Microsatellite Development and Application in Pigeonpea (*Cajanus cajan* (L.) Millsp.)

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Willhelms-Universität Bonn

Vorgelegt von

Damaris Achieng Odeny

aus

Kisumu, Kenya

Bonn 2006

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Referentin: Prof. Dr. Dorothea Bartels

2. Koreferentin: Priv. Doz. Dr. Christiane Gebhardt

Tag der Promotion: 23/02/2006

To the memory of my beloved father, WILFRED ODENY KAYI,

my role model for hard work and perseverance

ERKLÄRUNG

Ich versichere, daß ich die von mir vorgelegte Dissertation selbstandig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit einschließlich Tabellen, Karten und Abbildungen, die anderen Werken in Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; daß diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat. Die von mir vorgelegte Dissertation ist von Prof. Dr. Dorothea Bartels betreut worden.

Bonn, February 2006 Damaris Achieng Odeny

ABSTRACT

Pigeonpea is a major legume crop grown in the semi-arid tropics but has been relatively neglected in terms of genomic research. This study aimed at developing molecular markers as a basic requirement towards initiating marker assisted breeding techniques in pigeonpea. Simple Sequence Repeat (SSR) loci of pigeonpea were isolated by screening non-enriched (library A) and enriched (library B) partial genomic libraries with SSR probes. Positive clones were sequenced and primers designed for 152 microsatellite loci, 39 from library A and 113 from library B. Optimisation of reaction conditions was achieved for 51% and 65% of primers designed from library A and B, respectively. For the purpose of exploiting the transferability of SSRs across genera within the legume species, 220 soybean primers were tested in pigeonpea, 39 of which amplified interpretable bands.

Nineteen out of 20 amplified primers from library A were polymorphic among 15 cultivated and 9 wild species. The diversity analysis revealed contrasted levels of variability within cultivated and wild accessions. A total of 98 alleles were detected at the 19 polymorphic loci with an average of 4.9 alleles per locus while the observed heterozygosity ranged from 0.17 – 0.80 with a mean of 0.42 per locus. Substantially less allelic variation (31 alleles) was observed within the cultivated species than across the wild species (92 alleles). Primers from library B were not tested for amplification in wild species but 35 out of the amplified 73 revealed polymorphism among 24 pigeonpea genotypes. The number of alleles detected ranged from 2 to 6 with a total of 110 alleles and an average of 3.1 alleles per locus. Only one of the 39 amplified soybean primers revealed polymorphism among 24 cultivated pigeonpea accessions. No significant relationship was detected between the class of repeats and heterozygosity values.

AT and TG class of repeats were the most abundant di-nucleotide repeats in library A and B respectively while TAA and GAA were the most abundant tri-nucleotide repeats in both libraries. Protein database searches provided putative functions for 21 SSR-containing pigeonpea sequences that would be useful in functional marker development. UPGMA and MDS cluster analysis revealed genetic relationships among recently bred varieties, old varieties and wild accessions. Nine of the markers developed were polymorphic to the parental lines of a F_6 *Fusarium* wilt RIL mapping population that had been developed by ICRISAT breeders. Analysis of allele segregation in the RIL population revealed that all the 9 SSRs segregated in the expected 1:1 ratio and were further tested for any possible linkage with a QTL for resistance to *Fusarium* wilt. All the polymorphic markers derived from this study are now being used for characterisation and evaluation of pigeonpea germplasm collection at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) headquarters, India. SSRs provide a powerful tool for genomic studies and are recommended for systematic fingerprinting of pigeonpea germplasm.

TABLE OF CONTENTS

Acronyms and Abbreviations

AFLP	Amplified Fragment Length Polymorphism
AGL	Applied Genomics Laboratory
APS	Ammonium per sulphate
BLUP	Best Linear Unbiased Prediction
BLAST	Basic Local Alignment Search Tool
CDS	Protein coding sequence
CMS	Cytoplasmic Male Sterility
CRBD	Completely Randomised Block Design
CTAB	Cetyltrimethyl-ammoniumbromide
DArT	Diversity Arrays Technology
dNTP	Deoxyribonucleotide triphosphate
EDTA	Ethylene-diamine-tetra-acetic acid
EST	Expressed Sequence Tag
FAOSTAT	Food and Agriculture Organisation Statistics
ICRISAT	International Crops Research Institute for the Semi-Arid tropics
IPTG	Isopropyl-beta-D-thiogalactopyranoside
LB	Luria Broth
LOD	Logarithm of Odds
MAS	Marker Assisted Selection
MCS	Multiple Cloning Site
MDS	Multidimensional Scaling
MPIZ	Max-Planck-Institut für Züchtungsforschung
NCBI	The National Centre for Biotechnology Information
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PIC	Polymorphism Information Content
PSI	Position Specific Iterated
QTL	Quantitative Trait Loci
RAPD	Random Amplified Polymorphic DNA
RIL	Recombinant Inbred Lines
RSA	Resistance Gene Analog
SAP	Shrimp Alkaline Phosphatase
SAT	Semi-Arid Tropics
SDS	Sodium dodecyl sulphate
SHAN	Sequential Hierarchical Agglomerative Non-overlapping
SMD	Sterility Mosaic Disease
SSC	Saline Sodium Citrate
SSR	Simple Sequence Repeats
TAE	Tris-acetate-EDTA
TBE	Tris-borate-EDTA
TEMED	Tetramethyl-ethylenediamine
TIGR	The Institute for Genomic Research
UPGMA	Unweighted Pair Group Method of Arithmetic Means
USA	United States of America
USDA	United States Department of Agriculture
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside

1 INTRODUCTION

1.1 Economic importance of pigeonpea

The fact that the world faces a water crisis has become increasingly clear in the last decade. Current predictions estimate that by the year 2050, at least 1 in every 4 people is likely to live in a water deficient area. An important challenge facing scientists is increasing food production with less water. Several reviews on procedures for improving water efficiency use have recently been published (Zwart and Bastiaanssen 2004). Several successful approaches to achieve high yielding drought tolerant crops through biotechnology have been reviewed (Van Camp 2005). Crops that were once considered "orphan" are now being incorporated into major breeding programs, as they seem to hold the key to the future. The importance of a drought-tolerant legume such as pigeonpea (*Cajanus cajan* (L.) Millsp.), which combines several desirable traits, cannot therefore be ignored.

Pigeonpea (*Cajanus cajan* (L.) Millsp.) is a grain legume belonging to the *Cajaninae* sub-tribe of the economically important leguminous tribe *Phaseoleae*. The tribe *Phaseoleae* also contains soybean (*Glycine max* L.), common bean (*Phaseolus vulgaris* L.) and mungbean (*Vigna radiata* L. Wilczek) (Young et al. 2003). The genus *Cajanus* comprises 32 species most of which are found in India and Australia although one is native to West Africa. Pigeonpea is the only cultivated food crop of the *Cajaninae* sub-tribe and has a diploid genome comprising 11 pairs of chromosomes (2n = 22) with a physical size estimated at about 0.853 pg (Greilhuber and Obermayer 1988).

India is the world's largest pigeonpea producer (Table 1.1) and grows over 77% of the total world production. Pigeonpea is now reported to be grown in 50 countries of Asia, Africa and the Carribbean, where its name "pigeon-pea" is thought to have originated. The current global annual production of pigeonpea is valued at more than US$ 1700 m (FAOSTAT 2005). The crop can be described as unique because it is a legume and a woody shrub. It has an inherent ability to withstand environmental stresses (especially drought) making it one of the most sought after crops in plant introduction trials aimed at bringing new areas under cultivation (Okiror 1986). It contributes to the C, N and P economy of the soil (Fujita et al. 2004; Kumar Rao et al.

1987; Rego and Nageswara Rao 2000) enhancing its performance even under marginal input. Pigeonpea is tolerant to low P supply and acid soils as well as having a high capacity for incorporation of external P into organic P (Fujita et al. 2004). Its critical requirement of P concentration for dry matter production is low compared to other major protein crops like soybean [*Glycine max* (L.)] (Adu-Gyamfi et al. 1990).

Table 1.1 World pigeonpea production (2005)

Country	**Area Harvested (Ha)**
Puerto Rico	165
Bahamas	180
Comoros	440
Grenada	520
Trinidad and Tobago	1,100
Jamaica	1,100
Burundi	2,000
Venezuela	2,500
Bangladesh	3,237
Panama	4,800
Haiti	6,000
Democratic Republic of Congo	8,000
Dominican Republic	13,000
Nepal	29,000
Tanzania	68,000
Uganda	84,000
Malawi	123,000
Kenya	200,000
Myanmar	540,000
India	3,500,000

Modified from FAOSTAT, 2006

Its deep root system allows extraction of moisture from deep layers of the soil and thus makes it a crop that produces biomass including protein-rich grain while utilizing residual moisture (Nene and Sheila 1990). It can be intercropped with cereals such as maize (*Zea mays* L.) and sorghum (*Sorghum bicolor* (L.) Moench) without a negative impact on the main crop. It is an important component in the integrated crop and livestock systems of the semi-arid tropics as it can be used as forage or hay. Pigeonpea adapts to different climates and soils except those that are excessively wet or experience frost (Troedson et al. 1990)

Pigeonpea plays an important role in food security, balanced diet and alleviation of poverty because it can be used in diverse ways as a source of food, feed, fodder (Rao et al. 2002), fuel wood, rearing *lac* insects (Zhenghong et al. 2001), hedges, windbreaks, soil conservation, green manuring and roofing. It is a major source of protein to about 20% of the world population (Thu et al. 2003) and is an abundant source of minerals and vitamins (Saxena et al. 2002). Its abundance in protein makes it an ideal supplement to traditional cereal-, banana- or tuber-based diets of resource poor farmers that are generally protein-deficient. The perennial nature of pigeonpea allows farmers to take multiple harvests with surpluses traded in both local and international markets.

1.2 Constraints to productivity

Despite its importance in the semi-arid tropics (SAT), little concerted research effort has been directed at either crop improvement or technology transfer. The production of pigeonpea has remained static over the last several years (Souframanien et al. 2003). The yield on farmer's fields is low and a number of factors are responsible. Farmers continue to grow their traditional landraces, which frequently suffer from several biotic and abiotic stresses due to lack of quality seed, with the result that productivity can be erratic across years. Poor production practices such as low plant densities, low soil fertility, insufficient weeding and insufficient/inappropriate use of fungicides and herbicides are other constraints. Environmental (frequent droughts, easily erodible soils with poor water holding capacity) and socio-economic (lack of roads, marketing infrastructure, and exploitation by middlemen) factors also affect productivity.

Important insect pests include the pod boring lepidoptera (*Helicoverpa armigera* Hübner*, Maruca vitrata* Geyer and *Etiella zinkenella* Treitsche), pod sucking bugs (*Clavigralla tomentosicollis* Ståll and *Clavigralla horrida* Germar) and podfly (*Melanagromyza chalcosoma* Spencer) (Minja et al. 2000). Though pigeonpea diseases have been reported to be of minor importance in the past, recent surveys indicate that *Fusarium* wilt (*Fusarium udum* Butler), sterility mosaic disease (SMD), leaf spot (*Mycovellosiella cajani*) and to a lesser extent powdery mildew (*Leveillula taurica*) are diseases of economic concern. *Fusarium* wilt is especially prevalent in India and East Africa, where field losses of over 50% are common (Marley and Hillocks 1996).

The crop's long life cycle and a heterozygous genome structure conserved by out-crossing (up to 70%) (Saxena et al. 1990) make breeding slow and expensive. Historically, desirable traits in pigeonpea have been selected for by farmers from landraces to suit their production systems and uses. The establishment of the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) in 1972 created a new focus and research interest leading to the recent development of cytoplasmic male sterile (CMS) lines (Saxena and Kumar 2003; Mallikarjuna and Saxena 2005) for commercial hybrid breeding of pigeonpea. However, specific cultivar improvement has been difficult due to the limited knowledge on the inheritance of important traits and lack of understanding on the level of inter- and intra-specific genetic diversity.

Wild relatives have now been reported to possess many agronomically important traits such as resistance to pests and diseases (Reddy et al. 1996; Sharma et al. 2003), salinity tolerance (Subbarao et al. 1991) and high protein content (Saxena et al. 1996), all of which would be useful in cultivated pigeonpea. As different needs and opportunities surface, pigeonpea breeders need to incorporate new genetic sources using various breeding methods aided with modern tools such as biotechnology. An approach with more perspective is marker assisted selection (MAS) (Ribaut and Hoisington 1998), which has emerged in recent years due to developments in molecular marker technology, especially those based on the polymerase chain reaction (PCR) (Powell et al. 1996; Bussell et al. 2005).

Molecular markers are DNA sequence variants that can readily be detected and whose inheritance can be monitored (Newbury and Ford-Lloyd 1999). Molecular marker technology can facilitate the precise determination of the number, chromosomal location and individual and interactive effects of genes that control traits (Peleman and van der Voort 2003). However, use of MAS requires detailed information on the plant genome. A basic pre-requisite for any molecular breeding program is a robust set of polymorphic markers for the species under investigation. Among the different marker systems available are Simple Sequence Repeats (SSRs) (Tautz and Rentz 1984).

1.3 Simple Sequence Repeat (SSR) marker development

Simple Sequence Repeat (SSR) markers, also known as microsatellites, are tandemly repeated motifs of 1-6 nucleotides found in all prokaryotic and eukaryotic genomes (Zane et al. 2002). According to Pupko and Graur (1999), any number of tandem repeats of a certain nucleotide combination may be regarded as a microsatellite (Fig. 1.1). These repeats are present in both coding and non-coding regions (Hancock 1995) and are usually characterized by a high degree of length polymorphism (Zane et al. 2002). Microsatellite loci are inherently unstable with high mutation rates, a phenomenon that is reported to be caused by DNA polymerase slippage and/or unequal recombination (Li et al. 2002). Due to their high mutability, SSRs play a significant role as molecular markers for evolutionary and population genetic studies.

TATTTATGGGAAACAAAATATCCCCTAGTCATGCGTATTGAATGAATTG
AACACACACACACACACACACACACACACACACACACACACACACACACA
CACACACACACACACACACACACACACACACACACACACACACACACACA

Figure 1.1 Pigeonpea sequence containing an AC-repeat (highlighted in red)

Microsatellites offer several advantages compared to other molecular markers: they are highly reproducible, highly polymorphic, PCR-based and readily portable within a species (Edwards et al. 1996). In a recent study comparing SSRs, RAPDs and AFLPs for the genetic analysis of yeast *(Saccharomyces cerevisiae)* strains, Gallego et al. (2005) reported that SSR analysis gave the highest level of information content. Similar results were reported earlier in soybean (Powell et al. 1996). Microsatellites have also attracted scientific attention because they have been shown to be part of or linked to some genes of agronomic interest (Yu et al. 2000). All these positive attributes coupled with their multi-allelic nature, co-dominant transmission, relative abundance, extensive genome coverage and requirement of only a small amount of template DNA have contributed to the extraordinary increase of interest in SSRs in many organisms (Zane et al. 2002).

In pigeonpea, however, only 20 SSRs have been developed so far of which only 10 are polymorphic in cultivated pigeonpea germplasm (Burns et al. 2001). In contrast, more than 1000 SSR loci have been mapped in soybean (Song et al. 2004), about 400 in chickpea (*Cicer arietinum* L.) (Lichtenzveig et al. 2005), over 100 in

common bean (*Phaseolus vulgaris* L.) (Blair et al. 2003) and groundnut (*Arachis hypogaea* L.) (Ferguson et al. 2004). Despite the reported high informative nature of SSRs, the high cost and time required for their development is a major limitation. This is especially the case in crops such as pigeonpea, for which no sequences exist in databases that could be directly searched for SSRs. In such species, microsatellites can only be isolated *de novo*.

The traditional and most simple procedure of microsatellite isolation involves the cloning of small genomic DNA fragments and the screening of clones through by colony hybridisation with repeat containing probes (Powell et al. 1996; Chen et al. 1997). This procedure works well for species that are abundant in SSRs but not in those that are SSR poor. To increase the chances of success, the use of enriched libraries was proposed and those based on selective hybridization (Karagyozov et al. 1993; Billotte et al. 1999; Edwards et al. 1996) have been the most successful.

The basic protocol involves DNA fragmentation followed by ligation of the fragments to a known sequence – a vector or an adaptor. The DNA is then hybridized with a repeat containing probe, which could be bound to a nylon membrane (Stajner et al. 2005) or 5'-biotinylated and bound to streptavidin-coated beads (Yaish and de la Vega 2003). Non-specific binding is reduced by several washes, after which the DNA is eluted and recovered by PCR amplification. The enriched DNA is finally cloned into a suitable vector. The recombinants could be directly sequenced (if efficiency of the procedure is high) or further screened for the presence of repeats using southern blotting or PCR strategies (Zane et al. 2002). The sequenced clones are searched for microsatellite motifs (Temnykh et al. 2001) and then primers are designed from the unique DNA that flanks microsatellite motifs (Glenn and Schable 2005). Subsequently, the primers are tested for amplification using DNA of the respective species.

Due to the reported even distribution of microsatellite markers across genomes (Li et al. 2002), SSRs developed using genomic DNA could be either from the coding or non-coding regions. Two types of microsatellites have been described; type I (genic SSR) and type II. Type I markers are associated with genes of known functions and are more useful for comparative gene mapping to study genome evolution (Vignal et al. 2002) while type II markers are of no known function. Type I markers are relatively rare

and generally less polymorphic than type II markers. Detection of markers located within genes and ESTs provides a possibility to convert type II markers into type I.

1.4 Microsatellites from coding regions of the genome

With the establishment of expressed sequence tag (EST) sequencing projects for gene discovery programs in several plant species, a wealth of DNA sequence information has been generated and deposited in online databases. The most common procedure for identification of type I SSRs uses computer programs to download sequence data for ESTs, genes and cDNA clones from genebank followed by scanning for identification of SSRs. Similarly, SSR-containing genomic sequences can be used to search for syntenic regions amongst well-annotated databases of closely related species for identification of putative genic SSRs.

For pigeonpea, the most useful databases would be those of *Medicago* and soybean, as well as that of *Arabidopsis thaliana* (a dicot model plant with a sequenced genome). The first step would involve trawling a sequence database with tools such as FASTA (Pearson 1998) or BLAST (McGinnis and Madden 2004), the latter being the most commonly used. Results from these searches would quickly reveal similarities between the query (in this case pigeonpea genomic sequence) and a range of database sequences. Ideally a search output should show exact similarity to a well-characterised protein over the full length of the query (Attwood 2000). However, this is rarely the case, especially with high possibilities of raw sequences having errors and repetitive regions. Furthermore, a high sequence similarity may only happen by chance and may not necessarily mean identical function.

The greatest challenge, therefore, is on how to come up with a reliable inference homology to be used in verifying a relationship. Identification of significant sequence alignment is usually carried out using a cut-off BLAST probability score – the expect ("e") value. The lower the "e" value, the stronger the similarity. This can be combined with a different criteria based on length alignment and percent identity (Salse et al. 2002) to strengthen results. Some authors (for example Bennetzen et al. 2004) have suggested that a combination of tools could yield a more reliable final product. A satisfactory homology between an SSR containing sequence and a defined protein from the database would give an indication of a potential type I SSR. However, such a study

would be incomplete until the possible linkage of the identified SSR to the putative gene is verified through functional genomic studies.

Both type I and II SSRs have increasingly come into prominence over the last decade because scientists have found them to be remarkably versatile molecular tools. According to Chambers and MacAvoy (2000), the key factor leading to their widespread adoption lies in the power that they provide to solve biological problems. In pigeonpea, microsatellites could be applied in a range of studies starting from identification of individuals to tracking the evolutionary history of populations.

1.5 Potential application of SSRs in pigeonpea

1.5.1 Geographical origin

There has been a major dispute on the possible origin of pigeonpea. Several conclusions have been made in favour of India given the presence of several wild relatives, the large diversity of the crop gene pool, ample linguistic evidence, a few archaeological remains and the wide usage in daily cuisine (Van der Maesen 1983). However, some authorities (Purseglove 1968; Rachie and Roberts 1974) considered Africa to be the centre of origin due to the presence of pigeonpea seeds in Egyptian tombs and a wild species (*Cajanus kerstingii*) in West Africa. The only certain way to resolve such disputes is through the study of the genus *Cajanus* at the DNA level as has been recently done in cassava (*Manihot esculenta* ssp. *esculenta*) (Olsen 2004) and apricot (*Prunus armeniaca* L.) (Zhebentyayeva et al. 2003). According to Heslop-Harrison (2000), synthetic oligonucleotide SSRs have been able to reveal that microsatellite sequences vary widely with regard to genomic organisation making them perfect for this kind of study.

1.5.2 Genotype identification and genetic diversity

The International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) currently has a collection of more than 13,000 pigeonpea germplasm accessions in the genebank. This germplasm has been morphologically characterised (Remanandan et al. 1988) and found to contain variation among accessions. Morphological studies alone do not provide sufficient information to understand genetic diversity within the species as well as its relatedness to other species. Molecular analysis using SSRs can provide

additional information on genetic diversity that would be useful for breeding programs through selection of diverse parents (Charcosset and Moreau 2004).

The current interest in the genetic potential of wild relatives (Sharma et al. 2003) could be further enhanced through the use of SSR markers in identification of the most closely related parents for inter-specific crossing. The ongoing breeding emphasis on development of hybrid pigeonpea will also require a quick and efficient way of predicting and identifying inbred lines that can produce highly heterotic hybrids precisely. Other aspects including seed certification, plant variety rights, and description and protection of germplasm of pigeonpea would also benefit from the availability of adequate SSRs.

1.5.3 Molecular linkage map and synteny

The concept of a linkage map first presented by Sturtevant (1913) in *Drosophila melanogaster,* has become a widespread and essential genetic tool for crop improvement and other biological studies (Svetleva et al. 2003). Mapping in pigeonpea has been hampered by the lack of appropriate and sufficient molecular markers. Microsatellites are the markers of choice for the development of a pigeonpea linkage map due to the genetic complexity of breeder's populations and high levels of heterozygosity in individual genotypes. In recent years, a number of practical examples have demonstrated the power of SSRs in development of genetic maps in legumes such as soybean (Song et al. 2004), common bean (Blair et al. 2003) and peas (*Pisum sativum* L.) (Loridon et al. 2005).

Comparative mapping will be important in transferring knowledge from extensively studied legumes (such as the model legume *Medicago sativa* L.) into the less studied genome of pigeonpea. Higher levels of synteny have been shown between common bean, mungbean, and soybean (Lee et al. 2001) and also between soybean and *Medicago* (Mudge et al. 2005). Such reports are encouraging in view of the fact that pigeonpea has been grouped in the tribe *Phaseoleae*, which also contains soybean, common bean and mungbean (Young et al. 2003).

1.5.4 Trait tagging and Marker-Assisted Selection (MAS)

Being a perennial crop, development of superior lines in pigeonpea using conventional methods has been very slow (see Section 1.2). Most of the important agronomic characters are controlled by several genes (Collard et al. 2005) (quantitative traits). The genetic factors responsible for a part of the observed phenotypic variation for a quantitative trait are called quantitative trait loci (QTL). A QTL (Salvi and Tuberosa 2005) indicates a region on the genome and could be comprised of one or more functional genes (Falconer and Mackay 1996). Traditionally, this problem has been dealt with by having several replicated trials, which allow identification of genotypic differences through statistical analysis. Such a procedure can be laborious and especially slow for a late-maturing crop like pigeonpea.

In other legumes, SSR markers have been used as a tool to identify major genes and QTLs and also to introduce new characters in elite germplasm (Asensio-S.-Manzanera et al. 2005). SSR markers linked to quantitative traits such as seed quality (Hyten et al. 2004) and yield (Wang et al. 2004) in soybean, and resistance to *Ascochyta* blight (Prioul et al. 2004) in common bean have been reported. Availability of adequate SSRs in pigeonpea would enable breeders to know the location of specific genes and QTLs making it possible to improve the efficiency of breeding through MAS.

1.5.5 Association mapping

Even though QTL linkage analysis remains a powerful tool in studying complex traits, one of its major limitations is the poor estimation of the QTL position on a genetic linkage map (Bink and Meuwissen 2004) as it samples only a small fraction of the possible alleles in the population. Moreover, the QTL effects observed in bi-parental crosses are often not representative of those encountered in elite cultivars (Simko et al. 2004). First developed for application in human genetics, association mapping is now becoming popular as a more accurate method for estimating gene positions on a map. This technique incorporates the effect of many past generations of recombination into a single analysis and provides an opportunity to detect allele-trait associations in larger sets of genotypes. Association mapping has only been recently used in plant genetics (Simko et al. 2004) but offers great opportunity for future mapping activities in pigeonpea because no mapping population needs to be created. However, adequate and

appropriate molecular markers must be available for such a study to be undertaken with accuracy (Salvi and Tuberosa 2005).

1.5.6 Map-based cloning

Map-based cloning is the process of identifying the genetic basis of a mutant phenotype by looking for linked markers of which the physical location in the genome is known (Jander et al. 2002). The procedure has been used successfully to isolate, for example, a CMS restorer gene in rice (*Oryza sativa*) (Akagi et al. 2004) and could be similarly done for several traits of importance in pigeonpea. Such a procedure would require a well-saturated linkage map to reduce the laborious efforts in chromosome walking. With the successful development of an efficient gene transfer procedure for pigeonpea (Dayal et al. 2003; Kumar et al. 2004), map-based cloning would enable scientists to incorporate and manipulate different novel traits that would otherwise be impossible using conventional means.

1.6 Objectives and outline of the thesis

Despite the existence of various morphological, physiological and agronomic traits in pigeonpea, no effective molecular breeding program has been developed to facilitate breeding of this crop. The genetics of most of the important traits are not known and mapping strategies are lacking to enhance efficient selection of desirable lines. There is an urgent need to develop a robust set of polymorphic markers and eventually a linkage map. This study was carried out with the overall objective of developing the basic requirements that would enable future molecular breeding in pigeonpea. The specific aims were to;

- Develop a robust set of genomic SSRs and design primers for their amplification.
- Characterise the SSRs by motif and test their ability to identify polymorphism in diverse pigeonpea germplasm.
- Compare genomic sequences generated from the genomic libraries and identify possible putative genic SSRs.
- Assess the transferability of the new markers to wild pigeonpea germplasm.
- Test the transferability of SSRs from soybean to pigeonpea.

- Develop a genetic linkage map for pigeonpea.
- Identify markers associated with *Fusarium* wilt resistance in pigeonpea.

The work described herein was carried out in two different institutions. The development of two genomic DNA libraries, screening and sequencing of clones was done at the Max Planck Institute for Plant Breeding Research (MPIZ), Cologne Germany. The amplification, characterisation and determination of the application of the microsatellites were done at the ICRISAT headquarters, Hyderabad, India.

2 MATERIALS AND METHODS

2.1 Plant material

All seeds used in this study were obtained either from the germplasm unit (for landraces and wild genotypes) or from the pigeonpea breeding section (for all breeding lines) of ICRISAT (Patancheru, Andhra Pradesh, India). Accession ICP 2376 (Table 2.1) was grown under glasshouse conditions at the University of Bonn, Department of Plant Pathology for the preparation of genomic DNA libraries. One-week-old seedlings were kept in the dark for 2 days after which tender leaves were harvested and freeze dried at –60ºC for 48 hours. ICP 2376 (susceptible to *Fusarium* wilt) is one of the parents, together with a resistant line ICPL 87119, of a *Fusarium* wilt mapping population developed at ICRISAT. The former is an improved landrace from India (Andhra Pradesh) while the latter is a breeding line. A detailed description of the parents is given in Table 2.1.

Table 2.1 Characteristics of parents of the mapping population

Trait	**ICPL 87119**	**ICP 2376**
No. of days to flower	124	118
No. of days to harvest	180	180
Plant height (cm)	195	185
Seeds/pod	3.8	3.8
100 seed weight	11.2	9.5
Resistance to wilt	Resistant	Susceptible
Resistance to sterility mosaic	Resistant	Resistant
Flower colour	Yellow	Yellow with dense red streaks
Growth habit	Non-determinate (NDT)	Non-determinate (NDT)
Seed colour	Brown	Cream

The two genotypes were crossed through hand emasculation and pollination to produce F_1 hybrids. Self-fertilizing the F_1 individuals under glasshouse conditions generated two F_2 segregating populations, one having 238 (Set A) and another 215 (Set B) individuals. Each generation of the mapping population was advanced using single seed descent to F_6. Three hundred and fifty F_2 plants and 40 F_5 families, both from Set

A, were used in a preliminary screening (Section 2.17) experiment. A random sampling of 94 and 121 F_6 recombinant inbred lines (RILs) from set A were chosen for genotyping (Section 2.14) and phenotyping (Section 2.19) respectively. The genotyping set was grown in 96-well format in rehydrated Jiffy-7 pellets (Jiffy, International, Norway) under controlled environment in ICRISAT's Applied Genomics Laboratory (AGL), India. Phenotyping was done in the glasshouse.

