African Journal of Biotechnology Vol. 11(97), pp. 16305-16312, 4 December, 2012
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB11.1110
ISSN 1684–5315 ©2012 Academic Journals

Full Length Research Paper

Developing DNA barcoding (*mat*K) primers for marama bean [*Tylosema esculentum* (Burchell) Schreiber]

Takundwa M.[1], Chimwamurombe P. M.[1]*, Kunert K.[2] and Cullis C. A.[3]

[1]Department of Biological Sciences, University of Namibia, P. Bag 13301, Windhoek, Namibia.
[2]Department of Plant Science, University of Pretoria, 0001 South Africa.
[3]Department of Biology, Case Western Reserve University, Cleveland, Ohio, USA.

Accepted 24 October, 2012

DNA barcoding is based on the premise that a short standardized DNA barcoding sequence can distinguish individuals of a species because the genetic variation between species exceeds that within species. Information on genetic variation of breeding materials helps to maintain genetic diversity and sustains long term selection gain. This information is a prerequisite for the genetic improvement of any plant species for effective use of germplasm in breeding and for conservation. Marama bean [*Tylosema esculentum* (Burchell) Schreiber] is found in the arid, dry parts of Southern Africa and due to the high nutrient value of the seeds and tubers, richness in protein, oil and starch, it is a potential crop for arid areas where few conventional crops can survive. The effective conservation and use of marama bean genetic resources for domestication involves investigating the extent of genetic variation. The *mat*K gene, formerly known as *orf*K, is emerging as a DNA barcoding gene with potential contribution to plant molecular systematics and evolution. The gene *mat*K, approximately 1500 base pairs (bp), is believed to code for a maturase-related protein based on structural similarities to other such genes. This gene was investigated for potential contribution in genetic variation studies of marama bean and also establishing a barcode for *T. esculentum*. The *mat*K gene was amplified in marama bean and we reported herein, the first record of sequences of this gene for the species that were found to be related to other legume *mat*K sequences deposited in GenBank. The homology found 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: Marama bean, DNA barcoding, genetic variation, maturase kinase.

INTRODUCTION

Marama bean and the need for molecular markers for this future crop

Marama bean [*Tylosema esculentum* (Burchell) Schreiber] occurs naturally in arid and dry parts of Southern Africa, including Botswana, Namibia and South Africa. In Botswana and Namibia, this species constitutes an important part of the diet of indigenous people in the remote and arid regions (Keegan and van Staden, 1981). Due to the high nutrient value of the seeds and tubers, richness in protein, oil and starch, marama bean has the potential to become a productive crop in arid areas where few conventional crops can survive (Ketshajwang et al., 1998). Marama bean is a wild tuber-producing and non-nodulating legume; the seeds and tubers are edible after roasting and cooking, respectively. Legumes are second only to Poaceae (grasses) in agricultural and economic importance. Efforts are underway to develop marama into a crop and further develop desirable cultivars that are high yielding and early maturing as it is still not yet

*Corresponding author. E-mail: pchimwa@unam.na.

cultivated with its potential contribution to food security (Chimwamurombe, 2008; Nepolo et al., 2009).

Information on genetic variation is a prerequisite for the improvement of any plant species by breeding programs. The natural populations of marama bean are under pressure from both overgrazing and human exploitation of the seeds, therefore a detailed knowledge of the genetic structure of these populations is required for developing the remaining wild germplasm (Naomab, 2004). The surge in application of molecular biology information to systematic and evolutionary questions has resulted in significant contributions to both plant and animal systematics and in the emergence of molecular systematics as a solid interdisciplinary field (Wolfe and Liston, 1998). Among the different approaches used in molecular systematics, DNA sequencing has become one of the most widely used, particularly above the genus level (Bidartondo, 2009). The proposed DNA barcoding gene, *mat*K was one of the DNA regions explored for potential contribution as a molecular marker for marama bean genetic diversity studies.

DNA barcoding and the *matK* gene

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 et al., 2003). 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 the compilation 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.

Species identification through barcoding is usually achieved by the retrieval of a short DNA sequence- the “barcode”- from a standard part of the genome (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 (a new halotype or geographical variant), or it can suggest the existence of a newly encountered species (Hebert et al., 2003). 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 (Saitou and Nei, 1987). In phylogenetic studies, DNA barcoding can be a starting point for optimal selection of taxa and barcode sequences can be added to the sequence dataset 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).

Kress et al. (2005) suggested that the use of the COI sequence “is not appropriate for most species of plants because the cytochrome C oxidase I gene evolves at a much slower rate in higher plants than in animals”. A series of experiments was then conducted to find a more suitable region of the genome for use in the DNA barcoding of flowering plants or the larger group of land plants). One 2005 proposal was the nuclear internal transcribed spacer region and the plastid trnH-psbA intergenic spacer, while other researchers advocated other regions such as *mat*K. Meanwhile, there has been no agreement on which region(s) should be used for barcoding land plants. Therefore, to provide a community recommendation on a standard plant barcode, the CBOL Plant Working Group compared the performance of the seven 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*I 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 (Hollingsworth et al., 2009). 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 towards the discovery of overlooked species of land plants (CBOL Plant Working Group, 2009).

