IDENTIFICATION OF NEW MOLECULAR MARKERS FOR DIVERSITY ANALYSIS AND BREEDING FOR EARLY MATURITY AND DETERMINATE LABLAB (*LABLAB PURPUREUS*) VARIETIES

BY

ELIEZAH MAINA KAMAU

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DOCTOR OF PHILOSOPHY IN PLANT BREEDING AND BIOTECHNOLOGY, SCHOOL OF AGRICULTURE AND BIOTECHNOLOGY, UNIVERSITY OF ELDORET, KENYA.

FEBRUARY, 2022

DECLARATION

Declaration by the candidate

This thesis is my original work and has not been presented for a degree in any other University. No part of this thesis may be reproduced without the prior permission of the author and/ University of Eldoret.

Kamau, Eliezah Maina

_____ _____

AGR/D.PHIL/007/09 **Date**

Declaration by supervisors

This thesis has been submitted for examination with our approval as University supervisors.

_____ _____

Prof. Miriam G. Kinyua **Date**

University of Eldoret, Eldoret, Kenya

_____ _____

Dr. Charles N. Waturu **Date**

Horticulture Research Institute, KALRO,

Thika, Kenya

DEDICATION

This thesis is dedicated to my dear wife, Julie and our lovely children Ivy and Samuel for their endless love, support and encouragement.

ABSTRACT

Lablab *(Lablab purpureus* (L.) Sweet) is a grain legume crop commonly grown in Africa and India and is used as human food, animal feed, in soil conservation, enhancing soil fertility and in weed management. In Kenya, most farmers grow landraces which are inherently low yielding and have other undesirable attributes like long maturity duration and indeterminate growth habit. The status of genetic diversity of the local lablab germplasm and how it relates to materials from other regions remain unclear. Unavailability of novel breeding selection tools such as molecular markers and lack of adequate information on the inheritance pattern of important traits have also hampered the crop improvement in Kenya. The objectives of the study were therefore: to develop new molecular markers for lablab; to assess the genotypic diversity of local, exotic and wild lablab accessions using simple sequence repeats, diversity array technology (DArT) and single nucleotide polymorphism (SNP) molecular markers; to identify the inheritance pattern of selected lablab qualitative traits and establish the linkage relationship of the genes controlling them; determine the heritability estimates, genetic gain and character association of important traits of determinate lablab. Transcriptome sequencing using 454 Titanium FLX system of mRNA isolated from leaves and shoots of lablab samples, was conducted to discover genic-SSRs and to develop SSR markers. Eight of these new developed SSR markers were used to characterize 189 lablab accessions. SilicoDArT and SNP markers were developed using DArTSeq technology and used to characterize 240 lablab accessions. The genetics of growth habit and other qualitative traits were studied in three generations (F_1, F_2 F_3) of eight lablab populations. Selected F_5 lines with determinate growth habit were grown using RCBD design at KALRO Thika and Katumani to determine heritability estimates, genetic gain and character association. Results indicated that there were 446 genic SSRs from 3140 assembled lablab contigs indicating an overall density of 202 SSR per Mbp. SSR primer pairs designed from the contigs sequences amplified on lablab genome. The gene diversity among the 189 accessions based on SSR loci ranged from 0.26 to 0.52 with an average of 0.38, with germplasm collected from Kenya showing a moderate genetic diversity of 0.36. Higher genetic diversity ($He<0.5$) was detected within the Ethiopian and South Africa populations. A total of 15,601 polymorphic DArT markers and 11,431 SNP markers were identified each with average reproducibility and genotype call rate of more than 90%. Based on both DArT and SNP markers the 240 lablab was of narrow genetic diversity with the expected mean heterozygosity of 0.030 (DArT) and 0.039 (SNP). However, genetic differentiation was most pronounced between the cultivated and the wild accessions. The growth habit in lablab is under control of three genes which could be temperature dependent. The genes controlling stem growth habit and time to flowering in lablab are linked. The study identified, moderate to high heritability, genetic advance estimates and significant positive correlations of pods per plant, raceme per plant, plant height, pod width, racemes per pod and number of flower nodes. The newly developed molecular markers are useful in grouping lablab genotypes into related clusters that breeders can use to enhance lablab productivity. Selection for high number of units of pods per plant, raceme per plant, plant height, pod width, racemes per pod and can be effective when targeting to develop high seed yielding determinate varieties.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

cDNA	- Complementary deoxyribonucleic acid
CTAB	- Cetyltrimethylammonium bromide
DArT	- Diversity array technology
DNA	- Deoxyribonucleic Acid
EDTA	- Ethylenediaminetetraacetic acid
GAM	-Genetic Advance as percentage of mean
GBK	- Gene bank of Kenya
GenAlex	– Genetic Analysis in Excel soft ware
GenStat	– GenStat Statistical software
KALRO	- Kenya Agricultural and Livestock Research Organization
ng	- Nanogram
NJ	- Neighbor joining
PCR	- Polymerase Chain Reaction
PIC	- Polymorphic information content
RFLP	- Restriction Fragment Length Polymorphism
SSRs	- Simple sequence repeats
SNP	- Single Nucleotide Polymorphism
TBE	- Tris Boric acid-EDTA
TE	- Tris EDTA
Tm^{o}C	- Anealing temperature
UNJ	- Unweighted Neighbour Joining

ACKNOWLEDGEMENTS

First to God the giver of all wisdom and knowledge.

It's my great pleasure to acknowledge the tireless concerted effort from my supervisor Professor Miriam Kinyua who during the execution of the project guided me regularly. She offered many professional insights which were always most encouraging and profitable to my training. I am indebted to my other supervisor, Dr. Charles Waturu who showed great interest in my practical and professional training and spent time to critique the entire write up. My gratitude is due to my employer, the Kenya Agricultural and Livestock Research Organization (KALRO) for granting study leave to undertake this study.

I am immensely grateful to Dr Catherine Taracha, Centre Director, KARLO Biotechnology for allowing me use the laboratory facilities for the molecular characterization experiment.

I wish to express my heartfelt gratitude to Kirkhouse Trust for providing financial and technical support in course of my PhD study. Kirkhouse Trust supported in payment of school fees, provided all reagents for molecular work and supported my visit to Sean Mayes Laboratory at Sutton Bonington Campus of University of Nottingham in Britain where I developed the SSR molecular markers. I shall always be grateful for their good gesture. I am grateful to Profesor Sean Mayes and Dr. Katie Mayes of Sean Mayes laboratory for their enormous support in the molecular work. I am grateful to Dr Alice Muchugi of ICRAF for her support in the development of DArT and SNP markers.

I am also indebted to Mr. Robert Kariba of ICRAF for the technical support in DNA samples preparation for DArTseq analysis and statistical analysis of data. I am grateful for the advice given by Mr. Bramwel Wanjala and Perpetua Wangari of KALRO

biotechnology on DNA samples preparation and statistical analysis of the data. Many thanks to Mr. Daniel Karangi for assisting in the general crop management. I also acknowledge Mr. Joseph Karoga and Mrs. Dorcas Githungo of KALRO Thika for their assistance in laying out the field experiments and in data collection. I am unable to find adequate words to express my gratitude to my wife, Julie Wanjiru, my daughter, Ivy Wanjiku and my son, Samuel Kamau, who consciously permitted me to be away to pursue this training. They supported me all along, and their trust and confidence in me were a great inspiration during the long hours of work endured during this training. Much thanks to my mum, Agnes Wanjiku for her prayer and encouragement. I also acknowledge my late father-in-law, Julius Kibuika for his moral support in my acadermic progress. My unreserved thanks are also due to all those people not mentioned here, who in one way or another made contributions towards the successful execution of this work.

CHAPTER ONE

INTRODUCTION

1.1 Background information

Lablab (*Lablab purpureus* (L.) Sweet) is a leguminous crop that is currently grown in Africa, India and parts of Southeast Asia (Maass *et al*., 2005; Kimani *et al*., 2012; Sennhenn, 2015). The crop is cultivated in diverse climatic conditions worldwide and displays wide genetic diversity. The production of lablab in India is about 0.030 million tonnes, grown on approximately 0.085 million hectares (Keerthi *et al*., 2014). In Kenya, lablab is grown in more than 10,000 hectares, with an average yield of less than 0.5 tons ha^{-1}(Kamotho *et al*., 2015). In Kenya, lablab can grow in different agro-ecological zones ranging from the lowlands of the coastal region to the highlands of Mt. Kenya region (Kamotho *et al*., 2015).

Lablab is an important pulse crop in some developing countries of Asia and Africa where it is mainly consumed as dry seeds and fresh green pods (Maass *et al*., 2005; Kimani *et al*., 2012; Sennhenn, 2015; Amkul *et al*., 2020). As food, it serves as a vital source of protein (20-28%), minerals such as Zinc (34mg/kg) and iron (57 mg/kg) (Onyango, 2011). The immature pods are commonly used as vegetable while the crop residue after harvest are utilized for making manure (Julius *et al*., 2021). It is also important as livestock forage and feed especially in Australia (Maass *et al*., 2010). The spreading lablab varieties are used in soil conservation and in weed management. Lablab has also been utilized in the management of *diabetes mellitus*, inflammations, coronary heart diseases and anaemia (Al-Snafi, 2017; Naeem *et al*., 2021). The ability of lablab to extract soil water at deep depth even at heavy-textured soil makes it a crop of choice for farmers in semi- arid areas (Kimani *et al*., 2012; Kilonzi *et al.,* 2017). In addition to subsistence use, the excess grains are sold in the local market where it

fetches higher prices than most of the other common legumes (AFFA, 2016). Therefore, there is a great potential for commercialization of this crop thereby giving an incentive for breeding and production.

In Africa, lablab has remained neglected and underutilized despite the crop being well adapted to arid and semiarid areas which are characterized by food insecurity (Maass *et al*., 2010). Like other underutilized crops in Africa, there is poor research attention given towards the improvement of lablab as evidenced by little literature on the crop and poor crop productivity. Poor research attention on this leguminous crop is partly because it is still regarded as subsistence crop, thereby not attracting donor research funding like the other major and commercial crops. Owing to the minimal research effort on this crop, the grain yield of lablab on farmers' fields in Kenya is low, ranging from 0.3 tons per hectare to 0.71 tons per hectare (Kamotho *et al.,* 2010; Tegemeo, 2010; Meru County, 2018; Lamu County, 2018). Higher grain yields of 1.5 tons per hectare have been reported under irrigation in semi-arid areas of Machakos in Kenya (Sennhen, 2015). In, Brazil and Bolivia, grain yield of upto one ton per hectare has been reported. In Columbia, 1.2 – 1.8 tons per hectare were reported in the high attitude areas (Kempanna *et al*., 2008). This indicates that, there is room for improvement of the current level of lablab productivity in Kenya through investment in breeding of high yielding varieties that meet the farmer's requirements.

Currently, most of the lablab production in Kenya is carried out in the semi-arid areas (Kamotho, 2015). The production of this legume is mainly done by smallholder farmers within diverse cropping systems, ranging from monoculture to intercropping, mainly with cereals. A majority of these farmers are resource-poor and faced by numerous challenges. However, little effort has been made to understand the challenges that limit

productivity among the small-scale lablab farmers in Kenya. In their survey in major lablab producing areas in Kenya, Kinyua and Kiplagat, (2012) identified low soil fertility, pests and disease infestation, cultivars with undesirable traits like low genetic yield potential, late maturity and indeterminate growth habit as important production constraints. Little effort has been made to address these challenges largely because the crop is still considered underutilized and therefore receive minimal attention from researchers and donors.

1.2 Statement of the problem

Production and productivity of lablab in Kenya has remained low (Boit *et al.*, 2013, Kinyua and Kiplagat, 2012, Tegemeo, 2010). Lack of cultivars which are high yielding and with desirable traits like earliness to maturity and determinate bush characters have been identified to contribute to the low adoption and productivity especially in short - season environments (Kinyua and Kiplagat, 2012). Lablab is mainly grown in the dryland areas of Kenya which frequently experience short rainfall duration (Daryanto *et al*., 2015). Under rain-fed conditions, long duration cultivars are subjected to end of season moisture stress leading to yield loss (Sennhenn, 2015). Yield losses of more than 43.5% has been reported in several legumes when grown under moisture stress (Daryanto *et al*., 2015). Even in drought tolerant crops like lablab, excess moisture stress during pods and seed development can result in formation of few pods, shivered grains and ultimately low grain yield (Fredrick *et al*., 2001; Nepomuscene *et al*., 2017). Currently, the cultivars and landraces available to farmers in Kenya all have medium to long maturity duration (Kamotho, 2015). To overcome the challenge of production of long maturity cultivars under short season environment, farmers are advised to plant lablab during the short season rainfall (October-December) but harvest their crop after the long season of the following year. The short season rainfall period is utilized to

germinate and establish the crop. The short season rainfall is not enough to allow full crop cycle of the long duration cultivars. Therefore, the crop is allowed to remain in the field during the dry months of January and February awaiting for the rains at March. During this period, the crop remain vegetative though it sheds most of the leaves. The vegetative lablab crop in field usually attract heavy infestation of insect pests such as aphids since during this dry period, there are only few other alternative hosts for the pests. After the onset of the long season rainfall, new shoots immediately develop followed by flowering and pod formation. The small holder farmers are discouraged from planting lablab due to the long duration (usually 7-8 months) they have to wait before harvesting of the grains. This problem can be overcome by farmers adopting early maturing cultivars which are able to escape drought. Drought escape is an adaptive mechanism which allows plants to develop rapidly and complete their full life-cycle before the onset of drought (Shavrukov *et al*., 2017). Early maturity under the conditions of terminal drought can reduce exposure to moisture stress during the delicate flowering and grain filling periods.

Majority of the lablab landraces and varieties frequently used by Kenyan farmers are not only late maturing but also have indeterminate growth habit thereby limiting their utilization especially under intercropping system (Kamotho *et al.,* 2015). Cereal/lablab intercropping system is very common among lablab small holder farmers with limited arable land (Kamotho *et al*., 2010). Intercropping enables the farmers to produce diverse crop species within the same field and season thus minimizing risks associated with crop failure (Bonginkosi *et al*., 2018). The current practice is where the farmers intercrop the cereals with indeterminate lablab cultivars. Under the cereal/lablab intercropping system, the indeterminate and aggressive lablab varieties become very vegetative and climb on, pull down and lodge the intercrop cereal resulting in yield

reduction of both the cereal and lablab crop (Kamotho, 2015; Kgasago, 2006). For instance, a reduction of 57% in maize leaf yield when maize and lablab were sown at the same time compared to the leaf yield of the maize crop under monoculture (Bonginkosi *et al*., 2018). To reduce the problem of competition and entanglement of indeterminate varieties on maize, farmers are usually advised to plant the lablab at least two weeks later than maize (Rapholo *et al*., 2020). This practice is not favourable especially for areas with short seasons. Inaddition, indeterminate lablab varieties usually produce high biomass at the expense of the grain yield (Nord *et al*., 2020) thereby resulting in low grain yield. Currently there is lack of determinate lablab varieties developed for grain production in Kenya. Therefore, development of varieties with determinate growth habit may be necessary to increase lablab productivity and especially under the cereals intercrop system.

The current problem of low lablab productivity in Kenya can be resolved through crop improvement. A thorough understanding of the level of genetic diversity of the available germplasm is important in selecting of the suitable parents to be used in the breeding programme. Although, there is a considerable amount of lablab genetic resources held by gene bank of Kenya (GBK) and other gene banks such as international livestock research institute (ILRI), the level of relatedness of these materials is not well understood (Shivashi *et al*., 2012, Kimani *et al*., 2012, Kamotho, 2015) limiting their utilization in breeding programs. Rapid characterization and subsequent utilization of lablab germplasm in the improvement of the farmer's cultivars can be achieved through the use of efficient biotechnology tools such as molecular markers. Research effort have been made to develop molecular markers in lablab (Guwen *et al*., 2013) but the numbers are still too low compared to other leguminous crops. Lack of adequate DNA markers in lablab is of particular concern because these

molecular tools have been efficient in characterization and genetic improvements of many other crop plants (Yunbi and Jonathan 2008, Ribaut *et al*., 2010, Guo-Liang Jiang, 2013; Raatz *et al*., 2019; Amkul *et al*., 2020).

This study therefore sought to address the problem of inadequate molecular tools for germplasm characterization, limited understanding of genetic relatedness of the available germplasm and inadequate understanding of characters association in lablab which is a prerequisite for the development and selection of early maturing determinate lablab varieties in Kenya.

1.3 Justification of the study

Lablab productivity in Kenya has remained low due to lack of adapted varieties especially to the short -season environments (Kinyua and Kiplagat, 2012). Crop improvement is one of the sustainable strategies that can be used to increase productivity of underutilized crops such as lablab. Success of developing improved crop varieties is dependent on the access to full range of germplasm resources. A good characterization of these genetic resources in terms of their relatedness is a prerequisite for their use in crop improvement (Kamenya *et al*., 2021). The characterization of lablab germplasm is important because it would allow simple grouping of accessions, development of core collections and in identification of valuable germplasm for breeding better adapted varieties. The inclusion of big worldwide lablab collections in this study is important because it enables better understanding of the relatedness of the local and exotic genetic materials. This information would enable the local breeding program to benefit from the lablab genetic resources from other countries and regions. Such materials can be introduced and hybridized with the local cultivars to broaden their genetic base.

Conventionally, diversity of crops is assessed through phenotypic qualitative traits such as growth habit or quantitative traits such as yield potential. However, this approach has a major setback because the expression of quantitative traits is subjected to strong environmental influence (Wamalwa *et al*., 2016) and is dependent on the developmental stage of the plant thereby reducing their efficiency in germplasm classification. Modern techniques which analyse diversity at molecular level have been developed and successfully applied in evolutionary and diversity studies of different crops (Nadeem *et al*., 2018). However, like in most other underutilized crops, there are limited molecular markers developed for lablab crop (Guwen *et al*., 2013). Simple sequence repeats (SSR) markers, which are single locus markers with multiple alleles, highly polymorphic, co-dominance of alleles, arbitrary dispensation in genome (Luo *et al*., 2018) can be effective for discriminating between lablab genotypes. Currently, Diversity array technology (DArT) molecular markers are becoming popular in characterizing large number of individuals because they are amenable to high-throughput genotyping and are cost-effective (Liu *et al*., 2018). No prior knowledge of DNA sequence is required for detection of polymorphisms with DArT markers therefore making them ideal for species with limited genome information such as lablab. Similar to SSR markers, single nucleotide polymorphism (SNP) markers are highly abundant, polymorphic and are co-dominant. Their abundance in genome and their low cost of development makes them popular in population structure and genetic diversity studies (Semagn *et al*., 2018). Therefore, the development of numerous molecular markers for lablab in this study can offer important tools to effectively investigate the genetic diversity in the huge local and introduced lablab germplasm, understand the structure of different populations and in classification of relatedness of these materials (Nadeem *et al*., 2018).

Most of the cultivars grown by farmers in Kenya have long maturity period of more than 120 days (Kamotho, 2015) and extends way beyond the rainy season which is usually less than 80 days (Sanhenn, 2015) thereby suffering from terminal drought. Breeding for early crop maturity is therefore essential to match crop duration with the period of favourable growing conditions, avoid losses caused by terminal drought and to stabilize the grain yield. Availability of determinate bush lablab varieties in Kenya can also stabilize yield especially under the intercrop system. Towards the development of improved lablab varieties, combination of multiple qualities such as high yield along with determinate growth habit and earliness to maturity is therefore essential. Clear understanding of the genetics of these traits is essential in determination of the suitable strategies for combining these traits into improved cultivars (Acquaah, 2015). In addition, the information on heritability, genetic advance and correlation of yield related traits is necessary to identify the traits to use in selection of individuals with the desired high grain yield, determinate growth habit and early maturing traits. Although this genetic information has been generated in lablab, there have been contrasting conclusions from diverse workers and most of the work has been undertaken in Asia with limited studies done in Africa. Unless reasonable information on inheritance pattern, genetic advance, correlation and heritability of important lablab traits is established under the local genetic background, the process of selection of improved lablab varieties for Kenyan farmers will be slow and less efficient.

1.4 Objectives

1.4.1 Broad objectives.

The study aimed to contribute to enhanced lablab productivity by improving the maturity duration and growth habit of Kenyan lablab varieties, hence contributing to expanded adoption, increased yields and income of small holder farmers.

1.4.2 Specific objectives.

1) To identify new lablab derived simple sequence repeat (SSR) molecular markers;
2) To identify the genetic diversity of some local and introduced lablab accessions using the new SSRs markers;
3) To assess the genetic diversity and population structure of some worldwide lablab accessions based on two ultra-high-throughput diversity array technology (DArT) markers (Silico DArT and SNP);
4) To determine the genetics of growth habit traits, days to flower initiation, and flower colour traits in lablab;
5) To assess genetic variation, heritability and genetic advance of important yield attributing traits of some advanced determinate lablab lines
6) To determine association among yield and yield related traits of determinate lablab.

1.5 The Hypotheses

1) SSRs are abundant in lablab genome and the Roche/454 next generation sequencing system produces long reads which have enough flanking regions to design primers.
2) The new SSRs are robust in revealing polymorphism between closely related lablab accessions.
3) DArTSeq platform can be applied to provide genetic differences of lablab germplasm.
4) Stem growth habit, flower colour and duration to flower initiation is controlled by major genes in lablab.

5) There are some yield contributing traits with high heritability and genetic advance which if identified could be useful in selection of high yielding determinate varieties in Kenya
6) There are some yield contributing characters that are highly correlated to grain yield and if identified could be useful in development of early maturing determinate lablab varieties in Kenya.

CHAPTER TWO

LITERATURE REVIEW

2.1 Origin and distribution of Lablab

Lablab purpureus (L) Sweet belongs to the family Fabaceae, sub-family *faboideae,* tribe *phaseolae*, and subtribe *phaseolineae* and genus lablab. Verdicort, (1970) classified the species L. *purpureus* into three subspecies namely*;* subspecies *unicinatus,* subspecies *bengalensis* and subspecies *purpureus*. These subspecies mainly differ in the shape and sizes of their pods. For instance, the subspecies *uncinatus* have small, scimitar-shaped pods of about 40 mm x15 mm and contains all the wild accessions. The subspecies *bengalensis*, have the longest (about 140mm) pod sizes among the others and the shape is linear oblong. The subspecies *purpureus* has larger, scimitar-shaped pods of 100 mm x 40 mm and consist of all accessions cultivated as pulse (Maass, 2005). Maass *et al.,* (2016), observed that although Verdcourt (1970), combined all wild morphotypes under subspecies *unicinatus,* there was clear difference in pod morphology of the two seeded and four seeded wild types. The subspecies *purpureus* and subspecies *bengalensis* are genetically very similar. Most of the domesticated accessions in majority of lablab growing regions belongs either to subspecies *purpureus* or subspecies *bengalensis* with subspecies *unicinatus* domesticated only in Ethiopia.

There are several theories suggesting different origin of lablab as reviewed by Kempanna *et al*., 2008. Verdcourt (1970) suggested the origin to be in east Africa because the wild ancestral forms of *L. purpureus;* ssp. *unicinatus* are found in East Africa. Pulseglove (1974) suggested the origin of lablab to be Asia because that is the place where it's widely cultivated. Fuller (2003) mentioned that lablab originated from India because its archeological findings in India at 3500 BC. Through AFLP studies Mass *et al*., (2005) showed that lablab originated from eastern or south Africa. Maas *et*

al., (2016) disputed the Indian origin and suggested eastern and southern Africa as the centre of origin because these are the only regions with occurrence of the wild plants. Like the other crops of Africa, lablab could have been taken to India by the ancient people of Arabia.

2.2 Morphology of Lablab

After germination, the plant forms a taproot, from which lateral and many adventitious roots develop. Lablab has primary leaves that are unifoliate and the subsequent leaves are trifoliate. The leaflets are ovate while the lateral leaflets are oblique and sometimes covered with soft hair. The inflorescence is a stiff axillary raceme with many flowers. Flowers are borne in the axillary and terminal racemes which may be one or many flowered. The colour of the flower may be white, pink, red or purple. The flowers contain ten stamens and a single multi-ovuled ovary which is predominantly self-fertilized (She & Jiang., 2015). Once fertilized it develops into a pod which could be flat or inflated and can vary from 5-20cm in length and are usually 1-5cm in breadth. Pods usually have 2-6 seeds. The seed shape varies from round, oval or flat while the colour ranges from white, cream, beige, red, brown or black or variously speckled. The hilum of the seed is white, usually extending to one third around the seed. One hundred seed weight ranges from 20-50g.

2.3 Ecology of lablab

Lablab is adapted to warm climate with an average of temperature ranging from 18-30^0C. Many cultivars are tolerant to high temperature while others can withstand short duration of frost. Cold weather of less than 15^0C is not suitable for this legume because it affects pollination and seed set. Lablab is sensitive to day length. Most genotypes require short days to initiate flowering, but long-day cultivars exist as well (Cook *et al*.,

2005). Sennhen, (2015) observed that some accessions depend on temperature and photoperiod to trigger the flowering response. Cultivars which can produce flowers regardless of the photoperiod have been reported in India (Keerthi *et al*., 2014). The legume is drought tolerant and grow well in areas with average annual rainfall is 600-900 mm. It is considered to be more drought tolerant than other legumes like common beans (*Phaseolus vulgaris*) and cowpea (*Vigna ungiculata*) (Maass *et al.,* 2010). However, it requires adequate moisture during the crop establishment but it's later on able to withstand moisture stress due to its well-developed tap root with many laterals and adventitious roots. Lablab can be grown in a wide range of soils provided they are well drained. The crop cannot withstand water logging. It prefers sandy loam soil with an average soil pH of 6.5. The crop cannot tolerate soils with high aluminum content.

2.4 Global lablab production

Lablab has been in cultivation in Asia, Africa, Australia and America for a long period of time (Raghu *et al.* 2018). In Asia, it's mainly grown in South India for use as a vegetable. The area under cultivation in India is 0.085 million hectares with a production of 0.030 million tonnes (Keerthi *et al*., 2014). Karnataka is the leading lablab production area contributing to about 90% of both area and production of this leguminous crop in India (Laxmi *et al.,* 2015). It's considered the third most important vegetable in the central and south-western parts of Bangladesh with a total production area in the region of approximately 48,000 ha (Maass *et al*., 2010). In USA, a lablab variety "Rio Verde" developed at the AgriLife Research and Extension Center at Overton in 2006 is widely cultivated as a forage in Texas and California (Smith *et al*., 2008). The forage varieties highworth, Rongai and endurance are cultivated in USA and Australia.

There is little information about the current national production acreage and volumes of lablab in Kenya. This is mainly because the crop is still categorized as minor and therefore receiving little attention from ministry of agriculture and other stakeholders (MOA, 2014). In their survey, Kamotho *et al*., (2010) reported that lablab was grown in different agro-ecological zones of Kenya ranging from coastal areas to the highlands. In Meru County, one of the leading lablab producer, the acreage under lablab production has increased from 4282 ha in 2012 to 6345 ha in 2017 (Meru county, 2018). At the coastal region, Lamu county is the leading producer with more than 433ha of land under lablab production in 2017 (Lamu county, 2018).

In inter-cropping systems, lablab is mainly cultivated in association with maize, millet or sorghum (Kimani *et al*. 2012). The practice of intercropping lablab with maize is more common in Lamu, Maragwa and Thika probably because of the small land sizes (Kamotho *et al*., 2010). In coastal Kenya, lablab/maize system, has demonstrated to increase maize grain yield by 79% compared to maize monoculture with no fertilizer inputs (Njunie, 2002). The actual grain yield at farmers field are low, ranging from 0.3 t/ha to 0.71 t/ha which is way below the crop potential of over 1.5 t/ha (Boit *et al.*, 2013, Kamotho *et al.*, 2010, Tegemeo, 2010; Meru county, 2018; Lamu county, 2018).

2.5 Utilization and consumption trend of lablab in Kenya

In Kenya, lablab is mostly utilized as dry grain. Sizeable number of people also utilize it either for livestock feed or in conservation agriculture. Local consumption of lablab as vegetable is still very low (Kinyua and Kiplagat, 2012).

The data on the current consumption volumes of lablab in Kenya is lacking. However, there is evidence of increased demand for this pulse as indicated by the high grain price in the market. For instance, in 2015, the wholesale prices of lablab ranged between Kes 7000-8,000 per 90 Kg bag (MOAL, 2018). The lablab price was higher than that of

most other common legumes like common beans, cowpeas and pigeon peas. Currently, Kenya is a net importer of grain lablab mainly from the neighbouring countries suggesting further high demand of the pulse. In 2014, Kenya imported 121,430 MT of lablab grain (AFFA, 2014) and the demand has been increasing. The awareness of the food value of lablab and other indigenous crops is increasing especially at the urban areas and is expected to increase the demand of this pulse.

2.6 Lablab production constraints

Lablab production in Kenya is constrained by insect pests, lack of adapted genetic cultivars, drought and low soil fertility (Kinyua and Kiplagat, 2012) which contribute to low grain yield. The major insect pests include pod borer (*Helicoverpa armigera*), aphid (*Aphis craccivora*) and lablab bug (*Coptosoma cribraria*). Majority of the farmers do not control the insect pests due to high cost of the chemicals resulting in high yield losses (Kamotho *et al*., 2010; Kinyua and Kiplagat, 2012).

In Kenya, the production of lablab is usually undertaken in marginal environments characterized by depressed amounts of rainfall and reduced duration of precipitation. Majority of the lablab cultivars face unique challenge in such environment. Due to their long duration period, the crop flower and mature during the period of moisture stress usually experienced at the end of the season. The seeds formed during the end of season moisture stress remain unfilled thereby affecting the yields and quality of grain produced.

Most of the lablab production in Kenya is done under cereals legume intercrop system (Kinyua and Kiplagat, 2012). The most common practise is where the farmers grow a row of lablab between every row of maize or sorghum crop (Kamotho *et al*., 2015). The indeterminate growth habit nature of lablab cultivars poses a challenge under intercrop environment. Under this environment, the indeterminate cultivars are excessively

vegetative, they climb on, pull down and lodge the intercrop cereal thereby affecting the grain yield (Kamotho, 2015; Sullivan, 2003; Kgasago, 2006 Bonginkosi *et al*., 2018; Rapholo *et al*., 2020)

2.7 Adaptation of long duration lablab cultivars in short season environments

Production of late maturing lablab cultivars have contributed to the low grain yields at farm level especially at the semi-arid areas (Kamotho *et al.*, 2010). The previous lablab breeding efforts in Kenya have not prioritized on improvement for earliness to maturity (Kamotho *et al.*, 2010; KEPHIS, 2015). Since farmers have no access to improved early maturing varieties, they continue to cultivate their landraces or the few available medium to long duration varieties. Under rain-fed conditions in semi-arid areas which is usually characterized by short season, late maturing cultivars are subjected to end of season moisture stress which leads to yield loss (Sennhenn, 2015). Most dryland areas experience terminal drought which result in low crop yields (Daryanto *et al*., 2015). For instance, yield losses of more than 43.5% was reported when the moisture stress occurred during the reproductive stage of various leguminous crops (Daryanto *et al*., 2015). Even in drought tolerant crops like lablab, excess moisture stress during pods and seed development results in formation of few pods, shriveled grains and ultimately low grain yield (Fredrick *et al*., 2001; Nepomuscene *et al*., 2017).

To overcome the challenge of production of long maturity cultivars under short season environment, farmers plant lablab during the short season rainfall (October-December) but harvest their crop after the long season of the following year. The short season rainfall period is utilized to germinate and establish the crop. The short season rainfall is not enough to allow full crop cycle of the long duration cultivars. Therefore, the crop is allowed to remain in the field during the dry months of January and February awaiting

for the rains at March. During this period, the crop remain vegetative though it sheds most of the leaves. The vegetative lablab crop in field usually attract heavy infestation of insect pests such as aphids since during this dry period, there are only few other alternative hosts for the pests. After the onset of the long season rainfall, new shoots immediately develop followed by flowering and pod formation. Harvesting is done from June indicating that farmers can only reap one crop per year. The pests accumulated during the dry period usually attack the new shoots and pods and if not controlled can potentially reduce the grain yields. Farmers are discouraged from planting lablab due to the long time they have to wait before harvest and also the low grain yield. In Kenya, lablab is produced by small holder farmers with small parcel of land of less than 2.5 acres (Kinyua and Kiplagat, 2012) Small holder farmers prefer to cultivate short term crop varieties which would not hold their land for a long period of time. Therefore, selection towards consistently early flowering determinate varieties is important for the production success in areas with short growing season such as semi-arid areas.

2.8 Indeterminate varieties under intercropping systems

Maize/legume intercropping is a significant solution for food security among small scale producers (Thobatsi, 2009). In intercropping systems two or more crops are grown in the same field at the same time (Sullivan 2003) resulting into increased yield per unit area and thereby improved food security. This cropping system is very common to lablab small holder farmers in Kenya who usually have limited arable land (Kamotho *et al*., 2010). The most common practise is where the farmers grow a row of lablab between every row of maize or sorghum crop (Kamotho *et al*., 2015). In intercropping system proper selection of cereal/legume varieties is key as it increases yields of the intercrops and minimizes intercrop competition (Makgoga, 2014). The indeterminate

varieties usually climb on, pull down and lodge the intercrop cereal resulting in yield reduction of the cereal crop (Kamotho, 2015; Sullivan, 2003; Kgasago, 2006; Bonginkosi *et al*., 2018).

Due to the lack of determinate varieties with compact growth habit, a wider plant spacing is adopted when indeterminate lablab varieties are intercropped with cereals. For instance, a spacing of 90cm is recommended between rows of maize and lablab in semi-arid areas of Kenya. Wide plant spacing can result in low productivity ((Makgoga, 2014). Varying the planting date of the intercrop species has been previously used to the reduced competition between intercrop plants (Egbe, 2010). However, in areas that receive low rainfall like where lablab is commonly grown, early planting is recommended to ensure that the crop utilize all the available water during the growing season.

The indeterminate lablab cultivars have prolonged harvest period resulting from non-uniform maturity of the pods (Keerthi *et al*., 2014). Such cultivars with prolonged harvest period are however not suitable for farmers with small parcels of land who prefer short season crops for maximum utilization of their land. Breeding for varieties with rapid, uniform maturation and determinate growth habit is an important strategy to enhance food security among the lablab farmers especially in the semi-arid areas of Kenya.

2.9 Importance of early maturity in lablab

Successful production of lablab in marginal areas and under cereal legume intercrop environment requires use of genotypes with bush growth habit and with early maturity characteristics. In legumes, the greatest grain yield losses is when moisture stress occur during the flowering and post flowering stages (Manjeru *et al*., 2007; Nepomuscene *et al*., 2017). Water stress at flowering and pod maturity period cause flower and embryo

arbotion, lower percentage of pods and seeds formed and shrivelled grains. Therefore matching crop phenology events such as flowering and maturity to environmental conditions, mainly water supply is an important strategy of minimizing the impact of end of season moisture stress.

Most of the cultivars grown by farmers in Kenya have long maturity period. Kamotho (2015) evaluated 45 landraces and found that they flowered within 88- 124 days and matured within 123 days to 190 days after planting. In Machakos County, a major lablab growing areas in eastern region of Kenya, the everage rainfall duration recorded between 1961 and 2012 was less than 82 days each for both the short and long rain season (Sanhenn, 2015). It's therefore important to reduce the duration of crop growth in lablab cultivars in order to increase and stabilize their grain yield.

2.10 Factors affecting timing of flowering

The number of days taken from sowing to onset of flowering (flowering time) is a major component of crop adaptation and may directly affect yield, particularly in rain-fed environments (Sennhenn, 2015). Successful transition from vegetative stage to flowering stage is important for good pod set, seed load and therefore affect the grain yield. Flowering time in legumes is affected by genotypic variation and regulated by environmental factors such as photoperiod and temperature (Weller, 2015). Photoperiod is an important external factor that influence flowering of crops grown at diverse latitudes. Plants perceive photoperiod through photoreceptors in the leaf. Consequently, a signal called florigen is translocated to the apex of the shoot resulting to floral initiation. This process is regulated by the internal circadian clock which measure seasonal change. This is important in facilitating the matching of flowering to environment (Matthew *et al*., 2010). Flowering in long day plant is accelerated with lengthening photoperiod while in a short-day plant is accelerated as the photoperiod

shortens. In lablab, both short-day and long-day photoperiod types have been reported (Sennhenn, 2015). However, photoperiod insensitive (PI) or day neutral plants (DNT) have been also been reported in India.

Flowering time response to photoperiodic is affected by the temperature. Sennhenn, (2015), observed that increase in temperature decreased the critical day-length in short day lablab accessions. However, she noted that below the critical day-length, time to flowering in lablab was affected by temperature only. In Kenya, lablab is mainly grown between latitude 2^0 south and 2^0 north of equator with 12 hours day length. Therefore, the main important factor affecting the time to flowering of lablab under Kenyan condition is basically the temperature.

Phenological responsiveness to ambient temperature is observed across plant species. In plant species, leaves have the capacity to measure the diurnal temperature. The ambient temperature regulate several physiological processes in plants. Flowering is one of such processes regulated through the plant ambient temperature signalling pathway (Matthew *et al*., 2010). In general, lower temperature result in elongation of time to flowering. For instance, delayed flowering under lower temperature has been reported in lablab (Sennhenn, 2015). The flowering genes may influence maturity date through their effects on the onset of reproduction and duration of reproductive phase (Kumar *et al.*, 2001).

2.11 Factors affecting determinate growth habit

The determinate growth habit is a spectacular architectural modification in grain legumes. In some legumes like Lupin *sp* determinacy is expressed by separate vegetative and reproductive phases. In other legumes like Soybean, Faba bean, common bean, lablab and pea, the determinate types don't completely stop vegetative growth upon flowering but the vegetative growth is highly reduced. In lablab, the determinate

growth habit is characterized by the apex of the main stem terminating into a flower bud formation usually at 8^{th} to 10^{th} node (Keerthi, 2014). Lateral shoots latter develop and terminate in the same sequence. The new flowers are at the apex of the pedicel while the older flowers are at the base. In determinate plants, the first flower to open are those located on the main terminal stem (Kim *et al*., 1992). There are several reports of environment-dependent stem determinate growth habit in legumes. Kim & Okubo, (1995) reported that high temperature of above 30^0C or/and long photoperiod (more than 13 hours) shifted the growth habit of lablab from determinate to indeterminate. In their study in soyabean, Inouye *et al*., (1979) also reported that some plants shifted from determinate growth habit at 20^0C to indeterminate habit at 30^0C - 35^0C. This highlights the role of temperature and photo period in control of growth habit.

2.12 Breeding for Early Maturity and growth habit in Lablab

Successful production of lablab in marginal areas and under cereal legume intercrop environment requires use of genotypes with bush growth habit and with early maturity characteristics. The early maturing varieties could be ready for harvest within 100 days thus allowing farmers to plant and harvest more than one crop per year. Varieties with determinate growth habit have synchronous flowering, more uniform pod maturity and thereby enabling cost-effective harvesting.

Crop improvement programmes are based on the existing differences within crop species. In other words, no crop improvement is possible in the absence of diversity within the crop species. Genetic diversity is the natural divergence present within different individuals of same species. The diversity is as a result of presence of different alleles of a gene in individuals resulting in contrasting phenotypes. The natural divergence between crops have been widely identified and used in development of new varieties of crops (Bhanu *et al*., 2017). Under the current climate change and its adverse

effects on weather patterns, genetic diversity is an important reservoir of many unique traits conferring adaptability to the emerging crop growing conditions. Genetic diversity is also crucial in providing novel traits to adapt to the changing cropping systems such as intercropping caused by the diminishing arable land sizes especially among the small holder farmers (Govindaraj *et al*., 2015). The goal of every crop improvement program is to identify the diverse germplasm required to improve the deficiency of the elite and popular cultivars. Various techniques have been used in discrimination of plant genotypes. These include the use of morphological markers, biochemical evaluation and DNA marker analysis (Arunga *et al*., 2015; Bhanu *et al*., 2017; Govindaraj *et al*., 2015).

2.12.1 Morphological characterization

Traditional methods of estimating genetic diversity relied on use of morphological traits to discriminate genotypes within crop species (Gepts, 2006). In morphological characterization, different genotypes are grown in the field and the discrimination of the entries is done based on the visually observable traits. Morphological markers are direct, easy to score and less expensive than other markers. However, they are limited in number, dependent on the growth stage and often masked by factors in the environment hence limiting their application in genetic diversity assessment (Nadeem *et al*., 2018; Cholastova *et al*., 2012). However, since most of the morphological characters are greatly influenced by environmental factors and the developmental stage of the plant, new techniques which analyse diversity at biochemical or molecular level have been developed and successfully applied in evolutionary and diversity studies of different crops.

2.12.2 Biochemical markers

Germplasm characterization using biochemical markers involves separation of proteins or their variants (isozymes) into specific banding patterns (Bhanu *et al*., 2017).The isozymes are detected through electrophoresis and specific staining. The isozymes are the products of different alleles and not the genes. Use of biochemical markers for assessing diversity is fast and requires smaller amount of plant tissue. However, like the morphological markers, biochemical markers are few in number and are affected by changes in environment (Govindaraj *et al*., 2015).

2.12.3 Molecular markers

2.12.3.1 Restriction fragment length polymorphisms (RFLP)

Restriction fragment length polymorphism (RFLPs) is a hybridization-based marker and was developed in the late 1970s .The RFLP technique is based on the use of restriction enzyme. Restriction enzyme is a protein that identifies a specific, short nucleotide sequence and cleaves the DNA only at a particular recognition sequence site. DNA experience mutations such as substitutions/ deletions/ insertions which alter the fragments sequence and lengths of particular fragment. Using RFLP process, the differences in the length of DNA fragments in genotypes can be detected using some labelled probes. The purified DNA is fragmented by digesting it using restriction enzymes. The enzymes identifies specific recognition sites, cleaves the DNA generating a number of fragments generated from digestion. The differences in the size of fragments produced is based on the number of recognition sites identified. The restriction fragments generated during DNA fragmentation are separated using gel electrophoresis. The fragments are then transferred to a nylon membrane, hybridized to specific DNA probes and visualized by autoradiography. RFLP are highly reproducible between laboratories, are transferable across populations are codominant (can

distinguish between homozygous and heterozygous individuals) and are numerous in number. However, they are highly expensive mainly due to the enzymes and the probes used, time consuming, labour intensive and require larger amounts of DNA (Bhau *et al.*, 2016).

2.12.3.2 Random Amplified Polymorphic DNA (RAPD)

RAPD markers are based on PCR amplification of random DNA segments with short, arbitrary primers. RAPD uses a short DNA sequence (primer) usually ten base pairs long to amplify many loci along the DNA of the samples. The amplified DNA fragments are separated using gel electrophoresis and visualised under ultraviolet light after staining (Kumar *et al.*, 2009). RAPDs require low amount of samples DNA, are fast and easy to use, do not require prior knowledge of the sequence information of the genome and are highly polymorphic. Polymorphisms of DNA in RAPD occur due to deletions of a priming site, insertions or insertions that change the size of a DNA segment without preventing its amplification. (Williams *et al*., 1990). RAPDs have previously been utilized in diversity studies of crops such as lablab (Liu, 1996); Bambara groundnut (Massawe *et al*., 2003) and wheat (Li *et al*., 2012). This marker system is restricted by very poor transferability between laboratories.

2.12.3.3 Amplified Fragment Length Polymorphism (AFLP)

The AFLP technique combines both the use of restriction enzymes and PCR amplification. The genomic DNA is cut using both rare cutting (*Eco*RI or *Pst*I) and frequent cutting (*Mse*I or *Taq*I) restriction enzymes (Vos *et al,* 1995). This is followed by ligation of double-stranded adapters and selective amplification of the restriction fragments with adapter specific primers (Loh *et al*., 1999). The marker has also been

used in genetic diversity analysis studies such as in common bean (Maciel *et al*., 2003) in cowpeas (Coulbaly *et al.,* 2002), and in bambara groundnut by Massawe *et al.,* (2002). Using AFLP, Maass *et al*., (2005) discriminated two sub species of lablab namely *purpureus* and *unicinatus*. AFLP diversity analysis by Venkatesha *et al.* (2007) indicated narrow genetic diversity local lablab accessions at Bangalore India. While using AFLP markers, Kimani *et al*., (2012) reported low genetic diversity of some 50 Kenyan lablab accessions. The advantages of AFLP markers are that they are less labour intensive compared to RFLP, require small amount of DNA and they have number of polymorphic bands (Kimani *et al*., 2012; Nadeem *et al*., 2018; Nurmansyah *et al*., 2020).

2.12.3.4 Single nucleotide polymorphism (SNP)

Single nucleotide polymorphism (SNPs) are single nucleotide differences in the DNA sequences of individuals in the populations. A SNP occurs when individuals of the same species differ in their genome with a single nucleotide such as adenine (A), thiamine (T), cyosine (C), or guamine (G). They originate from a single point mutation event. SNPs can occur in both coding and non-coding regions of the genome. Within the coding regions, SNPs can either affect or not affect the amino acid sequence. Like SSRs, SNPs are abundant in genome of most organisms, though the latter is more abundant than the former. In contrast to multi-allelic markers, the bi-allelic SNP markers are suitable for automation thus making them to be very effective compared to other methods of DNA analysis. In addition, thousands of SNPs can be analysed at the same time through the use of DNA microarrays. Using 1200 SNP markers, Buckler *et al*., (2009) reported that flowering time in maize is controlled by small additive QTL. A soybean aphid resistance gene *Rag1*, was mapped between two SNP markers that

corresponded to a physical distance of 115 kb (Kim *et al*., 2010). Similarly, another aphid resistance gene, *Rag2*, was fine mapped to a 54 kb interval using SNP markers (Kim *et al*., 2010).

There are numerous SNP genotyping platforms which vary in terms of their allele discrimination approaches, detection methods and their reaction formats. The reaction formats include solution phase, liquid phase, gel electrophoresis and the next generation sequencing. The next generation sequencing reaction format is the most commonly adopted approach in SNP genotyping due to its cost effectiveness (Semagn *et al*., 2018). There are several SNP genotyping platforms that are currently being used. They include GoldenGate™, GeneChip™, GenFlex™, TaqMan™, Gold™ Assay and Mass-ARRAY™ and Kompetitive Allele Specific PCR, or KASP™ (Broccanello *et al*., 2018; Semagn *et al*., 2018).

SNPs have been used in germplasm characterization of several crops. For instance, Hamblin *et al.,* (2007) used 847 SNPs to study the structure of populations and the genetic relationships of 259 maize inbred lines. Van Inghelandt *et al*. (2010) used over 8000 SNP and 359 SSR markers to genotype 1,537 elite maize inbred lines and indicated that SSRs were better at clustering the inbred lines and more SNPs than SSRs were required for effective characterization of population structure and genetic diversity. Blair *et al*., (2012) used GoldenGate array technology to score 736 SNPs on 236 common bean genotypes and concluded that though the SNPs polymorphism was low, they were useful in discriminating the common bean races. Cichy *et al*. (2015) characterized an Andean diversity panel (ADP) comprising of 349 Andean, 21 Mesoamerican, and 26 Andean–Mesoamerican admixed accessions using the Illumina BARCBean6K_3 SNP chip. They reported that the SNPs were able to differentiate the

different race groups and grain types. A set of 708 common beans genotypes collected from breeders in Africa and CIAT breeding program were characterized using over 800 SNPs and were able to reveal duplicated lines with different names (Raatz *et al*., 2019). In rice, Adeboye *et al*., (2020) studied the population structure and genetic diversity of 176 accessions adapted to the upland ecology using 7063 genome-wide SNP markers from DArTseq and reported average polymorphism information content and heterozygosity of 0.25 and 0.03, respectively. There is no literature on development of SNPs or their application in lablab.

2.12.3.5 Diversity Array Technology (DArT)

DArT^{TM} is a high-throughput marker system first developed in 2001 by Diversity Array Technology Pty Ltd (DArT, Canberra, ACT, Australia (Jaccoud *et al.*, 2001). In DArT, DNA samples are pooled together, cut with chosen restriction enzyme and ligated to adaptors. A complexity reduction method using PCR primers with selective overhangs is carried out, amplified DNA fragments cloned and individual inserts arrayed onto a microarray to form a discovery array assay (Alam *et al*., 2018). DArT markers have been used for characterizing large number of individuals because they are amenable to high-throughput genotyping and are cost-effective.

Recently, DArTseq platform has been developed for genome-wide marker discovery. The platform uses restriction enzyme to reduce the genome complexity followed by sequencing of the DNA fragments on the Next Generation Sequencing (NGS) platforms (Edet *et al*., 2018). In most cases, almost all the DArTseq markers align matchlessly to the reference genome. This is attributed to the fact that this genotyping technology target areas of DNA with less repetitive sequences. DArTseq produces two types of markers namely silico DArT and SNP markers. SilicoDArT markers are microarray

markers that are dominant while SNPs are co-dominant markers (Alam *et al*., 2018). DArTseq markers have been applied in genomic studies of several crops such as, common beans (Nemli *et al*., 2017) macadamia (Alam *et al*., 2018), wheat (Edet *et al*., 2018) and pinneaple (Kilian *et al.*, 2016). Alam *et al*., (2018) characterized the genetic diversity and population structure in 80 macadamia cultivars using 3,956 DArTseq-Based single nucleotide polymorphism (SNP) markers. The average polymorphic information content (PIC) and gene diversity values were 0.21 and 0.412 respectively suggesting the efficiency of this marker in germplasm characterization. In a recent study, 186 chickpea genotypes were characterized using DArTseq-based single nucleotide polymorphism (SNP) markers. The markers revealed high genetic diversity among the chickpea genotypes demonstrating the usefulness of DArTseq-based SNP for genetic analysis of large number of germplasm (Farahani *et al*., 2019). However, this useful technology has not been used in understanding of genetic diversity and population structure in lablab germplasm.

2.12.3.6 Simple sequence repeats (SSR)

Microsatellites, or simple sequence repeats (SSR), are DNA stretches of di-, tri-, tetra- and penta nucleotides units repeated in tandem (Andrea *et al*., 2012). SSR polymorphism arise due to both unequal crossing over, mutation and DNA slippage occurring randomly in stretches of repetitive sequence (Yong-jin *et al.*, 2009). Mutation at the microsatellite loci occur by insertions or deletions and the rate of these mutations increase with the length of the repeats.

Microsatellites are preferred in breeding applications due to their co-dominant inheritance that allows distinguishing of homozygotes and heterozygotes in segregating populations (Kalia *et al.* , 2011; Wang *et al*., 2017; Li & Zhang, 2015; Zhao *et al*.,

2016). The SSRs are powerful genetic makers because of their high reproducibility from one laboratory to another and their ability to distinguish between heterozygotes and homozygotes (Gupta and Varshney, 2000). The hyper-variable nature of SSRs make them good at distinguishing closely related individuals (Kumar *et al*., 2009; Amkul *et al*., 2020). Microsatellites have been used in comparative and association studies, genetic diversity, individual identification, marker-assisted selection, population and evolutionary studies (Andrea *et al*., 2012). Microsatellites have been used in genetic studies of various crops. For instance, Wang *et al.,* (2018) evaluated 184 mung bean accessions obtained from the germplasm resources information center of the USDA using 38 polymorphic SSR markers. The SSRs revealed high diversity with an average of 4.2 alleles per locus and a PIC value of 0.65 per locus. Cabral *et al*., (2011) evaluated the genetic diversity of 57 dry bean accessions using common beans SSR primers. They found that 13 SSRs were polymorphic, giving 29 polymorphic alleles. The PIC varied from 0.11 to 0.5. Currently, there are very few studies on use of SSRs in diversity studies of lablab in Kenya. Shivashi *et al*., (2012) used 21 dry beans SSRs to evaluate 13 lablab accessions from Kenya. He recorded genetic distance of 0.0-0.62 with majority having less than 0.4 indicating narrow genetic diversity. Kamotho *et al*., 2015 evaluated 96 lablab accessions from Kenya using 10 SSRs. The mean amplification was 4.3 allele per primer, PIC of 0.63 and average heterozygosity of 0.38. The study revealed that the accessions had a narrow genetic base.

Numerous number of SSRs are now available for many crops such as Soya bean (Li *et al*., 2011), chickpea (Sethy *et al.,* 2003), cow pea (Gupta *et al*., 2010), pigeonpea (Saxena *et al.,* 2010), mung bean (Sing *et al*., 2013) and common bean (Blair *et al.,* 2011). In lablab, Zhang *et al*., (2013) identified 22 expressed sequence tags (EST) SSRs and used them to genotype 19 Chinese accessions. Chapman, (2015) mined SSRs from

transcriptome data and identified 2427 lablab loci. However, only 12 primers were developed and tested for amplification in lablab. Recently, (Keerthi *et al.,* 2018) used 55 lablab EST based SSR markers to characterize 16 lablab genotypes. The number of SSR markers developed in lablab is so far fewer than those reported for other legumes.

2.13. Strategies of SSR Development

There are several methods and protocols developed for the isolation of SSR loci. These methods include, Construction and screening of SSR enriched genomic libraries (Kalia *et al.,* 2011), screening of available sequenced EST databases (Tang *et al.,* 2008), testing of transferability of markers from other related species (Satya *et al*., 2016) or sequencing of whole or parts of the genome using high-throughput technologies (Zane *et al*., 2002). Isolation of SSRs through enriched genomic libraries is time consuming and produce few microsatellites. Searching for SSRs from EST databases is a cheaper way of isolation but is largely limited to those species or close relatives for which there is a sufficiently large number of ESTs available. For underutilised crops like lablab, little number of EST are available due to minimal research on these crops. For instance, in their search for SSRs from lablab EST at NCBI database, Zhang *et al*., (2013), obtained 459 EST and isolated only 22 SSRs. Microsatellites have poor transferability which is generally limited to closely related species. Shivakumar *et al.,* (2017) screened 275 SSRs from various legumes and reported 45.8% transferability to lablab. In other studies lower rate of transferability has been reported between genus of leguminous crops (Garcia *et al*., 2011).

The advent of high throughput sequencing technologies such as next generation sequencing (NGS) have produced millions of nucleic sequences in less time than the traditional methods (Ma *et al*., 2019). This makes the NGS platforms suitable for large-

scale discovery of genetic markers at reduced prices. There are different NGS technologies platforms developed including 454 FLX (Roche), Solexa (Illumina), SOLiD (Applied Biosystems) and HeliScope True Single Molecule Sequencing (Helicos) (Taheri *et al*., 2018; Krishna *et al*., 2019). Roche 454 and Illumina are among the NGS platforms widely used for developing SSR markers (Vieira *et al*., 2016). Roche 454 technology has been used in marker discovery in crops such as pigeon peas (Dutta *et al*., 2011), chick peas (Hiremath *et al*., 2011), winged bean (Vatanparast *et al*., 2016), faba bean (Suresh *et al*., 2013), finger millet (Gimode *et al*., 2016) and lentils (Kaur *et al*., 2011).

The development of SSR using NGS utilize either genomic DNA sequences or DNA synthesised from single-strand RNA (cDNA). Direct sequencing using DNA instead of RNA is straighter forward, as it does not require library construction and normalization. However, transcriptome sequencing (RNA-Seq) is more preferred because it can discover markers embedded in function gene sequences and directly associated with transcribed gene (Taheri *et al*., 2018). Transcriptome sequencing is particularly important for non-model plant species like lablab that lack adequate background genomic information (Ma *et al*., 2019). RNA-seq has successfully been applied for SSR development in many leguminous plant species such as sesame (Wei *et al*., 2011), pea and faba beans (Luke *et al*., 2012), peanut (Yin *et al*., 2013), common beans (Wu *et al*., 2014) and mung bean (Liu *et al*., 2016). There is however limited application of RNA-Seq for development of SSRs and generation of genomic information in lablab. The only available information on application of RNA-Seq in lablab is by Chapman, (2015) who identified 2427 SSRs from RNA sequences and managed to develop and test amplification of only 12 primers.

2.14 Marker Assisted Selection

Phenotypic selection is a crucial tool of conventional plant breeding that has successfully been utilized in improvement of genetic traits (Acquaah, 2012). Plant breeders are usually concerned with the phenotype (physical appearance) of a plant. Phenotype of a plant is determined by the genetic constitution of the individual and the environmental conditions under which the individual grows (Arnel, 2011). The environmental influence and genetic interactions can mask the presence of specific genes, making it challenging for breeders to select the plants that they really seek. Improvement of polygenic traits require breeders to develop a large population to provide a possibility of combining desired alleles at several loci into a single line (Luby & Douglas, 2000). This is followed by evaluation of hundreds of the lines in multiple environments to identify the recombinants with desired traits. Though phenotypic selection has been used in improvement of many genetic traits, phenotypic evaluation can be costly and time consuming. Molecular markers may provide a potential solution to some of these problems

Molecular Markers are short pieces of DNA that correspond to particular sequences of DNA in the plant genome (Guo-Liang Jiang, 2013). Markers can inform breeders about the presence of desirable alleles in individual plants. If the presence of the desirable alleles can be confirmed through the use of markers, breeders do not have to resort to costly and time consuming phenotypic evaluation to determine whether or not the alleles are present. The selection of target trait indirectly using molecular markers that are closely linked to the underlying genes is referred to as marker assisted selection (MAS) (Yunbi and Jonathan 2008, Ribaut *et al*., 2010, Guo-Liang Jiang, 2013). Availability of adequate number of polymorphic markers and knowledge of marker-trait association are prerequisite for successful MAS (Babu *et al*., 2004). In MAS,

breeders are interested with those markers that are closely associated with the target traits or tightly linked to the genes for they provide sufficient guarantee for the success in practical breeding (Guo-Liang Jiang, 2013). Like in many other underutilised crops, lack of suitable markers has been identified as one of the most restricting factors for the lablab bean breeding studies (Zhang, 2013). Among the many different types of molecular markers that can be used in plant breeding, simple sequence repeat (SSR) markers are the markers of choice due to their simplicity in use, reproducibility, and most of all, they are a co-dominant in nature (Sulieman *et al*, 2016).

This study attempted to fill the gap of lack of molecular markers for lablab by using the new NGS platforms to develop more lablab based molecular markers. The availability of molecular markers will benefit the breeding program of this less studied crop. Once developed, the markers that are closely associated with important traits like growth habit should then be identified for use in the selection process.

2.15 Inheritance of Qualitative Traits

2.15.1 Inheritance of time to flowering in lablab

In legumes, Flowering time (*Ft*) genes have been identified to be responsible for integrating the environmental cues for flowering at the leaves and signalling the production of flowering hormones at the shoot apex (Weller &Ortega, 2015). The genes influence maturity through their effects on the onset of reproductive growth and then the subsequent duration of the reproductive phase. In many cases the flowering loci have been identified as QTL, but in other cases as major loci (Weller & Ortega, 2015). In *Phaseolus* species like common beans, recessive allele on locus *Ppd* which is on linkage group one (LG1) has been reported to control photoperiod response and early flowering under long days condition (Koinange *et al.*,1996). Four non-allelic genes (*ef-*

1, ef-2, ef-3, ef-4) have been identified to control flowering time in chickpea (Gaur *et al.,* 2015). A major recessive gene 'sn' was identified to be responsible for early flowering in lentil variety *Precoz* (Sarker & Erskine, 2006). Murfet, (1973) indicated that duration to flowering in *Pisum* was under control of two genes (S_2 & E) with the earliness being recessive. In India, a photoperiod insensitive early maturing lablab cultivars have been developed. A single gene is responsible for the photoperiod response to flowering time in these genotypes (Keerthi *et al.,* 2014; Keerthi *et al*., 2016). The recessive allele at this locus confer early flowering and maturity. However, the inheritance pattern of flowering time under local Kenyan environment and local genetic background is still lacking.

2.15.2 Inheritance of growth habit in lablab

Determinate growth habit in many legumes has been reported to be under the control of major genes. In Soybean two genes were involved in the control of determinate growth habit. In faba bean and garden pea, recessive monogenetic control was reported (Filipetti, 1986., Singer *et al*., 1991) while two genes were reported to control determinate growth habit in Chickpea. . The information on genetics of growth habit in lablab is limited and conflicting. For instance, Keerthi *et al*., (2014) reported that growth habit is controlled by three genes, GH1, GH2 and GH3 of which, one (GH1) is independent and the other two (GH2 and GH3) are complementary. However, in Bengalulu India, Keerthi *et al*., (2016) indicated that growth habit was under control of two genes which exhibited classical complementary epistasis. Earlier on, Gowda *et al*., (2008) had reported that growth habit was controlled by single dominant gene but while using different genetic populations they indicated that the trait was controlled by two dominant genes. Peeta *et al*., (2017) reported that determinate bush growth habit in lablab is controlled by single recessive gene. Notably, all these studies were conducted

in Southern Asia and used their local genetic populations. In Kenya, the information of genetics of growth habit under the local environment and using local genetic populations is still lacking.

2.16 Character Association

One of the main thrust in crop improvement programme is to enhance yield. Breeders have to seek for effective and efficient ways of selecting superior plants in breeding program. Yield is a complex trait and difficult to manipulate but can be best understood if it's broken down into components (Jiaqin *et al*., 2009). The understanding of how various yield components are related to one another will allow an indirect selection of yield based on those characters.

The knowledge of how crop characters are associated to one another is crucial for successful breeding. The expression of a qualitative trait like yield in a plant is a product of the interaction of numerous physiological processes. These processes include developmental, reproductive and morphological features which contribute to the yield of a plant. A good understanding of the relationship of plant characters and the yield is essential because the final yield is the sum total of effects of all its related traits (Verma *et al*., 2015). Breeders seek to break down a complex trait like yield into yield components, aiming to find suitable selection criteria to improve the complex trait. Correlation coefficient analysis enables the understanding of the nature and degree of association between any two measurable characters. Such knowledge is useful in identifying the traits suitable for indirect selection of yield thereby improving the efficiency of the selection process.

Strong positive phenotypic and genotypic correlation between the grain weight yield and the number of pods per plant has been reported in lablab (Salim *et al*., 2014; Das

et al., 2015). This suggested that the grain weight of lablab is greatly determined by the number of pods per plant. The plants with high number of pods per plant produced high total grain weight and vice versa. The number of pods per plant could be valuable index for effective selection towards higher yield. Ali *et al.,* (2005) observed positive correlation of yield per plant with pod length and pod diameter but observed a negative significant correlation of pod weight with flowers per inflorescence and number of pods per inflorescence. Noorjahan *et al*., (2019) studied 31 genotypes of lablab bean and observed a positive and strong correlation of days to flowering and days to first pod harvest, pod yield per plant and pod length. The pod length could be contributing to the pod yield per plant through the number of seeds per pod or seed size. This is because longer pods are likely either to have more seeds per pod or have larger seeds. A positive significant correlation of days to flowering and days to first pod harvest suggest that the early flowering plants were also early in maturity. Early flowering, can therefore be used to select for the early maturing genotypes.

Though simple correlations coefficients have been used by plant breeders to understand the extent of association of characters, they are however limited in that they do not measures the magnitude of direct and indirect contribution of a component character to a complex character. Path analysis is a powerful method for partitioning the direct effect of a trait on yield and its indirect effects through other traits (Hama *et al*., 2016) thereby providing information on which characters have direct and indirect contribution towards yield.

2.17 Path Analysis for relatedness of Characters

In path analysis, correlation coefficients is partitioned into two components. The first component is the path coefficient that measures the direct effect of a predictor variable

upon its response variable, with the second component being the indirect effect(s) of a predictor variable on the response variable through another predictor variable (Wang *et al*., 2011). Plant breeders have used path analysis to identify traits which could be used as selection criteria for improvement of crop yield. For instance, Muhammad *et al.*, (2004) identified that biological yield and harvest index had high direct effect on grain yield of chickpea and recommended that the two traits could be used as selection criteria for yield. However, in a different study using path analysis, Ozveren *et al*., (2006) demonstrated that number of seeds per plant and number of fully developed pods per plant had the highest direct influence on yield. In mung bean, Moushree *et al*. (2014) identified positive direct effect between seed yield and 100 seed weight and days to 50% flowering and recommended that these two traits to be considered as selection criteria for yield improvement in mungbean breeding programs. A positive relationship between grain yield and number of pods per plant, pod weight, 100 seed weight and pod size was shown in lablab (Devmore *et al*., 2015; Das *et al*., 2014; Parmar *et al*., 2013; Magalingam *et al.*, 2013; Chaudhari *et al*., 2013). They suggested that selection for yield in lablab should be based on these traits. However, no study has been carried out to determine the relationship of the yield and yield parameters of the determinate lablab cultivars especially under various growing areas in Kenya.

2.18 Genetic Variability, Heritability and Genetic Advance

The understanding of the extent and nature of variability existing in plant is of paramount significance as it provides the basis for effective selection. The phenotype of a plant is determined by its genetic composition, the environment in which the plant is grown, and the interaction of genotype with environment. Genotypic variability on the other hand is the variation that can be attributed to genes that encode specific traits and can be transmitted from one generation to the next. This type of variability that can

be consistently expressed from one generation to another is very useful for the breeders in their selection processes.

The genetic coefficient of variation (GCV) provides a good measure for comparing genetic variability in quantitative traits. However, high GCV alone is not sufficient for the determination of the heritable variation as it simply measures the extent of genetic variability present for the character (Magalingam, *et al*., 2013). Heritability is the proportion of phenotypic variance that is due to genetic variance. Heritability is therefore of great value to breeders since it represent the variation that is transferred from one generation to another (Chaitanya *et al*., 2013). Large heritability values indicate the relative ease with which selection can be made based on the phenotype. However, the practical application of heritability in crop improvement is better if accompanied by high genetic advance (Hailu *et al*., 2016).

Genetic variability can be created through crossing individual parents followed by population advancement through selection. Selection identifies superior individuals with high genetic potential thereby changing the population mean of the trait in a positive way in the next generation (Acquaah, 2012). Genetic advance (GA) is the difference between the mean phenotypic value of the offspring of the selected parents and the whole of the parental generation before selection (Hailu *et al*., 2016). The GA depends on the total phenotypic variation of the population, heritability and the selection pressure imposed by the plant breeder (Acquaah, 2012).

The understanding of the GCV and heritability estimates is important as it enlightens on the expected genetic advances to be made through selection (Magalingam, *et al*., 2013). However, better information is obtained when these parameters are considered together rather than individually. For example, a trait which combines high GCV, GA

and heritability estimates are desirable because it indicates the trait is under additive gene action and therefore simple selection for the trait would be effective (Hailu *et al*., 2016). However, traits possessing low genetic advance with high heritability indicates the presence of non-additive gene action, thus simple selection procedure in early segregating population will not be efficient for screening desirable traits.

Considerable genetic variability for important quantitative traits has been reported in Lablab and other legumes. Moderate genotypic variation (GCV, 35-60%) for number of pods per plant was reported by Chaudhari *et al*., (2013), Parmar *et al*., (2013) Parmar, *et al*., (2014), Das *et al*., (2015) while Magalingam *et al*., (2013), Pawar and Prajapati (2013) recorded low genotypic variation of the trait. In common beans, low to moderate genetic variability for number of pods per plant was recorded by Roy *et al*., (2006). Molosiwa (2012) reported high genotypic variation for pods per plant in Bambara groundnut. Moderate genetic variability for pod yield per plant was recorded in lablab by Pamar, *et al*., (2013) and Chaudhari *et al*., (2013). Low genetic variability for days to 50%, days to maturity, no of seeds per plant, pod length and 100 seed weight was reported in lablab by Chaudhari *et al*., (2013) and Parmar *et al*., (2013). Similar results were obtained for these traits in *Phaseolus vulgaris* by Roy *et al*., (2006) and Prakash and Ram, (2014).

In lablab, high heritability for number of pods per cluster (Magalingam *et al.*, 2013), number of pods per plant (Magalingam *et al.*, 2013; Chaudhari *et al*., 2013), green pod length (Magalingam *et al.*, 2013) green pod width (Magalingam *et al.*, 2013; Parmar *et al*., 2014), pod weight (Magalingam *et al.*, 2013), protein content (Magalingam *et al.*, 2013; Parmar *et al*., 2014; Chaudhari *et al*., 2013), number of branches per plant

(Chaudhari *et al*., 2013; Moushree *et al*. 2014) and plant height (Moushree *et al*. 2014) has been reported. However, Patel *et al*., (2011) while evaluating 63 lablab genotypes noted low heritability for pedicel length (23.41%) and number of pods per plant (23.40%). In common beans, Roy *et al*., (2006) recorded high heritability estimates for pod length, duration of flowering, % protein content and days to 50% flowering while low values were recorded for seed per pod and plant height. While evaluating some bambara groundnuts genotypes, Molosiwa (2012) recorded different heritability estimates for days to maturity in a greenhouse experiment (10%) and field experiment (95%). However, in both experiments most of the characters recorded more than 70% heritability.

