IDENTIFICATION OF SPECIFIC MARKERS LINKED TO REGIONAL DIFFERENTIATION OF Warburgia ugandensis IN KENYA

OCHIENG’ NOEL ONYANGO (BSc.)
156/13839/05

A thesis submitted in partial fulfilment of the requirements for the award of the Degree of Master of Science (Biotechnology) of Kenyatta University.
DECLARATION

I, Ochieng’ Noel Onyango, duly declare that this thesis is my original work and has not been presented for a degree in any other university or any other award.

Signature: Ochieng’ Noel Onyango
Date: 25/8/2011

We confirm that the work reported in this thesis was carried out by the candidate under our supervision

Supervisors

Dr. Alice Muchugi,
Department of Biochemistry and Biotechnology,
Kenyatta University,
P. O Box 43844,
Nairobi, Kenya.

Signature: Dr. Alice Muchugi
Date: 25/8/2011

Dr. Ramni Jamnadass,
Tree Genetic Resources and Conservation (GRP1),
World Agroforestry Centre-ICRAF,
P. O Box 30677-00100,
Nairobi, Kenya.

Signature: Dr. Ramni Jamnadass
Date: 25/8/2011

Dr. Bonaventure Omondi Aman,
Department of Plant Protection Sciences,
Swedish University of Agricultural Sciences,
P. O. Box 102 SE 230 53,
Alnarp, Sweden.

Signature: Dr. Bonaventure Omondi Aman
Date: 18th August 2011
DEDICATION

I dedicate this work to my father Gershon Onyango Owuoth who never had a chance to read it and my entire family members for patience.
ACKNOWLEDGEMENTS

I am extremely grateful to very special people who made this strenuous work possible. My supervisors Dr. Ramni Jamnadass of World Agroforestry Centre-ICRAF, Dr. Bonaventure Omondi Aman of Swedish University of Agricultural Sciences and Dr. Alice Muchugi of Kenyatta University who offered exemplary advice and assistance. I would like to also thank, Dr. Paul Mireji, Dr. Steve Ger and Dr. Daniel Masiga of International Centre of Insect Physiology Ecology-ICIPE, Dr. Rob Skilton of International Livestock Research Institute-ILRI and Biosciences Eastern and Central Africa-SEGOLI, Dr. Edward Muge of University of Nairobi. I thank you all.

Lots of thanks to Dr. Steve Runo of Kenyatta University, Dr. Francis MacOdimba of Aga Khan University, Dr. Okoth Ayugi of Kenya polytechnic University college, Dr. Ian Dawson, Dr. Roeland Kindt, Hellen Ochieng’, Brenda Kingolla, Kennedy Olale, Moses Mujuga, Celine Oduor, Hilda Kegode, Evelyne Ndenga, Emma Akinyi, Morine Aringo, Anne Omollo, Anne Kuria, Agnes Gachuiri, Philip Kinga, Agnes Were and Nelly Mutio all of World Agroforestry Centre-ICRAF, Julius Osaso of Applied Biosystems International-Africa (ABI-Africa) and Delia Ouko Wasawo of International Livestock Research Institute-ILRI for giving valuable scientific advice during the study.

With profound gratitude and great humility I would like to appreciate the very much assistance from the International Centre of Insect Physiology Ecology (ICIPE) in DNA cloning, Biosciences Eastern and Central Africa- BECA for genome sequencing and World Agroforestry Centre-ICRAF through Tree Genetic Resources and Conservation for supporting this work. Without these institutions this work could have not have been a success.

This work was funded by the World Agroforestry Centre-ICRAF, through Tree Genetics Resources and Conservation.

Last but not least, special thanks to my entire family but more to my mum Ascar Onyango, brother in Law Dr. Okoth Ayugi, brother Dr. Clayton Onyango, sisters Emily Onyango, Beryl Onyango, Juliet Onyango, Joan Okoth and nephew Cyril Otumba for their continued support during my research work.
# TABLE OF CONTENTS

DECLARATION ...........................................................................................................................(i)

DEDICATION ...................................................................................................................................(ii)

ACKNOWLEDGEMENTS ...................................................................................................................(iii)

TABLE OF CONTENTS ...................................................................................................................(iv)

LIST OF TABLES ...........................................................................................................................(ix)

LIST OF FIGURES ...........................................................................................................................(xi)

LIST OF SEQUENCES .......................................................................................................................(xiii)

ABBREVIATIONS AND ACRONYMS ............................................................................................(xiv)

ABSTRACT ...........................................................................................................................................(xv)

CHAPTER ONE ......................................................................................................................................(1)

INTRODUCTION .................................................................................................................................(1)

1.1 Background of the study ...........................................................................................................(1)

1.2 Statement of the problem and justification ..............................................................................(4)

1.3 Research questions ...................................................................................................................(5)

1.4 Research hypothesis ...................................................................................................................(5)

1.5 Objectives of the study .............................................................................................................(6)

1.5.1 General objectives ...................................................................................................................(6)

1.5.2 Specific objectives ...................................................................................................................(6)
CHAPTER TWO ................................................................. 7

LITERATURE REVIEW ......................................................... 7

2.1 Taxonomy and distribution of species in the genus Warburgia .......... 7

2.2 Uses of Warburgia ugandensis ........................................... 7

2.3 Warburgia ugandensis species natural’s habitat ......................... 9

2.4 Genetic differentiation and evolutionary diversification ............... 13

2.5 Drivers of Genetic Evolution in Plant’s ................................... 15

2.5.1 Forest destruction in the Afro-alpine: a consequence to speciation 16

2.6 Geographical barriers are hindrance to gene flow ...................... 18

2.7 Mutagens (Transposable Elements) a cause of plant speciation ...... 19

2.8 Bulk segregant analysis (BSA) ............................................ 22

2.9 Marker assisted selection .................................................. 23

2.10 Random amplified polymorphic DNA .................................. 24

2.11 Cloning, transformation and sequencing of RAPD markers ........ 27

2.12 Bioinformatics tools for comparative sequence genornics .......... 28

2.12.1 Protein domains and sequences patterns in evolution ........... 29

2.13 Phylogenetic Analysis ..................................................... 31

2.13.1 Genetic distance ...................................................... 32

2.13.2. Disparity index .................................................... 33

CHAPTER THREE .................................................................. 35
<table>
<thead>
<tr>
<th>MATERIALS AND METHODS</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Sample collection</td>
<td>35</td>
</tr>
<tr>
<td>3.2 DNA extraction by Cetyltrimethyl Ammonium Bromide (CTAB)</td>
<td>36</td>
</tr>
<tr>
<td>3.3 RNase digestion</td>
<td>38</td>
</tr>
<tr>
<td>3.4 DNA Quantification</td>
<td>39</td>
</tr>
<tr>
<td>3.4.1 Agarose gel electrophoresis and ethidium bromide fluorescence</td>
<td>39</td>
</tr>
<tr>
<td>3.5 Bulk Segregrant Analysis</td>
<td>40</td>
</tr>
<tr>
<td>3.6 Random Amplified Polymorphic DNA Analysis</td>
<td>40</td>
</tr>
<tr>
<td>3.7 PCR products resolution, bands selection and gel recovery</td>
<td>41</td>
</tr>
<tr>
<td>3.7 Purification of amplified products</td>
<td>42</td>
</tr>
<tr>
<td>3.8 Adenylation</td>
<td>42</td>
</tr>
<tr>
<td>3.9 DH5α Competent cells preparation</td>
<td>44</td>
</tr>
<tr>
<td>3.10 Cloning and transformation of PCR products</td>
<td>45</td>
</tr>
<tr>
<td>3.11 Analysis of recombinants by PCR</td>
<td>47</td>
</tr>
<tr>
<td>3.12 Plasmid DNA preparation</td>
<td>47</td>
</tr>
<tr>
<td>3.14 DNA sequencing</td>
<td>48</td>
</tr>
<tr>
<td>3.15 Sequence similarity search by BLAST</td>
<td>48</td>
</tr>
<tr>
<td>3.16 Multiple Sequence Alignment</td>
<td>49</td>
</tr>
</tbody>
</table>

CHAPTER FOUR

RESULTS

4.1 Identification of RAPD markers showing East and West differentiations
4.2 Cloning, plasmid purification and analysis of the transformants ....57
4.3 NCBI similarity searches results ............................................57
4.3.2 Homologous sequence analysis data information..............57
4.3.1 Homologous sequence information in NCBI ....................59
4.4 Phylogenetic analysis of W. ugandensis and homologous sequences 70

CHAPTER FIVE .............................................................................78

DISCUSSION, CONCLUSION AND RECOMMENDATIONS ..........78

5.1 Discussion............................................................................78
5.2 Conclusion ..........................................................................82
5.3 Recommendations ..............................................................83

REFERENCES ...........................................................................84

APPENDICES .............................................................................104

APPENDIX I .............................................................................104
Primers used for RAPD screening and transformants analysis ....104
Annex 1 ....................................................................................104
Annex 2 ....................................................................................104

Appendix II ...............................................................................105
Sequence’s retrieved from Warburgia ugandensis ..................105

Appendix III .............................................................................107

pGEMT®- T easy vector cloning maps ......................................107
Appendix IV ................................................................. 108

pGEM®-T Easy Vector circle map and sequence reference points ........ 108

Appendix V ..................................................................... 109

Media preparation ................................................................ 109

Appendix VI ................................................................. 111

Purification of amplified products ........................................ 111
LIST OF TABLES

Table 1 Location of *W. ugandensis* samples in the study .....................37

Table 2 RAPD-PCR Loading order from of samples .................................43

Table 3 Five arbitrary primers used in the study ......................................44

Table 4 Amplified band products gel excised for cloning ..........................52

Table 5 *W. ugandensis* BLASTX and TBLASTX homologous sequences ....59

Table 6 *WarburgiaIC55W* TBLASTX homologous sequence features .........61

Table 7 *WarburgiaIC15W* TBLASTX homologous sequence features in NCBI .................................................................62

Table 8 *WarburgiaIC28E* TBLASTX homologous sequence features in NCBI .................................................................63

Table 9 *WarburgiaIC55W, WarburgiaIC15W* and *WarburgiaIC28E* TBLASTX homologous sequence features in NCBI .................................64

Table 10 Cellular features and biological functions observed in NCBI database searches of significant homologs ......................................................65

Table 11 Multiple Sequence Alignment of *WarburgiaIC55W* and homologous sequences ..............................................................................66
Table 12 Multiple Sequence Alignment of *Warburgia IC15W* and homologous sequences ................................................................. 68

Table 13 Multiple Sequence Alignment of *Warburgia IC28E* and homologous sequences ................................................................. 69

Table 14 Multiple Sequence Alignment of *Warburgia ugandensis* and 3 significant homologous sequences ........................................... 69

Table 15 Colours used to represent different amino acids in the MSA ........ 70

Table 17 Pair-wise distances of *Warburgia IC55W* and disparity index ....... 71

Table 18 Pair-wise distances for *Warburgia IC15W* and disparity index ...... 73

Table 19 Pair-wise distances for *Warburgia IC28E* and disparity index ...... 75

Table 20 Pair-wise distances for all sequences and disparity pattern index ... 77
LIST OF FIGURES

Figure 1 Warburgia ugandensis herbal products available in East Africa ........8

Figure 2 Warburgia salutaris herbal products available from South Africa .......8

Figure 3 W. ugandensis Sesquiterpenes’s (a) Warburganal (b) Polygodial ......9

Figure 4 An illustration of species richness in the Afro-alpine belt .............11

Figure 5 Natural habitat fragmentations in the Great Rift Valley .............14

Figure 7 Warburgia ugandensis debarked for medicinal purposes .............18

Figure 8 Typical habitat of W. ugandensis in Laikipia District, Kenya .........18

Figure 9 Locations of W. ugandensis material used in the study ...............36

Figure 10 RAPD profiles of W. ugandensis using primer IC21. ...............52

Figure 11 RAPD profiles of W. ugandensis using primer IC21 replicate gel.53

Figure 12 RAPD profiles of W. ugandensis using primer IC11. ...............54

Figure 13 RAPD profiles of W. ugandensis using primer IC55. ...............55

Figure 14 RAPD profiles of W. ugandensis using primer IC28. ...............56

Figure 15 Alignments of Warburgia55W sequence against homologues........66

Figure 16 Alignments of Warburgia11W sequence against homologues........67

Figure 19 Phylogram of 3 homologous sequences of WarburgiaIC55W .......71
Figure 20 Phylogram of 3 homologous sequences of WarburgiaIC15W......72

Figure 21 Phylogram of 3 homologous sequences of WarburgiaIC28E......74

Figure 22 Phylogram of the homologous sequences to all W. ugandensis significant sequences. .................................................................76
LIST OF SEQUENCES

Sequence contig 1 WarburgiaIC28E (736bp)..................................................105

Sequence contig 2 WarburgiaIC11W (338bp)..................................................105

Sequence contig 3 WarburgiaIC55E (796bp)..................................................105

Sequence contig 4 WarburgiaIC21W (292bp)..................................................105

Sequence contig 4 WarburgiaIC11W (358bp)..................................................106

Sequence contig 6 WarburgiaIC15W (845bp)..................................................106
# ABBREVIATIONS AND ACRONYMS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFLP</td>
<td>Amplified Fragment Length Polymorphism</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td>CTAB</td>
<td>Cetyltrimethyl ammonium bromide</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>DNTPs</td>
<td>Deoxyribonucleotides triphosphates</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylene diamine tetra-acetic acid</td>
</tr>
<tr>
<td>EST</td>
<td>Express Sequence Tags</td>
</tr>
<tr>
<td>GIS</td>
<td>Geographic Information System</td>
</tr>
<tr>
<td>GRP1</td>
<td>Global Research Priority1</td>
</tr>
<tr>
<td>ICRAF</td>
<td>International Centre for Research in Agroforestry</td>
</tr>
<tr>
<td>KEFRI</td>
<td>Kenya Forestry Research Institute</td>
</tr>
<tr>
<td>KWS</td>
<td>Kenya Wildlife Service</td>
</tr>
<tr>
<td>RAPD-PCR</td>
<td>Random Amplified Polymorphic DNA-Polymerase Chain Reaction</td>
</tr>
<tr>
<td>RFLP</td>
<td>Restriction Fragment Length Polymorphism</td>
</tr>
<tr>
<td>Seq.</td>
<td>Sequence</td>
</tr>
<tr>
<td>SOB</td>
<td>Super Optimal Broth</td>
</tr>
<tr>
<td>SOC</td>
<td>Super Optimal Culture</td>
</tr>
<tr>
<td>Rpm</td>
<td>Revolutions per minute</td>
</tr>
</tbody>
</table>
ABSTRACT

*Warburgia ugandensis* Sprague is an important medicinal tree with a restricted distribution in the tropical afro-alpine environment of Africa. The species population in Kenya has shown a high genetic differentiation with respect to the Rift Valley. The kind of differentiation indicated a possible allopatric speciation which may confer species adaptation advantages. The objective of the study was to identify specific genetic markers linked to the regional differentiation of *W. ugandensis* within and across Kenyan Rift Valley. Nine populations of *W. ugandensis* were analysed by an adaptation of Bulk Segregant Analysis employing Random Amplified Polymorphic DNA (RAPD) marker technique. Five primers showed putative East and West genetic differentiation. The diagnostic markers were isolated, cloned into a pGEM-T® vector and confirmed by sequencing. The resulting sequences were analysed by searching for similarity from the genomic databases of National Centre of Biotechnology Institute (NCBI) using BLAST algorithm. Of the 5 sequencing products, only 3 (*WarburgiaIC15E, WarburgiaIC55E* and *WarburgiaIC28W*) showed significant association with plants and bacterial-like chromosomal sequence homolog’s with very low E-values. A sequence alignment from CLUSTALW showed significant conserved protein regions or domain of both plants and some bacteria like sequences. Phylogenetic analysis revealed high rates of genetic distances (of an average 0.8) and a low rate of disparity indices of (0), suggesting some evolutionary forces behind the demographic differentiation. These results imply that the genetic differentiation observed among *W. ugandensis* population in Kenya may be as a result of genetic mutants (eg Transposable Elements) in certain domains of the chromosome. This may have some implication on genome functionality, although not confirmed in this study.
CHAPTER ONE

INTRODUCTION

1.1 Background of the study

Anthropogenic destruction of forests mainly by fire, cultivation, overgrazing, timber logging and fragmentation of the formerly continuous natural landscape is on the increase and has become a significant threat to the maintenance of biodiversity in many terrestrial ecosystems of important tree species (Young et al., 1996). Over-reliance on some plants for medicinal purposes has reshaped the original state of the species habitat, leading to a restricted and scattered remnant (Ryder, 2009). Despite their medicinal value, the depletion of these plants through an expanded local and global market has increased pressure on a resource that is already shrinking in wild habitats (Gurib-Fakim, 2006).

