EVALUATION OF NEW EXPRESSED SIMPLE SEQUENCE REPEAT MARKERS FOR GENETIC STUDIES AND IDENTIFICATION OF CASSAVA BROWN STREAK DISEASE RESISTANCE

BY

GERADINA PANTALEO MZENA.

01 OCT 2013

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE IN CROP SCIENCE OF SOKOINE UNIVERSITY OF AGRICULTURE. MOROGORO, TANZANIA.

2010

ABSTRACT

Evaluation for effectiveness of new expressed simple sequence repeat (ESSR) marker is an important component in the genetic improvement of cassava crop for agricultural development. ESSRs are often the markers of choice for genetic studies and identification of CBSD resistance. This study was aimed at: (i) Determining the optimum PCR conditions on annealing temperature and $MgCl_2$ concentrations. ii) Identifying polymorphic loci for ESSR markers for genetic studies of cassava (iii) Identifying polymorphic loci of ESSR markers linked to CBSD. PCR amplification of each ESSR locus was optimized for annealing temperature and $MgCl_2$ concentration. Variation at each ESSR locus was assessed in 24 cassava genotypes from 11 different countries all over the world. Successful amplification was achieved from a total of 68 primer pairs of which 53 showed polymorphism within the diversity of cassava genotypes panel. A total of 231 alleles were observed with an average of 4.35 alleles per marker and a range of 2 to 11 alleles. The PIC values measured by the frequency of alleles ranged from 0.173 to 0.844 with average allele frequency of 4.91. Fourteen (14) primer pairs were polymorphic to identification of genotype linked to CBSD resistance total of 63 alleles with average of 4.5 alleles per locus. These markers will be useful for application in genetic conservation and plant breeding. The genetic diversity observed by dendrogram revealed four major clusters and one sub clusters, varieties most closely related were Per 458 and Mex 55 (D=0.1732), TME 1389, and TME 1368 (D=0.1982), Mkita and Kalolo (D=0.3981) followed by Pesangani and Kigoma (0.4912) while distantly related genotype were these varieties BRA 990 and Mex 55 (0.7089), Albert and Mkita (0.6036), BRA255 and Cub1 (0.6054), TME 230 and TME 539 (D=0.5914), and Namikonga and Albert (0.5139).

DECLARATION

I, Geradina Pantaleo Mzena, do hereby declare to the Senate of Sokoine University of Agriculture that this dissertation is my own original work and that it has neither been submitted nor being concurrently submitted for a degree award in any other institution.

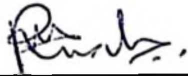
08/11/2010

Geradina Pantaleo Mzena

Date

(MSc Candidate)

The above declaration is confirmed

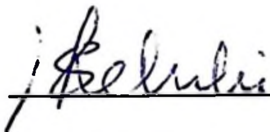


08/11/2010

Dr. P. Kusolwa

Date

(Supervisor)



08/11/2010

Prof. S. O. W. M Reuben

Date

(Supervisor)

COPYRIGHT

No part of this dissertation may be reproduced, stored in any retrieval system, or transmitted in any form or by any means without prior written permission of the author or Sokoine University of Agriculture in that behalf.

ACKNOWLEDGEMENT

I wish to express my profound gratitude to my supervisors Dr. P. Kusolwa and Prof. S. O. W. M Reuben of Department of Crop Science and production of Sokoine University of Agriculture and Dr. M. Ferguson of the International Livestock Research Institute for their intellectual motivation, guidance and valuable comments without which this work would have been difficult to accomplish. May Almighty God bless you all. Special thanks are also due to the entire staff of the Department of Crop Science and Production for their co-operation and assistance. Their critical comments during the time of the entire study significantly contributed to the success of this study. Acknowledgements are extended to the different administrators of the ILRI- Nairobi for granting permission to undertake this study. Mr. I. Nzuki, the Lab manager of ILIRI is highly acknowledged for his support.

I would also like to express gratitude to my family, my parents, the rest of the family and relatives for their love, constant support, encouragement and inspiration during the whole time that I was undertaking this study. My gratitude also goes to the Ministry of Agriculture, Food Security and Cooperatives and BeCa for sponsoring my studies at Sokoine University of Agriculture. I would also like to thank my employer Ministry of Agriculture, Food Security and Cooperatives for granting me a study leave to enable me pursue the same. Thanks are also due to my fellow postgraduate students in the Department of Crop Science and Production for their moral and material support during the course of this study.

DEDICATION

This work is dedicated to my Almighty God who allowed me to accomplish this work without him this work could not be possible. Most to my parents Pantaleo Dominic Mzena and Agatha Bruno Mzena. I also dedicate to my lovely husband Elia A. Kibona, my lovely son Lusekelo E Kibona and my lovely daughter Nitike E. Kibona for their support throughout my study period.

TABLE OF CONTENTS

ABSTRACT ii
DECLARATION iii
COPYRIGHT iv
ACKNOWLEDGEMENT v
DEDICATION vi
TABLE OF CONTENTS vii
LIST OF TABLES x
LIST OF FIGURES xi
LIST OF APPENDICES xii
LIST OF ABBREVIATIONS xiii

CHAPTER ONE 1
1.0 INTRODUCTION 1
1.1 Background Information 1
1.1.1 Cassava production in Africa 2
1.1.2 Cassava production in Tanzania 2
1.1.3 Production constraints of cassava in Tanzania 4
1.1.3.1 Insect- Pests and Diseases 4
1.1.3.1.1 Cassava brown streak disease (CBSD) 4
1.1.3.1.2 Cassava mosaic diseases (CMD) 5
1.1.3.2 Shortage of planting materials and continuous use of cassava varieties with low genetic potential. 5
1.1.4 Cassava breeding in Tanzania 5

1.2 Justification 6
1.3 Objectives 7
1.3.1 Overall objective 7
1.3.2 Specific objectives 7

CHAPTER TWO 8
2.0 LITERATURE REVIEW 8
2.1 Cassava: The Origin and Distribution of the Plant 8
2.1.1 Origin 8
2.1.2 Introduction and spread of cassava in Africa and particularly Tanzania 8
2.2 Identification of Polymorphic Loci for Genetic Diversity and Relatedness of Cassava 9
2.3 Optimization PCR Condition on Annealing T° and $Mgcl_2$ Concentrations 11
2.4 Uses of Markers in Cassava 12
2.4.1 QTL mapping 12
2.4.2 Marker assisted selection (MAS) 12
2.5 Useful Molecular Markers in Cassava 13
2.5.1 Restriction Fragment Length Polymorphism (RFLP) 13
2.5.2 Randomly Amplified Polymorphic DNA (RAPDs) 13
2.5.3 Amplified Fragment Length Polymorphisms (AFLPs) 14
2.5.4 Microsatellites (Simple Sequence Repeats) (SSRs) 14
2.5.5 Expressed sequence tags 15
2.6 Identification of Polymorphic Loci of ESSR Markers linked to CBSD resistance 16
2.7 Suitability of ESSR Markers for Genetic Diversity and Resistance to CBSD 16

CHAPTER THREE 18

3.0 METHODOLOGY 18

3.1 Planting Materials 18

3.2 Genomic DNA Extraction 19

3.3 PCR Optimization 20

3.4 Suitability of ESSR Primers for Genetic Diversity 21

3.5 Identification of ESSR Loci for Resistance to Cassava Brown Streak Diseases 22

CHAPTER FOUR 23

4.0 RESULTS AND DISCUSSION 23

4.1 Optimization of PCR Condition 23

4.2 Identification of Polymorphic Loci for Genetic Diversity and Relatedness of Cassava 23

4.3 Genetic Relatedness of Cassava Genotypes 23

4.4 Identification of ESSR Loci linked to Resistance to Cassava Brown Streak Disease 27

4.5 Discussion 30

CHAPTER FIVE 36

5.0 CONCLUSIONS AND RECOMMENDATIONS 36

5.1 Conclusions 36

6.2 Recommendations 37

REFERENCES 38

APPENDICES 47

LIST OF TABLES

Table 1: The major cassava producing countries worldwide 2

Table 2: List of 24 cassava genotypes collected from different countries 18

Table 3: Genetic distances between genotypes as revealed by markers in this genetic diversity array 26

Table 4: Allelic diversity of ESSRs loci differentiating Namikonga (resistant) and Albert (susceptible) to CBSD 27

Table 5: Number of alleles, allele size, diversity, degree of availability and PIC of ESSR markers observed in genotypes of cassava with respect to CBSD resistance 29

LIST OF FIGURES

Figure 1: Cassava-producing areas of Tanzania.. 3

Figure 2: Phylogenetic relationships among cassava genotypes on the basis of Wright's (1978) modification of Rogers' genetic distance using UPGMA clustering. 95% Bootstrap value was performed to test the robustness of the dendrogram topology. 25

Figure 3: Namikonga (resistance genotype) plot of marker ESSR 76 showing heterozygote individuals. 28

Figure 4: Albert (susceptible genotype) plot of marker ESSR 76 showing homozygote individuals. 28

LIST OF APPENDICES

Appendix 1: List of ESSR primers and target repeats and sequences used in the study 47

Appendix 2: Optimized PCR conditions for amplification of ESSR primers 50

Appendix 3: Number of alleles, allele size, diversity, degree of availability and PIC of ESSR markers observed in different genotypes of cassava 52

Appendix 4: Peaks showed the Effect of Magnesium Chloride concentrations 56

Appendix 5: Appendices 4 List of allele, frequency (freq), variance and standard (SD) deviation 57

LIST OF ABBREVIATIONS

ABI	-	Applied Biosytems
ACMV	-	African Cassava Mosaic Virus
AFLP	-	Amplified Fragment Length Polymorphism
BecA		Biosciences for Eastern and Central Africa
CAD	-	Cassava Anthracnose Disease
CBB	-	Cassava Bacterial Blight
CBSD	-	Cassava Brown Steak Disease
CCD	-	Charged Coupled Device
CGM	-	Cassava Green Mite
CIAT	-	International Centre for Tropical Agriculture
CM	-	Cassava Mealy Bug
CMD	-	Cassava Mosaic Diseases
COSCA	-	Collaborative Study of Tropic in Africa
cpDNA	-	Chloroplast DNA
DNA	-	Deoxyribonucleic Acid
dUTP	-	Deoxyuridine Triposphate
EACMV	-	East African Cassava Mosaic Virus
EDTA	-	Ethylenediamine Tetraacetic Acid
ESSR	-	Expressed Simple Sequence Repeat
ESTs	-	Expressed Sequence Tags
F/R	-	Forward and Reverse
FAO	-	Food and Agriculture Organization
GCP	-	Generation Challenge Project

IGR	-	Institute of Genomic Research
IITA	-	International Institute of Tropical Agriculture
ILRI	-	International Livestock Research Institute
IPGRI	-	International Plant Genetic Resource Institute
$MgCl_2$	-	Magnesium Chloride
NaCl	-	Sodium Chloride
PAGE	-	Polyacrlamide Gel Electrophoresis
PCR	-	Polymerase Chain Reaction
PIC	-	Polymorphic Information Content
QTL	-	Quantitative traits loci
RAPDs	-	Randomly Amplified Polymorphic DNA's
RFLP	-	Restriction Fragment Length Polymorphism
rfu	-	Relative Fluorescent Units
SDS	-	Sodium Dedocly Sulfate
SSCP	-	Single Strand Conformation Polymorphism
SSR's	-	Simple Sequence Repeats
STR's	-	Short Tandem Repeats
UPGMA	-	Unweighted pair group method with arithmetic averages algorithm
VNTR's	-	Variable number of tandem repeats

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background Information

Cassava (*Manihot esculenta crantz*) is an essential tropical root crop that belongs to Euphorbiaceae family containing 28 wild spp that evolved from inter-specific hybridization among wild species. It is a major source of carbohydrate for million of people in the tropics (FAO, 2006). Cassava is mostly produced by small scale farmers on marginal land in humid and subhumid tropics. The crop is efficient in carbohydrate production, thrives well on infertile soils and tolerates to adverse climatic conditions and low input management (FAO, 2006). In sub-Saharan Africa, the average consumption exceeds 300kg per person per year and about five hundred million people consume cassava daily, making it one of the most important food crops in sub-Saharan Africa (FAO, 2002) Cassava gives about 40% of carbohydrate production per hectare which is 61.53% higher than rice and 25% more than maize. As a result cassava is considered the cheapest source of calories for both human nutrition and animal feeding. In most cases the root tubers are used for making chips and flour. Both in developed and developing countries the crop is also utilized for industrial purposes, in preparing starch flakes and pearls and as a source of animal feed. Currently the role of cassava in Tanzania, Nigeria and many parts of Asia and Latin America has been changed from traditional human food to an efficient industrial crop (FAO, 2002).

In Tanzania the crop has gained economic importance through transforming it from a rural subsistence staple to a cash crop. This situation has led to rapid increase in production and marketing as a result it was been given a second priority in national ranking of crops in Tanzania (FAO, 2006).

1.1.1 Cassava production in Africa

Global cassava production in 2009 was forecasted at 242 million tones, 4% above the record of the previous year (FAO, 2009). According to FAO (1999), the rate of increase of cassava production has been higher than any other crop in Africa since the 1980s. It is now one of Africa's major starch staples second to maize, feeding 500 million people (FAO, 2002).

Table 1: The major cassava producing countries worldwide

Country	Production (tons)
Nigeria	45,000,000
Thailand	30,088,000
Brazil	26,000,000
Indonesia	20,500,000
Democratic Republic of Congo	15,036,000
Ghana	10,000,000
Mozambique	9,200,000
India	9,200,000
China	8,700,000
Tanzania	6,500,000
Uganda	4,500,000
Paraguay	5,400,000

Source: FAO (2009).

1.1.2 Cassava production in Tanzania

Tanzania is the fifth largest producer of cassava in Africa after Nigeria, Democratic Republic of Congo, Ghana and Mozambique, with an estimated total root production of about 6.5 million tones (FAO, 2009). The main cassava producing areas in Tanzania are the coastal strip along the Indian Ocean, around Lake Victoria, Lake Tanganyika and along the shores of Lake Malawi About 48.8% of total production is produced from coastal areas, 23.7% comes from around Lake Victoria and 13.7% from around Lake Nyasa in southern Highlands of Tanzania (Fig.1)

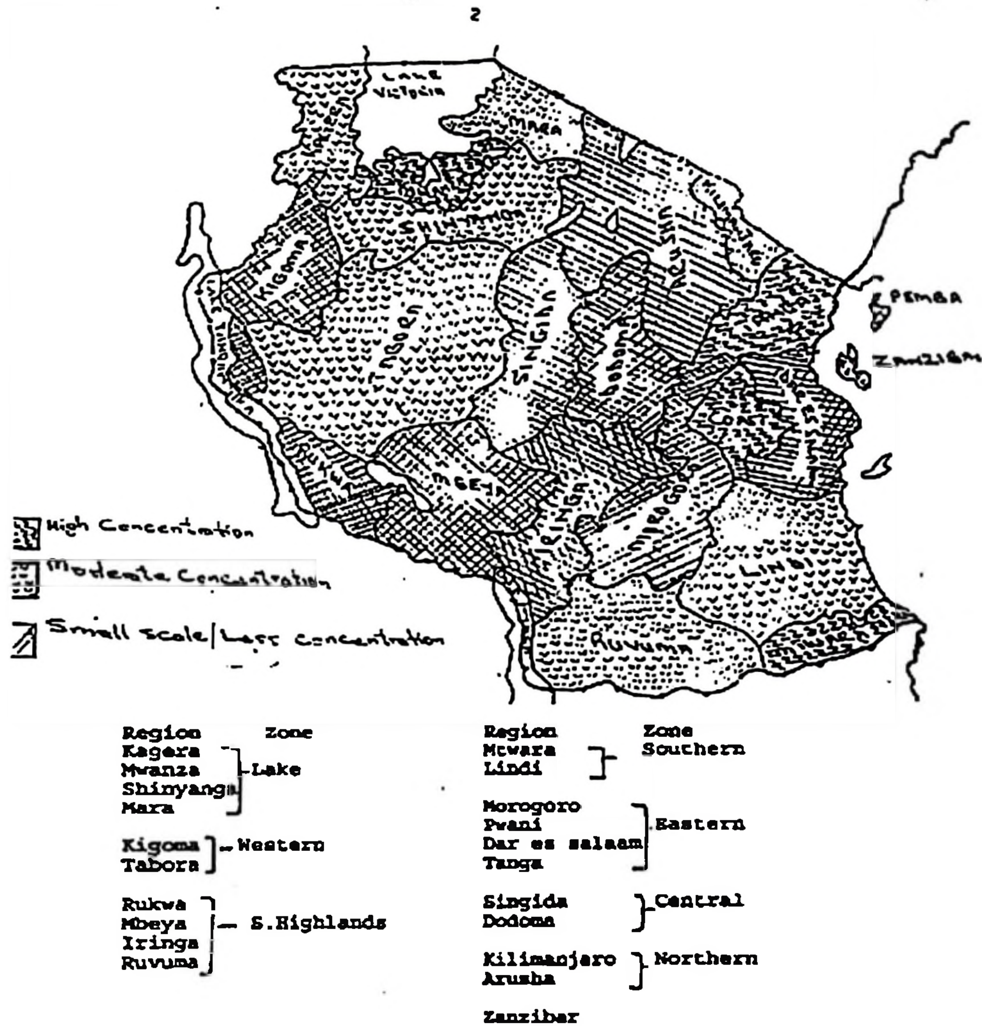


Figure 1: Cassava-producing areas of Tanzania.