A total of 24 diverse accessions (Table 2.2) including 9 wild pigeonpea genotypes were selected based on phenotypic analysis (Reddy et al. 2005) for assessment of polymorphism of markers developed from the non-enriched library (Section 2.7.1). Twelve other diverse ICRISAT breeding lines (ICPL 87091, ICP 7035, ICPL 151, Kat 60/8, HPL 24, LD Dwarf, ICPL 99066, MN 5, ICPL 332, ICPA 2068, ICPA 2032, ICP 13092) were used to replace all of the wild genotypes and 3 landraces (ICP 9267, ICP 11181, ICP 14144) while screening for polymorphism in all other markers used in this study. Eight seedlings of each cultivar were grown in 96-well format in rehydrated Jiffy-7 pellets as already described. All the wild-type pigeonpea genotypes were grown in 6-inch pots in the glasshouse at 25°C.

2.2 Cloning vectors and bacterial strain

The vector pBluescript II SK +/- (Stratagene, Heidelberg) was used for the preparation of the non-enriched library (Section 2.7.1). The vector is 2961 bp long and has a multiple cloning site (MCS) flanked by T3 and T7 promoter sequences. pBluescript II carries the β-galactosidase gene for blue/white screening and the β–lactamase gene that confers resistance to ampicillin.

pGEM®-T vector (Promega, Madison, USA) was used for the preparation of the enriched library (Section 2.7.2). This vector is designed for convenient cloning of PCR products as it contains single 3'-T overhangs at the cloning site that prevent circularisation of the vector and provide a compatible overhang for a PCR product. The vector contains a multiple cloning region within the α–peptide coding region of the enzyme β-galactosidase. T7 and SP6 promoters flank the MCS.

The bacterial strain used for replicating the vector was *E. coli* genotype DH-10B (Invitrogen Life Technologies, Freiburg).

Table 2.2 Genotypes used to characterise polymorphism

Accession number	Species	Description/ genepool	Country of Origin
ICPW 1	*C. acutifolia*	Secondary[1]	Australia
ICPW 13	*C. albicans*	Secondary[1]	India (Karnataka)
ICPW 28	*C. cajanifolia*	Secondary[1]	India (Madhya Pradesh)
ICPW 39	*C. latisepala*	Secondary[1]	Australia
ICPW 41	*C. lineata*	Secondary[1]	India (Tamil Nadu)
ICPW 68	*C. platycarpa*	Tertiary[1]	India (Uttar Pradesh)
ICPW 74	*C. reticulata*	Secondary[1]	Australia
ICPW 119	*C. scarabaeoides*	Secondary[1]	Philippines
ICPW 162	*C. sericea*	Secondary[1]	Australia
ICP 9267	*C. cajan*	Landrace[1]	Guyana
ICP 11181	*C. cajan*	Landrace[1]	India (Andhra Pradesh)
ICP 14144	*C. cajan*	Landrace[1]	Jamaica
ICP 13575	*C. cajan*	Landrace[1&2]	Sierra Leone
ICP 15145	*C. cajan*	Landrace[1&2]	Zaire
ICP 9266	*C. cajan*	Landrace[1&2]	Guyana
ICP 4167	*C. cajan*	Landrace[1&2]	India (Uttar Pradesh)
ICP 14576	*C. cajan*	Landrace[1&2]	Thailand
ICP 12058	*C. cajan*	Landrace[1&2]	Tanzania
ICP 14352	*C. cajan*	Landrace[1&2]	Venezuela
ICP 1514	*C. cajan*	Landrace[1&2]	India (Andhra Pradesh)
ICP 7543	*C. cajan*	Landrace[1&2]	India (Madhya Pradesh)
ICP 7852	*C. cajan*	Landrace[1&2]	India (Karnataka)
ICP 13092	*C. cajan*	Breeding line[2]	Kenya (Kirinyaga)
ICP 7035	*C. cajan*	Breeding line[2]	India (Madhya Pradesh)

1 – Used for screening non-enriched library only; 2 – used for screening enriched library only; 1&2 – Used for screening both libraries

2.3 Fungal isolate

A single isolate of *Fusarium udum* from ICRISAT fields (Patancheru) was used throughout this study as a source of inoculum. Repeated single colony sub-culturing created pure cultures of the isolate.

2.4 Standard primers, adaptors and oligonucleotide probes

All standard primers and biotinylated oligonucleotides were obtained from Metabion (Martinsried, Germany) while the adaptors (*RSA* 21 and *RSA* 25) and non-biotinylated oligonucleotides were from Qiagen (Hilden, Germany).

2.4.1 Primers used during the study

Name	Sequence
T7	5’-GTAATACGACTCACTATAGGGC-3’
T3	5’-AATTAACCCTCACTAAAGGG-3’
SP6	5' ATTTAGGTGACACTATAGAATAC 3'
RSA 21	5’-CTCTTGCTTACGCGTGGACTA-3’
RSA 25	5’-pAGTCCACGCGTAAGCAAGAGCACA-3’
CA_{15}	5’-CACACACACACACACACACACACACACA-3’
CT_{15}	5’-CTCTCTCTCTCTCTCTCTCTCTCTCTCT-3’
AAT_{10}	5’-AATAATAATAATAATAATAATAATAAT-3’
GCC_{10}	5’-GCCGCCGCCGCCGCCGCCGCCGCCGCC-3’
CAA_{10}	5’-CAACAACAACAACAACAACAACAACAA-3’
$BioGA_8$	5’-Biotin-GAGAGAGAGAGAGAGA-3’
$BioGT_8$	5’-Biotin-GTGTGTGTGTGTGT-3’
$BioGAA_8$	5’-Biotin-GAAGAAGAAGAAGAAGAAGAA-3’
$BioTAA_8$	5’-Biotin-TAATAATAATAATAATAATAA-3’

2.5 Enzymes and solutions

2.5.1 Enzymes

Type II restriction endonucleases

Name	Restriction site	Supplier
*Eco*R1	5’-G↓AATTC-3’	Roche, Mannheim
*Tsp*5091	5’-↓AATT-3’	Roche, Mannheim
*Sau*3A 1	5’-↓GATC-3’	New England Biolabs

Other enzymes

Name	Supplier
T4 DNA Polymerase	New England Biolabs
T4 DNA Ligase	New England Biolabs
T4 Polynucleotide kinase	New England Biolabs

EXOSAP	Amersham (USB)
Shrimp Alkaline Phosphatase	Roche, Mannheim
RNAse A	Qiagen, Hilden
Taq DNA polymerase	Invitrogen Life Technologies (Freiburg); Bioline (UK)

2.5.2 Media and basic solutions

Media and buffers (unless otherwise stated) were dispensed into the appropriate containers and sterilised by autoclaving for 20 min at 121^{o}C (118 KPa nominal steam pressure). Equipments (except those sensitive to heat) were sterilised by placing in enclosed containers or covering with aluminium foil before autoclaving under the same conditions. All other equipment were sterilised by soaking overnight in 0.5% (v/v) Clorax or 70% ethanol.

Laboratory reagents used in this study were obtained from Biozym (Epicenter, USA), Invitrogen Life Technologies (Freiburg), Sigma (Deisenhofen), Merck (Germany or India), Difco (USA), Roth (Karlsruhe), Serva (Heidelberg), Qualigenes (India), Pharmacia Biotech (Sweden).

The buffers, solutions and media were prepared as follows:

Ammonium acetate 3 M (500 ml)

Dissolve 115.62 g of ammonium acetate in 500 ml water

Ampicillin solution

50 mg ml^{-1} ampicillin dissolved in sterile distilled water

Sterile filtered through 0.22-micron filter

Ammonium per sulphate (APS)	10%
Chloroform/Isoamylalcohol	24:1 (v/v)

Ethylenediamine tetra-acetic acid (EDTA) (20 mM) (1 l)

Dissolve 7.4 g of EDTA in 1 l of water and adjust pH using NaOH solution

Ethidium Bromide 10 mg ml^{-1}

Dissolve 0.2 g ethidium bromide in 20 ml water.

IPTG solution

200 mg ml^{-1} IPTG in sterile distilled water

Sterile filtered through 0.22-micron filter and stored at -20ºC

Loading dye (10X)

25 mg Bromophenol blue

25 mg Xylene Cyanol

1 ml 10X TAE

3 ml glycerol

6 ml sterile distilled water

Luria Broth (LB) media for 1 litre (pH 7.5)

10 g Bacto trypton

10 g NaCl

5 g yeast extract

15 g Agar (add to solid LB medium)

Polyacrylamide gel (6%) 75 ml

52.5 ml water

7.5 ml 10x TBE

15 ml Acrylamide/bisacrylamide (29:1) (w/w)

450 µl APS

100 µl TEMED

RNAase A (10 mg ml^{-1}) 10 ml

Dissolve 100 mg in 10 ml of 10 mM Tris pH 7.5. Heat in hot block for 15 min

SDS 20% (w/v)

Dissolve 200 g of SDS in 1 litre of water while stirring and heating

Sodium acetate 3 M (1 l)

Dissolve 408 g of sodium acetate in water and adjust pH to 5.2 with 3 M acetic acid.

Sodium Chloride 1.4 M

Dissolve 82 g of sodium chloride in 1 l sterile distilled water.

Sodium Hydroxide 0.4 M (1 l)

Dissolve 16 g of sodium hydroxide in 1 l sterile distilled water.

20x SSC

3 M NaCl

0.3 M Sodium citrate

20x SSPE (5 l)

3 M NaCl (876.6 g)

0.2 M NaH_2PO_4 (138 g)

20 mM EDTA (37.2 g)

1x TAE Buffer	For 1 l
40 mM Tris base	4.84 g
20 mM hydroxyacetate	1.142 ml
5 mM EDTA	0.372 g

1x TBE Buffer (pH 8.0)	For 1 l
89 mM Tris base	10.8 g
89 mM Boric Acid	5.5 g
2mM EDTA	0.83 g

Tris-EDTA ($T_{10}E_1$) Buffer

10mM Tris-Hcl

1mM EDTA, pH adjusted to 8.0 and autoclaved

X-gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside)

20 mg ml^{-1} dissolved in dimethylformamide

All *in vitro* manipulations including surface sterilisation, inoculation and sub culturing of cells, were performed under aseptic conditions in a laminar air-flow cabinet providing a minimum airflow of 0.5 m sec^{-1}. All work involving genetically engineered bacterial strains was carried out in a class II biological containment vertical laminar airflow cabinet.

2.6 DNA extraction

2.6.1 Isolation of genomic DNA for library preparation

Approximately 0.4 g of freeze dried leaf material was homogenised to a fine powder and incubated in CTAB buffer (Table 2.3) at 74°C for 20 min. DNA was extracted as described by Oberhagemann et al. (1999). The DNA was further purified by caesium chloride gradient centrifugation according to Sambrook et al. (1989).

Table 2.3 CTAB buffer composition for various DNA extractions

Reagent	Strength/Quantity		
	Library preparation	All other cultivars	Wild species
Tris	100mM pH 9.5	100mM pH 8	200mM pH 8
NaCl	1.4 M	1.4 M	1.4 M
EDTA	20mM	20mM	20mM
CTAB	2%	3%	2%
β-mercaptoethanol	0.25%	2.5%	2%
PEG 6000	1%	-	-

2.6.2 Isolation of genomic DNA for detecting polymorphism and genotyping

DNA was extracted from leaves of 1-week-old seedlings using a CTAB-based procedure with modifications according to Mace et al. (2003) for all the cultivated accessions including the mapping population. For the wild species, the protocol described by Sivaramakrishnan et al. (1997) was used. These two isolation procedures were different with respect to CTAB buffer composition (Table 2.3) as well as in the procedure. The former employed a small-scale preparation procedure designed for high-throughput facility while the latter was used for larger quantities of tissue (3 - 4 g).

2.6.3 Isolation of plasmid DNA

The vector pBluescript (Stratagene, CA, USA) was isolated from *E. coli* using the Qiagen plasmid midi protocol (Qiagen, Hilden, Germany) and purified by caesium chloride gradient centrifugation using the procedure described by Sambrook et al. (1989).

2.6.4 DNA quantification

Except for the mapping population, DNA was quantified spectrophotometrically by reading the 260 nm absorbance. The absorbance ratio of 260nm/280nm was also read to assess the purity of the DNA. The acceptable ratio of the absorbance levels for pure DNA was 1.7 – 2.0. For the mapping population, DNA quantification was done using PicoGreen (Juro Supply Gmbh, Switzerland) according to the manufacturer's

instructions. The quantity and integrity of the DNA was confirmed by resolving on a 0.8% (w/v) agarose gel and comparing with known quantities of lambda DNA.

2.7 Development of microsatellites

2.7.1 Non-enriched library

Optimised digestion conditions were determined by varying genomic DNA concentration, enzyme units per µg DNA and digestion period. Finally, 5 µg of purified genomic DNA from accession ICP 2376 was partially digested for half an hour with 5 units of *Tsp*5091. Digestion Products were size-fractioned on a 1.5% agarose gel. Fragments between 700 bp – 1000 bp long were recovered using Qiagen Gel extraction kit (Qiagen, Hilden, Germany). Eight hundred ng of plasmid DNA were digested with 10 units of *Eco*R1 in a total volume of 15 µl for 4 hours. After checking for complete digestion, 500 ng of the vector were dephosphorylated using 1 unit of SAP at 37ºC for 40 min before finally stopping the reaction at 65ºC for 20 min.

The digested dephosphorylated vector was checked on a 1% agarose gel. Recovered pigeonpea DNA fragments were ligated into the vector at a molar ratio of 3:1 using T4 DNA ligase and incubated at 16ºC overnight. Ligation products were precipitated by adding $1/20^{th}$ volume of 3 M ammonium acetate and 2 volumes of absolute alcohol, then incubating at -70ºC for 1 hour. The precipitated products were centrifuged for 30 min at 13000 rpm, washed with 70% ethanol, air-dried and dissolved in sterile distilled water. Ten ng of ligation products were mixed with 40 µl competent DH-10B cells in a clean 0.1 cm cuvette ensuring no bubbles before subjecting the mixture to electroporation (Gene Pulser, BioRad) at 1.8 KV/cm. The transformed cells were immediately mixed with 900 µl of SOC medium (Invitrogen, Freiburg, Germany) in a 2 ml tube and incubated at 37^{o}C for an hour with vigorous shaking at 250 rpm before plating out on X-Gal/IPTG/ampicillin LB-agar plates at the same temperature overnight for blue/white selection.

White colonies were carefully scraped with sterile toothpicks and used as templates for colony PCR (94ºC for 3 min, 45 cycles of 92ºC for 1 min, 55ºC for 1 min, 72ºC 1 min and final extension of 72ºC for 5 min) using T7 and T3 primers in a total volume of 25 µl. Each PCR reaction contained 5 pM of dNTP, 0.4 µM of each primer, 1.5 units of Taq polymerase, 1.6 mM $MgCl_2$ and 1x buffer (Invitrogen). Clone insert

lengths were confirmed through 1.2% agarose gel electrophoresis using 5 µl of PCR product. Clones with single inserts ranging in length from 700 bp – 1000 bp were selected for microsatellite screening. PCR products (15 µl) of selected colonies were subjected to 1.2% agarose gel electrophoresis at 100 volts for 1 hour. The DNA was denatured by incubating the gel for 30 min in 0.4 M NaOH and then transferred overnight on Hybond N^+ membrane (Boehringer, Mannheim). Eight membranes, each representing a 96-well plate PCR product, were utilised. The membranes were air-dried and DNA covalently cross-linked using Stratalinker (Stratagene, Heidelberg) at 1200 µjoules x 100. The membranes were further dried at 80°C for 1 hour and stored aseptically at room temperature.

Synthetic oligonucleotide repeats $(CA)_{15}$, $(CT)_{15}$, $(AAT)_{10}$, $(GCC)_{10}$, $(CAA)_{10}$ were end-labeled with γ-(^{32}P) dATP (Amersham Buchler, Braunschweig) in a 50 µl reaction mixture containing 50 pmol of the respective oligonucleotide, 2 µl (20 units) of T4 polynucleotide kinase, 5 µl of T4 polynucleotide kinase buffer, and 5 µl of γ-(^{32}P) dATP (10 µCi/µl) (Amersham Buchler, Braunschweig). The mixture was incubated at 37^{o}C for 1 hour after which the reaction was stopped by adding 50 µl of TE buffer and incubating at 65^{o}C for 20 min. The labelled probes were hybridised to the membranes in 20x SSPE, 100x Denhart's solution, 20% SDS and 100 µg ml^{-1} herring sperm DNA (minimises non-specific binding). Hybridisation was carried out for 16 hours at 63^{o}C for $(CA)_{15}$ and $(CT)_{15}$, 65^{o}C for $(GCC)_{10}$, 55^{o}C for $(CAA)_{10}$ and 43^{o}C for $(ATT)_{10}$. Post-hybridisation washes were carried out at 40^{o}C with stringent SSC washes. The membranes were exposed to X-ray films (Kodak) with intensifying screens for 2 hours at -70^{o}C and thereafter processed using Kodak M35A X-OMAT automated processor.This library is also referred to as library A.

2.7.2 Enriched library

Restriction of DNA with the enzyme *Sau*3AI was optimised. Finally 10 µg DNA was partially digested with 50 enzyme units in a volume of 100 µl overnight at 37°C. The size of the DNA fragments was monitored by separating a 10 µl aliquot of the restricted DNA on a 1.5% agarose gel. The rest of the DNA fragments were ligated by T4 DNA ligase onto *MluI* self-complementary adaptors *RSA* 21 and phosphorylated *RSA* 25 according to Billotte et al. (1999). Twenty-five ng of the adaptor-ligated DNA was

amplified through PCR in a 25-µl-reaction mixture that contained 1 µM *RSA* 21, 2 µl of 2.5mM dNTP, 1x buffer, 1.5 mM $MgCl_2$ and 2.5 units of *Taq* polymerase.

The PCR program involved initial denaturation at 95ºC for 1 min followed by 28 cycles of 94ºC for 40 sec, 60ºC for 1 min and 72ºC for 2 min, and a final extension at 72ºC for 5 min. Ten µl of each PCR product was viewed on a 1.2 % agarose gel to confirm amplification. The rest of the PCR product was purified through microcolumns (Qiagen, Hilden) according to the manufacturer's instructions and resuspended in 100 µl of water. Microsatellite sequences were selected using biotinylated GA_8, GT_8, GAA_8 and TAA_8 oligonucleotide probes and streptavidin-coated magnetic beads following the hybridisation based capture methodology adapted from Billotte et al. (1999).

The selected fragments were amplified using the same PCR program as in ligation step and 10 µl of each product viewed on a 1.2% gel. The rest of the products were purified using Qiagen PCR purification kit (Qiagen, Hilden), and concentrations adjusted to approximately 25 ng μl^{-1} before cloning into pGEM-T vector according to the supplier's instructions. The ligation reaction consisted of 5 µl of 2x buffer, 50 ng of vector, 3 µl of PCR product and 1 µl of T4 DNA ligase in a total volume of 10 µl at 16ºC overnight. Two µl of the ligated product was transformed into *E. coli* competent cells by electroporation as already described. The procedure for plating out and selecting white colonies was described in Section 2.7.1. White colonies were handpicked and subjected to colony PCR using T7 and SP6 primers. Clone insert lengths and amplifications were confirmed on a 1.2% agarose gel. All products with an insert size range of 300 – 1100 bp were selected for sequencing. This library is also referred to as library B.

2.8 Sequencing, sequence analysis and primer design

Positive and selected clones from library A and library B were respectively purified for DNA sequencing using EXOSAP according to the manufacturer's instructions. Approximately 25 to 50 ng of each amplicon was used as a template for DNA sequencing. DNA sequences were determined by the Max Planck Institute for Plant Breeding (MPIZ) DNA core facility (ADIS) on Applied Biosystems (Weiterstadt, Germany) ABI Prism 377, 3100 and 3730 sequencers using Big Dye-terminator v3.1

chemistry. Premixed reagents were from Applied Biosystems. A total of 2,356 genomic DNA clone inserts were sequenced from both ends using the appropriate primers.

All raw sequences were trimmed of unwanted fragments, and vector sequences using SEQUENCHER v4 (Gene Codes Corp, Ann Arbor, MI) software. MSE adaptors were also trimmed from both ends of the insert sequences from library B. Redundant sequences were determined using CAP3 sequence assembly web interface (http://deepc2.zool.iastate.edu/aat/cap/cap.html) (Huang and Madan 1999). All non-redundant trimmed sequences were screened for repeat motifs using the Simple Sequence Repeat Identification Tool (SSRIT) (Temnykh et al. 2001) available on the GRAMENE web page http://ars-genome.cornell.edu/rice/tools.html. Sequences containing at least 6 di-nucleotide repeats and 4 tri-nucleotide (or larger) repeats or larger were selected as a microsatellite. All motifs having continuous uninterrupted repeats were classified as "perfect", those with a few base pair interruptions were classified as "interrupted" and those motifs containing more than one class of repeats were classified as "compound". Mononucleotide repeats were excluded from analysis. Lengths of various repeats were defined both in terms of base pairs and number of repeats.

Primers were designed for each SSR locus using Primer 3 web interface program (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). The following criteria was used for primer design: primer length of 15 – 24 bp, GC content 35 – 55, T_m 54 – 65, product size 90 – 500 bp, $\Delta T \leq 4$. Possible hairpins, dimers and cross dimers were identified for each primer pair using NETPRIMER (http://www.premierbiosoft.com/netprimer/index..html). A total of 153 primer pairs were designed and synthesised by GenoMechanix (Florida, USA).

2.9 Functional analysis of pigeonpea sequences

All sequences from the 2 libraries were clustered together and low complexity DNA and simple sequence repeats were masked with RepeatMasker software (http://www.repeatmasker.org/cgi-bin/WEBRepeatmasker). Non-redundant sequences were used for searching various databases. The basic local alignment search tool (tBLASTx) family of programs was used to compare all sequences with public databases of *Arabidopsis thaliana*, *Medicago truncatula* and *Glycine max* L. tBLASTx

analyses were carried out using non-redundant protein database of the Institute for Genomic Research (TIGR) (http://tigrblast.tigr.org/tgi) and *Medicago* (http://www.genome.ou.edu/medicago_blast.html). The threshold for declaring significant homology was set at "e" (expect) value of ≥10, >30% identity and ≥20% alignment.

Open reading frames (ORFs) likely to encode proteins were identified by GENSCAN (http://genes.mit.edu/GENSCAN.html). The settings for the search were as follows: Organism: *Arabidopsis*; Sub-optimal exon cut-off: 1.00; Print options: Predicted CDS and peptides. Only sequences with ORFs of >99 amino acids were selected. Sequences that had no satisfactory "hits" were translated in all six frames with TRANSEQ (http://bioweb.pastuer.fr/seqana/interfaces/transeq.html) and submitted to Position-Specific Iterated (PSI) BLAST (Altschul et al. 1997) using 3 iterations with an e-value threshold of 0.001.

2.10 PCR optimisation and product electrophoresis

PCR conditions were optimised for each primer pair using a modified Taguchi method (Cobb and Clarkson 1994) as described by Buhariwalla and Crouch (2004). Three concentrations each of primer (0.2, 0.3, 0.5 pM), Mg^{++} (1.0, 1.5, 2.0 pM), and enzyme (0.2, 0.3, 0.5 U), and two concentrations each of DNA (5, 10 ng) and dNTP (0.1, 0.2) were varied in 5 different protocols (Table 2.4). Amplifications were performed in a final volume of 10 µl with 1x reaction buffer (BioLine, UK).

Table 2.4 Orthogonal arrays of permutations used for optimising PCR conditions

Reaction[a]	Primer (pmoles)	DNA (ng)	Mg^{++} (mM)	Enzyme (U)	dNTP (mM)
1	0.2 (A)	5 (A)	1 (A)	0.2 (A)	0.1 (A)
2	0.2 (A)	10 (B)	2 (C)	0.2 (A)	0.15 (B)
3	0.3 (B)	5 (A)	1.5 (B)	0.5 (C)	0.15 (B)
4	0.3 (B)	10 (B)	2 (C)	0.3 (B)	0.1 (A)
5	0.5 (C)	10 (B)	1 (A)	0.5 (C)	0.15 (B)

[1]*Total reaction volume = 10 µl*

Three different programs of "Touchdown" PCR (Don et al. 1991) with base annealing temperature ranges of 55°C - 45°C, 60°C - 55°C, and 65°C - 60°C and optimum annealing temperatures of 48°C, 56°C and 59°C respectively were used to reduce spurious amplification. The initial denaturation step was performed at 95°C for 3 min followed by 1 cycle of 94°C for 20 s, 20 s at the appropriate annealing temperatures and 30 s at 72°C (extension). In this initial annealing step, the annealing temperature was decreased by 1°C in each subsequent cycle until reaching the lowest temperature within each range. Products were thereafter amplified for 30 cycles at the appropriate optimum annealing temperature with a final extension of 20 min.

PCR products were separated on non-denaturing 6% 29:1 (w/w) acrylamide/bisacrylamide gels followed by silver staining. Silver staining consisted of 3-5 min in deionised water, 20 min in 0.1% (w/v) CTAB, 15 min in 1.3% (v/v) ammonia solution, 15 min in a solution of 250 mM NaOH and 0.1% (w/v) silver nitrate (solution titrated with ammonia). The gel was rinsed for 1 min in deionised water and thereafter developed in a 1.5% (w/v) sodium carbonate solution with 0.02% (v/v) formaldehyde solution. Gels were rinsed in deionised water for 5-10 s and then fixed in a 1.5% (v/v) glycerol solution. Reactions with combinations of reagents giving the most discrete and easy-to-score bands were selected as optimum.

2.11 Other sources of markers

Ten additional pigeonpea SSR primers (Burns et al. 2001), 6 pigeonpea resistance gene analogs (RGAs), and 72 soybean primers were available at the AGL, ICRISAT, India. Dr. Perry Cregan of USDA Laboratories kindly provided 100 soybean SSR primers that were previously mapped in soybean (Cregan et al. 1999). Eleven of the primers provided by Dr. Cregan overlapped those that were already present in the AGL reducing the number of both lots from 172 to 161 primer pairs, all of which were of ATT-origin. Primers were also designed for 59 other soybean EST-SSRs from public databases. Thus, a total of 220 soybean SSRs were tested for amplification in pigeonpea using soybean cultivar Williams' DNA as a control. Optimisation procedure for all these primers was similar to that already described (Section 2.10).

2.12 Microsatellite nomenclature

All new pigeonpea SSRs that amplified a product were named following the procedure of Yu et al. (2000) for beans (*Phaseolus vulgaris* L.). All SSR names began with uppercase CC (stands for *Cajanus cajan*) followed by the repeat motif in lower case and an arbitrary number starting from 001 for each distinct repeat motif. A repeat of TAA_8 would therefore be named CCtaa001 if it were the first one in that category. For imperfect and compound repeats, the motif with the highest repeat was used. In case both motifs had equal repeats, the repeat on the 5' end was considered. A repeat motif of "ag, ga, ct, tc" or "gaa, aga, aag, tcc, tgt and ttg" were considered of the same class. This nomenclature deviates from the earlier one used for the first identified pigeonpea SSRs and takes into consideration future expansion.

Except for the 59 soybean primers designed from the public databases, the rest of the soybean primers had been given names. For the 15 of the 59 EST-SSRs that amplified a PCR product, a temporary name beginning with "SP" followed by the respective numbers was given.

2.13 Screening for polymorphism

All markers that showed no more than 2 distinct bands after optimisation were characterised for several alleles per locus and gene diversity using 33 accessions of *Cajanus* spp. Conservation of the SSRs isolated was examined by testing for amplification of the SSR loci in other *Cajanus* species. Markers from library A were screened against 15 cultivars and 9 wild pigeonpea lines. All the rest of the markers were screened against 24 diverse cultivars with no wild type genotypes included. In each case, the genotypes used included each of the parents of the mapping population to allow easy evaluation of polymorphism between them. DNA was extracted from single plants and DNA from a single plant represented a genotype during each screening. PCR reactions were prepared in 10 µl volumes and amplification was confirmed in 1.2% agarose gels by loading 5 µl PCR products of randomly selected samples.

Amplification products were thereafter visualised on non-denaturing 6% 29:1 (w/w) acrylamide/bisacrylamide gels followed by silver staining as already described in Section 2.10. The bands visualized on the gels were scored using a binary code for presence ("1") or absence ("0") of bands (alleles) for every SSR locus.

2.14 Genotypic analysis

Primers that consistently showed different banding patterns in the two parents in at least 2 different amplification events were used to genotype 94 RILs. A single plant represented a family of the RIL, assuming complete homozygosity within families. PCR reactions were performed in volumes of 5µl and amplification products were visualised as described in Section 2.10. Gels were scored as "A" (having allele of parent A), or "B" (having allele of parent B), or "H" (having both alleles).

2.15 Preparation of fungal isolate, pots and seeds

A single conidial culture of the isolate was multiplied on 100 ml of potato dextrose broth in a flask that was placed on a rotary shaker for 10 days at 25^{o}C, 125 rpm. The content of the flask was diluted with sterile distilled water to different final inoculum concentrations depending on the experiment. Black soil was mixed with sand at a ratio of 1:3 respectively. The mixture was steam-sterilized for 24 hours, allowed to cool and then placed in 6-inch diameter pots in the glasshouse. Pre-germination of all seeds evaluated in this study was achieved by sowing in sterile riverbed sand in the glasshouse 10 days before the anticipated date of inoculation. Seeds were sterilised for 3 min in 5% (v/v) Clorax solution and then rinsed twice in distilled water before sowing. The pots containing soil and sand mixture were thoroughly watered a day before transplanting. Moisture level was maintained at 15-20%.