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). 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 et al., 1997). 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 and Hongping, 1997). The presence of the gene in the parasitic *Epifagus,* a taxon that has lost about 65% of its chloroplast genes, speaks for the functional significance of the *mat*K gene in

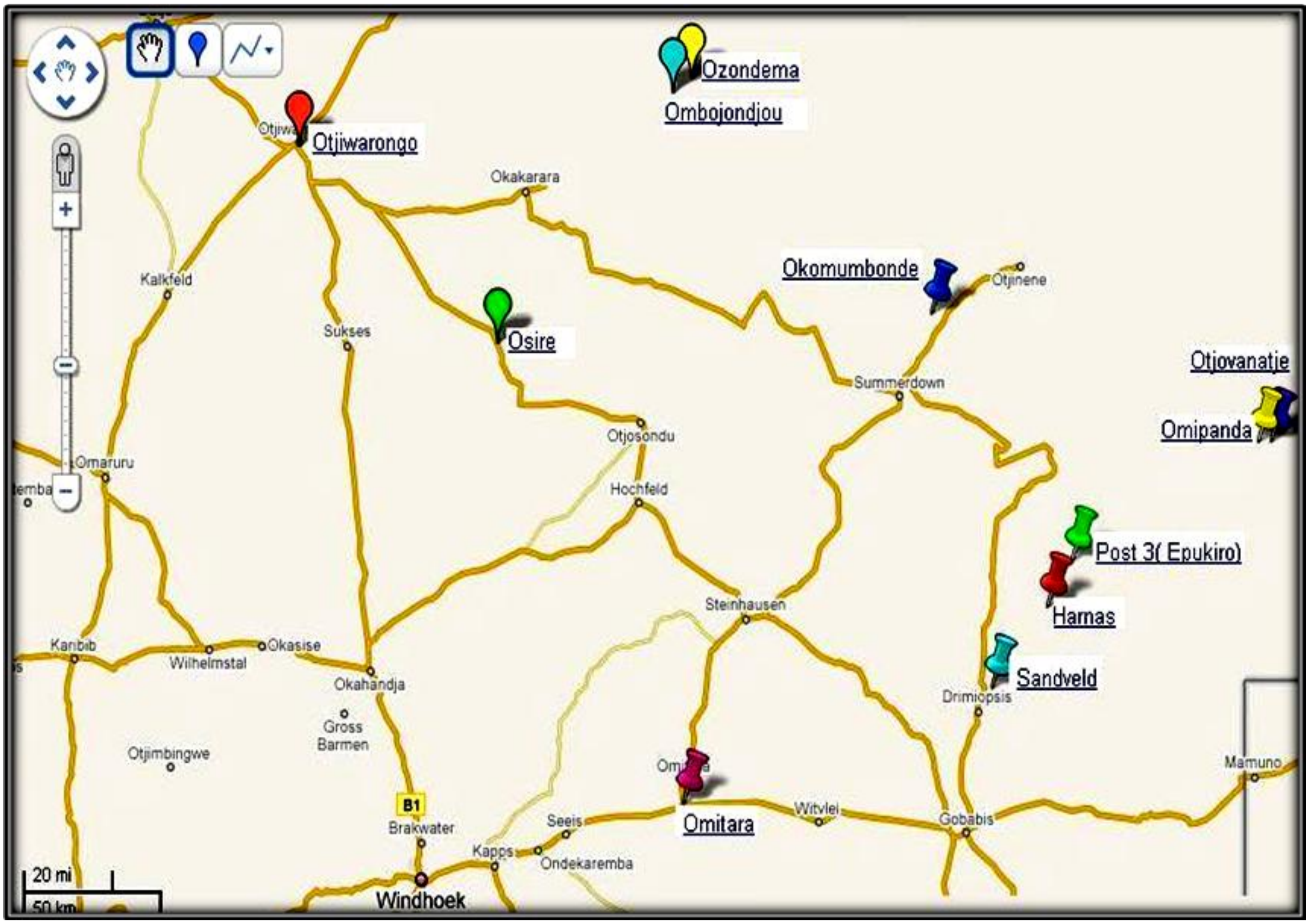


Figure 1. A map representing the chosen study sites in Namibia.

plants. In fact the two exons of the *trn*K gene that flank the *mat*K were lost leaving the gene intact (Wolfe and Liston, 1998).

In addition, the gene contains high substitution rates within the species and is emerging as potential candidate to study plant systematics and evolution (Notredame et al., 2000). The *mat*K gene may be able to contribute to genetic diversity studies in marama bean. It has been found that when one plant species is closely related to another, differences are usually detected in the *mat*K DNA (Lahaye et al., 2008,). The *mat*K gene for *Tylosema esculentum* had not been previously characterized, and so this region in the genome of the species was explored.

Objectives of the study

The objectives of this study were to amplify and characterize the *mat*K gene in *T. esculentum* and to determine the potential of this gene in contributing to studying genetic diversity in marama bean.

