CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Exploring Agro-morphological Variability in Diverse Soybean (*Glycine max*) Germplasm

by

Hina Areej Zafar

A thesis submitted in partial fulfillment for the
degree of Master of Science

in the
Faculty of Health and Life Sciences
Department of Bioinformatics and Biosciences

2024

I dedicate this thesis to my loving and supportive family and friends who have fully helped me in achieving my life goals.

CERTIFICATE OF APPROVAL

Exploring Agro-morphological Variability in Diverse Soybean (*Glycine max*) Germplasm

by
Hina Areej Zafar
(MBS223020)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Sayed Afzal Shah	NUMS, Rawalpindi
(b)	Internal Examiner	Dr. M. Asad Anwar	CUST, Islamabad
(c)	Supervisor	Dr. Sohail Ahmad Jan	CUST, Islamabad

Dr. Sohail Ahmad Jan
Thesis Supervisor
September, 2024

Dr. Syeda Marriam Bakhtiar
Head
Deptarment of Bioinfo. & Biosciences
September, 2024

Dr. Sahar Fazal
Dean
Faculty of Health & Life Sciences
September, 2024

Author's Declaration

I, **Hina Areej Zafar** hereby state that my MS thesis titled "**Exploring Agro-morphological Variability in Diverse Soybean (*Glycine max*) Germplasm**" is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

Hina

(Hina Areej Zafar)
Registration No: MBS223020

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "**Exploring Agro-morphological Variability in Diverse Soybean (*Glycine max*) Germplasm**" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

Hina

(Hina Areej Zafar)
Registration No: MBS223020

Acknowledgement

I am very grateful to Allah Almighty for his countless blessings bestowed upon me. I am grateful for all that I have received and will receive in my life. I am ever so thankful for the countless miracles that took place just for me. I am thankful for the infinite guidance and support of **my parents**. They have been a wonderful blessing for me.

I'm thankful to **Capital University of Science and Technology**. This has been a second home for me for last 6 years. The environment provided here helped me to discover and develop numerous new skills. I would like to acknowledge the help and support of my supervisor **Dr. Sohail Ahmad Jan**. His invaluable advice assisted me in conduction and writing of this thesis. I am also thankful to National Agricultural Research Centre (NARC), for providing essential resources and expertise that facilitated the completion of this project.

I also like to acknowledge help of my friends in conduction of this thesis. The bonds that were created and the memories that were made will be cherished forever. The unexpected experiences and redirected paths we took together were priceless. I appreciate the help of **Zahra Khan** as well. My thesis would not have been completed without the existence of my friends.

Special shout out to the infinite amount of tea and coffee provided by **my siblings** to make this journey easier. While I was busy burning the midnight oil, my family provided me with lots of love, care and kindness. They helped me keep the hope alive. They motivated me and lift me up whenever I needed them. Thank you so much for being ever so selfless.

Hina

(Hina Areej Zafar)

Abstract

Morphological based diversity help in the characterization and identification of promising genotypes in a large set of population. Soybean is one of the important oilseed crop having multiple applications in food industry. Agromorphological based diversity for key quantitative traits helped in the identification of key soybean genotypes. Therefore, the current study was conducted to explore agro-morphological based variability among 33 diverse local Pakistani Soybean genotypes.Augmented design was used and nine key quantitative traits i.e. Days to flower initiation (DFI), Days to 50% flowering (DF 50%), Days to Flower completion (DFC), Pods per Plant (P/P), Plant Height (PH), Branches per Plant (B/P), Days to Maturity (DM), 100-Seed weightr (100 SW) and Yield per plant (Y/P) was recorded. The first five principal components (PCs) i.e. PC1, PC2, PC3, PC4 and PC5 with eigen value unity or more (>1) were considered important which accounted for 89.16% of overall phenotypic variation in population. High positive correlation was observed between DFC (Days to flower completion) and DF 50% (Days to 50% flowering) i.e. 0.877694 and high negative correlation was observed between P/P (Pods per plant) and DM (Days to maturity) i.e. -0.578701. Three genotypes i.e. GB-220, PJB-2030 and PJB-2006 were considered best and further suggest for the breeding programs.

Contents

Author's Declaration **iv**

Plagiarism Undertaking **v**

Acknowledgement **vi**

Abstract **vii**

List of Figures **xi**

List of Tables **xii**

Abbreviations **xiii**

1 Introduction **1**
- 1.1 Problem Statement . . . 5
- 1.2 Scope . . . 6
- 1.3 Aim and Objectives . . . 6
  - 1.3.1 Objectives . . . 6

2 Literature Review **7**
- 2.1 Genetic Diversity in Soybean . . . 7
  - 2.1.1 Importance of Genetic Diversity . . . 8
    - 2.1.1.1 Sources of Genetic Diversity . . . 8
  - 2.1.2 Conservation of Genetic Diversity . . . 9
- 2.2 Techniques for Assessing Genetic Diversity . . . 9
- 2.3 Agro-morphological Parameters in Soybean . . . 9
- 2.4 Methods for Evaluating Diversity . . . 10
  - 2.4.1 Morphological Evaluation . . . 10
  - 2.4.2 Biochemical Markers . . . 12
  - 2.4.3 Molecular Markers . . . 12
  - 2.4.4 Genomic and Transcriptomic Approaches . . . 14
  - 2.4.5 Phenotypic and Genotypic Data Integration . . . 16
- 2.5 Factors Influencing Diversity . . . 18
  - 2.5.1 Evolutionary Processes . . . 18
  - 2.5.2 Mutation . . . 18