In their study, Das *et al*., (2015), observed both high heritability and high genetic gain for number of pods per plant, and pod weight and suggested that the traits could be effectively used for direct selection of high pod yield genotypes in the lablab improvement programs. Similarly, both high genetic advance as percentage of mean (> 40%) and heritability (> 75%) was also recorded in number of pods per cluster, number of pods per plant, green pod length, width, weight and protein content by Magalingam *et al*., (2013). In Mung bean, high heritability estimates together with genetic advance was observed on 100 seed weight and days to 50% flowering (Moushree *et al*., 2014). There is limited amount of work on heritability and genetic advance on the yield characters of determinate lablab genotypes and especially under the Kenyan conditions hence this study.

CHAPTER THREE

Development of Simple Sequence Repeat (SSR) Molecular Markers for Lablab (*Lablab purpureus*)

3.1 Abstract

Lablab (*Lablab purpureus*) is one of the important pulse crop used as human food, feed, in soil conservation, in enhancing soil fertility and weed management. However, the crop has remained underutilised in many regions. Lack of adequate and specific molecular markers has hindered rapid improvement of this crop through molecular breeding. Development of such markers would hasten and enhance lablab variety development. In this study, we isolated mRNA from leaves and shoots samples, conducted transcriptome sequencing using 454 Titanium FLX system to discover genic-SSRs and develop molecular SSR markers for lablab. BLASTX analysis revealed that the largest number of lablab transcriptome sequence matches were on genus Vigna (V. angulasis & V. radiata) 45.2% followed by Phaseolus vulgaris 22% and genus Glycine (G. max and G. soja) at 11.8%. In this study, 446 genic SSRs were identified from 3140 assembled contigs. The overall density of the genic SSR was 202 SSR per Mbp. Tri-nucleotide (51.3%) and di-nucleotide (46%) were the most abundant repeat units. After several filtering procedures, 145 SSR primer pairs were designed from the unigene sequences and 75 of these SSRs were validated for amplification. The marker validation showed that 70 (93.3%) of the genic SSR markers were amplifiable on lablab genome. Majority of the primers (60%) did not produce secondary products indicating their high quality standard. Eighteen of these PCR based markers were utilised to investigate diversity of 10 lablab accessions collected from different regions of the globe. The 18 markers yielded 42 alleles, with an average of 2.3 alleles per locus and a moderate PIC of 0.34. The sequence information and DNA markers generated in this study are useful for genetic improvement of this underutilised leguminous crop.

3.2 Introduction

Lablab (*Lablab purpureus)* is one of the important pulse crops produced and consumed mostly in Asia and Africa. The dry seeds and fresh green pods serves as vital source of protein (20-28%), minerals such as Zinc (34mg/kg) and iron (57 mg/kg) (Maass *et al*., 2005; Kimani *et al*., 2012; Sennhenn, 2015).

Successful improvement of lablab requires availability of wide genetic variation which consist of reserviour of genes required for development of profitable and sustainable varieties of the future. The genetic resources are screened to identify those germplams containing desirable traits which are used as parents in crossing to combine the positive traits into improved varieties. The conventional approaches of improvement of

important plant traits have mainly relied on phenotypic and pedigree information which are influenced by environment and are time and labour intensive. However, the implementation of conventional selection procedures combined with the use of innovative tools like molecular markers should considerably accelerate the breeding process.

Molecular markers are fixed marks in the genome and are found at specific locations of the genome and are used for identification of specific genetic differences. They have previously been used for various molecular breeding related activities such as assessing genetic diversity, back crossing of gene of interest, gene/QTL mapping and marker assisted selections (Nadeem *et al*., 2018, Wang *et al*., 2017). There is a range of molecular markers systems available for genetic studies of various crops. Simple sequence repeats (SSR) markers are tandem repeats of di- to tetra- nucleotides sequence motifs found both in coding and non-coding region of the genome (Molosiwa, 2012). They exhibit high polymorphism due to their variation in the number of repeat units that arise due to slippage-like events that happen at the repetitive sequences during the DNA replications. The microsatellites are able to detect large number of allelic variants because of their multi-allelic nature. Microsatellites are preferred in breeding applications due to their co-dominant inheritance that allows distinguishing of homozygotes and heterozygotes in segregating populations (Kalia *et al*. , 2011; Wang *et al*., 2017; Li & Zhang, 2015; Zhao *et al*., 2016).

One major challenge towards application of molecular tools in the breeding of lablab is due to the inadequate number of codominant molecular markers such as SSR (Zhang *et al*., 2013). To date, only few SSR markers have been developed and tested on lablab genome. For instance, Chapman, (2015) developed and tested for amplification 12 SSR

in lablab, Keerthi *et al.*, (2018) used 55 lablab EST based SSR markers to characterize 16 lablab genotypes while Zhang *et al*., (2013) identified 22 SSRs and used them to genotype 19 Chinese lablab accessions. The number of SSR markers developed in lablab is so far fewer than those reported for other legumes, such as cow peas (Chen *et al*., 2017b), mung bean (Chen *et al*., 2015), Common bean (Blair *et al*., 2011), chickpea (Choudhary *et al*., 2009) and pigeon pea (Bohra *et al*., 2017).

There are several methods and protocols developed for the isolation of SSR loci. These methods include, Construction and screening of SSR enriched genomic libraries (Kalia *et al*., 2011), screening of available sequenced EST databases (Tang *et al.*, 2008), testing of transferability of markers from other related species (Satya *et al*., 2016) or sequencing of whole or parts of the genome using high-throughput technologies. The introduction of high throughput sequencing technologies such as next generation sequencing (NGS) have produced millions of nucleic sequences in less time than the traditional methods (Ma *et al*., 2019). This makes the NGS platforms suitable for large-scale discovery of genetic markers at reduced prices. There are different NGS technologies developed including 454 Roche, Illumina genome analyzer, ABI SOLID, Pacific Bioscience, Ion Torrent, Oxford Nanopore and Qiagen GeneReader (Taheri *et al*., 2018). Among the NGS platforms, Roche 454 and Illumina have been widely used for developing SSR markers. Roche 454 technology has been used in marker discovery in crops such as pigeon peas (Dutta *et al.*, 2011), chick peas (Hiremath *et al*., 2011), winged bean (Vatanparast *et al*., 2016), faba bean (Suresh *et al*., 2013) and lentils (Kaur *et al*., 2011).

The development of SSR using NGS utilize either genomic DNA sequences or DNA synthesised from single-strand RNA (cDNA). Transcriptome sequencing (RNA-Seq)

is more prefered because it can discover markers embeded in function gene sequences and directly associated with transcribed gene (Taheri *et al*., 2018). Transcriptome sequencing is particularly important for non model plant species like lablab that lack adequate background genomic information (Ma *et al*., 2019). RNA-seq has successfully been applied for SSR development in many leguminous plant species such as sesame (Wei *et al*., 2011), pea and faba beans (Luke *et al*., 2012), peanut (Yin *et al*., 2013), common beans (Wu *et al*., 2014) and mung bean (Liu *et al*., 2016). There is however limited application of RNA-Seq for development of SSRs and generation of genomic information in lablab. The only available information on application of RNA-Seq in lablab is by Chapman, (2015) who identified 2427 SSRs from RNA sequences and managed to develop and test amplification of only 12 primers.

The generation of expressed sequences resources through transcriptome sequencing and development of more reliable and informative molecular markers for lablab is important for enhancing of our understanding about this underutilized crop and to facilitate molecular breeding. In this study, we conducted transcriptome sequencing in lablab by using 454 Titanium FLX system to discover genic-SSRs and develop molecular SSR markers.

3.2. Objectives

The broad objective was to identify some SSR markers in lablab. The specific objectives included:

(1) To generate lablab EST data set and determine functional annotation of the ESTs.
(2) To assess frequency and distribution of genic SSRs in lablab genome
(3) To determine and validate novel SSR markers for lablab

3.3 Materials and Methods

3.3.1 Plant Materials

Lablab genotype GBK 012026 was used for the development of EST SSRs. The genotype is gene bank of Kenya (GBK) accession collected from the coastal region of Kenya. It is a late maturing accession that has white flowers and with indeterminate growth habit. The stem has no anthocyanin and the leaves are light green in colour. Its seeds are brown in colour and are usually three to four in a pod.

3.3.2 RNA Isolation

The seeds of the accession were sown in pots in a greenhouse at Sutton Bonington Campus, University of Nottingham in Britain. At pre-flowering stage, young leaves and terminal shoots were harvested and immediately frozen in liquid nitrogen and later stored at -80^0C freezer. Total RNA was extracted using RNeasy plant mini kit (Qiagen) according to manufacturer's protocol. The RNA was extracted as follows: About 100mg of the frozen leaves and shoots were weighed, placed in liquid nitrogen and grounded thoroughly with mortar and pestle. 450 ul of buffer RLC containing β-ME was added to the 100mg tissue powder and vortexed vigorously. The lysate were transferred to a QlAshredder spin column which was placed on 2 ml collection tube. 0.5 volume of 96% ethanol was added to the lysate and mixed by pipetting. The sample was transferred to RNeasy spin column placed on 2ml collection tube and centrifuged at 10,000 rpm and the flow through discarded. The spin column membrane was washed with 500ul of buffer RPE. The RNA was eluted by adding 50ul of RNase free water directly to the spin column.

3.3.3 Construction and Sequencing of cDNA Library

The total RNA was delivered to Centre for genomics and genetics, University of Nottingham for mRNA purification and cDNA libray construction and sequencing.

The mRNA was isolated from total RNA using Oligotex mRNA kit (Qiagen) according to manufacturer's protocol. A cDNA library was constructed following the standard Roche Diagnostics protocol (Roche 454, 2009). The protocol consists of the following steps: fragmentation of the RNA, denaturation of RNA, first strand cDNA synthesis, second strand cDNA synthesis, double stranded cDNA purification, fragment end repair, adaptor ligation, small fragment removal and library quantification.

Sequencing of the cDNA was done using 454 Titanium FLX system (Roche diagnostics) at the Centre for genomics and genetics, University of Nottingham. The generated sequence reads were assembled into, contigs, isotigs and isogroups using the GS denovo assembler (Newbler 2.3). In this study only a ¼ plate of the lablab cDNA library was sequenced mainly because of limitation of funds.

3.3.4 Gene Annotation and analysis.

The assembled contigs were utilized as BLASTx queries for homology searches against NCBI non redundant (NR) (http://www.ncbi.nlm.nih.gov/) protein databases using BLAST2GO program. The BLASTx hits were set at below E-value of 1E-10. Mapping was done to identify and retrieve the gene ontology (GO) terms associated with the obtained BLAST hits. The GO terms were finally assigned to the hit sequences. The GO terms were classified according to molecular function, biological process and cellular component ontologies.

3.3.5 Microsatellite discovery

The assembled contigs were screened for microsatellite using perl script MIcroSAtellite (MISA) program. The misa.pl script was downloaded together with misa.ini file which contains the search parameters. The following parameters were used for search of

microsatellite in the assembled lablab contigs: at least 6 repeats for di- and 5 repeats for tri-, 5 repeats for tetra-, penta- and hexa nucleotide SSRs.

In this study, the number of SSRs obtained from the assembled contigs were few to obtain numerous polymorphic SSR markers. We therefore extended our search for more SSRs to the unassembled raw read sequences. In this search, microsatellite were defined as those having at least six repeats for dinucleotide and with five minimum number of repeats for tri-, tetra-, penta- and hexa-nucleotides.

3.3.6 Primer Design

All the primers were designed from the flanking sequences of SSR using Primer3 web interface program (http://fokker.wi.mit.edu/primer3/input.htm). Primer pairs were designed using the following parameters: The amplicon range was 100-400 bp, the primer annealing temperature was restricted to $50–65^0$ C, the (guanine and cytosine) GC content was in the range of 40–60% while the primer length was 18–27 bp. The designed primers were synthesized by Europhin MWG Operon Company.

3.3.7. PCR gradient optimisation for primer annealing temperature.

The polymerase chain reaction (PCR) involves *in vitro* amplification of DNA through a series of three polymerization cycles, the DNA denaturation, primer templates annealing and DNA synthesis by thermostable DNA polymerase. Optimization of PCR (gradient PCR) involves testing of different annealing temperatures to identify the best primer annealing temperature. In this study, 75 primer pairs were screened and optimised for annealing temperature.

3.3.8 DNA Isolation

DNA extraction was done at Kenya Agricultural and Livestock Research Organization (KALRO) Biotechnology Research Centre, Nairobi. A panel of five diverse lablab genotypes were selected to be used in the process of optimization of the 75 primer pairs. The five accessions include KDD, ILRI 13692, ILRI 21048, DL1002 and GBK 028663 as shown in Table 3.1 below. DNA was isolated from each of the five accessions using modified cetrimonium bromide method (CTAB) (Doyle 1987). The modification done was addition of 2% of polyvinylpyrrolidone (PVP) into the CTAB extraction buffer. The extraction buffer (2% CTAB) was prepared using 2% CTAB, 100mM Tris-Hydrochloric acid (pH 8.0), 1.4M Sodium chloride (NaCl), 50mM Ethylenediaminetetraacetic acid (EDTA), 2% Polyvinylpyrolidone (PVP) and 10µl of 2% β-mercaptoethanol. Approximately 0.3g of the leaf tissues were placed in self-standing tubes each containing a ceramic bead. Nine hundred micro-liters (900µl) of extraction buffer was added to the leaf tissue containing the ceramic bead. The samples were crushed by geno-grinder machine (Benchtop homogenizer) set at revolution speed of 4 meter per second for one minute. The samples were then incubated in a water bath with constant shaking at 65^0C for 15 minutes. The tubes were then centrifuged at 14000 rpm for five minutes. Six hundred microliters (600µl) of the supernatant were transferred into a fresh eppendorf tube and an equal volume of chloroform: isoamyl alcohol (24:1) added. The above step was repeated. Four hundred microliters (400µl) of the aqueous phase was transferred to a fresh tube and an equal volume of ice-cold isopropanol added and mixed by inverting several times and incubated at -20^{o}C for 2 hours to precipitate the DNA. The tubes were centrifuged at 14000rpm for 5 minutes. The supernatant was decanted leaving the DNA pellet at the bottom of the tube. The pellets were washed using 500µl of 70% ethanol and centrifuged for one

minute at 14000rpm before they were air dried for one hour. The dried pellets were re-suspended in 50µl of sterile distilled water. RNA was removed by adding two microliters (2µl) of pancreatic ribonuclease A (RNase A) (10mg/ml) and incubating the samples for 30 minutes at 37^{o}C. The samples were stored at -20^{o}C.

Table 3.1: Lablab Genotypes used in Determining Optimum PCR Annealing Temperature of the New SSR Molecular Markers

Genotype	Maturity group	Seed testa colour	Growth habit	Flower colour	Seed size	Sub species
KDD	Early	Cream	Determinate	White	Small	*purpureus*
ILRI 13692	Medium	dirty brown-cream	Indeterminate prostrate	Purple	Very large	*bengalensis*
ILRI 21048	Very late	Black	Indeterminate prostrate	Purple	Very small	*unicinatus*
DL 1002	Medium	Black	indeterminate	Purple	Medium	*purpureus*
GBK 028663	Medium	brown-cream	indeterminate	Purple	Medium	*purpureus*

3.3.9 DNA quantification and quality determination.

The quantity and quality of genomic DNA was examined by comparing the template DNA isolated from samples with a DNA ladder (gene ruler) of one kilo base (1 kb) in a 0.8% agarose gel using 1x TBE buffer and viewed in a gel box (G: Box, Syngene). The concentration and quality was further determined at optical density (OD) readings of 260 nm and 280 nm using a Nanodrop spectrophotometer (Thermo Scientific NanoDrop 2000C). The concentrations were used to guide the normalization of DNA of each sample at a concentration of 20 ng/ μL. Additionally, the ratio of OD 260/280 was provided by the Nanodrop and this gave an indication of purity of the samples. Pure DNA has OD260/OD280 value of 1.8 and a deviation from this signifies the presence of contaminants that inhibit PCR reaction.

The DNA was later diluted to a concentration of 20ng/uL for use in the PCR.

3.3.10 Dilution of primers.

The primers were synthesized by Europhin MWG Operon Company, UK and were supplied as lyophilized powder. The primers were diluted in TE buffer (1 mM Tris-HCL pH8.0/0.01 mM EDTA) to give a stock solution with concentration of 200 pmol/ µL. A working solution of 2pmol/µl was then prepared by diluting 1 part of the stock 200 pmol/µl with 990 µl of sterile water and stored at -20^{o}C.

3.3.11 Gradient PCR

Polymerase chain reaction (PCR) optimization was carried out using a composite DNA samples obtained from the five genotypes. Each of the five DNA samples were diluted to 20ng concentration. Equal amount of each the diluted DNA samples was mixed to form a composite sample which was used for PCR reaction.

The following were used in the 25ul PCR; Lyophilised master mix bead (easy to use beads) containing (dNTPs, $MgCl_2$, Tris-HCl (pH 9.0), KCl and Taq), 19µL of sterile distilled water, 2µL of 2 pmoles/ul forward primer, 2µL of 2 pmoles/ul reverse primer and 2µL of template of 20ng/ul DNA. Amplification was carried out in a Thermocyler machine (Techne-TC 412, Applied Biosystems Veriti systems) programmed with the following regime and 35 cycles: 94°C for 3 minutes, 35 cycles of 94°C for 1 minute, 8 temperatures ranging between 50-65^{o}C (50°C, 51°C, 52.9°C, 55.7°C 59.1°C, 62°C,63.8°C 65°C) for 1 minute, 72°C for 1 minute and final extension at 72°C for 10 minutes and final hold at 4^{0}C. A total of 75 lablab EST SSR markers (Appendix I) were screened.

3.3.12 Gel electrophoresis of PCR products.

The PCR products for each marker were separated on 2% agarose gel at 80V for one hour. Agarose powder was dissolved in Tris-borate EDTA (1x TBE) buffer by slowly

boiling in a microwave oven. The agarose was allowed to cool and 1.5uL of 1mg/ml ethidium bromide was added to the gel. The warm agarose solution was then poured into the gel tray in which combs were inserted to form sample wells. The gel was allowed to solidify for 30 minutes before immersing in the electrophoresis tank containing 100ml TBE buffer. The samples were run alongside 3.0μL 1kb DNA ladder at 80 volts for 60 minutes. The amplified products were viewed under UV light in a gel box (G: Box, Syngene).

3.3.13 Genotype characterization using SSR primers.

Eighteen SSR primers that showed good amplification were tested for their informativeness in lablab genome. The study was carried out at biotechnology research centre, Kabete. The primers were amplified on genome of 10 lablab accessions selected from different groups of origin, seed colour, maturity duration, growth habit among others (Table 3.2). The following were used in the 25ul PCR; Lyophilised master mix bead (easy to use beads) containing (dNTPs, MgCl2, Tris-HCl (pH 9.0), KCl and Taq), 19µL of sterile distilled water, 2µL of 2 pmoles/ul forward primer, 2µL of 2 pmoles/ul reverse primer and 2µL of template of 20ng/ul DNA. Amplification was carried out in a Thermocyler machine (Techne-TC 412, Applied Biosystems Veriti systems) programmed with the following regime and 35 cycles: 94°C for 3 minutes, 35 cycles of 94°C for 1 minute, annealing temperature of 57°C for 1 minute, 72°C for 1 minute and final extension at 72°C for 10 minutes and final hold at 4^0C.

Table 3. 2: Important attributes of 10 lablab accession used to amplify the 18 LabRRT primers

Accession name	Sub species	Growth habit	Flower colour	Floweri ng time	Seed colour	Seed size
KDD	Purpureus	Determinate	White	Early	Cream	Small
ILRI 13692	Bengalensis	Indeterminate	Purple	Medium	Brown	Very big

ILRI 011615	Purpureus	Indeterminate	Purple	Early	Black	Medium
ILRI 14411	Purpureus	Indeterminate	Purple	Late	Black	Small
ILRI 21081	Unicinatus	Indeterminate	Purple	V. late	Black	Very small
ILRI 14440	Unicinatus	Indeterminate	Purple	Late	Black	Very small
GBK 028663	Purpureus	Indeterminate	Purple	Medium	Brown	Medium
Njoro	Purpureus	Indeterminate	Purple	Medium	Black	Medium
Kahuro	Purpureus	Indeterminate	Purple	Medium	Black	Medium
ILRI 21048	Unicinatus	Indeterminate	Purple	V.Late	Black	Very small

The PCR products were size separated using PolyAcrylamide gel (PAGE). In this experiment, 6% PAGE was used. The 6% gel was prepared as follows; 53.8 ml of distilled water was put in a clean 500 glass beaker, 18.8 ml of 40% acrylamide/ bis-acrylamide 19:1 was added to the beaker, 2.5ml of 50x TAE (Tris acetate EDTA) was then added and mixed by stirring, one tablet containing ammonium persulfate (APS) and TEMED was dissolved in 50ml distilled water in a 50 ml falcon tube, the APS/TEMED solution was then added to the glass beaker and gently stirred without introducing any air bubble, the solution was immediately poured into the gel tray with in-built combs and was covered with glass plate treated with silane solution and allowed to gel for 1 hour. The glass plate with the attached PAGE was then transferred to the horizontal PAGE (hPAGE) (Cleaver Scientific) electrophoresis tank containing 1X TAE solution. Two micro liters (2uL) of samples were loaded into the wells alongside 3.0µL 1kb DNA ladder and run at 140V for 3hrs. The glass plate was removed from the gel tank and placed in another glass tray for post staining of the gel. To stain the PAGE, 6uL of ethidium bromide was dissolved in 200 ml of water and poured into the glass tray to cover the glass plate containing the PAGE. The tray was kept in a dark room for 1 hour. The staining solution was then drained off and the gel de-stained using clean water for 30-45 minutes. The amplified products were then viewed under UV light in a gel box.

3.3.11 Data collection and analysis

The gradient PCR products for each marker were separated on 2% agarose gel at 80V for one hour. The amplified products were viewed under UV light in a gel box and captured by a camera. The intensity of the amplified band was scored. The clearly visible and blight bands were scored as strong bands, those that were faint and not clear were scored as weak bands while absence of amplified band was scored as nil band. The annealing temperature that produced the strongest band was recorded for each marker and was considered as the optimal annealing temperature for that marker.

The PCR products arising from genotyping of 10 accessions using 18 SSRs were size separated on 6% PolyAcrylamide gel (PAGE). The amplified products were then viewed under UV light in a gel box. The images were captured using a camera connected to the gel box and saved in the computer. To score the bands, the images were pasted in MS power point program. On the gel image, a straight line was drawn across the bands that had migrated the longest distance. The accessions that had bands at that position were scored as 1 while those that didn't have were scored 0. The straight line was moved to another allele and scoring was done. The process continued until all the bands were scored.

The number of alleles per locus, major allele frequency per marker, gene diversity, heterozygosity and the polymorphic information content (PIC) for each locus was computed using PowerMarker version 3.25 (Liu and Muse, 2005).

3.4 Results

3.4.1 Sequencing and de novo assembly of reads.

Sequencing of cDNA library was done using FLX Titanium system (Roche 454) at Deep Seq in the University of Nottingham and assembly done using GS denovo assembler (Newbler 2.3). In this study, a ¼ plate of the lablab cDNA library was used

and produced 144,619 set of high quality cDNA sequence reads with total size of 2,198,045 bp. These sequences were assembled into 3140 isotigs. The length of the isotigs ranged from 25 bp to 9905 bp with an average length of 700 bp. A total of 1184 (37.7%) had lengths not longer than 500 bp. Majority of the isotigs 1627 (51.8%) were 500-1000 bp long. The rest of the isotigs (316) (10%) had lengths longer than 1000 bp (Figure 3.1).

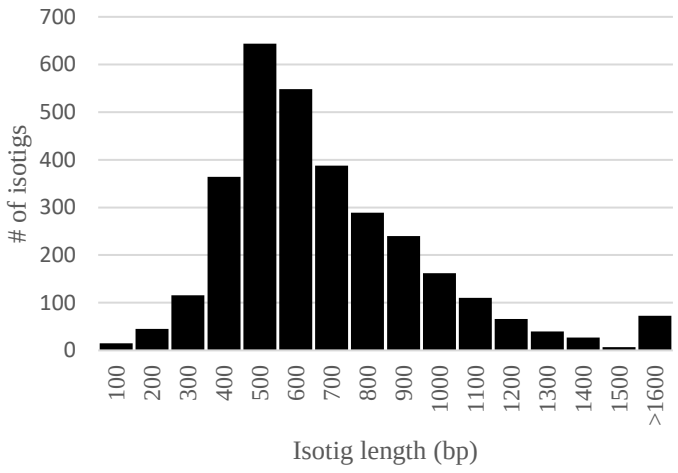


Figure 3. 1: The length distribution of isotigs assembled from lablab cDNA sequence reads

3.4.2 Sequence annotation

An important use of genomic data is in associating the individual sequences with known biological functions in order to make them more meaningful. This process of DNA sequence annotation is important for relating the genome data to functions of different proteins. In this study, the assembled isotigs sequences were used for homology searches against the NCBI non redundant (NR) protein databases. The results of the number of contigs allocated to different annotation categories are shown in figure 3.2 below. Of the 3140 isotigs queried against the NR database, 2527 (80.4%) had

significant hits at greater than 1×10^{-10} stringency. This meant that approximately 80.4% of isotigs could be assigned to putative protein functions. About 19.4% (611) of the isotigs did not have similar sequences in the database.

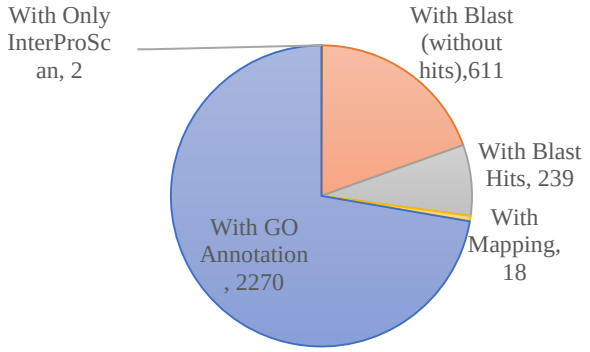


Figure 3. 2: The number of lablab cDNA assembled contigs allocated to different annotation categories

The number of gene ontology (GO) terms assigned to each of these significant sequences ranged from 1-22. However, majority of them (96%) were assigned less than 10 GO terms (Figure 3.3).

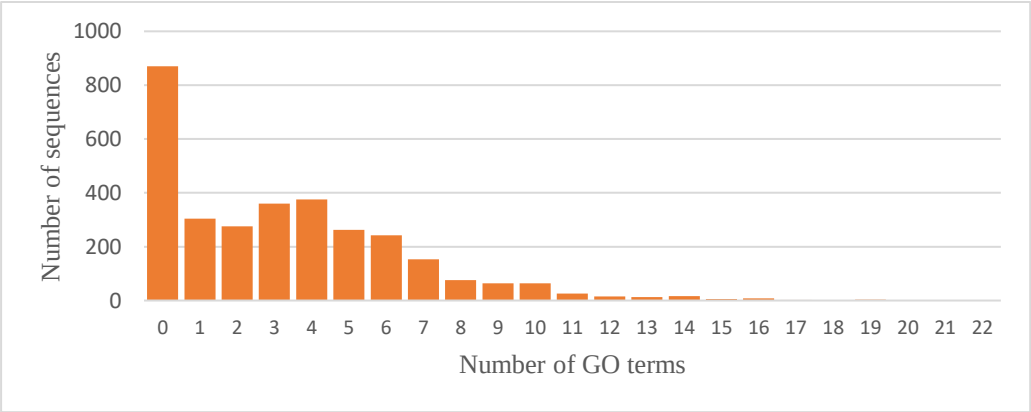


Figure 3. 3: The annotation distribution of GO terms

Approximately 63.78% of the aligned isotigs, showed very strong homology to sequences in the non-redundant (Nr) protein database (E-value < 1.0E-50), whereas the

remaining 36.22% had weak homology (E-value between 1.0E-5 and 1.0E-50) (Figure 3.4).

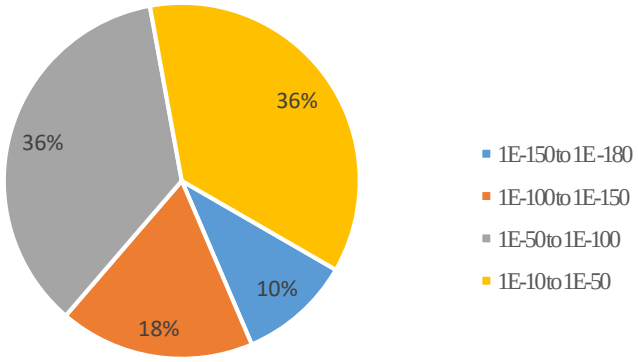


Figure 3. 4: E value distribution of BLAST hits for the assembled isotigs in NR database

The results of the similarity distribution of the aligned unigenes compared to the sequences in the NR database are shown in Figure 3.5 below. About 92% of unigenes had a similarity > 80%, 6.7% of the unigenes had a similarity between 60% and 80% while only 0.4% of the unigenes had < 60% similarity with NR sequences. This suggests that for the majority of these isotigs, highly homologous sequence annotation information could be obtained from the NR database.

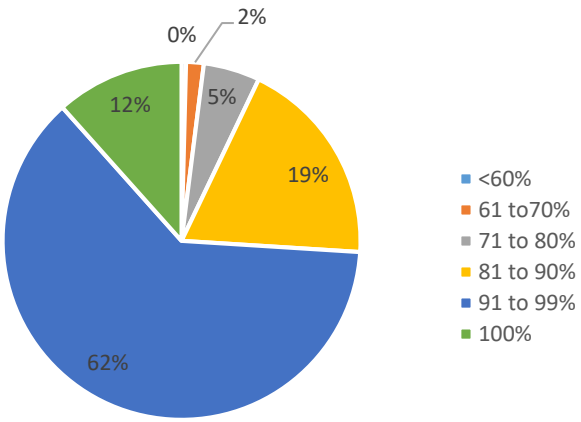


Figure 3. 5: Similarity distribution of contigs with sequences in the Nr database

The top blast hits of annotation of lablab transcriptomes are given in Figure 3.6 below. Among the leguminosae species, *Phaseolus vulgaris* was ranked first with 703 (22.4%) of the top blast hits followed by *Vigna angulasis* (16.7%) *Vigna radiata* var radiata (15.1%), *Vigna angularis* var angularis (13%), *Glycine max* (6.9%), *Glysne soja* (4.8%), *Cajan cajan* (3.3%) and *Mucuna pruriens* (3.1%) (Figure 4). In other words, 75.2% of the assembled sequences matched with sequences from just seven leguminosae species. This further demonstrate that the sequencing and assembly strategy used in this study was quite efficient.

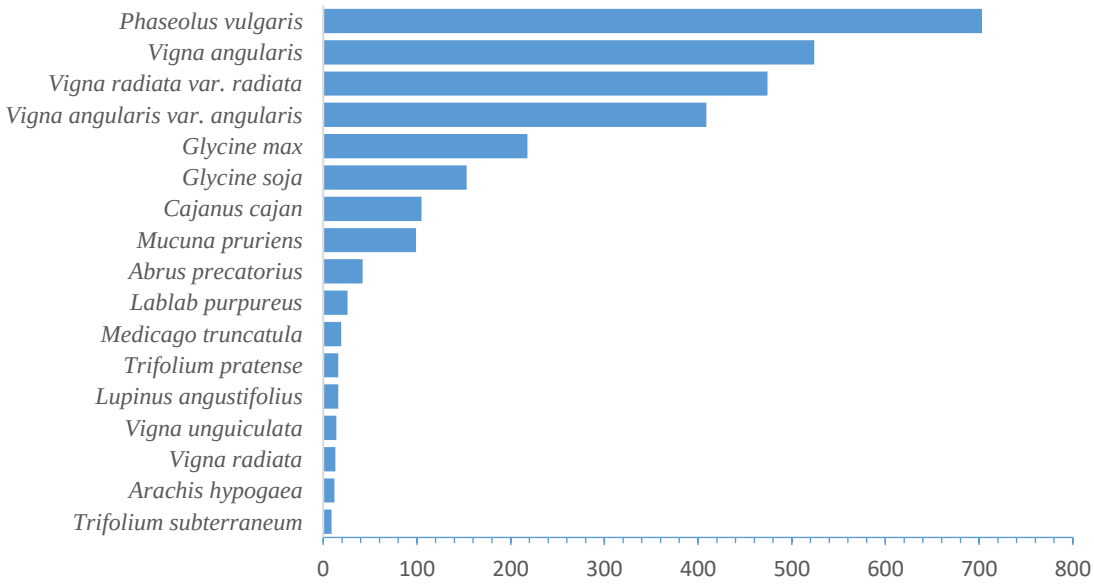


Figure 3. 6: The species distribution of the top BLAST hits of annotation of lablab transcriptomes

The contigs generated in this study captured a broad range of different types of transcripts, as indicated by the variety of Gene Ontology (GO) terms assigned. The contigs were annotated to various gene ontology categories including molecular function, biological process, and cellular component. A total of 3572 terms were allocated to the molecular function, 3539 under biological process, and 3232 under the

cellular component. Among the contigs annotated to biological process, approximately 74% of these contigs accounted for cellular process, metabolic process and biological regulation (Figure 3.7). In the category of cellular components, about 69% of the contigs were annotated to cell, cell parts, organelle and cell membrane components (Figure 3.8). Majority of the contigs (92%) annotated to molecular functions were associated with catalytic and binding activities (Figure 3.9).

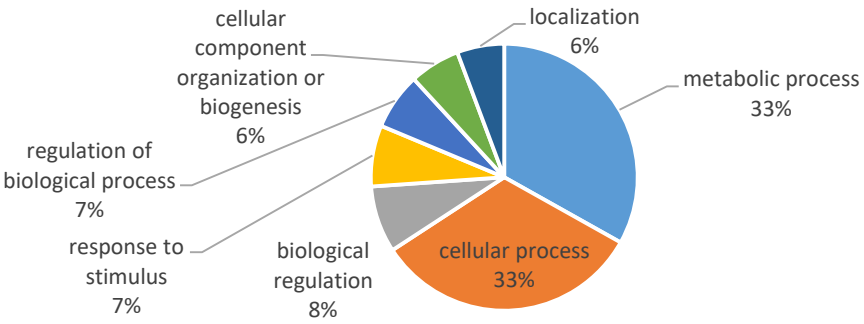


Figure 3. 7: GO category assignment: Biological process

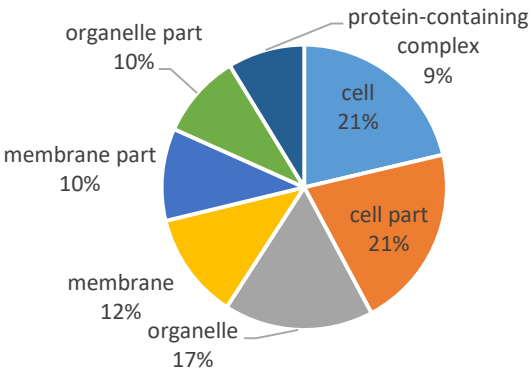


Figure 3. 8: GO category assignment: Cellular process

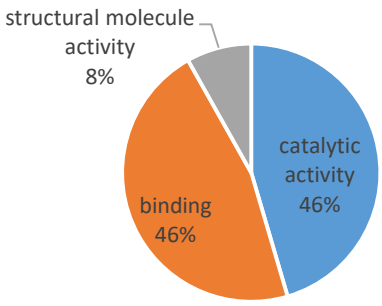


Figure 3. 9: GO category assignment: Molecular function

Majority (82.3%) of the isotigs from where primer pairs were successfully designed were associated with 1 to 14 GO terms (Appendix 3.2). The highest number of GO terms (14) were associated with the isotig01972 (Lab T11). The high number of isotigs with GO terms of homologs retrieved by BLAST in this study suggest that the quality of assembly of these isotigs was high.

3.4.3 Frequency and distribution of SSR from assembled isotigs

A search for SSRs in the 3140 contigs was done using MISA software. The results showed that a total of 446 SSRs were identified from 396 SSR containing sequences. This represent an overall density of 202 SSR/Mbp (i.e., one SSR found in every 4.9 kbp) across the genome. Tri-nucleotide (229, 51.3%) was the most abundant repeat unit, followed by di- (205, 46%), tetra- (6, 1.3%), penta- (3, 0.7%), and hexa- (3, 0.7%) nucleotides (Figure 3.10). This indicate that both di-nucleotide and tri-nucleotide constituted about 97% of the motifs. The total length of di- to hexa nucleotide repeats accounted for about 0.24% (5402/2198045 bp) of the assembled genome (Figure 3.10).

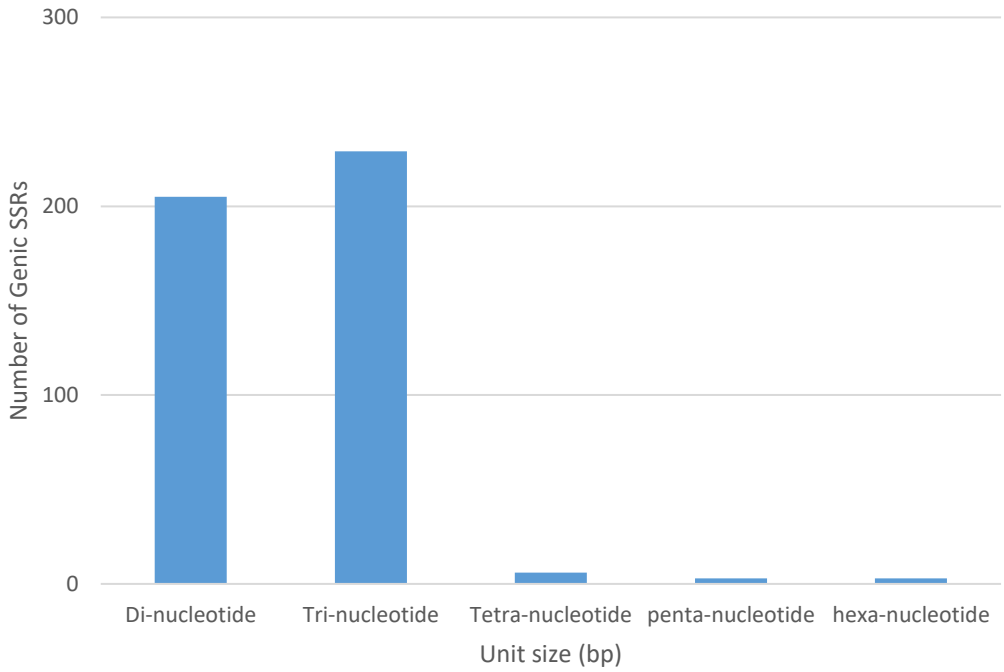


Figure 3. 10: Frequency distribution of the genic-SSR motif length from the lablab isotigs

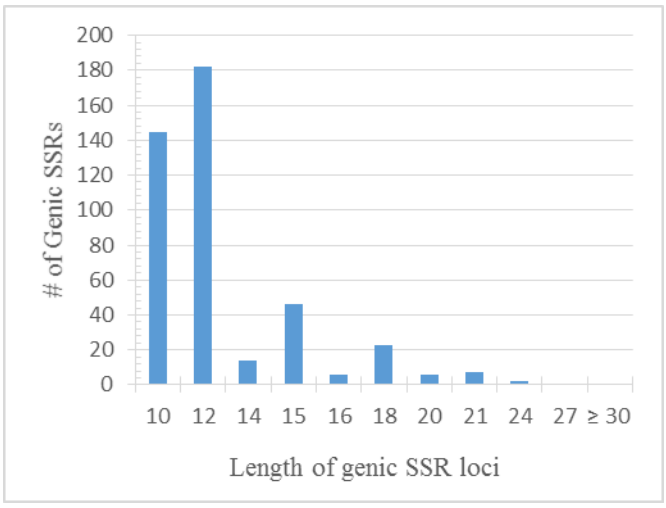


Figure 3. 11: Frequency distribution of the of genic-SSR repeat unit derived from lablab isotigs

The number of SSR repeats ranged from 4 to 78. It was observed that those SSRs with lower number of repeats were more abundant than those with many repeats. For instance, the SSRs with five repeats (199, 44.7%) were the most abundant (Figure 3.11),

followed by those with four tandem repeats (154, 34.6%), six tandem repeats (48, 10.7%), and seven repeats (23,5.2%).

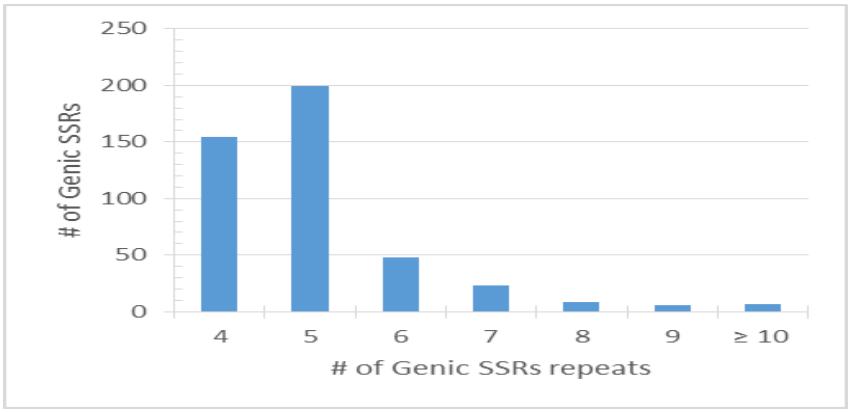


Figure 3. 12: Frequency distribution of the number of genic-SSR repeat unit derived from lablab isotigs

The most common single repeat loci were those containing 12 bp (182, 42.0%) followed by those with 10 bp (145 33.5%), 15 bp (46, 10.6%) and 18 bp (23, 5.3%). The longest SSR locus was 156 bp.

A total of 23 repeat motif types were identified, based on sequence complementarity. Tri-nucleotides were the most frequent motifs (54.6%) followed by di-nucleotides (42.7%), while the other nucleotides were observed at 2.7 %. The AT/AT repeat motif (29.4%) was the most common dinucleotide. The most common tri-nucleotide repeat motifs were AAG/CTT (12.1%) and ATC/ATG (11.5%) respectively (Table 3.3).

Table 3. 3: Frequency distribution of the 12 most frequent genic-SSR repeat motifs in the lablab transcriptome and the number of repeats within each motif

Order	Repeats	4	5	6	7	8	9	≥ 10	Total	%
1	AT/AT	-	93	19	8	4	4	3	131	29.4
2	AG/CT	-	39	9	5	2	1	3	59	13.3
3	AAG/CTT	34	7	8	4	-	1	-	54	12.1
4	ATC/ATG	37	10	4	-	-	-	-	51	11.5

5	AAT/ATT	20	7	2	2	1	-	-	32	7.2
6	ACC/GGT	18	4	2	1	-	-	-	25	5.6
7	AGC/CTG	17	6	1	-	-	-	-	24	5.4
8	AGG/CCT	12	3	-	-	1	-	-	16	3.6
9	AC/GT	-	13	-	1	-	-	-	14	3.1
10	AAC/GTT	6	5	-	-	-	-	-	11	2.5
11	CCG/CGG	5	2	-	-	-	-	-	7	1.6
12	ACT/AGT	2	2	1	-	-	-	1	6	1.3
13	ACG/CGT	3	-	-	-	-	-	-	3	0.7
14	Others	-	8	2	2	-	-	-	12	2.7
	Total	154	199	48	23	8	6	7	445	100

3.4.4 Length distribution of SSR motifs mined from the raw reads

A search for more SSRs was extended to the raw (single trimmed 454 run transcriptome) reads. About 144,614 raw reads sequences were examined and 3882 sequences which contained SSRs were identified. Of these 1904 sequences, 396 had more than one SSR while 551 SSRs were in compound formation. Tri-nucleotide (1553, 40%) was the most abundant repeat unit, followed by di- (1182, 30%), tetra- (300, 8%), penta- (410, 11%), and hexa- (437, 11.2%) nucleotides (Figure 3.12). The most frequent SSRs were those with less than 10 repeats (Figure 3.13). Generally, a negative correlation existed between abundance and microsatellite length.

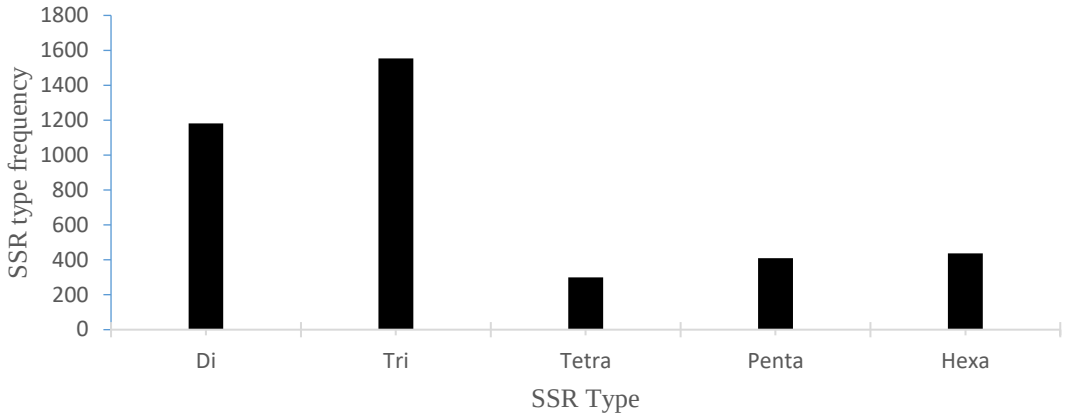


Figure 3. 13: Frequency distribution of different types of genic-SSRs obtained from the lablab raw reads.

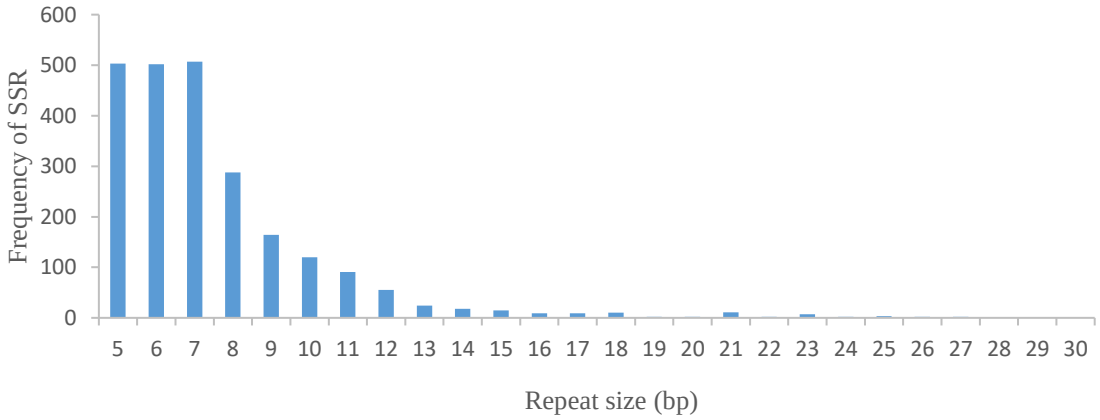


Figure 3. 14: Frequency distribution of the genic-SSRs of different repeat sizes obtained from the lablab raw reads.

The AT/AT repeat motif (73%) was the most common dinucleotide, followed by (AG/CT)n. This study did not identify any (GC)n SSR. The most abundant tri-nucleotide repeat motifs was ACT/AGT (18%) followed by AAC/GTT (16%), ATC/ATG (15%), AAG/CTT (15%) and AAT/ATT (15%). More than 90% of the tetra-, penta- and hexa-nucleotide were dominated by the AT rich motifs (AAAT/ATTT, AAAAT/ATTTT, AAAAAT/ATTTTT).

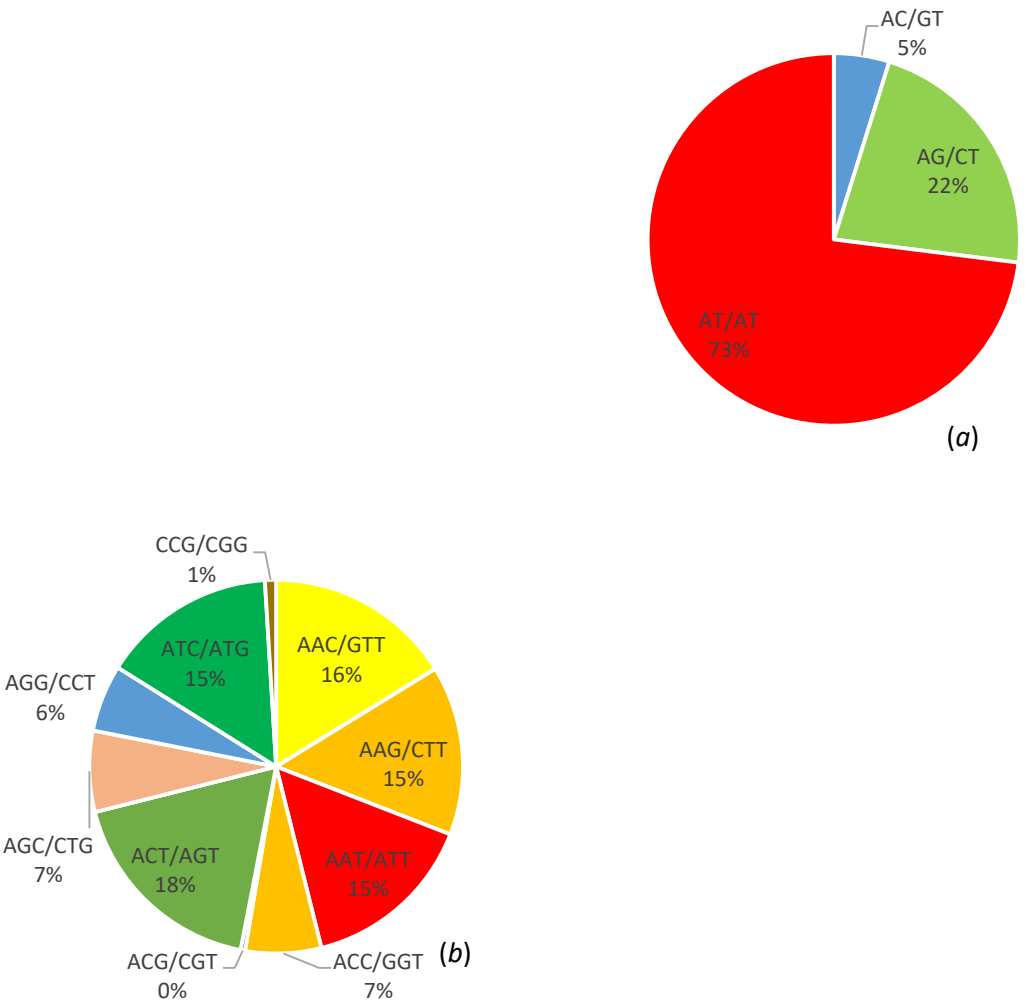


Figure 3. 15: Frequency distribution of the genic-SSR repeat motifs obtained from the lablab raw reads. (a) di-nucleotide repeat motif (b) tri-nucleotide repeat motif.

3.4.5 Design of Primers from assembled contigs and unassembled raw reads

Not all the fragments containing SSRs were suitable for designing primers. This is because some SSRs were located too close to the end of the fragments and therefore did not have enough flanking region to accommodate primer design. For others, the base composition of the flanking sequence was unsuitable for primer design. For instance, the composition of the flanking region was considered unsuitable for primer design where the GC content was outside the range 40–60%. In this study, from the 446

SSRs identified from the 3140 isotigs only 34 unique primers pairs were designed using the Primer 3 software (Table 3.4). The majority (52.9%) of those primers designed in this study were for tri-nucleotide repeat motif followed by di-nucleotide (10).

Table 3. 4: Characteristic of the SSR sequences identified from assembled cDNA library of L. purpureus and designed primer pairs

Repeat unit	Repeat class	Number of the motifs	cDNA (%)	Designed SSR
Di-	AC/GT	14	7	1
	AG/CT	59	29	5
	AT/AT	132	64	4
	Total	**205**	**-**	**10**
Tri-	AAC/GTT	11	5	0
	AAG/CTT	54	24	6
	AAT/ATT	32	14	2
	ACC/GGT	25	11	3
	ACG/CGT	3	1	1
	ACT/AGT	6	3	3
	AGC/CTG	24	10	0
	AGG/CCT	16	7	1
	ATC/ATG	51	22	2
	CCG/CGG	7	3	0
	Total	**229**		**18**
Tetra-	AAAT/ATTT	2	33	0
	AACC/GGTT	1	17	1
	ACTC/AGTG	2	33	1
	ATGC/ATGC	1	17	0
	Total	**6**		**2**
Penta-	AAAAT/ATTTT	2	33	
	AGCAT/ATGCT	1	17	1
Hexa-	AAAAGG/CCTTTT	1	17	1
	AACACC/GGTGTT	1	17	1
	AGATAT/ATATCT	1	17	1
	Sub-total	**6**		**4**
	Total	**446**		**34**

To design more SSR primers from the transcript sequence data, we extended the search of SSR loci to the raw (unassembled 454 transcriptome) reads. About 144,614 raw reads sequences were examined and 2454 SSRs were identified. The 2454 SSRs were contained in 1904 sequences. Only 600 raw reads had ample flanking sequence length suitable for designing primers. One hundred and eleven (111) unique primers pairs were thereafter designed from these 600 raw reads. The majority (54%) of the primers

successfully designed were for tri-nucleotide repeat motif, followed by di- (28.8%) while the least number of primers were for penta- (4.5%) and hexa- (2.7%) (Table 3.5). The details of the designed primers are given in appendix I.

Table 3. 5: Characteristic of the SSR sequences identified from raw reads of cDNA library of L. purpureus and number of designed primers

Repeat unit	Repeat class	# of designed SSR	Repeat unit	Repeat class	# of designed SSR
Di-	AC/GT	3	Tetra-	ACGC/CGTG	1
	AG/CT	25		ACTC/AGTG	1
	AT/AT	4		AGAT/ATCT	2
	Total	**32**		**Sub-Total**	**11**
Tri-	AAG/CTT	20	Penta-	AACGT/ACGTT	1
	AAT/ATT	8		AAGAG/CTCTT	2
	ACC/GGT	10		AGCAT/ATGCT	2
	ACG/CGT	3		**Sub-total**	**5**
	ACT/AGT	4	Hexa-	AAAAAC/GTTTTT	0
	AGC/CTG	3		AAAAGG/CCTTTT	1
	AGG/CCT	6		AACACC/GGTGTT	1
	ATC/ATG	6		AGATAT/ATATCT	1
	Sub-total	**60**		**Sub-total**	**3**
Tetra-	AAAC/GTTT	1		**Total**	**111**
	AAAG/CTTT	1			
	AAAT/ATTT	1			
	AACC/GGTT	2			
	AAGT/ACTT	1			
	ACAT/ATGT	1			

3.4.6 Optimization of primers

Among the designed primers (from raw reads and contigs), 75 were selected to determine their potential as molecular markers in lablab. The annealing temperature of each of these primers was optimised using gradient PCR. The gel pictures of the PCR products amplified by the primers is presented in plate 3.1.

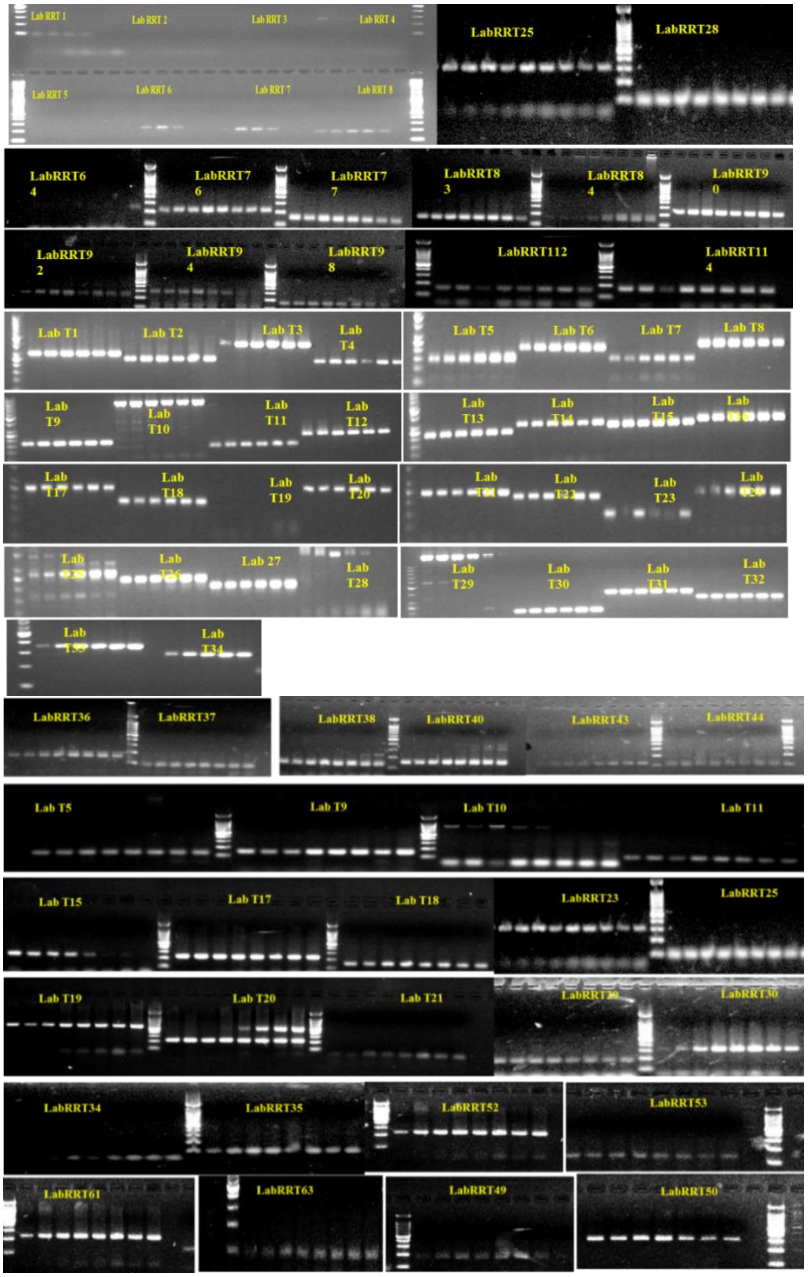


Plate 3.1: Agarose gel picture showing PCR amplification of various lablab SSR markers at different amplification temperatures. The marker loci is indicated in yellow, the 1 Kb ladder is indicated by a Column of bands, the single bands across the gel represents the PCR product amplified for each locus at 8 temperatures ranging between 50-65˚C (50˚C, 51˚C, 52.9˚C, 55.7˚C 59.1˚C, 62˚C,63.8˚C 65˚C).

The result showed that 70 (93.3%) primers pairs amplified lablab genome (Appendix. 3.3). Primers LabRRT 64, LabRRT 34, LabRRT 5, LabRRT 2 and LabRRT 3 did not amplify lablab genome. Most of the primers were able to amplify DNA products at more than one annealing temperature. However, 28 (40%) of these primers pairs produced secondary products in at least one of the annealing temperatures tested. Notably, the frequency of secondary PCR products was lower among the primers designed from assembled isotigs (Lab T primers) but higher in those designed from raw reads (LabRRT primers). The intensity of the amplified products varied from faint to strong bands. Among the primers with successful amplification, 63 (90%) produced strong bands while the rest produced faint but visible bands. It was observed that the most of the primer pairs (68) produced amplification products of the expected size. However, two primer pair's (LabT10 and LabT19) generated PCR products larger than expected. In this study, we suggested the optimum annealing temperatures (TA) of the 70 SSR primers based on the strength of the band produced and the absence/presence of the secondary products. The optimum TA of primers ranged from 51.9^0C to 62^0C but majority were ranging between 55^0C and 59^0C. Majority of the primers had their TA 3^0C to 7^0C lower than their primer melting temperature (TM).

3.4.7 Effectiveness of SSR markers in detecting allele availability and polymorphism in lablab

In this study, 18 SSR primers were used to genotype 10 lablab accessions. The results indicated that 16 out of the 18 lablab primer pairs were polymorphic (Table 3.6), among the 10 lablab accessions. In total, 42 alleles were revealed by the 16 polymorphic SSR loci. The number of alleles per locus ranged from two to four with an average of 2.62 alleles. The highest number of allele's amplified products was observed in LabRRT90. The PIC values of the polymorphic primers were calculated, based on the allelic

variation in the 10 lablab accessions. The PIC values across 16 loci ranged from 0.16 (LabRRT 76) to 0.54 (LabRRT 90). The average PIC was moderate at 0.34. Moderate gene diversity of > 0.4 was detected at 10 (62.5%) SSR loci. The highest gene diversity was detected at LabRRT 90 (0.58) while the lowest was at LabRRT 38 (0.20).

Table 3. 6: Characteristics of the 18 Lablab SSR markers (isolated from unassembled raw data) indicating major allele frequency, number of alleles, expected heterozygosity and polymorphism information content (PIC).

Marker	**Major. Allele. Frequency**	**Genotype No (Ng)**	**Sample Size**	**No. of obs**	**Allele No**	**Availability**	**Gene Diversity**	**Heterozygosity**	**PIC**
LabRRT 36	0.83	3	10	9	3	0.90	0.29	0.22	0.27
LabRRT 37	0.78	2	10	9	2	0.90	0.35	0.00	0.29
LabRRT 38	0.89	2	10	9	2	0.90	0.20	0.00	0.18
LabRRT 40	0.80	2	10	10	2	1.00	0.32	0.00	0.27
LabRRT 44	0.75	3	10	8	3	0.80	0.41	0.00	0.37
LabRRT 49	0.63	3	10	8	3	0.80	0.53	0.00	0.47
LabRRT 50	0.71	3	10	7	3	0.70	0.45	0.00	0.41
LabRRT 53	0.70	2	10	10	2	1.00	0.42	0.00	0.33
LabRRT 61	0.70	3	10	10	3	1.00	0.46	0.00	0.41
LabRRT 63	0.50	2	10	4	2	0.40	0.50	0.00	0.38
LabRRT 76	0.90	2	10	10	2	1.00	0.18	0.00	0.16
LabRRT 72	0.80	2	10	10	2	1.00	0.32	0.00	0.27
LabRRT 77	0.56	2	10	9	2	0.90	0.49	0.00	0.37
LabRRT 83	0.70	3	10	10	3	1.00	0.46	0.00	0.41
LabRRT 90	0.60	4	10	10	4	1.00	0.58	0.00	0.54
LabRRT 98	0.67	2	10	9	2	0.90	0.44	0.00	0.35
LabRRT 112	1.00	1	10	10	1	1.00	0.00	0.00	0.00
LabRRT 103	1.00	1	10	10	1	1.00	0.00	0.00	0.00
Mean	**0.75**	**2.3**	**10.0**	**9.0**	**2.3**	**0.9**	**0.36**	**0.01**	**0.34**

3.5 Discussion

In this study, Roche 454 GS FLX titanium was used to develop a collection of expressed sequence reads from lablab leaves and shoots and mined for genic-SSR markers. We generated 144,614 high quality cDNA sequence reads which were assembled to 3140 isotigs of lablab. Our analysis identified 446 genic SSRs from the isotigs and further

1908 genic SSRs from the raw reads. The approximately 2.1 x 10^6 base pairs of data and genic SSRs produced here represent a substantial resource for lablab and will contribute to genetic diversity studies, linkage mapping, QTL mapping and marker assisted selection in lablab.

The 454 transcriptome pyrosequencing involves construction of cDNA library, attachement of the library to beads via adapter sequences and amplification in water-in-oil emulsion PCR (emPCR). The cDNA-coated beads are then washed over a picoliter reaction plate wells where pyrosequencing then occurs. The wells containing suitably clonally-coated beads will produce reads of varying base pair lengths (bp) (Heather & Chain, 2016). In the present study, a total of 144,614 sequence reads were produced and assembled into 3140 unigenes. The average length of the unigenes was 700 bp while 61.8% of these unigenes being > 500 bp in length. The unigenes length reported here compares well with those reported in legumes such as field pea (719 bp), faba bean (615 bp) (Luke *et al*., 2012), sesame (629 bp) (Wei *et al*., 2011) and winged bean (798 bp) (Wong *et al*., 2017). However, the unigenes length was shorter than those reported in common bean (813 bp) (Hiz *et al*., 2014), mung bean (874 bp) (Chen *et al*., 2015), rice bean (986 bp) (Chen *et al*., 2016), lablab (999 bp) (Chapman, 2015) but was longer than that of black gram (443 bp) (Souframanien & Reddy, 2015). This difference may reflect a variation of species and of the assembly algorithm used (Chen *et al*., 2017). About 80.4% of the isotigs in this study showed significant hits in the NCBI non rendundant protein databases (Figure 3.2) and 100% (34) of the SSR markers developed from these isotigs had PCR amplification of lablab genome. This suggest that assembled contigs were of high quality. In this study, the percentage of contigs with positive hits in the NCBI nr database was higher that those reported for instance in mung bean (41%) and black gram (73.9%) (Souframanien & Reddy, 2015). This could

be attributed by the higher number (61.8%) of contigs with long sequences > 500 bp compared to 24.1% in black gram and 52.6% in mung bean. In sequence annotation, the length of a querry sequence is an important factor which determine whether the sequence simmilarity will be significant or not (Souframanien & Reddy, 2015). Therefore the longer isotigs obtained in this study could have contributed to more BLAST matches in the protein databases. BLASTX analysis revealed that the largest number of matches were to genus Vigna (*V. angulasis* & *V. radiata*) 45.2% followed by *Phaseolus vulgaris* 22% then genus Glycine (G. *max* and *G. soja) 11.8%* and other plant species. This result is consistent with known phylogenetic relationships, as genus lablab is classified together with genus vigna and phaseolus in subtribe phaseolinae while genus glycine is in a different subtribe glycininae (Pasquet & Padulosi., 2013).

This study produced only 3140 unigenes of 2.1 x 10^6 bp size which represents a small fraction of lablab transcriptome. Here, we carried out small scale run of 454 pyrosequencing equivalent to ¼ Pico Titer plate. This could have contributed to the small number of sequenced reads reported because in ¼ Picotiter plate run, smaller number cDNA-coated beads are initially captured than in a full Picotiter plate. This may indicate that ¼ plate pyrosequencing of leaf and shoot transctritome of lablab is not enough to obtain the desired large number of genic SSRs. Higher number of contigs have been obtained in other plant species while using 454 pyrosequencing. For instance, (Luke *et al*., 2012) obtained 6370 and 13602 contigs of faba bean and field pea respectively from ¼ Picotiter plate run, Grahnen *et al*., (2010) reported 57,086 contigs of lodgepole pine (*Pinus contorta*) from ½ plate run while (Moe *et al*., 2011) identified 98,716 contigs of mung bean from a full plate run.

In this study, a total of 446 genic SSR markers were identified from 396 SSR-containing unigene sequences. Previous studies in lablab had identified a total of 2567 SSRs from

15,740 unigene sequences (Chapman, 2015). Higher number of genic SSRs have been identified in many legume species, including 14, 637 in medicago (Gupta & Prasad, 2009), 1840 in black bean (Souframanien & Reddy, 2015), 13134 in mung bean (Chen *et al*., 2015), 5560 in cowpeas (Chen *et al*., 2017b) and 3011 in rice bean (Chen *et al*., 2016). The small number of SSRs identified here could be attributed to the limited number of unigenes assessed. In terms of overall SSR density, this study observed 202 SSR/Mbp (446 SSR/ 2,198,045 bp) or 0.20 SSR in every one Kbp. The results compares well with reports in other legumes such as soya bean (0.29 SSR/Kb), medicago (0.29 SSR/Kb) and Lotus (0.30 SSR/Kb) (Jayashree *et al.*, 2006). This indicates that like in other legume crops, SSRs are equally abundant in the coding region of the lablab genome and can be identified and developed into molecular markers for use in breeding process.

