (BAP), 0.05 mg/L GA3, 0.02 mg/L NAA, 100 mg/L myo-inositol, 1 mg/L Thiamine-HCl , and 0.7 % Agar (Fisher Scientific, Fair Lawn, NJ), adjusting the pH of the medium to 5.75 with KOH 0.1 N before autoclaving at 121 °C for 20 min to 15 psi (Appendix G). Media was dispensed at 50 mL per magenta box (60 x 60 x 10 cm) and 5 ml per test tube (16 mm x 125 mm). Shoot cultures were incubated at 27 ± 2˚C in a growth cabinet (Percival Scientific™ Inc., Perry, IA) and were partially covered from light with butter paper for two weeks to induce etiolation. Afterwards cultures were exposed to light under a 16 h/day photoperiod (75 µmoles m^{-2} s^{-1}, Phillips fluorescent tube lamps) following the protocols described by Escobar (1991). The plants were micropropagated by nodal cuttings every four months in the same medium, and young leaves were used for induction of somatic embryogenesis based on the procedure described by Szabados et al. (1987) and Siritunga and Sayre (2003).

Figure 9. Cassava plants at the Corozal Experimental Station field (a) and grown in pots at the Department of Biology of the University of Puerto Rico (b).

3.2.2 Induction of somatic embryogenesis

Immature leaves 1-6 mm in length were excised from the maternal plant and placed on solid embryogenesis induction medium (MS8) in 10 cm diameter Petri dishes. MS8 medium is the basal culture medium described by Murashige and Skoog (1962), supplemented with 8 mg/L 2, 4-dichlorophenoxyacetic acid, 2 % (w/v) sucrose, 10 ml/L of 100X Gamborg's B-5 vitamins (Gamborg et al. 1968), 50 mg/L casein, 2 µM $CuSO_4$, and 0.2 % Phytagel (Sigma Aldrich, St. Louis, MO), adjusting the pH of the medium to 5.75 with 0.1 N KOH before autoclaving at 121 °C for 20 min to 15 psi. Petri dishes with nine explants per dish were sealed with Parafilm to avoid contamination with fungi and bacteria. Plant cultures were partially covered with cheese cloth to reduce light and to optimize the generation of embryogenic structures and then were incubated at 27 ± 2 °C for four weeks. Development of embryogenic tissues was scored for each explant and Petri dish per accesion, and then the explants with developed somatic embryos were transferred to solid regeneration medium RM1 [MS basal medium supplemented with 1.0 mg/L thiamine, 100 mg/L myo-inositol, 2 %

sucrose, 0.1 mg/L BAP, 0.5 mg/L GA_3 and 0.2 % Phytagel at pH 5.75] in order to encourage plant development (Appendix G).

3.3 Results

3.3.1 Cassava micropropagation

Stem cultures showed a faster released of buds after one week (Figure 10). The rate of nodes obtained per shoot cultured on 4E medium for the 23 accessions (Amarillo not included because it was not introduced to *in vitro* tissue culture and because it was found late in 2008) and the control MCol2215, after 8 weeks was approximately 1:4 (shoot: nodes) depending upon the genotype (Table 1). Generally low light regimens during the first two weeks improved the sprouting of apical buds.

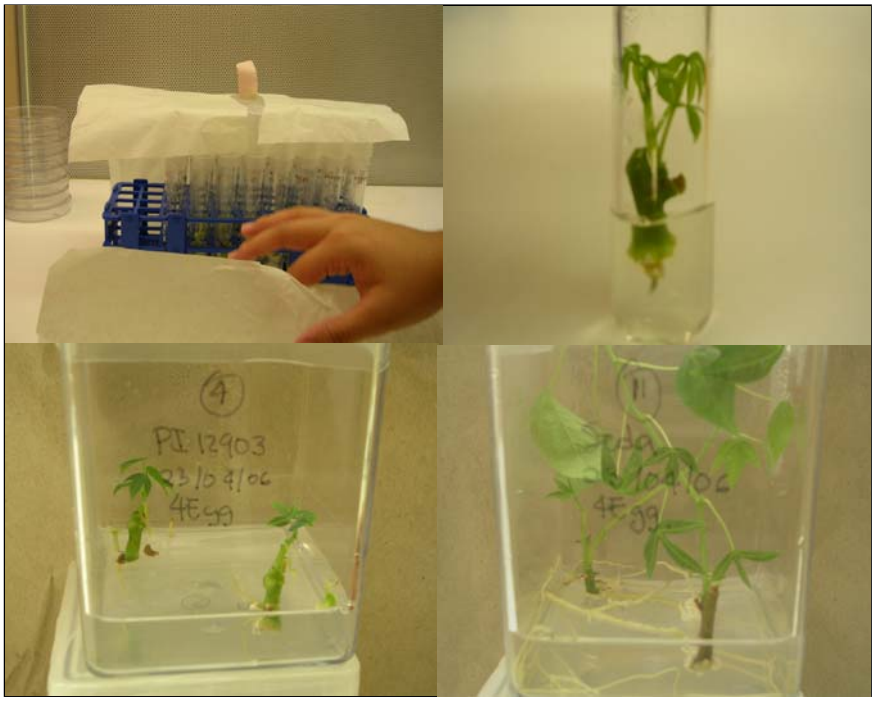

Figure 10. *In vitro* cassava explants growing in tubes and magenta boxes

Table 8. Number of cassava nodes produced per shoot by each genotype after two months growing in 4E medium.

No.	Accession	Nodes per shoot
1	PI12902	4
2	Jamaica 18	3
3	Cubana	5
4	Pi12903	3
5	Pi12900	5
6	Forastera	5
7	Llanera	3
8	Trinidad 1456	3
9	Brava	3
10	Tremesiana	3
11	Seda	8
12	Serralles	5
13	Senon	3
14	CM3064	7
15	CM3311	4
16	CM3380	7
17	CM4484	5
18	SG804	4
19	SM494	3
20	SM523	4
21	Abuelo	3
22	Chilena	3
23	Valencia	5
24	MCol2215	4

3.3.2. Induction of somatic embryogenesis and embryo germination

After one week on the induction medium (MS8), the leaf lobes turned from green to creamish-yellow and began the formation of callus, which was in most of the cases friable callus. A wide variation of color callus was found; white, cream-white, green, brown and black. Globular somatic embryos developed from the nodular embryogenic tissues and appeared on the upper surface of the leaf lobe explant between days 15 and 20. The first torpedo shaped embryos were observed after one month (Figure 11).

