NUCLEOTIDE DIVERSITY OF COMMON BEAN PHASEOLIN (α-PHS) GENE AND ITS ASSOCIATION WITH SEED PROTEIN CONTENT

Barasa Sonia Achieng

I56/11087/2018

A thesis submitted in partial fulfillment of the requirements for the award of degree of Master of Science in Bioinformatics in the Department of Biochemistry of the University of Nairobi

November, 2022

DECLARATION

This thesis is entirely my work and has not been submitted elsewhere for examination, award of degree or publication.

Barasa Sonia Achieng

Registration No: I56/11087/2018

Signature:__________ Date: 16th November,2022

Supervisors' Approval

The research thesis has been submitted for examination with our approval as university supervisors:

Dr. Evans Nyaboga

Department of Biochemistry, University of Nairobi

Signature:__________ Date: 16th November, 2022

Dr. Christine Adhiambo

Department of Biochemistry, University of Nairobi

Signature:__________ Date: 16th November, 2022

DEDICATION

I dedicate this thesis to my adorable parents Mr. & Mrs. Elisha Barasa who encouraged me to go on every adventure, especially this one. To my siblings Hellen Barasa and Joy Barasa for being a core part of my life.

ACKNOWLEDGMENTS

I am grateful to God through whom all things are possible. I wish to sincerely acknowledge my supervisors, who have been my biggest inspiration and great mentors. Dr. Evans Nyaboga, my deepest gratitude goes to you for your patience with me during this time and unwavering professional support. I wish to thank you for the listening ear you granted me during the many consultations I made in your office. Thank you for your patience in going through my work, guiding and encouraging me and for your input in every step of the way. I am also grateful to you Sir, for facilitating my needs in the laboratory necessary for the generation of the data and ensuring I achieved my objectives. To Dr. Christine Adhiambo, I wish to thank you sincerely for your immense, unwavering support and encouragement towards the realization of this research project. Despite your very busy schedule, you spared time going through my work keenly and giving your input all the way. Your constant inquiry on my progress pushed me to finish this work. May God richly bless you for your kind mentorship during the entire project period.

To my dear father, Mr. Elisha Barasa, your unconditional support, determination and selflessness has been incomprehensible. You have been my greatest cheerleader to complete this work. Thank you for consistently believing in me and encouraging me to continue to the end even when I slacked. I will always be grateful to you dad, for your emotional and financial support. To my mother, Mrs. Elizabeth Barasa, your endless prayers and support kept me going throughout this project phase. To my siblings Hellen and Joy, thank you for your unconditional support, in your own ways you made this work possible. My gratitude also goes to the Head Department of Biochemistry (University of Nairobi), Dr. Gerald Juma and Research Coordinator CEBIB, Prof.

George Obiero for the conducive environment they offered while doing the project. I wish to thank my friend Dr. Kevin Ochwedo for his technical advice on the analysis. To my colleagues who helped me in one way or another, I am truly grateful to all of you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND ACRONYMS

%	Percent
°C	Degree Celsius
µL	Microliter
AFLP	Amplified Fragment Length Polymorphism
AM	Arbuscular Mycorrhiza
BC	Before Christ
BSA	Bovine Serum Albumin
cDNA	Complementary DNA
CEBIB	Center for Biotechnology and Bioinformatics
DNA	Deoxyribonucleic Acid
DnaSP	DNA Sequence Polymorphism
dNTPs	Deoxyribonucleotide Triphosphates
dwb	Dry weight basis
EDTA	Ethylenediaminetetraacetic acid
HCl	Hydrochloric acid
KDa	Kilodalton
M	Molar
MCL	Maximum Composite Likelihood
MEGA	Molecular Evolutionary Genetic Analysis
mg	Milligram

mL	Milliliter
MLM	Mixed Linear Model
MSA	Multiple Sequence Alignment
MUSCLE	Multiple Sequence Comparison by Log-Expectation\
NJ	Neighbor joining
NCBI	National Center for Biotechnology
PCR	Polymerase Chain Reaction
pH	Potential of Hydrogen
RFLP	Restriction Fragment Length Polymorphism
Rpm	Revolutions Per Minute
SDS	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SNP	Single Nucleotide Polymorphisms
SSRs	Simple Sequence Repeats
T CS	Transitive Consistency Score
TASSEL	Trait Analysis by Association
TBE	Tris Boric Ethylenediaminetetraacetic acid
UPGMA	Unweighted Pair Group Method with Arithmetic Mean
UV	Ultra violet
V	Voltage
V/v	Volume by Volume
W/v	Weight by Volume
WHO	World Health Organization

ABSTRACT

Phaseolin (α-Phs) is the most abundant protein reserve in seeds of common beans accounting for 40 – 50 % of the total seed protein. Despite having low methionine (0.5 - 0.80 %) content, phaseolin is the primary source of amino acids in common bean seeds. More than 40 genetic variants of phaseolin differing in amino acid composition have been reported. Therefore, phaseolin gene diversity could be used as a strategy to improve the nutritional value of protein in common beans. To date, no information is available on the relationship between natural nucleotide polymorphisms of α-Phs gene and seed protein content in common beans. This study was conducted to determine natural nucleotide polymorphism in α-Phs gene and their association with protein content in dry seeds of common bean. Eleven selected common bean accessions were planted in plastic pots in the greenhouse. Young leaves of 4-week-old plants were used for extraction of genomic DNA, followed by polymerase chain reaction (PCR) amplification using primers specific to different fragments of the phaseolin gene. Amplified PCR products were sequenced, sequences edited and analysed for nucleotide polymorphisms to infer levels of genetic variability, genetic diversity indices and other evolutionary analyses including haplotype diversity, neutrality tests, linkage disequilibrium and recombination events using DNA Sequence Polymorphism (DnaSP) software. Amino acid/codon changes occurring on sequenced α-Phs gene of the common bean accessions were elucidated using Codon Code Aligner software. Dry mature seeds of the selected common bean accessions were harvested and analysed for the total protein content using Lowry protein method. The association of α-Phs gene sequence polymorphisms and protein content was determined. The full-length sequence of α-Phs gene revealed a total of 41 genetic variants which consisted of 24 single nucleotide polymorphisms (SNPs) and 17 indels/parsimony informative sites. Ninety percent of the segregated sites in the coding region of the gene resulted in non- synonymous mutations. The coding region polymorphisms classified the α-Phs gene into 9 distinct haplotypes. The full-length sequence had a nucleotide diversity of π = 0.00271. Some mutated positions of the α-Phs gene were in positive or negative linkage disequilibrium and 6 paired informative sites had a history of recombination. The computed Tajima’s D was significantly less than 0 indicating presence of purifying selection. The association analysis revealed that three non-synonymous indels on the coding region were significantly associated with protein content. The findings from this study indicate that the polymorphisms detected in α- Phs gene can be used for discrimination of the genetic relationships among common bean germplasm. The genetic variants associated with seed protein content in the common bean accessions, could be explored in molecular breeding as well as potential genetic markers in the improvement of protein content in common beans.

CHAPTER ONE: INTRODUCTION

1.1 Background to the study

Protein dietary malnutrition is the most dangerous form of malnutrition affecting many people worldwide, mostly children, due to insufficient protein in the diet (Schönfeldt & Hall, 2012). Plant foods that are protein-rich such as common beans have the potential to provide solutions for malnutrition more so in low income countries in the world where there is low intake of animal protein. Common bean (*Phaseolus vulgaris* L.) is an important leguminous crop used by humans for direct nutritional purpose and serves as an important source of dietary protein to more than one billion people globally (Lioi *et al*., 2019). It serves as a major source of vegetable protein (Bitocchi *et al*., 2011). In addition, its seeds contain significant amounts of other valuable nutrients including vitamins, energy, fiber, minerals and low content of fat (Celmeli & Sari, 2018). Common beans are also known to have substantial health promoting properties such as reducing the risk of coronary heart disease, renal and diabetes type- II diseases, protecting against many cancer types as well as controlling overweight and obesity (Mullins & Arjmandi, 2021).

The main protein components of common beans include globulins (54 – 79 %) and albumins (12 – 30 %). However, common bean is known to have a poor balance of essential amino acids relative to human nutritional requirement. The nutritional quality of common beans is therefore considered low as a protein source due to the presence of sub-optimal amounts of Sulphur amino acids i.e. methionine, cysteine and S-homocysteine. Furthermore, common bean proteins are poorly digested even after cooking (low protein digestibility) because of the components of its protein fractions, presence of anti-nutritional compounds such as phytic acid, proteinase inhibitors as well as the presence of oligosaccharides, that are water soluble, hence causing accumulation of gas in the alimentary canal, resulting in eructation (Montoya *et al*., 2013).

Phaseolin (encoded by α-Phs gene) is the most abundant protein reserve in beans, constituting 40 – 50 % of the total protein content and hence a major source of amino acids in the common bean seeds (Montoya *et al*., 2010). Phaseolin protein is an important genetic marker specifically in the understanding of genetic- based diversity/variability of different landraces of common bean (Fuente *et al*., 2012). Previous studies have utilized the molecular diversity of the α-Phs locus to distinguish genetic variation/evolutionary relationships across the species and it has not been considered as a parameter to improve the protein nutritional value in the common beans using genetic approaches (Qureshi *et al*., 2019). More than 40 genetic variants/forms of phaseolin differing in amino acid composition have been reported both in the wild and domesticated beans based on their composition of polypeptides and the most common ones include: Tendergreen (T), Sanilac (S) and Contender (C) (Gepts *et al*., 1986). The genetic variability of phaseolin gene may hence be used as an opportunity to enhance protein nutritional value in common bean seeds (Yildiz *et al*., 2017). Modern genetic approaches for the improvement of nutritional quality of proteins in common bean seeds, require the knowledge of nucleotide diversity to understand the general composition, structure and organization of its genetic based diversity in the different landraces.

The diversity of phaseolin glycoprotein has been studied in many genotypes of common beans using Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) (Fuente *et al*., 2012). Gel electrophoretic band patterns of phaseolin have revealed the diversity/variation of common beans as being organized into distinct eco geographic pools of genes mainly Mesoamerican and Andean (Bitocchi *et al*., 2011). In the present study, nucleotide polymorphisms of α-Phs gene was analysed among 11 common bean accessions of agronomic importance in Kenya and association of polymorphisms with seed protein content evaluated.

1.2 Problem statement

Common bean (*Phaseolus vulgaris* L.) is a main food legume used directly by humans as a protein source mainly because it contains high amount of protein. However, the protein quality in the different common bean germplasm is variable. For example, some common bean accessions have sub-optimal amounts of sulphur-containing essential amino acids mainly cysteine, methionine and S-homocysteine. These amino acids are important to humans ; methionine is used in initiating the synthesis of amino acids in proteins while cysteine, plays a significant role in protein-folding pathways and structure (Brosnan & Brosnan, 2006) and is a component of antioxidants such as glutathione. Phaseolin, its main seed storage protein contains only 0.5 - 0.80 % methionine content, which is below the human nutritional need (Aylor *et al*., 2008; Celmeli & Sari, 2018). The suggested nutritional requirements for methionine-cysteine in the human diet are between 2.5 and 2.6 % which is equal to between 26 and 25 mg methionine-cysteine gram per protein (McLarney *et al*., 1996; Millward, 2015).Thus depending on common bean diet entirely, especially from common bean germplasm with low protein can result in malnutrition. Previous studies have utilized variability of the α-Phs locus to distinguish genetic variation/evolutionary relationships in common bean germplasm, while the improvement on the quality of protein, has been neglected (Qureshi *et al.*, 2019). Polymorphisms present in the α-Phs locus gene in common beans have been scarcely studied, even though its variants can be used as potential genetic markers for improving the nutritional quality in common bean germplasm. Phaseolin as a dietary protein is also faced with the problem of being poorly digested in its original form; due to its inability to be degraded by gastrointestinal tract enzymes of various monogastric animals. Common beans protein quality is further compromised by anti-nutritional compounds such as phytic acid, proteinase inhibitors and oligosaccharides that are water soluble and can cause accumulation of gas in the alimentary canal, resulting in eructation (Montoya *et al*., 2010).

1.3 Justification of the study

Common beans are considered to be the main legume /grain food in many low income and developing countries especially in sub-Saharan Africa and Latin America (Bitocchi *et al.*, 2011). In order to improve the quality of seed protein in beans for nutritional enhancement; an understanding of the nucleotide diversity of α-Phs locus is essential. Seed protein content and quality varies in different common bean germplasm, with different landraces exhibiting varied protein fractions and amino acid composition (Bernal *et al*., 2014). Phaseolin as a main protein reserve in the beans is known to be genetically diverse. The variability /genetic-based diversity of phaseolin can be considered as an important strategy to target enhancement of protein in common bean using marker-assisted breeding or genetic engineering approaches (Montoya *et al.*, 2010). Modern genetic improvement methods of protein quantity and quality require the knowledge of nucleotide diversity of genes encoding for protein in different common bean germplasm.

1.4 Objectives

1.4.1 General objective

To determine the nucleotide diversity of common bean phaseolin (α-Phs) gene and association with seed protein content in selected germplasm.

1.4.2 Specific objectives

The specific objectives of the study were:

(i) To determine protein content in mature seeds of selected common bean accessions of agronomic importance in Kenya

(ii) To analyse the nucleotide polymorphisms of phaseolin (α-Phs) locus in selected common bean accessions of agronomic importance in Kenya.

(iii) To evaluate the association between the nucleotide polymorphisms of common bean α-Phs gene and mature seed protein content.

1.5 Null hypotheses

(i) There is no variability in seed protein content of selected common bean accessions of agronomic importance in Kenya

(ii) There is no nucleotide polymorphism of phaseolin (α-Phs) locus in common bean accessions of agronomic importance in Kenya.

(iii) There is no relationship between nucleotide polymorphisms of common bean α-Phs gene and seed protein content.

CHAPTER TWO: LITERATURE REVIEW

2.1 Worldwide domestication, distribution and production of common beans

The common bean (*Phaseolus vulgaris* L.) is an important legume grain used for dietary nutrition in many parts of the world. It emanated from both South and Central America, from where its cultivation was started in the early 6000 BC in Peru and 5000 BC in Mexico. It's most characterized and important species have been identified distinctly in two gene pools, the Andean and Mesoamerican (Figure 2.1 **A**). The two gene pools differ in many aspects i.e. genetic diversity and characteristics that are phenotypically expressed as well as different dynamics on evolution (Mamidi *et al*., 2013). The cultivation of common bean has become widespread and it is considered as a main food crop in various parts of the world. Main producing areas include the sub-Saharan Africa, Latin America, Middle East, China, Europe, Australia, United States, and Canada (Mohammed, 2013). Latin America remains the leading grower of common bean as it is considered a traditional and an important part of the diet in the region. In sub-Saharan Africa, the common beans are mainly cultivated for livelihood purposes, with regions surrounded by the lakes having the greatest mean consumption per individual. Common beans are considered as a major source of protein nutrient in the diet more so in the African countries of Tanzania, Kenya, Malawi, Uganda, and Zambia (Katungi *et al*., 2009). Between the year 2000 and 2013, Africa showed the highest production growth across all the other continents with close to up to 75%, in comparison to Asia (33%) and American continents (6%) (Figure 2.1 B). Production is high generally in tropical countries and worldwide high production continues to be experienced in Latin America, Brazil and Mexico (Bitocchi *et al*., 2011). In Africa, common bean crop production lies at approximately 25 % of the total world's production, with 70 % of production in Eastern Africa (Katungi *et al*., 2009a). Lead producers of common bean in the sub-Saharan African region include Tanzania, Uganda and Rwanda, Cameroon, Kenya and Ethiopia(Beebe *et al*., 2012).

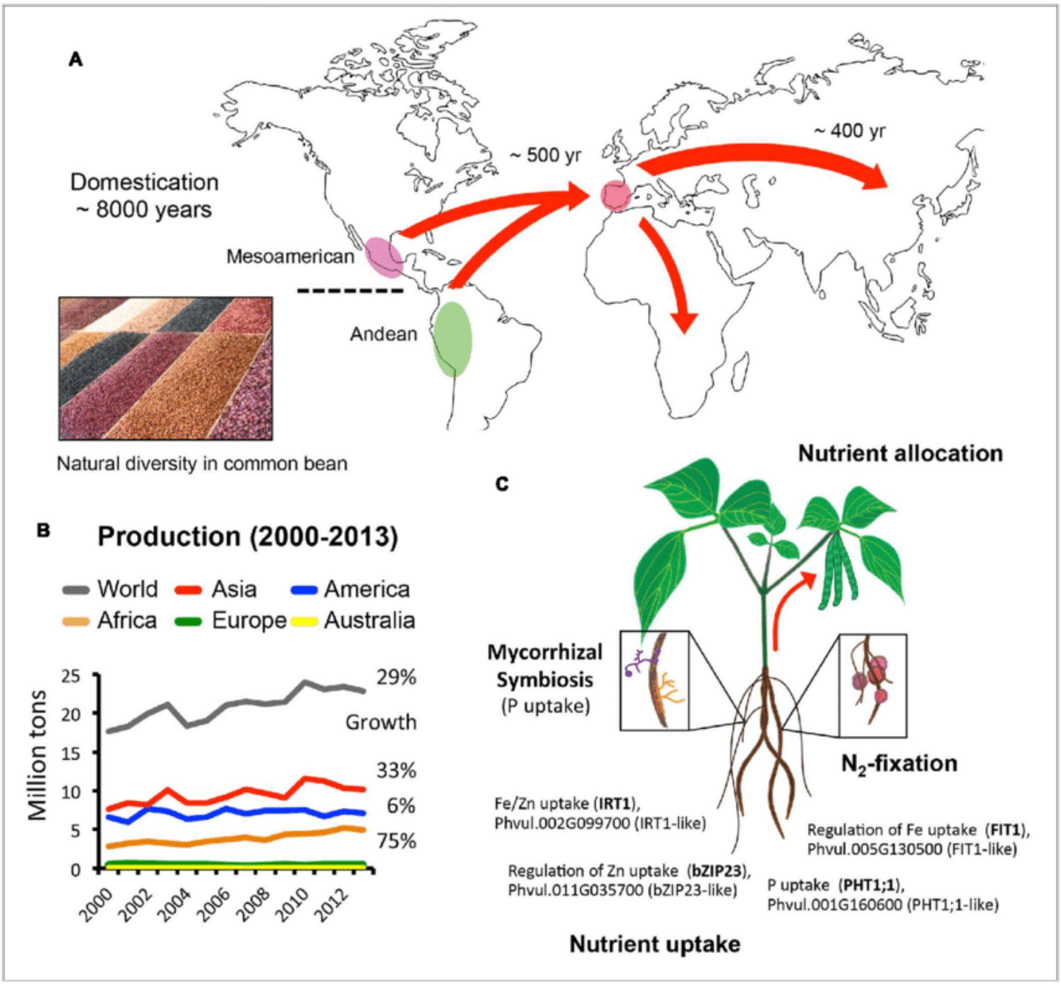


Figure 2.1: Worldwide cultivation/domestication, production and distribution of common beans. (A) Geographic distribution of common bean following domestication; (B) Worldwide production of the common bean in before omics era (2000-2013) and (C) Common bean relationship with symbiotic endomycorrhiza and Rhizobia (Castro-guerrero *et al.*, 2016).

According to report by FAO in 2020, common bean production was 28.9 million tons in 2019 and global harvested area was 33.1 million ha. Production in Asian continent has risen to 50% of the global production of common bean (Figure 2.2), and Myanmar, India, Brazil, China, America were the top five dry bean producing countries in the world in 2000–2019.

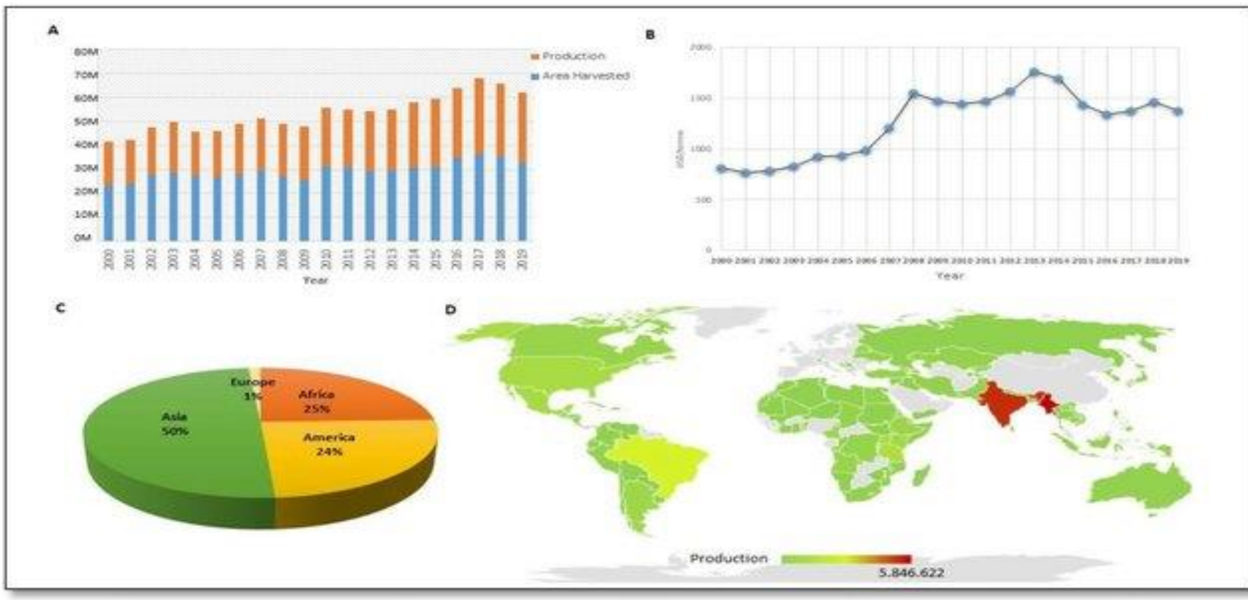


Figure 2.2: Common bean production in the world. (A) Evolution of common bean seed production and area under cultivation from 2000 to 2019. (B) Evolution of sale prices of common bean seed from 2000 to 2019. (C) Production share of common bean seed by continent in 2019. (D) Map of production quantities of common bean seed by country in 2019 (Source: Food and Agriculture Organization Statistical Databases ((FAOSTAT, 2020).

2.2 Production of common bean in Kenya

In Kenya, the bean crop is grown in almost all regions as a staple food for subsistence, where its growth is associated with a large number of small-holdings of not more than one ha per household (Barkutwo *et al*., 2020). The major common bean producing areas in Kenya includes counties in Eastern, Nyanza, Central, Western and Rift valley regions. Kenya is a leading producer of common bean in East Africa region with 300,000-500,000 hectares of land under coverage of the crop, producing about 40,000–150,000 metric tons per year (Gichangi *et al*., 2019). The crop is mainly for subsistence use with approximately 40% of production for commercial purposes (Margaret *et al*., 2014). Some of the most common bean varieties in Kenya have different expected yields and they include: Chelalang (840-980 kg/hectare), Red Haricot (GLP585) "Wairimu" (870-1110 kg/hectare), Tasha (625-870 kg/hactare), Ciankui (625-825 kg/hectare), KK15 (756-900 kg/hectare), Canadian (613-672 kg/hectare) and Rosecoco (560-935 kg/hectare).

Yields obtained from these varieties are actually lower than the standard yields. Common bean yields on farmers' fields is on the decline and range from 0.14 to 0.77 to 110 (Ayuke *et al*., 2018). Common bean production Kenya is constrained by low yielding varieties and inadequate technical knowledge among farmers on crop management (Nay *et al*., 2019). Bean varieties developed and released in the recent past with better yields and adaptation have not been widely tested and disseminated in the country. Availability of adaptable varieties with high yields can enhance bean production by farmers and total earnings in the agricultural sector on which country's population is dependent on for their livelihood (Barkutwo *et al*., 2020).

2.3 Nutritional and economic significance of common beans

The common bean is an important grain legume in the human diet at global level, as it provides protein, complex carbohydrates and valuable micronutrients for more than 300 million people in the tropics (Brigide *et al.*, 2014). *Phaseolus* species are a significant source of protein, carbohydrates, vitamins, fiber, minerals and low content of fat and sodium, more so for resource limited populations throughout the world (Celmeli & Sari, 2018). They are also rich in unsaturated fatty acids, such as linoleic and oleic acids. The common beans are thus considered to be the main legume /grain food in many low incomes and developing countries such as Latin America and sub-Saharan Africa (Bitocchi *et al.*, 2011). Common beans are also known to have substantial health promoting properties such as reducing the risk of acquiring diseases like coronary heart disease, renal and diabetes type-II diseases, protecting against many cancer types as well as controlling overweight and obesity (Tc *et al.*, 2018). Moreover, they are rich in phytochemicals, flavonoids and antioxidants. Common beans are able to counter constipation, thus reducing risks of cancers that affect the alimentary canal, for instance colon cancer (Campos-Vega *et al*., 2013).

Common beans contribute up to 65 % of the total protein consumption and 32 % of the total energy (Katungi *et al*., 2009). Besides the consumption of their dry mature seeds, they are also cultivated for their green leaves and pods, which are cooked as vegetables as well as their residues used as fodders for animals. In Kenya, common bean is an important supplement to the country's carbohydrate rich diet such as maize, cassava, sweet potato and wheat (Kimani & Warsame, 2019). Besides its nutritional significance and health promoting capability, common bean plants are grown in many areas around the world because of their vast known economic importance. The crop also generates revenue in Kenya through export and income to small scale rural farmers who sell the crop to the urban residents (Katungi *et al*., 2009). In addition, common beans have the ability to create symbiotic relations with endomycorrhiza (Arbuscular mycorrhizal fungi) and bacteria with nitrogen fixing potential (Rhizobia), (Figure 2.1C) thus make uptake of essential nutritional elements such as phosphorous and nitrogen by plants easier (Venkateshwaran *et al*., 2013). The associations that are symbiotic further can reduce the use of organic fertilizers dramatically and hence minimize the emission of Nitrous oxide a major greenhouse gas as their by-product. Common beans are therefore enriching to the soil and can consequently help in reducing the need to use organic commercial nitrogen fertilizers, which not only deteriorate the environment but are also expensive for the small holder farmer. Common beans also play an essential role in preventing soil erosion; as they are substantially good cover crops (Anunda *et al*., 2019).

2.1 Major seed storage protein constituents in common beans

Seed storage proteins in common bean seeds have been investigated broadly because of their economic significance as important protein sources for animals, and also their importance as biochemical model systems essential for gene isolation, molecular characterization and expression (Ersland *et al*., 1983). Seed storage proteins supply nutrients for the germinating seeds and in consequence make up a big proportion of the entire protein content in the mature seed. The storage proteins are categorized based on the specific time when they accumulate in seed during development, level of nitrogen contents, and whether they occupy protein bodies in the cotyledon cells (Shewry & Halford, 2002). In the common beans, the major seed storage proteins are globulins accounting for up to 50 % of the entire protein content (Zheng *et al*., 1995). Globulins are separated into two fractions, legumin and vicilin, on the basis of their solubility.

The other seed protein storage in the common beans are albumins which constitute 30% of entire nitrogenous content. In addition, Common beans are known to contain high amounts of glutenin (20 – 30 %). Phaseolin which is the main protein reserve in beans constitutes the major globulin fraction i.e. 7S fraction, representing 40 – 50 % of the entire amount of protein in the seed (García-Cordero *et al*., 2021). Phaseolin has undergone post translation modifications thus is considered as a glycoprotein. Globulin (11S fraction) on the other hand represents only 10 % of the entire amount of protein. The rest of the nitrogen content of the common beans include prolamin (2 - 4 %) and free-amino acids (5-9 %) (Montoya *et al*., 2010). The bean seed amino acid composition exists in different fractions inclusive of those that are available in insignificant levels such as cysteine, S-homocysteine and methionine.

2.5 Phaseolin protein

Phaseolin is a group of polypeptides that comprise the major storage glycoproteins in the seeds of common bean (Fuente *et al*., 2012). It is a constituent glycoprotein of the 7S vicilin class. It contributes greatly to the quality of protein for nutritional purpose in the seeds of common bean as it constitutes 40 - 50 % fraction of the entire protein content (Montoya *et al*., 2010). Phaseolin as a dietary protein is faced with the problem of being poorly digested in its original form due to its inability to be degraded by enzymes of the alimentary canal of various animals with a single- chambered stomach (Fuente *et al*., 2012). Phaseolin protein is also lacking the necessary amounts of essential amino acids known to contain Sulphur i.e. methionine and cysteine hence effort to enhance its essential amino acid composition promises to improve quality of storage protein in the common bean plant (Montoya *et al.*, 2010). Phaseolin protein has been reported to play other important roles in the food industry because of its biochemical & physical attributes; for instance, it has been found useful in them formulation of food, beverage production industry, as biopolymer films components and preservatives in storage of bread because of its ability to inhibit activity by fungal species (Yin *et al.*, 2010).