This study observed predominance of TA motifs among di-nucleotides in both unigenes (68%) and raw reads (73%). This observation was similar to that of (Blair *et al*., 2014) in common bean and (Song *et al*., 2010) in soya bean. However it was different from other reports in legume species like lablab (Chapman, 2015), mucuna (Sathyanarayana *et al*., 2017), winged bean (Wong *et al*., 2017), and cowpeas (Chen *et al*., 2017) who observed the predominance of AG/CT motif among the di-nucleotide repeats. The abundance of the TA motifs in this study could be as a result of high frequency of certain amino acids present in the leaves and shoots of tissues which were sequenced. No CG/GC dinucleotide repeat motif were observed in this study which agrees with reports by Souframanien & Reddy, (2015) and Wong *et al*., (2017) who observed that GC motif repeats were very rare in dicots. In this study, tri-nucleotide (51.3%) and dinucleotide (46%) were the most abundant repeats. The dominance of trinucleotide SSRs over other repeats in coding regions may be due to the suppression of non-trimeric

SSRs in coding regions to avoid the risk of frameshift mutations. The non-trifold SSRs in coding regions can cause frameshifts thereby resulting to inactivation of gene expression or coding for shorter protein sequence (Qi *et al.,* 2018; Kalia *et al*., 2011). The relative abundance of di- and tri-nucleotide repeats in ESTs sequences has also been observed in many legumes including lablab (Chapman, 2015), winged bean (Wong *et al*., 2017), mung bean (Moe *et al*., 2011) and mucuna (Sathyanarayana *et al*., 2017).

The tri-nucleotide AAG/CTT was the most common followed by ATC/ATG. This is similar to many previous reports in legumes such as winged bean (Wong *et al*., 2017), cow peas (Chen *et al*., 2017) and common beans (Blair *et al*., 2014) further suggesting a shared origin among the Phaseoleae tribe.

Primer pairs were designed for all the SSR motifs detected. Among the 446 SSR motifs, only small percentage (7.6%) of the motifs were successfully used for primer design. In the process of molecular marker development, not all SSR motifs can be designed as primer pairs mainly due to lack of sufficient long flanking sequences for primer design (Moe *et al*., 2012). Chapman, (2015), also reported that large portion of the lablab SSRs (43%) were located too close to either end of the sequences and therefore not suitable for primer design. Our results suggest that to develop a large number of genic SSRs, there is need to equally sequence a large number of cDNA fragments.

During the SSR marker development not all the identified sequences with repeats become informative markers. Only the SSR loci which can be amplified in lablab genome and which reveal different alleles are useful for molecular breeding. In this study most (93.3%) of the designed primer pairs successfully amplified lablab genome. This high success rate could be attributed to the use of high quality sequence information for primer development and absence of large in-dels in the sample genome

DNA (Chen *et al*., 2016). Success rate of 14–90% amplification for both genomic and EST-SSRs has been reported in different studies (Chen *et al*., 2016; Choudhary *et al*., 2009; Chen *et al*., 2017; Blair *et al*., 2014; Wei *et al*., 2011 & Ma *et al*., 2019).

In this study, majority of the primers (60%) did not produce secondary products indicating that primers were carefully designed. This was achieved by selecting only those primers with the lowest self-3'-complementarity score at the primer design stage using Primer 3 software. Primer secondary products are produced due to intermolecular interactions between the two (forward and reverse) primers to form intermolecular dimers instead of hybridizing to the target DNA. Presence of primer dimers can cause reduction of PCR product yield.

To determine the polymorphism levels of the developed genic SSR markers, we evaluated 18 genic SSR loci where 16 (88%) primer pairs were polymorphic. This level of polymorphism reported here compares well with those from previous reports in different plant species, including lablab (68.75%), (Zhang *et al*., 2013), lablab (94.55%) (Keerthi *et al*., 2018), sesame (88%) (Wei *et al*., 2011), common beans (77.5%) (Hanai *et al*., 2007), black gram (58.2%) (Souframanien & Reddy, 2015), rice bean (53.5%) (Chen et al., 2016) and *Actinidia chinensis* (93.5%) as reviewed in (Varshney *et al*., 2005).

In estimating Polymorphic information content (PIC), we need to take into account both the number of alleles expressed and the relative frequency of those alleles. The PIC provides a better estimator of the informativeness of the SSR loci. In this study, the PIC values were moderate at an average of 0.34. The PIC values in this study compares well with that reported in another study in lablab (0.28) (Zhang *et al*., 2013), but lower that the one reported on lablab (0.34-0.68) by (Keerthi *et al*., 2018). Polymorphic information content values range from zero to one and the higher the PIC value, the

more informative is the SSR marker. In this study primers LabRRT90 (0.54), LabRRT49 (0.47), LabRRT50 (0.41), LabRRT61 (0.41) and LabRRT83 (0.41) were found to have moderate PIC values. This indicate that some of the SSR loci identified in this study had several and relatively frequent alleles and therefore might be suitable for germplasm characterization, variety identification and for selection of materials for breeding purposes.

In summary, 3140 high-quality unigene sequences were assembled and some 446 genic SSRs were identified. The overall SSR density in the coding region of lablab was 0.20 SSR in every one Kbp. About 80.4% of the isotigs in this study showed significant hits in the NCBI non rendundant protein databases. A total of 145 SSR primer pairs were designed. About 93.3% of the genic SSR markers were amplifiable on lablab genome. Sixteen of the 18 markers (88.8%) tested showed a moderate average polymorphism (PIC=0.34) among 10 lablab accessions.

CHAPTER FOUR

Genetic Diversity and Population Structure of Local and Exotic Lablab Accessions in Kenya as Revealed by Microsatellites

4.1 Abstract

A comprehensive understanding of genetic diversity of any crop is important for its prudent exploitation in crop improvement. Molecular markers like simple sequence repeats are effective tools used in discrimination of plant genotypes. Kenya has a wide collection of lablab accessions preserved at gene bank of Kenya and by the lablab growers in their field. Previous effort to characterize the germplasm had limited coverage of the available local germplasm, did not include exotic materials and wild accessions. The study was therefore undertaken to determine genetic diversity and population structure of local and exotic lablab accessions and constitute a core collection for the local breeding program. Characterization was carried out at a molecular laboratory in KALRO's Biotechnology Research Centre, Nairobi. Eight SSR primer pairs were used to genotype the 189 lablab accessions originating from eight populations. A total of 39 alleles were revealed by the eight SSR loci with an average of 4.88 alleles per polymorphic loci. The average PIC of the eight loci was 0.419. The gene diversity among the accessions ranged from 0.257 to 0.518 with an average of 0.38.The highest gene diversity was recorded on Ethiopia (He=0.518) and South Africa (He=0.508) populations. The average Neis unbiased genetic distance was highest between Uganda population and the other populations (0.19 -0.317). A cluster analysis based on unweighted neighbour joining method classified the 189 accessions into four groups, while the Bayesian model-based Structure method clustered the accessions into three groups. In both cluster methods, accessions from Kenya fell in all the groups suggesting that local germplasm is of wide diversity and is a significant representation of diversity available globally. The wild accessions clustered with those from Southern and Eastern Africa confirming that lablab is of African origin. A core collection consisting of 45 accessions mainly from Africa and India and belonging to all the three sub species of lablab was suggested. The core represent a set of diverse materials of priority for future crop breeding attempts.

4.2 Introduction

The existence of genetic diversity within a crop species or population is a major resource for improvement of performance of plant traits. The genetic variability of of genotypes within crop species is important because it is the reservoir of genes of many unique traits. With genetic diversity, breeders can select new varieties directly or identify parents for hybridization to obtain plant recombinants of superior performance (Bhanu *et al*., 2017). With the current changing climatic conditions, genetic diversity is essential for adaptability of crops to impacts of climate change such as pest and disease (Govindaraj *et al*., 2015). A thorough understanding of genetic diversity of any crop is therefore important in order to exploit them prudently.

Various techniques have been used in discrimination of plant genotypes. This include the use of morphological markers, biochemical evaluation and DNA marker analysis (Arunga *et al*., 2015; Bhanu *et al*., 2017; Govindaraj *et al*., 2015). In morphological characterization, different genotypes are grown in the field and the discrimination of the entries is done based on the visually observable traits. Morphological markers are direct, easy to score and less expensive than other markers. However, they are limited in number, dependent on the growth stage and often masked by factors in the environment hence limiting their application in genetic diversity assessment (Nadeem *et al*., 2018; Cholastova *et al*., 2012). The advent of molecular markers more than three decades ago provided an alternative and more reliable approach of discriminating even closely related genotypes. Molecular markers are heritable DNA sequences found at specific locations within the genome and have the advantage of being independent of environmental influence (Li *et al*., 2002). Diversity among individuals in a population is caused by the differences in DNA sequences and the environmental effect. However,

it's only the variation at the DNA sequences that is transferrable from the parents to their off springs.

Various types of molecular markers have been used in assessing the level of genetic variations of plant populations. These include Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP), Simple Sequence Repeats (SSR), Sequence Characterized Amplified Region (SCAR), Sequence Tagged Site (STS) and Single Nucleotide Polymorphism (SNP) (Idrees & Irshad, 2015). Previously, few studies have reported on genetic diversity in lablab using RAPD (Liu, 1996) and AFLP (Venkatesha *et al*., 2007; Kimani *et al*., 2012).

Simple sequence repeats (SSR) or microsatellites are short tandem repeats motifs usually with varying number of repeats at a given locus. This variation in the number of repeats is as a result of the high mutation rate at SSR loci caused by DNA slippage during the DNA replication process (Li *et al*., 2002). Recently, SSRs have been used widely in genetic analysis of crops due to their wide distribution in genome, they are single locus and can reveal multiple allelles, are co-dominant and are highly reproducible (Jun *et al*., 2011; Arunga *et al*., 2015; Ali *et al*., 2015; Lucia *et al*., 2016). Few studies have been reported on genetic diversity analysis in lablab using SSRs (Wang *et al*., 2007; Shivachi *et al*., 2012; Robotham & Chapman, 2015, Kamotho *et al*., 2016). Wang *et al*., (2007) used SSR markers developed for Soybean (*Glycine max*), barrel medic (*Medicago truncatula* Gaertn and cowpea (*Vigna unguiculata* (L) Walp.) while Shivachi *et al*., (2012) employed SSRs developed from common beans (*Phaseolus vulgaris*). Only Robotham and Chapman, (2015) and Kamotho *et al*., (2016) who have employed SSRs designed from lablab transcriptome in genetic diversity analysis of this leguminous crop. The low number of SSR markers available for this crop has contributed to their limited application in molecular studies. To address

this challenge of limited number of SSRs, efforts have been made to design more SSRs and has resulted in identification of tens of new SSRs as reported in chapter three above. There is need to utilize the new SSR loci to discriminate the lablab germplasm in Kenya in order to get more information.

Kenya has a wide collection of more than 330 lablab accessions preserved at gene bank of Kenya (GBK) (Kinyua and Kiplagat, 2012). Few of these collections have been characterized morpho-agronomically or with molecular markers. For instance, Kimani *et al*. (2012) evaluated 55 lablab accessions using AFLP, Kamotho *et al*, (2016) 96 accessions using both morpho-agronomy and SSRs, while Shivachi *et al*., (2012) assessed 13 accessions using SSRs. Whereas the studies provided valuable information for our germplasm conservation and plant breeding programs, they only had limited coverage of the available local germplasm in Kenya. In addition, the studies reported existence of low genetic diversity among the Kenyan lablab germplasm assessed and recommended introduction of exotic germplasm from other countries to diversify the genetic base of Kenyan accessions (Kimani *et al*., 2012). Therefore, further investigations of lablab accessions representing a wider coverage of the local landraces is essential. This is because landraces/local cultivars constitute of a genepool of unexplored alleles which if introgressed into breeding programs could contribute to the broadening of the genetic base and to the development of improved varieties (Kyratzis *et al*., 2019). In addition, assessment of other exotic materials such as those at International Livestock Research Institute (ILRI) would establish the extent at which the local lablab breeding program will benefit from the genetic resources from other countries and regions.

4.2.1 Objectives

The objectives of this study are:

1. To determine the diversity and population structure of worldwide cultivated and wild lablab collection;
2. To identify a core collection appropriate for future lablab breeding program in Kenya

4.3 Materials and methods

4.3.1 Plant materials

The materials used included accessions previously collected from farmers' fields and markets in various lablab growing regions of Kenya (40 accessions), some advanced determinate early maturing lines from Kenya Agriculture and Livestock Research Organization (KALRO) lablab breeding program (8 lines), released variety (1 variety) and accessions preserved at the GBK (43 accessions). The advanced breeding lines were pedigrees of crosses between accession KDD, Njoro and GBK 028663. KDD is an accession collected from a local market in Nairobi, has white flowers, determinate growth habit, early maturing and with cream seed coat colour. Njoro and GBK 028663 have indeterminate growth habit, purple flowers, medium maturing and with black seed coat. Also included was a collection from the International Livestock Research Institute Forage Germplasm (ILRI-FG) which comprised of cultivars from other parts of Africa, USA, Asia, Australia and accessions of unknown origin (Figure 4.1). The ILRI-FG was established and has been maintained since 1982. In total, 189 accessions were evaluated in this study (Appendix II).

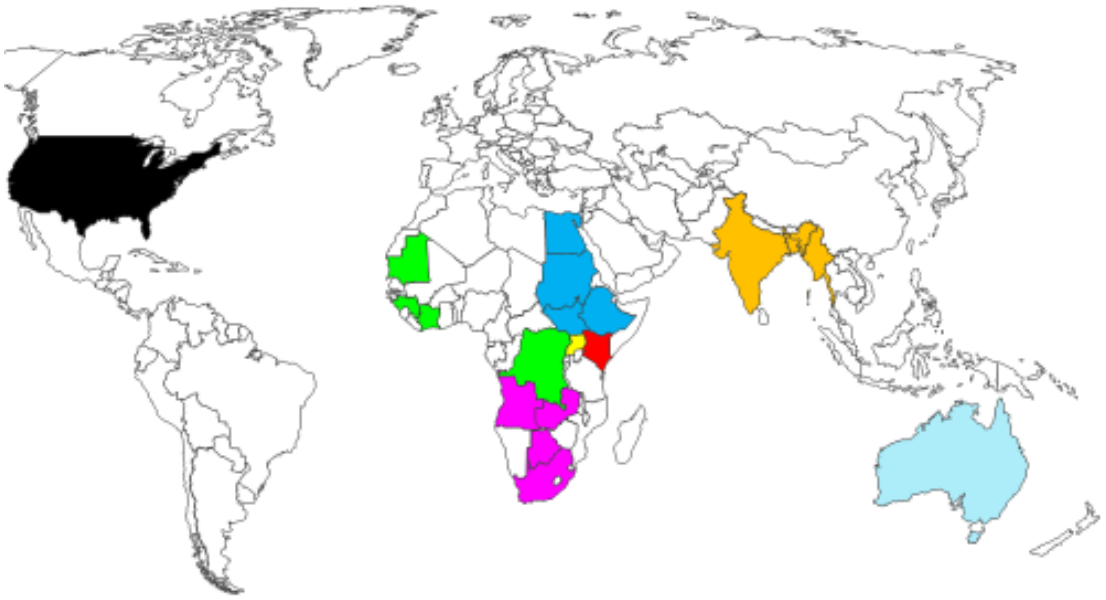


Figure 4. 1: Map showing regions from where the lablab accessions used in the study were collected. Red: Kenya, Yellow: Uganda, Purple: Southern Africa, Green: Western Africa, Sky blue: Australia, Dark blue: Ethiopia region Orange: India, Black: USA.

4.3.2 DNA isolation

The activity was undertaken at Kenya Agricultural and Livestock Research Organization (KALRO) Biotechnology Research Centre, Nairobi. Seeds of the 189 accessions were planted in a greenhouse. At two leaves stage, approximately one gramme of leaf tissue of each genotype was harvested, placed on labelled zip lock papers and immediately placed on ice box. DNA was isolated using modified cetriamonium bromide method (CTAB) (Doyle 1987). Approximately 0.3g of the leaf tissues were placed in self-standing tubes each containing a ceramic bead. Nine hundred micro-liters (900µl) of extraction buffer was added to the leaf tissue containing the ceramic bead. The samples were crushed by geno-grinder machine (Benchtop homogenizer) set at revolution speed of 4 meter per second for one minute. The samples were then incubated in a water bath with constant shaking at 65^0C for 15 minutes. The

tubes were then centrifuged at 14000 rpm for five minutes. Six hundred microliters (600µl) of the supernatant were transferred into a fresh eppendorf tube and an equal volume of chloroform: isoamyl alcohol (24:1) added. The above step was repeated. Four hundred microliters (400µl) of the aqueous phase was transferred to a fresh tube and an equal volume of ice-cold isopropanol added and mixed by inverting several times and incubated at -20^{o}C for 2 hours to precipitate the DNA. The tubes were centrifuged at 14000rpm for 5 minutes. The supernatant was decanted leaving the DNA pellet at the bottom of the tube. The pellets were washed using 500µl of 70% ethanol and centrifuged for one minute at 14000rpm before they were air dried for one hour. The dried pellets were re-suspended in 50µl of sterile distilled water. RNA was removed by adding two microliters (2µl) of pancreatic ribonuclease A (RNase A) (10mg/ml) and incubating the samples for 30 minutes at 37^{o}C. The samples were stored at -20^{o}C.

4.3.3 DNA quantification and quality determination

The extracted genomic DNA was quantified by comparing with a DNA ladder (gene ruler) of one kilo base (1 kb) in a 1% agarose gel using 1x TBE buffer. The bands were then viewed in a gel box (G: Box, Syngene). The concentration of the sample DNA was further determined at optical density (OD) readings of 260 nm and 280 nm using a Nanodrop spectrophotometer (Thermo Scientific NanoDrop 2000C). The samples were then diluted with molecular water to a final concentration of 20 ng/ μL for use in the PCR.

4.3.4 Polymerase chain reaction and Polyacrylamide gel electrophoresis

In this study, eight SSR markers were used to discriminate the 189 lablab accessions. The eight markers were part of the new Lablab specific SSR markers reported in chapter three. These markers were selected based on their ability to reveal many clear

polymorphic bands on a panel of 19 diverse lablab accessions. The eight markers are listed in Table 4.1 below.

Table 4. 1: List of 8 lablab specific SSR markers used to evaluate 189 accessions

Primer designed name	Forward primer	Reverse primer	type	motif	Exp. prod size (bp)	Observed product size (bp)	Primer melting temp. (Tm) in ^{0}C
Lab T2	GTGCGCGTCACTTATTAGTTCTTA	CAATATCTTCACGTAACCACGGTA	p6	(TATATC)7	224	180-240	59.3- 61.0
Lab RRT23	GGGAGTGTGAAATAGAGAATCAGTT	CAGCACTATCCACACCTGCAATAC	p2	(AG)12	136	120-150	59- 62.7
Lab RRT28	AATCGAACAAAGTGAAGTGCCTTG	AAATAGCCTCCAACTTCTCCCACT	p2	(AG)8	96	90-110	59.3- 61.0
Lab RRT44	AAGCTTCGTTGTTTCTGCGATTAG	CGAGCTTTAAACCAATCAGGACAC	p3	(ATT)6	92	90-110	59.3- 61.0
Lab RRT90	ATAACTCTGGCTCGCTCTGTGG	GTGCATTTGATTTGGTGGGAAA	p4	(GAGT)5	234	220-250	56.5– 62.1
Lab RRT53	ACACCACATCACACACTTATTC	CTTGCTGACTGTTCTCCATT	p3	(CAT)7	176	140-180	58.4– 59.0
Lab RRT77	CTTTCTCCTTCTCTTTCTCACTC	GAAGACGGGTAGTTCCTAGTTAT	p3	(CTT)9	181	160-190	59.3– 61.3
Lab T12	CACCACCTCCAACTTCTACGGTTA	TGACCTCCATTATGGGATTCAGAT	p2	(TA)11	193	180-220	59.3– 62.7

The following were used in the 25ul PCR; Lyophilised master mix bead (easy to use beads) containing (dNTPs, $MgCl_2$, Tris-HCl (pH 9.0), KCl and Taq), 19µL of sterile distilled water, 2µL of 2 pmoles/ul forward primer, 2µL of 2 pmoles/ul reverse primer and 2µL of template of 20ng/ul DNA. Amplification was carried out in a Thermocyler machine (Techne-TC 412, Applied Biosystems Veriti systems) programmed with the following regime and 35 cycles: 94˚C for 3 minutes, 35 cycles of 94˚C for 1 minute, annealing temperature of 57ºC for 1 minute, 72˚C for 1 minute and final extension at 72˚C for 10 minutes and final hold at 4^{0}C.

4.3.5 Data collection

The PCR products were size separated using 6% PolyAcrylamide gel (PAGE). The PCR products were size separated using 6% PolyAcrylamide gel (PAGE). The 6% gel was prepared as follows; 53.8 ml of distilled water was put in a clean 500 glass beaker, 18.8 ml of 40% acrylamide/ bis-acrylamide 19:1 was added to the beaker, 2.5ml of 50x TAE (Tris acetate EDTA) was then added and mixed by stirring, one tablet containing ammonium persulfate (APS) and TEMED was dissolved in 50ml distilled water in a 50 ml falcon tube, the APS/TEMED solution was then added to the glass beaker and gently stirred without introducing any air bubble, the solution was immediately poured into the gel tray with in-built combs and was covered with glass plate treated with silane solution and allowed to gel for 1 hour. The glass plate with the attached PAGE was then transferred to the horizontal PAGE (hPAGE) (Cleaver Scientific) electrophoresis tank containing 1X TAE solution. Two micro liters (2uL) of samples were loaded into the wells alongside 3.0µL 1kb DNA ladder and run at 140V for 3hrs. The glass plate was removed from the gel tank and placed in another glass tray for post staining of the gel. To stain the PAGE, 6uL of ethidium bromide was dissolved in 200 ml of water and poured into the glass tray to cover the glass plate containing the PAGE. The tray was kept in a dark room for 1 hour. The staining solution was then drained off and the gel de-stained using clean water for 30-45 minutes.

The amplified products were then viewed under UV light in a gel box. The images were captured using a camera connected to the gel box and saved in the computer (Appendix 4.2). To score the bands, the images were pasted in MS power point program. On the gel image, a straight line was drawn across the bands that had migrated the longest distance. The accessions that had bands at that position were scored as 1

while those that didn't have were scored 0. The straight line was moved to another allele and scoring was done. The process continued until all the bands were scored.

4.3.6 Data analysis

The number of alleles per locus, major allele frequency per marker, gene diversity, heterozygosity and the polymorphic information content (PIC) for each locus was computed using PowerMarker version 3.25. GenAlEx version 6.501 (Peakall and Smouse, 2012) was used to estimate, the number of different alleles per locus, number of effective alleles per locus, expected heterozygosity, Shannons Information Index, percentage of polymorphic loci and analysis of molecular variance (AMOVA). Dissimilarity matrix was generated using simple matching coefficient of DARwin software version 6.0.17. The dissimilarity matrix was used to construct Principal Coordinate Analysis (PCoA) scatter plot. A core collection representing a subsample of the 189 accessions was created using the maximum length sub tree function of DARwin software. A neighbour joining tree threshold value of 0.5% was used to remove the synonymous accessions.

The Bayesian based STRUCTURE software version 2.3.3 was used to investigate population structure of the 189 lablab accessions. The number of clusters (K) in STRUCTURE were set from 2 to 8. Each run consisted of a burn-in period of 5000 steps followed by 50000 Monte Carlo Markov Chain (MCMC) replicates. We assumed an admixture model and uncorrelated allele frequencies. To determine the most likely number of clusters we used delta value (ΔK) method proposed by Elvanno *et al*., (2005) using the STRUCTURE Harvester available online at http://taylor0.biology.ucla.edu/struct_harvest/. To assign accessions to various groups, the accessions with estimated memberships above 0.8 were allocated to the matching group while those with below 0.8 were assigned to a mixed group.

4.4 Results

4.4.1 SSR Polymorphism

A total of 39 alleles were revealed by the eight SSR loci with an average of 4.88 alleles per polymorphic loci (Table 4.2). The number of alleles per locus ranged from 3 for locus Lab RRT77 to 7 for Locus Lab T12. The PIC ranged from 0.148 (Lab RRT77) to 0.772 (Lab T12) with an average of 0.49. Half of the markers were highly polymorphic with PIC values of over 0.5. The gene diversity varied among the loci with an average of 0.45 recorded. Most of the markers revealed moderately high gene diversity of more than 0.4. In this study, all the markers had lower observed heterozygosity (Ho) compared to expected heterozygosity (He).

Table 4. 2: Summary of PowerMarker data analysis of 189 lablab accessions with 8 SSR markers

Marker	MAF	SS	No.	AN	Avai.	GD	Obs. Heter	PIC	f
Lab T12	0.2971	189	170	7	0.8995	0.8008	0.0176	0.7722	0.9781
Lab T2	0.5772	189	162	6	0.8571	0.6146	0.0062	0.5783	0.9900
Lab RRT23	0.6872	189	187	4	0.9894	0.4712	0.0588	0.4174	0.8758
Lab RRT44	0.9181	189	177	5	0.9365	0.1550	0.0621	0.1514	0.6009
Lab RRT53	0.6044	189	182	5	0.9630	0.5843	0.0000	0.5461	1.0000
Lab RRT77	0.9144	189	187	3	0.9894	0.1581	0.0107	0.1485	0.9327
Lab RRT28	0.5027	189	186	5	0.9841	0.6360	0.0054	0.5763	0.9916
Lab RRT90	0.9091	189	187	4	0.9894	0.1696	0.0267	0.1630	0.8431
Mean	**0.6763**	**189**	**179**	**4.88**	**0.9511**	**0.4487**	**0.0234**	**0.4191**	**0.9480**

MAF: Major allele frequency; SS: Sample size; No.: Number of observations; AN: Allele number; Avai: Availability; GD: gene diversity; PIC: Polymorphic information content; f: Inbreeding coefficient.

The average observed heterozygosity was low with an average of 0.02. The locus Lab RRT53 did not display any observed heterozygosity. The inbreeding coefficient (***f***) of the SSR loci was high with an average of 0.94. Both Hardy-Weinberg Chi square and exact test for all populations revealed that all the eight loci exhibited significant deviation from HWE as revealed by the p-value (Table 4.3). Therefore all the eight markers were used in diversity analysis of the 189 accessions.

Table 4. 3: The 8 SSR markers used in the diversity analysis of 189 lablab subjected to Chi square and HWE exact test

Marker	X^2 value	X^2 d.f	X^2 p-value	Exact p-value
Lab T12	830.94	21	0.0000	0.0000
Lab T2	794.56	15	0.0000	0.0000
Lab RRT23	359.76	6	0.0000	0.0000
Lab RRT44	423.60	10	0.0000	0.0000
Lab RRT53	728.00	10	0.0000	0.0000
Lab RRT77	232.66	3	0.0000	0.0000
Lab RRT28	730.27	10	0.0000	0.0000
Lab RRT90	330.77	6	0.0000	0.0000

4.4.2 Genetic diversity within and among 189 lablab accessions.

The summary of genetic diversity indices of the nine lablab populations are shown in Table 4.4. The average number of different alleles (Na) were highest for Kenyan population (4.375), South Africa population (3.875) and Ethiopia population (3.750) but lowest for Uganda population at 1.625. The number of effective loci (Ne) ranged from 2.452 to 1.570 with an overall mean of 1.930. Private allelic richness was only detected in Kenya population (0.25), advanced lines (0.125) and South Africa (0.125) lablab populations. The remaining lablab populations did not reveal any private allele in the loci studied. High gene diversity of > 0.5 was recorded only in Ethiopia and South Africa populations. However, moderate gene diversity of 0.3-0.4 was observed for advanced lines, Australia, India, Kenya and West Africa populations of lablab. The average expected heterozygosity was moderate at 0.38. The proportion of polymorphic loci ranged from 50 % (Uganda) to 100% (Kenya and South Africa), with an average of 80.56%. The highest Shannon's Index (I) was observed on Ethiopia (0.964) and South Africa (0.929) populations while the least was in Uganda population (0.661). The observed heterozygosity (Ho) values were very low for all the populations studied indicating the inbreeding nature of the lablab crop. However, the highest observed

heterozygosity of 0.09 was recorded on advanced lines suggesting that these lines could still be segregating at some loci.

Table 4. 4: Mean number of different loci (Na), number of effective loci (Ne) Shannon index (I), Expected heterozygosity (He), Private alleles richness (PAR), and percentage of polymorphic Loci (% poly) among 189 lablab accessions studied.

Pop		Na	Ne	I	Ho	He	F	% poly	PAR
Adv line	Mean	2.875	1.950	0.666	0.096	0.362	0.569	87.5	0.125
	SE	0.549	0.381	0.191	0.021	0.099	0.157	-	0.125
Australia	Mean	2.000	1.731	0.521	0.000	0.328	1.000	62.5	0.000
	SE	0.327	0.241	0.162	0.000	0.100	0.000	-	0.000
Ethiopia	Mean	3.750	2.452	0.964	0.031	0.518	0.914	100	0.000
	SE	0.491	0.392	0.151	0.021	0.071	0.071	-	0.000
India	Mean	2.750	1.880	0.685	0.022	0.396	0.822	87.5	0.000
	SE	0.313	0.245	0.143	0.015	0.084	0.140	-	0.000
Kenya	Mean	4.375	1.940	0.735	0.018	0.363	0.872	100	0.250
	SE	0.324	0.421	0.178	0.008	0.090	0.062	-	0.164
South Afr	Mean	3.875	2.372	0.929	0.015	0.508	0.942	100	0.125
	SE	0.350	0.322	0.139	0.007	0.076	0.032	-	0.125
Uganda	Mean	1.625	1.575	0.383	0.000	0.257	1.000	50	0.000
	SE	0.263	0.255	0.154	0.000	0.100	0.000	-	0.000
USA	Mean	1.750	1.570	0.429	0.083	0.285	0.651	62.5	0.000
	SE	0.250	0.203	0.137	0.055	0.089	0.188	-	0.000
West Afri	Mean	2.250	1.897	0.639	0.036	0.402	0.905	75	0.000
	SE	0.313	0.224	0.147	0.036	0.090	0.082	-	0.000
Total	Mean	2.806	1.930	0.661	0.033	0.380	0.849	80.56	0.056
	SE	0.160	0.103	0.054	0.009	0.030	0.035	6.29	0.046

4.4.3 Population differentiation.

In this study, the 189 lablab accessions were divided into nine subpopulations based on their origin of collection. The populations included Kenya (95), Ethiopia (27), Uganda (3), USA (3), South Africa (27), West Africa (7), India (11), Australia (8) and advanced lines (8). Wrights F statistics and AMOVA were used to partition the genetic diversity to within and among the nine populations. In this study, AMOVA revealed 94% of the

allele diversity was attributed to individuals within populations (94%) while only 6% was distributed among the populations (Table 4.5). A low value of ΦPT of 0.061 was observed confirming that there was only a small differentiation among the populations.

Table 4. 5: Analysis of molecular variance (AMOVA) for 9 populations of Lablab purpureus and partitioning of the total diversity into population components

Source	df	SS	MS	Est. Var.	% molecular variation	P value	PhiPT
Among Pops	8	121.280	15.160	0.477	6	0.01	0.061
Within Pops	180	1313.339	7.296	7.296	94		
Total	188	1434.619		7.773	100		

The F statistics of Wright, (1951) indicated that the genetic differentiation among populations ($F_{ST)}$ ranged from 0.088 to 0.284 with an overall mean of 0.164 (Table 4.6) suggesting low differentiation between the populations studied. Among the SSR loci studied, only Lab RRT53, Lab T2, Lab RRT28 and Lab 12 had F_{ST} values of above 0.18, while the rest had low Fst values of less than 0.16. Majority of the SSR loci had high inbreeding coefficient (F_{IS}) close to one with exception of Lab RRT44 which had F_{IS} of 0.326. The average F_{IS} value was 0.851 suggesting high genetic inbreeding within the populations. Notably, Lab RRT53 had F_{IS} value of 1 indicating that for this locus, no observed heterozygous allele was detected within the subpopulations. The coefficient of gene flow was moderately high at an average of 1.514 indicating substantial movement of alleles into and out of the populations.

Table 4. 6: F-statistics and estimates of differentiation of Lablab purpureus populations for each locus

Locus	Fis	Fit	Fst	Nm
Lab T12	0.969	0.975	0.183	1.118
Lab T2	0.968	0.975	0.197	1.018
Lab RRT23	0.831	0.848	0.098	2.306
Lab RRT44	0.326	0.398	0.107	2.076
Lab RRT53	1.000	1.000	0.284	0.631
Lab RRT77	0.985	0.988	0.159	1.321
Lab RRT28	0.976	0.980	0.194	1.039
Lab RRT90	0.751	0.772	0.088	2.599
Mean	0.851	0.867	0.164	1.514
SE	0.081	0.073	0.023	0.252

The Fst values among pairs of populations were found to range from 0.028 to 0.278 (Table 4.7). Population differentiation was highest between Uganda and USA populations (Fst=0.278).

Table 4. 7: Pairwise unbiased Nei's genetic distance (GD) and population differentiation (Fst) between lablab accessions of origin in Africa Asia Australia and USA

	GD								
	Al.	Aus.	Eth	India	Kenya	SA	Uganda	USA	WA
Al.	-	0.081	0.064	0.094	0.025	0.108	0.317	0.156	0.120
Aus.	0.075	-	0.050	0.042	0.031	0.059	0.243	0.062	0.011
Eth.	0.058	0.064	-	0.110	0.050	0.062	0.209	0.123	0.091
India	0.068	0.051	0.068	-	0.099	0.078	0.319	0.097	0.079

Kenya	0.028	0.034	0.043	0.060	-	0.110	0.246	0.084	0.078
SA	0.076	0.064	0.031	0.054	0.066	-	0.190	0.199	0.119
Uganda	0.214	0.228	0.151	0.217	0.164	0.133	-	0.316	0.315
USA	0.134	0.123	0.111	0.119	0.083	0.134	0.278	-	0.021
WA	0.092	0.057	0.072	0.079	0.067	0.082	0.256	0.095	-
	Fst								

Unbiased Neis genetic distance (GD) in upper diagonals and genetic differentiation (Fst) in the lower diagonals.

The Uganda and USA populations were more highly differentiated from the other populations. The average Neis unbiased genetic distance was highest between Uganda and the other populations (0.19 -0.317). The highest genetic distance between Kenya population and others was recorded with Uganda (0.246) and South Africa (0.110). The least genetic distance was recorded between accessions collected from Kenya and the advanced lines (0.025).

High level of gene flow was observed between Kenya and advanced lines (8.67), Kenya and Australia (7.81), Kenya and neighboring Ethiopia (5.56) and Ethiopia and South Africa (7.81). Some moderate level of gene flow was also observed between Kenya and South Africa (3.54), Kenya and West Africa (3.48), Kenya and India (3.92), Ethiopia and Australia (3.66) and Ethiopia and India (3.43) (Table 4.8). Least gene flow was between Uganda and the other countries.

Table 4. 8: Pairwise gene flow (Nm) between lablab populations of origin in Africa Asia Australia and USA

	Al.	Aus.	Eth	India	Kenya	SA	Uganda	USA	WA
Al.	-								
Aus.	3.08	-							
Eth.	4.06	3.66	-						
India	3.42	4.65	3.43	-					
Kenya	8.67	7.10	5.56	3.92	-				

SA	3.03	3.66	7.81	4.38	3.54	-			
Uganda	0.91	0.85	1.41	0.90	1.27	1.63	-		
USA	1.61	1.78	2.00	1.85	2.76	1.62	0.65	-	
WA	2.46	4.14	3.22	2.91	3.48	2.80	0.73	2.38	-

4.4.4 Genetic associations among accessions

The genetic distances among 189 lablab accessions were subjected to clustering using unweighted neighbour joining method of DARwin software version 6.0.17. The neighbour joining tree (NJ) separated the lablab accessions into four main clusters of (Figure 4.2). The four clusters however did not reflect on the geographic origin of the 189 lablab accessions indicating possibility of gene flow between regions.

The number of accessions allocated to each of the four (I, II, III and IV) NJ tree clusters varied. Cluster I comprised of materials from Africa, India and Australia. Majority of the accessions allocated to this group were from Kenya and Ethiopia. Notably, three early flowering accessions (ILRI 6930, ILRI 14411, ILRI 14437) placed in group 1, were previously clustered together in RB2/Zwai2 group of Pengelly and Maass (2001) core collection. A popular released variety in Kenya, DL1002 and two accessions (Njoro and GBK 028663) used as parents to develop some advanced lines included in the present study were also clustered in group 1. The NJ cluster II consisted of eight accessions originating from African countries (Kenya, South Africa and Ethiopia).

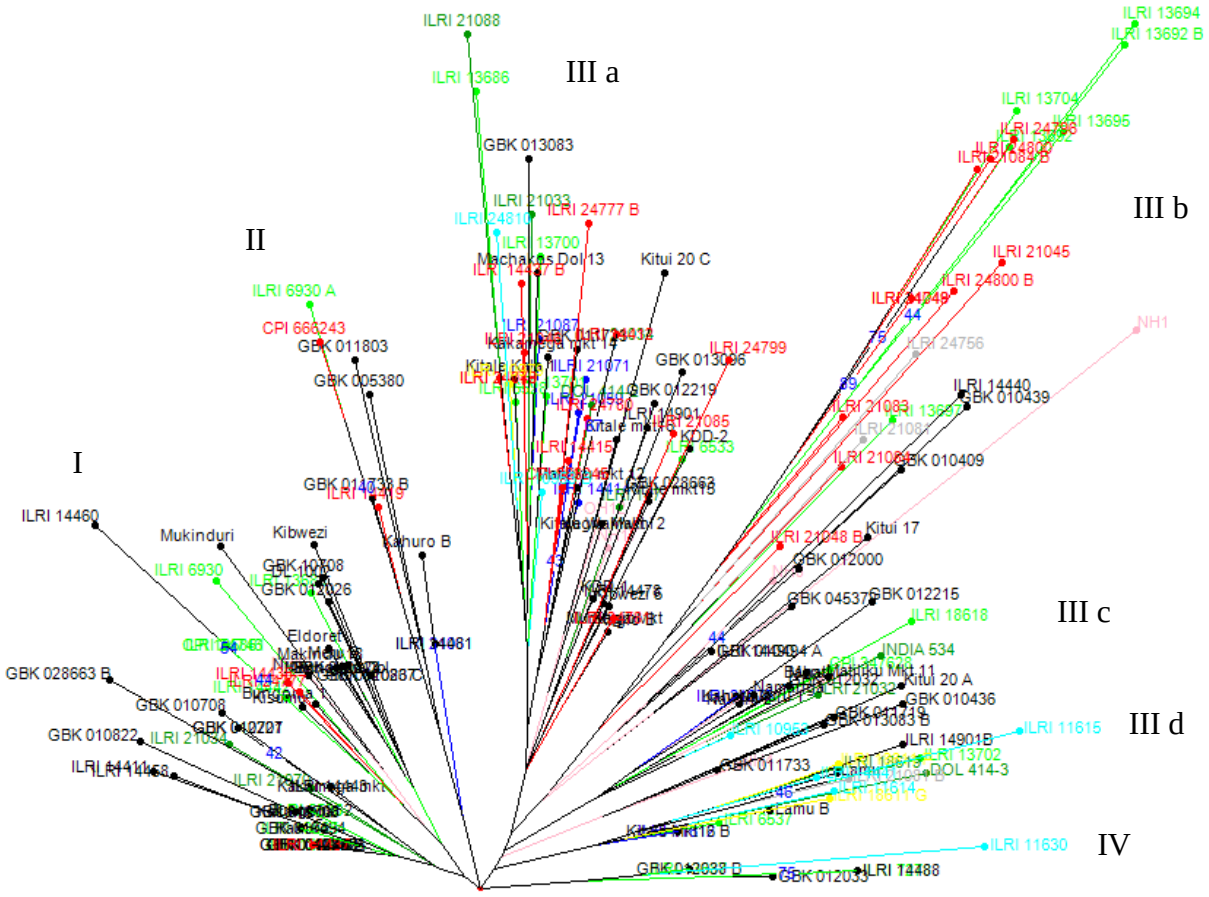


Figure 4. 2: Unweighted Neigbour joining tree of 189 lablab accessions constructed from dissimilarity matrix obtained from genotyping accessions using 8 polymorphic SSR markers. Different colours have been used to differentiate the sources of the accessions. Black (Kenya), dark red (South Africa), Light green (Ethiopia), dark green (India), light blue (West Africa), Blue (Australia), grey (Uganda), light red (advanced lines) and yellow (USA).

The NJ cluster III was allocated the highest number of accessions. The cluster was subdivided into 4 distinct sub clusters IIIa, IIIb, IIIc and IIId. Sub-cluster IIIa consisted of 50 accessions originating from six countries, but majority were from Kenya and South Africa. Sub cluster IIIb was fully dominated by accessions originating from Africa (South Africa, Kenya, Ethiopia and Uganda). Notable, all the accessions belonging to subspecies unicinatus were allocated to sub cluster IIIb with exception of

accession ILRI 24778. Three late maturing accessions (ILRI 13701, ILRI 6533 and ILRI 13700) in Pengelly and Maass (2001) core collection were also included in this group. Cluster IIIc consisted of 18 accessions with majority from Kenya (72%) while the rest were from West Africa, India and Ethiopia. The fourth sub cluster IIId had 11 accessions from all the countries except South Africa. Two large seeded accessions (ILRI 13702 and ILRI 11615) in Pengelly and Maass (2001) core collection were included in this group. Cluster IV was allocated only six accessions of which three originated from Kenya, two from Ethiopia and one from West Africa.

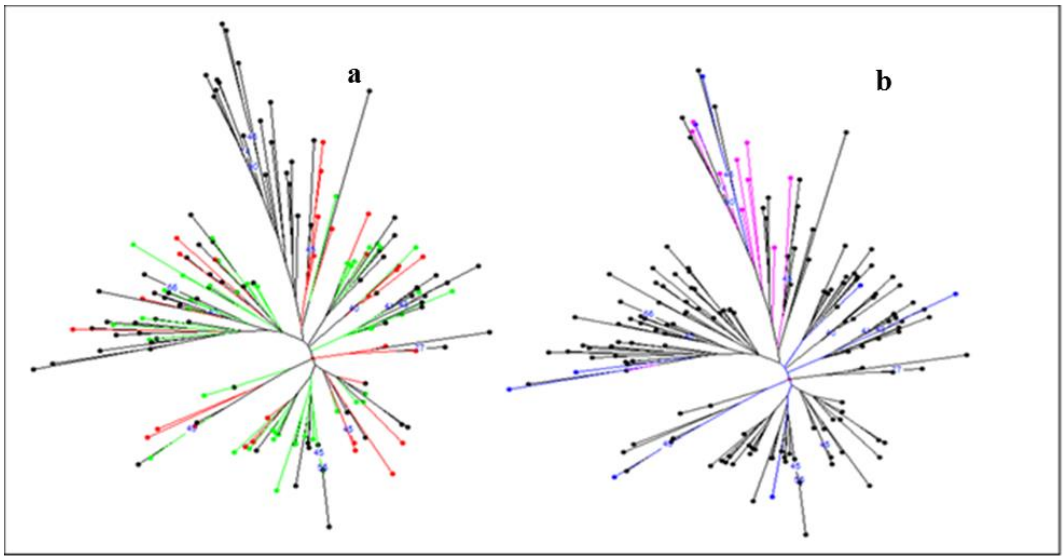


Figure 4. 3: Unweighted Neigbour joining tree showing clustering of 189 lablab accessions. Different colours have been used to differentiate the sources of the accessions and type of subspecies. (a) Red represents 41 Kenya accessions conserved at GBK, blue represents 49 Kenyan accessions collected from farmers' fields and markets; black represents accessions from the rest of the world and conserved at ILRI Ethiopia. (b) Blue represents subspecies bengalensis, purple represents unicinatus while black represents purpureus.

Seven out of the eight advanced early maturing breeding lines (with prefix NH) from Kenya were grouped in cluster III while the other line (NH13) grouped together with the released variety DL1002 in cluster I. The accessions from Kenya were mainly

obtained from farmers' fields/local markets (the accessions with local names) and from gene bank of Kenya (the accessions with GBK prefix). These Kenyan accessions were split across all the four clusters of NJ tree (Figure. 4.3a). Similarly, the accessions obtained from International Livestock Research Institute (ILRI) (the accessions with ILRI prefix) were also divided in all the four clusters. The eight accessions belonging to sub species bengalensis did not clustered together and were split in three of the four clusters of NJ tree (Figure 4.3b).

4.4.5 Population structure analysis using Bayesian-model Structure.

We explored population structure in the 189 *Lablab purpureus* germplasm using the Bayesian model-based Structure method. According to Evano *et al* (2005), the real number of population is detected at the modal value or the upper most value of Delta K.

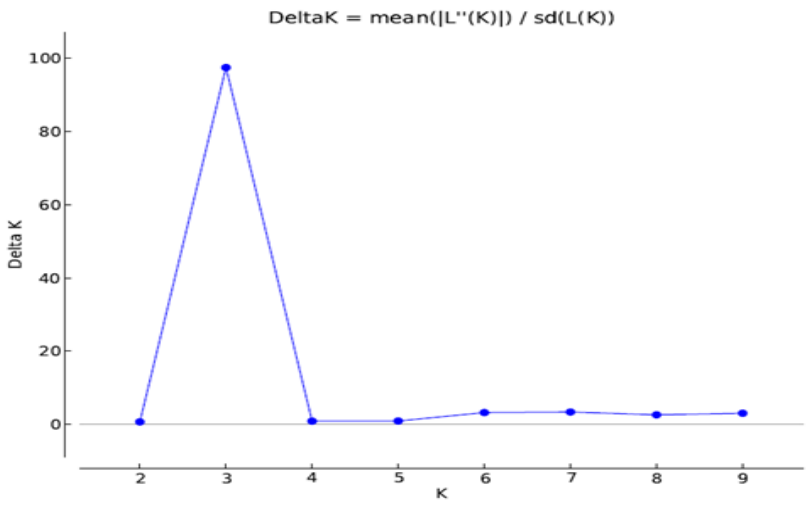


Figure 4. 4: Plot of Delta K against the likely sub populations (K) generated according to Evanno et al. (2005).

In this study, the highest value of delta K was observed at K=3 (Figure 4.4) suggesting that the 189 lablab accessions are assigned to three sub-groups.

The 3 subpopulations detected using STRUCTURE are shown in Figure 4.5. Each lablab accession is represented by a vertical bar. The bars are segmented into different coloured fragments which represented the estimated membership of a certain subpopulation.

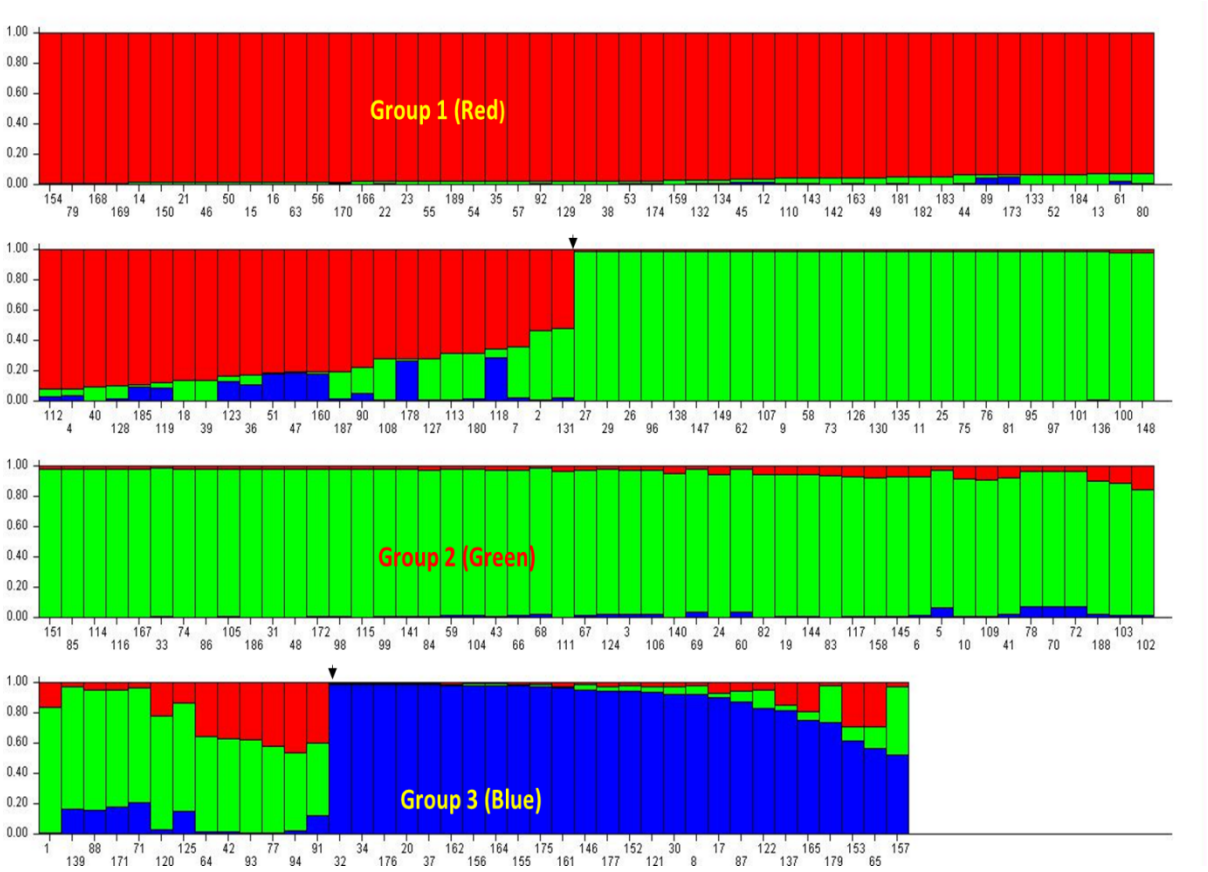


Figure 4. 5: Population structure of 189 accessions of Lablab purpureus as determined by STRUCTURE analysis based on SSR allelic data at 8 loci. The subpopulations 1, 2 & 3 are denoted by red, green and blue colours respectively. The values on the x axis represents the accessions code, y axis value represent the estimated group membership of accessions and the small black arrow separates the groups.

The STRUCTURE cluster 1 comprised of 68 accessions while clusters 2 and 3 had 77 and 22 accessions respectively. The STRUCTURE cluster 1 contained accessions from

Kenya, Ethiopia, SA, Australia, India, USA and West Africa. The accessions from Kenya included those collected from gene bank of Kenya (GBK 12219, 012215, 011723, 011719, 013096), from central Kenya (Kagio, Muranga), coastal Kenya (Lamu, Kibwezi), western Kenya (Kakamega, Kitale) and those preserved at ILRI (ILRI 14901, ILRI 14445). The eleven accessions from SA included in STRUCTURE cluster 1 were mainly from sub species purpureus. Materials from Ethiopia allocated to cluster 1 included ILRI 13686, ILRI 13688, ILRI 13700, ILRI 13701, ILRI 6528, ILRI 6533, ILRI 6537 and ILRI 6930, those from Australia included ILRI 21071, ILRI 21059, ILRI 14414, ILRI 21087 and ILRI 11617. All accessions from India with exception of one bengalensis (ILRI 21032) were also grouped in STRUCTURE cluster 1. Similarly, included in cluster 1 were all the accessions from USA and five accessions from West Africa which consisted of two accessions (ILRI 11615 and ILRI 24810) belonging to sub species bengalensis.

STRUCTURE cluster 2 was assigned accessions from Kenya, Ethiopia, SA, Australia, West Africa and India. The Kenyan materials in this cluster originated from all major lablab growing areas and markets such as central region (Kahuro), Rift valley region (Nakuru, Njoro, Bahati, Namanga, Eldoret), western region (Kisumu and Bungoma) and eastern region (Kitui and Meru). Majority (79%) of the Kenyan accessions from the gene bank of Kenya were included in this cluster. Cluster 2 also included 11 accessions from Ethiopia such as ILRI 6535, ILRI 6536, ILRI 7278, ILRI 13685, ILRI 13688, ILRI 13687, five from South Africa (ILRI 14437, ILRI 14435, ILRI 14419, ILRI 24777, CPI 666243), from Australia (ILRI 21061, ILRI 21076, ILRI 11617), from West Africa (ILRI 11630, ILRI 10953) and India (ILRI 21032).

All the accessions allocated to cluster 3 were of African (South Africa, Ethiopia, Kenya, Uganda) origin. The accessions from South Africa included the two seeded wild accessions (ILRI 21045, ILRI 24800 and ILRI 21083), four seeded wild accessions (ILRI 21048, ILRI 24749) and cultivated sub species purpureus (ILRI 14437, ILRI 21084, ILRI 21085). The accessions from Kenya consisted of a two seeded wild germplasm (ILRI 14440) and five local collection. The accessions included in cluster 3 from Ethiopia consisted of a bengalensis (ILRI 13692) and other three sub species purpureus (ILRI 13694, ILRI 13695, and ILRI 13704). This group also included accession ILRI 21081 and a four seeded wild germplasm (ILRI 24756) from Uganda. A total of 22 accessions showed some mixed ancestry (the membership value of less than 80% of any sub population). The 15 accessions in Pengelly and Maass (2001) core collection included in this study were distributed in all the three STRUCTURE clusters. For instance, accessions ILRI 13700, ILRI 13701, ILRI 6533, ILRI 11615, ILRI 13702, ILRI 6930 and ILRI 20134 were grouped in STRUCTURE cluster 1. Included in cluster 2 were ILRI 11630, ILRI 13687, ILRI 14411 and ILRI 14437. The third cluster consisted of accessions ILRI 14440, ILRI 13692, ILRI 13694 and ILRI 13695. All the accessions from sub species unicinatus with exception of accession ILRI 24778 were grouped in the STRUCTURE cluster 3 which was dominated by accessions from Africa. The accessions belonging to sub species bengalensis were split into all the three STRUCTURE clusters. All the seven advanced lines (AL lines) from a breeding program in Kenya were grouped in clusters 1 and 2 together with their parental lines (Njoro, GBK 028663B, KDD).

PCoA analysis was done using DARwin 6.0.17 software to further understand the population structure of the 189 lablab accessions. The PCoA analysis separated the accessions into three clusters thereby confirming the results obtained with

STRUCTURE. The degree of distribution of the accessions in the PCoA scatter plot varied across the clusters (Figure 4.6).

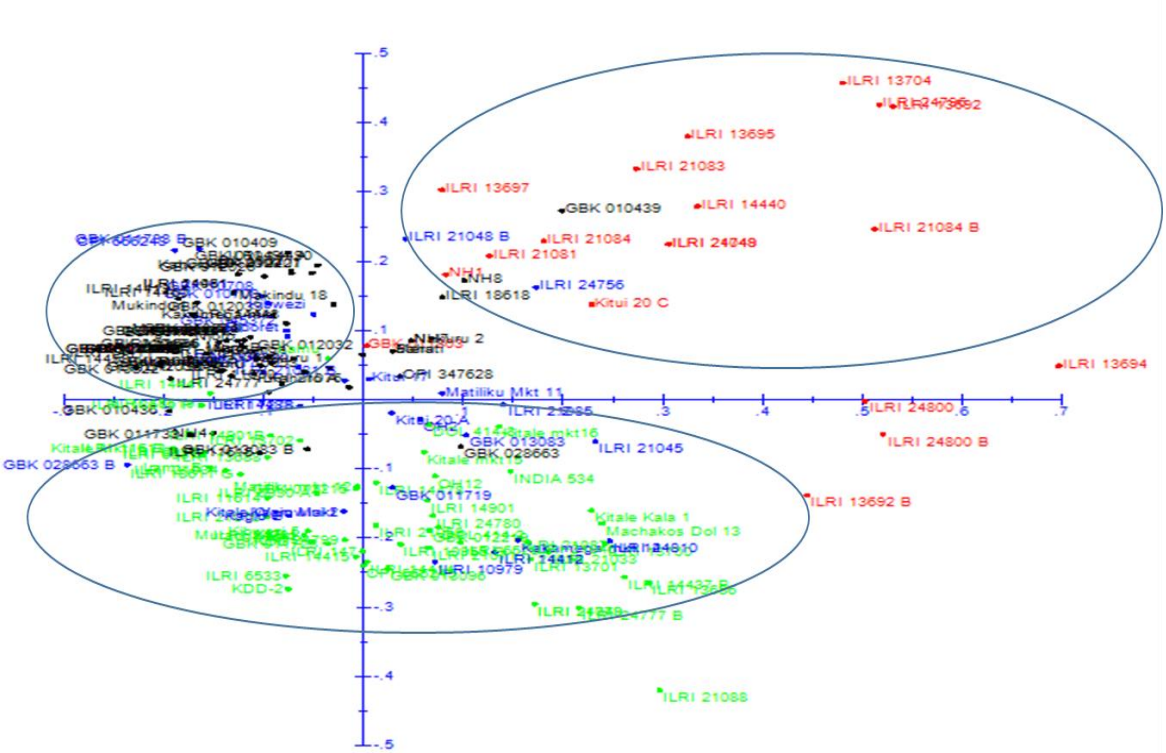


Figure 4. 6: Principal coordinates analysis (PCoA) of 189 lablab accessions using 8 selected SSR. The cluster 1identified in STRUCTURE is shown in green, cluster II in black, cluster III in red and the admixed cluster in blue.

For instance, accessions in cluster 1 (green) and cluster 3 (red) were widely dispersed in their distribution, while those in cluster 2 (black) were concentrated on the left hand side of the plot. This suggest that the accessions within cluster 1 and 3 are more diverse than those in cluster 2. The accessions in blue colour represents mixed group which could not match the three groups according to STRUCTURE classification. In PCoA scatter plot, these accessions (admixed group) did not form any definite grouping and were widely distributed in all the four quadrants of the plot.

4.4.6 Establishment of lablab Core collection

A neighbour joining sub tree representing the core collection was developed from the dissimilarity matrix between the 189 lablab accessions using the 'maximum length sub tree' function of DARwin 6.0.17 (Figure 4.7). Some advanced breeding lines, accessions with known valuable morphological traits or those popular in some lablab growing regions in Kenya were included in the core collection using DARwin 'forced' option.

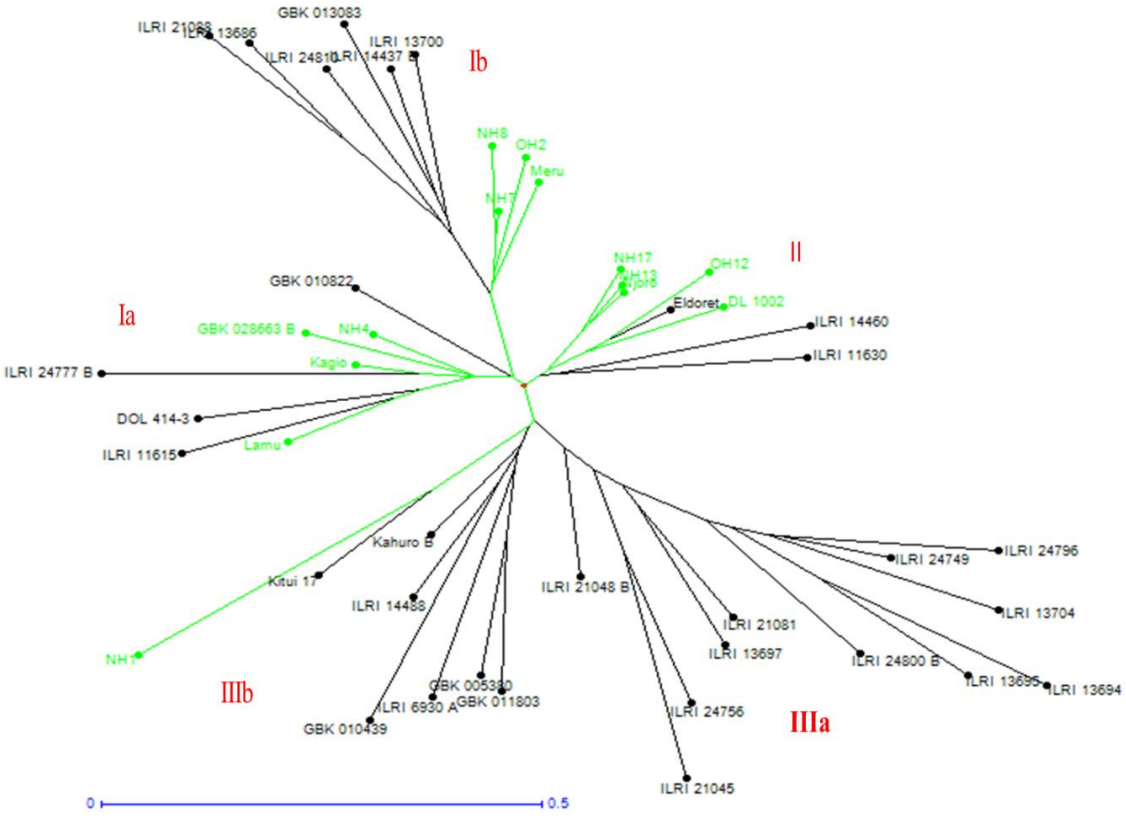


Figure 4. 7: Neighbour joining tree of the 45 lablab accessions identified for inclusion in germplasm core collection. The green coloured accessions are those which were 'forced' into the collection.

In this study, 45 accessions were selected to be included in the core collection. Among these, 14 were forced into the selection and these included seven advanced early

maturing lines (NH1, NH4, NH7, NH8, NH13, NH17, OH12 and OH2), Meru, Lamu, Kagio, Njoro, GBK 028663 and variety DL1002. The latter six were retained because they represent accessions popularly grown in the major lablab growing areas in Kenya. Using the group membership from the original DARwin clustering of the 189 accessions, the core collection comprised of nine accessions from cluster I, three accessions from cluster II, 31 accessions from cluster III and two accessions from cluster IV. The selected core collection originated from six countries. However, majority of them originated from Kenya (56%), Ethiopia (16%) and South Africa (16%). Among those selected from Kenya, five were from gene bank of Kenya (GBK 011803, GBK 005380, GBK 010822, GBK 013983, and GBK 010439). Majority of these accessions were collected from the eastern and coastal regions and they represent some of the most diverse accessions of the collection at gene bank of Kenya. Seven accessions (ILRI 13700, ILRI 14437, ILRI 11630, ILRI 11615, ILRI 6930, ILRI 13695 and ILRI 13694) comprised in the Pengelly and Maass (2001) core collection were also included in the current core collection. These seven accessions were also split in all the 3 clusters of this core collection. All the three subspecies were represented in the core collection with majority (78%) being from subspecies purpureus and subspecies unicinatus (13%) (Figure 4.8).

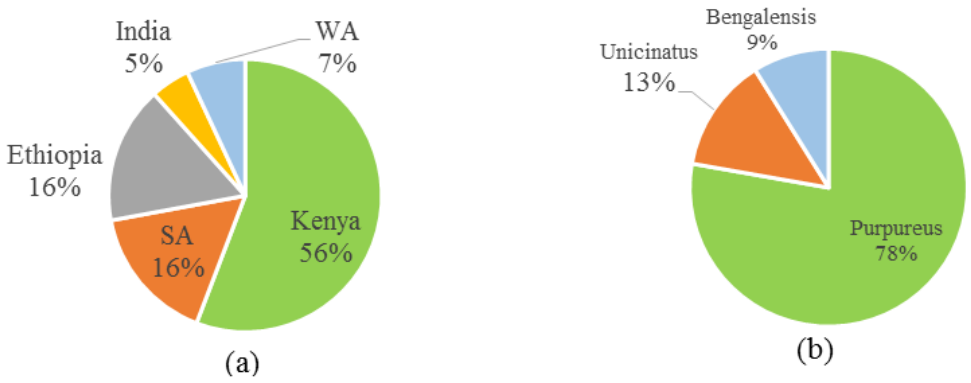


Figure 4. 8: The distribution of the accessions selected to represent core collection by (a) country of origin (b) the subspecies.

4.5 Discussion

The understanding of genetic diversity available in germplasm is important for any plant breeding program. It enables the breeders to select from large sets of genotypes the suitable parents for crossing to create new superior genetic variability. Genetic markers are preferred to morphological markers in germplasm characterization because they provide variability at genetic level thereby providing better estimate of genetic diversity (Shivashi *et al*., 2012). This study utilized lablab derived SSRs to understand the genetic diversity of lablab.

In the present study, 8 SSR markers were used for genetic diversity study in 189 lablab germplasm accessions. The microsatellite marker analysis found some moderate level of polymorphism in the worldwide lablab collection. The number of alleles per locus detected by the microsatellites were moderate ranging from three to seven with mean 4.88 respectively. The number of alleles suggests the richness of the population (Aljumaili *et al*., 2018). The higher the allelic richness at a locus, the higher the degree of diversity. The intensity of the alleles per locus identified in this study compares well with those reported in lablab by Wang et al., (2007), Shivachi *et al*., (2012) and Kamotho *et al*., (2016). However, allelic richness in this study is lower than 7.4 alleles

per locus reported in lablab by Robotham and Chapman (2015). The lower number of alleles per locus observed here could have been attributed to the difference in genetic population tested. In addition, the fragment analysis method used was not the same, manual acrylamide gels were run in this study while Robotham and Chapman (2015) used automatic capillary sequencing which could explain the differences.

Polymorphic information content (PIC) is a good estimator of the informativeness of molecular marker loci because it takes into account both the number alleles and the relative frequencies of the alleles. The higher the PIC value, the more informative is the SSR marker (Kamotho *et al*., 2016). According to Luo *et al*., (2019) molecular markers with PIC > 0.5 were considered to be highly informative; those with a PIC value of 0.25 to 0.5 were moderately informative while those with PIC value less than 0.25 were slightly informative. This study revealed polymorphisms of moderate informativeness (mean PIC value 0.419) suggesting their suitability for diversity analysis. Microsatellites are more informative than other markers due to their high mutations that evade correction during DNA mismatch repair system thereby allowing formation of new alleles in those loci (Vieira *et al*., 2016). The average PIC value in this study compares with 0.492 by Robotham and Chapman (2015) in lablab and 0.56 in Ali *et al*., (2015) but lower than 0.63 by Kamotho *et al*., (2016) in lablab.

The total gene diversity (GD) or average expected heterozygosity (H), is a good descriptor of the diversity of the alleles revealed in germplasm by the molecular markers (Bhandari *et al*., 2017). The GD value ranges from 0 to one with values close to one indicating very high allelic diversity. In this study, majority of the markers recorded moderate to high expected heterozygosity of more than 0.4 suggesting that the markers were able to reveal fairly high number of equally frequent alleles.

It is a common practice in genetic studies to test molecular markers for Hardy-Weinberg equilibrium (HWE). A population is said to be in HWE when the genotype and allele frequencies remain unchanged from one generation to the other (Graffelman & Weir, 2016). In this study, significant deviation from HWE was observed for all the SSRs. The Hardy Weinberg disequilibrium is expected considering the inbreeding nature of lablab. Self-fertilization create deviation from HWE through the reduction of genotypic frequencies of heterozygotes (Laird and Lange, 2011). Given the foregoing, it's clear that the SSRs markers used in this study were very efficient in revealing allelic diversity and could be useful for future investigations into genetic diversity in breeding and germplasm conservation programs.

Previous genetic studies of Kenyan accessions have reported low to moderate genetic diversity and have suggested the need to diversify the genetic base of the local accessions by introducing the wild cultivars and exotic germplasm (Kimani *et al*., 2012; Shivachi *et al*., 2012). The diversity of Kenyan germplasm in relation to those from the region and other continents remain unknown. This study considered diversity of a wide range of accessions collected from Kenya, other parts of Africa and beyond the continent. The present study showed that the heterozygosity expected under Hardy Weinberg Equilibrium (He) was moderate (mean He=0.38) for the 189 lablab accessions studied across eight SSR loci. In addition, the low number of different allele (Na=2.8) and average effective locus (Ne=1.93) further reflected the low to moderate genetic diversity of these accessions. The expected heterozygosity, number of different allele and effective number of locus are important indicators of population diversity and its potential for adaptability (Greenbaum *et al*., 2014). The higher the value of these parameters the higher the degree of diversity present in the germplasm under study.

