MOLECULAR PHYLOGENY OF KENYAN *Dioscorea species* (YAMS) AND THE QUANTIFICATION OF THEIR DIOSCIN LEVELS

KARIUKI SAMWEL MUIRURI
156/77082/2009
B. Sc. Hons. (Nairobi)

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF

MASTERS OF SCIENCE
IN
PLANT PHYSIOLOGY AND BIOCHEMISTRY

In the School of Biological Sciences
College of Biological and Physical Sciences
University of Nairobi
DECLARATION

This is my original study and has not been presented for a degree in any other university or institution.

Signature 03/05/2012

Date

This thesis has been submitted with this approval as supervisors

Dr. Saifuddin Fidahussein Dossaji
School of Biological Sciences (University of Nairobi)

Signature 03/05/2012

Date

Dr. Jacques Muturi Kabaru
School of Biological Sciences (University of Nairobi)

Signature 3/5/2012

Date
DEDICATION

To Kariuki, this is for you. To my entire Family I say thank you for believing in me. Dad and Mum, Simon, David, Zipporah, Ian, Ann, Terry and Gabriel thank you for believing in me. you are always a source of constant inspiration.
ACKNOWLEDGEMENT

I would like to acknowledge my supervisors Dr. Dossaji and Dr. Kabaru and Dr. Kim who have guided me throughout the course of my research study. To Dr. Dong-Jin Kim whose support has been great, both financially, in Project development and implementation. To my Fellow students and Laboratory mates Eric Magembe, Titus Kathurima, Kimani, and all the members of Laboratory number 7 for the support they have given me throughout the project. I also like to acknowledge International Institute of Tropical agriculture (IITA) for the funds to carry out my research and to Biosciences Eastern and Central Africa (BECA) for hosting me as a student in their laboratories.

To my lecturers in School of Biological Sciences University of Nairobi; Dr. Akunda, Dr. Misra, Dr. Amugune, Dr. Ndegwa, Dr. Githaiga, Dr. Gichuki all who encouraged me throughout the first year of my Masters degree. I would also like to acknowledge my class mates in the course study, Hinga, Ngatia, Juma and the rest of the 2009 class whose encouragement has been insurmountable, guys the journey isn’t over yet!
# TABLE OF CONTENTS

DECLARATION.......................................................................................................................1
DEDICATION........................................................................................................................ 1
ACKNOWLEDGEMENT.........................................................................................................III
TABLE OF CONTENTS..........................................................................................................IV
LIST OF FIGURES................................................................................................................VI
LIST OF TABLES.................................................................................................................. VIII
ABBREVIATIONS................................................................................................................ IX
ABSTRACT........................................................................................................................... x

## CHAPTER ONE...................................................................................................................1

1. **0** INTRODUCTION ...................................................................................................... 1

1.1 LITERATURE REVIEW............................................................................................... 3
   1.1.1 Importance and Agricultural uses of *Dioscorea*................................................. 3
   1.1.2 SAPONIN BIOLOGICAL AND MEDICINAL IMPORTANCE............................... 5
   1.1.3 HPLC QUANTIFICATION OF DIOSCIN ............................................................... 7
   1.1.4 *Dioscorea* spp IN KENYA ................................................................................. 8
   1.1.5 STATUS OF *Dioscorea* spp TAXONOMY ........................................................ 9
   1.1.6 SPECIES RELATEDNESS USING MOLECULAR PHYLOGENY ......................... 11
1.2 JUSTIFICATION .......................................................................................................... 15
   1.2.1 Statement of the problem .................................................................................. 16
1.3 AIM AND OBJECTIVES ............................................................................................. 17
   1.3.1 OVERALL OBJECTIVE .................................................................................... 17
   1.3.2 SPECIFIC OBJECTIVES .................................................................................. 17
   1.3.3 HYPOTHESIS ................................................................................................. 17

## CHAPTER TWO..................................................................................................................18

2. **0** MATERIALS AND METHODS ............................................................................. 18

2.1 COLLECTION OF PLANT MATERIALS....................................................................... 18
2.2 DEOXYRIBONUCLEIC ACID EXTRACTION, POLYMERASE CHAIN REACTION AMPLIFICATION AND Dideoxy SEQUENCING ................................................................. 18
   2.2.1 DNA quantification ......................................................................................... 20
2.3 SEQUENCE ASSEMBLY, ALIGNMENT AND SPECIES POSITIONING .................... 20
2.4 SAMPLES FOR HPLC ANALYSIS .......................................................................... 21
   2.4.1 HPLC analysis ................................................................................................. 22

## CHAPTER THREE..............................................................................................................23

3. **0** RESULTS ................................................................................................................ 23

3.1 *Dioscorea* spp TUBERS AND THEIR MORPHOLOGICAL VARIABILITY .................. 23
3.2 MOLECULAR SPECIES PHYLOGENIES .................................................................. 24
   3.2.1 Gene bank species information for the three target markers ......................... 24
   3.2.2 DNA extraction and quantification .................................................................. 24
   3.2.3 Phylogenies of the gene bank species ............................................................. 28
3.3 MARKERS matK, trnF-L AND rbcL IN MOLECULAR SPECIES PHYLOGENY ............ 33
3.4 APPLICATION OF THE MOLECULAR PHYLOGENY APPROACH TO UNKNOWN SPECIES .............................................................................................................. 41
   3.4.1 Species phylogeny of unknown species based on known African species .......... 41
3.5 DIOSCIN IDENTIFICATION AND QUANTIFICATION .............................................. 50
   3.5.1 HPLC dioscin standard curve ......................................................................... 51
   3.5.2 Identification and quantification of dioscin from *Dioscorea* samples .............. 55

## CHAPTER FOUR...............................................................................................................59
LIST OF FIGURES

Figure 1: The chemical structure of dioscin (Takechi et al., 1991) .............................................6
Figure 2: A map showing Dioscorea spp growing areas in Kenya .............................................9
Figure 3: Illustration of the steps involved in the Dioscorea spp. confirmation through phylogeny .................................................................14
Figure 4: Morphological variability Dioscorea tubers collected from Kenyan markets ............24
Figure 5: Gel-based DNA quantification .......................................................................................26
Figure 6: Gel Image on the PCR amplification testing the different Primers .........................27
Figure 7: PCR amplification of Dioscorea samples with the primers matK – Red, rbcL – yellow and trnF_L – green .................................................................27
Figure 8: Phylogenetic tree for Dioscorea species based on the marker matK using NCBI gene bank sequences .................................................................29
Figure 9: Phylogenetic tree based on the marker rbcL NCBI gene bank sequences .............31
Figure 10: Phylogenetic tree based for the marker trnF_L using NCBI gene bank ...............33
Figure 11: Phylogenetic tree showing usability of molecular phylogeny based on marker rbcL .................................................................................................36
Figure 12: Phylogenetic tree showing usability of the molecular approach based on the marker matK .......................................................................................38
Figure 13: Phylogenetic tree showing position of African yams in respect genebank data based on the marker trnF_L .................................................................40
Figure 14: A phylogenetic tree showing positioning of Kenyan unknown species based on the marker rbcL .................................................................42
Figure 15: species identification based on the known species for the marker trnF_L ...............44
Figure 16: species identification based on the known species sequences for the marker matK .................................................................................................45
Figure 17: Molecular phylogeny of Unknown Kenyan species based on gene bank and known African for marker rbcL .............................................................................47
Figure 18: Unknown species identification/phylogeny based on gene bank species using the marker matK .................................................................................................48
Figure 19: Molecular phylogeny of unknown species based on the marker trnF_L ...............49
Figure 20: Chromatogram for dioscin at a concentration of 1ng/ml ........................................53
Figure 21: A calibration curve for 8 dioscin concentration standards .......................................54
Figure 22: HPLC run for dioscin extract for Kenyan Dioscorea sample KDio71 detected at 265nm, mobile phase; Methanol: Water; Acetic acid (70:30:0.2) ............58
Figure 23: Chromatogram for dioscin at 0.75ng/ul concentration ...........................................82
Figure 24: HPLC chromatogram for dioscin concentration of 0.5ng/ul .....................................83
Figure 25: HPLC run for dioscin at a concentration of 0.25ng/ul .............................................84
Figure 26: HPLC results for the Kenyan Dioscorea sample KDio58 detected at 265nm, mobile phase; Methanol: Water: Acetic acid (70:30:0.2) ...........................................85
Figure 27: Chromatogram for Kenyan Dioscorea sample KDio74 detected at 265nm, mobile phase; Methanol: Water: Acetic acid (70:30:0.2) ...........................................86
Figure 28: HPLC chromatogram results for the Kenyan Dioscorea sample KDio60 detected at 265nm, mobile phase; Methanol: Water: Acetic acid (70:30:0.2) ..............87
Figure 29: HPLC chromatogram for the Kenyan Dioscorea sample number KDio82 detected at 265nm, mobile phase; Methanol: Water: Acetic acid (70:30:0.2) ..................88
Figure 30: HPLC chromatogram result for Kenyan Dioscorea sample number KDio77 detected at 265nm, mobile phase; Methanol: Water: Acetic acid (70:30:0.2) .............89
Figure 31: HPLC chromatogram for the Kenyan *Dioscorea* sample KDio72 detected at 265nm, mobile phase; Methanol: Water: Acetic acid (70:30:0.2) ................................................. 90

Figure 32: HPLC dioscin run for the Kenyan *Dioscorea* sample KDio56 detected at 265nm, mobile phase; Methanol: Water: Acetic acid (70:30:0.2) ......................................................... 91

Figure 33: HPLC chromatogram for the Kenyan *Dioscorea* sample KDio22 detected at 265nm, mobile phase; Methanol: Water: Acetic acid (70:30:0.2) ......................................................... 92
LIST OF TABLES

Table 1: Comparative annual Dioscorea production, Area harvested and yield 1999-2009..... 4
Table 2: PCR profile used in the sample amplification............................................................19
Table 3: Primer sequences used in PCR amplification..............................................................20
Table 4: Spectrophotometer readings for some of the samples under study .......................25
Table 5: Relative numbers of nucleotide substitutions for the three markers rbcL, matK and
tronF  L ........................................................................................................................................34
Table 6: Major characterized African yam species with sequences for the different markers in
the NCBI genebank......................................................................................................................34
Table 7: Dioscin concentrations for calibration curve at Retention time (T_r) 2.3± 0.3 minutes
..................................................................................................................................................50
Table 8: Quantities of dioscin in Different Kenyan Dioscorea samples ...............................57
ABBREVIATIONS

ABI – Applied Biosystems
BecA – Biosciences Eastern and Central Africa
BP – Bootstrap
CL – Compound leaves
DNA – Deoxyribonucleic acid
FAO – Food and Agricultural organization
HPLC – High Performance Liquid Chromatography
IITA – International Institute of Tropical Agriculture
matK – maturase K
ML – Maximum likelihood
MP – Maximum Parsimony
mVs – Microvolt
NJ – Neighbor Joining
PCR – Polymerase chain reaction
rbcL – ribulose 1,5 carboxylase Large subunit
trnF L – transfer RNA Leucin-Phenylalanine
WHO – World Health Organization
Dioscorea (Yams) in Kenya is a neglected crop, notwithstanding its potential as food and in pharmaceutics as a drug source. Research on neglected crops specifically Dioscorea is a sure way of sensitizing the scientific community and policy makers about it. This study used three universal molecular markers (matK, rbcL, trnF |L) to investigate the species position of cultivated Kenyan yams and identify their relatedness to major African species and to species in the gene bank. The study also used Reverse Phase High Performance Liquid Chromatography (RP-HPLC) to identify and quantify dioscin content in Kenyan yams. Dioscin was targeted due to its importance as a source of steroidal drugs.

DNA was extracted from tubers and leaf samples and Polymerase Chain Reaction (PCR) carried out on the DNA samples with direct sequencing of the PCR products. DNA sequences were assembled with sequencher® and multiple sequence alignment through clustalW in megalign from which phylogenetic trees were drawn. Dioscin was extracted from freeze dried tubers, extracted in methanol under sonication and quantified in HPLC UV/Vis light detector.

Results indicated that taxonomic position of Kenyan cultivated yams collected in this study belong to 7 major species represented by Dioscorea cayenensis, D. alata, D. rotundata, D. mangenotiana, D. dumetorum, D. schimperiana and D. bulbifera. Two (D. schimperiana and D. mangenotiana) of the seven have never been documented to be under cultivation in Kenyan before. Dioscin was identified in the Kenyan yam samples under study, albeit at low concentration (v/v). The quantities of dioscin were variable with the least being 0.884 parts per billion (ppb) and the highest 5.12 ppb. Dioscin quantities did not relate to the species observed and was found to vary in even the same species. This work opens up research into yams which is still facing neglect from both the scientific community and the farmers.

**KeyWords**: Dioscorea, PCR, species, dioscin, HPLC
CHAPTER ONE

1.0 INTRODUCTION

*Dioscorea* in Kenya is among the neglected crops with a lot of potential not only as a source of food but also for pharmacological purposes. Yam in Kenya is majorly cultivated in the central highlands, coastal region and western regions. Worldwide, West Africa is the main yam growing belt, where it is among the the main staples. Major global production of yams is for food. *Dioscorea* is a good source of starch and can be used as a supplement or replacement to other starch giving crops like cassava, maize and rice. Being a perennial crop, yam is also known for its abiotic advantages. *Dioscorea* is drought tolerant and is thus a potential food source during drought. Pharmacologically, *Dioscorea* is a superior to other crops in consideration of health promoting molecules exemplified by dioscin which is an antioxidant. Dioscin and dioscorin are the main known antioxidant in yams, with the former being a secondary metabolite and the later a protein. Dioscin is classified as a saponin under the subgroup steroidal saponin. The most important role that the compound diocin has played is in the product of the steroidal hormone progesterone. Apart from being used in the production of steroidal hormones, dioscin has other health promoting effects such as antioxidant effects and reduction of postmenopausal symptoms in women (Wang *et al.*, 2002). It is therefore essential to quantify the amounts of dioscin in *Dioscorea* both cultivated and wild.

