DEVELOPMENT OF MICROSATELLITE (SSR) AND DNA BARCODING (*MAT*K) PRIMERS FOR MARAMA BEAN [*TYLOSEMA ESCULENTUM* (BURCHELL) SCHREIBER] FOR GERMPLASM CHARACTERIZATION OF NAMIBIAN GENOTYPES

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

OF

THE UNIVERSITY OF NAMIBIA

BY

MUTSA M. TAKUNDWA

JANUARY 2010

Main supervisor: Dr. P.M. Chimwamurombe (Department of Biological Sciences, University of Namibia)

Co-supervisor: Professor K. Kunert (Forestry and Agricultural Biotechnology Institute, University Of Pretoria, South Africa)

ABSTRACT

Information on genetic variation of breeding materials helps maintain genetic diversity and sustains long term selection gain. Marama bean [*Tylosema esculentum* (Burchell) Schreiber] occurs naturally in arid, dry parts of Southern Africa and due to the high nutrient value of the seeds and tubers, rich in protein, oil and starch, it is a potential crop for arid areas where few conventional crops can survive. Microsatellites are becoming the molecular marker system of choice because they are multiallelic and generally more informative. Recently, the development of SSR enrichment techniques has increased the efficiency of SSR characterization in new species. The overall aim of the project was to develop SSR's for detection of polymorphisms in marama bean. The microsatellite regions of the genome were the main focus and the *mat*K gene was also explored for its potential to carry out a marama bean genetic diversity study. Microsatellite loci were isolated from the marama bean germplasm using a modified FIASCO enrichment technique. Nine marama bean microsatellite libraries enriched for $(AAG)_7$, $(GTT)_7$, $(AGG)_7$, $(GAG)_7$, $(CA)_{10}$, $(CT)_{10}$, $(TCC)_7$, $(CA)_{15}$ and $(CAC)_7$ were created. Of the 80 primers designed, 76% were able to detect polymorphism. All polymorphic primers will be applied in the future to establish a trait linkage map for marama. Four of the SSR's were used for a genetic variation analysis and have proved to be useful and informative for genetic diversity studies. The *mat*K gene product in marama bean was found to be half the expected size of the gene found in other legumes. The found homology with *Tylosema fassoglensis* (*trn*K gene) and *Pisum sativum* (*mat*K gene) suggests that an identical region was amplified for *Tylosema esculentum*. A phylogenetic tree was constructed based on the *mat*K sequences and the results suggest that the *mat*K region can also be used in determining levels of genetic variation and for barcoding.

Key words: *Tylosema esculentum*, microsatellites, SSR, *mat*K, genetic diversity.

TABLE OF CONTENTS

Abbreviations

AFLP-	Amplified Fragment Length Polymorphism
BLAST-	Basic Local Alignment Search Tool
bp -	Base pairs
CBOL-	Consortium for the Barcode of Life
cpDNA-	Chloroplast DNA
D_L	Discriminatory power
DNA-	Deoxy-ribonucleic acid
FIASCO-	Fast Isolation by AFLP of Sequences Containing repeats
MAS-	Molecular Marker assisted selection
*mat*K-	Maturase kinase
mM-	Millimolar
mtDNA-	Mitochondrial DNA
NCBI-	National Centre of Biotechnology Information
ng -	Nanogram
PCR-	Polymerase Chain Reaction
RAPD-	Random Amplified Fragment Length Polymorphism
RFLP-	Restriction Fragment Length Polymorphism

SSR-	Simple Sequence Repeat
SSRIT-	Simple Sequence Repeat Identification Tool
SNP-	Single nucleotide polymorphism
Taq-	*Thermophillus aquaticus*
VNTR-	Variable Number Tandem Repeat
µl-	Microlitre

Figures

Tables

Acknowledgements

This study was funded by the Kirkhouse trust, United Kingdom through a research grant (to PC) and by a World-Wide Learning Experience grant, a McGregor Fund initiative in the College of Arts and Sciences (to CAC). I am grateful to the Kirkhouse Trust which has supported my studies through the full scholarship awarded to me.

I am very grateful to my main supervisor, Dr P.M. Chimwamurombe for the opportunity and trust he gave me to carry out this study. His level of support and guidance throughout this study was immense and unlimited. He was determined to see me excel in my academic life and he has always encouraged me to work hard. He was always willing to advice and assist whenever it was necessary. I have come across few people who are willing to help others unlimitedly like he does. I have come to learn a lot about research through his mentorship and I have grown a love for this field of science.

I am greatly indebted to my co-supervisor, Professor K. Kunert for his support and critical questions during my study. His advice and words of encouragement were immense and unreserved.

I am thankful to Professor C. Cullis for his contributions in the laboratory work for this project. His seasoned expertise in the field of plant genetics was invaluable. A special thanks to Margret Cullis and Cory Bickel who assisted in the lab work on a visit with Professor Cullis to our University.

I am thankful to Mr. J. D. Uzabakhiriho and Mr. A.V. Hijarunguru for their unlimited technical assistance in the field and laboratory work. I am also thankful to Mr. E. Nepolo, for his unreserved support during this study. He was an outstanding colleague and I have enjoyed working with him.

I appreciate the help and support from the MSc. and PhD students (2008/2009) in the Karl Kunert laboratory at the Forestry and Agricultural Biotechnology Institute (FABI) in Pretoria during my visit to their laboratory.

I am particularly indebted to several individuals who were helpful and gave me emotional support and comfort throughout this study: Connie Tjiurutue, Petrina Kapewangolo, Efe Agbamu-Pote, Pauline Muandingi, Laimy Shikongo, Marius Hedimbi, Jay Bolin Prof. Enos Kiremire and Dr. Daves Mumbengegwi. The Hoveka family in the Omaheke region of Namibia and the Hijarunguru family in the Otjozondjupa region were invaluable guides throughout the field trips.

I would also like to extend my gratitude to the staff members of the Department of Biological Sciences and the Department of Chemistry and Biochemistry at The University of Namibia who extended a helping hand during this study. I would like to thank my family, and siblings who are my pillar of strength and have always supported my academic endeavours without any reservations. I am greatly indebted to my aunts Ms. R. Mucheri and Mrs. G. Bakasa whose support over the years has brought me this far. Lastly I would like to thank God who is my provider and has given me this life. All the opportunities and experiences I have had are all in his plan for my life.

Dedication

I dedicate this thesis and the work that went into it, in loving memory of my late mother Margret Gloria Mucheri-Takundwa who left us in May, 2000. Your hard work and persevering spirit have encouraged me over the years. May your soul rest in peace.

Declarations

I, Mutsa Takundwa, declare hereby that this study is a true reflection of my own research, and that this work, or part thereof has not been submitted for a degree in any other institute of higher education.

…… Date……..

Mutsa Takundwa

Publications and conference proceedings

The following are articles that have been published, submitted for publication or presented in conference proceedings based on work carried out in this thesis:

1. Nepolo, E., **Takundwa, M**., Chimwamurombe, P.M., Cullis, C.A. & Kunert, K. (2009). A review of geographical distribution of marama bean [*Tylosema esculentum* (Burchell) Schreiber] and genetic diversity in the Namibian germplasm. *African Journal of* Biotechnology, 8(10), 2088-2093.

2. **Takundwa, M.**, Chimwamurombe, P.M., Kunert, K. & Cullis, CA. (2009, July 22-25). Developing DNA barcoding (*mat*K) and SSR primers for Marama bean [*Tylosema esculentum* (Burchell) Schreiber]. Paper presented at the Tropical Crop Biotechnology Conference 2009, Hazyview, Mpumalanga, South Africa. Awarded a prize in the "Best poster" category.

3. **Takundwa, M**., Chimwamurombe, P.M., Kunert, K. & Cullis, C.A. (2010). Isolation and characterization of microsatellite repeats in Marama bean (*Tylosema esculentum*). *African Journal of Agricultural Research* 5(7), 561-566.

4. **Takundwa, M**., Nepolo, E., Chimwamurombe, P.M., Cullis, C.A., Kandawa-Schultz, M. & Kunert, K. (2010). Development and use of microsatellites markers for genetic variation analysis, in the Namibian germplasm, both

within and between populations of Marama bean (*Tylosema esculentum*). *Journal of Plant Breeding and Crop Science* 2(8), 233-242.

5. **Takundwa, M**., Chimwamurombe, P.M., Kunert, K. & Cullis, C.A. (2010). Developing DNA barcoding (*mat*K) primers for Marama bean [*Tylosema esculentum* (Burchell) Schreiber]. Submitted to the *Namibian Development Journal*.

CHAPTER 1: INTRODUCTION

1. 1. General introduction

Marama bean [*Tylosema esculentum* (Burchell) Schreiber] occurs naturally in the drier areas of Southern Africa, including Botswana, Namibia and South Africa, where it is to a small extent harvested as a wild plant for human consumption (Amarteifio & Moholo, 1998, p. 329). Due to the potential of this plant as an arid agricultural crop, there is increasing interest in its possible cultivation. Despite its traditional use as a food source in Botswana and Namibia, little is known about the germplasm diversity, genomic variability and relationships between the different ecotypes (National Academy of Sciences, 1979, p. 29).

The seeds of marama have traditionally been an important source of protein for the inhabitants of the Kalahari region, however, the plant has never been cultivated by these people but it is harvested from extensive populations which are scattered throughout the natural range (Halloran & Monaghan, 1996, p. 287). The marama beans have a protein content of about 36% and like other legumes are rich in the amino acid lysine, but low in the amino acid methionine and also contain some proteins with pharmaceutical potential such as protease inhibitors (Elfant, Bryant & Starcher, 1985, p. 329).

The oil content which ranges between 40% and 43% recorded in deshelled marama seeds approaches that in peanuts (Müseler & Schönfeldt, 2006, p. 7). This amount of nutrients compare favourably with the nutritional values of many existing legume crops hence marama is a plant of considerable potential for semi-arid agriculture. *Tylosema esculentum* has been successfully cultivated in trials carried out in America and Australia (Francis & Campbell, 2003, p. 2) and recently a cultivation program has been initiated in Namibia (Chimwamurombe, 2009, p. 2).

Halloran and Monaghan (1996, p. 290) reported that effective conservation and use of plant genetic resources for domestication involves investigating the extent of genetic variation. Over the years the methods of detecting genetic variation have expanded from Mendelian analyses of discrete morphological variants, to biometrical approaches, biochemical techniques based upon protein and isozyme profiles, to methods based on DNA sequence variation (Prasad, Varchney, Roy, Balyan & Gupta, 2000, p. 584). There is now a huge array of molecular marker techniques and each type of marker has an appropriate use in addressing key issues in germplasm conservation.

According to Naomab (2004, p. 5), to date there have been few investigations of either the amount of genetic diversity present or of the distribution of variability within and between populations of marama bean. Naomab (2004, p. 7) used RAPD and AFLP analysis in collecting this type of data and found that genetic variation in

marama bean was within rather than between marama populations. An earlier study by Halloran and Monaghan also used RAPD markers to assess variation within and between natural populations in Southern Africa (Halloran & Monaghan, 1996, p. 287). The study revealed variation occurred within rather than between populations in the three populations assessed.

Information on genetic variation is a prerequisite for the genetic improvement of any plant species for effective use of germplasm in breeding and for conservation. The natural populations of marama bean are under pressure from both grazing and human exploitation of the seeds, therefore, a detailed knowledge of the genetic structure of these populations is important for developing a strategy for conserving and developing the remaining wild germplasm (Naomab, 2004, p. 10).

The main objective of this study therefore was to identify and isolate microsatellites from marama bean and design primers based on the microsatellites that can be used as a possible molecular tool for germplasm characterization of the marama populations in Namibia. The primers might also be useful in future breeding programs through Molecular Marker Assisted Selection (MAS). The *mat*K gene, which is emerging as a DNA barcoding gene (Hollingsworth et al., 2009, p. 448) was also investigated in this study for possible contribution in genetic diversity studies and as a contribution to the greater land plants barcoding effort being undertaken by the Consortium for the Barcode of Life (CBOL).

1.2 Key features of Marama bean (*Tylosema esculentum*)

Tylosema esculentum is widespread, with large populations in Botswana (around the central Kgalagadi), Eastern parts of Namibia, and smaller populations in the provinces of Limpopo, North-West and Gauteng of South Africa (Castro, Silveira, Pereira-Coutinho, & Figueiredo, 2005, p. 99). It grows at altitudes of between 1000 and 1500 m with 300 to 700 mm rainfall, and at a minimum temperature above 15 ^{o}C and a maximum of approximately 33^{o}C (Müseler & Schönfeldt, 2006, p. 7).

The plant is adapted to the harsh conditions of Botswana and Namibia, which are characterized by low rainfall and nutritionally poor soils (Hartley, Tshamekeng, & Thomas, 2002, p. 67). Marama bean is dormant in winter and re-grows from the tuber in spring. The plant grows in well drained, fine, generally calcareous sands but also in regions of harder calcareous conglomerates, at pH 6-8 with very little organic matter, nitrate or phosphate (Lawlor, 2004, p. 1). The rainfall in these areas is typically below 800mm per annum as shown on the map of annual rainfall data on the African content. The area circled in black (Figure 1) indicates the areas where *Tylosema esculentum* is endemic.

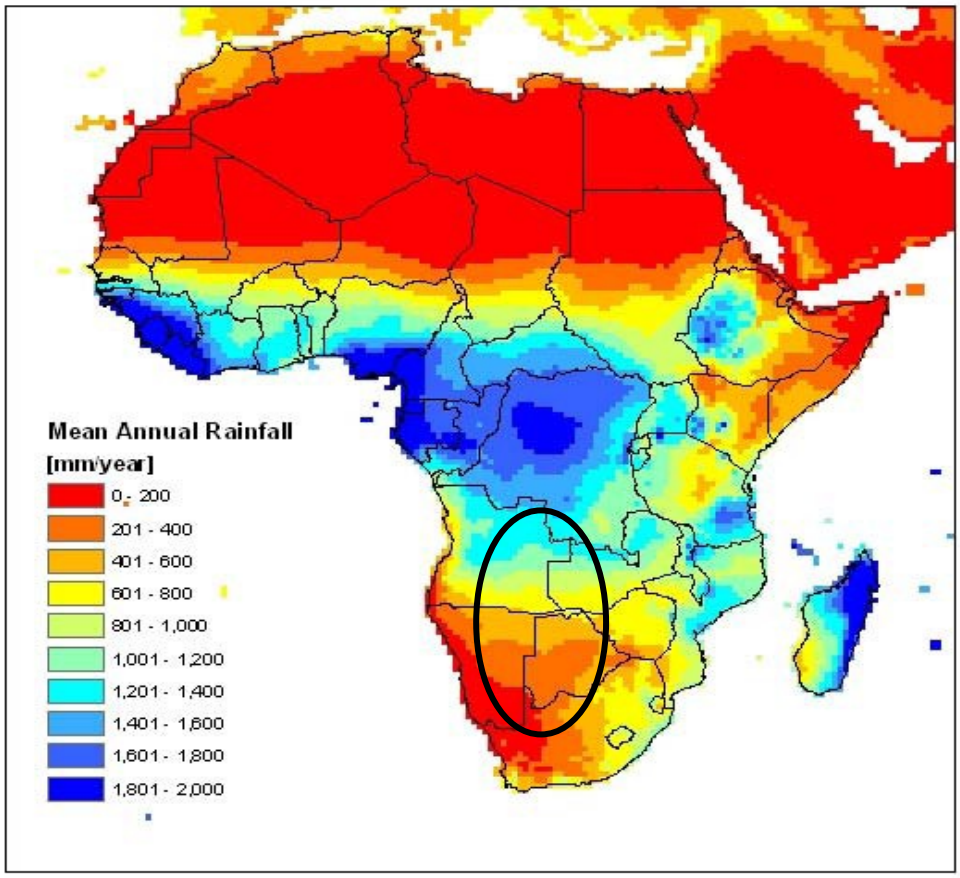


Figure 1: Mean annual rainfall distribution map for Africa. Adapted from Spelling (2008).

The genus *Tylosema* (Schweinf.) Torre & Hillc. (Leguminosae, Caesalpinioideae) comprises four currently accepted species [*T.esculentum* (Burch.) A.Schreib., *T.fassoglense* (Schweinf.) Torre & Hillc., *T.argenteum* (Chiov.) Brenan and *T. humifusum* (Pic.Serm & Roti Mich.) Brenan] occurring in eastern and central tropical Africa (Coetzer & Ross, 1977, p. 77; Thulin, 1993, p. 360). This genus has the following characteristics: presence of only two fertile stamens, the remaining seven or eight stamens being sterile and variously shaped and coloured and presence of a lobed non-spothaceous calyx limb (Castro et al., 2005, p. 99).

The leaves of marama bean are bi-lobed, opening up in cooler conditions, but closing in hot dry conditions to conserve water and moisture. The plant uses a tuber to store water. This plant is considered to be potentially a valuable crop for semi arid lands, particularly with breeding improvements (Keegan & van Staden, 1981, p. 387).

Marama bean is an excellent source of good quality protein and compares well with other protein foods including soybeans (Bower, Hertel, Oh, & Storey, 1988, p. 533). Its oil is rich in mono- and di-unsaturated fatty acids and contains no cholesterol (Mmnonatau, 2005, p. 3). It is also a good source of calcium, iron, zinc, phosphate, magnesium, the B complex vitamins and folate (Hartley et al., 2002, p. 68).

Owing to the tough outer shell of the seed, the bean can be harvested for months after ripening. The seeds are contained in pods that dehisce when dry (Naomab, 2004, p. 5). The seeds are roasted and eaten as a snack by the native Ovaherero people who dominate the Otjozondjupa and Omaheke regions of Namibia where several sub-populations of *Tylosema esculentum* are found. In their language, they refer to the plant as “Ovirema” and the seed that they roast as “Ombanui”. Figure 2 shows some marama beans ripe and ready for harvest.

Figure 2: The marama bean seeds dehisce from the pods in the field in the Omaheke Region, Namibia.

1. 3. Plant domestication and the potential of *Tylosema esculentum*

In the developing world, especially Africa, agriculture dominates the economy and is the primary source of employment (Keegan & Van Staden, 1981, p. 390). The majority of African populations are subsistence or communal farmers. However, food deficiency is still a major problem faced by many African countries including Namibia. Future demands for food will not be met without improved and new crop varieties that offer higher and more stable yields, greater economic returns to the farmers and less costly but more nutritious food for the consumers (Azam-Ali, 1996, p. 2).

Domestication refers to the outcome of a selection process that leads to increased adaptation of plants and animals to cultivation or rearing and utilization by humans (Gepts, 2004, p. 1). Domestication of the marama bean could contribute to food security in Namibia and in Sub-Saharan Africa as plants can be taken from the wild, selected and used to provide a possible food source for the local population (Gepts, 2004, p. 13). Marama bean is deserving of research attention towards its domestication. There is an increasing demand for alternative crops and more so for arid environments like Namibia.

This study was initiated to investigate physiological and agronomic aspects of marama bean to understand the genetic diversity and genetic constitution of the species and mechanisms by which the plant can be improved through selection and breeding (Travlos, Economou, & Karamanos, 2007, p. 501). According to Ruane and Sonnino (2006, p. 130) molecular characterization helps determine the breeding behaviour of species, individual reproductive success and existence of gene flow, that is, the movement of alleles within and between populations of the same or related species, and its consequences.

CHAPTER 2: LITERATURE REVIEW

2.1 Plant domestication

2.1.1 Domestication and genetic variation in plants

Crop plants evolved in parallel with human societies and most of our major crops have spread with man over large areas of the globe. The plants have been taken from the wild and domesticated. The origins of agriculture have been traced back, from archaeological evidence, to the Neolithic period some 10 000 years ago. Prior to this time and in some areas in present day, man subsisted as a hunter of wild animals and gatherer of wild roots, fruits and seeds (Gepts, 2004, p. 3).

In Namibia, the San people are still found as hunters/gatherers in the Kalahari desert regions where marama bean is endemic. The seeds of marama bean are still gathered from the wild in present day and this may lead to genetic erosion of some ecotypes of the species. Marama bean has been identified as a future crop but is still not yet cultivated (National Research Council, 2006, p. 235).

The characteristics of plants are determined by the genes located in their DNA. The expression of many of these genes is subject to the influence of the environment and plant characteristics are generally thought of as the result of gene and environment interaction (Hughes, 1996, p. 168). Genes are themselves subject to constant, though

infrequent changes due to events which are both external (e.g. irradiation) and internal (e.g. errors in the replication and repair of DNA). These random changes in the linear code of the DNA affect and produce the variation in the form and function of the plant on which adaptation depends. Organisms evolve with new potentialities and attributes which differ in their fitness under the selection pressures of a particular environment (Gepts, 2004, p. 931).

Plants have evolved and conserved the ability to exchange genetic variation among individuals. In higher plants, the mechanism of exchange –sexual reproduction-takes many forms often of great complexity (Holden, Peacock & Williams., 1993, p. 4). Sexual reproduction ensures the spread and mixing of genetic variation within the group of individuals with exchange of genes (Hughes, 1996, p. 32). The consequence is that in each generation, all individuals differ in their genetic potential and hence their fitness to the environment. Some will be better adapted than others and will leave more progeny. These progeny in turn will go through the same process of genetic reassortment when they in turn reproduce.

The reproductive biology and breeding systems in *Tylosema esculentum* have been investigated in wild and experimental populations growing in Botswana (Hartley et al., 2002, p. 67). Field observations confirmed that the species is heterostylous with the pistil and anthers exhibiting reciprocal heights in the two morphs, although pollen size and sculpturing did not vary. The wet, non papillates stigma characteristic of the

species was the first to be reported in Caesalpinioideae. A diallelic self incompatibility was demonstrated to exist by *in vivo* and *in vitro* diallel crossing experiments (Hartley et al., 2002, p. 67).

2.1.2 What makes marama bean a candidate for domestication?

Marama bean has high nutritional value and can grow in hostile environments, therefore it can be cultivated as a crop. There is much interest at present in the development of new crops and we can recognise various motives for this. There is need for:

- Alternative crops to replace those which are in overproduction, as is the case for most major crops of Western Europe and America.
- New crops to extend farming into hitherto marginal or hostile environments in order to increase total food production.
- The improvement of cultivated species which are scarcely different from wild plants. This is the case for a great number of species grown by poor subsistence farmers in tropical areas to provide nutritional supplements and dietary diversity. The potential here for ‘new’ crops is immense.
- New crops to fulfil new needs.