2.6 The genetic diversity and variation of phaseolin gene

Phaseolin contains neutral sugars which results in a high variation within its molecular weight subunits. Its genetic diversity is commonly used as a pointer to the evolution of the domestication of bean crop in Argentina, Bolivia, Chile, Colombia, Ecuador, Peru, Venezuela and America (Gepts, 1988). Phaseolin belong to a group of proteins of the same family, although they are diverse to some degree in their polypeptide arrangement, because of their sequence level differences, as well as post-translational modifications. Phaseolin protein is assumed to be represented by a family of genes comprising approximately 6-10 co-dominant genes that are interrupted in each haploid genome embedded on chromosome 7 in a single cluster (Bernal *et al*., 2014). Both Phaseolin complementary deoxyribonucleic acid (cDNA) and genomic clones have indicated that the gene family of phaseolin can be distinctly grouped into two different types of genes: - α and β (Bernal *et al*., 2014). Each of the gene types i.e. α and β is glycosylated during co -translational protein modification in the central cavity of endoplasmic reticulum via the shift oligosaccharides that contain complex high -mannose; or N-acetylglucosamines to the amide functional group that is specific for residues of Asparagine, resulting into oligosaccharide side chains that are N-linked.

The genetic variability of phaseolin can be used as an opportunity to determine protein compositions of the bean seeds both quantitatively and qualitatively for nutritional enhancement (Yildiz *et al*., 2017). Further, the diversity of phaseolin can provide archaeological, historical and botanic information because of its environmental stability, biochemical complexity features. Bean domestication studies have indicated that two main phaseoline types: Sanilac (S) and T (Tendergreen) are commonly found in the cultivated beans. S phaseolin type is more common in the Central America cultivars, Mexico and North of Colombia cultivars (Chacón *et al*., 2005). The T phaseolin on the other hand is common in the Andean cultivars, mainly: Bolivia, Chile, Argentina and South of Peru (Rodiño *et al*., 2006). Other forms of phaseolin have also been reported in the wild cultivars within each center of domestication.

The electrophoretic profile when considered indicates that phaseolin proteins are generally composed of polypeptides ranging between 2- 6; that differ distinctly in their molecular weight i.e. 40-54 kDa and in their isoelectric point (Montoya *et al*., 2010). Phaseolin hence belongs to a family of glycoproteins that differ in polypeptide composition, molecular weight and isoelectric point, the differences which reflect the divergence in their genomic DNA sequences encoding for different families of polypeptides (Fuente *et al*., 2012). Different precursor subunit profiles for the different phaseolin types: Sanilac (S), Contender (C) and Tender-green (T), have been reported based on procedures such as mass spectrometry (Yin *et al*., 2010). The subunits have nucleotide sequence differences of the α and β gene types for each of phaseolin type (Kami & Gepts, 1994). The differences in molecular weight can also be linked to modifications in both pre- and post-translational processes, resulting in the polypeptides differentiation in phaseolin, small insertions and deletions, duplications as well as nucleotide substitutions. Further, composition of carbohydrate and the phosphate binding sites of phaseolin also do contribute to diversity in molecular weight observed on the same precursor of protein (Montoya *et al*., 2013).

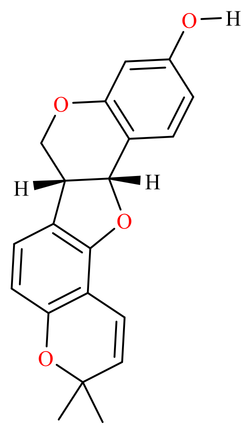


Figure 2.3: Two-Dimensional structure of phaseolin (C20H18O4) (Edy Susanto, 2019).

2.7 Genome architecture of the phaseolin gene

Phaseolin polypeptides are encoded as a small, homologous, group of genes that are closely linked and may evolve by continuous replication and diversification from an ancestral gene (Talbot *et al*., 1984). The gene comprises of both exon and intron regions and is 4502 base pairs in full length. The two distinct gene families encoding phaseolin protein includes the alpha-and beta-type (Phs) phaseolin genes. Both alpha-and beta-types of the genes contrast in their coding regions due to the presence or absence of two different direct repeat sequences. The alpha - phaseolin type contain 2 small (15 and 27 base pairs) direct repeats, while the Beta -phaseolin type of genes do not contain any of these repeats (Slightom *et al*. 1985). The genes encoding the alpha-type phaseolin polypeptides therefore contain direct repeat sequences which result in approximately 14 additional amino acids in their proteins, thus preferred for the improvement of protein quality (Kami & Gepts, 1994). Aside from the differences in the direct repeat of sequences, the phaseolin genes encoding the alpha-and beta- type forms show a high level of similarity/ homology (up to 98%) which is a confirmation that these genes have evolved from a common ancestral origin. The heterogeneity found in the phaseolin polypeptides to a great extent, appear to be because of post-translational processing (Talbot *et al.*, 1984). The alpha-form of phaseolin gene therefore may contain a few amino acid replacement substitutions which may result in codon changes, and this is lacking in the beta-type genes.

2.8 Genetic improvement of nutritional content of phaseolin in common bean seed

Phaseolin, a seed protein globulin in common beans is known to exhibit abundant diversity among its various subtypes (Gepts & Bliss, 1986). Possibly, this diversity is increased by natural outcrosses between wild and cultivated beans. The structural and functional features of phaseolin, make it a useful marker for improvement of nutritional value in common beans (Lioi & Piergiovanni, 2013). Phaseolin in many different common bean germplasms has been shown to be very diverse and this may be significant in understanding the various constituents of phaseolin and the relevance of phaseolin in genetic improvement of the protein content in accessions with low protein content (Emani & Hall, 2014). Many bean cultivars have a high variation in seed protein with a varied range of protein concentrations, thus a considerable opportunity for protein quality improvement (Bliss & Brown, 1983). Genetic improvement is one of the most pivotal strategies to obtain bean accessions that have improved agronomic traits, such as yield and the best nutritional features to satisfying consumer needs (Bailey-Serres *et al.*, 2019). Exploiting the diversity of phaseolin gene in common beans, thus represent a very efficient strategy to increase its nutritive value in common bean germplasm. The nutritive value of phaseolin is often limited by increasingly resistance to *in vitro* hydrolysis and *in vivo* digestion as a result of glycosylation, which makes its chemical structure rigid and compact (Oliveira *et al*., 2017). The digestibility can however increase considerably after thermal treatment to between 80 – 90%.

2.9 Genetic markers and their use in crop diversity for genetic improvement of crops

Genetic markers are a valuable advancement in plant breeding. They are known part of a DNA sequence with a known chromosome location and control a specific trait (Nadeem *et al*., 2018). DNA molecular markers such as restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), simple sequence repeats (SSRs), and single-nucleotide polymorphisms, can reveal polymorphisms in DNA sequences; across different species arising due to mutations on the DNA (Collard & Mackill, 2008). DNA molecular markers are therefore polymorphic and can help distinguish between and within individual species, depending on whether they are co-dominant or dominant as they can differentiate whether they are heterozygotes and homozygotes.

The molecular markers are broadly categorized based on detection method i.e., hybridization, DNA sequence-based and PCR (Govindaraj *et al*., 2015). The markers that are PCR- based such as AFLPs, and SSRs usually only amplify specific regions of the DNA, which are detected and separated by gel electrophoresis. The most recent advancements have resulted in the generation of markers such as SNPs, InDel polymorphic molecular markers and Diversity arrays technology, which have high-throughput performance and can be detected by automation. They are on the actual DNA sequence and can be obtained by editing sequences generated by various sequencing methods, such as next generation sequencing methods; as well as automated systems to detect existing polymorphisms

(Govindaraj *et al*., 2015).

Molecular markers can be used to establish genetic diversity of a particular species and many of such assessments have already been reported in common bean accessions, for instance (Kumar *et al*., 2008) used AFLPs markers. Genetic diversity gives detailed information on the evolution and genomics of various species, hence can help elucidate phylogenetic, evolutionary relationships and population structure within and between species. Molecular plant breeding technology make use of molecular markers in the study of heterosis, during crossing so as to determine the performance of the developed progeny and identify the parental crosses that can result in superior genotypes (Mogesse *et al*., 2020). Selection and breeding based on molecular markers is a more superior breeding technique compared to the conventional breeding. In Molecular marker assisted Selection for instance, the entire process of selection targets specifically the trait of interest; which therefore ensures accuracy and specificity, and time saving (Collard & Mackill, 2008). SNPs have been tested and validated across well-chosen common bean (*Phaseolus vulgaris* L.) germplasm for assessment diversity, and parents of mapping populations (Cortés *et al*., 2011). The use of molecular markers will enhance the understanding of the genetic factors that are responsible for agronomic traits that would assist in the selection of superior genotypes.

CHAPTER THREE: MATERIALS AND METHODS

3.1 Plant materials and establishment of plants in the glasshouse

Eleven common bean accessions commonly grown in Kenya in different geographical regions were used in the study (Table 3.1). The common bean accessions were selected based on their agronomic importance including high yield and resistance to bacterial wilt and anthracnose diseases. Common bean seeds (3 mature seeds, Figure 3.1) were planted in plastic pots (4.5" Diameter, 4" Height) containing sterile organic soil under glasshouse (400 sq ft (20' x 20') conditions to maturity. The experimental soil used was obtained from the forest at the Faculty of Science and Technology, University of Nairobi. The soil was sterilized by autoclaving and left to cool overnight. The plastic pots used were clearly labelled using the sample identities of the various common bean accessions and filled with the sterile soil. The seeds of each of the accessions were then established in the specific pots in triplicates, watered and left to germinate. To minimize pest and bug infestation, the greenhouse surrounding was cleaned often and lighting installed. After germination, the seedlings were watered regularly after every three days till maturity. Young leaves were collected from four-weeks-old bean plants and used for genomic DNA extraction. The common bean plants were left to grow to maturity and the harvested dry seeds were used for determination of protein content. The bean seeds of the various accessions were harvested at different times as they all matured at different times, between 60 to 90 days of planting.

Table 3.1: Accession name seed size and color of common bean landraces used in the study.

Sample ID	Accession name	Seed size	Seed color
1	Ekebure	Small	White
2	KATB1	Medium	Yellow
3	Kidney bean	Medium	Dark brown
4	Mbeere 1	Large	Brown-red patches
5	Mbeere 2	Medium	Brown-red patches
6	Mbeere 3	Medium	Brown with cream stripes
7	Mbeere 4	Large	Purple with white spots
8	19	Small	Wine red
9	23	Small	Red
10	Unknown	small	Light green
11	VAX4	small	Black

Figure 3.1: Seed coat colour of the various accessions used in the study.

3.2 Protein extraction from dry mature seeds of common bean accessions

To extract the proteins from dry mature bean seeds, 1 gram of seeds of each common bean accession used in the study were first soaked overnight and ground using mortar and pestle. The resulting ground paste was dissolved in 10 mL distilled water and allowed to settle, after which was centrifuged at 14000 rpm (Eppendorf Centrifuge Rotor-Marshall Scientific, USA). The resulting supernatant (1 mL) was used to determine the seed protein content using Lowry protein procedure.

3.3 Determination of seed protein content in common bean accessions

Protein content in seeds of the 11 dry common bean accession was determined using Lowry protein assay method (Geiger & Bessman, 1972). Lowry's reagent A consisted of 1 % copper sulfate (CuSO4·5H2O). Lowry's reagent B consisted of 2 % sodium potassium tartrate in a total volume of 100 mL. Lowry's reagent C consisted of 2 % sodium carbonate (Na_2CO_3) dissolved in 0.1 M sodium hydroxide (NaOH). Lowry's reagent D (freshly prepared) comprised of a combination of 0.5 mL of copper sulfate (reagent A), 0.5 mL of sodium potassium tartrate and 49 mL of the reagent C. Lowry's reagent E comprised of Folin-phenol (The *Folin*–Ciocâlteu reagent) mixed in the ratio of 1: 3 with sterile distilled water. A total of 1 mL of the extracted protein of each sample was added to reagent D (4 mL) and incubated for 10 minutes. After incubation, 1 mL of reagent E was added to each of the tubes and incubated for 30 minutes with continuous stirring in a dark room at room temperature. The amounts of the protein in the samples were determined using a spectrophotometer (Thermo Fisher Scientific, USA). The Lowry protein assay method involved the use of 0.2 mL Bovine Serum Albumin (BSA) as the protein standard in 5 different test tubes, which were serially diluted to 1 mL using distilled water. The test tube with 1 mL distilled water served as blank during the spectrophotometry reading. The absorbance was measured at 650 nm and the values were plotted to develop the standard graph/ curve.

3.4 PCR amplification using phaseolin (α-Phs) gene-specific primers

3.4.1 Extraction of genomic DNA

Genomic DNA was isolated from fresh young leaves of four-week-old common bean plants grown under greenhouse conditions. DNA extraction reagents were first prepared (Appendix 1). Modified cetyltrimethylammonium bromide (CTAB) extraction protocol was used for DNA extraction as described by Bijay *et al*. (2020). Briefly, 1 gram of fresh leaves from each of different common bean accessions were crushed using 1 mL of CTAB extraction buffer to a homogenous paste. A total of 530 µL of the crushed mixture was then transferred into a sterile well labelled micro centrifuge tube and 10 µL of proteinase-K (20 mg/mL) added to the mixture. A total of 60 µL of 10 % SDS was then added to the mixture to give a final concentration of 1 % SDS. The tube was gently mixed and incubated in a water bath at 65 °C for one hour, then left to cool to room temperature. A total of 100 µL of 5 M NaCl was added to the tube and mixed gently to get a homogenous solution, after which 80 µL of CTAB/ NaCl solution (2 % CTAB/0.7 M NaCl) was added to the mixture and mixed gently to obtain a homogenous solution. The mixture was incubated in a water bath at 65 °C for 15 minutes to ensure complete dissolution of CTAB in order to precipitate polysaccharides. The mixture was allowed to cool to room temperature (25°C). Equal volumes (700 µL) of chloroform: isoamyl alcohol (24:1) was added to the samples and centrifuged at 12000 rpm (Eppendorf Centrifuge Rotor, Marshall Scientific, USA). Following complete dissolution of the DNA, 5 µL of RNase A was added to the solution and incubated at 37 °C for 1 hour. Gel electrophoresis was carried out on a 0.8 % agarose gel mixed with 0.5 µg/mL Ethidium bromide and the genomic DNA was visualized using UV illuminator (ThermoFisher

Scientific, USA).

3.4.2 Quantification and quality assessment of DNA

In order to quantify and ascertain quality of the DNA samples, 0.8 % agarose gel stained with ethidium bromide was prepared. The gel was prepared by weighing 0.8 % agarose powder (0.8% w/v); and dissolving in 80 mL Tris/borate/EDTA (TBE buffer), followed by heating gently on a hot plate for 3 minutes, to enhance dissolution of the agarose powder and allowed to cool to approximately 45 °C. A total of 4 µL of ethidium bromide stain (0.5 µg/mL) was added to the gel as ethidium bromide binds to the DNA and makes visualization of the DNA possible under ultraviolet (UV) light. The gel was then poured on a casting gel tray already fitted with well combs and left to solidify for 30 minutes at room temperature. Once solidified, the agarose gel was submerged into an electrophoretic tank containing 1× TBE buffer and combs gently removed. The DNA samples were allowed to thaw for 1 hour at 4 ° C. A total of 5 µL of the DNA was mixed with bromophenol blue dye on a flat sterile surface and loaded on the wells carefully. The lambda DNA molecular weight marker (Thermo-Fisher Scientific, USA) was pipetted into the first well to help estimate the molecular weight of the DNA samples. The gel electrophoresis was run at 80 Voltage (V) for 45 minutes, then visualized under a gel imaging Doc (Thermo-Fisher Scientific, USA) to view the bands.

To determine the extracted DNA yield, DNA concentration was estimated by measuring the absorbance at 260nm, adjusting the A_{260} measurement for turbidity and multiplying by the dilution factor, using the relationship that an A_{260} of 1.0 = 50µg/ml pure dsDNA. Concentration of the DNA (µg/ml) was determined by the relationship (A_{260} reading – A_{320} reading) × dilution factor × 50µg/ml. Total yield was therefore obtained by multiplying the DNA concentration by the final total purified sample volume. DNA yield (µg) = DNA concentration × total sample volume (mL).

3.4.3 PCR amplification of the extracted DNA

The PCR amplification of the DNA samples was carried out using 10 primer pairs (Table 3.2) targeting different fragments of phaseolin gene in order to amplify the full-length gene. The primer pairs were synthesized by Inqaba biotech limited in South Africa. A total of 10 primer pairs were used to generate separate overlapping fragments of the full length phaseolin gene. To optimize the PCR conditions for the various pairs of primers, standard criteria of gradient PCR was adopted for annealing Temperature (T_a) from 55 °C to 64 °C depending on the primer pair used. The PCR reaction mixture (50 µL total volume) comprised of 25 µL of 2× Ampliqon mix (Stenhuggervj 22, Denmark), 2 µL of each primer pair (labelled Phs1-10) i.e. both reverse and forward primers, 19 µL sterile Nuclease free water and 2 µL of (30 ng) DNA template. Reactions were conducted using a thermocycler (Invitrogen™-Thermo-fisher) with the following cycling conditions: 94 °C for 5 min for initial denaturation, followed by 35 cycles of 94 °C for 30 sec, 55 °C to 64 °C for 1 minute (depending on the primer pair used), 72 °C for 1 minute; and final extension at 72 °C for 5 minutes, and held at 4 °C. The resulting PCR products of each primer pair were electrophoresed on 1 % agarose gel stained with 4 µL ethidium bromide (0.5 µg/mL). The gel was run at 80 Voltage (V) for 1.5 hours, then visualized under a gel imaging Doc (Thermo-Fisher Scientific, USA). The band sizes were compared in reference to the DNA molecular weight marker used.

Table 3.2: Forward and reverse primers targeting fragments 1-4 of Phaseolin gene.

Primer code	Primer sequence (5'-3')	T_a (°C)	Size (bp)
Phs1	Forward- CCCCAACCAAGATGAACAC	64	650
	Reverse- TTGTCATGTGTTGACCCTTG		
Phs2	Forward- CACCCAACCAAAATAGCTTC	58	510
	Reverse- CCTTTTTCCTGTGTTCTTACC		
Phs3	Forward- GGAACAAAAACGGAACGAAC	58	500
	Reverse- CAAAGTGTCCAACACCTCG		
Phs4	Forward- GTGAAAACCATCACCGTCC	59	600

Note: T_a is the annealing temperature of the primers. The primers used in this study were obtained from a study by Diniz *et al*. (2014).

Table 3.3: Forward and reverse primers targeting fragments 5-10 of Phaseolin gene.

	Reverse- AACGACTACAGAGTGCCATG		
Phs5	Forward- ATGCCATTCAAAACCCAC	61	550
	Reverse- GGTGGCGGAAGAGAATG		
Phs6	Forward- ACCATCTCAACCCACACAC	61	550
	Reverse- TGGACCTGAACTCCACAAG		
Phs7	Forward- GAGGTTCGACCAACAATCC	61	500
	Reverse- GTTGGGCTTCTGTGCTAG		
Phs8	Forward- AACAACCCTCAGATTCATG Reverse-	52	650
	TTTTCTAACCAAGACGACAG Forward-		
Phs9	GACCGATAACTCCTTGAATG	52	600
	Reverse- AGAGAACGTAAGCCCCAAC		
Phs10	Forward- AGAGAACGTAAGCCCCAAC	50	600
	Reverse- GAAGGTAAGACGGACAATG		

Note: T_a is the annealing temperature of the primers. The primers used in this study were obtained from a study by Diniz *et al*. (2014).

3.5 Purification of PCR products

The PCR products were purified using the GeneJet Gel Extraction Kit (Thermo-Scientific, USA) which comprised of binding buffer solution, wash buffer solution, and elution buffer (10Mm Tris-HCl, pH 8.5). The purification procedure was necessary to ensure unused primers and dNTPs during PCR were all degraded and/or removed for subsequent seamless sequencing procedure. The purification procedure was carried out using the GeneJet Gel Extraction Kit (Thermo- Scientific, USA) according to the manufacturer's instructions. The procedure involved adding equal volumes of the binding buffer to the PCR product; i.e. 30 µL equal volumes of the binding buffer and amplicons. The contents in the tube was mixed thoroughly by inversion until a yellow colour was achieved, an indication that the optimal pH for DNA binding was reached. Equal volume (60 µL) of isopropanol was added to the mixture and mixed evenly to facilitate the precipitation of the DNA molecules, allowed to settle for 5 minutes after which transferred to the GeneJET purification column. The tube with the purification column was centrifuged, the flow- through discarded and the column placed back in the same micro-centrifuge collection tube. A total of 700 µL of wash buffer (diluted with ethanol) was added to the Gene-JET purification column and centrifuged for 1 minute at 13000 rpm (Eppendorf Centrifuge Rotor-Marshall Scientific, USA). Flow-through was discarded again, and the Gene-JET purification column returned to the same collection tube. The empty Gene-JET purification column was repeatedly centrifuged at 13000 rpm (Eppendorf Centrifuge Rotor-Marshall Scientific, USA) for 1 minute to completely remove any traces of residual wash buffer and ethanol in the purified DNA solution which may inhibit downstream processing during sequencing. The Gene-JET purification column was transferred into a new sterile 1.5 mL micro-centrifuge tube and 30 µL of elution buffer added at the centre of the purification column membrane and centrifuged for 1 minute to collect the eluted DNA sample.

3.6 Gel electrophoresis for the PCR products

A total of 5 µL of the cleaned PCR products of each primer pair were electrophoresed on 1 %

agarose gel stained with 4 µL ethidium bromide (0.5 µg/mL). The results were examined for the bands of the expected size i.e. checked based on the DNA molecular weight marker as a reference. In order to prepare 1 % agarose gel: 1 g of agarose was weighed, dissolved in 100 mL of TBE buffer and the procedure described in section 3.2.2 was followed. The gel was run at 80 Voltage (V) for 1.5 hours, then visualized under a gel imaging Doc (Thermo-Fisher Scientific, USA) to confirm the band sizes in reference to the DNA molecular weight marker used.

3.7 DNA sequencing

The products of PCR were sequenced at the Molecular and Infectious Diseases Research Laboratory (MIDR) of the University of Nairobi, Faculty of Health Sciences, using Sanger sequencing method.

3.8 Sequence data and diversity analysis

3.8.1 Pre-processing (quality check of the sequences)

A total of 110 paired end reads (both forward and reverse) generated from the sequencing procedure were pre-processed for quality check before downstream sequence analysis. Quality of the reads was first checked manually using Bioedit software, (version 7.2 Bioedit Ltd, United Kingdom) which helps to determine whether peaks of the sequences are evenly spaced and with only one color; i.e. to know whether there is base line noise. The quality scores associated with each sequence chromatogram were viewed on Geneious Prime software (version 2021.2.2, Bio-matters Tech Ltd, New Zealand) through chromatogram graph options that display a quality measure (Phred quality score) for each base as assessed by the base calling program. Reads that had a good phred quality score of above 30 were selected and used for analysis.

3.8.2 Assembly of reads into consensus sequence

Having ascertained the quality of the reads; the high-quality paired end reads were edited and assembled into contigs; which were then used to obtain full length consensus sequences for each sample using Geneious prime software version 2021.2.2. Both de-novo assembly and assembly using a reference sequence were used. *Phaseolus vulgaris* cultivar BRS Vereda alpha-phaseolin (Phs) gene, complete cds (sequence ID: KJ544115.1); obtained from both phytozome and NCBI databases was used as a reference sequence to help identify potential alignment sites and retrieval of genomic coordinates.

3.8.3 Multiple sequence alignment

All of the assembled consensus sequences of each of the common bean accessions were aligned together to obtain a multiple sequence alignment (Appendix 2). The sequences were aligned to a reference sequence using MUltiple Sequence Comparison by Log- Expectation (MUSCLE) embedded on the Molecular Evolutionary Genetics Analysis (MEGAX) software version 11.09. Prior to the alignment, the sequences were trimmed to uniform lengths using Jalview version 2.11.0. Multiple sequence alignment (MSA) of translated transcript of each sample i.e. the amino acid sequences of the phaseolin gene was also done using MUSLE software.

3.8.4 Single nucleotide polymorphism and allelic diversities

Sequence polymorphisms i.e. single nucleotide polymorphisms (SNPs) were screened across the sequenced ά-Phs gene for all the 11 common bean accessions used in this study. The polymorphisms were analyzed using DNA Sequence Polymorphism (DnaSP) software version 6.12 (Posada, 2009). In order to confirm the SNPs as true mutations, Codon-Code Aligner software (version 10.0.2, Codon-Code Corporation) was used to visualize all the SNP and indels substitutions as it confirms actual loci positions as well as amino acid codon changes.

3.8.5 Neutrality test analysis

Neutrality test analyses including Tajima D, Fu and Li's D* and F* were performed on DnaSP software, on the aligned sequences (Korneliussen *et al*., 2013).

3.8.6 Linkage disequilibrium and recombination events

Linkage disequilibrium and recombination analysis was determined by considering loci associations using dnaSP software. The investigation of linkage disequilibrium was done between pairwise segregating sites so as to predict the genetic marker density and resolution needed to identify genetic variants that tag causal variants (Xu *et al*., 2014).

3.8.7 Phylogenetic analysis

Phylogenetic analysis was done on MEGA X software to generate a phylogenetic tree which helps to depict the genetic relationships of the phaseolin gene in the various accessions studied. Neighbor-Joining method was applied to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) so as to obtain the initial tree (s) for the heuristic search. Maximum likelihood was preferred because it evaluates the probability that the chosen evolutionary model will have generated the observed sequences. Evolutionary rate differences was determined by applying a discrete Gamma distribution among the sites (5 categories (+G, parameter = 0.0500), which enabled for some sites to be evolutionarily invariable ([+I], 49.32% sites). This analysis involved 12 different nucleotide sequences. The evolutionary history was inferred by using the Maximum Likelihood method and Kimura 2-parameter model (Kimura, 1980). The tree that showed the highest log likelihood (- 1571.26) was chosen.

3.8.8 Haplotype and gene diversity analysis

Haplotype diversity analysis of the common bean accessions was done using DnaSP software version 6.12.03; which was able to generate haplotype data files. Arlequin software version 3.5.2.2; was then used to infer the haplotype frequencies (Excoffier and Lischer, 2010).

3.8.9 Pairwise individual genetic distances

Molecular Evolutionary Genetics Analysis (MEGA X) software was used was used to estimate the sequence differences between the common bean accessions; which yielded statistical pairwise nucleotide differences between the accessions used in this study.

3.9 Protein sequence analyses

The composition of amino acids of the predicted phaseolin (α-Phs) protein for all common bean accessions studied was calculated using PEPSTATS (Chojnacki *et al*., 2017). The chemical and physical parameters of protein such as molecular weight, theoretical isoelectric point (pI), total number of positively and negatively charged residues, extinction coefficient ($M^{-1}cm^{-1}$), instability index, aliphatic index and grand average of hydropathicity (GRAVY) were calculated, according to the amino acid scale values by Kyte & Doolittle, (1982) using ProtParam tool at ExPASy (Walker *et al*., 2005).

3.9.1 Structural analysis of α-Phs protein

Structural analysis of the predicted α-Phs protein was done by a system of neural networks on Predict-Protein software (Yachdav *et al.*, 2014). Different secondary structure states were predicted including: helix (H; includes alpha-, pi- and 3_10-helix), (beta-) strand (E = extended strand in beta-sheet conformation of at least two residues length) and other (O).

3.9.2 Modelling of α-Phs protein

The predicted phaseolin amino acid sequence was used to build three-dimensional protein structure using homology modelling approach on SWIISS MODEL software (Bordoli *et al.*, 2009). This was achieved based on a sequence alignment between the target protein and the template structure. SWISS-MODEL software uses experimentally determined structures of related family members as templates. Homology modelling was preferred because it is currently the most accurate method to generate reliable three-dimensional protein structure models.

3.10 Correlation analysis for the exonic polymorphisms associated with protein content

The association between polymorphisms seen on the exonic region of the phaseolin gene was done using Tassel software version 5.0 with mixed linear model (MLM) (Bradbury *et al*., 2007). A total of 10 polymorphisms/variants on the exonic region including 7 SNPs and 3 parsimony informative sites were used for the analysis. The correlation was further done by comparing the protein content values against the shared polymorphisms for the common bean accessions used in this study and multivariate correlation coefficient analysis approach.