Warburgia ugandensis which is an important medicinal tree that belongs to the genus Warburgia of the family Canellaceae, a small family of tropical tree species with restricted distribution in afro-alpine environment, faces an evolutionary and possibly extinction consequences and therefore needs conservation (Muchugi et al., 2008). These kind of genetic differences observed in the East of the Kenyan Rift Valley compared to its genetic diversity could have been brought about by the past historical and ecological causes and consequences (Hughes et al., 2008), either habitat fragmentation and decline in population size leading to genetic drift causing loss of important
alleles in a stochastic fashion (Suzuki et al., 1989). Genetic drift is a random alteration in frequency of alleles that can weaken the survival of species populations consequently causing speciation (Guo et al., 2009).

The population of *W. ugandensis* found to the Western region of the Kenyan Rift Valley compared to the rest of the population under study has shown a significant genetic differentiation indicating speciation (Muchugi et al., 2008). Similar genetic disjunction had been observed in other important indigenous tree species within the afro-alpine region of the Kenyan Rift Valley and Ethiopia such as *Prunus africana* (Muchugi et al., 2006) and *Lobelia giberroa* (Mulugeta et al., 2007). Indigenous important species distribution in the tropics of sub-Saharan Africa generally, has been determined by prevailing soils, varying climates with demonstrated cycles of varying temperature and humidity as well as odographic changes such as elevated terrains (Trauth et al., 2005).

Despite these changes, mountains of the afro-alpine tropical East Africa have provided a relatively stable habitat where older species survive by altitudinal range shifts with new lineages generated (Fjeldså and Lovett, 1997). The flora occurs in isolated enclaves on the upper parts of the high East African and Ethiopian mountainous regions that are home to a number of geographical and ecologically isolated temperate “islands” (Mulugeta et al., 2007). The isolated and fragmented occurrence of tropical mountain vegetation promotes high speciation and an exceptionally high endemism at the species and genus level (Sklenár and Ramsay, 2001). Therefore, the isolation of these high mountains...
and their enclaves provides an ideal model system for studying speciation in connection with geographical isolation and adaptation to extreme environmental conditions (Mulugeta et al., 2007).

The high elevations of this region creates greater diversity for plants from both orographic effects of steep ecological gradients and elevated (Cochrane, 2011). This in return causes variations in wind intensity, precipitation, cloudiness, and insolation inducing extreme values of sharp differences both spatially and temporally (Young and Leon, 2007). Agricultural extension pressures caused by man and the erratic weather patterns caused by climate change has led to a decrease in plant population, with the remnant stands being over-harvested (Malla, 2008). The impact felt in the long run can only be reversed by a proper management of tree species. This is because without tree species management for adaptation, the current potential of tropical forests to both remove and store atmospheric carbon may diminish therefore, resulting in a positive feedback of carbon emissions (Gitz and Ciais, 2004). Tree conservation through agroforestry is a process of domestication that involves characterisation, selection and adoption of desirable germplasm.

Genetic differentiation among populations provides useful information for planning actions such as reinforcement of existing populations, reintroduction and or seed collection, with an aim of preserving the maximum amount of genetic diversity (Holsinger and Gottlieb, 1991). Exchange of genetic resource material may be important to counteract the effects of selection as well as
genetic drift in small groups of species with disrupted gene flow (Still et al., 2005).

This study aimed at identifying adaptive regional genetic makers in W. ugandensis of the Kenyan Rift Valley. The application of marker assisted selection in addition to Bulk Segregant Analysis (BSA) quantifies desired characteristics within the genome and can be used to reveal important RAPD diagnostic markers in a species (Welsh et al., 1990; Michelmore et al., 1991). Finally use of bioinformatics through computational comparative genomics tools by BLAST similarity searches, Multiple Sequence alignment (MSA) and Phylogenetic analysis gives an elaborate evolutionary history of genomic sequences (Rodriguez et al., 2007).

1.2 Statement of the problem and justification

There is genetic a disjunction in W. ugandensis species found in the East compared to those in the western populations in the Kenyan Rift Valley (Muchugi et al., 2008), possibly resulted by an allopatric speciation. Habitat loss through high genetic differentiation may result in loss of important functional traits that confer species with important adaptation advantages (Booy et al., 2005). This kind of regional adaptation may be important in climate change mitigation by conserving trees in the regions they can survive (Burdon and Wilcox, 2007). Some studies have shown that undisturbed ecosystems may develop and change species composition and other features over longer periods (Uhlig, 1988).
Genetic disjunction studies provide important insights on adaptive tree molecular markers showing differences linked to traits of economic importance (Upadhyaya et al., 2008). The whole concept creates improved and better managed plants germplasm of identifiable accessions for the future (Jamnadass et al., 2009). The routine is especially important for climate change mitigation challenges through a well-managed food security, health care and sequestration of carbon (Pye-Smith, 2010, Kremer and Neale, 2011). Efficient utilisation and conservation strategies are also important in understanding species biology and genetic diversity within its natural population (Fraxigen, 2005). Sustainable utilisation of plants will also improve the local population economy through bioprospects of the resources from medicinal plants extracts (Beattie and Ehrlich, 2004).

1.3 Research questions

i. Which genetic markers can be used to link *W. ugandensis* within the Kenyan Rift Valley?

ii. If present are they linked to important adaptation traits of the species?

iii. Can the markers be used in further development and regulation of important germplasm?

1.4 Research hypothesis

This research was guided by the following hypotheses:

i. There are RAPD diagnostic markers showing regional differentiation.

ii. These markers are linked to the species’ adaptive traits.
1.5 Objectives of the study

1.5.1 General objectives

To identify population specific RAPD markers linked to *W. ugandensis* regional differentiation across the Kenyan Rift Valley.

1.5.2 Specific objectives

i. To find regional differentiation across the *W. ugandensis* species range in Kenya using RAPD markers.

ii. To identify sequence(s) linked to regional diagnostic markers.

iii. To find sequence similarity for possible homology from the database/GenBank.
CHAPTER TWO

LITERATURE REVIEW

2.1 Taxonomy and distribution of species in the genus Warburgia

Warburgia ugandensis, Sprague belongs to the genus Warburgia, together with other members that include Pepper-back tree (Warburgia salutaris Chiov.), Mkaa (W. stuhlmannii Engl.) and W. elongata, Verde. They are distinguished by simple un-toothed, gland-dotted leaves; bisexual flowers, stamens united into a tube and with a fruit that is a berry with leathery glandular skin characteristics (Dale and Greenway, 1961). Warburgia ugandensis is an example of a heavily exploited tree species mainly for medicinal products and is indigenous to East and Central Africa, with a wide distribution across dry upland forests (including wooded riverines), and is restricted in its geographical distribution to Tanzania, Uganda and Kenya (Beentje, 1994). In Kenya, it is widely distributed across the Kenyan highlands with a varying ecological distribution (Beentje, 1994).

2.2 Uses of Warburgia ugandensis

This species is valued for timber and medicinal compounds harvested from stems, leaves, bark and roots. The medicinal compounds possess both antibacterial and antifungal properties and are widely used by the local communities living within its natural range to cure several ailments such as malaria (Olila et al., 2001; Njoroge and Bussmann, 2006; Nanyingi et al.,
2008; Were et al., 2010). The curative efficacies of extracts from the species have been scientifically tested both in vivo and in vitro trials after isolating major bioactive ingredients (Rabe and Van-Staden, 2000; Were et al., 2010). Phytochemical extracts discovered in *W. ugandensis* include alkaloid skimmianine and Sesquiterpene lactones muzigadial, polygodial (Olila et al., 2001; Wube et al., 2005). *W. ugandensis* bioprospected commercial products available in the market for different ailments are shown in Figure 1. While Figure 2 is *W. salutaris* products and Figure 3 is a *W. ugandensis* bioactive extracted compounds.

Figure 1 *Warburgia ugandensis* herbal products available in East Africa

Figure 2 *Warburgia salutaris* herbal products available from South Africa.
2.3 *Warburgia ugandensis* species natural’s habitat

*Warburgia ugandensis* has a restricted distribution in the afro-alpine environment of the tropical East Africa, floristic regions in the African *phytochoria* also known as the Afromontane *phytochoria* (Grimshaw, 2001). In Africa, Afro-alpine ecosystems occur as small and isolated patches in the mountains along the Great Rift Valley (Frankfurt Zoological Society, 2007). This region is characterised by a number of isolated high mountains like Ruwenzori, Aberdare’s, Mt. Elgon, Mt. Kilimanjaro, Mt. Kenya, and the Bale with altitudes of 4000 m or more above sea level and is categorised as either Eastern or Western regions of the Rift Valley (Finch *et al.*, 2009). The *phytochoria* is known for species richness, endemism and internal consistency with the environment (Linder *et al.*, 2005).
These traits are manifested in a small, but consistent number of tree species, whereby the richness decreases rapidly from East Africa to the Cape of South Africa (White, 1981). Therefore, the phytochoria shows that the montane forests and Afro-alpine belt are important genetic reservoirs for many endemic species yet also fragile ecosystems that serve an important protective and productive function. The vegetation displays a more distinct differentiation with altitude with elevational range of the vegetation belts becoming narrower giving it an abrupt transition in the species composition (Uhlig, 1988). Along the diversity there is a change in competition in the stands with gradual transition from interspecific to intraspecific competition (Uhlig, 1988). This kind of competition results in a conspicuous zonation which starts with a montane forest belt followed by an ericaceous belt, and finally an afro-alpine belt having its lower limit at 3000- 4000 m above sea level (Bussmann, 2006).

Forests are found in the lower most part of the mountain vegetation region therefore, making it look like a transition zone connecting the Afromontane and the lowland phytochoria (White, 1983). Display species richness in the Afro-alpine belt of the African Phytochoria in show in Figure 4. The climate of the afro-alpine belt is varied but of a very significant type, having night frosts all year round with intense sunshine during the daytime (Finch et al., 2009). Most vascular plants species within the afro-alpine region are adapted to some significant types of climates (Castañeda et al., 2009). This is through physiological modification adaptations such as slow growth rate due to decrease of carbon dioxide partial pressure (Hedberg, 1995).
Retreat of glaciers during the last ice age and climatic fluctuations of the Pleistocene have also contributed to the shaping of species' geographic population structure (Avise, 1994). Patterns of species' distribution and richness modelled in part by climatic fluctuation on a geologic time scale have played a major role in the evolution of tropical alpine plants (Smith and Young, 1987). Phylogeographical pattern studies of *L. giberroa* (a medicinal plant within the African *phytochoria*) showed large range by random long-distance colonization dispersal (Mulugeta *et al.*, 2007). This due to the nature of montane forest bridges as responsible for an allopatric speciation and not geographical processes such as earthquakes and volcanoes (Mulugeta *et al.*, 2007).
In the afro-alpine belt *W. ugandensis* mostly occur in lowland rainforest, upland dry evergreen forest and swamp forest soils, receiving on an average annual rainfall of between 100-2200 mm. The tree can also survive in relatively tougher environment as those plants that have undergone considerable vegetation change such as secondary bush land, grassland and swamp forest as remnants of the formerly larger population therefore becomes a priority to forest conservation (World Agroforestry Centre, 1992). Within Kenya, *W. ugandensis* populations are mostly found in the ravine with active or in active stream a long a down slope channel in the afro-alpine belt of the Rift Valley.

The major unique geographical feature between the western and central Kenyan highlands is the Great Rift Valley. The mountain enclaves receive much rain with a mean annual rainfall of 1000 mm but lesser in the lowland transitional vegetation (Friis and Sebsebe, 2001). The flora occurs in isolated enclaves on the upper parts of the high East African and Ethiopian mountains (Linder *et al.*, 2005). The soil types of the mountains are of volcanic origin and of unequal ages ranging from the Miocene to late Pleistocene (periods of varying warmer global climates and repeated glaciations of many million years in the past), with excessive internal drainage and good soil porosity (Castañeda *et al.*, 2009). The flora occurs in isolated enclaves on the upper parts of the high varied topology of East African Mountains which inhabits a number of geographically and ecologically isolated temperate “islands” (Hedberg, 1995).
Generally isolated habitats exhibit variable latitudinal and elevational climate which are comparable to oceanic islands due to the unique separation between the montane forests and savanna vegetation (Jonas and Geber, 1999; Clevering et al., 2001). With increasing altitude changes in species composition, diversity, structure of the stands, and morphology of the plants occur. These depend on the various limiting factors which characterise the ecological conditions (Uhlig, 1988). Despite the climatic differences, the surrounding lowlands provide the most consistent source of new species, showing both Afro-tropical and Palaearctic influences (Kingdon, 1989). This has been a case for many hundreds of years, and a predominant reason for a widespread loss of natural vegetation in the ecoregion brought about by both human activities mainly for agricultural reasons and animals’ dependence for grazing (Tilahun et al., 1996).

2.4 Genetic differentiation and evolutionary diversification

Genetic differentiation is a balance between the local ecological and population processes acting on the genetic variation in a habitat and regional processes that determine gene flow and migration that normally cause genetic and fitness differentiation of species (Saccheri and Hanski, 2006; Fussmann et al., 2007). Whereas high migration rates between populations will prevent or slow down differentiation, random drift, selection, founder effects, and bottlenecks can cause populations to differentiate (Charlesworth, 2009). Evolutionary diversification can occur as a consequence of geographical divergence; this process happens by isolating mechanisms, in part due to the
restriction of gene flow between populations (Still et al., 2005). Human encroachment to the belt for agricultural expansion results in a widespread destruction of wildlife and habitats. The fact that sufficient moisture availability and usual altitudinal lapse rate control most of the vegetation patterns, the plant communities in the region seem to depend on much more edaphic modification components (Bussmann, 2006). An example of Kenyan Rift Valley showing a degraded of the African afro-montane region is shown in Figure 5.

Figure 5 Natural habitat fragmentations in the Great Rift Valley

Among subpopulations of a widespread species, different ecological environments and independent evolution of populations through genetic drift may lead to divergence (Bomblies and Weigel, 2010). However, if gene flow
is not restricted by geographic barriers adaptation to new environments by plants will be minimal. Therefore, restricted gene flow through geographic barriers, coupled with differential selection pressures in species range provides an opportunity for allopartric speciation (Fickel et al., 2007). Allopatric speciation is a method by which new species arise due to geographic separation, a phenomenon whereby biological populations are physically isolated by an extrinsic barrier and evolve intrinsic genetic reproductive isolation, such that if the barrier ever vanishes, individuals of the populations can no longer interbreed (Vidya et al., 2005). Allopatric speciation can result from divergent evolution of geographically isolated populations (Whitaker, 2006).