MATERIALS AND METHODS

Amplification of the *matK* gene and DNA barcoding

DNA was extracted from each of the plant samples collected from the 12 sampling sites (Figure 1) using the DNeasy mini protocol for purification of total DNA from plant tissue. The manufacturer's protocol was followed to obtain DNA from the plant tissue (Quiagen, 2006) and the DNA was stored in clearly labelled microcentrifuge tubes at -20°C. 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. The *mat*K primer pairs known to amplify the *mat*K region in legumes (Wojciechowski et al., 2004) were used for amplification of the *mat*K gene from the geographically diverse marama bean germplasm collection from the 12 localities namely: Ozondema (OZO), Ombujondjou (OMB), Osire (OSI), and Otjiwarongo (OTR) in Otjozondjupa region, Omitara (OMI) in Khomas region, Sandveld (SAN), Otjovanatje (OTJ), Omipanda (OMP), Post 3/Epukiro (EPK), Harnas (HAR), Okomombonde (OKO) in Omaheke region and Pretoria (PTA) in South Africa.

The primers used were *trn*K686, *trn*K2, *mat*K 4La, *mat*K 1100L, *mat*K 1932Ra, *mat*K 832R. Polymerase chain reaction (PCR) amplifications were performed in 25 µL reaction volumes, with a 2X

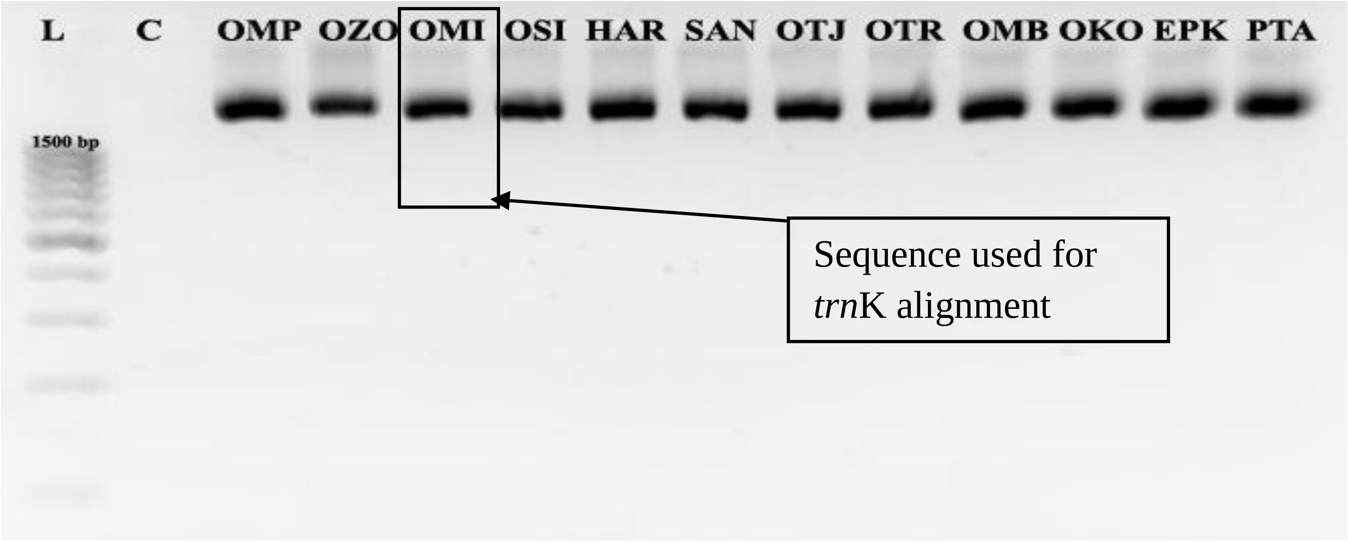


Figure 2. Amplification of the *trn*K gene in *Tylosema esculentum*. C, Negative control; L, 100-bp ladder. The rest of the lanes contain samples from each of the 12 localities, and products were more than 1500 bp.

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°C for 4 min, followed by 35 cycles of denaturation at 95°C for 30 s, an annealing at 55°C for 60 s and an extension at 72°C for 2 min, a final extension at 72°C for 5 min and then held at 4°C. Agarose gel (2.5%) visualization of PCR products was then used to determine if PCR products were 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*.