Source: Tanzania Ministry of Agriculture, Food and Cooperatives, Planning and Marketing Division Dar-es-Salaam (1996).

1.1.3 Production constraints of cassava in Tanzania

In Tanzania, the low average yields of about 2 metric tons per hectare is caused by many factors including susceptibility of commonly grown varieties to major diseases and pests (Muhana and Mtunda, 2002). According to Thresh *et al.* (1997) eight diseases caused by viruses have been observed on cassava in Sub-Saharan Africa, such diseases are East African Cassava Mosaic Virus (EACMV), and the African Cassava Mosaic Virus (ACMV), cassava brown steak disease (CBSD), cassava bacterial blight (CBB), cassava green mite (CGM), cassava mealy bug (CM) and nematodes. Other constraints include Shortage of planting materials and continuous use of low genetic potential cassava varieties

1.1.3.1 Insect- Pests and Diseases

1.1.3.1.1 Cassava brown streak disease (CBSD)

CBSD is a viral disease, described by Storey (1936) who recorded it in the foothills of Usumbara Mountains in Tanzania. Nichols (1950) later reported that the disease was associated with chlorosis in roots and leaves, thus the disease is readily introduced into newly planted areas through the use of infected planting materials. Storey (1936) believed that the disease was caused by an insect-borne virus and that the most likely vector was a whitefly (*Bemisia spp.*) Research at Kibaha and Naliendele in Tanzania has revealed as high as 55% loss of root yield and quality due to CBSD (Muhana and Mtunda, 2002). The disease also now considered the greatest threat to food security of people who depend on cassava in countries like Uganda and Tanzania. A number of genotypes with high levels of tolerance to CBSD have been identified, such as Nase3 and MH96/2961 and these are currently being used in cassava breeding as parents to introgress genes for resistant/tolerance to cassava brown streak disease (Fregene *et al.*, 2003).

1.1.3.1.2 Cassava mosaic diseases (CMD)

CMD is widely distributed all over the country and is mostly transmitted through infected cutting (81%) and only 19% by whitefly vector (Thresh *et al.*, 1997). Surveys conducted throughout the major growing areas by COSCA showed that CMD was next to cassava green mite in spreading symptoms which were observed in about 70 percent of the villages (Thresh *et al.*, 1997). Recently, CMD incidences were observed along the coast of the Indian Ocean particularly in Mtwara and Lindi regions. The disease is more devastating because heavy attacks by CMD can result in high magnitude yield loss and storage root quality.

1.1.3.2 Shortage of planting materials and continuous use of cassava varieties with low genetic potential.

Lack of adequate and clean planting materials is one of major constraint to cassava production. Consequently, farmers plant any materials they come across. Most of the varieties grown by farmers have been selected mainly on their characteristic such as sweet taste. But most of such varieties have low genetic potential in term of yields/or resistance to the major insect pests and diseases (Fregen *et al.*, 2003).

1.1.4 Cassava breeding in Tanzania

Currentl breeding programme in Tanzania emphasizes on increase of cassava production to serve as the main food security in sub- Saharan countries. One of the best methods to achieve that is through the development of better varieties that are resistant to diseases, insect pests, drought, as well as high yielding (Moyib *et al.*, 2007). However, it is difficult to improve these traits in cassava using conventional methods because most of quantitative traits, are controlled by many alleles and influenced by environment (Muhana and Mtunda, 2002). Therefore molecular markers which detect variation at the DNA

level, provide a method of accurately characterizing germplasm more accurately as they are less influenced by environmental effects. The most prominent molecular marker systems include; Restriction Fragment Length Polymorphism (RFLPs) (Southern, 1975), Amplified Fragment Length Polymorphisms (AFLPs) (Vos *et al.*, 1995), Randomly Amplified Polymorphic DNA (RAPDs) (Williams *et al.*, 1990), and SSRs (Weber and May, 1989). Among the recently developed molecular markers in cassava, SSRs markers are being considered as the markers of choice as they detect the variation in allele frequency at many linked and unlinked loci, high level of heterozygosity and codominance nature (Wanger *et al.,* 1994). In cassava most available sequences are expressed sequence tags (ESTs). Here I report evaluation of effectiveness of ESSR marker for genetic diversity and identification of CBSD resistance.

1.2 Justification

Based on importance of cassava worldwide efforts are being made to develop varieties with high resistance to CBSD for increased production and quality in all coastal areas of Tanzania (Muhana and Mtunda, 2002). Success in this depends on the adequate amount of reliable marker system and identification of ESSRs that are linked to CBSD. These markers are expected to be used in different cassava research stations to enhance breeding programmes (Ferguson, 2004). Recently, through the Generation Challenge Programme (GCP) funded project, the International Institute of Tropical Agriculture (IITA) and Institute of Genomic Research (IGR) have generated new ESSR markers in cassava. These markers need to be validated for genetic studies and polymorphism with respect to CBSD from among different cassava varieties. Consequently, stable and high fidelity ESSR markers will be used for genetic studies and breeding for CBSD resistance and other traits in Tanzania.

1.3 Objectives

1.3.1 Overall objective

The overall objective of this study is to identify reliable new expressed simple sequence repeat markers (ESSRs) for genetic studies and identification of cassava brown streak disease resistance.

1.3.2 Specific objectives

The specific objectives of the study are

(i) To assess the optimum PCR condition on annealing T° and $MgCl_2$ concentrations

(ii) To identify polymorphic loci of ESSR markers for genetic diversity of cassava

(iii) To identify polymorphic loci of ESSR markers with respect to CBSD resistance

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Cassava: The Origin and Distribution of the Plant

2.1.1 Origin

Cassava is originally from South America, probably domesticated in the Amazon region (Olsen & Schaal, 2001). The process of cassava domestication involved selection for root size, growth habit, number of stems and the ability of clonal propagation through stem cuttings. About 98 species of Manihot are known, of which 28 wild species evolved through interspecific hybridization. Sexual barriers within the genus appear to be weak, indicating a recent evolution of the genus (Roa *et al.,* 1997).

2.1.2 Introduction and spread of cassava in Africa and particularly Tanzania

Cassava was first taken to Africa from Latin America in early 15th century by European traders as a potentially useful food crop. From 15th century cassava spread through Africa by a number of mechanisms. The most important appear to have been initial contacts with the Portuguese-Brazilian culture at the African coasts where the crop infiltrated the inland through riverine trade routes. Cassava arrived and diffused first along the West African coast, while it arrived at the East African coast during the 18th century (Fregene *et al.,* 2001). In Tanzania cassava first reached the shores of Lake Tanganyika from West Africa by the Congolese farmers, from there cassava moved to the inland Tanganyika through farmers to farmers diffusion (Cartel *et al.,* 1992). Both bitter and non-bitter varieties were grown at an early stage of introduction. Distinction between early maturing and late-maturing varieties was important and the latter were used in areas where famine was a frequent threat. Many, if not all, late-maturing varieties were bitter.

2.2 Identification of Polymorphic Loci for Genetic Diversity and Relatedness of Cassava

According to IPGRI and Cornell University genetic diversity is the product of interplay of biotic factors, physical environment, artificial and plant characters such as size, mating system, mutation, migration and dispersal which result into variation. The genetic variation is the results of selection, mutation, migration, genetic drift and /or recombination. All these phenomena influence genetic composition and genetic diversity in populations. Variation in allele frequency at many unlinked loci is the preferred way to assess genetic diversity and differentiation and to estimate the strengths of the various forces shaping the diversity (Wu and Tansley, 1993). In cassava, genetic diversity holds promise of providing ways around breeding obstacles of long growth cycle because selection can be made earlier in the growth cycle, even at the seedling stage.

Genetic markers have become fundamental tools for understanding the inheritance and diversity of natural variation. The earliest markers in cassava were morphological such as shape and root colour. The second generations of markers were biochemical, such as isozymes of which they provided a useful tool for genetic fingerprinting and diversity studies of cassava. Isozymes have been applied to characterizing relationships among accessions of African cassava germplasm and fingerprinting of the international cassava collection held at International Centre for Tropical Agriculture (CIAT).

The greatest genetic diversity for cassava exists in Latin America; although there is substantial diversity in Africa there are some factors that reduce diversity. Cassava being mainly a vegetative by propagate crop, its genetic diversity has been reduced by accumulation of systemic pathogens and viral diseases example CBSD this influence the diversity.

Currently molecular markers are recognized as significant tools to orient plant genetic resource conservation management, providing a means to accurately estimate the genetic diversity and genetic structure for a species of interest (Hamrick and Godt, 1997). Analysis of genetic diversity of African cassava was evaluated using restriction fragment length. Polymorphisms (RFLP) (Beeching *et al.*, 1993), random amplified polymorphic DNA (RAPD) markers (Marmey *et al.*, 1994) and simple sequence repeat markers (SSR) markers (Fregene *et al.*, 2003). The success of crop improvement program thus is highly reliant on the power and efficiency with which the genetic variability in the different populations can be manipulated. Genetic diversity analyzed by different markers such as (SSRs) markers was measured in terms of number of alleles per locus and Nei's unbiased estimate of gene diversity (H) also known as expected heterozygosity (He) (Nei, 1987) which is important in cassava breeding program.

However, RFLP, AFLP and SSR markers stand out as most effective in detecting polymorphism in cassava (Weising *et al.*, 1998). Given the large amount of DNA required for RFLP detection and the difficulties in automating RFLP analysis and the dominance of AFLP markers and the requirement of high quality DNA to ensure complete restriction (Weising *et al.*, 1998), SSRs are the markers of choice as they are highly polymorphic and almost found in every organism. Therefore ESSR, are more easily transferred across taxa and more likely linked to genes for traits of interest (Saha *et al.*, 2004). ESSR are short DNA molecules (300 - 500 bp) reverse-transcribed from a cellular mRNA population (MacIntosh *et al.*, 2001). They are generated by single-pass sequencing of randomly picked cDNA clones and have proven to be efficient and rapid means to identify novel genes (Adams *et al.*, 1991). ESSR thus represent informative source of partial expressed genes and provide a sequence resource that can be exploited for large-scale gene discovery (Whitefield *et al.*, 2002).

Currently the pattern of genetic diversity within large cassava collections at the International Centre for Tropical Agriculture (CIAT), Colombia and the International Institute for Tropical Agriculture (IITA), Nigeria is well characterized using SSRs markers but this does not necessarily reflect the extent of genetic structure of varieties under cultivating conditions.

2.3 Optimization PCR Condition on Annealing T° and Mgcl₂ Concentrations

Optimization is a balance between good amplification and nonspecific amplification. PCR is sensitive to a number of parameters including magnesium ion concentration, template DNA quality and concentration, primer concentration, and annealing temperature. These parameters need to be optimized to avoid nonspecific amplification products such as primer-dimers and fragments of heterogeneous size in cassava plants (Blanchard *et al.*, 1993). The annealing temperature of 62°C falls within the range of 55°C to 72°C that is generally recognized to produce best results in cassava (Innis and Gelfand, 1990).

Magnesium ions are an essential co-factor for DNA polymerase in PCR (Innis and Gelfand, 1990) and its concentration must be optimized for every primer system. Many components of the reaction bind magnesium ions, including primers, template, PCR products and dNTPs (Innis and Gelfand, 1990). Mg^{++} bind tightly to the phosphate sugar backbone of nucleotides and nucleic acids, and variation in the $MgCl_2$ concentration has strong effects on nucleic acid interactions (Blanchard *et al.*, 1993). It is necessary for free Mg^{++} to serve as an enzyme co-factor in PCR. In general, magnesium ion should be varied in a concentration series from 1.0–3.0 mM.

2.4 Uses of Markers in Cassava

2.4.1 QTL mapping

One of the great uses of DNA markers has been in the construction of genetic linkage maps. Linkage maps indicate the position and relative genetic distance between markers along chromosomes, which is analogous to signs or landmarks along a highway (Collard *et al.*, 2005). However the effective use of QTL will require, first, a validation of the markers through testing them in a larger population, in many environments and thereby achieving a more accurate performance regarding the stability of the QTL across environments (Fregene *et al.*, 2001). QTL is efficient used in mapping gene resistance to cassava bacterial blight (Jorge *et al.*, 2000, 2001). The potential use of the identified QTL in cassava will be in pyramiding the disease resistance genes together with those for agronomic traits in cassava genotype.

2.4.2 Marker assisted selection (MAS)

Marker assisted selection (MAS) is the use of DNA markers for selection of desired individuals within a population, by linking the molecular marker to a desired trait (Williams *et al.*, 1990). Marker-assisted selection enables the breeder to eliminate inferior or superior genotypes at an early stage and increases the efficiency of selection by allowing the breeder to concentrate on fewer genotypes. In cassava, application of MAS has been developed more recently compared to other major crops. MAS for breeding CMD resistance has successfully been applied for introducing resistance into elite gene pools at CIAT (Fregene & Mba, 2004) and also to introgress resistance to cassava green mite (CGM) and CMD in local Tanzanian varieties (Kullaya *et al.*, 2004).

2.5 Useful Molecular Markers in Cassava

2.5.1 Restriction Fragment Length Polymorphism (RFLP)

RFLP technique is a non- PCR based method. It was the first technique to be developed and enabled the detection of polymorphism at the sequence level (Southern, 1975). Beeching *et al.* (1993) used RFLP to assess the genetic diversity within a collection of cassava germplasm. RFLPs have been applied in studies of phylogenetic relationships of species within the genus *Manihot* and also in tagging resistance genes for diseases such as cassava green mites. RFLP analysis is co-dominant, thus enabling identification of heterozygotes from homozygotes. It is also reproducible within and between laboratories. Limitations of RFLP analysis include usage of probes, high cost, and impossibility of automation.

2.5.2 Randomly Amplified Polymorphic DNA (RAPDs)

The PCR based-RAPD technique has been used to provide means for genetic analysis and identifying molecular marker for resistance to cassava anthracnose disease CAD. Candidate genetic markers that can distinguish between resistant and susceptible plants have been identified in the previous study by the method of bulked segregate analysis. Where 80 -100% polymorphism observed in the arbitrary primers suggested the potential use of RAPD markers in cassava (Akano *et al.*, 2002). Relative position of all the RAPD markers linked to the putative anthracnose resistance locus was identified in a segregating population derived from a cross TMS 92/0326 x TME117.

These results demonstrate that RAPD analysis can still be a suitable tool for detection and mapping of major genes for marker-assisted selection in cassava breeding. Despite these merits, RAPDs are dominant markers, and cannot be distinguished from homozygotes the heterozygote. The technique causes incidence of non-inherited bands that are probably

PCR artefacts and it is not reproducible. This limits utilization of RAPDs in cassava genotyping (Welsh and McClelland, 1990).