The global yam species taxonomy at both inter-species and intraspecific levels has been a concept of controversy, with different authors placing samples into different sections, genus and species. The controversies in yam inter-species and intraspecific taxonomy can be attributed to inherent human errors as well as complex morphological characters. Globally, multiple studies have been conducted on yam taxonomy. Using morphological, molecular
and isoenzyme studies taxon have been reviewed with variable results. Some of these studies have resulted in more confusion than clarity. The dependence on morphological features in species identification is complex because the expertise required is not readily available taxonomy. Furthermore, morphological characterization is time consuming. Another approach that can complement morphological characterization is molecular. Molecular characterization is much cheaper and can work to supplement morphological characterization. Molecular characterization uses DNA to differentiate species.

Studies on Kenyan yams have not been comprehensive either with only a few major studies done. These studies have only been at the intraspecific level concentrating mainly on morphological characters. Moreover, no single molecular study has been conducted on Kenyan yams. This attests to the fact that yam is a neglected crop both at the research level and in cultivation by farmers. Yams have not been given utmost importance especially in the tropics with major studies being only done in western Africa the crop is of a higher importance. It is in this note that this study was made. This study aimed at first making inroads into the studies of yam in Kenya. Second, through this study, identification of the phylogenetic position of collected Kenyan samples at the species level was targeted. Third, due to the importance of dioscin a compound found in yams, the quantities of this was quantified. To quantify dioscin, samples were first phylogenetically positioned in relation to other major cultivated African yam species based on three markers.
1.1 LITERATURE REVIEW

1.1.1 Importance and Agricultural uses of Dioscorea

*Dioscorea* spp are grown for their tubers or storage organs, which are subterranean (*D. rotundata, D. alata*) or aerial (e.g. *D. bulbifera*), and serve a dual agricultural function as source of food and planting material (Coursey, 1967; Hahn, 1995). Yams are widely distributed throughout the tropics with only a few members in the temperate regions of the world (Eka, 1998). Four of the edible *Dioscorea* species which include *D. cayanensis, D. dumetorum, D. bulbifera* and *D. rotundata* have their origin in tropical West Africa; *D. alata* on the other hand has its origin in south Asia (FAO, 1990).

*Dioscorea* are primarily grown for food the tubers of this crops are a source of starch. Several *Dioscorea* species are economically important as staple food, (Coursey, 1967, Degras, 1993) these are actually grown and consumed as primary, secondary or tertiary staple foods (Okwu, and Ndu, 2006). There are many ways which *Dioscorea* tubers can be consumed: they can be dried, ground into flour and stored for use, the flour can also be moistened, molded, boiled and then pounded into “fufu” and eaten with soup (Okwu and Ndu, 2006). *Dioscorea* flour is definitely preferred to the cassava analogue (Obioha, 1972). According to FAO the world *Dioscorea* production and area under harvest have been on the rise with the production over the period of 1999 – 2009 rising by slightly above 15 million tons (FAO, 2011). Total African production rose by 14 million with East African production rising with 18 thousand tonnes (Table 1).

Contrary to the global and regional trends, Kenya *Dioscorea* production has been on a decline with the production decreasing by over 6000 tones in a span of 10 years 1999 - 2009 (FAO, 2011) (Table 1). The decline in *Dioscorea* production in Kenya is a cause of worry for the second most important tuber crop. Farmers are further alienating *Dioscorea* and
preferring other cash crops like maize (Maundu, et al., 1999). There is a need for further research in *Dioscorea* to boost its productivity and increase the area under cultivation.

### Table 1: Comparative annual *Dioscorea* production, Area harvested and yield 1999-2009

<table>
<thead>
<tr>
<th>Countries</th>
<th>1999</th>
<th>2001</th>
<th>2003</th>
<th>2005</th>
<th>2007</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kenya</td>
<td>1087</td>
<td>1000</td>
<td>952</td>
<td>835</td>
<td>925</td>
<td>882</td>
</tr>
<tr>
<td>World (Total)</td>
<td>3955032</td>
<td>4100420</td>
<td>4334657</td>
<td>4552398</td>
<td>4854582</td>
<td>5035358</td>
</tr>
<tr>
<td>Africa (Total)</td>
<td>3810697</td>
<td>3950945</td>
<td>4175546</td>
<td>4378360</td>
<td>4642445</td>
<td>4803781</td>
</tr>
<tr>
<td>--Eastern Africa (Total)</td>
<td>35137</td>
<td>34420</td>
<td>34632</td>
<td>31738</td>
<td>44491</td>
<td>36313</td>
</tr>
<tr>
<td>Kenya</td>
<td>92318</td>
<td>78980</td>
<td>84107</td>
<td>86682</td>
<td>74648</td>
<td>45170</td>
</tr>
<tr>
<td>World (Total)</td>
<td>98485</td>
<td>97537</td>
<td>102071</td>
<td>107852</td>
<td>97714</td>
<td>107436</td>
</tr>
<tr>
<td>Africa (Total)</td>
<td>98267</td>
<td>97161</td>
<td>101956</td>
<td>108208</td>
<td>98151</td>
<td>108475</td>
</tr>
<tr>
<td>--Eastern Africa (Total)</td>
<td>68296</td>
<td>68022</td>
<td>73947</td>
<td>64780</td>
<td>72210</td>
<td>71049</td>
</tr>
<tr>
<td>Kenya</td>
<td>10035</td>
<td>7898</td>
<td>8007</td>
<td>7238</td>
<td>6905</td>
<td>3984</td>
</tr>
<tr>
<td>World (Total)</td>
<td>38951201</td>
<td>39994519</td>
<td>44244575</td>
<td>49098720</td>
<td>47436315</td>
<td>54098112</td>
</tr>
<tr>
<td>Africa (Total)</td>
<td>37446888</td>
<td>38388064</td>
<td>42572460</td>
<td>47377397</td>
<td>45566219</td>
<td>52109183</td>
</tr>
<tr>
<td>--Eastern Africa (Total)</td>
<td>239973</td>
<td>234135</td>
<td>256095</td>
<td>205601</td>
<td>321270</td>
<td>258002</td>
</tr>
</tbody>
</table>

Source: FAOSTAT 15 February 2011
1.1.2 Saponin biological and medicinal importance

Saponins are a group of secondary metabolites which are known to have antioxidant activity. They comprise a large family of structurally diverse compounds containing a steroidal or triterpenoid aglycone linked to one or more oligosaccharide moieties. The aglycone or nonsaccharide portion of the saponin molecule is called the genin or sapogenin. The genin is derived from the saponin through concentrated acid hydrolysis. In the case of dioscin, its hydrolysis in acid results to diosgenin, whose structure does not have the sugar backbone (Hostettmann and Marston, 1995) (Figure 1). Saponins occur constitutively in a great many plant species, in both wild plants and cultivated crops (Fenwick et al., 1991).

Depending on the type of genin present, the saponins can be divided into three major classes: Triterpenoid glycosides, steroidal glycosides which include spirostanol saponin, furastanol saponin, nugatigenin saponin and polipodo saponin and finally the steroid alkoloid glycosides (glykoalkoloids) (Hostettmann and Marston, 1995). In cultivated crops the triterpenoid saponins are generally predominant, while steroid saponins are common in plants used as herbs or for their health-promoting properties, but this is not exclusive because there are some crops like Dioscorea which constitute human food and are also known to have considerable amounts of Saponins.

Dioscin is a steroidal saponin found in among other crops, yams. Dioscin and other saponins are valuable compounds to humans due to their use in pharmacy, industry, cosmetics, agriculture and the food market (Fenwick and Oakenfull, 2006). They have been shown to form insoluble complexes with cholesterol (lindahl et al., 1957). Saponins equally form micelles with most sterols such as bile acids in addition to cholesterol. The interaction is in such way that; sapogenin, the hydrophobic portion of saponin associates with the hydrophobic sterol nucleus, in a stacked micellar aggregation (Oakenfull and Sidhu, 1989).
Diosgenin, derived by acid hydrolysis of dioscin is the primary furostanol saponin found in several plants including *Dioscorea* species (*Dioscorea*), fenugreek and *Costus speciosus* (Sautour et al., 2004). The steroidal sapogenin diosgenin has been used as a precursor for the synthesis of steroid drugs (Oliver et al., 2006). This compound has further been shown to be important in maintaining healthy blood cholesterol levels. The estrogenic and anti-inflammatory effects of diosgenin have been previously hypothesized due to the molecule's structural similarity to estrogen (Higdon et al., 2001). Interestingly, sustained delivery of diosgenin has been shown, by means of bone histomorphometric analysis and three-point bending analysis, to be able to significantly prevent bone loss to the same extent as estrogen, whilst not contemporaneously altering the material properties of bone for ovariectomized mice (Higdon et al., 2001).

![Figure 1: The chemical structure of dioscin (Takechi et al., 1991)](image)

While edible plant species should not have very high amounts of steroids, a preferably high content is important for those species which serve as raw material for the isolation of diosgenin derivatives (Oliver et al., 2006).

From the early 1940, diosgenin contents of various *Dioscorea* species and several different organs were repeatedly estimated, but the values published by several authors are to be compared with caution and should only partially be considered as contribution to an overall
diosgenin screening of the whole genus, since analytical methods as well as the tissue samples varied considerably (Akahori, 1965).

The importance of *Dioscorea* as a source of food and in pharmaceutical industry, have already been highlighted through other studies (Oliver *et al.*, 2006, Kato *et al.*, 1991, Santos *et al.*, 1997, Aquino *et al.*, 1991, Sauvaire *et al.*, 1995, Hu *et al.*, 2007, Peng *et al.*, 1996). The presence of bioactive compounds in *Dioscorea* makes it an important taxon. Just like other storage tubers *Dioscorea* tubers have proteins, these are important not only to the physiology of the plant but also to the consumers (*Dioscorea*) tubers contain dioscorins as the major storage protein accumulated in the vacuoles and accounting for about 85% of total protein (Arango *et al.*, 2010). Dioscorins are present as two classes, with about 70% similarity and with significant homology to a-carbonic anhydrase from various sources. Dioscorin from *Dioscorea* has been shown to be bioactive. Pei-Fen Su *et al*., (2010) in a study on dioscin in *D. batatas* concluded that it may be considered for further development as a dietary supplement for use alongside chemotherapy during cancer treatment due to its bioactivity. Dioscin from *D. batatas* has been shown to possess antioxidant properties, and has dehydroascorbate reductase and mono-dehydroascorbate reductase activities, and an ability to scavenge against both 1, 1-diphenyl-2-picrylhydrazl radicals and hydroxyl radicals (Liao *et al*., 2006).

1.1.3 HPLC quantification of dioscin

Reverse phase HPLC (RP-HPLC) has been previously used in the quantification of steroidal Saponins in many plants (Jane *et al.*, 2005, David *et al.*, 2005). Hubert *et al* (2005) used a simplified HPLC–UV in the analysis of soyasaponins variability in different cultivars (Jane *et al.*, 2005). This method has also been used in model plants; HPLC with a photodiode array detector (PDA) has been used in the quantification of saponins in both the aerial and subterranean tissues of *Medicago truncatula* (David *et al.*, 2005). Furthermore, HPLC has
been used in similar studies in *Dioscorea* species (Zhao, 2007) they used HPLC with evaporative light detection (ELD) in fingerprinting of *Dioscorea nipponica*.

### 1.1.4 *Dioscorea* spp in Kenya

In Kenya, not much information is available on *Dioscorea* and it is actually a neglected crop. However, there are still areas where *Dioscorea* cultivation is still going on (Figure 2); around the central highlands where the major species is *D. minutiflora* Engl. This species is cultivated at an altitude of 1,500-2,400 m above sea level and in areas with more than 700 mm rainfall (Maundu et al., 1999). Western Kenya (Kitale area) and Coastal Kenya (Taita Taveta) are some other areas where *Dioscorea* cultivation occurs in Kenya. *Dioscorea bulbifera* L. and *Dioscorea dumetorum* (Kunth) Pax. are cultivated in Western and Coastal areas respectively (Maundu et al., 1999). Other species of edible *Dioscorea* cultivated in East Africa include the West African *Dioscorea* (*D. cayenensis* or yellow Guinea *Dioscorea*), which are cultivated in Uganda and probably in western Kenya (Maundu et al., 1999). *D. alata* (white *Dioscorea*, originally from India) is also cultivated in Uganda and Tanzania but reportedly not yet in Kenya (Maundu et al., 1999).
1.1.5 Status of *Dioscorea* spp taxonomy

*Dioscorea* belong to the monocotyledonous family dioscoreaceae and genus *Dioscorea* with about 600 species growing in the tropical and subtropical regions of the world. They are twining climbers, often prickly and arising from a tuberous rootstock (Maundu *et al.*, 1999).

Studies on *Dioscorea* taxonomy have been carried out. The first major study on *Dioscorea* systematic was by Knuth (1924). This study had recognized about 600 *Dioscorea* species.