It seems scarcely necessary to reiterate the point that progress in any of these areas is dependent on the availability of genetic diversity. However the availability of genetic

diversity is not in itself enough to ensure success of potential new cash crops. There also needs to be a strong demand and market interest for the product (Holden et al., 1993, p. 17; Falk & Holsinger, 1991, p. 3).

Legumes are second only to *Poaceae* (the grasses) in agricultural and economic importance. The family includes horticultural varieties and many species harvested as crops and for oils, fibre, fuel, timber, medicines and chemicals. Marama bean is a legume that has been under-utilized to date. Unfortunately marama bean is not yet cultivated but remains a wild plant with so much potential. The plant needs to be developed into a crop and further developed into desirable cultivars that are high yielding and early maturing (Chimwamurombe, 2008, p. 6; Nepolo, Takundwa, Chimwamurombe, Cullis, & Kunert, 2009, p. 2089).

2.2 Molecular characterization of plants

2.2.1 Molecular markers

For quite a long period of time, allozymes have been the molecular marker of choice. In recent years, however, attention has increasingly focused on the DNA molecule as a source of informative polymorphism. Because each individuals DNA sequence is unique, this sequence information can be exploited for any study of genetic diversity or relatedness between organisms. A genetic marker is any difference in DNA, no

matter how it is detected, whose pattern of transmission can be tracked. Molecular markers are also linked to traits of interest in some cases. Markers for detecting polymorphisms among individuals based on DNA fragments that are independent of the growing environment and can be unambiguously scored are known as molecular markers in modern genetic analysis (Cullis, 2006, p. 1; Hartl & Jones, 2009, p. 39).

Molecular markers are advantageous in accelerating the introduction of desirable traits into crop plants and for assessing genetic diversity. Primer sites for molecular markers are generally conserved regions and that is what makes molecular markers useful for use in genetic diversity studies. Since nuclear DNA is inherited essentially unchanged from generation to generation, markers based on DNA fragments are ideal (Cullis, 2006, p. 1). Although an extensive number of molecular marker technologies are available, DNA-based technologies have greatly overtaken technologies based on proteins or enzymes, because the latter can be influenced by the environment. According to Ford-Lloyd (2001, p. 60), DNA-based technologies can be classified into three different categories:

1. Non-PCR-based methods: Restriction fragment length polymorphism analysis (RFLP) and variable number of tandem repeats (VNTRs) are some examples of this category (Ford-Lloyd, 2001, p. 60). The numerous differences in organisms DNA that make them genotypically distinct result in restriction fragment polymorphism and VNTR's. The Polymerase chain reaction (PCR) technique is not used here.

2. Arbitrary (or semi-arbitrary) primed techniques: This is a PCR-based category that uses random primers during PCR amplification. The most well known and widely used of these methods are the "Random Amplified Polymorphic DNA" (RAPD) and "Amplified Fragment Length Polymorphisms" (AFLP). Random amplification of DNA with short arbitrary primers is a useful technique in phylogenetics. AFPL analysis is able to detect high levels of polymorphism and has high repeatability and speed of analysis. This technique has a very high diversity index, resulting in a limited number of primer combinations to screen a whole genome (Ford-Lloyd, 2001, p. 60).

3. Site-targeted PCR: In this class, primers that amplify specific regions of the DNA are used during the PCR reaction. Examples are single nucleotide polymorphisms (SNP's), microsatellites or short sequence repeats (SSR's). These DNA sequences appear to be hypervariable and often present high levels of inter- and intra-specific polymorphism (Ford-Lloyd, 2001, p. 61).

Breeders are abandoning the traditional approach of selection based on phenotype in favour of the much faster, more highly discriminating and less costly approach of using molecular technologies. By using DNA rather than morphological markers, breeders can already select for desirable traits at the seedling stage, rather than at the adult stage. This saves time from an entire growing season (months to years depending on the plant). Molecular markers might also prove invaluable in breeding of marama bean as its complete development time is 18 months.

Karp et al., (1997, p. 626) also indicated that DNA-based technologies can be useful in defining an accession identity; defining the degree of similarity among individuals of an accession or a group of accessions; and defining the presence of a particular allele or nucleotide sequence in an individual, population, or taxon. A number of PCR–based DNA technologies have been developed in recent years to evaluate genetic variation at the intraspecific and interspecific levels (Wolfe & Liston, 1998, p. 44). Developed molecular markers using DNA-based techniques allow the selection of desired traits based on genotype and can therefore complete and accelerate plant breeding programs. They can also be used for the early selection process of traits, which are not expressed during the juvenile phase such as persistence, competitive ability, pod dehiscence character and seed yield (Kölliker, Jones, Drayton, Dupal, & Forster, 2001, p. 420).

Detecting genetic polymorphism can be approached using different methods. The different approaches are in use because no single method is ideal for all applications, each method has its own advantages and limitations and new methods are continually being developed. The five commonly used marker systems are compared in Table 1 below. The choice on which marker system to use depends on the genetic questions to be answered and also on the costs involved. Microsatellites are currently the favoured choice (Hartl & Jones, 2009, p. 65).

Table 1: A comparison of some molecular markers

Feature	RFLPs	RAPDs	AFLPs	Microsats	SNPs
Quality of DNA required	high	high	moderate	moderate	high
PCR-based	no	yes	yes	yes	yes
Number of polymorphic loci analyzed per analysis	1.0-3.0	1.5-50	20-100	1.0-3.0	1.0
Ease of use	not easy	easy	easy	easy	easy
Amenable to automation	low	moderate	moderate	high	high
Reproducibility	high	unreliable	high	high	high
Development cost	low	low	moderate	high	high
Cost per analysis	high	low	moderate	low	low

The following properties would be generally desirable for a molecular marker:

1. Highly polymorphic behaviour.
2. Codominant inheritance (which allows us to discriminate homo- and heterozygotic states in diploid organisms).
3. Frequent occurrence in the genome.
4. Even distribution throughout the genome.
5. Selectively neutral behaviour (i.e. no pleiotropic effects).
6. Easy access (e.g., by purchasing or fast procedures).
7. Easy and fast to assay (e.g., by procedures amenable to automation)

8. High reproducibility.

9. Easy exchange between laboratories.

No molecular markers are available yet that fulfil all of these criteria. However, according to the kind of study to be undertaken, one can already choose between the varieties of marker systems available, each of which combines at least some of the above mentioned properties. A major advantage of measuring DNA variation is that sampling can often be done non-invasively and genotyping done following amplification using PCR. Since extremely small samples of DNA can be amplified by around a million times by PCR, only small biological samples are needed to conduct molecular genetic analyses (Wolfe & Liston, 1998, p. 44).

PCR is used to amplify (generate multiple copies of) DNA from tiny samples (as little as a single molecule). PCR is essentially an extra cellular version of natural DNA replication, except that it replicates the region of DNA of interest only. DNA is extracted and purified from the biological sample and added to a reaction mix containing all necessary reagents. The reagents include DNA oligonucleotide primers, a heat-resistant DNA replicating enzyme (*Taq* polymerase), magnesium, the four DNA nucleotides (dATP, dGTP, dTTP and dCTP) and buffer (Singer & Berg, 1991, p. 20).

The primers are homologous to DNA sequences on either side of (flanking) the DNA sequences to be amplified (i.e. the locus of interest). A DNA polymerase cannot

elongate a DNA strand. It is not possible for DNA polymerase to initiate synthesis of a new strand, even when a template molecule is present. One important implication of this principle is that DNA synthesis requires a pre-existing segment of nucleic acid that is hydrogen-bonded to the template strand. This segment is the oligonucleotide primer (Sambrook & Russell, 2001, p. 5).

The *Taq* polymerase enzyme replicates DNA, the nucleotides are the building blocks of the new DNA and magnesium and the buffer are required for the enzyme to work. Repeated cycles are used to denature the DNA (separate the strands), allow the DNA primers to attach to the flanking sequence (anneal), and to replicate the DNA sequence between the two primers (extend). Each cycle doubles the quantity of DNA of interest as shown in Figure 3 below. All other DNA in the sample becomes so relatively rare that is it irrelevant in subsequent analyses (Figure 3).

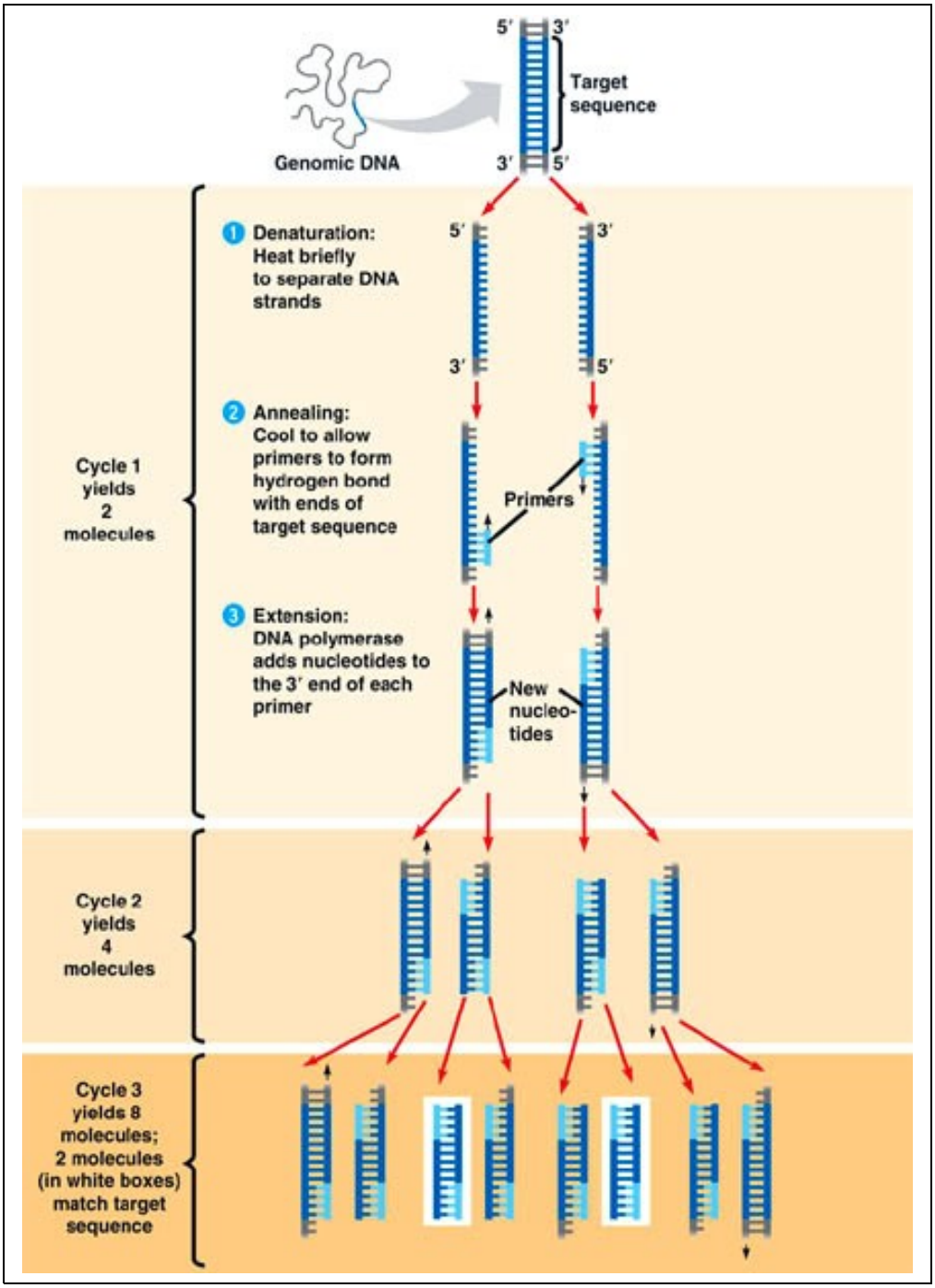


Figure 3: The PCR reaction and how it works. Adapted from http://www.bio.miami.edu/~cmallery/150/gene/Taq.htm. Retrieved on 25 November, 2009.

2.2.2 The use of molecular markers in Molecular marker assisted selection and germplasm characterization.

According to Dreher et al. (2000, p. 2), molecular markers are transmitted by the standard laws of inheritance from one generation to the next. They may be located in or near genes. Since DNA is the same in every cell, the molecular markers can be identified by a DNA test regardless of the developmental stage, age, or environmental challenges experienced by the organism. Molecular markers can be useful tools to both facilitate breeding programmes (in Molecular Marker Assisted Selection - MAS) and to aid in characterization of collections of germplasm (or varieties).

During conventional breeding, breeders cross two parent lines and select the offspring which have the most favourable combinations of traits. Breeders can map the relative positions of genes controlling visible characteristics on the chromosomes. Such a genetic map can help them to identify markers linked to traits that are selected for (Gepts, 2004, p. 5). Advances in gene technology and the growing knowledge of DNA sequences of plants offer opportunities for further genetic improvement of crops and for increasing our knowledge of crop science, production and management (Hartley, 1997, p. 3).

Scientists in many fields such as agronomy, plant breeding and quality control are encountering and using such information regularly (Azam-Ali, 1996, p. 5). The techniques of molecular biology are being used to define a more certain link between the genome and a trait of agronomic interest and to generate new combinations of genes and traits that would be unlikely to occur by conventional breeding (Gepts, 2004, p. 20).

The knowledge of genetic diversity is useful in a breeding program as it facilitates efficient sampling and utilization of germplasm resources. The breeder can use genetic variation information to make informed decisions regarding the choice of genotypes to cross for the development of populations with desired qualities (Gepts, 2002, p. 1780). The information can also be used to facilitate the identification of diverse parents to cross in hybrid combinations in order to maximize the expression of heterosis (Gepts, 2004, p. 22). Molecular markers are useful tools for assessing genetic diversity and potentially can be applicable in characterizing marama bean germplasm collections.

2.2.3 Microsatellites

Microsatellites or simple sequence repeats (SSR's) consists of direct tandem repeats of short (2-6) nucleotide motifs according to Tautz, Trick, and Dover, (1986, p. 652). For a wide range of genetic and population studies, SSR markers are a suitable

choice based on cost, labour and genetic informativeness. Microsatellites are commonly identified through:

1. Screening of small- insert or microsatellite enriched genomic libraries by hybridization with oligonucleotide primers followed by sequencing and/or
2. Searching DNA sequence databases. Database searching is only suitable for the development of SSR markers in plant species well represented in databases (Sharopova et al., 2002, p. 465).

Microsatellites have become one of the most widely used molecular techniques for genetic studies in recent years (Sharopova et al., 2002, p. 463). Enriching the AFLP or specific adaptor-amplified DNA fragments is a simple and efficient approach for SSR isolation and has been successfully applied to a number of plant genomes. Microsatellites are typically neutral, co-dominant and are used as molecular markers which have wide ranging applications in the field of genetics, including kinship and population studies (Brown, 1998, p. 287).

One common example of a microsatellite is a $(CA)_n$ repeat as shown in Figure 4 below, where n is variable between 'alleles' (alternative forms of a genetic locus); a single allele for each locus is inherited separately from each parent. In the human genome, at a locus for eye colour, the allele might result in blue or brown eyes. These SSR markers often present high levels of inter and intra specific polymorphism, particularly when tandem repeats number ten or greater.

Microsatellites are ideal for determining paternity, population genetic studies and recombination mapping (Brown, 1998, p. 287).

Microsatellite diversity is detected by amplifying DNA using PCR. Unique conserved regions, primers, flanking microsatellites are used to define the DNA segment that is to be amplified. The resulting DNA fragments are separated according to size using electrophoresis on acrylamide or agarose gels. If an individual is heterozygous for two microsatellite alleles with different numbers of repeats, then two different sized bands will be detected as shown in Figure 5 below. Three genotypes, two different homozygotes and a heterozygote are illustrated together with their banding pattern. **X** and **Y** are invariant (conserved) DNA sequences, primer sites, flanking the microsatellite repeat.

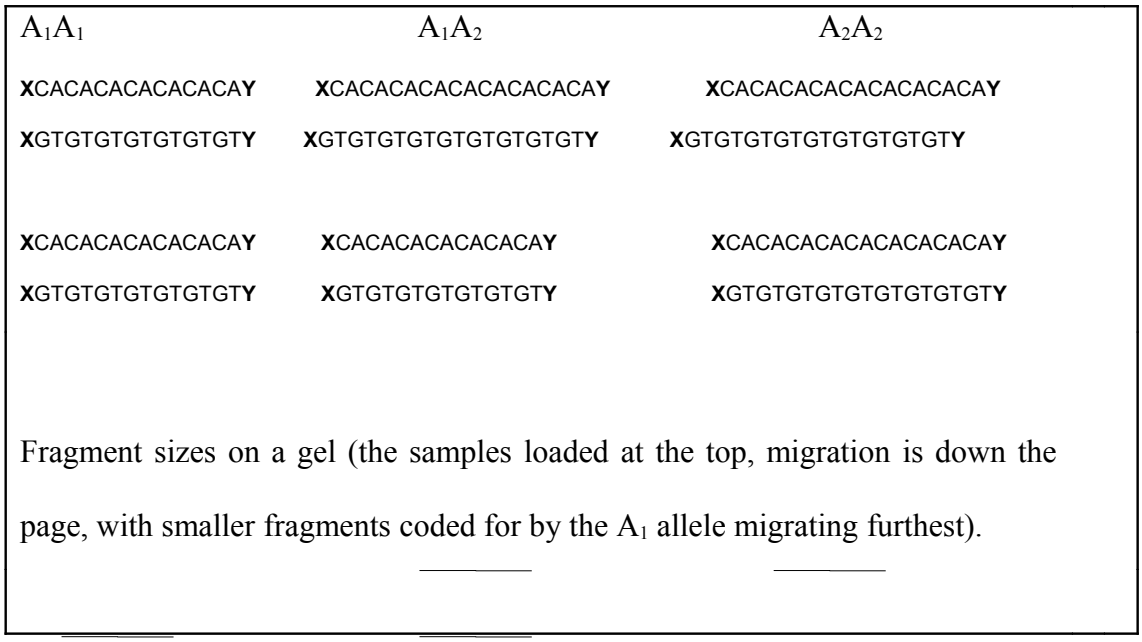


Figure 4: Allele migration for two alleles of the CA repeat.

2.2.4 FIASCO and the development of microsatellites.

Traditionally, microsatellites have been isolated from partial genomic libraries of the species of interest, screening several thousands of clones through colony hybridization with repeat containing probes (Rassmann, Schlötterer & Tautz, 1991, p. 113). Although relatively simple, especially for microsatellite rich genomes, this approach can turn out to be extremely tedious and inefficient for species with low microsatellite frequencies. Therefore several alternative strategies have been devised in order to reduce the time invested in microsatellite isolation and to significantly increase yield. The most recently proposed Fast Isolation by AFLP of Sequences Containing Repeats (FIASCO) technique (Zane, Bargelloni & Patarnelo, 2002, p. 2) was used in this study with some modifications.

In 2002, Zane et al., proposed the FIASCO procedure. In their tests, the percentage of clones containing dinucleotide repeats varied from a minimum of 50% (*Passera lagia*) to a maximum of 95% (*Sparus aurata*). The method is fast and simple and many unnecessary steps in previously proposed methods have been eliminated. The protocol relies on the extremely efficient digestion ligation reaction of the AFLP procedure. DNA is simultaneously digested with restriction enzymes and ligated with adaptors. The digestion ligation mixture is diluted and amplified. Following amplification, DNA is hybridized with biotinylated microsatellite probes. DNA molecules hybridized to biotynylated probes are selectively captured by streptavidin beads (Zane et al., 2002, p. 13).

The beads-probe-DNA complex is separated by a magnetic field from the hybridization buffer which is discarded. Non specific DNA is removed by non stringency and stringency washes. The DNA is then separated from the beads-probe complex and stored. The elutions harbouring increased proportions of DNA fragments containing selected repeats are amplified with the respective adaptor and agarose gel visualizations display smears.

The PCR products of the elution steps are the best candidates for producing highly enriched microsatellite libraries, because they are likely to contain the largest proportion of repeat containing fragments. PCR products can then be cloned before sequencing or sequenced directly (Zane et al., 2002, p. 14; Santana et al., 2009, p. 217).

2.3.1 Chloroplast DNA and chloroplast genes

In plant and fungal cells, DNA is not only contained in the nucleus, but also in the cytoplasm. Plants posses a chloroplast (cpDNA) and a mitochondrial genome (mtDNA), fungi only have mtDNA. These cytoplasmic genomes are much smaller than the nuclear genome. Plant cpDNA has a size of about 150 kb. Chloroplast chromosomes lie within the stroma and a number of features of their structure

resemble prokaryotic chromosomes. The circular DNA molecules are unlike nuclear chromosomes and are not complexed with histones. The entire chloroplast chromosomes of a number of plants including tobacco, rice and the liverwort (*Marchantia polymorpha*) have been sequenced (Hughes, 1996, p. 37).

Chloroplast DNA in plants and mitochondrial DNA in animals and plants are both uniparentally transmitted in contrast to the biparentally transmitted nuclear DNA. In most cases, cpDNA and mtDNA are transmitted maternally; one well-known exception is the paternal transmission of cpDNA in gymnosperms. Sequence analysis of the chloroplast genome has proved to be very valuable for taxonomic and phylogenetic studies at an interspecific or intergeneric level. However only rarely has intraspecific variation been encountered at levels that are high enough for population studies. The chloroplast genome for *Cuscuta exaltata* is shown in below (Figure 5).