2.16 Inoculation procedure

During inoculation the seedlings were gently removed from the sand, roots were cleaned, then trimmed with a sterile surgical blade and dipped into the inoculum for 5 min before finally transplanting into the pots. Controls of both susceptible and resistant lines were used for every batch. The controls were divided into two lots: one inoculated and another non-inoculated. Pots were kept in the glasshouse for 1 month and wilting of the host observed. Plants that wilted before disease incubation period of 1 week were discarded as their wilting could not be attributed to the pathogen. The degree of variation for resistance to *Fusarium* wilt was measured and quantified by evaluating the host-parasite interaction in terms of disease incidence (DI) (%). Low values for DI indicated the degree of resistance of the genotype whereas higher values indicated

degree of susceptibility. The pathogen was re-isolated from the wilted plants and its pathogenicity re-confirmed.

2.17 Preliminary screening

A preliminary study involving 350 F_2 plants and 40 F_5 progenies of the mapping population was done before starting the main experiment. Each pot contained 5 seedlings and for the progenies, each pot represented 1 replication. Fifteen seeds per progeny were evaluated for disease resistance in a CRBD.

2.18 Optimisation of inoculum concentration

Five different inoculum concentrations were varied to determine the correct concentrations for screening the remaining progenies for accurate phenotypic data. The pathogen was multiplied as described in Section 2.15 and finally diluted into concentrations of 1 x 10^4, 2.5 x 10^4, 5 x 10^4, 1 x 10^5, 2 x 10^5 conidia ml^{-1} including control. Each of the parental lines including controls was screened side by side with a resistant (ICP 8863) and susceptible check (ICP 2376 – pathology source) from the pathology laboratory, ICRISAT. The experiment was done in a CRBD with 3 replicates. Each replicate consisted of 15 seedlings. Inoculation was done as already described and disease progress monitored till all the susceptible controls had wilted.

2.19 Evaluation of disease reaction phenotypes

Once the optimum concentration was established, 121 F_6 progenies were screened under the same conditions. An average of 15 seedlings per family was screened alongside parental lines for ease of scoring. The disease reaction phenotype scores were evaluated by recording the total number of plants wilting due to *Fusarium* wilt upto the 17^{th} day after inoculation. F_6 progenies were categorised as resistant (0 - 10% mortality), moderately resistant (10.1 – 30% mortality) and susceptible (30.1 – 100% mortality).

2.20 Statistical analyses

DNA bands were counted by considering those in the range of best resolution in the polyacrylamide gel (approximately 100 to 500 bp). Amplification bands were numbered according to their migration within the gel. It was assumed that bands of the same

molecular weight in different genotypes were identical. For each genotype, the presence or absence of each band was determined and designated "1" if present and "0" if absent. The genetic distance between individuals was estimated by using the markers that produced the expected size (100 – 500 bp) of amplification product. Polymorphism information content (PIC) was calculated as described by Botstein et al. (1980) using the formula below;

$$PIC = 1 - \left[\sum_{i=1}^{n} p_i^2\right] - \left[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 2 p_i^2 p_j^2\right]$$

Where p_i equals the frequency of the i^{th} allele and p_j the frequency of the $(I + 1)^{th}$ allele. This computation was done using Genstat 7.10. For diversity analysis, only data from polymorphic SSR loci was used. Genetic diversity was estimated by computing the mean number of pairwise differences over each locus among SSR binary phenotypes using Genstat 7.10 software. Similarities between any 2 genotypes were estimated according to Nei and Li (1979) as;

$$S_{ij} = 2\, N_{ij}\, (N_i + N_j),$$

Where N_{ij} is the number of bands in common accessions i and j, N_i and N_j are the total number of bands in common between any 2 accessions and may range from 0 (no common bands) to 1 (identical band profile for the 2 accessions).

Based on genetic distance estimates $Dij = 1 - Sij$, multidimensional scaling (MDS) (Kruskal and Wish, 1978) was performed to see whether the observed molecular variation indicated any evidences of clustering among accessions. A dendrogram was constructed based on the S*ij* values by adopting the sequential hierarchical agglomerative non-overlapping (SHAN) clustering technique of unweighted pair group method of arithmetic means (UPGMA), which is a variant of the average linkage clustering algorithm (Sneath and Sokal, 1973). The dendrogram was truncated at a similarity threshold value at which well-separated clusters, as indicated by MDS were obtained. These computations were performed using the statistical analysis package NTSYS-pc v2.10t (Rohlf, 1994).

For phenotypic analysis, data was evaluated using MS-Excel to determine the cut-off point for taking observations. Three percentile rankings of the observed disease severity were used to determine three rankings (susceptible, moderately resistant, and resistant). The chi-square method was used to test the goodness of fit of the segregating F_2 and F_6 populations with the expected phenotypic ratios using the formula;

$$\chi^2 = \Sigma(O - E)^2 / E$$

Where, O – Observed values, E – expected values

The segregation ratios of the 2 homozygous classes at each marker were tested for 1:1 expected proportion. The best linear unbiased prediction (BLUP) of the performance of the genotypes was obtained for *Fusarium* wilt using restricted maximum likelihood (GenStat 7.10). Linkage analysis was performed using JoinMap 3.0. A LOD score threshold of 3 and a maximum recombination fraction of 0.25 were employed as general criteria to establish linkage groups. Centimorgan (cM) units were calculated using the Kosambi mapping function (Kosambi 1944)

For marker-trait analysis, a single-marker analysis approach was used based on the G^2 statistic for independence in a 2-way contingency table as below;

$$G^2 = 2\sum O \log (O/E)$$

Where, O – observed values; E – expected number of individuals in a cell; log – natural logarithm.

3 RESULTS

3.1 Isolation and characterisation of microsatellite loci from pigeonpea

A non-enriched genomic DNA library (library A) and an enriched one (library B) were generated and screened for various microsatellite motifs (Section 2.7). The average insert size for library A was around 963 bp (Fig. 3.1) while that of library B was about 500 bp (Fig 3.2). The digestion and selection procedures used for the two libraries could explain the differences in sizes. The digestion period was longer (overnight) for library B (Section 2.7.2) and more enzyme units (5u/µg) were used compared to 35 min digestion period and 1 unit of enzyme per µg of DNA in library A (Section 2.7.1). No size selection was done for library B and the subsequent enrichment procedure also favoured selection of smaller fragments.

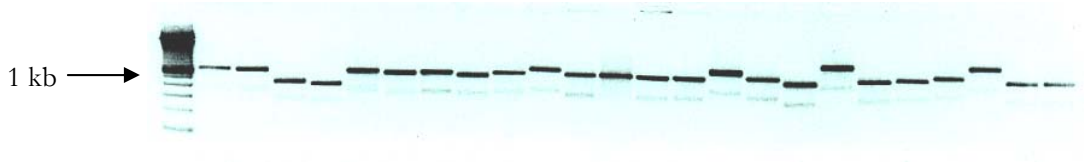


Figure 3.1 A gel picture showing sizes of randomly selected clones from library A

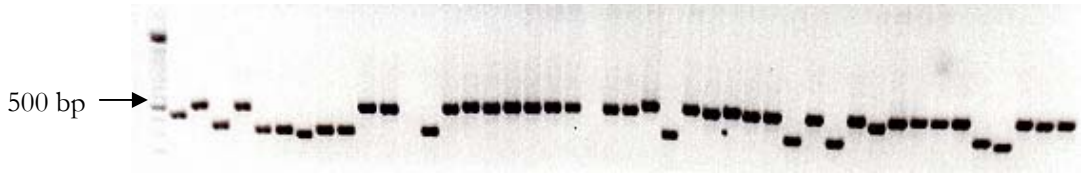


Figure 3.2 Randomly selected clones from library B showing average insert sizes of 500 bp. Lanes with no bands indicate either absence of inserts or inserts that are too small to be seen

More than half (54%) of the clones sequenced from library A were more than 1 kb long while 81% clones from library B were less than 500 bp long. Fig. 3.3 gives a detailed graphical representation of the distribution of clone sizes from respective libraries.

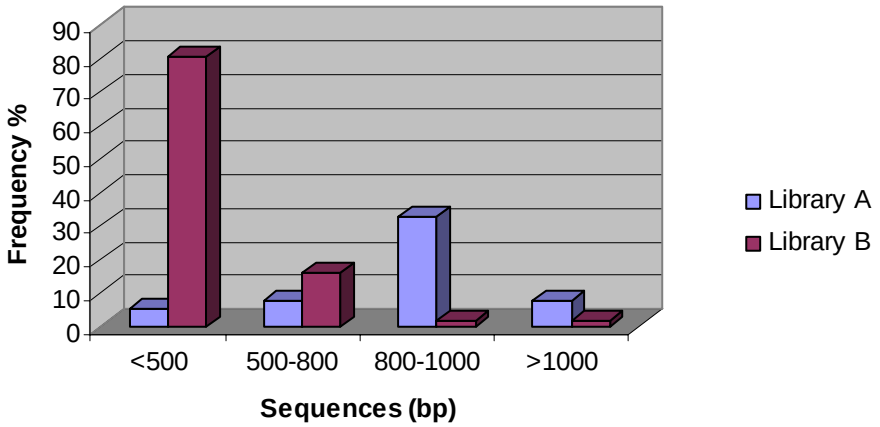


Figure 3.3 A graph showing size distribution of clones sequenced from the two libraries

Two hundred and eight clones from library A were sequenced in both directions resulting into 416 sequences. After alignment and clustering, 263 non-redundant clones were identified (Table 3.1). Ideally, all the complementary sequences should have aligned to give a final 208 sequences in the absence of redundant clones. However, 175 sequences failed to align and were thus treated as singletons while the remaining 241 sequences clustered into 88 contigs. A high number of false positives (77%) were identified in this library with only 48 out of the expected 208 sequences containing a microsatellite.

A total of 2,131 aligned forward and reverse sequences were generated from library B, 966 of which contained microsatellites. An enrichment success of 45% was therefore observed in library B before clustering. The total number of sequences was further reduced to 641 contigs after clustering reflecting an overall redundancy level of 70% (Table 3.1). The number of clones containing a microsatellite motif was similarly reduced to 117 indicating a redundancy level of 88% among microsatellite-containing sequences. The actual proportion of sequences that contained an SSR motif was further reduced to 20%. A summary of the total number of sequences generated in both libraries is shown in Table 3.1. The two libraries thus covered an estimated 253,269 bp and 286,527 bp for A and B respectively translating into approximately 0.54 Mbp of pigeonpea genome.

Table 3.1 Comparison of sequences generated from the 2 libraries

Library	No. of colonies screened	No. of colonies sequenced	Clones without inserts	Total sequences after alignment	Level of redundancy	Average size of sequences (bp)
A	753	208 (x2)	-	263	-	963
B	-	2148 (x2)	17	641	70%	447

AT-based motifs (AT and ATT) were the most abundant (44%) in library A followed by TC repeats (Table 3.2). In contrast, TAA class of repeats were the least successful enrichment (4.3%) in library B compared to TG (32.5%) and GAA (27.4%), which were the most abundant (Table 3.2). The most successful enrichments in library B were also the most redundant with 70% of the redundant clones containing either TG or GAA motifs. On the whole, di-nucleotide repeats were the longest (average 27 bp long) and also the most abundant (Table 3.2) followed closely by tri-nucleotide repeats (average 25 bp long) in both libraries. Tetra-nucleotide and hexa-nucleotide repeats formed a negligible proportion (Table 3.2) and were only a matter of chance as they were neither screened nor enriched for. No penta-nucleotide repeats were identified in both libraries. Table 3.2 shows the abundance and breakdown of the clones into nucleotide repeat classes.

Table 3.2 Distribution and characterisation of isolated microsatellite motifs

Repeat Group	Repeat family	SSRs identified		No. of primers designed		No. of primers amplifying	
		Lib A	Lib B	Lib A	Lib B	Lib A	Lib B
Dinucleotides	AT	10	2	8	2	3	2
	TC	9	16	5	15	2	8
	TG	2	38	1	37	1	28
	TA + TG	1	3	1	3	1	2
	Total	**22**	**59**	**15**	**57**	**7**	**40**
Trinucleotides	TTA	11	5	10	5	5	3
	CAT	1	1	1	1	1	1
	CAA	2	3	2	3	1	2
	GAT	-	5	-	5	-	3
	GAA	1	34	1	32	1	19
	GGT/GGA/GCT/GCA	3	4	3	4	1	1
	GGC/GCC	3	-	2	-	1	-
	Total	**21**	**52**	**19**	**50**	**10**	**29**
Tetranucleotides	CTTC/ CCTC/ GAAA/ TTAT	3	3	3	3	1	2
Hexanucleotides	TACCCG/ AAACCC/ GAAAAA/ GGGAGA	2	3	2	3	2	2
TOTAL		**48**	**117**	**39**	**113**	**20**	**73**

Sixty percent of the motifs identified were less than 20 bp long. Perfect repeats were the majority (77%) compared to interrupted (15%) and compound (8%) repeats. Specifically, library A had a higher proportion (87.1%) of perfect repeats but less of interrupted (10.3%) and compound repeats (2.6%) than library B which contained 73.5%, 16.8% and 9.7% of the respective motifs. The longest motif from library A in absolute terms (number of base pairs) was a 69 bp interrupted AT motif while a 258 bp perfect hexa-nucleotide (AAACCC) repeat was the longest in library B. In terms of repeat units, an AT with 17 uninterrupted repeats and a TG with 74 uninterrupted repeats were the longest from library A and B, respectively.

Further details of all motifs for which primers were designed are given in Appendix 6.1. Due to the high percentage of false positives identified in library A that reflected inefficiency of the hybridisation procedure, the frequency of microsatellites in the pigeonpea genome was not estimated. Library B could not be used for this purpose due to the enrichment procedure that would have resulted in over-estimation.

3.2 Primer Evaluation

3.2.1 Pigeonpea microsatellites

A total of 152 primers were designed after discarding 12 and 13 sequences from library A and B, respectively, for having bad sequences or insufficient flanking region. Three of the remaining 36 sequences from library A and 9 of 104 sequences from library B contained two different SSR motifs. Optimum conditions were established for 93 (Table 3.3) microsatellite loci resulting in amplification of 61% of the primers that were designed. Primers that were not optimised either produced little or no PCR product, or could not amplify reproducible banding patterns. The “touchdown” PCR program of 60-55°C (T_a = 56°C) was ideal for more than half of the primers (52.7%) while 65-60°C (T_a = 59°C) and 55-45°C (T_a = 48°C) both amplified the same number of markers (23.6%) (Table 3.3). Most of the bands generated under the program of 55-45°C were not sharp and often characterised with a smear or spurious bands presumably due to the low annealing temperature (T_a = 48°C). Fig. 3.4 shows some reaction results as viewed on a 6% PAGE gel. More information on the primer sequences including the reactions and PCR programs is available in Appendix 6.1. The specific sequences are available on

NCBI's Genbank database (ncbi.nlm.nih.gov) and also on ICRISAT's local intranet on: http://www.icrisat-intranet.org/gt1/pigeonpea/pigeonpea.asp.

Reaction condition 1 led to successful amplification of 42 (45%) of the primers and was the most successful followed by reaction condition 2 (21.5%) (Table 3.3). Both reaction conditions 1 and 2 (Table 2.4) made use of only 0.2 pmol (the lowest) of primer while reaction condition 2 required double concentrations of DNA, Mg^{**} and dNTP. The least useful reaction condition was 3 and this could be attributed to its low requirement of DNA concentrations with respect to all other PCR components.

Table 3.3 Optimisation of PCR amplification using various combinations of reactions and temperature ranges

Reaction condition[a]	No. of primers amplifying	Touchdown Program		
		65–60°C	60-55°C	55-45°C
1	42	8	28	6
2	20	5	11	4
3	4	1	-	3
4	10	1	4	5
5	17	7	6	4
Total	**93**	**22**	**49**	**22**

[a]*The concentrations of PCR substances for these reactions are given in Table 2.4*

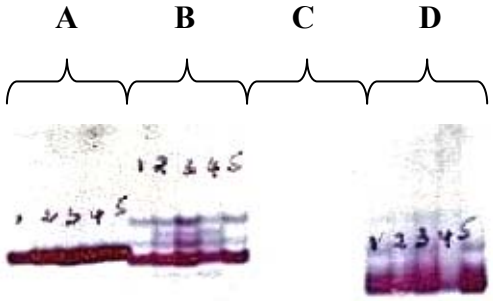


Figure 3.4 Optimisation of amplification of 4 primers, A, B, C and D. Primer C failed to amplify at this temperature (60-55°C). Each reaction is indicated as 1, 2, 3, 4 or 5. Primer A shows the best amplification under reaction 1. Primer B shows spurious amplification while primer D shows a smear. Both B and D were further optimised by varying the temperature.

In reaction conditions where bands still appeared faint after increasing the DNA concentration, substantial increase in primer and enzyme concentration (Reaction condition 5) improved amplification especially for primers that required higher annealing temperatures. A plot on the trend of optimum amplification with varying

concentrations of PCR substances showed an initial negative response in the entire "touchdown" PCR programs with increasing primer concentration. At lower annealing temperatures (48°C), amplification results did not seem to vary much irrespective of the concentrations of substances used (Fig. 3.5). The best combination of various reaction conditions and annealing temperature was that of reaction 1 under program 60-55°C (Fig. 3.5).

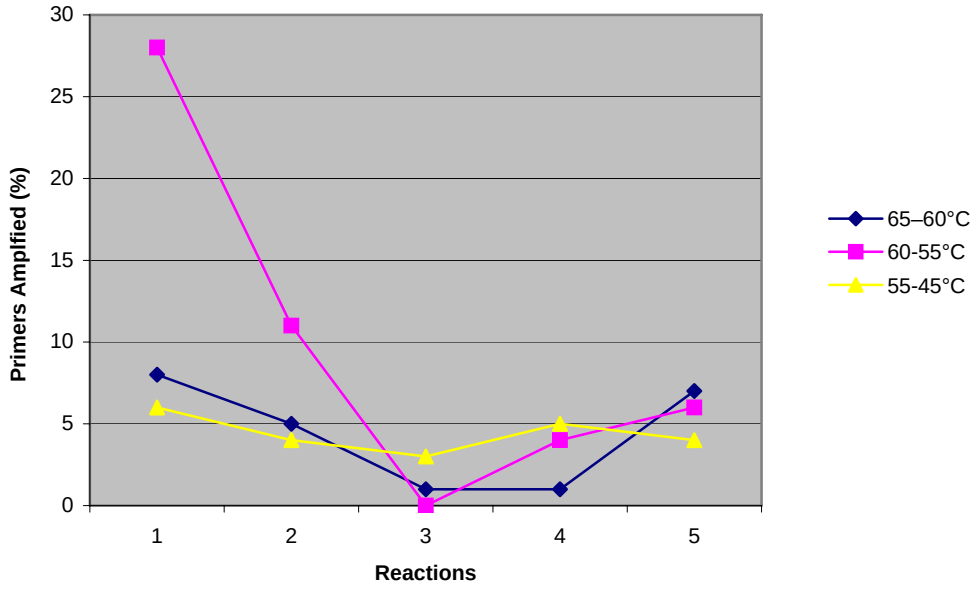


Figure 3.5 Output of "touchdown" temperature regimes in combination with various reaction conditions during primer optimisation

Amplification was achieved in 51% of all the primers designed from library A compared to 65% from the enriched library B. A lower proportion (46.7%) of di-nucleotide repeats for which primers were designed could be amplified compared to tri-nucleotide repeats (52.6%) in library A. This trend was reversed in library B where a higher proportion of di-nucleotide repeats (70.1%) than tri-nucleotide repeats (58%) amplified out of all motifs for which primers were designed.

In general, more di-nucleotide (65.3%) than tri-nucleotide repeats (56.5%) amplified – a fact that may have been influenced by the proportion of TC and AT-based repeats present in each class. Only half (52%) of the AT-based repeats amplified compared to 63% amplification of non-AT based motifs among di- and tri-nucleotide repeats. TC-based repeats amplified very poorly (50%) in both libraries in comparison

to other non-AT based di-nucleotide motifs (76.2%). Amongst the two most abundant repeats, TG amplified better (76.3%) than GAA (60.6%) repeats. Amplification varied with the length of repeats (Fig. 3.6.A) and whether the repeat was perfect, interrupted or compound (Fig. 3.6.B).

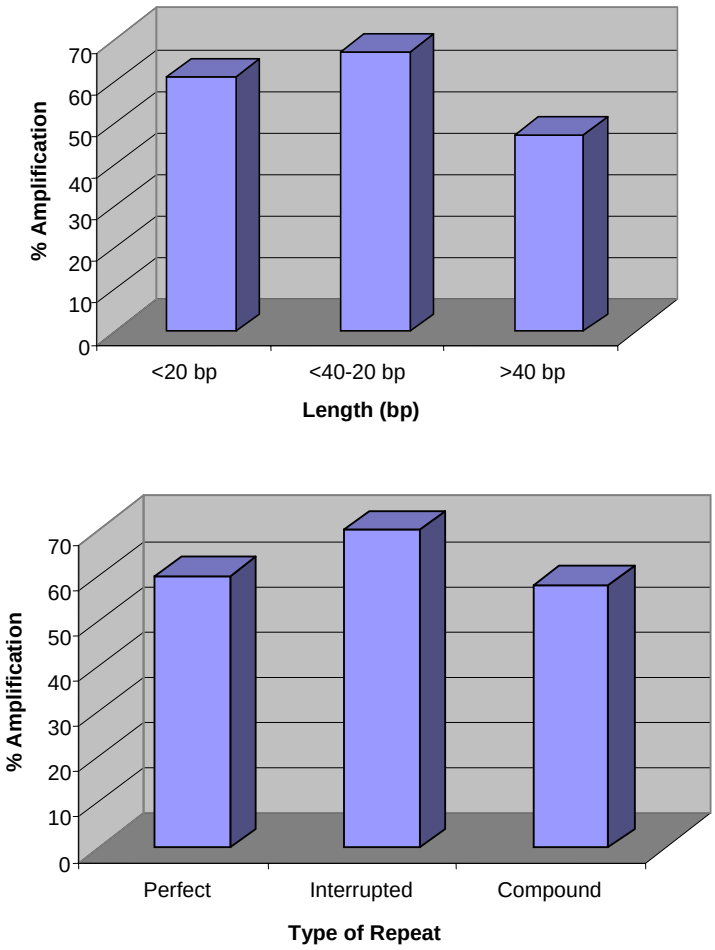


Figure 3.6 Graphs showing amplification of various types of motifs isolated from both enriched and non-enriched libraries for which primers were designed. **A.** Amplification of motifs from various length groups and **B.** Amplification of various motif types

3.2.2 Soybean microsatellites

Soybean microsatellites were tested for amplification in pigeonpea because of the presumably close relationship reported (Young et al. 2003) between the two legumes. The first set (those with names beginning with "Satt") had been derived from genomic DNA (Cregan et al. 1999) while the second set (those with names beginning with SP) were searched from soybean EST databases and primers designed (Table 3.4).

Optimum primer conditions were established for 39 out of 220 (17.7%) soybean primers tested on cultivated pigeonpea. The common problems in the interpretation of amplification products for the soybean SSRs were the appearance of excess number of bands (Fig. 3.7), smears, and amplification failure.

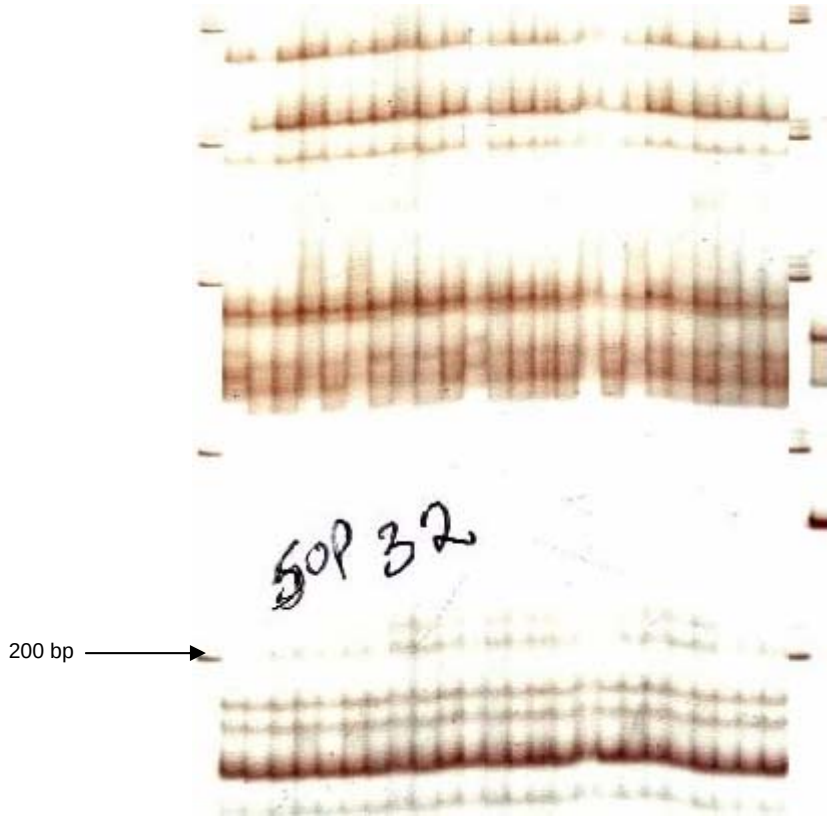


Figure 3.7 A figure showing spurious amplification of a soybean EST-SSR (SP 32) during optimisation

Table 3.4 lists all soybean microsatellites that amplified including the motifs and allele size range. All primer sequences of amplifying soybean microsatellites identified from the database are given in Appendix 6.2. The rest of the primer information can be found at http://129.186.26.94/SSR.html. A higher proportion (25.4%) of the EST-SSRs designed amplified pigeonpea DNA compared to 9.2% of the genomic SSRs tested. This situation could have also been worsened by the AT-richness of the genomic SSRs, which were all of the ATT class of repeats. The EST-SSRs had relatively shorter lengths compared to the genomic soybean SSRs (Table 3.4). Due to the low amplification rate of the soybean SSRs, very little further analysis (Section 3.5) was done on them.

Table 3.4 Optimised PCR conditions for soybean microsatellites

Locus[a]	Motif	Reaction conditions[b]	Touchdown program	Alleles size
Satt 173	(ATT)18	4	55-45	185
Satt 225	(ATT)13	4	55-45	150
Satt 239	(ATT)22	5	60-55	145
Satt 258	(ATT)9	4	55-45	115
Satt 259	(ATT)14	2	60-55	115
Satt 293	(AT)13(ATT)	5	55-45	170
Satt 308	(ATT)21	4	55-45	195
Satt 312	(ATT)11	3	60-55	250
Satt 328	(ATT)11	3	60-55	305
Satt 331	(ATT)14	3	60-55	150
Satt 336	(ATT)14	5	55-45	165
Satt 337	(ATT)19	4	55-45	255
Satt 339	(ATT)26	2	55-45	320
Satt 340	(ATT)12	3	55-45	230
Satt 343	(ATT)13	3	55-45	165
Satt 371	(ATT)11	5	55-45	390
Satt 403	(ATT)10	3	55-45	240
Satt 409	(ATT)27	4	55-45	315
Satt 442b	(ATT)35	3	55-45	400
Satt 541	(ATT)22	4	55-45	280
Satt 563b	(ATT)18	2	55-45	295-320
Satt 595	(ATT)9(GTT	2	55-45	445
Satt 599	(ATT)10	2	55-45	245
Satt 712	(ATT)21	1	60-55	350
SP 001	TGCTC(3)	5	55-45	175
SP 003	TGA(4)	5	55-45	500
SP 004	AT(26)	4	55-45	130
SP 010	CAA(6)	4	55-45	285
SP 028	TTA(4)	4	55-45	190
SP 032	CAA(5)	4	55-45	135
SP 040	CT(6)	4	55-45	400
SP 041	ATTC(5)	4	55-45	400
SP 047	TC(12)	2	55-45	145
SP 048	GAGAA(3)	2	55-45	270
SP 050	TATAT(3)	4	55-45	500
SP 053	CT(8)	4	55-45	300
SP 055	CCA(4)	4	55-45	230
SP 057	TA(10)	4	55-45	310
SP 059	CGCA(5)	4	55-45	200

[a] *All loci with names starting with "SP" are EST-SSRs identified from the public database.*
[b] *The concentrations of PCR substances for these reaction conditions are given in Table 2.4*

3.3 Functional characterisation of pigeonpea genomic sequences

3.3.1 Putative type I microsatellites

tBlastx analysis of all non-redundant microsatellite-containing sequences gave 20 significant "hits" after searching the TIGR's *Arabidopsis, Medicago* and soybean databases (Table 3.5). ORF search yielded one significant result (Table 3.5). The motifs of these putative genic SSRs were mainly characterised by short, perfect and/or

interrupted repeats. Five of the putative genic SSRs failed to amplify a PCR product (Table 3.5).