2.5.3 Amplified Fragment Length Polymorphisms (AFLPs)

Amplified fragment length polymorphism (AFLP) is a reliable, PCR-based marker system for obtaining quantitative estimates of genetic relationships (Vos *et al.*, 1995). Ninety-three varieties of *M. esculenta* (Crantz) were assessed by AFLPs for genetic diversity and for resistance to CBB (Lozano and Sequeira 1974). AFLP analysis was performed using two primer combinations and a 79.2% level of polymorphism was found. The phenogram obtained showed between 74% and 96% genetic similarity among all cassava accessions analyzed (Jorge *et al.*, 2000, 2001).

The results demonstrate that resistance to cassava bacterial bright CBB is broadly distributed in cassava germplasm and that AFLP analysis is an effective and efficient means of providing quantitative estimates of genetic similarities among cassava accessions. Since polymorphism is detected as the presence or absence of amplified restriction fragments, AFLPs are usually considered dominant markers which limit the use of AFLP.

2.5.4 Microsatellites (Simple Sequence Repeats) (SSRs)

Microsatellites or simple sequence repeats (SSR) are the markers of choice for molecular genetic mapping and marker-assisted selection in cassava. Simple sequence repeat (SSR) markers are powerful tools for assessing genetic diversity, studying population genetics and for marker applications in plant breeding because of their ubiquity, codominant behaviour, reproducibility, amenability to high throughput visualization and high level of polymorphism (Milbourne *et al.*, 1997; Witsenboer *et al.*, 1997). One drawback is that in

general SSRs need to be identified and flanking primers developed specifically at each target locus within each species, although some SSR primers are transferable across species and even genera (Wang *et al.*, 2005). In cassava, the discovery, inheritance and variability of 14 GA SSRs were described by Chavariaga *et al.* (1998). Later Mba *et al.* (2001) reported 172 largely dinucleotide SSR markers derived from microsatellite-enriched genomic libraries. Thirty-six of these were placed on the genetic linkage map of cassava. Roa *et al.* (2000) found ten SSR primer pairs to be generally compatible across seven species within the genus *Manihot.*

2.5.5 Expressed sequence tags

ESTs are short DNA molecules (300 - 500 bp) reverse-transcribed from a cellular mRNA population (MacIntosh *et al.,* 2001). They are generated by single-pass sequencing of randomly picked cDNA clones and have proven to be efficient and rapid means to identify novel genes (Adams *et al.*, 1991).

The identification of ESSR through mining EST databases has become a fast, efficient and low-cost option for SSR discovery in many plant species (Han *et al.*, 2004; Thiel *et al.*, 2003). ESSRs are more easily transferred across taxa and more likely linked to genes for traits of interest (Saha *et al.*, 2004). In cassava most publically available sequences are either expressed sequence tags (ESTs) or full-length cDNAs. ESTs thus represent informative source of partial expressed genes and provide a sequence resource that can be exploited for large-scale gene discovery (Whitefield *et al.*, 2002). Recently, a large number of cassava ESTs has been sequenced. Results of some of these studies have been published (Anderson *et al.*, 2004; Lopez *et al.*, 2004).

2.6 Identification of Polymorphic Loci of ESSR Markers linked to CBSD resistance

The effectiveness of selection may be more limited by the reliability of the screening method for the trait than by any other factor. Thus the breeding of a new variety can take between eight and twelve years and even then the release and adoption of an improved variety cannot be guaranteed. This study provide eight 8 ESSR markers for identifying resistance to CBSD. Most important among the biotic factors affecting cassava production are viral diseases, specifically cassava brown streak disease (CBSD), which has long been recognized as a major limiting factor to cassava production in Africa. Polymorphic loci of ESSR linked to CBSD resistance are important due to the fact that DNA markers can be used to identify genotype with resistance to CBSD earlier than the time such traits may be assessed phenotypically. Effort has been made to use ESSR in both polymorphic characterization and linking to CBSD by IITA - Roots and Tubers Programme.

The IITA branch in Dar es Salaam and its partners namely Agricultural Research Institute (ARI), Tanzania, and the National Agricultural Research Organization (NARO), Uganda, have devised a concerted program supported by Bill & Melinda Gates Foundation to identify and use molecular markers for faster and more accurate breeding of cassava varieties resistant to Cassava Brown Streak Disease (CBSD). In identification for resistance to CBSD, accurate sizing of ESSRs allele is crucial and the detection system needs to be able to differentiate a 1bp difference (Buchanan *et al.*, 1994).

2.7 Suitability of ESSR Markers for Genetic Diversity and Resistance to CBSD

PIC is the measure of the usefulness of each marker in distinguishing one individual from another. PIC value is influenced by the number and frequency of alleles and it is a valuable indicator of fidelity of marker for polymorphism (Liu *et al.*, 2000). A high PIC value indicates that the marker is highly informative and can distinguish closely

related genotypes. PIC values are affected by the number and frequency of alleles in the population under study. Buchanan *et al.* (1994) reported that loci with a large number of different alleles may have high PIC values, but if one or two alleles dominate, then the PIC value may still be relatively small. Up to date IITA and ARI have identified a few varieties with some level of resistance to the disease and seek to identify the DNA markers associated with the resistance gene in these varieties and integrate marker assisted selection into cassava breeding programmes.

CHAPTER THREE

3.0 METHODOLOGY

3.1 Planting Materials

The plant materials used in this study consisted of 24 different cassava genotypes obtained from different countries. Of the 24 genotypes, 9 were from South America (Brazil, Colombia, Peru, Cuba, Mexico and Venezuela), 9 from Tanzania comprising resistant and susceptible genotypes to CBSD and 6 from Uganda, Ghana, Togo and Liberia (Table 2).

Table 2: List of 24 cassava genotypes collected from different countries

S/No	DNA genotype	Country
1	TME - 539	Uganda
2	TME - 1389	Ghana
3	TME - 1368	Ghana
4	TME - 396	Togo
5	TME - 230	Togo
6	TME - 589	Liberia
7	BRA - 990	Brazil
8	BRA – 255	Brazil
9	Col - 297	Colombia
10	PER - 458	Peru
11	CUB 1	Cuba
12	MEX - 55	Mexico
13	VEN - 77	Venezuela
14	MCOL - 1468	Colombia
15	MCOL - 1734	Colombia
16	Namikonga (resistance to CBSD)	Tanzania
17	Albert (susceptible to CBSD)	Tanzania
18	Pesangani	Tanzania
19	Kigoma	Tanzania
20	Mwaya	Tanzania
21	Faraja	Tanzania
22	Mkita	Tanzania
23	Kalolo	Tanzania
24	Kibaha	Tanzania

3.2 Genomic DNA Extraction

Genomic DNA was extracted following the modified method of Dellaporta (Dellaporta *et al.*, 1983). Leaf tissue of 0.5g was ground to fine powder in liquid nitrogen using mortar and pestle. The powdered tissue was transferred to a frozen 1.5ml eppendorf tube containing 800µl of extraction buffer (100mM Tris-HCl, 50mM EDTA, 500mM NaCl, 1% PVP, 700µl of B-mercaptoethanol in one litre of extraction buffer) incubated at 65°C. Thereafter, 50µl of 20% SDS was added to the mixture sample and put on a water bath at 65°C with intermittent mixing for 30min. After incubation, the tubes were cooled at room temperature (25-27°C) and then 250µl of ice-cold 5M potassium acetate added with gentle mixing by inverting the tubes after which they were incubated on ice for 20 minutes.

The solution was centrifuged at 12 000rpm for ten minutes in eppendorf table centrifuge. The supernatant was transferred to a fresh 1.5ml eppendorf tube and one volume of ice-cold isopropanol added with gentle mixing followed by incubation at -80°C for one hour before centrifugation at 12 000rpm. The supernatant was poured off and last drops of isopropanol removed by placing the tubes facedown on paper towels. The pellet was re-suspended in 500µl of TE buffer (50mM Tris-HCl /10mM EDTA and pH 8.0). .

The pellet was re-extracted and precipitated by addition of one volume of ice-cold isopropanol followed by incubation and centrifugation. One hundred microlitres of 10mM Tris-HCl, 1mM EDTA containing 10mg/ml RNase was added and stored at 4°C overnight to dissolve the pellet. The dissolved DNA was transferred to fresh appendorf tube and stored at -20°C until used for PCR. DNA quality and quantity was determined using spectrophotometer. One microlitre of DNA samples was loaded into a nano drop spectrophotometer, which had been blanked with triple distilled water. Absorbance measurements were made at 260nm (A260), and 280nm (A280). The DNA concentration

was calculated automatically; the ratios A260/A280, which is an estimate of the DNA purity, were also computed.

3.3 PCR Optimization

The optimal annealing temperature for each primer pair was determined using a standard PCR protocol and gradient PCR between the temperatures of 55.8°C and 68.1°C were most of ESSR amplified in broad range (62.7- 68) (Appendix 2). Standard PCR reaction condition were set using 50 ng DNA; 0.8 pmole (Forward and Reverse) primer; 0.375U *Taq* polymerase; 1.5 mM $MgCl_2$; 0.2 mM dNTPs, 1x PCR buffer (50 mM KCl; 10 mM Tris-HCl (pH 8.3). The reaction was done in volume of 10 µl and the PCR profile consisted of 5 min initial template denaturation at 95°C, 35 cycles of 30 seconds at 95°C, 2 min at primer annealing temperature and 30 seconds at 72°C.

This was followed by a final primer extension of 72°C for 5 minutes. To optimize the magnesium chloride concentration all primers were subjected to the above PCR conditions using one DNA sample, but with a varying range of $MgCl_2$ concentrations (1.0 mM, 1.5 mM, 2.0 mM, 2.5 mM and 3.0 mM) which shows a clear peak Appendix 4. The reaction conditions that gave a clear single peak with relative fluorescent units (rfu) of above 2000 and minimum of stutter peaks was selected as the optimal PCR condition for that particular ESSR primer. In a similar way the amount of fluorescent tail to be added for analysis on the capillary sequencer was optimised across three concentrations: 0.125, 0.150 and 0.175 pmole/µl. The optimal PCR conditions were performed using capillary electrophoresis for high throughput genotyping. The results are summarized in Appendix 2.

3.4 Suitability of ESSR Primers for Genetic Diversity

Selection of ESSR primers was based on peak quality detected by ABI 3730 sequencer. ESSR primers that gave strong clear single peaks with minimum stray peaks were initially selected and subsequently subjected for polymorphism test. Polymorphic primers were identified by screening all the optimized ESSR primers against 24 cassava genotypes. Amplification was carried out using the optimized PCR conditions and fragment analysis performed on the ABI 3730 (Applied Biosystems). The peaks were sized and the alleles obtained using the Genemapper software (Applied Biosytems). Genetic diversity was measured in terms of number of alleles per locus and Nei's unbiased estimate of gene diversity (H) also known as expected Heterozygosity (He) (Nei, 1987). Distance matrices were calculated on the basis of Euclidean distance among genotypes and geographic origins. The distance matrices were subjected to UPGMA (Unpaired Group Method of Arithmetic Averages) and clustering was done resulting in a construction of dendrogram using Darwin software. Relative validation was performed to validate the dendrogram. Polymorphic Information Content (PIC) for the polymorphic ESSRs was calculated according to Nei (1987) as

$$PICi = 1-\sum_{j=1}^{n} (Pij)^2 \text{ where}$$

PICi = polymorphic information content of a marker i

Pij = frequency of the jth pattern for marker i and the summation extends over n patterns. Distance matrices were calculated on the basis of modification of Rogers' genetic distance by (Botstein *et al.* 1980).

3.5 Identification of ESSR Loci for Resistance to Cassava Brown Streak Diseases

A total of 14 ESSR were used for identification of CBSD loci. The amplified ESSR's were co-loaded based on their dye label, fragment size and the relative fluorescence unit (rfu) on the ABI-3730. The pooled PCR products for each co-loaded group were centrifuged and 1.0µl drawn from the mix and added to 9.0µl of HIDI ® formamide-LIZ ® 500 standard mix. This cocktail was denatured at 95°C for three minutes and quickly chilled on ice before loading on ABI-3730. Samples were introduced to the 48 capillaries by electrokinetic inoculation migrates according to size, by electrophoresis through POP4 (performance optimized polymer) matrix containing urea, along a 36cm capillary array. A run module which consists of 60°C run temperature, 10 seconds injection time, 15KV run voltage and 20 min run time, were used for all samples. As the DNA fragments passed through the detection cell, the fluorescent dyes were excited by an argon laser and emission spectra of the various dyes were detected by charge coupled device (CCD) camera and the captured electrons were represented as relative fluorescence units (rfu). Each electrophoresis run consisted of 48 samples that were simultaneously analyzed. After the fragments were separated using ABI 3730 capillary electrophoresis machine (Applied Biosystems), the peaks were sized and the alleles obtained through Genemapper software (Applied Biosystems).

The results of two varieties of Namikonga (resistant) and Albert (susceptible) were confirmed by observation of electropherograms of sample plot after analysis using Genemapper software. The allele intensities were fairy good ranging between 4000 and 30 000. ESSR primers that gave a different allelic richness between Namkonga and Albert were initially selected and subsequently subjected for polymorphism test.

CHAPTER FOUR

4.0 RESULTS AND DISCUSSION

4.1 Optimization of PCR Condition

An annealing temperature of 62°C was found to be optimum for all primer pairs except for ESSR 72 and ESSR 110 in which 57°C was found to be optimum. 0.175 pmole/µl fluorescent nucleotides were found to be appropriate for all reactions. Optimum magnesium chloride concentrations varied and are provided in Appendix 2. Of the 80 markers selected for optimisation 68 (85%) was successfully amplified while 12 (15%) of markers failed to amplify.

4.2 Identification of Polymorphic Loci for Genetic Diversity and Relatedness of Cassava

During this study 68 ESSR primer pairs, which were found to amplify well were screened across 24 diverse genotypes. Of these fifteen (15) markers were monomorphic as presented in Table 5 ESSR73, ESSR 80, ESSR 81, ESSR 83, ESSR 84, ESSR 85, ESSR 93, ESSR 108, ESSR 109, ESSR 119, ESSR123, ESSR 125, ESSR134, ESSR 138 and ESSR 139. Fifty three primers were polymorphic Appendix 3. On the other hand a total of 231 alleles were amplified by 53 ESSR, the number of alleles per primer ranged from 2 to 11 Appendix 3. The mean number of alleles per primer was 4.35. The PIC values ranged from 0.174 for ESSR 104 to as high as 0.844 for ESSR 91.

4.3 Genetic Relatedness of Cassava Genotypes

Four distinct groups of phylogenetically related cassava genotypes were identified in this study. Phylogenetic relationship demonstrated by UPGMA, genetic distance grouping indicated that cassava genotypes from the Rest of Africa and those from South America fall in the same group presented by cluster 1 (Fig. 2). In this group/ cluster, four (4)

genotypes from Rest of Africa namely TME 230, TME 1383, TME 1368 and TME 589 are phylogenetically identical to BRA 990 and BRA 255 from S. America. The second group distancing from genotypes in cluster one are Mkita, Kalolo, Mwaya from Tanzania and TME 539, TME396 from the Rest of Africa at a distance of 0.5116.

Also the UPGMA revealed that cassava genotypes from Tanzania and those from South America fall in the same group presented by cluster 3 (Fig. 2). In this group three 3 genotypes from South America namely COL 297, MCOL1734, MCOL 1488 are phylogenetically identical to Kibaha from Tanzania. The second group distancing from other genotypes from cluster three 3 are CuB1, Venezuela from South America and Namikonga from Tanzania at a distance of 0.5170. UPGMA revealed that cassava genotypes from Tanzania and those from South America fall in the same group presented by cluster four. In this group four 4 genotypes from Tanzania namely Pesangani, Kigoma, Albert, and Faraja, are phylogenetically identical to PERU 458 and MEX 55 from South America at a distance of 0.5677.

