Genetic Diversity and Population structure analysis of bambara groundnut [*Vigna subterranea* (L.) Verdc.] Landraces using Morpho-agronomic Characters and SSR Markers

ODIRELENG OZIE MOLOSIWA

Thesis submitted to the

University of Nottingham for the degree of

Doctor of Philosophy

March 2012

School of Biosciences

Division of Plant and Crop Sciences

The University of Nottingham Sutton Bonington Campus

Loughborough, Leicestershire

LE12 5RD

UK

Dedication

In memory of my grandmother, Twaambo Molosiwa and my mother Mmaletsatsi Molosiwa who had to go, during the course of my studies.

'We only part to meet again'

Acknowledgements

I would like to express my sincere gratitude to my supervisors, Dr Sean Mayes and Professor Sayed Azam-Ali for their guidance, constructive advice, and assistance given to me during the course of my studies. Many thanks for giving me this opportunity to study at the University of Nottingham. I would also like to thank Dr Erik Murchie for the guidance during my studies.

I highly appreciate the supervision I got from Professor Abu Sesay from Botswana College of Agriculture (BCA) and Dr Sununguko Wata Mpoloka of Biological Science Department (University of Botswana) during field work studies in Botswana. I am also grateful to all the technical assistance and encouragement from Dr Geleta Legesse Fite and Mr Chiyapo Gwafila both from the Department of Agricultural Research (Ministry of Agriculture).

I am very thankful to the technical advice and assistance for laboratory work experiments from Dr Katie Mayes, Dr Shravani Basu, Fiona Wilkinson, Dr GraciaRibas-Vargas, for their sincere help and appreciation. A special thanks to Jim Craigon for his statistical input and interest in my project. Further, EU-BAMLINK project team assistance is acknowledged.

Many thanks from Mark Meacham, John Alcock, Matthew Tovey, and David Hodson for their support in glasshouse and growth room experiments. In addition the help from Emma Hooley, Sonoko Mitsui-Angwin, Diane Jones, Sue Flint, Sue Golds and Linda Staniford is highly appreciated.

I am grateful and thankful to my colleagues and friends we share a journey in our quest for knowledge, Stanley Noah, Endah Sri Redjeki, Nariman Ahmad, Presidor Kendabie, Chai Hui Hui, Kidist Bogale Kibret, Ghaliya Al-Mamari, Upendra Kumar Devisetty, Chanate Malumpong and Steve Hayes

I am grateful to the Commonwealth Scholarship and Fellowship Plan (CSFP) for awarding me this scholarship, and the Botswana Government (Ministry of Agriculture) to award me a study leave.

Finally, I thank my ‘big family’ of the Nazareth Church, my friends and my family especially, Makgopho Nthebolang, Boemo Molosiwa and Moagi Molosiwa for their kindness and support.

Abstract

Bambara groundnut is an indigenous African legume grown mainly in sub-Saharan Africa; it is an important source of protein to the rural majority. There are no established varieties and subsistence farmers grow locally adapted landraces which are generally low yielding. Bambara groundnut is a predominantly self-pollinating crop and is expected to exist as non-identical inbred lines, although the previous lack of co-dominant markers has prevented a formal assessment of heterozygosity within bambara groundnut genotypes.

A total set of 75 microsatellites that were characterised in this study were used to investigate the genetic diversity of a set of 24 bambara groundnut landraces, to provide an evaluation of the markers for polymorphism and provide a link with DArT marker data that were previously analysed.

Sixty eight microsatellites were identified that were found to be consistent and reproducible, from which a set of markers were selected and used for genetic variability studies of bambara groundnut, to compare the use of molecular markers with morphological markers, and to investigate using SSR markers in pure line selection.

The genetic diversity of bambara groundnut was assessed based on morphological characters for two seasons; in a glasshouse experiment at the University of Nottingham, Sutton Bonington Campus, UK and in a field experiment that was conducted at the Botswana College of Agriculture (Notwane farm), Gaborone in a randomised block design with three replicates. The landraces were characterised for 24 quantitative and 13 qualitative characters. The results indicated considerable variation for quantitative characters, while significant morphological differences were also recorded for most characters. Multivariate data analysis was conducted using principal component analysis, cluster analysis and heritability estimates were developed. The low cost, simplicity and agricultural relevance of morphological characterisation makes it an important tool in germplasm genetic variation studies.

Thirty four lines from field experiments were investigated for genetic diversity based on 20 microsatellites. The expected heterozygosity (He) had an average of

1in agreement with the fact that bambara groundnut is predominantly self-pollinating. Both cluster analysis and principle component analysis (PCoA) grouped landraces based mainly on their areas of origin.

A thorough molecular analysis of genetic and morphological variation in bambara groundnut was conducted to investigate the relationship between the two assessment techniques. This comparison will assist in breeders making informed decisions as to which approach is best to use in germplasm characterisation and plant breeding and how best to apply such knowledge in practical situations. DNA markers could then aid with the selection of germplasm for breeding, quality control within breeding programmes and, potentially, direct selection via Marker Assisted Selection (MAS). Euclidean distance estimates for morphological data and (Nei's 1972) genetic distance estimates for SSR data were strongly correlated ($r = 0.7$; $P < 0.001$) in the agronomy bay and ($r = 0.6$; $P< 0.001$) in the controlled growth room. These results suggest the two approaches are generating the same pattern of genetic diversity, and as such can be used as a surrogate for each other.

Table of Contents

Dedication ..... **ii**

Abstract ..... **iv**

List of Figures ..... **xvi**

List of Tables ..... **xx**

Abbreviations ..... **xxiv**

CHAPTER ONE: Introduction **1**

1.1 Leguminous crops 1

1.2 Bambara groundnut taxonomy 4

1.3 The bambara groundnut crop 6

1.4 Potential of bambara groundnut 7

1.4.1 Some interesting agronomic characters of the crop. 8

1.4.2 Bambara groundnut yield potential 8

1.4.3. Some uses of bambara groundnut 9

1.4.4 Genetic diversity resources 9

1.4.5. Potential areas of expansion 9

1.5 Constraint to bambara groundnut production 10

1.5.1 Influence of sowing date/photoperiod 10

1.5.2 Low moisture, pests and diseases 11

1.5.3 Anti-nutritional factors in bambara groundnut 11

1.5.4 Genetic resources 12

1.5.5 Mating systems 12

1.6 Diversity evaluation using both morphological and molecular markers 12

1.6.1 Genetic markers 12

1.6.2 Morpho-agronomic markers 13
1.6.3 Biochemical markers 15
1.6.4 Molecular Markers 16
1.6.4.1 Hybridisation (Sequence dependent) 16
1.6.4.1.1 Random Amplified Fragment Polymorphism (RFLP) 16
1.6.4.1.2 Diversity Array Technology (DArT) 17
1.6.5 PCR-based molecular markers 19
1.6.5.1 Random Amplified Polymorphic DNA (RAPD) 19
1.6.5.2 Amplified Fragment Length Polymorphism (AFLP) 20
1.6.5.3 Microsatellites: Simple Sequence Repeats (SSR) 21
1.7 Microsatellites development and application 22
1.8 Potential application of microsatellites in bambara groundnut 23
1.8.1 Conservation of genetic resources 23
1.8.2 Molecular mappings 25
1.8.3 Marker Assisted Selection and QTL 25
1.9 Comparison of genetic diversity estimates methods 28
1.9.1 The objectives of the study 30
CHAPTER TWO: Materials and Methods 31
2.1 Introduction 31
2.1.1. Standard solutions 31
2.1.2 List of plant materials 32
2.1.3 Overview of experiments. 35
2.2 Methodology for Marker and DNA techniques 37
2.2.1 Introduction 37

2.2.2 Plant materials 37
2.2.3 DNA extraction 38
2.2.3.1 Sigma DNA extraction Kit 38
2.2.4 DNA quantitation 39
2.2.5 Microsatellite development 39
2.2.6 PCR gradient optimisation for primer annealing temperature 40
2.2.7 Gel electrophoresis of PCR products 41
2.2.8 Three primer systems 41
2.2.9 PCR amplification of microsatellites 42
2.3.0 Gel electrophoresis of PCR products of Tagged primers 43
2.3.1 Capillary electrophoresis 43
2.3.2 Analysis of microsatellites 44
2.4 Potential genotyping errors and some mitigating strategies in microsatellite analysis 44
2.4.1 DNA degradation 44
2.4.2 PCR based sources of error 44
2.4.3 Interpretation of capillary electrophoresis 45
2.4.4 Spectral overlap 45
2.4.5 Stutter and A-addition 47
2.4.6 Short allele dominance 49
2.4.7 Allele size binning (Automated binning) 49
2.4.8 Deviation from Hardy Weinberg equilibrium 54
2.5 Data analysis 55
2.5.1 Data analysis for microsatellites, development and characterisation 55

2.5.1.1 Microsatellites marker analysis..... 55

2.5.1.2 Principal component analysis (PCO)..... 55

2.5.1.3 Cluster analysis..... 55

2.5.1.4 Comparison of DArT and SSR genetic estimates..... 56

2.5.2 Population structure and genetic diversity of bambara groundnut..... 56

2.5.2.1 Estimation of genetic diversity in the population..... 56

2.5.2.2 Estimation of genetic diversity within and among bambara groundnutpopulations..... 56

2.5.2.3 Estimation of population structure..... 57

2.5.3 Genetic diversity of bambara groundnut based on SSR markers and the comparison with morpho-agronomic characters..... 58

2.5.3.1 Polymorphism of microsatellites in bambara groundnut..... 58

2.5.3.2 Principal component (PCO) and cluster analysis..... 58

2.5.3.3 Analysis of Molecular Variance (AMOVA)..... 58

2.5.3.4 Morphological data analysis..... 58

2.5.3.5 Comparison of SSR marker and morphological marker data..... 59

2.6 Morpho-agronomic characterisation and evaluation of bambara groundnut..... 59

2.6.1 Introduction:..... 59

2.6.2 Glasshouse experiment..... 59

2.6.3 Plant materials..... 60

2.6.4 Experimental design..... 60

2.6.5 Crop management..... 60

2.6.6 Morpho-agronomic traits measurements collected in the greenhouse..... 63

2.6.6.1 Quantitative traits measurements in the green house..... 64

2.6.6.2 Qualitative traits measurements in the glasshouse 64

2.7 Field work experiment in Botswana 66

2.7.1 Introduction 66

2.7.2 Field site and experimental preparation 66

2.7.3 Plant material 66

2.7.4 Experimental design 66

2.7.5 Crop management 67

2.7.6 Agro-morphological traits measurements in the field experiment 70

2.7.7 Statistical analysis of agronomic traits 70

2.7.8 Data analysis of agronomic traits 70

2.7.8.1 Descriptive characteristics 70

2.7.8.2 Principal component analysis 71

2.7.8.3 Cluster analysis 71

2.7.8.4 Shannon-Weaver diversity 72

2.7.8.5 Correlation coefficient 72

2.7.8.6 Quantitative variances 72

2.7.8.7 Selection index (SI) and Duncan Multiple Range Test (DMRT) 74

CHAPTER THREE: Microsatellites, development and characterisation 76

3.1 Introduction 76

3.1.1 Breeding systems in bambara groundnut 76

3.1.2 Floral biology of bambara groundnut 77

3.1.3 Seed dissemination systems 78

3.1.4 Analysis of breeding systems in bambara groundnut 78

3.1.5 Breeding system studies in other leguminous species 79

3.1.6 Applications of microsatellites in this study 80

3.2 Materials and Methods 81

3.2.1 DArT marker screening 81

3.3 Results 83

3.3.1 Microsatellites marker analysis 83

3.3.1.1 Hardy Weinberg Equilibrium (HWE) 85

3.3.1.2 Estimation of Null alleles 87

3.3.2 Principal Component Analysis (PCO) 90

3.3.3 Cluster analysis 92

3.4 Discussions: 97

3.5 Conclusions 99

CHAPTER FOUR: Phenotypic diversity for morphological and agronomic characters of bambara groundnut 101

4.1 Introduction 101

4.1.1 Correlation analysis studies 102

4.1.2. Selection of lines for breeding 106

4.1.3 The objectives of this study 107

4.2 Results 108

4.2.1 Qualitative analysis of the genotypes 108

4.3.2 Shannon Weaver (H') diversity analysis 111

4.3.3 Descriptive analysis of the genotypes 113

4.3.4 Principal component analysis: 117

4.3.5 Cluster analysis 122

4.3.6 Correlation coefficients among traits 125

4.3.7 Quantitative variance analysis 131

4.3.8 Comparison of agronomy bay and field experiment..... 135

4.3.8 Selection for breeding bambara groundnut..... 137

4.4 Discussions 140

4.5 Conclusions 145

CHAPTER FIVE: Population structure and genetic diversity of bambara groundnut..... 146

5.1 Introduction 146

5.1.1 Genetic diversity in bambara groundnut..... 147

5.1.2 Genetic diversity and population structure of other legumes..... 147

5.1.3The objectives of the study 149

5.2 Materials and Methods 149

5.2.1 Phenotypic data analysis..... 149

5.3 Results..... 150

5.3.1 Genetic diversity analysis 150

5.3.2 Genetic diversity within and among regions..... 151

5.3.3 Principal coordinates analysis (PCoA) 152

5.3.4 Cluster analysis..... 153

5.3.5 Genetic differentiation based on F_{ST}..... 155

5.3.5.1 Pairwise comparison..... 155

5.3.6 Analysis of Molecular Variance Analysis (AMOVA)..... 156

5.3.7 Comparison of molecular markers with pod and seed characters..... 157

5.4 Discussion..... 162

5.5 Conclusions 165

CHAPTER SIX: Genetic diversity of bambara groundnut based on SSR markers and the comparison with morpho-agronomic characters 166

6.1 Introduction 166

6.1.1 Genetic diversity of bambara groundnut..... 167

6.1.2 Genetic diversity in other leguminous crops..... 167

6.1.3 Efficiency of molecular and morphological markers in genetic diversity estimates 169

6.2 Materials and methods..... 171

6.2.1 Plant Materials used..... 171

6.3 Results..... 171

6.3.1 Polymorphism of microsatellites in bambara groundnut 171

6.3.2 Principal Component Analysis (PCO) 176

6.3.3 Comparison of SSR and morphological markers..... 178

6.3.3.1 Principal Component Analysis 178

6.3.4 Genetic distance estimates between landraces..... 183

6.3.5 Correlation between molecular and morphological distance estimates 183

6.3.6 Molecular variance among bambara groundnut landraces..... 185

6.3.7 Breeding strategy 186

6.4 Discussion..... 189

6.5 Conclusions 193

CHAPTER SEVEN: General discussions 194

7.1 Introduction 194

7.2 Recap of the study 196

7.3 Microsatellites development and characterisation 197

7.4 Morphological characterisation 199

7.5 Genetic diversity based on SSR markers and a comparison withmorphological characters 203

7.6 Population structure analysis 208

7.7 Impact of the findings and future work 210

7.8 Future work 211

REFERENCE 212

APPENDICES 242

Appendix 1: Preparations of standard solutions 242

Appendix 2: List of characterised 75 primers used in bambara groundnut diversity, and list of primer combinations used in multiple experiments. 243

Appendix 3: Estimated repeat length of alleles and adjustment for the characterisation of 75 markers used in the analysis of 24 landraces 245

Appendix 4: A comparison of Nei and Li, (1979) similarity estimates for DArT marker (upper) and SSR markers (bottom) matrices calculated using MVSP version 3.1 for the 24 bambara groundnut landraces. 247

Appendix 5: Mean values for the characters of the 35 landraces grown in the agronomy bay experiment (UK) 248

Appendix 6: Mean values for the characters of the 34 lines grown in the field experiment (Botswana). 250

Appendix 7: Range of classes for the quantitative traits used for both the glasshouse and the field experiment 252

Appendix 8: Hardy Weinberg Equilibrium (HWE) and the exact p-values estimated using PowerMarker (Version 3.25) 253

Appendix 9: Cluster analysis, genetic similarity among the 105 bambara groundnut genotypes, analysis using 141 variables and 105samples/cases 254

Appendix 10: Scatter plots for morpho-agronomic markers on (Euclidean distance estimates) and molecular markers on (Nei's 1972) conducted using Mantel's test on

NTSYS, Pearson correlation and Spearman's rank correlations on SPSS in the Agronomy bay and controlled growth room experiment: Appendix 10.1 and 10.2. .. 257

Appendix 11: Mean for the phenotypic measures of 5 lines from the controlled growth room experiment 259

List of Figures

Figure 1.1: Phylogenetic trees showing the relationship between Vigna species from various Vigna subgenus and section. Adapted from (Wang *et al.*, 2008).....3

Figure 1.2: A typical bambara groundnut crop in the field, unshelled pods, flowering and pod initiation, and bambara groundnut seeds.....6

Figure 2.0: Diagrammatic representation of the setup of the project.....36

Figure 2.4.4: A pair of capillary electrophoresis traces of PCR products for blue labelled. The genuinely labelled blue PCR products, has bled through into spectrum of the green labelled PCR product and false peaks are shown for sample M18 and M19 (green).....46

Figure 2.4.5: Capillary electrophoresis showing a potential scoring error due to the effects of stutter band and overlap on sample PR 45-H. H11 and 44.E05.....48

Figure 2.4.6: Capillary electrophoresis showing limited short allele dominance for marker PR 15 top and marker PR 42 bottom, since both are clearly visible and complete drop out did not occur, correct calling of the peaks could be done…....49

Figure 2.4.7: Capillary electrophoresis showing potential sources of mistyping errors due to rounding off alleles during binning.....51

Figure 2.4.8: Capillary electrophoresis showing some potential miscalling errors, therefore the use of allele shapes, their height and size ranges are set as standard way to identify genuine peaks for correct allele calling.....52

Figure 2.4.9: A graphical output of the cumulative allele length for marker 16, it illustrates an example of an accurately binned marker with clearly defined colours for different alleles as red and blue. The analysis was conducted with FLEXIBIN (Automated binning) using a one unit repeat.....53

Figure 2.6.5.1: The maximum and minimum temperature in the agronomy bay (Glasshouse) experiment for the 119 bambara groundnut landraces grown in the 2008 season days from sowing.....62

Figure 2.6.5.2: The maximum and minimum relative humidity in the agronomy bay (Glasshouse) experiment for the 119 bambara groundnut landraces grown in the 2008 season.....63

Figure 2.7.5.1: The amount and distribution of rainfall in bambara groundnut field experiment at (Notwane) Sebele, in 2008- 2009 season.....68

Figure 2.7.5.2: Maximum and minimum temperature in the field experiment for the 34 bambara groundnut landraces grown at (Notwane) Sebele in the 2008 - 2009 season.....69

Figure 2.7.5.3: Maximum and minimum relative humidity in the field experiment for the 34 bambara groundnut landraces grown at (Notwane) Sebele in the 2008 - 2009 season.....69

Figure 3.1.2 Bambara groundnut flower, showing the floral morphology.....77

Figure 3.2: The first two axes of the PCO case scores, generated from the 24 landraces using MVSP for figure 3 (a) DArT Axis 1 represents 28.45% and Axis 2 represents 8.84 % of the molecular variation, figure 3 (b) SSR markers; Axis 1 represents 10.91 % and Axis 2 represents 8.61% of the molecular variation in the 24 selected bambara groundnut landraces.....91

Figure 3.3:(a) Cluster analysis of the 24 bambara groundnut landraces. The UPGMA dendrogram is based on the similarity matrix obtained from 201 DArT markers using the Nei and Li, (1979). The number at the nodes of branches represents the percentage bootstrap support of individual nodes at resampling at 1000.....93

Figure 3.3:(b) Cluster analysis based on the 24 bambara groundnut landraces, the dendrogram was obtained based on 68 SSR markers, the UPGMA tree is based on the Nei and Li, (1979) similarity coefficient. The number at the nodes of branches represents the percentage bootstrap support of individual nodes at resampling at 1000.....93

Figure 3.4: (a) A scatter plot produced based on the matrix for DArT and SSR markers genetic distance estimates from Nei and Li, 1979 (Appendix 4) using Pearson product-moment coefficient correlation based on SPSS version 16 (b) A scatter plot based on the matrix for DArT and SSR produced from the same genetic distance estimates in Appendix 4 using Mantel-matrix correspondence test on NTSYS pc version 2.1 program (MXCOMP module) based on 1000 permutation.....96

Figure 4.2.2: Dendrogram of 35 bambara groundnut landraces showing a (UPGMA) Euclidean cluster analysis based on 34 agro-morphological markers in glasshouse experiment. The colour code for West Africa = Green, Southern Africa =Red, East Africa =Yellow, Central Africa = Blue, Indonesia = Purple. The number at the nodes of branches represents the percentage bootstrap support of individual nodes resampling at 1000……....123

Figure 4.2.3: Dendrogram of 34 bambara groundnut lines showing genetic similarities based on 37 morpho-agronomic traits 24 quantitative traits and 13 qualitative traits, using the UPGMA cluster analysis (field experiment Botswana). The number at the nodes of branches represents the percentage bootstrap support of individual nodes resampling at 1000……...124

Figure 4.2.4: A regression analysis plot of mean over all the genotypes for the 24 variables recorded from agronomy bay (UK) and field experiment in Botswana.....136

Figure 5.1.0 A PCO scatter plot for the 123 bambara groundnut genotypes from Africa and Indonesia generated from 12 microsatellites with MVSP program

with a molecular variation of 16.15 %, with axis 1 contributing (9.87%) and while Axis 2 explained (6.28 %). The two cluster groups were hand drawn on Microsoft Word…….....153

Figure 5.2.1: Cluster analysis of bambara groundnut landraces from five regions, from Africa and Indonesia (Asia). The dendrogram is based on 12 SSR markers. The Unweighted pair group method with arithmetic averages (UPGMA) tree was based on Nei and Li's coefficient of genetic similarity generated from the presence/absence binary matrix on 123 bambara groundnut landrace accessions.....154

Figure 5.3.0: A PCO scatter plot for the 87 bambara groundnut that produced pods and seeds among 119 bambara groundnut planted. The data is based on 7 pod and 8 seed characters analysed using the MVSP program. The percentage variation for Axis 1 represents 33.59% and the Axis 2 represent 16.8 % with a cumulative percentage of 49.87% for the first two Axes.....158

Figure 5.4.0: A PCO scatter plot on case scores for the 87 bambara groundnut that produced pods and seed based on 12 microsatellites, generated on MVSP program. The cumulative percentage of variation explained for the first two Axes is 16.08 %, Axis 1 contributes 9.45% and Axis 2 contributes 6.63%.....159

Figure 5.5.0: Scatter plot of correlation for morphological marker genetic distances estimate based on standard Euclidean and SSR marker genetic distance based on Nei's 1972, the analysis was conducted on (a) Mantel test correspondence test on NTSYS and (b) on Pearson correlation on SPSS version 16……..160

Figure 6.1.1: UPGMA dendrogram of 105 bambara groundnut genotypes revealed by UPGMA cluster analysis of 20 SSR markers based on Nei and Li, 1979 similarity estimate. Bootstrap values of 1000 replications more than 50% are shown on corresponding nodes.....175

Figure 6.2.1: The first two axes of the PCO case scores, generated from the 105 bambara groundnut genotypes based on 20 SSR markers generated on MVSP, the first Axis accounts for 8.69 % while Axis 2 represent 6.27 % and together explain a cumulative 14.95 % of the molecular variation. Figure a: shows a PCO plot demarcated on 5 symbols to identify three individuals from one landrace while Figure b: shows the grouping of the five regions into two major groups. The two cluster groups were hand drawn on Microsoft Word…….....…..177

Figure 6.3.1: The first two axes of the PCO case scores, generated from the 34 bambara groundnut landraces using MVSP for figure 6.3.1 (a) SSR marker Axis 1 represent 9.90 % and Axis 2 represent 8.35 %, figure 6.3.1 (b) Morphology marker; Axis 1 represent 13.56 % and Axis 2 represent 8.80 % molecular variation with a cumulative % of 18.25 % and 22.36 % respectively.....179

Figure 6.4.1: Cluster analysis of 34 bambara groundnut analysis with Unweighted pair group method with arithmetic method (UPGMA) were generated using NTSYS version 2.1, Figure 6.4.1 (a) is SSR marker dendrogram generated from 20 microsatellites markers based on Nei's 1972 distance estimates, Figure (b) is a

morphology dendrogram generated on 37 morpho-agronomic traits generated on Euclidean distance estimates.....182

Figure 6.5.1:A scatter plot of correlation for morpho-agronomic and molecular marker based on Pearson, Spearman (rank) and Mantel test, analysis conducted on (A) NTSYS pcversion 2.1 and (B) on SPSS version 16, the morphological markers were based on Euclidean distances estimates while the molecular marker were on Nei's 1972 coefficient……...184

Figure 6.5.2: Schematic diagram showing the selection strategy for the three round of selection of bambara groundnut.....187

List of Tables

Table 1.1: Summary of Vigna classifications based on 6 sub-genera, and some examples from each section, adapted from African Vigna.....2

Table 2.1.2.1: List of selected landraces used for the characterisation of SSR markers and DArT analysis, their areas of origin and the clusters where the landraces were selected. The selection was on the basis of a study conducted by Singrün and Schenkel (2003), where a total of 223 bambara groundnut landraces were analysed for genetic diversity using enzyme system *Eco*Ri/*Mse*I amplified fragment length polymorphism (AFLP).....32

Table 2.1.2.2:A list of 123 bambara groundnut accessions, source and their areas of origin used in the experiment; 105 bambara groundnut accessions selected in the greenhouse (35 x 3) samples, and 34 accessions that were selected and planted in the field experiment in (Botswana).....33

Table 2.3.1: A summary of Flexibin analysis for marker 16, showing repeat length, standard deviation and count of each repeat length.....54

Table 2.6.1: Amount of irrigation water (mm) applied in the bambara groundnut experiment in the agronomy bay (Glasshouse) expressed in days after sowing (DAS) for the duration of the experiment in 2008 season.....61

Table 2.6.6: Quantitative and qualitative traits recorded and brief description as listed from (IPGRI, 2000).....65

Table 2.7.5: Amount of irrigation water (mm) applied in the bambara groundnut experiment in the field experiment in Botswana, expressed in days after sowing (DAS) for the duration of the experiment in the 2008/ 2009 season.....67

Table 3.1: A summary of PowerMarker data analysis of 24 bambara groundnut landraces, based on 68 microsatellites.....83

Table 3.2: The 68 markers used in the 24 bambara groundnut analysis were subjected to Chi square and HWE exact test using MVSP version 3.25, with the exception of nine non-polymorphic markers.....86

Table 3.3: Estimation of null allele frequencies for each locus, using the population inbreeding model (PIM) and the individual inbreeding model (IIM) using INEst (Chybicki and Burczyk, 2009).....88

Table 3.4: PCO case scores for the population structure of the selected 24 bambra groundnut landraces, determined based on 201 DArT markers.....90

Table 3.5: PCO case scores for the population structure of the selected 24 bambra groundnut landraces, determined based on 65 SSR markers.....90

Table 3.6: A comparison of the distribution of the 24 bambara groundut landraces based on the UPGMA clustering analysis done using a set of 201 DArT markers and 65 SSR markers.....94

Table 3.7 Pearson, Spearman and Mantel test correlations between the genetic similarity matrices based on the two markers systems (DArT vs SSR).....95

Table 4.1.1: A comparison of correlations between yield components; seed yield per plant, number of pods per plant, seed yield per hectare and 100 seed weight and a number of characters, sourced from Karikari and Tabona, (2004); Misangu *et al.,* (2007); Ouedraogo *et al.,* (2008); Goli *et al.,* (1995); Jonah *et al.,* (2010); Karikari, (2000), and (Oyiga and Uguru, 2011).....103

Table 4.1.2: Descriptor, classes and frequency distribution among the 35 landraces planted in the agronomy bay and 34bambara groundnut lines selected and planted in the field in Botswana.....109

Table 4.1.3: Shannon-Weaver index on the phenotypic diversity of 24 quantitative characters in the agronomy bay experiment and the field experiment (Botswana).....111

Table 4.1.4: Shannon weaver index on phenotypic diversity of qualitative characters for the studied landraces in agronomy bay and field experiment.....112

Table 4.1.5: Descriptive characteristics for the 35 bambara groundnut planted UK (Agronomy bay, 2008) from three replications.....114

Table 4.1.6: Descriptive characteristics for the 34 bambara groundnut planted in field (Notwane, Botswana, 2008/2009 season) with three replications, derived from seed from single plant from the agronomy bay experiment. These exclude 1 landrace which had few seeds.....115

Table 4.1.7: Principal components, matrix of eigenvalues and vectors for 24 quantitative characters of bambara groundnut landraces planted in the agronomy bay (UK).....119

Table 4.1.8: Principal component, matrix of eigenvalues and vectors for 24 quantitative characters of bambara groundnut lines planted in Botswana.....121

Table 4.1.9: Correlation coefficients among the 24 traits based on the 35 bambara groundnut planted in the Agronomy bay (UK), traits were measured 10 weeks after planting.....127

Table 4.2.1: Correlation coefficient for 24 quantitative traits of the 34 bambara groundnut planted in the field experiment in (Botswana) traits were measured 10 weeks after planting.....129

Table 4.2.2: Quantitative variances based on phenotypic coefficient of variability (PCV), genotypic coefficient of variability (GCV), broad sense heritability (h^2B) and genetic advance (GA) in the 35 landraces in the agronomy bay (UK).....132

Table 4.2.3: Quantitative variances based on phenotypic coefficient of variability (PCV), genotypic coefficient of variability (GCV), broad sense heritability (h^2B) and genetic advance (GA) in the 34 lines (Field experiment).....133

Table 4.2.4: A summary of analysis for the relationship between the agronomy (UK) experiment and the field experiment in (Botswana), computed on Genstat version 13.0.....136

Table 4.2.5: The Duncan multiple range test and the selection index of bambara groundnut based on the vegetative and yield characters (Agronomy bay, UK)..138

Table 4.2.6: The Duncan multiple range test and the selection index of bambara groundnut based on the vegetative and yield (field experiment in Botswana).....139

Table 5.1: PowerMaker summary data analysis for the 12 microsatellites used amplified from 123 bambara groundnut landraces (118 from Africa and 5 from Asia/Indonesia).....150

Table 5.2: A comparison of the genetic diversity estimates for the among the five regions of Africa and Asia (Indonesia) analysis conducted using FSTAT 2.9.3 for all the 123 bambara groundnut landraces.....151

Table 5.3: Principal Coordinate analysis (PCoA) from the investigation of population structure of 118 bambara groundnut landraces collected from Africa and 5 from Indonesia based on MVSP program.....152

Table 5.4: Genetic differentiation of the 123 bambara groundnut landraces from 4 regions of Africa and also Asia (Indonesia), estimated using Weir and Cockerham (1984) on Genepop version 4.0.....155

Table 5.5: Pairwise genetic distance based on F_{ST} values between populations, calculated on 12 microsatellites based on five regions of Africa including Asia (Indonesia).....156

Table 5.6: Analysis of Molecular Variance for the 123 bambara groundnut landraces based on 12 SSR markers using Arlequin version 3.1.....157

Table 5.7: Principal Coordinate analysis (PCoA) for 15 characters of pods and seeds for 87 landraces that set reasonable seed numbers among the 119 landraces planted in the agronomy bay for bambara groundnut germplasm characterisation.....157

Table 5.8: Principal Coordinate Analysis (PCoA, Euclidean) for 87 bambara groundnut landraces that set seed, based on 12 microsatellites.....158

Table 5.9: Correlation of molecular marker distance matrices, based on Pearson correlation, Spearman rank correlation and Mantel test for the 12 qualitative character and 12 molecular markers.....161

Table 6.1: Summary of PowerMarker data analysis for the 35 bambara groundnut landraces using 20 microsatellites analysis conducted on each of the 105 individual genotype.....172

Table 6.2: Intra-landrace diversity among the 35 genotypes conducted on each of three genotypes per landrace using 20 SSR markers based on Arlequin version 3.1.....173

Table 6.3: PCO case scores for the population structure of the 105 genotypes determined from each of the three samples of the 35 bambara groundnut landraces based on 20 SSR markers.....176

Table 6.4: PCO case scores for the population structure of the 34 bambara groundnut selected for field studies in Botswana, analyses based on 20 SSR markers.....178

Table 6.5: PCO case scores for the population structure of the 34 bambara groundnut based on 37 morpho-agronomic characters, from the field experiment conducted in Botswana.....178

Table 6.6: Correlation between morpho-agronomic markers and molecular markers for the 35 and 34 bambara groundnut genotypes based on 20 microsatellites and 37 morph-agronomic characters, and for 5 lines based on 12 markers and 22 morpho-agronomic characters.....185

Table 6.7: Analysis of molecular variance (AMOVA) for the 105 bambara groundnut genotypes for the comparison based on the five selected regions, analysis using Arlequin version 3.5.....186

Table 6.8: Mean and range of the genetic distances values for three different selection cycles of bambara groundnut from single seed descent estimated based on 12 microsatellites markers using Popgene version1.31 (Yeh and Boyle, 1997).....188

Abbreviations

ANOVA	Analysis of Variance
AMOVA	Analysis of Molecular Variation
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleotide Triphosphates (usually mix of dATP/dTTP/dCTP/dGTP)
dATP	deoxyadenosine Triphosphate
dTTP	deoxythymidine Triphosphate
dCTP	deoxycytidine Triphosphate
dGTP	deoxyguanine Triphosphate
DAR	Department of Agricultural Research
cM	centiMorgan
CV	Coefficient of variation
DAS	Days after sowing
EDTA	ethylene diamine Tetracetic Acid
f	Inbreeding coefficient
F_{IS}	Inbreeding coefficient
F_{IT}	Overall fixation index
F_{ST}	Fixation index
H`	Gene diversity
GA	Genotypic advance
GVC	Genotypic coefficient of variation
G_{ST}	Nei's total fixation index
He	Expected heterozygosity
Ho	Observed heterozygosity
HWE	Hardy-Weinberg equilibrium
IBPGR	International Board for Plant Genetic Resources

IITA	International Institute of Tropical Agriculture
ISSR	Inter Simple Sequence Repeat
MAS	Marker Assisted Selection
MVSP	Multivariate Statistical Package
NTSYS	Numerical Taxonomy and Multivariate Analysis System
PCoA	Principle coordinates analysis
PCV	Phenotypic coefficient of variation
PCO	Principal component analysis
PIC	Polymorphic information content
PCR	Polymerase chain reaction
RAPD	Randomly Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
SS	Sizestandard
SLS	Sample loading solutions
SSR	Simple Sequence Repeat
UPGMA	Unweighted pair group method with arithmetic means

CHAPTER ONE: Introduction

1.1 Leguminous crops

The genus *Vigna* is a member of the family Leguminosae (= Fabaceae), subfamily Papilionoideae, tribe Phaseoleae. Leguminosae are morphologically diverse and include a number of trees and some aquatic plants such as in the genus *Neptunia*in the subfamily Mimosoideae (Polhill and Raven, 1981) which consist of a number of species that are aquatic. It is the third largest family of flowering plants behind orchids (Orchidaceae) and asters (Asteraceae) and consists of approximately 650 genera and 18,000 species (Polhill *et al*., 1981; Doyle and Luckow, 2003). In terms of agricultural importance Leguminosae comes second to cereals. The Leguminosae have been divided into three major groups mainly on the basis of their morphological and floral differences, that is the Caesalpinioideae, Mimosoideae and Papilionoidaeae (Doyle and Luckow, 2003).Papilionoideae with approximately 70% of the Leguminosae species is the largest subfamily, it includes most of the crops and major model legume species (Doyle and Luckow, 2003; Cannon *et al.,* 2009) it is subdivided into four large groups the galegoid, millettioids, dalbergioids and genistoids (Doyle and Luckow, 2003).

The galegoid contains the robinioid clade with several forages and trees (*Sesbania* and *Robinia*); it also consists of inverted-repeat-loss clade (IRLC) which includes clovers (*Trifolium* spp.), vetch (*vicia* spp.), pea(*Pisum sativum*), chickpea (*Cicer arietinum*), lentil (*Lens culinaris*) and alfalfa(*Medicago sativum*) (Doyle and Luckow, 2003).The milletioid clade consists of common bean (*Phaseolus vulgaris*), soybean (*Glycine max*), cowpeas (*Vigna unguiculata*), pigeonpea (*Cajanus cajan*), mungbean (*Vigna radiata*), adzuki bean (*Vigna angularis*), tepary bean (*Phaseolus acutifolius*), lima bean (*Phaseolus lunatus*), hyacinth bean (*Lablab purpureus*), bambara groundnut (*Vigna subterranea*) and hausa groundnut (*Macrotyloma geocarpum*). The dalbergioid clade consists of a number of tropical trees (eg *Dalbergis* spp) and peanut (*Arachis hypogaea*).The genistoid contains many tropical and temperate genera for example the lupins (*Lupinus* spp) (Cannon *et al.,* 2009).

Vigna consists of approximately 80 species that are grouped into six subgenera: *Vigna*, *Ceratotropis*, *Plectotropis*,*Sigmoidotropis*,*Lasiosporon* and *Haydonia*.

Subgenus *Vigna* comprises 39 species, and it includes some important agricultural species such as, Cowpea (*Vigna unguiculata* L. Walp), bambara groundnut (*Vigna subterranea* L. Verdc) and mungbean (*Vigna radiata*) (Goel *et al.*, 2002; Vijaykumar *et al*, 2009). These species are of considerable importance in many developing countries with cowpea and bambara groundnut from Africa, while mungbean is from Asia (Smartt, 1985; Doi *et al.*, 2002).

Table 1.1: Summary of Vigna classifications based on 6 sub-genera, and some examples from each sectionadaptedfrom (African Vigna, bioversityinternational.org)

Subgenus	Section	Specie
Vigna	Vigna	*V. subterranea*
	Comosae	*V. comosa , V. haumaniana*
	Macrodontae	*V. somaliensis*
	Reticulatae	*V. reticulata*
	Liebrechtsia	*V. frutescens*
	Catiang	*V. anguiculata*
Haydonia	Haydonia	*V.monophylla*
	Microspermae	*V.microsperma*
	Glossostylus	*V. nigritia*
Plectotropis	Plectotropis	*V. vexillata*
	Pseudoliebrechtsia	*V. nuda*
Ceratotropis	Ceratotropis	*V. mungo ,V. radiata*
	Aconitifoliae	*V. aconitifolia*
	Angulares	*V. angularis*
Lasiospron	Lasiospron	*V. longifolia*
Sigmoidotropis	Sigmoidotropis	*V. elegans*
	Pedunculares	*V. peduncularis*
	Caracallae	*V. caracalla*
	Condylostylis	*V. venusta*
	Leptospron	*V. adenantha*

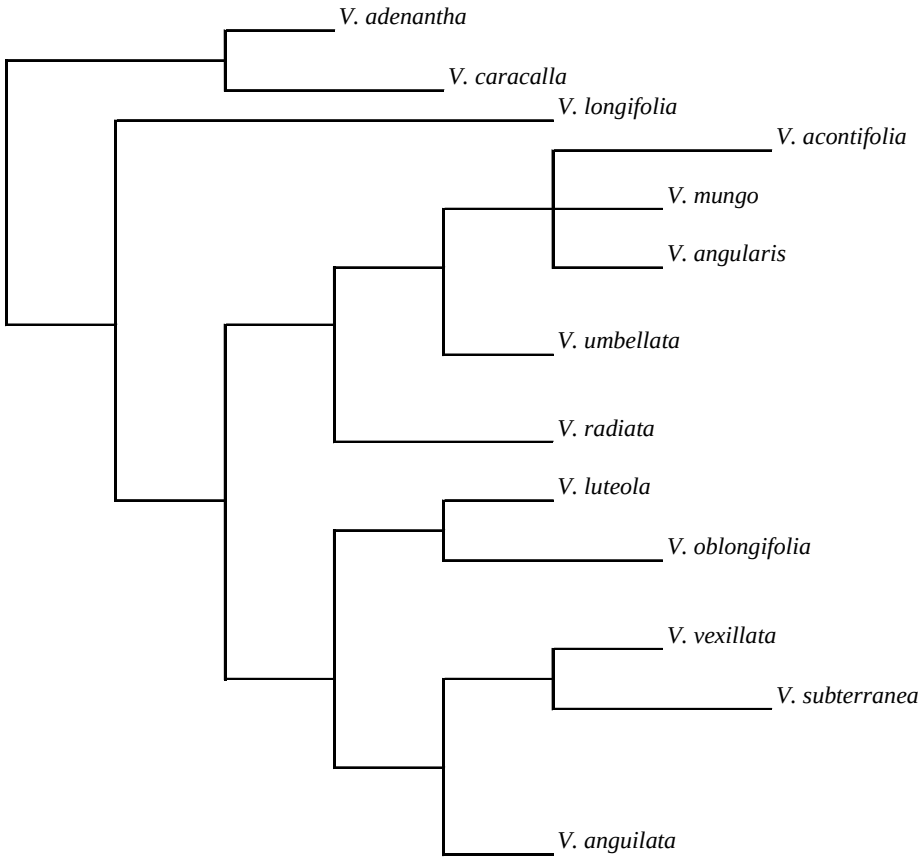


Figure 1.1: Phylogenetic trees showing the relationship between Vigna species from various Vigna subgenus and section. Adapted from (Wang *et al.*, 2008)

Legumes are an important part of subsistence agriculture as they provide protein-rich food, and ameliorate the soil by improving the structure (Sato *et al*., 2010). They are a source of oil production for human consumption and production of animal feed. Thanks to their symbiosis with nitrogen-fixing bacteria, legumes can be grown without the addition of nitrogen fertilizer and most can be grown on poor soils (Sandal *et al*., 2002). Leguminous plants have two model species, *Lotus japonicas* and *Medicago truncatula,* together with significant research in soybean (*Glycine max*), and these have been chosen to represent the diverse legume family. A number of studies have been conducted on model species to develop genome resources, such as the production of cDNA libraries, DNA marker

production and analysis, high density genetic linkage map production and genome sequencing. The experience gathered from model species can facilitate basic genetics in crop legumes and accelerate crop breeding (Sato *et al*., 2010).

1.2 Bambara groundnut taxonomy

Bambara groundnut (*Vigna subterranea* (L.)Verdc) synonym [*Voandzeia subterranea* (L.) Thouars] is a herbaceous, self-pollinating plant with an indeterminate growth habit. The domesticated bambara groundnut landraces have quite a distinct tap root and numerous short lateral stems on which the trifoliate leaves are borne, while the wild forms have a limited number of elongated lateral stems with no clear tap root. The petiole is long, stiff and grooved with a base of a wide range of colours such as green, purple or brown (Swanevelder, 1998). The species *subterranea* is further divided into two groups: var. *spontanea,* comprising the wild forms, found in a small area around northern Cameroon and Nigeria, and var. *subterranea* comprising the cultivated forms in parts of the tropics, mostly in sub-Saharan Africa (Pasquet *et al.,* 1999; Basu *et al.,* 2007). The chromosome number in both wild and cultivated plants is $2n = 2x = 22$ (Forni-Martins, 1986).

The wild bambara groundnut landraces usually have a spreading growth habit, compared to the compact type of domesticated landraces (Swanevelder, 1998). The other major difference between the two types is that of pod size, with domesticated landraces having bigger seeds which do not wrinkle upon drying, compared to the wild type (Pasquet, 2003, Basu *et al.,* 2007). The germination of cultivated forms is rapid and uniform while in the wild forms it is erratic and takes longer, from approximately 15 to 30 days to germinate (Basu *et al.,* 2007).Generally, the domestication of crops involves a number of major steps, with the development of altered plant architecture and also of harvest ability traits, so that a wild form plant can be domesticated and made more amenable to intensive agriculture (Basu *et al.,* 2007). Both morphological and isozyme data were used to demonstrate that wild bambara groundnut (*spontanea*) is the true progenitor of domesticated bambara groundnut (*subterranea*) by Pasquet *et al.,* (1999).

Domestication of bambara groundnut is believed to have occurred within the area where the wild forms are found, which is the Jos plateau and Yola regions of

northern Nigeria through to Garuoa in Cameroon and, possibly, as far as the Central African Republic (Hepper, 1963), with some authors including the areas from Nigeria to Sudan, which includes Cameroon, Chad, and the Central African Republic (Pasquet *et al*., 1999; Hanelt, 2001). Recently, Olukolu *et al.,* 2011, using both DArT molecular markers and phenotypic descriptors, provided evidence that pointed out Cameroon/Nigeria as the putative area of origin of bambara groundnut. The region showed a higher phenotypic diversity for both quantitative and qualitative characters compared to regions of East Africa, Central Africa and a combination of other countries in West Africa. The crop is believed to have been brought first to East Africa and Madagascar, then later to South and South East Asia, with the slave trade to Suriname, Brazil and later to the New World (Hanelt, 2001). It is reported to be cultivated in South and Central America, India, Indonesia, Malaysia, the Philippines, Sri Lanka and parts of northern Australia (Linnemann and Azam-Ali, 1993; Suwanprasert *et al.*, 2006).

Bambara groundnut is related to cowpea and has a podding habit similar to that of peanut (*Arachis hypogaea*) in that the pale yellow flower stalk bends downward after fertilization, pushing the young pod into the soil, where it develops and matures (Doku and Karikari, 1970; Uguru and Ezere, 1997), however, it is not believed to require complete coverage with soil for the pods to develop.

Figure 1.2: A typical bambara groundnut crop in the field, unshelled pods, flowering and pod initiation, and bambara groundnut seeds. Scale bar = 20 mm

1.3 The bambara groundnut crop

Bambara groundnut is an important food legume crop, cultivated mainly in sub-Saharan Africa. Through many years of successive cultivation, farmers have selected for desirable traits such as growth habit and seed colour (Linnemann and Azam-Ali, 1993). Farmers prefer the stable, reliable and low yield of bambara groundnut to high yields of groundnut, which has been associated with more yield volatility (Linnemann, 1994).

Bambara groundnut is adapted to wide climatic zones, it can be cultivated from sea-level to up to 1600 m altitude, and an average temperature of 20-28^{o}C is considered ideal for the crop. A growth period of 110 to 150 days is required for the crop to develop, although some records of reduced growth cycle landraces of approximately 90 days have been recorded in Ghana (Berchie *et al.,* 2010). It is usually grown in mixed intercropping systems with no addition of fertilizers (Karikari *et al.,* 1995). The crop does well on poor soils which are low in

nutrients; however the application of phosphorus results in better nitrogen fixation and an increase in stover and kernel yield (Ellah and Singh, 2008). It grows well on well-drained soils, but sandy loams with a pH of 5.0 to 6.5 are most suitable (Swanevelder, 1998).

The seed makes a complete food as it contains sufficient protein, carbohydrate, fat and micronutrients (Poulter and Caygill, 1980). The seeds are consumed in a variety of ways, as fresh pods or boiled with salt and pepper, or eaten as a snack or mixed with maize seeds or with maize flour as a relish. Nutritional composition undertaken by several researchers revealed that on average the seeds contain 63 % carbohydrates, 19% protein, and 6.5 % oil (Ijarotimi and Esho, 2009; Nwokolo, 1987; Borough and Azam-Ali, 1992). The protein is of high quality having a good balance of the essential amino acids and a relatively high lysine (6.8%) and methionine (1.3%) content (Ellah and Singh, 2008). The gross energy has been reported to be higher than that of other pulses including cowpea, lentils and pigeonpea (Poulter, 1980). The high nutritional value of bambara groundnut provides a cheap source of protein to poorly-resourced farmers in semi-arid areas (Doku *et al*, 1978; Borough and Azam-Ali, 1992; Amarteifio *et al.*, 2006) making it a good supplement to a cereal-based diet.

The production records of bambara groundnut in some countries are scanty since it is recorded together with other pulses and sometimes records are not easy to get because it has not entered the formal market (Mbewe *et al.,* 1995). According to the Food and Agriculture Organisation of the United Nations: FAOSTAT (2009) most of bambara groundnut production is taking place in West African countries with Burkina Faso producing (44712) metric tonnes (MT), Mali (25165) MT, Cameroon (24000) MT, and Democratic Republic of Congo (1000) MT.

1.4 Potential of bambara groundnut

Most African countries rely on rainfed agriculture, but such agriculture is particularly vulnerable to climate change. In addition, there are usually other concerns such as poverty, soil degradation and recurring drought (Mendelsohn, 2000). In most countries in sub-Saharan Africa that are prone to drought, unreliable rainfall, poor soils and poor crop productivity, the production of more

drought tolerant, indigenous crops, such as bambara groundnut are encouraged. There is evidence that demonstrates that the crop is more resilient to adverse environmental conditions as it tolerates low soil fertility soils and low rainfall, and it is one of the most favoured crops by indigenous people (Azam-Ali *et al.*, 2001).

1.4.1 Some interesting agronomic characters of the crop.

Bambara groundnut landraces have been shown to be able to tolerate drought as they can sustain leaf turgor pressure by employing a combination of osmotic adjustment, leaf area reduction and effective stomatal regulation of water loss (Collinson *et al.*, 1997). Some changes in the leaf orientation, which assist the crop to reduce incident radiation on the leaf surface, are reported in droughted landraces such as DipC from Botswana and DodR from Tanzania, reducing water loss through transpiration (Collinson *et al.*, 1999). Recently, drought response mechanisms of bambara groundnut were revealed in two landraces, one from a drought-prone environment (Namibia), S19-3, and from a high rainfall area (Swaziland), UniswaRed. UniswaRed had a relatively higher transpiration rate under drought conditions compared to S19-3 which showed a delay in reduction in transpiration. This mechanism allowed S19-3 to maximise its water use and escape drought better than UniswaRed (Jørgensen *et al.*, 2010). The crop is endowed with the advantages of being relatively resistant to pests and diseases, and has substantial morphological diversity, with good adaptation to marginal areas and poor conditions (Azam-Ali*et al.*, 2001; Sesay *et al.*, 1996). It also contributes to the soil fertility through biological nitrogen fixation making it beneficial in crop rotations and intercropping (Mukumbira, 1985, Karikari *et al*, 1995), hence farmers do not normally apply chemical fertilizers to bambara groundnut (Mkandawire, 2007).

1.4.2 Bambara groundnut yield potential

The crop managed to outperform groundnut in controlled environment experiment to survive and produced some pods when groundnut failed, which is a clear indication of the crop potential (Azam-Ali *et al*, 2001). Bambara groundnut landraces produced good yield in controlled environment and field experiments, such as the University of Nottingham's Tropical Crops Research Unit (TCRU) where pods yields as high as 4 tha^{-1} were obtained (Collinson *et al.*, 1999). In the

fields in Swaziland, Sesay *et al.*, (2008) recorded seed yield of 2.6tha^{-1} while in Côte d' Ivoire (Kouassi and Zoro, 2009) recorded seed yield as high as 4 tha^{-1}. If these landraces are developed further to produce cultivars and varieties they could possibly produce even greater yields. The fresh seed of bambara groundnut often have a high market price, with demand outweighing supply in many areas (Coudert, 1984). In Botswana it is more expensive than cowpea and groundnut (Botswana Agricultural Marketing Board, 2008), making it a good source of income.

1.4.3. Some uses of bambara groundnut

In Botswana, soybean (*Glycine max*) is the ingredient for most weaning foods, although bambara groundnut has been found to be promising in initial results as a replacement, but has not yet been fully explored (Wambete and Mpotokwane, 2003; Ohiokpea, 2003). In Kenya, it is slowly replacing peanut as a substitute for weaning food (Mkandawire, 2007). In recent years, bambara groundnut's importance as a cash crop has increased, as it is now being canned at commercial levels in Zimbabwe (Makanda *et al.*, 2009). The haulm for bambara groundnut is also a valuable source of animal feed (Tibe *et al*, 2007).

1.4.4 Genetic diversity resources

There are substantial bambara groundnut genetic resources for the future improvement of the crop since there are approximately 2000 seed accessions in gene banks held by International Institute of Tropical Agriculture (IITA), and about 972 accessions in the Southern Africa Development Community (SADC) countries (Massawe *et al.,* 2005). This provides a good opportunity for bambara groundnut variety development and improvement of yields, which are still relatively low.

1.4.5. Potential areas of expansion

Bambara groundnut has wide adaptability, since it is able to grow in ecological zones of varying climates, ranging from areas with annual rainfall as low as 300 mm annually in Botswana to high annual rainfall of 1250 mm in Swaziland (Azam-Ali *et al.,* 2001). By scrutinising the world for potential bambara groundnut production using Geographic Information Systems (GIS) technology,

Azam-Ali *et al.,* (2001) identified some areas in America, Australia, Asia, as well as in Africa, where it could produce significant pod yields, and some areas in the Mediterranean were it is predicted to have the potential of producing yields as high as 8.5tha^{-1}

1.5 Constraint to bambara groundnut production

The introduction of peanut- groundnut (*Arachis hypogaea*) in many developing countries has replaced bambara groundnut as a major crop (Azam-Ali *et al.,* 2001; Pasquet, 2003). Since bambara groundnut is grown by smallholders, especially women, in drier regions (Linneman and Azam-Ali 1993) with limited resources there is more likely to be poor management of the crop and thus yields are usually low. In addition there are no established bambara groundnut varieties as yet, and farmers are using landraces which are a mixture of genotypes (Zeven, 1998). Bambara groundnut, as an underutilized species, until recently has been largely ignored by research. To some extend this is due to lack of funds for research in developing countries where the crop is grown (Azam-Ali *et al*, 2001).

1.5.1 Influence of sowing date/Photoperiod

Sowing date has been reported to influence the yield and yield variability, through the effects of temperature and day length on plant development (Collinson *et al.,* 1996; Sesay *et al.,* 2008). It is a short day species; in most bambara groundnut genotypes the onset of flowering is not affected by photoperiod while the onset of podding is adversely affected by photoperiod (Brink, 1997).There is variation among genotypes in regard to response to photoperiod both at onset of flowering and onset of podding with landraces Ankap 4, Yola and Ankap 2 from Nigeria appearing to show different responses to photoperiod sensitivity to onset of podding (Linnemann and Craufurd, 1994). This suggests that the crop generally has a facultative response to photoperiod (Jackson, 2008). No genetic studies as yet have been undertaken on the photoperiod response of bambara groundnut to identify genomic regions affecting the response of the crop. This is despite photoperiod being an important characteristic, in attempts to further adapt the crop, particularly in countries away from the equator.

1.5.2 Low moisture, pests and diseases

Low yields of approximately 700 $kgha^{-1}$ and as low as 200$kgha^{-1}$ have been recorded in bambara groundnut. Despite the crop being tolerant to drought (Balole *et al.,* 2003), dry matter production and yield of bambara groundnut are adversely affected by soil moisture stress (Collinson *et al.,* 1996; Mwale *et al.,* 2007) as is the case with all crops. Even though it is a hardy crop and susceptible to few pests and diseases some had been observed to cause damage to the crop. Some diseases such as leaf spot and blight (*Phoma exigua* var. *exigua*), root rot (*Pythium parocandrum*) and wilt (*Fusarium solani* and *F*. o*xysporium*) and root knot nematode (*Meloidogyne javanica*) have been observed in Swaziland and Botswana (Magagula *et al.*, 2003; Karikari *et al*., 1995). Aphids, which in turn spread rosette viral diseases and groundnut plant hopper (*Hilda patruelis*) which feeds on pegs and pods have been reported on bambara groundnut (Mkandawire, 2007). In storage, shelled bambara groundnut seeds are susceptible to attack by bruchids (*Callosobruchus maculata*s), the shelled ones being less susceptible (Munthali and Ramoranthudi, 2003). Observations have shown that the crop is vulnerable to fungal disease attacks caused by *Colletotrichum capsici* with adverse effects on grain yields (Obagwu, 2003).

1.5.3 Anti-nutritional factors in bambara groundnut

Even though bambara groundnut is an important source of protein in developing countries, research has revealed the presence of condensed tannins, especially among the brown, tan and red coloured landraces. Tibe *et al.,* (2007) in their study found 13 out of 27 landraces from Namibia, Botswana and Swaziland to contain tannin content below the allowed limit of 0.1% in weaning food in Botswana. The condensed tannin content ranged from 0.02 % to 0.49%, the cream coloured landraces recorded levels well below the allowed limit and are recommended to be used as weaning formula. However, Akaninwor and Okechukwu (2004), in Nigeria found that tannin content in bambara groundnut can be reduced by approximately 50% through processing techniques, such as soaking, dehulling, drying and autoclaving. Farmers also claim that the cream coloured seed requires shorter cooking time and taste better compared to red and dark coloured seed (Ramolemana *et al*., 2003).

1.5.4 Genetic resources

There are substantial amount of genetic resources held by the IITA and various gene banks in SADC countries. Despite these abundant genetic resources, at the moment there is no Consultative Group on International Agricultural Research Institution (CGIAR) that has a mandate to do research on bambara groundnut (Mayes *et al.,* 2009). IITA list their legume crops as cowpea and soybean (http://www.iita.org) while International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) list their legume crops as chickpea, pigeonpea and groundnut (http://www.icrisat.org). The genetic potential of bambara groundnut is not yet fully exploited, but with the introduction of biotechnology, new techniques such as molecular markers will assist researchers to better understand the genetics of bambara groundnut.

1.5.5 Mating systems

Bambara groundnut produces perfect flowers, it is self-pollinating and the fertilization of the ovule occurs at the day of anthesis (Linnemann, 1994). It is difficult to undertake hybridisation, with several attempts at artificial hybridisation reported as unsuccessful (Suwanprasert *et al.,* 2006) and a few reported cases achieved (Massawe *et al.,* 2003). Therefore, relatively few studies have been undertaken on the inheritance of yield and related traits in bambara groundnut (Basu *et al.,* 2007), hence no breeding programme aimed at improving bambara groundnut has so far been initiated to develop cultivars or varieties (Oyiga *et al.,* 2010).

1.6 Diversity evaluation using both morphological and molecular markers

1.6.1 Genetic markers

Crop genetic diversity is important for crop adaptation to withstand pests and diseases and it is an important precondition for plant breeders to enhance the progress of traits of economic value such as yield. Various methods are available for use in estimating the genetic diversity of crops, such as morphological, biochemical and molecular markers. Measurements of genetic diversity can be generated using conserved accessions in gene banks (Gilbert *et al.,* 1999; Parzies

et al., 2000). DNA-based molecular markers have several advantages over the conventional phenotypic markers since their presence is not dependent on the growth stage of the crop and can be found in all tissues (Mondini *et al.,* 2009).

1.6.2 Morpho-agronomic markers

The morphological method is the oldest and considered the first step in description and classification of germplasm (Hedrick, 2005). Evaluation of genetic diversity through morphological traits is direct, inexpensive and easy. However, morphological estimations are more dependent on environment and are more subjective than other measurements (Li *et al.,* 2009). Morphological variability depends on a limited number of genes, and may not access much of the potential variability for the agronomic traits present in a crop (Mayes *et al.,* 2009). The use of morphological and agronomic traits is a standard way of assessing genetic variation for many species, especially under-researched crops such as bambara groundnut (Azam-Ali *et al.,* 2001).

Since bambara groundnut is an underutilised crop, studies of its genetic diversity are scarce. However, Goli *et al.,* (1995), characterized and evaluated approximately 1400 bambara groundnut accessions at the International Institute of Tropical Agriculture (IITA) in Nigeria based on 38 characters, which included both quantitative and qualitative traits. Substantial agro-morphological diversity was revealed, which they recommended to be confirmed using molecular markers. Ntundu *et al.,* (2006) identified some vegetative traits that had prominent loadings in principal components analysis, and these are useful in distinguishing bambara groundnut landraces. Similar traits, like seed weight, internode length, petiole length, leaflet length, leaflet width, were identified as important traits in distinguishing between wild and domesticated bambara groundnuts when analysed with isozyme markers (Pasquet *et al.,* 1999). In addition morphological characters which can be highly correlated to grain yield give breeders the choice to make decisions as to which traits to select for in bambara groundnut landraces (Karikari, 2000).

Morphological markers have been used for phenotypic diversity studies in a number of crops. Several numerical taxonomic techniques have been successfully employed to classify and measure the patterns of genetic diversity in the

germplasm collection by other researchers working on crops such as black gram (*Vigna mungo*) and Mungbean (*Vigna radiata*) (Ghafoor *et al.,* 2001), soybean (*Glycine max*) (Cater *et al.,* 2001) and wheat (*Triticum aestivum*) (Bechere *et al.,* 1996). The comparison of phenotypic and genotypic variation within and between several other crops has been examined to provide accurate taxonomic and genetic differentiation in *Musa* spp, (Crouch *et al.,* 2000), cowpeas (*Vigna unguiculata*) (Omiogui *et al.*, 2006) and sorghum (*Sorghum bicolor*)(Can and Yoshida, 1999). Agronomic and morphological characters have been used to identify traits contributing to important traits such as yield in crops like bambara groundnut (Makanda *et al.,* 2009) and soybean (Malik *et al.,* 2007).

In a strategy to develop what they termed phenotypic similarity index (PS), Cui *et al.,* (2001) used morphological and agronomic traits to study the phenotypic diversity of Chinese and North American soybean. A total of 47 Chinese and 25 North American cultivars were assessed for 25 characters. Their results showed more phenotypic diversity among the Chinese cultivars, than the North American cultivars, they also found clear differences between the two groups. From the use of morphological markers they managed to come up with a strategic plan to broaden the North American germplasm by the introgression of Chinese cultivars, especially those from different clusters.

Swamy *et al*., (2003) used 20 agronomic characters to study the phenotypic diversity and identify traits with higher loadings in principal component analysis (PCA) for use as best descriptors in the core collection of Asian groundnut (*Arachis hypogaea*). A total of 504 accessions which consist of 274 accessions of subs. *fastigiata* (var. *fastigiata* and *vulgaris*) and 230 subs. *hypogaea* (var. *hypogaea*) were evaluated. A significant difference between *fastigiata* and *hypogaea* groups was found, and the principal component analysis showed that all the traits contribute significantly to variation for both groups except pod yield per plant, which did not appear in the first five principal components for both groups. Low variation in the pod yield per plant indicated that it was not significantly contributing as a descriptor in these accessions.

In studies to determine the selection criteria for cowpea (*Vignaunguiculata*) breeding, Omoigui *et al*., (2006) analysed the genetic variability and heritability

of reproductive traits of cowpeas. They found a substantial amount of genotypic coefficient of variation (GCV) and broad sense of heritability (h^2) among the selected cultivars on a number of traits. Higher heritability for 100-seed weight (0.98), plant height (0.94), days to flowering (0.83) and days to maturity (0.77) were recorded which was an indication that progress could be achieved in selecting these traits for cowpea improvement

1.6.3 Biochemical markers

Isozyme analysis was the first technique used in the estimate of genetic variance developed by Lewinton and Hubby in (1966). Isozymes are protein molecules with different charges, and can be separated by gel electrophoresis based on their molecular sizes, weight and electrical charges (Hedrick, 2005). The use of isozyme is simple and cheap, since no DNA or sequence information, primers and expensive PCR machines are need as in other marker types. Isozyme markers have the advantage of being co-dominant, giving them an advantage over other markers such as RAPDs, which are dominant markers and they are reproducible (Spooner *et al.*, 2005). The main disadvantage is that there are few isozyme assays per species, and the enzymatic loci account for a small and non-random part of the entire genome. Isozyme analysis is also affected by plant tissue and plant developmental stage (Mondini *et al*, 2009). Different tissues in the same plant can reveal different isozyme variation.

Koenig and Gepts, (1989) employed nine polymorphic isozyme loci to study the genetic diversity of 83 wild common beans (*Pharsalus vulgaris*)from both the Mesoamerican and Andean regions. The study was able to confirm the existence of the two gene pools, Mesoamerican and Andean accessions. Genetic diversity of Dst, Hs, and Ht were estimated. Dst estimates the total gene diversity distributed among populations, Hs estimates mean heterozygosity with the population, while Ht measures the mean heterozygosity in the entire population. The level of genetic diversity within the wild species was Ht =0.13, and non-significant within accession of Hs =0.006, and between accessions a moderate between Dst = 0.126 was recorded. Pasquet *et al.,* (1999) used isozymes to investigate the population structure of bambara groundnut and partition the genetic diversity between domesticated and wild forms. A high genetic Nei genetic identity of 0.948

between the wild and domesticated bambara groundnut landraces lead to a conclusion that the wild bambara groundnut is the progenitor of domesticated landraces.

To augment the initial description based on morphological markers, biochemical markers were introduced and later replaced by DNA molecular markers which are more robust as compared to both morphological and biochemical markers.

1.6.4 Molecular Markers

Molecular markers are fixed marks in the genome found at specific locations of the genome, there are used to identify specific genetic differences. In order to precisely identify traits of interest, the marker must be close to the gene of interest so that the allele of both the marker and the gene could be inherited together. DNA markers are passed on from one generation to another through the laws of inheritance (Semagn *et al.,* 2006). Several markers are available to choose for genetic diversity studies. The selection criteria could be based on cost, technical labour, level of polymorphism, reproducibility, locus specificity and genomic abundance (Garcia *et al.,* 2004). Molecular markers are useful in the development of genetic and physical maps, and have increased the efficiency of indirect selection of marker linked traits, generally markers are classified into hybridisation based DNA markers and PCR-based DNA markers (Gupta *et al*., 1999).

1.6.4.1 Hybridisation (Sequence dependent)

1.6.4.1.1 Random Amplified Fragment Polymorphism (RFLP)

RFLP was the first DNA marker system which was widely used and is based on sequence differences which affect restriction enzyme recognition sequences. Anumbers of steps are required in RFLP analysis. Restriction enzyme digested genomic DNA is size fractionated by gel electrophoresis then transferred to a hybridisation membrane. A ‘DNA probe’, a short fragment of labelled DNA, is hybridised to the filter (Saiki *et al.,* 1985; Kumar *et al.,* 2009).The differences are caused by evolutionary processes, spontaneous mutations and unequal crossing over(González-Chavira *et al.,* 2006). RFLP can also result from differences in

DNA sequences (additions or deletions, or gross chromosomal changes such as inversions or translocations) and these changes the fragment sizes detectable as restriction fragment length polymorphisms (Michelmore and Hubert, 1987). Velasquez and Gepts, (1994) employed RFLP for diversity analysis of 85 common bean accessions in their center of origin. The accessions were classified into two major groups the Middle America and the Andes. The genetic diversity they recorded (Ht =0.38) was twice that they found when using isozyme markers. Overall their analysis of both RFLP and Isozyme showed that RFLP revealed more polymorphism.

However, RFLP has a number of disadvantages. It is time consuming, often uses radioactive reagents, and requires large quantities of high quality genomic DNA (Mondini *et al*, 2009). The RFLP technique has a problem of detecting low polymorphism and few loci per assay; it is also not amenable to automation (Semagn *et al.*, 2006). The limitations in terms of routine use of RFLP lead to the development of other markers such as RAPDs (Roy, 2000). Garcia *et al*., (2004), compared the efficiencies of random amplification of polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), amplified fragments length polymorphism (AFLP) and simple sequence repeat (SSR) to assess the genetic diversity of 18 tropical maize inbred lines. They employed a total number of 774 (AFLPs), 262 (RAPDs), 185 (RFLP) and 68 SSR markers for genetic diversity studies. The estimates of genetic distance correlation was higher for AFLP and RFLP (r =0.87), followed by AFLP and SSR (r =0.78), and RAPDs and SSR (r =0.33). The higher similarity between AFLP and RFLP markers are attributed to the fact that the two techniques are based on restriction site changes and both produced relatively higher polymorphism among the selected maize inbred lines.

1.6.4.1.2 Diversity Array Technology (DArT)

DArT is a micro-array hybridisation based technique that enables whole genome, high throughput and screening (Jaccourd *et al.*, 2001). In DArT, DNA samples from a representative sampling of the germplasm are assembled to make up a diversity panel. A complexity reduction method is carried out for the genomic DNA of the representative germplasm. The genomic representation derived is then cloned and individual inserts arrayed onto a microarray to form a discovery

array. The labelled DNA representations from individual test samples are hybridised to the discovery array. The polymorphic DArT markers can be identified as present or absent (Wenzel *et al.,* 2004; Semagn *et al,* 2006) in a single genotype. Various complexity reduction methods can be applied. A number of DNA based molecular markers available are hampered by their dependence on gel electrophoresis, therefore resulting in lower throughput. DArT is a genetic marker system which requires low quantities of DNA and can provide comprehensive genome coverage in organisms without prior DNA sequence information (Jaccoud *et al*., 2001).

DArT markers revealed low levels of genetic diversity between cultivated and wild pigeonpea (*Cajanus cajan*) (Yang *et al.,* 2006), in bambara groundnut DArT markers revealed a higher genetic diversity among a subset of 40 accessions selected from a representative of the 124 landraces (Olukolu *et al.,* 2011).Genetic diversity and mapping have also been carried out in crops such as barley (Wenzl *et al.,* 2004; Zhang *et al.,* 2009), wheat (Akbari *et al.,* 2006), and sorghum (Mace *et al.,* 2009). They have also been used in QTL analysis of root-lesion nematode resistance in barley (Sharma *et al.,* 2011), and mapping kernel characteristics in hard red spring whear lines (Tsilo *et al*., 2010). Recently, Briñez *et al.,* (2011), used DArT markers to assess the genetic diveristy of 89 common bean accesssions. The Neighbour-Joining distance matrices was employed to distinguish two major gene pools of common beans, the Mesoamerican and the Andean, which was in agreement with previous studies conducted, based on morphological markers, biochemical, and other molecular markers such as AFLP.

DArT are dominant markers thus are unable to differentiate heterozygous loci from homozygous, but have the advantage of high locus specificity, due to their detection by hybridisation (Jaccoud *et al*., 2001). While SSR markers have an advantage over DArT markers because they are co-dominant, highly polymorphic and widely distributed in the genome (Yang *et al*., 2006) they have the disadvantage that they require substantial sequence information to generate.

1.6.5 PCR-based molecular markers

The 'Southern transfer process' has been almost replaced by the polymerase chain reaction (PCR) (Mullis, 1990). PCR is useful in studying DNA sequence variation as it provides amplification of the DNA between two specific priming sites in the genome. Polymerase chain reaction based markers require less DNA per assay than RFLP and are higher throughput.

1.6.5.1 Random Amplified Polymorphic DNA (RAPD)

RAPD markers offered an opportunity to reduce the time and expense taken in RFLP for genetic diversity and molecular mapping. It is based on PCR amplification of random DNA segments with short, arbitrary primers (William *et al.*, 1990). An oligonucleotide used for RAPDs is usually ten base pairs long and amplifies many loci simultaneously and therefore a number of multiple markers can be assayed within a single PCR reaction. The amplified DNA is visualised after ethidium staining and there is no need for hybridisation with labelled probes as in RFLPs (Kumar *et al.,* 2009).

The technique has been used for identification and mapping QTLs conferring resistance to *Aschochyta blight* in chickpea (Santra *et al.,* 2000) and identification of the Uvf-1 gene which confers resistance against rust in *Vicia faba* (Avila *et al.,* 2003). $RAPD_S$ have been used in bambara groundnut for some landraces in a genetic diversity assessment (Amadou *et al.,* 2001; Massawe *et al.,* 2003; Mine *et al.,* 2003). High levels of polymorphism were reported among landraces using RAPDs markers in contrast to isozyme markers used by Pasquet *et al*., (1999).

Twenty-one RAPDs and 29 SSR markers were used to assess the genetic variation and relationships between subspecies and botanical varieties of cultivated peanut (*Arachis hypogaea*) and their relationships with the wild peanut species of the genus *Arachis, Heteanthae, Rhizomatae* and *Procumbentes.* A high polymorphism of 42.7% for RAPDs and 54.4 % for SSR was recorded for the 13 Arachis selected, which was relatively high genetic variation for peanut as it is considered to generally have a lower genetic variation (Raina *et al*., 2001).

The RAPDs technique is simple and inexpensive and can be used in laboratories with limited resources. Some short comings of RAPDs include its poor

reproducibility and when used in linkage map production, the same loci may not be detectable in different populations. The false positives observed in RAPDs emanates from the rearrangement of fragments produced by primer binding sites and intrastrand annealing and interactions during PCR reactions (Semagn *et al.,* 2006). RAPDs are dominant markers and do not differentiate between homozygous and heterozygous markers. The inherent problems of reproducibility make RAPDs unsuitable markers for transferability of results.

1.6.5.2 Amplified Fragment length polymorphism (AFLP)

AFLP was developed to overcome some of the shortcomings of reproducibility of RAPDs as the technique combines the digestion of DNA with some specific restriction endonucleases with a PCR-based technique (Sandal *et al*., 2002). AFLP analysis involves the restriction digestion of genomic DNA with a combination of rare cutting (*Eco*RI or *Pst*I) and frequent cutting (*Mse*I or *Taq*I) restriction enzymes (Vos *et al,* 1995). Only DNA fragments with nucleotides that flank the restriction sites that match the selective nucleotides of the primer are amplified during PCR (Loh *et al.*, 1999). The technique is amenable to high-throughput analysis which is an added advantage. It is also more efficient and reproducible as compared to the RAPD (Semagn *et al*., 2006). AFLPs are highly effective markers and could be useful in genetic resource exploitation and identification of novel traits (Crouch and Ortiz, 2004).

AFLP has been used in genetic diversity analysis studies such as in common bean (*Phaseolus vulgaris*) (Maciel *et al*., 2003) in cowpeas (*Vigna unguiculata*) (Coulbaly *et al.*, 2002), and in bambara groundnut by Massawe *et al.*, (2002) and Ntundu *et al.*, (2004). AFLP has also been used in the mapping of the nodulation loci *sym9* and *sym10* in pea (*Pisum sativum*) (Schneider *et al*., 2002). The first outline linkage map of bambara groundnut was developed mostly from AFLP markers; 67 AFLP and one SSR (Basu *et al.*, 2007).

1.6.5.3 Microsatellites: Simple Sequence Repeats (SSR)

Microsatellites, or simple sequence repeats (SSR), are tandem di- to tetra-nucleotides sequence motifs flanked by sequences and are present in most eukaryotes genomes (McCouch *et al.,* 1997). They arise due to slippage-like events occurring randomly in stretches of repetitive sequence (Tautz, 1989). This makes microsatellite a more powerful genetic maker and because of their high reproducibility and co-dominance they are the marker of choice (Gupta and Varshney, 2000; Reusch, 2001). Microsatellites are mostly useful in comparative and association studies, genetic diversity, marker-assisted selection, population and evolutionary studies (Nunome *et al.,* 2006; Shi *et al.,* 2011). Because of their high variability they are especially good at distinguishing closely related individuals (Kumar *et al.,* 2009). A number of microsatellites are now available for a wide range of crops, such as groundnut (*Arachis hypogaea*) (He *et al.,* 2003; Cuc *et al.,* 2008), pigeonpea (*Cajanus cajan*) (Odeny *et al.,* 2007; Saxena *et al.,* 2010), bambara groundnut (Basu *et al.,* 2007), chickpea (*Cicer arietinum*) (Sethy *et al.,* 2003) and common bean (*Phaseolus vulgaris*) (Blair *et al.,* 2011).

The technical simplicity, small amount of DNA required and high power of genetic resolution had led to SSR markers slowly replacing other markers. The microsatellite amplification protocol is easy, once primers have been designed for a specific locus. After amplification of microsatellites by PCR, the products are separated by capillary gel electrophoresis and detection of amplified allele can be achieved by a laser induced fluorescence detection system. The use of fluorescence labelled primer and laser detection (automated genotyping), improves throughput, accuracy of call. The cost of fluorescent label attached to each primer, which could be prohibitive, could be reduced by the three primer procedure (Schueke, 2000). Multiple loci can be analysed simultaneously through multiplexing. The major problem with microsatellites is that they need to be isolated *de novo* from each species (Zane *et al.,* 2002). In addition, there is poor transferability of markers developed for one taxon to another (Ellis and Burke, 2007).