CHAPTER FOUR: RESULTS

4.1 Seed protein content in selected common bean accessions

A broad range of protein content was observed in mature seeds of the 11 studied common bean accessions (Figure 4.1). The protein content of the common bean accessions ranged from 126.7 to 429.00 mg/g of sample. The analysis of variance showed that the accessions have significantly different protein content in the mature seeds (Table 4.1). The highest and lowest protein contents were observed in the accessions 19 and Mbeere4, respectively with a significant difference at $p \leq 0.05$. The common bean accession 19 had 238.6% more protein content than accession Mbeere4. The protein concentration of accession 19 was higher than Mbeere3 however, it was not significantly different.

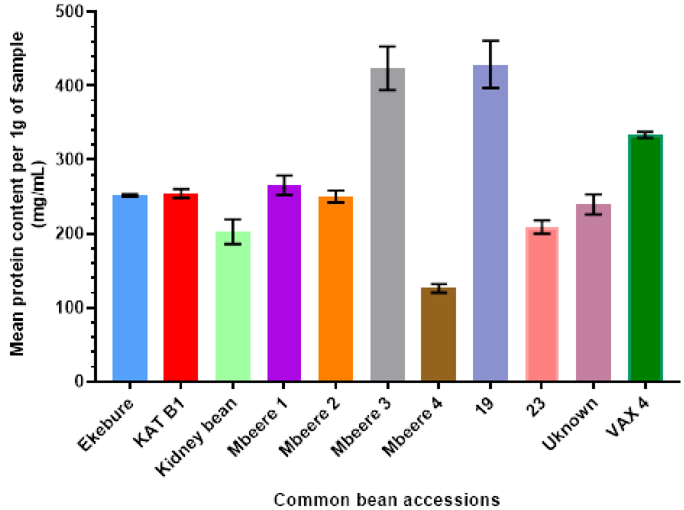


Figure 4.1: Protein concentration in mature seeds of 11 studied common bean accessions using Lowry method. Results are showing mean values ± SD of three biological replicates. The significance values at 95% confidence levels among the pairs are illustrated in Table 4.1.

Table 4.1: Significance values at the 95% confidence level in protein content values among the paired common bean accessions.

Tukey's multiple comparisons test	Mean Difference	95% CI	Adjusted P-Value	
Ekebure vs. Mbeere 4	125.6	90.83 to 160.4	0.0028	A-G
Ekebure vs. 23	42.8	0.2593 to 85.34	0.0494	A-I
Ekebure vs. VAX 4	-81.8	-110.5 to -53.08	0.0077	A-K
KAT B1 vs. Mbeere 3	-169.2	-310.8 to -27.56	0.0354	B-F
KAT B1 vs. Mbeere 4	128.3	85.62 to 171.1	0.0066	B-G
KAT B1 vs. 19	-174.6	-332.5 to -16.59	0.0412	B-H
KAT B1 vs. VAX 4	-79.06	-126.8 to -31.33	0.018	B-K
Kidney bean vs. Mbeere 1	-62.7	-112.8 to -12.65	0.0322	C-D
Kidney bean vs. Mbeere 3	-220.8	-319.3 to -122.3	0.0118	C-F
Kidney bean vs. 19	-226.2	-333.3 to -119.1	0.0126	C-H
Kidney bean vs. VAX 4	-130.7	-214.2 to -47.24	0.0201	C-K
Mbeere 1 vs. Mbeere 3	-158.1	-252.5 to -63.66	0.0176	D-F
Mbeere 1 vs. Mbeere 4	139.4	48.42 to 230.4	0.021	D-G
Mbeere 1 vs. 19	-163.5	-273.6 to -53.44	0.0225	D-H
Mbeere 2 vs. Mbeere 4	124.3	55.61 to 192.9	0.0154	E-G
Mbeere 2 vs. 23	41.45	24.90 to 58.00	0.0101	E-I
Mbeere 2 vs. VAX 4	-83.15	-128.5 to -37.76	0.0151	E-K
Mbeere 3 vs. Mbeere 4	297.5	114.8 to 480.2	0.0186	F-G
Mbeere 4 vs. 19	-302.9	-502.6 to -103.2	0.0215	G-H
Mbeere 4 vs. 23	-82.8	-148.3 to -17.28	0.0316	G-I
Mbeere 4 vs. Uknown	-113.5	-198.5 to -28.46	0.0282	G-J
Mbeere 4 vs. VAX 4	-207.4	-266.8 to -148.0	0.0034	G-K

Note: Common bean accessions that do not share a letter are significantly different according to Fisher's pairwise test ($p \leq 0.05$) .CI = confidence of interval. The confidence interval for the difference between the mean protein contents range which does not include zero, indicate that the difference between these means is statistically significant.

4.2 Phaseolin gene amplification and sequencing

Ten primer pairs were used in the PCR amplification of the full-length phaseolin (α-Phs) gene. All the 10 primer pairs produced clear and reproducible bands of the respective fragments of α- Phs gene. Single bands of sizes ranging from 500 to 650 bp were obtained for all the DNA samples of the 11 common bean accessions. Figure 4.2 represents the amplification profile of the 10 primer pairs with the respective band sizes using DNA sample from common bean accession (KATB1).

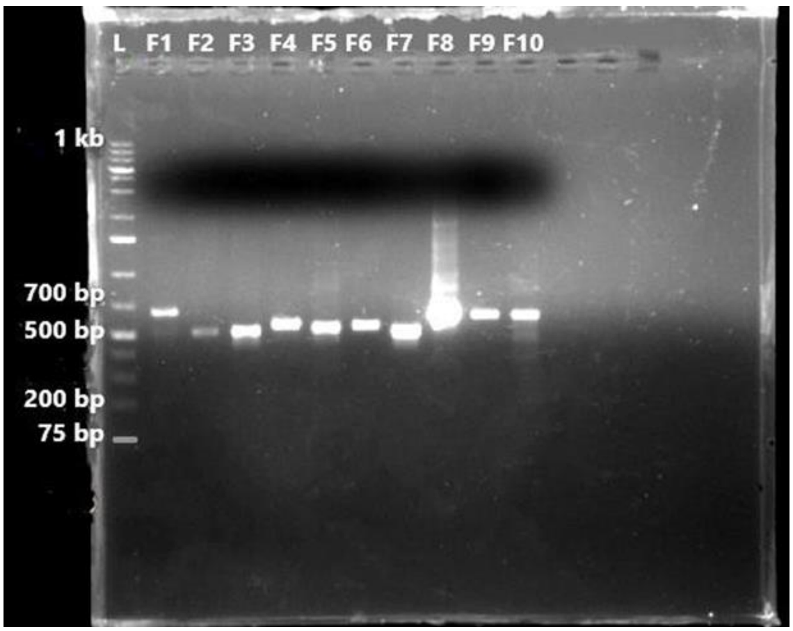


Figure 4.2: A representative gel image of PCR amplification profile for common bean accession KATB1 using 10 primer pairs targeting fragments of phaseolin gene. L in the diagram represents the DNA molecular weight marker used to estimate the band sizes of the PCR products. The primer pairs used are labelled as 1 to 10 (all the 10 primer pairs used on KATB1 accession, represented as F). All the primer pairs generated amplicon sizes of between 500 - 650

base pairs (bp).

The sequencing rate was 100% for all the amplified PCR products. All the purified PCR products of 10 fragments of α-Phs gene for the 11 common bean accessions were successively sequenced. Figure 4.3 is a representative chromatogram displaying peaks generated for Mbeere2 common bean accession.

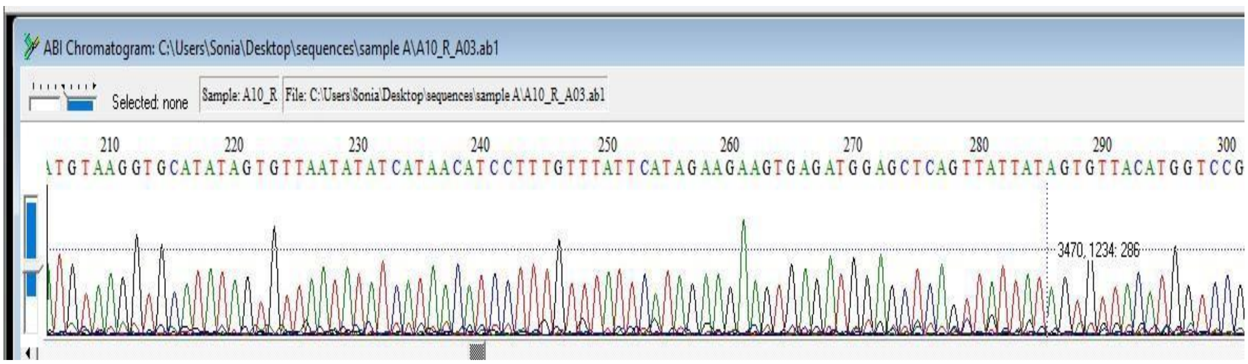


Figure 4.3: Representative chromatogram showing reads generated from sequencing of PCR product of Mbeere2 common bean accession. The single clear peaks with one color depict the quality of reads generated from the sequencing procedure. The colors of the peaks are corresponding to the different DNA nucleotides.

For each of the 11 common bean accessions, 10 paired end reads corresponding to fragments of the phaseolin (α-Phs) gene, were assembled into contigs (overlapping fragments) and consensus sequences (Figure 4.4). The sequences of the amplified products obtained using 10 primer pairs when overlapped covered the entire length of the α-Phs gene (Figure 4.4).

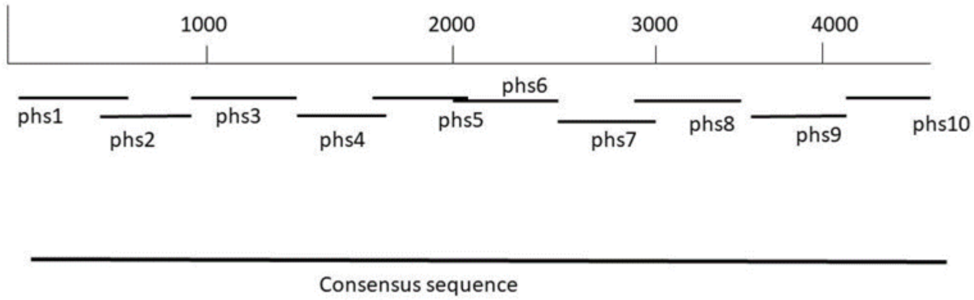


Figure 4.4: Schematic representation of the contigs of the phaseolin (α-Phs) gene. The lines labelled phs 1 - 10 represent the overlapping fragments of the α-Phs gene amplified using 10 primer pairs targeting different fragments of phaseolin gene.

4.3 Phaseolin (α-Phs) gene sequence diversity

4.3.1 Phaseolin (α-Phs) gene structure

To define the exon and intron regions of α-Phs gene, the full-length consensus sequences were compared to the corresponding coding DNA sequences (CDS) of the reference sequence (Phaseolus vulgaris; KJ544115.1.) retrieved from the NCBI. The gene was approximately 4502 bp long and revealed 5 exons with a varied range of nucleotides (Figure 4.5) and 6 distinct introns. Based on the consensus sequences of the α-Phs gene of the 11 common bean accessions, exon 1 was between positions (2316 - 2624), exon 2 (2697 -2888), exon 3 (2977 – 3057), exon 4 (3182 to 3410) and exon 5 (3539 - 3762). The α-Phs exon/transcript region was 1035 bp long.

All the 6 introns occupied different gene positions including the first intron between 1-2315 bp, second intron between 2625 – 2697 bp, third intron between 2889 – 2971 bp, fourth intron between 3058 – 3182 bp, fifth intron between 3411 – 3538 bp and the sixth intron between 3763 – 4502 bp.

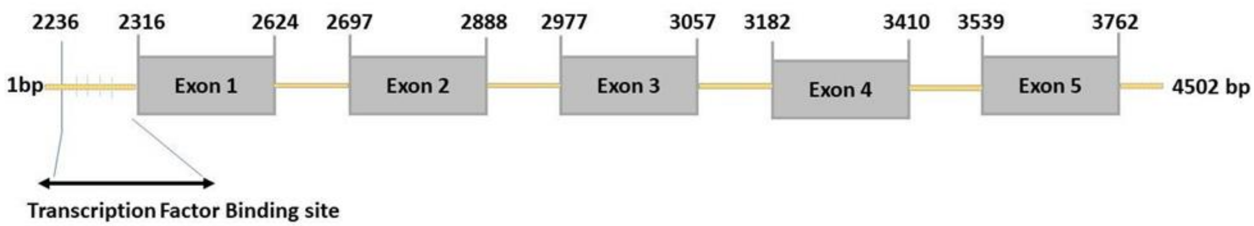


Figure 4.5: Gene structure (organization structure) of the Phaseolus vulgaris α-Phs gene for common bean accession KATB1 (representing α-Phs gene structure in all common bean accessions).

4.3.2 Sequence polymorphism

The multiple sequence alignment of the 11 consensus sequences and the reference sequence from NCBI (KJ544115.1.) revealed a total of 41 mutations/polymorphic sites within the entire gene region (both intron and exon regions), an average of 1 mutation for every 110 bp (Appendix 3). Out of the 41 mutations, 24 were SNPs and 17 were parsimony informative sites among the 11 common bean accessions studied. Single nucleotide polymorphism was present in different positions of α-Phs gene which include 419, 439, 477, 480, 500, 593, 598, 759, 849, 925, 982, 1234, 1334, 2214, 2566, 2567, 2759, 2796, 2828, 2872, 2878, 3508, 3766, and 3846 across the 11 common bean accessions studied. The indels/parsimony sites were in positions 359, 465, 476

543, 806, 815, 872, 974, 1051, 1124, 1539, 2045, 2129, 2313, 3041, 3216 and 3400. The analyzed α-Phs gene full length sequence (4502 bp) is 66% AT rich. The nucleotide distribution on the α-Phs gene was: A (35 %, 1580 bp), T (31 %, 1336 bp); G (16%, 733 bp), and C (18 %, 853 bp). Regarding the base mutations, transverse and transition mutations were 51.22 % and 48.78 %, respectively. The C/T substitutions were the highest with 20.73 % (Table 4.2).

Table 4.2: Nucleotide changes in the full length α-Phs gene of *P. vulgaris* based on the consensus sequences.

Type of base substitution	Nucleotide changes	No. of nucleotide substitution	% of nucleotide substitution	No. of amino acid substitution
Transition	C>T	17	20.73	4
	A>G	8	9.75	2
	T>C	3	3.66	2
	G>A	12	14.63	3
	Sub-total	40	48.78	11
Transversion	A>C	10	12.19	3
	A>T	16	19.51	5
	T>A	3	3.66	3
	G>C	6	7.3	2
	C>A	7	8.54	3
	Sub-total	42	51.22	16
	Total	82	100	27

The mutations on the intron regions were 31, occupying different positions (Table 4.3) in the different common bean accessions. The mutations were present on introns 1, 5 and 6. The fifth intron spanning loci 3411 - 3538 had only one polymorphic site in the sequence of common bean accession sample 23 as a result of transversion mutation (T>A) at position 3508. The sixth intron (3763 - 4502) had two mutation sites at base position 3766 (A>C) and 3846 (T>C) in the sequence of common bean accession Kidney bean as a result of transversion and transition, respectively.

There were no mutations in introns 2, 3 and 4.

Table 4.3: Mutated loci on all intron regions of α-phaseolin gene for each analyzed common bean accession.

Intron region	Accession ID(s)/Name(s)	Mutation site (Base substitution)	Type of substitution
Intron 1 (1-2315 bp)	Mbeere 1, Mbeere 4	359 (C>T)	Transition
	Unknown	419 (C>T) 982 (A>C) 1334 (C>T)	Transition Transversion Transition
	VAX4	439 (T>A) 500 (A>C	Transversion Transversion
	Mbeere 1, Kidney bean	465 (A>G)	Transition
	Mbeere 1, Mbeere 4	476 (A>T)	Transversion
	KATB1	477 (A>T)	Transversion
	Mbeere 1, Kidney bean, Mbeere 4	543 (A>C) 806 (G>A)	Transversion Transition
	Mbeere 2	593 (G>A) 598 (C>T) 2313 (A>T)	Transversion Transition Transversion
	Mbeere 4	759 (G>A)	Transition
	Mbeere 1, Kidney bean, KATB1	815 (G>A) 1051 (C>T)	Transition Transition
	Kidney bean	849 (A>G)	Transition

	Mbeere 1, Kidney bean, KATB1, Mbeere 4	872 (C>T)	Transition
	Mbeere 1	925 (A>T)	Transversion
	Mbeere 1, KATB1, VAX4	974 (A>G)	Transition
	Mbeere1, Unknown, Kidney bean, KATB1	1124 (C>A)	Transversion
	Ekebure	1234 (G>A)	Transition
	Mbeere1, Unknown, Kidney bean, KATB1, Mbere4	1539 (A>T)	Transversion
	Mbeere1, Unknown, Kidney bean, Mbeere4	2129 (G>C), 2045(A>T)	Transversion Transversion
	Kidney bean	2214 (C>T)	Transition
Intron 5	23	3508 (T>A)	Transversion
Intron 6	Kidney bean	3766 (A>C)	Transversion
	Kidney bean	3846 (T>C)	Transition

Mutations were also observed on the various positions of the exons of α-phaseolin gene except exon 5 (Table 4.4). No mutation was observed in exon 5 (3539 – 3762 bp). A total of 10 mutations which include 7 SNPs and 3 indels/parsimony informative sites were detected on the 4 exons (exons 1 to 4). Of the 10 mutations, 1 was synonymous, while 9 were non-synonymous and led to amino acid replacements. Exon 1 (2316 – 2624 bp) had two polymorphic sites on sequence of common bean accession, sample Unknown. Exon 2 (2697 – 2888 bp) had 5 polymorphic sites. Exons 3 (2977 – 3057 bp) and 4 (3182 to 3410 bp) had 1 and 2 polymorphic sites, respectively.

Table 4.4: Mutation positions on exons of α-phaseolin gene for each of the analyzed common bean accessions

Region/motif	Sample ID	Mutation sites (Base substitution)	Type of substitution	Codon change	Type of mutation
Exon 1	Unknown	2566 (C>A), 2567 (C>A)	Transversion Transversion	Pro856Lys	Non-synonymous
Exon 2	Mbeere 1	2759 (G>C)	Transversion	Gly920Ala	Non-synonymous
Exon 2	Mbeere 3	2756 (G>C)	Transversion	Leu932Leu	Synonymous
Exon 2	Mbeere 3	2828 (T>C)	Transition	Ile943Thr	Non-synonymous
Exon 2	VAX4	2872 (C>A)	Transversion	Pro958Thr	Non-synonymous
Exon 2	Mbeere1	2878 (T>C)	Transition	Phe960Leu	Non-synonymous
Exon 3	Mbeere3, Unkown,19	3041(C>T)	Transition	Pro1014Leu	Non-synonymous
Exon 4	Mbeere2,	3216 (A>C)	Transversion	STP1072Cys	Non-synonymous
	KATB1, Kidney bean				
Exon 4	Unknown, Kidneybean, 23	3400 (G>A)	Transition	Gly1134Arg	Non-synonymous

4.4 Predicted amino acid changes

The predicted amino acid changes were visualized on Codon Code Aligner software. The mutations resulted to 10 amino acid changes on the coding region among the 11 common bean accessions (Table 4.4). Among the 10 amino acid changes, 9 were conservative (non-synonymous mutations leading to a replacement of an amino acid and one was non-conservative (resulting in amino acid changes with similar biochemical properties).

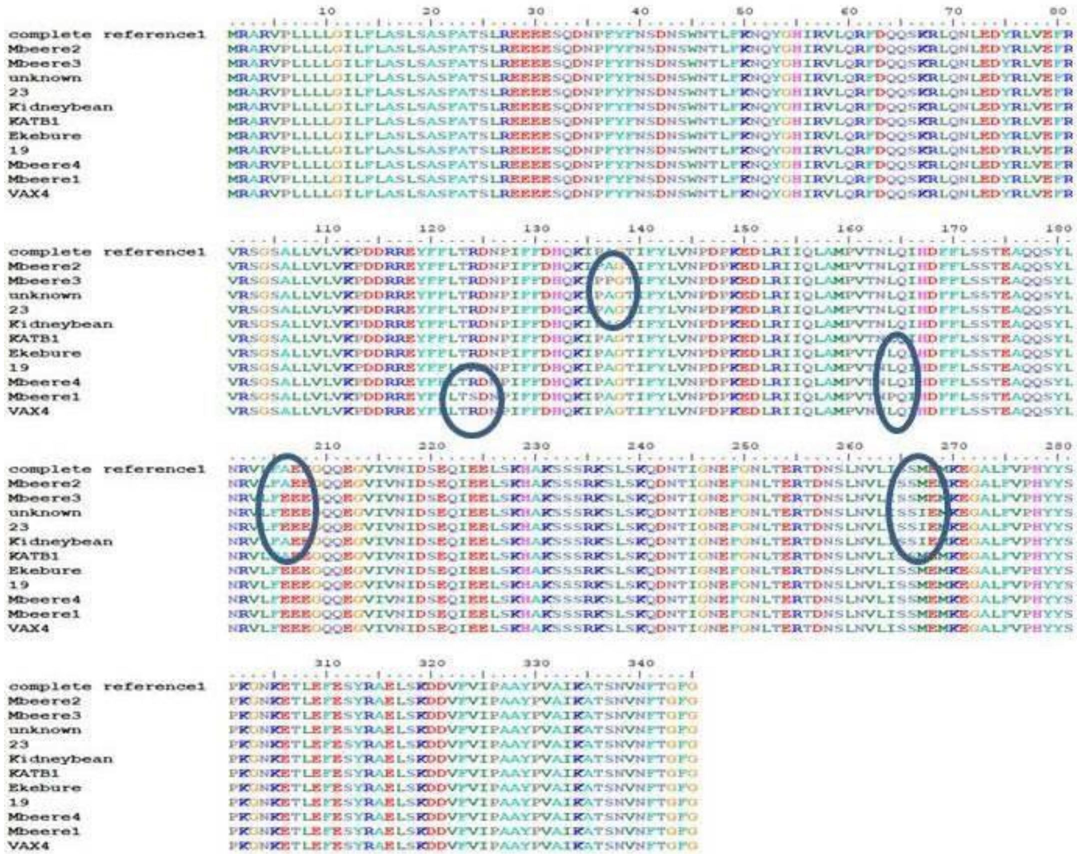


Figure 4.6: Multiple sequence alignment of the predicted protein/ amino acid sequence (345bp) of the transcript region of α-phaseolin gene from the 11 common bean accessions. All the sequences were aligned together with the reference sequence (KJ544115.1.) obtained from NCBI database. The eclipses indicate to the mutated sites. Red color represents highly conserved residue while blue color represents weakly conserved residue.

4.5 Genetic diversity and differentiation

The full-length sequence had a nucleotide diversity (π) of 0.00271. There was higher nucleotide diversity in the intron region (π = 0.002751) than in the exon regions (π = 0.00256). Among all the exons, the highest nucleotide diversity was observed in exons 3 and 4, with π = 0.00629 and π= 0.00473, respectively. The sequences had high percentage sequence conservation, C: 0.991 (99.1%). The coding region (1035 bp) had nucleotide diversity (π) of 0.00256, haplotype diversity index (Hd) of 0.964 and a conservation index (C) of 0.990 (99.0 %).

A total of 9 haplotypes were observed from the coding regions of the α-Phs gene sequences of the 11 tested common bean accessions. The most frequent haplotypes were Hap_1 and Hap_6, both which contained two common bean accessions. Haplotype distributions included Hap_1 comprising of Mbeere2 and KATB1, Hap_2 comprising of Mbeere3, Hap_3 comprising of Unknown common bean accession, Hap_4 comprising of common bean accession identified as 23, Hap_5 comprising of Kidney bean, Hap_6 comprising of Ekebure and Mbeere4 accessions, Hap_7 comprising of common bean accession identified as sample 19, Hap_8 comprising of Mbeere1 and Hap_9 comprising of VAX4. The highest haplotype distribution frequency was observed in exons 2 and 4, each with 4 haplotypes. The lowest haplotype diversity was found on exon 1 (Hd = 0.1818), while exon 4 showed the highest variability (Hd = 0.6909). The overall haplotype diversity (Hd) index was 0.9636 (Table 4.5). Based on the full length (4502 bp) sequence of the α-Phs gene in all the common bean accessions analyzed, a total of 11 haplotypes were detected with each sample corresponding to a particular haplotype distribution.

Table 4.5: Nucleotide and haplotype diversity in the coding region and conserved motifs of the α-Phs gene of *P. vulgaris.*

Region/motif	Fragment length (bp)	No. of SNPs	No. of indels	Sequence conservation (C)	Haplotype No.	Haplotype diversity (Hd)	Nucleotide diversity (π)
Exon 1	309	2	None	0.994	2	0.1818	0.00118
Exon 2	192	5	None	0.974	4	0.4909	0.00473
Exon 3	81	None	1	0.988	2	0.5091	0.00629
Exon 4	229	None	2	0.991	4	0.6909	0.00381
Exon 5	224	None	None	None	1	0.000	0.000
Entire exon region	1035	7	3	0.990	9	0.9636	0.00256

4.6 Pairwise genetic distances

Pairwise genetic distances analysis revealed the number of base substitutions per site between the sequences analyzed. The highest pairwise genetic distance was observed between common bean accessions Kidney bean and Mbeere2 with a value of 0.004463 (Table 4.6). The lowest pairwise genetic distance value was 0.000457 between common bean accessions 19 and Mbeere3.

Table 4.6 Estimates of genetic divergence between sequences.

	Mbeere2	Mbeere3	Unknown	23	Kidney bean	KATB1	Ekebure	19	Mbere4	Mberee1
Mbeere3	0.001612									
Unknown	0.003424	0.003137								
23	0.001360	0.001120	0.002931							
Kidney bean	0.004463	0.004176	0.004645	0.003970						
KATB1	0.002940	0.002653	0.003123	0.002448	0.003532					
Ekebure	0.001143	0.000949	0.002761	0.000697	0.003800	0.002277				
19	0.001155	0.000457	0.002679	0.000662	0.003718	0.002196	0.000492			
Mbeere4	0.003376	0.003089	0.003559	0.002884	0.002541	0.002445	0.002713	0.002632		
Mbeere1	0.004413	0.004126	0.004596	0.003921	0.002746	0.003482	0.003751	0.003669	0.002492	
VAX4	0.001841	0.001554	0.003177	0.001349	0.004216	0.002693	0.001178	0.001097	0.003129	0.004167

4.7 Deviation from a standard neutral model

To investigate if α-Phs gene can be used in common bean improvement and if selection was taking course, full length sequence (4502 bp) and coding regions were tested by neutrality tests, including the D* and F* of Fu & Li's and Tajima's D statistics. The values for the D* and F* of Fu and Li's and Tajima's D of common bean α-Phs were significantly less than 0 indicating abundance of rare alleles and purifying selection, thus this gene may be selected for improvement process (Table 4.7). The computed Tajima's D test statistic was -0.71864, Fu and Li's D* was -1.00960 and Fu and Li's and F* -0.96413 for full length gene.

Table 4.7: Neutrality test statistical values on the coding region.

Parameters	Exon 1	Exon 2	Exon 3	Exon 4	Exon 5	Entire exon region
Fu and Li's D*	-1.65766	-2.1251	0.77552	1.07842	0.00000	-1.00960
Fu and Li's F*	-1.79737	-2.09823	0.97943	0.99697	0.00000	-3.645
Tajima's D	-1.42961	-1.79107	1.18560	0.85048	None	-0.94712

On the coding region Tajima's D and Fu's Fs statistics were -0.94712 and -3.645, respectively, and were significant (P-value < 0.025) suggesting excess frequency of rare alleles while Fu and Li's D* test statistic was -1.00960 non-significant (P value > 0.10).

4.8 Mutation as a result of recombination

The polymorphic sites in the entire sequence were used to detect recombination events. A minimum of 6 recombination events (Rm) were found to be responsible for the polymorphism of α-Phs gene. Recombination events were detected in the informative sites found in pairs of mutated positions (359, 465), (465, 476), (872, 974) and (974, 1539) all on intron1 gene region, mutated positions (2129, 3041) located between intron 1 and exon 3 gene regions, and mutated positions (3216, 3400) on exon 4.