Table 3.5 SSR-containing sequences corresponding genes of described function

Accession no.	SSR name	Motif	Database[1]	Homologue	Expect (e) value
CZ681995	CCac016	(ac)6aag(ctaa)3	*Medicago*	P1 P700 chlorophyll A apoprotein	8.7e-30
CZ681933	CCttc008	(aga)4	Soybean	Cytochrome B6	1.2e-26
CZ681957	CCttc011	(aga)4	*Medicago*	*SNF2/SW12* family global transcription factor	2.2e-32
CZ681962	CCac007	(tg)(tc)2(tg)7	Soybean	Unknown factor	4.9e-14
CZ445540	CCggt001	(ggt)4	*Medicago*	Predicted protein	1.7e-35
CZ682014	CCac030	(cata)3ta(tg)6, (ac)8(at)7acat	*Arabidopsis*	Chloroplast 30S ribosomal protein	9.9e-25
CZ681997	N/A	(ag)6g(c)9*	Soybean	Similar to LRRGT00012 (*Rattus norvegicus*) (hypothetical protein)	2.2e-20
CZ445554	CCgtt001	(aac)4	*Arabidopsis*	ATP dependent RNA helicase-like protein	4.9e-65
CZ681965	CCgaaaaa001	(gaaaaa)5	Soybean	Soybean|CF922419 (hypothetical protein)	1.1e-11
CZ681967	CCttc015	(gaa)2gagg(gaa)4gag(gaa)2	Soybean	NADH dehydrogenase	2.7e-11
CZ445521	CCat004	(ta)4(gatag)(at)4	Soybean	Hypothetical protein 91 – garden pea chloroplast	2.0e-21
CZ681970	CCtc007	(tc)6, (ctt)4	*Medicago*	*Ycf3* protein	4.5e-12
CZ681980	N/A	(ttc)5*	Soybean	Similar to Emb|CAB83157.1 (hypothetical protein)	2.2e-14
CZ445542	CCggc001	(ggc)4	*Arabidopsis*	Phospholipase-like protein	3.0e-27
CZ681999	CCac017	(caccac)(a)5(ca)6c(a)4	*Medicago*	Protein kinase-like protein	1.2e-17
CZ445543	N/A	(tat)4*	*Medicago*	Ribosomal protein S14	4.3e-25
CZ681990	CCac013	(tg)6(agtg)3	Soybean	Hypothetical protein	1.0e-13
CZ445544	CCttat001	(ttat)4	Soybean	Similar to (At2g39450) (*Arabidopsis thaliana*) – unknown protein.	1.4e-27
CZ445547	N/A	(ga)6*	*Medicago*	*RPE15* protein (hypothetical protein)	5.6e-32
CZ445549	N/A	(aat)4*	*Medicago*	Transposon	2.1e-35
CZ445539	CCcat001	(cat)4	Genscan	ORF 220 bp long	N/A

[1]*In cases where all the databases had similar "hits" of the same sequences, the database with the highest "hit" was used.*

**Did not amplify a PCR product*

About half (10 out of 21) of the deduced proteins were either hypothetical proteins or unknown factors. Six (CCggt001, CCggc001, CCttat001, CCgtt001, CCat004, CCcat001) out of the remaining 15 putative type I SSRs were tested for cross-species amplification and all of them amplified various wild pigeonpea genotypes (Table 3.7, Table 3.8).

3.3.2 Coding capacity of sequences generated

A total of 216 out of 764 non-SSR containing sequences examined by tBLASTx analyses produced database matches with *P(N)* values $\leq 10^{-10}$. These "hits" suggested putative identities or functions for the hypothetical proteins encoded by these genomic sequences. Those "hits" giving complete homologies to genes of described functions are given in Table 3.6.

Table 3.6 Pigeonpea genomic sequences similar to genes of described function in important physiological processes

Database	Query Id	Homologue[a]	Functional Class	E-value
Arabidopsis thaliana (TIGR) database	DU169719	Chloroplast 30S ribosomal protein S7	Protein synthesis	1.7e-36
	DU169728	Receptor-like protein kinase	Signal transduction	6.0e-21
	DU169766	Chloroplast 30S ribosomal protein *S7*	Protein synthesis	1.6e-11
	DU169775	*Ndh*B	Electron transport	2.5e-11
	DU169797	Photosystem II P680 chlorophyll A apoprotein (*CP-47* protein)	Energy metabolism	6.9e-25
	DU169807	*ycf*2 protein	Chloroplast ORF	5.2e-35
	DU169812	P700 chlorophyll A apoproteinA1 (PsaA) (PSI-A)	Energy metabolism	4.7e-56
	DU169828	*ycf*2	Chloroplast ORF	1.8e-32
	DU169831	*ycf*2	Chloroplast ORF	3.4e-38
	DU169848	Photosystem II 44 kDa reaction centre protein (P6 protein) (CP43)	Energy metabolism	1.5e-65
	DU169851	Photosystem I P700 chlorophyll A apoproteinA1 (*PsaA*) (*PSI-A*)	Energy metabolism	8.0e-54
	DU169859	(Photosystem II protein D1)	Energy metabolism	3.1e-33
	DU169870	*ycf*2 {*Arabidopsis thaliana*}	Chloroplast ORF	2.9e-13
	DU169886	Photosystem I P700 chlorophyll A apoproteinA1 (*PsaA*) (*PSI-A*)	Energy metabolism	1.7e-42
	DU169892	Photosystem I P700 chlorophyll A apoproteinA1 (*PsaA*) (*PSI-A*)	Energy metabolism	5.5e-46

[a]*In cases where all the databases had similar hits of the same sequences, the database with the highest hit was used*

Table 3.6 Continued

Database	Query ID	Homologue	Functional Class	E-value
Glycine max (TIGR) database	DU169695	Maturase K (Intron maturase)	Protein synthesis	3.2e-31
	DU169710	Acetyl-coenzyme A carboxylase carboxyltransferase subunit beta (ACCASE beta chain)	Lipid metebolism	2.2e-17
	DU169741	Photosystem II protein D1 precursor	Energy metabolism	6.3e-43
	DU169814	Ribosomal protein L2	Protein synthesis	1.4e-44
	DU169823	Chloroplast envelope membrane protein	Transport	3.2e-16
	DU169837	Acetyl-coenzyme A carboxylase carboxyltransferase subunit beta (ACCASE beta chain).	Lipid metabolism	3.3e-16
	DU169839	*Rpo*C2 {Glycine max}	Protein synthesis	6.3e-17
	DU169852	*Rpo*C1 {Glycine max}	Protein synthesis	1.5e-38
	DU169853	Photosystem II protein D1 precursor	Energy metabolism	2.4e-19
Medicago truncatula database (University of Oklahoma)	DU169729	Ribose-phosphate pyrophosphokinase Isoform II	Nucleotide metabolism	1e-40
	DU169782	Soybean RFLP_A458	Marker	5e-45
	DU169783	Mt R_gene homolog	Stress/defense	4e-30
	DU169808	Mt RGA-MtA4(AZ758017)	Marker	3e-49
	DU169810	*Rho* (ras-related) GTP-binding protein, *rac* G-Protein (*rac*1 gene)	Signal transduction	6e-55
Medicago truncatula (TIGR)	DU169688	NADH-plastoquinone oxidoreductase	Electron transport	1.5e-19
	DU169727	NADH-plastoquinone oxidoreductase	Electron transport	1.1e-26
	DU169743	Photosystem II P680 chlorophyll A apoprotein (CP-47 protein)	Energy metabolism	9.6e-55
	DU169816	Cytochrome B6	Electron transport	1.3e-27

[1]*In cases where all the databases had similar hits of the same sequences, the database with the highest hit was used*

Sequences from the enriched library predicted a higher proportion of putative proteins (34%) compared to sequences from the non-enriched library (19%). This probably had to do with the restriction endonucleases used for the 2 libraries. *Sau*3A1 is sensitive to cytosine methylation and therefore expected to preferentially cut in hypo-methylated gene-rich regions as opposed to *Tsp*5091. Chloroplast encoded genes (Table 3.6) formed the highest proportion of the deduced proteins having complete homology with genes from other species. Some of the deduced proteins shown in Table 3.6 may

have been identified in all the databases searched but only the database giving the lowest "e" value is shown.

Searches to the *Medicago truncatula* database (University of Oklahoma) gave only 17 significant "best hits", compared to 95, 70 and 43 from TIGR's soybean, *Arabidopsis* and *Medicago* databases respectively. Of the 17 "hits" from the former database, 10 were unknown factors/hypothetical proteins while 3 were markers (1 RGA and 2 RFLPs). The TIGR's soybean database gave the most "best hits", suggesting that soybean could be the closest to pigeonpea of all the crops of which the databases were searched. *Arabidopsis* database searches gave more significant "best hits" than the TIGR's *Medicago* even though the latter is a legume and therefore was expected to be more closely related to pigeonpea. ORF search yielded 26 significant results (more than 100 amino acids) while PSI search gave 1 result of a putative retro-element with an e-value of $2e^{-04}$.

3.4 Cross-Species amplification of pigeonpea microsatellites

Pigeonpea SSRs from library A for which optimum PCR conditions were successfully established (20) were tested for cross-species amplification among 9 wild pigeonpea types. All microsatellites allowed allele detection on all the genotypes and also showed variation in allele sizes except in one instance (CCtta001). Fig 3.8 shows the variation in sizes of alleles for 5 randomly chosen microsatellite loci tested.

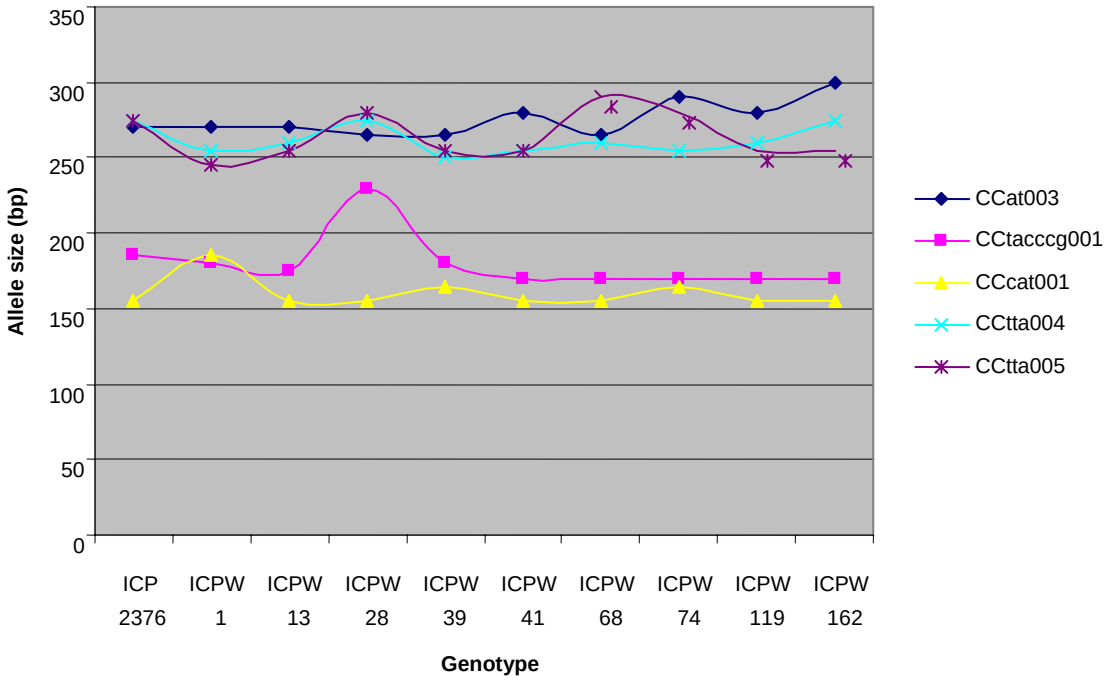


Figure 3.8 Allele sizes (in bp) of various loci in different genotypes of the genus *Cajanus*

3.5 Polymorphism of microsatellites in pigeonpea

While amplification was possible for 39 soybean microsatellites in pigeonpea (Section 3.2.2), only 1 was polymorphic among 24 diverse cultivated pigeonpea germplasm. The polymorphic microsatellite was an ATT-repeat (ATT_{18}) and showed 4 alleles ranging from 290 to 335 bp. All amplifying SSR markers from library A detected polymorphisms amongst 24 diverse accessions (including wild pigeonpea) except CCtta001 (Table 3.7). Thirty-five microsatellite loci from library B showed polymorphism in 24 cultivated pigeonpea germplasm. Detailed analysis of the allelic variation from each library is given in Section 3.5.1 and 3.5.2.

Table 3.7 Properties of pigeonpea microsatellite markers from library A

SSR Name	Core Motif	Allele size	N_ac	N_aw	N_ac+w	PIC_c	PIC_w	PIC_{c+w}
CCttc001	$(ttc)_5$	185-220	1	9	9	0	0.85	0.57
CCtta001	$(tta)_{10}$	220	1	1	1	0	0	0.00
CCat001	$(ta)_8$	280-340	5	7	7	0.64	0.78	0.69
CCat002	$(ta)_{10}(tg)_9$	355-500	6	9	10	0.47	0.82	0.76
CCat003	$(at)_{11}$	265-300	1	5	5	0	0.74	0.48
CCtta002	$(tat)_9$	240-320	5	3	5	0.56	0.55	0.58
CCtacccg001	$(tacccg)_4$	170-230	2	9	10	0.37	0.83	0.80
CCtacccg002	$(cgggta)_4$	305-320	3	3	3	0.46	0.46	0.47
CCtta003	$(tta)_4$	180-190	1	3	3	0	0.53	0.28
CCcat001	$(cat)_4$	155-185	1	3	3	0	0.44	0.21
CCggt001*	$(ggt)_4$	205-210	1	2	2	0	0.16	0.07
CCtc001	$(ct)_6$ tt$(ct)_2$	160-225	1	6	6	0	0.73	0.49
CCggc001*	$(ggc)_4$	200-260	2	5	5	0.12	0.7	0.49
CCttat001*	$(ttat)_4$	210-250	1	6	6	0	0.75	0.43
CCtta004	$(tta)_4$	250-260	1	4	4	0	0.66	0.44
CCtta005	$(aat)_4$	245-290	1	6	6	0	0.73	0.39
CCac001	$(tg)_6$	250-290	1	3	3	0	0.45	0.17
CCtc002	$(ga)_{12}$	185-200	3	3	3	0.47	0.50	0.49
CCgtt001*	$(aac)_4$	170-190	2	3	4	0.19	0.41	0.31
CCat004*	$(ta)_4(gatag)(at)_4$	210-220	3	3	3	0.21	0.31	0.26

N_ac – Number of alleles in cultivated genotypes; N_aw- Number of alleles in wild relatives; N_ac+w – total number of alleles in both wild and cultivated; PIC_c – PIC for cultivated genotypes; PIC_w – PIC for wild relatives; PIC_{c+w} – Total PIC for both wild and cultivated.
**Putative genic SSRs*

3.5.1 Allelic diversity for markers generated from library A

A greater diversity in terms of allele size range was observed among the wild genotypes as compared to cultivated pigeonpea (Fig. 3.9). A total of 98 alleles were amplified with an average number of 4.9 alleles per locus. The PIC value ranged from 0.17 to 0.80 with 5 (25% of the amplified primers) SSRs having a PIC value more than 0.5 (Table 3.6). Most of the polymorphic SSR loci contained di- (35%) and tri- (50%) repeats and only 3 (15%) had tetra- and hexa- repeats.

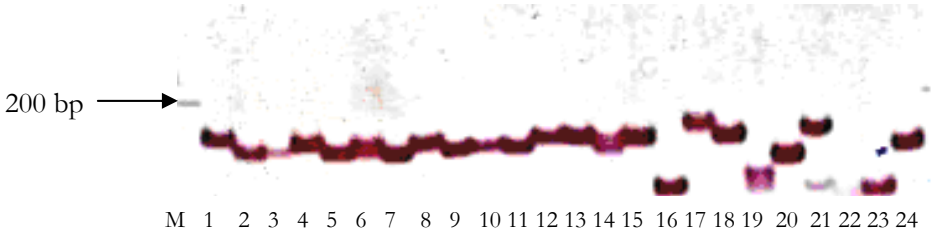


Figure 3.9 A gel picture showing amplification of a polymorphic microsatellite. Only 2 alleles are shown within the cultivated species (1-15) compared to 5 within the wild species (16-24).

Substantially less variation was detected within the cultivated species with only 9 markers detecting polymorphism and a total of 31 alleles. The average number of alleles was 3.4 with an average PIC of 0.39 per locus. Five alleles out of 31 were unique to cultivated species and were found in loci CCat002, CCtta002, CCtacccg001, and CCgtt001. The two unique alleles from locus CCtta002 were only found in accession ICP 13575, which is a landrace from Sierra Leone. There was more variation within the wild species with 19 polymorphic loci having 92 different alleles. The average number of alleles and PIC were 4.8 and 0.60 respectively. A total of 56 alleles were unique to the wild species.

All the polymorphic SSRs (19) were used to analyse and group the 24 *Cajanus* genotypes. Both the MDS plot (Fig. 3.10) and UPGMA dendrogram (Fig. 3.11) gave a similar clustering pattern of all the genotypes studied. There was one major cluster of the cultivated genotypes while the wild species remained individually distinct from each other (Fig. 3.10). There was a remarkable similarity between *C. cajanifolius* and cultivated pigeonpea types (Fig. 3.11). Specifically, the landrace ICP 14352 from Venezuela had the highest similarity (81%) to *C. cajanifolius*. This level of similarity between a specific cultivated species and a wild species might be indicative of the low levels of improvement that has been done on the landrace, but could also give clues to the evolution of cultivated pigeonpea. ICP 7543 and ICP 14144 showed the highest genetic similarity (98%) even though the 2 landraces were collected from India and Jamaica respectively. The only ICRISAT-developed breeding line (ICPL 87119) used in this analysis did not cluster very closely with the other landraces, which could be indicative of its superiority.

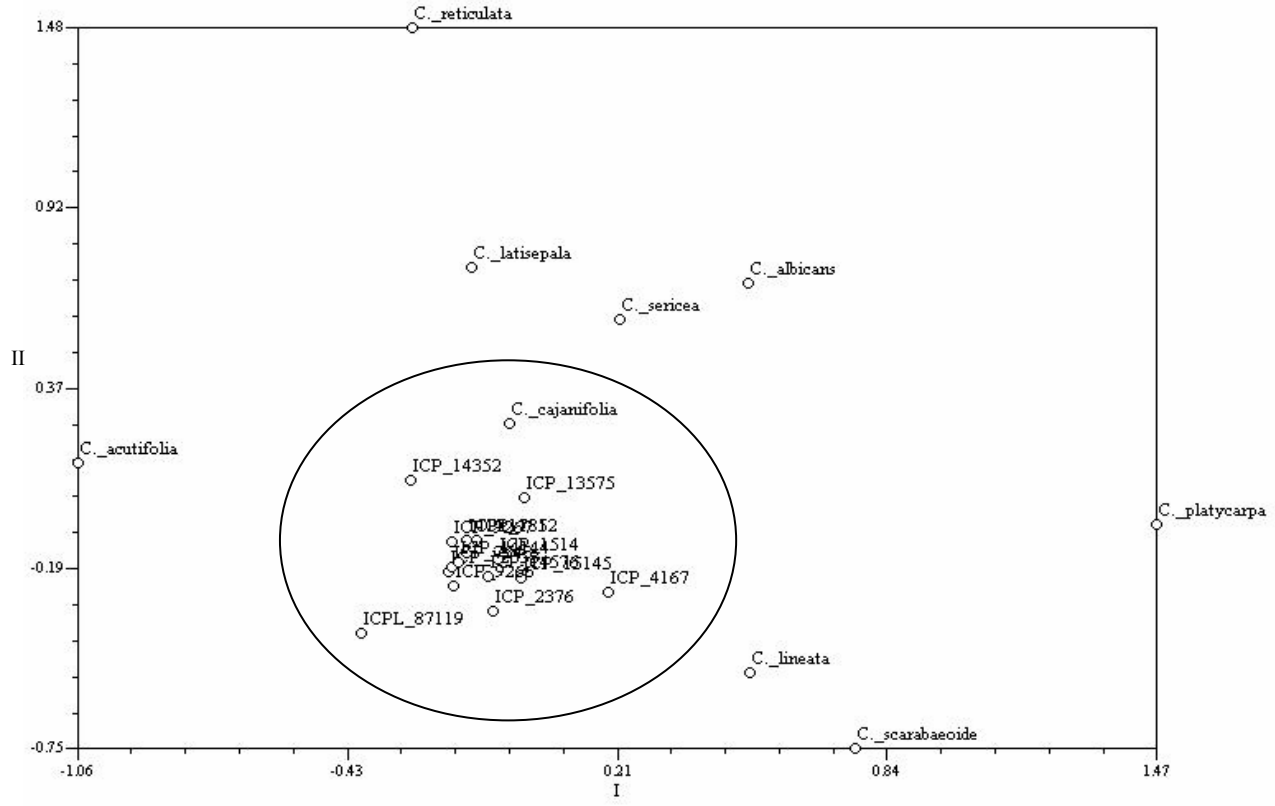


Figure 3.10 Multidimensional 2D plot of *Cajanus* accessions generated from SSR data. The circle shows a cluster of cultivated species but which also includes *C. cajanifolia.* The scale shows various coordinate levels

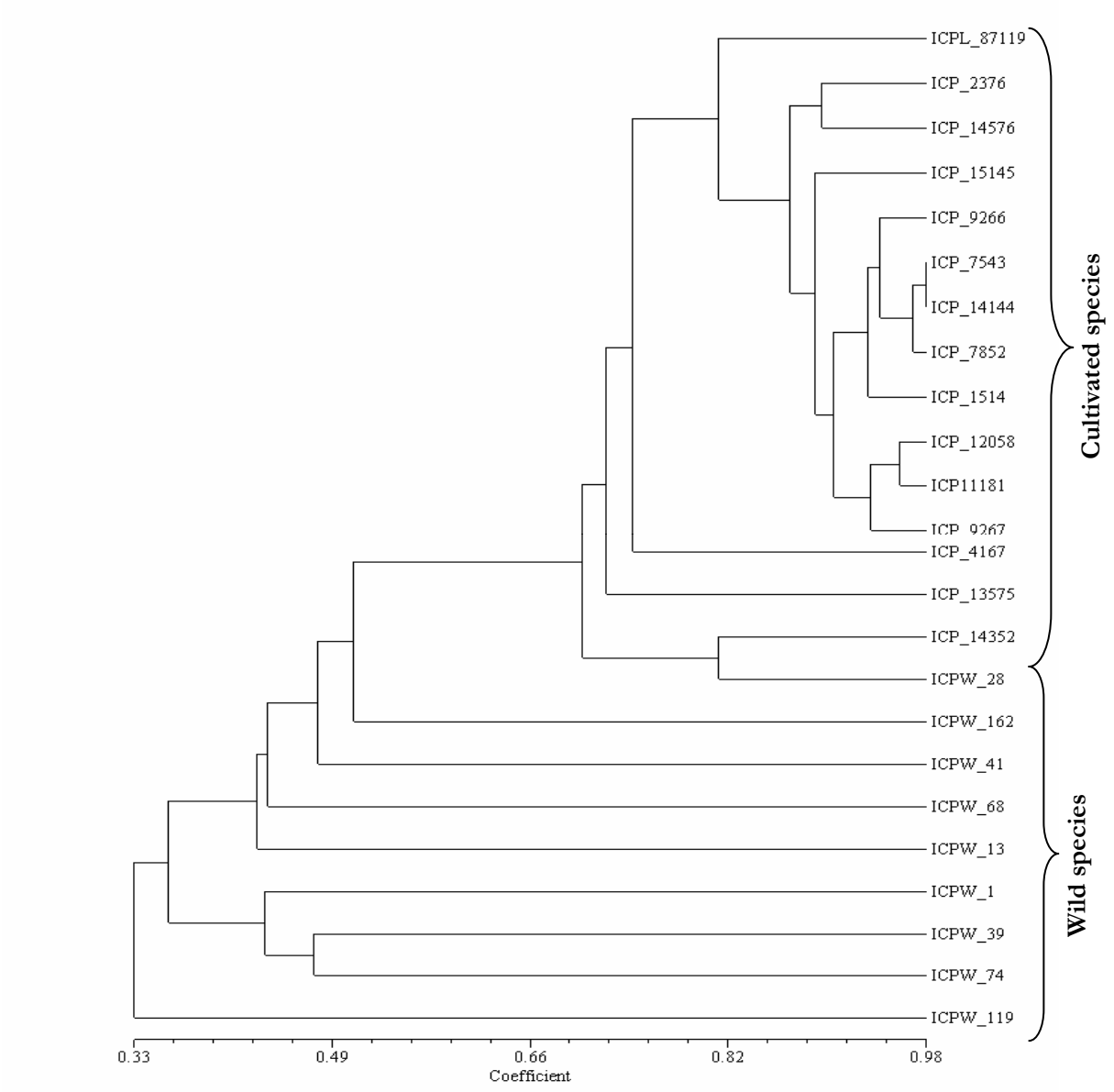


Figure 3.11 Dendrogram resulting from UPGMA cluster analysis of 24 pigeonpea genotypes based on data of 19 microsatellite primer pairs. The scale indicates similarity levels

3.5.2 Allelic diversity for markers generated from library B

All the 73 SSR markers from library B were used to characterise 24 cultivated genotypes. A total of 148 alleles were identified from 73 markers, 38 of which were from non-polymorphic markers. Table 3.8 gives a detailed characterisation of all the polymorphic markers from library B.

Table 3.8 Properties of polymorphic microsatellite markers from library B

SSR Name	Core Motif	Allele size	No. of alleles	PIC
CCttc002	(gaa)5g(gaa)5	184-215	5	0.63
CCttc004	(gaa)6	240-250	2	0.08
CCac003	(ca)8	175-198	5	0.72
CCttc005	(gaa)11gag(gaa)5gaggaagag(gaa)17	305-320	5	0.76
CCac004	(ta)5(tg)7ta(tg)4	243-245	2	0.08
CCttc007	(aga)5	265-275	4	0.54
CCtc003	(tc)8	175-190	2	0.07
CCtc005	(ag)20	165-185	4	0.59
CCtta006	(att)21	290-310	5	0.71
CCcttc001	(cttc)4	260-270	4	0.28
CCgaaa001	(cttt)4	220-225	2	0.37
Ccctа001	(gat)5(tct)(gat)4	275-302	3	0.29
CCac006	(ca)10cg(ca)6	295-350	5	0.49
CCgtt002	(tgt)4	210-235	3	0.40
CCttc012	(ttc)7	170-190	3	0.42
CCac007[a]	(tg)(tc)2(tg)7	275-280	2	0.08
CCgtt003	(ttg)5(ttc)7	165-180	2	0.37
CCtc007[a]	(tc)6	310-330	4	0.68
Ccac010	(ca)7	191-198	3	0.50
Ccac011	(gt)7	227-270	2	0.36
CCttc017	(aga)11(ggag)(gaa)4ga(gga)3a(gaa)16	145-150	3	0.37
Ccat006	(ta)7(ca)6	220-277	3	0.54
Ccctа003	(gat)4	420-450	3	0.39
CCac015	(ac)4aa(ac)38c(ca)7	135-145	2	0.14
CCtc009	(tc)6	200-220	4	0.45
CCac017[a]	caccac(a)5(ca)6c(a)4	215-227	2	0.24
CCac018	(ac)6a	200-210	3	0.54
CCac019	(tg)6	130-135	2	0.36
CCac026	(ac)7	278-295	4	0.64
CCac030	(cata)3ta(tg)6	236-244	2	0.37
CCttc018	(aga)5	275-288	2	0.35
CCac027	(tg)7	295-300	2	0.21
CCttc019	(aag)13	220-245	6	0.66
CCttc020	(ctt)8	236-242	2	0.37
CCac029	(caa)(ca)6caa	160-180	3	0.33

[a]*Putative genic SSRs*

The number of alleles detected ranged from 2 - 6 at each of the 35 polymorphic loci with a total of 110 alleles and an average of 3.1 alleles per locus. Gene diversity values ranged from 0.07- 0.76 with an average of 0.41. Di-nucleotide repeats formed the highest proportion of polymorphic markers followed by tri-nucleotide repeats. None of the hexa- and only two of tetra-nucleotide repeats were polymorphic. While TG class of repeats formed the highest proportion (40%) of all the polymorphic loci, the highest number of alleles (6) was observed from a perfect tri-nucleotide repeat

(CCttc019). The most informative marker (PIC = 0.76) was a tri-nucleotide compound repeat (CCttc005), which was also the longest motif among those analysed for polymorphism.

Ten (ICPL 87119, ICP 2376, ICP 13575, ICP 9266, ICP 14576, ICP 14352, ICP 1514, ICP 7543, ICPL 332, LD Dwarf) accessions harboured unique alleles, 7 of which were landraces. The accession ICP 9266 contained the highest number (4) of unique alleles followed by ICPL 87119 (3 unique alleles). The former accession is a landrace from Guyana (South America) while the latter is an ICRISAT developed line. Five alleles found at loci CCttc005 (2 alleles), CCcttc001, CCttc012, CCcta003 were present only in a few improved lines (ICPL 87091, LD Dwarf, ICPL 332, ICPA 2068, ICPL 151, HPL 24, ICPL 87119, ICPA 2032) and absent in all landraces. ICPL 332 harboured all the alleles that were unique within improved cultivars. Five alleles on loci CCcta001, CCac006, CCttc012, CCtc007 and CCac026 were unique to the landraces. The accessions were clustered using MDS (Fig. 3.12) and UPGMA (Fig. 3.13) as already described for library A.

