EVALUATION OF THE GENETIC DIVERSITY OF MALAWIAN PIGEONPEA USING SIMPLE SEQUENCE REPEATS MARKERS

by

VINCENT NJUNG'E MICHAEL

submitted in accordance with the requirements for

the degree of

MASTER OF SCIENCE

in the subject

AGRICULTURE

at the

UNIVERSITY OF SOUTH AFRICA

SUPERVISOR: DR S M DE VILLIERS

CO-SUPERVISOR: DR K NTUSHELO

AUGUST 2013

DECLARATION

This thesis is my original work and has not been presented for a degree in any other university.

Signed ……. Date…….

Vincent Njung'e Michael

This thesis has been submitted to the Board of Post graduate studies with the consent and approval of the following supervisors

Signature…… Date……

Dr. Santie de Villiers

International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) Nairobi

Signature…… Date……

Dr. Khayalethu Ntushelo

Department of Agriculture and Animal Health

DEDICATION

This work is dedicated to my wonderful wife, Evalyne and lovely children, Clare and Herman.

ACKNOWLEDGEMENTS

I wish to express my gratitude to all without whom this work could not have been done.

Firstly to God who gave me the opportunity to study and provided me with countless graces throughout.

Secondly, to my supervisor, dr. Santie de Villiers (who was more than a supervisor to me) and my co-supervisor, dr. Khayalethu Ntushelo of the University of South Africa, for their patience and help. I am deeply grateful for their critical and astute contributions to this thesis.

I would also like to thank the ICRISAT-Nairobi staff in the Biotechnology laboratory for their support during my study. My gratitude goes especially to Geoffrey Mugambi and Wilson Kimani for their technical advice in this project and for training me in molecular techniques.

Special gratitude goes to Irish Aid for funding of this research under the Malawi Seed Industry Development Project, year 2 (2010/11).

Finally, I thank my wife, Evalyne for her love, support and encouragement and who together with my children, Clare and Herman, sacrificed their time and resources to enable me to do this work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

ABBREVIATIONS

µl	microlitre
AFLP	Amplified Fragment Length Polymorphism
CGIAR	Consultative Group on International Agriculture Research
CTAB	Cetyl trimethyl ammonium bromide
DArT	Diversity Array Technology
DNA	Deoxyribonucleic Acid
dNTPs	deoxynucleotide triphosphates
doi	Digital object identifier
EDTA	Ethylenediaminetetraacetic acid
FAOSTAT	Food and Agriculture Organization Statistical databases
g	gram
GCP	Generation Challenge Programme
ICRISAT	International Research Institute for Semi-Arid Tropics
M	Molar
mg	milligram
ng	nanogram
PCR	Polymerase Chain Reaction
PIC	Polymorphic Information Content
RAPD	Random Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
RFU	Relative Fluorescence Unit
RNA	Ribonucleic acid
RNAse	Ribonuclease
SNP	Single Nucleotide Polymorphism
SSR	Simple Sequence Repeats
Taq	*Thermus aquaticus*
TBE	Tris/Borate/EDTA
TE	Tris EDTA
UV	Ultra Violet light

ABSTRACT

Pigeonpea (*Cajanus cajan* (L.) Millsp.) is a drought tolerant legume of the *Fabaceae* family in the order *Fabales* and the only cultivated species in the genus *Cajanus*. It is mainly cultivated in the semi-arid tropics of Asia and Oceania, Africa and America. In Malawi, one of the top producers of pigeonpea in Africa, it is grown by small scale farmers as a source of food and income and for soil improvement in intercropping systems. However, varietal contamination due to natural outcrossing causes significant yield losses for farmers. In this study, 48 polymorphic SSR markers were used to assess diversity in all pigeonpea varieties cultivated in Malawi with the aim of developing a genetic fingerprint to distinguish the released varieties. SSR alleles were separated by capillary electrophoresis on an ABI 3700 automated sequencer and allele sizes determined using GeneMapper 4.0 software. Allelic data was analysed with PowerMarker. A total of 212 alleles were revealed averaging 5.58 alleles per marker with a maximum number of 14 alleles produced by CCttc019 (Marker 40). Polymorphic information content (PIC) ranged from 0.03 to 0.89 with an average of 0.30. DARwin software was used to generate a neighbour-joining tree that displayed three major clusters with two sub clusters in Cluster I. The released varieties were scattered across all the clusters observed, indicating that they generally represent the genetic diversity available in Malawi, although it was observed that there is substantial variation that can still be exploited through further breeding. Screening of the allelic data associated with five popular pigeonpea varieties for which a DNA fingerprint was to be developed, revealed 6 markers – CCB1 (Marker 1), CCB7 (Marker 2), Ccac035 (Marker 7), CCttc003 (Marker 15), Ccac026 (Marker 37) and CCttc019 (Marker 40)– which gave unique allelic profiles for each of the five varieties. With further tests needed for its robustness, this

genetic fingerprint can be used for seed certification to ensure only genetically pure seeds are delivered to Malawi farmers.

CHAPTER 1

1. INTRODUCTION

1.1 Background information

Pigeonpea, *Cajanus cajan* [L] Millsp. is a drought tolerant crop and one of the most important legumes grown in the tropics and sub tropics. As a rich source of protein for humans (Saxena et al., 2002), pigeonpea is largely used in diets to supplement cereals, which are protein deficient. Moreover, its high nutritional value has also made pigeonpea a good source of fodder. In the southern Great Plains of the United States of America, pigeonpea provides primary or supplementary high quality forage at a time when other forages are less productive (Rao et al., 2002). There are reports of improved performance of poultry fed with pigeonpea (Amafule and Obioha, 2005; Abdelati et al., 2009). Pigeonpea fixes atmospheric nitrogen and improves the quality and structure of soils (Kumar et al., 1983). Perennial pigeonpea types provide material for fuel wood, basket weaving, and roofing in African villages (Odeny, 2007). Due to its versatility, pigeonpea is an established and valued crop among small scale farmers in Malawi.

Analyzing genetic relationships in species is important for revealing genetic diversity. In addition to showing variability among cultivars (Mohammadi and Prasanna, 2003) genetic diversity provides valuable information for successful breeding programs (Sneller et al., 2005; Varshney et al., 2005). Molecular markers have been useful tools in studying genetic diversity of various crops and among them simple sequence repeats (SSRs) are more popular since they reveal more variation e.g. in pea (Loridon et al., 2005), rice (Jin et al., 2010), maize (Chakraborti et al., 2011) and wheat (Emon et al., 2010).

Most available molecular markers have been employed to study genetic diversity in pigeonpea e.g. random amplified polymorphic DNA (RAPD) (Ratnaparkhe et al., 1995; Choudhury et al., 2008), amplified fragment length polymorphism (AFLP) (Panguluri et al., 2006; Wasike et al., 2005) diversity array technology (DArT) (Yang et al., 2006) and restriction fragment length polymorphism (RFLP) (Sivaramakrishnan et al., 2002).

However, development of SSR markers for pigeonpea has led to their increased use in pigeonpea diversity studies (Burns et al., 2001; Odeny et al., 2007; Odeny et al., 2009; Saxena et al., 2010b, Saxena et al., 2010c, Varshney et al., 2010). In fact SSRs in pigeonpea are bound to be more informative as most have now been mapped in the pigeonpea genome (Bohra et al., 2012).

Pigeonpea in Malawi, especially in southern Malawi, is mainly grown for household consumption on almost every small holder farm and garden. It is an economically important crop as it provides food security, high nutrition, improves the soils and also serves as a valuable cash crop (Soko et al., 2000). However, production suffers greatly due to low quality seeds, which are a result of seed mixing and/or contamination with pathogen propagules. Moreover, access to good quality seed is limited (Jones et al., 2002). It is therefore important to determine the general level of purity of each pigeonpea variety available in Malawi and to what extent varieties become mixed. This is useful in determining how seed purity can be maintained to ensure that farmers are provided with genetically pure seeds.

1.2 Problem statement and justification

Pigeonpea production in Malawi has increased from 64 kilotonnes in 2005 to 193 kilotonnes in 2010 making Malawi Africa's top pigeonpea producer in 2010 (FAOSTAT, 2010). Traditional varieties are largely cultivated and there is greater potential for production if farmers have access to improved high yielding varieties (Simtowe et al., 2009). Natural outcrossing, which can be as

high as 45%, is the major source of varietal contamination in pigeonpea (Saxena et al., 1990). This causes significant yield losses for farmers in Malawi. It is further worsened by the lack of effective channels to avail sufficient high quality seeds to farmers by various stakeholders (Simtowe et al., 2010). Besides genetic purity that directly affects pigeonpea yields, access to pure seeds is of greater importance (Jones et al., 2002). This study assessed the diversity of all known pigeonpea varieties cultivated in Malawi with the aim to develop a genetic fingerprint to distinguish the released varieties. Moreover, it will set a basis for tracking dissemination and adoption of improved and released varieties.

1.3 Objectives

- Assess the level of diversity in all pigeonpeas grown in Malawi and in the Malawi gene bank.
- To develop a genetic fingerprint for identification of commonly cultivated pigeonpea varieties of Malawi.

CHAPTER 2

2. LITERATURE REVIEW

2.1 Pigeonpea classification

Pigeonpea (*Cajanus cajan* (L.) Millsp.) is a drought tolerant legume of the *Fabaceae* family in the order *Fabales*. Other common names are red gram, Congo pea, Gungo pea, Gunga pea, and no-eye pea. It is the only cultivated species in the genus *Cajanus*. Initially, members of this genus were spread between two main genera; *Atylosia* and *Cajanus*. With evidence emerging from morphological, cytological and chemo-taxonomical studies, many *taxa* of *Atylosia*, found to be congeneric with *Cajanus*, were reclassified into *Cajanus* (van der Maesen, 1981). This genus now comprises 32 species from Asia, Africa and Australia.

Pigeonpea is cultivated in the semi-arid tropics, indicated in Figure 1, of Asia and Oceania, Africa and America. India and East Africa hold the largest diversity of pigeonpea and hence it was thought pigeonpea originated from either location. However, van der Maesen, (1980) proposed India as the primary origin from where it was distributed to East Africa and then to the rest of the world. It is mostly grown in the tropical regions of South Asia, Sub-Saharan Africa and Americas.

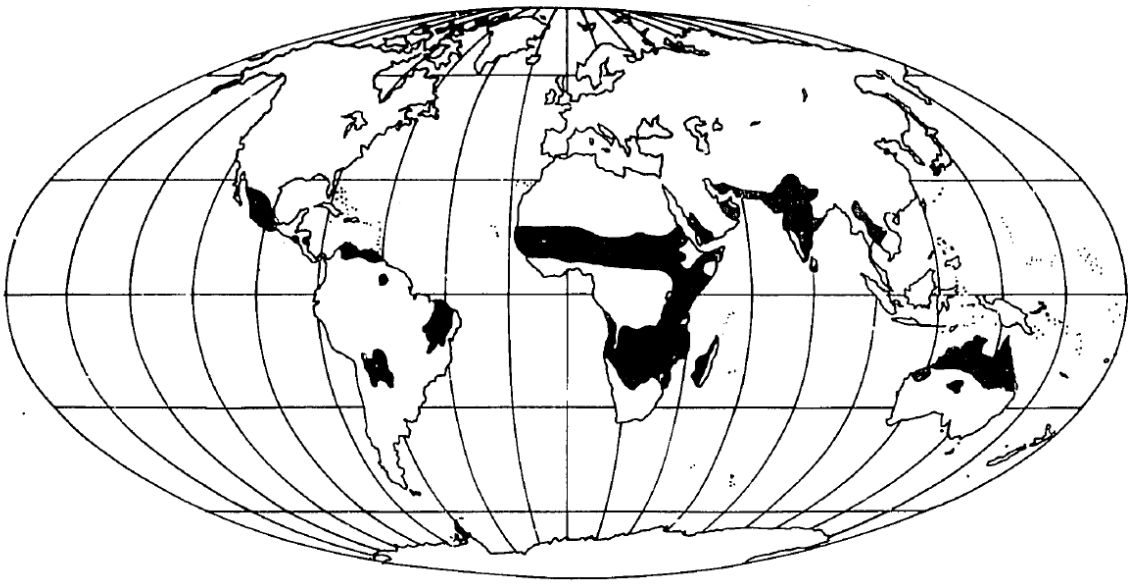

Figure 1. Map showing the semi-arid tropics. (van der Maesen, 1983)

2.2 Global production

Pigeonpea production in the world has increased from 1.9 metric tonnes (Mt) in 1972 to 4.4Mt in 2011. Asia is the highest producer accounting for more than two-thirds of global production. In 2010, production stood at 3.7Mt with India (2.46Mt) as the largest producer followed by Myanmar (0.7Mt), Malawi (0.2Mt), Kenya (0.19Mt), and Tanzania (0.1Mt) (FAOSTAT, 2010), as indicated in Figure 2. However, actual production is higher than that reported since pigeonpea is often intercropped, grown as hedges or as garden plants and also often used as green vegetables for home consumption (van der Maesen, 1983).

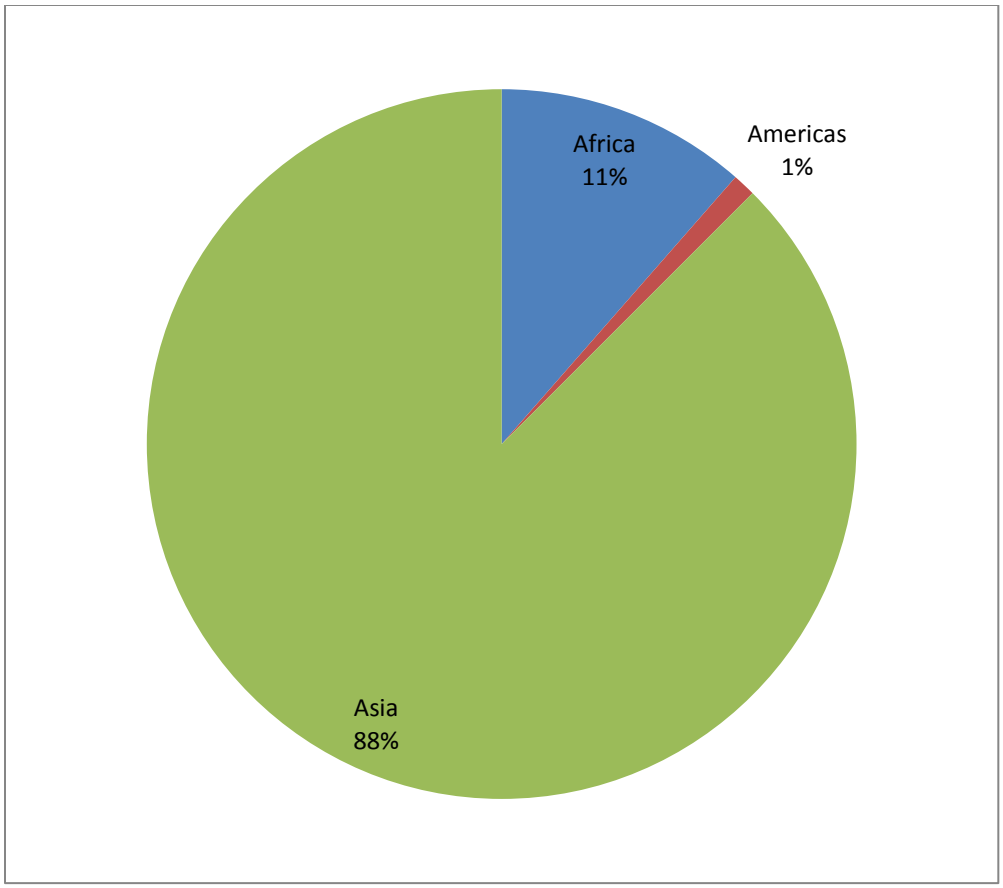

Figure 2. Global pigeonpea production share by region 1992-2011

(From FAOSTAT, 2011)

2.3 Uses of the Pigeonpea

The versatility of pigeonpea has made it an important crop in the semi-arid tropics. As indicated in Table 1, it is a nutritious legume with high levels of amino acids, which are largely digestible and is mainly used to supplement carbohydrate rich foods e.g. maize, cassava, and rice (Saxena et al., 2010a; Faris and Singh, 1990). Pigeonpea is also a rich source of minerals such as potassium, phosphorus, calcium and magnesium (Nwokolo, 1986). Consumption is commonly as dry or green seeds and green pods. In India and Asian countries, pigeonpea is consumed as *dhal*

which is the dry seed excluding the seed coat. Flour milled from the seeds is also popular in several African countries (Eneche, 1999; Oshodi and Ekperigin, 1989).

Pigeonpea provides quality animal feed. The dry leaves, left over pods and milling products form feed for livestock (Saxena et al., 2002). Pigeonpea plants produce a high biomass and provide edible forage or grazing vegetation in dry seasons (Whiteman and Norton, 1981).

The green leaves and pods are used for medicinal purposes. Extracts from the leaves are rich in anti-oxidants and widely used in traditional medicine to treat several diseases (Wu et al., 2009). Extracts from the roots are used as a relief for fever and as an anthelminthic (Chaohong et al., 2001).

Similar to most members of the *Fabaceae* family, pigeonpea has root nodules and it helps improve soil quality by nitrogen fixation (Rao and Mathuva, 2000; Abunyewa and Karbo, 2005). Due to this, pigeonpea is cultivated in intercropping systems with maize and sorghum leading to reduced need for commercial nitrogen and phosphorus fertilizers (Rao and Willey, 1980; Adu-Gyamfi et al., 2007)

Table 1. Distribution of nutrients in mature pigeonpea seed

Constituent	Whole seed	Cotyledons	Embryo	Seed coat
Carbohydrates%	64.2	66.7	31	58.7
Protein%	20.5	22.2	49.6	4.9
Fat%	3.8	4.4	13.5	0.3
Fiber%	5	0.4	1.4	31.9
Ash%	4.2	4.2	6	3.5
Lysine[1]	6.8	7.1	7	3.9
Threonine[1]	3.8	4.3	4.7	2.5
Methionine[1]	1	1.2	1.4	0.7
Cystine[1]	1.2	1.3	1.7	-
Calcium[2]	296	176	400	917
Iron[2]	6.7	6.1	13	9.5
Thiamine[2]	0.63	0.4	-	-
Riboflavin[2]	0.16	0.25	-	-
Niacin[2]	3.1	2.2	-	-

(Adapted from Faris and Singh (1990))
1: g per 100g protein
2: mg per 100g dry matter

2.4 Genetic Diversity and Molecular markers

Genetic diversity is carried out for several purposes including phylogeny, breeding, germplasm conservation and variety identification. Morphological traits and biochemical markers were used before the advent of molecular markers. However, molecular markers, such as RFLPs, AFLPs, RAPDs, SSRs and SNPs are free from environmental influence and can be scored at any stage in plant growth (Gupta et al., 1999). They are now extensively used for genetic mapping, germplasm characterization, and to improve the efficiency of conventional breeding (Rafalski and Tingey, 1993).

