MOLECULAR AND MORPHOLOGICAL IDENTIFICATION OF PLANTS CONSUMED BY YELLOW BABOONS IN AMBOSELI, KENYA

RISPAH NYAMBURA NG'ANG'A

BSc. (Biology), University of Nairobi

I56/82866/2015

A thesis submitted in partial fulfilment of the requirement for the award of a degree in Master of Science in Genetics at the School of Biological Sciences in the University of Nairobi, Kenya.

2019

DECLARATION

This is my original work, which has never been presented for a degree in any other university or institution.

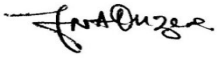
Rispah Nyambura Ng'ang'a,

Signature:  Date: 13- November -2019

School of Biological Sciences, University of Nairobi.

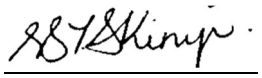
The thesis has been submitted with the approval of my supervisors:

Dr. Nelson O. Amugune,

Signature:  Date: 13- November -2019

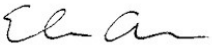
School of Biological Sciences, University of Nairobi,

Dr. Mercy Y. Akinyi,

Signature:  Date: 13- November -2019

Department of Animal Science, Institute of Primate Research,

Professor Elizabeth A. Archie,

Signature:  Date: 13- November -2019

Department of Biological Sciences, University of Notre Dame, USA.

DEDICATION

To my parents Silas and Gladwell Ng'ang'a; siblings Rahab Mwihaki, Simon Gachoya, Meshack Mwangi, and James Gaitho; and all my dear friends. I am beyond grateful for your support and encouragement throughout this research period. May God bless you all!

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincerest gratitude to my supervisors: Dr. Nelson Amugune, Dr. Mercy Akinyi, and Professor Elizabeth Archie, for their guidance during the entire period of this study. I am eternally grateful for always being available for consultations, and could not have imagined having better advisors for my dissertation.

I am exceptionally grateful to the University of Nairobi for allowing me to pursue this Master's degree through a scholarship. This opportunity has provided a great deal of exposure.

I want to acknowledge the Amboseli Baboon Research Project team for allowing me to conduct my research with them. I am eternally grateful for this opportunity and financial support. I would also like to thank Professor Archie, Mr. Kinyua Warutere and Ms. Mauna Dasari, all of whom aided in the collection and preservation of the plant materials from the study site in Amboseli.

I would also like to thank the Institute of Primate Research (IPR), Kenya, for allowing me to conduct my research at their laboratories. Additionally, I am very grateful to the barcoding team at the Molecular Genetics Laboratory, situated at the National Museums of Kenya, for their assistance and advice.

Finally, I am exceedingly grateful to the Almighty God for giving me strength, guidance, and good health throughout this research period.

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND ACRONYMS

°C	degrees Celsius
µL	microliter
ABRP	Amboseli Baboon Research Project
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
BSA	Bovine serum albumin
CBOL	Consortium for the Barcode of Life
*CO*1	cytochrome c oxidase 1 gene
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphates
EDTA	Ethylene diamine tetra-acetic acid
g	grams
IPR	Institute of Primate Research
ITS	Internal transcribed spacer region
IUCN	International Union for Conservation of Nature
*mat*K	maturase K gene
mg	milligrams
$MgCl_2$	Magnesium chloride
min	minutes
mL	millilitre
mM	millimolar
NCBI	National Centre for Biotechnology Information
NEB	New England BioLabs
PCR	Polymerase chain reaction
*rbc*L	ribulose 1,5-bisphosphate carboxylase gene

RNA	Ribonucleic acid
rpm	revolutions per minute
TAE	Tris-acetate-EDTA
*trn*L (UAA) intron	Leucine transfer RNA intron
*trn*L-P6	P6 loop of the chloroplast *trn*L (UAA) intron
TRPNR	Tana River Primate National Reserve
v/v	volume per volume
w/v	weight per volume

ABSTRACT

Plant species can be identified based on their morphological characteristics and molecular properties. Molecular identification utilizes specific regions in the plant genome, followed by comparison against reference databases. The flora in the Amboseli ecosystem has only been characterized based on its morphology, and no genetic data exists in the public databases. The general objective of this research was to identify the plants consumed by yellow baboons (*Papio cycnocephalus*) in Amboseli, Kenya, based on morphological and molecular analyses.

Eighty plants (40 monocotyledons and 40 dicotyledons) were collected from Amboseli Baboon Research Project's study site in the Amboseli ecosystem, Kenya, in three periods, namely June 2016, January 2018 and May/June 2018. Twenty-three plants whose scientific names were uncertain were deposited at the University of Nairobi herbarium. DNA was extracted from all the samples using Qiagen's DNeasy Plant Mini Kit, followed by the amplification of five barcoding genes, namely: *ITS*1, *ITS*1-Poaceae, the *trn*L (UAA) intron, *trn*L-P6, and the 18S ribosomal DNA locus. The amplicons were sequenced using the Sanger sequencing method then analysed using BLAST and phylogenetic approaches to determine the consensus identities.

The amplification success rate of all the extracted DNA was generally higher in monocotyledons (93.06%) than in dicotyledons (87.67%). With regards to the candidate markers, the *ITS*1 locus had the highest amplification success rate (100%) followed by the *trn*L (UAA) intron (95%), *ITS*1-Poaceae (90%), the 18S rDNA locus (82.05%), and lastly, *trn*L-P6 (80.49%). Two-hundred and eighty-nine amplicons were sent for sequencing at Macrogen Netherlands (Europe), and high-quality sequences were generated for 182 samples. Generally, more plants were identified at both the genus and species levels using GenBank® than in the BOLD database. Furthermore, more monocotyledons were identified using BLAST analysis than dicotyledons, whereas the phylogenetic analysis was more successful in the identification of dicotyledons than monocotyledons. With regard to the consensus identities, 66 out of the 80 plants were identified. Specifically, 50 plants were determined to only the genus level, while 16 samples were distinguished to the genus and species levels.

The use of multiple markers - from both the nuclear and chloroplast regions - was very crucial in the overall high identification success rate achieved in this study. The data generated from this work can be used as a reference for future studies relating to the characterization of plants in the Amboseli ecosystem and by extension, in Kenya. Furthermore, because the selected plants are those that are eaten explicitly by the Amboseli baboons, the data will be used to conduct a diet metabarcoding study.

CHAPTER ONE: INTRODUCTION

1.1 Background to the study

The yellow baboons (*Papio cynocephalus*) are non-human primates that are broadly distributed across Eastern, Central, and Southern Africa (Altmann, 1974). They are found in a wide variety of habitats, mainly in savannahs, which have variable tree cover with a dominant grass cover (Altmann and Altmann, 1970; Altmann, 1974). Baboons live in complex, mixed-sex social groups (Altmann and Altmann, 1970). They are omnivorous, but their diet is predominantly composed of plant material, including fruits, leaves, roots, and seeds. Besides, baboons also consume insects such as grasshoppers, and vertebrates including hares, vervet monkeys, and young gazelles (Post *et al.*, 1980; Post, 1982). In general, yellow baboons are highly opportunistic foragers, that feed on any suitable items that they come across (Altmann and Altmann, 1970). The International Union for Conservation of Nature and Natural Resources (IUCN) has classified the yellow baboons as "least concern species" on the Red List of Threatened Species (Kingdon *et al.*, 2008). This designation is because baboons are a widespread and common species and only a few threats could negatively affect their populations (Kingdon *et al.*, 2008). The plants that were used in this study were selected based on earlier direct observations of the food items consumed by yellow baboons living in the Amboseli ecosystem (Post *et al.*, 1980; Post, 1982; Altmann *et al.*, 1987; Altmann, 1998).

The identification of a plant refers to its assignment to a given taxonomic group (Hagedorn *et al.*, 2010; Hassoon *et al.*, 2018). The taxonomy of any plant can be determined based on either its morphological characteristics or molecular properties (Harris and Harris, 1994; Vijayan and Tsou, 2010; Wilson *et al.*, 2014; Purty and Chatterjee, 2016; Waldchen *et al.*, 2018). Key morphological characters required for plant identification include its leaves, stem, flowers, fruits, seeds, and habit (Hagedorn *et al.*, 2010; Santos *et al.*, 2012; Liu *et al.*, 2013; Hassoon *et al.*, 2018). The naming process involves the use of pictures and illustrations; utilization of identification keys in botanical books; and consulting the experts at the herbarium (Carrière, 2002; Hagedorn *et al.*, 2010; Culley, 2013; Felger *et al.*, 2014; Wilson *et al.*, 2014).

DNA regions used for identification purposes are referred to as molecular markers (Korzun, 2003). Standardized molecular markers, termed 'DNA barcodes,' have been adopted to aid in the identification and characterization of biodiversity (Hebert *et al.*, 2003; CBOL Plant Working Group, 2009; China Plant BOL Group, 2011; Hollingsworth *et al.*, 2016). The cytochrome c oxidase 1 (*CO*1) gene was the first barcoding gene to be universally used for identification across animal species (Hebert *et al.*, 2003). In plants, barcoding involves utilization of multiple loci from both the chloroplast and nuclear regions such as *mat*K, *rbc*L, the *trn*L(UAA) intron, and the *ITS* gene (CBOL Plant Working Group, 2009; China Plant BOL Group, 2011; Hollingsworth *et al.*, 2011; Kress, 2017; Tahir *et al.*, 2018; Wu *et al.*, 2019).

1.2 Problem statement

The different types of flora in the Amboseli ecosystem are yet to be genetically characterized. Previous research work on the flora existing in this ecosystem has only been based on phenotypic methods. One major constraint of morphological identification of plants is that it requires specific expertise in the taxonomic field.

Generally, Africa lags in the molecular characterization of its plants. A search on GenBank® shows that the database contains less than 60000 plant sequences each from Kenya, Uganda, Tanzania, and South Africa, whereas a developed country like Japan has deposited over 9 million plant sequences. This data is insufficient because current ecological concerns require highly precise taxonomic information to be present in the reference DNA databases in order to address issues such as what wild animals consume.

1.3 Justification of the study

Molecular analyses are efficient in the identification and distinguishing plant species because genetically, each species, and each individual, is unique in the fact that no one genome is identical to the next. In the recent past, molecular tools have been adopted for species identification as they provide easy, less laborious means for assigning known and unknown plant taxa. These techniques answer many new evolutionary and taxonomic queries, which are not possible with only morphological characterization because DNA sequences are more reliable in

capturing species differences and evolutionary relationships. Furthermore, molecular identification can be performed without having complete plant organs, and anyone could do it, regardless of their expertise in the field of taxonomy.

The 18S rDNA locus, *ITS*1, and *ITS*1-Poaceae barcodes are derived from the nuclear ribosomal DNA region, whereas the *trn*L (UAA) intron and *trn*L-p6 are chloroplast regions. The combination of the nuclear and plastid genomic information will confer more reliability to the data set. Furthermore, *ITS*1 and the *trn*L (UAA) intron are easily amplified in diverse plants, and the two barcodes were recently used to assess the diet of various herbivores such as elephants, impalas, dik-diks, buffaloes, zebras and cattle in Laikipia, Kenya (Kartzinel *et al.*, 2015). The adoption of the same barcodes is essential for the creation of a shared community resource. Therefore, this research will add on to the data generated by the aforementioned study.

1.4 Objectives

1.4.1 General objective

To identify the plants consumed by yellow baboons (*Papio cycnocephalus*) in Amboseli, Kenya, based on morphological and molecular analyses.

1.4.2 Specific objectives

1) To identify the selected plants based on their morphological characteristics and deposit them as voucher specimens.
2) To compare the discriminatory power of five barcodes namely, *ITS*1, *ITS*1-Poaceae, the I8S rDNA region, the *trn*L (UAA) intron, and *trn*L-p6.
3) To test the identification efficiencies of BLAST and phylogenetic analyses based on both GenBank® and BOLD databases.

1.5 Hypotheses

1) The plants of interest can be correctly identified based on their morphological characteristics.
2) The selected molecular markers have a high taxonomic resolution and can differentiate the plants into their given taxonomic groups.
3) BLAST and phylogenetic analyses based on both GenBank® and BOLD databases can efficiently discriminate the plants into their given taxonomic groups.

CHAPTER TWO: LITERATURE REVIEW

2.1 Plant diet analysis of yellow baboons

Yellow baboons (*Papio cynocephalus*) are described as "opportunistic omnivores" because they feed on different plants, insects, and animals, depending on their availability (Shefferly, 2004; Kingdon *et al*., 2008). In Kenya, observational studies on yellow baboons' diet have been conducted in Amboseli (Post *et al*., 1980; Post, 1982; Altmann, 1998) and Tana River (Bentley-Condit, 2009; Bentley-Condit and Power, 2018). However, Amboseli and the Tana River Primate National Reserve (TRPNR) are quite different baboon habitats because the former has much higher annual rainfall (Bentley-Condit, 2009).

The baboons in Amboseli live in an open savannah habitat, which has been defined as an area in which perennial grasses form the primary ground cover and in which trees occur at low density (Altmann, 1998). The wild baboons in this region have been noted to consume at least 44 plants, which include trees, shrubs, forbs, grasses, and sedges (Post *et al*., 1980; Post, 1982; Altmann, 1998). Examples of these food choices include seeds and flowers of *Acacia* spp., *Abutilon* sp., and *Rhamphicarpa montana*; fruits of *Azima tetracantha*, *Withania somnifera* and *Commicarpus plumbagineus*; and, corms and blades of *Sporobolus* spp. and *Cynodon* spp. A significant component of the baboons' diet is provided by grasses and sedges (Altmann, 1998). However, it was noted that the following plants - although commonly found in Amboseli - are not eaten by the baboons: *Volkensinia prostrata*

(that is also known as *Dasyphaera prostrata*), *Dicliptera albicaula, Leucas stricta* and *Solanum incanum*.

The yellow baboons that exist in TRPNR forage in both the savannah and riverine forests (Bentley-Condit, 2009). A study conducted by Bentley-Condit (2009) listed fifty plants eaten by these baboons, including fruits of *Alangium salviifolium*, *Saba comorensis,* and *Cordia sinsensis*; corms and shoots of *Cyperus* spp. and *Brachiaria* spp.; and, flowers of *Acacia robusta* and *Hibiscus micranthus.* Bentley-Condit and Power (2018) analysed the dietary macronutrient and mineral content of these baboons, mainly based on the food items on the previous research by Bentley-Condit (2009). The study compared the results for thirty-four forest species and twenty-four savannah species, which represented fifty-six flora species. The results indicated the highly selective dietary choices made by wild baboons.

2.2 Traditional methods of analysing plant composition in diets

Plants consumed by animals can be evaluated simply by directly observing their foraging behaviour. However, this process is very prolonged and impractical in some circumstances such as, when an animal consumes numerous plant sources that exist in the same space, or when the animal feeds at night or underground and cannot be observed (Valentini *et al.*, 2009). A second approach is to extract the stomach extrusa following anaesthesia, or through analysis of the gut contents after killing

an animal (Hyslop, 1980; Mcinnis *et al.*, 1983; Solé *et al.*, 2007). This method is impossible when it is neither ethical nor feasible to kill the animal of interest.

Other approaches include the morphological analysis of plant cuticle fragments in faecal matter via microscopy, and analysis of the natural alkanes of plant cuticular wax (Johnson *et al*., 1983; Stevens *et al.*, 1987; Shrestha and Wegge, 2006; de Iongh *et al.*, 2011; Garnick *et al.*, 2018). However, microscopic identification is not always reliable, mainly when the food items have been fully digested.

2.3 Genetic identification of plants

DNA-based methodologies provide precise tools that can be used to identify plant species and classify them into their specific taxonomic groups (Vijayan and Tsou, 2010; Ali *et al.*, 2014; Patwardhan *et al.*, 2014; Leache and Oaks, 2017; Mishra *et al.*, 2017). Sequence data exist for conserved loci, common to a wide range of organisms, which allow relevant genes to be amplified without any prior knowledge of the genome of the target species using universal primers (Arif *et al.*, 2010; Rydberg, 2010; Hollingsworth *et al.*, 2011; Santos and Pereira, 2016). The primers target highly conserved regions in angiosperms and gymnosperms, preventing strong bias due to primer mismatch in the efficiency of amplifications among species (Kress *et al.*, 2005; Li *et al.*, 2015; Staats *et al.*, 2016; Erickson *et al.*, 2017). Moreover, the DNA-based approach is particularly well-suited for large scale

analyses of plant material (Dong *et al.*, 2013; Angers-Loustau *et al.*, 2016; Fahner *et al.*, 2016).

DNA barcoding and metabarcoding techniques exploit short, standardized genetic markers, termed as "barcodes" to identify species (Hebert *et al.*, 2003). For barcoding, DNA is extracted from single specimens, followed by the amplification of organism-specific barcodes followed by direct Sanger sequencing (Taberlet *et al.*, 2012; Cristescu, 2014; Dechbumroong *et al.*, 2018; Wu *et al.*, 2019). In metabarcoding, DNA is extracted from environmental materials such as faecal samples, water, or soil which basically comprises DNA from different organisms (Cristescu, 2014; Fahner, 2015; Deiner *et al.*, 2017). Organism-specific barcodes are used to amplify the different DNA types, which are then sequenced using high-throughput (next-generation) sequencing leading to multiple species identification (Hajibabaei *et al.*, 2011; Taberlet *et al.*, 2012; Bell *et al.*, 2017; Mallott *et al.*, 2018).

The standard region for DNA barcoding in animals is mitochondrial cytochrome c oxidase 1 (*CO1*) gene, which allows researchers to distinguish between closely related animal species (Hebert *et al.*, 2003). However, in plants, the mitochondrial genome has evolved very slowly; therefore, *CO1* cannot provide sufficient nucleotide differentiation to differentiate species (Fazekas *et al.*, 2008; Hollingsworth *et al.*, 2011). Instead, plant DNA barcoding is primarily based on the chloroplast (plastid) genome; wherein multiple loci are used to enhance species resolution (Fazekas *et al.*, 2008, 2012; CBOL Plant Working Group, 2009;

Hollingsworth *et al.*, 2011). Nuclear and plastid DNA barcodes provide better species discrimination when used together (Chase *et al.*, 2005; Hollingsworth *et al.,* 2011; Fazekas *et al.*, 2012).

The degree of species resolution among plants, using accepted DNA plastid-based barcoding regions such as *mat*K and *rbc*L, is generally lower than the resolution typically attained by the mitochondrial *CO1* gene in animals (Fazekas *et al.*, 2009; Ali *et al.*, 2015). This low resolution occurs because the animal mitochondrion exhibits more variability in comparison to the plant plastid. Indeed, animal *CO1* has approximately 10-30 times more nucleotide substitutions than the plant plastid region (Wolfe *et al.*, 1987). The decreased variability in the plant genome results from various processes, including hybridization and polyploidy, which gives rise to similar haploid genotypes of the plastid in differentiated species (Fazekas *et al.*, 2009). However, these concerns are not common to every plant group, and for this reason, DNA barcoding markers have been successfully used to distinguish species in various plant groups (Ali *et al.*, 2015; Iwanowicz *et al.*, 2016; Bell *et al.,* 2017; Hosein *et al.*, 2017).

2.4 Applications of DNA (meta)barcoding

2.4.1 Dietary analysis

Plant-based diets of various primates have been analysed through DNA metabarcoding. Bradley *et al.* (2007) characterized the diets of wild western gorillas (*Gorilla gorilla*) and black and white colobus monkeys (*Colobus guereza*) using *rbc*L and *ITS2* barcodes. Quemere *et al.* (2013) and Srivathsan *et al.* (2014) exploited the *trn*L approach to evaluate the diets of the golden-crowned sifaka (*Propithecus tattersalli*) and red-shanked doucs langurs (*Pygathrix nemaeus*), respectively. Srivathsan *et al.* (2016) analysed the diet composition of the banded leaf monkey (*Presbytis femoralis*) by combining *rbc*L, *mat*K, and *trn*L-F markers. In the diet analysis of white-faced capuchins (*Cebus capuchins*), *trn*L barcodes outperformed *rbc*L and yielded more significant numbers of sequences with equal sequencing effort, higher resolution taxonomic identifications, and identified a greater number of families than the observed diet (Mallott *et al.*, 2018). However, the plant diet of the yellow baboons in Amboseli is yet to be characterized using molecular markers, including the *trn*L (UAA) intron, *ITS*1, and the 18S rDNA gene, hence the need for the study.

2.4.2 Characterizing biodiversity

Many plant species have yet to be characterized using molecular analyses (Pauls *et al.*, 2010; Thomsen and Willerslev, 2015; Hosein *et al.*, 2017). Furthermore,

existing taxonomic records for some plants need to be reconciled and updated so that unidentified organisms are correctly assigned to their taxonomic groups (Su *et al.*, 2016; Bezeng *et al.*, 2017; Hosein *et al.*, 2017). DNA barcoding reduces the ambiguity of species identification (Pettengill and Neel, 2010; Lopez-Alvarez *et al.*, 2012; Bączkiewicz *et al.*, 2017) and has also resulted in new species being found (Nguyen and Seifert, 2008; Pauls *et al.*, 2010; Liu *et al.*, 2013). DNA (meta)barcoding has also been employed to evaluate species richness within various regions (Fazekas *et al.*, 2008; Heise *et al.*, 2015; Pei *et al.*, 2017).

2.5 Plant barcodes

2.5.1 Plastid-based markers

Barcodes from the chloroplast genome comprise either protein-coding or non-coding regions.

2.5.1.1 Protein-coding barcodes

The most commonly used plastid-based barcodes are those encoding ribulose- 1,5-bisphosphate carboxylase (*rbc*L) and maturase K (*mat*K). A coding region (exon) is a locus in a gene that is transcribed and translated into protein. Such regions mutate slowly and hence are relatively more conserved than their non-coding counterparts. CBOL Plant Working Group (2009) recommended that the *rbc*L and *mat*K markers be adopted as the core DNA barcodes for land plants. This recommendation arises from the fact that the *rbc*L locus can be recovered easily, and the *mat*K locus results

in high resolution of interspecific and intraspecific relationships. However, the discriminatory power of the *rbc*L+*mat*K is secondary to that of the mitochondrial *CO1* gene in animals (CBOL Plant Working Group, 2009; Hollingsworth *et al.*, 2011). Moreover, in some plants, the *mat*K region is somewhat challenging to amplify with existing primers (CBOL Plant Working Group, 2009; Hollingsworth *et al.*, 2011). These shortcomings necessitate the adoption of alternative or supplementary markers from both the coding and non-coding regions of the chloroplast and, also, from the nuclear ribosomal DNA. These alternative markers are described below.

2.5.1.2 Non-coding barcodes

The non-coding loci in the chloroplast genome comprise introns and intergenic spacers. Generally, these loci are highly variable as compared to protein-coding regions. The most widely used non-coding barcodes from the chloroplast include the *psb*A-*trn*H intergenic spacer region (*trn*H-*psb*A), the $tRNA^{Leu}$ (UAA) intron sequence (also known as *trn*L UAA), and the intergenic spacer between the *trn*L (UAA) and *trn*F (GAA) genes. The *trnH-psb*A spacer is highly variable and is easily exploited in numerous land plants (Kress *et al.*, 2005; CBOL Plant Working Group, 2009; Pang *et al.*, 2012). Furthermore, the published primers appear to likely be universal in use (Kress *et al.*, 2005; Shaw *et al.*, 2007; Bolson *et al.*, 2015). Besides, this marker can be used to amplify DNA from degraded herbarium specimens (Shaw

et al., 2007). Nevertheless, sequencing *trn*H-*psb*A can sometimes present challenges due to the existence of micro-inversions and multiple mononucleotide repeats, which result in unidirectional reads (Devey *et al.*, 2009; Whitlock *et al.*, 2010).

The chloroplast *trn*L (UAA) intron, also referred to as $tRNA^{Leu}$ (UAA) intron sequence (Figure 2.1), is located between the *trn*F (GAA) and *trn*T (UGU) genes.

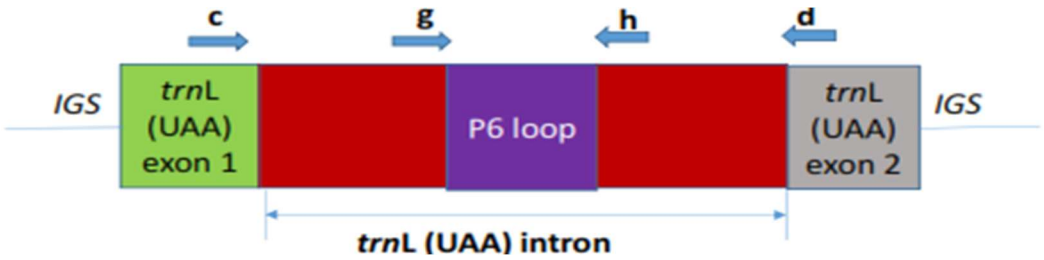


Figure 2.1: The chloroplast *trn*L (UAA) gene.

(Taberlet *et al*., 2007). Key: *IGS* – intergenic spacer (non-coding DNA sequences).

This gene has a conserved secondary structure, with alternating conserved and variable regions; hence, it is classified as the sole group I intron in the chloroplast (Taberlet *et al.*, 2007). This locus has been extensively studied since the beginning of the 1990s (Taberlet *et al.*, 1991). The intron contains a short stem-loop structure, referred to as the P6 loop (Figure 2.1). Both the entire *trn*L (UAA) intron and P6 loop have been successfully exploited in barcoding. However, when compared with several other non-coding chloroplast markers, these barcodes have a lower species resolution due to a smaller intraspecific variation (Taberlet *et al.*, 2007; Valentini *et*

al., 2009). The main advantage of this barcoding locus is the existence of the following universal primers: c and d, designed by Taberlet *et al.* (1991) to amplify the entire *trn*L (UAA) intron, and, g and h, designed by Taberlet *et al.* (2007) for the P6 loop (Table 3.2). Additionally, the P6 loop has been widely used in plant research regarding mixed template and/or degraded DNA samples such as faeces (Taberlet *et al.*, 2007; Valentini *et al.*, 2009; Hollingsworth *et al.*, 2011).

Valentini *et al.* (2009) developed the *trn*L approach, which exploits the P6 loop of the chloroplast *trn*L (UAA) intron – also referred to as *trn*L-P6 - to analyse the diet composition of animals. In the research above, approximately half of the taxa could be described to the species level. Various studies have adopted the *trn*L approach to study diets of certain herbivores (Soininen *et al.*, 2009; Rayé *et al.*, 2011; Kartzinel *et al.*, 2015) and birds (Ando *et al.*, 2013).

Kartzinel *et al.* (2015) employed both the *trn*L-P6 and *ITS*1 regions to investigate the diets of herbivores such as elephants, impalas, dik-diks, buffaloes, zebras, and cattle in Laikipia, Kenya. In this study, 77% of the *trn*L-P6 sequences corresponded to a single species/morphospecies, which indicated that this approach yielded high-resolution identifications and hence, consistent with prior evaluations of this marker. In a recent study, *trn*L-P6 outperformed the *rbc*L gene in that, it produced more significant numbers of sequences with equal sequencing effort, higher resolution taxonomic identifications (albeit with a more extensive reference database), and

identified a higher number of families also found in the observed diet (Mallott *et al.*, 2018).

2.5.2 Nuclear-based markers

Ribosomal DNA (rDNA) is the gene coding for ribosomal RNA (rRNA), which is found in the nucleus and is essential for protein synthesis in all living organisms (Rogers and Bendich, 1987). In eukaryotes, rDNA (Figure 2.2) exists in tandem repeats of genes (that is, 18S, 5.8S, and 26/28S) that are thousands of copies long, each divided by intergenic spacers (Rogers and Bendich, 1987).

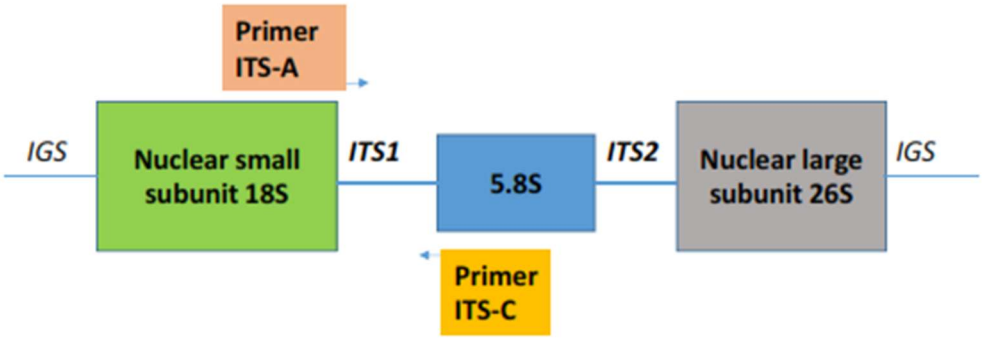


Figure 2.2: A nuclear ribosomal DNA repeat unit.

(adapted from Blattner, 1999) Key: *IGS* – intergenic spacer; *ITS* – internal transcribed spacer.

The 18S rDNA gene is a component of the small eukaryotic ribosomal subunit, while 5.8S and 26/28S are components of the large ribosomal subunit (Srivastava and Schlessinger, 1991; Olsen *et al.*, 1992). The internal transcribed spacer (*ITS*) is

the spacer DNA bounded by the small subunit rRNA (18S) and large subunit rRNA (26/28S) genes in the nuclear genome (Baldwin, 1992). The *ITS*1 gene is found in the middle of the 18S and 5.8S rRNA genes, and *ITS2* occurs between 5.8S and 26S rRNA genes (Baldwin, 1992).

The ribosomal DNA markers are generally based on the 18S, 26/28S, and the *ITS* loci. The 18S and 26S rRNA genes have been widely used for phylogenetic reconstruction at higher taxonomic levels in plants since the 19th century (Hamby and Zimmer, 1988; Mishler *et al.*, 1994; Kuzoff *et al.*, 1998; Soltis *et al.*, 1999).

The *ITS* locus (*ITS*1-5.8S-*ITS2*) has the highest species discriminatory power when compared to the existing coding and non-coding plastid markers (China Plant BOL Group, 2011; Cheng *et al.*, 2016), which is due to the high degree of sequence variation even within closely related species and high copy number of rRNA genes (Alvarez and Wendel, 2003; Chase *et al.*, 2005; China Plant BOL Group, 2011). Despite having high resolution, this locus was previously discounted as a barcode due to the following concerns: (1) paralogy and presence of pseudogenes within individuals can result in sequencing difficulties in numerous plant groups; and (2) fungal contamination can confuse species identifications, especially in instances where plants consist of fungal endophytes (Hollingsworth *et al.*, 2011; Fazekas *et al.*, 2012).

To reduce these limitations with amplification and sequencing, portions of the *ITS* assemblage, namely *ITS*1 and *ITS2*, have been individually accepted for barcoding

(Chen *et al.*, 2010; Han *et al.*, 2013; Mishra *et al.*, 2016). Wang *et al.* (2014) suggested that for barcoding eukaryotic species, *ITS*1 loci should be used instead of *ITS2*. Moreover, in terms of DNA sequencing and amplification efficiencies, the *ITS*1 region has several advantages, including having a set of primers that work in many plant groups; the length of the amplification product is shorter, and the GC content is lower. The *ITS*1 barcode has also been found to exhibit superior species discrimination to other commonly used barcodes, such as *mat*K, *rbc*L and *trn*H-*psb*A, applied singly or in combination (Wang *et al.*, 2014).

Kartzinel *et al.* (2015) employed the *ITS*1 locus to assess the robustness of the *trn*L-P6 marker in a dietary study of herbivores. In this research, three plant-family specific *ITS*1 markers (that is, *ITS*1-Asteraceae, *ITS*1-Cyperaceae, and *ITS*1-Poaceae) were chosen because they provided a greater species-level taxonomic resolution of plant sequences within the specified families. The *ITS* data independently validated conclusions based on *trn*L-P6 about the relative dietary importance of species within each of the three plant families. For instance, the grasses most frequently detected by *ITS*1, including *Pennisetum* spp., were also often identified by *trn*L-P6.

2.6 DNA barcoding database

With the emergence of barcoding approaches, several scientists recognized a need to develop an open-access and secure reference database that can be used to store,

organize, and query DNA barcoding records. Existing public DNA databases, such as GenBank®, have a large number of misidentified specimens, which then leads to erroneous identifications (Shen *et al.*, 2013). For instance, using public databases, it is at times difficult to discriminate partial sequences from those covering the whole *ITS* locus (Wang *et al.*, 2014). This problem led to the development of the Barcode of Life Data system (BOLD), which provides an integrated bioinformatics platform, which is the pillar of all stages of the analytical procedures from specimen collection to the comparison of sequences with existing barcodes (Ratnasingham and Herbert, 2007).

In addition to public databases, local, project-specific databases are also essential. Studies by Valentini *et al.* (2009) revealed that "by constructing a comprehensive database that comprises a majority of the plants occurring in a study site, about 50% of these plants will be identified to their species levels, whereas 90% will be discriminated to the genus levels. The degree of identification to the species level is lower when the sequences are matched to public databases as compared to population-specific, local databases. This greater accuracy is due to the higher occurrence of closely related species that exhibit the same P6 loop sequence in public databases". On the other hand, Nakahara *et al.* (2015) noted that a local database of the P6 loop might be limited in its taxonomic discrimination when a larger number of plant species are included.

CHAPTER THREE: MATERIALS AND METHODS

3.1 Study area

The study area of approximately 100 km^2 in size was located inside the Amboseli basin; at the south-western border of the Amboseli National Park (Figure 3.1).

Figure 3.1: The location of Amboseli National Park in Kajiado, Kenya.

(Source: https://www.expressvacationtours.co.ke/kenya.html)

The Amboseli National Park (Figure 3.1) occupies an area of 392 km^2 and lies at a latitude of 2º 40' 0S, the longitude of 37º 16' 60E, and an altitude of 1100 m (Mbane, 2012; Markham, 2017). It is located in Kajiado County, which is 250 km southeast of Nairobi, Kenya. The name "Amboseli" is a Maasai word that means "salty dust" (KWS, 2008), which is in reference to the volcanic ash from Mount Kilimanjaro eruptions that occurred a millennium ago (KWS, 2008).

The Amboseli basin spreads approximately 8,000 – 8,500 km^2 from Kenya to Tanzania, at the northern base of Mount Kilimanjaro (Altmann *et al.*, 1985; Gara *et al.*, 2016; Markham, 2017). It is an arid to semi-arid savanna environment (Altmann *et al.,* 1985; KWS, 2008) that receives less than 400 mm of rainfall per year (Kinuthia, 2002; Mbane, 2012; Gara *et al.*, 2016; Markham, 2017). The temperature ranges from 20 – 30 ºC (Gara *et al.*, 2016).

The Amboseli ecosystem is characterized by spatial and temporal variation in hydrology, and surface water is found in a few rivers, streams, and swamps (KWS, 2008; Okello *et al.*, 2016). These water resources are predominantly a result of the hydrological influence of Mount Kilimanjaro (Kemunto, 2013; Gara *et al.*, 2016). The extent of Lake Amboseli depends on the level of rainfall, which makes it be seasonal most of the time (Kinuthia, 2002).

The soils in the Amboseli basin are made up of volcanic ash deposits that resulted from the volcanic activities that formed Mt. Kilimanjaro and the adjacent Chyulu hills (Williams, 1972; Mbane, 2012). The hot and dry climate with its high

evapotranspiration resulted in the upward movement of salts in the soil, creating varying degrees of salinity and alkalinity conditions that support limited vegetation growth (Post, 1982; Altmann *et al.*, 1985; Kinuthia, 2002; Markham, 2017). The main vegetation types in the Amboseli ecosystem are open grasslands, shrublands and woodlands (Kinuthia, 2002; KWS, 2008; Mbane, 2012), which support the pastoralist lifestyle of the local Maasai and a wide array of wildlife that is the cornerstone of tourism in this region (KWS, 2008).

3.2 Collection of plant materials

A total of eighty plants were used in this study, and they were selected based on previous behavioural observations of the plant foods consumed by baboons. Precisely, thirty-eight of these plants (Table 3.1) are listed by Altmann (1998) and were identified in the field using key morphological characteristics described in appendices 1 and 2.

Forty-two plants, whose scientific names were uncertain, were selected based on additional observational data collected over the recent years by ABRP's field observers (E. A. Archie, personal communication). This batch consisted of six shrubs and herbs (indicated as, plants A, B, C, D, E, and G) and thirty-six monocotyledons (namely, grasses A – T; and AA – AR), collected in June 2016, and January and May/June 2018. Each of the plants was first photographed before collection (appendices 1 and 2).

Table 3.1: Plants used in this study.

Sample ID	Family	Common / alternate name
Abutilon mauritianum (Jacq.) Medik.	Malvaceae	Country mallow
*Acacia tortilis (*Forssk.) Hayne	Fabaceae	Umbrella tree; *Vachellia tortilis (*Forssk.)
Acacia xanthophloea Benth.	Fabaceae	Fever tree; *Vachellia xanthophloea* (Benth.)
Achyranthes aspera Linn.	Amaranthaceae	Devil's horse whip
Asparagus '*asparagii*'	Asparagaceae	Asparagus fern
Azima tetracantha Lam.	Salvadoraceae	Bee sting bush
Balanites pedicellaris Mildbr. & Schltr.	Zygophyllaceae	Soap berry tree
Balanites sp.	Zygophyllaceae	Soapberry tree
Cadaba farinosa Forssk	Capparaceae	Kadhab
Capparis tomentosa Lam.	Capparaceae	Caper bushes
Cassia italica (Mill) Lam.	Fabaceae	*Senna italica* Mill.
Commicarpus plumbagineus (Cav.) Standl.	Nyctaginaceae	Sticky fruit plants
Cordia monoica (Roxb.)	Boraginaceae	Sandpaper tree
Dasyphaera prostrata (Volk. ex Gilg) Cavaco	Amaranthaceae	*Volkensinia prostrata* (Volkens ex Gilg) Schinz
Drake-brockmania somalensis Stapf.	Poaceae	
Euclea schimperi (A.DC.) Dandy.	Ebenaceae	Blue guarri, Bush guarri, Magic guarri
Ficus sp.	Sapotaceae	Fig tree
Hibiscus 'lila'	Malvaceae	Lila
Lantana camara L.	Verbenaceae	Tickberry
Lycium europaeum L.	Ebenaceae	Trumpet flower bush
Maerua angolensis DC.	Capparaceae	Bead-bean
Maerua crassifolia Forssk.	Capparaceae	Jega/Agargar
Maerua sp.	Capparaceae	Bead-bean
Odyssea paucinervis (Nees) Stapf.	Poaceae	Spiky grass
Rhamphicarpa montana N.E. Br.	Orobanchaceae	*Cycnium tubulosum* (L.f.) Engl.
Rhus natalensis Bernh.	Anacardiaceae	Natal rhus
Ruellia patula Jacq.	Acanthaceae	Popping seed plant
Salvadora persica L. (Meswak)	Salvadoraceae	Witches' broom
Scutia myrtina (Burm. f.) Kurz.	Rhamnaceae	Cat-thorn
Setaria verticillata (L.) P. Beauv.	Poaceae	Burr grass
Solanum dubium Fresen.	Solanaceae	*Solanum coagulans* Forssk.; African nightshade
Solanum incanum L.	Solanaceae	African nightshade, Thorn apple, Bitter apple
Solanum nigrum L.	Solanaceae	Blackberry nightshade
Sporobolus consimilis Fresen.	Poaceae	
Suaeda monoica Forssk.	Amaranthaceae	Common seablite
Trianthema ceratosepalum Volkens & Irmscher	Aizoaceae	Giant pigweed
Tribulus terrestris L. (TT)	Zygophyllaceae	Devil's thorn
Withania somnifera (L.) Dunal	Solanaceae	Poison gooseberry

3.2.1 Preparation of voucher specimens

Twenty-three of the plants whose scientific names were uncertain were collected as a whole with their roots, stem, leaves, fruits and/or flowers still intact. Collected plants were placed between sheets of newspapers, which were then transferred to a wooden plant press (Figure 3.2). The plant press was secured tightly using a belt and left undisturbed for a month to allow the complete drying of the plants. These plants were then taken to the University of Nairobi herbarium for identification by the curator.

Figure 3.2: A wooden plant press used to flatten and dry plants.

3.3 Molecular analysis

For molecular analysis, tissues such as fruits, seeds, buds and young growing leaves were collected because they are known to contain high concentrations of DNA and low concentrations of secondary metabolites such as tannins, alkanoids, phenolics, and terpenes, which interfere with DNA isolation procedures (Michiels *et al.*, 2003; Souza *et al.*, 2011; Sahu *et al.*, 2012; Inglis *et al.*, 2018). The tissues were stored in duplicate that is, in silica gel beads and absolute 95% (v/v) ethanol (Bressan *et al.*, 2014), then later transported to the Institute of Primate Research (IPR) in both January and June 2018, for molecular analysis.

3.3.1 DNA extraction

Genomic DNA was extracted from the collected plant materials using Qiagen's DNeasy Plant Mini Kit as per a modified version of the manufacturer's protocol (Qiagen, 2015). Plants stored in silica gel were prioritized, and the ethanol-preserved replicates were used as a back-up.

In preliminary trials, four methods of extracting genomic DNA were tested using grass samples collected at IPR. However, it should be noted that these fresh samples were not stored in either silica gel or absolute ethanol before use, which was the only difference from the ones collected in Amboseli. The samples were pulverized using a mortar and pestle as per one of the following procedures: (1) ground first in liquid nitrogen then 400 µL of lysis buffer was added; (2) ground using only 1200 µL lysis buffer (without any prior addition of liquid nitrogen); (3) the samples were

placed in cryovial tubes which were then inserted in liquid nitrogen before grinding the samples; and (4) the samples were crushed without the addition of any reagents/buffers.

Based on the preliminary results, the modifications to the DNA extraction protocol (Qiagen, 2015) were as follows: the liquid nitrogen was excluded; the lysis buffer was increased from the recommended 400 µL to 1200 µL, and the incubation time for lysis (at 65 ºC) was increased from 10 min to overnight. Additionally, the pulverization of all the plant material was conducted using a homogenizer known as the 2010 Geno/Grinder® that was situated at the Barcoding laboratory in the National Museums of Kenya. This equipment reduced the chances of cross-contamination, which is one of the issues experienced when crushing multiple plants using mortars and pestles. Additionally, the process was very time-efficient because forty-eight samples were ground in 6 minutes.

Before pulverizing the plant tissues, the samples were prepared differently depending on the preservation method. For the specimens preserved in silica gel, approximately 100 mg tissue (dried weight) was removed from the envelopes containing silica gel and transferred into microcentrifuge tubes. Next, the samples were placed in the 2010 Geno/Grinder® and pulverized at 1610 rpm for 6 min. For the specimens preserved in ethanol, ~ 100 mg of each sample was first air-dried before being crushed using the 2010 Geno/Grinder® at 1610 rpm for 6 min. For samples that resisted pulverization, as was the case for some fruits and grasses, they

were first to cut into smaller pieces, and the pulverization process repeated as described above.

Next, 1200 µL of the lysis buffer AP1 and 4 µL of RNase A were added to the powdered samples before vortexing vigorously. The mixture was incubated overnight at 65°C with intermittent mixing. One hundred and thirty microliters (130 µL) of Buffer P3 was added to the lysate, mixed, and then incubated for 30 min on ice. The lysate was then centrifuged for 5 min at 14,000 rpm then pipetted into a QIAshredder mini spin column inserted in a 2 mL collection tube before being centrifuged for 2 min at 14,000 rpm. The flow-through fraction was transferred into a new 2 mL microcentrifuge tube without disturbing the cell-debris pellet.

Next, 1.5 volumes of Buffer AW1 was added to the cleared lysate and mixed by pipetting. Six hundred and fifty microliters of the mixture were pipetted into a DNeasy mini spin column, which was placed in a 2 mL collection tube. Any precipitate that may have formed on the collection was neither disturbed nor transferred to the spin column. The samples were centrifuged for 1 min at 8000 rpm, and any flow-through was discarded. This step was repeated until all the contents in the microcentrifuge tube had been transferred through the spin column. The samples were centrifuged for 1 min at 8000 rpm, followed by discarding of the flow-through and collection tube.

The DNeasy mini spin column was placed into a new 2 mL collection tube followed by the addition of 500 µL of Buffer AW2. The samples were then centrifuged for 1

min at 8000 rpm before discarding the flow-through. Five hundred microliters of Buffer AW2 was added to the DNeasy Mini spin column and then centrifuged for 2 min at 14,000 rpm and the flow-through discarded. Without adding any buffer, the spin column was centrifuged again for 1 min at 14,000 rpm to dry the membrane.

The DNeasy mini-spin column was transferred to a new 2 mL microcentrifuge tube, and 25 µL of Buffer AE was added directly onto the DNeasy membrane. The sample was then incubated for 1 hour at room temperature (15–25 °C) and then centrifuged for 10 min at 13000 rpm to elute. Once more, 25 µL of Buffer AE was added directly onto the DNeasy membrane, followed by an hour-long incubation at room temperature (15–25 °C). The sample was again centrifuged for 10 min at 13000 rpm to elute. Lastly, the DNeasy Mini spin column was removed from the tube, and the eluted genomic DNA solution was visualized on 1.5% (w/v) agarose gel as described in section 3.3.2, then stored at – 20 °C.

3.3.2 Agarose gel electrophoresis

DNA extracts and PCR products were visualized on a 1% (w/v) and 1.5% (w/v) agarose gels, respectively. This gel was prepared by mixing either 1 g or 1.5 g of agarose powder, respectively with 100 mL 1X TAE buffer (protocol described in Appendix 3). The mixture was placed in a microwave for 30 s so as to dissolve the powder, then left to cool before adding ethidium bromide. The solution was poured onto a gel tray with combs and left to solidify. The solid gel was then placed in a

gel tank containing TAE buffer, and 6 µL of the DNA extracts or PCR amplicons were loaded after first mixing with 1 µL of loading dye. The samples were left to run for about 1 hour at 100 V then observed under an ultraviolet (UV) transilluminator. A gel image was then recorded using a phone's camera.

The lengths of the fragments were estimated by comparing with the given measurements of the 100 bp NEB ladder.

3.3.3 Polymerase chain reaction (PCR)

Five genes, namely the *trn*L (UAA) intron, *trn*L-P6, *ITS*1, *ITS*1-Poaceae, and the 18S rDNA locus, were each targeted for amplification using their specific primer sequences (Table 3.2).

Table 3.2: Primers used in this study and their sequences.

Gene	Primer	Sequence 5'- 3'	Citation
*trn*L (UAA) intron	*trn*L(UAA)c	CGAAATCGGTAGACGCTACG	(Taberlet *et al.*, 1991)
	*trn*L(UAA)d	GGGGATAGAGGGACTTGAAC	
*trn*L-P6	*trn*L(UAA)g	GGGCAATCCTGAGCCAA	(Taberlet *et al.*, 2007)
	*trn*L(UAA)h	CCATTGAGTCTCTGCACCTATC	
*ITS*1	*ITS*-A	GGAAGGAGAAGTCGTAACAAGG	(Blattner, 1999)
	ITS-C	GCAATTCACACCAAGTATCGC	
*ITS*1-Poaceae	*ITS*1-F	GATATCCGTTGCCGAGAGTC	(Ait Baamrane *et al.*, 2012)
	*ITS*1Poa-R	CCGAAGGCGTCAAGGAACAC	
18S rDNA	18S-1510R	CCTTCYGCAGGTTCACCTAC	(Amaral-Zettler *et al.*, 2009)
	18S-1380F	NNNNNCCCTGCCHTTTGTACACAC	

The reaction mixture for the *trn*L (UAA) intron and *trn*L-P6 was identical and consisted of the following: 2.5 mM $MgCl_2$, 200 µM of each dNTP, 0.1 mg/mL BSA, 4% DMSO, 0.2 µM each of the forward and reverse primers, 0.2 µL Invitrogen™ Platinum™ Taq polymerase, and 2 µL of the DNA extract, prepared in a 12.5 µL PCR reaction mixture as described by Kartzinel *et al.* (2015). With regards to the *ITS*1 and *ITS*1-Poaceae regions, their reaction mixture was also identical and consisted of the following: 2.5 mM $MgCl_2$, 400 µM of each dNTP, 0.1 mg/mL BSA, 4% DMSO, 0.2 µM each of the forward and reverse primers, 0.2 µL Invitrogen™ Platinum™ Taq polymerase, and 2 µL of the DNA extract, prepared in a 12.5 µL PCR reaction mixture as described by Kartzinel *et al.* (2015). For the 18S rDNA locus, the reaction mixture constituted of the following: 2.5 mM $MgCl_2$, 200 µM of each dNTP, 0.3 µM forward and reverse primers, 1X PCR Buffer, 0.2 µL Invitrogen™ Platinum™ Taq polymerase, and 2 µL of the DNA extract, prepared in a 12.5 µL PCR reaction mixture (Hua *et al.*, 2018; Xue *et al.*, 2018).

Amplification for each barcoding region was conducted using SimpliAmp™ thermal cycler as per the thermocycling programs listed on Tables 3.3a–b (Kartzinel *et al.*, 2015) and Table 3.3c (Hua *et al.*, 2018; Xue *et al.*, 2018). PCR products were then visualized on 1.5% (w/v) agarose gel, as described in section 3.3.2, and then stored at – 20 °C.

Table 3.3 a: Thermal cycling programs for *trn*L loci.

Program/primer:	***trn*L(UAA)c/ *trn*L(UAA)d**	***trn*L(UAA)g/ *trn*L(UAA)h**	**Cycles**
Initial denaturation:	95°C for 4 min	95°C for 10 min	1
Denaturation:	94 °C for 30 s	95 °C for 30 s	35
Annealing:	50 °C for 30 s	55 °C for 30 s	
Extension:	72 °C for 1 min	72 °C for 30 s	
Final extension:	72 °C for 5 min	72 °C for 2 min	1

Table 3.3 b: Thermal cycling programs for *ITS*1 loci.

Program/primer:	**ITS-A/ITS-C**	***ITS*1-Poa**	**Cycles**
Initial denaturation:	95°C for 2 min	95°C for 10 min	1
Denaturation:	95 °C for 30 s	94 °C for 30 s	35
Annealing:	55 °C for 30 s	58 °C for 30 s	
Extension:	72 °C for 45 s	72 °C for 45 s	
Final extension:	72 °C for 5 min	72 °C for 2 min	1

Table 3.3 c: Thermal cycling program for the 18S rDNA locus.

Program/primer:	**18S**	**Cycles**
Initial denaturation:	98 °C for 1 min	1
Denaturation:	98 °C for 10 s	35
Annealing:	50 °C for 30s	
Extension:	72 °C for 30 s	
Final extension:	72 °C for 5 min	1

3.3.4 PCR product purification and sequencing

PCR products were cleaned using Qiagen's QIAquick PCR purification kit, as per the manufacturer's protocol (Qiagen, 2008), with a few modifications. First, five volumes of Buffer PB were mixed with one volume of the PCR sample. This mixture was then loaded onto a QIAquick spin column and centrifuged at 13000 rpm for 1 min. The flow-through was loaded onto the spin column twice, centrifuged again at

13000 rpm for 1 min, then eventually discarded. The QIAquick spin column was placed onto the same collection tube, and 750 µL of Buffer PE was added then centrifuged at 13000 rpm for 1 min. The flow-through was loaded onto the spin column twice, centrifuged again at 13000 rpm for 1 min, then eventually discarded.

Next, the QIAquick spin column was placed onto the same collection tube and centrifuged for an additional 1 min. The spin column was placed in a sterile 2 mL microcentrifuge tube to elute the DNA, and 13 µL of Buffer EB was added to the centre of the QIAquick membrane. The column was left to stand for 1 hour at room temperature before being centrifuged at 13000 rpm for 10 min. An additional 13 µL of Buffer EB was added to the centre of the QIAquick membrane, incubated at room temperature then centrifuged again. The DNA was visualized on 1.5% agarose gel, as described in section 3.3.2. The samples were then sent for Sanger sequencing at Macrogen Netherlands (Europe).

3.4 Data analysis

The amplicons for all five primer regions used in this study (Table 3.2) were sequenced in both directions. The chromatogram files were loaded into Geneious Prime® 2019.0.4 software (http://www.geneious.com, Kearse *et al.*, 2012), which was used to trim and clean the sequences. First, using Geneious builder, forward and reverse sequences obtained for each primer pair were aligned with sequences obtained from the BOLD database, and the primer regions were removed. Next, the

trimmed forward and reverse sequences were aligned using Geneious alignment, and the chromatogram was scrutinized to ensure that each base was called accurately. After confirming that there were no gaps or misreads, a consensus sequence was generated. Only bidirectional sequences that had a pairwise identity greater than 95% were used in section 3.5.

3.5 Species identification

Two different methods – BLAST and tree-building – were adopted to assess the success of species identification for each of the five loci (CBOL Plant Working Group, 2009; China Plant BOL Group, 2011; Elansary *et al.*, 2017; Tahir *et al.*, 2018).

3.5.1 Sequence similarity analysis

The consensus sequences generated (section 3.4) for each marker were queried via the BLASTn algorithm against both the BOLD and GenBank® databases. The identification of specimens was based on the BLAST1 method (Ross *et al.*, 2008), which stipulates that the correct identity is that of the genus or species associated with the best BLAST hit and e-value based on the threshold. This corresponded to choosing the top hit in the BLAST results (Ross *et al.*, 2008; Elansary *et al.*, 2017; Tahir *et al.*, 2018).

In this study, the threshold was set at 95%, and the top matching hit was used as the identity of the specimen queried. Precisely, (1) successful identifications occurred when the highest-scoring hit of the query was assigned to only a single genus or species (that is, one organism); (2) ambiguous identifications occurred when the highest-scoring hit was assigned to multiple genera or species (that is, many organisms); and, (3) the identity was considered as "unidentifiable" if the highest-scoring hit of the query was below 95%.

3.5.2 Phylogenetic analysis

The goal of this analysis was to investigate the placement of individual plant specimens for taxonomic identification, rather than a determination of plant phylogenetic relationships; hence, phylogenetic congruence among markers was not assessed. For each primer region, sequences representing dicotyledons (that is, trees, shrubs, and herbs) were separated from monocotyledons (specifically, grasses and sedges). These sequences were then combined with closely matching sequences from GenBank®, and BOLD identified via BLAST. Next, these sequences were aligned using Geneious Prime® 2019.0.4's inbuilt MUSCLE (Edgar, 2004) plugin, as per the default settings. The multiple alignments were then edited to remove inaccurate gaps and misreads.

Molecular phylogenies were then constructed using the Maximum-Likelihood (ML) criterion in the following software: MEGA X (http://www.megasoftware.net,

Kumar *et al.*, 2018), Garli 2.0 (Zwickl, 2006; Bazinet *et al.*, 2014) and RAxML version 8.2.1.1 (Stamatakis, 2014, 2016) plugins in Geneious Prime®. For the MEGA X analysis, multiple alignments generated in Geneious were exported to MEGA X. Next, the best-fitting models of molecular evolution were determined for all loci as per the following parameters: (i) Neighbour-Joining tree; (ii) maximum-likelihood statistical method; and, (iii) 50% partial deletion of gaps/missing data. The best model was selected by comparing the Bayesian Information Criterion (BIC) score, Akaike Information Criterion, corrected (AICc) value, and Maximum Likelihood value (*lnL*) among models (Kumar *et al.*, 2018). Maximum-likelihood reconstructions were conducted using 1000 bootstrap replicates, following the settings indicated by the initial model test. Gaps/missing data on each alignment were partially (50%) deleted. Garli 2.0 was run as per the recommended settings, while RaxML specifications were as follows: (a) GTR+GAMMA model of evolution; (b) rapid bootstrapping and search for best-scoring ML tree; and (c) 1000 bootstrap iterations (Birch *et al.*, 2017). All the images were generated using Geneious version 2019.0 created by Biomatters and edited in FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree).

For specimen identification based on phylogenetic trees, the criteria according to the "liberal tree-based method" of Meier *et al.* (2006) was applied. Specimen identifications to genus rank were categorized as follows: (1) successful identifications occurred when the query sequence was placed in a clade exclusively consisting of congeneric individuals, sister to a clade with congeneric individuals,

or in a polytomy with congeneric individuals; (2) ambiguous identifications occurred if the individual was placed in a polytomy with allogeneric individuals, or sister to a clade with allogeneric individuals, and (3) "unidentifiable" if no individuals were included in the dataset (Meier *et al.,* 2006; Ross *et al.*, 2008; Birch *et al.*, 2017).

Species identifications were categorized as follows: (1) successful identifications occurred when the query sequence was placed in a clade exclusively consisting of one individual/species; (2) ambiguous identifications occurred if the individual was placed in a polytomy with allospecific individuals, or sister to a clade with allospecific individuals; and (3) "unidentifiable" or if no individuals were included in the dataset (Meier *et al.,* 2006; Ross *et al.*, 2008; Birch *et al.*, 2017).

CHAPTER FOUR: RESULTS

4.1 Morphological identification

Twenty-three plants (Figures 4.0 a-t) whose scientific names were uncertain during collection were identified at the University of Nairobi's herbarium then given the voucher reference numbers listed in Tables 4.0 a-b.

Table 4.0 a: Voucher reference information for dicotyledons deposited at the University of Nairobi herbarium.

Herbarium ID	**Family**	**Voucher reference no.**	**Locality; Habitat; Collector; Collection ID; Date**
Asparagus setaceus (Kunth) Jessop	Asparagaceae	EAA2018/01	Amboseli, Kenya; semi-arid savannah; Elizabeth A. Archie; *Asparagus asparagii*;28-May-18
Barleria masaiensis L. Darbysh	Acanthaceae	EAA2018/02	Amboseli, Kenya; semi-arid savannah; Elizabeth A. Archie; Plant C; 30-May-18
Boerhavia erecta L.	Nyctaginaceae	EAA2018/03	Amboseli, Kenya; semi-arid savannah; Elizabeth A. Archie; Plant B; 29-May-18
Ipomoea obscura (L.) Ker Gawl.	Convolvulaceae	EAA2018/15	Amboseli, Kenya; semi-arid savannah; Elizabeth A. Archie; Plant D; 30-May-18
Tephrosia pumila (Lamb.) Pers. va pumila	Fabaceae	EAA2018/20	Amboseli, Kenya; semi-arid savannah; Elizabeth A. Archie; Plant E; 1-Jun-18

Table 4.0 b: Voucher reference information for monocotyledons deposited at the University of Nairobi herbarium.