The highest gene diversity of > 0.5 was detected for Ethiopia and South Africa populations. This is not unexpected because the origin of lablab has been suggested to be Southern and Eastern Africa while Ethiopia is considered as a Centre of diversity (Maass *et al*., 2016). Regions of origin and diversification are expected to contain larger genetic diversity (Bernard, *et al*., 2018). Another reason for high genetic diversity in the South Africa population could be due to the high number of wild accessions included in the population. Among the 27 accessions originating from South Africa, nine accessions belonged to sub species unicinatus. Pengelly & Maass, (2001) reported that the wild accessions from eastern and southern Africa represented a far wider diversity than the cultivated accessions. Maass *et al*., (2005) further classified the wild accessions belonging to sub species unicinatus into two type's namely 4-seeded pod and 2-seeded pod and noted that the 2-seeded pod types were distinct from the other crop germplasm. In this study, out of the nine accessions of sub species unicinatus included in the South Africa population, four belonged to the 2-seeded pod type further explaining the high gene diversity in this Southern Africa population. The moderate gene diversity (He=0.363) for Kenya population observed in this study compares well with that of Kamotho *et al*., (2016) in lablab. Whereas there are substantial amount of genetic diversity in Kenyan lablab accessions, the country breeding programs can benefit more from introducing germplasm from other countries especially those from South Africa and Ethiopia because they contain many different loci.

Fst is an important measure in understanding the extent of population differentiation due to genetic structure. Frankham *et al*., (2002) suggested that an Fst value greater than 0.15 is significant for differentiating populations. In this study, the Fst values among majority of the populations with exception of that from Uganda was lower than 0.15 indicating low genetic differentiation between these populations. The small Fst

values observed between the populations in this study suggest presence of high similarity of allele frequencies within the sub populations. Fst is highly related to the variation of allele frequency among populations (Holsinger *et al*., 2009). Relatively high Fst (>0.15) values were obtained between Uganda population and others implying that Uganda population was more dissimilar than the others. However, the high estimate of Fst for this population could also be attributed to the small number of individuals (three) included in this population. In population structure analysis, reliable estimation of population differentiation depends on the presence of either large population sample size (> 20) or large number of markers (Willing *et al*., 2012). More accessions from Uganda should be further studied before concluding that the population has different allele frequencies from those in other regions.

In the present study, the low genetic differentiation among the populations as indicated by the Fst values was further confirmed by the AMOVA results which showed that variation within sub populations accounted for the majority (94%) of the total variation. The gene flow between populations was also significant and with values ranging from 0.73 to 8.67 and majority of the populations having values of > 2. The significant gene flow reported here could have contributed to the transfer of alleles among the populations thereby resulting in low genetic differentiation between the populations. There is a long history of exchange of lablab germplasm within Africa and other regions of the world as reviewed by Maass *et al.,* (2016). For instance, before 2000 B.C lablab was transferred from Africa to India through Sabaean lane. During the slave trade lablab was transferred from West Africa to Brazil and Caribbean region where it was cultivated. After the abolishment of slave trade, the Indians carried lablab from the Caribbean region back to India sub-continent. In the present era, gene flow between populations could also be attributed to sharing of germplasm conserved in the gene

banks. Currently, a huge collection of more than 3000 lablab accessions collected from various regions of the world are conserved in over 13 gene banks worldwide (Maass *et al*., 2010). These materials are available for the research and development projects across different countries thereby contributing to gene transfer across regions. The cross boarder grain trade between countries could also have contributed to the high gene flow between populations. For instance, in 2014, Kenya imported lablab grain worth 124,000 Mt from the region (AFFA, 2014). Some of these grains could have been used as seeds by small holder farmers who purchase grain in the market for planting (Kamotho, 2015).

To further understand the population structure in the 189 world wide lablab accessions, Neigbour Joining phylogenetic trees, STRUCTURE and Principal Coordinate Analysis (PCoA) were used to identify genetically similar individuals based on the SSR allele frequencies. In Neighbour joining method, the phylogenetic trees are constructed by finding pairs of neighbours with minimized branch length and clustering them at each stage (Saitou & Nei, 1987). It would be expected that accessions originating from geographical areas that are distantly located to each other for instance South Africa and India to be clustered separately from each other due to the wide spatial range separating them. However, the results in this study indicated that the NJ tree clustered the accessions into four groups which did not fit well to their geographical origin. Similarly, the Bayesian model-based STRUCTURE analysis method clustered the accessions into three groups not related to their location of origin. The overlay in clustering between individuals from different regions and countries observed in this study suggest widespread genetic exchange among the regions (Luo *et al*., 2019). The results were consistent with that of Maass *et al.,* (2005) and Zhang *et al*., (2013). The study of Maass *et al.,* (2005) showed that using AFLP markers, 103 lablab accessions from Africa and

Asia were clustered into 12 groups that represented plant types and not the geographical origin. The study of Zhang *et al*., (2013) used 22 SSRs to screen 24 lablab accessions from China and Africa and showed that some materials from the two regions were clustered together in one group. Notably, in both the NJ tree and STRUCTURE analysis clusters, only accessions from African origin were present in all the groupings suggesting huge genetic variation from this continent.

The proportion of the accessions allocated to the STRUCTURE admixture cluster in this study was small (22 accessions out of 189). This suggest that even though some evidence of widespread genetic exchange in the region exist, there is low intraspecific hybridization occurring among the populations. This would support the argument by Maass *et al*., (2010) that lablab as an underutilized species has undergone limited crop improvement and that the few available improvement programs are based on limited genetic diversity. Lablab is predominantly self-fertilizing crop (Kamotho *et al*., 2015), and therefore any population admixture can significantly occur through the organized crossing of parents in breeding programs.

The present study showed that in all the three population structure analysis approaches (NJ tree dendogram, STRUCTURE analysis and PCoA) the accessions belonging to sub species unicinatus were clustered together and shared the clusters only with individuals originating from Africa. The accessions which clustered closely to sub species unicinatus were all from Southern and Eastern Africa (Ethiopia, Uganda and Kenya) regions suggesting the origin of lablab to these areas. Previous studies have indicated Africa as origin for lablab (Maass *et al*., 2005; Verdcourt, 1970; Robotham and Chapman, 2015).

The accessions collected from gene bank of Kenya and those obtained from ILRI gene bank were split across all the four clusters of NJ tree and the three clusters of

STRUCTURE indicating their wide diversity. This suggests that these gene banks are important reserviours of lablab allelic diversity that is much needed for the genetic improvement of this underutilized legume. Gene banks are centres of *ex-situ* conservation of wide range of plant genetic diversity necessary for improvement of crops. Lablab improvement programs should therefore introduce and integrate into their breeding programs the unique germplasm collection held at various gene banks (Maass *et al.,* 2010).

Previous molecular diversity studies of lablab accessions from Kenya reported narrow genetic diversity and suggested the need to introduce exotic germplasm from other countries to broaden the genetic base (Kimani *et al*., 2012; Kamotho *et al*., 2016). The Kenyan accessions evaluated in this study were distributed in all clusters of both NJ tree and STRUCTURE but some clusters had more membership than others. The overlap of the Kenyan genotypes with other accessions from the rest of Africa and beyond the continent in the population structure indicate that the local germplasm is a significant representation of the genetic diversity available globally. An important consideration in plant breeding is in the selection of very divergent parents for use in artificial crosses to generate superior recombinant genotypes (Bertan *et al*., 2007). The SSR loci used in this study was successful classifying the Kenyan accessions into different clusters which can be useful for plant breeders in selection of parents for their crossing program. For instance, selection of parents to improve the local popular variety DL1002 would be made from either cluster II, III or IV of the NJ tree or group 1 or 3 of STRUCTURE.

Plant breeders are faced with a problem of choosing which accessions to use for their breeding program among the large germplasm collections usually containing duplicate accessions. The precise understanding of the genetic diversity in germplasm is

important for its utilization in crop improvement programs (Campoy *et al*., 2016). A small sample representing the highest genetic diversity of the large collection can be identified and could become a source of promising parents for cross breeding. In the present study, a core collection was constructed from a worldwide collection of accessions from Africa, India, Australia and USA. A good core collection should contain few accessions representing a sub sample of the other accessions, with maximum genetic variation and representative of all geographical regions (Bernard *et al*., 2018). The representative sample of the wider germplasm can be developed using either morphological, biochemical or genetic data (Ndjiondjop *et al*., 2017). The current study used SSR allelic data to develop a core collection consisting of 45 accessions mainly from Africa and India and represented all the three sub species of lablab. Whereas these accessions don't replace the existing germplasm, they represent a set of diverse materials that are of priority to the local breeding program. Lablab is an important source of grain in the semi-arid areas of Kenya where it's grown. However, with the increasing effects of drought especially in semi-arid areas where lablab is grown under rain fed and with the dwindling land sizes has led to interest in developing cultivars that are early maturing and more drought tolerant. The core collection developed in this study provide an improved understanding of the diversity existing within germplasm collection that has potential to impact on the local lablab breeding program. For instance, the current core collection consist of a drought tolerant wild accession ILRI 24796 (Robotham and Chapman, 2015) and early maturing materials ILRI 11615, ILRI 6930 and ILRI 14437 (Pengely and Maass, 2001) which if incorporated in the local breeding program may provide improved cultivars. Our study divided the core collection into three clusters with majority of the accessions from Kenya clustered in group 1 and 2 suggesting that successful improvement of local

materials should target the exotic materials in cluster 3 as parents for cross breeding. Cluster 3 of this core collection contain five accessions (ILRI 21048, ILRI 21045, ILRI 24800, ILRI 24749 and ILRI 24796) belonging to subspecies unicinatus implying its importance as a source of high genetic diversity. This agrees with previous studies that have suggested that the two forms of sub species unicinatus require to be integrated in the lablab breeding program because they carry valuable gene resources (Maass *et al*., 2005). The wild accessions can be crossed with local cultivars to produce superior recombinants. Successful introgression of traits in a wild accession into a commercial cultivar to produce a high yielding variety has been done in other countries (Maass *et al*., 2010).

Evidently, some materials selected for the present core collection were also included in previous core collections. A case in point are the seven accessions (ILRI 13700, ILRI 14437, ILRI 11630, ILRI 11615, ILRI 6930, ILRI 13695 and ILRI 13694) selected for the current core collection and had previously included in the Pengely and Maass, (2001) core collection developed using agro-morphological traits. This confirm that the SSR markers used in this study were efficient in clearly discrimination of genotypes. The three large seeded Ethiopian accessions (ILRI 13694, ILRI 13695 and ILRI 13704), 4-seeded wild accession (ILRI 24796) and two 2-seeded wild accessions (ILRI 21045 and ILRI 24800) from South Africa were present grouped together. In addition, the moderate flowering accession ILRI 13700 was presently classified with accession ILRI 13686 in another group. This is in accordance with previous results based on different EST SSRs data (Robotham and Chapman, 2015). However, the grouping of accessions in the current study was different from that of Pengely and Maass, (2001) who allocated the early flowering accessions with long pods (ILRI 6930 and ILRI 14437) in ‘Zwai 2’ group while the two accessions (ILRI 13700 and ILRI 11630) with vigorous climbing

stem and moderate maturity period were placed in different group (‘Zwai 3’). The big seeded and early maturing accessions (ILRI 11615, ILRI 13694 and ILRI 13695) were allocated into different groups ‘Zwai 5’, ‘Zwai 7’ and ‘Zwai 8’ respectively. In current study, however, the early flowering and large seeded accessions (ILRI 13694 and 13695) were placed together (group 3) with the long podded accession ILRI 6930, while accessions ILRI 14437, ILRI 11615 and ILRI 13700 were clustered together in group 1. This inconsistencies suggests that the two studies picked and estimated different variations present in the accessions probably due to the different markers applied in both studies. Molecular markers and morphological traits complement each other in providing reliable insight into the genetic diversity of species (Kumar *et al*., 2009).

While breeding legumes for human consumption, seed colour and grain quality are important selection criteria (Pengelly and Maass, 2001). Farmers can use seed colour to select for varieties that are preferred in the market and with acceptance in the local culture (Ishikawa *et al*., 2019). In Kenya, the most preferred lablab in the market include black, brown, cream and tan seeded types although the seed colour preferences varies from region to region (Kamotho, 2015). The core collection developed here consist of accessions with varying seed colour and seed sizes and therefore offer new potential for improving traits required at different market segments.

In this study, SSR allelic data was used to assess genetic diversity and population structure among a worldwide lablab collection and possibility of developing a core collection for crop improvement program. Based on our data, the SSRs markers were efficient in revealing allelic diversity in lablab germplasm. Moderate genetic diversity was established in the panel of accessions studied. The highest gene diversity was detected for Ethiopia and South Africa populations. Kenyan accessions were of moderate genetic diversity with mean expected heterozygosity of 0.36. A core

collection consisting of 45 accessions mainly from Africa and India and including all the three sub species of lablab was suggested. This set of diverse materials is of priority to the local breeding program.

CHAPTER FIVE

Genetic Relationships and Population Structure of Worldwide Dolichos (*Lablab purpureus*) based on DArTseq Technology

5.1 Abstract

Lablab purpureus (L) sweet is a legume species that has been cultivated for many decades but has remained underutilized in many regions. Vast collection of lablab germplasm is maintained in various international and local gene banks and by farmers. However, there is lack of comprehensive understanding of the genetic diversity of this genetic resource. Diversity Array Technology (DArT) is a high-throughput molecular marker system developed for genome-wide marker discovery. This study was undertaken to determine the genetic diversity and population structure of 240 lablab accessions collected from various regions of the globe using DArTSeq derived molecular markers. A total of 15,601 polymorphic DArT markers were identified in a set of 240 lablab accessions. The average reproducibility and call rate of the markers was high at 99% and 95.5% respectively. The PIC values of the DArT markers ranged from 0.49 to 0.008 and an average of 0.06. In addition, 11,431 SNP markers were identified in this study with an average reproducibility and call rate of 99% and 91.9% respectively. The PIC of SNP markers ranged from 0 to 0.5 with an average of 0.06 with only 149 (0.13%) markers being monomorphic. Nei & Li (Dice) genetic dissimilarity matrices generated based on DArT markers ranged from 0.00 to 0.940. Wild accessions from South Africa and Uganda displayed exceptionally highest genetic dissimilarity (> 0.8) from the other accessions suggesting they are distantly related to the cultivated subspecies. Based on both DArT and SNP markers the germplasm evaluated had a narrow genetic diversity with the expected mean heterozygosity for DArT 0.030 and SNP at 0.039. Cluster analysis based on unweighted neighbour joining method using DArT and SNP markers separately, classified the accessions into seven groups each while six clusters were inferred by STRUCTURE method. The grouping showed that lablab genetic differentiation was most pronounced between the cultivated and the wild accessions. Based on STRUCTURE clustering, accessions collected from the farmers' fields and local markets fell in only two out the six populations suggesting that farmers in Kenya are cultivating lablab of narrow genetic diversity. The two markers have successfully classified lablab accessions into clusters which forms basis for selecting the most suited parents for improving the local cultivars and for conservation efforts.

5.2 Introduction

Lablab purpureus (L) sweet commonly known as dolichos is a legume species that has been cultivated for many decades. Lablab is a staple food for some communities in Africa and Asia providing a cheap source of protein in their diets (Kilonzi *et al.*, 2017).

The young and immature pods and leaves are consumed as vegetables. The foliage is used as animal feed while the crop residues after harvest are used as manure (Pengelly & Maass, 2001).

Lablab purpureus belongs to the family fabaceae and is acknowledged as monotypic genus (Maass, 2016). According to Maass *et al*., (2016), the origin of lablab is eastern and southern Africa because, these are the only regions where the wild accessions naturally occur. Lablab is one of the most agro-morphologically diverse legume (Maass *et al*., 2010). Based on agro-morphological traits, Verdcourt, (1970) recognised three subspecies of *L.purpureus* namely; cultivated subspecies, subsp. *purpureus* and subsp. *bengalensis* (Jacq.) Verdc and wild subspecies u*ncinatus* Verdc. Pengelly & Maass, (2001) and Verdcourt, (1970) observed that though subsp. *purpureus* and subsp. *bengalensis* were morphologically different, the two subspecies were closely related and freely interbreed. The subspecies *unicinatus* has two distinctive morphotypes namely two seeded or four seeded types.

Success of any crop improvement program depends on the availability of diverse germplasm of the crop. This is because crop germplasm contain important genes which enable the species to respond to changes in climate, pathogen populations and agricultural practices (Manifesto *et al.,* 2001). The genebanks and landraces maintained by farmers are the major sources of genetic resources for crop improvement (Kyratzis *et al*., 2019). In lablab, there is a huge collection of germplasm maintained in various international and local genebanks. For instance, Maass *et al*., (2010), reported that more than 3000 lablab accessions were maintained in Africa, Asia, Oceania, America and Europe. In Kenya, about 300 accessions are maintained at the Gene Bank of Kenya (Kinyua and Kiplagat, 2012). The understanding of the genetic diversity of such a huge

collection of lablab accession is crucial for their effective management and utilization in breeding programs.

Lablab germplasm have been characterized based on morphological markers (Pengelly & Maass, 2001; Maass & Usongo, 2007; Kamotho, 2016; Vaijayanthi *et al*., 2016), molecular markers such as random amplified polymorphic DNA (RAPD) (Liu, 1996), amplified fragment length polymorphism (AFLP) (Maass *et al*., 2005; Venkatesha *et al*., 2007; Kimani *et al*., 2012), simple sequence repeats (SSR) (Wang *et al*., 2007; Shivachi *et al*., 2013; Robotham & Chapman, 2015; Kamotho *et al*., 2016). The number of primer/ primer pairs used in most of these studies were few because of the limited number of molecular markers available for this leguminous crop. The introduction of new sequencing technologies has enhanced the development of high marker density with better genome coverage at a less cost (Desalegne *et al*., 2017). Currently, sequence-based single nucleotide polymorphism (SNP) markers are fast becoming marker of choice in crop improvement programs because of their abundance, stability, efficiency and cost effectiveness. The other high throughput genotyping system becoming popular recently is the diversity array technology (DArT).

DArT is a high-throughput marker system first developed in 2001 by Diversity Array Technology Pty Ltd (DArT, Canberra, ACT, Australia (Jaccoud *et al.,* 2001). The DArT technology applies microarray hybridization approach and solid-state surfaces to produce numerous polymorphic markers in a single assay (Alam *et al*., 2018). Recently, DArTseq platform has been developed for genome-wide marker discovery. The platform uses restriction enzyme to reduce the genome complexity followed by sequencing of the restriction fragments (Edet *et al*., 2018). DArTseq produces two types of markers namely silico DArT and SNP markers. SilicoDArT markers are microarray markers that are dominant while SNPs are co-dominant markers (Alam *et al*., 2018).

DArTseq markers have been applied in genomic studies of several crops such as, common beans (Nemlı *et al*., 2017) macadamia (Alam *et al*., 2018), wheat (Edet *et al*., 2018) strawberry (SaÂnchez-Sevilla *et al*, 2015) and pinneaple (Kilian *et al.*, 2016). However, this useful technology has not been used in understanding of genetic diversity and population structure of lablab germplasm. A better understanding of genetic diversity among genotypes is crucial for breeding programs because it enables exploitation of germplasm resources and in selection of parental materials useful for development of populations.

5.2.1 Objectives

The objectives of this study were:

1. To determine the genetic diversity and population structure of worldwide lablab collection using silicoDArT markers;
2. To determine the genetic diversity and population structure of worldwide lablab collection using DArTSeq derived SNP markers.

5.3 Materials and Methods

5.3.1 Plant materials

The plant materials used in this study included accessions previously collected from farmers' fields and markets in various lablab growing regions in Kenya, some advanced determinate early maturing lines from KALRO lablab breeding program and accessions preserved at the GBK. The regions from where the accessions were collected included rift valley, coast, western and central Kenya. The advanced breeding lines were pedigrees of crosses between accession KDD, Njoro and GBK 028663. KDD is an accession collected from a local market at Nairobi, has white flowers, determinate growth habit, early maturing and with cream seed coat colour. Njoro and GBK 028663 have indeterminate growth habit, purple flowers, medium maturing and with black seed

coat. Also included was a collection from the International Livestock Research Institute Forage Germplasm (ILRI-FG) which comprised of cultivars from other parts of Africa, USA, Asia, Australia and accessions of unknown origin (Fig 4.1). The ILRI-FG was established and has been maintained since 1982. In total, 240 accessions were evaluated in this study (Appendix 5.1).

5.3.2 DNA isolation

The seeds of the 240 lablab accessions were sowed in a field at KALRO Thika Centre to raise seedlings to obtain leaves samples for DNA extraction. Leaves samples were harvested from the emerging plants at first trifoliate growth stage. The samples were put in labelled zip lock papers and immediately placed on ice box transported to the laboratory for DNA extraction. The DNA isolation was undertaken at world agroforestry (ICRAF) molecular laboratory in Nairobi. About one gram of the fresh sample was grind in pestle and mortal in liquid nitrogen and immediately transferred the ground sample into a falcon tube placed on ice. 20ml of high concentration extraction buffer was added to the fresh ground sample. The high extraction CTAB buffer contained 100mM Tris-HCL (pH 8.0), 20mM EDTA (pH 8.0), 3M Nacl, 3% PVP, 1% B-Mecaptol-ethanol and 3% CTAB. The samples were kept in water bath at 65^0C for one hour with intermittent shaking. The samples were then cooled in ice and 4ml of 5M Nacl was added, mixed thoroughly and centrifuged for 15 minutes at 3500 rpm. The supernatant were transferred into a new falcon tube and equal volume of chloroform: iso amyl alcohol (24:1) was added and thoroughly mixed and centrifuged for 15 minutes at 3500 rpm. The supernatant was transferred to a new 50 ml falcon tube and equal volume of chilled isopropanol was added and allowed to precipitate for one hour. The tubes were then centrifuged for 15 minutes at 3500 rpm and the supernatant was carefully discarded leaving behind the DNA pellet. The DNA pellet was washed

with 70% alcohol and air dried until all the alcohol had evaporated. The DNA pellet was dissolved in 50ul double distilled water. 10ul of Rnase-A (10mg/ml) was added and incubated at 50^0C with intermittent shaking. The DNA was then stored at -20^0 C.

5.3.3 DNA quantification and quality determination

The extracted genomic DNA was quantified by comparing with a DNA ladder (gene ruler) of one kilo base (1 kb) in a 0.8% agarose gel using 1x TBE buffer. The bands were then viewed in a gel box. The concentration of the sample DNA was further determined using Qubit florometry. The DNA samples were then normalized to 100ng/ul. 50ul of each sample was then sent for DArTseq analysis.

5.3.4 DArTseq analysis

The DNA samples were sent to Diversity Array Technology (Canberra Australia) for DArTSeq analysis. DArTseq analysis involves genome complexity reduction through use of a combination of restriction enzymes and sequencing of resulting representations on next generation sequencing platforms (Nadeem *et al*., 2018). The restriction enzymes used in DArTSeq analysis facilitates the selection of genome fractions corresponding to various active genes and avoids the repetitive fraction of the genome. DArTseq generated two types of data including SilicoDArT and SNPS. The SilicoDArT data are score for presence or absence of sequences obtained from genomic representations. The SNP data represent the SNPs in the fragments present in the genomic representations. For each data set, different parameters were included such as call rate, polymorphic information content (PIC), reproducibility and one ratio. The SNPs and DArTs from the DArTseq approach were assigned a value of 1 if present in the genomic representation of the sample and 0 if absent.

5.3.6 Data analysis.

5.3.6.1 Genetic diversity analysis

To determine the genetic relationships of the lablab accessions, DARwin software version 6.0.17 was used to generate Dice (Nei & Li) dissimilarity matrices based on silicoDArT markers data. The Dice dissimilarity coefficient was chosen mainly because of absence of information on ancestry of accessions studied (Kosman & Leonard, 2005). Dendograms were generated using Unweighted Neighbour-joining method of cluster analysis based on both the SNP and DArT markers data. The trees were bootstrapped with 2000 replicates.

5.3.6.2 Population structure analysis

The germplasm was divided into 10 different groups based on mainly geographical origin namely Kenya, Uganda, West Africa, South Africa, Australia, USA, India and Ethiopia. Included also were a group of accessions of unknown origin and advanced determinate bush breeding lines from a Kenyan breeding program. AMOVA was conducted on the DArT markers to assess the level of variation among the groups of accessions. This was implemented in GenAlEx 6.5 software. To reveal the population differentiation, the PhiPT values among the populations were computed. In addition, the gene flow among the populations was also estimated through the number of migrants per generation using the same software. Principal Coordinate Analysis (PCoA) of the DArT data matrix for the presence or absence of each allele was performed to reveal gentic relationships among the 240 accessions using GenAlEx 6.5.

The genetic structure of the accessions based on SNP markers was determined using the Bayesian based STRUCTURE software version 2.3.3 (Pritichard *et al.*, 2001). The

number of clusters (K) in STRUCTURE were set from 2 to 10. Each run consisted of a burn-in period of 50,000 steps followed by 100,000 Monte Carlo Markov Chain (MCMC) replicates. We assumed an admixture model and uncorrelated allele frequencies. To determine the most likely number of clusters we used delta value (ΔK) method proposed by Elvanno *et al*., (2005) using the STRUCTURE Harvester (Earl and Vonholdt, 2012).

5.4 Results

5.4.1 Quality of molecular markers generated

A total of 15,601 polymorphic DArT markers were identified in a set of 240 lines of lablab originating from different parts of the globe (Appendix 5.1). Most of the markers (15,461) showed high reproducibility of 99% (Figure 5.1). The average call rate of the DArT markers was 95.5% with all the identified markers having a call rate value of more than 80% (Figure 5.1). The average one ratio value of the silico DArT markers was 0.37 with 6565 markers having values more than 0.05. The PIC values of the markers ranged from 0.49 to 0.008 and an average of 0.06. Among the 15601 informative markers, only 3.6% had PIC values of more than 0.3 while 83.2% had values of less than 0.1. A threshold values of >96% for reproducibility, >95% for call rate, >0.004 for one ratio and >0.008 for PIC were used to select the markers for the study. In total, 9,672 silicoDArT markers with an average PIC 0.04 were picked using the above criteria. The distribution of PIC values of these DArT markers are shown in Figure 5.2.

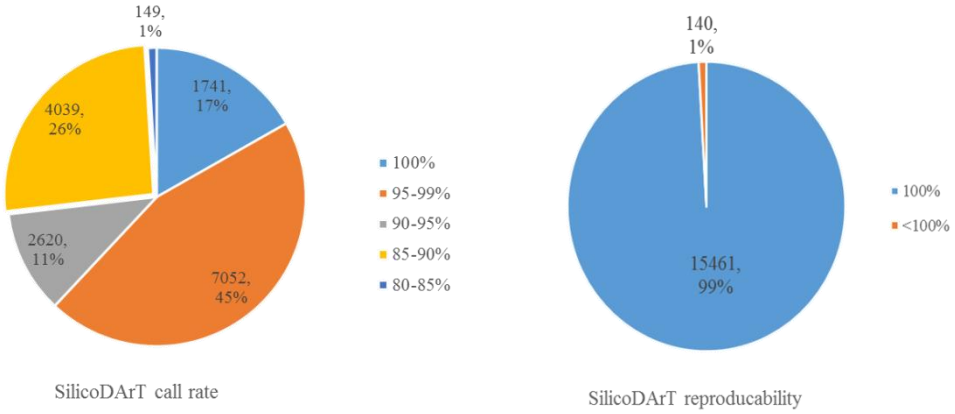


Figure 5. 1: Distribution of silicoDArT marker data for call rate and reproducibility parameters

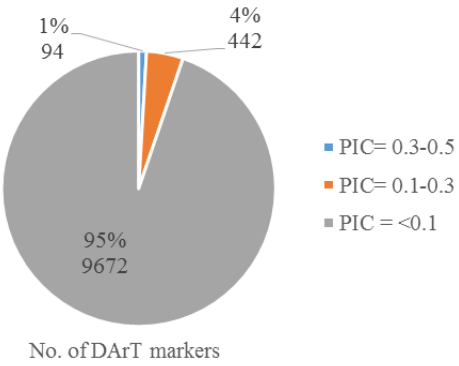


Figure 5. 2: Distribution of PIC values of DArT markers used for genomic studies in lablab

A total of 11,431 SNP markers were identified in this study. The average reproducibility of these markers was >99% with 94% of the markers having 100% reproducibility (Figure 5.3). The average call rate of the SNP markers was high at 91.9%. About 9288 of the markers displayed call rate values of more than 90%. The average one ratio ranged from 1 to 0.004 with an average value of 0.27. However, only 4221 (36.9%) of the SNP markers displayed a one ratio of more than 0.05. The average call rate value in this study was high at 0.93. The PIC values for the SNP markers ranged from 0 to 0.5 with an average of 0.06. Only 149 (0.13%)

SNP markers were monomorphic. About 3.5% of the total SNP markers had PIC values of more than 0.3 while > 86% displayed PIC values of less than 0.1 (Figure 5.4). Based on the high PIC and one ratio values, 5075 SNP markers were selected for genotyping lablab germplasm. The distribution of PIC values of these 5075 SNP markers are shown in Figure 5.5. It was noted that in many cases SNPs could not be called for wild lablab accessions. Due to the high number of missing SNP data for wild accessions, the genotyping of wild accessions was only done using DArT markers.

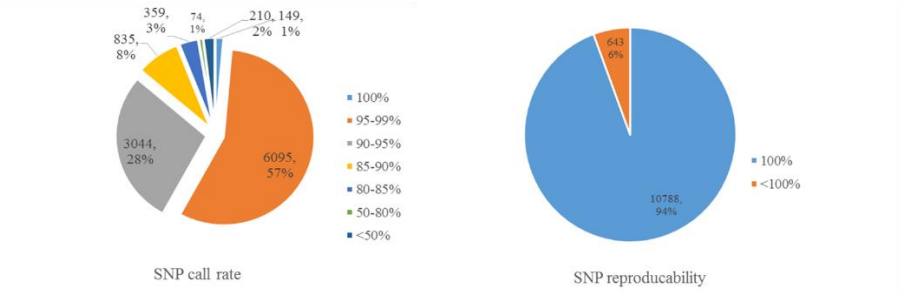


Figure 5. 3: Distribution of SNP marker data for call rate and reproducibility parameters

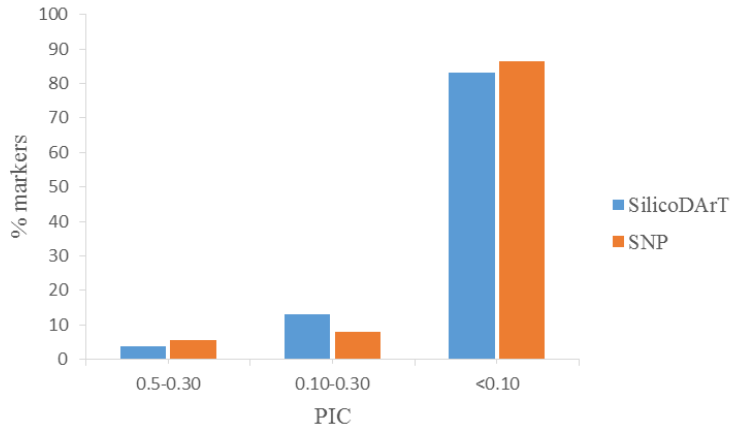


Figure 5. 4: Distribution of PIC values of silicoDArT and SNP markers developed

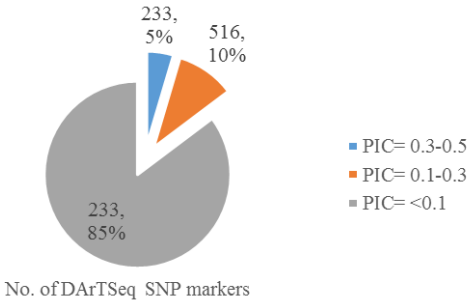


Figure 5. 5: Distribution of PIC values of DArTSeq SNP markers used for genomic studies in lablab

5.4.2 Genetic relationships among cultivars based on silicoDArT markers

DArT markers were used to understand the diversity of lablab accessions. Nei & Li (Dice) genetic dissimilarity matrices was generated to establish the level of relatedness of the lablab accessions. The genetic dissimilarity indices ranged from 0.00 to 0.940 with the 5 percentile at 0.46 and 95 percentile at 0.6023 (Appendix 5.2). This served as an indicator of diversity in the accessions studied. As expected, the replicate accessions were among those which revealed the least amount of genetic dissimilarity. For instance, accessions ILRI 11630 and ILRI 11630_1, ILRI 14441 and ILRI 14441_1, ILRI 21076 and ILRI 21076_1, ILRI 15436 and ILRI 15436_1 revealed genetic distance of 0.00. However, there are some accessions with different names and/ or origin but with very small genetic distance. For instance, the two accessions originating from West Africa, ILRI 14441 and ILRI 11630 revealed a genetic distance of 0.00 suggesting that the two could be duplicate accessions. Others with 0.00 genetic distance include two Kenyan accessions ILRI 14478 and ILRI 14443; GBK 10392 and GBK 10396. The study also revealed zero genetic distance between the Kenyan accession GBK 012033 and West African accession ILRI 14441. The local accessions collected from Meru Mukinduri and from Muranga market had 0.00 genetic distance with the

local variety DL1002. Other Kenyan accessions with close genetic distance with DL1002 include Fc Bahati (0.158), Fc Bungoma (0.153), Fc Kagio market (0.150), Fc Meru (0.156), GBK 010822 (0.156), GBK 12000 (0.163), GBK 013083 (0.162) GBK 045372 (0.120), ILRI 14443 (0.150), ILRI 14471 (0.158) and ILRI 24777 (0.144). The accessions belonging to sub species unicinatus from South Africa (ILRI 21048, ILRI 21083, ILRI 24749, ILRI 24800) and Uganda (ILRI 24756) displayed exceptionally highest genetic dissimilarity (> 0.8) from the other accessions suggesting they are distantly related to the cultivated subspecies. Notably, the accessions ILRI 21081 from Uganda and ILRI 21076 from unknown origin belonging to sub species purpureus also displayed high genetic dissimilarity (> 0.8) from the other accessions. Two Kenyan breeding lines (NH13 and NH4) developed through hybridization of determinate accession KDD and local cultivar Njoro exhibited very narrow genetic distance (0.084) between themselves indicating they are from the same gene pool.

The dendogram obtained with silioDArT markers revealed 9 clusters of related accessions (Figure 5.6). The clades of this sub group 1 were of almost the same height suggesting the tight closeness of the accessions in this group. This subgroup consisted of only accessions from gene bank of Kenya and from ILRI gene bank. Sub group 2 consisted of accessions obtained from ILRI gene bank but originating from outside Kenya with exception of ILRI 14901. Four South African accessions (ILRI 24781, 24780, 21085 and 21049) belonging to sub species purpureus were placed closely in this sub group 2. Similarly, three accessions from India (ILRI 18619, ILRI 21029 and ILRI 21033) were clustered together in this group suggesting they are from close gene pool. The neighbour-joining tree identified a cluster (Sub group 3) which consisted of the wild accessions belonging to sub species unicinatus. The tree clades of this sub

group had the longest height suggesting that these wild accessions are distantly related to the rest of accessions.

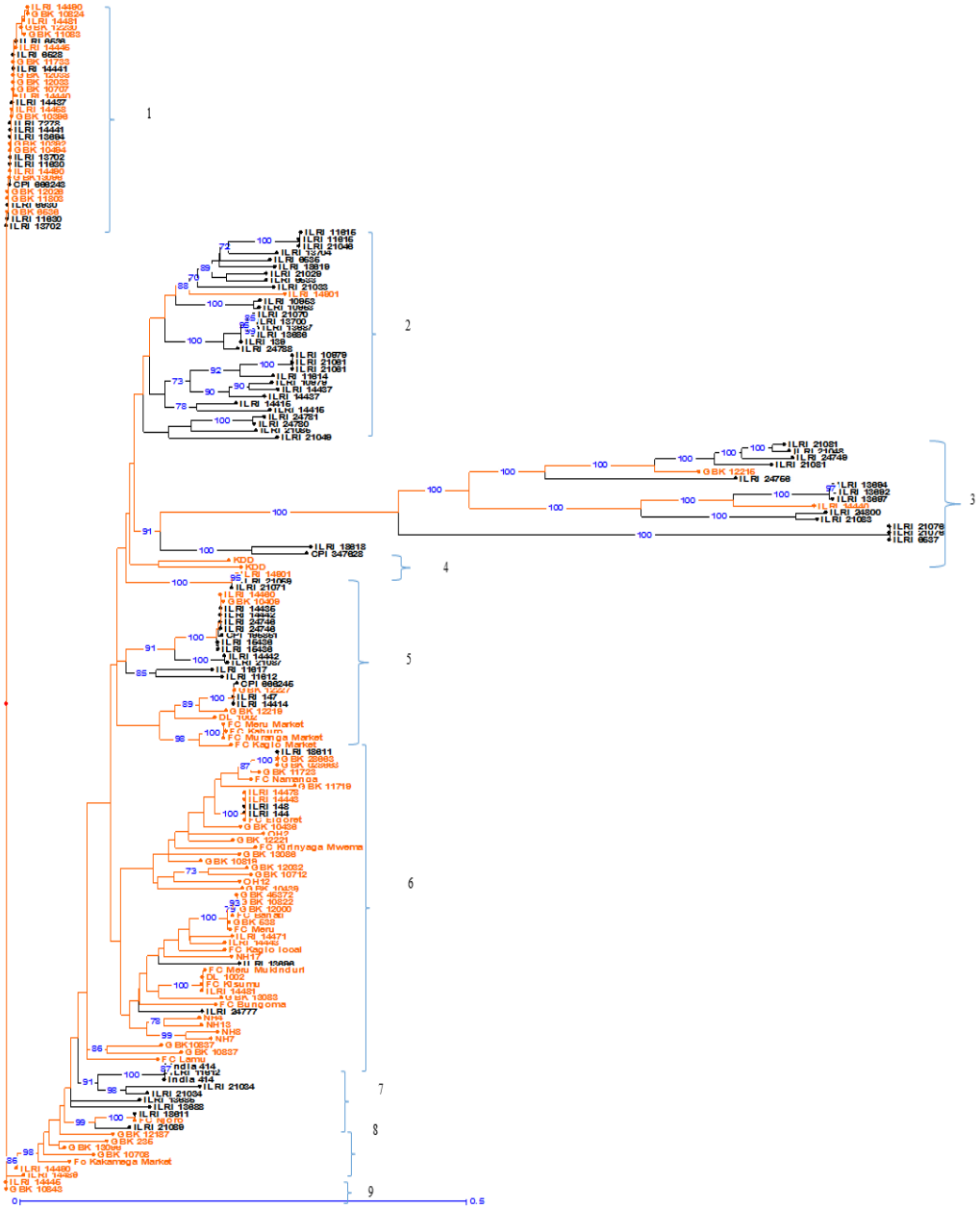


Figure 5. 6: Unweighted neighbour-joining dendrogram of 174 lablab accessions based on 9672 silicoDArT markers. The accessions in red font are of Kenyan origin while those in black font are from outside Kenya

The three big seeded accessions from Ethiopia (ILRI 13694, ILRI 13692 and ILRI 13697) and one accession from the gene bank of Kenya (GBK 012215) were placed together in this subgroup. This suggest that this Kenyan accession (GBK 012215) could be having some unique alleles that are different from the other local accessions.

Sub group 5 mainly included those accessions collected from ILRI gene bank but originating from Australia, Kenya, Ethiopia and South Africa. Notably, six accessions (ILRI 21059, ILRI 21071, ILRI 21067, ILRI 11617, ILRI 11612 and ILRI 14414) from Australia were included in this cluster indicating that these accessions are closely related. Sub group 6 consisted mainly accessions collected from farmer's field, markets and gene bank of Kenya. Four advanced lines (NH13, NH4, NH8, and NH7) which shared the same parents were also placed together in this sub group. The sub group 7 included accessions from India, Australia, Ethiopia and USA. The other two groups (8 &9) consisted of accessions from Kenya.

5.4.3 Population structure analysis based on SilicoDArT markers

In this study, the lablab accessions were divided into 10 subpopulations based on their origin of collection. The genetic diversity within each population was estimated through the expected heterozygosity. The overall mean expected heterozygosity estimate (He) derived from the 9672 DArT markers for the 10 populations was 0.03 while the Shannon index was 0.049. The expected heterozygosity and Shannon index were highest for the population of accessions of unkwown origin (0.13 and 0.212) and lowest for the advanced lines (0.005 and 0.007) (Table 5.1).

Table 5. 1: Mean expected heterozygosity (He) and Shannon Index (I) over DArT Loci for 10 populations of L. purpureus.

Population	Expected heterozygosity (He)[a]	Shannon Index (I)[a]
Advanced lines (AL)	0.005 ± 0.001	0.007 ± 0.001
Australia	0.006 ± 0.001	0.010 ± 0.001
Ethiopia	0.044 ± 0.010	0.092 ± 0.001
India	0.006 ± 0.010	0.010 ± 0.001
Kenya	0.009 ± 0.001	0.018 ± 0.001
Southern Africa	0.038 ± 0.001	0.071 ± 0.001
Uganda	0.043 ± 0.001	0.049 ± 0.002
Unknown	0.130 ± 0.001	0.212 ± 0.002
USA	0.008 ± 0.01	0.011 ± 0.001
West Africa	0.006 ± 0.001	0.009 ± 0.001
Mean overall loci and populations	**0.03 ± 0.001**	**0.049 ± 0.001**

[a]Means followed by the standard errors.

AMOVA was undertaken to partition the genetic diversity to within and among the 10 populations. The AMOVA revealed high and significant differences among the 10 populations of lablab, which accounted for 15% (probability of 0.023) of the total silicoDArT loci variance (Table 5.2).

Table 5. 2: Analysis of molecular variance (AMOVA) for 10 populations of Lablab purpureus and partitioning of the total diversity into population components

Source	df	SS	MS	Estimated Variance	Percentage of variance	PHIPT Stats	Pa
Among Pops	9	6950.118	772.235	38.567	15		
Within Pops	164	35438.152	216.086	216.086	85		
Total	173	42388.270		254.654	100	0.151	0.023

Pa - Probability values based on 999 permutations

The PHiPT values among pairs of populations were found to range from 0.00 to 0.831 (Table 5.3). Population differentiation was highest between Uganda and Kenya populations (PHiPT=0.831). The Uganda population was more highly differentiated from the other populations. The highest genetic distance between Kenya population and

others was recorded with Uganda (0.831), population from unknown origin (0.461) and South Africa (0.136). The least genetic distance was recorded between accessions collected from Ethiopia and those from India, USA and Australia. High level of gene flow was observed between Ethiopia and Australia (49.75), Ethiopia and South Africa (22.30) and Ethiopia and India (15.38). Notably, there was low gene flow between Kenya and Uganda (0.05), Kenya and Unknown population (0.29) and Kenya and South Africa (1.59). Generally, the least gene flow (< 1) was between Uganda and the other countries.

Table 5. 3: Pairwise population differentiation (PhiPT) values (below diagonal) and Nm values (above diagonal) between lablab accessions of origin in Africa Asia Australia and USA

	AL	AUS	ETH	IND	KEN	SA	UGA	UNK	USA	WA
AL	**0.000**	1.49	13.64	1.63	6.04	10.66	0.08	2.48	1.23	1.00
AUS	0.143	**0.000**	49.75	9.68	3.67	9.03	0.07	1.52	14.46	4.30
ETH	0.000	0.000	**0.000**	15.38	3.27	22.30	0.32	1.79	3.08	1.31
IND	0.133	0.025	0.000	**0.000**	3.85	10.09	0.07	1.72	2.25	3.94
KEN	0.040	0.064	0.071	0.061	**0.000**	1.59	0.05	0.29	3.53	5.68
SA	0.023	0.027	0.011	0.024	0.136	**0.000**	0.32	1.24	4.65	9.45
UGA	0.757	0.771	0.437	0.775	0.831	0.435	**0.000**	0.83	0.14	0.07
UNK	0.092	0.141	0.122	0.127	0.461	0.168	0.233	**0.000**	24.75	1.65
USA	0.169	0.000	0.000	0.100	0.066	0.000	0.646	0.000	**0.000**	3.57
WA	0.199	0.055	0.000	0.060	0.042	0.026	0.777	0.131	0.065	**0.000**

A PCoA analysis based on the PhiPT values clearly identified the lablab accessions into seven groups (Fig. 5.7). One group included three large seeded accessions from Ethiopia (ILRI 13692, ILRI 13694 and ILRI 13697). The second group contained three accessions belonging to subspecies unicinatus. The two seeded wild accessions in this group originated from Kenya (ILRI 14440) and South Africa (ILRI 24800 & ILRI 21083). The third group consisted of two South African accessions (ILRI 24749 & ILRI

21048) with four seeds per pod and belonging to subspecies unicinatus. Also included in this group was a Ugandan accession ILRI 21081. A wild accession from Uganda (ILRI 24756) was grouped together with a Kenyan accession GBK 12215. Two accessions (ILRI 21076 and ILRI 6537) were placed together in the 5th group. The five accessions originating from West Africa were all placed together in group six suggesting that they are closely related. All the other accessions were clustered together in one major group. The first and the second principal coordinates explained 53.76% of the total variation for the SilicoDArT markers.

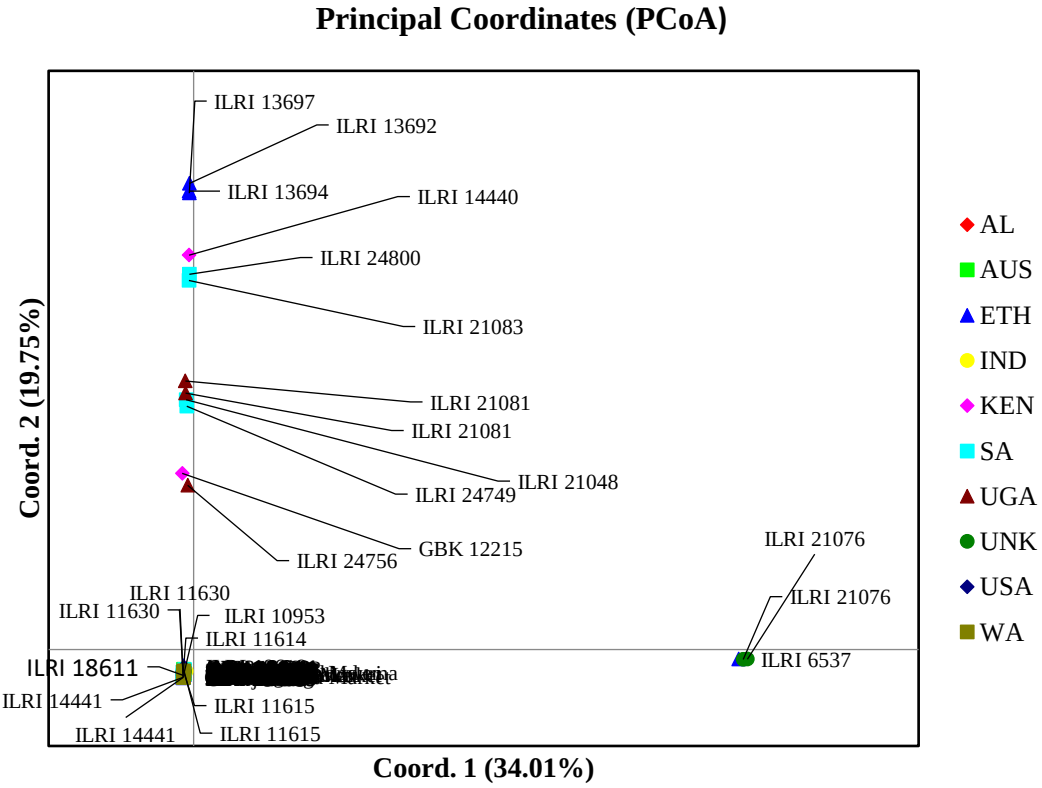


Figure 5. 7: Principal coordinates analysis based on PhiPT values for 10 lablab populations. Percentages in parentheses indicate the proportion of the total variation explained by each principal coordinate.

5.4.4 Genetic relationships among lablab germplasm based on SNP markers

The dendogram obtained with SNP markers revealed 7 clusters of related accessions (Figure 5.8). The subgroup 1 consisted of accessions originating from Kenya with exception of accessions ILRI 6536 (Ethiopia), ILRI 13696 (Ethiopia), ILRI 24777 (South Africa) and ILRI .

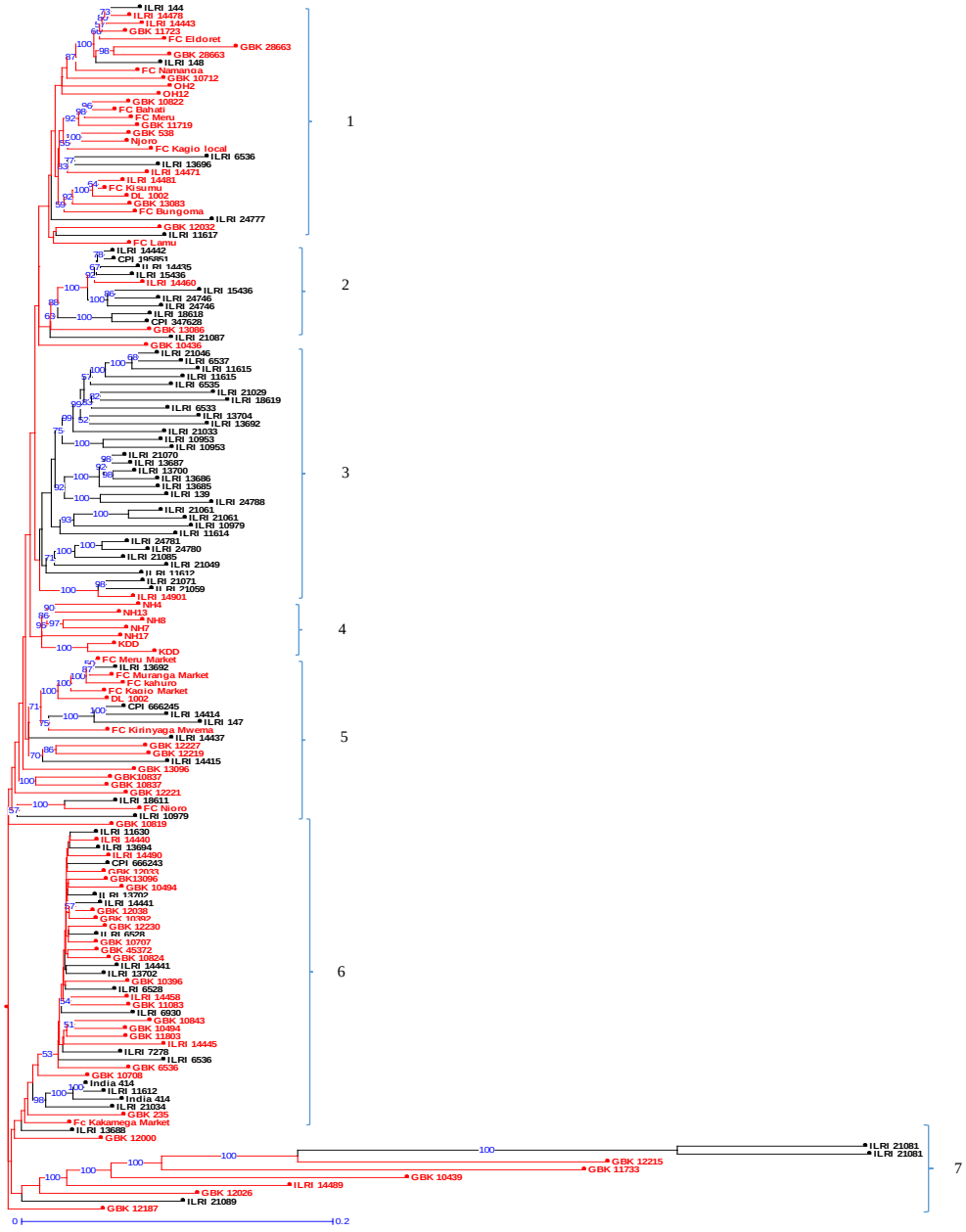


Figure 5. 8: Unweighted neighbour-joining dendrogram of lablab accessions based on 5075 SNP markers. The accessions in red font are of Kenyan origin while those in black font are from outside Kenya

11617 (Australia). Among the accessions included in this subgroup 1 from Kenya were collected from farmers' fields and markets at coast region (FC Lamu), Rift valley region (Namanga, Bahati, Njoro and Eldoret) and western region (FC Kisumu and FC

Bungoma). The other local farmers/market accessions (Meru Market, Kagio market, Muranga market, Kahuro and Kirinyaga Mwema) collected from central and eastern region of Kenya clustered together in a separate subgroup (subgroup 6). Two advanced breeding lines (OH 2 and OH12) clustered closely with one of their parent GBK 028663 in this sub group. The other dwarf advanced breeding lines (NH8, NH7, NH4, NH13 and NH17) clustered in the same group (subgroup 4) as their common parent (KDD). Sub group 2 consisted mainly of accessions originating from Ethiopia. Most of the accessions originating outside Kenya were clustered in subgroup 3. It was observed that in subgroup 3, accessions originating from the same country tended to cluster together suggesting their close similarity. For instance, four South African accessions (ILRI 24781, 24780, 21085 and 21049) were placed closely in this sub group 3 while accessions ILRI 13687, ILRI 13700, ILRI 13686 and ILRI 13685 from Ethiopia also clustered closely. Similarly, Australian accessions (ILRI 11612, ILRI 21071 and 21059) and Indian accessions (ILRI 21029 and ILRI 18619) formed two separate clusters in subgroup 3. Accessions ILRI 21081 originating from Uganda and Kenyan accessions (GBK 012215, GBK 011733, GBK 010439, GBK 012026 and ILRI 21089) formed a separate cluster (subgroup 7).

5.4.5 Population structure based on SNP markers

We explored population structure in the *Lablab purpureus* germplasm using the model-based Bayesian Structure method. According to Evano *et al* (2005), the real number of population is detected at the modal value or the upper most value of Delta K (Δ K). In this study, the value of Δ K was highest at K=6 (Figure 5.9) suggesting 228 lablab accessions could be assigned to six distinct groups. The six sub populations are denoted by Pop1, Pop2, Pop3, Pop4, Pop5 and Pop6. Pop1 and Pop6 contained the highest percentage of the total accessions at 28.7% and 23.4% respectively while the least

number (1.4%) was in Pop5 (Table 5.4). The genetic diversity of the populations as revealed by average expected heterozygosity was low at 0.0349 and ranged from 0.0022 (Pop1) to 0.1550 (Pop5).

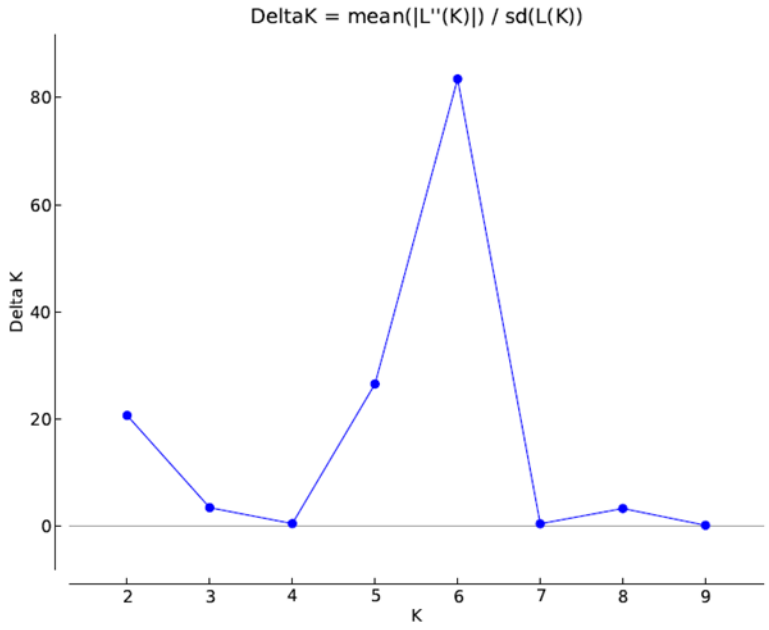


Figure 5. 9: Estimation of number of populations (K) of 228 Lablab accessions using SNP marker data, as estimated using the model-based Bayesian algorithm implemented in the STRUCTURE program.

Table 5. 4: The proportion of membership and expected heterozygosity of the 6 lablab populations based on SNP markers

Population	Proportion of membership	Expected heterozygosity (He)
Pop1	0.287	0.0022
Pop2	0.186	0.0236
Pop3	0.099	0.0038
Pop4	0.180	0.0149
Pop5	0.014	0.1550
Pop6	0.234	0.0104
Average	**-**	**0.0349**

The genetic divergence among the populations as revealed by Nei's net nucleotide distance (D) showed that Pop5 was the most distantly related to the rest of the populations (Table 5.5). The least genetic distance (D=0.0081) was observed between Pop4 and Pop6.

Table 5. 5: The genetic divergence among 6 lablab populations revealed by Nei's net nucleotide distance (D)

	Pop1	Pop2	Pop3	Pop4	Pop5
Pop1	-				
Pop2	0.0276	-			
Pop3	0.0248	0.0272	-		
Pop4	0.0252	0.0166	0.0258	-	
Pop5	0.1737	0.1662	0.1696	0.1677	-
Pop6	0.0246	0.0158	0.0276	0.0081	0.1699

The six populations detected through STRUCTURE analysis are shown in Figure 5.10. Each lablab accession is represented by a vertical bar. The full list of the names and codes of the accessions analysed using STRUCTURE are given in appendix 5.7. The bars are segmented into different coloured fragments which represent the estimated membership of a certain population. Majority of the accessions consisted in Pop1 originate from Kenya (maintained at either ILRI gene bank or at the Gene Bank of Kenya). Accessions ILRI 6528, ILRI 13694, ILRI 13688 and ILRI 13702 from Ethiopia were also included in this group. There are some accessions included in this Pop1 such as GBK 012187, GBK 012000, GBK 013096, GBK 011803, GBK 010837 and ILRI 18611 which showed intermediate to high mixed genetic composition. Pop 2 consisted of accessions originating from outside Kenya (Ethiopia, India, Australia, USA and South Africa). A determinate early maturing accession (KDD) collected from a market in Nairobi, Kenya was also included in this group. Two accessions (ILRI 13692 and ILRI 11615) belonging to the subspecies bengalensis were among the 11 accessions

entirely assigned to Pop2 (Figure 5.9). The rest of the accessions in Pop2 had mixed genetic composition of varying degree. The accessions with the highest genetic admixture in this group included ILRI 24780 (SA), KDD (Kenya), ILRI 10979 (USA), ILRI 13685 (Ethiopia), ILRI 14437 (SA), ILRI 14415 (SA) and NH8 (AL). The advanced line NH8 was clustered in Pop2 with together one of its parent, KDD. The Kenyan accessions collected from Muranga, Meru, Kirinyaga, Katumani regions together with GBK 012227, GBK 011804 and GBK 012215 were grouped in Pop3.The South African accession CPI 666245 and Australian collection ILRI 144414 were also clustered with the Kenyan accessions in Pop3. Pop4 consisted of mainly Kenyan and Ethiopian accessions. Two determinate early maturing lines (OH2, and OH12) from Kenya breeding program were placed together with a common parent, GBK 028663.

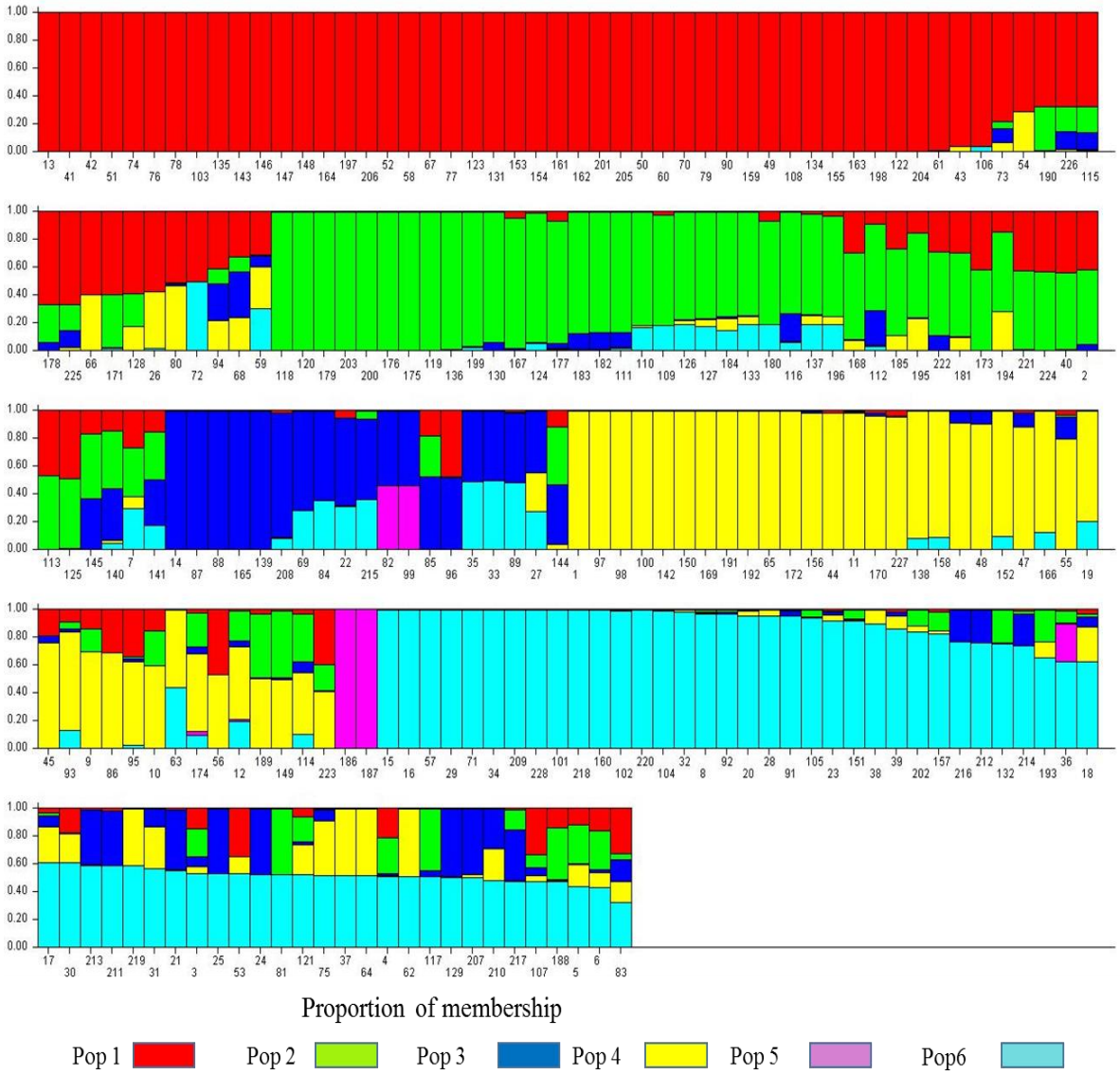


Figure 5. 10: Population structure of 243 Lablab accessions as determined by STRUCTURE analysis based on SNP allelic data at 5075 loci. The numbers on the x-axis are the accessions code, the scale on the y axis is the proportion of estimated membership.

Accession ILRI 21081 from Uganda was entirely assigned to Pop5. However, only GBK 012215 and FC Namanga from Kenya which possessed some significant genetic information from Pop5. Pop6 was dominated by Kenyan accessions mainly collected from the farmers' fields and local markets. The variety DL1002 released by Kenyan

breeding program was assigned to Pop6 together with the other local cultivars maintained by farmers. The advanced breeding lines (NH7, NH4, NH17 and NH13) which are progenies of Njoro X KDD were assigned together with their parent Njoro in this group.

5.5 Discussion

The understanding of the available genetic diversity in the germplasm collection is crucial for actual conservation and their utilization in crop improvement (Roorkiwal *et al*., 2014). Efforts have been made to develop large number of different types of molecular markers which can be used for understanding the genetic diversity and identifying germplasm with important traits for use in plant breeding programs. The present study has demonstrated the suitability of DArTseq technology as an effective tool in revealing the genetic diversity of lablab. The technology has yielded large number of polymorphic markers. A total of 15,601 polymorphic DArT markers were identified of which 9,672 markers were used to reveal genetic architecture of lablab accessions. In addition, the technology revealed 11,431 DArTSeq SNPs of which 5075 were informative. To the best of our knowledge, this is the first attempt to develop DArT marker system for lablab and to exploit these makers in understanding the genetic diversity of worldwide lablab accessions.

A good range of PIC value (0.04 – 0.49) was found in this study for both the DArT and SNP markers, which agrees with previous works on other legumes. Nadeem *et al*., (2018) found PIC values ranging from 0.10-0.50 , Nemlı *et al*., (2017) found values of 0.005 to 0.5 using DArTSeq SNP markers in common beans while Julius *et al*., (2021) reported PIC values of 0.02- 0.5 in lablab. Using 2763 DArT markers across 94 chickpea genotypes, Roorkiwal *et al*., (2014) reported PIC values of 0.02 to 0.37. In

pigeon peas, (Mir *et al*., (2012), found PIC values of 0.02 to 0.5 for both SNP and DArT markers. However, the average PIC obtained for DArT (0.06) and SNP (0.06) in this study was much lower than 0.4 (Nadeem *et al*., 2018) and 0.25 (Nemlı *et al*., 2017) in common beans using DArT and SNP markers respectively. This lower than expected average PIC reported here could be as a result of germplasm choice. Although the accessions selected for this study were picked to signify maximum diversity, majority of them were not previously subjected to any molecular analyses. The lower average PIC could also be contributed by the fact that in this study, the accessions were all from one purpureus species. However, the lower PIC values of DArT markers seem to be beneficial in giving defined grouping of accessions from distant genetic origin (Badea *et al*., 2011). For instance, the DArT PCoA was able to separate the sub species unicinatus from the rest of the sub species purpureus. The average genotype call rate and reproducibility of the markers identified in this study was over 91% suggesting that the markers were of high quality. This means that there was only less than 9% of the missing data for each marker. This call rate value compares well with 92% reported by Valdisser *et al*., (2017) while using DArTSeq SNPs in common beans. The genotype call rate is the proportion of genotypes per marker with no missing data. Notably, large number of SNPs could not be called for wild lablab accessions. This is probably because majority of the genotypes used to design the DArT and SNP markers in this study were from cultivated lablab and only few (less than 4% of the total accession) were from wild accessions. This could have resulted in low representation of the wild genome in the diversity panel. Similar results of low DArTSeq SNP call rate on wild accessions was also reported in chickpea by Roorkiwal *et al*., (2013). This study revealed the significance of DArTSeq platform as a dependable tool for assessing diversity lablab germplasm. The DArTSeq platform combines the DArT microarray platform, which

achieves the genome complexity reduction using combinations of restriction enzymes with next generation sequencing thereby producing large number of markers (Sánchez-Sevilla *et al*., 2015). Lack of suitable and numerous molecular markers for lablab has in the past been identified as the most limiting factor for breeding studies of this underutilized crop (Zhang *et al*., 2013). Previous studies in lablab using other markers used very few molecular markers, for instance, Shivachi *et al*.,(2013) utilised only 21 microsatellite primer pairs from common beans to investigate genetic diversity in the 13 lablab genotypes while Kamotho *et al*., (2016), utilized 10 SSR markers to study diversity of 96 lablab accessions. Therefore, the numerous number of both DArT and SNP markers produced in this study will be useful for diversity studies in lablab. The DArTSeq markers are not only cost effective but are known to have a wider genome coverage (Robbana *et al*., 2019). This suggest that these new markers are also valuable resource for future association mapping studies in lablab.