There have been however lots of discontent with the infrageneric taxons of Knuth. Through the study of Wilkin (1999) and Wilkin and Cardick (2000) it was shown that many of the infrageneric taxon of Knuth were either paraphyletic or polyphyletic and the study was therefore limited in that aspect. Many more attempts had been tried to improve on the systematic studies of Knuth but most of these have equally been limited. The first attempt was by Burkill (1939, 1960); although this study had improvements on Knuth’s, there were
Burkill’s emphasis on seed characters, and addition of underground organ morphology and development and male inflorescence morphology to the character set use resulted to a more complex classification to that of Knuth (Wilkin, 2005). Milne-Redhead tried to improve on Burkill’s study of 1939, this study found as unsatisfactory the varieties of *D. hirtiflora* by Burkill, which had been based on the reduction of stamens to staminodes, this study also insist that the naming of Burkill was not right because it was not done in Latin (Milne-Redhead, 1972). Further study by Huber (1998) had done a lot to improve on inherent errors of Knuth and Burkill, whose Taxa were not right (Wilkin, 2005). Huber only concisely repeated Knuth’s and Burkill’s taxa and included all the dioecious taxa of *Dioscoreaceae* in subfamily Dioscoreoideae as He termed this as “genera and genus-equivalent sections”. Despite having the new approach of “genera and genus equivalent sections”, Huber’s work used the same.

The complexity of taxonomic characters used and their variability within a species require a method that can readily distinguish them. For species that are difficult to discriminate morphologically, nucleotide sequences based studies are vital (Leliaert et al., 2009. With the development of molecular systematics, more and more researchers use DNA sequences to solve complicated taxonomic problems and to infer phylogenetic relationships among organisms, including species in *dioscoreaceae* (Gao et al., 2008). *Dioscorea genus* is the only dioecious genus in the family *dioscoreaceae* and comprises of about 600 species (Wilkin et al., 2005). Out of the 600 known species only a few are edible and these include: *D. alata, D. rotundata, D. bulbifera D. esculenta D. trifida, D. dumetorum D. opposita* and *D. cayenensis*. 

limitations too. Burkill’s knowledge of species from China, Madagascar, and the New World was restricted to herbarium material and as well avoided the rank of sub genus in his infrageneric classification of the Old World taxa (Wilkin, 2005).
Although a very important food and medicinal crop, *Dioscorea* has presented a challenge to systematists for many years due to its complex morphological diversity, dioecy, and small flowers (Wilkin *et al.*, 2005). Being the most fundamental unit of comparison in virtually all subfields of biology, from anatomy to behavior, development, ecology, evolution, genetics, molecular biology, paleontology, physiology, systematics and Plant breeding, species (Queiroz, 2005) was chosen as the unit of study. A species describes the smallest taxonomic unit in a population which is capable of producing viable offsprings (Clive, 1938).

Precision in taxonomic identification at species level is crucial not only to plant breeders, but also to pathologists, phytochemists, and other applied biologists (Wilkin, 2005). Plant breeding depends on a form of taxonomy called the ‘special purpose taxonomy’, which is based on degree of inter-fertility between plants (Clive, 1938). It is therefore imperative that accurate taxonomic units are clearly distinguished to enable successful breeding. In this study three markers namely: *matK*, *rbcL* and *trnF_L* were used. The study was carried out using *Discorea* sequences derived from characterized and novel species derived from gene bank and Kenyan *Discorea* species. This study ascertains the applicability of DNA-based approach in distinguishing the morphologically complex *Discorea* *spp* taxon.

1.1.6 Species relatedness using molecular phylogeny

Phylogenetics is the science of estimating the evolutionary past of organisms, in the case of molecular phylogeny, based on the comparison of DNA or protein sequences (Baldauf, 2003). Molecular data, specifically DNA sequences, have received a great deal of attention as a potential source of "phylogenetically informative" characters that are putatively less ambiguous than non-molecular characters (Chase, *et al.*, 1993). Phylogenetics has recently been used in the delimitation of species in living organisms. With gene sequencing, molecular phylogeny has emerged as a major approach for differentiating complex species.
(Baldauf, 2003). Phylogenetic relatedness in species is evaluated from Phylogenetic trees drawn from DNA or protein sequences. A Phylogenetic tree is composed of branches (edges) and nodes, branches connect nodes with a node being the point at which two (or more) branches diverge (Page and Holmes, 1998). Branches and nodes can be internal or external (Baldauf, 2003). A node and everything arising from it is called a 'clade' or a 'monophyletic group'. A monophyletic group is a natural group and all its members derived from a unique common ancestor and having a set of unique common traits (Baldauf, 2003).

Methods for calculating phylogenetic trees are divided into distance-matrix methods, also known as clustering or algorithmic methods, and discrete data methods, also known as tree searching methods (Page and Holmes, 1998). Distance uses a single statistic, the distance which is calculated for all pair wise combinations and the distances are assembled into a tree. Discrete data methods examine each column of the alignment separately and look for the tree that best accommodates all of this information. Distance methods are faster than discrete data methods although they yield little information other than the tree. Discrete data analyses, however, are information rich having a hypothesis for every column in the alignment, evolution can therefore be traced at specific sites in the molecule (Baldauf, 2003). The accuracy of the trees drawn in molecular phylogenetics is of paramount importance. The simplest test of Phylogenetic accuracy is the bootstrap (Felsenstein, 1985). Bootstrapping is essentially a test of whether the whole dataset is supporting the tree, or if the tree is just a marginal winner among many nearly equal alternatives (Felsenstein, 1995). This is done by taking random subsamples of the dataset, building trees from each of these and calculating the frequency with which the various parts of your tree are reproduced in each of these random subsamples (Baldauf, 2003).
Wilkin et al. (2005) viewed that the best approach to revising the systematics of the genus *Dioscorea* was by cladistic analysis of DNA sequences of some species in the genus. Their study was aimed at discovering phylogenetic pattern of relationship in *Dioscorea spp.* and hence producing monophyletic infrageneric entities, they also wanted to study on the *Dioscorea spp.* sections of Knuth and Burkill (1924, 1939) for possible clarification. Wilkin's study had a bias towards Madagascar which has a high *Dioscorea* species per unit area (Wilkin, 2005) which has a high *Dioscorea* species per unit area with one species per 14,500 km$^2$ relative to a species per 53,000 km$^2$ in Zambia. Gao et al. (2008) concentrated on the section *stenophora* which has over 90% of the species being endemic in Asia with a few in Europe and America. African and especially Kenyan yam have therefore not had major studies especially at the molecular level. Three markers were used in the study by Gao (*matK, rbcL* and *trnF_L*) while that of Wilkins (2005) used only two (*rbcL* and *matK*). Two of the regions used in this study have been accepted as the DNA barcodes for plants. Studies that have been done on DNA based taxonomy shows that this method can be used to solve complex phylogenetic problems like in the case of *Dioscorea spp.*
Figure 3: Illustration of the steps involved in the *Dioscorea* spp. confirmation through phylogeny.
1.2 Justification

*Dioscorea* is an important crop both for food and due to its medicinal potential. *Dioscorea* taxonomy both in Kenya and globally has been very challenging to taxonomists based on morphological characterization. Several reasons have been linked to the difficulty in *Dioscorea* species taxonomy. First, *Dioscorea* taxonomy is challenging due to complexity and plasticity of taxonomic character within a species hence difficulty in their identification. Second, *Dioscorea* has been a neglected crop by researchers globally. Clear differentiation of species is of utmost importance for breeding purposes. Mapping species of a given organism is essential not only for conservation purposes but also for breeding purposes. Molecular phylogeny can be used to position species in morphologically complex taxa like *Dioscorea*. Molecular phylogeny is easier to use and easily available. It is with this in mind that this project was carried out.

*Dioscorea* was chosen in this study to demonstrate the applicability of this approach mainly due to the complexity that exists in the morphological characters within this species, the dispute that exist in the current taxonomic groups as well as a considerable number of species with the target genes available in the gene bank. This study aimed at showing the applicability of DNA-based taxonomic approach in species phylogeny with an emphasis on *Dioscorea*. In this study we used three gene regions chloroplastid (*matK*), chloroplast (*rbcL*) and intergenic spacer region *trnF*-*L*. The aim of this study was to find whether the twenty *Dioscorea* with species information already in the gene can be picked out to be of that species using at least one of the three markers and if so can we be able to use phylogeny with certainty. Figure 3 indicates the steps involved in the molecular species phylogeny.

Secondary metabolites in plants have been used by communities over the years for medicinal purposes, food preservation and as pesticides. Identification and quantification of secondary metabolites is essential for further extractions. A secondary metabolite dioscin is found in
yams and is of medicinal and economic importance. Dioscin is an essential component in the manufacture of progesterone, a hormone used in birth control. This project used molecular phylogeny to position Kenyan yam species. The project also used HPLC in the identification and quantification of dioscin in yams tubers.

1.2.1 Statement of the problem

The question of *Dioscorea* species under cultivation in Kenya formed the basis of this study. This study aimed at solving this through molecular phylogeny using DNA nucleotide sequences relating them to known African species and species with information in the genebank.
1.3 Aim and objectives
The aim of the project was to authenticate and establish the taxonomic status of Kenyan Dioscorea species using DNA-based approach and consequently establish and quantify dioscin content.

1.3.1 Overall objective
To phylogenetically identify the species position of Kenya yams, identify and quantify dioscin levels.

1.3.2 Specific objectives
i. To establish the taxonomic status of Kenyan Dioscorea spp under cultivation using molecular phylogenetics.

ii. To determine the presence of dioscin in Kenyan Dioscorea species

iii. To quantify the amount of dioscin in Kenyan Dioscorea species

1.3.3 Hypothesis
Yam samples under study are of the same species, do not contain dioscin and the dioscin quantities do not vary.
CHAPTER TWO

2.0 MATERIALS AND METHODS

2.1 Collection of Plant Materials

Both *Dioscorea* tubers and leaf samples were used in this study. Seven from Madagascar and an out group *Tacca pinnatifida*, seven from International Institute of Tropical Agriculture (IITA) and ten samples collected from markets Embu, Nairobi (Ngara and City markets), Nkubu, Meru, Kikuyu and Limuru markets in Kenya. One hundred and thirty nine samples from Kenya genebank were also used in this work. The samples from the Kenyan genebank were collected from all parts of the country. Kenyan samples used in this study are denoted with the letter K as the initial character. The seven *Dioscorea* samples from IITA were of different species, the Madagascar species were also of different species. The samples collected were divided into two sets, the first set comprised of the Madagascar and IITA collection herein designated Known African yams. The second set was the Kenyan yam collection both from the gene bank and the markets, this collection constitute the Unknown Kenyan yams.

2.2 Deoxyribonucleic acid extraction, Polymerase chain reaction amplification and dideoxy sequencing

Twenty milligrams of silica gel-dried leaf material or 150 milligrams of lyophilized tuber powder was ground in liquid nitrogen. DNA extraction was done using the Qiagen DNAeasy® plant extraction kit (QIAGEN, Inc., Valencia, Calif.), according to the manufacturers protocol with slight modifications (Cheng et al, 2007; Kumar et al 2007; Nováková and Bártá, 2009). The modifications were as follows: 350μl washing buffer instead of 500μl twice in the washing step. The centrifugation speed was set at 6000g for 1
minute in the first two washes and a speed of 6000g for 3 minutes in the final wash followed by a final 12000g centrifugation for 5 minutes. Modification in the elution step involved heating of elution buffer to 60°C instead of using it at room temperature and not waiting for 5 minutes after pipetting the elution buffer to the binding column but rather, centrifuging immediately. Elution was done with 200 µl volume of buffer.

Synthetic oligo-nucleotides for polymerase chain reaction (PCR) primers (*rbcL*m-3, *trnF* _L_ and _matK_) were ordered from Biooneer and DNA amplification done using these 3 primer pairs. PCR amplification profile was as set in the table 2, with the annealing temperature set at 57 °C. The markers used code for ribulose 1,5 biphosphate carboxylase large subunit (*rbcL*), Maturase K (*MatK*) and transfer RNA intergenic spacer (*trnF* _L_). The primers sequences were as shown in the Table 3. A Bionneer premix 10ul reaction volume PCR was made with addition of 8ul of milli Q water, 0.5 ul of 5pmole/ul primer and 1 ul of 5ng/ul DNA.

**Table 2: PCR profile used in the sample amplification**

<table>
<thead>
<tr>
<th>Step</th>
<th>Time</th>
<th>Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Premelt</td>
<td>2 mins 30 secs</td>
<td>94°C</td>
</tr>
<tr>
<td>Denaturation</td>
<td>1 min</td>
<td>94°C</td>
</tr>
<tr>
<td>Annealing</td>
<td>45 secs</td>
<td>57°C</td>
</tr>
<tr>
<td>Extension</td>
<td>30 secs</td>
<td>72°C</td>
</tr>
<tr>
<td>Final extension</td>
<td>7 mins</td>
<td>72°C</td>
</tr>
<tr>
<td>Number of cycles</td>
<td>35</td>
<td></td>
</tr>
</tbody>
</table>

PCR products were cleaned using the Bioneer® PCR purification kit. PCR amplification products were sequenced by ABI’s (Applied Biosystems) dideoxy cycle sequencing technique using the Bigdye terminator kit V.3.1. Sequencing reactions were run on an ABI3730 and ABI3130 automated sequencers. All these steps (cleaning, amplification and sequencing) were done according to the manufacturer’s protocols. Sequences were then assembled in Sequencher 4.6 software.
Table 3: Primer sequences used in PCR amplification

<table>
<thead>
<tr>
<th>Position</th>
<th>Primer name</th>
<th>Sequence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward</td>
<td><em>matK</em> MF</td>
<td>ATT TGC GAT CTA TTC ATT CAA T</td>
<td>Gao et. al., 2008</td>
</tr>
<tr>
<td>Reverse</td>
<td><em>matK</em> MR</td>
<td>TGA GAT TCC GCA GGT CAT T</td>
<td>Gao et. al., 2008</td>
</tr>
<tr>
<td>Forward</td>
<td><em>rbcL</em> m3F</td>
<td>TAT CTT AGC GCC ATT CCG AGT A</td>
<td>Gao et. al., 2008</td>
</tr>
<tr>
<td>Reverse</td>
<td><em>rbcL</em> m4R</td>
<td>CGC GGA TAA TTT CAT TAC CTT C</td>
<td>Gao et. al., 2008</td>
</tr>
<tr>
<td>Forward</td>
<td><em>trnF</em> L cF</td>
<td>CGA AAT CGG TAG ACG CTA CG</td>
<td>Gao et. al., 2008</td>
</tr>
<tr>
<td>Reverse</td>
<td><em>trnF</em> L f R</td>
<td>ATT TGA ACT GGT GAC ACG AG</td>
<td>Gao et. al., 2008</td>
</tr>
</tbody>
</table>

2.2.1 DNA quantification

DNA quantification was done using both gel electrophoresis and by spectrophotometer (nanodrop). Quantification using gel electrophoresis was by running a known quantity of the ladder in parallel with the extracted DNA.