1.7 Microsatellites development and application

Bambara groundnut has genetic resources that offer potential for food security, but the lack of molecular marker systems for their diversity assessment poses a challenge for its genetic improvement and promotion as a crop (Yu *et al.,* 2009). Microsatellites have proven to be the marker of choice for genetic studies. Despite their usefulness for many applications, the difficulty, expense and time in obtaining microsatellite markers is a major hindrance in their use (Zane *et al.,* 2002). It is important that more microsatellites markers are developed. Basically there are two strategies used for microsatellite development: microsatellites markers can be sourced based on DNA sequence information deposited in the databases (mining in public libraries/databases) or through screening of genomicDNA libraries specifically constructed for discovery of repeated sequences in the genome (Ritschel *et al*., 2004).

1.7.1 *Microsatellites markers sourced from databases*

The development of SSR markers had been reported through searching expressed sequence tags (ESTs) databases. An EST is a DNA segment representing the sequence from a cDNA clone that is derived by reverse transcription from an mRNA molecule, or a part of it (Gupta *et al.,* 1999). *In silico* mining of microsatellites for the plant of interest can be done in the available DNA sequence databases at the National Centre for Biotechnology Information (NCBI) and European Molecular Biology Laboratory (EMBL) (Gupta and Varshney, 2000).

The sequences, after having been downloaded and aligned, can be used to identify unique flanking sequence for microsatellite marker development. The markers developed have been found to have the same utility as those derived from an enriched genomic library (Sharma *et al.,* 2007). This marker has been developed in crops such as peas (*Pisumsativum*) (Moreno and Polans, 2006), mungbean (*Vigna radiata*) (Seehalak *et al.,* 2009), lima bean (*Phaseolus lunatus*) (Gaitán-Solís *et al*., 2002), common bean (*Phaseolus vulgaris*) Garcia *et al.,* (2011) and chickpea (*Cicer arietinum*) (Qadir *et al.,* 2007; Varshney *et al.,* 2007). The transferability of EST-SSRs has been found to be relatively better compared to non-ESTs SSR markers (Ellis and Burke, 2007). EST-SSRs markers derived from *Medicago truncatula* revealed a significant transferability among other three

pulses, peas (*Pisum sativum*), faba bean (*Vicia faba*) and chickpea (*Cicer arietinum*) (Gutierrez *et al.,* 2005) and SSR markers have been used between cultivated peanut and wild peanut (Liang *et al*., 2009).

1.7.2 *Construction of genomic library*

The simple approach or 'traditional method' of obtaining microsatellites has been to create small insert in a plasmid library then to screen the clones by repeated rounds of filter hybridisation using oligonucleotides (Akkaya *et al.,* 1992; Strus and Plieske, 1998). This method was found to be laborious, time consuming and had low efficiency. The numbers of microsatellites discovered are low, and range from approximately 0.04 to 12%, especially in those species with low levels of microsatellite repeat (Nunome *et al.,* 2006).

The technique has since been improved through selective hybridisation. There are several approaches used to enrich the genomic library for microsatellites, detailed in Zane *et al.,* (2002). The approach of using enrichment for genomic library with microsatellites has been modified by Edwards *et al.,* (1996) and has proved to be popular and applied by many researchers e.g. pigeonpea (*Cajaus cajan*) Burns *et al.,* (2001), bambara groundnut (*Vigna subterranea*) Basu *et al.,* (2007) and pigeonpea (*Cajanus cajan*) Odeny *et al.,* (2009).

Factors such as the cloning efficiency, the need to increase the throughput by sequencing large clones, and hybridisation limit the scope of microsatellites. The advent of next generation sequencing is most likely to resolve these problems (Santana *et al.,* 2009). Microsatellites were chosen as the preferred method for studying the genetic diversity of bambara groundnut. Even though there have been great advances in genomic technology in several crops species, the availability of molecular tools such as microsatellites have been limited in bambara groundnut.

1.8 Potential application of microsatellites in bambara groundnut

1.8.1 Conservation of genetic resources

The management and characterisation of germplasm is a starting point for crop improvement. The germplasm collection is usually too large to be easily accessed

by plant breeders; hence the concept of core collection was developed so that a few representative accessions are selected for use (Glaszmann *et al.,* 2010).

There is a substantial number ofbambara groundnut accessions held by respective countries in sub-Saharan Africa, some of which have undergone some morphological characterisation such as those from Burkina Faso (Ouedraogo *et al.,* 2008) and Tanzania (Ntundu *et al.,* 2006).Some genotypes from IITA have undergone field characterisation and evaluation using phenotypic markers (Goli *et al.,* 1995). The use of microsatellites in particular should be able to identify clusters among closely related materials and identify genotypes distantly related for selection for breeding purposes. The use of SSR markers can also be helpful in adding more data to the IITA current passport data (Mayes *et al,* 2009). In germplasm some redundancies can occur and microsatellites can be used to identify the redundant or closely related accessions.

The International Crops Research Institute for Semi-Arid Tropics (ICRISAT) Upadhyaya *et al.,* (2008) employed 48 microsatellites to analyse a core collection of 3000 accessions of chickpea (*Cicer arietum*).They managed to divide the accessions into four manageable subsets of Desi, kabuli, peas shaped and wild accessions of *Cicer* among these accessions.

Genetic improvement has been successfully achieved in other leguminous crops such as in common bean and soybean. Similar approaches could be applied on bambara groundnut, which is lagging behind in terms of genomic research. The availability of microsatellites in bambara groundnut would enable breeders to target genes of interest. The application of marker assisted selection (MAS) techniques in bambara groundnut could be helpful in tagging those traits that are economically important, for example early maturity, photoperiod insensitive and high yield, understand drought tolerance in bambara groundnut and identify traits necessary to enhance water use efficiency. The genomics research already done (especially ‘omic’) in model legumes can be useful for studying ‘orphan’ crop such as bambara groundnut.

1.8.2 Molecular mappings

Genetic mapping assist in identifying simply inherited markers which are close to genetic factors affecting quantitative traits (QTLs). Molecular markers allow high density DNA marker maps to be made for a number of crops, and this provides the structure needed for the application of MAS. The traits could be genetically simple or complex quantitative traits, which involve many genes the quantitative traits loci (QTL) (Doerge, 2003).Sato *et al.,* (2010), observed that MAS and genomics have not yet been practically deployed significantly for underutilised crops, even though there is a lot of potential to have a significant impact on these crops.

The limited availability of microsatellites developed in some leguminous crops has been attributed to a number of factors, such as low variability in the crop, to lack of resources for marker development such as in chickpea (Millan *et al.,* 2006), groundnut (Varshney *et al.,* 2009), pigeonpea (Yang *et al.,* 2006) and bambara groundnut (Basu *et al.,* 2007).

For cultivated groundnut, the construction of linkage maps for the crop was only recently reported by Varshney *et al.,* (2009) using 318 recombinant inbred lines (RIL) population derived from a cross between two cultivated genotypes (RIL-1 :TAG 24 x ICGV 86031). The map consists of 191 marker loci on 22 linkage groups covering a total of 1785.40 cM with an average distance of 9.24cM. The mapping population segregates for drought tolerance traits like transpiration efficiency, specific leaf area and SPAD chlorophyll meter reading (SCMR). Even though several QTLs were identified, none revealed a high phenotypic variation that could be used in marker-assisted selection (Varshney *et al.,* 2009).

1.8.3 Marker Assisted Selection and QTL

QTL mapping assists in identifying most heritable variation attributed to the interaction between two or more genes and their environment. The knowledge acquired is useful in designing crosses that may lead to improvements in crop breeding (Collard *et al.,* 2005). Genetic markers have made it possible to identify regions of the genome (QTL) that contribute to the variation of traits of economic importance in crops. Such markers can be useful in introgression and to facilitate backcrossing which would otherwise take several years using just morphological

markers (Charcosset and Moreau, 2004). Some major achievements have been recorded in chickpea and pigeonpea, with tremendous crop improvement (Kumar *et al.*, 2011).

The present map for bambara groundnut developed by Basu *et al.*, (2007) is based on F_2 population derived by crossing *V. subterranea var. subterranea* (cultivated) x *V. subterranea var.spontanea* (wild). Sixty-seven AFLP and one SSR markers were mapped on 20 linkage groups spanning a total length of 516cM.Four major QTLs have been located on the map for seed weight, specific leaf area (SLA), carbon isotope discrimination (CID; an indicator of water use efficiency in other species) and number of stems per plant (Basu *et al.*, 2007). More traits of economic importance have to be studied.

A study in groundnut was undertaken by Khedikar *et al.*, (2010) for quantitative trait locus (QTL) analysis for late leaf spot (LLS) and rust, which are two major foliar diseases in groundnut which cause yield losses of approximately 50-70% in the crop. Parental genotypes TAG 24 and GPBD 24 were screened with 67 SSR markers which were found to be polymorphic out of a set of 1,039 SSR markers. 56 of these markers produced 14 linkage groups, spanning 462.4 cM with an average of 8.25 cM. The 268 recombinant inbred lines of TAG 24 and GPBD 24 were used in the QTL analysis, 11 QTL were produced for late leaf spot with 1.7 to 6.5 % phenotypic variation, 12 QTLs were produced for rust with 1.7 to 55.2 % phenotypic variation. In this study, they identified a candidate SSR marker (IPAHM 103) which is linked with a major QTL (rust 01), 55.2%. The marker was validated for use in marker assisted selection in rust disease in a large number of germplasm lines (Khedikar *et al.*, 2010).

Some important abiotic and biotic stress, pests and diseases that cause damage and losses to bambara groundnut have already been intensively studied in other leguminous crops and their QTLs mapped successfully. Detection and mapping of major locus resistance for fusarium wilt in common bean (Fall *et al*., 2001), resistance to bacterial blight (Singh and Muñoz, 1999), white mold resistance (Ender and Kelly, 2005), and phosphorus acquisition ability (Beebe *et al.*, 2006) have been reported. In groundnut, QTLs linked to drought resistance had been identified (Ravi *et al.*, 2011),

High resolution maps and ability to determine marker order is largely dependent on population size. The smaller populations sizes often results in detection of few QTLs which could have large phenotypic effects (Semagn *et al.,* 2010).The latest trend has been to combine QTL mapping with methods in functional genomics. More saturated maps that include SNPs, ESTs derived markers, and STSs provides a good opportunity for QTL mapping of highly saturated maps and could be useful in MAS and comparative mapping (Collard *et al.,* 2005). Expressed sequence tag collections provide a platform for microarray technology that gives and provides a potential source of candidate genes.

In a study to identify transcribed portion of the pigeonpea (*Cajanuscajan*) genome for genes associated with *Fusarium* wilt (FW) and Sterility Mosaic disease (SMD), 16 cDNA libraries were generated from *Fusarium* infected root tissues from four genotypes ICPL 20162 and ICP2376 for FW and ICP7035 and TTB7 for SMD. A total of 5,860 expressed (ESTs) for FW and 3,788 for SMD tissues were also discovered and deposited in the NCBI. This is a good opportunity for marker development, gene discovery and functional studies for other orphan crops (Varshney *et al.,* 2009).

The study of rice as a model species for cereal crops has indicated that individual rice chromosomes were largely collinear with those of other crops species such as maize, rye, sorghum, barley and wheat and other important agricultural grasses, at least at a gross level. Researchers identified QTL controlling important agronomic traits, such as shattering and plant height that had been mapped to collinear regions among grass species (Xu *et al.,* 2005).

A significant collinearity in gene order had also been reported in a number of legumes such as common bean and soybean. Yang *et al*., (2010) undertook a study to evaluate the efficacy of using soybean gene chip for transcript profiling in common bean. They hybrised cRNAs from nodule, leaf and roots for soybean and common bean in triplicate on a soybean Gene Chip, their results revealed that genes for basic cellular functions and metabolism were highly conserved between the two species. Their result is an indication of a functional orthologs between this species, and the study could be extended to other legumes for crop improvement.

An example is reported by Zhu *et al.*, (2005) when information from a model specie *Medicago truncatula* has been used to map the nodulation receptor kinase (*NORK*) gene which is responsible for both bacterial and fungal symbiosis in other legumes.

1.9 Comparison of genetic diversity estimates methods

A limit to plant breeding has been due to the lack of robust markers such as molecular markers, previous work was based on pedigree data, morphological, physiological and cytological measurements (Garcia *et al*., 2004). The advent of molecular markers has meant that plant breeders could estimate genetic diversity faster and easier. Since different marker types differ in their properties, it is possible they give different estimates of genetic diversity (Rauf *et al*., 2010). The comparison of molecular markers for estimating genetic diversity could show how useful a marker is for a plant breeding purpose(Franco *et al.*, 2001). The estimates of genetic diversity of makers can be done based on correlation, regression, scatter plots and cluster analysis (Weir, 1996).

The efficiency and utility of six primer combinations for AFLP and RAPD, 100 RFLP and 36 SSR markers were investigated in12 soybean germplasm by Powell *et al.,* (1996.). The study consisted of a total of 12 genotypes of *Glycine max* of which 2 are wild type *Glycine soja,* the similarity matrices for the markers were compared, it revealed that the average similarity matrix was lower for SSR(0.341) while the other markers were similar AFLP (0.655), RFLP (0.639), RAPD (0.664). The Mantel test was used to determine the correlation between the markers and found significant correlation between all maker types ($P<0.001$). The highest correlation was between SSR and AFLP (0.855) while the lowest was recorded between RAPD and RFLP (0.744). Both markers proved to be useful in the assessment of the selected genotypes.

Lu *et al.,* (1996) compared PCR based methods (RAPDs, AFLP, microsatellites-AFLP, and inter-SSR) with DNA based RFLP to determine the most informative, and useful in genetic diversity studies based on ten pea genotypes. Their results revealed that the PCR based method were more informative than RFLP, and trees

derived from PCR based markers were significantly correlated with the exception of inter-SSR derived tree.

Other studies for the comparison of markers were conducted in other crops as well, Pejic *et al.*,(1998) investigated the efficiency of RAPD, SSR and AFLP in the analysis of maize, inbred lines. Garcia *et al*., (2004), compared the utility of RAPDs, RFLP, AFLP and SSR markers to find the best marker suitable for maize inbred lines selection. In wheat, Stodart *et al.,* (2005) compared AFLP and SSR markers to determine their utility in genetic diversity measurements among the 44 bread wheat landraces from different regions.

The geneticdistanceestimates compared in leguminous crops, include the one from Maras *et al.,* (2008) who evaluated the ability of AFLP and SSR to detect genetic diversity among 29 common bean (*Phaseolus vulgaris*) accessions. Ten primer combinations of AFLP produced 112 polymorphic bands, while 14 SSR markers produced 100 polymorphic bands and both markers were able to separate the two gene pools of Andean and Mesoamerican origin. Jaccard coefficient of similarity was employed to generate similarity matrix in both markers, the two genetic distances GS_{AFLP} and GS_{SSR} were evaluated for correlation using the Mantel correspondence test (Mantel, 1967), and a significant correlation r =0.67 was found, which shows a good similarity between the two markers.

In comparison of the morphological and RAPDs markers in estimating the differences among 15 common beans (*Phaseolus vulgaris*), Dursun *et al.,* (2010), employed 8 RAPDs and 16 morpho-agronomic markers. The difference between the two markers was revealed in the displaying of clusters as they differed in topology. The Euclidean matrix produced by the morphological marker and the Dice similarity matrix from the RAPDs markers were compared using Mantel matrix correspondence tests, the results showed no correlation between the two markers. This lack of correlation was thought to be possibly incorrect measurements for morphological traits and few samples sizes for RAPDs used in the study (Dursun *et al.,* 2010). However, in most of the studies conducted to reveal the genetic distances estimates the relationship between molecular and morphological markers had been observed to show non-significant correlations (Burstin and Charcosset, 1997).

No study has been conducted to compare the genetic distance estimates of markers in the germplasm of bambara groundnut. Therefore this study aims to determine the genetic diversity among the selected bambara groundnut germplasm employing both morpho-agronomic (qualitative and quantitative) markers, and molecular markers, and determine the relationship between the two techniques.

1.9.1 The objectives of the study

- To develop and characterise microsatellites markers; the development of markers will have a major impact on the genetic analysis and breeding of bambara groundnut, particularly in genetic diversity, population structure analysis implementation of pure line selection.
- To characterise selected landraces based on morpho-agronomic characters and to determine the agro-morphological diversity among landrace and consequently produce a genetic distances estimate to correlate with the genetic distance estimates based on SSR.
- To conduct a genetic diversity estimate based on SSR markers, which will consequently produce a genetic distance estimates to correlate with the morphological marker distance estimates.
- To compare morpho-agronomic markers with the SSR markers and identify any significant correlations and evaluate which is more informative and whether the costs associated with molecular analysis are justified.
- To establish the genetic similarity among bambara groundnut landraces sampled across a vast area of sub-Saharan Africa using microsatellites (SSR) markers since there is little information about this germplasm. There is constant movement of bambara groundnut germplasm between various neighbouring countries, and among farmers within the same country.
- The existence of landraces in bambara groundnut means that there are likely to be multiple genotypes planted in any trial for a landrace. This will add genetic variability to the already existing environmental variability and interaction (i.e. $V_P = V_G + E_G + V_{GXE}$). Co-dominant microsatellite markers will allow us to determine whether this is more of a problem in some landraces than others

CHAPTER TWO: Materials and Methods

2.1 Introduction

This chapter is divided into two sections: molecular biology (DNA and marker techniques) and phenotypic (morpho-agronomic) assessment of the germplasm. Materials and methods that are common in each section are described. Those procedures described that are specific to some experiments are described under appropriate chapters. The procedures described were used to carry out experiments at the University of Nottingham, Sutton Bonington, Campus, (UK) and Botswana College of Agriculture, Sebele, and Department of Biological Science, University of Botswana (Botswana).

2.1.1. Standard solutions

A list of standard solutions that were used in the molecular biology experiments are in appendix 1, while section 2.1.2 is a list of plant materials used in the phenotypic assessment and molecular biology experiments.

2.1.2 List of plant material

Table 2.1.2.1: List of selected landraces used for the characterisation of SSR markers and DArT analysis, their areas of origin and the cluster where the landraces were selected. The selection was on the basis of a study conducted by Singrün and Schenkel (2003), where a total of 223 bambara groundnut landraces were analysed for genetic diversity using enzyme system *Eco*Ri/*Mse*I amplified fragment length polymorphism (AFLP).

Landraces	Origin	Region	Cluster
DodR	Tanzania(TZA)	East Africa	6
DodC	Tanzania (TZA)	East Africa	1
AS17	South Africa (RSA)	Southern Africa	15
DipC	Botswana (BWA)	Southern Africa	12
SwaziRed	Swaziland (SWA)	Southern Africa	14
TicaNicuru	Mali (MLI)	West Africa	Core
Ramayana	Indonesia(IND)	Asia	1
LunT	Sierra Leone (SLA)	West Africa	Core
Vssp6	Cameroon (CMR)	West Africa	Core
Nav 4	Ghana (GHA)	West Africa	Core
Nav red	Ghana (GHA)	West Africa	Core
Mahenene black	Namibia (NAM)	Southern Africa	Core
S19/3	Namibia (NAM)	Southern Africa	Core
S19-3	Namibia (NAM)	Southern Africa	8
UniswaRed	Swaziland (SWA)	Southern Africa	Core
SB16 5A	Namibia (NAM)	Southern Africa	1
AHM968	Namibia (NAM)	Southern Africa	15
NAM 1761/3	Namibia (NAM)	Southern Africa	8
Malawi 3	Malawi (MW)	Southern Africa	Core
Tvsu 569	Cameroon (CMR)	West Africa	4
Tvsu 610	Nigeria (NGA)	West Africa	6
Tvsu 747	Zambia (ZMB)	Southern Africa	7
GabC	Botswana (BWA)	Southern Africa	16
Tvsu 999	Zimbabwe (ZWE)	Southern Africa	17

Nine core accessions, have been previously used in BAMLINK experiments. (BAMLINK-Molecular, Environmental and Nutritional Evaluation of Bambara Groundnut (*Vigna subterranea*L. Verdc.) for Food Production in Semi-Arid Africa and India.

Table 2.1.2.2:A list of 123 bambara groundnut accessions, source and their areas of origin used in the experiment; 105 bambara groundnut accessions selected in the greenhouse (35 x 3) samples, and 34 accessions that were selected and planted in the field experiment in (Botswana).

Tag No.	Accession	Origin	Source	Regions	Grown in greenhouse			Field
					Sample 1	Sample 2	Sample 3	Selected
1	(Wild type) 1	Nigeria	IITA	West Africa				
2	(Wild type) 13	Nigeria	IITA	West Africa				
3	9	Nigeria	IITA	West Africa	A	B	C	C
4	(Wild type) 144	Ghana	IITA	West Africa	A	B	C	C
5	191	Benin	IITA	West Africa				
6	289	Benin	IITA	West Africa	A	B	C	B
7	85	Burkina Faso	IITA	West Africa				
8	292	Burkina Faso	IITA	West Africa				
9	308	Burkina Faso	IITA	West Africa				
10	1276	Ivory Coast	IITA	West Africa	A	B	C	B
11	1284	Central A.Republic	IITA	Central Africa				
12	1288	Central A.Republic	IITA	Central Africa				
13	1307	Central A.Republic	IITA	Central Africa				
14	1315	Central A.Republic	IITA	Central Africa				
15	1324	Central A.Republic	IITA	Central Africa				
16	1329	Central A.Republic	IITA	Central Africa				
17	1337	Central A.Republic	IITA	Central Africa				
18	(Wild type) 1206	Burkina Faso	IITA	West Africa				
19	1352	Central A.Republic	IITA	Central Africa				
20	118	Ivory Coast	IITA	West Africa	A	B	C	A
21	438	Cameroon	IITA	West Africa				
22	440	Cameroon	IITA	West Africa				
23	447	Cameroon	IITA	West Africa				
24	448	Cameroon	IITA	West Africa				
25	(Wild type) 1164	Burkina Faso	IITA	West Africa				
26	460	Cameroon	IITA	West Africa				
27	467	Cameroon	IITA	West Africa				
28	472	Cameroon	IITA	West Africa				
29	473	Cameroon	IITA	West Africa				
30	476	Cameroon	IITA	West Africa	A	B	C	C
31	480	Cameroon	IITA	West Africa				
32	483	Cameroon	IITA	West Africa				
33	484	Cameroon	IITA	West Africa	A	B	C	C
34	492	Cameroon	IITA	West Africa				
35	501	Cameroon	IITA	West Africa				
36	502	Cameroon	IITA	West Africa				
37	503	Cameroon	IITA	West Africa				
38	506	Cameroon	IITA	West Africa				
39	529	Cameroon	IITA	West Africa				
40	536	Cameroon	IITA	West Africa	A	B	C	C
41	210	Ghana	IITA	West Africa				
42	214	Ghana	IITA	West Africa				
43	216	Ghana	IITA	West Africa				
44	229	Ghana	IITA	West Africa				
45	231	Gambia	IITA	West Africa	A	B	C	C
46	243	Gambia	IITA	West Africa				
47	246	Gambia	IITA	West Africa				
48	790	Zambia	IITA	Southern Africa	A	B	C	C
49	793	Kenya	IITA	East Africa	A	B	C	N/A
50	792	Kenya	IITA	East Africa	A	B	C	A
51	799	Madagascar	IITA	Southern Africa				
52	806	Madagascar	IITA	Southern Africa				
53	808	Madagascar	IITA	Southern Africa				
54	810	Madagascar	IITA	Southern Africa				
55	88	Mali	IITA	West Africa				
56	89	Mali	IITA	West Africa	A	B	C	A
57	91	Mali	IITA	West Africa				
58	23	Nigeria	IITA	West Africa				
59	25	Nigeria	IITA	West Africa				
60	32	Nigeria	IITA	West Africa	A	B	C	C
61	33	Nigeria	IITA	West Africa				
62	119	Nigeria	IITA	West Africa				

Table 2.1.2.2: continued

Tag No.	Accession	Origin	Source	Regions	Grown in greenhouse			Field
					Sample 1	Sample 2	Sample 3	Selected
63	120	Nigeria	IITA	West Africa				
64	172	Nigeria	IITA	West Africa				
65	395	Cameroon	IITA	West Africa				
66	275	Nigeria	IITA	West Africa				
67	278	Nigeria	IITA	West Africa				
68	283	Nigeria	IITA	West Africa				
69	286	Nigeria	IITA	West Africa	A	B	C	B
70	329	Nigeria	IITA	West Africa	A	B	C	C
71	330	Nigeria	IITA	West Africa				
72	331	Nigeria	IITA	West Africa				
73	334	Nigeria	IITA	West Africa				
74	335	Nigeria	IITA	West Africa	A	B	C	A
75	348	Nigeria	IITA	West Africa				
76	390	Sudan	IITA	Central Africa	A	B	C	B
77	391	Sudan	IITA	Central Africa				
78	369	Tanzania	IITA	East Africa				
79	371	Tanzania	IITA	East Africa				
80	379	Tanzania	IITA	East Africa				
81	385	Tanzania	IITA	East Africa	A	B	C	C
82	682	Zambia	IITA	Southern Africa				
83	683	Zambia	IITA	Southern Africa				
84	696	Zambia	IITA	Southern Africa	A	B	C	B
85	754	Zambia	IITA	Southern Africa	A	B	C	B
86	757	Zambia	IITA	Southern Africa				
87	1033	Zimbabwe	IITA	Southern Africa				
88	AHM753	Namibia	UoN	Southern Africa	A	B	C	A
89	DipC	Botswana	UoN	Southern Africa				
90	S19-3	Namibia	UoN	Southern Africa	A	B	C	C
91	UNIS R	Swaziland	UoN	Southern Africa	A	B	C	C
92	AHM968	Namibia	UoN	Southern Africa	A	B	C	B
93	AS17	South Africa	UoN	Southern Africa				
94	Dod C	Tanzania	UoN	East Africa				
95	Dod R	Tanzania	UoN	East Africa				
96	GAB C	Botswana	UoN	Southern Africa				
97	JAC B	Botswana	UoN	Southern Africa				
98	KABCA 4	Sierra Leone	UoN	West Africa				
99	SB4-2	Namibia	UoN	Southern Africa				
100	SB16 5A	Namibia	UoN	Southern Africa				
101	UNIS C	Swaziland	UoN	Southern Africa				
102	V5 6O A	Botswana	UoN	Southern Africa				
103	TICANICARU	Mali	UoN	West Africa				
104	S-1913	Namibia	UoN	Southern Africa	A	B	C	A
105	MaheneneBlack	Namibia	UoN	Southern Africa	A	B	C	A
106	YOLA	Nigeria	UoN	West Africa				
107	NAV-RED	Ghana	UoN	West Africa				
108	NAV-4	Ghana	UoN	West Africa				
109	BOTS 1	Botswana	DAR	Southern Africa	A	B	C	A
110	BOTS 2	Botswana	DAR	Southern Africa				
111	BOTS 3	Botswana	DAR	Southern Africa				
112	BOTS 4	Botswana	DAR	Southern Africa				
113	BOTS 5	Botswana	DAR	Southern Africa	A	B	C	A
114	CS37(RP)	Kenya	UoN	East Africa				
115	CS129 (RP)	Kenya	UoN	East Africa				
116	VSSP 11	Cameroon	UoN	West Africa				
117	VSSP 6	Cameroon	UoN	West Africa	A	B	C	C
118	RAMAYANA	Indonesia	UoN	Asia	A	B	C	B
119	Hybrid	UoN	UoN	Southern Africa	A	B	C	B
120	BC	Indonesia	Bogor	Asia	only DNA samples sourced			
121	BH	Indonesia	Bogor	Asia	only DNA samples sourced			
122	GC	Indonesia	Gresik	Asia	only DNA samples sourced			
123	GH	Indonesia	Gresik	Asia	only DNA samples sourced			

N/A: Not applicable for accession 49-Acc793KEN which was not planted in the field due to shortage of seeds. A, B, C: represent plant/genotype 1, 2, 3 respectively, for the same landrace

2.1.3 Overview of experiments.

Laboratory experiments were conducted to investigate the genetic diversity of bambara groundnut, using a core of 24 landraces. These landraces were selected based on a study conducted by Singrün and Schenkel (2003) (Table 2.1.2.1). These formed an initial test set to evaluate marker polymorphism and provided a link to previous DArT analysis. A total of 75 pairs of microsatellites were characterised (Appendix 2).

The second aspect of the project involved the characterisation and evaluation of bambara groundnut landraces for quantitative and qualitative characters. This was initiated in the agronomy bay glasshouse at the University of Nottingham using 119 bambara groundnut landraces (87 sourced from International Institute of Tropical Agriculture, 27 from The University of Nottingham, and 5 from Department of Agricultural Research, Ministry of Agriculture, Botswana) the experiment was conducted at the School of Biosciences, UK (27May planting to 4November 2008 harvest). DNA was extracted from the 119 accessions which were planted in the glasshouse and sent for DArT analysis (Mayes et al., 2009) as well as a subset of samples being analysed with microsatellite markers characterised in this project.

Field work was conducted on 34 lines derived from seed from single plants selected from the previous year's experiment (among the 119 bambara groundnut landraces) (Table 2.1.2.2)and planted at the Botswana College of Agriculture, Notwane (Sebele) field (11December 2008 to 11May, 2009). Extraction of DNA was carried out at the University of Botswana, Biological Science Department from (20May to 5June, 2009).

A selection of the five best lines from the field experiment for use as potential varieties to use in (Botswana), and a growth room experiment was conducted for characterisation, evaluation and genetic analysis of these set of lines

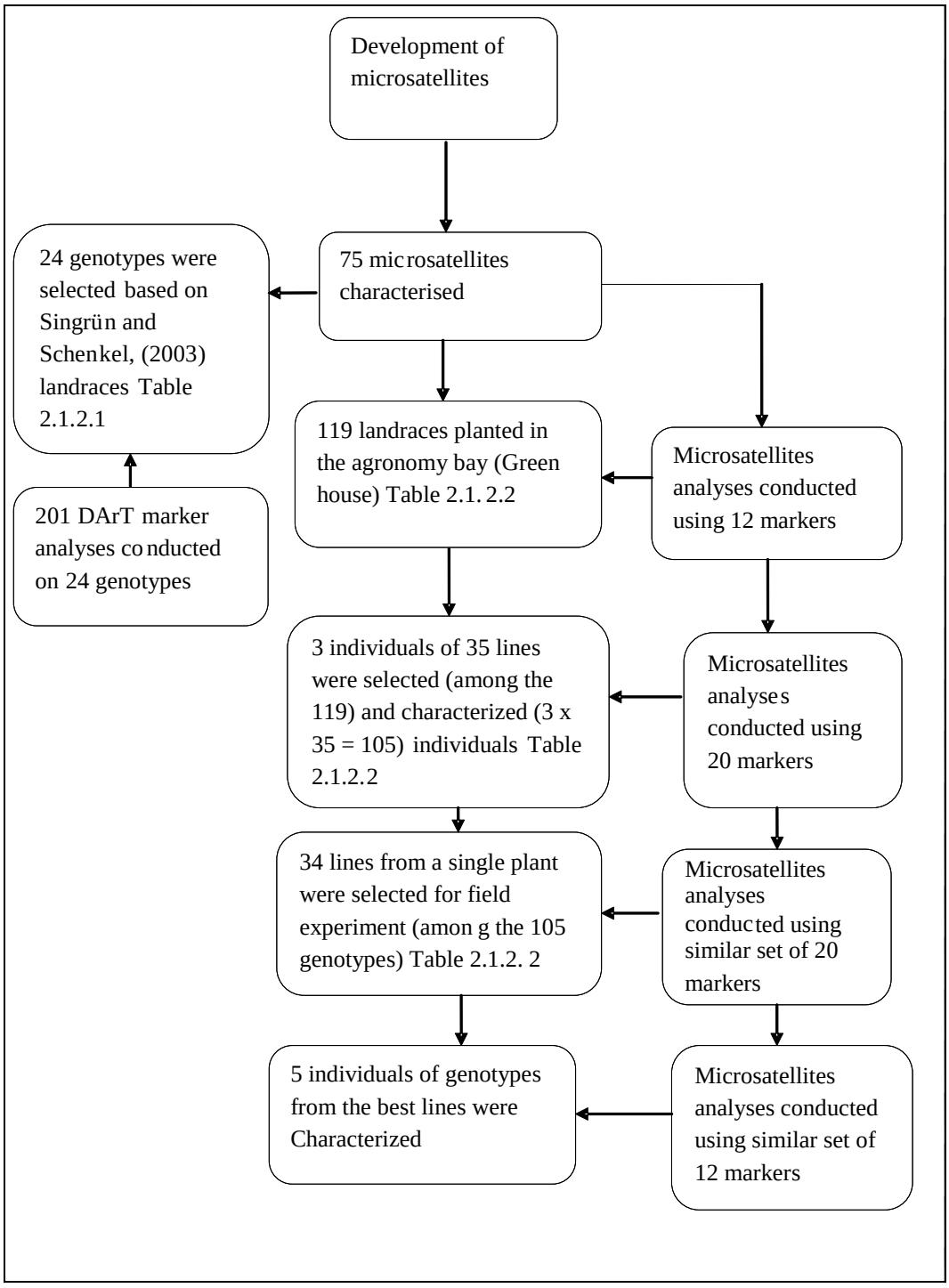


Figure 2: Diagrammatic representation of the setup of the project

2.2 Methodology for Marker and DNA techniques

2.2.1 Introduction

Characterisation of molecular markers is dependent on the amplification of DNA extracted from samples based on available markers. This section describes the procedures used to extract DNA from the plant materials listed in section 2.1.2.2, the quantitation and amplification employed. The development and characterisation of markers listed in Appendix 2, and their subsequent analysis are described. The use of capillary electrophoresis to size the amplified fragments and the potential genotyping errors and some mitigating strategies in microsatellites analysis are described.

2.2.2 Plant materials

2.2.2.1 Plant materials for microsatellites characterisation and DArT analysis

Isolation of bambara groundnut DNA was undertaken for use in the characterisation of microsatellites. A core set of 24bambara groundnut landraces listed in (Table 2.1.2.1) formed an initial test to evaluate the marker polymorphism and provide a linkage to DArT analysis.

2.2.2.2 Plant materials used for population structure analysis

A total of 119 bambara groundnut accessions were planted in the agronomy bay, 118 are from Africa while 5 are originally from Indonesia (4 were directly sourced from Indonesia, and 1 is from The University of Nottingham Stock) (Table 2.1.2.2.).African accessions were sourced as follows; eighty-seven accessions were sourced from the International Institute of Tropical Agriculture (IITA; Nigeria) while 27 accessions were sourced from the University of Nottingham stocks and five were supplied from Department of Agricultural Research, Ministry of Agriculture, Botswana. The origin of the complete accession list was derived from five major regions**;** 5 genotypes are from Asia (Indonesia), 11 from Central Africa region (Central African Republic and Sudan), 11 from East Africa (Kenya and Tanzania), 29 from Southern Africa (Botswana, Namibia, Madagascar, Swaziland, Zambia and Zimbabwe) and 67 from West Africa (Nigeria, Ghana, Benin, Burkina Faso, Ivory Coast, Cameroon and Sierra Leone).

2.2.2.3 Plant materials used for genetic diversity analysis

A total of 35 bambara groundnut landraces in (Table 2.1.2.2) were used in the study. Twenty one accessions were sourced from IITA (Nigeria), while 12 were from the University of Nottingham stocks and two were sourced from Botswana. Three plants per accession which makes 105 genotypes were used for the genetic diversity study analysis to estimate the genetic diversity within landraces.

2.2.3 DNA extraction

In this experiment the intention was to get good quality genomic DNA to use in PCR for the optimisation of the 75 available primer pairs that has been developed and characterised in this experiment section 2.2.5. The 119 landraces were used in population structure analysis, while the 105 genotypes were used for genetic diversity analysis. The GenElute Plant Genomic DNA kit (Sigma Aldrich) was used in the DNA extraction, since it has been shown to produce high quality DNA for PCR (Basu *et al.*, 2007).

2.2.3.1 Sigma DNA extraction Kit

The GenElute Plant Genomic DNA kit (Sigma Aldrich) protocol was followed for DNA extraction. Fresh young growing leaves were picked and collected in 50 mL Falcon tubes (Sarstedt) and placed on liquid nitrogen. About200mg of leaf tissue was ground with a pestle in a mortar under liquid nitrogen until sample became fine powder and transferred into a pre-chilled Eppendorf tubes. Then 350µL of lysis A solution and 50µL of lysis B solutions were added and mixed thoroughly by vortexing. The mixture was then incubated at 65^{o}C for 10 minutes with occasional inversion. 130µL of precipitation solution was added to the mixture and mixed by inversion before incubating on ice for 5 minutes. The tube was centrifuged at 13,000 rpm for 5 minutes to separate debris, proteins and polysaccharides. The supernatant was transferred to a GenElute filtration column, and centrifuged at 13,000 rpm for 1 minute. 700µL of binding solution was added to the flow through in the collection tube and mixed by inversion. 500µL of column preparation solution was added to the binding column to activate the retention of DNA and then centrifuged at 13000 rpm for 1 minute and the flow through discarded. 700µL of supernatant was added into the prepared column and centrifuged for 13,000 rpm for 1 minute, and the flow through was discarded.

The binding column was placed into another tube for the first wash with 500µL of wash solution and centrifuged at 13000 rpm for 1 minute. The washing with 500µL of wash solution was repeated but now centrifuged at 13,000 rpm for 3 minutes. The flow through was discarded while the binding column was transferred to a new collection tube, and 100 µL of pre-heated elution solution at 65^{o}C is added to each column and centrifuged at 13,000 rpm for 1 minute. The genomic DNA was quantitated (as below) and stored at -20^{o}C for later use.

2.2.4 DNA quantitation

Agarose gel visualisation under (UV light) was used to estimate the quantity and approximate size (quantity) of the DNA. 5 µL of each sample of isolated DNA was loaded onto a 1% Agarose Molecular Grade (Bioline) gel in 0.5 x TBE buffer alongside a range of uncut lambda DNA standards containing 500 to 25 ng DNA. The DNA was stained by adding (2.5 µL of a 10mg/mL) ethidium bromide before pouring and quantification of the DNA was achieved by comparing the intensity of the DNA bands from the DNA extraction with the intensity of the bands from the lambda DNA standards. Approximate DNA loading (± 20ng) can be obtained by comparing band intensities by the eye. After quantitation DNA samples were diluted to 10ng/µL for PCR.

2.2.5 Microsatellite development

Development of microsatellites libraries was undertaken based on the method of Edwards *et al.,* (1996). The technique involves the hybridisation of restriction digested and PCR amplified genomic DNA to small filters carrying simple sequence repeat oligonucleotides (SSRs), followed by the elution and amplification. Rather than cloning, a mixture of enriched libraries was pyrosequenced (Roche 454).

The basic approach is given in Basu *et al.,* (2007). Sequences containing microsatellites repeat motifs were identified using the MISA.pl Perl script. Primers were then designed flanking the motifs with the aid of the Primer3 web interface program (http://fokker.wi.mit.edu/primer3/input.htm) (Roven and Skaletsky, 2000). A total of 75 primer sets (Appendix 2) were designed, PCR amplification and optimum annealing temperatures were determined. The following criteria were used for primer design: primer length of 18-27, GC

content 20-80, Tm 57-63^{o}C, product size 70-300 bases. Primers were designed and synthesized by MWG Eurofins. Microsatellites were not directly labelled with WellRed dyes from Beckman Coulter, but they were labelled using a three-primer 'tagged' reaction (Schuelke, 2000).

2.2.6 PCR gradient optimisation for primer annealing temperature

The polymerase chain reaction (PCR) involves *in vitro* amplification of DNA through a series of three polymerization cycles, the DNA denaturation, primer templates annealing and DNA synthesis by thermostable DNA polymerase. Optimization of PCR involves testing a number of factors, such as annealing temperature (T_a), poor results showing multiple bands on agarose are reflected when the $T_a$is too low, even when $T_a$is too high the desired products quality is also reduced due to the poor annealing of primers (Rychlik *et al.,* 1990). Gradient PCR helps to identify the optimal annealing temperature for pairs of primers. The range of annealing temperatures over which amplification occurs also gives an indication of how reliable the primer pairs are in PCR.Seventy five primer pairs were screened and optimised for annealing temperature using the genomic DNA extracted from the 24 genotypes in (Table 2.1.2.1**)** this was done to ensure optimal primer performance and to identify the best primers for tagging.

The PCR reaction mixtures (20µL final volumes) that contained approximately 10ng (2µL) of template DNA were constructed as given below in 96-well plates (Thermo Scientific): PCR Buffer (New England BioLabs; includes $MgCl_2$ to 1.5mM final (2µL). 20 µM Forward primer (0.5µL), 20 µM Reverse primer (0.5µL). 10 mM dNTPs (0.4µL) (Promega corporation), *Thermus aquaticus* polymerase (Taq) (New England BioLabs) (0.2µL) and 14.4 µL of sterilized distilled water. The plate was briefly centrifuged at 3,700 rpm in an Eppendorf refrigerated centrifuge (5180) to bring down the contents and sealed with Thermowell® Sealing mat (Fisher Scientific). Amplification was carried out in a Thermo HybaidPCR gradient machine (Thermo Hybaid Express) programmed with the following cycling regime: 94^{o}C for 3 minutes, 35 cycles of 94^{o}C for 1 minute, 12 temperatures ranging between 45-60^{o}C for 1 minute, 72^{o}C for 1 minute and final extension at 72^{o}C for 10 minutes. The optimised annealing temperatures for each primer are shown in Appendix 2.

2.2.7 Gel electrophoresis of PCR products

5µL of 6x loading buffer (standard reagent) was added to each sample and gently mixed before being given few seconds spin at 3,700 rpm in an Eppendorf refrigerated centrifuge (5180). After amplification, the reaction products were analysed by gel electrophoresis alongside a 2-log ladder (New England Biolabs) on a 2 % agarose gel in 0.5 x TBE, with ethidium bromide (Promega corporation) (2.5 µL of a 10mg/stock added before pouring, using 26-well combs (Biorad Maxi gel, model). After running the gel at 90 Voltage for 45 min, it was visualised by illumination with UV light and images taken using a Biorad (Gel DOC 2000), and hard copies of images were printed on a thermal printer (Mitsubishi P91) for the analysis of bands. The optimal annealing temperatures were determined based on the strongest band intensity temperature, that the product was approximately the expected size and also that the amplification was reasonably clean with little track background.

2.2.8 Three primer systems

To reduce the costs and facilitate screening of large numbers of potential microsatellites, a three primer system was used. One of these primers carries a fluorescent label and these fluorescent labels are relatively expensive. To overcome this especially when using a large number of microsatellites for genotyping, Schuelke (2000) devised a three primer system method, whereby, a sequence-specific (M13) tail is added at the 5' of the forward primer to give the 'Tagged-Forward' primer. A sequence–specific reverse primer is used in the reaction, together with a labelled M13 sequence primer. The amount of the tagged forward primer should be roughly 1/10th of the reverse primer. The remaining $9/10^{th}$ of the forward reaction primer is made from the fluorescently-labelled M13 primer. The PCR conditions are set in such a way that during the early PCR cycles, the specific forward primer with its M13 (-21) sequence is incorporated into the accumulating PCR products. As the tagged forward primer is exhausted in the PCR reaction, the universal M13 takes over as the forward primer due to the PCR products now having this priming sequence and this incorporates the fluorescent dye into the final PCR product. The M13 (-21) primer genotyping protocol provides a cheaper way to use commercially available fluorescent labelled dye primers.

The ‘Tag’ CAC GAC GTT GTA AAA CGA C sequence was fused to each forward primer at the 5’ end that had optimised well in the first round of untagged priming. In this study 69 primers were tagged out of the 75. Example of Tagged Forward primer D1: 5’ CAC GAC GTT GTA AAA CGA CTG CTT CTT CAA GGA GGA AGT AAG T 3’ where the underlined sequence represents the M13 Universal Tag, while the rest of the sequence is the bambara groundnut specific microsatellite primer sequence. Tagged forward primers were ordered directly from MWG Eurofins. The main disadvantage of this approach is that extending the 5’ end of the forward primer with a non-specific sequence of 18bp can destabilize the reaction and make successful amplifications less likely (Basu *et al.,* 2007). Primers that showed multiple bands and or for which no clear amplification product occurred were not selected for further use.

2.2.9 PCR amplification of microsatellites

A total of 68 microsatellite markers in (Appendix 2) were assayed against 24 diverse bambara groundnut genotypes listed (Table 2.1.2.1), which were expected to be a reasonable representation of the diversity present in bambara and had differences in morphological traits and collection sites. The 24 accessions were selected from 17 clusters identified from the analysis of 223 bambara groundnut accessions, selective amplification was conducted using *Eco*RI and M*seI* primer, 10 AFLP primer combinations and one heterologous SSR primer were used (Singrun and Schenkel, 2003).The distribution of the 24 accessions among the 17 clusters is given on (Table 2.1.2.1).

The M13 labelling primer was chosen to label the PCR products with blue, green or black fluorescent dye (WellRED primers; Sigma). For the preparation of a 10x working stock, 10µL from the x1000 (200pmol/µL) stock was mixed with 990µL of dH_2O. This gave all working primer stocks for the three primer reactions as x10 (including M13). The following components were used in the polymerase reaction: dH_20 (11.4µL), PCR Buffer (2µL), 20 µM Forward primer (0.2µL), 20µM Reverse primer (2µL), M13 Tag (1.8 µL), 10 mM dNTPs (0.4µL) (NEB) Taq (0.2 µL), 10ng/µL bambara genomic DNA (2µL). After sample mixing primer reactions were dispensed into a 96 well plate and spun briefly in an Eppendorf 5180 refrigerated centrifuge for a few seconds at 3700 rpm. The plate

was then placed in the PCR System 9700 (Applied Biosystems), programmed with the following cycling regimes: 94^{o}C for 3 minutes, 35 cycles of (94^{o}C 1 min, selected primer annealing temperature for 1 min, 72^{o}C for 2 min) and final extension at 72^{o}C for 10 minutes. The PCR amplification products were checked using a 2% agarose gel before size analysis using a CEQ 8000 fragment analyser (Beckman coulter inc, USA).

2.3.0 Gel electrophoresis of PCR products of Tagged primers

After amplification, the reaction products were analysed by gel electrophoresis alongside a 2-log ladder (NEB) on a 2 % agarose gel as described in (section 2.2.7). After running the gel at 90 Voltage for 45 minutes, the gel was visualised by illumination with UV light and images taken for analysis of bands. The bands intensities were used to determine the amount of PCR product to pool for multiplexed fragment analysis on the Beckmann CEQ 8000. An individual sample which gave a strong amplification product is reduced in a pool, while one which gave weaker amplification is increased to give a better overall balance. PCR amplifications which did not work well were repeated.

2.3.1 Capillary electrophoresis

A single PCR product or a set of two to four PCR products were pooled together, PCR products used usually ranged from (2 – 6 µL) per primer depending on the intensity of bands recorded on agarose gel. The sample loading solution (SLS) (Beckman coulter Inc, Fullerton, USA) was mixed with the size standard (SS) (Beckman Coulter Inc, Fullerton, USA) in the ratio of 1:100 (v/v) and 25µL of the mix was loaded into each well in a new PCR plate. 4µl of the PCR product from the different genotypes was added to the SLS:SS mix and covered with a drop of mineral oil (Beckman Coulter, Inc Fullerton, USA). All PCR products (SSR fragments) were sized on CEQ^{TM} Genetic Analysis System with a 400 bp size standard. Fragments were analysed with the CEQ^{M} 8000 Fragments Analysis Software Version 8 (Beckman Coulter Inc., Fullerton, USA) the sizes were manually scored.

2.3.2 Analysis of microsatellites

Microsatellite's hypervariability, abundance and co-dominance has led to them being employed in many research fields such as population genetics, linkage map construction, quantitative trait locus (QTL) mapping and molecular marker-assisted selection (MAS). When scoring microsatellites a number of errors can occur, the most common being those due to stuttering, large-allele dropout and null alleles (Bonin *et al.,* 2004). The occurrence of genotypic errors in the data can be limited by undertaking some mitigating measures. In a protocol for estimating error rates it is recommended that these measures be systematically reported to attest the reliability of published genotyping studies (Pompanon *et al*., 2005). All types of molecular markers are prone to genotyping error and they occur when the observed genotype of an individual does not correspond to the true genotype (Bonin *et al.*, 2004).

2.4 Potential genotyping errors and some mitigating strategies in microsatellite analysis

2.4.1 DNA degradation

The quality of DNA can deteriorate during sampling, extraction and during storage. A low number of DNA molecules in an extract due to extreme dilution or degradation leads to low numbers of intact molecules of DNA template for amplification and this favours allelic drop out and false alleles (Pompano *et al*, 2005). To minimise DNA degradation DNA stocks were stored in TE buffer and stored at -20^{o}C. Poor quality DNA produces poor amplification of PCR products, which could lead to some missing data. In this study the quality of DNA was tested by analysis on a 1% agarose gel electrophoresis with ethidium bromide, and stocks were diluted to 10 ng for working stock for all samples.

2.4.2 PCR based sources of error

PCR inhibitors in DNA preparations can contribute to genotyping errors. Low quality reagents, high temperature and high concentrations of PCR products have been reported to cause allelic dropout (Pompano *et al*., 2005). Marker assays were conducted using standard protocols for bambara groundnut of (Basu *et al.,* 2007). Microsatellites were optimised through the use of an annealing temperature

gradient from (45 ^{o}C to 60^{o}C; Hybaid PCR Express) and the optimal temperature was determined and used in the amplification experiments. Markers were screened individually using the blue M13 WellRed Dye to provide information on peak patterns and size ranges of alleles. This helps to avoid unidentified large-allele dropout in multiplexing, identify primers with difficult to interpret patterns and helps to spot SSRs which are amplifying from more than one locus. Multiplexing combinations of PCR products were as described in (section 2.3.0). The amount of PCR product used in a pooled sample was adjusted based on how strong the signal was after capillary analysis for the single samples. Generally, twice as much D3 (green) labelled product was added as D4 (blue) labelled product. It is during the process of identification isolation and amplification by PCR that some errors, such as null alleles, stuttering and large allele drop out occur (Van Oosterhout *et al.,* 2004).

2.4.3 Interpretation of capillary electrophoresis

Agarose and slab gel polyacrylamide gel electrophoresis which were widely used for microsatellites have limitations, particularly in terms of accurate sizing of alleles (Wang *et al.,* 2009). The CEQ^{TM} 8000 (CEQ 8000: Genetic Analysis System, Beckman Coulter, USA) provides automated and accurate estimates of allele sizes, together with the use of a combination of three fluorescently labelled primers (and a fourth Dye for ladder) (Hirst and Illand, 2001). While the CEQ^{TM}8000 contains automated binning wizard software which can be used to determine the sizes of alleles of markers, the allele size determination was done manually to avoid errors that can be brought about by the automated sizing, which may not differentiate between stutters and true peaks. Visual inspection of electrophoretic patterns is highly recommended during screening of markers in order to solve problems of stutter patterns, low height large alleles, which may not be detected by automatic automation. Dewoody *et al.,* (2006) recommend the use of both automated calling software and human inspection.

2.4.4 Spectral overlap

Primers were fluorescently labelled and PCR products were multiplexed with a general guide that D4 (blue) dye labelled products are diluted more than the D3 (green) or D2 (black) labelled PCR products. Despite that, spectral overlaps

sometimes create false peaks as shown by figure 2.4.4. The genuinely labelled D4 (blue) labelled PCR product, has bled through into the spectrum of the D3 (green) labelled PCR product leading to the formation of false peak. This problem was resolved by multiplexing PCR products with expected size differences, and scoring multiplexed PCR products simultaneously so that false peaks can be identified. Generally, the false signal is at a far lower intensity than the genuine signal, so can be distinguished without major problems.

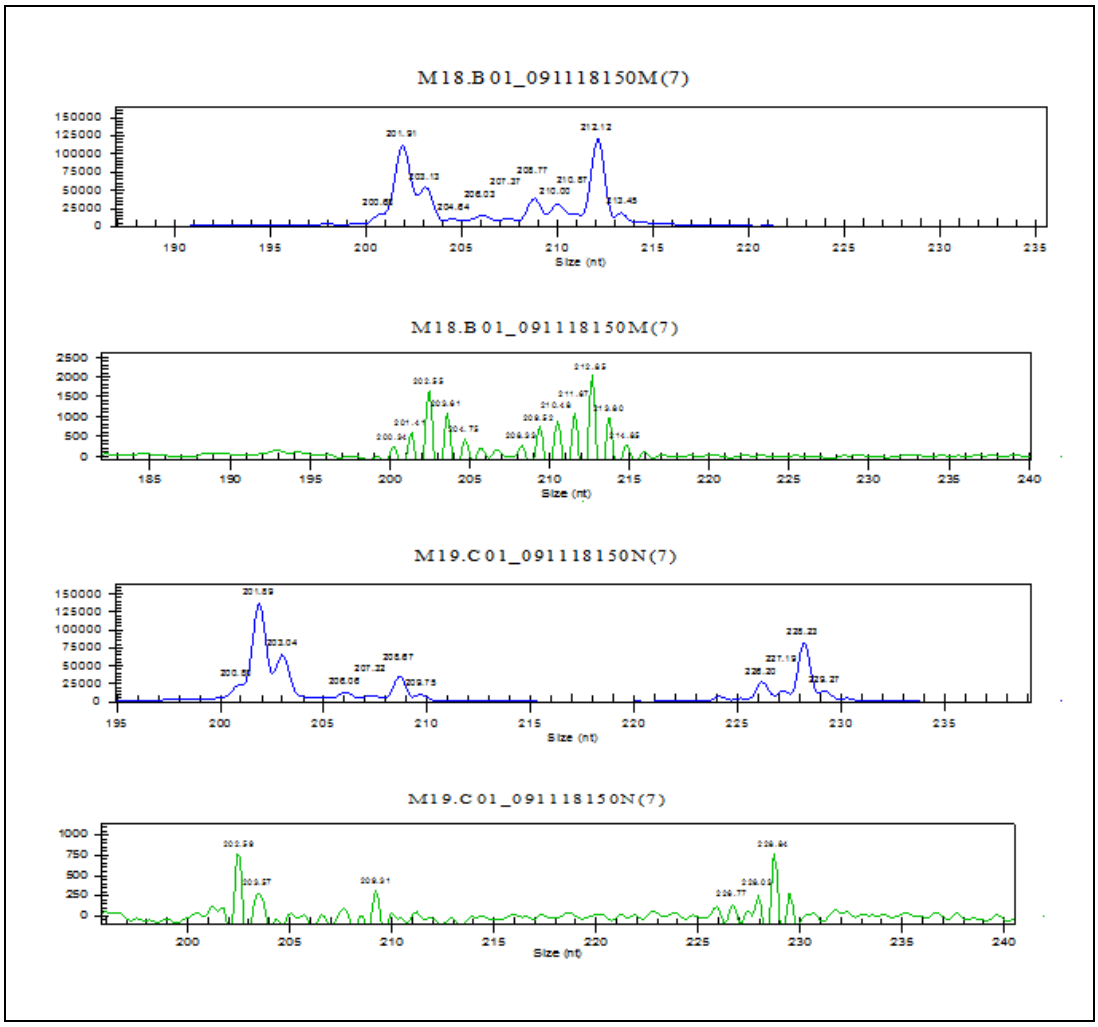


Figure 2.4.4: A pair of capillary electrophoresis traces of PCR products for blue labelled. The genuinely labelled blue PCR products, has bled through into spectrum of the green labelled PCR product and false peaks are shown for sample M18 and M19 (green).

2.4.5 Stutter and A-addition

Most loci tend to produce ‘stutter’ bands due to slipping of the polymerase with respect to template during the *Taq* polymerase extension step. In addition, *Taq* polymerase also has a tendency to add a non-template adenine to the 3’ end of the newly synthesised strand (Pompano *et al.,* 2005; Bonin *et al.,* 2007). Interpreting stutter loci can be difficult since heterozygotes can be scored as homozygote for difficult to interpret alleles which are shallow and contain many stutter alleles. This consistent mistyping will bias allele frequencies (Dewoody *et al.,* 2006). Figure 2.4.5; demonstrates one of the potential errors that can occur due to stutter bands when two alleles with different sizes overlap. Sample at PR 45-H. H11 in the upper pane and sample E44.E 05 in the lower pane are not easy to interpret on their own. The alleles appear to be heterozygotes with complex stutter bands, but it becomes clearer when compared with other samples from the same locus with similar pattern and shapes. For sample PR 45-H. H11 it becomes clear that it is a heterozygote when analysing it together with sample PR45-G-G11, while sample E-44.E05 is clearly resolved to be a heterozygote when compared with E-47.G05 and E51.H05.

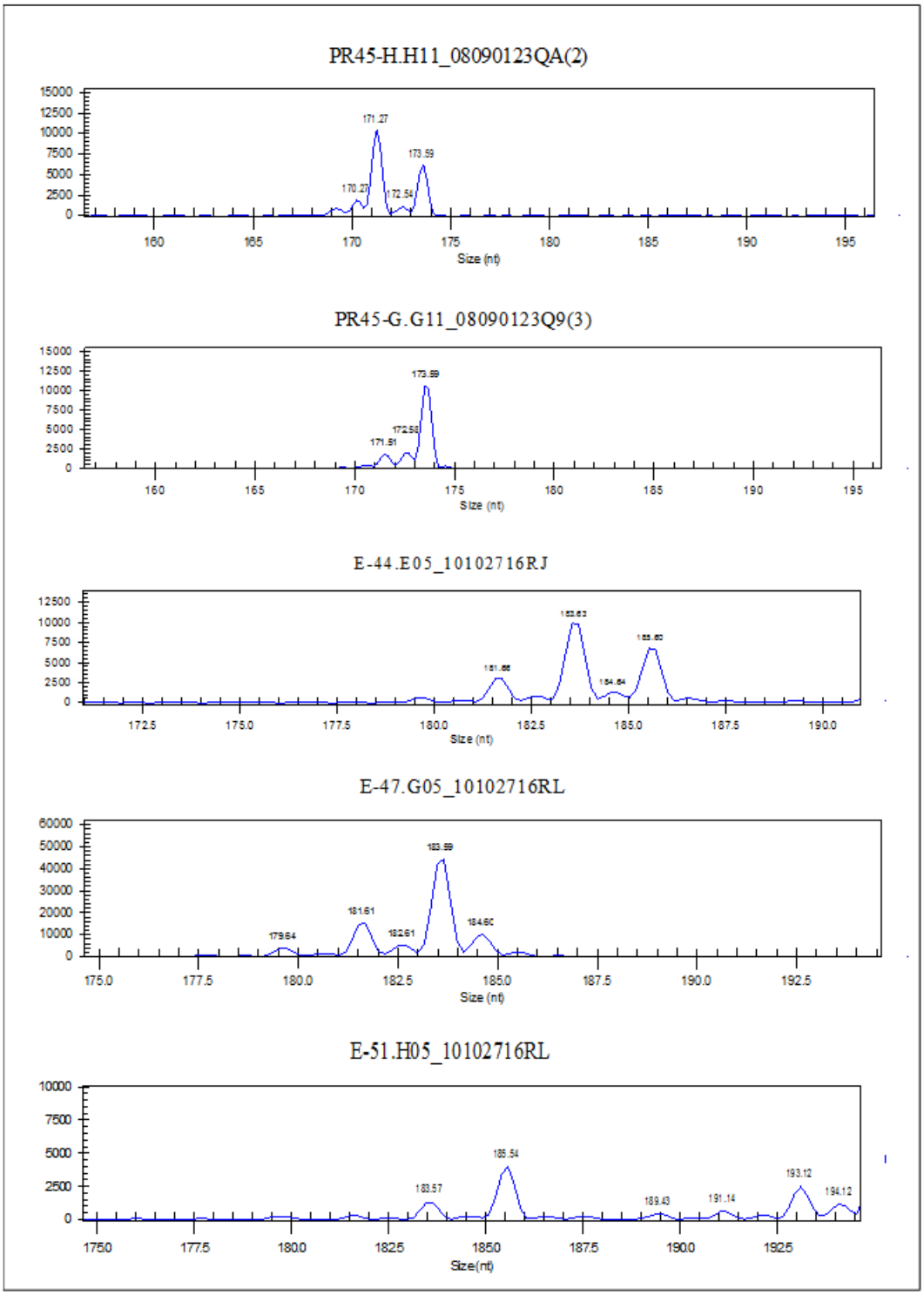


Figure 2.4.5: Capillary electrophoresis showing a potential scoring error due to the effects of stutter band and overlap on sample PR 45-H. H11 and 44.E05

2.4.6 Short allele dominance

Short allele dominance (large allele dropout) when not detected can lead to a decrease in sample heterozygosity in microsatellites analysis. Large allele dropout occurs when during amplification smaller alleles amplify better than the larger alleles and the larger allele occasionally fails to appear altogether. This phenomenon can be prevalent in loci with large differences in allele sizes (Dewoody *et al,* 2006). Two examples of short allele dominance are shown in figure 2.4.6 for Primer 15 and Primer 42.

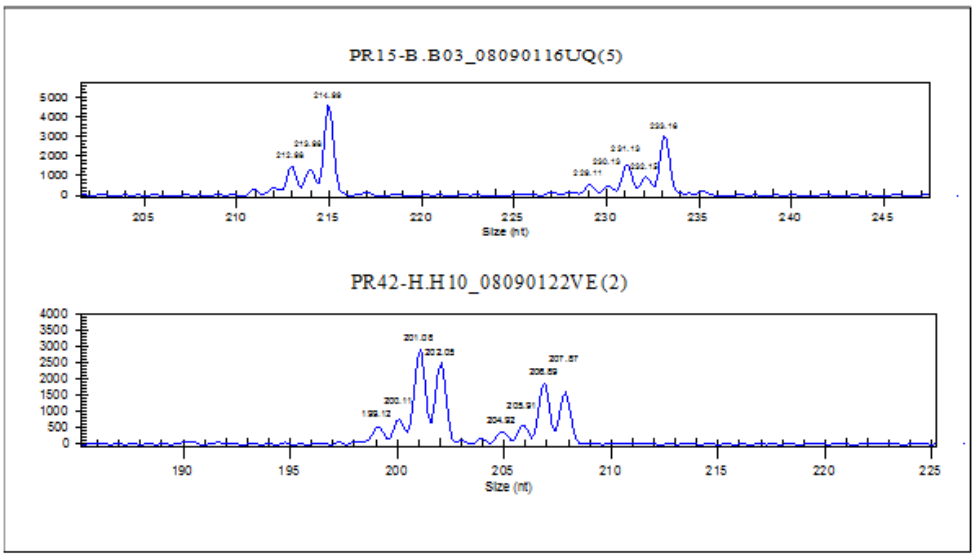


Figure 2.4.6: Capillary electrophoresis showing limited short allele dominance for marker PR 15 top and marker PR 42 bottom, since both are clearly visible and complete drop out did not occur, correct calling of the peaks could be done.

2.4.7 Allele size binning (Automated binning)

Binning is sorting allele lengths into discrete classes and there could be errors associated with this process. Several strategies are in place to undertake allele calling, such as comparing raw data to the database of the expected length and assign it to the closest data in automated binning and this has been found not to be a suitable approach in non-model species with no reference data set, such as bambara groundnut (Amos *et al.,* 2007).

Automated binning has been resolved with the software FLEXIBIN which uses least square minimization and allows allelic drift (Amos *et al.,* 2007). The common practice of rounding to the nearest whole number usually result in miscalls and most likely under estimation of allelic richness. In addition the issue of 'allelic drift' also makes it difficult to undertake automatic binning of alleles (Matschiner and Salzburger, 2009). Allelic drift is a source of some errors because it is tendency of true allele bins to display a slightly different value from the known repeat length (Idury and Cardon, 1997). A fixed repeat length is set so that only allele bins with a specific difference are allowed. Each dinucleotide repeat unit contributes an effective repeat unit of length in the range of 1.7 to 2.3bp (Amos *et al.,* 2007; Idury and Cardon, 1997).

Called alleles can be binned using Flexibin (Amos *et al,* 2007) to produce a graphical output that allows any potential problems to be identified and rectified. Figure 2.4.9 shows marker 16, which has a clear discrete allele distribution. For using microsatellite genotyping, allele sizes should be whole numbers but the genotype software (CEQ 8000) creates output to two decimals, which is partly a reflection of the effect of different base composition having slightly different molecular weights. Their conversion to integer values or 'alleles' poses a potential danger of mis-typing. The usual way of rounding off integers (i.e. <X.4999 = X.0 and Y.5000 = Y+1) was not followed since it could introduce errors. In practice, the calling needs to reflect the movement of the microsatellite 'shape' up and down the size range.

An example of a potential source of error in rounding off is illustrated in figure 2.4.7. Sample D92 would be called differently at 194 from the rest of the samples which would be recorded as 195 for D95 and D99 on the basis of rounding, but in this case all the alleles were recorded as 195. Size calling was done using FLEXIBIN (Amos *et al.,* 2007), which forces all alleles into a one base pair periodicity.

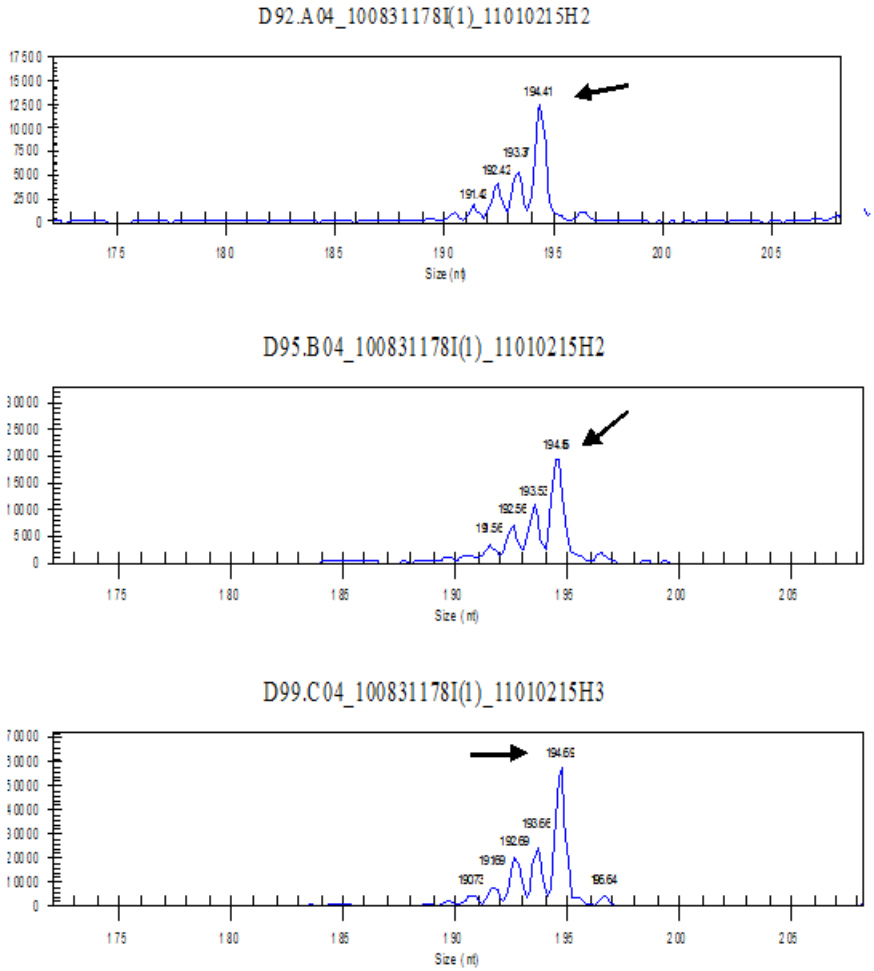


Figure 2.4.7: Capillary electrophoresis showing potential sources of mistyping errors due to rounding off alleles during binning. Analysis based on CEQ 8000 software.

Another potential source of genotyping error is illustrated in figure 2.4.8. The peak shapes that make up the microsatellite were carefully observed to set up a standard way to identify genuine peaks and to correct allele calling, based on the shape of peak, their height and size ranges. Sample I3 B09 has a higher peak with a smaller recording of 211.99 it has a similar pattern to sample I20E09 which has lower peak with a recording of 212.03. This difference could lead to the samples called differently, but looking at sample I40G09 it gives a characteristic shape and height of these alleles and all are called at 213. A similar potential genotyping error is observed in sample L21 and L23 with a call size of 257 different from L24, but their similar shape and height led to them being called at 257.

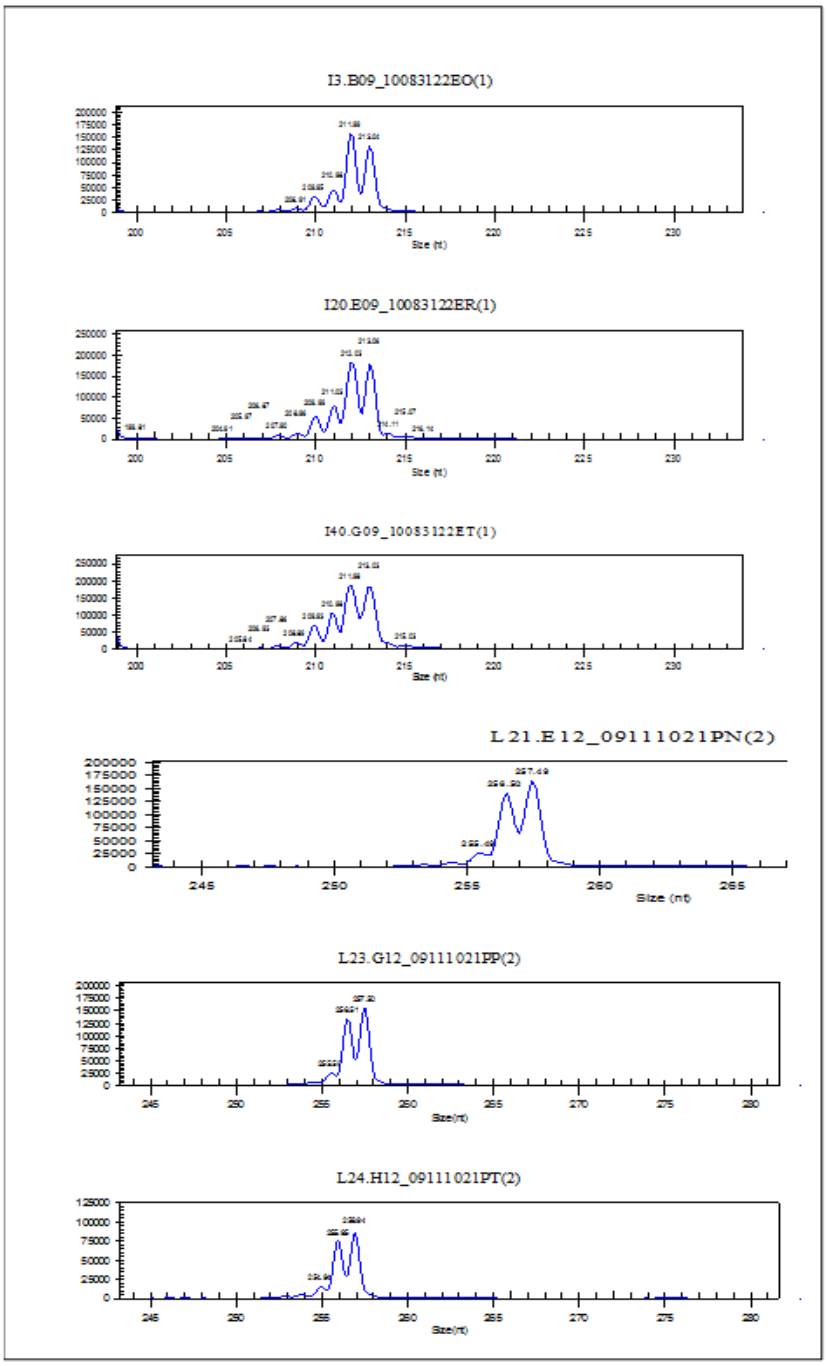


Figure 2.4.8: Capillary electrophoresis showing some potential miscalling errors, therefore the use of allele shapes, their height and size ranges are set as standard way to identify genuine peaks for correct allele calling.

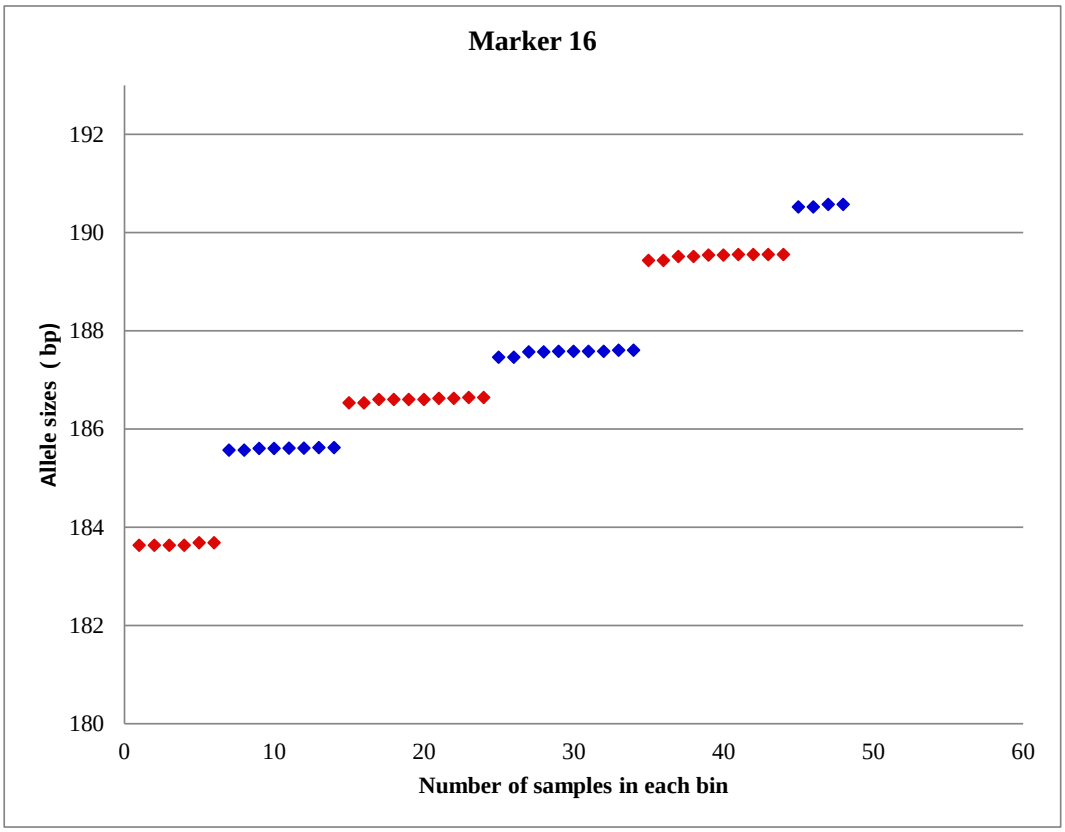


Figure 2.4.9: A graphical output of the cumulative allele length for marker 16, which illustrates an example of an accurately binned marker with clearly defined colours for different alleles as red and blue. The analysis was conducted with FLEXIBIN (Automated binning) using a one unit repeat**.**

The automated binning for marker 16 shows measures of allele size ranging from 183.65 to 190.51 grouped into 6 group repeats (Figure 2.4.9). A Summary of Flexibin analysis for marker 16 estimated repeat length, standard deviation and counts of repeats of each length are given in (Table 2.3.1). The estimated repeat length summary for each allele and the adjustment factors are listed in appendix 3.

Table 2.3.1: Summary of Flexibin analysis for marker 16, showing repeat length, standard deviation and count of each repeat length

Repeats	Length	Mean bp	s.d.	Count
3	183.65	183.65	0.026	6
5	185.61	185.60	0.020	8
6	186.59	186.60	0.039	10
7	187.57	187.56	0.053	10
9	189.53	189.52	0.048	10
10	190.51	190.55	0.029	4

bp= base pairs sd = standard deviation

2.4.8 Deviation from Hardy Weinberg equilibrium

Testing for Hardy-Weinberg equilibrium has become an important quality control in genetic data under the assumption that a high error rate will generate some disequilibrium. However other causes lead to disequilibrium, including selection, inbreeding and population admixtures through migration or fusion. Genotyping errors are another primary suspect in any observed deviations from HWE, and if genotypic error can be ruled out, other possibilities such as admixture should be investigated (Chen e*t al*, 2005). According to (Pompano *et al*., 2005) errors can cause disequilibrium, such as null alleles and allelic dropout.