Analysis of genetic distances between the genotypes (Table 3) showed that the varieties most closely related were Peru 458 and Mexico 55 (D=0.1732), TME 1389 and TME 1368 (D=0.1982), Mkita and Kalolo (D=0.3981) followed by Pesangani and Kigoma (0.4912) while distantly related genotypes were BRA 990 and Mex 55 (0.7089), Albert and Mkita (0.6036), BRA255 and CuB1 (0.6054), TME 230 and TME 539 (D=0.5914), Namikonga and Albert (0.5139).

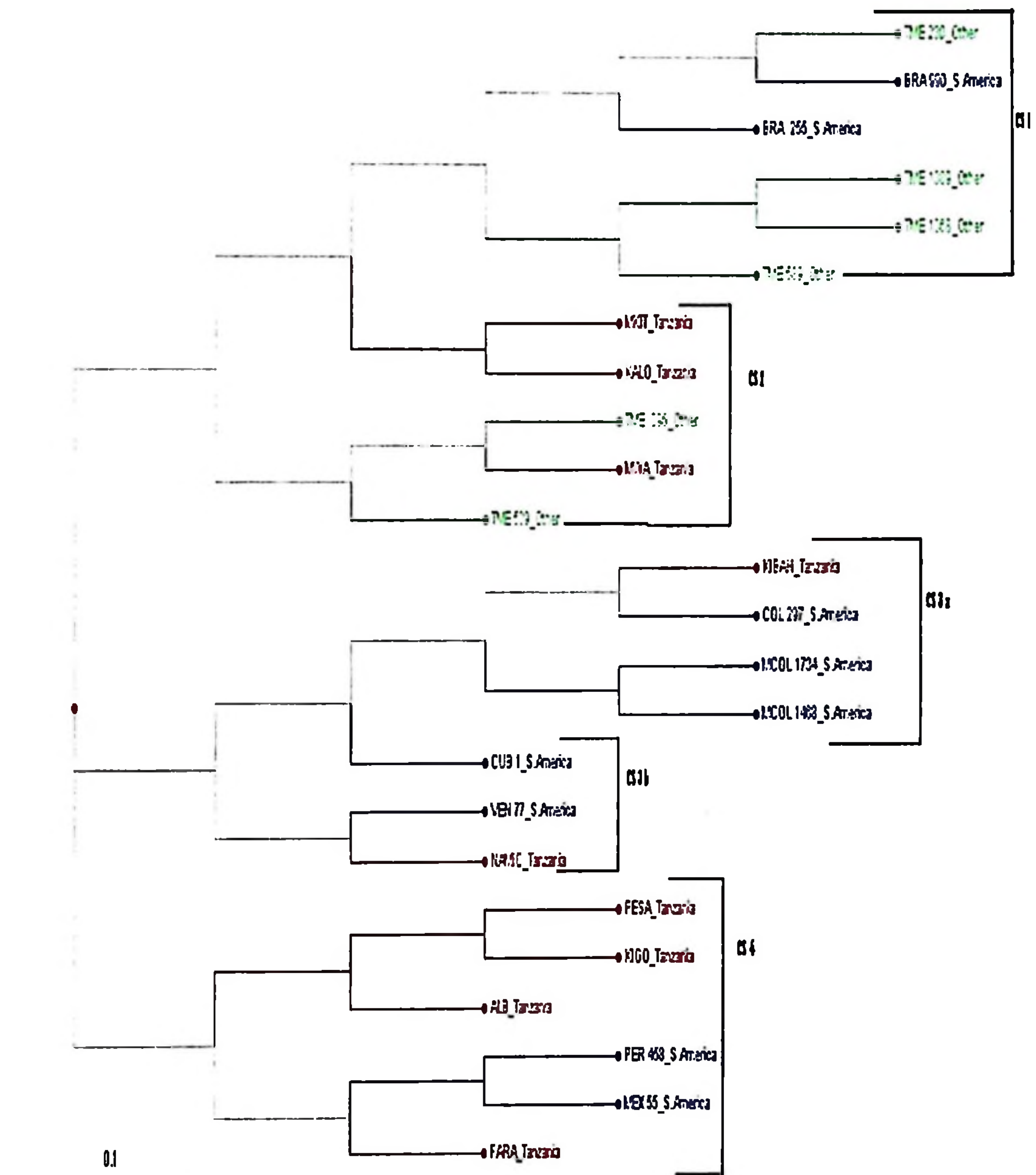


Figure 2: **Phylogenetic relationships among cassava genotypes on the basis of Wright's (1978) modification of Rogers' genetic distance using UPGMA clustering. 95% Bootstrap value was performed to test the robustness of the dendrogram topology.**

Key: CS1= 1st Cluster, CS2 = 2nd Cluster, CS3a, b = 3rd a= main cluster and b= sub cluster and CS4 = 4th Cluster.

Table 3: Genetic distances between genotypes as revealed by markers in this genetic diversity array

		ALB	BRA 255	BRA 990	COL 297	CUB I	FARA	KALO	KIBAH	KIGO	MCOL 1468	MCOL 1734	MEX 55	MKIT	MWA	NAMIC	PER 458	PESA	TME 396
1	ALB																		
2	BRA 255	0.6212																	
3	BRA 990	0.6629	0.5847																
4	COL 297	0.5653	0.5436	0.6205															
5	CUB I	0.5361	0.6054	0.5783	0.4361														
6	FARA	0.5893	0.5501	0.6131	0.5614	0.5271													
7	KALO	0.6495	0.5587	0.6593	0.5160	0.5230	0.5523												
8	KIBAH	0.5609	0.5914	0.6155	0.4417	0.5304	0.6166	0.6742											
9	KIGO	0.5260	0.6855	0.7038	0.6004	0.6449	0.5984	0.5941	0.5907										
10	MCOL 1468	0.5948	0.5095	0.6122	0.4261	0.4272	0.5737	0.5663	0.4263	0.6569									
11	MCOL 1734	0.5966	0.5333	0.5777	0.4697	0.4480	0.6064	0.5770	0.5062	0.6500	0.2218								
12	MEX 55	0.5980	0.6499	0.6374	0.5336	0.5602	0.4864	0.6870	0.5958	0.7096	0.4968	0.5563							
13	MKIT	0.6036	0.5012	0.6503	0.5849	0.5237	0.5643	0.3981	0.6345	0.6261	0.6105	0.6421	0.6174						
14	MWA	0.5889	0.6402	0.6664	0.5776	0.5314	0.5646	0.5936	0.5658	0.5307	0.4663	0.5143	0.6189	0.5778					
15	NAMIC	0.5139	0.5749	0.6742	0.4739	0.4894	0.5370	0.5233	0.5170	0.5444	0.5496	0.5999	0.5054	0.5005	0.5658				
16	PER 458	0.5251	0.6026	0.6170	0.5557	0.5460	0.5442	0.7086	0.5684	0.6667	0.5026	0.5807	0.1732	0.6333	0.6212	0.4759			
17	PESA	0.6374	0.5302	0.6730	0.5871	0.5843	0.5677	0.6345	0.6089	0.4912	0.5789	0.6143	0.6560	0.5616	0.5796	0.5894	0.6645		
18	TME 396	0.5864	0.6264	0.5854	0.6408	0.5965	0.6110	0.6446	0.6306	0.6741	0.5342	0.5336	0.5827	0.6291	0.4943	0.5587	0.6086	0.6446	
19	TME 1368	0.6662	0.5498	0.6723	0.5757	0.5464	0.6585	0.6112	0.6979	0.7115	0.5536	0.5698	0.6136	0.6244	0.7130	0.6373	0.6586	0.6544	[illegible]
20	TME 1389	0.6645	0.5215	0.6730	0.6222	0.5346	0.6255	0.5533	0.6884	0.6673	0.5485	0.5921	0.5745	0.5952	0.6605	0.6081	0.6419	0.6270	[illegible]
21	TME 230	0.6585	0.4325	0.5022	0.6033	0.6119	0.5723	0.5762	0.6457	0.7582	0.5519	0.5605	0.6858	0.6353	0.6457	0.6509	0.7109	0.6483	0.5415
22	TME 539	0.6669	0.5316	0.6700	0.5549	0.6220	0.6119	0.5354	0.6002	0.6558	0.4722	0.4895	0.5778	0.6159	0.4777	0.5561	0.6367	0.5854	[illegible]
23	TME 589	0.5881	0.5299	0.7293	0.5562	0.5957	0.6100	0.5389	0.7176	0.7032	0.5570	0.5577	0.6838	0.5987	0.6945	0.6614	0.7354	0.6618	[illegible]
24	VEN 77	0.5369	0.5791	0.6822	0.4038	0.3914	0.5474	0.4349	0.4705	0.6025	0.3412	0.4362	0.4419	0.5658	0.4556	0.3357	0.4965	0.5609	0.5548

BRA – 990=Brazil, BRA – 255=Brazil, Col – 297=Colombia, PER – 458=Peru, CUB 1=Cuba, MEX – 55=Mexico, VEN – 77=Venezuela, MCOL – 1468=Colombia, MCOL – 1734=Colombia, Mkit= Mkita, Kalo=Kalolo, Mwa= Mwaya, Kibah= Kibaha, Namic= Namikonga, Pesa=Pesangani, Kigo=Kigoma, Alb= Albet, Fara=Faraja, TME 539, TME 1389, TME1368, TME 396, TME 230, TME 589

4.4 Identification of ESSR Loci linked to Resistance to Cassava Brown Streak Disease

The results reveal a set of 14 ESSR markers which are able to discriminate resistant and susceptible individuals to brown streak diseases. The size of genotype plot that was produced in the ABI 3730 electropherogram indicates that alleles ranged in size from 100 to 3000bp with the majority being between 400 and 2000bp. The number of alleles detected by the different markers is shown in Table 4.

Table 4: Allelic diversity of ESSRs loci differentiating Namikonga (resistant) and Albert (susceptible) to CBSD

Markers	Observed Allele size for CBSD resistant genotype	Observed Allele for susceptible
ESSR105	182, 188	198
ESSR110	183, 192	187
ESSR115	117	109,121
ESSR121	132,138	181
ESSR130	102,105	140
ESSR136	202,208	190
ESSR71	171,177	209
ESSR75	186,189	166
ESSR76	207, 210	191
ESSR88	123,129	131
ESSR91	180,186	183
ESSR95	111,117	186
ESSR96	109,112	216
ESSR116	191,182	175
Total	27	15
Mean	1.93	1.07

Allelic difference from ESSR's loci confirmed by observation of electrophelogram of sample plot after analysis using Genemapper software. ESSR markers detected two alleles in resistant genotype Namikonga (heterozygotes) (Fig. 3) and one allele in susceptible genotypes (homozygote) (Fig. 4). This indicates the variation between the two varieties.

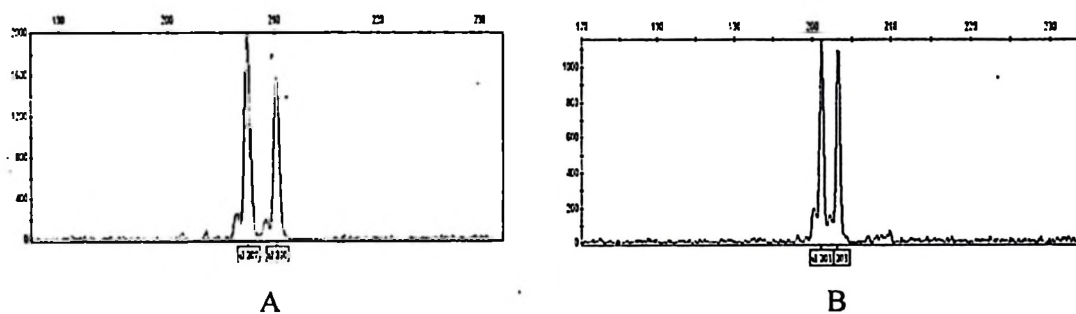


Figure 3: Namikonga (resistance genotype) plot of marker ESSR 76 showing heterozygote individuals.

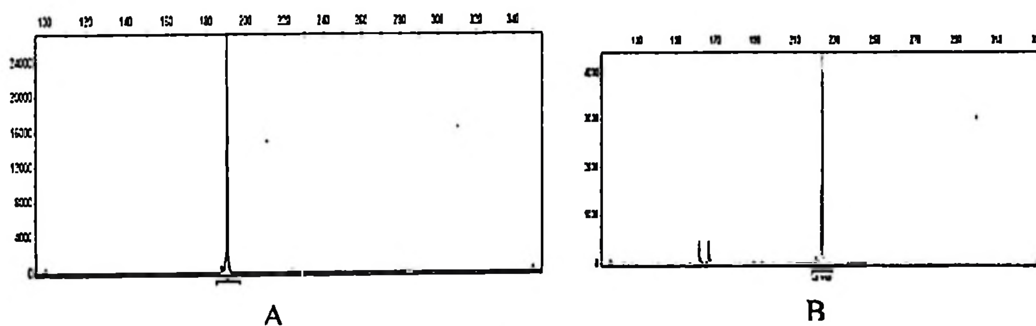


Figure 4: Albert (susceptible genotype) plot of marker ESSR 76 showing homozygote individuals.

The number of allele per locus among 24 cassava genotype were detected, a total of 63 alleles were detected in this study (Table 5). The number of alleles per primer ranged between 2 and 8, and the mean number of alleles per primer was 4.5. The PIC values ranged from 0.239 to 0.844. Eight 8 markers namely ESSR91, ESSR116, ESSR96, ESSR130, ESSR95, ESSR75, ESSR121 and ESSR110 had high PIC values, implying their suitability for CBSD studies and genetic characterization for further research to find gene that link to CBSD disease resistance.

Table 5: Number of alleles, allele size, diversity, degree of availability and PIC of ESSR markers observed in genotypes of cassava with respect to CBSD resistance

Marker	Genotype Number	Allele Number	Observe alleles bp	Gene Diversity	Heterozygosity	Availaability	PIC
ESSR71	2	2	210, 181	0.2778	0.3333	1.0000	0.2392
ESSR76	5	3	180, 192, 154	0.5720	0.9167	1.0000	0.4783
ESSR110	6	3	182, 187, 191	0.6016	0.5833	1.0000	0.5344
ESSR115	5	3	188, 194, 196	0.5512	0.6250	1.0000	0.4657
ESSR136	4	3	201, 204, 168	0.3793	0.3750	1.0000	0.3475
ESSR75	6	4	201, 203, 205, 177	0.6146	0.6667	1.0000	0.5568
ESSR88	4	4	198, 201, 204, 191	0.3568	0.3333	1.0000	0.3360
ESSR105	6	5	182, 188, 191, 194, 198	0.5217	0.4583	1.0000	0.4528
ESSR130	7	5	174, 177, 183, 192, 192	0.6649	0.6667	1.0000	0.5998
ESSR121	7	5	199, 205, 208, 209, 175	0.6345	0.6667	1.0000	0.5733
ESSR95	7	5	204, 206, 213, 215, 161	0.6510	0.6667	1.0000	0.5976
ESSR116	9	6	186, 190, 192, 196, 202, 209	0.6832	0.7500	1.0000	0.6344
ESSR96	10	7	166, 170, 173, 176, 176, 178, 204	0.6467	0.6667	1.0000	0.6224
ESSR91	15	8	192, 196, 198, 202, 204, 206, 207, 163	0.8602	0.9167	1.0000	0.8442
Mean	6.642	4.5		0.576446	0.616079	1.0000	0.520171
Total allele			63				

4.5 Discussion

PCR is sensitive to a number of parameters including magnesium ion concentration, template DNA quality and concentration, primer concentration, and annealing temperature. These parameters need to be optimized to avoid nonspecific amplification products such as primer-dimers and fragments of heterogeneous size. Therefore optimization is important for accurate and robust genotyping, particularly when amplification products are to be post-PCR co-loaded onto capillary sequencers to reduce unit costs (Beaulieu *et al.*, 2001). One of the specific objectives of the study was to determine the optimum annealing temperature and $MgCl_2$ concentration. This study revealed an optimum annealing temperature of 62°C. This temperature has proved to be optimum and gave reproducible results during the evaluation of total of 80 ESSR, using 24 cassava genotypes from South America, Tanzania and other African countries The results agree with the suggestions made by Innis and Gelfand, (1990) that annealing temperature of 62°C falls within the range of 55°C to 72°C that is generally recognized to produce best results.