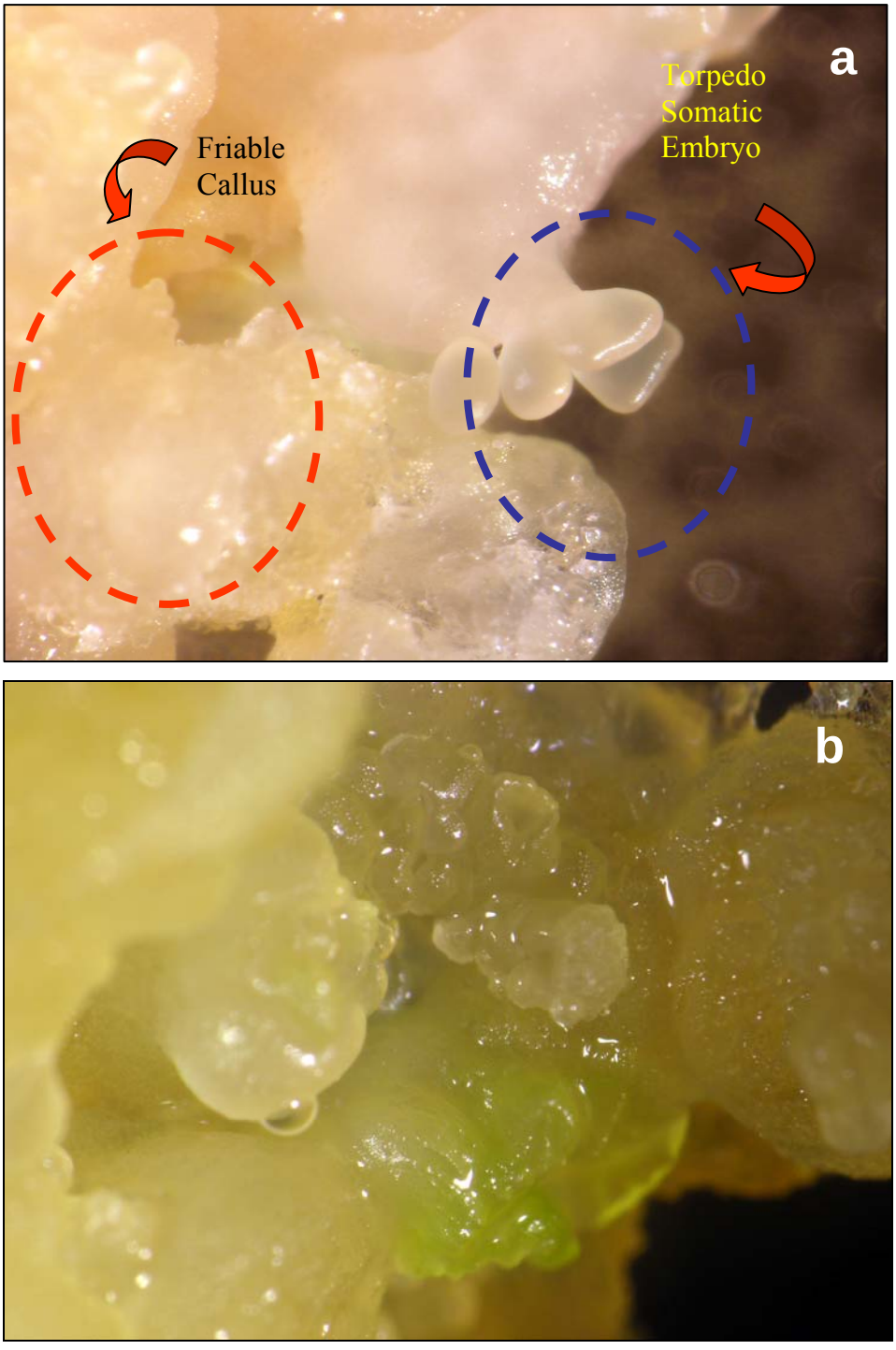

Figure 11. Cassava friable calli and somatic embryos in mature torpedo shape developing on the embryogenic tissue (a) and somatic embryos in globular shape (b) (200X).

A total of 3,104 leaf explants were induced with MS8 medium (Appendix H). The best response to the embryogenesis induction was observed in the accessions SG804, Tremesiana, Seda, MCol2215, PI12903, CM3064 and Senon (Figure 12). Other cassava accessions showed lower or no embryogenic response under the same treatment or developed adventitious roots during the formation of callus, such as the accessions Valencia and PI12902. Few accessions showed clumps during their embryogenic growth, and those with a torpedo shape developed into germinated embryos easily. However, those that showed single embryos were difficult to germinate (Table 9).

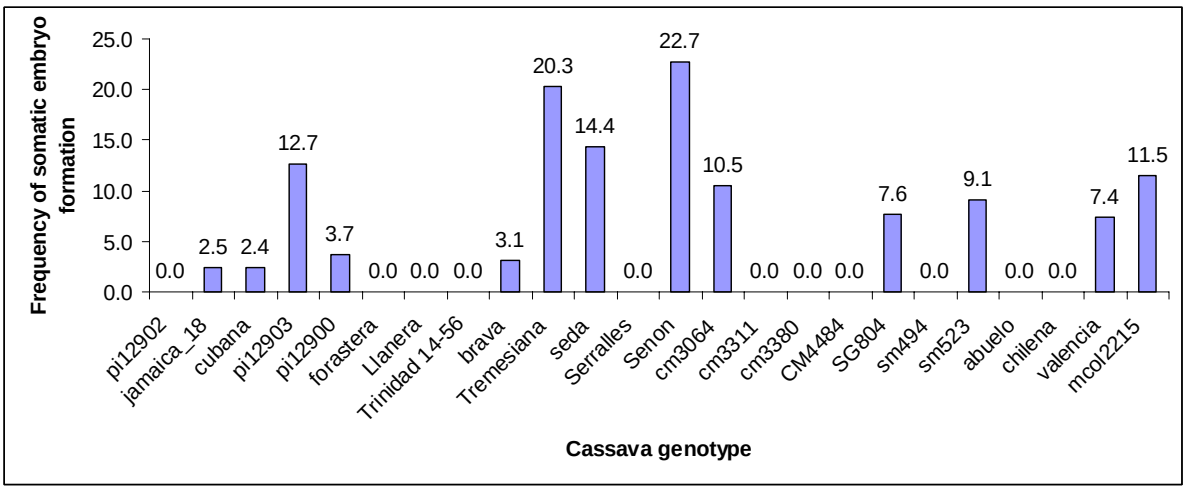

Figure 12. Percentage of cassava somatic embryos developed on the surface calli by the 23 accessions from Puerto Rican collection and one control (MCol2215) under the effect of 8 mg/L of 2,4-D treatment.

Table 9. Cassava embryogenic cultures growing as clumps or single embryos on young leaf explants.

Accession Name	Embryogenic Cultures
PI12902	-
Jamaica 18	Single embryos
Cubana	Single embryos
PI12903	Single embryos
PI12900	Clumps of embryos
Forastera	-
Llanera	-
Trinidad 14-56	-
Brava	Single embryos
Tremesiana	Clumps of embryos
Seda	Clumps of embryos
Serralles	-
Senon	Single embryos
CM3064	Clumps of embryos
CM3311	-
CM3380	-
CM4484	-
SG804	Clumps of embryos
SM 494	-
SM523	Single embryos
Abuelo	-
Chilena	-
Valencia	Single embryos
MCol2215	Clumps of embryos

- No embryo formation.

Forty-eight percent of the accessions did not produce organized embryogenic structures, showing their poor potential for embryogenesis. To promote the differentiation of the organized somatic embryos into plants (Figure 13), they were transferred to RM1 medium and incubated under continuous light (75 μmoles m^{-2} s^{-1}). The highest regeneration percentages were observed for MCol2215, followed by CM3064, Tremesiana, PI12900, SG804, PI12903 and Seda (Figure 14).

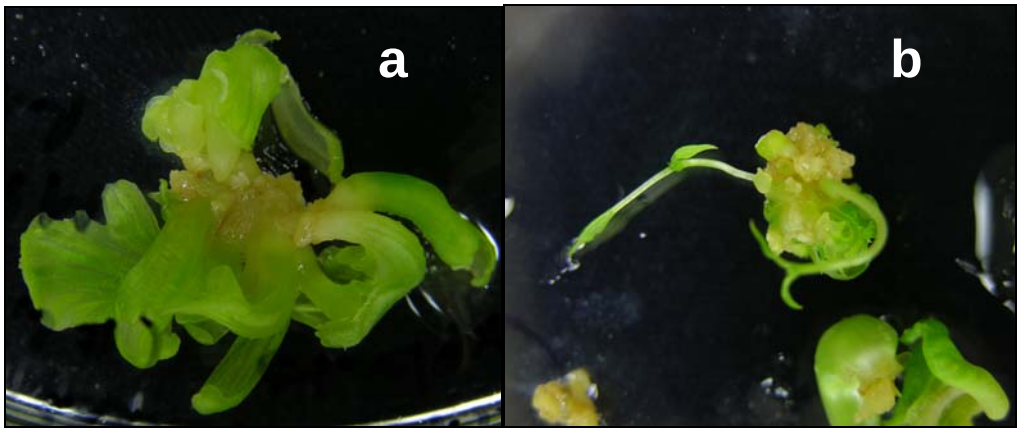

Figure 13. Germinating cassava somatic embryos of SG804 growing on RM1 medium. a) Germinating embryos in green cotyledon (mature) stage; b) shoot development from germinating embryos.