2.2.1.1 Preparation of agarose gels

Standard 2% and 0.8% (w/v) were prepared by dissolving 2 and 0.8 grams of agarose in 100ml of 1×TBE (0.001M tris-borate, 0.001M EDTA, pH 8.0) electrophoresis buffer. The agarose in TBE buffer was heated in a microwave to allow complete dissolving at 90°C. The gel was cooled to about 20°C. Five microlitres of gelred® DNA staining dye was added.

2.3 Sequence assembly, alignment and species positioning

Sequences for all the markers were imported to Sequencher® and assembly done at 85% minimum percentage match and 20% minimum overlap. Both ends of the contigs were trimmed to rid off dirty sequences. Manual editing of ambiguities was done in correlation to
the base calls in the chromatograms. Sequences were exported per marker and each of the three group of sequences (Gene bank sequences, Known African Dioscorea and the unknown Kenyan Dioscorea) saved separately. Multiple sequence alignment and visualization was done using clustalW. Three multiple sequence alignments were done for sequences from each of the three markers (rbcL, matK and trnFL).

Progressive multiple sequence alignment and visualization was done using clustalW (Larkin et al., 2007) in MEGA 5. Separate alignments were done for each of the three markers (rbcL, matK and trnFL). The multiple alignments were performed with the default setting: a 15.00 gap opening penalty, Gap penalty of 6.66, DNA transition weight at 0.05, delay of divergent sequences at 30% and the DNA weight matrix being IUB. Phylogeny reconstruction for each of the markers was done using Maximum likelihood (ML) and Neighbor Joining (NJ) algorithms. Bootstrap method was used as the test of phylogeny with 1000 number of bootstrap replicates. The model used in both cases was Kimura-2-parameter and the rate among sites was set as uniform. Missing data was treated by complete deletion while the ML heuristic method was by Nearest-Neighbor-Interchange (NNI).

2.4 Samples for HPLC analysis

Dioscorea tubers were collected from major markets including Meru, Nkubu, Embu, Muranga, Thika, kikuyu and Ngara markets while others were also kindly provided by the Kenya gene bank. Western African Dioscorea were brought in from IITA Ibadan-Nigeria. Tuber collection from Kenyan markets was based on the local names given by traders; these were treated as different samples. The tubers were washed, peeled and sliced into small pieces of approximately one by one centimeter. These were then immersed into liquid nitrogen in readiness for freeze drying. Freeze drying was done to avoid the loss or change in the nature of the target compound. The Samples were then ground into fine powder and stored in air tight containers in readiness for dioscin extraction.
One gram of powdered tuber was weighed into a 10ml falcon tube and 8ml of the 100% methanol added into this powder. Extraction was done under sonication for 1 hour and the extract centrifuged at 1000 g for 10 minutes. The solvent was evaporated under reduced pressure to yield methanolic extracts. Solvent extraction was sequentially carried out with 4ml ethyl acetate (EtOAc) and 4ml of n-butanol three times each. The extracts were then filtered through 0.25um millipore filters and evaporated then re-suspended in 100% methanol to give dioscin extract. The dioscin extract was now ready for analysis in the HPLC machine.

2.4.1 HPLC analysis

A Younglin 9100® HPLC system equipped with a 4.6 * 150mm C18 column, a YL.9110® Quaternary pump, a Vacuum degasser and a UV/Vis detector was used. Samples and standards were filtered through Millipore® PTFE filters to remove particulate matters that would otherwise clog the column. A 20 µL volume of the filtered standard or sample solution was directly injected through the Rheodyne® 7725 manual injector with the knob at INJECT position. Loading was then done by turning the knob to the position LOAD. The analysis was performed on a gradient system with Methanol: water: Acetic acid in the ratio 70:30:0.01. The HPLC separation column temperature was set at 37°C and the flow rate at 1ml/min. YLClarity® chromatography chromatogram analysis software was used to display the integrated data and peaks. The solvent system was degassed daily before use, organic solvents were degassed using a filter membrane through a vacuum with stirring. Water was degassed by sonication for 1 hour before use.

The total dioscin quantity from the initial Dioscorea powder of 5g was obtained by calculating the total amount of dioscin in ng. The total volume of the sample dissolved in (µl) was multiplied by the concentration in ng to give the total dioscin quantity in ng (Total quantity in ng = concentration (ng/µl) * total dissolving volume).
CHAPTER THREE

3.0 RESULTS

3.1 *Dioscorea* spp. tubers and their morphological variability

Tubers collected from the markets were found to be varying in both the morphology and sizes. These tubers were being referred to by different local names in different regional markets. This was due to the differences in dialects of the traders in the different areas. The pictures in (Figure 4) indicate variability in the tuber morphologies and sizes. The tuber morphologies varied from oval/oblong (Figure 4 diagrams 4 and 6) and elongated and cylindrical (Figure 4 diagrams 5 and 2). These tuber morphological characters have been used by the IPGR in yam species taxonomy. Other characteristic that were observable included branching which occurred at either the base of the tuber or in the middle (3 and 1 respectively). Skin texture was also observed to vary, for the samples pictured in figure 4, samples 1 and 5 had course skin texture while the rest's were less course. Other samples (Not included in the picture) were observed to be smooth in their tuber skin texture. Samples numbered 1 to 6 are represented by the samples 1 – KEMBUA, 2 – KMBEUM, 3 – KMXK, 4 – KDio57, 5 – KDio62 and 6 – KDio56.
3.2 Molecular species Phylogenies

3.2.1 Gene bank species information for the three target markers

The search carried out in the National Center for Biotechnology Information (NCBI) for *Dioscorea* resulted in 124 *Dioscorea* species. Three different markers were targeted in the gene bank search due to their use in similar research (Wilkins, 2005). 75 sequences were identified for the marker ribulose 1, 5 carboxylase large subunit (*rbcL*) (Fig. 9), 69 for marker maturase K (*matK*) (Fig. 8) and 56 for transfer RNA phenylalanine and leucine (*trnF_L*) (Fig 10), all these sequences were species specific.

3.2.2 DNA extraction and quantification

The spectrophotometer readings for DNA quantification were as indicated in the Table4. The ratio 260/280 indicates DNA contamination with proteins and the ideal value is at 2.0, and that of 260/230 is 1.8. The quantities of DNA from the samples were ranging between 1 and 26 ng/ul. DNA extracted from tubers resulted in lower quantities relative those from leaf samples. Total DNA quantities were ranging between 218 and 5268 ng in 200 μl of elution buffer.
Table 4: Spectrophotometer readings for some of the samples under study

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>EMBU_A_2_UNQ</td>
<td>8.55</td>
<td>1710</td>
<td>0.171</td>
<td>0.112</td>
<td>1.53</td>
<td>0.42</td>
</tr>
<tr>
<td>DIO 1</td>
<td>1.56</td>
<td>312</td>
<td>0.031</td>
<td>0.032</td>
<td>0.99</td>
<td>0.41</td>
</tr>
<tr>
<td>DIO 28</td>
<td>3.15</td>
<td>630</td>
<td>0.063</td>
<td>0.041</td>
<td>1.54</td>
<td>0.59</td>
</tr>
<tr>
<td>DIO 43</td>
<td>1.41</td>
<td>282</td>
<td>0.028</td>
<td>0.019</td>
<td>1.49</td>
<td>0.7</td>
</tr>
<tr>
<td>DIO 51</td>
<td>13.23</td>
<td>2646</td>
<td>0.265</td>
<td>0.171</td>
<td>1.55</td>
<td>0.67</td>
</tr>
<tr>
<td>Y-075</td>
<td>26.34</td>
<td>5268</td>
<td>0.527</td>
<td>0.296</td>
<td>1.78</td>
<td>1.04</td>
</tr>
<tr>
<td>Y-323</td>
<td>5.26</td>
<td>1052</td>
<td>0.105</td>
<td>0.063</td>
<td>1.66</td>
<td>0.51</td>
</tr>
<tr>
<td>Y-475</td>
<td>11.48</td>
<td>2296</td>
<td>0.23</td>
<td>0.122</td>
<td>1.89</td>
<td>0.73</td>
</tr>
<tr>
<td>Y-723</td>
<td>18.16</td>
<td>3632</td>
<td>0.363</td>
<td>0.207</td>
<td>1.75</td>
<td>0.56</td>
</tr>
</tbody>
</table>

Quantification was also done on an agarose gel using lambda DNA. The images were captured on a camera under UV light and the image observed is as indicated in Figure 5. The samples were quantified by relating to the lambda DNA intensity. Numbers 2 and 12 are lambda DNA intensities produced by loading 100ng of lambda DNA. The relative intensity of DNA samples is indicated in the Figure 5 and amount in nanograms (ng). Most of the samples were found to be 15ng or below.
3.2.2.1 Polymerase chain reaction (PCR)

Polymerase chain reaction was conducted on the DNA quantified with the different primers that had different amplification sizes (Figure 6). The PCR products for the primer *rbcL* were 1300 base pairs (bp) long while *matK* and *trnF_L* had 1000 and 900bp respectively. The different sizes were identified from 100bp marker (100 bp because each band of the marker is separated from the next with 100 base pairs). More amplification of DNA samples results are as shown in the Figure 7. Here, 24 DNA samples were amplified with the 3 primers, red indicates amplification with *matK*, yellow with *rbcL* and green with *trnF_L*.
Figure 6: Gel Image on the PCR amplification testing the different Primers

Ninety five percent of all the samples had successful amplifications. DNA sample numbers 3, 7 did not amplify with the marker \textit{matK}. Samples numbers 7 and 17 did not amplify with the marker \textit{rbcL}. The marker \textit{trnF\_L} did not result in amplification for the samples numbers 1, 3, 7 and 11. The annealing temperature for the three markers was at 57°C. The multiple amplifications observed across all the markers were resolved by increasing the annealing temperature to 60°C. The marker targeting the region 16srDNA was not used in consequent analysis this is because it resulted in multiple amplifications even after increasing the annealing temperature to 60°C.

Figure 7: PCR amplification of \textit{Dioscorea} samples with the primers \textit{matK} – Red, \textit{rbcL} – yellow and \textit{trnF\_L} – green.
3.2.3 Phytogenies of the gene bank species

Phytogenies for the gene bank sequences resulted in specificity with closely related species clustering in a clade, this was consistent across all the three markers. The result of the phylogenetic results for the marker *matK* based on gene bank sequences is shown (Fig. 8). Monophyletic groups based on this marker segregated majority into Asian and African yams species. The African group was represented by two major groups namely Madagascar and other African yams. Madagascar yams include the group with the species *D. sosu*, *D. fandra*, *D. maciba* and *D. birmanica*. Madagascar yams were well supported in the marker *matK* at a 79 bootstrap value (BP). Other two clades that had some African and Malagasy representative Malagasy clade include first the clades represented by the species *Dioscorea daunen*, *D. birmanica* and *D. sansibarensis*. The second sister clade is one with the species *D. antaly*, *D. bulbifera* and the known African yam *D. dumetorum*. *Dioscorea* species evolutionary history for the marker *matK* was inferred using Maximum parsimony (MP) method, which involves the assertion that differences between two species are well explained by their simplest difference. Tree number 22 out of 154 most parsimonious trees (trees with the simplest evolutionary difference) (length = 224 the tree with the shortest distance is the most parsimonious) is shown in Figure 8. The MP tree was obtained using the Close-Neighbor-Interchange algorithm with search level 1 in which the initial trees were obtained with the random addition of sequences (10 replicates). The analysis for the marker *matK* involved 58 nucleotide sequences, this is so because of these, 11 sequences were from different varieties but of the same species. They were aligned and the consensus sequence used as the representative for that species. Codon positions included were 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup> and noncoding. All positions containing gaps and missing data were eliminated. There were a total of 618 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura et
Species specificity was in line with the expectations and this marker could therefore be used in species phylogeny.

Figure 8: Phylogenetic tree for *Dioscorea* species based on the marker *mutK* using NCBI gene bank sequences.

Based on the marker *rbcL*, sequences obtained from the gene bank indicated regional congruence. The evolutionary history based on the marker *rbcL* (Fig. 9) was inferred using
the MP method. Similar to the case of the marker \textit{matK}, the marker \textit{rbcL}, resulted in regional monophyletic groups represented by Asian, two African (Malagasy and rest of Africa) and European groups. The marker \textit{rbcL} resulted in two clades within the African yams, one with species purely from Malagasy and the other with a mixture of both Malagasy and African species \textit{D. bulbifera} and \textit{D. rotundata}. The Malagasy species \textit{D. antaly} is especially closely related to \textit{D. dumetorum} with the marker \textit{rbcL} too. The evolutionary history for the marker \textit{rbcL} was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 0.18617000 is shown (Fig.9). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown above the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The rate variation among sites was modeled with a gamma distribution (shape parameter = 0.6). The analysis involved 74 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 478 positions in the final dataset.
Figure 9: Phylogenetic tree based on the marker *rbcL* NCBI genebank sequences.
The analysis for the marker *trnF*L involved 41 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 413 positions in the final alignment. This marker had the common clades; African and Asian yams. The African yams clade had subclades with Malagasy and other African species including the species *D. praehensilis* (Fig. 10). The Asian clade was well supported at a BP of 100. The species in this clade included *Dioscorea tokoro, D. nipponica, D. himifolia* and *D. gracillima*.