Herbarium ID	Family	Voucher reference no.	Locality; Habitat; Collector; Collection ID; Date
Brachiaria dictyoneura (Fig. & De Not.) Stapf	Poaceae	EAA2018/04	Amboseli, Kenya; semi-arid savannah; Elizabeth A. Archie; Grass AA, AQ; 28-May-18
Cenchrus ciliaris L.	Poaceae	EAA2018/05	Amboseli, Kenya; semi-arid savannah; Elizabeth A. Archie; Grass AD, AG; 29-May-18
Chloris virgata Swartz	Poaceae	EAA2018/06	Amboseli, Kenya; semi-arid savannah; Elizabeth A. Archie; Grass AE; 29-May-18
Cynodon aethiopicus Clayton & Harlan	Poaceae	EAA2018/07	Amboseli, Kenya; semi-arid savannah; Elizabeth A. Archie; Grass AM; 29-May-18
Cynodon dactylon (L.) Pers.	Poaceae	EAA2018/08	Amboseli, Kenya; semi-arid savannah; Elizabeth A. Archie; Grass AI; 29-May-18
Cyperus kilimandscharicus Kük.	Cyperaceae	EAA2018/09	Amboseli, Kenya; semi-arid savannah; Elizabeth A. Archie; Grass AK; 29-May-18
Cyperus teneriffae Poir.	Cyperaceae	EAA2018/10	Amboseli, Kenya; semi-arid savannah; Elizabeth A. Archie; Grass AR; 1-Jun-18
Dactyloctenium aegyptium (L.) Willd.	Poaceae	EAA2018/11	Amboseli, Kenya; semi-arid savannah; Elizabeth A. Archie; Grass AB; 28-May-18
Drake-brockmania somalensis Stapf	Poaceae	EAA2018/12	Amboseli, Kenya; semi-arid savannah; Elizabeth A. Archie; *Drake-brockmania somalensis*; 28-May-18
Enneapogon cenchroides (Roem. & Schult.) C.E. Hubb.	Poaceae	EAA2018/13	Amboseli, Kenya; semi-arid savannah; Elizabeth A. Archie; Grass AH; 29-May-18
Eragrostis cilianensis (All.) Lut.	Poaceae	EAA2018/14	Amboseli, Kenya; semi-arid savannah; Elizabeth A. Archie; Grass AP; 30-May-18
Kyllinga comosipes (Mattf. & Kük.) Napper	Cyperaceae	EAA2018/16	Amboseli, Kenya; semi-arid savannah; Elizabeth A. Archie; Grass AL; 29-May-18
Sporobolus ioclados (Trin.) Nees\[	Poaceae	EAA2018/17	Amboseli, Kenya; semi-arid savannah; Elizabeth A. Archie; Grass AJ, AN; 29-May-18
Sporobolus quadratus W. D. Clayton	Poaceae	EAA2018/18	Amboseli, Kenya; semi-arid savannah; Elizabeth A. Archie; Grass AF; 29-May-18
Sporobolus stapfianus Gand.	Poaceae	EAA2018/19	Amboseli, Kenya; semi-arid savannah; Elizabeth A. Archie; Grass AC; 28-May-18

Asparagus setaceus (Kunth) Jessop (Figure 4.0 a; Table 4.0 a) is a scrambling perennial herb with sturdy green stems, which may reach several metres in length. The leaves are leaf-like cladodes, which arise in clumps of up to 15 from the stem, making a delicate, soft green fern-like foliage.

Figure 4.0 a: *Asparagus setaceus* (Kunth) Jessop

(Source: E. A. Archie)

Barleria masaiensis L. Darbysh (Figure 4.0 b; Table 4.0 a) is an erect prickly shrub that is usually single-stemmed. It has ellipsoid leaves with the base protected by three to five sharp spines. The tubular flowers are yellow-orange with several long protruding stamens.

Figure 4.0 b: *Barleria masaiensis* L. Darbysh.

(Source: E. A. Archie)

Boerhavia erecta L. (Figure 4.0 c; Table 4.0 a) is an annual to short-lived perennial herb. The stem branches mainly from the base, and it is fleshy; green and often flushed with red; lower parts are thinly and hairy while the upperparts are glabrous with swollen nodes. The leaves are opposite, simple, about equal; stipules absent; blade broadly lanceolate to ovate. The inflorescence is an axillary, small, often congested umbel.

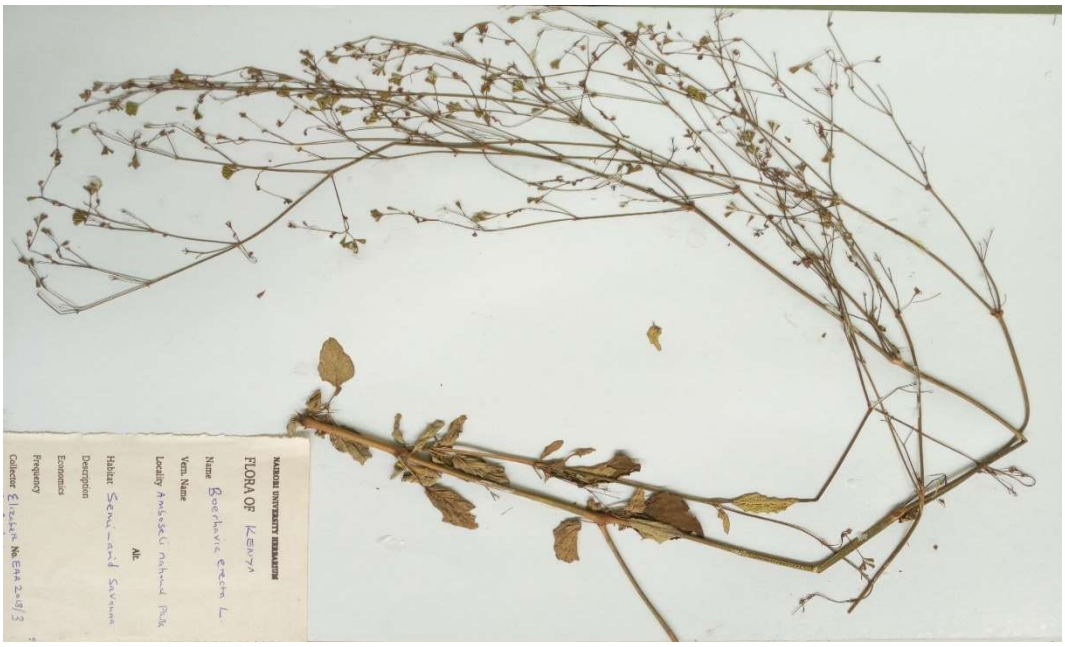


Figure 4.0 c: *Boerhavia erecta* L.

(Source: E. A. Archie)

Ipomoea obscura (L.) Ker Gawl. (Figure 4.0 d; Table 4.0 a) is an annual or perennial herb with slender, twining or prostrate stems. Its inflorescence is a simple cyme or reduced to 1 or 2 flowers.

Figure 4.0 d: *Ipomoea obscura* (L.) Ker Gawl.

(Source: E. A. Archie)

Tephrosia pumila (Lamb.) Pers. va pumila (Figure 4.0 e; Table 4.0 a) is an annual or short-lived perennial with procumbent or straggling branches. The leaf-rhachis includes a petiole with stipules narrowly triangular or subulate. Its flowers are white, pale pink or purplish in short terminal or leaf-opposed pseudo-racemes and upper leaf-axils. The bracts are narrowly triangular.

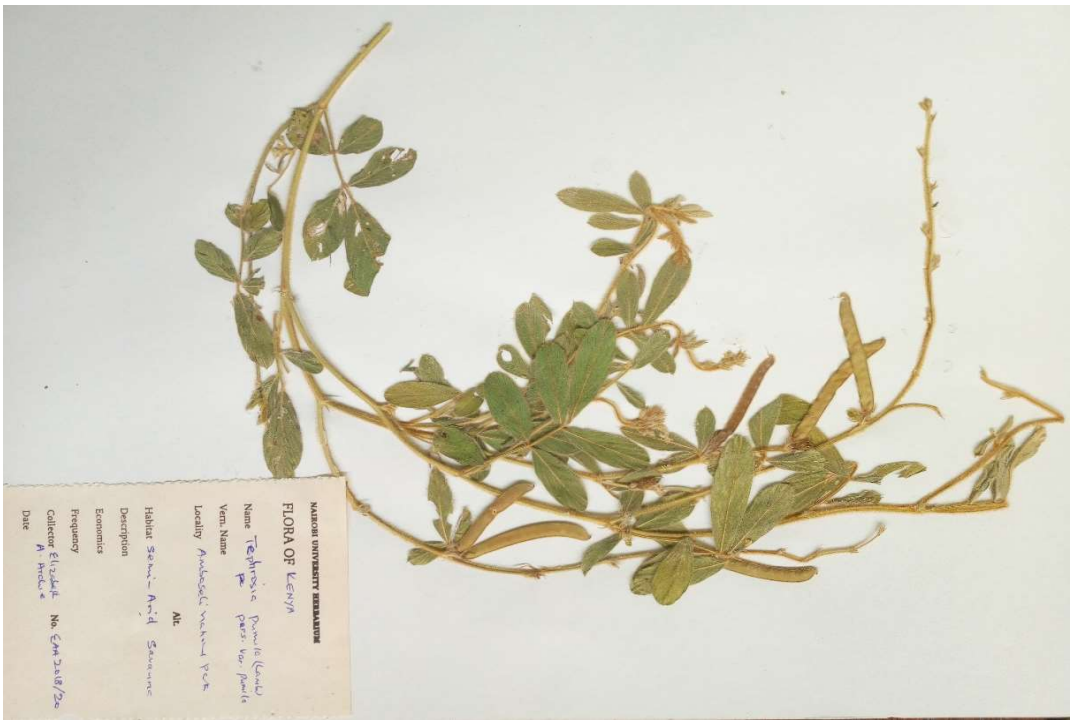


Figure 4.0 e: *Tephrosia pumila* (Lamb.) Pers. va pumila.

(Source: E. A. Archie)

Brachiaria dictyoneura (Fig. & De Not.) Stapf. (Figure 4.0 f; Table 4.0 b) is a densely tufted, semi-erect, stoloniferous perennial with short rhizomes and with stems. The stolons are slender but strong and of a reddish colour. The leaves are linear to lanceolate, with strongly denticulate margins. The inflorescence consists of 3-8 racemes on an axis, bearing spikelets in two rows; spikelets elliptic.

Figure 4.0 f: *Brachiaria dictyoneura* (Fig. & De Not.) Stapf.

(Source: E. A. Archie)

Cenchrus ciliaris L. (Figure 4.0 g; Table 4.0 b) is a perennial grass with linear leaves and flowers produced in a panicle. The inflorescence is a bristly false spike, straw- or purple-coloured; all bristles are joined at the base below spikelet cluster to form a small inconspicuous disc.

Figure 4.0 g: *Cenchrus ciliaris* L.

(Source: E. A. Archie)

Chloris virgata Swartz (Figure 4.0 h, Table 4.0 b) is an annual grass with tufted culms. The basal leaf sheaths are strongly keeled; leaf blades are flat or folded, glabrous, adaxial surface scabrous, and apex acuminate. The racemes are digitate, silky, pale brown, or tinged pink or purple. It has spikelets with 2 or 3 florets.

Figure 4.0 h: *Chloris virgata* Swartz.

(Source: E. A. Archie)

Cynodon aethiopicus Clayton & Harlan (Figure 4.0 i; Table 4.0 b) is a coarse stoloniferous perennial without rhizomes. Its stolons are stout, lying flat on the ground, whereas the culms are very robust, hard, shining, and woody. The leaf-blades are wide, stiff and harsh, glaucous, scaberulous, glabrous or with a few scattered hairs. The racemes occur in 2–5 whorls (rarely 1), are stiff and spreading. The spikelets are strongly pigmented with red or purple.

Figure 4.0 i: *Cynodon aethiopicus* Clayton & Harlan.

(Source: E. A. Archie)

Cynodon dactylon (L.) Pers. (Figure 4.0 j; Table 4.0 b) has blades that are a grey-green colour with rough edges. The stems are slightly flattened, often tinged purple. The seed heads are produced in a cluster of 2 to 6 spikes together at the top of the stem. This grass has a deep root system.

Figure 4.0 j: *Cynodon dactylon* (L.) Pers.

(Source: E. A. Archie)

Cyperus kilimandscharicus Kük (Figure 4.0 k; Table 4.0 b) is an annual or perennial sedge. The culms are usually simple, triangular and leafy. It has perfect flowers with the inflorescence being involucrate in dense spikes or clusters, capitate, or on rays, which are often compound. Its spikelets are flat or subterete.

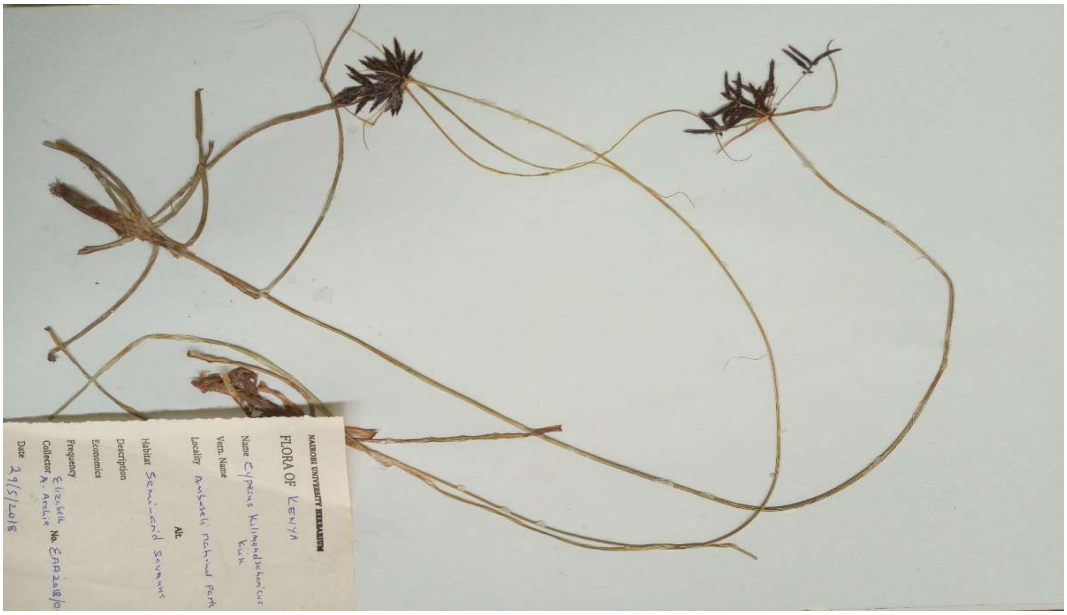


Figure 4.0 k: *Cyperus kilimandscharicus* Kük.

(Source: E. A. Archie)

Cyperus teneriffae Poir (Figure 4.0 l; Table 4.0 b) is an annual sedge with fibrous roots. It has few leaves that are weak, flat or conduplicate, gradually acuminate, smooth or scaberulous at the top. Its inflorescence is a single, hemispherical, or subglobose head.

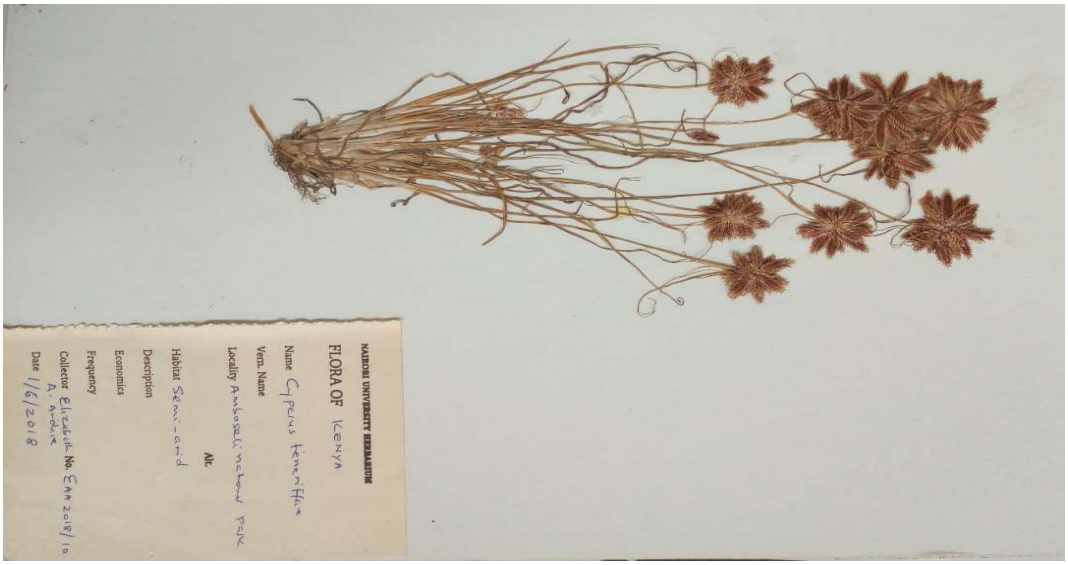


Figure 4.0 l: *Cyperus teneriffae* Poir.

(Source: E. A. Archie)

Dactyloctenium aegyptium (L.) Willd (Figure 4.0 m; Table 4.0 b) is a slightly stoloniferous and tufted short-lived perennial or annual grass, consisting of many branches. The stems are slender, ascending and geniculate or erect. The stolons root from the lower nodes and may creep. The roots are horizontal, while the leaves are broadly linear. The inflorescences are borne at the apex of the stem and are characteristically digitate or sub-digitate and arranged in two to six single, horizontal spikes.

Figure 4.0 m: *Dactyloctenium aegyptium* (L.) Willd.

(Source: E. A. Archie)

Drake-brockmania somalensis Stapf. (Figure 4.0 n; Table 4.0 b) is an annual grass. Its culms are prostrate, spreading, sometimes rooting at the nodes. It is highly-branched with flowering occurring on the ascending lateral branches. The inflorescence is often sub-capitate, composed of 2–6 spikes on a central axis.

Figure 4.0 n: *Drake-brockmania somalensis* Stapf.

(Source: E. A. Archie)

Enneapogon cenchroides (Roem. & Schult.) C.E. Hubb. (Figure 4.0 o; Table 4.0 b) is an annual grass with intact basal sheaths. The panicle is loosely contracted, often lobed at the base, and hairy on the back.

Figure 4.0 o: *Enneapogon cenchroides* (Roem. & Schult.) C.E. Hubb.

(Source: E. A. Archie)

Eragrostis cilianensis (All.) Lut. (Figure 4.0 p; Table 4.0 b) is an annual grass that forms tufts. The stems are generally erect but may droop or bend. The stems have glandular tissue near the nodes, and the long leaves are often dotted with glands as well. The branching inflorescences have one to several spikelets per branch. Each spikelet is greenish-brown, sometimes very slightly purple-tinted.

Figure 4.0 p: *Eragrostis cilianensis* (All.) Lut.

(Source: E. A. Archie)

Kyllinga comosipes (Mattf. & Kük.) Napper (Figure 4.0 q; Table 4.0 b) is a herbaceous plant with culms of grass-like leaves growing from a long, slender rhizome that creeps horizontally under or close to the ground surface.

Figure 4.0 q: *Kyllinga comosipes* (Mattf. & Kük.) Napper.

(Source: E. A. Archie)

Sporobolus ioclados (Trin.) Nees (Figure 4.0 r; Table 4.0 b) is a tussocky perennial, often with creeping stolons. Its leaves are flat or rolled, harsh or soft, often pungent. The basal sheaths are persistent, chartaceous, often keeled and flabellate, and the panicle is narrowly ovate to pyramidal.

Figure 4.0 r: *Sporobolus ioclados* (Trin.) Nees.

(Source: E. A. Archie)

Sporobolus quadratus W. D. Clayton (Figure 4.0 s; Table 4.0 b) is a tufted perennial with basal sheaths that are fairly broad, usually papery to sub-coriaceous. Its leaf-blades are convolute; panicle spiciform; primary branches appressed to the central axis, densely spiculate. The spikelets are grey-green.

Figure 4.0 s: *Sporobolus quadratus* W. D. Clayton.

(Source: E. A. Archie)

Sporobolus stapfianus Gand (Figure 4.0 t; Table 4.0 b) is a densely caespitose perennial. The basal leaf-sheaths form a compacted mass of fine fibres with age. The leaf-blades are convolute while the sheath-margins are tomentose with curly hairs. The branches are capillary and tinged with red. Its spikelets are greyish-green or sometimes dark green.

Figure 4.0 t: *Sporobolus stapfianus* Gand.

(Source: E. A. Archie)

4.2 Molecular analysis

4.2.1 DNA extraction

DNA was extracted from 62 samples preserved in silica gel, and 85.48% of the samples produced a distinct DNA band on an agarose gel (i.e., 53 out of 62 plants). In contrast, out of 54 samples preserved in ethanol, only 62.96% produced a distinct DNA band on an agarose gel (i.e., 34 of 54 plants). Representative gel images of DNA extracts for the samples preserved in silica gel and ethanol are shown in Figures 4.1 and 4.2.

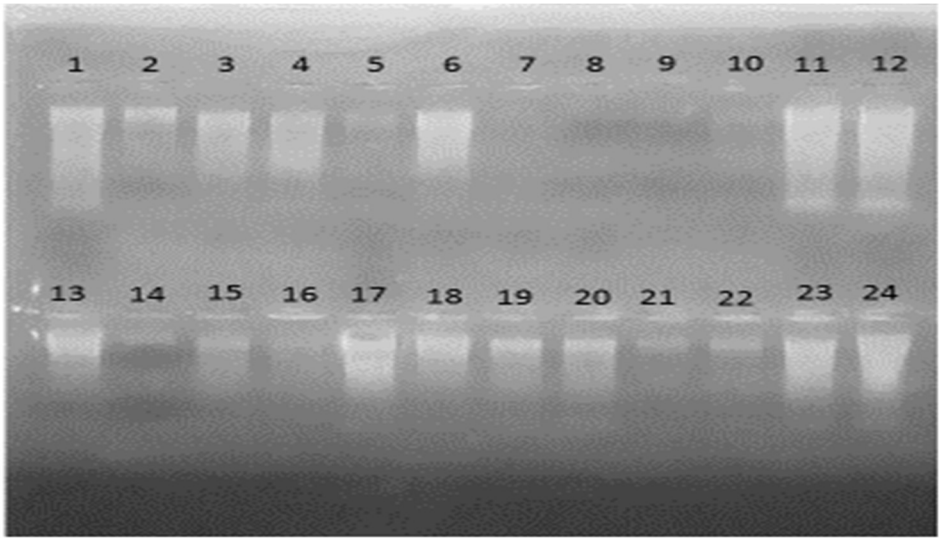


Figure 4.1: An agarose gel image of genomic DNA extracted from plants preserved in silica gel.

Key: Gel lanes: 1,2 - *Rhamphicarpa montana*; 3,4 - *Euclea schimperi*; 5,6 - *Tribulus terrestris*; 7,8 - *Lycium europaeum*; 9,10 - *Dasyphaera prostrata*; 11,12 - *Cordia monoica*; 13,14 - *Solanum dubium*; 15,16 - *Sporobolus consimilis*; 17,18 - *Withania somnifera*; 19,20 - Grass D; 21,22 - *Maerua crassifolia*; 23,24 - *Scutia myrtina.*

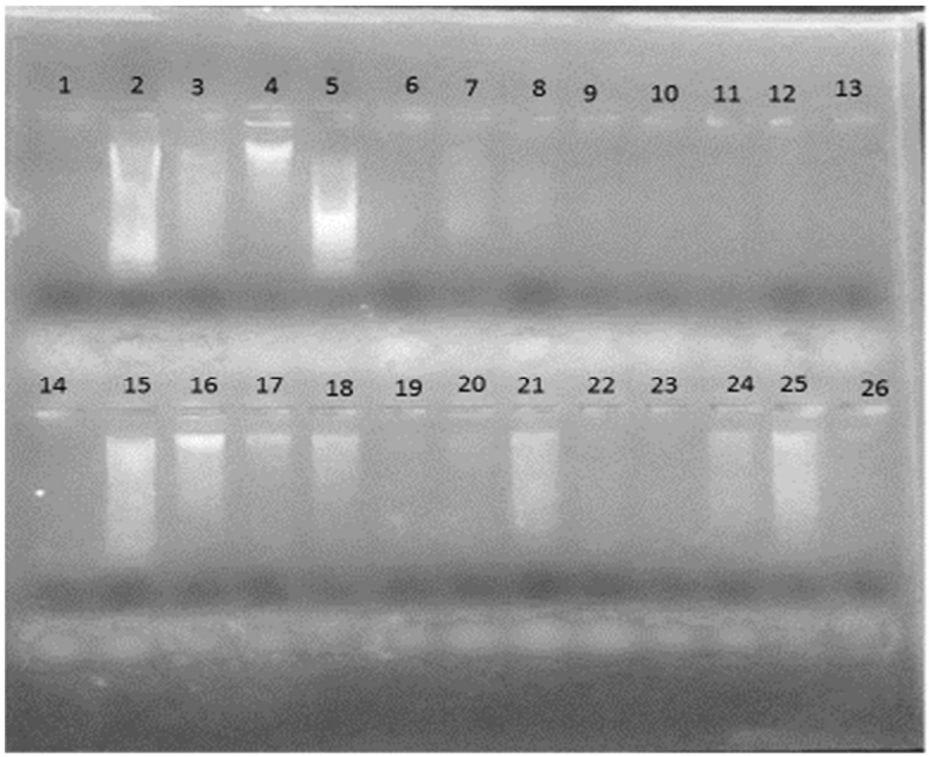


Figure 4.2: An agarose gel image of genomic DNA extracted from preserved in ethanol.

Key: Gel lanes: 1 - *Achyranthes aspera*; 2,3 - *Azima tetracantha*; 4 - *Capparis tomentosa*; 5,6 - *Commicarpus plumbagineus*; 7 - *Cordia monoica*; 8 - *Dasyphaera prostrata*; 9-*Euclea schimperi*; 10-*Ficus sp.*; 11-*Rhus natalensis*; 12-*Ruellia patula*; 13 - *Scutia myrtina*; 14 - *Withania somnifera*; 15 - *Balanites sp.*; 16 - *Cadaba farinosa;* 17 - *Cassia italica;* 18 - Grass G; 19 - Grass H; 20 - Grass I; 21 - Grass K; 22 - Grass L; 23 - Grass M; 24 - Grass N; 25 - Grass O; 26 - Grass P.

4.2.2 Polymerase chain reaction (PCR)

4.2.2.1 The amplification success rates of the candidate barcoding genes

Three-hundred and nineteen samples were evaluated, and they generated 289 amplicons (Table 4.1). The *ITS*1 locus was successfully amplified in all the plants,

whereas the *trn*L (UAA) intron had a success rate of 95% (76 out of 80 samples; Table 4.1). The *ITS*1-Poaceae region was amplified in 90% of all the monocotyledons (36 out of 40 samples; Table 4.1), while the amplification rate for the 18S rDNA region was 82.05% (64 out of 78 samples; Table 4.1). Lastly, *trn*L-P6 had the lowest amplification rate at 80.49% (33 out of 41 plants; Table 4.1). The bands for this region were very faint even after several optimization attempts; hence, the PCRs were discontinued, and not all the 80 plants were tested. It was additionally noted that this locus was a loop within the *trn*L (UAA) intron; hence, the data generated by the latter locus could be sufficient.

It was noted that the samples that showed no genomic DNA on an agarose gel after extraction still produced successful amplification of the target genes. This observation suggested that the concentrations of DNA were too low to be visualized on an agarose gel.

Table 4.1: Plants successfully amplified by each barcode

Barcode	**No. of samples tested**	**No. of samples amplified**
*ITS*1	80	80 (100%)
*trn*L(UAA) intron	80	76 (95%)
*ITS*1-Poaceae	40 (monocotyledons only)	36 (90%)
18S	78	64 (82.05%)
*trn*L-P6	41 (the samples produced very faint bands; hence PCR was not done for the rest of the plants)	33 (80.49%)
Total:	**319**	**289 (90.60%)**

4.2.2.2 The amplification success rates in monocotyledons and dicotyledons

Out of the 289 amplicons that were generated, 161 amplicons (55.71%) belonged to grasses and sedges, henceforth referred to as monocotyledons (Table 4.2), while 128 amplicons (44.29%) were for the rest of the plants that included trees, shrubs, and herbs, hereafter referred to as dicotyledons (Table 4.2). The amplification success rate was generally higher in monocotyledons (93.06%; Table 4.2) as compared to dicotyledons (87.67%; Table 4.2). For candidate barcodes, the amplification success rates ranged from 87.50 – 100% in monocotyledons, whereas it was from 75 – 100% in dicotyledons (Table 4.2).

The *ITS*1 locus was amplified in all monocotyledons and dicotyledons (Table 4.2), whereas the *trn*L (UAA) intron was the second most amplified locus in both monocotyledons and dicotyledons (Table 4.2). As for the *trn*L-p6 region, it was the third most successfully amplified region in monocotyledons (92.31%), and the least amplified region in dicotyledons (75%; Table 4.2). The 18S rDNA locus was the least amplified region in monocotyledons (87.50%), and the third successfully amplified locus in dicotyledons (75%).

Table 4.2: Monocotyledons and dicotyledons successfully amplified by the candidate barcodes

Monocotyledons (i.e., grasses and sedges)		
Barcode	**No. of samples tested**	**No. of samples amplified**
*ITS*1	40	40 (100%)
*trn*L(UAA) intron	40	38 (95%)
*trn*L-P6	13	12 (92.31%)
*ITS*1-Poaceae	40	36 (90%)
18S	40	35 (87.50%)
Total:	**173**	**161 (93.06%)**
Dicotyledons (i.e., trees, shrubs and herbs)		
Barcode	**No. of samples tested**	**No. of samples amplified**
*ITS*1	40	40 (100%)
*trn*L(UAA) intron	40	38 (95%)
18S	38	29 (76.32%)
*trn*L-P6	28	21 (75%)
Total:	**146**	**128 (87.67%)**

Representative gel images for the amplification results of the five regions of interest are shown in Figures 4.3, 4.4, 4.5, 4.6, and 4.7.

The *ITS*1 region was amplified in all the 80 plants after the optimization of the protocol. This locus was approximately 300 - 400 bp in length (Figure 4.3).

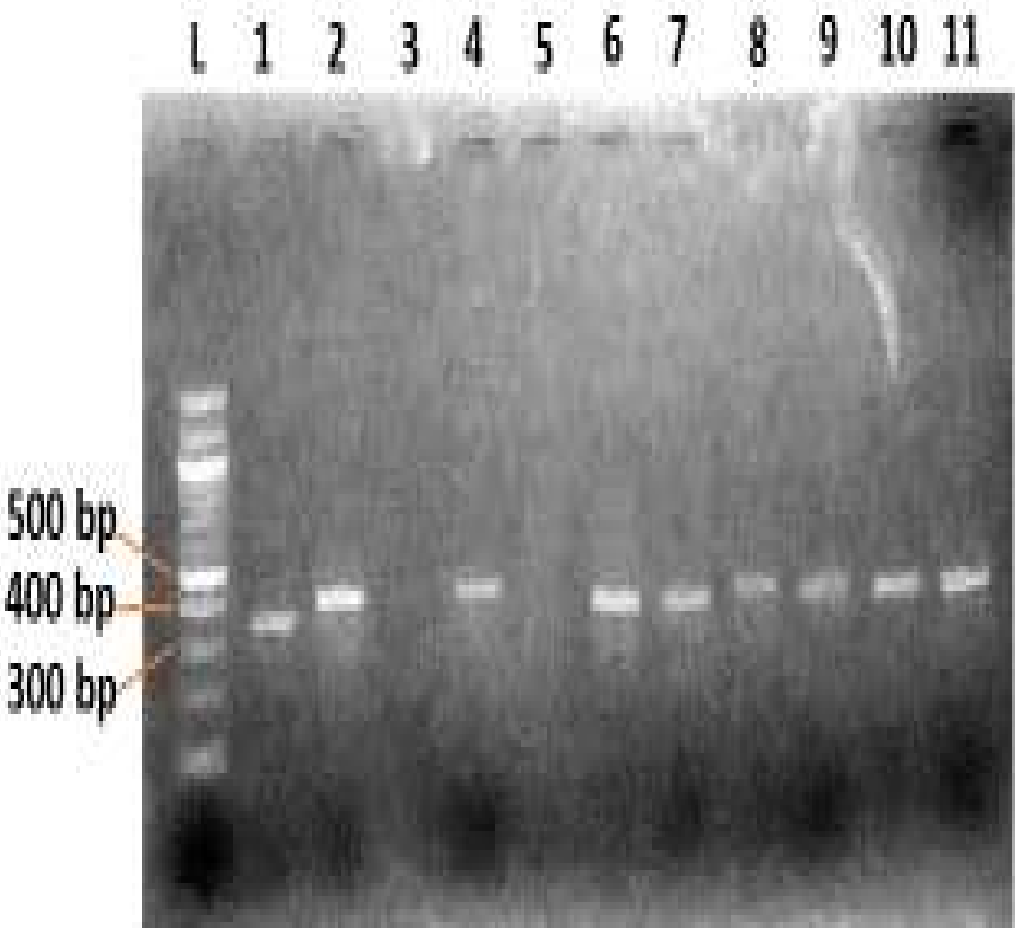


Figure 4.3: An agarose gel image for the PCR amplicons of the *ITS*1 region.

Key: - Gel lane: L - 100 bp NEB ladder, 1 - *Rhamphicarpa montana*; 2 - *Rhus natalensis*, 3 - *Ruellia patula*, 4 - *Salvadora persica*, 5 - *Scutia myrtina*, 6 - *Solanum dubium*, 7 - *Sporobolus consimilis*, 8 - *Suaeda monoica*, 9 - Grass B, 10 - Grass C, 11 - Grass D.

The *trn*L (UAA) intron was amplified in 76 out of 80 plants (95%). This locus was approximately 500 bp in length (Figure 4.4).

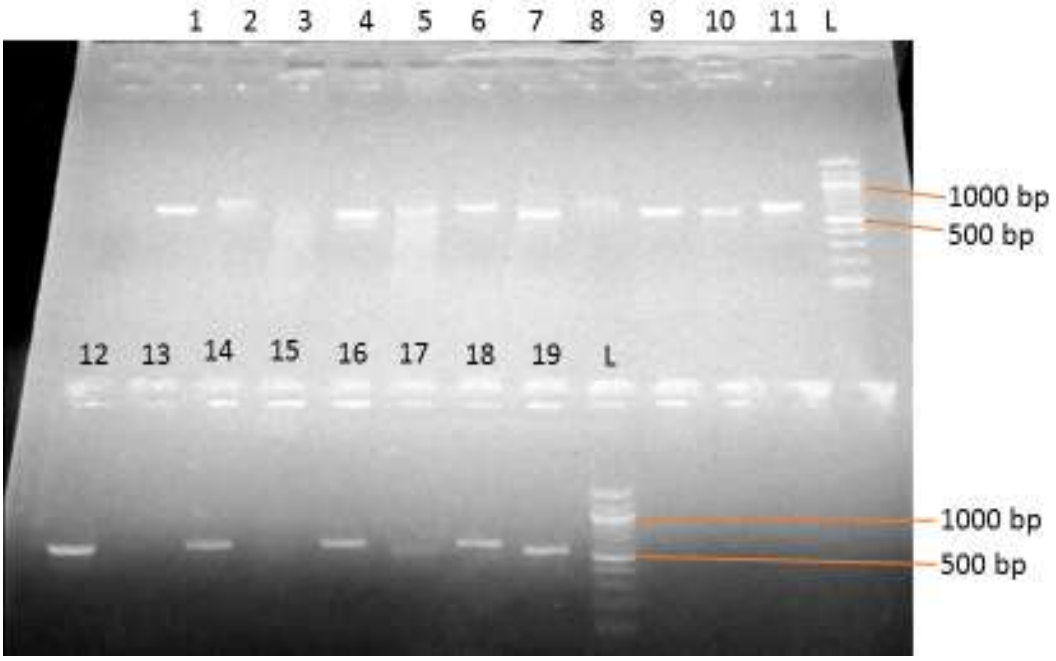


Figure 4.4: An agarose gel image for the PCR amplicons of *trn*L(UAA) intron.

Key: - Gel lanes: 1 - *Acacia xanthophloea*, 2 - *Achyranthes aspera*, 3 - *Azima tetracantha*, 4 -*Balanites pedicellaris*, 5 - *Capparis tomentosa*, 6 - *Commicarpus plumbagineus*, 7 - *Cordia monoica*, 8 - *Dasyphaera prostrata*, 9 - *Euclea schimperi*, 10 - *Ficus* sp., 11 - *Odyssea paucinervis*, 12 -*Rhamphicarpa montana*, 13 - *Rhus natalensis*, 14 - *Salvadora persica*, 15 - *Scutia myrtina*, 16 -*Setaria verticillata*, 17 - *Trianthema ceratosepala*, 18 - Grass A, 19 - *Withania somnifera*, L - 100 bp NEB ladder.

The *ITS*1-Poaceae region was amplified in 36 out of the 40 monocotyledons (90%). This locus was approximately 100 bp in length (Figure 4.5).

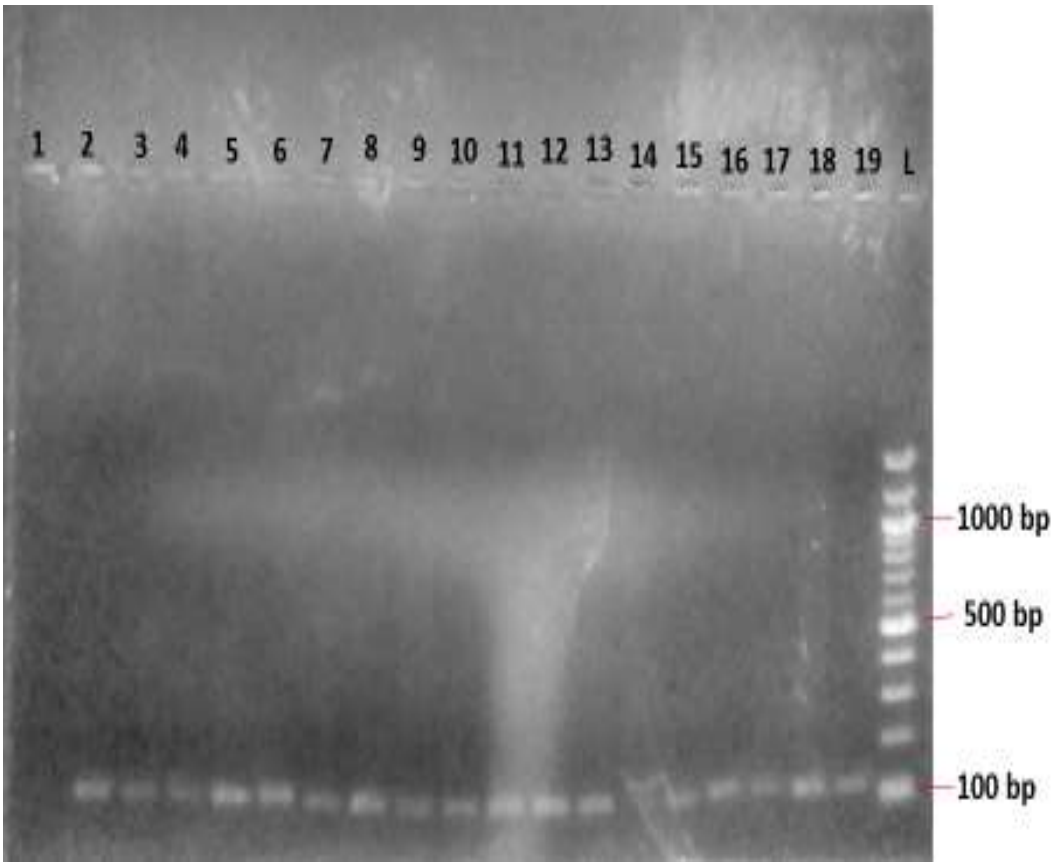


Figure 4.5: An agarose gel image for the PCR amplicons of *ITS*1-Poaceae region.

Key: - Gel lanes: 1 - Grass G; 2 - Grass H; 3 - Grass I; 4 - Grass K; 5 - Grass L; 6 - Grass M; 7 - Grass N; 8 - Grass O; 9 - Grass P; 10 - Grass Q; 11 - Grass R; 12 - Grass AA; 13 - Grass AB; 14 - Grass AC; 15 - Grass AD; 16 - Grass AE; 17 - Grass AF; 18 - Grass AG; 19 - Grass AH; L - 100 bp NEB ladder.

The 18S rDNA region was amplified in 64 out of 78 plants (82.05%). This locus was approximately 100 bp in length (Figure 4.6).

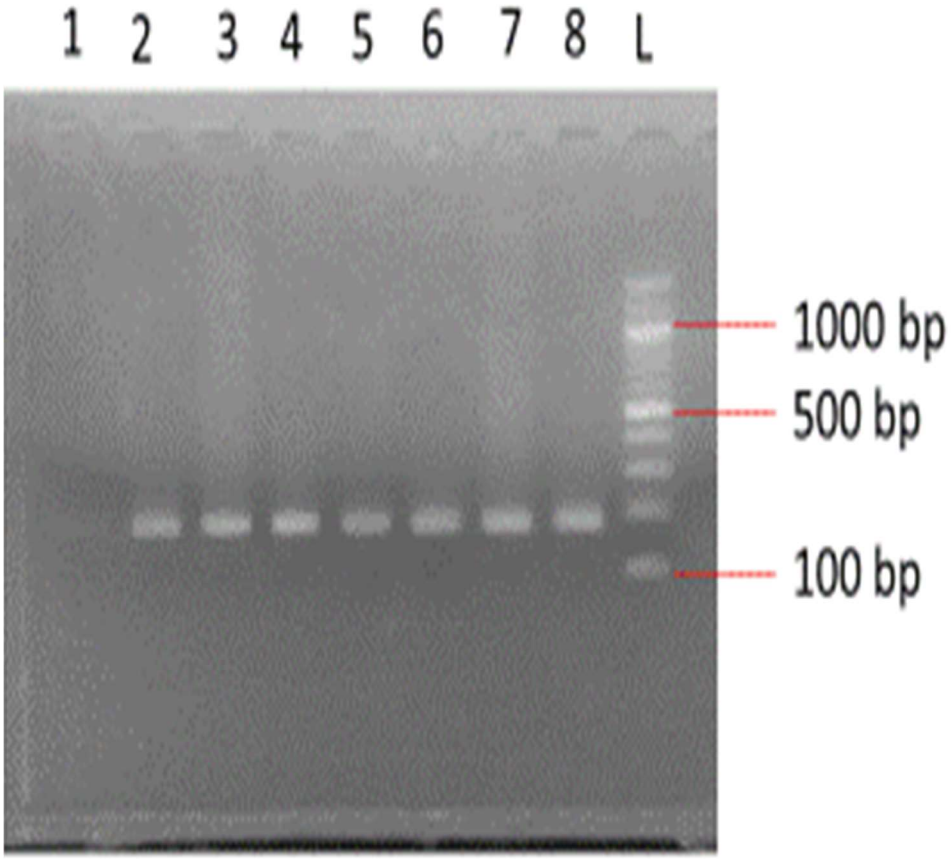


Figure 4.6: An agarose gel image for the PCR amplicons of 18S rDNA region.

Key: - Gel lanes: 1 - Negative control; 2 – Plant B; 3 – *Abutilon mauritianum*; 4 - Grass AF; 5 - Grass AG; 6 - Grass AH; 7 - Grass AI; 8 - Grass AJ, L- 100 bp NEB ladder.

The *trn*L-P6 region was amplified in 33 out of 40 plants (80.49%). This locus was approximately 100bp in length (Figure 4.7).

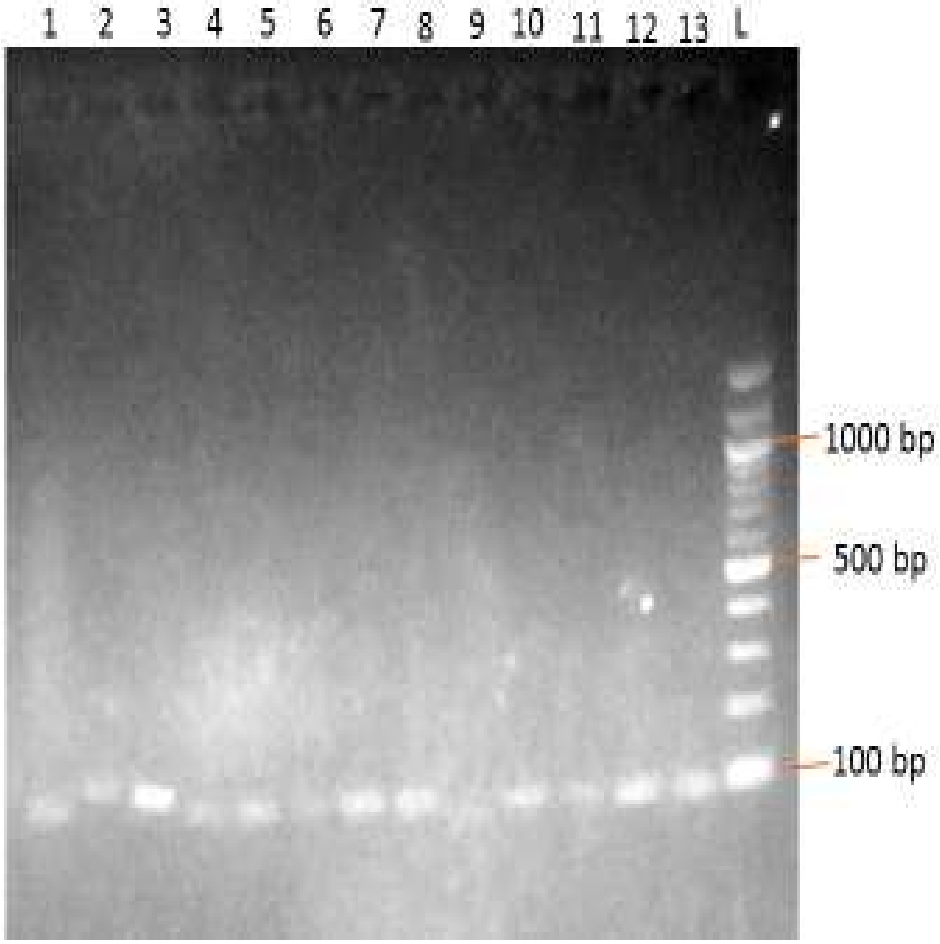


Figure 4.7: An agarose gel image for the PCR amplicons of *trn*L-P6 region.

Key: - Gel lanes: 1 - *Cordia monoica,* 2 - *Dasyphaera prostrata,* 3 - *Hibiscus lila,* 4 - *Maerua angolensis,* 5 - *Maerua crassifolia,* 6 - *Rhus natalensis,* 7 - *Setaria verticillata,* 8 - *Sporobolus consimilis,* 9 - *Suaeda monoica,* 10 - Grass B, 11 - Grass C, 12 - Grass D, 13 - Grass J, L - 100 bp NEB ladder.

4.3 Sanger sequencing

Out of the 289 amplicons sent for sequencing (Table 4.1), high-quality sequences were generated for 182 samples (Table 4.3). Specifically, the *ITS*1 gene had 71 sequences; the *trn*L (UAA) intron had 64 sequences; the *ITS*1-Poaceae barcode had 19 sequences; the 18S rDNA region had 24 sequences; and lastly, the *trn*L-P6 locus had 4 sequences (Table 4.3).

Table 4.3: High-quality DNA sequences for monocotyledons and dicotyledons generated for each candidate barcode using Sanger sequencing

Barcode	**Monocotyledons**	**Dicotyledons**	**Total**
*ITS*1	37	33	**71**
*trn*L intron	31	34	**64**
18S	15	9	**24**
*ITS*1-Poaceae	19	N/A	**19**
*trn*L-P6	2	2	**4**
Total	**104**	**78**	**182**

KEY: N/A – The barcode was specific for the Poaceae family (monocotyledons).

4.3.1 Length distribution of sequences

The length in base pairs of bidirectional sequences of the highest quality ranged from 90 - 593 bp and varied across the different barcoding loci (Table 4.4).

Table 4.4: Length distribution of DNA sequences generated for each barcode

Barcode	**Longest sequence (bp)**	**Shortest sequence (bp)**	**Median length (bp)**
*trn*L intron	593	389	572
*ITS*1	383	252	310
18S rDNA	184	179	180
*ITS*1-Poaceae	121	113	114
*trn*L-P6	109	90	92

The longest sequence for the *trn*L(UAA) intron (593 bp) belonged to Grass C, while the shortest (389 bp) were for grass samples AM, E and I (Table 4.4). The longest *ITS*1 sequence (383 bp) was produced for *Hibiscus* 'lila', while the shortest (252 bp) belonged to *Acacia xanthophloea* (Table 4.4). The longest sequences for the 18S rDNA locus (184 bp) were for grass samples AI and F, while the shortest (179 bp) belonged to both *Azima tetracantha* and grass E2 (Table 4.4). For the *ITS*1-Poaceae locus, the longest sequence (121 bp) belonged to grass AP, while the shortest (113 bp) were for grasses C, T, and AN (Table 4.4). Lastly, the longest sequence for the *trn*L-P6 region (109 bp) belonged to *Abutilon mauritianum*, while the shortest sequence (90 bp) was for *Acacia xanthophloea* (Table 4.4).

4.4 Species identification

4.4.1 Sequence similarity analysis based on BLAST algorithm

4.4.1.1 The identification of monocotyledons based on BLAST analysis

In monocotyledons, more sequences were successfully identified to the genus level using GenBank® (77.88%; Table 4.5) than by the BOLD database (62.50%; Table 4.5). Furthermore, the top BLAST hits for 18.27% of the sequences queried in GenBank® (Table 4.5) identified multiple genera, whereas this was the case for only 1.92% of the sequences queried in BOLD database. Lastly, the top BLAST hits for 35.58% of the sequences queried in BOLD databases were below the threshold (which was set at 95%). Such identities occurred for only 3.85% of the sequences queried in GenBank® (Table 4.5).

Using GenBank® (Table 4.5), the barcodes with the most successful identification rate at the genus level were both *ITS*1 (100%) and *trn*L-P6 (100%), followed by *ITS*1-Poaceae (94.74%), *trn*L(UAA) intron (74.19%), and lastly the 18S rDNA locus (6.67%). In comparison, based on the BOLD database (Table 4.5), the barcode with the most successful identification rate at the genus level was *trn*L-P6 (100%), followed by the *trn*L(UAA) intron (83.87%), *ITS*1 (70.27%), *ITS*1-Poaceae (57.89%), and lastly, the 18S rDNA locus, which had no successful identification.

Table 4.5: BLAST analysis of monocotyledons' DNA sequences for identification at the genus level

GenBank – Genus				
Barcode	**Sample size**	**Successful identifications**	**Ambiguous identifications**	**Unidentifiable**
*ITS*1	37	37 (100%)	0	0
*trn*L-P6	2	2 (100%)	0	0
*ITS*1-Poaceae	19	18 (94.74%)	0	1 (5.26%)
*trn*L(UAA) intron	31	23 (74.19%)	5 (16.13%)	3 (9.68%)
18S	15	1 (6.67%)	14 (93.33%)	0
Total:	**104**	**81 (77.88%)**	**19 (18.27%)**	**4 (3.85%)**
BOLD – Genus				
Barcode	**Sample size**	**Successful identifications**	**Ambiguous identifications**	**Unidentifiable**
*trn*L-P6	2	2 (100%)	0	0
*trn*L(UAA) intron	31	26 (83.87%)	2 (6.45%)	3 (9.68%)
*ITS*1	37	26 (70.27%)	0	11 (29.73%)
*ITS*1-Poaceae	19	11 (57.89%)	0	8 (42.11%)
18S	15	0	0	15 (100%)
Total:	**104**	**65 (62.50%)**	**2 (1.92%)**	**37 (35.58%)**

With regards to the identification at the species-level, more sequences were successfully identified using GenBank® (60.58%; Table 4.6) than the BOLD database (59.62%; Table 4.6). Furthermore, the top BLAST hits for 35.58% of the sequences queried in GenBank® (Table 4.6) identified multiple species, whereas this was the case for only 4.81% of the sequences queried in BOLD database. Lastly, the top BLAST hits for 35.58% of the sequences queried in BOLD databases were below the threshold (which was set at 95%). Such identities occurred for only 3.85% of the sequences queried in GenBank® (Table 4.6).

Using GenBank®, the barcode with the most successful identification rate at the species level was *ITS*1 (89.19%), followed by *trn*L(UAA) intron (74.19%), *ITS*1-Poaceae (31.58%), the 18S rDNA locus (6.67%), and lastly, *trn*L-P6 that had no correct identifications. In comparison, the barcode with the most successful identification rate at the species level in BOLD database (Table 4.6) was *trn*L-P6 (100%), followed by the *trn*L(UAA) intron (77.42%), *ITS*1 (70.27%), *ITS*1-Poaceae (52.63%), and lastly, the 18S rDNA locus, which had no successful identification.

Table 4.6: BLAST analysis of monocotyledons' DNA sequences for identification at the species level

GenBank – Species				
Barcode	**Sample size**	**Successful identifications**	**Ambiguous identifications**	**Unidentifiable**
*ITS*1	37	33 (89.19%)	4 (10.81%)	0
*trn*L(UAA) intron	31	23 (74.19%)	5 (16.13%)	3 (9.68%)
*ITS*1-Poaceae	19	6 (31.58%)	12 (63.16%)	1 (5.26%)
18S	15	1 (6.67%)	14 (93.33%)	0
*trn*L-P6	2	0	2 (100%)	0
Total:	**104**	**63 (60.58%)**	**37 (35.58%)**	**4 (3.85%)**
BOLD – Species				
Barcode	**Sample size**	**Successful identifications**	**Ambiguous identifications**	**Unidentifiable**
*trn*L-P6	2	2 (100%)	0	0
*trn*L(UAA) intron	31	24 (77.42%)	4 (12.90%)	3 (9.68%)
*ITS*1	37	26 (70.27%)	0	11 (29.73%)
*ITS*1-Poaceae	19	10 (52.63%)	1 (5.26%)	8 (42.11%)
18S	15	0	0	15 (100%)
Total:	**104**	**62 (59.62%)**	**5 (4.81%)**	**37 (35.58%)**

Generally, more monocotyledons were identified at both the genus and species levels using GenBank® (77.88% and 60.58%, respectively; Tables 4.5 and 4.6) than by BOLD database (62.50% and 59.62%, respectively; Tables 4.5 and 4.6).

4.4.1.2 The identification of dicotyledons based on BLAST analysis

In dicotyledons, more sequences were successfully identified to the genus level using GenBank® (71.79%; Table 4.7) than by BOLD database (53.85%; Table 4.7). Furthermore, the top BLAST hits for 14.10% of the sequences queried in GenBank® identified multiple genera, whereas there was no such occurrence in the BOLD database (Table 4.7). Lastly, the top BLAST hits for 46.15% of the sequences queried in BOLD databases were below the threshold (which was set at 95%). Such identities occurred for 14.10% of the sequences queried in GenBank® (Table 4.7).

Using GenBank® (Table 4.7), the barcodes with the most successful identification rate at the genus level were *trn*L(UAA) intron (84.85%), followed by *ITS*1 (70.59%), *trn*L-P6 (50%), and lastly, the 18S rDNA locus (33.33%). In comparison, based on the BOLD database (Table 4.7), the barcode with the most successful identification rate at the genus level was *trn*L-P6 (100%), followed by the *trn*L(UAA) intron (81.82%), *ITS*1 (38.24%), and lastly, the 18S rDNA locus, which had no successful identification.

Table 4.7: BLAST analysis of dicotyledons' DNA sequences for identification at the genus level

GenBank – Genus				
Barcode	**Sample size**	**Successful identification**	**Ambiguous identification**	**Unidentifiable**
*trn*L(UAA) intron	33	28 (84.85%)	4 (12.12%)	1 (3.03%)
*ITS*1	34	24 (70.59%)	0	10 (29.41%)
*trn*L-P6	2	1 (50%)	1 (50%)	0
18S	9	3 (33.33%)	6 (66.67%)	0
Total:	**78**	**56 (71.79%)**	**11 (14.10%)**	**11 (14.10%)**
BOLD – Genus				
Barcode	**Sample size**	**Successful identification**	**Ambiguous identification**	**Unidentifiable**
*trn*L-P6	2	2 (100%)	0	0
*trn*L(UAA) intron	33	27 (81.82%)	0	6 (18.18%)
*ITS*1	34	13 (38.24%)	0	21 (61.76%)
18S	9	0	0	9 (100%)
Total:	**78**	**42 (53.85%)**	**0**	**36 (46.15%)**

With regards to the identification at the species-level, more sequences were successfully identified using GenBank® (69.23%; Table 4.8) than the BOLD database (53.85%; Table 4.8). Furthermore, the top BLAST hits for 16.67% of the sequences queried in GenBank® (Table 4.8) identified multiple species, whereas this was the case for only 3.85% of the sequences queried in BOLD database. Lastly, the top BLAST hits for 42.31% of the sequences queried in BOLD databases were below the threshold (which was set at 95%). Such identities occurred for only 14.10% of the sequences queried in GenBank® (Table 4.8).

Using GenBank® (Table 4.8), the barcode with the most successful identification rate at the species level was *trn*L (UAA) intron (81.82%), followed by *ITS*1 (67.65%), *trn*L-P6 (50%), and lastly, the 18S rDNA locus (33.33%). In comparison, the barcode with the most successful identification rate at the species level in the BOLD database (Table 4.8) was *trn*L (UAA) intron (78.79%), followed by *ITS*1 (47.06%), and lastly, both the 18S rDNA locus and *trn*L-P6, which had no successful identifications.