Our research showed that based on both the DArT and SNP markers the studied lablab material had a narrow genetic diversity. The expected mean heterozygosity (He) was low for both DArT (He = 0.030) and SNP (He= 0.035). The expected heterozygosity accounts for the frequency of the different types of alleles and their evenness in the population (Kimani *et al*., 2012). Higher mean heterozygosity (He) have been reported in lablab other studies. For instance, Kimani *et al*., (2012) reported mean heterozygosity of 0.189 in 50 Kenyan lablab accessions using AFLP. Kamotho *et al*., (2016) reported mean He=0.38 when using 10 SSRs while Zhang *et al*., (2013) found mean He=0.34 using 11 SSR markers in lablab. The low expected mean heterozygosity (He) in this study could be explained by the fact that the DArTSeq platform targets the genome fraction corresponding predominantly to active genes and therefore could be less variable in comparison to SSRs that are usually located in non-coding regions (Liu *et*

al., 2016). However, the high density DArT markers in the present study represents wider genome coverage which make these markers more reliable when it comes to understanding genome-wide genetic diversity (García *et al*., 2018). The low average gene diversity could also be attributed to the self-pollinating nature of this leguminous crop.

The DArT markers were effective in establishing genetic relationship of the lablab accessions, grouping related samples together. Using these markers, as expected, the replicate accessions showed the least genetic dissimilarity suggesting the efficiency of the markers in classifying the genotypes. The markers were also able to identify some accessions with different names but with very small or zero genetic distances amongst themselves suggesting duplicate germplasm. The duplicate accessions were detected both within the countries and between countries. For instance, this study found zero genetic distance between two accessions GBK 010392 and GBK 010396 both preserved at gene bank of Kenya and also between Kenyan material GBK 012033 and West African germplasm ILRI 14441. Germplasm collections regularly comprise duplicate accessions both within and between gene banks. This may be due to error when registering materials, collection from identical areas, breeding lines originating from common background and exchange of materials between gene banks. Plant breeders are faced with a problem of choosing which accessions to use for their breeding program among germplasm collections containing duplicate accessions. The understanding of the genetic relationship of the lablab germplasm generated in this study is therefore crucial for utilization of genetic resources in crop improvement programs (Campoy *et al*., 2016).

Wild relatives of many crops remain to be an important genetic resource. In the present study, the neighbour joining phylogeny tree and PCoA generated from the DArT markers data showed that lablab genetic differentiation was most pronounced between the cultivated and the wild accessions. The NJ tree and PCoA based on DArT markers were able to separate the two seeded wild accessions (ILRI 14440, ILRI 24800 & ILRI 21083) from the four seeded (ILRI 24749, ILRI 21048 & ILRI 24756) wild types. These two forms of sub species unicinatus need to be integrated in the lablab breeding program to maximize the diversity of the cultivated accessions. The wild accessions such as ILRI 24796 which has drought tolerance (Robotham and Chapman, 2015) could be useful in expanding the tolerance of cultivated lablab accessions to this important abiotic stress. Successful introgression of traits from a wild lablab accession into a commercial cultivar to produce a high yielding variety has been achieved in Australia (Maass *et al*., 2010). The cultivated accessions which clustered closely to sub species unicinatus were all from Eastern Africa including Ethiopia (ILRI 13692, ILRI 13694, ILRI 13697, ILRI 6537), Uganda (ILRI 21081) and Kenya (GBK 012215) suggesting the origin of lablab to these areas. Previous studies have indicated Africa as origin for lablab (Maass *et al*., 2005; Verdcourt, 1970; Robotham and Chapman, 2015). A similar distinct clustering pattern based on subspecific taxonomic organization, i.e., subspecies purpureus and subspecies uncinatus, as well as to cultivated and wild forms was reported by Maass *et al*., (2005) using AFLP markers.

The NJ tree based on SNP markers developed clusters of related accessions that were consistent with the results given by DArT markers. Although the clustering of germplasm was not exactly the same for both markers, accessions with common parents were grouped together. For instance, advanced lines with common parents like NH4, NH8, NH17 and NH8 were placed closely in both dendogram. Both markers identified

ILRI 21081 and GBK 012215 as among the most distinct accessions. Ethiopian accessions ILRI 13700, ILRI 13687, ILRI 13686, ILRI 139 were all grouped together in both studies. The consistency in determining accession relationship indicated that both the SNP and DArT markers were highly reliable for genetic diversity of lablab.

The genetic similarity determined in this study using DArT markers showed little geographic differentiation of the lablab accessions. The accessions collected from Kenya and other parts of the world were distributed in almost all the subgroups. The results were consistent with that of Maass *et al.*, (2005) and Zhang *et al*., (2013) who did not find clustering of lablab genotypes based on geographical origin while using AFLP and SSR markers respectively. The level of gene flow between the countries in the present study was high supporting the little geographic differentiation observed. Notably, there was high gene flow between Ethiopia, Australia and India suggesting widespread genetic exchange among the regions (Luo *et al*., 2019). The high gene flow from Ethiopia and India is probably because the two are considered as centres of domestication (Maass, 2016) and could have experienced a lot of sharing of germplasm between the two regions.

The SNP markers based NJ phylogeny tree of the lablab accessions suggested seven groups while six clusters were inferred by STRUCTURE. The group membership by the two clustering method were consistent. For instance, the accessions in Pop1, Pop2, Pop3, Pop4 and Pop6 in STRUCTURE were similar to those of sub group 6, group 3, group 5, group 2 and group 1 of NJ tree respectively. The populations in STRUCTURE were presented by different colours, red (Pop1), green (Pop2), blue (Pop3), yellow (Pop4), purple (Pop5) and sky blue (Pop6) and different number of accessions were discretely assigned to each (Figure 5.9). Majority of the accessions consisted in Pop1

originate from Kenya (maintained at either ILRI gene bank or at the Gene Bank of Kenya) and were entirely assigned to blue genome without admixture of genome. The fact that the genome of these accessions were not admixed suggest that they have not either experienced outcrossing with diverse germplasm or undergone any artificial hybridization targeting crop improvement. Gene banks as custodians of various genetic resources apply stringent protocols in the field to avoid genetic contamination of the germplasm which could explain why most of the materials from gene bank were not admixed. Notably, with exception of few, the accessions with entirely blue genome (Pop1) had white flowers and had either brown or cream seed coat thus indicating their similarity even at phenotypic level. Pop 2 (green genome) consisted of accessions originating from outside Kenya (Ethiopia, India, Australia, USA and South Africa) and was among the populations with the highest genetic diversity (He=0.0236). Future breeding programs in Kenya should target this population to select parents for improvement of their local cultivars since it may contain many different alleles which are not available in Kenyan accessions.

The majority of the lablab accessions collected from the farmers' fields and local markets were distributed in only two populations (Pop6 and Pop3) indicating the narrow genetic diversity of cultivated germplasm by farmers in Kenya. This could be attributed by the limited research and development of the crop in Africa (Maass *et al*., 2010) thereby leaving the farmers with limited choice of producing narrow genetic diverse landraces and a few commercial cultivars. In this study, DArT and SNP markers have been successfully used to classify lablab accessions into similar clusters which provides good basis for selecting the most suitable germplasm to expand the genetic base of the local cultivars.

Most of the populations with exceptions of Pop5 (purple) had a certain proportion of the accessions with admixture genome. Though lablab is predominantly self-pollinating crop some out-crossing has been reported (Kimani *et al*., 2012) explaining the possible reason for admixture. All improved varieties and advanced lines (DL1002, NH8, NH4, NH13, NH7, OH12 and OH2) had admixed genome from the parents populations indicating that artificial crossing could have also contributed to the admixture of genome. The Pop5 consisted of a single accession (ILRI 21081) which originated from Uganda. The STRUCTURE analysis based on SNP allelic data identified this population as the most diverse and had some similarity with the clustering pattern of "Darwin" which had grouped it together with the wild accessions. This suggest that this unique accession ILRI 21081 could be included in the breeding programs to expand the genetic base of this multipurpose crop.

The present study was one of the earliest diversity analysis in lablab using thousands of silicoDArT and SNP markers. The average genotype call rate and reproducibility of the markers identified in this study was over 91% indicating that the markers were of high quality. Both the silicoDArT and SNP markers were successful in identifying genetic diversity within the worldwide lablab collection. Narrow genetic diversity was observed among the studied germplasm using both markers. However, greater genetic differentiation was observed between the cultivated and wild accessions suggesting the need to integrate the sub species unicinatus in the breeding programs. Majority of the accessions currently cultivated by the farmers in Kenya and included in this study were found to have narrow genetic diversity.

CHAPTER SIX

Inheritance of growth habit and other qualitative traits and their linkage in local Kenyan Lablab (*Lablab purpureus*) populations

6.1 Abstract

An understanding of inheritance of qualitative and quantitative traits of crops and their relationship is important for developing effective breeding strategy. In lablab, flowering time and growth habit are traits of economic importance as they determine the adaptability of the plant to its growing environment. The information on inheritance pattern of these traits under the Kenyan environment and using local genetic populations is still lacking. Investigation was carried out with an aim of identifying the inheritance pattern of four lablab qualitative traits and establish the possible linkage relationship of the genes controlling these traits. The genetics of growth habit and the other qualitative traits were studied in three generations (F_1, F_2 F_3) of eight lablab populations. Six F_2 populations segregating for growth habit were also investigated for their segregation ratios under two different environments at KALRO Thika and KALRO Kiboko. Goodness of fit of the observed ratio of the traits to the expected ratio of one, two and three gene model was tested. A chi square test was done to determine the possibility of linkage of genes for the various traits while the frequency of recombination of the linked genes were determined using product-ratio method. All the F_1 plants had indeterminate growth habit suggesting that indeterminate is dominant to determinate. The segregation of growth habit under Thika environment fitted to the expected ratio of 3:1 but was 57:7 under Kiboko environment suggesting that growth habit in lablab is under control of three genes, one basic and two complementary genes which could be temperature dependent. A single dominant gene was associated with the presence of purple flowers while segregation ratio of 9:7 for stem anthocyanin suggested di-genic control with complementary epistasis. The segregation of the F_2 progenies revealed a monogenic inheritance control for days to flower initiation under Thika environment with the early flowering being recessive to late flowering. Distorted segregation from the expected ratio of 9:3:3:1 was observed between stem growth habit and time to flowering of F_2 populations of two crosses KDD x GBK 028663 and KDD x DL1002. The recombination fraction for the two traits was small at 0.05 and 0.14, while the mapping distance between the traits was short at 7.24cM and 14.07cM for the crosses KDD x GBK 028663 and KDD x DL1002 respectively which suggest possible linkage of the two trait pairs. The implications of the study findings on the breeding of lablab and seed production in Kenya are discussed.

6.2 Introduction

Lablab bean (*Lablab purpureus (L) sweet*) is a leguminous species that belongs to the family fabaceae. Lablab is grown in many regions of the world including Africa, Asia, Australia and USA (Maass *et al*., 2005; Kimani *et al*., 2012; Sennhenn, 2015). The crop is multipurpose; mainly grown for its young immature pods and leaves which are

consumed as vegetables, the dry seeds are used for various food preparations, leaves and stems for fodder, hay, silage and cover crop for soil conservation management (Kamotho *et al*., 2015).

The legume can grow in different agro-ecological zones ranging from the lowlands of the coastal region to the highlands of Mt. Kenya (Kamotho *et al*., 2015). However, due to its drought tolerance, (Kimani *et al*., 2012) production is mainly done in the semi-arid environments where other legume crops cannot perform well (Kinyua and Kiplagat, 2012).

In Kenya, lablab is predominantly grown by smallholder farmers usually in less than two acres of land. Due to their small land sizes, most of the farmers prefer to intercrop lablab with other crops mainly cereals.

Lablab improvement in Kenya is constrained by lack of appropriate genetic sources of early maturity and less spreading growth habit which is suitable for short season semi-arid environments and intercropping systems. Due to the effect of climate change, the early maturing lablab varieties like DL1002 which were developed for semi-arid areas of Kenya are now taking longer to mature than the length of the growing season in the area. The crops are therefore subjected to end of season moisture stress which leads to yield loss (Sennhenn, 2015). Even in drought tolerant crops like lablab, excess moisture stress especially during pods and seed development results in formation of few pods , shriveled grains and ultimately low grain yield (Fredrick *et al*., 2001).

Intercropping maize/ sorghum/ millet with lablab is very common to lablab small holder farmers in Kenya (Kamotho *et al*., 2010; Kimani *et al*., 2012). The most common practise is where the farmers grow a row of lablab between every row of maize or sorghum crop (Kamotho *et al*., 2015). In intercropping system proper selection of cereal/legume varieties is key as it increases yields of the intercrops and minimizes

intercrop competition ((Makgoga, 2014). In intercrop system, the indeterminate and vigorous climbing lablab varieties are not desirable because they climb on, pull down and lodge the intercrop cereal resulting in yield reduction of the cereal crop (Kamotho, 2015; Sullivan, 2003).

A plant displaying indeterminate growth habit has a terminal shoot meristem that remains in a vegetative state even after the initiation of production of reproductive structures. However, the determinate types are characterized by the main axis terminating into an inflorescence (Repinski *et al*., 2012; Keerthi, 2014) usually in the early stages of development. This results in reduced number of nodes, leaves and shorter plant height which is especially desirable for cereals intercrop system. Determinate growth habit allows more synchronous flowering, more uniform pod maturity and thereby enabling cost-effective harvesting. Plants with determinate growth habit are associated with earliness to flowering and maturity.

Development of early maturing lablab varieties in Kenya could be achieved through selection of simple traits like earliness to flowering and determinate growth habit. This is because in legumes, determinate growth habit is associated with earliness to flowering and maturity. In countries like India, selection for these two traits has resulted in development of varieties which have become popular among farmers in areas with short season environments (Keerthi, 2014). In pigeon peas, ICRISAT developed some super early varieties through selection for early flowering (< 49 days) and determinate lines (Vales *et al.,* 2012). In several other legume crops, varieties with decreased plant biomass and with enhanced photosynthesates partitioning to the pods have been developed through selection for determinate growth habit (Kwak *et al*., 2012).

The understanding of the genetics of traits with economic importance like flowering time and growth habit is important for formulation of efficient breeding strategy for developing early maturing genotypes. On the other hand, there are other traits like flower colour and presence or absence of anthocyanin on the stem which are of no economic importance to lablab but are useful markers in germplasm characterization, maintenance and in identifying true hybrid in crossing program. The information on the genetics of such traits is therefore equally important for development of suitable strategies for combining these traits into improved cultivars.

The information on genetics of growth habit in lablab is limited and conflicting. For instance, Keerthi *et al*., (2014) reported that growth habit is controlled by three genes, GH1, GH2 and GH3 of which, one (GH1) is independent and the other two (GH2 and GH3) are complementary. However, in Bengalulu India, Keerthi *et al*., (2016) reported that growth habit is controlled by two genes which exhibit classical complementary epistasis. Peeta *et al*., (2017) reported that determinate bush growth habit in lablab is controlled by single recessive gene. Keerthi *et al.,* (2016) reported photoperiod induced sensitivity to flowering time in lablab was under the control of a single gene. Notably, all these studies were conducted in Southern Asia and used their local genetic populations. In Kenya, the information of genetics of these traits under the local environment and using local genetic populations is lacking.

6.2.1 Objectives

The objectives of the present study were to:

1. Identify the inheritance pattern of four lablab qualitative characters (growth habit, flower colour, time to flower initiation and stem anthocyanin),

2. Identify linkage relationship of the genes controlling the four qualitative traits and their implication in breeding lablab varieties.

6.3 Materials and methods

6.3.1 Experimental materials.

The materials used in this experiment included farmer's accessions (Kahuro, Kagio, Eldoret), germplasm from the Gene bank of Kenya (GBK 011719, GBK 028663), accessions from International forage, ILRI (ILRI 18611, ILRI 14440), a released variety (DL 1002) and an accession collected from a Nairobi market (KDD). Kahuro and Kagio are black seeded, moderately early maturing accessions collected from central region of Kenya (Table 6.1).

Table 6. 1 : Morphological Descriptors of the Parents Used in Making Crosses in Lablab

	Genotype	Source	Growth habit	Flowering time	Flower colour	Stem pigmentation
1	GBK 028663	Gene bank	ID	Late	Purple	Present
2	DL 1002	Variety	ID	Late	Purple	Present
3	Kahuro	Farmers collection	ID	Late	Purple	Present
4	Kagio	Farmers collection	ID	Late	Purple	Present
5	Eldoret	Farmers collection	ID	Late	Purple	Present
6	GBK 011719	Gene bank	ID	Late	Purple	Present
7	ILRI 18611	ILRI	ID	Late	Purple	Present
8	ILRI 14440	ILRI	ID	Late	Purple	Present
9	KDD	Local market	D	Early	White	Absent

ID- indeterminate growth habit, D- determinate growth habit

The genotype Eldoret is a late maturing high yielding accession from western Kenya. Accessions GBK 028663 was collected from eastern Kenya, has variegated seed colour and is a high yielder. Accession ILRI 18611 is an aggressive climber with purple coloured stem and originated from USA. The accession ILRI 14440 is an aggressive climber belonging to the sub species unicinatus. The genotype KDD has a determinate

growth habit, white flowers, cream seed coat colour and the origin is a Nairobi city market.

6.3.2 Generation of population

Eight genotypes DL1002, GBK 028663, Kahuro, Eldoret, GBK 011719, ILRI 18611, ILRI 14440 and KDD were planted in a screen house at KALRO Thika in October 2015. These parents were contrasting for growth habit (Indeterminate and determinate), flower colour (purple and white), stem anthocyanin (with or without anthocyanin), days to flower initiation (early or late). Planting of the seeds was done at different dates to synchronize the flowering time of the parental genotypes. Mature flower buds of the female parent (which were likely to open within the next 24hrs) were carefully selected. Using a forceps, a cut was made on the ventral side of the bud. The undehiscent anthers of the male parent were carefully removed. Freshly opened flowers of the male parent were harvested and anthers gently rubbed on the stigma of the female parent.

To start with, two genotypes (DL1002 and GBK 028663) were used as male parents while accession KDD as female parent to make the following crosses, KDD x DL1002 and KDD x GBK 028663. The F_1 seeds of the two crosses were harvested separately and planted in a screen house during the long rain season of 2016. The plants were allowed to self-pollinate to generate F_2 seeds which were then harvested separately.

6.3.3 Field evaluation of the F_2 populations

Two F_2 populations (KDD x DL1002 and KDD x GBK 028663) were evaluated at KALRO Thika during the short rain season of 2016. About, 500 F_2 seeds of each of the two populations were planted on 20m by 10m plots at a spacing of 100cm between the rows and 30cm between the plants. The standard crop production and pest control practices were used to raise the plants. Data collection was done on the growth habit at

50% flowering stage, flower colour, presence of anthocyanin colour on the stem at 50% flowering stage and number of days to flower initiation. The data was collected on all the plants. The plants were allowed to grow to maturity to produce F_3 seeds.

6.3.4 Field evaluation of F_3 families

To confirm the inheritance of the qualitative traits, 200 F_3 families of KDD x DL1002 and KDD x GBK 028663 each consisting of 20 plants were grown at KALRO Thika field during the long rain season of 2017. Data on the growth habit, flower colour and stem anthocyanin were recorded. The data was recorded on the number of plants with a particular trait (for instance purple flower colour) per family row.

6.3.5 Inheritance of growth habit at different growing altitude

In the course of seed multiplication of our lablab breeding lines at KALRO Kiboko, we observed that some lines that had consistently expressed determinate growth habit at KALRO Thika were showing indeterminate growth habit at this site. KALRO Thika and Kiboko are located in more 200Km away from each other. KALRO Thika is situated at an attitude of 1586 M.a.s.l, the soils are nitosols, mean annual rainfall of 1000mm and temperature of 14^0 C to 25^0 C. KALRO Kiboko is situated at Makueni County, at an attitude of 900 M.a.s.l, latitude of 2° 20' South, longitude of 37° 40' East, mean annual rainfall of 464mm and temperature of 23.5^0 C to 29.4^0 C. To further understand the inheritance pattern of growth habit in lablab under different altitudes, we developed six (KDD x Kahuro, KDD x Kagio, KDD x Eldoret, KDD x GBK 011719, KDD x ILRI 18611 and KDD x ILRI 14440) populations segregating for growth habit. In all the crosses, the determinate genotype KDD was used as the female parent. The F_1 seeds were harvested separately and planted in a screen house at KALRO Thika during the long rain season of 2017. The plants were allowed to self-

pollinate to generate F_2 seeds. The F_2 seeds were harvested separately and subdivided into two portions. One portion was planted at KALRO Thika while the other was planted at KALRO Kiboko on 29th November 2018. About 180 F_2 seeds of each population was planted at a spacing of 100cm by 30cm in each location. The crop at Kiboko was given supplemental irrigation at third and fourth week after planting. Data was collected on the growth habit of each individual F_2 plants at 50% flowering and at physiological maturity growth stages. The temperature and rainfall recorded in the two sites during the experiment are presented in Table 6.2 below.

Table 6. 2: Rainfall, humidity and temperature recorded at KALRO Kiboko and Thika during the short rain season of 2018

Site	Environment attribute	Nov. 2018	Dec. 2018	Jan. 2019	Feb. 2019	Mean
Kiboko	Rainfall total (mm)	171	323	44	0.5	
	Humidity	82.5	85.5	89.2	82.8	
	Temperature					
	Maximum (^{0}C)	31.5	29.9	31.4	33.1	31.4
	Minimum (^{0}C)	18.0	18.0	17.4	17.1	17.6
Thika	Rainfall total (mm)	63.2	247.6	19.3	0.6	
	Humidity mean (%)	64.7	71.6	62.4	58.1	
	Temperature					
	Maximum (^{0}C)	26.5	26.1	26.6	26.9	26.5
	Minimum (^{0}C)	14	14	14	14	14.0

6.3.6 Data collection

For the crop planted during the short rain season of 2016, data was collected on growth habit (determinate or indeterminate) at flowering stage, flower colour (purple or white), stem anthocyanin colour (present or absent) on the stem at flowering stage and flowering time (Days from planting to when flowers start to open). The genotypes which formed flowers by 55 days of planting at KALRO Thika field were considered to be early while the others were considered late flowering. The duration of 55 days

was set because it was the maximum time that the early maturing parent KDD took to attain flowering under this environment.The above data was recorded for every individual F_2 plants derived from both KDD x DL1002 and KDD x GBK 028663 populations. Data on the number F_3 families breeding true and those segregating for flower colour, growth habit, presence of anthocyanin on the stem and flowering time were recorded. This was done by counting the number of plants for instance with purple flowers and white flowers in each F_3 family. Those families whose plants were exhibiting both forms of the two traits were considered segregating while those with only one form of the trait were considered to be breeding true. For the F_2 populations planted at Thika and Kiboko sites in 2018, data was collected on the number of determinate plants (those plants whose main stem axis terminated into a flower raceme) and indeterminate plants (those plants whose main stem axis remained vegetative even after the onset of flowering).

6.3.7 Statistical data analysis

A chi square analysis to determine the goodness of fit of the observed ratio of the indeterminate vs determinate growth habit, white vs purple flowers, stem anthocyanin vs without anthocyanin and early vs late flowering plants in the F_2 populations with the expected ratio of one, two and three gene model was tested. The expected ratios for one gene model tested for F_2 generation was 1:3. The ratio tested for the two genes model were 1:15, 3:13 and 9:7. Genetic ratio of 57:7 and 1:63 were also tested to determine whether the traits were under the control of three genes. The genetic ratio tested on the F_3 families, were 1:2:1 for one gene model, 1:8:7 and 7:6:3 for two gene model.

A chi square test was done according to Mather, (1951) to determine the possibility of linkage of genes for growth habit, flower colour, stem anthocyanin and flowering time. The frequency of recombination (r) of the linked genes were determined using product-ratio method as described by Kuspira & Bhambhani, (1984).

6.4 Results

6.4.1 Genetics of growth habit

Genetics of growth habit in lablab was studied in eight crosses in three separate experiments. The first experiment involved two crosses, KDD x GBK 028663 and KDD x DL1002. All the F_1 plants in both crosses were indeterminate (Table 6.3). The F_2 plants of KDD x GBK 028663 segregated into 338 indeterminate: 113 determinate and those of KDD x DL1002 segregated into 267 indeterminate: 95 determinate. The segregation of the F_2 plants in both KDD x GBK 028663 and KDD x DL1002 were in good fit with the ratio of 3 indeterminate: 1 determinate stem habits ($\chi^2 = 0$, $P = 1$ and $\chi^2 = 0.22$, $P = 0.639$) respectively. The F_3 families derived from F_2 plants in both KDD x GBK 028663 and KDD x DL1002 segregated into ratio of 1 family homozygous for indeterminate: 2 families segregating for both indeterminate and determinate: 1 family segregating for determinate growth habit.

Table 6.3: Segregation for growth habit in F_2 and F_3 of lablab crosses involving indeterminate and determinate parents at KALRO Thika during short rain season of 2016

Cross	Observed number of F_2 plants/F_3 fam.					Expected number of F_2 plants/F_3 fam.					
	Parental /Gen	Total	ID	SEG	D	Exp. ratio	ID	SEG	D	X^2 stats	P value
KDD x GBK 028663	KDD	15	0	-	15	-	-	-	-	-	-
	GBK 028663	15	15	-	0	-	-	-	-	-	-
	F_1	15	15	-	0	-	-	-	-	-	-
	F_2	451	338	-	113	3:1	338	-	113	0	1.000
	F_2	451	338	-	113	9:7	254	-	197	65.58	0.0001
	F_2	451	338	-	113	15:1	423	-	28	275.11	0.0001
	F_2	451	338	-	113	13:3	366	-	85	11.36	0.0007
	F_2	451	338	-	113	57:7	402	-	49	54.97	0.0001
	F_3	108	33	54	21	1:2:1	27	54	27	2.66	0.2644
KDD x DL1002	KDD	15	0	-	15	-	-	-	-	-	-
	DL1002	15	15	-	0	-	-	-	-	-	-
	F_1	15	15	-	0	-	-	-	-	-	-
	F_2	362	267	-	95	3:1	271	-	91	0.22	0.6390
	F_2	362	267	-	95	9:7	204	-	158	44.57	0.0001
	F_2	362	267	-	95	15:1	339	-	23	240.68	0.0001
	F_2	362	267	-	95	13:3	294	-	68	13.19	0.0002
	F_2	362	267	-	95	57:7	322	-	40	85.01	0.0001
	F_3	188	48	100	40	1:2:1	47	92	47	0.737	0.6917

The inheritance of growth habit in lablab was further studied in six F_2 crosses in two environments (Thika and Kiboko) which are located in different altitudes. One of the crosses planted (KDD x ILRI 14440) had poor germination at Thika site and was therefore not included in the analysis. The F_1 plants from all the other five crosses (KDD x GBK 011719, KDD x ILRI 18611, KDD x Eldoret, KDD x Kahuro, KDD x Kagio) at Thika site were indeterminate (Table 6.4). The F_2 plants derived from all the five crosses segregated to the expected ratio of 3 indeterminate: 1 determinate.

Table 6. 4: Test of significance of segregation for genes controlling GH in F2 generations in lablab at KALRO Thika during Short rains season of 2018

Cross	Observed number of F_2 plants/F_3 fam.					Expected number of F_2 plants/F_3 fam.					
	Parental /Gen	Total	ID	SEG	D	Exp. ratio	ID	SEG	D	X^2 stats	P value
KDD x GBK 011719	KDD	10	0	-	10	-	-	-	-	-	-
	GBK 011719	10	10	-	0	-	-	-	-	-	-
	F_1	10	10	-	0	-	-	-	-	-	-
	F_2	66	51	-	15	3:1	49	-	17	0.31	0.577
KDD x ILRI 18611	KDD	10	0	-	10	-	-	-	-	-	-
	ILRI 18611	10	10	-	0	-	-	-	-	-	-
	F_1	10	10	-	0	-	-	-	-	-	-
	F_2	88	70	-	18	3:1	66	-	22	0.96	0.327
KDD x Eldoret	KDD	10	0	-	10	-	-	-	-	-	-
	Eldoret	10	10	-	0	-	-	-	-	-	-
	F_1	10	10	-	0						
	F_2	105	84	-	21	3:1	79	-	26	1.27	0.259
KDD x Kahuro	KDD	10	0	-	10	-	-	-	-	-	-
	Kahuro	10	10	-	0	-	-	-	-	-	-
	F_1	10	10	-	0	-	-	-	-	-	-
	F_2	193	149	-	44	3:1	145	-	48	0.44	0.500
KDD x Kagio	KDD	10	0	-	10	-	-	-	-	-	-
	Kagio	10	10	-	0	-	-	-	-	-	-
	F_1	10	10	-	0	-	-	-	-	-	-
	F_2	109	84		25	3:1	81	-	28	0.43	0.511

In Kiboko site, all F_1 plants from the five crosses (KDD x GBK 011719, KDD x ILRI 18611, KDD x Eldoret, KDD x Kagio and KDD x ILRI 14440) had indeterminate growth habit. All the F_2 plants derived from the five crosses showed goodness of fit to a ratio of 57 indeterminate: 7 determinate (Table 6.5).

Table 6. 5: Segregation ratio for genes controlling growth habit in F_2 generations in lablab at KALRO Kiboko during Short rains season of 2018

Cross	Observed number of F_2 plants/F_3 fam.					Expected number of F_2 plants/F_3 fam.					
	Parental /Gen	Total	ID	SEG	D	Exp. ratio	ID	SEG	D	X^2 statis	P value
KDD x GBK 011719	KDD	10	0	-	10	-	-	-	-	-	-
	GBK 011719	10	10	-	0	-	-	-	-	-	-
	F_1	10	10	-	0						
	F_2	109	99	-	10	3:1	82	-	27	14.22	0.0001
	F_2	109	99	-	10	13:3	89	-	20	6.12	0.013
	F_2	109	99	-	10	57:7	97	-	12	0.37	0.543
KDD x ILRI 18611	KDD	10	0	-	10	-	-	-	-	-	-
	ILRI 18611	10	10	-	0	-	-	-	-	-	-
	F_1	10	10	-	0						
	F_2	160	144	-	16	3:1	120	-	40	15.20	0.0001
	F_2	160	144	-	16	13:3	111	-	49	32.02	0.0001
	F_2	160	144	-	16	57:7	143	-	17	0.065	0.79
KDD x Eldoret	KDD	10	0	-	10	-	-	-	-	-	-
	Eldoret	10	10	-	0	-	-	-	-	-	-
	F_1	10	10	-	0						
	F_2	125	112	-	13	3:1	94	-	31	13.89	0.0001
	F_2	125	112	-	13	13:3	102	-	20	8.67	0.003
	F_2	125	112	-	13	57:7	111	-	14	0.08	0.77
KDD x Kagio	KDD	10	0	-	10	-	-	-	-	-	-
	Kagio	10	10	-	0	-	-	-	-	-	-
	F_1	10	10	-	0	-	-	-	-	-	-
	F_2	132	116	-	16	3:1	99	-	33	11.66	0.0006
	F_2	132	116	-	16	13:3	107	-	25	17.81	0.0001
	F_2	132	116	-	16	57:7	118	-	14	0.31	0.577
KD1 x ILRI 14440	KDD	10	0	-	10	-	-	-	-	-	-
	ILRI 14440	10	10	-	0	-	-	-	-	-	-
	F_1	10	10	-	0	-	-	-	-	-	-
	F_2	96	87	-	9	3:1	72		24	12.5	0.004
	F_2	96	87	-	9	13:3	78		18	5.53	0.018
	F_2	96	87	-	9	57:7	86		10	0.11	0.74

6.4.2 Flowering time

The days to flower in F_1 plants of both crosses was much later than the early maturing parent (KDD) and similar to the late flowering parents (GBK 028663, DL1002). The days to flowering for the F_2 plants showed some normal distribution in both crosses with majority of the plants flowering between 60-67 days after planting (Figure 6.1). The F_2 were classified into two groups, early maturing ($\leq$ 55 days) and late maturing ($\geq$

56 days). The duration of 55 days was set because it was the maximum time that the early maturing parent KDD took to attain flowering under this environment. There were many transgressive segregants on the late flowering side of distribution.

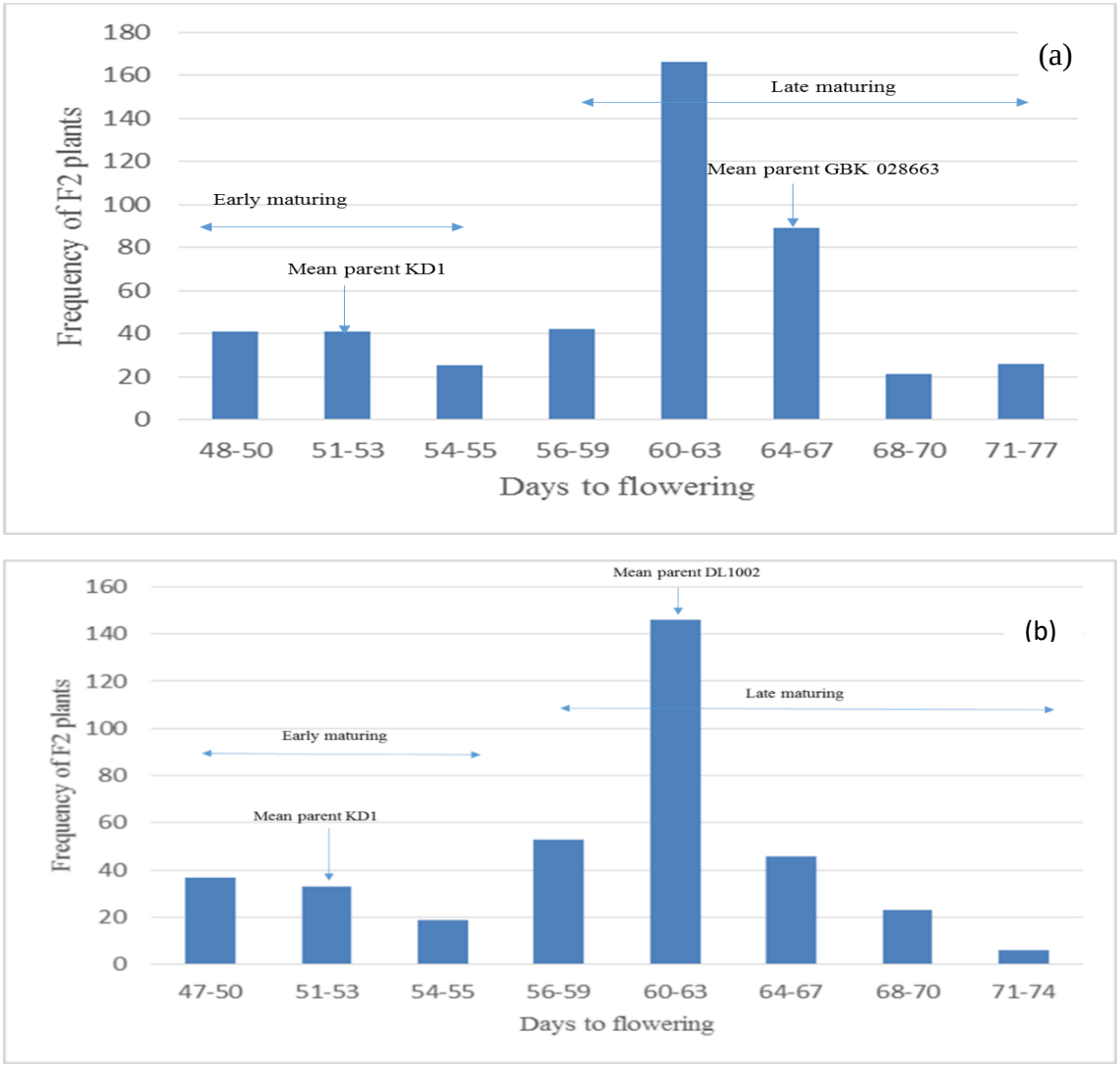


Figure 6. 1: Frequency distribution of days to flower in parents and (a) KDD x 028663 F_2 plants (b) KDD x DL1002 F_2 plants at Thika, Kenya

The F_2 plants of KDD x GBK 028663 segregated into 343 late: 108 early and those of KDD x DL1002 segregated into 274 late: 88 early (Table 6.7). The segregation of the F_2 plants in both KDD x GBK 028663 and KDD x DL1002 were in good fit with the

ratio of 3 late flowering: 1 early flowering (χ^2 = 0.29, *P* = 0.59 and χ^2 = 0.13, *P* = 0.717) respectively.

Table 6. 6: Segregation ratio for genes controlling flowering time in F_2 generation in lablab at KALRO Thika during short rain season of 2016

Cross	Observed number of F_2 plants.					Expected number of F_2 plants.					
	Parental /Generation	Total	Late	SEG	early	Expected ratio	late	SEG	early	X^2 stat	P value
KDD x GBK 028663	KDD	20	0	-	20	-	-	-	-	-	-
	GBK 028663	20	20	-	0	-	-	-	-	-	-
	F_1	14	14	-	0	-	-	-	-	-	-
	F_2	451	343	-	108	3:1	338	-	113	0.29	0.5900
	F_2	451	343	-	108	9:7	254	-	197	71.38	0.0001
	F_2	451	343	-	108	15:1	423	-	28	243.7	0.0001
	F_2	451	343	-	108	13:3	366	-	85	7.66	0.005
	F_2	451	343	-	108	57:7	402	-	49	79.69	0.0001
KDD x DL1002	KDD	15	0	-	15	-	-	-	-	-	-
	DL1002	15	15	-	0	-	-	-	-	-	-
	F_1	15	15	-	0						
	F_2	362	274	-	88	3:1	271	-	91	0.13	0.717
	F_2	362	274	-	88	9:7	204	-	158	55.02	0.0001
	F_2	362	274	-	88	15:1	339	-	23	196.1	0.0001
	F_2	362	274	-	88	13:3	294	-	68	7.24	0.007
	F_2	362	274	-	88	57:7	322	-	40	64.75	0.0001

6.4.3 Flower colour

Two crosses involving a white flowered parent (KDD) and purple flowered parents (GBK 028663 and DL1002) were used to study the inheritance of flower colour in lablab. All the F_1 progenies of both crosses produced purple flowers. This indicate the dominance of gene (s) controlling purple flowers over the white flowers. In KDD x GBK 028663, the F_2 plants segregated into 318 purple: 125 white, while those in KDD x DL1002 into 261 purple: 98 white (Table 6.7). These numbers are in good fit with the ratio of 3 purple: 1 white suggesting that flower colour in lablab could be under control of one gene.

Table 6. 7: Segregation ratio of genes controlling Flower colour in F_2 generation in lablab at KALRO Thika during the short rains of 2016

Cross	Observed number of F_2 plants					Expected number of F_2 plants/F_3 fam.					
	Parental /Gen	**Total**	**Purple**	**SEG**	**White**	**Expect. ratio**	**Purple**	**SEG**	**White**	**χ^2 stats**	**P value**
KDD x GBK 028663	KDD	15	0	-	15	-	-	-	-	-	-
	GBK 028663	15	15	-	0	-	-	-	-	-	-
	F_1	15	15	-	0						
	F_2	443	318	-	125	3:1	332	-	111	1.69	0.193
	F_2	443	318	-	125	13:3	360	-	83	26.1	0.0001
	F_2	443	318	-	125	57:7	395	-	48	138.5	0.0001
KDD x DL1002	KDD	15	0	-	15	-	-	-	-	-	-
	DL1002	15	15	-	0	-	-	-	-	-	-
	F_1	15	15	-	0						
	F_2	359	261	-	98	3:1	269		90	0.94	0.332
	F_2	359	261	-	98	13:3	292		67	17.63	0.0001
	F_2	359	261	-	98	57:7	320		39	100.12	0.0001

6.4.4 Stem pigmentation

All the F_1 plants from the two crosses (KDD x GBK 028663 and KDD x DL1002) produced stems with purple pigmentation. This suggests that presence of purple pigmentation on the stem of lablab was dominant to the absence of the pigmentation on stem. The F_2 population of KDD x GBK 028663 segregated into 236 plants with stem anthocyanin: 211 plants without anthocyanin, while that of KDD x DL1002 segregated into 205 plants with stem anthocyanin: 156 plants without anthocyanin (Table 6.8).

Table 6.8: Segregation ratio of genes controlling stem pigmentation in F_2 and F_3 generations in lablab at KALRO Thika during the short rains of 2016

Cross	Observed number of F_2 plants/F_3 fam.					Expected number of F_2 plants/F_3 fam.					
	Parental /Gen	Total	Presen. Anthoc.	Seg	Absen. anthoc.	Expect. ratio	Presen. anthoc	Seg	Absen. anthoc	χ^2 stats	P value
KDD x GBK 028663	KDD	14	0	-	14	-	-	-	-	-	-
	GBK 028663	15	15	-	0	-	-	-	-	-	-
	F_1	15	15	-	0	-	-	-	-	-	-
	F_2	447	236	-	211	3:1	335	-	112	116.7	0.0001
	F_2	447	236	-	211	9:7	251	-	196	2.03	0.154
	F_2	447	236	-	211	15:1	419	-	28	1275	0.0001
	F_2	447	236	-	211	13:3	363	-	84	236.4	0.0001
	F_2	447	236	-	211	57:7	398	-	49	601.5	0.0001
	F_3	182	15	96	71	1:8:7	11	91	79	2.53	0.282
KDD x DL1002	KDD	12	0	-	12	-	-	-	-	-	-
	DL1002	14	14	-	0	-	-	-	-	-	-
	F_1	15	15	-	0						
	F_2	361	205	-	156	3:1	271	-	90	64.47	0.0001
	F_2	361	205	-	156	9:7	203	-	158	0.44	0.5071
	F_2	361	205	-	156	15:1	338	-	23	821.4	0.0001
	F_2	361	205	-	156	13:3	293	-	68	140.3	0.0001
	F_2	361	205	-	156	57:7	322	-	39	393.5	0.0001
	F_3	106	9	54	43	1:8:7	7	53	46	0.78	0.677

The numbers observed in the two crosses were in good fit with the ratio of 9:7 at χ^2 = 2.03, P= 0.15 and χ^2 =0.44, P=0.50 respectively. The results were further confirmed by the segregation ratio of the F_3 families which were derived from the F_2 population. The F_3 families from KDD x GBK 028663 segregated into; all 15 families with stem anthocyanin: 96 families segregating for both stem anthocyanin and absence of anthocyanin: 71families without stem anthocyanin. In the cross KDD x DL 1002 the F_3 families segregated into 9 families with stem pigmentation only: 54 families with presence and absence of pigmentation: 43 families without stem anthocyanin. Again the number observed in the two crosses were in good fit to ratio 1:8:7 with the probability of χ^2 being more than 0.05 in both crosses. This result suggests that the stem anthocyanin in lablab genotypes is under control of two genes with complementary epistasis.

6.4.5 Joint segregation of four qualitative traits

The inheritance of growth habit, flower colour and flowering time traits in the lablab populations studied above indicated that these traits are under single gene control. However, the inheritance of stem anthocyanin was revealed to be under control of two complementary gene. Analysis of joint segregation was undertaken to identify the possible linkage among these four characters. The segregation ratio of growth habit and flower time traits observed from the two crosses KDD x GBK 028663 and KDD x DL1002 showed some significant departure from the expected ratio of 9 indeterminate late: 3 indeterminate early: 3 determinate late: 1 determinate early (Table 6.10). This suggests that the two traits are linked. The recombination fraction and mapping distance for the two traits was 0.05 and 0.14, 7.24cM and 14.07cM for KDDx GBK 028663 and KDD x DL1002 respectively. In the cross KDD x GBK 028663, the observed joint segregation ratio of growth habit and stem anthocyanin indicated that the two traits are linked. However the joint segregation of the two traits in a cross between KDD and DL1002 showed that the two traits were unlinked. This study further revealed that the characters growth habit and flower colour, flower colour and time to flowering, time to flowering and stem anthocyanin were unlinked in both crosses. The mapping distance between these traits was more than 27cM in both crosses suggesting that the genes controlling the pair of traits were distant apart.

Table 6.9: Joint segregation of four qualitative traits in two F_2 populations in lablab

Cross	Number of observation/ plants										
	Loci	+/+	+/-	-/+	-/-	Ratio tested	χ^2	P value	Link. status	r	Dist.(cM)
KDD x GBK 028663	GG:TT	338	10	6	97	9:3:3:1	339.13	0.000	Linked	0.05± 0.007	7.24
	GG:FF	245	91	69	38	9:3:3:1	6.76	0.07	unlinked	0.44 ± 0.02	49.71
	GG:SS	190	148	46	63	27:21:9:7	9.31	0.02	Linked	0.42 ± 0.02	24.91
	FF:TT	242	72	96	33	9:3:3:1	3.36	0.33	Unlinked	0.47 ± 0.022	45.41
	TT:SS	187	153	49	58	27:21:9:7	5.97	0.11	Unlinked	0.19 ± 0.03	27.95
KDD x DL1002	GG:TT	246	22	28	67	9:3:3:1	147.44	0.0001	Linked	0.14 ± 0.03	14.07
	GG:FF	195	72	66	26	9:3:3:1	1.34	0.719	Unlinked	0.50 ± 0.03	46.17
	GG:SS	158	110	45	48	27:21:9:7	3.07	0.381	Unlinked	0.18 ± 0.01	27.07
	TT:SS	158	115	45	43	27:21:9:7	1.225	0.74	Unlinked	0.21 ± 0.016	31.44

Homologous recessive and homologous dominant are denoted -/- and +/+ are whereas -/+ and +/- are heterozygous. GG denotes growth habit, TT- flowering time, SS- stem anthocyanin and FF- flower colour. r- denotes the recombination fraction.

6.5 Discussion

The findings from this study showed that segregation of growth habit in lablab under Thika environment fitted to the expected ratio of 3:1 while the segregation fitted to a ratio of 57:7 under Kiboko environment. All the F_1 plants were indeterminate suggesting that indeterminate is dominant to determinate. The segregation ratio of F_2 populations derived from seven crosses and evaluated under Thika environment suggested that growth habit was under monogenic control but when the same populations were evaluated at Kiboko the segregation ratio suggested that the trait was under control of three genes.

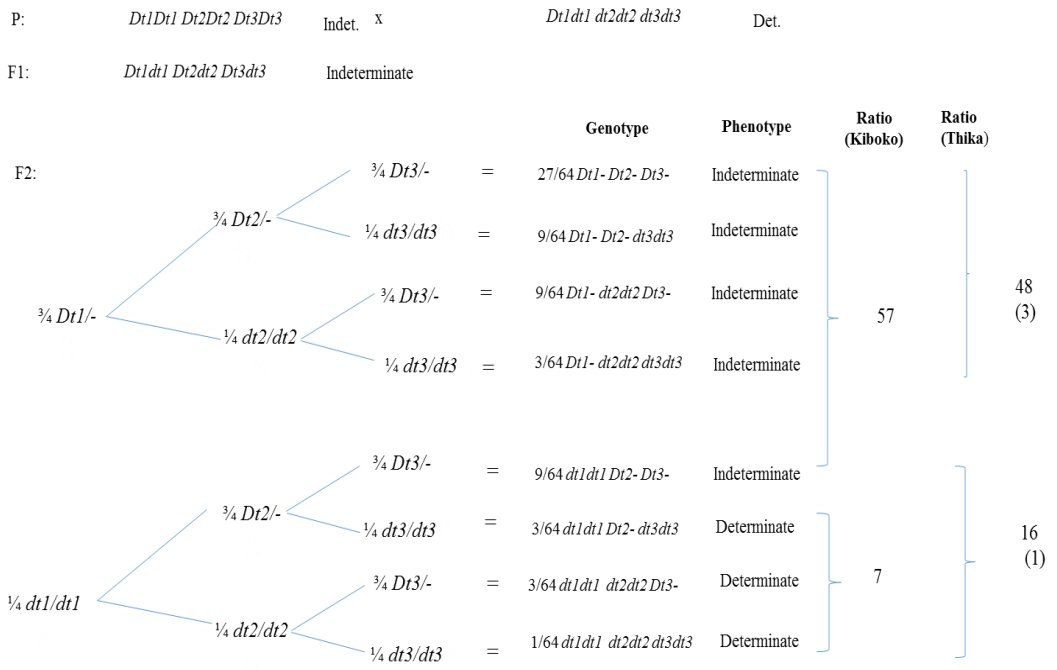


Figure 6. 2: Genotypes, phenotypes and expected ratio tested for goodness of fit in F2 generation segregating for growth habit in lablab at Kiboko and Thika sites

From this results, we propose that growth habit in lablab is controlled by three epistatic genes. We designate the three genes as *Dt1/dt1*, *Dt2/dt2* and *Dt3/dt3*. Dt1/dt1 is the basic gene while the other two (*Dt2/dt2*, *Dt3/dt3*) are complementary to the first one. The *Dt1* allele either in homozygous or heterozygous state (*Dt1- Dt2- Dt3-*, *Dt1- Dt2- dt3dt3*, *Dt1- dt2dt2 Dt3-*, *Dt1- dt2dt2 dt3dt3*) regardless of the state of the other two alleles (*Dt2/dt2 and Dt3/dt3*) produce indeterminate growth habit (Figure 6.2).

When the allele *dt1* is in homozygous recessive state and at least one of the other alleles was also in homozygous recessive state (*dt1dt1 Dt2- dt3dt3*, *dt1dt1 dt2dt2 Dt3-*, *dt1dt1 dt2dt2 dt3dt3*) determinate growth habit was produced. However, when the allele dt1 was in homozygous recessive state and both of the other alleles are either in homozygous dominant or heterozygous state produce indeterminate growth habit (*dt1dt1 Dt2- Dt3-*) giving a ratio of 57:7. Under Thika environment, the gene Dt2 and

Dt3 are inactive thereby reducing the number of indeterminate plants to give a segregation ratio of 3 indeterminate: 1 determinate. Under Kiboko environment, the two genes Dt2 and Dt3 are activated, thereby increasing the number of indeterminate plants and modifying the ratio to 57:7. This suggests that the number of plants expressing determinate growth habit are determined by the presence of homozygous recessive gene *dt1dt1* in first locus but is modified by the presence of homozygous dominant or heterozygous gene in the second and third loci. The expression of the second and third genes in control of growth habit in lablab is environment dependent. In our case, Kiboko and Thika site are located within the latitude of $0\text{-}2^0$ south of equator and therefore experience similar photoperiod of about 12 hours. Kiboko site experienced higher day and night mean temperature (31.4^0C and 17.6^0C) compared to Thika site (26.5^0C and 14^0C) during the time of the experiment. Temperature was therefore the main environmental factor distinguishing the two sites. Therefore, we put forward that the growth habit in lablab is controlled by three genes, one basic and two complementary genes which could be temperature dependent. At high temperature (could be above 30^0C) the two gene are activated while lower temperature deactivate the genes. In their study in soyabean, Inouye *et al*., (1979) also reported that some plants shifted from determinate growth habit at 20^0C to indeterminate habit at 30^0C - 35^0C.

Other studies have reported that growth habit in lablab is under control of three genes. For instance, Keerthi *et al*., (2014) studied two lablab crosses (HA 4 × GL 103 and HA 4 × GL 37) and reported that growth habit was under control of three genes (*GH*1, *GH*2 and *GH*3) with *GH*1 being independent while *GH*2 and *GH*3 were complementary. Using different genetic background, Girish & Gowda, (2009) also reported three genes model for control of growth habit in Lablab. Our results differed from those of Keerthi *et al* (2016) who reported that growth habit in lablab was under digenic control. The

difference could be due to the different genetic background and environment used in these studies. We suggest that the results obtained here for segregation ratio of growth habit under Kiboko environment to be confirmed in F_3 families. The environmental conditions responsible for shifting determinate growth habit to indeterminate should be further studied in Kenya.

Our results indicated that a single dominant gene was associated with the presence of purple flowers. We observed similar segregation ratio in the F_2 generations of both crosses tested which indicate that the two purple flowered parents GBK 028663 and DL1002 could be having similar flower colour allele. Our results of a monogenic control of flower colour is consistent with that of Keerthi *et al*, (2016). The segregation of stem anthocyanin in F_2 population fitted to a ratio of 9:7 suggesting the trait is under control of two genes with complementary epistasis. The presence of purple pigmentation in all the F_1 plants of both crosses indicated that stem anthocyanin is dominant to the absence of the pigmentation. A plant could only express stem anthocyanin if it had dominant alleles at the two loci. Any other combination of alleles in the two loci resulted in inhibition of expression of pigmentation on the stem of the lablab plants. Harland (1920) in lablab, Phippen, (2000) in *Ocimum basilicum* and Ghose *et al*., (1963) in rice also reported digenic control of stem anthocyanin. In lablab, anthocyanin was observed on either the entire stem or localized at the stem nodes and in the veins under the leaves. The understanding of inheritance of anthocyanin colouration on foliage and flower colour generated in this study is important for lablab breeding. For instance, when designing crosses, breeders may wish to select genotypes without anthocyanin pigmentation on the foliage as the female parent while those with anthocyanin as the male parent. With the understanding that anthocyanin pigmentation is completely dominant to absence of the pigmentation, breeders can use the trait as

morphological marker to identify the true F_1 hybrids and remove the selfed plants even at seedling stage.

In this study, the F_2 plants obtained from the crosses between the early flowering accession (KDD) and late flowering genotypes (GBK 028663 & DL1002) fitted well into two distinct classes (late and early). This is probably because of the large difference for days to flowering (> 11 days) between the parents. The segregation of the F_2 progenies in both crosses revealed a monogenic inheritance control for days to flowering with the early flowering being recessive to late flowering. Keerthi *et al.,* (2014) and Keerthi *et al*., (2016) have also reported monogenic control of flowering time in lablab. In common beans, recessive allele on locus *Ppd* which is on linkage group one (LG1) has been reported to control photoperiod response and early flowering under long days condition (Koinange *et al.,*1996). Four non-allelic genes (*ef-1, ef-2, ef-3, ef-4*) have been identified to control flowering time in chickpea (Gaur *et al.,* 2015). A major recessive gene 'sn' was identified to be responsible for early flowering in lentil variety *Precoz* (Sarker & Erskine, 2006). Murfet, (1973) indicated that duration to flowering in *Pisum* was under control of two genes (S_2 & E) with the earliness being recessive.

Our study also revealed a continuous distribution of days to flowering of the F_2 progenies of both crosses suggesting that polygenic system is also in control of days to flowering in lablab. This indicate that flowering time could be under both major and minor genes. Flowering time loci has been identified in many legumes. In many cases the flowering loci have been identified as QTL, but in other cases as major loci (Weller & Ortega, 2015). For instance, a major gene for flowering time *ef-1* was initially identified from the desi kabuli chickpea cultivar ICCV 2. However, using the same

parent, one QTL for days to flowering was identified (Cho *et al*., 2002) suggesting that in some cases a QTL may match up to a major loci.

Our study did not identify transgressive segregants for earliness to flowering. In other words, crossing the early flowering parent KDD with late flowering parents GBK 028663 and DL1002 did not produce recombinant progenies with lesser days to flowering than parent KDD. This suggest that the late flowering parents did not contribute gene for earliness during recombination to produce transgressive segregants for earliness to flowering. Other sources of genes for earliness to flowering in lablab need to be identified, crossed with KDD and selection for early flowering transgressive segregants done. Induced mutagenesis can also be used to reduce the flowering time in the accession KDD.

This investigation did not evaluate the F_3 families to confirm the segregation ratio of flowering time to one gene model. In addition, the evaluation of the F_2 progenies in this study was only done in one site and therefore could not account for the effect of interaction of flowering time genes with the environment. This is because transition from vegetative to flowering stage in legumes is affected by environmental cues especially the photoperiod and temperature (Weller and Ortega, 2015). Therefore, we recommend validation of our reported monogenic gene control of flowering time in diverse genetic backgrounds and multiple environments. We also recommend that the early maturing parent KDD to be used in generating mapping population for tagging the gene(s) for time to flowering using molecular markers. This will facilitate the application of the environmental invulnerable molecular markers in the selection of earliness to flowering in lablab.

Joint segregation of pairs of qualitative traits (growth habit, time to flowering, stem anthocyanin and flower colour) was conducted to test for deviations from various

independent assortment ratios. Most of the character pairs did not deviate from the tested ratios suggesting that the traits are unlinked. In both crosses, distorted segregation from the expected ratio of 9:3:3:1 was observed between growth habit and time to flowering suggesting possible linkage of the trait pairs. In other words, the gene governing growth habit and flowering time in lablab could be residing in the same chromosome. The chances of crossing over and recombination to take place between two genes is determined by their location on the chromosome. The genes that are closely located will have a smaller chance of crossing over. In this study, the recombination fraction which is an indicator of possible number of synapsis between the traits was small (0.05 and 0.14) for both crosses indicating possible linkage. The short (less than 15cM) mapping distance estimated here between the two characters further supports a possible linkage. Linkage of growth habit and flowering time has also been reported by Keerthi *et al.*, (2014) and Keerthi *et al*., (2016) in lablab. However, the recombination fraction and mapping distance reported in the two studies were different from the one obtained in this study probably because of the different genetic background and environment used. In common beans, flowering loci PvTFL1y and PvTFL1z were mapped near determinacy loci (Kwak *et al*., 2008) indicating the two characters are linked. Erickson, (1992) also reported linkage between growth habit and flowering time in lima bean.

Through plant breeding, it's possible to accumulate desirable alleles of important morphological characters into one cultivar with enhanced adaptability to environment. This strategy can be feasible for improvement of lablab in Kenya where production is constrained by lack of acceptable cultivars of determinate growth habit, early maturity and high yielding.

In the present study, we have established that growth habit locus is linked to that of flowering time. This implies that lablab breeders can indirectly select for early maturing progenies using the more easily observable determinate growth habit trait. The parent KDD used in this study could be a good source of gene for earliness to flowering and determinate growth habit in Kenya. Integrating this accession into the breeding program can improve the local cultivars for earliness to maturity. In the present investigation, we observed progenies of early flowering determinate F_2 plants with purple flowers suggesting that it's possible to combine determinate growth habit and other important traits in lablab. Determinacy is useful both under conditions of excessive vegetative growth and severe drought (Hegde, 2011). Integrating genes for determinacy into lablab cultivars may increase and stabilize lablab yields under intercrop system environments where indeterminate types produces excessive vegetative growth thereby lodging the intercrop plant resulting in reduced yield (Kinyua and Kiplagat, 2012).

Lablab production in Kenya is carried out in wide agro-ecological zones with ranging environmental conditions such as temperature, soils and moisture (Kamotho *et al*., 2015). Breeders should therefore target to breed and release cultivars with wide adaptability to increase chances of varietal adoption by many farmers. The present investigation has observed different segregation ratio for growth habit of F_2 progenies grown in Kiboko and Thika sites and concluded that the discrepancies was temperature dependent. The implication of this is that some varieties selected for determinate growth habit in areas with low temperatures like Thika may express indeterminate growth habit when produced in high temperature areas like lower eastern and coastal Kenya. This can pose a serious challenge especially during multiplication of breeder's/ nuclear/certified seed under KEPHIS inspection and can result to rejection of the crop on the basis of non-conformity to the descriptor. We therefore recommend that selection

for determinate growth habit to be done in early generations and under high temperature conditions (> 30^0C). Such environment is useful to select against those progenies which would express as determinate under low temperature but show indeterminate growth habit under high temperature.

Growth habit in lablab is controlled by three genes, one basic and two complementary genes which are temperature dependent. Flower colour and time to flowering characters are each under monogenic control while stem anthocyanin pigmentation is under control of two complementary genes. Joint segregation of gene controlling growth habit and flowering time revealed linkage between the two genes. There was no evidence of linkage between genes for growth habit and flower colour, growth habit and stem anthocyanin, stem anthocyanin and flowering time and flowering time with flower colour. The parent KDD is a good source of gene for earliness to flowering and determinate growth habit in Kenya and should be integrated in breeding program for development of varieties with better adaptation to growing environments.

CHAPTER SEVEN

Genotypic Variability and Path Analysis of Yield Components of Determinate Lablab Genotypes in Kenya

7.1 Abstract

Lablab is an important food and cash crop in Kenya. However, productivity is low and little research has been done on the crop. Early maturing determinate varieties adapted to emerging growing conditions and cropping systems are lacking. This study was conducted to obtain phenotypic and genetic variability, heritability estimates, expected response to selection and character association of several vegetative and reproductive characteristics. Forty lablab F_5 lines and an accession with determinate growth habit were evaluated at Kenya Agricultural and Livestock Research Organization, Katumani Centre in Machakos and Kandara Centre in Thika during 2017 using a randomized complete block design with two replications. Significant differences ($P< 0.05$) were observed among the genotypes for all the characters. There was wide trait genotypic variation for grain yield and its secondary traits. The phenotypic coefficient of variation ranged from 37.33% (seed yield per plant at Thika) to 2.09% (for days to maturity at Katumani). The genotypic coefficient of variation ranged from 27.99% for grain yield to 3.02% (for days to flowering). Heritability for the characters differed between the two sites and among the 11 traits and were moderate to high. Most of the characters recorded broad sense heritability of >0.5 at Thika and >0.6 at Katumani. There was higher broad sense heritability estimates at Katumani experiment for days to flowering (0.80), days to maturity (0.81), plant height (0.85) and 100SW (0.83) compared to 0.61, 0.40, 0.49 and 0.51 respectively at Thika. The expected responses to selection was highest for seed weight per plant at 48.53% and 37.80% for Thika and Katumani respectively whereas the lowest response was recorded for days to maturity (3.37%) at Thika and days to flowering (5.55%) at Katumani. The maturity related traits (days to 50% flowering and days to maturity) and pod length consistently showed lowest response to selection of less than 12% in both sites despite their high heritability. All the characters were positively and significantly ($P> 0.05$) correlated to seed weight per plant with exception of 100 seed weight and pod length. Pods number ($r = +0.87$) and raceme number ($r = +0.81$) had the highest positive significant association with seed yield per plant. Low but significantly negative correlations was recorded for 100 seed weight and the number of pods per raceme ($r= -0.28$). Pods per plant, 100 seed weight, number of racemes per plant, plant length and plant width, number of flowering nodes, plant height and days to maturity had positive direct effect on seed yield per plant. Pods per plant (0.68), racemes per plant (0.25), pods per raceme (0.13), plant width (0.12) and 100 seed weight (0.11) had the largest direct effect on seed yield. Fairly high indirect positive effects on seed yield through pods per plant was observed for raceme per plant (0.51), pods per raceme (0.28), number of flowering nodes per raceme (0.33) and plant height (0.30). Improvement of determinate lablab varieties through phenotypic selection is possible in pods per plant, raceme per plant, plant height, pod width, pods per raceme, number of flower nodes and seed yield.

7.2 Introduction

In Kenya, Lablab is mainly grown in the semi-arid areas most of which are characterized by decreasing seasonal rainfall (Bosire *et al*., 2019) . Breeding for high yielding and rapidly maturing varieties will help in cushioning farmers from the impacts of climate change such as complete crop failure due to moisture stress. This can be achieved by development of varieties with determinate growth habit. In lablab, the determinate growth habit is characterized by the main axis terminating into inflorescence and has been associated with reduction of flowering and maturity period (Keerthi, 2014; Keerthi *et al*., 2016) which is beneficial in growing conditions with moisture stress. However, determinate early maturing varieties are still lacking in Kenya despite all their potential benefits to the farmers. Conventional introgression of determinate growth habit and earliness to maturity traits into our local cultivars will generate population from where selection of improved varieties can be made.

Subsequent to generating genetic variability, the breeder is required to advance the population through selection. Selection, involves discriminating among the diverse population to identify and pick a number of individuals to form the next generation. A breeder expects to change the population mean of the traits in the next generation positively through the selection of those individuals with high genetic potential. The response to selection (genetic gain or genetic advance) is the change of population mean between generations following selection (Acquaah, 2012). The genetic gain attained through selection depends on the available phenotypic variation, heritability of the trait being selected and the selection pressure imposed by the breeder. Heritability is the proportion of the phenotypic variance that is due to genetic effects. A high heritability is likely to contribute to high response to selection and thereby advancing the population in the desired direction of change (Acquaah, 2012). As the breeder continue

to advance their materials to higher generations the genetic variation and genetic gain from one generation to the next declines, but at the same time, the mean value of the trait being improved increases. Therefore, the understanding of genetic parameters such as heritability and genetic advance estimates are important to predict the gains from selection (Holland *et al*., 2010) . Broad heritability and genetic gain estimates of yield contributing attributes in lablab has been reported by several authors (Venkatesha *et al*., 2016; Parmar *et al*., 2013; Singh *et al*., 2015; Salim *et al*., 2014; Sadak *et al*., 2018). However, the information on heritability and genetic advance of important agronomic traits of local lablab populations under Kenyan conditions is lacking. Heritability and genetic advance of crop traits are affected by the genetic composition of the population and the growing environment (Holland *et al*., 2010).

The understanding of the relationships among important agronomic traits in plant population is crucial for enhanced progress in crop improvement. Complex traits like grain yield are important but difficult to manipulate for crop improvement (Jiaqin *et al*., 2009). However identifying other characters that correlate with yield will allow an indirect selection of yield based on those characters. Path coefficient analysis (PCA) gives more information than simple correlations by partitioning both direct and indirect effects, thereby revealing the importance of each component in determining the trait of interest (Gelalcha & Hanchinal, 2013).

The use of correlation coefficients jointly with path coefficient analysis to understand trait associations has been widely reported (Hassan *et al*., 2013; Machikowa & Laosuwan, 2011; Vu *et al*., 2019; Sayo *et al*., 2017 ; Salim *et al*., 2014). Salim *et al*., (2014) found that days to flowering, number of pods/plant, pod yield/plant, pod length, number of seeds/pod, number of seeds/plant, 100-seed weight influenced lablab seed

yield/plant directly in positive direction. Pramod & Prakash (2011) reported that number of pods per plant, pod length, pod width and seed length had positive effect on pod yield per plant while days to first flowering had negative effect. However, through path coefficient analysis, they found highest indirect effect on pod yield/plant was through days to first flower, days to first picking and per cent fruit set/cluster.

However, there is no literature on association of yield related traits of early maturing determinate lablab genotypes in Kenya. Several breeding lines have been generated through crossing of local landraces and early maturing determinate accession which if evaluated for trait variability and association would generate valuable information for lablab breeding in Kenya.

7.2.1 Objectives

The objective of this study was

1. To determine genetic variation, heritability, genetic advance and to estimate correlation between yields and yield attributing traits;
2. To determine the direct and indirect effects of these components in grain yield of determinate lablab lines in Kenya.

7.3 Materials and methods

7.3.1 Plant materials.

Five local lablab indeterminate genotypes (Njoro, GBK 028663, DL1002, Kagio and Kahuro) and a determinate accession (KDD) were used to generate the population used in this study. Njoro, Kagio and Kahuro are landraces with indeterminate growth habit, moderate maturity duration and with large black seeds. Genotype GBK 028663 is an accession from gene bank of Kenya (GBK) which has indeterminate growth habit, moderate maturity and with brown spotted seeds. The cultivars Njoro, GBK 028663, Kagio and Kahuro were collected from Nakuru, Embu, Kirinyaga and Muranga

counties respectively. DL1002 is a released variety in Kenya and has indeterminate growth habit, moderate maturity with black seeds. The accession KDD was collected from a market in Nairobi, has determinate growth habit, early maturity and with small cream seeds.

7.3.2 Crossing and selection scheme

The six parental genotypes were planted in plastic pots in a screen house at KALRO Thika in April 2014. The parents were planted three times at one week interval in order to synchronise the flowering of the genotypes. Crosses were made between the determinate parent and all the other indeterminate parents using the technique described by Rangaswami and Nambiar (1935). The F_1 seeds from all the crosses were harvested and bulked together to form a composite bulk. The F_1 seeds were planted in a screen house in October of 2014 and allowed to self-pollinate to generate F_2. All the F_2 seeds were then space planted in the field during the long rains of 2015 to raise F_2 plants. All the F_2 plants with determinate growth habit were selected and F_3 seeds harvested and bulked. A sample of determinate F_3 seeds were planted during the short rains of 2015 to raise F_3 plants. About 400 F_3 determinate plants were selected and their seeds harvested individually. Progeny rows of the F_{3-4} seeds were raised in the field during the long rains of 2016. A spacing of 50cm between rows and 20 cm within rows was used. Each progeny row was 2 m long with about 10 plants. At maturity, 39 progeny rows were selected based on maturity period, number of pods per plant, plant height and seed colour. The selection tried to capture as much variability of these traits as possible. The F_5 seeds of the selected rows were harvested separately. These seeds were used to study the relationship between grain yield and its related traits.