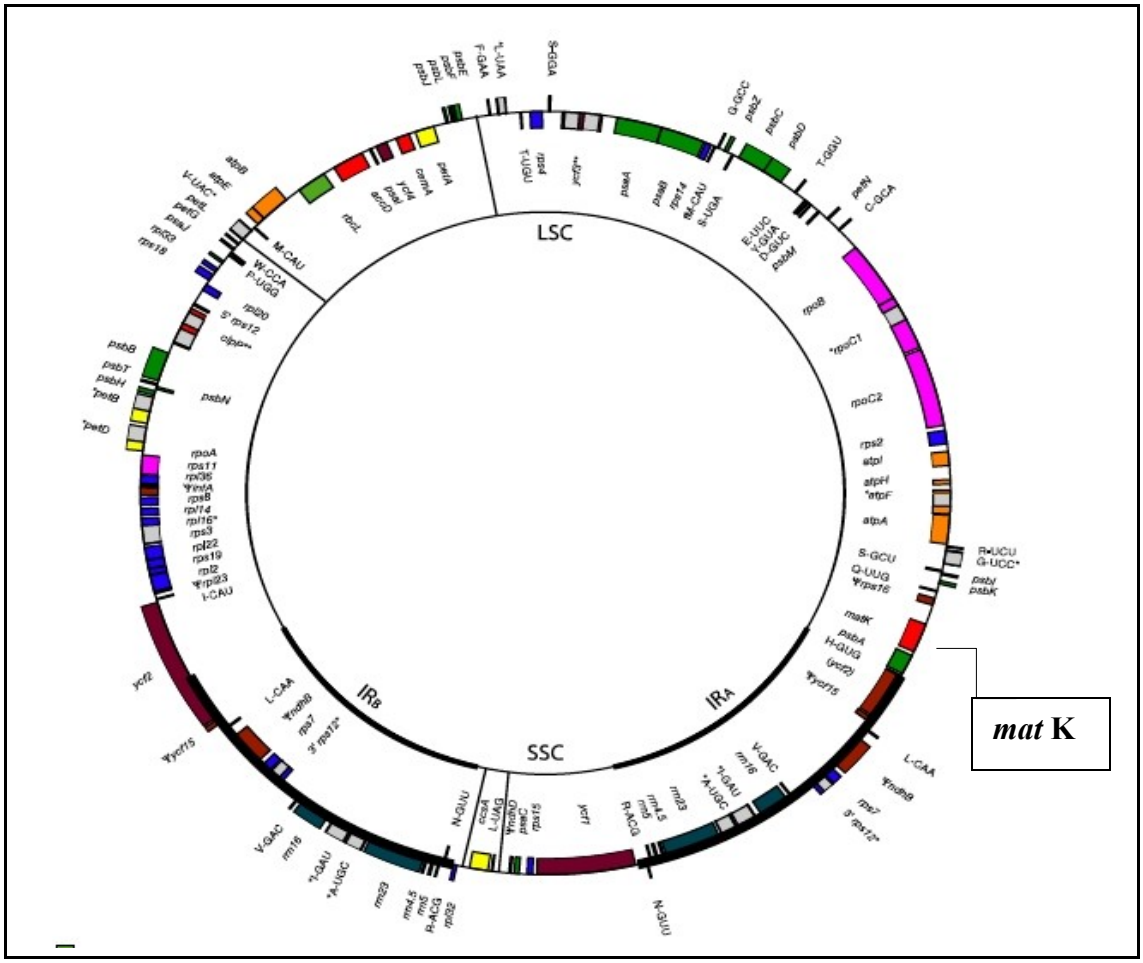


Figure 5: A genome map representing the chloroplast genes in *Cuscuta exaltata* (McNeal, Kuehl, Boore & de Pamphilis, 2007, p. 5).

2.3.2 DNA barcoding and *mat* K

The Consortium of the Barcode of Life (CBOL) is an international collaboration of natural history museums, herbaria, biological repositories and biodiversity inventory sites together with academic and commercial experts in genomics, taxonomy electronics and computer science. The mission of CBOL is to rapidly accelerate compiling of DNA barcodes of known and newly discovered plant and animal

species, establish a public library of sequences linked to named specimens and promote the development of portable devices for DNA barcoding. The *mat*K gene of marama was sequenced and characterized as a contribution to the greater global effort of barcoding.

DNA barcoding is based on the premise that a short standardized DNA barcoding sequence can distinguish individuals of a species because genetic variation between species exceeds that within species (Hebert, Cywinska, Ball & de Waard, 2003, p. 313). Species identification through barcoding is usually achieved by the retrieval of a short DNA sequence- the "barcode"- from a standard part of the genome (i.e. a specific gene region) from the specimen under investigation. The barcode sequence from each unknown specimen is then compared with a library of reference barcode sequences derived from individuals of known identity. A specimen is identified if its sequence closely matches one in the barcode library. Otherwise, the new record can lead to a novel barcode sequence for a given species (i.e. a new halotype or geographical variant), or it can suggest the existence of a newly encountered species (Hebert et al., 2003, p. 314).

A 650- base fragment of the 5' end of the mitochondrial gene cytochrome c oxidase I (COI, cox1) has gained designation as the barcode region for animals. Barcoding projects typically involve gathering specimens of a given taxonomic group (identified by conventional taxonomic methods such as morphology), cataloguing

them together with collateral data such as photographs and locality information and assembling the barcode library (i.e. a 650-base segment of the COI gene) (Hajibabaei et al., 2005, p. 1960).

The analysis of DNA barcoding data is usually performed by a clustering method such as distance based neighbour-joining (NJ) and by evaluating genetic distances within and between species. In phylogenetic studies, DNA barcoding can be a starting point for optimal selection of taxa and barcode sequences can be added to the sequence data set for phylogenetic analysis. In population genetics investigations, DNA barcodes can provide a first signal of the extent of population divergences and will facilitate comparative studies of population diversity in many species (Hebert et al., 2003, p. 314).

There has been no agreement on which region(s) should be used for barcoding land plants. To provide a community recommendation on a standard plant barcode, the CBOL Plant Working Group compared the performance of the 7 leading candidate plastid DNA regions (*atp*F- *atp*H spacer, *mat*K gene, *rbc*L gene, *rpo*B gene, *rpo* C1 gene, *psb*K- *psb*l spacer and *trn*H-*psb*A spacer). Based on assessments of recoverability, sequence quality and levels of species discrimination, the 2-locus combination of *rbc*L and *mat*K was recommended as the plant barcode. This core 2-locus barcode will provide a universal framework for the routine use of DNA sequence of DNA sequence data to identify specimens and contribute toward the

discovery of overlooked species of land plants (CBOL Plant Working Group, 2009, p. 12794).

Many chloroplast, mitochondrial and nuclear genes have been utilized for studying sequence variation at genus level. Among these genes *rbcL* gene sequences have been analysed by various workers to address plant systematics (Chase et al., 1993, p. 528). DNA sequences of the gene '*mat*K' differ among plant species, but are nearly identical in plants of the same species. This means that the *mat*K gene can provide scientists with an easy way of distinguishing between different plants, even closely related species that may look the same to the human eye. The *matK* gene of chloroplasts is 1500 bp long, located within the intron of the *trnK* and codes for a maturase like protein, which is involved in Group II intron splicing, *matK* is the only maturase of higher plant plastids (Vogel, Hübschmann, Börner & Hess, 1997, p. 179).

A homology search for this gene indicates that the 102 amino acids at the carboxyl terminus are structurally related to some regions of a maturase-like polypeptide (Khidir & Hongping, 1997, p. 830). The presence of the gene in the parasitic *Epifagus,* a taxon that lost about 65% of its chloroplast genes, speaks for the functional significance of the *mat*K gene in plants. In fact the two exons of the *trn*K gene that flank the *mat*K were lost leaving the gene intact (Wolfe, Morden, & Palmer, 1992, p. 10648). The gene contains high substitution rates within the species

and is emerging as a potential candidate to study plant systematics and evolution (Notredame, Higgins & Heringa., 2000, p. 217). The *mat*K gene may be able to contribute to genetic diversity studies in marama bean.

The *matK-trnK* gene complex is commonly used for plant evolution studies and addresses the solution for various taxonomic levels. The *matK* gene has ideal size, high rate of substitution, large proportion of variation at nucleic acid level at first and second codon position, low transition/transversion ratio and the presence of mutationally conserved sectors. These features of the *matK* gene are exploited to resolve family and species level relationships (Selvaraj, Sarma & Sathishkumar, 2008, p. 24).

Polymorphism of chloroplast DNA especially *trnK*, *matK* and intergenic *trnL* - *trnF* regions has been used to study the phylogeny of various plants. Combined analyses of *matK* and other rapidly evolving DNA regions with available multigene data sets have strong potential to enhance resolution and internal support in deep level plant phylogenetics and provide additional insights into plant evolution. It has been found that when one plant species is closely related to another, differences are usually detected in the *mat*K DNA (Lahaye, van der Bank & Bogarin, 2008, p. 2932).

Biologists often identify species by comparing their DNA sequences with those in GenBank. The public DNA database GenBank is a fundamental resource for

molecular biology, evolutionary ecology and ecology. At the moment only the original depositors of a DNA sequence accession in GenBank can modify the data associated with that DNA (Bidartondo, 2009, p. 6). The *mat*K gene is among the genes put forward as one of the suggested genes to be used in plant barcoding. The *mat*K gene for *Tylosema esculentum* had not been characterized to date so this region in the genome of the species was also explored.

2.4 Statement of the problem

Before large scale cultivation or domestication of marama bean could be undertaken, more basic biological information about the growth characteristics and genetic variability of the Namibian marama germplasm was needed. In particular, for assessing possible domestication, information on the genetic diversity for important yield parameters including flowering time and flower number, seeds per pod and pods per plant are needed. Therefore, there was an urgent need to develop and apply modern biotechnology tools with the aim of the conservation of the biodiversity of this native plant and to achieve the domestication of marama bean with the subsequeunt development of useful cultivars for planting in the Southern Africa region and even globally as some accessions may be disappearing in the wild due to human exploitation.

The development of microsatellite DNA into useful markers for a genetic diversity study sought therefore to initially address the problem of identification of genetic

diversity of the marama bean in Namibia. This will be part of a strategy to domesticate marama bean as a new cultivated crop in Namibia and Southern Africa in the Kalahari agro-ecology areas especially in the face of the current food shortages and climate change in the region. It is important to emphasize that before this study microsatellites in marama bean had not been extensively isolated or characterized. The molecular markers required to accelerate germplasm characterization of marama bean were not available. There had been about 6 microsatellites isolated from marama bean (Cullis, 2008, personal communication). The AFLP and RAPD techniques used in previous projects by Naomab (2004, p. 5) as well as Halloran and Monaghan (1996, p. 289) were unable to differentiate populations. Variation was found within but not between populations.

A possible reason for the problem with marama bean germplasm characterization is that the markers used could not detect between population variation. Microsatellites were anticipated to be able to detect between population variation and thus differentiate the different sub-populations because a typical microsatellite marker has more variants than those from other marker systems (Dreher et al., 2000, p. 12). Furthermore, the initial material used in previous studies on marama bean came from a small area and few sub-populations were assessed. The work done with RAPDs and AFLPs gave the typical results expected of an outcrosser (Dolanska & Curn, 2004, p. 95) that is more variation within rather that between populations.

2.5 Objectives

The overall aim of the project was to identify and isolate microsatellites and the *mat*K gene from marama for the detection of genetic diversity in marama bean. Specific objectives were to:

1. Isolate microsatellites (SSR's) from *Tylosema esculentum* using the Fast Isolation by AFLP of Sequences Containing Repeats (FIASCO) technique.

2. Use the isolated microsatellites to design microsatellite primers for marama bean DNA amplification.

3. Compare the products of amplification with the microsatellite primers using isolated genomic DNA from marama collected in Namibia and South Africa and identify polymorphic microsatellite loci.

4. Isolate and characterize the *mat*K gene in *Tylosema esculentum*.

In this study 11 sub populations of marama bean were assessed from a wide geographic range in the north Eastern regions of Namibia where the plant has previously been reported to be occurring. A collection from Pretoria was later added to the collection to make at total of 12 sub populations. This project isolated microsatellites from *Tylosema esculentum* using individuals from the Namibian germplasm in selected sites of the Otjozondjupa, Khomas and Omaheke regions of Namibia. This was done in order to establish a microsatellite library. In addition, the

*mat*K gene was also explored for its potential to study the genetic diversity in marama bean.

2.6 Research hypothesis

The research hypothesis that was investigated in this study is that microsatellites could be developed into useful molecular markers for *Tylosema esculentum* as all eukaryotic species contain microsatellite regions in their genomes and similar studies have been carried out in other species. It was further hypothesized that the *mat*K gene which is a chloroplast gene and has been characterized in other plant species and used in molecular systematics to resolve phylogenies would be present in *Tylosema esculentum* as well and possibly be able to contribute to genetic variation studies of marama bean and establish a DNA barcode for marama bean.

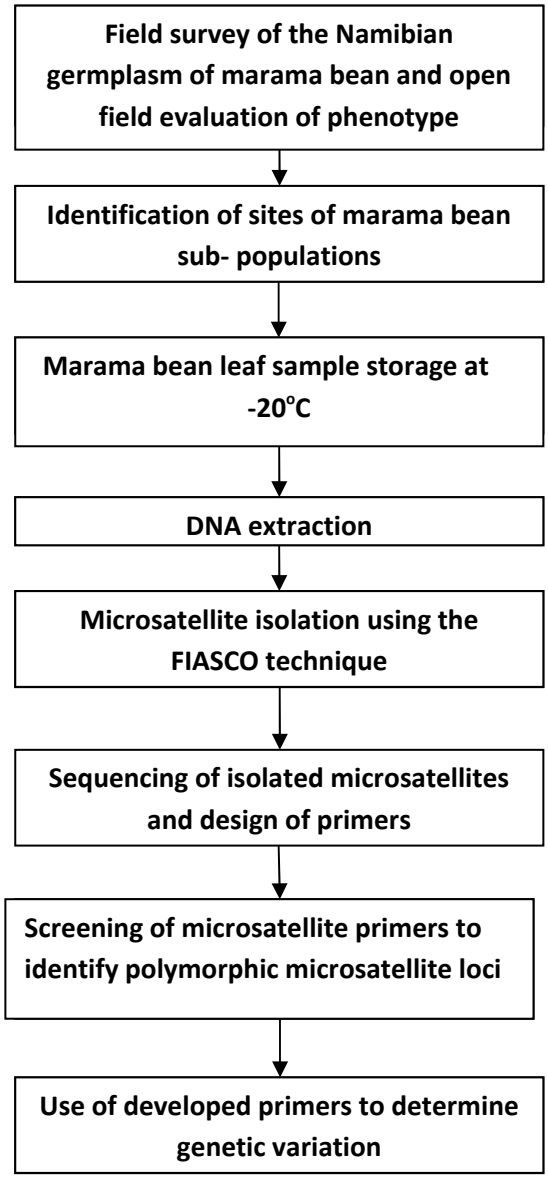
2.7 Significance of the study

The microsatellite or SSR technique has gained rapid acceptability because of its codominant nature, reproducibility and high information content. In this study an attempt was be made to isolate microsatellites from the marama bean germplasm and 80 primer pairs were developed. The polymorphic primers obtained for the isolated microsatellites will be used in diversity studies, germplasm characterization and

genotype identification. The primers that were designed were needed as tools for molecular marker assisted selection in the domestication of marama bean. If domestication of marama bean is to be realized, this baseline data is required to see if high genetic variation exists between populations from the different regions or significant variation is only within the sub-populations. This information is needed to be able to determine if breeding for particular traits will require crossing individuals from different localities or there is sufficient variation within the populations.

CHAPTER 3: MATERIALS AND METHODS

The flow chart below gives an outline of the research design that was used to isolate and characterize microsatellite loci in this study:



3.1 Sampling sites

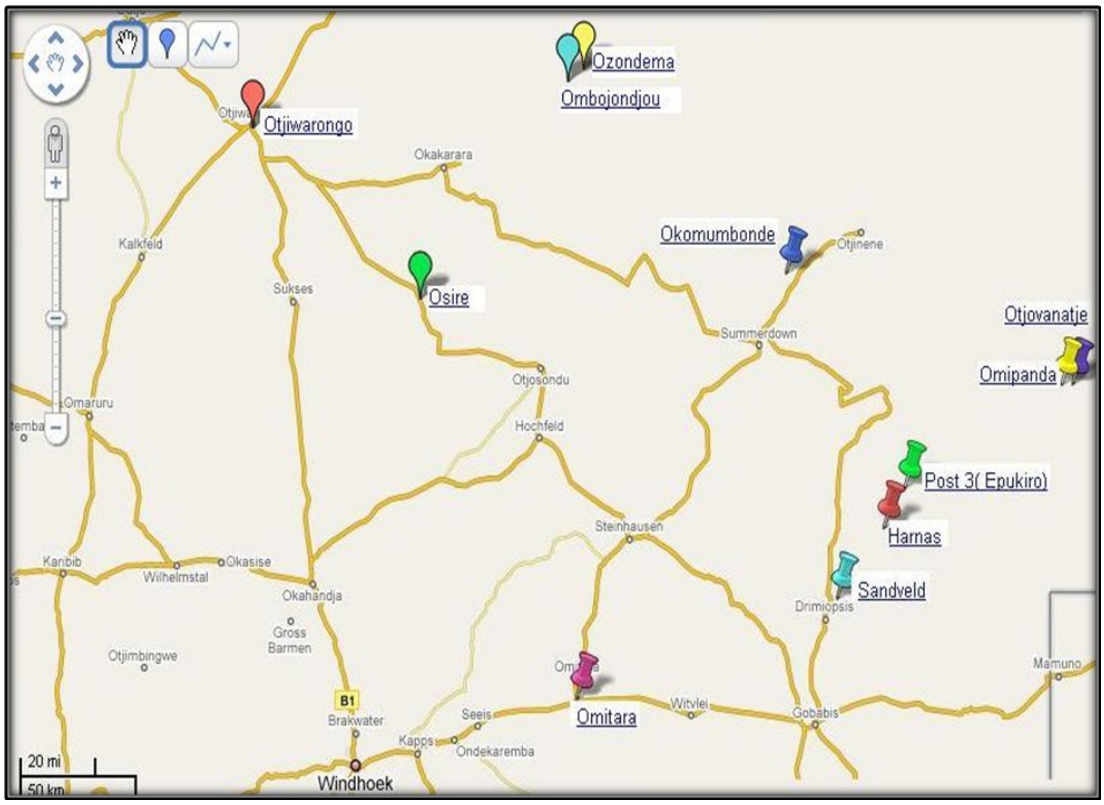


Figure 6: A map showing the study sites in Namibia.

The sampling area covered a wide geographic range of about 166 250 km^2 in the north Eastern regions of Namibia. The sampling sites were near or around village or farming settlements.

3.2 Sampling strategy

Leaf material was collected from the Omaheke, Khomas and Otjozondjupa regions of Namibia where *Tylosema esculentum* is known to occur. The sites sampled were as follows:

Otjozondjupa: Ozondema, Ombujondjou, Osire, Otjiwarongo, Khomas: Omitara, Omaheke: Sandveld, Otjovanatje, Omipanda, Post 3, Harnas, Okomumbonde. At least 20 individual plants within each sub-population were sampled for leaf material. Plants that were at least 10 metres apart were sampled to avoid multiple sampling and plants with as many phenotypic differences visible to the eye were selected. The sampling areas at each locality were about 1 km^2. The phenotypic differences included were, overall size of the plant, length of the creepers, leaf size, internode length and number of seeds per pod. A total of 361 *T. esculentum* individuals representing 11 populations described in Nepolo et al. (2009) were sampled in the Namibian germplasm (Figure 6, Table 2) and a twelfth site Pretoria was also included following a visit to the University of Pretoria. Leaf material was stored at –20 °C in the laboratory following field collections.

Table 2: List of accessions used in this study and their places of origin.

Site	GPS	Number of accessions	Country
Omipanda (OMP)	S21 19.355 E20 04.553	31	Namibia
Ozondema (OZO)	S20 15.921 E18 02.490	26	Namibia
Omitara (OMI)	S22 21.596 E18 02.476	19	Namibia
Osire (OSI)	S21 02.031 E17 21.244	60	Namibia
Harnas (HAR)	S21 47.705 E19 19.921	25	Namibia
Sandveld (SAN)	S22 01.751 E19 08.009	21	Namibia
Otjovanatje (OTJ)	S20 27.393 E16 39.443	20	Namibia
Otjiwarongo (OTR)	S20 46.092 E16 65.123	40	Namibia
Ombujondjou (OMB)	S20 18.600 E17 58.525	40	Namibia
Okomumbonde (OKO)	S20 57.000 E18 55.000	44	Namibia
Epukiro (EPK)	S21 39.642 E19 25.092	30	Namibia
Pretoria farm (UP) , (PTA)	S25 45.490 E28 11.368	5	South Africa

3.3 DNA extraction

DNA was extracted from each of the plant samples collected from the 12 sampling sites using the DNeasy mini protocol for purification of genomic DNA from plant tissue (Appendix). The manufacturer's protocol was followed to obtain DNA from the leaves (Qiagen, 2006) and the DNA was stored in clearly labelled microcentrifuge tubes at -20 ^{o}C. DNA with a concentration of 25- 250 µg/µl was collected. The concentration was determined on a 1% agarose gel stained with ethidium bromide using known molecular weight standards and also using a

spectrophotometer. DNA samples were then diluted accordingly to get equal concentrations of 10 ng/µl.

3.4 Microsatellite isolation using a modified FIASCO technique

The genomic DNA extracted was enriched for microsatellites by a modified FIASCO technique (Zane et al., 2002) without cloning as described below. Nine marama bean microsatellite libraries enriched for $(AAG)_7$, $(GTT)_7$, $(AGG)_7$, $(GAG)_7$, $(CA)_{10}$, $(CT)_{10}$, $(TCC)_7$, $(CA)_{15}$ and $(CAC)_7$ were created using the modified FIASCO technique described below:

3.4.1 Restriction enzyme digests and purification

One microgram of DNA was digested with the enzymes, *Msp* 1, *Csp* 6I and *Sau 3A* as per the supplier's instructions. The digest was then cleaned using the Qiagen PCR purification kit.

3.4.2 Adaptor ligation and amplicon preparation

The ligation was performed with 500 ng of the restriction-digested DNA, 1 µl of 12 mer adapter, 1 µl 24 mer adapter, 3 µl ligation buffer in a final volume of 28 µl. The reaction mixture was heated to 72 °C for 3 min then cooled by one degree per minute to 4 °C. 2 µl ligase was added and the reaction incubated at 4 °C for 16 hours. The 12 bp adaptors were removed by heating to 72 °C for five minutes to melt off the 12 mer followed by purification using the QIAquick PCR purification kit. Next, the

ligated DNA was amplified with PCR by combining 5 µl PCR buffer, 5 µl (20 mM) $MgCl_2$, 4 µl (10 mM) dNTPs, 2 µl adapter (100 µM), 34.75 µl water, and 1 µl ligated DNA. The reaction was heated at 72 °C for 5 min, 5 units of *Taq* polymerase enzyme was added incubated for 5 min at 72 °C. The DNA was amplified for 20 cycles of 95 °C for 30 seconds, and 72 °C for 90 seconds with a final hold at 72 °C for five minutes. Following the PCR, another QIAquick PCR cleanup was performed.