Table 4.8: BLAST analysis for dicotyledons' DNA sequences for identification at the species level

GenBank – Species				
Barcode	**Sample size**	**Successful identification**	**Ambiguous identification**	**Unidentifiable**
*trn*L(UAA) intron	33	27 (81.82%)	5 (15.15%)	1 (3.03%)
*ITS*1	34	23 (67.65%)	1 (2.94%)	10 (29.41%)
*trn*L-P6	2	1 (50%)	1 (50%)	0
18S	9	3 (33.33%)	6 (66.67%)	0
Total:	**78**	**54 (69.23%)**	**13 (16.67%)**	**11 (14.10%)**
BOLD – Species				
Barcode	**Sample size**	**Successful identification**	**Ambiguous identification**	**Unidentifiable**
*trn*L(UAA) intron	33	26 (78.79%)	1 (3.03%)	6 (18.18%)
*ITS*1	34	16 (47.06%)	0	18 (52.94%)
18S	9	0	0	9 (100%)
*trn*L-P6	2	0	2 (100%)	0
Total:	**78**	**42 (53.85%)**	**3 (3.85%)**	**33 (42.31%)**

Generally, more dicotyledons were identified at both the genus and species levels using GenBank® (71.79% and 69.23%, respectively; Tables 4.7 and 4.8) than by BOLD database (53.85% and 53.85%, respectively; Tables 4.7 and 4.8).

4.4.2 Tree-based identification

The best-fit substitution model for each alignment (Table 4.9) was produced in MEGA X (http://www.megasoftware.net, Kumar *et al.*, 2018) before building the phylogenetic trees.

Table 4.9: The best-fit nucleotide substitution model for each candidate barcode

Alignment	**Substitution model, rates and patterns**
*ITS*1 monocotyledons	K2+G
*ITS*1-Poaceae	K2+G
*trn*L monocotyledons	T92+G
*trn*L-P6 monocotyledons	JC
18S rDNA monocotyledons	K2+G
*ITS*1 dicotyledons	K2+G
*trn*L dicotyledons	T92+G
*trn*L-P6 dicotyledons	JC
18S rDNA dicotyledons	K2+G+I

Abbreviations*:* K2: Kimura 2-parameter; G: Gamma distribution; I: invariable sites; T92: Tamura 3-parameter; JC: Jukes-Cantor.

The trees generated by *ITS*1, *ITS*1-Poaceae, *trn*L(UAA) intron, *trn*L-P6, and the 18S rDNA locus for monocotyledons and dicotyledons are shown in Figures 4.8 – 4.16 and appendices 4 – 12.

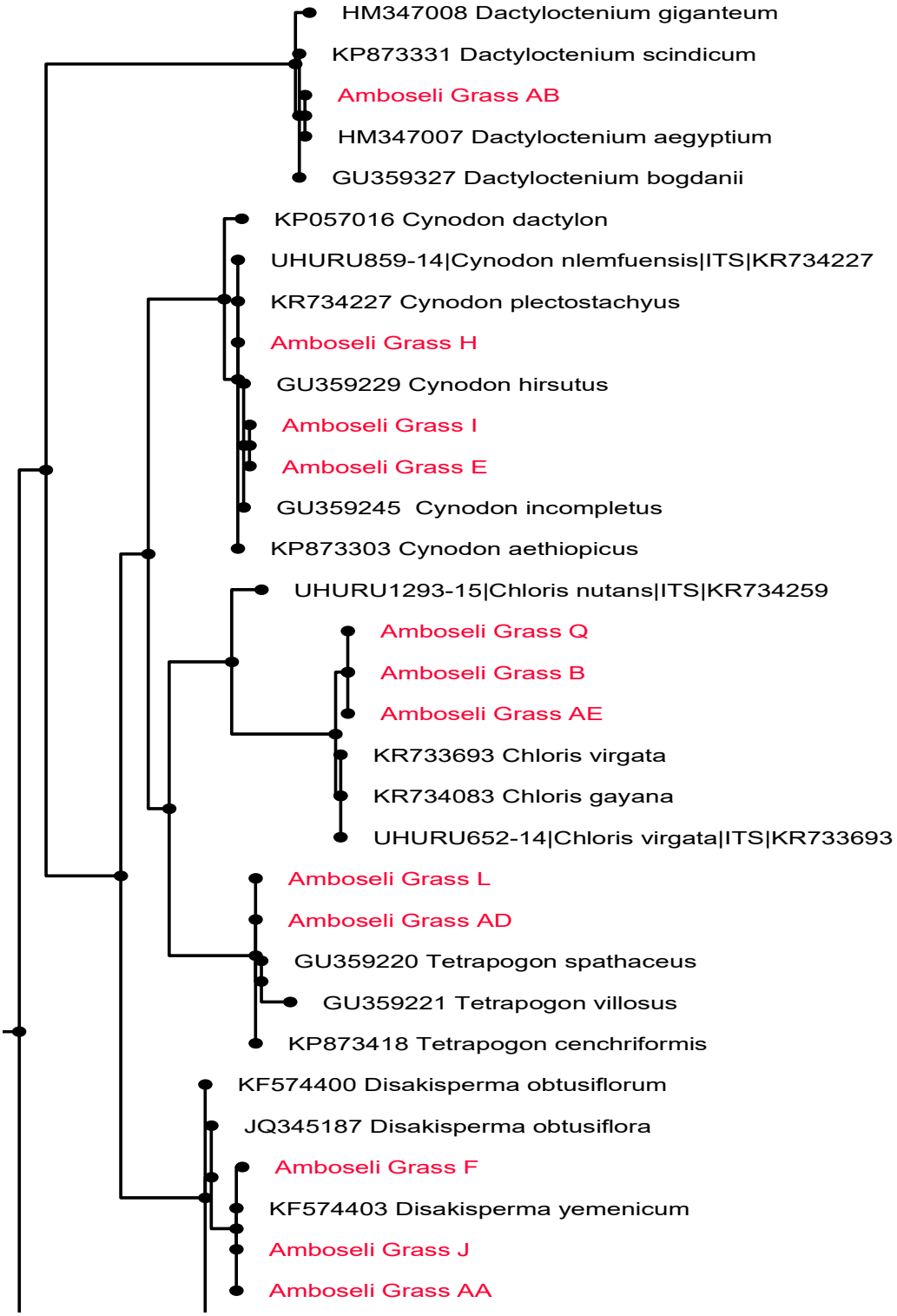


Figure 4.8: Phylogenetic tree for monocotyledons based on the *ITS*1 gene.

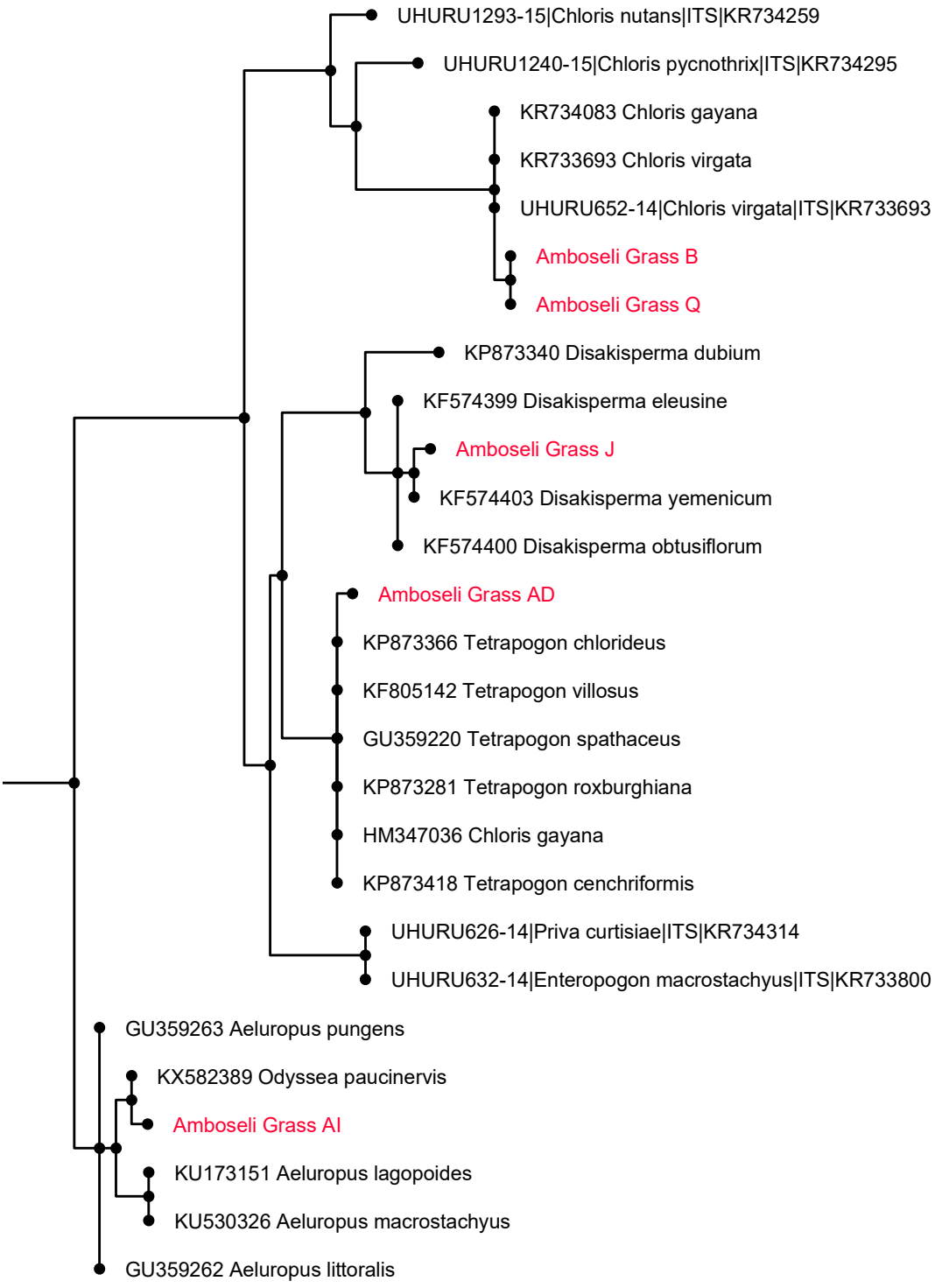


Figure 4.9: Phylogenetic tree for monocotyledons based on the *ITS*1-Poaceae barcode.

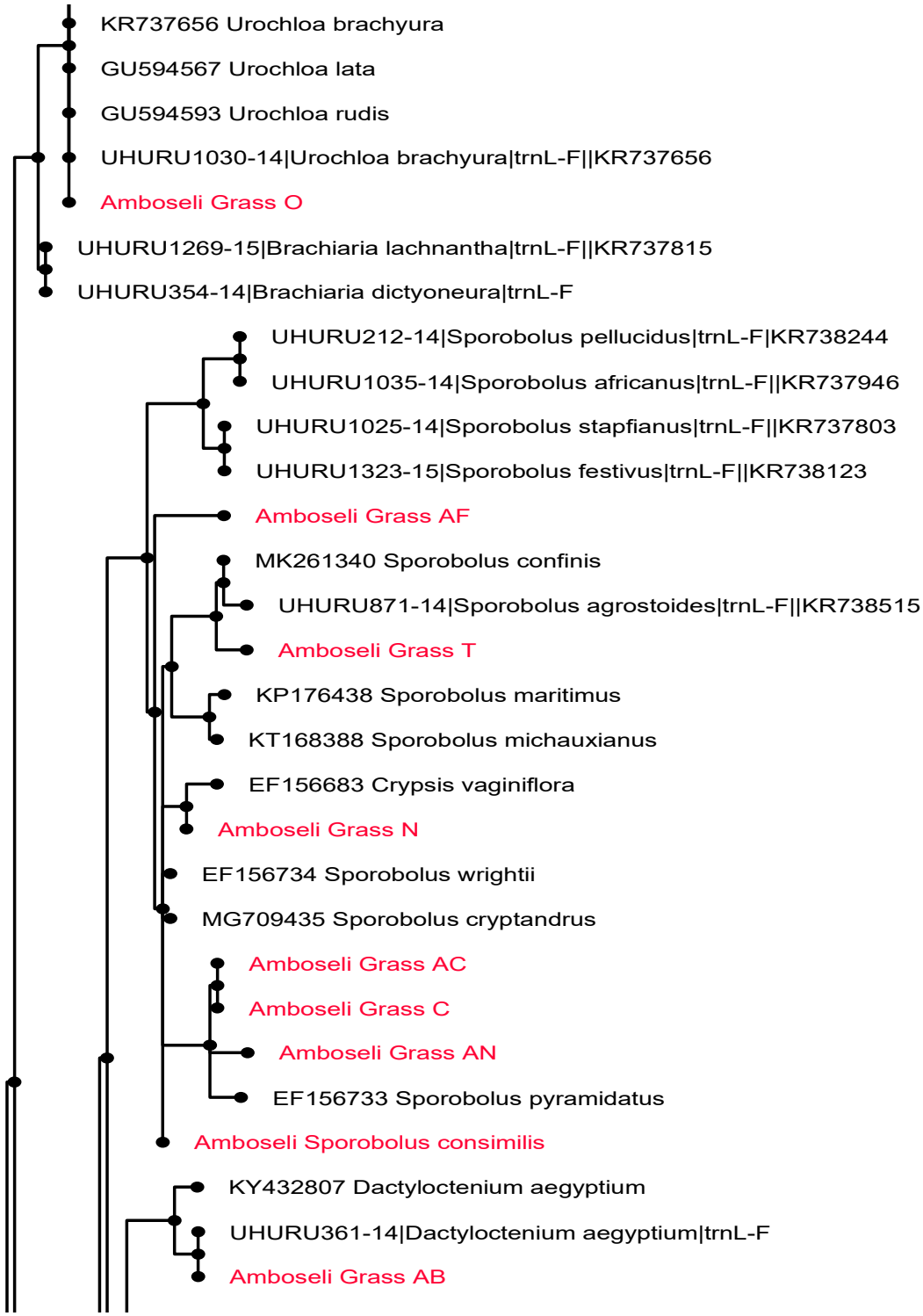


Figure 4.10: Phylogenetic tree for monocotyledons based on the *trn*L(UAA) intron.

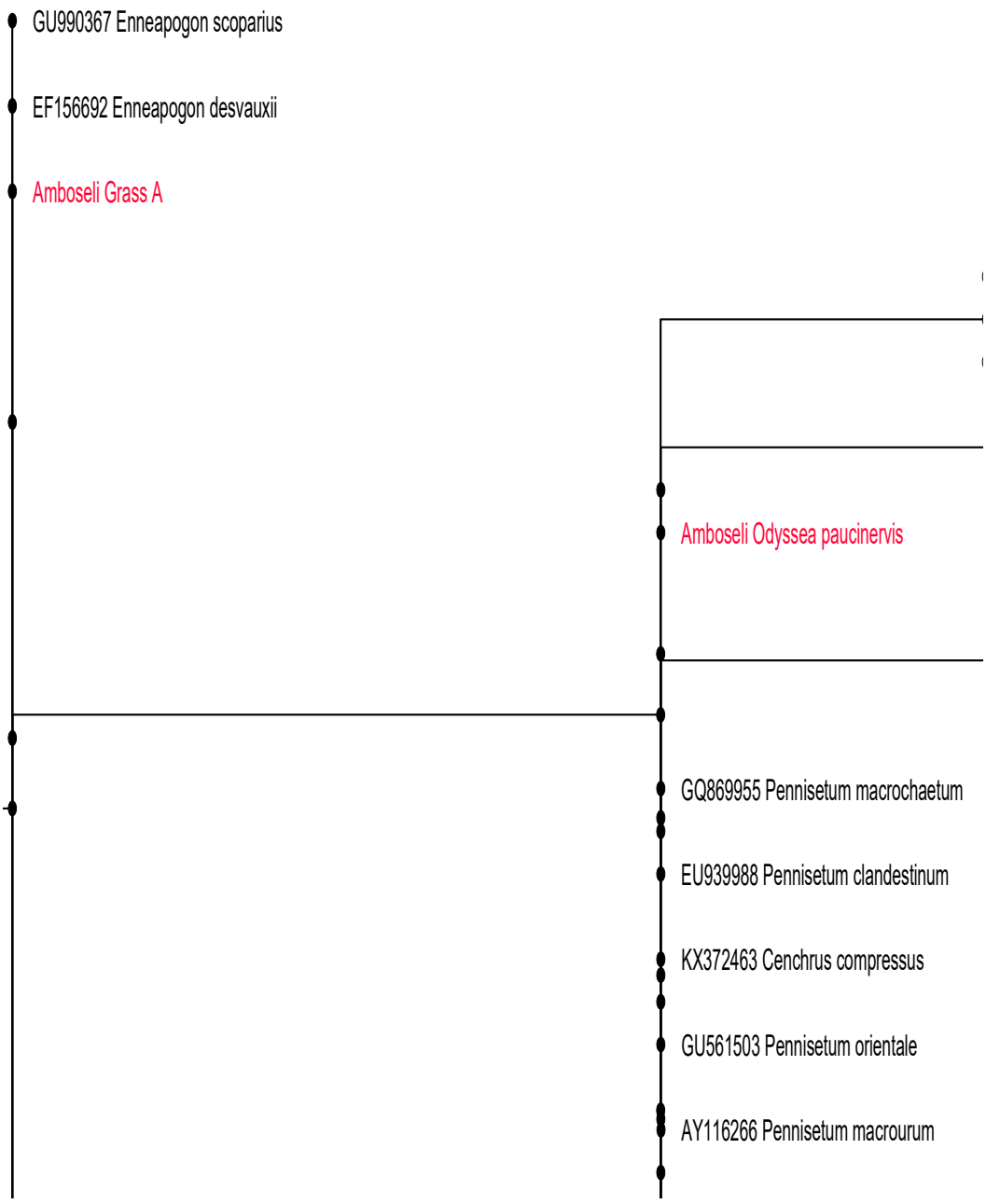


Figure 4.11: Phylogenetic tree for monocotyledons based on the *trn*L-P6 barcode.

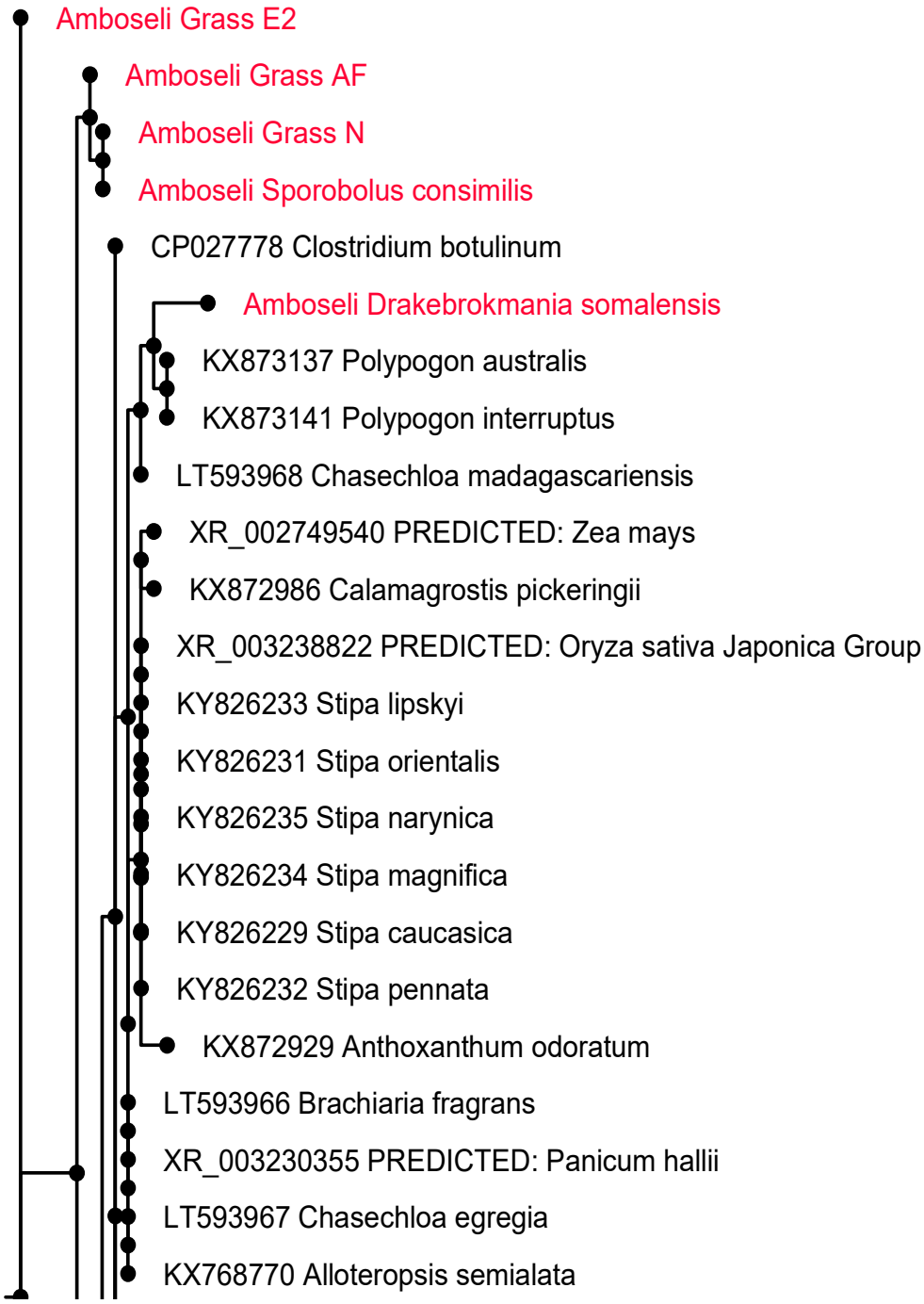


Figure 4.12: Phylogenetic tree for monocotyledons based on the 18S rDNA region.

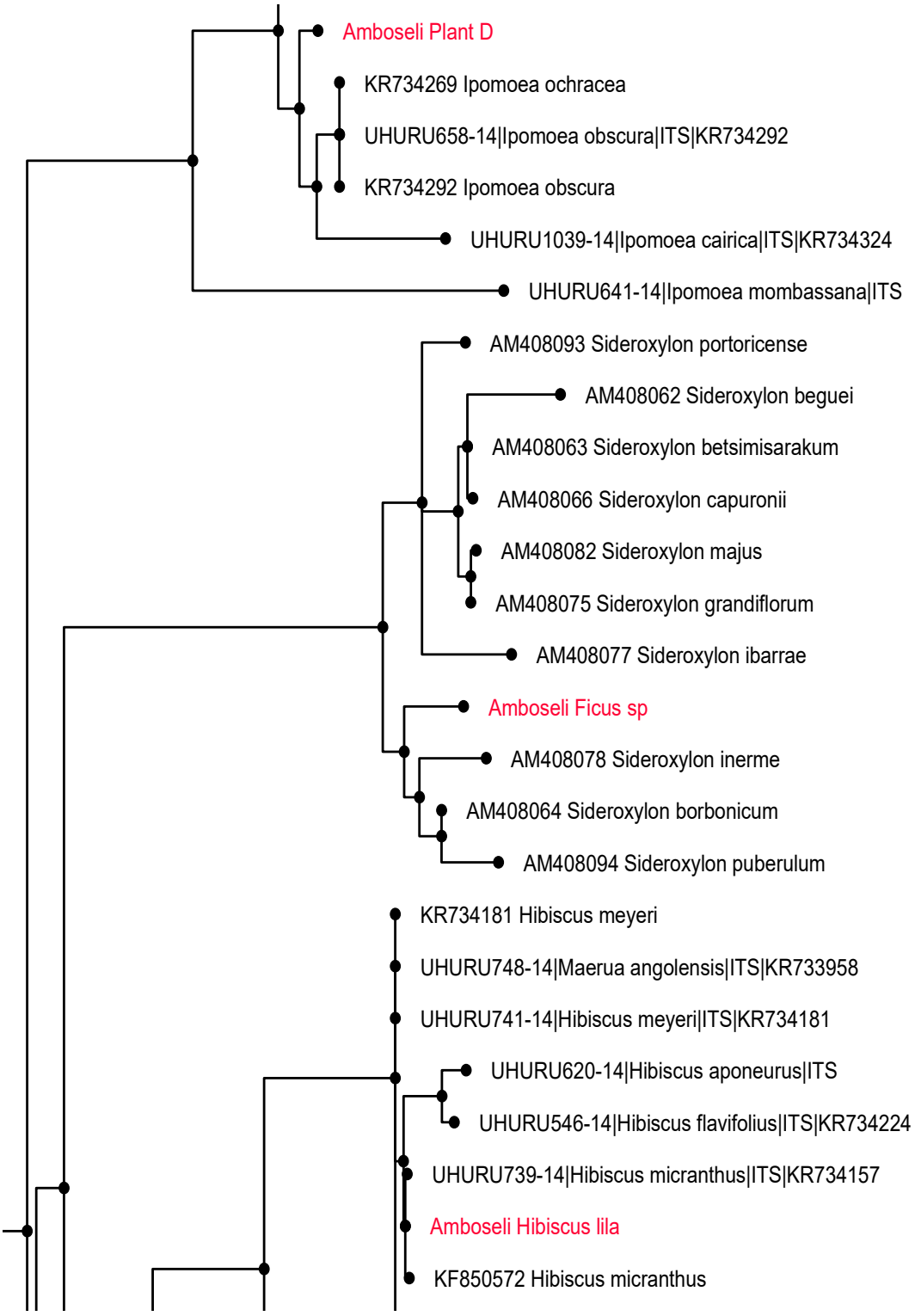


Figure 4. 13: Phylogenetic tree for dicotyledons based on the *ITS*1 gene.

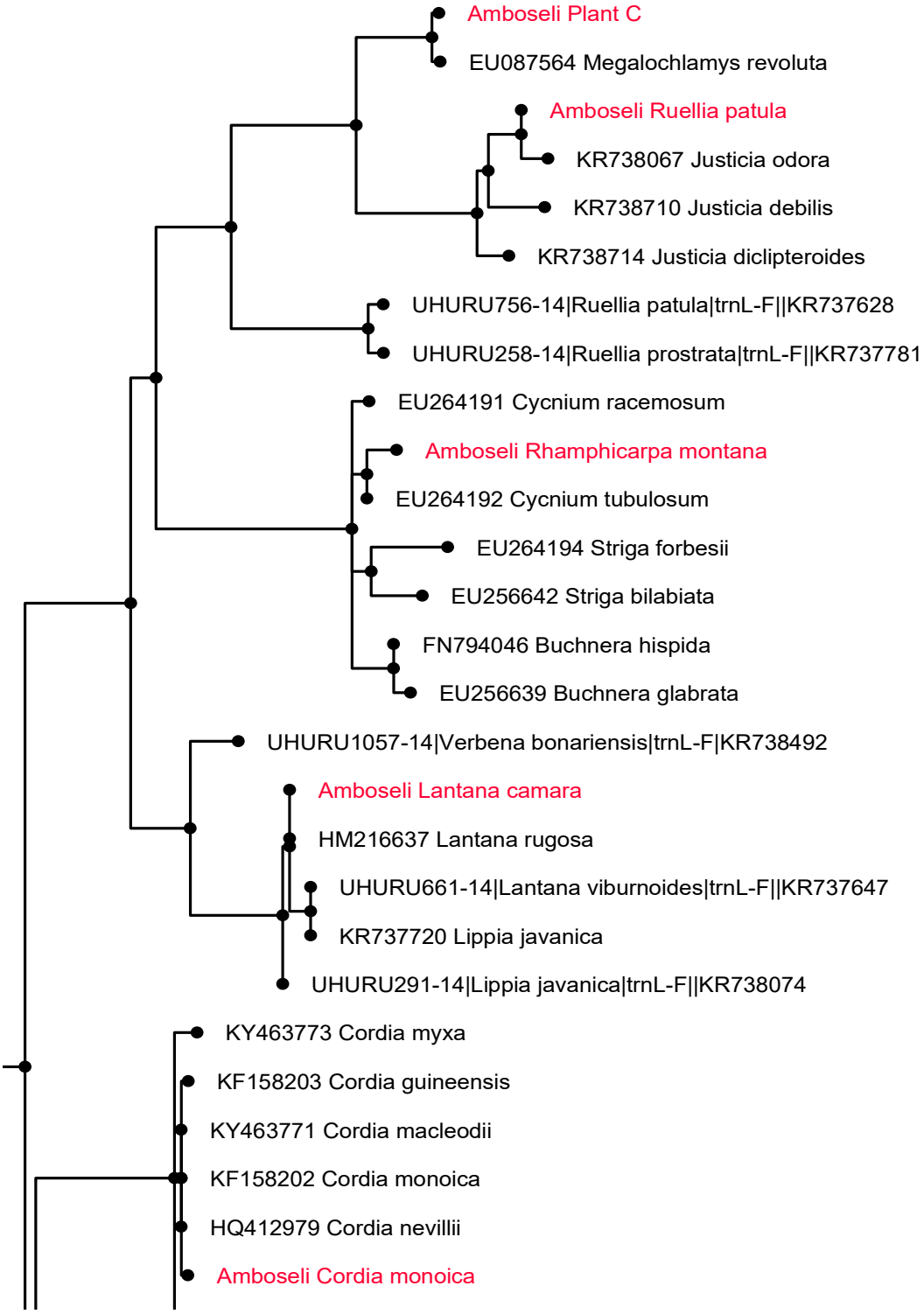


Figure 4.14: Phylogenetic tree for dicotyledons based on the *trn*L(UAA) intron.

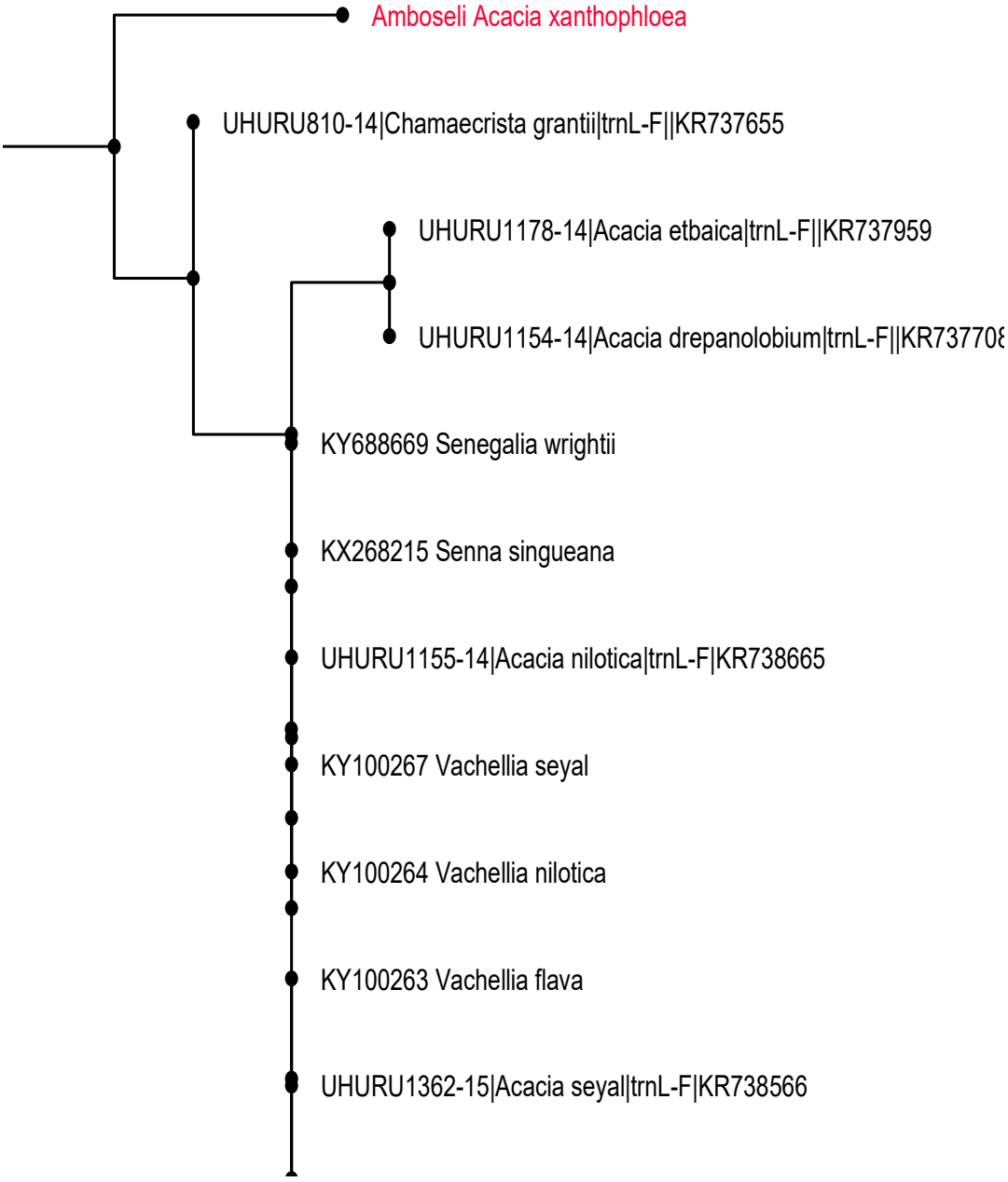


Figure 4.15: Phylogenetic tree for dicotyledons based on the *trn*L-P6 barcode.

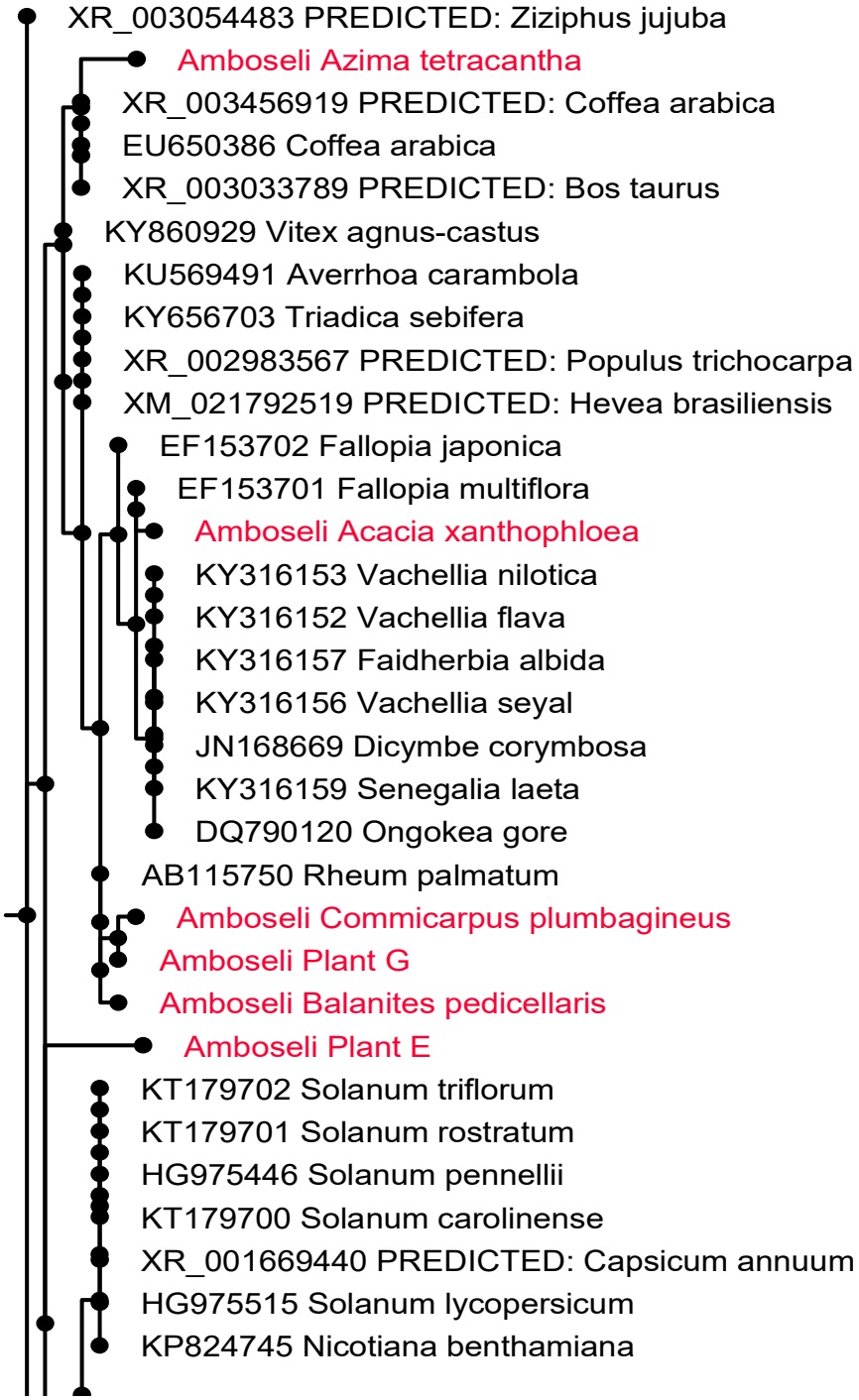


Figure 4.16: Phylogenetic tree for dicotyledons based on the 18S rDNA region.

4.4.2.1 The identification of monocotyledons based on phylogenetic analysis

In monocotyledons, successful identifications at the genus level occurred for 76.92% of the samples, while 38.46% were distinguished to the species taxa (Table 4.10). However, 1.92% of the samples were assigned to multiple genera; hence, their identities were ambiguous (Table 4.10), which was also the case for 40.38% of the samples that were identified at the species level (Table 4.10). Lastly, 21.25% of the samples could not be identified because the query sequences were not related to any clade or polytomy on the phylogenetic trees (Figures 4.8 – 4.16).

With reference to the individual barcodes (Table 4.10), the *ITS*1 gene had the highest identification rate at the genus level (100%), followed by *ITS*1-Poaceae (89.47%), the *trn*L(UAA) intron (77.42%), *trn*L-P6 (50%), and lastly, the 18S rDNA locus (6.67%). For species determination (Table 4.10), the *ITS*1 barcode still had the highest identification success rate (59.46%), followed by the *trn*L(UAA) intron (45.16%), *ITS*1-Poaceae (21.05%), and lastly, both *trn*L-P6 and the 18S rDNA locus that had no successful identification.

Table 4.10: Identification of monocotyledons based on maximum-likelihood tree analysis

Genus				
Barcode	**Sample size**	**Successful identification**	**Ambiguous identification**	**Unidentifiable**
*ITS*1	37	37 (100%)	0	0
*ITS*1-Poaceae	19	17 (89.47%)	1 (5.26%)	1 (5.26%)
*trn*L(UAA) intron	31	24 (77.42%)	1 (3.23%)	6 (19.35%)
*trn*L-P6	2	1 (50 %)	0	1 (50%)
18S	15	1 (6.67%)	0	14 (93.33%)
Total:	**104**	**80 (76.92%)**	**2 (1.92%)**	**22 (21.15%)**
Species				
Barcode	**Sample size**	**Successful identification**	**Ambiguous identification**	**Unidentifiable**
*ITS*1	37	22 (59.46%)	15 (40.54%)	0
*trn*L(UAA) intron	31	14 (45.16%)	11 (35.48%)	6 (19.35%)
*ITS*1-Poaceae	19	4 (21.05%)	14 (73.68%)	1 (5.26%)
18S	15	0	1 (6.67%)	14 (93.33%)
*trn*L-P6	2	0	1 (50%)	1 (50%)
Total:	**104**	**40 (38.46%)**	**42 (40.38%)**	**22 (21.15%)**

4.4.2.2 The identification of dicotyledons based on phylogenetic analysis

In dicotyledons, successful identifications at the genus level occurred for 80.77% of the samples, while 58.97% were distinguished to the species taxa (Table 4.11). However, 3.85% of the samples were assigned to multiple genera; hence, their identities were ambiguous (Table 4.11), which was also the case for 25.64% of the samples that were identified at the species level (Table 4.11). Lastly, 15.38% of the samples could not be identified because the query sequences were not related to any clade or polytomy on the phylogenetic trees (Figures 4.8 – 4.16).

Regarding the individual barcodes (Table 4.11), the *trn*L(UAA) intron had the highest identification rate at the genus level (93.94%), followed by the *ITS*1 gene (91.18%), *trn*L-P6 (50%), and lastly, the 18S rDNA locus that had no successful identification. For species determination (Table 4.11), the *trn*L(UAA) intron still had the highest identification rate (69.70%), followed by the *ITS*1 gene (67.95%), and lastly, both *trn*L-P6 and 18S rDNA locus that had no successful identification.

.

Table 4.11: Identification of dicotyledons based on maximum-likelihood tree analysis

Genus				
Barcode	**Sample size**	**Successful identification**	**Ambiguous identification**	**Unidentifiable**
*trn*L(UAA) intron	33	31(93.94%)	1 (3.03%)	1 (3.03%)
*ITS*1	34	31 (91.18%)	1 (2.94%)	2 (5.88%)
*trn*L-P6	2	1 (50%)	1 (50%)	0
18S	9	0	0	9 (100%)
Total:	**78**	**63 (80.77%)**	**3 (3.85%)**	**12 (15.38%)**
Species				
Barcode	**Sample size**	**Successful identification**	**Ambiguous identification**	**Unidentifiable**
*trn*L(UAA) intron	33	23 (69.70%)	9 (27.27%)	1 (3.03%)
*ITS*1	34	23 (67.65%)	9 (26.47%)	2 (5.88%)
18S	9	0	0	9 (100%)
*trn*L-P6	2	0	2 (100%)	0
Total:	**78**	**46 (58.97%)**	**20 (25.64%)**	**12 (15.38%)**

4.5 Consensus identities based on morphological and molecular analyses

Sixty-six out of 80 plants of interest (82.50%) were identified using both molecular and morphological studies (Table 4.12). Specifically, 50 plants were determined to only the genus level, while 16 samples were further distinguished to the species level (Table 4.12).

Table 4.12: Monocotyledons and dicotyledons successfully identified at both the genus and species levels

	Genus-level only	Species	**Total**
Monocotyledons	30/40 (75%)	3/40 (7.50%)	33/40 (82.50%)
Dicotyledons	20/40 (50%)	13/40 (32.50%)	33/40 (82.50%)
Total	**50/80 (62.50%)**	**16/80 (20%)**	**66/80 (82.50%)**

Thirty monocotyledons were only identified to the genus-level, and three grasses were further distinguished to the species rank (Tables 4.12 and 4.13). The identities of six monocotyledons were ambiguous because multiple genera were identified for each specimen (Appendix 1). For instance, the BLAST results for Grass AD based on the *ITS*1 gene identified it as a *Chloris* sp., while the *trn*L (UAA) intron matched it with multiple genera, including *Cynodon*, *Astrebla*, *Chloris* and *Enteropogon* spp. (Appendix 1). Furthermore, 'Grass AM' was not identified because the results did not reach the required threshold stated in the data analyses.

Table 4.13: Consensus scientific names of monocotyledons based on both morphological and molecular analyses.

No	Sample ID	Consensus ID	Family
1	*Drake-brockmania somalensis*	Ambiguous	Poaceae
2	*Odyssea paucinervis*	Ambiguous	Poaceae
3	*Setaria verticillata*	*Setaria verticillata*	Poaceae
4	*Sporobolus consimilis*	*Sporobolus* sp.	Poaceae
5	Grass A	*Enneapogon sp.*	Poaceae
6	Grass AA	*Disakisperma* sp.	Poaceae
7	Grass AB	*Dactylotenium aegyptium*	Poaceae
8	Grass AC	*Sporobolus* sp.	Poaceae
9	Grass AD	Ambiguous	Poaceae
10	Grass AE	*Chloris* sp.	Poaceae
11	Grass AF	*Sporobolus* sp.	Poaceae
12	Grass AG	*Pennisetum/Cenchrus* sp.	Poaceae
13	Grass AH	*Enneapogon* sp.	Poaceae
14	Grass AI	*Odyssea paucinervis*	Poaceae
15	Grass AJ	*Sporobolus* sp.	Poaceae
16	Grass AK	*Cyperus* sp.	Cyperaceae
17	Grass AL	*Cyperus* sp.	Cyperaceae
18	Grass AM	Unidentifiable	Unidentifiable
19	Grass AN	*Sporobolus* sp.	Poaceae
20	Grass AP	*Eragrostis* sp.	Poaceae
21	Grass AR	*Cyperus* sp.	Cyperaceae
22	Grass B	*Chloris* sp.	Poaceae
23	Grass C	*Sporobolus* sp.	Poaceae
24	Grass D	*Pennisetum/Cenchrus* sp.	Poaceae
25	Grass E	*Cynodon* sp.	Poaceae
26	Grass E2	*Cynodon* sp.	Poaceae
27	Grass F	Ambiguous	Poaceae
28	Grass G	*Pennisetum/Cenchrus* sp.	Poaceae
29	Grass H	*Cynodon* sp.	Poaceae
30	Grass I	*Cynodon* sp.	Poaceae
31	Grass J	Ambiguous	Poaceae
32	Grass K	*Enneapogon* sp.	Poaceae
33	Grass L	Ambiguous	Poaceae
34	Grass M	*Sporobolus* sp.	Poaceae
35	Grass N	*Sporobolus* sp.	Poaceae
36	Grass O	*Urochloa/Brachiaria/Eriochloa* sp.	Poaceae
37	Grass P	*Eragrostis* sp.	Poaceae
38	Grass Q	*Chloris* sp.	Poaceae
39	Grass R	*Tricholaena /Melinis* sp.	Poaceae
40	Grass T	*Sporobolus* sp.	Poaceae

KEY: Ambiguous – multiple species identified; Unidentifiable – the BLAST and phylogenetic identities fell below the threshold. (Reference: Appendix 1)

Twenty dicotyledons were only identified to the genus-level, and thirteen plants were further distinguished to the species rank (Tables 4.12 and 4.14). Four samples had ambiguous identities because multiple genera were identified for each specimen (Appendix 2). For instance, the BLAST results for Plant C based on the *ITS*1 gene identified it as *Megalochlamys revoluta*, while the phylogeny matched it with multiple genera. Furthermore, the consensus identities for three samples could not be determined because the results did not reach the required threshold stated in the data analyses.

Table 4.14: Consensus scientific names of dicotyledons based on both morphological and molecular analyses.

No	Sample ID	Consensus ID	Family
1	*Abutilon mauritianum*	*Abutilon sp.*	Malvaceae
2	*Acacia tortilis*	*Acacia/Vachellia tortilis*	Fabaceae
3	*Acacia xanthophloea*	*Acacia/Vachellia* sp.	Fabaceae
4	*Achyranthes aspera*	*Achyranthes* sp.	Amaranthaceae
5	*Asparagus asparagii*	*Asparagus* sp.	Asparagaceae
6	*Azima tetracantha*	Unidentifiable	Unidentifiable
7	*Balanites pedicellaris*	*Balanites* sp.	Zygophyllaceae
8	*Balanites sp.*	*Balanites* sp.	Zygophyllaceae
9	*Cadaba farinosa*	*Maerua triphylla*	Capparaceae
10	*Capparis tomentosa*	*Capparis* sp.	Capparaceae
11	*Cassia italica*	*Senna* sp.	Fabaceae
12	*Commicarpus plumbagineus*	*Commicarpus* sp.	Nyctaginaceae
13	*Cordia monoica*	*Cordia* sp.	Boraginaceae
14	*Dasyphaera prostrata*	*Volkensinia prostrata*	Amaranthaceae
15	*Euclea schimperi*	*Euclea* sp.	Ebenaceae
16	*Ficus sp.*	Ambiguous	Sapotaceae
17	*Hibiscus lila*	*Hibiscus micranthus*	Malvaceae
18	*Lantana camara*	*Lantana* sp.	Verbenaceae
19	*Lycium europaeum*	*Lycium* sp.	Solanaceae
20	*Maerua angolensis*	Unidentifiable	Unidentifiable
21	*Maerua crassifolia*	*Maerua triphylla*	Capparaceae
22	*Maerua sp.*	*Maerua triphylla*	Capparaceae
23	*Rhamphicarpa montana*	*Cycnium tubulosum*	Orobanchaceae
22	*Rhus natalensis*	*Searsia sp.*	Anacardiaceae
24	*Ruellia patula*	*Justicia odora*	Acanthaceae
26	*Salvadora persica*	*Salvadora* sp.	Salvadoraceae
27	*Scutia myrtina*	*Scutia myrtina*	Rhamnaceae
28	*Solanum dubium*	*Solanum coagulans*	Solanaceae
29	*Solanum incanum*	*Solanum* sp.	Solanaceae
30	*Solanum nigrum*	*Solanum nigrum*	Solanaceae
31	*Suaeda monoica*	*Suaeda* sp.	Amaranthaceae
32	*Trianthema ceratosepala*	Ambiguous	Aizoaceae
33	*Tribulus terrestris*	*Tribulus terrestris*	Zygophyllaceae
34	*Withania somnifera*	*Withania somnifera*	Solanaceae
35	Plant A	Ambiguous	Ambiguous
36	Plant B	*Boerhavia erecta*	Nyctaginaceae
37	Plant C	Ambiguous	Acanthaceae
38	Plant D	*Ipomoea* sp.	Convolvulaceae
39	Plant E	*Tephrosia* sp.	Fabaceae
40	Plant G	Unidentifiable	Unidentifiable

KEY: Ambiguous – multiple species identified; Unidentifiable – the BLAST and phylogenetic identities fell below the threshold. (Reference: Appendix 2)

CHAPTER FIVE: DISCUSSION

5.1 Morphological identification

The taxonomic assignment of plants is done using both morphology and molecular analysis (Hollingsworth *et al.*, 2016; Su *et al.*, 2016). In this study, the morphological identities of twenty-three plants were determined by the University of Nairobi's herbarium. It was noted that a critical challenge to morphological identification was if the external characters of the plants were damages due to improper specimen handling (Chan *et al.*, 2014).

5.2 Molecular analysis

The amplification rates of all the five barcodes used in this study were above 80%, with the *ITS*1 gene being the most successfully amplified locus (100%) followed by the *trn*L (UAA) intron (95%), *ITS*1-Poaceae (90%), the 18S rDNA locus (82.05%), and lastly, *trn*L-P6 (80.49%). These success rates were comparable to other studies that used these barcodes. For instance, Wang *et al.* (2014) tested whether *ITS*1 was better than *ITS*2, and the former region had a PCR success of 97.2%. Mishra *et al.* (2016) used the *ITS*1 barcode to resolve the genera of subtribe Cassiinae (Fabaceae), which resulted in an amplification success rate of 90%. Cheng *et al.* (2016) examined four different primer sets for the *ITS*1 locus in angiosperms, and the PCR success rate ranged from 65.8 – 96.4%. Madesis *et al.* (2012) utilized the *trn*L region to discriminate against Mediterranean leguminous crops and attained a PCR success

rate of 100%. The amplification success for the *trn*L intron in a barcoding study on xerothermic plants was 92.10% (Heise *et al.*, 2015). As for 18S (Banaras *et al.*, 2012) and *ITS*1-Poaceae (Ait Baamrane *et al.*, 2012; Banaras *et al.*, 2012; Kartzinel *et al.*, 2015; Lopes *et al.*, 2015), the specific amplification rates were not specified by the studies that utilized those barcodes.

The amplification success rate was higher in monocotyledons (93.06%) than in dicotyledons (87.67%). The amplification success rates for the various barcodes ranged from 87.50 – 100% in monocotyledons, and 75 – 100% in dicotyledons. It could be that the primer regions for some barcodes were designed based on the conserved loci in monocotyledons, which might be variable in some dicotyledons. If the primers were not specific to some dicotyledons, this could be the cause of the unsuccessful amplification.

5.3 Sanger sequencing

High-quality sequences were generated for 62.98% of the amplicons that were sent for Sanger sequencing at Macrogen Netherlands (Europe). The *ITS*1 gene was the most successfully sequenced region followed by the *trn*L (UAA) intron, the *ITS*1-Poaceae barcode, the 18S rDNA region, and lastly, the *trn*L-P6 locus. For both the 18S rDNA locus and the *trn*L-P6 barcode, high-quality sequences were generated for less than 50% of the amplicons that were sent for sequencing. The main issue could have been the fact that these two barcodes had a low PCR yield, which was

represented by very faint bands on the agarose gels. A major limitation of Sanger sequencing is the need for high-target amplicon yield (Shokralla *et al.*, 2014; Batovska *et al.*, 2017). Furthermore, the P6 marker has been noted to be highly A-T rich and contains many short mononucleotide repeats that may increase the sequencing error rates, which then hampers the correct taxonomic assignment (Erickson *et al.*, 2017).

Other issues that hinder direct Sanger sequencing and could have affected the sequencing of the genes used in this study include the following: loss of DNA during purification; non-target contamination; co-amplification of nuclear mitochondrial pseudogenes; and instances of intra-individual variability that is also known as, heteroplasmy (Shokralla *et al.*, 2014).

The length distribution of the sequences for all barcodes was within the documented range. The sequences for *trn*L (UAA) intron were 389 – 593 bp in length, and the primers used (*trn*L-c + *trn*L-d) are known to amplify the locus of 254 - 767 bp in plants (Taberlet *et al.*, 2007). The sequences for *ITS*1 locus were 252 – 383 bp in length, and the primers used (*ITS*-A + *ITS*-C) are known to amplify the locus of ~ 200 – 386 bp in plants (Kartzinel *et al.*, 2015). The sequences for the 18S rDNA region were 179 – 184 bp in length, and the primers used (1380F + 1510R) are known to amplify a locus of 87 - 186 bp in eukaryotes (Amaral-Zettler *et al.*, 2009). The sequences for *ITS*1-Poaceae were 113 – 121 bp in length and the primers used (*ITS*1-F + *ITS*1-Poa_R) are known to amplify the locus of ± 100 bp in plants (Ait

Baamrane *et al.*, 2012). The sequences for *trn*L-P6 were 90 – 109 bp in length, and the primers used (*trn*L-g + *trn*L-h) are known to amplify the locus of 10 - 143 bp in plants (Taberlet *et al.*, 2007).

5.4 Species identification

There is no single optimal method that is used to determine the identity of organisms for all taxa based on DNA sequences (Austerlitz *et al.*, 2009; Casiraghi *et al.*, 2010). Different approaches exist for matching an unknown query sequence with sequences in a reference database or library, and these approaches tend to be based on ad hoc criteria, which may include the frequency of the highest hits; percentage sequence similarity; bootstrapping; BLAST scores; or tree-based clustering assessment (CBOL Plant Working Group, 2009; China Plant BOL Group, 2011).

Caution is warranted in strictly relying on one approach, as errors in the curation of sequences in publicly available databases can propagate through the analysis and lead to misidentification of sequences (Deiner *et al.*, 2017). If possible, a combination of approaches should be used, and when possible, the resultant species assignments should be scrutinised with independent data based on the known distribution and ecology of the species (Deiner *et al.*, 2017). In this study, two commonly used methods, namely, sequence similarity (using BLAST algorithm) and tree-based clustering (using maximum-likelihood estimation), were utilized for species identification.

Different criteria can be used to interpret the BLAST output – the most common being accepting the top hits as the identity for the query sequence (Ross *et al.*, 2008; Yao *et al.*, 2010; Cheng *et al.*, 2016; Elansary *et al.*, 2017; Ghorbani *et al.*, 2017; Schilling and Floden, 2018; Tahir *et al.*, 2018). Ross *et al.* (2008) analysed the BLAST output using four different criteria, and the one that led to many correct identifications was the one where the top hits were accepted as the specimen identification. This was the BLAST criterion that was adopted in this study.

At the genus level, more monocotyledons and dicotyledons were identified using GenBank® (77.88% and 71.79%, respectively) than in the BOLD database (65.20% and 53.85%, respectively) using BLAST analysis. This was the same case for the identifications at the species level where more monocotyledons and dicotyledons were distinguished using GenBank® (60.58% and 69.23%, respectively) than by BOLD database (59.62% and 53.85%, respectively). The higher identification success rates in GenBank® in contrast to the BOLD database could be due to the fact that the former database generally contains very many plant sequences deposits because it has been in existence since the 1980s (Choudhuri, 2014), while BOLD was developed in early 2000s (Ratnasingham and Herbert, 2007). The very high number of sequences in the GenBank® database thus increases the likelihood of finding similar matches for any query.

Another primary reason why BOLD has lower success rates is that it is much more selective in which sequences it accepts (Ratnasingham and Herbert, 2007; Collins

and Cruickshank, 2012; Macher *et al*., 2017). BOLD database has fewer sequences, but they are much higher quality than GenBank®, which has no criteria for accepting sequences or taxonomic identities, therefore has more errors (Meier *et al.*, 2006; Valentini *et al.*, 2009; Kress *et al.*, 2015; Macher *et al.*, 2017). Additionally, plant barcoding studies have in the past focused on *mat*K and *rbc*L markers (CBOL Plant Working Group, 2009; Bafeel *et al.*, 2011; China Plant BOL Group, 2011; Peterson *et al.*, 2014; Bolson *et al.*, 2015; Wattoo *et al.*, 2016; Maloukh *et al.*, 2017; Mishra *et al.*, 2017), hence it could be that some of the plants of interest in this study are yet to be analysed by any of the chosen barcodes.

Based on GenBank®, the *ITS*1 gene had the highest success rates in identifying monocotyledons at both the genus and species levels (100% and 89.19%, respectively), whereas in dicotyledons, the *trn*L (UAA) intron was the most successful at both taxa (84.85% and 81.82%). On the other hand, the 18S rDNA region was the least successful (6.67%) in distinguishing the monocotyledons at the genus level, whereas the *trn*L-P6 gene could not determine the species. As for the dicotyledons, the 18S rDNA region had the least success rates at both the genus and species levels. Searching the query "18S" in the GenBank® database resulted in 1,365,445 hits, while *ITS*1 = 835,831, *trn*L = 285,947, *ITS*1 Poaceae = 10858, and lastly *trn*L-P6 = 8215 items. This means that GenBank® contains more sequences for the 18S rDNA locus as compared to all the other barcodes used in this study. Nonetheless, this region had the lowest successful identification rates for both monocotyledons and dicotyledons. It could be that the plants of interest in this study

have not yet been analysed by this marker, and hence, the sequences are not present in this database.

With reference to the BOLD database, the *trn*L-P6 gene successfully identified all monocotyledons at both the genus and species levels, whereas in dicotyledons, the *trn*L-P6 gene and *trn*L (UAA) intron had the highest success rates at the genus and species levels, respectively. The *trn*L locus had the highest identification success rates, probably due to having more sequences deposited in the public databases than the rest of the barcodes used. This locus has generally been very popular in diet barcoding studies, hence generating more references in the BOLD database (Soininen *et al.*, 2009; Valentini *et al.*, 2009; Ait Baamrane *et al.*, 2012; Kartzinel *et al.*, 2015; Mallott *et al.*, 2018). Additionally, the *trn*L markers were used to identify plants found in Laikipia, Kenya (Kartzinel *et al.*, 2015), some of which were similar to this study's plants of interest that were sampled from Amboseli, Kenya.