7.3.3 Field experiments for yield analysis

Replicated trials were conducted in two locations, KALRO Katumani in Machakos County, and KALRO Thika (formerly KARI Thika) in Muranga County. Different characteristics of the two sites is given in the Table 7.1 below.

Table 7. 1: Different characteristics of test locations

Testing location	Altitude (M.a.s.l)	Location		Annual rainfall (mm)	Annual Temperature (0C)		Soil type
		Latitude	Longitude		Min	Max	
Thika	1549	10 09’ S	370 04’E	900	14	25	Nitosols
Machakos	1600	10 35’ S	370 14’E	1000	12	27	Chromic luvisols

In each site, the 39 F_5 lines together with their determinate parent were planted using randomized complete block design (RCBD) with two replications. The plot size was two rows of 2 m long each. A basal application of Di-ammonium phosphate (DAP) fertilizer at a rate of 150 Kgha-^{1} was done in the planting furrows during planting. One seed was planted in the furrow at a spacing of 50 x 25 cm between and within rows respectively giving a population of 8 plants per m^2. Hand weeding was done two times at the first trifoliate and pre-flowering growth stages. The crop was sprayed with recommended insecticides to control insect pests.

7.3.4 Data collection

For each trial, plant height at physiological maturity stage, number of pods per plant at maturity, number of racemes per plant at maturity, pod width and length at pod filling stage, number of primary branches per plant, number of pods per raceme (main raceme) at maturity, number of flowering nodes at 50 % flowering, days to 50% flowering, days to 75% pod maturity, 100 seed weight and grain yield per plant were recorded. The data on days to 50% flowering, days to 75% pod maturity and 100 seed weight was on

plot basis while the other traits were taken from six randomly selected plants per plot. The grain yield weight was obtained at approximately 13% moisture content.

7.3.5 Data analysis

Data for each trait was subjected to analyses of variance (ANOVA) using a GENSTAT ed. 15 statistical program to estimate the genetic variability of the selected genotypes and to partition the phenotypic variability into components due to genetic and environmental factors. Measures of variability such as genotypic coefficient of variability (GCV), phenotypic coefficient of variability (PCV), broad sense heritability (h_2), and genetic advance (GA) based on percentage of the mean were estimated.

The genetic parameters were estimated using formulas adapted from Allard, (1960), Singh and Chadhary, (1985) as follows:

σ^2G = [Mean Square Genotype-Mean Square Error/*r*]

σ^2P = [Mean Square Genotype/*r*]

Ve = [Mean Square Error/*r*]

r is the number of replications. σ^2P is the phenotypic variance while σ^2G is the genotypic variance. The Mean Square Genotype (MSG) and Mean Square Error (MSE) are variance components estimated as functions of the mean square estimates from ANOVA table. Mean square genotype (MSG): estimates genotypic variance, this value is observed variance among the line means, while mean square error (MSE) measures variance from plot residuals. Phenotypic (PCV) and genotypic (GCV) coefficient are estimated using the following formulas:

PCV = ($\sqrt{}$Vp/X) x 100

GCV = ($\sqrt{}$Vg/X) x 100

Vp represents the phenotypic variance: Vg represents the phenotypic variance, while X represents the mean.

Heritability (h^2B) expressed as the percentage of the ratio of the genotypic variance (Vg) to the phenotypic variance (Vp) was estimated based on the genotypic mean.

Expected genetic advance (GA) was estimated using a formula of Allard, (1960) as

$GA = K (Sp) h^2B$,

GA (as % of mean) = (GA/X) x100

Where h^2B and Sp is the heritability ratio and the phenotypic standard deviation ($\sqrt{Vp}$) and K is a selection differential that varies depending on the selection intensity. In the present analysis 2.06 was considered for K, which is 5% selection intensity.

Correlation coefficient (*r*) was used in the study to determine inter-relations between 11 quantitative characters. Pearson correlation coefficients between traits were generated using the IBM SPSS statistics version 20 procedure over the two locations. The significance test for correlation coefficient was tested on a two-tailed test on the same program.

Using the same software, a linear regression analysis was carried out to estimate the relationship between seed yield per plant (dependent variable) and the other 10 independent variables. Validity of P values for the t-test was determined by testing the normality of residuals. A test on multicollinearity among the various predictors was also done.

Path coefficient analysis for seed yield per plant was carried out as demonstrated by Dewey and Lu (1959). Ten characters, 100 seed weight, pod per plant, raceme per plant, pod length, pod width, pods per raceme, flowering nodes per plant, plant height, days to 50% flowering and days to maturity were included in the path coefficient analysis for single plant yield. Simultaneous equations were drawn as per Dabholkar (1992) as below:

r10 = P10 + P20r12 + P30r13 + P40r14 + P50r15 + P60r16 + P70r17 +P80r18 + P90r19 +P100r110

r20 = P10r21 + P20 + P30r23 + P40r24 + P50r25 + P60r26 + P70r27 +P80r28 + P90r29 +P100r210

r30 = P10r31 + P20r32 + P30 + P40r34 + P50r35 + P60r36 + P70r37 +P80r38 + P90r39 +P100r310

r40 = P10r41 + P20r42 + P30r43 + P40 + P50r35 + P60r46 + P70r47 +P80r48 + P90r49 +P100r410

r50 = P10r51 + P20r52 + P30r53 + P40r54 + P50 + P60r56 + P70r57 +P80r58 + P90r59 +P100r510

r60 = P10r61 + P20r62 + P30r63 + P40r64 + P50r65 + P60 + P70r67 +P80r68 + P90r69 +P100r610

r70 = P10r71 + P20r72 + P30r73 + P40r74 + P50r75 + P60r76 + P70 + P80r78 + P90r79 +P100r710

r80 = P10r81 + P20r82 + P30r83 + P40r84 + P50r85 + P60r86 + P70r87 + P80 + P90r89 +P100r810

r90 = P10r91 + P20r92 + P30r93 + P40r94 + P50r95 + P60r96 + P70r97 + P80r98 + P90 +P100r910

r100 = P10r101 + P20r102 + P30r103 + P40r104 + P50r105 + P60r106 + P70r107 + P80r108 + P90r109 +P100

Where 0 = Dependant variable = Yield; 1 - 10 were independent variables;

1 = 100 SW, 2 = pods per plant, 3 = raceme per plant, 4 = pod length, 5 = pod width, 6 = pods per raceme, 7 = flowering nodes number, 8 = plant height, 9 = days to 50% flowering, 10 = days to maturity, respectively.

r = Pearson Correlation Coefficient between traits

P = Path Coefficient (regression standardized coefficients)

7.4 Results

7.4.1 Genotypic and phenotypic variance of determinate lablab characters

Significant differences were observed among the genotypes for all the characters suggesting presence of high amount of variability for the characters studied. The mean sum of squares for 9 characters in 39 genotypes and one check accessin of lablab are presented in Table 7.2.

Table 7. 2: Analysis of variance of 9 yield and yield related traits of 40 determinate F_5 genotypes evaluated at Thika and Katumani sites during the LR 2017

Mean square										
Site	S.O.V.	D.F	SWPP	100Sw	PPP	RPP	PL	PW	PH	DTF	DTM
Thika	Rep	1	211.25	0.45	781.20	22.76	0.27	0.003	92.02	1.51	1.80
	Genotype	39	115.54***	13.15**	239.20*	5.60**	0.27***	0.06***	187.61*	20.91*	14.33**
	Error	39	41.14	6.50	125.70	3.14	0.08	0.009		12.54	5.56
	Total	79									
Katumani	Rep	1	337.57	1.80	147.60	13.41	1.61	0.14	18.24	0.05	36.45
	Genotype	39	105.89***	19.91***	268.8**	4.42**	0.32***	0.09***	325.52***	7.15***	85.5***
	Error	39	36.46	3.44	106.1	2.53	0.12	0.018	50.33	1.43	16.17
	Total	79									

SWPP= seed weight per plant, 100SW=100 seed weight, PPP=pods per plant, RPP= raceme per plant, PL=pod length, PW= pod width, PH=plant height, DTF= days to flowering, DTM= days to maturity,

***, **, * is significant at the 0.001, 0.01, 0.05 level respectively.

The estimates of genotypic variance (σ2G), phenotypic variance (σ2P), phenotypic coefficient of variability (PCV) and genotypic coefficient of variability (GCV), heritability (in a broad sense), and genetic advance as a percentage of the mean were analysed (Table 7.3) for the Thika site (Table 7.4) for Katumani site.

High range of variation in Thika site was recorded for seed weight per plant are 5 g – 42 g, pods per plant 9-68, plant height 26 cm – 82 cm and for 100 seed weight 14 g – 30 g, while in the Katumani experiment the ranges for seed weight per plant is 13 g – 51 g, pods per plant 20 – 88, plant height 46 cm – 143 cm and for 100 seed weight is 14 g – 30 g. Low range of variation was observed for days to flowering and days to maturity in both sites.

Table 7. 3: Range, mean, variance, broad sense heritability, genotypic and phenotypic coefficient of variations and genetic advance as percent of mean for characters of lablab genotypes studied at Thika

Traits	σ^2G	σ^2P	σ^2E	Mean	Range	PCV	GCV	ECV	H^2	GA	GAM (%)
Seed weight per plant	35.18	55.75	20.57	20.00 ± 1.00	5 - 42	37.33	29.66	22.68	0.63	9.71	48.53
100 SW	3.32	6.57	3.25	20.88 ± 0.35	14 - 30	12.28	8.73	8.63	0.51	2.67	12.78
Pods per plant	56.75	119.60	62.85	32.00 ± 1.62	9 - 68	34.18	23.54	24.77	0.47	10.69	33.41
Raceme per plant	1.25	2.82	1.57	5.00 ± 0.24	1 - 10	33.56	22.32	25.06	0.44	1.53	30.57
Pod length	0.10	0.14	0.04	5.10 ± 0.05	3.9 – 6.9	7.20	6.04	3.92	0.70	0.53	10.44
Pod width	0.030	0.034	0.00	1.70 ± 0.02	1.3 – 2.0	10.60	9.84	3.95	0.86	0.32	18.82
Plant height	45.92	93.81	47.89	51.00 ± 1.32	26 - 82	18.99	13.29	13.57	0.49	9.77	19.15
Days to flowering	4.19	10.46	6.27	50.54 ± 0.45	46 - 59	6.40	4.05	4.95	0.40	2.67	5.28
Days to maturity	4.38	7.16	2.79	100.00± 0.35	94 - 104	2.68	2.09	1.67	0.61	3.37	3.37

S.E Mean= Standard error of the mean, σ^2G = Genotypic variance, σ^2E = Environmental variance, σ^2P = Phenotypic variance, H^2 (%) = Broad sense heritability, GCV (%) = Genotypic coefficient of variation, PCV (%) = Phenotypic coefficient of variation, (%) ECV= Environmental coefficient of variation, (%) GA= Genetic advance, GAM= Genetic advance as percent of mean.

In the present investigation, the σ^2P was higher than the corresponding σ^2G for all the characters evaluated at Thika site (Table 7.3). For most of the characters studied, the two values differed moderately suggesting that the expression of these characters was reasonably influenced by the environment. However, higher differences between σ^2P and σ^2G was observed for DTF, RPP, PPP and PH at Thika site. The same trend of high σ^2P compared to their corresponding σ^2G with the former being higher than latter was observed for all traits at Katumani site (Table 7.4).

PCV and GCV was high for number of racemes per plant at 33.5% (PCV) compared to 22.3% (GCV) in Thika and 21.24% (PCV) compared to 14% (GCV) at Katumani.

Table 7.4: Range, mean, variance, broad sense heritability, genotypic and phenotypic coefficient of variations and genetic advance as percent of mean for characters of lablab genotypes studied at Katumani

Traits	σ^2G	σ^2P	σ^2E	Mean	Range	PCV	GCV	ECV	H^2	GA	GAM (%)
Seed weight per plant	34.72	52.95	18.23	26 ± 0.97	13 - 51	27.99	22.66	16.42	0.66	9.83	37.80
100 SW	8.23	9.95	1.72	21.30 ± 0.38	14 - 30	14.81	13.47	6.16	0.83	5.37	25.23
Pods per plant	81.35	134.40	53.05	42 ± 1.53	20 - 88	27.60	21.47	17.34	0.61	14.46	34.42
Raceme per plant	0.96	2.21	1.25	7.00 ± 0.21	4 - 13	21.24	14.00	15.97	0.43	1.33	19.00
Pod length	0.10	0.16	0.06	5.1 ± 0.05	4 – 6.3	7.84	6.20	4.80	0.63	0.52	10.10
Pod width	0.04	0.05	0.01	1.7 ± 0.03	1.3 – 2.1	12.89	11.62	5.58	0.81	0.37	21.57
Plant height	137.6	162.7	25.17	70.4 ± 1.52	46 - 143	18.12	16.66	7.13	0.85	22.22	31.56
Days to flowering	2.85	3.58	0.73	55.8 ± 0.23	53 - 62	3.39	3.02	1.53	0.80	3.10	5.55
Days to maturity	34.6	42.75	8.09	99.08 ± 0.80	90 - 120	6.60	5.94	2.87	0.81	10.92	11.02

S.E Mean= Standard error of the mean, σ^2G = Genotypic variance, σ^2E = Environmental variance, σ^2P = Phenotypic variance, H^2 (%) = Broad sense heritability, GCV (%) = Genotypic coefficient of variation, PCV (%) = Phenotypic coefficient of variation, (%) ECV= Environmental coefficient of variation, (%) GA= Genetic advance, GAM= Genetic advance as percent of mean.

In this site, the highest difference between σ^2P and their corresponding σ^2G was recorded for PPP (134.4 for σ^2P and 81.35 for σ^2G) and RPP (2.21 for σ^2P and 0.96 for σ^2G). The differences between the σ^2P and σ2G for most of the characters was lower at Katumani than at Thika site.

The highest PCV at Thika site was observed in the seed weight per plant at 37.33% and the lowest was on the days to maturity at 2.68%. In the Katumani experiment, the PCV ranged from high for seed weight per plant at 27.99% to low for days to 50% flowering

at 3.39%. The PCV was relatively higher at Thika experiment for raceme per plant at 33.5%, seed weight per plant at 37.3% compared to 21.24% for raceme per plant and 27.99% for seed weight per plant at Katumani field. The PCV was low for the days to flowering and maturity at both Katumani and Thika sites.

The GCV at Thika experiment ranged from seed weight per plant (29.66%) to 2.09% for days to maturity, while at Katumani the same trait (seed weight per plant) showed the highest GCV (27.99%) and the lowest GCV of 3.02% was recorded for days to flowering. High GCV values (>20%) were recorded on pods per plant, seed weight per plant and racemes per plant in both sites indicates that these traits could lead to good progress in crop improvement. Similarly, the lower GCV values (<7%) shown by, days to flowering, days to maturity and pod length in both sites suggest that these traits will be less responsive to improvement through selection. The values of both PCV and GCV in Thika and Katumani compared well though the values at Thika site were slightly higher than those of Katumani for majority of the traits. Heritability (H^2) expressed as the percentage of the ratio of the genotypic variance (σ^2G) to the phenotypic variance (σ^2P) was estimated for each character in the two sites (Table 7.3 and Table 7.4). At Thika site, broad sense heritability estimates ranged from 0.86 for pod width to 0.4 for days to flowering, whereas heritability was highest for plant height at 0.85 and lowest for number of racemes per plant at 0.43. Most of the characters recorded broad sense heritability of more than 0.5 at Thika while the values were more than 0.6 for most of the characters at Katumani site. There was higher broad sense heritability estimates at Katumani experiment for days to flowering (0.81), days to flowering (0.80), plant height (0.85) and 100SW (0.83) compared to 0.61, 0.40, 0.49 and 0.51 respectively for the same characters at Thika experiment.

The genetic advance (GA) shows the extent of gain in a trait which is attained under a particular selection pressure. In this study, selection pressure of 5% was considered. The genetic advance was expressed as percentage of the mean (GAM) of the trait for easier comparison amongst traits with different measurements units. The results showed highest GAM for seed weight per plant of 48.53% and 37.80% at Thika and Katumani respectively whereas the lowest GAM was recorded for days to maturity (3.37%) at Thika and days to flowering (5.55%) at Katumani. In addition, pods per plant and raceme per plant showed relatively high estimates of genetic advance of 33.4% and 30.57% respectively, at Thika site. The other traits that displayed higher estimates GAM at Katumani experiment were pods per plant (34.42%), 100 seed weight (25.23%) and plant height (31.56%). The maturity related traits (days to 50% flowering and days to maturity) and pod length consistently showed lowest GAM of less than 12% in both sites.

7.4.2 Correlation coefficients among traits

The Pearson correlations coefficients of 11 traits observed in this study are presented in Table 7.5 below. The correlations were based on the calculated means of the genotypes for the Thika and Katumani experiments. The study showed that all the characters were positively and significantly ($P > 0.05$) correlated to seed weight per plant with exception of 100 seed weight and pod length (Table 7.5). Pods number ($r = +0.87$) and raceme number ($r = +0.81$) seems to contribute significantly to seed weight per plant which suggests that indirect selection for grain weight per plant can be effectively realized by selecting for the number of pods and raceme.

Table 7.5: Correlation coefficients among 11 characters estimated from 40 genotypes of lablab bean.

	SWPP	100 sw	PPP	RPP	PL	PW	PPR	NFN	PH	DTF	DTM
Seed weight per plant (SWPP)	1	0.086	0.869**	0.811**	0.025	0.223**	0.203**	0.461**	0.515**	0.254**	.182*
100 sw		1	-0.13	0.056	-0.038	0.102	-0.279**	-0.068	0.128	0.061	0.087
Pods per plant (PPP)			1	0.752**	0.019	0.059	.411**	0.487**	0.435**	0.208**	0.079
Raceme per plant (RPP)				1	-0.024	0.12	0.13	0.429**	0.593**	0.386**	0.072
Pod length (PL)					1	0.019	-0.085	0.013	0.021	0.065	0.072
Pod width (PW)						1	-0.058	-0.019	0.086	0.127	0.275**
Pods per raceme (PPR)							1	0.584**	0.124	-0.06	-0.165*
No. flower nodes (NFN)								1	0.410**	0.086	-0.061
Average plant height (PH)									1	0.500**	0.138*
Days to flowering (DTF)										1	0.426**
Days to maturity (DTM)											1

***, **, * Correlation is significant at the 0.001, 0.01, 0.05 level respectively

In addition, moderate significant correlations was observed for seed weight per plant and number of nodes per raceme (r=+0.46) and plant height (r=+0.52). The genotypes that produced many pods per plant also had many racemes per plant as shown by the high positive correlations (r=+0.75) between the two traits. Similarly, tall plants tended to produce many racemes as indicated by relatively high positive corrections (r=+0.60)

between the two traits. Moderate positive correlations were recorded between pods per plant and pods per raceme (r=+0.41), number of nodes per raceme (r=+0.49) and plant height (r=+0.44). Genotypes with many racemes per plant, pods per raceme and big plant height tended to have many flowering nodes per raceme as indicated by their positive moderate correlation of 0.43, 0.58 and 0.41 respectively. The days to flowering exhibited some moderate positive correlations of r=+0.50 and r=+0.43 for plant height and days to maturity respectively. Low but significantly negative correlations was recorded for 100 seed weight and the number of pods per raceme (r= -0.28).

7.4.3 Path analysis

The analysis of variance (ANOVA) indicated that the regression model of the 10 independent variable on seed weight of the lablab genotypes was significant at P< 0.001 (Table 7.6). The regression coefficients from the model were therefore used for undertaking the path analysis. The adjusted R square was high at 0.869.

Table 7. 6: ANOVA and summary of regression model of 10 independent variables on seed weight per plant of 40 determinate lablab genotypes

ANOVA of Seed weight per plant				**Regression Model summary**			
Model	Sum of Squares	DF	Mean Square	R	R Square	Adj. R Square	Std. Error of the Estimate
Regression	14286.209	10	1428.621***	0.932	0.869	0.860	3.803
Residual	2155.485	149	14.466				
Total	16441.694	159					

*** is significant at the 0.001 level

The regression coefficients and collinearity statistics of the causal factors on seed weight per plant of lablab genotypes is presented in Table 7.7 below. The beta standardized coefficients ranged from 0.68 for pods per plant to -0.08 for days to flowering. The regression coefficients for all the independent variables were significant at p< 0.05 except for pod length and plant height. Collinearity implies that two variables

are linear combinations of each other. As the collinearity increases, the regression coefficient estimates become unstable. Collinearity statistics consists of tolerance and variance inflation factors (VIF). The tolerance and VIF statistics for the independent variables considered in this study were within the acceptable range of > 0.10 for tolerance and < 10 for VIF.

Table 7.7: Regression coefficients and collinearity statistics of 10 independent variables on seed weight per plant of 40 determinate lablab genotypes

Model	Unstandardized coefficient		Standardized coefficient	T	Collinearity statistics	
	B	Std. err	Beta		Tolerance	VIF
Constant	-23.494	7.284	-	-3.225***	-	-
100 SW	.337	0.100	0.108	3.380***	0.863	1.159
Pods per plant	.442	0.035	0.681	12.745***	0.309	3.241
Raceme per plant	1.027	0.234	0.248	4.392***	0.276	3.626
Pod length	.145	0.674	0.007	.216ns	0.961	1.041
Pod width	5.561	1.446	0.120	3.846***	0.903	1.108
Pod per raceme	-.346	0.117	-0.126	-2.948***	0.482	2.076
No of flower nodes	.395	0.174	0.096	2.272*	0.490	2.041
Plant height	.034	0.026	0.054	1.299ns	0.509	1.963
Days to flowering	-.201	0.096	-0.082	-2.094*	0.574	1.743
Days to maturity	.148	0.065	0.080	2.270*	0.709	1.410

***, **, * regression coefficients is significant at the 0.001, 0.01, 0.05 level respectively

A path coefficient analysis which gives indirect effects for each independent variable on the dependent variable is presented in Table 7.8. Among the 10 independent traits (causal), eight of them had positive direct effect on seed yield per plant (dependent). These were 100 seed weight, pods per plant, racemes per plant, pod length and pod width, number of flowering nodes, plant height and days to maturity. Pods per raceme and days to maturity had negative direct effect on grain yield per plant. Pods per plant, racemes per plant, pod width and 100 seed weight had the largest direct effect on seed weight yield respectively.

Table 7. 8: Path coefficient analysis showing direct effect (bold) and indirect effects of 10 characters on seed yield per plant in lablab bean

Character	100SW	PPP	RPP	PL	PW	PPR	NFN	PH	DTF	DTM
100 seed weight (100SW)	**0.108**	-0.085	0.014	-0.001	0.012	0.035	-0.007	0.007	-0.005	0.007
Pods per plant (PPP)	-0.085	**0.681**	0.187	0.000	0.007	-0.052	0.047	0.023	-0.017	0.006
Raceme per plant (RPP)	0.006	0.512	**0.248**	0.000	0.014	-0.016	0.041	0.032	-0.032	0.006
Pod length (PL)	-0.004	0.013	-0.006	**0.007**	0.002	0.011	0.001	0.001	-0.005	0.006
Pod width (PW)	0.011	0.040	0.030	0.000	**0.120**	0.007	-0.002	0.005	-0.010	0.022
Pod per raceme (PPR)	-0.030	0.280	0.032	-0.001	-0.007	**-0.126**	0.056	0.007	0.005	-0.013
No. of flower nodes (NFN)	-0.007	0.331	0.106	0.000	-0.002	-0.074	**0.096**	0.022	-0.007	-0.005
Plant height (PH)	0.014	0.296	0.147	0.000	0.010	-0.016	0.039	**0.054**	-0.041	0.011
Days to flowering (DTF)	0.007	0.142	0.096	0.000	0.015	0.008	0.008	0.027	**-0.082**	0.034
Days to maturity (DTM)	0.009	0.054	0.018	0.000	0.033	0.021	-0.006	0.007	-0.035	**0.080**

Note: Bold and diagonal figures indicate direct effect; Residual effect: 0.14

Raceme per plant, pods per raceme, number of flowering nodes per raceme and plant height had fairly high indirect positive effects through pods per plant (0.51, 0.28, 0.33, 0.30) respectively. The effects of these four traits on seed weight through pods per plant was larger than their respective direct effects. Even though the direct effect of days to flowering is negative, its indirect effects through pods per plant is positive and substantial. Pods per plant and number of flowering nodes had small indirect negative effect of -0.05 and -0.07 respectively through pods per raceme. Small negative indirect negative effect of pods per plant through 100 seed weight (-0.09) was recorded.

7.5 Discussion

Genotypic variability is prerequisite for progress of any breeding program. In conventional plant breeding, variability is created through crossing (hybridization) of

plants which carry the desired genes followed by discrimination among the variability (selection) to identify the most desirable recombinant (Acquaah, 2012). Success in improvement of the desired traits require presence of many recombinants with wide range of targeted character from which selection can be done (Oduori, 2008). In this study, the variability of some determinate recombinant lablab lines derived from crossing early maturing determinate accession with the local indeterminate landraces were evaluated. The highly significant difference in mean squares implied that there is distinguishable evidence of inherent genetic variability among the lablab lines with respect to seed weight per plant, 100 seed weight, pods per plant, raceme per plant, plant height, days to 50% flowering and days to maturity. Wide range of variation was shown for most of the traits except for days to 50% flowering and maturity. The presence of variability for most traits offers adequate variation upon which to establish a breeding program. The narrow range of variation observed on maturity traits could be explained by the fact that all the lines evaluated in this study were of determinate growth habit. Selection of genotypes with determinate growth habit at early generation population could also have resulted in selection of early flowering and maturing genotypes. The mean number of days to flowering (50-55) and maturity (99-100) of the determinate lines tested in this study are lower than those reported on indeterminate local landraces by Kamotho (2015). This indicates that selection of determinate growth habit in lablab can result in selection for early flowering and maturing genotypes. Earliness to flowering and maturity has been reported to be linked to determinate growth habit in some legumes (Keerthi *et al*., 2014; Repinski *et al*., 2012; González *et al*., 2016).

The study showed that the phenotypic variance was higher than the genotypic variance in all the traits studied across the sites. Progress from selection depends on the

availability of genetic variability in the population and selection is more effective when there is high genetic variation in relation to environmental variation. In this study, the magnitude of the genotypic variance for seed weight per plant, 100 seed weight, pod length, pod width and days to maturity yield components were consistently higher than their respective environmental variance in both sites. This implies that significant improvement for these traits can be achieved through phenotypic selection (Manggoel *et al*., 2012). Variability in quantitative characters of lablab was also reported by Verma *et al*., (2015) and Salim *et al*., (2014).

The genotypic and phenotypic variance estimates and the range of mean values of traits can give a rough estimate about the magnitude of variation present among different genotypes. However, genotypic coefficient of variation is better in revealing the extent of variability present within the genetic materials and its estimate gives good implication for genetic potential in crop improvement through selection (Burse *et al*., 2015). According to Hailu, *et al*., (2016) PCV and GCV values > 20% are regarded as high, while values between 10% and 20% medium and values < 10% are considered low. In this study, high PCV and GCV values of > 20% were observed for seed weight per plant and pods per plant in all the sites. In addition to this, these characters recorded medium to high heritability suggesting the presence of more additive gene effects for possible improvement. Singh *et al*., (2015) and Parmar *et al*., (2013) in lablab and Asante *et al*., (2009) in mung bean have reported high PCV and GCV and moderate heritability for number of pods per plant. However, heritability values reported for pods per plant in this study (0.49 for Thika and 0.61 for Katumani) is lower than those reported by Pramod *et al*., (2011), Verma *et al.*, (2015) and Salim *et al.*, (2014) in lablab but was higher than that of Chaitanya *et al*., (2014). This discrepancies of heritability values could be due to nature of test materials and environment where

experiments were conducted. The low PCV and GCV values and moderate to high heritability observed in this study agrees with that of Singh *et al.*, (2015) in lablab. Generally, the differences between PCV and their corresponding GCV values for all characters with exception of racemes per plant was small to moderate indicating that these characters were less influenced by the environment. The high difference between GCV and PCV for raceme per plant and its corresponding moderate heritability in both sites suggest that selection for this trait using observed variation may be less effective since the proportion of additive gene effect is low.

Successful selection based on phenotype is possible in characters with large heritability estimates, however use of heritability together with genetic advance estimates is more reliable in guiding selection (Ayalew *et al*., 2011). Genetic advance (GA) is the gain of genotypic value of a new population compared with the base population resulting from one cycle of selection in a given selection intensity (Hailu *et al*., 2016). In this study, high heritability coupled with relatively high genetic advance as percent of mean was observed for seed weight per plant, pods per plant, pod width and plant height. This suggest that these traits are under additive gene action and selection based on the phenotypes will be effective. Our results were similar with the findings of Chaitanya *et al*., (2014), Veerendra *et al.*, (2014) and Singh *et al*., (2013) for pods per plant, Pramod *et al*, (2011) and Parmar *et al*., (2013) for pod width, Singh *et al*., (2015) and Verma *et al*., (2015) for plant height, Salim *et al*., (2014) and Singh *et al*., (2013) for seed yield per plant. Moderate to high heritability coupled by low genetic advance as percentage of mean was observed for days to 50% flowering and maturity. The low genetic advance of these two maturity related traits could be as a result of the small phenotypic variance of the test materials. This is because genetic advance has direct relationship with standard deviation of the population and heritability. In this study, all the test materials

are of determinate growth habit, a trait that is associated with earliness to maturity (Keerthi *et al*., 2014) and this could have contributed to narrow phenotypic variability for these maturity traits and consequently low genetic advance. However, the moderate to high heritability and low genetic advance of these two traits could also be as a result of influence of non- additive gene action and considerable influence of the environment. This suggests that based on the evaluated determinate lablab lines in this study, days to flowering and maturity characters can only be partialy improved by selection. Simmilar results of moderate to high heritability coulped by low genetic advance for days to 50% flowering and maturity has been reported by Singh *et al*., (2015) and Verma *et al*., (2015) in lablab and Veerendra *et al*., (2014) in pigeon peas.

Plant breeders are rarely interested with one character and therefore there is need to study the association between various characters and especially between yield and other traits (Tadele *et al*., 2014). A good understanding of the relationship of plant characters and the yield is essential because the final yield is the sum total of effects of all its related traits (Verma *et al*., 2015). The knowledge of the association of grain yield and its related traits of the determinate lablab genotypes will allow an indirect selection of yield based on those characters. In the present study, correlation coefficient (*r*) was used to determine associations of yield related traits of the determinate lablab genotypes. Correlation coefficient ranges between -1 and 1 with correlation of 0 implying that there is no linear relationship between the variables while -1 or 1 suggests total linear relationship.

This study identified that the number of pods and racemes had high significant correlation with seed weight per plant. The results are in agreement with those of Salim *et al*., (2014) in lablab, Tadele *et al.*, (2014) in lentil and Manggoel *et al.*, (2012) in

cowpeas. In determinate lablab genotypes, seeds are contained in pods borne on the axillary or terminal racemes and therefore genotypes with many pods and racemes are likely to contain many seeds and therefore high seed yield. This suggests that in breeding programs, selection for high grain yield can be effectively realized through the indirect selection for traits that are visually easy to score like high number of pods and racemes. Other traits with moderate to low significant positive correlation with grain yield per plant include days to 50% flowering, maturity pod width, plant height and flowering nodes per raceme. This suggests that those lines which, flowered and matured late, with longer stem height, many flowering nodes and many racemes and pods produced high grain yield. Generally, in absence of production constraints such as moisture stress, late flowering and maturing varieties are likely to have an advantage in gaining more plant height and grain productivity since the total amount of photosynthesates received from the leaves is greater than for early maturing varieties (Yamada *et al*., 2012).

Through path coefficient analysis, the total correlation between traits are partitioned into direct and indirect effects which forms a better basis for selection to improve yield. (Manggoel *et al*., 2012). Direct effects measure the sensitivity of the dependent variable to changes in the independent variable by one unit while all other factors are held constant. In contrast, the indirect effect quantify the changes on dependent variable when the independent variable is held constant and the intermediary variable changes by the amount it would have changed had the independent variable increased by one unit (Pearl, 2001). In this study, the number of pods per plant had the highest direct positive effect on seed yield per plant. High direct effects of number of pods per plant on seed yield has been reported by other authors Salim *et al*., (2014), Singh *et al*., (2011), Singh *et al*., (2015) and Verma *et al*., (2015) in lablab, Veerendra *et al*., (2014)

in French beans and Machikowa & Laosuwan (2011) in Soybean. Moderate positive indirect effect on seed yield through pods per plant was detected for raceme per plant, pods per raceme, flowering nodes per plant and plant height. This suggest that selection of high number of units of these characters can be effective in selection of high seed yield in determinate lablab genotypes. The negative direct effect of days to flowering and 100 seed weight on grain yield per plant observed in the current study is in agreement with the findings of Singh *et al*., (2011) in lablab. However, the negative effect of these two traits is cancelled by positive indirect effects through other traits. The overall positive association (sum of direct and indirect effects) of plant height, days to flowering and maturity and yield suggests that there will be a problem when combining dwarfness, earliness to flowering and maturity with high seed yield. Development of early maturing lines is suitable for drought prone areas or where early cessation of rainfall is prevalent. Short determinate lablab lines would fit an intercropping system.

The advanced determinate inbred lines used in this study had substantial genetic variability. Plant characters such as seed weight per plant, pods per plant, pod width and plant height had high heritability coupled with relatively high genetic advance in F_5 determinate inbred lines. There was significant positive correlations between number of pods per plant, raceme per plant, plant height, pod width, pods per raceme and number of flower nodes. The same traits also had high direct and indirect effects on seed yield. These traits could therefore be used as suitable selection criteria for effective improvement in yield of determinate lablab genotypes.

CHAPTER EIGHT

General Discussion Conclusion and Recommendation

Lablab is a grain legume crop that is commonly grown in Africa and India. In Africa, lablab has remained neglected and underutilized despite the crop being well adapted to semiarid areas which are characterized by food insecurity. In Kenya, farmers grow landraces which are inherently low yielding and have other undesirable attributes like long maturity duration and indeterminate growth habit. Comprehensive understanding of the genetic diversity and population structure is important to identify genotypes suitable to be used as parental materials in breeding programs. The availability of novel tools such as molecular markers which are suitable for germplasm characterization and marker assisted selection are necessary for successful crop improvement efforts. The study on genetics and association among important traits would assist in identifying the best strategies for identifying and selecting lablab for breeding purposes.

This study was therefore focused on generation of new genetic information and molecular marker tools for breeding of early maturing determinate lablab varieties in Kenya.

The following research areas were investigated:

- To identify new lablab derived simple sequence repeat (SSR) molecular markers and their characterization. The markers form additional molecular tools for use in the genetic diversity, population structure and genetic analysis of the crop.
- To assess the genetic diversity of some local and introduced lablab accessions using the new SSRs markers. This would aid in selecting of parental materials for future breeding and conservation efforts.

- To determine the genetic diversity and population structure of some worldwide lablab accessions based on two ultra-high-throughput diversity array technology (DArT) markers (Silico DArT and SNP). This would provide some additional variety of molecular markers and generate further information on the dissimilarity of local and regional lablab germplasm.
- To analyze the genetics of earliness to maturity and growth habit in lablab. This would help in identifying best selection strategies for developing these traits.
- To determine the genotypic variability for yield and other agronomic traits, and the association among traits of determinate lablab. This was done to evaluate the level of genotypic variability, estimate the genetic advance possible and identify yield related traits with the aim of selecting suitable determinate early maturing varieties in Kenya.

The following studies were undertaken to address the above mentioned objectives.

To address the problem of inadequate molecular markers in lablab, new simple sequence repeat markers were developed (Chapter 3). Total RNA was extracted from young leaves and terminal shoots, mRNA isolated and sequenced. The generated sequence reads were assembled, screened for microsatellite, primers designed from the flanking sequences of SSR and synthesized by Europhin MWG Operon Company.

To understand the usefulness of the new SSRs in classifying lablab germplasm, eight SSRs were used to discriminate 189 lablab accessions (Chapter 4). The accessions included both the local Kenyan collections and those originating from other countries. DNA was extracted from leaves young leaves and PCR amplification carried out. The

PCR products were size separated using 6% PolyAcrylamide gel (PAGE) and viewed under UV light in a gel box.

To further understand the genetic diversity of lablab germplasm, DArTseq based molecular markers were developed and used to genotype the accessions (Chapter 5). The accessions included both the local Kenyan collections and those originating from other countries. DNA was extracted from leaves young leaves and samples sent to Diversity Array Technology (Canberra Australia) for DArTSeq analysis.

The fourth part of the study (Chapter 6) was to understand the genetic inheritance of growth habit and other qualitative traits and their linkage in Kenyan local lablab populations. The inheritance of traits (growth habit, flower colour, stem anthocyanin and flower initiation) was studied in six F_2 populations and F_3 families in two varying growing environments (Thika and Kiboko).

The last part of the study (Chapter 7) was conducted to obtain phenotypic and genetic variability, heritability estimates, expected response to selection and character association of several vegetative and reproductive characteristics of lablab lines with determinate growth habit. Forty lablab F_5 lines and variety with determinate growth habit were evaluated at Machakos and Thika environments in replicated trials.

8.1 Molecular markers development

In the first study (Chapter 3) 145 SSR markers were designed from lablab transcriptome and 75 were tested for amplification on lablab genome. A set of 70 (93.3%) SSR markers were found to consistently amplifying the lablab genome suggesting that the approach used to design and develop the markers was quite effective. In the second study (Chapter 5) a total of 15,601 polymorphic DArT markers were identified in a set

of 240 lablab accessions. In addition, 11,431 DArTSeq SNP markers were also identified. The quality of the DArTSeq markers developed in this study were high based on their high average genotype call rate and reproducibility (>91%). This suggest the reliability of these markers in genetic analysis. These markers could therefore be useful tools for genetic diversity, population structure analysis and marker assisted selection of lablab.

8.2 Genetic diversity and population structure based on SSR, DArT and SNP markers

Eight preselected polymorphic SSR markers were used to study the genetic diversity of lablab accessions. Based on the 8 SSR loci, moderate genetic diversity was revealed among the 189 worldwide lablab accessions. The average expected heterozygosity was moderate at 0.38. However, evaluation of the same genotypes using thousands of both the DArT and SNP markers revealed a narrower genetic diversity. The expected mean heterozygosity (He) was low for both DArT (He = 0.030) and SNP (He= 0.039). In other words, the three molecular markers types had varying capability in estimating the genetic distances of the germplasm. This can be attributed to the varying number of markers used and due to the fact that different types of molecular markers target different regions of the genome. However, although the clustering of germplasm was not exactly the same for the three markers, accessions with closely related accessions were grouped together. For instance, the accessions belonging to subspecies unicinantus were always grouped together while using different markers. This consistency of clustering of accessions suggest that the three types of molecular markers (SNP, SSR and DArT) developed in this study were highly reliable for genetic diversity of lablab.

Classification of lablab germplasm based on the three marker types showed that the accessions originating from Kenya were diverse and were split across all the clusters of the phylogenetic trees. However, majority of the accessions which were collected from the farmers' fields and local markets were not distributed in all the clusters. This indicate that though the lablab germplasm in Kenya is a fair representation of the diversity existing in other regions of the world, the current cultivars by farmers are of narrow genetic diversity. The local breeding program should therefore target to use both the accessions at the gene bank of Kenya and the introductions from outside the country to improve the farmer's varieties. The breeding program should in particular target the accessions from Ethiopia and South Africa populations which recorded the highest gene diversity in this study. Based on the SSR loci, the present study was able to develop a core collection that represent a set of diverse materials that are of priority to the local breeding program. The core collection that contained 45 accessions was grouped into three clusters with majority of the accessions from Kenya clustered in group 1 and 2. This indicate that successful improvement of local materials should target the exotic materials in cluster 3 as parents for cross breeding. Majority of the accessions in cluster 3 belong to subspecies unicinatus implying its importance as a source of high genetic diversity for crop improvement.

8.3 Genetic inheritance of growth habit and other important qualitative traits in lablab

Genetics of growth habit and other qualitative traits were studied in three generations (F_1, F_2 F_3) of eight lablab populations. To further understand the inheritance pattern of growth habit in lablab under different altitudes, six F_2 populations segregating for growth habit were studied under Thika and Kiboko environments. Chi square (χ^2)

statistical procedure was used to determine inheritance pattern of these qualitative traits. Chi-square statistic measures the deviations of the observed frequencies from that of expected frequencies. Joint segregation of pairs of qualitative traits (growth habit, time to flowering, stem anthocyanin and flower colour) was conducted to detect any possible linkage of traits.

The findings from this study suggest that segregation of stem growth habit of F_2 lablab populations under Thika environment fitted to the expected ratio of 3:1 while the segregation fitted to a ratio of 57:7 under Kiboko environment. Stem growth habit in lablab is therefore controlled by three genes, one basic and two complementary genes which are temperature dependent. At high temperature (could be above 30^0C) the two genes are activated while lower temperature deactivate the genes. Similar reports were made by Keerthi *et al*., (2014) in lablab, who observed that growth habit was under control of three genes (*GH*1, *GH*2 and *GH*3) with *GH*1 being independent while *GH*2 and *GH*3 were complementary. Kim & Okubo, (1995) had previously observed that high temperature of above 30^0C or/and long photoperiod (more than 13 hours) caused a shift of lablab stem growth habit from determinate to indeterminate type. We therefore suggest that selection for determinate growth habit to be done in early generations and under high temperature conditions (> 30^0C). This is because once identified at early generations, the determinate stem growth habit will be fixed and therefore will breed true to type in later generations. Selection of determinacy under high temperature conditions is useful for picking against those progenies which would express as determinate under low temperature but show indeterminate growth habit under high temperature.

Our results indicated that a single dominant gene was associated with the presence of purple flowers which was consistent with that of Keerthi *et al*, (2016). The segregation of stem anthocyanin in F_2 population fitted to a ratio of 9:7 suggesting the trait is under control of two genes with complementary epistasis. The segregation of the F_2 progenies revealed a monogenic inheritance control for days to flower initiation with the early flowering being recessive to late flowering. Previously, Keerthi *et al.,* (2014) and Keerthi *et al*., (2016) had also reported monogenic control of flowering time in lablab while four non-allelic genes (*ef-1, ef-2, ef-3, ef-4*) were responsible of controlling flowering time in chickpea (Gaur *et al.,* 2015). With the understanding that anthocyanin pigmentation and purple flower colour are completely dominant to absence of the pigmentation and white colour, breeders can use these traits as morphological markers to identify the true F_1 hybrids and remove the selfed plants even at early growing stage.

Distorted segregation from the expected ratio of 9:3:3:1 was observed between stem growth habit and time to flowering of F_2 populations suggesting possible linkage of the trait pairs. The recombination fraction which is an indicator of possible number of synapsis between the traits was small (0.05 and 0.14) further indicating possible linkage. The short (less than 15cM) mapping distance estimated between the two characters further supports a possible linkage. This implies that lablab breeders can indirectly select for early maturing progenies using the more easily observable determinate growth habit trait. The parent KDD used in this study could be a good source of gene for earliness to flowering and determinate growth habit in Kenya. This is because besides having determinate growth habit, it flowered earlier that the local accessions. Integrating this accession into the breeding program can improve the local cultivars for earliness to maturity.

8.4 Genetic variability, heritability, genetic advance and character association of determinate lablab lines

In the present study, investigation was carried out to obtain phenotypic and genetic variability, heritability estimates, expected response to selection and character association of several vegetative and reproductive characteristics of lablab genotypes with determinate growth habit. Forty lablab F_5 lines and an accession with determinate growth habit were evaluated at Kenya Agricultural and Livestock Research Organization, Katumani Centre in Machakos and Kandara Centre in Thika, Kenya during 2017 using a randomized complete block design with two replications. Measures of variability such as genotypic coefficient of variability (GCV), phenotypic coefficient of variability (PCV), broad sense heritability (h2), and genetic advance (GA) based on percentage of the mean were estimated. Heritability (h^2B) expressed as the percentage of the ratio of the genotypic variance (Vg) to the phenotypic variance (Vp) was estimated based on the genotypic mean. Pearson correlation coefficients (*r*) was used to determine inter-relations between the quantitative characters. The direct and indirect relationship of 10 lablab characters, (100 seed weight, pod per plant, raceme per plant, pod length, pod width, pods per raceme, flowering nodes per plant, plant height, days to 50% flowering and days to maturity) with single plant grain yield was determined using path coefficient analysis as demonstrated by Dewey and Lu (1959) and Dabholkar (1992).

Significant differences were observed among the genotypes for grain yield and most of its secondary traits. This suggest that introgression of determinate growth habit into the local accessions can produce diverse progenies from which plant breeders can select those progenies which combine both the determinate growth habit and other desired

traits like high grain yield. Most of the F_5 progenies were early to flower and mature with the mean number of days to flowering ranging from 50-55 days and maturity period ranging from 99-100 days (Table 7.3 and Table 7.4). This indicates that selection of determinate growth habit in lablab can result in selection for early flowering and maturing genotypes. Earliness to flowering and maturity has been reported to be linked to determinate growth habit in some legumes (Keerthi *et al*., 2014; Repinski *et al*., 2012; González *et al*., 2016).

The magnitude of the genotypic variance for seed weight per plant, 100 seed weight, pod length, pod width and days to maturity yield components were consistently higher than their respective environmental variance in both sites. This implies that significant improvement for these traits can be achieved through phenotypic selection (Manggoel *et al*., 2012). Progress from selection depends on the availability of genetic variability in the population and selection is more effective when there is high genetic variation in relation to environmental variation.

Most of the characters recorded moderate broad sense heritability of more than 0.5 at both sites. High heritability coupled with relatively high genetic advance as percent of mean was observed for seed weight per plant, pods per plant, pod width and plant height (Table 7.3 and Table 7.4). This suggest that these traits are under additive gene action and selection based on their phenotypes will be effective. Moderate to high heritability coupled by low genetic advance as percentage of mean was observed for days to 50% flowering and maturity. The low genetic advance of these two maturity related traits could be as a result of the small phenotypic variance of the test materials, influence of non- additive gene action or considerable influence of the environment. This suggests that based on the evaluated determinate lablab lines in this study, days to flowering and

maturity characters can only be partialy improved through phenotypic selection. Other sources of earliness to maturity should be included in the crossing program to create and select transgressive segregants which are earlier maturing than their parents.

Pearson correlation coefficient analysis is one of the most commonly used measure of association between variables. It describes the direction and degree to which one variable is linearly related to one another. Pods number ($r = +0.87$) and raceme number ($r = +0.81$) seems to contribute significantly to seed weight per plant which suggests that indirect selection for grain weight per plant can be effectively realized by selecting for the number of pods and raceme. Other traits with moderate to low significant positive correlation with grain yield per plant include days to 50% flowering, maturity, pod width, plant height and flowering nodes per raceme. This suggests that those lines which, flowered and/or matured late, with longer stem height, many flowering nodes and many racemes and pods produced high grain yield.

Path coefficient analysis revealed that the number of pods per plant had the highest direct positive effect on seed yield per plant. Moderate positive indirect effect on seed yield through pods per plant was detected for raceme per plant, pods per raceme, flowering nodes per plant and plant height. This suggest that selection of high number of units of these characters can be effective in selection of high seed yield in determinate lablab genotypes. The overall positive association (sum of direct and indirect effects) of plant height, days to flowering and maturity and grain yield suggests that there will be a problem when combining dwarfness, earliness to flowering and maturity with high seed yield.

The study has established that

- Introgression of determinate growth habit trait in accession KDD into the local lablab accessions can produce determinate progenies with wide genetic variability.
- The plant characters; number of pods per plant, raceme per plant, plant height, pod width, pods per raceme and number of flowering nodes per plant are useful traits for selection of high yielding determinate lablab based on their high heritability, high genetic advance, direct and indirect effects on seed yield per plant.
- The overall positive association (sum of direct and indirect effects) of plant height, days to flowering and maturity and grain yield suggests that there will be a problem when combining dwarfness, earliness to flowering and maturity with high seed yield.

8.5 Conclusion

- This work has confirmed that it's possible to develop large number of useful molecular markers for underutilized crops such as lablab using the next generation technology.
- It has been confirmed that substantial genetic variability, the foundation of breeding exists in the available local and exotic lablab accessions.
- Majority of the accessions currently cultivated by the farmers in Kenya (Farmers collection) are of narrow genetic diversity and therefore will require to be improved with the materials at the gene bank of Kenya and exotic materials from regions such as Ethiopia and South Africa.

- Wild accessions (sub *species unicinatus*) is a major source of genetic diversity in lablab and should be included in the breeding programs.
- Growth habit in lablab is controlled by three genes, one basic and two complementary genes which are possibly dependent on temperature. Pyramiding determinate growth habit, early flowering, and other desirable characters is possible in development of genotypes with better adaptation to growing environments.
- After hybridisation, identification of high yielding determinate cultivars is possible through selection of high units of number of pods per plant, racemes per plant, plant height, pod width, pods per raceme and number of flower nodes.

8.6 Recommendations

The following recommendations were made:

1. Development of large number of lablab genic SSRs markers using 454 pyrosequencing approach will require at least a full Pico Titer plate run in order to capture and sequence large number of cDNA fragments from where the microsatellites will be mined.
2. The large number and different types of molecular markers developed here could be useful to design a saturated lablab genetic map. To develop diverse mapping population, crossing should be done between accessions belonging to subspecies unicinatus and purpureus.
3. Majority of the accessions currently cultivated by the farmers in Kenya are of narrow genetic diversity and therefore will require to be improved with the materials at the gene bank of Kenya and exotic materials from regions such as Ethiopia and South Africa.

4. Morphological characterization of the the genotypes identified to form the core collection in this study should be done to identify the unique traits in them which could be utilized in improvement of the local lablab cultivars.
5. The subspecies unicinatus is a major source of genetic diversity in lablab and should be included in the breeding programs to develop more adapted cultivars.
6. Further research should be carried out to determine why there was low wild (sub species unicinatus) genotype call rate for most of the DArTSeq SNP markers.
7. The accession KDD is a good source of gene for earliness to flowering and determinate growth habit in Kenya and therefore it's recommended to be integrated into the breeding program to improve the local cultivars for earliness to maturity and stem determinacy.
8. The study confirmed that the genes controlling growth habit and flowering time in lablab are closely linked. Therefore, indirect selection for early maturing progenies using the more easily observable determinate growth habit trait is possible. Pyramiding determinate growth habit, early flowering, and other desirable characters is also possible.
9. Growth habit in lablab was found to be controlled by three genes, one basic and two complementary genes which are environmental dependent. Further investigation on the environmental conditions responsible for shifting determinate to indeterminate growth habit in lablab should be studied in Kenya.
10. Molecular markers linked to growth habit in lablab should be identified to assist in selection of plants with determinate growth habit phenotype during the breeding process.

11. Identification of high grain yielding determinate lablab cultivars is possible through indirect selection of high units of number of pods per plant, racemes per plant, plant height, pod width, pods per raceme and number of flower nodes.
12. The promising F_5 determinate lines need to be advanced through to F_7 and the high yielding lines tested in multiple environments with a view to release the best adapted lines to the farmers

REFERENCES

Acquaah, G. (2015). Conventional Plant Breeding Principles and Techniques. In J.M. Al-Khayri et al. (eds.), Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools. *Springer International Publishing*, Switzerland. https://doi.org/10.1007/978-3-319-22521-0_5

Acquaah, George. (2012). Principles of plant genetics and breeding (2nd Ed.). Oxford, UK: Wiley-Blackwell

Adeboye, K., Oyedeji, O., Alqudah, A., Börner, A., Oduwaye, O., Adebambo, O., & Daniel, I. (2020). Genetic structure and diversity of upland rice germplasm using diversity array technology (DArT)-based single nucleotide polymorphism (SNP) markers. *Plant Genetic Resources:* Characterization and Utilization, 18(5), 343-350. https://doi:10.1017/S1479262120000362

Agriculture Fisheries and Food Authority. (2014). AFFA year book of statistics 2014. Agriculture, Fisheries and Food Authority Tea House. Nairobi, Kenya. 62.

Alam, M., Neal, J., Connor, K. O., Kilian, A., & Topp, B. (2018). Ultra-high-throughput DArTseq-based silicoDArT and SNP markers for genomic studies in macadamia. *PLoS ONE*, 13(8), 1–20.

Ali, F., Sikdar, F., Roy, A.K & Joarder, O. I. (2005). Correlation and genetic variation of twenty different genotypes of Lablab bean, *Lablab purpureus* (L.) sweet. Bangladesh journal of Botany, 34 (2):125-128

Ali, Z. B., Yao, K. N., Odeny, D. A., Kyalo, M., Skilton, R., & Eltahir, I. M. (2015). Assessing the genetic diversity of cowpea [*Vigna unguiculata* (L.) Walp.] Accessions from Sudan using simple sequence repeat (SSR) markers. *African Journal of Plant Science*, 9(7), 293–304. https://doi.org/10.5897/ajps2015.1313

Aljumaili, J., Rafii, M. Y., Latif, M. A., Sakimin, S. Z., Arolu, I. W., & Miah, G. (2018). Genetic Diversity of Aromatic Rice Germplasm Revealed by SSR

Markers. *BioMed Research International*, 2018. https://doi.org/10.1155/2018/7658032

Allard, R.W. (1960), Principles of plant breeding, John Wiley and sons, New York

Al-Snafi, A. Esmail. (2017). the pharmacology and medical importance of Dolichos lablab (Lablab purpureus) - A review. *IOSR Journal of Pharmacy*, 7(2) 22-30. www.iosrphr.org (e)-ISSN: 2250-3013

Amkul, J., Moongkanna, S., & Prakit, S. (2020). Genetic diversity and structure of landrace of lablab (*Lablab purpureus* (L.) Sweet) cultivars in Thailand revealed by SSR markers. *Breeding Science Preview* 1-4. doi: 10.1270/jsbbs.20074

Andrea, A.H., Juliana, P. B., Paula, M. N. & Karina, A. M. (2012). Microsatellite as a tool for genetic diversity analysis, genetic diversity in microorganisms, *Mahmut Caliskan* (Ed.), ISBN: 978-953-51-0064-5.

Arnel, R. Hallauer (2011). Evolution of plant breeding. *Crop Breeding and Applied Biotechnology* 11: 197-206.

Arunga, E. E., Kinyua, M., Ochuodho, J. & Emmy, C. (2015). Genetic diversity of determinate French beans grown in Kenya based on morpho-agronomic and simple sequence repeat variation. *Journal of Plant Breeding and Crop Science*, 7(8), 240–250. https://doi.org/10.5897/jpbcs2015.0503

Asante, I.K., Addey, R., & Carson, G.C. (2009). Correlation and Path-Coefficient Analysis of Component of Agronomic Characters in Lima Bean (Phaseolus lunatus). *Ghana Journal of Science* 49(0):51-56.

Ayalew, H., Genet, T., Dessalegn, T. & Wendale, L. (2011). Multivariate diversity, heritability and genetic advance in tef landraces in Ethiopia. *African Crop Science Journal*, 19(3), 201–212.

Babu, R., Sudha, K., Prasanna, B. & Gupta, H. S. (2004) Integrating marker-assisted selection in crop breeding – Prospects and challenges use of molecular markers to indirectly select for target traits. Current Science, 87 (5).

Badea, A.F., Eudes, Salmon, S., Tuvesson, A., Vrolijik, C.T., Larsson, V., Huttner, E. & Andre, L. (2011). Development and assessment of DArT markers in triticale. *Theor Appl Genet*. 122:1547–1560. DOI 10.1007/s00122-011-1554-3

Becker, J., Vos, P., Kuiper, M., Salamini, F., & Heun, M. (1995) Combined mapping of AFLP and RFLP markers in barley. *Mol. Genet,* 249:65-73.

Bernard, A., Barreneche, T., Lheureux, F., & Id, E. D. (2018). Analysis of genetic diversity and structure in a worldwide walnut (Juglans regia L.) germplasm using SSR markers. *PLoS ONE* 13(11): e0208021. https://doi.org/10.1371/journal.pone.0208021.

Bertan, I., Carvalho, F. De, & Oliveira, A. C. De. (2007). Parental Selection Strategies in Plant *Breeding Programs. Journal of Crop Science and Biotechnology*, 10(4), 211–222.

Bhandari, H.R, Bhanu, A., Srivastava, K., Singh, M.N. Shreya, A.M. (2017). Assessment of Genetic Diversity in Crop Plants - An Overview. *Advances in Plants & Agriculture Research*, 7(3), 279–286.

Bhanu, A. N. (2017). Assessment of Genetic Diversity in Crop Plants - An Overview. *Advances in Plants & Agriculture Research,* 7(3), 279–286. https://doi.org/10.15406/apar.2017.07.00255

Blair, M. W., Andres, J.C. & Doug, R.C. (2012). A high-throughput SNP marker system for parental polymorphism screening, and diversity analysis in common beans (*Phaseolus vulgaris* L.). *Theoretical and applied genetics*, 126, 535-548.

Blair, M.W., Córdoba, J. M., Muñóz, C., & Yuyó, D. K. (2014). BAC-end microsatellites from intra and inter-genic regions of the common bean genome and their correlation with cytogenetic features. *PLoS ONE*, 9(9). https://doi.org/10.1371/journal.pone.0101873

Blair, Matthew W, Hurtado, N., Chavarro, C. M., Muñoz-torres, M. C., Giraldo, M. C., Pedraza, F., & Wing, R. (2011). Gene-based SSR markers for common bean (Phaseolus vulgaris L.) derived from root and leaf tissue ESTs : an integration of the *BMc series*. 1–10.

Boit, L., Kinyua, M. & Kiplagat, O. (2013). Screening dolichos genotypes for resistance to pod borers. *African Crop Science Conference Proceedings*. 11, pp. 471 - 473. Kampala, Uganda: African Crop Science.

Bonginkosi E. M., Terry M. E. & Colin S. E. (2018). Intercropping maize (Zea mays L.) with lablab (Lablab purpureus L.) for sustainable fodder production and quality in smallholder rural farming systems in South Africa. Agroecology and Sustainable Food Systems, 42:4, 362-382. https://doi.org:10.1080/21683565.2017.1393649

Bosire, E., Gitau, W., Karanja, F., & Ouma, G. (2019). Analysis of Climate Variability, Trends and expected implication on crop production in a Semi-Arid Environment of Machakos County, Kenya 1. *Advances in Agricultural Sciences*, 7(01), 99–115. https://doi.org/ISSN: 2588-3801

Broccanello, C., Chiodi, C. & Stevanato, P. (2018). Comparison of three PCR based assays for SNP genotyping in plants. *Plants Methods*, 14(28).

Buckler, E. S., Holland, J. B. & Bradbury, P. J. (2009). The genetic architecture of maize flowering time. *Science*, 325(5941), 714–718.

Burse, N., Thorat, A., Gawande, P. P., Lakhote, S. J., & Tayade, S. D. (2015). Genetic Variability, Heritability and Genetic Advance for Yield Contributing Characters in Okra. 39(July), 41–43.

Cabral, S., Soares, C., Lima, A. & Miranda, F. (2011). Genetic diversity in local and commercial dry bean (*Phaseolus vulgaris*) genotypes based on microsatellite markers. *Genetic and molecular research*, 10(1), 140-149.

Campoy, J. A., Lerigoleur-balsemin, E., Christmann, H., Beauvieux, R., Girollet, N., Quero-garcía, J., & Barreneche, T. (2016). Genetic diversity, linkage disequilibrium, population structure and construction of a core collection of Prunus avium L. Landraces and bred cultivars. BMC *Plant Biology*, 16(49), 1–15. https://doi.org/10.1186/s12870-016-0712-9

Campoy, J. A., Lerigoleur-balsemin, E., Christmann, H., Beauvieux, R., Girollet, N., Quero-garcía, J., … Barreneche, T. (2016). Genetic diversity, linkage disequilibrium, population structure and construction of a core collection of *Prunus avium* L. Landraces and bred cultivars. BMC *Plant Biology*, 16(49), 1–15. https://doi.org/10.1186/s12870-016-0712-9

Chaitanya, V., Reddy, R. V. S. K., & Kumar, P. A. (2014). Variability, heritability and genetic advance in indigenous dolichos bean (*Dolichos lablab* L. var Typicus) genotypes. *Plant Archives*, 14(1), 503–506.

Chapman, M. A. (2015). Transcriptome Sequencing and Marker Development for Four Underutilized Legumes. *Applications in Plant Sciences*, 3(2), 1400111. https://doi.org/10.3732/apps.1400111

Chaudhari, P., Patel, A., Kadam, Y. & Patel, J. (2013). Variability correlation and path analysis study in vegetable Indian bean (*Lablab purpureus* (L.) Sweet). *Crop Res.* 45 (3), 229-236.

Chen, H., Chen, X., Tian, J., Yang, Y., Liu, Z., Hao, X., … Cheng, X. (2016). Development of gene-based ssr markers in rice bean (*Vigna umbellata* L.) based on transcriptome data. *PLoS ONE*, 11(3), 1–13. https://doi.org/10.1371/journal.pone.0151040

Chen, H., Wang, L., Liu, X., Hu, L., Wang, S., & Cheng, X. (2017). De novo transcriptomic analysis of cowpea (*Vigna unguiculata* L. Walp.) for genic SSR marker development. *BMC Genetics*, 18(1), 1–12. https://doi.org/10.1186/s12863-017-0531-5

Chen, H., Wang, L., Wang, S., Liu, C., Blair, M. W., & Cheng, X. (2015). Transcriptome sequencing of mung bean (*Vigna radiate* L.) genes and the identification of EST-SSR markers. *PLoS ONE*, 10(4), 1–15. https://doi.org/10.1371/journal.pone.0120273

Cho, S., Kumar, J., Shultz, J. L., Anupama, K., Tefera, F., & Muehlbauer, F. J. (2002). Mapping genes for double poding and other morphological traits in chickpea. *Euphytica*, 128(2), 285–292. https://doi.org/10.1023/A:1020872009306

Cholastova T, Knotova D. (2012). Using morphological and microsatellite markers to assess the genetic diversity in Alfalfa (Medicago sativa L). *International Journal of Biological, Food ,Veterinary and Agricultural Engineering*, 6(9), 146–152.

Choudhary, S., Sethy, N. K., Shokeen, B., & Bhatia, S. (2009). Development of chickpea EST-SSR markers and analysis of allelic variation across related species. *Theoretical and Applied Genetics,* 118(3), 591–608. https://doi.org/10.1007/s00122-008-0923-z

Cichy, K.A., Porch, G.T., Beaver, S.J., Cregan, P., et al (2015). A *Phaseolus vulgaris* diversity panel for Andean bean improvement. *Plant genetic resources*, 55 (5).

Coulibaly, S., Pasquet, R.S., Papa, R. & Gepts, P. (2002), AFLP organisation analysis of the phonetic organization and genetic diversity of *Vigna unguiculata* L. Walp. Reveals extensive gene flow between wild and domesticated types, *Theoretical and Applied Genetics* 104: 358 – 366

Dabholkar, A.R. 1992. Elements of biometrical genetics. Concept Publishing Company, New Delhi, India.

Daryanto S, Wang L, Jacinthe P. (2015). Global synthesis of drought effects on food legumes production. *PLoS ONE*, e0127401. https://doi.org/:10.1371/journal.pone.0127401

Das, T. S., Durwas, S.V., Dutta, S., Chattopadhyay, A. & Singh, B. (2015). Gene action and combining ability for yield and yield component traits in Dolichos bean (Dolichos Lablab Var. Typicus). *Sabrao Journal of Breeding and Genetics*, 46 (2) 293-304.

Desalegne, B. A., Dagne, K., Melaku, G., Ousmane, B., & A, F. C. (2017). Efficiency of SNP and SSR-based Analysis of Genetic Diversity, Population Structure, and Relationships among Cowpea (Vigna unguiculata (L .) Walp .) Germplasm from East Africa and IITA Inbred Lines. *J. Crop Sci. Biotech*, 20(2), 107–128.

Devmore, P. K., Sawardekar, S., Bhave, S., & Palshetkar, M. (2015) Correlations and path coefficient analysis for yield and yield component in segregating (F4) generation of Lablab bean Lablab purpureus L. sweet. *IJABPT.*, 11 (3), 237-240

Dreher, K. (2000) Is marker assisted selection cost-effective compared to conventional plant breeding methods? The case of Quality Protein Maize. In Proceedings of the 4th Annual Conference of the International Consortium on Agricultural Biotechnology, Ravello, Italy, 2000.

Dutta, S., Kumawat, G., Singh, B. P., Gupta, D. K., Singh, S., Dogra, V., & Singh, N. K. (2011). Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. *BMC Plant Biology*, 11. https://doi.org/10.1186/1471-2229-11-17

Earl, D.E. & Vonholdt, B.M. (2012). STRUCTURE harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method.

Conservation Genetic Resource. 4(2):359-361.http://taylor0.biology.ucla.edu/structureHarvester

Edet, O. U., Gorafi, Y. S. A., Shuhei, N., & Tsujimoto, H. (2018). DArTseq-based analysis of genomic relationships among species of tribe Triticeae. *Scientific Reports*, 8(1), 1–11. https://doi.org/10.1038/s41598-018-34811-y

Egbe, O.M. (2010). Effects of plant density of intercropped soybean with tall sorghum on competitive ability of soybean and economic yield at Otobi, Benue State, Nigeria. *Journal of Cereals and Oilseeds* 1:1-10.

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

Farahani, S., Mojdeh, M., Rahim, M., Homayoun, K., Armin, S., Jacqueline, B. & Reza, T. (2019). Whole Genome Diversity, Population Structure, and Linkage Disequilibrium Analysis of Chickpea (Cicer arietinum L.) Genotypes Using Genome-Wide DArTseq-Based SNP Markers. *Genes,* 10, 676. https://doi.org/10.3390/genes10090676

Fatokun, C.A., Menancio-Hautea, D.I., Danesh, D., & Young, N.D. (1992). Evidence for orthologous seed weight genes in cowpea and mung bean based on RFLP mapping. *Genetics*, 132 3, 841-846

Filipetti, A. (1986) Inheritance of determinate growth habit induced in Vicia faba major by ethyl methane sulfonate (EMS), *Fabis Newsl*, 15, 12-14

Frankham, R., Ballou, J. D., & Briscoe, D. A. (2002). Introduction to Conservation genetics. *Cambridge University Press*, Cambridge doi: 10.1017/CBO9780511808999

Fredrick, J. R, Camp, C. R. & Bauer, P. J. (2001). Drought stress effects on brach and mainstem seed yield and yield components of determinate soyabean. *Crop Science*, 41(3), 759-763.

García, C., Erwan, G. & Arndt, H. (2018). A comparative analysis between SNPs and SSRs to investigate genetic variation in a juniper species (Juniperus phoenicea ssp. turbinata). *Tree Genetics & Genomes*.14 (87). https://doi.org/10.1007/s11295-018-1301-x

Garcia, R. V., Rangel, P. N., Brondani, C., Martins, W. S., Melo, L. C., Carneiro, M. S., … Brondani, R. P. V. (2011). The characterization of a new set of EST-derived simple sequence repeat (SSR) markers as a resource for the genetic analysis of *Phaseolus vulgaris*. *BMC Genetics*, 12(May). https://doi.org/10.1186/1471-2156-12-41

Gaur, P. M., Samineni, S., Tripathi, S., Varshney, R. K., & Gowda, C. L. L. (2015). Allelic relationships of flowering time genes in chickpea. *Euphytica*, 203(2), 295–308. https://doi.org/10.1007/s10681-014-1261-7

Gelalcha, S., & Hanchinal, R.R. (2013). Correlation and path analysis in yield and yield components in spring bread wheat (*Triticum aestivum* L.) genotypes under irrigated condition in Southern India. *African Journal of Agricultural Research*, 8(24), 3186–3192. https://doi.org/10.5897/AJAR2012.6965

Gepts, P. (2006). Plant genetic resources conservation and utilization: The accomplishments and future of a societal insurance policy. *Crop Science*, 46, 2278 – 2292.