2.5 Drivers of Genetic Evolution in Plant’s

Evolution is not always discrete with clearly defined boundaries that pinpoint the origin of a new species, nor is it a steady continuum. Possible causes of plant evolution may arise through mechanical agents such as human activity, animals’ destruction and geographical barriers to gene flow that may alter the species adaptation (Giraud et al., 2010). Evolution requires genetic variation which results from changes within a gene pool, the genetic make-up of a specific population (NCBI, 2011). A gene pool is the combination of all the allele’s alternative forms of a genetic locus-for all traits that population may exhibit (Hartl, 2007). These changes in a gene pool can result from mutation-variation within a particular gene-or from changes in gene frequency of the proportion of an allele in a given population (Young et al., 1996).
2.5.1 Forest destruction in the Afro-alpine: a consequence to speciation

Human anthropogenic destruction of forests and natural habitat fragmentation of a small size forest population cover has increased and has now become a significant threat to the maintenance of biodiversity in many terrestrial ecosystems (Young et al., 1996). The most striking changes in the ecosystems have been observed to be as a result of man's interference through clearance, fire and grazing. These processes affect large areas of the montane forest belt by degradation of the forests leading to loss of biodiversity and genetic diversity (Lienert, 2004). Natural habitat fragmentation affects the genetic partitioning among plants population with variation within the population either decreasing or increasing its differentiation levels (Pluess and Stoecklin, 2005). Differentiation in species fitness due to abiotic and biotic conditions like altitudinal gradients leads to major modifications in selective forces on plant life history traits resulting in their adaptation to local conditions (Cody and Overton, 1996).

Smaller populations are more vulnerable to demographic as well as environmental stochasticity leading to genetic erosion (Young et al., 1996; Menges, 2000). Small plant populations can alter particular alleles by chance over generations though geographical barriers also contributing to speciation of species through loss of gene flow (Still et al., 2005). Plants natural habitats may also be fragmented to a smaller or larger extent resulting in suitable habitat 'islands' for a species surrounded by non-suitable habitats (Bossuyt et al., 2003; Pither et al., 2003). The East African mountain forest bridges
suggest to have been important for species dispersal, though the theory seems not to hold for many forests across the West, Central Kenya and Ethiopian highlands (Mulugeta et al., 2007).

The distribution of many plant species across isolated African highland islands is thought to be a combination of long distance dispersal events and direct overland migrations facilitated by climate fluctuations (Muchugi et al., 2006; Mulugeta et al., 2007). In smaller isolated populations, species persistence is very much affected by enhanced random genetic drift, increased inbreeding, limited gene flow, and reduced mate availability (Young et al., 1996). *W. ugandensis* debarked by herbalists for medicinal purposes as shown in Figure 7, while Figure 8 shows a picture of *W. ugandensis’s* natural habitat undergoing destruction by animals as developed pathways for grazing grounds. These kinds of destructions and unutilised mode of consumption leads to a decrease in the species and other important species population creating a separate contact zone, which can however, lead to genetic drift (Futuyma, 1998).

Genetic drift is change in allele frequency from one generation to another within a population, due to the sampling of finite numbers of genes that is inevitable in all finite sized populations (Norman et al., 1993). Genetic drift is selectively neutral and both deleterious, therefore, advantageous alleles can get lost, while the level of heterozygosity decreases (Lacy, 1987). This kind of drift can therefore contribute in species divergence also known as “allopatry”.
When the population is small, the greater will be genetic drift, with the result that some alleles are lost and genetic diversity is reduced (Futuyma, 1998).

![Figure 7 Warburgia ugandensis debarked for medicinal purposes](image)

**Figure 7** *Warburgia ugandensis* debarked for medicinal purposes

![Figure 8 Typical habitat of *W. ugandensis* in Laikipia District, Kenya.](image)

**Figure 8** Typical habitat of *W. ugandensis* in Laikipia District, Kenya. (In set arrows show remnant trees habitat destruction by elephants)

**2.6 Geographical barriers are hindrance to gene flow**

Geographical divergence has been hypothesised as one means of evolutionary diversification in which the process occurs by isolating mechanisms, in part due to the restriction of gene flow between populations (Still *et al.*, 2005). Simple geographical isolation is not seen as the only widespread cause of
speciation and the interruption of gene flow between populations; nevertheless, both larger and smaller features of geography lead to isolation of populations, He et al., (2010) were able to show an example where the genetic structure of *Banksia hookeriana* species, not solely being dependent on the structure landscape, the habitat located on sand dune crests physically separated by uninhabitable hollows.

The theory does not also hold to Kadu et al., (2011) in their study on Phylogeography of the Afromontane *Prunus Africana* revealing a former migration corridor between East and West African highlands. Insights into phylogeny and implications of habitat diversification have been observed in *Farfugium* (Asteraceae) as a monophyletic group associated with a range of habitats in an archipelago in East Asia (Nomura et al., 2010). In a nutshell their suggestions were unbiased to isolation of islands and subsequent parallel adaptation events follows migration over Quaternary land-bridges along the distribution range. The events might have been introduced by uninformative DNA sequence variation coupled with highly divergent morphologies for adaptive diversification as rapid (Hurst and Warren, 2001: Heslop-Harrison, 2010).

### 2.7 Mutagens (Transposable Elements) a cause of plant speciation

Significant candidates involved in plant population speciation could be genetic mutagens such as Selfish Genetic Elements (SGE’s) in form Transposable Elements (TE’s) (Hurst and Warren, 2001). Transposable elements are
fragments of DNA that can insert into new chromosomal locations and often make duplicate copies of themselves in the process on large proportions of many eukaryotic genomes (Dvorak, 2009). They also are inform of repetitive genomic sequences capable of moving or translocating within the chromosomal genome making and making up a substantial proportion of themselves in the plant genomes in the form of Inserted Sequence (IS) groups (Bennetzen, 2000: Fedoroff, 2000: Wicker et al., 2007).

In the process of TE’s translocation they can rearrange genomes and alter individual gene structure and regulation through any of the activities to promote insertion, excision, chromosome breakage, and ectopic recombination (Slotkin and Martienssen, 2007: Kordis, 2009). Many genes may have been assembled or amplified through the action of transposable elements (Watson et al., 2004). They have an ability to amplify their copy number within the genome via transposition, thereby providing a selectable function that can make them selfish or parasitic DNAs (Doolittle and Sapienz, 1980: Orgel and Crick, 1980: Hun-Van et al., 2011). Transposable Elements can have a role in reproductive isolation by increasing the rate at which new chromosomal rearrangements (inversions and translocations) arise and fixation in different populations of different chromosomes (Pulletikurti et al., 2009). These elements can be subjected to weak selection for their efficacy varies as a function of demographic factors, which is a major driving force on their distribution in the genome (Lockton et al., 2008). Transposons have remarkably become diverse molecular tools for whole-genome and single-
gene studies in all organisms in the studies of gene regulation, expression, signal transduction and functions which can recognize and epigenetically silence them (Fontanillas et al., 2007; Lisch, 2009; Muotri et al., 2007).

Therefore, loss of genetic diversity can reduce plant performance and lower population viability in the short term, (Ellstrand and Elam, 1993). While in the long term, reduced genetic diversity may limit the potential for further adaptive evolution (Falconer and Mackay, 1996). Although population densities of *W. ugandensis* trees in the sampled populations were found to be quite low in the ecological survey, their genetic diversity was high within the populations as compared to other molecular studies on other afro-alpine indigenous tree species (Dawson and Powell, 1999; Muchugi et al., 2008). Preliminary studies on the species of the genus *Warburgia*, showed significant genetic differentiation between *W. ugandensis* population from the West of Rift Valley compared to the rest (Muchugi et al., 2008). Similar genetic variations were also observed in *Prunus africana* (Muchugi et al., 2006) and *Lobelia giberroa* showing a few clear diagnostic markers (Mulugeta et al., 2007). Findings from previous studies suggest that the markers could be linked to the species adaptation; a milestone to management and conservation of the biodiversity for sustainability and mitigation of global climate challenges (Muchugi et al., 2008). Genetic diversity and differentiation has fascinated evolutionary scientists of the world today for their economic prominence in healthcare and agriculture (Linhart and Grant, 1996).
2.8 Bulk segregant analysis (BSA)

Bulk segregant analysis is a technique used on investigation of diagnostic markers closely linked to specific important traits (Michelmore et al., 1991). It provides for a rapid, technically simple alternative method for identifying markers linked to specific genes (Michelmore et al., 1991). BSA compares two pooled DNA samples of individuals from a segregating population, which contrast for a trait. Population of similar species DNA is screened for a character of interest done by pooling together individuals in a batch to form a representative identical genotypic bulk. Therefore, each bulk contains individuals selected to be having identical genotypes for a particular genomic region but random genotypes at loci unlinked to the selected region. The two resultant bulks that differ genetically in the selected region are then tested for the presence or absence of molecular markers. Any marker polymorphism between the two bulks indicates the linkage between the marker and character of interest attributed by noticeable contrasting alleles contributing in positive and negative effects (Mwase et al., 2007; Eugenia et al., 2008).

Bulk segregant analysis has proved to be more effective for tagging qualitative and quantitative traits that can be classified into two discrete classes since it is dependent on accurate phenotyping and is therefore popular in Marker assisted selection by screening many loci (Michelmore et al., 1991). The resultant unique banding are then subjected to sequencing and then annotated to reveal meaningful information using bioinformatics tools such as BLAST for similar sequence homologous of genes studied in other organisms (Eugenia et al.,
2008). The overall information obtained is of economic importance, since the species adaptive traits, functional and biochemical synthesis is well defined and understood. These characterised traits can be in form of gene functionality as well as protein regulation that can be used as the plants environmental adaptation ability as well as production of the phytochemicals common in species (Olila et al., 2001; Cheng-Wei and Bor-Sen, 2010).

Therefore, the information can lead to understanding of the different phytochemical compounds found in each of the W. ugandensis provenance in the tree species through the possibility of ascertaining highly conserved genes across the species adaptation (Wang et al., 2010; Mwase et al., 2007). Genes that affect the expression of adaptive traits fundamentally influence an individual organism’s capacity to survive and reproduce. Therefore, the maintenance of adaptive trait variability in plant gene pools is emerging as a primary objective of biodiversity conservation (Jones and Monaco, 2009). This is made especially important by the environmental and climatic changes that impose substantial pressures on plant adaptive mechanisms especially in forest tree species. Adaptive trait variability has thus become a conservation and research priority in forest biodiversity management especially in subdivided species (Ennos et al., 1998; Lynch et al., 1999).

2.9 Marker assisted selection

Marker assisted selection (MAS) is an indirect selection technique where a trait of interest is selected not based on the trait itself but on a locus linked to it
(Welsh et al., 1990; Gupta et al., 1999; Barros et al., 2002). The development of DNA-based molecular markers has clearly allowed the direct comparison of the genetic material of two individual plants or more and to detect allele polymorphism. Often a unique allele (DNA sequence) with an occurrence in proximity to the gene or locus of interest can be identified. Some of these molecular markers are: hybridization based DNA markers such as restriction fragment length polymorphism (RFLP) and PCR based DNA markers such as random amplified polymorphic DNA (RAPD) and amplified fragment length polymorphisms (AFLP).

Molecular marker technologies have been used as a tool because of their high resolution power, compared to traditional breeding methods for differentiation and evaluation of germplasm accessions (FAO, 2007; Eenennaam and Weller, 2008). This is because DNA-based assays are robust, speedy, information may be obtained from little amounts of plant material at any stage of development and it is not affected by environmental conditions. Then the unique alleles are subjected to complete sequencing of genome to gather information for comparative and functional analysis (Mwase et al., 2007).

2.10 Random amplified polymorphic DNA

The development of polymerase chain reaction (PCR) for amplifying DNA has led to a revolution in the applicability of molecular methods and a range of new technologies developed to overcome many of the technical limitations related to the routine use of RFLPs (Kumar, 1999). The method therefore, is
popular for comparing the DNA of biological systems that have not had much attention in scientific research or systems where relatively few DNA sequences are compared (Aman, 1995). The technique is based on the amplification of unknown DNA sequences containing several primer binding sites closer to one another in an inverted orientation, using short oligonucleotide primers of arbitrary sequence (Erlich, 1992). The technique essentially scans a genome for small inverted repeats and amplifies intervening DNA sequences of variable length (Hadrys et al., 1992).

The main problem encountered in the use of RAPD analysis is limited repeatability and the amplification of repetitive DNA sequences (Devos and Gale, 1992; Penner et al., 1993). The stochastic nature of DNA amplification with random sequence primers, suggests a need of a highly maintained optimization and consistent reaction conditions for reproducible DNA amplification (Aman, 1995). Though RAPD-PCR analysis is associated with inherent problems of low reproducibility of results and extreme sensitivity to contamination and changes in reaction conditions (Guirao et al., 1997), carefully developed laboratory protocols overcomes this problem.

Reproducibility problems may also be overcome if care is taken to ensure consistent reaction conditions during amplification. There is also the problem of co-migrating bands being assumed homologous without this necessarily being the case. This is because gel electrophoresis separates DNA fragments quantitatively (according to size), but not qualitatively (according to base sequence) (Black, 1993; Ford-Lloyd and Painting, 1996). RAPD's has
therefore become valuable in trace, characterization, and the phylogeny of diverse plant and animal species. In this study, the value of RAPD-PCR technique in genetic differentiation of a species has been realized since DNA fragments that were generated from RAPD-PCR technique through BSA were highly polymorphic and distinguishable. RAPD analysis of the amplified products are separated on an agarose gels in the presence of ethidium bromide and visualised under ultraviolet light (Williams et al., 1990).

The analysis is by polymorphisms between individuals normally occurring as a presence or absence of a specific fragment (Williams et al., 1990). The technology is well suited to DNA fingerprinting (Elavazhagan et al., 2009) even though it does suffer from a certain lack of reproducibility due to mismatch annealing (Karp et al., 1996). The impact of these limitations can be reduced by scoring an appropriate number of RAPD fragments usually > 30 (Lynch and Milligan, 1994). RAPD-PCR technique is suitable in plant breeding and conservation especially where molecular markers are required to establish the genetic basis of important traits or in plant gene banking where molecular characterization of accessions can facilitate efficient management of germplasm collections (Passos et al., 2005). RAPD assay has been used for research as an efficient tool in identification of markers linked to agronomically important traits introgressed during the development of near isogenic lines (Gracia et al., 1996).
2.11 Cloning, transformation and sequencing of RAPD markers

DNA cloning refers to the process of creating multiple copies of an isolated DNA fragment or fragments by *in vitro* or *in vivo* methods. It is possible to clone entire gene fragments, random portions of DNA fragments or specific DNA sequences. Apart from DNA cloning, two other main cloning types are reproductive cloning, which is concerned with human and animal cloning, and therapeutic cloning that is concerned with embryonic cloning to harvest stem cells for research and potential medical treatment purposes. Other methods for accomplishing this include different procedures like PCR, agarose gel electrophoresis and DNA sonication (Sambrook *et al.*, 1989).

For the study cloning DNA inserts in plasmid vector mainly bacterial vectors is carried by using a shot gun strategy which is very straightforward method by means of electroporation (Sambrook *et al.*, 1989). Sequence determination of a genome is the first step and remains the only method toward understanding genome organization, gene structure, gene expression patterns, and other features scientific and commercial values (Skuse and Du, 2008). The approaches for shotgun-based sequencing of organism’s genomes are now well-established, and have resulted in the generation of numerous draft whole-genome sequence assemblies (Blakesley *et al.*, 2010). Shot gun sequencing strategy produces accurate consensus sequences from many overlapping a gel reading (readable frames) then reassembles the short fragments to reconstruct the target DNA sequence (Idury and Waterman, 1995). These reconstructed sequences are used for comparative genome analysis.
2.12 Bioinformatics tools for comparative sequence genomics

Bioinformatics tools provide for computational biological information of sequences derived from techniques of MAS such as RAPD-PCR finger prints. It is an integral part of recombinant technology that involves the incorporation of DNA from one organism into the genetic material of another organism for the rapid processing of enormous volumes of biological data Gene-bank (Benson et al., 2010). The method involves the use of biology, mathematics, statistics and computer science, including data mining, machine learning algorithms and visualization (Hogeweg, 2011). Bioinformatics is used in a number of applications, including sequence analysis, comparative and functional genomics, analysis of protein expression and analysis of organisms’ mutations (Brunner et al., 2004). The main aim of bioinformatics studies through a well-defined molecular fact sheet is the provision of useful biological data such as protein domains within sequences for comparative genetic studies.