Based on the BOLD database, the 18S rDNA region could not distinguish any monocotyledon or dicotyledon at both the genus and species levels. It should be noted that very few plant sequences for this marker exist in this database, likely indicating that the region has not been majorly adopted as a barcode. Moreover, it may be that none of the plants of interest has been analysed based on this region, hence the lack of similar references in this database. Its lack of use could be because the 18S rDNA region is highly conserved (Soltis and Soltis, 1997; Patwardhan *et*

al., 2014) hence less variable than the *trn*L and *ITS*1 loci, therefore making it not very useful for distinguishing plant species.

Various methods exist for interpreting phylogenetic trees, most notably the liberal and strict criteria. For correct identification, the liberal-tree based process requires that the query sequence be either within or sister to a monospecific clade, whereas the strict tree-based requirement that the query is within, but not sister to a single-species clade (Meier *et al.*, 2006; Ross *et al*., 2008). The liberal tree-based method generally has a higher rate of correct identifications as compared to the strict tree-based criterion (Ross *et al.*, 2008; Pettengill and Neel, 2010; Wilson *et al.*, 2011).

Based on the liberal-tree method, 76.92% of the monocotyledons were successfully distinguished to the genus level, while 38.46% were assigned to their species ranks. In contrast, 80.77% of the dicotyledons were successfully identified to their genus ranks, while 58.97% were distinguished to their species levels. This meant that the criteria had much higher success in identifying dicotyledons than monocotyledons at both the genus and species levels. The liberal-tree based method has been documented as being unable to accurately handle sequences that do not have a conspecific in the reference database (Pettengill and Neel, 2010). It could be that some of the monocotyledons used in this study lack conspecifics in both GenBank® and BOLD databases; hence, the lower success rates of identification.

With reference to the individual barcodes, the trees for the *ITS*1 gene had the highest identification rates at both the genus and species levels (100% and 59.46%,

respectively) in monocotyledons. In contrast, for dicotyledons, the trees for the *trn*L locus had the highest identification rates at both the genus and species levels (93.94% and 67.65%, respectively). In all instances, the trees for the 18S rDNA region had the lowest identification success rates. In comparison, phylogenetic analyses by Mishra *et al.* (2016) resulted in the *ITS*1 barcode efficiently identifying 90% of the plant species. In a study of Amazonian trees based on the *trn*L (UAA) intron, the monophyletic genera and species recovered were 63% and 53%, respectively (Gonzalez *et al.*, 2009). It has been documented that the slow rate and pattern of 18S rDNA evolution across land plants may limit the usefulness of this gene for phylogeny reconstruction at deep levels of plant phylogeny (Soltis and Soltis, 1997; Soltis *et al.*, 1999; Patwardhan *et al.*, 2014).

5.5 Consensus identities

For consensus identification, only data generated by *ITS*1 and the *trn*L (UAA) intron were used because the targeted *ITS*1-Poaceae region is found within the *ITS*1 gene, whereas the *trn*L-P6 locus is located within the *trn*L (UAA) intron. Furthermore, the two larger loci had better identification success rates using BLAST and phylogenetic analyses. The 18S rDNA locus was unsuccessful using all methods; hence, not much useful data was generated for identification purposes.

Sixty-six (82.50%) out of 80 plants of interest were identified using both molecular and morphological analyses. Specifically, 50 plants were identified to only the genus

level, while 16 samples were further distinguished to the species level. Out of the remaining 14 plants whose consensus scientific names could not be ascertained, the identities for 10 samples were ambiguous in that multiple genera were identified by various analyses; whereas the analyses for remaining 4 plants did not reach the specified threshold. In a few cases, the names of the samples given based on morphology differed from those assigned through molecular analyses. For instance, '*Cadaba farinosa'* was identified as *Maerua triphylla* based on all *ITS*1 analyses; and *'Ruellia patula'* as *Justicia odora*. However, a search for the plant '*Cadaba farinosa*' in GenBank® gave no results. Moreover, only the sequences for *mat*K and *rbc*L genes exist in the BOLD database.

Monocotyledons and dicotyledons had the same identification success rate (82.50%). However, the majority of monocotyledons (30/40 plants) could only be distinguished to the genus level, and only 3 were identified at the species level. These were *Dactylotenium aegyptium*, *Odyssea paucinervis* and *Setaria verticillata*. Furthermore, nine grasses belonged to the *Sporobolus* genera, but the analyses could not ascertain the species name. For instance, the BLAST analysis for 'Grass N' identified it as *Sporobolus virginicus* based on the *ITS*1 gene, whereas the *trn*L (UAA) intron matched it with *S. michauxianus* and *S. africanus*.

Furthermore, the *ITS*1 phylogeny showed that it was related to multiple *Sporobolus* species, whereas the *trn*L (UAA) intron identified it as *Crypsis vaginiflora*. This was the case for the other grasses classified in this genus because there was no

consensus species name identified by all the analyses. Therefore, there is a need to revise the names given to the sequences deposited in the public database for this genus because it could be that a particular species has been given multiple identities hence the confusion in the reference.

Peterson *et al.* (2014) analyzed the *Sporobolus* species in the public databases and created a new subgeneric classification for this genus. In this study, grasses 'AC', 'AJ', 'C', and 'M' identify the following species: *Sporobolus marginatus*, *S. coromandelianus*, *S. pyramidatus* and *S. cordofanus,* which were reclassified to *Sporobolus* subsect *Pyramidati* P.M. Peterson (Peterson *et al.*, 2014). Grasses 'C', 'N' and 'AC' identify *Sporobolus africanus*, which was reclassified to the subgenus *Sporobolus* subsect. *Sporobolus* (Peterson, Romaschenko, Arrieta, *et al.*, 2014). Grass 'AC' identifies *Sporobolus stapfianus* that was also reclassified to the subgenus *Sporobolus* subsect. *Sporobolus* (Peterson, Romaschenko, Arrieta, *et al.*, 2014). Grass 'N' identified the grass *Crypsis vaginiflora* that was reclassified to the subgenus *Sporobolus* subsect. *Crypsis* (Peterson *et al.*, 2014). Grass 'T' identified both *Sporobolus agrostoides* and *S. fimbriatus* that were reclassified to the subgenus *Sporobolus* subsect. *Fimbriatae* (Peterson, Romaschenko, Arrieta, *et al.*, 2014). The classification for *Sporobolus consimilis* was denoted as "incertae sedis" by Peterson *et al.* (2014), meaning that its true identity was uncertain. In this study, the '*Sporobolus consimilis*' sample was matched to multiple other *Sporobolus* species such as *S. cryptandrus* and *S. agrostoides.*

Three monocotyledons were matched to the *Cenchrus* or *Pennisetum* genera. These two genera are synonymous based on the book called "CRC world dictionary of Grasses: Common Names, Scientific Names, Eponyms, Synonyms, and Etymology" by Quattrocchi (2006). Furthermore, a study discovered that the sequences deposited for *Cenchrus* sp., *Odontelytrum* sp., and *Pennisetum* sp. in GenBank® were similar, hence proposed that the species of both *Pennisetum* and *Odontelytrum* be renamed to *Cenchrus* species (Chemisquy *et al.*, 2010). However, the correct species name for the grasses in this genus could not be ascertained. For instance, the BLAST analysis for Grass 'D' matched it with *Cenchrus mezianus* and *Pennisetum mezianum* based on the *ITS*1 gene, whereas the *trn*L (UAA) intron identified *Cenchrus mezianus*, *C. ciliaris* and *Pennisetum mezianum*. Furthermore, the *ITS*1 phylogeny matched it with *Cenchrus ciliaris,* while the *trn*L (UAA) intron placed it a clade with multiple *Cenchrus* and *Pennisetum* species. Grasses 'G' and 'AG' also had the same results.

The remaining monocotyledons in this study were assigned to the following genera: 4 *Cynodon* species; 3 *Chloris* species; 3 *Cyperus* species; 3 *Enneapogon* species; 2 *Eragrostis* species; and 1 sample each for *Disakisperma*, *Trichlolaena* or *Melinis*, *Urochloa* or *Eriochloa* or *Brachiaria*. For instance, the BLAST analysis for the grass 'A' identified it as *Enneapogon cenchroides* based on the *ITS*1 gene, whereas the *trn*L (UAA) intron matched it with *Enneapogon scoparius*, *E. cenchroides* and *E. persicus*. Furthermore, the phylogenetic analyses identified this grass with *Enneapogon scoparius*. Hence, as noted earlier on, only the genus could be

ascertained for most of the monocotyledons used in this study since multiple species (within the same genus) were identified.

The consensus identities for four monocotyledons were ambiguous in that the plants were matched to multiple genera. For instance, grasses 'L' was identified as *Chloris mossambicensis* and *C. nutans* based on the *ITS*1 BLAST analysis, while the *trn*L (UAA) intron matched it with *Cynodon dactylon* and *Chloris roxburghiana*. Furthermore, the *ITS*1 phylogeny identified it as *Tetrapogon cenchriformis*. *Tetrapogon* and *Chloris* genera are listed as synonyms by Quattrocchi (2006). Peterson *et al.* (2012) noted that the sequences deposited in the GenBank® for *Tetrapogon* sp. were positioned in a similar clade to *Enteropogon* sp. and *Saugetia* sp. Furthermore, Peterson *et al.* (2015) conducted a phylogenetic analysis of the grass sequences deposited in GenBank® and found that some of the *Chloris* species were embedded in clades consisting of *Tetrapogon* species – and this led to the renaming of some of the *Chloris* species as *Tetrapogon* species. It would be very beneficial if the sequences in the public databases for these two genera are revised as per Peterson *et al.* (2015) findings.

The grass sample '*Drake-brockmania somalensis*' was identified as *Dinebra haareri* by *ITS*1 analyses while the *trn*L (UAA) intron placed it with *Leptochloa virgata* and *Enteropogon macrostachyus*. The name of this grass sample is documented as being synonymous to *Eleusine somalensis* by "CRC World Dictionary of Grasses" by Quattrocchi (2006). Furthermore, Peterson *et al.* (2012)

conducted phylogenetic analyses of 130 grasses present in GenBank® and revealed that the genus *Leptochloa s.l* was polyphyletic, and *Dinebra* sp. and *Drake-brockmania somalensis* were all embedded within the *Leptochloa* clade. The authors noted that the species of both *Dinebra* and *Leptochloa* are similar in morphology, hence related, while *Drake-brockmania* sp. and *Leptochloa* sp. are slightly dissimilar in appearance (Peterson *et al.*, 2012). The study concluded that *Drake-brockmania somalensis* was similar to *Dinebra somalensis,* while *Drake-brockmania haareri* was similar to *Dinebra haareri,* hence proposed that *Drake-brockmania* be renamed to *Dinebra* species (Peterson *et al.*, 2012). The correct scientific name for the grass '*Drake-brockmania somalensis*' used in this study is still uncertain.

As for dicotyledons, 13 plants had ambiguous identities; for instance, '*Ficus* sp.' was identified by the *trn*L (UAA) intron as *Faucherea thouvenotii* and *Manikara zapota*, whereas the *ITS*1 gene matched it with multiple *Sideroxylon* species. The sample '*Trianthema ceratosepalum*' was matched to *Tetragonia schenckii* based on the *trn*L (UAA) intron phylogenetic tree, whereas *ITS*1 placed it in a polytomy with multiple *Trianthema* species. In such instances, the correct scientific names of these samples remain uncertain.

Three dicotyledons and one monocotyledon were not identified because their results did not reach the threshold given for the molecular analyses. These were *Azima tetracantha*, *Maerua angolensis*, plant 'G' and grass 'AM'. A search for the plant

'*Azima tetracantha*' in GenBank® resulted in no hits. Moreover, only sequences for *mat*K and *rbc*L genes exist in the BOLD database for this plant indicating that the specimen is yet to be analysed based on any of the five barcodes used in this study. This means that there existed no reference for the sequence analyses, hence the lack of identification. As for the sample *Maerua angolensis*, only the *trn*L (UAA) intron PCR was successful in this study. Furthermore, the identity of the query sequence did not match any of the *Maerua angolensis* sequences deposited in the public databases. It seems that there was no reference in the public databases for the identification of the four plants mentioned above.

The following chages were also noted: *Acacia* species were reassigned to *Vachellia* species (Dyer, 2014); *Cassia italica* is also known as *Senna italica* (Okeyo and Bosch, 2007); *Cenchrus* species are otherwise known as *Pennisetum* species (Quattrocchi, 2006; Chemisquy *et al.*, 2010); *Dasyphaera prostrata* is also known as *Volkensinia prostrata* (Altmann, 1998); *Drake-brockmania somalensis* was reclassified to *Dinebra somalensis* (Quattrocchi, 2006; Peterson *et al.*, 2012). *Rhamphicarpa montana* was renamed to *Cycnium tubulosum* (Staner, 1938); *Tetrapogon* species are also known as *Chloris* species (Quattrocchi, 2006; Petersonn *et al.*, 2015); *Solanum dubium* is synonymous to *Solanum coagulans* (Altmann, 1998); and species of *Tricholaena* and *Melinis* are equivalent (Quattrocchi, 2006). Additionally, the following four genera are considered to be a monophyletic group based on phylogenetic analysis: *Brachiaria, Urochloa, Eriochloa,* and *Melinis* (González and Morton, 2005).

CHAPTER SIX: CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

1) The morphological identities of sixty-six plants were in agreement with genera identified through molecular analyses. The combination of morphological and molecular analyses presented an effective method of identifying the plants of interest.
2) The *ITS*1 gene and the *trn*L (UAA) intron had the highest discriminatory power out of all the five barcodes used. The utilization of markers from both the nuclear and chloroplast regions was crucial to the overall high discriminatory efficiency (82.50%) achieved in this study.
3) The plant identification success rate was higher in GenBank® at both the genus and species levels (75.27% and 65.93%, respectively) than in the BOLD database (58.24% and 54.95%, respectively).
4) The BLAST analysis was more successful in the identification of monocotyledons (77.88%) than dicotyledons (71.79%). In contrast, the phylogenetic analysis was more successful in the identification of dicotyledons (80.77%) than monocotyledons (76.92%).

6.2 Recommendations

1) The data generated from this work can be used as a reference for further studies relating to the characterization of plants in the Amboseli ecosystem and by extension, Kenya.
2) Because the selected plant samples are those that are eaten explicitly by the yellow baboons in Amboseli, the data generated can be used to conduct a metabarcoding study of the baboon plant-diet composition. By analysing the faecal matter, future research could confirm whether the analysed plants form part of the baboons' diet.
    a) Furthermore, it is well-known that diet affects the composition of the gut microbiome; hence, the data generated from the metabarcoding study could be useful in any research relating to the composition of the Amboseli baboons' gut microbiome.
    b) Moreover, the data generated from the metabarcoding study could be used to show how social behaviour (specifically, dominance rank) is linked to inter-individual variation in diet composition.
3) Lastly, the macronutrient and mineral content of these plants could be analysed in a bid to evaluate the nutritional value of these plants to the yellow baboons in Amboseli.

REFERENCES

Ait Baamrane, M. A., Shehzad, W., Ouhammou, A., Abbad, A., Naimi, M., Coissac, E., Taberlet, P. and Znari, M. (2012) 'Assessment of the food habits of the Moroccan dorcas gazelle in M'Sabih Talaa, west central Morocco, using the trnL approach', *PLoS ONE*, 7(4). doi: 10.1371/journal.pone.0035643.

Ali, M. A., Gyulai, G. and Al-Hemaid, F. M. A. (2015) *Plant DNA Barcoding and Phylogenetics*. Edited by M. A. Ali, G. Gyulai, and F. M. A. Al-Hemaid. LAP LAMBERT Academic Publishing.

Ali, M. A., Gyulai, G., Hidvegi, N., Kerti, B., Al Hemaid, F. M. A., Pandey, A. K. and Lee, J. (2014) 'The changing epitome of species identification – DNA barcoding', *Saudi Journal of Biological Sciences*, 21, pp. 204–231. doi: 10.1016/j.sjbs.2014.03.003.

Altmann, J., Hausfater, G. and Altmann, S. A. (1985) 'Demography of Amboseli Baboons, 1963-1983', *American Journal of Primatology*, 8, pp. 113–125.

Altmann, S. A. (1974) 'Baboons, Space, Time, and Energy', *Amer. Zool.*, 14, pp. 221–248.

Altmann, S. A. (1998) *Foraging for Survival: Yearling Baboons in Africa*, *University of Chicago Press, Chicago*.

Altmann, S. A. and Altmann, J. (1970) *Baboon Ecology: African Field Research*. University of Chicago Press, Chicago.

Altmann, S. A., Post, D. G. and Klein, D. F. (1987) 'Nutrients and toxins of plants in Amboseli , Kenya', *Afr. J. Ecol.*, 25, pp. 279–293.

Alvarez, I. and Wendel, J. F. (2003) 'Ribosomal ITS sequences and plant phylogenetic inference', *Molecular Phylogenetics and Evolution*, 29, pp. 417–434. doi: 10.1016/S1055-7903(03)00208-2.

Amaral-Zettler, L. A., McCliment, E. A., Ducklow, H. W. and Huse, S. M. (2009) 'A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes', *PLoS ONE*, 4(7), pp. 1–9. doi: 10.1371/journal.pone.0006372.

Ando, H., Setsuko, S., Horikoshi, K., Suzuki, H., Umehara, S., Inoue-Murayama, M. and Isagi, Y. (2013) 'Diet analysis by next-generation sequencing indicates the frequent consumption of introduced plants by the critically endangered red-headed wood pigeon (Columba janthina nitens) in oceanic island habitats', *Ecology and Evolution*, 3(12), pp. 4057–4069. doi: 10.1002/ece3.773.

Angers-Loustau, A., Petrillo, M., Paracchini, V., Kagkli, D. M., Rischitor, P. E.,

Gallardo, A. P., Patak, A., Querci, M. and Kreysa, J. (2016) 'Towards Plant Species Identification in Complex Samples : A Bioinformatics Pipeline for the Identification of Novel Nuclear Barcode Candidates', *PLoS ONE*, 11(1), pp. 1–15. doi: 10.1371/journal.pone.0147692.

Arif, I. A., Bakir, M. A., Khan, H. A., Al Farhan, A. H., Al Homaidan, A. A., Bahkali, A. H., Al Sadoon, M. and Shobrak, M. (2010) 'A Brief Review of Molecular Techniques to Assess Plant Diversity', *International Journal of Molecular Sciences*, 11, pp. 2079–2096. doi: 10.3390/ijms11052079.

Austerlitz, F., David, O., Schaeffer, B., Bleakley, K., Olteanu, M., Leblois, R., Veuille, M. and Laredo, C. (2009) 'classification methods', *BMC Bioinformatics*, 10. doi: 10.1186/1471-2105-10-S14-S10.

Bączkiewicz, A., Szczecińska, M., Sawicki, J., Stebel, A. and Buczkowska, K. (2017) 'DNA barcoding, ecology and geography of the cryptic species of Aneura pinguis and their relationships with Aneura maxima and Aneura mirabilis (Metzgeriales, Marchantiophyta)', *PLoS One*, 12(12). doi: 10.1371/journal.pone.0188837.

Bafeel, S. O., Arif, I. A., Bakir, M. A., Khan, H. A., Al Farhan, A. H., Al-homaidan, A. A., Ahamed, A. and Thomas, J. (2011) 'Comparative evaluation of PCR success with universal primers of maturase K (matK) and ribulose-1, 5-bisphosphate carboxylase oxygenase large subunit (rbcL) for barcoding of some arid plants', *POJ*, 4(4), pp. 195–198.

Baldwin, B. G. (1992) 'Phylogenetic Utility of the Internal Transcribed Spacers of Nuclear Ribosomal DNA in Plants: An Example from the Compositae', *Molecular Phylogenetics and Evolution*, 1(1), pp. 3–16.

Banaras, S., Aman, S., Zafar, M., Khan, M., Abbas, S., Shinwari, Z. K. and Shakeel, S. N. (2012) 'Molecular Identification and Comparative Analysis of Novel 18S Ribosomal RNA Genomic Sequences Of A Wide Range Of Medicinal Plants', *Pak. J. Bot.*, 44(6), pp. 2021–2026.

Batovska, J., Cogan, N. O. I., Lynch, S. E. and Blacket, M. J. (2017) 'Using Next-Generation Sequencing for DNA Barcoding : Capturing Allelic Variation in ITS2', *g3*, 7, pp. 19–29. doi: 10.1534/g3.116.036145.

Bazinet, A. L., Zwickl, D. J. and Cummings, M. P. (2014) 'A Gateway for Phylogenetic Analysis Powered by Grid Computing Featuring GARLI 2.0', *Syst. Biol.*, 63(5), pp. 812–818. doi: 10.1093/sysbio/syu031.

Bell, K. L., Loeffler, V. M. and Brosi, B. J. (2017) 'An rbcL Reference Library to Aid in the Identification of Plant Species Mixtures by DNA Metabarcoding', *Applications in Plant Sciences*, 5(3). doi: 10.3732/apps.1600110.

Bentley-Condit, V. K. (2009) 'Food Choices and Habitat Use by the Tana River

Yellow Baboons (Papio cynocephalus): A Preliminary Report on Five Years of Data', *American Journal of Primatology*, 71, pp. 432–436. doi: 10.1002/ajp.20670.

Bentley-Condit, V. K. and Power, M. L. (2018) 'The nutritional content of Tana River yellow baboon (Papio cynocephalus) foods in a partially forested habitat', *PLoS ONE*, 13(11), p. e0207186. doi: doi. org/10.1371/journal.pone.0207186.

Bezeng, B. S., Davies, T. J., Daru, B. H., Kabongo, R. M., Maurin, O., Yessoufou, K., Bank, H. Van Der and Bank, M. Van Der (2017) 'Ten years of barcoding at the African Centre for DNA Barcoding', *Genome*, 60, pp. 629–638. doi: doi.org/10.1139/gen-2016-0198.

Birch, J. L., Walsh, N. G., Cantrill, D. J., Holmes, G. D. and Murphy, D. J. (2017) 'Testing efficacy of distance and tree-based methods for DNA barcoding of grasses (Poaceae tribe Poeae) in Australia', *PLoS ONE*, 12(10), p. e0186259. doi: https://doi.org/10.1371/ journal.pone.0186259.

Blattner, F. R. (1999) 'Direct Amplification of the Entire ITS Region from Poorly Preserved Plant Material Using Re- combinant PCR', *BioTechniques*, 27, pp. 1180–1186.

Bolson, M., Smidt, E. D. C., Brotto, M. L. and Silva-, V. (2015) 'ITS and trnH-psbA as Efficient DNA Barcodes to Identify Threatened Commercial Woody Angiosperms from Southern Brazilian Atlantic Rainforests', *PLoS one*, 10(12), p. e0143049. doi: 10.1371/journal.pone.0143049.

Bradley, B. J., Stiller, M., Doran-Sheehy, D. M., Harris, T., Chapman, C. A., Vigilant, L. and Poinar, H. (2007) 'Plant DNA Sequences from Feces: Potential Means for Assessing Diets of Wild Primates', *American Journal of Primatology*, 69, pp. 1–7. doi: 10.1002/ajp.

Bressan, E. A., Rossi, M. L., Gerald, L. T. S. and Figueira, A. (2014) 'Extraction of high-quality DNA from ethanol-preserved tropical plant tissues', *BMC Research Notes*, 7, p. 268. Available at: http://www.biomedcentral.com/1756-0500/7/268.

Carrière, S. (2002) 'Photographic Key for the Microhistological Identification of some Arctic Vascular Plants', *ARCTIC*, 55(3), pp. 247–268.

Casiraghi, M., Labra, M., Ferri, E., Galimberti, A. and Mattia, F. De (2010) 'DNA barcoding : a six-question tour to improve users ' awareness about the method', *Brief Bioinform. 2010;*, 11(4), pp. 440–453. doi: 10.1093/bib/bbq003.

CBOL Plant Working Group (2009) 'A DNA barcode for land plants', *PNAS*, 106(31), pp. 12794–12797. Available at:

www.pnas.org/cgi/doi/10.1073/pnas.0905845106.

Chan, A., Chiang, L., Hapuarachchi, H. C., Tan, C., Pang, S., Lee, R., Lee, K., Ng, L. and Lam-Phua, S.-G. (2014) 'DNA barcoding : complementing morphological identification of mosquito species in Singapore', *Parasites & Vectors*, 7(569). doi: 10.1186/s13071-014-0569-4.

Chase, M. W., Salamin, N., Wilkinson, M., Dunwell, J. M., Kesanakurthi, R. P., Haidar, N. and Savolainen, V. (2005) 'Land plants and DNA barcodes : short-term and long-term goals', *Phil. Trans. R. Soc. B*, 360, pp. 1889–1895. doi: 10.1098/rstb.2005.1720.

Chemisquy, M. A., Giussani, L. M. G., Scataglini, M. A., Kellogg, E. A. and Morrone, O. (2010) 'Phylogenetic studies favour the unification of Pennisetum , Cenchrus and Odontelytrum (Poaceae): a combined nuclear , plastid and morphological analysis , and nomenclatural combinations in Cenchrus', *Annals of Botany*, 106, pp. 107–130. doi: 10.1093/aob/mcq090.

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., Zhu, Y., Ma, X., Gao, T., Pang, X., Li, Y., Li, X., Jia, X., Lin, Y. and Leon, C. (2010) 'Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species', *PLoS One*, 5(1). doi: 10.1371/journal.pone.0008613.

Cheng, T., Xu, C., Lei, L., Li, C., Zhang, Y. and Zhou, S. (2016) 'Barcoding the kingdom Plantae: new PCR primers for ITS regions of plants with improved universality and specificity', *Molecular Ecology Resources*, 16(1), pp. 138–149. doi: 10.1111/1755-0998.12438.

China Plant BOL Group (2011) 'Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants', *PNAS*, 108(49), pp. 19641–19646. doi: 10.1073/pnas.1104551108.

Choudhuri, S. (2014) 'Data, Databases, Data Format, Database Search, Data Retrieval Systems, and Genome Browsers', in *Bioinformatics for beginners*, pp. 77–181.

Collins, R. A. and Cruickshank, R. H. (2012) 'The seven deadly sins of DNA barcoding', *Molecular Ecology Resources*. doi: 10.1111/1755-0998.12046.

Cristescu, M. E. (2014) 'From barcoding single individuals to metabarcoding biological communities : towards an integrative approach to the study of global biodiversity', *Trends in Ecology & Evolution*. Elsevier Ltd, pp. 1–6. doi: 10.1016/j.tree.2014.08.001.

Culley, T. M. (2013) 'Why Vouchers Matter in Botanical Research', *Applications in Plant Sciences 2013*, 1(11), p. 1300076. doi: 10.3732/apps.1300076.

Dechbumroong, P., Aumnouypol, S., Denduangboripant, J. and Sukrong, S. (2018) 'DNA barcoding of Aristolochia plants and development of species-specific multiplex PCR to aid HPTLC in ascertainment of Aristolochia herbal materials', *PLoS ONE*, 13(8), p. e0202625.

Deiner, K., Bik, H. M., Elvira, M., Seymour, M., Lacoursiere-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D. M., Vere, N. De, Pfrender, M. E. and Bernatchez, L. (2017) 'Environmental DNA metabarcoding : Transforming how we survey animal and plant communities', *Molecular Ecology.*, pp. 1–24. doi: 10.1111/mec.14350.

Devey, D. S., Chase, M. W. and Clarkson, J. J. (2009) 'A stuttering start to plant DNA barcoding: microsatellites present a previously overlooked problem in non-coding plastid regions', *Taxon*, 58(1), pp. 7–15.

Dong, W., Cheng, T., Li, C., Xu, C., Long, P., Chen, C. and Zhou, S. (2013) 'Discriminating plants using the DNA barcode rbcL b : an appraisal based on a large data set', *Molecular Ecology Resources*, pp. 1–8. doi: 10.1111/1755-0998.12185.

Dyer, C. (2014) 'New names for the African Acacia species in Vachellia and Senegalia', *Southern Forests: a Journal of Forest Science*, 76(4), pp. iii–iii. doi: 10.2989/20702620.2014.980090.

Edgar, R. C. (2004) 'MUSCLE : multiple sequence alignment with high accuracy and high throughput', *Nucleic Acids Research*, 32(5), pp. 1792–1797. doi: 10.1093/nar/gkh340.

Elansary, H. O., Ashfaq, M., Ali, H. M. and Yessoufou, K. (2017) 'The first initiative of DNA barcoding of ornamental plants from Egypt and potential applications in horticulture industry', *PLoS ONE*, 12(2). doi: 10.1371/journal.pone.0172170.

Erickson, D. L., Reed, E., Ramachandran, P., Bourg, N. A., Mcshea, W. J. and Ottesen, A. (2017) 'Reconstructing a herbivore ' s diet using a novel rbcL DNA mini-barcode for plants', *AoB PLANTS*, 9. doi: 10.1093/aobpla/plx015.

Fahner, N. A. (2015) *Assessing Belowground Plant Diversity in Wetland Soil through DNA Metabarcoding : Impact of DNA Marker Selection and Analysis of Temporal Patterns*. Master's Thesis, The University of Guelph.

Fahner, N. A., Shokralla, S., Baird, D. J. and Hajibabaei, M. (2016) 'Large-Scale Monitoring of Plants through Environmental DNA Metabarcoding of Soil : Recovery , Resolution , and Annotation of Four DNA Markers', *PLoS ONE*, 11(6). doi: 10.1371/journal.pone.0157505.

Fazekas, A. J., Burgess, K. S., Kesanakurti, P. R., Graham, S. W., Newmaster, S.

G., Husband, B. C., Percy, D. M., Hajibabaei, M. and Barrett, S. C. H. (2008) 'Multiple Multilocus DNA Barcodes from the Plastid Genome Discriminate Plant Species Equally Well', *PLoS One*, 3(7). doi: 10.1371/journal.pone.0002802.

Fazekas, A. J., Kesanakurti, P. R., Burgess, K. S., Percy, D. M., Graham, S. W., Barrett, S. C. H., Newmaster, S. G., Hajibabaei, M. and Husband, B. C. (2009) 'Are plant species inherently harder to discriminate than animal species using DNA barcoding markers ?', *Molecular Ecology Resources*, 9, pp. 130–139. doi: 10.1111/j.1755-0998.2009.02652.x.

Fazekas, A. J., Kuzmina, M. L., Newmaster, S. G. and Hollingsworth, P. M. (2012) 'DNA Barcoding Methods for Land Plants', in Kress, W. J. and Erickson, D. L. (eds) *DNA Barcodes. Methods in Molecular Biology (Methods and Protocols)*. Totowa, NJ: Humana Press, pp. 223–252. doi: 10.1007/978-1-61779-591-6.

Felger, R. S., Rutman, S. and Malusa, J. (2014) 'Ajo Peak to Tinajas Altas: A flora of southwestern Arizona. Part 6. Poaceae – grass family.', *Phytoneuron*, 35(March), pp. 1–139.

Gara, T. W., Wang, T., Skidmore, A. K., Zengeya, F. M., Ngene, S. M., Murwira, A. and Ndaimani, H. (2016) 'Understanding the effect of landscape fragmentation and vegetation productivity on elephant habitat utilization in Amboseli ecosystem, Kenya', *African Journal of Ecology*, p. DOI: 10.1111/aje.12346. doi: 10.1111/aje.12346.

Garnick, S., Barboza, P. S. and Walker, J. W. (2018) 'Assessment of Animal-Based Methods Used for Estimating and Monitoring Rangeland Herbivore Diet Composition', *Rangeland Ecology & Management*, 71(4), pp. 449–457.

Ghorbani, A., Saeedi, Y. and Boer, H. J. de (2017) 'Unidentifiable by morphology : DNA barcoding of plant material in local markets in Iran', *PLoS ONE*, 12(4).

González, A. M. T. and Morton, C. M. (2005) 'Molecular and morphological phylogenetic analysis of Brachiaria and Urochloa (Poaceae)', *Molecular Phylogenetics and Evolution*, 37, pp. 36–44. doi: 10.1016/j.ympev.2005.06.003.

Gonzalez, M. A., Baraloto, C., Engel, J., Mori, S. A., Petronelli, P., Riera, B., Roger, A., Thebaud, C. and Chave, J. (2009) 'Identification of Amazonian Trees with DNA Barcodes', 4(10), p. e7483. doi: 10.1371/journal.pone.0007483.

Hagedorn, G., Rambold, G. and Martellos, S. (2010) 'Types of identification keys', in Nimis, P. L. and Vignes Lebbe, R. (eds) *Tools for Identifying Biodiversity: Progress and Problems*. EUT Edizioni Università di Trieste, pp. 59–64.

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A. C. and Baird, D. J. (2011)

'Environmental Barcoding: A Next-Generation Sequencing Approach for Biomonitoring Applications Using River Benthos', *PLoS One*, 6(4). doi: 10.1371/journal.pone.0017497.

Hamby, R. K. and Zimmer, E. A. (1988) 'Ribosomal RNA sequences for inferring phylogeny within the grass family (Poaceae)', *Plant Systematics and Evolution*, 160(1/2), pp. 29–37.

Han, J., Zhu, Y., Chen, X., Liao, B., Yao, H., Song, J., Chen, S. and Meng, F. (2013) 'The Short ITS2 Sequence Serves as an Efficient Taxonomic Sequence Tag in Comparison with the Full-Length ITS', *BioMed Research International*.

Harris, J. G. and Harris, M. W. (1994) *Plant Identification Terminology: An Illustrated Glossary*. Spring Lake Publishing.

Hassoon, I. M., Kassir, S. A. and Altaie, S. M. (2018) 'A Review of Plant Species Identification Techniques', *International Journal of Science and Research (IJSR)*, 7(8), pp. 326–328. doi: 10.21275/ART2019476.

Hebert, P. D. N., Cywinska, A., Ball, S. L. and DeWaard, J. R. (2003) 'Biological identifications through DNA barcodes', *Proc. R. Soc. Lond. B*, 270, pp. 313–321. doi: 10.1098/rspb.2002.2218.

Heise, W., Babik, W., Kubisz, D. and Kajtoch, L. (2015) 'A three-marker DNA barcoding approach for ecological studies of xerothermic plants and herbivorous insects from central Europe', *Botanical Journal of the Linnean Society*. doi: 10.1111/boj.12261.

Hollingsworth, P. M., Graham, S. W. and Little, D. P. (2011) 'Choosing and Using a Plant DNA Barcode', *PLoS One*, 6(5). doi: 10.1371/journal.pone.0019254.

Hollingsworth, P. M., Li, D., Bank, M. Van Der, Twyford, A. D., Botanic, R., Edinburgh, G., Row, I., Eh, E. and Hollingsworth, P. M. (2016) 'Telling plant species apart with DNA : from barcodes to genomes'.

Hosein, F. N., Austin, N., Maharaj, S., Johnson, W., Rostant, L., Ramdass, A. C. and Rampersad, S. N. (2017) 'Utility of DNA barcoding to identify rare endemic vascular plant species in Trinidad', *Ecology and Evolution*, pp. 1–23. doi: 10.1002/ece3.3220.

Hua, C., Wang, Y., Xie, Z., Guo, Z., Zhang, Y., Qiu, Y. and Wang, L. (2018) 'Sciences in Cold and Arid Regions Effects of intercropping on rhizosphere soil microorganisms and root exudates of Lanzhou lily (Lilium davidii var . unicolor)', *Sciences in Cold and Arid Regions*, 10(2), pp. 0159–0168. doi: 10.3724/SP.J.1226.2018.00159.Effects.

Hyslop, E. J. (1980) 'Stomach contents analysis - a review of methods and their application', *J. Fish. Biol.*, 17, pp. 411–429.

Inglis, P. W., Pappas, M. D. C. R., Resende, L. V. and Grattapaglia, D. (2018) 'Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high- throughput SNP genotyping and sequencing applications', *PLoS ONE*, 13(10), p. e0206085. Available at: https://doi.org/10.1371/journal. pone.0206085.

de Iongh, H. H., de Jong, C. B., van Goethm, J., Klop, E., Brunsting, A. M. H., Loth, P. E., Prins, H.H. T. and Prins, Herbert H T (2011) 'Resource partitioning among African savanna herbivores in North Cameroon: the importance of diet composition, food quality and body mass', *Journal of Tropical Ecology*, 27, pp. 503–513. doi: 10.1017/S0266467411000307.

Iwanowicz, D. D., Vandergast, A. G., Cornman, R. S., Adams, R., Kohn, J. R., Fisher, R. N. and Brehme, C. S. (2016) 'Metabarcoding of Fecal Samples to Determine Herbivore Diets : A Case Study of the Endangered Pacific Pocket Mouse', *PLoS ONE*, 11(11). doi: 10.1371/journal.pone.0165366.

Johnson, M. K., Wofford, H. and Pearson, H. A. (1983) *Microhistological Techniques for Food Habits Analyses*, *Research Paper*.

Kartzinel, T. R., Chen, P. A., Coverdale, T. C., Erickson, D. L., Kress, W. J., Kuzmina, M. L., Rubenstein, D. I., Wang, W. and Pringle, R. M. (2015) 'DNA metabarcoding illuminates dietary niche partitioning by African large herbivores'. doi: 10.1073/pnas.1503283112.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. and Drummond, A. (2012) 'Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data', *Bioinformatics*, 28(12), pp. 1647–1649. doi: 10.1093/bioinformatics/bts199.

Kemunto, E. A. (2013) *Role of Geoinformation in Enhancing Tourism in National Parks Case Study: Amboseli National Park*. Bachelor's Project Report, Department of Geospatial and Space Technology at the University of Nairobi.

Kingdon, J., Butynski, T. M. and De Jong, Y. (2008) 'Papio cynocephalus , Yellow Baboon', *The IUCN Red List of Threatened Species*.

Kinuthia, N. Z. (2002) *Ecotourism and Its Effects on the Livelihoods of the Host Community and Natural Resource Management: A Case Study of Amboseli, Kenya*. Master's Thesis, Faculty of Agriculture at the University of Nairobi.

Korzun, V. (2003) 'Molecular Markers and Their Applications in Cereals Breeding', *Cell and Molecular Biology Letters*, 7, pp. 811–820. doi: 10.1007/978-3-319-31703-8_6.

Kress, W. J. (2017) 'Plant DNA barcodes : Applications today and in the future', *Journal of Systematics and Evolution*, 55(4), pp. 291–307. doi: 10.1111/jse.12254.

Kress, W. J., Garcıa-Robeldo, C., Uriarte, M. and Erickson, D. L. (2015) 'DNA barcodes for ecology, evolution, and conservation', *Trends in Ecology & Evolution*, 30(1). doi: 10.1016/j.tree.2014.10.008.

Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A. and Janzen, D. H. (2005) 'Use of DNA barcodes to identify flowering plants', *PNAS*, 102(23), pp. 8369–8374.

Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. (2018) 'MEGA X : Molecular Evolutionary Genetics Analysis across Computing Platforms', *Mol. Biol. Evol.*, 35(6), pp. 1547–1549. doi: 10.1093/molbev/msy096.

Kuzoff, R. K., Sweere, J. A., Soltis, D. E., Soltis, P. S. and Zimmer, E. A. (1998) 'The Phylogenetic Potential of Entire 26S rDNA Sequences in Plants', *Mol. Biol. Evol.*, 15(3), pp. 251–263.

KWS (2008) 'Amboseli Ecosystem Management Plan, 2008-2018'. Nairobi, Kenya.

Leache, A. D. and Oaks, J. R. (2017) 'The Utility of Single Nucleotide Polymorphism (SNP) Data in Phylogenetics', *Annual Review of Ecology, Evolution, and Systematics*, 48, pp. 69–84. doi: 10.1146/annurev-ecolsys-110316-022645.

Li, X., Yang, Y., Henry, R. J., Rossetto, M., Wang, Y. and Chen, S. (2015) 'Plant DNA barcoding : from gene to genome', *Biol. Rev.*, 90, pp. 157–166. doi: 10.1111/brv.12104.

Liu, Yu-jing, Newmaster, S. G., Wu, X., Liu, Yue, Ragupathy, S., Motley, T. and Long, C.-L. (2013) 'Pinellia hunanensis (Araceae), a new species supported by morphometric analysis and DNA barcoding', *Phytotaxa*, 130(1), pp. 1–13.

Lopes, C. M., Barba, M. De, Boyer, F., Mercier, C., Filho, S., Heidtmann, L. M., Galiano, D., Kubiak, B. B., Langone, P., Garcias, F. M., Gielly, L., Coissac, E., Freitas, T. R. O. De and Taberlet, P. (2015) 'DNA metabarcoding diet analysis for species with parapatric vs sympatric distribution : a case study on subterranean rodents', *Heredity*. Nature Publishing Group, pp. 1–12. doi: 10.1038/hdy.2014.109.

Lopez-Alvarez, D., Lopez-Herranz, M. L., Betekhtin, A. and Catalan, P. (2012) 'A DNA Barcoding Method to Discriminate between the Model Plant Brachypodium distachyon and Its Close Relatives B . stacei and B . hybridum (Poaceae)', *PLoS One*, 7(12). doi: 10.1371/journal.pone.0051058.

Macher, J., Macher, T. and Leese, F. (2017) 'Combining NCBI and BOLD databases for OTU assignment in metabarcoding and metagenomic datasets : The BOLD _ NCBI _ Merger', *Metabarcoding and Metagenomics*, 1, pp. 1–8. doi: 10.3897/mbmg.1.22262.

Madesis, P., Ganopoulos, I., Ralli, P. and Tsaftaris, A. (2012) 'Barcoding the major Mediterranean leguminous crops by combining universal chloroplast and nuclear DNA sequence targets', *Genet. Mol. Res.*, 11(3), pp. 2548–2558.

Mallott, E. K., Garber, P. A. and Malhi, R. S. (2018) 'trnL outperforms rbcL as a DNA metabarcoding marker when compared with the observed plant component of the diet of wild white-faced capuchins (Cebus capucinus)', *PLoS ONE*, pp. 1–16. doi: 10.1371/journal.pone.0199556.

Maloukh, L., Kumarappan, A., Jarrar, M., Salehi, J., El-wakil, H. and Lakshmi, T. V. R. (2017) 'Discriminatory power of rbcL barcode locus for authentication of some of United Arab Emirates (UAE) native plants', *3 Biotech*. Springer Berlin Heidelberg, 7, p. 144. doi: 10.1007/s13205-017-0746-1.

Markham, A. C. (2017) 'Amboseli', *The International Encyclopaedia of Primatology*. doi: 10.1007/978-3-642-.

Mbane, J. O. (2012) *Land Use and Tenure Changes and Their Impact On the Kitenden Wildlife Corridor, Amboseli Ecosystem, Kenya*. Master's Thesis, School of Biological Sciences at the University of Nairobi.

Mcinnis, M. L., Vavra, M. and Krueger, W. C. (1983) 'A Comparison of Four Methods Used to Determine the Diets of Large Herbivores', *Journal of Range Management*, 36(3).

Meier, R., Shiyang, K., Vaidya, G. and Ng, P. K. L. (2006) 'DNA Barcoding and Taxonomy in Diptera : A Tale of High Intraspecific Variability and Low Identification Success', *Systematic Biology*, 55(5), pp. 715–728.

Michiels, A., Ende, W. Van Den, Tucker, M., Riet, L. Van and Laere, A. Van (2003) 'Extraction of high-quality genomic DNA from latex-containing plants', *Analytical Biochemistry*, 315, pp. 85–89. doi: 10.1016/S0003-2697(02)00665-6.

Mishler, B. D., Lewis, L. A., Buchheim, M. A., Renzaglia, K. S., Garbary, D. J., Delwiche, C. F., Zechman, F. W., Kantz, T. S. and Chapman, R. L. (1994) 'Phylogenetic Relationships of the " Green Algae " and " Bryophytes "', *Annals of the Missouri Botanical Gargden*, 81(3), pp. 451–483.

Mishra, P., Kumar, A., Nagireddy, A., Shukla, A. K. and Sundaresan, V. (2017) 'Evaluation of single and multilocus DNA barcodes towards species delineation in complex tree genus Terminalia', *PLoS ONE*, 12(8), pp. 1–18.

Mishra, P., Kumar, A., Rodrigues, V., Shukla, A. K. and Sundaresan, V. (2016) 'Feasibility of nuclear ribosomal region ITS1 over ITS2 in barcoding taxonomically challenging genera of subtribe ... Feasibility of nuclear ribosomal region ITS1 over ITS2 in barcoding taxonomically challenging genera of subtribe Cassiinae (Fabaceae)', *PeerJ*, 4, p. e2638. doi: 10.7717/peerj.2638.

Mishra, P., Kumar, A., Sivaraman, G., Shukla, A. K., Kaliamoorthy, R., Slater, A. and Velusamy, S. (2017) 'Character-based DNA barcoding for authentication and conservation of IUCN Red listed threatened species of genus Decalepis (Apocynaceae)', *Scientific Reports*. Springer US, 7. doi: 10.1038/s41598-017-14887-8.

Nakahara, F., Ando, H., Ito, H., Murakami, A., Morimoto, N., Yamasaki, M., Takayanagi, A. and Isagi, Y. (2015) 'The applicability of DNA barcoding for dietary analysis of sika deer', *DNA Barcodes*, 3, pp. 200–206. doi: 10.1515/dna-2015-0021.

Nguyen, H. D. T. and Seifert, K. A. (2008) 'Description and DNA barcoding of three new species of Leohumicola from South Africa and the United States', *Persoonia*, 21, pp. 57–69. doi: 10.3767/003158508X361334.

Okello, M. M., Kenana, L., Maliti, H., Kiringe, J. W., Kanga, E., Warinwa, F., Bakari, S., Ndambuki, S., Massawe, E., Sitati, N., Kimutai, D., Mwita, M., Gichohi, N., Muteti, D., Ngoru, B. and Mwangi, P. (2016) 'Population density of elephants and other key large herbivores in the Amboseli ecosystem of Kenya in relation to droughts', *Journal of Arid Environments*. Elsevier Ltd, 135, pp. 64–74. doi: 10.1016/j.jaridenv.2016.08.012.

Okeyo, J. M. and Bosch, C. H. (2007) *Senna italica Mill.*, *[Internet] Record from PROTA4U*.

Olsen, G. J., Overbeek, R., Larsen, N., Marsh, T. L., McCaughey, M. J., Maciukenas, M. A., Kuan, W.-M., Macke, T. J., Xing, Y. and Woese, C. R. (1992) 'The Ribosomal Database Project', *Nucleic Acids Research*, 20(suppl), pp. 2199–2200. Available at: http://dx.doi.org/10.1093/nar/20.suppl.2199.

Pang, X., Liu, C., Shi, L., Liu, R., Liang, D., Li, H., Cherny, S. S. and Chen, S. (2012) 'Utility of the trnH – psbA Intergenic Spacer Region and Its Combinations as Plant DNA Barcodes : A Meta-Analysis', *PLoS One*, 7(11), pp. 1–9. doi: 10.1371/journal.pone.0048833.

Patwardhan, A., Ray, S. and Roy, A. (2014) 'Molecular Markers in Phylogenetic Studies-A Review', *Journal of Phylogenetics & Evolutionary Biology*, 2(2). doi: 10.4172/2329-9002.1000131.

Pauls, S. U., Blahnik, R. J., Zhou, X., Wardwell, C. T. and Holzenthal, R. W. (2010) 'DNA barcode data confirm new species and reveal cryptic diversity in Chilean Smicridea (Smicridea) (Trichoptera: Hydropsychidae)', *J. N. Am. Benthol. Soc.*, 29(3), pp. 1058–1074. doi: 10.1899/09-108.1.

Pei, N., Chen, B. and Kress, W. J. (2017) 'Advances of Community-Level Plant DNA Barcoding in China', *Frontiers in Plant Science*, 8(225). doi: 10.3389/fpls.2017.00225.

Peterson, P. M., Romaschenko, K. and Arrieta, Y. H. (2015) 'A molecular phylogeny and classification of the Eleusininae with a new genus , Micrachne (Poaceae : Chloridoideae : Cynodonteae)', *Taxon*, 64(3), pp. 445–467.

Peterson, P. M., Romaschenko, K., Arrieta, Y. H. and Saarela, J. M. (2014) 'A molecular phylogeny and new subgeneric classification of Sporobolus (Poaceae : Chloridoideae : Sporobolinae)', *TAXON*, 63(6), pp. 1212–1243.

Peterson, P. M., Romaschenko, K., Snow, N. and Johnson, G. (2012) 'A molecular phylogeny and classification of Leptochloa (Poaceae : Chloridoideae : Chlorideae) sensu lato and A molecular phylogeny and classification of Leptochloa (Poaceae: Chloridoideae: Chlorideae) sensu lato and related genera', *Annals of Botany*, 109, pp. 1317–1329. doi: 10.1093/aob/mcs077.

Peterson, P. M., Romaschenko, K. and Soreng, R. J. (2014) 'A laboratory guide for generating DNA barcodes in grasses : a case study of Leptochloa s . l . (Poaceae : Chloridoideae)', *Webbia: Journal of Plant Taxonomy and Geography*. Taylor & Francis, 69(1), pp. 1–12. doi: 10.1080/00837792.2014.927555.

Pettengill, J. B. and Neel, M. C. (2010) 'An Evaluation of Candidate Plant DNA Barcodes and Assignment Methods In Diagnosing 29 Species In The Genus Agalinis (Orobanchaceae)', *American Journal of Botany*, 97(8), pp. 1391–1406. doi: 10.3732/ajb.0900176.

Post, D. G. (1982) 'Feeding Behavior of Yellow Baboons (Papio cynocephalus) in the Amboseli National Park, Kenya', *International Journal of Primatology*, 3(4).

Post, D. G., Hausfater, G. and McCuskey, S. A. (1980) 'Feeding Behavior of Yellow Baboons (Papio cynocephalus): Relationship to Age, Gender and Dominance Rank', *Folia primatol.*, 34, pp. 170–195.

Purty, R. and Chatterjee, S. (2016) 'DNA Barcoding : An Effective Technique in Molecular Taxonomy', *Austin J Biotechnol Bioeng. 2016*, 3(1), p. 1059.

Qiagen (2008) 'QIAquick PCR Purification Kit Protocol', in *QIAquick Spin Handbook*, pp. 19–20.

Qiagen (2015) *DNeasy ® Plant Handbook DNeasy Plant Mini Kit and tissues , or fungi Sample & Assay Technologies QIAGEN Sample and Assay Technologies*.

Quattrocchi, U. (2006) *CRC World Dictionary of Grasses: Common Names, Scientific Names, Eponyms, Synonyms, and Etymology*.

Quemere, E., Hilbert, F., Miquel, C., Lhuillier, E., Rasolondraibe, E., Champeau, J., Rabarivola, C., Nusbaumer, L., Chatelain, C., Gautier, L., Ranirison, P., Crouau-Roy, B., Taberlet, P. and Chikhi, L. (2013) 'A DNA Metabarcoding Study of a Primate Dietary Diversity and Plasticity across Its Entire Fragmented Range', *Plos ONE*, 8(3). doi: 10.1371/journal.pone.0058971.

Ratnasingham, S. and Herbert, P. D. N. (2007) 'BOLD: The Barcode of Life Data System (www.barcodinglife.org)', *Molecular Ecology Notes*, 7, pp. 355–364. doi: 10.1111/j.1471-8286.2006.01678.x.

Rayé, G., Miquel, C., Coissac, E., Redjadj, C., Loison, A. and Taberlet, P. (2011) 'New insights on diet variability revealed by DNA barcoding and high-throughput pyrosequencing: chamois diet in autumn as a case study', *Ecol Res*, 26, pp. 265–276. doi: 10.1007/s11284-010-0780-5.

Rogers, S. O. and Bendich, A. J. (1987) 'Ribosomal RNA genes in plants: variability in copy number and in intergenic spacer', *Plant Molecular Biology*, 9, pp. 509–520. doi: 10.1007/BF00015882.

Ross, H. A., Murugan, S. and Li, W. L. S. (2008) 'Testing the Reliability of Genetic Methods of Species Identification via Simulation', *Syst. Biol.*, 57(2), pp. 216–230. doi: 10.1080/10635150802032990.

Rydberg, A. (2010) *DNA barcoding as a tool for the identification of unknown plant material: A case study on medicinal roots traded in the medina of Marrakech*. Master's thesis, Uppsala Universitet.

Sahu, S. K., Thangaraj, M. and Kathiresan, K. (2012) 'DNA Extraction Protocol for Plants with High Levels of Secondary Metabolites and Polysaccharides without Using Liquid Nitrogen and Phenol', *ISRN Molecular Biology*, p. doi:10.5402/2012/205049. doi: 10.5402/2012/205049.

Santos, C. and Pereira, F. (2016) 'Design and evaluation of PCR primers for amplification of four chloroplast DNA regions in plants', *Conservation Genetics Resources*. Springer Netherlands. doi: 10.1007/s12686-016-0605-0.

Santos, R. C., Pires, J. L. and Correa, R. X. (2012) 'Morphological characterization of leaf , flower , fruit and seed traits among Brazilian Theobroma L . species', *Genet Resour Crop Evol*, 59, pp. 327–345. doi: 10.1007/s10722-011-9685-6.

Schilling, E. E. and Floden, A. (2018) 'Barcoding the Asteraceae of Tennessee, tribes Heliantheae and Millerieae', *Phytoneuron*, 83, pp. 1–9.

Shaw, J., Lickey, E. B., Schilling, E. E. and Small, R. L. (2007) 'Comparison Of Whole Chloroplast Genome Sequences To Choose Noncoding Regions For Phylogenetic Studies In Angiosperms : The Tortoise And The Hare III', *American Journal of Botany*, 94(3), pp. 275–288.

Shefferly, N. (2004) *Papio cynocephalus*, *(On-line) Animal Diversity Web*.

Shen, Y.-Y., Chen, X. and Murphy, R. W. (2013) 'Assessing DNA Barcoding as a Tool for Species Identification and Data Quality Control', *PLoS ONE*, 8(2). doi: 10.1371/journal.pone.0057125.

Shokralla, S., Gibson, J. F., Nikbakht, H., Janzen, D. H., Hallwachs, W. and Hajibabaei, M. (2014) 'Next-generation DNA barcoding : using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens', *Molecular Ecology Resources*, 14, pp. 892–901. doi: 10.1111/1755-0998.12236.

Shrestha, R. and Wegge, P. (2006) 'Determining the Composition of Herbivore Diets in the Trans-Himalayan Rangelands : A Comparison of Field Methods', *Rangeland Ecology and Management*. Elsevier, 59(5), pp. 512–518. doi: 10.2111/06-022R2.1.

Soininen, E. M., Valentini, A., Coissac, E., Miquel, C., Gielly, L., Brochmann, C., Brysting, A. K., Sønstebø, J. H., Ims, R. A., Yoccoz, N. G. and Taberlet, P. (2009) 'Analysing diet of small herbivores : the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures', *Frontiers in Zoology*, 6, p. 16. doi: 10.1186/1742-9994-6-16.

Solé, M., Beckmann, O., Pelz, B., Kwet, A. and Engels, W. (2007) 'Studies on Neotropical Fauna and Environment Stomach-flushing for diet analysis in anurans : an improved protocol evaluated in a case study in Araucaria forests , southern Brazil *', *Studies on Neotropical Fauna and Environment*, 0521. doi: 10.1080/01650520400025704.

Soltis, D. E. and Soltis, P. S. (1997) 'Phylogenetic Relationships in Saxifragaceae Sensu Lato: A Comparison of Topologies Based On 18S rDNA and rbcL Sequences', *American Journal of Botany*, 84(4), pp. 504–522. doi: 10.2307/2446027.

Soltis, P. S., Soltis, D. E., Wolf, P. G., Nickrent, D. L., Chaw, S. and Chapman, R. L. (1999) 'The Phylogeny of Land Plants Inferred from 18S rDNA Sequences : Pushing the Limits of rDNA Signal?', *Mol. Biol. Evol.*, 16(12), pp. 1774–1784.

Souza, S. De, Andressa, B. I., Generoso, L., Dantas, I. B. and Soares, F. (2011) 'An improved method for genomic DNA extraction from strawberry leaves', *Ciência Rural*, 41(8), pp. 1383–1389.

Srivastava, A. K. and Schlessinger, D. (1991) 'Structure and organization of ribosomal DNA', *Biochimie*, 73(6), pp. 631–638. doi: https://doi.org/10.1016/0300-9084(91)90042-Y.

Srivathsan, A., Ang, A., Vogler, A. P. and Meier, R. (2016) 'Fecal metagenomics for the simultaneous assessment of diet , parasites , and population genetics of an understudied primate', *Frontiers in Zoology*. Frontiers in Zoology, pp. 1–13. doi: 10.1186/s12983-016-0150-4.

Srivathsan, A., Sha, J., Vogler, A. P. and Meier, R. (2014) 'Comparing the effectiveness of metagenomics and metabarcoding for diet analysis of a leaf-feeding monkey (Pygathrix nemaeus)'. doi: 10.1111/1755-0998.12302/full.

Staats, M., Arulandhu, A. J., Gravendeel, B., Holst-jensen, A., Scholtens, I., Peelen, T., Prins, T. W. and Kok, E. (2016) 'Advances in DNA metabarcoding for food and wildlife forensic species identification', *Analytical and Bioanalytical Chemistry*. Analytical and Bioanalytical Chemistry, 408, pp. 4615–4630. doi: 10.1007/s00216-016-9595-8.

Stamatakis, A. (2014) 'RAxML version 8 : a tool for phylogenetic analysis and post-analysis of large phylogenies', *Bioinformatics*, 30(9), pp. 1312–1313. doi: 10.1093/bioinformatics/btu033.

Stamatakis, A. (2016) 'RAxML-VI-HPC : Maximum Likelihood-Based Phylogenetic Analyses with Thousands of Taxa and Mixed Models RAxML-VI-HPC : maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models', *Bioinformatics*, 22(21), pp. 2688–2690. doi: 10.1093/bioinformatics/btl446.

Staner, P. (1938) 'Révision des espèces congolaises des genres Cycnium et Rhamphicarpa (Scrophulariacées)', *Bulletin du Jardin botanique de l'État a Bruxelles*, 15(2), pp. 147–151.

Stevens, E. J., Stevens, S. J., Gates, R. N., Eskridge, K. M. and Waller, S. . (1987) 'Procedure for Fecal Cuticle Analysis of Herbivore Diets', *Journal of Range Management*, 40(2), pp. 187–189.