2.5 Data analysis

2.5.1 Data analysis for microsatellites, development and characterisation

2.5.1.1 Microsatellites marker analysis

A total of 68 microsatellite markers in (Appendix 2) were assayed against 24 diverse bambara groundnut genotypes listed (Table 2.1.2.1). A summary of SSR statistics such as number of alleles, observed heterozygosity (Ho), expected heterozygosity (He), polymorphic information content and inbreeding coefficient (*f*) for each locus were computed using the program PowerMarker version 3.25 (Lui and Muse, 2005).All alleles were binary coded as 1 or 0 for their presence or absence in each genotype and used for data analysis.

The proportion of alleles shared between two genotypes averaged over loci was used as a measure of similarity for both markers(DArT and SSR) based on Nei and Li, (1979)similarity coefficient. The estimation was based on the formula*:* $GS_{ij} = 2N_{ij}/(Ni + Nj)$, where *Nij* represents the number of fragments shared by accession *i* and *j*, *Ni* represents amplified fragments in sample *i* and *Nj* represents amplified fragments in sample *j* (Nei and Li, 1979).

2.5.1.2 Principal component analysis (PCO)

To examine the genetic relationship between and within all individual landraces of bambara groundnut, principal coordinate analysis (PCoA) was used on the data set using multivariate statistical package (MVSP) (Kovach, 2006). The ordination does not make any assumptions about the distribution of variates or the population genetics of the population (Kloda *et al.*, 2007).

2.5.1.3 Cluster analysis

A similarity matrix produced was used to generate dendrograms based on the unweighted pair group method with arithmetic averages (UPGMA) cluster analysis was performed using multivariate statistical analysis (MVSP) (Kovach, 2006), and dendrogram were produced to show the similarities and differences between bambara groundnut genotypes. The cophenetic correlation between the genetic similarity and dendrogram generated was estimated to validate the relation

of the original similarity estimates and the binary data matrix analysed using NTSYS pc version 2.1 (Rohlf, 2000). Cluster analysis was also conducted using Winboot (Yap and Nelson, 1996) bootstrap analysis with 1000 replications.

2.5.1.4 Comparison of DArT and SSR genetic estimates

The genetic similarity obtained from DArT and SSR were compared by measuring the degree of correlation between them using the Matrix correlation correspondence test Mantel Z statistics based on 1000 permutations. The computations were conducted on NTSYSversion 2.1 (Rohlf, 2000).The correlations between the similarity matrices was also analysed based on Pearson product-moment correlation coefficient and Spearman's rank correlation coefficient using SPSS version 16.0. These comparisons were conducted to investigate whether there are any similarities between the genetic distance estimates generated by these markers.

2.5.2 Population structure and genetic diversity of bambara groundnut

2.5.2.1 Estimation of genetic diversity in the population

For population structure analysis of 123 bambara groundnut accessions, analysis was based on 12 markers in (Appendix 2). The standard parameters of genetic diversityas described in (section 2.5.1.1) were analysed based on PowerMarker Version 3.25 (Lui and Muse, 2005).

2.5.2.2 Estimation of genetic diversity within and among bambara groundnut populations

To evaluate the relationships between the 123 bambara groundnut accessions, Principal Coordinate Analysis (PCoA) and cluster analysis were employed based on the binary matrices that were generated for the presence or absence of alleles at each locus. The PCoA reveals the major components of molecular differentiation. The accessions were also colour and shape coded according to origin to help reveal any relations between the geographical location and genetic differentiation present in each dataset for PCoA.

Another measure of genetic variation in a population is gene diversity, sometimes referred to as average heterozygosity; however the two genetic measures are not

identical. Gene diversity measures the frequency of alleles at a gene locus, while average heterozygosity estimates the mean proportion of heterozygosity over all loci studied (Bergmann and Ruetz, 1991).Inbred populations show few heterozygotes, but mostly different homozygotes, thus the use of gene diversity estimates is more appropriate (Weir, 1990). Genetic diversity estimate gene diversity per locus and the data is calculated from the sample andestimated using unbiased estimator (Nei, 1987) on FSTAT version 2.9.3.

Different numbers of genotypes/samples were assayed from various countries. The estimate of observed number of alleles in a sample is dependent on the sample size. This problem was resolved through the use of FSTAT software to calculate the allelic richness in each population based on smallest number of individual samples (Leberg, 2002). The program estimates allelic richness (R_s) independent of the sample sizes, and this allows a comparison of genetic diversity between populations with different sample sizes. It estimates the expected number of alleles in a sub-sample of 2n genes, given that 2N have been sampled ($N \geq n$). In FSTAT, n is fixed as the smallest number of individuals typed in a sample

$$R_S = \sum_{i=1}^{n} \left[\frac{(2N - Ni)}{\left(\frac{2N}{2n} \right)} \right]$$

where *Ni* is the number of alleles of type *i* among the 2N gene (Goudet, 2001)

2.5.2.3 Estimation of population structure

To quantify the structure of the populations F-statistics, F_{ST} (Wright, 1978) was calculated using Arlequin version 3.1 based on Weir and Cockerham, (1984) and pairwise genetic distance among populations (Excoffier *et al*, 2005) were generated. The significance threshold of F_{ST} was generated by 1000 permutation testing to get an unbiased P-value for the test data.

To investigate the genetic structure of bambara groundnut landraces, analysis of molecular variance (AMOVA) was conducted using Arlequin version 3.1. The accessions were grouped into a three level hierarchy according to the classification based on PCoA structure analysis in figure 5.1.0.

2.5.3 Genetic diversity of bambara groundnut based on SSR markers and the comparison with morpho-agronomic characters

2.5.3.1 Polymorphism of microsatellites in bambara groundnut

To determine the genetic relationships within and between populations of bambara groundnut samples, three plants per landrace were used (Table 2.1.2.2). All three samples of the 35 accessions were counted as individual cases for the construction of a binary matrix, scored as presence (1) or (0) for absence for each possible allele to make a total of 105 samples. The 0/1 matrix was used for the calculation of genetic distances and the generation of cluster data to determine how the selected bambara groundnut were related.

2.5.3.2 Principal component (PCO) and cluster analysis

The matrix generated with genetic similarity estimates was used to cluster genotypes and the generation of principal components and principal coordinate analysis as described in section 2.5.1.2 and 2.5.1.3, and to examine the genetic relationship between and among all individual genotypes based on MVSP program (Kovach, 2006).

2.5.3.3 Analysis of Molecular Variance (AMOVA)

Since three genotypes per landrace were studied, analysis of genetic diversity within each landrace was conducted using an AMOVA analysis with Arlequin 3.5 (Excoffier and Lishcher, 2010). The total variance among genotypes was partitioned into variance among populations, among individuals within populations and within populations; the populations were defined based on the two groups in figure 6.2.1. The significance of the partitioning of the genetic variance components was tested using 1000 permutations.

2.5.3.4 Morphological data analysis

Thirty five genotypes and 34 bambaragroundnut lines selected (Table 2.1.2.2) were studied for variation of morphological and agronomic traits following the IPGRI descriptors (IITA, BAMNET, 2000) (section 2.6.6; Table 2.6.6).The description of morpho-agronomic data analysis and the generation of cluster analysis are described in (section 2.6.6 and 2.7.8).

2.5.3.5 Comparison of SSR marker and morphological marker data

For the comparison of morpho-agronomic and molecular (SSR)markers, PCO analysis, cluster analysis and correlation matrix was conducted on both data set based on 20 SSR markers and 34 and 37 morpho-agronomic traits recorded in the agronomy bay and field experiment respectively.

Nei's 1972 genetic distance was estimated for SSR markers while the Euclidean distances were estimated for morphological marker. The estimated means were tested by means of Matrix correspondence test (Mantel, 1967), which uses 1,000 permutations to estimate the correlations significance between distance matrices and this was calculated using the NTSYS pc software. Simple Pearson product-moment coefficient correlation and Spearman rank's coefficient correlation were used to test the correlations based on SPSS version 16. In addition the results for cluster and PCO analysis were compared to identify any similarities between the two marker types.

2.6 Morpho-agronomic characterisation and evaluation of bambara groundnut

2.6.1 Introduction:

Among the 119 accessions planted in the agronomy bay (greenhouse), morpho-agronomic assessment on the germplasm was conducted on three individuals of the 35 bambara groundnut genotypes listed on (Table 2.1.2.2).A field work experiment was conducted on 34 lines derived from seed from single plants selected from the greenhouse (Table 2.1.2.2.). The details of the experiment procedures are described below.

2.6.2 Glasshouse experiment

The experiment was set up in an unheated agronomy bay glasshouse at the University of Nottingham, School of Biosciences, Sutton Bonington, in the United Kingdom. The dimension of the glasshouse is 10.1 m x 4.7 metres wide and 2.3 metres high. The glasshouse is made up of conventional aluminium and glass and it had vent for manual regulation of heat inside the glasshouse.

2.6.3 Plant materials

Eighty seven bambara groundnut accessions from the 119 accessions were sourced from the International Institute of Tropical Agriculture (IITA; Nigeria), while 27 were from the University of Nottingham and five were brought from Botswana.

2.6.4 Experimental design

Seed bed preparation was done by digging and raking the soil and applying 290 kg/ha of Ammonium Nitrate fertilizer. Soil was raked to level, and the seedbed was covered with black plastic to suppress weed growth before planting. The glasshouse was fitted with a Tiny Tag (Gemini Data Loggers, UK) to measure temperature and relative humidity every 10 minutes for the entire duration of the experiment.

Accessions were planted in a randomised complete block design replicated three times. Two seeds were sown per hole at a depth of 5 cm and spacing of 30 cm x 30 cm (inter- and intra-row) giving 15 plants per row. Seeds were surface sterilised with 15% by volume NaClO (Sodium hypochlorite) for 15 minutes and rinsed 3 times in sterile water before sowing which was done on the 27May 2008 and later thinned to 1 plant 21 days after sowing. Each landrace was represented once per replication, thus each replication was used in emergence counts.

2.6.5 Crop management

2.6.5.1 Photoperiod

Bambara groundnut is a short day length crop, and the glasshouse was receiving natural long day light, that could adversely affect the pod formation of the crop. The crop received natural daylight with no supplementary lighting. Day length was controlled at 12hrs per day by covering with a black polythene screen fitted over a metal frame above the crop starting at 2000hrs and uncover at 0800hrs to maintain a 12 hrs photoperiod, from 20June 2008.

2.6.5.2 Crop protection

Phytoseilus persimilis was used as a biological pest control against red spider mite (*Tetranychus urticae*) every two weeks.

2.6.5.3 Irrigation

The trickle irrigation system was used which consists of PVC micro-porous tubing placed at each row. Crops received non limiting moisture on a weekly basis starting from day 0 to 112 days after sowing and approximately a total of 330 mm of water was supplied (Table 2.6.1)

Table 2.6.1: Amount of irrigation water (mm) applied in the bambara groundnut experiment in the agronomy bay (Glasshouse) expressed in days after sowing (DAS) for the duration of the experiment in 2008 season.

DAS	Amount (mm)
0	20
8	20
10	10
17	10
22	20
26	20
35	10
42	20
47	20
56	20
61	20
68	20
75	10
84	10
90	20
97	20
102	20
107	20
112	20
Total	330

2.6.5.4 Climatic factors

Air temperature and relative humidity were recorded every 10 minutes automatically on the tiny tag, the average minimum and maximum temperatures of 10.9^{o}C and 29.5^{o}C respectively were recorded with an average of 17.4^{o}C (Figure 2.6.5.1). The average relative humidity recording was 78.7 %, with a maximum and minimum of 95.6 % and 78.7 % respectively (Figure 2.6.5.2).

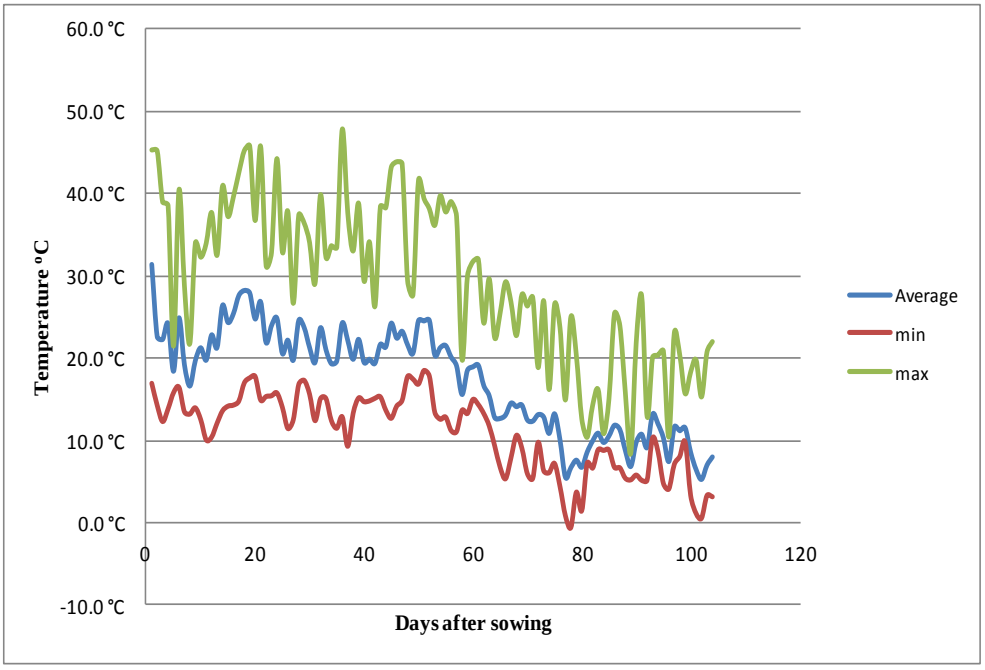


Figure 2.6.5.1: The maximum and minimum temperature in the agronomy bay (Glasshouse) experiment for the 119 bambara groundnut landraces grown in the 2008 season

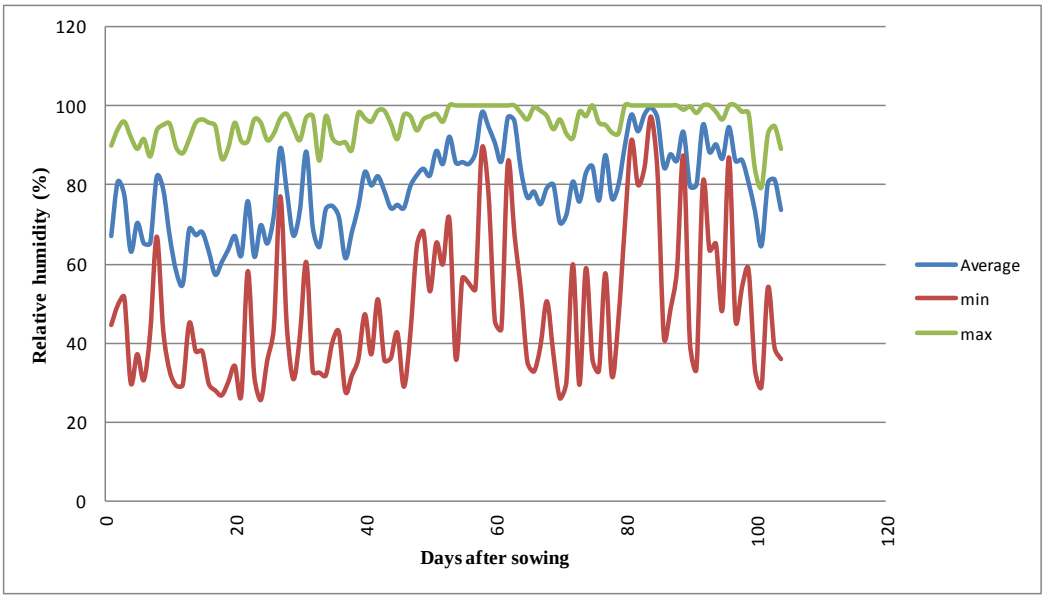


Figure 2.6.5.2 The maximum and minimum relative humidity in the agronomy bay (Glasshouse) experiment for the 119 bambara groundnut landraces grown in the 2008 season**.**

2.6.6 Morpho-agronomic traits measurements collected in the greenhouse

In the agronomy bay (green house) experiment 35 accessions (Table 2.1.2.2) that emerged from all three replications were followed through for data collection. The accessions were evaluated for 24 quantitative and 10 qualitative characters, according to the bambara groundnut descriptor list (IPGRI, IITA, BAMNET 2000)and measured both during vegetative growth and after harvesting (Table 2.6.6).

Additional characters measured were *days toseedling emergence* and *leaf area.* A seedling was considered to have emerged when the first true leaf become visible.Non-destructive *leaf area* assessment was determined in the glasshouse using the measurement of the middle leaflet width and length using the equation: Leaf Area = 0.74 x 3 x leaf number x (leaflet Length x Width x π/4)] developed by Deswarte (2001). The equation was confirmed by Cornelissen (2004) using the leaf area meter (LI-COR 3000), and was applied in field experiments in Swaziland by Edje and Sesay, (2003).

Days to maturity was observed with the yellowing and browning of leaves. The date of final harvest was based on the observation of leaf senescence, then *shoot dry weight* was measured by oven drying the above ground of selected plants.The final*pod yield* was determined from each single plant. The pods were oven dried at 37^{o}C for one week while the plant biomass was oven dried for 48 hrs at 72^{o}C.

2.6.6.1 Quantitative traits measurements in the green house

Data for the 24 quantitative traits consists of; *days to emergence, days to 50% flowering, number of leaves, plant spread*(Canopy) (cm), *leaflet length* (mm), *leaflet width* (mm), *plant height* (mm), *internode length* (mm), *petiole length* (mm), *petiole-internode ratio*, *petiolule length* (mm), *peduncle length* (mm) and *number of stems per plant*, were recorded at 10 weeks after sowing. Yield characters scored after harvest include; *number of pods per plant, pod length* (mm), *pod width* (mm), *pod dry weight* (g), *number of seeds per plant, seed weight per plant* (g), *seed length* (mm) and *seed width* (mm) (Table 2.6.6).

2.6.6.2 Qualitative traits measurements in the glasshouse

For qualitative data, individual plants were recorded per plot to represent each genotype. Thirteen qualitative characters recorded were for *testa colour, eye pattern, testa pattern, pod colour, pod texture, pod shape, seed shape, terminal leaflet colour, stress susceptibility, leaf shape* and *stem hairiness* and *leaf colour at germination* and *growth habit.*

In the agronomy bay experiment only 10 qualitative data were recorded with exception of *stem hairiness, leaf colour at germination*and *growth habit*, while in the field experiment all the 13 characters were recorded.

Table 2.6.6: Quantitative and qualitative traits recorded and brief description as listed from (IPGRI, 2000).

Characters	**Characters and description**
Days to emergence (DAE)	Number of days from sowing to when the first fully expanded leave appears in a plot
Days to 50% flowering (DAF)	Number of days from sowing to first flower opening on 50% of plants per plot
Number leaves per plant (LNO)	Total number of leaves per plant at 10 weeks after planting (WAP)
Plant spread (SPRD)	Widest point between two opposite points recorded at 10 WPA
Leaflet length (LL)	Length of median leaflet at the fourth node recorded at 10 WPA
leaflet width (LW)	Width of median leaflet at the fourth node recorded at 10 WPA
Leaf Area(LA)	Multiply leaflet width X length and number of leaves and use a formula at 10WPA
Plant height (PHT)	Measured from the ground level to the tip of the highest point recorded at 10WPA
Internode length (ITN)	Length of fourth internode of the longest stem, recorded at 10WPA
Petiole length (PTL)	Measured from the stem node to the junction of the three leaflets at the longest stem at the fourth node at 10WPA
Petiole-Internode ratio (PITN)	The ratio of the measurement of the petiole and internode
Petiolule length (PTLL)	Recorded on the base of the leaflet of the longest petiolule at the fourth node at 10WPA
Peduncle length (PNL)	Recorded on the fourth internode of the longest stem recorded at 10WPA
Number of stems (STEM)	Number of stems recorded from selected plants at 10WPA
Days to maturity (DAM)	Number of days from planting to maturity
Shoot dry weight (SDW)	Weight of above ground biomass of harvested plants
Number of pods per plant (POD)	Average of 5 plants recorded per plot recorded within two months of harvest
Pod dry weight (PDW)	Average weight of pods taken from 5 plants per plot within two months of harvest
Pod length (PODL)	Average length of pod taken from 5 plants per plot within two months of harvest
Pod width (PODW)	Average width of pod taken from 5 pods per plant within two months of harvest
Seeds per plant (SNO)	Average number of seeds taken from 5 plants per plot within two months of harvest
Seed length (SL)	Average of seed length from 5 seeds taken per plot
Seed width (SW)	Average seed width from 5 seeds taken per plot
Seed weight (SWE)	Average weight of seed from 5 plants taken with a plot after drying, within two months of harvest

2.7 Field work experiment in Botswana

2.7.1 Introduction

A field experiment was conducted at Botswana College of Agriculture (Notwane farm) Sebele, Botswana from 11 December 2008 to 11 May, 2009 (2008/2009). Detailed descriptions of the study site, experimental design, and crop management are given below.

2.7.2 Field site and experimental preparation

The field experiment was undertaken at Botswana College of Agriculture (Notwane farm) Sebele in Botswana approximately at latitude $24^{o}33'S$ and longitude $25^{o}54'E$, 994 metres above sea level. The analysis of soils in Sebele have been recorded as; shallow, ferruginous tropical soils, medium to coarse grain sands and sandy loams with a low water holding capacity and subject to crusting after heavy rains (Baker, 1987). After tractor ploughing and harrowing each plot was hand harrowed to make a fine seedbed and a basal application of single superphosphate at a rate of 25 P kg ha^{-1} of fertiliser was applied just before planting.

2.7.3 Plant material

Thirty four bambara groundnut lines, with seed selected from a single plant of 35 accessions planted in the glasshouse. One individual among the three plants selfed in the greenhouse was selected for field experiment (Table 2.1.2.2) except for landrace 49-Acc 793 from Kenya, which produced low number of seeds for the field experiment.

2.7.4 Experimental design

The design of the experiment was a randomised complete block design with three replicates, and each bambara groundnut line was assigned randomly to the plots with sowing done on the 11December 2008. Individual plot sizes were 3.2 m x 0.4 m, with 0.5 m guard row surrounding each experimental plot. Seeds were sown with 10 cm space between plants and later thinned to 30 cm between plants at 21 days after sowing to remain with 10 plants per row.

2.7.5 Crop management

2.7.5.1 Crop protection

During sowing a nematicide (Nemacur 10 GR, Bayer AG) was applied to each row at a rate of 1.5 gm^{-1} to prevent the infestation of root-knot nematodes. Plants were sprayed with insecticide Malathion 50% EC (*S*-1,2–bis (ethoxycarbonyl)ethyl*O,O*-dimethyl phosphorodithioate) and fungicide Eria (triazole, binzimidazole) (Republic of South Africa) using knapsack sprayer to control aphids and diseases as needed.

2.7.5.2 Irrigation

The watering regime for the experiment was synchronised with the BAMLINK project, which had the same planting date and connected through same source of water supply. Trickle irrigation system was used and PVC micro-porous tubing placed in each row. Irrigation was applied up to 79 DAS and a total amount of 156 mm of water was applied (Table 2.7.5). In addition the crops received a total amount of366.87 mm of rainwater, which was recorded through the Hobo Weather Station Data logger (Weather Tempcon L.t.d.) installed at the site, the rainfall distribution is shown on (Figure 2.7.5.1).

Table 2.7.5: Amount of irrigation water (mm) applied in the bambara groundnut experiment in the field experiment in Botswana, expressed in days after sowing (DAS) for the duration of the experiment in the 2008/ 2009 season

DAS	Amount (mm)
2	18
8	18
9	18
14	36
15	18
21	18
28	15
79	15
Total	156

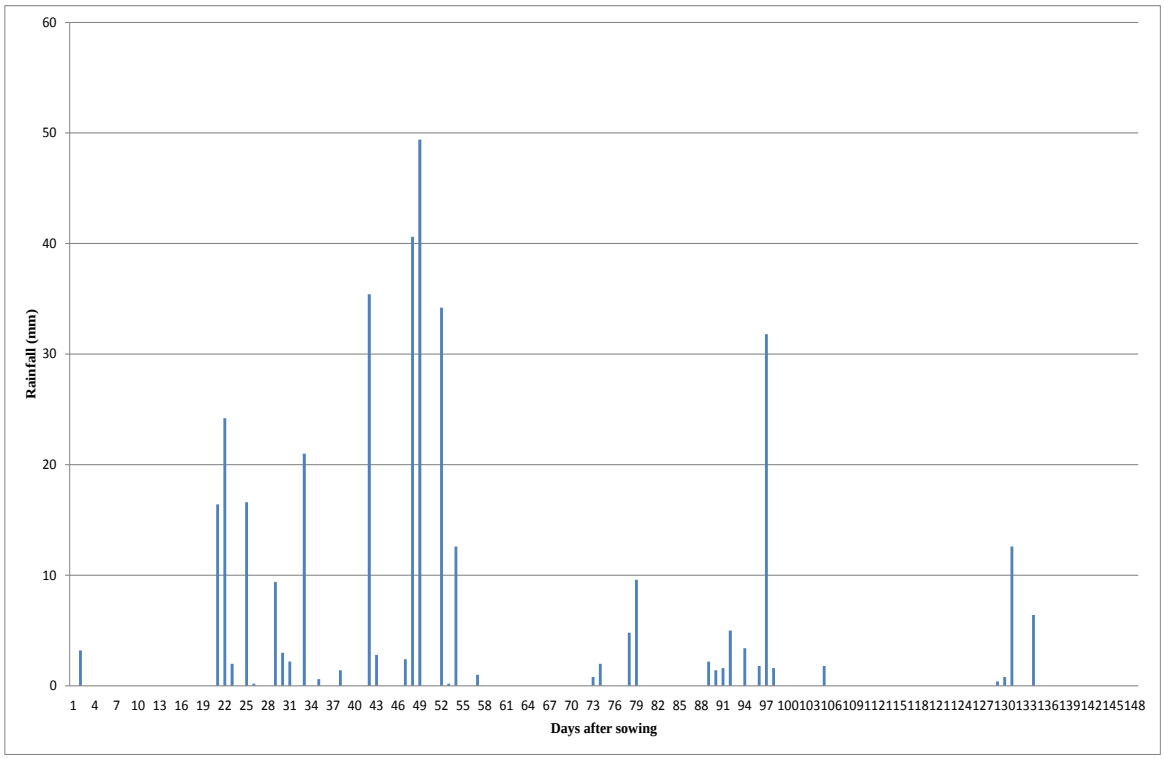


Figure 2.7.5.1: The amount and distribution of rainfall in bambara groundnut field experiment at (Notwane) Sebele, in 2008- 2009 season.

2.7.5.3 Climatic factors

A total amount of approximately 522 mm of moisture was received by the crop in the field including a combination of both applied moisture and precipitation received. The average temperature was 21.7^{o}C, while the minimum and maximum were 15.6^{o}C and 28.9 ^{o}C respectively as shown in (Figure 2.7.5.2). The relative humidity recorded an average of 76%, and a minimum and maximum of 41.6% and 98% respectively (Figure 2.7.5.3), and these were fairly similar to the recordings found in the agronomy bay experiment in UK.

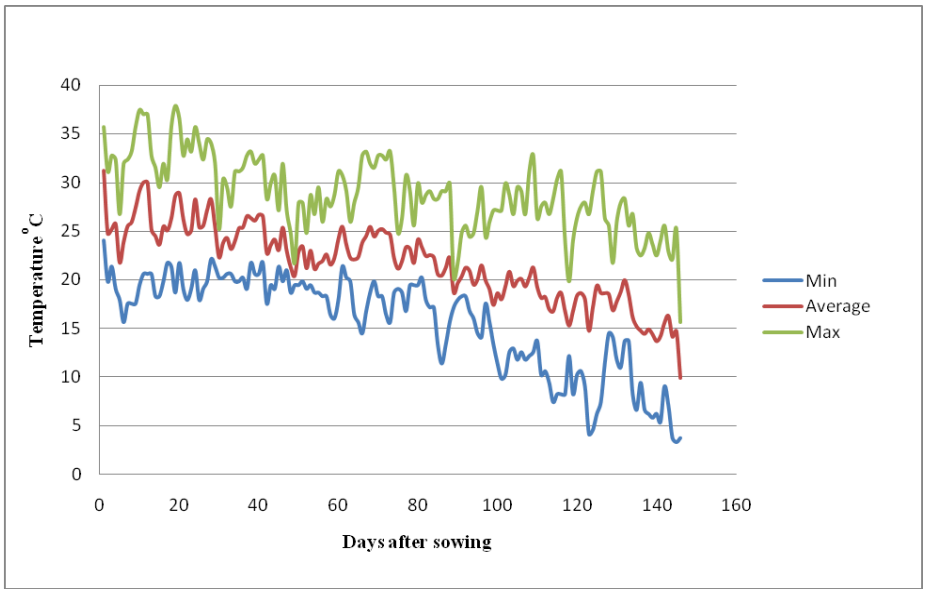


Figure 2.7.5.2: Maximum and minimum temperature in the field experiment for the 34 bambara groundnut landraces grown at (Notwane) Sebele in the 2008 - 2009 season.

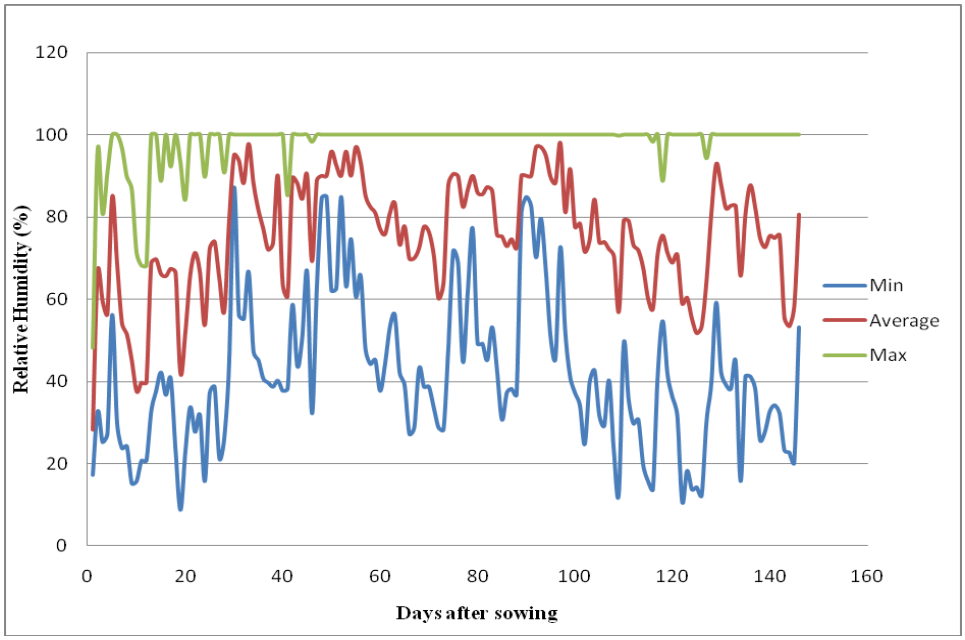


Figure 2.7.5.3: Maximum and minimum relative humidity in the field experiment for the 34 bambara groundnut landraces grown at (Notwane) Sebele in the 2008 - 2009 season.

2.7.6 Agro-morphological traits measurements in the field experiment

Data was collected from 15 plants per line, 5 plants from each replication. The number of pods and seeds per plant were recorded from each harvested plant. The date of final harvest was based on the observations of leaf senescence and the final pod yield was determined from five plants per replication. *Leaf area* (cm^2), measurements were determined in the field based on the middle leaflet width and length as described in (section 2.6.6).

2.7.7 Statistical analysis of agronomic traits

Similar statistical analysis was conducted for both agronomy house experiment and the field experiment. To identify any structures on the bambara groundnut landraces based on the phenotypic diversity, cluster and principle coordinate analysis (PCoA) were conducted. In order to identify characters that are contributing more to morphological diversity, principal component analysis (PCA) was undertaken and Eigenvalues were examined, and for selection of the best landraces heritability estimates were done, to try to identify characters which are expected to respond more to selection (Roy, 2000).

2.7.8 Data analysis of agronomic traits

2.7.8.1 *Descriptive characteristics*

Data analysis for all quantitative characters were subjected to analysis of variance (ANOVA) using the Genstat version 13.0 (Lawes Agricultural Trust, Rothamsted Experimental Station, UK) General linearized models (GLM) package to determine the statistical differences on the traits for the given genotypes. The mean values, ranges standard error of means and coefficient of variation were calculated on the 24 agro-morphological data.

The measurements for each trait for all the genotypes were standardized by subtracting the mean from respective traits and dividing by the standard deviation using Genstat version 13.0 in order to reduce the influence of the scale differences (Upadhyaya, 2003). The transformed values for each character had an average of zero and standard deviation of one and these values were used for principal component analysis and cluster analysis. The standardized data was used to

estimate the matrix of distances between all pairwise combinations of genotypes. The dendrograms were constructed from measurements of a combination of both the qualitative and quantitative characters (Hill *et al.*, 1998).

2.7.8.2 Principal component analysis

Principal component analysis (PCA) is a technique that summarises patterns of correlations among observed variables and reduces a large number of observed variables to a smaller number of components with several linear combinations called principal components (Tabachnick and Fidel, 2007) with each principle component or Eigenvalues being independent of other components. The importance of PCA is to extract maximum variance from the data set with each component. The new sets of transformed uncorrelated variables are close to the original variables but arranged in decreasing order of variance. PCA further enables plotting data in two dimensions to look at outliers, groups or clusters based on biological data (Chatfield and Collins, 1980). The quantitative and qualitative data were analysed using Principal Component Analysis in Genstat version 13, based on the correlation matrix, which basically gives traits equal weightings.

Principal Component analysis (PCA) was used to reveal traits that account for most variation between lines. The Eigenvalues ≥ 1 were selected and used to define the agro-morphological diversity. Principal component analysis was constructed using MVSP (Kovach, 2006).

2.7.8.3 Cluster analysis

Cluster analysis, basically aims to find groupings in a set of individuals, objects or units such that individuals within a group are similar to each other but individuals in a different group are dissimilar to others. The quantitative and the qualitative data were converted to binary data. For qualitative characters the absence of a trait was recorded as 0 and the presence of a trait as 1. For quantitative data genotypes that were significantly different were scored 1 and those not significantly different were scored 0. Then the unweighted pair group method with arithmetic averages (UPGMA) cluster analysis was performed using NTSYC version 2.1 (Rholf, 2000) using average linkage based on Euclidean distance and dendrograms

produced to show the similarities and differences between bambara groundnut genotypes.

2.7.8.4 Shannon-Weaver diversity

The Shannon-Weaver diversity index (H') of Shannon and Weaver (1949) was estimated to measure the phenotypic diversity between traits. In the calculations of Shannon Weaver index (H'), in order to avoid bias of the contribution of the individual trait the diversity of the mean, phenotypic range are arbitrarily divided into similar number of classes for both the glasshouse and field experiment (Engels, 1994) (Appendix 7)

The Shannon-Weaver diversity index (H') was calculated using the phenotypic frequencies to assess the phenotypic diversity for each trait. Where (H'= $\sum$ (*Pilnp)*where pi is the proportion of accessions in the ith class of an n- class character and n is the number of phenotypic classes for a character. Pi calculates the abundance of the given categories for each character, which is then multiplied by the natural log. The traits measurements were divided by their maximum values, log n, and normalised to keep the values between 0 and 1, on Genstat version 13 based on ECDIVERSITY procedure. The analysis was conducted on both the quantitative and qualitative traits.

2.7.8.5 Correlation coefficient

Correlation coefficient (*r*) was used in the study to determine inter-relations between all quantitative characters (Table 2.6.6).Correlation ranges between -1 and 1 and measures the extent to which two variable scores increase at the same time in positive correlation while in negative correlation the other goes down while the other increases. A correlation of 0 implies that there is no linear relationship between the variables. Pearson correlation was conducted using SPSS version 16.0 and the significance test for correlation coefficient was tested on a two-tailed test on the same program.

2.7.8.6 Quantitative variances

Data for each trait was subjected to analyses of variance to estimate the genetic variability of the selected genotypes and to partition the phenotypic variability into components due to genetic and environmental factors. Measures of variability

such as genotypic coefficient of variability (GCV), phenotypic coefficient of variability (PCV), broad sense heritability (h^2), and genetic advance (GA) based on percentage of the mean were estimated.

There are a number of methods available for estimating heritability (h^2), which includes using the resemblance among relatives, from generations derived from a cross between two pure-breeding lines, experimental mating designs and from components calculated from replication experiments (Hill *et al.*, 1998). The latter method was selected for use and each trait that was subjected to analyses of variance was used to estimate the genetic variability of the selected genotypes and to partition the phenotypic variability into components due to genetic and environmental factors. The genetic parameters were estimated using formulas adapted from Allard, (1960), Singh and Chadhary, (1985) as follows:

Vg = [Mean Square Genotype-Mean Square Error/*r*]

Vp = [Mean Square Genotype/*r*]

Ve = [Mean Square Error/*r*]

r is the number of replications

The Mean Square Genotype (MSG) and Mean Square Error (MSE) are variance components estimated as functions of the mean square estimates from ANOVA table. Mean square genotype (MSG): estimates genotypic variance, this value is observed variance among the line means, while mean square error (MSE) measures variance from plot residuals.

Phenotypic (PCV) and genotypic (GCV) coefficient are estimated using the following formulas

PCV = (√Vp/X) x 100

GCV = (√Vg/X) x 100

Vp represents the phenotypic variance: Vg represents the phenotypic variance, while X represents the mean.

Heritability (h^2B) expressed as the percentage of the ratio of the genotypic variance (Vg) to the phenotypic variance (Vp) was estimated based on the genotypic mean

Expected genetic advance (GA) was estimated using a formulaof Allard, (1960) as GA = K (Sp) h^2B, GA (as % of mean) = (GA/X) x100

where h^2B and Sp is the heritability ratio and the phenotypic standard deviation ($\sqrt{Vp}$) and K is a selection differential that varies depending on the selection intensity. In the present analysis 2.06 was considered for K, which is 5% selection intensity. The phenotypic standard deviations among traits were calculated using Genstat version 13.

2.7.8.7 Selection index (SI) and Duncan Multiple Range Test (DMRT)

For a development of new varieties it is important that selection of the best genotypes is conducted. Usually, it is those traits of interest or economic value that breeder select for. Plant breeders could decide to select for one or more traits at a time and this is referred to as multiple trait selection. And appropriate weight is given to each character, for example its heritability (h^2), genetic and phenotypic correlations between different characters of interest could be used. The component characters are then combined into a score, or a selection index (Falconer and Mackay, 1996). For multi-trait selection the classical selection index (I) proposed by Smith (1936) and Hazel (1943) which is a linear combination of traits of interest could be used with the formula:

$I = b_1x_1 + b_2x_2 + c_3x_3 + ... + b_nx_n$

where x_1, x_2, $x_3...x_n$ are the phenotypic performance of different traits, while b_1, b_2, b3, are relative weight attached to each traits.

The weights attached depend on the economic importance attached to traits depending on their heritability and correlation between various traits (Hill *et al.*, 1998).

In this study, a similar index was used, based on four characters of importance, *shoot dry weight*, *leaf area*, *seeds number plant* and *pod number per plant*.

Economic weight attached to these characters was based on the genetic advance (GA) (5% of the mean), which is described in section 2.7.8.6.

$SI= (X_1xW_1) + (X_2xW_2) + (X_3xW_3) + (X_nxW_n)$

where W_1, W_2, $W_3...W_n$ are the respective weights for each variable. Since the variables are measured in different units with large differences in magnitude, the variables were standardized with the following formula;

$Xi = (Xi - \mu)/$ st. Dev

X1 = *shoot dry weight*, X2 = *leaf area* X3 = *seed number plant* X4 = *pod number per plant.*

To identify lines which have a potential to produce higher yields in a Botswana environment based on four selected characters, the selection index was used with a weighting of the genetic advance found in the field experiment in Chapter 4, table 4.2.7. $SI = (X_1 \times 0.378) + (X_2 \times 0.424) + (X_3+0.828) + (X_4 \times 0.881)$. X_1 = *Leaf area*, X_2 =*Shoot dry weight*, X_3 =*Seed number per plant*, X_4 = *Pod number plant*. Selection index ranks values were obtained using Microsoft Excel.

Duncan Multiple Range Test (DMRT) was used to identify genotypes that were significantly different from each other on selected traits and calculated based on Genstat version 13.0. Based on Duncan Multiple Range Test and selection index the best performing genotypes were identified and ranked.

CHAPTER THREE: Microsatellites, development and characterisation

3.1 Introduction

Bambara groundnut (*Vigna subterranea* (L.) Verdc) is an important indigenous leguminous crop that is cultivated especially by women in most regions of sub-Saharan Africa (Azam-Ali *et al.*, 2001). In some countries, for example in Botswana, bambara groundnut is usually grown both for home consumption and for sale and has considerable importance for subsistence farmers for their local market and commercialisation on a small scale. Despite the fact that bambara groundnut has a potential to contribute to food security in Africa, it has no established varieties. Resource poor farmersgrow crops that are adapted to local environmental conditions (some commercial) landraces, which are genetically diverse populations selected under low-input agriculture (Zeven, 1998).

3.1.1 Breeding systems in bambara groundnut

Knowledge of the mating systems of species or plant populations is important for establishing controlled breeding programs, as breeding methods for self-pollinated crops can be different from the cross-pollinated ones, and those with mixed mating systems (Ferriara *et al.*, 2000). The mechanisms by which plants produce their offspring has a far reaching impact on how the diversity is partitioned and spread within and between populations. Outcrossing species are genetically variable with lower genetic differentiation, while inbreeding plants are less variable, with higher local structure and diversity between populations (Rymer *et al.*, 2002). Breeding systems can significantly affect population ecology and evolution in several ways, since it determines the homozygosity/ heterozygosity of individuals. Inbreeding in plants can be measured by using genetic markers, by estimating the frequencies of homozygotes and heterozygotes (Charlesworth, 2006). Inbreeding also leads to a reduction in effective population size and lowers the genetic recombination that occurs within the population (Lui *et al*., 1999). Inbreeding leads to an increased homozygosis within a population and random changes in gene frequencies from subsequent generations (Robertson, 1961). The mating systems can be investigated using molecular markers such as isozymes, RAPDs, AFLP and microsatellites.

3.1.2 Floral biology of bambara groundnut

The crop has perfect flowers, with stamens and pistil borne in the same flower. Flowers are borne on a raceme on long, hairy peduncles which arise from the nodes (Doku and Karikari, 1970). The flower has a pair of hairy epicalyces.The calyx consists of five hairy sepals, four on the upper side and the lower sepal is free. The standard petal encloses the wing and keel petal until the flower opens. It is usually bright golden yellow and wraps around keel. Generally the wing petals are yellow and enclose the stigma, style and stigma. The stamens are diadelphous which means that there are nine partly fused filaments. In young flower buds, the stigma is slightly above the anthers, while in mature flowers the filaments elongate to place the anthers at a level with the stigma (Massawe *et al.*, 2003).

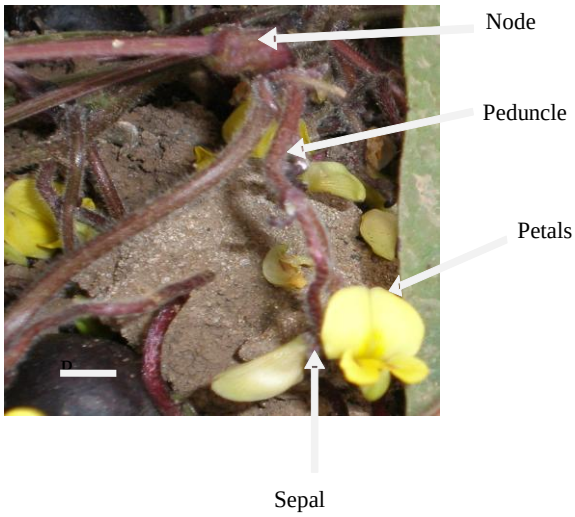


Figure 3.1.2 Bambara groundnut flower, showing the floral morphology. Scale bar = 1 cm

Bambara groundnut is believed to be mainly self-pollinated and anthers dehisce as the stigma becomes receptive even before the flowers open and sometimes fertilisation takes place on the same day as anthesis (Linnemann, 1994). A similar observation was made by Doku and Karikari, (1971), who noticed that pollen maturity and stigma receptivity occurs just before or immediately after the flower opens. The flower structure of bambara groundnut discourages outcrossing since the staminate and pistilate parts are covered by a bract of the cap-like operculum. In addition,the rapid loss of pollen viability reduces the transfer of viable pollen grains (Chijioke *et al.*, 2010). Although there is a lack of detailed studies on the breeding system of bambara groundnut, it appears to be preferably cleistogamous

and would be expected to be inbreeding. However, Doku and Karikari, (1970) reported that ants can also facilitate self- and cross- pollinate bambara groundnut. Mkandawire (2007) reported that self polination in bambara groundnut is mainly found in bunched plants while cross pollination occurs in spreading types.

3.1.3 Seed dissemination systems

Seed dissemination is important when investigating the potential for migration and geneflow between populations. The higher demand for bambara groundnut seed in countries in southern Africa cannot be met within country, thus farmers are sourcing seeds from Zimbabwe, which exports some of its seeds to countries such as Botswana, South Africa and Swaziland (Azam-Ali *et al.*, 2001). A survey carried out in Botswana (Brink *et al.*, 1996) showed that while most of the farmers prefered to use the previous season's harvest as their seed stock, they do also exchange seeds with friends and family members. This movement of seeds across regions is likely to have a impact on the genetic diversity and population structure of bambara groundnut.

3.1.4 Analysis of breeding systems in bambara groundnut

Since bambara groundnut is a self-pollinating crop, intra-landrace variation might be expected to be low. Pasquet *et al.*, (1999) investigated the genetic diversity and population structure of bambara groundunt using 79 domesticated and 21 wild accessions. They employed a total of 41 isozyme markers representing 23 enzyme systems and reported a higher genetic diversity for wild accessions (Ht = 0.087) and a lower genetic diversity for the domesticted type (Ht = 0.052) with 14 and 7 polymophic loci each, respectively. However, their results revealed a relatively higher intrapopulation diversity among the domesticated accessions (Hs=0.033) and lower levels in the wild type (Hs =0.025) and both accessions showed very low levels of heterozygosity which was attributed to the self-pollination nature for both wild and domesticated bambara groundnut.

Massawe, *et al.*, (2002) employed AFLP while Massawe *et al.*, (2003) used RAPDs to determine the heterogeneity within bambara groundnut landraces. They found significant variation among landraces and also among individuals within each landrace. This observation was attributed to the autogamous breeding system of bambara groundnut.

Investigations of intra-landrace genetic diversity were conducted in 10 landraces and 15 individual genotypes of the bambara groundnut landrace by Singrun and Schenkel (2003) who used *Eco*RI/*Mse*I amplified fragment length polymorphism (AFLP) and the heterologous primer pair of AG81 from soybean. Their results demonstated that none of the landraces consisted of a single genotype. Although these studies have shown high levels of variability in bambara groundnut and shed some light on the mating system of the crop, because of the relatively limited polymorphism of isozymes and the dominant marker nature of RAPD and AFLP analysis, there is still more work to be done, particularly in relation to the levels of heterozygosity present within individuals. Initial microsatellite work has confirmed the presence of multiple genotypes within bambara groundnut landraces which has implications for the breeding of the crop. However, determining the level of heterozygosity present within individual genotypes is important for coming up with possible breeding routes available, especially in the production of pure line seed (Basu *et al.,* 2007; Mayes *et al.,* 2009).

3.1.5 Breeding system studies in other leguminous species

Plant mating systems have been generally divided into three main sections, that is predominantly outcrossing, mixed self fertilizing and out crossing and predominantly self fertilizing (Hedrick, 2005). Some leguminous species like *Medicago trunculata* (Kamphius *et al.,* 2007), common bean (*Phaseolus vulgaris*) Tosti and Negri, (2005) and pigeonpea (*Cajanus cajan*) Songok *et al.,* (2010) are predominantly self pollinating, but with a low level of cross-pollination. Although it had been previously believed that wild soybean *Glycine soja* was autogamous, as is cultivated soybean (*Glycine max*), a mean multilocus outcrossing rate estimate of 13% showed that it is also cross-pollinated (Ohara and Shimamoto, 2002). Chickpea (*Cicer arietum*) also commonly known to be a self pollinating crop was shown to have the capability to cross-pollinate with other wild Cicer species such as, *Cicer echinospermum* and *Cicer reticulatum* (Upadhayaya *et al.,* 2008).

Maquet *et al*., (1997) used isozymes markers to confirm the self-pollinating mating sytem of lima bean (*Phaseolus lunatus*). They studied a collection of 235 lima bean accessions originating from Latin America and the Carib zone using 10

allozyme markers. The study revealed a high inbreeding coefficient (f = 0.891), a low intrapopulation gene diversity (Hs = 0.032) as compared to a higher interpopulation gene diversity (D_{ST} =0.235).

When using 48 SSR markers to analyse the genetic diversity of 39 parental lines of mung bean (*Vigna radiata*), Somta *et al*., (2009) observed lower observed heterozygosity of*Ho* = 0.04 compared to a higher expected heterozygosity of *He* =0.39 which was an indication of the inbreeding nature of mung bean.

3.1.6 Applications of microsatellites in this study

The aim of this part of the study was to develop a comprehensive set of microsatellites for bambara groundnut and select the best markers for fingerprinting and other breeding applications. Microsatellites have desirable features that makes them well suited for this application, compared to other markers. Microsatellites or simple sequence repeats (SSRs), are tandem arrays of nucleotide repeats (one to six bases motifs) with SSR loci spread all over the genome. They are a marker of choice due to their higher information content and other features such as high reproducibility and their co-dominant nature (Gupta and Varshney, 2000). They are multi-allelic, highly abundant, analysis is simple and methodologies are easily transferable. Therefore, they are more useful than RAPD or AFLPs and can yield twice as much information per locus as the AFLP and three times as much as RAPDs, according to Gallego *et al*., (2005).

In this study the characterisation of microsatellites has been undertaken using 24 bambara groundnut landrace accessions. The 24 bambara groundnut landraces were selected on the basis of a study conducted by Singruin and Schenkel (2003), where a total 223 bambara groundnuts were analysed for genetic diversity, 46 accessions were originaly from West Africa (Benin, Ghana and Nigeria), 6 from East Africa (Kenya and Tanzania), 7 from Madagascar, 4 from Indonesia, while the rest (160) were from Southern Africa (Botswana, Namibia, Swaziland, Zambia and Zimbabwe). Analysis was undertaken using *Eco*RI/*Mse*I amplified fragment length polymophisms (AFLP) and one heterologous SSR primer pair AG81 derived from soybean. Their results produced 17 clusters and the 24 landraces were used selected from these clusters and are listed on (Table 2.1.2.1).

The same 24 landraces were analysed with Diversity Arrays Technology (DArT) markers for comparison of the efficiency of the two techniques.

Diversity Arrays Technology was developed as a hybridisation-based technology and is valued for the high level of data production due to its microarray platform. It can type thousands of loci in a single assay, and generates whole genome fingerprints of genomic representations, generated from sub-samples of genomic DNA (Jaccourd *et al.,* 2001). DArT has the advantage of low cost, high throughput and it does not require sequencing, this makes it more suited for use in 'orphan' crops such as bambara groundnut, as compared to SSR markers which requires prior sequence information (Yang *et al*., 2006). A number of marker types have been employed in crop breeding studies with different efficacy and ease of use to quickly develop or assay large number of markers (Akbari *et al*., 2006). The importance of comparing different marker systems is to assist in making informed decisions as to which marker is best to use in germplasm characterisation and plant breeding. The aim of this part of the study is to compare the use of DArT and SSR in assessing the genetic diversity of 24 bambara groundnut landraces and genetic diversity analysis of bambara groundnut.

3.2 Materials and Methods

3.2.1 DArT marker screening

DArT marker screening and genotyping were undertaken by Diversity Array Pty, Ltd, Yarralumla, Australia as described by Jaccoud *et al.,* 2001. This basically consists of three major steps; array development, genotyping, and data analysis. The number of markers that can be obtained does not only depend on the levels of genetic diversity in the germplasm but also on the combination of restriction enzymes used to generate the representation used to produce the clones.Therefore a suitable complexity reduction method has to be identifed. Two restriction endonucleases combinations were tested by DArT Pty Ltd. for the treatment of combined DNA samples, with *Pst*I used as the rare cutter (restriction 6 bp), while enzymes *Alu*I, *Ban*II, *Bso*BI, *Bst*NI, *Mse*I, *Ras*I, *Taq*I and *Tsp*5091) sourced from (New England Biolabs Ltd., Pickering Canada) with a 4 bp were tested as frequent cutter. Gel electrophoresis suggested that *Alu*I was a suitable 4 bp cutter

as it produced a homogenous smear without repetitive bands after visualisation with Ethidium bromide stained agarose gel, thus it was selected to develop intial Discovery Array. Further details on the development of the DArT array for bambara groundnut is described in Stadler, (2009).

A full genotyping array containing 7,680 clones was generated from two complexity reduction methods, using *Pst*I/*Alu*I and *Pst*I/*Taq*I. The restriction endonuclease *Pst*I/*Alu*I produced 157 polymorphic clones, while in the second complexity reduction method *Pst*I/*Taq*I produced 168 polymorphic clones. When data sets were combined, and after removing all repeated discrimation patterns a final remainder of 296 polymorphic clones were used for DArT genetic diversity analysis, based on the initial 94 genotypes. However, when these polymorphic markers were used in the large scale analysis of 342 bambara groundnut a total of 201 robust markers polymorphic across all samples analysed remained and these were used in the analysis of the full bambara groundut germplasm. It is this dataset that the 24 bambara groundut accessions were drawn from, the same accessions having also been analysed using 68 SSR markers by the author.

3.3 Results

3.3.1 Microsatellites marker analysis

From the initial set of 75 markers listed in appendix 2, seven markers had poor amplification; producing smeared/complex bands or no PCR product at all. As a result they were discarded. Therefore 68 microsatellites were used to characterise and evaluate the genetic diversity of 24 bambara groundnut landraces (Table 3.1).

Table 3.1: Summary of PowerMarkers data analysis of 24 bambara groundnut landraces, based on 68 microsatellites.

Marker	MAF	GN	SS	No.	AN	Avail.	GD	Het.	PIC	*f*
Primer 1	0.71	3	24	24	3	1	0.45	0.00	0.40	1.00
Primer 2	0.69	4	24	24	3	1	0.45	0.04	0.38	0.91
Primer 3	0.88	3	24	24	3	1	0.23	0.00	0.21	1.00
Primer 4	0.54	4	24	24	4	1	0.62	0.00	0.57	1.00
Primer 5	0.50	4	24	24	4	1	0.57	0.00	0.48	1.00
Primer 6	0.50	4	24	24	4	1	0.64	0.00	0.58	1.00
Primer 7	0.29	6	24	24	6	1	0.74	0.00	0.69	1.00
Primer 8	0.71	2	24	24	2	1	0.41	0.00	0.33	1.00
Primer 9	0.92	2	24	24	2	1	0.15	0.00	0.14	1.00
Primer 10	0.46	4	24	24	4	1	0.61	0.00	0.53	1.00
Primer 11	0.92	2	24	24	2	1	0.15	0.00	0.14	1.00
Primer 12	0.54	3	24	24	3	1	0.56	0.00	0.47	1.00
Primer 13	0.92	2	24	24	2	1	0.15	0.00	0.14	1.00
Primer 14	1.00	1	24	24	1	1	0.00	0.00	0.00	NaN
Primer 15	0.17	15	24	24	14	1	0.90	0.13	0.89	0.87
Primer 16	0.21	6	24	24	6	1	0.82	0.00	0.79	1.00
Primer 17	0.67	5	24	24	5	1	0.52	0.00	0.48	1.00
Primer 18	0.46	5	24	24	5	1	0.71	0.00	0.67	1.00
Primer 19	0.29	9	24	24	9	1	0.82	0.00	0.80	1.00
Primer 20	1.00	1	24	24	1	1	0.00	0.00	0.00	NaN
Primer 21	0.83	3	24	24	3	1	0.29	0.00	0.26	1.00
Primer 22	1.00	1	24	24	1	1	0.00	0.00	0.00	NaN
Primer 23	0.81	4	24	24	3	1	0.32	0.04	0.30	0.88
Primer 24	0.88	4	24	24	4	1	0.23	0.00	0.22	1.00
Primer 25	1.00	1	24	24	1	1	0.00	0.00	0.00	NaN
Primer 26	0.40	12	24	24	12	1	0.80	0.04	0.79	0.95
Primer 27	0.75	5	24	24	5	1	0.42	0.04	0.39	0.90
Primer 28	0.50	5	24	24	5	1	0.63	0.00	0.57	1.00
Primer 29	1.00	1	24	24	1	1	0.00	0.00	0.00	NaN
Primer 30	0.29	6	24	24	6	1	0.80	0.00	0.77	1.00
Primer 31	0.88	3	24	24	3	1	0.23	0.00	0.21	1.00
Primer 32	0.25	11	24	24	10	1	0.86	0.04	0.84	0.95

Table 3.1 (Continued)

Marker	MAF	GN	SS	No.	AN	Avail.	GD	Het.	PIC	*f*
Primer 33	0.29	9	24	24	9	1	0.82	0.00	0.79	1.00
Primer 34	0.96	2	24	24	2	1	0.08	0.00	0.08	1.00
Primer 35	1.00	1	24	24	1	1	0.00	0.00	0.00	NaN
Primer 36	1.00	1	24	24	1	1	0.00	0.00	0.00	NaN
Primer 37	0.33	7	24	24	7	1	0.75	0.00	0.71	1.00
Primer 38	0.58	5	24	24	5	1	0.61	0.00	0.58	1.00
Primer 40	0.63	2	24	24	2	1	0.47	0.00	0.36	1.00
Primer 41	0.83	3	24	24	3	1	0.29	0.00	0.26	1.00
Primer 42	0.63	4	24	24	4	1	0.54	0.04	0.49	0.93
Primer 43	0.42	5	24	24	5	1	0.70	0.00	0.64	1.00
Primer 44	0.42	5	24	24	5	1	0.74	0.00	0.70	1.00
Primer 45	0.73	3	24	24	2	1	0.39	0.04	0.32	0.90
Primer 48	0.29	13	24	24	13	1	0.83	0.00	0.82	1.00
mBam2co80	0.17	12	24	24	12	1	0.90	0.00	0.89	1.00
D1	0.21	11	24	24	11	1	0.89	0.00	0.87	1.00
D2	0.67	5	24	24	5	1	0.52	0.00	0.48	1.00
D3	0.81	3	24	24	3	1	0.31	0.04	0.27	0.87
D4	0.79	4	24	24	4	1	0.36	0.00	0.34	1.00
D5	0.42	7	24	24	7	1	0.74	0.00	0.71	1.00
D6	0.83	3	24	24	3	1	0.29	0.00	0.26	1.00
D7	0.96	2	24	24	2	1	0.08	0.00	0.08	1.00
D8	1.00	1	24	24	1	1	0.00	0.00	0.00	NaN
D9	0.46	7	24	24	6	1	0.71	0.04	0.68	0.94
D10	0.71	6	24	24	6	1	0.48	0.00	0.45	1.00
D11	0.42	8	24	24	7	1	0.76	0.04	0.73	0.95
D12	0.42	9	24	24	9	1	0.76	0.00	0.74	1.00
D13	0.96	2	24	24	2	1	0.08	0.00	0.08	1.00
D14	0.23	12	24	24	11	1	0.86	0.08	0.85	0.91
D15	0.27	11	24	24	10	1	0.83	0.04	0.81	0.95
E1	0.79	3	24	24	3	1	0.34	0.00	0.31	1.00
E3	0.96	2	24	24	2	1	0.08	0.00	0.08	1.00
E5	0.94	3	24	24	3	1	0.12	0.04	0.12	0.66
E7	0.52	5	24	24	4	1	0.61	0.08	0.54	0.87
E9	0.96	2	24	24	2	1	0.08	0.00	0.08	1.00
E10	1.00	1	24	24	1	1	0.00	0.00	0.00	NaN
E11	0.88	3	24	24	3	1	0.23	0.17	0.21	0.28
Mean	**0.65**	**5**	**24**	**24**	**5**	**1**	**0.45**	**0.01**	**0.42**	**0.97**

MAF-Major Allele Frequency
GN- Genotype number
SS- Sample Size
No.- Number of observations for a marker
AN- Allele number
Avail. Availability
GD- Gene diversity or expected heterozygosity, the probability that two randomly chosen alleles from population are different
Het. The proportion of heterozygous individuals in the population
PIC-Polymorphic Information Content
f-inbreeding coefficient

A total of 313 alleles were identified, with nine markers non-polymorphic (marker 14, marker 20, marker 22, marker 25, marker 29, marker 35, marker 36, marker D8 and marker E10). The number of alleles per marker ranged from 1 for non-polymorphic markers to 14 in marker 15, with a mean of 5 alleles per marker (Table 3.1). The polymorphic information content (PIC) values ranged from 0.08 to 0.89 from primer D7, primer D13, primer E9, primer E3, primer 34 to marker 15 and marker mBam2co80with an average of 0.42, with nine markers been monomorphic. 39.7 % of the markers were highly polymorphic with PIC values ranging from 0.5 to 0.89, while 33.8 % of the markers were just informative with PIC values ranging from 0.21 to 0.49. The remaining 26.5% include nine markers that were monomorphic, and eight markers with low polymorphic information content and a range of 0.08 to 0.15.A one sample t-test conducted on the data, for the average He(0.45) and PIC (0.42) revealed no significant difference to those obtained by Basu *et al.,* (2007) for He(0.50) and PIC (0.47).

Bambara groundnut is an inbreeding crop, so as expected all the markers showed a lower observed heterozygosity (*Ho*) compared to the expected heterozygosity (*He*). However, 23.5% of the markers showed some heterozygosity with a range of 0.04 to 0.17. Markers E5 and E11showed a low inbreeding coefficient at 0.66 and 0.28 respectively, while the rest of the markers had an inbreeding coefficient of 1. As both markers E5 and E11 appear to be outliers, it seems possible that these primers detected more than a single locus. Interestingly, both are derived from Roche 454 sequence (average read length 92bp) derived from RNA.

3.3.1.1 Hardy Weinberg Equilibrium (HWE)

HWE was tested using PowerMarker, which uses three different methods to test for Hardy-Weinberg equilibrium, the Chi square statistics and the permutation version of the exact test given (Table 3.2). All markers are highly significant at ($P<0.05$), an indication that the population is not in HW, as would be expected for structured accessions of a germplasm collection derived from an inbreeding species.

Table 3.2: The 68 markers used in the 24 bambara groundnut analysis were subjected to Chi square and HWE exact test using MVSP version 3.25, with the exception of nine non-polymorphic markers.

Marker	X^2 value	X^2 d.f.	Exact p-value
Primer 1	48.00	3	0.00
Primer 2	43.13	3	0.00
Primer 3	48.00	3	0.00
Primer 4	72.00	6	0.00
Primer 5	72.00	6	0.00
Primer 6	72.00	6	0.00
Primer 7	120.00	15	0.00
Primer 8	24.00	1	0.00
Primer 9	24.00	1	0.00
Primer 10	72.00	6	0.00
Primer 11	24.00	1	0.00
Primer 12	48.00	3	0.00
Primer 13	24.00	1	0.00
Primer 15	254.43	91	0.00
Primer 16	120.00	15	0.00
Primer 17	96.00	10	0.00
Primer 18	96.00	10	0.00
Primer 19	192.00	36	0.00
Primer 21	48.00	3	0.00
Primer 23	38.39	3	0.00
Primer 24	72.00	6	0.00
Primer 26	240.07	66	0.00
Primer 27	72.96	10	0.00
Primer 28	96.00	10	0.00
Primer 30	120.00	15	0.00
Primer 30	48.00	3	0.00
Primer 32	202.36	45	0.00
Primer 33	192.00	36	0.00
Primer 34	24.00	1	0.02
Primer 37	144.00	21	0.00
Primer 38	96.00	10	0.00
Primer 40	24.00	1	0.00
Primer 41	48.00	3	0.00
Primer 42	48.20	6	0.00
Primer 43	96.00	10	0.00
Primer 44	96.00	10	0.00
Primer 45	19.20	1	0.00

Table 3.2 (Continued)

Marker	X^2 value	X^2 d.f.	Exact p-value
Primer 48	288.00	78	0.00
mBam2co80	264.00	66	0.00
D1	240.00	55	0.00
D2	96.00	10	0.00
D3	24.02	3	0.00
D4	72.00	6	0.00
D5	144.00	21	0.00
D6	48.00	3	0.00
D7	24.00	1	0.02
D9	102.59	15	0.00
D10	120.00	15	0.00
D11	125.23	21	0.00
D12	192.00	36	0.00
D13	24.00	1	0.02
D14	208.76	55	0.00
D15	204.55	45	0.00
E1	48.00	3	0.00
E3	24.00	1	0.01
E5	24.01	3	0.02
E7	40.13	6	0.00
E9	24.00	1	0.02
E11	24.22	3	0.03

3.3.1.2 Estimation of Null alleles.

Deviation from HWE in a population can be caused by selection, non-random mating, inbreeding/self-fertilisation, migration and the presence of null alleles. The presence of null alleles was investigated in the data set (Table 3.3). Estimation of null alleles was conducted using the INEst (Inbreeding /Null allele Estimation) software, which takes into account the possibility of inbreeding within a population during estimation of null frequencies (Chybicki and Burczyk, 2009). The population inbreeding model (PIM) and the individual inbreeding model (IIM) are calculated. The PIM estimate uses the jacknife algorithm, while the IIM estimates uses the Gibbs sample command which uses a number of run-in steps of approximately 10,000. The estimates of PIM and IIM are given in table 3.3.

Table 3.3: Estimation of null allele frequencies for each locus, using the population inbreeding model (PIM) and the individual inbreeding model (IIM) using INEst (Chybicki and Burczyk, 2009).

Marker	PIM	IMM	Ho	He	F(Wright index)
Primer 1	0.000	0.001	0.000	0.050	1.000
Primer 2	0.000	0.001	0.000	0.621	1.000
Primer 3	0.000	0.001	0.000	0.082	1.000
Primer 4	0.000	0.001	0.000	0.642	1.000
Primer 5	0.000	0.001	0.000	0.369	1.000
Primer 6	0.000	0.001	0.083	0.909	0.908
Primer 7	0.000	0.001	0.000	0.000	1.000
Primer 8	0.000	0.001	0.042	0.839	0.950
Primer 9	0.000	0.001	0.000	0.553	1.000
Primer 10	0.000	0.001	0.083	0.852	0.902
Primer 11	0.000	0.001	0.000	0.684	1.000
Primer 12	0.000	0.001	0.000	0.858	1.000
Primer 13	0.000	0.001	0.042	0.329	0.873
Primer 14	0.000	0.001	0.000	0.723	1.000
Primer 15	0.000	0.001	0.000	0.684	1.000
Primer 16	0.000	0.001	0.000	0.656	1.000
Primer 17	0.000	0.001	0.042	0.776	0.946
Primer 18	0.000	0.001	0.083	0.883	0.906
Primer 19	0.000	0.001	0.042	0.488	0.915
Primer 20	0.000	0.001	0.000	0.911	1.000
Primer 21	0.000	0.001	0.042	0.042	0.901
Primer 22	0.000	0.001	0.000	0.082	1.000
Primer 23	0.000	0.001	0.000	0.284	1.000
Primer 24	0.000	0.001	0.000	0.223	1.000
Primer 25	0.000	0.001	0.000	0.383	1.000
Primer 26	0.000	0.001	0.000	0.000	1.000
Primer 27	0.000	0.001	0.000	0.000	1.000
Primer 28	0.000	0.001	0.000	0.000	1.000
Primer 29	0.000	0.001	0.000	0.156	1.000
Primer 30	0.000	0.001	0.000	0.156	1.000
Primer 31	0.000	0.001	0.000	0.000	1.000
Primer 32	0.000	0.001	0.000	0.528	1.000
Primer 33	0.000	0.001	0.000	0.755	1.000
Primer 34	0.000	0.001	0.000	0.000	1.000
Primer 35	0.000	0.001	0.000	0.000	1.000
Primer 36	0.000	0.001	0.000	0.230	1.000
Primer 37	0.000	0.001	0.000	0.000	1.000
Primer 38	0.000	0.001	0.042	0.692	1.000
Primer 40	0.000	0.001	0.000	0.000	1.000

PIM: Population Inbreeding Model: (IIM) Individual Inbreeding Model

Table 3.3 (Continued)

Marker	PIM	IMM	Ho	He	F(Wright index)
Primer 41	0.000	0.001	0.000	0.574	1.000
Primer 42	0.000	0.001	0.000	0.000	1.000
Primer 43	0.000	0.001	0.000	0.000	1.000
Primer 44	0.000	0.001	0.000	0.000	1.000
Primer 45	0.000	0.001	0.000	0.000	1.000
Primer 48	0.000	0.001	0.000	0.610	1.000
mBam2co80	0.000	0.001	0.000	0.000	1.000
D1	0.000	0.001	0.000	0.294	1.000
D2	0.000	0.001	0.083	0.536	0.845
D3	0.000	0.001	0.000	0.730	1.000
D4	0.000	0.001	0.042	0.403	0.897
D5	0.000	0.001	0.000	0.691	1.000
D6	0.000	0.001	0.000	0.883	1.000
D7	0.000	0.001	0.042	0.042	0.000
D8	0.000	0.001	0.042	0.042	0.000
D9	0.000	0.001	0.000	0.230	1.000
D10	0.000	0.001	0.000	0.741	1.000
D11	0.000	0.001	0.000	0.082	1.000
D12	0.000	0.001	0.000	0.000	1.000
D13	0.000	0.001	0.042	0.728	0.943
D14	0.000	0.001	0.000	0.305	1.000
D15	0.000	0.001	0.000	0.723	1.000
E1	0.000	0.001	0.000	0.000	1.000
E3	0.000	0.001	0.000	0.284	1.000
E5	0.000	0.001	0.000	0.000	1.000
E7	0.000	0.001	0.000	0.000	1.000
E9	0.000	0.001	0.000	0.000	1.000
E10	0.000	0.001	0.000	0.000	1.000
E11	0.000	0.001	0.167	0.156	-0.068

PIM: Population Inbreeding Model: (IIM) Individual Inbreeding Model

Both the population inbreeding model (PIM) and individual inbreeding model (IIM) showed that none of the markers have null alleles which usually leads to erroneous interpretation of the data (Wagner *et al.*, 2006). However, marker E5 with unusual low inbreeding coefficient (0.66) was excluded from the data analysis. Other two markers which were excluded are markers E11 and 15 with high heterozygous values of 0.17 and 0.13 which maybe more indicative of the SSR amplifying from two loci, rather than true heterozygosity at the locus.

Therefore only 65 SSR markers were used in the finale analysis of the 24 bambara groundnut landraces.

3.3.2 Principal Component Analysis (PCO)

Eigenvalues and the cumulative percentage of the principal component case scores were used for the analysis of the population structure for the selected 24 bambara groundnut for the comparison of DArT and SSR , the results are given in (Table 3.4 and Table 3.5). Data for the two markers looking at the first two axes suggests that DArT marker is revealing more variation at (37.3%) compared to SSR markers at (19.5 %) in the first two axes.

Table 3.4: PCO case scores for the population structure of the selected 24 bambra groundnut landraces, determined based on 201 DArT markers

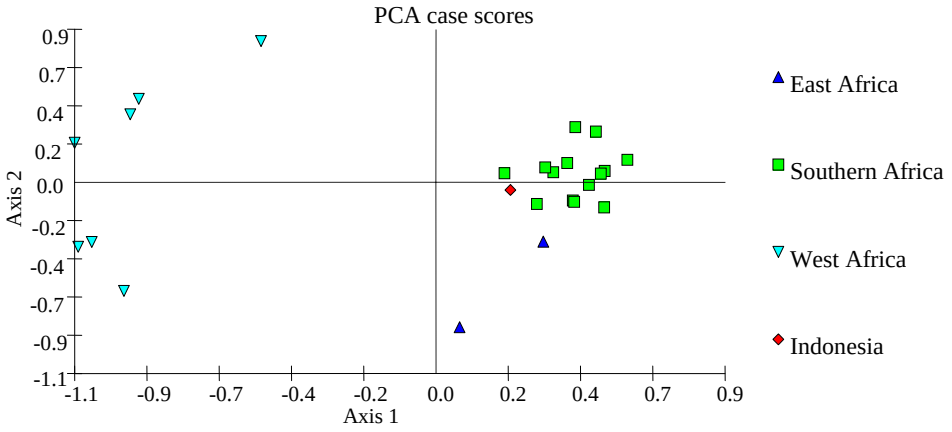
	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5	Axis 6	Axis 7	Axis 8	Axis 9	Axis 10
Eigenvalues	9.58	2.98	2.15	2.02	1.79	1.71	1.53	1.39	1.35	1.22
Percentage	28.45	8.84	6.39	6.00	5.32	5.07	4.54	4.13	4.01	3.61
% Accumulation	28.45	37.30	43.69	49.68	55.00	60.07	64.61	68.74	72.75	76.36

Table 3.5: PCO case scores for the population structure of the selected 24 bambra groundnut landraces, determined based on 65 SSR markers.

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5	Axis 6	Axis 7	Axis 8	Axis 9	Axis 10
Eigenvalues	2.58	2.04	1.87	1.80	1.57	1.47	1.27	1.21	1.13	1.05
Percentage	10.91	8.61	7.90	7.60	6.63	6.19	5.37	5.13	4.77	4.44
% Accumulation	10.91	19.51	27.42	35.02	41.64	47.83	53.20	58.33	63.10	67.53

The PCO analysis with each marker type indicates that both DArT and SSR analysis show differentiation between the selected 24 bambara groundnut landraces. However, it was DArT which clearly gave a separation of landraces according to their areas of origin, West African landraces were separated from the East African landraces and the Southern Africa landraces, but Ramayana from Indonesia formed a group with the Southern African landraces (Figure 3.2 a). The SSR marker analysis did not display such a clear separation of landraces based on place of origin (Figure 3.2 b). The landraces from Southern Africa are scattered across both axes while East African landraces grouped together. Ramayana from Indonesia-Asia shows a close proximity to the Southern African landraces.

a) DArT marker



b) SSR marker

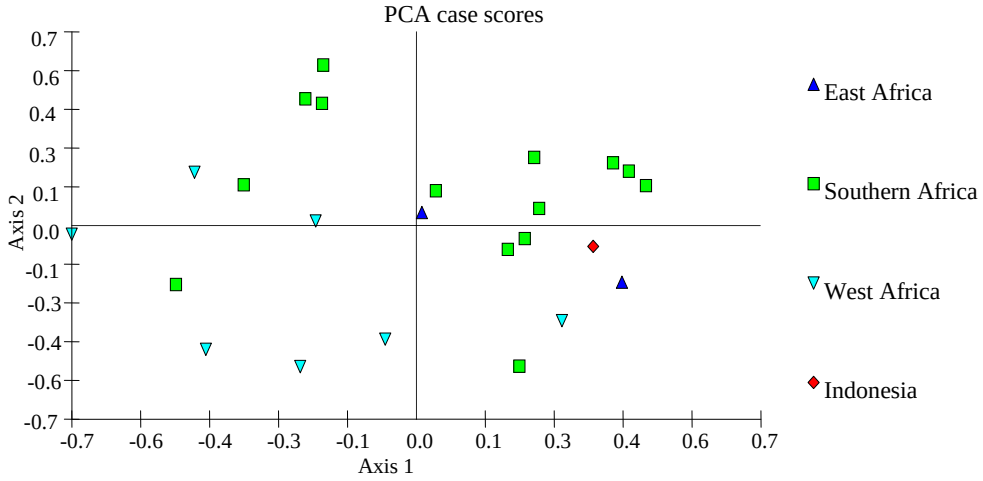


Figure 3.2: The first two axes of the PCO case scores, generated from the 24 landraces using MVSP for figure 3.2 (a) DArT Axis 1 represents 28.45% and Axis 2 represents 8.84 % of the molecular variation, figure 3.2 (b) SSR markers; Axis 1 represents 10.91 % and Axis 2 represents 8.61% of the molecular variation in the 24 selected bambara groundnut landraces.