The MDS plot did not give a clear cluster of individuals even on the 3-dimensional view. The UPGMA dendrogram gave 4 major clusters based on geographical adaptation (Fig. 3.1.3). A landrace collected from Thailand was the most distant of all the accessions tested at about 81% similarity co-efficient compared to the highest similarity of 94% between ICP 12058 and ICP 13092 (Fig. 3.13). The 2 genotypes with the highest similarity co-efficient were both from East Africa with the former being native to Tanzania and the latter from Kenya.

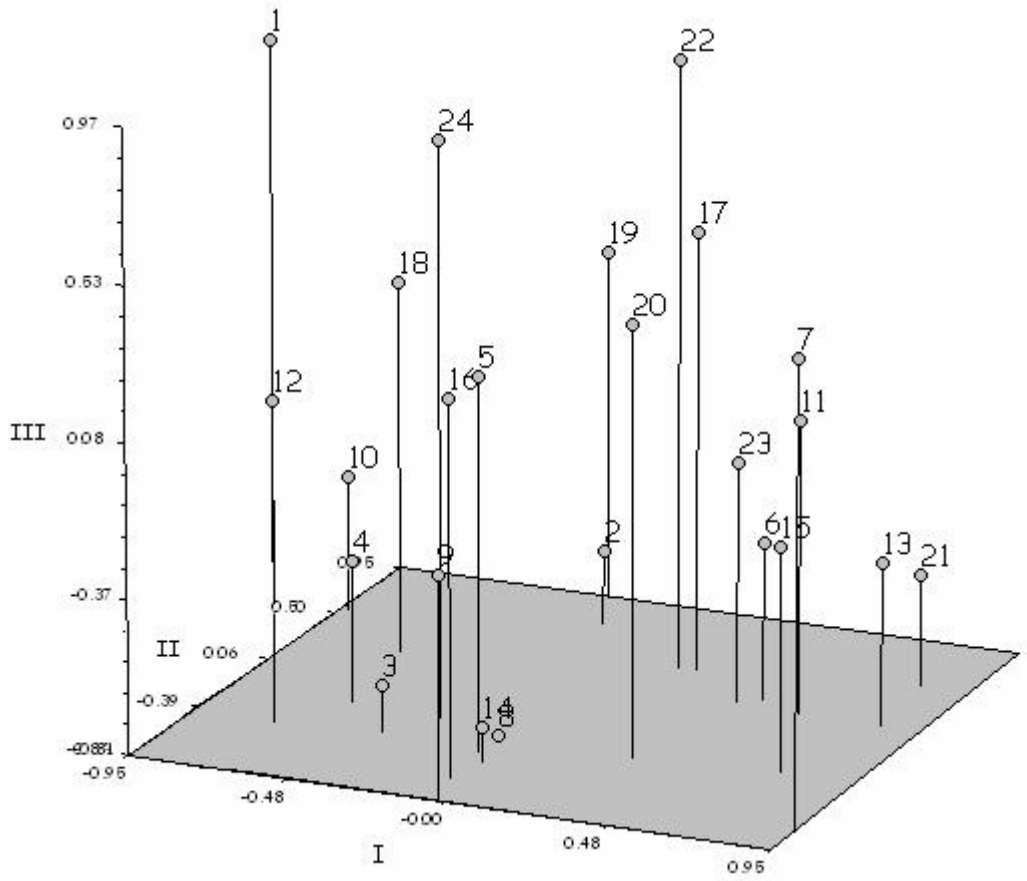


Figure 3.12 Multidimensional 3D plot showing associations among 24 cultivated pigeonpea genotypes. The scale shows various coordinate levels. The observations are coded as follows: 1-ICPL 87119, 2-ICP 2376, 3-ICP 13575, 4-ICP 15145, 5-ICP 9266, 6-ICP 4167, 7-ICP 14576, 8-ICP 12058, 9-ICP 14352, 10-ICP 1514, 11-ICP 7543, 12-ICP 7852, 13-ICPL 151, 14-ICP 13092, 15-ICPL 87091, 16-Kat 60/8, 17-ICP 7035, 18-HPL 24, 19-LD Dwarf, 20-ICPL 99066, 21-MN 5, 22-ICPL 332, 23-ICPA 2068, 24-ICPA 2032

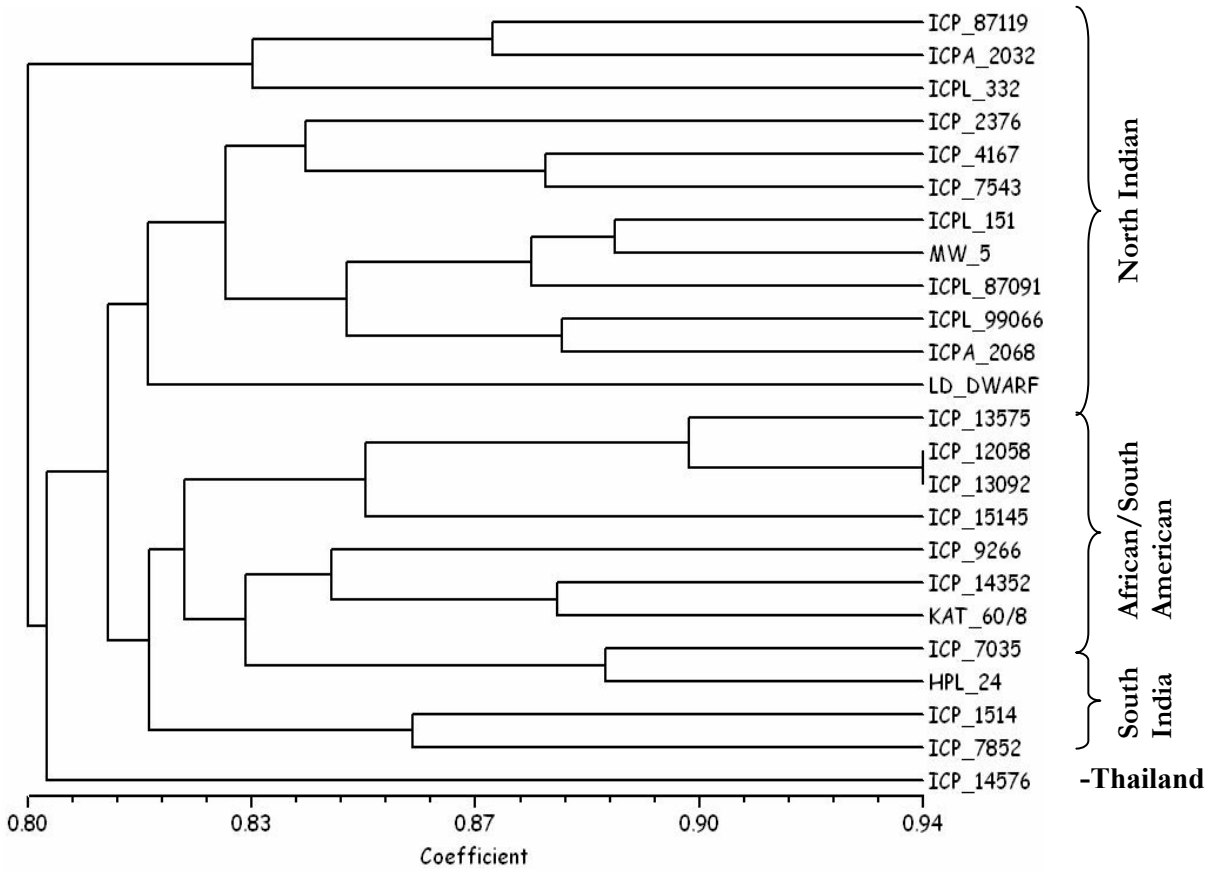


Figure 3.13 Dendrogram resulting from UPGMA cluster analysis of 24 cultivated pigeonpea genotypes based on 44 SSR primer pairs. The scale shows similarity levels.

3.5.3 Comparison of procedures used to develop microsatellites

To avoid bias estimates, markers that were polymorphic only in wild species in the case of library A were regarded in this analysis as monomorphic. Library A turned out to be the most efficient source of polymorphic microsatellites for cultivated pigeonpea despite the high number of false positives (Table 3.9). Use of soybean primers was the least efficient even though it involved less effort. Despite having the highest absolute number of polymorphic primers, efficiency of library B was reduced by the high redundancy level as well as low enrichment success. This enriched library was also the most expensive considering the high number of sequencing that it involved.

Of all the pigeonpea primers designed (152), 44 (47.3%) were polymorphic among cultivated germplasm detecting 141 alleles while the PIC values ranged from 0.07 - 0.76. The average number of alleles per locus for all primers polymorphic among cultivated germplasm was 3.2 while the PIC was 0.41. A further comparison of allelic

variation revealed that the proportion of polymorphic markers to primers that amplified a product was not significantly different for the two libraries (45% for library A and 48% for library B) (Table 3.9). Comparable results were achieved for both libraries in terms of number of alleles and heterozygosity (PIC) values (Table 3.9). Polymorphism was found in all classes of repeats and also in various lengths of motifs (Table 3.10).

No correlation was found between the number of alleles and core motif (Table 3.10), although these results will need to be verified with a larger database for tetra- and hexa-nucleotide repeats. There was neither any significant correlation between the average number of alleles (and hence the PIC) and length of repeats. The shortest polymorphic motifs were 12 bp long with PIC values as high as 0.68 while one of the longest motifs (101 bp long) (CCac015) detected only 2 alleles with a PIC value of 0.14.

Table 3.9 Comparison of microsatellite sources and levels of attrition

Stage	**Source of Microsatellites**			**Level of Attrition (%)**		
	Lib A	Lib B	Soybean	Lib A	Lib B	Soybean SSR
Successfully sequenced clones	208	2131	N/A	N/A	N/A	N/A
Unique clones	263[a]	641	N/A	-	**70.1**	N/A
SSR-containing clones	48	125	N/A	**76.9**	24	N/A
Loci for primer design	39	113	220[b]	4.3	0.6	N/A
Amplifying loci	20	73	39	9.1	1.9	**82.3**
Polymorphic loci[e]	9 **(4.3%)**	35 **(1.6%)**	1 **(0.5%)**	5.3	1.8	17.3
Range of alleles	2-6	2-6	4[c]			
Average no. of alleles per locus	3.40	3.14	N/A			
PIC range	0.12-0.64	0.07-0.76	N/A			
Average PIC	0.39	0.41	N/A			

[a]*This number is higher than the number of clones sequenced because some reverse and forward sequences did not align and were therefore treated as unique clones.*
[b]*161 primers were obtained already designed. Primers were designed for 59 soybean EST-SSRs*
[c]*This was based on 1 microsatellite (n=1) hence not comparable to other results.*
[d]*Numbers in brackets give the final percentage of polymorphic primers obtained.*
Cells in grey are not applicable

A higher proportion of repeats that were at least 20 bp long (60%) were polymorphic compared to only 38% polymorphism observed among short repeats (Table 3.10). Compound repeats were the most polymorphic (71.4%) of all primers that amplified as compared to interrupted (62.5%) and perfect repeats (41.4%). This probably had to do with the length of the various motifs because 91% of all the compound and interrupted repeats were at least 20 bp long compared to only 26% of perfect repeats that were longer than 20 bp.

Table 3.10 Average number of alleles for the 44 loci

SSR Type		**No. of loci**	**Average no. of alleles**	**Average PIC**	**% Polymorphic[a]**
SSR class of repeat	Perfect	29	3.2	0.43	41.4
	Compound	5	3	0.31	62.5
	Interrupted	10	3.3	0.38	71.4
SSR core repeats	Dinucleotide	22	3.2	0.40	49
	Trinucleotide	18	3.3	0.42	43.6
	Tetranucleotide	2	3.0	0.33	66.7
	Hexanucleotide	2	2.5	0.42	50
SSR length	<20 bp	21	3.0	0.39	38
	≥20	23	3.4	0.42	60

[a] *This defines the proportion of markers that were polymorphic out of all markers that amplified for each group described.*

3.6 Phenotypic data

3.6.1 Optimum inoculum concentration

The optimum inoculum concentration was determined by plotting data on disease incidence for various inoculum concentrations against number of days to wilt (Fig 3.14). The most reliable cut-off point was chosen at culture concentrations of 10^5 conidia ml^{-1} and day 17 of observation (Fig 3.14. E).

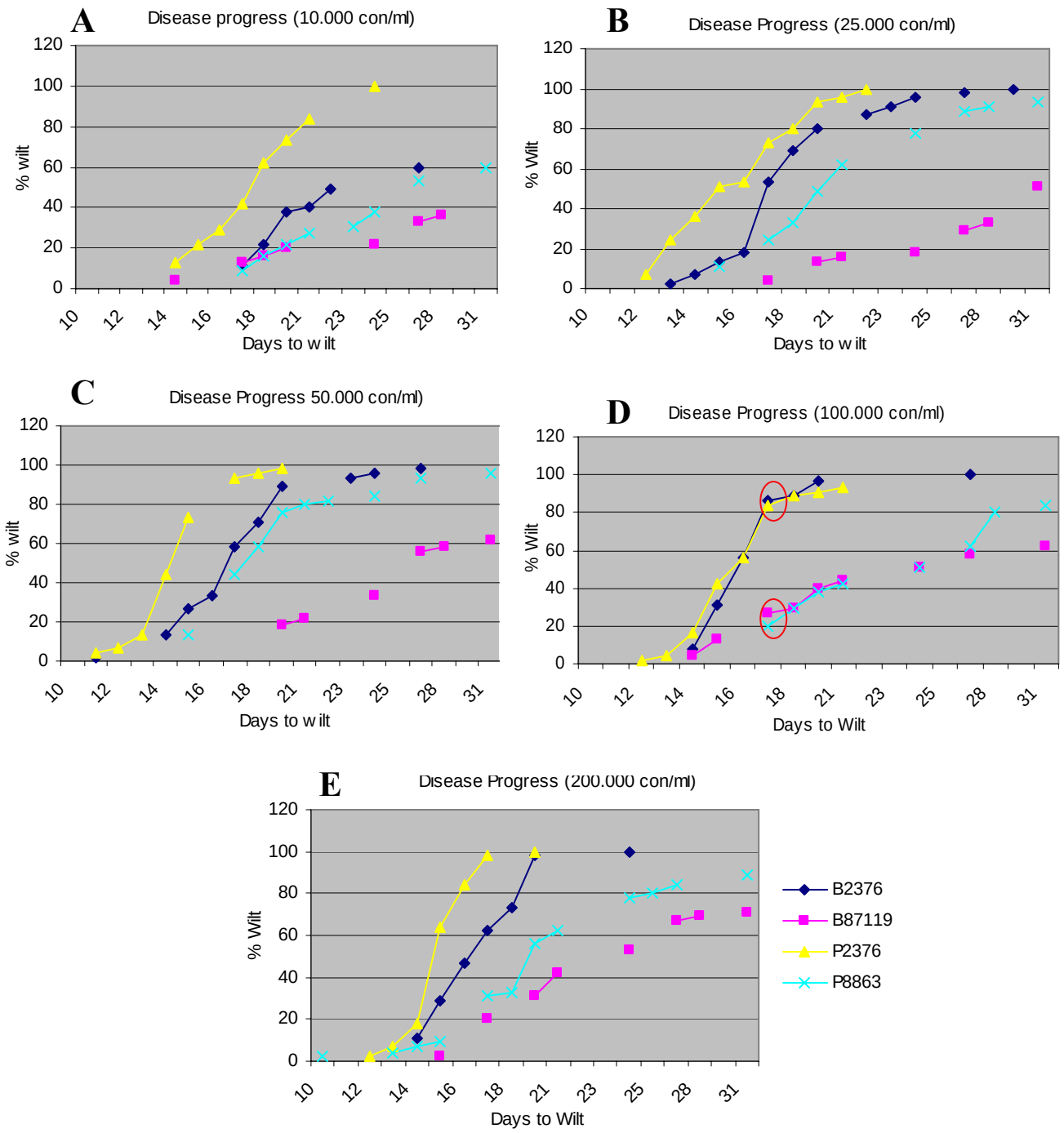


Figure 3.14 Optimisation data showing effects of different doses of *Fusarium udum* on wilt development in pigeonpea. The legend for the cultivars represented by different colours is shown in E. Day 17 of Figure D was chosen as the best cut-off day (disease intensity circled in red for susceptible and resistant parents) for taking observation while testing the mapping population of ICPL 87119 x ICP 2376.

On the 17th day after inoculation, about 20% of both the resistant sources wilted as compared to 80% of the susceptible varieties (Fig. 3.15). Observations would still be accurate for ICPL 8863 but not for ICPL 87119 at 2 x 10^5 conidia ml^{-1}, if observations were all taken by the 16th day. These results showed that the toxicity of the culture is not just influenced by fungal concentrations but also by the interactions between the fungus and the specific genotype used.

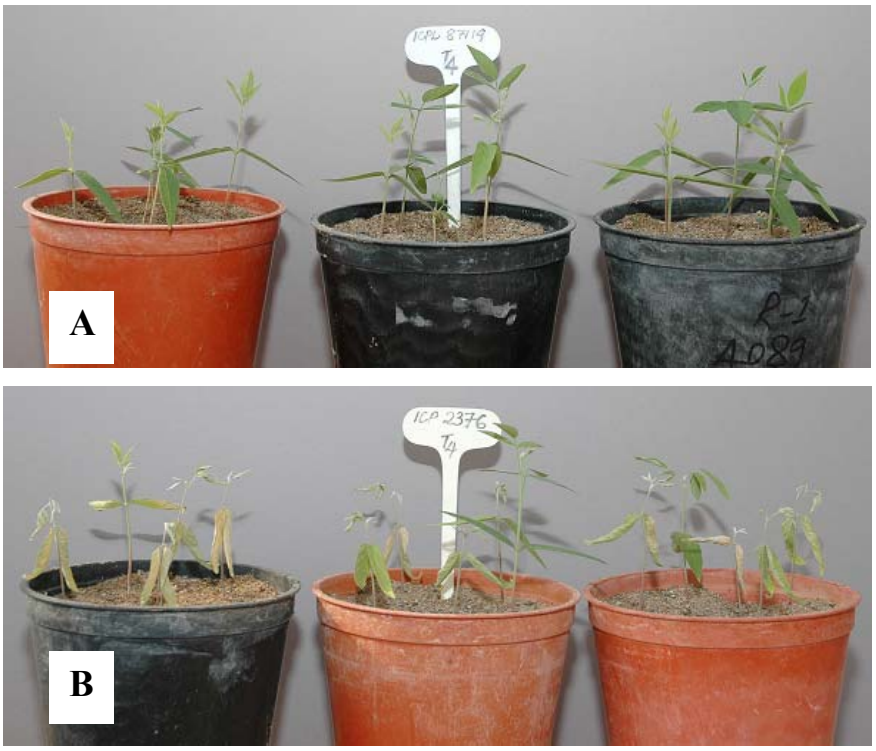


Figure 3.15 **A**. A picture showing the resistance of ICPL 87119 at 10^5 conidia ml^{-1} 17 days after inoculation. **B**. Most of the susceptible seedlings have wilted on the same day with similar inoculum level.

3.6.2 Hypothesised genetics of resistance to *Fusarium* wilt

Analysis of the F_2 population in the glasshouse indicated that *Fusarium* wilt is a quantitative trait. Although the sample size was small, a Chi-square analysis of F_2 favored the segregation of a quantitative trait (10:6 resistant to susceptible) more than that of a qualitative one (3:1 resistant to susceptible) (Table 3.11). The results suggested partial dominance with additive effect at both gene pairs. Though high homozygosity levels were expected at F_6, considerable variation was observed within families confirming the possible quantitative control of the trait. Chi square analysis of the F_6

population neither fitted the expected 1:1 segregation for a major gene confirming possible quantitative control (Table 3.11).

Table 3.11 Chi-square calculation of segregation ratios in F_2 and F_6 progenies

Generation	Observed		Expected								
			3:1			9:7			10:6		
	R	S	R	S	χ^2	R	S	χ^2	R	S	χ^2
F_2	220	115	251	84	15.3	188	147	12.4	209	126	1.5*
F_6 set I	45	76	60.5	60.5	7.94						

*R- Resistant; S-Susceptible. All χ^2 calculated at 1 df and 5% level of confidence is 3.84. *gives non-significant difference between observed and expected values. Cells in grey are not applicable.*

3.7 Recombinant Inbred Line (RIL) analysis

All RGAs and 8 out of 10 other pigeonpea microsatellites amplified well during optimisation but showed no polymorphism between the two parents of the *Fusarium* wilt mapping population. A total of 9 newly developed markers showed discrete difference (Fig. 3.16) between the 94 progenies and were scored. JoinMap analysis clustered 8 of the markers into 3 linkage groups (Table 3.12). A complete table showing all loci analysed and their recombination frequencies is shown in Appendix 6.3.

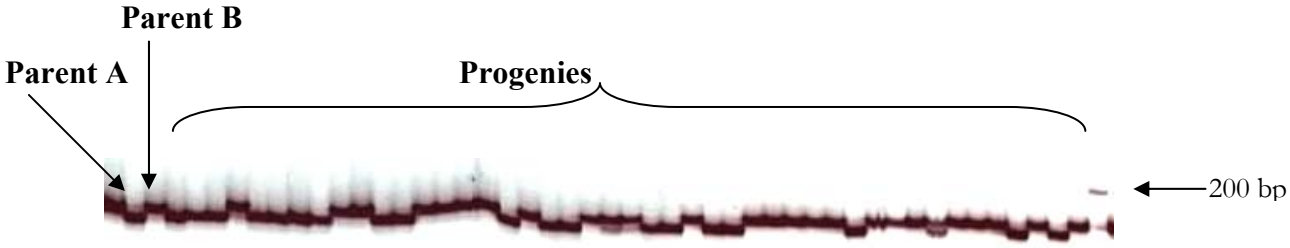
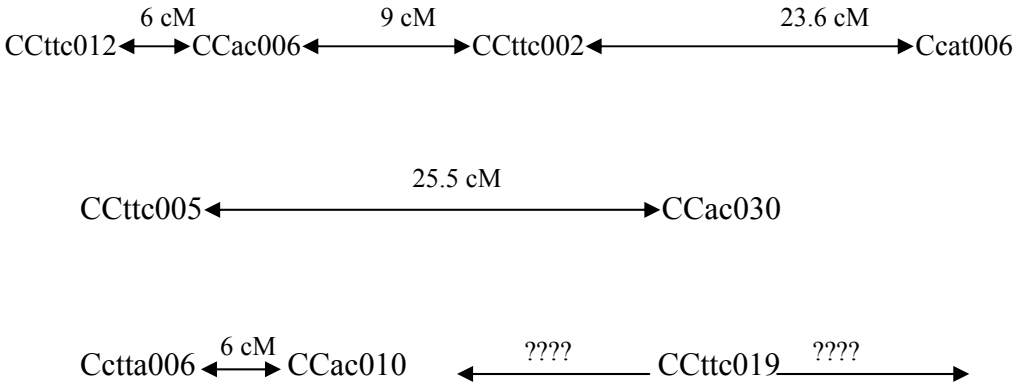


Figure 3.16 A gel picture showing segregation of progenies with respect to the parents. A is the resistant parent and B the susceptible one.

Table 3.12 Detected Linkage groups and corresponding recombination frequencies

Linkage Group	Locus 1	Locus 2	Recomb. Freq.	Mapping units (cM)	LOD
1	CCttc002	CCac006	0.09	9.0	10.67
	CCttc002	CCttc012	0.10	10.0	9.01
	CCttc002	CCat006	0.22	23.6	3.50
	CCac006	CCttc012	0.06	6.0	12.70
2	CCttc005	CCac030	0.23	25.5	3.04
3	CCtta006	CCac010	0.06	6.0	12.55

The loci studied could be theoretically localised as follows;



3.8 QTL analysis

No linkage was detected between the 9 SSR markers and the trait after performing a single marker analysis. This was to be expected since efficient and reliable mapping requires a saturated molecular map covering all the chromosomes. The 9 markers used here represented only a very small fraction of the pigeonpea genome.

4 DISCUSSION

4.1 Efficiency of pigeonpea microsatellite isolation

In this study, 93 new pigeonpea microsatellites have been developed using different methods with varying results. This is also the first report on attempts to utilise soybean microsatellites in pigeonpea. The results have demonstrated the essential need to make use of all tools at hand in attempts aimed at developing a larger quantity of markers, if marker assisted breeding is to be realised. Various lessons learnt from this work form the basis for future research and are therefore discussed in detail in this chapter.

When developing the non-enriched library for pigeonpea, 753 colonies were screened, 208 identified as positive while only 48 (23%) actually contained a microsatellite motif when sequenced. Whereas many traditional microsatellite isolation studies have reported an average of only 2.3% final positive clones (Zane et al. 2002) after sequencing, the average percent of false positives in such studies has been estimated at 48.7% (Squirrel et al. 2003). This study reports 6.4% clones containing a microsatellite but only after getting rid of 77% false positives. Even though the final proportion of microsatellite-containing clones appears to be high, the presence of such a high number of false positives after radioactive hybridisation is a matter of concern.

The short length of oligonucleotides used as probes is a possible reason for the resulting high proportion of false positives. The length of the probe applied for hybridisation can influence the type and length of microsatellites captured. Armour et al. (1994) found that longer oligonucleotide probes not only favoured the isolation of relatively long arrays but also highly eliminated mismatches. While the protocol in this study made use of oligonucleotides 28-30 bp long, other more successful studies (for example, Stajner et al. 2005) used much longer probes (200-550 bp). In other investigations (Lopes et al. 2002; Senda et al. 2004), 2 rounds of hybridisation have been performed to eliminate false positives. A second hybridisation is especially recommended if inefficient labeling of oligonucleotide probes (by using incorrect concentration of probes, any form of contamination, or inadequate hybridisation and wash steps) was suspected in the first round of screening. In pearl millet [*Pennisetum glaucum* (L.) R. Br.] (Budak et al. 2003) for example, a second round of hybridisation managed to discard more than 50% of clones (46 out of 80) as false positives.

However, the occurrence of a low number of positive clones is not an uncommon finding for non-enriched genomic libraries. Similar results have been reported (Winter et al. 1999; Ferguson et al. 2004) and several authors (Edwards et al. 1996; Panaud et al. 1996) have suggested enrichment of genomic libraries to enhance identification of microsatellites. Enrichment ratios published in the literature vary from 20-95% depending mainly on the procedure and method used (Stajner et al. 2005). Indeed, the enriched library prepared in this study resulted in a higher percentage of microsatellites (45%) but redundancy problems further reduced the actual percentage to 20. Problems of redundancy in enriched libraries have been reported in other investigations (Rallo et al. 2000; Mba et al. 2001) including pigeonpea (Burns et al. 2001). Some authors have argued that redundancy can be helpful in resolving ambiguities in a DNA sequence and that some duplicated clones with a variation in their nucleotide composition could be a source for the finding of Single Nucleotide Polymorphisms (SNPs) or for the detection of duplicated loci (Jones et al. 2001). Nevertheless, redundancy largely limits the number of effective microsatellite loci that can be developed, as is manifested in this study.

The overall redundancy in this library was 70%, which was higher than the average redundancy (37%) assessed for 71-plant microsatellite isolations (Squirrel et al. 2003) but comparable to those reported in other legumes including beans (70%) (Gaitán-Solis et al. 2002) and groundnut (67%) (He et al. 2003). The high level of redundancy found in this library is presumed to have arisen from the PCR amplification carried out after the affinity capture prior to cloning. This step was aimed at increasing the quantity of DNA, which is quite critical for successful recovery of clones. Billotte et al. (1999) have suggested the reduction in number of amplification cycles during PCR prior to cloning to reduce redundancy and chimeric clones.

4.2 Characteristics of pigeonpea SSR markers

The abundance of the TG/AC class of repeats in the enriched library and hence in the entire study was apparent. This result is consistent with those of previous studies in hop (*Humulus lupulus* L.) (Stajner et al. 2005), white clover (*Trifolium repens* L.) (Kölliker et al. 2001) and *Lolium temulentum* (Senda et al. 2004), for all of which enrichment was employed. Although database surveys have shown that TG/AC motifs are relatively rare

in plants (Wang et al. 1994), the high success of their isolation after enrichment in these studies may suggest that they occur at a significant frequency in the genomes of most plants.

In the absence of enrichment however, the most prevalent di-nucleotide repeat was AT even though it was not probed for. AT repeats have been reported to be the most abundant in plants (Morgante and Olivieri 1993; Cardle et al. 2000; Morgante et al. 2002) but previous screening studies (Nunome et al. 2003) have excluded them due to the problems with self-complementarity and difficulties in amplification (Su et al. 1996; Temnykh et al. 2001). Studies in sunflower (*Helianthus annuus* L.) that involved enrichment of di-nucleotide repeats also identified AT repeats as the most abundant non-target SSRs (Tang et al. 2002).

The TC class of repeats was the next most abundant di-nucleotide repeat in both the enriched and non-enriched libraries. In other legumes (Ferguson et al. 2004; Lichtenzveig et al. 2005) TC repeats have been reported as the most abundant di-nucleotide repeat even though in both cases AT repeats were not probed for. In a study analysing wheat (*Triticum aestivum* L.), rice (*Oryza sativa* L.), maize (*Zea mays*) and soybean ESTs, TC repeats were identified as the most frequent di-nucleotide repeat in general (Gao et al. 2003) suggesting that TC repeats could be the most abundant di-nucleotide repeats in coding regions of most plant genomes. Most AT repeats on the other hand have been reported to reside in non-coding regions (Temnykh et al. 2001). No GC repeats were isolated, consistent with reports that these repeats are extremely rare in most genomes (Cui et al. 2005). Lower frequencies of GC repeats have been attributed to methylation of cytosine which in turn increases the chance of mutation to thymine by deamination (Schorderet and Gartlar 1992).