Magnesium ions are an essential co-factor for DNA polymerase in PCR. The study revealed that $MgCl_2$ concentration varied from 1.0–3.0 mM (that means it shows a clear peak when $MgCl_2$ concentration is at 1.0, 1.5, 2.0, 2.5 and 3.0). The optimum $MgCl_2$ concentration in the reaction mix was 2.0 mM for all primers. Variations in $MgCl_2$ concentration below 4 mM affect primer specificity in PCR (higher concentrations lower the specificity, lower concentrations raise primer specificity). In general, magnesium ion should be varied in a concentration should range from 1.0–3.0 mM. In this work the optimum $MgCl_2$ concentration was 2.0Mm and therefore recommended for ESSR amplification studies in cassava.

The second specific objective of the study was to identify polymorphic loci of ESSR markers for determining genetic diversity among 24 cassava genotypes. The average PIC value for all 53 ESSR marker loci were fairly high this indicates that the markers that had high number of alleles had generally high PIC values indicating that some markers are more useful for differentiating between closely related genotypes than others. Buchanan *et al.* (1994) indicated that loci with a large number of different alleles may have high PIC values, but if one or two alleles dominate, then the PIC value may still be relatively small. The effect of allele frequency can be demonstrated by the marker with the lowest allele frequency of the dominating allele eg ESSR91 (191bp) (18.7%) has the highest PIC (0.844), while loci ESSR104 with the highest allele frequency (90%) for the predominating allele (190bp), has the lowest PIC value (0.173) Appendix 5. PIC values found in this study ranged from (0.1736) to (0.844). Therefore PIC value which is influenced by the number and frequency of alleles is a valuable indicator of fidelity of marker for identifying polymorphism with respect to CBSD resistance and for genetic studies in general (Liu *et al.*, 2000). High PIC and large number of allele will enable the breeders to determine whether the desired genes of the disease resistance have been successfully transferred from the parents to the offspring at the seedling stage.

Twenty nine (29) ESSR markers were highly informative in genetic studies. The highest PIC value of 0.8442 was observed for the marker ESSR 91. This marker also produced the highest number of alleles (8 alleles) suggesting that it could be used to distinguish closely related cassava genotypes. On the other hand 24 ESSR markers gave low PIC value and low allele frequencies. The lowest PIC value of 0.1736 was observed for the marker ESSR 104. These results agree with the findings from Buchanan *et al.* (1994) that loci with a large number of different alleles may have high PIC values, but if one or two alleles

dominate, then the PIC value may still be relatively small because of low genetic diversity. Analysis of genetic distance among 24 genotypes revealed that Tanzania and South America genotypes are well spread across the dendrogram. This indicates a high genetic diversity due to high gene flow between genotypes from Tanzania and South America. This indicates that there has been extensive germplasm was efficient exchange of cassava varieties between Tanzania and South America as well as extensive distribution of varieties in Amani breeding program from the 1920s to the 1960s throughout Tanzania. The 53 ESSR markers reported here are therefore appropriate for further applications in cassava germplasm characterization and plant breeding. Also dendrogram revealed that, the genotypes from the other countries in Africa appear only in 1st and 2nd cluster, indicating a narrow genetic diversity found in these countries. In this case there is an urgent need to collect, conserve and utilize germplasm from both Tanzania and South America for broadening the genetic base of cassava in the other African countries

The close genetic similarities observed among genotype Namikonga from Tanzania and Venezuela from South America (0.3357) followed by Kibaha from Tanzania and Colombia 297 from South America (0.4417) may be due to the historical events of cassava introduction in Africa as described in section 2.1.2.

The genotypes BRA990 and Mex 55 from South America showed the highest genetic distance (0.7089) as compared to the genotypes Albert and Mkita from Tanzania (0.6036) and TME230 and TME 539 from other African countries (0.5913). This is an expected result; although there is substantial diversity in Africa there are some factors that reduce the diversity. Being mainly a vegetatively propagated crop, a reduction in genetic diversity is due to the accumulation of systemic pathogens and the spread of a few, vigorous, adapted landraces could be expected. This together with possible founder effect (loss of

genetic variation) during the crop's spread to Africa further influence the diversity and the genetic base.

The study has also identified genetically closely related genotypes such as genotype MEX 55 and PER (D=0.1732), followed by TME 1389 and TME 1368 with 0.1982, MCOL1734 and MCOL1468 MKITA and KALOLO with 0.3981, TME 396 and Mwaya (D= 0.483) as indicated in (Fig. 2 and Table 6) are the most closely related genotype since they had the smallest genetic distance between them and clustered together in the dendrogram. This similarity can be explained based on the common geographical area from which the planting materials were collected. The most distantly related genotypes were TME 230 and kigoma at a distance of 0.7582, BRA 990 and MEX 55 (0.7089), Albert and Mkita (D= 0.6036), BRA255 and CuB1 (D=0.6054), TME 230 and TME 539 (D=0.5914), Namikonga and Albert (D=0.5139).

This dissimilarity observed between cassava genotypes is also possibly a result of evolutionary adaptations and divergence over time that might have occurred, which cause accumulation of allele unique to Rest of Africa, Tanzania and S.A. The accumulation of novel variation suggest that germplasm from Rest of Africa has been separated from that in Tanzania. Also this dissimilarity is due to the historical events on introduction of cassava in Africa from South America. It is known that there were two independent arrivals of cassava from South America to Africa. Cassava arrived and diffused earlier at the West African coast by the15th century compared with the 18th century in East Africa (Carteret *al*., 1992). This might cause accumulation of allele unique to Rest of Africa, Tanzania and South. America. The analysis revealed that genotype Albert is quite separated in the phylogenetic cluster to which it belongs. This may be due to limited introgression of Albert with other genotypes because of its high susceptibility to CBSD.

The third and last specific objective was to identify polymorphic loci of ESSR markers associated with resistance to cassava brown streak disease (CBSD). Eight (8) ESSR markers showed fidelity for effectiveness and efficiency in distinguishing resistant and susceptible genotypes to CBSD. Thus providing tools for detection and mapping of major genes for marker-assisted selection in cassava breeding programmes. Therefore eight ESSR markers have proven to be effective in overcoming the limitations of traditional breeding methods. These ESSR markers will enable the breeder to identify and eliminate at an early stage susceptible genotypes of cassava. Eight (8) number of loci reported in this study demonstrates high variability between resistant and susceptible parents to CBSD. Therefore researchers should use these markers to speed up and improve the efficiency of their breeding programmes.

The highly PIC value indicates that the markers were informative and can distinguish resistance and susceptible genotypes to CBSD. In this study eight (8) markers namely ESSR 110, ESSR 121, ESSR 130, ESSR 75, ESSR 91, ESSR 95, ESSR 96 and ESSR 116 demonstrated high PIC values (Table 5),.are therefore highly informative linked to CBSD resistance. The 8 markers preliminary indicate their ability to identify genotypes or individuals that are resistant to CBSD. Further work is required to develop a population with a pool of individuals that are resistant to CBSD and a pool for susceptible in order to link specific markers to resistance that can be use in future breeding work. Alternatively segregating populations of general hundreds of cassava genotypes from a cross between Resistant and Susceptible will need to be constituted in order to develop identify QTLs and markers that will be linked to resistance to CBSD. However, to date gene expression for CBSD is still unknown. The use of makers is expected to shorten breeding cycles because selection of resistance can be made earlier in the growth cycle, even at the

seedling stage without waiting for root maturity. In contrast to conventional breeding methods that rely on the direct selection based on phenotypic effect, DNA markers enable indirect selection by identification of desirable genotypes for quantitative traits like resistance to CBSD, and markers for QTL are not influenced by environmental conditions.

Most important among the biotic factors affecting cassava production are viral diseases, specifically CBSD which has recently been recognized as a major limiting factor to cassava production in Africa. Genotype (TME 230, Mkita, Kibaha, Pesangan, Kigoma, and Per 458 BRA 990 and Mwaya, and Mex 55, BRA 255 and TME 533, Albert, Namikonga and Faraja were found to be genetically more distantly related. However, this study has not resolved fully the varying levels of resistance of the above genotypes except for Namikonga and Albert. Albert which predominated in Mtwara District up until 1998, still considered in genetic mapping due to its desirable traits such as high yielding, and palatability (sweet taste). Since the resistant genotype (Namikonga) and susceptible genotype (Albert) are demonstrated to be unrelated genotype at a distance of 0.5139 and did not cluster together.

However the investigation of varying level of resistance to other genotypes should be followed in for future research and more comprehensive use of markers be included in National cassava improvement programme. Therefore the results observed in this study indicates that the 53 ESSR markers reported here should be appropriate for further applications in cassava germplasm characterization and plant breeding.

CHAPTER FIVE

5.0 CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

An annealing temperature of 62°C was found to be optimum for all primer pairs except for ESSR72 and ESSR101 in which 57°C was found to be optimum. Optimum magnesium chloride concentrations varied and are provided in Appendix 2. 0.175 pmole/µl fluorescent nucleotides were found to be appropriate for all reactions. Of the 80 markers selected for optimisation 68 markers were successfully amplified.

The findings from this study showed that, markers used were highly in determining the distinctness of cassava germplasm. ESSR 91 was the most informative with eight (8) of alleles and highest PIC values. This marker was able to distinguish closely related genotypes. The number and frequency of alleles at different DNA marker loci found in this study had not been previously reported. This has allowed for the first time an analysis of fidelity of ESSR for cassava genetic study and identification to CBSD resistance. This study provides reference data with representative alleles across 53 polymorphic ESSR markers that allow comparison of cassava accession for future genetic assessments. Also the findings revealed 14 ESSRs which are able to discriminate resistant and susceptible genotypes to CBSD making them useful markers for inclusion in mapping work for disease resistant loci. The Phylogenetic relationship revealed four major clusters and one minor cluster. Cassava genotype from Tanzania and those from South America are well spread across the dendrogram indicating that the diversity from South America is well captured and represented in all clusters. The 53 ESSR markers reported in this study should be appropriate for further applications in germplasm characterization and plant breeding.

The findings in this study indicated that Albert is in a quite separated Phylogenetic cluster, this may be due to limited introgression of Albert with other genotype because of its susceptibility to CBSD. Since the two genotypes which are Resistance to CBSD (Namikonga) and susceptible (Albert) genotype are genetically unrelated I thereby advocated to be included in National cassava improvement programs in the country genetically dissimilar thus these cultivars can be used for breeding to introgress gene resistance to CBSD.

6.2 Recommendations

Because of the desirable traits of cassava genotype Albert (high yielding and sweet test) there is also a need to introgress genes of resistance to CBSD in order to develop variety with high yield, sweet taste and resistant to CBSD by crossing it with Namikonga which is resistant to CBSD and distantly related.

The ESSRs presented in this study showed to have average polymorphic information content values ranging from 0.173 to 0.844 and can be used to improve the efficiency in tagging agronomically important genes and in diversity studies. There are therefore very useful in implementing cassava genetic improvement research. However further work is required to be undertaken with a wider spectrum of genotypes consisting of varying levels of resistance for more comprehensive use of markers.

REFERENCES

Adams, M. D., Kelly, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulus, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., Kerlavage, A. R., McCombie, W. R. and Venter, J. C. (1991). Complementary DNA sequencing: expressed sequence tags and human genome projects. *Science* 252: 1651 - 1656.

Akano, A. O., Dixon, A. G. O., Mba, C., Barrera, E. and Fregene, M. (2002). Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. *Theoretical and Applied Genetics* 105: 521 – 525.

Anderson, J. V., Fregene, M. A., Jorge, V., Mba, C., Lopez, C., Restrpo, S., Soto, M., Piegu, B., Verdier, V., Cooke, R., Thome, J. and Horvath, P. D. (2004). An EST resource for cassava and other species of Euphorbiacea. *Plant Molecular Breeding* 56: 527 - 539.

Beaulieu, M., Larson, G. P., Geller, L., Flanagan, S. D. and Krontiris, T. G. (2001). PCR candidate region mismatch scanning: adoption to quantitative, high-throughput genotyping. *Nucleic Acid Research* 29: 1114 - 1124.

Beeching, J. R., Marmey, P., Gavalda, M. C., Noirot, M., Haysom, H. R., Hughes, M. A. and Charrier, A. (1993). An assessment of genetic diversity within a Collection of cassava (*M. esculenta* Crantz) Germplasm using molecular markers. *Journal of Botany 72*: 515 - 520.

Blanchard, M. M., Miller, P. and Nowotny, P. (1993). PCR buffer optimization with uniform temperature regime to facilitate automation. *PCR Methods Application* 2: 234 - 240.

Botstein, D., White, R. L., Skolnick, M. and Davis, R. W. (1980). Construction of genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32: 314 - 331.

Buchanan, F.C., Adams, L. J., Littlejohn, R. P., Maddox, J. F. and Crawford, A. M. (1994). Determination of evolutionary relationships among breeds using microsatellites. *Genomics* 22: 397 - 403.

Carter, S. E., Fresco, L. O., Jones, P. G. and Fairbairn, J. N. (1992). *An Atlas of Cassava in Africa Historical, Agroecological and Demographic Aspects of Crop Distribution.* CIAT Ltd., Cali. 585 pp.

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

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B. and Pary, E. C. K. (2005). Introduction to markers, quantitative trait loci (qtl) mapping and marker-assisted selection for crop improvement. *Euphytica 142*: 169 - 196.

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

FAO (1999) Food Statistics, Situation Reports and Crop Forecasts [.http://www.fao.org/docrep/newsroom/news/ao75oeoo.html] site visited on 13/05/2009.

FAO (2002). Partnership formed to improve cassava: The staple food of 600 million people. [.http://www.fao.org/english/newsroom/news/2002/10541-en.html] site visited on 17/10/2007.

FAO (2006). Introduction and spread of cassava in Africa and particularly in Tanzania [.http://www.fao.org/waicent/0is/news/Press_NE/Toc99E.html] site visited on 27/06/2008.

FAO (2009). Food and energy security endeavors: The propel global production [hit://www.fao.org/docrep/012/ak341e/ak341e06.htm.] sited visited on 12/02/2010.

Ferguson, M. E., Burrow, M. D., Schulze, S. R., Bramel, P. J., Paterson, A. H., Kresovich, S. and Mitchell, S. (2004). Microsatellite identification and characterization in peanut (A.hypogaea L.). *Theoretical and Applied Genetics* 108: 1064 - 1070.

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

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

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

Fregene, M., Suarez, M., Mkumbira, J., Kulembeka, H., Ndedya, E., Kullaya, A. (2003) Diversity in cassava landraces and differentiation in asexual propagated crop. *Theoretical Applied Genetics* 107: 1083-1093.

Gibbons, R. W., Bunting, A. H. and Smartt, J. (1972). The classification of groundnut (*Arachis hypogaea .L.*). *Euphytica* 21: 78 - 85

Gupta, P. K., Balyan, H. S., Sharma, P. C. and Ramesh, B. (1996). Microsatellite in plants: a new class of molecular marker. *Plant Science* 70: 45 - 54.

Hamrick, J. L. and Godt, M. J. W. (1997). Allozyme diversity in cultivated crops. *Crop Science 37*: 26 - 30.

Han, Z. G., Guo, W. Z., Song, X. L. and Zhang, T. Z. (2004). Genetic mapping of EST-derived microsatellites from the diploid *Gossypium Arboreum* in allotetraploid cotton. *Molecular Genetics and Genomics* 272: 308 - 327.

Innis, M. A. and Gelfand, D. H. (Eds.) (1990). *Optimization of PCR's protocols: A Guide to Methods and Applications.* Academic press., New York. 312pp.

