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) with no 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).