3.3.3 Cluster analysis

The UPGMA dendrogram (Figure 3.3a and Figure 3.3b) show a moderate genetic variability for both markers. DArT markers showed three clear and distinct clusters (Figure 3.3a) and grouped landraces based on their areas of origin. Cluster 1 consists of landraces mainly from West Africa, while cluster 2 consists of a mixture of three Southern Africa landraces, this group also includes Ramayana originally from Indonesia-Asia, and this could be a reflection of the origin of this landrace from Africa, this cluster also consists of wild landrace VSSP6 from Cameroon and DodR from East Africa. Cluster three is mainly landraces from Southern Africa with the exception of DodC from Tanzania located on the edge of the cluster.

The SSR marker dendrogram figure 3.3 (b) grouped the landraces into three clusters with, two major groups and one which consist of Tiganicuru from Mali and DodR from Tanzania only. The landraces are largely grouped based on their areas of origin like the DArT markers since the majority of landraces within a cluster originate from the same region. Cluster one, consists of a mixture of four landraces from West Africa and four from Southern Africa. Cluster two consist mostlylandraces from Southern Africa with the exception of Nav Red and Tvsu 569 from West Africa, DodC from East Africa and Ramayana from Indonesia. The comparison for the two markers is tabulated in Table 3.6.

DArT markers had more bootstrap values more than 50% as compared to the SSR marker DArT which indicates the reliability of the marker. Landraces S19-3 and S19/3, both from Namibia, were identified by DArT at 97% similarity while SSR marker showed minor differences at 99% similarity. It is interesting to note that DArT markers found GabC from Botswana to be identical to AHM968 from Namibia. This could probably be caused by the seeds exchange between these two neighbouring countries or potentially a mistake has been made during seed handling.

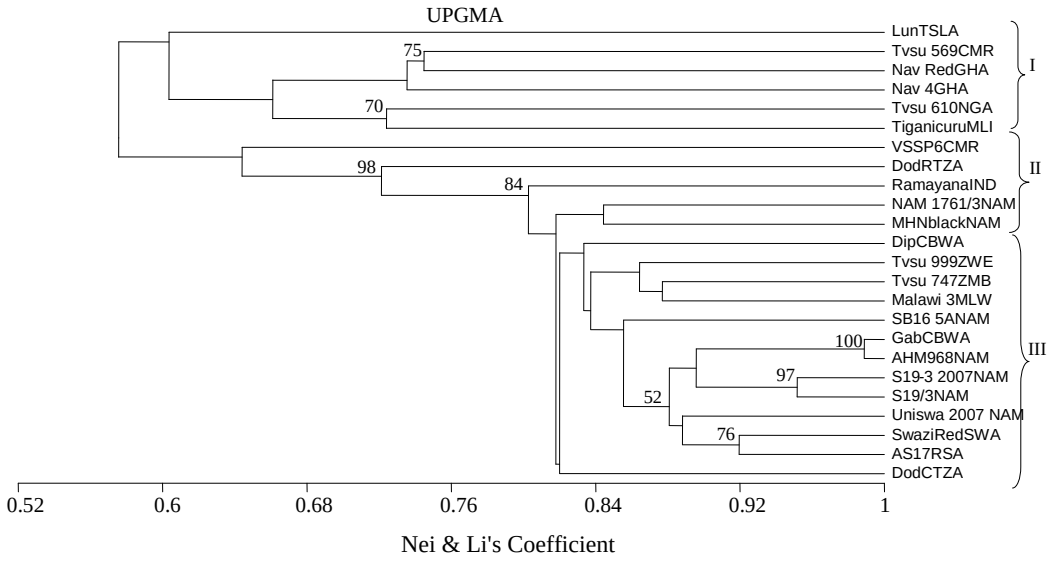


Figure 3.3 (a) Cluster analysis of the 24 bambara groundnut landraces. The UPGMA dendrogram is based on the similarity matrix obtained from 201 DArT markers using the Nei and Li, (1979). The number at the nodes of branches represents the percentage bootstrap support of individual nodes at resampling at 1000.

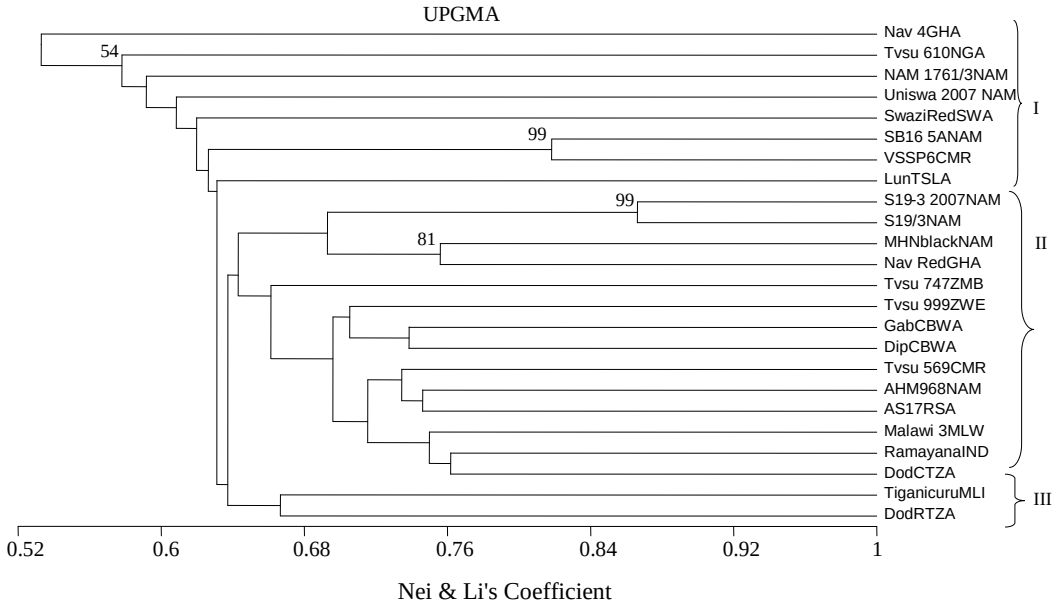


Figure 3.3 (b) Cluster analysis based on the 24 bambara groundnut landraces, the dendrogram was obtained based on 65 SSR markers, the UPGMA tree is based on the Nei and Li, 1979 similarity coefficient. The number at the nodes of branches represents the percentage bootstrap support of individual nodes at resampling at 1000.

Table 3.6: A comparison of the distribution of the 24 bambara groundnut landraces based on the UPGMA clustering analysis done using a set of 201 DArT markers and 65 SSR markers.

a) DArT Marker

Cluster I	Cluster II	Cluster III
LunT SLA	VSSP6 CMR	DodC TZA
Tvsu 569 CMR	DodR TZA	S19/3 NAM
Nav 4 GHA	Ramayana IND	S19-3 2007 NAM
Nav Red GHA	NAM 1761/3 NAM	Uniswa 2007 NAM
Tvsu 610 NGA	Mahenene black NAM	SB16 5A NAM
Tiganicuru MLI		AHM968 NAM
		Malawi 3 MLW
		Tvsu 747 ZMB
		Gabc BWA
		Tvsu 999 ZWE
		AS17 RSA
		DipC BWA
		SwaziRedSWA

b) SSR marker

Cluster I	Cluster II	Cluster III
Nav 4 GHA	S19/3 NAM	Tiganicuru MLI
Tvsu 610 NGA	S19-3 2007 NAM	DodR TZA
Uniswa 2007 NAM	Mahenene black NAM	
NAM 1761/3 NAM	Nav Red GHA	
SwaziRedSWA	Tvsu 747 ZMB	
SB16 5A NAM	Tvsu 999 ZWE	
VSSP6 CMR	Gabc BWA	
LunT SLA	DipC BWA	
	Tvsu 569 CMR	
	AHM968 NAM	
	AS17 RSA	
	Malawi 3 MLW	
	Ramayana IND	
	DodC TZA	

DArT markers revealed higher similarities between the landraces with an average of 0.71 and a wider range of 0.48 to 0.99 between the landraces compared to SSR markers which had a genetic similarity estimate mean of 0.65 and a range of 0.43 to 0.87 (Appendix 4). The lowest difference between the landraces according to DArT markers is between AHM968 from Namibia and GabC from Botswana

(0.99), and the largest difference is between landraces DodR from Tanzania and Tvsu 610 from Nigeria (0.48). For SSR markers the least difference was between S19-3 and S19/3 at 0.87, and the largest genetic distance is between Nav 4 and Tvsu 999 at 0.43. The differences in genetic distance/similarity estimates by markers has been attributed to the extent of distribution of genome coverage by markers and their evolutionary different properties and the individual loci used for analysis (Geleta *et al*., 2005). SSR markers are also likely to show more intra-landrace variability, as they are likely to evolve at higher rates than DArT markers.

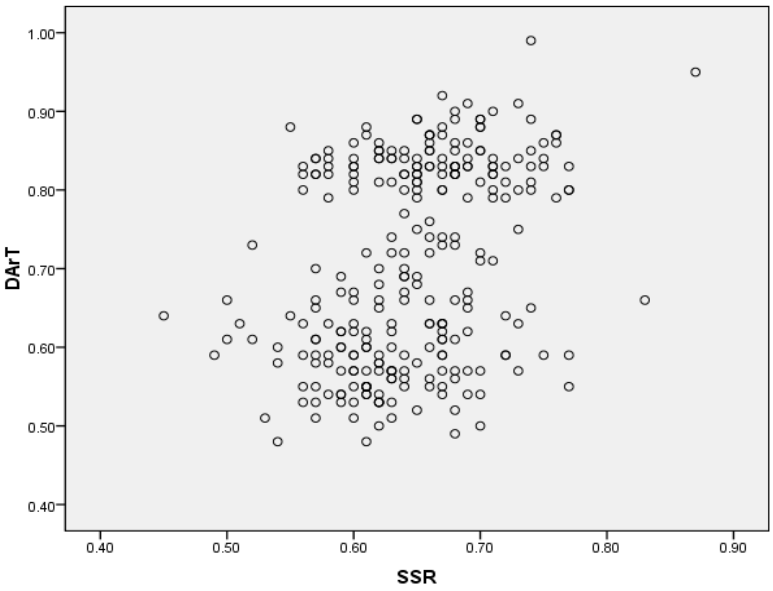
The cophenetic estimates measures how the dendrogram produced clearly reveals pairwise distances in the original data. Cophenetic values for the two marker types displayed significant values with r = 0.97 for DArT and r = 0.83 for SSR which is an indication of a very good fit and a good fit, respectively (Figure 3.3a and Figure 3.3b). The correlation between genetic similarity estimates for the two markers were highly significant, using the correlations estimates produced using Pearson product-moment coefficient correlation, Spearman rank's coefficient correlation and Mantel tests correlation (Table 3.7) and (Figure 3.4 a and Figure 3.4 b).

Table 3.7: Pearson, Spearman and Mantel test correlations between the genetic similarity matrices based on the two markers systems (DArT vs SSR).

Marker	Pearson product-moment coefficient Correlation	Spearman(rank) coefficient correlation	Mantel test
	DArT	DArT	DArT
DArT	1	1	1
SSR	0.346	0.336	0.354
N	276	276	276
P value	0.01	0.01	0.005

N = Number of observations in matrix

a)



b)

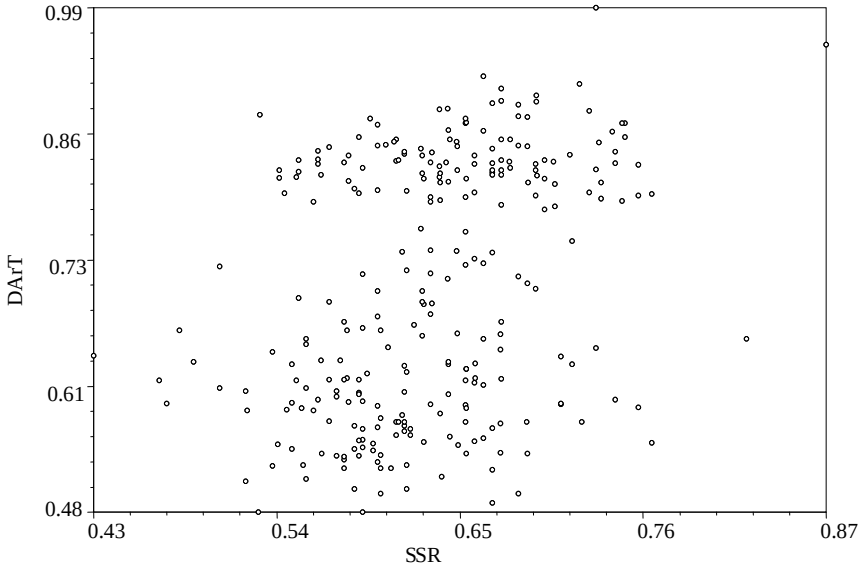


Figure 3.4 (a) A scatter plot produced based on the matrix for DArT and SSR markers genetic distance estimates from Nei and Li, 1979 (Appendix 4) using Pearsonproduct-moment coefficient correlation based on SPSS version 16 (b) A scatter plot based on the matrix for DArT and SSR produced from the same genetic distance estimates in Appendix 4using Mantel-matrix correspondence test on NTSYS pc version 2.1 program (MXCOMP module) based on 1000 permutation.

3.4 Discussions:

Diversity in bambara groundnut.

Tests for deviations from Hardy Weinberg proportions are usually used to check for random mating in populations which will in turn be used to estimate the inbreeding coefficient (Robertson and Hill, 1984). The results showed that all markers did not conform to HWE. There was a deficit of heterozygosity observed among the 24 bambara groundnut landraces as confirmed by the (*f)* inbreeding coefficient average of0.97, which could be the most likely reason to account for the deviations from Hardy Weinberg equilibrium. The number of alleles per marker ranged from 1 to 14, with an average of 5 per marker and the observed heterozygosity (Ho) was lower than the expected heterozygosity, which would be consistent with the clear deviation observed from HWE. The heterozygosity of the SSR markers is very low, reflecting the genetic composition and mating behaviour of the tested landraces, as inbreeding, together with the lack of null alleles in these markers suggesting that this marker type has usefully revealed very low levels of out-crossing in bambara groundnut.

Even though there was a large inbreeding estimate in this study at an average of 0.97, the accessions showed far higher expected heterozygosity (*He* = 0.45) than observed. The selected landraces are originally from 13 countries, with the majority of landraces from Southern Africa (14), seven from West Africa, while two are from (East Africa) Tanzania and one is from (Asia) Indonesia. A similar observation was made in other highly self-fertilizing species. Siol *et al*., (2008) set up a study on *Medicagotruncatula*, to find out the reason behind the high genetic diversity among the selfing species. Seven microsatellite loci were used and showed between two and five alleles per loci per locus, and an average observed heterozygosity of 0.011 against an expected heterozygosity of 0.457.

Buso *et al*., (2006) carried out a similar exercise for common bean (*Phaseolus vulgaris*) which is thought to have a similar breeding system to bambara groundnut. They found that from 20 SSR markers evaluated using 85 accessions, the number of alleles per locus ranged from 3 to 10, with a mean of 7. They also recorded a lower observed heterozygosity (Ho) of 0.026 compared to the expected heterozygosity (He) of 0.622, suggesting that it is also an inbreeding crop.

Principal Component Analysis (PCoA); in this study DArT markers were able to clearly differentiate the 24 bambara groundnut landraces in a way that corresponded to their areas of origin. Similar findings by Massawe *et al*., (2002) using AFLP and Amadou *et al.,* (2001) using RAPDs on studying bambara groundnut landraces. However, Yang *et al.,* (2006) when using DArT markers on the analysis of pigeonpea, found that they could not be differentiated according to their place of origin, but the markers were related to their morphological characters. The DArT markers showed more molecular variation among the West African bambara groundnut landraces compared to the Southern African landraces, which may be a reflection of the domestication pattern of bambara groundnut. The West African landraces as the putative area of origin has been identified as more diverse using morphological markers (Pasquet *et al*., 1999). In contrast, principle component analysis for the SSR markers did not clearly show the differentiation of the 24 landraces based on their areas of origin, it showed that there is greater genetic differentiation among the southern African landraces than the West African materials. This shows that the two markers reveal different levels of discrimination (Jaccoud *et al*., 2001), possibly due to their differing mutation rates. However, the DArT markers explained the greater proportion of the molecular variability in the first two axes among the 24 bambara groundnut landraces at 37.3% as compared to SSR with 19.5 %.

Both DArT and SSR markers cluster analysis fits well with the dendrogram produced. This was revealed by high cophenetic coefficients for each marker type 0.97 for DArT and 0.83 for SSR. Other researchers have recorded similar magnitudes of cophenetic correlation, Giancola *et al.,* (2002) recorded cophenetic coefficient r = 0.701 for SSR marker among 100 soybean cultivars when using 33 SSR markers, in 12 soybean accessions, Powell *et al.* (1996) recorded r = 0.958 for 36 SSR markers, Raman *et al.* (2008) recorded a similar cophenetic coefficient of 0.97 for DArT markers, in a set of 94 genotypes of *Lupinus albus* L.

A highly significant Pearson product-moment correlation coefficient (r =0.35), Spearman rank correlation coefficient (r =0.34) and Mantel matrix correspondence test (r = 0.35) which was low showed that both techniques, even though there are targeting different parts of the genome, their results could still be inferred to some extent from one to the other (Table 3.7). The comparison made

on wheat using Mantel test correspondence showed a relatively higher correlation between DArT and SSR markers. Mantovani *et al.,* (2008) found a correlation between the genetic distance matrices of DArT and SSR of r = 0.68 among a set of 31 accessions using 1,315 DArT markers and 103 SSR markers, which indicated an agreement between the two markers. Stodart *et al.,* (2007) observed a strong positive correlation (r =0.84) between DArT and SSR markers when using 256 DArT markers and 63 SSR markers on 44 accessions of bread wheat (*Triticum aestivum* L.) using Mantel test correspondence.

The DArT marker as in the PCO was able to clearly separate landraces based on their areas of origin. The use of these relationships revealed by the PCO analysis and UPGMA dendrogram could assist in formulating a breeding program for bambara groundnut, for example, by selecting genetically far apart landraces for cross breeding in this case which could combine the best attributes of landraces from Southern Africa with those from West Africa, taking into account agro-ecological zones.

Both molecular techniques DArT and SSR, showed a relatively similar cluster pattern. However it was the DArT marker which consistently showed more efficiency, revealing higher PCA score values, bootstrap values and clearly structured PCO and clusters consistent with known origins. Similar studies on the comparison of DArT and SSR by Mantovani *et al*., (2008) revealed that cluster classification for DArT was more robust as compared to the one obtained through SSR markers or at least, is more functionally useful. Which they suggested could be due to the relatively medium to high numbers of polymorphic markers for DArT that can be identified and it is difficult to get a similar numbers for SSR markers

3.5 Conclusions

Genetic analysis for the two techniques broadly showed a similar pattern of clustering; grouping landraces was mainly based on their areas of origin. Both clustering and PCoA, DArT markers consistently defined landrace on their areas of origin. SSR marker revealed a higher differentiation among the landraces, shown by lower average genetic similarities. Comparatively the initial costs for the two markers are relatively similar (Hurtado *et al.,* 2008), but the DArT

markers have the advantage of less cost per assay and this makes DArT markers more attractive. In this study DArT markers proved to be relatively superior, however fewer genotypes were used for the comparison of the two markers.

A summary of the achievements in this chapter is the development and characterisation of bambara groundnut and the comparison of SSR markers and DArT. A set of 68 markers were characterised and found to be suitable for later use in genetic studies of bambara groundnut

CHAPTER FOUR: Phenotypic diversity for morphological and agronomic characters of bambara groundnut

4.1 Introduction

Bambara groundnut grows well in low input cropping systems hence it is one of the legume crops preferredby many subsistence farmers (Harris and Azam-Ali, 1993). Physiological variation in bambara groundnut has been recorded for different responses to photoperiod (Linnemann *et al*., 1995), sowing date (Sesay *et al*., 2008), moisture deficit (Collison *et al*., 1999; Mwale *et al*., 2007) and growth rate (Massawe *et al*., 2003; Collinson *et al*., 1996). Confirming there could mean single genotypes have different responses but this provides a good opportunity for bambara groundnut variety development. However, at the moment yields are unreliable and low; due to the lack of developed varieties and farmers are still planting landraces (Zeven, 1998).

Crop genetic diversity is important for crops to withstand pest and diseases and is useful for plant breeders to enhance the breeding progress of traits of economic value such as yield. A wide range of phenotypes provide more insurance and during harsh climatic changes crops with favourable characters survive better than poor ones, therefore enough genetic diversity will ensure survival of the crop/species.

Knowledge of phenotypic diversity has been employed in crop improvement in developing breeding lines that have stable yields and across various environments; for example in chickpea (Yaghatipoor and Farshdfar, 2007);and in breeding for disease tolerant and drought tolerant genotypes in groundnut (Puttha *et al*., 2008; Painawadee *et al*., 2009). Durán *et al.*, (2005) used morphological characteristics to estimate phylogenetic relationships among lines of Caribbean bean landraces within the Andean and Mesoamerican gene pools. Morphological characters managed to identify two clusters, one with Mesoamerican characteristics which includes red mottled lines, while the Andean characteristics included all the lines from Puerto Rico and the Dominican Republic. These examples demonstrate that even though morphological markers can be influenced by the environment, their application especially to underutilized crops is still very important.

Yield is usually a complex trait and controlled by a number of genes as well as influenced by the environment. The application of correlation analysis - association between two or more characters - is important to understand how an improvement in one character could cause simultaneous changes to other characters (Falconer and Mackay, 1996). In order to develop high yielding varieties it is important to study the genetic variation for yield and yield components which are in turn influenced by the genetic and environmental causes (Maniee *et al*., 2009). There is limited amount of work on heritability and genetic advance on the quantitative characters of bambara groundnut.

4.1.1 Correlation analysis studies

Yield is an important and complex trait difficult to manipulate for crop improvement (Shi *et al.,* 2009), however yield such as *seed number per plant*, *seed yield per hectare*, *pods number per plant* and *100 seeds weight* could be correlated to other characters. This will then allow an indirect selection of yield based on those characters.

Relatively low correlations were found among most of the traits among 1384 bambara groundnut accessions at the International Institute of Tropical Agriculture (IITA) Nigeria, by Goli *et al.,* (1995). They found a strong correlation of *seed yield per plant* to a number of characters which they identified as potential characters to select for bambara groundnut improvement. They recorded correlations of *seed yield per plant* to *number of seeds per pod* of(r = 0.88), and to *pods per plant (r* = 0.30*).* A positive correlation of r =0.13 was observed between *100 seed weight* and*seed yield per plant* which is a good indication that these two characters could be used effectively in the selection of bambara groundnut. *Number of stems per plant* was positively correlated with *days to maturity*, which was an indication that plants with more stems matured late, a negative correlation was found between *days to maturity* and *cercospora* virus index, which could indicate that fast maturing plants have a lower probability of infection; however no heritability studies were undertaken.

Table 4.1.1: A comparison of correlations between yield components; seed yield per plant, number of pods per plant, seed yield per hectare and 100 seed weight and a number of characters , sourced from Karikari and Tabona, (2004); Misangu *et al.*, (2007); Ouedraogo *et al.*, (2008); Goli *et al.*, (1995); Jonah *et al.*, (2010); Karikari, (2000), and Oyiga and Uguru, 2011.

No.	Characters	Materials used in the study	Correlation character	Correlation Values	Reference
1	Seed yield per plant	12 landraces	Number of pods per plant	0.764	Karikari and Tabona, 2004
		9 landraces		0.83	Misangu *et al.*, 2007
		310 accessions		0.852	Ouedraogo *et al.*, 2008
		1384 accessions		0.3	Goli *et al.*, 1995
		12 landraces		0.33	Jonah *et al.*, 2010
		12 landraces	100 seed weight per plant	0.415	Karikari and Tabona, 2004
		9 landraces		0.16	Misangu *et al.*, 2007
		310 accessions		0.257	Ouedraogo *et al.*, 2008
		12 landraces		0.06	Jonah *et al.*, 2010
		12 landraces	Shelling percentage	0.587	Karikari and Tabona, 2004
		310 accessions		0.275	Ouedraogo *et al.*, 2008
		12 Landraces		-0.1	Jonah *et al.*, 2010
		9 landraces	Plant height	0.38	Misangu *et al.*, 2007
		310 accessions		0.026	Ouedraogo *et al.*, 2008
		1384 accessions		0.08	Ouedraogo *et al.*, 2008
		12 Landraces		0.42	Jonah *et al.*, 2010
		9 landraces	Days to maturity	-0.41	Misangu *et al.*, 2007
		1384 accessions		-0.01	Ouedraogo *et al.*, 2008
		9 landraces	Seed width	0.11	Misangu *et al.*, 2007
		310 accessions		0.224	Ouedraogo *et al.*, 2008
		12 Landraces		-0.23	Jonah *et al.*, 2010
		9 landraces	Seed length	-0.08	Misangu *et al.*, 2007
		310 accessions		0.295	Ouedraogo *et al.*, 2008
		12 Landraces		-0.05	Jonah *et al.*, 2010
		9 landraces	Number of leaves per plan	0.34	Misangu *et al.*, 2007
		1384 accessions		0.12	Goli *et al.*, 1995
		9 landraces	Leaf width	0.11	Misangu *et al.*, 2007
		1384 accessions		0.13	Goli *et al.*, 1995
		9 landraces	Leaf length	0.5	Misangu *et al.*, 2007
		1384 accessions		0.14	Goli *et al.*, 1995
		310 accessions	Pod length	0.194	Ouedraogo *et al.*, 2008
		1384 accessions		0.1	Goli *et al.*, 1995
		12 Landraces		0.13	Jonah *et al.*, 2010
		310 accessions	Pod width	0.092	Ouedraogo *et al.*, 2008
		1384 accessions		0.08	Goli *et al.*, 1995
		12 Landraces		0.15	Jonah *et al.*, 2010
		310 accessions	Canopy spread	0.231	Ouedraogo *et al.*, 2008
		1384 accessions		0.15	Goli *et al.*, 1995
		12 landraces	Number seeds per pod	0.202	Karikari and Tabona, 2004
		1384 acessions		0.04	Goli *et al.*, 1995
		12 landraces	Number of seeds per plan	0.882	Karikari and Tabona, 2004
		12 landraces	Seed yield per hectare	0.8	Jonah *et al.*, 2010
2	Number of pods per pl	1384 accessions	100 seed weight	0.12	Goli *et al.*, 1995
		12 landraces		-0.71	Jonah *et al.*, 2010
		12 landraces	Shelling percentage	0.35	Jonah *et al.*, 2010
		310 accessions		-0.172	Ouedraogo *et al.*, 2008
		9 landraces	Plant height	0.55	Misangu *et al.*, 2007
		1384 accessions		0.23	Goli *et al.*, 1995
		13 genotypes		0.66	Oyiga and Uguru, 2011
		12 landraces		0.25	Jonah *et al.*, 2010
		310 accessions	Seed width	0.023	Ouedraogo *et al.*, 2008
		12 landraces		-0.74	Jonah *et al.*, 2010
		310 accessions	Seed length	0.073	Ouedraogo *et al.*, 2008
		12 landraces		0.51	Jonah *et al.*, 2010
		9 landraces	Number of leaves per plan	0.31	Misangu *et al.*, 2007
		1384 accessions		0.36	Goli *et al.*, 1995
		13 genotypes		0.663	Oyiga and Uguru, 2011
		9 landraces	Leaf length	0.56	Misangu *et al.*, 2007
		1384 accessions		0.27	Goli *et al.*, 1995
		12 landraces	Pod length	0.67	Jonah *et al.*, 2010
		310 accessions		0.03	Ouedraogo *et al.*, 2008
		1384 accessions		0.25	Goli *et al.*, 1995
		12 landraces	Pod width	-0.67	Jonah *et al.*, 2010
		310 accessions		-0.079	Ouedraogo *et al.*, 2008
		1384 accessions		0.18	Goli *et al.*, 1995
		12 landraces	Seed yield per plant	0.33	Jonah *et al.*, 2010
		310 accessions		0.852	Ouedraogo *et al.*, 2008

Table 4.1.2(Continued)

No.	Characters	Materials used in the study	Correlation character	Correlation Values	Reference
3	Seed yield per hectare	9 landraces	100 seed weight	0.88	Karikari , 2000
		12 landraces		0.16	Jonah *et al.,* 2010
		9 landraces	Shelling percentage	0.82	Karikari , 2000
		310 accessions		-0.054	Ouedraogo *et al.,* 2008
		12 landraces		-0.11	Jonah *et al.,* 2010
4	100 seeds weight	310 accessions	Number of pods per plant	-0.054	Ouedraogo *et al.,* 2008
		1384 accessions		0.12	Goli *et al.,* 1995
		310 accessions	Shelling percentage	0.187	Ouedraogo *et al.,* 2008
		9 landraces		0.88	Karikari, 2000
		12 Landraces		0.1	Jonah *et al.,* 2010
		310 accessions	Plant height	0.096	Ouedraogo *et al.,* 2008
		1384 accessions		0.11	Goli *et al.,* 1995
		12 landraces		0.27	Jonah *et al.,* 2010
		9 landraces	Days to maturity	-0.8	Karikari, 2000
		1384 accessions		-0.14	Goli *et al.,* 1995
		310 accessions	Seed width	0.524	Ouedraogo *et al.,* 2008
		9 landraces		0.85	Jonah *et al.,* 2010
		310 accessions	seed length	0.529	Ouedraogo *et al.,* 2008
		12 landraces		0.79	Jonah *et al.,* 2010
		310 accessions	Pod length	0.44	Ouedraogo *et al.,* 2008
		1384 accessions		0.29	Goli *et al.,* 1995
		12 landraces		0.81	Jonah *et al.,* 2010
		310 accessions	Pod width	0.491	Ouedraogo *et al.,* 2008
		1384 accessions		0.48	Goli *et al.,* 1995
		12 landraces		0.62	Jonah *et al.,* 2010
		310 accessions	Canopy spread	0.23	Ouedraogo *et al.,* 2008
		9 landraces		-0.86	Karikari, 2000
		1384 accessions		0.16	Goli *et al.,* 1995

Karikari, (2000) when studying the variability of Botswana and Zimbabwean landraces in a field experiment he found a significant correlation of *days to flowering* (0.84), *100 seeds weight*(r = 0.88), and *shelling percentage* (r = 0.82), to *grain yield kg ha*$^{-1}$ of bambara groundnut. A negative correlation was recorded between *grain yield kg ha*$^{-1}$to both *canopy spread* at (r = -0.85), and *days to maturity* at (r = -0.63). He also carried out a heritability analysis to identify the best traits for selection in the Botswanan environment. He recorded heritability (h^2) values of 0.72 for *grain yield*, 0.25 for 100 *seed weight*, 0.38 for *shelling percentage*, and 0.36 for *plant dry matter* at harvest. Ouedraogo, *et al.,* (2008) characterised and evaluated 310 accessions from Burkina Faso, *yield per plant* was correlated against a number of characters and revealed a positive correlation to*days to flowering* (0.06),*100seeds weight*(r = 0.257) but low correlations. In contrast *canopy spread* had a positive correlation(r = 0.231) while *shelling percentage* had a negative correlation (r = -0.199) to *yield per plant.* The characters were not subjected to heritability analysis.

Karikari and Tabona (2004) undertook a study on 12 bambara groundnut landraces to identify characters associated most with adaptation to drought in the Botswanan environment. Their results showed *canopy spread*, *100 seed weight*,

root-shoot ratio, and *number of seeds per pod* as the most suitable characters. A highly significant correlation was found between *seed yield per plant* with *number of pods* at r = 0.76, and between *seed yield per plant* with *number of seeds per plant* at r = 0.88. They emphasised the importance of *root-shoot ratio* in the semi-arid environment of Botswana, and they found that the *root-shoot ratio* has significant correlation with *shelling percentage* and *seed yield per plant* at r = 0.296 and r = 0.398, respectively.

Jonah *et al*., (2010), investigated the genetic correlations between yield and yield related characters in 12 bambara groundnut landraces in Nigeria. Highly positive correlations were found between *seed yield per hectare* and *pod yield per plant* (r =0.87), and between *seed yield per hectare* and *seed yield per plant* (r = 0.91), and between *seed yield per plant* and *plant height at 8 weeks after sowing*(r =0.77). A high correlation was also identified between *pod length* and *pod width* at (r =0.89) and *seed length* and *seed width* at (r =0.82), which is potentially useful for selecting genotypes with bigger seeds. However they found a negative correlation between *pod number per plant* with *100 seed weight* at (r= -0.74) which implies that selecting landraces for higher pods numbers could lead to, leaner pods produced in turn.

Wigglesworth, (1996), undertook a field trial on six bambara groundnut landraces in Botswana in order to study the genotypic variation and heritability of *pod numbers*, *100 seed weight*, *seed weight per plant* and to find some correlation between the traits. The results recorded a significant phenotypic correlation between *pod numbers* and *seed weight per plant* (r = 0.77), and between *100 seed weight* and *seed weight per plant* (r = 0.52). Heritability values recorded were lower for *seed weight per plant* (0.25), *pod number* (0.39) and higher for *100 seed weight* at (0.94). Therefore *100 seed weight* was singled out as an important trait to select for, among the local landraces.

Thirteen bambara groundnut populations were evaluated for floral structure in Nigeria in a field experiment byOyiga*et al*., (2010). They recorded an *anther diameter* correlation with number of *pods per plant* (r =0.41), and to *seed weight per plant* (r =0.51) which is an indication of the relationship between seed number and the biomass synthesized during the growth stages of seed formation (Jeuffroy

and Chabanet, 1994). Selection for *anthers* with larger diameter was identified as a strategy for yield improvement in bambara groundnut. They also found a negative correlation between *stigma anther* separation with *seed weight per plant* at (r = -0.59), which they thought the *stigma-anther* separation is an important factor in the production of low seed *weight* in bambara groundnut.

In another study, Jonah *et al*., (2010), carried out a phenotypic diversity study on 12 bambara groundnut landraces. They undertook some broad sense heritability and genetic advance estimates and recorded high heritability and genetic advance in *pod yield per plant* (0.75; 16%), for *seed width* (0.85; 16%), and for *100 seed weight* (0.70; 12%) the high heritability and genetic advance suggest that these are selectable traits. However, they reported moderate value for *seed yield per hectare* at 0.54, which indicate a limitation to the improvement of this trait.

4.1.2. Selection of lines for breeding

For a robust plant breeding program the selection process should be effective enough to capture those individuals with which are superior in a number of traits (Strefeler and Wehner, 1986). It is usually genotypes with the superior characters that are recommended over others for crop improvement and usually the selection indices are used to identify those genotypes. In many breeding programs more than one trait is been selected for at the same time, and thus multiple selection indexes are used. Several simultaneous selection indices such as those by Smith (1936) and Hazel (1943) are used (Tardin *et al.,* 2007). The selection index method is expected to be faster in generating benefits, the method assigns suitable weights for each trait depending on its importance (Eshghi *et al.,* 2011). The concept of selection index as was developed by Smith (1936) was found to have some difficulties like determining relative economic values thus several modifications have been made (Baker, 1974).

The efficiency of selection indices does not depend only on the estimation ofthe coefficient, but also on the crop and characters under study. Monirifar, (2010) adds after evaluating and constructing some selection indices for use in alfalfa (*Medicago sativa*). In this study, *leaf area, shoot dry weight* (*biomass), yield* such as *pod number per plant* and *seed number per plant* were characters selected for use in the selection index. Canopy development is reported as an important

determinant of crop radiation capture and is mainly influenced by temperature (Massawe *et al*., 2003) while Collinson *et al*., (1999) reported the effect of soil moisture deficit and its impact on leaf area development and yield reduction in three bambara groundnut landraces.

The selection index (SI) adopted from Monirifar, (2010) comes up with one value from several variables that had been selected as of economic value by the breeder. This technique also requires the use of some weight for each variable to include in the equation, but with different weight assigned depending on its importance as deemed by the plant breeder (Baker, 1974). The selection index is often used in the breeding program of cassava at the International Center for Tropical Agriculture (CIAT) (Ceballos *et al.,* 2007). The selection index (SI) has recently been utilised in soybean germplasm evaluation for acid tidal swamp tolerance (Kuswantoro *et al.,* 2010). A greenhouse and a field experiment were conducted based on assessing 17 genotypes for six characters. One genotype with the highest ranking because of its root and shoot dry weight was identified. Ojulong *et al.,* (2010) used a similar selection index to evaluate cassava seedlings developed for yield characteristics, and identified traits which they suggested could be included in selection of the crop such as root weight and root weight per tree.

4.1.3 The objectives of this study were

- To evaluate and characterise bambara groundnut landraces based on agro-morphological characters
- To assess the genetic diversity of bambara groundnut landraces based on Shannon weaver index, genotypic variability, phenotypic variation, heritability and genetic advance
- To classify bambara groundnut genotypes by means of cluster and principal component analysis in order to select genotypes suitable for further breeding
- To identify better performing lines in a Botswana environment based on seed yield and biomass production based on the selection index and Duncan Multiple Range Test

4.2 Results

4.2.1 Qualitative analysis of the genotypes

The frequency distribution among the 35 bambara groundnut planted in the glasshouse and 34 bambara groundnut lines planted in the field experiment, shows that majority of the seed colour of the landraces were reddish in colour (77.2% of the landraces planted in the agronomy bay, and 85.1% of the genotypes that were planted in the field experiment) based on classes 3, 4, 5, and 6 for *testa colour* (Table 4.1.2). Twolandraces had cream coloured seeds while three had black seedsin both the glasshouse and field experiment. Most landraces had no *eye pattern* (91.4%; 92.9%) and no *testa pattern* (82.9%; 82.1%) and only 2 seeds had dotted spot. Even though farmers have been observed to plant a mixture of colours for bambara groundnut, two surveys done in Swaziland and Botswana revealed that the most preferred landraces are the cream coloured ones (Sesay *et al.*, 2003; Brink *et al.*, 1996). In this study all pod colour classes were observed with the exception of black ones, and most of them were those which are *pointed with a nook* (60%; 64.2%)(Table 4.1.2).

The crops experienced little stress in both the experiments (88.6%; 85.3%) the temperature recordings in the glasshouse had an average of 21.7^{o}C while in the field an average of 29.5^{o}C was observed and these are ideal temperatures for bambara groundnut (Swanevelder, 1998). Three types of plant growth habit were observed 47.1% were bunch type, 38.2% were semi-spreading and 14.7% were spreading types, which shows that farmers are mostly selecting for the bunch and semi spreading types.

Table 4.1.2: Descriptor, classes and frequency distribution among the 35 landraces planted in the agronomy bay and 34 bambara groundnut lines selected and planted in the field in Botswana

Descriptor and Classes	Frequency of class (%)								
	0	1	2	3	4	5	6	7	8
Testa colour		**5.7**	**0**	**14.3**	**11.4**	**28.6**	**22.9**	**8.6**	**8.6**
1= Cream 2= Grey 3=Light red 4= Dark red 5= Brownish red 6= Dark brown 7=Dark purple 8= Black		7.1	0.0	14.3	10.7	42.9	17.9	0.0	7.1
Eye pattern	**91.4**	**5.7**	**0.0**	**0.0**	**2.9**	**0.0**	**0.0**		
0=No eye pattern 1= Butterfly 2=Triangular 3= Mottled 4= Thick dotted lines 5=Circular 6=Thin lines	92.9	3.6	0.0	0.0	3.6	0.0	0.0		
Testa pattern	**82.9**	**5.7**	**0.0**	**0.0**	**5.7**	**2.9**	**0.0**	**2.9**	**0.0**
0=no pattern 1= Entire 2=Striped 3=Marbled	82.1	0.0	0.0	0.0	7.1	3.6	0.0	7.1	0.0
4=Dotted 5=Little rhomboid one side spotting 6=Little rhomboid two side spotting 7=Much rhomboid 8= Holstein									
Pod colour		**21.4**	**39.3**	**21.4**	**17.9**	**0.0**			
1=Yellowish brown 2= Brown 3=Reddish brown 4 =Purple 5=Black		21.4	39.3	21.4	17.9	0.0			
Pod texture		**42.9**	**31.4**	**17.1**	**8.6**				
1= Smooth 2= Little grooved 3=Much grooved 4 = Much grooved		46.4	28.6	17.9	7.1				
Pod shape		**8.6**	**28.6**	**60.0**	**2.9**				
1=Without point 2=Pointed 3=Pointed and nooked 4=Pointed both sides		3.6	28.6	64.3	3.6				

Table 4.1.2 continued

Seed shape				**14.3**	**85.7**			
1 =Round	2= Oval			21.4	78.6			
Terminal leaflet colour				**74.3**	**0.0**	**25.7**		
1=Green	2=Red	3=Purple		73.5	0.0	26.5		
Stress susceptibility				**88.6**	**2.9**	**5.7**	**2.9**	**0.0**
1= No visible sign	2=Low	3= high		85.3	2.9	5.9	5.9	0.0
4= High	5= very high							
Leaf shape				**0.0**	**65.7**	**31.4**	**2.9**	
1=Round	2=Oval	3=Lanceolate		0.0	58.8	38.2	2.9	
4= Elliptic								
Stem hairiness			50.0	47.1	2.9			
0=absent	3=Sparse	5=Dense						
Leaf colour at germination				41.2	58.8			
1=Green	2=Purple							
Growth habit				47.1	38.2	14.7		
1=Bunch	2=Semi-bunch	3=Spreading						

The ones in bold: Glasshouse experiment results

4.3.2 Shannon Weaver (H') diversity analysis

Knowledge of variation of characters is important to plant breeders since they should know which population is more varied for which characters. Shannon Weaver (H') within population index (Hennink and Zeven 1991), therefore (H') can be useful in identifying those traits that warrant the attention of breeders to improve. The estimates of Shannon weaver (H'), was relatively high with a similar mean diversity 0.70 in UK and 0.69 in Botswana (Table 4.1.3). The diversity ranged from 0.19 (*Leaflet length*) to 0.97 (*petiole-internode ratio*) in UK and in Botswana it ranged from 0.19 (*leaflet width*) to 0.99 (*pod width*).

Table 4.1.3: Shannon-Weaver index on the phenotypic diversity of 24 quantitative characters in the agronomy bay experiment and the field experiment (Botswana).

	UK	BOTSWANA
Characters	H'	H'
Days to emergence	0.86	0.32
Days to 50% flowering	0.96	0.43
Leaf number.	0.71	0.71
Spreading	0.77	0.74
Leaflet length	0.19	0.73
Leaf width	0.97	0.19
Leaf Area	0.88	0.84
Plant height	0.79	0.92
Internode	0.67	0.54
Petiole	0.79	0.92
Pet-Internode	0.97	0.86
Petiolule	0.24	0.73
Peduncle	0.19	0.73
Stem number.	0.96	0.94
Days to maturity	0.42	0.79
Shoot dry weight	0.93	0.88
Pod numbers per plant	0.93	0.61
Pod dry weight	0.94	0.56
Pod length	0.70	0.77
Pod width	0.32	0.99
Seed number.	0.90	0.56
Seed length	0.32	0.59
Seed width	0.51	0.86
Seed weight	0.95	0.28
Mean diversity (H')	**0.70**	**0.69**

The characters that showed greatest variation between the two sites were the *leaflet length*, *leaflet width*, *seed weight*, *and pod width*. This variance was reflected in *pods number per plant* and *seeds number per plant* which was drastically affected in the field experiment.

The Shannon-Weaver diversity index (H') was calculated on the qualitative characters to compare the genetic diversity among characters both in UK and Botswana. The most diverse characters were *pod colour* (0.94) and *testa colour* (0.93) in UK, *leaf colour at emergence* (0.98) and *pod colour* (0.96) in the Botswana. In both UK and Botswana the least diverse characters was *eye pattern* 0.32 and 0.28 respectively.

Table 4.1.4: Shannon weaver index on phenotypic diversity of qualitative characters for the studied landraces in agronomy bay and field experiment

	UK	Botswana
Character	H'	H'
Pod texture	0.89	0.87
Pod colour	0.94	0.96
Pod shape	0.70	0.63
Seed shape	0.59	0.75
Testa colour	0.93	0.87
Testa pattern	0.47	0.42
Eye pattern	0.32	0.28
Leaflet colour	0.82	0.83
Leaf shape	0.67	0.71
Stress susceptibility	0.34	0.41
Leaf colour at emergence	*	0.98
Stem hairiness	*	0.73
Growth habit	*	0.91
Mean diversity (H')	0.67	0.72

* Not recorded

These considerable variation, identified based on (H') is important for bambara groundnut improvement, however there is non-significant and low correlation(r = 0.168) between diversity analysis values in UK and Botswana field experiment for quantitative characters, compared to a highly significant correlation (r = 0.953) for

the qualitative characters, based on Pearson correlation analysis. These also reflect the lower effect of the environment on qualitative compared to quantitative characters.

4.3.3 Descriptive analysis of the genotypes

The average genetic diversity was slightly higher in Botswana (0.71) compared to UK (0.67) and substantial variability among the genotypes was revealed in most of the 24 characters (Table 4.1.5, and 4.1.6) for both experiments in the UK (agronomy bay) and Botswana (field experiment). For example, the minimum and maximum shown in the UK agronomy for *shoot dry weight* are 6.9 g -113.9g, *leaf area* 1341 cm^2 – 4489 $cm^{2,}$*pod numbers per plant* 7 - 182 and for *seeds numbers per plant*7 – 155, while in the field experiment the ranges for *shoot dry weight* is 12.8 g - 113.7g, *leaf area* 52 cm^2 - 3304 cm^2, *pod number per plant* 2 - 138 and for *seed number per plant* 1- 140. This show there is great potential for selection in these traits for further crop improvement.

Table 4.1.5: Descriptive characteristics for the 35 bambara groundnut planted UK (Agronomy bay, 2008) from an average of three plants, the vegetative characters recorded at 10 weeks after planting, while the yield characters are recorded after harvest.

Characters	Minimum	Maximum	Mean	Stdev	CV	F test
Days to emergence (d)	8	18	10.1	1.3	12.9	*
Days to 50% flowering (d)	38	54	42.4	2.4	5.8	***
Number leaves per plant	33	293	79.4	23.5	29.6	***
Canopy width (cm)	7	51	19.5	4.7	24.3	***
Leaflet length (cm)	6.8	10.9	8.6	0.5	5.2	***
Leaflet width (cm)	2.3	5.3	3.7	0.3	8.4	***
Leaf Area (cm^2)	1341	4489	17034	1795.7	10.5	***
Plant height (cm)	22	46	33.5	2.3	6.8	***
Internode length (mm)	0.8	6	2.3	0.4	15.7	***
Petiole length (cm)	8.5	25.5	16.8	1.5	8.7	***
Petiole-Internode ratio	3.3	17.6	8	1.5	18.1	***
Petiolule length (mm)	0.9	4.8	2.3	0.3	14.4	***
Peduncle length (mm)	1	4.9	2.3	0.5	19.8	***
Number of stem	4	22	9.9	2.5	25.4	**
Days to maturity (d)	109	161	155.3	7.5	4.9	ns
Shoot dry weight (g)	6.9	113.9	33.3	11.5	34.4	***
Number of pods plant	7	182	54.4	19.8	36.3	***
Pod dry weight (g)	2.1	97.5	32.7	12	36.7	***
Pod length (mm)	12.3	24	18.8	1.5	8	***
Pod width (mm)	7.8	15.8	12.1	0.9	7.6	***
Number of seed plant	7	155	59.1	18.4	31.2	***
Seed length (mm)	6.6	13.4	10.7	0.9	8.9	***
Seed width (mm)	5.3	10.3	8.4	0.7	8.2	***
Seed weight (g)	1	69.8	23.1	8.6	37.2	***

*, **, *** Significant at 5%, 1% and 0.1% respectively, ns = non-significant

Table 4.1.6: Descriptive characteristics for the 34 bambara groundnut planted in field (Notwane, Botswana, 2008/2009 season) with three replications, the vegetative characters recorded at 10 weeks after planting, while the yield characters are recorded after harvest.

Characters	Minimum	Maximum	Mean	Stdev	CV	F test
Days to emergence (d)	10	19	15.1	1.7	11	***
Days to 50% flowering (d)	42	67	56.7	2.8	5	***
Number leaves per plant	34.5	233.8	112.7	20.4	18	***
Canopy width (cm)	98.4	577.8	240.6	31.8	13	***
Leaflet length (cm^2)	22.4	94	66.9	6.6	10	***
Leaflet width (cm)	14.8	40	24.9	2.2	9	***
Leaf Area (cm)	52	8369	3304	857	25.9	***
Plant height (cm)	183	361.8	278.1	22	8	***
Internode length (mm)	11.2	86.6	22.7	6.3	28	***
Petiole length (cm)	63	191.4	139.4	15.5	11.1	***
Petiole-Internode ratio	2.3	14.5	7.2	1.2	16	***
Petiolule length (mm)	6.5	38	17.9	3.2	18	***
Peduncle length (mm)	6.1	37.6	20.2	4.8	24	***
Number of stem	4	16.2	8.3	1.7	20	***
Days to maturity (d)	126.8	155	136.9	2.5	2	***
Shoot dry weight (g)	12.8	113.7	41.8	9.4	23	***
Number pods plant	1.5	137.5	16	12.7	79	***
Pod dry weight (g)	0.3	54.1	8.2	6.9	83	ns
Pod length (mm)	12	26	18.2	1.8	10	***
Pod width (mm)	7.3	14	10.5	1.1	10	***
Number of seed plant	1	140.5	15.6	12.4	79	**
Seed length (mm)	7	16	11	1.2	11	***
Seed width (mm)	4.7	10.4	8.1	0.9	11	**
Seed weight (g)	0.1	30.2	5.3	4.2	80	**

*, **, *** Significant at 5%, 1% and 0.1% respectively, ns = non-significant

The estimate of coefficient of variation (CV) was used to compare the variability of different characters of the 35 genotypes grown in the UK (agronomy bay experiment) and 34 genotypes in Botswana (field experiment), Table 4.1.5 and 4.1.6). The 34 bambara groundnut lines were derived from seeds from single plants selected among the 35 accessions from the agronomy bay experiment (Table 2.1.2.2). The characters that revealed higher variations are the *yield* and yield related traits, with a relatively higher CV in the greenhouse experiment, such as *pod numbers* (36%), *pod weight* (37%), *seed numbers* (31%) and *seed weight* (22%). While in the field the *coefficient of variation* was higher for similar traits

pod numbers (79%), *pod weight* (83%), *number of seed* (79%) and *seed weight* (80%).

Lower coefficients of variation were revealed in characters such as *days to emergence* (4.9%), *leaflet length* (5.2%), *days to flowering* (5.8%), days *to 50% flowering* (5.8%) and *plant height* (6.8%) in the agronomy bay experiment. In the field experiment lower coefficients of variation were recorded in *days to 50% flowering* (5.75%), *leaflet length* (5.23%), *leaflet width* (8.38%), *plant height* (6.81%), *petiole length* (8.70%), *days to maturity* (4.85%), *pod width* (7.56%), *pod length* (7.95%), *seed length* (8.88%) and *seed width* (8.24) which indicates the small variation in both environments for these traits, it also suggests that these are more likely to be highly heritable traits.

Highly significant differences ($P<0.001$) were detected among genotypes in most of the 24 characters that were analysed in the Agronomy bay (glasshouse) and in the field experiment in Botswana. With the exception of *days to maturity* in the glasshouse, and *pod dry weight* in the field experiment (Table 4.15 and 4.1.6). This is an indication that there is a substantial amount of genetic variability among the traits analysed for this set of accessions.

The largest value for *seed number* and *pod number per plant* were shown by genotype 88-AHM753 from Namibia, in UK (agronomy bay experiment) with 121 seeds per plant, and 121 pods per plant followed by genotype 91UNIS R from Swaziland with 107 seeds per plant and genotype 85 Acc754 from Zambia with 106 pods numbers per plant. The largest *shoot dry weight* was shown by 118-Ramayana from Indonesia with 81 g, while the largest *leaf area* was attained by 48Acc790 from Kenya at 10369cm^2, both were followed by genotype 85 Acc 754 from Zambia with 76 g *shoot dry weight* and 9997cm^2 *leaf area*. The *two leaf areas* although different, there are not significantly different from each other according to Duncan's range multiple test (Table 4.2.5).

The genotypes with the lowest *seed number* and *pod number per plant* were 70Acc330 and 69Acc286 both from Nigeria with 15 and 12 seeds each, and only 10 and 11 seeds each, respectively. The lowest *shoot dry weight* and *leaf area* were shown by genotypes 56Acc89 from Mali at 10g *shoot dry weight* and 1357

cm^2 leaf area, followed by 6Acc289 from Benin for *shoot dry weight* at 10g, and 50Acc792 from Kenya with 1887 cm^2 for *leaf area.*

For the field experiment in Botswana, line 90S19-3 from Namibia produced the largest *number of pods per plant* and *seeds per plant* of 68 pods and 66 seeds, respectively. The second highest yielding line was 84Acc696 from Zambia with an average of 36 pods and 33 seeds per plant. The greatest *shoot dry weight* was recorded from 76Acc390 from Sudan with an average 103 g per plant followed by 81Acc385 from Tanzania with an average of 83 g per plant, both lines also had the largest leaf area, of 5986 cm^2 for 76Acc390, with 81Acc385 coming second with 5437 cm^2. However, these two lines produced a relatively low number of pods with an average of 10 pods per plant. This shows that they concentrated most of the assimilates in leaf formation instead of pods and could indicate issues with fertility/pod/seed set.

The lowest *shoot dry weights* were recorded from lines 69Acc286 and 70Acc329 from Nigeria, and 45Acc231 from Ghana, both lines failed to produce any pods or seed. Fewer numbers of pods were recorded in lines 95DODRfrom Tanzania and 113Bots5 from Botswana. In a drought experiment conducted in the glasshouse by Collinson *et al., (*1999) 95DODR compared favourablywell withone landrace from Botswana (DIPC), in terms of pod number produced and final harvest biomass. Its poor performance in Botswana could imply that it is not well adapted to the Botswanan environment.

The line 113Bots5 was acquired from the Francistown market in the north eastern part of Botswana in 2004/2005 season, in an effort to study and improve some bambara groundnut landraces. The landrace was not included in field trials before; its poor performance shows that it may be originally from a different environment from the Botswanan environment.

4.3.4 Principal component analysis:

The 24 characters assessed in the Agronomy bay experiment were subjected to principal component analysis to identify characters that accounted for most of the variance in the selected genotypes. Only the first six principal components with eigenvalues more than one were selected, giving an accumulated total variation of

79.24% (Table 4.1.7). The first principal component (PC 1) contributed 35% of variation and characters with higher loadings were *canopy width, plant height, shoot dry weight, petiole length, internode length, pod number* and *leaf area*. The second principal component accounted for 14% of the total variation and higher loadings were observed from mostly the vegetative part of the crop*, petiole-internode ratio, petiolule length, leaf area, petiolelength* and*shoot dry weight. Leaflet length, days to 50% flowering, days to emergence, pod length and peduncle length* accounted for most of the 11.88% identified at the third principal component. Characters contributing most variation among the genotypes in the fourth principal component were *days to maturity, leaf numbers, days to 50% flowering, internode length,* and *days to emergence*. Principal components (PC5) and (PC6) contributed approximately 6% and 5 % respectively, with *pod dry weight, days to emergence, petiolule, petiole-internode ratio* and *days to maturity* accounting for most of the 6 % in principal component 5. Similarly *leaflet length, day to emergence, pod dry weight, number of stems* and *internode length*are contributing higher loadings in principal component 6. Generally when observing PC1 and PC2 in the agronomy bay experiment it shows that characters that were able to separate the genotypes are mainly vegetative, with *shoot dry weight* and *petiole length* appearing in both PC1 and PC2, while *pod number* contributedonly in the first principal component.

Table 4.1.7: Principal components, matrix of eigenvalues and vectors for 24 quantitative characters of bambara groundnut landraces planted in the agronomy bay (UK).

	Principal components (PC)					
	PC1	PC2	PC3	PC4	PC5	PC6
Eigenvalues Variance	8.459	3.359	2.851	1.74	1.384	1.223
% Total contribution	35.25	14.00	11.88	7.25	5.77	5.09
%Accumulated	35.25	49.25	61.13	68.38	74.15	79.24
Days to emergence	-0.108	0.095	0.264	0.115	0.282	0.357
Days to 50% flowering	-0.029	0.122	0.339	0.321	0.164	-0.326
Days to maturity	0.233	-0.181	-0.057	0.322	0.247	-0.242
Plant height	0.295	0.081	0.157	0.055	-0.168	0.104
Internode length	0.255	-0.102	0.045	0.165	-0.229	0.197
Leaflet length	0.102	0.018	0.343	0.101	0.126	0.384
Leaf number per plant	0.233	-0.181	-0.057	0.322	0.247	-0.242
Leaflet width	0.209	0.017	0.142	0.115	0.159	0.171
Leaf Area	0.249	0.265	0.052	-0.029	-0.120	-0.119
Pod dry weight	0.044	-0.237	-0.078	0.013	0.538	0.349
Peduncle length	0.033	-0.337	0.204	0.081	-0.265	0.119
Petiole length	0.275	0.168	0.144	0.106	-0.131	0.054
Petiolule length	0.162	0.291	0.154	-0.362	0.266	-0.136
Pet-Internode ratio	0.162	0.291	0.154	-0.362	0.266	-0.136
Pod length	0.033	-0.281	0.229	-0.387	0.114	-0.056
Pod width	0.152	-0.313	0.184	-0.070	0.002	-0.301
Pod number per plant	0.253	0.064	-0.359	0.071	0.053	0.009
Shoot dry weight	0.281	0.162	0.042	0.053	-0.084	-0.117
Seed Length	0.198	-0.355	0.121	-0.210	-0.055	0.034
Seed width	0.149	-0.318	0.057	-0.242	-0.061	-0.167
Seed number plant	0.243	0.058	-0.372	0.025	0.088	0.037
Canopy width	0.295	0.081	0.157	0.055	-0.168	0.104
Stem numbers	0.215	0.039	-0.142	-0.229	-0.164	0.271
Seed weight	0.239	-0.100	-0.324	-0.14	0.154	0.099

The principal component analysis for the 24 quantitative characters was also conducted among the 34 lines planted in the field (Botswana) experiment. The first six principal components with eigenvalues over one, which accounted for 74.15% of the total variation, were selected to analyse the characters for the selected lines (Table 4.1.8). The first component (PC1) explained 23.64 % of the overall variability among the lines with most of the variation coming from *petiole length, leaf area, petiolule length, shoot dry weight,* and *plant height*. High

loadings for the second component (PC2) which described 17.08 % of the variance was accounted for by *pod weight, seed weight, seed number, pod number* and *pod width.petiole-internode ratio, days to 50% flowering, seed length, leaf number,* and *pod length* contributed most of the 11.98% explained at principal component three (PC3). Principal component 4 separates the lines based mainly on the *seed length, seed width, pod length, pod width* and *leaflet length* with a total variation of 8.68%. Principal component five and six contributed 6.82% and 5.95 % respectively. *Days to 50% flowering, leaflet length, plant height, days to emergence* and *pod length* contributed approximately 7% of total variance to (PC5), while *days to maturity, days to 50% flowering, pod length, peduncle length* and *leaf numbers* accounted for most of the 5.95 % of total variance in (PC 6). Observations made on the field experiment revealed that the vegetative characters contributed most of the PC1 while seed characters significantly contributed in PC2 separating the lines. However, it was the vegetative traits in both experiments that appear as the main characters that can be used to clearly separate the selected genotypes.

Table 4.1.8: Principal component, matrix of eigenvalues and vectors for 24 quantitative characters of bambara groundnut lines planted in Botswana.

	Principal components (PC)					
	PC1	PC2	PC3	PC4	PC5	PC6
Eigenvalues Variance	5.67	4.1	2.88	2.08	1.64	1.43
% Total contribution	23.64	17.08	11.98	8.68	6.82	5.95
%Accumulated	23.64	40.72	52.7	61.38	68.2	74.15
Days to emergence	-0.245	0.156	-0.041	-0.047	0.275	0.123
Days to 50% flowering	0.005	-0.033	0.299	-0.126	0.348	0.430
Days maturity	0.109	-0.211	0.056	0.020	0.038	0.506
Plant height	0.292	-0.020	0.050	0.165	0.293	-0.286
Internode length	0.233	0.066	-0.433	0.017	0.034	0.055
Leaflet length	0.251	0.145	-0.019	0.181	0.329	-0.153
Leaf number	0.089	-0.191	0.233	-0.07	-0.485	0.201
Leaf width	0.281	0.032	-0.022	-0.178	0.201	0.147
Leaf Area	0.326	-0.056	0.176	-0.017	-0.185	0.135
Pod weight	0.040	0.443	0.078	-0.125	0.008	0.153
Peduncle length	0.219	0.135	-0.202	0.178	-0.135	0.215
Petiole length	0.341	-0.095	0.099	-0.065	0.061	-0.313
Petiolule length	0.308	-0.099	0.092	0.098	-0.022	-0.147
Pet-Inter ratio	-0.013	-0.120	0.472	-0.157	-0.03	-0.259
Pod length	-0.046	0.157	0.226	0.371	0.268	0.215
Pod width	-0.186	0.253	0.154	0.241	0.007	-0.013
Pod number per plant	0.111	0.351	0.18	-0.329	-0.089	-0.007
Shoot dry weight	0.306	-0.112	0.211	0.018	0.113	0.105
Seed length	0.082	0.190	0.239	0.442	-0.115	0.016
Seed width	-0.014	0.238	0.098	0.404	-0.336	-0.052
Seeds number per plant	0.120	0.377	0.174	-0.290	-0.074	-0.046
Canopy width	0.285	0.031	-0.244	0.067	-0.191	0.186
Stems numbers	0.111	-0.067	-0.083	0.083	0.096	0.032
Seed weight	0.120	0.389	-0.161	-0.201	-0.053	-0.025

The characters that could potential reveal greater diversity among germplasm and possibly as identified by higher loadings in the PC1 in both agronomy bay and field experiment are *spreading length, shoot dry weight, petiole length, petiolule length* and *leaf area.*

4.3.5 Cluster analysis

Figure 4.2.2 shows a dendrogram for the Euclidean cluster analysis for the 35 bambara groundnut evaluated for a combined analysis for 24 quantitative and 10 qualitative characters in the glass house (UK). The landraces were grouped into three clusters, mainly on their areas of origin since cluster I consist mostly landraces from West Africa, while cluster II consists of landraces mainly from Southern Africa. The landraces from Central Africa were grouped with those from West Africa, while those from East Africa were grouped with those from Southern African together with Ramayana from Indonesia. Three landraces from Southern Africa 90S19-3, 109BOTS1 and 119Hybrid showed some morphological similarities with those in West African. The morphological characters used could not distinguish between 3Acc9NGA and 45A231GHA from Nigeria and Ghana respectively.

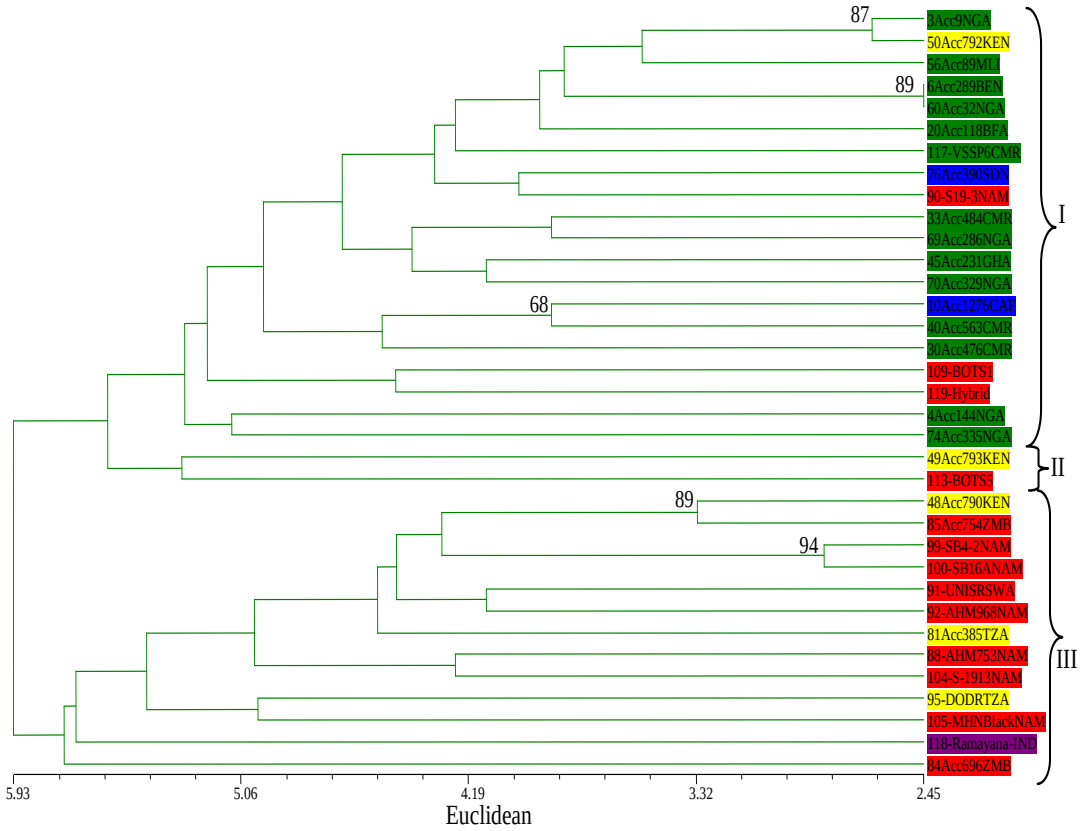


Figure 4.2.2: Dendrogram of 35 bambara groundnut landraces showing a (UPGMA) Euclidean cluster analysis based on 34 agro-morphological markers in glasshouse experiment. The colour code for West Africa = Green, Southern Africa =Red, East Africa =Yellow, Central Africa = Blue, Indonesia = Purple. The number at the nodes of branches represents the percentage bootstrap support of individual nodes resampling at 1000

A dendrogram for cluster analysis performed in the field experiment (Botswana) using 24 quantitative and 13 qualitative characters (37 agro-morphological) on 34 lines of bambara groundnut produced three clusters traits (figure 4.2.3). Cluster I consists mostly of lines from West Africa, but with a mixture of some lines from Southern Africa, like 92-AHM968NAM, 91-UNISRSWA, 99-SB4-2NAM and 104-S-193NAM which also include Ramayana from Indonesia. Cluster II consists of mixture of four lines which are from West Africa, three from Southern Africa and one from Central Africa. Lines 40Acc563CMR and 69Acc286NGA did not produce any pods, but are clustered together with the Southern African lines mainly due to their higher *petiole length*, and *leaf area* and these are some of the traits that had higher loadings in PC 1 (Table 4.1.8).

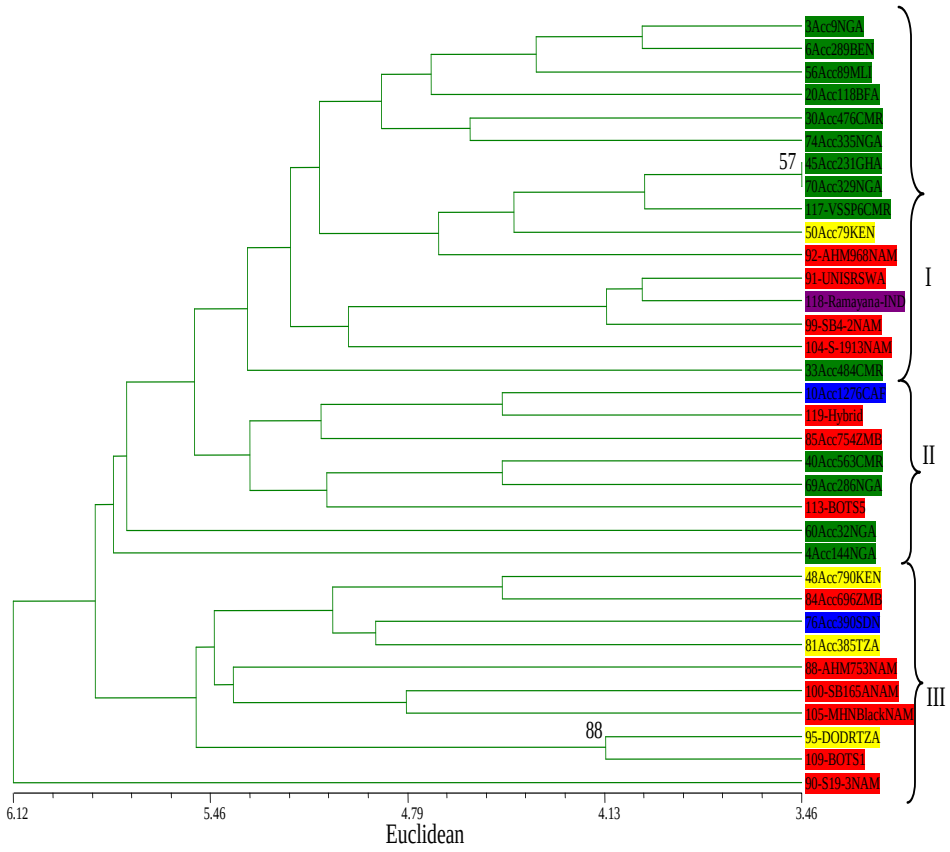


Figure 4.2.3: Dendrogram of 34 bambara groundnut lines showing genetic similarities based on 37 morpho-agronomic traits 24 quantitative traits and 13 qualitative traits, using the UPGMA cluster analysis (field experiment Botswana). The number at the nodes of branches represents the percentage bootstrap support of individual nodes resampling at 1000

Three East African lines were clustered together in cluster III, which also contains six lines from Southern Africa. Line 76Acc390 from Sudan was morphologically similar to the Southern African lines and performed relatively well in terms of *pod number per plant, leaf area* and had larger *leaf area* and *plant height*. The poor adaptation of the West African lines was revealed by lower number of *pods per plant* produced. East African lines that produced comparatively higher number of *pods per plant* like those from Southern Africa were grouped together such as 81-Acc 385 from Tanzania, 95DODR from Tanzania and 48-Acc790 from Kenya. Those lines which originally are from Southern Africa but, produced low *number of pods per plant* like 99-SB4-2 and 92-AHM968 both from Namibia were clustered together with the West African lines in Cluster I. Even though the lines

were not clearly separated based on their areas of origin, individual clusters consist of majority of lines from one region.

A comparison of the two clusters revealed that higher genetic distance estimates were observed in the agronomy bay at 3.38 compared to 2.66 in the field experiment. The tree from the agronomy bay experiment clearly defined landraces according to their areas of origin much more than in the field experiment with 5 bootstrap values above 50% compared to only 2 for the field experiment.

4.3.6 Correlation coefficients among traits

To determine the relationship among the 24 measured characters, Pearson correlations based on the mean of the genotypes were generated (Table 4.1.9 and Table 4.2.1) for the agronomy bay experiment and field experiment, respectively. In the agronomy bay experiment a number of the characters were positively correlated to both *pod number* and *seed number*, while a number of traits such as *days to emergence*, *days to flowering*, *peduncle length*, *leaflet length*, *petiole-internode ratio* and *pod length* showed a negative correlation. This suggest that those landraces which, emerged late ended up having a lower number of leaves and most of those landraces affected were the spreading ones, eventually producing a lower number of seeds. *Seed weight* appears to contribute significantly to both *pod number* and *seed number* at (r = +0.86) and (r = +0.88) respectively, and this implies that indirect selection for *pod number* and *seed number* can be successfully achieved by selecting for the *seed weight* character. There were moderate correlations of *shoot dry weight* and *pod number* and *pod number* to *leaf area* at (r = +0.61 and r = +0.52), respectively, while *leaf area* and *shoot dry weight* were highly correlated to each other at (r = +0.94). *Pod number* and *seed number* were also highly correlated at (r = +0.94). The genotypes that produced large *shoot dry weight* and *leaf area* managed to produce higher *pod number* and *seed number*, as is reflected by the positive correlation between the *shoot dry weight*, *leaf area*, *pod number* and *seed number per plant*. This could be affected by the absence of plant to plant competition.

A similar trend was revealed in the field experiment, many of the characters were positively correlated to *pod number* and *seed number*, except *days to flowering*, *number of stems*, *shoot dry weight*, *leaf numbers* and *days to maturity*, which

suggests that the effect of *days to flowering* and *days to maturity* can be detrimental on *pod number* and *seed number* produced. This is mainly because both the *leaf numbers*, *number of stems* and *shoot dry weight* of the crop is reduced. The late flowering landraces had less time to develop pod/ flowers. *Seed weight* and *pod dry weight* showed a high correlation to *pod number* with (r = +0.84) and (r = +0.83), respectively. A lower correlation was observed between *shoot dry weight* and *pod number* at (r = +0.17) and *leaf area* and *pod number* at (r = +0.25), but there was a moderate relatively high correlation between *shoot dry weight* and *leaf area* at (r = +0.68), while the correlation between *seed number* and *pod number* was a perfect correlation (r = +1.0) implying a fixed number of seeds per pod.

Table 4.1.9: Correlation coefficients among the 24 traits based on the 35 bambara groundnut planted in the Agronomy bay (UK), traits were measured 10 weeks after planting.