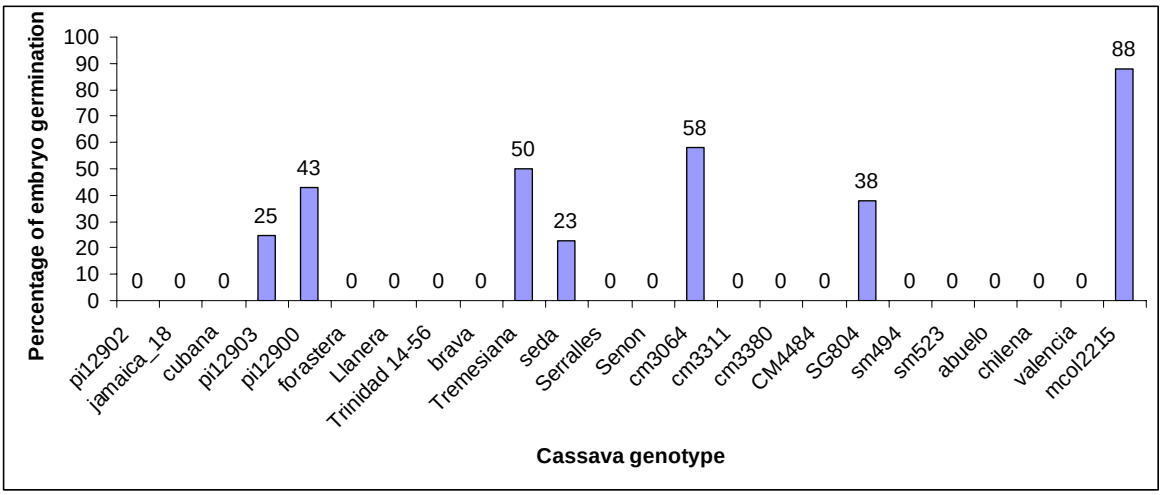

Figure 14. Percentage of cassava embryo germination on RM1 medium for the 23 accessions and the control.

Using the germination induction medium (RM1), we found that only 26 % of the somatic embryos from the total of the accessions finally germinate. All accessions tested produced embryos at a lower germination rate compared to control MCol2215.

3.4. Discussion

In the present study 52 % of the accessions produced organized embryogenic structures, but only six of the 23 accessions (SG804, CM3064, Tremesiana, PI12900, PI12903 and Seda) showed germination of those somatic embryos on RM1 medium, suggesting they could be used in further embryogenesis studies. Additionally, our results indicate that the number of somatic embryos produced was higher than that which already germinated, which agrees with the findings reported by Szabados et al. (1987). The cotyledon-like structure was the most conspicuous structure observed during germination process. However, shoots and adventitious roots were also observed.

The differences found among the explants from Puerto Rican accessions in their capacity to form embryogenic structures could be explained as the result of their intrinsic genetic characteristics. Our results support the thought that production of embryogenic structures in cassava varies among different accessions. Additionally, differences in the production of embryos among explant tissues could be attributed to the induction system used, because on solid medium the explants only make basal contact with the medium, and the nutrient and hormone uptake is then limited, forming gradients in the embryogenic calli, that promote variation and reduce the growth of the embryogenic structures (Raemakers et al. 1993; Szabados et al. 1987). To improve the efficiency of generation of somatic embryos and its germination, we suggest the use of liquid medium in those accessions that showed embryogenic capacity and also implementation of desiccation treatment to permit the synchronous development of shoots and roots (Mathews et al. 1993).

4. Conclusions

- Twenty-three cassava accessions were satisfactorily introduced to an *in vitro* system. Their response to propagation and their conversion to a plant were highly dependent on the genotype. Generally, low light regimes during the first two weeks after the explant being cultured improved the sprouting of apical buds.
- For all SSR loci studied, a very high genetic polymorphism (96.85 % ± 6.86) across all country groups was observed, which confirm that the cassava microsatellite kit is an excellent tool to access the diversity of this crop.
- The diversity assessment of the Puerto Rican cassava collection with 34 SSR markers revealed that there is a high genetic diversity in the collection and also showed a low genetic differentiation between Puerto Rican accessions and those from Brazil and Colombia, suggesting exchange of material between the island and these centers. The high diversity found in the Puerto Rican collection could be explained as the result of the large diversity present in the original populations, the high individual heterozygosity and the allogamous nature of cassava.
- Cluster analysis also revealed that Puerto Rican accessions grouped with South American landraces. Additionally, two misnamed accessions in the Puerto Rican collection were detected due to the high resolution of the technique used. Furthermore, it led to the identification of possible closest cultivated relatives to some of the collection's materials.
- One of the principal findings of this work is the confirmation that there is a low level of differentiation [F_{ST} (theta) = 0.047± 0.005] among country groups. Even though there is enough genetic distance among accessions to separate them according to their country of origin. These results agree with those obtained in previous cassava diversity studies.

- The separation observed between Neotropical and African accessions and the highly differentiated landraces from Guatemala confirm the results reported by Fregene et al. (2003) and support the idea that Guatemalan germplasm represents an independent event of domestication or that an introgression with wild accessions has occurred.
- Using tissue culture techniques, we found that 52 % of the accessions produced organized embryogenic structures but only 26 % of the total produced germinated somatic embryos. All Puerto Rican collection accessions produced embryos at a lower germination rate than the control MCol2215.

5. Suggestions and perspectives

- The differentiation patterns in cassava are very important for the development of a cassava breeding program. The evaluation of the genetic variation of new accessions from different countries will help confirm or find new differentiation patterns.
- At the moment most of the cassava diversity studies have been focused on South American and African accessions, and there is a little known about the cassava genetic present in the Mesoamerican regions. Thus we suggest that future work may include the systematic collection of new accessions in the archipelago of Puerto Rico and adjacent islands as part of a conservation strategy for the cassava genetic resources for the Caribbean region.
- The high genetic diversity found in the Puerto Rican collection and the process involved in the maintaining of this diversity have an important application in the *ex situ* and *in vitro* cassava conservation programs.
- To improve the generation of somatic embryos and their germination, we suggest the use of liquid medium and embryo desiccation treatments to increase the production of embryogenic structures and permit their synchronous development into plants.

6. Bibliography

Allem A (1994) The origin of *Manihot esculenta* Crantz (Euphorbiaceae). Genetic Resources and Crop Evolution 41: 133-150.

Baverstock P, Moritz X (1996) Project design. In: Hillis D, Moritz C, Mable B (eds) Molecular Systematics. Sinauer Associates, Sunderland MA, 17-27.

Beeching J, Marmey P, Gavalda M, Noirot M, Haysom H, Hughes M, Charrier A (1993) An assessment of genetic diversity within a collection of cassava (*Manihot esculenta* Crantz) germplasm using molecular markers. Annals of Botany 72: 515-520.

Best R, Hargrove T (1994) Cassava: the latest facts about an ancient crop. CIAT publication. Cali, Colombia.

Castelblanco W, Perea M, Fregene M (2004) Diversidad y diferenciación genética de la yuca (*Manihot esculenta* Crantz) con marcadores microsatélites en poblaciones de África y Latinoamérica. Acta Biológica Colombiana. 9 (2).

Ceballos H, Iglesias C, Perez J, Dixon A (2004) Cassava breeding: opportunities and challenges. Plant Molecular Biology 56: 503-516.

Centro Internacional de Agricultura Tropical, CIAT (2002) La Yuca en el tercer milenio: sistemas modernos de producción, procesamiento, utilización y comercialización. Cali, Colombia. 586 pp.