Distance analysis and evolutionary tree construction

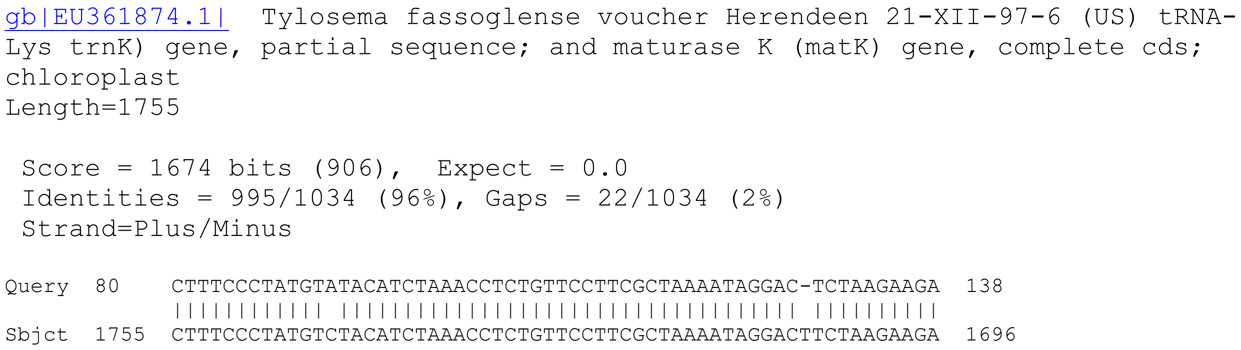
The basic local alignment search tool (BLAST), which is available at (http://www.ncbi.nlm.nih.gov/BLAST/) was used to find homologous sequences to those obtained for marama bean. BLAST uses the query sequence, which was that for marama bean in this case to search various databases to look for similar 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. When the sequences had been confirmed to be *mat*K and *trn*K sequences, a phylogenetic tree based on the *mat*K gene was constructed to investigate if the *mat*K gene would be useful in genetic diversity studies of *T. esculentum*. The species *Tylosema fassoglensis*, which is in the same genus, was also included in the analysis together with other members of the Cercideae tribe in the Fabaceae family to determine sequence divergence among the samples. The sequences were downloaded from NCBI GenBank. A ClustalW multiple sequence alignment was run in MEGA 4.0 and distance analysis was also carried out with MEGA 4.0. The Jukes-Cantor (one parameter) model was used to calculate distance and to calculate the tree and construct the phylogeny; the neighbour-Joining clustering algorithm with bootstrap of 1000 replications was used. Primer 3 software available online was used to design a marama specific primer pair for the *mat*K gene based on the consensus sequence.

RESULTS

*mat*K and *trn*K amplifications and BLAST alignments

The PCR products for the *trn*K genes were expected in the range of 2500 bp, while for *mat*K the expected product size was 1500 bp. Figure 2 shows the amplification products obtained. The BLAST searches gave matches with legume sequences from the NCBI database. The partial *trn*K sequence of *T. esculentum* aligned with that of *T. fassoglensis* with 96% identities. An alignment of the amino acid translations showed that 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 *T. fassoglensis* (Figure 3). The partial *mat*K sequence of *T. esculentum* aligned with that of *Pisum sativum* with 99% identities. An alignment of the amino acid translations showed that the *T. esculentum* sequence was at the N-terminus of the *mat*K gene when aligned with that of *P. sativum*. 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 *mat*K region it was 637 base pairs (Figures 4 to 6). The conserved regions are marked with an asterisk in Figure 7 and the phylogenetic tree constructed based on the *mat*K sequences for *T. esculentum* generated in this study is shown in Figure 8. The sequence for a member of the tribe Cercideae, *Bauhinia variegata* was used as an out-group to root the tree.

Figure 8 also shows the evolutionary relationships of 12 marama bean sequences based on the *mat*K gene. The evolutionary history was inferred using the neighbour-joining method. The bootstrap consensus tree inferred



```
gb|EU361874.1|  Tylosema fassoglense 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 3. The BLAST result for the *trn*K sequence alignment.

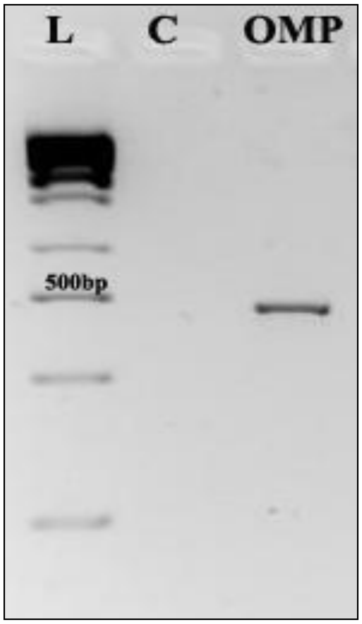
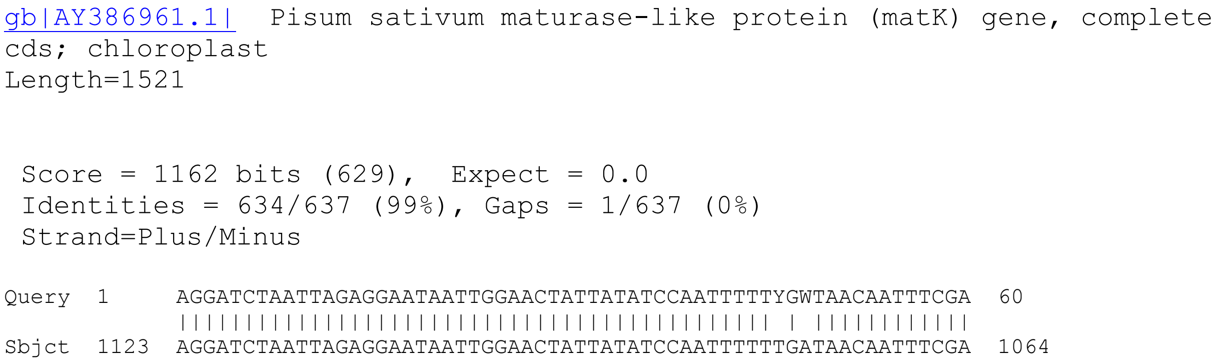


Figure 4. Amplification of the *mat*K gene in *Tylosema esculentum*. C, Negative control; L, 100-bp ladder. The lane OMP contains a sample from Omipanda which gave the longest *mat*K sequence of 637 bp.