Characters	DAE	DAF	LEAFN	SPREADL	LEAFL	LEAFW	LEAFAREA	HEIGHT	INTERNODE	PETIOLE	PETINTERN	PETIOLOULE
DAE	1.00											
DAF	0.39*	1.00										
LEAFN	-0.25	0.04	1.00									
SPREADL	-0.31	-0.17	0.44**	1.00								
LEAFL	0.20	0.21	-0.16	0.17	1.00							
LEAFW	0.05	0.07	0.22	0.46**	0.42*	1.00						
LEAFAREA	-0.17	0.09	0.92**	0.56**	0.13	0.48**	1.00					
HEIGHT	-0.17	0.04	0.49**	0.59**	0.44**	0.50**	0.67**	1.00				
INTERNODE	-0.16	-0.08	0.27	0.80**	0.25	0.43*	0.42*	0.68**	1.00			
PETIOLE	-0.12	0.08	0.53**	0.58**	0.36*	0.52**	0.71**	0.91**	0.63**	1.00		
PETINTERN	0.19	0.09	-0.14	-0.63**	0.14	-0.19	-0.16	-0.11	-0.69**	-0.07	1.00	
PETIOLOULE	0.00	0.13	0.48**	0.15	0.23	0.29	0.57**	0.43*	0.12	0.46**	0.20	1.00
PEDUNCLE	0.04	0.07	-0.22	0.30	0.13	0.05	-0.12	0.09	0.41*	-0.06	-0.48**	-0.26
STEMSNO	-0.23	-0.29	0.49**	0.37*	0.10	0.21	0.50**	0.50**	0.46**	0.37*	-0.20	0.31
DAM	-0.28	0.09	0.23	0.50**	0.14	0.36*	0.29	0.48**	0.53**	0.41*	-0.42*	0.08
SDW	-0.19	0.06	0.84**	0.67**	0.19	0.52**	0.94**	0.70**	0.54**	0.75**	-0.25	0.47**
PODNO	-0.38*	-0.25	0.57**	0.60**	-0.10	0.31	0.52**	0.47**	0.48**	0.46**	-0.31	0.21
PDW	0.07	-0.22	-0.22	0.02	0.13	0.16	-0.19	-0.04	0.07	-0.07	-0.13	-0.09
PODL	0.08	-0.03	-0.08	-0.17	0.04	-0.06	-0.06	0.07	0.07	-0.04	-0.01	0.11
PODW	-0.22	0.08	-0.01	0.25	0.16	0.35*	0.12	0.31	0.34*	0.22	-0.18	0.03
SEEDNO	-0.33	-0.29	0.51**	0.55**	-0.14	0.34*	0.47**	0.42*	0.46**	0.40*	-0.31	0.23
SEEDL	-0.26	-0.21	0.03	0.39*	0.24	0.39*	0.14	0.43**	0.54**	0.28	-0.26	0.08
SEEDW	-0.29	-0.17	-0.09	0.24	0.11	0.27	0.03	0.29	0.27	0.14	-0.09	0.06
SEEDWEIGHT	-0.35*	-0.43**	0.28	0.47**	-0.05	0.32	0.29	0.39*	0.44**	0.31	-0.24	0.22

* Correlation is significant at the 0.05 level; **Correlation is significant at the 0.01 level

Table 4.1.9 (Continued)

Characters	PEDUNCLE	STEMSNO	DAM	SDW	PODNO	PDW	PODL	PODW	SEEDNO	SEEDL	SEEDW	SEEDWEIGHT
PEDUNCLE	1.00											
STEMSNO	0.10	1.00										
DAM	0.19	0.21	1.00									
SDW	-0.02	0.48**	0.45**	1.00								
PODNO	-0.19	0.59**	0.54**	0.61**	1.00							
PDW	0.10	0.06	0.31	-0.06	0.13	1.00						
PODL	0.33	0.08	0.07	-0.05	-0.25	0.26	1.00					
PODW	0.36*	0.08	0.43*	0.31	0.08	0.15	0.46**	1.00				
SEEDNO	-0.21	0.58**	0.51**	0.53**	0.98**	0.13	-0.21	0.05	1.00			
SEEDL	0.49**	0.31	0.43*	0.28	0.21	0.26	0.61**	0.65**	0.21	1.00		
SEEDW	0.32	0.24	0.34*	0.15	0.16	0.10	0.362*	0.64**	0.18	0.71**	1.00	
SEEDWEIGHT	-0.06	0.58**	0.50**	0.40*	0.86**	0.32	0.03	0.27	0.88**	0.45**	0.421*	1.00

* Correlation is significant at the 0.05 level; **Correlation is significant at the 0.01 level

DAE: days to emergence; DAF: days to 50% flowering; LEAFN: Leaf number per plant; SPREADL: Canopy size; LEAFW = Leaflet width HEIGHT =plant height

PETINTERN =Petiole-Internode ratio, STEMSNO =Number of stems per plant DAM=Days to maturity, SDW = shoot dry weight PODNO = pods number per plant

PDW =Pod dry weight SEEDNO = seeds number per plant SEEDW = seed width

Table 4.2.1: Correlation coefficient for 24 quantitative traits of the 34 bambara groundnut planted in the field experiment in (Botswana) traits were measured 10 weeks after planting.

Characters	DAE	DAF	LEAFN	SPRDIN	LEAFW	LEAFL	LEAFAREA	HEIGHT	INTERNODE	PETIOLE	PETINTER	PETIOLULE
DAE	1.00											
DAF	0.12	1.00										
LEAFNO	-0.25	0.01	1.00									
SPRDIN	-0.44*	-0.18	0.17	1.00								
LEAFW	-0.44*	-0.05	-0.01	0.40*	1.00							
LEAFL	-0.31	0.00	-0.18	0.40*	0.62**	1.00						
LEAFAREA	-0.48**	-0.04	0.68**	0.46**	0.62**	0.52**	1.00					
HEIGHT	-0.35*	-0.04	-0.04	0.31	0.60**	0.66**	0.48**	1.00				
INTERNODE	-0.21	-0.34	0.04	0.69**	0.37*	0.37*	0.37*	0.38*	1.00			
PETIOLE	-0.50*	-0.08	0.24	0.45**	0.58**	.054**	0.62**	0.82**	0.35*	1.00		
PETINTER	-0.12	0.25	0.13	-0.35*	0.00	-0.10	0.02	0.08	-0.72**	0.30	1.00	
PETIOLULE	-0.42*	0.02	0.25	0.34	0.24	0.35*	0.43*	0.55**	0.29	0.64**	0.17	1.00
PEDUNCLE	-0.05	-0.24	0.27	0.47**	0.23	0.16	0.424*	0.34*	0.53**	0.25	-0.43*	0.33
STEMSNO	-0.49**	0.00	-0.02	0.25	0.28	0.31	0.22	0.20	0.18	0.30	-0.03	0.15
DAM	-0.25	0.30	0.19	0.28	0.34	-0.11	0.24	0.08	0.02	0.13	0.09	0.19
SDW	-0.34*	0.10	0.52**	0.30	0.29	0.24	0.65**	0.43*	0.28	0.56**	0.10	0.53**
PODNO	0.03	0.16	-0.02	0.03	0.28	0.16	0.21	0.01	0.00	0.14	0.13	0.06
PDW	0.21	0.14	-0.23	0.11	0.16	0.21	0.01	-0.04	0.12	-0.12	-0.10	-0.10
PODL	0.19	0.32	-0.16	-0.12	-0.12	0.23	-0.08	0.00	-0.24	-0.18	0.04	-0.04
PODW	0.37	-0.05	-0.21	-0.30	-0.40*	-0.03	-0.34	-0.15	-.38*	-0.42*	0.08	-0.35
SEEDSNO	-0.03	0.10	-0.07	0.08	0.25	0.22	0.19	0.05	0.04	0.18	0.14	0.06
SEEDL	-0.05	0.03	0.10	0.04	0.04	0.31	0.30	0.23	-0.07	0.07	0.05	0.13
SEEDW	-0.06	-0.27	-0.09	0.07	-0.17	0.03	-0.05	0.01	-0.11	-0.16	-0.06	0.04
SEEDWEIGHT	0.06	-0.17	-0.27	0.389*	0.28	0.33	0.07	0.05	0.43*	0.08	-0.30	-0.02

* Correlation is significant at the 0.05 level; **Correlation is significant at the 0.01 level

Table 4.2.1: (Continued)

Characters	PEDUNCLE	STEMSNO	DAM	SDW	PODNO	PDW	PODL	PODW	SEEDSNO	SEEDL	SEEDW	SEEDWEIGHT
PEDUNCLE	1.00											
STEMSNO	-0.03	1.00										
DAM	0.07	0.14	1.00									
SDW	0.34	0.28	0.37	1.00								
PODNO	0.15	-0.02	-0.25	0.13	1.00							
PDW	0.20	-0.11	-0.18	-0.05	0.74**	1.00						
PODL	0.01	0.05	-0.06	0.07	0.01	0.27	1.00					
PODW	-0.05	-0.16	-0.20	-0.24	0.08	0.46*	0.41*	1.00				
SEEDSNO	0.14	-0.02	-0.25	0.16	0.96**	0.80**	0.09	0.18	1.00			
SEEDL	0.18	0.01	-0.11	0.14	0.19	0.25	0.52**	0.27	0.20	1.00		
SEEDW	0.24	-0.16	-0.08	-0.13	0.20	0.33	0.26	0.44*	0.21	0.68**	1.00	
SEEDWEIGHT	0.35	-0.01	-0.29	-0.14	0.63**	0.75**	-0.04	0.14	0.71**	0.07	0.14	1.00

* Correlation is significant at the 0.05 level; **Correlation is significant at the 0.01 level

DAE: days to emergence; DAF: days to 50% flowering; LEAFN: Leaf number per plant; SPREDL: Canopy size; LEAFW = Leaflet width HEIGHT =plant height PETINTERN =Petiole-Internode ratio STEMSNO =Number of stems per plant DAM=Days to maturity, SDW = shoot dry weight PODNO = pods number per plant PDW =pod dry weight SEEDNO = seeds number per plant SEEDW = seed width

4.3.7 Quantitative variance analysis

Bambara groundnut improvement is not only dependent on the magnitude of phenotypic variation of the crop, but also on the extent of how the traits are heritable. Therefore it is important to quantify the heritable and non-heritable component from the phenotypic variation observed. Assessing the genotypic coefficient of variation (GCV), heritability and genetic advance (as percentage of the mean) at the same time gives a good estimation of the amount of advance expected in selection (Baye, 2002). Since one objective of this study is to identify the best performing lines, it is important the genotypic coefficient of variation (GCV), phenotypic coefficient of variation (PCV), heritability, and genetic advance (as percentage of the mean) are estimated in the selected genotypes.

The estimates of phenotypic coefficient of variability (PCV) and genotypic coefficient of variability (GCV), heritability (in a broad sense), and genetic advance as a percentage of the mean were analysed (Table 4.2.2) for the glasshouse experiment (Table 4.2.3) for the field experiment in Botswana. The highest *phenotypic coefficient of variation* (PCV) in the glasshouse experiment was observed in the number of *pods per plant* at 51.73% and the lowest was on the *days to maturity* at 1.13%. In the field experiment, the range of the *phenotypic coefficient of variation* ranged from a high for *number of pods per plant* at 82.17% to a low for *days to maturity* at 5.6%. The *phenotypic coefficient of variation* was relatively high in the agronomy bay experiment for *shoot dry weight* at 55.1%, *leaf area* at 54.4% compared to 45.2% for *shoot dry weight* in the field and 31.5% for *leaf area* in the field. But for the seed numbers *phenotypic coefficient of variation* was much higher in the field at 79.0% compared with 51.5% in the agronomy bay experiment.

Table 4.2.2: Quantitative variances based on phenotypic coefficient of variability (PCV), genotypic coefficient of variability (GCV), broad sense heritability (h^2B) and genetic advance (GA) in the 35 landraces in the agronomy bay (UK).

Traits	MSG	MSE	PCV%	GCV%	h^2B	GA %of mean
Days to emergence	4.18	2.57	7.23	11.65	0.39	10.3
Days to 50% flowering	33.79	8.93	6.78	7.91	0.74	8.7
Number leaves per plant	4758	826.2	45.59	50.16	0.83	50.3
Canopy spread	181.02	33.8	35.85	39.75	0.81	40.6
Leaflet length	1.28	0.311	6.61	7.60	0.76	8.2
leaflet width	0.77	0.14	12.56	13.88	0.82	14.1
Leaf area	17918884	4836578	46.52	54.44	0.73	60.2
Plant height	56.41	7.79	12.02	12.95	0.86	12.1
Internode length	1.77	0.193	31.52	33.40	0.89	28.8
Petiole length	31.66	3.23	18.33	19.35	0.90	16.1
Petiole-Internode ratio	13.98	3.16	23.71	26.95	0.77	28.9
Petiolule length	1.028	0.167	23.50	25.67	0.84	25.0
Peduncle length	0.742	0.321	16.08	21.34	0.57	23.2
Number of stems	17.8	9.45	16.90	24.68	0.47	24.6
Days to maturity	94.48	85.24	1.13	3.61	0.10	1.0
Shoot dry weight	1011.1	197	49.47	55.13	0.81	57.1
Number pod per plant	2960.5	584.9	51.73	57.75	0.80	60.0
Pod dry weight	778.7	215.4	41.90	49.27	0.72	54.6
Pod length	8.6	3.35	7.03	9.00	0.61	10.0
Pod width	4.39	1.24	8.49	10.02	0.72	11.2
Number of seeds per plant	2815.3	508.6	46.92	51.83	0.82	52.6
Seed weight	449.8	110.8	46.0	53.0	0.75	57.8
Seed length	2.94	1.34	6.84	9.27	0.54	10.0
Seed width	1.49	0.72	6.05	8.42	0.52	8.8

Mean square genotype (MSG): estimates genotypic variance, this value is observed variance among the line means, while mean square error (MSE) measures variance from plot residuals.

Mean Square Error (MSE) are variance components estimated as functions of the means square estimates from ANOVA table.

Genetic advance are estimates percentage based on the mean

Table 4.2.3: Quantitative variances based on phenotypic coefficient of variability (PCV), genotypic coefficient of variability (GCV), broad sense heritability (h^2B) and genetic advance (GA) in the 34 lines (Field experiment).

Traits	MSG	MSE	PCV%	GCV%	h^2B	GA %of mean
Days to emergence	7.36	3.65	10.36	7.36	0.50	11.41
Days to 50% flowering	137.15	10.24	11.89	11.44	0.93	9.32
Number leaves per plant	3045.1	552.2	28.91	26.16	0.82	31.26
Canopy spread	19641	1335	33.74	32.57	0.93	25.42
Leaflet length	234.57	57.7	13.19	11.46	0.75	15.27
leaflet width	35.74	6.40	13.81	12.51	0.82	14.87
Leaf area	3240130	972383	26.31	31.45	0.70	37.84
Plant height	3133.3	652.5	11.62	10.34	0.79	13.10
Internode length	208.74	53.23	37.22	32.12	0.74	43.35
Petiole length	1504.4	316.4	16.09	14.30	0.79	18.06
Petiole-Internode ratio	9.8	1.83	25.08	22.62	0.81	27.34
Petiolule length	54.49	13.32	23.95	20.82	0.76	27.75
Peduncle length	86.19	30.83	27.08	21.70	0.64	32.25
Number of stems	11.8	3.66	23.58	19.58	0.69	28.21
Days to maturity	176.39	9.57	5.60	5.45	0.95	3.63
Shoot dry weight	1067.6	117.5	46.38	43.75	0.89	42.46
Number pod per plant	511.4	237.6	82.17	60.12	0.54	88.15
Pod dry weight	97.89	69.97	68.65	36.66	0.29	48.46
Pod length	25.98	4.57	16.06	14.58	0.82	16.32
Pod width	3.94	1.75	10.84	8.08	0.56	11.82
Number of seeds per plant	453.3	225	79.49	56.41	0.50	82.86
Seed weight	43.49	26.16	72.52	45.78	0.40	66.01
Seed length	4.62	2.07	11.36	8.44	0.55	12.34
Seed width	2.22	1.16	10.61	7.33	0.48	10.68

Mean square genotype (MSG): estimates genotypic variance, this value is observed variance among the line means, while mean square error (MSE) measures variance from plot residuals.

Mean Square Error (MSE) are variance components estimated as functions of the means square estimates from ANOVA table.

Genetic advance are estimates percentage based on the mean

The *genotypic coefficient of variation* in the agronomy bay experiment ranged from number of *pods per plant* from 57.75% to 3.61% for *days to maturity*, while in the field experiment the same traits revealed lowest *genotypic coefficient of variation for days to maturity* (5.45%) andhighest *genotypic coefficient of variation*60.12%. High *genotypic coefficient of variation* values as shown by *pods per plant*, *number of seeds per plant*, *seed weight* in both sites (UK and in

Botswana) indicates that these traits could lead to good progress in crop improvement. Lower *genotypic coefficient of variation* shown by, *days to flowering*, *days to maturity*, *seed width* that are lower in both sites indicate that these traits are less amenable to improvement through selection. Relatively low values of both *phenotypic coefficient of variation* and *genotypic coefficient of variation* were in agronomy bay experiment than in the field experiment.

To quantify the amount that is heritable for the 24 characters, broad sense heritability was estimated in (Table 4.2.2 and Table 4.2.3). In the agronomy bay experiment estimates of broad sense heritability ranged from 0.9 for *petiole length* to 0.1 for *days to maturity*, most of the characters have a heritability of more than 0.7. Whereas, in the field experiment the estimates of broad sense heritability was highest for *days to maturity* at 0.95 and lower for *days to pod dry weight* 0.29, with most of the characters showing more than 0.7 heritability. There was higher broad sense heritability in agronomy bay experiment for *pod number per plant* and *seed number per plant* 0.8 and 0.82 compared to 0.54 and 0.50 respectively for both characters in the field experiment. While the *shoot dry weight* increased from 0.81 in the agronomy bay to 0.89 in the field experiment and the *leaf area* was the same at 0.7 for both sites. The higher estimates of heritability in agronomy bay experiment and field experiment for *number of leaves*, *plant spread*, *leaflet width*, and *plant height* shows that these traits may not have been affected by environment.

Genetic advance (GA) expected when selecting at 5% based on the percentage of the mean reached a maximum of 60.2% for internode length and a minimum of 1% for *days to maturity* in the agronomy bay experiment, while for the field experiment it ranged from 88.2% for *number of pods per plant* to 3.6 % for *days to emergence.* Other characters that showed relatively high estimates of *genetic advance*were *number of seeds per plant* (54.6%), *number of leaves per plant* (57.1%), *petiolule* (57.8 %) and *plant height* (60%) respectively, in the agronomy bay. The characters that showed higher genetic advance in the field experiment were *internode length* (43.3%), *pod dry weight* (48.5%), *seed weight* (66%) and *number of seeds per plant* (82.9%). Lower *genetic advance* recorded in the field were for *days to 50% flowering* (9.3%), *seed width* (10.7%), *days to emergence*

(11.4%) *and pod width* (11.8%), respectively, which suggest that these traits could be difficult to select for in semi-arid Botswana environment.

A number of characters such as *shoot dry weight*, *number of pods per plant*, *number of leaves per plant*, and *plant spread* have consistently shown higher *phenotypic coefficients of variation, genotypic coefficients of variation* and estimates of broad sense heritability in both experimental sites, which suggest that these characters could be useful for selection in bambara groundnut. In this study, *pod number*, *seed number*, *leaf area* and *shoot dry weight* (biomass) were selected as the basis for multiple selection in bambara groundnut.

4.3.8 Comparison of agronomy bay and field experiment

Bambara groundnut evaluation and characterisation was conducted at two experimental sites one in UK and one in Botswana, with different environmental conditions. The performances of the genotypes in the two sites were compared especially since the UK materials are used as the initial selection for breeding purpose.

A regression analysis was conducted between the agronomy bay and field experiment, a linear relation between the two experiments is shown on figure 4.2.4. The relationship between the two experiments is highly significant ($P < 0.001$), which suggest that, the agronomy bay experiment could be useful in the selecting of materials to plant in the field (Table 4.2.4).

Table 4.2.4: A summary of analysis for the relationship between the agronomy (UK) experiment and the field experiment in (Botswana), computed on Genstat version 13.0

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	1	58646	58646.1	209.81	<.001
Residual	22	6150	279.5		
Total	23	64796	2817.2		

Percentage variance accounted for 90.1
Standard error of observations is estimated to be 16.7.

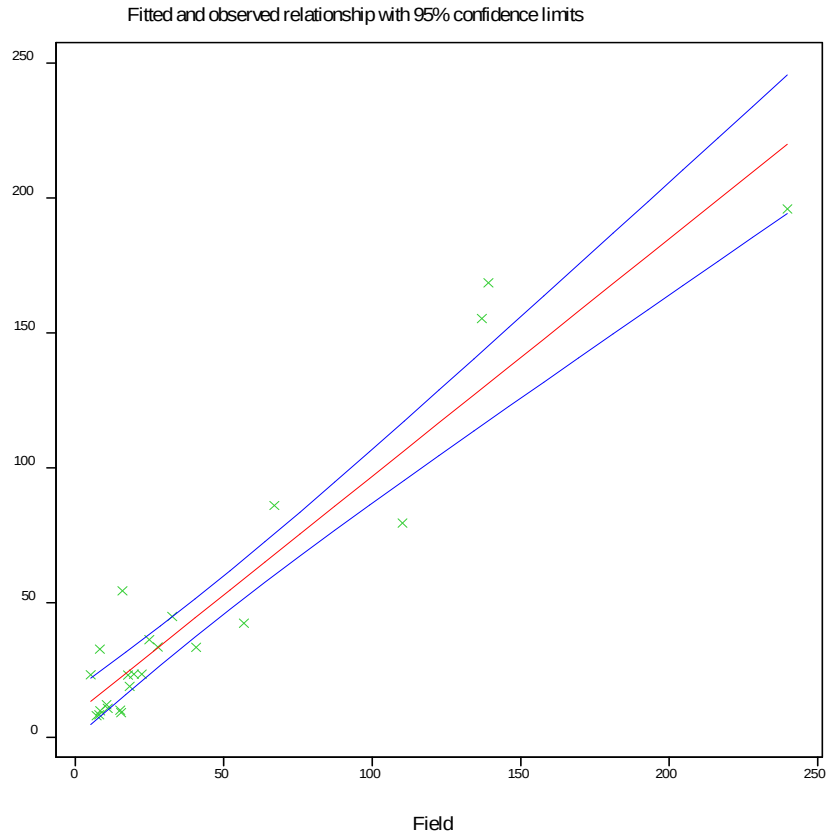


Figure 4.2.4: shows a regression analysis plot of mean over all the genotypes for the 24 variables recorded from agronomy bay (UK) and field experiment in Botswana

The data from the two sites for the 24 characters, show that they are well correlated (Figure 4.2.4).

4.3.8 Selection for breeding bambara groundnut

To identify lines which have a potential to produce higher yields in a Botswanan environment based on four selected characters, the selection index was used with a weighting of the genetic advance found in the field experiment: A linear equation for SI = (X_1 x 0.378) + (X_2 x 0.424) + (X_3+0.828) + (X_4 x 0.881) was derived using the genetic advance from the fields study (Table 4.2.6). This index put more emphasis on yield in a multiple trait selection. X_1 = *Leaf area*, X_2 =*Shoot dry weight*, X_3 =*Seed number per plant*, X_4 = *Pod number plant*.

Selection index (SI) analysis produced single values for each genotype. These values were then ranked for the agronomy bay and field experiment respectively. The Duncan multiple range test (DMRT) was also performed to identify genotypes with different characters. Table 4.2.8 and Table 4.2.9 show the ranking of the genotypes using the selection index (SI) and (DMRT).

Table 4.2.5: The Duncan multiple range tests and the selection index of bambara groundnut based on the vegetative and yield characters (Agronomy bay, UK).

Landraces	Leaf Area	DMRT	SDW	DMRT	PODS	DMRT	SEEDS	DMRT	**SI**	RANK
3Acc9NGA	2956	Abcd	17.2	Ab	36	Abcd	54	abcdefg	-1.33	25
4Acc144NGA	2953	Abcd	30.2	Abcd	46	Abcdef	50	abcdef	-0.8	18
6Acc289BEN	2587	Abcd	10.8	A	17	A	25	Abc	-2.47	34
10Acc1276CAF	4316	Abcd	35.8	Abcd	46	Abcdef	52	abcdef	-0.25	15
20Acc118BFA	2512	Abcd	17.2	Ab	34	Abcd	43	abcde	-1.66	26
30Acc476CMR	3391	Abcd	24.2	Abc	54	Abcdef	60	bcdefgh	-0.56	16
33Acc484CMR	3470	Abcd	23.4	Abc	50	Abcdef	56	abcdefg	-0.71	17
40Acc563CMR	4896	Abcd	33.5	Abcd	20	Ab	25	Abc	-1.13	23
45Acc231GHA	3318	Abcd	35.5	Abcd	40	Abcde	33	Abcd	-0.93	20
48Acc790KEN	10369	F	68.3	Fgh	106	Gh	105	Ij	4.3	**1**
49Acc793KEN	2281	Abc	20.1	Abc	52	Abcdef	55	abcdefg	-1.08	22
50Acc792ZWE	1887	Ab	14.6	Ab	33	Abcd	38	Abcd	-1.99	30
56Acc89MLI	1357	A	9.9	A	28	Ab	40	Abcd	-2.33	33
60Acc32NGA	2613	Abcd	20.2	Abc	22	Ab	30	Abc	-2	31
69Acc286NGA	2953	Abcd	13.2	A	10	A	15	Ab	-2.62	35
70Acc329NGA	4049	Abcd	28.9	Abcd	11	A	12	A	-1.89	28
74Acc335NGA	2944	Abcd	22.6	Abc	20	Ab	26	Abc	-1.93	29
76Acc390SDN	1955	Ab	11.2	A	30	Ab	39	Abcd	-2.13	32
81Acc385TZA	6713	Cdef	42.7	Bcdef	86	Efgh	102	Hij	2.19	**8**
84Acc696ZMB	9939	Ef	54.1	Defg	35	Abcd	50	abcdef	1.45	**10**
85Acc754ZMB	9997	Ef	76.3	Gh	106	Gh	94	Fghij	4.27	**2**
88-AHM753NAM	4347	Abcd	33.7	Abcd	121	H	133	J	2.53	**5**
90-S19-3NAM	3364	Abcd	21.5	Abc	48	Abcdef	48	abcde	-0.95	21
91-UNISRSWA	6059	Bcde	53.2	Defg	105	Gh	107	Ij	2.79	**3**
92-AHM968NAM	4103	Abcd	35.7	Abcd	87	Efgh	96	Ghij	1.24	12
95-DODRTZA	3822	Abcd	34.2	Abcd	79	Cdefgh	86	efghi	0.79	14
99-SB4-2NAM	6500	Cdef	46.8	Cdef	100	Gh	105	Ij	2.58	**4**
100-SB16ANAM	6429	Cdef	48.1	Cdef	67	Bcdefg	64	cdefghi	1.26	11
104-S-1913NAM	4240	Abcd	36.8	Abcde	92	Fgh	103	Ij	1.52	**9**
105-MahenèneBlackN	6794	Def	62.5	Efgh	81	Defgh	76	defghi	2.29	**7**
109-BOTS1	2393	Abcd	20.7	Abc	47	Abcdef	53	abcdef	-1.15	24
113-BOTS5	4557	Abcd	31.7	Abcd	84	Efgh	77	defghi	0.85	13
117-VSSP6CMR	2704	Abcd	20.5	Abc	31	Abc	30	Abc	-1.77	27
118-Ramayana-IND	9583	Ef	80.5	H	48	Abcdef	50	abcdef	2.45	**6**
119-Hybrid	4751	Abcd	29.7	Abcd	31	Abc	40	Abcd	-0.82	19

Means with similar letters in column are not significantly different at 5% Duncan Multiple Range test.

DMRT (Duncan Multiple Range Test)

Table 4.2.6: The Duncan multiple range tests and the selection index of bambara groundnut based on the vegetative and yieldcharacters (field experiment in Botswana).

Lines	Leaf Area	DMRT	Shoot	DMRT	Seeds No	DMRT	Pod No	DMRT	SX	RANK
3Acc9NGA	4329	Fghi	35.7	Bcdefghi	7	A	7	Ab	-0.91	24
4Acc144NGA	2878	Bcdefg	31	Abcdefghi	9	A	8	Ab	-1.31	29
6Acc289BEN	3391	Bcdefg	28.6	Abcdefgh	7	A	9	Ab	-1.24	27
10Acc1276CAF	2157	Abcde	44.8	Efghij	8	A	8	Ab	-1.32	30
20Acc118BFA	3988	Defgh	21.5	Abc	12	A	11	Ab	-0.72	20
30Acc476CMR	2861	Bcdefg	37.9	Cdefghi	14	A	13	Ab	-0.49	17
33Acc484CMR	2811	Bcdefg	27.1	Abcdefg	22	A	23	Ab	**0.47**	**9**
40Acc563CMR	3614	Defgh	45.4	Efghij	*	A	*	Ab	0.21	13
45Acc231GHA	1446	Ab	18.7	Abc	*	A	*	Ab	-1.08	26
48Acc790KEN	4586	Ghi	61.4	Jk	11	A	11	Ab	0.25	12
50Acc792ZWE	3059	Bcdefg	48.2	Ghij	*	A	*	Ab	0.09	14
56Acc89MLI	1614	Abc	21.9	Abc	6	A	6	Ab	-2.26	34
60Acc32NGA	2583	Bcdef	38.6	Cdefghi	13	A	12	Ab	-0.69	19
69Acc286NGA	4319	Fghi	16	Ab	*	A	*	Ab	-0.17	15
70Acc329NGA	387	A	13.7	A	*	A	*	Ab	-1.54	33
74Acc335NGA	3786	Defgh	23.4	Abcd	10	A	11	Ab	-0.9	23
76Acc390SDN	5986	I	82.9	L	12	A	10	Ab	**1.19**	**6**
81Acc385TZA	5437	Hi	102.9	M	8	A	10	Ab	**1.17**	**7**
84Acc696ZMB	4343	Fghi	50.8	Ij	33	A	36	B	**3.12**	**2**
85Acc754ZMB	4098	Efgh	72.3	Kl	20	A	17	Ab	**1.33**	**4**
88-AHM753NAM	3879	Defgh	47.2	Fghij	31	A	35	Ab	**2.68**	**3**
90-S19-3NAM	4019	Defgh	48.3	Ghij	66	B	68	C	**7.33**	**1**
91-UNISRSWA	2442	Bcdef	36.6	Bcdefghi	9	A	9	Ab	-1.26	28
92-AHM968NAM	2062	Abcd	24.7	Abcde	18	A	21	Ab	-0.23	16
95-DODRTZA	3152	Bcdefg	38	Cdefghi	5	A	5	A	-1.53	32
99-SB4-2NAM	3256	Bcdefg	29.4	Abcdefghi	21	A	21	Ab	**0.46**	**10**
100-SB16ANAM	3428	Cdefg	45.1	Efghij	18	A	19	Ab	**0.53**	**8**
104-S-1913NAM	2715	Bcdefg	21	Abc	22	A	23	Ab	0.31	11
105-MaheneneBlack	3306	Bcdefg	44.3	Defghij	9	A	9	Ab	-0.81	22
109-BOTS1	3982	Defgh	69.9	Kl	18	A	19	Ab	**1.25**	**5**
113-BOTS5	3865	Defgh	49.8	Hij	5	A	5	Ab	-1.04	25
117-VSSP6CMR	2252	Bcde	26.6	Abcdef	*	A	*	Ab	-0.63	18
118-Ramayana-IND	2609	Bcdefg	31.8	Abcdefghi	7	A	8	Ab	-1.51	31
119-Hybrid	2383	Bcdef	46.3	Fghij	12	A	11	Ab	-0.75	21

Means with similar letters in a column are not significantly different at (5%) Duncan Multiple Range test.

DMRT (Duncan Multiple Range Test)

The selection index (SI) as revealed by the ranking was able to identify the best performing landraces in the agronomy bay experiment as, (1) 48-Acc790 from Kenya, (2) 85-Acc754, (3) 91-UNISR from Swaziland, (4) 99-SB4-2 from Namibia, (5) 88-AHM753 from Namibia, (6) 118-Ramayana from Indonesia, (7)

105-Mahenene black, (8) 81-Acc385 from Tanzania, (9) 104-S-1913 from Namibia and (10) 84Acc696 from Zambia.

In the field experiment (Botswana), (1) 90-S19-3 from Namibia, (2) 84-Acc 696 from Zambia, (3) 88-AHM753, (4) 85-Acc754 from Zambia, (5) 109-BOTS1 from Botswana, (6) 76-Acc390 from Sudan, (7) 81-Acc385 from Tanzania, (8) 100SB16A from Namibia, (9) Acc3348 from Cameroon, and (10) 99-SB4-2 from Namibia. Interestingly genotypes from Namibia, from a drier environment performed well in the agronomy bay environment and similarly produced more yield in the Botswanan environment.

At 5 % selection using Duncan multiple range test, in the agronomy bay experiment no particular landrace stood outas revealed by the selection index, on all traits. In the field experiment it was line 90-S19-3 from Namibia which surpassed the rest on number of *pods per plant* and number of *seeds per plant*.

4.4 Discussion

Qualitative characters observed

Qualitative characters showed a substantial amount of variability in *growth habit*, *leaflet shape* and in pod and seed characters. *Pod texture*, *pod shape*, *pod colour*, *testa colour* and *test pattern* showed considerable variation while *eye pattern* and *stress susceptibility* had low variation. The predominant *testa colour*wasthe reddish colour, which implies that these are the ones farmers have selected for. All the three plant growth habits were identified among the selected landraces, which reflect the cropping system of bambara groundnut by farmers. The spreading type landraces are useful in mixed cropping with other crops such as cereals while the semi bunch and the bunch are good for monoculture.

Descriptive characters observed

The phenotypic ranges of various characters found in this study are in line with findings by other researchers when evaluating bambara groundnut (Goli *et al*., (1995); Karikari and Tabona, (2004); Ntundu *et al*., (2006); Ouedraogo *et al*.,

(2008). Generally, characters which showed a greater range of difference were observed for *number of pods per plant*, *seed weight*, and *seed number*. The broad range of differences among these traits shows that there is a good possibility for crop improvement.

Knowledge on the genetic variation, heritability and correlation between bambara groundnut agro-morphological characters is important to initiate a feasible breeding program. In this study considerable variation in all the characters was observed, except for *pod dry weight* and *days to maturity*. Adeniji *et al*., (2008) found significant differences for 11 characters among the 18 traits they measured in their study for 10 bambara groundnut accessions to evaluate the inter-relationship for pod and seed yield characters sourced from North-Eastern Nigeria. Similarly, Ntundu *et al.,* (2006) observed highly significant morphological variation for 13 characters on 27 of the characters they measured for the two seasons in Tanzania when they were studying the morphological diversity among 100 Tanzania bambara groundnut accessions.

The application of principal component analysis

Principal component analysis was performed to further find out which characters are important in explaining the variation among the selected genotypes. In the agronomy bay experiment the first two principal components (PC 1 and PC 2) separate mainly on the basis of vegetative traits such as the *spreading length, shoot dry weight*, *petiole length*, *petiolule length*, *leaf area* and *pod number per plant*. *Shoot dry weight* and *petiole length* had higher loadings in both (PC1 and PC2). In the field experiment overall variability among the lines came primarily from vegetative traits in PC 1 such as *petiole length, leaf area, petiolule length*, *shoot dry weight*, *plant height* and *spreading length*, while PC 2 higher loadings were mainly pod and seed characters *pod dry weight*, *seed weight*, *seed number per plant*, *pod number per plant*, and *pod width.* This indicates the importance of these characters in identifying bambara groundnut landraces. Ntundu *et al.*, (2006) observed similar patterns of loading in their study on 100 landraces in Tanzania, whereby the high loading within principal component one was mainly due to vegetative characters while the second was mainly seed characters.

Clusters analysis

Cluster analysis revealed that bambara groundnut is phenotypically distinct across the regions of Africa and can easily be separated, based on the important characters, in this case *pod number per plant* and *shoot dry weight* differences. The degree of genetic differentiation in a crop is dependent on its breeding systems, life history and geographical distribution (Roy, 2000). In bambara groundnut the influence of rainfall and its distribution pattern on the productivity was reported (Azam-Ali *et al.,* 2001), while Collinson *et al.,* (1999) identified moisture as one important factor in the biomass production on bambara groundnut. Sesay *etal*., (2008) observed that in the sub-tropical regions of Africa sowing date had an impact on the final yield of bambara groundnut mainly due to the effects of temperature and varying day-length.

The formation of clear clusters in bambara groundnut (figure 4.2.2 and figure 4.2.3) was expected, as bambara groundnut is an in-breeder it should reveal greater inter-population diversity and less intra-population. Cui *et al.*, (2001) used phenotypic traits and clearly separated Chinese and North American soybean cultivars, while Upadhyaya, (2003) used 16 morphological and 32 agronomic traits in groundnut to differentiate subspecies, *fastiga* (var. *fastiga,vulgaris*, *aequatorian*, *peruviana*) and subsp. *Hypogaea* (var. *hypogaea*, *hirsuta*).

However, there were some overlaps noticed from various regions, especially between Southern Africa and East African lines, as revealed in both studies in the agronomy bay and field experiment. Confirming the importance of seed sources in bambara groundnut (Brink *et al.*, 1996; Massawe *et al*., 2005)

Correlations coefficient

Correlations of greater than approximately (r = +0.7) or less than (r = -0.7) are the ones likely to be of biological importance (Hill *et al.*, 1998). Such correlations were found between, *petiole length* and *plant height* (r = +0.91), *shoot dry weight* and *plant height* (r = +0.70), *leaf area* and *petiole length* (r = +0.71) for the agronomy bay experiment. While for the field experiment correlation coefficients of r = +0.7 or greater were recorded between; *pod length* and *days to maturity* (r = 0.93), *leaf area* and *leaf number* (r = +0.71) and plant *height* and *petiole length* (r = +0.82).

Yield related traits recorded higher positive correlations; *seed weight* and *pod dry weight* were highly correlated to *pod number* with r = +0.84 and r = +0.83 respectively, while *seed number per plant* was highly correlated with *pod number* at r =+0.98 in the agronomy bay and r = +1.0 in the field experiment in Botswana. This could be possibly due to sick plant producing few pods which are shrivelled, and of low weight. Similar findings were found in bambara groundnut by Ofori (1996), Karikari (2000), Karikari and Tabona (2004), Jonah *et al.,* (2010) and Onwubiko *et al*., (2011), which suggest that these characters can be selected for in order to improve bambara groundnut yield.

The lower correlations of *leaf area*,*shoot dry weight* to both *seed number* and *pod number* reveals that most of the lines which are not adapted to the Botswana environment concentrated their growth into vegetative growth rather than in pod and seed production, some high biomass producing lines are not necessarily those that produced more seeds and pods. For example lines 76-Acc390 from Sudan, 81-Acc385 from Tanzania, 48-Acc790 from Kenya, 84-Acc696 from Zambia, 3-Acc9, 69-Acc286 from Nigeria and 85-Acc754 from Zambia produced high *leaf area* and *shoot dry weight* (biomass) more so than 90-S9-3 from Namibia which produced the highest number of pods and seeds. However, this still reflects the importance of *leaf area* and *shoots biomass* production in the final yield of bambara groundnut and the different adaptation of landraces to their original climatic environment (Mwale*et al.,* 2007).

Phenotypic coefficient of variation and genotypic coefficient of variation

Among the characters assessed both in the agronomy bay experiment and in the field experiment in Botswana, the number of *pods per plant* showed a high *phenotypic coefficient of variation* and *genotypic coefficient of variation*, while broad sense heritability was relatively low. The *phenotypic coefficient of variation* for *pod number per plant* ranged from 51.7% in the agronomy bay experiment to 82.1% in the field experiment, while the *genotypic coefficient of variation* ranged from 57.8 % from the agronomy bay experiment to 60.1% in the field experiment. Wigglesworth (1996) found a *phenotypic coefficient of variation* of 20.2 %, a *genotypic coefficient of variation* of 12.6 % and a relatively lower heritability of 0.39 in *pod number* and lower *phenotypic coefficient of variability* and *genotypic*

coefficient of variability of *100 seed weight* at 18 % and 17.7% , but higher heritability of 0.94 when he assessed 10 landraces, six from Botswana, one from Zimbabwe, and three as checks from West Africa (Mali, The Gambia , Niger) for their potential under irrigation using sewage water systems in Botswana environment. In contrast, Karikari, (2000) reported heritability of bambara *grain yield* at 0.71, *100 seed weight* at 0.25, and *shoot dry weight*at0.36 when assessing the adaptability of local and exotic landraces in a Botswana environment. Their results suggest that selecting for *100 seed weight* in Botswana would lead to a yield increase as compared to selecting for lower heritable traits like *seed weight, shoot dry weight* and *pod number per plant*.

In Nigeria, Jonah *et al.*,(2010) reported heritability for *pod yield per plant* as 0.69, *seed yield per plant* 0.72, for *100 seed weight* 0.92 and the genetic mean advance was relatively low at 16.2% for *pod number* and 22.5% for *seed number*, which suggests that these traits will be difficult to select for in the Nigerian environment. Hill *et al.,* (1998) argued that even though heritability is an important tool in the selection of potential material, the estimates may also depend on the environment upon which materials are under test, and can sometimes differ within the same crops. In this study, number of *pods per plant* and *number of seed* revealed moderate genetic advance of the mean at 60% and 52.6% in the agronomy bay experiment and different values to those in the field experiment, at 88.2% for *pod number per plant* and 42.5% for *seed number per plant* (Table 4.2.2 and Table 4.2.3). This is an indication that these traits are under additive genetic control, thus selection of these traits can lead to an improvement in bambara groundnut.

Selections of the best lines

According to the selection index ranks obtained as shown in (Table 4.2.5 and Table 4.2.6), genotype 90-S19-3 from Namibia was the highest ranked with an index value of 7.33 followed by 84-Acc696 from Zambia with an index value of 3.12. Genotype 109-BWA1 from Botswana was used as a check as it was one of the lines among the five selected lines for stable seed colour, uniform leaf morphology, and yield stability in various regions of Botswana for more than five seasons (Chui *et al.,* 2003), but it recorded an index value of 1.25 at the fifth

position. This shows that among the selected lines, at least four more could be potentially adopted for selection and breeding in Botswana environment. The yield production of 18 g per plant for 90-S19-3, 13.5 g for 84-Acc696 compares favourably well with those found by other researchers, Berchie *et al.*, (2010) in Ghana recorded yield per plant for landrace Zebra (23.6 g), and Burkina at (17.7 g).

4.5 Conclusions

There was a substantial amount of variation found in the selected material, and this indicates that there is a great potential for crop improvement. This was also shown by higher genotypic coefficient of variation, heritability and genetic advance (5 % of the mean) in seed and pod characters. GCV, h^2b, GA for *number ofpods per plant* (82.9%; 60.12%; and 88%) is a good estimation that selection could lead to crop improvement. The use of *shoot dry weight, leaf area, seed number per plant* and *pod number per plant* in the selection index, successfully managed to identify 5 genotypes that have the potential to use as varieties in Botswana environment. The selection index also identified Namibian landraces as the best performers since 4 were selected among the best in 10 in the greenhouse experiment in UK and 5 were selected among the best 10 in the field experiment in Botswana.

Contributions in this chapter

- Multiple selection of characters was effectively employed in bambara groundnut genotype selection
- A combination of *leaf area, shoot dry weight, seed number per plant* and *seed weight*, are important traits that can be useful in determining the selection indices of landraces to identify best performing lines.

CHAPTER FIVE: Population structure and genetic diversity of bambara groundnut

5.1 Introduction

Many different landraces of bambara groundnut have been cultivated for a long time by farmers all over sub-Saharan Africa and have experienced a wide range of environmental conditions (Sesay, 2009). Farmers have also kept a large number of genetic resources on their farms.As a minor crop relatively little attention has been given to its genetic structure, despite a large germplasm collection being held at International Institute of Tropical Agriculture (IITA) and a number of countries in sub-Saharan Africa. An example is a gene bank that was established in 1988 by 15 member countries of Southern African Development Community (SADC), the SADC Plant Genetic Resources Centre based in Zambia (www.spgrc.org.zm). An appropriate application of genetic analysis requires a detailed knowledge of the genetic and historical relationships among and within landraces. This knowledge also assists in identifying inbred lines that have maximal diversity for use in breeding programmes (Liu *et al*., 2003).

Genetic resources for crop breeding are comprised from populations of genotypes collected from various places. It is important to maintain maximum genetic variability as well as identify origins or genotypes which are particularly useful for breeding. Plant genetic resources are the backbone of agriculture and play an important role in development of new cultivars (Malik and Singh, 2006).

Knowledge of genetic and trait diversity within a population and among populations is also important for conservation management and for identifying rare traits or genetic origins within species and to determine which could merit special attention (Zhuravlev *et al*., 2010). Genetic variability is important for landraces to adapt to environmental changes for their future survival, and for genetic and trait improvement in crop breeding (Upadhyaya*et al*., 2008; Mwale *et al.*, 2007).

Diversity and population estimations are useful measures for estimating different aspects of genetic structure in population studies (Gregorius, 2010). The population structure of a species can arise due to numerous factors such as the breeding system of the crop, effects of cultivation, breeding history and usage

since they can have significant effects on the partitioning of genetic diversity within and among populations (Hamrick and Godt, 1996). Population differentiation is affected by evolutionary processes, such as genetic drift, population size, selection founder effects and migration (Hedrick, 2005; Roy, 2000).

5.1.1 Genetic diversity in bambara groundnut

Early work done on bambara groundnut using isozyme markers by Pasquet *et al*., (1999) to investigate the population structure and partitioning of the observed genetic diversity between the wild and domesticated accessions lead to the conclusion that genetic diversity is present in both wild and domesticated bambara groundnut. The other major findings in their study are the almost complete absence of heterozygotes in both wild and domesticated forms and the highgenetic identities between the wild and domesticated forms. This led them to conclude that wild bambara groundnut is a true progenitor of the domesticated type.

Further genetic diversity studies were conducted by Ntundu *et al.,* (2004) they used AFLP markers to assess the genetic diversity among 100 bambara groundnut landraces from diverse geographical regions of Tanzania. They used 11 informative AFLP primer combinations and generated 49 scorable polymorphic fragments across all the selected accessions. Genetic distances between accessions ranged from 0.1 to 0.68 based on a Jaccard variability index, while the cluster analysis revealed that bambara groundnut consist of two major groups based on their geographic origins in Tanzania. These markers provided evidence that there is substantial genetic diversity within bambara groundnut.

5.1.2 Genetic diversity and population structure of other legumes.

A comparison of genetic diversity and population structure was conducted in 88 pigeonpea (*Cajanus cajan)* accessions from India and East Africa, using 6 microsatellites (Songok *et al*., 2010). Since India is the putative centre of origin while East Africa is the presumed secondary centre of diversity, as expected more diversity were recorded in India as compared to East Africa. Higher number of alleles (42) were observed in India and Nei's unbiased estimates of gene diversity (H) of (0.55) compared to East Africa with 31 alleles and (H) of 0.228.

Using 18 microsatellites Lazrek *et al*., (2009) investigated the genetic diversity of 136 lines of *Medicago truncatula*, of 10 populations from Tunisia. They detected an average of 4.2 alleles per locus, and an average gene diversity of 0.35. Population structure results based on analysis of molecular variance (AMOVA) showed that the majority (53.6%) of the variation present is mainly from the differences between populations. The significant variation was attributed to the difference between thenorthern and southern part of the country mainly due to the influences of eco-environmental factors.

Liu *et al.*, (2008) in China undertook a study on a total of 440 lentil(*Lens culinaris*)from the National Genebank (Chinese Academy of Agricultural Sciences, Beijing) with 204 accessions originally from China while 132 were introduced into China ('foreign') and the rest 104 were designated 'alien' with no traceable records. Fourteen SSR markers were used to investigate the genetic diversity and population structure of all three lentil accession groups. A total of 87 alleles were detected among the 440 accessions with an average of 6.2 alleles per locus, a mean observed heterozygosity (Ho) of 0.08 and an expected heterozygosity (H_E) of 0.56 were recorded. The researchers employed Principle Coordinate Analysis (PCoA) and cluster analysis for population structure analysis, with both techniques in agreement with each other as they separated the germplasm into three accession groups, with the 'foreign' materials proving to be the most diverse.

A population genetic structure analysis was conducted in 604 common bean accessions from the International Center for Tropical Agriculture (CIAT) using 36 SSR markers. A total of 679 alleles were detected with an average of 18.4 alleles per locus. The use of PCoA divided the collection into two main gene pools Mesoamerican and Andean (Blair *et al*., 2009). They conducted an analysis of molecular variance (AMOVA) to determine variation in gene pools, races, subgroups, and the difference between primary and secondary center of diversity for the crop. More variability was assigned to the genepools (36.77%) and races(32.57%) as compared to subgroups (32.09) but most of the variation remained within each subpopulation (Blair *et al*., 2009).

In the present study a set of 12 preselected SSR markers were employed to investigate the genetic diversity and population structure among 123 bambara groundnut landraces. One hundred and eighteen are African landraces originally from most parts of sub-Saharan Africa which covers four regions, namely; Central Africa, East Africa, Southern Africa and West Africa. Five landraces are from Indonesia (Asia); four were sourced directly from Indonesia while one was sourced from The University of Nottingham seed stock.

5.1.3 The objectives of the study:

- To analyse the population structure and genetic diversity of bambara groundnut based on pod and seed related characters and SSR markers
- To determine the association of morpho-agronomic markers based on (qualitative characters of seed and pods) with SSR marker.

5.2 Materials and Methods

5.2.1 Phenotypic data analysis

Eight-seven landraces from the 119 that were planted in the agronomy bay experiment (Table 2.1.2.2) reached maturity and produced reasonable pod and seed numbers. The pods and seed characterisation was based on IPGR descriptors (IITA, 2000). The measurements were taken on 10 seeds per plant using a Vernier calliper (Mitutoyo) for *pod length, pod width, seed length,* and *seed width.* Other measures were taken for *pod texture, pod colour, pod shape, seed testa colour, testa pattern,* and *eye pattern.* The *pod weight* and *seed weight per plant* were measured on an electronic balance and values were standardized (normalised), to remove scalar effects.

5.3 Results

5.3.1 Genetic diversity analysis

From an initial 75 microsatellites, 12 markers were selected for further use, due to their good amplification and ability to detect high levels of polymorphism in the initial analysis of SSR markers. These markers are used throughout for population structure analysis of bambara groundnut. A detailed characterisation of the 12 microsatellites on 123 bambara groundnut landraces revealed that all markers were polymorphic (Table 5.1).

Table 5.1: PowerMaker summary data analysis for the 12 microsatellites used in the analysis of 123 bambara groundnut landraces (118 from Africa and 5 from Asia/Indonesia).

Marker	**MAJ**	**GN**	**SS**	**No.**	**AN**	**Avail.**	**He**	**Ho**	**PIC**	***f***
Primer 7	0.61	6	123	123	5	1	0.55	0.01	0.49	0.99
Primer 15	0.30	21	123	123	16	1	0.80	0.08	0.78	0.90
Primer 16	0.60	8	123	123	8	1	0.59	0.00	0.55	1.00
Primer 19	0.14	24	123	123	23	1	0.93	0.01	0.93	0.99
Primer 23	0.64	6	123	123	6	1	0.51	0.00	0.45	1.00
Primer 33	0.40	16	123	123	15	1	0.75	0.02	0.72	0.97
Primer 37	0.38	13	123	123	12	1	0.78	0.01	0.76	0.99
Primer 44	0.89	5	123	123	5	1	0.21	0.00	0.20	1.00
mBam3co18	0.15	16	123	123	15	1	0.89	0.01	0.87	0.99
D11	0.32	19	123	123	17	1	0.83	0.02	0.81	0.97
D14	0.11	32	123	123	29	1	0.95	0.03	0.94	0.97
E7	0.66	5	123	123	4	1	0.47	0.03	0.39	0.93
Mean	**0.43**	**14**	**123**	**123**	**13**	**1**	**0.69**	**0.02**	**0.66**	**0.97**

MAJ: major allele frequency; **GN:** genotype number; **SS:** sample size;
No.: number of observation; **AN**: number of alleles; **Avail**: availability;
He: exp. heterozygosity;**Ho**: observed heterozygosity; **PIC**: polymorphic information content
f: inbreeding coefficient

The 12 microsatellites had high average polymorphic information content (0.66) and managed to detect a total of 155 alleles with an average of 14 alleles per marker. Polymorphic information content ranged from 0.20 for marker 44to 0.94 for maker D14. The average observed heterozygosity was 0.02 leading to a corresponding inbreeding coefficient of 0.97. The observed heterozygosity was much lower than the expected heterozygosity with departures from Hardy Weinberg equilibrium (Appendix 8) detected across all markers and estimates of inbreeding coefficient (*f*) between roughly 0.90 and 1.0 for all the markers. This is typical for a strongly inbreeding crop.

5.3.2 Genetic diversity within and among regions

A comparison of the genetic diversity of 123 bambara groundnut was compared among the 5 regions, based on estimates of total number of alleles, number of alleles per locus and two estimates of allelic richness and Nei unbiased estimates of gene diversity (H`) (Table 5.2). Higher genetic diversity was observed in Southern African populations at 0.70 and lower diversity in Asian populations at 0.18 probably due to fewer number of samples in this population.

Table 5.2: A comparison of the genetic diversity estimates among the five regions of Africa and Asia (Indonesia) analysis conducted using FSTAT 2.9.3 for all the 123 bambara groundnut landraces.

Regions	N	At	NA	Rs_1	Rs_2	Ho	H`
Asia (Indonesia)	5	1.50	18	1.50	N/A	0.00	0.18
Central Africa	10	3.58	43	3.06	3.58	0.02	0.48
East Africa	10	4.67	56	4.01	4.67	0.01	0.69
Southern Africa	31	7.42	9	4.31	5.73	0.04	0.70
Western Africa	67	9.83	118	4.37	6.06	0.01	0.65

N: Number of samples/genotypes **NA**: Number of alleles per locus **At** = Total number of alleles
Ho = Observed heterozygosity **N/a** = Estimates not calculated**H`** = Gene diversity
Rs_1 = Allelic richness based on sample size of 5 individuals
Rs_2= Allelic richness based on sample size of 10 individuals

To compare genetic diversity based on allelic richness, two estimates were calculated. The first allelic richness (Rs_1) was standardized based on the smallest number of landraces from Asia (5), while the second allelic richness estimate (Rs_2) was based on small samples from Central Africa and East Africa of 10 samples each (Table 5.2). The first and second estimates of allelic richness show West Africa to have a higher diversityat 4.37 and 6.06 respectively followed by Southern Africa at (Rs_1) of 4.31 and (Rs_2) of 5.73.

5.3.3 Principal coordinates analysis (PCoA)

Principal coordinate analysis was used to investigate the population structure of the 123 bambara groundnut genotypes. The results show a cumulative percentage of 16.15 %, for the first two axes with a variation of 9.87 % for Axis 1 and 6.28 % for Axis 2 as shown in Table 5.3.

Table 5.3: Principal Coordinate analysis (PCoA) from the investigation of population structure of 118 bambara groundnut landraces collected from Africa and 5 from Indonesia based on MVSP program.

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5	Axis 6	Axis 7	Axis 8	Axis 9	Axis 10
Eigenvalues	0.78	0.49	0.47	0.42	0.36	0.30	0.28	0.27	0.26	0.23
Percentage	9.87	6.28	6.01	5.31	4.51	3.79	3.53	3.42	3.27	2.94
Cumulative %	9.87	16.15	22.16	27.48	31.98	35.77	39.30	42.73	45.99	48.93

The PCoA for the 123 bambara groundnut landraces is shown in figure 5.1.0 for the first two axes, the population structure of the landraces are clearly demarcated based on their areas of origin. The most evident separation is the West African-Central African separation from the Southern Africa-East Africa Indonesian landraces which are clearly distinguished as two groups (Figure 5.1.0). There are also four landraces from the Southern African origin that share some potential characters/introgression with the West African landraces.

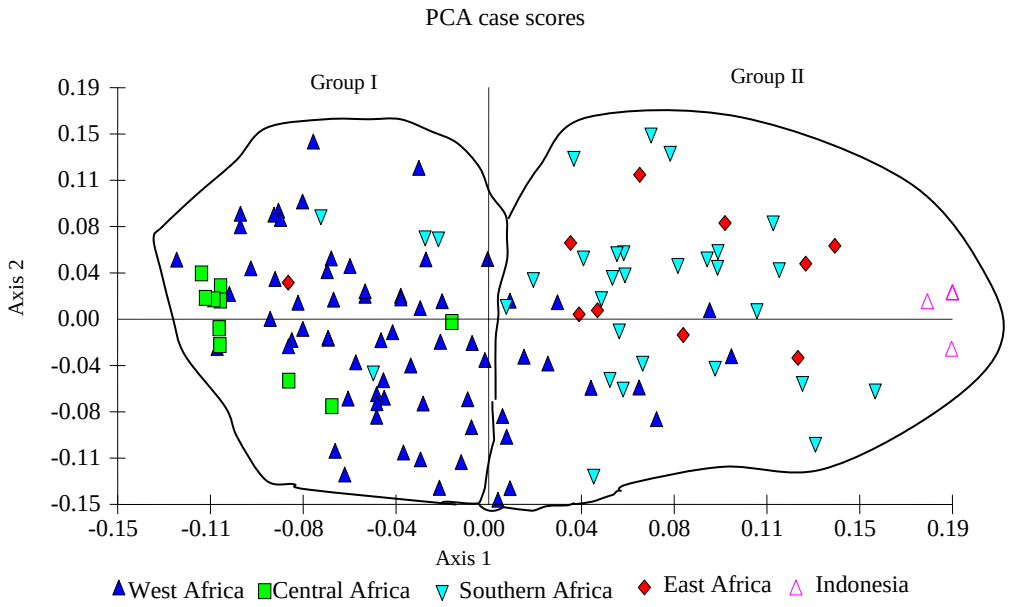


Figure 5.1.0 A PCO scatter plot for the 123 bambara groundnut genotypes from Africa and Indonesia generated from 12 microsatellites with MVSP program with a molecular variation of 16.15 %, with axis 1 contributing (9.87%) and while Axis 2 explained (6.28 %). The two cluster groups were hand drawn on Microsoft Word.

5.3.4 Cluster analysis

A dendrogram which shows a population analysis based on the four regions for Africa and one for Indonesia is shown in Figure 5.2.1. The cluster analysis is largely in agreement with the PCoA coordinates which clearly demarcated landraces based on their areas of origin. Landraces from one region are mostly grouped together, with some exception where some mixture are visible, as some dark blue traces could be found among the green colour coded landraces. The landraces from Southern Africa, East Africa and Indonesia are also clustered together while the West African and the Central African landraces are grouped together

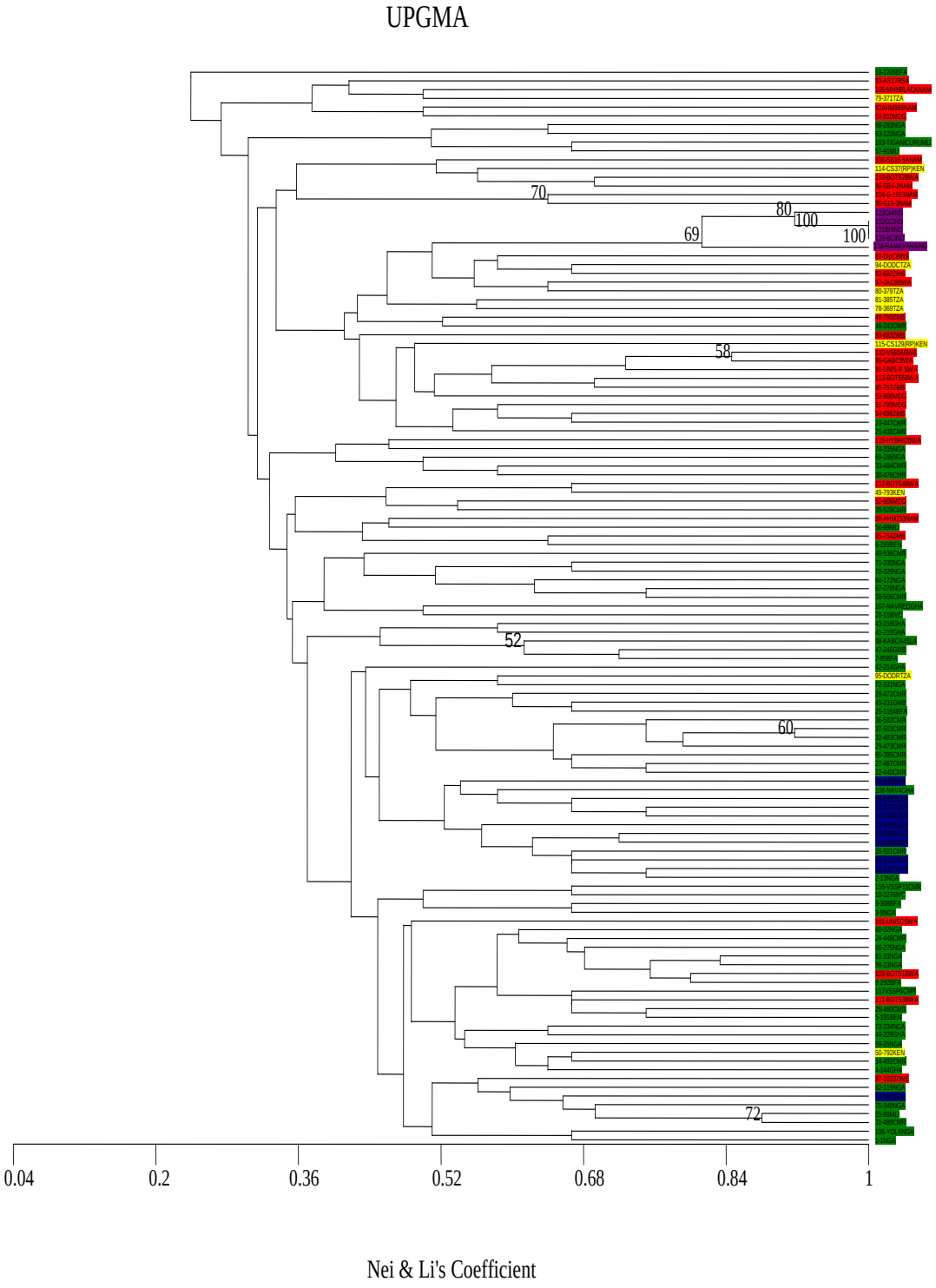


Figure 5.2.1. Cluster analysis of bambara groundnut landraces from five regions, from Africa and Indonesia (Asia). The dendrogram is based on 12 SSR markers. The Unweighted pair group method with arithmetic averages (UPGMA) tree was based on Nei and Li, (1979) coefficient of genetic similarity generated from the presence/absence binary matrix on 123 bambara groundnut landraces on MVSP. Colour codes :(**Purple: Indonesia), (Yellow: East Africa), (Red: Southern Africa), (Dark blue: Central Africa) and (Green: West Africa).**

5.3.5 Genetic differentiation based on F_{ST}

Genetic differentiation was generally high ranging from 0.610 for West Africa to 0.645 for East Africa, while in Central Africa was moderate at 0.440. The diversity was lower in the Indonesian landraces with 0.267 (Table 5.4).

Table 5.4: Genetic differentiation of the 123 bambara groundnut landraces from 4 regions of Africa and also Asia (Indonesia), estimated using Weir and Cockerham (1984) on Genepop version 4.0

Differentiation level	FST
Asia (Indonesia)	0.267
Central Africa	0.440
East Africa	0.645
Southern Africa	0.630
Western Africa	0.610

5.3.5.1 Pairwise comparison

F_{ST}- based genetic differentiation revealed significant differentiation ($P < 0.05$) among the landraces in all the regions except between the East African and the Southern African landraces and also between the East African and Asian landraces (Table 5.5). The results are consistent with PCoA results and cluster analysis, where the landraces from the two regions showed no clear separate groups. There was low but significant genetic differentiation between the West Africa landraces with the Central, East African and Southern African landraces possibly due to some mixture of the landraces between regions which was also revealed by PCoA and cluster analysis. The highest genetic distance was observed between Asian landraces and Central African landraces, these populations also have lower genetic differentiation

Table 5.5: Pairwise genetic distance based on F_{ST} values between populations, calculated on 12 microsatellites based on five regions of Africa including Asia (Indonesia).

Regions	Central Africa	East Africa	Southern Africa	West Africa
Asia (Indonesia)	0.510**	0.178^{ns}	0.166^{ns}	0.276**
Central Africa		0.204**	0.202**	0.095**
East Africa			0.027^{ns}	0.088**
Southern Africa				0.103**

**Significant at (P <0.05) and (ns) are not significant

5.3.6 Analysis of molecular variance analysis

The analysis of molecular variance (AMOVA) for all 123 bambara groundnut landraces was partitioned into a three-level of hierarchy which consists of variation in between the 2 groups as revealed in Figure 5.1.0. Group 1 consist of most of the West African landraces with the exception of 13, one from East Africa, four from Southern Africa and all landraces from Central Africa. Group 2 is made up of all Southern African landraces except four, all East African landraces except one, 13 landraces from West Africa and all the five landraces from Indonesia (Asia). AMOVA identified a highly significant (P <0.000) and meaningful variation at all the three hierarchy levels. The difference between the two groups was significant at 12.45 %, but the majority of variation was among individual landraces at 84.5% with little but significant variation within individuals at 3.05% (Table 5.8).Similar observations were made in the PCoA, cluster analysis and F_{ST}, where genetic differentiation was found between regions, countries and within genotypes.

Table 5.6: Analysis of Molecular Variance for the 123 bambara groundnut landraces based on 12 SSR markers using Arlequin version 3.1

Source of variation	df	Sum of squares	Variance components	Percentage variation	P value
Among populations	1	69.133	0.514Va	12.45	<0.000
Among individuals within populations	121	859.578	3.489Vb	84.50	<0.000
Within individuals	123	15.500	0.126Vc	3.05	<0.000
Total	245	944.211	4.129		

5.3.7 Comparison of molecular markers with pod and seed characters

For the morphological markers for the first two axes, the cumulative variation explained was 49.9% (Table 5.7) which was higher (based on 87 bambara groundnut accessions) compared to variation explained by the first two axes using SSR markers (16.08 %; Table 5.8). The PCoA for the pod and seed characters was able to group the West African landrace together, while the landraces from other regions do not show a clear pattern of separation. Most of the landraces were clustered together at the centre of the graph, which shows that they share common characters (Figure 5.3.0). The PCoA for the SSR marker data shows a greater dispersion among landraces, with clear separation for landraces from different regions (Figure 5.4.0).

Table 5.7: Principal Coordinate analysis (PCoA, Euclidean) for 15 characters of pods and seeds for 87 landraces that set reasonable seed numbers among the 119 landraces planted in the agronomy bay for bambara groundnut germplasm characterisation.

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5	Axis 6	Axis 7	Axis 8	Axis 9	Axis 10
Eigenvalues	3.98	1.93	1.07	0.68	0.58	0.4	0.37	0.33	0.29	0.26
Percentage	33.59	16.3	9.04	5.71	4.88	3.37	3.16	2.79	2.48	2.23
Cumulative %	33.59	49.9	58.9	64.6	69.5	72.9	76	78.8	81.28	83.51

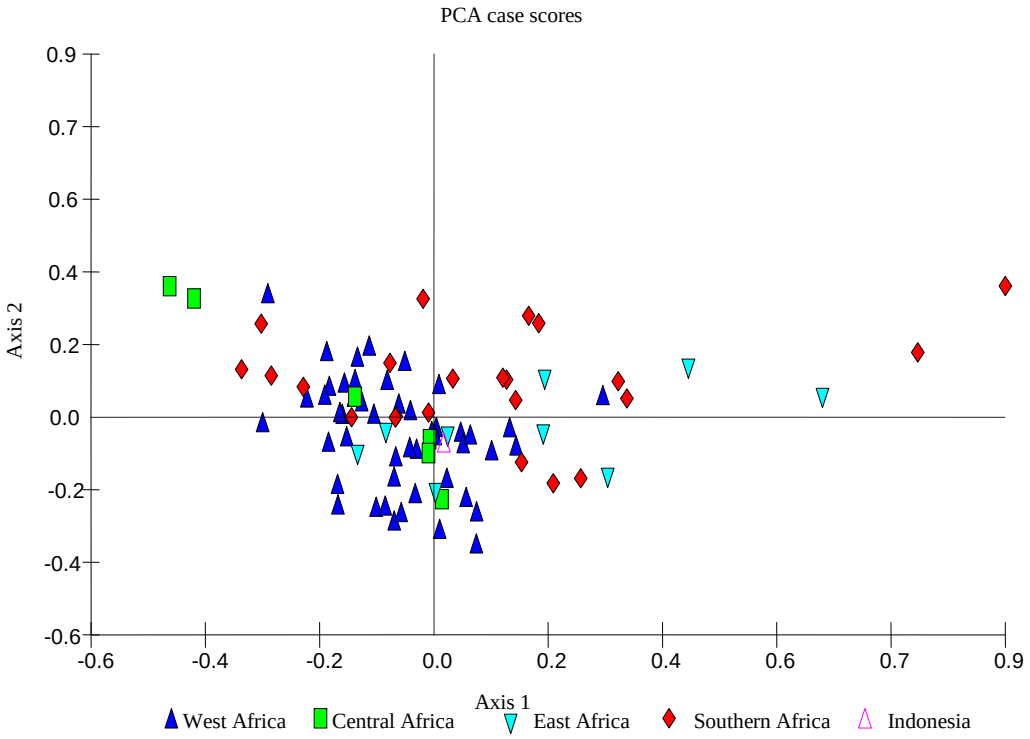


Figure 5.3.0: A PCO scatter plot for the 87 bambara groundnut that produced pods and seeds among 119 bambara groundnut planted. The data is based on 7 pod and 8 seed characters analysed using the MVSP program. The percentage variation for Axis 1 represents 33.59% and the Axis 2 represent 16.8 % with a cumulative percentage of 49.87% for the first two Axes.

Table 5.8: Principal Coordinate Analysis (PCoA) for 87 bambara groundnut landraces that set seed, based on 12 microsatellites

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5	Axis 6	Axis 7	Axis 8	Axis 9	Axis 10
Eigenvalues	0.73	0.51	0.46	0.43	0.35	0.33	0.31	0.30	0.27	0.25
Percentage	9.45	6.63	5.95	5.57	4.57	4.29	4.03	3.88	3.53	3.16
Cumulative %	9.45	16.08	22.03	27.60	32.17	36.46	40.50	44.38	47.91	51.07

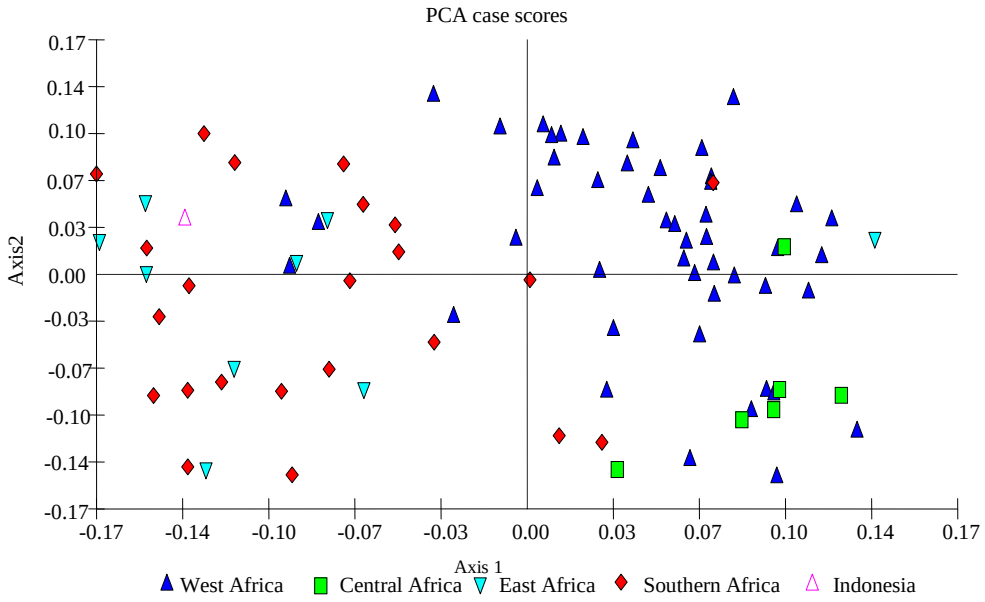
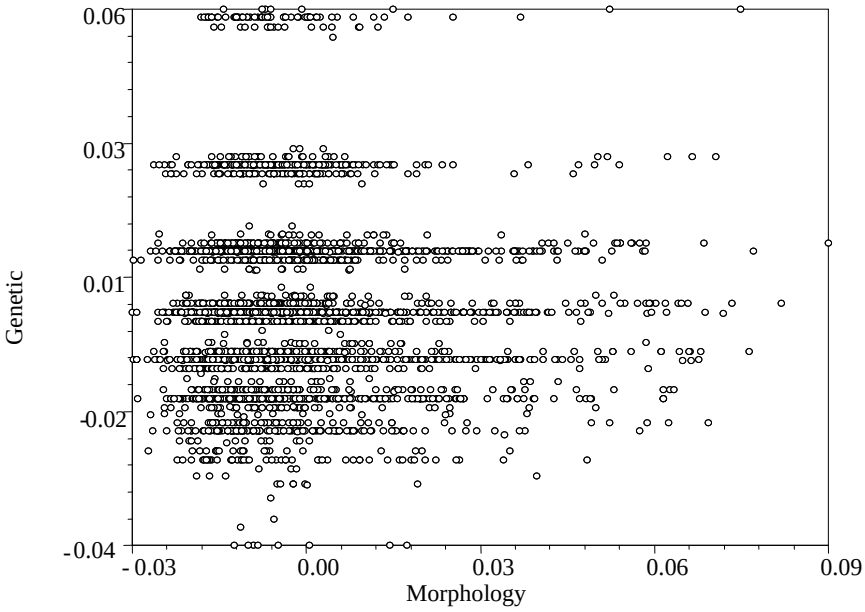


Figure 5.4.0.A PCO scatter plot on case scores for the 87 bambara groundnut that produced pods and seed based on 12 microsatellites, generated on MVSP program. The cumulative percentage of variation explained for the first two Axes is 16.08%, Axis 1 contributes 9.45% and Axis 2 contributes 6.63%.

The Southern African landraces together with the East African and Indonesia landrace are grouped together, while the West African and the Central African are grouped together. These groups are clearly separate from each other, with the exception of 4 landraces from Southern Africa and 7 from West Africa which may reflect some exchange of material or other gene flow between these regions, although simple errors in the accession records or samples order could potentially give similar effects.

The relationship between the Euclidean distance matrix based on the 13 morpho-agronomic characters of pod and seed and 12 SSR markers based on (Nei's 1972) genetic distance matrix were tested using a Mantel test correspondence test, Spearman rank correlation coefficient and examined through Pearson product-moment correlation coefficient on NTSYS and SPSS 16 (Table 5.9 and Figure 5.5.0).

a)



b)

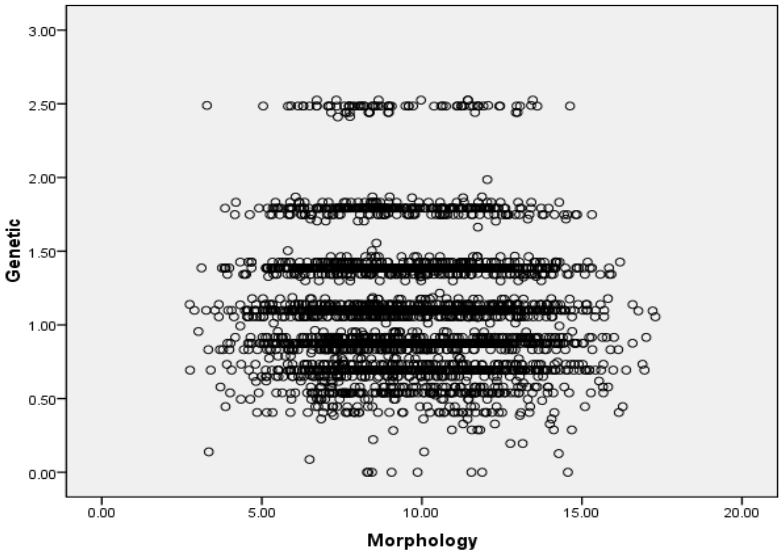


Figure 5.5.0 Scatter plot of correlation for morphological marker genetic distances estimate based on standard Euclidean and SSR marker genetic distance based on Nei's 1972, the analysis was conducted on (a) Mantel test correspondence test on NTSYS and (b) on Pearson correlation on SPSS version 16

A low, negative and non-significant correlation between the genetic distance matrices were recorded on Mantel correspondence test while Pearson correlation and the Spearman rank correlation recorded a low but highly significant correlations. This could be an indication that the two markers are explaining different variation in the selected materials.

Table 5.9: Correlation of molecular marker distance matrices, based on Pearson correlation, Spearman rank correlation and Mantel test for the 12 qualitative character and 12 molecular markers.

Marker	Pearson	Spearman	Mantel test
	Morphology	Morphology	Morphology
SSR	-0.048**	-0.038*	-0.0016
Morphology	1	1	1
N	3741	3741	3741
P value	0.003	0.021	0.488

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

5.4 Discussion

Genetic diversity in the 123 landrace accessions

The 123 bambra groundnut accessions used in this study from 21 countries covered all the regions of Africa, and also included 5 accessions from Indonesia. The collection of landraces represented is expected to represent the breadth of the diversity in bambara groundnut.