4.9 Phylogenetic tree analysis

The evolutionary history inferred by using the Maximum Likelihood Method and Kimura 2-parameter model (Kimura, 1980), generated a phylogenetic tree with branch lengths in proportion to the genetic distances used to infer the tree (Figure 4.7). The phylogenetic tree construction for both protein and nucleotide sequence revealed two major clusters, A and B (Figure 4.7). Within the Cluster A, there was a unique sub-cluster between Mbeere3 and 19 common bean accessions with the highest protein contents.

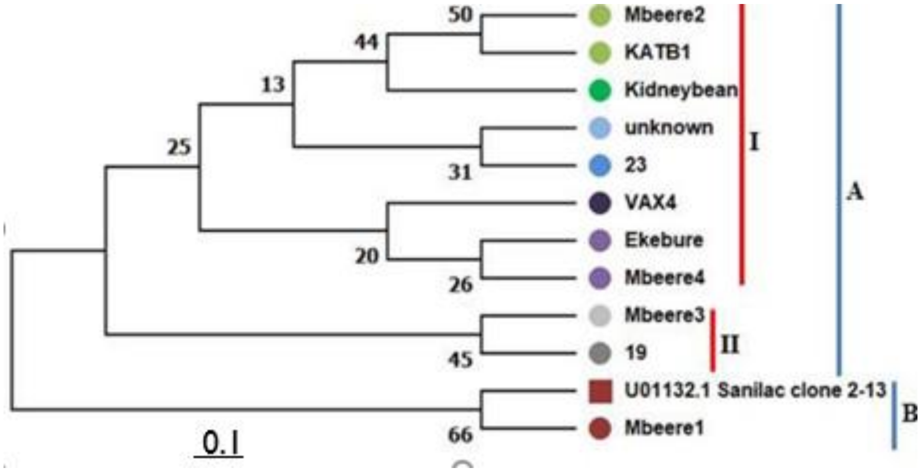


Figure 4.7: Phylogenetic tree based on the full length α-phaseolin gene sequences inferred by Maximum Likelihood method. The percentage of replicate trees in which the associated accession clustered together in the bootstrap test (1000 replicates) is shown next to the corresponding node branches.

4.10 Linkage disequilibrium

The linkage disequilibrium analysis performed on DnaSP software revealed significantly positive LD (D′) results with coefficient of correlation values ($R^2 \geq 1$) between some of the mutated site pairs including (359, 476), (439, 500) (Table 4.8). All the pairs of mutated positions whose coefficient of correlation (R^2) was equal to 1.0 were considered significant according to Fischer's exact values described by Kulinskaya and Lewin, (2009). Linkage disequilibrium (LD) was calculated in all the polymorphic/ informative sites (Appendix 4). The α-Phs with very fast LD decay is suitable for use in further candidate gene association analysis. LD decay over nucleotide was found to be a steady decline as shown by significant linkage disequilibrium ($R^2 \geq 1$), which was maintained over a 110 bp distance within the 4502 bp nucleotide sequences for the α-Phs gene (Figure4.8).

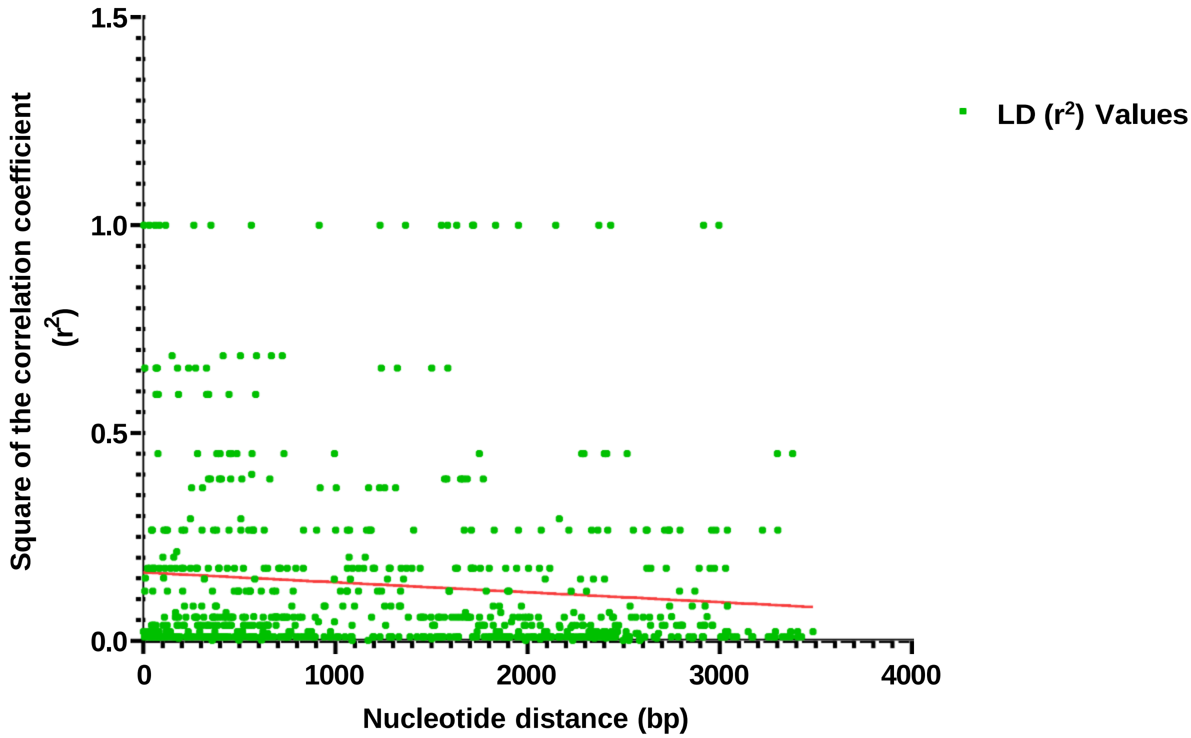


Figure 4.8: Plot of R^2 (the linkage disequilibrium statistic) versus nucleotide sequence for phaseolin gene among 11 common bean accessions. The green dots depict pairwise comparisons among informative sites of the gene. Values of Fischer's exact test that showed significant linkage are those whose coefficient of correlation (R^2) = 1.

Table 4.8: Significant linkage disequilibrium output on paired informative sites.

Site1	Site2	Dist	D	D'	R	Fisher
359	476	117	0.139	1.000	1.000	0.015*
419	982	563	0.076	1.000	1.000	0.083
419	1334	915	0.076	1.000	1.000	0.083
419	2566	2147	0.076	1.000	1.000	0.083
419	2567	2148	0.076	1.000	1.000	0.083
439	500	61	0.076	1.000	1.000	0.083
439	2872	2433	0.076	1.000	1.000	0.083
477	480	3	0.076	1.000	1.000	0.083
500	2872	2372	0.076	1.000	1.000	0.083
543	806	263	0.188	1.000	1.000	0.005
593	598	5	0.076	1.000	1.000	0.083
815	872	57	0.222	1.000	1.000	0.002
849	2214	1365	0.0076	1.000	1.000	0.083
849	3766	2917	0.076	1.000	1.000	0.083
849	3846	2917	0.076	1.000	1.000	0.083
925	2796	1834	0.0761	1.000	1.000	0.083
925	2878	1953	0.076	1.000	1.000	0.083
982	2566	1584	0.076	1.000	1.000	0.083
982	2567	1585	0.076	1.000	1.000	0.083
1334	2566	1232	0.076	1.000	1.000	0.083
1334	2566	1233	0.076	1.000	1.000	0.083
2045	2129	84	0.056	1.000	0.426	0.002
2214	3766	1552	0.076	1.000	1.000	0.083
2214	3846	1632	0.076	1.000	1.000	0.083
2759	2878	119	0.076	1.000	1.000	0.083
2796	2828	32	0.076	1.000	1.000	0.083
3766	3846	80	0.076	1.000	1.000	0.83

Note: The asterisk indicates one linked site on the α-Phs gene that mapped far from all other informative sites.

4.11 Estimated gene genealogy using transitive consistency score (TCS)

Transitive consistency score (TCS) network analysis of the relationship of haplotypes showed the number of mutations leading to the occurrence of a particular haplotype (Figure 4.9). The analysis confirmed 9 distinct haplotype distributions. Haplotypes comprising of one common bean accession were Hap_2 (Mbeere3), Hap_3 Unknown), Hap_4 (accession 23), Hap_7 (accession 19), Hap_8 (Kidney bean) and Hap_9 (VAX4). The numbers of hatch marks were equivalent to the number of mutations resulting in a specific haplotype. Each node size was directly proportional to the frequency of the haplotypes in that specific node indicating that the larger the node the more the number of common bean accessions.

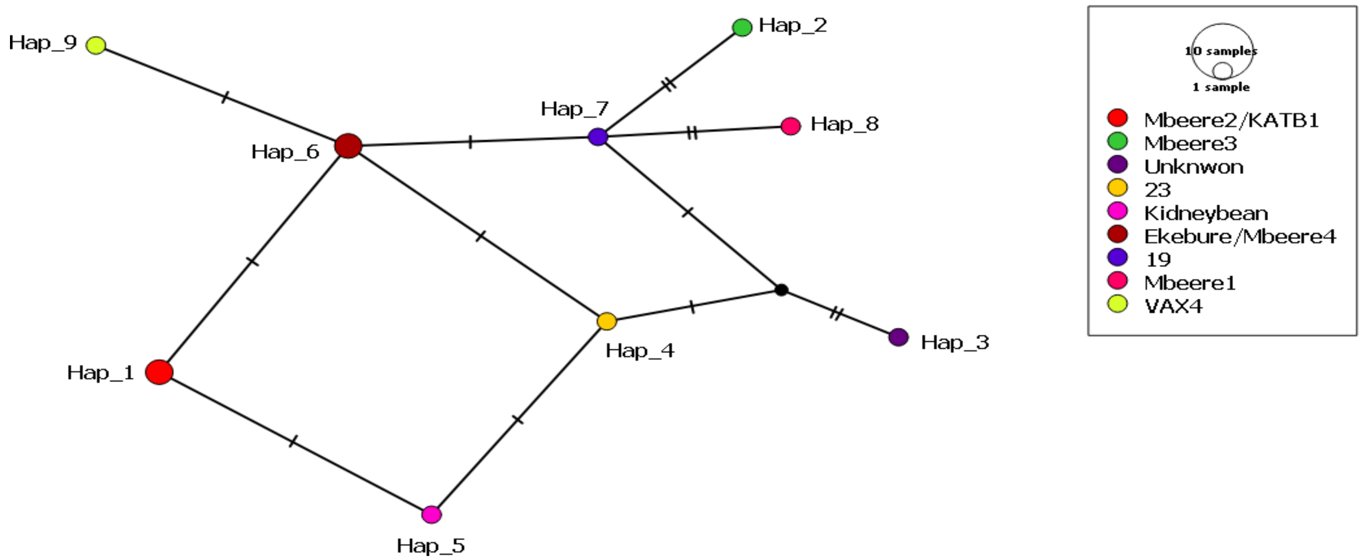


Figure 4.9: Haplotype networks for α-Phs using transitive consistency score (TCS). A total of 9 haplotypes were revealed from the genealogy analysis using the protein coding regions. The hatch marks on the TCS represents polymorphic sites. The size of the circle/node equates to the frequency of the observed haplotypes. The unlabeled node between Hap_3 and Hap_4 is a transition node.

4.12 Physiochemical features (biochemical characteristics) of α-Phs in *P. vulgaris*

Primary structural analysis of the translated protein of the α-phaseolin gene revealed 20 amino acid residues to be present in the predicted protein sequence. Leucine was the most abundant amino acid (Table 4.9). The frequency of amino acids in α-Phs were determined to establish whether there were differences between common bean accessions in terms of protein quality (Appendix 5). The percentage of Leucine (11.6% - 11.9%) was not significantly different among the tested common bean accessions. The rare amino acids were tryptophan and methionine. Tryptophan was significantly higher in kidney bean common bean accession as compared to all the other 10 amino acids (Table 4.9).

Table 4.9: Relative frequency of predicted amino acids in α-Phs from mature seeds of 11 common bean accessions.

Common bean accession	Abundant amino acid	Rare amino acids	Missing amino acids
Kidney bean	Leu (11.59 %)	Trp (1 %), Met (0.8 %)	Cys, Glu
KATB1	Leu (11.6 %)	Trp (0.3 %) Met (0.9 %)	Cys
Ekebure	Leu (11.6 %)	Met (1.2 %)	Cys
Mbeere1	Leu (11.6%)	Trp (0.3%), Met (1.2%)	Cys
Mbeere2	Leu (11.7 %)	Trp (0.3 %)	Cys
Mbeere3	Leu (11.9 %)	Trp (0.3 %) Met (1.2%)	Cys
Mbeere4	Leu (11.7 %)	Trp (0.3 %) Met (1.2%)	Trp Met
19	Leu (11.89 %)	Trp(0.29%) Met (1.15 %)	Cys
Unknown	Leucine (11.9 %)	Trp (0.3 %) Met (0.9%)	Cys
23	Leucine (11.6 %)	Trp (0.3%) Met (0.9%)	Cys
VAX4	Leucine (11.6 %)	Trp (0.3%) Met (0.9%)	Cyst

Note: Leu = Leucine, Met = Methionine, Cys = Cysteine, Trp = Tryptophan, Glu = Glutamic acid

The molecular weight of Phaseolus vulgaris α-Phs translated protein varied from 39306.38 (Kidney bean) to 39424.53 (Mbeere3) g/mol each with 345 amino acid residues. The predicted isoelectric point of α-Phs protein ranged from 4.96 (for Mbeere1) to 5.07 (for Mbeere2). The average predicted isoelectric point was 5.01 in all common bean accessions studied. Number of positively and negatively charged side chains in α-Phs protein are 50 and 35, respectively, in all common bean accessions, except KATB1 and Kidney bean with 49 negatively charged side chains and Mbeere1 with 34 positively charged side chain. Grand Average of Hydropathy (GRAVY) ranged from − 0.305 to − 0.336. Individual hydrophobicity values of all the amino acids were confirmed and predicted according to Kyte & Doolittle, (1982) (Figure 4.10). Aliphatic index mean was 92.56 for all common bean accessions and it ranged from 91.83 to 94.09. The molar extinction coefficient recorded at 280 nm measured in water, considering all pairs of cysteine ranged from 0.517 to 0.519 M^{-1} cm^{-1} in all accessions. Molar extinction coefficient for α-Phs protein in common bean accessions was 20400 (Table 4.10).

Table 4.10: Biochemical characteristics/properties of predicted protein by Pritparam.

Common bean accession	MW (g/mol)	pI	Molar Extinction coefficient	Extinction coefficient ($M^{-1}cm^{-1}$)	Charge	Average residue weight	II	GRAVY	AI	R-	R+
Ekebure	39382.45	5.02	20400	0.518	-11.5	114.152	53.05	-0.328	91.83	50	35
KATB1	39324.42	4.899	20400	0.519	-10.5	113.984	52.10	-0.312	92.12	49	35
Kidney bean	39306.38	5.07	20400	0.519	-10.5	113.932	53.37	-0.305	93.25	49	35
Mbeere1	39313.35	4.96	20400	0.519	-11.5	114.190	52.71	-0.317	91.83	50	34
Mbeere2	39324.42	5.07	20400	0.519	-10.5	113.984	52.10	-0.312	92.12	49	35
Mbeere3	39424.53	5.02	20400	0.517	-11.5	114.274	53.18	-0.322	92.67	50	35
Mbeere4	39382.45	5.02	20400	0.518	-11.5	114.152	53.05	-0.328	92.80	50	35
19	39398.50	5.02	20400	0.518	-11.5	114.199	53.18	-0.312	92.96	50	35
23	39364.42	5.02	20400	0.518	-10.5		54.32	-0.320	92.6	50	35
Unknown	39411.48	5.02	20400	0.518	-11.5	114.12	55.42	-0.310	94.09	50	35
VAX4	39395.45	5.02	20400	0.518	-11.5	114.190	53.73	-0.336	91.83	50	35

Note: MW (Molecular Weight), pI (Isoelectric Point), GRAVY (Grand Average of Hydropathy), A (Aliphatic Index), and II (Instability Index) were predicted based on amino acid composition.

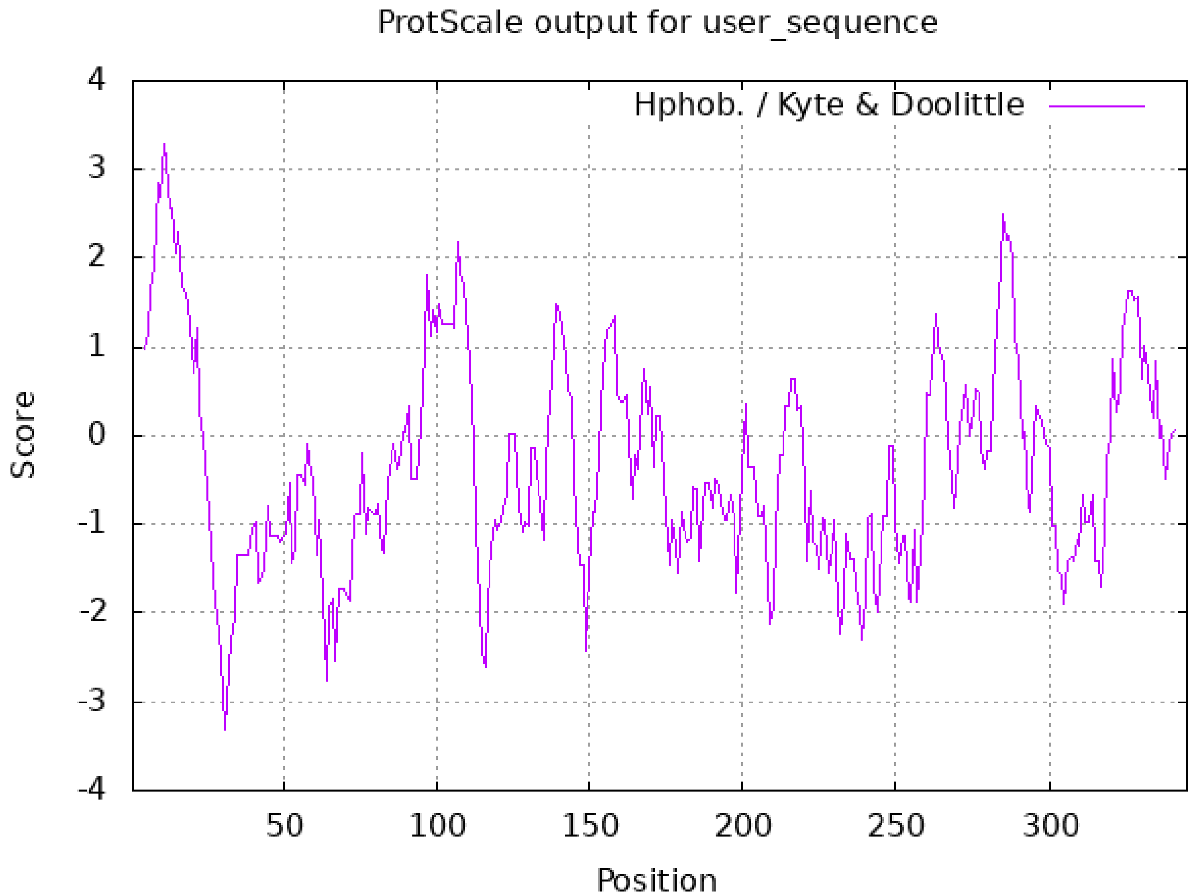


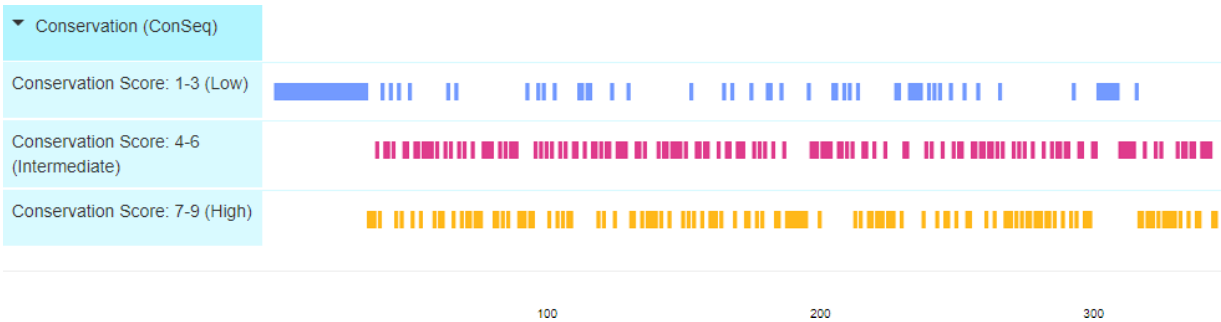
Figure 4.10: Hydrophobicity values of all the amino acids according to Kyte & Doolittle, (1982).

4.13 Secondary and tertiary structure (3D modelling) of the phaseolin protein

Structural characteristics of α-Phs in common bean accessions including secondary structure of the protein, conservation score and disordered and DNA binding regions were predicted by Protein predict software (Figure 4.11). The protein structure showed helix and other protein strand structural elements. The predicted protein length was 345 bp long with bi cupin protein regions. Loops (53 - 55.1%) were the most abundant secondary structure in the protein sequences followed by extended strands (27.3 - 29.3 %) and helices (13.3 - 15.7%). Transmembrane helices occupied signal peptide regions (1 - 20) in all the protein sequences. There were no disulfide bond strands as per the predicted secondary structure. Table 4.11 gives detailed information on the secondary and tertiary structures of common bean α-Phaseolin.

(A)

(B)



(C) Disordered and DNA binding regions

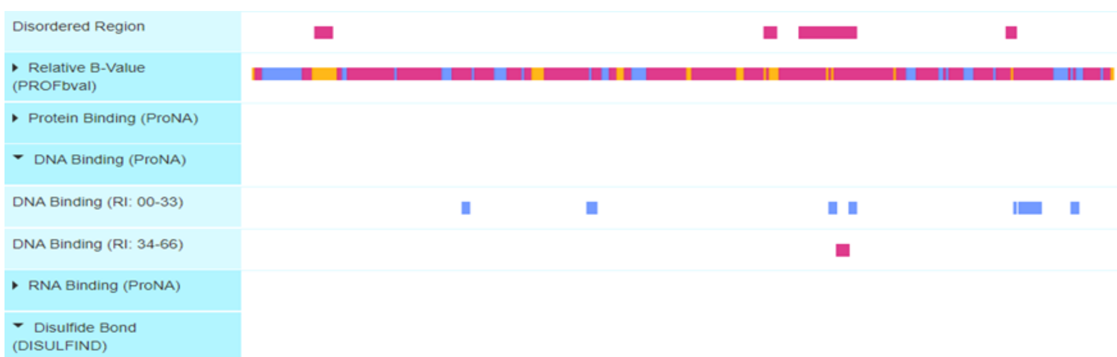


Figure 4.11: Structural characteristics of α-Phs in common bean accessions. (A) Secondary structure of the protein as predicted by protein-predict software. (A) Secondary structure of the protein as predicted by protein-predict software. (B) Conservation Score predicted by Protein-Predict software; (C) Disordered and DNA binding regions.

Protein domain predicted from the amino acid sequences indicated regions of the protein to be moderately conserved.

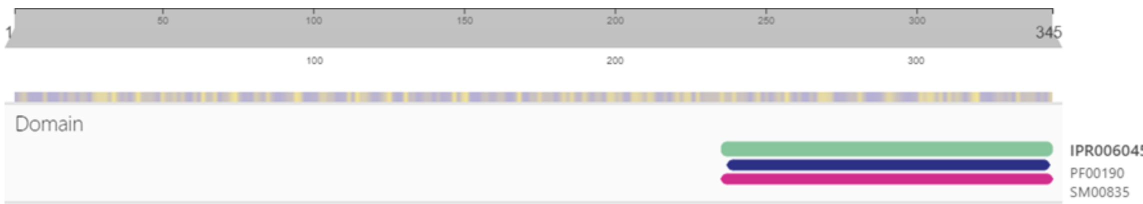


Figure 4.12: Protein domain prediction from the amino acid sequences predicted from the coding region. Colors indicate specific amino acid classes. Colors indicate specific amino acid classes: Green and yellow represent hydrophobic residues, while blue indicate hydrophilic polar uncharged.

The predicted phaseolin amino acid sequence generated from transcript regions of Mbeere2, Mbeere3, Mbeere4 and sample 19 common bean accessions yielded a representation of three-dimensional protein structure of phaseolin protein using homology modeling approach on SWIISS MODEL (Figure 4.13). The predicted 3D structure was almost similar for the common bean accessions analyzed indicating similarities in biochemical properties of the protein. Common bean accession 19 had the best confidence score of 0.94, Template modeling (TM) - score of 0.913±0.08 and Root-mean-square deviation (RMSD of 1.34 ± 1.8 Å, while Mbeere4 had the lowest confidence score of 0.81, Template modeling (TM)-score of 0.90 ± 0.06 and Root-mean-square deviation (RMSD) of 1.23±1.4 Å. Mbeere3 and Mbeere 2 samples were included for comparison purposes in terms of structure, since they recorded high and low protein concentrations, respectively.

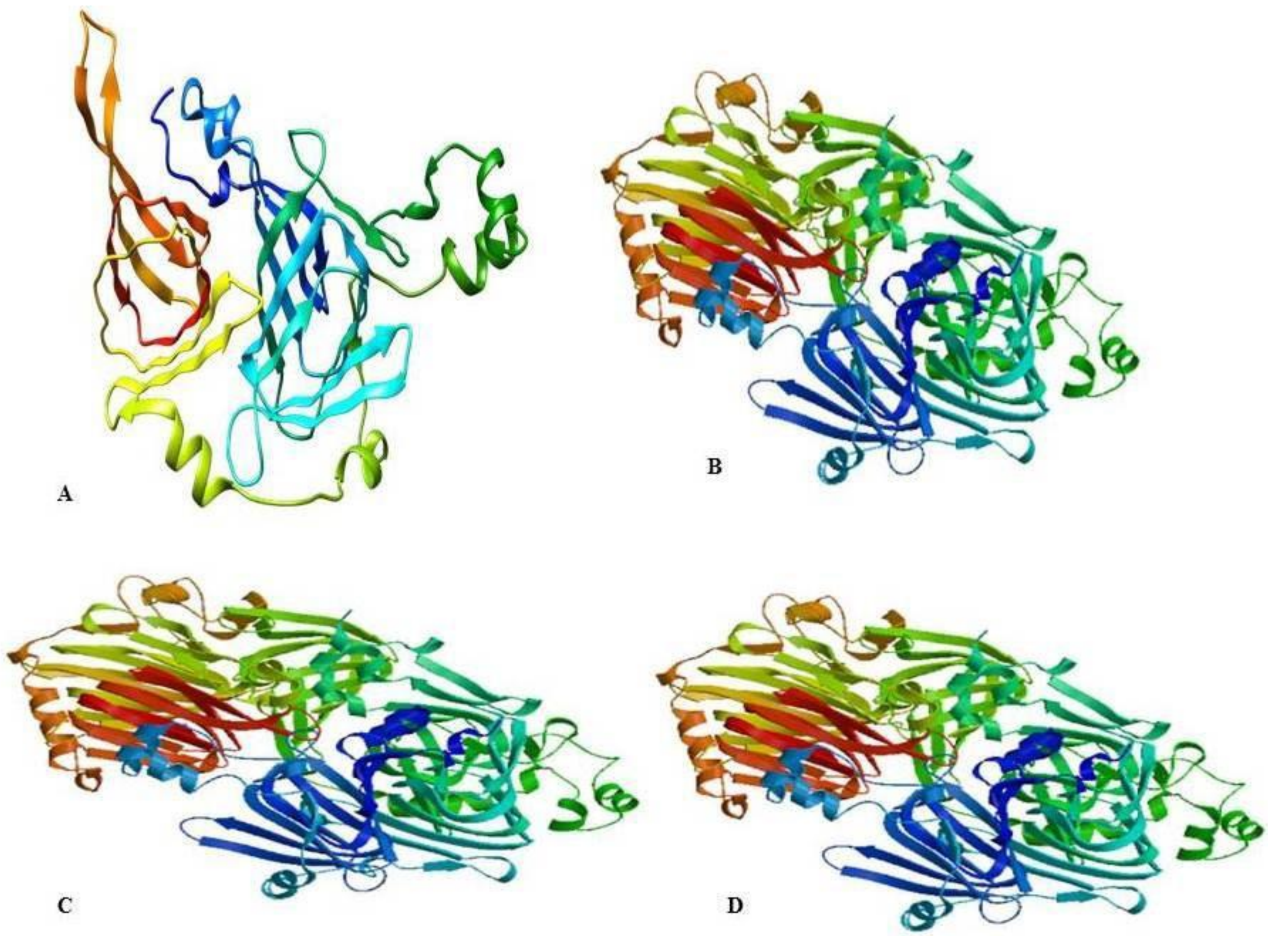


Figure 4.13: Three dimensional (3D) modelled structure selected common bean accessions. A= Mbeere2 common bean, B: Mbeere3 common bean, C: 19 common bean accession, D: Mbeere 4. Structure of phaseolin at 2.2 angstroms resolution, an implication for a common legumin structure and the genetic engineering of seed storage proteins. Red and blue colors in the ribbon represent basic and acid residues. Green colors represent predicted binding ligand positions on the protein.