Comparative genomics is a field in which use information provided by signatures in genomic sequences left by evolution (Proost et al., 2009). Exploitation of similarities and differences on the sequence level between species we are able to assign functional elements and their changes in evolution through computer simulations (Rasmussen and Kellies, 2010: Katzman et al., 2010). Common descent with modification, in evolution is reflected in genomic sequences (Sussman, 2009). Therefore, by comparing genomic sequences between species, modifications can also be traced (Mayer
et al., 2011). This is significance of comparisons data is to determine common ancestry of the elements studied. Elements sharing common ancestry are called homologous (Altschul, 1990), and is reflected in sequence similarity between genes and is an important concept in comparative genomics analysis.

There are many methods that have been developed to mine genomic data, and sequence comparison for analysis, but one of the best known and used methods is the Basic Local Alignment Search Tool (BLAST) (Lobo, 2008). This method uses heuristics and a generic model describing preferred substitutions in sequences and able to rapidly search large databases of sequence for sequence similarity (McGinnis and Madden, 2004). It has a faster computing power by rapidly searching large databases of sequences for similar sequences in a few seconds (Cameron et al., 2007). Though BLAST is a very powerful tool providing useful metrics of statistic value for sequence similarity like E-values, it does not homology for example that these sequences share common ancestry (Wong et al., 2010). Homology is determined best by manual sequence comparisons rather than using predetermined cut-off values for sequence similarity searches (Dong et al., 2007).

2.12.1 Protein domains and sequences patterns in evolution

The best sequences to align are those of protein-coding genes because proteins diverge more slowly than DNA sequences (Dame, 2005; Hsieh et al., 2006). This is because detection of positive selections in protein-coding genes usually
relies on codon models because they permit distinguishing between synonymous and non-synonymous substitutions, which are directly exposed to the action of selection (Anisimova and Liberles, 2007). These models can be inform of protein domains, since a domain is a continuous block of sites in a sequence alignment that can assigned to genes and protein-coding such as exons or non coding such as, introns (Ashurst and Collins, 2003). A protein domain (protein structural domain) is a discrete portion of a protein assumed to fold independently of the rest of the protein, and to possess evolutionary, structural, and functional significance (Ponting and Russell, 2002). It is for these properties, that protein domains have been used recently in the study of genomic and species evolution on wide a scale (Lin and 2000; Yang et al., 2005).

Abundant proteins are multi-domain linked by interactions between their domains, this gives them a complex but instrumental function, a concept that is also integral in creating new proteins by shuffling their domains (Bagowski et al., 2010). The importance of domains is that they cannot be divided into smaller units they represent fundamental building blocks that can be used to understand the evolution and function of proteins (NCBI, 2011). Genomic information by mining databases for homology of domains and not through the whole Protein is important in obtaining clues to functionality of genes because they are evolutionarily conserved (Thompson et al., 2011). For this reason protein sequence patterns frequently represent highly conserved parts of domains that can be considered as the true core signal of evolution,
structure, and function (Islam et al., 1995). These domains when mapped to species trees, the evolutionary history of domains and domain combinations is revealed, and the general evolutionary trend of domain and combination is analysed (Yang and Bourne, 2009).

2.13 Phylogenetic Analysis

Molecular phylogenetic also known as molecular systematics is the science of estimating the evolutionary past of among groups of species organisms and genes. It is as emerged as a major tool for inference of divergence within and among genes over time probably through mutations by providing overwhelming amount of information (Mount, 2004). The main objectives of phylogenetic analysis are to determine the closest relative of organism of interest, predict protein-protein interactions, determine structure of protein, estimate time of divergence, follow the changes in rapidly changing species and predict a function of a gene through BLAST (Chen et al., 2004). The analytical method is based on the comparison of DNA or protein sequences with the results presented in a phylogenetic tree by using a model such as maximum likelihood estimation (MLE) for gene divergence (Liao et al., 2010). MLE is a preferred parameter for estimation in statistics in particular in non-linear modelling with non-normal data (Myung, 2002). Maximum likelihood method accounts for all the information in a data set with a good power and accuracy to detect positive selection (Wong et al., 2004; Zhang et al., 2005). Closely related organisms generally have a high degree of
agreement in the molecular structure of its substances, while molecules of organisms distantly related show a pattern of dissimilarity (Liao et al., 2010).

Reliable phylogenies provide information on the sequence of evolutionary events that have generated in the present day diversity of genes and species and understanding the mechanisms and history of evolution of organisms (Tamura et al., 2007). Phylogenetic analysis though being the best diagnostics of the evolutionary events within organisms as both intuitive and extremely informative (Sankoff and Nadeau, 2003), it also has some limitations due to minimised amount of change, which is apparent when analyzing distantly related taxa (Swoford et al., 2001). Phylogenies even on studies outside the realm of sciences like linguistics and forensics apart from evolutionary processes of organisms through molecular sequences, therefore, gaining a lot acceptance (Howe et al., 2001: Gray and Atkinson, 2003: Salas et al., 2007).

2.13.1 Genetic distance

Genetic distance refers to the genetic divergence between species or between populations within a species done by considering a variety of parameters used to measure the variation (Shriver et al., 1995). Smaller genetic distances indicate a close genetic relationship whereas large genetic distances indicate a more distant genetic relationship (Nei, 1978). Genetic distance can be used to compare the genetic similarity and measures of divergence between different plant species and sub-species (Khan et al., 2011). Conserved sequences, such as mitochondrial DNA, are expected to accumulate mutations over time, and
assuming a constant rate of mutation providing a molecular clock for dating divergence (Bromham, 2002).

Molecular phylogeny uses the molecular clock assumption that quantitative similarity of genotype is a sufficient measure of the recency of genetic divergence (Aurahs et al., 2011). Using MEGA4 software package (Tamura et al., 2007) menu provides access to the test for examination of the substitution pattern homogeneity between sequences (Kumar and Gadagkar 2001). Genetic distance is therefore, a good indicator of how long ago two groups went their separate ways. When sequences of two nucleic acids or protein molecules are found in different organisms, they are likely to have been derived from a common ancestor (Scotland, 2010). To compare the significant sequences similarity based on genetic distances Multiple Sequence Analysis (MSA) is used based on maximum parsimony (MP) method that is useful for small comparisons of sequences in a number of nucleotide differences per site (Nei, 1991: Saitou, 1996). The main study targets comparison of both genetic distance and disparity index of the gene (Huang, 2008).

2.13.2. Disparity index

Disparity index is a simple statistic to measure and test the homogeneity of Substitution Patterns between molecular sequences (Kumar and Gadagkar, 2001). It is a common assumption in comparative sequence analysis that sequences evolve with the same pattern of nucleotide substitution (Galtier and Gouy, 1998: Tarrio et al., 2000). Disparity Index measures the observed
difference in substitution patterns for a pair of sequences (Kumar and Gadagkar, 2001). MEGA 4 software are computated and presents the Disparity Index per site, which is given by the total disparity index between two sequences divided by the number of positions compared, excluding gaps and missing data (Tamura et al., 2007). It is therefore, observed to be more powerful than a chi-square test of the equality of base frequencies between sequences (Kumar and Filipski, 2001).

It works by comparing the nucleotide or amino acid frequencies in given pair of sequences and using the number of observed differences between sequences. In identification of genes and species the two sequences can be defined as deviating from normal expectations of patterns change (Kumar and Gadagkar, 2001). A useful fact for elucidating the evolutionary mechanism responsible for observed differences. It is normally expected that sequences that have evolved with the same substitution process have similar nucleotide and or amino acid compositions. Therefore, differences in the substitution process among lineages can be detected by comparing the observed patterns of nucleotide frequencies in the extant sequences (Kumar and Gadagkar, 2001). This test can be used as a diagnostic tool to identify genes and lineages that have evolved with substantially different evolutionary processes as reflected in the observed patterns of change (Kumar and Gadagkar, 2001).
CHAPTER THREE

MATERIALS AND METHODS

3.1 Sample collection

Young leaf materials of about two years old were collected from 20 individuals at a minimum distance of about 100 m apart from nine natural populations of *W. ugandensis* from the Kenyan side of the Eastern Great Rift Valley, namely Kitale, Cherangani, Laikipia, Maasai Mara, Londiani, Karura, Taita, and two reference samples from Kibale and Lushoto from Uganda and Tanzania respectively. In order to study the genetic differentiation of the two regions sampled, locations were selected to cover a wide range of the natural distribution of the species. This was done with a focus on different geographical positions, elevations and altitudes based on latitude and longitude. The region was split into two as East and West of the Rift Valley region in Kenya as shown in Figure 9 and Table 1 indicates the GIS locations. The leaves were cleaned with 70% ethanol to sterilise their surface, then placed in snap-top plastic bags containing silica gel and left to dry at room temperature overnight to reduce the moisture content from the leaves, followed by storage at −20 °C prior for DNA extraction (Milligan, 1992).
Figure 9 Locations of *W. ugandensis* material used in the study

3.2 DNA extraction by Cetyltrimethyl Ammonium Bromide (CTAB)

Total genomic DNA was extracted from the dried leaf material from samples collected in locations shown in red points in Figure 9 and elevations in Table 1 using a modified cetyltrimethyl ammonium bromide (CTAB) method by Doyle and Doyle (1987). Approximately 1 cm² dry leaf material was placed in a sterile microfuge tube and the leaf made brittle by submerging the tube in liquid nitrogen. A small amount of Polyclar 100 SB® (BDH) was added and
the frozen leaf material ground to a fine powder using eppendorf® disposable grinders. Briefly, the procedure involved the following steps;

### Table 1 Location of *W. ugandensis* samples in the study

<table>
<thead>
<tr>
<th>Population</th>
<th>Country</th>
<th>Population in relation to the Rift Valley</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kitale</td>
<td>Kenya</td>
<td>West</td>
<td>01 00.631N</td>
<td>035 01.595E</td>
</tr>
<tr>
<td>Cherangani</td>
<td>Kenya</td>
<td>West</td>
<td>00 56.804N</td>
<td>035 17.850E</td>
</tr>
<tr>
<td>Kibale</td>
<td>Uganda</td>
<td>West</td>
<td>00 50.233N</td>
<td>030 43.388E</td>
</tr>
<tr>
<td>Laikipia</td>
<td>Kenya</td>
<td>Within</td>
<td>00 07.615N</td>
<td>036 25.717E</td>
</tr>
<tr>
<td>Maasai Mara</td>
<td>Kenya</td>
<td>Within</td>
<td>01 31.791S</td>
<td>035 19.074E</td>
</tr>
<tr>
<td>Londiani</td>
<td>Kenya</td>
<td>Within</td>
<td>00 07.091S</td>
<td>035 49.244E</td>
</tr>
<tr>
<td>Karura</td>
<td>Kenya</td>
<td>East</td>
<td>01 14.248S</td>
<td>036 49.546E</td>
</tr>
<tr>
<td>Taita</td>
<td>Kenya</td>
<td>East</td>
<td>03 30.030S</td>
<td>038 35.230E</td>
</tr>
<tr>
<td>Lushoto</td>
<td>Tanzania</td>
<td>East</td>
<td>04 37.553S</td>
<td>038 12.126E</td>
</tr>
</tbody>
</table>

Five hundred microlitres of extraction CTAB buffer (12.1 g Tris, 5.8 g Na₂EDTA, 81.9 g NaCl, and 20 g CTAB dissolved in a litre of water) containing 100 µl of mercaptoethanol per 100 ml was added to the finely ground leaves and incubated while shaking at 65 °C in a water bath with gentle shaking for 30 minutes. The suspension was mixed with 700 µl chloroform/Isoamylalcohol (24:1) and the mixture placed on a rotator Swirl (Fisher Company, U.S.A) for 15 minutes at room temperature, followed by centrifugation at 10,000 x g for 2 minutes. The supernatant was pipette out and placed in a fresh tube and mixed with 350 µl isopropanol, vortexed for 15 seconds and incubated at 4 °C for 5 minutes. The tube was centrifuged at 10,000 x g for 5 minutes at room temperature, the supernatant decanted and 450 µl of the aqueous layer transferred to a fresh microfuge tube. An equal
volume of ice-cold absolute isopropanol (450 μl) was added, mixed by vortexing for 15 seconds and incubated at -20 °C for 15 minutes.

The microfuge tube contents were centrifuged at 10,000 x g for 5 minutes at room temperature to pellet the DNA. The pellet was washed with 70 % ethanol, followed by 2 minutes centrifugation at 10,000 x g. The ethanol was removed and pellet air dried in a safety cabinet for 45 minutes. The pellet was resuspended in 100 μl TE buffer (10 mM Tris HCl, 1 mM EDTA pH 8.0) with 2 μl of 0.2 mg RNAase. Finally, the solution was vortexed for 15 sec and allowed to stand at RT for ~ 30 min in order to resuspend the DNA.

3.3 RNase digestion

The RNase digestion step was necessary to remove RNA from the isolated genomic DNA. RNase digestion was carried out as described in Sambrook et al., (2000) by adding 10 μl of 10 mg/ml RNase TO 50 μl of resuspended DNA pellet and then incubated at 37 °C for 30 minutes. Two volumes of chilled absolute ethanol was added to each sample and then centrifuged at 10,000 x g for 10 minutes to re-precipitate the DNA. The supernatant was poured off and the DNA pellets air-dried and resuspended in 50 μl of molecular grade water and stored at -20 °C.
3.4 DNA Quantification

The DNA concentration and purity was determined using ethidium bromide methods as described by Sambrook et al., (1989). Briefly, 1% (w/v) agarose gel was prepared by adding 300 ml of 1 x TBE buffer (TBE: 89 mM Tris-HCl [pH 8], 89 mM Boric acid and 2 mM EDTA) this was heated till the agarose dissolved then allowed to polymerise by cooling at room temperature.

3.4.1 Agarose gel electrophoresis and ethidium bromide fluorescence

Before the gel completely cooled 0.005 μg/ml ethidium bromide was added to the gel then swirled gently to evenly mix. Ten microliters of each DNA sample loaded as shown in then mixed on a clean sample-loading tray with 5 μl of 1 x gel loading buffer III (0.25% bromophenol blue, 0.25% xylene cyanol and 30% glycerol). Samples were loaded by placing 10 μl of dye mixed sample in each well with standards of 10 and 20 ng/μl method on comparing to known concentration of uncut lambda (λ) DNA marker (Promega, Madison-USA) to give sufficient quantity and quality for RAPD analysis. Gels were viewed under UV illumination (312 nm) and photographed using UVI gel documentation (UVI Gel documentation, USA). DNA concentration was estimated by comparing the band fluorescence with that of the standards.
3.5 Bulk Segregrant Analysis

Two bulks through Bulk Segregrant Analysis method were established representing the East and West groups in relation to populations in the Rift Valley. For each bulk, 12.5 ng DNA from individual in a region but of different populations of the two regions (East or West of Rift Valley) sampled were bulked by mixing based on the locality of the region and tagged numerically as seen in Table 2.

