ADDIS ABABA UNIVERSITY

INSTITUTE OF BIOTECHNOLOGY

Genetic Diversity Study of Napier Grass (*Cenchrus Purpureus* L.) Collections from Different Part of the World and Progeny Plants

BY: TADELECH BIZUNEH

A Thesis Submitted to the Institute of Biotechnology, Addis Ababa University, in Partial Fulfillment of the Requirements for the Degree of Master of Science in Biotechnology

ADDIS ABABA, ETHIOPIA
FEBRUARY, 2021

ADDIS ABABA UNIVERSITY
INSTITUTE OF BIOTECHNOLOGY

EXAMINERS' APPROVAL SHEET

We certify that Tadelech Bizuneh's MSc. Thesis Research entitled "**Genetic Diversity Study of Napier Grass (*Cenchrus Purpureus* L.) Collections from different part of the World and Progeny Plants"** has been conducted under our direct supervision. We recommended that the thesis has been accepted as fulfilling the requirement for the Degree of Master of Science in Biotechnology.

Advisors	**Signature**	**Date**
Meki Shehabu (PhD), ILRI	____________	_________
Tileye Feyissa (PhD, Associ. Prof.), IoB, AAU	____________	_________
Ermias Habte (PhD), ILRI	____________	_________

We, the undersigned, members of the Board of Examiners of the final open defense by Tadelech Bizuneh have read and evaluated her thesis entitle"**Genetic Diversity Study of Napier Grass (*Cenchrus Purpureus* L.) Collections from different part of the World and Progeny Plants"**and examined the candidate. This is therefore to certify that the thesis has been accepted in partial fulfillment of the requirement for the degree Master of Science in Biotechnology.

Demissachew Gade (PhD)	Signature	Date
Name of Examiner one	____________	____________
Teklehaimanot Hailesilassie (PhD)	Signature	Date
Name of Examiner two	____________	____________
Tesfaye Sisay (PhD)	Signature	Date
Director, Institute of Biotechnology	____________	____________

Final approval and acceptance of the thesis is contingent upon the submission of the final copy of the thesis to Institute of Biotechnology.

List of figures

Figure 1: Frequency distribution of PIC values of SilicoDArT and SNP markers 23

Figure 2: The Distribution of SilicoDArT (a) and SNP (b) markers 24

Figure 3: Genome wide distribution of Discriminant Analysis of Principal Components (DAPC) 25

Figure 4: a) Delta k that shows picks at K = 2 and K = 10, suggesting possible number of populations/ sub-clusters and the admixtures 28

Figure 5: Discriminant Analysis of Principal Components (DAPC) 347 Napier grass genotypes. 30

Figure 6: A dendrogram showing clusters and sub-clusters using hierarchal clustering 34

List of tables

Table 1: Progenies raised from the 13 Napier grass accessions. 20

Table 2: List of genotypes under different K groups from DAPC 31

Table 3: The list of genotypes as detected by hierarchal clustering analysis 35

Table 4: The Nei's genetic distance among the ten sub-clusters of Napier grass 37

Table 5: AMOVA among the ten sub-clusters (sub-populations) of Napier grass genotypes 38

Table 6: Mean values of Fst for the ten sub-clusters detected by STRUCTURE analysis 38

Table 7: Divergence based on allele-frequency computed using the STRUCTURE software 38

List of Abbreviations

AFLP	Amplified Fragment Length Polymorphism
DArT	Diversity Array Technology
EMBRAPA	Brazilian Agricultural Research Corporation
EST	Expressed Sequence Tag
ICRISAT	International Crop Research Institute for the Semi_Arid Tropics
GenAlEx	Genetic Analysis in Excel
NGS	Next Generation Sequencing
NTSYSpc	Numerical Taxonomy System for personal computer
PCA	Principal Component Analysis
RAPD	Random Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
SNP	Single Nucleotide Polymorphism
SSR	Simple Sequence Repeat
UPGMA	Unweighted Paired Group Method Using Arithmetic Mean
UPGMC	Unweighted Paired Group Method Using Centroid
ILRI	International Livestock Research Institute

Acknowledgements

First of all, I thank the almighty GOD and his Mother Virgin Mariam whose blessings have enabled me to accomplish my thesis work successfully. Next, it is my proud privilege to express my sincere thanks and deep sense of gratitude to my advisors Dr. Meki Shehabu, Dr. Tileye Feyissa and Dr. Ermias Habte for their unreserved guidance, encouragement, invaluable advices, and relentless corrections of this thesis from preparation of the proposal and the conceptualization of its structure up to the final accomplishment of the thesis. Their constant encouragement, friendly approach, motivation, and professional support throughout my study were indeed special. I am also very grateful to Dr. Meki Shehabu for his follow up in each and every procedure was so special and unforgettable.

I also extend my cheerful thanks to Mr. Yilkal Assefa for his brotherly approach, advices, help and guidance during the laboratory work.

Moreover, I would like to express my deep gratitude to international livestock research institute (ILRI), especially the feed and forage development program (FFD) headed by Dr. Chris Stephen Jones, for hosting me as a graduate fellow and funding the project. I am grateful to Dr. Chris Stephen Jones for reading, correcting, commenting, and suggesting in this thesis. Next, I would like to extend my thanks to Integrated Genotyping Service and Support (IGSS), for providing me the genotyping (sequence) service.

Finally, I want to express my thanks to my home institute Ethiopian Institute of Agricultural Research for giving me study leave and Institute of Biotechnology, Addis Ababa University for accepting me to the Graduate program.

Tiltle: Genetic Diversity Study of Napier Grass (*Cenchrus purpureus* L.) Collections from Different Part of the World and Progeny Plants

By: Tadelech Bizuneh

Year: 2021, Addis Ababa University

Abstract

There are limited studies on characterization of Napier grass forage through the application of currently available advanced molecular techniques. Furthermore, ILRI genebank held collection of this forage grass from different part of the world with little information. Therefore, the objective of this study was to evaluate the extent of the genetic diversity of Napier grass accessions collected from different parts of the world and progeny plants raised from naturally produced seeds. The genomic DNA was extracted from leaves of 347 Napier grass genotypes and genotyped by the DArTseq platform. A total of 96,454 Silico DArT and 96,321 SNP markers were generated, of which highly informative 1001 SNP markers were selected for diversity analysis after screening. The average polymorphic information content (PIC) values of Silico DArT and SNP markers were 0.21 and 0.15 and Average heterozygosity of 0.26 and 0.18 respectively. Two major groups and ten sub-clusters were identified from STRUCTURE, Hierarchal clustering population stratification and diversity analysis. The sub-clusters obtained by cluster analysis were also confirmed through discriminant analysis of principal component (DAPC). Analysis of molecular variance (AMOVA was highly significant (P<0.001) with 60% variation among the populations. The mean values of fixation index (Fst) per cluster ranged from 0.3398 in cluster VI that consist progeny plants to 0.7621 in cluster VII that consist most of the ILRI collections and the largest divergence (0.3777) was also between sub-cluster IV and VII. All these parameters showed the presence of high diversity and genetic differentiation among the assayed genotypes. This result indicates the potential of progenies and introduced genotypes for being used as parental plant in future hetrotic breeding and suitability of the studied population in Molecular genetic studies.

Key word: *Cenchrus purpureus*, DArTseq markers, Differentiation, Divergence, Progeny plants

Contents

List of figures iii

List of tables iv

List of Abbreviations v

Acknowledgements vi

Abstract vii

1 Introduction 1

1.2. Objectives of the study 3

Specific objectives 3

2 LITERATURE REVIEW 4

2.1 Origin and distribution of Napier grass 4

2.2 Taxonomy and botanical descriptions of Napier grass 4

2.2.1 Taxonomy 4

2.2.2 Botanical descriptions and ecology of Napier grass 4

2.3 Mode of propagation/reproduction 5

2.3.1 Sexual propagation of Napier grass 5

2.3.2 Asexual propagation of Napier grass 7

2.4 Economic value of Napier grass 7

2.4.1 Challenges of Napier grass cultivation 8

2.5 Genetic diversity studies of Napier grass 9

2.6 Genetic markers and their applications for genetic diversity analysis in Napier grass 9

2.6.1 Morphological markers 10

2.6.2 Biochemical marker /protein based molecular marker 10

2.6.3 DNA based molecular markers 10

2.6.4 DArTseq markers 13

2.7 Parameters for genetic diversity analysis 14

2.7.1 Multivariate statistics to estimate genetic diversity 16

2.8 Some statistical tools for genetic diversity analysis 17

3 Materials and Methods 20

3.1 Plant material 20

3.2 DNA extraction and genotyping 21

3.3 Data analysis 21

4 Results 23
4.1 Diversity and genome-wide distribution of the DArTseq markers 23
4.1.1 Marker selection for Diversity study 25
4.2 Genetic diversity in Napier grass populations 26
4.2.1 Genetic diversity revealed by the Bayesian algorithm in STRUCTURE software 26
4.2.2 Genetic diversity revealed by the Discriminant Analysis of Principal Components (DAPC) 29
4.2.3 Genetic diversity revealed by the hierarchal cluster analysis 33
4.3 Population differentiation and divergence analysis 37
5 Discussion 39
5.1 Marker diversity and genome-wide distribution 39
5.2 Diversity among Napier grass 40
5.2.1 Population differentiation and genetic divergence 41
6 Conclusions and Recommendations 42
6.1 Conclusions 42
6.2 Recommendations 42
7 References 43
Appendices 54
Appendix 1: (Supplementary table 1, S1 54
Appendix 2: (Supplementary table 2, S2) 66
Appendix 3: (Supplementary table 3, S3) 69
Appendix 4: loading plot 73
Appendix 5: Compo plot that shows membership probabilities of individual genotypes from DPAC analysis 74

1 Introduction

Napier grass (*Cenchrus purpureus (*Schumach) *Morrone* L.syn. *Pennisetum purpureum Schumach.)* or elephant grass is a monocot perennial grass that belongs to the family *Poaceae* (grass family) and genus *Cenchrus*. The genus *Cenchrus* is very diverse consisting of a heterogeneous group of species with different basic chromosome number and ploidy level. Among these species Napier grass (*Cenchrus purpureus*) is an allotetraploid species with a basic chromosome number of 7 (2n =4x=28) (Singh and Obeng, 2013). Napier grass is widely cultivated as a forage crop in the tropical and subtropical regions of the world. It is mainly used as animal feed due to its ease of establishment and management, high palatability and high dry matter that can provide up to 78 tons of dry matter/ ha/ year (Rj *et al.,* 2014; Alemayehu Teressa *et al.,* 2017; Oliveira *et al.,* 2017). In addition, Napier grass has the potential to produce biofuels such as alcohol, ethanol and butanol, and methane since it has high cellulose content that can be used as carbon (energy) source (Yasuda *et al.,* 2013; Romero *et al.,* 2019; Roslan *et al.,* 2020). It is also, considered as one of the exceptional potential phytoremediator plant to tackle heavy metal and chemical environmental pollutions due to its rapid growth rate, ease of establishment, low cost of management system and high level of biomass that can withstand and enable it to accumulate large amount of pollutants (contaminants) in their system; ability to grow on different environmental situations even on poor soil condition (Lotfy and Mostafa, 2014; Tananonchai and Sampanpanish, 2018; Roslan *et al.,* 2020). According to Yang *et al.* (2020) through the use of Liquid extraction method that directly remove heavy metals from plants; the content of heavy metal in Napier grass was lowered (detoxified) while the crude protein was retained to be used safely as animal feed or energy production materials. Moreover, some Napier grass cultivars were identified as potential plant in “push-pull insect management strategy” and used to trap African stem borer, *Busseola fǘsca* Fuller (*Lepidoptera*: *Noctuidae*) an insect that causes production loss in maize and sorghum (Khan *et al.,* 2007).

The international livestock research institute (ILRI) forage genebank holds more than 130 Napier grass accessions collected from different parts of the world. These collection consist from different sources where 60 accessions are ILRI’s in trust collection which represents a diverse set of genotypes assembled from a range of environments and origins, 8 accessions are hybrids of *P. purpureum* × *P. glaucum* crosses and the remaining 62 accessions were introduced to ILRI from different institutions, including the Brazilian Agricultural Research Corporation (EMBRAPA), the Crop Genetics and

Breeding Research Unit, USDA–ARS, Tifton, Georgia and International Crop Research Institute for the Semi-Arid Tropics (ICRISAT), India. These collections were maintained in-situ at the Bishoftu and Ziway (Batu) sites in Ethiopia (Alemayehu Teressa *et al.*, 2017). The collection is characterized by a high amount of genetic diversity and is very variable in terms of agronomic and morphological characters (Meki Shehabu *et al.*, 2019). Though the collections represent diverse sets of genotypes that are variable in genetic and phenotypic traits, still the diversity and population size is very limiting in selecting different traits for East Africa's different agroecological zones. Furthermore, study by Meki Shehabu *et al.* (2019) indicated presence of long haplotype blocks and less linkage disequilibrium (LD) decay in ILRI entrust collections which were maintained via vegetative propagation than the EMBRAPA materials that passed through active breeding program suggesting breeding the ILRI collection would increase the existing genetic variability. Thus, developing breeding population would be utilized for molecular genetic studies such as quantitative trait loci (QTL) mapping and genetic diversity analysis. Introducing new materials from international institutes would be good and easiest way to increase the diversity, even though the use of the introduced materials is restricted by research compliances and copy rights. The other way to increase the diversity is by crossing most genetically distant genotypes and analysing the progenies to identify unique genotypes.

The application of currently available advanced molecular tools and characterization of this forage crop is limited. Utilization of molecular technologies is essential to assess and identify the variability found in this forage crop accurately in a short time and can support improvement and conservation efforts. But so far, a handful of publications that are based on the use of low density molecular markers mostly for diversity analysis (Azevedo *et al.*, 2012; Wanjala *et al.*, 2013; Paudel *et al.*, 2018) and a couple of publications (Zhou *et al.*, 2018; Meki Shehabu *et al.*, 2019) that are based on the advanced sequencing technologies have been produced on Napier grass genotyping. Zhou *et al.* (2018) used Illumina-based sequencing technology to develop about 50 EST-SSR markers and 6 SNP markers through transcriptome analysis to facilitate the genetic diversity study in Napier grass. There is also a recent report by Meki Shehabu *et al.*(2019) that used GBS method of the DArTseq platform for the development of genome-wide sequence-based molecular markers (dominant (SilicoDArT) and co-dominant (SNP) markers) for 105 Napier grass accessions.

The current study evaluated the level and patterns of genetic diversity in Napier grass accessions and progenies maintained in the ILRI forage gene-bank using genome-wide markers from the DArT-seq platform to generate baseline information for breeding, conservation and its future utilization.

1.2. Objectives of the study

General objective

To evaluate the level and patterns of genetic diversity in Napier grass accessions and progeny plants maintained in the ILRI forage gene-bank using genome-wide markers from the DArT-seq platform to generate baseline information for breeding, conservation and its wise use.

Specific objectives

- To assess genetic variation among and within accessions of Napier grass (*Cenchrus purpureus*) collections and progeny plants.
- To identify potential genotypes that could be used in the future Napier grass breeding program
- To identify potential duplicate and divergent genotypes

2 LITERATURE REVIEW

2.1 Origin and distribution of Napier grass

Napier grass (*Cenchrus purpureus* (Schumach.) Morrone syn. *Pennisetum purpureum* Schumach.), commonly called elephant grass has originated from tropical region of sub-Saharan Africa (Clayton *et al.,* 2013) and distributed as forage crop into most tropical and sub-tropical regions all over the world. It was introduced into the USA in 1913 (Burton, 1990), into Brazil from Cuba around 1920 (Daher *et al.,* 2002), into Central and South America and the West Indies in the 1950s and into Australia in the 1960s. Currently, it is naturalized to these areas and sometimes becomes invasive (CABI, 2014). It is often considered as a weed in crops, growing along roadsides, waterways, wetlands, floodplain, swamps, forest edges, disturbed areas and wastelands (Francis, 2004). Since it withstands drought, it is a pioneer species in arid lands such as the Galapagos Islands (CABI, 2014).

2.2 Taxonomy and botanical descriptions of Napier grass

2.2.1 Taxonomy

Elephant grass or Napier grass (*C. purpureus*) was originally described and classified as *Pennisetum purpureum* Schumach (Stapf and Hubbard, 1934). However, the taxon *Cenchrus purpureus* (Schumach.) Morrone was proposed in 2010 as a replacement for *Pennisetum purpureum* Schumach (Chemisquy *et al.,* 2010).

Cenchrus is the richest genus that consists about 140 species, including important cultivated species such as Napier grass, Pearl millet and Kikuyu grass (Kikuyu and Mithen, 1987). The species belonging to this genus constitute a heterogeneous assemblage with different basic chromosome numbers 5, 7, 8 and 9, varying ploidy levels from diploid to octoploid with sexual or apomictic reproductive behavior and annual, biennial or perennial life cycles (Alemayehu Teressa *et al.,* 2017).

2.2.2 Botanical descriptions and ecology of Napier grass

Elephant grass (*C. purpureum)* is a perennial and one of the highest yielding monocot C4 major tropical grasses. It is a very versatile species that can be grown under a wide range of ecological conditions and systems: dry or wet conditions, smallholder or large-scale agriculture. It is a valuable forage and very popular throughout the tropics, as cut-and-carry animal feed forage (Alemayehu Teressa *et al*., 2017).

Napier grass is a summer-growing grass that grows from sea level up to an altitude of 2100 m. It is more productive in places where temperatures range from 25 °C to 40 °C, annual rainfall over 1500 mm and on rich, deep soils, such as friable loams (Skerman and Riveros 1990; FAO, 2015). It is not able to grow below 15 °C and is sensitive to frost, though it can regrow from the stolons if the soil is not frozen (Duke, 1983). It has some level of tolerance to drought and can grow in areas where the rainfall range is 200-4000 mm (Singh *et al.,* 2013) but, it undergoes some morphological changes including leaf rolling, reducing stomatal conductance and enhanced water use efficiency to withstand water stress conditions (Alemayehu Teressa *et al.,* 2017). On the other hand, this grass doesn't tolerate flooding; it prefers well-drained soils. When there is a poor drainage system, it is best to grow it on raised beds (Göhl, 1982). But it is able to grow on poorly drained clays, with a fairly heavy texture, or excessively drained sandy soils with a pH ranging from 4.5 to 8.2 (FAO, 2015). Elephant grass is perennial full day sunlight species that can still grow under partial shade but, does not withstand complete shade under a dense tree canopy (Francis, 2004).

Morphologically it is described as a robust, grass with a vigorous root system that penetrates deep into the soil, developing from the nodes of its rhizomes; forms dense thick clumps, up to 1 m across. The culms are coarse, perennial, and up to 4-7 m in height, branched overhead. The leaves are flat, linear, and hairy at the base and up to 100-120 cm length and 1-5 cm width, with a bluish-green color, leaf margin is finely toothed and leaf blade has a prominent midrib (Singh *et al.,* 2013). The inflorescence is a stiff terminal bristly spike, up to 15-20 cm in length, yellow-brown to purplish in color. The spikelets are 4-6 mm long, surrounded by 2 cm long plumose bristles and arranged around a hairy axis that fall at maturity. There is little or no seed formation due to the grass's self-incompatibility and exogamous nature. When seeds are present, they are small (3,000,000 seed/kg) (Mannetje, 1992; Francis, 2004).

2.3 Mode of propagation/reproduction

2.3.1 Sexual propagation of Napier grass

The sexual propagation of this grass is limited due to their cross-pollination nature which results in seeds of a mixed lot, thus the plants produced from them will not be uniform and their performance is also unpredictable. In addition, seeds germinate poorly and seedlings are weak even when it grows fully, poor seed-setting and shattering make seed availability a problem. Still, most of the ILRI Napier grass accessions growing in the Bishoftu and Ziway sites produce seeds, which need further investigation whether the seeds have been produced through apomixis or sexual reproduction. A preliminary genetic diversity analysis among progeny plants raised from seeds suggested that sexual reproduction is possibly the major mechanism for the seeds production. In addition, Napier grass is naturally cross-compatible with pearl millet (*Pennisetum glaucum*, 2n = 2x = 14) and their crosses resulted in production of triploid hybrids, which are sterile and can be propagated vegetatively by means of stem cuttings (Alemayehu Teressa *et al*., 2017).

According to Souza *et al*. (2019), about 95% of Napier grass seeds predominantly derived from cross-fertilization and the rate outcrossing in their study reveals it is an allogamous grass with multilocus (tm) and unilocus outcrossing (rates (ts) of 0.953 and 0.895, respectively). Their result is also consistence with the previous reports by Harris *et al.* (2010); Azevedo *et al. (*2012) which described Napier grass as allogamous species with high heterozygosity since outcrossing guarantees genetic variability, and thus new combinations of alleles within a species.

Napier grass, being strictly out breeder (95%), the seeds they set show low germination index and the resulting plants have low vigor (Pereira *et al*., 2010; Souza *et al*., 2019) since most of them are self-incompatible and produce few viable seeds when self- fertilized (Hanna *et al.*, 2004). Self-incompatibility varies greatly among accessions and is directly affected by environmental factors such as temperature, humidity, different environments and artificial pollinating techniques (Souza *et al*., 2019).

In addition, protogyny, which affects the production of sexual seeds, has already been described in this species. According to Pereira *et al*. (2010), stigmas are receptive between 3 to 5 days prior to the pollen grains release which makes fertilization within the same inflorescence unfeasible. Difficulties in self-fertilization limit the development of hybrids which offer opportunities for greater uniformity, higher selection intensities, absolute parental control and maximum exploitation of heterosis in Napier grass (Canto *et al.*, 2016).

2.3.2 Asexual propagation of Napier grass

Asexual or vegetative propagation of Napier grass is the most common method of propagation. Asexual propagation of Napier grass through stem chopping consists of at least 3 nodes in which two of these nodes are buried in the soil. The planting row width ranges from 50 to 200 cm and distance within rows ranges from 50 to 100 cm (Mannetje, 1992). After planting, elephant grass grows vigorously and can be as tall as 4 m in three months (Skerman *et al.,* 1990). It is fast-growing and has a high annual productivity that depends on the climate, especially temperature and rainfall (Artus-Poliakoff *et al.,* 1991; Aroeira *et al.,* 1999). Elephant grass requires high levels of fertilizer and regular water supply (Mannetje, 1992) and the yields range from 20 to 80 t DM/ha/year under high fertilizer inputs (Skerman *et al.,* 1990; Francis, 2004), while 2-10 t DM/ha/year when there is no, or inadequate, fertilizer input (Bogdan, 1977). Cuttings can be made at 45-90 days intervals, depending on location (FAO, 2015).

2.4 Economic value of Napier grass

Napier grass has a lot of economic importance being primarily used as a fodder crop for small scale farmers since it grows with little nutrient supply and ease of management. The perennial nature of the grass makes it the first choice as a forage crop. It is also used as an alternative bioenergy source, as a phytoremediator plant, and as pest management in push and pull management system (Roslan *et al.*, 2020).

A number of traits, including high dry matter production, ease of establishment and regeneration, persistence, and enhanced water use efficiency make Napier grass the primary forage of choice by small-scale dairy farmers (Nymbati *et al.,* 2010). Napier grass follows C4 photosynthetic pathway and considered to have a competitive advantage over C3 grass species when grown in tropical and sub-tropical regions (Taylor *et al.,* 2011). It also has the capacity to reduce shoot dry matter and maximize carbon assimilation during times of water stress, making it a desirable forage crop in areas prone to droughts (Cardoso *et al.,* 2015). It is often fed fresh in cut-carry systems. It can also be cut into pieces prior to feeding to reduce coarseness of leaves and stems by the animal. Moreover, wilting after chopping in the sun for several hours reduces moisture, facilitates rumination, stimulates appetite and thus improves forage utilization (Moran, 2011).

Napier grass (*C. purpureum)* genotypes were also evaluated by (de Morais *et al.,* 2009) for bioenergy production and they indicated that genotypes such as Cameroon, CNPGL F 06-3 and Bag 02 promising for bioenergy production purposes. Other studies in Brazil's semiarid area using gypsum by (dos Santos *et al.,* 2015) also state that elephant grass mainly of Cameroon and Gramafante varieties as great potential energy crop. Also according to (Favare *et al.*, 2019), Elephant grass is excellent alternative for bioenergy production mainly, due to its high percentage of stem and dry matter yield.