Table 4.11: Detailed information on the secondary and tertiary structures of common bean α-Phaseolin.

Accession	Template	QMEAN	GMQE	Range	Information of α-helices		Information of β-sheets	
					Location	Overall ratio	Location	Overall ratio
19	2phl.1.A	0.88	0.85	30-345	5-16 7-21, 180-180 186-192	45.00	47-50, 54-60 75-82 262-269	5.14
Mbeere4	2phl.1.A	0.87	0.84	34-345	7-21, 5- 16, 63- 63, 180- 180 186-192 197-205 220-230	46.00	37-39 78-82 152-159 262-269	4.90
Mbeere 3	2phl.1A	0.88	0.85	34-345	7-21 26-32 180-180 186-192	47.00	54-60 152-159 176-179 193-196	5.12
					195-205		262-269	
Mbeere 2	2phl.1A	0.87	0.85	33-345	7-21	45.00	278-82	5.0
					180-180		152-159	
					186-193		197-999	
					222-232			

4.14 Relationship between α-Phs gene polymorphism and seed protein content

Association analysis was performed to find possible relationship between nucleotide polymorphisms in α-Phs gene and protein content. Correlation was observed between overall mutations observed per exon on 2 of the exons and total protein content. The correlation coefficient values obtained from the analysis indicated positive correlations on exons 2 (correlation coefficient (r) of 0.48) and 3 (correlation coefficient (r) =0.6 respectively (Figure 4.14). Although mutations on exon 2 had a positive correlation coefficient of 0.48, the p-value registered was statistically not significant (p=0.137). Mutation on exon 3 (position 3041) with non-synonymous mutation revealed a strong correlation coefficient of 0.6 when compared with protein content in each of its samples (Table 4.12). Exon 5 did not have any mutations thus no correlation.

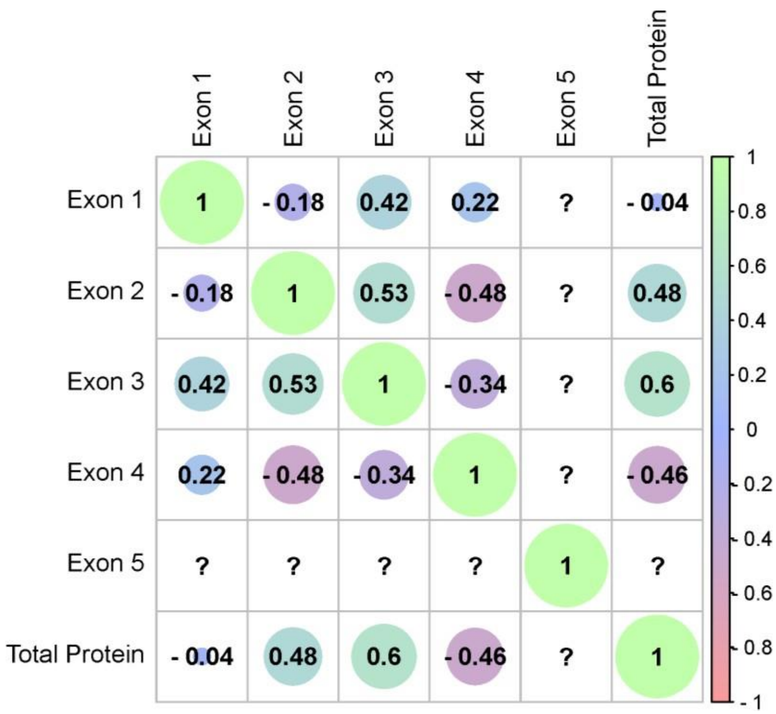


Figure 4.14: **Output of association analysis**. The result indicates positive correlations on exons 2 and 3 (correlation coefficients of 0.48 and 0.6, respectively) when compared with overall protein content. The color distribution suggests level of correlation, green color represents strong correlation towards the value of 1, and the blue color represents middle correlation whereas the purple color represents weak correlation.

Exon 4 (positions 3216 and 3400) had mutations comprising of common bean accessions with close range protein contents. Samples belonging to the same haplotypes (similar mutations) had nearly identical amounts of protein. Common bean accessions KATB1 and Mbeere2 belonging to Hap_1 had 250.35 and 254 (mg/g) protein content per 1g of sample, respectively and shared non-synonymous mutation on position 3126 (exon 4). The non-synonymous mutation on position 3400 was shared between Kidney bean (202.80 mg/g) and accession 23 (208.90 mg/g) both with close range protein contents. Parsimony informative mutation on position 3041 of exon 3 was common for Mbeere3 bean accession (423.60 mg/g) and accession 19 (429.00 mg/g) which recorded the highest protein contents. Overall, three non-synonymous mutations were found to be significantly associated with protein content. The mutations on positions 3041, 3126, and 3400 bp in the genomic region of α-Phs were significantly associated with protein content.

Table 4.12: Significant associations between mutation of α-Phs and protein content of common bean accessions.

Mutation position	Region	r	P-value	Nucleotide substitution	Codon change	Common bean accession ID	Protein content (mg/g)
3041	Exon 3	0.6	0.050	C>T*	Pro1014Leu	Mbeere3, 19	423.60 429.00
3126 and 3400	Exon 4			C>T*	Pro1014Leu	KATB1 Mbeere2	250.35 254

Note: r is the correlation coefficient between the exon and protein content. df is the degree of freedom for each data point. *Indicates the mutation was significantly ($P \leq 0.05$) associated with protein content as per the p- value.

CHAPTER FIVE: DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

5.1 Discussion

5.1.1 Protein content in mature seeds of selected common bean accessions

Common bean is a major grain legume crop with an estimated production of 28.9 million metric tons globally (Nadeem *et al.*, 2021). The grains are an important source of protein for both humans and animals. In the present study, wide variability with respect to seed protein content was observed in 11 common bean accessions grown under similar environmental conditions. These results indicate that it is possible to identify promising parents for exploitation in breeding programs with a view to increasing the overall seed protein content. In general, the protein content ranged from 126.7 to 429.00 mg/g of sample. These values are within the estimated range by Kotue *et al*. (2018) and Ceyhan (2006) in common bean germplasm from USA and Turkey respectively.

Typically, most dry beans contain 15 to 25% protein on a dry weight basis (dwb). Water-soluble albumins and salt-soluble globulins constitute between 10 to 30% and 45 to 70% of the total proteins (Sathe, 2002). A single 7S globulin dominates dry globulins fraction and may account for up to 50 to 55% of the total proteins in the dry beans. Most dry bean proteins are deficient in sulfur amino acids, methionine, and cysteine, and therefore are of lower nutritional quality when compared with the animal proteins. Some of the analysed common bean had protein amounts that are below human nutritional need. For instance, some of the samples registered as low as 0.3 Methionine content, which is considered deficient relative to human requirement. The suggested nutritional requirements for methionine-cysteine in the human diet are between 2.5 and 2.6 %

which is equal to between 26 and 25 mg methionine-cysteine gram per protein (McLarney *et al.*, 1996 ; Millward, 2015).; Millward, 2015). The protein content is one of the most relevant food grain legume traits for breeding, as the interest in plant-based protein increases in developed countries to provide healthier diets, and the need for cheap protein sources to fight malnourishment in developing countries remains (Considine *et al*., 2017; Magrini *et al*., 2016). Consequently, breeding of enhanced protein content is an important objective of common bean genetic improvement. Screening genetic variability for protein content in common bean germplasm is the first step to identifying high protein grain varieties for development of high yielding and high protein cultivars.

5.1.2 Nucleotide diversity of phaseolin gene in selected common bean accessions

In grain legumes, genes encoding seed storage proteins such as albumin, globulin, legumin, vicilin, glutelin and phaseolin are involved in protein content in legumes. Of these, phaseolin gene plays a pivotal role in modifying the protein content in common beans (Montoya *et al*., 2018). Considering the paramount importance of phaseolin gene in regulating the seed protein content in common beans, this study sought to characterize the phaseolin gene and the encoded protein at molecular level and compared among selected common bean accessions. This information could aid in the genetic improvement through SNPs/mutations linked to the high protein content or directly through a transgenic strategy.

In the current study, nucleotide sequence analysis revealed polymorphic sites in the intron and coding regions of the common bean phaseolin gene. By aligning the sequences of phaseolin gene, 24 SNPs and 17 parsimony informative sites were detected which were distributed unequally across the DNA fragments. The results showed that SNPs were more frequent than parsimony informative sites/indels on the noncoding than the coding region of phaseolin gene. This discrepancy was expected since noncoding regions are known to be neutral and thus retain point mutations, and they usually show higher diversity. In addition, the decrease in nucleotide polymorphism in the exon region suggests that it is heavily influenced by greater selection pressure. Petrovski *et al.*, (2013) reported that coding regions of functional genes tend to be conserved due to their specificity for and affinity with other types of molecules. Coding

indels usually result in stronger purifying selection when compared to SNPs. SNPs have been reported to be the most common form of DNA sequence variation between alleles in several plant species (Morgil *et al.*, 2020). The polymorphisms registered nucleotide substitutions that were both transversion and transition mutation types. Transition mutations can be linked to selection and occur when a pyrimidine base substitutes for another pyrimidine base or when a purine base substitutes for another purine base. Transversion mutations affect genetic diversity and occur when a purine is substituted for a pyrimidine or vice versa.

SNPs, a substitution, insertion or deletion of a single nucleotide are fundamental genetic resources in plant breeding programs (Zhu *et al*., 2008). The analysis of genetic diversity for functional genes is important for understanding the genetic background of phenotypic variation for crop improvement. The number of polymorphisms identified in the current study were fewer as compared to the previous report by Diniz *et al*. (2014) on phaseolin gene, which can be attributed to the unusual pattern of gene sequence variation and other aspects of common bean evolution such as convergent phenotypic evolution, the sample size used and the differences in the bean gene pool. The differences could also be explained by the geographic origin of the germplasm used in the study.

Polymorphism detected at the gene level of phaseolin in the present study can influence polypeptide variability in the different accessions. A study by Slightom *et al*. (1985), elaborated that one of the reasons for the existence of polymorphisms observed among the phaseolin polypeptides was due to existence of differences in DNA sequences, which encode two different polypeptide sub-families namely α- and β-phaseolin. Despite the presence of polymorphic sites, the gene was highly conserved, with a 99.9% conservation rate for the full-length gene. The high sequence conservation rate suggests that the gene could have been subjected to weak selection pressure over time.

Protein functions and gene expressions are related to the location of SNPs in coding regions or regulatory sequences (Coolon *et al*., 2015). The abundant genetic variations is important in plant breeding because it allows for the selection of crop varieties that are more superior and suited to different agricultural systems (Xu *et al*., 2014).

All the 10 mutations identified in the coding region were non-synonymous except one synonymous mutation at position 2796. The synonymous mutations are attributed to selection pressure acting on the phaseolin gene. The non-synonymous mutations in the coding region predicted codon/ amino acid changes which could have a negative or positive impact on the structure and function of the phaseolin glycoprotein and as a result, affect protein content in common bean germplasm. For instance, codon change on position 2759 i.e. Gly920Ala (G→A) has the potential to affect the phaseolin protein structure and function. Glycine is greatly preferred at the amino-terminal (N) and carboxyl-terminus (C) caps at internal positions while alanine stabilizes the structure of the protein helices. The amino acid alanine has a methyl group in the side chain (R group) and is generally hydrophobic hence contributes to closeness in protein folding. The nucleotide substitution at position 2566 (Pro856Lys (C→A), resulted in change from proline to lysine amino acid thus might increase helix forming propensity as lysine tend to stabilize the secondary protein structure by their capability of forming hydrogen bonds with negatively charged non-protein atoms. Proline on the other hand disrupts protein secondary structure by inhibiting the formation of alpha-helices or beta-sheets. Non-synonymous mutations have been shown to have a direct and definite effect on the functions of many genes in plants, whereby the genes evolve to increased functional diversity at the targeted trait (Bitocchi *et al.*, 2017).

The sequence diversity levels were investigated through the calculation of nucleotide variability (π). Nucleotide variability/ diversity depends on the source of polymorphism, and the set of common bean accessions under study (Ma *et al*., 2010). In this study, nucleotide diversity value (π) was 0.00271 for the full-length gene sequence. There was slightly higher nucleotide diversity in the intron region ($\pi = 0.002751$) than in the exon regions ($\pi = 0.00256$) indicating that the intron region is less conserved than the exons. Nucleotide diversity can be influenced by factors such as life history characteristics and anthropogenic activity and the low values registered suggest the sequences analyzed are highly conserved (Loeuille *et al*., 2021). Purifying selection is also associated with low values of nucleotide diversity (0.00271) and Tajima's D because only low frequency polymorphisms can avoid being eliminated by widespread direction selection.

The presence of natural selection was also confirmed by neutrality test statistics including Tajima's D values as well as Fu and Li's F* statistic. Tajima's D test statistic indicated that purifying selection was affecting the majority of sites within the phaseolin gene, reducing genetic diversity (Cvijović *et al*., 2018). The selection pressure was however generally weak because the computed p-value for Tajima's D statistic was not significant for full length α-Phs gene. Tajima's D value as well as Fu and Li's F* statistics were statistically significant for the coding region indicating an excess of intermediate frequency polymorphism confirming the presence of balancing or diversifying selection.

Phylogenetic analysis establishes a balance of evolutionary and phylogenetic relationships of various species of organisms, so as to define the organism's genetic diversity over time and space (Pace *et al*., 2012). Phylogenetic analysis of the α-Phs gene sequences showed that there were variations among the common bean accessions analyzed. The variations could be attributed to the genetic differences of the common bean accessions. Classification of the common bean accessions led to the formation of two clusters and a unique sub-cluster was formed consisting of only two accessions with significantly high seed protein content. These results showed that phaseolin, the major seed storage protein of common bean is an important DNA marker in studies of genetic diversity and evolution of common bean populations due to its functional and structural properties (De La Fuente *et al*., 2012).

Recombination and linkage disequilibrium were identified as factors that demonstrate a systematic association between allelic polymorphisms at two different informative sites. The presence of positive linkage disequilibrium between 27 paired informative sites as well as a history of recombination events on 6 different pairwise positions, confirmed the history of selection pressure. From the sequence analysis it can be deduced that phaseolin gene is experiencing moderately fast evolution as a result of a combination of factors such as genetic drift and selective sweeps. Common beans are known to have a wide range of both phenotypic and morphological differences, as well as an abundant genetic diversity (Long *et al*., 2020). The crop has a high effective frequency of recombination. In this study, negative linkage disequilibrium suggested a history of random drift, which reduces the number of variants while increasing homozygosity and thus influence the loss of favorable mutations (Guzmán Díaz *et al*., 2001).

The haplotype diversity obtained was 0.9636 for the full-length gene sequence among the common bean accessions, indicating that the phaseolin gene was highly polymorphic. Haplotype computation for individual exon regions varied with the highest haplotype diversity indices being observed in exons 2 and 4. The higher values of haplotype diversity corresponded to the frequency of haplotypes in each exon. The presence of haplotypes among the common bean accessions analyzed provides strong evidence for comparable patterns of evolution in their domestication processes and seed storage protein contents (Reddy *et al.,* 2017). The Haplotype network distribution by transitive consistency score showed a total of nine haplotypes, comprising of a set of nearby genomic structural variations, such as polymorphic SNPs, with a strong linkage disequilibrium (LD) between them.

Based on the molecular weight, (39313.35 to 39398.50 g/mol) α-Phs can be considered as a small protein. In all the common bean accessions studied, the isoelectric point (pI) is less than 7.0, indicating their precipitation in acidic buffers and this is important information as it can be used in the purification of phaseolin protein. The size of the protein and pI were generally more conserved in all the common bean accessions. The most abundant amino acid in all the tested common bean accessions was leucine (11.6% - 11.9%) and was not significantly different among the tested common bean accessions while the rarest amino acids were Cysteine, Methionine and Tryptophan. This suggests that amino acid concentrations in beans vary according to genetic factors (Flores-Sosa *et al.*, 2020). The predicted protein sequence in α-Phs gene indicated both positively and negatively charged side chains chains which can bond with one another to hold a length of protein in a certain shape or conformation. Grand Average of Hydropathy (GRAVY) values in all the protein sequences were below 0 implicating that phaseolin protein can have interactions with water and is therefore hydrophilic. Aliphatic index mean was 92.56 for all common bean accessions suggesting that the phaseolin protein is thermostable over a wide range of temperature(Atsushi, 1980). The aliphatic index was contributed by the aliphatic amino acids such as alanine, glycine, isoleucine, leucine, proline, and valine present in phaseolin protein amino acid chain. The molar absorption coefficients values ranged between 0.517 to 0.519 $M^{-1}cm^{-}$ in all the phaseolin protein sequences, which is considerably low absorptivity. This can be linked to the absence of disulfide bonds as a result of low fractions of Cysteine amino acids and Tryptophan. The absorbance of a protein at 280 nm depends on the content of Tryptophan, Tyrosine, and cysteine (disulfide bonds) (Pace *et al.*, 1995). The molar coefficient values are significant in measuring epsilon values for phaseolin protein during its characterization.

The most abundant secondary structure in the predicted protein sequences was loops followed by extended strands, and helices signifying the possibility of phaseoiln protein interaction with other biological molecules. Absence of disulfide bonds on the protein structure suggests that then phaseolin protein may be a cytoplasmic protein, hence cysteine residues are unlikely to form disulfides. Protein modeling is an important way to decide the three-dimensional structure, interactions, protein function, domain structure and ligand binding site (Moturu *et al*., 2018). Three-dimensional structure of phaseolin protein yielded a 3D structure at 2.2 angstroms resolution, which is considered high resolution value an indication that templates used were a good match with the targeted phaseolin protein sequences during homology modeling.

5.1.3 Effect of phaseolin sequence polymorphism on seed protein content

In the current study, there was an association between α-Phs sequence variations/polymorphisms and seed protein content of the analyzed common bean accessions. The high protein contents in the two common bean accessions (19 and Mbeere3) which recorded the highest protein concentrations could be explained by the mutations within exon 3 gene region. The presence of a $3041^{C\rightarrow T}$ mutation located in exon 3 leads to a Pro1014Leu substitution in the protein sequence. Proline is an amino acid with a unique cyclic structure that facilitates the folding of many proteins but also disrupts the protein secondary structure by inhibiting the backbone to conform to an alpha helix or beta sheet formation. On the other hand, leucine is a nonpolar amino acid which provides hydrophobic bulk but limit internal flexibility.

Mutation at position $2566^{C\rightarrow A}$, Pro856Lys located in exon 1 did not show any association with protein content. The change from proline to lysine amino acid maintained the stability of the protein. Lysine amino acid is a positively charged basic amino acid often found on protein surfaces. They take part in protein stability by several interactions such as the hydrogen bond formations, ionic interactions, and interaction with water.

Association analysis is a powerful method to explore the relationship between sequence polymorphisms and phenotypic variation (Zhu *et al*., 2008). Many studies have shown a correlation between phaseolin type and seed weight, seed size, growth habit precocity and antiparasitic traits (P. Gepts & Bliss, 1986; Koenig *et al*., 1990; Johnson *et al*., 1996; Escribano *et al*., 1998).

5.2 Conclusions

The conclusions from this study are:

(i) The selected 11 common bean accessions showed significant variations in seed protein content. The variability in protein content can be explained by the genetic characteristics of the common bean accessions under study.

(ii) The α-phaseolin gene in the studied common bean accessions showed high nucleotide polymorphisms and genetic diversity.

(iii) Genetic variants on exon 3 of α-Phs gene showed significant association with seed protein content in studied common bean accessions. The SNPs can be applied as molecular markers in breeding programs to improve protein content in common beans germplasm after validation.

5.3 Recommendations

The recommendations drawn from this study are:

(i) The genetic variants that showed significant correlation with protein content should be explored as potential genetic markers in marker-assisted selection and breeding of common beans.

(ii) Amino acid changes/codon changes determined in the present study should be validated in future studies to determine if they have a deleterious or favourable effect on the three-dimensional structure of phaseolin protein.

REFERENCES

Atsushi, I. (1980). Thermostability and aliphatic index of globular proteins. *Journal of Biochemistry*, *88*(6), 1895–1898.

Aylor, M. E. T., Hapman, R. A. C., & Eyaert, R. O. B. (2008). Seed storage protein deficiency improves sulfur amino acid content in common bean (*Phaseolus vulgaris* L.): Redirection of Sulfur from γ -Glutamyl- S -methyl-cysteine. *The Journal of Agricultural and Food Chemistry, 56*(14)5647–5654.

Barkutwo, J., Koech-kifuko, M., Ndungu-magiroi, K., Mutoko, M., Kamidi, M., Wanyama, J., Wanjekeche, E., Ouma, V., Miriti, J., Nyongesa, D., Githunguri, C., Thuranira, E., Esilaba, A., Okoti, M., & Lusike, W. (2020). *Onsite* , *Bean Variety and Fertilization Regime on Bean Yields in Kenya*. *10*(5), 109–114.

Beebe, S., Chirwa, R., … J. R.-S. seasons of, & 2007, U. (2012). Enhancing common bean productivity and production in Sub-Saharan Africa. *Tropicallegumeshub.Com*, 64–101.

Brosnan, J., & Brosnan, M. (2006). 5th Amino Acid Assessment Workshop. *The Journal of Nutrition*, *136*(6), 16365–16405.

Bailey-Serres, J., Parker, J. E., Ainsworth, E. A., Oldroyd, G. E. D., & Schroeder, J. I. (2019). Genetic strategies for improving crop yields. *Nature, 575*(7781), 109–118.

Koenig, R. L., Singh, S. P., & Gezrs, P. (1990). (*Phaseolus vulgaris*). *Economic Botany, 44*(July 1988), 50–60.

Bernal, J., Franco, D., & Zapata, C. (2014). Evaluating two-dimensional electrophoresis profiles of the protein phaseolin as markers of genetic differentiation and seed protein quality in common bean (*Phaseolus vulgaris* L.). *The Journal of Agricultural and Food Chemistry, 62,* 7200−7208.

Bijay, R. S., Giri, R. T., & Ram, P. C. (2020). DNA isolation and optimization of PCR protocol for ISSR analysis of *Girardinia diversifolia*: a medicinal and economic plant species from Nepal Himalaya. *African Journal of Biotechnology, 19*(10), 747–753.

Bitocchi, E., Nanni, L., Bellucci, E., Rossi, M., Giardini, A., & Spagnoletti, P. (2011). Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proceedings of the National Academy of Sciences, 109*(14).

Bitocchi, E., Rau, D., Benazzo, A., Bellucci, E., Goretti, D., Biagetti, E., Panziera, A., Laidò, G., Rodriguez, M., Gioia, T., Attene, G., McClean, P., Lee, R. K., Jackson, S. A., Bertorelle, G., & Papa, R. (2017). High level of nonsynonymous changes in common bean suggests that selection under domestication increased functional diversity at target traits. *Frontiers in Plant Science, 7*(January), 1–15.

Bliss, F. A., & Brown, J. W. S. (1983). Breeding Common Bean for Improved Quantity and Quality of Seed Protein. *Plant Breeding Reviews*, 59–102.

Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., & Schwede, T. (2009). Protein structure homology modeling using SWISS-MODEL workspace. *Nature Protocols*, *4*(1), 1– 13.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635.

Brigide, P., Canniatt-Brazaca, S. G., & Silva, M. O. (2014). Nutritional characteristics of biofortified common beans. *Food Science and Technology (Campinas)*, *34*(3), 493–500.

Broughton, W. J., Hern, Aacute, Ndez, G., Blair, M., Beebe, S., Gepts, P., & Vanderleyden, J. (2003). Beans (*Phaseolus* spp.) - model food legumes. *Plant and Soil*, *252*(1), 55–128.

C., P. N., Felix, V., Lanette, F., Gerald, G., & Theronica, G. (1995). How to measure and predict the molar absorption coefficient of a protein. *Prorein Science*, *4*, 2411–2423.

Campos-Vega, R., Oomah, B. D., Loarca-Piña, G., & Vergara-Castañeda, H. (2013). Common beans and their non-digestible fraction: Cancer inhibitory activity—an overview. *Foods*, *2*(3), 374–392.

Castro-guerrero, N. A., Isidra-arellano, M. C., Mendoza-cozatl, D. G., González-guerrero, M., & Grusak, M. A. (2016). Common bean: a legume model on the rise for unraveling responses

and adaptations to Iron, Zinc, and Phosphate deficiencies. *Frontiers in Plant Science 7*(May), 1–7.

Celmeli, T., & Sari, H. (2018). The nutritional content of common bean (*Phaseolus vulgaris* L.) landraces in comparison to modern varieties. *Agronomy*, 8,166.

Ceyhan, E. (n.d.). Variations in grain properties of dry bean.pdf. *International Journal* of *Agricultural* Research, *1*(2)116-121

Chacón S, M. I., Pickersgill, B., & Debouck, D. G. (2005). Domestication patterns in common bean (*Phaseolus vulgaris* L.) and the origin of the Mesoamerican and Andean cultivated races. *Theoretical and Applied Genetics*, *110*(3), 432–444.

Chojnacki, S., Cowley, A., Lee, J., Foix, A., & Lopez, R. (2017). Programmatic access to bioinformatics tools from EMBL-EBI update: 2017. *Nucleic Acids Research*, *45*(W1), W550–W553.

Collard, B. C. Y., & Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1491), 557–572.

Considine, M. J., Siddique, K. H. M., & Foyer, C. H. (2017). Nature's pulse power: Legumes, food security and climate change. *Journal of Experimental Botany*, *68*(8), 1815–1818.

Coolon, J. D., Stevenson, K. R., Mcmanus, C. J., Yang, B., Graveley, B. R., & Wittkopp, P. J. (2015). Molecular mechanisms and evolutionary processes contributing to accelerated divergence of gene expression on the Drosophila X chromosome. *Molecular Biology and Evolution*, *32*(10), 2605–2615.

Cortés, A. J., Chavarro, M. C., & Blair, M. W. (2011). SNP marker diversity in common bean

(*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics*, *123*(5), 827–845.

Cvijović, I., Good, B. H., & Desai, M. M. (2018). The effect of strong purifying selection on genetic diversity. *Genetics*, *209*(4), 1235–1278.

De La Fuente, M. D., López-Pedrouso, M., Alonso, J., Santalla, M., De Ron, A. M., Álvarez, G., & Zapata, C. (2012). In-depth characterization of the phaseolin protein diversity of common bean (*Phaseolus vulgaris* L.) based on two-dimensional electrophoresis and mass spectrometry. *Food Technology and Biotechnology*, *50*(3), 315–325.

Diniz, A. L., Zucchi, M. I., Santini, L., Benchimol-reis, L. L., Fungaro, M. H. P., Lucia, M., & Vieira, C. (2014). Nucleotide diversity based on phaseolin and iron reductase genes in common bean accessions of different geographical origins. *Genome, 57*(February) , 69–77.

Edy Susanto, M. (2019). Synthesis of the Phytoalexin (&)-Phaseollin: 3-Phenylthiochromans as Masked 2H-Chromenes and o-Prenyl Phenols. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699.

Emani, C., & Hall, T. C. (2014). Phaseolin: Structure and evolution. *The Open Evolution Journal*, *1*(1), 66–74.

Ersland, D. R., Brown, J. W. S., Casey, R., & Hall, T. C. (1983). The Storage proteins of *Phaseolus vulgaris* L., Vicia faba L. and Pisum sativum L. *Seed Proteins*, 355–375.

Escribano, M. R., Santalla, M., Casquero, P. A., & De Ron, A. M. (1998). Patterns of genetic diversity in landraces of common bean (*Phaseolus vulgaris* L.) from Galicia. *Plant Breeding*, *117*(1), 49–56.

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567.