2.4.1 Restriction Fragment Length Polymorphism (RFLPs)

RFLPs were the first molecular markers to be developed and used for plant genome analysis. Variability of RFLPs in plants is caused by processes that result in the addition or elimination of

restriction sites in the genome (Helentjaris et al., 1986). In their development, genomic DNA is subjected to restriction enzymes and the resultant fragments separated by gel electrophoresis, followed by transfer to a filter by Southern blotting and probed. RFLP markers are codominant, reproducible, labour intensive and difficult to automate.

2.4.2 Random Amplified polymorphic DNA (RAPDs)

RAPD profiling is one of the polymerase chain reaction (PCR) based molecular markers. It involves the use of arbitrary primers to amplify DNA at discreet random sequences (Williams et, al. 1990). Varietal differences are assigned due to the presence or absence of PCR products visualized on a gel. RAPDs development is easy to automate and does not require prior knowledge of the target sequences used to design primers. It is a dominant marker but its main disadvantage is that its reproducibility is very low.

2.4.3 Amplified Fragment Length Polymorphisms (AFLPs)

AFLPs are based on the selective amplification of restriction fragments. After digestion of the target DNA, specific double stranded adapters are ligated to the restriction fragments. Primers designed to bind to the adapters, the adjacent restriction sites and a few selective bases at the 3' ends of the adaptors are then used for PCR amplification (Vos et al., 1995). The selective bases allow amplification of specific restriction fragments and generate enough bands for polymorphism detection. Differences between samples are due to the presence or absence of restriction sites in the area of PCR amplification. AFLPs are dominant markers and highly reproducible.

2.4.4 Simple sequence repeats (SSRs)

Simple sequence repeat markers are also known as microsatellites or short sequence repeats. They are tandem repeats of between two to six bases occurring throughout all plant genomes. They are abundant across the genome and are developed using primers designed to flank the repeat sequences. SSRs are also highly polymorphic (i.e. show up differences between different individuals analysed), easy to use and automate, codominant (i.e. can discriminate between individuals that are heterozygous and homozygous) and multi allelic i.e. (produce a number of different sized PCR products for a single pair of primers across a range of individuals) (Powell, 1996). They are thus one of the most used molecular markers in genetic analyses of many plants. As they became popular, development of SSRs was confined mostly to important food crops, namely, rice, maize and wheat and the so-called 'orphan crops', like pigeonpea were neglected (Varshney et al., 2009). However there has been an increase in the number of polymorphic SSRs now available due to a recent initiative in pigeonpea genomics (Varshney et al., 2010). Moreover about 330 of these polymorphic SSRs were used to construct a consensus pigeonpea genetic map and their exact location in the chromosomes is now known (Bhora et al., 2012). Therefore because of their ease of use and high polymorphism, SSRs were selected for use in this study.

2.4.5 Single nucleotide polymorphisms (SNPs)

A single nucleotide polymorphism occurs when a single base in a DNA sequence is different between individuals. Therefore when using SNPs alleles are determined by sequence variation and not length variation like the SSRs. SNPs can occur in coding or non-coding regions and tend to be more in repetitive sequences. Although there are several methods of SNP discovery and genotyping they basically make a distinction between a probe of known sequence and the target DNA, which contains the SNP site. SNPs are bi-allelic (there usually exist only two possible

alleles at a target site) and abundant across plant genomes, occurring at least once every 1000 bp (Gupta et al., 2001). A recent study on pigeonpea revealed over 6500 SNPs in conserved orthologous sequence loci with over 750 amenable to high throughput and parallel "oligo pool all" (OPA) genotyping assays (Varshney et al., 2010). These SNPs were used by Kassa et al., (2012) to study the phylogenetic and domestication history, genetic structure, patterns of genetic diversity, gene flow and historical hybridization between *Cajanus cajan* (pigeonpea) and its wild relatives. Moreover, there are also SNPs that were identified to be associated with disease resistance. These SNPs can be used for marker assisted breeding for disease resistance in pigeonpea (Kassa, 2011). Although SNPs are useful in genetic mapping and diversity studies, high costs are incurred since a large number is needed to compensate for their bi-allelic nature and increase genome coverage (Mammadov et al., 2012). While these costs are lowered through genotyping by sequencing, extensive investment in equipment and manpower is required to compute, process and store the large amount of sequencing data generated with this approach (Semagn et al., 2006; Mammadov et al., 2012).

2.5 Diversity studies in Pigeonpea

Realizing the importance of this orphan crop, diversity studies have been carried out for various purposes such as phylogeny, breeding programmes, genetic fingerprinting or variety identification. Earlier studies, which employed morphological traits in pigeonpea were expensive, as they required grow-out tests and carry the risk of environmental interference. This piloted the use of biochemical and DNA-based markers.

2.5.1 Biochemical markers

In evolutionary studies, pigeonpea and *Atylosia* have always been judged as closely related. Seed protein profiles facilitated establishment of phylogenetic relationships among pigeonpea,

Atylosia, and *Rhynchosia* species (Ladizinsky and Hamel, 1980; Pundir and Singh, 1985) and later between pigeonpea and its wild relatives (Kollipara et al., 1994; Jha and Ohri, 1996; Panigrahi et al., 2007). Although seed proteins are stable and reproducible they reveal low polymorphism and are labour intensive (Doveri et al., 2008). The environment also influences biochemical markers since they are the products of expressed genes.

2.5.2 DNA-based Markers

Several more diversity studies in pigeonpea exploited DNA-based markers such as RFLPs and RAPDs (Nadimpalli et al., 1992; Ratnaparkhe et al., 1995, Choudhuray et al., 2008). These markers cover larger regions of the genome in comparison to proteins and some have been linked to a resistant gene for *Fusarium* wilt, a major disease in pigeonpea (Kotresh et al., 2006). Even though they remain the markers of choice in some studies, RAPDs have low reproducibility and RFLPs are difficult to develop, assay and automate. Therefore the subsequent discovery of AFLPs and SSRs, both PCR based, promised to overcome these constraints (Powell et al., 1996; Muller and Wolfenbarger, 1999). Consequently, a number of diversity and phylogeny studies in this crop have utilized AFLPs (Long et al., 2004; Panguluri et al., 2006; Ganapathy et al., 2011).

2.5.3 SSRs

Large numbers of SSR markers have been developed and applied in pigeonpea diversity studies (Burns et al., 2001; Odeny et al., 2007; Odeny et al., 2009; Saxena et al., 2010b, Saxena et al., 2010c). Unlike AFLPs, SSRs are co-dominant and abundant across the genome. They are also multi-allelic, amenable to high-throughput applications and detect more polymorphism (Gupta and Varshney 2000; Salgado et al., 2006). Furthermore, due to their robustness, SSRs are resourceful in assessing genetic purity and can even distinguish between pigeonpea hybrids (Saxena et al., 2010d; Datta et al., 2010; Upadhaya et al., 2011).

2.6 SSRs and genetic diversity

SSRs are tandem repeats that are between two and six bases long and occurring abundantly in a genome. Diversity at the SSR loci is due to the variable number of repeat units. This variation is caused by slip-strand mispairing, which occurs during DNA synthesis and results in a gain or loss of one or more repeat units (Semagn et al., 2006). To allow PCR amplification, primers are designed to flank the SSR loci. The amplification products are separated on silver-stained polyacrymide gels or by capillary electrophoresis, which incorporates fluorescence detection systems. For separation with fluorescent detection systems the primers (usually the forward primer) are synthesized with a fluorochrome attached to the 5' end.

Alternatively, two different forward primers can be used in the PCR step. The first primer, used in the first few PCR cycles, is designed to contain an M13 sequence at the 5' end, in addition to the unique primer sequence (Shuelke, 2000).

Numerous studies have led to development and subsequent utilization of many SSR markers in pigeonpea diversity analyses (Burns et al., 2001; Odeny et al., 2007; Odeny et al., 2009; Saxena et al., 2010b, Saxena et al., 2010c).

The information on diversity revealed in this study across the cultivated and genebank pigeonpea genotypes will be important for future breeding programmes, germplasm conservation efforts and seed certification by a DNA fingerprint.

A tool that will allow the DNA fingerprinting of, especially, the most popular cultivated and newly developed and/or released varieties, will allow scientists, seed producers and seed producer organizations as well as bodies that regulate the quality control of seed purity etc, to accurately determine if a variety is pure and accurately labeled (Soko et al., 2000; Jones et al.,

2002). The Malawi Seed Alliance, which includes the official seed certification unit in Malawi, has recently expressed a need for such a tool as it will allow them to better track how well seed purity and identity is being maintained and will allow for improved confidence in testing, maintaining and providing pure, good quality seed to farmers (ICRISAT-Lilongwe, personal communication). For this reason, this study investigated the possibility to identify a small number of polymorphic SSR markers that can provide a DNA fingerprint for the most important cultivated and released varieties of pigeonpea in Malawi, i.e. ICP 9145, Mtawajuni, ICPV 87105, ICEAP 00040 and ICEAP 00057.

CHAPTER 3

3. Materials and Methods

3.1 DNA Extraction

Seventy nine varieties (listed in Appendix A and B) representing all accessions held in the Malawi gene bank as well as released varieties of pigeonpea in Malawi, were obtained with the assistance of ICRISAT-Lilongwe and planted in Nairobi, Kenya in a screen house.

Two weeks after germination, DNA was extracted from leaves from 5 individual seedlings of each genotype, to ensure statistically sound representation from each accession, according to the protocol described by Mace et al., (2003), omitting the phenol: chloroform extraction step. Two steel beads were inserted in each well of a strip tube, secured in a 96-well rack (Green tree Scientific, USA), together with the leaf samples cut into small pieces to ease maceration of the samples and increase the surface area for detergent activity. Prior to grinding with a 2000 Geno/Grinder© (SpexCertiPrep Inc., USA) 450µl of pre-heated (65^0C) extraction buffer containing 3% (w/v) CTAB, 1.4M NaCl, 0.2% (v/v) β-Mercapto-ethanol and 20mM EDTA was added to the leaf samples. The macerated samples were then incubated for 15 minutes at 65^0C in a water bath with occasional mixing. Solvent extraction was done by adding 450µl of chloroform: isoamyalcohol (24:1) to each sample followed by thorough mixing by inverting the tubes two to four times. The tubes were centrifuged at 4000 rpm for 10 minutes at 24^0C using an Allegra^{TM} 25R centrifuge (BECKMAN COULTER Inc., USA) and approximately 400µl of the upper aqueous layer transferred into clean tubes. Cold isopropanol (0.7 volume) was added and gently mixed to precipitate the DNA. The tubes were centrifuged at 4000rpm for 15 minutes at

4^0C using the AllegraTM 25R centrifuge (BECKMAN COULTER^{TM}) after 30-60 minutes incubation at -20^0C. The supernatant was decanted and the pellet air-dried for 30 minutes. To each pellet, 200µl of low salt TE buffer (1mM Tris and 0.1mM EDTA, pH 8.0) with 3µl of RNase A (10mg/ml) was added and incubated at 45^0C in a water bath to digest RNA. A second solvent extraction step was performed by adding 200µl of chloroform: isoamyalcohol (24:1) to each sample and centrifuged after inverting twice to mix. The aqueous layer (about 180µl to 190µl) was transferred into clean tubes. Ethanol (315µl) and 1/10 volume of 3M sodium acetate solution (pH 5.2) was added to each sample followed by incubation at -20^0C for 5 minutes to allow precipitation. The tubes were then centrifuged at 4000 rpm using an Allegra^{TM} 25R centrifuge (BECKMAN COULTER Inc., USA) for 5 minutes and the supernatant discarded. To wash the DNA pellet, 200µl of 70% ethanol was added and centrifuged at 4000 rpm for 5 minutes. The ethanol was decanted and the DNA pellet was air-dried for 60 to 90 minutes and then re-suspended in 100µl of low salt TE (10mM Tris, 1mMEDTA pH 8.0) buffer.

For all the seeds that failed to germinate, DNA was extracted from the seeds using the protocol described by Sharma et al., (2003). However the homogenization solution was modified to contain 5M NaCl, 2% (w/v) Sarcosyl, 100mM Tris and 20mM EDTA.

3.2 DNA Quality check and quantification

DNA quality was determined by electrophoresis using 0.8% (w/v) agarose gel electrophoresis stained with 5µl/100ml Gel Red® (Biotium Inc., USA). A mixture of 4µl of DNA and 2µl of loading buffer (25mg bromophenol blue (0.25%), 25mg xylene xyanol (0.25%), 4g sucrose (40%)), was electrophoresed for 1 hour at 80 volts in a 1 x TBE buffer (0.1M Tris base, 0.1M boric acid and 0.02M EDTA; pH 8.0). The fragments were visualized under UV light and photographed using a Scion camera (Scion Corporation, USA).

The DNA quantity was determined by spectrophotometry using a Nanodrop© 1000 (Thermo Scientific, USA). Nucleic acids absorb light at both 260nm and 280nm wavelengths and proteins absorb at 230 nm. The spectrophotometer was programmed to measure absorbance (A) from 220 to 350 nm and display the DNA concentration. The ratio of absorbance at these wavelengths is an indicator of DNA purity. For pure DNA, the $A_{260/280}$ should be between 1.8 to 2.0 and the $A_{260/230}$between 1.8 to 2.3. Lower values of the latter ratio indicate the presence of contaminating proteins. All the DNA samples were then diluted to 10ng/µl and used for PCR.

3.3 Polymerase Chain Reaction

Polymerase chain reaction was done using 48 publicly available polymorphic markers (Appendix D). All the forward primers contained an M13 tag (CACGACGTTGTAAAACGAC to allow incorporation of a fluorochrome during the PCR process (Shuelke, 2000). A second fluorescent-labeled forward primer was also incorporated in each reaction, consisting only of the M13 sequence, which subsequently generated labeled PCR amplification products for capillary electrophoresis, that also has the M13 sequence incorporated at the 5' end in the final PCR product (Shuelke, 2000).

The fluorescent labels used were 6-Carboxyfluorescein (6-FAM), and NED®, VIC® and PET® (Life Technologies Corporation, Carlsbad, USA). During capillary electrophoresis the amplification products passed through a detection window and a light excited the fluorescent dye. The fluorescence was thereafter visualized using a computer programme as relative fluorescent unit (RFU) against fragment length in base pairs. An allele was scored for each data point as length in base pairs at the highest RFU peak. The 48 SSRs were selected from among the most polymorphic SSRs reported in several works (Burns et al., 2001; Odeny et al., 2007; Odeny et al., 2009; Saxena et al., 2010b, Saxena et al., 2010c). Moreover, the Generation

Challenge Programme (GCP) created by the Consultative Group of International Agriculture Research (CGIAR) characterized 1000 composite pigeonpea accessions with 20 of these SSRs (GCP-Bioinformatics Registry, http://gcpcr.grinfo.net/index.php?app=datasets&inc=files_list).

Each PCR contained 1x PCR buffer (20mM Tris-HCl (pH7.6); 100mM KCl; 0.1mM EDTA; 1mM DTT; 0.5% (v/v) Triton X-100; 50% (v/v) glycerol), 2mM $MgCl_2$, 0.16mM dNTPs, 0.16µM of a labeled M13-primer, 0.04µM M13-forward primer, 0.2µM reverse primer, 0.2 units of *Taq* DNA polymerase (SibEnzyme Ltd, Russia) and 30ng of template DNA. The volume for each PCR was topped to 10µl with sterile water. The concentrations and volumes for components in each PCR are shown in Table 2 below.

Table 2. Concentrations and volumes for each PCR reagent in a single PCR reaction

PCR Component	Stock Concentration	Final Concentration	Volume for one PCR reaction in µl
PCR Buffer without $MgCl_2$	10x	1x	1
$MgCl_2$	50mM	2mM	0.4
dNTPs	2mM	0.16mM	0.8
M13-Fluorescent forward primer	2µM	0.16µM	0.8
Forward primer	2µM	0.04µM	0.2
Reverse primer	2µM	0.2µM	1
Taq DNA Polymerase	5U	0.2U	0.04
Sterile Water			2.76

Reactions were performed on a thermocycler (GeneAmp PCR system 9700®, Applied Biosystems, USA) with initial denaturation of 94^{o}C for 5 minutes, followed by 35 cycles of 94^{o}C for 30 seconds, 59^{o}C for 1 minute and 72^{o}C for 2 minutes followed by final elongation at 72^{o}C for 20 minutes. PCR conditions were optimized by changing the annealing temperatures to ensure that all SSR markers were amplified. For markers that did not amplify with this PCR

protocol, changes were made only to the annealing temperature using the published annealing temperatures for the respective markers, followed by testing annealing temperatures calculated using the SSR primer sequences in the first step of BioMath Calculators (http://www.promega.com/techserv/tools/biomath/calc11.htm). For a group of SSRs primers that failed to amplify after these annealing temperature adjustments, a gradient PCR using Techne TC-5000 Thermo cycler®, (Bibby Scientific Group, United Kingdom), was used to determine their annealing temperatures. Gradient PCR is done on a gradient PCR machine, which allocates different annealing temperatures to each column in a 96-well PCR plate. The temperatures used for this study were between 48.8°C in column 1 to 61.1°C in column 12.

Amplification was confirmed by electrophoresis using a 2% (w/v) agarose gel stained with GelRed® (Biotium, USA) and visualized under UV light. Depending on the efficiency of amplification, 2.5µl – 3.5µl of 3 to 4 different amplification products were co-loaded along with the internal size standard, GeneScan™ –500 LIZ® (Applied Biosystems, USA) and Hi-Di™ Formamide (Applied Biosystems, USA) and separated by capillary electrophoresis using an ABI Prism® 3730 Genetic analyzer (Applied Biosystems, USA) (Kuomi et al., 2004).