On other hand, eight Napier grass (*C. purpureum*) varieties, were evaluated for their potential role as trap plants in the management of the African stemborer, *Busseola fusca* Füller (Lepidoptera: Noctuidae) in a push–pull strategy (Khan *et al.,* 2007). They concluded Napier grass cultivar that called Bana had potential for use as a trap plant in the management of *B. fusca* in a 'push–pull' strategy, but the effectiveness of such a strategy would strictly depend on proper establishment and management of these companion plants.

In addition, Napier grass is also considered as one of the exceptional phytoremediator plant to tackle environmental pollutions due to its fast growth rate, high level of Biomass yield, ease of establishment, low management cost and tolerance to pollutions; that enable it to accumulate large amount of pollutants (contaminants) such as cobalt, chromium, cadmium and EDTA in their system and able to degrade it (Lotfy and Mostafa, 2014; Tananonchai and Sampanpanish, 2018; Roslan *et al.,* 2020).

2.4.1 Challenges of Napier grass cultivation

Napier grass as a major fodder for a dairy farm in East and central Africa, grown by more than 70 percent of smallholder dairy farmers in Kenya (Mulaa *et al.,* 2013); Uganda (Kabirizi *et al.,* 2007) and Tanzania (Pallangayo *et al.,* 2008). It constitutes between 40 to 80% of the forage for smallholder dairy farmers (Staal *et al.*, 1997). In general, ease of establishment, management and regeneration, and its enhanced water use efficiency make it primary forage of choice in the regions of tropics and subtropics for smallholder dairy farmers (Alemayehu Taressa *et al.,* 2017). However, the production of this forage crop is currently facing challenges from abiotic and biotic conditions such as climate change, Napier stunt and head smut diseases, caused by phytoplasma and a fungus *Ustilago kameruniensis*, which spread by wind, farm tools, infected plants, water, and animal manure, respectively, have caused forage yield reduction of up to 90% (Mulaa *et al.,* 2013), and these biotic factors are currently the biggest threats to forage production and hence dairy sector in the East and Central Africa region.

Thus, information on the range of diversity of the available genetic resources of this important forage crop will help in the development of varieties that can overcome this production and productivity constraints.

2.5 Genetic diversity studies of Napier grass

Precise assessment of the level of genetic diversity can be invaluable in plant breeding for diverse applications including: analysis of genetic variability in cultivars, identifying diverse parental combinations to create progenies with maximum genetic variability for further selection, and introgressing desirable genes from diverse germplasm into the available genetic base (Mohammadi and Prasanna, 2003). Therefore, a major focus of research in molecular genetics has been to determine the amount of genetic variation in populations and describing the possible mechanisms of maintaining such variability in meeting new environmental challenges (Weir, 1996).

Specifically, diversity information that can be derived from molecular analyses of Napier grass germplasm will help to determine the degree of relationships of the different germplasm within and among breeding populations and the germplasms held at ILRI genebank, used in future association study to identify outstanding morphological and agronomic characters present in the germplasm, Also, it will provide the basic data for the effective utilization of the germplasm in the grass breeding programs for its improvement and conservation.

2.6 Genetic markers and their applications for genetic diversity analysis in Napier grass

Analysis of genetic diversity within and among populations involves the use of different genetic markers. Now, genetic markers are used in both basic plant research and plant breeding programs to characterize plant germplasm for gene isolation, marker-assisted introgression of favorable alleles, to produce improved varieties (Henry, 2001), and to obtain information about the genetic variation of populations for conservation purposes. There are three main classes of genetic markers for genetic diversity analysis. These are morphological markers, biochemical or protein markers, and DNA based molecular markers and are briefly discussed below.

2.6.1 Morphological markers

Qualitative traits are usually controlled by a single locus and their expression is the same over a range of environmental condition. These traits can be used as marker as they are usually visually described. These traits include seed shape, seed color and flower color (Bagali *et al.,* 2010). The advantage of using these markers for genetic diversity analysis within and among populations is that it does not require any sophisticated equipment and it is the most direct method. However, they are limited in number (Mondini *et al.,* 2009). In Napier grass morphological markers were used by Van De Wouw *et al.* (1999) to assess variation among 53 accessions and indicated that some accession could not be distinguished from each other by any of the characters observed. As general, morphological markers are not as such a reliable marker to measure genetic differences related to productivity since most of the yield related traits are quantitate trait that are controlled by muliti-loci and gene expression is influenced by different factors including environmental factors.

2.6.2 Biochemical marker /protein based molecular marker

To overcome the limitation of morphological markers, protein-based biochemical markers and DNA based markers have been developed (De Vicente and Fulton, 2003). Biochemical marker-based analysis of genetic diversity is the separation of proteins into specific banding patterns. It is a fast method that requires only small amounts of isozymes or protein. Napier grass polymorphism based on isozymes and total proteins (TP) was assessed by Bhandari *et al*. (2006) and could distinguish variations among the assed accessions. Since only a limited number of enzymes are available and not evenly distributed throughout the genome of the organism, they do not clearly show diversity at the genome level (Mondini *et al.,* 2009).

2.6.3 DNA based molecular markers

Molecular marker uses naturally occurring polymorphism in DNA sequences due to error during DNA replication, recombination or mutation (Gupta *et al*., 1999). Molecular markers are highly polymorphic and heritable, relatively simple to detect, distributed throughout the genome and completely independent of environmental conditions. It enables to detect genetic variability at any stage of plant development. Main disadvantage of this technique is it requires complex equipment and procedures. DNA based molecular markers are classified into two namely, hybridization-based marker and PCR based markers.

2.6.3.1 Restriction Fragment Length Polymorphism (RFLP)

Restriction Fragment Length Polymorphism (RFLP) is a hybridization-based marker and an important tool for plant genome mapping (Jiang, 2013). It involves the digestion of genomic DNA into fragments with specific restriction enzymes and utilizes the variation in fragment length to assess differences. Variations in fragment length arise due to change (mutation) at an enzyme recognition site, single base pair substitution, insertion of genetic materials, such as transposable element, or by tandem duplications deletions, translocations, or other rearrangements (Gupta *et al.,* 2002).

Smith *et al.* (1993) used RFLP and RAPD to study genetic diversity of Napier grass. RFLP is co-dominant locus-specific marker and is highly reproducible which can be used for different related organism since they allow synteny (conserved order of genes between related organisms). The main shortcoming of this marker includes tedious and time-consuming procedure.

2.6.3.2 PCR based markers

I. Random Amplified Polymorphic DNA (RAPD)

RAPD marker uses usually 10 nucleotides long single arbitrary primer to amplify the genomic DNA at different amplicon sizes. The difference in fragment length due to some changes in the sequences found in primer annealing sites and between the two priming sites (Williams *et al.,* 1990). As stated by (Alemayehu Teressa *et al.,* 2017) Smith and co-works were the first to use AFLP and RAPD markers to analyze diversity in Napier grass and able to clearly distinguish Napier grass accessions from its hybrid and pearl millet. The main limitation of RAPD is low reproducibility and its dominant nature (Williams *et al.,* 1990).

II. Amplified Fragment Length Polymorphism (AFLP)

This marker was originally developed for universal DNA fingerprinting analysis (Vos *et al.,* 1995). It is robust and relatively insensitive to PCR reaction conditions and highly reproducible. AFLP is used for the rapid screening of genetic diversity and intraspecific variation (Russell *et al*, 1997). Wanjala *et al*. (2013) used AFLP to study genetic diversity of Napier grass collections from east African region and ILRI gene bank collections and they were able to discriminate and group these collections efficiently. The limitations of AFLP include difficulties in interpretation of band profiles in terms of loci, allele, and dominance.

III. **Simple Sequence Repeat (SSR)**

Simple sequence repeat (SSR) markers also called microsatellite markers are very short DNA (mono, di-, tri-, tetra-, penta-, and hexanucleotide) motifs usually characterized by high degree of repetition even in thousands. They are evenly distributed throughout the genome of the organism (Singh *et al.*, 2010) and their polymorphism reflects differences in simple repetitive sequences of defined regions of the genome. Product of different sizes or lengths can be amplified with a pair of primers flanking different microsatellite regions. Typically, only single locus is amplified resulting in single or double bands depending on the homozygosity or heterozygosity of that specific locus (Singh *et al.*, 2010). In general, SSR markers are locus-specific and co-dominant, simple to operate, abundant throughout the genome, highly reproducible. It has high rates of transferability from species to species. The cross-amplification rate of microsatellite markers between species of the same genus can vary from 50 to 100% (Azevedo *et al.*, 2012).

This transferability property was assessed for closely related species of *Pennisetum glaucum* and *Pennisetum purpureum* by Azevedo *et al.* (2012) to evaluate the genetic diversity and phylogenetic relationships among Napier grass accessions using SSR markers which are originally developed for pearl millet (*P. glaucum*). According to their study, about 55.5% of the primers showed successful cross-amplification to study the diversity of Napier grass and enabled them to cluster 107 accessions into three main groups. Kandel *et al.* (2016) also studied the genetic diversity of Napier grass (*Pennisetum purpureum* Schum.) using microsatellite, Single-Nucleotide Polymorphism and Insertion-Deletion markers which are developed for Pearl Millet (*Pennisetum glaucum*). They were also able to show cross-amplification of these markers within a genus. The main limitation of this marker is the requirement for prior sequence information for the primer flanking regions to develop a pair of primers.

IV. Inter Simple Sequence Repeat (ISSR)

Inter Simple Sequence Repeat marker was introduced in 1994 to asses genetic variation below the species level; mainly for studying population structure and differentiation of cultivated plants (Gupta *et al.*, 1994).This marker technique involves the amplification of DNA segment present between two identical microsatellite repeat regions oriented in opposite direction (Joshi *et al.*, 2000).

In Napier grass (Babu *et al.,* 2009) used RAPD and ISSR markers for analysis of genetic diversity in Napier grass to detect DNA polymorphism among thirty Napier grass collections from wide geographical distribution and able to identify the variation among these assayed genotypes.

V. Single-nucleotide polymorphism (SNP)

Through SNP markers differences in DNA sequence among genotypes are identified through an expensive and laborious DNA sequencing process. At first, SNPs revolutionize biomedicine, but since the technology depends on intensive genomic sequencing and a high cost of analysis, it was very expensive for applications in agriculture or basic research. However, the development of a solid-state, open-platform method for DNA polymorphism analysis which is called Diversity Array Technology, offered a low-cost high-throughput, robust system with minimal DNA sample requirement capable of providing high quality genome coverage information even in organisms without any DNA sequence information such as Napier grass (Meki shehabu *et al.*, 2019).

2.6.4 DArTseq markers

Some of the above mentioned marker systems such as SSR and ISSR depends on prior sequence information while most are gel-based assay system; that result in high cost per assay and low throughput. The DArTseq technology uses a combination of genome complexity reduction using restriction enzymes and next generation sequencing (NGS) and produces high-density genome-wide dominant (SilicoDArT) and co-dominant (SNP) markers. This technology offers Napier grass breeding programs and genetic resources conservation at the ILRI forage gene bank an alternative and excellent approach to whole-genome profiling (Meki shehabu *et al.,* 2019) as in any other species.

2.7 Parameters for genetic diversity analysis

Genetic diversity can be measured using different statistical methods. It can be in terms of the Coefficient of Correlation, genetic distance and heterozygosity. Coefficient of Correlation is defined as the probability that alleles of two individuals are identical by descent or by state. The value of coefficient of correlation ranges from zero, when that genotype is completely unrelated, to one, when the two individuals have many alleles in common or almost identical (Martin *et al.,* 1991). This Coefficient of Correlation is used to cluster genotypes in to similar groups (Bered *et al.,* 2002). Coefficient of correlation (rxy) can be computed for all pairwise combinations of genotypes from pedigree information using formula given by Falconer and Mackay (1996):

$$r\text{xy} = 2\text{f}\,\text{xy}|\sqrt{(1+Fx)(1+Fy)}$$

Where:- fxy= a coefficient of co-ancestry, Fx and Fy = inbreeding coefficients of X and Y, respectively.

Genetic distance

Genetic distance is used to measure the genetic divergence between species or populations within a species, whether the distance measures degree of differentiation or time from common ancestor. Many types of estimation of genetic distance are available, appropriate choice of a genetic distance measure, on the basis of the type of the variable and the scale of measurement is an important component in analysis of genetic diversity among genotypes. For the first time Nei defined genetic distance as the difference between two entities that is explained by allelic variation (Nei, 1972). Later on, it was defined in a more comprehensive way by Beumont *et al.* (1998) as any quantitative measure of genetic difference at either sequence or allele frequency level calculated between genotype individuals or populations. Currently there are different methods that are used to measure genetic diversity from molecular data. Some of them include Euclidean distance, Roger's distance Fixation index (Fst), AMOVA (analysis of molecular variance). Allelic diversity is used when genetic marker data or molecular marker data are interpreted in terms of locus/allele model. This allelic diversity may be expressed as a) the percentage of polymorphic loci (p), b) mean number of alleles per locus (n), c) total gene diversity or expected heterozygosity (He), and polymorphic information content (PIC)

(Bhanu, 2017). Percentage of polymorphic loci (p) gives an estimate of number of polymorphic loci with respect to total loci including polymorphic and monomorphic loci and is calculated as

$$p = \frac{Np}{Nt} * 100$$

N_p is the number of polymorphic loci and N_t is the number of total loci

Polymorphic information content (PIC) or indirect estimate of number of alleles per locus can be calculated as

$He = 1 - \sum_{i=1}^{n}(Pi)^2$ And $PIC = 1 - \Sigma p_i^2 - \Sigma \Sigma p_i^2 p_j^2$

P_i and p_j are the frequency of ith and jth allele at any particular locus

The PIC is a good index for genetic diversity evaluation and used to evaluate the level of gene variation, with values ranging from zero to one. PIC value > 0.5, indicates locus with high diversity; PIC value <0.25, locus with low diversity and PIC value between 0.25 and 0.5 for locus with intermediate diversity (David *et al.*, 1980).

Hardy-Weinberg Equilibrium (HWE): Hardy-Weinberg equilibrium explains that both gene and genotype frequencies will be constant from generation to subsequent next generations under some assumptions (Labate, 2000). Those assumptions include: genetic drift, mutation and natural selection do not happen and the population is closed (no gene flow) and has random mating pattern.

F-statistics (Fst): fixation indices are the measures of standardized variances in allele frequencies that detect departure from HWE caused by biased inbreeding out breeding or population subdivision and drift (Wright, 1950). The F statistics is a measure of the difference between the mean heterozygosity. Fst is considered to be the most informative statistic for examining the overall level of genetic divergence among subpopulations. Fst can be calculated as: $\mathrm{Fst} = \frac{(\mathrm{HT} - HS)}{HT}$

Where Fst is reduction in heterozygosity, H_T is average heterozygosity in population and H_S average heterozygosity in subpopulation.

The Fst value ranges between zero and one. When it is equal to zero, it means complete sharing of genetic material and when it is one, no sharing (the populations are fixed).

Analysis of Molecular Variance (AMOVA) is a method of estimating variance components within and among population directly from molecular data and testing hypotheses about differentiation (Excoffier *et al.*, 1992). AMOVA treats molecular data as a vector q_i which is a matrix of 1s and 0s, 1 indicating the presence of a marker and 0 its absence.

Euclidean distance: between pairs of vectors are calculated by subtracting the vector of one haplotype from another, according to the formula ($\mathbf{q}_j - \mathbf{q}_k$). If $\mathbf{q}_j$ and $\mathbf{q}_k$ are visualized as points in *n*-dimensional space indicated by the intersections of the values in each vector, with *n* being equal to the length of the vector, then the Euclidean distance is simply equal to the shortest distance between those two points. The distance between points p and q is the length of the line segment connecting them ($\overline{pq}$) (Aremu, 2012).

$$d(p,q) = d(q,p) = \sqrt{(q1 - p1)2} + (q2 - p2)2 + \cdots (qn - pn)2$$

$$= \surd\sum_{i=1}^{n}(qi - pi)2$$

For molecular data allele frequencies calculated and genetic distance between individual p and q can be estimated using a formula: $d(p,q) = constant(\sum_{i=1}^{n}|Xi\,p - Xiq|r)^{1/r}$

Where, X_{ip} is the frequency of allele i for individual p, n is the number of alleles per locus and r is constant based on the coefficient used. Then when r=1, genetic distance can be

$D_{1p,\,q} = 1/2\sum_{i=1}^{n}|Xip - Xiq|$

When r = 2, it is referred as Roger's (1972) measure of genetic distance (R_D) and can be calculated as:

$R_{D(}p,q) = 1/2[\sum(Xip - Xiq)2\ ]1/2$

2.7.1 Multivariate statistics to estimate genetic diversity

Genetic diversity among different varieties and/or of a species can be assessed using multivariate statistics. It provides reliable information on the real genetic distances between genotypes hence they are a tool for assessment of genetic diversity (Bhanu, 2017). Some of the multivariate techniques include:

I. Cluster analysis

These techniques depict a pattern of similarity/relatedness between genotypes based on their evolutionary relationships and group the similar ones in the same group while differentiating the others. This method is mainly based on Unweighted paired group method using arithmetic mean (UPGMA) to provide precise grouping information on breeding materials used in accordance with pedigrees and calculated results found in agreement with known heterotic groups than the other clusters (Aremu *et al.*, 2007).

II. Principal component analysis (PCA)

Principal component analysis (PCA) can be described as the quantitative type of data reduction technique. This technique transforms multi-correlated variables into different set of uncorrelated variables for further study (Bhanu, 2017). The new variables are in linear combinations with original variables. It is based on the development of characteristic values and mutually independent principal components arranged in a decreasing order of variance. The technique is most suitable when different variables have same unit and difficult for different scales. This difficulty is avoided by standardizing all the variables and to do it each variable is divided by its estimated standard deviation. This technique is a mean for further analysis not an end (Bhanu, 2017).

2.8 Some statistical tools for genetic diversity analysis

I. Numerical Taxonomy System for personal computer (NTSYSpc)

NTSYSpc is used to analyze genetic diversity from different molecular marker data and works based on similarity indices as 0, 1 matrix of genotypic data. It is used for cluster analysis, principal component analysis, principal coordinate analysis (Rohlf, 1998).

II. Genetic Analysis in Excel (GenAlEx)

GenALEx is an Excel add-in easy and user-friendly program designed for use of SSR, SNP, AFLP, allozyme, multi locus markers and sequencing DNA data in genetic diversity analysis. It works with three data type's codominant data, dominant, and geographic data. Analysis performed includes: observed and expected heterozygosity, marker index, fixation index, allelic patterns, haploid diversity by population, haploid diversity by locus, haploid disequilibrium and, Nei's genetic distance, principal component analysis, Shannon index (Bhanu, 2017).

III. Popgene

Popgene software is user-friendly package developed for the analysis of genetic diversity among and within natural populations. Codominant data, dominant and quantitative traits data are the data type it works on and performs population genetic structure analysis using the target markers/ traits, gene frequency, number of allele, polymorphic loci, gene diversity, Shannon index, homozygosity, gene flow, genetic distance (based on Nei coefficient) and dendrogram (based on UPGMA and neighbor-joining method) and neutrality (Bhanu, 2017).

IV. R statistical software

R is a free, open-source; user friendly software and programming language developed in 1995 at the University of Auckland as an environment for statistical computing and graphics (Ikaha and Gentleman, 1996). Currently, several scientific disciplines, including medicine, agriculture, soil science, and ecology preferentially use R software due to its graphical capabilities and availability of different packages that are designed for analyzing different data types.

The free software R is a standard and preferable for the analysis of genetic data, offering packages that are dedicated to population genetics (Paradis, 2010), phylogenetic (Schliep, 2011) or genome-wide association studies (Clayton and Leung, 2007). When the R software is known as a standard for the genetic data analysis, classical population genetics tools are being challenged by the increasing availability of genomic sequences since dedicated tools are needed for harnessing the large amount of information generated by currently available sequencing technologies.

I. The R package Adegenet for the multivariate analysis

The R package Adegenet is contributed and used to implement classes and functions to facilitate the multivariate analysis of genetic markers. This led to define new formal classes for genotypes (genind) or groups of genotypes (genpop), which can be used as input to multivariate methods proposed in the R software. A lot of functions are also implemented to manipulate and analyse these objects, including recent development in spatial genetics and data simulation. By assuring a good interoperability of data, adegenet contributes to making the R software a unifying platform for the analysis of genetic markers (Jombart *et al.*, 2008).

II. The R package Poppr for populations genetic analysis

The R package Poppr gives significant, accessible tools for analysis of clonal, partially clonal, and sexual populations available in one environment on all major operating systems. The capacity to analyze data for multiple populations across a user-defined hierarchy and provide novel functionality in R. Combined with R's graphing abilities, publication-ready figures is thus obtained conveniently (Kamvar *et al.*, 2014).

III. The R package Ape for Phylogenetics and Evolution Analysis

Analysis of Phylogenetic and Evolution (APE) is a package written in the R language for use in molecular evolution and phylogenetics. APE enables the manipulation of phylogenetic trees, as well as several advanced methods for phylogenetic and evolutionary analysis such as comparative and population genetic methods (Paradis *et al.,* 2004, 2019).

IV. R package synbreed

The package synbreed is developed within the synbreed project for synergistic plant and animal breeding (www.synbreed.tum.de). This package executes data processing, data analysis, and visualization since it contains a collection of functions that are embedded within the framework of a single, unified data that required for genomic analysis (Wimmer *et al.*, 2012). Moreover, its implementation is flexible with respect to different ranges of data formats.

V. Pegas R package for AMOVA analysis

The package Pegas functions data reading, writing, plotting, analyzing and manipulating allelic and haplotypic data. It is used for analysis of linkage disequilibrium, population structure (fixation index (Fst), Analysis of Molecular variance (AMOVA)) and Hardy-Weinberg Equilibrium (HWE) (Knaus and Winter, 2020).

3 Materials and Methods

3.1 Plant material

A total of 363 Napier grass accessions that comprise 60 accessions from the ILRI gene bank collection: 45 (25 Centro Nacional de Pesquisa de Gado de Leite (CNPGL), EMBRAPA's elite lines and 20 Brazilian Active Germplasm Bank of Napier Grass (BAGCE) accessions) from the Brazilian Agricultural Research Corporation (EMBRAPA), Brazil, 22 accessions from the United States Department of Agriculture (USDA), Tift, Georgia, USA, 31 accessions from ICRISAT gene-bank, Patancheru, India (supplementary table 1) and a total of 205 progeny plants raised from seeds of 13 ILRI accessions (Table.1), were included in this genetic diversity study.

From the seed bearing Napier grass accessions approximately 20 to 30 seeds were pre-germinated on agar medium containing potassium nitrate. The germinated seeds were transplanted into soil-filled pots and maintained in the screen house until they produce 3 to 4 leaves for sample collection. The soil was in composition of 3:2:1 ratio of soil, compost and sand. On average, about 16 progeny plants were taken from each of the 13 accessions that produced seeds to determine the level of diversity within and among progeny plants. On the other hand, recently introduced accessions from USA and India were propagated through stem cuttings. As indicated by Mannetje (1992) Napier grass stems were chopped into fragments consisting of at least 3 nodes in which two of these nodes were buried in the soil at an angle of 45^0 during planting. The seedlings that were raised in greenhouse were watered two to three times a day.

Table 1**:** Progenies raised from the 13 Napier grass accessions. The detail of each progeny and accessions used in the study is shown in supplementary table S1.

Acc. No.	**Species**	**Progenies**	**total No. of progenies**
ILRI_1026	*Cenchrus purpureum*	ILRI_NS_1	15
ILRI_16789	*Cenchrus purpureum*	ILRI_NS_2	7
ILRI_16839	*Cenchrus purpureum*	ILRI_NS_3	20
ILRI_16783	*Cenchrus purpureum*	ILRI_NS_4	17

ILRI_14983	*Cenchrus purpureum*	ILRI_NS_5	17
ILRI_16835	*Cenchrus purpureum x P. glaucum*	ILRI_NS_6	19
ILRI_16837	*Cenchrus purpureum x P. glaucum*	ILRI_NS_7	18
ILRI_16803	*Cenchrus purpureum*	ILRI_NS_8	14
ILRI_16821	*Cenchrus purpureum*	ILRI_NS_9	10
ILRI_16818	*Cenchrus purpureum*	ILRI_NS_10	13
ILRI_16810	*Cenchrus purpureum*	ILRI_NS_11	19
ILRI_14984	*Cenchrus purpureum*	ILRI_NS_12	18
ILRI_16790	*Cenchrus purpureum*	ILRI_NS_13	18

3.2 DNA extraction and genotyping

Young leaf samples were collected from each individual plant using 2ml eppendof tube on ice and on -80^0c freezer before, freeze-dried for about 48 hours and ground into fine powder using a tissue grinder (tissue lyser). Genomic DNA was extracted from the leaf powder using a DNeasy® Plant Mini Kit (250)(Qiagen Inc.,Valencia, CA) following the manufacturer's procedures. The genomic DNA concentration and quality was checked using a Nano-drop spectrometer (DeNovix DS-11 FX spectrophotometer). The concentration of the DNA was adjusted to 50 to 100 ng/µl and sent for DArTseq sequence genotyping under Integrated Genotyping Service and Support (IGSS) platform at BecA-ILRI, Nairobi, Kenya.