Flores-Sosa, Á. R., Aquino-Bolaños, E. N., Cardador-Martínez, A., Chávez-Servia, J. L., Vera-Guzmán, A. M., Carrillo-Rodríguez, J. C., & Jiménez, J. E. A. (2020). Variation in protein and amino acids content among landraces of common bean (*Phaseolus vulgaris* L.). *Emirates Journal of Food and Agriculture*, *32*(10), 750–760.

Fuente, M. D. La, López-pedrouso, M., Alonso, J., Santalla, M., Ron, A. M. De, Álvarez, G., & Zapata, C. (2012). In-Depth characterization of the phaseolin protein diversity of common bean (*Phaseolus vulgaris* L.) Based on Two-dimensional electrophoresis and Mass Spectrometry. *Food Technology* and *Biotechnology*,50(3), 315–325.

Gichangi, A., Birachi, E., Wambua, S., Kavoi, J., Karanja, D., Muriithi, F., & Mutua, M. (2019). Factors influencing smallholder farmers' inputs use in major bean production corridors in Kenya. *Journal of Agriculture and Food Sciences*, *17*(1), 136.

García-Cordero, J. M., Martínez-Palma, N. Y., Madrigal-Bujaidar, E., Jiménez-Martínez, C., Madrigal-Santillán, E., Morales-González, J. A., Paniagua-Pérez, R., & Álvarez-González, I. (2021). Phaseolin, a protein from the seed of *phaseolus vulgaris*, has antioxidant, antigenotoxic, and chemopreventive properties. *Nutrients*, *13*(6).

Geiger, P. J., & Bessman, S. P. (1972). Protein determination by Lowry's method in the presence of sulfhydryl reagents. *Analytical Biochemistry*, *49*(2), 467–473.

Gepts, P., & Bliss, F. A. (1986). Phaseolin variability among wild and cultivated common beans (Phaseolus vulgaris) from Colombia. *Economic Botany*, *40*(4), 469–478.

Gepts, P., Osborn, T. C., Rashka, K., & Bliss, F. A. (1986). Phaseolin-protein Variability in Wild Forms and Landraces of the Common Bean (Phaseolus vulgaris): Evidence for Multiple Centers of Domestication. *Economic Botany*, *40*(4), 451–468.

Gepts, Paul. (1988). *Phaseolin as an evolutionary Marker*. *Genetic Resources of Phaseolus Beans*,215–241.

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*(Figure 1).

Guzmán Díaz, F. A., Toro Chica, O., Ocampo Nahar., C. H., Sánchez, I., Cárdenas, H., & Debouck, D. G. (2001). Observations about risks of genetic erosion and drift during multiplication and regeneration of germplasm, using wild common bean as model. *Bean Improvement Cooperative. Annual Report (USA)*, January, 44:29-30.

Henry, N. A., Evans, N. N., & Nelson, O. A. (2019). Genetic diversity of common bean (Phaseolus vulgaris L.) landraces from South Western Kenya for resistance to Pythium root rot disease. *African Journal of Biotechnology*, *18*(15), 316–324.

Johnson, W. C., Menéndez, C., Nodari, R., Epimaki, M. K., Magnusson, S., Singh, S. P., & Gepts, P. (1996). Association of a seed weight factor with the phaseolin seed storage protein locus across geno- types, environments, and genomes in *Phaseolus-vigna* spp.: Sax (1923) revisited. *Quantitative. Trait Loci*, *2*(1923), 000–000.

Johnson, Y., Kinama, J., Ayuke, F. O., & Sijali, I. V. (2018). Determining the water use efficiency (WUE) of drought tolerant common bean varieties Determining the water use efficiency (WUE) of drought tolerant common bean varieties Department of Land Resource Management and Agricultural Technology (LARMAT), De. *International Journal of Agronomy and Agricultural Research (IJAAR)ISSN: 2223-7054 (Print) 2225-3610 (Online) Http://Www.Innspub.Net*, *Vol. 12*(3), 35–39.

Kami, A., & Gepts, P. (1994). Phaseolin nucleotide sequence diversity in *Phaseolus*. L. Intraspecific diversity in *Phaseolus vulgaris*. *Genome,* 37: 751-757.

Katungi, E., Farrow, A., Chianu, J., Sperling, L., & Beebe, S. (2009a). Common bean in Eastern and Southern Africa: a situation and outlook analysis. June.

Katungi, E., Farrow, A., Chianu, J., Sperling, L., & Beebe, S. (2009b). International Centre for Tropical Agriculture: Common bean in Eastern and Southern Africa: a situation and outlook analysis. June, 1–44.

Kimani, P. M., & Warsame, A. H. M. E. D. (2019). Breeding second-generation bio fortified bean varieties for Africa. *Food and Energy Security*, *8*(4), 1–17.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *16*(2), 111–120.

Korneliussen, T. S., Moltke, I., Albrechtsen, A., & Nielsen, R. (2013). Calculation of Tajima' s D and other neutrality test statistics from low depth next-generation sequencing data. *BioMed Central Bioinformatics*, *2013*, 14:289.

Kulinskaya, E., & Lewin, A. (2009). Testing for linkage and Hardy-Weinberg disequilibrium. *Annals of Human Genetics, 73*(2), 253–262.

Kumar, V., Sharma, S., Kero, S., Sharma, S., Sharma, A. K., Kumar, M., & Venkataramana Bhat, K. (2008). Assessment of genetic diversity in common bean (*Phaseolus vulgaris* L.) germplasm using amplified fragment length polymorphism (AFLP). *Scientia Horticulturae*, *116*(2), 138–143.

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, *157*(1), 105–132.

Lioi, L., & Piergiovanni, A. R. (2013). European Common Bean. *Genetic and Genomic Resources of Grain Legume Improvement*, *1583*, 11–40.

Lioi, L., Zuluaga, D. L., & Pavan, S. (2019). *Genotyping-by-Sequencing Reveals Molecular Genetic Diversity in Italian Common Bean Landraces*. 1–14.

Loeuille, B., Thode, V., Siniscalchi, C., Andrade, S., Rossi, M., & Pirani, J. R. (2021). Extremely low nucleotide diversity among thirty-six new chloroplast genome sequences from Aldama (*Heliantheae*, *Asteraceae)* and comparative chloroplast genomics analyses with closely related genera. *PeerJ*, *9*.

Long, J., Zhang, J., Zhang, X., Wu, J., Chen, H., Wang, P., Wang, Q., & Du, C. (2020). Genetic Diversity of Common Bean (*Phaseolus vulgaris* L.) Germplasm Resources in Chongqing, Evidenced by Morphological Characterization. *Frontiers in Genetics*, *11*(July), 1–9.

Magrini, M. B., Anton, M., Cholez, C., Corre-Hellou, G., Duc, G., Jeuffroy, M. H., Meynard, J. M., Pelzer, E., Voisin, A. S., & Walrand, S. (2016). Why are grain-legumes rarely present in cropping systems despite their environmental and nutritional benefits? Analyzing lock-in in the French agrifood system. *Ecological Economics*, *126*, 152–162.

Mamidi, S., Rossi, M., Moghaddam, S. M., Annam, D., Lee, R., Papa, R., & McClean, P. E. (2013). Demographic factors shaped diversity in the two gene pools of wild common bean Phaseolus vulgaris L. *Heredity*, *110*(3), 267–276.

Mogesse, W., Zelleke, H., & Nigussie, M. (2020). General and Specific Combing Ability of Maize (*Zea mays* L.) Inbred line for grain yield and yield related traits using 8×8 Diallel Crosses. *American Journal of BioScience*, *8*(3), 45.

Mohammed, A. (2013). An Overview of Distribution, Biology and the Management of Common *Bean Anthracnose Plant Pathology & Microbiology*. *4,* 4–9.

Montoya, C. A., Lallès, J., Beebe, S., & Leterme, P. (2010). Phaseolin diversity as a possible strategy to improve the nutritional value of common beans (*Phaseolus vulgaris*). *Food Research International*, *43*(2), 443–449.

Montoya, C. A., Lalles, J., Beebe, S., Souffrant, W. B., Molle, D., Leterme, P., Montoya, C. A., Lalles, J., Beebe, S., Souffrant, W. B., & Molle, D. (2013). Susceptibility of phaseolin

(*Phaseolus vulgaris*) subunits to trypsinolysis and influence of dietary level of raw phaseolin on protein digestion in the small intestine of rats. *British Journal of Nutrition*, *101* (9), 1324-1332.

Morgil, H., Can Gercek, Y., & Tulum, I. (2020). Single Nucleotide Polymorphisms (SNPs) in Plant Genetics and Breeding. *The Recent Topics in Genetic Polymorphisms*, 1–12.

Moturu, T. R., Thula, S., Singh, R. K., Nodzyński, T., Vařeková, R. S., Friml, J., & Simon, S. (2018). Molecular evolution and diversification of the SMXL gene family. *Journal of Experimental Botany*, *69*(9), 2367–2378.

Mullins, A. P., & Arjmandi, B. H. (2021). Health benefits of plant-based nutrition: Focus on beans in cardiometabolic diseases. *Nutrients*, *13*(2), 1–16.

Margaret, N., Tenywa, J. S., Otabbong, E., Mubiru, D. N., & Basamba, T. A. (2014). Development of common bean (Phaseolus Vulgaris L.) production under low soil phosphorus and drought in Sub-Saharan Africa: A review. *Journal of Sustainable Development*, *7*(5), 128–139.

McLarney, M. J., Pellett, P. L., & Young, V. R. (1996). Pattern of amino acid requirements in humans: An interspecies comparison using published amino acid requirement recommendations. *Journal of Nutrition*, *126*(7), 1871–1882. Millward, D. J. (2015). Protein: Requirements. *Encyclopedia of Food and Health*, *March*, 538–547.

Nay, M. M., Souza, T. L. P. O., Raatz, B., Mukankusi, C. M., Pastor-Corrales, M. A., Abreu, A. F. B., & Melo, L. C. (2019). A review of angular leaf spot resistance in common bean.

Crop Science, *59*(4), 1376–1391.

Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., Hatipoğlu, R., Ahmad, F., Alsaleh, A., Labhane, N., Özkan, H., Chung, G., & 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.

Nadeem, M. A., Yeken, M. Z., Shahid, M. Q., Habyarimana, E., Yılmaz, H., Alsaleh, A., Hatipoğlu, R., Çilesiz, Y., Khawar, K. M., Ludidi, N., Ercişli, S., Aasim, M., Karaköy, T., & Baloch, F. S. (2021). Common bean as a potential crop for future food security: an overview of past, current and future contributions in genomics, transcriptomics, transgenics and proteomics. *Biotechnology and Biotechnological Equipment*, *35*(1), 758–786.

Oliveira, A. P., Andrade, G. F., Mateó, B. S. O., & Naozuka, J. (2017). Protein and Metalloprotein Distribution in Different Varieties of Beans (*Phaseolus vulgaris* L.): Effects of Cooking. *International Journal of Food Science*, *2017*.

Pace, N. R., Sapp, J., & Goldenfeld, N. (2012). Phylogeny and beyond: Scientific, historical, and conceptual significance of the first tree of life. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(4), 1011–1018.

Petrovski, S., Wang, Q., Heinzen, E. L., Allen, A. S., & Goldstein, D. B. (2013). Genic Intolerance to Functional Variation and the Interpretation of Personal Genomes. The Public Library of

Science *Genetics*, *9*(8).

Posada, D. (2009). Bioinformatics for DNA sequence analysis. Preface. *Methods in Molecular Biology*, *537*, 337–350.

Qureshi, A. M. I., Dar, Z. A., & Wani, S. H. (2019). Quality breeding in field crops. *Current Knowledge and Future Perspectives*, *5*(Chapter 5), 1–275.

Reddy, U. K., Nimmakayala, P., Abburi, V. L., Reddy, C. V. C. M., Saminathan, T., Percy, R. G., Yu, J. Z., Frelichowski, J., Udall, J. A., Page, J. T., Zhang, D., Shehzad, T., & Paterson,

H. (2017). Genome-wide divergence, haplotype distribution and population demographic histories for Gossypium hirsutum and Gossypium barbadense as revealed by genome- anchored SNPs. *Scientific Reports*, *7*(December 2016), 1–10.

Rodiño, A. P., Santalla, M., González, A. M., De Ron, A. M., & Singh, S. P. (2006). Novel genetic variation in common bean from the Iberian Peninsula. *Crop Science*, *46*(6), 2540– 2546.

Sathe, S. K. (2002). Dry bean protein functionality. *Critical Reviews in Biotechnology*, *22*(2), 175–223.

Schönfeldt, H. C., & Hall, N. G. (2012). Dietary protein quality and malnutrition in Africa. *British Journal of Nutrition*, *108*(SUPPL. 2)

Shewry, P. R., & Halford, N. G. (2002). Cereal seed storage proteins: Structures, properties and role in grain utilization. *Journal of Experimental Botany*, *53*(370), 947–958.

Slightom, J., Drong, R., & Klassy, R. (1985). *Nucleic Acids Research*. *13*(18), 47–48.

Talbot, D. R., Adang, M. J., Slightom, J. L., & Hall, T. C. (1984). Size and organization of a multigene family encoding phaseolin, the major seed storage protein of (*Phaseolus vulgaris* L.) *Molecular & General Genetics*, *198*(1), 42–49.

TC, K., M, J., & LY, A. (2018). Nutritional properties and nutrients chemical analysis of common beans seed. *MOJ Biology and Medicine*, *3*(2), 41–47.

Venkateshwaran, M., Volkening, J. D., Sussman, M. R., & Ané, J. M. (2013). Symbiosis and the social network of higher plants. *Current Opinion in Plant Biology*, *16*(1), 118–127.

Walker, J. M., Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Analysis Tools on the ExPASy Server 571: The Proteomics protocols handbook edited protein identification and analysis tools on the ExPASy Server. *The Proteomics Protocols Handbook*, 571–607. 571–607.

Xu, S., Yang, Z., Zhang, E., Jiang, Y., Pan, L., Chen, Q., Xie, Z., & Xu, C. (2014). Nucleotide diversity of maize ZmBT1 gene and association with starch physicochemical properties. *The Public Library of Science*, *9*(8).

Yachdav, G., Kloppmann, E., Kajan, L., Hecht, M., Goldberg, T., Hamp, T., Hönigschmid, P., Schafferhans, A., Roos, M., Bernhofer, M., Richter, L., Ashkenazy, H., Punta, M., Schlessinger, A., Bromberg, Y., Schneider, R., Vriend, G., Sander, C., Ben-Tal, N., & Rost, B. (2014). PredictProtein - An open resource for online prediction of protein structural and functional features. *Nucleic Acids Research*, *42*(W1), 337–343.

Yin, S. W., Tang, C. H., Wen, Q. B., & Yang, X. Q. (2010). Functional and conformational properties of phaseolin (*Phaseolus vulgris* L.) and kidney bean protein isolate: A comparative study. *Journal of the Science of Food and Agriculture*, *90*(4), 599–607.

Yildiz, S., Awale, H., & Kelly, J. (2017). *Determination of Phaseolin Types in Common Bean (Phaseolus vulgaris)* Varieties from Turkey. *Greener Journal of Agricultural Sciences 4*(March 2014), 39–45.

Zheng, Z., Sumi, K., Tanaka, K., & Murai, N. (1995). The bean seed storage protein β-phaseolin is synthesized, processed, and accumulated in the vacuolar type-11 protein bodies of transgenic rice endosperm. *Plant Physiology*, *109*(3), 777–786.

Zhu, C., Gore, M., Buckler, E. S., & Yu, J. (2008). Status and Prospects of Association Mapping in Plants. *The Plant Genome*, *1*(1), 5–20.

APPENDICES

APPENDIX 1: DNA extraction reagents preparation.

Genomic DNA extraction reagents that were prepared included: CTAB extraction buffer by dissolving 100 ml of 1 M Tris HCl (pH 8.0), 280 ml of 5 M Sodium chloride, 40 ml of 0.5 M EDTA, 20 g of CTAB (cetyltrimethyl ammonium bromide) in 1000 ml double distilled sterile water. One (1) M Tris HCl (pH8.0) was prepared by dissolving 121.1 g of Tris base first in 700 ml of distilled H_2O; followed by concentrated HCl to adjust its pH to 8.0; before adjusting to a total volume of 1 liter with distilled water. The 0.5 M EDTA was prepared by dissolving 86.12 g EDTA in 700 ml distilled water first, followed by adding 18 g of sodium hydroxide (NaOH) pellets, to adjust its pH to 8 (optimum pH for EDTA dissolution), and volume adjusted to 1 liter with distilled water. The 5 M NaCl was prepared by dissolving 292.2 g of NaCl first in 700 ml distilled water; and volume adjusted to 1 liter with distilled water. Other reagents for DNA extraction procedure included 10 % SDS, RNase A, 70 % Ethanol, proteinase -k, CTAB/NaCl solution; Isopropanol, Chloroform- Isoamyl alcohol, and 10 x TBE buffer prepared in 1-liter volume by dissolving 55 g Boric, 108 g Tris base Hcl, 9.3 g EDTA mixed in a conical flask and heated on a hot plate to boiling to enhance complete dissolution of the mixture of reagents.

APPENDIX 2 : Representation of multiple sequence alignment for all consensus sequences.

```
                              10        20        30        40        50        60        70        80
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|.........|.
complete    reference1       CCCCAACCAAGATGAACACTCACTTGACATTTCCATGAAAAGATTCCCTTGGTCTGTGCAGGAAGTGTAATGTCGATCACC
Mbere2                    ------------------------------------------------------------------------TCGATCACC
Mbere3                    ------------------------------------------------------------------------TCGATCACC
unknown                   ------------------------------------------------------------------------TCGATCACC
23                        ------------------------------------------------------------------------TCGATCACC
Kidneybean                ------------------------------------------------------------------------TCGATCACC
KATB1                     ------------------------------------------------------------------------TCGATCACC
Ekebure                                                                                           TCGATCACC
19                                                                                                TCGATCACC
Mbere4                    ------------------------------------------------------------------------TCGATCACC
Mbere1                    ------------------------------------------------------------------------TCGATCACC
VAX4                      ------------------------------------------------------------------------TCGATCACC
                          ------------------------------------------------------------------------
                          ------------------------------------------------------------------------
```

```
                              220       230       240       250       260       270       280       290
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|.........|.
complete    reference1       CCCTCCAAAACCAACACGGACACTCACTTGACATTTCAGAAAAAATTTCCCTGGGTATGTGCAGGAAGTCTAATGTGGGTC
Mbere2                    CCCTCCAAAACCAACACGGACACTCACTTGACATTTCAGAAAAAATTTCCCTGGGTATGTGCAGGAAGTCTAATGTGGGTC
Mbere3                    CCCTCCAAAACCAACACGGACACTCACTTGACATTTCAGAAAAAATTTCCCTGGGTATGTGCAGGAAGTCTAATGTGGGTC
unknown                   CCCTCCAAAACCAACACGGACACTCACTTGACATTTCAGAAAAAATTTCCCTGGGTATGTGCAGGAAGTCTAATGTGGGTC
23                        CCCTCCAAAACCAACACGGACACTCACTTGACATTTCAGAAAAAATTTCCCTGGGTATGTGCAGGAAGTCTAATGTGGGTC
Kidneybean                CCCTCCAAAACCAACACGGACACTCACTTGACATTTCAGAAAAAATTTCCCTGGGTATGTGCAGGAAGTCTAATGTGGGTC
KATB1                     CCCTCCAAAACCAACACGGACACTCACTTGACATTTCAGAAAAAATTTCCCTGGGTATGTGCAGGAAGTCTAATGTGGGTC
Ekebure                   CCCTCCAAAACCAACACGGACACTCACTTGACATTTCAGAAAAAATTTCCCTGGGTATGTGCAGGAAGTCTAATGTGGGTC
19                        CCCTCCAAAACCAACACGGACACTCACTTGACATTTCAGAAAAAATTTCCCTGGGTATGTGCAGGAAGTCTAATGTGGGTC
Mbere4                    CCCTCCAAAACCAACACGGACACTCACTTGACATTTCAGAAAAAATTTCCCTGGGTATGTGCAGGAAGTCTAATGTGGGTC
Mbere1                    CCCTCCAAAACCAACACGGACACTCACTTGACATTTCAGAAAAAATTTCCCTGGGTATGTGCAGGAAGTCTAATGTGGGTC
VAX4                      CCCTCCAAAACCAACACGGACACTCACTTGACATTTCAGAAAAAATTTCCCTGGGTATGTGCAGGAAGTCTAATGTGGGTC
```

```
                              430       440       450       460       470       480       490       500
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|.........|.
complete    reference1       TGCCCCACCCAACCAAAATAGCTTCAGAAAATGTATTTTTTTTGAAGAAAAAAAAAAACTAAAAAAGCATACATAGTTACAC
Mbere2                    TGCCCCACCCAACCAAAATAGCTTCAGAAAATGTATTTTTTTTGAAGAAAAAAAAAAACTAAAAAAGCATACATAGTTACAC
Mbere3                    TGCCCCACCCAACCAAAATAGCTTCAGAAAATGTATTTTTTTTGAAGAAAAAAAAAAACTAAAAAAGCATACATAGTTACAC
unknown                   TGCCCCACCCAACCAAAATAGCTTCAGAAAATGTATTTTTTTTGAAGAAAAAAAAAAACTAAAAAAGCATACATAGTTACAC
23                        TGCCCCACCCAACCAAAATAGCTTCAGAAAATGTATTTTTTTTGAAGAAAAAAAAAAACTAAAAAAGCATACATAGTTACAC
Kidneybean                TGCCCCACCCAACCAAAATAGCTTCAGAAAATGTATTTTTTTTGGAGAAAAAAAAAAACTAAAAAAGCATACATAGTTACAC
KATB1                     TGCCCCACCCAACCAAAATAGCTTCAGAAAATGTATTTTTTTTGAAGAAAAAAAAAATCTTAAAAAGCATACATAGTTACAC
Ekebure                   TGCCCCACCCAACCAAAATAGCTTCAGAAAATGTATTTTTTTTGAAGAAAAAAAAAAACTAAAAAAGCATACATAGTTACAC
19                        TGCCCCACCCAACCAAAATAGCTTCAGAAAATGTATTTTTTTTGAAGAAAAAAAAAAACTAAAAAAGCATACATAGTTACAC
Mbere4                    TGCCCCACCCAACCAAAATAGCTTCAGAAAATGTATTTTTTTTGAAGAAAAAAAAATACTAAAAAAGCATACATAGTTACAC
Mbere1                    TGCCCCACCCAACCAAAATAGCTTCAGAAAATGTATTTTTTTTGGAGAAAAAAAAATACTAAAAAAGCATACATAGTTACAC
VAX4                      TGCCCCACCCAACCAAAAAAGCTTCAGAAAATGTATTTTTTTTGAAGAAAAAAAAAAACTAAAAAAGCATACATAGTTACCC
```

```
                              640       650       660       670       680       690       700       710
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|.........|.
complete    reference1       ACTCATACAAGGGTCAACACATGACAAATAACACAGATCACCCAATACAAAGCACAATATGAATGTTTCACTTAATAAAGA
Mbere2                    ACTCATACAAGGGTCAACACATGACAAATAACACAGATCACCCAATACAAAGCACAATATGAATGTTTCACTTAATAAAGA
Mbere3                    ACTCATACAAGGGTCAACACATGACAAATAACACAGATCACCCAATACAAAGCACAATATGAATGTTTCACTTAATAAAGA
unknown                   ACTCATACAAGGGTCAACACATGACAAATAACACAGATCACCCAATACAAAGCACAATATGAATGTTTCACTTAATAAAGA
23                        ACTCATACAAGGGTCAACACATGACAAATAACACAGATCACCCAATACAAAGCACAATATGAATGTTTCACTTAATAAAGA
Kidneybean                ACTCATACAAGGGTCAACACATGACAAATAACACAGATCACCCAATACAAAGCACAATATGAATGTTTCACTTAATAAAGA
KATB1                     ACTCATACAAGGGTCAACACATGACAAATAACACAGATCACCCAATACAAAGCACAATATGAATGTTTCACTTAATAAAGA
Ekebure                   ACTCATACAAGGGTCAACACATGACAAATAACACAGATCACCCAATACAAAGCACAATATGAATGTTTCACTTAATAAAGA
19                        ACTCATACAAGGGTCAACACATGACAAATAACACAGATCACCCAATACAAAGCACAATATGAATGTTTCACTTAATAAAGA
Mbere4                    ACTCATACAAGGGTCAACACATGACAAATAACACAGATCACCCAATACAAAGCACAATATGAATGTTTCACTTAATAAAGA
Mbere1                    ACTCATACAAGGGTCAACACATGACAAATAACACAGATCACCCAATACAAAGCACAATATGAATGTTTCACTTAATAAAGA
VAX4                      ACTCATACAAGGGTCAACACATGACAAATAACACAGATCACCCAATACAAAGCACAATATGAATGTTTCACTTAATAAAGA
```

```
                              850       860       870       880       890       900       910       920
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|.........|.
complete    reference1       AAAATAAAAAAGAATGCGCACACACTTGCAACATCCCATGGAGAATTAAGGGGAACAAAAACGGAACGAACAAAAAATGAA
Mbere2                    AAAATAAAAAAGAATGCGCACACACTTGCAACATCCCATGGAGAATTAAGGGGAACAAAAACGGAACGAACAAAAAATGAA
Mbere3                    AAAATAAAAAAGAATGCGCACACACTTGCAACATCCCATGGAGAATTAAGGGGAACAAAAACGGAACGAACAAAAAATGAA
unknown                   AAAATAAAAAAGAATGCGCACACACTTGCAACATCCCATGGAGAATTAAGGGGAACAAAAACGGAACGAACAAAAAATGAA
23                        AAAATAAAAAAGAATGCGCACACACTTGCAACATCCCATGGAGAATTAAGGGGAACAAAAACGGAACGAACAAAAAATGAA
Kidneybean                AAAATAAAGAAGAATGCGCACACACTTGCAATATCCCATGGAGAATTAAGGGGAACAAAAACGGAACGAACAAAAAATGAA
KATB1                     AAAATAAAAAAGAATGCGCACACACTTGCAATATCCCATGGAGAATTAAGGGGAACAAAAACGGAACGAACAAAAAATGAA
Ekebure                   AAAATAAAAAAGAATGCGCACACACTTGCAACATCCCATGGAGAATTAAGGGGAACAAAAACGGAACGAACAAAAAATGAA
19                        AAAATAAAAAAGAATGCGCACACACTTGCAACATCCCATGGAGAATTAAGGGGAACAAAAACGGAACGAACAAAAAATGAA
Mbere4                    AAAATAAAAAAGAATGCGCACACACTTGCAATATCCCATGGAGAATTAAGGGGAACAAAAACGGAACGAACAAAAAATGAA
Mbere1                    AAAATAAAAAAGAATGCGCACACACTTGCAATATCCCATGGAGAATTAAGGGGAACAAAAACGGAACGAACAAAAAATGAA
```

APPENDIX 3: Mutations registered across common bean accessions.