3.4 Fragment Analysis

Fragment analysis was performed with Gene Mapper 4.0 (Applied Biosystems, USA) and allelic data for each marker analyzed with PowerMarker V3.25 (Liu and Muse, 2005), DARwinV.5.0.158 (Dissimilarity analysis and representation for Windows®) software (Perrier and Jacquemound-Collet, 2006). Powermarker® and DARwin are statistical analysis softwares. Powermarker produces summary statistics such as allele number, gene diversity and/or polymorphic information content, inbreeding coefficient; estimation of allelic, genotypic and

haplotypic frequency; Hardy-Weinberg disequilibrium and linkage disequilibrium. Polymorphic information content, which is a measure of diversity, is calculated using the formula

$$\widehat{PIC}_l = 1 - \sum_{u=1}^{k} \tilde{p}_{lu}{}^2 - \sum_{u=1}^{k-1} \sum_{v=u+1}^{k} 2\tilde{p}_{lu}^2 \tilde{p}_{lv}^2$$

where P_{lu} is the allele population frequency at the *l*th locus and P_{lv} is the genotype population frequency at the *l*th locus.

Dissimilarity was calculated by Darwin software using the formula

$$d_{ij} = 1 - \frac{1}{L} \sum_{l=1}^{L} \frac{m_l}{\pi}$$

where *dij* is the dissimilarity between units *i* and *j*, *L* is the number of loci, π is the ploidy and m_l is the number of matching alleles for locus *l*. DARwin was also used to display dendograms using the dissimilarity matrix.

3.5 Genetic fingerprint

Allelic results were investigated to identify markers with the potential to provide a DNA fingerprint for cultivated and released pigeonpea varieties from Malawi. The ideal fingerprinting markers were considered to be those that can unambiguously discern all the varieties from one another. It was highly unlikely that a single marker would fit these criteria and more likely that a set of markers would have to be considered together for this purpose. In order to identify such a set of markers, the following steps were followed. Firstly, the allelic data for the target varieties were selected from the complete dataset and considered in isolation from the gene bank and reference data. Secondly, the data were screened to eliminate all the markers that had low success in PCR amplification (and therefore presented ≥40% missing data), were monomorphic and heterogeneous (provided multiple different alleles within a population). If a marker presented a different allele for two individuals of an accession, it was considered heterogeneous

and not included. If a marker presented a different allele for only a single individual, it was considered homogeneous and included, provided that it was polymorphic across all the accessions.

CHAPTER 4

4. RESULTS

4.1 Genomic DNA Extraction

Leaf tissue was sampled as illustrated in Figure 3 and DNA extracted individually from five plants from each of the 72 accessions studied that are listed in Appendices A and B. At least 10 seeds of each plant were planted and after two weeks, most accessions showed good germination, except for 26 that had produced fewer than five seedlings or did not germinate at all (Table 3). All seeds that failed to germinate were obtained from the Malawi gene bank.

Table 3. Accessions that germinated poorly and from which fewer than 5 seedlings per genotype were obtained

Number of seedlings that germinated	0	1	2	3	4
Genotypes	MW 765, MW 587, MW 648, MW 2240, MW 2289.	MW 480, MW 2243, MW 2327	MW 470, MW 2281, MW 2287, MW 2298	MW326, MW 454, MW 786, MW 2270, MW 2286, MW 2313	MW 690, MW 2238, MW 2245, MW 2265, MW 2283, MW 2295, MW 2305, MW 2333

Figure 3. Sampling pigeonpea leaves for genomic DNA extraction from 14-day old seedlings

4.1.1 DNA Quality

All the samples extracted from fresh leaves contained good quality, high molecular weight DNA even though the phenol: chloroform extraction step was omitted as illustrated in Figure 4. Genomic DNA in lanes 1, 2, 3 and 93 shows high quality, intact genomic DNA while that in lanes 6 and 24 show some degradation. Figure 5 shows that most of the samples extracted from seeds showed high degradation compared to the samples shown in Figure 4 from leaves except for those in lane 1, 2, 3 and 21, which showed intact DNA.

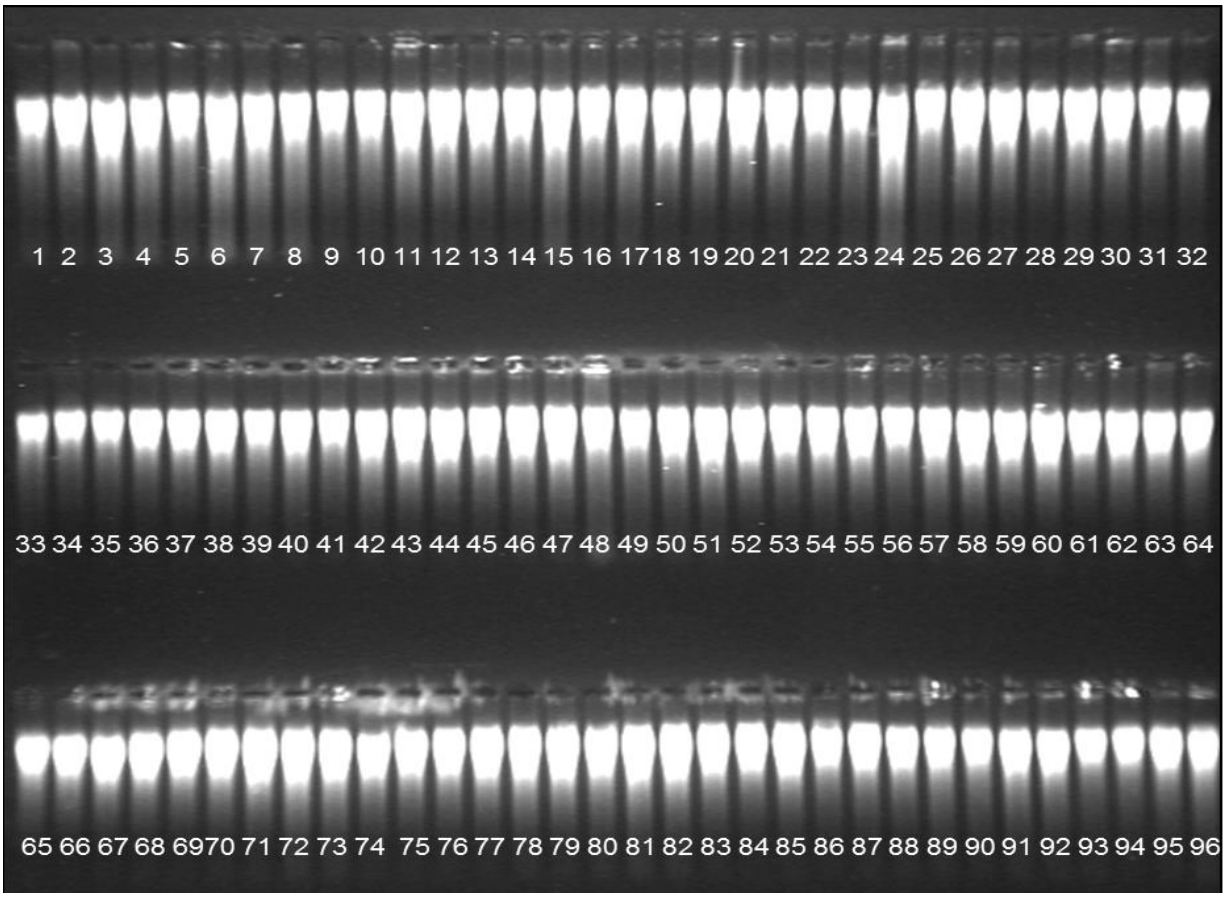

Figure 4. Agarose gel (0.8% w/v) image of extracted high quality genomic DNA obtained from fresh 14-day old leaf material.

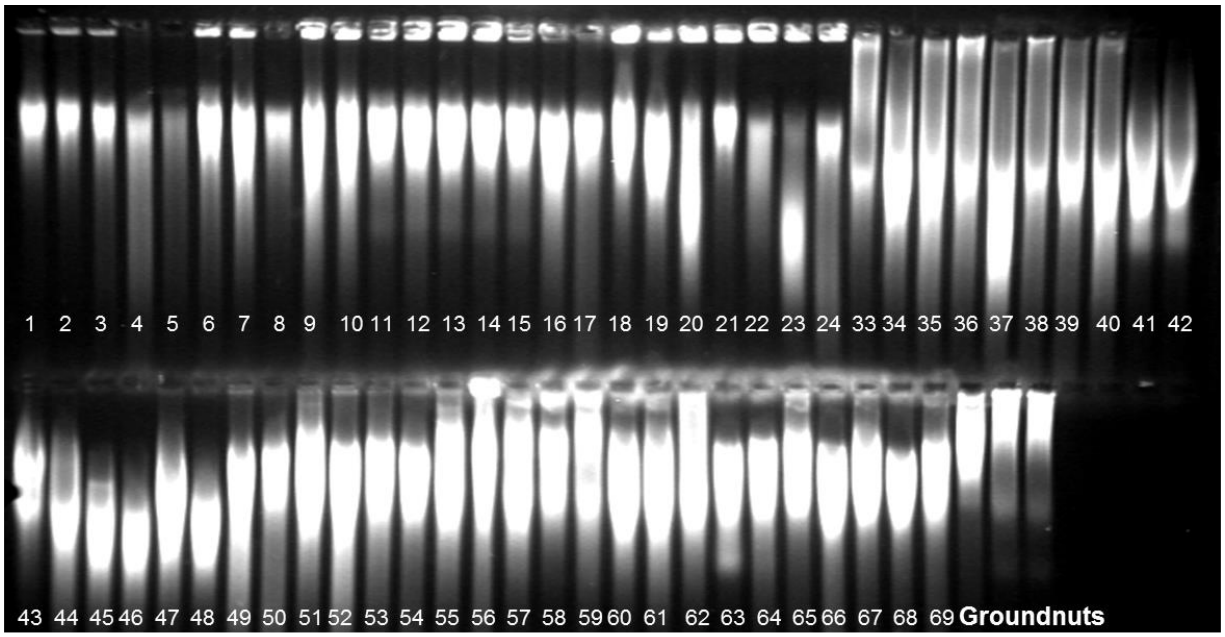

Figure 5. Agarose gel (0.8% w/v) image of extracted DNA that was mostly degraded and obtained from seeds.

4.1.2 DNA Quantity

For DNA quantification, the Nanodrop© 1000 (Thermo Scientific, USA) spectrophotometer was programmed to measure absorbance (A) from 220 to 350 nm and display the DNA concentration, as illustrated in Figure 6. All DNA samples extracted from fresh leaves in this study were of high purity with an average $A_{260/280}$ of 1.95, ranging between 1.72 and 2.10. The average concentration of the DNA from leaf samples was 573.77 ng/µl, ranging from 65.99 ng/µl to 1342.21 ng/µl. The DNA extracted from the seeds was less pure and achieved a mean $A_{260/280}$ of 1.62, ranging from 1.19 to 2.00 and average $A_{260/230}$ of 0.70, ranging from 0.21 to 1.19. For DNA from seeds, the average concentration obtained was 457.82 ng/µl, ranging from 96.30 ng/µl to 688.52 ng/µl. A few examples of the spectrophotometer outputs are presented in Table 4 and a complete list is presented in Appendix C.

Table 4. Nanodrop© spectrophotometer outputs from a selection of extracted DNA samples.

Sample Used	Sample ID	DNA concentration in ng/µl	$A_{260/280}$	$A_{260/230}$
	KAT 60/8_5	633.85	2.00	2.15
	ICP 2309-2	232.01	1.95	2.08
	ICP 13076_1	953.00	1.98	2.38
Fresh leaves	ICEAP 00068_2	683.36	1.97	2.40
	MW 765-5	406.92	1.51	0.61
	MW 648-3	457.28	1.8	0.86
	MW 2289-3	440.02	1.88	1.02
seeds	MW 454-5	684.4	1.69	0.45

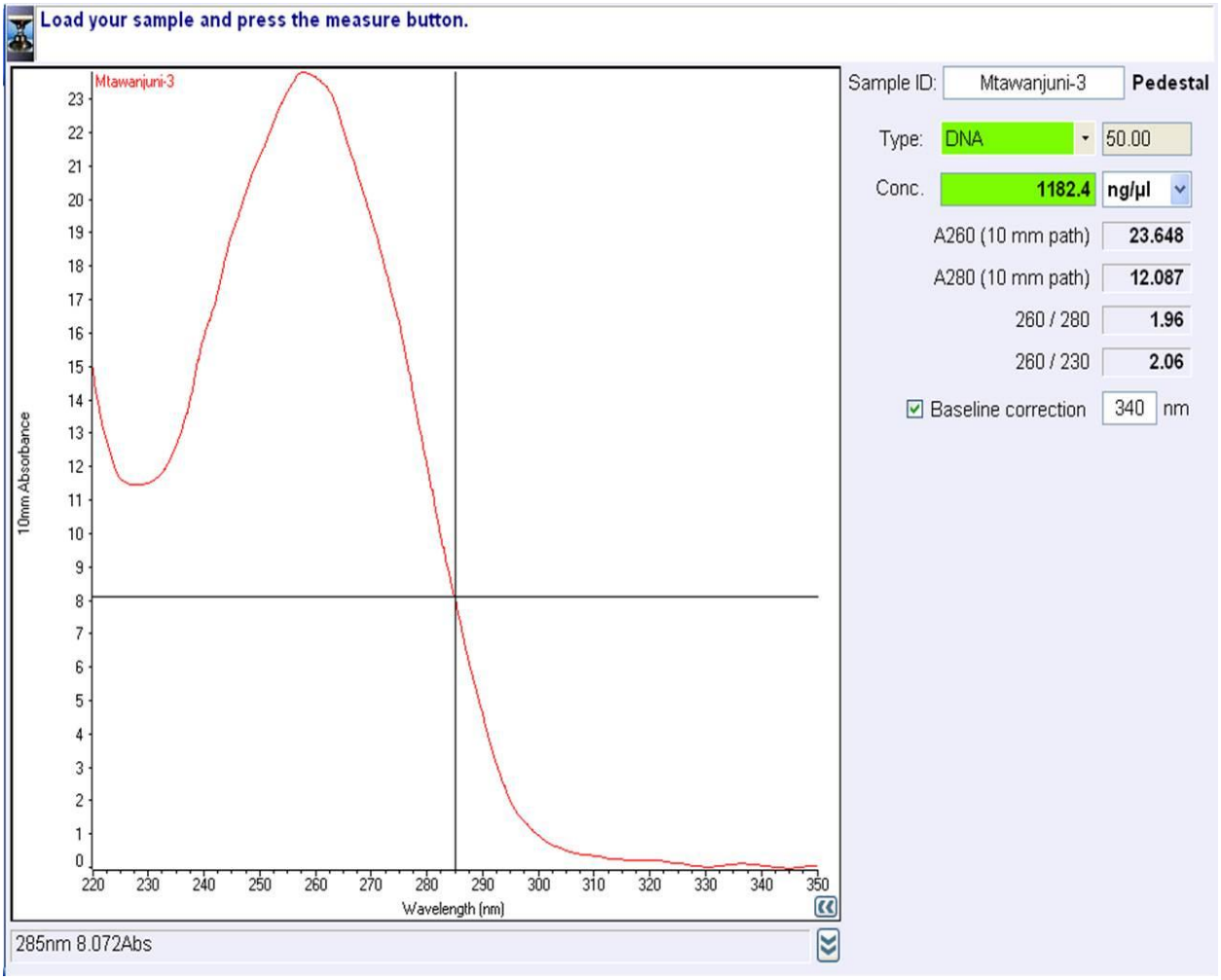

Figure 6. A computer screen shot of the Nanodrop© output

4.2 Polymerase Chain Reaction

For 37 of the SSRs tested, there were bands showing successful amplification by PCR while no bands appeared for the remaining 11 SSRs, which indicated non-amplification (Figure 7). To ensure amplification for these 11 SSRs, PCR was optimized by adjusting the annealing temperatures. First, using 8 random DNA samples, annealing temperatures were examined and PCR was performed using the published annealing temperatures for the respective markers, followed by testing annealing temperatures calculated using the SSR primer sequences in the first step of BioMath Calculators (http://www.promega.com/techserv/tools/biomath/calc11.htm)

and finally, with annealing temperatures obtained from a gradient PCR set with annealing temperatures ranging from 48.8°C degrees to 61.1°C. The results are presented in Figure 8. After these adjustments in the PCR conditions a further eight SSRs were amplified whereas three SSRs, Marker 25 (CCcttc001), marker 34 (Cccta003) and Marker 46 (CCttc007) still did not amplify and were not further used in this study. It was further noted that three of the eight successful SSRs, Marker 22 (CCttc006), Marker 28 (CCttc012) and Marker 44 (CCtc020) amplified in less than half of the total samples. Further optimization through the reduction of the fluorescent dye for one of these markers (Marker 22) resulted in amplification (Figure 9). However, the fluorescent signals from these amplification products could not be detected during capillary electrophoresis to allow allele scoring on GeneMapper® software and this marker was also excluded from further use.