The most abundant tri-nucleotide repeats were TAA/ATT and GAA/CTT in the non-enriched and enriched libraries respectively. ATT has been reported to be the most common tri-nucleotide motif in other legumes including soybean (Akkaya et al. 1995; Cregan et al. 1999), groundnut (Ferguson et al. 2004) and chickpea (Udupa et al. 1999; Lichtenzveig et al. 2005). A previous study in *Arabidopsis* (Loridon et al. 1998) reported high proportions of GAA repeats. Song et al. (2005) noted the superiority of TAA motifs to all other tri-nucleotide repeats for the successful development of microsatellite markers in wheat. The results of this study are therefore consistent with

those of other investigators including Lagercrantz et al. (1993) and Morgante and Olivieri (1993), who reported that TAA and GAA are the most frequent tri-nucleotide repeat motifs in plants.

Despite being the most abundant tri-nucleotide repeat in the non-enriched library, TAA enrichment was the least successful yielding only 5 positive clones. A similar study in common bean (Gaitán-Solis et al. 2002) reported the absence of TAA repeats in sequenced clones despite its enrichment. They suggested that this failure could have to do with selective enrichment of specific microsatellite sequences. However, the high success in enrichment of similar motifs in a separate study in the same plant (common bean) using similar enrichment (Métais et al. 2002) strategy invalidates this point.

The failure of TAA enrichment in these studies could be attributed to the low melting temperature of these repeats and their likely association with interspersed repetitive elements (Temnykh et al. 2001). Song et al. (2005) reported that the efficiency of ATT marker development could be increased by creating libraries from sheared rather than enzyme digested DNA fragments as well as the arbitrary addition of a clamp such as GCG, CCC, or GGG to the 5' end of one or both primers.

The abundance of AT-based motifs (AT and ATT) in the non-enriched library could have been a result of the restriction enzyme used to size-fractionate the library. The choice of restriction endonucleases has been shown to affect the microsatellite content of genomic libraries (Billotte et al. 1999; Hamilton and Fleischer 1999). A similar effect was observed by Kölliker et al. (2001). *Sau*3A1 (▼GATC) is methylation-sensitive and is expected to be intermediate in its target unlike *Tsp*5091 (▼AATT). DNA methylation provides epigenetic modifications that help to regulate genome imprinting, gene expression and DNA repair. The advantage of using a methylation sensitive enzyme such as *Sau*3A1 is that it would cut in hypomethylated gene-rich DNA producing libraries that are enriched for low copy sequences (Moretzsohn et al. 2005).

Efficient digestion of genomic DNA for library preparation could also be inhibited by the presence of polysaccharides (Reiter 2002) and/or DNA methylation (Finnegan et al. 1998). To avoid biases that may be caused by enzymes, physical methods and particularly sonication (Deininger 1983) has been recommended for its even coverage of the genome. However, sonication also suffers problems of

irreproducibility, requires relatively large amounts of DNA and yields a broad size distribution with a low yield of fragments useful for cloning and sequencing. The use of more than one restriction enzyme in library construction has been suggested (Kölliker et al. 2001) to result in maximum yield of potentially useful SSRs and to support even distribution of SSRs across the genome.

Although di-nucleotide repeats are the most commonly used class of microsatellite markers in plants, there has been concern over the difficulty experienced in genotyping them due to a high frequency of strand slippage artefacts (Levinson and Gutman 1983). In contrast, tri- and tetra-nucleotide repeat based markers have been shown to produce a higher proportion of discrete PCR products due to reduced level of stuttering (Gastier et al. 1995). Tri-nucleotide repeats have also been of particular interest since AAG together with AAT have been found to be common in intron sequences (Smulders et al. 1997). In human, these motifs have been found in genes that are associated with several neurological diseases such as the X syndrome (Jin and Warren 2000), Huntingdon's disease and several forms of ataxia (Sermon et al. 2001) and myotonic dystrophy (Timchenko et al. 2001). These disorders are caused by excessive expansions of triplet repeats located near or within coding regions (Fujimori et al. 2003).

Most of the motifs isolated were short, a third of which were interrupted. This probably had to do with the short oligonucleotide probes (Section 4.1) used both in screening of the non-enriched library and in selective hybridisation of the enriched library. It has been suggested that short probes for hybridisation tend to select fragments with shorter interrupted repeats. Stajner et al. (2005) used long probes (200-550 bp) and more than 50% of isolated clones had a microsatellite length longer than 16 repeat units with only 2% of clones having less than 5 repeat units. Short repeats have been reported to produce monomorphic PCR products or displayed very low PIC value (Qi et al. 2001) and this is partially true for this study (Section 4.3). Other studies have also reported that the proportion of interrupted repeats is related to the type of enrichment used (Van de Wiel et al. 1999) and the nature of repeat isolated (Milbourne et al. 1998).

About 39% of all primers designed failed to amplify despite the efforts on optimisation. The abundance of AT-based motifs in the non-enriched library is highly likely the cause of much lower amplification rates compared to enriched library.

Palacios et al. (2002) have reported a modified selective amplification technique for AT-based repeats that could enhance their amplification. Poor amplification can also be caused by the divergence in the sequences flanking SSRs thus creating null alleles (Smulders et al. 1997). Null alleles are presumably caused by DNA polymorphisms in primer sites (Tang et al. 2002) especially in non-coding regions (Mogg et al. 2002). This phenomenon was unlikely responsible for the lower amplification success since interrupted SSRs, which are believed to be degraded perfect SSRs (Taylor et al. 1999), showed a higher rate of amplification (68%) than the perfect ones (59%). Null alleles could also be caused by the deletion of a microsatellite at a specified locus (Callen et al. 1993), presence of large introns, deriving primers from chimeric clones or the possible extension of one or both primers across a splice site (Varshney et al. 2005).

Unsuitable primer pair sequences and/or improper PCR conditions (Akagi et al. 1996) is another possible cause of amplification failure. This could have hardly been the case since the sequences used were relatively long, especially in the non-enriched library where the percentage of amplification was much lower (51%). The sequencing was also done in both directions in order to increase the quality of sequences for primer design after alignment. The correct PCR conditions were established through optimisation although Squirrel et al. (2003) have indicated that optimisation in itself could be a source of attrition. Other investigations have pointed out the fact that PCR optimisation may lead to insignificant improvement in band interpretability (Ashworth et al. 2004).

This study observed that varying “touchdown” programs greatly improved band clarity and hence eased scoring. However, the 5 reactions used for optimisation could be further reduced to three without significant reduction in quality of bands. It did not appear to add any value when increasing primer and enzyme concentrations compared to increasing the DNA concentration. The advantage of being able to optimize down the enzyme and primer concentrations has great cost reduction implications for similar projects in the future.

4.3 Microsatellite allelic diversity in pigeonpea

The allelic variation observed at microsatellite loci varies greatly between different species. The average number of alleles per locus in cultivated pigeonpea genotypes was 3.40 and 3.14 for library A and B respectively in cultivated pigeonpea and 4.8 with inclusion of wild relatives. A previous diversity analysis of cultivated pigeonpea species reported an average of 3.10 for 10 polymorphic loci (Burns et al. 2001), which is similar to the present results. This is lower than in other legume crops such as soybean (11-26 alleles per locus) (Rongwen et al. 1995) and common bean (6 alleles per locus) (Gaitán-Solis et al. 2002) but comparable in groundnut (average 4.25 alleles per locus) (He et al. 2003). The low genetic variability in cultivated species compared to wild relatives has also been observed in other plants (Nunome et al. 2003; Moretzsohn et al. 2004; Zhao et al. 2005) suggesting changes in allele frequencies at specific loci due to natural and artificial selection.

In the non-enriched library, 98 different alleles were observed for the 20 loci examined with over 50% of the alleles (56) unique to wild species while in library B, 7 out of 10 accessions harboured unique alleles. The number of unique alleles is a simple measure of genetic distinctiveness (Kalinowski 2004) and this is clearly reflected within the wild species from the dendrogram generated (Fig. 3.11). This analysis showed that while some alleles have been lost in the process of evolution, selection and breeding, new ones have entered the germplasm constitution of modern varieties. Future studies will therefore need to consider the contribution of single individuals to the overall genetic diversity level, and to evaluate whether there still exists a dynamic balance of diversity.

The average PIC among di-nucleotide and tri-nucleotide repeats were 0.48 and 0.33 respectively with the inclusion of wild pigeonpea genotypes in the analysis. There are similar reports of higher levels of polymorphism for di-nucleotide repeats than tri-nucleotide repeats for common bean (Yu et al. 2000) and in avocado (*Persea americana* Mill.) (Ashworth et al. 2004) but not for soybean where tri-nucleotide repeats were more polymorphic than di-nucleotide repeats (Rongwen et al. 1995). It is important to note that there was no significant variation in the average PIC among di- (0.40) and tri-nucleotide repeats (0.42) when cultivated species alone were considered. This suggests

that allelic variation with respect to various classes of motifs depends on available genetic diversity within the genotypes studied.

Di-nucleotide repeats have been reported to reside outside of coding regions of genes (Temnykh et al. 2001) and are characterized by more repeat numbers (Li et al. 2004) making them the best source of highly polymorphic SSR markers. In contrast, tri-nucleotide repeats are more abundant in protein coding regions (Tóth et al. 2000; Subramanian et al. 2003) with relatively small repeat numbers (Thiel et al. 2003). Despite their frequent polymorphism, the use of di-nucleotide repeats may be limited by the occurrence of stutter bands in the amplification products that may lead to incorrect scoring of alleles (Ashworth et al. 2004; Wang et al. 2005).

An increase in the level of polymorphism with increasing SSR array length was observed. Although a higher proportion (60%) of microsatellites with a length of at least 20 bp were polymorphic as compared to a mere 38% among shorter repeats, the differences in PIC were rather small (0.39 and 0.44 respectively). This was possibly due to the small range of data analysed. While some studies (Budak et al. 2003) have discarded SSRs with less than 20 bp in the past, some of the shortest SSRs (12 bp) in the current study were also polymorphic detecting up to 4 alleles with a PIC as high as 0.68. In common bean, no clear relationship was observed between repeat length and degree of polymorphism (Yu et al. 2000). The same study in common bean detected polymorphism of even shorter repeats (10 bp long) than the ones included in the present investigation. This demonstrates that shorter SSR sequences are also important sources for developing polymorphic SSR markers in plants. Theoretically, the number of repeats is correlated with the mutation rate (Li et al. 2002). This means that more recently evolved microsatellites would have fewer polymorphisms because of fewer occasions for mutations even if they have longer repeats (Budak et al. 2003).

4.4 Comparative genomics application

Primers designed for cultivated pigeonpea were able to amplify the DNA of wild pigeonpea genotypes. All the genotypes tested were from the same genus and from primary or secondary gene pool except *C. platycarpa*, which was from the tertiary gene pool. Conservation of regions flanking microsatellite repeats in legumes has been shown within *Phaseolus* (Guerra-Sanz 2004), *Arachis* (Moretzsohns et al. 2004),

Glycine (Zou et al. 2004), *Cicer* (Choumane et al. 2004) and also across leguminous species (Wang et al. 2004) confirming their importance for the design of locus-specific primers (Pandian et al. 2000).

The high amplification success rate of primers among wild relatives of pigeonpea is noteworthy because all these species have been shown to contain special traits (Section 1.2) that could be introgressed into the pigeonpea genome. The possibility of using SSR markers developed for one species in genetic evaluation of other species greatly reduces the cost of analysis since the development of microsatellite markers is still expensive and time consuming (Moretzsohns et al. 2004). The SSR markers developed in this study are potentially very useful for genetic analysis of wild species of pigeonpea including comparative genome mapping, population genetic structure and phylogenetic inferences among species.

Out of 220 soybean primers tested, 39 amplified pigeonpea DNA. Although successful amplification alone does not guarantee the presence of the repeat motif within the sequence (Decroocq et al. 2003), potential transferability of microsatellite markers has been demonstrated across several species (Peakall et al. 1998; Pandian et al. 2000; Kuleung et al. 2004). According to Choumane et al. (2004), the generation of amplification products from a defined locus requires at least, that the 3' terminal nucleotides of the target sequence are perfectly complementary to the primers. Amplification across genera boundaries implies that the respective loci are conserved between the various species in question. The low amplification percentage in this case could be indicative not only for the evolutionary distance between pigeonpea and soybean but also for the rate of evolution of the genomic sequences where the primers are located (Choumane et al. 2004).

So far, no studies have been done to confirm the level of synteny between soybean and pigeonpea. A review by Young et al. (2003) clustered soybean close to pigeonpea making soybean the closest intensively studied legume to pigeonpea and hence the best target in comparative genome mapping. This is especially because of the rich availability of information on soybean genetics at the molecular level. Higher amplification of soybean SSRs in pigeonpea and better polymorphism was therefore expected than was the outcome in this study. Palop et al. (2000) observed similar discouraging results while working on *Limonium* (*Plumbaginaceae*). The choice of SSR

primers could be a contributing factor, since most of the soybean SSR primers tested for cross amplification in pigeonpea were from genomic sequences.

Genomic SSR markers have been reported to be biased towards genome specificity and generally do not transfer well to other species (La Rota 2005). In contrast, genic SSRs are transferable among distantly related species with reasonably high polymorphism levels (Varshney et al. 2005). Accordingly, the proportion of soybean EST-derived microsatellites that amplified (25.4%) was higher than that of genomic microsatellites (9.2%). The current study optimised all of the soybean primers in pigeonpea perhaps with negative effects. Although modifying the PCR protocol may increase transferability, false positives can appear and the fragments amplified may not be the ones expected. Decroocq et al. (2003) recommended no optimisation of the new primers into the new plant while testing amplification across taxa.

Beyond the importance of genomic sequences for use in SSR identification, this study has demonstrated the possible utilisation of such sequences to search for syntenic regions amongst well-annotated databases. Genomic sequences have the advantage of identifying genes that are expressed at very low levels, or those that may have rare mRNAs but still important as key transcriptional regulators or developmental sensors (Whitton et al. 2004). They can also be searched for ORFs or signals for regulation (Tyagi et al. 2004). This study used a relatively high expect (e) value ($\leq 10^{-10}$) compared to that used in other studies for example $\leq 10^{-8}$ (Strong and Nelson 2000) and $\leq 10^{-5}$ (Buhariwalla et al. 2005; Rensink et al. 2005). The criteria based on length alignment and percent identity was included to strengthen the results. A similar approach has been reported elsewhere (Salse et al. 2002; Rensink et al. 2005).

The institute for genomic research (TIGR) soybean database resulted in most of the "best hits", suggesting that soybean could be the closest relative of pigeonpea compared to all the crops of which databases were searched. *Arabidopsis* database searches gave more significant hits than the TIGR *Medicago* database even though the latter is a legume. The major reason for the higher number of hits in *Arabidopsis* could be the larger and more advanced database as compared to *Medicago's* even though recent studies (Mudge et al. 2005) have also reported high levels of synteny (up to 60%) between other legumes and *Arabidopsis*.

Putative repetitive elements were found in high numbers confirming reports on their abundance in nuclear DNA content (Kumar and Bennetzen 1999). In barley, 41% of 290 genomic clones were found to show direct or indirect association with known repetitive elements (Ramsay et al. 1999). Retrotransposons are potentially useful for gene tagging as they generate stable insertions and integrate into unlinked sites (Ramachandran and Sundaresan 2001). They have been used in *Vitis vinifera* L. for genotyping (Labra et al. 2004); in *Medicago* for mutagenesis (d'Efurth et al. 2003) and in mammalian systems to resolve deep rooted phylogenetic relationships (Shimamura et al. 1997). The potential use of retrotransposons adds a powerful tool for the study of pigeonpea genome.

A total of 21 putative type I SSRs were identified, all of which were short with repeats of less than 10. Identification of SSRs in gene sequences of plant species has apparently been carried out as early as 1993 by Morgante and Olivieri (Varshney et al. 2005). Most type I SSRs however, have been identified by searching public EST databases. This study applied the principle that genomic SSRs are found in both coding and non-coding regions and worked backwards to identify putative genic SSRs. Yaish and de la Vega 2003 followed a similar procedure and reported a polymorphic microsatellite locus linked to a MADS box gene in common bean.

The repeat numbers and total lengths of SSRs in protein coding regions or ORFs are relatively small (Thiel et al. 2003) with most studies reporting the abundance of tri-nucleotide repeats (Töth et al. 2000; Gao et al. 2003; Li et al. 2004) followed by dinucleotide repeats. Such EST-SSRs have been shown to be highly robust and produce high quality bands (no stuttering) (Varshney et al. 2005). Our results, though indicative of the possible location of these putative type I SSRs, will not be conclusive until functional studies are carried out to confirm the functions of the respective orthologues in pigeonpea. Nevertheless, these results are a starting point for other more comprehensive studies on functional and comparative genomics in pigeonpea. The putative genes would be useful also in the development and validation of functional markers. Functional marker development requires allele sequences of functionally characterised genes from which polymorphic, functional motifs affecting plant phenotype can be identified. Such markers, once developed, are superior to random DNA markers (Andersen and Lübberstedt 2003).

This study made a first attempt to introduce comparative genome mapping in pigeonpea as this appears to be the fastest way forward for future molecular breeding in pigeonpea. Despite the little work that has been done on this crop, comparative mapping offers possibilities of combining the research efforts in other closely related species as has been done in grasses (Van Deynze et al. 1995) and between *Arabidopsis* and *Brassica* spp (Parkin et al. 2005).

4.5 Genetic diversity in pigeonpea germplasm

Early studies on phylogenetic relationships in pigeonpea were examined by use of morphology (Remanandan et al. 1988), protein and isozyme analyses (Ladizinsky and Hamel 1980; Krishna and Reddy 1982), RAPDs (Ratnaparkhe et al. 1995) and RFLP mtDNA markers (Sivaramakrishnan et al. 2002). In our study, 19 SSRs detected genetic distance between various genotypes of the genus *Cajanus*. Many cluster procedures have been reported for molecular data analysis (Lubbers et al. 1991; Grabau et al. 1992; Kresovich et al. 1994; Griffin and Palmer 1995; Gizlice et al. 1996; Thompson et al. 1998). This study used the hierarchical clustering method of Unweighted Paired Group Method using Arithmetic Averages (UPGMA) for visualising relationships in the data. Additionally, the Multidimensional Scaling (MDS) procedure was used to confirm the various groupings since a single clustering method might not always be optimal or effective in revealing genetic associations (Mohammadi and Prasanna 2003).

In the first set of diversity study that included wild species, the two clustering procedures confirmed the same results. The smallest genetic distance and hence the largest similarity coefficient among the various species of *Cajanus* studied was between *C. cajanifolius* and cultivated pigeonpea. This revealed a very close relationship between the 2 species consistent with the results of Sivaramakrishnan et al. (2002) using mtDNA RFLP analysis. *C. cajanifolius* has been reported as the nearest wild relative of pigeonpea (Van der Maesen 1990). Morphologically, *C. cajanifolius* resembles pigeonpea in all traits except the presence of a prominent strophiole (Van der Maesen 1990).

Within the landraces, ICP 7543 and ICP 14144 showed the highest genetic similarity (98%) even though the 2 landraces were collected from India and Jamaica respectively. High similarities (94%) were also detected between ICP 12058 and ICP

13092, which are from Tanzania and Kenya respectively. This level of similarity could be indicative of the presence of genetic redundancy in the collections. If these accessions are indeed genetically identical, they could be pooled with no loss in the overall amount of genetic variation. At least 13,548 pigeonpea accessions mainly from India, Africa, Australia and the Caribbean are being maintained at the ICRISAT gene bank in Hyderabad, India. Such a collection is numerically impressive but can be financially wasteful with a lot of genetic redundancy if poorly characterised. This study therefore calls for the need for more comprehensive assessments of genetic redundancy which should lead to a more efficient conservation and use of these resources.

Although the number of accessions sampled was small and not fully representative of total available diversity within each region, the results also support the theory that the South American landraces were likely introduced from Africa whereas the African gene pool was probably derived from relatively few introductions from South India. Identifying a genetic structure within pigeonpea germplasm is usefull for establishing strategies for sampling and managing germplasm. Obtaining genetic patterns from molecular analyses such as this in future studies will help in managing the germplasm and assist breeders make better choices when selecting among the large numbers of accessions available. For a precise analysis of intra- and inter-specific microsatellite variability in pigeonpea, more accessions should be analysed from a wider range of species.

4.6 Mapping options – towards MAS in pigeonpea

One aim of this study was to detect QTL for resistance to *Fusarium* wilt. Recombinant Inbred Lines (RIL) instead of an F_2 population were used because they allow a high resolution map due to higher chances for recombination. Despite the efforts, no successful mapping results could be obtained, partly due to the inadequate number of polymorphic markers generated from this study, and partly due to the wrong sampling procedures followed. The successful application of molecular markers in trait mapping is greatly associated by the availability of a genetic map for the species under investigation.

While successful mapping attempts could still be carried out in the absence of a linkage map as has been done in chickpea (Chandra et al. 2003) one would require a

representative number of individuals (Deorge 2001). Unfortunately, this study failed to synchronise the correct individuals for phenotyping and genotyping leaving only 45 individuals with common data. Furthermore, of the 44 SSR loci that were polymorphic among the cultivated germplasm, only 9 were polymorphic for the 2 parental lines of the *Fusarium* wilt population. It was not possible to detect a QTL with this small sample size and inadequate markers. Future studies need to focus on more individuals as well as the incorporation of other markers such as AFLPs.

4.7 Concluding remarks

Plant breeders have become increasingly interested in marker assisted selection for efficient and precise transfer of genes conditioning important agronomic traits. The current results clearly demonstrate that pigeonpea SSRs constitute efficient sources of molecular markers for both cultivated and other *Cajanus* species. The current interest in genetic potential of wild relatives will benefit from the use of SSR markers in identification of the most closely related parents for inter-specific crossing and also in identifying new sources of male sterility. With the ongoing breeding emphasis on development of hybrid pigeonpea, MAS would lead to a quick and efficient way of predicting and identifying inbred lines that can produce highly heterotic hybrids precisely. Pigeonpea germplasm collection at ICRISAT is already benefiting from the current study by utilising the markers developed here to characterise a representative collection. The markers developed here will be beneficial in future for comparative genome analysis between the different *Cajanus* species for more efficient exploitation of the desirable characteristics therein.

A larger number of markers would still be required in future to enable MAS in pigeonpea. The construction of large-insert bacterial artificial chromosome (BAC) libraries, as has been done in chickpea (Lichtenzveig et al. 2005), will be necessary in pigeonpea for their potential wide genome coverage. The use of more than one restriction enzyme in library construction as well as targeting longer motifs are other options likely to maximise the yield of potentially useful SSRs across the genome. With the current efforts to make DArT technology available in pigeonpea (Huttner et al. 2005) and the falling prices in DNA sequencing and SNP assays (Rafalski 2002), more

superior markers will undoubtedly be incorporated to complement the current efforts and enhance molecular marker technology in pigeonpea.

5 REFERENCES

Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W. and Lipman D.J. 1997. Gapped blast and PSI-blast: a new generation of protein database search programs. Nucl. Acids Res. 25: 3389-3402.

Adu-Gyamfi J.J., Fujita K. and Ogata S. 1990. Phosphorus fractions in relation to growth in pigeonpea (*Cajanus cajan* (L) Millsp.) at various levels of P supply. Soil Sci. Plant Nutr. 36: 531-543.

Akagi H., Nakamura A., Yokozeki-Misono Y., Inagaki A., Takahashi H., Mori K. and Fujimura T. 2004. Positional cloning of the rice *Rf-1* gene, a restorer of BT-type cytoplasmic male sterility that encodes a mitochondria-targeting PPR protein. Theor. Appl. Genet. 108: 1449-1457.

Akagi H., Yokozeki Y., Inagaki A. and Fujimura T. 1996. Microsatellite DNA markers for rice chromosomes. Theor. Appl. Genet. 93: 1071-1077.

Akkaya M.S., Shoemaker R.C., Specht J.E., Bhagwat A.A. and Cregan P.B. 1995. Integration of simple sequence repeat DNA markers into a soybean linkage map. Crop Sci. 35: 1439-1445.

Andersen J.R. and Lübberstedt T. 2003. Functional markers in plants. Trends Plant Sci. 8: 554-560.

Armour J.A., Neumann R., Gobert S. and Jeffreys A.J. 1994. Isolation of human simple repeat loci by hybridization selection. Hum. Mol. Genet. 3: 599-565.

Asensio-S.-Manzanera M.C., Asensio C. and Singh S.P. 2005. Introgressing resistance to bacterial and viral diseases from the middle American to Andean common bean. Euphytica 143: 223-228.

Ashworth V.E.T.M., Kobayashi M.C., De La Cruz M. and Clegg M.T. 2004. Microsatellite markers in avocado (*Persea americana* Mill.): development of dinucleotide and trinucleotide markers. Sci. Hortic. 101: 255-267.

Attwood T.K. 2000. The quest to deduce protein function from sequence: the role of pattern databases. Int. J. Biochem. Cell Biol. 32: 139-155.

Bennetzen J.L., Coleman C., Liu R., Ma J. and Ramakrishna W. 2004. Consistent over-estimation of gene number in complex plant genomes. Curr. Opin. Plant Biol. 7: 732-736.

Billotte N., Lagoda P.J.L., Risterucci A.-M. and Baurens F.-C. 1999. Microsatellite-enriched libraries: applied methodology for the development of SSR markers in tropical crops. Fruits 54: 277-288.

Bink M.C.A.M. and Meuwissen T.H.E. 2004. Fine mapping of quantitative trait loci using linkage disequilibrium in inbred plant populations. Euphytica 137: 95-99.

Blair M.W., Pedraza F., Buendia H.F., Gaitán-Solís E., Beebe S.E., Gepts P. and Tohme J. 2003. Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.). Theor. Appl. Genet. 107: 1362-1374.

Botstein D., White R.L., Skolnick M., Davis R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am. J. Hum. Genet. 32: 314-331.

Budak H., Pedraza F., Cregan P.B., Baenziger P.S. and Dweikat I. 2003. Development and utilization of SSRs to estimate the degree of genetic relationships in a collection of pearl millet germplasm. Crop Sci. 43: 2284-2290.

Buhariwalla H.K., Jayashree B., Eshwar K. and Crouch J.H. 2005. Development of ESTs from chickpea roots and their use in diversity analysis of the *Cicer* genus. BMC Plant Biol. 5:16 doi:10.1186/1471-2229-5-16

Buhariwalla H.K., Crouch J.H. 2004. Optimisation of marker screening protocol to assess the degree and distribution of genetic diversity in landraces of pigeonpea. In: Bramel P.J. (ed), Assessing the risk of losses in biodiversity in traditional cropping systems. ICRISAT, Patancheru, pp. 57-74.

Burns M.J., Edwards K.J., Newbury H.J., Ford-Lloyd B.V. and Baggott C.D. 2001. Development of simple sequence repeat (SSR) markers for the assessment of gene flow and genetic diversity in pigeonpea (*Cajanus cajan*). Mol. Ecol. Notes 1: 283-285.

Bussell J.D., Waycott M., Chappill J.A. 2005. Arbitrarily amplified DNA markers as characters for phylogenetic inference. Perspect. Plant Ecol. Evol. Systemat. 7: 23-26.

Callen D.F., Thompson A.D., Shen Y., Philips H.A., Richards R.I., Mulley J.C. and Sutherland G.R. 1993. Incidence and origin of "null" alleles in the (AC)n microsatellite markers. Am. J. Hum. Genet. 52: 922-927.

Cardle L., Ramsay L., Milbourne D., Macaulay S.M., Marshall D and Waugh R. 2000. Computational and experimental characterisation of physically clustered simple sequence repeats in plants. Genetics 156: 847-854.

Chambers G.K. and MacAvoy E.S. 2000. Microsatellites: consensus and controversy. Comp. Biochem. Physiol. B 126: 455-476.

Chandra S., Buhariwalla H.K., Kashiwagi J., Hari Krishna S., Rupa Sridevi K., Krishnamurthy L., Serraj R. and Crouch J.H. 2004. Identifying QTL-linked markers in marker-deficient crops. 4th International Crop Science Congress, Brisbane, Australia (http://www.regional.org.au/au/cs/2004/poster/3/4/1/795_chandras.htm)

Charcosset A., Moreau L. 2004. Use of molecular markers for the development of new cultivars and the evaluation of genetic diversity. Euphytica 137: 81-94.

Chen X., Temnykh S., Xu Y., Cho Y.G. and McCouch S.R. 1997. Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza staiva* L.). Theor. Appl. Genet. 95: 553-567.

Choumane W., Winter P., Baum M. and Kahl G. 2004. Conservation of microsatellite flanking sequences in different taxa of *Leguminosae*. Euphytica 138: 239-245.

Cobb B.D. and Clarkson J.M. 1994. A simple procedure for optimising the polymerase chain reaction (PCR) using modified Taguchi methods. Nucleic Acids Res. 22: 3801-3805.

Collard B.C.Y., Jahufer M.Z.Z., Brouwer J.B. and Pang E.C.K. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142: 169-196.