Su, X., Liu, Y. P., Chen, Z. and Chen, K. L. (2016) 'Evaluation of candidate barcoding markers in Orinus (Poaceae)', *Genetics and Molecular Research*, 15(2), p. gmr.15027714.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. and Willerslev, E. (2012) 'Towards next-generation biodiversity assessment using DNA metabarcoding', *Molecular Ecology*, 21, pp. 2045–2050.

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermat, T., Corthier, G., Brochmann, C. and Willerslev, E. (2007) 'Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding', *Nucleic Acids Research*, 35(3). doi: 10.1093/nar/gkl938.

Taberlet, P., Gielly, L., Pautou, G. and Bouvet, J. (1991) 'Universal Primers For Amplification Of 3 Noncoding Regions Of Chloroplast Dna', *Plant Molecular Biology*, 17, pp. 1105–1109. doi: 10.1007/BF00037152.

Tahir, A., Hussain, F., Ahmed, N., Ghorbani, A. and Jamil, A. (2018) 'Assessing universality of DNA barcoding in geographically isolated selected desert medicinal species of Fabaceae and Poaceae', *PeerJ*. doi: 10.7717/peerj.4499.

Thomsen, P. F. and Willerslev, E. (2015) 'Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity', *Biological Conservation*. Elsevier Ltd, 183, pp. 4–18. doi: 10.1016/j.biocon.2014.11.019.

Valentini, A., Miquel, C., Nawaz, M. A., Bellemain, E., Coissac, E., Pompanon, F., Gielly, L., Cruaud, C., Nascetti, G., Wincker, P., Swenson, J. E. and Taberlet, P. (2009) 'New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach', *Molecular Ecology Resources*, 9, pp. 51–60. doi: 10.1111/j.1755-0998.2008.02352.x.

Vijayan, K. and Tsou, C. H. (2010) 'DNA barcoding in plants : taxonomy perspective', *Current Science*, 99(11), pp. 1530–1541.

Waldchen, J., Rzanny, M., Seeland, M. and Mader, P. (2018) 'Automated plant species identification — Trends and future directions', *PLoS Comput Biol*, 14(4), p. e1005993.

Wang, X., Liu, C., Huang, L., Bengtsson-Palme, J., Chen, H., Zhang, J., Cai, D. and Li, J.-Q. (2014) 'ITS1: a DNA barcode better than ITS2 in eukaryotes?', *Molecular Ecology Resources*. doi: 10.1111/1755-0998.12325.

Wattoo, J. I., Saleem, M. Z., Shahzad, M. S., Arif, A., Hameed, A. and Saleem, M. A. (2016) 'DNA Barcoding: Amplification and sequence analysis of rbcl and matK genome regions in three divergent plant species', *Advancements in Life Sciences*, 4(1), pp. 3–7.

Whitlock, B. A., Hale, A. M. and Groff, P. A. (2010) 'Intraspecific Inversions Pose a Challenge for the trnH-psbA Plant DNA Barcode', *PLoS one*, 5(7). doi: 10.1371/journal.pone.0011533.

Williams, L. A. J. (1972) *Geology of the Amboseli Area*.

Wilson, B. L., Brainerd, R. E., Otting, N., Knaus, B. J. and Nelson, J. K. (2014)

‘IDENTIFICATION AND TAXONOMIC STATUS OF CORDYLANTHUS TENUIS SUBSP. PALLESCENS (OROBANCHACEAE)’, *Madroño*, 61(1), pp. 64–76. Available at: https://www.jstor.org/stable/43288668.

Wilson, J. J., Rougerie, R., Schonfeld, J., Janzen, D. H., Hallwachs, W., Hajibabaei, M., Kitching, I. J., Haxaire, J. and Hebert, P. D. N. (2011) ‘When species matches are unavailable are DNA barcodes correctly assigned to higher taxa ? An assessment using sphingid moths’, *BMC Ecology*, 11, p. 18.

Wolfe, K. H., Li, W.-H. and Sharp, P. M. (1987) ‘Rates of nucleotide substitution vary greatly among plant mitochondrial , chloroplast , and nuclear DNAs’, *Proc. Nati. Acad. Sci. USA*, 84, pp. 9054–9058.

Wu, F., Li, M., Liao, B., Shi, X. and Xu, Y. (2019) ‘DNA Barcoding Analysis and Phylogenetic Relation of Mangroves in Guangdong Province, China’, *Forests*, 10(56), pp. 1–13. doi: 10.3390/f10010056.

Xue, Y., Chen, H., Yang, J. R., Liu, M., Huang, B. and Yang, J. (2018) ‘Distinct patterns and processes of abundant and rare eukaryotic plankton communities following a reservoir cyanobacterial bloom’, *The ISME Journal*. Springer US. doi: 10.1038/s41396-018-0159-0.

Yao, H., Song, J., Liu, C., Luo, K., Han, J., Li, Y., Pang, X., Xu, H., Zhu, Y., Xiao, P. and Chen, S. (2010) ‘Use of ITS2 Region as the Universal DNA Barcode for Plants and Animals’, *PLoS ONE*, 5(10). doi: 10.1371/journal.pone.0013102.

Zwickl, D. J. (2006) *Genetic Algorithm Approaches for The Phylogenetic Analysis of Large Biological Sequence Datasets*. PhD Dissertation, Faculty of the Graduate School of the University of Texas at Austin.

APPENDICES

Appendix 1: Molecular and morphological identities of monocotyledons

(Source: E. A. Archie)

Sample and herbarium ID: *Drake-brockmania somalensis* Stapf. (aka *Dinebra somalensis* (Stapf) P.M.Peterson & N.Snow; Quattrocchi, 2006; Peterson *et al.*, 2012).

- An annual plant; culms prostrate, spreading, sometimes rooting at the nodes, much branched, flowering on the ascending lateral branches. Inflorescence often subcapitate, composed of 2–6 spikes on a central axis.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Dinebra haareri*	97.90%	KP873335
BOLD	Unidentifiable	<95.00%	N/A
***trn*L BLAST**			
GenBank®	*Leptochloa virgata*	99.10%	KY432784
BOLD	*Enteropogon macrostachyus*	99.10%	UHURU632-14|Enteropogon macrostachyus|trnL-F|KR737769

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Dinebra haareri	Ambiguous

* Ambiguous – multiple genera were identified.

Sample ID: *Odyssea paucinervis* (Nees) Stapf.

- Perennial mat grass, with long stout deeply penetrating rhizomes bearing dense tufts of spiny glaucous shoots at the nodes. Leaf-blades linear-lanceolate, pilose on both sides, often only sparingly, margins scabrid especially towards the tip. Inflorescence narrowly elliptic to elliptic-oblong, rarely almost linear, with 3–14 spikelets. Spikelets 4–9-flowered, elliptic to narrowly elliptic-oblong.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Odyssea paucinervis*	97.90%	KX582389
BOLD	Unidentifiable	<95.00%	N/A
***trn*L BLAST**			
GenBank®	*Triodia wiseana*	99.80%	NC_037161
	Halopyrum mucronatum	99.80%	KY432780
BOLD	*Enteropogon macrostachyus*	99.70%	UHURU632-14|Enteropogon macrostachyus|trnL-F|KR737769

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Odyssea paucinervis	Unidentifiable

Sample ID: *Setaria verticillata* (L.) P. Beauv.

- An annual grass with erect or decumbent stems that grows up to one metre in height. The leaf blades have a long sheath around the stem. The inflorescence is a dense panicle which tapers at both ends. It contains many small spikelets and bristles. The bristles have tiny backwards-pointing barbs.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Setaria verticillata*	100.00%	KR733690
BOLD	*Setaria verticillata*	100.00%	UHURU1345-15|Setaria verticillata|ITS|KR734307
***trn*L BLAST**			
GenBank®	*Setaria verticillata*	100.00%	KR738477
BOLD	*Setaria verticillata*	100.00%	UHURU1369-15|Setaria verticillata|trnL-F|KR738477

***ITS*1 phylogeny**	***trn*L phylogeny**
Setaria verticillata	*Setaria verticillata*

Sample ID: *Sporobolus consimilis* Fresen.

***trn*L BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	*Sporobolus cryptandrus*	99.50%	MG709435
BOLD	*Sporobolus agrostoides*	98.20%	UHURU871-14|Sporobolus agrostoides|trnL-F|KR738515

***trn*L phylogeny**
Sporobolus spp.

NB. The sequencing of the *ITS*1 region was unsuccessful.

Sample ID: Grass A

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Enneapogon cenchroides*	100.00%	KR734310
BOLD	*Enneapogon cenchroides*	100.00%	UHURU1373-15|Enneapogon cenchroides|ITS|KR733746
***trn*L BLAST**			
GenBank®	*Enneapogon scoparius*	99.50%	DQ655895
BOLD	*Enneapogon cenchroides*	99.10%	UHURU908-14|Enneapogon cenchroides|trnL-F|KR738224
	Enneapogon persicus	99.10%	UHURU1251-15|Enneapogon persicus|trnL-F|KR738619

***ITS*1 phylogeny**	***trn*L phylogeny**
Enneapogon scoparius	*Enneapogon scoparius*

Sample ID: Grass B

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Chloris gayana*	99.70%	KR734190
	Chloris virgata	99.70%	KR734150
BOLD	*Chloris virgata*	99.70%	UHURU655-14|Chloris virgata|ITS|KR733910
***trn*L BLAST**			
GenBank®	*Chloris virgata*	100.00%	KX765279
BOLD	*Chloris virgata*	99.90%	UHURU648-14|Chloris virgata|trnL-F|KR738355

***ITS*1 phylogeny**	***trn*L phylogeny**
Chloris spp.	*Chloris* spp.

Sample ID: Grass C

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Sporobolus marginatus*	97.90%	KM010441
	Sporobolus coromandelianus	97.90%	KM010403
BOLD	Unidentifiable	<95.00%	N/A
***trn*L BLAST**			
GenBank®	*Sporobolus pyramidatus*	99.00%	EF156733
BOLD	*Sporobolus africanus*	97.80%	UHURU1035-14|Sporobolus africanus|trnL-F|KR737946

*Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Sporobolus spp.	*Sporobolus pyramidatus*

NB. - *Sporobolus marginatus*, *S. coromandelianus*, *S. pyramidatus* and *S. cordofanus* were reclassified to the subgenus *Sporobolus* subsect. *Pyramidati* P.M. Peterson (Peterson *et al.*, 2014).
- *Sporobolus africanus* was reclassified to the subgenus *Sporobolus* subsect. *Sporobolus* (Peterson, Romaschenko, Arrieta, *et al.*, 2014).

Sample ID: Grass D

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Cenchrus mezianus*	97.30%	KR733720
BOLD	*Pennisetum mezianum*	97.30%	UHURU603-14|Pennisetum mezianum|ITS|KR734250
***trn*L BLAST**			
GenBank®	*Cenchrus mezianus*	99.80%	KR737645
BOLD	*Pennisetum mezianum*	99.80%	UHURU603-14|Pennisetum mezianum|trnL-F|KR738535
	Cenchrus ciliaris	99.80%	UHURU356-14|Cenchrus ciliaris|trnL-F

***ITS*1 phylogeny**	***trn*L phylogeny**
Cenchrus ciliaris	*Pennisetum/Cenchrus* spp.

Sample ID: Grass E

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Cynodon plectostachyus*	99.70%	KR734227
BOLD	*Cynodon nlemfuensis*	99.70%	UHURU859-14|Cynodon nlemfuensis|ITS|KR734227
***trn*L BLAST**			
GenBank®	Unidentifiable	<95.00%	N/A
BOLD	Unidentifiable	<95.00%	N/A

***** Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Cynodon spp.	Unidentifiable

Sample ID: Grass E2

***trn*L BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	*Cynodon plectostachyus*	99.60%	KR738310
BOLD	*Cynodon nlemfuensis*	99.60%	UHURU284-14|Cynodon nlemfuensis|trnL-F|KR738310

***trn*L phylogeny**
Cynodon spp.

NB. The sequencing of the *ITS*1 region was unsuccessful.

Sample ID: Grass F

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Disakisperma yemenicum*	97.80%	KF574403
BOLD	*Enteropogon macrostachyus*	96.20%	UHURU634-14|Enteropogon macrostachyus|ITS|KR733941
***trn*L BLAST**			
GenBank®	*Cynodon dactylon*	99.40%	MG709452
	Astrebla pectinata	99.40%	KT168391
BOLD	*Enteropogon macrostachyus*	99.30%	UHURU632-14|Enteropogon macrostachyus|trnL-F|KR737769

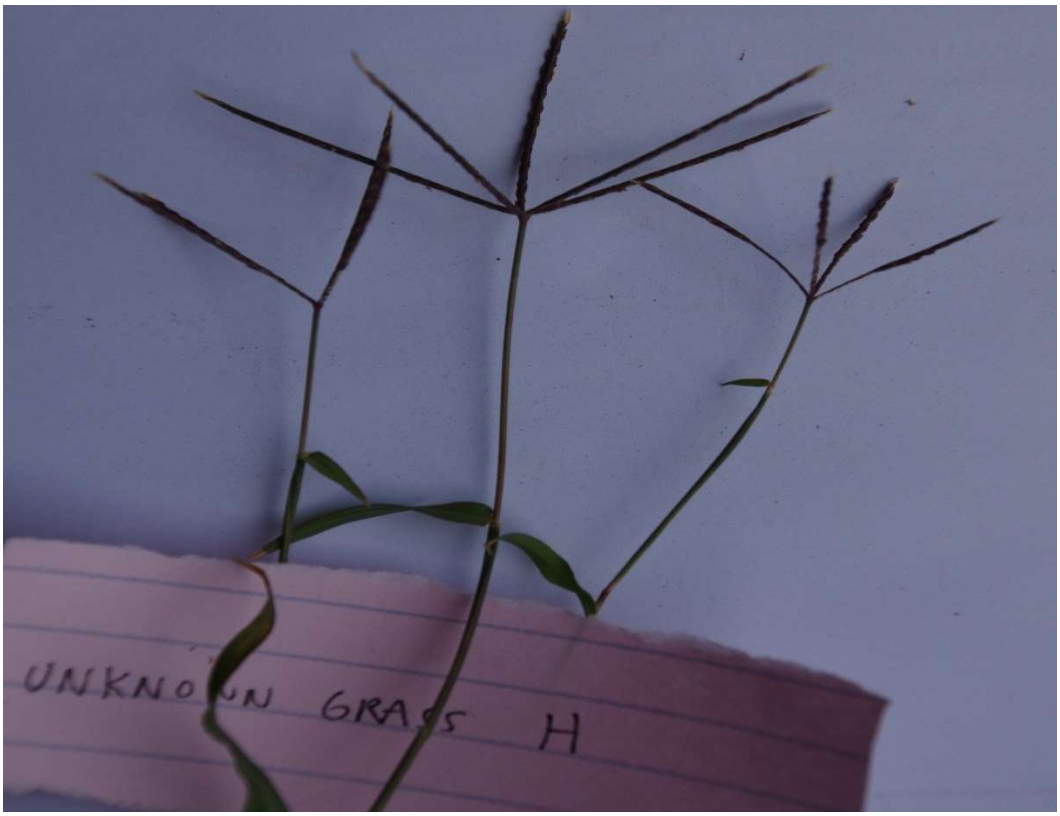
***ITS*1 phylogeny**	***trn*L phylogeny**
Disakisperma yemenicum	*Disakisperma dubium*

No image available.

Sample ID: Grass G

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Cenchrus mezianus*	97.30%	KR733720
BOLD	*Pennisetum mezianum*	97.30%	UHURU603-14|Pennisetum mezianum|ITS|KR734250
***trn*L BLAST**			
GenBank®	*Cenchrus mezianus*	99.70%	KR737645
BOLD	*Pennisetum mezianum*	99.70%	UHURU603-14|Pennisetum mezianum|trnL-F|KR738535

***ITS*1 phylogeny**	***trn*L phylogeny**
Cenchrus ciliaris	*Cenchrus setiger*



Sample ID: Grass H

***ITS*1 BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	*Cynodon plectostachyus*	100.00%	KR734227
BOLD	*Cynodon nlemfuensis*	100.00%	UHURU859-14|Cynodon nlemfuensis|ITS|KR734227

***ITS*1 phylogeny**
Cynodon spp.

NB. The sequencing of the *trn*L region was unsuccessful.

Sample ID: Grass I

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Cynodon plectostachyus*	99.70%	KR734227
BOLD	*Cynodon nlemfuensis*	99.70%	UHURU859-14|Cynodon nlemfuensis|ITS|KR734227
***trn*L BLAST**			
GenBank®	Unidentifiable	<95.00%	N/A
BOLD	Unidentifiable	<95.00%	N/A

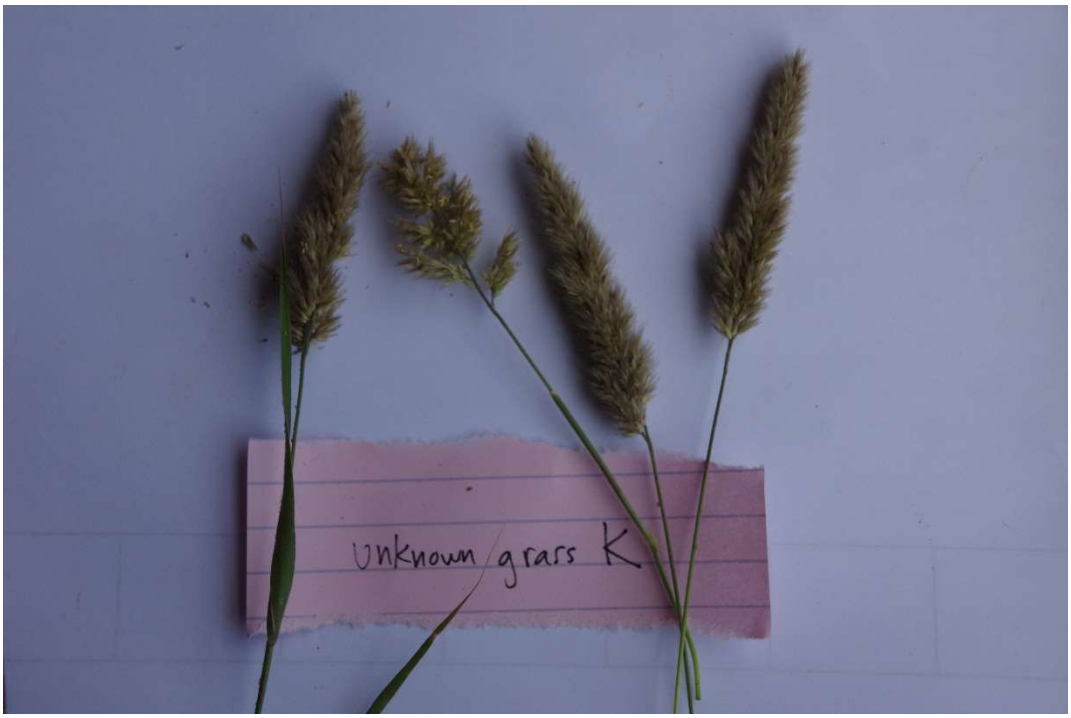
* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Cynodon spp.	Unidentifiable

Sample ID: Grass J

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Disakisperma yemenicum*	97.90%	KF574403
BOLD	*Enteropogon macrostachyus*	96.30%	UHURU634-14|Enteropogon macrostachyus|ITS|KR733941
***trn*L BLAST**			
GenBank®	*Cynodon dactylon*	99.40%	KY024482
	Astrebla pectinata	99.40%	KT168391
BOLD	*Enteropogon macrostachyus*	99.30%	UHURU632-14|Enteropogon macrostachyus|trnL-F|KR737769

***ITS*1 phylogeny**	***trn*L phylogeny**
Disakisperma yemenicum	*Disakisperma dubium*



Sample ID: Grass K

***ITS*1 BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	*Enneapogon cenchroides*	99.70%	KR734310
BOLD	*Enneapogon cenchroides*	99.80%	UHURU1373-15|Enneapogon cenchroides|ITS|KR733746

***ITS*1 phylogeny**
Enneapogon scoparius

NB. The sequencing of the *trn*L region was unsuccessful.

Sample ID: Grass L

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Chloris mossambicensis*	98.90%	HM347036
BOLD	*Chloris nutans*	95.60%	UHURU1317-15|Chloris
***trn*L BLAST**			
GenBank®	*Cynodon dactylon*	99.10%	MG709452
BOLD	*Chloris roxburghiana*	99.10%	UHURU516-14|Chloris roxburghiana|trnL-F

***ITS*1 phylogeny**	***trn*L phylogeny**
Tetrapogon cenchriformis	Unidentifiable

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

NB. *Tetrapogon* spp. are also known as *Chloris* spp. (Quattrocchi, 2006; Petersonn *et al.*, 2015).

No image available.

Grass M.

***ITS*1 BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	*Sporobolus marginatus*	97.90%	KM010441
BOLD	Unidentifiable	<95.00%	N/A

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**
Sporobolus spp.

NB.: - The sequencing of the *trn*L region was unsuccessful.

- *Sporobolus marginatus*, *S. cordofanus* and *S. coromandelianus* were reclassified to *Sporobolus* subsect. *Pyramidati* P.M. Peterson (Peterson *et al.*, 2014).

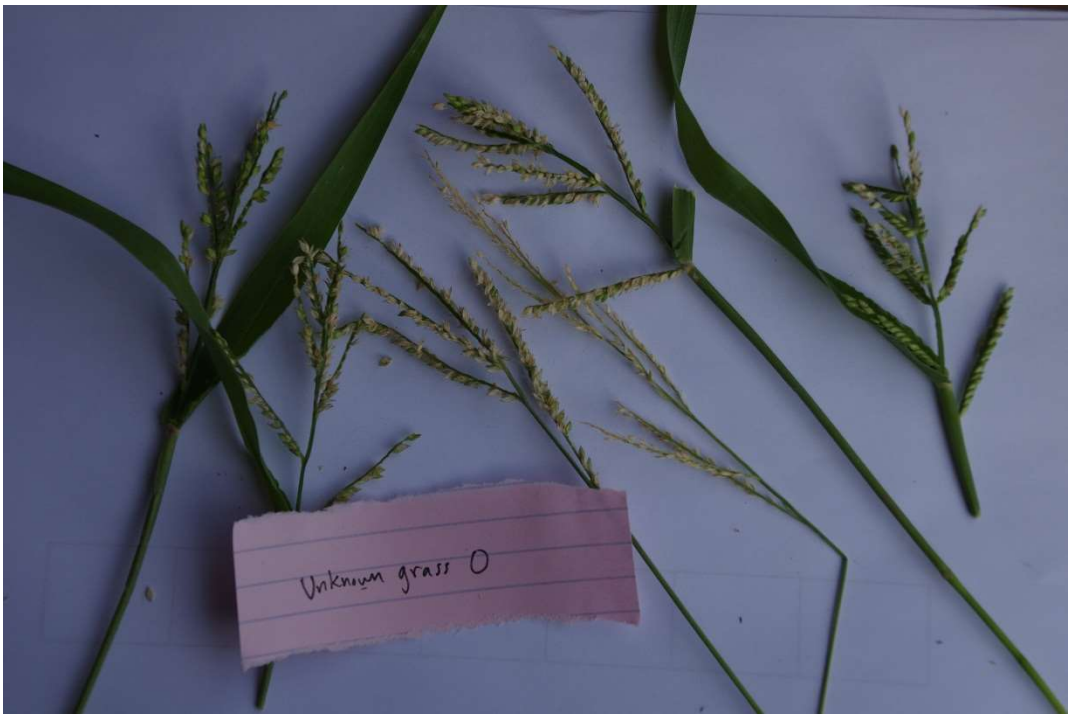
Sample ID: Grass N

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Sporobolus virginicus*	97.70%	MF029711
BOLD	Unidentifiable	<95.00%	N/A
***trn*L BLAST**			
GenBank®	*Sporobolus michauxianus*	99.30%	MG709435
BOLD	*Sporobolus africanus*	98.10%	UHURU1035-14|Sporobolus africanus|trnL-F|KR737946

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Sporobolus spp.	*Crypsis vaginiflora*

NB. *Crypsis vaginiflora* was reclassified to the subgenus *Sporobolus* subsect. *Crypsis* (Peterson *et al.*, 2014).

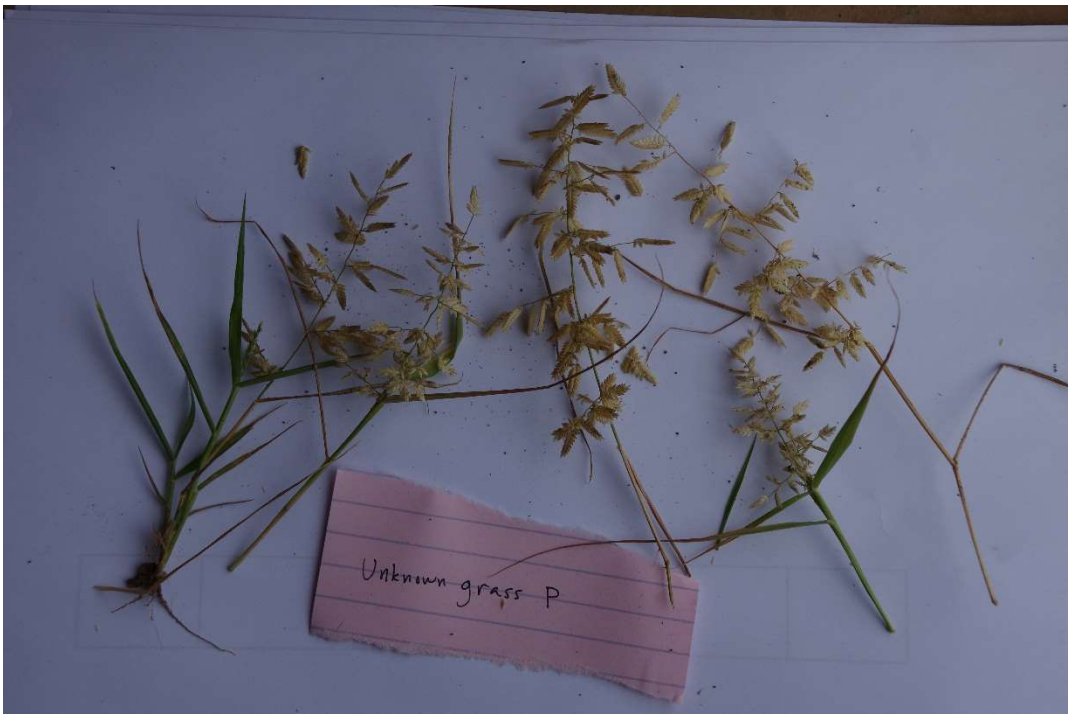


Sample ID: Grass O

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Brachiaria deflexa*	100.00%	KR734067
BOLD	*Eriochloa fatmensis*	100.00%	UHURU1209-15|Eriochloa fatmensis|ITS|KR734067
***trn*L BLAST**			
GenBank®	*Urochloa brachyura*	99.90%	KR737656
BOLD	*Urochloa brachyura*	99.90%	UHURU1032-14|Urochloa brachyura|trnL-F|KR738144

***ITS*1 phylogeny**	***trn*L phylogeny**
Brachiaria/Eriochloa spp.	*Urochloa/Brachiaria/Eriochloa* spp.

NB. *Brachiaria, Urochloa, Eriochloa* and *Melinis* are considered to be a monophyletic group based on phylogenetic analysis (González and Morton, 2005).



Sample ID: Grass P

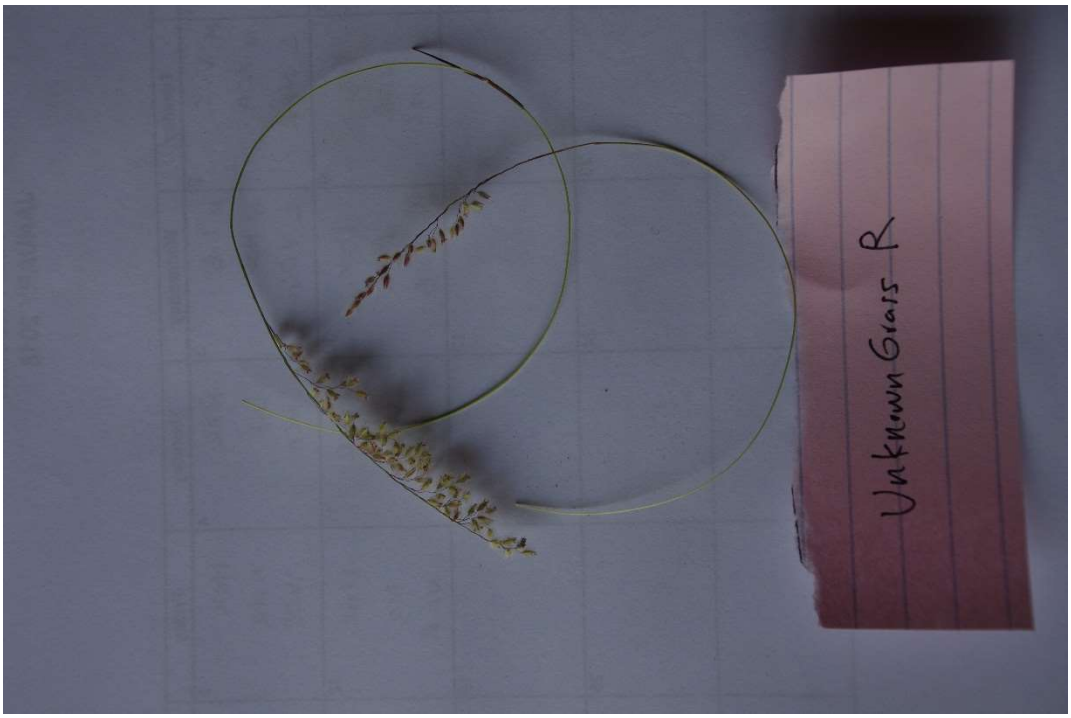
	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Eragrostis minor*	100.00%	KP057036
BOLD	*Eragrostis papposa*	98.20%	UHURU1170-14|Eragrostis papposa|ITS|KR734013
***trn*L BLAST**			
GenBank®	*Eragrostis secundiflora*	100.00%	MG709400
BOLD	*Sporobolus rangei*	99.80%	UHURU1231-15|Sporobolus rangei|trnL-F|KR737763

***ITS*1 phylogeny**	***trn*L phylogeny**
Eragrostis minor	*Eragrostis papposa*

Sample ID: Grass Q

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Chloris gayana*	99.70%	KR734190
	Chloris virgata	99.70%	KR734150
BOLD	*Chloris virgata*	99.70%	UHURU655-14|Chloris virgata|ITS|KR733910
***trn*L BLAST**			
GenBank®	*Chloris virgata*	100.00%	KX765279
BOLD	*Chloris virgata*	99.90%	UHURU648-14|Chloris virgata|trnL-F|KR738355

***ITS*1 phylogeny**	***trn*L phylogeny**
Chloris spp.	*Chloris* spp.



Sample ID: Grass R

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Tricholaena monachne*	99.50%	HM347025
BOLD	*Melinis repens*	97.10%	UHURU1166-14|Melinis repens|ITS|KR733716
***trn*L BLAST**			
GenBank®	*Tricholaena monachne*	99.90%	KY432799
BOLD	*Melinis repens*	99.20%	UHURU1165-14|Melinis repens|trnL-F|KR737732

***ITS*1 phylogeny**	***trn*L phylogeny**
Melinis repens	*Tricholaena monachne*

NB. *Tricholaena* spp. and *Melinis* spp. are synonymous (Quattrocchi, 2006).

Sample ID: Grass T

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Sporobolus agrostoides*	99.80%	KR734236
BOLD	*Sporobolus agrostoides*	99.80%	UHURU871-14|Sporobolus agrostoides|ITS|KR734236
***trn*L BLAST**			
GenBank®	*Sporobolus cryptandrus*	98.60%	MG709435
BOLD	*Sporobolus agrostoides*	98.10%	UHURU871-14|Sporobolus agrostoides|trnL-F|KR738515

***ITS*1 phylogeny**	***trn*L phylogeny**
Sporobolus fimbriatus	*Sporobolus agrostoides*

NB. *Sporobolus agrostoides* and *S. fimbriatus* were reclassified to the subgenus *Sporobolus* subsect. *Fimbriatae* (Peterson, Romaschenko, Arrieta, *et al.*, 2014).

Sample ID: Grass AA; herbarium ID: *Brachiaria dictyoneura* (Fig. & De Not.) Stapf.

- A densely tufted, semi-erect, stoloniferous perennial with short rhizomes and with stems; stolons slender but strong and of reddish colour. Leaf linear to lanceolate, glabrous and with strongly denticulate margins. Inflorescence consisting of 3-8 racemes on an axis, bearing spikelets in two rows; spikelets elliptic.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Disakisperma yemenicum*	97.90%	KF574403
BOLD	*Enteropogon macrostachyus*	96.30%	UHURU634-14|Enteropogon macrostachyus|ITS|KR733941
***trn*L BLAST**			
GenBank®	*Cynodon dactylon*	99.40%	KY024482
	Astrebla pectinata	99.40%	KT168391
BOLD	*Enteropogon macrostachyus*	99.30%	UHURU632-14|Enteropogon macrostachyus|trnL-F|KR737769

***ITS*1 phylogeny**	***trn*L phylogeny**
Disakisperma yemenicum	*Disakisperma dubium*

Sample ID: Grass AB; Herbarium ID: *Dactyloctenium aegyptium* (L.) Willd

- A slightly stoloniferous and tufted short-lived perennial or annual grass, consisting of many branches. The stems are slender, ascending and geniculate or erect. The stolons root from the lower nodes and may creep. The roots are horizontal, while the leaves are broadly linear. The inflorescences are borne at the apex of the stem and are characteristically digitate or sub-digitate and arranged in two to six single, horizontal spikes.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Dactyloctenium aegyptium*	100.00%	HM347007
BOLD	Unidentifiable	<95.00%	N/A
***trn*L BLAST**			
GenBank®	*Dactyloctenium aegyptium*	99.40%	KY432807
BOLD	*Dactyloctenium aegyptium*	99.80%	UHURU361-14|Dactyloctenium NC_037161

***ITS*1 phylogeny**	***trn*L phylogeny**
Dactyloctenium aegyptium	*Dactyloctenium aegyptium*

Sample ID: Grass AC; Herbarium ID: *Sporobolus stapfianus* Gand

- A densely caespitose perennial, the basal leaf-sheaths forming a compacted mass of fine fibres with age. Leaf-blades convolute, sheath-margins tomentose with curly hairs. The branches are capillary and tinged with red. Spikelets are greyish green or sometimes dark green.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Sporobolus marginatus*	97.90%	KM010441
	Sporobolus coromandelianus	97.90%	KM010403
	Sporobolus cordofanus	97.90%	KM010400
BOLD	Unidentifiable	<95.00%	N/A
***trn*L BLAST**			
GenBank®	*Sporobolus pyramidatus*	99.00%	EF156733
BOLD	*Sporobolus africanus*	97.80%	UHURU1035-14|Sporobolus africanus|trnL-F|KR737946

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Sporobolus spp.	*Sporobolus pyramidatus*

NB. - *Sporobolus marginatus* and *S. pyramidatus* were reclassified to the subgenus *Sporobolus* subsect. *Pyramidati* P.M.Peterson (Peterson, Romaschenko, Arrieta, *et al.*, 2014).

**Sporobolus stapfianus* and *S. africanus* were reclassified to the subgenus *Sporobolus* subsect. *Sporobolus* (Peterson, Romaschenko, Arrieta, *et al.*, 2014).

Sample ID: Grass AD; Herbarium ID: *Cenchrus ciliaris* L.

- This is a perennial grass with linear leaves and flowers are produced in a panicle. The inflorescence is a bristly false spike, straw- or purple-coloured; all bristles are joined at base below spikelet cluster to form a small inconspicuous disc.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Chloris mossambicensis*	98.90%	HM347036
BOLD	*Chloris nutans*	95.60%	UHURU1317-15|Chloris nutans|ITS|KR734326
***trn*L BLAST**			
GenBank®	*Cynodon dactylon*	99.10%	MG709452
	Astrebla pectinata	99.10%	KT168391
BOLD	*Chloris roxburghiana*	99.10%	UHURU516-14|Chloris roxburghiana|trnL-F
	Enteropogon macrostachyus	99.10%	UHURU632-14|Enteropogon macrostachyus|trnL-F|KR737769

***ITS*1 phylogeny**	***trn*L phylogeny**
Tetrapogon cenchriformis	Unidentifiable

***** Unidentifiable – the identity did not meet the criteria stated in the analyses.

NB. *Tetrapogon* spp. are also known as *Chloris* spp. (Quattrocchi, 2006; Petersonn *et al.*, 2015).

Sample ID: Grass AE**;** Herbarium ID: *Chloris virgata* Swartz.

- This is an annual grass, culms tufted, erect or geniculately ascending and slightly flattened. Basal leaf sheaths strongly keeled, glabrous; leaf blades flat or folded, glabrous, adaxial surface scabrous, apex acuminate. Racemes digitate, silky, pale brown or tinged pink or purple; rachis scabrous or hispid. Spikelets with 2 or 3 florets.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Chloris gayana*	99.70%	KR734190
	Chloris virgata	99.70%	KR734150
BOLD	*Chloris virgata*	99.70%	UHURU655-14|Chloris virgata|ITS|KR733910
***trn*L BLAST**			
GenBank®	*Chloris virgata*	100.00%	KX765279
BOLD	*Chloris virgata*	99.90%	UHURU648-14|Chloris virgata|trnL-F|KR738355

***ITS*1 phylogeny**	***trn*L phylogeny**
Chloris spp.	*Chloris* spp.

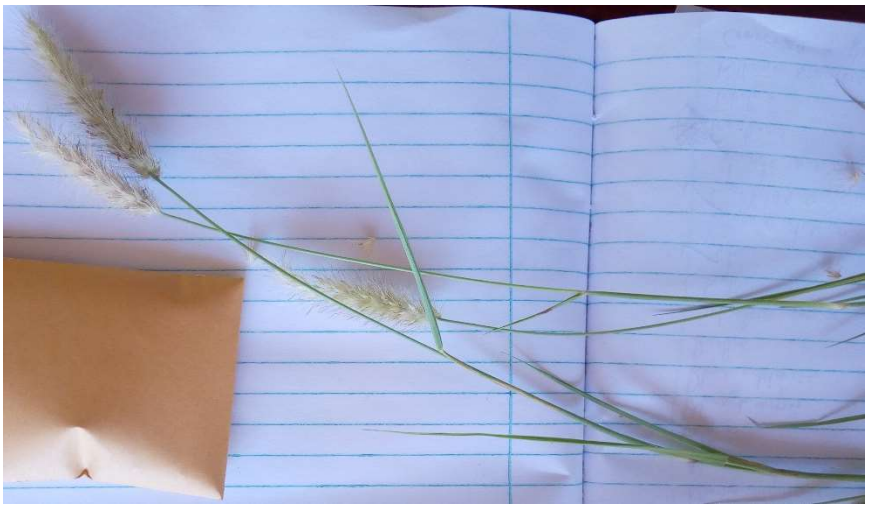
Sample ID: Grass AF ; Herbarium ID: *Sporobolus quadratus* W. D. Clayton

- This is a tufted perennial, the basal sheaths fairly broad, usually papery to subcoriaceous. Leaf-blades convolute; panicle spiciform; primary branches appressed to the main axis, densely spiculate. Spikelets are grey-green.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Sporobolus spicatus*	97.70%	KM010496
BOLD	Unidentifiable	<95.00%	N/A
***trn*L BLAST**			
GenBank®	*Sporobolus cryptandrus*	99.00%	MG709435
BOLD	*Sporobolus africanus*	98.10%	UHURU1035-14|Sporobolus africanus|trnL-F|KR737946

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Sporobolus cryptandrus	*Sporobolus* spp.



Sample ID: Grass AG; Herbarium ID: *Cenchrus ciliaris* L.

- This is a perennial grass with linear leaves and flowers are produced in a panicle. The inflorescence is a bristly false spike, straw- or purple-coloured; all bristles are joined at base below spikelet cluster to form a small inconspicuous disc.

***ITS*1 BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	*Cenchrus mezianus*	96.80%	KR733720
BOLD	*Pennisetum mezianum*	96.80%	UHURU603-14|Pennisetum mezianum|ITS|KR734250

***ITS*1 phylogeny**
Cenchrus ciliaris

NB. - The sequencing of the *trn*L region was unsuccessful.

- *Cenchrus* spp. are also known as *Pennisetum* spp. (Quattrocchi, 2006; Chemisquy *et al.*, 2010).

Sample ID: Grass AH; Herbarium ID: *Enneapogon cenchroides* (Roem. & Schult.) C.E. Hubb.

- An annual grass; basal sheaths remaining intact. Panicle loosely contracted, often lobed at the base. Hairy on the back all over; awns ciliate.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Enneapogon cenchroides*	100.00%	KR734310
BOLD	*Enneapogon cenchroides*	100.00%	UHURU1373-15|Enneapogon cenchroides|ITS|KR733746
***trn*L BLAST**			
GenBank®	*Enteropogon scoparius*	99.50%	DQ655895
BOLD	*Enneapogon persicus*	99.10%	UHURU1251-15|Enneapogon persicus|trnL-F|KR738619
	Enneapogon cenchroides	99.10%	UHURU908-14|Enneapogon cenchroides|trnL-F|KR738224

***ITS*1 phylogeny**	***trn*L phylogeny**
Enteropogon scoparius	*Enteropogon scoparius*

Sample ID: Grass AI; Herbarium ID: *Cynodon dactylon* (L.) Pers**.**

- The blades are a grey-green colour with rough edges. The stems are slightly flattened, often tinged purple in colour. The seed heads are produced in a cluster of two to six spikes together at the top of the stem. It has a deep root system

***ITS*1 BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	*Odyssea paucinervis*	97.90%	KX582389
BOLD	Unidentifiable	<95.00%	N/A

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**
Odyssea paucinervis

NB. The sequencing of the *trn*L region was unsuccessful.

Sample ID: Grass AJ; Herbarium ID: *Sporobolus ioclados* (Trin.) Nees.

- This is a tussocky perennial, often with creeping stolons; leaves flat or rolled, harsh or soft, often pungent; basal sheaths persistent, chartaceous, often keeled and flabellate. Panicle narrowly ovate to pyramidal; primary branches in whorls.

***ITS*1 BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	*Sporobolus marginatus*	97.90%	KM010441
BOLD	Unidentifiable	<95.00%	N/A

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**
Sporobolus spp.

NB. - The sequencing of the *trn*L region was unsuccessful.

- *Sporobolus ioclados, S. marginatus*, *S. cordofanus* and *S. coromandelianus* were reclassified to the subgenus *Sporobolus* subsect. *Pyramidati* P.M.Peterson (Peterson, Romaschenko, Arrieta, *et al.*, 2014).

Sample ID: Grass AK; Herbarium ID: *Cyperus kilimandscharicus* Kük.

- Annuals or perennials, the culms simple, usually triangular and leafy; inflorescence involucrate in dense spikes or in clusters, capitate or on rays which are often compound; spikelets flat or subterete; flowers perfect.

***ITS*1 BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	*Cyperus aggregatus*	97.70%	KF193566
BOLD	*Cyperus rubicundus*	95.50%	UHURU1042-14|Cyperus rubicundus|ITS|KR734104

***ITS*1 phylogeny**
Cyperus spp.

NB. The amplification of the *trn*L region was unsuccessful.

Sample ID: Grass AL; Herbarium ID: *Kyllinga comosipes* (Mattf. & Kük.) Napper

- A herbaceous, perennial plant with culms of grass-like leaves frowing from a long, slender rhizome creeping horizontally under or close to the ground surface.

***ITS*1 BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	*Cyperus babakan*	96.70%	LS999525
BOLD	*Cyperus rubicundus*	96.10%	UHURU1042-14|Cyperus rubicundus|ITS|KR734104

***ITS*1 phylogeny**
Cyperus spp.

NB. The sequencing of the *trn*L region was unsuccessful.

Sample ID: Grass AM; Herbarium ID: *Cynodon aethiopicus* Clayton & Harlan.

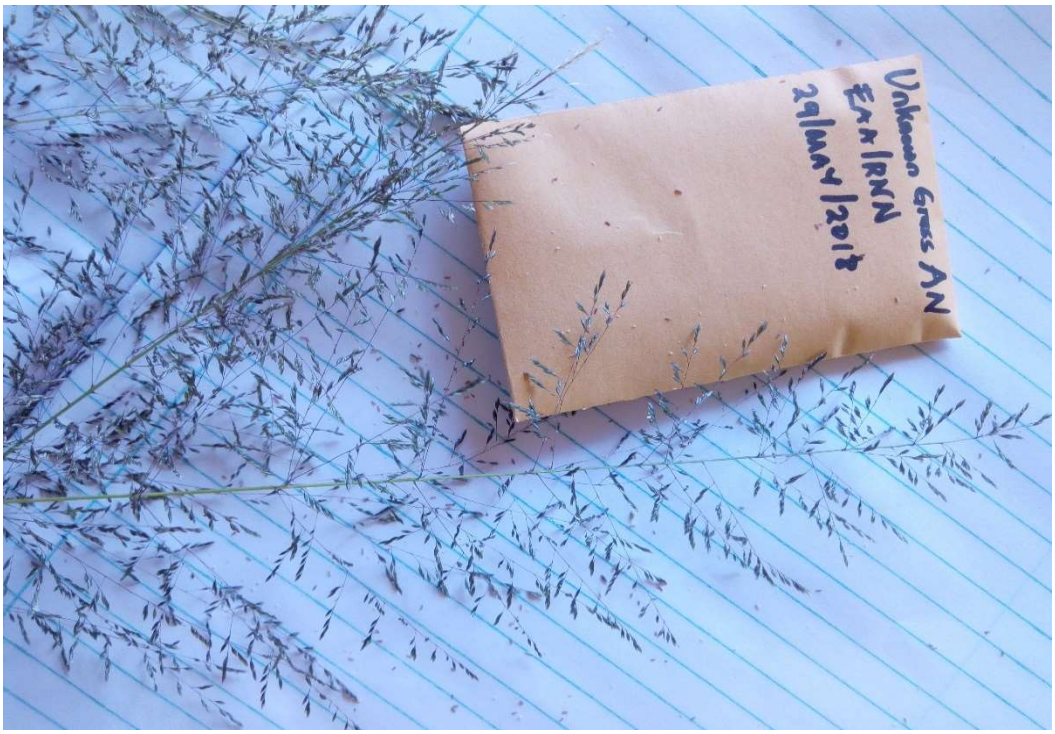
- Coarse stoloniferous perennial without rhizomes; stolons stout, lying flat on the ground; culms very robust, hard, shining and woody. Leaf-blades wide, stiff and harsh, glaucous, scaberulous, glabrous or with a few scattered hairs; ligule a scarious rim. Racemes in 2–5 whorls (rarely 1), stiff and spreading. Spikelets strongly pigmented with red or purple; glumes narrowly lanceolate in side view.

***trn*L BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	Unidentifiable	<95.00%	N/A
BOLD	Unidentifiable	<95.00%	N/A

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***trn*L phylogeny**
Unidentifiable

NB. The sequencing of the *ITS*1 region was unsuccessful.



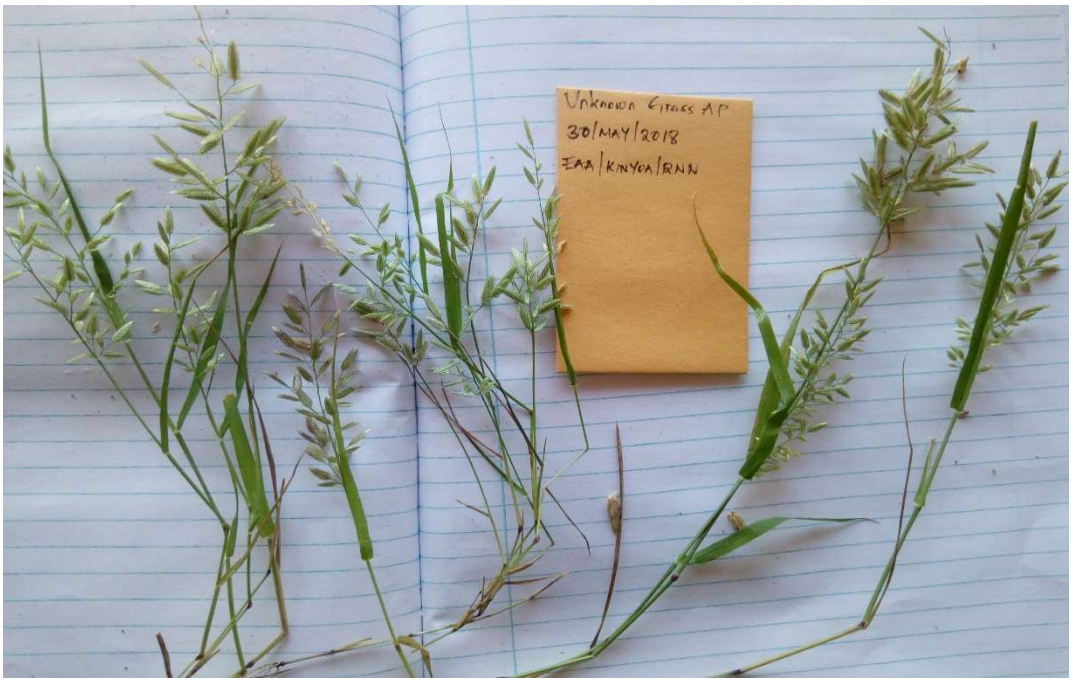
Sample ID: Grass AN; Herbarium ID: *Sporobolus ioclados* (Trin.) Nees.

- This is a tussocky perennial, often with creeping stolons; leaves flat or rolled, harsh or soft, often pungent; basal sheaths persistent, chartaceous, often keeled and flabellate. Panicle narrowly ovate to pyramidal; primary branches in whorls.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Sporobolus marginatus*	97.70%	KM010441
BOLD	Unidentifiable	<95.00%	N/A
***trn*L BLAST**			
GenBank®	*Sporobolus pyramidatus*	98.10%	EF156733
BOLD	*Sporobolus africanus*	96.90%	UHURU1035-14|Sporobolus africanus|trnL-F|KR737946

*Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Sporobolus sp.	*Sporobolus* sp.



Sample ID: Grass AP; Herbarium ID: *Eragrostis cilianensis* (All.) Lut.

- This is an annual bunchgrass forming tufts. The stems are generally erect but may droop or bend. The stems have glandular tissue near the nodes and the long leaves are often dotted with glands as well. The plants have a strong scent. The branching inflorescences have one to several spikelets per branch. Each spikelet is greenish brown, sometimes very slightly purple-tinted.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Eragrostis minor*	99.80%	KP057036
BOLD	*Eragrostis papposa*	98.10%	UHURU1170-14|Eragrostis papposa|ITS|KR734013
***trn*L BLAST**			
GenBank®	*Eragrostis minor*	99.90%	MG709400
BOLD	*Sporobolus rangei*	99.70%	UHURU1231-15|Sporobolus rangei|trnL-F|KR737763

***ITS*1 phylogeny**	***trn*L phylogeny**
Eragrostis spp.	*Eragrostis papposa*

Sample ID: Grass AR; Herbarium ID: *Cyperus teneriffae* Poir.

- An annual sedge with fibrous roots. Leaves few, weak, flat or conduplicate, gradually acuminate, smooth or scaberulous at the top. Inflorescence a single, hemispherical or subglobose head.

***ITS*1 BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	*Cyperus rubicundus*	99.10%	KR733903
BOLD	*Cyperus rubicundus*	99.10%	UHURU1042-14|Cyperus rubicundus|ITS|KR734104

***ITS*1 phylogeny**
Cyperus aggregatus

NB. The amplification of the *trn*L region was unsuccessful.

Appendix 2: Molecular and morphological identities of dicotyledons

(Source: E. A. Archie)

Sample ID: *Abutilon mauritianum* (Jacq.) Medik.

- A perennial herb or shrub up to 2m tall; simple leaves that are alternate; bisexual flowers that are solitary in leaf axil.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Abutilon sp.*	99.70%	KR733990
BOLD	*Abutilon hirtum*	99.70%	UHURU144-14|Abutilon hirtum|ITS|KR733990
***trn*L BLAST**			
GenBank®	*Abutilon mauritianum*	100.00%	KR738346
BOLD	*Abutilon mauritianum*	100.00%	UHURU950-14|Abutilon mauritianum|trnL-F|KR738346

***ITS*1 phylogeny**	***trn*L phylogeny**
Abutilon spp.	*Abutilon* spp.

Sample ID: *Acacia tortilis (*Forssk.) Hayne (aka *Vachellia tortilis*; Dyer, 2014)

-A small to medium-sized evergreen tree or shrub that grows up to 21 m tall; leaves are glabrous to densely pubescent; inflorescence has globose heads; pods are variable, indehiscent, spirally twisted or rarely almost straight.

***trn*L BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	*Vachellia tortilis*	100.00%	KY100266
BOLD	*Acacia tortilis*	100.00%	UHURU1134-14|Acacia tortilis|trnL-F|KR738113

***trn*L phylogeny**
Vachellia tortilis

NB. The sequencing of the *ITS*1 region was unsuccessful.

Sample ID: *Acacia xanthophloea* Benth. (aka *Vachellia xanthophloea*; Dyer, 2014)

- Medium-sized tree up to 25 m tall; bark is lemon yellow to greenish yellow; crown open, with spreading branches; alternate leaves; bisexual flowers; linear-oblong pod.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Acacia xanthophloea*	98.80%	JQ265831
BOLD	Unidentifiable	<95.00%	N/A
***trn*L BLAST**			
GenBank®	*Vachellia seyal*	100.00%	KY100267
	Vachellia nilotica	100.00%	KR737618
	Acacia nilotica	100.00%	AF522973
BOLD	*Acacia seyal*	100.00%	UHURU1362-15|Acacia seyal|trnL-F|KR738566
	Acacia nilotica	100.00%	UHURU1157-14|Acacia nilotica|trnL-F|KR737618

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Acacia spp.	*Acacia/Vachellia seyal; nilotica*

Sample ID: *Achyranthes aspera* Linn.

- An erect or spreading long-lived (perennial) herb which can grow up to 2 m tall. Its stems become woody at the base. It short-stalked leaves (dark green above and paler below) are opposite, simple and egg-shaped with broad end at base (ovate) up to 10 cm long by 8 cm wide, densely to sparsely hairy (pubescent) tapering to a point at both ends and shortly stalked. The small greenish-white flowers form narrow, elongated terminal spikes up to 60 cm long.

***ITS*1 BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	*Achyranthes aspera*	100.00%	LT992996
BOLD	*Achyranthes aspera*	97.80%	UHURU875-14|Achyranthes aspera|ITS|KR734237

***ITS*1 phylogeny**
Achyranthes splendens

NB. The amplification of the *trn*L region was unsuccessful.

Sample ID: *Asparagus asparagii*

Herbarium ID: *Asparagus setaceus* (Kunth) Jessop

- A scrambling perennial herb with tough green stems, which may reach several metres in length. The leaves are actually leaf-like cladodes, which arise in clumps of up to 15 from the stem, making a fine, soft green fern-like foliage.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Asparagus altissimus*	98.50%	HE602411
BOLD	*Asparagus falcatus*	97.30%	UHURU823-14|Asparagus falcatus|ITS|KR734049
***trn*L BLAST**			
GenBank®	*Asparagus globicus*	99.80%	LC309027
BOLD	*Asparagus falcatus*	99.70%	UHURU823-14|Asparagus falcatus|trnL-F|KR738201

***ITS*1 phylogeny**	***trn*L phylogeny**
Asparagus altissimus	*Asparagus* spp.

Sample ID: *Azima tetracantha* Lam.

- Dioecious, erect shrub up to 90 cm tall with spines in each leaf axil; leaves are decussately opposite, simple and entire; unisexual flowers.

***ITS*1 BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	Unidentifiable	<95.00%	N/A
BOLD	Unidentifiable	<95.00%	N/A

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**
Salvadora spp.

NB. The sequencing of the *trn*L region was unsuccessful.

Sample ID: *Balanites pedicellaris* Mildbr. & Schltr.

- A shrub or small tree that grows up to 6m tall; branches are yellowish or greyish-green, bearing simple green spines; leaves are alternate or grow on the spines; greenish-white flowers; the fruit is a drupe, which is round or ellipsoid and normally flattened on either end.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	Unidentifiable	<95.00%	N/A
BOLD	Unidentifiable	<95.00%	N/A
***trn*L BLAST**			
GenBank®	*Balanites glabra*	100.00%	MH990655
BOLD	*Balanites rotundifolia*	100.00%	UHURU357-14|Balanites rotundifolia|trnL-F

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Balanites glabra	*Balanites glabra*

Sample ID: *Balanites* sp.

- A shrub or small tree that reaches 10 m in height with a generally narrow form. The branches have long, straight green spines arranged in spirals. The dark green compound leaves grow out of the base of the spines and are made up of two leaflets which are variable in size and shape.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	Unidentifiable	<95.00%	N/A
BOLD	Unidentifiable	<95.00%	N/A
***trn*L BLAST**			
GenBank®	*Balanites aegypticus*	98.80%	KR738666
BOLD	*Balanites aegyptica*	98.80%	UHURU388-14|Balanites aegyptica|trnL-F|KR738639

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Balanites spp.	*Balanites* spp.