Chavarriaga-Aguirre P, Maya M, Bonierbale M, Kresovich S, Fregene M, Tohme J, Kochert G (1998) Microsatellites in cassava (*Manihot esculenta* Crantz): discovery, inheritance and variability. Theoretical and Applied Genetics 97: 493-501.

Cock J (1985) Cassava: new potential for a neglected crop. Boulder: Westview Press, London.

Dellaporta S, Wood J, Hicks J (1983) A plant DNA minipreparation: Version II. Plant Molecular Biology Reporter 1 (14): 19-21.

Escobar, R (1991). Estudio comparativo de dos métodos de propagación de la yuca (*Manihot esculenta* Crantz) *in vitro*. Tesis B. S. Universidad Santiago de Cali. Cali, Colombia. 115 pp.

Fregene M, Vargas J, Ikea J, Angel F, Tohme J, Asiedu R, Akorada M, Roca W (1994) Variability of chloroplast DNA and nuclear ribosomal DNA in cassava (*Manihot esculenta* Crantz) and its wild relatives. Theoretical and Applied Genetics 89:719–727.

Fregene M, Angel F, Gómez R, Rodríguez F, Chavarriaga P, Roca W, Tohme J, Bonierbale M (1997) A molecular genetic map for cassava (*Manihot esculenta* Crantz). Theoretical and Applied Genetics 95: 431-441.

Fregene M, Okogbenin E, Mba C, Angel F, Suárez M, Gutierrez J, Chavarriaga P, Roca W, Bonierbale M, Tohme J (2001) Genome mapping in cassava improvement: challenges, achievements and opportunities. Euphytica 120: 159-165.

Fregene M, Suárez M, Mkumbira J, Kulembeka H, Ndedya E, Kulaya A, Mitchel S, Gullberg U, Rosling H, Dixon A, Dean R, Kresovich S (2003) Simple sequence repeat marker diversity in cassava landraces: genetic diversity and differentiation in an asexually propagated crop. Theoretical and Applied Genetics 107: 1083-1093.

Gamborg O, Miller R, Ojima K (1968) Nutrient requirement of suspension cultures of soybean and carrot cells. Experimental Cell Research 50: 151-158.

Goenaga R, Rivera-Amador E, Chardon U (2002) Yield performance of introduced cassava clones in an ultisol in Puerto Rico. Journal of Agriculture of the University of Puerto Rico 86: 27-33.

Goudet J (1995) FSTAT (version 1.2): A computer program to calculate F-statistics. Journal of Heredity 86: 485-486.

Gregorius H (1987) The relationship between the concepts of genetic diversity and differentiation. Theoretical and Applied Genetics 74: 397-401.

Hankoua B, Ng S, Fawole I, Pounti-Kaerlas J, Pillay M, Dixon A (2005) Regeneration of a wide range of African cassava genotypes via shoot organogenesis from cotyledons of maturing somatic embryos and conformity of the field-established regenerants. Plant Cell, Tissue and Organ Culture 82: 221-231.

Hedrick P (1942) Genetics of populations. Third ed. Jones and Bartlett Publishers. USA., 737pp.

Heysom H, Chris-Chan T, Hughes M (1994) Phylogenetic relationships of *Manihot* species revealed by restriction fragment length polymorphism. Euphytica. 76: 227-234.

Hokanson S, Szewc-McFadden A, Lamboy W, McFerson J (1998) Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus x domestica* borkh. Core subset collection. Theoretical and Applied Genetics 97: 671-683.

Hurtado P, Olsen K, Buitrago C, Ospina C, Marin J, Duque M, de Vicente C, Wongtiem P, Wenzel P, Killian A, Adeleke M, Fregene M (2008) Comparison of simple sequence repeat (SSR) and diversity array technology (DArT) markers for assessing genetic diversity in

cassava (*Manihot esculenta* Crantz) Plant Genetic Resources: Characterization and Utilization. 6 (3): 208-214.

Judd W, Campbell C, Kellogg E, Stevens P (1999) Plant systematics: a phylogenic approach. Sinauer Associates, Inc. Sunderland, Massachusetts. 464 pp.

Kawano K (2003) Thirty years of cassava breeding for productivity-biological and social factors for success. Crop Science 43: 1325-1335.

Kijas J, Fowler J, Thomas M (1995) An evaluation of sequence-tagged microsatellite site markers for genetic analysis within *Citrus* and related species. Genome 38: 349-355.

Kizito E, Bua A, Fregene M, Egwang T, Gullberg U, Westerberg A (2005) The effect of cassava mosaic disease on the genetic diversity of cassava in Uganda. Euphytica 146: 45-54.

Kizito E (2006) Genetic and root growth studies in cassava (*Manihot esculenta* Crantz): implications for breeding. Thesis Ph.D. Swedish University of Agricultural Sciences. 47 pp.

Kumar L (1999) DNA markers in plant improvement: an overview. Biotechnology Advances 17: 143-182.

Lavi U, Akkaya M, Bhagwat A, Lahav E, Cregan P (1994) Methodology of generation and characteristics of simple sequence repeat DNA markers in Avocado (*Persea americana* M.) Euphytica 80: 171-177.

Marín J (2002) Mapeo genético y análisis asistido por marcadores microsatélites de la genética del rendimiento en una población F_2 de yuca (*Manihot esculenta* Crantz). Tesis B.S. Universidad del Tolima, Facultad de Ciencias, Ibagué, Colombia, 111 pp.

Mathews H, Schopke C, Carcamo R, Chavarriaga P, Fauquet C, Beachy R (1993) Improvement of somatic embryogenesis and plant recovery in cassava. Plant Cell Reports 12: 328-333.

Mba R, Stephenson P, Edwards K, Melzer S, Nkumbira J, Gullberg U, Apel K, Gale M, Tohme J, Fregene M (2001) Simple sequence repeat (SSR) markers survey of the cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. Theoretical and Applied Genetics 102: 21-31.

Morgante M, Olivieri A (1993) PCR-amplified microsatellites as markers in plant genetics. The Plant Journal 3: 175-182.

Moyib O, Odunola O, Dixon A (2007) SSR markers reveal genetic variation between improved cassava cultivars and landraces within a collection of Nigerian cassava germplasm. African Journal of Biotechnology 6: 2666-2674.

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue culture. Plant Physiology 15: 473-497.

Nei M (1972) Genetic Distance between Populations. American Naturalist 106: 283-292

Nei M (1973) Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Science 70: 3321-3323.

Nei M (1977) F-Statistics and analysis of gene diversity in subdivided populations.
Annals of Human Genetics 41:225-233.

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

Nei M and Chesser R (1983) Estimation of fixation indices and gene diversities. Annals of Human Genetics 47: 253-259.

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York, NY.

Oetting W, Lee H, Flanders D, Wiesner G, Sellers T, King R (1995) Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13-tailed primers. Genomics 30: 450-458.

Olsen K and Schaal B (1999) Evidence on the origin of cassava: phylogeography of *Manihot esculenta*. Proceedings of the National Academy of Science 96: 5586-5591.

Olsen K and Schaal B (2001) Microsatellite variation in cassava (*Manihot esculenta*, Euphorbiaceae) and its wild relatives: further evidence for a southern Amazonian origin of domestication. American Journal of Botany 88: 131-142.

Perry B (1943) Chromosome number and phylogenetic relationships in the Euphorbiaceae. American Journal of Botany 30: 527-543.

Phillips T, Taylor D, Sanni L, Akoroda M (2004) A cassava industrial revolution in Nigeria: the potential for a new industrial crop. IFAD, FAO, Rome.

Pujol B, David P, McKey D (2005) Microevolution in agricultural environments: how a traditional Amerindian farming practice favors heterozygosity in cassava (*Manihot esculenta* Crantz, Euphorbiaceae). Ecology Letters 8: 138-147.