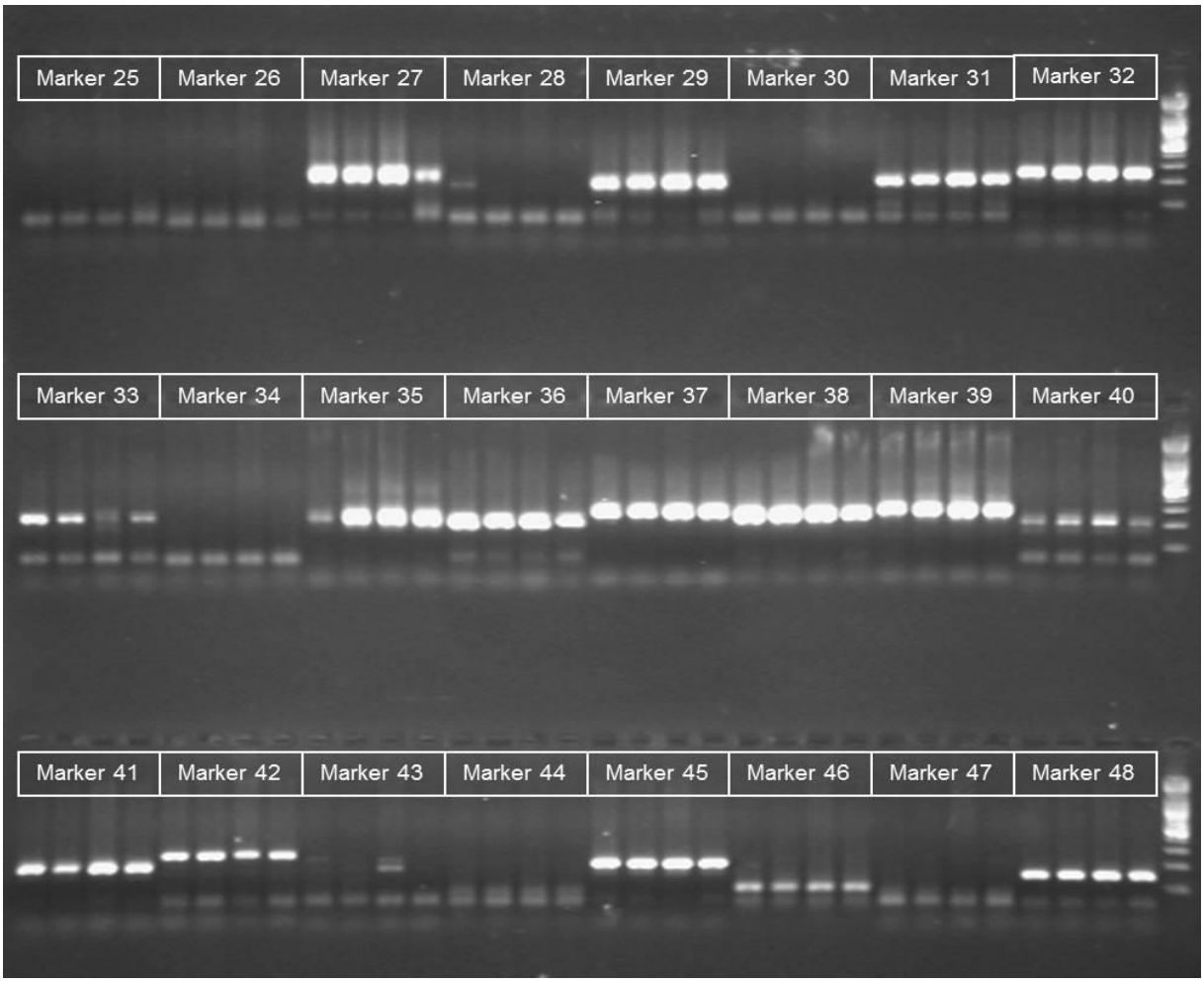

Figure 7. Agarose gel (2.0% w/v) analysis of SSR PCR amplification products for the last 24 markers on selected samples at annealing temperature 59°C. The DNA ladder used was 100bp. Markers 42, 45 and 48 showed good amplification products while Markers 25, 34 and 44 showed no amplification.

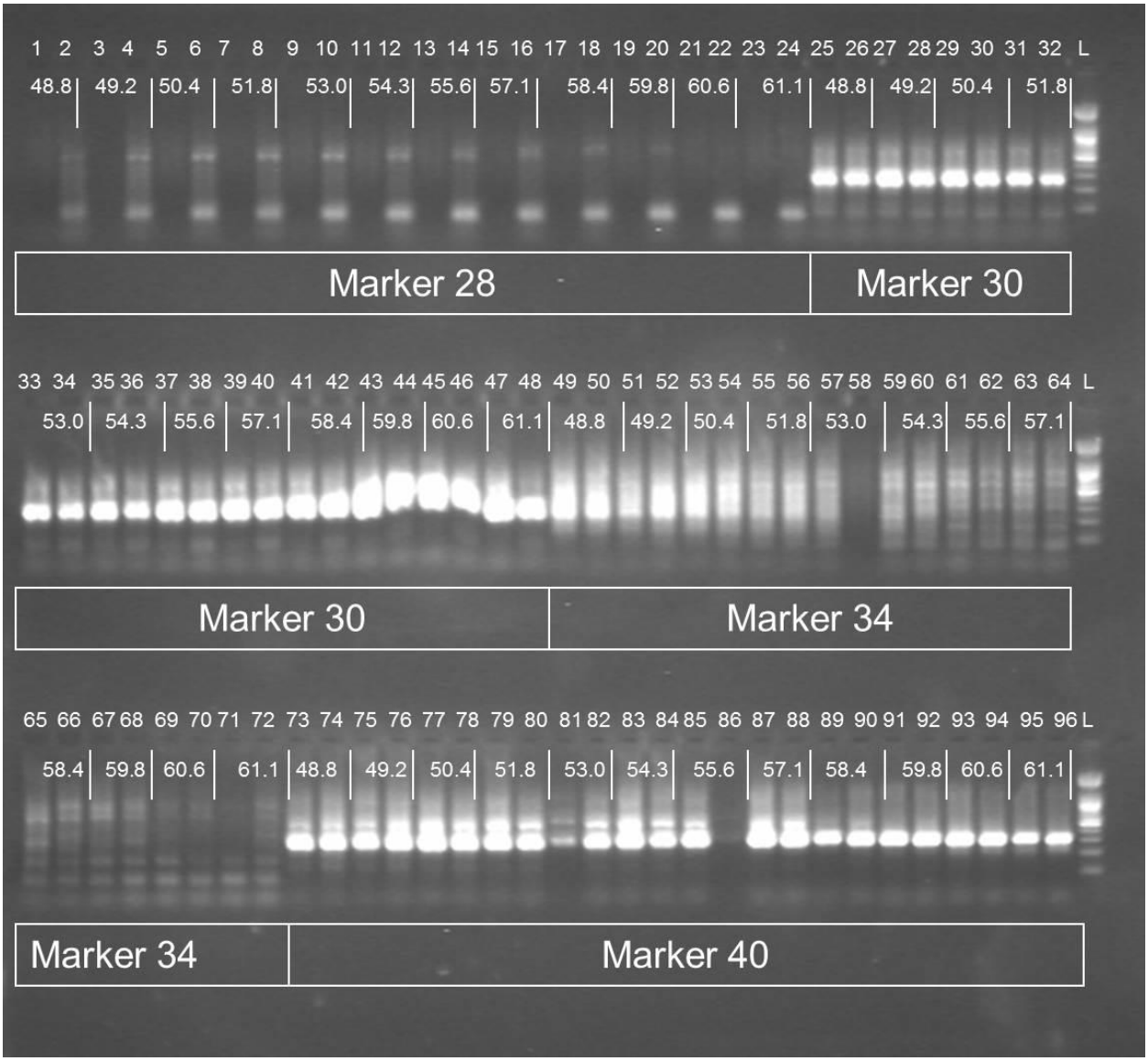

Figure 8. Agarose (2.0% w/v) gel image illustrating the success or not of applying gradient PCR for 4 markers across two samples. Lanes 1 to 24 were two pigeonpea accessions tested at different annealing temperatures (ranging from 48.8°C to 61.1°C) as indicated with Marker 28. The same accessions were tested at similar temperatures for Marker 30 (lanes 25 to 48), Marker 34 (lanes 49 to 72) and Marker 40 (lanes 73 to 96). The products were run at 120V for 30 minutes. Lane L contains 100bp ladder. Marker 30 and Marker 40 were successful at annealing temperatures 51.8°C (lanes 31 and 32) and 61.1°C(lanes 95 and 96) respectively. Marker 28 and 34 did not amplify.

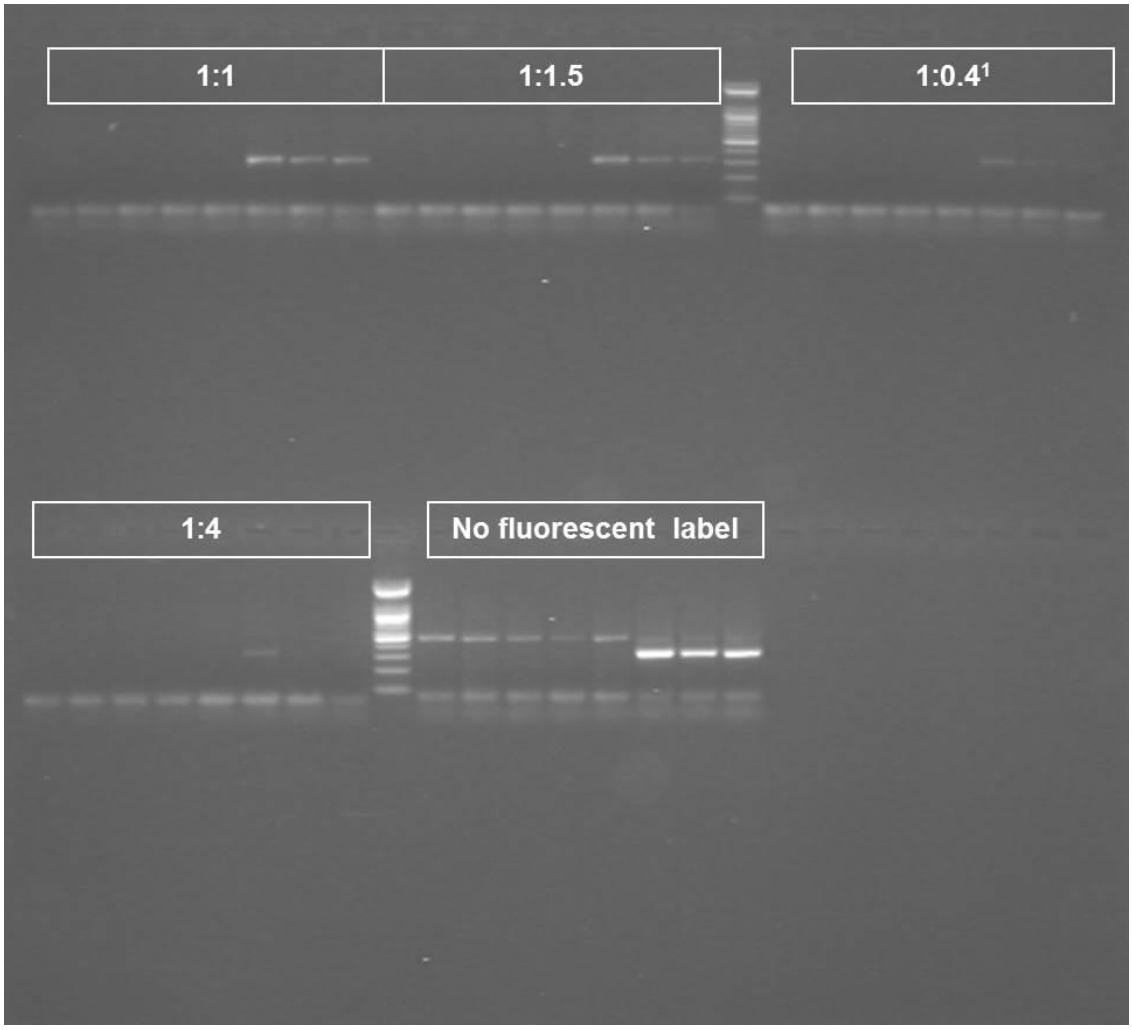

Figure 9.Agarose (2.0% w/v) gel image illustrating amplification for Marker 22 with different fluorescent label concentrations. A concentration of 0.04µM of the forward primer and 0.16µM for the fluorescent labeled primer (1:4) was used. This is similar to all the other markers that were successful. The other ratios were derived from these concentrations.

4.3 Allele scoring and analysis.

Following PCR, the amplified DNA fragments were successfully separated by capillary electrophoresis on an ABI3730 automatic sequencer and the allele (fragment) sizes could be scored using GeneMapper® software as presented in Figure 10.

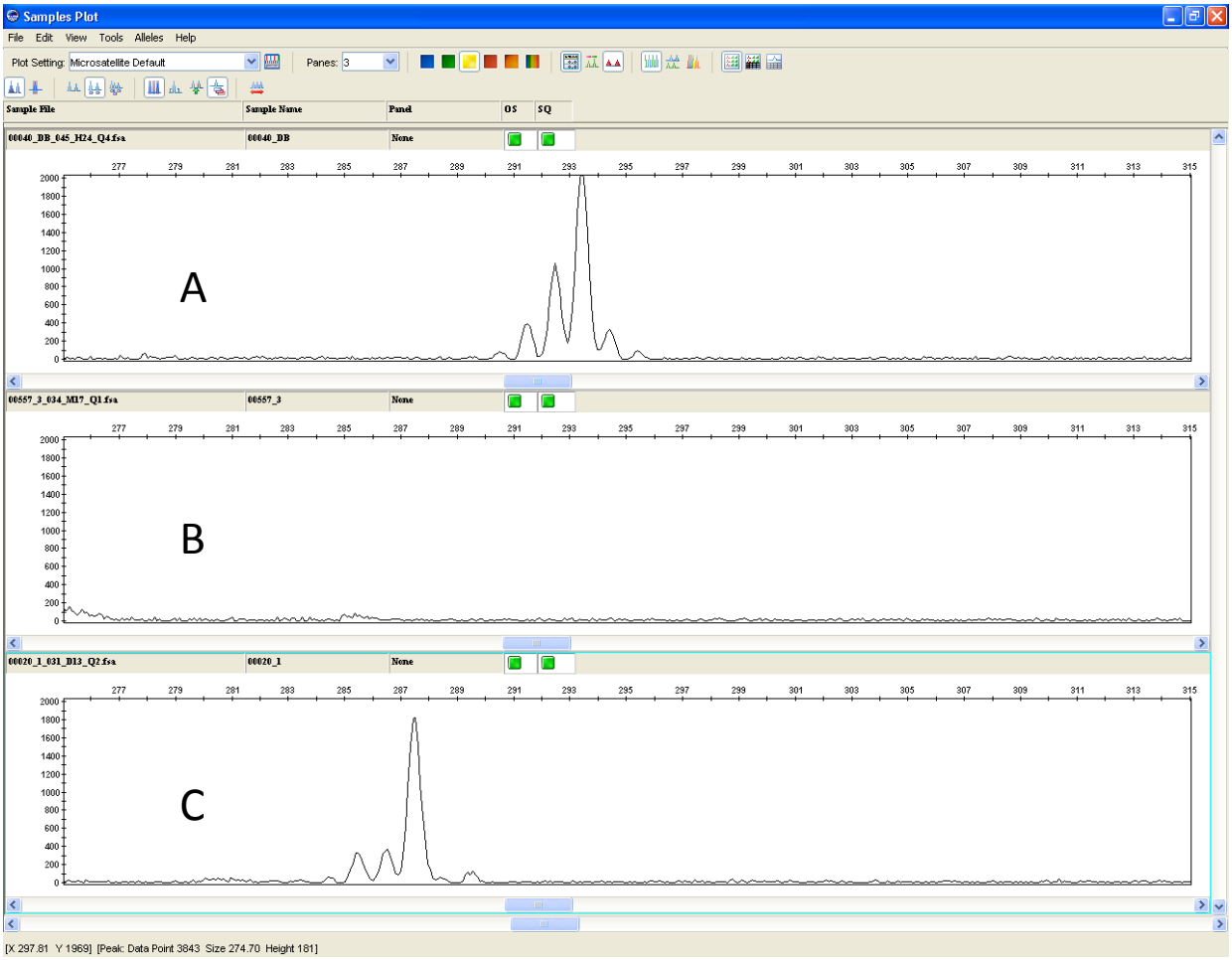

Figure 10. Computer screen shot of GeneMapper® peaks. Samples A and C showed different sizes (alleles) for the same SSR marker while sample B amplified no peak at this locus.

Initial analysis of the allelic data using PowerMarker® software confirmed that three markers (Marker 25, Marker 34 and Marker 46) did not amplify during PCR, and three more markers (Marker 22, Marker 28 and Marker 44) failed to amplify in more than 50% of the samples. One marker, Marker 6 (CCttc008), amplified two different loci, which was evident in that each sample produced two distinct alleles – one 255bp and the other ranging from 251bp to 255bp long - of which the former one was monomorphic, i.e. it amplified the same allele (255bp) in all samples and was therefore not useful for discerning genetic diversity in this germplasm. Marker 33 (Ccat011 (Ccat006)) was highly heterozygous, i.e. each sample produced two different alleles

and these were often different amongst the individuals within an accession, which complicated the interpretation of the allelic data for this marker within this germplasm set and this marker was therefore excluded from further data analysis. The DNA samples from the genotypes MW 2243_3, MW 2243_4 and MW 2355_7 did not amplify successfully with most SSRs during PCR as only 33% of allelic data was available. These genotypes were also excluded from subsequent analysis. It was noted that DNA samples for these three genotypes were obtained from seeds.

After data curation the data matrix obtained was for 38 markers and 392 genotypes. These polymorphic markers, their primer sequences and repeat motifs are listed in Appendix D. The allelic data was analysed by PowerMarker® to produce a table of allele frequencies, heterozygosity, allele number and polymorphic information content, presented in Table 5. Two hundred and twelve alleles were revealed with an average of 5.58 alleles per marker and a maximum number of 14 alleles produced by Marker 40 (CCttc019). Polymorphic information content (PIC), an indicator of how well a marker is able to distinguish the samples tested due to the diversity of alleles detected across the samples, ranged from 0.03 to 0.89 with an average PIC of 0.30.

Darwin software was used to produce a dissimilarity matrix, which was displayed in a neighbour-joining tree or dendogram, illustrated in Figure 11. DARwin software was further used for principle coordinate analysis. There were three major clusters and two sub clusters in Cluster I (A and B). Two of them, cluster I and II comprised of released varieties and some gene bank materials. A large part of the gene bank materials was grouped together in cluster III. Landraces were also grouped in cluster III, apart from ICP 13076, which was in Cluster I sub-cluster B.