IPGRI and Cornell University (2003). Molecular marker: the technology in studies on plant genetic diversity [http://www.igd.cornell.ed/training and Education.html] site visited on 13/11/2009.

Jorge, V., Fregene, M. A., Duque, M. C., Bonierbale, M., Tohme, J. and Verdier, V. (2000). Genetic mapping of resistance to bacterial blight disease. *Theoretical and Applied Genetics* 101: 865 - 872.

Jorge, V., Fregene, M., Velez, C. M., Duque, M. C., Tohme, J. and Verdier, V. (2001). QTL analysis of field resistance to *Xanthomonas asonopodis* pv. *manihotis* in cassava. *Theoretical and Applied Genetics* 102: 564 – 571.

Kullaya, A., Mtunda, K., Kulembeka, H., Ferguson, M., Marin, J., Ospina, C., Barrera, E., Jarvis, A., Morante, N., Ceballos, H., Tohme, J. and Fregene, M. (2004). Molecular Marker Assisted and Farmer Participatory Improvement of Cassava Germplasm for Farmer/Market Preferred Traits in Tanzania. In: *Proceedings on the Sixth International Scientific Meeting of the Cassava Biotechnology Network.* (Edited by Alves, A. and Tohme, J.), *8-14 March 2004*, Cali, Colombia. pp. 70-81.

Liu, S., Cantrell, R. G., McCary, J. C. and Sterwart, J. (2000). Simple sequence repeat-based assessment of genetic diversity in cotton race stock accessions. *crop Science* 40: 1459-1469.

Lopez, C., Jorge, V. and Piegu, B. (2004). A unigene catalogue of 5700 expressed genes in cassava. *Plant Molecular Biology* 56: 541–554.

Lozano, J. C. and Sequeira, L. (1974). Bacterial blight of cassava in colombia: epidemiology and control. *Phytopathology* 64: 83 – 88.

MacIntosh, G. C., Wilkerson, C. and Green, P. J. (2001). Identification and analysis of analysis of Arabidopsis expressed sequence tags characteristic of non-coding Rnas. *Plant Physiology.* 127(3): 765 - 776.

Marmey, P., Beeching, J. R., Hamon, S. and Charrier, A. (1994). Evaluation of cassava (*manihot esculenta* crantz) germplasm collection using rapd markers. *Euphytica* 14: 203 - 209.

Mba, R. E. C., Stephenson, P., Edwards, K., Melzer, S., Mkumbira, J., Gullberg, U., Apel, K., Gale, M., Tohme, J. and Fregene, M.A. (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.

Milbourne, D., Meyer, R. and Bradshaw, J. E. (1997). comparison of PCR-based marker systems for the analysis of genetic relationships in cultivated potato. *Molecular Breeding* 3: 127 - 136.

Moyib, O. K., Odunola, O. K. and Dixon, A. G. O. (2007). SSR markers reveal genetic variation between improved cultivar and landraces within a collection of Nigeria cassava germplasm. *African Journal of Biotechnology* 6(23): 2666 - 2674.

Muhana, M. and Mtunda. K. J. (2002). The effect of cassava brown streak virus disease yield and quality of cassava. *Journal of Phytopathology* 149: 389 - 394.

Nei, M. (1983). Analysis of gene diversity in subdivided populations. *Proceedings of Thenational Academy of Science* 70: 3321-3323.

Nei, M. and Saitoh, N., (1987). The neighbor-joining method: a new method for constructing phylogenetic trees. *Molecular Biology Evolution* 4(4): 406 - 425.

Nichols, R.F.W. (1947). Breeding cassava for virus resistance. *East African Agricultural Journal 15*: 154 - 160.

Nichols, R. F. J. (1950).The brown streak disease of cassava. *East African Agricultural Journal* 15: 154 –160.

Olsen, K.M. and Sachaal, B. (2001). Microsatellite variation in cassava *Manihot Esculenta, Euphorbiaceae* and its wild relatives: Evidence for a Southern Amazonian origin of domestication. *American Journal of Botany* 131 - 142.

Tanzania Ministry of Agriculture and Cooperatives (1996). *Planning and Marketing Divisio in Tanzania Annual Report.* Government Printer, Dar-es-Salaam, Tanzania. 12pp.

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

Roa, A. C., Chavarriaga, A., guirre, P., M. C., Duque, M. M., Maya, M. W., Bonierbale, C. and Tohme, J. (2000). Cross species amplification of cassava (*Manihot Esculenta*) (Euphorbiaceae) microsatellites: Allelic polymorphism and degree of relationship. *American Journal of Botany* 87(11): 1647 - 1655.

Saha, M. C., Mian, M. A., Eujayl, I., Zwonitzer, J. C., Wang, L., May, G. D. (2004). Tall fescue EST-SSR markers with transferability across several grass species. *Theoretical and Application Genetics* 109: 783 - 791.

Southern, E. M. (1975). Southern hybridization detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* 98: 503 – 507.

Storey, H. H. (1936). Virus diseases in east african plants. *East African Agricultural Journal* 2: 34 - 39.

Thiel, T., Michalek, W., Varshney, R. and Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived ssr-markers in barley (*Hordeum Vulgare* L.). *Theoretical and Application Genetics* 106: 411- 422.

Thresh, J. M., Nape, G. W., Legg, J. P. and Fargette, D. (1997). African cassava mosaic virus disease: the magnitude of the problem. *African Journal of Root and Tuber Crops* 2: 36 - 53.

Vos, R., Hogers, M., Reijens, T., Van der Lee, M., Horns, A., Frijters, J., Pot, J., Peleman, M., and. Zabeau, M. (1995). AFLP: A new technique for DNA finger printing. *Nucleic Acid Research* 2: 4407 - 441.

Wang, M. L., Barkley, N. A., Yu, J. K., Dean, R. E., Newman, M. L., Sorrells, M. E. and Pedersonl, G.A. (2005). Transfer of simple sequence repeat (ssr) markers from major cereal crops to minor grass species for germplasm characterization and evaluation. *Plant Genetic Research* 3: 45 – 57.

Wanger, Z., Weber, J. L., Zhong, G. and Tanksley, S. D. (1994). Survey of plant short tandem DNA repeat. *Theoretical Applied Genetics* 88:1- 6.

Weber, J. L. and May, P.E. (1989). Abundant class of human dna polymorphisms, which can be typed using the polymerase chain reaction. *American Journal of Human Heredity* 44: 388 - 396.

Weising, K., Winter, P., Huttel, B. and Kohl, G. (1998). Microsatellite markers for molecular breeding. *Journal of Crop Production* 1:113-143.

Whitefield, C. W., Band, M. R., Bonaldo, M. F., Kumar, C. G., Liu, L., Pardinas, J. R, Robertson, M., Soares, M. B. and Robinson G. E. (2002). Annotated expressed sequence tags and cdna micro arrays for studies of brain and behaviour in the honey bee. *Genome Research* 12: 555 -566.

Witsenboer, H., Vogel, J. and Michelmore, R. (1997) Identification, genetic localization and allelic diversity of selectively amplified microsatellite polymorphic loci in lettuce and wild relatives (*Lactuca spp*). *Genome* 40: 923 - 936.

Welsh, J. and McClelland, M. (1990). Fingerprinting genomes using pcr with arbitrary primers. *Nucleic Acid Research* 18: 7213 – 7218.

Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A. and Tingey, S. V. (1990). DNA polymorphism amplified by arbitrary primers is useful as genetic markers. *Nucleic Acid Research* 18: 6531 - 6535.

Wu, K. S. and Tanksley, S. D. (1993). Abundance, polymorphism and genetic mapping of microsatellites in rice. *Molecular Genetics and Genomics 241*: 225 – 235.

APPENDICES

Appendix 1: List of ESSR primers and target repeats and sequences used in the study

IITA Name	Forward primer	Reverse Primer	Nucleotide and no of repeats	Product size (bp)
IITA ESSR71	TGCTCCGCCAATTCAAGCCG	GGTGGAGGCGGAGGGCAGTA	(CCA)5	183
IITA ESSR72	TCATCACCTCTCGCAAAACAAAAGGC	GCAACATCCTGGCTTTAGGAGTGTCC	(GAA)5	190
IITA ESSR73	ACGATCGAGCTGCAACCCCAA	CGGTGTAGAAGTTGAAGCCGTTGAACC	(TCA)5	167
IITA ESSR74	TGGCAAAGACATGGGCGGTC	TGCGAGGCGTGTTACAGCTCTAATTG	(GGT)6	174
IITA ESSR75	TTGCAGCTTCCAAATGATGAGTGTTCA	CATCTCTTGCTCATCGAAGTCAAGGC	(ATG)5	180
IITA ESSR76	ACAGCTGGTCCTTTGCTCCGTTG	TGAAGCGCAGTGATTTTGGCTCC	(TCT)7	152
IITA ESSR77	GAGCGTCTCTCCGGTGACGTTG	CGGAGCAAATTATCATCATCGAACCA	(TCT)10	158
IITA ESSR78	TTCCTTCAATTCTTGGCCATCCCC	GCATGTTGAGCCTGGGCTTGC	(GCA)5	157
IITA ESSR79	CTACCTCAAGGACAAAAAGTGAGGCCA	CGGGGAACTTCTCAATGTTTTTCTTGG	(GCA)5	171
IITA ESSR80	CATCAATACCGGCGCTAATTCTTCCA	TCCACTACAGACGCCTCTCCGTTC	(GAT)5	168
IITA ESSR81	CTGCTTCCTCCGTCTTCCAAACGC	ATCGGCGACGCTGAGATTCCC	(AGA)5	150
IITA ESSR82	TGAAGAGAATGAGGAGGGTGGATTGA	TCCTCCTTAGGTGCATCCTTCTTTGG	(TGA)5	174
IITA ESSR83	TCCTGGAGTTTGCTCCATTTGTTTGA	CCATGCCATGTCGATAACGATGC	(TTC)5	185
IITA ESSR84	TGAAGGGAAGGCTCAAATGAAGAGCA	CACTCCTCTCTGCACCACCACCC	(GGA)5	184
IITA ESSR85	AAGTCACCACCCCCACCCCC	CACTGCATCAACGTTCGATCTCGTCT	(CCA)5	185
IITA ESSR86	CAAAAGATGGGTGTTGGAGGAACTTTG	CATTCTGAACCTTGAGCTCACAGCCAT	(AAG)5	/151
IITA ESSR87	AAATCTGTGGCCTCTACGATTCACCC	CCCGGGTATTCTTGTCAACAAAGACAG	(TCT)5	154
IITA ESSR88	TCAGAGCCTTTTGGGTAAAACCAAGC	GGGCCTGGACTGGAACAGAAGAGAA	(ATC)5	175
IITA ESSR89	TGGAACTGGTTTCGGCGGTG	TCACGTCCTCTCTTTTGTCCAATCCAG	(TCT)6	165
IITA ESSR90	TGGCTGACCTGCCAATGGTTTTG	GCTAGCTCAGTCACCCATGTCTCCAC	(AGA)5	181
IITA ESSR91	TTCATTTCCAGGAGACCACCCTTTCA	CCAAAGGAACACTCTCTGGGCCG	(CTC)7	172
IITA ESSR92	TTGCAATGAGCTCCCTGATTATGCTG	GCCTTCACATGGCACGGTCG	(TGA)6	165
IITA ESSR93	CTTGGAGCTGCCTCTCACTACGGC	GGGTTGACGACTGGGTGGCG	(CCA)6	169
IITA ESSR94	AAGGTGGCAACAGCAAATGGATGG	GCAAAAGGGTGGATAGTTTTGGTTTGG	(TCT)8	162
IITA ESSR95	CCATCTCCAGAGCTCTAAACCGCCA	CGAGGAAGCATTCGAGACGCA	(CTT)6	172

IITA Name	Forward primer	Reverse Primer	Nucleotide and no of repeats	Product size (bp)
IITA ESSR96	TTTTTCCCTTTTCACCGCTCTGGTT	TTGAATCACCGCTGCTCTGATGG	(AGA)8	150
IITA ESSR97	CAAACCCTTTTCTCCCATCGCCT	GCGTCAACGGTTGCAGATTCTGACT	(TCT)6	200
IITA ESSR98	CGAAGTTGGTGTCACAGAAGATCACGA	TTGGAGGAGGACCACCTTCTTTGC	(GAA)6	184
IITA ESSR99	TCACTCATGCAGCTAACAAGGGCG	TGCTTCCACTTCCACCTGAGGTTTACA	(TTC)5	176
IITA ESSR100	CTACCTTCAAAGCTTCAGCCTTCAGCA	AGACCTGCAGCTGTTTTGAGGCCTTT	(AAG)5	150
IITA ESSR101	GAAGGAGCAACGTGATTACAGGGCA	CGAGGAGGCATTGCTTGGCG	(AGA)8	166
IITA ESSR102	TGGAATTGTTGGACGCCGGA	CCGCGTAGAATCCTGAATCGCC	(ACC)6	173
IITA ESSR103	TCCCATGTTGGAGACCCCATCG	CCTCTGGTGGCTGTTGCCCA	(GAA)6	176
IITA ESSR104	TGCCTCCTCCTCCTCCACTGC	GGATCCGGAAGAGAGAAAAGAGGACC	(TTC)11	159
IITA ESSR105	CCGACGGCAAGGCTCAGCTC	CCCAGCATCATCTTGAAAGCGAGG	(TTC)7	71
IITA ESSR106	CGAGCTCCGCCTCTTTCCCA	TGCCTGAAGCTCTGTAGCATCGC	(GCA)5	166
IITA ESSR107	GGGGCTGGGTTTTGGGGGA	TGGGCTTCATGGCCAAAGGA	(TCT)5	164
IITA ESSR108	GGTGGAGCTGGAGGGAGTCGTG	AACTCAATCCCAGGCGGCGG	(GCT)5	150
IITA ESSR109	TGCAACCCCAAAGATCAGGGC	GCATCTCGGTGTAGAAGTTGAAGCCG	(TCA)5	163
IITA ESSR110	TTTATAAATACTCCGCCCTTCCCATCG	CGCCGCTTTGGACGAGTCAAC	(TCG)5	164
IITA ESSR111	GATGAAGCTGTTGAAACTGCAAAGCAA	GCTCGCCGAAGCTCGGGAATA	(GCT)5	157
IITA ESSR112	TTCCGTCGTTCAAATGCTTCCG	CGGCGGTTGCTTGACAGCTC	(CTT)5	156
IITA ESSR113	TCAAGCGAAGAGCATCAGAGCAAATCT	TGTCAGGATTGTAGGGATTGACGGC	(GCA)9	175
IITA ESSR114	GGGATCTGCGCCCATCACAGA	CAGCTGCACGCTCTTCCGCA	(TGA)5	152
IITA ESSR115	TACGAGCATCGCTCTCTTCCCTCA	CGTGTAGTGCTCCGCAAACTACCGA	(CAA)6	165
IITA ESSR116	CGCCTCTTTCCCACCTCCGA	GAATGGCTTTCGCCTGAAGCTCTG	(GCA)7	175
IITA ESSR117	CTTTGAAAGAACTCCAAATGGCCACC	TGGAGGTGGTGGGTTTGAGGAAAA	(CAC)8	150
IITA ESSR118	GTTGTGAAAGCCGAGAAGACGCC	TCTCCGGCAATGCAATTCTACTGCTT	(AGA)7	151
IITA ESSR119	GATGGAGAGGATGCTTATGATATGCGG	GCTTTTGCCTCTGATTTTGGGTCTGG	(CAT)6	152
IITA ESSR120	AACGGCGGGTGTCTCACCTCC	GGTTGATGATAGGAGCAGGAGCAGC	(CTT)5	166
IITA ESSR121	GAAGGACCAAGGGTTTTGCAGATGG	TCCTCAAAGATTCCTCTGCCTGTCTTG	(TCT)8	179
IITA ESSR122	CCATTCTTTGCTACACATGGAGGGC	TCAGGAGTACCCTTCTGAAAGGGGAAA	(TTC)5	190