Gimode D., Odeny D. A., de Villiers E.P., Wanyonyi S., Dida M.M., Mneney E.E., et al. (2016). Identification of SNP and SSR Markers in Finger Millet Using Next Generation Sequencing Technologies. *PLoS ONE* 11(7): e0159437. https://doi.org/10.1371/journal.pone.0159437

Girish, G. & Gowda, M. B. (2009) Inheritance of qualitative characters in dolichos bean *Lablab purpureus* L. Sweet. *Environment and Ecology*, 27 (2). 571-580

Govindaraj, M., Vetriventhan, M., & Srinivasan, M. (2015). Importance of genetic diversity assessment in crop plants and its recent advances: An overview of its analytical perspectives. *Genetics Research International*, https://doi.org/10.1155/2015/431487

Govindaraj, M., Vetriventhan, M., & Srinivasan, M. (2015). Importance of genetic diversity assessment in crop plants and its recent advances: An overview of its analytical perspectives. Genetics Research International, 2015 https://doi.org/10.1155/2015/431487

Gowda, M. B. (2008). Genetic enhancement of Dolichos bean through integration of conventional breeding and molecular approaches. Kirkhouse funded annual progress report April 2007-March 2008. University of Agric. Sci. Bangalore.

Graffelman, J., & Weir, B. S. (2016). Testing for Hardy-Weinberg equilibrium at biallelic genetic markers on the X chromosome. *Heredity*, 116(6), 558–568. https://doi.org/10.1038/hdy.2016.20

Grahnen, J. A., Benkman, C. W., Buerkle, C. A., Geist, K. S., & Parchman, T. L. (2010). Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC Genomics*, 11(1), 180. https://doi.org/10.1186/1471-2164-11-180

Greenbaum, G., Templeton, A. R., Zarmi, Y., & Bar-David, S. (2014). Allelic richness following population founding events - A stochastic modeling framework incorporating gene flow and genetic drift. *PLoS ONE*, 9(12), 1–23. https://doi.org/10.1371/journal.pone.0115203

Guo-Liang, J. (2013). Plant Breeding from Laboratories to Fields: Molecular Markers and Marker-Assisted Breeding in Plants http://dx.doi.org/10.5772/52583

Gupta P.K., Varshney R.K., Sharma P.C. and Ramesh B. (2000), Molecular markers and their applications in wheat breeding, *Plant breeding*, 118, 369-390

Gupta, S., & Prasad, M. (2009). Development and characterization of genic SSR markers in Medicago truncatula and their transferability in leguminous and non-leguminous species. *Genome*, 52(9), 761–771. https://doi.org/10.1139/g09-051

Hailu, A., Alamerew, S., Nigussie, M., & Assefa, E. (2016). Genetic Variability, Heritability and Genetic Advance for Yield and Yield Related Traits in Barley (Hordeum vulgare L.) Germplasm Oxfam America, Horn of Africa Regional Office. World Journal of Agricultural Sciences, 12(1), 36–44. https://doi.org/10.5829/idosi.wjas.2016.12.1.1894

Hailu, A., Alamerew, S., Nigussie, M., & Assefa, E. (2016). Genetic variability, heritability and genetic advance for yield and yield related traits in barley (Hordeum vulgare L.) germplasm. *World Journal of Agricultural Sciences*, 12(1), 36–44. https://doi.org/10.5829/idosi.wjas.2016.12.1.1894

Hamblin MT, Warburton ML, & Buckler ES (2007) Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. *PLoS ONE* 2:e1367

Hanai, L. R., de Campos, T., Camargo, L. E. A., Benchimol, L. L., de Souza, A. P., Melotto, M., … Vieira, M. L. C. (2007). Development, characterization, and comparative analysis of polymorphism at common bean SSR loci isolated from genic and genomic sources. *Genome*, 50(3), 266–277. https://doi.org/10.1139/g07-007

Harland, S.C. (1920). Inheritance in Dolichos lablab. *J. genetics* 10:219-226.

Hassan, S. F., Iqbal, M. S., Rabbani, G., Riaz, M., & Noorka, I. R. (2013). Correlation and path analysis for yield and yield components in sunflower (Helianthus annus. L). *African Journal of Biotechnology*, 12(16), 1968–1971. https://doi.org/10.5897/AJB12.817

Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics,* 107(1), 1–8. https://doi.org/10.1016/j.ygeno.2015.11.003

Hiremath, P. J., Farmer, A., Cannon, S. B., Woodward, J., Kudapa, H., Tuteja, R., & Varshney, R. K. (2011). Large-scale transcriptome analysis in chickpea (Cicer arietinum L.), an orphan legume crop of the semi-arid tropics of Asia and Africa. *Plant Biotechnology Journal*, 9(8), 922–931. https://doi.org/10.1111/j.1467-7652.2011.00625.x

Hiz, M. C., Canher, B., Niron, H., & Turet, M. (2014). Transcriptome analysis of salt tolerant common bean (*Phaseolus vulgaris* L.) under saline conditions. *PLoS ONE*, 9(3). https://doi.org/10.1371/journal.pone.0092598

Holland, J. B., Nyquist, W. E., & Cervantes-Martínez, C. T. (2010). Estimating and Interpreting Heritability for Plant Breeding: An Update. *Plant Breeding Reviews* 22 (1) 1-10. https://doi.org/10.1002/9780470650202.ch2

Holsinger, K. E., Weir, B. S., & Biology, E. (2009). Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nat Rev Genet*., 10(9), 639–650. https://doi.org/10.1038/nrg2611.

Inouye, J., Shanmugasundaram, S. & Masuyama, T. (1979). Effects of temperature and daylength on the flowering of some photo insensitive soybean varieties. *Japan. J. Trop. Agr*. 22, 167-171.

Jaccoud, D., Peng, K.M., Feinstein, D. & Kilian, A. (2001) Diversity Arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Research*.; 29(4):e25. PMID: 11160945

Jayashree, B., Punna, R., Prasad, P., Bantte, K., Hash, C. T., Chandra, S., … Varshney, R. K. (2006). A database of simple sequence repeats from cereal and legume expressed sequence tags mined in silico: survey and evaluation. In *Silico Biol*, 6(6), 607–620.

Jiaqin, S., Ruiyuan, L., Dan, Q., Congcong, J., Yan, L., Morgan, C., & Jinling, M. (2009). Unraveling the complex trait of crop yield with quantitative trait loci mapping in Brassica napus. *Genetics*, 182(3), 851–861.

Julius P. S., Siraj I. K., Abasi K., Muhammad K., Clementine N., Kato W., James B., Allen M. & Swidiq M. (2021). Genome-wide diversity and structure variation among lablab (Lablab purpureus (L.) Sweet) accessions and their implication in a Forage breeding program. *Genet Resour Crop Evol.* https://doi.org/10.1007/s10722-021-01171-y

Kalia, R. K., Rai, M. K., Kalia, S., Singh, R., & Dhawan, A. K. (2011). Microsatellite markers: An overview of the recent progress in plants. Euphytica, 177(3), 309–334. https://doi.org/10.1007/s10681-010-0286-9

Kalia, R. K., Rai, M. K., Kalia, S., Singh, R., & Dhawan, A. K. (2011). Microsatellite markers: An overview of the recent progress in plants. *Euphytica*, 177(3), 309–334. https://doi.org/10.1007/s10681-010-0286-9

Kamenya, S. Ndagire; Erick, O. Mikwa; Bo Song & Damaris A. Odeny. (2021). Genetics and breeding for climate change in Orphan crops. *Theoretical and Applied Genetics* https://doi.org/10.1007/s00122-020-03755-1

Kamotho, G. N. (2015). Evaluation of adaptability potential and genetic diversity of Kenyan dolichos bean (*Lablab purpureus* (L.) sweet) germplasm. PhD thesis, School of Agriculture, University of Eldoret.

Kamotho, G. N., Kinyua, M. G., Muasya, R. M., Gichuki, S. T., & Wanjala, B. W. (2016). Assessment of Genetic Diversity of Kenyan Dolichos Bean (Lablab purpureus L. Sweet) Using Simple Sequence Repeat (SSR) Markers. International Journal of Agriculture, Environment and Bioresearch, 1(01), 26–43.

Kamotho, G. N., Kinyua, M. G., Muasya, R. M., Gichuki, S. T., Kamau, E.M & Wanjala, B. W. (2016). Assessment of Genetic Diversity of Kenyan Dolichos Bean (*Lablab purpureus* L. Sweet) Using Simple Sequence Repeat (SSR) Markers. *International Journal of Agriculture, Environment and Bioresearch*, 1(01), 26–43.

Kamotho, G.N., Kinyua, M.G., Muasya, R.M., Orwa, D.O., & Kimani, E.N. (2010). A baseline survey on production, Utilization and Marketing Constraints of Lablab bean: Impact on Lablab Bean Improvement in Kenya, *International Journal of Professional Pactice*, Vol.1, pp. 21-29.

Kaur, S., Cogan, N. O. I., Pembleton, L. W., Shinozuka, M., Savin, K. W., Materne, M., & Forster, J. W. (2011). Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigene assembly and SSR marker discovery. *BMC Genomics*, 12(June). https://doi.org/10.1186/1471-2164-12-265

Keerthi, C. M., Ramesh, S., Byregowda, M., & Vaijayanthi, P. V. (2018). Simple Sequence Repeat (SSR) Marker Assay-Based Genetic Diversity among Dolichos Bean (Lablab purpureus L. Sweet) Advanced Breeding Lines Differing for Productivity per se Traits. International Journal of Current Microbiology and Applied Sciences, 7(05), 3736–3744.

Keerthi, C. M., Ramesh, S., Byregowda, M., & Vaijayanthi, P. V. (2018). Simple Sequence Repeat (SSR) Marker Assay-Based Genetic Diversity among Dolichos Bean (Lablab purpureus L. Sweet) Advanced Breeding Lines Differing for Productivity per se Traits. *International Journal of Current Microbiology and Applied Sciences*, 7(05), 3736–3744. https://doi.org/10.20546/ijcmas.2018.705.433

Keerthi, C. M., Ramesh, S., Byregowda, M., Rao, A. M., & Prasad, B. R. (2016). Further evidence for the genetic basis of qualitative traits and their linkage relationships in dolichos bean (Lablab purpureus L.). *J. Genet.*, 95(1), 89–98.

Keerthi, C. M., Ramesh, S., Byregowda, M., Rao, A. M., Rajendra Prasad, B. S., & Vaijayanthi, P. V. (2014). Genetics of growth habit and photoperiodic response to flowering time in dolichos bean (*Lablab purpureus* (L.) Sweet). *Journal of Genetics,* 93(1), 203–206. https://doi.org/10.1007/s12041-014-0336-5

Keerthi, C.M., Ramesh, S., Byregowda, M., Rao, A.M., & Prasad, B.R. (2016). Further evidence for the genetic basis of qualitative traits and their linkage relationships in dolichos bean (*Lablab purpureus* L .). J. Genet., 95(1), 89–98.

Keerthi, C.M., Ramesh, S., Byregowda, M., Rao, A.M., Rajendra Prasad, B.S., & Vaijayanthi, P. V. (2014). Genetics of growth habit and photoperiodic response to flowering time in dolichos bean (Lablab purpureus (L.) Sweet). *Journal of Genetics*, 93(1), 203–206. https://doi.org/10.1007/s12041-014-0336-5

Kenya Plant Health Inspectorate Services (KEPHIS). (2015). National crop variety list- Kenya. pp 76-80 at http://*www.kephis.org*

Kgasago, H. (2006). Effects of planting dates and density on yield and yield components of short and ultra-short growth period maize (Zea mays L.), Masters Dissertation.. University of Pretoria, South Africa.

Kilian, A., Sanewski, G., & Ko, L. (2016). The application of DArTseq technology to pineapple. *Acta Horticulturae,* 1111, 181–188.

Kilonzi, S. M., Makokha, A. O., & Kenji, G. M. (2017). Physical characteristics, proximate composition and anti-nutritional factors in grains of lablab bean (Lablab purpureus) genotypes from Kenya. *Journal of Applied Biosciences*, 114(1), 11289. https://doi.org/10.4314/jab.v114i1.2

Kim, K. S., Bellendir S., K. & Hudson, A. (2010) Fine mapping the soybean aphid resistance gene Rag1 in soybean, *Theoretical and Applied Genetics,* vol. 120, no. 5, pp. 1063–1071.

Kim, S. E., & Okubo, H. (1995). Control of growth habit in determinate lablab bean (Lablab purpureus) by temperature and photoperiod. *Scientia Horticulturae*, 61(3–4), 147–155. https://doi.org/10.1016/0304-4238(94)00740-7

Kimani, E. N., Wachira, F. N., & Kinyua, M. G. (2012). Molecular Diversity of Kenyan Lablab Bean (Lablab purpureus (L.) Sweet) Accessions Using Amplified Fragment Length Polymorphism Markers. *American Journal of Plant Sciences*, 3, 313–321. https://doi.org/10.4236/ajps.2012.33037

Kinyua, M.G., & Kiplagat, O.L. (2012). Lablab (Lablab purpureus L. Sweet) bean Improvement using mutation and biotechnological techniques. *Dansten Agencies*, Nairobi, Kenya.

Koinange, E. M. K., Singh, S. P., & Gepts, P. (1996). Genetic control of the domestication syndrome in common bean. *Crop Science*, 36(4), 1037–1045. https://doi.org/10.2135/cropsci1996.0011183X003600040037x

Kosman, E. & Leonard, K. J. (2005). Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploid species. *Molecular Ecology* 14, 415–424. doi: 10.1111/j.1365-294X.2005.02416.x

Kovach W. (2006). MVSP-Multivariate Statistical Package Version 3.1. Kovach Computing Services: Anglesey, Wales

Krishna, B. Meera., Munawwar, A. Khan. & Shams, T. Khan (2019). Next-Generation Sequencing (NGS) Platforms: An Exciting Era of Genome Sequence Analysis. In V. Tripathi et al. (eds.), *Microbial Genomics in Sustainable Agroecosystems*, Vol 2. pp 89-108, Singapore, Springer nature.

Kumar V., Sharmar S., Sharma K.A., Sharma S. And Bhat V.K. (2009), Comparative analysis of diversity based on morpho- agronomic traits and microsatellite markers in common bean, *Euphtyica* 170: 249 -262

Kwak, M., Toro, O., Debouck, D. G., & Gepts, P. (2012). Multiple origins of the determinate growth habit in domesticated common bean (*Phaseolus vulgaris*). *Annals of Botany*, 110(8), 1573–1580. https://doi.org/10.1093/aob/mcs207

Kyratzis, A.C., Nikoloudakis, N, & Katsiotis A. (2019). Genetic variability in landraces populations and the risk to lose genetic variation. The example of landrace 'Kyperounda' and its implications for ex situ conservation. *PLoS ONE* 14(10): e0224255. doi:10.1371/journal.pone.0224255

Laird, N.M. & Lange, C. (2011). The Fundamentals of Modern Statistical Genetics. Springer Science & Business Media, New Yolk.

Lamu County Government (2018). Lamu County integrated development plan 2018-2022. LamuCounty.https://academia-ke.org/library/download/lamu-county-integrated-development-plan-fy-2018-2022/. Accessed on 24th May 2021.

Laxmi, K., Vaijayanthi, P.V., Keerthi, C.M., Shivakumar, M.S., Ramesh, S. & Mohan, R. A. (2015). Genotype-dependent photoperiod-induced sensitivity to flowering time in Dolichos bean (Lablab purpureus L.) sweet var. lignosus. *Banglandesh Journal of Botany*, 45 (3): 471-476.

Li, Y. C., Korol, A. B., Fahima, T., Beiles, A., & Nevo, E. (2002). Microsatellites: Genomic distribution, putative functions and mutational mechanisms: A review. *Molecular Ecology*, 11(12), 2453–2465. https://doi.org/10.1046/j.1365-294X.2002.01643.x

Li, Y., & Zhang, W. (2015). Isolation and Characterization of Microsatellite Markers for Jasminum sambac (Oleaceae) Using Illumina Shotgun Sequencing. *Applications in Plant Sciences*, 3(10), 1500063.

Liu K, Muse SV. (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*. 21: 2128–2129. PMID: 15705655.

Liu, C. J. (1996). Genetic diversity and relationships among Lablab purpureus genotypes evaluated using RAPD as markers. *Euphytica* 90(1), 115–119. https://doi.org/10.1007/BF00025167

Liu, C., Fan, B., Cao, Z., Su, Q., Wang, Y., Zhang, Z., … Tian, J. (2016). A deep sequencing analysis of transcriptomes and the development of EST-SSR markers in mungbean (Vigna radiata). *Journal of Genetics*, 95(3), 527–535. https://doi.org/10.1007/s12041-016-0663-9

Liu, C., Fan, B., Cao, Z., Su, Q., Wang, Y., Zhang, Z., … Tian, J. (2018). A deep sequencing analysis of transcriptomes and the development of EST-SSR markers in mungbean (Vigna radiata). *Journal of Genetics,* 95(3), 527–535. https://doi.org/10.1007/s12041-016-0663-9

Liu, C.J. (1996) Genetic diversity and relationships among Lablab purpureus genotypes evaluated using RAPD as markers. *Euphytica* 90: 115–119.

Liu, Y., Zhonghu, H. & Xianchun, X. (2012). Functional markers in wheat: current status and future prospects. *Theoretical and applied Genetics*, 125, 1-10.

Luby, J. James & Douglas, V. Shaw. (2000). Does Marker-assisted Selection Make Dollars and Sense in a Fruit Breeding Program? *Hortisci*. 36(5): 87.

Lucia, M., Vieira, C., Santini, L., Diniz, A. L., & Munhoz, C. D. F. (2016). Microsatellite markers : What they mean and why they are so useful Microsatellite markers : what they mean and why they are so useful. *Genetics and Molecular Biology*, 39(3), 312–328. https://doi.org/10.1590/1678-4685-GMB-2016-0027

Luke, W., Noel, O. I., Keith, W., White, J., Kaur, S., Pembleton, L. W., … Paull, J. (2012). Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. *BMC Genomics*, 13(1), 104. https://doi.org/10.1186/1471-2164-13-104

Luo Chang; Chen Dongliang; Cheng Xi; Liu Hua; Li Yahui & Huang Conglin. (2018). SSR Analysis of genetic relationship and classification in Chrysanthemum germplasm collection. *Hort. Plant journal* 4 (2) 72-83. https://doi.org/10.1016/j.hpj.2018.01.003

Luo, Z., Brock, J., Dyer, J. M., Kutchan, T., Schachtman, D., Augustin, M., … Abdel-Haleem, H. (2019). Genetic diversity and population structure of a Camelina sativa spring panel. *Frontiers in Plant Science*, 10 (February), 1–12. https://doi.org/10.3389/fpls.2019.00184

Ma, S., Dong, W., Lyu, T., & Lyu, Y. (2019). An RNA Sequencing Transcriptome Analysis and Development of EST-SSR Markers in Chinese Hawthorn through Illumina Sequencing. *Forests*, 10(2), 82. https://doi.org/10.3390/f10020082

Maass, B. L. (2016). Domestication, origin and global dispersal of *Lablab purpureus* (L.) Sweet (Fabaceae): current Origin , domestication and global dispersal of *Lablab purpureus* (L .) Sweet (Fabaceae): Current understanding. *Legume Persipectives*, (13), 5–8.

Maass, B. L., & Usongo, M. (2007). Changes in seed characteristics during the domestication of the lablab bean (*Lablab purpureus* (L.) Sweet : Papilionoideae) Changes in seed characteristics during the domestication of the lablab bean (*Lablab purpureus* (L .) Sweet : Papilionoideae). *Australian Journal of Agricultural Research,* 58(January), 9–19. https://doi.org/10.1071/AR05059

Maass, B. L., Jamnadass, R. H., Hanson, J., & Pengelly, B. C. (2005). Determining sources of diversity in cultivated and wild *Lablab purpureus* related to provenance of germplasm by using amplified fragment length polymorphism. *Genetic Resources and Crop Evolution*, 52(6), 683–695. https://doi.org/10.1007/s10722-003-6019-3

Maass, B. L., Jamnadass, R. H., Hanson, J., & Pengelly, B. C. (2005). Determining sources of diversity in cultivated and wild Lablab purpureus related to provenance of germplasm by using amplified fragment length polymorphism. *Genetic Resources and Crop Evolution*, 52(6), 683–695. https://doi.org/10.1007/s10722-003-6019-3

Maass, B. L., Knox, M. R., Venkatesha, S. C., & Pengelly, B. C. (2010). Lablab purpureus — *A Crop Lost for Africa ? Tropical Plant Biol*., 1, 1–13. https://doi.org/10.1007/s12042-010-9046-1

Machikowa, T., & Laosuwan, P. (2011). Path coefficient analysis for yield of early maturing soybean. *Songklanakarin Journal of Science and Technology*, 33(4), 365–368.

Maciel, F.L., Echeverrigaray, S., Gerald, L.T. & Gobbi-Grazziotin, F. (2003) Genetic relationships and diversity among Brazilian cultivars and landraces of common beans (*Phaseolus vulgaris* L.) revealed by AFLP markers. *Gen. Res. Crop Evol*. 50, 887-893.

Magalingam, V., Mohammed, Y., Kumar, S & Ramesh, S. (2013) Genetic variability and character association in dolichos bean. *SAARC J. Agri*., 11(2), 161-171

Makgoga, M. W. (2014). Influence of Lablab (Lablab purpureus) and dry bean (Phaseolus vulgaris) intercrops with maize (Zea mays) on maize grain yield and soil fertility status, PhD Thesis, University of Limpopo South Africa.

Manggoel, W., Uguru, M. I., Ndam, O. N., & Dasbak, M. A. (2012). Genetic variability, correlation and path coefficient analysis of some yield components of ten cowpea [Vigna unguiculata (L.) Walp] accessions. *Journal of Plant Breeding and Crop Science*, 4(5), 80–86. https://doi.org/10.5897/JPBCS12.007

Manifesto, M.M., Schlatter, A.R. & Hopp, H.E. (2001) Quantitative evaluation of genetic diversity in wheat germplasm using molecular markers. Crop Science 41(3):682–690

Manjeru, P., Madanzi, T., Makeredza, B., Nciizah, A. & Sithole, M. (2007). Effect of water stress at different growth stages on components and grain yield of common bean (*Phaseolus vulgaris*). *African crop science proceedings*, 8, 299-303.

Massawe F.J., Robert J.A., Azam-Ali S.N. & Davey M. R. (2003), Genetic diversity in Bambara groundnut (Vigna subterranea (L.) Verdc) landraces assessed by Random Amplified Polymorphic DNA (RADPs) markers, Genetic Resources and Crop Evolution 50: 737-741

Massawe F.J., Robert J.A., Azam-Ali S.N. & Davey M. R. (2003), Genetic diversity in Bambara groundnut (Vigna subterranea (L.) Verdc) landraces assessed by

Random Amplified Polymorphic DNA (RADPs) markers, *Genetic Resources and Crop Evolution* 50: 737-741

Massawe, F.J., Dickinson, M., Robert, J.A. & Azam-Ali S.N. (2002), Genetic diversity in Bambara groundnut (Vigna subterranea (L.) Verdc) landraces revealed by AFLP markers. *Genome* 45, 1175 – 1180

Matthew, N. N., Jens D. B. & William, E. (2010). Flowering time control in annual legumes: prospects in a changing. CAB Reviews: *Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources*, 5, 1-15. DOI: 10.1079/PAVSNNR20105017

Meru County Government (2018). Meru County integrated development plan, 2018-2022. Meru.

Ministry of Agriculture and livestock (MOAL) (2018). Monthly wholesale average prices of scheduled food crops January-November 2018. Ministry of Agriculture.

Ministry of Agriculture Annual Report, (2014). Annual Crops Report for 2014. Ministry of Agriculture, Kenya.

Mir, R.R., Rachit, K.S., Kulbhushan, S., Hari, U., Andrezej, K., Douglas, R.C. and Rajeev, K.V. (2012). Whole-genome scanning for mapping determinacy in Pigeonpea (Cajanus spp.*). Plant Breeding* 1: doi:10.1111/j.1439-0523.2012.02009.x

Moe, K. T., Chung, J. W., Cho, Y. Il, Moon, J. K., Ku, J. H., Jung, J. K., … Park, Y. J. (2011). Sequence information on simple sequence repeats and single nucleotide polymorphisms through transcriptome analysis of mungbean. *Journal of Integrative Plant Biology*, 53(1), 63–73. https://doi.org/10.1111/j.1744-7909.2010.01012.x

Moe, K. T., Hong, W. J., Kwon, S. W., & Park, Y. J. (2012). Development of cDNA-derived SSR markers and their efficiency in diversity assessment of

Cymbidium accessions. *Electronic Journal of Biotechnology*, 15(2), 40–61. https://doi.org/10.2225/vol15-issue2-fulltext-4

Molosiwa, O. O. (2012). Genetic diversity and population structure analysis of Bambara groundnuts (Vigna subterranea (L.) Verdc.) Landraces using morpho- agronomic characters and SSR markers, PhD Thesis, University of Nottingham, Sunton Bonigham Campus.

Moushree, S., Sanhita, G., & Kundagrami, S. (2014). Genetic variability and character association of yield and yield components in Mungbean [vigna radiate (L.) Wilczek). *Journal Agroecological and natural resource management*, Vol. 1(3) 161-165.

Muhammad, A., Bakahsh, A. & Abdul G. (2004). Path coefficient analysis in Chickpea (*Cicer arientinum* L.) under rainfed conditions. *Pak. J. Bot*., 36(1), 75-81.

Murfet, I. C. (1973). Flowering in pisum. Hr, a gene for high response to photoperiod. *Heredity*, 31(2), 157–164.

Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., … Baloch, F. S. (2018). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnology and Biotechnological Equipment*, 32(2), 261–285.

Naeema M., Asfia S., Abid A., Tariq A., Masroor A. K. & Moin U. (2021). Hyacinth bean (Lablab purpureus L.) – An underutilised crop with future potential. *Scientia Hortic*., 272:1. https://doi.org/10.1016/j.scienta.2020.109551

Ndjiondjop, M. N., Semagn, K., Gouda, A. C., Kpeki, S. B., Tia, D. D., Sow, M., … Warburton, M. L. (2017). Genetic variation and population structure of Oryza glaberrima and development of a mini-core collection using DArTseq. *Frontiers in Plant Science*, 8(October), 1–19. https://doi.org/10.3389/fpls.2017.01748

Nemlı, S., Kaygisiz Aşçioğul, T., Ateş, D., Eşıyok, D., & Tanyolaç, M. B. (2017). Diversity and genetic analysis through DArTseq in common bean (Phaseolus vulgaris L.) germplasm from Turkey. *Turkish Journal of Agriculture and Forestry*, 41(5), 389–404. https://doi.org/10.3906/tar-1707-89.

Nepomuscene N., Richard N. Onwonga, Rolf Sommer, Clare M. Mukankusi, John Mburu & Jean C. R. (2017) Effect of excessive and minimal soil moisture stress on agronomic performance of bush and climbing bean (*Phaseolus vulgaris* L.), *Cogent Food & Agriculture*, 3:1. https://doi.org/ 10.1080/23311932.2017.1373414

Njunie, M. N. (2002). Evaluation of forage legumes for soil fertility improvement in maize/cassava, (PhD Thesis). North Carolina State University North Carolina.

Noorjahan, A.M., Deshmukh, J.D., Bagade, A.B., Kalpande, H.V & Vijay, R. (2019). Path analysis studies in dolichos bean (Lablab purpureus L.). J*ournal of pharmacognosy and phytochemistry*, 8(2), 41-42.

Nord A, Miller NR, Mariki W, Drinkwater L and Snapp S. (2020). Investigating the diverse potential of a multi-purpose legume, *Lablab purpureus* (L.) Sweet, for smallholder production in East Africa. *PLoS ONE* 15(1). https://doi.org /10.1371/0227739

Nurmansyah S., Alghamdi S., Hussein M. M., Muhammad A. K. & Muhammad A. (2020). AFLP-Based Analysis of Variation and Population Structure in Mutagenesis Induced Faba Bean. *Diversity* 12(303). https://doi.org/10.3390/d12080303

Oduori, C. O. A (2008). Breeding investigations of finger millet characteristics including blast disease and Striga resistance in Western Kenya (PhD thesis). University of KwaZulu-Natal, South Africa.

Onyango, S. Omondi. (2011).The potential for njahi (Lablab purpureus L.) in improving consumption adequacy for protein, iron and zinc in households: A case for Nandi south district, Kenya. (MSc Thesis). University of Nairobi, Kenya.

Ozveren, D., Anlarsal, A., & Yucel, C. (2006). Genetic variability correlation and Path analysis of yield and yield components in Chickpea (*Cicer arietinum* L.). *Turk J. Agric* 30: 183-188

Parmar, A., Singh, A., Dhillon, P. & Jamwal, M. (2013). Genetic diversity of morphological and yield traits in Dolichos bean (*Lablab purpureus* L.). *African Journal of Agric. Res*. 8 (12), 1022-1027.

Pasquet R.S. and Padulosi S. (2013). Taxonomy : Current Status and Prospects. *Innovative Research*, 66–87.

Patel, K., Nandan, M., Sharma, G.L. & Sarnaik, D.A. (2011). Genetic divergence analysis in dolichos bean (Dolichos lablab L.) for Chhattisgarh plains. *Vegetable Science* 38(2): 241-242

Pawar, R. and Prajapati, R. (2013). Genetic variability correlation and path analysis in Indian bean (Lablab purpureus L. sweet). *International Journal of Agric. Sci*. Vol 9 (2), 615-619.

Peeta, G., Satyanarayana, A., Rama Krishna, A. & Krs, S. (2017). Studies on the inheritance of bush habit and photo insensitivity in garden bean (*Lablab purpureus*). *Journal of Crop Research and Fertilizers*, 2, 2–4.

Pengelly, B. C., & Maass, B. L. (2001). *Lablab purpureus* (L.) Sweet – diversity, potential use and determination of a core collection of this multi-purpose tropical legume. *Genetic resource & crop evolution*, 261–272.

Phippen, W. B. (2000). Anthocyanin inheritance and instability in purple basil (*Ocimum basilicum* L.). *Journal of Heredity*, 91(4), 289–296. https://doi.org/10.1093/jhered/91.4.289

Qi, W., Xue-Mei, J; Chao-Chao, Y., Wan-Qing, Z., Guo-Sheng, X., Bi-Song, Y. & Cai-Quan, Z. (2018) Distribution patterns and variation analysis of simple sequence repeats in different genomic regions of bovid genomes. *Scientific Reports* 8:14407. DOI: 10.1038/s41598-018-32286-5.

Raatz, B., Mukankusi, C., David, J., Male, A., Chisale, V., Amsalu, B. et al. (2019). Analyses of African common beans (Phaseolus vulgaris) germplasm using a SNP fingerprinting platform: diversity, quality control and molecular breeding. *Genetic resources and crop evolution*, 66, 707-722.

Raghu, B.R., Samuel, D.K., Mohan, N. & Aghora, T.S. (2018). Dolichos bean: An underutilized and unexplored crop with immense potential. International Journal of recent advances in multidisciplinary research, 5(12), 4338-4341.

Repinski, S. L., Kwak, M., & Gepts, P. (2012). The common bean growth habit gene PvTFL1y is a functional homolog of Arabidopsis TFL1. *Theoretical and Applied Genetics*, 124(8), 1539–1547. https://doi.org/10.1007/s00122-012-1808-8

Ribaut, J.M., Vicente, M.C. & Delannay, X. (2010). Molecular breeding in developing countries: challenges and perspectives. *Current Opinion in Plant Biology*, 13,1-6.

Robotham, O., & Chapman, M. (2015). Population genetic analysis of hyacinth bean (Lablab purpureus (L.) Sweet, Leguminosae) indicates an East African origin and variation in drought tolerance. *Genetic Resources and Crop Evolution*, 64(1), 139–148. https://doi.org/10.1007/s10722-015-0339-y

Roorkiwal, M., Wettberg, E.J., Upadhyaya, H.D., Warschefsky, E., Rathore, A. (2014). Exploring germplasm diversity to understand the domestication process in Cicer spp. using SNP and DArT Markers. *PLoS ONE* 9(7): e102016. doi:10.1371/journal.pone.0102016

Roy, S., Karim, M., Islam, A., Bari, Md., Miani, M & Tetsushi, H. (2006). Relationship between yield and its component characters on bush beans (Phaseolus vulgaris L.). *South pacific studies,* 27 (1).

SaÂnchez-Sevilla, J.F, Horvath A, Botella MA, Gaston A, Folta K, Kilian A, et al. (2015) Diversity Arrays Technology (DArT) marker platforms for diversity analysis and linkage mapping in a complex crop, the octoploid cultivated strawberry (Fragaria× ananassa*). PLoS One*. 2015; 10(12):e0144960. https://doi.org/10.1371/journal.pone.0144960 PMID: 26675207

Saitou, N., & Nei, M. (1987). The Neighbor-joining Method : A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution*, 4(4), 406–425.

Salim, M., Hossain, S., Alam, S., Rashid, J., & Islam, S. (2014). Variability, correlation and path analysis in lablab bean (Lablab purpureus L.). *Bangladesh Journal of Agricultural Research*, 38(4), 705–717. https://doi.org/10.3329/bjar.v38i4.19662

Sarker, A., & Erskine, W. (2006). Recent progress in the ancient lentil. *Journal of Agricultural Science*, 144(1), 19–29. https://doi.org/10.1017/S0021859605005800

Sathyanarayana, N., Pittala, R. K., Tripathi, P. K., Chopra, R., Singh, H. R., Belamkar, V., … Egan, A. N. (2017). Transcriptomic resources for the medicinal legume Mucuna pruriens: De novo transcriptome assembly, annotation, identification and validation of EST-SSR markers. *BMC Genomics*, 18(1). https://doi.org/10.1186/s12864-017-3780-9

Satya, P., Paswan, P. K., Ghosh, S., Majumdar, S., & Ali, N. (2016). Confamiliar transferability of simple sequence repeat (SSR) markers from cotton (Gossypium hirsutum L.) and jute (Corchorus olitorius L.) to twenty two Malvaceous species. *Biotech*, 6(1), 1–7. https://doi.org/10.1007/s13205-016-0392-z

Saxena, R.K., Prathima, C., Saxena, K.B., Hoisington, D.A., Singh, N.K., & Varshney, R.K. (2010). Novel SSR markers polymorphism detection in pigeonpea (Cajanus spp). *Plant breeding*, 129(2).

Sayo, S., David, K. O., Omolayo, J. A., Silvestro, M., Lawrence, S. F., Ayoola, O. O., & Adebiyi, O. O. (2017). Correlation and path coefficient analysis of top-cross and three-way cross hybrid maize populations. *African Journal of Agricultural Research*, 12(10), 780–789. https://doi.org/10.5897/ajar2016.11997

Schaaffhausen, R V., (1963). Economical methods for using the legume Dolichos lablab for soil improvement, food and feed. Turrialba. 13:172-178.

Segmagn, K., Zhang, J. & Arnaud, G. (2018). Development of species diagnostic SNP markers for quality control genotyping in four rice species. *Mol. Breeding*, 38 (11).

Sennhenn, A. (2015). Exploring niches for short-season grain legumes in semi-arid Eastern Kenya. (PhD Thesis), Apolda, Germany.

She C-W, Jiang X-H. (2015). Karyotype analysis of Lablab purpureus (L.) sweet using fluorochrome banding and fluorescence in situ hybridisation with rDNA probes. *Czech J. Genet. Plant Breed*. 51:110–116.

Shivachi, A., Kiplagat, K., & Kinyua, G. (2013). Microsatellite analysis of selected Lablab purpureus genotypes in Kenya. *Rwanda Journal*, 28(1), 39–52. https://doi.org/10.4314/rj.v28i1.3

Shivakumar, M. S., Ramesh, S., Rao, A. M., Udaykumar, H. R., & Keerthi, C. M. (2017). Cross Legume Species/Genera Transferability of SSR Markers and their Utility in Assessing Polymorphism among Advanced Breeding Lines in Dolichos Bean (*Lablab purpureus* L.). *International Journal of Current Microbiology and Applied Sciences*, 6(8), 656–668. https://doi.org/10.20546/ijcmas.2017.608.083

Sing, R.K. & Chadhary, S.D. (1985), Biometrical methods in quantitative genetic analysis. Kalian publishers, New Delhi, India.

Singer S.R., Hsiung L.P., Huber S.C. (1991) Determinate (det) mutant of Pisum sativum L. (luguminosae: papilonideae) exhibits an indeterminate growth pattern, *Am. J. Bot* 77:1330-1335

Singh, J., Fiyaz, R. A., Kumar, S., Ansari, M. A., & Gupta, S. (2013). Genetic variability, correlation and path coefficient analysis for yield and its attributing traits in pigeonpea (Cajanus cajan) grown under rainfed conditions of Manipur. *Indian Journal of Agricultural Sciences*, 83(8), 852–858.

Singh, N., Singh, H., & Nagarajan, P. (2013). Development of SSR markers in mung bean, *Vigna radiata* (L) Wilezek using in silico methods. *Journal of crop and weed*, 9(1), 69-74.

Singh, Rai, N., Lal, H., Bhardwaj, D. R. ., Singh, R., & Singh, A. P. (2011). Correlation, path and cluster analysis in hyacinth bean (*Lablab purpureus* L. Sweet). *Journal of Agricultural Technology,* 7(4), 1117–1124.

Singh, S., Singh, P. K., Singh, D. R., Pandey, V. B., & Srivastava, R. C. (2015). Genetic variability and character association study in dolichos bean. *Indian Journal of Horticulture*, 72(3), 343–346. https://doi.org/10.5958/0974-0112.2015.00067.5

Smith, G.R., Rouquette, F.M. & Pemberton, I. J. (2008). Registration of 'Rio Verde' lablab. *Journal of plant registrations*, 2 (1).

Song, Q., Geofeng, J., Youlin, Z., David, G. & Nelson, R. T. (2010). Abundance of SSR Motifs and Development of Candidate Polymorphic SSR Markers (BARCOYSSR_1.0) in soybean. *Crop Sci.* 50, 1950-1960. http://digitalcommons.unl.edu/agronomyfacpub/791

Souframanien, J., & Reddy, K. S. (2015). De novo assembly, characterization of immature seed transcriptome and development of genic-SSR markers in black gram [*Vigna mungo* (L.) Hepper]. *PLoS ONE*, 10(6). https://doi.org/10.1371/journal.pone.0128748

Suliemani, A., Al-Faifi, A., Hussein, M., Migdadi, S. et al (2016). Development, characterization and use of genomic SSR markers for assessment of genetic diversity in some Saudi date palm (*Phoenix dactylifera* L.) cultivars, *Electronic Journal of Biotechnology*, 21, 18–25.

Sullivan, P. (2003). Intercropping Principles and Production Practices. In P. Williams, Appropriate Technology Transfer for Rural Areas (ATTRA). Fayetteville: Arkansas.

Suresh, S., Park, J. H., Cho, G. T., Lee, H. S., Baek, H. J., Lee, S. Y., & Chung, J. W. (2013). Development and molecular characterization of 55 novel polymorphic cDNA-SSR markers in faba bean (*Vicia faba* L.) using 454 pyrosequencing. *Molecules*, 18(2), 1844–1856. https://doi.org/10.3390/molecules18021844

Tadele, T., Teshome, L., Behailu, M., & Gashaw, S. (2014). Correlation and path coefficient analysis of yield and yield components in lentil (Lens *culinaris Medik*.) germplasm in the highlands of Bale, Ethiopia. *International Journal of Biodiversity and Conservation*, 6(1), 115–120. https://doi.org/10.5897/ijbc2013.0618

Taheri, S., Abdullah, T. L., Yusop, M. R., Hanafi, M. M., Sahebi, M., Azizi, P., & Shamshiri, R. R. (2018). Mining and development of novel SSR markers using Next Generation Sequencing (NGS) data in plants. *Molecules*, 23(2). https://doi.org/10.3390/molecules23020399

Tang, J., Baldwin, S. J., Jacobs, J. M. E., van der Linden, C. G., Voorrips, R. E., Leunissen, J. A. M., … Vosman, B. (2008). Large-scale identification of polymorphic microsatellites using an in silico approach. *BMC Bioinformatics*, 9, 1–13. https://doi.org/10.1186/1471-2105-9-374

Tegemeo. (2010). Kenya Baseline Study, Report. Nairobi: Alliance for a Green Revolution in Africa.

Thobatsi, T. (2009). Growth and yield response of maize (Zea mays L.) and Cowpea (Vigna unguiculata L.) in intercropping system. (*Masters dissertation*). University of Pretoria, South Africa.

Valdisser, P.A., Pappas GJ Jr, de Menezes IP, Mu¨ ller BS, Pereira WJ, Narciso MG, et al. (2017). SNP discovery in common bean by Restriction-Associated DNA (RAD) sequencing for genetic diversity and population structure analysis. *Mol Genet Genomics* 291, 1277–91. https://doi.org/10.1007/s00438-016-1182-3

Vales, M. I., Srivastava, R. K., Sultana, R., Singh, S., Singh, I., Singh G., Patil, S. B. & Saxena, K. B. (2012).Breeding for Earliness in Pigeon pea: Development

of New Determinate and Non determinate Lines. *Crop Science*, 52, November-December.

Van Inghelandt, D., Melchinger, A., Lebreton, C. & Stich, B. (2010). Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theor. Appl. Genet*. 120, 1289–1299

Vatanparast, M., Shetty, P., Chopra, R., Doyle, J. J., Sathyanarayana, N., & Egan, A. N. (2016). Transcriptome sequencing and marker development in winged bean (*Psophocarpus tetragonolobus*; Leguminosae). *Scientific Reports*, 6(April), 1–14. https://doi.org/10.1038/srep29070

Veerendra, K., Verma, A., Kumar, J.A., Amit, P.C. (2014). Genetic divergence, path coefficient and cluster analysis of French bean (*Phaseolus vulgaris* L.) genotypes. *Indian Journal of Agricultural Sciences*, 84(8):925-930.

Venkatesha, S. C., Byre Gowda, M., Mahadevu, P., Mohan Rao, A., Kim, D. J., Ellis, T. H. N., & Knox, M. R. (2007). Genetic diversity within Lablab purpureus and the application of gene-specific markers from a range of legume species. *Plant Genetic Resources*: Characterisation and Utilisation, 5(3), 154–171. https://doi.org/10.1017/S1479262107835659

Venkatesha, S., Ramanjini, C., Gowda, P.H., Ganapathy, K.N., Byre Gowda, M., Ramachandra, R., Girish G., Channamallikarjuna, V., Shantala, L. & Gowda, T.S. (2010). Genetic Fingerprinting in Dolichos Bean using AFLP Markers and Morphological Traits *International Journal of Biotechnology and Biochemistry*, 6 (3), 395–404

Venkatesha, S.C., Gowda, M.B., Mahadevu, P., Rao, A.M., Kim, D., Ellis, T.N., & Knox, M.R. (2007). Genetic diversity within *Lablab purpureus* and the application of gene-specific markers from a range of legume species. *Plant Genetic Resources*: Characterization and Utilization, 5(3), 154–171.

Verdcourt, B. (1970). Lablab Adans. In: Studies in the Leguminosae- Papilionoideae for the 'Flora of Tropical East Africa': III. *Kew Bull*. 24: 409–411.

Verma, A. K., Uma Jyothi, K., & Dorajee Rao, A. V.D. (2015). Variability and character association studies in dolichos bean (*Lablab purpureus* L.) genotypes. *Indian Journal of Agricultural Research*, 49(1), 46–52. https://doi.org/10.5958/0976-058X.2015.00006.2.

Vieira, M. L. C., Santini, L., Diniz, A. L., & Munhoz, C. de F. (2016). Microsatellite markers: What they mean and why they are so useful. *Genetics and Molecular Biology*, 39(3), 312–328. https://doi.org/10.1590/1678-4685-GMB-2016-0027

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van De Lee, T., Hornes, M., Frijters, A., Pot J., Peleman, J. & Kuiper, M. (1995), AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* 11: 4407 – 4014.

Vu, H., Le, T., Vu, D., Nguyen, T., & Ngoc, T. (2019). Correlations and Path Coefficients for Yield Related Traits in Soybean Progenies. *Asian Journal of Crop Science*, 11(2), 32–39. https://doi.org/10.3923/ajcs.2019.32.39

Wamalwa, N., John, M. & Clabe, W. (2016). Genetic Diversity of Cowpea (Vigna unguiculata (L.) Walp.) Accession in Kenya Gene Bank Based on Simple Sequence Repeat Markers. *International Journal of Genomics* 1, 1-5, http://dx.doi.org/10.1155/2016/8956412

Wang, L., Peng, B., Xingxin, Y., Honglin, C., Suhua, W., Xin, C., & Xuzhen, C. (2018). Genetic diversity assessment of a set of introduced mung bean accessions (*Vigna radiata* L.) *The crop journal*, 6, 207-213.

Wang, M. L., Morris, J. B., Barkley, N. A., Dean, R. E., Jenkins, T. M., & Pederson, G. A. (2007). Evaluation of genetic diversity of the USDA Lablab purpureus germplasm collection using simple sequence repeat markers. *Journal of Horticultural Science and Biotechnology*, 82(4), 571–578. https://doi.org/10.1080/14620316.2007.11512275

Wang, Y., Ghouri, F., Shahid, M.Q., et al. (2017).The genetic diversity and population structure of wild soybean evaluated by chloroplast and nuclear gene sequences. *Biochem Syst Ecol*.71, 170–178.

Wang, Z., Raymer, P., & Chen, Z. (2017). Isolation and Characterization of Microsatellite Markers for Stenotaphrum Trin. Using 454 Sequencing Technology. HortScience, 52(1), 16–19. https://doi.org/10.21273/hortsci10521-16

Wei, W., Qi, X., Wang, L., Zhang, Y., Hua, W., Li, D., … Zhang, X. (2011). Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. *BMC Genomics*, 12(September). https://doi.org/10.1186/1471-2164-12-451

Weller, J. L., & Ortega, R. (2015). Genetic control of flowering time in legumes. *Frontiers in Plant Science*, 6(APR), 1–13. https://doi.org/10.3389/fpls.2015.00207

Weller, O. (2015). Genetic control of flowering time in legumes. Front. *Plant Sci*., 6(206), 1-13. doi:10.3389/fpls.2015.00207

Willing, E., Dreyer, C., & Oosterhout, C. Van. (2012). Estimates of Genetic Differentiation Measured by F ST Do Not Necessarily Require Large Sample Sizes When Using Many SNP Markers. *PLoS ONE*, 7(8), 1–7. https://doi.org/10.1371/journal.pone.0042649

Wong, Q. N., Tanzi, A. S., Ho, W. K., Malla, S., Blythe, M., Karunaratne, A., … Mayes, S. (2017). Development of gene-based SSR markers in winged bean

(*Psophocarpus tetragonolobus* (L.) DC.) for diversity assessment. *Genes*, 8(3), 1–12. https://doi.org/10.3390/genes8030100

Wu, J., Wang, L., Li, L., & Wang, S. (2014). De novo assembly of the common bean transcriptome using short reads for the discovery of drought-responsive genes. *PLoS ONE*, 9(10). https://doi.org/10.1371/journal.pone.0109262

Wu, J., Wang, L., Li, L., & Wang, S. (2014). De novo assembly of the common bean transcriptome using short reads for the discovery of drought-responsive genes. *PLoS ONE*, 9(10). https://doi.org/10.1371/journal.pone.0109262

Yamada, T., Hajika, M., Yamada, N., Hirata, K., Okabe, A., Oki, N., Ishimoto, M. (2012). Effects on flowering and seed yield of dominant alleles at maturity loci E2 and E3 in a Japanese cultivar, Enrei. *Breeding Science*, 61(5), 653–660.

Yin, D., Wang, Y., Zhang, X., Li, H., Lu, X., Zhang, J., … Chen, S. (2013). De Novo Assembly of the Peanut (Arachis hypogaea L.) Seed Transcriptome Revealed Candidate Unigenes for Oil Accumulation Pathways. *PLoS ONE*, 8(9), 1–11. https://doi.org/10.1371/journal.pone.0073767

Yong-jin, P., Ju Kyong, L. & Nam-Soo, K. (2009). Simple sequence repeat polymorphisms for evaluation of molecular diversity and germplasm classification of minor crops. *Molecules* 14: 4546-4569.

Yunbi, X. & Jonathan, H. C. (2008) Marker-Assisted Selection in Plant Breeding: From Publications to Practice. *Crop Science*, 48, 391–407

Zane, L., Bargelloni, L., & Patarnello, T. (2002). Strategies for microsatellite isolation: a review. *Molecular Ecology*, 11(1), 1–16. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11903900

Zhang, G., Xu, S., Mao, W., Gong, Y., & Hu, Q. (2013). Development of EST-SSR markers to study genetic diversity in hyacinth bean (Lablab *purpureus* L.). *Plant Omic Journal*, 6(4), 295–301.

Zhao, J., Tong, Y., Ge, T., & Ge, J. (2016). Genetic diversity estimation and core collection construction of *Sinojackia huangmeiensis* based on novel microsatellite markers. *Biochemical Systematics and Ecology*, 64(388), 74–80.

APPENDICES

Appendix I. The names, sequence, motif and expected product sizes of the developed lablab derived SSR primers

Primer name	Forward primer	Reverse primer	SSR type	SSR motif	Expected Product size in bp
LabRRT 1	TGGATTCTACAGTTTC GATGACGA	GTCAGACGGTGGTTTC TGCCTTAT	p3	(GAC)6	91
LabRRT 2	GCCATGTTCTGAAAG ATGTAACAGTG	GGCAAGCAGTCATATC CAGAAACT	p4	(AAAG)5	183
LabRRT 3	GGCATGGTGAAGATT GAAGAAGAG	AGAAGCAGAGGACAG GTGAATTGT	p4	(AAAT)5	275
LabRRT 4	TACTGTTCATCTCCAC AACCCTCA	CTCATCGTAGGCAATC CTTCAGTT	p4	(AACA)5	272
LabRRT 5	GACAGCAGCTCTAGC TCTTCGAGT	CCTCAGTTTCGACAAC CTCAACTT	p3	(AAG)6	136
labRRT 6	TTCAGCCTGCTCCTG AAGTT	CAGCTGCCTTAGCCTT TGAG	p3	(AAG)7	243
LabRRT 8	TCAGAACTCTACTTTC TGAGCTTGA	ATCATACAGTCCGTGT TGTTCG	p3	(AAG)9	168
LabRRT 9	TAGCTCTGTGTAAGA TGTGCTGCT	AGATTTGGTGAGAACT GTGCCTA	p3	(AAT)6	184
LabRRT 11	CGGCGATATTACGAC TCTAACTG	AACACCTCTTGCCAGT AATTCTCT	p3	(AAT)7	253
LabRRT 12	GGGATATCTCCTTGG TTATACTGG	GTAAAGAAGCCTCAAT TCTTCTGC	p3	(AAT)8	247
LabRRT 13	GTCCTCCTTTACAGA AAGGGTAGC	GGAACAGCAATCAGA GACAGAATA	p2	(AC)7	198
LabRRT 14	CTAAATCCCAAGCAC ATAAACTCC	CTACCCATTATTTGCT TTGAGACC	p2	(AC)9	152
LabRRT 18	TGGGGATCATTCAAC TAACTTCTC	CTCCACAACTCCAACA CTAACAAC	p3	(ACT)8	188
labRRT 19	AATTGACAAGTGCTC TACCGAATC	GAGGAAGCAACAGAT ATTGGAGTT	p4	(ACTT)5	320
LabRRT 20	AACAGCCAAACGAAA CTAAGTCAG	CTACCAGAATAATGGC TAGTGCAG	p2	(AG)10	336
LabRRT 21	CGCCAATTAGCAAAG TGAGAAT	AAACGTCAAAGAGTC GGATTGT	p2	(AG)10	174
LabRRT 22	CAAGATTTGCGGCAG AGTAAACAG	ACGAGAGTCTGAGAC CTGCCTTC	p2	(AG)11	302
LabRRT 23	GGGAGTGTGAAATAG AGAATCAGTT	CAGCACTATCCACACC TGCAATAC	p2	(AG)12	136
LabRRT 24	GGAAACAACGAGCAT CAAGAGAAC	AGGCAAACTAACAAG GTGAACAGG	p2	(AG)7	214
LabRRT 25	CCACAAATAACCCTC CATTCATTG	TCTGCCTCACAGTCTA TGGGATCT	p2	(AG)7	301
LabRRT 26	AGCAAGCACGACGAG TAGGACTAA	CAGATGAACTTGATTG CTTCCAGA	p2	(AG)7	321
LabRRT 27	GATAGGATCCTGAAA TGCGGACT	ACCAACTCAGCCATCG TAACAACT	p2	(AG)7	105
LabRRT 28	AATCGAACAAAGTGA AGTGCCTTG	AAATAGCCTCCAACTT CTCCCACT	p2	(AG)8	96
LabRRT 29	ACAGAGCAGAAACA GCTGCACTT	GATTCAAAGGAGGGA AAGTGTGAA	p2	(AG)8	187

LabRRT 30	GCAGTTGTGAGACAC AGAAATTAAGC	CGAAGGAGGCAACAT TACCTAGAA	p2	(AG)8	182
LabRRT 31	CAGAGCTGTTGGAAA TACGAGGTT	GCAGAAGAACAAGCA GAAATGGAT	p2	(AG)9	271
LabRRT 32	CCTTCTCCACCCAAG AAACAAG	ACGAACACGCGTCTGT AATCCT	p3	(AGA)6	311
LabRRT 33	ATTATGGCTATCCAT CCAGAAGCA	AGAGGCTTGTCAACGT GGTTACTC	p3	(AGA)6	325
LabRRT 34	GACCACTACACCAAA CGCAACTTA	TTCATCTCTGAGGATT GGTTTTCC	p3	(AGA)7	108
LabRRT 35	GAGGGTCGAGATCTC ATGATCG	ATTGGCTAGCAACCCA TCTTCA	p4	(AGAT)6	101
LabRRt 36	CcAcACGTCATCGCTA TCAG	tTgGGtTGATTCtgAGTg GA	P2	(AG)9	120
LabRRT 37	CGAAGAATGCAATCG TTATCGT	TCCGAACCCAGCAACA GTAAAT	p3	(AGT)6	89
LabRRT 38	CGAAGATTTGCGTCG TAATTTG	GACAAGTGCTCTACCG AATCAGC	p4	(AGTA)5	125
LabRRT 39	CCTTTGGAATTCTCCA TTCCATC	TAAGGCTGTGAGAAC AGGCACAG	p3	(ATC)7	150
LabRRT 40	GCTCAGCCTCGGAGT TACTTGTAG	AGAGGACACCATCACT CCATCAC	p3	(ATG)10	148
LabRRT 41	CGAGAATGTCGGCTT GAGTTTC	GCACTATCACTGCTTT GTTCGTCT	p3	(ATG)6	180
LabRRT 42	GTTGCCGTAGAATAA AGACGTTGC	GCGATTCAAACCCAGA TTTACAAC	p3	(ATG)7	245
LabRRT 43	GTTCAGCCACCAAAT TACTCGTTT	CCATTCAGTATTCTTT CCATCCACA	p5	(ATGAC) 6	117
LabRRT 44	AAGCTTCGTTGTTTCT GCGATTAG	CGAGCTTTAAACCAAT CAGGACAC	p3	(ATT)6	92
LabRRT 45	TTGGTGGTAGAAGTT AAAGTGGAAG	TTAGTGTCATCAGAAG TACCTGCTA	p3	(ATT)7	145
LabRRT 46	TGCGTCTTAGCCAGT GTAATTTTC	GAACAATTGCTCATCA CCTAGCAA	c	(CA)7tc(T A)7	80
LabRRT 47	TTTGTACTATTGGGC ACACTCCAA	TCAAGCATTCCTCTAT TACAGCAGA	p3	(CAC)6	302
LabRRT 48	GTCTCCTTCAACTGTG TCCACTGA	TCAATCGTTGTTGGAA GAGGGTAT	p3	(CAC)6	187
LabRRT 49	CATGCTCTCAAGCTG TTCATCAAT	GAGTCCAACGTTGTTA GCGAGAGT	p6	(CACCA A)5	234
LabRRT 50	TCACAGAGCCAAAGA CAAACTCA	GATGAGGAGCCTCGTT GAATTG	p4	(CACG)5	138
LabRRT 51	GTGGTGACAAACcAG AGATTGATG	AATTCCCAGATCACgG TCAATAAC	p3	(AAT)6	123
LabRRT 52	CAGGTTTGTGATTCG CATGAGTC	TGGTGACTATTCATGA TGGGAATG	p3	(CAG)6	93
LabRRT 53	GGAAGAACTAAGGTC ATCATGC	GATCGCAATGATCTTC CAAAGG	p3	(CAT)7	176
LabRRT 54	GAGTAGTCCAAAGAA AGGGTGAA	TTTCAGGAATTGGGAA CAATGG	p4	(CATG)5	293
LabRRT 55	CACGGTCTCTACCAC CTCCATAAC	CAGATAAGGGTGGAG ATCCTGAGA	p3	(CCT)6	279
LabRRT 57	CGAAGCAAAATAATG GGTGTGAGT	GGCCTTTCAACAGTAC CAAGTTGT	p2	(CT)10	335
LabRRT 58	TGTTTGTTGTGTCGGA ACTGTTG	GGTGTTGAGACTGATG GAGATGG	p2	(CT)11	275
LabRRT 59	AAGATGATGAAGCCA GTGCAGAG	CATCAATCTCAGTCAT CATCCTCA	p2	(CT)12	129

LabRRT 60	CTAACCTTGACCGCCTTGaGAGT	CCAATCaAGATcCAAATCATGC	p2	(CT)7	106
LabRRT 61	TGCGAGGCTCTTTTGACTCTACTT	AGTTATTCCACTGCATTGCATCCT	p2	(CT)7	303
LabRRT 62	TCCGACTCCTACGATAGTGCCTAC	GATCAGGAGAGGGACAGAATACGA	p2	(CT)7	148
LabRRT 63	GACATGTACAGTCTTTGAGAAAGC	TGGCCTTGAACTCAAGATACAACC	c	(CT)7(AT)7	333
LabRRT 64	AGGATGGATTTAGAAATGGGAAGC	GCTGAACACAGCTAACAATCATGC	p2	(CT)8	251
LabRRT 65	ACTGGTTTTCCCTTGGTTCATGT	GGCCCGTATGTTATAGCCAAAAG	p2	(CT)8	202
LabRRT 66	ACGCCTAGAGGCTTATGATGAACA	CATGTTCTCCATTTAGGGTTCCAA	p2	(CT)9	146
LabRRT 67	TCGACTCAGCTACTCAACAACGAC	ACGAACATGGGTATGTGCTTTGTA	p2	(CT)9	164
LabRRT 68	TTCAGAACCTCGTCATCCATGTT	GGAAGAATTGGGATGGATTTGAG	p5	(CTATG)6	326
LabRRT 69	CCGGAACCACTTCCTAGTCGTAA	AATGGGATCTCGTCTCTGGATCT	p3	(CTC)6	209
LabRRT 70	TGGAAGTACAAGCATTAACCAACG	ACTGAGAACCAAAACCATGAGGAA	p3	(CTG)6	308
LabRRT 71	CTTTAACAAAACACGCGCCAAC	AACGACAGTAGGAACCCTCCTTG	p3	(CTT)15	162
LabRRT 72	GGTGGCATAATAGTCCTCGTCATC	TTACTGTGAGTTGGATTGGGAATG	p3	(CTT)6	249
LabRRT 73	TCAGTCTTCTTGTGCCCTGGTAG	AGAGAAGCCTCAGGAAGAGGTGA	p3	(CTT)6	312
LabRRT 74	CATTGTTTGTTGTGTCGGAACTGT	AGGAAAATACCAATGCAGAGTCCA	p3	(CTT)6	117
LabRRT 75	ACTGAAGTGCAACCCCTATTTCC	GTTCTGCTTCACCTTCCGAAACT	p3	(CTT)7	307
LabRRT 76	GGGAGGTAATTTTCAACGATTCC	GGAAAATTCTAGGCCGAAGATCA	p3	(CTT)8	217
LabRRT 77	AGTTTAGCACACCGATCAAATGGT	CACAAACCTCCATTACTCTCAGTCA	p3	(CTT)9	156
LabRRT 78	TGCTCGTCATTTGAGCTAAGTTTG	AATGAGTGAATGCAGCAGCAGTAG	p6	(CTTTTC)5	228
LabRRT 79	TAAGCCACTTTCACCGGAGTGTAT	TCTAATTGTAATGCTGCCCTGAGA	p2	(GA)10	202
LabRRT 80	GGCGGAGAATCAAAATTAGGAGTA	AGTGCCCAATTCTCTTCCAAGATT	p2	(GA)7	167
LabRRT 81	ACCTTTTCTGGGCATAAATCAAGC	GGAAGGAAGACGAGTGATTGAAAC	P2	(GA)8	336
LabRRT 82	AGGGAGCTGAAACTGATGTTTGAC	TTTACCTTGATCCCAACTCATTGC	p2	(GA)8	139
LabRRT 83	ATAAGAAGATCGCTTGTCGCCTTC	TCTGAGTTTTGGGTCGTTTAATCC	p3	(GAA)7	150
LabRRT 84	CAGTCGAGGGAACGGTTAATCTG	GTACATGGGGATCTTCACCACAAT	p5	(GAAGA)5	149
LabRRT 85	ACGAAGAGGATTTTGATGACGAAG	AGCCAAACAACTTCACAGGGTAAT	p3	(GAG)7	287
LabRRT 86	ATGATGAAGATGATGAAGCCAGTG	TGGAGCCAACAAAAGAAGAGATTC	p3	(GAG)7	223
LabRRT 87	TTATCCAGTGCTGAAAAGCTCCAT	AGATAATAGCCACCCTCACCAACA	p3	(GAG)8	284
LabRRT 88	TGAACAGAAAGACGATACGGATGA	AGATAGCTCCACCTGCTTAACACC	p3	(GAG)8	333

LabRRT 89	GTTATCTGGGTTGGTTGCAGTTG	AGGCTGGTAGGGAGATCTTGTTG	p5	(GAGAA)5	225
LabRRT 90	ATAACTCTGGCTCGCTCTGTGG	GTGCATTTGATTTGGTGGGAAA	p4	(GAGT)5	232
LabRRT 91	GGTGGCATAATAGTCCTCGTCATC	TTACTGTGAGTTGGATTGGGAATG	p3	(TTC)6	255
LabRRT 92	CTCATTGGCGAAATTCTCAGGTA	TCTTCACGTAACCACGGTAGGATT	p6	(TATATC)7	249
LabRRT 93	CCGAGTTAACCAGCTGGACTGTA	CACTCCAACAAGTTCAATCAACCA	p3	(GAT)6	109
LabRRT 94	GGTGACACAGCAAAGACAACCTAA	CGTTATAATCAAGTACAGGGTCTCACA	P3	(GAT)7	299
LabRRT 95	GTTGGTGGCCCAATATGAATATGTA	CTTGGCTTAAGGACACTCCATCAC	p3	(GCA)6	204
LabRRT 96	TGAAGAGCTCAGAGAAAGGGCTAA	GCGTTCAAAACCGTGTAAGTTCTC	p5	(GCATA)6	251
LabRRT 97	CTCAGCTCAAAATGAGGCTACTGATAA	TTGCAGAAGGGTACACATTAAGACTTC	p3	(GCT)6	112
LabRRT 98	AAAGGGAAAAGAGAGTGGTGGTG	AATCCAACCAAGTTGTCCAGTGA	p3	(GGT)6	167
LabRRT 99	TAGAAGCAGGGGTTTTGGGTTT	CAAGCAAGATTACCCACATGGA	p3	(GGT)7	230
LabRRT 100	GGATATCTGGCTGAACCTAACTGT	TAGCTTTAGAGGCCATACTCATCA	p4	(GGTT)7	228
LabRRT 101	CAGCTCAAGTCTGACTAATCCAAA	ATCAATGGAACGGAGTACTTCAAC	p2	(GT)10	251
LabRRT 102	TGTACTAATGAGAAGGCATGCAAC	CCCTACTCATGTGACCTAATTGCT	p3	(GTG)6	244
LabRRT 103	GATGAAAGAACCTCAAAGCTTGTC	GTATAAACCCTTCTACCATGAACAG	p2	(TA)10	206
LabRRT 104	CTCTCACGCTAGAAGCAGATGTAG	CCCTTGTCATACATGTGGAATACT	p2	(TA)10	352
LabRRT 105	ACAACACTCGGAAACCGTTACATT	CACCACCTCCAACTTCTACGGTTA	p2	(TA)11	168
LabRRT 106	CACCACCTCCAACTTCTACGGTTA	TGACCTCCATTATGGGATTCAGAT	p2	(TA)12	197
LabRRT 107	TACCACCAATTCATCTTGAGAGCA	TAGGGTCAAAGGACAAACATTTCC	p3	(TAA)6	193
LabRRT 108	CTTTCCAATGAAGATCACCTTTCC	CCGGCGATATTACGACTCTAACTG	p3	(TAT)6	300
LabRRT 109	GTGCAAGAAGATCAAATCCCAAAC	CAGACCAAACTTCCCAGCTTCTTA	p3	(TCA)6	258
LabRRT 110	GTTGAGTCGCAGCTGAACTTGAT	GAAGGAGTTGCAAAAGCAAGAGAG	p3	(TCT)6	277
LabRRT 111	GCAGCAGCAGCAACAAGTAAAAC	AACTCAAGGAAGGGTCGTTTCAG	p3	(TGG)6	288
LabRRT 112	CATAAGGGAAAGGGACTTTGACAG	TCAAAACTGTCATCACTCCAAACC	p3	(TGG)8	153
LabRRT 113	TCGTGGTGACGAGTCAAATTTCT	AGAACGTATGGATCAAAGGCAGA	p4	(TGGT)5	239
LabRRT 114	TGGTAGGTGAGGAGGCCATAGTTA	TCCACGTCTCAACTAGTGGTCATT	p4	(TGTA)5	157
Lab T1	ACCAGAATGGTTTCTCAAGTTCCT	GGTGAACCTTCCTACACCATGACT	p2	(TA)7	273
Lab T2	GTGCGCGTCACTTATTAGTTCTTA	CAATATCTTCACGTAACCACGGTA	p6	(TATATC)7	224

Lab T3	CAGATCGATTGGTagCTGGATTtC	CCTCCTTACAGAAAGGGTAGCCTAGT	p2	(TG)7	194
Lab T4	ATGTTTGGATGTACGTGATGGTTG	TCTTGAAATTACATCTCCTGCTTGC	p3	(TTC)7	206
Lab T5	CGGTTGTTTCTAATAAGCCTGGTC	GACAACCAAACCACACATGTACTGA	p3	(TAT)8	194
Lab T6	TCAATCGTTGTTGGAAGaGGGTAT	GTCTCCTTCAACTGTGTCCACTGA	p3	(TGG)6	187
Lab T7	CAGCAGTGTTGCCTCATACAGAAC	TGTACTTAGCCAAGATCAGGCACA	p3	(ATG)6	188
Lab T8	ATGATTCGGGAACTCTGTTGAGTC	CATCTTTCTATTCCTGGTCCaACC	p3	(TCT)6	216
Lab T9	AACAAGCAGAAATGGATTTCCAAG	ATGGGCAACaACTGGTACACCATA	p2	(TC)9	230
Lab T10	CAAGCTCTGCAGATGATGATGAAT	AGATAATAGCCACCCTCACCAACA	p3	(GAG)8	191
Lab T11	CCGAGTTAACCAGCTGGACTGTA	ACTCCAACAAGTTCAATCAAACCA	p3	(GAT)6	236
Lab T12	CACCACCTCCAACTTCTACGGTTA	TGACCTCCATTATGGGATTCAGAT	p2	(TA)11	193
Lab T13	ACGTGATTGTCCCAATGATCCTAT	TGCAATCCCTACTCATGTGACCTA	p3	(GTG)6	200
Lab T14	GGCATGGTGAAGATTGAAGAAGAG	AGAAGCAGAGGACAGGTGAATTGT	p2	(GA)8	255
Lab T15	GCTTCAGTGTTGATAATGCCAGTG	TCGTCATCCATGTTAATGGTCAAC	p5	(GCATA)6	235
LabT 16	ATTCCGAGGACTGGAACAaGAaG	CATGCTCTCAAGCTGTTCATCAAT	p6	(GTGTTG)5	263
Lab T17	TGAAAGATGGAGTGATGACACATTG	GTGTTGTTCCtTTCCCTTGAGATG	p4	(CATG)5	222
Lab T18	CAACTGGGATTCTTGCACTTTAGG	AATTCCCAGATCACgGTCAATAAC	p3	(AAT)6	233
Lab T19	CATTGTTTGTTGTGTCGGAACTGT	GCACCCAAACAGACTAATGATGGT	p3	(CTT)6	216
Lab T20	CTGAGTCACCATCATCGTCTTCAT	TCGTCAAAGTACTTAACgGCCATC	p3	(GAA)6	205
Lab T21	ACCAGGAGgTTTGCCTAGAAGC	CACTGCTTTGTTCGTCTTCAGAAA	p2	(AG)7	289
Lab T22	CTtGCtGAAGAGTAAACCGAGGAT	CCAATCaAGATcCAAATCATGC	p2	(AG)7	226
Lab T23	CAAATCActATGACGGCTCaaacc	CCAGATGCATCACCTTCTagTCCT	p3	(GAT)6	173
Lab T24	GATCAGCTCCAGACTGCTGACG	TAACCCTCCATTCATTGTCCATTC	p2	(TC)7	202
Lab T25	GGGTTGAagCTCACACAAATTCTT	CCAATGATGGTTGTATGAGTAGCAC	p4	(TGGT)5	190
Lab T26	CTTAGCTATTTGCCCAACCAAACA	TGCATGTTCAcACAAACTTCTCCT	p2	(AT)8	211
Lab T27	TACTGTCATCCtCcTTTCCATTGC	GTaACTGATGAGGCTGTGGAGGTT	p3	(CAG)6	250
Lab T28	CTTTCTCCATGCAGACCAAACTTC	CCTGTAAATAACTGTCCTGGGAAGC	p3	(ATG)6	204
Lab T29	TGGTGCTACTGCTACCACTGTCTC	GAAGAATGCAATCGTTATCGTTGAG	p3	(ACT)6	162
Lab T30	AGGTCTGCTTCTACCCATCCATGT	ACGTTCCTCCGCAATTAGCATA	p2	(AT)11	211

Lab T31	CTTGGTCTCCAGGGATTCCATTAT	CAGCATGATCATCCAAATCTTCAA	p3	(AAG)6	201
Lab T32	CTAGCTTCCACGACATTTCCTtgT	CTCGGAGTGAGAGTTCATCTCGTT	p3	(ACC)7	169
Lab T33	CTAACCATGGCCTTGAGTGGTACT	AATGAGTGAATGCAGCAGCAGTAG	p6	(CTTTTC)5	228
Lab T34	ATGCTTTGGCACTTtCTTATCAGC	ATGATAAAGTTCAGCCTGCTCCTG	p3	(CTT)7	271

Appendix II.The expect value (E), gene ontology identity (GO ID) and names obtained by BLASTX search of isotigs sequences from where the 34 lablab primers were designed.