Table 5. PowerMarker® Summary statistics output for the selected 38 markers across 392 individual DNA samples

Marker	Major.Allele.Frquency	Allele No	Heterozygosity	PIC
1	0.92	8.00	0.06	0.14
2	0.58	6.00	0.03	0.39
3	0.94	6.00	0.02	0.12
4	0.85	5.00	0.06	0.25
5	0.97	2.00	0.01	0.05
6	0.96	2.00	0.00	0.07
7	0.94	2.00	0.00	0.11
8	0.63	7.00	0.24	0.40
10	0.94	3.00	0.03	0.10
11	0.99	2.00	0.02	0.03
13	0.91	5.00	0.00	0.17
14	0.50	8.00	0.93	0.55
15	0.86	5.00	0.03	0.23
16	0.91	4.00	0.01	0.16
17	0.90	6.00	0.03	0.17
18	0.98	2.00	0.00	0.04
19	0.97	3.00	0.04	0.06
20	0.84	3.00	0.00	0.23
21	0.39	7.00	0.67	0.65
23	0.91	9.00	0.00	0.17
24	0.42	8.00	0.11	0.67
26	0.13	11.00	0.03	0.89
27	0.47	3.00	0.31	0.51
29	0.53	7.00	0.28	0.51
30	0.93	3.00	0.00	0.13
31	0.90	3.00	0.05	0.18
32	0.57	7.00	0.67	0.60
33	0.45	14.00	0.03	0.73
35	0.72	4.00	0.14	0.41
36	0.92	3.00	0.00	0.15
37	0.86	7.00	0.01	0.23
39	0.82	10.00	0.05	0.30
40	0.59	14.00	0.82	0.58
41	0.80	3.00	0.00	0.31
42	0.47	7.00	0.07	0.56
43	0.94	7.00	0.01	0.11
47	0.73	4.00	0.05	0.40
48	0.93	2.00	0.01	0.13
Mean	**0.76**	**5.58**	**0.13**	**0.30**

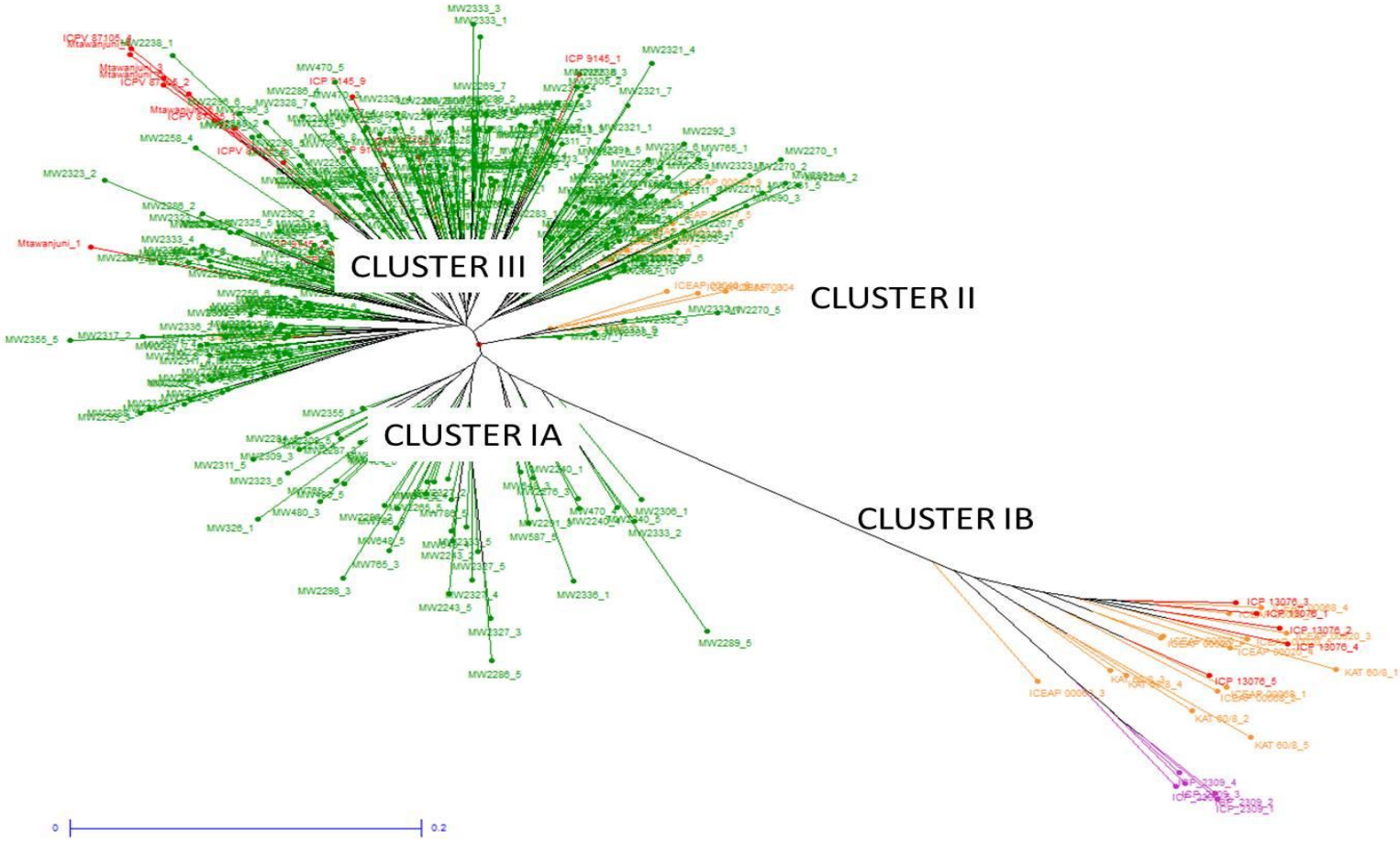

Figure 11. Dendogram showing different pigeonpea clusters

Orange	Released varieties
Green	Gene bank materials
Red	Landraces
Violet	Reference variety

4.4 DNA fingerprint.

Screening of the allelic data associated with the selected varieties for which a DNA fingerprint was to be developed, revealed that 6 markers – CCB1 (Marker 1), CCB7 (Marker 2), Ccac035 (Marker 7), CCttc003 (Marker 15), Ccac026 (Marker 37) and CCttc019 (Marker 40) - met all the set criteria (described in section 3.5 in Chapter 3). Of the other 39 (29 if the final PowerMarker data set is considered) markers, 15 presented ≥40% missing data, 16 were monomorphic and 10 were heterogeneous and were not considered. The fingerprint developed with the 6 markers listed above, for the most important cultivated varieties in Malawi, are presented in Table 6.

Table 6. Genetic fingerprint for five pigeonpea varieties using 6 SSR markers.

		Allelic sizes				
Marker Number	**Marker Name**	**ICEAP 00040**	**ICP9145**	**ICEAP 00557**	**ICPV 87105**	**Mtawanjuni**
1	**CCB1**	222	222	222	222	220
2	**CCB7**	172	174	172	174	174
7	**CCac035**	267	267	267	265	265
15	**CCttc003**	196	196	196	193	193
37	**CCac026**	268	268	268	266	266
40	**CCttc019**	222	219	225/234	219	219/225/228

CHAPTER 5

5. Discussion

5.1 DNA Extraction and PCR amplification

High quality DNA extraction was achieved in this study, even without using the prescribed phenol: chloroform extraction described by Mace et al., (2003). This made the extraction both safer and cheaper by eliminating the use of phenol, which is hazardous and expensive to dispose of (Marechal-Drouard and Guillemaut, 1995). Moreover, some recently published extraction protocols exclude both phenol and chloroform and achieve quality results even from seeds (Meru et al., 2013). This method could not be used in this study as it was published after the work was completed. Such methods will be investigated for pigeonpea in future work to further simplify DNA extraction.

DNA extracted from seeds was degraded and of lower quality than that obtained from leaf material. Pigeonpea seed contains polysaccharides and polyphenols (Saxena et al., 2002). These compounds precipitate with the DNA after the adding of isopropanol/ethanol: sodium acetate and cause contamination of DNA (Sharma et al., 2003). Moreover polysaccharides inhibit *Taq* DNA polymerase action and the three DNA samples, which had the least successful amplification, were obtained from seeds (Pandey et al., 1996). Degradation of DNA is mainly caused by endonucleases (Sahu et al., 2012). However the pigeonpea seeds also contain phenolic terpenoids, which are also known to cause DNA degradation by binding to the DNA after cell lysis (Kim et al., 1997). Although the DNA obtained from seeds was degraded, it could still be used for PCR since SSR markers do not require high molecular weight DNA (Jones et al., 1997).

The total amount of DNA obtained from leaves ranged from a minimum of 6.33µg to 128.85µg with a mean of 55.08µg. This is much higher than the mean reported for pigeonpea (7.50µg) in the protocol used (Mace et al., 2003). Total DNA extracted from the seed samples ranged from 9.24µg to 66.10µg with an average of 43.95µg. The protocol used to extract DNA from the seeds reported an average yield of 50µg from soybean seeds and 35µg from chickpea seeds. The amount of DNA required for PCR with all the 48 primers is 1.14µg, so the extracted DNA from each sample was adequate for all the reactions.

PCR optimization is an important step to ensure the successful amplification of the target DNA fragment. Many aspects of any PCR protocol are recommended for consideration in optimization (Roux, 2009). However, this study focused only on the annealing temperature and primer concentration (Caetano-Anollés, 1998). This decision was made mainly because of ICRISAT-Nairobi experience with these and other primers used to analyse crops such as sorghum, groundnuts and pigeonpea in which optimizing annealing temperature usually solved non-amplification issues. Moreover, annealing temperature can be affected by primer concentration and adjustment of either could result in amplification (Roux, 2009). At first, amplification for 37 of the 48 primer pairs was successful using a fixed annealing temperature of 59°C. Eight of the remaining 11 primer pairs amplified the targeted SSRs when the annealing temperature was adjusted. It was only after increasing the amount of forward primer in the PCR reaction mixture and reducing the amount of fluorescently labeled M13 tag concentrations, that the last three primers show amplification bands. However, with the reduced fluorescent M13 tag, the resultant fragments did not incorporate enough fluorescence and could not be detected by the laser of the capillary electrophoresis machine. This has been experienced before in other studies that used labeled M13 sequences (Deshpande Santosh, pers. comm.). This problem can be avoided by

using directly labeled forward primers. However, due to limited funds and time in the current project, this was not done in this study. In addition, 45/48 markers (94%) of the markers tested, did amplify by PCR and this was considered sufficient for this study (Odeny et al., 2007; Saxena et al., 2010b). However, not all markers amplified equally well and another 8 had to be excluded from analysis. Although this is a large amount of data that was excluded from the analysis, the final number of 38 good markers compare well with other published studies on genetic diversity analysis where 30 to 40 SSR markers are typically considered adequate e.g. in pigeonpea (Saxena et al., 2010c), in groundnut (Tang et al., 2007), in wheat (Rousell et al., 2005) and in rice (Chakravathi and Naravaneni, 2006).

5.2 Allelic data analyses

As indicated in Table 5, allelic data analysis showed an average of 5.58 alleles per marker. This was higher than other pigeonpea diversity studies published to date, which used similar markers on cultivated varieties (Burns et al., 2001; Odeny et al., 2007). The major allele frequencies were generally high with a mean value of 0.76 (maximum possible value is 1), indicating that most alleles that occurred most frequently for each marker, occurred in a high proportion of the samples as opposed to a low major allele frequency, which would indicate that several alleles occur at a locus and is fairly evenly spread across the set of samples. Heterozygosity for the selected markers was generally low at mean 0.13 (minimum possible value of 0), indicating that at most marker loci, the same allele was observed on both chromosomes, an indication that these loci are stable and not prone to high outcrossing frequency or alternatively that the materials tested were genetically pure. Therefore these were good markers to use in genetic diversity studies as there should be little variation amongst the five individuals of each accession at these loci.

Diversity in cultivated pigeonpea is generally considered to be low (Saxena et al., 2002; Varshney et al., 2012). This was observed even when other types of markers were used (Yang et al., 2006; Panguluri et al., 2006). Consequently, studies that have included wild species reported higher PIC and allele number averages (Odeny et al., 2009). However there is potential to detect more polymorphism within cultivated varieties using newly developed SSRs from the pigeonpea genome project (Varshney et al., 2010). Despite the relatively low polymorphism, the markers grouped the genotypes clearly into three major groups, one containing a sub-group. Most of the released varieties (e.g. ICEAP 00040, ICEAP 00020, KAT60/8, ICEAP 00068 and ICEAP 00557) were developed from Kenyan and Tanzanian varieties and introduced to Malawi (Silim et al., 2005; Gwata et al., 2007). ICEAP 00068 and ICEAP 00557 are released varieties improved in Tanzania but in this study, they grouped in different clusters. The released varieties that were developed in Kenya (ICEAP 00040, ICEAP 00020, and KAT60/8) grouped together except for ICEAP 00040, which is in cluster III. All these released varieties have different durations of maturity and were selected and improved for traits such as disease resistance, high yields or drought tolerance (Silim et al., 2001). ICEAP 00040 and ICEAP 00020 are medium and long duration maturity genotypes, respectively, which are resistant to *Fusarium* wilt while ICEAP 00068, a medium duration, is susceptible but is popular with farmers as it yields large grains (Gwata et al., 2007). The genotypes ICPV 9145, ICP 13076 are ICRISAT-India accessions collected from Kenya although they group in different clusters. Both genotypes and ICPV 87105 have moderate resistance to *Fusarium* wilt (Silim et al., 1994). The obvious genetic differences observed between ICPV 9145 and ICP 13076 in this study could indicate possible different sources of *Fusarium* wilt resistance in these two varieties. This should be further investigated in

studies that link markers to the resistance genes to confirm if this is the case so that this diversity can be exploited in future in breeding programmes.

Although individuals of the same genotype grouped together for the most part, some were scattered e.g. ICP 9145 and ICEAP 00040. This is probably due to contamination/mixture of the seeds. Incidentally, these two are both long duration varieties, which exhibit some resistance to *Fusarium* wilt. Two landraces, Mtawanjuni and ICP 9145 were grouped with gene bank materials. Mtawanjuni is a popular traditional cultivar in Malawi. It is a high yielding medium duration variety, which farmers prefer due to its relatively good insect resistance. ICP 9145 is a Kenyan landrace and one of the first varieties to be introduced to Malawi in 1987. It is high yielding and has resistance to *Fusarium* wilt (Soko, 2000).

Natural outcrossing, due to insect pollination, is high in pigeonpea and is difficult and expensive to control in the fields since plants have to be isolated under insect-proof nets if outcrossing is to be avoided (Saxena et al., 1990). In Malawi, this is the cause of contamination of seeds in farmers' fields since many farmers plant more than one variety on their farms or have neighbours who plant different varieties whose flowering times overlap. For example, after obtaining pure Mtawanjuni seeds used in this study from breeders, other seeds of this variety were obtained randomly from different Malawi farmers. The seeds obtained from the farmers had five different seed coat colours and none was similar to seeds obtained from breeders. Such contamination can cause yield losses due to loss or dilution of insect or *Fusarium* wilt resistance and often closes market opportunities when mixtures give rise to different seed colours or seed size (Jones et al., 2002).

5.3 Genetic diversity

From the neighbour-joining tree (Figure 11) it is clear that there is substantial genetic diversity present in the germplasm analysed in this study. The released varieties were spread across all the clusters observed, indicating that they generally represented the genetic diversity available in Malawi. However, the major clusters showed only a single released variety and there was substantial variation that could still be exploited through further breeding. The markers used in this study were not linked to any traits of interest and this should be the next step in pigeonpea genomics to allow visualization of which varieties harbour important traits such as the different maturity duration, flowering times during a season, high yields, large, cream seeds, insect resistance (especially pod borers) and *Fusarium* wilt resistance (Bohra et al., 2012). Markers linked to these traits will allow scientists to determine if there are multiple sources – and therefore different mechanisms – that control these traits and which germplasm have the traits in order to transfer these to the best yielding and most popular varieties (Varshney et al., 2005; Odeny and Gebhardt, 2009). Markers linked to these traits will also allow pyramiding these traits into a select few varieties. The recent sequencing of the pigeonpea genome is a major step in this direction (Varshney et al., 2011).

5.4 DNA fingerprint

To my knowledge, there was no available software that could screen allelic data and identify markers suited for a DNA fingerprint. Therefore, this study attempted a logical approach to identify markers that will provide such a fingerprint and the criteria were developed accordingly (Saxena et al., 2010d). The six markers identified for the DNA fingerprint, generally had low heterozygosity and intermediate to high PIC scores according to the PowerMarker results of the

entire dataset (see Table 5 – presented before in Results section). Since the resulting number of markers and genotypes were both small, the fingerprint could be determined visually and is presented in table 6 above. In all cases, at least four out of the five individuals always presented the same alleles, except for individual ICEAP00557/3 and marker 37 where missing data reduced this number to 3/5. CCttc019 (Marker 40) was a heterozygous marker, which presented a monomorphic allele of 196bp for all individuals across all the released varieties. This allele was excluded for the fingerprint and only the second, polymorphic alleles from all varieties were included. When the combination of alleles for each variety across the six markers are considered, this preliminary DNA fingerprint for pigeonpea can discern each variety with confidence. However, this fingerprint will need to be further tested for robustness, repeatability and ability to discern admixtures due to cross-pollination.

CHAPTER 6

6. CONCLUSIONS AND RECOMENDATIONS

This study set out to investigate the level of genetic diversity in all cultivated Malawi pigeonpea varieties with SSR markers. While this was successful, it was observed that the level of diversity is low and further studies should exploit newly available SSR markers. It is also recommended that such studies include wild pigeonpea genotypes as they could reveal a new genetic resource. It was however noted that the released varieties are generally representative of the genetic diversity available in Malawi pigeonpea germplasm.

With a small number of markers it was possible to create a genetic fingerprint of the five most important pigeonpea varieties in Malawi. Although this needs to be tested further, it shows the potential of using SSR markers to discern pigeonpea varieties. Moreover, use of more polymorphic markers will increase the number of genotypes that can be discerned with the fingerprint. This can be used to detect seed contamination, a major cause of low yields, and ensure availability of high quality seeds for Malawi farmers.

Adequate high quality DNA was obtained from leaves despite omitting the phenol: chloroform extraction step. This and the advent of new methods that eliminate use of hazardous substances during DNA extraction show clearly that DNA extraction is becoming safer and cheaper.

CHAPTER 7

7. REFERENCES

Abdelati KA, Mohammed HAR, Ahmed ME (2009). Influence of feeding processed pigeonpea (*Cajanus cajan*) seeds on broiler chick performance. Journal of Poultry Science 8 (10): 971-975

Abunyewa AA, Karbo KN (2005). Improved fallow with pigeonpea for soil fertility improvement and to increase maize production in a smallholder crop–livestock farming system in the sub-humid zone of Ghana. Land Degradation and Development. 16:447–454. doi: 10.1002/ldr.672

Adu-Gyamfi JJ, Myaka F A, Sakala WD, Odgaard R, Vesterager JM, Høgh-Jensen H (2007). Biological nitrogen fixation and nitrogen and phosphorus budgets in farmer-managed intercrops of maize–pigeonpea in semi-arid southern and eastern Africa. Plant and Soil 295 (1-2), 127-136.