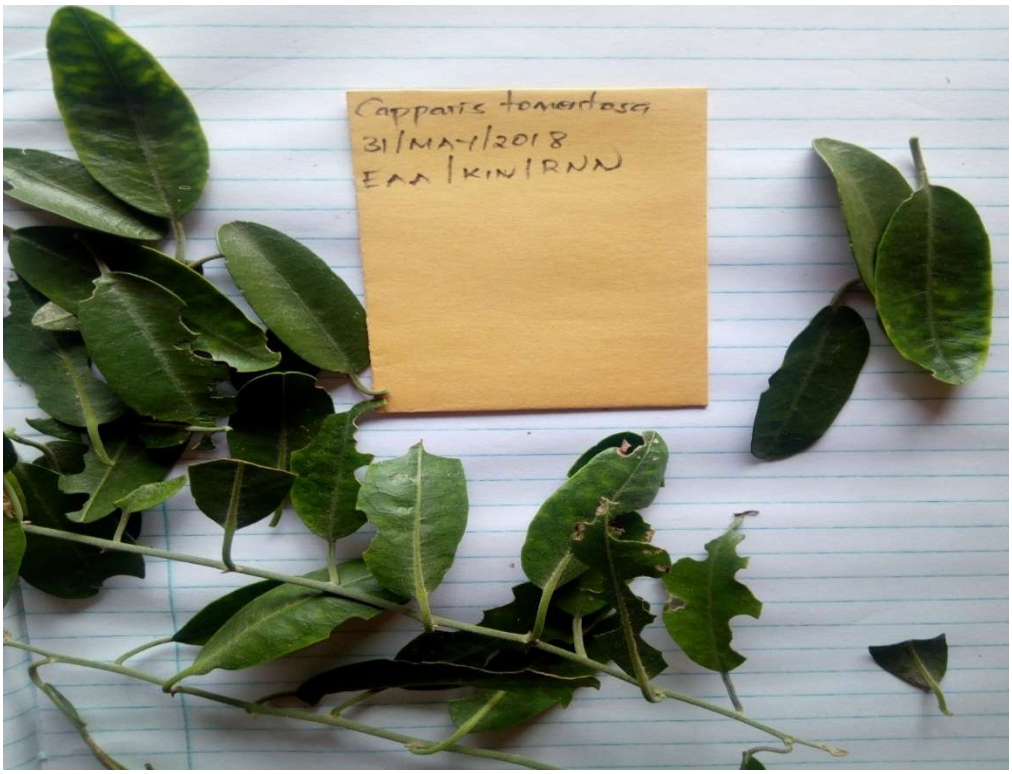
Sample ID: *Cadaba farinosa* Forssk.

- An evergreen shrub or small tree; simple ovate leaves with entire margins; whitish or pinkish flowers, and is covered in powdery hairs or scales.

***trn*L BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	*Maerua triphylla*	99.90%	KR738342
BOLD	*Maerua triphylla*	99.90%	UHURU382-14|Maerua triphylla|trnL-F|KR738342

***trn*L phylogeny**
Maerua triphylla

NB. The sequencing of the *ITS*1 region was unsuccessful.



Sample ID: *Capparis tomentosa* Lam.

- Mostly a robust woody climber; the stem has sharp, paired, hooked spines; leaves form between the spines; flowers form in clusters.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	Unidentifiable	<95.00%	N/A
BOLD	Unidentifiable	<95.00%	N/A
***trn*L BLAST**			
GenBank®	*Capparis tomentosa*	99.80%	KR738333
BOLD	*Capparis tomentosa*	99.80%	UHURU1047-14|Capparis tomentosa|trnL-F|KR738333

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Capparis sikkimensis	*Capparis tomentosa*

Sample ID: *Cassia italica* (Mill) Lam. (aka *Senna italica* (Mill.); Okeyo and Bosch, 2007)

- Perennial herb with several prostrate to decumbent, branched stems up to 40 cm high; leaves variable in size: flowers are actinomorphic.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Senna italica*	100.00%	KX057899
BOLD	Unidentifiable	<95.00%	N/A
***trn*L BLAST**			
GenBank®	*Senna wislizeni*	98.60%	AF365028
BOLD	*Chamaecrista grantii*	97.20%	UHURU811-14|Chamaecrista grantii|trnL-F|KR737998

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Senna italica	*Senna italica*

Sample ID: *Commicarpus plumbagineus* (Cav.) Standl.

- A herb with long branched stems, up to several metres, growing from a woody root-stock. Stems may be woody near the base; leaves are ovate; inflorescences in irregular umbels of white trumpet-shaped flowers with long exerted stamens.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Commicarpus pedunculosus*	98.90%	KR734330
BOLD	*Commicarpus pedunculosus*	98.90%	UHURU1283-15|Commicarpus pedunculosus|ITS|KR734034
***trn*L BLAST**			
GenBank®	*Acleisanthes obtusa*	98.80%	MH286321
BOLD	Unidentifiable	<95.00%	N/A

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Commicarpus plumbagineus	*Commicarpus pedunculosus*

Sample ID: *Cordia monoica* (Roxb.)

- A multi-stemmed shrub or tree to 6 m; leaves are broadly oval to almost round; yellow flowers in dense clusters.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Cordia monoica*	99.30%	MK261116
BOLD	*Cordia monoica*	99.30%	UHURU250-14|Cordia monoica|ITS
***trn*L BLAST**			
GenBank®	*Cordia sagotii*	98.30%	FJ039222
BOLD	*Heliotropium steudneri*	96.80%	UHURU785-14|Heliotropium steudneri|trnL-F|KR737947

***ITS*1 phylogeny**	***trn*L phylogeny**
Cordia monoica	*Cordia monoica*

Sample ID: *Dasyphaera prostrata* (Volk. ex Gilg) Cavaco (aka *Volkensinia prostrata* (Volkens ex Gilg) Schinz; Altmann, 1998)

- Much-branched perennial herb or subshrub with a tough fibrous rootstock and brittle branches. Leaves ovate or broadly ovate to narrowly or oblong-lanceolate. Partial inflorescences dense; peduncles slender. Bracts deltoid-ovate, membranous with a narrow green very shortly excurrent midrib. Bracteoles of ultimate triads similar to the bracts or rather smaller. Flowers bright carmine-red.

***ITS*1 BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	*Volkensinia prostrata*	98.90%	LT995183
BOLD	Unidentifiable	<95.00%	N/A

***** Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**
Volkensinia prostrata

NB. The amplification of the *trn*L region was unsuccessful.

Sample ID: *Euclea schimperi* (A.DC.) Dandy.

- Leaves thinly leathery, obovate or oblanceolate, up to 4 times as long as broad, distinctly broadest in upper half. Ovary covered with stiff hairs.

***trn*L BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	*Euclea divinorum*	100.00%	KR738166
BOLD	*Euclea divinorum*	100.00%	UHURU777-14|Euclea divinorum|trnL-F|KR738166

***trn*L phylogeny**
Euclea spp.

NB. The sequencing of the *ITS*1 region was unsuccessful.

Sample ID: *Ficus* sp

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Sideroxylon inerme*	95.90%	AM408078
BOLD	Unidentifiable	<95.00%	N/A
***trn*L BLAST**			
GenBank®	*Faucherea thouvenotii*	99.60%	KC479316
	Manikara zapota	99.60%	AJ430885
BOLD	*Euclea divinorum*	97.90%	UHURU777-14|Euclea divinorum|trnL-F|KR738166

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Sideroxylon spp.	Ambiguous

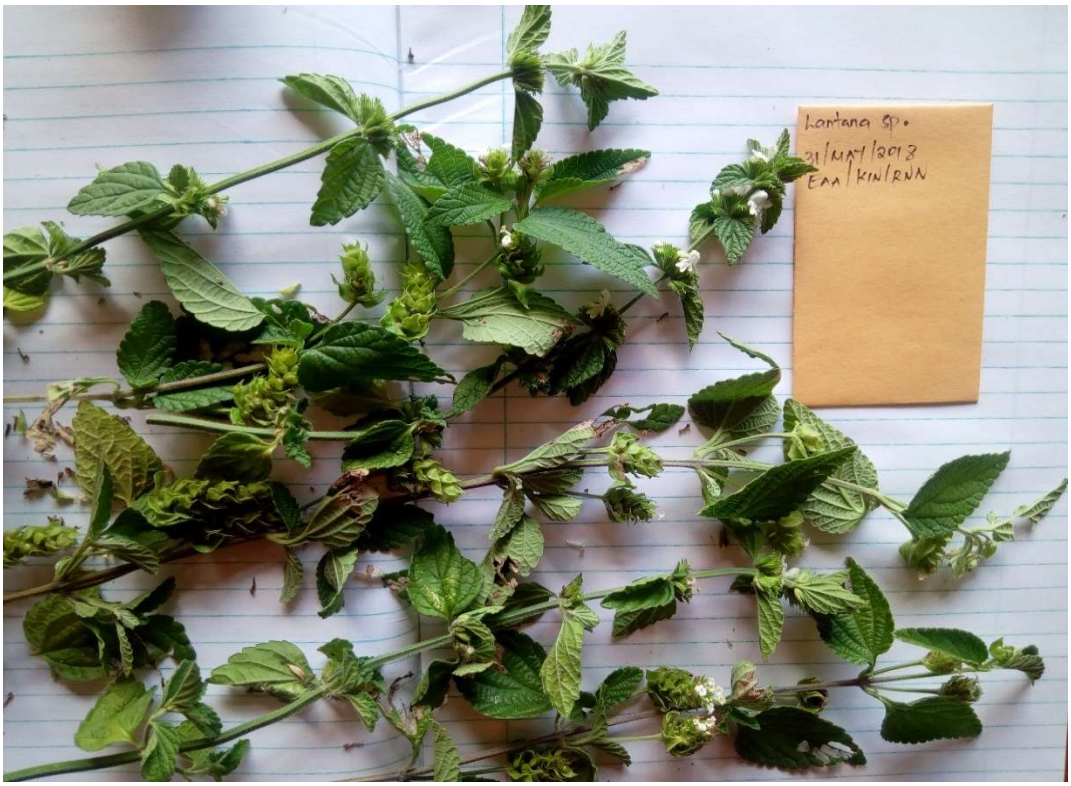
*Ambiguous – multiple genera were identified.

Sample ID: *Hibiscus* '*lila*'

- A slender shrub growing up to 2.5 metres tall.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Hibiscus micranthus*	99.90%	KR734203
BOLD	*Hibiscus micranthus*	99.90%	UHURU738-14|Hibiscus micranthus|ITS|KR734203
***trn*L BLAST**			
GenBank®	*Hibiscus ovalifolius*	99.50%	MK261741
	Hibiscus micranthus	99.50%	KR737979
BOLD	*Hibiscus ovalifolius*	99.50%	UHURU509-14|Hibiscus ovalifolius|trnL-F
	Hibiscus micranthus	99.50%	UHURU737-14|Hibiscus micranthus|trnL-F|KR737979

***ITS*1 phylogeny**	***trn*L phylogeny**
Hibiscus micranthus	*Hibiscus micranthus*



Sample ID: *Lantana camara* L.

- A small perennial shrub which can grow to around 2 m tall. has small tubular shaped flowers, which each have four petals and are arranged in clusters in terminal areas stems. Flowers come in many different colours, including red, yellow, white, pink and orange, which differ depending on location in inflorescences, age, and maturity.

***trn*L BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	*Lippia javanica*	100.00%	KR738559
	Lantana rugosa	100.00%	HM216637
BOLD	*Lantana viburnoides*	100.00%	UHURU068-14|Lantana viburnoides|trnL-F|KR738559

***trn*L phylogeny**
Lantana rugosa

NB. The sequencing of the *ITS*1 region was unsuccessful.

Sample ID: *Lycium europaeum* L.

- A shrub, often thorny, growing 1 to 4 meters tall. The leaves are small, narrow, and fleshy, and are alternately arranged, sometimes in fascicles. Flowers are solitary or borne in clusters. The funnel-shaped or bell-shaped corolla is white, green, or purple in colour.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Lycium texanum*	96.90%	FJ439761
BOLD	Unidentifiable	<95.00%	N/A
***trn*L BLAST**			
GenBank®	*Lycium* sp.	100.00%	KU323942
BOLD	*Lycium europaeum*	100.00%	UHURU763-14|Lycium europaeum|trnL-F|KR737902

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Lycium sp.	*Lycium europaeum*

Sample ID: *Maerua angolensis* DC.

- The tree has a rounded crown and smooth grey bark flaking to reveal yellowish-orange patches. The twigs and branches display prominent lenticels. Leaves are soft and drooping, with petioles equal to the leaves in length, and visibly thicker or inspissate at their extremities. Leaves are alternate and broadly elliptic to ovate, with rounded or notched apex and a terminal bristle (mucronate).

***ITS*1 BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	Unidentifiable	<95.00%	N/A
BOLD	Unidentifiable	<95.00%	N/A

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**
Unidentifiable

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

NB. The sequencing of the *trn*L region was unsuccessful.

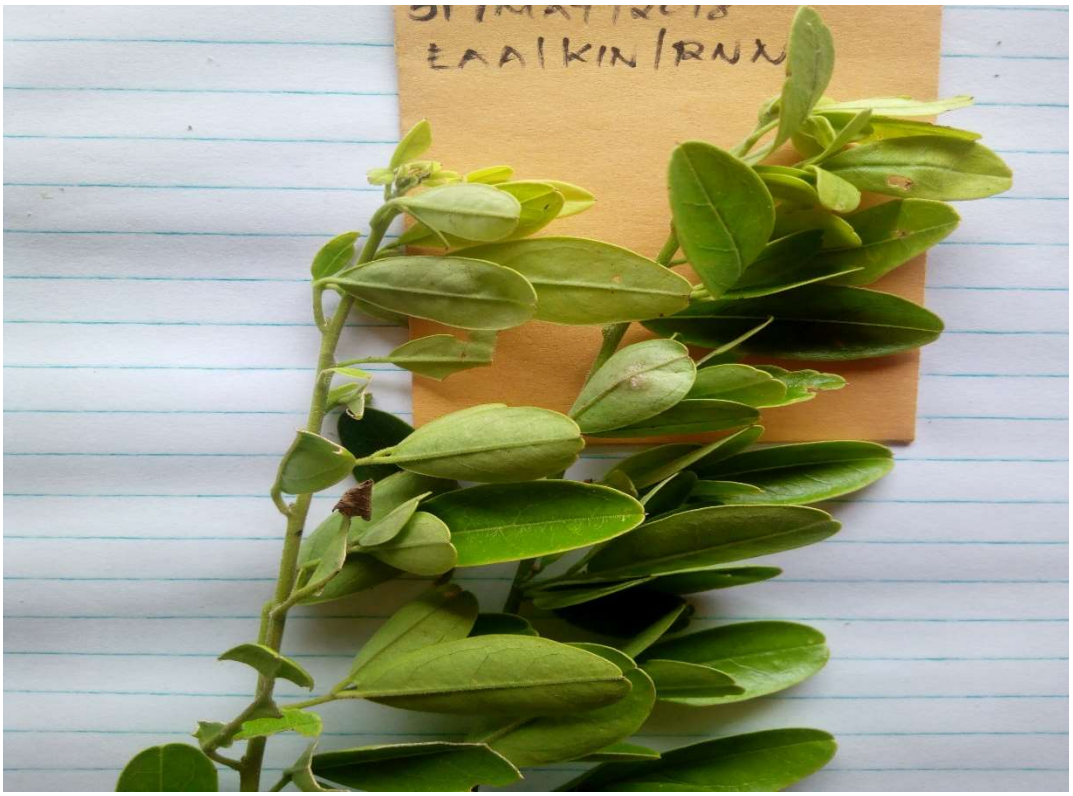
Sample ID: *Maerua crassifolia* Forssk.

- A small tree to 10 m high.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	Unidentifiable	<95.00%	N/A
BOLD	Unidentifiable	<95.00%	N/A
***trn*L BLAST**			
GenBank®	*Maerua triphylla*	100.00%	KR738342
BOLD	*Maerua triphylla*	100.00%	UHURU382-14|Maerua triphylla|trnL-F|KR738342

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Unidentifiable	*Maerua triphylla*

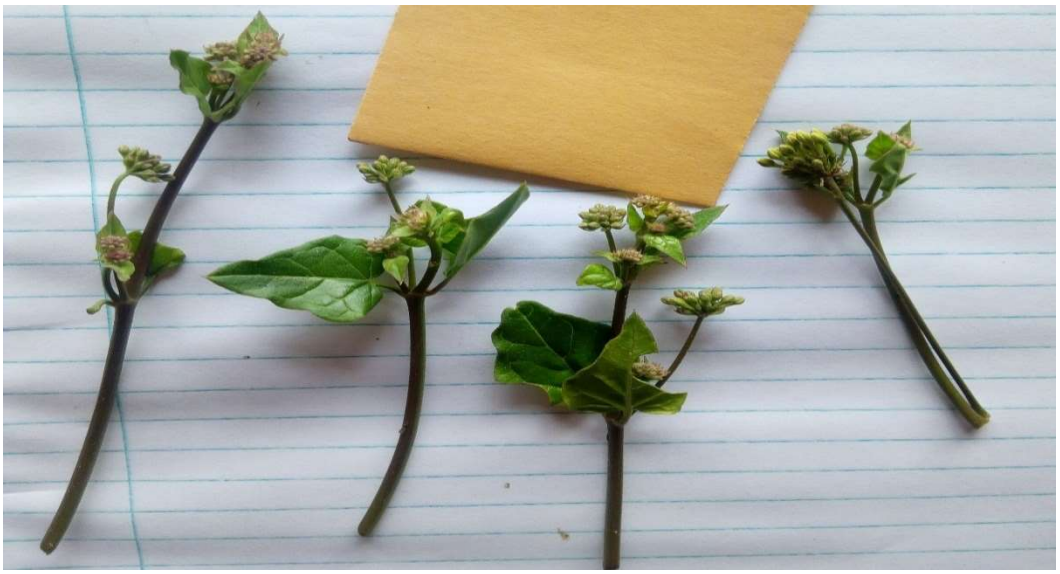


Sample ID: *Maerua* sp.

***trn*L BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	*Maerua triphylla*	100.00%	KR738342
BOLD	*Maerua triphylla*	100.00%	UHURU382-14|Maerua triphylla|trnL-F|KR738342

***trn*L phylogeny**
Maerua triphylla

NB. The sequencing of the *ITS*1 region was unsuccessful.



Sample ID: *Rhamphicarpa montana* N.E. Br. (aka *Cycnium tubulosum* (L.f.) Engl.; Staner, 1938)

- A hairless or nearly hairless, hemiparasitic perennial herbaceous plant, with angular stems having four flat sides, which are creeping, straggling or upright, that may have side branches or not, and sometimes there are a few glandular hairs. These stems carry few distanced leaves arranged oppositely or nearly so, which are approximately linear, widest at midlength, with a pointed cartilaginous tip, with a very short leaf stalk.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Cycnium tubulosum*	97.20%	KC480329
BOLD	Unidentifiable	<95.00%	N/A
***trn*L BLAST**			
GenBank®	*Cycnium tubulosum*	99.60%	EU264192
BOLD	Unidentifiable	<95.00%	N/A

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Cycnium tubulosum	*Cycnium tubulosum*

Sample ID: *Rhus natalensis* Bernh.

- A shrub up to 3m high.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Searsia leptodictya*	99.30%	AY641515
BOLD	Unidentifiable	<95.00%	N/A
***trn*L BLAST**			
GenBank®	*Searsia leptodictya*	99.60%	AY640466
BOLD	Unidentifiable	<95.00%	N/A

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Searsia tenuinervis	*Searsia leptodictya*

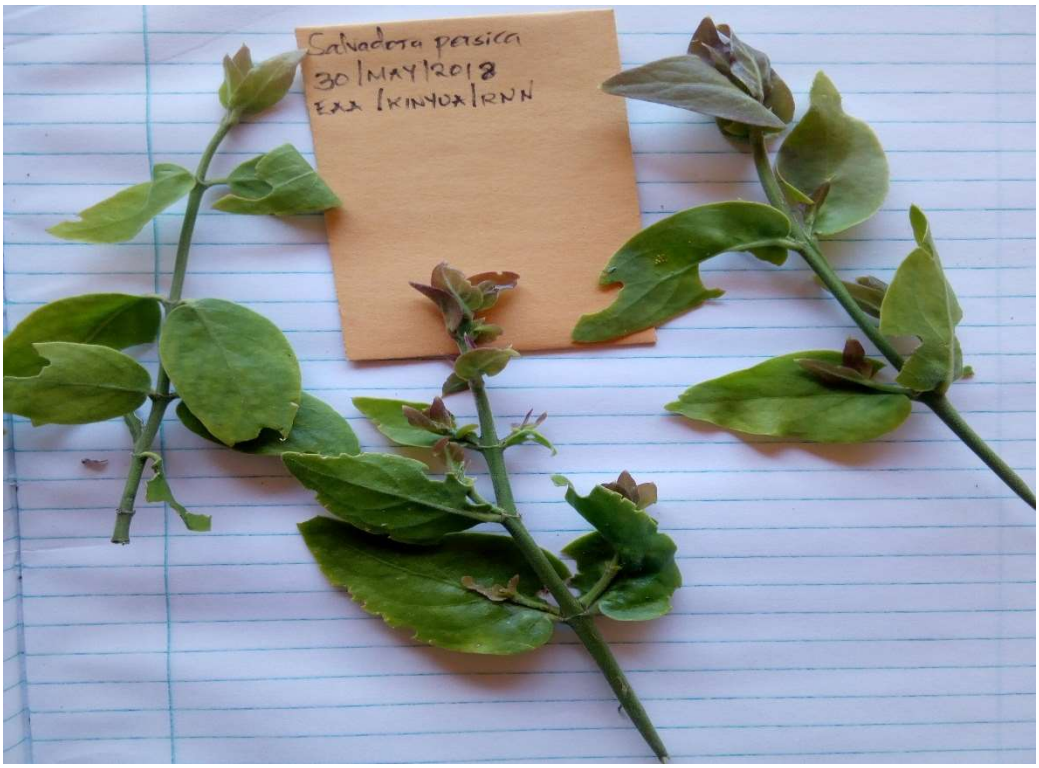
Sample ID: *Ruellia patula* Jacq.

- A perennial groundcover that produces annual, spreading shoots, from a woody rootstock.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	Unidentifiable	<95.00%	N/A
BOLD	Unidentifiable	<95.00%	N/A
***trn*L BLAST**			
GenBank®	*Justicia odora*	100.00%	KR737658
BOLD	*Justicia odora*	100.00%	UHURU996-14|Justicia odora|trnL-F|KR738537

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Justicia odora	*Justicia odora*



Sample ID: *Salvadora persica* L. (Meswak)

- A small tree or shrub with a crooked trunk, typically 6–7 metres in height. Its bark is scabrous and cracked, whitish with pendulous extremities. The root bark of the tree is similar in colour to sand, and the inner surfaces are an even lighter shade of brown.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	Unidentifiable	<95.00%	N/A
BOLD	Unidentifiable	<95.00%	N/A
***trn*L BLAST**			
GenBank®	*Salvadora angustifolia*	99.90%	KC479309
BOLD	Unidentifiable	<95.00%	N/A

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Salvadora oleiodes	*Salvadora* spp.

Sample ID: *Scutia myrtina* (Burm. f.) Kurz.

- A shrub or tree of 2-10 m tall with trunk diameter to 30 cm or often a scandent liane, climbing by means of thorns.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Scutia myrtina*	99.90%	KR734188
BOLD	*Scutia myrtina*	99.90%	UHURU1370-15|Scutia myrtina|ITS|KR733767
***trn*L BLAST**			
GenBank®	*Salvadora angustifolia*	97.00%	KC479309
BOLD	Unidentifiable	<95.00%	N/A

***** Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Scutia myrtina	Unidentifiable

Sample ID: *Solanum dubium* Fresen. (aka *Solanum coagulans* Forssk.; Altmann, 1998)

- A shrub with dense stellate tomentum on the branches, petioles, underside of leaves, and outside of calyx and corolla; branches terete; spines few or many; leaves ovate or ovate-elliptic; flowers solitary or few together.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Solanum coagulans*	99.40%	KR733860
BOLD	*Solanum coagulans*	99.40%	UHURU1120-14|Solanum coagulans|ITS|KR733794
***trn*L BLAST**			
GenBank®	*Solanum coagulans*	99.90%	KR737876
	Solanum pubescens	99.80%	KU719788
	Solanum sp.	99.80%	KR738609
BOLD	*Solanum coagulans*	99.90%	UHURU1120-14|Solanum coagulans|trnL-F|KR737759

***ITS*1 phylogeny**	***trn*L phylogeny**
Solanum sp.	*Solanum coagulans*

Sample ID: *Solanum incanum* L.

- A herb or soft wooded shrub up to 1.8 m in height with spines on the stem, stalks and calyces and with velvet hairs on the leaves. The flowers are often borne in the leaf axils, sometimes solitary or in clusters of a few flowers.

***trn*L BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	*Solanum incanum*	100.00%	MH283721
	Solanum rigidum	100.00%	MH283706
	Solanum sp.	100.00%	HQ721920
	Solanum campylacanthum	100.00%	HQ721908
	Solanum panduriforme	100.00%	EU176143
BOLD	*Solanum campylacanthum*	99.90%	UHURU123-14|Solanum campylacanthum|trnL-F|KR738245

***trn*L phylogeny**
Solanum sp.

NB. The sequencing of the *ITS*1 region was unsuccessful.

Sample ID: *Solanum nigrum* L.

- A short-lived perennial shrub or herbaceous plant. Its leaves are ovate to heart-shaped, with wavy or large-toothed edges; both surfaces hairy or hairless. The flowers have petals greenish to whitish, recurved when aged and surround prominent bright yellow anthers.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Solanum physalifolium*	100.00%	KY968826
	Solanum nigrum	100.00%	KR734145
	Solanum villosum	100.00%	KC540791
BOLD	*Solanum nigrum*	100.00%	UHURU266-14|Solanum nigrum|ITS|KR734020
***trn*L BLAST**			
GenBank®	*Solanum villosum*	100.00%	KT820839
	Solanum nigrum	100.00%	KT820820
BOLD	*Solanum nigrum*	100.00%	UHURU731-14|Solanum nigrum|trnL-F

***ITS*1 phylogeny**	***trn*L phylogeny**
Solanum nigrum	*Solanum nigrum*

Sample ID: *Suaeda monoica* Forssk.

- A shrub with leaves linear to linear–oblong. Flowers unisexual, usually quite numerous, sometimes contiguous into dense spikes.

***ITS*1 BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	*Suaeda fruticosa*	96.30%	KF848716
BOLD	Unidentifiable	<95.00%	N/A

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**
Suaeda monoica

NB. The sequencing of the *trn*L region was unsuccessful.

Sample ID: *Trianthema ceratosepalum* Volkens & Irmsch.

- Annuals or perennials generally characterized by fleshy, opposite, unequal, smooth-margined leaves, a prostrate growth form, flowers with five perianth segments subtended by a pair of bracts, and a fruit with a winged lid.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Trianthema parvifolium*	96.00%	KY657359
BOLD	*Zaleya petandra*	95.80%	UHURU1288-15|Zaleya pentandra|ITS|KR734226
***trn*L BLAST**			
GenBank®	Unidentifiable	<95.00%	N/A
BOLD	Unidentifiable	<95.00%	N/A

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Trianthema sp.	*Tetragonia schenckii*

No image available.

Sample ID: *Tribulus terrestris* L. (TT)

- A taprooted herbaceous plant. The flowers have five lemon-yellow petals, five sepals, and ten stamens.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Tribulus terrestris*	100.00%	KR734183
BOLD	*Tribulus terrestris*	100.00%	UHURU587-14|Tribulus terrestris|ITS|KR734173
***trn*L BLAST**			
GenBank®	*Tribulus terrestris*	100.00%	KR738417
BOLD	*Tribulus terrestris*	100.00%	UHURU585-14|Tribulus terrestris|trnL-F|KR738417

***ITS*1 phylogeny**	***trn*L phylogeny**
Tribulus terrestris	*Tribulus terrestris*

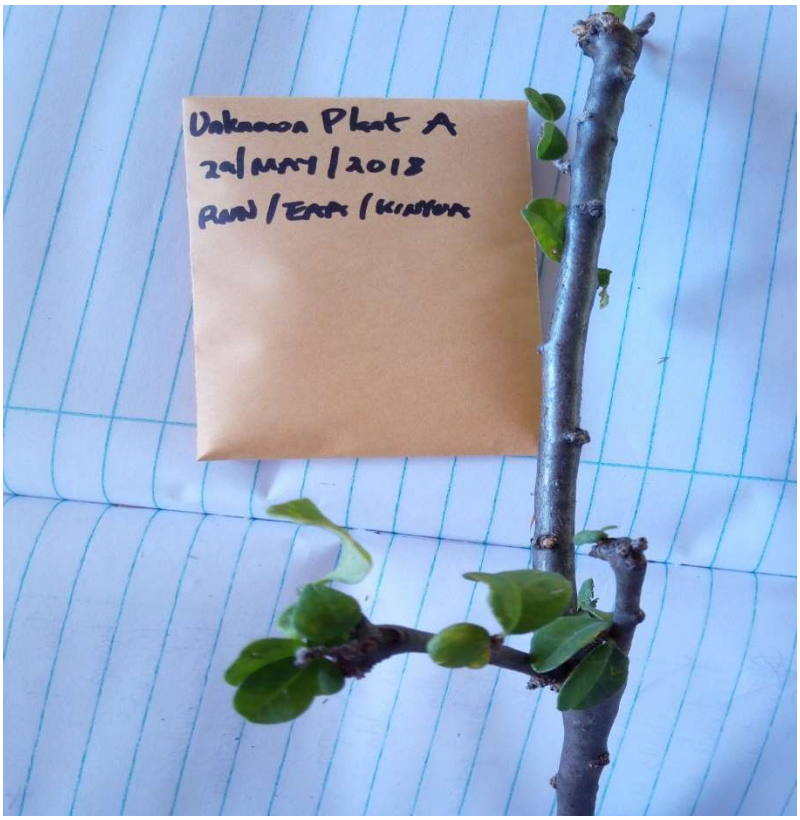
Sample ID: *Withania somnifera* (L.) Dunal

- A small shrub to 2 m high and to 1 m across. Almost the whole plant is covered with short, fine, silver-grey, branched hairs. The stems are brownish and prostrate to erect, sometimes leafless below.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Withania* sp.	99.40%	HM627273
	Withania somnifera	99.40%	HM627272
BOLD	Unidentifiable	<95.00%	N/A
***trn*L BLAST**			
GenBank®	*Withania somnifera*	100.00%	KR738304
BOLD	*Withania somnifera*	100.00%	UHURU390-14|Withania somnifera|trnL-F|KR738232

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Withania somnifera	*Withania somnifera*



Sample ID: Plant A

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	Unidentifiable	<95.00%	N/A
BOLD	Unidentifiable	<95.00%	N/A
***trn*L BLAST**			
GenBank®	*Maerua triphylla*	100.00%	KR738342
BOLD	*Maerua triphylla*	100.00%	UHURU382-14|Maerua triphylla|trnL-F|KR738342

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Euphorbia scatorhiza	*Maerua triphylla*

Sample ID: Plant B; Herbarium ID: *Boerhavia erecta* L.

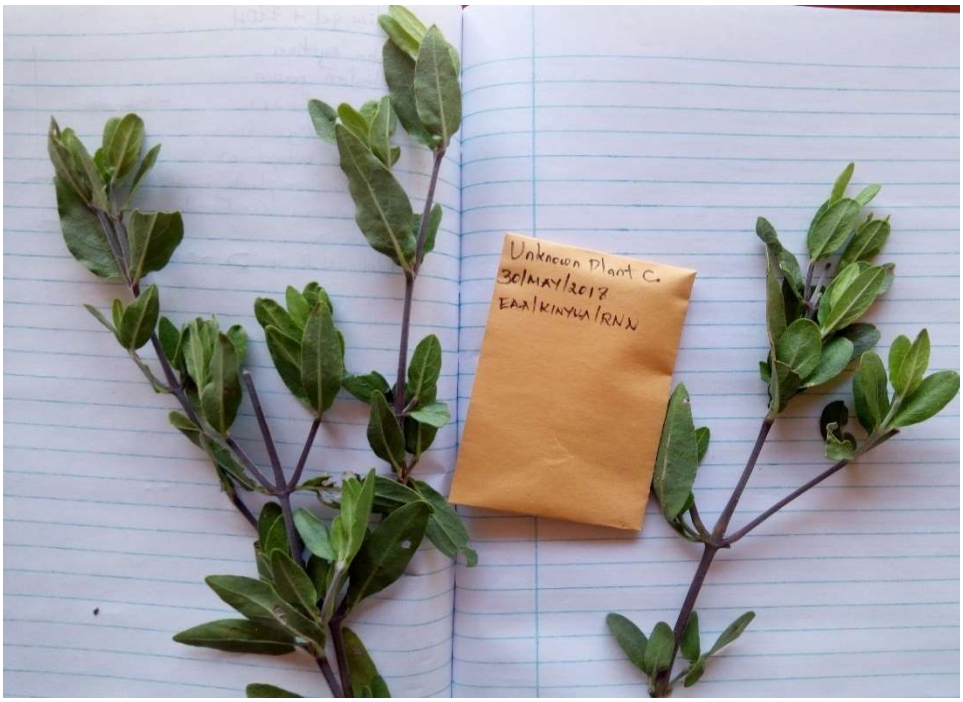
- Annual to short-lived perennial herb; stem branching mainly from the base, ascending to erect, fleshy, green, often flushed with red, lower parts thinly hairy, upper parts glabrous, nodes swollen. Leaves opposite, simple, about equal; stipules absent; blade broadly lanceolate to ovate. Inflorescence an axillary, small, often congested umbel.

***ITS*1 BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	*Boerhavia erecta*	100.00%	DQ317080
BOLD	Unidentifiable	<95.00%	N/A

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**
Boerhavia erecta

NB. The sequencing of the *trn*L region was unsuccessful.



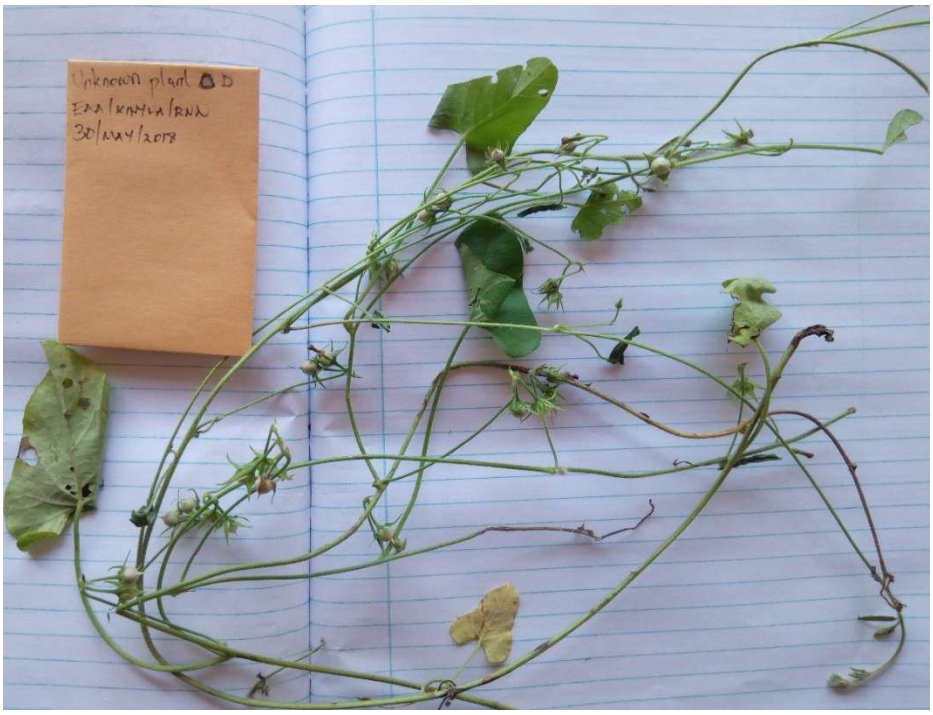
Sample ID: Plant C; Herbarium ID: *Barleria masaiensis* L. Darbysh.

- An erect, perennial, prickly shrub, usually single-stemmed. The leaves are ellipsoid. The base of the leaves is protected by three to five sharp spines. The yellow-orange tubular flowers with several long protruding stamens.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Megalochlamys revoluta*	97.10%	EU087473
BOLD	Unidentifiable	<95.00%	N/A
***trn*L BLAST**			
GenBank®	*Megalochlamys revoluta*	99.70%	EU087564
BOLD	*Justicia calyculata*	96.20%	UHURU563-14|Justicia calyculata|trnL-F|KR738575

***** Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Ambiguous	*Megalochlamys revoluta*



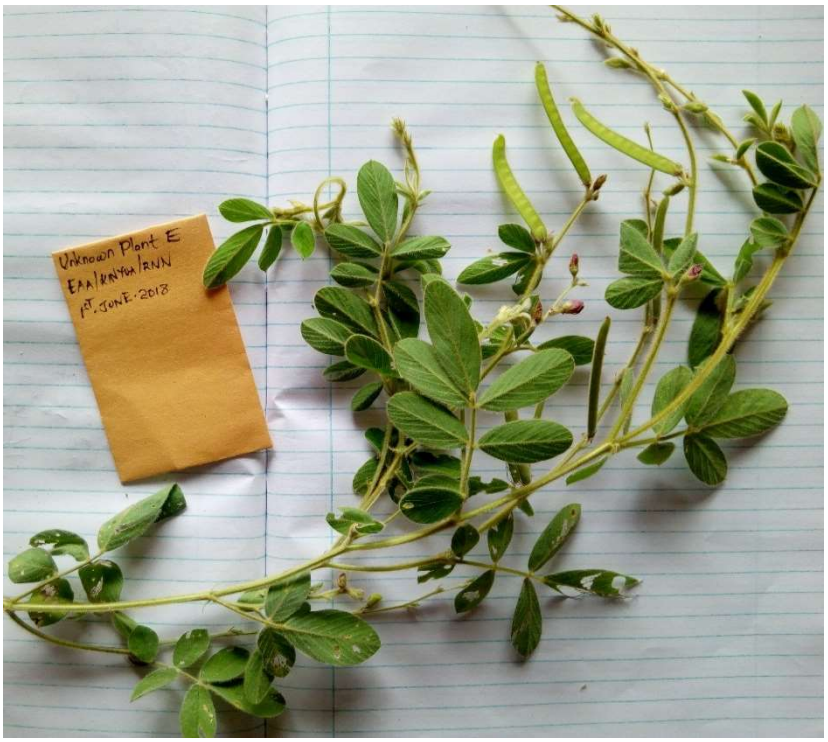
Sample ID: Plant D; Herbarium ID: *Ipomoea obscura* (L.) Ker Gawl.

- An annual or perennial herb with slender, twining or prostrate stems. Inflorescence a simple cyme or reduced to 1 or 2 flowers.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Ipomoea wrightii*	97.60%	KR734166
BOLD	*Ipomoea mombassana*	99.30%	UHURU640-14|Ipomoea mombassana|ITS
***trn*L BLAST**			
GenBank®	*Ipomoea sinensis*	100.00%	KR738586
BOLD	*Ipomoea mombassana*	100.00%	UHURU643-14|Ipomoea mombassana|trnL-F|KR738571

***** Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**	***trn*L phylogeny**
Ipomoea cairica	*Ipomoea* sp.



Sample ID: Plant E; Herbarium ID: *Tephrosia pumila* (Lamb.) Pers. va pumila.

- Annual or short-lived perennial; branches procumbent or straggling. Leaf-rhachis including a petiole; stipules narrowly triangular or subulate. Flowers white, pale pink or purplish in short terminal or leaf-opposed pseudoracemes and upper leaf-axils; bracts narrowly triangular, persistent.

	Description	**Grade**	**Accession no.**
***ITS*1 BLAST**			
GenBank®	*Tephrosia uniflora*	97.00%	KR046200
BOLD	*Tephrosia reptans*	96.40%	UHURU1142-14|Tephrosia reptans|ITS|KR734134
***trn*L BLAST**			
GenBank®	*Indigofera* sp.	98.60%	KR738170
BOLD	*Tephrosia emeroides*	98.60%	UHURU1355-15|Tephrosia emeroides|trnL-F|KR738170

***ITS*1 phylogeny**	***trn*L phylogeny**
Tephrosia obovata	*Tephrosia pedicellata*

Sample ID: Plant G

***ITS*1 BLAST**	**Description**	**Grade**	**Accession no.**
GenBank®	Unidentifiable	<95.00%	N/A
BOLD	Unidentifiable	<95.00%	N/A

* Unidentifiable – the identity did not meet the criteria stated in the analyses.

***ITS*1 phylogeny**
Euphorbia scatorhiza

NB. The sequencing of the *trn*L region was unsuccessful.

Appendix 3: Protocol for making the TAE Buffer

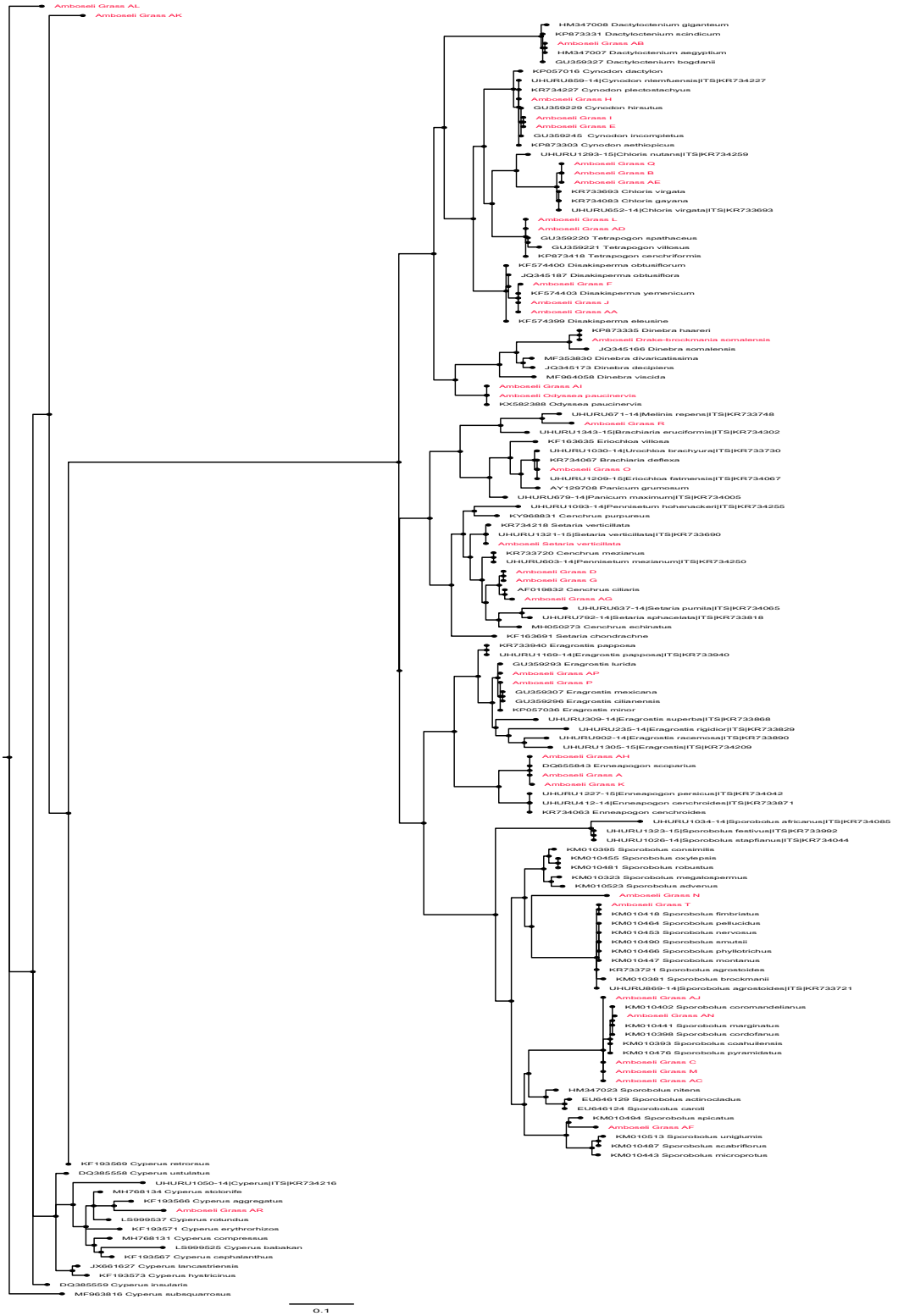
Stock solution: 10X TAE buffer recipe:

1) 900 mL distilled water
2) 48.4 g of Tris base [tris(hydroxymethyl)aminomethane]
3) 11.4 mL of glacial acetic acid
4) 3.72 g of EDTA disodium salt; dissolves when the pH is 8.
5) Adjust water to 1 litre

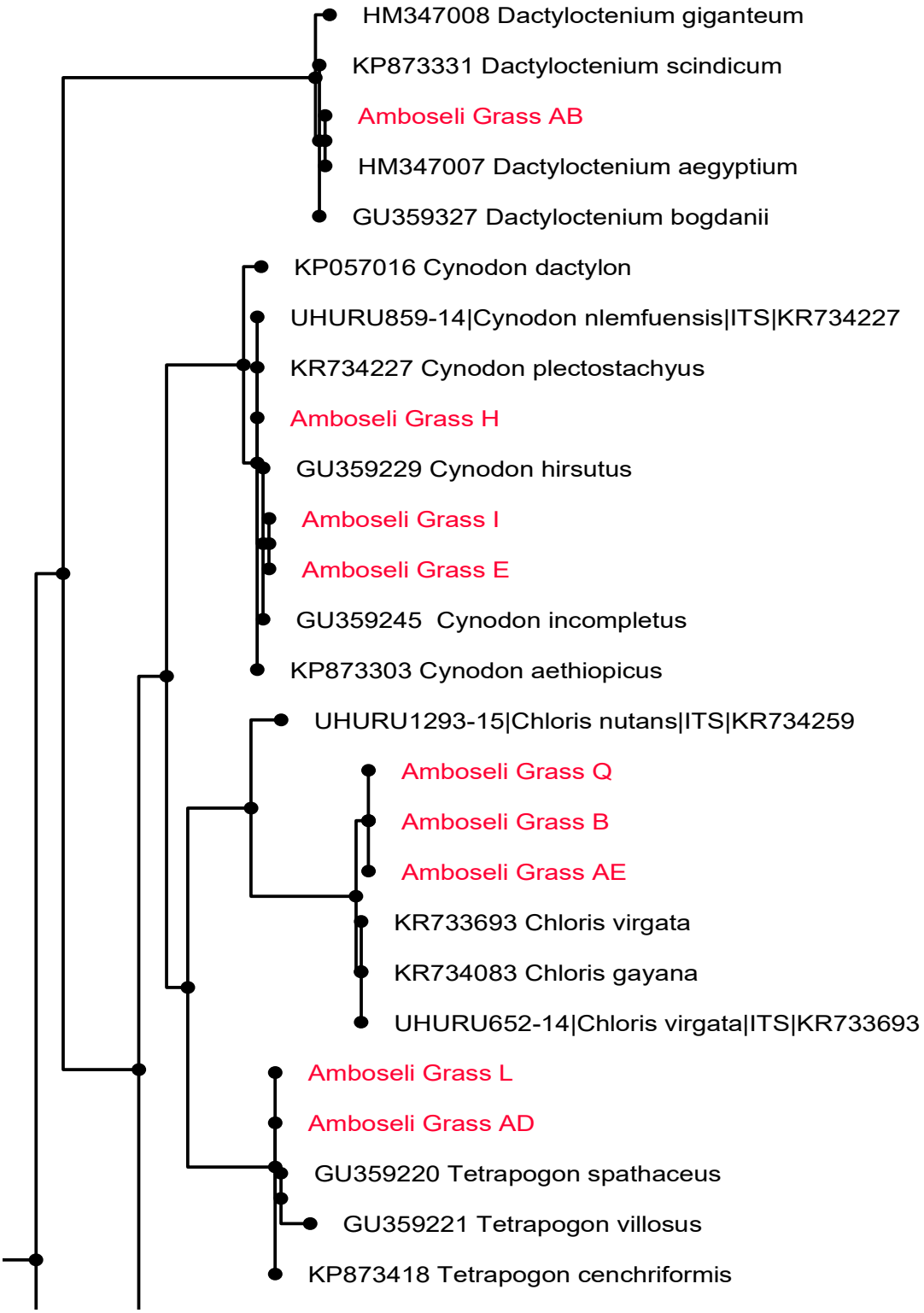
1X TAE recipe:

- 100 mL of 10X TAE buffer + 900 mL of distilled water.

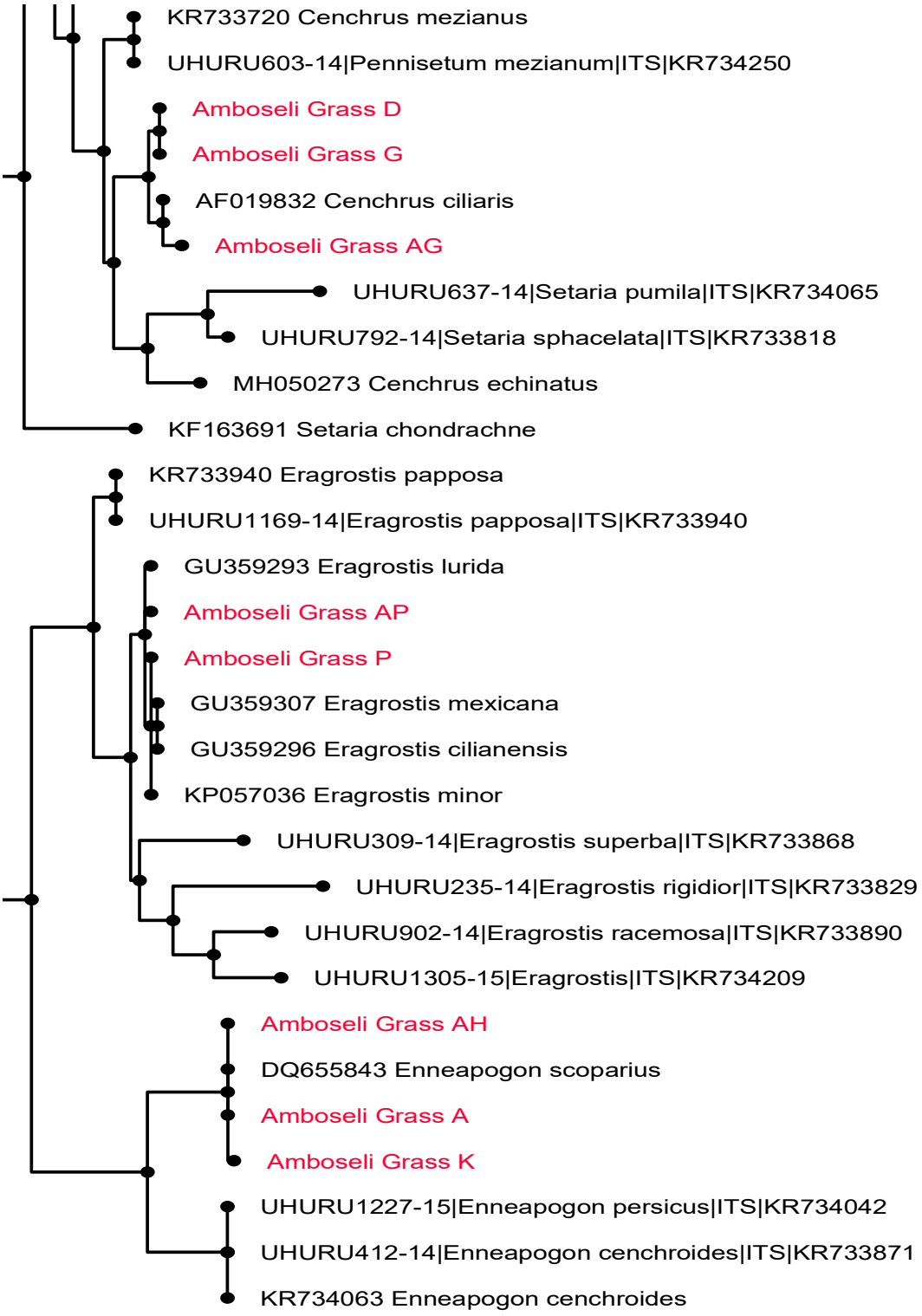
Appendix 4: Phylogenetic tree for monocotyledons based on the *ITS*1 gene.



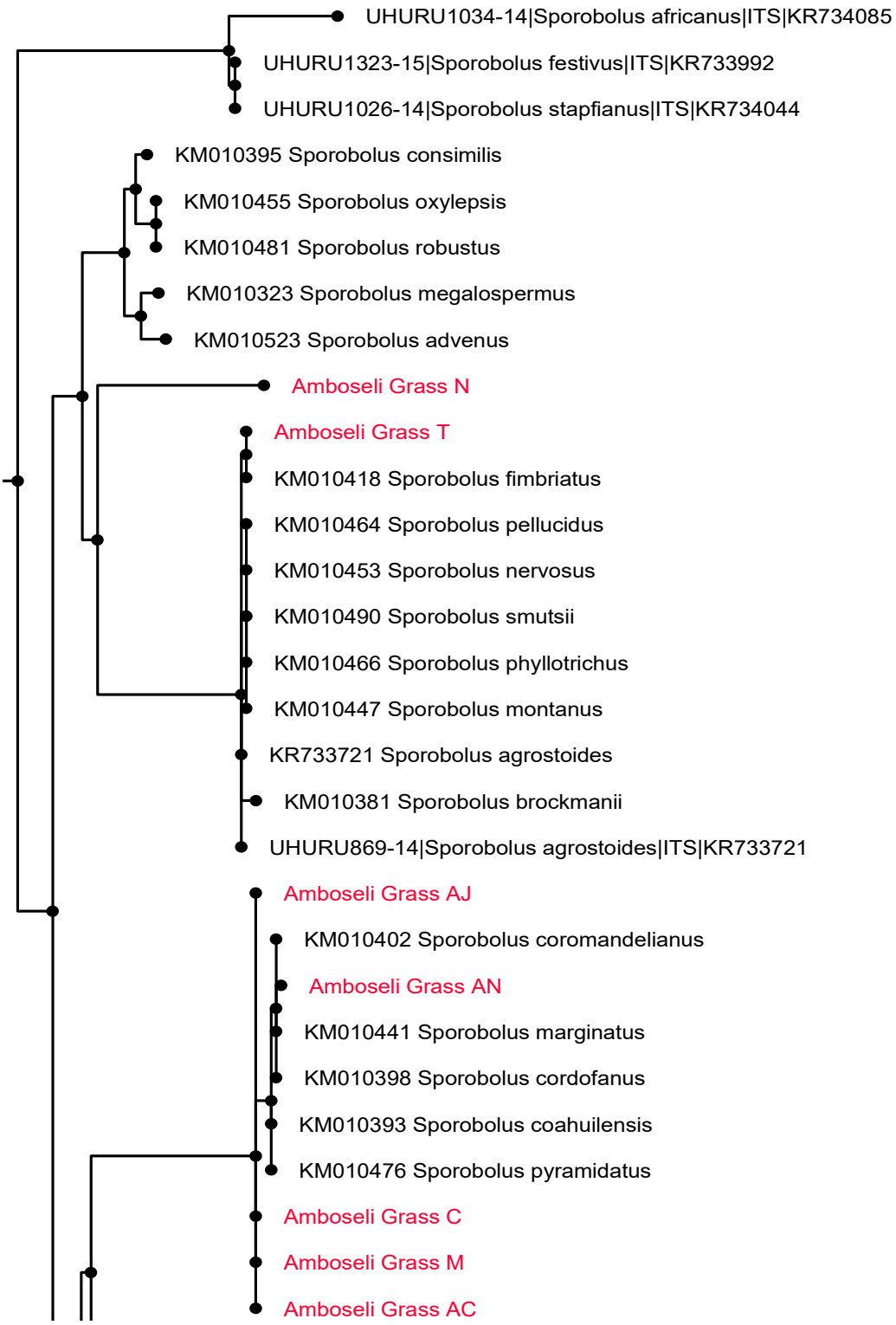
The phylogenetic tree for monocotyledons based on the *ITS*1 gene.



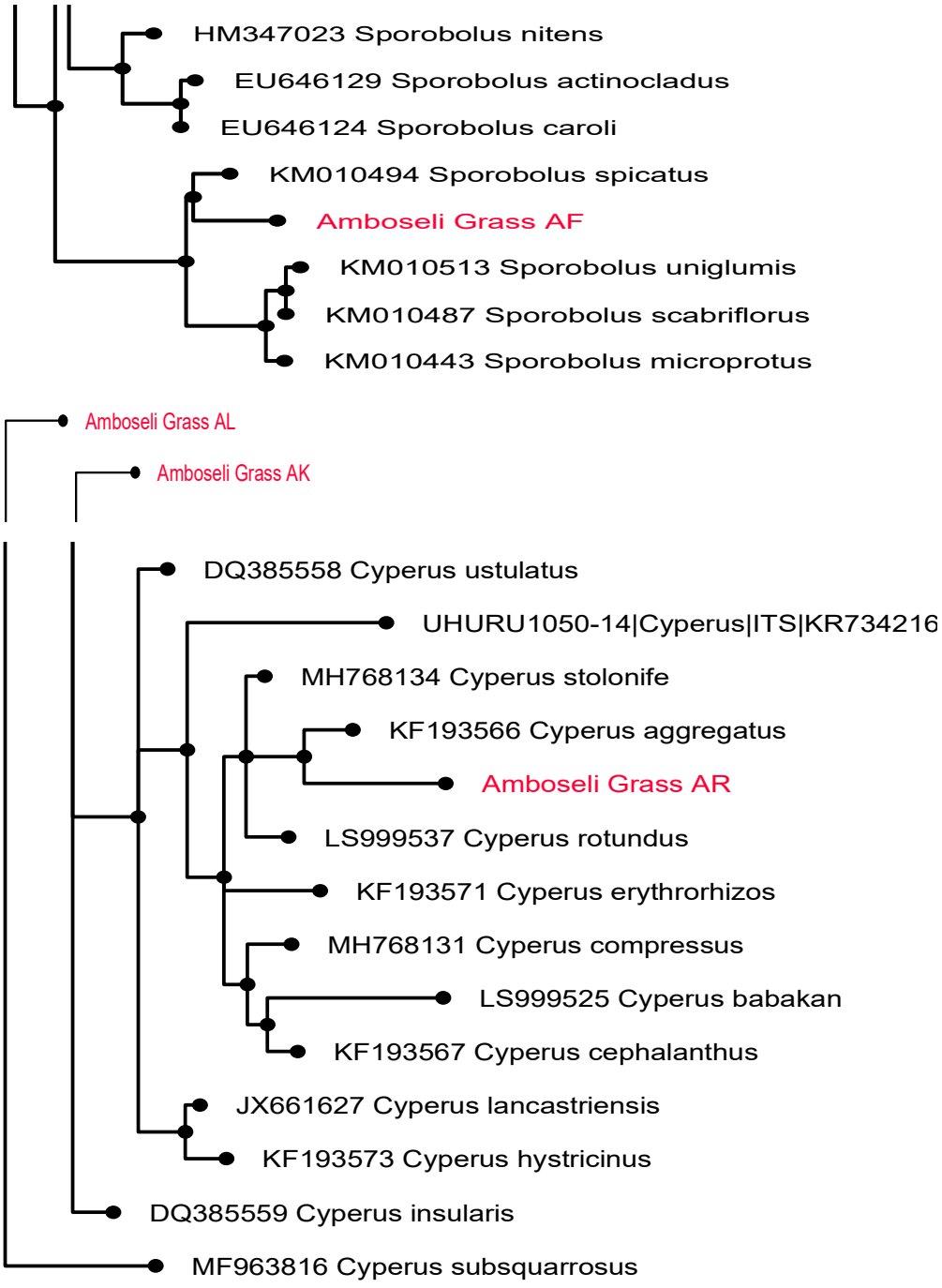
The phylogenetic tree for monocotyledons based on the *ITS*1 gene.