The evolutionary history for the marker *trnF*L was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 0.24358836 is shown (Fig 10). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The rate variation among sites was modeled with a gamma distribution (shape parameter = 0.6). The analysis involved 41 nucleotide sequences. All positions containing gaps and missing data were eliminated.
3.3 Markers *matK, trnF*L and *rbcL* in molecular species phylogeny

The known species information set was of two different origins, the IITA set and the Madagascar collection. Presence of known species sequences for the three markers was confirmed (Table 6) The species *D. minutiflora* only had sequence for the marker *trnF*L in the gene bank; this was the same case with *D. cayenensis* and *D. praehensillis*. *D. mangenotiana* did not have any sequence for the three markers. Most of the species within
the reference set had nucleotide sequences for at least one of the three markers with some species having for all the genes. Phylogeny of the sequences obtained resulted in regional specific clusters, the Asian group includes species

<table>
<thead>
<tr>
<th>Table 5: Relative numbers of nucleotide substitutions for the three markers rbcL, matK and trnF_L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marker</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>rbcL</td>
</tr>
<tr>
<td>matK</td>
</tr>
<tr>
<td>trnF_L</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 6: Major characterized African yam species with sequences for the different markers in the NCBI genebank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dioscorea species</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>D. alata</td>
</tr>
<tr>
<td>D. minutiflora</td>
</tr>
<tr>
<td>D. antaly</td>
</tr>
<tr>
<td>D. bemarivensis</td>
</tr>
<tr>
<td>D. bulbifera</td>
</tr>
<tr>
<td>D. cayanensis</td>
</tr>
<tr>
<td>D. domotorium</td>
</tr>
<tr>
<td>D. fandra</td>
</tr>
<tr>
<td>D. maciba</td>
</tr>
<tr>
<td>D. mangenotiana</td>
</tr>
<tr>
<td>D. praehensilis</td>
</tr>
<tr>
<td>D. rotundata</td>
</tr>
<tr>
<td>D. sansibarensis</td>
</tr>
<tr>
<td>D. soso</td>
</tr>
<tr>
<td>Tacca pinnatifida</td>
</tr>
</tbody>
</table>

Alignment of the sequences from known African Dioscorea with the gene bank’s were done to establish the usability of the molecular approach. Figure 11 indicates the alignment for the
marker \textit{rbcl}. The phylogenetic tree from \textit{rbcl} resulted in two major clades denoted as Asian and African yams. Asian clade had a BP support of 51, this clade branched into two sister clades. The first Asian sister clade was branching with BP support value of 50, this was represented by the among others the species \textit{D. bimifolia}, \textit{D. prazeri}, \textit{D. nipponica} and \textit{D. tokoro}. The second Asian sister clade was observed to a BP support of of 7, this is the clade with the species \textit{D. dannea}, \textit{D. nittens}, \textit{D. suhecalva} and \textit{D. hemsleyi} (Fig 11).
Figure 11: Phylogenetic tree showing usability of molecular phylogeny based on marker rbcL.

The African clade using the marker rbcL segregated into main sub-clades. Each sub-clade had its own sister clade.
The two main African clades are denoted in figure 11 as Malagasy and other African yams and the second one as other African yams and Malagasy. The first clade is supported at a BP value of 41 while the second is at 21. The clade at 41 BP comprises of the species *D. dumetorum* and *D. hispida* with *D. antaly* occurring as a sister species but within the same clade. The clade supported at 21 has the species *D. delaveyi D. kamoenensis* and *D. esquirili* as a clade and the species *D. hanzhuana, D. arachnida* and *D. bulbifera* being sister species to them. The second African clade denoted in figure 11 as other African and Malagasy yams is supported at 15 BP. This clade equally has subclades within it. The subclade with the species *D. alata* is a main clade and this is supported at a BP of 64. Other species within this clade segregate as sister species to *D. alata*. Among these are the economically important African species *D. rotundata* and *D. cayanensis*.

The evolutionary history for the marker *rbcL* was inferred using the Neighbor-Joining method. The total alignment resulted in 318 positions this is exclusive of gaps and missing data which was eliminated. A total number of 86 positions were used in the construction of the phylogenetic tree. The optimal tree with the sum of branch length = 0.17591694 is shown (Fig 11). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown above the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The rate variation among sites was modeled with a gamma distribution (shape parameter = 0.6). The analysis involved 86 nucleotide sequences. Codon positions included were 1\textsuperscript{st}, 2\textsuperscript{nd}, 3\textsuperscript{rd} and Noncoding were used in this phylogeny.
Figure 12: Phylogenetic tree showing usability of the molecular approach based on the marker matK
Figure 12 is the resultant tree for the gene bank and known African Dioscorea species based on maximum likelihood. The optimal tree with the sum of branch length = 0.40397045 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown below the branches. The evolutionary distances were computed using the p-distance method and are in the units of the number of base differences per site. The analysis involved a total of 79 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 577 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura et al, 2011).

Based on the marker matK clades segregated in a similar manner as with the markers rbcl. and trnF_L earlier discussed. African yams were observed to segregate into one major clade, this is the clade with the species D. alata, D. calcicola, D. mangenotiana, D. cayanensis and D. rotundata this clade was supported at a BP support of 45. Then clade with African yams is equally divided into two subclades one supported at 63 and the other at 26. The second clade with African yams is one with the species D. bulbifera, D. japonica and D. dumetorum. This clade is supported at a BP of 33. With the marker rbcl., there was observed also the clade with many of the Madagascar yam species, this is the clade supported at a BP value of 2 with the species D. hemsrivensis, D. arcuatinives, D. sasx, D. trichantha and D. sandra represented.

The Phylogenetic tree for the trnF_L for the gene bank sequences and the sequences from known African Dioscorea is as indicated in the Figure 13. Three major clades supported at 37, 100 and 99 BP values were observed in the phylogenetic tree based on the marker trnF_L.

The clade supported at 37 has subclades with different species, the first subclade has the species D. togoensis, D. minutiflora, D. cayanensis and D. burkilliana. The other subclade comprise the species D. alata and D. fordii. Other subclades are the clade with the species D. 
rotundata and D. praehensilis, the other with the species D. dumetorum and D. trifida all the other species are representes as sister species to the subclades mentioned.

Figure 13: Phylogenetic tree showing position of African yams in respect to genebank data based on the marker trnF_L.
3.4 Application of the molecular phylogeny approach to unknown species

On identifying the usability of the molecular approach based on the known African Dioscorea species vis-à-vis gene bank species, the approach was then to be applied on “unknown species” that need phylogeny or confirmation. The aim of this approach is to apply the approach in real life situation like the case of Kenyan Dioscorea.

3.4.1 Species phylogeny of unknown species based on known African species

The known African species were first to be used as a reference in the species identification. These were aligned to the unknown Kenyan species as it was previously done in the usability test. The alignments for the three markers were as well done separately. The reason for use of the few known African Dioscorea species was to first authenticate and or confirm the Kenyan Dioscorea species based on these few species and this could result to further identification based on gene bank species. Based on all the markers Kenyan yams under study were found to be strongly supported to D. cayanensis with some linked to D. rotundata, D. dumetorum and D. alata. Based on the marker \textit{rbcL}, three major clades were observed, these were supported at 87, 85 and 96 BP values. The clade with 87 BP support had three major subclades supported at 60, 86 and 62. These subclades contained the species D. cayanensis, D. alata and D. rotundata respectively. The clade supported 85 BP had the species D. dumetorum forming the main sub-clade with D. antaly being the sister species. This subclade was supported by a strong BP of 96 (Fig 14). The final clade in the \textit{matK} phylogenetic tree was the clade supported at 96 and extending the species D. hemarivensis and D. fandra.
Figure 14: A phylogenetic tree showing positioning of Kenyan unknown species based on the marker \( rbcL \).

The species position of other Kenyan samples KDio155 using the marker \( rbcL \) could not be cluster with any known African yam specie used in this study. The Kenyan samples that were strongly supported included K70, 76 and K149 these were supported to the species \( D. \) rotundata.
The marker *trnF L* based on known African yams resulted in Kenyan species positioning into *D. cayanensis* supported at 62 BP. The Kenyan sample KDio60 was strongly supported to the species *D. alata* (Fig. 15). *D. rotundata* and *D. dumetorum* were not supported together with any Kenyan sample, the reason being the marker *trnF L* resulted in low quality reads during sequencing.

Based on the marker *matK*, four main clades were observed. The four were supported at BP values of 65, 88 another 65 and 99 from top of the tree downwards. The first Clade supported at 65 BP value had three subclades branching from it. The first subclade supported at a BP value of 69 had within it a subclade supported at a BP value of 56 with *D. cayanensis* being the known African yam species. In this clade, the Kenyan samples K18, K19, K13, K132 to name a few were supported. The Kenyan sample number KDio59 the lone sample that does not fall at the *D. cayanensis* subclade.

The subclade supported at a BP of 84 links Kenyan Yam samples to the known African species *D. mangenotiana* and *D. rotundata*. In this cluster, Kenyan Yam samples like KW1, KW2, K115 and KDio44 among others are represented. The marker *matK* did not clearly differentiate some of the Kenyan samples K13, K116, K123 and K155 between the species *D. mangenotiana*, *D. cayanensis* or *D. rotundata*. 
Samples closely related to the species *D. cayanensis*

Samples linked to the species *D. alata*

Figure 15: species identification based on the known species for the marker *trnF* 1.
Samples linked to the species *D. cayamensis*

**Figure 16:** species identification based on the known species sequences for the marker *matK*
3.4.1.1 Further Identification based on gene bank species

Since there were only few species within the reference set and due to a possibility of an imperfect match, in the species identification based on the known African *Dioscorea* species, further analysis based on the gene bank sequences was necessitated. Kenyan samples that had not been phylogenetically positioned using Known African yam samples were positioned using the three markers.

Based on the marker *rbcL*, Kenyan sample K155 strongly clustered to the species *D. bulbifera* at a BP of 62 (Fig 17). The sample numbers KW8, K70, K149 and K76 clustered to the species *D. rotundata* with a strong BP value of 59. With the marker *matK*, three occurrences were observed for the samples that had previously been unidentified using Known African species. The sample number KW10 formed a clade with the species *D. schimperiana* at a BP value of 62 (Fig 18). The other K155, K1452 and K1458 were strongly supported to the species *D. bulbifera* at a strong BP value of 86. The sample number KW7 clustered to the species *D. dumetorum* while samples K115, K70, K144 and K144, KW1, KW2, KW3, KW4 and KW9 strongly supported to both species *D. rotundata* and *D. mangenotiana*, these two species could not be differentiated further. Based on the marker *trnF_L*, there was major identification since all the samples had earlier been positioned using known African yam species (Fig 19).
Figure 17: Molecular phylogeny of Unknown Kenyan species based on gene bank and known African for marker rbcL.
Figure 18: Unknown species identification/phylogeny based on gene bank species using the marker \textit{matK}
Samples linked to
D. Cayanensis

Figure 19: Molecular phylogeny of unknown species based on the marker trnL-F.
3.5 Dioscin identification and quantification

Kenyan *Dioscorea* samples were identified to be *D. cayanaensis*, and due to the limited amount of sample for HPLC analysis, only 15 out of study samples were used in dioscin quantification (Table 8). Dioscin was identified in all the 15 Kenyan *Dioscorea* samples albeit at low quantities (Table 9); the retention time of the dioscin identified was at 2.3 ± 0.3 minutes. The dioscin quantities were quantified using the external standard method (ES11) where a calibration curve at different concentrations was used in drawing of a concentration curve. The loading volume was 20ul and the total dioscin quantity loaded was obtained by multiplying the concentration by the loading volume (Table 8)

Table 7: Dioscin concentrations for calibration curve at Retention time (t_r) 2.3± 0.3 minutes

<table>
<thead>
<tr>
<th>Response (area MvS)</th>
<th>Loading volume (ul)</th>
<th>Concentration (ng/ul)</th>
<th>Loaded dioscin quantity (ng)</th>
<th>Linear response</th>
<th>Linear amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>166383.494</td>
<td>20</td>
<td>1</td>
<td>20</td>
<td>166383.494</td>
<td>20</td>
</tr>
<tr>
<td>118105.465</td>
<td>20</td>
<td>0.75</td>
<td>10</td>
<td>118105.465</td>
<td>10</td>
</tr>
<tr>
<td>44534.034</td>
<td>20</td>
<td>0.5</td>
<td>5</td>
<td>44534.034</td>
<td>5</td>
</tr>
<tr>
<td>22520.285</td>
<td>20</td>
<td>0.25</td>
<td>2.5</td>
<td>22520.285</td>
<td>2.5</td>
</tr>
<tr>
<td>11018.144</td>
<td>20</td>
<td>0.125</td>
<td>1.25</td>
<td>11018.144</td>
<td>1.25</td>
</tr>
<tr>
<td>5738.569</td>
<td>20</td>
<td>0.0625</td>
<td>0.625</td>
<td>5738.569</td>
<td>0.625</td>
</tr>
<tr>
<td>3311.265</td>
<td>20</td>
<td>0.03125</td>
<td>0.3125</td>
<td>3311.265</td>
<td>0.3125</td>
</tr>
<tr>
<td>1804.382</td>
<td>20</td>
<td>0.015625</td>
<td>0.1563</td>
<td>1804.382</td>
<td>0.1563</td>
</tr>
</tbody>
</table>
3.5.1 HPLC dioscin standard curve

The standard dioscin was dissolved in methanol and the respective dilutions made. The standard curves had 8 points with 0.15 - 20ng total dioscin content. The HPLC was run at a ratio of 90:10 Acetonitrile to water. Eight dioscin concentrations were used in the calibration curve this are indicated with the corresponding response factor which is area in mVs (Table 8). Four of the peaks for the dioscin standard run are shared in this thesis.