3.6 Random Amplified Polymorphic DNA Analysis

The polymerase chain reactions contained 10-mer arbitrary primers (200 nM) with approximately 2.5 ng of genomic DNA, 200 μM each of dNTPs (Boehringer Mannheim), 10 x PCR Buffer II (Perkin Elmer), 25 mM MgCl₂ (Perkin Elmer) and one unit of Taq DNA polymerase (Amplitaq Gold™, Applied Biosystems) in a final volume of 20 µl. Amplification was carried out in a MicroAmp® Optical 96 Well Reaction Plate (Applied Biosystems) using an MJ Research Inc. PTC-200 Thermal cycler programmed as follows: Initial denaturation step of 15 minutes at 94 °C, followed by 45 cycles of one minute at 94 °C, one minute at 30 °C, two minutes at 72 °C with a further final extension step of five minutes at 72 °C. As a negative control, sterile distilled water was added to the PCR master mix as a template in every PCR run. The reproducibility of RAPDs was tested by repeating a subset of PCR runs across subset of samples from all populations.
3.7 PCR products resolution, bands selection and gel recovery

PCR products were resolved by electrophoresis on 2% (w/v) agarose gel in 1 x TBE buffer by running the products at 5 v/cm for three and 0 half hours in a submarine electrophoresis tank. 20 μl of the PCR products mixed with 1 μl Bromophenol blue in xylene cyanol loading dye were loaded in a 2% (w/v) gel wells and run. The products were sized against a 100 bp ladder (Promega, Madison-USA) visualised under UV light after staining for 30 minutes in 0.5 μg/ml of ethidium bromide in 300 ml of TBE buffer, and then photographed using UVI gel documentation system (UVI Gel documentation, USA).

The unique fragments to the east and west regions were excised from the gel and the DNA purified for further analysis. Table 3 shows 10-mer RAPD oligonucleotide primers used in the different analysis. The table shows the order in which loading of the bulked DNA samples from different regions (West and East) in the Rift Valley was done to find the RAPD discriminate banding patterns for analysis. Table 2 show samples loading order in the gel, for the East and West region as used in the RAPD analysis. The table shows the oligonucleotide reference, sequence, Melting temperature (TM), and the length of the primer. The primers were synthesised from Operon Biotechnologies). Only fragments present exclusively in either East or West sample group were tested across the samples from the six populations (60 individuals) and five each from the reference populations (Kibale and Lushoto).
3.7 Purification of amplified products

Discriminate bands were excised from the agarose gels and purified using a QIAquick® gel extraction kit [QIAGen Inc., Valencia, CA] following manufacturer’s instructions, brief application as described in Appendix 6.

3.8 Adenylation

Since Taq DNA polymerase has a non-template-dependent terminal transferase activity that adds a single deoxyadenosine (A) to the 3’ ends of the PCR products (Innis et al., 1990), poly A tailing was performed to the RAPD amplicons for overhangs 3’ deoxthymidine (T) residues ligation of the pGEM®-T Easy (Kobs, 1997). 1-7 µl of purified PCR products to 1 µl Taq DNA polymerase in the presence of 10 µl 10 X Reaction Buffer with MgCl₂. 1 µl of dATP was added to a final concentration of 0.2 mM. Five units of Taq DNA polymerase was added to the mix of a volume of 1 µl, then the final volume was made up to 10 µl with addition of molecular grade water. This was then incubated at 70°C for 30 minutes for the formation of the polyA tails (Ausubel et al., 1987; Sambrook et al., 1989). The vector of choice was (Promega, Madison-USA)’s pGEM®-T Easy Vector kit for its convenience as a system for cloning of PCR products, since it is a linearised vector with a single 3’-terminal thymidine at both ends.
<table>
<thead>
<tr>
<th>Lane</th>
<th>East and West regions of the Rift Valley</th>
<th>Sampled populations</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>-----------------------------</td>
<td>100 bp ladder (Promega, Madison-USA)</td>
</tr>
<tr>
<td>1</td>
<td>West</td>
<td>Bulked samples; Kitale1 + Cherangani</td>
</tr>
<tr>
<td>2</td>
<td>West</td>
<td>Bulked samples; Kitale2 + Cherangani</td>
</tr>
<tr>
<td>3</td>
<td>West</td>
<td>Bulked samples; Kitale4+ Kibale</td>
</tr>
<tr>
<td>4</td>
<td>West</td>
<td>Individual sample as control; Cherangani</td>
</tr>
<tr>
<td>5</td>
<td>East and within</td>
<td>Individual sample as control; Londiani</td>
</tr>
<tr>
<td>6</td>
<td>East and within</td>
<td>Bulked samples; Karura + Laikipia</td>
</tr>
<tr>
<td>7</td>
<td>East and within</td>
<td>Bulked samples; Karura + Laikipia</td>
</tr>
<tr>
<td>8</td>
<td>East</td>
<td>Bulked samples; Karura + Taita</td>
</tr>
<tr>
<td>9</td>
<td>With East and out of the R.V</td>
<td>Bulked samples; Laikipia + Lushoto</td>
</tr>
</tbody>
</table>
Table 3 Five arbitrary primers used in the study

<table>
<thead>
<tr>
<th>ICRAF Ref</th>
<th>Sequence</th>
<th>TM (°C)</th>
<th>Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>IC15</td>
<td>5'-GGG ACGTCT C-3'</td>
<td>34.0</td>
<td>RAPD's</td>
</tr>
<tr>
<td>IC11</td>
<td>5'-GCA TGG AGC T-3'</td>
<td>32.0</td>
<td>RAPD's</td>
</tr>
<tr>
<td>IC28</td>
<td>5'-CCC GGC TTG T-3'</td>
<td>34.0</td>
<td>RAPD's</td>
</tr>
<tr>
<td>IC21</td>
<td>5'-GCT AGG GCG G-3'</td>
<td>36.0</td>
<td>RAPD's</td>
</tr>
<tr>
<td>IC55</td>
<td>5'-GTA GAC CCG T-3'</td>
<td>32.0</td>
<td>RAPD's</td>
</tr>
</tbody>
</table>

3.9 DH5α Competent cells preparation

DH5α chemical competent cells were prepared from stub cultures that had been stored at -20°C in the laboratory. Briefly this was performed by spreading 2 µl Escherichia coli cells on Luria-Bertani agar (LB) (1.0% tryptone, 0.5% Yeast Extract, 1.0% NaCl pH 7.0 and 100 µg/ml ampicillin as fully described in Appendix 4) and grown overnight at 37°C. A single colony of transformed cells was then inoculated in 50 ml of super optimal broth (SOB) (2% bacto tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl and 1/100th volume of sterile filtered 1 M MgSO₄ added before use), in a 500 ml flask and incubated overnight at 37°C with constant shaking ~115 x g in horizontal cooled orbital incubator shaker (Gallenkamp economy incubators, USA). From the overnight culture, 500 ml of cells were inoculated into fresh 500 ml of SOB broth media in a 2000 ml bottle and incubated at 18°C with
constant shaking in a Environ-vertical shaker 3597-1 Instruments (Labline
vertical shaker, Inc., Melrose Park, ILL.) until the culture attained an optical
density of 600. This was then chilled on ice for 10 minutes before spinning at
a speed of 2500 x g for 5 minutes at 4°C.

The pellet obtained was re-suspended gently in 80 ml TB (Tris-borate) then
placed on ice for 10 minutes. The cells were pelleted by spinning again as
described, but this time with an addition of 1.4 ml DMSO and 18.6 ml ice cold
TB. The pellet was then resuspended in 20 ml TB-DMSO (7% DMSO in TB),
then placed on ice for 10 minutes. Aliquots of cells were made on ice (in pre-
cooled tubes) and frozen immediately in liquid nitrogen. The competent cells
were then stored at -80°C in New Brunswick Scientific, Ultra low temperature
freezer made in England ready for use.

3.10 Cloning and transformation of PCR products

Purified PCR products were cloned in pGEM-T Easy vector (pGEM-T® Easy
vector Systems kit, (Promega, Madison-USA) Corp., Madison, WI, U.S.A;
map shown in Appendix 3) using a short gun strategy (Cantor and Smith,
1999). The map shows multiple cloning sites found in a pGEM-T Easy vector
and the vector size in terms of base pairs. Ligations were then performed in a
total volume of 10 ml. Purified PCR product 3 µl mixed with 1 µl of 50ng
pGEM-T Easy vector 5 µl of 2 X rapid ligation buffer, 1 µl of T4 DNA ligase
and 3 µl of ddH₂O then mixed by pipetting to and incubated at 4°C overnight
to produce maximum number of transformants.
Transformation of DH5α high efficiency competent cells was carried out using
the ligation mixture. In the transformation reaction, 50 μl aliquots of DH5α
cells were thawed on ice for 5 minutes and mixed with the 2 μl of each
ligation reaction and because of heat shocking and need for faster temperature
change the cloning mix were aliquoted in PCR tubes. It was then incubated on
ice for 20 minutes, heat-shocked at 42°C for 1 minute to enhance
transformation and immediately placed in ice for 1 minute. 950 μl SOC media
was added and the mix' incubated for 1 hour at 37°C under constant shaking in
a horizontal cooled orbital incubator shaker (Gallenkamp economy incubators,
USA) at 150 rpm.

The cells were then pelleted by centrifugation at approximately ≥10,000 x g
(~13,000 rpm) for 5 minutes at 4°C and the supernatant discarded table-top
microcentrifuge (eppendorf® centrifuge 55417 R, GmbH, Germany). The
supernatant layer of SOC media was decanted leaving approximately 100 μl,
mixed by tapping the tube and spread onto prepared LB (Luria-Bertani media)
medium (with agar) + ampicilin + X-gal + IPTG plates prepared for blue/
screening as stated in Appendix 6 and then incubated at 37°C overnight for
the transformed cells to grow. White colonies were picked into 50 μl PCR mix
for colony PCR screening. The major difficulty was to distinguish between
plasmids that have contained sequences of foreign sequences. However, this
was overcome by maintaining, reducing and adjusting the concentrations of
the foreign DNA and vector DNA during the ligation during the
recircularization process of the plasmid. These were carried out using T4 DNA
ligase which is the most versatile and commonly used ligase for DNA cloning because it is an ATP dependent enzyme. This enzyme covalently joins blunt or compatible cohesive ends, as well as nicks in double-stranded DNA. Adenine 5'-phosphoryl end group is required for ligation to a 3'-hydroxyl (Sambrook et al., 2000).

3.11 Analysis of recombinants by PCR

Recombinants were analysed by PCR using universal primers M13 show in Appendix 4 that flank the multiple cloning sites and allow direct sizing of the PCR products primers (Messing, 1983). PCR cocktail was prepared consisting 5 µl of 10 X PCR buffer, 0.5 µl of 50mM dNTPs, 1µM of 100 ng/µl M13 primers and 1unit of DNA Taq Polymerase. PCR mix was made to 20 µl with 0.6 µg/ml DNA added as template by touching the white colony with a pipette tip and aspirating in the mix and confirmed for the desired size by PCR. The products were electrophoresed on a 1% (w/v) gel and visualised under UVI gel documentation system and photographed. Few recombinant colonies were picked and cultured overnight in Luria-Bertani (LB) medium containing 100 µg/ml Ampicillin. The recombinants were that constituted of plasmid and the insert were then was sized against 1 kb Gene Ruler™ (Fermentas Life Sciences, UK).

3.12 Plasmid DNA preparation

Plasmid DNA was purified using QIAquick® gel extraction kit [QIAGen Inc., Valencia, CA] according to manufacturer’s protocol.
3.14 DNA sequencing

The purified plasmid DNA of a molecular size’s > 3015 bp from an a (1% (w/v) agarose analysis comparison) were subjected to DNA sequencing with ABI 3100 Genetic analyser (Applied Biosystems) using ABI Prism Big Dye Terminator V3.1 Ready Reaction Cycle Sequencing Kit (Applied Biosystems). Sequencing was performed from both directions using universal primers M13 primers and an additional primer Sp6 for confirmation (Butler and Chamberlain, 1982: Messing, 1983) in Appendix 3 and 4. The chromatograms from each of the sequenced fragments were scanned manually to confirm the sequences. The corrected sequence readings obtained were trimmed off vector sequences, contigs between forward and reverse reads built for each clone and assigned tags bearing species name, primer number and the geographical region from which the marker is found. An example was *WarburgiaIC28W* to stand for a *W. ugandensis* species sequence showing differentiation on the Western side of the Rift Valley.

3.15 Sequence similarity search by BLAST

Sequence data were used to make predictions of the functions of newly identified genes, estimate evolutionary distance in phylogeny reconstruction, determine protein functions, regulations and construct novel mutations (Goodstat and pointing, 2006: Dikow, 2011). These sequences were then compared against the Gen-bank non-redundant protein database National Center for Biotechnology Information (NCBI) using the BLASTX and
TBLASTX algorithms (Altschul et al., 1990). Vector DNA contamination on these sequences and the primer sequences end reads editing was done by NCBI’s VecScreen and ABI sequencher program, sequence (http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html) Scanner Software (http://bioinformatics.unc.edu/software/sequencher/seq_ABI_files_mac.htm) for fidelity conformation and clean sequences for Basic Local Alignment Sequence Tool BLAST similarity queries.

They were also used for sequence alignments to correct any mistakes, the corresponding sequences’ after trimming off the vector and oligonucleotide primer sequence reads and a few chromatogram’s that were observed in the analysis were in the following orientation (Reads from 5’ to 3’). The generated amino acid sequences were queried with the higher plants and other organisms’ database at the NCBI to determine E-values (the probability that the alignment is due to chance) at a statistical threshold of 0.0001 and considering meaningful biological information. Nucleotide sequences of cloned inserts were translated to protein using the translation tool at Swiss Bioinformatics Institute website – Expasy (http://us.expasy.org) (Bairoch, 1991). The amino acid sequences were then used to query various public protein databases to identify conserved domains.

3.16 Multiple Sequence Alignment

The first step in MSA was to translate DNA to amino acid sequences, then perform the alignment at the protein level, before back-translation to the DNA
alignment in a final step by CLUSTALW (http://www.ebi.ac.uk/clustalw/) website. The procedure avoids inserting gaps in the final DNA alignment that are not multiple of three and that would disrupt the reading frame. The amino acid sequences were also compared with other flowering plants (taxid: 3398), green plants (taxid: 33090), and land plants (taxid: 3193) via alignment with CLUSTALW (Altschul et al., 1997; Thompson et al., 2011). ClustalW is a general purpose In addition biological sequence alignments were edited by editor-BioEdit, (Tom hall, Ibis Therapeutics Carlsbad CA.) of (Hall, 1999) and Needleman-Wunsch global alignment was also used to find consensus alignment (NeedleN) (Needleman and Wunsch, 1970; Kruskal, 1983; Rice et al., 2000).
CHAPTER FOUR

RESULTS

4.1 Identification of RAPD markers showing East and West differentiations

DNA with good quality and quantity from the yield extracted was used for PCR amplifications to obtain RAPD polymorphic bands. A total of 20 RAPD primers sets were tested. Only five primers sets showed clear banding pattern with good resolution of East and West population differentiation in the Kenyan Rift Valley shown in Table 4 and Figures 10-14. The results in the study showed western populations differentiation by use of three RAPD primers IC21, IC55 and IC11 revealed at different bands levels.

Primer IC21 showed the smallest western population differentiated bands at 300 bps, while IC11 bands were shown at 320 and 360 bps as shown in Figures 10-13. Primer IC55 is the only primer that showed large populations differentiation bands though for the Western side at 800 and 1400 bps as shown in Figure 14. Eastern populations’ differentiation was only observed by using one RAPD primer IC28 with two different band levels (380 bps and 750 bps) as shown in Figure 10.
Table 4 Amplified band products gel excised for cloning

<table>
<thead>
<tr>
<th>RAPD Primer</th>
<th>Approximate band size in Bp excised</th>
</tr>
</thead>
<tbody>
<tr>
<td>IC11</td>
<td>360 bp</td>
</tr>
<tr>
<td>IC21</td>
<td>300 bp</td>
</tr>
<tr>
<td>IC28</td>
<td>750 bp</td>
</tr>
<tr>
<td>IC15</td>
<td>850 bp</td>
</tr>
<tr>
<td>IC55</td>
<td>800 bp</td>
</tr>
</tbody>
</table>

Figure 10 RAPD profiles of *W. ugandensis* using primer IC21.

(X and X₁ arrows show RAPD bands revealing population differences, while M is a molecular size maker of 100 bps.)
Figure 11 RAPD profiles of *W. ugandensis* using primer IC21 replicate gel.