Cregan P.B., Jarvik T., Bush A.L., Shoemaker R.C., Lark K.G., Kahler A.L., Kaya N., VanToai T.T., Lohnes D.G., Chung J. and Specht J.E. 1999. An

integrated genetic linkage map of the soybean genome. Crop Sci. 39: 1464-1490.

Cui L., Fan Q., Hu Y., Karamycheva S.A., Quackenbush J., Khuntirat B., Sattabongkot J. and Carlton J.M. 2005. Gene discovery in *Plasmodium vivax* through sequencing of ESTs from mixed blood stages. Mol. Biochem. Parasit. 144: 1-9.

Dayal S., Lavanya M., Devi P. and Sharma K.K. 2003. An efficient protocol for shoot regeneration and genetic transformation of pigeonpea (*Cajanus cajan* (L.) Millsp.) using leaf explant. Plant Cell Rep. 21: 1072-1079.

Decroocq V., Favé M.G., Hagen L., Bordenave L. and Decroocq S. 2003. Development and transferability of apricot and grape EST microsatellite markers across taxa. Theor. Appl. Genet. 106: 912-922.

Deininger P.L. 1983. Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. Anal. Biochem. 129: 216-223.

d'Erfurth I., Cosson V., Eschstruth A., Lucas H., Kondorosi A. and Ratet P. 2003. Efficient transposition of the *Tnt1* tobacco retrotransposon in the model legume *Medicago truncatula.* Plant J. 34: 95-106.

Don R.H., Cox P.T., Wainwright B.J., Baker K. and Mattick J.S. 1991. "Touchdown" PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res. 19: 4008

Edwards K.J., Barker J.H.A, Daly A., Jones C. and Karp A. 1996. Microsatellite libraries enriched for several microsatellite sequences in plants. Biotechniques 20: 758-759.

Falconer D.S. and Mackay T.F.C. 1996. Introduction to quantitative genetics. 4th edn. Longman Group, Essex, 464 pp.

FAOSTAT 2005. http://faostat.fao.org/faostat/collections?version=ext&hasbulk=0&subset=agriculture.

Ferguson M.E., Burrow M.D., Schulze S.R., Bramel P.J., Paterson A.H., Kresovich S. and Mitchell S. 2004. Microsatellite identification in peanut (*A. hypogea* L.). Theor. Appl. Genet. 108: 1064-1070.

Finnegan E.J., Genger R.K., Peacock W.J. and Dennis E.S. 1998. DNA methylation in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 49: 223-247.

Fujimori S., Washio T., Higoc K., Ohtomod Y., Murakamid K., Matsubarad K., Kawaie J., Carnincie P., Hayashizakie Y., Kikuchic S. and Tomitaa M. 2003. A novel feature of microsatellites in plants: a distribution gradient along the direction of transcription. FEBS Lett. 554: 17-22.

Fujita K., Kai Y., Takayanagi M., El-Shemy H., Adu-Gyamfi J.J. and Mohapatra P.K. 2004. Genotypic variability of pigeonpea in distribution of photosynthetic carbon at low phosphorus level**.** Plant Sci. 166: 641-649.

Gao L., Tang J., Hongwei L. and Jia J. 2003. Analysis of microsatellites in major crops assessed by computational and experimental approaches. Mol. Breed. 12: 245-261.

Gastier M.J., Pulido J.C., Suden S., Brody T., Kenneth H.B., Murray J.C., Weber J.L., Hudson T.J., Sheffield V.C. and Duyk G.M. 1995. Survey of trinucleotide repeats in the human genome: assessment of their utility as genetic markers. Hum. Mol. Genet. 4: 1829-1836.

Gaitán-Solis E., Duque M.C., Edwards K.J. and Tohme J. 2002. Microsatellite repeats in common bean (*Phaseolus vulgaris*): Isolation, characterisation and cross-species amplification in *Phaseolus* ssp. Crop Sci. 42: 2128-2136.

Gallego F.J., Pérez M.A., Núñez Y. and Hidalgo P. 2005. Comparison of RAPDs, AFLPs and SSR markers for the genetic analysis of yeast strains of *Saccharomyces cerevisiae.* Food Microbiol. 22: 561-568.

Gizlice Z., Carter Jr T.E. and Burton J.W. 1996. Genetic diversity patterns in North American public soybean cultivars based on coefficient of parentage. Crop Sci. 36: 753-765.

Glenn T.C. and Schable N.A. 2005. Isolating microsatellite DNA loci. Methods Enzymol. 395: 202-222.

Grabau E.A., Davis W.H., Phelps N.D. and Gengenbach B.G. 1992. Classification of soybean cultivars based on mitochondrial DNA restriction fragment length polymorphisms. Crop Sci. 32: 271-274.

Greilhuber J. and Obermayer R. 1998. Genome size variation in Cajanus cajan (Fabaceae): a reconsideration. Plant Syst. Evol. 212: 135-141.

Griffin J.D. and Palmer R.G. 1995. Variability of 13 isozyme loci in the USDA soybean germplasm colections. Crop Sci. 35: 897-904.

Guerra-Sanz J.M. 2004. New SSR markers of *Phaseolus vulgaris* from sequence databases. Plant Breed. 123: 87-89.

Hamilton M.B. and Fleischer R.C. 1999. Cloned microsatellite repeats differ between 4-base restriction endonucleases. J. Hered. 90: 561–563.

Hancock J.M. 1995. The contribution of slippage-like processes to genome evolution. J. Mol. Evol. 41: 1038-1047.

He G., Meng R., Newman M., Gao G., Pittman R.N. and Prakash C.S. 2003. Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L.). BMC Plant Biol. 3: 3.

Heslop-Harrison J.S. 2000. Comparative genome organisation in plants: From sequence and markers to chromatin and chromosomes. Plant Cell 12: 617-636.

Huang X., Madan A. 1999. CAP3: A DNA sequence assembly program. Genome Res. 9: 868-877.

Huttner E., Wenzl P., Akbari M., Caig V., Carling J., Cayla C., Evers M., Jaccoud D., Peng K., Patarapuwadol S., Uszynski G., Xia L., Yang S. and Kilian A. 2005. Diversity arrays technology: A novel tool for harnessing the genetic potential of orphan crops. In: Serageldin I, Persley GJ (eds) Discovery to Deliver: BioVisionAlexandria. Proceedings of the 2004 Conference of The World Biological Forum, CABI Publishing.

Hyten D.L., Pantalone V.R., Sams C.E., Saxton A.M., Landau-Ellis D., Stefaniak T.R. and Schmidt M.E. 2004. Seed quality QTL in a prominent soybean population. Theor. Appl. Genet. 109: 552-561.

Jander G., Norris S.R., Rounsley S.D., Bush D.F., Levin I.M. and Last R.L. 2002. *Arabidopsis* map-based cloning in the post-genome era. Plant Physiol. 129: 440-450.

Jin P. and Warren S.T. 2000. Understanding the molecular basis of fragile X syndrome. Hum. Mol. Genet. 9: 901-908.

Jones E.S., Dupal M.P., Kölliker R., Drayton M.C. and Forster J.W. 2001. Development and characterisation of simple sequence repeats (SSR) markers for

perennial ryegrass (*Lolium perenne* L.). Theor. Appl. Genet. 102: 405-413.

Kalinowski S.T. 2004. Counting alleles with rarefaction: private alleles and hierarchical sampling designs. Conserv. Genet. 5: 539-543.

Karagyozov L., Kalcheva I.D. and Chapman V.M. 1993. Construction of random small-insert genomic libraries highly enriched for simple sequence repeats. Nucleic Acids Res. 21: 3911-3912.

Kosambi D.D. 1944. The estimation of map distances from recombination values. Annu. Eugen. 12: 172-175.

Kölliker R., Jones E.S., Drayton M.C., Dupal M.P. and Forster J.W. 2001. Development and characterisation of simple sequence repeat (SSR) markers for white clover (*Trifolium repens* L.). Theor. Appl. Genet. 102: 416-424.

Kresovich S., Lamboy W.F., Li R., Ren J., Szewc-McFadden A.K. and Bliek S.M. 1994. Application of molecular methods and statistical analysis for discrimination of accessions and clones of vetiver grass. Crop Sci. 34: 805-809.

Krishna T.G. and Reddy L.J. 1982. Species affinities between *Cajanus cajan* and some *Atylosia* species based on esterase isozymes. Euphytica 31: 709-713.

Kruskal J.B. and Wish M. 1978. Multidimensional Scaling. Sage Publications Inc, CA, 96 pp.

Kuleung C., Baenziger P.S. and Dweikat I. 2004. Transferability of SSR markers among wheat, rye, and triticale. Theor. Appl. Genet. 108: 1147-1150.

Kumar A. and Bennetzen J.L. 1999. Plant retrotransposons. Ann. Rev. Genet. 33: 479-532.

Kumar S.M., Kumar B.K., Sharma K.K. and Devi P. 2004. Genetic transformation of pigeonpea with rice chitinase gene. Plant Breed. 123: 485-489.

Kumar Rao J.V.D.K., Thompson J.A., Sastry P.V.S.S., Giller K.E. and Day J.M. 1987. Measurement of N_2-fixation in field-grown pigeonpea [*Cajanus cajan* (L.) Millsp.] using ^{15}N-labelled fertilizer. Plant Soil 101: 107–113.

La Rota M., Kantety R.V., Yu J.-K. and Sorrells M.E. 2005. Non-random distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. BMC Genomics 6: 23 doi:10.1186/1471-2164-6-23.

Labra M., Imazio S., Grassi F., Rossoni M. and Sala F. 2004. *Vine-1* retrotransposon-based sequence-specific amplified polymorphism for *Vitis vinifera* L. genotyping. Plant Breed. 123: 180-185.

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

Lagercrantz U., Ellegren H. and Andersson L. 1993. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res. 21: 1111-1115.

Lee J.M., Grant D., Vallejos C.E. and Shoemaker R.C. 2001. Genome organization in dicots. II. *Arabidopsis* as a 'bridging species' to resolve genome evolution events among legumes. Theor. Appl. Genet. 103: 765-773.

Levinson G.L. and Gutman G.A. 1987. High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K12. Nucleic Acids Res. 15: 5323-5338.

Li Y.-C., Korol A.B., Fahima T., Beiles A. and Nevo E. 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol. Ecol. 11: 2453-2465.

Li Y.-C., Korol A.B., Fahima T. and Nevo E. 2004. Microsatellites within genes: structure, function, and evolution. Mol. Biol. Evol. 21: 991-1007.

Lichtenzveig J., Scheuring C., Dodge J. and Abbo S. 2005. Construction of BAC and BIBAC libraries and their applications for generation of SSR markers for genome analysis of chickpea, *Cicer arietinum* L. Theor. Appl. Genet. 110: 492-510.

Lopes M.S., Sefc K.M., Laimer M. and Machado A.D. 2002. Identification of microsatellite loci in apricot. Mol. Ecol. Notes 2: 24-26.

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

Loridon K., Cournoyer B., Goubely C., Depeiges A., Picard G. 1998. Length polymorphism and allele structure of trinucleotide microsatellites in natural accessions of *Arabidopsis thaliana*. Theor. Appl. Genet. 97: 591-604.

Lubbers E.L., Gill K.S., Cox T.S. and Gill B.S. 1991. Variation of molecular markers among geographically diverse accessions of *Triticum tauschii*. Genome 34: 354-361.

Mace E.S., Buhariwalla H.K. and Crouch J.H. 2003. A high throughput DNA extraction protocol for tropical molecular breeding programs. Plant Mol. Biol. Rep. 21: 459a-459h.

Mallikarjuna N. and Saxena K.B. 2005. A new cytoplasmic nuclear male-sterility system derived from cultivated pigeonpea cytoplasm. Euphytica 142: 143-148.

Marley P.S. and Hillocks R.J. 1996. Effect of root-knot nematodes (*Meloidogyne* spp.) on *Fusarium* wilt in pigeonpea (*Cajanus cajan*). Field Crop Res. 46: 15-20.

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

McGinnis S. and Madden T.L. 2004. BLAST: At the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 32: W20-W25.

Métais I., Hamon B., Jalouzot R. and Peltier D. 2002. Structure and level of genetic diversity in various bean types evidenced with microsatellite markers isolated from a genomic enriched library. Theor. Appl. Genet. 104: 1346-1352.

Milbourne D., Meyer R.C., Collins A.J., Ramsay L.D., Gebhardt C. and Waugh R. 1998. Isolation, characterization and mapping of simple sequence repeat loci in potato. Mol. Gen. Genet. 259: 233-245.

Minja E.M., Shanower T.G., Silim S.N. and Karuru O. 2000. Efficacy of different insecticides for pigeonpea pest management in Kenya. ICPN 7: 30-43.

Mogg R., Batley J., Hanley S., Edwards D., O'Sullivan H. and Edwards K. 2002. Characterisation of the flanking regions of *Zea mays* microsatellites reveals a large number of useful sequence polymorphisms. Theor. Appl. Genet. 105: 532-543.

Mohammadi S.A. and Prasanna B.M. 2003. Analysis of genetic diversity in crop plants – salient statistical tools and considerations. Crop Sci. 43: 1235-1248.

Moretzsohn M.C., Hopkins M.S., Mitchell S.E., Kresovich S., Valls J.F.M. and Ferreira M.E. 2004. Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hypervariable regions of the genome. BMC Plant Biol. 4: 11 doi:10.1186/1471-2229-4-11.

Moretzsohn M.C., Leoi L., Proite K., Guimarães P.M., Leal-Bertioli S.C.M., Gimenes M.A., Martins W.S., Valls J.F.M., Grattapaglia D. and Bertioli D.J. 2005. A microsatellite-based, gene-rich linkage map for the AA genome of *Arachis* (Fabaceae). Theor. Appl. Genet. 111: 1060-1071.

Morgante M., Hanafey M. and Powell W. 2002. Microsatellites are preferentially associated with non-repetitive DNA in plant genomes. Nat. Genet. 30: 194-200.

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

Mudge J., Cannon S.B., Kalo P., Oldroyd G.E.D., Roe B.A., Town C.D. and Young N.D. 2005. Highly syntenic regions in the genomes of soybean, *Medicago truncatula*, and *Arabidopsis thaliana*. BMC Plant Biol. 5: 15.

Nei M. and Li W.H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA 76: 5269-5273.

Nene Y.L. and Sheila V.K. 1990. Pigeonpea: Geography and importance. In: Nene Y., Hall S.D., Sheila V.K. (eds), The Pigeonpea. CABI, Wallingford, pp. 1-14.

Newbury J. and Ford-Lloyd B. 1999. The production of user-friendly molecular marker for studying plants. Biotechnol. News 40: 5-6.

Nunome T., Suwabe K., Iketani H. and Hirai M. 2003. Identification and characterization of microsatellite in eggplant. Plant Breed. 122: 256-262.

Oberhagemann P., Chatot-Balandras C., Schäfer-Preg R., Wegener D., Palomini C., Salamini F., Bonnel E. and Gebhardt C. 1999. A genetic analysis of quantitative resistance to late blight in potato: towards marker-assisted selection. Mol. Breed. 5: 399-415.

Okiror M.A. 1986. Breeding for Resistance to *Fusarium* wilt. PhD thesis, University of Nairobi

Olsen K.M. 2004. SNPs, SSRs and inferences on cassava's origin. Plant Mol. Biol. 56: 517-526.

Palacios G., Bustamante S., Molina C., Winter P. and Kahl G. 2002. Electrophoretic identification of new genomic profiles with a modified selective amplification of microsatellite polymorphic loci technique based on AT/AAT polymorphic repeats. Electrophoresis 23: 3341-3345.

Palop M., Palacios C. and González-Candelas F. 2000. Development and across-species transferability of microsatellite markers in the genus *Limonium* (*Plumbaginaceae*). Conserv. Genet. 1: 177-179.

Panaud O., Chen X. and McCouch S.R. 1996. Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). Mol. Gen. Genet. 252: 597-607.

Pandian A., Ford R. and Taylor P.W.J. 2000. Transferability of sequence tagged microsatellite site (STMS) primers across four major pulses. Plant Mol. Biol. Rep. 18: 395a-395h.

Parkin I.A.P., Sigrun M.G., Andrew G.S., Lewis L., Martin T., Thomas C.O. and Derek J.L. 2005. Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. Genetics 171: 765-781.

Peakall R., Gilmore S., Keys W., Morgante M. and Rafaske A. 1998. Cross-species amplification of soybean (*Glycine max*) simple sequence repeat (SSRs) within the genus and other legume: implications for the transferability of SSRs in plants. Mol Biol Evol 15: 1275-1287.

Pearson W.R. 1998. Empirical statistical estimates for sequence similarity scores. J. Mol. Biol. 276: 71-84.

Peleman J.D. and van der Voort J.R. 2003. Breeding by Design. Trends Plant Sci. 8: 330-334.

Powell W., Machray G.C. and Provan J. 1996. Polymorphism revealed by simple sequence repeats. Trends Plant Sci. 1: 215-222.

Prioul S., Frankewitz A., Deniot G. and Morin G. 2004. Mapping of quantitative trait loci for partial resistance to *Mycosphaerella pinodes* in pea (*Pisum sativum* L.), at the seedling and adult plant stages. Theor. Appl. Genet. 108: 1322-1334.

Pupko T. and Graur D. 1999. Evolution of microsatellites in the yeast *Saccharomyces cerevisiae*: Role of length and number of repeated units. J. Mol. Evol. 48: 313-316.

Purseglove J.W. 1968. Tropical Crops – Dicotyledons 1. Longmans, London, pp. 236-241.

Qi X., Lindup S., Pittaway T.S., Allouis S., Gale M.D. and Devos K.M. 2001. Development of simple sequence repeat markers from bacterial artificial chromosomes without subcloning. Biotechniques 31: 355-362.

Rachie K.O. and Roberts L.M. 1974. Grain legumes of the lowland tropics. Adv. Agron. 26: 1-32.

Rafalski J.A. 2002. Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant Sci. 162: 329-333.

Rallo P., Dorado G. and Martin A. 2000. Development of simple sequence repeats (SSRs) in olive tree (*Olea europaea* L.). Theor. Appl. Genet. 101: 984-989.

Ramachandran S. and Sundaresan V. 2001. Transposons as tools for functional genomics. Plant Physiol. Biochem. 39: 243-252.

Ramsay L., Macaulay M., Cardle L., Morgante M., Ivanissevich S., Maestri E., Powell W. and Waugh R. 1999. Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. Plant J. 17: 415-425.

Rao S.C., Coleman S.W. and Mayeux H.S. 2002. Forage production and nutritive value of selected pigeonpea ecotypes in the southern great plains. Crop Sci. 42: 1259-1263.

Ratnaparkhe M.B., Gupta V.S., Ven Murthy M.R. and Ranjekar P.K. 1995. Genetic fingerprinting of pigeonpea (*Cajanus cajan* (L.) Millsp.) and its wild relatives using RAPD markers. Theor. Appl. Genet. 91: 893-898.

Reddy L.J., Upadhyaya H.D., Gowda C.L.L. and Sube S. 2005. Development of a core collection in pigeonpea [*Cajanus cajan* (L) Millspaugh] using geographic and qualitative morphological descriptors. Genet. Res. Crop Evol. (In press)

Reddy M.V., Raju T.N. and Sheila V.K. 1996. *Phytophthora* blight resistance in wild pigeonpea. ICPN 3: 52-53.

Rego T.J. and Nageswara Rao V. 2000. Long-term effects of grain legumes on rainy-season sorghum productivity in a semi-arid tropical vertisol. Exp. Agr. 36: 205–221.

Reiter W.-D. 2002. Biosynthesis and properties of the plant cell wall. Curr. Opin. Plant Biol. 5: 536-542.

Remanandan P., Sastry D.V.S.S.R. and Mengesha M.H. 1988. ICRISAT pigeonpea germplasm catalog (in two companion volumes): Evaluation and analysis. ISBN 92-9066-152-6. Passport information. ISBN 92-9066-153-4. Patancheru, A.P. 502 324, India: ICRISAT.

Rensink W.A., Lee Y., Liu J., Iobst S., Ouyang S. and Buell C.R. 2005. Comparative analyses of six solanaceous transcriptomes reveal a high degree of sequence conservation and species-specific transcripts. BMC Genomics 6: 124.

Ribaut J.-M. and Hoisington D. 1998. Marker-assisted selection: new tools and strategies. Trends Plant Sci. 3: 236-239.

Rohlf F.J. 1994. NTSYS-PC*:* Numerical Taxonomy and Multivariate Analysis System. New York: Exeter Software.

Rongwen J., Akkaya M.S., Bhagwat A.A., Lavi U. and Cregan P.B. 1995. The use of microsatellite DNA markers for soybean genotype identification. Theor. Appl. Genet. 90: 43-48.

Salse J., Piégu B., Cooke R. and Delseny M. 2002. Synteny between *Arabidopsis thaliana* and rice at the genome level: a tool to identify conservation in the ongoing rice genome sequencing project. Nucl. Acids Res. 30: 2316-2328.

Salvi S. and Tuberosa R. 2005. To clone or not to clone plant QTLs: present and future challenges. Trends Plant Sci. 10: 297-304.

Sambrook J., Fritsch E.F. and Maniatis T. 1989. Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Saxena K.B. and Kumar R.V. 2003. Development of a cytoplasmic nuclear male-sterility system in pigeonpea using *C. scarabaeoides* (L.) Thouars. Indian J. Genet. Plant Breed. 63: 225-229.

Saxena K.B., Kumar R.V. and Rao P.V. 2002. Pigeonpea nutrition and its improvement. In: Basra A.S., Randhawa I.S. (eds), Quality Improvement in Field Crops. Food Products Press, pp. 227-260.

Saxena K.B., Rao A.N., Singh U. and Remanandan P. 1996. Intraspecies variation in *Cajanus platycarpus* for some agronomic traits and crossability. ICPN 3: 49-51.

Saxena K.B., Singh L. and Gupta M.D. 1990. Variation for natural out-crossing in pigeonpea. Euphytica 46: 143-148.

Schorderet D.F. and Gartlar S.M. 1992. Analysis of CpG suppression in non-methylated species. Proc. Natl. Acad. Sci. USA 89:.957-961.

Senda T., Kubo N., Hirai M. and Tominaga T. 2004. Development of microsatellite markers and their effectiveness in *Lolium temulentum*. Weed Res. 44: 136-141.

Sermon K., Seneca S., De Rycke M., Goossens V., Van de Velde H., De Vos A., Platteau P., Lissens W., Van Steirteghem A. and Liebaers I. 2001. PGD in the lab for triplet repeat diseases-myotonic dystrophy, Huntingdon's disease and fragile-X syndrome. Mol. Cell Endocrinol. 183: 77-85.

Sharma H.C., Pampapathy G. and Reddy L.J. 2003. Wild relatives of pigeonpea as a source of resistance to the pod fly (*Melanagromyza obtuse* Malloch) and pod wasp (*Tanaostigmodes cajaninae* La Salle). Genet. Res. Crop Evol. 50: 817-824.

Shimamura M., Yasue H., Oshima K., Abe H., Kato H., Kishiro T., Goto M., Munechika I. and Okada N. 1997. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. Nature 388: 666-670.

Simko L., Costanzo S., Haynes K.G., Christ B.J. and Jones R.W. 2004. Linkage disequilibrium mapping of a *Verticillium dahliae* resistance quantitative trait locus in tetraploid potato (*Solanum tuberosum*) through a candidate gene approach. Theor. Appl. Genet. 108: 217-224.

Sivaramakrishnan S., Seetha K., Nageshwar Rao A. and Singh L. 1997. RFLP analysis of cytoplasmic male-sterile lines of pigeonpea [*Cajanus cajan* (L.) Millsp.] developed by interspecific crosses. Euphytica 93: 307-312.

Sivaramakrishnan S., Seetha K. and Reddy L.J. 2002. Diversity in selected wild and cultivated species of pigeonpea using RFLP of mtDNA. Euphytica 125: 21-28.

Smulders M.J.M., Bredemeijer G., Rus-Kortekaas W., Arens P. and Vosman B. 1997. Use of short microsatellites from database sequences to generate polymorphism among *Lycopersicon esculentum* cultivars and accessions of other *Lycopersicon* species. Theor. Appl. Genet. 97: 264-272.

Sneath P.H.A. and Sokal R.R. 1973. Numerical Taxonomy. Freeman, San Francisco, 573 pp.

Song Q.J., Marek L.F., Shoemaker R.C., Lark K.G., Concibido V.C., Delannay X., Specht J.E. and Cregan P.B. 2004. A new integrated genetic linkage map of the soybean. Theor Appl Genet 109: 122-128.

Song Q.J., Shi J.R., Singh S., Fickus E.W., Costa J.M., Lewis J., Gill B.S., Ward R. and Cregan P.B. 2005. Development and mapping of microsatellite (SSR) markers in wheat. Theor. Appl. Genet. 110: 550-560.

Souframanien J., Manjaya J.G., Krishna T.G. and Pawar S.E. 2003. Random amplified polymorphic DNA analyses of cytoplasmic male sterile and male fertile pigeonpea (*Cajanus cajan* (L.) Millsp.). Euphytica 129: 293-299.

Squirrell J., Hollingsworth P.M., Woodhead M., Russell J., Lowe A.J., Gibby M. and Powell W. 2003. How much effort is required to isolate nuclear microsatellites from plants? Mol. Ecol. 12: 1339-1348.

Stajner N., Jakse J., Kozjak P. and Javornik B. 2005. The isolation and characterisation of microsatellites in hop (*Humulus lupulus* L.). Plant Sci. 168: 213-221.

Strong W.B. and Nelson R.G. 2000. Preliminary profile of the *Cryptosporidium parvum* genome: an expressed sequence tag and genome survey analysis. Mol. Biochem. Parasit. 107: 1-32.

Sturtevant A.H. 1913. The linear arrangement of six sex-linked factors in *Drosophila* as shown by their mode of association. J. Exp. Zool. 14: 43-59.

Su X.-Z., Wu Y., Sifri C.D. and Wellems T.E. 1996. Reduced extension temperatures required for PCR amplification of extremely A+T-rich DNA. Nucl Acids Res 24: 1574-1575.

Subbarao G.V., Johansen C., Jana M.K. and Kumar Rao J.V.D.K. 1991. Comparative salinity responses among pigeonpea accessions and their relatives. Crop Sci. 31: 415-418.

Subramanian S., Mishra R.K. and Singh L. 2003. Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. Genome Biol 4: R13 http://genomebiology.com/2003/4/2/R13

Svetleva D., Velcheva M. and Bhowmik G. 2003. Biotechnology as a useful tool in common bean (*Phaseolus vulgaris* L.) improvement. Euphytica 131: 189-200.

Tang S., Yu J.-K., Slabaugh M.B., Shintani D.K. and Knapp S.J. 2002. Simple sequence repeat map of the sunflower genome. Theor. Appl. Genet. 105: 1124-1136.

Tautz D. and Rentz P. 1984. Simple sequences are ubiquitous repetitive componenets of eukaryotic genome. Nucleic Acids Res. 12: 4127-4138.

Taylor J.S., Durkin J.M.H. and Breden F. 1999. The death of a microsatellite: a phylogenetic perspective on microsatellite interruptions. Mol. Biol. Evol. 16: 567-572.

Temnykh S., DeClerck G., Lukashova A., Lipovich L., Cartinhour S. and McCouch S. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genet Res. 11: 1441-1452.

Thiel T., Michalek W., Varshney R.K. and Graner A. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor. Appl. Genet. 106: 411-422.

Thompson J.A., Nelson R.L. and Vodkin L.O. 1998. Identification of diverse soybean germplasm using RAPD markers. Crop Sci. 38: 1348-1355.

Thu T.T., Mai T.T.X., Dewaele E., Farsi S., Tadesse Y., Angenon G. and Jacobs M. 2003. *In vitro* regeneration and transformation of pigeonpea (*Cajanus cajan* (L.) Millsp.). Mol. Breeding 11: 159-168.

Timchenko N.A., Iakova P., Cai Z.J., Smith J.R. and Timchenko L.T. 2001. Molecular basis for impaired muscle differentiation in myotonic dystrophy. Mol. Cell Biol. 21: 6927-6938.

Tóth G., Gáspári Z. and Jurka J. (2000) Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res. 10: 967-981.

Troedson R.J., Wallis E.S. and Singh L. 1990. Pigeonpea: adaptation. In: Nene Y, Hall SD, Sheila VK (eds), The Pigeonpea. CABI, Wallingford, pp. 159-177.

Tyagi A.K., Khurana J.P., Khurana P., Raghuvanshi S., Gaur A., Kapur A., Gupta V., Kumar D., Ravi V., Vij S., Khurana. P. and Sharma S. 2004. Structural and functional analysis of rice genome. J. Genet 83: 79-99.

Udupa S., Robertson L., Weigand F., Baum M. and Kahl G. 1999. Allelic variation at $(TAA)_n$ microsatellite loci in a world collection of chickpea (*Cicer arietinum* L.). Mol. Gen. Genet. 261: 354-363.

Van Camp W. 2005. Yield enhancement genes: seeds for growth. Curr Opin Biotechnol 16: 147-153.

Van de Wiel C., Arens P. and Vosman B. 1999. Microsatellite retrieval in lettuce (*Lactuca sativa* L.) Genome 42: 139-149.

Van der Maesen L.J.G. 1983. World distribution of pigeonpea. *ICRISAT Information Bulletin* No.14. Patancheru, AP, India: ICRISAT, 40 pp.