Figure 8 (appendix) indicates the resulting peak for 20ng of dioscin loaded. The retention time for dioscin was detected at 2.3± 0.3 minutes. The area under the dioscin peak was 166383.494mVs.

The correlation between the dioscin quantity loaded and the area under the peak for that quantity were found to be directly proportional. When the amount of dioscin loaded reduced the resulting peak height and area were lower. Figure 21 indicates the peak height obtained when 15ng of dioscin were loaded. Comparing the height obtained in Figure 20 to that in Figure 21, the peak height is noticed to have declined marginally. The peak height in Figure 20 was 1.6 Vs whereas that in Figure 21 is above 600mVs. Equally the area under the peak in Figure 21 is slightly above double that in Figure 20.

Dioscin concentration of 0.5ng/μl resulted in a peak area of 44534.034mVs (Table 8) on the other hand, the peak height was 500mV (Figure 25, appendix). This is comparable to the other peak heights and areas in Figures 20 and 21. Comparison of peak height in Figure 21 and Figure 22 indicates that Figure 21's is higher. This disparity in height is attributed to the variable dioscin quantities loaded. The retention time was found to be consistent throughout all the runs and with a variation of ±0.2 minutes. Like the trend with the earlier concentrations, 0.25ng/μl run resulted in an even lower peak height and area. The resultant peak height and area was 220mV and 22520.285mV.s (Fig 23) respectively. The trend in the
proportionate reduction in peak area and height was clear in all the standard dilutions (Table 8).
Figure 20: Chromatogram for dioscin at a concentration of 1ng/μl
The areas recorded for each of the eight dioscin standard runs were plotted against the corresponding quantity; this resulted in a linear correlation (Fig. 24, appendix). Table 8 shows the results of the area under the curve for the eight dioscin concentrations. It is from the calibration curve obtained that the dioscin quantities for the sample were obtained if a peak was detected, this was done by the HPLC YLclarity® software from the equation \( y = 9012.0838981x \). A strong correlation factor \( R^2 = 0.9637 \) indicated a significance level of correlation between the area under the peak and dioscin quantities loaded.

![Graph showing calibration curve for 8 dioscin concentration standards]

Figure 21: A calibration curve for 8 dioscin concentration standards
3.5.2 Identification and quantification of dioscin from *Dioscorea* samples

HPLC run of the dioscin-rich extracts resulted in identification of variable but minimal dioscin quantities. Dioscin in the sample was identifiable by having the same retention time ($T_r$) as that obtained from the standards (2.3±0.2 minutes). Dioscin was identified in all the samples screened, by having same retention time with that obtained from the standard. The total quantity of dioscin for the 20ul loaded varied with the lowest sample having 0.088ng/20ul and the highest with 0.512ng in 20ul (Table 9). To identify the sample concentration per micro liter, the concentration obtained in 20ul was divided by 20 (Table 9).

The initial weight of the sample for extraction was 5g; the extract from this had been dissolved in 1ml of methanol. To obtain the total dioscin yield from the sample, the amount in ng/ul was multiplied by 1000 (Table 9). From this calculation the resultant values indicated that the sample with the highest dioscin quantity was Dio74 with 25.6ng. The sample that had the least amount was Dio71 with 4.42ng (Table 9). The species names identified for each of the samples through phylogenetic analysis are also indicated in the Table 9. There was no species-specific trend in dioscin identified from the samples screened. This is because samples identified to be within the species *D. cayanensis* were found to have higher dioscin quantities than those from different species example *D. alata*. Figure 25 indicates the HPLC run for the sample KDio71 which is in *D. cayanensis* species.

The peak number 5 in Figure 25 (Appendix) resulted after running dioscin-rich extract from sample KDio71. This sample had a dioscin peak area of 796.491mVs. This peak was obtained at a retention time of exactly 2.3 minutes. The corresponding dioscin amount for this sample was determined to be 4.419 parts per billion (ppb). HPLC run for sample KDio58 resulted in dioscin being detected at peak number 3 at a retention time ($T_r$) of 2.3 minutes. The dioscin peak area of 3219.68 mVs identified for KDio58 was correlating to a dioscin quantity.
of 17.864 ppb (Fig. 26). The sample KDio74 resulted in dioscin peak number 4 with an area of 4613.53mVs corresponding to a quantity of 25.5975 ppb. Samples KDio60, KDio82, KDio77, KDio72, KDio56 and KDio22 resulted in peak areas 1381.34, 3269.87, 3123.36, 2240.22, 3489.82 and 2588.75mVs with corresponding dioscin quantities of 7.664, 18.1405, 17.3295, 12.4295, 19.3625, and 14.3635 ppb respectively (Table 9).
<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Local name</th>
<th>species phylogeny</th>
<th>Area under curve (mV)</th>
<th>Experiment 2</th>
<th>Experiment 3</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Standard error</th>
<th>Total dioscin loaded (Quantity in 20μl)</th>
<th>Concentration of dioscin (ng/μl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dio78</td>
<td>Gaceru</td>
<td>D. cayanensis</td>
<td>2813.65</td>
<td>2830.75</td>
<td>2841.98</td>
<td>2828.793333</td>
<td>14.2659957</td>
<td>±3.77703530574186</td>
<td>0.31222</td>
<td>0.015611</td>
</tr>
<tr>
<td>Dio74</td>
<td>Gakwa</td>
<td>D. cayanensis</td>
<td>4613.53</td>
<td>4628.32</td>
<td>4639.55</td>
<td>4627.133333</td>
<td>13.05052617</td>
<td>±3.61255119977927</td>
<td>0.51195</td>
<td>0.0255975</td>
</tr>
<tr>
<td>Dio60</td>
<td>Karuguaci</td>
<td>D. alata</td>
<td>1381.34</td>
<td>1395.15</td>
<td>1406.38</td>
<td>1394.29</td>
<td>12.54213299</td>
<td>±3.54148739831958</td>
<td>0.15328</td>
<td>0.007664</td>
</tr>
<tr>
<td>Dio62</td>
<td>Mbeu</td>
<td>D. cayanensis</td>
<td>1460.32</td>
<td>1467.11</td>
<td>1478.34</td>
<td>1468.59</td>
<td>9.100708764</td>
<td>±3.0167380994081</td>
<td>0.16205</td>
<td>0.0081025</td>
</tr>
<tr>
<td>Dio62</td>
<td>Mbeu</td>
<td>D. cayanensis</td>
<td>1156.39</td>
<td>1160</td>
<td>1171.23</td>
<td>1162.54</td>
<td>7.739192464</td>
<td>±2.78194041351068</td>
<td>0.12832</td>
<td>0.006416</td>
</tr>
<tr>
<td>Dio72</td>
<td>Mucara</td>
<td>D. cayanensis</td>
<td>2240.22</td>
<td>2260.32</td>
<td>2271.55</td>
<td>2257.363333</td>
<td>15.87288989</td>
<td>±3.98407955317024</td>
<td>0.24859</td>
<td>0.0124295</td>
</tr>
<tr>
<td>Dio75</td>
<td>Mucharage</td>
<td>D. cayanensis</td>
<td>2819.24</td>
<td>2821.92</td>
<td>2841.24</td>
<td>2821.92</td>
<td>5.725058915</td>
<td>±2.39270954178282</td>
<td>0.31298</td>
<td>0.015649</td>
</tr>
<tr>
<td>Dio77</td>
<td>Muhika</td>
<td>D. cayanensis</td>
<td>3125.403333</td>
<td>3125.403333</td>
<td>3125.403333</td>
<td>3125.403333</td>
<td>5.887243271</td>
<td>±2.42636420831566</td>
<td>0.34659</td>
<td>0.0173295</td>
</tr>
<tr>
<td>Dio56</td>
<td>Muntu</td>
<td>D. cayanensis</td>
<td>3498.563333</td>
<td>3498.563333</td>
<td>3498.563333</td>
<td>3498.563333</td>
<td>9.426697902</td>
<td>±3.07029280392581</td>
<td>0.38725</td>
<td>0.0193625</td>
</tr>
<tr>
<td>Dio58</td>
<td>Ndamarari</td>
<td>D. cayanensis</td>
<td>3229.21</td>
<td>3239.59</td>
<td>3239.59</td>
<td>3239.59</td>
<td>9.982179121</td>
<td>±3.15945867526351</td>
<td>0.35728</td>
<td>0.017864</td>
</tr>
<tr>
<td>Dio58</td>
<td>Ndiathi</td>
<td>D. cayanensis</td>
<td>2040.57</td>
<td>2045.34</td>
<td>2045.34</td>
<td>2040.57</td>
<td>5.796829593</td>
<td>±2.40766060591873</td>
<td>0.2267</td>
<td>0.01131</td>
</tr>
<tr>
<td>Dio71</td>
<td>Ngandau</td>
<td>D. cayanensis</td>
<td>832.45</td>
<td>827.203333</td>
<td>832.45</td>
<td>827.203333</td>
<td>18.99564731</td>
<td>±2.89016627125254</td>
<td>0.08858</td>
<td>0.0064419</td>
</tr>
<tr>
<td>Dio22</td>
<td>Nigerin</td>
<td>D. cayanensis</td>
<td>2600.82</td>
<td>2592.92</td>
<td>2592.92</td>
<td>2592.92</td>
<td>6.676062091</td>
<td>±2.58889040927242</td>
<td>0.21772</td>
<td>0.01435</td>
</tr>
<tr>
<td>Dio22</td>
<td>Ngamukus</td>
<td>D. cayanensis</td>
<td>2504.57</td>
<td>2504.57</td>
<td>2504.57</td>
<td>2504.57</td>
<td>2504.57</td>
<td>±2.54184739842593</td>
<td>0.21772</td>
<td>0.01435</td>
</tr>
<tr>
<td>Dio67</td>
<td>Nguni</td>
<td>D. cayanensis</td>
<td>3668.556667</td>
<td>3668.556667</td>
<td>3668.556667</td>
<td>3668.556667</td>
<td>7.683792297</td>
<td>±2.77196506653555</td>
<td>0.40864</td>
<td>0.020432</td>
</tr>
</tbody>
</table>
Figure 22: HPLC run for dioscin extract for Kenyan *Dioscorea* sample KDio71 detected at 265nm, mobile phase; Methanol: Water: Acetic acid (70:30:0.2).
4.0 Discussion

4.1 Variability in *Dioscorea* tuber morphologies

Based on all the markers, the Kenyan samples Kmkk, Kmbeum, KdioM and KlbbuA (Figure 4) were found to be closely related to the species *Dioscorea cayanensis* despite the great morphological variability in the tuber shapes. Traders in the market were able to differentiate the *Dioscorea* samples collected based on morphological characters such as twining mode, tuber development (deep or shallow), tuber taste when cooked, leaf morphology and sizes, availability of thorn and their sizes, sizes of vines, emergence of vines from the corm and presence or absence of hairs on tubers. An earlier check in the University of Nairobi herbarium resulted little information on the species *Dioscorea*, this is why the information the traders gave was vital in this work.

Confirmation based on the International Plant Genetic Resources Institute (IPGRI) was only possible based on tubers as this was the only morphological part accessible from the markets. Based on the IPGRI standard yam descriptor list for *Dioscorea* lists the following tuber characteristics in species description: tuber shape which may be oval, oval-oblong, cylindrical, flattened and irregular) tuber branching (slightly or highly branched) and point of branching (upper third, middle third or lower third) (IPGR/ITA, 1997). The use of tuber morphology cannot be independently used in the species taxonomy placement in yams, other morphological characteristics have to be included too. Root characteristics and pollen morphology have been used in the systematic studies of African *Dioscorea* species (Burkett, 1960 and Wilkin, 1999). Based on these characteristics, African yam species were taxonomically placed into different sections. Key among these, the section emmimphylum.

From this study, using three molecular markers the tuber pictured in Figure 4 were found to be of the same species, *Dioscorea cayanensis*. 

59
4.2 Molecular phylogeny of genebank sequences

The diversity observed in the *Dioscorea* species-specific markers from the gene bank was used in the discrimination of both the known African species and the unknown Kenyan samples into the right position. The tendency of the markers *rbcL* and *matK* to portray more diversity relative to the marker *trnF*-L can be explained by longer fragment lengths on sequencing averaging 800 clean basepair. Moreover, the marker *matK* was observed to have indels that were giving trouble during alignment. The percentage DNA sequence identity was varying with the marker and the species. It ranged between 45 - 100% for the species with the least similarity and that with perfect match respectively.

The variable percentage sequence identity is attributable to the multiple substitutions observed within the gene bank sequences for the different primers. For the marker *matK*, substitutions from C to T were the highest totaling 139; substitutions from G to C were the least in number with only 5 substitutions (Table 5). Based on the marker *rbcL*, the highest substitutions were from A to G and from G to A, with both having 166, the lowest number of substitutions were observed to be from G to C. The total number of nucleotide substitutions for *rbcL* was 1240, 597 for *matK* and 295 for *trnF*-L (Table 5) hence making the marker *trnF*-L the least informative and *rbcL* the most informative based on the nucleotide substitutions.

Transversions are substitution of a purine by a pyrimidine or vice versa, while a transition is the substitution of a pyrimidine by a pyrimidine or purine by a purine (Tamura, 2011). In this study, the marker *matK* was observed to have a higher rate of transitions than there are transversions. The number of transitions was also higher in the marker *rbcL*. The ratio of transitions to transversions are an indicator to the species evolutionary relationship (Tamura, 2011) and this explains why the marker *trnF*-L was said to be less informative this is because the higher the ratio, the higher the informativeness the marker *trnF*-L had the lowest ratio.
and was hence less informative (Table 5). There were some species based on the sequences from the genebank that had different species names but with 100% sequence similarity species.