3.4.3 SSR enrichment

The SSR enrichments were repeated four times (with each enzyme of the preparation steps) with $(AAG)_7$, $(GTT)_7$, $(AGG)_7$, $(GAG)_7$, $(CA)_{10}$, $(CT)_{10}$, $(TCC)_7$, $(CA)_{15}$ and $(CAC)_7$ biotinylated primers. First the amplified digest was denatured and annealed to the biotinylated primer by combining 20 µl PCR product (200 µg) and 1 µl primer (10 µM) and heating at 95 °C for 5 min, followed by incubation at room temperature for 30 min. Before combining the primed DNA with streptavidin beads, 10 µl of unrelated DNA (sheared herring sperm at 1 mg/ml) was added to minimize non-specific binding. The annealed DNA mixture was then added to 1 mg of magnetic beads and incubated for 30 minutes at room temperature, allowing the streptavidin beads to join with the biotinlyated primers. Five washes with TEN100 (Tris/EDTA/NaCl) and 5 washes with SSC 0.2X SDS 0.1 % were performed to remove non-specific DNA. Next, two denaturation steps were performed to separate DNA containing SSRs from the beads. The first denaturation was done by adding 50 µl of TE (Tris-HCl 10 mM, EDTA 1 mM) and heating to 95 °C for 5 minutes. The remaining solution was separated magnetically and stored. The second denaturation used 12 µl 0.5N NaOH, which was neutralized with 12 µl 0.5N HCl and separated magnetically. Each denaturation product (2 per enzyme) was amplified separately

with PCR by adding 5 µl PCR buffer, 3 µl $MgCl_2$ (25 mM), 4 µl dNTPs (each 2.5 mM), 2 µl adapter (10 µM, 34.75 µl water, 0.25 µl r*Taq* (2 units) and 1 µl DNA into a PCR tube. The mixture was cycled 20 times from 95 °C to 72 °C. Electrophoresis was then performed on each of the denaturation products with a 1.5% agarose gel.

3.4.4 Sequencing

All the microsatellite-enriched genomic DNA pools were combined and 5 µg of the extracted microsatellites were sequenced and analyzed with the Roche 454 GS-FLX platform at Inqaba Biotech (Pretoria, South Africa). Sample preparation and analytical processing such as base calling, were performed at Inqaba Biotech using the manufacturer's protocol as described previously (Santana et al., 2009). The sequence information gathered was sorted into contig and single read files that were used for SSR identification in the sequences and subsequently in primer design.

3.4.5 Microsatellite discovery and Primer design

Primers were then developed around the SSR sites, identified with the Simple Sequence Repeat Identification Tool (SSRIT) software. Primer 3 was used to design the microsatellite primers and the primers were synthesized by Inqaba Biotech. SSRIT, which is an SSR finding program,

(http://www.gramene.org/db/markers/ssrtool) was used to identify the microsatellites in the contig files obtained from 454 sequencing. The search was made to include SSR's up to pentamers with a minimum of 5 repeats required for each type of microsatellite picked out. Sequences containing perfect microsatellites were used to design PCR primers complementary to the flanking region of the microsatellites. Eighty microsatellite primers pairs were designed from the sequence data of the

libraries using Primer 3 software available online (http://frodo.wi.mit.edu/primer3). The primers were synthesized by Inqaba Biotechnology laboratory in South Africa and then used for amplification of marama DNA from the 11 different Namibian sites as well as 1 location in South Africa (Pretoria). Each microsatellite primer marker was given a name consisting of the prefix "MARA" followed by a number (001-080).

3.5 SSR Primer screening

Each of the primer sets was used to screen individual marama DNAs. Two sets of screening populations were used. The screening was accomplished by using pooled DNA from the 12 locations (11 Namibian, Table 1, plus Pretoria, South Africa) as template in one population. The 19 individual plants from the Omitara sub-population made up the second screening population used. PCR amplifications were performed in 25 µl reaction volumes, with a 2X PCR master mix from Fermentas. Each PCR reaction contained 1 µl template genomic DNA, 1 µl of SSR forward primer, 1 µl of SSR reverse primer, 12.5 µl of the 2X PCR master mix and 9.5 µl nuclease free water. The PCR reaction profile used involved an initial denaturation step of 95 ^{o}C for 4 minutes, followed by 35 cycles of denaturation at 95 ^{o}C for 30 sec, an annealing at between 55 ^{o}C and 65 ^{o}C (primer sequence dependent) for 60 sec and an extension at 72 ^{o}C for 2 minutes, a final extension at 72 ^{o}C for 5 minutes and then held at 4 ^{o}C. Agarose gel (2.5%) visualization of PCR products was then used to determine if a primer pair was polymorphic or monomorphic based on its separation of amplification products and banding patterns generated on the agarose gels in the

different DNA templates. A total of 80 microsatellite primers were screened and described as polymorphic, monomorphic or unable to amplify. The sequences from which the monomorphic primers were derived from were searched for using the BLAST tool at NCBI to see if there was any sequence homology with any genes with known functions.

3.6 SSR Data analysis

In this study, the value discriminatory power (D_L) was used to compare the efficiency of the microsatellites to differentiate among individuals using the Omitara sub-population amplification profiles. The D_L value represents the probability that two randomly chosen individuals show different allelic patterns at the same microsatellite locus and thus are distinguishable from one another. That is if p_i is the proportion of the population carrying the *i*th banding patterns at the *j*th primer and if p_i were calculated for each pattern generated by the primer (Tessier, David, This, Boursiquot & Charrier, 1999, p 172), then $D_L = 1 - \sum p_i^2$. This is an extension of the polymorphism information content (PIC) (Anderson, Churchill, Autroque, Tanksley & Sorrels, 1993, p 183), available from the frequencies of the different banding patterns (or genotypes) generated by a primer. The discriminatory power approach was applied here as was done for the most important grain legume for human consumption in the world, *Phaseolus vulgaris* (Gaitán-Solis, Duque, Edwards & Tohme, 2002, p. 2129).

3.7 Amplifying the *matK* gene in marama bean

The *mat*K primer pairs known to amplify the *mat*K region in legumes (Wojciechowski, Lavin & Sanderson, 2004, p. 1847) were used for amplification of the *mat*K gene from the geographically diverse marama bean germplasm collection from the 12 localities. Ozondema, Ombujondjou, Osire, and Otjiwarongo in Otjozondjupa region, Omitara in Khomas region, Sandveld, Otjovanatje, Omipanda, Post 3, Harnas, Okomombonde in Omaheke region and Pretoria in South Africa. The primers used were *trn*K686, *trn*K2, *mat*K 4La, *mat*K 1100L, *mat*K 1932Ra, *mat*K 832R. PCR amplifications were performed in 25 µl reaction volumes, with a 2X PCR master mix from Fermentas. Each PCR reaction contained 1 µl template genomic DNA, 1 µl of the forward primer, 1 µl of the reverse primer, 12.5 µl of the 2X PCR master mix and 9.5 µl nuclease free water. The PCR reaction profile used involved an initial denaturation step of 95 ^{o}C for 4 minutes, followed by 35 cycles of denaturation at 95 ^{o}C for 30 sec, an annealing at 55 ^{o}C for 60 sec and an extension at 72 ^{o}C for 2 minutes, a final extension at 72 ^{o}C for 5 minutes and then held at 4 ^{o}C. Agarose gel (2.5%) visualization of PCR products was then used to determine if PCR products above 1000 base pairs obtained. The PCR products were sequenced to characterize the gene for marama bean and design primers that can be used to identify the marama bean plants as part of the plant barcoding effort and possibly amplify the gene in other species in the genus *Tylosema*. A phylogenetic tree based on the *mat*K gene was also constructed to investigate if the *mat*K gene would be useful in genetic diversity studies of *Tylosema esculentum*.

CHAPTER 4: RESULTS

Eighty microsatellite primer pairs were designed from sequence data obtained using microsatellite enriched DNA fragments. The microsatellite primers were screened using a panel of DNA samples from the Omitara sub- population and pooled DNA samples from each of the 12 sub-populations. The results of the screening of the primers for polymorphic loci are presented in this section. In addition, the *mat*K gene was amplified using PCR and the gene was characterized for marama bean. A phylogenetic tree was constructed based on the *mat*K gene and this indicated that the *mat*K gene can also be used in genetic diversity studies in marama bean.

4.1 DNA extraction

DNA was extracted and diluted to a standard concentration of 10 ng/µl for all samples. Figure 7 below shows the DNA extracts from some of the samples collected in Otjovanatje. A total of 361 DNA samples were used in this study from the 361 accessions collected at the different localities throughout Namibia and the collection from Pretoria.

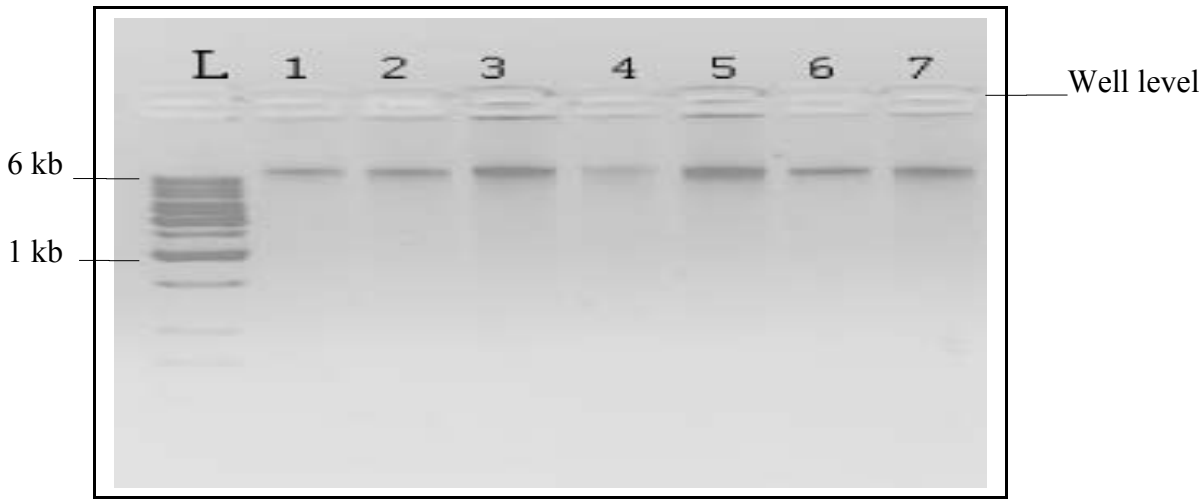


Figure 7: The DNeasy miniprotocol for purification of DNA was used to extract DNA from each of the plant samples. The picture shows sample 1-7 from Otjovanatje following electrophoresis on a 1% agarose gel. Lane L is a 1 kb ladder

4.2 Microsatellite isolation and enrichment

The enrichment products obtained following the FIASCO protocol described above are shown in Figure 8 below. The products appear as smears as they carry many different sized fragments enriched for the microsatellite repeats.

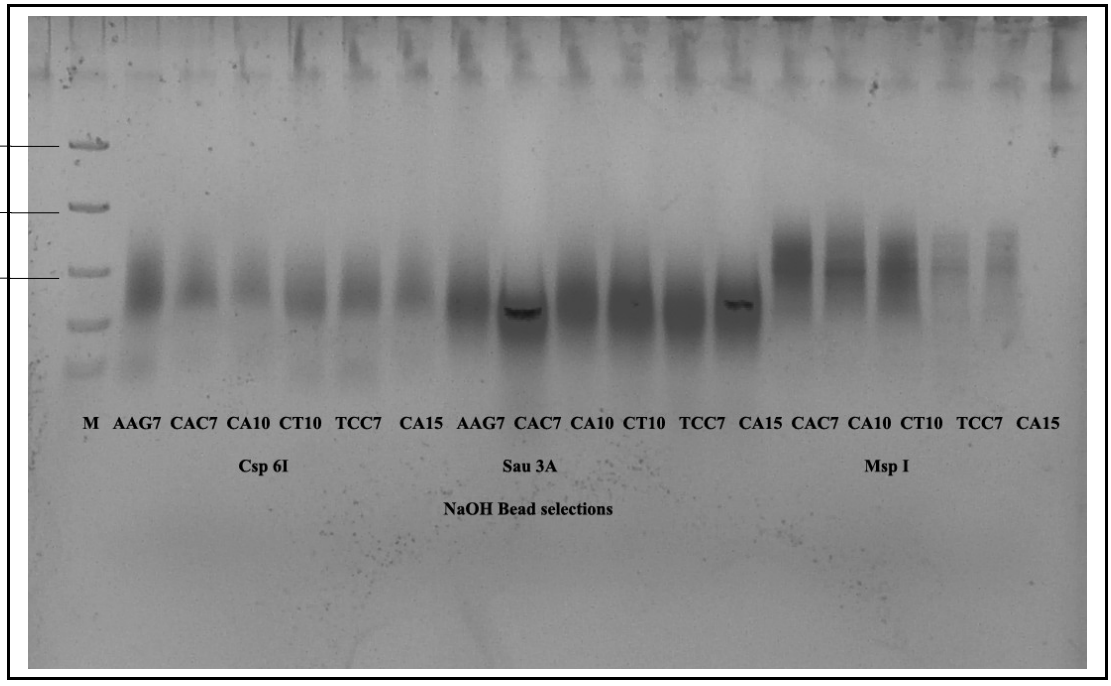


Figure 8: Agarose gel electrophoresis of microsatellite enriched DNA. A 100 bp ladder was used in lane M and the rest of the lanes contain enriched fragments carrying microsatellite DNA. The smears show the many different sized fragments present.

4.3 Microsatellite primer design

Microsatellite loci were isolated from the marama bean germplasm using the modified FIASCO enrichment technique and used to design 80 microsatellite primers based on perfect microsatellites. The output files from the primer design process were as shown in Figure 9 below. Each of the contig or single read files that had microsatellites with sufficient flanking sequences for primer design on both the 5' and 3' ends were used to design primers.

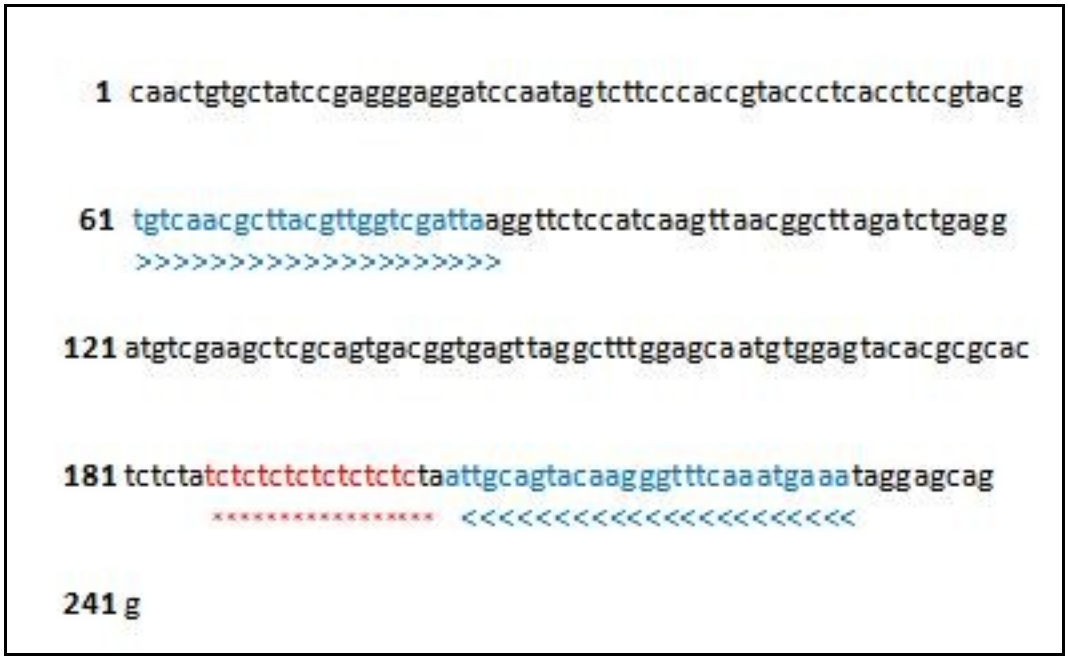


Figure 9: Primer 3 output for the design of primer pair MARA 011F and MARA011R to amplify the $(TC)_8$ microsatellite locus.

4.4 Microsatellite primer screening (amplification profiles)

A total of 80 primers were screened for polymorphism. The primers were then grouped according to the type of amplification patterns generated. The primers were screened using the panel of 12 ecotypes of marama bean and the Omitara sub-population. Microsatellite loci were described (Figure 10, Figure 11, Figure 12, Figure 13, Table 2) as: Group1- Monomorphic (for example *MARA 004);* Group 2- Polymorphic in typical SSR type pattern (for example *MARA 001)*; Group 3- Polymorphic but not SSR type pattern (for example *MARA 045)*; Group 4-Not amplifying some individuals or poor quality amplification products (for example *MARA 067)*. Figure 10, Figure 11, Figure 12 and Figure 13 below show the different gel profiles.

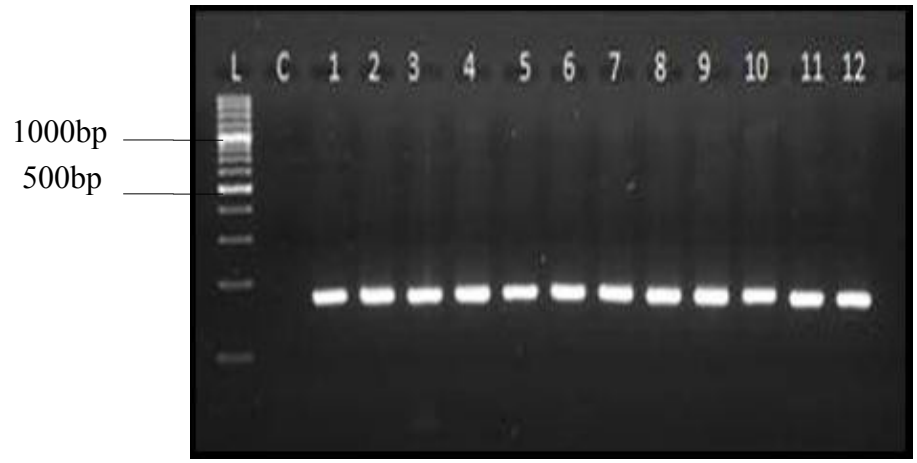


Figure 10: Primer MARA 004 a typical monomorphic (Group 1) microsatellite locus. Sample 1-12 are pooled DNA samples, lane L contains a 100 bp ladder and the gel was 2.5 %. Electrophoresis was at 100 V for 1 hour.

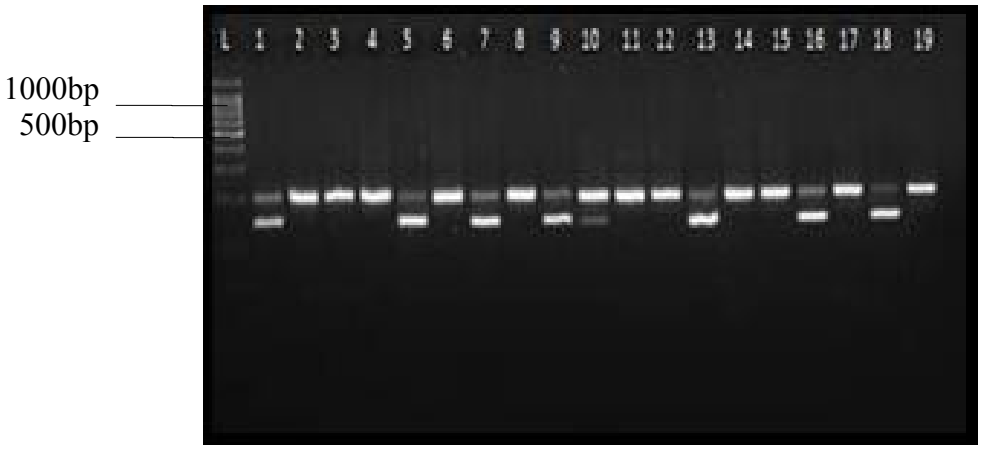


Figure 11: Primer MARA 001 a typical polymorphic (Group 2) microsatellite locus. Sample 1-19 are individuals from Omitara, lane L contains a 100 bp ladder and the gel was 2.5 %. Electrophoresis was at 100 V for 1 hour.

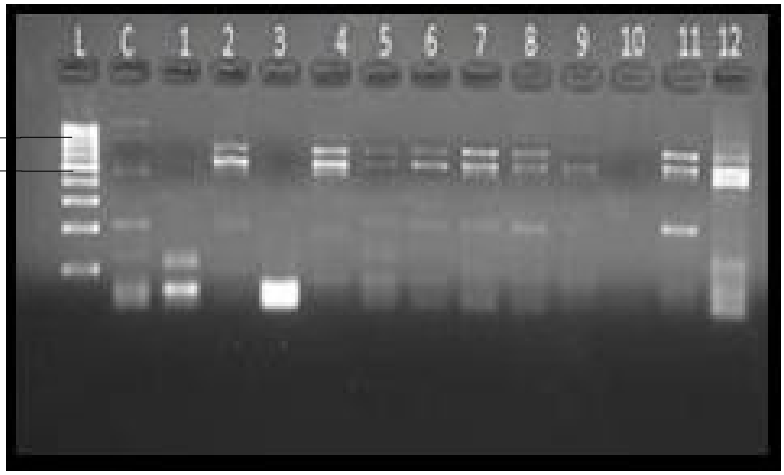


Figure 12: Primer MARA 045 a typical polymorphic (Group 3) microsatellite locus. Sample 1-12 are pooled DNA samples, lane L contains a 100 bp ladder and the gel was 2.5 %. Electrophoresis was at 100 V for 1 hour.

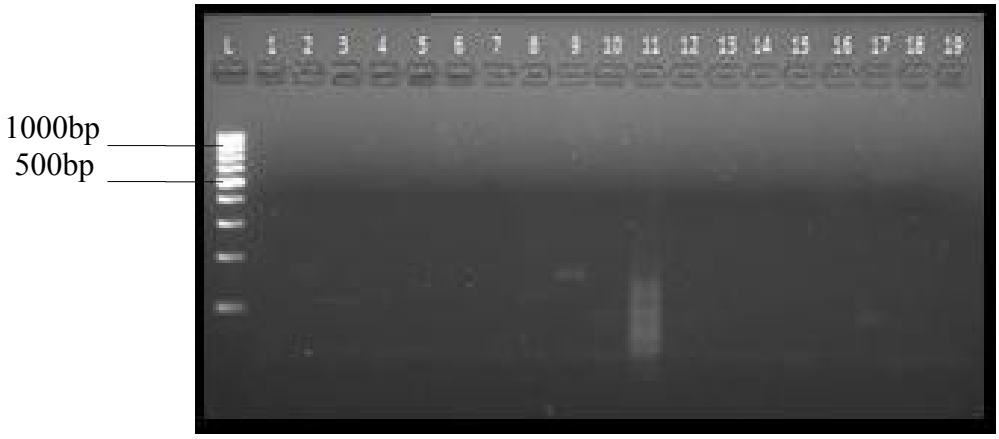


Figure 13: Primer MARA 067 a typical microsatellite locus that could not be scored or did not give any amplification. Sample 1-19 are individuals from Omitara, lane L contains a 100 bp ladder and the gel was 2.5 %. Electrophoresis was at 100 V for 1 hour.