IITA Name	Forward primer	Reverse Primer	Nucleotide and no of repeats	Product size (bp)
IITA ESSR124	GCAGCTGCCGTTTTATCCTCTTGC	TGCTCATCGACAAGCACCCTTTG	(CTT)5	178
IITA ESSR125	TTCTCCTGGATATCAAGCTCCGCA	ACGACCACGTCCTCGGCCAT	(CAG)6	154
IITA ESSR126	GGGGATGAGAGCGGTCTTGGTTTT	TGGTGCTGAGGTAATGAAGGCATGG	(TTC)5	166
IITA ESSR127	CTCGCCAACTCTTTGGCCGC	TGCCATTGCTGGGAACTTGTCC	(ATC)5	177
IITA ESSR128	ACGAGGTACGGCCGGGGAGA	GGGAAGGTTGGTAAGCAGTTGCATGA	(TCT)5	171
IITA ESSR129	CCAACACATGTTTACATGGCTCAGCA	TCAACGGTTGCAGATTCTGACTGGA	(TCT)6	163
IITA ESSR130	GCAAGCAACAACGTTTCTGAGAGCTG	GTTTCTTGGCCGTCCTGGCG	(TTC)10	157
IITA ESSR131	GTTGACTCTGGGGTGACCAAGATTCCT	AACCCGCTTACCATAACTTTCAATGCC	(CAG)5	174
IITA ESSR132	CTCCCATACCACTGGAACAGCCTCA	TCCCAGTTGGCCGGTGGTTG	(AGC)7	179
IITA ESSR133	TGTCCGCGGAGGGATCGTCT	CGCTATTCTTTCCGATGGACGACG	(AGG)6	155
IITA ESSR134	TACGAGTTCATTGGCTGCAGTTATGGC	AAGGATGAAGGTGGTTTGGTTACAGCA	(ATG)5	153
IITA ESSR135	AAGGCTGAGGACCAAGGTGCTGAAT	CGCAAGATAAAAATCCTCCAAACCCA	(GAA)8	193
IITA ESSR136	GCTCATCGGGTCAACAACCGC	CTCGGGTCCTCTGGGCCCTC	(GAA)6	179
IITA ESSR137	CCGAGAATGCTTCCAGGACCCC	TGAGCTCGGCCCTTTTGCCA	(GAG)5	168
IITA ESSR138	TGCCTGTGTGCACTGAGCGG	GCAACTCCAACAAGAACAAGCCCA	(GGT)5	172
IITA ESSR139	GGAAATTGCATTAAAGCGTGTGTTTGG	GCCTGGCTCCATCCTGTCTGC	(TC)8	158
IITA ESSR140	GGGTCCCCATGTAAATTGTAAAACCCC	TGGCTAGAACCACTCCCTGTGAGCA	(TTC)10	152
IITA ESSR141	ACGGGTTACTGAGTTTTGATTGCATGG	GATCATCGTCACCGTCCAATAATCCA	(AGA)9	150
IITA ESSR142	GGGGATTCATGGGTCGTGCC	TCCTGTTGAAAGAAAAACCGACCCAA	(GAA)6	152
IITA ESSR143	CATTCCCGTCAACCAACCCTTTCTC	TCCATGCAAAGTCCATCCGCA	(CT)6	171
IITA ESSR144	AGAAGAACATTCAAAATGCAAAGGGGG	TCTGCAGGAGGGAGGCCAGC	(GA)7	169
IITA ESSR145	GTCCCTCAAACCCCAAAACTCCCC	TTCTTCTGCCTGGGTCGCCA	(TC)6	162
IITA ESSR146	CCCCAATTTGCAGCAGCAAGGA	GTGGCAAACCCGCCAAGCAT	(CAG)6	181
IITA ESSR147	GAGAATTATGGCTATGCCCCTGCTCC	TCATCAGTTGAAAGCACAAGAATGCCA	(CAG)9	163
IITA ESSR148	TTGGATCTCTTGCTTTGCAATTTGAGG	ACAGCCAACAAAACCAAACAGCTCC	(TTC)6	162
IITA ESSR149	TGGGCTTGAAGAAGGATGGCAA	TCCCTGGCCCTTACTCATGGAAAA	(GAA)6	195
IITA ESSR150	GCGGAGGGATCGTCTGCGTC	TGACTCTCTCCCCGACAGATCGC	(AGG)6	170

Appendix 2: Optimized PCR conditions for amplification of ESSR primers

Primer	Obcrved size (bp)	Expected size	Optimum $MgCl_2$ mM	Best Annealing (TA) °C
IITA_ESSR71	208	183	1.0 - 2.5	62
IITA_ESSR72	215	190	2	57
IITA_ESSR73	189	167	1.5- 2.5	62
IITA_ESSR74	-	174	-	-
IITA_ESSR75	205	180	1 - 3	62
IITA_ESSR76	177	152	2	62
IITA_ESSR77	172	158	1- 2	62
IITA_ESSR78	179	157	1-3	62
IITA_ESSR79	191	171	1-3	62
IITA_ESSR80	190	168	2	62
IITA_ESSR81	173	150	1-3	62
IITA_ESSR82	476	174	1-3	62
IITA_ESSR83	212	185	1.5	62
IITA_ESSR84	209	184	1-2	62
IITA_ESSR85	293	185	2	62
IITA_ESSR86	177	151	2	62
IITA_ESSR87	160	154	1-1.5	62
IITA_ESSR88	198	175	1 -3	62
IITA_ESSR89	-	165	-	-
IITA_ESSR90	-	181	-	-
IITA_ESSR91	196	172	2	62
IITA_ESSR92	428	165	2	62
IITA_ESSR93	192	169	1-3	62
IITA_ESSR94	188	162	1.5 and 3	62
IITA_ESSR95	196	172	1-1.5	62
IITA_ESSR96	161	150	1-2.5	62
IITA_ESSR97	224	200	1.5	62
IITA_ESSR98	205	184	1-2.5	62
IITA_ESSR99	-	176	-	-
IITA_ESSR100	196	150	-	-
IITA_ESSR101	187	166	1.0 - 2.5	57
IITA_ESSR102	44	173	2	62
IITA_ESSR103	-	176	1.5- 2.5	62
IITA_ESSR104	59	159	-	-
IITA_ESSR105	129	171	1 - 3	62
IITA_ESSR106	192	166	2	62
IITA_ESSR107	554	164	1- 2	62
IITA_ESSR108	175	150	1-3	62
IITA_ESSR109	189	163	1-3	62
IITA_ESSR110	179	164	2	62
IITA_ESSR111	170	157	1-3	62
IITA_ESSR112	-	156	1-3	-
IITA_ESSR113	-	175	1.5	-
IITA_ESSR114	400	152	1-2	62

Primer	Obcrved size (bp)	Expected size	Optimum $MgCl_2$ mM	Best Annealing (TA) °C
IITA_ESSR116	197	175	2	62
IITA_ESSR117	141	117	2	62
IITA_ESSR118	-	151	1.0 - 2.5	-
IITA_ESSR119	175	152	2	62
IITA_ESSR120	191	166	1.5- 2.5	62
IITA_ESSR121	-	179	-	-
IITA_ESSR122	215	190	1 - 3	62
IITA_ESSR123	180	155	2	62
IITA_ESSR124	62	178	1- 2	62
IITA_ESSR125	179	154	1-3	62
IITA_ESSR126	193	166	1-3	62
IITA_ESSR127	181	177	2	62
IITA_ESSR128	169	171	1-3	62
IITA_ESSR129	189	163	1-3	62
IITA_ESSR130	177	157	1.5	62
IITA_ESSR131	195	174	1-2	62
IITA_ESSR132	192	179	2	62
IITA_ESSR133	186	155	2	62
IITA_ESSR134	180	153	1.0 - 2.5	62
IITA_ESSR135	217	193	2	62
IITA_ESSR136	137	179	1.5- 2.5	62
IITA_ESSR137	-	168	-	-
IITA_ESSR138	197	172	1 - 3	62
IITA_ESSR139	181	158	2	62
IITA_ESSR140	169	152	1- 2	62
IITA_ESSR141	173	150	1-3	62
IITA_ESSR142	177	152	1-3	62
IITA_ESSR143	193	171	2	62
IITA_ESSR144	195	169	1-3	62
IITA_ESSR145	201	162	1-3	62
IITA_ESSR146	195	181	1.5	62
IITA_ESSR147	69	163	1-2	62
IITA_ESSR148	202	162	2	62
IITA_ESSR149	69	195	2	62
IITA_ESSR150	199	170	2	62

Appendix 3: Number of alleles, allele size, diversity, degree of availability and PIC of ESSR markers observed in different genotypes of cassava

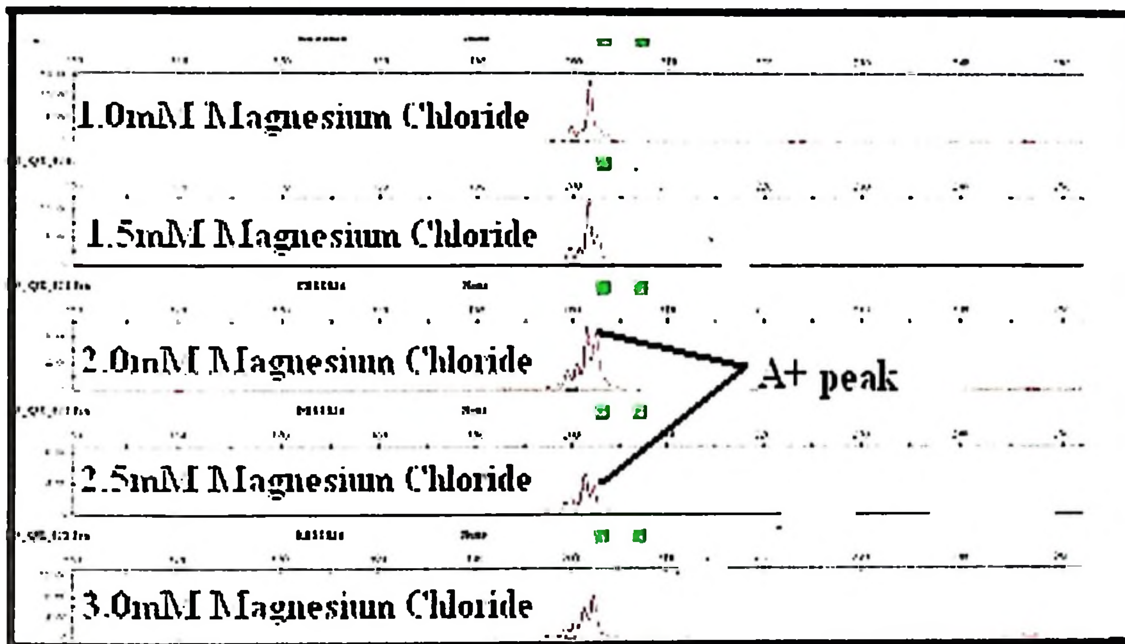
Marker	Genotype Number	Number of alleles	Observed alleles bp	Gene Diversity	Heterozygosity	Availability	PIC
ESSR71	2	2	210, 181	0.2778	0.3333	1.0000	0.2392
ESSR78	2	2	256, 132	0.4783	0.7917	1.0000	0.3639
ESSR79	3	2	192, 147	0.4274	0.5238	0.8750	0.3361
ESSR87	3	2	190, 185	0.2778	0.2500	1.0000	0.2392
ESSR109	2	2	187, 189	0.4861	0.0000	1.0000	0.3680
ESSR122	2	2	216, 187	0.4991	0.9583	1.0000	0.3746
ESSR124	3	2	194, 175	0.4592	0.1429	0.2917	0.3538
ESSR127	3	2	198, 124	0.2188	0.1667	1.0000	0.1948
ESSR133	2	2	184, 181	0.4297	0.6250	1.0000	0.3374
ESSR76	5	3	180, 192, 154	0.5720	0.9167	1.0000	0.4783
ESSR86	3	3	130, 157, 181	0.2143	0.2381	0.8750	0.1993
ESSR94	4	3	189, 192, 197	0.4028	0.4167	1.0000	0.3633
ESSR97	3	3	167, 176, 223	0.3809	0.4783	0.9583	0.3343
ESSR104	3	3	183, 190, 222	0.1838	0.2000	0.8333	0.1736
ESSR107	4	3	165, 166, 173	0.4753	0.4444	0.3750	0.4035
ESSR110	6	3	182, 187, 191	0.6016	0.5833	1.0000	0.5344
ESSR115	5	3	188, 194, 196	0.5512	0.6250	1.0000	0.4657
ESSR120	6	3	187, 190, 198	0.6085	0.6667	1.0000	0.5282
ESSR136	4	3	201, 204, 168	0.3793	0.3750	1.0000	0.3475
ESSR140	3	3	170, 176, 681	0.5747	0.1739	0.9583	0.4819
ESSR72	6	4	191, 214, 217, 168.	0.6156	0.6667	0.8750	0.5613

Marker	Genotype Number	Number of alleles	Observed alleles bp	Gene Diversity	Heterozygosity	Availability	PIC
ESSR75	6	4	201, 203, 205, 177	0.6146	0.6667	1.0000	0.5568
ESSR88	4	4	198, 201, 204, 191.	0.3568	0.3333	1.0000	0.3360
ESSR98	6	4	182, 188, 191, 194	0.5331	0.4783	0.9583	0.4609
ESSR102	6	4	168,175, 187, 193	0.6388	0.4500	0.8333	0.5791
ESSR113	5	4	177, 185, 194, 204	0.4696	0.5417	1.0000	0.3931
ESSR114	6	4	158, 170, 173, 179	0.5868	0.6250	1.0000	0.5146
ESSR126	7	4	178, 181, 194, 124	0.6710	0.7917	1.0000	0.6022
ESSR129	7	4	184, 187, 189, 172	0.6753	0.6667	1.0000	0.6068
ESSR145	7	4	197, 203, 211, 195	0.6593	0.2632	0.7917	0.5973
ESSR146	6	4	202, 205, 210, 207	0.6188	0.5455	0.9167	0.5548
ESSR77	6	5	164, 173, 179, 182, 180	0.7283	1.0000	1.0000	0.6812
ESSR95	7	5	204, 206, 213, 215, 161	0.6510	0.6667	1.0000	0.5976
ESSR101	6	5	174, 180, 189, 194, 200	0.5694	0.6250	1.0000	0.4822
ESSR105	6	5	182, 188, 191, 194, 198	0.5217	0.4583	1.0000	0.4528
ESSR117	6	5	102, 118, 155, 170, 102	0.6059	0.7083	1.0000	0.5366

Marker	Genotype Number	Number of alleles	Observed alleles bp	Gene Diversity	Heterozy gosity	Availability	PIC
ESSR121	7	5	199, 205, 208, 209, 175	0.6345	0.6667	1.0000	0.5733
ESSR130	7	5	174, 177, 183, 192, 192	0.6649	0.6667	1.0000	0.5998
ESSR131	8	5	194, 196, 197, 216, 103	0.7378	0.7083	1.0000	0.6920
ESSR141	6	5	171, 174, 180, 184, 192.	0.4584	0.4783	0.9583	0.4303
ESSR142	7	5	203, 207, 211, 221, 195	0.6736	0.9091	0.9167	0.6156
ESSR82	7	6	157, 181, 214, 222, 227, 105	0.6597	0.8696	0.9583	0.6201
ESSR92	8	6	166, 169, 172, 175, 215, 186	0.6250	0.9167	1.0000	0.5531
ESSR106	10	6	101, 185, 190, 192, 198, 212	0.7250	0.7391	0.9583	0.6817
ESSR111	10	6	121, 136, 187, 196, 209, 260	0.7335	0.7083	1.0000	0.6882
ESSR116	9	6	186, 190, 192, 196, 202, 209	0.6832	0.7500	1.0000	0.6344
ESSR135	8	6	216, 218, 219 227, 236, 197	0.6137	0.4583	1.0000	0.5812