3.3 Data analysis

The genotype data was analyzed using different statistical tools; the missing percentage of data, expected heterozygosity ($He = 1 - \sum_{i=1}^{n}(Pi)^2$) and polymorphic information content ($PIC = 1 - \Sigma p_i^2 - \Sigma \Sigma p_i^2 p_j^2$) of the markers were calculated in Excel (Microsoft Excel for Office 365). The distributions of these markers were also analyzed and visualized using the R package synbreed (version 0.12-12) (Wimmer *et al.*, 2012). Discriminant analysis of principal components (DAPC) (Jombart and Ahmed, 2011), using a subset of robust SNP markers selected based on their expected hetrozygosity (He), polymorphic information content (PIC), missing value percentage and genome wide distribution were employed to select the best contributing markers. Euclidean genetic distance

$$d(p,q) = d(q,p) = \sqrt{(q1 - p1)2} + (q2 - p2)2 + \cdots (qn - pn)2$$

$$= \surd \sum_{i=1}^{n} (qi - pi)2$$

and neighbor-joining (NJ), Unweighted paired group method using arithmetic mean (UPGMA), Hierarchal clustering were calculated using the R functions dist and hclust and dendrogram or phylogenetic trees (based on Hierarchal clustering) were analysed using different R packages such as Adegenet (Jombart *et al.,* 2008) and Poppr (Kamvar *et al.,*2014*)* that are available in R statistical software. The degree of genetic similarity between genotypes were analysed using Nei's genetic distance (Nei, 1972; Pagnotta, 2018). Genetic diversity and population stratification were analyzed by the Bayesian algorithm implemented in STRUCTURE software (Pritchard, 2009; Corander *et al.*, 2013) and Discriminant analysis of principal components (DAPC) (Jombart and Ahmed, 2011), using a subset of robust SNP markers as described above. In addition, major clusters and sub clusters were detected by using hierarchical clustering with complete linkage analysis (Mohammadi and Prasanna, 2003; Kamvar *et al.*, 2014) and Bayesian model based Population Structure analysis was carried out using STRUCTURE software and the true value of K was determined using the method of Evano *et al. (*2005). AMOVA (Excoffier *et al.*, 1992) and fixation index (Meirmans and Hedrick, 2011) were employed to determine the extent of population differentiation among different groups and sub groups that were obtained from population structure analyses using the R package Pegas (Knaus and Winter, 2020). Diversity among and within accessions and progeny plants was analyzed using statistics on R software. The quality and distribution of the markers were checked using callrate , one ratio and distribution of the markers was visualized using the R package synbreed. From the generated SNP markers highly informative markers were selected for genetic diversity study based on the following criteria:

I. Markers with missing data percentage less than or equal to ten (NA% ≤10);
II. Minor allele frequency greater than or equal to five percent (MAF ≥5% or 0.05);
III. Polymorphic information content (PIC) greater than or equal to zero point two (PIC ≥0.2);
IV. Expected heterozygosity greater than or equal to zero point two (He≥0.2);
V. Distribution of the markers across the Linkage groups (genome wide distribution);
VI. Markers contribution to diversity using the loading function in discriminant analysis of principal components (DAPC).

4 Results

4.1 Diversity and genome-wide distribution of the DArTseq markers

A total of 363 Napier grass genotypes from different parts of the world were genotyped using DArTseq platform and 96,454 SilicoDArT and 96,321 SNP markers were generated. Prior to further diversity analysis, genotype missing percentage was checked and 16 genotypes (6 progenies, 2 EMBRAPA collections, 1 EMBRAPA elite line, 3 ICRISAT, 3 USDA and 1 ILRI accession), which had high missing value (≥ 50%) were excluded from further analysis hence a total of 347 Napier grass genotypes were used for diversity analysis.

The percentages of missing values of the markers ranged from 0.0 to 29.0 % for SilicoDArT markers and 0 to 83.7 % for SNP markers. The expected heterozygosity (He) and polymorphic information content (PIC) of the markers ranged from 0 to 0.5 and 0 to 0.38, respectively for both SilicoDArT and SNP markers while the average He was 0.26 and 0.18, and the average PIC was 0.21 and 0.15 for the SilicoDArT and SNP markers, respectively (Fig. 1).

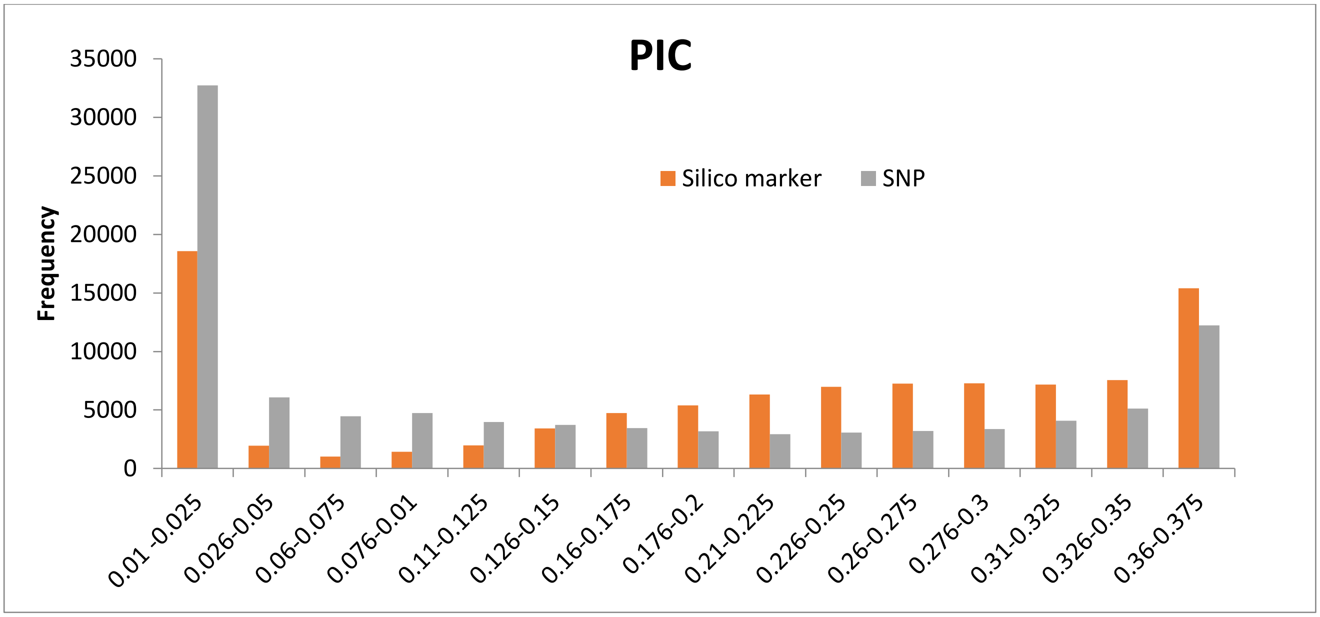


Figure 1: Frequency distribution of PIC values of SilicoDArT and SNP markers

In line to the distribution of SilicoDArT markers throughout the genome of Napier grass, out of 96,453 markers,76.28 % (73,573) mapped on to the fourteen linkage groups of the genome, while 1.73 % (1,669) mapped onto scaffolds and 21.99% (21,211) were found to be unmapped (unknown location). The highest number of markers were mapped in linkage group two (LG2) (7.32% (7,061) and (LG3) 7.14% (6,889) followed by LG1 (6.87% (6,623)), LG7 (6.61% (6,375)) and LG6 (6.50% (6,267)), while LG14 contained the least number of markers 3.36% (3,242) (Fig. 2a).

From the 96,400 SNP markers, 88.82% (85,619) were aligned on to the fourteen linkage groups while 1.87% (1,802) mapped on to different scaffolds and 9.31% (8,979) were not able to be mapped on to the draft genome. Out of the aligned 88.82% markers, 8.74% were mapped on LG2 and 8.51% on LG3. The percentage of mapping for LG1 (7.69% (7,416)), LG7 (7.40% (7,129)), LG6 (7.32% (7,058)), LG4 (6.50% (6,264)) and LG13 (6.15% (5,925)) and the lowest numbers of markers 3.79% (3,655) were mapped on LG14 (Fig.2b).

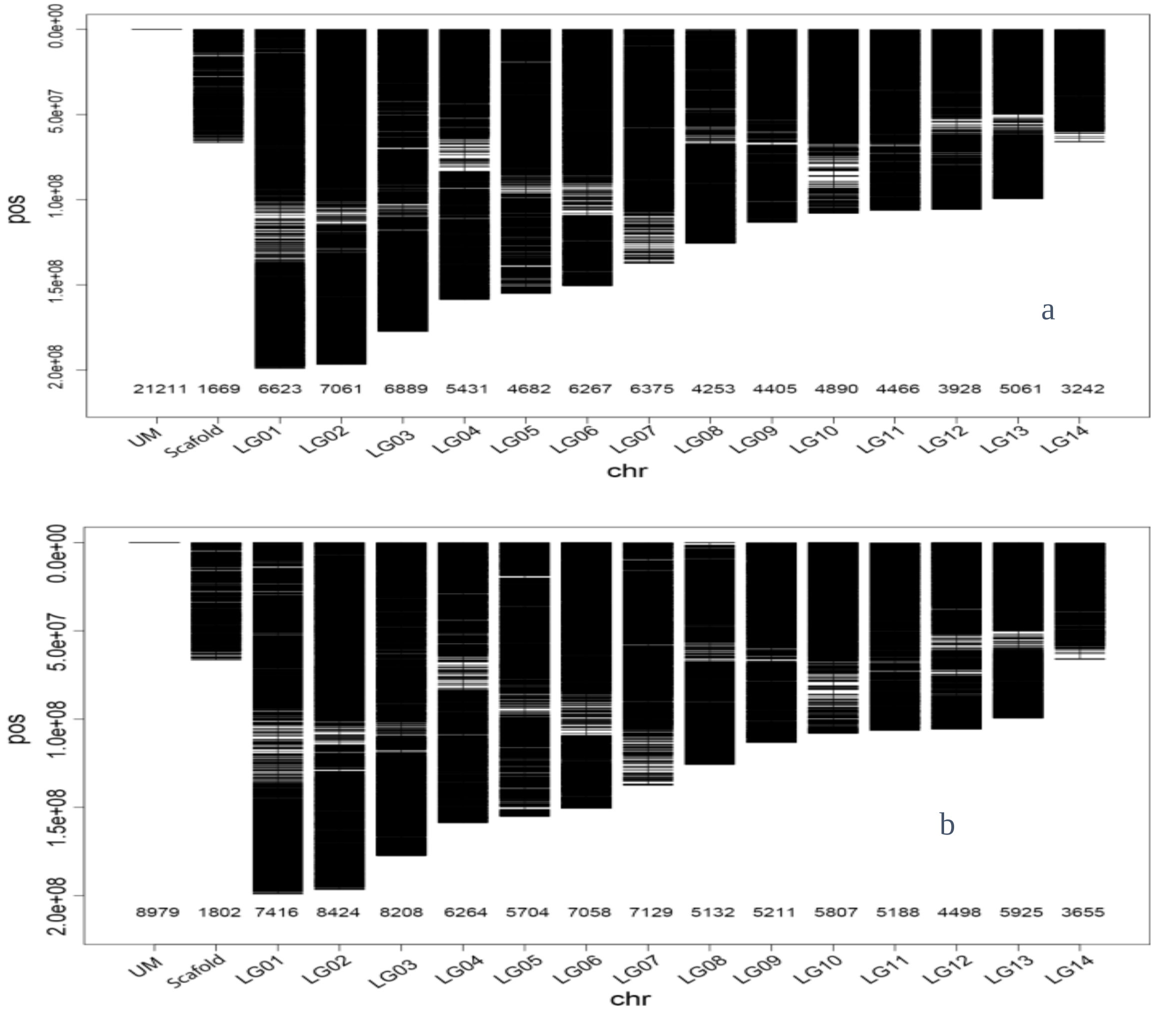


Figure 2: The Distribution of SilicoDArT (a) and SNP (b) markers throughout the fourteen chromosomes of the Napier grass (*C. purpureus*) draft genome.

4.1.1 Marker selection for Diversity study

Stepwise marker selection was conducted in such a way that each criterion is taken in to account after the previous criteria are considered. Hence, initially markers were selected based on missing value percentage (NA %) in which markers with NA % less than or equal to ten were retained while the rest removed so that 79,831 SNP markers were retained. Thus, based on the second criterion 3,629 SNP makers were retained. Further the third and fourth criteria reduced the markers to 2,357 SNP markers. Then using the R package Adegenet to identify and select the best contributing markers based on DAPCs, approximately 1000 best contributing makers, with an average PIC value of 0.31, were selected out of the 2,357 SNPs markers for diversity and population structure study. The distribution of these DAPCs selected markers were also analyzed and visualized using the R Package synbreed (version 0.12-12) and out of these selected markers, 1.8% (18) were unmapped, 0.9% (9) mapped to scaffolds, and the rest were distributed across the linkage groups with the highest distribution (map) number on LG2, (9.2% or 92 markers) and lowest number on LG14, which was 4.6% or 46 markers (Fig. 3).

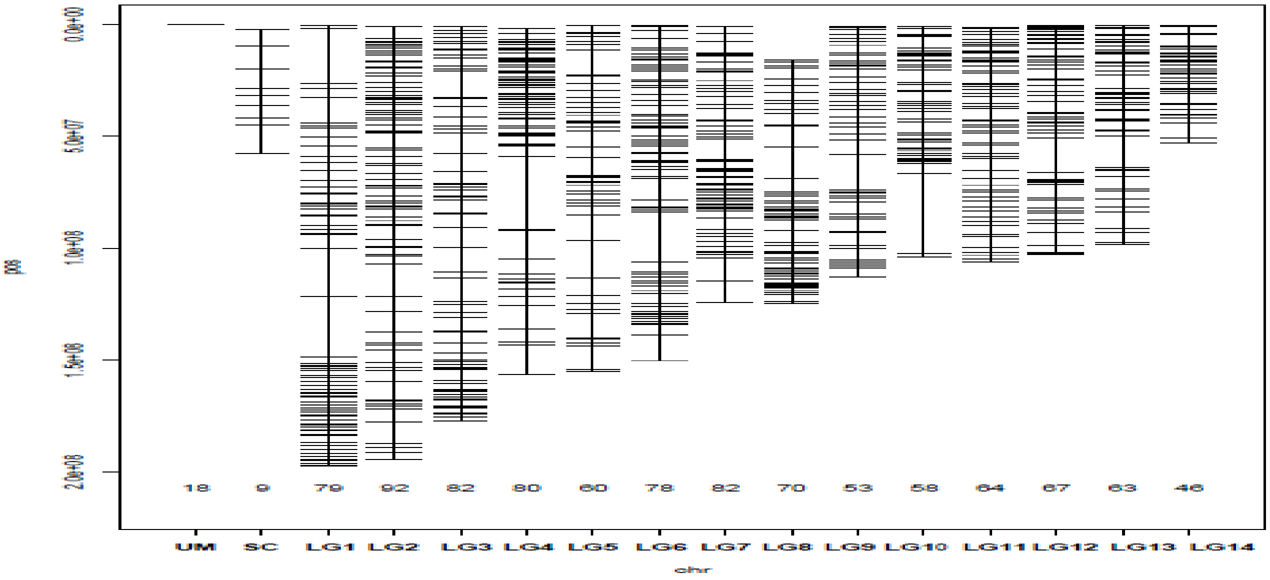


Figure 3: Genome wide distribution of Discriminant Analysis of Principal Components (DAPC) selected 1001 SNP markers

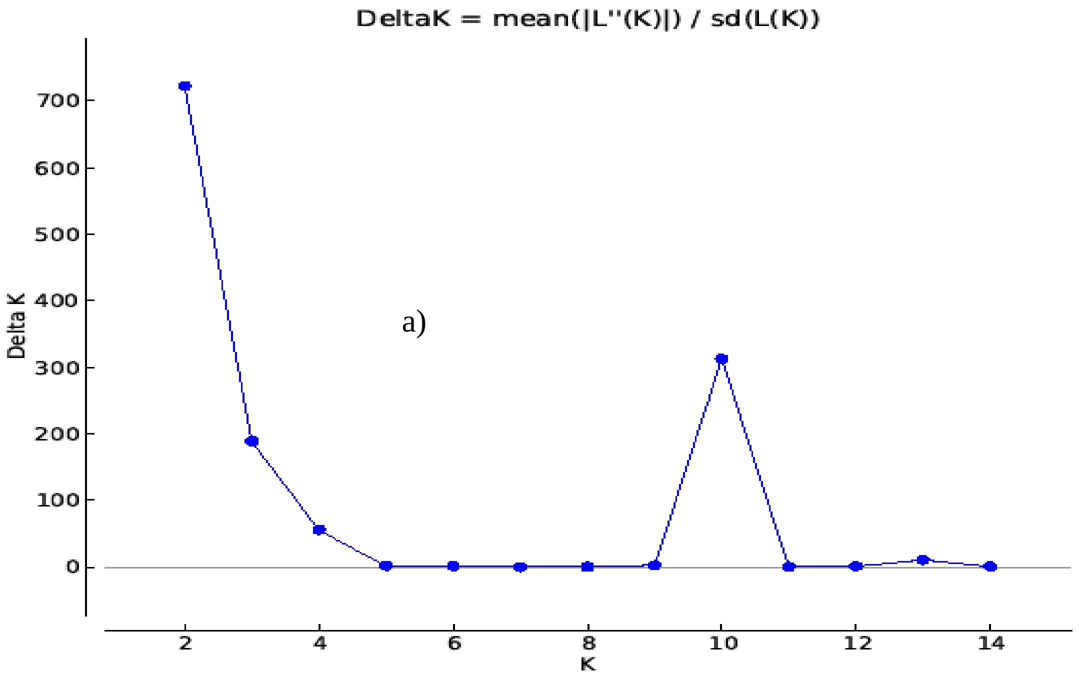
4.2 Genetic diversity in Napier grass populations

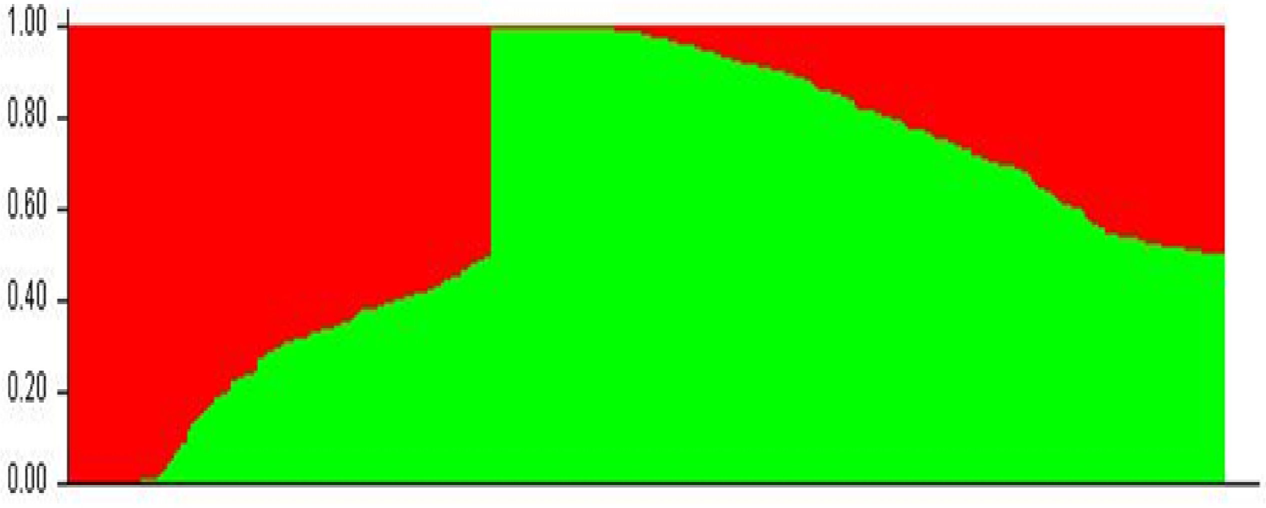
Genetic diversity and population stratification were detected by using DAPC, Structure software, and hierarchal clustering approaches and a significant amount of genetic diversity and population stratifications in the Napier grass collections and progeny plants with a high degree of consistency were obtained. The degree of genetic similarity between genotypes was identified using Nei's genetic distance (Nei, 1972). Furthermore, high amount of genetic diversity and the extent of variation among clusters and sub clusters were confirmed by AMOVA, and the fixation index (Fst).

4.2.1 Genetic diversity revealed by the Bayesian algorithm in STRUCTURE software

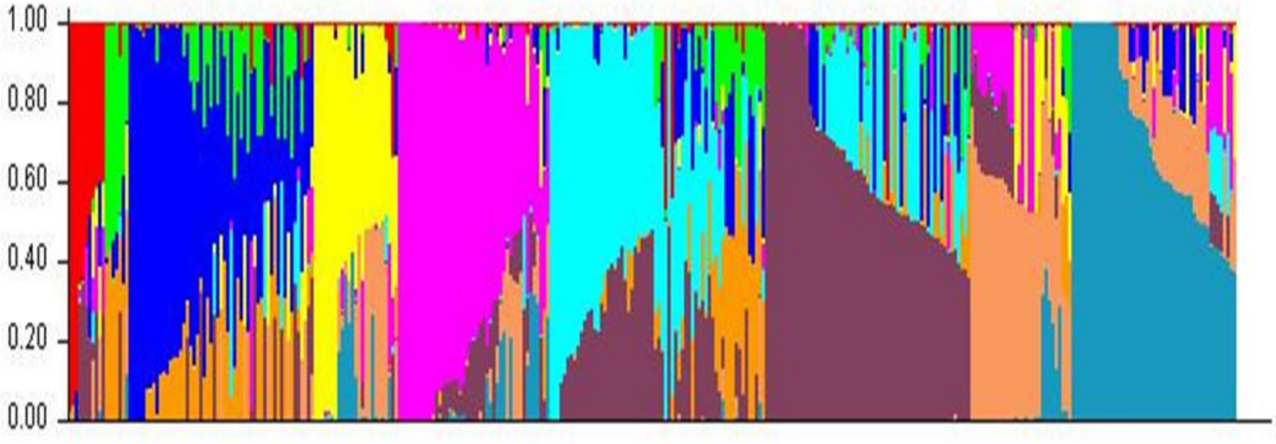
Population structure analysis using the structure software revealed two major clusters with ten sub clusters as suggested by the delta k (Dent, 2012) (Fig.4a). Cluster I comprises almost half of the progeny plants and most of the EMBRAPA collection (BAGCEs), some accessions from the ICRISAT collection, some from the ILRI collection, two of the EMBRAPA elite lines (CNPGLs), and just one from the USDA collection. Cluster II is composed mainly of genotypes from the ILRI collection, progeny plants, most of the EMBRAPA elite lines (CNPGLs) and the USDA collection.