In the group of the 80 microsatellite primers screened, 16 belonged to Group 1, 17 were in Group 2, 44 were in Group 3 and 3 belonged to Group 4

4.5 SSR data analysis

Of the 80 primers screened (76 %) were able to detect polymorphism (Group 2 and Group 3) and 20 % of them gave monomorphic bands (Group 1). The polymorphic primers can be useful in trait linkage for the marama breeding program. The remaining 3 primers out of the 80 primers did not give clearly scorable amplifications or they gave no product at all (Table 3).

The discriminatory power of each polymorphic primer D_L was calculated and ranged from 0.188 - 0.947 with an average D_L of 0.482 (Table 4). A primer with a D_L greater that 0.7 was considered to have a higher probability to be able to discriminate between two individuals. Four primer pairs (MARA 001, MARA 065, MARA 068 and MARA 077) out of the 80 primer pairs screened that gave reproducible polymorphic patterns were used for analysis of genetic variation of marama bean. Migration distance on a gel is inversely proportional to fragment length, and the standard fragments of known length ran on the gels provided a means of calibrating the gels.

Some of the sequences contained microsatellite regions but the flanking sequences on either the right or left side or both were too short for primer design. All the primers designed were made to amplify perfect microsatellites that were 10 bases or longer. Of the microsatellites targeted, two were pentanucleotide repeats, three were

tetranucleotide repeats and the remaining 75 microsatellites were dinucleotide and trinucleotide repeats. The longest microsatellite amplified was a $(CA)_{20}$ repeat amplified by primer MARA 058.

4.6 Reproducibility of SSR amplifications

It was demonstrated that SSR marker amplifications were reproducible in this study as shown by primer MARA 001 when amplifying DNA from individuals from Omitara. The PCR reactions A and B were run on different days but with the same reaction profile and same PCR ingredients.

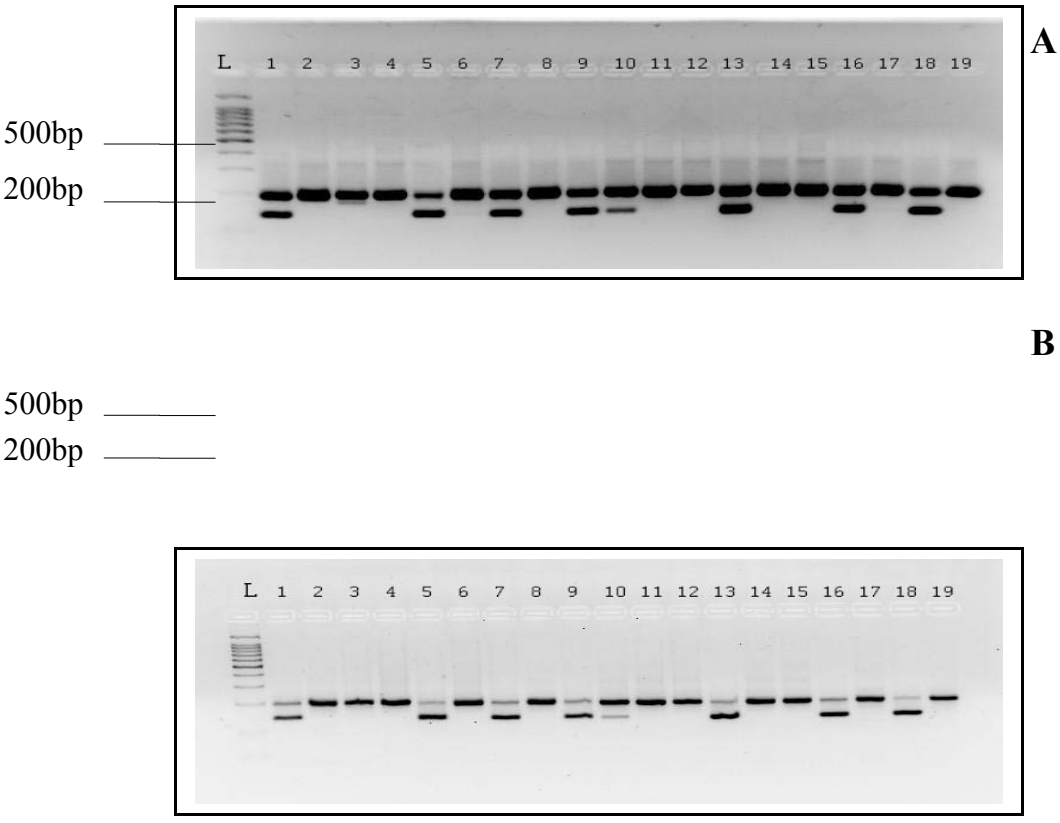


Figure 14: Reproducibility test for primer pair MARA 001. Lane 1-19 contain amplification products from Omitara DNA and Lane L contains a 100bp ladder in both A and B.

4.7 *mat*K and *trn*K amplification

The PCR products for the *trn*K genes were expected in the region of 2500bp, while for *mat*K the expected product size was in the region of 1500bp. A phylogenetic tree based on the *matK* gene was constructed to see if the gene has evolved or is the same at the 12 study sites. Figure 15, Figure 16 and Figure 17 below shows the amplification products obtained for the *trn*K and *mat*K gene following PCR amplification.

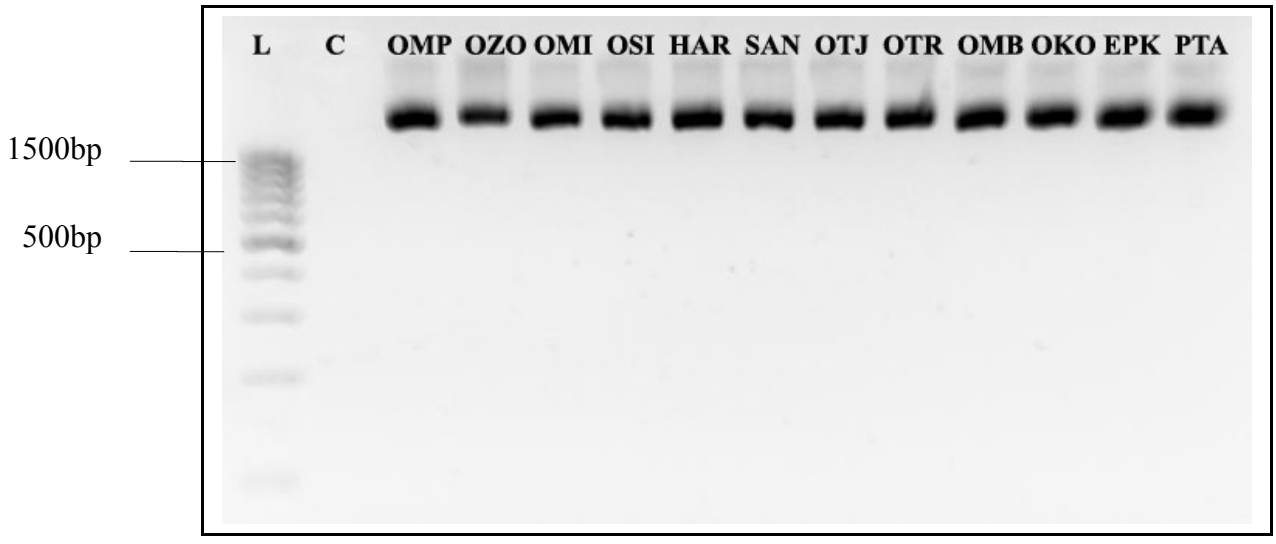


Figure 15: Amplification of the *trn*K gene in *Tylosema esculentum*. C is a negative control and L is a 100bp ladder. The rest of the lanes contain a sample from each of the 12 localities and products were more than 1500bp. The

expected product size was around 2500bp.

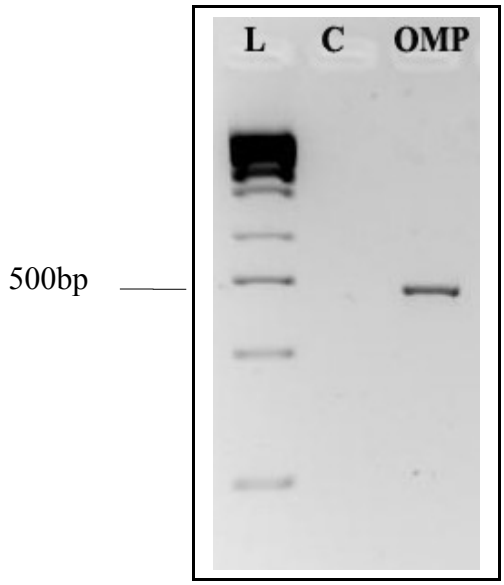


Figure 16: Amplification of the *mat*K gene in *Tylosema esculentum*. C is a negative control and L is a 100bp ladder. The lane OMP contains a sample from Omipanda which gave the longest *mat*K sequence of 637 bp.

All the sequences obtained for both the *mat*K and *trn*K region were partial sequences. The longest sequence read for the *trn*K region was 1167 base pairs long and for the matK region it was 637 base pairs. Figure 18 below shows the multiple sequence alignment of the *mat*K region in different marama bean individual plants.

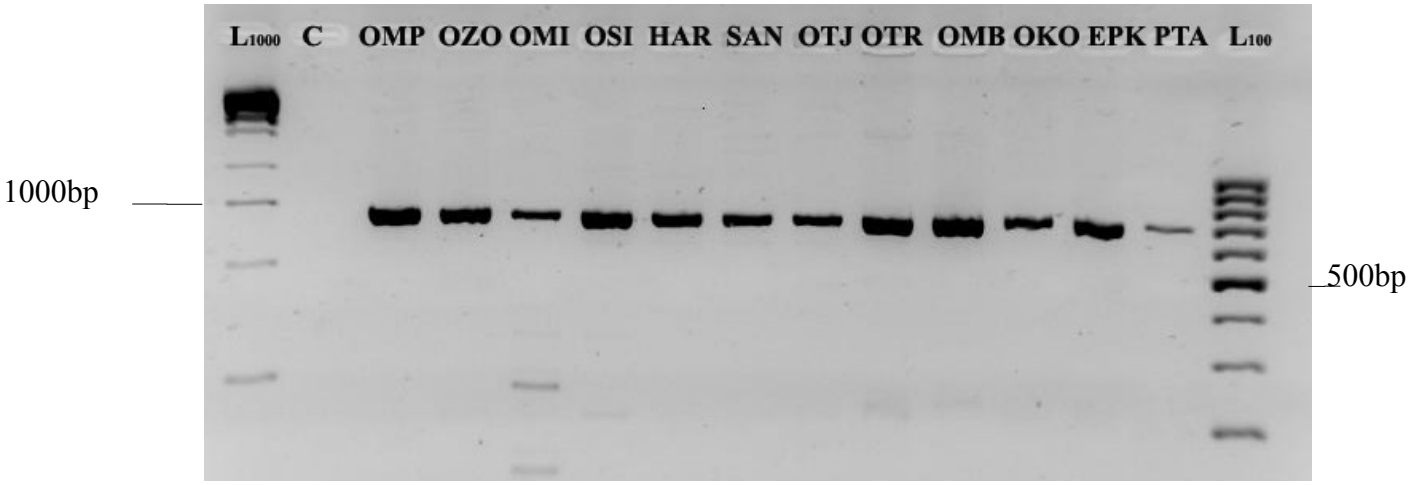


Figure 17: Amplification of the *mat*K gene in *Tylosema esculentum*. C is a negative control and L_{1000} is a 1kb ladder and L_{100} is a 100bp ladder. The rest of the lanes contain a sample from each of the 12 study sites.

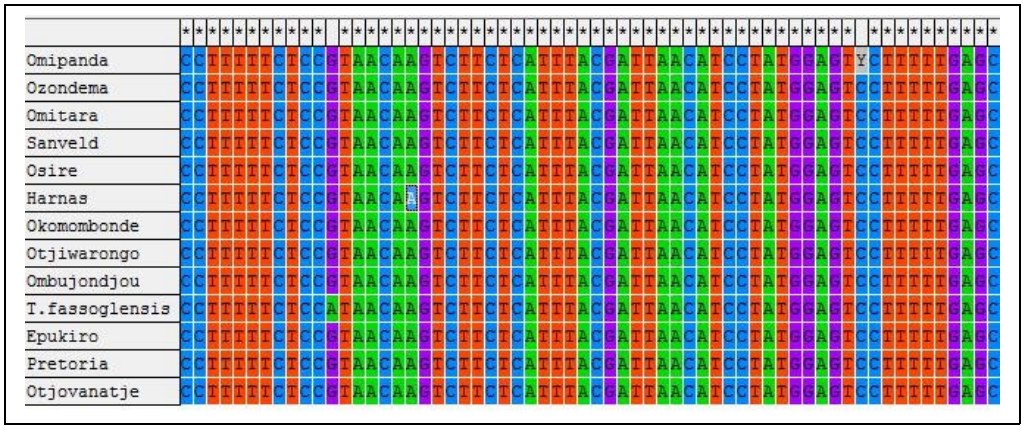


Figure 18: A section of the multiple sequence alignment of the *mat*K sequences of *Tylosema esculentum* from different localities and *Tylosema fassoglensis*.

A phylogenetic tree was constructed based on the multiple sequence alignment given in Figure 18. The phylogenetic tree is shown in Figure 19 below.

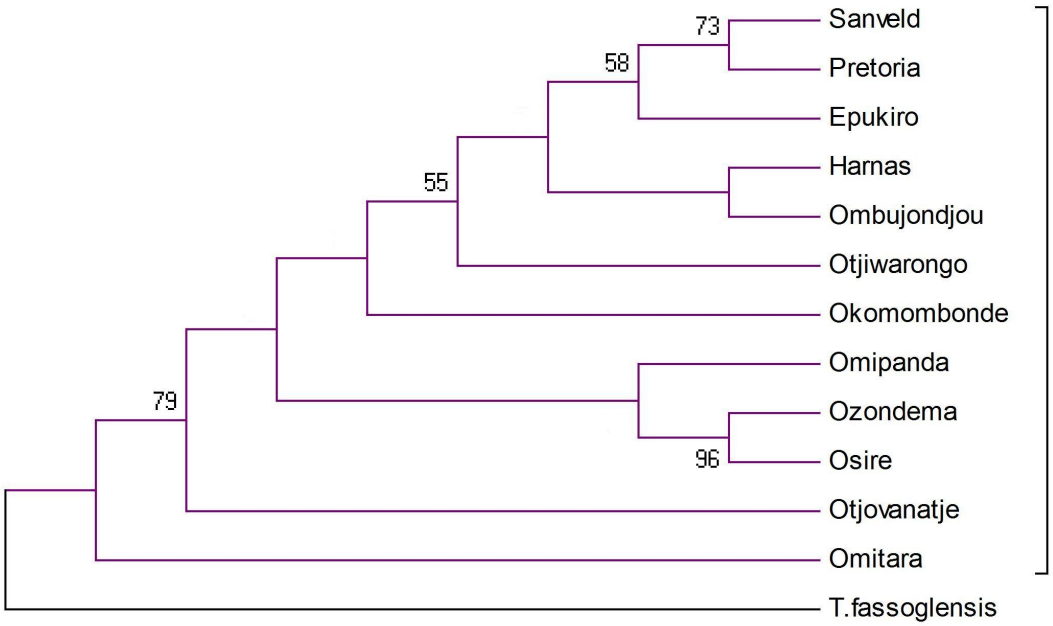


Figure 19: A phylogenetic tree for *Tylosema esculentum* based on the *mat*K gene. The bootstrap values show the confidence in the groupings as a percentage. The sequence for *Tylosema fassoglensis* was used as an out-group to root the tree generated with the software MEGA 4.0.

The phylogenetic tree was constructed based on the *mat*K sequences for *Tylosema esculentum* generated in this study. The sequence of the closely related *Tylosema fassoglensis* obtained from NCBI was used as an out-group to root the tree. The amplified fragments were about 700 base pairs long (Figure 17). The multiple sequence alignment and construction of the tree were performed using MEGA 4.0 (Figure 18 and Figure 19).

4.8 *mat* K and *trn*K BLAST alignment

The Basic Local Alignment Search Tool (BLAST) which is available at (http://www.ncbi.nlm.nih.gov/BLAST/) was used to find homologous sequences to those obtained for marama bean. BLAST uses the query sequence, which was that for marama bean in this case to search various databases to look for similar sequences. A similarity matrix was then used to measure the similarity between sequences in the database and the query sequence. Both Blastx and Blastn searches were performed.

The partial *trn*K sequence of *T. esculentum* aligned with that of *Tylosema fassoglensis* with 96% identities as shown in Figure 20 below. An alignment of the amino acid translations showed the *T. esculentum* sequence was in the middle of the C-terminus and N-terminus of the *trn*K gene with equal gaps on either side when aligned with that of *Tylosema fassoglensis*. The partial *mat*K sequence of *T. esculentum* aligned with that of *Pisum sativum* with 99% identities as shown in Figure 21 below. An alignment of the amino acid translations showed the *T. esculentum* sequence was at the N-terminus of the *mat*K gene when aligned with that of *Pisum sativum*.

```
gb|EU361874.1|  Tylosema fassoglensis voucher Herendeen 21-XII-97-
6 (US) tRNA-Lys trnK) gene, partial sequence; and maturase K
(matK) gene, complete cds; chloroplast
Length=1755

 Score = 1674 bits (906),  Expect = 0.0
 Identities = 995/1034 (96%), Gaps = 22/1034 (2%)
 Strand=Plus/Minus

Query  80    CTTTCCCTATGTATACATCTAAACCTCTGTTCCTTCGCTAAAATAGGAC-TCTAAGAAGA  138
             ||||||||||||| |||||||||||||||||||||||||||||||||| ||||||||||
Sbjct  1755  CTTTCCCTATGTCTACATCTAAACCTCTGTTCCTTCGCTAAAATAGGACTTCTAAGAAGA  1696
```

Figure 20: The BLAST result for the *trn*K sequence alignment. The full alignment is shown in the appendices.

```
gb|AY386961.1|  Pisum sativum maturase-like protein (matK) gene,
complete cds; chloroplast
Length=1521

 Score = 1162 bits (629),  Expect = 0.0
 Identities = 634/637 (99%), Gaps = 1/637 (0%)
 Strand=Plus/Minus

Query  1     AGGATCTAATTAGAGGAATAATTGGAACTATTATATCCAATTTTTYGWTAACAATTTCGA  60
             |||||||||||||||||||||||||||||||||||||||||||||| | ||||||||||||
Sbjct  1123  AGGATCTAATTAGAGGAATAATTGGAACTATTATATCCAATTTTTTGATAACAATTTCGA  1064
```

Figure 21: The BLAST result for the *mat*K sequence alignment. The full alignment is shown in the appendices.

CHAPTER 5: DISCUSSION

Microsatellites that can be used for germplasm characterization in marama been were isolated and described for the first time for the species. The products of amplification showed that a greater proportion of these microsatellites (76 %) can be used as markers for genetic diversity studies as they were able to detect polymorphism in the germplasm collection used in this study. The *mat*K gene which was sequenced and characterized for the first time as well for the species *Tylosema esculentum* can also serve as a molecular marker in genetic diversity studies of marama bean since it was able to separate the different individuals from different localities showing that there is genetic diversity between individuals from different sub-populations in the Namibian marama bean germplasm.

Microsatellite were successfully isolated and used for genetic variation analysis in *Tylosema esculentum*. A total of 80 primers were designed. Out of the 80 microsatellites primers developed, 5% of the SSR's were used for genetic variation analysis (Nepolo, 2010, in press) and these SSR primers have proved that microsatellites are useful and informative for assessing intra-specific and inter-specific variability of marama bean. The microsatellites were able to detect between population variation better than the RAPD and AFLP markers used before. The information reported here is the first for perfect microsatellites whose primers were designed and screened in this study.

Microsatellites have become one of the most widely used molecular markers for genetic studies in recent years. The FIASCO technique that was used was able to develop a number of polymorphic microsatellite primers. This further demonstrated that enriching the AFLP or specific adaptor-amplified DNA fragments is a simple and efficient approach to SSR isolation and has been successfully applied to a number of plant genomes. By employing modified procedures from other plant species, corresponding isolating protocols were established in marama bean for the development of microsatellites. The study also marked one of the first that employed Roche 454 technology for sequencing DNA libraries for microsatellites as was also done by Santana et al., 2009.

The number of samples that were used to represent the Namibian population depended on what was found in the field at the time samples were collected. Natural factors such as animal grazing, flowering time or seed germination of the plant could not be controlled by collectors. Since each individual has a unique DNA make-up, a sample of DNA isolated from an individual can be used to make a direct assessment of sequence variation in genotypes or measure the degree of relatedness. Molecular analysis is applicable at any stage in the organism's life cycle and it can make use of any part of the organism. Molecular markers such as the microsatellite markers developed here are able to determine the extent of these genetic differences (Brown, 1998, p. 23; Tait, 1997, p. 25).

SSR primers have shown high levels of polymorphism in many plants including important crops like rice (*Oryza sativa* L.) (Chen, Temnykh, Xu, Cho & McCouch, 1997, p. 553), wheat (*Triticum aestivum* L.) (Devos, Bryan, Collins, Stephenson & Gale, 1995, p. 247; Roder et al., 1995, p. 327), barley (*Hordeum vulgare* L.) (Liu, Biyashev & Saghai-Maroof, 1996, p. 869), oat (*Avena sativa* L.) (Li, Rossnagel & Scoles, 2000, p. 1259), maize (*Zea mays* L.) (Senior, Murphy, Goodman & Stuber, 1998), sorghum [*Sorghum bicolor* (L.) Moench] (Brown et al., 1996, p. 190), soyabean [*Glycine max* (L.) Merr.] (Akkaya, Bhagwat & Cregan, 1992, p. 1131, beans (*Phaseolus* and *Vigna*) (Yu, Park, Poysa, 1999, p. 27), alfafa (*Medicago* spp.) (Diwan, Bryan, Collins & Stephenson, 1997, p. 887), sunflower (*Helianthus annuus* L.) (Brunel, 1994, p. 397) and tomato (*Lycopersicon esculentum* Mill.) (Smulders, Bredemeijer, Rus-Lortekaas, Arens & Vosma, 1997, p. 264). The present study showed that microsatellite markers were also highly polymorphic in marama bean. They could distinguish between different individuals in the different accessions collected at different places within Namibia for example individual 1 and 2 from Omitara (Figure 14).