(X and $X_1$ arrows show RAPD bands revealing population differences, while M is a molecular size maker of 100 bps.)
Figure 12 RAPD profiles of *W. ugandensis* using primer IC11.

(X and $X_1$ arrows show RAPD bands revealing population differences, while M is a molecular size maker of 100 bps.)
Figure 13 RAPD profiles of *W. ugandensis* using primer IC55.

(X and X1 arrows show RAPD bands revealing population differences, while M is a molecular size maker of 100 bps.)
Figure 14 RAPD profiles of *W. ugandensis* using primer IC28.

(X and X1 arrows show RAPD bands revealing population differences, while M is a molecular size maker of 100 bps.)
4.2 Cloning, plasmid purification and analysis of the transformants

A total of 20 transformants were analysed by PCR using M13 forward and reverse oligonucleotide primers and the resultant obtain sequences (For other sequence reads see Appendix 3) were used for NCBI BLAST database searches.

4.3 NCBI similarity searches results

Data mining on the origin of the sequence homology and predicting their protein translates, domains and any publications wrote on their behalf in Tables 5-10 by TBLASTX by the *W. ugandensis* sequences above. *W. ugandensis* homologous sequences found in the NCBI similarity searches by BLAST revealing species, sequences genbank accession numbers, the locus at which they were found, the expected length of the sequences, E-values and identities shown in Table 5. Predictions of protein functions, products, domains and related source publications found in the genbank and their material source the homologous species sequences are shown in Table 6-10. The amino acid sequences obtained from Expasy (http://us.expasy.org) and NCBI searches revealed some homology with protein functions in some regions of chromosome inform of transposable elements therefore, revealing a kind of divergence in the species.

4.3.2 Homologous sequence analysis data information

Homologous sequences from NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi)
BLAST searches supported the statistic of low E-values since they attained probability values of between 0.000-4e-33 and high percent identity in from the genbank as shown in Table 5. Results obtained in the MSA of CLUSTALW (http://www.ebi.ac.uk/clustalw/) amino acid substitution matrices shown in Figures 15-17 revealed significant alignments of WarburgiaIC55W, WarburgiaIC15W and WarburgiaIC28E. The score results shown in Tables 11-13 did not show high scores of pair-wise alignments with W. ugandensis as compared to its homolog’s.

The tables contains pair-wise sequence alignments of W.ugandensis sequences and the homologous sequences by length and scores from the matrix. The only high scores were obtained of long stretches of amino acid pair-wise alignments between W. ugandensis and its homolog’s are shown in Table 14. These alignments features were explained in the MSA by use of colour, amino acid symbols and their class.
### 4.3.1 Homologous sequence information in NCBI

#### Table 5 W. ugandensis BLASTX and TBLASTX homologous sequences

<table>
<thead>
<tr>
<th>Homologues</th>
<th>Gene Bank Accession No.</th>
<th>Locus</th>
<th>Expected size</th>
<th>E-value</th>
<th>Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WarburgiaIC55W (800 bp) homologous sequences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Artemisia annua</em> strain <em>Artemis</em></td>
<td>DQ115326.1</td>
<td>DQ115326</td>
<td>1166 bp</td>
<td>4e-33</td>
<td>100%</td>
</tr>
<tr>
<td><em>Climacium americanum</em></td>
<td>DQ648742.1</td>
<td>DQ648742</td>
<td>1994 bp</td>
<td>1e-08</td>
<td>88%</td>
</tr>
<tr>
<td><em>Siparuna decipiens</em></td>
<td>DQ008779</td>
<td>DQ008779</td>
<td>2316 bp</td>
<td>4e-08</td>
<td>78%</td>
</tr>
<tr>
<td><strong>WarburgiaIC11W (850 bp) homologous sequences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Physcomitrella patens</em> subsp. <em>Patens</em></td>
<td>XP_001758096</td>
<td>XP_001758096</td>
<td>552 aa</td>
<td>0.057</td>
<td>59%</td>
</tr>
<tr>
<td><em>Ricinus communis</em></td>
<td>XP_002515023</td>
<td>XP_002515023</td>
<td>1987 aa</td>
<td>0.034</td>
<td>19%</td>
</tr>
<tr>
<td><em>Populus trichocarpa</em></td>
<td>XP_002315807</td>
<td>XP_002315807</td>
<td>894 aa</td>
<td>0.098</td>
<td>25%</td>
</tr>
<tr>
<td><strong>W. ugandensisIC28E (750 bp) homologous sequences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Arthrobacter sp.</em> FB24,</td>
<td>YP-829959</td>
<td>YP_829959</td>
<td>272 aa</td>
<td>2e-49</td>
<td>95%</td>
</tr>
<tr>
<td><em>Streptomyces sp.</em> C</td>
<td>ZP_05511487.1</td>
<td>ZP_05511487</td>
<td>256 aa</td>
<td>3e-48</td>
<td>92%</td>
</tr>
<tr>
<td><em>Rhodococcus erythropolis</em></td>
<td>NP_898661</td>
<td>NP_898661</td>
<td>268 aa</td>
<td>2e-49</td>
<td>90%</td>
</tr>
<tr>
<td>Homologues species</td>
<td>Gene Bank Accession No.</td>
<td>Locus</td>
<td>Expected length</td>
<td>E-value</td>
<td>Identity</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>-------------------------</td>
<td>--------</td>
<td>-----------------</td>
<td>---------</td>
<td>----------</td>
</tr>
<tr>
<td>Highly significant sequences WarburgiaIC55E, WarburgiaIC11W and WarburgiaIC28W homologs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Artemisia annua strain Artemis</td>
<td>EZ153264.1</td>
<td>EZ153264</td>
<td>402 aa</td>
<td>1e-19</td>
<td>100%</td>
</tr>
<tr>
<td>Collimonas fungivorans</td>
<td>AY593480.1</td>
<td>AY593480</td>
<td>11502 aa</td>
<td>5e-16</td>
<td>90%</td>
</tr>
<tr>
<td>Candidatus Glomeribacter gigasporarum</td>
<td>AJ561042.3</td>
<td>AJ561042</td>
<td>946 aa</td>
<td>2e-14</td>
<td>86%</td>
</tr>
</tbody>
</table>
# Table 6 WarburgiaIC55E TBLASTX homologous sequence features

<table>
<thead>
<tr>
<th>Species</th>
<th>Sources</th>
<th>Protein</th>
<th>Region/product</th>
<th>Domain/note</th>
<th>Published Journal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>C. americanum</em></td>
<td>genomic</td>
<td>***</td>
<td>Large subunit rRNA</td>
<td></td>
<td>The deepest divergences in land plants inferred from phylogenomic evidence (Qui <em>et al.</em>, 2006)</td>
</tr>
<tr>
<td></td>
<td>DNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>S. decipiens</em></td>
<td>genomic</td>
<td>***</td>
<td>Large subunit rRNA</td>
<td></td>
<td>Phylogentic analyses of basal Angiosperms based on nine plastid (Qui <em>et al.</em>, 2006)</td>
</tr>
<tr>
<td></td>
<td>DNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism</td>
<td>Source</td>
<td>Protein</td>
<td>Region/product</td>
<td>Domain/Note</td>
<td>Journal published Title</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-----------------</td>
<td>-----------------------------</td>
<td>------------------</td>
<td>----------------------</td>
<td>----------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><em>P. patens</em> subsp. <em>patens</em></td>
<td>Grandsen ecotype</td>
<td>Hypothetical protein</td>
<td>Myosin HC-like pfam07058</td>
<td>Myosin heavy chain like</td>
<td>The <em>physcomitrella</em> genome reveals evolutionary insights into the conquest of land by plants (Rensing <em>et al.</em>, 2008).</td>
</tr>
</tbody>
</table>
Table 8 WarburgiaIC28W TBLASTX homologous sequence features in NCBI

<table>
<thead>
<tr>
<th>Organism</th>
<th>Source</th>
<th>Protein</th>
<th>Region/product</th>
<th>Domain(note)</th>
<th>Journal published Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arthrobacter sp. FB24</td>
<td>genome</td>
<td>T1 B ATP binding domain-containing</td>
<td>ATP binding domain</td>
<td>transposase provisional PRK06526</td>
<td>High-level chromate resistance in Arthrobacter sp. strain FB24 requires previously uncharacterised accessory genes (Henne et al., 2009)</td>
</tr>
<tr>
<td>R. erythropolis</td>
<td>genome</td>
<td>IS sequence</td>
<td>ATP binding domain</td>
<td>transposase PRK06526</td>
<td>DNA sequence and genetic organization of 210Kb linear plasmid of R. Erythropolis (Stecker et al., 2003).</td>
</tr>
</tbody>
</table>
**Table 9 WarburgiaIC55E, WarburgiaIC15W and WarburgiaIC28W TBLASTX homologous sequence features in NCBI**

<table>
<thead>
<tr>
<th>Organism</th>
<th>Source</th>
<th>Protein</th>
<th>Region/product</th>
<th>Domain/note</th>
<th>Journal published Title</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Artemisia annua</em> strain Artemis</td>
<td>genome</td>
<td>*****</td>
<td>Contig12670, mRNA sequence</td>
<td>*****</td>
<td>Unpublished</td>
</tr>
<tr>
<td><em>Collimonas fungivorans</em></td>
<td>genome</td>
<td>*****</td>
<td>fosmid CFUFOS26</td>
<td>ATP binding domain</td>
<td>Phylogeny-function analysis of DNA libraries: screened for expression of rRNA genes by large-insert library fluorescent in situ hybridization (Leveau <em>et al.</em>, 2004).</td>
</tr>
<tr>
<td><em>Candidatus Glomeribacter gigasporarum</em></td>
<td>genome</td>
<td>*****</td>
<td>23S rRNA gene</td>
<td>*****</td>
<td>Vertical transmission of endobacteria in the arbuscular mycorrhizal fungus <em>Gigaspora margarita</em> through generation of vegetative spores (Bianciotto <em>et al.</em>, 2004).</td>
</tr>
</tbody>
</table>
Table 10 Cellular features and biological functions observed in NCBI database searches of significant homolog's

<table>
<thead>
<tr>
<th>Features</th>
<th>Name</th>
<th>Region or domain</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>RRM</td>
<td>RNA recognition motif protein</td>
<td>protein</td>
<td>Regulate post-transcriptional gene expression (Maris et al., 2005).</td>
</tr>
<tr>
<td>SMC protein</td>
<td>Structural Maintainace Protein Large family of ATPases</td>
<td>Protein</td>
<td>Participating in many chromosomal organization and dynamics (Hirano, 2002).</td>
</tr>
<tr>
<td></td>
<td>with diverse cellular Activities</td>
<td></td>
<td>Range of processes, including protein degradation, membrane fusion, microtubule severing, peroxisome biogenesis, signal transduction and the regulation of gene expression (Koonin et al., 2004).</td>
</tr>
<tr>
<td>AAA+ ATPases</td>
<td>Associated with diverse cellular Activities</td>
<td>Protein</td>
<td>Multifunctionality of plant ABC transporters: more than just detoxifiers (Martinoia et al., 2002).</td>
</tr>
<tr>
<td>ABC</td>
<td>ATP-Binding Cassette</td>
<td>Protein</td>
<td>Multifunctionality of plant ABC transporters: more than just detoxifiers (Martinoia et al., 2002).</td>
</tr>
<tr>
<td>1st B ATP</td>
<td>1st B ATP binding protein</td>
<td>Protein</td>
<td>Multifunctionality of plant ABC transporters: more than just detoxifiers (Martinoia et al., 2002).</td>
</tr>
<tr>
<td>binding</td>
<td>protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IS</td>
<td>Insertion sequences</td>
<td>small DNA</td>
<td>Insertion sequences (Mahillon and Chandler, 1998).</td>
</tr>
<tr>
<td></td>
<td>segments</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 15 Alignments of *Warburgia55E* sequence against homologues.

(The symbols represent: (-) means many gaps, ("."+) very few conserved substitutions, (".".) semi-conserved substitutions and ("*".*) very few identical amino acids.)

Table 11 Multiple Sequence Alignment of *WarburgiaIC55W* and homologous sequences

<table>
<thead>
<tr>
<th>SeqA Name</th>
<th>Len(aa)</th>
<th>SeqB Name</th>
<th>Len(aa)</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>WarburgiaIC55E</em></td>
<td>264</td>
<td>2</td>
<td>388</td>
</tr>
<tr>
<td>1</td>
<td><em>WarburgiaIC55E</em></td>
<td>264</td>
<td>3</td>
<td>664</td>
</tr>
<tr>
<td>1</td>
<td><em>WarburgiaIC55E</em></td>
<td>264</td>
<td>4</td>
<td>767</td>
</tr>
<tr>
<td>2</td>
<td>Gossypium</td>
<td>388</td>
<td>3</td>
<td>664</td>
</tr>
<tr>
<td>2</td>
<td>Gossypium</td>
<td>388</td>
<td>4</td>
<td>767</td>
</tr>
<tr>
<td>2</td>
<td>Climacium</td>
<td>664</td>
<td>4</td>
<td>767</td>
</tr>
<tr>
<td>Genus</td>
<td>Sequence 1</td>
<td>Sequence 2</td>
<td>Sequence 3</td>
<td>Sequence 4</td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td><em>Physcomitrella</em></td>
<td>QVDDLNQEMKQIEINQRIEAEKNDKMR---R---QKIAEIEKLSATVAZLEEA</td>
<td>LQEECINDQEMAYLADKBEAIVESQESAESKIANAEKQEEVELEEST</td>
<td>SWEECNEKHDYRLYARASAYQERHRE---EKSEARIDHM</td>
<td>KBERQRKIERAEEYEBGLLREYREK---QRQYKRDKERKRRE</td>
</tr>
<tr>
<td><em>Ricinus</em></td>
<td>INVLEKQEMEDEEHERMIRESQEOQLELQCLQCLRTVEPIVDGNTISVQEDPI</td>
<td>TSDEKYTELRSQCRYRQAVRQE---EKSEARIDHM</td>
<td>ILHDEEEDDDSRKPRRAELEDRREKREKEKEDDLAESEEIEEAE---</td>
<td></td>
</tr>
<tr>
<td><em>Warburgia</em></td>
<td>VLAGGAANAAADYQRQAQELLEGKKTLERELARAKITANVRVAUVVAN---</td>
<td>EAMVHEVRNIAASAAATLDKEKSRVETGSSPCRCLNSVQQLMKREKDFLSLAAG</td>
<td>LKVLEEKVAPVTTTASGSQGTADKRLLVqAER---TAKNEQSLIERQFMR</td>
<td></td>
</tr>
<tr>
<td><em>Populus</em></td>
<td>ILHDEEEDDDSRKPRRAELEDRREKREKEKEDDLAESEEIEEAE---</td>
<td>ILHDEEEDDDSRKPRRAELEDRREKREKEKEDDLAESEEIEEAE---</td>
<td>ILHDEEEDDDSRKPRRAELEDRREKREKEKEDDLAESEEIEEAE---</td>
<td></td>
</tr>
<tr>
<td><em>Physcomitrella</em></td>
<td>WKDANQKEDYRPPPQRQQQELLEGKKTLERELARAKITANVRVAUVVAN---TAKNEQSLIERQFMR</td>
<td>SRRQRSAVTQLRSMASQTLRATNGRTSKEPFGGIDGLDAGTRELDPS---</td>
<td>SRRQRSAVTQLRSMASQTLRATNGRTSKEPFGGIDGLDAGTRELDPS---</td>
<td></td>
</tr>
<tr>
<td><em>Ricinus</em></td>
<td>SRRQRSAVTQLRSMASQTLRATNGRTSKEPFGGIDGLDAGTRELDPS---</td>
<td>SRRQRSAVTQLRSMASQTLRATNGRTSKEPFGGIDGLDAGTRELDPS---</td>
<td>SRRQRSAVTQLRSMASQTLRATNGRTSKEPFGGIDGLDAGTRELDPS---</td>
<td></td>
</tr>
<tr>
<td><em>Warburgia</em></td>
<td>LDEEEQEVATASGVDREAPFFKASIATAT---DKSHCRAEQ----</td>
<td>LDEEEQEVATASGVDREAPFFKASIATAT---DKSHCRAEQ----</td>
<td>LDEEEQEVATASGVDREAPFFKASIATAT---DKSHCRAEQ----</td>
<td></td>
</tr>
<tr>
<td><em>Populus</em></td>
<td>LDEEEQEVATASGVDREAPFFKASIATAT---DKSHCRAEQ----</td>
<td>LDEEEQEVATASGVDREAPFFKASIATAT---DKSHCRAEQ----</td>
<td>LDEEEQEVATASGVDREAPFFKASIATAT---DKSHCRAEQ----</td>
<td></td>
</tr>
<tr>
<td><em>Physcomitrella</em></td>
<td>LDEEEQEVATASGVDREAPFFKASIATAT---DKSHCRAEQ----</td>
<td>LDEEEQEVATASGVDREAPFFKASIATAT---DKSHCRAEQ----</td>
<td>LDEEEQEVATASGVDREAPFFKASIATAT---DKSHCRAEQ----</td>
<td></td>
</tr>
<tr>
<td><em>Ricinus</em></td>
<td>LDEEEQEVATASGVDREAPFFKASIATAT---DKSHCRAEQ----</td>
<td>LDEEEQEVATASGVDREAPFFKASIATAT---DKSHCRAEQ----</td>
<td>LDEEEQEVATASGVDREAPFFKASIATAT---DKSHCRAEQ----</td>
<td></td>
</tr>
<tr>
<td><em>Warburgia</em></td>
<td>LDEEEQEVATASGVDREAPFFKASIATAT---DKSHCRAEQ----</td>
<td>LDEEEQEVATASGVDREAPFFKASIATAT---DKSHCRAEQ----</td>
<td>LDEEEQEVATASGVDREAPFFKASIATAT---DKSHCRAEQ----</td>
<td></td>
</tr>
<tr>
<td><em>Populus</em></td>
<td>LDEEEQEVATASGVDREAPFFKASIATAT---DKSHCRAEQ----</td>
<td>LDEEEQEVATASGVDREAPFFKASIATAT---DKSHCRAEQ----</td>
<td>LDEEEQEVATASGVDREAPFFKASIATAT---DKSHCRAEQ----</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 16** Alignments of *Warburgia* sequence against homologues.