Amaefule KU, Obioha FC (2005). Performance of pullet chicks fed raw or processed pigeonpea (*Cajanus cajan*) seed meal diets. Livestock Research for Rural Development 17:29

Bohra A, Saxena RK, Gnanesh BN, Saxena K, Byregowda M, Rathore A, KaviKishor PB, Cook DR, Varshney RK (2012). An intra-specific consensus genetic map of pigeonpea [*Cajanus cajan* (L.) Millspaugh] derived from six mapping populations. Theoretical and Applied Genetics 125(6):1325-1338

Burns MJ, Edwards KJ, Newbury HJ, Ford LBR, Baggot CD (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

Caetano-Anollés G (1998). DAF optimization using Taguchi methods and the effect of thermal cycling parameters on DNA amplification. Biotechniques 25:472-480

Chakraborti M, Prasanna BM, Hossain F, Mazumdar S, Singh AM, Guleria S, Gupta HS (2011). Identification of kernel iron- and zinc-rich maize inbreds and analysis of genetic diversity using microsatellite markers. Journal of Plant Biochemistry and Biotechnology 20(2):224-233. DOI: 10.1007/s13562-011-0050-9

Chakravarthi BK, Naravaneni R (2006). SSR marker based DNA fingerprinting and diversity study in rice (*Oryza sativa*. L). African Journal of Biotechnology 5(9):684-688

Chaohong Z, Zhenghong L, Saxena KB, Jianqiu Z, Yong G, Shiying Y, Xuxiao Z (2001). Traditional and alternative uses of pigeonpea in China. International Chickpea and Pigeonpea Newsletter (8). pp. 55-57. ISSN 1023-4861

Choudhury PR, Singh IP, George B, Verma AK, Singh NP (2008). Assessment of genetic diversity of pigeonpea cultivars using RAPD analysis. Biologia Plantarum 52(4): 648-653

Data J, Lal N, Kaashyap M, Gupta PP (2010). Efficiency of Three PCR based Marker Systems for Detecting DNA Polymorphism in *Cicer arietinum* L and (*Cajanus cajan* (L) Millsp.). Genetic Engineering and Biotechnology Journal Vol 2010: GEBJ-5

Doveri S, Lee D, Maheswaran M, Powell W (2008). Molecular Markers - History, features and applications. In principles and practices of Plant Genomics Vol 1 Eds Kole C and Abott AG. Science Publishers, New Hampshire

Eneche EH (1999). Biscuit-making potential of millet/pigeonpea flour blends. Plant Foods for Human Nutrition 54:21-27

Emon RM, Gustafson JP, Nguyen H, Musket T, Jahiruddin M, Islam MA, Haque MS, Islam MM, Begum SN, Hassan MM (2010). Molecular marker-based characterization and genetic diversity of wheat genotypes in relation to Boron use efficiency. Indian Journal of Genetics and Plant Breeding 70(4):339-348

FAOSTATS (2010). http://faostat.fao.org/site/567/DesktopDefault.aspx?PageID=567#ancor

FAOSTAT (2011). http://faostat3.fao.org/home/index.html#VISUALIZE_BY_DOMAIN

Faris DG, Singh U (1990). Pigeonpea: nutrition and products. In The Pigeonpea (Nene, Y.L., Hall, S.D., and Sheila, V.K., eds.). Wallingford, Oxon, UK: CAB International. Pp 409 -434

Ganapathy KN, Gnanesh BN, Gowda MB, Venkatesha SC, Gomashe SS, Channamallikarjuna V (2011). AFLP analysis in pigeonpea (*Cajanus cajan* (L.) Millsp.)revealed close relationship of cultivated genotypes with some of its wild relatives. Genetic Resources Crop Evolution 58:837-847

Gupta PK, Roy JK, Prasad M (2001). Single nucleotide polymorphisms (SNPs): a new paradigm in molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Current Science 80(4): 524-535.

Gupta PK, Varshney RK (2000). The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113(3):163-185

Gupta PK, Varshney RK, Sharma PC, Ramesh B (1999). Molecular markers and their applications in wheat breeding. Plant breeding 118:369-390

Gwata ET, Mligo JK, Silim SN (2007). Registration of pigeonpea cultivar Tumia. Crop Science 47(1): 436

Helentjaris T, King G, Slocum M, Sidenstrant C, Wegman S (1985). Restriction fragment length polymorphism as probes for plant diversity and their development as tools for applied plant breeding. Plant Molecular Biology 5:109-118

Jha SS, Ohri D (1996). Phylogenetic relationships of *Cajanus cajan* (L.) Millsp. (pigeonpea) and its wild relatives based on seed protein profiles. Genetic Resources and Crop Evolution 43:275-281

Jin L, Lu Y, Xiao P, Sun M, Corke H, Bao J (2010). Genetic diversity and population structure of a diverse set of rice germplasm for association mapping. Theoretical Applied Genetics 121:475-487

Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, van de Wiel C, Bredemeijer G, Vosman B, Matthes M, Daly A, Brettschneider R, Bettini P, Buiatti M, Maestri E, Malcevschi A, Marmiroli N, Aert R, Volckaert G, Rueda J, Linacero R, Vazquez A, Karp A (1997). Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Molecular Breeding 3:381-390

Jones R, Freeman HA, Le Monaco G (2002). Improving the access of small farmers in Eastern and Southern Africa to global pigeonpea markets. Agricultural Research Network Paper no 120, January 2002

Kassa, MT (2011). Molecular analysis of genetic diversity in domesticated pigeonpea (*Cajanus cajan* (L.) Millsp.) and wild relatives. PhD thesis, Rhodes University

Kassa MT, Penmetsa RV, Carrasquilla-Garcia N, Sarma BK, Datta S, Upadhyaya HD, Varshney RK, von Wettberg EJB, Cook DR (2012). Genetic Patterns of Domestication in Pigeonpea (*Cajanus cajan* (L.) Millsp.)and Wild *Cajanus* Relatives. PLoS ONE. 7(6): e39563. DOI:10.1371/journal.pone.0039563

Kim CS, Lee CH, ShinJS, Chung YS, Hyung NI (1997). A simple and rapid method for isolation of high quality genomic DNA from fruit trees and conifers using PVP. Nucleic Acids Research 25(5):1085–1086

Kollipara KP, Singh L, Hymowitz T (1994). Genetic variation of trypsin and chymotrypsin inhibitors in pigeonpea (*Cajanus cajan* (L) Millsp.) and its wild relatives. Theoretical and Applied Genetics 8:983-986

Kotresh H, Fakrudin B, Punnuri SM, Rajkumar BK, Thudi M, Paramesh H, Lohithaswa H, Kuruvinashetti MS (2006). Identification of two RAPD markers genetically linked to a recessive allele of a Fusarium wilt resistance gene in pigeonpea (*Cajanus cajan* L. Millsp.). Euphytica 149:113–120

Koumi P, Green HE, Hartley S, Jordan D, Lahec S, Livett RJ, Tsang KW, Ward DM (2004). Evaluation and validation of the ABI 3700, ABI 3100, and the MegaBACE 1000 capillary array electrophoresis instruments for use with short tandem repeat microsatellite typing in a forensic environment. Electrophoresis 25:2227–2241

Kumar Rao JVDK, Dart PJ, Sastry PVSS (1983). Residual effect of pigeonpea (Cajanus cajan (L.) Millsp.) on yield and nitrogen response of maize. Experimental Agriculture 19:131–141

Ladizinsky G, Hamel A (1980). Seed protein profiles of pigeonpea (*Cajanus cajan* (L.) Millsp.) and some *Atylosia* species. Euphytica 29:313-317

Liu K, Muse VS (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21(9): 2128-2129

Long Y, Jian-ping G, Xu-xiao Z (2004). Optimization of AFLP fingerprinting and screening of primer pairs in genetic diversity analysis of pigeonpea germplasm. Journal of Plant Genetic Resources 5:342-345

Loridon K, McPhee K, Morin J, Dubreuil P, Pilet-Nayel ML, Aubert G, Rameau C, Baranger A, Coyne C, Lejeune-Henaut I, Burstin J (2005). Microsatellite marker polymorphism and mapping in pea (*Pisum sativum* L.). Theoretical and Applied Genetics 111:1022-1031

Mace ES, Buhariwalla KK, Buhariwalla HK, Crouch JH (2003). A high-throughput DNA extraction protocol for tropical molecular breeding programs. Plant Molecular Biology Reporter 21(4):459-460. DOI:10.1007/BF02772596

Mammadov J, Aggarwal R, Buyyarapu R, Kumpatla S (2012). SNP markers and their impact on plant breeding. International Journal of Plant Genomics DOI:10.1155/2012/728398

Maréchal-Drouard L, Guillemaut P (1995). A powerful but simple technique to prepare polysaccharide-free DNA quickly and without phenol extraction. Plant Molecular Biology Reporter 13(1):26-30.

Meru G, McDowell D, Waters V, Seibel A, Davis J, McGregor C (2013). A non-destructive genotyping system from a single seed for marker-assisted selection in watermelon. Genetic and Molecular Research 12(1):702-709

Mohammadi SA, Prasanna BM. (2003). Analysis of genetic diversity in crop plants—salient statistical tools and considerations. Crop Science 43: 1235–1248

Mueller UG, Wolfenbarger LL (1999). AFLP genotyping and fingerprinting. Trends in Ecology and Evolution14 (10):389–394

Nadimpalli BG, Jaret RL, Pathak SC, Kochert G (1992). Phylogenetic relationships of the pigeonpea (*Cajanus cajan*) based on nuclear restriction fragment length polymorphism. Genome 36:216-223

Nwokolo E (1987). Nutritional evaluation of pigeonpea meal. Plant Foods for Human Nutrition 37(4): 283-290

Odeny DA (2007). The potential of pigeonpea (*Cajanus cajan* (L.) Millsp.) in Africa. Natural Resources Forum 31:297–305

Odeny DA, Gebhardt C (2009). Towards molecular breeding of pigeonpea (*Cajanus cajan* (L.) Millsp.): a case for increased production in Africa. Acta Horticulturae (ISHS) 806:151-154

Odeny DA, Jayashree B, Ferguson M, Hoisington D, Cry LJ, Gebhardt C (2007). Development, characterization and utilization of microsatellite markers in pigeonpea. Plant Breeding 126:130–136

Odeny DA, Jayashree B, Gebhardt C, Crouch J (2009). New microsatellite markers for pigeonpea (*Cajanus cajan* (L.) Millsp.). BMC Research Notes 2:35

Oshodi AA, Ekperigin MM (1989). Functional properties of pigeonpea (*Cajanus cajan*) flour. Food Chemistry 34(3):187-191

Panguluri DK, Janaiah K, Govil JN, Kumar PA, Sharma PC (2006). AFLP fingerprinting in pigeonpea (*Cajanus cajan* (L.) Millsp.) and its wild relatives. Genetic Resources and Crop Evolution 53: 523–531

Pandey RN, Adams RP, Flournoy LE (1996). Inhibition of random amplified polymorphic DNAs (RAPDs) by plant polysaccharides. Plant Molecular Biology Reporter 14(1):17–22

Panigrahi J, Kumar DR, Mishra M, Mishra RP, Jena P (2007). Genomic relationships among 11 species in the genus *Cajanus* as revealed by seed protein (albumin and globulin) polymorphisms. Plant Biotechnology Reporter 1:109–116

Perrier X, Jacquemoud-Collet JP (2006). DARwin software. http://darwin.cirad.fr/

Powell W, Machray GC, Provan J (1996). Polymorphism revealed by simple sequence repeats. Trends in Plant Science 1:215–222.

Pundir RPS, Singh RB (1985). Biosystematic relationships among *Cajanus*, *Atylosia*, and *Rhynchosia* species and evolution of pigeonpea (*Cajanus cajan* (L.) Millsp.). Theoretical and Applied Genetics 69:531-534

Rafalski JA, Tingey SV (1993). Genetic diagnostics in plant breeding: RAPDS, Microsatellites and Machines. Trends in Genetics 9:275-280

Rao MR, Mathuva MN (2000). Legumes for improving maize yields and income in semi-arid Kenya. Agriculture Ecosystems and Environment 78:123–137

Rao MR, Willey RW (1980). Evaluation of yield stability in intercropping: studies on sorghum/pigeonpea. Experimental Agriculture.16 (2), 105-116

Rao SC, Coleman, SW, Mayeux HS (2002). Forage production and nutritive value of selected pigeonpea ecotypes in the southern Great Plains. Crop Science 42:1259–1263

Ratnaprkhe MB, Gupta VS, Ven Murthy MR, Ranjekar PK (1995). Genetic fingerprinting of pigeonpea [*Cajanus cajan* (L.) Millsp.] and its wild relatives using RAPD markers. Theoretical and Applied Genetics 91:893-898

Roussel V, Leisova L, Exbrayat F, Stehno Z, Balfourier F (2005). SSR allelic diversity changes in 480 European bread wheat varieties released from 1840 to 2000. Theoretical and Applied Genetics 111(1):162-170

Roux KH (2009). Optimization and troubleshooting in PCR. Cold Spring Harbour Protocols 2009.4: pdb-ip66

Sahu SK, Thangaraj M, Kathiresan K (2012). DNA extraction protocol for plants with high levels of secondary metabolites and polysaccharides without using liquid nitrogen and phenol. International Scholarly Research Network Molecular Biology DOI:10.5402/2012/205049

Salgado, KCPC, Vieira MGGC, Pinho EVRV, Guimaraes CT, Pinho RGV, and Souza LV (2006). Genetic purity certificate in seeds of hybrid maize using molecular markers. Revista Brasileira de Sementes 28:169-175

Saxena K, Kumar R, Sultana R (2010a). Quality nutrition through pigeonpea - a review. Health 2:1335-1344

Saxena KB, Kumar RV, Rao PV (2002). Pigeonpea Nutrition and Its Improvement. Journal of Crop Production 5:1-2

Saxena KB, Singh L Gupta MD (1990). Variation for natural out-crossing in pigeonpea. Euphytica 46:143-148

Saxena RK, Prathima C, Saxena K, Hoisington DA, Singh NK, Varshney RK (2010b). Novel SSR Markers for Polymorphism Detection in pigeonpea (*Cajanus* spp.). Plant Breeding 129:142–148

Saxena RK, Saxena K, Varshney RK (2010d). Application of SSR markers for molecular characterization of hybrid parents and purity assessment of ICPH 2438 hybrid of pigeonpea [*Cajanus cajan* (L.) Millspaugh]. Molecular Breeding 26:371-380

Saxena RK, Saxena KB, Kumar RV, Hoisington DA, Varshney RK (2010c). Simple sequence repeat-based diversity in elite pigeonpea genotypes for developing mapping populations to map resistance to Fusarium wilt and sterility mosaic disease. Plant Breeding 129:135-241

Semagn K, Bjonstad A, Ndjiondjop MN (2006). An overview of molecular marker methods for plants. African Journal of Biotechnology 5(25):2540-2568

Sharma AD, Gill PK, Singh P (2003). DNA isolation from dry and fresh samples of polysaccharide-rich plants. Plant Molecular Biology Reporter 20(4): 415.

Shuelke M. (2000). An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology.18:233-234

Silim SN, Gwata ET, Mligo JK, Siambi M, Karuru O, King SB, Omanga P (2005). Registration of pigeonpea cultivar 'ICEAP 00040'. Crop Science 45(6):2647

Silim SN, Tuwafe S, Singh L eds. (1994). Pigeonpea Improvement in Eastern and Southern Africa - Annual Research Planning Meeting 1993, 25-27 Oct 1993.International Crops Research Institute for the Semi-Arid Tropics, Patancheru, Andhra Pradesh, India ISBN 9290662956

Simtowe F, Asfaw S, Shiferaw B, Siambi M, Monyo E, Muricho G, Abate T, Silim S, Ganga Rao NVPR and Madzonga O (2010). Socioeconomic Assessment of Pigeonpea and Groundnut Production Conditions – Farmer Technology Choice, Market Linkages, Institutions and Poverty in Rural Malawi. Research Report no.6. Patancheru 502 324, Andhra Pradesh, India: International Crops Research Institute for the Semi-Arid Tropics 92 pp

Simtowe F, Bekele S, Menale K, Monyo E, Silim S, Muricho G (2009). Assessment of the Current Situation and Future Outlooks for the pigeonpea Sub-Sector in Malawi. ICRISAT-Nairobi

Sivaramakrishnan S, Kannan S, Reddy LJ (2002). Diversity in selected wild and cultivated species of pigeonpea using RFLP of mtDNA. Euphytica 125:21-28

Sneller CH, Nelson RL, Carter TE, Cui Z (2005). Genetic diversity in crop improvement: The soybean experience. Journal of Crop Improvement 14:1-2

Soko HN (2000). Pigeonpea Research: Technology Exchange and Seed Production in Malawi. In: Silim SN, Mergeai G, Kimani PM (Eds.). Status and Potential of pigeonpea in eastern and southern Africa: Proceedings of a regional workshop, 12-15 Sept. 2000, Nairobi, Kenya. ICRISAT, pp 232

Tang R, Gao G, He L, Han Z, Shan S, Zhong R, Zhou C, Jiang J, Li Y, Zhuang W (2007). Genetic diversity in cultivated groundnut based on SSR markers. Journal of Genetics and Genomics 34(5):449-459

Upadhyaya HD, Reddy KN, Sharma S, Varshney RK, Bhattacharjee R, Singh S, Gowda LL (2011). Pigeonpea composite collection and identification of germplasm for use in crop improvement programmes. Plant Genetic Resources 9:97-108

van der Maesen LJG (1980). India is the native home of the pigeonpea. In: Liber Gratulatorius in nonerem H.C.D. de Wit. (Arends, J.C., Boelema, G., de Groot, C.T., and Leeuwenberg, A.J.M., eds.) Landbouwhoge school Miscellaneous Paper no.19, Wageningen, The Netherlands: H. Veenman and B.V. Zonen. pp 257-262

van der Maesen LJG (1981). ICRISAT (International Crops Research Institute for the Semi-Arid Tropics). Proceedings of the International Workshop on Pigeonpea, Volume 2, pp. 15-19 December 1980, Patancheru, A.P., India

Van der Maesen LJG (1983). World distribution of pigeonpea. Information bulletin No.14 Patacheru, A.P., India: International Crops Research Institute for The Semi-Arid Tropics

Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MT, Azam S, Fan G, Whaley AM, Farmer AD, Sheridan J, Iwata A, Tuteja R, Penmetsa RV, Wu W, Upadhyaya HD, Yang SP, Shah T, Saxena KB, Michael T, McCombie WR, Yang B, Zhang G, Yang H, Wang J, Spillane C, Cook DR, May GD, Xu X, Jackson SA (2011). Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nature Biotechnology 30(1):83-89

Varshney RK, Close TJ, Singh NK, Hoisington DA, Cook DR (2009). Orphan legume crops enter the genomics era! Current Opinion in Plant Biology 12:202-210

Varshney RK, Graner A, Sorrells ME (2005). Genomics-assisted breeding for crop improvement. Trends in Plant Science 10(12):621-30

Varshney RK, Kudapa H, Roorkiwal M, Thudi M, Pandey MK, Saxena RK, Chamarthi SK, Mohan SM, Mallikarjuna N, Upadhyaya H, Gaur PM, Krishnamurthy L, Saxena KB, Nigam SN and Pande S (2012). Advances in genetics and molecular breeding of three legume crops of semi-arid tropics using next-generation sequencing and high-throughput genotyping technologies. Journal of Biosciences 37:811–820.