The discriminatory power of each polymorphic primer (D_L) was calculated in this study for *Tylosema esculentum* and ranged from 0.188 - 0.947 with an average D_L of 0.482 (Table 4). In the study of *Phaseolus vulgaris* microsatellites, this value ranged from 0.09 – 0.94 with an average of 0.73 (Gaitán-Solis et al., 2002, p. 2134). The average in marama was lower than that in common bean, however, 23 of the primers

had a discriminatory power greater than 0.7. This was the first report of discriminatory power of microsatellite markers for marama bean.

The polymorphism observed can be due to nucleotide substitution in DNA or variation in the number of nucleotides (DNA length) for a specific DNA region. Length polymorphisms were characterized for marama bean and one important class of these DNA length polymorphisms is those due to deletion and insertion of relatively small numbers of nucleotides. These polymorphisms are usually observed in the noncoding regions of DNA but sometimes occur also in the coding regions. Variation in PCR product length is a function of the number of SSR units hence the different size fragments that were observed following electrophoresis at the polymorphic loci were due to the presence of different numbers of the SSR units at the different loci (Nei & Koehn, 1983, p. 65; Arnold, Rossetto, McNally & Henry, 2002, p. 22).

The second class of DNA length polymorphisms is those with respect to duplicate genes, which are caused by unequal crossing over. In this case, the number of copies of a particular gene or a particular DNA sequence may vary from individual to individual (Nei & Koehn, 1983, p. 65; Arnold et al., 2002, p. 22). Since the genome of *Tylosema esculentum* is not yet well characterized the polymorphisms observed at the microsatellite loci screened here could fall into either of the two classes since

microsatellites are known to be occurring in both coding and non coding regions of eukaryotic genomes.

Polymorphism was also observed in band intensity of amplicons, where bands were either bright or faded. In the SSR screening, there were no differences noted with regard to band intensity whether pooled DNA was used or individuals from the Omitara subpopulation, polymorphism due to different band intensities were commonly observed throughout (Figure 10, Figure 11, Figure 12, and Figure 13). The loci that reflected differences in band intensity could be linked to some traits of interest.

Compared to previous reports on peanut microsatellites (Hopkins et al., 1999, p. 1234) for example, in which five polymorphic markers from 26 primers pairs (19%) were identified using library screening methods and a later study by He et al. (2003, p. 1), where 19 of 56 primer pairs (34%) showed a polymorphism, the enrichment procedure used for marama bean was more efficient in identifying SSR markers as 61 out of the 80 primer pairs (76%) showed polymorphism.

The molecular data gathered in this study based on microsatellite loci will subsequently need to be linked to phenotypic characters. The “phenotype”, i.e. the organism defined by the characters made manifest, must be distinguished from the “genotype” or genetic constitution, which alone can transmit changes to the

offspring. The best marama bean would tentatively be one with a high number of seeds per pod or per plant, high oil content, high protein content, high tuber starch, early flowering and less shattering as these are the traits directly or indirectly linked to high yield and high nutritional value.

The polymorphic microsatellites identified through this study can potentially be applied to identify marama microsatellite loci that affect phenotype in a selection and breeding program. Some of the phenotypes observed in field collections were different internode lengths, germination times and leaf sizes which could directly or indirectly linked to yield traits in marama bean. Plants showing differences in these traits would be selected in the field then screened with the microsatellite markers to see if for a particular trait a microsatellite marker could be linked with that trait by being able to distinguish individuals at the molecular level. Microsatellites have become one of the most widely used molecular markers for genetic studies in recent years (George et al., 2006, p. 919; Rongwen et al., 1995, p. 43) and they have proved useful in evaluating genetic variation in the Namibian marama bean germplasm.

Microsatellites isolated from *Tylosema esculentum* and found to be polymorphic enabled genetic diversity studies to be carried out. The microsatellite primers developed through the use of the FIASCO technique were used to analyze the genetic variation of the selected populations of *Tylosema esculentum.* In addition several gaps in the knowledge of marama bean germplasm in Namibia were filled. Before

this study it was not known if microsatellites were abundant in the genome of marama bean and if they could be able to detect any polymorphisms and be useful as molecular markers for marama bean.

The RAPD system suffers from a lack of reproducibility and that is why microsatellites have become the marker of choice. In this study the amplification profiles were shown to be reproducible as shown by MARA 001 (Figure 14). This demonstrated the clear advantage of microsatellites over RAPD's. Multiplexing PCR products offers the prospect of rapid determination of the allelic constitution of three or more loci in one PCR reaction. Resolution of different alleles (which only differ from one another by a few bases if the polymorphism is due to changes in the length of the SSR region) on agarose gel systems is not the highest even though agarose is the cheapest system.

Another less costly approach would be the use of non-denaturing polyacrylamide gels stained with ethidium bromide or silver which can resolve bands differing from each other by a few base pairs. The use of higher temperature and decreasing the number of cycles in the PCR reactions from 35 to between 25 and 30 are suggested as ways to optimize the PCR profiles obtained in the amplifications. An increase in annealing temperature would reduce non-specific binding particularly in Group 3 type primers. Optimization experiments are required to enable easier band scoring as some of the gels were difficult to score. Several parameters influence the specificity

of PCR and these include the temperature profile of the thermocycler, the annealing temperature, the activity and amount of the polymerase, concentrations of primers, template DNA and Mg^{2+}.

Many technical questions remain with regard to the use of SSR sequences. However, the informativeness of this type of marker, the rapid detection by PCR and the potential of tens of thousands of SSR sequences per genome suggest that in a plant genetics study, consideration of the use of SSR markers as genetic markers remains a feasible option. The use of modern biotechnology tools is essential to reduce the time necessary to achieve the germplasm characterization and subsequent breeding cycles. (Schlötterer, Orsini, & Huttunen, 2004, p. 161). The microsatellites designed here demonstrated the feasibility of this molecular marker system and trying to get the best resolution for identifying useful markers associated with traits of interest may be the best approach when using these SSR's.

In this study, the *mat*K gene in marama was found to be half the expected size of the gene. This could be due some deletions of sections of the gene. The homology with *Tylosema fassoglensis* in the case of the *trn*K gene confirmed that this was the region amplified for *Tylosema esculentum* (Figure 20). The homology with *Pisum sativum* in the case of the *mat*K gene confirmed that this was the region that had been amplified in the PCR reaction (Figure 21).

An amino acid composition analysis of the *matK* gene revealed high amounts of alanine (39.7%), cysteine (13.9%), glycine (15.9%) and threonine (29.9%). Traces of tyrosine (0.3%) and tryptopan (0.2%) were also present (Figure 22). The gene has been characterized for marama bean for the first time and if *mat*K is agreed upon as the standard barcode for land plants, the sequence for marama bean is now available.

A phylogenetic tree was constructed based on the *mat*K sequences for samples from the 12 localities and the results suggest that the *mat*K region can also be used in determining levels of genetic variation in marama bean collected from different localities (Figure 19). The tree suggests there has been evolution in this gene as the individuals from different localities formed different groups. Pretoria grouped with Sanveld and it was interesting to see where the South African type fell in relation to the Namibian relatives. It is not surprising that the individual from Pretoria was closest to that from Sanveld as that is one of the closest Namibian localities to South Africa geographically. The Namibian localities however formed groups that were not necessarily related to geographic distances (Figure 6, Figure 19). *mat*K may be a better choice for genetic diversity studies or may be able to complement work done with microsatellite markers.

CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS

This is the first report of large scale molecular marker development for *Tylosema esculentum*. The molecular markers reported here will be of particular value in the ongoing efforts to characterize the marama bean germplasm. The development of the microsatellite markers is an important first step towards the development of a saturated genetic linkage map and the development of marker assisted selection programs for traits of agronomic importance in marama bean.

There was a wide range of primers designed as seen in the discriminatory power of each of the primers that ranged from 0.188 - 0.947. A high level of polymorphism exists in the marama bean germplasm as demonstrated by 76% of the loci screened being polymorphic. The FIASCO technique was applicable to marama bean and was efficient in producing many microsatellite enriched fragments that enabled the design of 80 microsatellite primers which exceeded the targeted 50 primers. Only 3 of these primers did not give clear amplification products and may simply require optimization.

The microsatellite loci described will provide a pool of polymorphism useful for population studies, genetic mapping and possibly application in other *Tylosema* or *Fabaceae*. The development of SSR markers linked to specific traits will facilitate the screening of marama bean plants at early growth stages, thus accelerating

selective breeding programs. The types and distribution of repeat motifs in *Tylosema esculentum* were investigated. In plant breeding programs, information on the genetic diversity within and among species is essential for a rational use of genetic resources. The tools to enable generation of this information were designed in this study.

The *mat*K gene was characterized for *Tylosema esculentum* as a contribution to the greater CBOL effort to barcode land plants. It appears there have been some deletions in the gene in *Tylosema esculentum*. The monomorphic loci of the microsatellites were investigated for any gene associations and several genes were discovered for marama bean through the exercise. The gene products included exonucleases, aminohydrolases, peptidase receptors, globulin precursors, topoisomerases, retrotransposons and methyltransferases.

The next steps in the marama breeding program will be made up of several phases. The first step will be to grade marama bean plants for specific traits and find out if any of the microsatellite markers differentiate two extremes of a trait (say longest internode from shortest internode for example). After crossing almost 90% of the off-spring with the shortest or the longest internode depending on what is selected as a useful trait should amplify the microsatellite with the correct size originated from the parent with the useful trait. If they just follow a 1:2:1 ratio regarding the microsatellite then the microsatellite is not useful as a marker for the particular trait. The microsatellites now need to be linked to phenotypic traits related to yield parameters and integrated into a molecular map for *Tylosema esculentum*.

CHAPTER 7: REFERENCES

Akkaya, M.S., Bhagwat, A.A. & Cregan, P.B. (1992). Length polymorphisms of simple sequence repeat DNA in soybean. *Genetics*, 132, 1131-1139.

Amarteifio, A. & Moholo, D. (1998). The chemical composition of four legumes consumed in Botswana. *Journal of Food Composition and Analysis*, 11,329 -33.

Anderson, J.A., Churchill, G.A., Autroque, S.D., Tanksley, S.D. & Sorrelis, M.E. (1993). Optimising parental selection for genetic linkages maps. *Genome*, 36, 181-186.

Arnold, C., Rossetto, M., McNally, J. & Henry, R.J. (2002). The application of SSR's characterized for Grape (*Vitis vinifera*) to conservation studies in *Viticea*. *American Journal of Botany*, 89, 22-28.

Azam-Ali, S.N. (1996). *Bambara groundnut: How did we get here?* Proceedings of the International Bambara Groundnut Symposium, University of Nottingham UK.

Bidartondo, M. (2009). DNA barcoding and sequencing news. *Kew Scientist*, 35, 6.

Bower, N., Hertel, K., Oh, J. & Storey, R. (1988). Nutritional evaluation of Marama bean (*Tylosema esculentum*, Fabaceae): analysis of the seed. *Economic Botany*, 42, 533-540.

Brown, S.M., Hopkins, M.S., Mitchell, S.E., Senior, T.Y., Wang, R.R., Duncan, F., et al. (1996) Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. *Theoretical and Applied Genetics*, 93, 190-198.

Brown, T. A. (1998). *Genetics, a molecular approach*. Cheltenham: Stanley Thornes.

Brunel, D. (1994). A microsatellite marker in *Helianthus annus* L. *Plant Molecular Biology*, 24, 397 - 400.

Castro, S., Silveira, P., Pereira-Coutinho, A. & Figueiredo, E. (2005). Systematic studies in *Tylosema* (Leguminosae). *Botanical Journal of the Linnean Society* 147, 99-115.

CBOL Plant Working group. (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Science*, 106 (31), 12792-12797

Chase, M.W., Soltis, D.E., Olmstead, R.G., Morgan, D., Les, D.H., Mishler, B.D., et al. (1993). Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanic Garden*, 80, 528-580.

Chen, X., Temnykh, S., Xu, Y., Cho, G. & McCouch, S.R. (1997). Development of a microsatellite framework map providing genome wide coverage in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics*, 95, 553-567.

Chimwamurombe, P. (2008). ABS and creation of an enabling environment for innovation, is it an issue for SADC countries? Marama bean domestication:

an ABS case. *Building Bridges for Poverty Reduction and Sustainable Development*, 3, 5-7.

Chimwamurombe, P. (2009, November). *Cultivation of Marama Bean: A new crop for Cash and Sustainable Land Management (COMB-CSLM)*. A proposal submitted to the UNDP through the Ministry of Environment and Tourism under the Country Pilot Partnership (CCP) Programme for Integrated Sustainable Land Management (ISLM), University of Namibia, Windhoek, Namibia.

Coetzer, L.A. & Ross, J.H. (1977). Tylosema. *Trees in South Africa* 28: 77-80.

Cullis, C.A. (2004). *Plant genomics and Proteomics*. New Jersey: John Wiley & Sons.

Cullis, C. (2006, June). *DNA and Molecular markers*. A paper presented in Windhoek, Namibia by Professor Christopher Cullis, a Francis Hobart Herrick Professor of Biology at the Case Western Reserve University, Cleveland, USA.

Devos, K.M., Bryan, G.J., Collins, A.J., Stephenson, P. & Gale, M.D. (1995). Application of two microsatellite sequences in wheat storage as molecular markers. *Theoretical Applied Genetics*, 90, 247-252.

Diwan, N., Bhagwat, A.A., Bauchan, G.B., Cregan, P.B. (1997). Simple sequence Repeat DNA markers in alfalfa and perennial and annual *Medicago* species. *Genome*, 40, 887-895.

Dolanska, L. & Curn, V. (2004). Identification of white clover (*Trifolium repens* L.) cultivars using molecular markers. *Plant Soil Environment*, 50 (3), 95-100.

Dreher, K., Morris, M., Khairallah, M., Ribaut, J.M., Panoley, S., & Srinivasan, G. (2000). *Is marker assisted selection cost-effective compared to conventional plant breeding methods? The case of quality protein maize*. Paper presented at the 4th ICABR Conference on Economics of Agricultural Biotechnology, Ravello, Italy.

Elfant, M., Bryant, L., & Starcher, B. (1985). *Isolation and characterization of a proteinase inhibitor from marama beans*. Proceedings of the Society for Experimental Biology and Medicine, 180 (3), 329-333.

Falk, D. & Holsinger, K.E. (1991). *Genetics and Conservation of Rare Plants*. New York: Oxford University Press.

Ford-Lloyd B.V. (2001). *Genotyping in plant genetic resources: Plant genotyping, the DNA fingerprinting of plants*. Lismore, Australia: CABI Publishing.

Francis, C.M & Campbell, M.C. (2003). New quality oil seed crops for temperate and tropical Australia. RIRDC Publication No 03/045 RIRDC Project No UWA-47A, p. 1-13. Report.

Gaitán-Solis E, Duquqe MC, Edwards KJ, & Tohme J (2002). Microsatellite repeats in Common Bean (*Phaseolus vulgaris*): Isolation, characterization and cross species amplification in *Phaseolus* ssp. *Crop Science*, 42, 2128-2136

George, J., Dobrowolski, M.P., De Jong, E., Cogan, N.O.I., Smith, K.F., & Forster, J.W. (2006). Assessment of genetic diversity in cultivars of white clover (*Trifolium repens* L.) detected by SSR polymorphisms. *Genome*, 49, 919-930.

Gepts, P. (2002). A comparison between crop domestication, classical plant breeding, and genetic engineering. *Crop Science*, 42, 1780-1790.

Gepts, P. (2004). Crop domestication as a long term experiment. *Plant Breeding Review*, 24, 1-44.

Hajibabaei, M., deWaard, J.R., Ivanova, N.V., Ratnasingham, S., Dooh, R.T., Kirk, S.L., et al. (2005) Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society of London: Biological Sciences,* 360, 1959–1967.

Halloran, G.M. & Monaghan, B.G. (1996). RAPD variation within and between natural populations of morama [*Tylosema esculentum* (Burchell) Schreiber] in Southern Africa. *South African Journal of Botany*, 62(6), 287-297.

Hartl, D.L. & Jones, E.W. 2009. *Genetics: Analysis of genes and genomes*. London: Jones and Bartlett Publishers.

Hartley, L.M. (1997). *Breeding systems and genetic variation in Tylosema esculentum* (PhD. Thesis). Falmer. Sussex University. Retrieved March 23, 2008, from Sussex University website.

Hartley, L.M., Tshamekeng, E. & Thomas, S.M. (2002). Functional heterostyly in *Tylosema esculentum* (Caesalpiniodeae). *Annals of Botany*, 89, 67- 76.

He, G., Meng, R., Newman, M., Gao, G., Pittman, R.N. & Prakash, C.S. (2003). Microsatellites as DNA markers in cultivated peanut (*Arachis hypogea* L.). *BMC Plant Biology*, 3, 3, 1-6.

Hebert, P.D.N., Cywinska, A., Ball, S.L. & deWaard, J.R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London: Biological Sci*ences, 270, 313–321.

Holden, J., Peacock, J. & Williams, T. (1993). *Genes, Crops and the environment*. Cambridge: Cambridge University Press.

Hollingsworth, M.L., Clark, A., Forrest, L.L., Richardson, J., Pennington, R.T., Long, D.G., et al. (2009). Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Molecular Ecology Resources*, 9, 2, 439 - 457.

Hopkins, M.S., Casa, A.M., Wang, T., Mitchell, S.E, Dean, R.E., Kochert, G.D., & Kresovich, S. (1999). Discovery and characterization of polymorphic simple sequence repeats (SSRs) in cultivated peanuts (*Arachis hypogea* L). *Crop Science*, 39, 1243-1247.

Hughes, M.A. (1996). *Plant Molecular genetics*. Essex: Longman.

Karp, A., Edwards, K.J.E., Bruford, M., Funk, S., Vosman, B., Morgante, M., et al. (1997). Molecular technologies for biodiversity evaluation: Opportunities and challenges. *Nature Biotechnology*, 15, 625-628.

Keegan, A.B. & Van Staden, J. (1981*). Tylosema esculentum*, a plant worthy of cultivation. *South African Journal of Science*, 77, 387-397.

Khidir W.H. & Hongping L. (1997). The *mat*K gene: sequence variation and application in plant systematics. *American Journal of Botany*, 19, 830-839.

Kölliker, R., Jones, E.S., Drayton, M.C., Dupal, M.P., & Forster, J.W. (2001). Development and characterization of simple sequence repeats (SSR) markers for white clover (*Trifolium rep*ens L.) *Theoretical Applied Genetics*, 102, 416-424.

Lahaye, R., van der Bank, M. & Bogarin, D. (2008). DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences, USA*, 105, 2923-2928.

Lawlor, D. (2004). Improvement of marama bean (*Tylosema esculentum*) - an under - utilized grain and tuber producing legume for southern Africa. ICA4-CT -200- 30010 Final Report. Rothamsted: European Union.

Li, C.D., Rossnagel, B.G. & Scoles, G.J. (2000). The development of oat microsatellite markers and their use in identifying *Avena* species and oat cultivars. *Theoretical Applied Genetics*, 101, 1259-1268.

Liu, Z.W., Biyashev, R.M. & Saghai-Maroof, M.A. (1996) Development of simple sequence repeat markers and their integration into a barley linkage map. *Theoretical Applied Genetics*, 93, 869-876.

McNeal, J.R., Kuehl, J.V., Boore, J.L., & de Pamphilis, C.W. (2007). Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. *BMC Plant Biology*, 7, 57, 1-22.

Mengoni, A., Gori, A., & Bazzicalupo, M. (2000). Use of RAPD and microsatellite (SSR) variation to assess genetic relationships among populations of tetraploid alfalfa, *Medicago sativa*. *Plant Breeding,* 119, 311-317.

Mmnonatau, Y. (2005). Flour from the Marama bean: composition and sensory properties in a Botswana perspective. A MSc thesis submitted to the University of Stellenbosch, South Africa. Retrieved March 24, 2008, from the University of Stellenbosch website.

Müseler, D.L & Schönfeldt, H.C. (2006). The nutrient content of the marama bean (*Tylosema esculentum*), an underutilized legume from Southern Africa. *Agricola*, 16, 7-13.

Naomab, E. (2004). *Assessment of genetic variation in natural populations of Marama Bean (Tylosema esculentum) using molecular markers*. A thesis submitted to the University of Namibia for the degree of Master of Science. Retrived April 04, 2008, from the University of Namibia Library, Special collections.

National Academy of Sciences. (1979). Tropical legumes: Resources for the future: A report of an Ad Hoc Panel of the Advisory Committee on Technology for International Development Commission on International Relations, National Council Research, Library of Congress Catalog Number 79-64185.

National Research Council. (2006). *Lost Crops of Africa: Volume II: Vegetables*. Washington: The National Academies Press.

Nei, M. & Koehn, R.K. (1983). *Evolution of genes and proteins*. Massachusetts: Sinauer Associates Inc. Publishers.

Nepolo, E. (2010). *Assessment of genetic variation within and between populations of Marama Bean [Tylosema esculentum (Burchell) Schreiber] based on microsatellites (SSR's) and Intergenic Spacer Length variation markers in the Namibian germplasm*. A thesis submitted to the University of Namibia for the degree of Master of Science. In press.

Nepolo, E., Takundwa, M., Chimwamurombe, P., Cullis, C.A., & Kunert, K. (2009). A Review of the geographical distribution of marama bean [*Tylosema esculentum* (Burchell) Schreiber] and genetic diversity in the Namibian germplasm. *African Journal of Biotechnology*, 8, 2088-2093.

Notredame, C., Higgins, D.G., & Heringa, J. (2000). T-Coffe: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, 302(1): 205-217.

Prasad, M., Varchney, R.K., Roy, J.K., Balyan, H.S., & Gupta, P.K. (2000). The use of microsatellites for detecting DNA polymorphism, genotype identification and genetic diversity in wheat. *Theoretical Applied Genetics*, 100, 584-592.

Quiagen (2006). *DNeasy Plant Handbook*. Hillden, Germany: Quiagen.

Rassman, K., Schlötterer, C., & Tautz, D. (1991). Isolation of simple sequence loci for use in polymerase chain reaction based DNA fingerprinting. *Electrophoresis*, 12, 113-118.