The collections were further grouped into ten sub-clusters, in which cluster I was further divided into six sub-clusters, while cluster II was divided into four sub-clusters (Supplementary table 2). Eighty one individual genotypes were found to be admixed (membership probability <50 %) among the sub-clusters (Supplementary table 3).The first sub cluster from the population structure analysis consisted of 47 progeny plants (13 of the ILRI_NS_11, 16 of the ILRI_NS_7 and ILRI_NS_9, 14 of the ILRI_NS_3 and ILRI_NS_6, 3 of the ILRI_NS_13 and 1 of the ILRI_NS_4). The second sub cluster consisted of 6 genotypes that were entirely ILRI accessions. The third sub cluster consisted of 7 genotypes, all progeny plants from ILRI_NS_6.The fourth sub-cluster consisted of 23 genotypes (11 of the ILRI collection, 9 genotypes of the ICRISAT collection and 3 genotypes of the EMBRAPA collection). The fifth sub cluster consists of 18 genotypes, out of these 12 were from the EMBRAPA collection, 3 were from the ILRI collection, 2 were EMBRAPA elite lines and one USDA genotype. The sixth sub cluster consisted of 37 genotypes, all of which were progeny plants (10 of ILRI_NS_12 and ILRI_NS_13, 6 of ILRI_NS_10, 7 of ILRI_NS_7, 4 of ILRI_NS_4 and 2 of ILRI_NS_6). The seventh sub-cluster consisted of 29 genotypes, of which 21(72.41%) were from the ILRI collection, 4 (23.79%) were EMBRAPA elite lines, two were from the USDA and one each were from the ICRISAT and EMBRAPA collections. The eighth sub-cluster consisted of 27 genotypes, mostly genotypes from the USDA collection and EMBRAPA elite lines. The ninth sub-cluster consisted of 7 genotypes of which six were from the ILRI collection and one was a USDA genotype. The last sub-cluster (sub-cluster X) consists of 66 genotypes, mostly represented by progeny plants (ILRI_NS_1, ILRI_NS_2, ILRI_NS_3, and ILRI_NS_5 and ILRI_NS_8) with only two genotypes from the ILRI collection and one from the ICRISAT collection.





b)



c)

Figure 4: a) Delta k that shows picks at K = 2 and K = 10, suggesting possible number of populations/ clusters; b) The two major clusters of the populations/structures; c) The ten populations/sub-clusters and the admixtures

4.2.2 Genetic diversity revealed by the Discriminant Analysis of Principal Components (DAPC)

DAPC clustered the population into ten clusters (K groups) (Fig. 5). The first cluster (K1) consists a total of 52 progeny plants from these 52 genotypes 38 (73.1%) were from ILRI_NS_12, ILRI_NS_13 and ILRI_NS_7. Generally, cluster (I) represents progeny plants while the second cluster consisted 52 genotypes, out of these 50 genotypes were progeny plants and the rest 2 from USDA and ICRISAT accession. The third cluster consisted 33 genotypes and of these 32 were progenies and one ILRI accession. The fourth cluster consisted 44 genotypes and of these 19 were progeny plants while the remaining were contributed by BAGCE (1), CNPGL (8), USDA (9) India (4) and ILRI (3) accessions. The fifth cluster consisted 28 genotypes where most of them were from ILRI and CNPGL genotypes (11 and 9, respectively). The remaining 4 were each from Indian and Tift genotypes. The sixth cluster consisted six genotypes which were entirely from ILRI accessions. The seventh cluster consisted of 30 genotypes while half (15) of these were ILRI genotypes, 11 were ICRISAT genotypes, 3 were BAGCE and 1 was USDA genotypes. The eighth cluster consisted of 48 genotypes while 46 were progeny plants and the remaining 2 were ILRI accessions. The ninth cluster consisted of 34 genotypes of BAGCE (13), ILRI (8), ICRISAT (5), USDA (4) and CNPGL (4) genotypes. The last cluster (cluster 10) consisted 20 genotypes of these 16 were ILRI accessions, 3 were EMBRAPA elite lines and 1 was EMBRAPA collection.

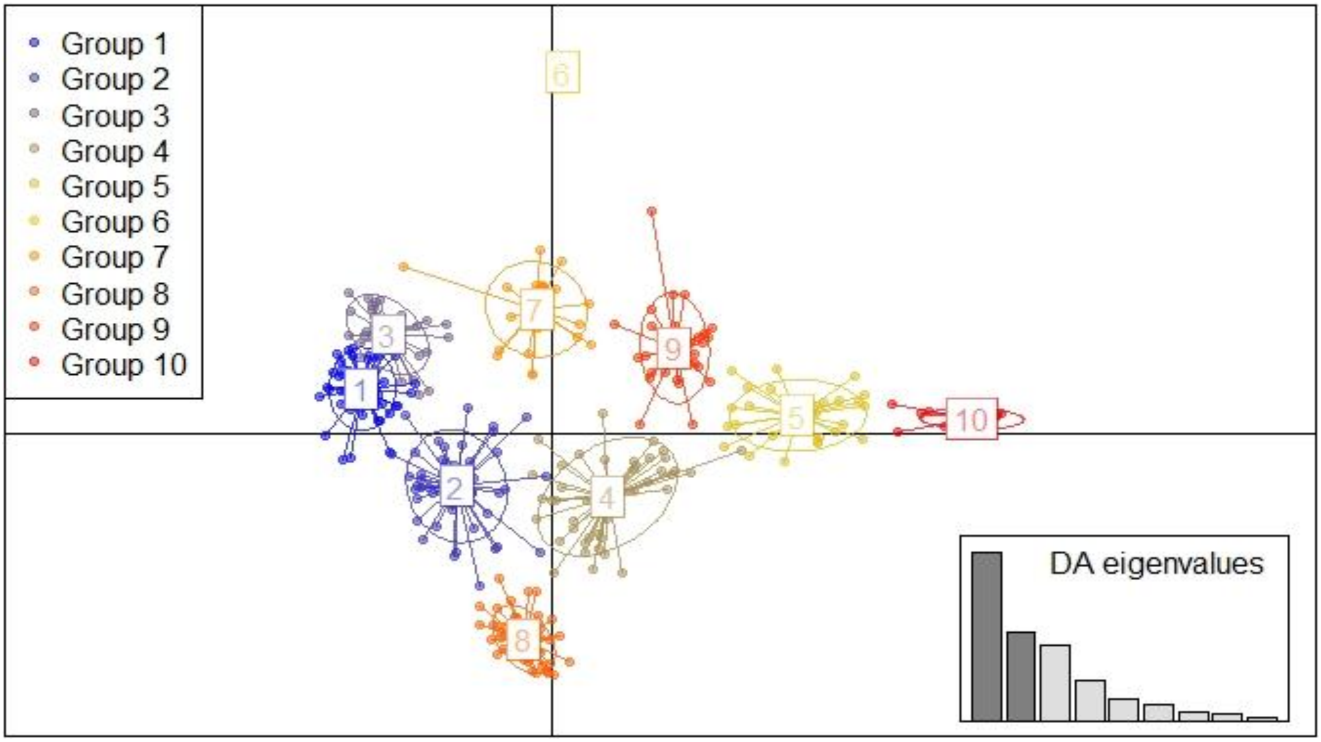


Figure 5: Discriminant Analysis of Principal Components (DAPC) for 347 Napier grass genotypes. The axes represent the first two discriminants. Each circle represents a cluster and each dot (point) represents an individual. Each group represents the different sub-clusters (K groups) identified by DAPC analysis.

Table 2: List of genotypes under different K groups from Discriminant Analysis of Principal Components (DAPC)

k1	k2	k3	k4	k5	k6	k7	k8	k9	k10
ILRI_NS_4.2	ILRI_NS_1.4	ILRI_NS_3.2	ILRI_NS_1.9	CNPGL_93.08.1	ILRI_16808	BAGCE_24	ILRI_Ns_1.1	BAGCE_16	BAGCE_63
ILRI_NS_4.10	ILRI_NS_1.16	ILRI_NS_6.5	ILRI_NS_3.11	CNPGL_93.32.2	ILRI_16809	BAGCE_80	ILRI_NS_1.2	BAGCE_56	CNPGL_92.38.2
ILRI_NS_4.11	ILRI_NS_1.18	ILRI_NS_6.8	ILRI_NS_5.5	CNPGL_94.07.2	ILRI_16810	Tift_N130	ILRI_NS_1.3	BAGCE_7	ILRI_14984
ILRI_NS_4_15	ILRI_NS_3.1	ILRI_NS_6.10	ILRI_NS_5.8	Tift_N147	ILRI_16818	India_149_21785	ILRI_NS_1.5	BAGCE_75	ILRI_16785
ILRI_NS_7.1	ILRI_NS_3.3	ILRI_NS_6.12	ILRI_NS_5.12	Tift_N109	ILRI_16821	India_126_22231	ILRI_NS_1.6	BAGCE_94	ILRI_16786
ILRI_NS_7.2	ILRI_NS_3.4	ILRI_NS_6.14	ILRI_NS_5.19	Tift_N131	ILRI_16822	India_118_22241	ILRI_NS_1.10	Tift_N37	ILRI_16787
ILRI_NS_7.4	ILRI_NS_3.6	ILRI_NS_6.15	ILRI_NS_6.4	Tift_N138		India_124_22233	ILRI_NS_1.12	Tift_N210	ILRI_16789
ILRI_NS_7.5	ILRI_NS_3.7	ILRI_NS_6.19	ILRI_NS_6.11	India_150_21784		India_125_22232	ILRI_NS_1.13	Tift_N71	ILRI_16792
ILRI_NS_7_6	ILRI_NS_3.8	ILRI_NS_7.3	ILRI_NS_6.13	India_144_21964		India_115_22243	ILRI_NS_1.14	Tift_N137	ILRI_16795
ILRI_NS_7.12	ILRI_NS_3.9	ILRI_NS_7.7	ILRI_NS_6.18	India_141_21967		India_132_22225	ILRI_NS_1.15	India_127_22230	ILRI_16798
ILRI_NS_7.13	ILRI_NS_3.10	ILRI_NS_7.10	ILRI_NS_6.20	India_147_21787		India_121_22237	ILRI_NS_2.1	India_146_21788	ILRI_16800
ILRI_NS_7.14	ILRI_NS_3.13	ILRI_NS_7.11	ILRI_NS_8.2	ILRI_14355		India_123_22234	ILRI_NS_2.2	India_142_21966	ILRI_16801
ILRI_NS_7.15	ILRI_NS_3.14	ILRI_NS_9.1	ILRI_NS_8.4	ILRI_14389		India_89_22236	ILRI_NS_2.3	India_131_22226	ILRI_16803
ILRI_NS_7.17	ILRI_NS_3.15	ILRI_NS_9.2	ILRI_NS_8.9	ILRI_14982		India_116_22242	ILRI_NS_2.6	India_151_21783	ILRI_16804
ILRI_NS_7.18	ILRI_NS_3.16	ILRI_NS_9.8	ILRI_NS_8.14	ILRI_15357		ILRI_1026	ILRI_NS_2.7	ILRI_15743.MOTT.	ILRI_16806
ILRI_NS_7.20	ILRI_NS_3.17	ILRI_NS_9.9	ILRI_NS_8.16	ILRI_16793		ILRI_16782	ILRI_NS_2.8	ILRI_16783	ILRI_16836
ILRI_NS_10.2	ILRI_NS_3.18	ILRI_NS_9.10	ILRI_NS_8.17	ILRI_16799		ILRI_16790	ILRI_NS_2.9	ILRI_16788	ILRI_18438
ILRI_NS_10.4	ILRI_NS_3.20	ILRI_NS_9.12	ILRI_NS_12.10	ILRI_16819		ILRI_16794	ILRI_NS_3.5	ILRI_16791	CNPGL_9279.2
ILRI_NS_10.5	ILRI_NS_4.1	ILRI_NS_9.13	ILRI_NS_12.19	ILRI_16837		ILRI_16796	ILRI_NS_3.12	ILRI_16802	CNPGL_92.66.3
ILRI_NS_10.6	ILRI_NS_4.3	ILRI_NS_9.15	BAGCE_1	ILRI_16902		ILRI_16797	ILRI_NS_3.19	ILRI_16812	Napier_Addis
ILRI_NS_10.7	ILRI_NS_4.4	ILRI_NS_9.18	CNPGL_91.06.2	ILRI_16840		ILRI_16805	ILRI_NS_4.14	ILRI_16813	
ILRI_NS_10_8	ILRI_NS_4.5	ILRI_NS_11.4	CNPGL_91.11.2	CNPGL_93.18.2		ILRI_16807	ILRI_NS_5.1	ILRI_16815	
ILRI_NS_10.9	ILRI_NS_4.6	ILRI_NS_11.5	CNPGL_91.25.1	CNPGL_93.01.1		ILRI_16814	ILRI_NS_5.3	BAGCE_97	

ILRI_NS_10.11
ILRI_NS_10.12
ILRI_NS_11_6
ILRI_NS_12.1
ILRI_NS_12.3
ILRI_NS_12.4
ILRI_NS_12.5
ILRI_NS_12.6
ILRI_NS_12.8
ILRI_NS_12.11
ILRI_NS_12.12
ILRI_NS_12.13
ILRI_NS_12.14
ILRI_NS_12.17
ILRI_NS_12.18
ILRI_NS_12.20
ILRI_NS_13.1
ILRI_NS_13.2
ILRI_NS_13.3
ILRI_NS_13.5
ILRI_NS_13.7
ILRI_NS_13.8
ILRI_NS_13.9
ILRI_NS_13.11
ILRI_NS_13.12
ILRI_NS_13.13
ILRI_NS_13.15
ILRI_NS_13.16
ILRI_NS_13.19

ILRI_NS_4.7
ILRI_NS_4.9
ILRI_NS_4.12
ILRI_NS_4.13
ILRI_NS_4.18
ILRI_NS_4.19
ILRI_NS_4.20
ILRI_NS_5.16
ILRI_NS_6.1
ILRI_NS_6.3
ILRI_NS_6.6
ILRI_NS_6.7
ILRI_NS_6.9
ILRI_NS_6.16
ILRI_NS_7.9
ILRI_NS_9.7
ILRI_NS_10.1
ILRI_NS_10.3
ILRI_NS_10.10
ILRI_NS_10.13
ILRI_NS_11.1
ILRI_NS_11.2
ILRI_NS_11.10
ILRI_NS_11.11
ILRI_NS_11.19
ILRI_NS_12.2
ILRI_NS_13.17
Tift_N128
India_128_22229

ILRI_NS_11.8
ILRI_NS_11.12
ILRI_NS_11.13
ILRI_NS_11.14
ILRI_NS_11.15
ILRI_NS_11.16
ILRI_NS_11.17
ILRI_NS_11.18
ILRI_NS_11.20
ILRI_16835

CNPGL_92.190.01
CNPGL_93.06.1
Tift_N200
Tift_N172
Tift_N43
Tift_N23
Tift_N225
Tift_N75
Tift_N8
Tift_N223
Tift_N68
India_145.1._21965
India_120_22238
India_119_22239
India_129_22228
ILRI_16784
ILRI_16811
CNPGL_96.21.1
CNPGL_93.37.5
CNPGL_93.04.2
mott_new2

CNPGL_96.27.3
CNPGL_92.133.3
CNPGL_92.56.2
CNPGL_94.13.1
Maralfalfa.1

ILRI_16816
ILRI_16817
ILRI_16834
ILRI_16838
ILRI_18448
ILRI_18662
BAGCE_17

ILRI_NS_5.2
ILRI_NS_5.4
ILRI_NS_5.6
ILRI_NS_5.9
ILRI_NS_5.10
ILRI_NS_5.13
ILRI_NS_5.14
ILRI_NS_5.15
ILRI_NS_5.17
ILRI_NS_5.18
ILRI_NS_6_2
ILRI_NS_8.3
ILRI_NS_8.7
ILRI_NS_8.8
ILRI_NS_8.10
ILRI_NS_8.11
ILRI_NS_8.15
ILRI_NS_8.19
ILRI_NS_8.20
ILRI_NS_11.3
ILRI_NS_11.9
ILRI_NS_12.7
ILRI_NS_12.15
ILRI_14983
ILRI_16839

CNPGL_00.1.1
BAGCE_30
BAGCE_53
BAGCE_90
BAGCE_81
BAGCE_34
BAGCE_86
CNPGL_96.23.1
CNPGL_92.198.7
PIONEIRO
BAGCE_100

4.2.3 Genetic diversity revealed by the hierarchal cluster analysis

Hierarchical clustering with complete linkage (hclust) analysis resulted into two major clusters and up to ten sub-clusters (Fig. 6), which is highly similar to the STRUCTURE clustering except that the genotypes identified as admixed by STRUCTURE are distributed across the different sub-clusters in this case.

Under the hierarchal cluster analysis, sub-cluster I represented 51 genotypes of these 50 were progenies (Table. 3); (9 progenies from each ILRI_NS7 and ILRI_NS9, 10 progenies from each ILRI_NS12 and ILRI_NS13, 4 progenies from each ILRI_NS_10 and ILRI_NS_11, 3 from ILRI_NS_4 and 1 ILRI_NS_6) and the rest were ILRI genotypes. The second sub-cluster consisted 34 genotypes in which all of them are progeny plants from (ILRI_NS_1, 10, 11, 12, 13, 3, 4, 6 and 7 with respective number of 1, 8, 2, 4, 4, 3, 6, 2, and 4). The third sub-cluster consist 27 of progenies from ILRI_NS_11, ILRI_NS_3, ILRI_NS_4, ILRI_NS_6 and ILRI_NS_7 in proportion of 8, 7, 2, 6 and 4, respectively. The fourth sub-cluster consists 25 genotypes and of these 17 were ILRI, 5 were ICRISAT and 3 BAGCE genotypes. The fifth sub-cluster consist 23 genotypes out of these 22 were progenies (6 ILRI_NS_3, 4 for each ILRI_NS_4 and ILRI_NS_6, (1 for each ILRI_NS_5, ILRI_NS_9 and ILRI_NS_10), 2 and 3 ILRI_NS_1 and ILRI_NS_11 respectively) and the rest 1 was ICRISAT genotypes. The sixth sub-cluster consist 46 genotypes from BAGCE (13), ILRI (11), ICRISAT (13), USDA (6), CNPGL (2) and ILRI_NS_4 (1).This sub-cluster generally represents most of the Napier grass collections. The seventh sub-cluster consist 48 genotypes of these 46 were progeny plants from (10ILRI_NS_1, 2were each ILRI_NS_11 and ILRI_NS_12, 7 were ILRI_NS_2, 3 were ILRI_NS_3, 1 were each ILRI_NS_4 and ILRI_NS_12 ILRI_NS_5 and 8 ILRI_NS_8) and the remaining 2 were ILRI collection. The eighth sub-cluster consist 26 genotypes of these 9 were (7 and 2 were USDA and ICRISAT) genotypes respectively. The remaining 17 genotypes were progeny plants of ILRI_NS_5, ILRI_NS_6 and ILRI_NS_8.Sub-cluster nine consist 36 genotypes out of these 27 were ILRI, 6 were CNPGL, 1 was BAGCE and2 were USDA genotypes. The last sub-cluster (sub-cluster ten) consist 31 genotypes most of these were from EMBRAPA; 16 CNPGL and 1 BAGCE. The rest 4 each were ILRI, ICRISAT and USDA genotypes and 2 ILRI_NS_12 progeny plant.

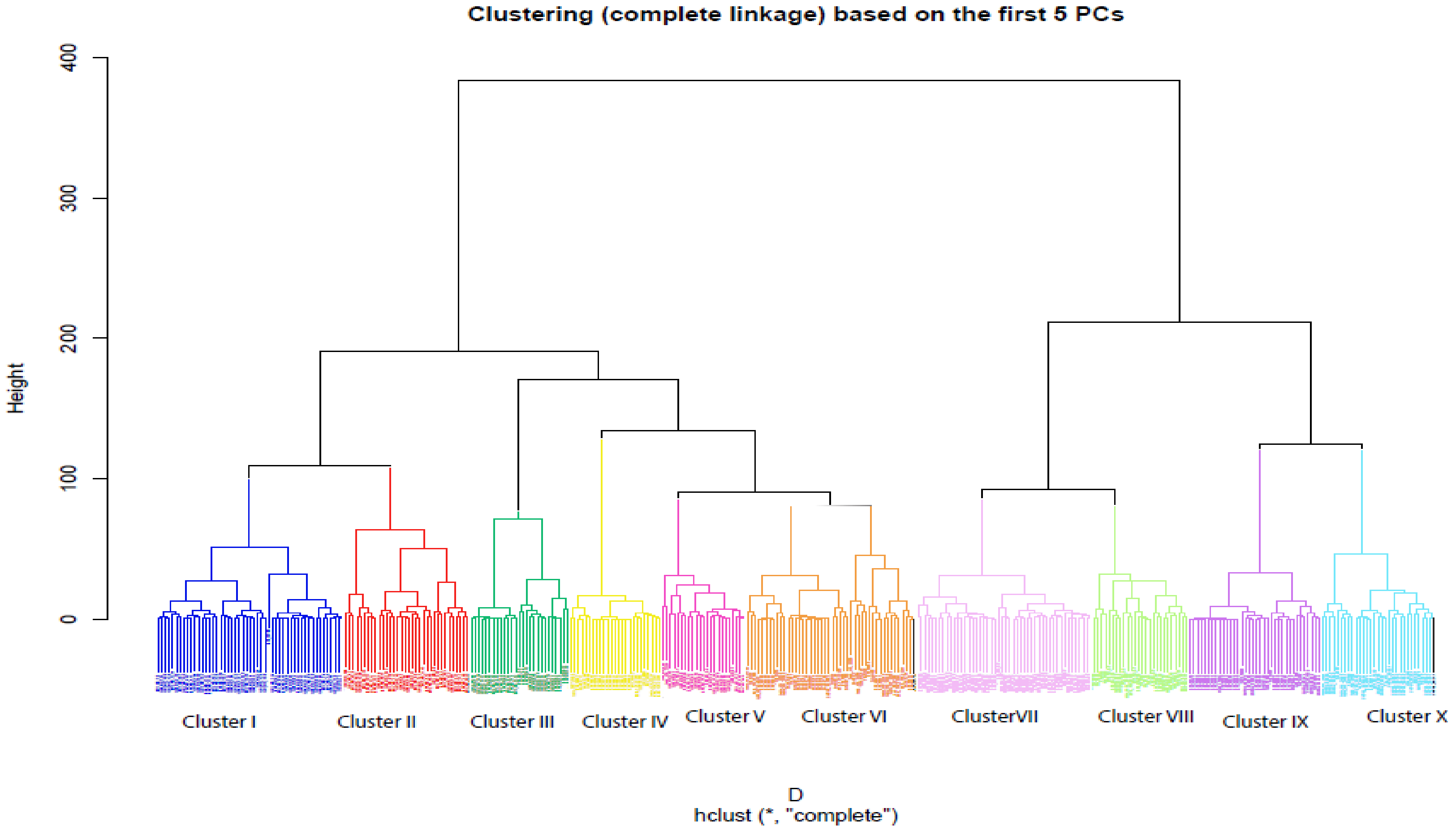


Figure 6: A dendrogram showing clusters and sub-clusters According to the hierarchal clustering based on complete linkage analysis

Table 3: The list of genotypes under different clusters and sub-clusters detected by hierarchal clustering analysis