(In the Figure the symbols represent: (-) means many gaps, (":") very few conserved substitutions, (".".) semi-conserved substitutions and ("*".) very few identical amino acids.)
Table 12 Multiple Sequence Alignment of *WarburgiaIC15W* and homologous sequences

<table>
<thead>
<tr>
<th>SeqA Name</th>
<th>Len(aa)</th>
<th>SeqB Name</th>
<th>Len(aa)</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>WarburgiaIC15E</td>
<td>281</td>
<td>2</td>
<td>Physcomitrella</td>
</tr>
<tr>
<td>1</td>
<td>WarburgiaIC15E</td>
<td>281</td>
<td>3</td>
<td>Ricinus</td>
</tr>
<tr>
<td>1</td>
<td>WarburgiaIC15E</td>
<td>281</td>
<td>4</td>
<td>Populus</td>
</tr>
<tr>
<td>2</td>
<td>Physcomitrella</td>
<td>552</td>
<td>3</td>
<td>Ricinus</td>
</tr>
<tr>
<td>2</td>
<td>Physcomitrella</td>
<td>552</td>
<td>4</td>
<td>Populus</td>
</tr>
<tr>
<td>3</td>
<td>Ricinus</td>
<td>2141</td>
<td>4</td>
<td>Populus</td>
</tr>
</tbody>
</table>

Figure 17 Alignments of *WarburgiaIC28E* sequence significant homologues.

(The symbols represents: (-) means many gaps, (":") very few conserved substitutions, (".")) semi-conserved substitutions and ("*")) very few identical amino acids.)
Table 13 Multiple Sequence Alignment of *Warburgia IC28E* and homologous sequences

<table>
<thead>
<tr>
<th>SeqA Name</th>
<th>Len(aa)</th>
<th>SeqB Name</th>
<th>Len(aa)</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 <em>Warburgia IC28W</em></td>
<td>244</td>
<td>2 <em>Streptomyces</em></td>
<td>256</td>
<td>3</td>
</tr>
<tr>
<td>1 <em>Warburgia IC28W</em></td>
<td>244</td>
<td>3 <em>Arthrobacter</em></td>
<td>272</td>
<td>4</td>
</tr>
<tr>
<td>1 <em>Warburgia IC28W</em></td>
<td>244</td>
<td>4 <em>Rhodococcus</em></td>
<td>268</td>
<td>7</td>
</tr>
<tr>
<td>2 <em>Streptomyces</em></td>
<td>256</td>
<td>3 <em>Arthrobacter</em></td>
<td>272</td>
<td>76</td>
</tr>
<tr>
<td>2 <em>Streptomyces</em></td>
<td>256</td>
<td>4 <em>Rhodococcus</em></td>
<td>268</td>
<td>76</td>
</tr>
<tr>
<td>3 <em>Arthrobacter</em></td>
<td>272</td>
<td>4 <em>Rhodococcus</em></td>
<td>268</td>
<td>80</td>
</tr>
</tbody>
</table>

Figure 18 Alignments of the 3 *W. ugandensis* and significant retrieved homologous sequences in the study.

Table 14 Multiple Sequence Alignment of *Warburgia ugandensis* and 3 significant homologous sequences

<table>
<thead>
<tr>
<th>SeqA Name</th>
<th>Len(aa)</th>
<th>SeqB Name</th>
<th>Len(aa)</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 <em>Warburgia</em></td>
<td>264</td>
<td>2 <em>Artemisin</em></td>
<td>402</td>
<td>43</td>
</tr>
<tr>
<td>1 <em>Warburgia</em></td>
<td>264</td>
<td>3 <em>collimonas</em></td>
<td>811</td>
<td>35</td>
</tr>
<tr>
<td>1 <em>Warburgia</em></td>
<td>264</td>
<td>4 <em>Candidatus</em></td>
<td>945</td>
<td>47</td>
</tr>
<tr>
<td>2 <em>Artemisin</em></td>
<td>402</td>
<td>3 <em>collimonas</em></td>
<td>811</td>
<td>51</td>
</tr>
<tr>
<td>2 <em>Artemisin</em></td>
<td>402</td>
<td>4 <em>Candidatus</em></td>
<td>945</td>
<td>56</td>
</tr>
<tr>
<td>3 <em>collimonas</em></td>
<td>811</td>
<td>4 <em>Candidatus</em></td>
<td>945</td>
<td>61</td>
</tr>
</tbody>
</table>
Table 15 Colours used to represent different amino acids in the MSA

<table>
<thead>
<tr>
<th>Colour</th>
<th>Amino acids</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>RED</td>
<td>AVFPMILW</td>
<td>Small (small+ hydrophobic</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(incl.aromatic -Y))</td>
</tr>
<tr>
<td>BLUE</td>
<td>DE</td>
<td>Acidic</td>
</tr>
<tr>
<td>MAGENTA</td>
<td>RK</td>
<td>Basic</td>
</tr>
<tr>
<td>GREEN</td>
<td>STYHCNGQ</td>
<td>Hydroxyl + Amine + Basic – Q</td>
</tr>
<tr>
<td>Others</td>
<td>Gray</td>
<td></td>
</tr>
</tbody>
</table>

4.4 Phylogenetic analysis of *W. ugandensis* and homologous sequences

Phylogenetic analysis results revealed two clusters with *W. ugandensis* sequences and ten orthologs as shown in Figures 19-22. Table 16-19 reveals the relations of the homologous sequences to *W. ugandensis* in terms of genetic distances and sequence homogeneity through disparity indices as shown in Tables 17-20. The phylogenetic tree showed clusters and genetic distances from the node, while the corresponding tables showed genetic distances of the sequences and their disparity indices. The results were of significant higher genetic pair-wise distance values of between 0.7 – 0.9 when a consensus of all the 3 most significant sequences of *W. ugandensis* (*WarburgiaIC55E, WarburgiaIC15W* and *WarburgiaIC28W*) that were earlier mined by NCBI and translated then aligned to its homologous sequences shown in Figure 22 and Table 20. While the homolog’s disparity indices were of quit low levels of substitutions (0) that predicts bigger conservations within the protein domains along the alignments.
Figure 19 Phylogram of 3 homologous sequences of WarburgiaIC55/W.

Table 17 Pair-wise distances of WarburgiaIC55/W and disparity index

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>WarburgiaIC55/W</td>
<td></td>
<td></td>
<td>WarburgiaIC55/W</td>
<td>(0.2)</td>
<td>(0.7)</td>
<td>(0.4)</td>
</tr>
<tr>
<td>2</td>
<td>Climacium</td>
<td>1.1</td>
<td></td>
<td>Climacium</td>
<td></td>
<td>0.1</td>
<td>(0)</td>
</tr>
<tr>
<td>3</td>
<td>Siparuna</td>
<td>1.6</td>
<td>1.7</td>
<td>Siparuna</td>
<td></td>
<td></td>
<td>(0)</td>
</tr>
<tr>
<td>4</td>
<td>Gossypium</td>
<td>2.2</td>
<td>2.0</td>
<td>Gossypium</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 20 Phylogram of 3 homologous sequences of *WarburgiaC15/W*. 
Table 18 Pair-wise distances for *Warburgia IC15E* and disparity index

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>Physcomitrella</em></td>
<td></td>
<td></td>
<td><em>Physcomitrella</em></td>
<td>(0.7)</td>
<td>(0.4)</td>
<td>(0.4)</td>
</tr>
<tr>
<td>2</td>
<td><em>Ricinus</em></td>
<td>1.5</td>
<td></td>
<td><em>Ricinus</em></td>
<td>(2.0)</td>
<td>(1.9)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td><em>Warburgia IC15E</em></td>
<td>2.1</td>
<td>2.0</td>
<td><em>Warburgia IC15E</em></td>
<td></td>
<td>(0.1)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td><em>Populus</em></td>
<td>2.2</td>
<td>2.4</td>
<td><em>Populus</em></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Figure 21 Phylogram of 3 homologous sequences of *Warburgia C28W*.
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 WarburgiaIC28/W</td>
<td></td>
<td></td>
<td></td>
<td>WarburgiaIC28/W</td>
<td>0</td>
<td>0.1</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>2 Rhodococcus</td>
<td>0.2</td>
<td></td>
<td></td>
<td>Rhodococcus</td>
<td></td>
<td>0</td>
<td>2.9</td>
<td></td>
</tr>
<tr>
<td>3 Streptomyces</td>
<td>0.2</td>
<td>0.3</td>
<td></td>
<td>Streptomyces</td>
<td></td>
<td></td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td>4 Arthrobacter</td>
<td>2.3</td>
<td>0.2</td>
<td>2.3</td>
<td>Arthrobacter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 22 Phylogram of the homologous sequences to all *W. ugandensis* significant sequences.
Table 20 Pair-wise distances for all sequences and their disparity pattern index

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Collimonas</td>
<td>Collimonas</td>
<td>(0.2)</td>
<td>(0.1)</td>
<td>(0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Artemisin</td>
<td>0.5</td>
<td>Artemisin</td>
<td>(0)</td>
<td>(0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Candidatus</td>
<td>0.3</td>
<td>0.4</td>
<td>Candidatus</td>
<td>(0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Warburgia</td>
<td>0.9</td>
<td>0.7</td>
<td>0.8</td>
<td>Warburgia</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER FIVE

DISCUSSION, CONCLUSION AND RECOMMENDATIONS

5.1 Discussion

Most of the RAPD-PCR discriminated bands were highly polymorphic suggesting a high diversity and differentiation within the species in the Kenyan Rift valley. The introduced replicates RAPD-PCR of the same primer products did not show significant reproducibility of bands except for the discriminate bands, even though they were subjected to the same conditions. The resultant bands at some points showed thick alleles suggesting co-migration therefore not revealing much of polymorphisms at those points. The pattern of the differences was based on either the presence or absence of an allele in one region as opposed to the other. From the five sets of oligonucleotide primers only three used in the study produced good significant biological similarities with the lowest E-values for east and west discrimination.

Primer products of IC11, 21 and 28 showed the highest polymorphism and with a high diversity in the given populations. Though the primer products of IC11, IC55 and IC28 showed high polymorphism there were also some clear indications of monomorphic band patterns observed at ~450, 500 and 1000 bp of RAPD-PCR products respectively. The three primers (IC11, IC28 and IC55), had also been used in an earlier study by Muchugi et al., (2008) and
had also revealed high genetic differentiation within and out of the Rift Valley. In this study, additional two primers (IC15 and IC21) screened all together showed a clear genetic difference within the species population East and West of the Rift Valley by producing varied amplification products with the test samples across the region.

Only three WarburgiaIC55W, WarburgiaIC15W, WarburgiaIC28E and their consensus sequences (WarburgiaIC55E, WarburgiaIC15W and WarburgiaIC28W) retrieved revealed East and Western significant homolog’s to plant and bacterial like sequences through BLASTX and TBLASTX. Sequence WarburgiaIC55W revealed plants like homologs to Artemisia annua strain Artemis, Siparuna decipiens and Climacium americanum. WarburgiaIC15W revealed sequence’s homology to plants sspecies like sequences of Populus balsamifera subsp. trichocarpa, Physcomitrella patens subsp. Patens and Ricinus communis (castor bean). The only sequence WarburgiaIC28E revealed bacterial like sequences homologs to Arthrobacter sp. FB24, Rhodococcus erythropolis and Streptomyces species.

The consensus sequences of the significant W. ugandensis (WarburgiaIC55W, WarburgiaIC15W and WarburgiaIC28E) revealed both plant and bacterial like sequences of Artemisia annua strain Artemis, Collimonas fungivorans fosmid and Candidatus Glomeribacter gigasporarum. All these W. ugandensis sequences obtained higher percent identity (above 75%) except WarburgiaIC15W that obtained a lower percent identity below 60%.
The only significant pair-wise sequences alignment were between *W. ugandensis* consensus sequences (*WarburgiaIC55E, WarburgiaIC15W* and *WarburgiaIC28W*) showing long conserved amino acid domains. Most of the amino acid sequences were either substituted by one another or gaps were added for a penalty. Most of the amino acid sequences were conserved to one another, with very few semi conserved regions. The highest scored alignments were between *W. ugandensis* with *Candidatus* and *Artemisia* sequences. *W. ugandensis* also had two orthologs namely *Climacium americanum* and *Populus balsamifera* subsp. *trichocarpa*. The species had ten paralogs namely *Siparuna decipiens, Gossypium, Physcomitrella patens subsp. Patens, Ricinus communis, Arthrobacter sp. FB24, Rhodococcus erythropolis, Streptomyces strain, Collimonas fungivorans fosmid, Artemisin annua* strain *Artemis* and *Candidatus Glomeribacter gigasporarum*.

The only significant results from the phylogenetic analysis relationship revealed 3 clusters; the one between *W. ugandensis* and *Collimonas and Artemisin* showing a low genetic distance and a lower disparity index (rates of amino acid substitutions) of 0.5 and (0.2) respectively. The other cluster was between *Candidatus* and the first cluster of *Collimonas and Artemisin* showing a genetic distance and a disparity index between *Candidatus* and *Artemisin* as 0.4 and (0) respectively. These were also high enough to suggest a gene function between the two sequences. Another cluster analysis was between *Candidatus* and *Collimonas* showing a genetic distance of 0.3 and a disparity
The kind of homologous sequence similarity functions retrieved were not sufficient enough to suggest *W. ugandensis* specific marker traits of adaptation within the Kenyan Rift Valley. Given that the rate of evolution of the sequences compared in the phylogenies is high with genetic distances (of an average 0.8) and a low disparity indices of (0), suggests some evolutionary forces behind the demographic differentiation (Qui *et al.*, 2005; Qui *et al.*, 2006; Rensing *et al.*, 2008). These kind of high genetic distances but low disparity can be as a result of mutational-variation within a particular gene or alleles in a given population (Young *et al.*, 1996). Simple geographical isolation is not seen as the only widespread cause of speciation or interruption of gene flow between populations (Mallet *et al.*, 2009).