Roder, M.S., Plaschke, J., Konig, S.U., Borner, A., Sorrels, M.E., Tanksley, S.D., et al. (1995). Abundance, variability and chromosomal location of microsatellites in wheat. *Molecular and General Genetics*, 246, 327-333.

Rongwen, J., Akkaya, M.S., Bhagwat, A.A., Lavi, U., & Cregan, P.B. (1995). The use of microsatellite DNA markers for soybean genotype identification. *Theoretical Applied Genetics*, 90, 43-48.

Ruane, J. & Sonnino, A. (2006). *The role of biotechnology in exploring and protecting agricultural genetic resources*. Rome: Food and Agriculture Organization of the United Nations.

Sambrook, J. & Russel, D. (2001). Molecular cloning: A laboratory Manual. New York: Cold Spring Harbour Laboratory Press.

Santanna, Q.C., Coetzee, M.P.A., Steenkamp, E.T., Mlonyeni, O.X., Hammond, G.N.A., Wingfield, M.J., & Wingfield, B.D. (2009). Microsatellite discovery by deep sequencing of enriched genomic libraries. *BoiTechniques*, 46, 217-223.

Schlötterer, C., Orsini, L., & Huttunen, S. (2004). A multilocus microsatellite phylogeny of the *Drosophila virilis* group. *Heredity*, 93, 161-165.

Selvaraj, D., Sarma, R.K., & Sathishkumar, R. (2008). Phylogenetic analysis of chloroplast *mat*K gene from Zingiberaceae for plant DNA barcoding. *Bioinformation*, 3(1), 24-27

Senior, M.L., Murphy, J.P., Goodman, M.M. & Stuber. C.W. (1998). Utility of SSR's for determining genetic similarities and relationships in maize using an agarose system. *Crop Science*, 38, 1088-1098.

Singer, M. & Berg, P. (1991). *Genes and genome*. Oxford: Blackwelll Scientific Publications.

Sharopova, N., McMullen, M.D, Schultz, L., Schroeder, S., Sanchez- Villeda, H., Gardiner, J., et al. (2002). Development and mapping of SSR markers for maize. *Plant Molecular Biology*, 48, 463-481.

Smulders, M.J.M., Bredemeijer, G., Rus-Kortekaas, W., Arens, P., Vosman, B. (1997). Use of short microsatellites from database sequences to generate polymorphisms among *Lycopersicon esculentum* cultivars and accessions of other *Lycopercicon* species. *Theoretical Applied Genetics*, 97, 264-272.

Spelling, F. (2008). *Land Degradation and Climate Change in Sub-Saharan Africa Building Resilience through Sustainable Land Management*. Retrieved from: http://www.terrafrica.org/file_view.asp?tbl=1130d4334i58&fid=9487114&t application/pdf.

Tait, R.C. (1997). *An introduction to Molecular biology*. Norfolk: Horizon Scientific Press.

Tautz, D., Trick, M. & Dover, G.A. (1986). Cryptic simplicity in DNA is a major source of genetic variation. *Nature*, 322, 652-656.

Tessier, C., David, J., This, P., Boursiquot, J.M. & Charrier, A. (1999). Optimization of the choice of molecular markers for varietal identification in *Vitis vinifera* L. *Theoretical Applied Genetics*, 98, 171-177.

Thulin M. (1993). Fabacea (Leguminosae). In: Thulin M, ed. Flora of Somalia, vol 1. Pteridophyta, Gymnospermae, Angiospermae (Annonacea- Fabacea). *Kew: Royal Botanic gardens*, 360-361.

Travlos, I.S., Economou, G. & Karamanos, A.I. (2007). Germination and emergence of the hard seed coated Tylosema esculentum (Burch) A. Schreib in response to different pre-sowing seed treatments. *Journal of Arid environments,* 68, 501-507.

Vogel, J., Hübschmann, T., Börner, T., Hess W.R. (1997). Splicing and intron internal RNA editing of *trnK-matK* transcripts in barley plastids: support for *matK* as an essential splice factor. *Journal of Molecular Biology*, 270, 179 -187.

Westman, A.L. & Kresovich, S. (1997). *Biotechnology and Plant genetic resources (Conservation and use).* Oxon: CAB International.

Wojciechowski, M.F., Lavin, M., & Sanderson, M.J. (2004). A phylogeny of legumes (*Leguminosae*) based on analysis of the plastid *mat*K gene resolves many well supported subclades within the family. *American Journal of Botany*, 91 (11), 1846- 1862.

Wolfe, K.H., Morden, C.W. & Palmer, J.D. (1992). Function and evolution of a minimal plastid genome from a non-photosynthetic parasitic plant. *Proceedings of the National Academy of Science, USA* 89: 10648-10652.

Wolfe, A. & Liston, A. (1998). *Contribution of PCR-based methods to plant systematics and evolutionary biology: Molecular Biology of Plants II DNA sequencing*. London: Kluwer Academic Publishing.

Yu, K., Park, S.J., Poysa, V. (1999). Abundance and variation of microsatellite DNA sequences in beans (*Phaseolus* and *Vigna*). *Genome*, 42, 27-34.

Zane, L., Bargelloni, L., & Patarnello, T. (2002). Strategies for microsatellite Isolation: a review. *Molecular ecology* 11: 1-16.

APPENDICES

5X TBE (TRIS-BORATE) STOCK

(1X=89mM Tris base, 89mM boric acid, 2mM EDTA)

(0.5X=45mM Tris borate, 1mM EDTA)

54.0 g Tris base

27.5 g Boric acid

3.72 g $Na_2EDTA.2H_2O$

To 1 litre with distilled water.

Microwave instructions for agarose preparation of gel concentrations ≤2% w/v

1. Choose a conical flask that is 2-4 times the volume of the solution.
2. Add room temperature 1X or 0.5X buffer to the beaker
3. Sprinkle the premeasured agarose powder
4. Heat the beaker in the microwave oven on HIGH power until bubbles appear.

CAUTION: Any microwaved solution may become superheated and foam over when agitated.

5. Remove the conical flask from the microwave oven.
6. Gently swirl the beaker to resuspend any settled powder and gel pieces.
7. Reheat the conical flask on high until the solution comes to a boil.
8. Hold at boiling point for 1 minute or until all of the particles are dissolved.
9. Remove the beaker from the microwave oven.
10. Gently swirl the beaker to mix the agarose solution thoroughly.
11. Mix thoroughly.
12. Cool the solution to 60^{o}C prior to casting.

Table 3: The 80 microsatellite loci and primers designed to screen for polymorphic microsatellite loci in this study.

Marama microsatellites designed and screened for polymorphism (80)							
Primer	**Left primer (L)**	**Right primer (R)**	**SSR**	**Repeats**	**Annealing (L/R)**	**Expected product (bp)**	**Polymorphic/ Monomorphic**
MARA001	GCACAACCAATTTCCTGCTT	TCCCTCACTGGCCTATATCC	gag	5	60.12/58.96	137	Polymorphic
MARA002	CTCCCTCCTCCTCCTCGTAG	CGGGAGCAAATAGACCCTTT	acc	8	60.34/60.44	106	Polymorphic
MARA003	TCTCACCGACCGGGTCTC	CCTCTATCCCGCTCCCTATC	ctc	5	62.30/60.02	160	Monomorphic
MARA004	TGCAGGCTTCACCAGAGTAA	TCTAAGACTGCGCACACAGC	ga	5	59.59/60.36	170	Monomorphic
MARA005	GCTATCCGAGGGAGGATCA	GTGTCTATGTGTGCGCGTGT	ag	6	60.13/60.82	128	Monomorphic
MARA006	GCTATCCGAGGGAGGATCA	TCCCATTCAGCCATTTTAGG	tg	7	60.13/59.89	171	--------
MARA007	TATCCGAGGGAGGATCATGT	TCACATCCTAAGACTCGAACTTCA	ac	6	59.32/60.29	150	Monomorphic
MARA008	GCTGGTCCATGGCTTCAT	TTTGTAATCGGTTGACACTTTGA	tg	5	59.59/59.54	185	Monomorphic
MARA009	GGGAGGATCAACCTCAACAA	TGTACAAAAAGCAGGCTCCA	gaa	5	59.90/59.46	216	Polymorphic
MARA010	TGTGCTATCCGAGGGAGGAT	ACGTCGCGATTAAACAAACC	aag	7	61.92/60.00	152	Polymorphic
MARA011	TGTCAACGCTTACGTTGGTC	TCATTTGAAACCCTTGTACTGC	tc	8	59.76/59.13	169	Polymorphic
MARA012	ATATGGTGGCTCGTCGATGT	GCACATAATTCGAACAGAACACA	ag	5	60.37/60.05	163	Monomorphic
MARA013	GCTTCTCGTACATGGGCTTT	GCATTTATCGCGAATACAGCA	tc	5	59.34/61.10	154	Polymorphic

Primer	**Left primer (L)**	**Right primer (R)**	**SSR**	**Repeats**	**Annealing (L/R)**	**Expected product (bp)**	**Polymorphic/ Monomorphic**
MARA014	GGTGGTGGTGTAGGAGGAGA	GACTTGAGTGCATGCCATTT	agg	5	59.96/58.73	167	Polymorphic
MARA015	ACTCATCCCGCTCCTAAGGT	AAACAGGCTCGTATTTTATCTTCG	tg	5	60.10/60.05	204	Polymorphic
MARA016	TTCAATTTTCTTCACCACAAACTC	ACAGGAAGGTCTTCCACAGC	ca	7	59.56/59.30	102	Polymorphic
MARA017	ACCCTTGAATTGTGGTAGGG	ACTGTGCTATCCGAGGGAGA	ct	6	58.76/59.83	105	Polymorphic
MARA018	ATTTTGGCTTTACCGCACAC	AGCACTCTCCAGCCTCTCAC	cttga	3	60.00/59.74	158	Polymorphic
MARA019	CCGGAACAGGAGAAGCTATG	TCAACTTTTGCAATGAACGAA	ctt	4	59.83/59.33	161	Monomorphic
MARA020	TGTCTTCCCCTCCTCTTCCT	TTGACACTTTGGGACTGCTG	cag	4	60.19/59.87	175	Polymorphic
MARA021	GAGGGAGGATCACCACTCAG	TGGCCATCAATCATGTTACG	tgt	4	59.64/60.34	166	Monomorphic
MARA022	CCCCTGTACCCAAGACTCTG	TCCATGAAGTCAGGAGAAGGA	tagc	3	59.57/59.79	171	Polymorphic
MARA023	ATGGGGATACTCCCGAAACT	AATGGGAGCAAGAATTTCCA	aaag	3	59.65/59.50	250	Monomorphic
MARA024	CCAAGAGTGGGGATGAAAGA	TTGGAATAGTTCCCCCTTCC	aga	4	60.04/60.12	227	Polymorphic
MARA025	CACGTGGGTTGTACTTATCTGC	TAATGTGTTGAGCGCCGTAG	tcttc	3	59.56/59.90	160	Monomorphic
MARA026	GCTGTTGGGAACCGTAGAAA	CCTATTGTCAGTGTAAGCAACCA	tc	4	60.11/59.21	208	Monomorphic
MARA027	TTGTTCCAAACCACAGGTCA	TGGCCATCTCCCAATTTTAC	ca	4	59.98/59.76	194	Monomorphic
MARA028	CTCCGCATCTGACTTCAAAA	CCTCCTCTCCCTGATTTTCC	aga	3	59.00/60.01	150	Monomorphic
MARA 029	CCGAGGGAGTAGTGCTTCAT	CGCCACTTAGCATTTGCTTT	tg	4	59.31/60.40	155	Monomorphic

MARA030	GAGCCAAAGCCATGATCCTA	CCCATGTTGTATATTCGTGGAA	caa	3	60.18/59.59	178	Monomorphic
Primer	**Left primer (L)**	**Right primer (R)**	**SSR**	**Repeats**	**Annealing (L/R)**	**Expected product (bp)**	**Polymorphic/ Monomorphic**
MARA031	CTCAGCACTCTCCAGCCTCT	CCGAGGGAGGATCATTAACA	gga	7	59.88/59.89	126	Polymorphic
MARA032	AGACGCACTCCCTCTCACC	GCTATCCGAGGGAGGATCA	aca	5	60.41/60.13	151	Polymorphic
MARA033	GCACTCAGGCAACTGTGCTA	AGCACTCTCCAGCCTCTCAC	aac	7	60.21/59.74	132	Polymorphic
MARA034	CTCAGCACTCTCCAGCCTCT	AGGGAGGATCACCTCCAAAC	gag	8	59.88/60.31	168	Polymorphic
MARA035	GACGCACTCAGCAACTCTCC	TCCAGCCTCTCACCGATTAC	acc	8	61.17/60.22	152	Polymorphic
MARA036	GACGCACTCAGGCAACTGT	CCGAGGGAGGATCAAAGAAT	gga	11	60.05/60.40	180	Polymorphic
MARA037	GGGAGGATCAATCTTCACCA	TCCGAGAGAAGAGGAGGAAA	gagt	5	59.86/59.09	165	Polymorphic
MARA038	TGTTGATGAACTAGTGCTAGTGGT	AGCACTCTCCAGCCTCTCA	tgg	7	58.02/58.79	126	Polymorphic
MARA039	TCATTAAAGGGCTCCATTGC	ATGCCCAAAATCACCAACAT	aga	7	60.04/60.06	176	Polymorphic
MARA040	GACGCACTCAGGCAACTGT	CTGGCCTATATCCCCTCCTC	gga	8	60.05/59.88	192	Polymorphic
MARA041	AGACGCACTCAGCACTCTCC	GCTATCCGAGGGAGGATCAC	gga	7	60.77/60.96	150	Polymorphic
MARA042	CAAATAGCCAAAGCCCGTTA	ACTCTCAAACCGTGGCACAT	agg	7	60.09/60.58	184	Polymorphic
MARA043	TGTTGATGAACTAGTGCTAGTGGT	AGCACTCTCCAGCCTCTCA	tgg	7	58.02/58.79	126	Polymorphic
MARA044	AGACGCACTCAGCATTCTCC	GGTCTCGTCTTCCCCTTCAT	gag	8	60.56/60.46	156	Polymorphic
MARA045	GACGCACTCAGGCAACTGT	CTGGCCTATATCCCCTCCTC	gga	8	60.05/59.88	192	Polymorphic
MARA046	GCACTCAGGCAACTGTGCTA	TGGCTGGCACTCTGATTAAG	cta	10	60.21/59.02	169	Polymorphic

MARA047	GCACTCAGGCAACTGTGCTA	TGACTAGTCCCCGTGATGGT	caa	7	60.21/60.39	188	Polymorphic
Primer	**Left primer (L)**	**Right primer (R)**	**SSR**	**Repeats**	**Annealing (L/R)**	**Expected product (bp)**	**Polymorphic/ Monomorphic**
MARA048	AGACGCACTCCACCACTGTA	TGCTGAAACCGTGAGAGAGA	ct	12	59.34/59.70	212	Polymorphic
MARA049	GCACTCAGGCAACTGTGCTA	GGCGAACTAGTGCTATCGAG	ct	13	60.21/57.75	183	Polymorphic
MARA050	AGACGCACTCAGCACTCTCC	TGTGCTATCCGAGGGAGGAT	cac	7	60.77/61.92	116	Polymorphic
MARA051	GCACTCAGGCAACTGTGCTA	AGCCTCTCACCGATTACTGC	ca	15	60.21/59.46	139	Polymorphic
MARA052	GCACTCAGGCAACTGTGCTA	CACGCCTCTCACAAGAAACA	ct	14	60.21/60.02	249	Polymorphic
MARA053	CTCAGCACTCTCCAGCCTCT	CCCTCATCTCCCTTTCCTTC	gga	9	59.88/60.01	238	--------
MARA054	GCACTCAGGCAACTGTGCTA	AGCACTCTCCAGCCTCTCAC	gtt	8	60.21/59.74	159	Polymorphic
MARA055	GACGCACTCAGCAACTCTCC	TCCAGCCTCTCACCGATTAC	acc	8	61.17/60.22	150	Polymorphic
MARA056	AGACGCACTCAGGCAACTGT	ATCCGAGGGAGGATCATTAAA	gga	8	61.07/59.75	150	Polymorphic
MARA057	GACGCACTCAGGCAACTGT	TGAAGATCCTCCCTCGGATA	gga	11	60.05/59.58	171	Monomorphic
MARA058	GCACTCAGGCAACTGTGCTA	ACGACGAACGTAGTCGTCTC	ca	20	60.21/57.96	246	Polymorphic
MARA059	GACGCACTCCTGTGCTATCC	ACGTCGCGATTAAACAAACC	aag	7	60.83/60.00	162	Polymorphic
MARA060	CTCAGCACTCTCCAGCCTCT	CCTTCGTGTTTTACAGTTGTCG	gtg	11	59.88/59.71	179	Polymorphic
MARA061	GAGGGAGGATCAAGGGACAC	AGCACTCTCCAGCCTCTCAC	ctc	7	60.86/59.74	183	Polymorphic
MARA062	GACGCACTCCTGTGCTATCC	TGCGCAAGGACAATGATTAC	aag	7	60.83/59.69	202	Polymorphic
MARA063	GTGCAAGACCCGTTTAGGAA	AGGACGAACACGTGCGTATC	ct	9	60.11/61.13	185	Monomorphic

MARA064	GGAGGAGGAGGAGGAGTTTG	GAGGATCCACTCCCTCACTG	gag	5	60.19/59.64	192	Monomorphic
Primer	**Left primer (L)**	**Right primer (R)**	**SSR**	**Repeats**	**Annealing (L/R)**	**Expected product (bp)**	**Polymorphic/ Monomorphic**
MARA065	TGGTGGTAGGGTGGTGGTAT	CCACTTTTCACAGGCAAACA	ttc	6	59.97/59.7\ 3	191	Polymorphic
MARA066	GCACTCAGGCAACTGTGCTA	GCTATCCGAGGGAGAGAGGA	cct	5	60.21/60.83	161	Polymorphic
MARA067	AGACGCACTCAGCCTCTCAC	CCTCTATCCCGCTCCCTATC	ctc	5	60.77/60.02	174	--------
MARA068	GGAGGAGGAGGAGGAGTTTG	GAGGATCCACTCCCTCACTG	gag	5	60.19/59.64	192	Polymorphic
MARA069	GGGAGGATCAACCTCAACAA	TGTACAAAAAGCAGGCTCCA	gaa	5	59.90/59.46	216	Polymorphic
MARA070	GACGCACTCAGGCAACTGT	GGGAGGATCACTTCCACTCTC	gga	5	60.05/60.07	168	Polymorphic
MARA071	CTCAGCACTCTCCAGCCTCT	CCTGTTGGGGAGTTGTTGTT	cac	6	59.88/59.86	202	Polymorphic
MARA072	CTCAGCACTCTCCAGCCTCT	GATTGCTGTTGTTGGCAGTG	caa	6	59.88/60.31	225	Polymorphic
MARA073	AGCACTCTCCAGCCTCTCAC	ATGTTGAGGCAGAGGAGGAA	cat	5	59.74/59.80	161	Polymorphic
MARA074	CCAGCCTCTCAACCGATTAC	AGGCACAGCCCTAGACTCCT	aag	6	59.69/60.41	250	Polymorphic
MARA075	GGTGGTGGTGTAGGAGGAGA	TCCAGCCTCTCACCGATTAC	agg	5	59.96/60.22	202	Polymorphic
MARA076	ATTTTTGCATCAGCAACAGC	ATCCGAGGGAGGATCCTATG	aca	5	58.92/60.25	193	Polymorphic
MARA077	CTCAGCACTCTCCAGCCTCT	GGGTTGGTTGAAGAGGGAGT	aag	5	59.88/60.35	197	Polymorphic
MARA078	GCACTCAGGCAACTGTGCTA	GTTCACCATCCCCTCTCTGA	gct	5	60.21/60.05	223	Polymorphic
MARA079	GACGCACTCCAACTGTGCTA	TCATTTGAAACCCTTGTACTGC	tc	8	60.06/59.13	238	Polymorphic

MARA080	ACCGACCGAGAGAATGAAGA	GAGTCCTCAACAGGGAGCTG	atc	6	59.80/59.99	153	Polymorphic

Table 4: The discriminatory power of the 80 microsatellite primers designed in this study.