```
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 5. The BLAST result for the *mat*K sequence alignment.

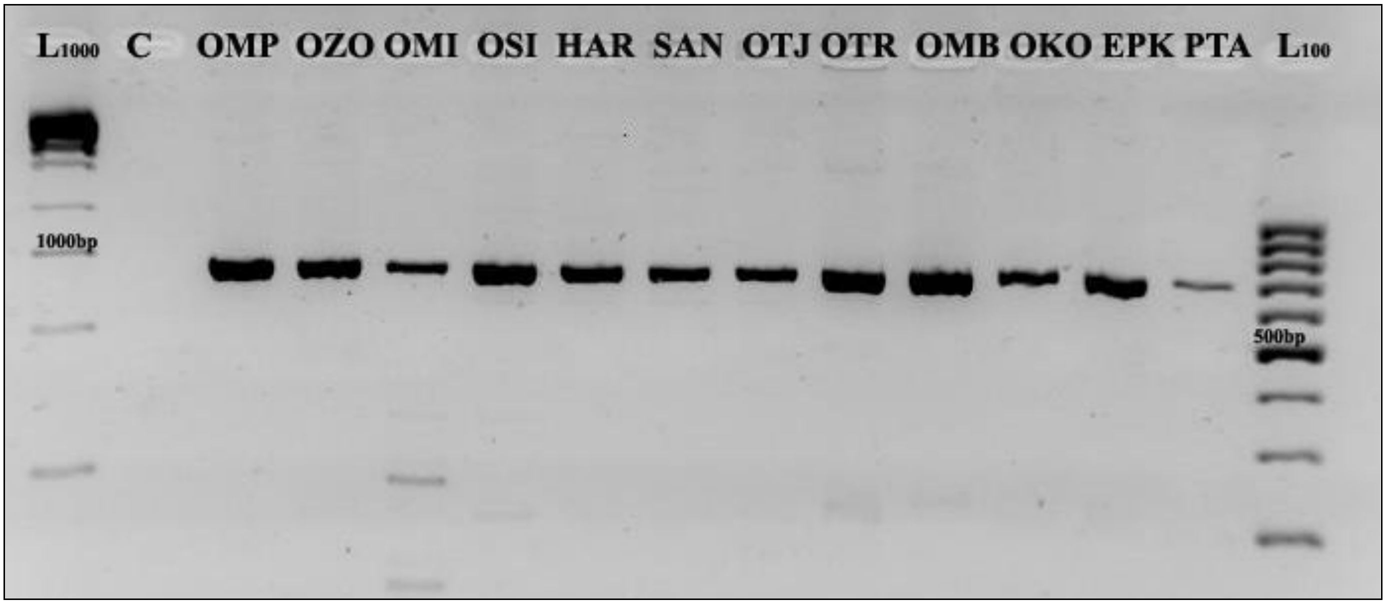


Figure 6. Amplification of the *mat*K gene in *Tylosema esculentum*. C, Negative control; L, 100-bp ladder. The rest of the lanes contain a sample from each of the 12 study sites.

from 1000 replicates was taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 10% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The evolutionary distances were computed using the Jukes-Cantor method and are in the units of the number of base substitutions per site. Codon positions included were 1st+2nd+3rd+noncoding. All positions containing alignment gaps and missing data were eliminated only in pair-wise sequence comparisons (Pair-wise deletion option). There were a total of 1111 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4. Three clades formed with bootstrap values higher than 80% were depicted by the colour coding of the branches. A primer pair for the consensus region was designed using Primer 3 software available online (http://frodo.wi.mit.edu/primer3).