Based on the marker rbcL, the species D. persimilis had 100% sequence identity to the species D. aspersa, D. tookoro and D. tenuiipes also had 100% sequence identity. With the marker matK, the species D. karatana and D. maciba, D. karatana and D. ovinala had 100% sequence similarity. Examples from the marker trnF L species species are D. abyssinica and D. liebrechtsiana. The species that shared 100% sequence similarity but had different species names is attributed to poor quality sequences deposited in the gene bank or possibly the same species being labeled as a different species due to morphological plasticity. The variability in these markers could be able to differentiate the species in question with the few variable nucleotides observed. This is consistent with other researchers who have used some of these markers in similar studies (Asahina, 2010, Wang et al., 1999, Douglas, et al., 1996) it was therefore feasible to go ahead with the three markers in species phylogeny. Moreover, there were some species that had different species names but matched with those sharing a different species name.

Based on the marker rbcL for example, the Species D. cirrhosa matched to the species D. polystachya, D. alatipes to D. pentaphylla, D. henzhuana to D. rocuti (Figure 9). The two species D. cirrhosa and D. polystachya just like in the marker rbcL, again matched in the marker matK, this shows some consistency in these two species. Species D. longibza and D. macvaughii, D. alata and D. exalata, D. persimilis and D. hamiltonii also matched to each other respectively in the marker matK (Figure 8). For the marker trnF L, there were not many species that had unspecific matches with only D. cayanensis matching to D. praehensilis, D. abyssinica to D. libbrechtschiana and D. burkilliana matching to D. minutthora (Figure 10).
The unspecific species matching can be attributed to one, presence of species which are more closely related with only up to one base pair difference in the nucleotide sequence in use, this means that the marker in use is quite conserved. Second it can also be attributed on the extreme to the species being incorrect and hence the same species given different names. The species name in such a case may be either the two. Alternatively, the two species may be the same but were classified as different possibly due to the general polymorphism in taxonomic characters within these species. The lack of perfect match in some of the species is a limitation in use of these markers.

The markers matK and rbcL were more conserved than the marker trnF_L, these markers showed more diversity within the gene bank species. In general however, the variability in these markers could be able to differentiate up to 95% of the species in question. This is consistent with other researchers who have used some of these markers in such kind of study (Asahina, 2010, Wang et. al., 1999, Douglas et al.).

After confirmation on the number of species-specific sequences for the three markers in the gene bank, the known species - specific sequences (IITA and Madagascar) were tested against the gene bank species to confirm on the usability of the markers in positioning of the species. The aim was to observe whether the reference sequences for all the three barcodes could be identified in comparison with the gene bank sequences. The species that did not have sequences in the gene bank for the respective markers included D. minutiflora, D. cayanensis, D. Mangenotiana, D. praehebeslis, D. rotundata and D. sansibolensis for the marker rbcL. D. minutiflora, D. cayanensis, D. mangenotiana and D praehebeslis were among the species where sequences were obtained using matK. Others were D. tena, D. sosj and D. dumetorum for the marker trnF_L (Fig. 13). Some of them were troublesome in PCR or sequencing and this was treated as missing data. There were new sequences to be deposited in the gene bank and these include rbcL sequences for D. rotundata, matK
sequences for *D. antaly* and *D. praehensilis*, *trnF L* sequences for *D. yassa*, *D. dumetorum* and *D. praehensilis*.

The findings of this study on the economically important *D. alata* are in consistency with studies done by (Wilkin, 2005) which observed that this species was not in the same clade with *D. hamiltonii* and *D. brevipetirolata*. The segregation of the different species in this study is generally in line with established taxonomic segregation of the *Dioscorea* species. The major cultivated yams belong to the yam section enantiophylum characterized by vines twining to the right, this section is of great taxonomic importance and include but not limited to the species *D. cayanensis*, *D. rotundata*, *D. dumetorum* and *D. alata*. The other cultivated yam species *D. hulbifera* is in the section Opsophyton and is a distinct species in the genus in that it produces aerial tubers and has two origins, Asia and Africa.

The species *D. cayanensis* and *D. rotundata* are similar than they are to *D. hulbifera*, this is in line with observations made by Tamiru et al (2007) on *Dioscorea* diversity based on AFLP markers. Kenyan *Dioscorea* species both cultivated and wild have been a subject of speculation, Maundu et al (1999) indicated that the diversity of cultivated yam species in Kenya comprise of the species *D. hulbifera*, *D. minutifolia* and *D. dumetorum* with the *D. minutiflora* being the major species in the central highlands. The close relationship of *D. rotundata* and *D. cayanensis* has resulted in the name *D. cayanensis/rotundata* complex (Mignouna et al, 2002).

An interesting observation in this analysis is the clade with the sample KW10 and the species *D. schimperiana*, these two are supported with a BP of 63 with the marker *matK*. *D. schimperiana* is a yam species of African origin classified under the section Asterotrichina (Burkill, 1939). This species is in cultivation in countries like Cameroon but is being neglected by farmers and is facing extinction (Leng, 2011). This study reveals a Kenyan cultivated variety closely related to the species *D. schimperiana*. Using molecular techniques.
**D** schimperiana has been shown to form sister clade to others of the section enantiophyllum (Wilkin et al., 2005). This finding was consistent although with a weak BP support. **D** schimperiana has been categorized as one of the many that contain apogenin and its probable geographical location was Uganda and Ethiopia (Abyssinia) (Franklin, 1969, Burkill, 1939). The analysis indicates that Kenyan Dioscorea species are not limited to the earlier thought species. **D**. dumetorum had a strong BP support to **D** hispida this can be explained since they are both entire leafed species and have the compound leafed **D**. *antaly* as a sister species an observation consistent with that made by Wilkin et al (2005).

Based on the three different markers used in this study, there was general congruence both at the regional and geographical level. This is in agreement with other studies by Mwirigi et al., 2009 and Wilkin et al., 2005 with only a few disagreements. For the marker *rbcl*, congruence was observed at the section taxon. In relation to the study of Wilkin et al who had used morphological characters like tuber morphology, twining mode and pollen size and shape congruence was observed. Wilkin et al., (2005) observed clades for the different Dioscorea species into section stenophora which is a purely rhizomatous and is sister to other Dioscorea species. This section included the species **D**. rockii, **D**. prazeri among others. There was also the section Enantiophyllum which is composed Dioscorea species with right twining stem. These include **D**. hamiltonii, **D**. alata and **D**. oryzetorum. The group with compound leaves (CL) was observed by Wilkin to include **D**. dumetorum, **D**. bulbifera, **D**. *antaly* (Wilkin, 2005). The sub-clade Malagasy was observed to have species like **D**. *swno*, **D**. *humarivensis*, and **D**. *fandra* among others. Wilkin et al (2005) separated world yams geographically linking them to the molecular data, these were based on what he called clades which are clusters of similar taxons in a tree. There was also the clade Birmanica comprising of species like **D**. birmanica and **D**. petelotii. The four subclades Malagasy, CL, birmanica and enantiophyllum together formed a bigger clade which Wilkin referred to as clade II which are
all from Africa. Wilkin further ordered the *Dioscorea* species into another clade with three subclades which had strong bootstrap percentages. The subclades included new world (NW) Southern/montane African subclade and European clade with only a few species (Wilkin, 2005). Other clades observed by Wilkin were the Burmanniaceae and the Hermaphrodite *Dioscorea*. The results observed in this study were in consistency with the work of Wilkin across all the markers used with slight differences. In this study, the marker *rubt* was able to differentiate most of the clades as they had been observed by Wilkin. The only incongruence between this work and Wilkin’s was with the Entatiophylum subclade, some of the species that Wilkin had observed to fall into this subclade (enantiophylum) were segregating differently in this study, an example of these are the species *D. lanata*, *D. hemslawi* and *D. schimperiana* which were observed not to segregate with the rest of the *Dioscorea* species to the subclade enantiophylum. Based on the marker *matK* in this study, four of the subclades were in congruence with the observations that had made concerning the Malagasy, African, enantiophylla, *stenophora* and NW subclades. Some species that Wilkin had observed as being in the same subclades differed in this study. The species *D. alata*, *D. elevata* and *D. sylvatica* in the enantiophyllum subclade exemplifies these. Others which did not conform to Wilkin’s in this study were *D. rotundata*, *D. mangenotiana* and *D. cayamensis* in the African subclade. The results indicate consistency with the revised *Dioscorea* classification this further boosts the usability of the Molecular approach used in this study.

### 4.4 Identification and quantification of dioscin

Through this research, Kenyan *Dioscorea* was found to contain dioscin. The hypothesis is therefore rejected that Kenyan yams contain dioscin. The sample that had the least amount of dioscin as observed in this study was Dio71 with 0.00000442mg. It is however interesting to note that through the molecular approach of species positioning, the sample with the least and that with the highest amount of dioscin were actually found to be closely associated with the
same species (Dio 74 and Dio 71 respectively) *D. cuyanensis*. Studies by Friedman (2006) established that following harvest, the secondary metabolite content in tubers can increase or decrease during storage and transportation and under the influence of light, heat, cutting, slicing, sprouting, and exposure to phytopathogens. The samples Dio 74 and Dio 71 had been collected from different geographical areas. There can be several explanations for the observed variability of the dioscin amounts within these two samples that had been observed to be of the same species. The first explanation is that the variability in the amount of dioscin may have been as a result of environmental effects since metabolites are known to vary with environmental conditions (Kirsten *et al.*, 2010). In this study however, the samples were collected from different localities and therefore did not have homogeneity in their resource environmental conditions. The environmental conditions in which the sample Dio74 grew may have promoted higher dioscin quantities while the conditions in which the sample Dio71 grew may have not been favorable for high dioscin quantities, since these were collected from the markets, their exact locality can not be quoted with authority. The second explanation for this disparity may be that the physiological conditions of these two samples were not the same during the extraction (*Rüdiger et al.*, 2004). One tuber may have been harvested longer before the other and this can also cause the differences in the amounts of dioscin observed.

### 4.5 Implications of the study and future prospects for Kenyan *Dioscorea*

There is continued decline in yam cultivation in Kenya according to the Food and Agriculture Organization statistics (FAO, 2009). The establishment that there is more yam species diversity in Kenyan than earlier thought is positive gesture. This is important to both yam conservation and breeding. Although the chloroplast markers used in this study could establish species specific clusters and at times intraspecific variability, a more robust approach is preferable in discern intraspecific variability, example the use of whole genome
in phylogenetics. This method has been successfully used in rice (Catherine et al., 2011). Such approaches are robust enough to identify both inter- and intra-specific diversity in Kenyan yams and other important crops. Future studies of Kenyan yam diversity should include wild yams that were not covered in this study.
CHAPTER FIVE

5.0 Conclusions

The approach used here is applicable to both species phylogeny and species phylogeny based on sequence information existing in the gene bank. Species phylogeny can be applicable in the taxonomy of morphologically plastic taxa like the *Dioscorea* species. Morphologically plastic taxa can result in a single species being taxonomically placed into different or two varieties to be identified as being different species. The approach used here is applicable for both species phylogeny and phylogenetic with enough sequence information existing in the gene bank. The use of many markers increases reliability. The work carried out here attests that it is possible to authenticate species when sequences are available in gene banks for the marker in use. Through this research, the status of cultivated *Dioscorea* species in Kenya especially in the central highlands was established, this is however a tip in the iceberg and there is need for further sample collection across the country with the inclusion of wild *Dioscorea* too. The use of more markers (nuclear and mitochondrial) would also be suggested in future studies to further enhance on the species identification.

This approach can therefore be used to differentiate the taxa under dispute to a single species or different species. The approach in species phylogeny is applicable in a situation where time and resources are limited to carry out morphological survey of a taxa or when the morphological species information for a geographical locality is unknown. The approach in species phylogeny can also apply in situations where there is expectation of more in terms of the taxa in question. The first step in the DNA phylogeny or taxon confirmation is gathering of information on species in question, collection of samples, extraction of DNA, PCR using the three gene regions and sequencing of the PCR products. Sequences are then aligned against what is in the gene bank. Comparison with the gene bank information will give an insight to the species information in terms of species phylogeny or species phylogeny.
5.1 Recommendations

From this research it would be recommended that further intensive studies on dioscin be through collection of more diverse plant materials including wild Dioscorea which exist in Kenya. Further work should be done to establish the bioactivity extent of dioscin. It is also recommended that metabolic profiling be done in future on Dioscorea tubers so that an array of compounds that can give a broader picture of the Dioscorea metabolite capability can be identified. Such metabolite profiling can involve a method like mass spectrophotometry (MS) in combination with other metabolite profiling approaches like Near Infra-red spectroscopy (NIR). Future research involving integration of metabolite and genetic aspects should be carried out with corresponding databases to act as reference for future studies. It is worthwhile to note that more study is essential especially to the African biodiversity at the genetic level. This is because a lot of African biodiversity whether cultivated or wild is not recognized at the gene level. A good example is the Dioscorea species Dioscorea alata despite being one of the top cultivated crops and Africa being the leading in its own production, there was information lacking for this species.
REFERENCES


Jacobo Arango, Bertha Salazar, Ralf Welsch, Felipe Sarmiento, Peter Beyer, Salim Al-Babili (2010) Putative storage root specific promoters from cassava and Dioscorea cloning and evaluation in transgenic carrots as a model system *Plant Cell Rep* 29:651-659

Jacobo Arango, Bertha Salazar, Ralf Welsch, Felipe Sarmiento, Peter Beyer, Salim Al-Babili (2010) Putative storage root specific promoters from cassava and Dioscorea cloning and evaluation in transgenic carrots as a model system *Plant Cell Rep* 29:651-659

Jane Hubert, Monique Berger,, and, Jean Daydé (2005) Use of a Simplified HPLC-UV: Analysis for Soyasaponin B Determination: Study of Saponin and Isoflavone Variability in Soybean Cultivars and Soy-Based Health Food Products *Journal of agricultural and food chemistry* 200553 (10), 3923-3930


72


Pei-Fen Su , Chin-Jin Li , Chih-Chien Hsu , Spencer Benson , Sheng-Yang Wang , Kandan Aravindaram , Sunney I. Chan , Shih-IHung Wu , Feng-Ling Yang , Wen-Ching Huang.