Cluster I						**Cluster II**			
sub-cluster I	**sub-cluster II**	**sub-cluster III**	**sub-cluster IV**	**sub-cluster V**	**sub-cluster VI**	**sub-cluster VII**	**sub-cluster VIII**	**sub-cluster IX**	**sub-cluster X**
ILRI_NS_9.9	ILRI_NS_1.16	ILRI_NS_11.12	BAGCE_24	ILRI_NS_5.16	BAGCE_100	ILRI_14983	ILRI_NS_1.9	BAGCE_63	BAGCE_1
ILRI_NS_9.8	ILRI_NS_10.1	ILRI_NS_11.14	BAGCE_17	ILRI_NS_1.18	BAGCE_16	ILRI_16839	ILRI_NS_3.11	CNPGL_92.133.3	CNPGL_00.1.1
ILRI_NS_9.2	ILRI_NS_10.10	ILRI_NS_11.16	BAGCE_80	ILRI_NS_1.4	BAGCE_30	ILRI_Ns_1.1	ILRI_NS_5.12	CNPGL_92.38.2	CNPGL_91.06.2
ILRI_NS_9.18	ILRI_NS_10.11	ILRI_NS_11.17	ILRI_16782	ILRI_NS_10.3	BAGCE_34	ILRI_NS_1.10	ILRI_NS_5.19	CNPGL_9279.2	CNPGL_91.11.2
ILRI_NS_9.15	ILRI_NS_10.12	ILRI_NS_11.18	ILRI_16794	ILRI_NS_11.11	BAGCE_53	ILRI_NS_1.12	ILRI_NS_5.5	CNPGL_93.01.1	CNPGL_91.25.1
ILRI_NS_9.13	ILRI_NS_10.13	ILRI_NS_11.4	ILRI_16796	ILRI_NS_11.19	BAGCE_56	ILRI_NS_1.13	ILRI_NS_5.8	CNPGL_93.32.2	CNPGL_92.190.01
ILRI_NS_9.12	ILRI_NS_10.5	ILRI_NS_11.5	ILRI_16797	ILRI_NS_11.2	BAGCE_7	ILRI_NS_1.14	ILRI_NS_6.11	CNPGL_94.13.1	CNPGL_92.56.2
ILRI_NS_9.10	ILRI_NS_10.6	ILRI_NS_11.8	ILRI_16805	ILRI_NS_3.13	BAGCE_75	ILRI_NS_1.15	ILRI_NS_6.13	ILRI_14355	CNPGL_92.66.3
ILRI_NS_9.1	ILRI_NS_10_8	ILRI_NS_3.1	ILRI_16807	ILRI_NS_3.15	BAGCE_81	ILRI_NS_1.2	ILRI_NS_6.18	ILRI_14389	CNPGL_93.04.2
ILRI_NS_7_6	ILRI_NS_11.1	ILRI_NS_3.10	ILRI_16808	ILRI_NS_3.16	BAGCE_86	ILRI_NS_1.3	ILRI_NS_6.20	ILRI_14982	CNPGL_93.06.1
ILRI_NS_7.5	ILRI_NS_11.10	ILRI_NS_3.18	ILRI_16809	ILRI_NS_3.20	BAGCE_90	ILRI_NS_1.5	ILRI_NS_6.4	ILRI_14984	CNPGL_93.08.1
ILRI_NS_7.4	ILRI_NS_12.11	ILRI_NS_3.2	ILRI_16810	ILRI_NS_3.4	BAGCE_94	ILRI_NS_1.6	ILRI_NS_8.14	ILRI_15357	CNPGL_93.18.2
ILRI_NS_7.20	ILRI_NS_12.13	ILRI_NS_3.3	ILRI_16816	ILRI_NS_3.9	BAGCE_97	ILRI_NS_11.3	ILRI_NS_8.16	ILRI_16785	CNPGL_93.37.5
ILRI_NS_7.2	ILRI_NS_12.2	ILRI_NS_3.6	ILRI_16818	ILRI_NS_4.1	CNPGL_92.198.7	ILRI_NS_11.9	ILRI_NS_8.17	ILRI_16786	CNPGL_94.07.2
ILRI_NS_7.18	ILRI_NS_12.20	ILRI_NS_3.8	ILRI_16821	ILRI_NS_4.18	ILRI_1026	ILRI_NS_12.15	ILRI_NS_8.2	ILRI_16787	CNPGL_96.21.1
ILRI_NS_7.15	ILRI_NS_13.11	ILRI_NS_4.13	ILRI_16822	ILRI_NS_4.6	ILRI_15743.MOTT.	ILRI_NS_12.7	ILRI_NS_8.4	ILRI_16789	CNPGL_96.23.1
ILRI_NS_7.14	ILRI_NS_13.12	ILRI_NS_4.4	ILRI_16834	ILRI_NS_4.9	ILRI_16783	ILRI_NS_2.1	ILRI_NS_8.9	ILRI_16792	CNPGL_96.27.3
ILRI_NS_7.1	ILRI_NS_13.17	ILRI_NS_6.10	ILRI_16838	ILRI_NS_6.1	ILRI_16788	ILRI_NS_2.2	India_129_22228	ILRI_16793	ILRI_16784
ILRI_NS_6.19	ILRI_NS_13.7	ILRI_NS_6.12	ILRI_18448	ILRI_NS_6.16	ILRI_16791	ILRI_NS_2.3	India_145.1._21965	ILRI_16795	ILRI_16811
ILRI_NS_4_15	ILRI_NS_3.14	ILRI_NS_6.14	ILRI_18662	ILRI_NS_6.6	ILRI_16802	ILRI_NS_2.6	Tift_N172	ILRI_16798	ILRI_16813
ILRI_NS_4.11	ILRI_NS_3.17	ILRI_NS_6.15	India_123_22234	ILRI_NS_6.7	ILRI_16812	ILRI_NS_2.7	Tift_N200	ILRI_16799	ILRI_NS_12.10
ILRI_NS_4.10	ILRI_NS_3.7	ILRI_NS_6.5	India_124_22233	ILRI_NS_9.7	ILRI_16814	ILRI_NS_2.8	Tift_N223	ILRI_16800	ILRI_NS_12.19
ILRI_NS_13.9	ILRI_NS_4.12	ILRI_NS_6.8	India_125_22232	India_128_22229	ILRI_16815	ILRI_NS_2.9	Tift_N225	ILRI_16801	India_141_21967

ILRI_NS_13.8
ILRI_NS_13.5
ILRI_NS_13.3
ILRI_NS_13.2
ILRI_NS_13.19
ILRI_NS_13.16
ILRI_NS_13.15
ILRI_NS_13.13
ILRI_NS_13.1
ILRI_NS_12.8
ILRI_NS_12.6
ILRI_NS_12.5
ILRI_NS_12.4
ILRI_NS_12.3
ILRI_NS_12.18
ILRI_NS_12.17
ILRI_NS_12.14
ILRI_NS_12.12
ILRI_NS_12.1
ILRI_NS_11_6
ILRI_NS_11.20
ILRI_NS_11.15
ILRI_NS_11.13
ILRI_NS_10.9
ILRI_NS_10.7
ILRI_NS_10.4
ILRI_NS_10.2
ILRI_16790

ILRI_NS_4.19
ILRI_NS_4.2
ILRI_NS_4.3
ILRI_NS_4.5
ILRI_NS_4.7
ILRI_NS_6.3
ILRI_NS_6.9
ILRI_NS_7.12
ILRI_NS_7.13
ILRI_NS_7.17
ILRI_NS_7.9

ILRI_NS_7.10
ILRI_NS_7.11
ILRI_NS_7.3
ILRI_NS_7.7

India_126_22231
India_149_21785

ILRI_16817
ILRI_16835
ILRI_NS_4.20
India_115_22243
India_116_22242
India_118_22241
India_119_22239
India_120_22238
India_121_22237
India_127_22230
India_131_22226
India_132_22225
India_142_21966
India_146_21788
India_151_21783
India_89_22236
PIONEIRO
Tift_N128
Tift_N130
Tift_N137
Tift_N210
Tift_N37
Tift_N71

ILRI_NS_3.12
ILRI_NS_3.19
ILRI_NS_3.5
ILRI_NS_4.14
ILRI_NS_5.1
ILRI_NS_5.10
ILRI_NS_5.13
ILRI_NS_5.14
ILRI_NS_5.15
ILRI_NS_5.17
ILRI_NS_5.18
ILRI_NS_5.2
ILRI_NS_5.3
ILRI_NS_5.4
ILRI_NS_5.6
ILRI_NS_5.9
ILRI_NS_6_2
ILRI_NS_8.10
ILRI_NS_8.11
ILRI_NS_8.15
ILRI_NS_8.19
ILRI_NS_8.20
ILRI_NS_8.3
ILRI_NS_8.7
ILRI_NS_8.8

Tift_N68
Tift_N75
Tift_N8

ILRI_16803
ILRI_16804
ILRI_16806
ILRI_16819
ILRI_16836
ILRI_16837
ILRI_16840
ILRI_16902
ILRI_18438
Maralfalfa.1
Napier_Addis
Tift_N131
Tift_N147

India_144_21964
India_147_21787
India_150_21784
mott_new2
Tift_N109
Tift_N138
Tift_N23
Tift_N43

4.3 Population differentiation and divergence analysis

The Nei's genetic distance reflected the diversity among the Napier grass collections and progeny plants and among clusters and sub-clusters. Based on the overall dataset, Nei's genetic distance ranged from 0.07071 for sub-cluster I and VI to 0.5118 for sub-clusters II and IX (Table. 4).The six ILRI collections in sub-cluster II showed high similarity among them, with 0.005 to 0.02 ranges for Nei's genetic distance. Analysis of molecular variance (AMOVA) was used to partition the existing genetic variation into different components. In the current study, analysis of one level molecular variance was carried out using the R package Pegas and the number of subpopulations which were determined with STRUCTURE software were used for AMOVA analysis. Variance components obtained by AMOVA was highly significant ($P < 0.001$) with 60% variation among populations (Table.5).

Estimated Mean value of fixation index (Fst) per cluster ranged from 0.3398 in cluster VI that contain progeny plants to 0.7621 in cluster VII, which mainly contains the ILRI collection (Table.6). Divergence among populations was estimated based on allele-frequency or Net nucleotide distance and the largest divergence (0.3777) was between sub-cluster IV and VII (Table. 7).

Table 4: The Nei's genetic distance among the ten sub-clusters of Napier grass collections and progeny plants

	CI	CII	CIII	CIV	CV	CVI	CVII	CVIII	CIX	CX
CI	0	0.20691	0.16473	0.11548	0.1695	**0.07071**	0.43496	0.24225	0.46183	0.18314
CII	0.20691	0	0.29624	0.11065	0.21632	0.24717	0.47953	0.29654	**0.5118**	0.29953
CIII	0.16473	0.29624	0	0.20794	0.18506	0.15965	0.30173	0.14491	0.29771	0.11329
CIV	0.11548	0.11065	0.20794	0	0.13455	0.15303	0.41527	0.21879	0.44283	0.22442
CV	0.1695	0.21632	0.18506	0.13455	0	0.17097	0.20149	0.10591	0.23945	0.16033
CVI	**0.07071**	0.24717	0.15965	0.15303	0.17097	0	0.36253	0.19599	0.36941	0.13015
CVII	0.43496	0.47953	0.30173	0.41527	0.20149	0.36253	0	0.13362	0.08839	0.21433
CVIII	0.24225	0.29654	0.14491	0.21879	0.10591	0.19599	0.13362	0	0.14961	0.11054
CIX	0.46183	**0.5118**	0.29771	0.44283	0.23945	0.36941	0.08839	0.14961	0	0.19898
CX	0.18314	0.29953	0.11329	0.22442	0.16033	0.13015	0.21433	0.11054	0.19898	0

Table 5: AMOVA using 1001 SNPs of the genetic variation among the ten sub clusters (sub populations) of Napier grass accessions and progeny plants,*at 1000 number of permutations for test of hypothesis

Source of variation	SSD	MSD	df	Variance components	Phi statistics	p.value*	CV
pops	13.757165	1.52857390	9	0.05969	0.70824	0.00	
Error	6.295303	0.02459103	256	0.02459			25.195
Total	20.052468	0.07566969	265				

Table 6: Mean values of Fst for the ten sub clusters detected by STRUCTURE analysis

Sub-cluster I	Sub-cluster II	Sub-cluster III	Sub-cluster IV	Sub-cluster V	Sub-cluster VI	Sub-cluster VII	Sub-Cluster VIII	Sub-cluster IX	Sub-cluster X
0.442	0.5041	0.3951	0.6142	0.4168	0.3398	0.7621	0.4052	0.5096	0.5854

Table 7: Divergence among the ten sub-clusters based on allele-frequency or nucleotide distance computed using the STRUCTURE software

	Sub-cluster I	II	III	IV	V	VI	VII	VIII	IX	X
I	-	0.1481	0.144	0.1736	0.1366	0.1395	0.2651	0.1093	**0.0927**	0.2204
II	0.1481	-	0.1545	0.1277	0.1747	0.1228	0.3126	0.1756	0.1455	0.2624
III	0.144	0.1545	-	0.2128	0.1005	0.1015	0.1887	0.097	0.1712	0.1493
IV	0.1736	0.1277	0.2128	-	0.1992	0.168	**0.3777**	0.2168	0.0957	0.3355
V	0.1366	0.1747	0.1005	0.1992	-	0.1354	0.2257	0.0891	0.1647	0.1851
VI	0.1395	0.1228	0.1015	0.168	0.1354	-	0.1645	0.1182	0.1518	0.1244
VII	0.2651	0.3126	0.1887	**0.3777**	0.2257	0.1645	-	0.1797	0.3146	0.1042
VIII	0.1093	0.1756	0.097	0.2168	0.0891	0.1182	0.1797	-	0.1528	0.1333
IX	**0.0927**	0.1455	0.1712	0.0957	0.1647	0.1518	0.3146	0.1528	-	0.2736
X	0.2204	0.2624	0.1493	0.3355	0.1851	0.1244	0.1042	0.1333	0.2736	-

5 5. Discussion

5.1 Marker diversity and genome-wide distribution

In this study, 363 Napier grass collection and progeny plants were considered to assess genetic diversity. However, 347 genotypes were used for the final investigation as 16 genotypes were excluded later from the study due to their high percentage of missing values (≥50%).

DArT-seq markers were utilized to investigate genetic variability and differentiation within and among the 347 Napier grass collection and progeny plants. A total of 96,454 Silico DArT and 96,321 DArT-SNP markers were generated, of which 2, 357 highly informative SNPs were selected for diversity study.

Previously, Napier grass reference genome sequence was not generated, and closely related pearl millet (*Pennisetum glaucum*) genome was used to identify the genomic position and genome-wide distribution of the SilicoDArT and SNP markers (Meki Shehabu *et al.*, 2019). Expected heterozygosity and PIC of the DArT seq generated markers ranged from 0 to 0.5 and 0 to 0.38 for both (silicoDArT and SNP) markers with the Average of He =0.26 and 0.18 and PIC = 0.21 and 0.15, respectively. The results of marker quality parameters were comparable with that of other species. The average PIC values of the generated Silico and SNPs markers was similar with Lesquerella and related species (0.21) (Cruz *et al.*, 2013) and lower than that of values identified in DArT markers of sorghum (0.41) (Emma *et al.*, 2008) and wheat (0.44). Mona *et al.* (2006) also found similar results with PIC value (0.212) of SSR markers that were originally developed for Pearl millet and used to asses genetic diversity in Napier grass due to the cross-species transferability of microsatellite markers (Kandel *et al*., 2016) and slightly greater than the previous report made on Napier grass using similar marker platform (Meki Shehabu *et al.*, 2019).

According to Botstein *et al.* (1980), the PIC values can be classified into highly informative (PIC value > 0.5), moderately informative (PIC value 0.25 ≤ 0.5) and slightly informative (PIC value ≤ 0.25). Therefore, the PIC value of SNP marker is considered as less informative markers due to the bi-allelic nature of the SNP which is restricted to the extreme PIC values of 0.5 (when the two alleles have the same frequencies). Moreover, high distribution of the marker throughout the draft genome were clearly indicated and the obtained result was also consistence with the previous report by Meki Shehabu *et al*. (2019) while slightly higher average He and PIC values were obtained in the current study indicating high marker diversity and distribution that could be attributed to the sample size difference of the studied population and the reference genome used.

5.2 Diversity among Napier grass

The patterns of genetic diversity were revealing in which STRUCTURE software identified two major classes and ten sub-clusters of the assayed genotypes (Figure 4a and 4b). Moreover, the result from hierarchal clustering in complete linkage was in accordance with result from structure analysis (Figure.6). The DAPC analysis resulted in ten K groups (clusters) indicating also highly consistent results with STRUCTURE and Hierarchal clustering for genetic diversity analysis. In addition, AMOVA showed that the two major groups and the ten sub groups detected were significantly different from each other. The high level of diversity (60% variation among populations) was observed and this high variation, and population stratification could be attributed to the outcrossing and self-incompatibility nature of Napier grass (Souza *et al.*, 2019), variation in geographic origin since some of the accessions collected from different parts of the world, selection and breeding system. In Napier grass genetic contribution of genetic drift and gene flow on genetic variation is expected to be low since this grass mostly propagated through stem cutting due to its low seed germination rate (Xie *et al.*, 2009; Wanjala *et al.*, 2013).

In this study, half of the progenies and most of the BAGCE accessions with some accessions from ICRISAT and ILRI were clustered under the main cluster one while most of ILRI collections, CNPGL (EMBRAPA elite lines), USDA collections and half of the progenies also clustered under the second main cluster. Most of the progenies clustered under sub-cluster I, III, VI, and X while most of the ILRI collections were clustered under sub-cluster II, IV, VII, and IX. Majority of the CNPGL and USDA collections were clustered under sub-cluster VIII while most of the BAGCE collections clustered under sub-cluster V. In some cases, progenies with the same maternal plants and clustered under more than one sub-cluster; implying they had different pollen sources. This clustering of the genotypes did not seem to be based on the geographical origin which is also consistence with the findings by Kandel *et al.* (2016); Alemayehu Teressa *et al.* (2018) and Meki Shehabu *et al.* (2019) while it is in contrast with reports made by Lowe *et al*. (2003) and Harris *et al.* (2009).

5.2.1 Population differentiation and genetic divergence

According to Weising *et al*. (2005), the extent of genetic variation in a species and its distribution among and within populations is determined by the interactions of various factors, including the evolutionary history of the species, genetic drift, mating system, gene flow, mutation, and selection. In general, outcrossing species retain most of their variation within populations whereas selfing species allocate more variation among populations. For DNA based markers, among population genetic differentiation is often estimated according to Nei (1972). However, AMOVA is nowadays even more widely used for the partitioning of genetic variation (Excoffier *et al*. 1992).The values of Fst from AMOVA can be grouped into four categories with respect to genetic differentiation; very high (Fst > 0.25), high (0.15 - 0.25), intermediate (0.05 - 0.15), and low (0.0 - 0.05) (De Vicente and Fulton, 2003).

In this study, the estimate of population differentiation in Napier grass collections and progeny plants using fixation index (Fst) revealed relatively high level of differentiation among the populations. Fst measures the degree of differentiation among populations in terms of allele frequencies (De Vicente and Fulton, 2003). Mean fixation index (Fst) per cluster of the current study ranged from 0.3398 in cluster VI that contain entirely progeny plants to 0.7621 in cluster VII that consisted most of the ILRI accessions, some EMBRAPA elite lines and USDA collections. More differentiation was observed among progeny plants and the accessions and this could be attributed to the outcrossing nature of the Napier grass (Souza *et al*., 2019) and the accessions were also collected from different ecological parts of the world.

In accordance with this, the one level AMOVA of the current study estimated the variance among the sub populations and high level genetic differentiation among the sub populations (60%) were revealed. From allele-frequency or Net nucleotide distance, the highest divergent result (0.3777) that was between populations of sub-cluster IV (consisted most ILRI accessions and few of ICRISAT and EMBRAPA's accessions) and VII consisted most of the progenies (2 ILRI accessions and 46 progeny plants), indicating the divergence between accessions and progenies and showed the potential of progeny plants, that are produced from open pollination, for enhancement of diversity of ILRI Napier grass collection with potential for future utilization in breeding program. Most of the introduced accessions were clustered away from the ILRI collections (Fig.6) hence; increase the genetic base for future use.

6 Conclusions and Recommendations

6.1 Conclusions

In this study, DArTseq-SNP markers were employed for genomic diversity and population structure analysis in Napier grass collections from different institutes and progenies raised from 13 ILRI accessions. In the studied Napier grass collections and progenies were obtained as revealed by the large number of SNPs with high distribution throughout the genome and high polymorphism and this is valuable information for genome wide identification of genetic diversity in Napier grass. These molecular markers are useful for detecting unique genotypes to enhance Napier grass diversity of the ILRI collection. Therefore, this genetic diversity information could be used for various purposes, including germplasm conservation, wise utilization, and implementation of effective breeding program.

Moreover, Results from the diversity parameters showed that the presence of up to ten sub populations with considerable variations among the sub populations. Sub-population IV and subpopulation VII were also found as most divergent and genotypes from this sub-population could be used in Napier grass hetrotic breeding program as potential parental plant. On the other hand, this result indicates the suitability of the populations in the future Quantitative Trait Loci (QTL) mapping and Genome wide Association Studies (GWAS) in Napier grass.

6.2 Recommendations

Based on the findings obtained in this study, the following recommendations were forwarded.

- Napier grass breeding program should focus on the divergent genotypes from progenies and accessions and also from distantly related population of ILRI and USDA accessions.
- The ILRI accession from sub-cluster II showed high similarity/low genetic diversity and hence, special attention should be given to conserve these genotypes or study their unique characters.
- Characterization of more populations is also very important and in-depth study should be made in the future including quantitative trait loci (QTL) mapping and genome wide association mapping.

7 References

Akbari, M., Wenzel, P., Ciag, V., Carling, J., Xia, L., Yang, S., Uszynski, G., Mohler, V., Lehmensiek, A., Kuchel, H. and Hayden, M.J.(2006). Diversity array technology (DArT) for high-throughput profiling of the hexaploid wheat genome. *Theor. Appl. Genet.* **113**:1409-1420.

Aktar, W., Sengupta, D. and Chowdhury, A. (2009). Impact of pesticides use in agriculture: Their benefits and hazards. *Interdiscip Toxicol*. **2**:1–12.

Alemayehu Teressa, Jorge, A., Hanson, J., Abel Teshome, Meki Shehabu, Azevedo, A. N. A. L. S., Lédo, F. J. S., Machado, J. C., and Chris, S. (2018). Molecular markers as a tool for germplasm acquisition to enhance the genetic diversity of a Napier grass (*Cenchrus purpureus* syn . *Pennisetum purpureum*) collection. *Trop. grassl.-Forrajes trop.* **6**: 58-69.

Alemayehu Teressa, Abel Teshome, Kumar, A., Hanson, J. and Jones, C. S. (2017). Opportunities for Napier Grass (*Pennisetum purpureum*) Improvement Using Molecular Genetics. *Agron J.* **7**:.28.

Aremu, C.O. (2012). Exploring statistical tools in measuring genetic diversity for crop improvement. J. *Plant Breed. Crop Sci.* p.340-348.

Aremu, C.O., Adebayo, M.A., Ariyo, O.J.and Adewale,B.B. (2007).Classification of genetic diversity and choice of parents for hybridization in cowpea vigna unguiculata (L.) walp for humid savanna ecology. *Afr. J. Biotechnol.* **6**.

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. *Int. J. Mol. Sci.* **11**: 2079-2096.

Aroeira, L.J.M., Lopes, F.C.F., Deresz, F., Vemeque, R.S., Dayrell, M.S., De Matos, L.L., Maldonado-Vasquez, H. and Vittori, A. (1999). Pasture availability and dry matter intake of lactating crossbred cows grazing elephant grass. *Anim. Feed. Sc.i Tech.* **78**:313-324.

Artus Poliakoff, F., Champannet, F. and Gayalin, M. (1991). Fodder production and sheep breeding in Martinique. *Revue d'Elevage et de Medecine veterinaire des pays tropicaux (France).*

Azevedo, A. L. S., Costa, P. P., Machado, J. C., Machado, M. A., Van der Pereira, A. and da Silva Lédo, F. J. (2012). Cross species amplification of *Pennisetum glaucum* microsatellite markers in *Pennisetum purpureum* and genetic diversity of Napier grass accessions. *Crop Sci.* **52**: 1776–1785.

Babu, C., Sundaramoorthi, J., Vijayakumar, G. and Ram, S.G.(2009). Analysis of genetic diversity in Napier grass (*Pennisetum purpureum* Schum) as detected by RAPD and ISSR markers. *J.Plant Biochem.Biotechnol.* **18**:181-187.

Bagali, P.G., Prabhu, P.A.H., Raghavendra, K., Bagali, P.G., Hittalmani, S. and Vadiyelu, J.S. (2010). Application of molecular markers in plant tissue culture.In *Proc. Ssia Pacific Conference on Plant Tissue and Agribiotechnology (APaCPA),***17** :21.

Baloch, F. S., Alsaleh, A., Shahid, M. Q., Miera, D., Aasim, M., Nadeem, A., Akta, H. and Hakan, O. (2017). A Whole Genome DArTseq and SNP Analysis for Genetic Diversity Assessment in Durum Wheat from Central Fertile Crescent. *Plos one.* **12**(1).

Beaumont, M.A., Ibrahim, K.M., Boursot, P. and Bruford, M.W. (1998). Measuring genetic distance. In: **Molecular tools for screening biodiversity**.pp. 315-325. Springer, Dordrecht

Bered, F., Barbosa-Neto, J.F.and de Carvalho, F.I.(2002). Genetic variability in common wheat germplasm based on coefficients of parentage. *Genet. Mol. Biol***. 25**:211-215.

Bhandari, A.P., Sukanya, D.H. and Ramesh, C.R. (2006). Application of isozyme data in fingerprinting Napier grass (Pennisetum purpureum Schum.) for germplasm management. *Genet. Resour. Crop. Evol.* **53**:253-264.

Bhanu, A. N. (2017). Assessment of Genetic Diversity in Crop Plants - An Overview. *Adv. Plants. Agric.Res.* **7**:.279–286.