Left primer: TCATCCAAGATTTTTCTTGTTCC
Right primer: CAAAAAGTAGAATGAATGCTCAGA

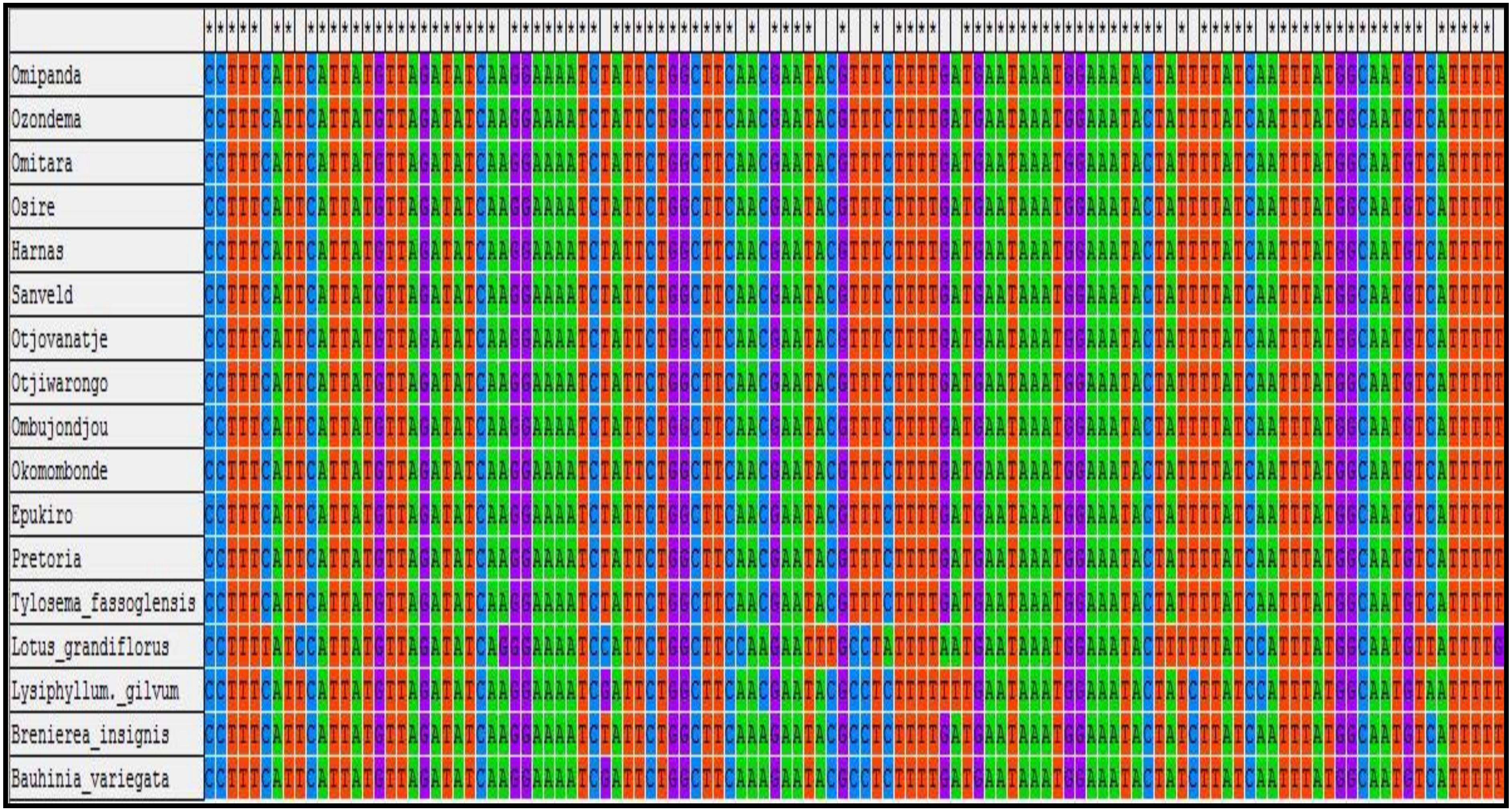
DISCUSSION

The *mat*K sequences were amplified and sequenced for the species *T. esculentum* as a contribution to DNA barcoding and efforts by the CBOL to barcode land plants. The gene was investigated in marama bean for possible contribution in genetic diversity studies and found to be useful. 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 *T. fassoglensis* in the case of the *trn*K gene confirmed that this was the region amplified for *T. esculentum* (Figure 3). The homology with *P. sativum* in the case of the *mat*K gene confirmed that this was the region that had been amplified in the PCR reaction (Figure 5). The primer pair designed can subsequently be tested in the genus *Tylosema* and used for barcoding of *T. esculentum*.

All the marama bean sequences from the 12 localities formed a clade together with *Lysiphyllum gilvum* and *T. fassoglensis* with the highest boot-strap value of 94% in the tree (Figure 8). The tree suggests there has been evolution in this gene as the individuals from different localities formed different groups from the collection of *T. esculentum* individuals.

However, the clustering with another species of the same genus *T. fassoglensis* and a member of another genus altogether, *L. gilvum* cautions the use of coding regions such as *mat*K in genetic studies as the rate of sequence divergence is generally low (Devey et al., 2009). High levels of sequence divergence allows for greater resolution at lower taxonomic levels (for example within families or genus). The *mat*K gene therefore could be useful to genetic diversity studies in *T. esculentum* to an extent, but the sequence divergence is low and the clade or monophyletic group support by bootstrap values formed in this study were low.

The low sequences divergence within *T. esculentum* could be due to the fact that marama bean in the Namibian geographic range exists in small sub-populations in the localities sampled here and the South African population in Pretoria at the University farm. The plant is thought to be pollinated by insects and seed are dispersed by small mammals, hence a coding gene such