Varshney RK, Penmetsa RV, Dutta S, Kulwal PL, Saxena RK, Data S, Sharma TR, Rosen B, Carrasquilla-Garcia N, Farmer AD, Dubey A, Saxena KB, Gao J, Fakrudin B, Singh MN, Singh BP, Wanjari KB, Yuan M, Srivastava RK, Kilian A, Upadhyaya HD, Mallikarjuna N, Town CD, Bruening GE, He G, May GD, McCombie R, Jackson SA, Singh NK, Cook DR (2010). Pigeonpea genomics initiative (PGI): an international effort to improve crop productivity of pigeonpea (*Cajanus cajan* L.). Molecular Breeding 26:393–408

Vos P, Hogers R, Reijans M, van de Lee T, Hornes M, Friters A, Pot J, Peleman J, Kupier M, Zabeau M (1995). AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23:4404-4414

Wasike S, Okori P, Rubaihayo PR (2005). Genetic variability of the Asian and African pigeonpea as revealed by AFLP. African Journal of Biotechnology 4(11):1228-1233

Whiteman PC, Norton BW (1981). Alternative uses of pigeonpea. In: Proceedings of the International Workshop on Pigeonpea, 15-19 Dec 1980, ICRISAT, India. Vol. 1. Patancheru, A.P. 502324, India: International Crops Research Institute for the Semi-Arid Tropics. pp 365-378

Williams JGK, Kubelik ARK, Livak JL, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by random primers are useful genetic markers. Nucleic Acids Research 18:6531-6535

Wu N, Fu K, Fu YJ, Zu YG, Chang FR, Chen YH, Liu XL, Kong Y, Liu W, Gu CB (2009). Antioxidant activities of extracts and main components of pigeonpea [*Cajanus cajan* (L.) Millsp.] leaves. Molecules 14:1032-1043

Yang S, Wen P, Ash G, Harper J, Carling J, Wenzl P, Huttner E, Zong X, Kilian A (2006). Low level of genetic diversity in cultivated pigeonpea compared to its wild relatives is revealed by diversity arrays technology. Theoretical and Applied Genetics 113:585-595

APPENDICES

Appendix A. Pigeonpea accessions used in this study

Accession	Type	Maturity
ICEAP 00040	Released Variety	Long
ICEAP 00557	Released Variety	Medium
ICEAP 00020	Released Variety	Long
ICEAP 00068	Released Variety	Medium
ICPV 87105	Landrace	Short
ICPV 9145	Landrace	Long
Mtawanjuni	Landrace	Medium
ICP 13076	Landrace	Long
Kat 60/8	Released Variety	Short
MW 326	Malawi Gene bank	
MW 454	Malawi Gene bank	
MW 470	Malawi Gene bank	
MW 480	Malawi Gene bank	
MW 587	Malawi Gene bank	
MW 648	Malawi Gene bank	
MW 690	Malawi Gene bank	
MW 765	Malawi Gene bank	
MW 786	Malawi Gene bank	
MW 793	Malawi Gene bank	
MW 2047	Malawi Gene bank	
MW 2097	Malawi Gene bank	
MW 2238	Malawi Gene bank	
MW 2240	Malawi Gene bank	
MW 2243	Malawi Gene bank	
MW 2244	Malawi Gene bank	
MW 2245	Malawi Gene bank	
MW 2251	Malawi Gene bank	
MW 2256	Malawi Gene bank	
MW 2258	Malawi Gene bank	
MW 2261	Malawi Gene bank	
MW 2263	Malawi Gene bank	
MW 2264	Malawi Gene bank	
MW 2265	Malawi Gene bank	
MW 2266	Malawi Gene bank	
MW 2267	Malawi Gene bank	
MW 2268	Malawi Gene bank	
MW 2269	Malawi Gene bank	
MW 2270	Malawi Gene bank	
MW 2271	Malawi Gene bank	
MW 2276	Malawi Gene bank	
MW 2279	Malawi Gene bank	

Accession	Type	Maturity
MW 2281	Malawi Gene bank	
MW 2282	Malawi Gene bank	
MW 2283	Malawi Gene bank	
MW 2284	Malawi Gene bank	
MW 2285	Malawi Gene bank	
MW 2286	Malawi Gene bank	
MW 2287	Malawi Gene bank	
MW 2288	Malawi Gene bank	
MW 2289	Malawi Gene bank	
MW 2291	Malawi Gene bank	
MW 2292	Malawi Gene bank	
MW 2295	Malawi Gene bank	
MW 2296	Malawi Gene bank	
MW 2298	Malawi Gene bank	
MW 2299	Malawi Gene bank	
MW 2300	Malawi Gene bank	
MW 2302	Malawi Gene bank	
MW 2303	Malawi Gene bank	
MW 2305	Malawi Gene bank	
MW 2306	Malawi Gene bank	
MW 2308	Malawi Gene bank	
MW 2309	Malawi Gene bank	
MW 2311	Malawi Gene bank	
MW 2313	Malawi Gene bank	
MW 2317	Malawi Gene bank	
MW 2321	Malawi Gene bank	
MW 2323	Malawi Gene bank	
MW 2324	Malawi Gene bank	
MW 2325	Malawi Gene bank	
MW 2326	Malawi Gene bank	
MW 2327	Malawi Gene bank	
MW 2328	Malawi Gene bank	
MW 2331	Malawi Gene bank	
MW 2332	Malawi Gene bank	
MW 2333	Malawi Gene bank	
MW 2336	Malawi Gene bank	
MW 2355	Malawi Gene bank	
MW 2869	Malawi Gene bank	

Appendix B. Additional information on the origin of released varieties and landraces

Genotype	Maturity	Origin (improved or landrace)
ICEAP 00040	Long	Kitui, Eastern Kenya; improved variety
ICP 9145	Long	India (Collected in Kenya by ICRISAT)
ICEAP 00557	Medium	Southern Tanzania, through ICRISAT-Nairobi; improved through selection
ICP 87105	Short	India (Collected in Kenya by ICRISAT)
Mtawajuni	Medium	Landrace, Malawi
ICEAP 00020	Long	Kitui, Eastern Kenya; Improved variety
ICEAP 00068	Medium	Tanzania, Masasi, through ICRISAT-Nairobi
ICP 13076	Long	India (Collected in Kenya by ICRISAT)
KAT 60/8	Medium	Katumani, Eastern Kenya; improved variety
ICP 2309		Nepal

Appendix C. Nanodrop© readings for extracted DNA from all the samples

Sample Name	DNA Concentrations in ng/µl	$A_{260/280}$	Sample used for Extraction
MW 786_5	443.77	1.64	Seed
MW 786_4	616.03	1.64	Seed
MW 765_5	406.92	1.51	Seed
MW 765_4	395.43	1.55	Seed
MW 765_3	368.52	1.60	Seed
MW 765_2	443.70	1.74	Seed
MW 765_1	393.75	1.87	Seed
MW 690_5	520.14	1.49	Seed
MW 648_5	444.89	1.70	Seed
MW 648_4	510.56	1.59	Seed
MW 648_3	457.28	1.80	Seed
MW 648_2	499.73	1.69	Seed
MW 648_1	436.52	1.67	Seed
MW 587_5	409.65	1.62	Seed
MW 587_4	510.03	1.63	Seed
MW 587_3	492.11	1.68	Seed
MW 587_2	392.99	1.65	Seed
MW 587_1	457.01	1.70	Seed
MW 480_5	311.89	1.92	Seed
MW 480_4	430.41	1.78	Seed
MW 480_3	688.52	1.97	Seed
MW 480_2	283.07	1.85	Seed
MW 470_5	378.26	1.38	Seed
MW 470_4	365.97	1.50	Seed
MW 470_3	405.71	1.54	Seed
MW 454_5	684.40	1.69	Seed
MW 454_4	537.66	1.62	Seed
MW 326_5	442.68	1.62	Seed
MW 326_4	604.56	1.64	Seed
MW 2333_5	537.87	1.61	Seed
MW 2327_5	427.17	1.33	Seed
MW 2327_4	631.92	1.30	Seed
MW 2327_3	420.27	1.25	Seed
MW 2327_2	592.81	1.29	Seed
MW 2313_5	594.10	1.61	Seed
MW 2313_4	506.15	1.54	Seed
MW 2305_5	656.96	1.64	Seed
MW 2298_5	605.68	1.62	Seed
MW 2298_4	593.89	1.45	Seed
MW 2298_3	596.22	1.71	Seed
MW 2295_5	623.69	1.64	Seed
MW 2289_5	373.69	1.74	Seed
MW 2289_4	307.42	1.87	Seed
MW 2289_3	440.02	1.88	Seed
MW 2289_2	477.50	1.90	Seed
MW 2289_1	96.30	2.00	Seed
MW 2287_5	465.46	1.58	Seed
MW 2287_4	346.88	1.58	Seed
MW 2287_3	302.84	1.60	Seed

Sample Name	DNA Concentrations in ng/µl	$A_{260/280}$	Sample used for Extraction
MW 2296_8	555.40	1.86	Leaf
MW 2296_6	546.81	1.86	Leaf
MW 2296_4	507.24	1.96	Leaf
MW 2296_3	492.76	1.97	Leaf
MW 2296_2	485.88	1.96	Leaf
MW 2295_4	383.68	1.93	Leaf
MW 2295_3	427.01	1.89	Leaf
MW 2295_2	305.57	1.90	Leaf
MW 2295_1	350.72	1.92	Leaf
MW 2292_5	364.18	1.90	Leaf
MW 2292_4	349.47	1.89	Leaf
MW 2292_3	437.86	1.89	Leaf
MW 2292_2	912.51	2.06	Leaf
MW 2292_1	1015.80	2.04	Leaf
MW 2291_5	477.10	1.80	Leaf
MW 2291_4	605.04	1.99	Leaf
MW 2291_3	263.32	1.81	Leaf
MW 2291_2	330.55	1.96	Leaf
MW 2291_1	398.17	1.85	Leaf
MW 2288_5	390.10	1.89	Leaf
MW 2288_4	283.05	1.91	Leaf
MW 2288_3	353.78	1.89	Leaf
MW 2288_2	167.96	1.90	Leaf
MW 2288_1	921.36	2.05	Leaf
MW 2287_2	309.70	1.88	Leaf
MW 2287_1	790.49	2.07	Leaf
MW 2286_3	138.16	1.90	Leaf
MW 2286_2	406.70	1.98	Leaf
MW 2286_1	360.46	1.95	Leaf
MW 2285_5	405.43	1.96	Leaf
MW 2285_4	606.85	2.01	Leaf
MW 2285_3	593.44	1.98	Leaf
MW 2285_2	581.91	1.96	Leaf
MW 2285_1	410.09	1.96	Leaf
MW 2284_5	783.08	1.90	Leaf
MW 2284_4	588.21	1.91	Leaf
MW 2284_3	764.33	1.93	Leaf
MW 2284_2	462.50	1.93	Leaf
MW 2284_1	1123.15	2.06	Leaf
MW 2283_4	114.30	1.90	Leaf
MW 2283_3	211.66	1.91	Leaf
MW 2283_2	503.15	1.98	Leaf
MW 2283_1	363.06	1.83	Leaf
MW 2282_9	518.06	1.88	Leaf
MW 2282_8	594.66	1.89	Leaf
MW 2282_7	581.02	1.89	Leaf
MW 2282_2	955.03	2.07	Leaf
MW 2282_1	1211.21	2.05	Leaf
MW 2281_2	230.90	1.91	Leaf

Sample Name	DNA Concentration in ng/µl	$A_{260/280}$	Sample used for Extraction
MW 2286_5	292.40	1.49	Seed
MW 2286_4	382.13	1.51	Seed
MW 2283_5	527.92	1.61	Seed
MW 2281_5	509.11	1.50	Seed
MW 2281_4	598.71	1.58	Seed
MW 2281_3	441.03	1.63	Seed
MW 2270_5	684.73	1.52	Seed
MW 2270_4	513.65	1.55	Seed
MW 2265_5	485.81	1.46	Seed
MW 2245_5	422.21	1.44	Seed
MW 2243_5	125.05	1.77	Seed
MW 2243_2	396.69	1.19	Seed
MW 2240_5	467.96	1.70	Seed
MW 2240_4	325.44	1.72	Seed
MW 2240_3	401.75	1.63	Seed
MW 2240_2	316.85	1.75	Seed
MW 2240_1	329.68	1.78	Seed
MW 2238_5	524.41	1.55	Seed
MW 793_7	565.27	1.91	Leaf
MW 793_6	384.40	1.89	Leaf
MW 793_5	416.66	1.89	Leaf
MW 793_2	1042.85	2.04	Leaf
MW 793_1	765.01	2.02	Leaf
MW 786_3	189.36	1.89	Leaf
MW 786_2	594.58	2.02	Leaf
MW 786_1	853.20	2.05	Leaf
MW 690_4	233.98	1.90	Leaf
MW 690_3	926.87	2.01	Leaf
MW 690_2	748.07	2.05	Leaf
MW 690_1	685.73	2.02	Leaf
MW 480_1	242.69	1.93	Leaf
MW 470_2	895.78	2.04	Leaf
MW 470_1	1342.21	2.04	Leaf
MW 454_3	345.47	1.90	Leaf
MW 454_2	189.31	1.92	Leaf
MW 454_1	276.45	1.92	Leaf
MW 326_3	224.16	1.93	Leaf
MW 326_2	624.08	2.02	Leaf
MW 326_1	951.48	2.01	Leaf
MW 2355_9	456.53	1.88	Leaf
MW 2355_8	455.71	1.89	Leaf
MW 2355_6	360.27	1.88	Leaf
MW 2355_5	289.83	1.87	Leaf
MW 2336_5	360.41	1.97	Leaf
MW 2336_4	461.26	1.91	Leaf
MW 2336_3	670.98	1.99	Leaf
MW 2336_2	581.06	1.84	Leaf
MW 2336_1	591.86	1.97	Leaf
MW 2333_4	189.52	1.97	Leaf
MW 2333_3	65.99	1.88	Leaf
MW 2333_2	92.76	1.94	Leaf

Sample Name	DNA Concentration in ng/µl	$A_{260/280}$	Sample used for Extraction
MW 2281_1	327.40	1.97	Leaf
MW 2279_6	408.75	1.88	Leaf
MW 2279_3	1236.22	2.05	Leaf
MW 2279_2	1313.73	2.06	Leaf
MW 2279_10	735.44	1.93	Leaf
MW 2279_1	913.80	2.08	Leaf
MW 2276_7	355.90	1.88	Leaf
MW 2276_6	306.58	1.91	Leaf
MW 2276_5	400.03	1.89	Leaf
MW 2276_4	327.91	1.94	Leaf
MW 2276_3	563.65	1.99	Leaf
MW 2271_9	423.73	1.88	Leaf
MW 2271_7	422.78	1.88	Leaf
MW 2271_6	401.04	1.89	Leaf
MW 2271_4	475.27	1.87	Leaf
MW 2271_3	578.64	1.82	Leaf
MW 2270_3	763.02	1.96	Leaf
MW 2270_2	679.59	1.93	Leaf
MW 2270_1	437.17	1.91	Leaf
MW 2269_9	507.44	1.86	Leaf
MW 2269_8	451.50	1.88	Leaf
MW 2269_7	394.59	1.88	Leaf
MW 2269_6	463.60	1.87	Leaf
MW 2269_5	494.16	1.88	Leaf
MW 2268_8	667.83	1.91	Leaf
MW 2268_7	481.91	1.94	Leaf
MW 2268_6	282.27	1.93	Leaf
MW 2268_5	553.61	1.91	Leaf
MW 2268_4	539.27	1.91	Leaf
MW 2267_6	672.38	2.09	Leaf
MW 2267_5	873.38	2.04	Leaf
MW 2267_4	618.63	2.00	Leaf
MW 2267_13	364.09	1.95	Leaf
MW 2267_12	687.43	1.98	Leaf
MW 2266_6	484.50	1.89	Leaf
MW 2266_5	672.90	2.00	Leaf
MW 2266_4	1034.86	2.10	Leaf
MW 2266_3	856.98	2.08	Leaf
MW 2266_2	1094.13	2.04	Leaf
MW 2265_4	692.54	2.01	Leaf
MW 2265_3	636.19	2.01	Leaf
MW 2265_2	595.79	2.00	Leaf
MW 2265_1	203.32	2.00	Leaf
MW 2264_9	781.92	1.97	Leaf
MW 2264_8	664.63	1.96	Leaf
MW 2264_7	638.27	1.96	Leaf
MW 2264_6	576.73	1.95	Leaf
MW 2264_2	610.82	1.94	Leaf
MW 2263_5	870.21	1.94	Leaf
MW 2263_4	677.41	1.91	Leaf
MW 2263_3	944.42	2.04	Leaf