Raemakers C, Schavemaker C, Jacobsen E, Visser R (1993) Improvements of cyclic somatic embryogenesis of cassava (*Manihot esculenta* Crantz). Plant Cell Reports. 12: 226-229.

Raven P, Fauquet C, Swaminathan M, Borlaug N, Samper C (2006) Where next for genome sequencing? Science 311: 468.

Renvoize B (1972) The area of origin of *Manihot esculenta* as a crop plant: a review of the evidence. Economic Botany 26: 352-360.

Roa A, Maya M, Duque M, Tohme J, Allem A, Bonierbale M (1997) AFLP analysis of the relationships among cassava and other *Manihot* species. Theoretical and Applied Genetics 95: 741-750.

Roa A, Chavarriaga-Aguirre P, Duque M, Maya M, Bonierbale M, Iglesias C, Tohme J (2000) Cross-species amplification of cassava (*Manihot esculenta*) (Euphorbiaceae) microsatellites: allelic polymorphism and degree of relationship. American Journal of Botany 87: 1647-1655.

Rocha O, Zaldivar M, Castro L, Castro E, Barrantes R (2008) Microsatellite variation of cassava (*Manihot esculenta* Crantz) in home gardens of chibchan Amerindians from Costa Rica. Conservation Genetics 9: 107-118.

Rogers D (1963) Studies on *Manihot esculenta* Crantz (cassava) and related species. Bulletin of the Torrey Botanical Club 99: 43-54.

Rogers D and Appan S (1973) Manihot and Manihotoides (Euphorbiaceae): a computer assisted study. Hafner Press, New York, pp 272.

Rohlf, F (1993) NTSYS-PC Numerical taxonomy and multivariate analysis system. Version 1.8 Exeter Publ., Setauket, New York.

Sarria R, Ocampo C, Ramírez H, Hershey C, Roca W (1992) Genetics of esterase and glutamate oxaloacetate transaminases isozymes in cassava. In: Roca WM, Thro AM (eds) Cassava Biotechnology Network. Proceedings of the first international scientific meeting of cassava biotechnology network held at CIAT. Cartagena de Indias, Colombia. 25-28 August. CIAT working doc. pp 62-64.

Scott L, Cross M, Shepherd M, Maguire T, Henry R (1999) Increasing the efficiency of microsatellite discovery from poorly enriched libraries in coniferous forest species. Plant Molecular Biology Reporter 17: 351–354.

Siritunga D (2002) Generation of acyanogenic cassava (*Manihot esculenta* Crantz): transgenic approaches. Thesis Ph.D. The Ohio State University, USA., 201 pp.

Siritunga D and Sayre R (2003) Generation of cyanogen-free transgenic cassava. Planta 217: 367-373.

Siritunga D, Arias-Garzón D, White W, Sayre R (2004) Over-expression of hydroxynitrile lyase in transgenic cassava roots accelerates cyanogenesis and food detoxification. Plant Biotechnology Journal 2: 37-43.

Sneath P and Sokal R (1973) Numerical taxonomy. Freeman, San Francisco, pp 573.

Stamp J and Henshaw G (1987) Somatic embryogenesis from clonal leaf tissues of cassava. Annals of Botany 59: 445-450.

Szabados L, Hoyos R, Roca W (1987) In vitro somatic embryogenesis and plant regeneration of cassava. Plant Cell Reports 6: 248-251.

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acid Research 17: 6463-6471.

Tonukari N (2004) Cassava and the Future of Starch. Electronic Journal of Biotechnology. http://www.ejbiotechnology.info/content/vol7/issue1/issues/2/

Vekemans X and Lefevre C (1997) On the evolution of heavy metal tolerant populations in *Armenia maritima*: evidence from allozyme variation and reproductive barriers. Journal of Evolutionary Biology 10: 175-191.

Weir B and Cockerham C (1984) Estimating *F*-statistics for the analysis of population structure. Evolution 38: 1358-1370.

Wright S (1965) The interpretation of population structure by *F*-statistics with special regard to systems of mating. Evolution 19: 395–420.

Wright S (1978) Evolution and the genetics of population, variability within and among natural populations. The University of Chicago Press, Chicago.

Zárate, L (2002) Mapeo genético de una Población F1 de yuca cultivada (*Manihot esculenta* Crantz) utilizando microsatélites. Tesis B.S. Universidad del Tolima Ibagué, Colombia, 76pp.

Zhang P (2000) Studies on cassava (*Manihot esculenta* Crantz) transformation: towards genetic improvement. Thesis Ph.D. Swiss Federal Institute of Technology Zürich, 129 pp.

Zho Y, Bui T, Auckland L, Williams C (2002) Direct fluorescent primers are superior to M13-tailed primers for *Pinus taeda* microsatellites. BioTechniques 32: 46-52

Appendix A. Dellaporta et al. 1983 modified protocol for DNA extraction from *in vitro* cassava leaves

1. Grind 0.15-0.2 g of *in vitro* cassava leaves with sterile sea sand and a pestle in a 1.5 ml Eppendorf tube containing 200 µl of (0.1 M of Tris-HCl pH 8.0, 0.05 M of EDTA, 0.5 M of NaCl, 1% of PVP and 0.07% of β-mercaptoethanol) extraction buffer. The temperature of the extraction buffer must be at 65 °C prior to adding it to each sample leaf.
2. When the tissue is completely ground it, add the additional 400 µl of extraction buffer and 50 µL of 20% SDS. Mix by inverting 5-8 times.
3. Incubate the samples at 65 °C in a water bath for 15 min, briefly vortexing every five min.
4. Let the samples cool down at room temperature for two min.
5. Add 250 µL of ice-cold (-20 ºC) 5 M potassium acetate to each sample, mixing by inverting 5-8 times. Then, incubate them in ice for 20 min.
6. Centrifuge the samples at 13,000 rpm for 15 min at room temperature.
7. Transfer the supernatant with a pipette to a new 1.5 mL eppendorf tube and add 700 µl of ice-cold (-20 ºC) isopropanol. Mix the samples softly by inverting 5-8 times.
8. Incubate the samples at -20 ºC for 1 hour or overnight.
9. Centrifuge the samples at 13,000 rpm for 15 min at room temperature.
10. Let the pellet dry at room temperature inverting the tubes on paper towels and wait until all isopropanol drops disappear from the walls of the tubes. Isopropanol may interfere in further PCR reactions.
11. Wash the pellet with 500 µL of ice-cold (-20 ºC) 70 % ethanol.
12. Centrifuge the samples at 13,000 rpm for 5 min.
13. Let the pellet dry at room temperature following the recommendations made in step 10.
14. Resuspend the pellet in 50 µL of Tris-EDTA 10:1 buffer and add 1 µL of 10mg/mL RNAse A.

15. Incubate the samples at 65 °C in a water bath for 10 min.

Table 10. Dellaporta et al (1983) DNA extraction buffer recipe.

DNA Extraction Buffer		
	1000 ml	**100 ml**
Final Concentration	**Initial Concentration**	
100mM Tris-HCl	100 ml of 1 M Tris–HCl pH 8	10 ml
50mM EDTA	100 ml of 0.5M EDTA	10 ml
500mM NaCl	100 ml of 5 M NaCl	10 ml
ddH_2O	700 ml	70 ml
Autoclave		
1 % PVP	10 g	1 g
beta-Mercaptoetanol	700 µl	70 µl

Appendix B. PCR amplification conditions.

Table 11 . PCR master mix used to amplify SSR loci.

Reagent	Stock concentration	Final concentration	Volume of the reaction
PCR Buffer	10X	1X	1.25 ul
MgCl2	25 mM	2.0 mM	1 ul
dNTPs	2.5 mM	0.2 mM	1 ul
Primer Forward	10uM	0.1 uM	0.125 ul
Primer Reverse	10uM	0.1 uM	0.125 ul
M13 primer	1 pmol/ul	0.5 pmol	0.5 ul
Taq polymerase	5 U/ul	~1U	0.2 ul
ddH2O			6.3 ul
DNA	10 ng/ul	20 ng	2 ul

Table 12. Thermocycling Program.