The genetic diversity and polymorphic information content recorded in this study gave averages for expected heterozygosity of 0.69, polymorphic information content of 0.66 and an average number of 13 alleles per locus (Table 5.1). Relatively high genetic diversity was observed possibly due to the fact that the material was from a wide geographic distribution from Africa and also Asia (Indonesia).

Diouf and Hilu, (2005) identified an average of 5.3 alleles per locus in 11 cowpea (*Vigna unguiculata*) varieties in Senegal, using 30 SSR markers. In common bean (*Phaseolus vulgaris*), Masi *et al.*, (2003) analysed 264 genotypes based on 30 SSR markers; they identified a total of 135 alleles and an average of 4.3 alleles per locus. Their study revealed lower number of alleles and average number of alleles per locus even though relatively higher number of SSR markers was used. However, it should be borne in mind that the SSR markers used here were deliberately selected after a pre-screen of an initial 75 showing polymorphism within bambara groundnut landraces.

Genetic diversity measures for the landraces were estimated, based on the total number of alleles and the number of alleles per locus. These measures are highly dependent on sample size and large samples usually contain more alleles than small samples (Kalinowski, 2004). The genetic analysis was conducted at a regional level, where landraces from several countries were grouped together. The problem of samples size was also resolved by the use of FSTAT software to calculate the allelic richness in each population based on smallest number of individual samples (Leberg, 2002). The program estimates allelic richness (R_s) independent of the sample sizes, and this allows a comparison of genetic diversity between populations with different sample sizes.

When allelic richness was measured based on a minimum of 5 samples (Rs_1) and 10 samples for (Rs_2), West African landraces revealed a slightly higher genetic diversity at 4.37 and 6.06 followed by the Southern African landraces with (Rs_1) 4.31 and (Rs_2) at 5.73(Table 5.2). For the gene diversity H`, Southern African and East African at 0.70 and 0.69 respectively revealed a slightly higher diversity compared to West Africa at 0.65.The relatively low diversity estimates for the West African landraces, which may be below what would be expected in the major centre of diversity compared to the Southern African landraces could be caused by the sampling strategy which biased the sampling to include more landraces from 2 countries among West African samples. A total of 67 genotypes that were sampled from West Africa, 45 came from Nigeria and Cameroon which is approximately 67 % of biased towards these two countries. While the Southern African countries contributing the majority of the landraces were Namibia and Botswana with 16 out of 30 genotypes, which is approximately 55 %. Nei unbiased estimates of gene diversity is mostly influenced by moderate allele number and allelic frequency rather than alleles of high or low frequency (Shete, 2003; Songok *et al.,* 2010).

Population structure and relationship among different landraces

Low F_{ST} measures were observed in the Indonesian landraces as compared to the African landraces, and this could be caused by long period of time the Indonesian landraces expansion to Indonesia took place. In addition the low differentiation from Indonesian landraces could also be caused by few entries of bambara groundnut from different places in Africa or the limited number of Indonesian accessions.

The SSR-based population structure analysis using principal coordinate (PCoA) analysis clearly defined landraces based on their areas of origin which is in agreement with previous work on bambara groundnut by Amadou *et al.,* (2001) and Massawe *et al.,* (2002). The landraces from West Africa tend to group with those from Central Africa, while those from Southern Africa clustered with those from East Africa and Indonesia. In fact there was no significant difference between the Southern Africa and East African landraces based on F_{ST} pairwise genetic differentiation between regions (Table 5.5). This is also a reflection of a

potentially extensive movement of seeds materials between these regions, by farmers.

F_{ST} estimates correlation between genes of different individuals in the same population and is used as a measure of genetic differentiation among populations F_{ST}= 0, between subpopulations indicates that they are identical in all allele frequencies, but when it is F_{ST} = 1 they are fixed for different alleles. In a pairwise comparison between the five regions, the F_{ST} ranged from 0.027 between East Africa and West Africa, to 0.510 between Indonesia and Central Africa. The East African landraces also show a lot of common alleles with the West African landraces at 0.088.

The cluster analysis, based on UPGMA (Nei and Li, 1979) displayed a minimum similarity of 24% among the 123 accessions and clustering occurred based on areas of origin, in a similar way to the PCoA results, although only two major groups were clearly differentiated.

The study revealed a clear structure for these 123 bambara groundnut landraces. The total analysis of molecular variance (AMOVA) results for the 123 landraces revealed that most of the variation was among individual genotypes (84.5%) followed by among populations (12.45%) and last is within individuals (3.05%). Massawe *et al.*, (2003) found high levels of polymorphism among landraces at 77.1% with 28.7% within genotypes using RAPDs markers. Wasike *et al.*, (2005) used AFLP to study the genetic diversity of 32 African and 9 Asian pigeonpea (*Cajanus cajan*) varieties, the analysis of molecular variance estimates between the two regions revealed a higher genetic variation of 92.16% within the populations while only 7.84 was among the two regions. In cowpeas, Zannou *et al.*, (2008) also reported a higher percentage of variation within accessions (73%) as compared to among groups (26%) which indicates a higher within population diversity for these species, possibly associated with the accessions originating from different ancestors.

Comparison of genetic distance estimates from morpho-agronomic and SSR markers

Morphological character assessment is the first step in the characterisation of germplasm, because breeding programmes rely on the magnitude of phenotypic variability in crops. Qualitative characters examined are mostly influenced by the consumer preference and the socio-economic conditions at a particular time (Gafoor *et al.,* 2002). In this study both morphological and SSR markers were able to differentiate the landraces. Correlation of the genetic distance estimates from the two marker types showed non-significant and negative correlation according to Mantel test, which indicates that, both marker typesdiscriminate differently among the genotypes. The lower levels of negative but highly significant correlations observed using Pearson and Spearman rank between morphological and DNA markers could also indicate some agreement between the phenotypic and molecular marker. However, the fewer number of markers used in the study may also contribute to lower correlations because of sampling of the genome is low (Vieira *et al.,* 2007) and here the sampling of the genome is relatively shallow (12 SSR). In addition morphological markers are less reliable, and efficient in clearly discrimination genotypes as compared to molecular markers in genetic relationships (Bayele *et al.,* 2005).

5.5 Conclusions

The East African landraces had a slightly higher differentiation to both Southern Africa and the West African landraces. AMOVA also revealed a low variance among the two major groups which are mainly clustered based on regions origins of (West African, Central Africa) and (Southern Africa, Indonesia, East Africa) as more variation was observed between genotypes. The poor correspondence between distance estimates based on morphological traits and SSR markers was observed. The correlation estimates for Pearson product –moment coefficient and spearman rank correlation coefficient gave negative and significant correlation while Mantel test gave a negative and non-significant correlation.

CHAPTER SIX: Genetic diversity of bambara groundnut based on SSR markers and the comparison with morpho-agronomic characters

6.1 Introduction

The application of molecular markers is widely accepted as a potentially powerful tool in crop improvement of a number of crops. Characterisation of plant genetic resources represents a good starting point to dissect allelic variation and identify variation in crops (Upadhyaya *et al*., 2008). Genetic diversity can be measured at various levels including within accessions (particularly for landraces), between accessions and also among species, with the phylogenetic relationships revealing how a group of species are related (Wang *et al*., 2009).

Molecular markers could be used to identify genetically different populations and use them in selecting parents so that inbreeding can be avoided as has been done in a highly inbreeding alfalfa (*Medicago sativa*)(Noeparvar *et al.,* 2008). DNA markers can be linked to agronomic characters and thus are useful in marker-assisted selection (MAS) in plant breeding. MAS could then be used in the selection of crops and make plant breeding more effective and efficient (Collard *et al*., 2005).

A number of molecular marker systems have been employed in bambara groundnut for genetic diversity assessment, including; RAPDs (Amadou *et al*., 2001), AFLP (Massawe *et al.,* 2002; Ntundu *et al.,* 2004); SSR markers (Basu *et al*., 2007) and isozymes (Pasquet *et al.,* 1999). Recently DArT markers and morphological markers have been used and compared (Olukolu *et al.,* 2011). A more robust approach to estimate genetic variation could be realised if both morphological and molecular techniques are simultaneously used (Parsaeian *et al.,* 2010).

Botswana is a country with semi-arid climatic conditions and usually low levels of cereal grain yield (mostly maize and sorghum) due to poor soils and low moisture. Well adapted crop genotypes in that environment such as bambara groundnut could be used to increase food production (Brink *et al.,* 2000). The first expedition for a bambara groundnut survey in Botswana was reported in 1947, followed by a second one in 1985 (Ministry of Agriculture, Botswana; Appa-Rao *et al.,* 1986), and these covered only the northern part of the country, which

covered only two, out of a potential ten, districts in the country (Karikari *et al.*, 1995). Few bambara groundnut landraces have been tested for adaptation and undergone selection for high yields in the Botswana environment. It is therefore important to identify genetic variation among bambara groundnut landraces, using both the molecular markers and morphological markers in the target environment.

6.1.1 Genetic diversity of bambara groundnut

Massawe *et al*., (2002) reported substantial genetic diversity among 16 bambara groundnut landrace single genotype samples when they used AFLP with a combination of seven primer pairs. Pairwise similarities between landraces were determined according to Jaccard coefficient and the matrices were used to produce dendrograms on unweighted pair-group method with arithmetic mean (UPGMA) cluster analysis. Landraces were grouped into three based on their geographic areas of origin and Southern Africa landraces DipC1995 and Malawi5 were grouped together, with no landraces samples identical. Amadou *et al*., (2001) used Random Amplified Random Amplified DNA (RAPD) to assess the genetic diversity of 25 single genotype landrace accessions collected from the International Institute of Tropical Agriculture (IITA). The landraces clustered into two main groups based on their areas of origin.

Massawe *et al*., (2003) also used RAPD markers on 12 bambara groundnut landraces with multiple genotypes per landrace. Data from individuals of each landrace was analysed to determine the level of heterogeneity within the landraces. AMOVA revealed highly significant variation ($P<0.001$) among landraces and also within each individual landrace. The partitioning of total genetic diversity showed that 71.25 % was explained by the landraces differences while 28.67% was among individuals within landraces. However, Stadler, (2009) used Diversity Array Technology markers (DArT) on bambara groundnut and found that intra-landrace diversity was lower among some landraces than others. This also shows that some landraces are 'purer' than others, which could be a good basis for selection of pure lines.

6.1.2 Genetic diversity in other leguminous crops

Several authors have reported a general low level of polymorphism among cultivated peanut germplasm (He *et al.*, 2003; Wang *et al.*, 2007). The narrow

gene pool of the cultivated peanut has been attributed to the evolution that occurred in South America through a limited number of interspecific hybridizations and polyploidization (Mace *et al.,* 2006) and this has led to limited genetic diversity of cultivated peanut, through a genetic bottleneck. Gimenes *et al.,* (2007) observed lower genetic diversity in groundnut (*Arachis hypogaea*) when studying 16 accessions of *A. hypogaea* and 38 accessions of eight other sections of *Arachis* using 13 microsatellites markers. They observed mean polymorphic loci (33%), mean number of alleles (4.02) and mean polymorphic information content (0.48) and He *et al.,* (2003) recorded a similar number of alleles per locus on 24 genotypes, when using 19 SSR markers of 4.25 alleles per locus.

The cultivated chickpea, as a self-pollinated crop with $2n = 2x = 16$ shows a lower genetic diversity as compared to the wild *Cicer*. Upadhyaya *et al.,* (2008) reported substantial genetic diversity based on the use of 48 SSR markers to analyse 2915 chickpea accessions (*Cicer arietinum*). They identified 1683 alleles, with a range of 14 to 67 alleles per locus and an average of 35. The polymorphic information content (PIC) ranged from 0.467 to 0.974 with an average of 0.854. These very high observations of genetic diversity was attributed to the large set of accessions analysed from the Mediterranean and African regions which are the center of origin and center of diversity, respectively (Upadhyaya *et al.,* 2008). Castro *et al.,* (2011) also observed higher genetic diversity among 32 commercial cultivars of chickpea using 15 microsatellites markers. They detected a total of 154 alleles, 10.3 mean number of alleles per locus and an average PIC of 0.78

A similar observation was found among 40 genotypes representing seven *Cajanus* species which consists of 32 cultivated type and 8 wild forms by Saxena *et al.,* (2010). They employed 16 microsatellites, which yielded a total of 72 alleles with an average of 5.5 alleles per marker in the germplasm. Allele numbers ranged from 2 to 8, PIC values for these markers ranged from 0.05 to 0.55, with an average of 0.32 per marker. Higher genetic diversity was observed in the wild type with a PIC of 0.64 and an average of 5 alleles compared to the cultivated form with a PIC of 0.15 and an average allele number of 2.08.

Yang *et al.,* (2006) used DArT markers in pigeonpea (*Cajanus cajan*) identified low levels of genetic diversity cultivated pigeonpea compared to its wild relatives. They evaluated 20 species of *Cajanus*, and identified a total of 700 markers that were polymorphic, but only 64 markers were polymorphic in cultivated accessions and this indicates the narrow genetic base of cultivated pigeonpea.

In soybean (*Glycine max*), Liu *et al.,* (2011) observed 250 alleles among 91 accessions at 35 SSR loci, and an average of 7.14 alleles per locus, and an average PIC of 0.74 in a study conducted in Shaanxi Province of China.

6.1.3 Efficiency of molecular and morphological markers in genetic diversity estimates

A number of approaches for measuring genetic distance such as the analysis of morphological characters or molecular markers have been widely used to try to measure crop diversity. The differences in DNA sequences between individuals detected when using molecular markers are often more informative compared to morphological markers (Tanksley *et al.,* 1989). There are several other advantages for molecular marker application; they are reliable, not influenced by environmental conditions and are essentially Mendelian markers. In some instances adequate levels of polymorphism are not available; therefore they can be limited in the evaluation of genetic diversity (Cupic *et al.,* 2009).Tantasawat *et al.,* (2010) compared the use of morphological and SSR marker for genetic diversity and relatedness studies in 17 mungbean (*Vigna radiata*) and 5 blackgram (*Vigna mungo*) accessions. The two species are mainly differentiated by seed colour, with some differences in seed shape and pod colour. In their findings morphological characters were not able to differentiate between the two Vigna species compared to SSR markers which were able to distinguish the two species, which is an indication that molecular markers can be more effective in differentiating the two species.

In common bean (*Phaseolus vulgaris*), Kumar *et al.,* (2009) compared the morpho-agronomic traits and microsatellites in genetic diversity analysis of 115 common bean. Seventy were Indian landraces, 24 released varieties and 21 exotic accessions. The Euclidean distance based dendrograms and the PCO were able to separate varieties from genotypes but based mainly on yield and yield related

traits while the microsatellites marker PCO and UPGMA clearly separated the genotypes into their respective groups. The two marker types were used to complement each other. However, the use of a Mantel test, revealed a good correlation between the morpho-agronomic distance and molecular marker genetic distance estimates (r = 0.876), which indicates that either of the marker can give a good reflection of genetic estimates from another marker (Kumar *et al.,* 2009), given this, morphological markers may well be simpler to apply in breeding situations.

Ntundu *et al.,* (2004) estimated phenotypic distances calculated on 20 quantitative and 7 qualitative traits, and also genetic distances based on 49 AFLP polymorphic markers in 100 bambara groundnut single genotype accessions in Tanzania to determine the relationship between the two markers types. A low correlation of r = 0.41 was recorded, while the clusters of accessions based on AFLPs compared well with that based on phenotypic characters.

Gomez *et al.,* (2004) studied the molecular and genetic diversity of common bean (*Phaseolus vulgaris*) landraces in Nicaragua using 14 traits measured in 12 individual landraces with seven SSR loci. The use of both morphological and SSR markers provided complementary information, since the variation at the molecular level was mostly between and within landraces, but did not reveal consistent differences between ecological zones, while the phenotypic variation corresponded to the ecological zones. The molecular differentiation of the landraces at F_{ST} = 0.34 was due to founder effects, while phenotypic differentiation was attributed to the effect of adaptation.

In this study a set of molecular data and morphological characters were recorded on the same landraces with an aim to evaluate the efficiency of these two techniques in bambara groundnut, so that either morpho-agronomic or DNA markers could be used, or both as a compliment to one another. The aim of this study was to assess the genetic diversity of bambara groundnut and to estimate the genetic correlation between the morphological genetic distance estimates and molecular (SSR) genetic distance estimates in bambara groundnut landraces.

6.2 Materials and methods

6.2.1 Plant Materials used

Thirty five bambara groundnut landraces were selected among the 119 accessions that were planted in the agronomy bay, listed in (Table 2.1.2.2). Three individuals were used which makes 105 genotypes per landrace.

6.2.2 Markers used

Twenty microsatellites which showed good amplification and had been previously shown to be polymorphic were selected from a pool of 75 markers and listed in Appendix 2.

6.3 Results

6.3.1 Polymorphism of microsatellites in bambara groundnut

A total of 105 genotypes were amplified with 20 microsatellites, a total of 231 alleles were identified with an average of 12 alleles per locus (Table 6.1). The highest number of alleles was recorded for marker D14 with 29 alleles and the lowest allele numbers were recorded for marker E7 with 3 alleles. Polymorphic information content (PIC) ranged from 0.07 to 0.95, (markers D8 and D14, respectively) with an average of 0.67. The genetic diversity detected using all microsatellites across the genotypes was high with a range of 0.07 to 0.95 and a mean of 0.69.

Table 6.1: Summary of PowerMarker data analysis for the 35 bambara groundnut landraces using 20 microsatellites analysis conducted on each of the 105 individual genotypes.

Marker	**MAJ**	**GN**	**SS**	**No.**	**AN**	**He**	**Ho**	**PIC**	***F***
Primer 1	0.74	6	105	105	6	0.43	0.00	0.41	1.00
Primer 7	0.61	6	105	105	6	0.59	0.00	0.55	1.00
Primer 10	0.33	10	105	105	10	0.78	0.00	0.75	1.00
Primer 15	0.13	20	105	105	20	0.92	0.00	0.91	1.00
Primer 16	0.41	8	105	105	8	0.75	0.00	0.72	1.00
Primer 19	0.14	21	105	105	21	0.92	0.00	0.91	1.00
Primer 21	0.50	7	105	105	7	0.62	0.00	0.55	1.00
Primer 23	0.76	6	105	105	6	0.40	0.00	0.38	1.00
Primer 30	0.69	6	105	105	6	0.49	0.00	0.45	1.00
Primer 31	0.39	9	105	105	9	0.75	0.00	0.71	1.00
Primer 32	0.17	21	105	105	19	0.92	0.02	0.91	0.98
Primer 33	0.30	16	105	105	15	0.84	0.01	0.82	0.99
Primer 37	0.39	14	105	105	12	0.76	0.03	0.73	0.96
Primer 44	0.51	8	105	105	8	0.67	0.00	0.64	1.00
D8	0.96	4	105	105	4	0.07	0.00	0.07	1.00
mBam2co80	0.18	16	105	105	16	0.89	0.00	0.88	1.00
D11	0.30	15	105	105	15	0.84	0.00	0.82	1.00
D14	0.10	31	105	105	29	0.95	0.02	0.95	0.98
D15	0.24	12	105	105	11	0.85	0.01	0.83	0.99
E7	0.67	3	105	105	3	0.45	0.00	0.36	1.00
Mean	**0.43**	**12**	**105**	**105**	**12**	**0.69**	**0.00**	**0.67**	**0.99**

MAF- Major allele frequency **GN**-Genotype number observed **SS** –Sample size

No. obs- Number of observations **NA**- Allele Number **GD**-(Expected heterozygosity)

Het.- Heterozygosity **PIC**- Polymorphic information content ***f***- Inbreeding coefficient

Table 6.2: Intra-landrace diversity among the 35 genotypes conducted on each of the three genotypes per landrace using 20 SSR markers based on Arlequin version 3.1

Landraces	No. observation	Polymorphic loci	Average no. alleles	Ho	He
3Acc 9	6	8	1.45	0.02	0.22
4Acc144	6	5	1.30	0.03	0.14
6Acc 289	6	5	1.30	0.00	0.15
10Acc 1276	6	6	1.35	0.00	0.17
20Acc118	6	3	1.15	0.00	0.08
30Acc 476	6	5	1.25	0.00	0.13
33Acc 484	6	3	1.20	0.00	0.09
40Acc 536	6	5	1.25	0.00	0.13
45Acc 231	6	6	1.35	0.00	0.17
48Acc790	6	11	1.80	0.00	0.35
49Acc793	6	11	1.65	0.02	0.32
50Acc 792	6	13	1.80	0.00	0.39
56Acc 89	6	9	1.50	0.00	0.25
60Acc 32	6	8	1.45	0.00	0.23
69Acc286	6	8	1.40	0.01	0.23
70Acc 329	6	6	1.30	0.00	0.16
74Acc335	6	7	1.35	0.00	0.19
76Acc390	6	8	1.40	0.00	0.21
81Acc385	6	14	1.80	0.02	0.39
84Acc696	6	5	1.35	0.00	0.16
85Acc 754	6	11	1.60	0.00	0.31
88AHM753	6	5	1.30	0.02	0.14
90S19-3	6	8	1.50	0.00	0.24
91UNIS R	6	14	1.95	0.00	0.44
92AHM968	6	4	1.20	0.00	0.11
95DODR	6	8	1.55	0.02	0.26
99SB4-2	6	6	1.35	0.00	0.17
100SB16 A	6	7	1.35	0.00	0.19
104S-1913	6	10	1.50	0.00	0.27
105MHN black	6	6	1.35	0.00	0.17
109BWA1	6	11	1.55	0.00	0.29
113BWA5	6	12	1.75	0.02	0.36
117VSSP6	6	2	1.10	0.00	0.05
118Ramayana	6	7	1.35	0.03	0.18
119Hyrid	6	7	1.40	0.07	0.19
Mean	**6**	**8**	**1.43**	**0.01**	**0.22**

All the markers showed an inbreeding coefficient close to 1, and the observed heterozygosity (*Ho*) is far lower than the expected heterozygosity (*He*) and close to 0, as shown in Table 6.1 and Table 6.2. Each landrace was analysed for intra-landrace diversity based on three genotypes sampled. The majority (28) landraces recorded narrow genetic diversity of an average of 1.5 number of alleles per locus and below. The more diverse landraces were 113-BWA 5 (Botswana), 48-Acc231 and 50-Acc 792 both from (Kenya), 81Acc385 (Tanzania) and Uniswa Red (Swaziland) that had at least 1.75 alleles per locus.

Three genotypes from each of the 35 landraces were investigated for genetic similarity and this made up a total of 105 individual genotypes samples that were analysed based on the Nei and Li (1979) similarity index. Then a cluster analysis was conducted based on the similarity matrix produced using UPGMA procedure on MVSP. The similarity of the genotypes ranges from 0.37 to 0.95 (Figure 6.1.1). Genetically similar genotypes were observed between 20Acc118 from Côte d Ivoire, while those genetically further from each other were104S-1913 from Namibia and 10Acc1276 from Central African Republic at 0.108.

Forty three points of high bootstrap values more than 50% indicated on the dendrogram showed that branches are well supported. Among the 43 bootstrap values 20 were supportingthree individuals within a genotype;this showed that 57 % of genotypes had individuals which are more similar to each other.

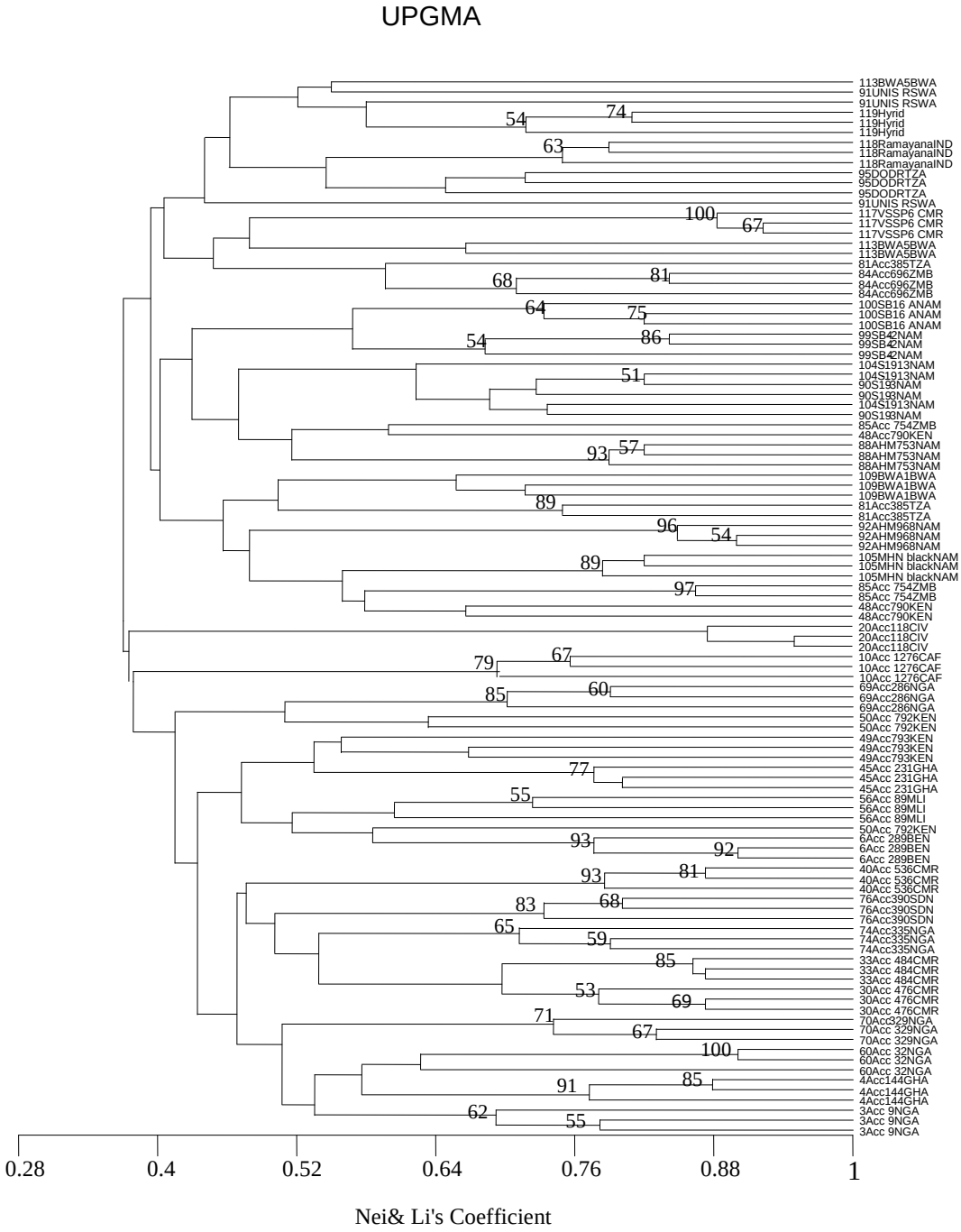


Figure 6.1.1: UPGMA dendrogram of 105 bambara groundnut genotypes revealed by UPGMA cluster analysis of 20 SSR markers based on Nei and Li, 1979 similarity estimate. Bootstrap values of 1000 replications more than 50% are shown on corresponding nodes.

6.3.2 Principal Component Analysis (PCO)

Eigenvalues and the cumulative percentage of the principal component case scores were used for the analysis of 105 genotypes. Data for the first two axes accounted for a total variation of 14.95 % (Table 6.3).

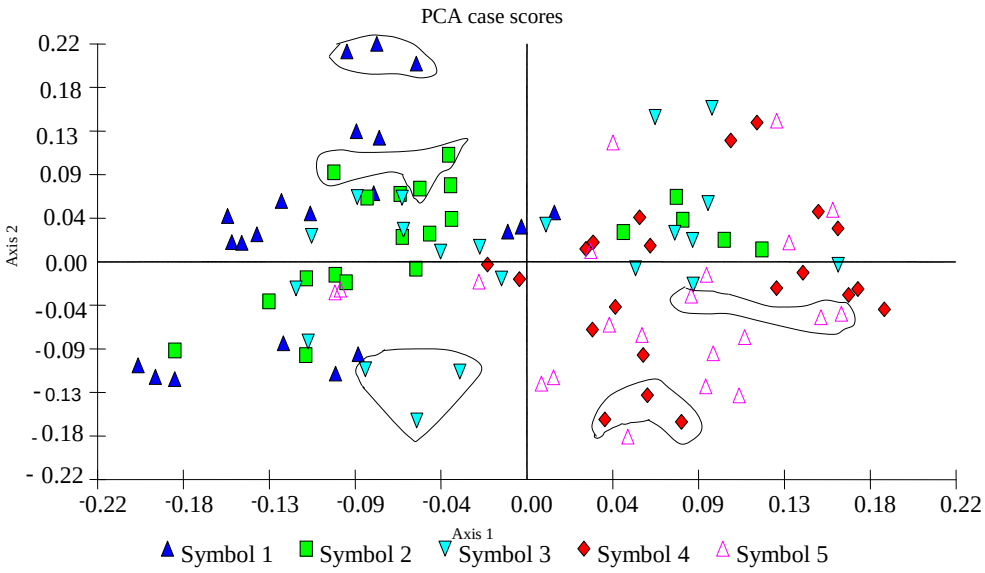
Table 6.3: PCO case scores for the population structure of the 105 genotypes determined from each of the three samples of the 35 bambara groundnut landraces based on 20 SSR markers

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5	Axis 6	Axis 7	Axis 8	Axis 9	Axis 10
Eigenvalues	1.02	0.73	0.67	0.59	0.53	0.47	0.47	0.42	0.41	0.36
Percentage	8.69	6.27	5.70	5.05	4.50	4.02	4.00	3.60	3.52	3.09
Cumulative %	8.69	14.95	20.65	25.70	30.19	34.22	38.21	41.81	45.33	48.42

Principle coordinate analysis allowed separation of genotypes mainly based on their areas of origin (Figure 6.2.1). All the West African landraces were in group one on the left panel of the diagram together with one landrace from Central Africa (76Acc390SDN), while the other one (10Acc1276CAF) grouped with the Southern African landraces in group 1. The Southern African, Indonesian and East African landraces are on the right panel of the diagram in group two with the exception of three individuals of 50Acc792 from Kenya and one individual of 49Acc793 from Kenya

Substantial variation was shown in landraces as they spread on the upper and lower panel of the diagram from their respective groups. All individuals from the landraces were able to be uniquely identified by the markers (figure 6.1.1 and Appendix 9).

a)



b)

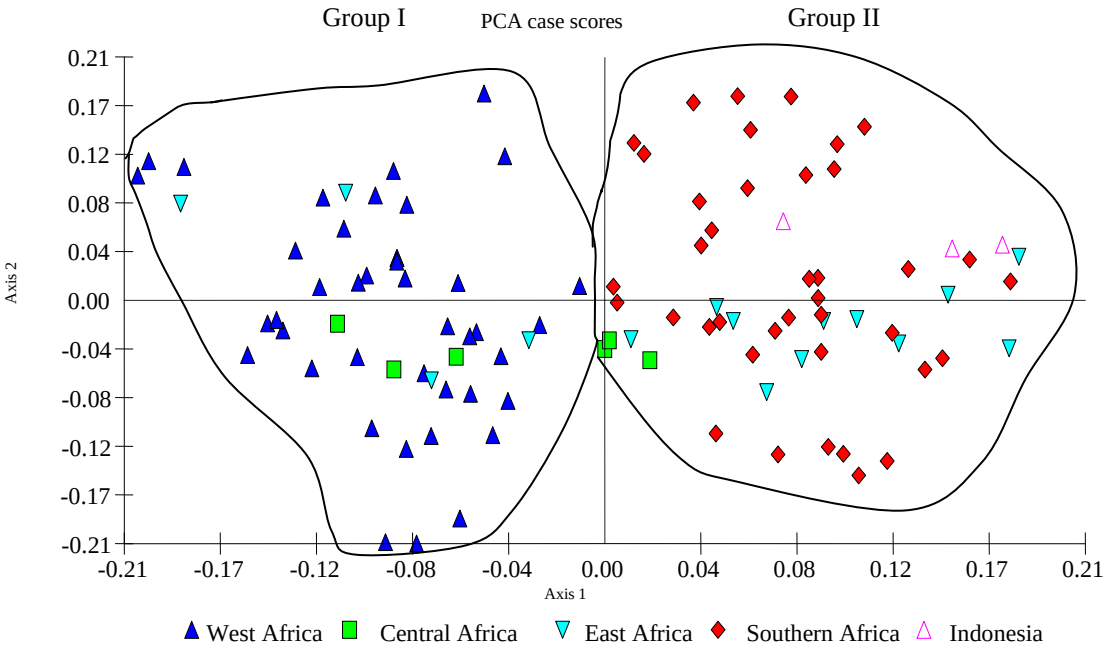


Figure 6.2.1: The first two axes of the PCO case scores, generated from the 105 bambara groundnut genotypes based on 20 SSR markers generated on MVSP, the first Axis accounts for 8.69 % while Axis 2 represent 6.27 % and together explain a cumulative 14.95 % of the molecular variation. Figure a: shows a PCO plot demarcated on 5 symbols to identify three individuals from one landrace while Figure b: shows the grouping of the five regions into two major groups. The two cluster groups were hand drawn on Microsoft Word.

6.3.3 Comparison of SSR and morphological markers

Since accessions are not homogenous as some groups of individuals from different accessions are more closely related than individuals within an accession (Figure 6.1.1), 34 lines derived from seed from a single plant were selected to study variation of morphological and agronomic traits following the IPGRI descriptors (IITA, BAMNET, 2000). For comparison of molecular (SSR) and morphological markers, PCO analysis, cluster analysis and correlation matrix was conducted on both data set based on 20 SSR markers and 37 morpho-agronomic traits.

6.3.3.1 Principal Component Analysis

The analysis of the percentage variation of principal component scores were used to reveal the differences between the two marker types. The results are shown in Table 6.4 and Table 6.5. Looking at the first two Axes suggests that morphological markers are revealing more variation (22.36 %) as compared to SSR markers (18.25 %) in the first two axes.

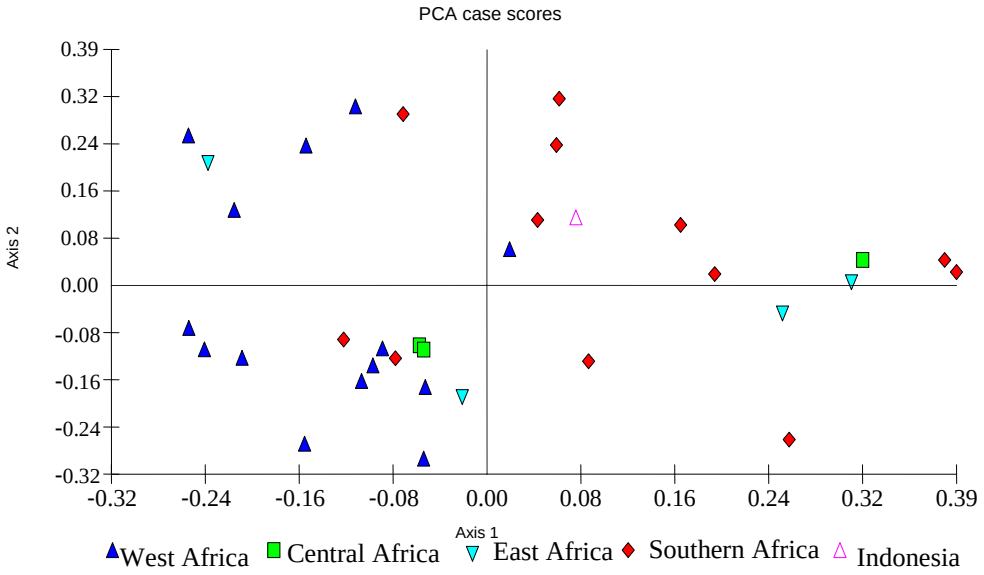
Table 6.4: PCO case scores for the population structure of the 34 bambara groundnut selected for field studies in Botswana, analyses based on 20 SSR markers

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5	Axis 6	Axis 7	Axis 8	Axis 9	Axis 10
Eigenvalues	1.16	0.98	0.92	0.72	0.62	0.60	0.58	0.53	0.49	0.45
Percentage	9.90	8.35	7.85	6.12	5.33	5.09	4.95	4.52	4.22	3.82
Cumulative %	9.90	18.25	26.10	32.21	37.54	42.64	47.59	52.11	56.32	60.14

Table 6.5: PCO case scores for the population structure of the 34 bambara groundnut based on 37 morpho-agronomic characters, from the field experiment conducted in Botswana

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5	Axis 6	Axis 7	Axis 8	Axis 9	Axis 10
Eigenvalues	2.13	1.38	1.24	1.10	0.99	0.85	0.74	0.70	0.67	0.61
Percentage	13.56	8.80	7.89	7.00	6.32	5.40	4.72	4.45	4.25	3.89
Cumulative %	13.56	22.36	30.25	37.25	43.57	48.97	53.68	58.13	62.38	66.27

a) SSR



b) Morphology

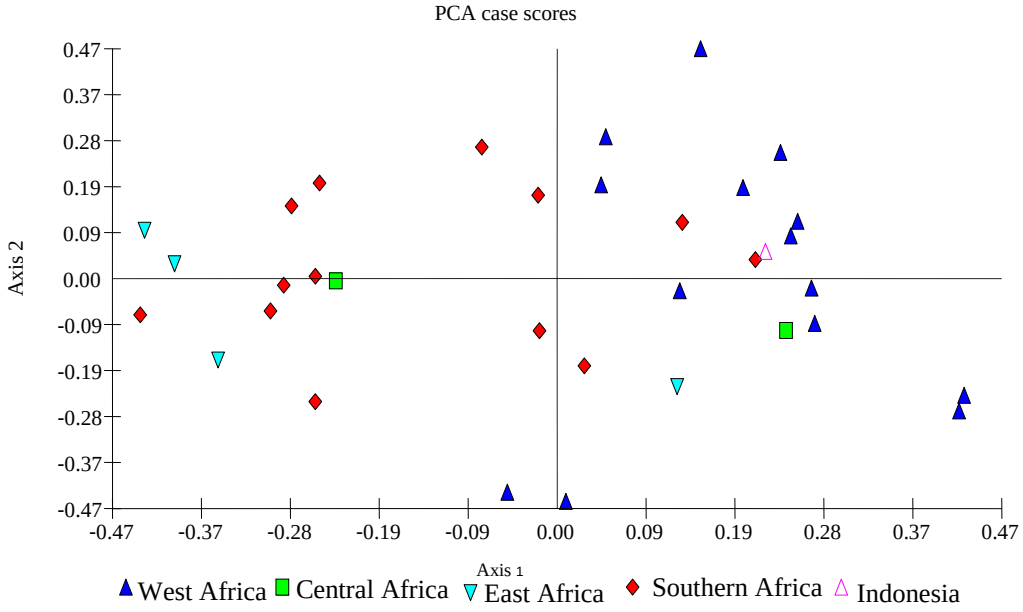


Figure 6.3.1: The first two axes of the PCO case scores, generated from the 34 bambara groundnut landraces using MVSP for figure 6.3.1 (a) SSR marker Axis 1 represent 9.90 % and Axis 2 represent 8.35 %, figure 6.3.1 (b) Morphology marker; Axis 1 represent 13.56 % and Axis 2 represent 8.80 % molecular variation with a cumulative % of 18.25 % and 22.36 % respectively.

The principal component analysis for the SSR markers explains 18.25 % of the variation in the 34 bambara groundnut lines for the first two axes. The genotypes from Southern Africa and West Africa were clearly defined into two distinct groups (Figure 6.3.1a). There are some few exceptions of 119Hybrid and two Namibian lines 99SB4 4 and 100SB 16 A which grouped with the West African lines and 40Acc536 from Cameroon which grouped with the Southern African lines. The West African lines on the left panel of the diagram also contain two of the lines from East Africa and two from Central Africa. Two East African lines and the Indonesian line are grouped with the Southern African lines which are scattered in the right pane of the graph in both the lower and upper panel.

Similarly the PCO score for the agronomic data separated the 34 bambara groundnut lines into two major groups as in the SSR marker PCO score data. The principle component explained a cumulative percentage for morphological variation of22.36 % among the lines for the first two axes (Figure 6.3.1 b). There is a demarcation between the Southern African lines from the West African lines with the only line 40Acc536 from Cameroon which is morphologically similar to the Southern African lines. The Southern African lines which grouped with the West African lines were 119 Hybrid, 92AHM968NAM, and 91UniswaRed SWA which generally produced narrower leaves with lower *leaf width* and lower *leaf area* observed (Chapter 4 results) which suggests these lines could do well in the West African environment with a relatively higher amount of rainfall compared to a Botswanan environment.

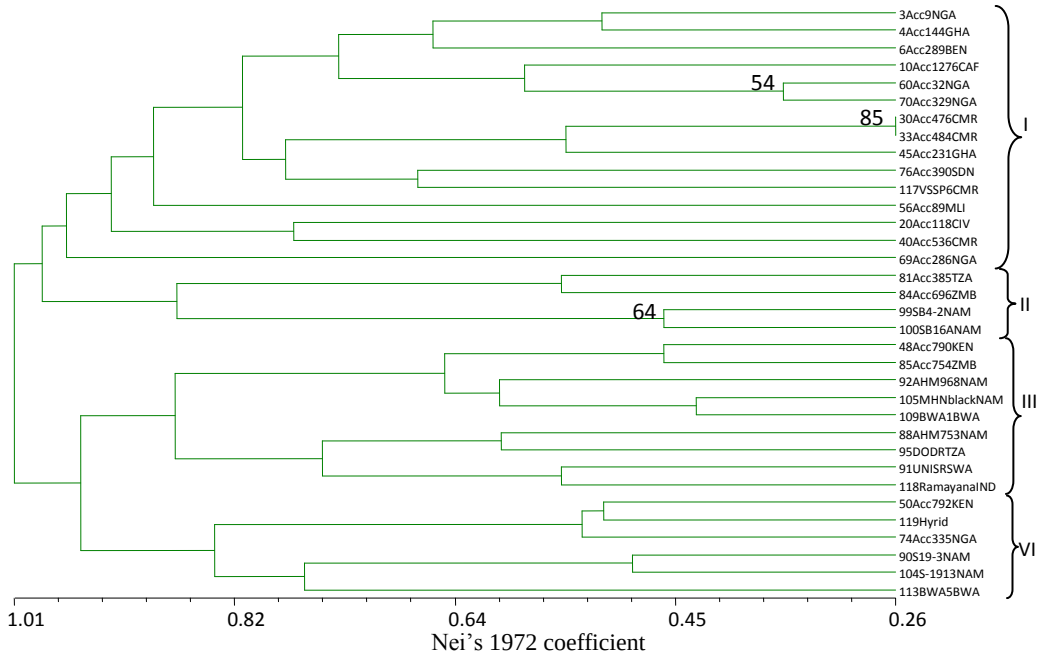
Cluster analysis using the UPGMA method based on the Nei's 1972, clustered the 34 bambara groundnut into four groups (figure 6.4.1a). Cluster one consists of a total of 15 lines, 13 are from West Africa while two 76Acc390 from Sudan and 10Acc1276 from Central African Republic are from Central Africa. Cluster two consists of lines from Southern Africa except 81Acc385 from Tanzania. The third and fourth clusters consists mostly of lines from Southern African, there are also mixed up with lines from East Africa and Indonesia with an exception of line 74Acc335 from Nigeria. This observation is largely in agreement with the PCO.

The Euclidean distance cluster analysis method based on the 37 agro-morphological markers grouped the 34 landraces into three groups (figure 6.4.1b).

The landraces are mainly separated based on characters which contribute more variation in bambara groundnut such as *shoot dry weight*, *pod number*, *plant height*, *seed number* and *canopy width* (Chapter four). Cluster 1 consists of lines mostly from West Africa with a mixture of Southern Africa lines. These lines performed poorly in terms of *pod yield per plants* and lines 70Acc329 from Nigeria, 50Acc792 from Kenya and 45Acc23 from Ghana, producing no yield at all. Cluster 2 is a mixture of lines from Southern Africa, West Africa and 10Acc1276 from Central African Republic from Central Africa, these lines had a higher *number of stems per plant,* relatively similar *plant height* and *shoot biomass*. Cluster 3 consists of 10 lines which performed relatively well in number of characters in Southern Africa (Botswana) environment and had lower *petiole-internode ratio* and high yield such as 88-AHM753 and 90-S19-3 from Namibia, 84Acc696 from Zambia, 81Acc385 from Tanzania and 109Bots1 from Botswana which produced highest number of *pods per plant*.

Morphological markers showed that they could to some extent separate landraces based on their areas of origin, which does reveal the importance of area of origin on the selection of bambara groundnut. There are some striking similarities between the SSR marker and the morphological marker cluster analysis; there was largely a clear demarcation between the Southern African landraces and the West African landraces.

a) SSR marker dendrogram



b) Morphological marker dendrogram

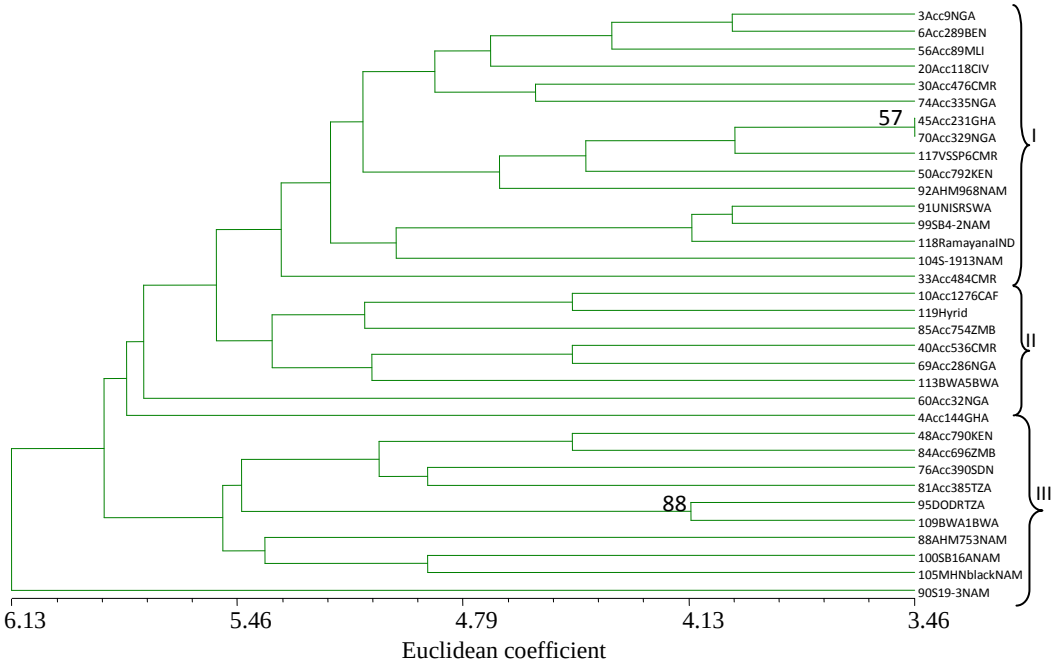


Figure 6.4.1: Cluster analysis of 34 bambara groundnut analysis with Unweighted pair group method with arithmetic method (UPGMA) were generated using NTSYS version 2.1, Figure 6.4.1 (a) is SSR marker dendrogram generated from 20 microsatellites markers based on Nei's 1972 distance estimates, Figure (b) is a morphology dendrogram generated on 37 morpho-agronomic traits generated on Euclidean distance estimates.

6.3.4 Genetic distance estimates between landraces

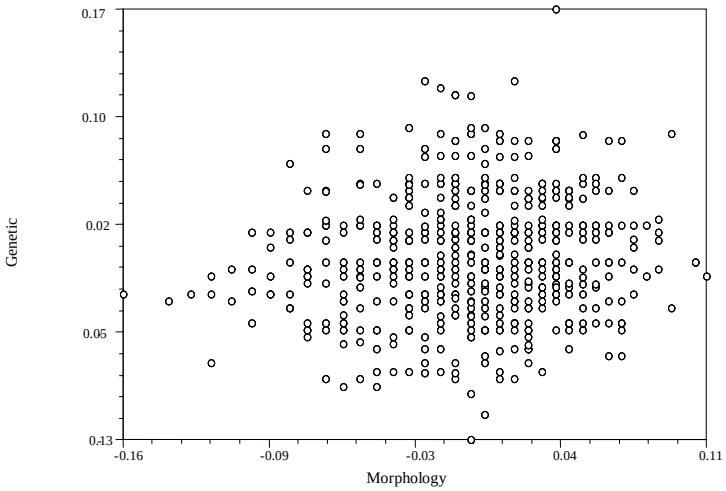
The genetic distance estimates from SSR marker were calculated using the Nei's 1972 coefficient on Popgene version 1.31. The coefficient ranged from 0.262 to 1.846. The lowest genetic distance was between 30Acc476 and 33Acc48 both from Cameroon while the highest genetic distance was found between 69Acc286 from Nigeria and 95DodRed from Tanzania.

The lowest genetic distance based on Euclidean was between 45-Acc231 from Ghana and 70-Acc329 from Nigeria at 12.00, and the highest genetic distance morphologically was between landraces 10Acc1276 from Central African Republic and 95DodRed from Tanzania at 49.00.

6.3.5 Correlation between molecular and morphological distance estimates

In carrying out a comparison of the distance estimates between the two marker types, a correlation between distance estimates matrices established by using Nei's1972 coefficient for SSR markers and Euclidean for morpho-agronomic traits was made using both the Mantel test, Pearson product-moment correlation coefficient and Spearman (rank) correlations coefficient. Correlation analysis was conducted for the 35 genotypes that were analysed with 20 SSR markers in the agronomy bay experiment, and 34 lines that were planted in the field and among the best 5 lines that were planted in the growth room experiment. A detailed chronology of how the experiment was conducted is given below in figure 6.5.2. Highly significant but low correlations were recorded in the agronomy bay ($r = 0.139$; $P < 0.006$), in the field experiment ($r = 0.122$; $P < 0.001$) while a relatively higher correlation ($r = 0.612$; $P < 0.001$), was observed in the controlled growth room experiment based on Mantel test (Table 6.6).

A)



B)

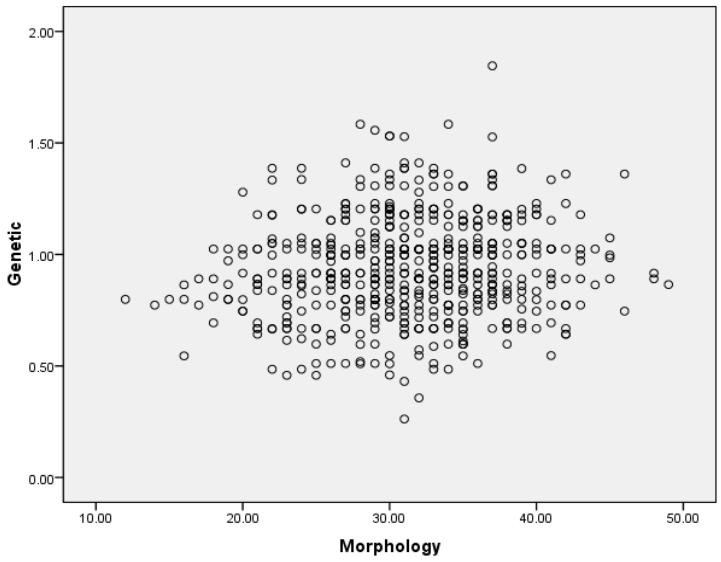


Figure 6.5.1: A scatter plot of correlation for morpho-agronomic and molecular marker based on Pearson, Spearman (rank) and Mantel test, analysis conducted on (A) NTSYS pc version 2.1 and (B) on SPSS version 16, the morphological markers were based on Euclidean distances estimates while the molecular marker were on Nei's 1972 coefficient.

Table 6.6: Correlation between morpho-agronomic markers and molecular markers for the 35 and 34 bambara groundnut genotypes based on 20 microsatellites and 37 morph-agronomic characters, and for 5 lines based on 12 markers and 22 morpho-agronomic characters.

a) Glasshouse	Pearson correlation	Spearman rank correlation	Mantel test
Marker	Morphology	Morphology	Morphology
SSR	0.767	0.771	0.139
Morphology	1	1	1
N	595	595	595
P-value	0.001	0.001	0.006
b) Field experiment	Morphology	Morphology	Morphology
SSR	0.112	0.105	0.122
Morphology	1	1	1
N	561	561	561
P-value	0.008	0.013	0.030
c) Growth room experiment	Morphology	Morphology	Morphology
SSR	0.665	0.461	0.612
Morphology	1	1	1
N	435	435	435
P-value	0.001	0.001	0.001

N = number of values in the matrix

6.3.6 Molecular variance among bambara groundnut landraces.

The partitioning of population diversity within and between populations were analysed on Analysis of Molecular Variance (AMOVA) based on the two groups (Figure 6.2.1). Group one consists of all genotypes from West Africa, Central Africa except one, and four genotypes from East Africa while group two consists of all genotypes from Southern Africa, all genotypes from Indonesia and most of the genotypes from East Africa. AMOVA revealed that most of the variation resides among individuals within populations (87.30 %; P <0.001), there is significant variation of 11.58% that exists among the two groups, while only 1.12 % is within individual genotypes (Table 6.7). Similar observations were made in the PCoA figure 6.2.1a and b, where most of the differentiation among genotypes was observed.

Table 6.7 Analysis of molecular variance (AMOVA) for the 105 bambara groundnut genotypes for the comparison based on the five selected regions, analysis conducted using Arlequin version 3.5

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation	P-value
Among populations	1	88.626	0.738Va	11.58	<0.001
Among individuals within populations	103	1152.85	5.560Vb	87.3	<0.001
Within individuals	105	7.5	0.017Vc	1.12	<0.001
Total	209				

6.3.7 Breeding strategy

From the 119 bambara groundnut accessions that were planted in the agronomy bay in 2008 season, three individual genotypes from each accession were selected from the 34 landraces that were analysed with a set of 20 microsatellites markers and this made up the first selection (Figure 6.5.2 a, b). Field work was then conducted on the 34 lines derived from seed from single plants selected from the previous year's experiment and planted in the field at Botswana College of Agriculture, (Botswana) and this made up the second cycle of selection for bambara groundnut lines (Figure 6.5.2 c). The third cycle of selection was conducted on the five best lines that were selected from a field experiment in Botswana, selected after a ranking analysis. A growth room experiment was conducted for characterisation, evaluation and genetic analysis of these set of lines. Five individual genotypes from these lines were analysed with a set of 12 markers (Figure 6.5.2 d).

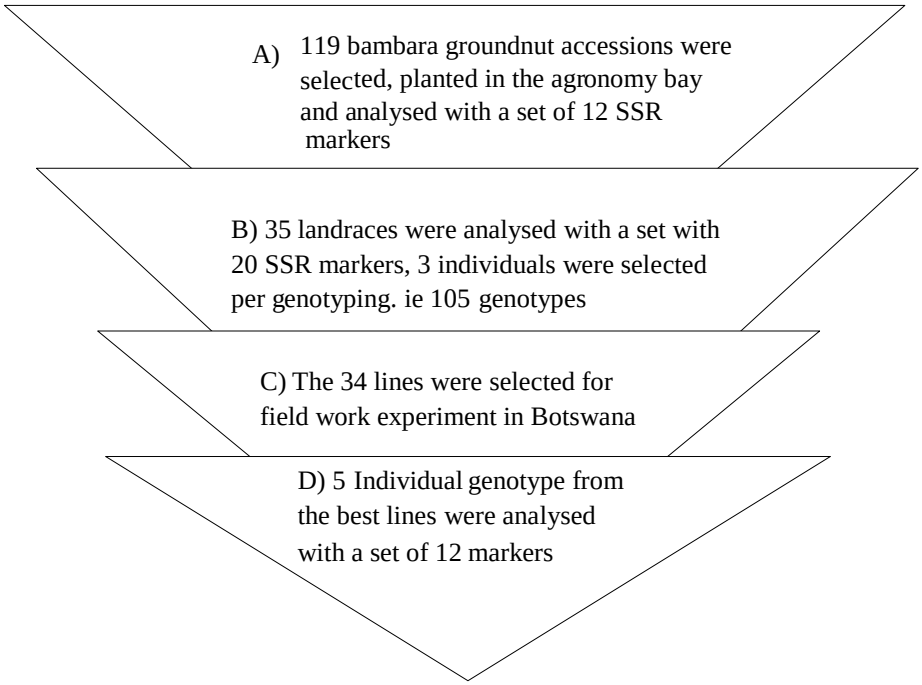


Figure 6.5.2: Schematic diagram showing the selection strategy for the three round of selection of bambara groundnut

The five landraces, 81-Acc385 from Tanzania, 84-Acc696 from Zambia, 88-AHM753 and 90S19-3 from Namibia and 109BWA1 from Botswana were followed through for three generations from the agronomy bay experiment, field experiment in Botswana and control growth room experiment. Twelve microsatellites markers were employed in the molecular analysis of the five landraces in the first season of selection and revealed an average genetic distance of 0.404 with a range of 0.222 for 88-AHM753 from Namibia to 0.751 for 81-Acc385 from Tanzania based on Nei, 1972 genetic distance estimates on Popgene (Table 6.8).

Table 6.8: Mean and range of the genetic distances values for three different selection cycles of bambara groundnut from single seed descent estimated based on 12 microsatellites markers using Popgene version1.31 (Yeh and Boyle, 1997).

	Genetic distance estimates											
	First cycle selection				Seecond cycle selection				Third cycle selection			
Selected lines	N	Mean	Ho-He	Range	N	Mean	Ho-He	Range	N	Mean	Ho-He	Range
81-Acc385TZA	3	0.751	0.000-0.356	0.287-1.049	7	0.000	0.000	0.000	6	0.000	0.000	0.000
84-Acc696ZMB	3	0.314	0.000-0.222	0.206-0.403	4	0.000	0.000	0.000	6	0.000	0.000	0.000
88-AHM753NAM	3	0.222	0.000-0.267	0.198-0.248	7	0.000	0.000	0.000	6	0.000	0.000	0.000
90-S19-3NAM	3	0.347	0.000-0.311	0.305-0.405	7	0.000	0.000	0.000	6	0.000	0.000	0.000
109-BWA1-BWA	3	0.389	0.000-0.311	0.331-0.431	7	0.000	0.000	0.000	6	0.000	0.000	0.000

N = Number of individual sample

Variability within bambara groundnut landraces has been reported before, and has been attributed to a range of causes, from low levels of outcrossing, followed by the natural development of inbred lines, through to the mixing of seeds during harvesting or from markets, especially those of same colour (Massawe *et al.,* 2005; Mayes *et al.,* 2009). As a breeding strategy for inbreeding crops like bambara groundnut, it is advantageous to obtain pure homozygous lines with good attributes. As expected, in the second and third round of selection pure lines were selected through single plants (Table 6.8). There was no observed or expected heterozygosity in the second and third round of selection. This data strongly suggests these genotypes are now essentially pure lines or effectively varieties. The data for the five lines are listed in (Appendix 11).

6.4 Discussion

The twenty selected microsatellites markers showed polymorphism among the set of 105 genotypes of the 35 bambara groundnut landraces. To determine the level of polymorphism in markers two distinct quantities can be used, that is heterozygosity and polymorphism information content (Shete *et al.,* 2000). Markers with polymorphic information content above 0.5 are considered highly informative (Botstein *et al.,* 1980). The average polymorphic information content found among the 20 selected markers of 0.67 shows that most of the markers were highly informative in this material. Approximately 75 % of the markers (15) had polymorphic information content more than 0.5 while only 15% (5) markers, marker D8, marker E7, marker 23, marker 21 and maker 30 had lower PIC values less 0.5.

Bambara groundnut is a self-pollinating crop, so it is not surprising to have shown an inbreeding coefficient close to 1(Table 6.1). Similar findings in bambara groundnut were observed by Basu *et al*., (2007). They found an inbreeding coefficient 1, among the 8 markers they studied except for two which had heterozygosity (*Ho*) of 0.28 and 0.06 against an expected heterozygosity *(He*) of 0.82 and 0.78, respectively. In other related crops, such as pigeonpea, Kuroda *et al.*,(2006) in Japan used 20 microsatellites to study the genetic diversity of 616 individuals of 77 wild soybean (*Glycine soja*) and 53 varieties of cultivated soybean (*Glycine max*). They recorded the expected heterozygosity (*He*) for wild soybean of 0.870 and 0.496 for cultivated accessions, and an observed heterozygosity (*Ho*) of 0.000 for cultivated and 0.018 for wild accessions, suggesting that both accessions are predominantly inbreeding. In common bean (*Phaseolus vulgaris*), Blair e*t al.,* (2009) observed an expected heterozygosity of 0.64 and observed heterozygosity of 0.049 among 604 genotypes analysed using a set of 36 SSR markers.

To quantify the genetic diversity among the selected bambara groundnut accessions both molecular analysis and agro-morphological data were analysed by cluster analysis using the UPGMA method and principal coordinates analysis.

For the cluster analysis of the 105 genotypes 43 nodes had bootstrap values of more than 50%, which could indicate that higher number of markers may be

requested to increase the robustness of the cluster analysis. Twenty nodes were supporting three individuals within a genotype which also indicate that some accessions are highly heterogeneous. The results were coherent with the origin or history of the landraces; in addition all genotypes were uniquely identified, with highest similarity among genotypes observed among landraces from the same country.

For the 34 bambara groundnut lines, the molecular marker cluster analysis on Nei's 1972 distance estimates and morphological markers based on Euclidean distances were compared. Molecular markers grouped landraces into four clusters while morphological markers grouped them into three clusters. Basically there were some similarities between the clusters produced, with most landraces grouping based on their areas of origin. Even when there were some mixtures of landraces within a cluster, most of the landraces found in that particular cluster had a common area of origin. Similar findings in bambara groundnut have been observed when using RAPDs by (Massawe *et al.,* 2003) and AFLP by Massawe *et al.,* (2002), where they showed that landraces were clearly grouped based on their areas of origin. This is an indication of the importance of adaptation on the genetic variation in bambara groundnut. A similar pattern of observation was also seen in the principle coordinate analysis.

One of the best options for crop improvement is through the hybridisation of genotypes with reasonable genetic distance and desirable agronomic traits (Parsaeian *et al.,* 2010). The molecular markers identified line 69Acc286 from Nigeria and line 95DodRed from Tanzania to be genetically far apart, while the Euclidean distance estimates identified lines 10Acc1276 from Central African Republic and 95DodRed from Tanzania as dissimilar. However, it is those landraces that have been found to be agronomically superior that could be used in a breeding programme as parents, such as S19-3 from Namibia, 76Acc390 from Sudan and 33Acc484 from Cameroon were some of the lines that produced higher number of pods per plant in Botswana environment.

The genetically distant lines from Southern Africa and West Africa, which had been found to cluster in different groups by both morpho-agronomic and SSR marker type, could significantly lead to an increase in bambara groundnut

performance in terms of yield, if the wide cross leads to good transgressive trait segregation which can be selected from to develop new varieties and possibly better adapted landraces to both environments (regions).

The correlations observed between morphological and genetic distance estimates could suggest that SSR markers could be used as substitute for phenotypic measurement (Ramakrishnan *et al.*, 2004). The highly significant and positive correlation observed for the agronomy bay experiment, field experiment and in the controlled growth room data matrices revealed that the morphological genetic distance could reflect the genetic distance estimates. The positive correlation found, is of significant importance to plant breeders especially for underutilized crops with lack of resources since morphological markers are the standard markers used regularly for crop improvement.

The positive correlation between SSR data marker genetic distance estimates and morphological markers have been observed in soybean at r = 0.31 (Priolli *et al.*, 2010). In bambara groundnut, Ntundu *et al*., (2004) recorded a positive correlation of r = 0.40 between morphological markers and AFLP among 100 bambara groundnut landraces in Tanzania, possibly due to low heterogeneity among the Tanzanian landraces.

The close agreement between molecular markers and SSR markers in the controlled growth room experimentsuggests for genetic diversity studies in bambara groundnut pure lines or varieties, SSR markers and for morphological markers were in good correspondence. The observation where there is a simultaneous increase in the phenotypic distance and molecular distance has been noted before and termed the ‘triangular relationship’ observed between furthest points (Burstin and Charcosset, 1997). This phenomenon was only observed in the correlation conducted in the growth room experiment (Appendix 10.2a).

Molecular markers proved to be more robust and reliable in genetic diversity analysis as shown in both cluster and PCO analysis where landraces were clearly defined (Figure 6.1.1 and Figure 6.2.1). A combination of molecular markersand morphological markers as revealed by high Pearson and Spearman correlations in the agronomy bay and in the controlled experiment shows that they could be used

to compliment the use of morphological markers in identifying landraces with desirable characters (Table 6.6 and Appendix 10).

AMOVA was conducted to investigate variation between two groups, group one consists of mainly countries from West Africa and Central Africa while group two consists mainly of countries from Southern Africa, Indonesia and East Africa. AMOVA revealed that most of the variation was among individual populations, and still some significant variation exists among the two groups. This implies that differentiation is skewed to more variation among genotypes. In other self-fertilising leguminous species higher within population variations were reported. Okori *et al.,* (2005) identified 92.16% within population and 7.82% among populations using AFLP in 41 pigeonpea (*Cajanus cajan*) landraces.

A number of factors have been attributed to the genetic diversity in bambara groundnut. There was a higher within landrace variation observed in this study, which is most likely due to the fact that bambara groundnut basically exists as 'inbred lines' due to its highly inbreeding nature. Farmers or consumers in different regions prefer specific landraces due to colour type and taste and various traditional beliefs differ in different regions of Africa (Sesay *et al.,* 2003).

Generally bambara groundnut is planted by small scale farmers with small hectares and mostly for family consumption, with a little for sale. Thus the exchange of seeds and their movement to other countries may not be as pronounced. Similar observation in bambara groundnut by Massawe *et al.,*(2003), when using RAPDs identified a significantly higher among landrace 71.25% compared to 28.67% to difference within individual landraces.

Analysis of landraces based on intra-landrace diversity identified 20% (7) landraces with at least one genotype different among the three selected genotype per landrace. The more diverse lines had a recording of more than 1.5 alleles per locus (Table 6.2).This character of bambara groundnut has implications in pure line selection for variety development (Basu *et al.,* 2007).

In the selection of pure lines of bambara groundnut 12 microsatellites were employed among five landraces. In the initial genetic diversity analysis of the landraces in the first cycle of selection the lowest residual heterogeneity were

recorded in lines 84Acc696 (0.222) from Zambia and the highestin81Acc385 from Tanzania (Table 6.8). However, when Stadler, (2009) investigated multiple samples of six bambara groundnut using DArT markers, S19-3 was identified as the genetically narrowest landrace. LandraceAHM753 has been used in previous bambara groundnut projects as core landrace, it has undergone a number of selections and thus it is relatively pure compared to other lines.

109-BWA1from Botswana, was selected from among four landraces from farmers and characterised in field experiments based mainly on seed colour, leaf morphology stability and in other seasons it was selected for grain yield and days to maturity (Chui *et al.*, 2003). The seeds were generally bulked and no pure line selection for the landrace was done, hence some relatively higher heterogeneity were discovered

6.5 Conclusions

In this study the extent of genetic diversity within and among 35 bambara groundnut accessions from wide geographic range has been investigated. The application of cluster and PCO analysis revealed that bambara groundnut individuals mainly grouped based on their area of origin. The genetic distance for both marker types could be used to identify those landraces that are genetically distant from each other, and there was a good correspondence between the two techniques.

Genetic variation in bambara groundnut is significantly higher within individual landraces compared to among populations, thus a number of landraces could be identified which are relatively pure for use in the selection as pure lines in bambara groundnut breeding.

CHAPTER SEVEN: General discussions

7.1 Introduction

Although bambara groundnut is cultivated throughout tropical regions of Africa, and is very important as a food security crop in the sub-Saharan Africa no concerted effort has been put in place to develop it into improved landraces or varieties. Farmers rely on landraces which can be inherently low yielding due to poor physical and genetic quality of seeds, low germination rates and poor crop management among other constraints. For a comprehensive breeding effort in bambara groundnut it is necessary that there are available molecular markers to use in the study of the genetics of bambara groundnut, even if only for quality control within the breeding programme. Ideally, there must be an understanding of the genetic diversity and population structure in order to be able to identify genotypes that could merit further selection and use as varieties and parental materials. A study of the morphology and genetic diversity of bambara groundnut would assist in identifying the best method for selecting bambara groundnut for breeding purposes and to identify accessions for further selection.

In this chapter, the aims of the research work are presented again, before discussion of the progress and problems encountered. The main areas that this research investigated are:-

I The development of microsatellite markers and their characterisation; these markers represent an additional tool for use in the genetic analysis of bambara groundnut, in mating systems studies, genetic diversity studies, population structure analysis for breeding and for conservation of the crop.

II Phenotypic diversity and morphological evaluation were conducted to assess the extent of phenotypic variability and to indicate the genetic advance possible with the aim of selecting landraces suitable to be grown in the semi-arid environment of Botswana

III The application of morphological markers is the standard and most frequently used tool in genetic diversity studies in underutilised species, including

bambara groundnut. For bambara groundnut, the application of molecular markers is not at an advanced stage and due to technical and costs implications their use is not yet widespread. Therefore, the comparison of the two techniques could bring an insight as to which is the most suitable to use and in what circumstances. The study aimed to investigate the genetic diversity and relationships among the selected landraces using morphological and SSR markers and also to assess any correlation between distance estimates based on morphological and molecular markers.

IV For a well-defined breeding programme a thorough knowledge of the population structure of the landraces needs to be understood. The movement of bambara groundnut landraces from the area of origin to other regions, such as Southern Africa, East Africa, Central Africa and even Indonesia is likely to have an impact on population differentiation. The informal movement of seeds material between farmers within the same countries, between countries and even across regions also affects the population structure of the crop. However, little information was known regarding the relationship of the landraces between regions and countries.

Brief description of the chapter and the main AIM of the experiment

One of the main aims of this study was comparing the utility of SSR and morphology techniques for genetic diversity analysis of bambara groundnut landraces and to establish the relationship between the two approaches, if one exists. The choice to use either morphological markers or molecular markers or even a combination of the two markers is an important consideration. In practice, it is often practically difficult and restrictively expensive to use molecular markers within a developing world context, unless there is a substantial gain to be made in breeding. The importance of a comparison of different marker systems is to assist in making informed decisions as to which marker is best to use in germplasm characterisation and plant breeding. However, the development of an understanding of breeding systems and germplasm population structures could allow a more focused breeding effort, even without further application of molecular markers. Potentially, markers could aid the selection of germplasm for breeding, quality control within breeding programmes and, potentially direct

selection via Marker Assisted Selection (MAS). In this study a number of experiments were conducted to make this evaluation possible.

7.2 Recap of the study

In the first parts of this thesis (Chapter 2 and 3) a total set of 75 microsatellites were characterised and used to investigate the genetic diversity of a set of 24 bambara groundnut landrace accessions. The markers were checked for the presence of null alleles, stutter-bands, spectral overlap and binning was also conducted as a precautionary measure, to identify suitable markers for further use (Chapter 2). A set of 68 markers were found to be polymorphic and produced robust amplification and consistent results. The markers were also compared to a DArT marker dataset that was previously generated from the same 24 genotypes (Chapter 3).

In the second part of the thesis (Chapter 4), phenotypic and morpho-agronomic diversity studies were undertaken on selected landrace accessions planted in the agronomy bay (greenhouse) in the UK, with a subset later taken for field studies in Botswana. Several analyses were undertaken to investigate the phenotypic diversity of the landraces, such as through the generation of Shannon weaver diversity indices, principal component analysis, cluster analysis and Pearson coefficient correlation studies.

The third part of the study (Chapter 5) determined the population structure of bambara groundnut in five regions, four from Africa and one from Asia, using a set of 12 pre-selected microsatellites developed and characterised in this study (Chapter 2 and 3).

The fourth part of the study (Chapter 6) assessed the use of both molecular and morphological markers in genetic diversity analysis of bambara groundnut. Morphological markers characterised in the Agronomy bay, in the field experiment and controlled growth room experiment were compared with molecular marker analysis from each respective experiment.

7.3 Microsatellites development and characterisation

A set of 68 markers were found to be polymorphic and had consistent amplification results. These markers represent a new tool for use in the genetic analysis of bambara groundnut, in mating system studies and for genetic diversity studies of population structure for breeding and conservation purpose of the crop.

The set of 24 genotypes were used to compare SSR and DArT markers for an assessment of overall bambara groundnut genetic diversity. The two techniques have proved useful in the genetic diversity analysis of the selected material. SSR markers showed slightly higher genetic differentiation between the landraces with lower similarity coefficients at an average of 0.65 compared to 0.71 for DArT markers. Similar findings were observed in cassava (*Manihot escuenta*) by Hurdato *et al.,* (2008) when comparing the utility of 1000 DArT markers and 36 SSR markers to assess 436 cassava accessions, 155 originally from Africa and 281 from Latin America.

In this study DArT markers appear to provide clearer genetic resolution when compared to geographical accession origins, as compared to SSR markers. DArT marker explained a higher percentage of the molecular variation in the first two PCO axes, show a wider range of genetic distance estimates between genotypes and more clearly defined grouping of individuals through both PCO and UPGMA analysis.

A highly significant correlation was found using the Pearson correlation coefficient (r =0.34), Spearman rank correlation (r = 0.34) and the Mantel test (r = 0.35) between the two techniques. This indicates a reasonable correspondence between the genetic distance estimates for the two markers types. Genetic distance estimates between genotypes are useful in the selection of crosses that could generate the best performing lines (Bertan *et al.,* 2010); therefore both marker types can be useful in the crop improvement of bambara groundnut.

Most of the comparisons between SSR markers and other markers suggest that SSR markers usually reveal a higher level of polymorphism while, other markers (for instance AFLP) have a higher marker index (requires less number of primer combinations to screen whole genome) (Spooner*et al.,* 2005). DArT markers are

relatively new markers so comparative data with other marker types is quite scanty in the literature and in bambara groundnut this comparison had not yet been conducted. Since both markers effectively differentiated the landraces they have proved to be useful in genetic diversity studies of bambara groundnut. The utility of the two marker types could be similar (Hurtado *et al.*, 2008), especially when more samples and a larger number of markers are employed.

Markers developed in this study will be used to investigate the genetic variability that exists in bambara groundnut both within and between landraces, and also to investigate correlations between genetic and morpho-agronomic traits in bambara groundnut.

A summary of achievements made in the chapters 2 and 3:

- Characterisation of 68 SSR markers for bambara groundnut
- Establishment of the relationship between and genetic variation detected by DArT markers and SSR markers in a set of genotypes representative of the available germplasm

7.4 Morphological characterisation

An assessment of morpho-agronomic trait variation of 35 bambara groundnut lines selected from the agronomy bay experiment (glasshouse) and the subset of 34 lines selected in the field experiment was explored using multivariate analysis (i.e cluster and principal component analysis), Shannon weaver diversity indices and Pearson correlation coefficients. These estimates potentially provide useful information that facilitates the exploitation of bambara groundnut germplasm. Estimations of heritability and genetic advance were conducted, to allow the use of selection indices in order to identify the best performing lines.

Multivariate techniques, such as cluster analysis and principal component analysis are efficient tools used in the estimation of quantitative variation in crops (Rahim *et al.,* 2008). These techniques have been explored in common bean to group accessions based on their yield performance (Salehi *et al.,* 2008).

Agro-morphological data for bambara groundnut was standardized to reduce the effect of scalar differences between traits on correlation estimates of morphological traits. Tabachnick and Fidel (2007) argue that conversion of data to standardized scores makes the data independent of scale measurement and sample sizes used. Correlation coefficient analysis measures the association between variables, the Pearson correlation coefficient is one of the most commonly used. It describes the direction and degree to which one variable is linearly related to one another (Bolboacă and Jäntschi, 2006).