Discriminatory power (D_L) of the microsatellites		
Locus	**Group**	**D_L**
MARA001	Group 2	0.488
MARA002	Group 3	0.626
MARA003	Group 1	0
MARA004	Group 1	0
MARA005	Group 1	0
MARA006	Group 1	0
MARA007	Group 1	0
MARA008	Group 1	0
MARA009	Group 2	0.332
MARA010	Group 3	0.366
MARA011	Group 3	0.582
MARA012	Group 3	0.188
MARA013	Group 3	0.488
MARA014	Group 3	0.931
MARA015	Group 3	0.493
MARA016	Group 3	0.188
MARA017	Group 3	0.620
MARA018	Group 3	0.377
MARA019	Group 1	0
MARA020	Group 2	0.681
MARA021	Group 1	0
MARA022	Group 2	0.499
MARA023	Group 1	0
MARA024	Group 2	0.548

Locus	**Group**	**D_L**
MARA025	Group 1	0
MARA026	Group 1	0
MARA027	Group 1	0
MARA028	Group 1	0
MARA 029	Group 1	0
MARA030	Group 1	0
MARA031	Group 3	0.587
MARA032	Group 3	0.920
MARA033	Group 3	0.947
MARA034	Group 3	0.737
MARA035	Group 3	0.687
MARA036	Group 3	0.626
MARA037	Group 2	0.499
MARA038	Group 2	0.554
MARA039	Group 2	0.499
MARA040	Group 3	0.659
MARA041	Group 3	0.864
MARA042	Group 3	0.698
MARA043	Group 2	0.360
MARA044	Group 3	0.831
MARA045	Group 3	0.909
MARA046	Group 2	0.277
MARA047	Group 3	0.853
MARA048	Group 3	0.637
MARA049	Group 3	0.615
MARA050	Group 3	0.776
MARA051	Group 3	0.565
MARA052	Group 2	0.825
MARA053	Group 4	-

MARA054	Group 3	0.582
Locus	**Group**	**D_L**
MARA055	Group 3	0.759
MARA056	Group 3	0.781
MARA057	Group 3	0.582
MARA058	Group 2	0.443
MARA059	Group 3	0.931
MARA060	Group 3	0.914
MARA061	Group 2	0.571
MARA062	Group 3	0.637
MARA063	Group 1	0
MARA064	Group 1	0
MARA065	Group 2	0.947
MARA066	Group 2	0.526
MARA067	Group 4	-
MARA068	Group 3	0.947
MARA069	Group 2	0.266
MARA070	Group 3	0.914
MARA071	Group 3	0.898
MARA072	Group 3	0.715
MARA073	Group 3	0.632
MARA074	Group 3	0.191
MARA075	Group 3	0.742
MARA076	Group 3	0.809
MARA077	Group 3	0.837
MARA078	Group 2	0.499
MARA079	Group 3	0.349
MARA080	Group 3	0.770

Protein: Tylosema esculentum_trnK_partial sequence
Length = 1167 amino acids
Molecular Weight = 97494.06 Daltons

Amino Acid		Number	Mol%
Ala	A	417	35.73
Cys	C	210	17.99
Asp	D	0	0.00
Glu	E	0	0.00
Phe	F	0	0.00
Gly	G	191	16.37
His	H	0	0.00
Ile	I	0	0.00
Lys	K	0	0.00
Leu	L	0	0.00
Met	M	0	0.00
Asn	N	0	0.00
Pro	P	0	0.00
Gln	Q	0	0.00
Arg	R	0	0.00
Ser	S	0	0.00
Thr	T	349	29.91
Val	V	0	0.00
Trp	W	0	0.00
Tyr	Y	0	0.00

Protein: Tylosema esculentum_matK_partial sequence
Length = 637 amino acids
Molecular Weight = 52785.91 Daltons

Amino Acid		Number	Mol%
Ala	A	253	39.72
Cys	C	89	13.97
Asp	D	0	0.00
Glu	E	0	0.00
Phe	F	0	0.00
Gly	G	101	15.86
His	H	0	0.00
Ile	I	0	0.00
Lys	K	0	0.00
Leu	L	0	0.00
Met	M	0	0.00
Asn	N	0	0.00
Pro	P	0	0.00
Gln	Q	0	0.00
Arg	R	0	0.00
Ser	S	0	0.00
Thr	T	191	29.98
Val	V	0	0.00
Trp	W	2	0.31
Tyr	Y	1	0.16

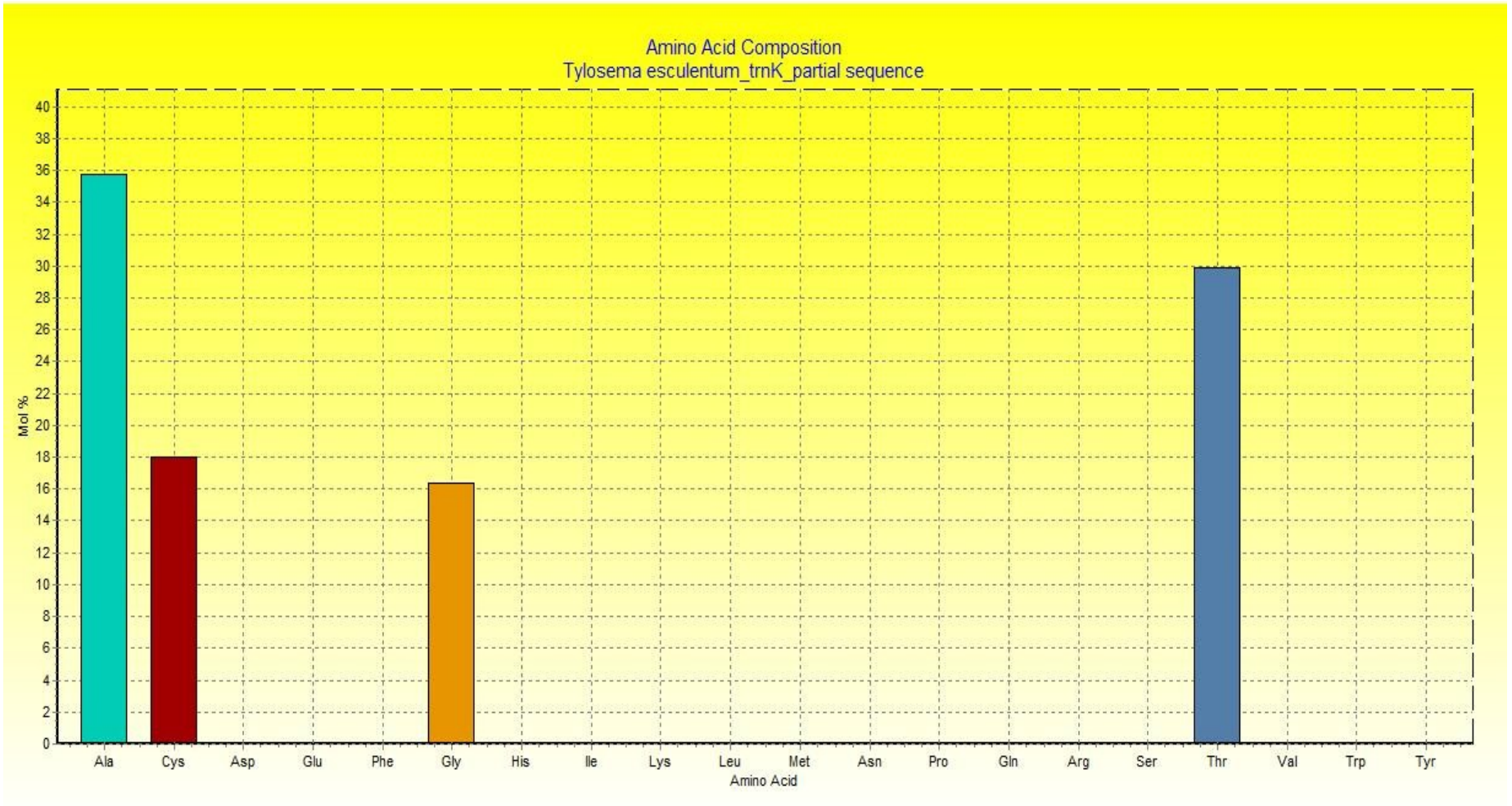

Amino Acid Composition
Tylosema esculentum_trnK_partial sequence
Mol %
40
38
36
34
32
30
28
26
24
22
20
18
16
14
12
10
8
6
4
2
0
Ala
Cys
Asp
Glu
Phe
Gly
His
Ile
Lys
Leu
Met
Asn
Pro
Gln
Arg
Ser
Thr
Val
Trp
Tyr
Amino Acid

Figure 22: Amino acid composition of the *Tylosema esculentum trn*K sequence.

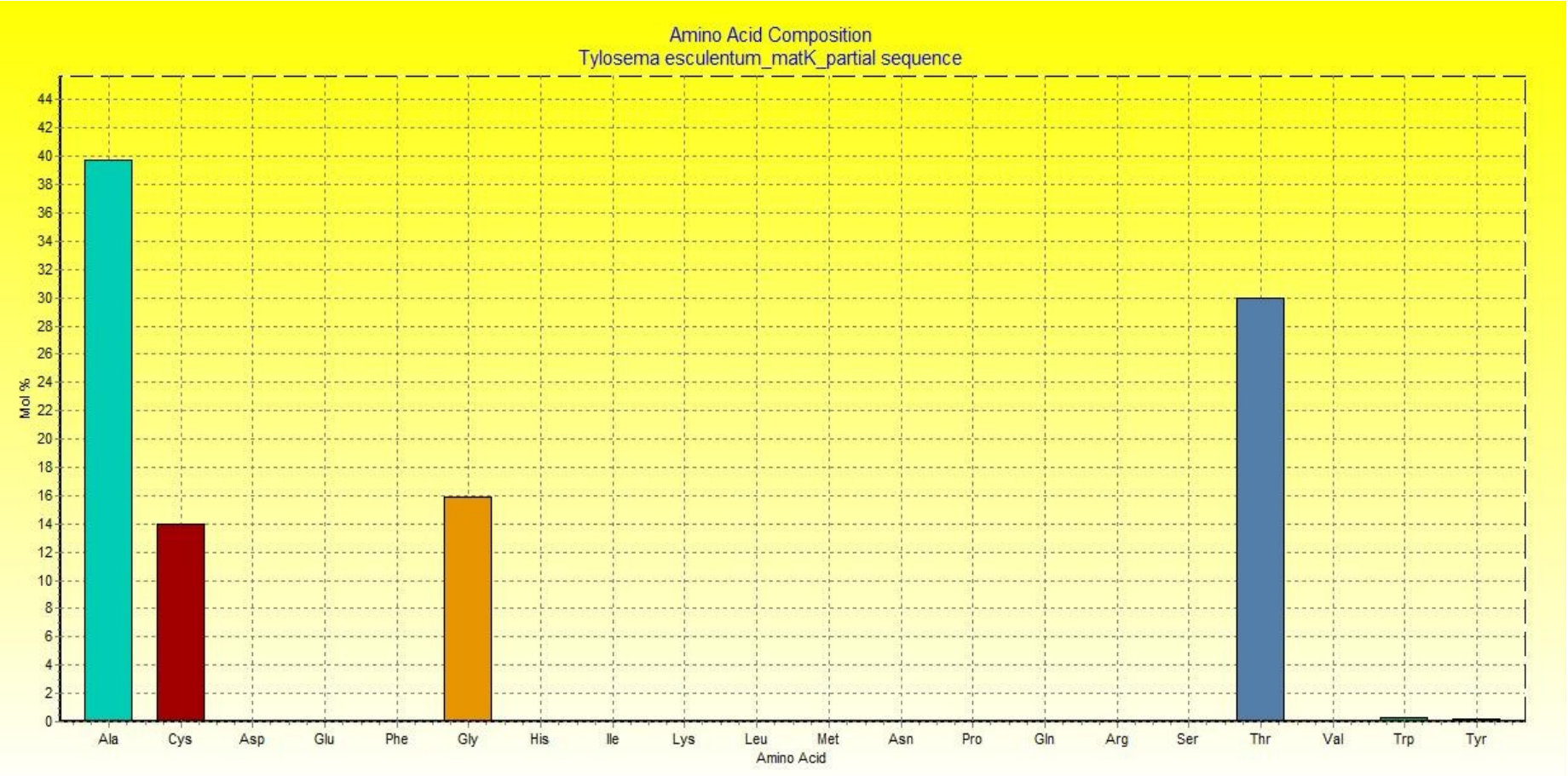


Figure 23: Amino acid composition of the *Tylosema esculentum mat*K sequence.

```
>gb|EU361874.1|  Tylosema fassoglensis voucher Herendeen 21-XII-97-6 (US) tRNA-Lys
(trnK) gene, partial sequence; and maturase K (matK) gene,
complete cds; chloroplast
Length=1755

 Score = 1674 bits (906),  Expect = 0.0
 Identities = 995/1034 (96%), Gaps = 22/1034 (2%)
 Strand=Plus/Minus

Query  80    CTTTCCCTATGTATACATCTAAACCTCTGTTCCTTCGCTAAAATAGGAC-TCTAAGAAGA  138
             |||||||||||| |||||||||||||||||||||||||||||||||| ||||||||||
Sbjct  1755  CTTTCCCTATGTCTACATCTAAACCTCTGTTCCTTCGCTAAAATAGGACTTCTAAGAAGA  1696

Query  139   AAATGGAATACTGAGTGGATCCATCGCTACTTCCTCTTACTATATGAGCATTTCATAATA  198
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1695  AAATGGAATACTGAGTGGATCCATCGCTACTTCCTCTTACTATATGAGCATTTCATAATA  1636

Query  199   GAATCATAATAGAAATGATTGCATTTGttttttttttCTCTGTATCATTTAAGAAGAATTT  258
             |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Sbjct  1635  GAATCATAATAGAAATGATTGCATTTGTTTTTTTTTTCTCTGTATCATTTAAGGAGAATTT  1576

Query  259   ACGGTGACAAAGTATCACAACCAATCATGAATGATTGACCAGATCATTTATGCAAATAAT  318
             || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1575  ACAGTGACAAAGTATCACAACCAATCATGAATGATTGACCAGATCATTTATGCAAATAAT  1516

Query  319   ATCCAAATACCAAATTCGACCTCCATAGAACCTCTGCAAAGTAAAAGAGGTTTTTCGGAA  378
             |||||||||||||||||||||| |||||||||||||||||| ||||||||||||||||
Sbjct  1515  ATCCAAATACCAAATTCGACCTCGATAGAACCTCTGCAAAGTAGAAGAGGTTTTTCGGAA  1456

Query  379   GATCAAAGAAAGAATCTCTTCTTCCTCTGGAAAGAATTCTTCCCATAATTCTGAACCTAA  438
             |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Sbjct  1455  GATCAAAGAAAGAATCTCTTCTTCCTCTGTAAAGAATTCTTCCCATAATTCTGAACCTAA  1396
```

```
Query  439   TCTTTTTAAAAAGGCGCGTACAGTACTTTTGTGTTTACGAGCCAAAGTTTTAATACAAGA  498
             ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1395  TCTTTTCAAAAAGGCGCGTACAGTACTTTTGTGTTTACGAGCCAAAGTTTTAATACAAGA  1336

Query  499   AAGCCGAAGTATATATTTTACTCGATACAAATTCtttttttttGGAGGATCCATTGTAATA  558
             |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Sbjct  1335  AAGCCGAAGTATATATTTTACTCGATACAAATTCTTTTTTTTTTGAGGATCCATTGTAATA  1276

Query  559   ATGAGAAAAATTTCTGCATATCCGCACAAATCGGTCAATAAGATCAAAATCAAATGAATC  618
             |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Sbjct  1275  ATGAGAAAAATTTCTGCATATCCGCACAAATCGGTCAATAAGATCAAAATCAGATGAATC  1216

Query  619   GGCCCAAACCGGTTTACTTATAGGATGACCTAATGCGTTACAAAATTTCGCTTTAGCCAA  678
             |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1215  GGCCCAGACCGGTTTACTTATAGGATGACCTAATGCGTTACAAAATTTCGCTTTAGCCAA  1156

Query  679   TAATCTAATTATGGGAATAATCGGAACTATTGTATTAAGCTTTTTCCTAACTTTTTCTAT  738
             |||||||||||| |||||||||||||||||||||||||||||||||| ||||||||
Sbjct  1155  TAATCTAATTATTGGAATAATCGGAACTATTGTATTAAGCTTTTTCCTAACATTTTCTAT  1096

Query  739   TAGAAATGAATTTTCCAGCATTCGACTCCGTACCACCGAAGGATTTAGCCGCACACTTGa  798
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1095  TAGAAATGAATTTTCCAGCATTCGACTCCGTACCACCGAAGGATTTAGCCGCACACTTGA  1036

Query  799   aaaataacccaaaaacaaaaagtaaaaGGAATGCTCAGAAAATTGGGTTTATATGGATTC  858
             ||||||||||||||||||||||| || |||||||||||| |||||| ||||||||||||
Sbjct  1035  AAAATAACCCAAAAACAAAAAGTAGAATGAATGCTCAGATAATTGG-TTTATATGGATTC  977

Query  859   TTCCTGGGTGAGACCAAACATAAAAATGACATTGGCCATAAATTGATAAAATAGTATTTC  918
             ||||||| |||||||||||||||||||||| |||||||||||||||||||||||||||||
Sbjct  976   TTCCTGGTTGAGACCAAACATAAAAATGACATTG-CCATAAATTGATAAAATAGTATTTC  918

Query  919   CATTTAT-CATCAAAAGAAACGTATTCGTTGAAGCCAGAATAGATTTTCCTTGATATCTA  977
             ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct  917   CATTTATTCATCAAAAGAAACGTATTCGTTGAAGCCAGAATAGATTTTCCTTGATATCTA  858

Query  978   A-A-A-TGAATGAAAGGGATCCTTAAATAAC-ATAAGGTCGACAGAAA-TCTTTAGCAAG  1032
             | | | |||||||||||| |||||||||||| |||||||||||||||| ||||||||||
Sbjct  857   ACATAATGAATGAAAGG-ATCCTTAAATAACCATAAGGTCGACAGAAAATCTTTAGCAAA  799

Query  1033  GACTTCT-CAAGATGT-CTGTTTT-C-ATAGAAA-AGATTCGCTCAAA--G-AC-C--TA  1081
             ||||||| |||||||| ||||||| | ||||||| |||||||||||||  | || |  ||
Sbjct  798   GACTTCTACAAGATGTTCTGTTTTTCCATAGAAATAGATTCGCTCAAAAAGGACTCCATA  739

Query  1082  GGATGTTA-TCGTA  1094
             |||||||| |||||
Sbjct  738   GGATGTTAATCGTA  725
```

```
>gb|AY386961.1|  Pisum sativum maturase-like protein (matK) gene, complete cds;
chloroplast
Length=1521

 Score = 1162 bits (629),  Expect = 0.0
 Identities = 634/637 (99%), Gaps = 1/637 (0%)
 Strand=Plus/Minus

Query  1     AGGATCTAATTAGAGGAATAATTGGAACTATTATATCCAATTTTTYGWTAACAATTTCGA  60
             |||||||||||||||||||||||||||||||||||||||||||| | ||||||||||||
Sbjct  1123  AGGATCTAATTAGAGGAATAATTGGAACTATTATATCCAATTTTTTGATAACAATTTCGA  1064

Query  61    TTAGAAATGTATTTTGCAGCATTTGACTCCGTACCACTGAACGATTTAGTCGCACATTTG  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1063  TTAGAAATGTATTTTGCAGCATTTGACTCCGTACCACTGAACGATTTAGTCGCACATTTG  1004

Query  121   AAAAATAGCCTAAAAGCTGAAATGAATGTTCGGATAATGGGTTTATATTGATCATTCTTG  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1003  AAAAATAGCCTAAAAGCTGAAATGAATGTTCGGATAATGGGTTTATATTGATCATTCTTG  944

Query  181   GTTGAGACCAAACATCaaaaaaaCATTGCCATAAATGGATAAAATAGTGTTTCCATTTAT  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  943   GTTGAGACCAAACATCAAAAAAACATTGCCATAAATGGATAAAATAGTGTTTCCATTTAT  884

Query  241   TCATCAAAAAAGGCGCATTCTTTGAAGCCAGAATGCATTTTCCTTGATATCGAACATAAT  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  883   TCATCAAAAAAGGCGCATTCTTTGAAGCCAGAATGCATTTTCCTTGATATCGAACATAAT  824

Query  301   GAATGTTAGGATCCTTGAAGAATGTTAAAGGATACGAAAAATCTTTAGCAAAAACTTTTC  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  823   GAATGTTAGGATCCTTGAAGAATGTTAAAGGATACGAAAAATCTTTAGCAAAAACTTTTC  764

Query  361   CAAGATGCTCTCTTTTTCCATaaaaaaaaattcgctcaaaaaaaatgctaaaaGATTTGA  420
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  763  CAAGATGCTCTCTTTTTCCATAAAAAAAAATTCGCTCAAAAAAAATGCTAAAAGATTTGA  704
Query  421  ATCCTAAATGAGAGGATTTATTACGTAGAAAAACGAGGATATATTCATATTCACATACAT  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  703  ATCCTAAATGAGAGGATTTATTACGTAGAAAAACGAGGATATATTCATATTCACATACAT  644

Query  481  AAAAATTATGGAGGAACAAAAATAATCTTGGATTACTTTTTGAAAAAGTCGAAATCGATT  540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  643  AAAAATTATGGAGGAACAAAAATAATCTTGGATTACTTTTTGAAAAAGTCGAAATCGATT  584

Query  541  TTTTGGTAGTAATAAAACTATTCCAATTACAAAAATGATAAAGAAACAGCCGTAATAAAT  600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  583  TTTTGGTAGTAATAAAACTATTCCAATTACAAAAATGATAAAGAAACAGCCGTAATAAAT  524

Query  601  GaaaaaaaGGGGCATCTTTCACCCAGTWATCGAAGGA  637
            |||||||||||||||||||||||||| |||||||||
Sbjct  523  GAAAAAAAGGGGCATCTTTCACCCAGT-ATCGAAGGA  488
```

DNeasy Plant Mini Protocol

The leaf tissue (≤ 100mg wet weight) was ground to a fine powder under liquid nitrogen using a mortar and pestle and transferred to 1.5mL microcentrifuge tube. The fine leaf powder had 400µL of buffer AP1 and 4µL of RNase A added to it. The tube and its contents were vortexed then incubated for 10 minutes at 65°C. The tube and its contents were inverted 3 times during the incubation period. The buffer AP2 (130 µL) was added to the microcentrifuge tube at the end of the incubation. The contents of the tube were mixed and then incubated on ice for 5 minutes. Following incubation, the tube and its contents were centrifuged for 5 minutes at 20 000 x g (14 000 rpm). The lysate was pipetted into a QIAshredder Mini spin column in a 2ml collection tube. Centrifugation at 20 000 x g (14 000 rpm) followed this step. The flow through fraction was transferred into a new tube without disturbing the pellet. Typically 450µL of lysate is recovered and 675µL (1.5 volumes of the lysate) buffer AP3/E were added to the cleared lysate. The contents of the tube were mixed by pipetting. A 650µL portion of the mixture was transferred into a DNeasy Mini spin column in a 2ml collection tube. The tube was centrifuged for 1 minute at 6 000 x g (8 000 rpm). The flow through was discarded. The centrifugation was repeated with the remaining portion of the lysate mixture. The spin column was placed into a new 2ml collection tube. A 500µL portion of buffer AW was added and the tube was centrifuged for 1 minute at 6 000 x g (8 000rpm). The flow through was discarded and the washing of the column was repeated with another 500µL portion of buffer AW but this time centrifugation was for 2 minutes at 20 000 x g (14 000 rpm). The spin column was transferred to a new 1.5ml microcentrifuge tube and 100 µL of Buffer AE was added for elution of the DNA. The tube was incubated at room temperature for 5 minutes. A centrifugation for 1 minute at 6 000 x g (8 000 rpm) followed. The elution was repeated and 2 aliquots of DNA were available. The DNA was quantified then stored at -20°C.

QIAquick PCR purification kit protocol

The PCR product had 5 volumes of Buffer PB added to 1 volume of the PCR sample and mix. A QIAquick spin column was placed in a 2ml collection tube. To bind DNA, the PCR product and buffer sample was applied to the QIAquick column and centrifuged at 20 000 x g (14 00rpm) for 45 seconds. The flow through was discarded. The QIAquick column was placed back into the same collection tube and 0.75ml of buffer PE were added to wash the column. The column and its contents were centrifuged at 20 000 x g (14 000rpm) for 45 seconds and the flow through was discarded. The QIAquick column was centrifuged for an additional 1 minute at maximum speed after which the column was placed in a clean 1.5ml microcentrifuge tube. A 30μL portion of buffer EB was added to the column and allowed to stand for 1 minute after which it was centrifuged for 1 minute to elute the DNA. The PCR product was clean and ready for its downstream use.