Sample ID	Mutated Loci	Mutated bases	Codon Changes
Mbere1	359	C>T	Ser120Leu
Mbere4	359	A>C	Ser120Leu
unknown	419	A>C	Ser120Leu
VAX4	439	A>C	STP147Lys
Mbere1	465	A>C	STP155Trp
Kidneybean	465	A>C	STP155Trp
Mbere1	476	A>C	Lys159Ile
Mbere4	476	A>C	Lys159Ile
KATB1	477	A>C	Lys159Asn
KATB1	480	A>C	Leu160Leu
VAX4	500	A>C	His167Pro
Mbere1	543	A>G	Ser181Ser
Kidneybean	543	A>G	Ser181Ser
Mbere4	543	A>G	Ser181Ser
Mbere2	593	A>G	Arg198Lys
Mbere2	598	A>G	Leu200Leu
Mbere4	759	A>G	Thr253Thr
Mbere1	806	A>G	Gly269Glu
Kidneybean	806	A>T	Gly269Glu
Mbere4	806	A>T	Gly269Glu
Mbere1	815	A>T	Arg272Lys
Kidneybean	815	A>T	Arg272Lys
KATB1	815	A>T	Arg272Lys
Mbere4	815	A>T	Arg272Lys
Kidneybean	849	A>T	Lys283Lys
Mbere1	872	A>T	Thr291Ile
Kidneybean	872	A>T	Thr291Ile
KATB1	872	A>T	Thr291Ile
Mbere4	872	A>T	Thr291Ile
Mbere1	925	A>T	Lys309STP
Mbere1	974	A>T	Lys325Arg
unknown	974	A>T	Lys325Arg
Kidneybean	974	A>T	Lys325Arg
KATB1	974	C>A	Lys325Arg
VAX4	974	C>A	Lys325Arg
unknown	982	C>A	Asn328His
Mbere1	1051	C>A	Pro351Ser
Kidneybean	1051	C>A	Pro351Ser

KATB1	1051	C>A	Pro351Ser
Mbere1	1124	C>A	Ser375STP
unknown	1124	C>T	Ser375STP
Kidneybean	1124	C>T	Ser375STP
KATB1	1124	C>T	Ser375STP
Ekebure	1234	C>T	Tyr412Asn
unknown	1334	C>T	Pro445Leu
Mbere1	1539	C>T	STP513Cys
unknown	1539	C>T	STP513Cys
Kidneybean	1539	C>T	STP513Cys
KATB1	1539	C>T	STP513Cys
Mbere4	1539	C>T	STP513Cys
Mbere1	2045	C>T	Asn682Ile
unknown	2045	C>T	Asn682Ile
Kidneybean	2045	C>T	Asn682Ile
Mbere4	2045	C>T	Asn682Ile
Mbere1	2129	C>T	STP710Ser
unknown	2129	C>T	STP710Ser
Kidneybean	2129	G>A	STP710Ser
Mbere4	2129	G>A	STP710Ser
Kidneybean	2214	G>A	Pro738Pro
Mbere2	2313	G>A	Leu771Leu
unknown	2566	G>A	Pro856Lys
unknown	2567	G>A	Pro856Lys
Mbere1	2759	G>A	Gly920Ala
Mbere3	2796	G>A	Leu932Leu
Mbere3	2828	G>A	Ile943Thr
VAX4	2872	G>A	Pro958Thr
Mbere1	2878	G>A	Phe960Leu
Mbere1	3041	G>A	Pro1014Leu
Mbere3	3041	G>C	Pro1014Leu
unknown	3041	G>C	Pro1014Leu
19	3041	G>C	Pro1014Leu
Mbere2	3216	G>C	STP1072Cys
Kidneybean	3216	G>C	STP1072Cys
KATB1	3216	G>C	STP1072Cys
unknown	3400	T>A	Gly1134Arg
23	3400	T>A	Gly1134Arg
Kidneybean	3400	T>A	Gly1134Arg
23	3508	T>C	Leu1170Ile
Kidneybean	3766	T>C	Asn1256His
Kidneybean	3846	T>C	Cys1282Cys

APPENDIX 4: Linkage disequilibrium output on the entire gene.

Site1	Site2	Dist	D	D'	R	Fisher
359	419	60	-0.014	-1.000	-0.135	1.000
359	439	80	-0.014	-1.000	-0.135	1.000
359	465	106	0.056	0.400	0.400	0.318
359	476	117	0.139	1.000	1.000	0.015*
359	477	118	-0.014	-1.000	-0.135	1.000
359	480	121	-0.014	-1.000	-0.135	1.000
359	500	141	-0.014	-1.000	-0.135	1.000
359	543	184	0.125	1.000	0.775	0.045
359	593	234	-0.014	-1.000	-0.135	1.000
359	598	239	-0.014	-1.000	-0.135	1.000
359	759	400	0.069	1.000	0.674	0.167
359	806	447	0.125	1.000	0.775	0.045
359	815	456	0.111	1.000	0.632	0.091
359	849	490	-0.014	-1.000	-0.135	1.000
359	872	513	0.111	1.000	0.632	0.091
359	925	566	0.069	1.000	0.674	0.167
359	974	615	0.014	0.143	0.076	1.000
359	982	623	-0.014	-1.000	-0.135	1.000
359	1051	692	0.042	0.333	0.258	0.455
359	1124	765	0.028	0.250	0.158	1.000
359	1234	875	-0.014	-1.000	-0.135	1.000
359	1334	975	-0.014	-1.000	-0.135	1.000
359	1539	1180	0.097	1.000	0.529	0.152
359	2045	1686	0.111	1.000	0.632	0.091
359	2129	1770	0.111	1.000	0.632	0.091
359	2214	1855	-0.014	-1.000	-0.135	1.000
359	2313	1954	-0.028	-1.000	-0.200	1.000
359	2566	2207	-0.014	-1.000	-0.135	1.000
359	2567	2208	-0.014	-1.000	-0.135	1.000
359	2759	2400	0.069	1.000	0.674	0.167
359	2796	2437	-0.014	-1.000	-0.135	1.000
359	2828	2469	-0.014	-1.000	-0.135	1.000
359	2872	2513	-0.014	-1.000	-0.135	1.000
359	2878	2519	0.069	1.000	0.674	0.167
359	3041	2682	0.028	0.250	0.158	1.000
359	3216	2857	-0.056	-1.000	-0.316	0.515
359	3400	3041	-0.042	-1.000	-0.258	1.000
359	3508	3149	-0.014	-1.000	-0.135	1.000
359	3766	3407	-0.014	-1.000	-0.135	1.000

359	3846	3487	-0.014	-1.000	-0.135	1.000
419	439	20	-0.007	-1.000	-0.091	1.000
419	465	46	-0.014	-1.000	-0.135	1.000
419	476	57	-0.014	-1.000	-0.135	1.000
419	477	58	-0.007	-1.000	-0.091	1.000
419	480	61	-0.007	-1.000	-0.091	1.000
419	500	81	-0.007	-1.000	-0.091	1.000
419	543	124	-0.021	-1.000	-0.174	1.000
419	593	174	-0.007	-1.000	-0.091	1.000
419	598	179	-0.007	-1.000	-0.091	1.000
419	759	340	-0.007	-1.000	-0.091	1.000
419	806	387	-0.021	-1.000	-0.174	1.000
419	815	396	-0.028	-1.000	-0.213	1.000
419	849	430	-0.007	-1.000	-0.091	1.000
419	872	453	-0.028	-1.000	-0.213	1.000
419	925	506	-0.007	-1.000	-0.091	1.000
419	974	555	0.049	1.000	0.357	0.417
419	982	563	0.076	1.000	1.000	0.083
419	1051	632	-0.021	-1.000	-0.174	1.000
419	1124	705	0.056	1.000	0.426	0.333
419	1234	815	-0.007	-1.000	-0.091	1.000
419	1334	915	0.076	1.000	1.000	0.083
419	1539	1120	0.049	1.000	0.357	0.417
419	2045	1626	0.056	1.000	0.426	0.333
419	2129	1710	0.056	1.000	0.426	0.333
419	2214	1795	-0.007	-1.000	-0.091	1.000
419	2313	1894	-0.014	-1.000	-0.135	1.000
419	2566	2147	0.076	1.000	1.000	0.083
419	2567	2148	0.076	1.000	1.000	0.083
419	2759	2340	-0.007	-1.000	-0.091	1.000
419	2796	2377	-0.007	-1.000	-0.091	1.000
419	2828	2409	-0.007	-1.000	-0.091	1.000
419	2872	2453	-0.007	-1.000	-0.091	1.000
419	2878	2459	-0.007	-1.000	-0.091	1.000
419	3041	2622	0.056	1.000	0.426	0.333
419	3216	2797	-0.028	-1.000	-0.213	1.000
419	3400	2981	0.063	1.000	0.522	0.250
419	3508	3089	-0.007	-1.000	-0.091	1.000
419	3766	3347	-0.007	-1.000	-0.091	1.000
419	3846	3427	-0.007	-1.000	-0.091	1.000
439	465	26	-0.014	-1.000	-0.135	1.000
439	476	37	-0.014	-1.000	-0.135	1.000

439	477	38	-0.007	-1.000	-0.091	1.000
439	480	41	-0.007	-1.000	-0.091	1.000
439	500	61	0.076	1.000	1.000	0.083
439	543	104	-0.021	-1.000	-0.174	1.000
439	593	154	-0.007	-1.000	-0.091	1.000
439	598	159	-0.007	-1.000	-0.091	1.000
439	759	320	-0.007	-1.000	-0.091	1.000
439	806	367	-0.021	-1.000	-0.174	1.000
439	815	376	-0.028	-1.000	-0.213	1.000
439	849	410	-0.007	-1.000	-0.091	1.000
439	872	433	-0.028	-1.000	-0.213	1.000
439	925	486	-0.007	-1.000	-0.091	1.000
439	974	535	0.049	1.000	0.357	0.417
439	982	543	-0.007	-1.000	-0.091	1.000
439	1051	612	-0.021	-1.000	-0.174	1.000
439	1124	685	-0.028	-1.000	-0.213	1.000
439	1234	795	-0.007	-1.000	-0.091	1.000
439	1334	895	-0.007	-1.000	-0.091	1.000
439	1539	1100	-0.035	-1.000	-0.255	1.000
439	2045	1606	-0.028	-1.000	-0.213	1.000
439	2129	1690	-0.028	-1.000	-0.213	1.000
439	2214	1775	-0.007	-1.000	-0.091	1.000
439	2313	1874	-0.014	-1.000	-0.135	1.000
439	2566	2127	-0.007	-1.000	-0.091	1.000
439	2567	2128	-0.007	-1.000	-0.091	1.000
439	2759	2320	-0.007	-1.000	-0.091	1.000
439	2796	2357	-0.007	-1.000	-0.091	1.000
439	2828	2389	-0.007	-1.000	-0.091	1.000
439	2872	2433	0.076	1.000	1.000	0.083
439	2878	2439	-0.007	-1.000	-0.091	1.000
439	3041	2602	-0.028	-1.000	-0.213	1.000
439	3216	2777	-0.028	-1.000	-0.213	1.000
439	3400	2961	-0.021	-1.000	-0.174	1.000
439	3508	3069	-0.007	-1.000	-0.091	1.000
439	3766	3327	-0.007	-1.000	-0.091	1.000
439	3846	3407	-0.007	-1.000	-0.091	1.000
465	476	11	0.056	0.400	0.400	0.318
465	477	12	-0.014	-1.000	-0.135	1.000
465	480	15	-0.014	-1.000	-0.135	1.000
465	500	35	-0.014	-1.000	-0.135	1.000
465	543	78	0.125	1.000	0.775	0.045*
465	593	128	-0.014	-1.000	-0.135	1.000

465	598	133	-0.014	-1.000	-0.135	1.000
465	759	294	-0.014	-1.000	-0.135	1.000
465	806	341	0.125	1.000	0.775	0.045*
465	815	350	0.111	1.000	0.632	0.091
465	849	384	0.069	1.000	0.674	0.167
465	872	407	0.111	1.000	0.632	0.091
465	925	460	0.069	1.000	0.674	0.167
465	974	509	0.097	1.000	0.529	0.152
465	982	517	-0.014	-1.000	-0.135	1.000
465	1051	586	0.125	1.000	0.775	0.045*
465	1124	659	0.111	1.000	0.632	0.091
465	1234	769	-0.014	-1.000	-0.135	1.000
465	1334	869	-0.014	-1.000	-0.135	1.000
465	1539	1074	0.097	1.000	0.529	0.152
465	2045	1580	0.111	1.000	0.632	0.091
465	2129	1664	0.111	1.000	0.632	0.091
465	2214	1749	0.069	1.000	0.674	0.167
465	2313	1848	-0.028	-1.000	-0.200	1.000
465	2566	2101	-0.014	-1.000	-0.135	1.000
465	2567	2102	-0.014	-1.000	-0.135	1.000
465	2759	2294	0.069	1.000	0.674	0.167
465	2796	2331	-0.014	-1.000	-0.135	1.000
465	2828	2363	-0.014	-1.000	-0.135	1.000
465	2872	2407	-0.014	-1.000	-0.135	1.000
465	2878	2413	0.069	1.000	0.674	0.167
465	3041	2576	0.028	0.250	0.158	1.000
465	3216	2751	0.028	0.250	0.158	1.000
465	3400	2935	0.042	0.333	0.258	0.455
465	3508	3043	-0.014	-1.000	-0.135	1.000
465	3766	3301	0.069	1.000	0.674	0.167
465	3846	3381	0.069	1.000	0.674	0.167
476	477	1	-0.014	-1.000	-0.135	1.000
476	480	4	-0.014	-1.000	-0.135	1.000
476	500	24	-0.014	-1.000	-0.135	1.000
476	543	67	0.125	1.000	0.775	0.045*
476	593	117	-0.014	-1.000	-0.135	1.000
476	598	122	-0.014	-1.000	-0.135	1.000
476	759	283	0.069	1.000	0.674	0.167
476	806	330	0.125	1.000	0.775	0.045*
476	815	339	0.111	1.000	0.632	0.091
476	849	373	-0.014	-1.000	-0.135	1.000
476	872	396	0.111	1.000	0.632	0.091

476	925	449	0.069	1.000	0.674	0.167
476	974	498	0.014	0.143	0.076	1.000
476	982	506	-0.014	-1.000	-0.135	1.000
476	1051	575	0.042	0.333	0.258	0.455
476	1124	648	0.028	0.250	0.158	1.000
476	1234	758	-0.014	-1.000	-0.135	1.000
476	1334	858	-0.014	-1.000	-0.135	1.000
476	1539	1063	0.097	1.000	0.529	0.152
476	2045	1569	0.111	1.000	0.632	0.091
476	2129	1653	0.111	1.000	0.632	0.091
476	2214	1738	-0.014	-1.000	-0.135	1.000
476	2313	1837	-0.028	-1.000	-0.200	1.000
476	2566	2090	-0.014	-1.000	-0.135	1.000
476	2567	2091	-0.014	-1.000	-0.135	1.000
476	2759	2283	0.069	1.000	0.674	0.167
476	2796	2320	-0.014	-1.000	-0.135	1.000
476	2828	2352	-0.014	-1.000	-0.135	1.000
476	2872	2396	-0.014	-1.000	-0.135	1.000
476	2878	2402	0.069	1.000	0.674	0.167
476	3041	2565	0.028	0.250	0.158	1.000
476	3216	2740	-0.056	-1.000	-0.316	0.515
476	3400	2924	-0.042	-1.000	-0.258	1.000
476	3508	3032	-0.014	-1.000	-0.135	1.000
476	3766	3290	-0.014	-1.000	-0.135	1.000
476	3846	3370	-0.014	-1.000	-0.135	1.000
477	480	3	0.076	1.000	1.000	0.083
477	500	23	-0.007	-1.000	-0.091	1.000
477	543	66	-0.021	-1.000	-0.174	1.000
477	593	116	-0.007	-1.000	-0.091	1.000
477	598	121	-0.007	-1.000	-0.091	1.000
477	759	282	-0.007	-1.000	-0.091	1.000
477	806	329	-0.021	-1.000	-0.174	1.000
477	815	338	0.056	1.000	0.426	0.333
477	849	372	-0.007	-1.000	-0.091	1.000
477	872	395	0.056	1.000	0.426	0.333
477	925	448	-0.007	-1.000	-0.091	1.000
477	974	497	0.049	1.000	0.357	0.417
477	982	505	-0.007	-1.000	-0.091	1.000
477	1051	574	0.063	1.000	0.522	0.250
477	1124	647	0.056	1.000	0.426	0.333
477	1234	757	-0.007	-1.000	-0.091	1.000
477	1334	857	-0.007	-1.000	-0.091	1.000

477	1539	1062	0.049	1.000	0.357	0.417
477	2045	1568	-0.028	-1.000	-0.213	1.000
477	2129	1652	-0.028	-1.000	-0.213	1.000
477	2214	1737	-0.007	-1.000	-0.091	1.000
477	2313	1836	-0.014	-1.000	-0.135	1.000
477	2566	2089	-0.007	-1.000	-0.091	1.000
477	2567	2090	-0.007	-1.000	-0.091	1.000
477	2759	2282	-0.007	-1.000	-0.091	1.000
477	2796	2319	-0.007	-1.000	-0.091	1.000
477	2828	2351	-0.007	-1.000	-0.091	1.000
477	2872	2395	-0.007	-1.000	-0.091	1.000
477	2878	2401	-0.007	-1.000	-0.091	1.000
477	3041	2564	-0.028	-1.000	-0.213	1.000
477	3216	2739	0.056	1.000	0.426	0.333
477	3400	2923	-0.021	-1.000	-0.174	1.000
477	3508	3031	-0.007	-1.000	-0.091	1.000
477	3766	3289	-0.007	-1.000	-0.091	1.000
477	3846	3369	-0.007	-1.000	-0.091	1.000
480	500	20	-0.007	-1.000	-0.091	1.000
480	543	63	-0.021	-1.000	-0.174	1.000
480	593	113	-0.007	-1.000	-0.091	1.000
480	598	118	-0.007	-1.000	-0.091	1.000
480	759	279	-0.007	-1.000	-0.091	1.000
480	806	326	-0.021	-1.000	-0.174	1.000
480	815	335	0.056	1.000	0.426	0.333
480	849	369	-0.007	-1.000	-0.091	1.000
480	872	392	0.056	1.000	0.426	0.333
480	925	445	-0.007	-1.000	-0.091	1.000
480	974	494	0.049	1.000	0.357	0.417
480	982	502	-0.007	-1.000	-0.091	1.000
480	1051	571	0.063	1.000	0.522	0.250
480	1124	644	0.056	1.000	0.426	0.333
480	1234	754	-0.007	-1.000	-0.091	1.000
480	1334	854	-0.007	-1.000	-0.091	1.000
480	1539	1059	0.049	1.000	0.357	0.417
480	2045	1565	-0.028	-1.000	-0.213	1.000
480	2129	1649	-0.028	-1.000	-0.213	1.000
480	2214	1734	-0.007	-1.000	-0.091	1.000
480	2313	1833	-0.014	-1.000	-0.135	1.000
480	2566	2086	-0.007	-1.000	-0.091	1.000
480	2567	2087	-0.007	-1.000	-0.091	1.000
480	2759	2279	-0.007	-1.000	-0.091	1.000

480	2796	2316	-0.007	-1.000	-0.091	1.000
480	2828	2348	-0.007	-1.000	-0.091	1.000
480	2872	2392	-0.007	-1.000	-0.091	1.000
480	2878	2398	-0.007	-1.000	-0.091	1.000
480	3041	2561	-0.028	-1.000	-0.213	1.000
480	3216	2736	0.056	1.000	0.426	0.333
480	3400	2920	-0.021	-1.000	-0.174	1.000
480	3508	3028	-0.007	-1.000	-0.091	1.000
480	3766	3286	-0.007	-1.000	-0.091	1.000
480	3846	3366	-0.007	-1.000	-0.091	1.000
500	543	43	-0.021	-1.000	-0.174	1.000
500	593	93	-0.007	-1.000	-0.091	1.000
500	598	98	-0.007	-1.000	-0.091	1.000
500	759	259	-0.007	-1.000	-0.091	1.000
500	806	306	-0.021	-1.000	-0.174	1.000
500	815	315	-0.028	-1.000	-0.213	1.000
500	849	349	-0.007	-1.000	-0.091	1.000
500	872	372	-0.028	-1.000	-0.213	1.000
500	925	425	-0.007	-1.000	-0.091	1.000
500	974	474	0.049	1.000	0.357	0.417
500	982	482	-0.007	-1.000	-0.091	1.000
500	1051	551	-0.021	-1.000	-0.174	1.000
500	1124	624	-0.028	-1.000	-0.213	1.000
500	1234	734	-0.007	-1.000	-0.091	1.000
500	1334	834	-0.007	-1.000	-0.091	1.000
500	1539	1039	-0.035	-1.000	-0.255	1.000
500	2045	1545	-0.028	-1.000	-0.213	1.000
500	2129	1629	-0.028	-1.000	-0.213	1.000
500	2214	1714	-0.007	-1.000	-0.091	1.000
500	2313	1813	-0.014	-1.000	-0.135	1.000
500	2566	2066	-0.007	-1.000	-0.091	1.000
500	2567	2067	-0.007	-1.000	-0.091	1.000
500	2759	2259	-0.007	-1.000	-0.091	1.000
500	2796	2296	-0.007	-1.000	-0.091	1.000
500	2828	2328	-0.007	-1.000	-0.091	1.000
500	2872	2372	0.076	1.000	1.000	0.083
500	2878	2378	-0.007	-1.000	-0.091	1.000
500	3041	2541	-0.028	-1.000	-0.213	1.000
500	3216	2716	-0.028	-1.000	-0.213	1.000
500	3400	2900	-0.021	-1.000	-0.174	1.000
500	3508	3008	-0.007	-1.000	-0.091	1.000
500	3766	3266	-0.007	-1.000	-0.091	1.000

500	3846	3346	-0.007	-1.000	-0.091	1.000
543	593	50	-0.021	-1.000	-0.174	1.000
543	598	55	-0.021	-1.000	-0.174	1.000
543	759	216	0.063	1.000	0.522	0.250
543	806	263	0.188	1.000	1.000	0.005**
543	815	272	0.167	1.000	0.816	0.018*
543	849	306	0.063	1.000	0.522	0.250
543	872	329	0.167	1.000	0.816	0.018*
543	925	382	0.063	1.000	0.522	0.250
543	974	431	0.063	0.429	0.293	0.523
543	982	439	-0.021	-1.000	-0.174	1.000
543	1051	508	0.104	0.556	0.556	0.127
543	1124	581	0.083	0.500	0.408	0.236
543	1234	691	-0.021	-1.000	-0.174	1.000
543	1334	791	-0.021	-1.000	-0.174	1.000
543	1539	996	0.146	1.000	0.683	0.045*
543	2045	1502	0.167	1.000	0.816	0.018*
543	2129	1586	0.167	1.000	0.816	0.018*
543	2214	1671	0.063	1.000	0.522	0.250
543	2313	1770	-0.042	-1.000	-0.258	1.000
543	2566	2023	-0.021	-1.000	-0.174	1.000
543	2567	2024	-0.021	-1.000	-0.174	1.000
543	2759	2216	0.063	1.000	0.522	0.250
543	2796	2253	-0.021	-1.000	-0.174	1.000
543	2828	2285	-0.021	-1.000	-0.174	1.000
543	2872	2329	-0.021	-1.000	-0.174	1.000
543	2878	2335	0.063	1.000	0.522	0.250
543	3041	2498	0.000	0.000	0.000	1.000
543	3216	2673	0.000	0.000	0.000	1.000
543	3400	2857	0.021	0.111	0.111	1.000
543	3508	2965	-0.021	-1.000	-0.174	1.000
543	3766	3223	0.063	1.000	0.522	0.250
543	3846	3303	0.063	1.000	0.522	0.250
593	598	5	0.076	1.000	1.000	0.083
593	759	166	-0.007	-1.000	-0.091	1.000
593	806	213	-0.021	-1.000	-0.174	1.000
593	815	222	-0.028	-1.000	-0.213	1.000
593	849	256	-0.007	-1.000	-0.091	1.000
593	872	279	-0.028	-1.000	-0.213	1.000
593	925	332	-0.007	-1.000	-0.091	1.000
593	974	381	-0.035	-1.000	-0.255	1.000
593	982	389	-0.007	-1.000	-0.091	1.000

593	1051	458	-0.021	-1.000	-0.174	1.000
593	1124	531	-0.028	-1.000	-0.213	1.000
593	1234	641	-0.007	-1.000	-0.091	1.000
593	1334	741	-0.007	-1.000	-0.091	1.000
593	1539	946	-0.035	-1.000	-0.255	1.000
593	2045	1452	-0.028	-1.000	-0.213	1.000
593	2129	1536	-0.028	-1.000	-0.213	1.000
593	2214	1621	-0.007	-1.000	-0.091	1.000
593	2313	1720	0.069	1.000	0.674	0.167
593	2566	1973	-0.007	-1.000	-0.091	1.000
593	2567	1974	-0.007	-1.000	-0.091	1.000
593	2759	2166	-0.007	-1.000	-0.091	1.000
593	2796	2203	-0.007	-1.000	-0.091	1.000
593	2828	2235	-0.007	-1.000	-0.091	1.000
593	2872	2279	-0.007	-1.000	-0.091	1.000
593	2878	2285	-0.007	-1.000	-0.091	1.000
593	3041	2448	-0.028	-1.000	-0.213	1.000
593	3216	2623	0.056	1.000	0.426	0.333
593	3400	2807	-0.021	-1.000	-0.174	1.000
593	3508	2915	-0.007	-1.000	-0.091	1.000
593	3766	3173	-0.007	-1.000	-0.091	1.000
593	3846	3253	-0.007	-1.000	-0.091	1.000
598	759	161	-0.007	-1.000	-0.091	1.000
598	806	208	-0.021	-1.000	-0.174	1.000
598	815	217	-0.028	-1.000	-0.213	1.000
598	849	251	-0.007	-1.000	-0.091	1.000
598	872	274	-0.028	-1.000	-0.213	1.000
598	925	327	-0.007	-1.000	-0.091	1.000
598	974	376	-0.035	-1.000	-0.255	1.000
598	982	384	-0.007	-1.000	-0.091	1.000
598	1051	453	-0.021	-1.000	-0.174	1.000
598	1124	526	-0.028	-1.000	-0.213	1.000
598	1234	636	-0.007	-1.000	-0.091	1.000
598	1334	736	-0.007	-1.000	-0.091	1.000
598	1539	941	-0.035	-1.000	-0.255	1.000
598	2045	1447	-0.028	-1.000	-0.213	1.000
598	2129	1531	-0.028	-1.000	-0.213	1.000
598	2214	1616	-0.007	-1.000	-0.091	1.000
598	2313	1715	0.069	1.000	0.674	0.167
598	2566	1968	-0.007	-1.000	-0.091	1.000
598	2567	1969	-0.007	-1.000	-0.091	1.000
598	2759	2161	-0.007	-1.000	-0.091	1.000

598	2796	2198	-0.007	-1.000	-0.091	1.000
598	2828	2230	-0.007	-1.000	-0.091	1.000
598	2872	2274	-0.007	-1.000	-0.091	1.000
598	2878	2280	-0.007	-1.000	-0.091	1.000
598	3041	2443	-0.028	-1.000	-0.213	1.000
598	3216	2618	0.056	1.000	0.426	0.333
598	3400	2802	-0.021	-1.000	-0.174	1.000
598	3508	2910	-0.007	-1.000	-0.091	1.000
598	3766	3168	-0.007	-1.000	-0.091	1.000
598	3846	3248	-0.007	-1.000	-0.091	1.000
759	806	47	0.063	1.000	0.522	0.250
759	815	56	0.056	1.000	0.426	0.333
759	849	90	-0.007	-1.000	-0.091	1.000
759	872	113	0.056	1.000	0.426	0.333
759	925	166	-0.007	-1.000	-0.091	1.000
759	974	215	-0.035	-1.000	-0.255	1.000
759	982	223	-0.007	-1.000	-0.091	1.000
759	1051	292	-0.021	-1.000	-0.174	1.000
759	1124	365	-0.028	-1.000	-0.213	1.000
759	1234	475	-0.007	-1.000	-0.091	1.000
759	1334	575	-0.007	-1.000	-0.091	1.000
759	1539	780	0.049	1.000	0.357	0.417
759	2045	1286	0.056	1.000	0.426	0.333
759	2129	1370	0.056	1.000	0.426	0.333
759	2214	1455	-0.007	-1.000	-0.091	1.000
759	2313	1554	-0.014	-1.000	-0.135	1.000
759	2566	1807	-0.007	-1.000	-0.091	1.000
759	2567	1808	-0.007	-1.000	-0.091	1.000
759	2759	2000	-0.007	-1.000	-0.091	1.000
759	2796	2037	-0.007	-1.000	-0.091	1.000
759	2828	2069	-0.007	-1.000	-0.091	1.000
759	2872	2113	-0.007	-1.000	-0.091	1.000
759	2878	2119	-0.007	-1.000	-0.091	1.000
759	3041	2282	-0.028	-1.000	-0.213	1.000
759	3216	2457	-0.028	-1.000	-0.213	1.000
759	3400	2641	-0.021	-1.000	-0.174	1.000
759	3508	2749	-0.007	-1.000	-0.091	1.000
759	3766	3007	-0.007	-1.000	-0.091	1.000
759	3846	3087	-0.007	-1.000	-0.091	1.000
806	815	9	0.167	1.000	0.816	0.018*
806	849	43	0.063	1.000	0.522	0.250
806	872	66	0.167	1.000	0.816	0.018*