In both the agronomy bay and field experiment a number of strong and positive correlations were observed within the vegetative characters such as *petiole length* and *plant height* (r = 0.91) and also within the seed yield related traits and seed *number per plant* and *pod number per plant* (r = 0.98). Similar observations were made by Kobraee *et al.,* (2010), in Chickpea who observed high correlation between *grain yield per plant* and *plant height* (r = 0.827), and a high correlation between *grain yield per plant* and*plant height* at (r =0.813). Farshadfar and Farshadfar, (2008) in chickpea (*Cicer arientinum*) where they observed that their highest correlation were between *seed yield per plant* and *pod number plant* (r = 0.78). In pigeonpea, Vange and Moses (2009) recorded their highest correlation between *grain yield* and *pod dry weight* (r = 0.87). This suggests that correlation

is also dependent on the characters observed. Correlations found between vegetative traits and seed yield were generally low even for those which are statistically highly significant. A higher and significant correlation of r = 0.61 was observed between *pods number per plant* to both *shoot dry weight* and *canopy spread* while a slightly higher and significant correlation r = 0.67 was observed between *shoot dry weight* and *canopy spread* in the agronomy bay experiment. Similar traits recorded lower correlations in the field experiment, *pod number per plant* and *shoot dry weight* had an r = 0.13, while *pod number per plant* and *canopy spread* recorded r = 0.03, *shoot dry weight* and *canopy spread* hadr = 0.30, this observation indicates that selecting for medium canopy spread lines (which are the bunch and the semi-bunch type) would not significantly increase *pod yield per plant* since the variation for this and correlations are very low.

The characters were subjected to Principal component analysis (PCA), using a correlation matrix, to identify characters showing the highest explanation for the major variation in morphological traits given in the first eigenvalues which could be used for characterising bambara groundnut landraces. PCA is an analysis of a matrix consisting of variances and covariances or correlations among variables to come up with smaller sets of components that summarise the correlations (Fenty, 2004). The first component extracted accounts for the maximum amount of total variation observed, subsequently followed by the second principal component (Jollie, 2002).

The PCs are uncorrelated they are orthogonal, therefore are affected by the sizes of correlations, and possibly some of the factors that affect correlation too, such as size of the data sets. The variables which are highly correlated will then tend to be concentrated into one component. Most of the variation was accounted for by the vegetative traits and seed yield traits, in the agronomy bay experiment for PC1 and PC2, respectively. Higher loadings of vegetative traits such as *shoot dry weight, leaf area, plant canopy* and *plant height,* together with *pod number per plant* were observed both in the agronomy bay and field experiment. Ntundu*et al.*, (2006) in a study in 100 bambara groundnut landraces also observed that most of the most of the vegetative traits were loaded in PC 1 while the seeds characters are highly loaded in PC 2, which is an indication that similar variation was observed between these two experiments. Therefore the vegetative, pod and seed

yield characters are some important traits useful for the characterisation of bambara groundnut germplasm.

Landraces were mainly clustered based on the vegetative, pods and seed traits but the vegetative traits had higher loadings such as *shoot dry weight*, *petiole length* and *leaf area*as they appeared in the first principle component (PC1) for both experiments (Table 4.17 and Table 4.1.8). These traits also differed based on regions of origin as shown in the cluster analysis in the agronomy bay experiment (figure 4.2.2). The cluster from the field experiment was slightly different, because landraces were also grouped based on their seed yield performance as the latter dominated the loadings in PC2 in the field experiment. Lower yield (*number of seeds per plant* and *number of pods per plant*) were observed in the field experiment. Overall the principal component and the clusters produced in the two experiments were slightly different, possibly due to the effect of environmental conditions in the field compared to the agronomy bay which reduced the variation expressed in some traits. In addition, the Botswanan material could have reduced phenotypic and genetic variability since they were selected from single plants, while the analysis of three individuals in the Agronomy bay experiment from each landrace almost certainly means that non-identical genotypes were examined.

Knowledge of heritability estimates gives an indication of the expected performance of progenies (Bertoldo *et al.,* 2010) thus it is a useful parameter for selection of desirable traits. Numerous indices are at the breeder's disposal to use for selection, each with different characteristics (Strefeler and Wehner, 1986). In this study a simple selection index (SI) was employed to select the best performing genotypes based on *leaf area*, *shoot dry weight*, *seed number per plant* and *pod number per plant* based on the same genotype from three replications. The genetic advance obtained from the field experiment was used as an economic weight in the SI and this makes the index more robust Chapter 4, (Table 4.2.3). The data set for the genotypes were then ranked in such a way that the best lines were concentrated at the top and the best five performing lines were identified.

Bertoldo *et al.,* (2010) used two selection index methods, the Smith (1936) and Hazel (1943) method and the one developed by Pesek and Baker (1969) in common bean to estimates genetic gain among 23 accessions based on 7

characters. The Smith (1936) and Hazel (1943) is a linear combination of traits of economic importance based on the estimation of correlation between characters. While in Pesek and Baker (1969), the technique replaces the economic weight of traits by those values determined by the breeders. The selection index (SI) used in this study has attributes of both techniques as it is a linear combination and also contains an option of using the economic weight of selected characters of interest for selection. It has been used successfully in the selection of cassava seedlings by Ojulong *et al.,* (2010).

Achievements made in the characterisation and evaluation of bambara groundnut

- *Shoot dry weight, leaf area, number of pods per plant* and *number of seeds per plants* have proved to be useful traits for the selection of bambara groundnut.
- The selection index (SI) and Duncan Multiple Range Test identified 5 lines that can potentially be used as varieties in Botswana. These lines have proved to have potential as they performed well in a Botswana environment and could be useful as new sources of germplasm for Botswana.
- The morphological markers grouped genotypes which appeared to relate to their regions of origin, and this was influenced by traits with higher loading in PCA especially the vegetative traits and seed and pod related characters.

7.5 Genetic diversity based on SSR markers and a comparison with morphological characters

The comparison of 20 SSR markers for genetic diversity with 37 morpho-agronomic characters was conducted in the agronomy bay and in the field experiment, respectively. The genetic distance estimates for the two techniques and the dendrograms produced based on the genetic distance estimates differ. The morphological markers revealed slightly higher genetic distance estimates and greater genetic diversity, while the molecular marker showed relatively smaller genetic distance estimates and lower genetic diversity. The principal component analysis revealed higher explanation of trait diversity in the first two components for morphological markers (33%) as compared to the genetic (SSR) characters (19%). This phenomenon was also observed in groundnut by Krishma *et al.*, (2004). The use of principle component (PCA) and cluster analysis clearly defined the 34 landraces tested based on areas of origin, Chapter 6, section 6.3.3.1 (figure 6.3.1a and figure 6.3.1b) with the major groupings based on a separation between the Southern African and the West African landraces.

However, some landraces were found in the West African cluster but morphologically were similar to the Southern landraces and vice versa. This could be an indication that some landraces are adapted to climatic conditions which occurs in both regions. For example a Sudanese landrace 76Acc390 is ranked number six in terms of the vegetative and seed yield in a Botswanan environment (Table 4.2.5) and thus it is grouped together with some highly ranked lines such as 81Acc385 from Tanzania, 90S-19 from Namibia and 109BWA 1 from Botswana. Selection of bambara groundnut for breeding purposes, based on area of origin could potentially be misleading as some landraces from different regions can be morphologically similar. In these cases the use of molecular analysis could be a better option. For instance, 91UniswaRed from Swaziland and 92AHM968 from Namibia were grouped together with the West African lines in PCoA, figure 6.3.1a but in the molecular PCoA they are ‘correctly’ placed with the Southern African lines. Such lines where there is a clear discordance between morphology and molecular markers could be of interest for breeding for particular environments, but maintaining maximal genetic diversity between parents.

Relatively high and highly significant correlations were detected between the genetic distance estimates derived from SSR markers (Nei's 1972) and Euclidean distances derived from agro-morphological measurements in the agronomy bay and growth room experiment using the Pearson correlation and Spearman's rank correlation. Mantel's test detected a weak but highly significant correlation ($r = 0.139$; $P< 0.006$) in the agronomy bay (glasshouse) experiment, and ($r = 0.122$; $P< 0.03$) in the field experiment.

In the growth room experiment the molecular and morphological markers revealed highly significant correlations that were tested for significance on a two-tailed test. The correlation was high for Pearson correlation $r = 0.665$, and Mantel's test $r = 0.612$ but moderate for Spearman's rank correlation $r = 0.461$.Low but highly significant correlation ($r =0.1$) were recorded in the field experiment. The low correlation could reflect low genomic coverage and high variability of the environment (Cheverud, 1988; Brown, 1997).

The distance estimate methods employed in the correlation tests clearly have some impact on the final analysis. In comparing the two examples of Euclidean distances both defined by Pythagoras theorem like simple matching coefficient and standardized Euclidean measures the two are likely to have a higher correlation. The simple matching coefficient is a distance measure derived from calculating the proportion of disagreements as a categorical measure, but as standardized Euclidean is a transformed variable to have the same variance for distance estimates. As Pearson product-moment correlation coefficient expects to find a linear relationship between these two, the fact that one is non-linear would be expected to reduce the detected correlations. Another effect could be coming from the inherent differences of the similarity coefficient used in genetic distance estimates. Ramakrishnan *et al.,* (2004) also observed that the scale of variation could affect correlation between marker types. The relationship estimated for phenotypic distances and molecular marker distances sometimes produces conflicting results. Some researchers report no correlation while other reports clear correlations (Burstin and Charcosset, 1997)

In order to determine the effect of origin of *Medicago sativa* spp. *falcata* Li *et al.,* (2009), conducted both Pearson product-moment correlation coefficient and a regression analysis on the genetic distance estimates, phenotypic distance and geographic distances. Seven SSR markers and 7 morphological characters were used to compare the two markers and produce an assessment of genetic diversity of the 12 *Medicago sativa* spp. *falcata.* Their study did not reveal any correlation between genetic distance estimates and phenotypic distance estimates, but a high correlation (r = 0.78) was found between the morphological distance with geographic estimates. The lack of correlation between the genetic and the morphological markers was attributed to low numbers of samples and the narrow range of altitudes where the samples were collected (Li *et al.,* 2009).

The low association generally observed between morphological distance estimates and molecular distance estimates has been attributed to a number of possible reasons. It could be because of the application of different number of markers and due to the fact that molecular markers target non-adaptive variation while morphological markers are highly influenced by environment (Veira *et al.,* 2007). Low levels of genetic variation in the germplasm and sometimes errors in recording could also contribute to poor correlation between markers (Karuri *et al.,* 2010; Giancola *et al.,* 2002).

Comparing genetic distance estimates between marker types could reveal how useful and efficient markers are for plant breeding purposes and identify those that are most informative (Franco *et al.,* 2001). Mantel tests are particularly useful in testing genetic distance matrices, as it tests for linear or monotonic independence between distances (Legendre and Fortin, 2010). In a comparison of the two techniques, Legendre and Fortin, (2010) established that Pearson correlation has more power in the analysis of linear relationships between raw data, while Mantel's tests have greater power in the analysis of transformed data for distance estimates.

The statistical tools used for estimating relationships among distance matrix variables (here, Pearson correlation and the Mantel test) have some instances where they give contrasting results, therefore further investigation into these two methods has been recommended (Dutilleul *et al.,* 2000).

In addition these methods assume linear relationships; however some evidence suggests that the relationship between marker distance estimates and phenotypic distance estimates display a triangular shape. The general trend though is that lower molecular marker distances are associated with lower phenotypic distances, but higher molecular marker distances could be associated with either low or high phenotypic distances (Burstin and Charcosset, 1997)

7.5.1 Pure line selection

The lack of co-dominant markers has hampered the formal assessment of heterozygosity within bambara groundnut genotypes and this has hampered the progress in bambara groundnut pure lines selection (Basu *et al.,* 2007) which could be one of the best options for variety development. Pure line selection could be a rapid and effective way for bambara groundnut improvement, especially given that artificial hybridization is difficult in this species (Suwaprasert *et al.,* 2006; Oyiga *et al.*, 2010).

Twelve microsatellites were employed among the samples of the five landraces. This set of lines (81-Acc385TZA, 84-Acc696ZMB, 88-AHM753NAM, 90S-19-3NAM and 109-Bots1-BWA) have previously been identified as the best performers in the field experiment and were assessed on the basis of *pods numbers per plant, seed numbers per plant, shoot dry weight* and *leaf area* (Chapter 4). The initial genetic diversity of the first cycle of selection between three independent seed from each landrace showed a lower average genetic diversity of 0.222 in84-Acc696 from Zambia and the highest (0.356) for 81Acc385 from Tanzania. However, as all the lines showed no signs of residual heterozygosity in the second round of selection; these lines could be pronounced as true varieties.

Achievements made in the comparison of genetic and morphological markers

- Substantial genetic and morphological diversity was observed in the bambara groundnut landraces

- There was concordance in the genetic differentiation of accessions by both the morphological and molecular markers based on PCoA and cluster analysis
- The PCoA and cluster analysis largely grouped landraces based on their areas of origin
- All the statistical measures detected relatively high and highly significant correlations in agronomy bay and growth room but detected low correlation in the field experiment, with the exception of the Mantel test which identified a highly significant but weak correlation in agronomy bay experiment.
- AMOVA revealed that the greatest amount of molecular variation was within landraces as compared toamong regions (identified groups)
- The application of microsatellites to confirm the genetic purity of single pure line selected varieties.

7.6 Population structure analysis

In the present study a set of 12 pre-selected SSR markers were employed to investigate the genetic diversity and population structure among 123 bambara groundnut accessions, sourced from four regions of Africa, and one region from Asia (Indonesia). As expected the highest genetic diversity among landraces was found from West African landraces which is the putative area of origin of bambara groundnut, Chapter 5, section 5.3.2 (Table 5.2). Principle coordinate analysis allowed the separation of landraces, based mainly on their areas of origin, and could identify those genotypes that show 'transfer' between the regions. Four landraces from Southern Africa were clustered with the West Africa lines, and 13 from West Africa showed some grouping with Southern African genotypes. This potentially indicates a low rate of recent movement of landraces between regions, as shown by lower levels of between regions variation in the AMOVA at 12 % Chapter 5, section 5.3.6 (Table 5.6). That much of the variation was between landraces (85%) might suggest continual movement across regions, with perhaps some subsequent selection to give the limited molecular variation between regions. Analysis of Molecular Variance (AMOVA) was employed to partition the 123 bambara groundnut landraces based on variation among two groups, whereby group one consists of (West Africa and Central Africa) regions, while group 2 consists of mainly (Southern Africa, East Africa and Indonesia)regions. The majority of the molecular variation is mainly found within landraces at 85 %. This also indicates that the substantial phenotypic diversity observed in bambara groundnut is also reflected at the genotypic level, an observation not seen as clearly with groundnut and pigeonpea, which display high levels of phenotypic diversity but reasonably low genetic diversity.

Population structure analysis was conducted among the 87 landraces that managed to produce enough pods and seeds to allow reasonable characterisation and these were co-analysed using the chosen set of 12 microsatellites. Genetic distance estimates for the 87 bambara groundnut landraces were compared with those obtained in the analysis of 15 qualitative traits of seed and pods and the two were tested for potential correlations. The qualitative traits are highly heritable and

clearly expressed in all environments (IITA, IPGIR, 2000). Differences in genetic diversity measures between qualitative and SSR marker data were detected, with a highly significant negative correlation, albeit very weak, analysed on Pearson and Spearman correlations, while Mantel's test recorded a low, negative and non-significant correlation, which suggests that the two markers may be giving different estimates of the diversity present. However, in common bean (*Phaseolus vulgaris*), Blair *et al.,* (2009) found some association between SSR marker and seed size characteristics.

Summary of achievements in population structure analysis of bambara groundnut

- Low levels of genetic differentiation between regions were found (based on the two groups), but one to note is the non-significant differentiation between Southern African landraces and the East African landraces, which could be a transition route for materials from other regions of Africa.
- The differentiation of landraces based on their regions shows that landraces could be traced to their regions of origin relatively easily, particularly with tools such as PCA.
- A relatively low genetic differentiation of Asian landraces from African landraces could possibly be a suggestion of a recentintroduction to Indonesia from Africa.

7.7 Impact of the findings and future work

This research is the first SSR-based study of genetic diversity among bambara groundnut from sub-Saharan Africa. Bambara groundnut breeders will benefit from the knowledge generated in this study, the genetic distances/similarity estimates for various landraces from different regions and countries are a potential source of parents for hybridisation breeding of the crop.

The use of morphological and molecular markers provides some complimentary information, especially where morphological markers fail to differentiate some landraces. Sometimes it may also not be feasible to undertake a morphological analysis; therefore the use of molecular markers may be unavoidable.

The use of these techniques is important in crop breeding since markers can be used in the prediction of variability, estimation of heterosis and for selecting the best lines for crosses and these may make breeding more efficient and effective. In this study SSR markers have been found to clearly differentiate landraces based on areas of their origin, which should assist breeders to track genotypes of interest and access genetically diverse, but environmentally matched accessions for breeding.

The techniques that have been used in this study - the application of SSR markers in pure line selection and the simultaneous use of both agro-morphological marker and molecular markers - could potentially be employed in other leguminous crop species. The availability of molecular markers will lead to a further exploitation of bambara groundnut germplasm and more bambara groundnut varieties will be developed and this should contribute in poverty alleviation in Sub-Saharan Africa.

A relatively high correlation between morphological and SSR marker suggests a good congruence between the two techniques, hence SSR markers could be used to infer morphological diversity in bambara groundnut.

7.8 Future work

- Establish the minimum number of markers that would be needed to establish significant correlations between morphological and DNA markers, if such correlation is currently limited by depth of genome coverage.
- To undertake a study on the selection index (SI) used in this study and compare it with others used in the literature, which could be used in the selection of more bambara groundnut varieties.
- To conduct a stability analysis on developed varieties in other parts of Botswana, so as to release them to farmers, with the exception of 109BWA1, this had undergone a number of field trials already.
- Conduct a study on the genetic diversity of landraces based on the climatic zones of Africa which is important to discern the effect of weather parameters on the genetic diversity of bambara groundnut and adaptation to local conditions.
- Conduct a more detailed comparison of genotypes from various regions, which would shed more light on the population structure of this landraces. Increasing the number of landraces per country and including more countries could give a detailed population structure analysis.
- To conduct within and between accession variation with an increased number of individuals per genotype.

REFERENCE

ADENIJI O.T., PETER J.M. and IBRAHIM B. (2008), Variation and interrelationships for pod and seed yield characters in bambara groundnut (*Vigna subterranea*) in Adamawa State, Nigeria, African Journal of Agricultural Research 3: 671 - 621

AKANINWOR J.O. and OKECHUKWU P. (2004), Comparative nutrient and anti-nutrient levels in commercial and formulated weaning mixtures, BIOKEMISTRI 16: 15 -21

AKBARI M., WENZL P., CAIG V., CARLING J., XIA L., YANG S., USZYNSKI G., MOHLER V., LEHMENSIEK A., KUCHEL H., HAYDEN M.J., HOWES N., SHARP P., VAUGHAN P., RATHMELL B., HUTTNER E. and KILIAN A. (2006), Diversityarrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome, Theoretical Applied Genetics 113: 1409 -1420

AKKAYA M.S., BHAGWAT A.A.and CREGAN P.B. (1992), Length polymorphism of simple sequence repeat DNA in soybean, Genetics 132: 1131 -1139

ALLARD R.W. (1960), Principles of plant breeding, John Wiley and sons, New York

AMADOU H.I., BEBELI P.J., and KALTSIKES P.J. (2001), Genetic diversity in bambara groundnut (*Vigna subterranea* L.) germplasm revealed by RAPD markers, Genome 44: 995 -999

AMARTEIFIO J.O., TIBE O. and NJUGO R.M. (2006), The composition of bambara groundnut (*Vigna subterranea* L. Verdc.) grown in Southern Africa, Africa Journal of Biotechnology 5: 2408 - 2411

AMOS W., HOFFMAN I.J., FRODSHAM A., ZHANG L., BEST S. and HILL A.V.S. (2007),Automated binning of microsatellites alleles: problems and solutions,Molecular Ecology Notes 7: 10 - 14

APPA-RAO S., MAZHANI L.M. and ATTERE A.F. (1986), Collecting bambara groundnut in Botswana, Plant Genetic Resources Newsletter 68: 27- 28

AVILA C.M., SILLERO J.C., RUBIALES D., MORENO M.T. and TORRES A.M. (2003), Identification of RAPD markers linked to the Urf-1 gene conferring hypersensitive resistance against rust (*Uromyces viciae-fabae*) in *ViciaFaba* L., Theoretical and Applied Genetics 107: 353-358

AZAM-ALI S.N., AGUILAR-MANJARREZ J. and BANNAYAN-AVVAL M. (2001), A globalmapping system for bambara groundnut production, Food and Agriculture Organization of The United Nations, Rome

AZAM-ALI S.N., SESAY A., KARIKARI S.K., MASSAWE F.J., AGUILAR-MANJARREZ J., BANNAYAN M, and HAMPSON K.J. (2001),

assessing the potential of an underutilized crop- a case study using bambara groundnut, Experimental Agriculture 37: 433 - 472

BAKER D.C. (1987), Arable farming development priorities in the central agricultural regions, Botswana: a farming system analysis, PhD thesis, Michigan States University, USA

BAKER R.J (1986), Selection indices in plant breeding, CRC Press, Boca Raton, Florida, USA

BAKER R.J. (1974), Selection indexes without economic weights for animal breeding, Canadian Journal of Animal Science 54: 1 - 8

BALOLE T.V., KARIKARI S.K., KHONGA E.B., RAMOLEMANA G. and LEGWAILA G. (2003), Effect of earthing on the yield of bambara groundnut (*Vignasubterranea* (L.)Verdc.) landraces grown in Botswana. In. Proceedings of the International bambara groundnut Symposium, Botswana College of Agriculture, Botswana, 8 -12 August 2003 pp153 - 158

BAMB, (2008), Botswana Agricultural Marketing Board, Annual Report, 2008

BASU S., MAYES S., DAVEY M., ROBERTS J.A., AZAM-ALI S.N., MITHEN R. and PASQUET R.S. (2007), Inheritance of 'domestication' traits in bambaragroundnut (*Vigna subterranea* (L.) Verdc.) Euphytica 157: 59-68

BASU S., ROBERTS J.A., AZAM-ALI S.N and MAYES S. (2007), Development of microsatellite markers for bambara groundnut (*Vigna subterranea* L. Verdc.) – an underutilized African legume crop species, MolecularEcology Notes 7: 1326 - 1328

BASU S., ROBERTS J.A., AZAM-ALI S.N. and MAYES S. (2007), Bambara groundnut. In: Genomic mapping and molecular breeding in plants, volume 3, Pulses,sugar and tuber crops(Eds. Kole C.), Springer, New York pp159 - 173

BAYE (2002), Genotypic and phenotypic variability in *Vernonia galamensis* germplasm collected from eastern Ethiopia, The Journal of Agriculture Science 139: 161 - 168

BECHERE E., BELAY G., MITIKU D. and MERKER A. (1996), Phenotypic diversity of tetraploid wheat landraces from northern and north-central regions of Ethiopia, Hereditas 124: 165-172

BEEBE E.S., ROJAS-PIERCE M., YAH X., BLAIR M.W., PEDRAZA F., MAÑOZ F., TOHME J. and LYNCH J.P. (2006), Quantitative trait loci for root architecture traits correlated with phosphorus acquisition in common bean, Crop Science 45: 413 - 423

BERCHIE J.N., SARKODIE –ADDO J., ADU-DAPAAH H., AGYEMANG A., ADDY S. ASSARE E. and DONKOR J. (2010), Yield evaluation of three early maturing bambara groundnut (*Vigna subterranea* L. Verdc.)

Landraces at the CSIR-Crops Research Institute, Fumesua-Kumasi, Ghana, Journal of Agronomy 9: 175 - 179

BERCHIE J.N., SARKODIE-ADDO J., ADU-DAPAAH H., AGYEMANG A., ADDY S., ASARE E. and DONKOR J. (2010), Yield evaluation of the three early maturing bambara groundnut (*Vigna subterranea* (L.) Verdc.) landraces at theCSIR-Crops Research Institute, Fumesua-Kumasi, Ghana, Journal of Agronomy 9: 111 - 118

BERGMANN F. and RUETZ W. (1991), Isozyme genetic variation and heterozygosityin random tree samples and selected orchard clones from the same Norway spruce populations, Forest Ecology and Management 46: 39 - 49

BERTAN I., DE CARVALHO F.I.F., and DE OLIVEIRA C.A. (2010), Parental selection strategies in plant breeding programs, Journal of Crop Science 4: 211 - 222

BERTOLDO G.J., BARILI D.L., DO VALE M.N., COIMBRA M.L.J., STÄHELIN D. and GUIDOLIN F. (2010), Genetic gain in agronomic traits of common bean inthe region of Planalto Catarinese, Euphytica 171: 381 - 388

BLAIR W.M., DıAZ M.L., BUENDıA F.H. and DUQUE M.C. (2009), Genetic diversity, seed size associations and population structure of common beans (*Phaseolus vulgaris* L.), Theoretical and Applied Genetics 119: 955 - 972

BLAIR W.M., HURTADO N., CHAVARRO C.M., MUŇOZ-TORRES C.M., GIRALDO M.E.,PEDRAZA F., TOMKIMS J. and WING R. (2011), Gene-based SSR markers for common bean (*Phaseolus vulgaris* L.) derived from root and leaf tissues ESTs: an integration of theBMC series, BMC Plant Biology 11: 50

BOLBOACĂ S.D. and JÄNTSCHI L. (2006), Pearson versus Spearman, Kendall's Tau correlation analysis on structure-activity relationship of biologic active compounds, Leornado Journal of Sciences 9: 179 - 200

BONIN A., BELLEMAIN E., EIDENSEN P.B., POMPANON F., BROCHMANN C. and TABERLET P.(2004), How to track and assess genotyping errors in population genetic studies, Molecular Ecology 13: 3261 - 3273

BONIN A., EHRICH D. and MANEL S. (2007), Statistical analysis of fragment length amplified length polymorphism data: a tool box for molecular ecologistsand evolutionists, Molecular Ecology 16: 3737 - 3758

BOTSTEIN D., WHITE R.L., SKOLNICK M.and DAVIS R.W. (1980), Construction ofgenetic linkage map in man using fragment length polymorphism, American of Journal Human genetics 32: 314 -331

BRIÑEZ B., BLAIR M., KILIAN A., CARBONELL S., CHIARATO A. and RUBIANO L. (2011), A whole genome DArT assay to assess germplasm collection diversity in common bean, Molecular Breeding 7: 1-13

BRINK M. (1997), Rates of progress towards flowering and podding in bambara groundnut (*Vigna subterranea*) as a function of temperature and photoperiod, Annals of Botany 80: 505 - 513

BRINK M., SIBUGA K.P., TARIMO A.J.P. and RAMOLEMANA G.M. (2000), Quantifying photothermal influences on reproductive development in bambara groundnut (*Vigna subterranea*): models and their validation, Field Crops Research 66: 1-14

BRINK M., COLLINSON S.T. and WIGGLESWORTH D.J (1996), Bambara groundnut (*Vigna subterranea*) cultivation in Botswana, Report of farmer's survey, pp. 1-52

BRINK M., COLLINSON S.T.andWIGGLESWORTH D.J. (1996), Characteristics of bambara groundnut cultivation in Botswana, In: Proceedings of the International Bambara groundnut Symposium, 23 - 25 July, University of Nottingham, UK pp. 133-142

BROWN R.P. (1997), Testing heterogeneity among phenotypic correlations: a comparison of methods using Monte Carlo simulations, Genetica 101: 67 - 74

BURNS M.J., EDWARDS K.J., NEWBURY H.J., FORD-LLOYD B.V. and BAGGOTT C.D. (2001), Development of simple sequence repeat (SSR) markers for the assessment of gene flow and genetic diversity in pigeonpea (*Cajanus cajan*), Molecular Ecology Notes 1: 283 - 285

BURSTIN J. and CHARCOSSET A. (1997), Relationship between phenotypic and marker distances: theoretical and experimental investigations, Heredity 79: 477 - 483

BUSO G.S.C., AMARAL Z.P.S., BRONDANI R.P.V and FERREIRA E. M. (2006), Microsatellites markers for the common bean (*Phaseolus vulgaris*), Molecular Ecology Notes 6: 252-254

CAN D.N. and YOSHIDA T. (1999), Genotypic and phenotypic variances and covariance in early maturing grain sorghum in a double cropping, Plant Production Science 2: 67-70

CANNON B.T.,MAY D.G., and JACKSON S.A. (2009), Three sequenced legume genomes and many crop species: rich opportunities for translational genomics, Plant Physiology 151: 970 - 977

CASTRO P., MILLÁN T., GIL J., MÉRIDA J., GARCIA M.L., RUBIO J. and FERNÁDEZ- RAMERO M.D.(2011), Identification of chickpea cultivars by microsatellites markers, Journal of Agricultural Sciences 149: 451-460

CEBALLOS H., PÉREZ J.C., CALLE F., JARAMILLO G., LENIS J.I., MORANTE N. and LÓPES J. (2007), A new evaluation scheme for cassava breeding at CIAT In: CIAT 2007.Cassava research and development in Asia. Exploring new opportunities for an ancient crop, pp.125 – 135 (Eds. Howeler H) CIAT, Colombia

CHARCOSSET A. and MOREAU L. (2004), Use of molecular marker for the development of new cultivars and the evaluation of genetic diversity, Euphytica 137: 81-94

CHARLESWORTH D. (2006), Evolution of plant breeding systems, Current Biology 16: 726 - 735

CHATFIELD C. and COLLLINS J.A. (1980), Introduction to multivariate analysis, Chapman and Hall, London, UK.

CHEN J.J, DUAN T, SINGLE R, MATHER K and THOMSON G.(2005), Hard-Weinberg testing of a single homozygous genotype, Genetics 170: 1439 - 1442

CHEVERUD J.M. (1988), A comparison of genetic and phenotypic correlations, Evolution 42: 958 - 968

CHIJIOKE B.O, IFEANYI M.U. and BLESSING C.A. (2010), Pollen behaviour andfertilisation impairment in bambara groundnut (*Vigna subterranea* (L.) Verdc), Journal of Plant Breeding and Crop Science 2: 12 - 23

CHUI J.N., LUZANI T., NKHORI S. and MBULAWA S.T. (2003), Bambara groundnut selection for various characteristics and response to row spacing and population density, In: Proceedings of the International bambara groundnut Symposium, Botswana College of Agriculture, Botswana 8-12 August, 2003 pp. 297 - 307

CHYBICKI J.I. and BURCZYK J. (2009), Simultaneous estimation of null alleles and inbreeding coefficient, Journal of Heredity 100: 106 -113

COLLARD B.C.Y., JAHUFER M.Z.Z., BROUWER J.B. and PANG E.C.K. (2005), An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts, Euphytica 142: 169 -196

COLLINSON S.T., CLAWSON E.J., AZAM-ALI S.N. and BLACK C.R. (1997), Effects of soil moisture deficits on the water relations of bambra groundnut (*Vigna subterranea* L. Verdc.), Journal of Experimental Botany 48: 877-884

COLLINSON S.T., AZAM-ALI S.N., CHAVULA K.M. and HODSON D.A. (1996), Growth, development and yield of bambara groundnut (*Vigna subterranea* L. Verdc.) in response to soil moisture, Journal of Agricultural Science 126: 307- 318

COLLINSON S.T., BERCHIE J. and AZAM-ALI S.N. (1999), The effect of soil moisture on light interception and the conversion coefficient for three landraces of bambara groundnut (*Vigna subterranea*L.Verdc.), Journal of Agricultural Science 133:151-157

CORNELISSEN E.J.L.R. (2004), Modelling variation in the physiology of bambara groundnut (*Vigna subterranea*L.Verdc.), PhD thesis, Cranfield University at Silsoe.

COUDERT M.J. (1984), Market openings in West Africa for cowpea and bambara groundnut, International Trade Forum 20: 14 - 29

COULIBALY S., PASQUET R.S., PAPA R.and GEPTS P. (2002), AFLP organisation analysis of the phonetic organization and genetic diversity of *Vigna unguiculata* L. Walp. reveals extensive gene flow between wild and domesticated types, Theoretical and Applied Genetics 104: 358 - 366

CROUCH H.K., CROUCH J.K., MADSEN S., VUYLSTEKE D.R. and ORTIZ R. (2000), Comparative analysis of phenotypic and genotypic diversity among plantain landraces (Musa spp., AAB group), Theoretical and Genetics 101: 1056-1065

CROUCH J.H. and ORTIZ R. (2004), Applied genomics in the improvement of crops in Africa, African Journal of Biotechnology 3: 489 - 496

CUC M.L., MACE E.S., CROUCH H.J., QUANG D.V., LONG D.T. and VARSHNEY K.R.(2008), Isolation and characterisation of novel microsatellites markers and their application for diversity assessment in cultivated groundnut (*Arachis hypogaea*), BMC Plant Biology 8: 55

CUI Z., CATER JR. E.M., BURTON J. W. and WELLS R. (2001), Phenotypic diversity of modern Chinese and North American soybean cultivars, Crop Science41: 1954-1967

CUPIC T., TUCAK M., POPOVIC S., BOLARIC S., GRLJUSIC S. and KOZUMPLIK V. (2009), Genetic diversity of pea (*Pisum sativum* L.) genotypes assessed by pedigree, morphological and molecular data, Journal of Food Agriculture and Environment 7: 343-348

DESWARTE J.C. (2001), Variation in the photosynthetic activity within and between three bambara groundnut Msc thesis, The University of Nottingham.

DEWOODY J., NASON J.D. and HIPKINS D.V. (2006), Mitigating scoring errors in microsatellites data from wild populations, Molecular Ecology Notes 6: 951 - 957

DIOUF D. and HILU K.W. (2005), Microsatellites and RAPDs markers to study genetic relationships among cowpea breeding lines and local varieties in Senegal, Genetic Resources and Crop Evolution 52: 1057 - 1067

DOERGE R.W. (2003), Mapping and analysis of quantitative trait loci in experimental populations, Genetics 3: 43 - 52

DOI K., KAGA A., TOMOOKA N. and VAUGHAN D.A. (2002), Molecular phylogeny of genus *Vigna* subgenus Ceratotropis based on rDNA ITS and *atpB –rbcL* intergenic spacer of cpDNA sequences, Genetica 114: 129 -145

DOKU E.V., HAMMONDS T.W. and FRANCIS B.J. (1978), On the composition of Ghanaian bambara groundnuts (*Voandzeia subterranea* (L.) Thours) Tropical Science 20: 263 - 269

DOKU E.V and KARIKARI S.K. (1970), Fruit development in bambara groundnut (*Voandzeia subterranea*), Annals of Botany 34: 951 - 956

DOKU E.V. and KARIKARI S.K. (1971), Bambara groundnut, Economic Botany 25: 255 - 262

DOYLE J.J. and LUCKOW A. (2003), The rest of the iceberg. Legume diversity and evolution in a phylogenetic context, Plant Physiology 131: 900 - 910

DURÁN L.A., BLAIR M. W, GIRALDO M. C., MACCHIAVELLI R., PROPHETE E., NIN J.C.and BEAVER J. S. (2005), Morphological and molecular characterization of common bean landraces and cultivars from Caribbean, Crop Science 45: 1320 - 1328

DURSUN A., HALILOGLU K. and EKINCI M. (2010), Characterization of breeding lines of common bean as revealed by RAPD and relationship with morphological traits, Pakistan Journal of Botany 42: 3839 - 3845

DUTILLEUL P., STOCKWELL D.J., FRIGON D. and LEGENDRE P. (2000), The Mantel test versus Pearson‘s correlation analysis; assessment of the differences forbiological and environmental studies, Journal of Agricultural, Biological and Environmental Statistics 5: 131 - 150

EDJE O.T. and SESAY A. (2003), Effect of seed source on performance and yield of bambara groundnut (*Vigna subterranea*) landraces, In: Promoting the conservation and use of underutilized and neglected crops. 9. Proceedings of the workshop on Conservation and Improvement of bambara groundnut (*Vigna subterranea* (L.) Verdc.) 14-16 November, 1995, Harare, Zimbabwe, pp 101 – 118 (Eds. Heller J., Begemann F. and Mushonga J.) International Plant Genetic Resources Institute, Rome, Italy pp.141 - 152

EDWARDS K.J., BARKER J.H., DALY A., JONES C. and KARP A. (1996), Microsatellites libraries enriched for several microsatellites sequences in plants, Biotechniques 5: 758 - 760

ELLAH M.M. and SINGH A. (2008), Bambara groundnut (*Vigna subterranea* L. Verdc.) yield as influenced by phosphorus and cultivars in the semi-arid savanna of Nigeria, Journal of Plant Sciences 3: 176 - 181

ELLIS J.R. and BURKE J.M. (2007), EST-SSRs as a resource for population genetic analyses, Heredity 99: 125 -132

ENDER M. and KELLY J.D (2005), Identification of QTL associated with white mold resistance in common bean, Crop Science 45: 2482 - 2490

ENGELS M.M.J. (1994), Genetic diversity in Ethiopian barley in relation to altitude,Genetic Resources and Crop Evolution: 41: 67 - 73

ESHGHI R., OJAGHI J. and SALAYEVA S. (2011), Genetic gain through selection indices in hulless barley, International Journal of Agriculture and Biology 13: 191 - 197

EXCOFFIER L., LAVAL G. and SCHNIEDER S. (2005), Arlequin version 3.1: An integrated software package for population genetic data analysis. Evolutionary Bioinformatics Online 1: 47 - 50

EXCOFFIER L. and LISCHER H.E.L. (2010), Arlequin suite version 3.5: A new series of programs to perform population genetics analyses under linux and windows, Molecular Ecology Resources 10: 564 - 567

FALCONER D.S. and MACKAY F.C. (1996), Introduction to quantitative genetics, Longman, England

FALL A.L., BRYNE P.F., JUNG G., COYNE D.P., BRICK M.A. and SWARTZ H.E. (2001), Detecting and mapping a major locus for fusarium wilt resistance in common bean, Crop Science 41: 1494 -1498

FAO, (2011), Food and Agriculture Organisation of the United Nations, FAO Statistical Database and Data set, http://faostat.fao.org/site/291 accessed 1 June 2011

FARSHADFAR M. and FARSHADFAR E. (2008), Genetic variability and path analysis of Chickpea (*Cicer arientinum* L.) landraces and lines, Journal of Applied Sciences 8: 3951-3956

FENTY J. (2004), Analysing distances, The Strata Journal 4: 1- 26

FORNI-MARTINS E.R. (1986), New Chromosome number in the genus *Vigna* Savi (*Leguminosae-Papilionoideae*), Bulletin Nationale Plantentium 56: 129 -133

FRANCO J., CROSSA J., RIBAUT J.M., BETRAN J., WARBURTON M.L. and KHAIRALLAH M. (2001), A method for combining molecular and phenotypic attributes for classifying plant genotypes, Theoretical and Applied Genetics 103: 994 - 952

GAITÁN-SOLÍS E., DUQUE M.C., EDWARDS K.J. and TOHME J. (2002), Microsatellites repeat in common bean (*Phaseolus vulgaris*): Isolation, characterisation, and cross-species amplification in phaseolus, Crop Science 42: 2128 -2136

GALLEGO J.F., PEREZ A.M., NUNEZ Y. and HILDALGO P. (2005), Comparison of RADPs, AFLPs and SSR markers for the genetic analysis of yeast of *Saccharomyces cerevisiae*, Food Microbiology, 561 -568

GARCIA A.A.F., BENCHIMOL L.L., BARBOSA M.M.A., GERALDI O.I., SOUZA JR. L.C. and DE SOUZA A.P. (2004), Comparison of RAPD, RFLP, AFLP and SSR markers for diversity studies in tropical maize inbred lines, Genetics and Molecular Biology 27: 579 - 588

GARCIA R.A.V., RANGEL N.P., BRONDANI C., MARTINS S.W., MELO C.L., CARNEIRO M.S., BORBA T.C.O. and BRONDANI R.P.V. (2011), The characterisation of new set of EST-derived simple sequence repeat

(SSR) markers as a resource for the genetic analysis of *Phaseolus vulgaris*, BMC Genetics12: 41

GHAFOOR A., AHMAD Z., QURESHI A.S. and BASHIR M. (2001), Genetic relationship in *Vigna mungo* (L.) Hepper and *V. radiata* (L.) R. Wilczek based onmorphological traits and SDS-PAGE, Euphytica 123: 367-378

GIANCOLA S., POLTRI M.S., LACABATE S.M. and HOPP H.E. (2002), Feasibility of integration of molecular markers and morphological descriptors in a real case study of a plant variety protection system for soybean, Euphytica, 127: 95 - 113

GILBERT J.E., LEWIS R.V., WILKINSON M.J. and CALIGARI P.D.S. (1999), Developing an appropriate strategy to assess genetic variability in plant germplasm, Theoretical and Applied Genetics 98: 1125 -1131

GIMENES A.M., HOSHINO A.A., BARBOSA V.G.A., PALMIERI A.D. and LOPES R.C.(2007), Characterization and transferability of microsatellite markers of cultivated peanut (*Arachis hypogaea*), BMC Plant Biology7:9

GLASZMANN J.C., KILIAN B., UPADHYAYA H.D. and VARSHNEY R.K. (2010), Accessing genetic diversity for crop improvement, Current Opinion in Plant Biology 13: 167 -173

GOEL S., RIANA S.N. and OGIHARA Y. (2002), Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of nuclearRibosomal DNA in the *Phaseolus-Vigna* complex, Molecular Phylogenetics and Evolution 22: 1-19

GOLI A. E., BEGEMANN F., and NG N.Q. (1995), Characterization and evaluation of IITA's bambara groundnut collection. In: Promoting the conservation and use of underutilized and neglected crops. 9. Proceedings of the workshop on Conservation and Improvement of bambara groundnut (*Vigna subterranea* (L.) Verdc.) 14-16 November, 1995, Harare, Zimbabwe, pp 101 – 118 (Eds. Heller J., Begemann F. and Mushonga J.) International Plant Genetic Resources Institute, Rome, Italy

GOMEZ O.J., BLAIR W.M., FRANKOW-LINDBERG B.E. and GULLBERG U. (2004), Molecular and phenotypic diversity of common bean landraces from Nicaragua, Crop Science 44: 1412 -1418

GONZÁLEZ-CHAVIRA M.M., TORRES-PACHEO I., VILLORDO-PINEDA E. and GUEVARA- GONZALEZ G.R. (2006), DNA markers, Advances in Agricultural and Food Biotechnology 6: 99 - 134

GOUDET J. (2001), FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/software/fstat.html

GREGORIUS R-H. (2010), Linking diversity and differentiation, Diversity 2: 370 - 394

GUPTA P.K. and VARSHNEY R.K. (2000), The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat, Euphtyica 113: 163 - 185

GUPTA P.K., VARSHNEY R.K., SHARMA P.C. and RAMESH B. (1999), Molecular markers and their applications in wheat breeding, Plant breeding 118: 369-390

GUTIERREZ M.C., VAN PATTOS M.C., HUGUET T., CUBERO J.I., MORENO M.T. and TORRES A.M. (2005), Cross-species amplification of *Medicago truncatula* microsatellites across three major pulse crops, Theoretical and Applied Genetics 110: 1210 - 1217

HAMRICK J. L. and GODT M.J.W. (1996), Effects of life history traits on genetic diversity in plants species, Philosophical Transactions: Biological Sciences351: 1291 - 1298

HANELT P. (2001), Leguminosae subfamily (Fabaceae). In Mansfield's Encyclopaedia of Agricultural and Horticultural Crops, 2: 635 -957 (Eds. Hanelt P.), Springer, Berlin, Germany

HARRIS D and AZAM-ALI S.N (1993), Implications of day length sensitivity in bambara groundnut (*Vigna subterranea* L. Verdc.) production in Botswana, Journal of Agricultural Science 120: 75 - 78

HAZEL L.N. (1943), The genetic basis for constructing selection indexes, Genetics28: 476 - 490

HE G., MENG R., NEWMAN M., GAO G., PITTMAN R.N. and PRAKASH C.S. (2003), Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea*L.), BMC Plant Biology 3: 1-6

HEDRICK P.W. (2005), Genetics of populations, Jones and Bartlett, London, UK.

HENNINK S. and ZEVEN A.C. (1991), Interpretation of Nei and Shannon-Weaver within population variation indices, Euphytica 51: 235 - 240

HEPPER F.N. (1963), Plant of the 1957-58 West Africa expedition II: The bambara groundnut (*Voandzeia subterranea*) and Kersting's groundnut (*Kerstingiella geocarpa*) wild in West Africa. Kew Bulletin 16: 395 - 407

HILL J., BECKER H.C. and TIGERSTEDT (1998), Quantitative and ecological aspects ofplant breeding, Chapman and Hall, London

HIRST L.G. and ILLAND M. (2001), Automated detection of microsatellite instability, Molecular Biotechnology 17: 239 - 247

HUBBY J.L. and LEWONTIN R.C. (1966), A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in*Drosophila pseudoobscura,* Genetics 54: 577 -594

HURTADO P., OLSEN K.M., BUITRAGO C., OSPINA C., MARIN J., DUQUE M., DE VINCENTE C., WONGTIEM P., WENZEL P., KILIAN A., ADELEKE M. and FREGENE M. (2008), Comparison of simple sequence repeat (SSR) and diversity (DArT) markers for assessing genetic diversity in cassava (*Manihot esculenta* Crantz), Plant Genetic Resources: Characterisation and utilization 6: 208 - 214

IDURY R.M. and CARDON L.R. (1997), A simple method for automated allele binning in microsatellites markers, Genome Research 7: 1104 - 1109

IJAROTIMI S.O. and ESHO R. T. (2009), Comparison of nutritional composition and anti-nutrient status of fermented, germinated and roasted bambara groundnut seeds (*Vigna subterranea*), British Food Journal 111: 376-386

IPGRI, IITA, BAMNET (2000), Descriptors for bambara groundnut (*Vigna subterranea*), International Plant Genetic Resources Institute, Rome, Italy; International Institute of Tropical Agriculture, Ibadan, Nigeria; International Bambara Groundnut Network, Germany.

JACCOUD D., PENG K., FEINSTEIN D. and KILIAN A. (2001), Diversity Arrays: a solid state technology for sequence information independent genotyping, Nucleic Acids Research 29: 1 - 7

JACKSON D.S. (2008), Plant responses to photoperiod, New Phytologist, 181: 517 -531

JEUFFROY M-H. and CHABANET C. (1994), A model to predict seed number per pod from early pod growth rate in pea (*Pisum sativum* L.) Journal of Experimental Botany 45: 709 - 715

JOLLIE I.T. (2002), Principal component analysis, Springer, New York.

JONAH P.M., ADENIJI O.T. and WAMMANDA D.T. (2010), Variability and genetic correlations for yield and yield characters in some bambara groundnut (*Vigna subterranea*) cultivars, International Journal of Agriculture and Biology 12: 303 - 307

JONAH P.M., ADENIJI O.T. and WAMMANDA D.T. (2010), Genetic correlations and path analysis in bambara groundnut (*Vigna subterranea*), Journal of Agriculture and Social Sciences 6: 1- 5

JØRGENSEN S.T., LUI F., OUÉDRAOGO M., NTUNDU W.H., SARRAZIN J. and CHRISTIANSEN L.J. (2010), Drought responses of two bambara groundnut (*Vigna subterranea* L. Verdc.) landraces collected from a dry and a humid area of Africa, Journal of Agronomy and Crop Science 196: 412 - 422

KALINOWSKI T.S. (2004), Counting alleles with rarefaction: private alleles and hierarchical sampling designs, Conservation Genetics 5: 539 -543

KAMPHIUS L.G., WILLIAMS H.A., D'SOUZA K.N., PFAFF T., SINGH B.K., OLIVER R.P. and LICHTENZVEIG J. (2007), The *Medicago truncatula* reference accession A17 has an aberrant chromosomal configuration, New Phytologist 174: 229-303

KARIKARI S.K. (2000), Variability between local and exotic bambara groundnut landraces in Botswana, African Crop Science Journal 8: 145 - 152

KARIKARI S.K. and TABONA T.T. (2004), Constitutive traits and selective indices of bambara groundnut (*Vigna subterranea* (L.) Verdc) landraces for drought tolerance under Botswana conditions, Physics and Chemistry of the Earth 29: 1029 - 1034

KARIKARI S.K., WIGGLESWORTH D.J., KWEREPE B.C., BALOLE T.V., SEBOLAI B. and MUNTALI D.C. (1995), Country report for Botswana. In: Promoting the conservation and use of underutilized and neglected crops. 9. Proceedings of the workshop onConservation and Improvement of bambara groundnut (*Vigna subterranea*(L.) Verdc.) 14-16 November, 1995, Harare, Zimbabwe, pp 11 – 18 (Eds. Heller J., Begemann F. and Mushonga J.) International Plant Genetic Resources Institute, Rome, Italy

KARURI H.W., ATEKA E.M., AMATA R., NYENDE A.B., MUIGAI A.W.T., MWASAME E. and GICHUKI S.T. (2010), Evaluating diversity among Kenyan sweet potato genotypes using morphological and SSR markers, International Journal of Agricultural Biology 12: 33 -38

KHEDIKAR P.Y., GOWDA M.C.V., SARVAMANGALA C., PATGAR K.V., UPADADHYAYA H.D. and VARSHNEY R.K. (2010), A QTL study on late leaf spot and rust revealed one major QTLformolecular breeding for rust in groundnut (*Arachis hypogaea* L.), Theoretical and Applied Genetics 121: 971 - 984

KLODA J.M., DEAN P.D.G., MADDREN C., MACDONALD D.W. and MAYES S. (2007), Using principle component analysis to compare genetic diversity across polyploidy levels within plant complexes: an example from British Restharrows (*Ononis spinosa* and *Ononis repense*), Heredity 100: 253 -260

KOBRAEE S., SHAMSI K., RASEKHI B. and KOBRAEE S. (2010), Investigation of correlation analysis and relationship between grain yield and other quantitative traits in Chickpea (*Cicer arietinum* L.), African Journal of Biotechnology 9: 2342-2348

KOENIG R. and GEPTS P. (1989), Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centres of genetic diversity, Theoretical and Applied Genetics 78: 809-817

KOUASSI N.J. and ZORO B. I.A. (2009), Effect of sowing density and seedbed typed on yield and yield components in bambara groundnut (*Vigna subterranea*) in woodland savannas of Côte d'Ivore, Experimental Agriculture 46: 99 -110

KOVACH W. (2006), MVSP-*Multivariate Statistical Package Version 3.1*. Kovach Computing Services: Anglesey, Wales

KRISHMA G.K., ZHANG J., BURROW M., PITTMAN R.N., LU Y. and PUPPALA N. (2004), Genetic diversity analysis in Valencia peanut (*Arachis hypogaea* L.) using microsatellite markers,Cellular and Molecular Biology Letters 9: 685-695

KUMAR J., CHOUDHARY A.K., SOLANKI R.K. and PRATAP A. (2011), Towards marker- assisted selection in pulses: a review, Plant breeding 130: 297 -313

KUMAR P., GUPTA V.K., MISRA A.K., MODI D.R. and PANDEY B.K. (2009), Potential of molecular markers in plant biotechnology, Plant Omics Journal 2: 141 - 162

KUMAR V., SHARMAR S., SHARMA K.A., SHARMA S. and BHAT V.K. (2009), Comparative analysis of diversity based on morpho-agronomic traits and microsatellite markers in common bean, Euphtyica 170: 249 -262

KURODA Y., KAGA A., TOMOOKA N. and VAUGHAN D.A. (2006), Population genetic structure of Japanese wild soybean (*Glycine soja*) based on microsatellite variation, Molecular Ecology 15: 959 -974

KUSWANTORO H., WIJANARKO A., SETYAWAN D., WILLIAM E., DADANG A. and MEJAYA J.M. (2010), Soybean germplasm evaluation for acid tidal swamp tolerance using selection index, International Journal of Plant Biology 1: e11

LAZREK F., ROUSSEL V., RONFORT J., CARDINET G., CHARDON F., AOUANI M.E. and HUGNET T. (2009), The use of the neutral and non-neutral SSRs to analyse the genetic structure of a Tunisian collection of *Medicago truncatula* lines and to reveal association with eco-environmental variables, Genetica 135: 391 - 402

LEBERG P.L. (2002), Estimating allelic richness: effect of sample size and bottlenecks, Molecular Ecology 11: 2445 - 2449

LEGENDRE P. and FORTIN M.J. (2010), Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data, Molecular Ecology Resources 10: 831 -844

LI P., WANG Y., SUN X.and HAN J. (2009), Using microsatellite (SSR) and morphological markers to assess the genetic diversity of 12 falcata (*Medicago sativa* spp. *falcata*) populations from Eurasia, African Journal of Biotechnology 8: 2102 - 2108

LIANG X., CHEN X., HONG Y., LIU H., ZHOU G., LI S. and GUO B. (2009), Utility of EST-derived SSR in cultured peanut (*Arachis hypogaea* L.) and *Arachis*wild species, BMC Plant Biology 9: 1 - 9

LINNEMANN A.R. and AZAM-ALI S.N. (1993), Bambara groundnut (*Vigna subterranea* (L.) Verdc). In: underutilised crop series. II. Vegetables and pulses, pp. 13 -57 (Eds. Williams, J.T), Chapman and Hall, London.

LINNEMANN A. R (1994), Phenological development in bambara groundnut (*Vigna subterranea*) at alternate exposure to 12 and 14 h photoperiods, Journal of Agricultural Science 123: 333 - 340

LINNEMANN A.R and CRAUFURD P.Q. (1994), Effects of temperature and photoperiod on phenological development in three genotypes of bambara groundnut (*Vigna subterranea*), Annals of Botany 74: 675 - 681

LINNEMANN A. R., WESTPHAL E. and WESSEL M. (1995), Photoperiod regulation of development and growth in bambara groundnut (*Vigna subterranea*), Field Crops Research 40: 39 - 47

LIU J., JIAN-PING G., DONG-XU X., XIAO-YANG Z., JING G. and XU-XIAO Z. (2008), Genetic diversity and population structure in Lentil (*Lens culinaris*Medik) Germplasm detected by SSR markers, Acta Agronomica Sinica 34: 1901 -1909

LIU M., ZHANG M., JIANG W., SUN G., ZHAO H. and HU S. (2011), Genetic diversity of Shaanxi soybean landraces based on agronomic traits and SSR markers, African Journal of Biotechnology 10: 4823 - 4837

LIU K., GOODMAN M., MUSE S., SMITH S.J., BUCKLER E. and DOEBLEY J. (2003), Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites, Genetics 165: 2117 - 2128

LOH P.J., KIEW R., KEE A., GAN H.L. and GAN Y. (1999), Amplified Fragment Length Polymorphism (AFLP) provides molecular markers for the identification of *Caladium bicolour* cultivars, Annals of Botany 84: 155-161

LU J., KNOX M.R., AMBROSE M.J., BROWN J.K.M. and ELLIS T.H.N. (1996), Comparative analysis of genetic diversity in pea assessed by RFLP- and PCR-based methods, Theoretical and Applied Genetics 93: 1103-1111

LUI F., CHARLESWORTH D. and KREITMAN M. (1999), The effect of mating system differences on nucleotides diversity at the phosphoglucose isomerase locus in the plant genus Leavenworthia, Genetics 151: 353 - 357

LUI K. and MUSE S. V. (2005), PowerMarker: Integrated analysis environment for genetic marker analysis. Bioinformatics 21: 2128 - 2129

MACE E.S., PHONG D.T.,UPADHAYAYA H.D., CHANDRA S. and CROUCH J.H. (2006), SSR analysis of cultivated (*Arachis hypogaea* L.) germplasm resistant to rust and late leaf spot diseases, Euphytica 152: 317 - 330

MACE E.S., RAMIN J-F, BOUCHET S., KLEIN P.E.,KLIEN R.R., KILIAN A., WENZEL P., XIA L., HALLORAN K.and JORDAN D.R. (2009), A consensus genetic map of sorghum that integrates multiple component maps and high-throughput DiversityArray Technology (DArT) markers, BMC Plant Biology 9: 13

MACIEL F.L., ECHEVERRIGARAY S., GERALD L.T.S. and GRAZZIOTIN G. F. (2003), Genetic relationships and diversity among Brazilian cultivars and landraces of common beans (*Phaseolus vulgaris* L.) revealed by AFLP markers, Genetic Resources and Crop Evolution 50: 887 - 893

MAGAGULA C.N., MANSUETUS A.B., SESAY A. and KUNENE I.S. (2003), Yield loss associated with pests and diseases on bambara groundnut (*Vigna subterranea* (L.)Verdc.) in Swaziland. In. Proceedings of the International bambara groundnut Symposium, Botswana College of Agriculture, Botswana, 8 -12 August 2003 pp.95- 105

MAKANDA I., TONGOONA P., MADAMBA R., ICISHAHAYO D. and DERERA J. (2009), Path coefficient analysis of bambara groundnut pod yield components at four planting dates, Research Journal of Agriculture and Biological Sciences 5: 287-292

MAKANDA I., TONGOONA P., MADAMBA R., ICISHAHAYO D. and DERERA J. (2009), Evaluation of bambara groundnut varieties for the production in Zimbabwe, African Crop Science Journal 16: 175-183

MALIK M.F.A., ASHRAF M., QURESHI A.F. and GHAFOOR A. (2007), Assessment of genetic variability, correlation and path analyses for yield and its components in soybean, Pakistan Journal of Botany 39: 405 - 413

MALIK S.S. and SINGH S.P. (2006), Role of plant genetic resources in sustainable agriculture, Indian Journal of Crop Science 1: 21 - 28

MANIEE M., KAHRIZI D. and MOHAMMADI R. (2009), Genetic variability of some morpho-physiological traits in durum wheat (*Triticum turgidum* var. *durum*), Journal of Applied Sciences 7:1383 - 1387

MANTEL N. (1967), The detection of disease clustering and generalised regression approach, Cancer Research 27: 209 -220

MANTOVANI P., MACCAFERRI M., SANGUITE M.C., TUBEROSA R., CATIZONE I., WENZEL P., THOMSON B., CARLING J., HUTTNER E., DEAMBROGIO E and KILIAN A (2008), An integrated DArT- SSR linkage map of durum wheat, Molecular Breeding 22: 629 - 648

MAQUET A., BI-ZORO I., DELVAUX M., WATHELET B.,and BAUDOIN P.J. (1997), Genetic structure of Lima bean base collection using allozymes markers, Theoretical and Applied Genetics 95: 980 - 991

MARAS M., ŠUšTAR-VOZLIČ J., JAVORNIK B. and MEGLIČ V. (2008), The efficiency of AFLP and SSR markers in genetic diversity estimation and genepool classification of common bean (*Phaseolus vulgaris* L.), Acta Agriculturae Slovenia 91: 8 - 96

MASI P., ZEULI S.L.P. and DONNI P. (2003), Development and analysis of multiplex microsatellite marker set in common bean (*Phaseolus vulgaris* L.), Molecular Breeding 11: 303 - 313

MASSAWE F.J., MWALE S.S., AZAM-ALI S.N. and ROBERTS J.A. (2005), Breeding in bambara groundnut (*Vigna subterranea* (L.) Verdc.): strategic considerations, African Journal of Biotechnology 4: 463 - 471

MASSAWE F.J., ROBERT J.A., AZAM-ALI S.N. and DAVEY M. R. (2003), Genetic diversity in bambara groundnut (*Vigna subterranea* (L.) Verdc) landraces assessed by Random Amplified Polymorphic DNA (RADPs) markers, Genetic Resources and Crop Evolution 50: 737-741

MASSAWE F.J., DICKINSON M., ROBERT J.A. and AZAM-ALI S.N. (2002), Genetic diversity in bambara groundnut (*Vigna subterranea* (L.) Verdc) landraces revealed by AFLP markers, Genome 45: 1175 - 1180

MASSAWE F.J., AZAM-ALI S.N. and ROBERTS J.A. (2003), Impact of temperature on leaf development in bambara groundnut landraces, Crop Science 43: 1375 -1379

MASSAWE F.J., SCHENKEL W., BASU S. and TEMBA E.M. (2003), Artificial hybridization in bambara groundnut (*Vigna subterranea* (L.) Verdc.)In. Proceedings of the International bambara groundnut Symposium, Botswana College of Agriculture, Botswana, 8 -12 August 2003 pp. 193 - 209

MATSCHINER M. and SALZBURGER W. (2009), TANDEM: Integrating automatic allele binning into genetics and genomics workflows, Bioinformatics 25: 1982 - 1983

MAYES S., BASU, S., MURCHIE E., ROBERTS J.A., AZAM-ALI S.N., STADLER F., MOHLER V., WENZEL G., MASSAWE F., KILIAN A., BONIN A., BEENA A. and SHESHSHAYEE M.S.(2009), BAMLINK – a Cross disciplinary programme to enhance the role of bambara groundnut(*Vigna subterranea* L. Verdc.) for food security in Africa and India. Acta Horticulturae 806: 137 - 150.

MBEWE D.N., MWALA M.S. and MWILA G.P. (1995), Country report for Zambia. In: Promoting the conservation and use of underutilized and neglected crops. 9. Proceedings of the workshop on Conservation and Improvement of bambara groundnut (*Vigna subterranea* (L.) Verdc.) 14-16 November, 1995, Harare, Zimbabwe, pp. 64 - 67 (Eds. Heller J., Begemann F. and Mushonga J.) International Plant Genetic Resources Institute, Rome, Italy

MCCOUCH S.R., CHEN X., PANAUD O., TEMNYKH S., XU Y., CHO G.Y., HUANG N., ISHII T. and BLAIR M. (1997), Microsatellite marker development, mapping and applications in rice genetics and breeding, Plant Molecular Biology 35: 89 -99

MENDELSOHN R (2000), Climate Change Impacts on African Agriculture, Ariel Dinar, World Bank, pp. 2 - 25

MICHELMORE R.W. and HUBERT S.H. (1987), Molecular markers for genetic analysis of phytopathogenic fungi, Annual Review of Phtyopathology 25: 383 - 404

MILLAN T., CLARKE H.J., SIDDIQUE K.H.M., BUHARIWALLA H.K., GAUR P.M., KUMAR J., GIL J., KAHL G. and WINTER P. (2006), Chickpea molecular breeding: New tools and concepts, Euphtyica 147: 81 - 103

MINE O.M., MOTLHABANE L.T. and BATLANG U. (2003), Preliminary assessment of genetic variation in bambara groundnut in Botswana using RAPD markers: a case of technology transfer. In: Proceedings of the International bambara groundnut symposium, Botswana College of Agriculture, Botswana, 8 -12, August, 2003 pp. 175 - 178

MINISTRY OF AGRICULTURE, GOVERNMENT OF BOTSWANA (1947), Jugo Beans. In: Review of Crop Experiments in Bechuanaland Protectorate, Department of Agricultural Research, Botswana pp. 127 - 128

MISANGU R.N., AZMIO A. and REUBEN S.O.W.M. (2007), Path coefficient analysis among components of yield in bambara groundnut (*Vigna subterranea* L. Verdc) landraces under screen house conditions, Journal of Agronomy 6: 317 - 323

MKANDAWIRE C.H. (2007), Review of Bambara groundnut (*Vigna subterranea* (L.) Verdc.) Production in Sub-Sahara Africa, Agricultural Journal 2: 464 -470

MONDINI L., NOORANI A. and PAGNOTTA M.A. (2009), Assessing plant genetic diversity by molecular tools, Diversity 1: 19 -35

MONIRIFAR H. (2010), Evaluation of selection indices for alfalfa (*Medicago sativa* L.), Notulae Scientia Biologicae 2: 84 - 87

MONTAVANI P., MACCAFERRI M., SANGUINETI M.C., TUBEROSA R., CATIZONE I., WENZEL P., THOMSON B., CARLING J., HUTTNER E., DEAMBROGIO E. and KILIAN A. (2008), An Integrated DArT-SSR linkage map of durum wheat,Molecular Breeding 22: 629 - 648

MORENO R.R and POLANS N.O. (2006), Development and characterisation of microsatellites loci in peas, Pisum Genetics 38: 10 -14

MUKUMBIRA L.M.(1985), Effects of rate of nitrogen fertilizer and previous grain legume crop on maize yield, Zimbabwe Agriculture Journal 82: 177 - 179

MULLIS K. (1990), The unusual origin of polymerase chain reaction, Scientific American April: 56 - 65

MUNTHALI D.C. and RAMORANTHUDI M. (2003), Susceptibility of bambara groundnut landraces to *Hilda patreulis*. In. Proceedings of the International bambara groundnut Symposium, Botswana College of Agriculture, 8 – 12 August2003 pp. 85-105

MWALE S.S., AZAM-ALI S.N. and MASSAWE F. J. (2007), Growth and development of bambara groundnut (*Vigna subterranea*) in response to soil moisture1.Dry matter and yield, European Journal Agronomy 26: 345 - 353

NATH U.K. and ALAM M.S. (2002), Genetic variability, heritability and genetic advance of yield and related traits of groundnut (*Arachis hypogaea* L.), Journal of Biological Sciences 2: 762 - 764

NEI M. and LI W. (1979), Mathematical model for studying genetic variation in terms of restriction endonucleases, Proceedings of the National Academic Science of United States of America 76: 5256 - 5273

NEI M. (1987), Molecular Evolutionary Genetics. Columbia University Press, New York

NOEPARVAR S., VALIZADEH M., MONIRIFAR H., HAGHIGHI R.A. and DARBANI B. (2008), Genetic diversity among and within alfalfa population native to Azerbaijan based on RAPD analysis, Journal of Biological Research- Thessaloniki 10: 159 -169

NTUNDU W.H., SHILLAH S.A., MARANDU W.Y.F. and CHRISTIANSEN J.L. (2006), Morphological diversity of bambra groundnut [*Vigna subterranea* (L.) Verdc.] landraces in Tanzania, Genetic Resources and Crop Evolution 53: 367 - 378

NTUNDU W.H., BACH I.C., CHRISTIANSEN J.L. and ANDERSEN S.B. (2004), Analysis of genetic diversity in bambara groundnut (*Vigna subterranea* L. Verdc)landraces using amplified fragment length polymorphism (AFLP)markers, African Journal of Biotechnology 3: 220 - 225

NUNOME T., NEGORO S., MIYATAKE K., YAMAGUCHI H. and FUKUOKA H. (2006), A protocol for the construction of microsatellite enriched genomic library, Plant Molecular Biology Reporter 24: 305 - 312

OBAGWU J. (2003), Evaluation of bambara groundnut (*Vigna subterranea* L. Verdc) lines for reaction to Cercospora Leaf Spot, Journal of Sustainable Agriculture 22: 193 - 100

ODENY D.A., JAYASHREE B.,FERGUSON M., CROUCH J. and GEBHARDT C. (2007),Development, characterisation and utilization of microsatellites markers in pigeonpea, Plant breeding 126: 130 -136

ODENY D.A., JAYASHREE B., GERBHARDT C. and CROUCH J. (2009), New microsatellite markers for pigeonpea (*Cajanus cajan* (L.) millsp.), BMC Research Notes 2: 35

OFORI K. (1996), Correlation and path-coefficient analysis of components of seed yield in bambara groundnut (*Vigna subterranea*), Euphytica 91: 103 -107

OFORI K., KUMANGA F.K. and Tonyigah A. (2010), Morphological characterization and agronomic evaluation of bambara groundnut (*Vigna subterranea* (L.) Verdc.) germplasm, PGR Newsletter, FAO-Biodiversity 145: 23 - 28

OHARA M. and SHIMAMOTO Y. (2002), Importance of genetic characterisation and conservation of plant genetic resources: The breeding system and geneticdiversity of wild soybean (*Glycine soja*), Plant Species Biology 17: 51-58

OHIOKPEA O. (2003), Food processing and nutrition: A vital link in agricultural development, Pakistan Journal of Nutrition 2: 204 - 207

OJULONG F.H., LABUSCHAGNE T.M., HERSELMAN L. and FREGENE M. (2010), Yield traits as selection indices in seedling populations of cassava, Crop Breeding and Applied Biotechnology 10: 191 -196

OKORI P.W.S. and RUBAIHAYO P.R. (2005), Genetic variability and relatedness of the Asian and African pigeonpea as revealed by AFLP, African Journal of Biotechnology 4: 1228 - 1233

OLUKOLU B.A., MAYES S., STADLER F., NG Q.N., FAWOLE I., DOMINIQUE D., AZAM- ALI S.N., ABOTT G. A. and KOLE C. (2011), Genetic diversity in bambara groundnut (*Vigna subterranea* (L.) Verdc.) as revealed by phenotypic descriptors and DArT marker analysis, Genetic Resources and Crop Evolution 59: 347 -358

OMOIGUI L.O, ISHIYAKA M.F, KAMARA A.Y, ALABI S.O and MOHAMMED S.G (2006), Genetic variability and heritability studies of some reproductive traits in Cowpea (*Vigna unguiculata*(L.) Walp.), African Journal of Biotechnology5: 1191-1195

ONWUBIKO N.I.C., ODUM O.B., UTAZI C.O. and POLY-MBAH P.C. (2011), Studies on the adaptation of bambara groundnut (*Vigna subterranea* L. Verdic.) in Owerri southeastern Nigeria, New York Science Journal 4: 60 - 67

OUEDRAOGO M., OEUDRAOGO J.T., TIGNERE J.B., BALMA D., DABIRE C.B. and KONATE G. (2008), Characterization and evaluation of accessions of bambara groundnut (*Vigna subterranea* L. Verdcourt) from Burkina Faso, Science and Nature 5: 191 - 197

OYIGA B.C and UGURU M.I. (2011) Interrelations among pod and seed yield traits in bambara groundnut (*Vigna subterranea* L. Verdc) in the derived savanna agro-ecology of south-eastern Nigeria under two planting dates, International Journal of Plant Breeding 5: 106 - 111

OYIGA B.C., UGURU M.I. and ARUAH C.B. (2010), Studies on the floral traits and their implications on pod and seed yield in bambara groundnut (*Vigna subterranea* (L.) Verdc.), Australian Journal of Crop Science 4: 91-97

PAINAWADEE M., JOGLOY S., KESMALA T., AKKASAENG C. and PATANOTHAI A. (2009), Heritability and correlations of drought resistance traits and agronomic traits in peanut (*Arachis hypogaea* L.), Asian Journal of Plant Sciences 8: 325 -334

PARSAEIAN M., MIRLONI A. and SAEIDI G. (2010), Study of genetic variation in Sesame (*Sesamum indicum* L.) using agro-morphological traits and ISSR markers, Russian Journal of Genetics 47: 359- 367

PARZIES H.K., SPOOR W. and ENNOS R.A. (2000), Genetic diversity of barley landrace accession (*Hordeum vulgare* ssp. *vulgare*) conserved for different length of time in ex situ gene banks, Heredity 84: 476 - 486

PASQUET R.S. (2003), Bambara groundnut and cowpea gene-pool organization and domestication: In. Proceedings of the International bambara groundnut Symposium, Botswana College of Agriculture, Botswana, 8- 12, August, 2003 pp.263 - 272

PASQUET R.S., SCHWEDE S. and GEPTS P. (1999), Isozyme diversity in bambara groundnut, Crop Science 39: 1228-1236

PEJIC I., AJMONE-MARSAN P., MORGANTE M., KOZUMPLICK V., CASTGLIONI P., TARAMINO G. and MOTTO M. (1998), Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs, and AFLPs, Theoretical and Applied Genetics 97: 1248 -1255

PESEK J. and BAKER R.J. (1969), Desired improvement in relation to selected indices, Canadian Journal of Plant Science 49: 803 - 804

POLHILL R.M., RAVEN P.H. and STIRTON C.H. (1981), Evolution and systematics of the legumes, In: Advances in legume systematics, pp. 1 – 26 (Eds Polhill R.M. and Raven P.H), Royal Botanic Garden, Kew, England.

POLHILL R.M. and RAVEN P.H. (1981), Advances in legumes systematics, part 1, Royal Botanic Gardens, Kew.

POMPANON F., BONIN A., BELLEMAIN E. and TABERLET P. (2005), Genotyping errors: causes, consequences and solutions, Genetics 6: 847 - 859

POULTER N.H. and CAYGILL J.C. (1980), Vegetable milk processing and rehydration characteristics of bambara groundnut (*Voandzeia*

subterranea (L.) Thouars),Journal of the Science of Food and Agriculture 31: 1158 - 1163

POWELL W., MORGANTE M., ANDRE C., HANAFEY M., VOGEL J., TINGEY S. and RAFALSKI A. (1996), The comparison of RFLP, RAPDs, AFLP and SSR (microsatellite) markers for germplasm analysis, Molecular Breeding 2:225 -238

PRIOLLI G.B.H.R., PINHEIRO B.J., ZUCCHI I.M., BAJAY M.M. and VELLO A. N. (2010), Genetic diversity among Brazilian soybean cultivars based on SSR loci and pedigree data, Brazilian Archives of Biology and Technology 53: 519 - 531

PUTTHA R., JOGLOY S., WONGKAEW S., SANITCHON J., KESMALA T. and PATANOTHAI A. (2008), Heritability, phenotypic and genotypic correlation of *Peanut bud necrosis virus* resistance and agronomic traits of peanut, Asian Journal of Plant Science 7: 276 - 283

QADIR S.A., DATTA S., SINGH N.P. and KUMAR S.(2007), Development of highly polymorphic markers for chickpea (*Cicer arietinum* L.) and their use as parental polymorphism, Indian Journal of Genetics 67: 329 -333

RAHIM A.M., MIA A.A., MAHMUD F. and AFRIN K.S. (2008), Multivariate analysis in some mungbean (*Vigna radiata* L. Wilczek) accessions on the basis of agronomic traits, American-Eurasian Journal of Scientific Research 3: 217 - 221

RAINA S.N., KOJIMA T., OGIHARA Y., SINGH K.P. and DEVAURUMATH R.M. (2001), RAPDs and ISSR fingerprints as useful genetic markers for analysis of genetic diversity, varietal identification, and phylogenetic relationships in peanut (*Arachis hypogaea*) cultivars and wild species, Genome 44: 763 -772

RAMAKRISHNAN P.A., MEYER E.S., WATERS J., STEVENS R.M., COLEMAN E.C. and FAIRBANKS D.J. (2004), Correlation between markers and adaptively significant variation in *Bromus tectorum* (Poaceae), an inbreeding annual grass, American Journal of Botany, 91: 797 - 803

RAMAN R., LUCKETT D.J. and RAMAN H. (2008), Estimating the genetic diversity in Albus Lupin (*Lupinus albus* L.), using DArT and genetic markers. In: Lupins for Health and Wealth' Proceedings of the 12th International Lupin Conference, 14 – 18 September 2008, pp. 236 – 241 (Eds. Palta J.A. and Berger J.B.), Fremantle, Western Australia, International Lupin Association, Canterbury, New Zealand.

RAMOLEMANA G.M., KWEREPE B.C., BALOLE T.V., KHONGA E.B. and KARIKARI S.K.(2003), Farmer and consumer perception on the bambara groundnut (*Vigna subterranea*) ideotype in Botswana, In: Proceedings of the International bambara groundnut Symposium, Botswana College of Agriculture, Botswana, 8 -12 August 2003 pp.17 - 24

RAUF S., JAIME A., DA SILVA T., KHAN A.A. and NAVEED A. (2010), Consequences of plant breeding on genetic diversity, International Journal of Plant Breeding 4:1 - 21.