Cycle	Temperature (ºC)	Time
1	95	5 min
2	94	30 sec
3	45 or 55	1 min
4	72	1 min
5	34 cycles since step 2	
6	72	5 min
7	10	forever

Appendix C. Polyacrylamide gel preparation and electrophoresis protocols

Reagent Preparation

Stock Bind Silane solution

Bind Silane is used to covalently bind the gel to the glass in the area where the comb is inserted. This treatment helps to maintain good well morphology when loading gels multiple times.

1. Add 50 µl of bind Silane to 10 ml of 100 % ethanol.
2. Mix well and store at 4 °C in an amber colored bottle.

Work Bind Silane solution

This solution is used to apply it over the area where the comb will be placed. Mix 25 µl of stock bind silane solution and 25 µl of 10% acetic acid in a 1.5 ml microcentrifuge tube. Mix thoroughly (pipette or vortex).

Ammonium persulfate (APS) solution

APS is a chemical that provides a source of free radicals needed for polymerization of the gel. A 10 % APS solution is made by adding 0.1 g APS powder to 1 ml ddH_2O in a 1.5 ml microcentrifuge test tube.

TBE 10X (1 L)

This solution is mainly used as a running buffer during the electrophoresis.

Tris Base	107.8 g
EDTA	7.44 g
Boric Acid	55 g

Complete with ddH_2O until 1 L.

Polyacrylamide 6.5 % KB Plus solution

1. Bring 20 ml of KB Plus to room temperature (takes 10-15 min) and meanwhile prepare glass plates for gel injection while KB Plus warms.
2. Add 150 µl of 10% APS and 15 µl TEMED when ready to inject gel solution.

Mix to homogenate and draw the gel solution into a 60 cc syringe with 14 gauge tip.

Cleaning the plates and preparing gel sandwich

Materials needed:

• 2% laboratory detergent solution made from a concentrate such as Micro 90 (International Products Corp., Burlington NJ).

• Spray bottle, containing 70% ethanol.

1. Pour a small amount of 2% Micro 90 solution onto the side of the plate that will contact the gel.
2. Work the solution into lather with a bristle brush and thoroughly scrub the entire plate. Remove any dried-on polyacrylamide and rinse well with tap water.
3. Buffer solution (TBE 1X) that has dried onto the plates can be removed with 1N NaOH.
4. Pour distilled water onto the plate working the water around the plate with a gloved hand and watch for bubble formation from detergent residue. Repeat until no bubbles form.
5. Rinse the plates with deionized distilled water.
6. Use a pump spray bottle to spray a fine mist of 70% ethanol/ddH_2O for a final rinse. Stand plates in a rack to air dry.
7. Wipe both plates (gel side) with 70% isopropanol using Kimwipes as needed to remove dust particles on the plates. Allow the plates to dry for 3 min.
8. Apply the work Silane solution in both plates (gel side) in the area where wells will form. Allow the solution to dry on the plates for 10 min.
9. Assemble the gel sandwich making sure that the rails fit tightly against the edges of both glass plates.
10. Prepare the polyacrylamide solution as described above and inject the gel into the sandwich inserting the sharktooth comb upside down.

11. Allow the gel to polymerize for 1. 5 hours (LI-COR 2004)

Table 13. Standardized conditions to run cassava SSR loci.

Parameter	Standardized Microsatellite Run
Plate Length	25 cm
Spacer Thickness	0.25 mm
Gel Composition	6.5 % LI-COR KB Plus
Pre-Run time	25 min
Run Time	1.45 hours
Reload Gel	1. 45 hours
Voltage	1500
Power	40
Current	40 mA
Temperature	45 °C

Loading samples

1. Dilute 0.5 µl of the PCR product in 5 µl of Blue Stop solution (LI-COR) and denature it at 95 °C for 3 min. Denature the samples few min before its loading on the gel.
2. Incubate the samples on ice until loading it in the well.
3. Pour on the well only 0.5 µl of the dilution.

Appendix D. Software description

FSTAT

FSTAT (Goudet 1995) is a computer package for PC which estimates gene diversity and differentiation statistic estimators from codominant genetic markers. The program computes both Nei and Weir and Cockerham estimators of gene diversities, and F-statistics and tests them using randomizing methods (bootstrapping and jackknifing resampling schemes).

GENSURVEY

This program estimates several classic population genetic parameters from individual genotype data (isozymes, RFLP, microsatellites). It also performs several statistical tests using numerical resampling methods. The program estimates for each population, each locus and each allele frequency, the observed heterozygosity and the inbreeding coefficient corrected for small samples. It also estimate for each population, averaged over loci, the within population variation, and makes an analysis of gene diversity which gives, for each polymorphic locus (at the 5% level),the gene diversity in the total set of populations (Ht), the average gene diversity within populations (Hs) and the average gene diversity between populations (Dst). Then shows the means of these estimates over loci and their standard deviations, and the value of the coefficient of gene differentiation (Gst=Dst/Ht), the interpopulational gene diversity (Dm) and the interpopulational gene diversity relative to the intrapopulational gene diversity (Rst=Dm/Hs). The same set of values is given using the corrections for small sample sizes and 95% confidence intervals for the means are computed by bootstrapping over loci. Additionally, the program gives the F-statistics computed according to Weir and Cockerham (1984) and also computes the pairwise genetic distances between pairs of populations using Nei (1972) and corrected Nei distances (Nei 1978), among others.

NTSYS

This is a computer package of programs that is used to find and display structure in multivariate data. The program was originally developed for use in biology in the context of the numerical taxonomy. The methods furnished in NTSYS are associated with the field of phenetics, but it also can be used in cladistics. It can be used to compute various measures of similarity or dissimilarity between all pairs of objects and then summarize this information either in terms of nested sets of similar objects (*cluster analysis*) or in terms of a spatial arrangement along one or more coordinate axes (*ordination analysis* or various types of *multidimensional scaling analysis*). The UPGMA procedure in the SAHN module is one of the most commonly used on molecular data.

Appendix E. FSTAT output

```
************************************************************************
************************************************************
*   The following results were generated the 2008-08-28 at 20:43:28 with
Fstat for windows, V2.9.3  2 (Feb. 2002) from file DavidFSTAT.dat. *
************************************************************************
Weir & Cockerham (1984) estimation of Fit (CapF), Fst (theta) and Fis
(smallF).
      relat is Relatedness estimated following Queller & Goodnight
(1989)
      relatc is relatedness inbreeding corrected following Pamilo (1984,
1985)
      sig_a, sig_b and sig_w are the component of variance
      among samples, among individuals within samples and within
individuals respectively.


       Jackknifing over loci.

              Capf       Theta      Smallf       Relat
       total       0.089       0.047       0.045       0.085       Means
                   0.024       0.005       0.025       0.010       Std.
Err.

***********************************************
               Bootstrapping over Loci.

               95% Confidence Interval.

              CapF       theta      Smallf       Relat
             0.042       0.036      -0.003       0.067
             0.137       0.057       0.093       0.104

               99% Confidence Interval.

              CapF       theta      Smallf       Relat
             0.028       0.033      -0.017       0.061
             0.151       0.061       0.107       0.110

Fst Values:

1         8         8
          AFR       BRA       COL       CR        ECU       GUA       MEX       PR
AFR       0         0.0627    0.0729    0.0396          NA  0.163     0.111     0.0891
BRA       0.0627    0         0.0109    0.0503          NA  0.0934    0.0468    0.0267
COL       0.0729    0.0109    0         0.0437          NA  0.1063    0.0481    -0.0024
CR        0.0396    0.0503    0.0437    0               NA  -0.0365   0.0528    0.0737
ECU            NA        NA        NA        NA    0               NA        NA        NA
GUA       0.163     0.0934    0.1063    -0.0365         NA  0         0.1342    0.1315
MEX       0.111     0.0468    0.0481    0.0528          NA  0.1342    0         0.0632
PR        0.0891    0.0267    -0.0024   0.0737          NA  0.1315    0.0632    0
```

Table 14. Proportion of differentiation among populations (G_{ST}=coefficient of gene differentiation) computed according to Nei and Chesser (1983) and *F*-statistics computed according to Weir & Cockerham (1984) over loci. 95% and 99% confidence intervals for the means were computed by bootstrapping (1000 bootstraps) over loci. The mean value over loci is given together with the mean and standard deviation obtained through jacknifing over loci (200 replications).