806	925	119	0.063	1.000	0.522	0.250
806	974	168	0.063	0.429	0.293	0.523
806	982	176	-0.021	-1.000	-0.174	1.000
806	1051	245	0.104	0.556	0.556	0.127
806	1124	318	0.083	0.500	0.408	0.236
806	1234	428	-0.021	-1.000	-0.174	1.000
806	1334	528	-0.021	-1.000	-0.174	1.000
806	1539	733	0.146	1.000	0.683	0.045*
806	2045	1239	0.167	1.000	0.816	0.018*
806	2129	1323	0.167	1.000	0.816	0.018*
806	2214	1408	0.063	1.000	0.522	0.250
806	2313	1507	-0.042	-1.000	-0.258	1.000
806	2566	1760	-0.021	-1.000	-0.174	1.000
806	2567	1761	-0.021	-1.000	-0.174	1.000
806	2759	1953	0.063	1.000	0.522	0.250
806	2796	1990	-0.021	-1.000	-0.174	1.000
806	2828	2022	-0.021	-1.000	-0.174	1.000
806	2872	2066	-0.021	-1.000	-0.174	1.000
806	2878	2072	0.063	1.000	0.522	0.250
806	3041	2235	0.000	0.000	0.000	1.000
806	3216	2410	0.000	0.000	0.000	1.000
806	3400	2594	0.021	0.111	0.111	1.000
806	3508	2702	-0.021	-1.000	-0.174	1.000
806	3766	2960	0.063	1.000	0.522	0.250
806	3846	3040	0.063	1.000	0.522	0.250
815	849	34	0.056	1.000	0.426	0.333
815	872	57	0.222	1.000	1.000	0.002**
815	925	110	0.056	1.000	0.426	0.333
815	974	159	0.111	0.571	0.478	0.222
815	982	167	-0.028	-1.000	-0.213	1.000
815	1051	236	0.167	1.000	0.816	0.018*
815	1124	309	0.139	0.625	0.625	0.067
815	1234	419	-0.028	-1.000	-0.213	1.000
815	1334	519	-0.028	-1.000	-0.213	1.000
815	1539	724	0.194	1.000	0.837	0.010*
815	2045	1230	0.139	0.625	0.625	0.067
815	2129	1314	0.139	0.625	0.625	0.067
815	2214	1399	0.056	1.000	0.426	0.333
815	2313	1498	-0.056	-1.000	-0.316	0.515
815	2566	1751	-0.028	-1.000	-0.213	1.000
815	2567	1752	-0.028	-1.000	-0.213	1.000
815	2759	1944	0.056	1.000	0.426	0.333

815	2796	1981	-0.028	-1.000	-0.213	1.000
815	2828	2013	-0.028	-1.000	-0.213	1.000
815	2872	2057	-0.028	-1.000	-0.213	1.000
815	2878	2063	0.056	1.000	0.426	0.333
815	3041	2226	-0.028	-0.250	-0.125	1.000
815	3216	2401	0.056	0.250	0.250	0.547
815	3400	2585	0.000	0.000	0.000	1.000
815	3508	2693	-0.028	-1.000	-0.213	1.000
815	3766	2951	0.056	1.000	0.426	0.333
815	3846	3031	0.056	1.000	0.426	0.333
849	872	23	0.056	1.000	0.426	0.333
849	925	76	-0.007	-1.000	-0.091	1.000
849	974	125	0.049	1.000	0.357	0.417
849	982	133	-0.007	-1.000	-0.091	1.000
849	1051	202	0.063	1.000	0.522	0.250
849	1124	275	0.056	1.000	0.426	0.333
849	1234	385	-0.007	-1.000	-0.091	1.000
849	1334	485	-0.007	-1.000	-0.091	1.000
849	1539	690	0.049	1.000	0.357	0.417
849	2045	1196	0.056	1.000	0.426	0.333
849	2129	1280	0.056	1.000	0.426	0.333
849	2214	1365	0.076	1.000	1.000	0.083
849	2313	1464	-0.014	-1.000	-0.135	1.000
849	2566	1717	-0.007	-1.000	-0.091	1.000
849	2567	1718	-0.007	-1.000	-0.091	1.000
849	2759	1910	-0.007	-1.000	-0.091	1.000
849	2796	1947	-0.007	-1.000	-0.091	1.000
849	2828	1979	-0.007	-1.000	-0.091	1.000
849	2872	2023	-0.007	-1.000	-0.091	1.000
849	2878	2029	-0.007	-1.000	-0.091	1.000
849	3041	2192	-0.028	-1.000	-0.213	1.000
849	3216	2367	0.056	1.000	0.426	0.333
849	3400	2551	0.063	1.000	0.522	0.250
849	3508	2659	-0.007	-1.000	-0.091	1.000
849	3766	2917	0.076	1.000	1.000	0.083
849	3846	2997	0.076	1.000	1.000	0.083
872	925	53	0.056	1.000	0.426	0.333
872	974	102	0.111	0.571	0.478	0.222
872	982	110	-0.028	-1.000	-0.213	1.000
872	1051	179	0.167	1.000	0.816	0.018*
872	1124	252	0.139	0.625	0.625	0.067
872	1234	362	-0.028	-1.000	-0.213	1.000

872	1334	462	-0.028	-1.000	-0.213	1.000
872	1539	667	0.194	1.000	0.837	0.010*
872	2045	1173	0.139	0.625	0.625	0.067
872	2129	1257	0.139	0.625	0.625	0.067
872	2214	1342	0.056	1.000	0.426	0.333
872	2313	1441	-0.056	-1.000	-0.316	0.515
872	2566	1694	-0.028	-1.000	-0.213	1.000
872	2567	1695	-0.028	-1.000	-0.213	1.000
872	2759	1887	0.056	1.000	0.426	0.333
872	2796	1924	-0.028	-1.000	-0.213	1.000
872	2828	1956	-0.028	-1.000	-0.213	1.000
872	2872	2000	-0.028	-1.000	-0.213	1.000
872	2878	2006	0.056	1.000	0.426	0.333
872	3041	2169	-0.028	-0.250	-0.125	1.000
872	3216	2344	0.056	0.250	0.250	0.547
872	3400	2528	0.000	0.000	0.000	1.000
872	3508	2636	-0.028	-1.000	-0.213	1.000
872	3766	2894	0.056	1.000	0.426	0.333
872	3846	2974	0.056	1.000	0.426	0.333
925	974	49	0.049	1.000	0.357	0.417
925	982	57	-0.007	-1.000	-0.091	1.000
925	1051	126	0.063	1.000	0.522	0.250
925	1124	199	0.056	1.000	0.426	0.333
925	1234	309	-0.007	-1.000	-0.091	1.000
925	1334	409	-0.007	-1.000	-0.091	1.000
925	1539	614	0.049	1.000	0.357	0.417
925	2045	1120	0.056	1.000	0.426	0.333
925	2129	1204	0.056	1.000	0.426	0.333
925	2214	1289	-0.007	-1.000	-0.091	1.000
925	2313	1388	-0.014	-1.000	-0.135	1.000
925	2566	1641	-0.007	-1.000	-0.091	1.000
925	2567	1642	-0.007	-1.000	-0.091	1.000
925	2759	1834	0.076	1.000	1.000	0.083
925	2796	1871	-0.007	-1.000	-0.091	1.000
925	2828	1903	-0.007	-1.000	-0.091	1.000
925	2872	1947	-0.007	-1.000	-0.091	1.000
925	2878	1953	0.076	1.000	1.000	0.083
925	3041	2116	0.056	1.000	0.426	0.333
925	3216	2291	-0.028	-1.000	-0.213	1.000
925	3400	2475	-0.021	-1.000	-0.174	1.000
925	3508	2583	-0.007	-1.000	-0.091	1.000
925	3766	2841	-0.007	-1.000	-0.091	1.000

925	3846	2921	-0.007	-1.000	-0.091	1.000
974	982	8	0.049	1.000	0.357	0.417
974	1051	77	0.146	1.000	0.683	0.045*
974	1124	150	0.194	1.000	0.837	0.010*
974	1234	260	-0.035	-1.000	-0.255	1.000
974	1334	360	0.049	1.000	0.357	0.417
974	1539	565	0.160	0.657	0.657	0.072
974	2045	1071	0.111	0.571	0.478	0.222
974	2129	1155	0.111	0.571	0.478	0.222
974	2214	1240	0.049	1.000	0.357	0.417
974	2313	1339	-0.069	-1.000	-0.378	0.470
974	2566	1592	0.049	1.000	0.357	0.417
974	2567	1593	0.049	1.000	0.357	0.417
974	2759	1785	0.049	1.000	0.357	0.417
974	2796	1822	-0.035	-1.000	-0.255	1.000
974	2828	1854	-0.035	-1.000	-0.255	1.000
974	2872	1898	0.049	1.000	0.357	0.417
974	2878	1904	0.049	1.000	0.357	0.417
974	3041	2067	0.028	0.143	0.120	1.000
974	3216	2242	0.028	0.143	0.120	1.000
974	3400	2426	0.063	0.429	0.293	0.523
974	3508	2534	-0.035	-1.000	-0.255	1.000
974	3766	2792	0.049	1.000	0.357	0.417
974	3846	2872	0.049	1.000	0.357	0.417
982	1051	69	-0.021	-1.000	-0.174	1.000
982	1124	142	0.056	1.000	0.426	0.333
982	1234	252	-0.007	-1.000	-0.091	1.000
982	1334	352	0.076	1.000	1.000	0.083
982	1539	557	0.049	1.000	0.357	0.417
982	2045	1063	0.056	1.000	0.426	0.333
982	2129	1147	0.056	1.000	0.426	0.333
982	2214	1232	-0.007	-1.000	-0.091	1.000
982	2313	1331	-0.014	-1.000	-0.135	1.000
982	2566	1584	0.076	1.000	1.000	0.083
982	2567	1585	0.076	1.000	1.000	0.083
982	2759	1777	-0.007	-1.000	-0.091	1.000
982	2796	1814	-0.007	-1.000	-0.091	1.000
982	2828	1846	-0.007	-1.000	-0.091	1.000
982	2872	1890	-0.007	-1.000	-0.091	1.000
982	2878	1896	-0.007	-1.000	-0.091	1.000
982	3041	2059	0.056	1.000	0.426	0.333
982	3216	2234	-0.028	-1.000	-0.213	1.000

982	3400	2418	0.063	1.000	0.522	0.250
982	3508	2526	-0.007	-1.000	-0.091	1.000
982	3766	2784	-0.007	-1.000	-0.091	1.000
982	3846	2864	-0.007	-1.000	-0.091	1.000
1051	1124	73	0.167	1.000	0.816	0.018*
1051	1234	183	-0.021	-1.000	-0.174	1.000
1051	1334	283	-0.021	-1.000	-0.174	1.000
1051	1539	488	0.146	1.000	0.683	0.045*
1051	2045	994	0.083	0.500	0.408	0.236
1051	2129	1078	0.083	0.500	0.408	0.236
1051	2214	1163	0.063	1.000	0.522	0.250
1051	2313	1262	-0.042	-1.000	-0.258	1.000
1051	2566	1515	-0.021	-1.000	-0.174	1.000
1051	2567	1516	-0.021	-1.000	-0.174	1.000
1051	2759	1708	0.063	1.000	0.522	0.250
1051	2796	1745	-0.021	-1.000	-0.174	1.000
1051	2828	1777	-0.021	-1.000	-0.174	1.000
1051	2872	1821	-0.021	-1.000	-0.174	1.000
1051	2878	1827	0.063	1.000	0.522	0.250
1051	3041	1990	0.000	0.000	0.000	1.000
1051	3216	2165	0.083	0.500	0.408	0.236
1051	3400	2349	0.021	0.111	0.111	1.000
1051	3508	2457	-0.021	-1.000	-0.174	1.000
1051	3766	2715	0.063	1.000	0.522	0.250
1051	3846	2795	0.063	1.000	0.522	0.250
1124	1234	110	-0.028	-1.000	-0.213	1.000
1124	1334	210	0.056	1.000	0.426	0.333
1124	1539	415	0.194	1.000	0.837	0.010*
1124	2045	921	0.139	0.625	0.625	0.067
1124	2129	1005	0.139	0.625	0.625	0.067
1124	2214	1090	0.056	1.000	0.426	0.333
1124	2313	1189	-0.056	-1.000	-0.316	0.515
1124	2566	1442	0.056	1.000	0.426	0.333
1124	2567	1443	0.056	1.000	0.426	0.333
1124	2759	1635	0.056	1.000	0.426	0.333
1124	2796	1672	-0.028	-1.000	-0.213	1.000
1124	2828	1704	-0.028	-1.000	-0.213	1.000
1124	2872	1748	-0.028	-1.000	-0.213	1.000
1124	2878	1754	0.056	1.000	0.426	0.333
1124	3041	1917	0.056	0.250	0.250	0.547
1124	3216	2092	0.056	0.250	0.250	0.547
1124	3400	2276	0.083	0.500	0.408	0.236

1124	3508	2384	-0.028	-1.000	-0.213	1.000
1124	3766	2642	0.056	1.000	0.426	0.333
1124	3846	2722	0.056	1.000	0.426	0.333
1234	1334	100	-0.007	-1.000	-0.091	1.000
1234	1539	305	-0.035	-1.000	-0.255	1.000
1234	2045	811	-0.028	-1.000	-0.213	1.000
1234	2129	895	-0.028	-1.000	-0.213	1.000
1234	2214	980	-0.007	-1.000	-0.091	1.000
1234	2313	1079	-0.014	-1.000	-0.135	1.000
1234	2566	1332	-0.007	-1.000	-0.091	1.000
1234	2567	1333	-0.007	-1.000	-0.091	1.000
1234	2759	1525	-0.007	-1.000	-0.091	1.000
1234	2796	1562	-0.007	-1.000	-0.091	1.000
1234	2828	1594	-0.007	-1.000	-0.091	1.000
1234	2872	1638	-0.007	-1.000	-0.091	1.000
1234	2878	1644	-0.007	-1.000	-0.091	1.000
1234	3041	1807	-0.028	-1.000	-0.213	1.000
1234	3216	1982	-0.028	-1.000	-0.213	1.000
1234	3400	2166	-0.021	-1.000	-0.174	1.000
1234	3508	2274	-0.007	-1.000	-0.091	1.000
1234	3766	2532	-0.007	-1.000	-0.091	1.000
1234	3846	2612	-0.007	-1.000	-0.091	1.000
1334	1539	205	0.049	1.000	0.357	0.417
1334	2045	711	0.056	1.000	0.426	0.333
1334	2129	795	0.056	1.000	0.426	0.333
1334	2214	880	-0.007	-1.000	-0.091	1.000
1334	2313	979	-0.014	-1.000	-0.135	1.000
1334	2566	1232	0.076	1.000	1.000	0.083
1334	2567	1233	0.076	1.000	1.000	0.083
1334	2759	1425	-0.007	-1.000	-0.091	1.000
1334	2796	1462	-0.007	-1.000	-0.091	1.000
1334	2828	1494	-0.007	-1.000	-0.091	1.000
1334	2872	1538	-0.007	-1.000	-0.091	1.000
1334	2878	1544	-0.007	-1.000	-0.091	1.000
1334	3041	1707	0.056	1.000	0.426	0.333
1334	3216	1882	-0.028	-1.000	-0.213	1.000
1334	3400	2066	0.063	1.000	0.522	0.250
1334	3508	2174	-0.007	-1.000	-0.091	1.000
1334	3766	2432	-0.007	-1.000	-0.091	1.000
1334	3846	2512	-0.007	-1.000	-0.091	1.000
1539	2045	506	0.194	1.000	0.837	0.010*
1539	2129	590	0.194	1.000	0.837	0.010*

1539	2214	675	0.049	1.000	0.357	0.417
1539	2313	774	-0.069	-1.000	-0.378	0.470
1539	2566	1027	0.049	1.000	0.357	0.417
1539	2567	1028	0.049	1.000	0.357	0.417
1539	2759	1220	0.049	1.000	0.357	0.417
1539	2796	1257	-0.035	-1.000	-0.255	1.000
1539	2828	1289	-0.035	-1.000	-0.255	1.000
1539	2872	1333	-0.035	-1.000	-0.255	1.000
1539	2878	1339	0.049	1.000	0.357	0.417
1539	3041	1502	0.028	0.143	0.120	1.000
1539	3216	1677	0.028	0.143	0.120	1.000
1539	3400	1861	0.063	0.429	0.293	0.523
1539	3508	1969	-0.035	-1.000	-0.255	1.000
1539	3766	2227	0.049	1.000	0.357	0.417
1539	3846	2307	0.049	1.000	0.357	0.417
2045	2129	84	0.222	1.000	1.000	0.002**
2045	2214	169	0.056	1.000	0.426	0.333
2045	2313	268	-0.056	-1.000	-0.316	0.515
2045	2566	521	0.056	1.000	0.426	0.333
2045	2567	522	0.056	1.000	0.426	0.333
2045	2759	714	0.056	1.000	0.426	0.333
2045	2796	751	-0.028	-1.000	-0.213	1.000
2045	2828	783	-0.028	-1.000	-0.213	1.000
2045	2872	827	-0.028	-1.000	-0.213	1.000
2045	2878	833	0.056	1.000	0.426	0.333
2045	3041	996	0.056	0.250	0.250	0.547
2045	3216	1171	-0.028	-0.250	-0.125	1.000
2045	3400	1355	0.083	0.500	0.408	0.236
2045	3508	1463	-0.028	-1.000	-0.213	1.000
2045	3766	1721	0.056	1.000	0.426	0.333
2045	3846	1801	0.056	1.000	0.426	0.333
2129	2214	85	0.056	1.000	0.426	0.333
2129	2313	184	-0.056	-1.000	-0.316	0.515
2129	2566	437	0.056	1.000	0.426	0.333
2129	2567	438	0.056	1.000	0.426	0.333
2129	2759	630	0.056	1.000	0.426	0.333
2129	2796	667	-0.028	-1.000	-0.213	1.000
2129	2828	699	-0.028	-1.000	-0.213	1.000
2129	2872	743	-0.028	-1.000	-0.213	1.000
2129	2878	749	0.056	1.000	0.426	0.333
2129	3041	912	0.056	0.250	0.250	0.547
2129	3216	1087	-0.028	-0.250	-0.125	1.000

2129	3400	1271	0.083	0.500	0.408	0.236
2129	3508	1379	-0.028	-1.000	-0.213	1.000
2129	3766	1637	0.056	1.000	0.426	0.333
2129	3846	1717	0.056	1.000	0.426	0.333
2214	2313	99	-0.014	-1.000	-0.135	1.000
2214	2566	352	-0.007	-1.000	-0.091	1.000
2214	2567	353	-0.007	-1.000	-0.091	1.000
2214	2759	545	-0.007	-1.000	-0.091	1.000
2214	2796	582	-0.007	-1.000	-0.091	1.000
2214	2828	614	-0.007	-1.000	-0.091	1.000
2214	2872	658	-0.007	-1.000	-0.091	1.000
2214	2878	664	-0.007	-1.000	-0.091	1.000
2214	3041	827	-0.028	-1.000	-0.213	1.000
2214	3216	1002	0.056	1.000	0.426	0.333
2214	3400	1186	0.063	1.000	0.522	0.250
2214	3508	1294	-0.007	-1.000	-0.091	1.000
2214	3766	1552	0.076	1.000	1.000	0.083
2214	3846	1632	0.076	1.000	1.000	0.083
2313	2566	253	-0.014	-1.000	-0.135	1.000
2313	2567	254	-0.014	-1.000	-0.135	1.000
2313	2759	446	-0.014	-1.000	-0.135	1.000
2313	2796	483	-0.014	-1.000	-0.135	1.000
2313	2828	515	-0.014	-1.000	-0.135	1.000
2313	2872	559	-0.014	-1.000	-0.135	1.000
2313	2878	565	-0.014	-1.000	-0.135	1.000
2313	3041	728	-0.056	-1.000	-0.316	0.515
2313	3216	903	0.111	1.000	0.632	0.091
2313	3400	1087	-0.042	-1.000	-0.258	1.000
2313	3508	1195	-0.014	-1.000	-0.135	1.000
2313	3766	1453	-0.014	-1.000	-0.135	1.000
2313	3846	1533	-0.014	-1.000	-0.135	1.000
2566	2567	1	0.076	1.000	1.000	0.083
2566	2759	193	-0.007	-1.000	-0.091	1.000
2566	2796	230	-0.007	-1.000	-0.091	1.000
2566	2828	262	-0.007	-1.000	-0.091	1.000
2566	2872	306	-0.007	-1.000	-0.091	1.000
2566	2878	312	-0.007	-1.000	-0.091	1.000
2566	3041	475	0.056	1.000	0.426	0.333
2566	3216	650	-0.028	-1.000	-0.213	1.000
2566	3400	834	0.063	1.000	0.522	0.250
2566	3508	942	-0.007	-1.000	-0.091	1.000
2566	3766	1200	-0.007	-1.000	-0.091	1.000

2566	3846	1280	-0.007	-1.000	-0.091	1.000
2567	2759	192	-0.007	-1.000	-0.091	1.000
2567	2796	229	-0.007	-1.000	-0.091	1.000
2567	2828	261	-0.007	-1.000	-0.091	1.000
2567	2872	305	-0.007	-1.000	-0.091	1.000
2567	2878	311	-0.007	-1.000	-0.091	1.000
2567	3041	474	0.056	1.000	0.426	0.333
2567	3216	649	-0.028	-1.000	-0.213	1.000
2567	3400	833	0.063	1.000	0.522	0.250
2567	3508	941	-0.007	-1.000	-0.091	1.000
2567	3766	1199	-0.007	-1.000	-0.091	1.000
2567	3846	1279	-0.007	-1.000	-0.091	1.000
2759	2796	37	-0.007	-1.000	-0.091	1.000
2759	2828	69	-0.007	-1.000	-0.091	1.000
2759	2872	113	-0.007	-1.000	-0.091	1.000
2759	2878	119	0.076	1.000	1.000	0.083
2759	3041	282	0.056	1.000	0.426	0.333
2759	3216	457	-0.028	-1.000	-0.213	1.000
2759	3400	641	-0.021	-1.000	-0.174	1.000
2759	3508	749	-0.007	-1.000	-0.091	1.000
2759	3766	1007	-0.007	-1.000	-0.091	1.000
2759	3846	1087	-0.007	-1.000	-0.091	1.000
2796	2828	32	0.076	1.000	1.000	0.083
2796	2872	76	-0.007	-1.000	-0.091	1.000
2796	2878	82	-0.007	-1.000	-0.091	1.000
2796	3041	245	0.056	1.000	0.426	0.333
2796	3216	420	-0.028	-1.000	-0.213	1.000
2796	3400	604	-0.021	-1.000	-0.174	1.000
2796	3508	712	-0.007	-1.000	-0.091	1.000
2796	3766	970	-0.007	-1.000	-0.091	1.000
2796	3846	1050	-0.007	-1.000	-0.091	1.000
2828	2872	44	-0.007	-1.000	-0.091	1.000
2828	2878	50	-0.007	-1.000	-0.091	1.000
2828	3041	213	0.056	1.000	0.426	0.333
2828	3216	388	-0.028	-1.000	-0.213	1.000
2828	3400	572	-0.021	-1.000	-0.174	1.000
2828	3508	680	-0.007	-1.000	-0.091	1.000
2828	3766	938	-0.007	-1.000	-0.091	1.000
2828	3846	1018	-0.007	-1.000	-0.091	1.000
2872	2878	6	-0.007	-1.000	-0.091	1.000
2872	3041	169	-0.028	-1.000	-0.213	1.000
2872	3216	344	-0.028	-1.000	-0.213	1.000

2872	3400	528	-0.021	-1.000	-0.174	1.000
2872	3508	636	-0.007	-1.000	-0.091	1.000
2872	3766	894	-0.007	-1.000	-0.091	1.000
2872	3846	974	-0.007	-1.000	-0.091	1.000
2878	3041	163	0.056	1.000	0.426	0.333
2878	3216	338	-0.028	-1.000	-0.213	1.000
2878	3400	522	-0.021	-1.000	-0.174	1.000
2878	3508	630	-0.007	-1.000	-0.091	1.000
2878	3766	888	-0.007	-1.000	-0.091	1.000
2878	3846	968	-0.007	-1.000	-0.091	1.000
3041	3216	175	-0.111	-1.000	-0.500	0.208
3041	3400	359	0.000	0.000	0.000	1.000
3041	3508	467	-0.028	-1.000	-0.213	1.000
3041	3766	725	-0.028	-1.000	-0.213	1.000
3041	3846	805	-0.028	-1.000	-0.213	1.000
3216	3400	184	0.000	0.000	0.000	1.000
3216	3508	292	-0.028	-1.000	-0.213	1.000
3216	3766	550	0.056	1.000	0.426	0.333
3216	3846	630	0.056	1.000	0.426	0.333
3400	3508	108	0.063	1.000	0.522	0.250
3400	3766	366	0.063	1.000	0.522	0.250
3400	3846	446	0.063	1.000	0.522	0.250
3508	3766	258	-0.007	-1.000	-0.091	1.000
3508	3846	338	-0.007	-1.000	-0.091	1.000
3766	3846	80	0.076	1.000	1.000	

APPENDIX 5: Amino acid composition on all accessions.

Ekebure		Mbeere1		Mbeere2		Mbeere3		Mbeere4		Kidney bean		KATB1		VAX4		19		23		Unknown	
AA	%	AA	%	AA	%	AA	%	AA	%	AA	%	AA	%	AA	%	AA	%	AA	%	AA	%
Ala	5.8	Ala	6.1	Ala	5.7	Ala	5.5	Ala	5.8	Ala	6.1	Ala	6.1	Ala	5.8	Ala	5.7	Ala	5.7	Ala	5.8
Arg	4.9	Arg	4.6	Arg	4.9	Arg	4.9	Arg	4.9	Cys	0.0	Arg	4.9	Arg	4.9	Arg	4.9	Arg	4.9	Arg	4.9
Asp	4.9	Asp	4.9	Asp	4.9	Asp	4.9	Asp	4.9	Asp	4.9	Asp	4.9	Asn	6.4	Asp	4.9	Asp	4.9	Asp	4.9
Asn	6.1	Asn	6.1	Cys	0	Cys	0	Cys	0	Glu	9.3	Asn	6.1	Asp	4.9	Cys	0	Cys	0	Cys	0
Glu	9.6	Cys	0.0	Glu	9.5	Glu	9.6	Glu	9.6	Phe	6.9	Cys	0.0	Cys	0.0	Glu	9.6	Glu	9.6	Glu	9.6
His	2.0	Gln	5.2	Gly	4.1	Gly	4.1	Gly	4.1	Gly	4.1	Gly	4.1	Gln	5.2	Gly	4.1	Gly	4.1	Gly	4.1
Ile	5.5	Glu	9.6	His	2	His	2.0	His	2.1	His	2.0	Glu	9.3	Glu	9.6	His	2.0	His	2.0	Ile	5.5
Leu	11.6	Gly	4.1	Ile	5.5	Ile	5.5	Ile	5.5	Ile	5.8	His	2.0	Gly	4.1	Ile	5.5	Ile	5.9	His	2.0
Lys	5.2	His	2.0	Lys	5.2	Leu	11.9	Leu	11.6	Leu	11.6	Ile	5.8	Ile	5.5	Leu	11.6	Leu	11.9	Leu	11.6
Met	1.2	Ile	5.5	Leu	11.6	Lys	5.2	Met	1.2	Lys	5.2	Lys	5.2	Leu	11.6	Lys	5.2	Lys	5.2	Lys	5.2
Phe	7.0	Leu	11.6	Met	1.2	Met	1.2	Phe	7.0	Met	0.9	Met	0.9	Lys	5.2	Met	1.2	Met	0.9	Met	1.2
Pro	4.3	Lys	5.2	Phe	7.0	Phe	7.0	Pro	4.1	Asn	6.09	Pro	4.3	Met	1.2	Phe	7.0	Phe	7.0	Phe	7.0
Ser	9.4	Met	1.2	Pro	4.06	Pro	4.3	Ser	9.0	Pro	4.3	Thr	3.8	Phe	7.0	Pro	4.0	Pro	4.1	Pro	4.06
Thr	3.8	Phe	7.0	Ser	9.0	Ser	9.0	Trp	0.3	Trp	0.3	Trp	0.3	Pro	4.3	Ser	9.0	Ser	9.0	Ser	9.0
Trp	0.3	Thr	3.8	Thr	3.9	Trp	0.3	Tyr	2.9	Tyr	2.9	Try	2.9	Ser	9.0	Thr	3.8	Thr	3.8	Thr	3.8
Val	6.7	Try	2.1	Try	2.9	Try	0.3	Thr	3.8	Val	6.7	Val	6.7	Tyr	2.9	Trp	0.3	Trp	0.3	Trp	0.3
Gly	9.6	Ser	9.1	Val	6.7	Val	6.7	Val	6.8	Ser	9.0	Ser	9.0	Thr	3.9	Try	2.9	Try	2.9	Tyr	2.9
Gln	5.2	Val	6.8	Trp	0.29	Asn	6.3	Asn	6.4	Phe	7.0	Phe	7.0	Trp	0.3	Val	6.7	Val	6.8	Val	6.7