Sample Name	DNA Concentration in ng/µl	$A_{260/280}$	Sample used for Extraction
MW 2333_1	555.77	1.96	Leaf
MW 2332_7	359.04	1.88	Leaf
MW 2332_5	438.48	1.86	Leaf
MW 2332_4	306.37	1.96	Leaf
MW 2332_3	502.26	1.96	Leaf
MW 2332_1	350.85	1.96	Leaf
MW 2331_9	352.83	1.86	Leaf
MW 2331_8	338.75	1.85	Leaf
MW 2331_6	405.86	1.84	Leaf
MW 2331_5	367.69	1.91	Leaf
MW 2331_4	388.49	1.96	Leaf
MW 2328_9	349.07	1.86	Leaf
MW 2328_8	401.09	1.86	Leaf
MW 2328_7	346.23	1.87	Leaf
MW 2328_6	457.09	1.85	Leaf
MW 2328_5	456.81	1.86	Leaf
MW 2327_1	470.20	1.85	Leaf
MW 2326_9	428.16	1.89	Leaf
MW 2326_8	425.05	1.89	Leaf
MW 2326_7	360.15	1.88	Leaf
MW 2326_6	295.57	1.89	Leaf
MW 2326_4	343.85	1.88	Leaf
MW 2325_8	439.27	1.89	Leaf
MW 2325_7	424.06	1.91	Leaf
MW 2325_6	919.79	2.06	Leaf
MW 2325_5	967.49	2.05	Leaf
MW 2325_4	580.93	2.00	Leaf
MW 2324_7	469.00	1.89	Leaf
MW 2324_6	335.41	1.93	Leaf
MW 2324_5	468.67	1.89	Leaf
MW 2324_4	517.30	2.02	Leaf
MW 2324_3	903.16	2.05	Leaf
MW 2323_7	245.59	1.88	Leaf
MW 2323_6	302.77	1.87	Leaf
MW 2323_4	423.30	1.88	Leaf
MW 2323_3	456.31	1.94	Leaf
MW 2323_2	460.26	1.97	Leaf
MW 2321_7	278.79	1.87	Leaf
MW 2321_6	332.73	1.90	Leaf
MW 2321_4	443.80	1.87	Leaf
MW 2321_2	743.68	1.89	Leaf
MW 2321_1	653.75	1.99	Leaf
MW 2317_5	412.89	1.90	Leaf
MW 2317_4	262.46	1.94	Leaf
MW 2317_3	199.76	1.97	Leaf
MW 2317_2	394.13	1.92	Leaf
MW 2317_1	1190.06	2.06	Leaf
MW 2313_3	521.86	1.90	Leaf
MW 2313_2	103.84	1.93	Leaf
MW 2313_1	808.65	2.04	Leaf
MW 2311_8	457.52	1.89	Leaf

Sample Name	DNA Concentration in ng/µl	$A_{260/280}$	Sample used for Extraction
MW 2263_2	926.90	2.01	Leaf
MW 2263_1	1008.37	2.03	Leaf
MW 2261_7	492.37	1.87	Leaf
MW 2261_6	806.34	1.96	Leaf
MW 2261_5	486.60	1.88	Leaf
MW 2261_2	559.43	1.93	Leaf
MW 2261_1	618.36	1.98	Leaf
MW 2258_8	660.57	1.95	Leaf
MW 2258_7	530.57	1.92	Leaf
MW 2258_6	641.89	1.95	Leaf
MW 2258_4	1151.35	2.04	Leaf
MW 2258_3	1182.75	2.05	Leaf
MW 2256_6	340.10	1.90	Leaf
MW 2256_5	440.15	1.90	Leaf
MW 2256_4	472.74	1.92	Leaf
MW 2256_2	750.38	1.98	Leaf
MW 2256_1	602.40	1.99	Leaf
MW 2251_8	485.90	1.92	Leaf
MW 2251_5	736.54	2.05	Leaf
MW 2251_4	744.72	2.03	Leaf
MW 2251_11	390.49	1.97	Leaf
MW 2251_10	390.82	1.92	Leaf
MW 2245_4	495.62	1.88	Leaf
MW 2245_3	674.91	1.92	Leaf
MW 2245_2	758.81	2.03	Leaf
MW 2245_1	661.91	2.01	Leaf
MW 2244_9	942.84	1.94	Leaf
MW 2244_8	623.10	1.95	Leaf
MW 2244_7	693.02	1.93	Leaf
MW 2244_6	514.57	1.89	Leaf
MW 2244_5	684.56	1.96	Leaf
MW 2243_1	516.88	1.89	Leaf
MW 2238_4	1012.78	2.03	Leaf
MW 2238_3	989.48	2.03	Leaf
MW 2238_2	1061.57	2.04	Leaf
MW 2238_1	603.26	2.02	Leaf
MW 2097_9	511.33	1.96	Leaf
MW 2097_7	468.92	1.94	Leaf
MW 2097_6	976.79	2.07	Leaf
MW 2097_3	904.38	2.03	Leaf
MW 2097_10	760.93	1.95	Leaf
MW 2047_9	712.87	1.97	Leaf
MW 2047_8	812.62	1.97	Leaf
MW 2047_7	727.84	1.95	Leaf
MW 2047_6	659.54	1.94	Leaf
MW 2047_4	976.21	2.04	Leaf
Mtawanjuni_6	200.82	1.89	Leaf
Mtawanjuni_5	286.08	1.87	Leaf
Mtawanjuni_4	161.97	1.84	Leaf
Mtawanjuni_3	1182.40	1.96	Leaf
Mtawanjuni_1	556.69	1.91	Leaf

Sample Name	DNA Concentration in ng/µl	$A_{260/280}$	Sample used for Extraction
MW 2311_7	574.18	1.88	Leaf
MW 2311_5	480.17	1.89	Leaf
MW 2311_4	1128.16	2.04	Leaf
MW 2311_3	940.07	2.03	Leaf
MW 2309_8	338.87	1.90	Leaf
MW 2309_5	297.87	1.87	Leaf
MW 2309_3	349.18	1.91	Leaf
MW 2309_2	451.58	1.94	Leaf
MW 2309_1	562.28	1.93	Leaf
MW 2308_5	427.60	1.96	Leaf
MW 2308_4	437.01	1.94	Leaf
MW 2308_3	423.28	1.96	Leaf
MW 2308_2	544.39	1.95	Leaf
MW 2308_1	651.04	2.00	Leaf
MW 2306_8	597.04	1.92	Leaf
MW 2306_7	467.70	1.86	Leaf
MW 2306_6	398.63	1.90	Leaf
MW 2306_5	355.47	1.92	Leaf
MW 2306_1	723.77	2.05	Leaf
MW 2305_6	150.49	1.91	Leaf
MW 2305_4	179.13	1.91	Leaf
MW 2305_3	1124.70	2.04	Leaf
MW 2305_2	1172.54	2.04	Leaf
MW 2303_7	490.28	1.86	Leaf
MW 2303_4	319.59	1.91	Leaf
MW 2303_3	405.05	1.83	Leaf
MW 2303_2	589.95	1.94	Leaf
MW 2303_1	502.43	1.94	Leaf
MW 2302_9	493.63	1.89	Leaf
MW 2302_8	429.30	1.91	Leaf
MW 2302_3	1101.58	2.03	Leaf
MW 2302_2	825.21	2.04	Leaf
MW 2302_10	629.80	1.93	Leaf
MW 2300_7	380.03	1.89	Leaf
MW 2300_6	450.38	1.87	Leaf
MW 2300_5	361.91	1.91	Leaf
MW 2300_4	300.60	1.92	Leaf
MW 2300_3	499.73	1.85	Leaf
MW 2299_5	377.39	1.88	Leaf
MW 2299_4	313.41	1.87	Leaf
MW 2299_3	606.21	2.02	Leaf
MW 2299_2	448.47	1.90	Leaf
MW 2299_1	826.13	1.99	Leaf
MW 2298_2	590.12	2.02	Leaf
MW 2298_1	801.19	2.07	Leaf
MW 2243_4	**248.74**	**1.75**	**Seed[1]**
MW 2243_3	**253.98**	**1.78**	**Seed[1]**

Sample Name	DNA Concentration in ng/µl	$A_{260/280}$	Sample used for Extraction
KAT 60/8_5	633.85	2.00	Leaf
KAT 60/8_4	584.26	1.96	Leaf
KAT 60/8_3	916.76	2.01	Leaf
KAT 60/8_2	474.54	1.95	Leaf
KAT 60/8_1	805.12	2.01	Leaf
ICPV 9145_9	361.88	1.90	Leaf
ICPV 9145_7	461.74	1.91	Leaf
ICPV 9145_6	533.36	1.88	Leaf
ICPV 9145_2	325.76	1.92	Leaf
ICPV 9145_1	346.48	1.96	Leaf
ICPV 87105_6	243.85	1.92	Leaf
ICPV 87105_4	426.08	1.93	Leaf
ICPV 87105_3	502.37	1.98	Leaf
ICPV 87105_2	319.16	1.95	Leaf
ICPV 87105_1	390.61	1.95	Leaf
ICP_2309_5	1079.28	1.99	Leaf
ICP_2309_4	1023.95	2.00	Leaf
ICP_2309_3	485.74	1.98	Leaf
ICP_2309_2	232.01	1.95	Leaf
ICP_2309_1	231.88	1.99	Leaf
ICP 13076_5	944.86	1.98	Leaf
ICP 13076_4	1079.71	2.00	Leaf
ICP 13076_3	1284.23	1.99	Leaf
ICP 13076_2	855.86	2.00	Leaf
ICP 13076_1	953.00	1.98	Leaf
ICEAP 00557_7	814.90	1.99	Leaf
ICEAP 00557_6	648.98	2.06	Leaf
ICEAP 00557_5	564.39	2.04	Leaf
ICEAP 00557_4	799.24	2.03	Leaf
ICEAP 00557_3	519.31	1.99	Leaf
ICEAP 00068_5	997.13	2.02	Leaf
ICEAP 00068_4	1140.72	2.05	Leaf
ICEAP 00068_3	290.50	2.00	Leaf
ICEAP 00068_2	683.36	1.97	Leaf
ICEAP 00068_1	1096.63	2.03	Leaf
ICEAP 00040_9	420.48	1.91	Leaf
ICEAP 00040_8	445.40	1.91	Leaf
ICEAP 00040_7	723.39	1.94	Leaf
ICEAP 00040_5	592.25	2.04	Leaf
ICEAP 00040_4	536.37	1.72	Leaf
ICEAP 00020_5	649.69	2.05	Leaf
ICEAP 00020_4	1070.64	2.08	Leaf
ICEAP 00020_3	1265.55	2.05	Leaf
ICEAP 00020_2	966.82	2.07	Leaf
ICEAP 00020_1	1099.66	2.04	Leaf
MW 2355_7	**387.88**	**1.87**	**Seed[1]**

[1] These samples were removed from analysis due to low amplification in all the markers

Appendix D. SSR markers that are polymorphic for cultivated pigeonpea

Marker No	SSR name	Motif	Primer Sequences	Reference
1	**CCB1**	(CA)10	F: AAGGGTTGTATCTCCGCGTG R: GCAAAGCAGCAATCATTTCG	Burns et al., (2001)
2	**CCB7**	(CT)16	F: CAACATTTGGACTAAAAACTG R: AGGTATCCAATATCCAACTTG	Burns et al., (2001)
3	**CCB8**	(CT)30	F: TGCGTTTGTAAGCATTCTTCA R: ACTTGAGGCTGAATGGATTTG	Burns et al., (2001)
4	**CCB9**	(CT)22	F: CACTTGGTTGGCTCAAGAAC R: GCCAATGAACTCACATCCTTC	Burns et al., (2001)
5	**CCB10**	(CA)15	F: CCTTCTTAAGGTGAAATGCAAGC R: CATAACAATAAAAGACCTTGAATGC	Burns et al., (2001)
6	**CCttc008**	(AC)7	F: TCACAGAGGACCACACGAAG R: TGGACTAGACATTGCGTGAAG	Saxena et al., (2010d)/ Odeny et al., (2009)
7	**CCac035**	(AC)7	F: TGAGAGGCAATGATGTTGGA R: TCTACAGGCACCCTTTGAAAAT	Saxena et al., (2010d)/ Odeny et al., (2009)
8	**CCac036**	(CATA)3ta(TG)6	F: ATCGGCTTTTGTCTTGATGA R: AAGCTACAAGGGATACACATGC	Saxena et al., (2010d)/ Odeny et al., (2009)
9	**CCttc031 (CCac019)**	(TG)6	F: CAAGGAATCACTTAAAAACCAAGC R: AGATGGCCAAGATTCCACAAC	Saxena et al., (2010d)/ Odeny et al., (2009)
10	**CCttc033 (CCttc020)**	(CTT)8	F: ATTCCCTCTCTATCTCAGACTTTT R: TCGTGATGGAACTCAAGATACACT	Saxena et al., (2010d)/ Odeny et al., (2009)
11	**CCac021**	(AC)6aag(CTAA)3	F: CACGATTCCATTGGTGGAG R: ACGGTTTCTGGGAGGGTCTA	Saxena et al., (2010d)/ Odeny et al., (2009)
12	**CCac009 (CCac007)**	(TG)(TC)2(TG)7	F: GGGAAACTCACCTATATTACCAA R: CACTACCGTCTACAGCCATCTC	Saxena et al., (2010d)/ Odeny et al., (2009)
13	**CCtc007**	(TC)8	F: CATTTATTTCTCTCTGGCATTCAC R: CGAGCTGCAAGCATAAACG	Saxena et al., (2010d)/ Odeny et al., (2009)
14	**CCttc005 (CCttc004)**	(GAA)6	F: ATCGCTTTGCATCCTTATC R: CTTCACGTACATTTTCGTTT	Saxena et al., (2010d)/ Odeny et al., (2009)
15	**CCttc003 (CCttc002)**	(GAA)5g(GAA)5	F: ACACCACCATGCTAAAGAACAAG R: CCAAGCAAGACACGAGTAATCATA	Saxena et al., (2010d)/ Odeny et al., (2009)
16	**CCttat001**	(TTAT)4	F: TACAGCAGCCACATCAAAGC R: TGAACCGTGAAAGTGGGATT	Odeny et al., (2007)

Marker No	SSR name	Motif	Primer Sequences	Reference
17	**CCtta007**	(ATT)4	F: ACCCATTATTGATTTGGGTA	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: CCAAATTTCACCCAAGAAA	
18	**CCggt001**	(GGT)4	F: ACGCTTCTGATGCTGTGTTG	Odeny et al., (2007)
			R: CATCAGCATCATCGTTACCC	
19	**CCtc002**	(GA)12	F: GACTCTTCACCTCACACTCATCAC	Odeny et al., (2007)
			R: ACCTCATACAACAACCCTAAGCAC	
20	**PKS30**		F: AAGTGTGACACCCTCTACCC	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: TGACATCGGGACATAGATAGAA	
21	**CCac003**	(CA)8	F: TGCTTCAAGTTGCCTACCAG	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: TCAAGGGAGGTGGACTACAAA	
22	**CCttc006 (CCttc005)**	(GAA)11gag(GAAa)5 gaggaagag(GAA)17	F: GTAGAGGAGGTTCCAAATGACATA	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: ATCTGTCTGGTGTTTTAGTGTGCT	
23	**CCttc008**	(AGA)5	F: TCACAGAGGACCACACGAAG	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: TGGACTAGACATTGCGTGAAG	
24	**CCtta011 (CCtta006)**	(ATT)21	F: TCAGGGGTAAATGCGGTATC	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: GAATTGCTTTTTGCTTCCTCA	
25	**CCcttc001**	(CTTC)4	F: TAAGGAAATGGCTGGGGTTG	Odeny et al. (2007)
			R: CACATAAATTTGGGGGTTCG	
26	**CCac006**	(CA)10cg(CA)6	F: ACATGTGTGGCGTAGTGTGA	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: GCAAAACCGTTCCATAAAAA	
27	**CCgtt002**	(TGT)4	F: TGGGCTGTGATCGATGAAT	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: CGACAACAACAACACCGACT	
28	**CCttc012**	(TTC)7	F: TAGAGCGTTGTCCCTTTTCTG	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: TCGAAGGACAACTCAAGCATT	
29	**CCgtt003**	(TTG)5(TTC)7	F: GTTCTTCTTGTTGTTGTTGTTG	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: AATTCGTGGAGTTCATTGG	
30	**CCtc013 (CCtc007)**	(TC)6	F: CTTCTCCCTGCCTCTTTTCC	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: CAAGTGGAGGGGAGTGAAGA	
31	**CCac012 (CCac010)**	(CA)7	F: ACCTTGCTTGTTTCGCTTTT	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: AAGGGAGGTGGACTACAAGGA	
32	**CCac013 (CCac011)**	(GT)7	F: GTGAGTGAGAGTGAGTGTATTTGTG	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: GCTCTGATGCCAAATGTTGA	

Marker No	SSR name	Motif	Primer Sequences	Reference
33	**Ccat011 (Ccat006)**	(TA)7(CA)6	F: TGCTCTAATGGCTAGTTCATCC	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: AAACACTCATGGGTTAGATTCTCC	
34	**Ccta003**	(GAT)4	F: TAGTATGGGCGTGGTAGAGGA	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: CGTGACAGAGTCAATCAGAAGC	
35	**CCtc009**	(TC)6	F: ACAAATCCGGTGACCCATAA	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: CCGAGAACAAAAACATTGAACA	
36	**CCac018**	(AC)6a	F: TCTTTCAGACGCAATGACCTT	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: CACTTATTTGTGGGGACCATC	
37	**CCac026**	(AC)7	F: TGAGAGGCAATGATGTTGGA	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: TCTACAGGCACCCTTTGAAAAT	
38	**CCac036 (CCac030)**	(TGT)(TTG)2(TG)7	F: TGATTTGTGCTTGTGCCTTG	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: GTCTTGCTTACGCGTGGACT	
39	**CCttc018**	(AGA)5	F: ACAATTACTCAAATGCTCTCAACG	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: TAAATGTCGCTTCCTATGATAGACC	
40	**CCttc019**	(AAG)13	F: TGAAATGAACAAACCTCAATGG	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: TGTATTGCACATTGACTTGGCTA	
41	**CCac029**	(CAA)(CA)6caa	F: CGTGGACTAATCATCCCGTAA	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: ATAATGCCAAAGGGGGAGAA	
42	**CCB4**	(CA)31	F: GGAGCTATGTTGGAGGATGA	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: CCTTTTTGCATGGGTTGTAT	
43	**CCcct004**	(CTC)4	F: ATCCTCCAAAAGTTCCACCA	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: CAAAGGAGGATTTCCACCAA	
44	**CCtc020**	(TC)13	F: CTAGGCCCTCGAGCTACATT	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: TCTTTTAGAGGTGCGCTGTG	
45	**CCtta015**	(AAT)4	F: AACACGCACCTCAATTCCA	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: GAATGAGGAATGAAGGGACAAA	
46	**CCttc007**	(GA)4ca(GA)4cagagt(GA)8	F: CTCTTGCTTACGCGTGGACT	Saxena et al., (2010d)/ Odeny et al., (2009)
			R: CTTTTGCTTTTGCGTGCTT	
47	**ICPM1E04**	(A)10	F: TTTTTATGGAATATTTATGAGTTAAC	Saxena et al (2010b)
			R: AAGAGTTTCCCAACCCTGCT	
48	**ICPM1G04**	(T)21	F: GCTCCAATTTTTCATTTCGG	Saxena et al (2010b)
			R: ATCAAACAATGCACCCATGA	