Locus	Gst	Fit (CapF)	Fst (Theta)	Fis (SmallF)
SSRY4	0.041	0.192	0.033	0.165
SSRY9	0.0365	0.004	0.027	-0.024
SSRY12	0.0434	0.043	0.053	-0.011
SSRY19	0.0456	0.063	0.039	0.025
SSRY20	0.0655	-0.006	0.051	-0.06
SSRY21	0.1286	-0.096	0.064	-0.17
SSRY34	0.1003	0.296	0.133	0.188
SSRY51	-0.021	0.129	0.038	0.095
SSRY52	0.0435	-0.057	0.052	-0.114
SSRY59	0.0761	0.169	0.085	0.092
SSRY63	0.1085	0.433	0.121	0.355
SSRY64	0.0127	0.074	0.041	0.034
SSRY69	0.0045	0.1	0.011	0.09
SSRY100	0.0899	0.122	0.06	0.066
SSRY102	-0.0062	0.017	-0.002	0.019
SSRY103	0.0095	0.088	0.041	0.049
SSRY106	0.007	0.033	0.031	0.002
SSRY108	0.0424	-0.149	0.033	-0.188
SSRY120	0.0803	-0.11	0.051	-0.171
SSRY135	0.1015	0.161	0.107	0.061
SSRY147	-0.0318	0.031	-0.02	0.049
SSRY148	0.1063	-0.265	0.03	-0.304
SSRY151	0.181	-0.02	0.049	-0.073
SSRY155	0.096	-0.079	0.066	-0.155
SSRY161	0.0522	0.221	0.079	0.154
SSRY164	0.0614	0.19	0.054	0.144
SSRY175	0.0604	0.071	0.05	0.022
SSRY177	0.0527	0.082	0.022	0.062
SSRY179	0.0565	0.182	0.03	0.156
SSRY181	-0.0371	0.132	-0.009	0.14
SSRY182	0.0152	0.165	0.046	0.124
NS189	-0.0807	0.109	-0.019	0.126
NS376	-0.0198	0.396	0.006	0.393
NS911	0.071	0.17	0.033	0.142
Mean BS	).0468 ± 0.053	-	-	-
Mean JK	-	0.089 ± 0.024	0.047 ± 0.005	0.045 ± 0.025
BS CI 95%	0.029 0.0646	0.042 0.137	0.036 0.057	-0.003 0.093
BS CI 99%	-	0.028 0.151	0.033 0.061	-0.017 0.107

Appendix F. Eigenvalues (NTSYS output)

SimQual: NTSYSpc 2.02i, (C) 1986-1998, Applied Biostatistics Inc.
Date & time: 8/29/2008 2:42:13 PM

Input parameters
Read input from file: D:\Matrices\Matrices_2008\Matriz final All markers and samples 2008\David\David final\DavidFinal.NTS
Compute by: cols
Save results in output file: DavidFSimQual
Coefficient: J
Positive: 1.0000
Negative: 0.0000

Matrix type =1, size =298 by 86, missing value code =9 (rectangular)
Result will be a 86 by 86 matrix
Results stored in file: DavidFSimQual

dcenter: NTSYSpc 2.02i, (C) 1986-1998, Applied Biostatistics Inc.
Date & time: 8/29/2008 2:47:13 PM

Input parameters
Read input from file: D:\Matrices\Matrices_2008\Matriz final All markers and samples 2008\David\David final\DavidFSimQual
Save results in output file: DavidFCen
Square matrix elements

Comments:
SIMQUAL: input=D:\Matrices\Matrices_2008\Matriz final All markers and samples 2008\David\David final\DavidFinal.NTS, coeff=J
by Cols, += 1.00000, -= 0.00000
Matrix type =3, size =86 by 86, missing value code ="none" (similarity)
Double-centered matrix stored in file: DavidFCen
eigen: NTSYSpc 2.02i, (C) 1986-1998, Applied Biostatistics Inc.
Date & time: 8/29/2008 2:51:53 PM

Input parameters

Read input from file: D:\Matrices\Matrices_2008\Matriz final All markers and samples 2008\David\David final\DavidFCen
Number of dimensions: 3
Save eigenvectors in output file: DavidFvec
Save eigenvalues in output file: DavidFval
Scaling: SQRT(LAMBDA)

Comments:
SIMQUAL: input=D:\Matrices\Matrices_2008\Matriz final All markers and samples 2008\David\David final\DavidFinal.NTS, coeff=J
by Cols, += 1.00000, -= 0.00000
DCENTER: input=D:\Matrices\Matrices_2008\Matriz final All markers and samples 2008\David\David final\DavidFSimQual type was = 3
Matrix type =3, size =86 by 86, missing value code ="none" (similarity)

i	Eigenvalue	Percent	Cumulative
1	4.44049536	6.8528	6.8528
2	3.42304049	5.2826	12.1355
3	2.51089031	3.8750	16.0105
4	2.03465918	3.1400	19.1505
5	1.97166680	3.0428	22.1933
6	1.92106178	2.9647	25.1580
7	1.81139535	2.7955	27.9534
8	1.61684845	2.4952	30.4486
9	1.57655733	2.4330	32.8817
10	1.57293226	2.4274	35.3091
11	1.43354733	2.2123	37.5215
12	1.39344240	2.1504	39.6719
13	1.31356883	2.0272	41.6991
14	1.24515693	1.9216	43.6207
15	1.17632353	1.8154	45.4361
16	1.12219210	1.7318	47.1679
17	1.11435854	1.7197	48.8877
18	1.04349284	1.6104	50.4980
19	1.00331558	1.5484	52.0464
20	0.97789358	1.5091	53.5556
21	0.95308643	1.4709	55.0264
22	0.93543471	1.4436	56.4701
23	0.92027284	1.4202	57.8903

24	0.90838386	1.4019	59.2921
25	0.86580689	1.3362	60.6283
26	0.84243864	1.3001	61.9284
27	0.82585854	1.2745	63.2029
28	0.80926802	1.2489	64.4518
29	0.78760986	1.2155	65.6673
30	0.76630518	1.1826	66.8499
31	0.74540876	1.1504	68.0003
32	0.73361921	1.1322	69.1325
33	0.71046247	1.0964	70.2289
34	0.69299771	1.0695	71.2984
35	0.68190807	1.0524	72.3507
36	0.66914941	1.0327	73.3834
37	0.65712012	1.0141	74.3975
38	0.64391557	0.9937	75.3913
39	0.63805612	0.9847	76.3759
40	0.61070608	0.9425	77.3184
41	0.60989910	0.9412	78.2597
42	0.60028052	0.9264	79.1860
43	0.58509523	0.9030	80.0890
44	0.57671011	0.8900	80.9790
45	0.56400107	0.8704	81.8494
46	0.55553633	0.8573	82.7068
47	0.54865748	0.8467	83.5535
48	0.54014478	0.8336	84.3871
49	0.52765789	0.8143	85.2014
50	0.51574485	0.7959	85.9973
51	0.50514520	0.7796	86.7769
52	0.49268462	0.7603	87.5372
53	0.48456162	0.7478	88.2850
54	0.47205086	0.7285	89.0135
55	0.46450188	0.7168	89.7304
56	0.45481555	0.7019	90.4323
57	0.44626308	0.6887	91.1210
58	0.43710050	0.6746	91.7955
59	0.42542148	0.6565	92.4521
60	0.41530043	0.6409	93.0930
61	0.41075656	0.6339	93.7269
62	0.40312964	0.6221	94.3490