Bogdan, A.V. (1977). Tropical pasture andfodder crops. *J. Trop. Agric. (Longman London).*

Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet*. **32** : 314.

Burton, G.W. (1990). Grasses: new and improved. In: **Advances in New Crops**. Janick, J. and Simon, J.E.(eds).Timber Press, Portland, Oregon, pp. 174-177.

CABI. (2014). Invasive Species Compendium. Wallingford, UK: CAB International.

Cardoso, J.A., Pineda, M., Jiménez, J.D.L.C., Vergara, M.F. and Rao, I.M. (2015).Contrasting strategies to cope with drought conditions by two tropical forage C4 grasses. *AoB Plants*. **7**.

Chemisquy, M.A., Gussani, L.M., 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. *Ann. Bot.* **106**:107-130.

Clayton, D. and Leung, H.T. (2007).An R package for analysis of whole-genome association studies. *Hum. Hered.* **64**: 45-51.

Clayton, W.D., Govaerts, R., Harman, K.T., Williamson, H. and Vorontsova, M. (2013).World checklist of Poaceae.*Richmond, UK: Royal Botanic Gardens, Kew*.

Corander, J., Cheng, L., Marttinen, P. and Tang, J. (2013). BAPS: Bayesian Analysis of Population Structure. *Department of Mathematics and statistics, University of Helsinki, Finland. Manual*, (**6**) pp. 1-28.

Daher, R.F., Pereira, M.G., Pereira, A.V. and Amaral Jr.AT.D.(2002). Genetic divergence among elephant grass cultivars assessed by RAPD markers in composit samples. *Sci Agric.* **59**: 623-627.

de Favare, H.G., de Abreu, J.G., de Barros, L.V., da Silva, F.G., Ferreira, L.M.M., Barelli, M.A.A., da Silva Neto, I.M., Cabral, C.E.A., Peixoto, W.M., da Silva Campos, F.I. and da Silva Ledo, F.J. (2019). Effect of Elephant grass genotypes to Bioenergy Production. *J. Exp. Agric. Int.* 1-11.

de Morais, R. F., de Souza, B. J., Leite, J. M., de Barros Soares, L. H., Rodrigues Alves, B. J., Boddey, R. M. and Urquiaga, S. (2009). Genótipos de capim-elefante para produção de bioenergia por combustão direta da biomassa. *Pesqui. Agropecu. Bras.* **44**: 133–140.

de Souza, F.R., da Silva Lédo, F.J., Machado, J.C., Machado, M.A., Pereira, A.V., de Oliveira, F.R. and Azevedo, A.L.S. (2019). Estimation of Outcrossing Rate in Napier grass. *Crop Sci.* **59**(3), p. 1030-1036.

de Vicente, M.C. and Fulton, T. (2003). Using Molecular marker technology in studies on plant genetic diversity.*Illus. Nelly Giraldo. IPGRI, Rome, Italy and Institute for Genetic Diversity, Ithaca, New York, USA.*

Do Canto, J., Studer, B. and Lubberstedt, T. (2016). Overcoming self-incompatibility in grasses:a pathway to hybrid breeding. *Theor. Appl. Genet.* **129**(10), p. 1815-1829.

dos Santos, R. L., Freire, F. J., da Rocha, A. T., da Silva, J. A. A., Tavares, J. A., Ferreira, E. G. B. de S. and de Oliveira, E. C. A. (2015). Elephant grass (*Pennisetum purpureum* Schum.) biomass production as promising alternative source of energy in Brazil's semiarid area using gypsum.

Aust. J. Crop Sci. **9**:1082–1088

Duke, J.A. (1983). *Pennisetum purpureum* K. Schumach.*Hand book of energy crops*.

Earl, D.A. and vonHoldt, B.M. (2012). Structure Harvester: A website and program for visualizing Structure output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**:359-361.

Evanno, G., Regnaut, S. and Goudet, J. (2005).Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. **14**: 2611-2620.

Excoffier, L., Smouse, P.E. and Quattro, J.M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics,***131**:479-491.

Falconer, D.S., Mackay, T.F.C. and Frankham, R. (1996).Introduction to quantitative genetics.4th (ed) London. *England: Longman group*, p. 463.

FAO. (2015). Grass land Index. A searchable catalogue of grass and forage legumes. *FAO, Rome. Italy*.

Favare, H. G. De, Abreu, J. G. De, Barros, L. V. de, Silva, F. G. Da, Ferreira, L. M. M., Barelli, M. A. A., Neto, I. M. da S., Cabral, C. E. A., Peixoto, W. M. *et al*.(2019). Effect of Elephant Grass Genotypes to Bioenergy Production. J. Exp. Agric. Int. **38**:1–11.

Flint-garcia, S. A., Thornsberry, J. M. and Iv, E. S. B. (2003). Structure of Linkage Disequilibrium in Plants. *Annu. Rev. Plant Biol.* **54**(1): 357-374.

Francis, J.K. (2004), wildland shrubs of the United States and its territories: Thamnic descriptions, volume 1. *Gen.Tech.Rep.IITF-GTR-26.San Juan, PR: US Department of Agriculture, Forest Service, International Institute of Tropical Forestry; Fort Collins, CO: USDepartment of Agriculutre, Forest Service, Rocky Mountain Research Station*.**830**:26.

Fujii, H., Cao, J. and Managi, S. (2015). Decomposition of Productivity Considering Multi-environmental Pollutants in Chinese Industrial Sector. *Rev. Dev. Econ.* **19**(1):75–84.

*G*öhl, B. (1982). Les aliments du bétail sous les tropiques. FAO, *Division de production et santé Animale, Roma, Italy.*

Gupta, P.K., Varshney, R.K. and Prasad, M. (2002). Molecular markers: principles and methodology. In: **Molecular techniques in crop improvement**. pp. 9-54. *Eds.* (Jain, S. M., Brar, D. S. and Ahloowalia, B. S.), Springer, York, NY, USA.

Gupta, P.K., Varshney, R.K., Sharma, P.C. and Ramesh, B. (1999).Molecular markers nd their applications in wheat breeding. *Plant Breed.* **118**(5):369-390.

Hanna, W.W., Chaparro, C.J., Mathews, B.W., Burns, J.C., Sollenberger, L.E. and Carpenter, J.R. (2004).***Perennial pennisetums*.*Warm-season (C4) grasses,*** **(warmseasonc4 gra)**, pp.503-535. (<u>Lowell, E. Moser, L. Burson, E. Sollenberger</u> Eds.), ASA-CSSA-SSSA, USA

Harris, K., Anderson, W. and Malik, R., (2010). Genetic relationships among Napier grass (*Pennisetum purpureumSchum*.) nursery accessions using AFLP markers. *Plant Genet. Resour***.** **8**(1):63-70.

Henry, R.J. (2001). Commercial applications of. Plant genotyping: the DNA fingerprinting of plants, CABI Publishing, pp. 265-273, Wallingford, Oxon.

Ihaka, R. and Gentleman, R. (1996). R : a language for data analysis and graphics. *J. Comput .Graph. Stat.* **5**(3):299-314.

Jaccourd, D., Peng, K., Feinstein, D. and Kilian, A. (2001). Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res.* **29**(4) :e25-e25.

Jiang, G.L. (2013). Molecular markers and marker-assisted breeding in plants. In: **Plant breeding from laboratories to fields**, pp. 45-83 (Sven Bode A. ed.), InTech, Croatia.

Jin, D., Seo, Y., Saito, T., Suzuki, H. and Ishii, Y. (2012). Ecotoxicology and Environmental Safety Uptake and translocation of cesium-133 in napiergrass (*Pennisetum purpureum* Schum .) under hydroponic conditions. *Ecotoxicol. Environ. Saf.*, **82**: 122–126.

Jombart, T. and Ahmed, I. (2011). A*degenet 1 . 3-1 : new tools for the analysis of genome-wide SNP data*. *Bioinformatics,* **27**(21):3070–3071.

Jombart, T., Lyon, D. and Biome, L. De. (2008). Adegenet : a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**(11):1403-1405.

Kabirizi, J., Nieslsen, S.L., Nicolaisen, M., Byenkya, S. and Alacai, T. (2007). Napier stunt disease in Uganda: farmers' perceptions and impact on fodder production. In 8th*African Crop Science Sociey Conference*, pp. 895-987. El-Minia, Egypt.

Kamvar, Z. N., Tabima, J. F. and Gr, N. J. (2014). Poppr : an R package for genetic analysis of populations with clonal , partially clonal , and / or sexual reproduction. *PeerJ, **2***: e281.

Kandel, R., Singh, H.P., Singh, B.P., Harris-Shultz, K.R. and Anderson, W.F. (2016). Assessment of genetic diversity in Napier grass (Pennisetum purpureum Schum.)using Microsatelite, single nucleotide polymorphism and insertion-deletion markers from millet (*Pennisetum glaucum* [L.] R.Br.). *Plant Mol. Biol. Rep,* **34**(1):.265-272.Katiyu, S. and Mithen, R. (1987).*Pennisetum* in Southern Africa. *Plant Genet. Resour***, 73**(74): 1-8.

Khan, Z.R., Midega, C.A., Wadhams, L.J., Pickett, J.A.and Mumuni, A. (2007). Evaluation of Napier grass (Pennisetum purpureum) varieties for use as trap plants for management of African stemborer (Busseola fusca) in a push-pull strategy. *Entomol.Exp.Appl.*, **124**(2):201-211.

Knaus, B. and Winter, D. (2020). *Package ' pegas'*

Labate, J.A. (2000). Software for population genetic analysis of Molecular data. *Crop Sci.*, **40**(6):1521-1528.

Lotfy, S. M. and Mostafa, A. Z. (2014). Phytoremediation of contaminated soil with cobalt and chromium. *J. Geochem. Explor*, **144***:*367–373.

Mace, E.S., Xia, L., Jprdan, D.R., Halloran, K., parh, D.K., Huttner, E., Wenzl, P. and Kilian, A. (2008). DArT markers: diversity analysis and mapping in Sorghum bicolor. *BMC genom*, **9**(1):1-11.

Mannetje, L.'t.(1992). *Pennisetum purpurreum Schumach.Record from proseabase*. Mannetje L.;t and Jones R.M.(Eds). PROSEA (plant Resources of South-east Asia) foundation, Bogor, Indonesia

Martin, J.M., Bllake, T.K. and Hockett, E.A. (1991). Diversity among North American spring barley cultivars based on coefficients of parentage. *Crop Sci.*, **31**(5):1131-1137.

Meki Shehabu, Teshome, A., Hanson, J., Negawo, A.T., Habte, E., Entfellner, J.B.D., Lee, K.W. and Jones, C.S. (2019). Genotyping by sequencing provides new insights into the diversity of Napier grass (*Cenchrus purpureus*) and reveals variation in genome-wide LD patterns between collections. *Sci. Rep*, **9**(1):6936.

Meirmans, P. G. and Hedrick, P. W. (2011). Assessing population structure: FST and related measures. *Mol. Ecol. Resour.*, **11**(1):5–18.

Mohammadi, S.A. and Prasanna, B.M. (2003). Analysis of genetic diversity in crop plantssalient statisticaltoolsandconsiderations. *Crop Sci.,* **43(**4):1235-1248.

Mondini, L., Noorani, A. and Pagnotta, M.A. (2009).Assessing plant gentic diversity by Molecular tools. *Diversity (Basel),* **1**(1):19-35.

Moran, J.B. (2011).Factors affecting high mortality rates of dairy replacement calves and heifers in the tropics and strategies for their reduction. *Asian-australas. J. Anim. Sci*, **24**(9):1318-1328.

Mulaa, M., Kabirizi, J., Pallangyo, B., Hanson, J., Proud, J., Mukiibi, E., Maeda, C., Wanjala, B.Awalla, B.J. and Namazzi, C. (2013).Diversity, biomass and resistance to stunt in Napier grass clones in East and Central Africa region. In: Enhancing the competitiveness of the livelihoods in Eastern and Central Africa (Ndikumana, J.; Mubiru, S.; Zziwa, E. and Tenywa, J.S. eds)..Association for Strengthening Agricultural Research in Eastern and Central Africa, Entebbe, pp. 27-34, Uganda.

Mushtaque, M., Ishaque, M. and Ahamd, M. (2010). Growth and herbage yield of Setaria sphacelata grass in response to varying clipping stages. *J. Anim. Plant Sci*, **20**(4):261-265.

Nei, M. (1972). Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci*, **70**(12):3321-3323.

Nielsen, N.H.,Backes, G., Stougaard, J.;,Andersen, S.U. and Jahoor, A.(**2014).** Genetic diversity and population structure Analysis of European hexaploid bread wheat (*Triticum aestivum* L.) varieties. *PLoS*. **9**: e94000

Nyambati, E.M., Muyekho, F.N., Onginjo, E. and Lusweti, C.M. (2010). Productioon, characterization and nutritional quality of Napier grass [Pennisetum purpureum (Schum.)] cultivars in western Kenya. *Afr. J. plant Sci.*, **4**(12):496-502.

O., C. (2012). Exploring Statistical Tools in Measuring Genetic Diversity for Crop Improvement. *Genet. Diver. Plants*, 340-348.

Oliveira, M. L. F., Daher, R. F., Menezes, B. R. S. and Vivas, M. (2017). Genetic diversity of elephant grass (Cenchrus purpureus [Schumach .] Morrone) for energetic production based on quantitative and multi-category traits. *Chil. J. Agric. Res.*, **77**(1):48-57.

Pagnotta, M. (2018). Comparison among Methods and Statistical Software Packages to Analyze

Germplasm Genetic Diversity by Means of Codominant Markers. *J*, **1**(1): 197-215.

Pallangyo, B., Maeda, C. and Mkonyi, S.(2008). Napier grass (Pennisetum purpurium Schum.) diversity, uses and diseases in Eastern, Northern and lake zones of Tanzania. In: 1st National Plant Protection Advisory Committee ad hoc meeting, sugarcane research institute, Kibaha Tanzania. 19th September (pp. 38-66).

Paradis, E. (2010). pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics*, **26**(3):419-420.

Paradis, E., Blomber, S., Bolker, B., Brown, J., Claude, J., Cuong, Hoa, S., Desper, R., Didier, G., Durand, B., Dutheil, J., Ewing, R., Gascuel, O., Guillerme, T., Heibl, C., Ives, A., Jones, B., Krah, F., Lawson, D., Lefort, V., de Vienne, D. (2019). Package "ape": Analyses of Phylogenetics and Evolution, version **2**(4).

Paradis, E., Claude, J. and Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics,* **20**(2):289–290.

Paudel, D., Kannan, B., Yang, X., Harris-shultz, K., Thudi, M., Varshney, R. K., Altpeter, F. and Wang, J. (2018). Surveying the genome and constructing a high-density genetic map of napiergrass (Cenchrus purpureus Schumach). *Sci. Rep.,* **8***(*1)–11.

Pereiara, A.V., Auad, A.M., Lédo, F.J.S. and Barbosa, S. (2010). Pennisetum perpureum. In: Fonseca, D.M. and Martuscello, J.A (ed) Plantas forrageiras. Editora UFV, Vicosa, MG. pp. 197-219.

Pereira, A.V., Ledo, F.D.S., Morenz, M.J.F., Leite, J.L.B., Brighenti, A.M., Martins, C.E. and Machado, J.C. (2016). BRS Capiacu: cultivar de capim-elefante de alto rendimento, para produção de silage, Embrapa Gada de Leite-Comunicado T*écnico (INFOTECA-E).*

Pritchard, J. K. (2009). Documentation for structure software*: Version 2 . 3*. University of Chicago,*USA.*

Pritchard, J. K., Stephens, M. and Donnelly, P. Assig test_Pritchard2000.pdf.

Qiagen. (2016). *DNeasy® Plant Mini Kit*. *March*, pp.1–2. https://doi.org/10.1016/j.imavis.2016.05.001

Rajpurohit, D. and Jhang, T., 2015.In situ and ex situ conservation of plant genetic resources and traditional knowledge. In: **Plant Genetic Resources and Traditional Knowledge for Food Security.**(pp. 137-162). Springer, Singapore.

Ramsey, J. and Schemske, D.W. (2002).Neopolyploidy in flowering plants. *Annu. Rev. Ecol Evol Syst.*, **33**(1):589-639.

Rj, G.-, Lorraine, M., Oliveira, F., Daher, R. F., Amara, G. De, Brito, V., Viana, A. P., Rodrigues, E. V., Shimoya, A., Teixeira, A. and Rafaela, B. (2014). Pre-breeding of elephant grass for energy purposes and biomass analysis in Campos *dos* Goytacazes-RJ, Brazil *Afr.J.Agri.Res.,* **9**(36), 2743–2758.

Rogers, J.S. (1972). Studies in genetics VII.*University of Texas Publication* 7213, p. 14153.

Rohlf, F.J. (1998). NTSYS-PC: Numerical taxonomy and multivariate analysis system. Version 2. *Applied Biostatistics, Inc. New York, NY.*

Romero, J., Carvalho, S. De, Marçal, T. D. S., Salvador, F. V., Carlos, A., Crescencio, P., Carneiro, S., Deon, M., Resende, V. De, Carneiro, C., Luisa, A., Azevedo, S. and Pereira, J. F. (2019). Unraveling candidate genes underlying biomass digestibility in elephant grass *(Cenchrus purpureus).MBC Plant Biol,* **19**(1): 1–12.

Roslan, A. M., Ibrahim, M. F. and Hassan, M. A. (2020). Potential use of Pennisetum purpureum for phytoremediation and bioenergy production : a mini review.*Sci. Rep,* **10**:6613.

Russell, J.R., Fuller, J.D., Macaulay, M., Hatz, B.G., Jahoor, A., Powell, W. and Waugh, R. (1997).Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. *Theor.Appl.Genet.*, **95**(4):714-722.

Schliep, K.P. (2011).Phangorn: phylogenetic analysis in R. *Bioinformatics*, **27**:592e593.

Singh, B.P., Singh, H.P. and Obeng, E. (2013).Elephant grass. In: **Biofuel crops: Production, physiology and genetics**, pp.271-291, (Signh B.P; CAB inetrnational Eds.),Front valley State University, GA, USA.

Singh,R.K., Mishra, S.K., Singh, S.P., Mishra, N. and Sharma, M.L. (2010). Evaluation of microsatellite markersfor genetic diversity analysis among sugarcane species and commercial hybrids. *Aust. J. Crop Sci.*, **4**(2):116.

Skerman, P.J. and Riveros, F. (1990).Tropical grasses. FAO Plant Production and Protection Series. FAO, Rome, Italy.

Smith, R.L., Schweder, M.E., Chowdhury, M.K.U., Seib, J.C. and Schank, S.C. (1993).Development and application of RFLP and RAPD DNA markers in genetic improvement of *Pennisetum* for biomass and forage production. *Biomass and Bioenergy*, **5**(1):51-62.

Souza, F. R. De, José, F. and Machado, J. C. (2019). Estimation of Outcrossing Rate in Napier Grass. *Crop Sci.*, **59**(3):1030–1036.

Staal, S.J., Chege, L., Kenyanjui, M., Kimari, A., Lukuyu, B.A., Njubi, D., Owango, M.O., Tanner, J.C., Thorpe, W.R. and Wambugu, M. (1997). Characterisation of dairy systems supplying the Nairobi milk market: a pilot survey in Kiambu district for the identification of target groups of producers.

Stapf, O., Hubbard, C.E. (1934). Pennisetum. In: prain, D. (ed) Flara of Tropical Africa, Vol 9, 6, pp. 954-1070, Reeve and Co, Ltd, Ashford, Kent, England.

Tananonchai, A. and Sampanpanish, P. (2018). Phytotolerance, Phytotoxicity and Phytoremediation Of Cd And EDTA Mixtures With Napier Grass. *Environ. Asia*. **11**(1): 157–167.

Taylor, S.H., Ripley, B.S., Woodward, F.I. and Osborne, C.P. (2011). Drought limitation of photosynthesis differs between C3 and C4 grass species in a comparative experiment. *Plant Cell Environ.*, **34**(1):65-75.

Tiwary, R. K. and Dhar, B. B. (1994). Environmental Pollution From Coal Mining Activities in Damodar River Basin, India. *Mine Water Environ.*, **13***(*1):1-10.

Van Mark, V.C., Kilian, A. and Dierig, D.A. (2013). Development of DArT marker platforms and genetic diversity assessment of the US collection of the new oilseed crop Lesquerella and related species. *PLos.one*, **8**(5) : e64062.

Van De Wouw, M., Hanson, J. and Luethi, S. (1999). Morphological and agronomic characterisation of a collection of napier grass (Pennisetum purpureum) and P. purpureum x P. glaucum. *Trop. Grassl.Forrajes trop.*, **33**(3):150–158.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T.V.D., Hornes, M., Friters, A., Pot, J., Paleman, J., Kuiper, M. and Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic acids Res.*, **23**(21):4407-4414.

Wanjala, B.W., Obonyo, M., Wachira, F.N., Muchugi, A., Mulaa, M., Harvey, J., Skilton, R.A., Proud, J. and Hanson, J. (2013). Genetic diversity in Napier grass (*Pennisetum purpureum*) cultivars: implications for breeding and conservation. *AoBPlants*,*5*.

Weir, S.B. (1996). Genetic data analysis II. Methods For Discrete Population Genetic Data. Sinauer Associates Inc., Sunderland.

Weising, K., Nybom, H., Wolff, K. and Kahl, G. (2005).**DNA Fingerprinting in Plants: Principles, Methods, and Applications** 2nd edn. pp. 444, CRC Press, New York..

Wenzl, P., Carling, J., Kudrna, D., Jaccoud, D., Huttner, E., Kleinhofs, A., & Kilian, A. (2004). Diversity Arrays Technology (DArT) for whole-genome profiling of barley. *Proc. Natl. Acad. Sci*, **101**(26): 9915–9920.

Williams, J.G., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic acids Res*, **18**(22):6531-6535.

Wimmer, A. V., Albrecht, T., Auinger, H., Schoen, C., Schaeffer, L., Erbe, M., Ober, U. and Reimer, C. (2012). *Package ' synbreed.*

Wright, S. (1950). Genetical structure of populations. *Nature*, **166**(4215):247-249.

Xie, X.M., Zhou, F., Zhang, X.Q. and Zhang, J.M. (2009). Genetic variability and relationship between MT-1 elephant grass and closely related cultivars assessed by SARP markers. *J.Genet.*, **88**(3):281-290.

Yasuda, M., Takeo, K., Nagai, H., Uto, T., Yui, T., Matsumoto, T., Ishii, Y. and Ohta, K. (2013). Enhancement of ethanol production from napiergrass (Pennisetum purpureum Schumach) by a low-moisture anhydrous ammonia pretreatment. *J. Sustain. Bioenergy Syst*, **3**(3):179.

Zailan, M.Z., Yaakub, H. and Jusoh, S. (2016). Yield and nutritive value of four Napier (Pennisetum purpureum) cultivars at different harvesting ages. *Agri. Biol. J.*, **7**: 213-219.

Zhou, S., Wang, C., Frazier, T. P., Yan, H. and Chen, P. (2018). The first Illumina-based de novo transcriptome analysis and molecular marker development in Napier grass (*Pennisetum purpureum*). *Mol. breed*, **38**(7):95

Appendices

Appendix 1: (Supplementary table 1, S1) List of Napier grass genotypes that were used in the study. The 13 ILRI accessions that were used to raise progenies and the progenies raised from them indicated by the same highlighted color in respective to their progenies. The 16 genotypes that were excluded from further analysis were yellow color highlighted.