	***** ** ***************** ******** *********** * **** * * **** ****************** * ***** ************** *****
Omipanda	CCTTTCATTCATTATGTTAGATATCAAGGAAAATCTATTCTGGCTTCAACGAATACGTTTCTTTTGATGAATAAATGGAAATACTATTTTATCAATTTATGGCAATGTCATTTTT
Ozondema	CCTTTCATTCATTATGTTAGATATCAAGGAAAATCTATTCTGGCTTCAACGAATACGTTTCTTTTGATGAATAAATGGAAATACTATTTTATCAATTTATGGCAATGTCATTTTT
Omitara	CCTTTCATTCATTATGTTAGATATCAAGGAAAATCTATTCTGGCTTCAACGAATACGTTTCTTTTGATGAATAAATGGAAATACTATTTTATCAATTTATGGCAATGTCATTTTT
Osire	CCTTTCATTCATTATGTTAGATATCAAGGAAAATCTATTCTGGCTTCAACGAATACGTTTCTTTTGATGAATAAATGGAAATACTATTTTATCAATTTATGGCAATGTCATTTTT
Harnas	CCTTTCATTCATTATGTTAGATATCAAGGAAAATCTATTCTGGCTTCAACGAATACGTTTCTTTTGATGAATAAATGGAAATACTATTTTATCAATTTATGGCAATGTCATTTTT
Sanveld	CCTTTCATTCATTATGTTAGATATCAAGGAAAATCTATTCTGGCTTCAACGAATACGTTTCTTTTGATGAATAAATGGAAATACTATTTTATCAATTTATGGCAATGTCATTTTT
Otjovanatje	CCTTTCATTCATTATGTTAGATATCAAGGAAAATCTATTCTGGCTTCAACGAATACGTTTCTTTTGATGAATAAATGGAAATACTATTTTATCAATTTATGGCAATGTCATTTTT
Otjiwarongo	CCTTTCATTCATTATGTTAGATATCAAGGAAAATCTATTCTGGCTTCAACGAATACGTTTCTTTTGATGAATAAATGGAAATACTATTTTATCAATTTATGGCAATGTCATTTTT
Ombujondjou	CCTTTCATTCATTATGTTAGATATCAAGGAAAATCTATTCTGGCTTCAACGAATACGTTTCTTTTGATGAATAAATGGAAATACTATTTTATCAATTTATGGCAATGTCATTTTT
Okomombonde	CCTTTCATTCATTATGTTAGATATCAAGGAAAATCTATTCTGGCTTCAACGAATACGTTTCTTTTGATGAATAAATGGAAATACTATTTTATCAATTTATGGCAATGTCATTTTT
Epukiro	CCTTTCATTCATTATGTTAGATATCAAGGAAAATCTATTCTGGCTTCAACGAATACGTTTCTTTTGATGAATAAATGGAAATACTATTTTATCAATTTATGGCAATGTCATTTTT
Pretoria	CCTTTCATTCATTATGTTAGATATCAAGGAAAATCTATTCTGGCTTCAACGAATACGTTTCTTTTGATGAATAAATGGAAATACTATTTTATCAATTTATGGCAATGTCATTTTT
Tylosema_fassoglensis	CCTTTCATTCATTATGTTAGATATCAAGGAAAATCTATTCTGGCTTCAACGAATACGTTTCTTTTGATGAATAAATGGAAATACTATTTTATCAATTTATGGCAATGTCATTTTT
Lotus_grandiflorus	CCTTTTATCCATTATGTTAGATATCAGGGAAAATCCATTCTGGCTTCCAAGAATTTGCCTATTTTAATGAATAAATGGAAATACTTTTTTATCCATTTATGGCAATGTTATTTTG
Lysiphyllum._gilvum	CCTTTCATTCATTATGTTAGATATCAAGGAAAATCGATTCTGGCTTCAACGAATACGCCTCTTTTTTTGAATAAATGGAAATACTATCTTATCCATTTATGGCAATGTAATTTTT
Brenierea_insignis	CCTTTCATTCATTATGTTAGATATCAAGGAAAATCTATTCTGGCTTCAAAGAATACGCCTCTTTTGATGAATAAATGGAAATACTATCTTATCAATTTATGGCAATGTCATTTTT
Bauhinia_variegata	CCTTTCATTCATTATGTTAGATATCAAGGAAAATCGATTCTGGCTTCAAAGAATACGCCTCTTTTGATGAATAAATGGAAATACTATCTTATCAATTTATGGCAATGTCATTTTT

Figure 7. A section of the multiple alignment of the *mat*K sequences of *Tylosema esculentum* from different localities and other Fabaceae. The alignment was generated with the software MEGA 4.0.

as *mat*K would be expected to be conserved between populations as indicated by the clustering of the marama bean samples into one clade. The sub-trees within this clade are not well supported as their bootstrap values are low; hence the focus in this discussion was on the clade with the higher bootstrap of 94%. This clade further supports the strongly held view that diversity in marama bean is within, rather than between populations (Takundwa et al., 2010; Nepolo et al., 2009; Naomab, 2004; Monaghan and Halloran, 1996).