Primer name	sequence name	E value	GO IDs	GO names
Lab T1	isotig02663	-	-	-
Lab T2	isotig02918	-	-	-
Lab T3	isotig02852	1.00E-19	F:GO:0004105; P:GO:0006657; F:GO:0031210	F:choline-phosphate cytidylyltransferase activity; P:CDP-choline pathway; F:phosphatidylcholine binding
Lab T4	isotig02277	8.31E-49	F:GO:0004124; C:GO:0005737 ; P:GO:0006535; F:GO:0030170	F:cysteine synthase activity; C:cytoplasm; P:cysteine biosynthetic process from serine; F:pyridoxal phosphate binding
Lab T5	isotig02486	-	-	-
Lab T6	isotig02335	-	-	-
Lab T7	isotig02317	1.06E-21	C:GO:0005634 ; P:GO:0006355; F:GO:0008270; P:GO:0009640	C:nucleus; P:regulation of transcription, DNA-templated; F:zinc ion binding; P:photomorphogenesis
Lab T8	isotig02270	-	-	-
Lab T9	isotig02269	6.05E-18	P:GO:0006508; F:GO:0008234; P:GO:0016311; F:GO:0016791	P:proteolysis; F:cysteine-type peptidase activity; P:dephosphorylation; F:phosphatase activity
Lab T10	isotig02122	2.39E-11	C:GO:0016021	C:integral component of membrane
Lab T11	isotig01972	9.02E-16	F:GO:0003700; F:GO:0003712; C:GO:0005634 ; P:GO:0006355; P:GO:0009723; P:GO:0009737; P:GO:0009863; P:GO:0009867; P:GO:0009938; P:GO:0010187; P:GO:0042538; F:GO:0043565; P:GO:2000033; P:GO:2000377	F:DNA-binding transcription factor activity; F:transcription coregulator activity; C:nucleus; P:regulation of transcription, DNA-templated; P:response to ethylene; P:response to abscisic acid; P:salicylic acid mediated signaling pathway; P:jasmonic acid mediated signaling pathway; P:negative regulation of gibberellic acid mediated signaling pathway; P:negative regulation of seed germination; P:hyperosmotic salinity response; F:sequence-specific DNA binding; P:regulation of seed dormancy process; P:regulation of reactive oxygen species metabolic process
Lab T12	isotig01905	-	-	-

Lab T13	isotig01632	7.17E-69	F:GO:0003676; F:GO:0008270; F:GO:0016491; P:GO:0055114	F:nucleic acid binding; F:zinc ion binding; F:oxidoreductase activity; P:oxidation-reduction process
Lab T14	isotig01239	1.05E-26	F:GO:0003700; C:GO:0005634; P:GO:0006355; F:GO:0043565	F:DNA-binding transcription factor activity; C:nucleus; P:regulation of transcription, DNA-templated; F:sequence-specific DNA binding
Lab T15	isotig01222	2.65E-41	F:GO:0004161; F:GO:0004311; F:GO:0004337; P:GO:0008299	F:dimethylallyltranstransferase activity; F:farnesyltranstransferase activity; F:geranyltranstransferase activity; P:isoprenoid biosynthetic process
LabT 16	isotig01196	3.13E-52	F:GO:0003743; P:GO:0006413; C:GO:0016021	F:translation initiation factor activity; P:translational initiation; C:integral component of membrane
Lab T17	isotig01193	-	-	-
Lab T18	isotig01303	-	-	-
Lab T19	isotig01128	4.04E-35	F:GO:0016874	F:ligase activity
Lab T20	isotig01116	2.27E-36	P:GO:0010112	P:regulation of systemic acquired resistance
Lab T21	isotig01099	8.93E-76	F:GO:0003735; P:GO:0006412; C:GO:0022625	F:structural constituent of ribosome; P:translation; C:cytosolic large ribosomal subunit
Lab T22	isotig00994	2.77E-131	P:GO:0005985; P:GO:0009877; F:GO:0016157	P:sucrose metabolic process; P:nodulation; F:sucrose synthase activity
Lab T23	isotig00970	2.26E-110	C:GO:0005854	C:nascent polypeptide-associated complex
Lab T24	isotig00904	3.66E-114	C:GO:0009535; C:GO:0009538; P:GO:0015979; C:GO:0016021; F:GO:0016491; P:GO:0055114	C:chloroplast thylakoid membrane; C:photosystem I reaction center; P:photosynthesis; C:integral component of membrane; F:oxidoreductase activity; P:oxidation-reduction process
Lab T25	isotig00789	-	-	-
Lab T26	isotig00746	-	-	-
Lab T27	isotig00724	2.22E-139	P:GO:0000495; F:GO:0003723; C:GO:0005829; C:GO:0009506; F:GO:0009982; P:GO:0031118; P:GO:0031120;	P:box H/ACA snoRNA 3'-end processing; F:RNA binding; C:cytosol; C:plasmodesma; F:pseudouridine synthase activity; P:rRNA pseudouridine synthesis; P:snRNA pseudouridine synthesis; C:box H/ACA snoRNP complex; P:mRNA pseudouridine synthesis

			C:GO:0031429; P:GO:1990481	
Lab T28	isotig00717	2.11E-102	C:GO:0016021	C:integral component of membrane
Lab T29	isotig00665	-	-	-
Lab T30	isotig00605	0	C:GO:0005737; P:GO:0006006; F:GO:0016779; F:GO:0080048	C:cytoplasm; P:glucose metabolic process; F:nucleotidyltransferase activity; F:GDP-D-glucose phosphorylase activity
Lab T31	isotig00585	0	C:GO:0005829; P:GO:0009228; C:GO:0009570; F:GO:0046872; P:GO:0052837	C:cytosol; P:thiamine biosynthetic process; C:chloroplast stroma; F:metal ion binding; P:thiazole biosynthetic process
Lab T32	isotig00520	1.68E-121	F:GO:0003723; C:GO:0009535; P:GO:1901259; C:GO:1990904	F:RNA binding; C:chloroplast thylakoid membrane; P:chloroplast rRNA processing; C:ribonucleoprotein complex
Lab T33	isotig00498	0	C:GO:0005777; F:GO:0008891; F:GO:0010181; P:GO:0055114	C:peroxisome; F:glycolate oxidase activity; F:FMN binding; P:oxidation-reduction process
Lab T34	isotig00493	5.61E-17	F:GO:0003743; F:GO:0003924; F:GO:0005525; P:GO:0006413	F:translation initiation factor activity; F:GTPase activity; F:GTP binding; P:translational initiation

Appendix III.The results of annealing temperature optimization for 75 SSR primer pairs amplified on lablab DNA

Unit	Primer name	TM (^{0}C)F	TM (^{0}C) R	Ann. Temp. optimization (^{0}C)	Presence of secondary products	Strength of band on agarose gel	Optim. Ann. temp. (^{o}C)
1	LabRRT 1	59.3	62.7	50, 53, 55, 57, 59, 62	N, N, Y,Y,Y,Y	Weak band	53
2	LabRRT 2	61.6	61.0	50, 53, 55, 57, 59, 62	na	No band	-
3	LabRRT 3	61.0	64.8	50, 53, 55, 57, 59, 62	na	No band	-
4	LabRRT 4	59.3	59.7	50, 53, 55, 57, 59, 62	N, N, N, na, na, na	Faint band	55
5	LabRRT 5	61.0	61.3	50, 53, 55, 57, 59, 62	Na	No band	-
6	LabRRT 6	59.3	62.7	50, 53, 55, 57, 59, 62	na, N, N, N, na, na	Faint band	57
7	LabRRT 7	62.7	61.0	50, 53, 55, 57, 59, 62	N, N, N, N, na, na	Faint band	57
8	LabRRT 8	61.0	61.0	50, 53, 55, 57, 59, 62	N, N, N, N, N, na	Faint band	59
9	LabRRT 12	61.0	59.3	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	Na, na, Y, Y, Y, Y, Y,Y	Strong bands	55.7
10	LabRRT 19	59.3	59.3	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	Y,Y,Y,Y,Y, Y,Y,Y	Faint bands	55-59
11	LabRRT 23	59.7	62.7	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	Y,Y,Y,Y,Y, Y,Y,Y	Faint bands	55-59
12	LabRRT 25	59.3	62.7	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	N,N,N,N,N, N,N,N	Strong bands	55-59
13	LabRRT 28	59.3	61.0	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	N,N,N,N,N, N,N,N	Strong band	55-59
14	LabRRT 29	60.6	59.3	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	Y,Y,Y,Y,Y, Y,Y,Y	Strong bands	55-59
15	LabRRT 30	61.6	61.0	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	Y,Y,Y,Y,Y, Y,Y,Y	Strong bands	55-59
16	LabRRT 34	61.0	59.3	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	na	No band	-
17	LabRRT 35	62.1	58.4	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	N,N,N,N,N, N,N,N	Strong bands	55-59
18	LabRRT 36	59.4	55.3	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	N,N,N,N,N, N,N,N	Strong bands	55-59
19	LabRRT 37	56.5	58.4	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	Y,Y,Y,Y,Y, N,Y,Y	Strong bands	55-59

20	LabRRT 38	56.5	62.4	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	Y,Y,Y,Y,Y, Y,Y,Y	Strong bands	55-59
21	LabRRT 40	64.4	62.4	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	N,N,N, Y,Y,Y,Y,Y	Strong bands	52.9
22	LabRRT 43	59.3	59.7	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	Y,Y,Y,Y,Y, Y,Y,Y	Faint bands	55-59
23	LabRRT 44	59.3	61.0	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	Y,Y,Y,Y,Y, Y,Y,Y	Faint bands	55-59
24	LabRRT 49	59.3	62.7	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	Y,N,Y,Y,Y, Y,Y,Y	Faint bands	51
25	LabRRT 50	58.9	60.3	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	N,N,N,N,N, N,N,N	Strong bands	55-59
26	LabRRT 52	60.6	59.3	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	N,N,N,N,N, Y,Y,Y	Strong bands	55-59
27	LabRRT 53	58.4	58.4	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	N,N,N,Y,Y, Y,Y,Y	Weak bands	52.9
28	LabRRT 61	61.0	59.3	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	N,N,N,N,N, Y,Y,Y	Strong bands	55-59
29	LabRRT 63	59.3	61.0	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	N,N,N,N,Y, Y,Y,Y	Weak bands	55-59
30	LabRRT 64	59.3	61.0	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	na	No band	-
31	LabRRT 76	58.9	58.9	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	Y,Y,Y,Y,Y, Y,Y,Y	Strong band	55-59
32	LabRRT 77	59.3	61.3	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	Y,Y,Y,Y,Y, Y,Y,Y	Strong band	55-59
33	LabRRT 83	61.0	59.3	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	N,N,N,N,N, N,N,N	Strong band	55-59
34	LabRRT 84	62.4	61.0	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	Y,Y,Y,Y,Y, Y,Y,Y	Weak bands	55-59
35	LabRRT 90	62.1	56.5	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	Y,Y,Y,Y,N, Y,Y,Y	Strong band	59
36	LabRRT 92	58.9	61.0	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	Y,Y,Y,Y,N, Y,Y,Y	Weak band	59
37	LabRRT 94	61.0	61.9	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	Y,Y,Y,Y,Y, N, na, na	Weak band	62
38	LabRRT 98	60.6	58.9	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	Y,Y,Y,Y,Y, Y,Y,Y	Weak band	55-59

39	LabRRT 103	59.3	59.7	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	N,N,N,N,na, na, na, Y	Strong bands	55.7
40	LabRRT 112	61.0	59.3	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	Y,Y,Y,Y,Y, Y,Y,Y	Strong bands	55-59
41	LabRRT 114	62.7	61.0	50, 51, 52.9, 55.7 59.1, 62, 63.8, 65	N.N.N.Y,Y, Y,Y,Y	Strong bands	52.9
42	Lab T1	59.3	62.7	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
43	Lab T2	59.3	59.3	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
44	Lab T3	61.0	64.8	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
45	Lab T4	59.3	59.7	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
46	Lab T5	61.0	61.3	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
47	Lab T6	59.3	62.7	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
48	Lab T7	62.7	61.0	50, 53, 55, 57, 59, 62	Y,N,Y,Y,Y, Y	Strong bands	53
49	Lab T8	61.0	61.0	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
50	Lab T9	57.6	61.0	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
51	Lab T10	59.3	61.0	50, 53, 55, 57, 59, 62	Y,Y,Y,Y,Y, Y	Strong bands, longer than expected	55-59
52	Lab T11	62.4	57.6	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
53	Lab T12	62.7	59.3	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
54	Lab T13	59.3	61.0	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
55	Lab T14	61.0	61.0	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
56	Lab T15	61.0	59.3	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
57	LabT 16	60.6	59.3	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
58	Lab T17	59.7	61.0	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
59	Lab T18	61.0	59.3	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
60	Lab T19	59.3	61.0	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	No band	-
61	Lab T20	61.0	61.0	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
62	Lab T21	62.1	59.3	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
63	Lab T22	61.0	56.5	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
64	Lab T23	61.0	62.7	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
65	Lab T24	64.0	59.3	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59

66	Lab T25	59.3	61.3	50, 53, 55, 57, 59, 62	Y,Y,Y,Y,Y, Y	Strong bands	55-59
67	Lab T26	59.3	59.3	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
68	Lab T27	61.0	62.7	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
69	Lab T28	61.0	63.0	50, 53, 55, 57, 59, 62	Y,Y,Y,Y,Y, Y	Stutter bands, longer than expected	-
70	Lab T29	64.4	59.7	50, 53, 55, 57, 59, 62	Y,Y,Y,Y,Y, Y	Strong bands, longer than expected	55-59
71	Lab T30	62.7	58.4	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
72	Lab T31	61.0	57.6	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
73	Lab T32	61.0	62.7	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
74	Lab T33	62.7	61.0	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59
75	Lab T34	59.3	61.0	50, 53, 55, 57, 59, 62	N,N,N,N,N, N	Strong bands	55-59

Tm(^{0}C) –melting (denaturing) temperature (Europhins MWG); Y- Presence of secondary products; N- absence of secondary products; n.a. – no amplification (no migration); Bp=base pairs

Appendix IV. List of 189 lablab accessions genotyped using 8 SSR markers

Code	Accession	Name of population	Region of collection	Sub species
1	NH13	Advan. Line	KALRO	Purpureus
2	NH17	Advan. Line	KALRO	Purpureus
3	NH7	Advan. Line	KALRO	Purpureus
4	OH12	Advan. Line	KALRO	Purpureus
5	NH8	Advan. Line	KALRO	Purpureus
6	NH4	Advan. Line	KALRO	Purpureus
7	OH2	Advan. Line	KALRO	Purpureus
8	NH1	Advan. Line	KALRO	Purpureus
9	ILRI 21061	Aus	ILRI	Purpureus
10	ILRI 21076	Aus	ILRI	Purpureus
11	ILRI 11617	Aus	ILRI	Purpureus
12	ILRI 21087	Aus	ILRI	Purpureus
13	ILRI 11612	Aus	ILRI	Purpureus
14	ILRI 21071	Aus	ILRI	Purpureus
15	ILRI 21059	Aus	ILRI	Purpureus
16	ILRI 14414	Aus	ILRI	Purpureus
17	ILRI 13692 B	Eth	ILRI	Bengalensis
18	ILRI 6930	Eth	ILRI	Bengalensis
19	ILRI 6536	Eth	ILRI	Purpureus
20	ILRI 13692	Eth	ILRI	Bengalensis
21	ILRI 13702	Eth	ILRI	Purpureus
22	ILRI 13701	Eth	ILRI	Purpureus
23	ILRI 13688	Eth	ILRI	Purpureus
24	ILRI 13687	Eth	ILRI	Purpureus
25	ILRI 13688 B	Eth	ILRI	Purpureus
26	ILRI 13685	Eth	ILRI	Purpureus
27	CPI 195851	Eth	ILRI	Purpureus
28	ILRI 13700	Eth	ILRI	Purpureus
29	ILRI 14442	Eth	ILRI	Purpureus
30	ILRI 13697	Eth	ILRI	Purpureus
31	CPI 347628	Eth	ILRI	Purpureus
32	ILRI 13704	Eth	ILRI	Purpureus
33	ILRI 24746	Eth	ILRI	Purpureus
34	ILRI 13694	Eth	ILRI	Purpureus
35	ILRI 13686	Eth	ILRI	Purpureus
36	ILRI 6930 A	Eth	ILRI	Bengalensis
37	ILRI 13695	Eth	ILRI	Purpureus

38	ILRI 6533	Eth	ILRI	Purpureus
39	ILRI 6537	Eth	ILRI	Purpureus
40	ILRI 6528	Eth	ILRI	Purpureus
41	ILRI 6535	Eth	ILRI	Purpureus
42	ILRI 7278	Eth	ILRI	Purpureus
43	ILRI 18618	Eth	ILRI	Purpureus
44	ILRI 18619	India	ILRI	Purpureus
45	INDIA 534	India	ILRI	Purpureus
46	DOL 414-2	India	ILRI	Purpureus
47	DOL 414-3	India	ILRI	Purpureus
48	ILRI 21032	India	ILRI	Bengalensis
49	ILRI 21034	India	ILRI	Purpureus
50	ILRI 147	India	ILRI	Purpureus
51	ILRI 21034	India	ILRI	Purpureus
52	ILRI 21070	India	ILRI	Purpureus
53	ILRI 21033	India	ILRI	Purpureus
54	ILRI 21088	India	ILRI	Bengalensis
55	KDD-1	Ken	Kenya	Purpureus
56	KDD-2	Ken	Kenya	Purpureus
57	ILRI 14901B	Ken	ILRI	Purpureus
58	ILRI 14445	Ken	ILRI	Purpureus
59	ILRI 14443	Ken	ILRI	Purpureus
60	ILRI 14490	Ken	ILRI	Purpureus
61	ILRI 14478	Ken	ILRI	Purpureus
62	ILRI 14481	Ken	ILRI	Purpureus
63	ILRI 14901	Ken	ILRI	Purpureus
64	ILRI 14488	Ken	ILRI	Purpureus
65	GBK 005380	Ken	Eastern	Purpureus
66	GBK 010409	Ken	Unknown	Purpureus
67	GBK 010436	Ken	Eastern	Purpureus
68	GBK 010439	Ken	Eastern	Purpureus
69	GBK 010494 A	Ken	Unknown	Purpureus
70	GBK 010707	Ken	Eastern	Purpureus
71	GBK 010708	Ken	Coast	Purpureus
72	GBK 010822	Ken	Coast	Purpureus
73	GBK 010824 A	Ken	Rift	Purpureus
74	GBK 010837	Ken	Rift	Purpureus
75	GBK 010843	Ken	Rift	Purpureus
76	GBK 012230	Ken	Unknown	Purpureus
77	GBK 10708	Ken	Coast	Purpureus
78	GBK 012221	Ken	Unknown	Purpureus
79	GBK 012219	Ken	Unknown	Purpureus
80	GBK 012215	Ken	Rift	Purpureus
81	GBK 012187	Ken	Coast	Purpureus

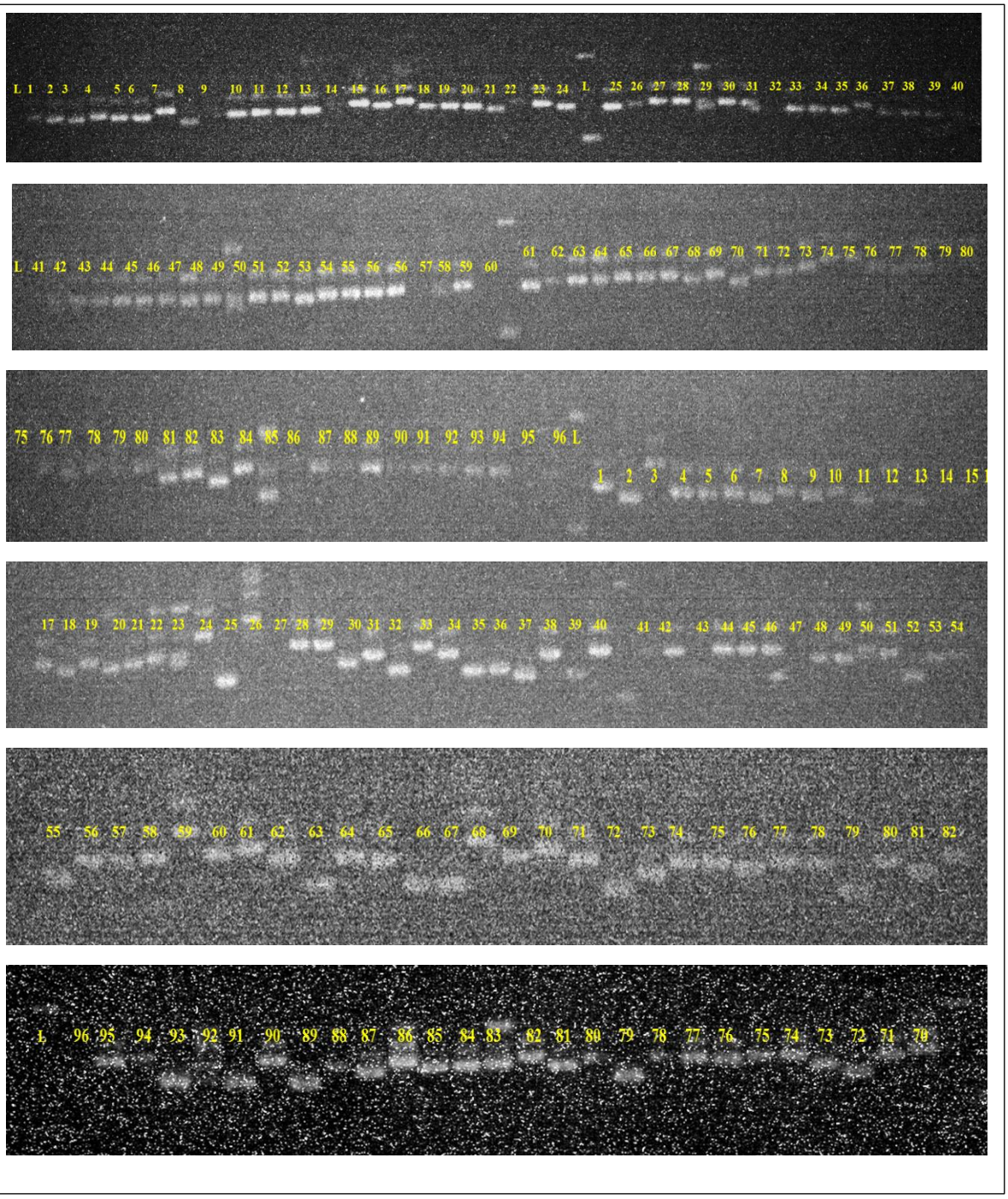
82	GBK 012038 B	Ken	Coast	Purpureus
83	GBK 012033	Ken	Unknown	Purpureus
84	GBK 012032	Ken	Unknown	Purpureus
85	GBK 012026 C	Ken	Coast	Purpureus
86	GBK 012000	Ken	Eastern	Purpureus
87	GBK 011803	Ken	Unknown	Purpureus
88	GBK 011733 B	Ken	Eastern	Purpureus
89	GBK 011723	Ken	Nairobi	Purpureus
90	GBK 011719	Ken	Eastern	Purpureus
91	GBK 013083	Ken	Eastern	Purpureus
92	GBK 013096	Ken	Eastern	Purpureus
93	GBK 028663 B	Ken	Eastern	Purpureus
94	GBK 045372	Ken	Eastern	Purpureus
95	GBK 010824	Ken	Rift	Purpureus
96	GBK 010494	Ken	Unknown	Purpureus
97	GBK 010707 B	Ken	Eastern	Purpureus
98	GBK 012026	Ken	Coast	Purpureus
99	GBK 010708	Ken	Coast	Purpureus
100	GBK 012038	Ken	Coast	Purpureus
101	GBK 010392	Ken	Unknown	Purpureus
102	GBK 011733	Ken	Eastern	Purpureus
103	GBK 013083 B	Ken	Eastern	Purpureus
104	Kakamega mkt.	Ken	West	Purpureus
105	Namanga	Ken	Rift	Purpureus
106	Nakuru 2	Ken	Rift	Purpureus
107	Bungoma 1	Ken	West	Purpureus
108	Kagio mkt.	Ken	Central	Purpureus
109	Kahuro A	Ken	Central	Purpureus
110	Muranga mkt.	Ken	Central	Purpureus
111	Kisumu	Ken	West	Purpureus
112	Lamu	Ken	Coast	Purpureus
113	Kagio B	Ken	Central	Purpureus
114	Bahati	Ken	Rift	Purpureus
115	Kahuro B	Ken	Eastern	Purpureus
116	Meru	Ken	Eastern	Purpureus
117	Mukinduri	Ken	Eastern	Purpureus
118	Kakamega mkt.14	Ken	West	Purpureus
119	Kitale mkt15	Ken	Rift	Purpureus
120	Kitui 20 A	Ken	Eastern	Purpureus
121	Kitui 20 C	Ken	Eastern	Purpureus
122	Kitui 17	Ken	Eastern	Purpureus
123	Kitale mkt.16	Ken	Rift	Purpureus
124	Makindu 18	Ken	Eastern	Purpureus
125	Matiliku mkt.11	Ken	Eastern	Purpureus

126	Machakos	Ken	Eastern	Purpureus
127	Kitale Wamuini 2	Ken	Rift	Purpureus
128	Kitale Kala 1	Ken	Rift	Purpureus
129	Kibwezi 5	Ken	Eastern	Purpureus
130	Bungoma 3	Ken	West	Purpureus
131	Kibwezi	Ken	Eastern	Purpureus
132	Matiliku mkt.12	Ken	Eastern	Purpureus
133	Kitale Mkt.16 B	Ken	Rift	Purpureus
134	Machakos 13	Ken	Eastern	Purpureus
135	Nakuru 1	Ken	Rift	Purpureus
136	DL 1002	Ken	Eastern	Purpureus
137	GBK 028663	Ken	Eastern	Purpureus
138	Kahuro	Ken	Central	Purpureus
139	Eldoret	Ken	Rift	Purpureus
140	Njoro	Ken	Rift	Purpureus
141	Kakamega	Ken	West	Purpureus
142	Lamu B	Ken	Coast	Purpureus
143	Kagio	Ken	Central	Purpureus
144	Bungoma 2	Ken	West	Purpureus
145	Meru B	Ken	Eastern	Purpureus
146	ILRI 14440	Ken	ILRI	Unicin 2 seeded
147	ILRI 14460	Ken	ILRI	Purpureus
148	ILRI 14458	Ken	ILRI	Purpureus
149	ILRI 14411	Ken	ILRI	Purpureus
150	ILRI 21046	SA	ILRI	Purpureus
151	ILRI 14437	SA	ILRI	Purpureus
152	ILRI 21084	SA	ILRI	Purpureus
153	ILRI 21045	SA	ILRI	Unicin 2
154	ILRI 14437 B	SA	ILRI	Purpureus
155	ILRI 24800	SA	ILRI	Unicin 2
156	ILRI 21048	SA	ILRI	Unicin 4
157	ILRI 21048 B	SA	ILRI	Unicin 4
158	ILRI 24777	SA	ILRI	Purpureus
159	ILRI 14415	SA	ILRI	Purpureus
160	ILRI 14412	SA	ILRI	Purpureus
161	ILRI 24800 B	SA	ILRI	Unicin 2 seeded
162	ILRI 24749	SA	ILRI	Unicin 4 seeded
163	ILRI 24781	SA	ILRI	Purpureus
164	ILRI 24796	SA	ILRI	Unicin 4 seeded
165	ILRI 21085	SA	ILRI	Purpureus
166	CPI 666245	SA	ILRI	Purpureus
167	ILRI 14435	SA	ILRI	Purpureus
168	ILRI 24778	SA	ILRI	Unicin 4 seeded
169	ILRI 21049	SA	ILRI	Purpureus

170	ILRI 24780	SA	ILRI	Purpureus
171	CPI 666243	SA	ILRI	Purpureus
172	ILRI 14419	SA	ILRI	Purpureus
173	ILRI 24777 B	SA	ILRI	Purpureus
174	ILRI 24799	SA	ILRI	Purpureus
175	ILRI 21083	SA	ILRI	Unicin 2 seeded
176	ILRI 21084 B	SA	ILRI	Purpureus
177	ILRI 21081	Uganda	ILRI	Purpureus
178	ILRI 21081 B	Uganda	ILRI	Purpureus
179	ILRI 24756	Uganda	ILRI	Unicin 4 seeded
180	ILRI 10979	USA	ILRI	Purpureus
181	ILRI 18611 G	USA	ILRI	Purpureus
182	ILRI 18611 P	USA	ILRI	Purpureus
183	ILRI 10953 B	WA	ILRI	Purpureus
184	ILRI 14441	WA	ILRI	Purpureus
185	ILRI 11615	WA	ILRI	Bengalensis
186	ILRI 10953	WA	ILRI	Purpureus
187	ILRI 24810	WA	ILRI	Bengalensis
188	ILRI 11630	WA	ILRI	Purpureus
189	ILRI 11614	WA	ILRI	Purpureus

WA: West Africa, SA: Southern Africa, Ken: Kenya, Eth: Ethiopia, Aus: Australia

Appendix V. Poly acrylamide gel picture showing PCR amplification of LabRRT 53 SSR loci on 189 lablab accessions. The accessions are indicated in yellow, the 1 Kb ladder is indicated by letter L.



Appendix VI. List of 240 lablab accessions genotyped using 9672 silicoDArT and 5075 SNP markers

Code	Genotype name	Population	Subspecies
1	GBK 028663	Ken	Pur
2	KDD	Ken	Pur
3	NH13	Al	Pur
4	NH17	Al	Pur
5	NH4	Al	Pur
6	NH7	Al	Pur
7	NH8	Al	Pur
8	FC Njoro	Ken	Pur
9	OH12	Al	Pur
10	OH2	Al	Pur
11	CPI 195851	Eth	Pur
12	CPI 347628	Eth	Pur
13	CPI 666243	SA	Pur
14	CPI 666245	SA	Pur
15	FC Bahati	Ken	Pur
16	FC Bahati_2	Ken	Pur
17	FC Bungoma	Ken	Pur
18	FC Bungoma_2	Ken	Pur
19	FC Eldoret	Ken	Pur
20	FC Kagio local	Ken	Pur
21	FC Kagio Market	Ken	Pur
22	FC Kahuro	Ken	Pur
23	FC Kahuro_1	Ken	Pur
24	FC Kahuro_2	Ken	Pur
25	FC Kahuro_3	Ken	Pur
26	Fc Kakamega Market	Ken	Pur
27	FC Kirinyaga Mwema	Ken	Pur
28	FC Kisumu	Ken	Pur
29	FC Kisumu_1	Ken	Pur
30	FC Lamu	Ken	Pur
31	FC Meru	Ken	Pur
32	FC Meru_1	Ken	Pur
33	FC Meru Market	Ken	Pur
34	FC Meru Mukinduri	Ken	Pur
35	FC Muranga Market	Ken	Pur
36	FC Namanga	Ken	Pur
37	FC Namanga_1	Ken	Pur
38	FC Njoro_1	Ken	Pur
39	FC Njoro_2	Ken	Pur
40	KDD_4	Ken	Pur

41	GBK 010392	Ken	Pur
42	GBK 010396	Ken	Pur
43	GBK 010396_1	Ken	Pur
44	GBK 010409	Ken	Pur
45	GBK 010436	Ken	Pur
46	GBK 010439	Ken	Pur
47	GBK 010439_1	Ken	Pur
48	GBK 010439_2	Ken	Pur
49	GBK 010494	Ken	Pur
50	GBK 010494_1	Ken	Pur
51	GBK 010707	Ken	Pur
52	GBK 010707_1	Ken	Pur
53	GBK 010708	Ken	Pur
54	GBK 010707_2	Ken	Pur
55	GBK 010712	Ken	Pur
56	GBK 010819	Ken	Pur
57	GBK 010822	Ken	Pur
58	GBK 010824	Ken	Pur
59	GBK 010837	Ken	Pur
60	GBK 010843	Ken	Pur
61	GBK 011083	Ken	Pur
62	GBK 011719	Ken	Pur
63	FC Eldoret_1	Ken	Pur
64	GBK 011719_1	Ken	Pur
65	GBK 011723	Ken	Pur
66	GBK 011733	Ken	Pur
67	GBK 011733_1	Ken	Pur
68	GBK 013096_1	Ken	Pur
69	GBK 011804	Ken	Pur
70	GBK 011803	Ken	Pur
71	GBK 012000	Ken	Pur
72	GBK 012000_1	Ken	Pur
73	GBK 012026	Ken	Pur
74	GBK 012026_1	Ken	Pur
75	GBK 012032	Ken	Pur
76	GBK 012033	Ken	Pur
77	GBK 012033_1	Ken	Pur
78	GBK 012038	Ken	Pur
79	GBK 012038_1	Ken	Pur
80	GBK 012187	Ken	Pur
81	GBK 010835	Ken	Pur
82	GBK 012215_1	Ken	Pur
83	GBK 012219	Ken	Pur
84	GBK 011804_1	Ken	Pur
85	GBK 012218	Ken	Pur

86	GBK 012221	Ken	Pur
87	GBK 012227	Ken	Pur
88	GBK 012227_1	Ken	Pur
89	GBK 012227_1	Ken	Pur
90	GBK 012230	Ken	Pur
91	GBK 013083	Ken	Pur
92	GBK 013083_1	Ken	Pur
93	GBK 013086	Ken	Pur
94	GBK 013096	Ken	Pur
95	GBK 013096_1	Ken	Pur
96	GBK 235	Ken	Pur
97	GBK 028663_1	Ken	Pur
98	GBK 028663_2	Ken	Pur
99	GBK 028663_3	Ken	Pur
100	GBK 028663_4	Ken	Pur
101	GBK 045372	Ken	Pur
102	GBK 045372_1	Ken	Pur
103	GBK 045372_2	Ken	Pur
104	GBK 538	Ken	Pur
105	GBK 538_1	Ken	Pur
106	ILRI 6536	Eth	Pur
107	GBK 010837	Ken	Pur
108	GBK 013096	Ken	Pur
109	ILRI 10953	Wa	Pur
110	ILRI 10953_1	Wa	Pur
111	ILRI 10979	Usa	Pur
112	ILRI 10979_1	Usa	Pur
113	ILRI 10979_1	Usa	Pur
114	ILRI 11612	Aus	Pur
115	ILRI 11612_1	Aus	Pur
116	ILRI 11614	Wa	Pur
117	GBK 010835_1	Ken	Pur
118	ILRI 11615	Wa	Ben
119	ILRI 11615_1	Wa	Ben
120	ILRI 11615_2	Wa	Ben
121	ILRI 11617	Aus	Pur
122	ILRI 11630	Wa	Pur
123	ILRI 11630_1	Wa	Pur
124	ILRI 13685	Eth	Pur
125	ILRI 13685_1	Eth	Pur
126	ILRI 13686	Eth	Pur
127	ILRI 13687	Eth	Pur
128	ILRI 13688	Eth	Pur
129	DL1002	Ken	Pur
130	ILRI 13692	Eth	Ben

131	ILRI 13694	Eth	Pur
132	ILRI 13696	Eth	Pur
133	ILRI 13700	Eth	Pur
134	ILRI 13702	Eth	Pur
135	ILRI 13702_1	Eth	Pur
136	ILRI 13704	Eth	Pur
137	ILRI 139	Unk	Pur
138	ILRI 144	Unk	Pur
139	ILRI 14414	Aus	Pur
140	ILRI 14415	SA	Pur
141	ILRI 14415_1	SA	Pur
142	ILRI 14435	SA	Pur
143	ILRI 14437	SA	Pur
144	ILRI 14437_1	SA	Pur
145	ILRI 14437_2	SA	Pur
146	ILRI 14440	Ken	Unic
147	ILRI 14441	Wa	Pur
148	ILRI 14441_1	Wa	Pur
149	ILRI 14442	Eth	Pur
150	ILRI 14442_1	Eth	Pur
151	ILRI 14443	Ken	Pur
152	ILRI 14443_1	Ken	Pur
153	ILRI 14445	Ken	Pur
154	ILRI 14445_1	Ken	Pur
155	ILRI 14458	Ken	Pur
156	ILRI 14460	Ken	Pur
157	ILRI 14471	Ken	Pur
158	ILRI 14478	Ken	Pur
159	ILRI 14481	Ken	Pur
160	ILRI 14481_1	Ken	Pur
161	ILRI 14489	Ken	Pur
162	ILRI 14490	Ken	Pur
163	ILRI 14490_1	Ken	Pur
164	ILRI 14490_2	Ken	Pur
165	ILRI 147	Ind	Pur
166	ILRI 148	Unk	Pur
167	ILRI 14901	Ken	Pur
168	ILRI 14901_1	Ken	Pur
169	ILRI 15436	Unk	Pur
170	ILRI 15436_1	Unk	Pur
171	ILRI 18611	Usa	Pur
172	ILRI 18611_1	Usa	Pur
173	ILRI 18611_2	Usa	Pur
174	ILRI 18618	Eth	Pur
175	ILRI 18619	Ind	Pur

176	ILRI 21029	Ind	Pur
177	ILRI 21033	Ind	Pur
178	ILRI 21034	Ind	Pur
179	ILRI 21046	SA	Pur
180	ILRI 21049	SA	Pur
181	ILRI 21059	Aus	Pur
182	ILRI 21061	Aus	Pur
183	ILRI 21061_1	Aus	Pur
184	ILRI 21070	Ind	Pur
185	ILRI 21071	Aus	Pur
186	ILRI 21081	Uganda	Pur
187	ILRI 21081_1	Uganda	Pur
188	ILRI 21085	SA	Pur
189	ILRI 21087	Aus	Pur
190	ILRI 21089	Unk	Pur
191	ILRI 24746	Eth	Pur
192	ILRI 24746_1	Eth	Pur
193	ILRI 24777	SA	Pur
194	ILRI 24780	SA	Pur
195	ILRI 24781	SA	Pur
196	ILRI 24788	SA	Pur
197	ILRI 6528	Eth	Pur
198	ILRI 6528_1	Eth	Pur
199	ILRI 6533	Eth	Pur
200	ILRI 6535	Eth	Pur
201	ILRI 6536	Eth	Pur
202	ILRI 6536_1	Eth	Pur
203	ILRI 6537	Eth	Pur
204	ILRI 6930	Eth	Ben
205	ILRI 7278	Eth	Pur
206	GBK 010738	Ken	Pur
207	DL 1002	Ken	Pur
208	Katumani_1	Ken	Pur
209	Machakos_1	Ken	Pur
210	Katumani_4	Ken	Pur
211	DL 1002_1	Ken	Pur
212	DL 1002_2	Ken	Pur
213	DL 1002_3	Ken	Pur
214	Katumani_1	Ken	Pur
215	Katumani_2	Ken	Pur
216	DL 1002_4	Ken	Pur
217	Katumani_3	Ken	Pur
218	Machakos_2	Ken	Pur
219	Katumani_3	Ken	Pur
220	Machakos_3	Ken	Pur

221	KDD_1	Ken	Pur
222	KDD_2	Ken	Pur
223	OH_10	Al	Pur
224	KDD_3	Ken	Pur
225	India 414	Ind	Pur
226	India 414_1	Ind	Pur
227	India 548	Ind	Pur
228	Machakos_4	Ken	Pur
229	ILRI 21076	Aus	Pur
230	ILRI 21076_1	Aus	Pur
231	ILRI 14440	Ken	Unic
232	ILRI 24749	SA	Unic
233	ILRI 21083	SA	Unic
234	ILRI 24756	Uga	Unic
235	ILRI 21046	SA	Unic
236	CPI 666243	SA	Unic
237	ILRI 21048	SA	Unic
238	ILRI 24800	SA	Unic
239	ILRI 14437	SA	Unic
240	ILRI 14440_1	Ken	Unic

WA: West Africa, **SA**: Southern Africa, **Ken**: Kenya, **Eth**: Ethiopia, Aus: **Uga**: Uganda, **Unk**: Unknown, **Al**: Advanced lines, **Ind**: India, **Aus**: Australia **Pur**: purpureus, **Unic**: unicinatus

Appendix VII. Selected Nei & Li (Dice) genetic dissimilarity matrices generated from DArT markers data and used to establish the level of relatedness of some lablab accessions

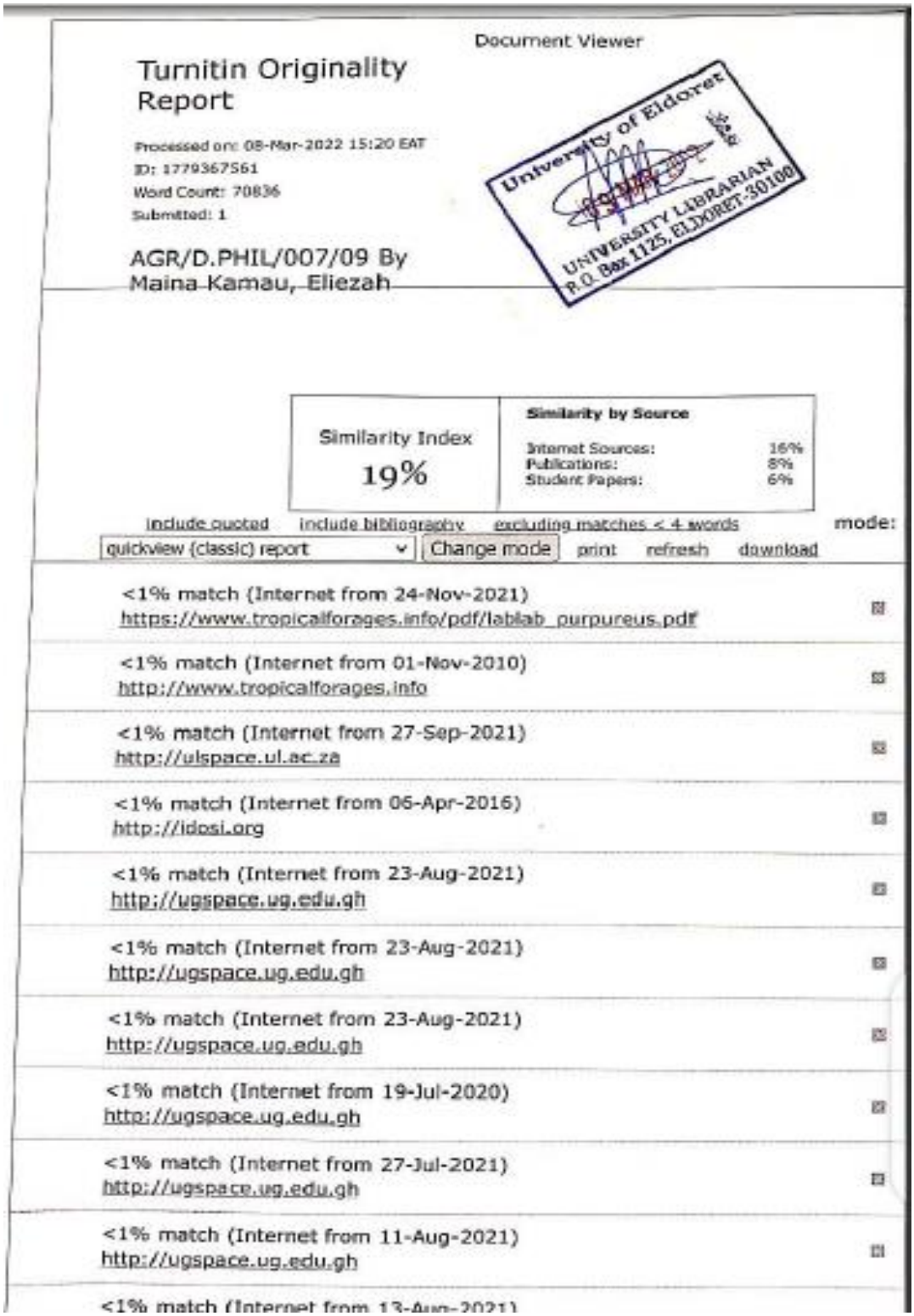
Genotypes code	109	141	129	130	175	119	140	40	186	27	116	94
141	0.20											
129	0.08	0.19										
130	0.14	0.21	0.13									
175	0.14	0.21	0.14	0.07								
119	0.23	0.26	0.22	0.24	0.24							
140	0.25	0.29	0.25	0.26	0.27	0.22						
40	0.23	0.26	0.23	0.24	0.25	0.27	0.30					
186	0.24	0.27	0.24	0.25	0.26	0.28	0.31	0.26				
27	0.22	0.25	0.22	0.23	0.24	0.26	0.29	0.14	0.25			
116	0.24	0.27	0.24	0.25	0.26	0.28	0.31	0.25	0.27	0.24		
162	0.25	0.28	0.24	0.26	0.26	0.29	0.31	0.25	0.28	0.24	0.27	
94	0.31	0.34	0.30	0.32	0.32	0.35	0.38	0.32	0.34	0.30	0.33	
184	0.31	0.34	0.30	0.32	0.32	0.35	0.38	0.32	0.34	0.30	0.33	0.00
150	0.24	0.27	0.24	0.25	0.26	0.28	0.31	0.25	0.27	0.24	0.26	0.31
63	0.23	0.27	0.23	0.25	0.25	0.28	0.30	0.22	0.27	0.21	0.26	0.32
64	0.23	0.26	0.22	0.24	0.24	0.27	0.30	0.25	0.19	0.24	0.26	0.33
134	0.23	0.26	0.23	0.24	0.25	0.27	0.30	0.22	0.26	0.21	0.25	0.32
181	0.32	0.36	0.32	0.33	0.34	0.37	0.39	0.33	0.36	0.32	0.34	0.37
124	0.23	0.26	0.22	0.24	0.24	0.27	0.29	0.25	0.19	0.24	0.26	0.33
137	0.26	0.30	0.26	0.27	0.28	0.31	0.33	0.27	0.30	0.26	0.28	0.28
113	0.27	0.30	0.27	0.28	0.29	0.31	0.34	0.28	0.30	0.27	0.29	0.28
102	0.24	0.27	0.23	0.25	0.25	0.28	0.31	0.26	0.20	0.25	0.27	0.34
235	0.92	0.95	0.91	0.93	0.93	0.96	0.99	0.92	0.95	0.91	0.94	0.96
105	0.91	0.95	0.91	0.92	0.93	0.96	0.98	0.92	0.95	0.91	0.93	0.96
185	0.22	0.26	0.22	0.23	0.24	0.27	0.29	0.25	0.18	0.24	0.26	0.33
96	0.20	0.18	0.20	0.22	0.22	0.27	0.29	0.27	0.28	0.26	0.28	0.35
147	0.91	0.95	0.91	0.92	0.93	0.96	0.98	0.92	0.95	0.91	0.93	0.96
169	0.26	0.30	0.26	0.27	0.28	0.30	0.33	0.27	0.30	0.26	0.28	0.28
117	0.22	0.25	0.22	0.23	0.24	0.26	0.29	0.24	0.18	0.23	0.25	0.32
221	0.22	0.26	0.22	0.23	0.24	0.27	0.29	0.25	0.18	0.24	0.26	0.33
115	0.29	0.32	0.29	0.30	0.31	0.33	0.36	0.30	0.32	0.29	0.31	0.30
62	0.23	0.26	0.23	0.24	0.25	0.27	0.30	0.22	0.26	0.21	0.25	0.32
103	0.23	0.26	0.22	0.24	0.24	0.27	0.30	0.22	0.26	0.20	0.25	0.31
180	0.33	0.36	0.33	0.34	0.35	0.37	0.40	0.34	0.36	0.33	0.35	0.38
108	0.23	0.26	0.22	0.24	0.24	0.27	0.30	0.22	0.26	0.20	0.25	0.31
148	0.23	0.26	0.22	0.24	0.24	0.27	0.30	0.22	0.26	0.20	0.25	0.31
39	0.23	0.26	0.22	0.24	0.24	0.27	0.30	0.25	0.19	0.24	0.26	0.33
163	0.28	0.31	0.28	0.29	0.30	0.32	0.35	0.29	0.31	0.28	0.30	0.29
3	0.28	0.32	0.28	0.29	0.30	0.33	0.35	0.29	0.32	0.28	0.30	0.30
75	0.23	0.26	0.23	0.24	0.25	0.27	0.30	0.25	0.19	0.24	0.27	0.33

155	0.98	1.01	0.97	0.99	0.99	1.02	1.05	0.98	1.01	0.97	1.00	1.03
198	0.22	0.25	0.22	0.23	0.24	0.26	0.29	0.24	0.18	0.23	0.26	0.32
133	0.22	0.26	0.22	0.23	0.24	0.27	0.29	0.25	0.18	0.24	0.26	0.33
19	0.24	0.27	0.24	0.25	0.26	0.28	0.31	0.26	0.00	0.25	0.27	0.34
214	0.24	0.27	0.24	0.25	0.26	0.28	0.31	0.25	0.27	0.24	0.00	0.33
138	0.29	0.32	0.29	0.30	0.31	0.33	0.36	0.30	0.32	0.28	0.31	0.30
157	0.28	0.31	0.28	0.29	0.30	0.32	0.35	0.29	0.31	0.28	0.30	0.29
160	0.29	0.32	0.28	0.30	0.30	0.33	0.36	0.29	0.32	0.28	0.31	0.30
121	0.22	0.25	0.22	0.23	0.24	0.26	0.29	0.24	0.13	0.23	0.25	0.32
168	0.28	0.31	0.28	0.29	0.30	0.32	0.35	0.30	0.19	0.29	0.31	0.38
125	0.26	0.30	0.26	0.27	0.28	0.31	0.33	0.27	0.30	0.26	0.28	0.28
37	0.24	0.27	0.24	0.25	0.26	0.28	0.31	0.26	0.00	0.25	0.27	0.34
207	0.16	0.16	0.16	0.17	0.18	0.23	0.25	0.23	0.24	0.22	0.24	0.31
77	0.25	0.28	0.25	0.26	0.27	0.29	0.32	0.26	0.28	0.25	0.27	0.31
5	0.22	0.25	0.22	0.23	0.24	0.26	0.29	0.23	0.25	0.22	0.11	0.31
218	0.20	0.15	0.19	0.21	0.21	0.26	0.29	0.26	0.27	0.25	0.28	0.34
53	0.18	0.19	0.18	0.19	0.20	0.24	0.27	0.24	0.26	0.23	0.26	0.32
212	0.23	0.26	0.23	0.24	0.25	0.20	0.15	0.28	0.29	0.27	0.29	0.35
18	0.19	0.14	0.19	0.20	0.21	0.25	0.28	0.25	0.26	0.24	0.27	0.33
65	0.24	0.27	0.24	0.25	0.26	0.28	0.31	0.25	0.27	0.24	0.16	0.33
219	0.23	0.27	0.23	0.24	0.25	0.28	0.30	0.24	0.27	0.23	0.16	0.32
84	0.22	0.25	0.22	0.23	0.24	0.26	0.29	0.24	0.18	0.23	0.26	0.32
1	0.24	0.28	0.24	0.25	0.26	0.21	0.19	0.29	0.30	0.28	0.30	0.37
13	0.16	0.16	0.16	0.17	0.18	0.23	0.25	0.23	0.24	0.22	0.24	0.31
215	0.21	0.24	0.21	0.22	0.23	0.25	0.28	0.23	0.19	0.22	0.24	0.31
193	0.19	0.15	0.19	0.20	0.21	0.26	0.28	0.26	0.27	0.25	0.27	0.34
61	0.23	0.26	0.23	0.24	0.25	0.27	0.30	0.24	0.26	0.23	0.15	0.32
30	0.17	0.16	0.16	0.18	0.18	0.23	0.26	0.23	0.24	0.22	0.24	0.31
87	0.23	0.26	0.23	0.24	0.25	0.27	0.30	0.24	0.26	0.23	0.15	0.32
194	0.24	0.27	0.24	0.25	0.26	0.21	0.16	0.28	0.29	0.27	0.30	0.36
190	0.23	0.27	0.23	0.24	0.25	0.28	0.30	0.26	0.19	0.25	0.27	0.33
152	0.27	0.30	0.27	0.28	0.29	0.24	0.18	0.31	0.32	0.30	0.33	0.39
223	0.22	0.26	0.22	0.23	0.24	0.27	0.29	0.25	0.18	0.24	0.26	0.33
71	0.23	0.26	0.22	0.24	0.24	0.27	0.30	0.25	0.19	0.24	0.26	0.33
78	0.23	0.26	0.23	0.24	0.25	0.27	0.30	0.22	0.26	0.21	0.25	0.32
48	0.23	0.26	0.22	0.24	0.24	0.20	0.14	0.27	0.28	0.26	0.28	0.35
206	0.23	0.26	0.23	0.24	0.25	0.19	0.23	0.27	0.28	0.26	0.29	0.35
220	0.22	0.26	0.22	0.23	0.24	0.27	0.29	0.25	0.18	0.24	0.26	0.33
213	0.23	0.26	0.22	0.24	0.24	0.27	0.30	0.25	0.19	0.24	0.26	0.33
201	0.23	0.27	0.23	0.24	0.25	0.28	0.30	0.26	0.19	0.25	0.27	0.34
9	0.24	0.27	0.24	0.25	0.26	0.16	0.24	0.28	0.29	0.27	0.30	0.36
47	0.18	0.22	0.18	0.19	0.20	0.15	0.14	0.23	0.24	0.22	0.24	0.31
8	0.20	0.16	0.20	0.21	0.22	0.27	0.29	0.27	0.28	0.26	0.28	0.35
22	0.24	0.28	0.24	0.25	0.26	0.29	0.31	0.27	0.20	0.26	0.28	0.34
24	0.24	0.27	0.23	0.25	0.25	0.28	0.31	0.26	0.22	0.25	0.27	0.34
232	0.22	0.25	0.22	0.23	0.24	0.26	0.29	0.24	0.18	0.23	0.25	0.32
56	0.24	0.27	0.24	0.25	0.26	0.28	0.31	0.26	0.20	0.25	0.27	0.34

225	0.29	0.32	0.29	0.30	0.31	0.26	0.20	0.33	0.34	0.32	0.35	0.41
202	0.25	0.28	0.25	0.26	0.27	0.22	0.16	0.29	0.30	0.28	0.31	0.37
89	0.23	0.26	0.22	0.24	0.24	0.27	0.30	0.25	0.19	0.24	0.26	0.33
222	0.22	0.25	0.22	0.23	0.24	0.26	0.29	0.24	0.18	0.23	0.26	0.32
58	0.20	0.16	0.20	0.21	0.22	0.26	0.29	0.27	0.28	0.26	0.28	0.35
196	0.22	0.25	0.22	0.23	0.24	0.26	0.29	0.24	0.18	0.23	0.26	0.32
20	0.23	0.27	0.23	0.24	0.25	0.16	0.23	0.28	0.29	0.27	0.29	0.36
205	0.23	0.26	0.22	0.24	0.24	0.27	0.30	0.25	0.19	0.24	0.26	0.33
33	0.23	0.26	0.22	0.24	0.24	0.27	0.30	0.25	0.19	0.24	0.26	0.33
44	0.22	0.25	0.22	0.23	0.24	0.26	0.29	0.24	0.18	0.23	0.26	0.32
195	0.76	0.80	0.76	0.77	0.78	0.81	0.83	0.77	0.80	0.76	0.78	0.81
236	0.23	0.27	0.23	0.24	0.25	0.28	0.30	0.24	0.27	0.23	0.07	0.32
82	0.22	0.25	0.22	0.23	0.24	0.19	0.13	0.26	0.27	0.25	0.28	0.34
11	0.24	0.28	0.24	0.25	0.26	0.29	0.31	0.25	0.28	0.24	0.00	0.33
36	0.24	0.27	0.23	0.25	0.25	0.28	0.31	0.26	0.20	0.25	0.27	0.34
12	0.18	0.18	0.18	0.20	0.20	0.25	0.27	0.25	0.26	0.24	0.26	0.33
32	0.23	0.26	0.22	0.24	0.24	0.20	0.19	0.27	0.28	0.26	0.28	0.35
68	0.19	0.22	0.19	0.20	0.21	0.23	0.26	0.21	0.15	0.20	0.23	0.29
72	0.24	0.27	0.24	0.25	0.26	0.28	0.31	0.26	0.20	0.25	0.27	0.34
31	0.27	0.30	0.27	0.28	0.29	0.24	0.18	0.31	0.32	0.30	0.33	0.39
81	0.20	0.16	0.20	0.21	0.22	0.27	0.29	0.27	0.28	0.26	0.28	0.35
200	0.19	0.15	0.19	0.20	0.21	0.26	0.28	0.26	0.27	0.25	0.27	0.34
233	0.22	0.25	0.22	0.23	0.24	0.26	0.29	0.24	0.18	0.23	0.26	0.32
224	0.21	0.25	0.21	0.22	0.23	0.26	0.28	0.24	0.19	0.23	0.25	0.31
209	0.22	0.26	0.22	0.23	0.24	0.27	0.29	0.25	0.18	0.24	0.26	0.32
111	0.89	0.93	0.89	0.90	0.91	0.94	0.96	0.90	0.93	0.89	0.91	0.94
231	0.23	0.26	0.23	0.24	0.25	0.27	0.30	0.25	0.19	0.24	0.27	0.33
85	0.19	0.14	0.19	0.20	0.21	0.25	0.28	0.25	0.27	0.24	0.27	0.33
142	0.23	0.26	0.23	0.24	0.25	0.20	0.15	0.28	0.29	0.27	0.29	0.35
38	0.23	0.26	0.23	0.24	0.25	0.27	0.30	0.25	0.19	0.24	0.26	0.33
167	0.22	0.25	0.22	0.23	0.24	0.26	0.29	0.24	0.18	0.23	0.25	0.32
107	0.23	0.26	0.22	0.24	0.24	0.27	0.30	0.25	0.19	0.24	0.26	0.33
104	0.23	0.26	0.22	0.24	0.24	0.27	0.30	0.22	0.26	0.20	0.25	0.31
118	0.20	0.15	0.19	0.21	0.21	0.26	0.29	0.26	0.27	0.25	0.28	0.34
16	0.23	0.26	0.23	0.24	0.25	0.20	0.15	0.28	0.29	0.27	0.29	0.35
28	0.24	0.27	0.24	0.25	0.26	0.28	0.31	0.26	0.20	0.25	0.27	0.34
234	0.16	0.16	0.16	0.17	0.18	0.23	0.25	0.23	0.24	0.22	0.24	0.31
176	0.24	0.27	0.24	0.25	0.26	0.28	0.31	0.26	0.20	0.25	0.27	0.34
4	0.24	0.28	0.24	0.25	0.26	0.29	0.31	0.27	0.20	0.26	0.28	0.34
239	0.21	0.24	0.21	0.22	0.23	0.25	0.28	0.23	0.17	0.22	0.24	0.31
241	0.22	0.26	0.22	0.23	0.24	0.27	0.29	0.25	0.18	0.24	0.26	0.33
2	0.30	0.33	0.30	0.31	0.32	0.34	0.37	0.31	0.33	0.30	0.32	0.31
173	0.25	0.28	0.24	0.26	0.26	0.29	0.31	0.25	0.28	0.24	0.27	0.31
29	0.24	0.27	0.23	0.25	0.25	0.28	0.31	0.24	0.27	0.23	0.26	0.29
149	0.22	0.26	0.22	0.23	0.24	0.27	0.29	0.25	0.18	0.23	0.26	0.32
97	0.28	0.32	0.28	0.29	0.30	0.32	0.35	0.29	0.32	0.28	0.30	0.26
144	0.24	0.28	0.24	0.25	0.26	0.29	0.31	0.25	0.28	0.24	0.26	0.22

136	0.23	0.26	0.22	0.24	0.24	0.27	0.30	0.22	0.26	0.20	0.25	0.31
26	0.23	0.26	0.22	0.24	0.24	0.27	0.30	0.25	0.19	0.24	0.26	0.33
110	0.28	0.31	0.27	0.29	0.29	0.32	0.35	0.28	0.31	0.27	0.30	0.20
183	0.29	0.32	0.29	0.30	0.31	0.33	0.36	0.30	0.32	0.29	0.31	0.21
146	0.31	0.35	0.31	0.33	0.33	0.36	0.38	0.32	0.35	0.31	0.34	0.33
165	0.86	0.90	0.86	0.88	0.88	0.91	0.93	0.87	0.90	0.86	0.89	0.91
182	0.29	0.32	0.29	0.30	0.31	0.33	0.36	0.30	0.32	0.29	0.31	0.32
139	0.90	0.93	0.89	0.91	0.91	0.94	0.97	0.90	0.93	0.89	0.92	0.95
159	0.27	0.30	0.26	0.28	0.28	0.31	0.34	0.28	0.30	0.26	0.29	0.29
114	0.87	0.90	0.87	0.88	0.89	0.91	0.94	0.88	0.90	0.87	0.89	0.92
100	0.16	0.18	0.16	0.17	0.18	0.23	0.25	0.23	0.24	0.22	0.24	0.31
126	0.26	0.30	0.26	0.28	0.28	0.31	0.33	0.27	0.30	0.26	0.29	0.29
135	0.28	0.31	0.27	0.29	0.29	0.32	0.35	0.28	0.31	0.27	0.30	0.30
171	0.25	0.28	0.24	0.26	0.26	0.29	0.32	0.25	0.28	0.24	0.27	0.26
179	0.91	0.94	0.90	0.92	0.92	0.95	0.97	0.91	0.94	0.90	0.93	0.95
153	0.84	0.88	0.84	0.86	0.86	0.89	0.91	0.85	0.88	0.84	0.87	0.89
154	0.86	0.89	0.86	0.87	0.88	0.90	0.93	0.87	0.89	0.86	0.88	0.91
145	0.80	0.84	0.80	0.82	0.82	0.85	0.87	0.81	0.84	0.80	0.83	0.85
170	0.25	0.28	0.24	0.26	0.26	0.29	0.31	0.25	0.28	0.24	0.00	0.33
112	0.25	0.28	0.25	0.26	0.27	0.29	0.32	0.26	0.28	0.25	0.27	0.26
101	0.23	0.26	0.23	0.24	0.25	0.20	0.15	0.28	0.29	0.27	0.29	0.35
178	0.23	0.26	0.23	0.24	0.25	0.20	0.15	0.28	0.29	0.27	0.29	0.35
188	0.22	0.26	0.22	0.23	0.24	0.27	0.29	0.21	0.26	0.20	0.24	0.31
227	0.22	0.26	0.22	0.23	0.24	0.27	0.29	0.21	0.26	0.20	0.24	0.31
51	0.98	1.01	0.97	0.99	0.99	1.02	1.05	0.98	1.01	0.97	1.00	1.03
166	0.98	1.01	0.97	0.99	0.99	1.02	1.05	0.98	1.01	0.97	1.00	1.03
86	0.31	0.34	0.30	0.32	0.32	0.35	0.38	0.32	0.34	0.30	0.33	0.00
120	0.28	0.32	0.28	0.29	0.30	0.33	0.35	0.29	0.32	0.28	0.30	0.21
203	0.27	0.30	0.27	0.28	0.29	0.24	0.18	0.31	0.32	0.30	0.33	0.39
242	0.23	0.27	0.23	0.24	0.25	0.28	0.30	0.26	0.19	0.25	0.27	0.33
50	0.27	0.30	0.27	0.28	0.29	0.31	0.34	0.28	0.30	0.27	0.29	0.28
99	0.27	0.31	0.27	0.28	0.29	0.31	0.34	0.28	0.30	0.27	0.29	0.29
177	0.29	0.32	0.28	0.30	0.30	0.33	0.35	0.29	0.32	0.28	0.31	0.10
95	0.31	0.35	0.31	0.33	0.33	0.36	0.38	0.32	0.35	0.31	0.34	0.33
208	0.32	0.35	0.31	0.33	0.33	0.36	0.39	0.33	0.35	0.31	0.34	0.33
174	0.22	0.25	0.22	0.23	0.24	0.26	0.29	0.24	0.18	0.23	0.26	0.32
204	0.22	0.26	0.22	0.23	0.24	0.27	0.29	0.25	0.18	0.24	0.26	0.33
74	0.23	0.26	0.22	0.24	0.24	0.27	0.30	0.25	0.19	0.24	0.26	0.33
131	0.22	0.26	0.22	0.23	0.24	0.27	0.29	0.25	0.18	0.24	0.26	0.33

Appendix VIII. Thesis originality report



Document Viewer

Turnitin Originality Report

Processed on: 08-Mar-2022 15:20 EAT
ID: 1779367561
Word Count: 70836
Submitted: 1

AGR/D.PHIL/007/09 By Maina Kamau, Eliezah

University of Eldoret
UNIVERSITY LIBRARIAN
P.O. Box 1125, ELDORET-30100

Similarity Index	Similarity by Source	
19%	Internet Sources:	16%
	Publications:	8%
	Student Papers:	6%

include quoted include bibliography excluding matches < 4 words mode:

quickview (classic) report Change mode print refresh download

<1% match (Internet from 24-Nov-2021)
https://www.tropicalforages.info/pdf/lablab_purpureus.pdf

<1% match (Internet from 01-Nov-2010)
http://www.tropicalforages.info

<1% match (Internet from 27-Sep-2021)
http://ulspace.ul.ac.za

<1% match (Internet from 06-Apr-2016)
http://idosi.org

<1% match (Internet from 23-Aug-2021)
http://ugspace.ug.edu.gh

<1% match (Internet from 23-Aug-2021)
http://ugspace.ug.edu.gh

<1% match (Internet from 23-Aug-2021)
http://ugspace.ug.edu.gh

<1% match (Internet from 19-Jul-2020)
http://ugspace.ug.edu.gh

<1% match (Internet from 27-Jul-2021)
http://ugspace.ug.edu.gh

<1% match (Internet from 11-Aug-2021)
http://ugspace.ug.edu.gh

<1% match (Internet from 13-Aug-2021)