No.	Acc. No.	Genus	Species/ cross	Origin	Collection	year acquired
1	ILRI_1026	Cenchrus	purpureus	Burundi	ILRI	1986
2	ILRI_14355	Cenchrus	purpureus	Ethiopia	ILRI	1985
3	ILRI_14389	Cenchrus	purpureus	Nigeria	ILRI	1985
4	ILRI_14982	Cenchrus	purpureus x glaucum	USA	ILRI	1986
5	ILRI_14983	Cenchrus	purpureus	USA	ILRI	1986
6	ILRI_14984	Cenchrus	purpureus	USA	ILRI	1986
7	ILRI_15357	Cenchrus	purpureus x glaucum	NA	ILRI	1986
8	ILRI_15743	Cenchrus	purpureus	USA	ILRI	1988
9	ILRI_16621	Cenchrus	purpureus	Namibia	ILRI	1991
10	ILRI_16782	Cenchrus	purpureus	Tanzania	ILRI	1992
11	ILRI_16783	Cenchrus	purpureus	Tanzania	ILRI	1992
12	ILRI_16784	Cenchrus	purpureus	Tanzania	ILRI	1992
13	ILRI_16785	Cenchrus	purpureus	Tanzania	ILRI	1992
14	ILRI_16786	Cenchrus	purpureus	Swaziland	ILRI	1992
15	ILRI_16787	Cenchrus	purpureus	Swaziland	ILRI	1992
16	ILRI_16788	Cenchrus	purpureus	Swaziland	ILRI	1992
17	ILRI_16789	Cenchrus	purpureus	Swaziland	ILRI	1992
18	ILRI_16791	Cenchrus	purpureus	Swaziland	ILRI	1992
19	ILRI_16792	Cenchrus	purpureus	Mozambique	ILRI	1992
20	ILRI_16793	Cenchrus	purpureus	Cuba	ILRI	1992
21	ILRI_16794	Cenchrus	purpureus	Mozambique	ILRI	1992
22	ILRI_16795	Cenchrus	purpureus	Zimbabwe	ILRI	1992
23	ILRI_16796	Cenchrus	purpureus	Zimbabwe	ILRI	1992
24	ILRI_16797	Cenchrus	purpureus	Zimbabwe	ILRI	1992

25	ILRI_16798	Cenchrus	purpureus	Zimbabwe	ILRI	1992
26	ILRI_16799	Cenchrus	purpureus	Zimbabwe	ILRI	1992
27	ILRI_16800	Cenchrus	purpureus	Zimbabwe	ILRI	1992
28	ILRI_16801	Cenchrus	purpureus	Zimbabwe	ILRI	1992
29	ILRI_16802	Cenchrus	purpureus	Zimbabwe	ILRI	1992
30	ILRI_16803	Cenchrus	purpureus	Zimbabwe	ILRI	1992
31	ILRI_16804	Cenchrus	purpureus	USA	ILRI	1992
32	ILRI_16805	Cenchrus	purpureus	USA	ILRI	1992
33	ILRI_16806	Cenchrus	purpureus	USA	ILRI	1992
34	ILRI_16807	Cenchrus	purpureus	USA	ILRI	1992
35	ILRI_16808	Cenchrus	purpureus	USA	ILRI	1992
36	ILRI_16809	Cenchrus	purpureus	USA	ILRI	1992
37	ILRI_16810	Cenchrus	purpureus	USA	ILRI	1992
38	ILRI_16811	Cenchrus	purpureus	USA	ILRI	1992
39	ILRI_16812	Cenchrus	purpureus	USA	ILRI	1992
40	ILRI_16813	Cenchrus	purpureus	USA	ILRI	1992
41	ILRI_16814	Cenchrus	purpureus	USA	ILRI	1992
42	ILRI_16815	Cenchrus	purpureus	USA	ILRI	1992
43	ILRI_16816	Cenchrus	purpureus	USA	ILRI	1992
44	ILRI_16817	Cenchrus	purpureus	USA	ILRI	1992
45	ILRI_16818	Cenchrus	purpureus	USA	ILRI	1992
46	ILRI_16819	Cenchrus	purpureus	USA	ILRI	1992
47	ILRI_16821	Cenchrus	purpureus	Zimbabwe	ILRI	1992
48	ILRI_16822	Cenchrus	purpureus	Malawi	ILRI	1992
49	ILRI_16834	Cenchrus	purpureus x glaucum	Zimbabwe	ILRI	1992
50	ILRI_16835	Cenchrus	purpureus x glaucum	Zimbabwe	ILRI	1992
51	ILRI_16836	Cenchrus	purpureus	Zimbabwe	ILRI	1992
52	ILRI_16837	Cenchrus	purpureus x glaucum	Zimbabwe	ILRI	1992
53	ILRI_16838	Cenchrus	purpureus x glaucum	Zimbabwe	ILRI	1992
54	ILRI_16839	Cenchrus	purpureus	Zimbabwe	ILRI	1992
55	ILRI_16840	Cenchrus	purpureus x glaucum	Zimbabwe	ILRI	1992

56	ILRI_16902	Cenchrus	purpureus	Zimbabwe	ILRI	1992
57	ILRI_16790	Cenchrus	purpureus	Swaziland	ILRI	1992
58	ILRI_18438	Cenchrus	purpureus	Tanzania	ILRI	1995
59	ILRI_18448	Cenchrus	purpureus	Tanzania	ILRI	1995
60	ILRI_18662	Cenchrus	purpureus x glaucum	South_Africa	ILRI	2006
61	BAGCE-1	Cenchrus	purpureus	Colombia	EMBRAPA_collection	1976
62	BAGCE-100	Cenchrus	purpureus	Brazil	EMBRAPA_collection	1993
63	BAGCE-16	Cenchrus	purpureus	Brazil	EMBRAPA_collection	1976
64	BAGCE-17	Cenchrus	purpureus	Costa Rica	EMBRAPA_collection	1976
65	BAGCE-22	Cenchrus	purpureus	NA	EMBRAPA_collection	1976
66	BAGCE-24	Cenchrus	purpureus	NA	EMBRAPA_collection	1976
67	BAGCE-25	Cenchrus	purpureus	India	EMBRAPA_collection	1976
68	BAGCE-30	Cenchrus	purpureus	Brazil	EMBRAPA_collection	1976
69	BAGCE-343	Cenchrus	purpureus	Brazil	EMBRAPA_collection	1976
70	BAGCE-53	Cenchrus	purpureus	Brazil	EMBRAPA_collection	1976
71	BAGCE-56	Cenchrus	purpureus	Brazil	EMBRAPA_collection	1989
72	BAGCE-63	Cenchrus	purpureus	Cuba	EMBRAPA_collection	1991
73	BAGCE-7	Cenchrus	purpureus	Brazil	EMBRAPA_collection	1976
74	BAGCE-75	Cenchrus	purpureus	Brazil	EMBRAPA_collection	1992
75	BAGCE-80	Cenchrus	purpureus	Brazil	EMBRAPA_collection	1992
76	BAGCE-81	Cenchrus	purpureus	Brazil	EMBRAPA_collection	1992
77	BAGCE-86	Cenchrus	purpureus	NA	EMBRAPA_collection	1992
78	BAGCE-90	Cenchrus	purpureus	NA	EMBRAPA_collection	NA
79	BAGCE-94	Cenchrus	purpureus	NA	EMBRAPA_collection	1993
80	BAGCE-97	Cenchrus	purpureus	NA	EMBRAPA_collection	1993
81	CNPGL_00-1-1	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
82	CNPGL_91-06-2	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
83	CNPGL_91-11_-2	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
84	CNPGL_91-25-1	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
85	CNPGL_92-133-3	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
86	CNPGL_92-198-7	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA

87	CNPGL_92-190-1	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
88	CNPGL_92-38-2	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
89	CNPGL_92-56-2	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
90	CNPGL_92-66-3	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
91	CNPGL_9279-2	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
92	CNPGL_93-01-1	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
93	CNPGL_93-04-2	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
94	CNPGL_93-06-1	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
95	CNPGL_93-08-1	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
96	CNPGL_93-18-2	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
97	CNPGL_93-32-2	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
98	CNPGL_93_-37-5	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
99	CNPGL_94-07-2	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
100	CNPGL_94-13-1	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
101	CNPGL_96-21-1	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
102	CNPGL_96-23-1	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
103	CNPGL_96-24-1	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
104	CNPGL_96-27-3	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
105	PIONEIRO	Cenchrus	purpureus	NA	EMBRAPA_elite_lines	NA
106	N 8	Cenchrus	purpureus K-12	PUERTO RICO	USDA	1975
107	N 19	Cenchrus	purpureus	PUERTO RICO	USDA	NA
108	N 23	Cenchrus	purpureus	SWAZILAND	USDA	NA
109	N 36	Cenchrus	purpureus	SCHANK	USDA	NA
110	N 37	Cenchrus	purpureus	SCHANK	USDA	NA
111	N 43	Cenchrus	purpureus	NA	USDA	NA
112	N 68	Cenchrus	purpureus	SCHANK	USDA	NA
113	N 71	Cenchrus	purpureus	SCHANK	USDA	NA
114	N 75	Cenchrus	purpureus	NA	USDA	1977
115	N109	Cenchrus	Purpureus (SPAIN NAPIER)	NA	USDA	1979
116	N128	Cenchrus	81-D62-1 (Dwarf Napier plant)	NA	USDA	NA
117	N130	Cenchrus	Purpureus (NB21 Museum Plot)		USDA	NA

118	N131	Cenchrus	purpureus	NA	USDA	NA
119	N137	Cenchrus	purpureus	BATORE	USDA	NA
120	N138	Cenchrus	purpureus	BATORE	USDA	NA
121	N147	Cenchrus	purpureus	NA	USDA	1983
122	N172	Cenchrus	purpureus	NA	USDA	NA
123	N200	Cenchrus	purpureus	NA	USDA	F2_1986
124	N210	Cenchrus	purpureus	NA	USDA	F2_1986
125	N223	Cenchrus	purpureus	NA	USDA	1988
126	N225	Cenchrus	purpureus	NA	USDA	1988
127	N228	Cenchrus	purpureus	NA	USDA	1988
128	145_1_21965	Cenchrus	purpureus	NA	ICRISAT	2019
129	120_22238	Cenchrus	purpureus	NA	ICRISAT	2019
130	127_22230	Cenchrus	purpureus	NA	ICRISAT	2019
131	117_22240	Cenchrus	purpureus	NA	ICRISAT	2019
132	149_21785	Cenchrus	purpureus	NA	ICRISAT	2019
133	126_22231	Cenchrus	purpureus	NA	ICRISAT	2019
134	150_21784	Cenchrus	purpureus	NA	ICRISAT	2019
135	118_22241	Cenchrus	purpureus	NA	ICRISAT	2019
136	128_22229	Cenchrus	purpureus	NA	ICRISAT	2019
137	124_22233	Cenchrus	purpureus	NA	ICRISAT	2019
138	125_22232	Cenchrus	purpureus	NA	ICRISAT	2019
139	115_22243	Cenchrus	purpureus	NA	ICRISAT	2019
140	119_22239	Cenchrus	purpureus	NA	ICRISAT	2019
141	146_21788	Cenchrus	purpureus	NA	ICRISAT	2019
142	132_22225	Cenchrus	purpureus	NA	ICRISAT	2019
143	144_21964	Cenchrus	purpureus	NA	ICRISAT	2019
144	141_21967	Cenchrus	purpureus	NA	ICRISAT	2019
145	142_21966	Cenchrus	purpureus	NA	ICRISAT	2019
146	131_22226	Cenchrus	purpureus	NA	ICRISAT	2019
147	147_21787	Cenchrus	purpureus	NA	ICRISAT	2019
148	151_21783	Cenchrus	purpureus	NA	ICRISAT	2019

149	121_22237	Cenchrus	purpureus	NA	ICRISAT	2019
150	129_22228	Cenchrus	purpureus	NA	ICRISAT	2019
151	123_22234	Cenchrus	purpureus	NA	ICRISAT	2019
152	89_22236	Cenchrus	purpureus	NA	ICRISAT	2019
153	136_21968	Cenchrus	purpureus	NA	ICRISAT	2019
154	116_22242	Cenchrus	purpureus	NA	ICRISAT	2019
155	G1 (Gaint Napier)	Cenchrus	purpureus	NA	ICRISAT	2019
156	Ns 1-1	Cenchrus	purpureus	ILRI_collection_collection	ILRI	2019
157	NS 1_2	Cenchrus	purpureus	ILRI_collection	ILRI	2019
158	NS 1_3	Cenchrus	purpureus	ILRI_collection	ILRI	2019
159	NS 1_4	Cenchrus	purpureus	ILRI_collection	ILRI	2019
160	NS 1_5	Cenchrus	purpureus	ILRI_collection	ILRI	2019
161	NS 1_6	Cenchrus	purpureus	ILRI_collection	ILRI	2019
162	NS 1_9	Cenchrus	purpureus	ILRI_collection	ILRI	2019
163	NS 1_10	Cenchrus	purpureus	ILRI_collection	ILRI	2019
164	NS 1_11	Cenchrus	purpureus	ILRI_collection	ILRI	2019
165	NS 1_12	Cenchrus	purpureus	ILRI_collection	ILRI	2019
166	NS 1_13	Cenchrus	purpureus	ILRI_collection	ILRI	2019
167	NS 1_14	Cenchrus	purpureus	ILRI_collection	ILRI	2019
168	NS 1_15	Cenchrus	purpureus	ILRI_collection	ILRI	2019
169	NS 1_16	Cenchrus	purpureus	ILRI_collection	ILRI	2019
170	NS 1_18	Cenchrus	purpureus	ILRI_collection	ILRI	2019
171	NS 2_1	Cenchrus	purpureus	ILRI_collection	ILRI	2019
172	NS 2_2	Cenchrus	purpureus	ILRI_collection	ILRI	2019
173	NS 2_3	Cenchrus	purpureus	ILRI_collection	ILRI	2019
174	NS 2_6	Cenchrus	purpureus	ILRI_collection	ILRI	2019
175	NS 2_7	Cenchrus	purpureus	ILRI_collection	ILRI	2019
176	NS 2_8	Cenchrus	purpureus	ILRI_collection	ILRI	2019
177	NS 2_9	Cenchrus	purpureus	ILRI_collection	ILRI	2019
178	NS 3_1	Cenchrus	purpureus	ILRI_collection	ILRI	2019
179	NS 3_2	Cenchrus	purpureus	ILRI_collection	ILRI	2019

180	NS 3-3	Cenchrus	purpureus	ILRI_collection	ILRI	2019
181	NS 3-4	Cenchrus	purpureus	ILRI_collection	ILRI	2019
182	NS 3-5	Cenchrus	purpureus	ILRI_collection	ILRI	2019
183	NS 3-6	Cenchrus	purpureus	ILRI_collection	ILRI	2019
184	NS 3-7	Cenchrus	purpureus	ILRI_collection	ILRI	2019
185	NS 3-8	Cenchrus	purpureus	ILRI_collection	ILRI	2019
186	NS 3-9	Cenchrus	purpureus	ILRI_collection	ILRI	2019
187	NS 3-10	Cenchrus	purpureus	ILRI_collection	ILRI	2019
188	NS 3-11	Cenchrus	purpureus	ILRI_collection	ILRI	2019
189	NS 3-12	Cenchrus	purpureus	ILRI_collection	ILRI	2019
190	NS 3-13	Cenchrus	purpureus	ILRI_collection	ILRI	2019
191	NS 3-14	Cenchrus	purpureus	ILRI_collection	ILRI	2019
192	NS 3-15	Cenchrus	purpureus	ILRI_collection	ILRI	2019
193	NS3-16	Cenchrus	purpureus	ILRI_collection	ILRI	2019
194	NS 3-17	Cenchrus	purpureus	ILRI_collection	ILRI	2019
195	NS 3-18	Cenchrus	purpureus	ILRI_collection	ILRI	2019
196	NS 3-19	Cenchrus	purpureus	ILRI_collection	ILRI	2019
197	NS 3-20	Cenchrus	purpureus	ILRI_collection	ILRI	2019
198	NS 4-1	Cenchrus	purpureus	ILRI_collection	ILRI	2019
199	NS 4-2	Cenchrus	purpureus	ILRI_collection	ILRI	2019
200	NS 4-3	Cenchrus	purpureus	ILRI_collection	ILRI	2019
201	NS 4-4	Cenchrus	purpureus	ILRI_collection	ILRI	2019
202	NS 4-5	Cenchrus	purpureus	ILRI_collection	ILRI	2019
203	NS 4-6	Cenchrus	purpureus	ILRI_collection	ILRI	2019
204	NS 4-7	Cenchrus	purpureus	ILRI_collection	ILRI	2019
205	NS 4-9	Cenchrus	purpureus	ILRI_collection	ILRI	2019
206	NS 4-10	Cenchrus	purpureus	ILRI_collection	ILRI	2019
207	NS 4-11	Cenchrus	purpureus	ILRI_collection	ILRI	2019
208	NS 4-12	Cenchrus	purpureus	ILRI_collection	ILRI	2019
209	NS 4-13	Cenchrus	purpureus	ILRI_collection	ILRI	2019
210	NS 4-14	Cenchrus	purpureus	ILRI_collection	ILRI	2019

211	NS 4_15	Cenchrus	purpureus	ILRI_collection	ILRI	2019
212	NS 4-18	Cenchrus	purpureus	ILRI_collection	ILRI	2019
213	NS 4- 19	Cenchrus	purpureus	ILRI_collection	ILRI	2019
214	NS 4-20	Cenchrus	purpureus	ILRI_collection	ILRI	2019
215	NS 5-1	Cenchrus	purpureus	ILRI_collection	ILRI	2019
216	NS 5-2	Cenchrus	purpureus	ILRI_collection	ILRI	2019
217	NS 5-3	Cenchrus	purpureus	ILRI_collection	ILRI	2019
218	NS 5 -4	Cenchrus	purpureus	ILRI_collection	ILRI	2019
219	NS 5-5	Cenchrus	purpureus	ILRI_collection	ILRI	2019
220	NS 5-6	Cenchrus	purpureus	ILRI_collection	ILRI	2019
221	NS 5-8	Cenchrus	purpureus	ILRI_collection	ILRI	2019
222	NS 5-9	Cenchrus	purpureus	ILRI_collection	ILRI	2019
223	NS 5-10	Cenchrus	purpureus	ILRI_collection	ILRI	2019
224	NS 5-12	Cenchrus	purpureus	ILRI_collection	ILRI	2019
225	NS 5-13	Cenchrus	purpureus	ILRI_collection	ILRI	2019
226	NS 5-14	Cenchrus	purpureus	ILRI_collection	ILRI	2019
227	NS 5-15	Cenchrus	purpureus	ILRI_collection	ILRI	2019
228	NS 5-16	Cenchrus	purpureus	ILRI_collection	ILRI	2019
229	NS 5-17	Cenchrus	purpureus	ILRI_collection	ILRI	2019
230	NS 5-18	Cenchrus	purpureus	ILRI_collection	ILRI	2019
231	NS 5-19	Cenchrus	purpureus	ILRI_collection	ILRI	2019
232	NS 6-1	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
233	NS 6_2	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
234	NS 6-3	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
235	NS 6-4	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
236	NS 6- 5	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
237	NS 6-6	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
238	NS 6-7	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
239	NS 6-8	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
240	NS 6-9	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
241	NS 6-10	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019

242	NS 6-11	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
243	NS 6-12	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
244	NS 6-13	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
245	NS 6-14	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
246	NS 6-15	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
247	NS 6-16	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
248	NS 6-18	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
249	NS 6-19	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
250	NS 6-20	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
251	NS 7-1	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
252	NS 7-2	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
253	NS 7-3	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
254	NS 7-4	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
255	NS 7-5	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
256	NS 7_6	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
257	NS 7-7	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
258	NS 7-8	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
259	NS 7-9	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
260	NS 7-10	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
261	NS 7-11	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
262	NS 7-12	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
263	NS 7-13	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
264	NS 7-14	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
265	NS 7-15	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
266	NS 7-17	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
267	NS 7-18	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
268	NS 7-20	Cenchrus	purpureus x glaucum	ILRI_collection	ILRI	2019
269	NS 8_2	Cenchrus	purpureus	ILRI_collection	ILRI	2019
270	NS 8_3	Cenchrus	purpureus	ILRI_collection	ILRI	2019
271	NS 8_4	Cenchrus	purpureus	ILRI_collection	ILRI	2019
272	NS 8_7	Cenchrus	purpureus	ILRI_collection	ILRI	2019

273	NS 8_8	Cenchrus	purpureus	ILRI_collection	ILRI	2019
274	NS 8_9	Cenchrus	purpureus	ILRI_collection	ILRI	2019
275	NS 8_10	Cenchrus	purpureus	ILRI_collection	ILRI	2019
276	NS 8_11	Cenchrus	purpureus	ILRI_collection	ILRI	2019
277	NS 8_14	Cenchrus	purpureus	ILRI_collection	ILRI	2019
278	NS 8_15	Cenchrus	purpureus	ILRI_collection	ILRI	2019
279	NS 8-16	Cenchrus	purpureus	ILRI_collection	ILRI	2019
280	NS 8-17	Cenchrus	purpureus	ILRI_collection	ILRI	2019
281	NS 8-19	Cenchrus	purpureus	ILRI_collection	ILRI	2019
282	NS 8-20	Cenchrus	purpureus	ILRI_collection	ILRI	2019
283	NS 9-1	Cenchrus	purpureus	ILRI_collection	ILRI	2019
284	NS 9-2	Cenchrus	purpureus	ILRI_collection	ILRI	2019
285	NS 9-7	Cenchrus	purpureus	ILRI_collection	ILRI	2019
286	NS 9-8	Cenchrus	purpureus	ILRI_collection	ILRI	2019
287	NS 9-9	Cenchrus	purpureus	ILRI_collection	ILRI	2019
288	NS9-10	Cenchrus	purpureus	ILRI_collection	ILRI	2019
289	NS 9-12	Cenchrus	purpureus	ILRI_collection	ILRI	2019
290	NS 9-13	Cenchrus	purpureus	ILRI_collection	ILRI	2019
291	NS 9-15	Cenchrus	purpureus	ILRI_collection	ILRI	2019
292	NS 9-18	Cenchrus	purpureus	ILRI_collection	ILRI	2019
293	NS 10-1	Cenchrus	purpureus	ILRI_collection	ILRI	2019
294	NS 10-2	Cenchrus	purpureus	ILRI_collection	ILRI	2019
295	NS 10-3	Cenchrus	purpureus	ILRI_collection	ILRI	2019
296	NS 10-4	Cenchrus	purpureus	ILRI_collection	ILRI	2019
297	NS 10-5	Cenchrus	purpureus	ILRI_collection	ILRI	2019
298	NS 10-6	Cenchrus	purpureus	ILRI_collection	ILRI	2019
299	NS 10-7	Cenchrus	purpureus	ILRI_collection	ILRI	2019
300	NS 10_8	Cenchrus	purpureus	ILRI_collection	ILRI	2019
301	NS 10-9	Cenchrus	purpureus	ILRI_collection	ILRI	2019
302	NS 10-10	Cenchrus	purpureus	ILRI_collection	ILRI	2019
303	NS 10-11	Cenchrus	purpureus	ILRI_collection	ILRI	2019

304	NS 10_12	Cenchrus	purpureus	ILRI_collection	ILRI	2019
305	NS 10-13	Cenchrus	purpureus	ILRI_collection	ILRI	2019
306	NS 11- 1	Cenchrus	purpureus	ILRI_collection	ILRI	2019
307	NS 11-2	Cenchrus	purpureus	ILRI_collection	ILRI	2019
308	NS 11-3	Cenchrus	purpureus	ILRI_collection	ILRI	2019
309	NS 11_4	Cenchrus	purpureus	ILRI_collection	ILRI	2019
310	NS 11-5	Cenchrus	purpureus	ILRI_collection	ILRI	2019
311	NS 11_6	Cenchrus	purpureus	ILRI_collection	ILRI	2019
312	NS 11-8	Cenchrus	purpureus	ILRI_collection	ILRI	2019
313	NS 11-9	Cenchrus	purpureus	ILRI_collection	ILRI	2019
314	NS 11_10	Cenchrus	purpureus	ILRI_collection	ILRI	2019
315	NS 11-11	Cenchrus	purpureus	ILRI_collection	ILRI	2019
316	NS 11_12	Cenchrus	purpureus	ILRI_collection	ILRI	2019
317	NS 11-13	Cenchrus	purpureus	ILRI_collection	ILRI	2019
318	NS 11_14	Cenchrus	purpureus	ILRI_collection	ILRI	2019
319	NS 11-15	Cenchrus	purpureus	ILRI_collection	ILRI	2019
320	NS 11-16	Cenchrus	purpureus	ILRI_collection	ILRI	2019
321	NS 11_17	Cenchrus	purpureus	ILRI_collection	ILRI	2019
322	NS 11-18	Cenchrus	purpureus	ILRI_collection	ILRI	2019
323	NS 11-19	Cenchrus	purpureus	ILRI_collection	ILRI	2019
324	NS 11-20	Cenchrus	purpureus	ILRI_collection	ILRI	2019
325	NS 12-1	Cenchrus	purpureus	ILRI_collection	ILRI	2019
326	NS 12-2	Cenchrus	purpureus	ILRI_collection	ILRI	2019
327	NS 12-3	Cenchrus	purpureus	ILRI_collection	ILRI	2019
328	NS 12-4	Cenchrus	purpureus	ILRI_collection	ILRI	2019
329	NS 12-5	Cenchrus	purpureus	ILRI_collection	ILRI	2019
330	NS 12-6	Cenchrus	purpureus	ILRI_collection	ILRI	2019
331	NS 12-7	Cenchrus	purpureus	ILRI_collection	ILRI	2019
332	NS 12-8	Cenchrus	purpureus	ILRI_collection	ILRI	2019
333	NS 12-10	Cenchrus	purpureus	ILRI_collection	ILRI	2019
334	NS 12-11	Cenchrus	purpureus	ILRI_collection	ILRI	2019