RAVI K., VADEZ V., ISOBE S., MIR R.R., GUO Y., NIGAM S.N., GOUDA M.V.C., RADHAKRISHMAN T., BERTIOLE D.J., KNAPP S.T. and VARSHNEY R.K. (2011), Identification of several small main-effect QTLs and a large number of epistatic QTLs for drought tolerance related traits in groundnut (*Arachis hypogaea* L), Theoretical and Applied Genetics 122: 1119 - 1132

REUSCH B.T. (2001), New markers- old questions: population genetics of sea grasses, Marine Ecology Progress Series 211: 261-274

RITSCHEL P.S., DE LIMA LINS T.C., TRISTAN R.L., BUSO G.S.C., BUSO J.A. and FERREIRA M.E (20040, Development of microsatellites markers from enriched genomic library for genetic analysis of melon (*Cucumis melo* L.), BMC Plant Biology 4: 9

ROBERTSON A. (1961), Inbreeding in artificial selection programmes, The International Journal for Genetics and Genomics Research 2: 189 - 194

ROBERTSON A. and HILL W.G. (1984), Deviations from Hardy-Weinberg proportions:sampling variables and use in estimation of inbreeding coefficient, Genetics 107: 703 - 718

ROHLF F.J. (2000), NTSYS-pc Numerical Taxonomy and Multivariate Analysis system version 2.1, Applied Biostatistics, New York

ROVEN S. and SKALETSKY H.J. (2000), Primer3 on WWW for general users and for biologists programmers. In: *Bioinformatics Methods and Protocols*: *Methods in Molecular Biology*, pp. 365 - 386 (Eds Krawetz S. and Misener S.) Humana Press, Totowa

ROY D.(2000), Plant breeding, analysis and exploitation of variation, Alpha Science International, UK.

RYCHLIK W., SPENCER W. and RHOADS R.E. (1990), Optimization of the annealing temperature for DNA amplification *in vitro*, Nucleic Acids Research 18: 6409 - 6412

RYMER D. P., MORRIS E. C. and RICHARDSON B.J. (2002), Breeding systems and population genetics of the vulnerable plant *Dillwynia tennuifolia*(Fabaceae), Austal Ecology 27: 241 - 248

SAIKI R.K., SCHARF S., FALOONA F., MULLIS K.B., HORN G.T., ERLICH H.A. and ARNHEIM N. (1985), Enzymatic amplification of beta- globin genomic sequences and restriction site analysis for diagnosis of sickle cell anaemia, Science 230: 1350 -1354

SALEHI M., TAJIK M. and EBADI A.G. (2008), The study of relationships between different traits in common bean (*Phaseolus vulgaris* L.) with

multivariate statistical methods, American Journal of Agriculture and Environmental Science 3: 806-809

SANDAL N., KRUSSELL L., RADUTOIU S., OLBRYT M., PEDROSA A., STRACKE S., SATO S., KATO T., TABATA S., PARNISKE M., BACHMAIR A., KETELSEN T. and STOUGAARD J. (2002), A genetic linkage map of the model legume *Lotus japonica* and strategies for fast mapping of new loci, Genetics 161: 1678 -1683

SANTANA C.Q., COETZEE M.P.A., STEENKAMP E.T. and MLONYENI X. O. (2009), Microsatellites discovery by deep sequencing of enriched genomic libraries, BioTechniques 46: 217 - 223

SANTRA K.D., TEKEOGLU M., RATNAPARKHE M., KAISER J.W. and MUEHLBAUER J.F. (2000), Identification and mapping of QTLs conferring resistance to *Aschochyta Blight* in chickpea, Crop Science 40: 1606 -1612

SATO S., ISOBE S. and TABATA S. (2010), Structural analyses of the genomes in legumes, Current Opinion in Plant Biology 13: 146 - 152

SAXENA R.K., PRATHIMA C., SAXENA K.B., HOISINGTON D.A., SINGH N.K. and VARSHNEY R.K. (2010), Novel SSR markers for polymorphism detection in pigeonpea (*Cajanus* spp.), Plant Breeding 129: 142 - 148

SCHNEIDER A., WALKER S.A., SAGAN M., DUC G., ELLIS T.H.N. and DOWIE J.A. (2002), Mapping of the nodulation loci *sym9* and *sym10*, Theoretical and Applied Genetics 104: 1312 -1316

SCHUELKE M. (2000), An economic method for the fluorescent labelling of PCR fragments. Nature Biotechnology 18: 233-234

SEEHALAK W., SOMTA P., SOMMANAS W. and SRINIVES P. (2009), Microsatellites markers for mungbean developed from sequence database, Molecular Ecology Resources 9: 862 - 864

SEMAGN K., BJøRNSTAD Å. and NDJIONDJOP M.N. (2006), An overview of molecular marker methods for plants, African Journal of Biotechnology 5: 2540 -2568

SEMAGN K., BJøRNSTAD Å. and XU Y. (2010), The genetic dissection of quantitative traits in crops, Electronic Journal of Biotechnology 13: 1 - 45

SESAY A. (2009), Influence of flooding on bambara groundnut (*Vigna subterranea*L.) germination: effect of temperature, duration and timing, African Journal of Agricultural Research 4: 100 - 106

SESAY A., COLLINSON S.T. and AZAM-ALI S.N. (1996), Where are we now with bambara groundnut? In: Proceedings of the international Bambara groundnut symposium, University of Nottingham, UK 23 – 25 July, 1996 pp.215 - 228

SESAY A., MAGAGULA C.N. and MANSUETUS A.B. (2008), Influence of sowing date and environmental factors on the development and yield of bambaragroundnut (*Vigna subterranea*) landraces in a sub-tropical region, Experimental Agriculture 44: 167 - 183

SESAY A., EDJE O.T. and MAGAGULA C.N. (2003), Working with farmers on the bambara groundnut (BAMFOOD) research project in Swaziland, In Proceedings of the International bambara groundnut Symposium, Botswana College of Agriculture, Botswana 8 - 12 August 2003 pp.3 - 15

SETHY K.N., SHOKEEN B. and BHATIA S. (2003), Isolation and characterisation of sequence-tagged microsatellite site markers in chickpea (*Cicer arietinum*L.), Molecular Ecology Notes 3:428 - 430

SHARMA C.P., GROVER A. and KAHL G. (2007), Mining microsatellites in eukaryotic genome, TRENDS in Biotechnology 25: 490 - 498

SHARMA S., SHARMA S., KOPISCH-OBUCH F.J., KEIL T., LAUBACH E., STEIN N., GRANER A. and JUNG C. (2011), QTL analysis of root-lesion nematode resistance in barley: 1. *Pratylechus neglectus* Theoretical and Applied Genetics 122: 1321 -1330

SHETE S. (2003), Uniformly minimum variance unbiased estimation of gene diversity, Journal of Heredity 94: 421- 424

SHETE S., TIWARI H. and ELSTON R.C. (2000), On estimating the heterozygosity and polymorphisminformation content value, Theoretical Population Biology, 57: 265 -271

SHI C., NAVABI A. and YU K. (2011), Association mapping of common bacterial blight resistance QTL in Ontario bean breedingpopulations, BMC Plant Biology 11: 52

SHI J., LI R., QUI D., JIANG C., LONG Y., MORGAN C., BACROFT I., Z HAO J. and MENG J. (2009), Unraveling the complex trait of crop yield with quantitative trait loci, Genetics 182: 851 - 861

SINGH R.K. and CHADHARY S.D. (1985), Biometrical methods in quantitative genetic analysis, Kalian publishers, New Delhi

SINGH S.P. and MUÑOZ C.G. (1999), Resistance to common bacterial blight among *Phaseolus* species and common bean improvement, Crop Science 39: 80 -89

SINGRUN C. and SCHENKEL W. (2003), Fingerprinting of bambara groundnut germplasm with molecular marker, In: Proceedings of the International Bambara groundnut Symposium, International Cooperation with Developing Countries, Botswana college of Agriculture, Botswana 8-12 August, 2003, Gaborone, Botswana pp.161 - 170

SIOL M., PROSPERI J.M., BONNIN I. and RONFORT J. (2008), How multilocus genotypic pattern helps to understand the history of selfing

populations: a case study of in *Medicago truncatula*, Heredity 100: 517 - 525

SMARTT J. (1985), Evolution of grain legumes. II. Old and New world pulses of lesser economic importance, Experimental Agriculture 21: 1 - 18

SMITH H.F. (1936), A discriminant function for plant selection, Annual Eugenic 7: 240 - 250

SOMTA P., SOMMANAS W. and SRINIVES P. (2009), Molecular diversity assessment of AVRDC – The World Vegetable Center elite – parental mungbeans, Breeding Science 59: 149 -157

SONGOK S., FERGUSON M., MUIGAI A.W. and SILIM S. (2010), Genetic diversity in pigeonpea (*Cajanus cajan* (L.) Millsp.) landraces as revealed by simple sequencerepeat markers, African Journal of Biotechnology 9: 3231 - 3241

SPOONER D., VAN TREUREN R. and DE VINCENTE M.C. (2005), Molecular markers for genebank management. IPGRI Technical Bulletin NO 10. International Plant Genetic Resources, Institute, Rome, Italy

STADLER F. (2009), Analysis of differential expression under water-deficit stress and genetic diversity in bambara groundnut (*Vigna subterranea* (L.) Verdc.) using novel high-throughput technologies, PhD thesis, TECHNISCHE UNIVESITÄT MÜNCHEN (TUM)

STRUSS D. and PLIESKE J. (1998), The use of microsatellites markers for detection of genetic diversity in barley population, Theoretical and Applied Genetics 97: 308 - 315

STODART J.B., MACKAY M. and RAMADAN H. (2005), AFLP and SSR analysis of genetic diversity among landraces of bread wheat (*Triticum aestivum* L. em Thell) from different geographic regions, Australian Journal of Agricultural Research 56: 691 - 697

STODART J.B., MACKAY M.C. and RAMADAN H. (2007). Assessment of molecular diversity in landraces of bread wheat (*Triticum aestivum* L.) held in an exsitu collection with Diversity Array Technology (DArT), Australian Journal of Agricultural Research 58: 1174 -1182

STREFELER M.S. and WEHNER T.C. (1986), Comparison of six methods of multiple-trait selection for fruit yield and quality traits in three fresh-market cucumber populations, Journal of the American Society for Horticultural Science 111: 792 - 798

SUWANPRASERT J, TOOJINDA T., SRINIVES P. and CHANPRAME S. (2006), Hybridization technique of bambara groundnut (*Vigna subterranea*), Breeding Science 56: 125 - 129

SWAMY M.B.P., UPADHAYAYA H.D., GOUDAR P.V.K., KULLAISWAMY B.Y. and SINGH S. (2003), Phenotypic variation for agronomic

characters in a groundnut core collection for Asia, Field Crops Research 84: 359 -371

SWANEVELDER C.J. (1998), Bambara- Food for Africa bambara groundnut (*Vigna Subterranea*), National Department of Agriculture, Republic of South Africa.

TABACHNICK B.G. and FIDEL S. L. (2007), Using multivariate statistics, PearsonInternational Edition, Boston

TANTASAWAT P., TRONGCHUEN J., PRAJONGJAI T., THONGPAE T., PETKHUM C., SEEHALAK W. and MACHIKOWA T. (2010), Variety identification and genetic relationships of mungbean and blackgram in Thailand based on morphological characters and ISSR analysis, African Journal of Biotechnology 9: 4452 -4464

TARDIN D.F., PEREIRA M.G., GABRIEL A.P.C., DO AMARAL JÚNIOR A.T. and DE SOUZA FILHO A. (2007), Selection index for and molecular markers in reciprocal recurrent selection in maize, Crop Breeding and Applied Biotechnology 7: 225 - 233

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

TIBE O., AMARTEIFIO J.O. and NJOGU R.M. (2007), Trypsin inhibitor activity and condensed tannin content in bambara groundnut (*Vigna subterranea* (L.) Verdc) grown in Southern Africa, Journal of Applied Science and Environmental Management 11: 159 - 164

TOSTI N. and NEGRI V. (2005), On-going on-farm microevolutionary process in neighbouring Cowpea landraces revealed by molecular markers, Theoretical and Applied Genetics 110: 1275 - 1283

TSILO T.J., HARELAND G.A., SIMSEK S., CHAO S. and ANDERSON J.A. (2010), Genome mapping of kernel characteristics in hard red spring wheat breeding lines, Theoretical and Applied Genetics 121: 717 - 730

UGURU M.I. and EZEH N. E. (1997), Growth, nodulation and yield of bambara groundnut (*Vigna subterranea* (L.) Verdc.) on selected Nigerian soils, Journal of Science of Food and Agriculture 73: 377 - 382

UPADHYAYA H.D. (2003), Phenotypic diversity in groundnut (*Arachis hypogaea*L.) core collection assessed by morphological and agronomic evaluations, Genetic Resources and Crop Evolutions 50: 539 - 550

UPADHYAYA D. H., DWIVEDI L.S., BAUM M., VARSHNEY R.K., UDUPA M.S., GOWDA L.L.C., HOISINGTON D. and SINGH S. (2008), Genetic structure, diversity and allelic richness in composite collection and reference set in chickpea (*Cicer arientum* L.) BMC Plant Biology 8: 106

VAN OOSTERHOUT C., HUTCHISON W. F., WILLS D.P.M. and SHIPLEY P. (2004), Micro-Checker: software for identifying and correcting genotyping errors in microsatellites data, Molecular Ecology Notes 5: 535 - 538

VANGE T. and MOSES E. O. (2009), Studies on genetic characteristics of pigeonpea germplasm at Otobi Benue State of Nigeria, World Agricultural Sciences 5: 714 -719

VARSHNEY R.K., BERTIOLI J.D., MORETZSOHN M.C., VADEZ V., KRISHNAMURTHY L., ARUNA R., NIGAM S.N., MOSS B.J., SEETHA K., RAVI K., HE G., KNAPP S.J. and HOISINGTON D.A. (2009), The first SSR-based genetic linkage map for cultivated groundnut (*Arachis hypogaea* L.), Theoretical and Applied Genetics 118: 729 - 739

VARSHNEY R.K., HORRES R., MOLINA C., NAYAK S., JUNGMANN R., SWANY R., WINTER P., JAYASHREE B., KAHL G. and HOISINGTON D.A, (2007), Extending the repertoire of microsatellite markers for genetic linkage mapping and germplasm screening in chickpea, Semi- Arid Tropic Journal 5: 1- 3

VARSHNEY R.K., PENMETSA R.V., DUTTA S., KULWAL P.L., SAXENA R.K., DATTA S., SHARMA T.R., ROSEN B., CARRASQUILLA-GARCIA N., FARMER A.D., DUBEY A., SAXENA B.K., GAO J., FAKRUDIN B., SINGH M.N., SINGH B.P., WANJARI K.B., YUAN M., SRIVASTAVA R.K., KILIAN A., UPADHYAYA H.D., MALLIKARJUNA N.,TOWN C.D., BRUENING G.E., HEG., MAY G.D., MCCOMBIE R., JACKSON A.S., SINGH N.K. and COOK D.R. (2009), Pigeonpea genomics initiative (PGI): an international effort to improve crop productivity (*Cajanus cajan*L.), Molecular Breeding 26: 393 - 408

VARSHNEY R.K., CLOSE T.J., SINGH N.K., HOISINGTON D.A and COOK D.R. (2009), Orphan legume crops enter the genomics era! Current Opinion in Plant Biology 12: 202 - 210

VELASQUEZ V.L. and GEPTS P. (1994), RFLP diversity of common bean (*Phaseolus vulgaris*) in its centres of origin, Genome 37: 256 -263

VIJAYKUMAR A., SAINI A. and JAWALI N. (2009), Phylogenetic analysis of subgenus *Vigna* species using nuclear ribosomal RNA ITS: evidence of hybridisation among *Vigna unguiculata* subspecies, Journal of Heredity 2: 177 - 178

VEIRA A.E., DE CARVAHLO F.I., BERTAN I., KOPP M.M.,ZIMMER D.P., BENIN G., DE SILVA G.A.J., HARTWIG I., MALONE G. and DE OLIVIERA C.A (2007), Association between genetic distances in wheat (*Triticum aestivum* L.) as estimated by AFLP and morphological markers, Genetics and molecular Biology 30: 392-399

VOS P., HOGERS R., BLEEKER M., REIJANS M, VAN DE LEE T., HORNES M., FRIJTERS A., POT J., PELEMAN J. and KUIPER M. (1995), AFLP: a new technique for DNA fingerprinting, Nucleic Acids Research 11: 4407 - 4014

WANG L.M.,BARKLEY A.N., GILLASPIE A.G and PEDERSON A.G. (2008), Phylogenetic relationships and genetic diversity of the USDA *Vigna* germplasm collection revealed by gene-derived markers and sequencing, Genetics Research 90: 467 -480

WANG T.C., YANG D.X., CHEN X.D., YU L.S., LIU Z.G., TANG Y.Y. and XU Z.J. (2007), Isolation of simple sequence repeat from groundnut, Electronic Journal of Biotechnology 10: 3

WANG X., RINEHART T.A., WADL P.A., SPIERS J.M., HADZIABDIC D., WINDHAM M.T., and TRIGIANO R.N. (2009), A new electrophoresis technique to separate microsatellite alleles, African Journal of Biotechnology 8: 2432 - 2436

WAGNER A.P., CREEL S. and KALINOWSKI S.T. (2006), Estimating relatedness and relationships using microsatellites loci with null alleles, Heredity 97: 336 -345

WAMBETE J. and MPOTOKWANE S. (2003), Investigating opportunities for bambara groundnut in the development of weaning foods. In. Proceedings of the International bambara groundnut Symposium, Botswana College of Agriculture, 8 – 12, August 2003 pp.273 - 279

WASIKE S., OKORI P. and RUBAIHAYO P.R. (2005), Genetic variability and relatedness of the Asian and African pigeonpea as revealed by AFLP, African Journal of Biotechnology 4: 1228 - 1233

WEIR B. S. (1996), Genetic Data Analysis II, Methods for Discrete Population Genetic Data, Sinauer Associates, Inc Publishers, Massachusetts, USA.

WEIR B.S. (1990), Genetic data analysis: methods for discrete population geneticdata, Sinauer: Massachusetts

WEIR B.S. and COCKERAM C. (1984), Estimating F-statistics for the analysis of population structure, Evolution 38: 1358 - 1370

WENZL P, CARLING J, KUDRNA D, JACCOUD D, HUTTER E, KLEINHOFS A and KILIAN A (2004), Diversity Arrays Technology (DArT) for whole-genome profiling of barley, Proceedings of the National Academy of Sciences of the United States of America 101: 9915 - 9920

WIGGLESWORTH D.J. (1996), The potential for genetic improvement of bambara groundnut (*Vigna subterranea* (L.) Verdc.) Botswana, In Proceedings of the International Bambara groundnut Symposium, University of Nottingham, UK, 23- 25 July, 1996 pp.181 - 191

WILLIAM J.G.K., KUBELIK A.R., LIVAK K.J., RAFALSKI J.A and TINGEY S.V (1990), DNA polymorphism amplified by arbitrary primers are useful as genetic markers, Nucleic Acids Research 18: 6531 - 6535

WRIGHT S. (1978), Evolution and genetics of population no. 14, variability within and among natural populations. University of Chicago press, Chicago

www.bioversityinternational.org African Vigna. Accessed 12.4.2011

www.icrisat.org Accessed 21.4.2011

www.iita.org) Accessed 21.4.211

www.spgrc.org.zm accessed 1 June, 2011

XU Y., MCCOUCH S.R. and ZHANG Q. (2005), How can we use genomics to improve cereals with rice as a reference genome? Plant Molecular Biology 59: 7 - 26

YAGHATIPOOR A. and FARSHDFAR E. (2007), Non-Parametric estimation and component analysis of phenotypic stability in chickpea (*Cicer arietinum*L.), Pakistan Journal of Biological Sciences 10: 2646 - 2652

YANG S., PANG W., ASH G., HARPE J., CARLING J., WENZEL P., HUTTNER E., ZONG X.and KILIAN A. (2006), Low level of genetic diversity in cultivated pigeonpea compared to its wild relatives revealed by diversity arrays technology, Theoretical and Applied Genetics 113: 585 - 595

YANG S.S., VALDÉS-LÓPEZ O., XU W.W., BUCCIARELLI B., GRONWALD J .W., HERNÁNDEZ G., and VANCE P.C. (2010), Transcript profiling of common bean (*Phaseolus vulgaris* L.) using the GeneChip®Soybean genome array: optimizing analysis by masking biased probes, BMC Plant Biology 10: 85

YAP V.M. and NELSON R.J. (1996), WinBoot: A program for performing bootstrap analysis of binary data to determine the confidence limits of UPGMA based dendrograms, International Rice Institute (IRRI), Manila, Philippines

YU K, PARK S.J, POYSA V. and GEPTS P. (2000), Integration of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris* L.), The Journal of Heredity 91: 429 - 424

YU W-J., DIXIT A., MA H-K., CHUNG W-J. and PARK J-Y. (2009), A study on relative abundance, composition and length of microsatellites in18 underutilized crop species, Genetic Resources and Crop Evolution 56: 237 - 246

ZANE L., BARGELLONI L. and PATARNELLO T. (2002), Strategies for microsatellites isolation: a review, Molecular Ecology 11: 1 – 16

ZANNOU A., KOSSOU D.K. AHNCHÉDÉ A., ZOUNDJIHÉKPON J., AGBICODO E., STRUIK P.C and SANNIN A (2008), Genetic variability of cultivated cowpea in Benin assessed by random amplified polymorphic DNA, African Journal of Biotechnology 7: 4407 - 4417

ZEVEN A. C. (1998), Landraces: A review of definitions and classifications, *Euphytica* 104: 127 – 139

ZHANG Y.L., MARCHAND S., TINKER N.A.and BELZILE F. (2009), Population structure and linkage disequilibrium in barley assessed by DArT markers, Theoretical and Applied Genetics 119: 43 - 52

ZHU H., CHOI H., COOK. D. and SHOEMAKER C.R. (2005), Bridging model and crop legumes through comparative genomics, Plant Physiology, 137: 1189 -1196

ZHURAVLEV N.Y., REUNOVA G.D., KATS L.I., MUZAROK T.I. and BONDAR A. A.(2010), Genetic variability and population structure of endangered *Panax ginseng* in Russia Primorye, Chinese Medicine 5: 21

APPENDICES

Appendix 1: Preparations of standard solutions

1.0M Tris pH 8.0
12.11 g Tris base dissolved in 1.0M HCl until pH 8.0
Final volume is adjusted to 100 mL with water

0.5 EDTA dissolved in 75 mL water
18.61 g EDTA dissolved in 75 mL water
2 g of NaOH pellets dissolved
pH adjusted to 8.0 with 1.0M NaOH solution
Final volume adjusted to 100 mL with water

5.0M NaCl
29.22 g NaCl dissolved in 70mL water
Final volume adjusted to 100 mL

50 x TAE DNA Electrophoresis buffer
242 g Tris base
57.1 mL glacial acetic acid
100 mL 0.5M EDTA pH 8.0
Final volume brought to 1 Litre

1 x TAE Buffer
20 mL 50x TAE buffer
Final volume adjusted to 1 litre

5 x TBE DNA Electrophoresis buffer
54 g Tris base
27.5 g Boric acid
20 mL 0.5M EDTA pH 8.0
Final volume brought to 1Litre

0.5 x TBE Buffer
100 mL 0.5x TBE buffer
Final volume adjusted to 1Litre

6 x loading buffer (for DNA gels)
To make 30% glycerol (15 mL + 35 MQ water)
0.025 g Bromophenol Blue
0.025g Xylene Cyanol

Appendix 2: List of characterised 75 primers used in bambara groundnut diversity, and list of primer combinations used in multiple experiments.

75 microsatellites characterised on 24 genotypes selected based on Singruin and Schenkel, (2003)				12 markers used in (123) population structute	20 markers used in (105) genetic diversity	12 markers used in (5) genetic anlaysis Chapter
Marker	**Sequence**	**Primer length**	**Tm (°C)**	analysis Chapter five	analysis Chapter six	six
PRIMER1F	AACTTGCCATACGTGGAAGG	20	59.0		Marker 1	
PRIMER1R	ACACGCTGCATAATTCACCA	20	59.0			
PRIMER2F	CGTGGATACCCATACCGTCT	20	46.3			
PRIMER2R	TAAGTCCATTTTGTCCGATTGA	22	46.3			
PRIMER3F	TGATGAATGAATGCAAAGTAAGA	23	59.0			
PRIMER3R	TTGGCTCATTGCCTAGTTCA	20	59.0			
PRIMER4F	CATTGTCTCTGCCACCATTTT	21	57.6			
PRIMER4R	CAGACTGGGATTTGCATGTG	20	57.6			
PRIMER5F	CTGCTGTGGTGAGCTTTTGT	20	59.0			
PRIMER5R	CTCCTTGCAGCTAAGCGTCT	20	59.0			
PRIMER6F	TACGGTCCTACACGGGAAAC	20	59.0			
PRIMER6R	ACCTGTCCAGCCGCAATTA	19	59.0			
PRIMER7F	GTAGGCCCAACACCACAGTT	20	55.3	Marker 7	Marker 7	Marker 7
PRIMER7R	GGAGGTTGATCGATGGAAAA	20	55.3			
PRIMER8F	GGAAGAGTGCGTTTTGGTGT	20	57.6			
PRIMER8R	CTGTGTGGACCCCAGAAAAT	20	57.6			
PRIMER9F	CCAGGAGTGAGGAGTGAGAAA	21	57.6			
PRIMER9R	ATGCATTTTCAGGGTCCAAG	20	57.6			
PRIMER10F	TCAGTGCTTCAACCATCAGC	20	55.3		Marker 10	
PRIMER10R	GACCAAACCATTGCCAAACT	20	55.3			
PRIMER11F	TGGAGGTGGAAATGATAACG	20	59.0			
PRIMER11R	TCCACCTTCACCTGCACT	18	59.0			
PRIMER12F	GTCTTGCAGGAAGGTTCAGC	20	57.6			
PRIMER12R	CAGATTACACACGCGCACTT	20	57.6			
PRIMER13F	CATTGCACGTCATAGAATTTGG	22	55.3			
PRIMER13R	GGGTGAACTACACCACCTTCA	21	55.3			
PRIMER14F	TGGTGGTAGAGAATTGGAGGA	21	57.6			
PRIMER14R	CACACAGAAACACAAACACAGC	22	57.6			
PRIMER15F	AGGAGCAGAAGCTGAAGCAG	20	55.3	Marker 15	Marker 15	Marker 15
PRIMER15R	CCAATGCTTTTGAACCAACA	20	55.3			
PRIMER16F	CCGGAACAGAAAACAACAAC	20	57.6	Marker 16	Marker 16	Marker 16
PRIMER16R	CGTCGATGACAAAGAGCTTG	20	57.6			
PRIMER17F	CAAAGCAACACAAACGATGG	20	55.3			
PRIMER17R	ATAACCATTGGCCGATTGAC	20	55.3			
PRIMER18F	TCTGCCACATTTCGCATAAG	20	55.3			
PRIMER18R	CGCTTCAAATCCGATGTTCT	20	55.3			
PRIMER19F	AGGCAAAAACGTTTCAGTTC	20	55.3	Marker 19	Marker 19	Marker 19
PRIMER19R	TTCATGAAGGTTGAGTTTGTCA	22	55.3			
PRIMER20F	CCCTTCACATACACTTAAGAACCA	24	59.0			
PRIMER20R	CCTCTTCCACGAGAACAAGC	20	59.0			
PRIMER21F	CAAACTCCACTCCACAAGCA	20	57.6		Marker 21	
PRIMER21R	CCAACGACTTGTAAGCCTCA	20	57.6			
PRIMER22F	TCCCAAAATGGGACCAACTA	20	55.3			
PRIMER22R	ATCCGACTGATTAAGCCTAAAA	22	55.3			
PRIMER23F	CAGTAGCCATAATTTGCTATGAACA	25	55.3	Marker 23	Marker 23	Marker 23
PRIMER23R	CGAATCACCATTCAATACGC	20	55.3			
PRIMER24F	TTGGGTTGAATGGAAGTATGAA	22	55.3			
PRIMER24R	CAGAAGATCCCTTTCGACCA	20	55.3			
PRIMER25F	GCTGGAACTGATCCACCTTT	20	59.0			
PRIMER25R	ATGTAGCAGTGCCACCAACA	20	59.0			
PRIMER26R	CGCTCATTTTAACCAGACCTC	21	57.6			
PRIMER26R	CAAACAAACCAACGGAATGA	20	57.6			
PRIMER27F	ACACCGCCATCATGAGATTT	20	60.2			
PRIMER27R	CATTTCAGGATTTGGGAGGA	20	60.2			
PRIMER28F	CAATGCTTCAACCATCAACC	20	47.7			
PRIMER28R	AGTGTATGGATGCCCAGACC	20	47.7			
PRIMER29F	TCTGACGCAAGCAAGAAGAA	20	55.3			
PRIMER29R	GGTTCGATCGGAAATCTGAA	20	55.3			
PRIMER30F	AATGCAAGATTTTGGCTTGG	20	59.0		Marker 30	
PRIMER30R	CCCACTCAAACCATACACCA	20	59.0			
PRIMER31F	GCTAAGGTGGAGTGGTGGAA	20	57.6		Maker 31	
PRIMER31R	CAATCATCTTTTGCGCTTCA	20	57.6			
PRIMER32F	TTCACCTGAACCCCTTAACC	20	57.6		Marker 32	
PRIMER32R	AGGCTTCACTCACGGGTATG	20	57.6			
PRIMER33F	ACGCTTCTTCCCTCATCAGA	20	57.6			
PRIMER33R	TATGAATCCAGTGCGTGTGA	20	57.6	Marker 33	Marker 33	Marker 33

Appendix 2: Continued

75 microsatellites characterised on 24 genotypes based on Singruin and Schenkel, (2003)				12 markers used in (123) population structure analysis Chapter five	20 markers used in (105) genetic diversity analysis Chapter six	12 markers used in (5) genetic analysis Chapter six
Marker	**Sequence**	**Primer length**	**Tm (^{o}C)**			
PRIMER34F	CATGTTGAAACCCGATGTCC	20	57.6			
PRIMER34R	ACCTCCTGGTGCATCTATGG	20	57.6			
PRIMER35F	CGTGCCTACCTTCGACTACC	20	55.3			
PRIMER35R	CGGTGGAAACTCCGATTAAA	20	55.3			
PRIMER36F	CGAAAGAACTTGACAGGCAGA	21	55.3			
PRIMER36R	TCAGCAGAATGATCCTCCAA	20	55.3			
PRIMER37F	CCGATGGACGGGTAGATATG	20	55.3	Marker 37	Marker 37	Marker 37
PRIMER37R	GCAACCCTCTTTTTCTGCAC	20	55.3			
PRIMER38F	TCACACTTGCAATGGTGCTT	20	53.3			
PRIMER38R	TCGTTGTTTCTCTTTTCATTGC	22	53.3			
PRIMER39F	TCGTACCGAATCACCATTCA	20	55.3			
PRIMER39R	CAGTAGCCATAATCTGCTATGAACA	25	55.3			
PRIMER40F	TGGACCATACCCATCTTCAAA	21	55.3			
PRIMER40R	TCAGGGACATTACCCAGACC	20	55.3			
PRIMER41F	ACACCGCCATCATGAGATTT	20	59			
PRIMER41R	CAGGATTTGCGAGGAGAGAG	20	59			
PRIMER42F	CCTTTCAGCTTCTCCAAACG	20	57.3			
PRIMER42R	TCAACCCACACAGAATCGAA	20	57.3			
PRIMER43F	ACTTGATGCTACCGAGAGAGAG	22	57.6			
PRIMER43R	AGGCTCCAACAATGCGATAG	20	57.6			
PRIMER44F	TGTGGGCGAAAATACACAAA	20	59.7	Marker 44	Marker 44	Marker 44
PRIMER44R	TCGTCGAATACCTGACTCATTG	22	59.7			
PRIMER45F	CGTGGATACCCATACCGTCT	20	55.3			
PRIMER45R	AAGTCCATTTTGTCCGATTGA	21	55.3			
PRIMER46BF	TTTGTCCGGTTCAACTGAATTA	22	51.4			
PRIMER46BR	TTGAAGATGGGTATGGTCCAC	21	51.4			
PRIMER47F	ACCCATTGCACGTCATAGAA	20	59			
PRIMER47R	GGGTGAACTACACCACCTTCA	21	59			
PRIMER48F	TACCTGCATTCGGGACAGTT	20	59			
PRIMER48R	TTCACTCTTTCTTGATCACATGC	23	59			
G33AB4-D1F	TGCTTCTTCAAGGAGGAAGTAAGT	24	59.0			
G33AB4-D1R	ACAAACATACGCACAACAGAGAAT	24	59.0			
G111AB4-D2F	AGGTTATGAGGTAAGCATTTCAGG	24	59.7			
G111AB4-D2R	TCAGATTGCATAATTTGCTTGATT	24	59.7			
G185AB4-D3F	CTCCACTCCACAAGCAATAAACTA	24	49.4			
G185AB4-D3R	CCATTTGTAAACCAACGACTTGTA	24	49.4			
G194AB4-D4F	CCCTTCAACCCTAGTTGAGATAGA	24	55.3			
G194AB4-D4R	TCCTATTTCTTTCGGCATATTTTT	24	55.3			
G196AB4-D5F	CCACGTTCTGGTTGTGAGTAGATA	24	49.4			
G196AB4-D5R	GTGCTTTCAGACCATTACTTGCTT	24	49.4			
G278AB4-D6F	TGGTTTTATAAATTGGGATTTTGG	24	57.6			
G278AB4-D6R	ACCTATAATTCACGCACACACG	22	57.6			
G331AB4-D7F	TCTTCTTTATTGGTGGACCATACC	24	45.5			
G331AB4-D7R	AAAACCAAGGACACAAATTCTAGC	24	45.5			
G372AB4-D8F	GCATCTTTACAGCAAGAGTTTCAA	24	59.0		Marker D8	
G372AB4-D8R	TGGATCTTCCTCATTGCAGTATAA	24	59.0			
G11-9-B2-D9F	ATCAAAATCAAGCAAATGAGA	21	53.3			
G11-9-B2-D9R	ACCTTTTACGCTCATTTTAACCAG	24	53.3			
G174B2-D10F	GTTTTAGGATCAAATGGTTTTGGA	24	59.0			
G174B2-D10R	TGCCTTTTATAATGATGTGCATTC	24	59.0			
G180B2-D11F	GAGGAAATAACCAAACAAACC	21	59.0	Marker D11	Marker D11	Marker D11
G180B2-D11R	CTTACGCTCATTTTAACCAGACCT	24	59.0			
G240-7-B2-D12F	TTTTGTTGTTGTATGAATCCAGTG	24	59.0			
G240-7-B2-D12R	CCTCATCAGACGCTCATCATT	21	59.0			
G326B2-D13F	AGAGGTGGAGGGGTTGGAT	19	59.0			
G326B2-D13R	CCTCAATAGCTGAATCCATTTCTC	24	59.0			
G240-9-B2-D14F	GAACGAAGCCAGGATAATGATAGT	24	59.0	Marker D14	Marker D14	Marker D14
G240-9-B2-D14R	CGAAAGCGACAACTCACTACTAAA	24	59.0			
G358B2-D15F	TGACGGAGGCTTAATAGATTTTTC	24	59.0		Marker D15	
G358B2-D15R	GACTAGACACTTCAACAGCCAATG	24	59.0			
mBam2co80	GAGTCCAATAACTGCTCCCGTTTG	24	59.0			
mBam2co80	ACGGCAAGCCCTAACTCTTCATTT	24	59.0	mBam2co80	mBam2co80	mBam2co80
E1F	TGTTGTGTCAACAAATTAAGATGAG	25	47.7			
E1R	ATGCTTCAAACTGTCCCTGA	20	47.7			
E2F	CATGTTCGTAATGATTTGAAGTGTT	25	59.0			
E2R	GCCAAAACAATATCTTCAAGAGG	23	59.0			
E3F	GGACGGAGTCCTTCAAACAA	20	47.7			
E3R	CCTTGTGCATACCCATAGTATCC	23	47.7			
E4F	CATGGCGAAGGAGGGCAGCGA	21	57.6			
E4R	AGCGATTACTGGGGTTGAGA	20	57.6			
E5F	CATGGAGTGCTATGTGGTGAT	21	51.4			
E5R	ATACGGTTGTGGCAGTGTCC	20	51.4			
E6F	CATGGACGAGGGATTAGCGCAG	22	47.7			
E6R	CCCTAGCCAAATGACCTACC	20	47.7			
E7F	CATGATTTGTTGTGATGATGAT	22	51.4	Marker E7	Marker E7	Marker E7
E7R	AACAACAAATGTACCAAAGAATCG	24	51.4			
E9F	CATGAGAAGGCCTTCTGATGAT	22	51.4			
E9R	CCACAAGTTCTTTTTATTCCCTTC	24	51.4			
E10F	CATGACTTTCTTCATTGGT	19	51.4			
E10R	TGCATTCCAATTAAATTCATAACAA	25	51.4			
E11F	CATGACCACCAGAGAAGATGT	21	47.7			
E11R	ATTCAGAATCCTCAAC	16	47.7			
E12F	CATGAAGGCGGAGACGGCGG	20	53.3			
E12R	CATGACCACCAGAGAAGATGT	21	53.3			

Appendix 3: Estimated repeat length of alleles and adjustment for the characterisation of 75 markers used in the analysis of 24 landraces

Landraces	Estimated repeat length	Adjustment
Primer 1	0.030	1.032
Primer 2	0.935	1.116
Primer 3	0.035	0.968
Primer 4	0.275	0.975
Primer 5	0.375	0.774
Primer 6	0.705	0.787
Primer 7	0.530	0.774
Primer 8	0.220	0.774
Primer 9	0.455	0.774
Primer 10	0.080	0.986
Primer 11	0.615	0.795
Primer 12	0.810	0.774
Primer 13	0.530	0.824
Primer 14	0.200	0.964
Primer 15	0.920	1.011
Primer 16	0.007	0.981
Primer 17	0.925	0.950
Primer 18	0.215	0.981
Primer 19	0.465	0.829
Primer 20	0.205	0.957
Primer 21	0.710	0.781
Primer 22	-0.035	1.060
Primer 23	-0.030	1.036
Primer 24	1.010	0.971
Primer 25	0.550	1.069
Primer 26	0.580	0.802
Primer 27	0.305	0.929
Primer 28	0.045	1.008
Primer 29	0.340	0.774
Primer 30	0.045	0.991
Primer 31	0.170	1.003

Appendix 3(continued)

Landraces	Estimated repeat length	Adjustment
Primer 32	0.030	0.985
Primer 33	0.980	1.031
Primer 34	0.665	0.777
Primer 35	0.460	0.903
Primer 36	-0.005	1.080
Primer 37	0.605	0.785
Primer 38	0.390	0.889
Primer 40	0.335	0.922
Primer 41	0.900	0.999
Primer 42	0.030	0.985
Primer 43	0.100	0.981
Primer 44	0.125	1.013
Primer 45	0.790	1.161
Primer 48	0.100	0.921
D1	-0.075	1.031
D2	0.130	0992
D3	0.680	0.774
D4	0.405	0.9089
D5	0.365	0.957
D6	0.115	0.975
D7	0.745	0.775
D8	0.105	1.079
D9	0.160	0.975
D10	0.385	0.842
D11	0.165	0.981
D12	0.035	0.985
D13	-0.020	1.100
D14	-0.015	1.027
D15	-0.005	1.021
E1	0.220	0.806
E3	0.565	0.775
E5	0.520	0.775
E7	0.440	0.866
E9	0.040	1.031
E10	0.160	0.957
E11	0.170	0.879
mBam2co80	0.345	0.859

Appendix 4: A comparison of Nei and Li, (1979) similarity estimates for DArT marker (upper) and SSR markers (bottom) matrices calculated using MVSP version 3.1 for the 24 bambara groundnut landraces.

	DodR	DodC	AS17	DipC	SwaziRed	Tiganicuru	Ramayana	LunT	VSSP6	Nav 4	Nav Red	MHNblack	S19/3	S19-3	Uniswa	SB16 5A	AHM968	NAM 1761/	Malawi 3	Tvsu 569	Tvsu 610	Tvsu 747	GabC	Tvsu 999
DodR	**1**	0.79	0.73	0.74	0.76	0.57	0.67	0.53	0.50	0.51	0.49	0.71	0.73	0.74	0.72	0.69	0.70	0.70	0.72	0.50	0.48	0.72	0.72	0.69
DodC	0.71	**1**	0.83	0.80	0.86	0.61	0.80	0.54	0.65	0.60	0.57	0.80	0.83	0.82	0.84	0.82	0.79	0.79	0.83	0.59	0.53	0.80	0.80	0.82
AS17	0.67	0.75	**1**	0.80	0.92	0.56	0.81	0.54	0.66	0.63	0.59	0.83	0.91	0.89	0.90	0.87	0.87	0.79	0.83	0.59	0.53	0.85	0.89	0.85
DipC	0.68	0.77	0.74	**1**	0.83	0.58	0.75	0.55	0.62	0.63	0.59	0.81	0.83	0.83	0.84	0.84	0.85	0.83	0.83	0.59	0.55	0.83	0.86	0.84
SwaziRed	0.66	0.68	0.67	0.68	**1**	0.57	0.80	0.53	0.62	0.61	0.58	0.82	0.89	0.87	0.88	0.84	0.86	0.80	0.84	0.59	0.51	0.85	0.87	0.84
Tiganicuru	0.68	0.68	0.63	0.62	0.61	**1**	0.59	0.63	0.57	0.70	0.68	0.57	0.56	0.59	0.57	0.56	0.58	0.60	0.57	0.71	0.72	0.60	0.57	0.60
Ramayana	0.69	0.77	0.74	0.73	0.73	0.67	**1**	0.56	0.66	0.58	0.57	0.77	0.82	0.81	0.82	0.81	0.81	0.74	0.86	0.63	0.55	0.83	0.82	0.79
LunT	0.63	0.69	0.70	0.64	0.59	0.67	0.68	**1**	0.52	0.59	0.61	0.51	0.55	0.55	0.59	0.53	0.55	0.54	0.57	0.57	0.61	0.53	0.55	0.53
VSSP6	0.62	0.69	0.64	0.60	0.59	0.69	0.66	0.65	**1**	0.67	0.63	0.62	0.67	0.66	0.67	0.66	0.69	0.65	0.66	0.66	0.54	0.63	0.69	0.66
Nav 4	0.57	0.54	0.58	0.51	0.50	0.57	0.54	0.56	0.59	**1**	0.74	0.66	0.62	0.61	0.63	0.63	0.65	0.66	0.61	0.73	0.59	0.59	0.64	0.64
Nav Red	0.68	0.63	0.60	0.62	0.62	0.65	0.64	0.67	0.66	0.63	**1**	0.55	0.57	0.57	0.58	0.55	0.60	0.58	0.57	0.75	0.68	0.56	0.60	0.61
MHNblack	0.70	0.65	0.65	0.65	0.64	0.60	0.64	0.63	0.63	0.57	0.77	**1**	0.84	0.82	0.80	0.81	0.81	0.84	0.82	0.60	0.48	0.80	0.83	0.82
S19/3	0.68	0.67	0.69	0.72	0.65	0.63	0.71	0.61	0.64	0.61	0.73	0.75	**1**	0.95	0.89	0.88	0.90	0.83	0.83	0.62	0.53	0.85	0.91	0.85
S19-3	0.67	0.68	0.65	0.68	0.66	0.67	0.72	0.60	0.68	0.57	0.67	0.68	0.87	**1**	0.89	0.83	0.88	0.82	0.82	0.63	0.52	0.85	0.89	0.85
Uniswa	0.64	0.64	0.68	0.62	0.55	0.62	0.58	0.60	0.60	0.60	0.58	0.64	0.68	0.70	**1**	0.87	0.87	0.82	0.86	0.59	0.54	0.84	0.88	0.84
SB16 5A	0.59	0.64	0.66	0.63	0.57	0.66	0.62	0.62	0.83	0.56	0.67	0.60	0.70	0.69	0.66	**1**	0.84	0.80	0.82	0.62	0.54	0.79	0.86	0.81
AHM968	0.64	0.76	0.76	0.70	0.60	0.65	0.70	0.66	0.65	0.57	0.61	0.65	0.71	0.70	0.67	0.65	**1**	0.83	0.83	0.65	0.55	0.85	0.99	0.84
NAM 1761/3	0.62	0.69	0.58	0.60	0.60	0.66	0.66	0.59	0.62	0.50	0.57	0.58	0.56	0.57	0.56	0.56	0.66	**1**	0.84	0.63	0.55	0.82	0.83	0.85
Malawi 3	0.70	0.77	0.65	0.71	0.62	0.60	0.76	0.63	0.60	0.52	0.63	0.68	0.69	0.67	0.62	0.57	0.74	0.67	**1**	0.59	0.51	0.88	0.83	0.87
Tvsu 569	0.70	0.77	0.75	0.72	0.64	0.71	0.73	0.70	0.69	0.52	0.65	0.59	0.69	0.67	0.58	0.67	0.74	0.66	0.72	**1**	0.62	0.61	0.64	0.63
Tvsu 610	0.61	0.57	0.56	0.56	0.53	0.63	0.61	0.57	0.61	0.49	0.62	0.54	0.60	0.68	0.59	0.58	0.61	0.57	0.60	0.59	**1**	0.54	0.54	0.54
Tvsu 747	0.66	0.67	0.68	0.68	0.58	0.61	0.66	0.62	0.63	0.57	0.64	0.67	0.66	0.63	0.60	0.65	0.64	0.60	0.67	0.67	0.61	**1**	0.85	0.86
GabC	0.61	0.71	0.74	0.75	0.61	0.59	0.65	0.61	0.64	0.55	0.63	0.60	0.73	0.70	0.61	0.66	0.74	0.58	0.71	0.72	0.62	0.74	**1**	0.84
Tvsu 999	0.64	0.71	0.62	0.71	0.63	0.59	0.72	0.62	0.62	0.45	0.61	0.65	0.66	0.70	0.57	0.63	0.69	0.62	0.76	0.67	0.67	0.69	0.73	**1**

Appendix 5: Mean values for the characters of the 35 landraces grown in the agronomy bay experiment (UK)

Landraces	DAE	DAF	LNO	SPRD	LL	LW	LA	PHT	ITN	PTL	PITN	PTLL
3Acc9NGA	9	39	82	12	7	3	2956	28	1	13	9	2
4Acc144NGA	10	41	49	34	10	4	2953	34	4	15	4	2
6Acc289BEN	12	41	57	13	8	3	2587	28	2	14	6	2
10Acc1276CAF	12	50	67	20	9	4	4316	38	3	22	8	3
20Acc118BFA	10	39	54	12	9	3	2512	30	2	13	7	2
30Acc476CMR	10	45	62	14	9	4	3391	30	1	15	11	3
33Acc484CMR	11	46	68	14	9	3	3470	32	2	16	10	2
40Acc563CMR	9	47	82	16	9	4	4896	33	2	14	8	2
45Acc231GHA	9	40	72	19	8	3	3318	24	1	14	10	2
48Acc790KEN	10	43	205	23	8	4	10369	36	2	18	8	3
49Acc793KEN	8	41	50	22	8	3	2281	34	3	16	7	3
50Acc792ZWE	12	49	49	12	8	3	1887	28	2	14	8	2
56Acc89MLI	10	39	35	11	8	3	1357	29	1	14	10	2
60Acc32NGA	10	42	54	14	8	3	2613	31	2	14	7	2
69Acc286NGA	10	47	47	11	9	4	2953	32	1	15	10	2
70Acc329NGA	11	45	84	12	8	3	4049	31	2	15	8	2
74Acc335NGA	10	43	52	12	8	4	2944	32	2	16	8	1
76Acc390SDN	13	41	39	9	9	3	1955	32	1	14	14	2
81Acc385TZA	8	38	104	27	8	4	6713	35	3	18	7	3
84Acc696ZMB	11	43	150	23	9	4	9939	38	3	21	8	4
85Acc754ZMB	8	45	198	25	8	3	9997	40	3	22	8	3
88-AHM753NAM	10	47	81	23	8	4	4347	32	3	16	5	2
90-S19-3NAM	9	39	55	19	9	4	3364	30	2	16	7	2
91-UNISRSWA	10	39	105	16	9	4	6059	36	2	17	9	3
92-AHM968NAM	10	39	79	22	8	4	4103	36	3	20	7	2
95-DODRTZA	10	39	68	34	8	4	3822	36	4	17	4	2
99-SB4-2NAM	10	38	116	32	8	4	6500	35	3	19	6	2
100-SB16ANAM	10	41	116	28	8	4	6429	36	3	19	8	2
104-S-1913NAM	10	39	65	21	9	4	4240	32	3	16	7	2
105-MHNblackAM	9	41	89	32	10	4	6794	45	5	25	6	2
109-BOTS1	11	41	42	11	9	4	2393	34	2	16	11	3
113-BOTS5	8	41	83	28	9	3	4557	39	2	19	9	2
117-VSSP6CMR	11	46	51	19	9	3	2704	29	2	13	6	2
118-Ramayana-IND	12	47	109	33	10	5	9583	39	3	22	7	3
119-Hybrid	10	43	62	13	10	4	4751	39	2	22	13	3

DAE: days to emergence, DAF: days to 50% flowering, LNO: number of leaves per plant, SPRD: plant spread/canopy, LL: leaflet length, LW: leaflet width, LA: leaf area, PHT: plant height, ITN: Internode length, PTL: petiole length, PITN: petiole-internode ratio, PTL: petiolule length, PNL: penduncle length, STEM: number of stems, DAM: days to maturity, SDW: shoot dry weight, POD: pod number per plant, PDW:Pod dry weight, PODL: pod lenght, PODW:pod width, SNO: seed number plant, SL: seed length, SW: seed width, SWE: seed weight

Appendix 5: continued

Landraces	PNL	STEM	DAM	SDW	POD	PDW	PODL	PODW	SNO	SL	SW	SWE
3Acc9NGA	2	8	152	17	36	32	19	12	54	10	8	20
4Acc144NGA	3	7	160	30	46	35	19	12	50	12	8	16
6Acc289BEN	3	9	151	11	17	58	21	10	25	10	7	8
10Acc1276CAF	2	7	161	36	46	56	20	13	52	11	8	21
20Acc118BFA	2	12	156	17	34	43	21	12	43	11	8	21
30Acc476CMR	2	9	153	24	54	48	21	13	60	11	8	26
33Acc484CMR	3	10	152	23	50	27	18	11	56	10	7	22
40Acc563CMR	3	11	155	34	20	27	21	15	25	12	10	13
45Acc231GHA	2	7	151	36	40	40	18	12	33	10	8	13
48Acc790KEN	2	14	160	68	106	34	20	11	105	11	8	32
49Acc793KEN	3	12	156	20	52	24	18	13	55	11	9	25
50Acc792ZWE	2	6	154	15	33	20	19	12	38	10	9	14
56Acc89MLI	2	7	148	10	28	19	21	11	40	11	9	21
60Acc32NGA	3	8	153	20	22	45	20	12	30	11	9	10
69Acc286NGA	2	6	152	13	10	36	16	10	15	9	7	3
70Acc329NGA	2	8	151	29	11	20	19	12	12	9	7	4
74Acc335NGA	3	7	156	23	20	26	21	14	26	12	9	11
76Acc390SDN	2	10	139	11	30	12	18	11	39	10	8	16
81Acc385TZA	2	9	160	43	86	17	20	13	102	12	9	46
84Acc696ZMB	2	14	143	54	35	4	20	11	50	10	8	13
85Acc754ZMB	2	11	160	76	106	7	18	12	94	11	8	29
88-AHM753NAM	2	12	160	34	121	17	16	11	133	10	8	36
90-S19-3NAM	2	9	143	22	48	21	15	11	48	10	8	21
91-UNISRSWA	2	14	160	53	105	60	19	13	107	11	9	49
92-AHM968NAM	2	10	161	36	87	23	18	13	96	10	8	32
95-DODRTZA	4	14	160	34	79	44	21	13	86	13	10	47
99-SB4-2NAM	2	10	157	47	100	47	15	11	105	10	8	29
100-SB16ANAM	2	12	155	48	67	27	17	12	64	11	8	25
104-S-1913NAM	2	13	161	37	92	68	18	12	103	11	9	46
105-MHNBlackNAM	3	14	160	63	81	47	22	14	76	13	9	34
109-BOTS1	2	8	156	21	47	60	21	13	53	12	9	33
113-BOTS5	2	10	160	32	84	39	18	12	77	10	9	33
117-VSSP6CMR	3	9	160	21	31	20	18	12	30	10	8	13
118-Ramayana-IND	3	8	161	81	48	27	18	14	50	11	9	19
119-HYBRID	2	11	161	30	31	17	18	11	40	11	9	13

DAE: days to emergence, DAF: days to 50% flowering, LNO: number of leaves per plant, SPRD: plant spread/canopy, LL: leaflet length, LW: leaflet width, LA: leaf area, PHT: plant height, ITN: Internode length, PTL: petiole length, PITN: petiole-internode ratio, PTL: petiolule length, PNL: penduncle length, STEM: number of stems, DAM: days to maturity, SDW: shoot dry weight, POD: pod number per plant, PDW:pod dry weight, PODL: pod lenght, PODW:pod width, SNO: seed number plant, SL: seed length, SW: seed width, SWE: seed weight

Appendix 6: Mean values for the characters of the 34 lines grown in the field experiment (Botswana).

Lines	DAE	DAF	LNO	SPRD	LL	LW	LA	PHT	ITN	PTL	PITN	PTLL
3Acc9NGA	16	57	208	27	6	2	4329	25	2	14	6	2
4Acc144NGA	15	53	85	49	8	3	2878	26	5	13	4	2
6Acc289BEN	16	51	113	19	7	3	3391	23	2	12	7	2
10Acc1276CAF	18	66	67	15	7	3	2157	26	2	102	6	2
20Acc118BFA	16	54	164	17	6	2	3988	25	2	12	8	2
30Acc476CMR	17	58	108	19	7	2	2861	26	2	12	7	1
33Acc484CMR	16	62	119	21	6	2	2811	28	2	14	8	2
40Acc563CMR	16	63	119	22	7	2	3614	29	2	16	10	2
45Acc231GHA	15	60	73	14	6	2	1446	22	1	10	9	2
48Acc790KEN	13	65	142	23	7	3	4586	29	1	15	10	2
50Acc792ZWE	14	61	147	25	6	2	3059	23	2	13	6	1
56Acc89MLI	16	52	105	13	4	2	1614	23	1	12	10	1
60Acc32NGA	14	51	125	23	6	2	2583	23	2	10	6	1
69Acc286NGA	11	59	57	29	10	4	4319	33	2	17	8	2
70Acc329NGA	18	59	23	17	5	2	387	25	1	11	8	2
74Acc335NGA	17	61	126	21	7	3	3786	26	2	12	7	1
76Acc390SDN	14	51	153	29	8	3	5986	32	2	16	10	2
81Acc385TZA	13	63	174	31	6	3	5437	28	3	17	7	2
84Acc696ZMB	12	56	121	26	8	3	4343	30	2	16	9	3
85Acc754ZMB	14	64	143	21	7	3	4098	32	2	18	11	3
88-AHM753NAM	14	54	118	26	7	3	3879	27	5	15	5	2
90-S19-3NAM	17	53	98	33	8	3	4019	30	3	16	6	2
91-UNISRSWA	15	53	88	18	7	2	2442	30	2	13	6	2
92-AHM968NAM	16	54	82	26	7	2	2062	26	2	14	9	2
95-DODRTZA	13	47	104	50	7	3	3152	32	4	16	5	2
99-SB4-2NAM	17	58	99	22	7	3	3256	29	2	15	8	2
100-SB16ANAM	17	56	95	25	7	3	3428	32	3	15	5	2
104-S-1913NAM	14	48	87	22	6	3	2715	28	3	13	5	1
105-MHNBlackNAM	14	64	94	24	8	3	3306	35	3	17	7	2
109-BOTS1	13	54	124	25	7	3	3982	28	2	14	7	2
113-BOTS5	15	47	120	18	8	2	3865	33	2	18	7	2
117-VSSP6CMR	16	64	86	30	7	2	2252	25	2	12	6	2
118-Ramayana-IND	16	61	97	19	6	2	2609	28	2	10	4	1
119-HYBRID	16	54	83	17	7	2	2383	30	2	16	8	2

DAE: days to emergence, DAF: days to 50% flowering, LNO: number of leaves per plant, SPRD: plant spread/canopy, LL: leaflet length, LW: leaflet width, LA: leaf area, PHT: plant height, ITN: Internode length, PTL: petiole length, PITN: petiole-internode ratio, PTL: petiolule length, PNL: penduncle length, STEM: number of stems, DAM: days to maturity, SDW: shoot dry weight, POD: pod number per plant, PDW:pod dry weight, PODL: pod lenght, PODW:pod width, SNO: seed number plant, SL: seed length, SW: seed width, SWE: seed weight

Appendix 6: Continued

Lines	PNL	STEM	DAM	SDW	POD	PDW	PODL	PODW	SNO	SL	SW	SWE
3Acc9NGA	3	8	138	36	7	3	16	10	7	10	7	2
4Acc144NGA	2	9	138	31	8	19	19	11	9	10	7	13
6Acc289BEN	2	8	140	29	9	4	20	9	7	11	8	2
10Acc1276CAF	1	8	144	45	8	5	25	11	8	10	7	2
20Acc118BFA	2	5	127	22	11	10	18	12	12	12	9	7
30Acc476CMR	1	7	138	38	13	7	18	11	14	11	8	4
33Acc484CMR	2	8	144	27	23	16	21	12	22	11	9	8
40Acc563CMR	2	11	*	45	*	*	*	*	*	*	*	*
45Acc231GHA	1	7	*	19	*	*	*	*	*	*	*	*
48Acc790KEN	2	8	155	61	11	4	16	10	11	11	8	3
50Acc792ZWE	1	12	*	48	*	*	*	*	*	*	*	*
56Acc89MLI	1	8	134	22	6	3	17	11	6	9	7	1
60Acc32NGA	2	11	127	39	12	7	23	13	13	13	10	4
69Acc286NGA	1	15	*	16	*	*	*	*	*	*	*	*
70Acc329NGA	1	7	*	14	*	*	*	*	*	*	*	*
74Acc335NGA	2	6	139	23	11	10	22	12	10	12	8	4
76Acc390SDN	2	7	142	83	10	7	18	11	12	12	8	4
81Acc385TZA	3	9	146	103	10	2	16	8	8	9	7	1
84Acc696ZMB	2	6	133	51	36	10	20	11	33	12	9	8
85Acc754ZMB	2	10	143	72	17	7	19	10	20	10	7	4
88-AHM753NAM	2	8	136	47	35	13	14	9	31	10	8	9
90-S19-3NAM	3	8	129	48	68	28	19	11	66	12	9	18
91-UNISRSWA	3	9	134	37	9	8	16	12	9	9	8	6
92-AHM968NAM	2	6	127	25	21	7	15	10	18	9	9	5
95-DODRTZA	3	9	155	38	5	3	16	9	5	11	9	2
99-SB4-2NAM	2	11	130	29	21	8	16	10	21	11	7	6
100-SB16ANAM	3	6	139	45	19	12	19	11	18	12	9	6
104-S-1913NAM	2	7	134	21	23	13	15	9	22	10	8	10
105-MHNBlackNAM	2	8	127	44	9	4	25	9	9	13	8	4
109-BOTS1	2	14	139	70	19	9	19	11	18	13	9	6
113-BOTS5	2	8	127	50	5	2	17	10	5	10	7	1
117-VSSP6CMR	2	6	*	27	*	*	*	*	*	*	*	*
118-Ramayana-IND	2	7	133	32	8	5	15	11	7	11	8	4
119-HYBRID	2	9	136	46	11	7	19	12	12	12	9	3

DAE: days to emergence, DAF: days to 50% flowering, LNO: number of leaves per plant, SPRD: plant spread/canopy, LL: leaflet length, LW: leaflet width, LA: leaf area, PHT: plant height, ITN: Internode length, PTL: petiole length, PITN: petiole-internode ratio, PTL: petiolule length, PNL: penduncle length, STEM: number of stems, DAM: days to maturity, SDW: shoot dry weight, POD: pod number per plant, PDW:pod dry weight, PODL: pod lenght, PODW:pod width, SNO: seed number plant, SL: seed length, SW: seed width, SWE: seed weight

Appendix7: Range of classes for the quantitative traits used for both the glasshouse and the field experiment

Characters	Range of class				
Days to emergence	<7	8-10	>11		
Days to 50% flowering	<30	31-35	36-40	41-45	>46
Leaf number	<50	51-100	101-150	151-200	>201
Canopy spread	<11	11-20	21-30	31-40	>41
leaflet length	<7	8-10	>11		
Leaflet width	<3	4-5	>6		
Leaf area	<30	31-45	46-55	56-65	>66
Plant height	<<25	26-30	31-35	36-40	>41
Petiole	<13	2-3	>4		
Internode	<1	14-16	17-19	20-22	>23
Pet-internode ratio	<7	8-9	>10		
Petiolule	<1	2-3	>4		
Peduncle	<1	2-3	>4		
Stem number	<7	8-10	>11		
Days to maturity	<100	101-115	116-130	131-145	>145
Shoot dry weight	<15	16-30	31-45	46-60	>61
Pod number	<20	21-40	41-60	61-80	>80
Pod dry weight	<10	11-20	21-30	31-40	>41
Pod length	<15	16-20	>21		
Pod width	<10	11-14	>15		
Seed number	<20	21-40	41-60	61-80	>81
Seed length	<9	10-14	>15		
Seed width	<7	8-10	>11		
Seed weight	<10	11-15	16-20	21-30	>31

Appendix 8: Hardy Weinberg Equilibrium (HWE) and the exact p-values estimated using PowerMarker (Version 3.25)

Marker	**X_2 value**	**X_2d.f.**	**Exact p-value**
Primer 7	984.00	36	0.0000
Primer 15	1633.06	120	0.0000
Primer 16	1095.73	45	0.0000
Primer 19	3321.00	378	0.0000
Primer 23	615.00	15	0.0000
Primer 33	1599.00	91	0.0000
Primer 37	1599.23	120	0.0000
Primer 44	615.00	15	0.0000
mBam3co18	2706.00	253	0.0000
Primer D11	1968.00	136	0.0000
Primer D14	3690.00	465	0.0000
Primer E7	482.33	10	0.0000

Appendix 9: Cluster analysis, genetic similarity among the 105 bambara groundnut genotypes, analysis using 141 variables and 105samples/cases UPGMA Nei and Li's coefficient

Node	Group 1	Group 2	Simil.	Objects in group
1	20Acc118CIV	20Acc118CIV	0.95	2
2	117VSSP6 CMR	117VSSP6 CMR	0.923	2
3	6Acc 289BEN	6Acc 289BEN	0.9	2
4	92AHM968NAM	92AHM968NAM	0.9	2
5	60Acc 32NGA	60Acc 32NGA	0.9	2
6	Node 2	117VSSP6 CMR	0.883	3
7	4Acc144GHA	4Acc144GHA	0.878	2
8	Node 1	20Acc118CIV	0.875	3
9	30Acc 476CMR	30Acc 476CMR	0.872	2
10	33Acc 484CMR	33Acc 484CMR	0.872	2
11	40Acc 536CMR	40Acc 536CMR	0.872	2
12	85Acc 754ZMB	85Acc 754ZMB	0.865	2
13	Node 10	33Acc 484CMR	0.861	3
14	Node 4	92AHM968NAM	0.849	3
15	84Acc696ZMB	84Acc696ZMB	0.842	2
16	99SB4-2NAM	99SB4-2NAM	0.842	2
17	70Acc 329NGA	70Acc 329NGA	0.829	2
18	100SB16 ANAM	100SB16 ANAM	0.821	2
19	88AHM753NAM	88AHM753NAM	0.821	2
20	105MHN blackNAM	105MHN blackNAM	0.821	2
21	90S19-3NAM	104S-1913NAM	0.821	2
22	119Hyrid	119Hyrid	0.81	2
23	45Acc 231GHA	45Acc 231GHA	0.8	2
24	76Acc390SDN	76Acc390SDN	0.8	2
25	88AHM753NAM	Node 19	0.79	3
26	74Acc335NGA	74Acc335NGA	0.789	2
27	69Acc286NGA	69Acc286NGA	0.789	2
28	118RamayanaIND	118RamayanaIND	0.789	2
29	40Acc 536CMR	Node 11	0.785	3
30	105MHN blackNAM	Node 20	0.785	3
31	3Acc 9NGA	3Acc 9NGA	0.78	2
32	Node 9	30Acc 476CMR	0.779	3
33	Node 3	6Acc 289BEN	0.775	3
34	Node 23	45Acc 231GHA	0.775	3
35	4Acc144GHA	Node 7	0.771	3

Appendix 9 Continued

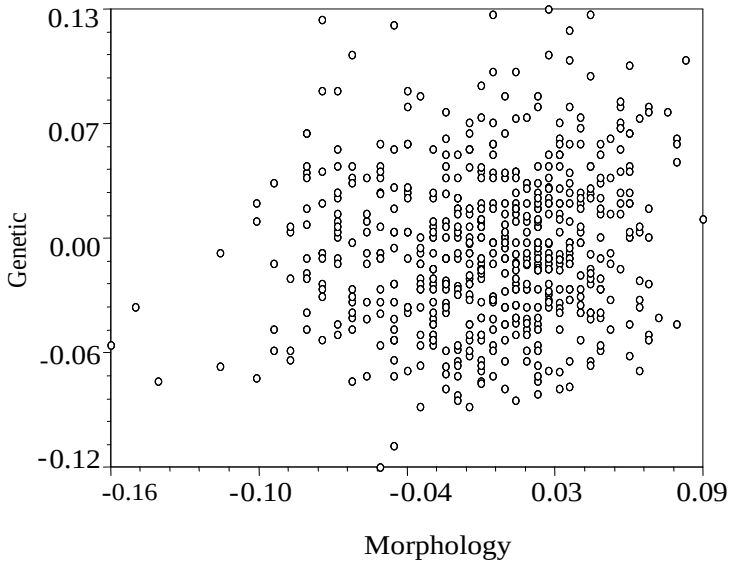
Node	Group 1	Group 2	Simil.	Objects in group
36	10Acc 1276CAF	10Acc 1276CAF	0.757	2
37	81Acc385TZA	81Acc385TZA	0.75	2
38	118RamayanaIND	Node 28	0.749	3
39	Node 17	70Acc 329NGA	0.74	3
40	90S19-3NAM	104S-1913NAM	0.737	2
41	Node 18	100SB16 ANAM	0.734	3
42	76Acc390SDN	Node 24	0.732	3
43	90S19-3NAM	Node 21	0.727	3
44	56Acc 89MLI	56Acc 89MLI	0.722	2
45	109BWA1BWA	109BWA1BWA	0.718	2
46	119Hyrid	Node 22	0.718	3
47	95DODRTZA	95DODRTZA	0.718	2
48	Node 26	74Acc335NGA	0.711	3
49	84Acc696ZMB	Node 15	0.71	3
50	69Acc286NGA	Node 27	0.7	3
51	Node 32	Node 13	0.696	6
52	10Acc 1276CAF	Node 36	0.693	3
53	Node 31	3Acc 9NGA	0.69	3
54	Node 40	Node 43	0.687	5
55	99SB4-2NAM	Node 16	0.683	3
56	48Acc790KEN	48Acc790KEN	0.667	2
57	49Acc793KEN	49Acc793KEN	0.667	2
58	113BWA5BWA	113BWA5BWA	0.667	2
59	Node 45	109BWA1BWA	0.658	3
60	95DODRTZA	Node 47	0.649	3
61	50Acc 792KEN	50Acc 792KEN	0.632	2
62	60Acc 32NGA	Node 5	0.625	3
63	Node 54	104S-1913NAM	0.624	6
64	56Acc 89MLI	Node 44	0.602	3
65	48Acc790KEN	85Acc 754ZMB	0.6	2
66	Node 49	81Acc385TZA	0.597	4
67	Node 33	50Acc 792KEN	0.583	4
68	Node 46	91UNIS RSWA	0.58	4
69	Node 56	Node 12	0.579	4
70	Node 35	Node 62	0.574	6
71	Node 55	Node 41	0.569	6

Appendix 9 Continued

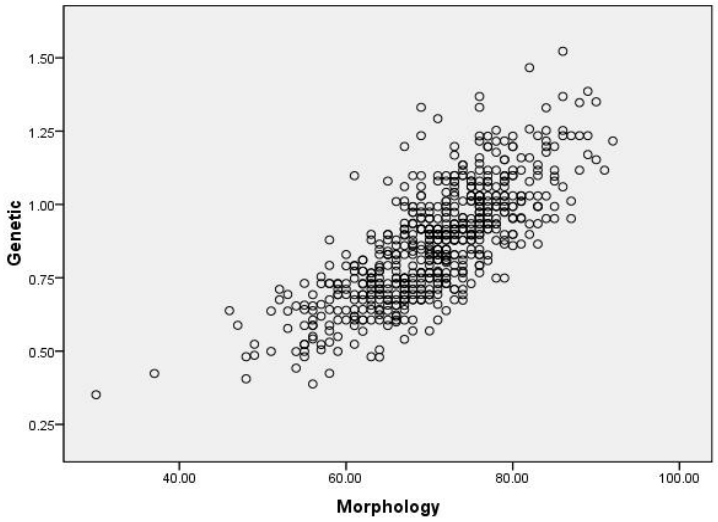
Node	Group 1	Group 2	Simil.	Objects in group
72	Node 69	Node 30	0.56	7
73	Node 57	49Acc793KEN	0.556	3
74	91UNIS RSWA	113BWA5BWA	0.55	2
75	Node 60	Node 38	0.545	6
76	Node 51	Node 48	0.537	9
77	Node 53	Node 70	0.533	9
78	Node 34	Node 73	0.533	6
79	Node 68	Node 74	0.521	6
80	Node 25	Node 65	0.517	5
81	Node 67	Node 64	0.514	7
82	Node 61	Node 50	0.507	5
83	Node 77	Node 39	0.505	12
84	Node 37	Node 59	0.505	5
85	Node 76	Node 42	0.499	12
86	Node 72	Node 14	0.48	10
87	Node 58	Node 6	0.48	5
88	Node 85	Node 29	0.474	15
89	Node 80	Node 63	0.47	11
90	Node 81	Node 78	0.47	13
91	Node 83	Node 88	0.466	27
92	Node 75	Node 79	0.462	12
93	Node 86	Node 84	0.457	15
94	Node 66	Node 87	0.449	9
95	91UNIS RSWA	Node 92	0.441	13
96	Node 91	Node 90	0.432	40
97	Node 89	Node 71	0.43	17
98	Node 96	Node 82	0.412	45
99	Node 94	Node 95	0.406	22
100	Node 93	Node 97	0.403	32
101	Node 100	Node 99	0.395	54
102	Node 98	Node 52	0.376	48
103	Node 102	Node 8	0.376	51
104	Node 103	Node 101	0.371	105

Appendix 10: Scatter plots for morpho-agronomic markers on (Euclidean distance estimates) and molecular markers on (Nei's 1972) conducted using Mantel's test on NTSYS, Pearson correlation and Spearman's rank correlations on SPSS in the Agronomy bay and controlled growth room experiment: Appendix 10.1 and 10.2.

a)

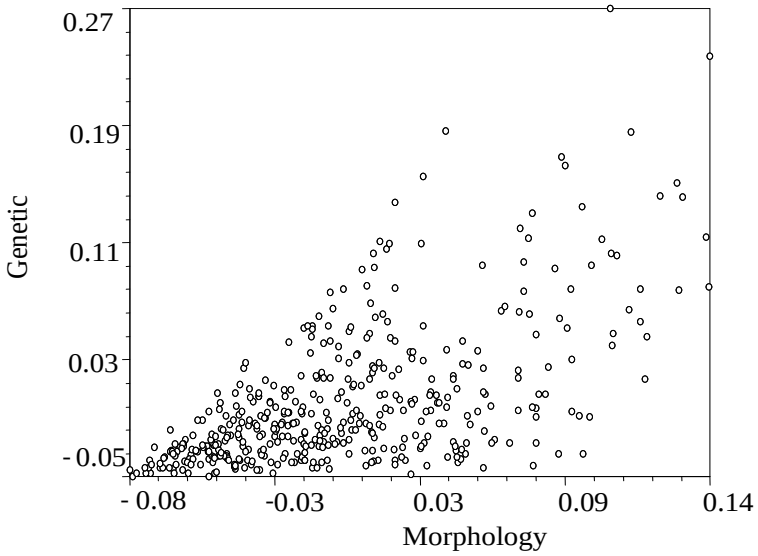


b)

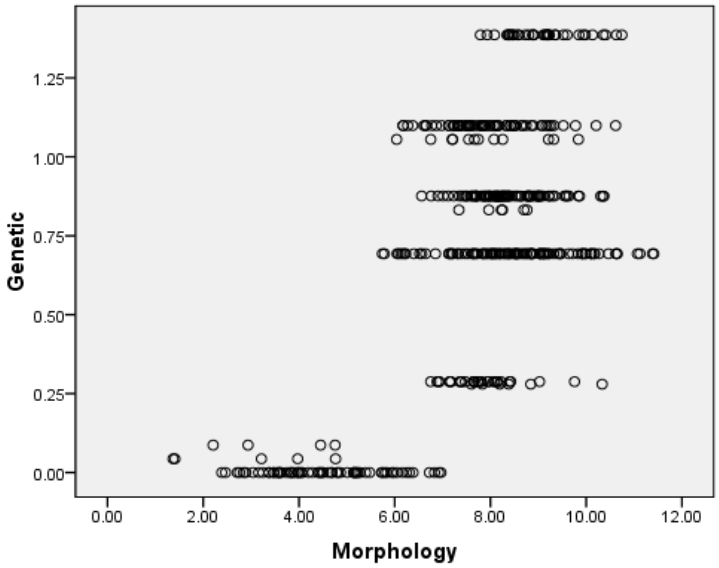


Appendix 10.1: A scatter plot for Agronomy bay experiment on34 morpho-agronomic markers and 20 SSR molecular marker, a) Mantel's test (r = 0.139; P<0.006) b) Pearson correlation (r =0.767: P<0.001) and Spearman rank correlation (r=771; P<0.001).

a)



b)



Appendix 10.2: A scatter plot for controlled growth experiment on 22 morpho-agronomic markers and 12 SSR molecular marker, a) Mantel's test (r = 0.612; P<0.001) b) Pearson correlation (r =0.665; P<0.001) and Spearman rank correlation (r=0.461; P<0.001).

Appendix 11: Mean for the phenotypic measures of 5 lines from the controlled growth room experiment

																		Pod colour				Pod texture		Pod shape		Seed Shape		Seed colour		
Lines	DAE	LNO	LL	LW	LA	PHT	INT	PTL	PITN	PNL	SDW	POD	PDW	PODL	SNO	SL	SWE	1	2	3	4	1	2	1	2	3	1	2	1	2	3
81-Acc385TZA	7	53	6	3	2143	15	2	13	6	1	13	19	6	12	16	8	4	1	0	0	0	1	0	0	0	1	0	1	1	0	0
84-Acc696ZMB	8	48	7	3	2106	18	2	15	8	1	15	10	4	12	10	8	2	0	0	0	1	1	0	0	1	0	1	0	0	1	0
88-AHM753NAM	7	47	6	3	1243	18	2	12	5	1	11	37	10	12	39	9	8	0	0	1	0	0	1	0	1	0	1	0	0	1	0
90-S19-3NAM	8	30	7	3	933	17	2	14	7	1	8	23	9	13	22	9	6	0	0	0	0	1	0	1	0	0	1	0	0	0	1
109-BWA1-BWA	7	47	8	3	1963	21	2	19	12	2	16	29	14	14	30	10	11	0	1	0	0	1	0	0	1	0	0	1	0	1	0

DAE: days to maturity, LNO: number of leaves per plant, LL: leaflet length, LW: leaflet width, LA: leaf area, PHT: plant height, ITN: Internode length, PTL: petiole length, PITN: petiole-internode ratio, PNL: penduncle length,SDW: shoot dry weight, POD: pod number per plant, PDW:pod dry weight, PODL: pod lenght, SNO: seed number plant, SL: seed length,SWE: seed weight