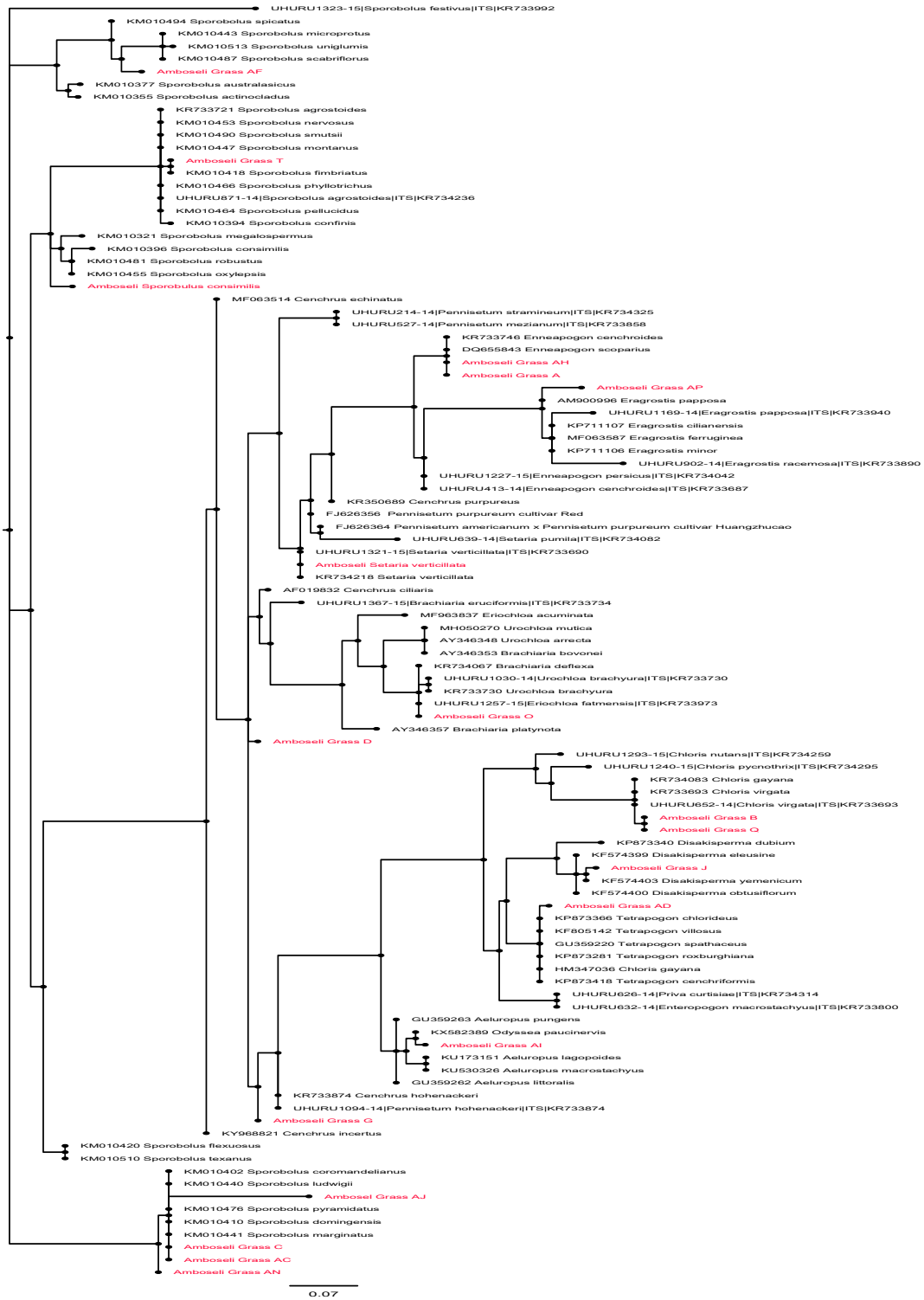
The phylogenetic tree for monocotyledons based on the *ITS*1 gene.



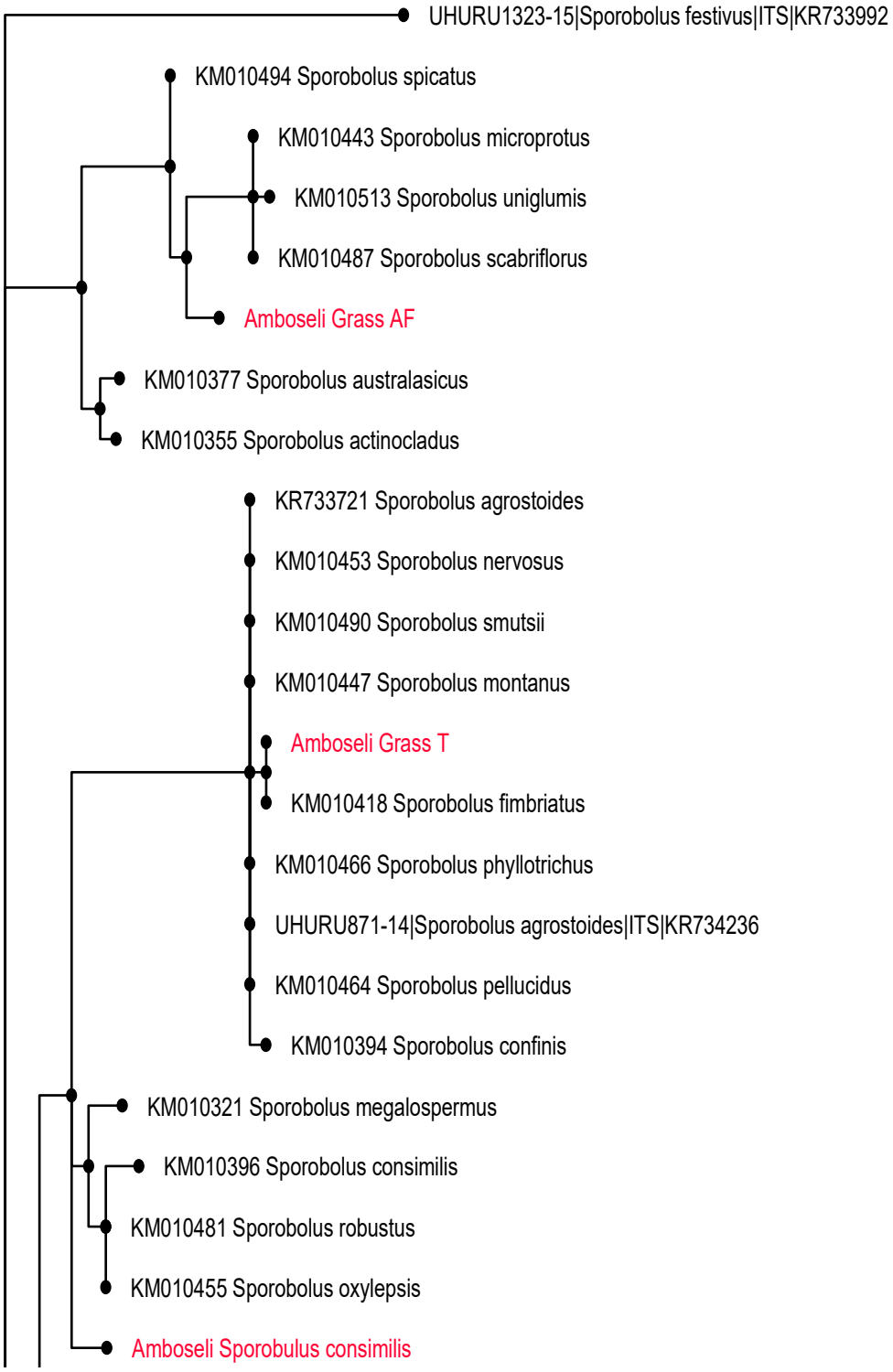
The phylogenetic tree for monocotyledons based on the *ITS*1 gene.



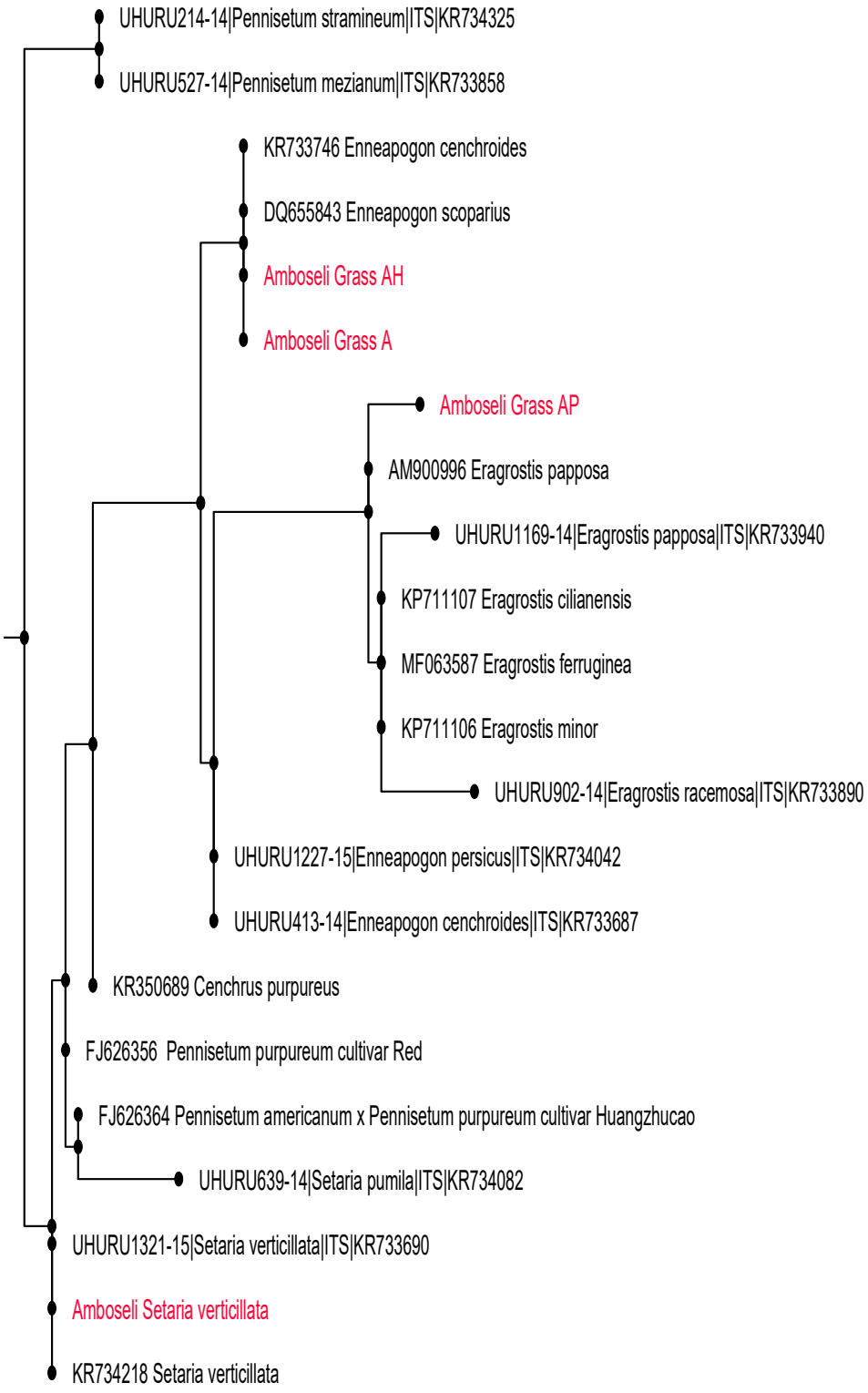
Appendix 5: Phylogenetic tree for monocotyledons based on the *ITS*1-Poaceae barcode.



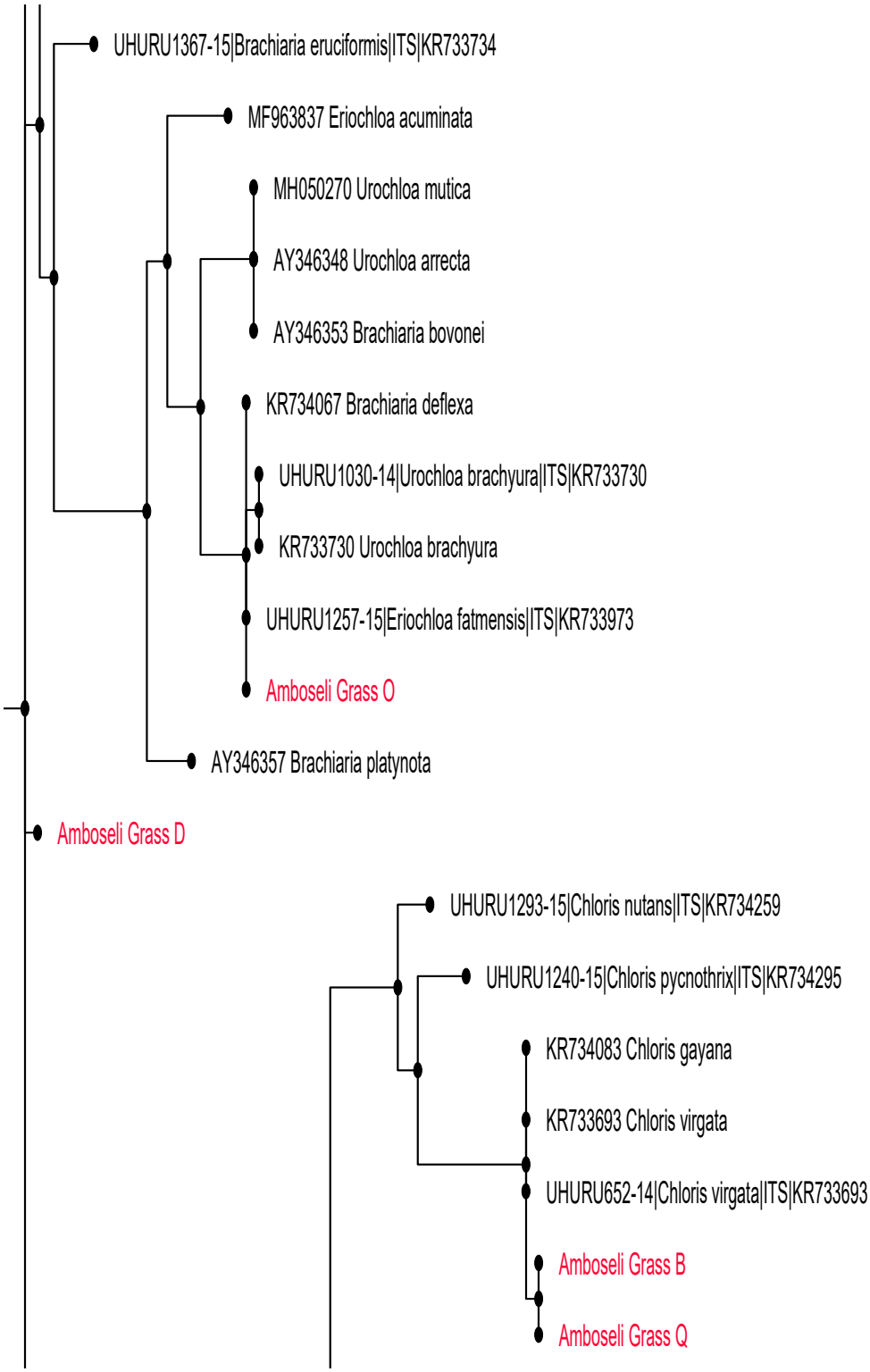
The phylogenetic tree for monocotyledons based on the *ITS*1-Poaceae barcode.



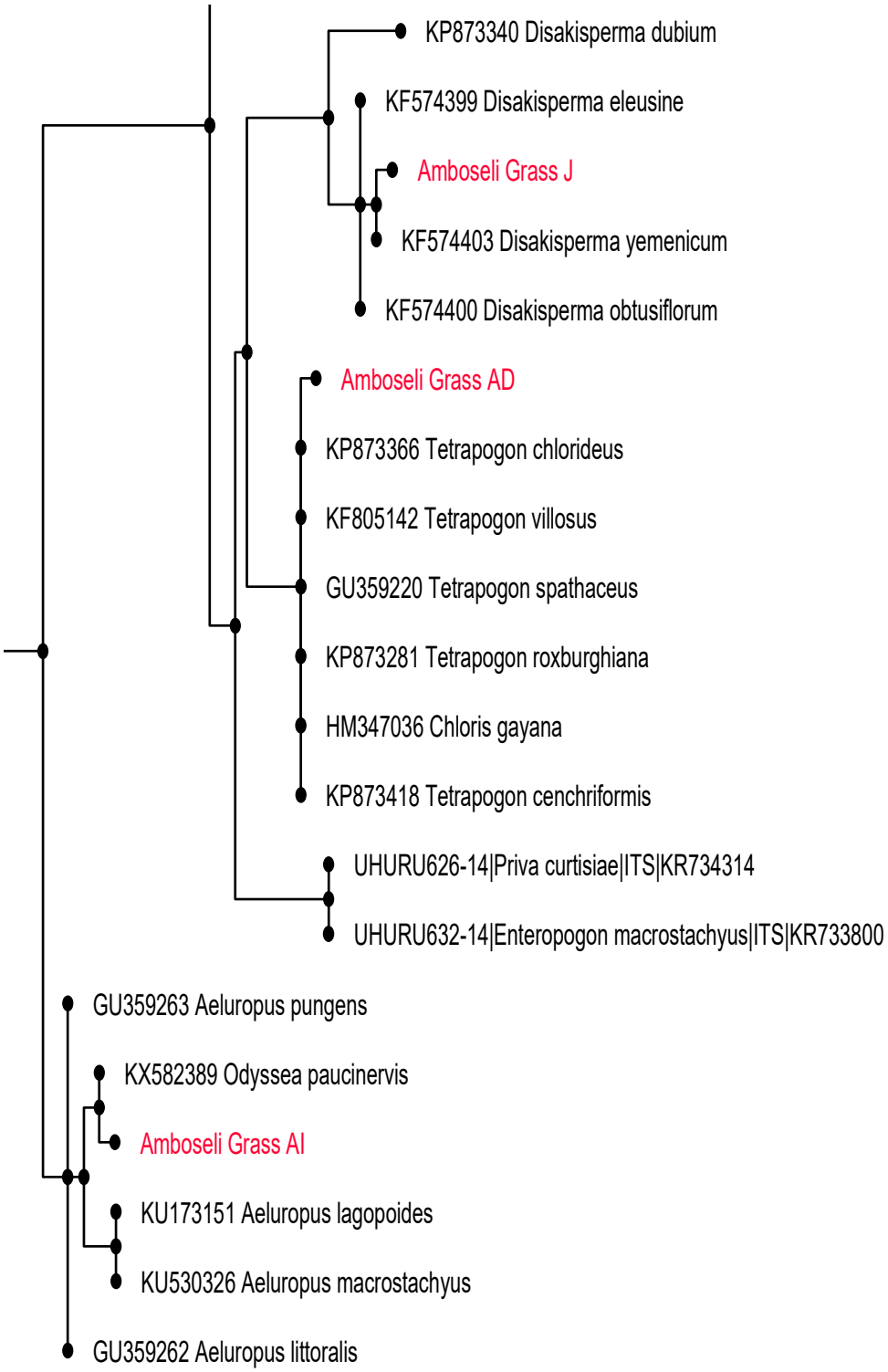
The phylogenetic tree for monocotyledons based on the *ITS*1-Poaceae barcode.



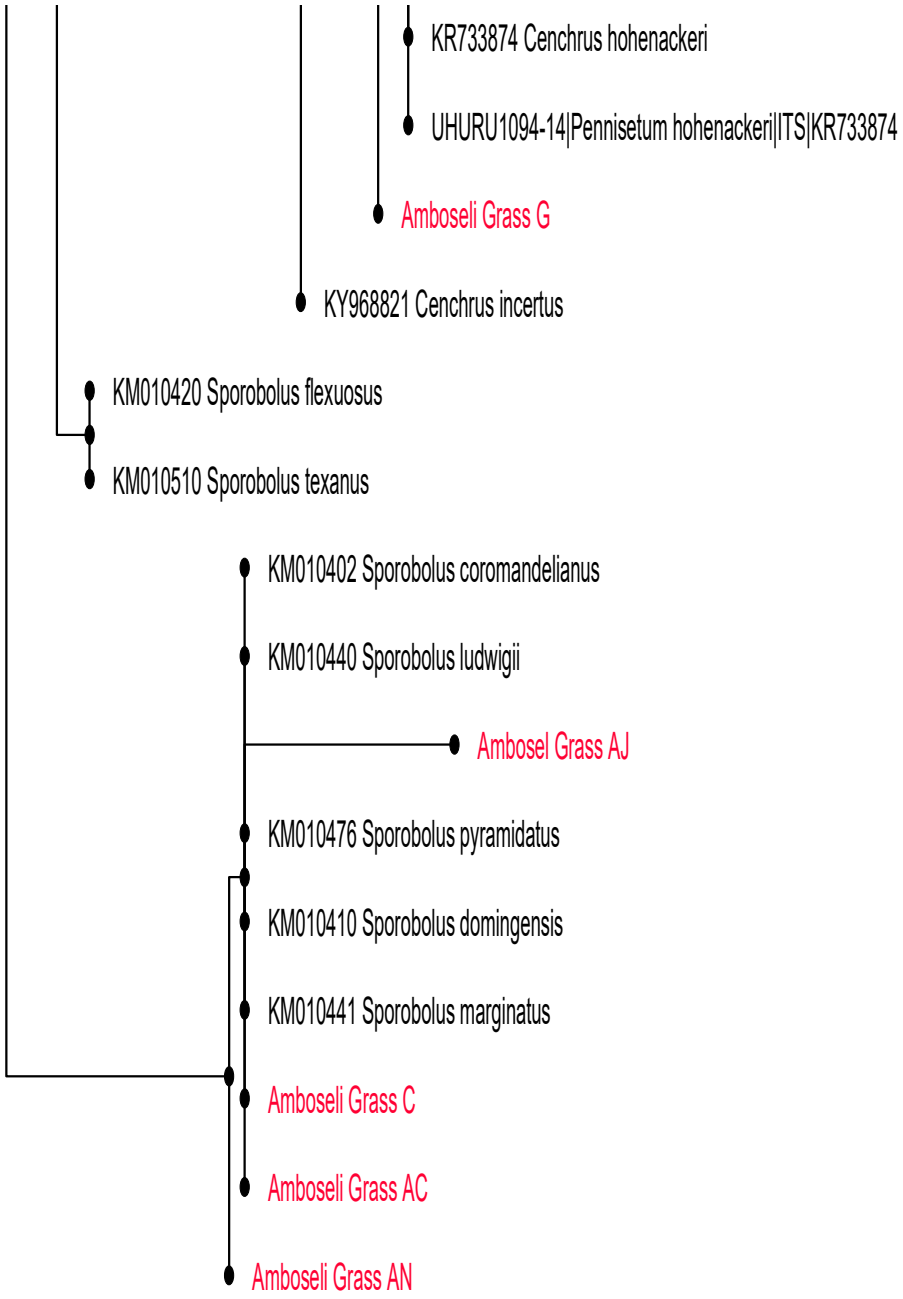
The phylogenetic tree for monocotyledons based on the *ITS*1-Poaceae barcode.



The phylogenetic tree for monocotyledons based on the *ITS*1-Poaceae barcode.

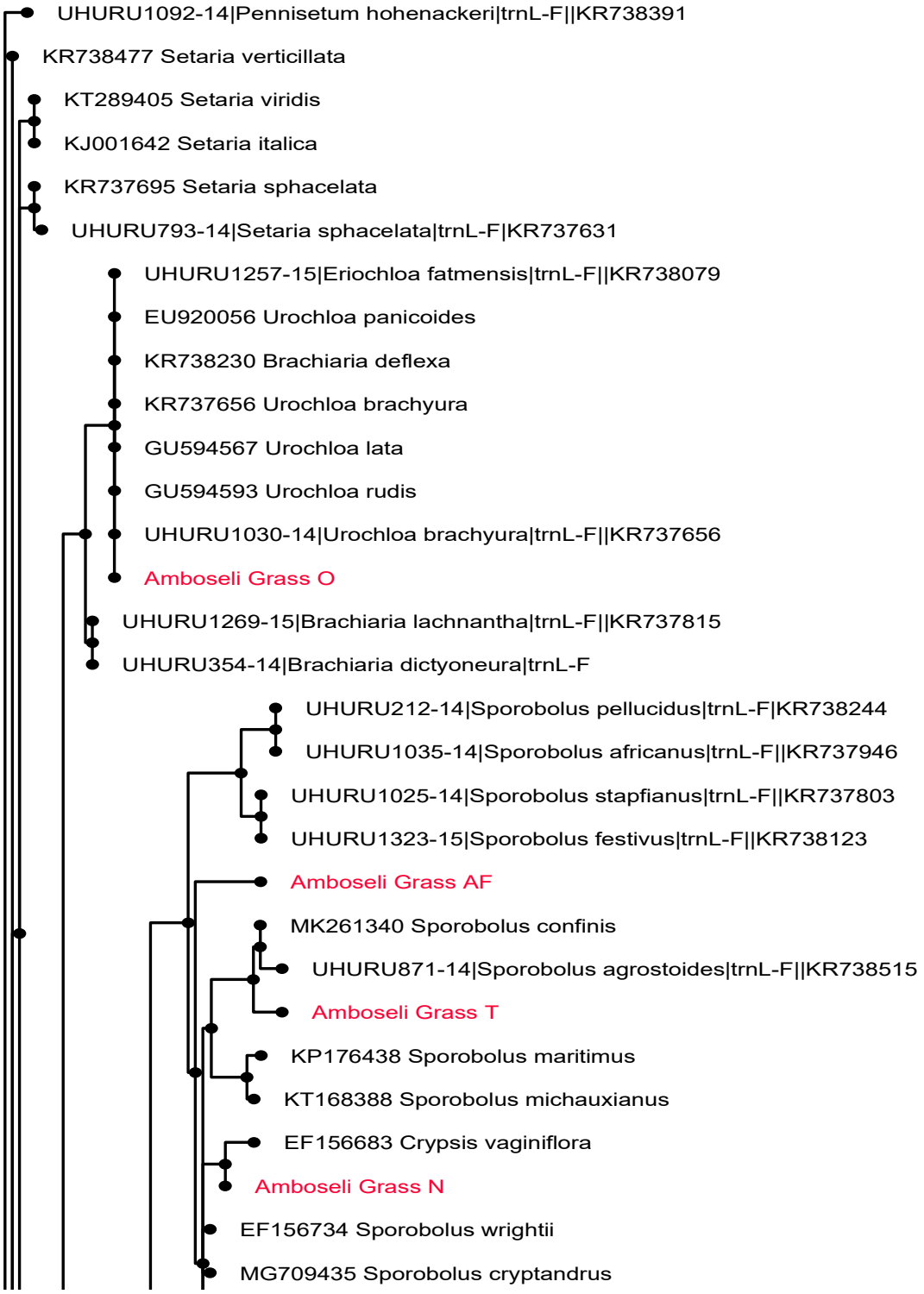


The phylogenetic tree for monocotyledons based on the *ITS*1-Poaceae barcode.

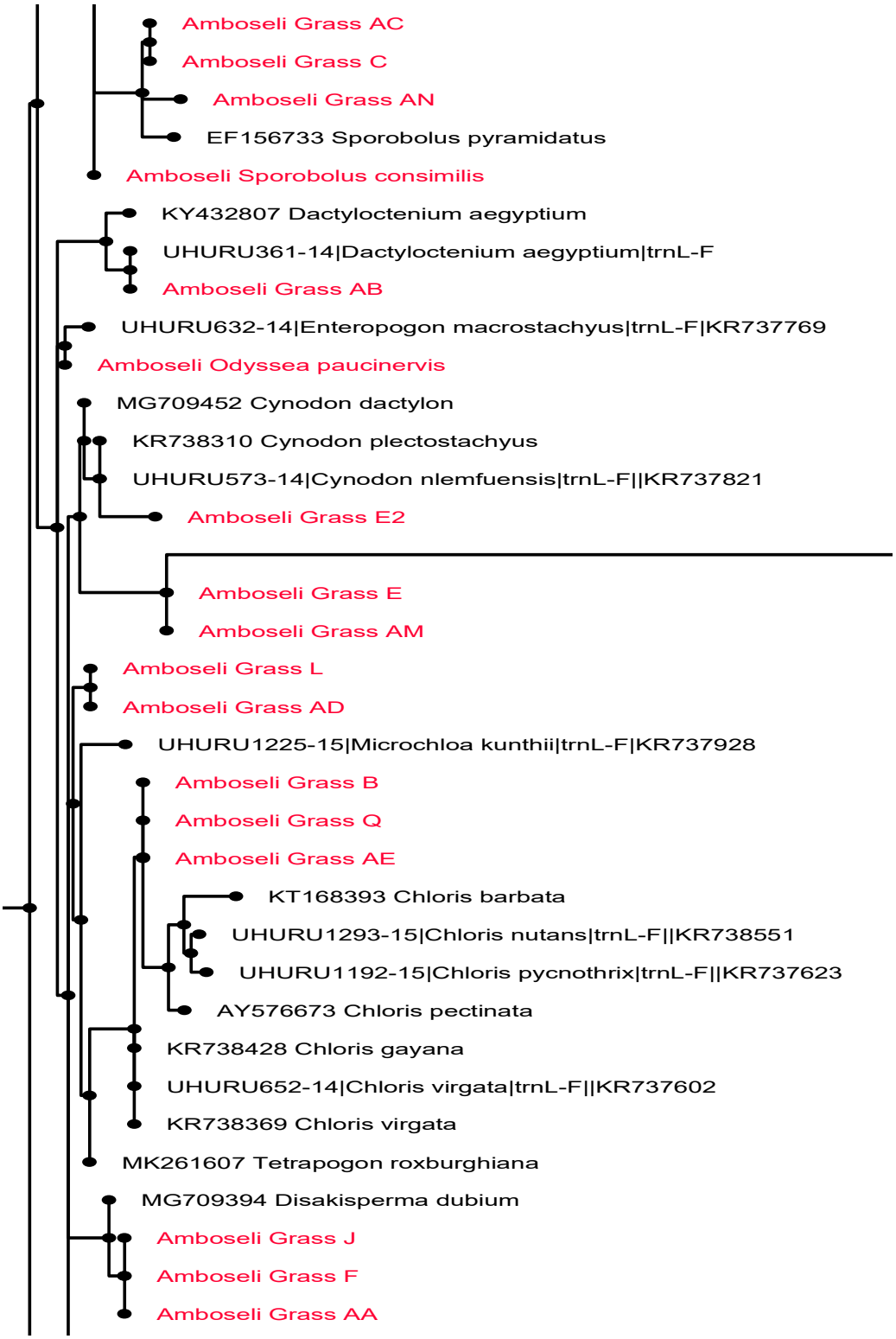


Appendix 6: Phylogenetic tree for monocotyledons based on the *trn*L (UAA) intron.

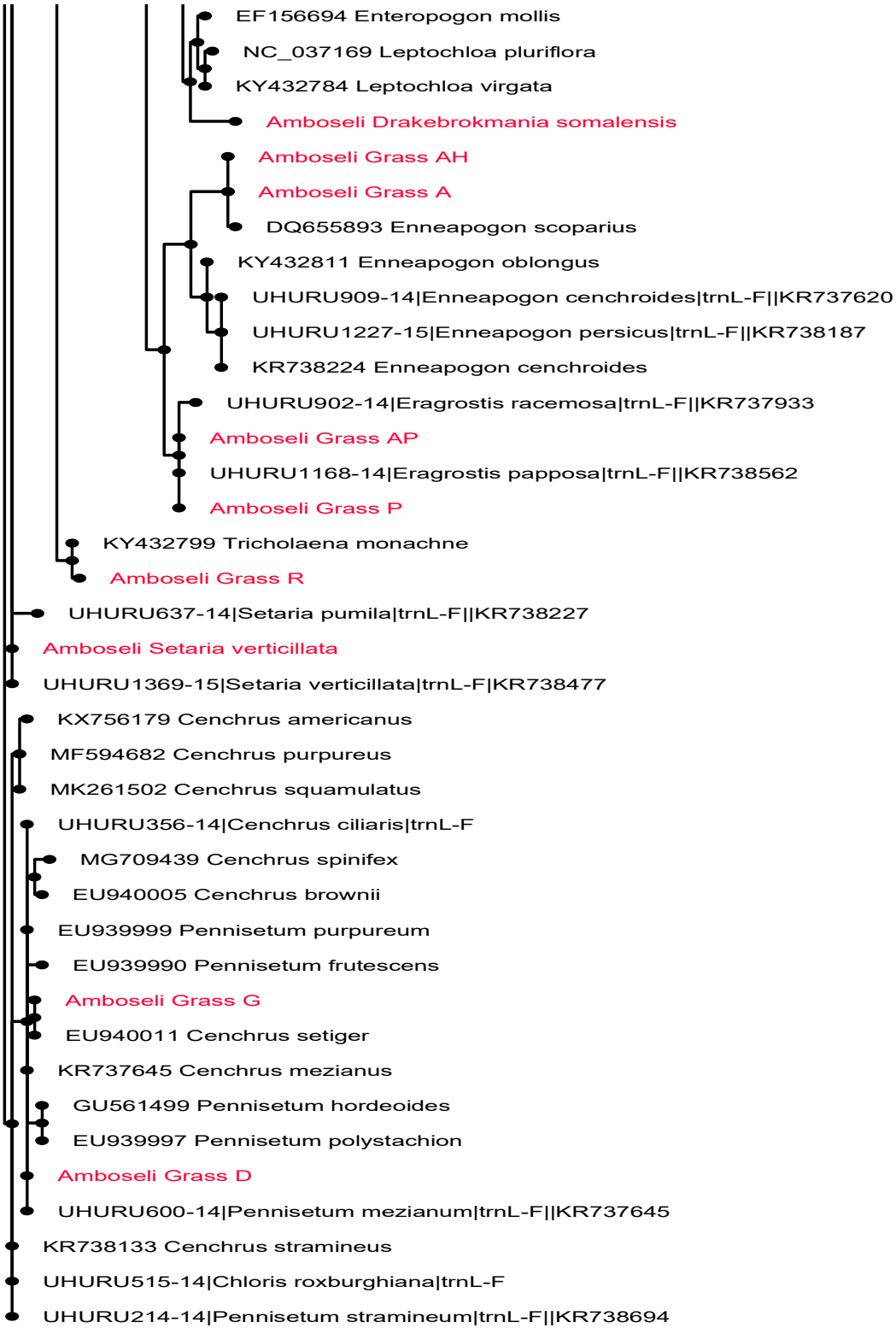
The phylogenetic tree for monocotyledons based on the *trn*L (UAA) intron.



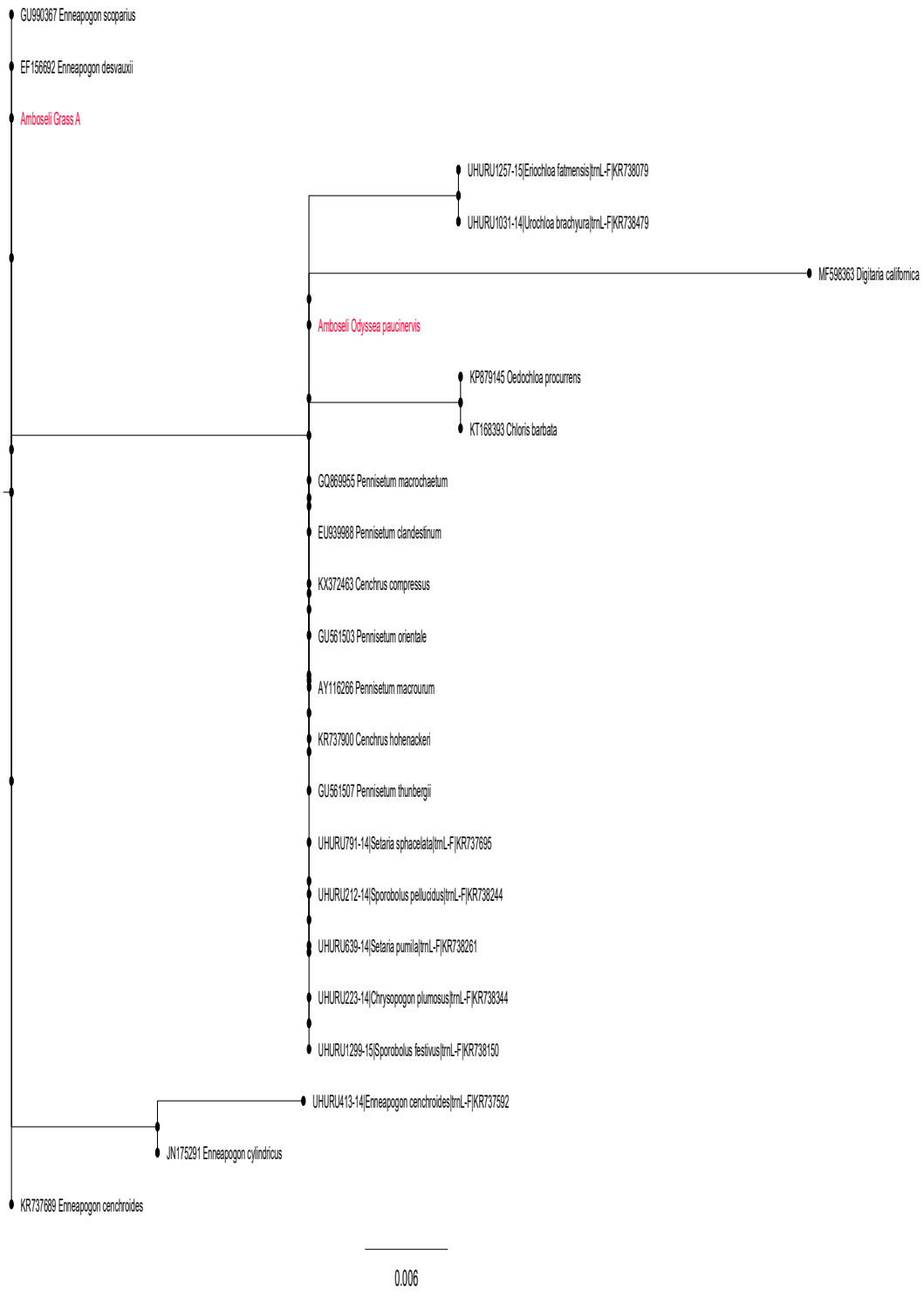
The phylogenetic tree for monocotyledons based on the *trn*L (UAA) intron.



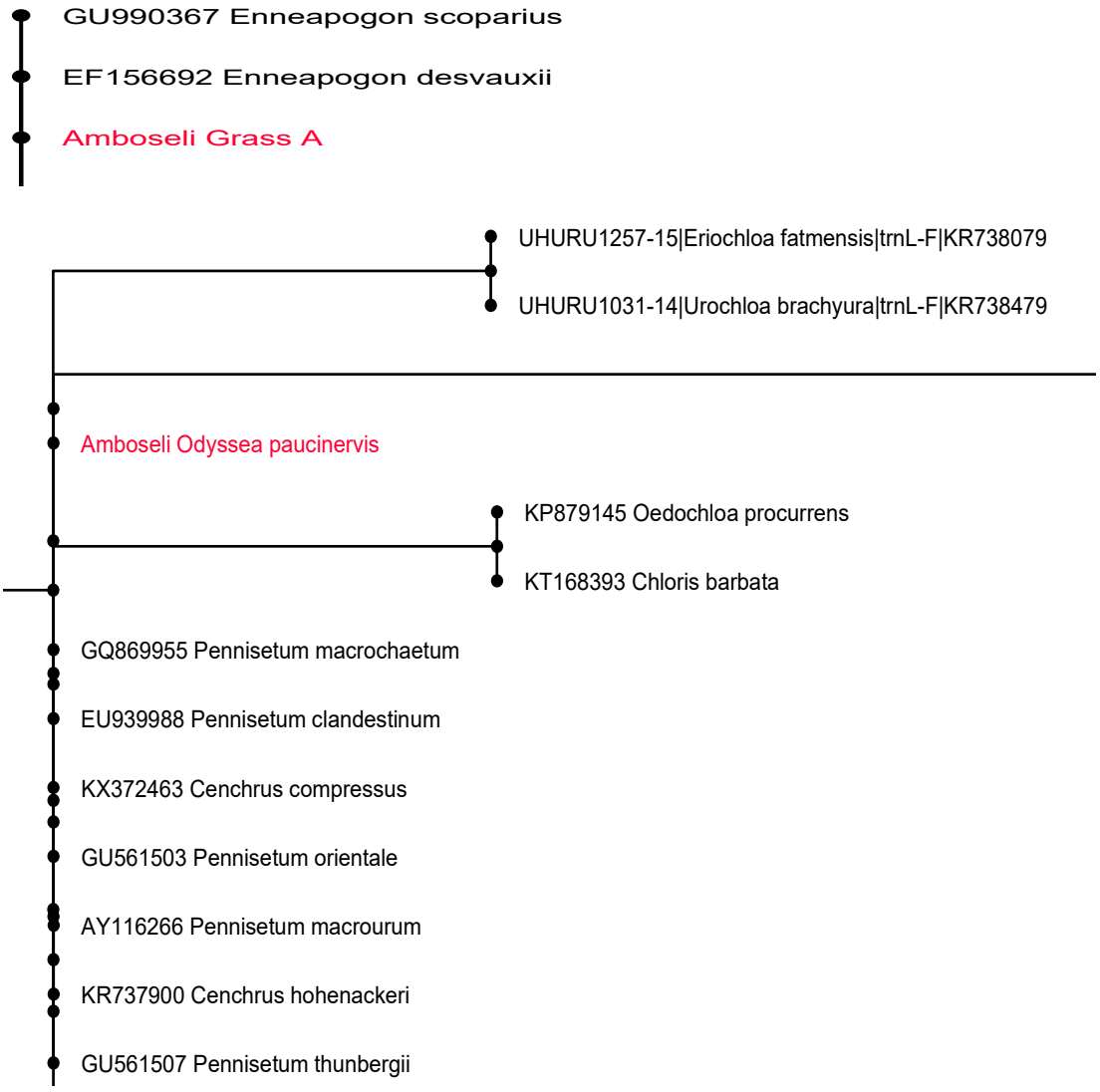
The phylogenetic tree for monocotyledons based on the *trn*L (UAA) intron.



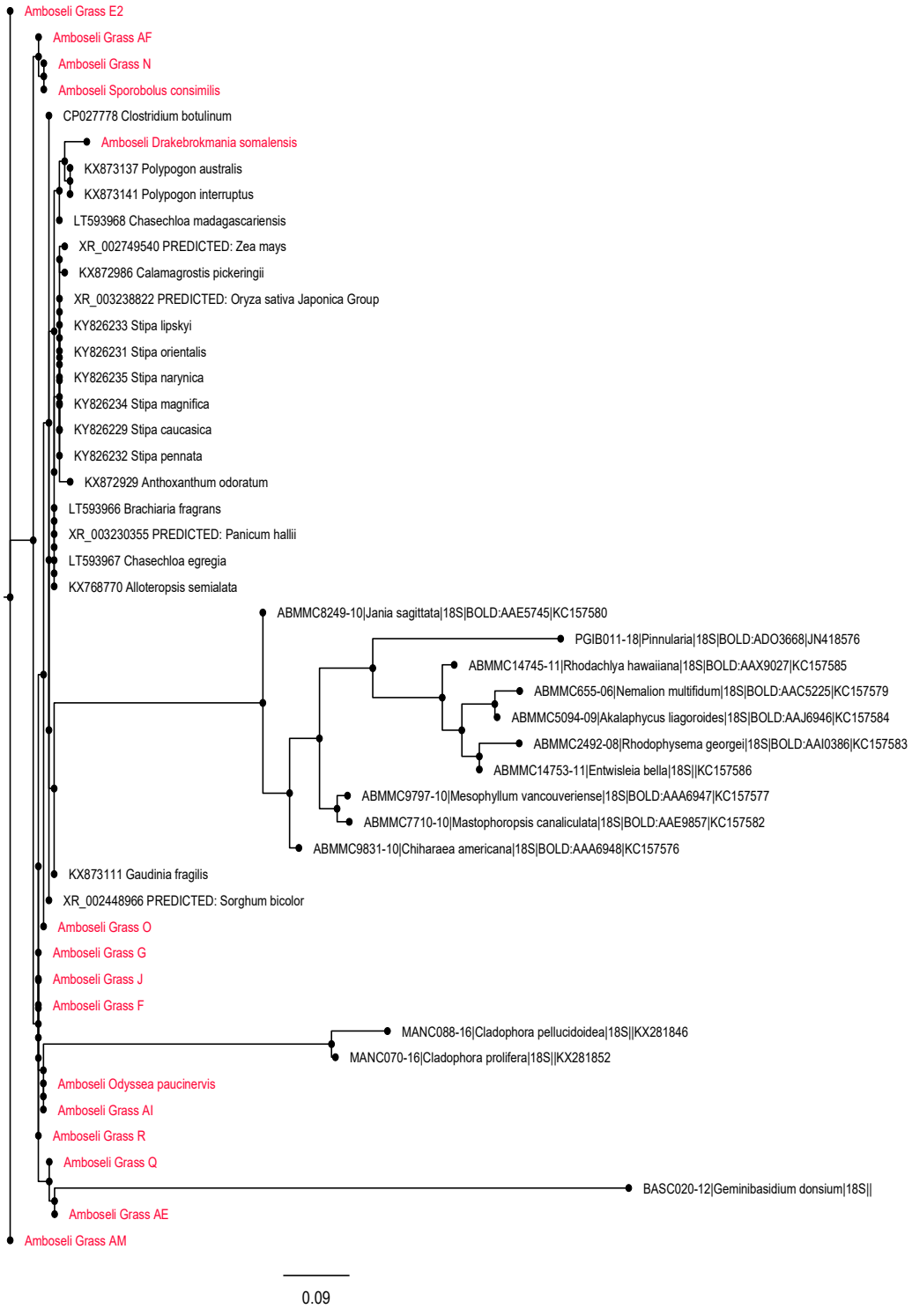
Appendix 7: Phylogenetic tree for monocotyledons based on the *trn*L-p6 locus.



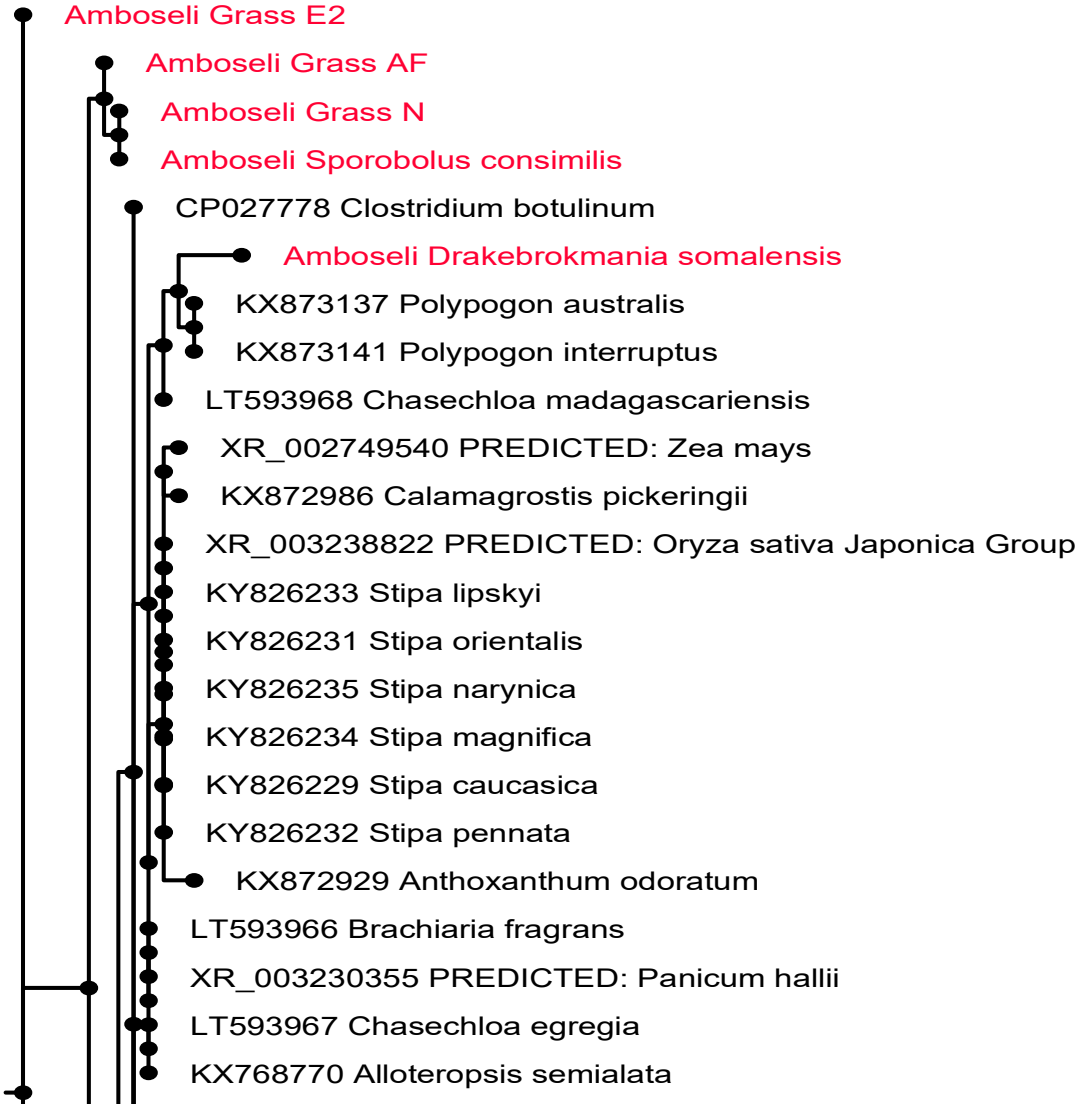
The phylogenetic tree for monocotyledons based on the *trn*L-p6 locus.



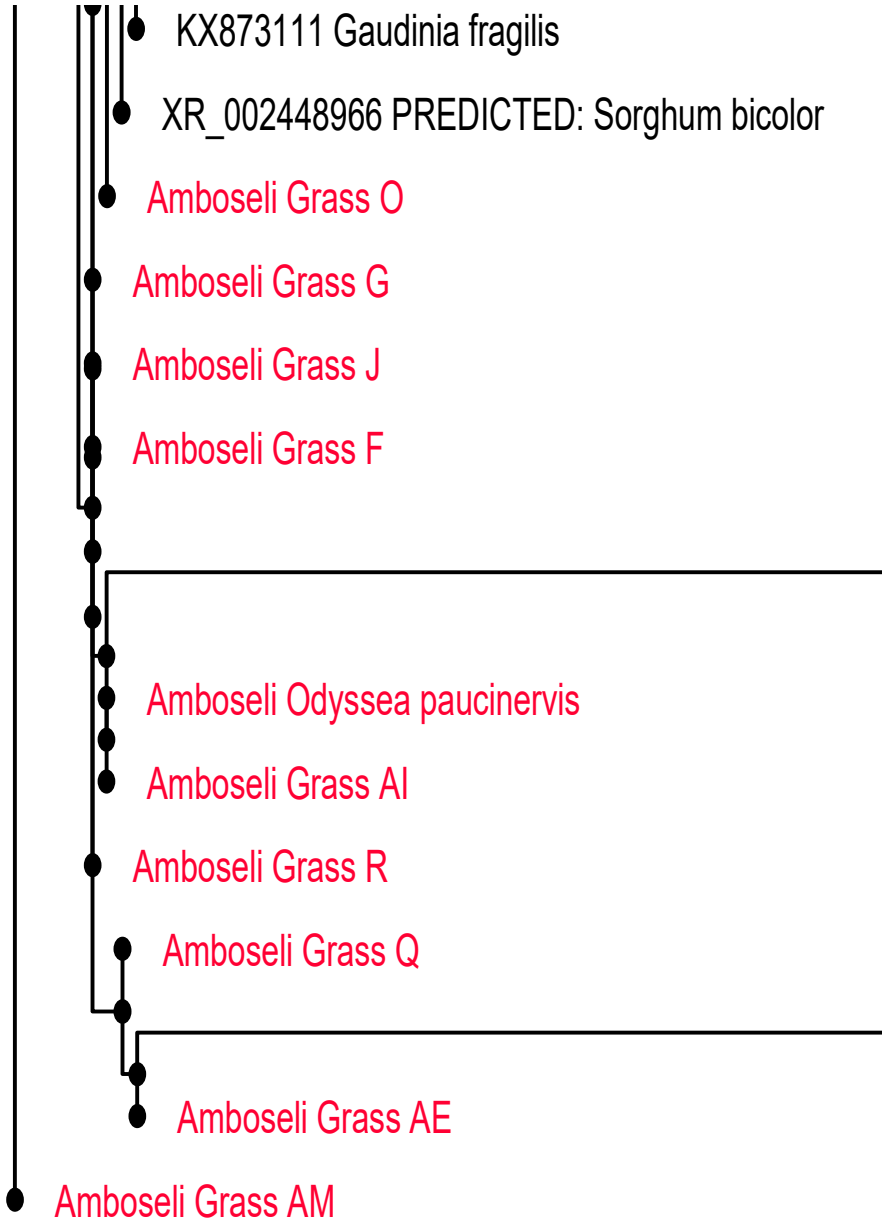
Appendix 8: Phylogenetic tree for monocotyledons based on the 18S rDNA region.



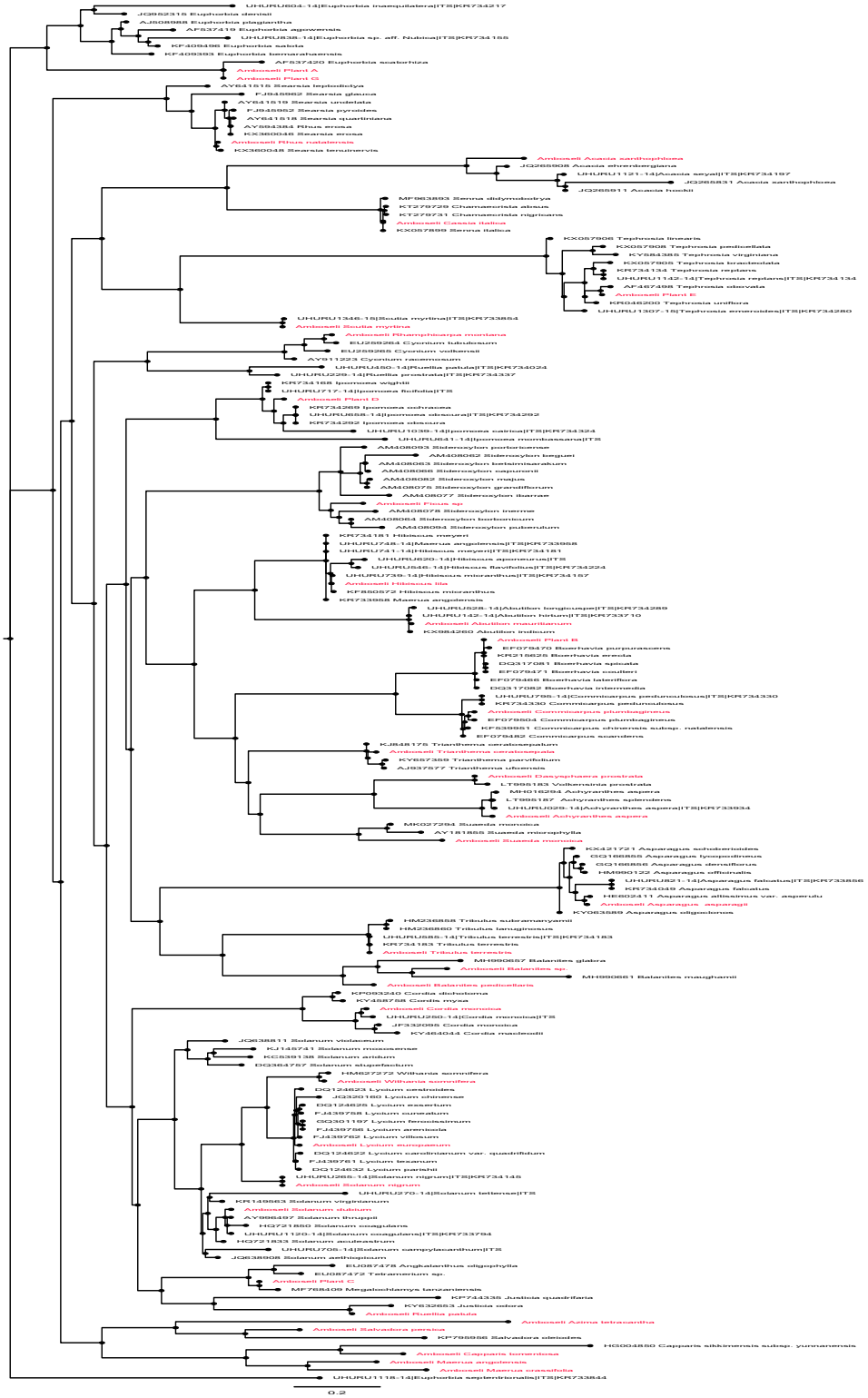
The phylogenetic tree for monocotyledons based on the 18S rDNA region.