Shu-Ling Fu, Ya-Hui Hsu, Pei-Yeh Lee, Wen-Chi Hou, Ling-Chien Hung, Chao-Hsiung Lin, Chiu-Ming Chen and Yu-Jing Huang (2006): Dioscorin isolated from Dioscorea alata activates TLR4-signaling pathways and induces cytokine expression in macrophages. Biochemical and Biophysical Research Communications Volume 339, Issue 1, Pg 137-144


Yu-Hsiu Liao, Chin-Yin Tseng, and Wenlung Chen (2006), Structural characterization of dioscorin, the major tuber protein of Dioscorea, by near infrared Raman spectroscopy IOP Publishing Ltd


APPENDIX

1.0 NCBI Dioscorea species sequences

<table>
<thead>
<tr>
<th>rbcL</th>
<th>matK</th>
<th>trnF_L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dioscorea sp. Newman</td>
<td>Dioscorea alatipes</td>
<td>Dioscorea liebrechtsiana</td>
</tr>
<tr>
<td>200110120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dioscorea sp. SH-2010</td>
<td>Dioscorea althaeoides</td>
<td>Dioscorea minutiflora</td>
</tr>
<tr>
<td>Dioscorea alata</td>
<td>Dioscorea amoonensis</td>
<td>Dioscorea rotundata</td>
</tr>
<tr>
<td>Dioscorea alatipes</td>
<td>Dioscorea aratana</td>
<td>Dioscorea trifida</td>
</tr>
<tr>
<td>Dioscorea althaeoides</td>
<td>Dioscorea aspersa</td>
<td>Dioscorea abyssinica</td>
</tr>
<tr>
<td>Dioscorea antaly</td>
<td>Dioscorea banzhuanal Dioscorea burkilliana</td>
<td></td>
</tr>
<tr>
<td>Dioscorea arachidna</td>
<td>Dioscorea bulbifera</td>
<td>Dioscorea smilacifolia</td>
</tr>
<tr>
<td>Dioscorea arcuatinervis</td>
<td>Dioscorea caucasica</td>
<td>Dioscorea togoensis</td>
</tr>
<tr>
<td>Dioscorea aspersa</td>
<td>Dioscorea cf. adenocarpa</td>
<td>Dioscorea praehensilis</td>
</tr>
<tr>
<td>Dioscorea banzhuanal</td>
<td>Dioscorea cirrhosa</td>
<td>Dioscorea alata</td>
</tr>
<tr>
<td>Dioscorea bimarivensis</td>
<td>Dioscorea cirrhosa var. cylindrica</td>
<td>Dioscorea althaeoides</td>
</tr>
<tr>
<td>Dioscorea biformifolia</td>
<td>Dioscorea collettii</td>
<td>Dioscorea aspersa</td>
</tr>
<tr>
<td>Dioscorea birmanica</td>
<td>Dioscorea collettii var. hypoglaucar</td>
<td>Dioscorea bimifolia</td>
</tr>
<tr>
<td>Dioscorea brachybotrya</td>
<td>Dioscorea communis</td>
<td>Dioscorea bulbifera</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Dioscorea brevipetiolata</td>
<td>Dioscorea composita</td>
<td>Dioscorea caucasica</td>
</tr>
<tr>
<td>Dioscorea buchananii</td>
<td>Dioscorea cotinifolia</td>
<td>Dioscorea cayenensis</td>
</tr>
<tr>
<td>Dioscorea bulbifera</td>
<td>Dioscorea decipiens</td>
<td>Dioscorea cirrhosa</td>
</tr>
<tr>
<td>Dioscorea caucasica</td>
<td>Dioscorea deltoidea</td>
<td>Dioscorea cirrhosa var cylindrica</td>
</tr>
<tr>
<td>Dioscorea cirrhosa</td>
<td>Dioscorea esquirii</td>
<td>Dioscorea colletii var hypoglauc</td>
</tr>
<tr>
<td>Dioscorea colletii</td>
<td>Dioscorea exalata</td>
<td>Dioscorea communis</td>
</tr>
<tr>
<td>Dioscorea communis</td>
<td>Dioscorea futschauensis</td>
<td>Dioscorea cotinifolia</td>
</tr>
<tr>
<td>Dioscorea composita</td>
<td>Dioscorea glomerulata</td>
<td>Dioscorea decipiens</td>
</tr>
<tr>
<td>Dioscorea cotinifolia</td>
<td>Dioscorea daunca</td>
<td>Dioscorea deltoidea</td>
</tr>
<tr>
<td>Dioscorea daunca</td>
<td>Dioscorea gracillima</td>
<td>Dioscorea deltoidea</td>
</tr>
<tr>
<td>Dioscorea decipiens</td>
<td>Dioscorea hamiltonii</td>
<td>Dioscorea deltoidea</td>
</tr>
<tr>
<td>Dioscorea delavayi</td>
<td>Dioscorea hemsleyi</td>
<td>Dioscorea esquirli</td>
</tr>
<tr>
<td>Dioscorea deltoidea</td>
<td>Dioscorea hexagona</td>
<td>Dioscorea esquirli</td>
</tr>
<tr>
<td>Dioscorea dunetorum</td>
<td>Dioscorea hispida</td>
<td>Dioscorea exalata</td>
</tr>
<tr>
<td>Dioscorea elephantipes</td>
<td>Dioscorea inopinata</td>
<td>Dioscorea futschauensis</td>
</tr>
<tr>
<td>Dioscorea esquirli</td>
<td>Dioscorea japonica</td>
<td>Dioscorea futschauensis</td>
</tr>
<tr>
<td>Dioscorea fandra</td>
<td>Dioscorea lanata</td>
<td>Dioscorea glabra</td>
</tr>
<tr>
<td>Dioscorea fandra</td>
<td>Dioscorea longirhiza</td>
<td>Dioscorea gracillima</td>
</tr>
<tr>
<td>Dioscorea fordii</td>
<td>Dioscorea maciba</td>
<td>Dioscorea hispida</td>
</tr>
<tr>
<td>Dioscorea futschauensis</td>
<td>Dioscorea mcvaughii</td>
<td>Dioscorea japonica</td>
</tr>
<tr>
<td>Dioscorea galeottiana</td>
<td>Dioscorea melanophyma</td>
<td>Dioscorea melanophyma</td>
</tr>
<tr>
<td>Dioscorea gigielli</td>
<td>Dioscorea membranacea</td>
<td>Dioscorea nipponica</td>
</tr>
<tr>
<td>Dioscorea glabra</td>
<td>Dioscorea namoro ensis</td>
<td>Dioscorea nipponica subs</td>
</tr>
<tr>
<td>Dioscorea gracillima</td>
<td>Dioscorea nipponica</td>
<td>Dioscorea nitens</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Dioscorea hamiltonii</td>
<td>Dioscorea nipponica</td>
<td>Dioscorea pentaphylla</td>
</tr>
<tr>
<td></td>
<td>subsp. rosthornii</td>
<td></td>
</tr>
<tr>
<td>Dioscorea hemsleyi</td>
<td>Dioscorea nitens</td>
<td>Dioscorea persimilis</td>
</tr>
<tr>
<td>Dioscorea hexagona</td>
<td>Dioscorea oppositifolia</td>
<td>Dioscorea polystachya</td>
</tr>
<tr>
<td>Dioscorea hispida</td>
<td>Dioscorea oryzerorum</td>
<td>Dioscorea sansibarensis</td>
</tr>
<tr>
<td>Dioscorea inopinata</td>
<td>Dioscorea ovinala</td>
<td>Dioscorea simulans</td>
</tr>
<tr>
<td>Dioscorea japonica</td>
<td>Dioscorea pentaphylla</td>
<td>Dioscorea sinoparviflora</td>
</tr>
<tr>
<td>Dioscorea kamooneensis</td>
<td>Dioscorea persimilis</td>
<td>Dioscorea sp Newman</td>
</tr>
<tr>
<td>Dioscorea lanata ,</td>
<td>Dioscorea petelotii</td>
<td>Dioscorea spongiosa</td>
</tr>
<tr>
<td>Dioscorea maciba</td>
<td>Dioscorea polystachya</td>
<td>Dioscorea subcalva</td>
</tr>
<tr>
<td>Dioscorea mcvaughii ,</td>
<td>Dioscorea prazeri</td>
<td>Dioscorea tokoro</td>
</tr>
<tr>
<td>Dioscorea melanophyma</td>
<td>Dioscorea pseudoniens</td>
<td>Dioscorea zingiberensis</td>
</tr>
<tr>
<td>Dioscorea membranacea</td>
<td>Dioscorea roci</td>
<td>Dioscorea banzhuana</td>
</tr>
<tr>
<td>Dioscorea nipponica</td>
<td>Dioscorea sansibarensis</td>
<td>Dioscorea colletti</td>
</tr>
<tr>
<td>subsp. rosthornii</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dioscorea nitens</td>
<td>Dioscorea schimperiana</td>
<td>Dioscorea kamooneensis</td>
</tr>
<tr>
<td>Dioscorea oppositifolia</td>
<td>Dioscorea septemloha</td>
<td></td>
</tr>
<tr>
<td>voucher</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dioscorea oryzerorum</td>
<td>Dioscorea simulans</td>
<td></td>
</tr>
<tr>
<td>Dioscorea pentaphylla</td>
<td>Dioscorea sinoparviflora</td>
<td></td>
</tr>
<tr>
<td>Dioscorea persimilis</td>
<td>Dioscorea soso</td>
<td></td>
</tr>
<tr>
<td>Dioscorea petelotii</td>
<td>Dioscorea spongiosa</td>
<td></td>
</tr>
<tr>
<td>Dioscorea polygonoides</td>
<td>Dioscorea subcalva</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------------</td>
<td></td>
</tr>
<tr>
<td>Dioscorea polystachya</td>
<td>Dioscorea subcalva var. submollis</td>
<td></td>
</tr>
<tr>
<td>Dioscorea prazeri</td>
<td>Dioscorea subhastata</td>
<td></td>
</tr>
<tr>
<td>Dioscorea pyrenaica</td>
<td>Dioscorea tentaculigera</td>
<td></td>
</tr>
<tr>
<td>Dioscorea schimperiana</td>
<td>Dioscorea tenuipes</td>
<td></td>
</tr>
<tr>
<td>Dioscorea septemloba</td>
<td>Dioscorea tokoro</td>
<td></td>
</tr>
<tr>
<td>Dioscorea simulans</td>
<td>Dioscorea trichantha</td>
<td></td>
</tr>
<tr>
<td>Dioscorea sinoparviflora</td>
<td>Dioscorea wallichii</td>
<td></td>
</tr>
<tr>
<td>Dioscorea spongiosa</td>
<td>Dioscorea zingiberensis</td>
<td></td>
</tr>
<tr>
<td>Dioscorea spp Kenya2</td>
<td>Dioscorea larecajensis</td>
<td></td>
</tr>
<tr>
<td>Dioscorea spp Kenya9</td>
<td>Dioscorea sylvatica</td>
<td></td>
</tr>
<tr>
<td>Dioscorea subcalva</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dioscorea subcalva var. submollis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dioscorea sylvatica</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dioscorea tenuipes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dioscorea tokoro</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dioscorea villosa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dioscorea</td>
<td>cf. apurimacensis PW-2004</td>
<td></td>
</tr>
</tbody>
</table>
2.0 HPLC dioscin standard concentration runs

Figure 23: Chromatogram for dioscin at 0.75ng/µl concentration
Figure 24: HPLC chromatogram for dioscin concentration of 0.5ng/ul.
Figure 25: HPLC run for dioscin at a concentration of 0.25ng/ul
3.0 Dioscin identification and quantification

Figure 26: HPLC results for the Kenyan *Dioscorea* sample KDio58 detected at 265nm, mobile phase; Methanol: Water: Acetic acid (70:30:0.2).
Figure 27: Chromatogram for Kenyan *Dioscorea* sample KDio74 detected at 265nm, mobile phase; Methanol: Water: Acetic acid (70:30:0.2).
Figure 28: HPLC chromatogram results for the Kenyan Dioscore asample KDio60 detected at 265nm, mobile phase; Methanol: Water: Acetic acid (70:30:0.2).
Figure 29: HPLC chromatogram for the Kenyan *Dioscorea* sample number KDio82 detected at 265nm, mobile phase; Methanol: Water: Acetic acid (70:30:0.2).
Figure 30: HPLC chromatogram result for Kenyan *Dioscorea* sample number KDio77 detected at 265nm, mobile phase; Methanol: Water: Acetic acid (70:30:0.2).
Figure 31: HPLC chromatogram for the Kenyan Dioscorea sample KDio72 detected at 265nm, mobile phase; Methanol: Water: Acetic acid (70:30:0.2).
Figure 32: HPLC dioscin run for the Kenyan *Dioscoreas* sample KDio56 detected at 265nm, mobile phase; Methanol: Water: Acetic acid (70:30:0.2).
Figure 33: HPLC chromatogram for the Kenyan *Dioscorea* sample KDio22 detected at 265nm, mobile phase: Methanol: Water: Acetic acid (70:30:0.2).