63	0.39649655	0.6119	94.9609
64	0.38380271	0.5923	95.5532
65	0.38171279	0.5891	96.1423
66	0.35843767	0.5532	96.6955
67	0.27401662	0.4229	97.1184
68	0.24453359	0.3774	97.4957
69	0.23357862	0.3605	97.8562
70	0.18734544	0.2891	98.1453
71	0.14858489	0.2293	98.3746
72	0.13272397	0.2048	98.5795
73	0.13036941	0.2012	98.7807
74	0.11934880	0.1842	98.9648
75	0.10136726	0.1564	99.1213
76	0.09625009	0.1485	99.2698
77	0.09256693	0.1429	99.4127
78	0.08141244	0.1256	99.5383
79	0.07583029	0.1170	99.6553
80	0.05780787	0.0892	99.7446
81	0.05323912	0.0822	99.8267
82	0.04304560	0.0664	99.8931
83	0.03735711	0.0577	99.9508
84	0.01728156	0.0267	99.9775
85	0.01459895	0.0225	> 100%
86	0.00000000	0.0000	> 100%

Sum of eigenvalues = 64.797815

Eigenvalues stored in file: DavidFval

Eigenvectors stored in file: DavidFvec

Average root: 0.7535

Proportions of variance expected using broken-stick model

Eigenvalue	**Expected**
1	5.8574
2	4.6946
3	4.1132
4	3.7256
5	3.4349

Eigenvalue	Expected
6	3.2024
7	3.0086
8	2.8425
9	2.6971
10	2.5679
11	2.4516
12	2.3459
13	2.2490
14	2.1596
15	2.0765
16	1.9990
17	1.9263
18	1.8579
19	1.7933
20	1.7321
21	1.6740
22	1.6186
23	1.5658
24	1.5152
25	1.4668
26	1.4202
27	1.3755
28	1.3325
29	1.2909
30	1.2508
31	1.2121
32	1.1746
33	1.1382
34	1.1030
35	1.0688
36	1.0356
37	1.0033
38	0.9718
39	0.9412
40	0.9114
41	0.8824
42	0.8540

Eigenvalue	Expected
43	0.8263
44	0.7993
45	0.7728
46	0.7470
47	0.7217
48	0.6970
49	0.6728
50	0.6490
51	0.6258
52	0.6030
53	0.5806
54	0.5587
55	0.5371
56	0.5160
57	0.4952
58	0.4748
59	0.4548
60	0.4351
61	0.4157
62	0.3966
63	0.3779
64	0.3594
65	0.3412
66	0.3234
67	0.3057
68	0.2884
69	0.2713
70	0.2544
71	0.2378
72	0.2214
73	0.2053
74	0.1894
75	0.1737
76	0.1582
77	0.1429
78	0.1278
79	0.1128

Eigenvalue	Expected
80	0.0981
81	0.0836
82	0.0692
83	0.0551
84	0.0410
85	0.0272
86	0.0135

Appendix G. Tissue culture media and stock solution preparation

Table 15. 4E medium for cassava micropropagation growth

4E-Shooty Micropropagation Medium			
Parameter/Reagent	**Stock Concentration**	**Final Concentration**	**Quantity g/ Volume (1L)**
MS Salts Powder	-	-	4.3 g
Thiamine-HCl	100 mg/L	1 mg/L	10 ml
Myo-Inositol	8000 mg/L	100 mg/L	12.5 ml
Sucrose	100 %	2 %	20 g
BAP	100 mg/L	0.04 mg/L	400 µl
GA_3	100 mg/L	0.05 mg/L	500 µl
NAA	100 mg/L	0.02 mg/L	200 µl
pH (KOH 1N, 0.1N)		5.7 - 5.8	
Agar Fisher	100 %	0.7 %	7 g

Table 16. MS8 (8 mg/l 2,4-D) medium for embryogenesis induction in Cassava

MS8-Embryogenesis Induction Medium			
Parameter/Reagent	**Stock Concentration**	**Final Concentration**	**Quantity g/ Volume (1L)**
MS Salts Powder	-	-	4.3 g
Vitamins B5	100 X	1 X	10 ml
2,4-D	10,000 mg/L	8 mg/L	800 µl
Sucrose	100 %	2 %	20 g
$CuSO_4$	2mM	2µM	1ml
Casein Hydrolysate	-	-	50 mg
pH (KOH 1N, 0.1N)		5.6 - 5.7	
Phytagel Sigma	100 %	0.2 %	2 g

Table 17. RM1 medium for the germination of somatic embryos.

RM1-Embryo Germination Medium			
Parameter/Reagent	**Stock Concentration**	**Final Concentration**	**Quantity g/ Volume (1L)**
MS Salts Powder	-	-	4.3 g
Thiamine-HCl	100 mg/L	1 mg/L	10 ml
Myo-Inositol	8000 mg/L	100 mg/L	12.5 ml
Sucrose	100 %	2 %	20 g
BAP	100 mg/L	0.1 mg/L	1 ml
GA_3	100 mg/L	0.5 mg/L	5 ml
pH (KOH 1N, 0.1N)		5.7 - 5.8	
Agar Fisher	100 %	0.7 %	7 g

Reagent Preparation

Hormones (10,000 mg/L)

All the stocks of hormones where prepared at 10,000 mg/L and dissolved in DMSO as follow:

0.1 g of hormone powder was diluted in 10 ml of DMSO and stored it at -20 °C. Work hormone solutions were prepared at 100 mg/L diluting 10 µl of the stock hormone solution in 990 µl of ddH_2O and storing it at 4 °C.

Thiamine-HCl (100 mg/L)

0.02 g of Thiamine-HCl were diluted in few drops of HCl 0.1N and the volume was completed with ddH_2O until 200 ml. Solution was stored at 4 °C

Myo-Inositol (8000 mg/L)

1.6 g were diluted in 200 ml of ddH_2O and stored it at 4 °C.

$CuSO_4$ - 5 H_2O (2 mM)

0.025 g of copper sulfate was diluted in 500 ml of ddH_2O and stored it at 4 °C.

Appendix H. Embryogenesis and germination of somatic embryos

Table 18. Embryogenesis and germination results

Accession	No. TE	No. EC	% EC	No. GE	%GE
PI12902	171	0	0.0	0	0
Jamaica_18	81	2	2.5	0	0
Cubana	126	3	2.4	0	0
PI12903	63	8	12.7	2	25
PI12900	189	7	3.7	3	43
Forastera	72	0	0.0	0	0
Llanera	63	0	0.0	0	0
Trinidad 14-56	54	0	0.0	0	0
Brava	97	3	3.1	0	0
Tremesiana	59	12	20.3	6	50
Seda	243	35	14.4	8	23
Serralles	63	0	0.0	0	0
Senon	216	49	22.7	0	0
CM3064	114	12	10.5	7	58
CM3311	54	0	0.0	0	0
CM3380	57	0	0.0	0	0
CM4484	54	0	0.0	0	0
SG804	485	37	7.6	14	38
SM494	81	0	0.0	0	0
SM523	242	22	9.1	0	0
Abuelo	63	0	0.0	0	0
Chilena	81	0	0.0	0	0
Valencia	54	4	7.4	0	0
Mcol2215	434	50	11.5	44	88

TE = Number or Total Explants used; EC = Number of explants with clumps of embryos; % EC = Percentage of embryo clumps; GE = Number of germinated somatic embryos; % GE = Percentage of germinated somatic embryos.