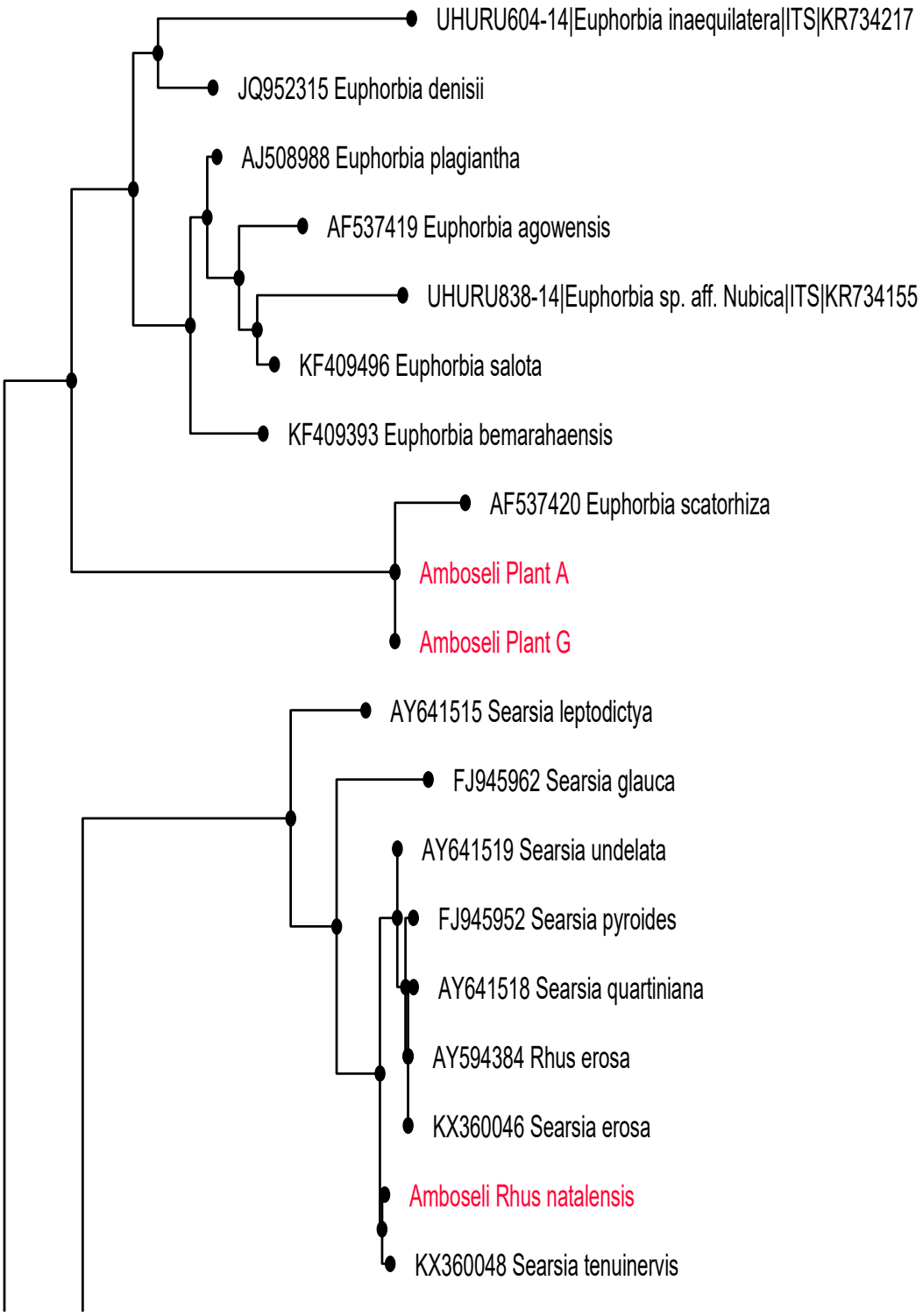
The phylogenetic tree for monocotyledons based on the 18S rDNA region.



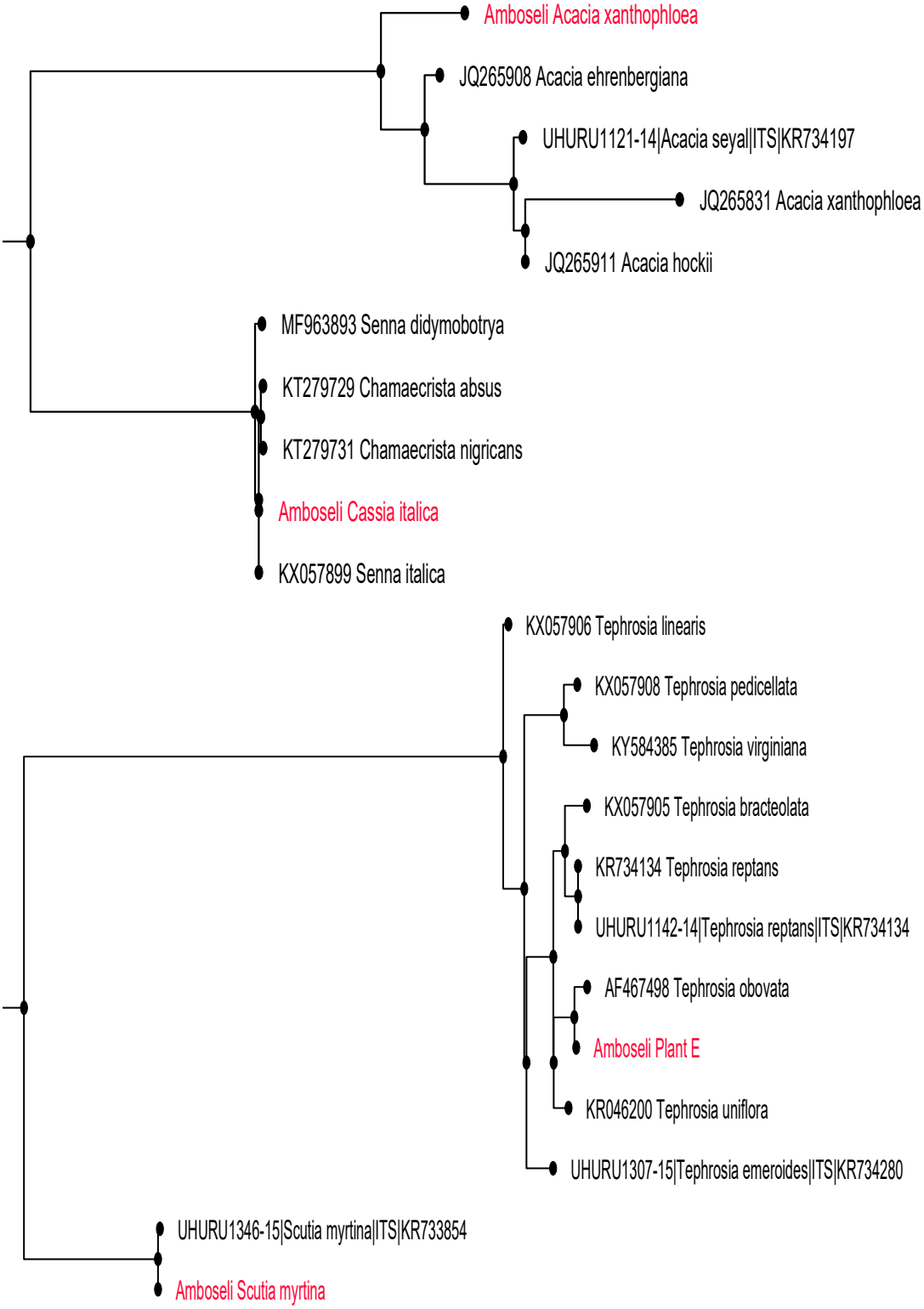
Appendix 9: Phylogenetic tree for dicotyledons based on the *ITS*1 gene.



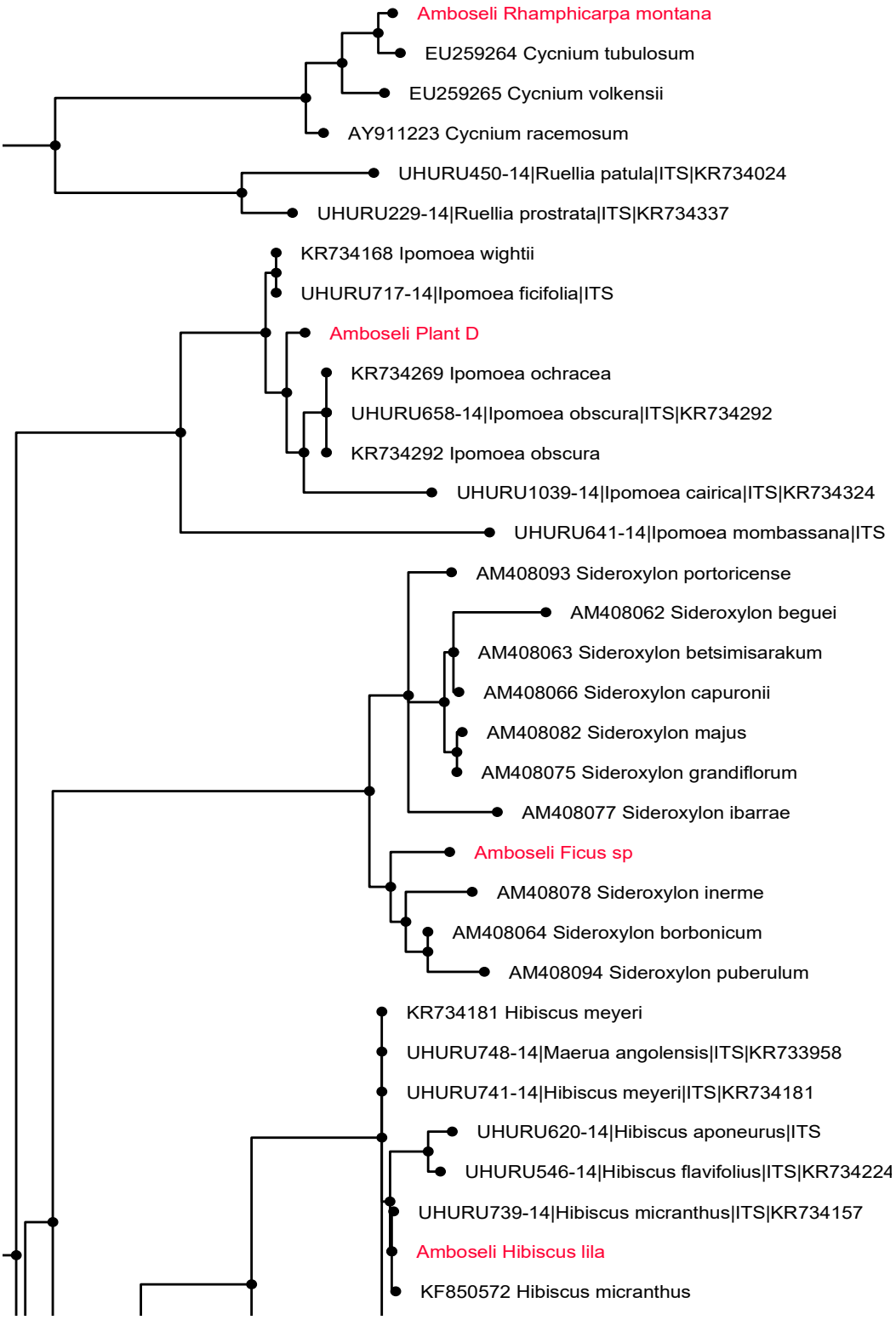
The phylogenetic tree for dicotyledons based on the *ITS*1 gene.



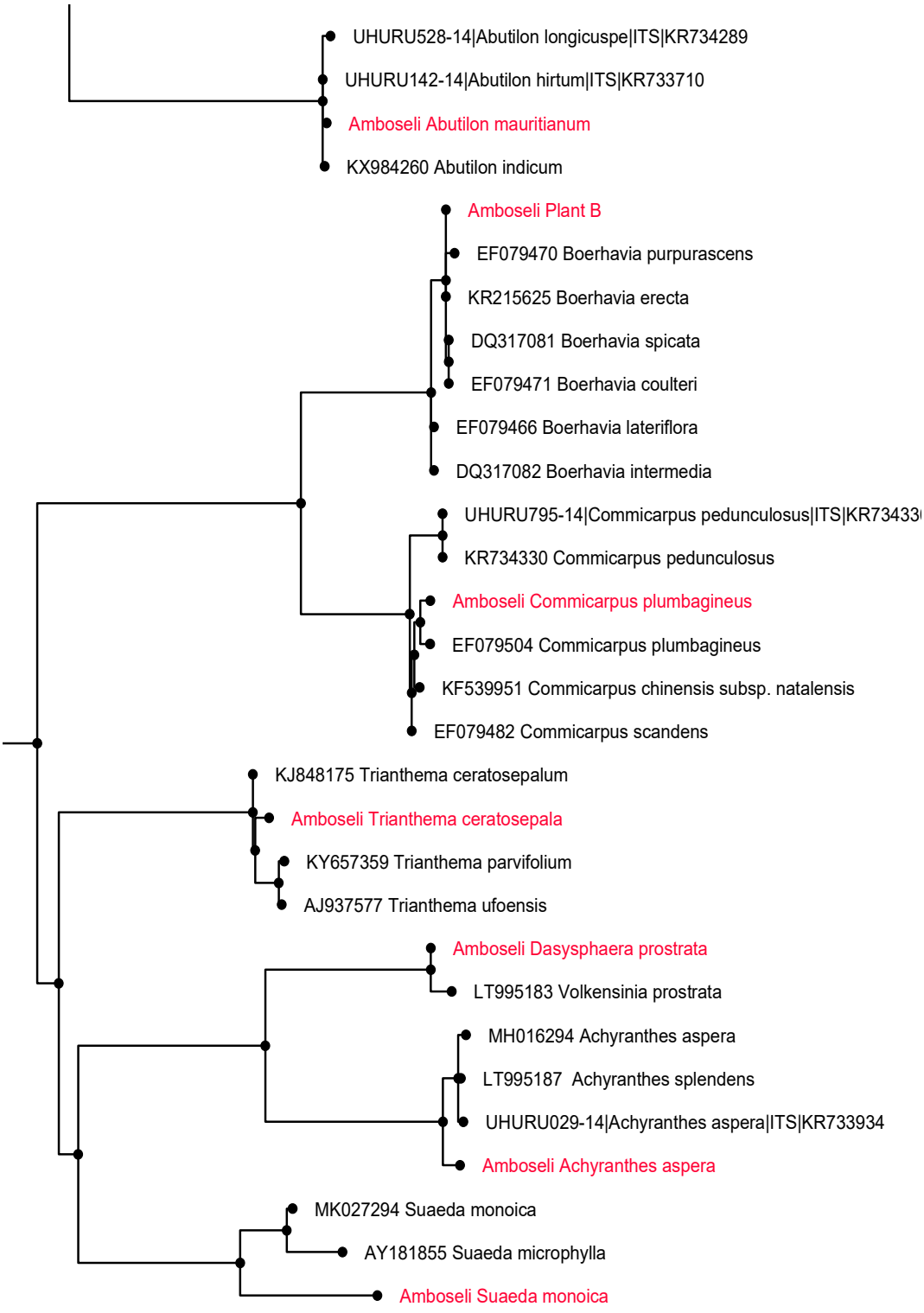
The phylogenetic tree for dicotyledons based on the *ITS*1 gene.



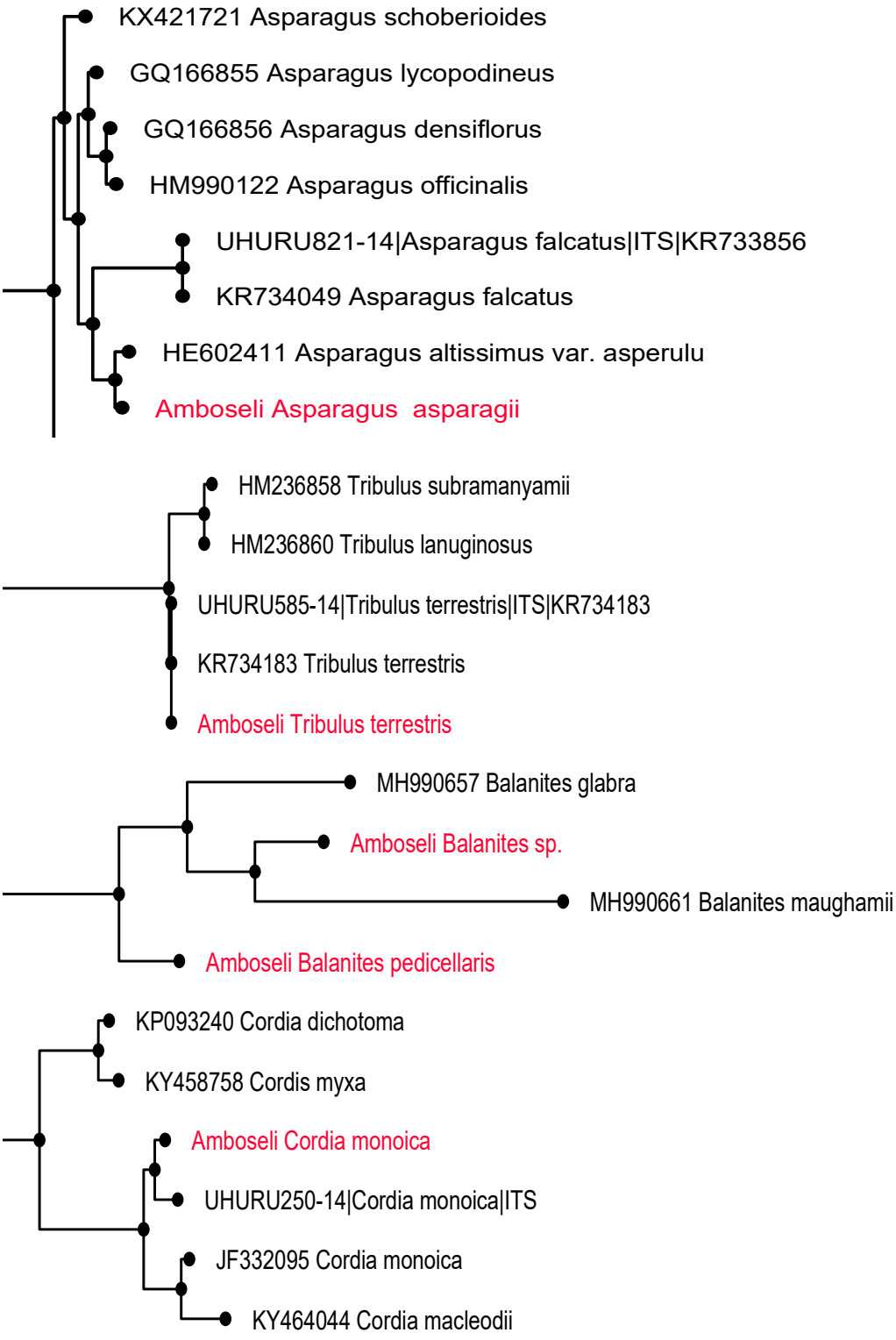
The phylogenetic tree for dicotyledons based on the *ITS*1 gene.



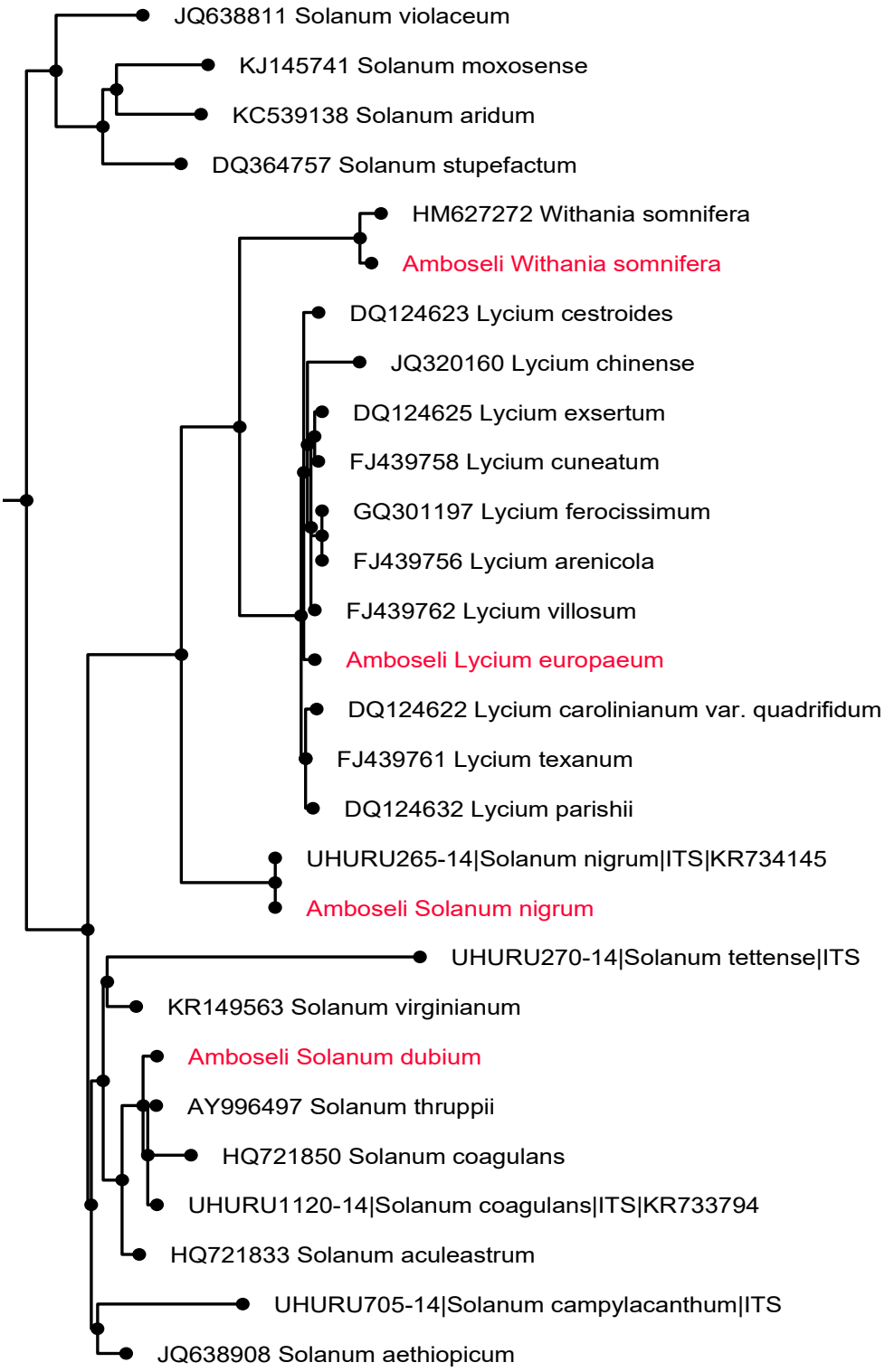
The phylogenetic tree for dicotyledons based on the *ITS*1 gene.



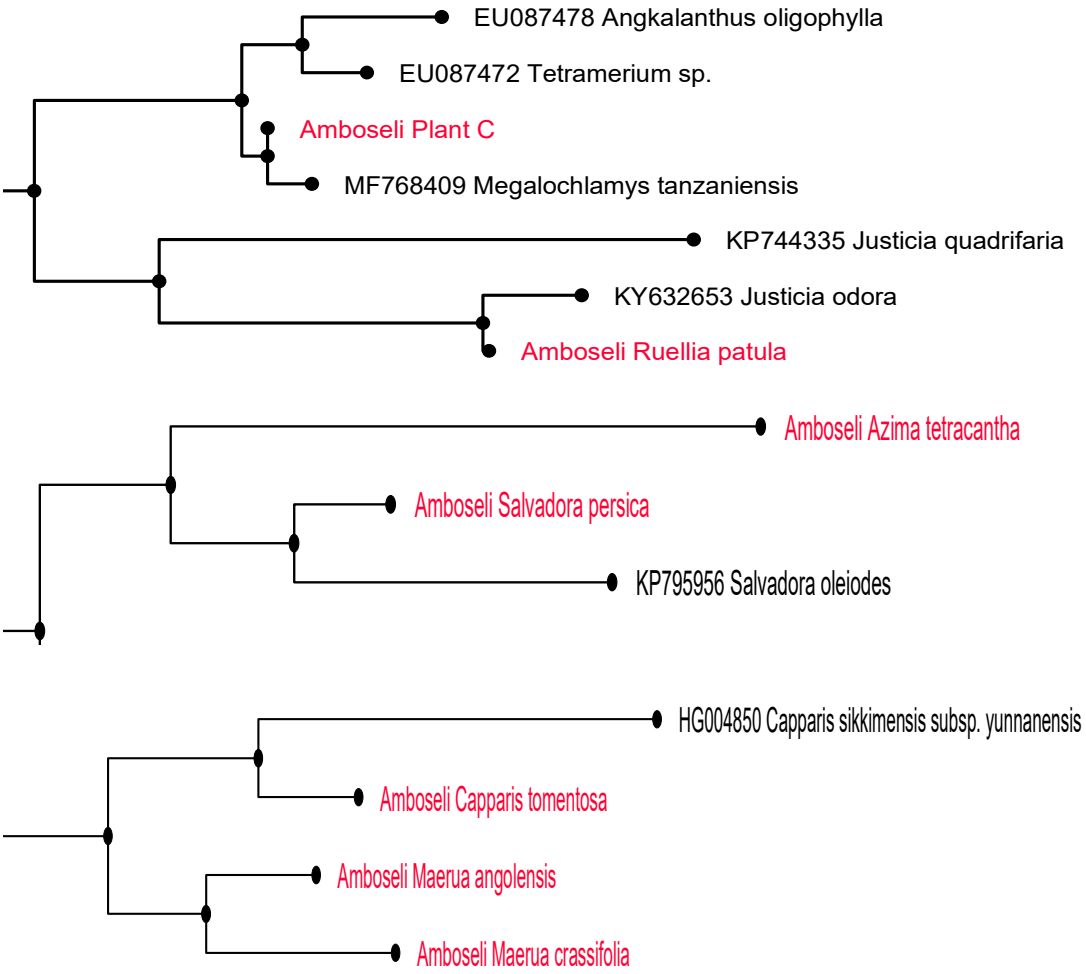
The phylogenetic tree for dicotyledons based on the *ITS*1 gene.



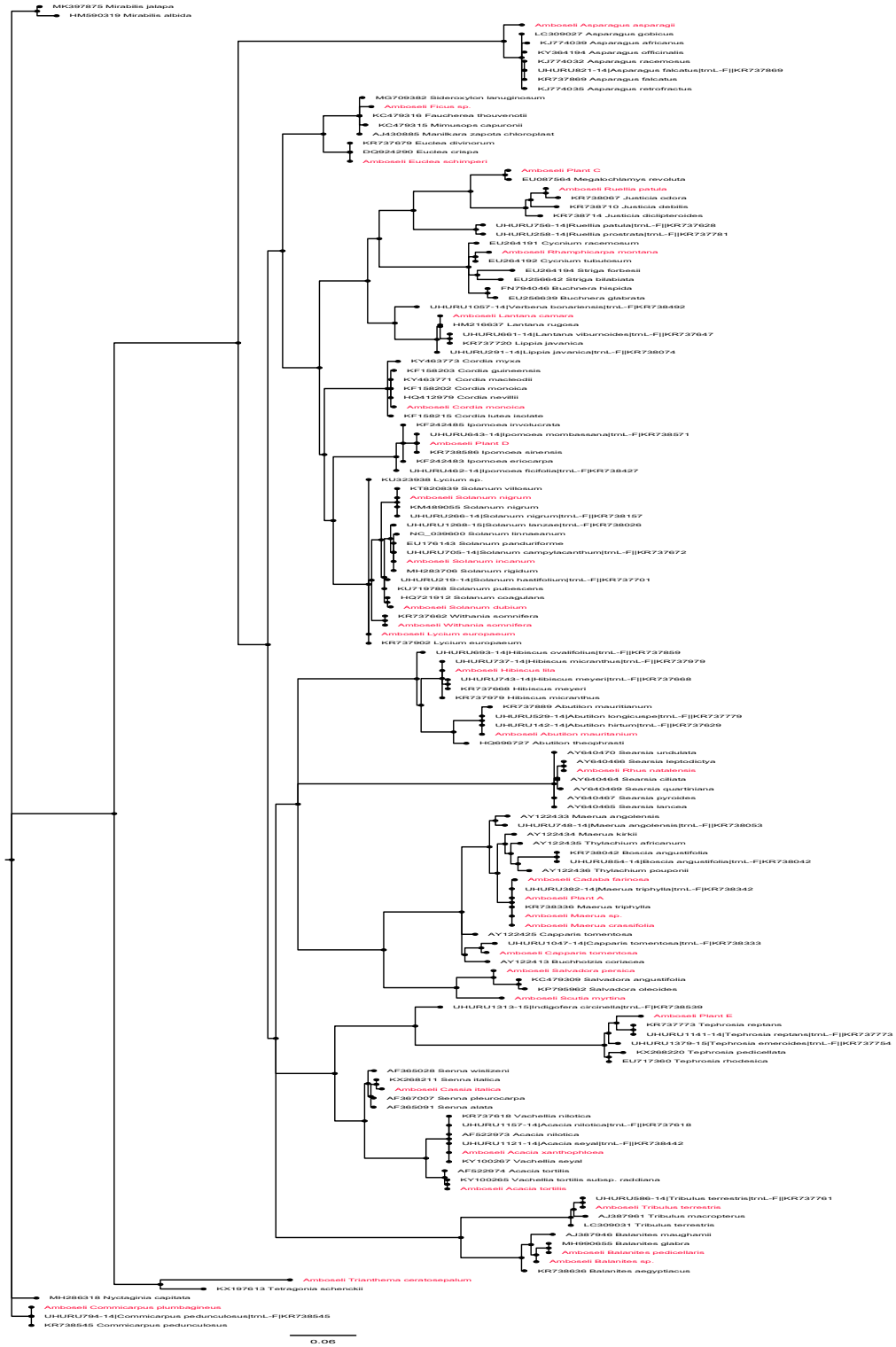
The phylogenetic tree for dicotyledons based on the *ITS*1 gene.



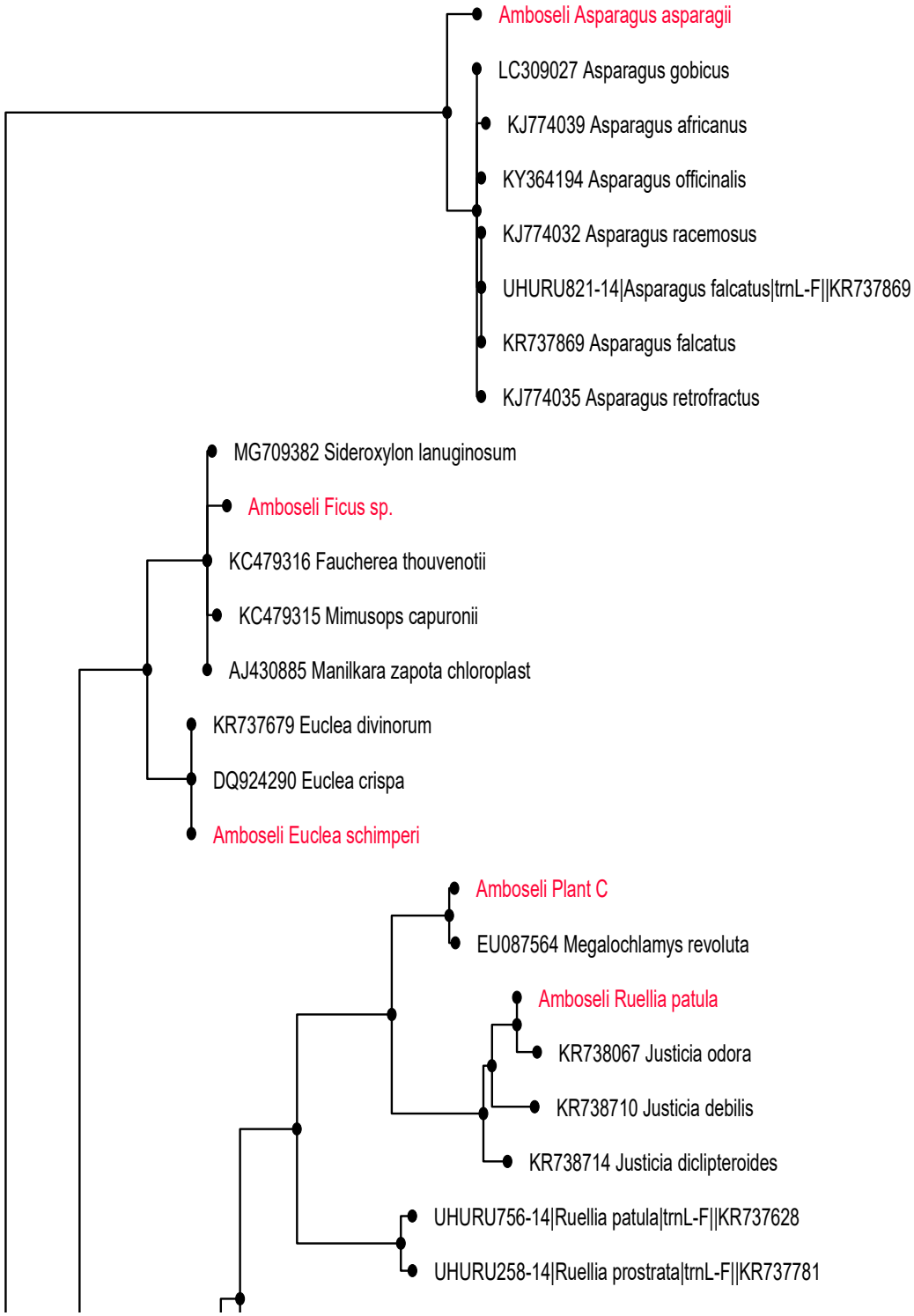
The phylogenetic tree for dicotyledons based on the *ITS*1 gene.



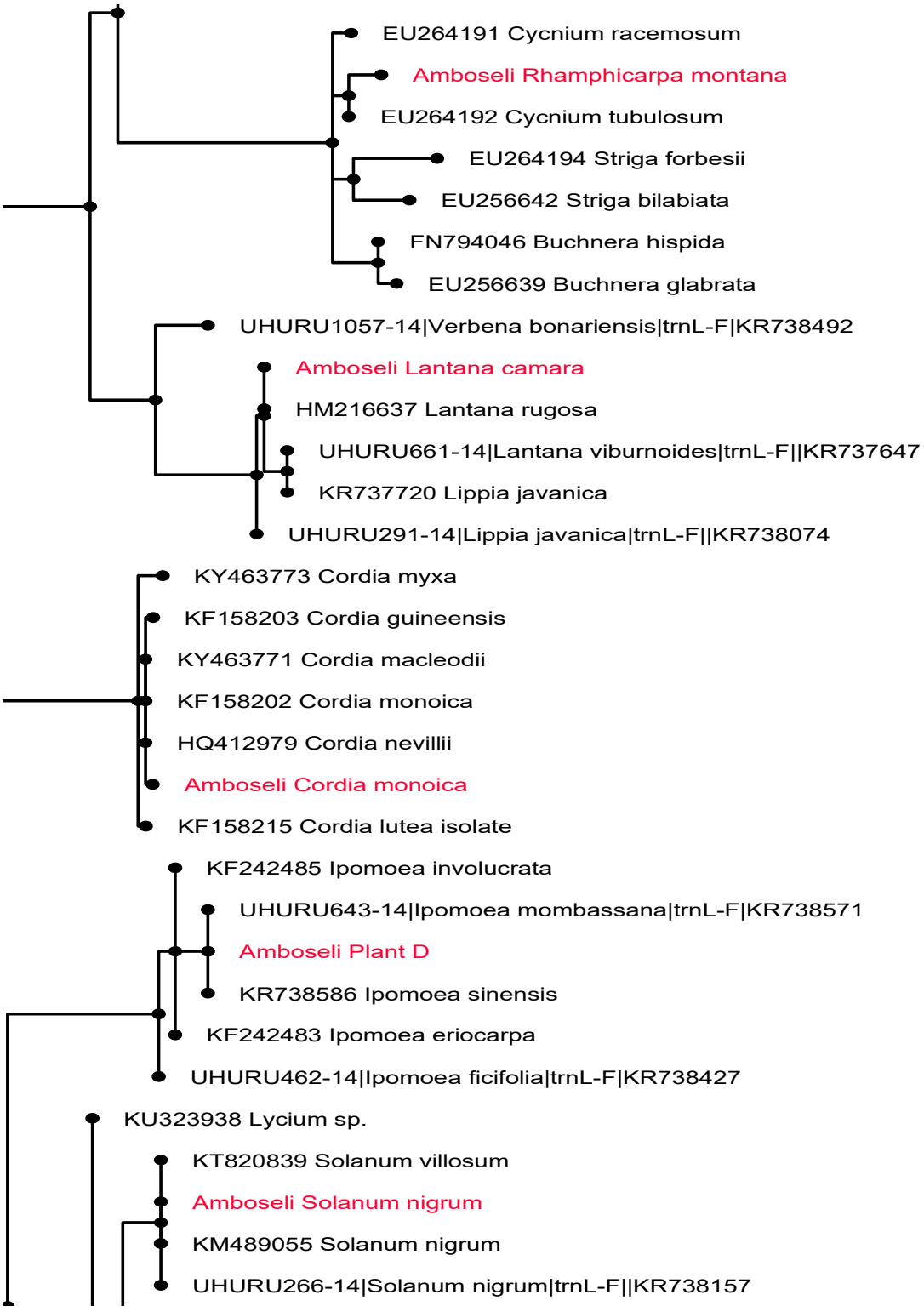
Appendix 10: Phylogenetic tree for dicotyledons based on the *trn*L (UAA) intron



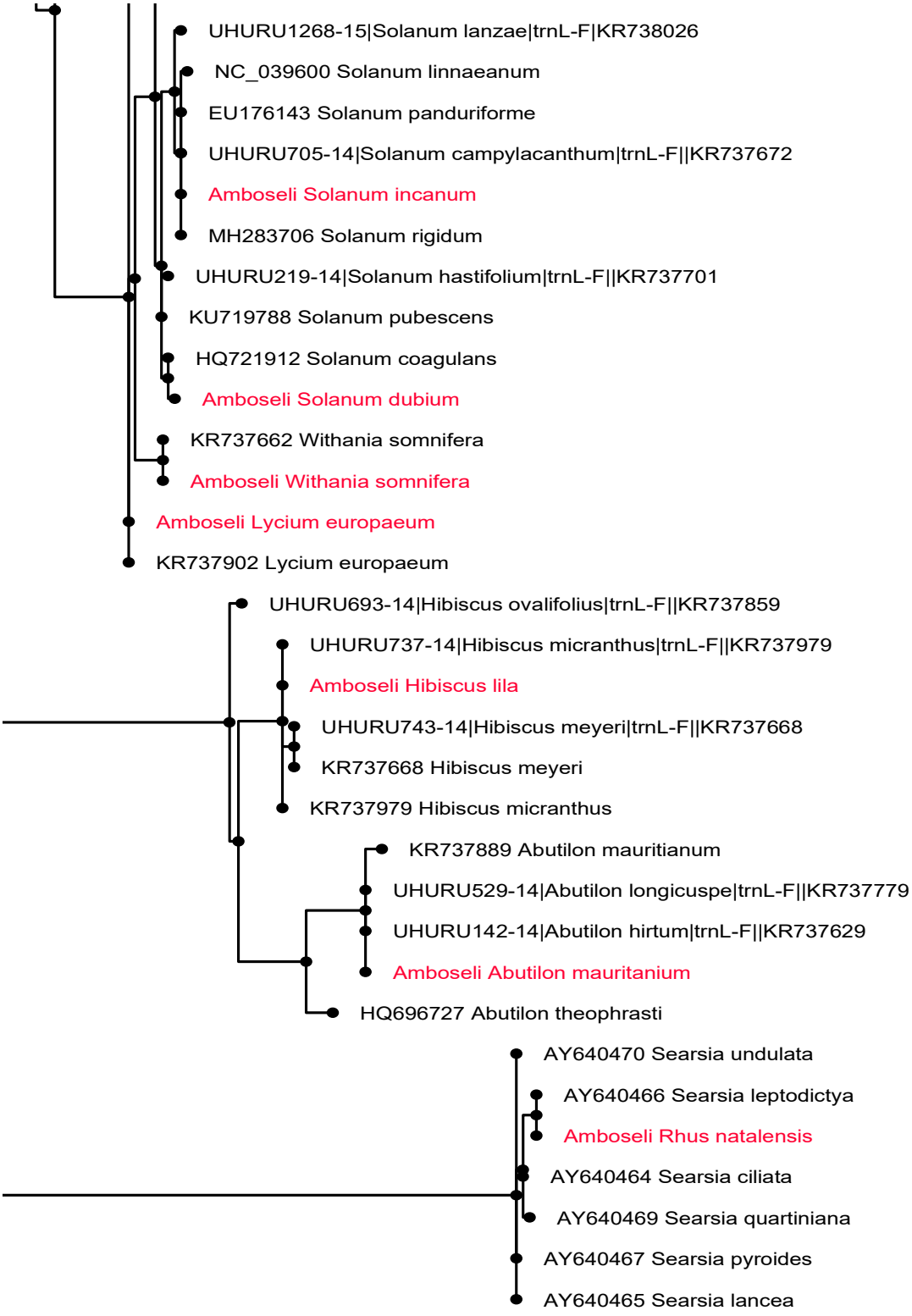
The phylogenetic tree for dicotyledons based on the *trn*L (UAA) intron.



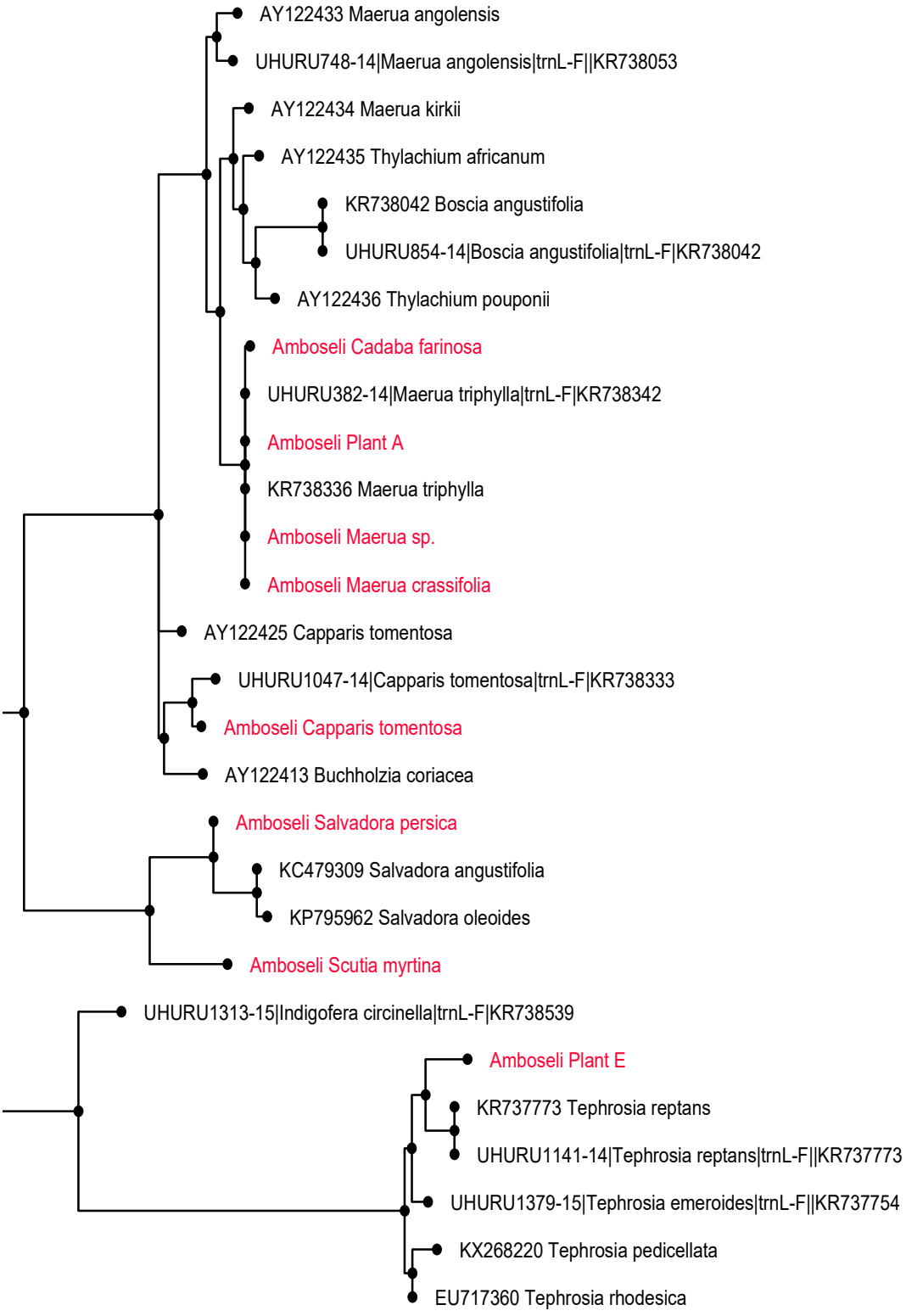
The phylogenetic tree for dicotyledons based on the *trn*L (UAA) intron.



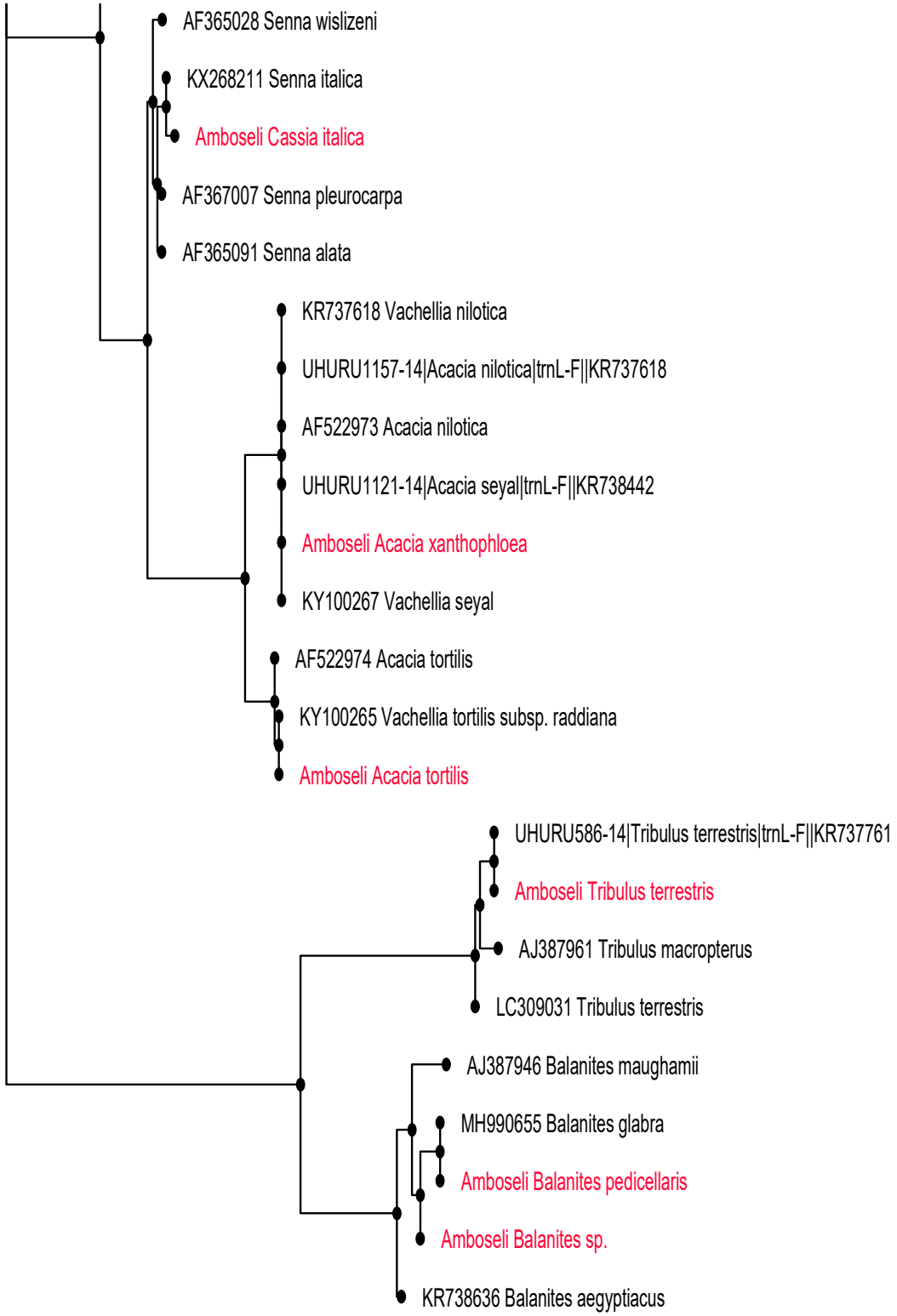
The phylogenetic tree for dicotyledons based on the *trn*L (UAA) intron.



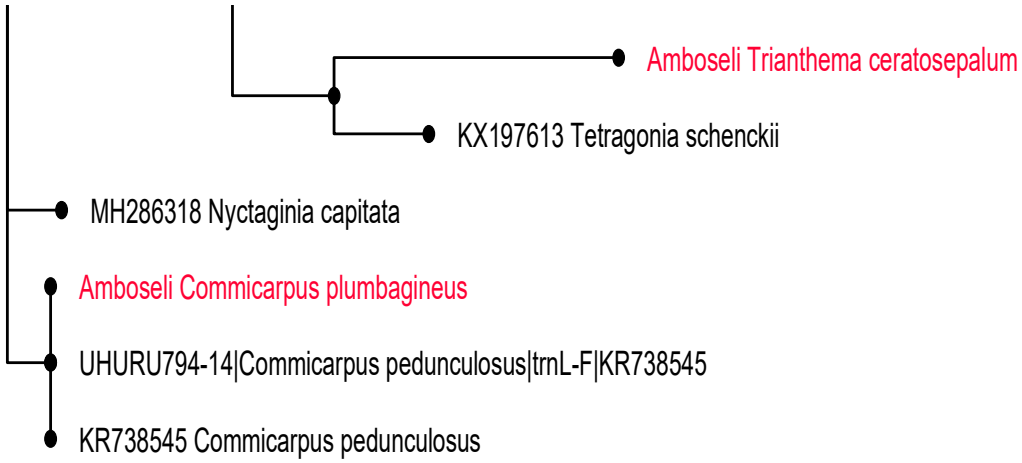
The phylogenetic tree for dicotyledons based on the *trn*L (UAA) intron.



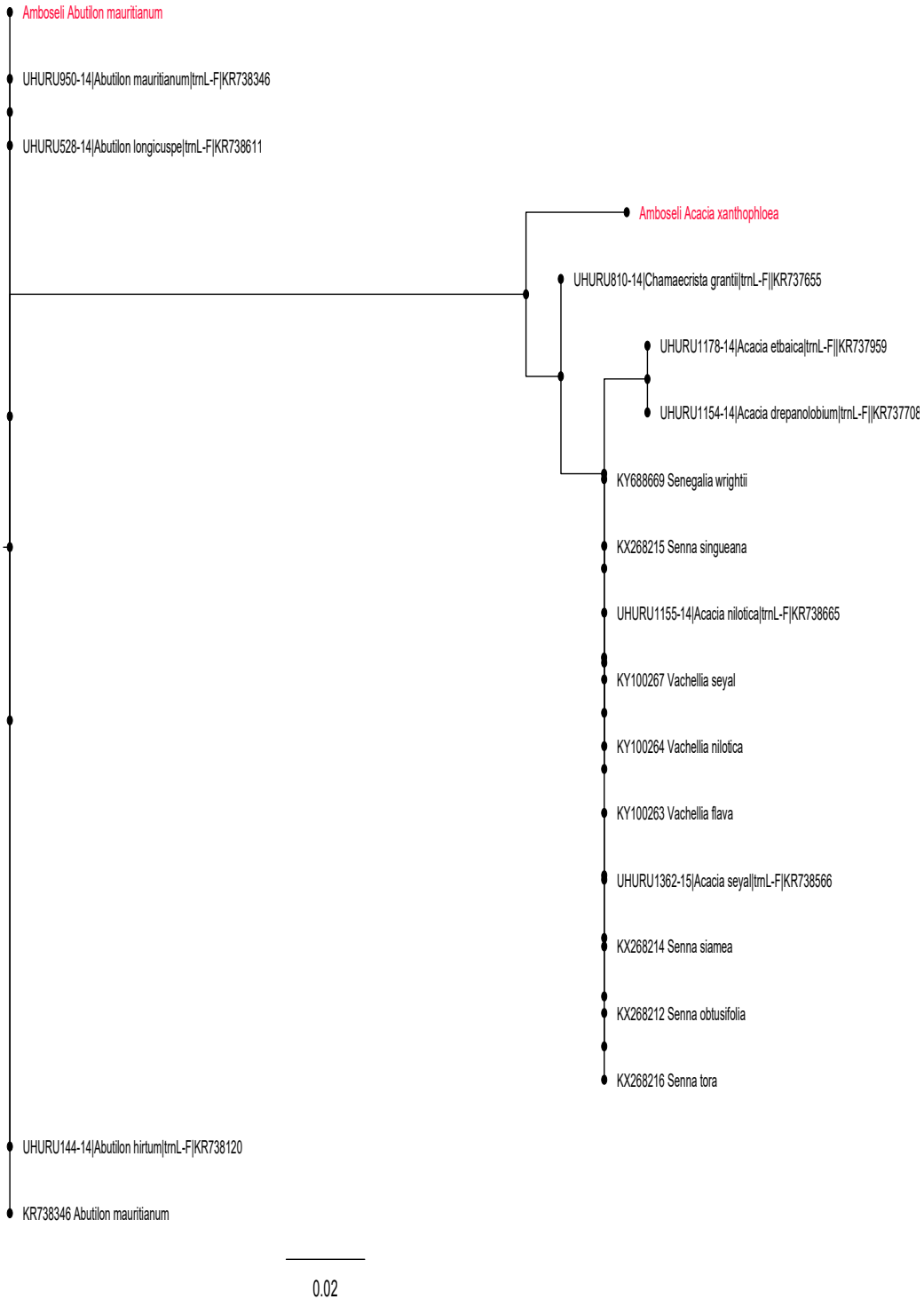
The phylogenetic tree for dicotyledons based on the *trn*L (UAA) intron.



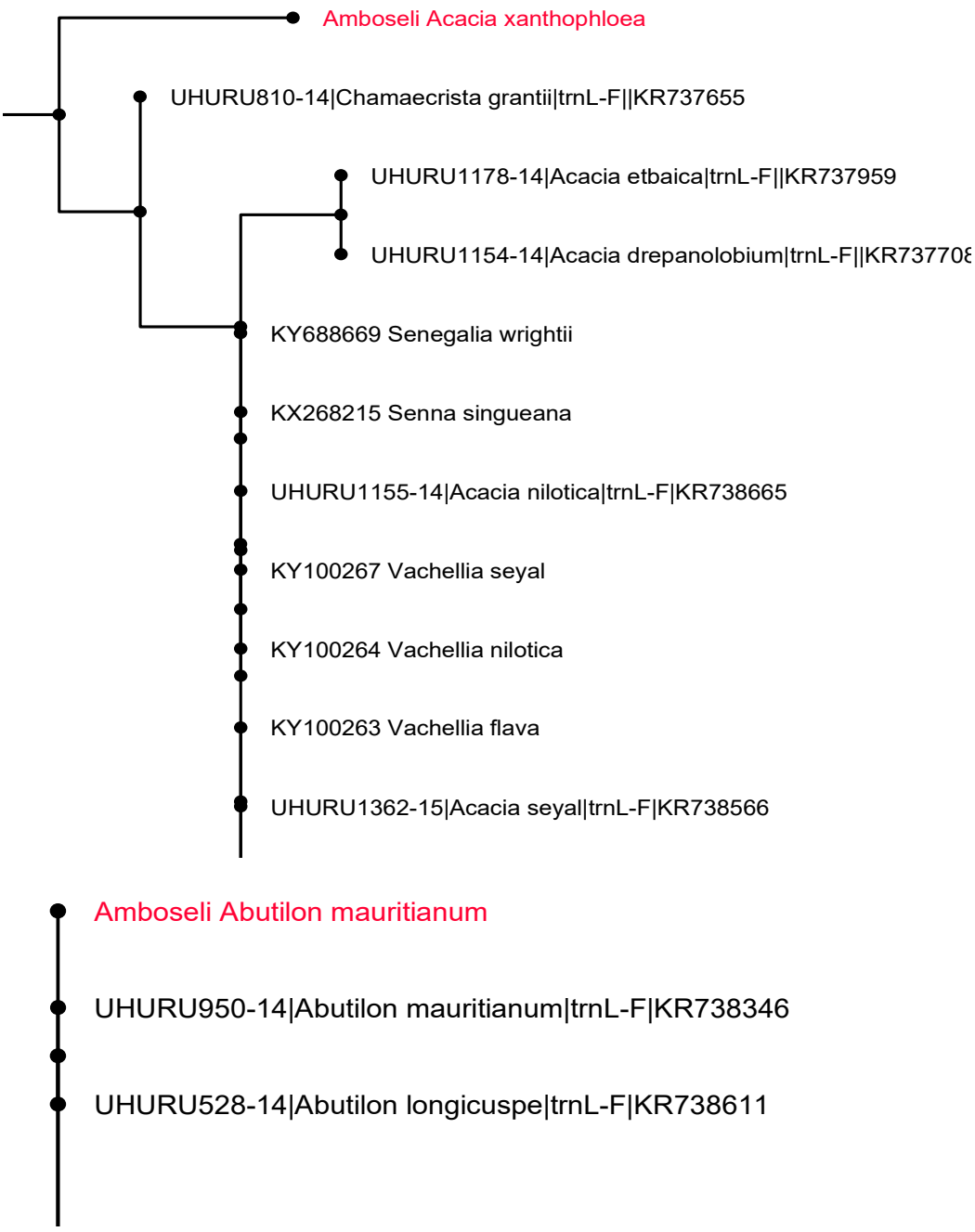
The phylogenetic tree for dicotyledons based on the *trn*L (UAA) intron.



Appendix 11: Phylogenetic tree for dicotyledons based on the *trn*L-p6 locus



The phylogenetic tree for dicotyledons based on the *trn*L-p6 locus.



Appendix 12: Phylogenetic tree for dicotyledons based on the 18S rDNA region

The phylogenetic tree for dicotyledons based on the 18S rDNA region.