He *et al.* (2010) showed that the genetic structure of *Banksia hookeriana*, not solely being dependent on the structure landscape, the habitat located on sand dune crests physically separated by uninhabitable hollows. Kadu *et al.* (2011) in their study on phylogeography of the Afrotomontane *Prunus Africana* revealed a former migration corridor between East and West African highlands rather than demographical influences. Insights into phylogeny and implications of habitat diversification have also been observed in *Farfugium* (Asteraceae) as a monophyletic group associated with a range of habitats in an archipelago of East Asia (Nomura *et al.*, 2010). In a nutshell their suggestions were unbiased to isolation of islands and subsequent parallel adaptation events.
following migration over Quaternary land-bridges along the distribution range. The events might have been introduced by uninformative DNA sequence variation coupled with highly divergent morphologies for adaptive diversification as rapid (Hurst and Warren, 2001; Heslop-Harrison, 2010). The results did not find any link of the sequences to *W. ugandensis* sequences, though not enough in term of its full genome sequence has been done and with only closely related species showing low genome sequences (Muge, 2008; Kadu *et al.*, 2011).

5.2 Conclusion

In conclusion, the study did not reveal RAPD-PCR genetic markers that distinguish *W. ugandensis* species populations either East or West categorised regions in the Kenyan Rift Valley. Although there were also some repeatable sequence variations between East and West populations, these were not specific genome markers. It is possible that the extent of genetic variation is only incipient, reducing the random chance that RAPD-PCR primers would amplify the variable regions, or that variation does not follow the geographical sampling pattern we adopted. Fine-scale markers would clarify this issue. The study suggests that they might have been some chromosomal insertions probably in the form elements such as transposable element’s (Mahillon and Chandler, 1998; Maris *et al.*, 2005; Dvorak, 2009). The results observed from the significant homologous sequences similarity analysis suggest presence of
transposable elements within the genome. The findings provide a basis for targeted study of fine-scale diversity such as through chloroplast DNA, transcriptome sequencing and analysis, Sequence Characterised Amplified Regions (SCARS) and Genome walking.

5.3 Recommendations

i. Future analysis of East and West differences using other marker techniques such as Sequence Characterised Amplified regions (SCARS) and Chloroplast DNA (cpDNA) to know the extent of the differentiation in the two regions.

ii. The conserved protein domain observed can be studied under X rays crystallography analysis to determine structure and functions.
REFERENCES


Sprague Canellaceae) and Zanthoxylum usambarense (Engl.) Kokwaro (Rutaceae) against Plasmodium knowlesi and Plasmodium berghei. *Journal of Ethnopharmacology*, **130**: 158-162.


APPENDICES

APPENDIX I

Primers used for RAPD screening and transformants analysis

Annex 1 RAPD primers used in screening specific markers in the study

<table>
<thead>
<tr>
<th>Arbitrary primer</th>
<th>ICRAF reference</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IC03</td>
<td>5’-CCG AAG CCC T-3’</td>
</tr>
<tr>
<td>2</td>
<td>IC05</td>
<td>5’-GTG CGG ACA G-3’</td>
</tr>
<tr>
<td>3</td>
<td>IC08</td>
<td>5’-CCA GGA AGC C-3’</td>
</tr>
<tr>
<td>4</td>
<td>IC11</td>
<td>5’-GCA TGG AGC T-3’</td>
</tr>
<tr>
<td>5</td>
<td>IC13</td>
<td>5’-CAG GGC CGC T-3’</td>
</tr>
<tr>
<td>6</td>
<td>IC15</td>
<td>5’-GGG ACG TCT C-3’</td>
</tr>
<tr>
<td>7</td>
<td>IC18</td>
<td>5’-CAG CCT CGT C-3’</td>
</tr>
<tr>
<td>8</td>
<td>IC21</td>
<td>5’-GCT AGG GCG G-3’</td>
</tr>
<tr>
<td>9</td>
<td>IC24</td>
<td>5’-CGC GAC GTG A-3’</td>
</tr>
<tr>
<td>10</td>
<td>IC27</td>
<td>5’-ACG CCC CCG A-3’</td>
</tr>
<tr>
<td>11</td>
<td>IC28</td>
<td>5’-CCC GGC TTG T-3’</td>
</tr>
<tr>
<td>12</td>
<td>IC32</td>
<td>5’-AAT CGG GCT G-3’</td>
</tr>
<tr>
<td>13</td>
<td>IC50</td>
<td>5’-TGC TCT GCC C-3’</td>
</tr>
<tr>
<td>14</td>
<td>IC48</td>
<td>5’-GGA CTG GAG T-3’</td>
</tr>
<tr>
<td>15</td>
<td>IC55</td>
<td>5’-GTA GAC CCC T-3’</td>
</tr>
<tr>
<td>16</td>
<td>IC16</td>
<td>5’-ACG GTG CGC C-3’</td>
</tr>
<tr>
<td>17</td>
<td>IC17</td>
<td>5’-GGA GGC CAC T-3’</td>
</tr>
<tr>
<td>18</td>
<td>IC18</td>
<td>5’-CAG CCT CGT C-3’</td>
</tr>
<tr>
<td>19</td>
<td>IC19</td>
<td>5’-TCC GGC CGC T-3’</td>
</tr>
<tr>
<td>20</td>
<td>IC20</td>
<td>5’-TTC CAG CGC A-3’</td>
</tr>
</tbody>
</table>

Annex 2 Colony PCR primer sequences

<table>
<thead>
<tr>
<th>Universal primer</th>
<th>Sequence</th>
<th>Melting temp (°C)</th>
<th>Utility</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>M13 F</td>
<td>5’-GTAAACGACGCGCCA-3’</td>
<td>50.0</td>
<td>forward</td>
<td>16</td>
</tr>
<tr>
<td>M13 R</td>
<td>3’-CAGGAAACAGCTATGAC-3</td>
<td>50.0</td>
<td>reverse</td>
<td>16</td>
</tr>
<tr>
<td>SP 6</td>
<td>5’ GTTGAAACGACGCGCCA-3’</td>
<td>50.0</td>
<td>forward</td>
<td>20</td>
</tr>
<tr>
<td>SP 6</td>
<td>5’ CACAGGAACAGCTATGACC-3’</td>
<td>50.0</td>
<td>reverse</td>
<td>20</td>
</tr>
</tbody>
</table>
Appendix II

Sequence’s retrieved from Warburgia ugandensis

Sequence contig 1 WarburgiaIC28E (736bp)

5’ GAA TGAGGCTTGTCCCGGACAGTAGGAAAGATGCTTGCAGCACCAGGGCGAGAGGCGAGAGGCGAGAGGCGATGACGTGGTGGCAGGCTTGCGGCTCAGTGGGCCGTTCTCGCGGAAGAGTCCCAAGGCTGCTTTCGCGATCCCAAGTCTGCCACGGTTTCACCCCTCGCCGTGACCACATCGTAACCGAAAGTGCAAAGTTCGCCATCTTTTGCGTTGGCAAAACGCAAAAGTACGGCTACATTGTCGTTACCGGCACGCAGTGGTCCCCAACTCAGAAGGACAGAGATGTTCCGCACCCGCGCTACGACGCAGTCTTCTTCGGCCGCTCCGGCTGCCGCCCTAGCA3’

Sequence contig 2 WarburgiaIC11W (338bp)

5’ GCT AGGGCGGAGAGGCGCTGGCTGATGGTCGACCGGGAGAGGCCGGACAGGGCGATGACGTGGTGGCAGGCTTGCGGCTCAGTGGGCCGTTCTCGCGGAAGAGTCCCAAGGCTGCTTTCGCGATCCCAAGTCTGCCACGGTTTCACCCCTCGCCGTGACCACATCGTAACCGAAAGTGCAAAGTTCGCCATCTTTTGCGTTGGCAAAACGCAAAAGTACGGCTACATTG TCATTTCTCGGTTGCGAAACCGAAAGATCGACTGCTGCTGTACCCGCCAGACAGCCCGAAGAATGTGCCTGAGTGGGCAGTAGACCCGCTATGCGGCTTACATTCAGTGGCAAGCTTAACCGATTAGGGAAGGCGTAGCGAAAGCGAGTCCGAATAGGGGTATCAGTCGCTGGGTGTAGACCCGAAACCAAGTGATCTATCCATGGCCAGGTTGAAGGTGGCTAACACGTACTGGAGGACCGAACCCACTAACGTTGAAAAGTTAGGGGATGAGCTGTGGATAGGGTGAAAGGCTAAACAAACTTGGAAATAGCTGGTTCTCTCCGAAAACTATTTAGGTAGTGCTGTCGTGTCTCACCTTCGGGGGTAGAGCACTGTCATGGTTGGGGGGTCTATTGCTGATTACCCCGCATAGCAAACTCCGAATACCGAAGAGTGCAAATCACGGGAGACAGACATCGGGTGCTAACGTCCCGGTGTCAAGAGGGAAACAACCCAGACCGCCAGCTAAGGTCCCTAAATATTGCTAAGTGGGAAAAGTAATAGCTCACTGATCGAGTCTGTCCTGCGCGGAAGATGTA3’

Sequence contig 3 WarburgiaIC55E (796bp)

5’ ATG CCGAAAGAACGTAATTATGGTAGACATGGTAGTACGACAAGTAGGGCGGAGAGGCGGAGAGGCGAGTAGGAAATCCTGCTGAGATGGGACCGCCATCCTCCAAAGGCTTAAATACCTTCTGTAATCTGCACCAGGGTATTGAAACGCTGCTACAGCACGCGTCCTAGTCAATGTCCTGGGAATGTGTCTTCGATGCAGGCGAGCCACGTCTTGTTGCCAAGGACACCCTGCGCATCCTGGATGAGTCGAAGCCTATAGTAAGCTTACGTGCACGCGATGCTAGCCCCAGGGGACGCGTCACTTGTGTGCGTCACGAGGTGCATCCGTGACAACTGGTTTTGGACGGGTCTACA3’

Sequence contig 4 WarburgiaIC21W (292bp)

5’ GTAGACCCCGTGCCTCCCATCTGCTAGGGTACAGTGGTACGACAAGTAGGGCGGAGAGGCGGAGAGGCGAGTAGGAAATCCTGCTGAGATGGGACCGCCATCCTCCAAAGGCTTAAATACCTTCTGTAATCTGCACCAGGGTATTGAAACGCTGCTACAGCACGCGTCCTAGTCAATGTCCTGGGAATGTGTCTTCGATGCAGGCGAGCCACGTCTTGTTGCCAAGGACACCCTGCGCATCCTGGATGAGTCGAAGCCTATAGTAAGCTTACGTGCACGCGATGCTAGCCCCAGGGGACGCGTCACTTGTGTGCGTCACGAGGTGCATCCGTGACAACTGGTTTTGGACGGGTCTACA3’
Sequence contig 4 *WarburgiaIC11W* (358bp)

```
5' GCTAGGGCCGCGGCTGGCTTTTAGAGATGGAGTGGCGACGTCGATGTTTGTTATTATAGGCTCTACCTAGTGAAGTTAGTGGATTAGTTATGAGGCGACGTCGCAATGTTAAGCATTAAGCTTTTTTTTAGTGTTGCATATACCATGCACTCTACTAAACTAGACAAAGACTTCTGCTCGCTCTCTTCTCTCTACTCTTGAGCGAACAACCAGCAAGCAACTTCTAATACTAATAAGGGCTAATAGTAACCCTATAGGGGTAGGCCCCTCTCACGCTACTTGACGTCATAAGGGAGAAGGGCCTTTATGGCCCTCCTCACGCCCTTCTCGCGTATAGTATCCCTCTCCTGCGCCCTAGCA3'
```
Appendix III

pGEMT®- T easy vector cloning maps

The top strand of the sequence shown corresponds to the RNA synthesized by T7 RNA polymerase. The bottom strand corresponds to the RNA synthesized by SP6 RNA polymerase.
Appendix IV

pGEM®-T Easy Vector circle map and sequence reference points

T7 RNA polymerase transcription initiation site (1), multiple cloning region (10–128), SP6 RNA polymerase promoter -17 to +3 (139–158), SP6 RNA polymerase transcription initiation site (141), pUC/M13 Reverse Sequencing Primer binding site (176–197), lacZ start codon (180), lac operator 200–216, β-lactamase coding region (1337–2197), phage f1 region (2380–2835), lac operon sequences (2836–2996, 166–395), pUC/M13 Forward Sequencing Primer binding site (2949–2972), T7 RNA polymerase promoter -17 to +3 (2999–3). Inserts can be sequenced using the SP6 Promoter Primer, T7 Promoter Primer, pUC/M13 Forward Primer, or pUC/M13 Reverse Primer following the company’s protocols.
Appendix V

Media preparation

1. Media Components

(a) LB-Ampicillin (per liter) from commercial prepare LB for growing colonies

- 1 litre of LB agar (10g Bacto-tryptone, 5g yeast extract, 10g NaCl, Adjust pH to 7.5 with NaOH, 15g of agar made up to a volume of 1000 ml)
- Autoclaved at 20 min at 15 psi
- Cool to 55°C
- Add 10 ml of 10 mg/ml filter-sterilized ampicillin. Pour petri dishes (to approximately 25 ml/100 mm plate)

(e) Agar plates for blue white colour screening

LB Ampicillin agar prepared as indicated above.

Ampicillin was added to a concentration of 50 µg/ml at 55 °C, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) to a final concentration of 80 µg/ml [prepared in dimethylformamide (DMF)] then Isopropyl-1-thio- β-D galactopyranoside (IPTG) was added to a final concentration of 20 Mm (prepared in sterile dH₂O. Alternatively, 100 µl of 10 mM IPTG and 100 µl of
2 % X-gal could have also been spread on solidified LB agar plates 30 minute prior to plating the transformations.

For consistent color development across the plate, X-gal and the IPTG were pipetted into 100 µl pool of SOC medium (2% Tryptone, 0.5% Yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM glucose in one litre of molecular grade water) and then the mixture spread across the plate.

**NOTE:** IPTG and X-gal should not be mixed before pipetting into the pool of SOC medium because these chemicals may precipitate.
Appendix VI

Purification of amplified products

The PCR amplified DNA products were excised from the agarose gel using a clean sharp scalpel and the resultant gel slices weighed. Three volumes of solubilisation buffer (QG), was added to 1 volume of the gel and incubated at 50 °C until the gel was completely dissolved. One volume of isopropanol was added to precipitation the DNA on the bond silica gel column and mixed. The sample was then applied to the QIAquick spin column and centrifuged at 10,000 x g (13000 rpm) for 1 minute. The flow through was discarded and QIAquick spin column placed back in the same collection tube leaving the DNA bound on the Silica membrane. 0.5 ml of buffer QG was added to QIAquick column, allowed to stand for 5 minutes and centrifuged for an additional 1 minute.

To wash, 0.75 ml of buffer PE was added to the QIAquick column centrifuged for an additional 1 minute at 10,000 x g. The QIAquick column was then placed in a clean 1.5 ml centrifuged tube and 10 µl of nuclease free Buffer EB (10 mM Tris-HCl, pH 8.5) in a clean RNAase free 1.5 ml microcentrifuge tube and stored at 4°C for cloning purposes. The column was allowed to stand for 1 minute and centrifuged at 10,000 x g for 1 minute to elute DNA. The centrifugations were done in a table-top microcentrifuge (Eppendorf centrifuge 55417 R, GmbH, Germany).