335	NS 12-12	Cenchrus	purpureus	ILRI_collection	ILRI	2019
336	NS 12-13	Cenchrus	purpureus	ILRI_collection	ILRI	2019
337	NS 12-14	Cenchrus	purpureus	ILRI_collection	ILRI	2019
338	NS12-15	Cenchrus	purpureus	ILRI_collection	ILRI	2019
339	NS 12-17	Cenchrus	purpureus	ILRI_collection	ILRI	2019
340	NS 12-18	Cenchrus	purpureus	ILRI_collection	ILRI	2019
341	NS 12-19	Cenchrus	purpureus	ILRI_collection	ILRI	2019
342	NS12-20	Cenchrus	purpureus	ILRI_collection	ILRI	2019
343	NS 13-1	Cenchrus	purpureus	ILRI_collection	ILRI	2019
344	NS 13-2	Cenchrus	purpureus	ILRI_collection	ILRI	2019
345	NS13-3	Cenchrus	purpureus	ILRI_collection	ILRI	2019
346	NS 13-4	Cenchrus	purpureus	ILRI_collection	ILRI	2019
347	NS 13-5	Cenchrus	purpureus	ILRI_collection	ILRI	2019
348	NS 13_6	Cenchrus	purpureus	ILRI_collection	ILRI	2019
349	NS 13-7	Cenchrus	purpureus	ILRI_collection	ILRI	2019
350	NS 13-8	Cenchrus	purpureus	ILRI_collection	ILRI	2019
351	NS 13-9	Cenchrus	purpureus	ILRI_collection	ILRI	2019
352	NS 13-10	Cenchrus	purpureus	ILRI_collection	ILRI	2019
353	NS 13-11	Cenchrus	purpureus	ILRI_collection	ILRI	2019
354	NS 13-12	Cenchrus	purpureus	ILRI_collection	ILRI	2019
355	NS 13-13	Cenchrus	purpureus	ILRI_collection	ILRI	2019
356	NS 13-15	Cenchrus	purpureus	ILRI_collection	ILRI	2019
357	NS 13-16	Cenchrus	purpureus	ILRI_collection	ILRI	2019
358	NS 13-17	Cenchrus	purpureus	ILRI_collection	ILRI	2019
359	NS 13-18	Cenchrus	purpureus	ILRI_collection	ILRI	2019
360	NS 13-19	Cenchrus	purpureus	ILRI_collection	ILRI	2019
361	Napier_Addis	Cenchrus	purpureus	NA	ILRI	NA
362	Maralfalfa_1	Cenchrus	purpureus	NA	ILRI	NA
363	mott_new2	Cenchrus	purpureus	NA	ILRI	NA

Appendix 2: (Supplementary table 2, S2) List of individual plants under the different clusters and sub-clusters according to the population structure analysis in STRUCTURE software

Cluster I						**Cluster II**			
sub-cluster I	**sub-cluster II**	**sub-cluster III**	**sub-cluster IV**	**sub-cluster V**	**sub-cluster VI**	**sub-cluster VII**	**sub-cluster VIII**	**sub-cluster IX**	**sub-cluster X**
ILRI_NS_6-10	ILRI_16808	ILRI_NS_6-11	BAGCE_17	BAGCE_100	ILRI_NS_7-20	BAGCE_63	CNPGL_91-06-2	ILRI_14355	ILRI_14983
ILRI_NS_11-14	ILRI_16809	ILRI_NS_6-13	BAGCE_24	BAGCE_16	ILRI_NS_7-18	CNPGL_92-38-2	CNPGL_91-11-2	ILRI_15357	ILRI_16839
ILRI_NS_11-17	ILRI_16821	ILRI_NS_6-4	BAGCE_80	BAGCE_30	ILRI_NS_7-15	CNPGL_92-66-3	CNPGL_91-25-1	ILRI_16799	ILRI_NS_10-1
ILRI_NS_11-16	ILRI_16822	ILRI_NS_6-6	ILRI_16782	BAGCE_34	ILRI_NS_7-13	CNPGL_9279-2	CNPGL_92-190-01	ILRI_16837	ILRI_NS_10-13
ILRI_NS_11-8	ILRI_16818	ILRI_NS_6-16	ILRI_16794	BAGCE_53	ILRI_NS_7_6	CNPGL_93-32-2	CNPGL_93-04-2	ILRI_16840	ILRI_Ns_1-1
ILRI_NS_11-5	ILRI_16810	ILRI_NS_6-1	ILRI_16796	BAGCE_56	ILRI_NS_6-9	ILRI_14389	CNPGL_93-06-1	Maralfalfa-1	ILRI_NS_1-10
ILRI_NS_6-8		ILRI_NS_6-7	ILRI_16797	BAGCE_75	ILRI_NS_6-3	ILRI_14982	CNPGL_93-37-5	Tift_N147	ILRI_NS_11-11
ILRI_NS_6-5			ILRI_16805	BAGCE_81	ILRI_NS_4-2	ILRI_14984	CNPGL_96-21-1		ILRI_NS_11-2
ILRI_NS_6-14			ILRI_16807	BAGCE_86	ILRI_NS_4-11	ILRI_16785	ILRI_16784		ILRI_NS_1-12
ILRI_NS_11-18			ILRI_16816	BAGCE_90	ILRI_NS_4-10	ILRI_16786	ILRI_16811		ILRI_NS_11-3
ILRI_NS_6-15			ILRI_16834	BAGCE_94	ILRI_NS_4_15	ILRI_16787	ILRI_NS_12-10		ILRI_NS_1-13
ILRI_NS_11-4			ILRI_16838	CNPGL_92-198-7	ILRI_NS_13-8	ILRI_16789	ILRI_NS_12-19		ILRI_NS_1-14
ILRI_NS_11-12			ILRI_18448	CNPGL_96-23-1	ILRI_NS_13-7	ILRI_16792	ILRI_NS_8-14		ILRI_NS_1-15
ILRI_NS_7-11			ILRI_18662	ILRI_16812	ILRI_NS_13-5	ILRI_16793	India_119_22239		ILRI_NS_1-16
ILRI_NS_7-7			India_116_22242	ILRI_16813	ILRI_NS_13-3	ILRI_16795	India_129_22228		ILRI_NS_11-9
ILRI_NS_7-3			India_121_22237	ILRI_16815	ILRI_NS_13-2	ILRI_16798	India_145(1) _21965		ILRI_NS_1-2
ILRI_NS_9-2			India_123_22234	PIONEIRO	ILRI_NS_13-19	ILRI_16800	mott_new2		ILRI_NS_12-15
ILRI_NS_3-2			India_124_22233	Tift_N71	ILRI_NS_13-16	ILRI_16801	Tift_N172		ILRI_NS_12-7
ILRI_NS_9-12			India_125_22232		ILRI_NS_13-15	ILRI_16803	Tift_N200		ILRI_NS_1-3
ILRI_NS_9-15			India_126_22231		ILRI_NS_13-13	ILRI_16804	Tift_N223		ILRI_NS_1-5
ILRI_NS_9-8			India_132_22225		ILRI_NS_13-11	ILRI_16806	Tift_N225		ILRI_NS_1-6
ILRI_NS_11-13			India_149_21785		ILRI_NS_12-6	ILRI_16819	Tift_N23		ILRI_NS_2-1

ILRI_NS_9-13	India_89_22236	ILRI_NS_12-5	ILRI_16836	Tift_N43	ILRI_NS_2-2
ILRI_NS_11-15		ILRI_NS_12-4	ILRI_16902	Tift_N68	ILRI_NS_2-3
ILRI_NS_6-19		ILRI_NS_12-3	ILRI_18438	Tift_N75	ILRI_NS_2-6
ILRI_NS_9-9		ILRI_NS_12-20	India_144_21964	Tift_N8	ILRI_NS_2-7
ILRI_NS_9-18		ILRI_NS_12-18	Napier_Addis		ILRI_NS_2-8
ILRI_NS_11-20		ILRI_NS_12-17	Tift_N109		ILRI_NS_2-9
ILRI_NS_4-4		ILRI_NS_12-13	Tift_N131		ILRI_NS_3-1
ILRI_NS_9-1		ILRI_NS_12-11			ILRI_NS_3-12
ILRI_NS_7-10		ILRI_NS_12-1			ILRI_NS_3-13
ILRI_NS_13-12		ILRI_NS_10-9			ILRI_NS_3-16
ILRI_NS_3-3		ILRI_NS_10-5			ILRI_NS_3-19
ILRI_NS_7-14		ILRI_NS_10-4			ILRI_NS_3-20
ILRI_NS_3-8		ILRI_NS_10-2			ILRI_NS_3-5
ILRI_NS_13-1		ILRI_NS_10-12			ILRI_NS_3-9
ILRI_NS_3-18		ILRI_NS_10_8			ILRI_NS_4-1
ILRI_NS_6-12					ILRI_NS_4-13
ILRI_NS_7-4					ILRI_NS_4-14
ILRI_NS_3-10					ILRI_NS_4-6
ILRI_NS_13-9					ILRI_NS_5-1
ILRI_NS_11-19					ILRI_NS_5-10
ILRI_NS_11_6					ILRI_NS_5-12
ILRI_NS_7-1					ILRI_NS_5-13
ILRI_NS_3-1					ILRI_NS_5-14
ILRI_NS_3-6					ILRI_NS_5-15
ILRI_NS_7-5					ILRI_NS_5-17
					ILRI_NS_5-18
					ILRI_NS_5-2

ILRI_NS_5-3

ILRI_NS_5-4

ILRI_NS_5-6

ILRI_NS_5-9

ILRI_NS_6_2

ILRI_NS_8-10

ILRI_NS_8-11

ILRI_NS_8-15

ILRI_NS_8-17

ILRI_NS_8-19

ILRI_NS_8-2

ILRI_NS_8-20

ILRI_NS_8-3

ILRI_NS_8-7

ILRI_NS_8-8

ILRI_NS_8-9

India_128_22229

Appendix 3: (Supplementary table 3, S3): Membership probability of individual genotypes under different clusters and sub clusters

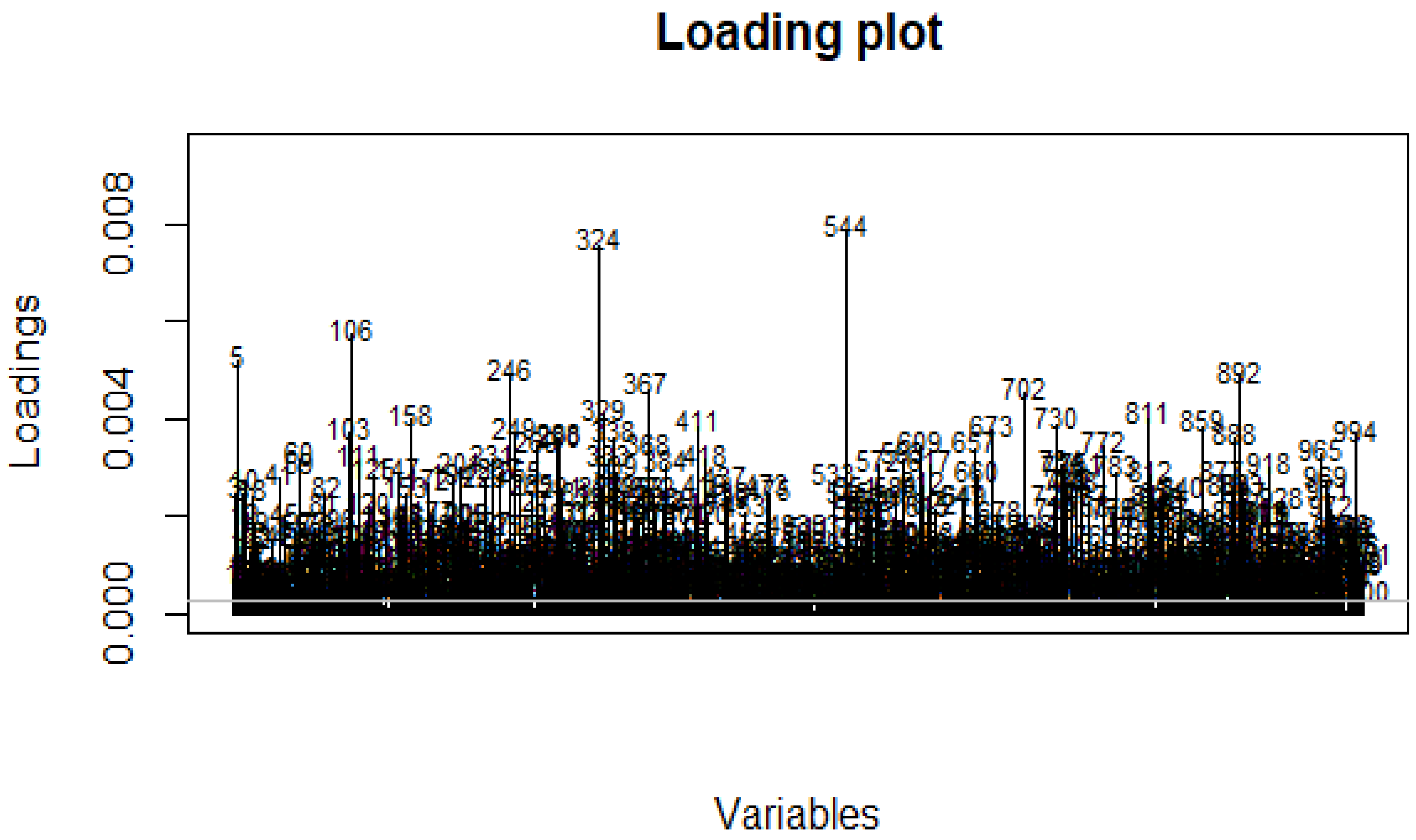
	Membership probability									
Genotype	**sub-cluster I**	**sub-cluster II**	**sub-cluster III**	**sub-cluster IV**	**sub-cluster V**	**sub-cluster VI**	**sub-cluster VII**	**sub-cluster VIII**	**sub-cluster IX**	**sub-cluster X**
BAGCE_1	0.00	0.00	0.00	0.00	0.29	0.00	0.17	0.43	0.09	0.00
BAGCE_7	0.00	0.00	0.00	0.35	0.47	0.00	0.17	0.01	0.00	0.00
BAGCE_97	0.00	0.00	0.00	0.40	0.41	0.00	0.13	0.00	0.06	0.00
CNPGL_00-1-1	0.01	0.01	0.00	0.01	0.48	0.00	0.17	0.32	0.00	0.00
CNPGL_92-133-3	0.00	0.00	0.00	0.00	0.03	0.00	0.47	0.24	0.26	0.00
CNPGL_92-56-2	0.00	0.00	0.00	0.06	0.21	0.00	0.28	0.24	0.21	0.00
CNPGL_93-01-1	0.00	0.00	0.00	0.11	0.06	0.00	0.45	0.00	0.38	0.00
CNPGL_93-08-1	0.00	0.00	0.00	0.00	0.22	0.00	0.38	0.20	0.20	0.00
CNPGL_93-18-2	0.00	0.00	0.00	0.00	0.16	0.00	0.39	0.29	0.16	0.00
CNPGL_94-07-2	0.00	0.00	0.00	0.00	0.12	0.00	0.44	0.38	0.05	0.00
CNPGL_94-13-1	0.00	0.00	0.00	0.00	0.12	0.00	0.45	0.11	0.32	0.00
CNPGL_96-27-3	0.00	0.00	0.00	0.00	0.44	0.00	0.34	0.00	0.22	0.00
ILRI_1026	0.01	0.20	0.00	0.44	0.00	0.01	0.16	0.02	0.15	0.00
ILRI_15743(MOTT)	0.11	0.01	0.00	0.27	0.00	0.17	0.44	0.00	0.00	0.00
ILRI_16783	0.12	0.01	0.00	0.26	0.00	0.15	0.45	0.00	0.00	0.00
ILRI_16788	0.13	0.00	0.00	0.26	0.00	0.16	0.44	0.00	0.00	0.00
ILRI_16790	0.28	0.01	0.00	0.48	0.00	0.23	0.00	0.00	0.00	0.00
ILRI_16791	0.10	0.01	0.00	0.26	0.01	0.19	0.43	0.00	0.00	0.00
ILRI_16802	0.11	0.01	0.00	0.26	0.01	0.21	0.41	0.00	0.00	0.00
ILRI_16814	0.27	0.01	0.44	0.22	0.00	0.00	0.00	0.06	0.00	0.00
ILRI_16817	0.28	0.00	0.46	0.23	0.00	0.00	0.00	0.04	0.00	0.00

ILRI_16835	0.43	0.06	0.29	0.21	0.00	0.00	0.00	0.00	0.00	0.00
ILRI_NS_10-10	0.18	0.00	0.00	0.00	0.00	0.49	0.00	0.00	0.00	0.33
ILRI_NS_10-11	0.28	0.00	0.01	0.00	0.00	0.41	0.00	0.08	0.00	0.21
ILRI_NS_10-3	0.26	0.00	0.03	0.01	0.00	0.29	0.00	0.00	0.00	0.42
ILRI_NS_10-7	0.40	0.00	0.00	0.00	0.00	0.46	0.00	0.00	0.00	0.13
ILRI_NS_11-1	0.15	0.05	0.00	0.01	0.00	0.37	0.00	0.00	0.00	0.42
ILRI_NS_11-10	0.12	0.03	0.00	0.01	0.00	0.34	0.09	0.10	0.00	0.31
ILRI_NS_1-18	0.45	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.48
ILRI_NS_12-12	0.43	0.00	0.00	0.00	0.00	0.47	0.00	0.00	0.00	0.10
ILRI_NS_12-14	0.38	0.00	0.02	0.00	0.00	0.47	0.00	0.00	0.00	0.13
ILRI_NS_12-2	0.13	0.00	0.00	0.00	0.00	0.43	0.00	0.00	0.00	0.44
ILRI_NS_12-8	0.45	0.00	0.01	0.00	0.00	0.49	0.00	0.00	0.00	0.05
ILRI_NS_13-17	0.04	0.01	0.01	0.00	0.00	0.47	0.00	0.00	0.00	0.47
ILRI_NS_1-4	0.43	0.01	0.00	0.01	0.00	0.10	0.00	0.00	0.00	0.45
ILRI_NS_1-9	0.00	0.00	0.00	0.00	0.06	0.00	0.16	0.34	0.00	0.44
ILRI_NS_3-11	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.37	0.00	0.49
ILRI_NS_3-14	0.09	0.02	0.02	0.00	0.00	0.47	0.01	0.07	0.01	0.31
ILRI_NS_3-15	0.39	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.07	0.47
ILRI_NS_3-17	0.10	0.07	0.00	0.03	0.00	0.35	0.00	0.00	0.00	0.45
ILRI_NS_3-4	0.35	0.00	0.02	0.00	0.00	0.15	0.00	0.00	0.00	0.48
ILRI_NS_3-7	0.29	0.00	0.00	0.00	0.00	0.24	0.00	0.00	0.00	0.46
ILRI_NS_4-12	0.18	0.00	0.00	0.00	0.00	0.49	0.00	0.00	0.00	0.33
ILRI_NS_4-19	0.37	0.00	0.00	0.00	0.00	0.38	0.00	0.00	0.00	0.24
ILRI_NS_4-20	0.22	0.00	0.00	0.17	0.01	0.00	0.06	0.19	0.01	0.34
ILRI_NS_4-5	0.22	0.00	0.00	0.00	0.00	0.48	0.00	0.00	0.00	0.29
ILRI_NS_4-7	0.21	0.01	0.00	0.04	0.00	0.36	0.00	0.00	0.00	0.38
ILRI_NS_4-9	0.39	0.00	0.00	0.00	0.00	0.23	0.00	0.00	0.00	0.38

ILRI_NS_5-16	0.00	0.00	0.10	0.44	0.00	0.00	0.00	0.00	0.00	0.46
ILRI_NS_5-19	0.00	0.00	0.00	0.00	0.03	0.00	0.09	0.43	0.01	0.45
ILRI_NS_5-5	0.00	0.00	0.00	0.00	0.00	0.00	0.19	0.38	0.01	0.43
ILRI_NS_5-8	0.00	0.00	0.06	0.00	0.07	0.00	0.02	0.38	0.00	0.46
ILRI_NS_6-18	0.06	0.00	0.44	0.09	0.03	0.06	0.01	0.21	0.10	0.00
ILRI_NS_6-20	0.00	0.00	0.47	0.00	0.00	0.00	0.14	0.29	0.05	0.04
ILRI_NS_7-12	0.31	0.00	0.00	0.00	0.00	0.43	0.00	0.00	0.00	0.25
ILRI_NS_7-17	0.32	0.01	0.00	0.03	0.01	0.39	0.00	0.00	0.00	0.24
ILRI_NS_7-2	0.49	0.00	0.00	0.04	0.01	0.46	0.00	0.00	0.00	0.00
ILRI_NS_7-9	0.42	0.00	0.00	0.00	0.02	0.30	0.00	0.03	0.02	0.20
ILRI_NS_8-16	0.00	0.00	0.00	0.00	0.01	0.00	0.14	0.43	0.00	0.41
ILRI_NS_9-10	0.48	0.01	0.00	0.19	0.00	0.31	0.00	0.00	0.00	0.00
ILRI_NS_9-7	0.29	0.03	0.08	0.14	0.00	0.04	0.00	0.00	0.00	0.42
India_115_22243	0.10	0.18	0.00	0.48	0.08	0.00	0.03	0.00	0.13	0.00
India_118_22241	0.14	0.12	0.02	0.36	0.01	0.01	0.12	0.19	0.04	0.00
India_120_22238	0.01	0.04	0.10	0.27	0.00	0.01	0.09	0.44	0.04	0.01
India_127_22230	0.00	0.13	0.00	0.43	0.04	0.00	0.17	0.01	0.21	0.01
India_131_22226	0.03	0.09	0.00	0.47	0.00	0.00	0.17	0.00	0.22	0.00
India_141_21967	0.00	0.00	0.00	0.06	0.01	0.00	0.37	0.23	0.33	0.00
India_142_21966	0.01	0.07	0.00	0.35	0.05	0.00	0.22	0.01	0.29	0.01
India_146_21788	0.02	0.09	0.00	0.06	0.42	0.00	0.26	0.02	0.14	0.00
India_147_21787	0.00	0.06	0.00	0.07	0.02	0.00	0.49	0.01	0.35	0.00
India_150_21784	0.01	0.00	0.01	0.16	0.02	0.00	0.38	0.08	0.33	0.02
India_151_21783	0.08	0.00	0.00	0.45	0.00	0.00	0.14	0.00	0.32	0.00
Tift_N128	0.19	0.06	0.36	0.03	0.00	0.04	0.00	0.32	0.00	0.00
Tift_N130	0.21	0.00	0.00	0.37	0.05	0.01	0.00	0.35	0.00	0.00
Tift_N137	0.00	0.00	0.02	0.35	0.08	0.00	0.21	0.02	0.29	0.02

Tift_N138	0.01	0.00	0.00	0.30	0.04	0.00	0.26	0.03	0.36	0.00
Tift_N210	0.17	0.00	0.00	0.47	0.03	0.01	0.32	0.00	0.00	0.00
Tift_N37	0.00	0.00	0.00	0.40	0.43	0.00	0.15	0.00	0.02	0.00
ILRI_NS_10-6	0.28	0.00	0.00	0.00	0.00	0.39	0.00	0.00	0.00	0.32
ILRI_NS_4-3	0.25	0.00	0.01	0.00	0.00	0.37	0.00	0.00	0.00	0.36
ILRI_NS_4-18	0.22	0.00	0.00	0.00	0.00	0.39	0.00	0.00	0.00	0.38

Appendix 4: loading plot for 1000 SNP markers that were selected from DPAC analysis



Appendix 5: Compo plot that shows membership probabilities of individual genotypes from DPAC analysis