Marker	Genotype Number	Number of alleles	Observed alleles bp	Gene Diversity	Heterozygosity	Availability	PIC
ESSR96	10	7	166, 170, 173, 176, 176, 178, 204	0.6467	0.6667	1.0000	0.6224
ESSR132	8	7	140, 190, 197, 201, 203, 316, 179	0.6033	0.7500	1.0000	0.5514
ESSR143	8	7	199, 203, 206, 207, 211, 221, 101	0.6911	0.9091	0.9167	0.6461
ESSR91	15	8	192, 196, 198, 202, 204, 206, 207, 163	0.8602	0.9167	1.0000	0.8442
ESSR128	11	10	142, 170, 198, 199, 213, 226, 274, 283, 319, 178	0.5822	0.5217	0.9583	0.5658
ESSR74	11	11	180, 192, 194, 197, 202, 203, 209, 213 225, 232, 179	0.5156	0.6250	1.0000	0.5044
Mean	5.9434	4.3585		0.5495	0.5784	0.9473	0.4912
Total allele			231				

Appendix 4: Peaks showed the Effect of Magnesium Chloride concentrations



Appendix 5: Appendices 4 List of allele, frequency (freq), variance and standard (SD) deviation

Marker	Allele	Count	Freq	Variance	SD
ESSR101	174	1	0.0208	0.00041594	0.0204
ESSR101	180	1	0.0208	0.00041594	0.0204
ESSR101	189	25	0.5208	0.00475622	0.0690
ESSR101	194	19	0.3958	0.00432219	0.0657
ESSR101	200	2	0.0417	0.00079572	0.0282
ESSR102	168	3	0.0750	0.00284375	0.0533
ESSR102	175	5	0.1250	0.00484375	0.0696
ESSR102	187	20	0.5000	0.00750000	0.0866
ESSR102	193	12	0.3000	0.00550000	0.0742
ESSR104	183	1	0.0250	0.00059375	0.0244
ESSR104	190	36	0.9000	0.00200000	0.0447
ESSR104	222	3	0.0750	0.00159375	0.0399
ESSR105	182	30	0.6250	0.00542535	0.0737
ESSR105	188	1	0.0208	0.00041594	0.0204
ESSR105	191	2	0.0417	0.00166377	0.0408
ESSR105	194	14	0.2917	0.00426794	0.0653
ESSR105	198	1	0.0208	0.00041594	0.0204
ESSR106	101	3	0.0652	0.00217802	0.0467
ESSR106	185	18	0.3913	0.00373962	0.0612
ESSR106	190	6	0.1304	0.00209583	0.0458
ESSR106	192	14	0.3043	0.00447933	0.0669
ESSR106	198	4	0.0870	0.00250678	0.0501
ESSR106	212	1	0.0217	0.00045204	0.0213
ESSR107	165	1	0.0556	0.00274348	0.0524
ESSR107	166	5	0.2778	0.01303155	0.1142
ESSR107	173	12	0.6667	0.01234568	0.1111
ESSR109	187	20	0.4167	0.01012731	0.1006
ESSR109	189	28	0.5833	0.01012731	0.1006
ESSR110	182	11	0.2292	0.00345414	0.0588
ESSR110	187	11	0.2292	0.00345414	0.0588
ESSR110	191	26	0.5417	0.00600405	0.0775
ESSR111	121	16	0.3333	0.00491898	0.0701
ESSR111	136	1	0.0208	0.00041594	0.0204
ESSR111	187	8	0.1667	0.00231481	0.0481
ESSR111	196	16	0.3333	0.00405093	0.0636
ESSR111	209	6	0.1250	0.00368924	0.0607
ESSR111	260	1	0.0208	0.00041594	0.0204
ESSR113	177	1	0.0208	0.00041594	0.0204
ESSR113	185	32	0.6667	0.00405093	0.0636
ESSR113	194	14	0.2917	0.00339988	0.0583
ESSR113	204	1	0.0208	0.00041594	0.0204
ESSR114	158	4	0.0833	0.00144676	0.0380
ESSR114	170	2	0.0417	0.00079572	0.0282
ESSR114	173	26	0.5417	0.00513600	0.0717
ESSR114	179	16	0.3333	0.00405093	0.0636

Marker	Allele	Count	Freq	Variance	SD
ESSR115	188	27	0.5625	0.00374349	0.0612
ESSR115	194	17	0.3542	0.00388817	0.0624
ESSR115	196	4	0.0833	0.00231481	0.0481
ESSR116	186	2	0.0417	0.00079572	0.0282
ESSR116	190	14	0.2917	0.00339988	0.0583
ESSR116	192	6	0.1250	0.00195313	0.0442
ESSR116	196	22	0.4583	0.00513600	0.0717
ESSR116	202	3	0.0625	0.00113932	0.0338
ESSR116	209	1	0.0208	0.00041594	0.0204
ESSR117	102	16	0.3333	0.00231481	0.0481
ESSR117	118	5	0.1042	0.00171803	0.0414
ESSR117	143	1	0.0208	0.00041594	0.0204
ESSR117	155	1	0.0208	0.00041594	0.0204
ESSR117	170	25	0.5208	0.00562428	0.0750
ESSR120	102	7	0.1458	0.00302011	0.0550
ESSR120	187	18	0.3750	0.00455729	0.0675
ESSR120	190	23	0.4792	0.00388817	0.0624
ESSR121	198	1	0.0208	0.00041594	0.0204
ESSR121	199	24	0.5000	0.00520833	0.0722
ESSR121	205	6	0.1250	0.00195313	0.0442
ESSR121	208	2	0.0417	0.00079572	0.0282
ESSR121	209	15	0.3125	0.00417752	0.0646
ESSR122	175	25	0.5208	0.00041594	0.0204
ESSR122	216	23	0.4792	0.00041594	0.0204
ESSR124	187	9	0.6429	0.02769679	0.1664
ESSR124	194	5	0.3571	0.02769679	0.1664
ESSR126	175	18	0.3750	0.00368924	0.0607
ESSR126	178	1	0.0208	0.00041594	0.0204
ESSR126	181	17	0.3542	0.00388817	0.0624
ESSR126	194	12	0.2500	0.00347222	0.0589
ESSR127	124	6	0.1250	0.00282118	0.0531
ESSR127	198	42	0.8750	0.00282118	0.0531
ESSR128	124	29	0.6304	0.00493137	0.0702
ESSR128	142	1	0.0217	0.00045204	0.0213
ESSR128	170	3	0.0652	0.00123284	0.0351
ESSR128	198	1	0.0217	0.00045204	0.0213
ESSR128	199	4	0.0870	0.00156160	0.0395
ESSR128	213	1	0.0217	0.00045204	0.0213
ESSR128	226	2	0.0435	0.00180817	0.0425
ESSR128	274	1	0.0217	0.00045204	0.0213
ESSR128	283	3	0.0652	0.00217802	0.0467
ESSR128	319	1	0.0217	0.00045204	0.0213
ESSR129	178	1	0.0208	0.00041594	0.0204
ESSR129	184	17	0.3542	0.00302011	0.0550
ESSR129	187	17	0.3542	0.00736039	0.0858
ESSR129	189	13	0.2708	0.00345414	0.0588
ESSR130	172	1	0.0208	0.00041594	0.0204
ESSR130	174	20	0.4167	0.00405093	0.0636

Marker	Allele	Count	Freq	Variance	SD
ESSR130	177	9	0.1875	0.00244141	0.0494
ESSR130	183	17	0.3542	0.00649233	0.0806
ESSR130	192	1	0.0208	0.00041594	0.0204
ESSR131	192	7	0.1458	0.00215205	0.0464
ESSR131	194	17	0.3542	0.00562428	0.0750
ESSR131	196	11	0.2292	0.00432219	0.0657
ESSR131	197	12	0.2500	0.00347222	0.0589
ESSR131	216	1	0.0208	0.00041594	0.0204
ESSR132	103	27	0.5625	0.00374349	0.0612
ESSR132	140	1	0.0208	0.00041594	0.0204
ESSR132	190	1	0.0208	0.00041594	0.0204
ESSR132	197	3	0.0625	0.00113932	0.0338
ESSR132	201	13	0.2708	0.00258608	0.0509
ESSR132	203	2	0.0417	0.00079572	0.0282
ESSR132	316	1	0.0208	0.00041594	0.0204
ESSR133	179	15	0.3125	0.00244141	0.0494
ESSR133	184	33	0.6875	0.00244141	0.0494
ESSR135	181	2	0.0417	0.00166377	0.0408
ESSR135	216	28	0.5833	0.00665509	0.0816
ESSR135	218	2	0.0417	0.00166377	0.0408
ESSR135	219	3	0.0625	0.00200738	0.0448
ESSR135	227	8	0.1667	0.00231481	0.0481
ESSR135	236	5	0.1042	0.00171803	0.0414
ESSR136	197	5	0.1042	0.00171803	0.0414
ESSR136	201	37	0.7708	0.00345414	0.0588
ESSR136	204	6	0.1250	0.00282118	0.0531
ESSR140	168	20	0.4348	0.01068464	0.1034
ESSR140	170	22	0.4783	0.00895866	0.0947
ESSR140	176	4	0.0870	0.00156160	0.0395
ESSR141	168	1	0.0217	0.00045204	0.0213
ESSR141	171	2	0.0435	0.00086299	0.0294
ESSR141	174	33	0.7174	0.00361634	0.0601
ESSR141	180	4	0.0870	0.00156160	0.0395
ESSR141	184	6	0.1304	0.00304101	0.0551
ESSR142	192	19	0.4318	0.00340440	0.0583
ESSR142	203	15	0.3409	0.00246525	0.0497
ESSR142	207	6	0.1364	0.00225394	0.0475
ESSR142	211	3	0.0682	0.00133828	0.0366
ESSR142	221	1	0.0227	0.00049305	0.0222
ESSR143	195	20	0.4545	0.00300526	0.0548
ESSR143	199	3	0.0682	0.00133828	0.0366
ESSR143	203	13	0.2955	0.00274699	0.0524
ESSR143	206	1	0.0227	0.00049305	0.0222
ESSR143	207	3	0.0682	0.00133828	0.0366
ESSR143	211	3	0.0682	0.00133828	0.0366
ESSR143	221	1	0.0227	0.00049305	0.0222
ESSR145	101	5	0.1316	0.00532148	0.0729
ESSR145	197	13	0.3421	0.00976819	0.0988

Marker	Allele	Count	Freq	Variance	SD
ESSR145	203	17	0.4474	0.01093454	0.1046
ESSR145	211	3	0.0789	0.00174953	0.0418
ESSR146	195	12	0.2727	0.00385049	0.0621
ESSR146	202	23	0.5227	0.00875751	0.0936
ESSR146	205	1	0.0227	0.00049305	0.0222
ESSR146	210	8	0.1818	0.00262960	0.0513
ESSR71	207	8	0.1667	0.00231481	0.0481
ESSR71	210	40	0.8333	0.00231481	0.0481
ESSR72	181	8	0.1905	0.00280747	0.0530
ESSR72	191	9	0.2143	0.00518303	0.0720
ESSR72	214	23	0.5476	0.00442717	0.0665
ESSR72	217	2	0.0476	0.00102581	0.0320
ESSR74	168	1	0.0208	0.00041594	0.0204
ESSR74	180	33	0.6875	0.00244141	0.0494
ESSR74	192	1	0.0208	0.00041594	0.0204
ESSR74	194	2	0.0417	0.00079572	0.0282
ESSR74	197	1	0.0208	0.00041594	0.0204
ESSR74	202	3	0.0625	0.00113932	0.0338
ESSR74	203	2	0.0417	0.00079572	0.0282
ESSR74	209	2	0.0417	0.00079572	0.0282
ESSR74	213	1	0.0208	0.00041594	0.0204
ESSR74	225	1	0.0208	0.00041594	0.0204
ESSR74	232	1	0.0208	0.00041594	0.0204
ESSR75	179	2	0.0417	0.00166377	0.0408
ESSR75	201	12	0.2500	0.00347222	0.0589
ESSR75	203	26	0.5417	0.00426794	0.0653
ESSR75	205	8	0.1667	0.00231481	0.0481
ESSR76	177	21	0.4375	0.00113932	0.0338
ESSR76	180	4	0.0833	0.00231481	0.0481
ESSR76	192	23	0.4792	0.00128400	0.0358
ESSR77	154	6	0.1250	0.00195313	0.0442
ESSR77	164	12	0.2500	0.00260417	0.0510
ESSR77	173	18	0.3750	0.00195313	0.0442
ESSR77	179	1	0.0208	0.00041594	0.0204
ESSR77	182	11	0.2292	0.00258608	0.0509
ESSR78	180	29	0.6042	0.00171803	0.0414
ESSR78	256	19	0.3958	0.00171803	0.0414
ESSR79	132	13	0.3095	0.00394126	0.0628
ESSR79	192	29	0.6905	0.00394126	0.0628
ESSR82	147	24	0.5217	0.00139722	0.0374
ESSR82	157	9	0.1957	0.00353415	0.0594
ESSR82	181	7	0.1522	0.00230131	0.0480
ESSR82	214	1	0.0217	0.00045204	0.0213
ESSR82	222	2	0.0435	0.00086299	0.0294
ESSR82	227	3	0.0652	0.00123284	0.0351
ESSR86	105	37	0.8810	0.00215959	0.0465
ESSR86	130	1	0.0238	0.00053990	0.0232
ESSR86	157	4	0.0952	0.00183565	0.0428

Marker	Allele	Count	Freq	Variance	SD
ESSR87	181	40	0.8333	0.00318287	0.0564
ESSR87	190	8	0.1667	0.00318287	0.0564
ESSR88	185	2	0.0417	0.00166377	0.0408
ESSR88	198	38	0.7917	0.00339988	0.0583
ESSR88	201	5	0.1042	0.00171803	0.0414
ESSR88	204	3	0.0625	0.00113932	0.0338
ESSR91	191	9	0.1875	0.00244141	0.0494
ESSR91	192	5	0.1042	0.00171803	0.0414
ESSR91	196	4	0.0833	0.00144676	0.0380
ESSR91	198	7	0.1458	0.00215205	0.0464
ESSR91	202	5	0.1042	0.00258608	0.0509
ESSR91	204	9	0.1875	0.00330946	0.0575
ESSR91	206	6	0.1250	0.00195313	0.0442
ESSR91	207	3	0.0625	0.00113932	0.0338
ESSR92	163	1	0.0208	0.00041594	0.0204
ESSR92	166	22	0.4583	0.00166377	0.0408
ESSR92	169	1	0.0208	0.00041594	0.0204
ESSR92	172	4	0.0833	0.00144676	0.0380
ESSR92	175	1	0.0208	0.00041594	0.0204
ESSR92	215	19	0.3958	0.00258608	0.0509
ESSR94	186	36	0.7500	0.00347222	0.0589
ESSR94	189	8	0.1667	0.00318287	0.0564
ESSR94	192	4	0.0833	0.00144676	0.0380
ESSR95	197	7	0.1458	0.00215205	0.0464
ESSR95	204	3	0.0625	0.00200738	0.0448
ESSR95	206	24	0.5000	0.00520833	0.0722
ESSR95	213	1	0.0208	0.00041594	0.0204
ESSR95	215	13	0.2708	0.00345414	0.0588
ESSR96	161	3	0.0625	0.00113932	0.0338
ESSR96	166	27	0.5625	0.00461155	0.0679
ESSR96	170	6	0.1250	0.00195313	0.0442
ESSR96	173	2	0.0417	0.00079572	0.0282
ESSR96	176	4	0.0833	0.00144676	0.0380
ESSR96	178	4	0.0833	0.00144676	0.0380
ESSR96	204	2	0.0417	0.00166377	0.0408
ESSR97	167	9	0.1957	0.00258897	0.0509
ESSR97	176	2	0.0435	0.00086299	0.0294
ESSR97	223	35	0.7609	0.00271225	0.0521
ESSR98	182	28	0.6087	0.00562998	0.0750
ESSR98	188	2	0.0435	0.00086299	0.0294
ESSR98	191	2	0.0435	0.00180817	0.0425
ESSR98	194	14	0.3043	0.00447933	0.0669