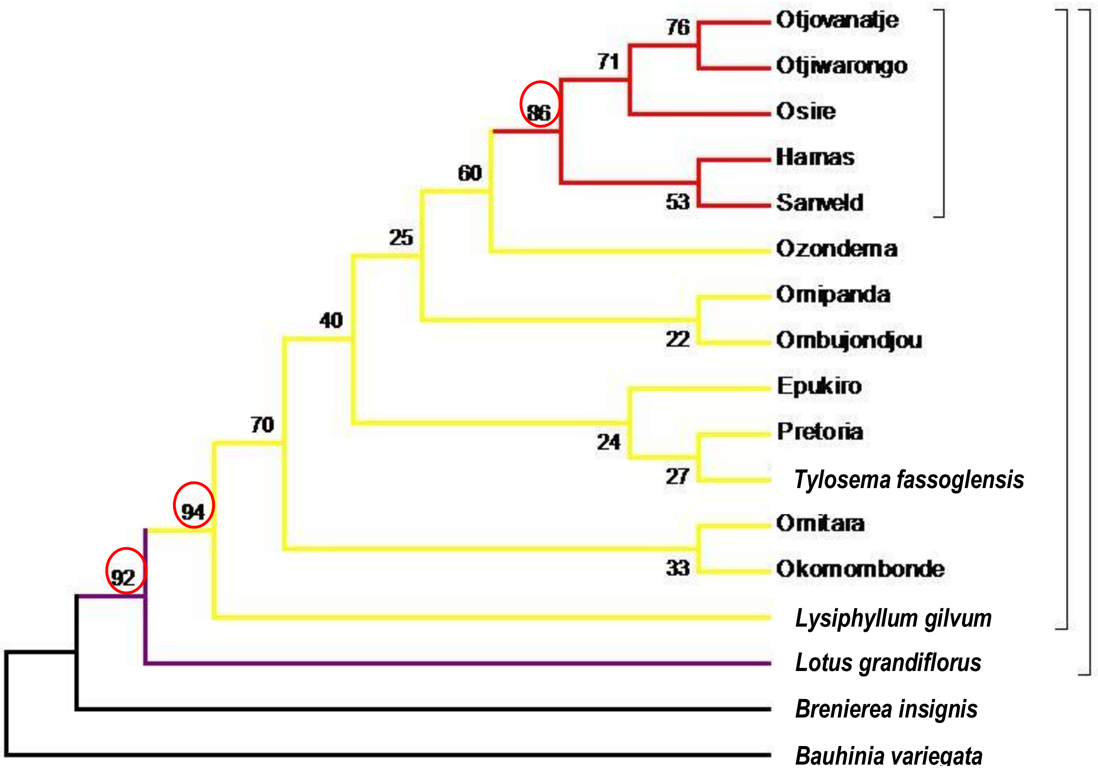


Figure 8. A phylogenetic tree for *Tylosema esculentum* based on the *mat*K gene. The bootstrap values show the confidence in the groupings as a percentage. The sequence for *Bauhinia variegata* was used as an out-group to root the tree generated with the software MEGA 4.0.

REFERENCES

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

CBOL Plant Working group (2009). A DNA barcode for land plants. Proc. Nat. Acad. Sci. 106(31):12792-12797.

Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD (1993). Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. Annals Missouri Botanic Garden 80:528-580.

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. Build. Bridges Poverty Reduct. Sustain. Dev. 3:5-7.

Devey DS, Chase MW, Clarkson JJ (2009). A stuttering start to plant DNA barcoding: microsatellites present a previously overlooked problem in non-coding plastid regions. Taxon 58:7-17.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003). Biological identifications through DNA barcodes. Proc. Roy. Soc. London: Biol. Sci. 270:313-321.

Hollingsworth ML, Clark A, Forrest LL, Richardson J, Pennington RT, Long DG, Cowan R, Chase MW, Gaudeul M, Hollingsworth PM (2009). Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. Mol. Ecol. 9(2):439-457.

Keegan AB, van Staden J (1981*). Tylosema esculentum*, a plant worthy of cultivation. S. Afr. J. Sci. 77:387-397.

Ketshajwang KK, Holmback J, Yeboah SO (1998). Quality and compositional studies of some edible Leguminosae seed oils in Botswana. J. Am. Oil Chem. Soc. 75:741-743.

Khidir WH, Hongping L (1997). The *mat*K gene: sequence variation and application in plant systematics. Am. J. Bot. 19:830-839.

Kress WJ, Wurdak KJ, Zimner EA, Weight LA, Janzen DH (2005). Use of DNA barcodes to identify flowering plants. Proc. Natl. Acad. Sci. USA 102(23):8369-8374.

Lahaye R, van der Bank M, Bogarin D (2008). DNA barcoding the floras of biodiversity hotspots. Proc. Natl. Acad. Sci. USA 105: 2923-2928.

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.

Nepolo E, Takundwa M, Chimwamurombe P, Cullis CA, Kunert K (2009). A review of the geographical distribution of marama bean [*Tylosema esculentum*(Burchell) Schreiber] and genetic diversity in the Namibian germplasm. Afr. J. Biotechnol. 8:2088-2093.

Notredame C, Higgins DG, Heringa J (2000). T-Coffe: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302(1):205-217.

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

Saitou N, Nei M (1987). The neighbour-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406-425.

Takundwa M, Chimwamurombe PM, Kunert K, Cullis CA (2010). Isolation and characterization of microsatellite repeats in Marama bean (*Tylosema esculentum*). Afr. J. Agric. Res. 5(7):561-566.

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

Wojciechowski MF, Lavin M, Sanderson MJ (2004). A phylogeny of legumes (*Leguminosae*) based on analysis of the plastid *mat*K gene resolves many well supported subclades within the family. Am. J. Bot. 91(11):1846-1862.

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.