Van der Maesen L.J.G. 1990. Pigeonpea: origin, history, evolution and taxonomy. In: Nene YL, Hall SD, Sheila VK (eds), The Pigeonpea. CABI, Wallingford, UK, pp. 15-46.

Van Deynze A.E., Nelson J.C., Yglesias E.S., Harrington S.E., Braga D.P., McCouch S.R. and Sorrells M.E. 1995. Comparative mapping in grasses: wheat relationships. Mol. Gen. Genet. 25: 744-54.

Varshney R.K., Graner A. and Sorrells M.E. 2005. Genic microsatellite markers in plants: features and applications. Trends Biotechnol. 23: 48-55.

Vignal A., Milan D., SanCristobal M. and Eggen A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genet. Sel. Evol. 34: 275-305.

Wang D., Graef G.L., Procopiuk A.M. and Diers B.W. 2004. Identification of putative QTL that underlie yield in interspecific soybean backcross populations. Theor. Appl. Genet. 108: 458-467.

Wang M.L., Gillaspie A.G., Newman M.L., Dean R.E., Pittman R.N., Morris J.B. and Pederson G.A. 2004. Transfer of simple sequence repeat (SSR) markers across the legume family for germplasm characterization and evaluation. Plant Genet. Resour. 2: 107-119.

Wang H., Li F. and Xiang J. 2005. Polymorphic EST–SSR markers and their mode of inheritance in *Fenneropenaeus chinensis.* Aquacul. 249: 107-114.

Wang Z., Weber J., Zhong G. and Tanksley S. 1994. Survey of short tandem DNA repeats. Theor. Appl. Genet. 88: 1-6.

Whitton C., Daub J., Quail M., Hall N., Foster J., Ware J., Mehul G., Slatko B., Barrell B. and Blaxter M. 2004. A genome survey of the filarial nematode Brugia malayi: repeats, gene discovery, and comparative genomics. Mol. Biochem. Parasit. 137: 215-227.

Winter P., Pfaff T., Udupa S.M., Hüttel B., Sharma P.C., Sahi S., Arreguin-Espinoza R., Weigand F., Muelbauer F.J. and Kahl G. 1999. Characterisation and mapping of sequence-tagged microsatellite sites in the chickpea (*Cicer arietinum* L.) genome. Mol. Gen. Genet. 262:.90-101.

Yaish M.W.F. and de la Vega M.P. 2003. Isolation of (GA)n Microsatellite Sequences and Description of a predicted MADS-box sequence isolated from common bean (*Phaseolus vulgaris* L.). Genet. Mol. Biol. 26:.337-342.

Young N.D., Mudge J. and Ellis T.N. 2003. Legume genomes: more than peas in a pod. Curr. Opin. Plant Biol. 6: 199-204.

Yu K., Park S.J., Poysa V. and Gepts P. 2000. Integration of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris* L.). J. Hered. 91: 429-434.

Zane L., Bargelloni L. and Patarnello T. 2002. Strategies for microsatellite isolation: a review. Mol. Ecol. 11: 1-16.

Zhao W., Miao X., Jia S., Pan Y. and Huang Y. 2005. Isolation and characterization of microsatellite loci from the mulberry, *Morus* L. Plant Sci. 168: 519-525.

Zhebentyayeva T.N., Reighard G.L., Gorina V.M. and Abbott A.G. 2003. Simple sequence repeat (SSR) analysis for assessment of genetic variability in apricot germplasm. Theor. Appl. Genet. 106: 435-444.

Zhenghong L., Saxena K.B., Chaohong Z., Jianyun Z., Yong G., Xuxiao Z. and Shiying Y. 2001. Pigeonpea: An excellent host for Lac production. ICPN 8: 58-60.

Zou J.J., Singh R.J. and Hymowitz T. 2004. SSR marker and ITS cleaved amplified polymorphic sequence analysis of soybean X *Glycine tomentella* subgeneric derived lines. Theor. Appl. Genet. 109: 769-774.

Zwart SJ, Bastiaanssen WGM (2005) Review of measured crop water productivity values for irrigated wheat, rice, cotton and maize. Agricult. Water. Manag. 69: 115-133.

6 APPENDICES

Appendix 6.1: SSR motifs, primer sequences and PCR amplification conditions

Genebank Accession number	Motif	Forward Primer	Reverse Primer	Amplification	Touchdown (^{0}C)	Reaction Condition
CZ445529	at16	cgagctcgaattgaccctat	tttgtttttgggctcattcc	No		
CZ445530	ttc5	cgggcttccttttcttctct	aaaaccccgaaaacaccatt	Yes	60-55	1
CZ445526	(ta)4c(at)15gtg(ta)5	gggggtttgttttagattgc	caagggaatggtttggttttt	No		
CZ445526	at17	gggtaccgagctcgaattat	aggttacttgcaggcgaga	No		
CZ445525	tta10	ttctggatccctttcatttttc	tgacacccttctaccccataa	Yes	60-55	1
CZ445522	ta8	cttcccccaactaagatcca	gttcgttctctttaattgacttgc	Yes	65-60	5
CZ445523	(ta)10(tg)9	tttcctgagccatcagtcg	aagcatcaacgtaccagcaa	Yes	55-45	4
CZ445531	ta11	tgaattgctgagaggacgttt	ctgttccaattccacggttt	Yes	60-55	1
CZ445532	tc9	ccgcactagatgattctcgtc	ggagatttaggtgctttgtgga	No		
CZ445533	taa6	atattgattgagcatgtgtgtgtg	aatggaatgatgacctaagagtgt	No		
CZ445520	tat9	cccatttagtgagggttaat	gactactccaggtcaaacacg	Yes	55-45	4
CZ445534	cctc4	cacctcgctgtcctcaactt	gagagggaagtggagaaggt	No		
CZ445535	tacccg4	gtcggggcgtgtaagtcata	ccgaaataaggatggcaaat	Yes	55-45	2
CZ445536	cgggta4	gtctttgagggacggaacc	ggggcggggaaagtacata	Yes	60-55	5
CZ445537	ggc4	gggctctggaggatgttgt	ccctcttgtgctccacca	No		
CZ445538	tta4	ccaagaaaaggtgctccaagt	ttgcttcttttctcgcttgc	Yes	60-55	2
CZ445539	cat4	tgatagggaccacaacgaca	agcgttgactcctccctctt	Yes	60-55	2
CZ445540	ggt4	acgcttctgatgctgtgttg	catcagcatcatcgttaccc	Yes	60-55	2
CZ445541	cct4	ccccttctgccatttacac	gagggtgaggatgaggaaca	No		
CZ445541	tcc4	gcactgctctcatccactca	gaagagagggaggagggtga	No		
CZ445519	(ct)6tt(ct)2	gactcttcacctcacactcatcac	acctcatacaacaaccctaagcac	Yes	60-55	2
CZ445527	aat4	agggaccacagaccaaggat	tgatgtaaatggaatgatgacc	No		
CZ445542	ggc4	ccattgtgcgtctttgtgtt	gcttttcctcttcctttctcg	Yes	60-55	1
CZ445543	tat4	tgggtctgatactctattagggaaa	ccttatagggatgggtataatggt	No		
CZ445544	ttat4	tacagcagccacatcaaagc	tgaaccgtgaaagtgggatt	Yes	65-60	1
CZ445553	tta4	acccattattgatttgggta	ccaaatttcacccaagaaa	Yes	60-55	2
CZ445545	aat4	tcttccattgcatggtgtt	gcatgatatgagatgatgacga	Yes	60-55	2
CZ445546	at7	ttaccatatgccatacaacgactct	ataagcaatctcccaagtgttga	No		

Genebank Accession number	Motif	Forward Primer	Reverse Primer	Amplification	Touchdown (^{0}C)	Reaction Condition
CZ445547	ga6	ggtcagagcgaacaatcaca	gcccctccacatttttcttt	No		
CZ445528	(at)6g(taa)3	aagtgtgacaccctctaccc	tgacatcgggacatagatagaa	No		
CZ445548	tat4	ttgttttggaccttattttgtactt	cccattttccttctcttctaacc	No		
CZ445524	tg6	ctgggcctctagcatagcaa	aaacttctggacgcaaaatga	Yes	60-55	1
CZ445555	ga12	ggaaaaccccgagacaaaag	gggcaacccataaaccctaa	Yes	60-55	5
CZ445551	gtt4	caacatgacatcctcctcca	ctacgccccaagaaacacaa	No		
CZ445554	aac4	ataggcccatctccaggttc	ttaatgcccagccaattctt	Yes	60-55	1
CZ445521	(ta)4(gatag)(at)4	ctacaatcccagggaaaagg	aacaaacgtaatctgtgttgatctc	Yes	60-55	1
CZ445549	aat4	ctaccttggccaaccattct	ggcacagttcttccaccatt	No		
CZ445552	gaaa4	gggtaccgagctcgaattat	ggagtagagatcatgagaacctttt	No		
CZ445552	tc6	tgaaaaggttctcatgatctct	ggctcactatagggcgaatta	No		
CZ681920	ca8	gcgggattctcttgcttac	tcacaaaacaatttggcaca	Yes	55-45	1
CZ681921	tct6	gggaattttgttggggtttt	tgcttacgcgtggactaatg	No		
CZ681922	(gaa)5g(gaa)5	acaccaccatgctaaagaacaag	ccaagcaagacacgagtaatcata	Yes	60-55	1
CZ681923	aag5	catcgcctacaatcatacaaaga	tcttgtccttttttcagtcatcgt	Yes	60-55	5
CZ681924	gaa6	atcgctttgcatccttatc	cttcacgtacattttcgttt	Yes	60-55	5
CZ681925	ca8	tgcttcaagttgcctaccag	tcaagggaggtggactacaaa	Yes	60-55	5
CZ681926	(gaa)11gag(gaa)5gaggaagag (gaa)17	gtagaggaggttccaaatgacata	atctgtctggtgttttagtgtgct	Yes	60-55	1
CZ681927	gaa16	ctcttgcttacgcgtggact	cttttgcttttgcgtgctt	Yes	55-45	1
CZ681928	(ta)5(tg)7ta(tg)4	tcttagcatgtcctctattttcgt	agtacatttcaaatccacacatcc	Yes	60-55	4
CZ681929	aga5	tcacagaggaccacacgaag	tggactagacattgcgtgaag	Yes	60-55	1
CZ681930	tc8	gcgctaagggaaaacaaaaa	aactcccttgttgtcatatggtg	Yes	60-55	4
CZ681931	ca9	cgtctatggagggttttcag	agacatttatcaatccaaggtg	No		
CZ681932	ttc9	ttcttcttcttctcctttcttc	ccatcatcttcaactgcgata	No		
CZ681933	aga4	agagggaaagggaagagaaga	tcaagcaactccaagaaattca	Yes	60-55	2
CZ681934	cttc(ctt)4	aaggcttttcaacaaataggg	agaagagaaaaagcataaaacttca	Yes	60-55	1
CZ681935	tc8	catttatttctctctggcattcac	cgagctgcaagcataaacg	Yes	60-55	1
CZ681936	ac53	gccattacttgagtgtgagttttg	gtgtgtgtgtgtgtgtgtgtgtgt	No		
CZ681937	ag20	tgcacagattcgaaggttcc	cctcaagattcctctttctctca	Yes	60-55	1
CZ681938	att21	tcaggggtaaatgcggtatc	gaattgctttttgcttcctca	Yes	65-60	5
CZ681939	gggaga4	gagaaatatgagaggcagagagaga	aagataattcattaggggtgga	Yes	60-55	1
CZ681940	cttc4	taaggaaatggctggggttg	cacataaatttgggggttcg	Yes	55-45	4

Genebank Accession number	Motif	Forward Primer	Reverse Primer	Amplification	Touchdown (^{0}C)	Reaction Condition
CZ681941	cttt4	ggacttgttactggggcact	aattcccatggtcattcg	Yes	55-45	4
CZ681942	gac4	cctatcggaaggagaaaaacatt	tcgctaaagtcttggtagataatgg	No		
CZ681943	(gat)5(tct)(gat)4	tgggcatggtagaggaagtt	cgtcatgaagcaacaggaga	Yes	55-45	4
CZ681944	ga9	cgtgaatggctgactctatgaa	tcttgagcttgcatcctcct	No		
CZ681945	aaaccc43	ccggtttagggtttagggttt	ggttggagggtttagggttg	No		
CZ681946	ctt4	taatcccattccgttgtcgt	cccaggaagagatgagacca	Yes	60-55	2
CZ681947	ca7	aggctttctcccttcaatcc	gccttttcaaacttttctcaca	Yes	60-55	1
CZ681948	tttc4	tcttcatcctcactcttccctaa	gaggtgcccaaggaagatag	No		
CZ681949	tc13	ccttttctttgtcggaatcactaa	cggaggctgttggatctagtattt	No		
CZ681950	agg4	ggggtgaatggtagtggaaa	tccctctctcctccccttat	No		
CZ681951	(ca)10cg(ca)6	acatgtgtggcgtagtgtga	gcaaaaccgttccataaaaa	Yes	65-60	2
CZ681952	tct29	ccagccggatcgttacacta	tggtagattttctcgtgactgc	No		
CZ681952	ata4	gcagtcacgagaaaatctaccac	ggttgattatcgaatgaaatggag	No		
CZ681953	gaa47	tgggcatggtagaggaagtt	catcataatcgtcttcatcacttg	No		
CZ681954	tc7	aaaaatttcgtccaaagctcct	ggaagattgaattacatacctctcg	No		
CZ681954	tc7	gaggattgcaccaagcaact	gcactgctggccttaccata	Yes	55-45	3
CZ681955	tgt4	tgggctgtgatcgatgaat	cgacaacaacaacaccgact	Yes	60-55	1
CZ681956	tgc)4(tga)2	cgggattctccttgccttac	gcagcatcatcatcactacga	No		
CZ681957	aga4	tgttccgtttcaagtggtca	cgacatttacccactcgttca	Yes	55-45	1
CZ681958	ttc7	tagagcgttgtccctttctg	tcgaaggacaactcaagcatt	Yes	65-60	1
CZ681959	aga9gaaagaa	tgggcatggtagaggaagtt	cccaccattaccaagcaagt	No		
CZ681960	at6	tcgtgggaatgctctacaac	aaccacaagtacacccacacc	Yes	65-60	2
CZ681961	aga10	atgggcatggtagaggaggt	cgctcatcatcgtcatcaaa	Yes	55-45	3
CZ681962	(tg)(tc)2(tg)7	gggaaactcacctatattaccaa	cactaccgtctacagccatctc	Yes	65-60	1
CZ681963	(ttg)5(ttc)7	gttcttcttgttgttgttgttg	aattcgtggagttcattgg	Yes	65-60	2
CZ681964	(ca)7aca(ta)3	gatagcacacacacacacaaca	taccttagggtcaccaacga	Yes	65-60	1
CZ681965	gaaaaa5	ctttgttcagagcggagcat	tttttaggacattgggaagca	Yes	65-60	2
CZ681966	(ttc)4tgc(ttc)3	agtcgatgtggaacatgagga	tgttgtaagccgtgggtagg	Yes	55-45	1
CZ681968	att4	caggattttaatggattctgcaa	gggtgaatactatttaaaaggatagg	Yes	65-60	5
CZ681967	(gaa)2gagg(gaa)4gag(gaa)2	aggtgcaaaggaagcactaat	cagctccactgtcttcaacg	Yes	60-55	1
CZ681969	act4	atcccagacttcatagggagatag	gtctagtcccaggtacaaagaggt	Yes	60-55	2
CZ681970	tc6	cttctccctgcctcttttcc	caagtggaggggagtgaaga	Yes	55-45	2
CZ681970	ctt4	tcctctctcctcttgtctttgtc	atggagaagtgaaagggatatgt	No		

Genebank Accession number	Motif	Forward Primer	Reverse Primer	Amplification	Touchdown (^{0}C)	Reaction Condition
CZ681971	ca8	aagttgcctactgggggttc	aaatagagctgtcaggggaggt	Yes	65-60	4
CZ681972	att5	tgcatgatatgagatgatggaga	ccctttttcacccaaaaatacaa	No		
CZ681973	gaa6	tggactaccaaacgcagaca	tcgtagctgcagagcatttt	Yes	65-60	1
CZ681974	ctc4	atcctccaaaagttccacca	caaaggaggatttccaccaa	Yes	55-45	3
CZ681974	tct4	cggccccttctatactgtca	gaaaagagaaaagaaggaaagagga	No		
CZ681975	cat6	acggtgccttgttgattgta	cggaacaggaggaaaaggtc	Yes	55-45	5
CZ681976	ag)5aac(ga)4	gcgaagagggtaaagggaaa	ccggtcacgagaaatgtgta	No		
CZ681977	ca7	accttgcttgtttcgctttt	aagggaggtggactacaagga	Yes	65-60	3
CZ681978	tc9	tgcaagcttgtattctatagtgtc	atggagatttaggtgctttgtg	No		
CZ681979	gt7	gtgagtgagagtgagtgtatttgtg	gctctgatgccaaatgttga	Yes	65-60	5
CZ681980	ttc5	catccattgggttgttctca	ggattaaagcgcaccatcat	No		
CZ681981	(tg)6cgagtgtga(gt)11gcaacta (tg)6	catcataatcatacatgtcaatgcta	ggttttatctttgtctccaattctg	Yes	55-45	2
CZ681982	ac)10c(ca)74	tgggaaacaaaatatcccctaa	agaggggtgtgatgaagcag	No		
CZ681983	(aga)11(ggag)(gaa)4ga(gga)3 (gaa)16	tgggcatggtagaggaagtt	tcagaagtcgatggcaagtg	Yes	55-45	5
CZ681983	tga11	gaggaggaggaagaagaagaaga	tcgtcgccgtatcactacaa	No		
CZ681984	ac53	catgcgtattgaatgaattg	tctcgtctgagtgggagtgt	No		
CZ681985	gt8	gcccctcttacacctttctt	ctcttgcttacgcgtggact	No		
CZ681986	(ta)7(ca)6	tgctctaatggctagttcatcc	aaacactcatgggttagattctcc	Yes	65-60	2
CZ681987	ttc8	tcttgcttacgcgtggacta	tggagaagggacacaaatgc	No		
CZ681988	ttccc)(ttc)3tcc(ttc)4taca(tct)7	ttacctgacgtgaagtgaatgg	cgtgcgacaggactacaatg	No		
CZ681989	gat4	tagtatgggcgtggtagagga	cgtgacagagtcaatcagaagc	Yes	55-45	2
CZ681990	(tg)6(agtg)3	caggtctgctactgccatca	agcccacttctgcatcactc	Yes	60-55	1
CZ681991	(ac)7(ca)3	ccacatccctcaacccatac	gaaaagcccttgatgacacc	Yes	60-55	1
CZ681992	ttc4	ttgtccgtagctctcgtttct	gctatgcagcggtaagtgtg	No		
CZ681993	(ga)4ca(ga)4cagagt(ga)8	atcatcagattcttcagccgta	ggttagaccaatccaatcaagc	Yes	60-55	1
CZ681994	(ac)4aa(ac)38c(ca)7	gggaaacaaaatatcccctaatc	taatcacacacatcacacctagca	Yes	55-45	5
CZ681995	(ac)6aag(ctaa)3	cacgattccattggtggag	acggtttctgggagggtcta	Yes	60-55	1
CZ681996	at6	ccacaagtacacccacacca	ttcgtgggaatgctctacaa	Yes	60-55	1
CZ681997	ag)6g(c)9	ttgggaaatgaaggttgagc	gcgtggagtaatccatgaaaa	No		
CZ681998	tc6	acaaatccggtgacccataa	ccgagaacaaaaacattgaaca	Yes	65-60	5
CZ681999	caccac(a)5(ca)6c(a)4	gactagaaaattcacctccgtctg	ttacaaaggctacattgatgagaac	Yes	65-60	1

Genebank Accession number	Motif	Forward Primer	Reverse Primer	Amplification	Touchdown (^{0}C)	Reaction Condition
CZ682000	tg7a(gt)23	ttgcttacgcgtggactaga	aacagtgggtgcatatgatttt	No		
CZ682001	ac6a	tctttcagacgcaatgacctt	cacttatttgtggggaccatc	Yes	60-55	1
CZ682002	tg6	caaggaatcacttaaaaaccaagc	agatggccaagattccacaac	Yes	60-55	4
CZ682003	ca6(ta)6(ca)3	catcaggcgttaggaactctc	ttgtggattgtgttatgtgtgc	No		
CZ682004	gt6	gccttttcaaacttttctca	catatgctttaagtgctttcct	Yes	60-55	1
CZ682005	ac6	tgtatgttcgtttagaggcttcc	gccccttttcacttttctca	Yes	65-60	5
CZ682006	ca6c	tgcctactaggggtttcgtg	tgaactatccagggaggtgag	Yes	65-60	1
CZ682007	tgt)(ttg)2(tg)7	tgatttgtgcttgtgccttg	gtcttgcttacgcgtggact	No		
CZ682008	ca8	aacgatgaaattcccaaacg	tgttagatgctcaacccaagg	Yes	60-55	1
CZ682009	tg7	agccacttaataaccaagcctttt	gtgtatgctttacttgctttccttt	Yes	65-60	5
CZ682010	aga6	tcttcgctttgaggggacta	gggaattttgttggggtttt	No		
CZ682011	gt7	aaattcaccaccatgatccaa	tcttcacttccgagacacaact	Yes	55-45	4
CZ682012	caa)g(ca)5cg(ca)(ta)2	tcaacacctgattaagatttgttcc	agggtttctcaagtggtaaggttt	No		
CZ682012	atg4	caagaaagcacccctcgtag	ataggagcatccgtcgacaa	No		
CZ682013	ac7	tgagaggcaatgatgttgga	tctacaggcacccctttgaaaat	Yes	60-55	1
CZ682014	(cata)3ta(tg)6	atcggcttttgtcttgatga	aagctacaagggatacacatgc	Yes	60-55	2
CZ682014	ac8(at)7acat	ggccaagtcactgtcgaatc	tgtagtccacgcgtaagcaa	No		
CZ682015	aga5	acaattactcaaatgctctcaacg	taaatgtcgcttcctatgatagacc	Yes	60-55	1
CZ682016	tg7	gacgtggtcattgaaagtagca	agacaaaaactacacgcactcaag	Yes	60-55	1
CZ682017	aag13	tgaaatgaacaaacctcaatgg	tgtattgcacattgacttggcta	Yes	60-55	5
CZ682018	tct9	ttggtcacatagttgtagagtgttg	aaagattactctgttgtcgtggat	No		
CZ682019	aat4	aacacgcacctcaattcca	gaatgaggaatgaagggacaaa	Yes	60-55	1
CZ682019	ctt8	attccctctctatctcagactttt	tcgtgatggaactcaagatacact	Yes	60-55	1
CZ682020	ag8	gcggtgaagatggatggat	ctcttgcttacgcgtggact	No		
CZ682021	tg8	ttagggtcaccagtgatgatatgt	tttcaggtgcagaaataaaggttag	Yes	65-60	2
CZ682021	(caa)(ca)6caa	cgtggactaatcatcccgtaa	ataatgccaaaaggggagaa	Yes	60-55	1
CZ682022	aac4	atcccgtaatgcacctttg	ttggtctgaattgtggcctat	No		
CZ682023	tc13	ctaggccctcgagctacatt	tcttttagaggtgcgctgtg	Yes	55-45	5

Appendix 6.2: SSR motifs, primer sequences and PCR amplification conditions

Genebank Accession number	Locus	Repeat	Left Primer	Right Primer
5058094	SP 001	tgctc(3)	attgtttgattccgcttggt	ccgagagctaagaagccaac
10843581	SP 003	tga(4)	gggatcttctccctccaaag	aatggtaggggtgtccact
7139589	SP 004	at(26)	gattcttctgcctcgattgc	gacagcgtatgcctgaacaa
4396518	SP 010	caa(6)	atcatcagcagcagcaagaa	ttgttgttgaggaggtgtgg
7029704	SP 028	tta(4)	tctccttcatcttggaagtgc	caaaacagtagggaagcaaaca
6747670	SP 032	caa(5)	acacactccctcctcctcct	tgttgttgctgctgttgttg
15815753	SP 040	ct(6)	ttcattgcatcaccgtcact	ctcctccttccattgctctg
15815753	SP 041	attc(5)	ttcattgcatcaccgtcact	ctcctccttccattgctctg
14125154	SP 047	tc(12)	ttgtgaccacaagccttcac	cgaggaaggtgacgtaggc
13563100	SP 048	gagaa(3)	gaggcggagacgaagaagt	ttgtagacgaggacgggaat
29844449	SP 050	tatat(3)	tgtaccaaaccgtgaaaacg	ccatgcttctccaggttcat
26057188	SP 053	ct(8)	atccctctcccctctctcac	tcctttgggttttggtcttg
26048070	SP 055	cca(4)	tcgcagtcatggtcaaagaa	tctaactcctcgccatgctt
26045009	SP 057	ta(10)	caagcagagatggcccttag	gaagcgaccaattaccaaaca
24136689	SP 059	cgca(5)	cttcactttccccaactcca	cgttgtgatcactcgcagac

Appendix 6.3: Recombination frequencies of all markers analysed (JoinMap output)

S/n	Nr1	Locus1	Nr2	Locus2	Recomb.freq.	LOD
1	1	CCttc002	2	CCttc005	0.4794	0.01
2	1	CCttc002	3	CCtta006	0.4049	0.19
3	1	CCttc002	4	CCac006	0.0916	10.67
4	1	CCttc002	5	CCttc012	0.1008	9.01
5	1	CCttc002	6	CCac010	0.4416	0.08
6	1	CCttc002	7	CCat006	0.2201	3.50
7	1	CCttc002	8	CCac030	0.4990	0.00
8	1	CCttc002	9	CCttc019	0.4990	0.00
9	2	CCttc005	3	CCtta006	0.4990	0.00
10	2	CCttc005	4	CCac006	0.4990	0.00
11	2	CCttc005	5	CCttc012	0.4990	0.00
12	2	CCttc005	6	CCac010	0.4990	0.00
13	2	CCttc005	7	CCat006	0.4990	0.00
14	2	CCttc005	8	CCac030	0.2351	3.04
15	2	CCttc005	9	CCttc019	0.4990	0.00
16	3	CCtta006	4	CCac006	0.4435	0.07
17	3	CCtta006	5	CCttc012	0.4990	0.00
18	3	CCtta006	6	CCac010	0.0618	12.55
19	3	CCtta006	7	CCat006	0.4990	0.00
20	3	CCtta006	8	CCac030	0.4990	0.00
21	3	CCtta006	9	CCttc019	0.3944	0.23
22	4	CCac006	5	CCttc012	0.0618	12.70
23	4	CCac006	6	CCac010	0.4404	0.07
24	4	CCac006	7	CCat006	0.3170	1.29
25	4	CCac006	8	CCac030	0.4594	0.07
26	4	CCac006	9	CCttc019	0.4990	0.00
27	5	CCac006	6	CCac010	0.4990	0.00
28	5	CCttc012	7	CCat006	0.3442	0.84
29	5	CCttc012	8	CCac030	0.4888	0.02
30	5	CCttc012	9	CCttc019	0.4886	0.00
31	6	CCac010	7	CCat006	0.4990	0.00
32	6	CCac010	8	CCac030	0.4990	0.00
33	6	CCac010	9	CCttc019	0.3559	0.60
34	7	CCat006	8	CCac030	0.4990	0.00
35	7	CCat006	9	CCttc019	0.4990	0.00
36	8	CCac030	9	CCttc019	0.4990	0.00

Columns highlighted in red indicate significant linkage between the respective markers

ACKNOWLEDGEMENT

I express my heartfelt gratitude to my supervisor, Prof. Dr. Dorothea Bartels, for all the support, encouragement and inspiration throughout this study. Her positive approach to academic problems helped me indeed during the trying times of this research. I am deeply indebted to Dr. Christiane Gebhardt for giving me the unique opportunity to work in her laboratory, for the supervision and for the many insights into new ideas and suggestions. I thank Prof. Dr. Paul Vlek for his collaborative role that helped me gain experience from some of the best institutions as well as provide a desirable multidisciplinary approach to my research. The supervision of Dr. Jonathan Crouch and Dr. David Hoisington during my research period in ICRISAT, India is greatly acknowledged. Many thanks to Dr. Perry Cregan for providing soybean SSR primers.

I express special thanks to Dr. Barry Shapiro, Dr. Farid Waliyar, Dr. Jayashree Balaji, Dr. KB Saxena, Dr. Subhash Chandra, Dr. Kulkarni, Ms. Seetha Kannan, Prashanth, Rupa, Sanjeev, Satish, VijayKumar, all former colleagues at ICRISAT especially from AGL, Bioinformatics department, pigeonpea breeding, LSU and the Christian group. Many thanks are due to all former colleagues at the Max Planck Institute for Plant Breeding (especially Dr. Agim Ballvora and Birgit Walkemeier), and all friends from ZEF. I am at a loss of words to describe unending love, financial support and blessings of my parents. The unquestioning love showered on me by my siblings Sophie, Ephraim, Hellen, Dinah and Obed contributed a lot to my accomplishments. Finally, the full financial support awarded to me by DAAD alongside the help of Dr. Manske, Rosemarie Zabel and Hannah Peters of ZEF, would not go unmentioned. Thanks are due to all my well wishers and friends.