2.5.3 Gene Flow . . . 19
2.5.4 Genetic Drift . . . 19
2.5.5 Natural Selection . . . 19
2.5.6 Breeding Practices . . . 20
2.5.7 Domestication and Cultivation . . . 21
2.5.8 Geographical and Environmental Factors . . . 21
2.5.9 Conservation Efforts . . . 22
2.5.10 Biotechnological Advances . . . 23
2.6 Agro-morphological Diversity in Soybean Genotypes . . . 23
2.6.1 Growth Parameters . . . 23
2.6.2 Yield Parameters . . . 24
2.6.3 Morphological Parameters . . . 24
2.6.4 Physiological Parameters . . . 25
2.6.5 Biotic and Abiotic Stress Parameters . . . 26
2.6.6 Agronomic Parameters . . . 27
2.6.7 Implications for Breeding and Crop Improvement . . . 28
2.6.8 Targeted Breeding for Yield Improvement . . . 29
2.7 Improving Nutritional and Industrial Qualities . . . 29
2.7.1 Incorporating Diverse Germplasm . . . 30
2.7.2 Integrating Advanced Technologies . . . 30
2.7.3 Addressing Socio-Economic Needs . . . 30
2.8 Challenges and Future Directions . . . 31
2.9 Soybean Imports and Exports in Pakistan . . . 33
2.9.1 Soybean Imports in Pakistan . . . 33
2.9.2 Soybean Exports in Pakistan . . . 34
2.10 Worldwide Production of Soybean . . . 35
2.10.1 Brazil . . . 36
2.10.2 United States . . . 36
2.10.3 Argentina . . . 36
2.10.4 China . . . 36
2.10.5 India . . . 37

3 Materials and Methods 38
3.1 Plant Materials . . . 38
3.2 Agro-morphological Characterization & Crop Management . . . 39
3.3 Data Analysis . . . 40

4 Results and Discussion 42
4.1 Evaluation of Morphological and Agronomic Traits in Diverse Accessions . . . 42
4.1.1 Flowering Time . . . 43
4.1.2 Yield Components . . . 43
4.1.3 Plant Height and Yield . . . 44

4.1.4 Branching 44
4.1.5 Maturity Time 44
4.2 Principal Components and Their Constituent Traits Contribution To Overall Variability Among Studied Germplasm 44
4.3 Principal Component Analysis of Agronomic and Morphological Traits in Diverse Accessions 47
4.4 Eigenvalues and Variance Explained by Principal Components . . . 52
4.5 Scree Plot Analysis 55
4.6 Principal Component Analysis (PCA) Biplot of Various Variables . 56
4.6.1 For instance 58
4.7 Factor Analysis Projection of Cases on Factor-Plane (1x2) 59
4.8 Factor Analysis Projection of Cases on Factor-Plane (2x3) 62
4.9 Correlations Before Key Morphological Traits of Soil 64

5 Conclusion **69**

Bibliography **71**

List of Figures

3.1 Graphical representation of the methodology used for exploring the agro morphological parameters of Soybean germplasm used. 41

4.1 Eigen value plot 53
4.2 Eigen values of correlation matrix 54
4.3 Principal Component Analysis Biplot. Factor 1 and Factor 2 56
4.4 Principal Component Analysis Biplot. Factor 2 and Factor 3 58
4.5 PCA scatter plot 59
4.6 PCA scatter plot 62

List of Tables

3.1 The detail of collected information of Soybean (*Glycine max*) Germplasm used . . . 38

4.1 Principal components and their constituent traits contribution to overall variability among studied germplasm. . . . 45
4.2 Various quantitative traits their mean values and standard deviation 45
4.3 Principal components and their constituent traits contribution to overall variability among studied germplasm . . . 47
4.4 Description of 3 agromorphologically diverse Soybean germplasm polulations revealed by PCA scatter plot. . . . 61
4.5 Correlation Matrix Analysis of Dataset Variables . . . 64

Abbreviations

B/P	Branches per plant
DFC	Days to flower Completion
DFI	Days to flower initiation
DM	Days to maturity
GWAS	Genome wide associated studies
P/P	Pods per plant
PCA	Principal component analysis
SSR	Simple Sequence Repeats
SW	Seed Weight
Y/P	Yield per plant
Y/P	Yield per plant

Chapter 1

Introduction

Protein (40%), oil (20%), and soluble carbohydrates (15%) make up soybean seeds, which are one of the most important crops in the world in terms of economic importance. Soybean contribute roughly 25% of the world's edible oil production, approximately two-thirds of the world's protein concentrate for livestock feed, and are vital raw materials in the food, pharmaceutical, and other industries. Compared to other grain legumes, soybeans are known for their higher productivity potential and adaptability. In Ethiopia, soybean productivity is currently below both its potential yield and the average global productivity [1].

Evaluating the genetic diversity of the current germplasm is essential to creating breeding plans that are successful. In order to facilitate hybridization and selection procedures, several populations with desired features are used. Finding the ideal parents for hybridization to produce new genetic lines for plant development is made easier by looking at genetic variety. Furthermore, assessing the genetic diversity of germplasm collections is essential to the long-term preservation and application of these resources. Genetically diverse parents combined in hybridization programs produce a wide range of diversity in subsequent generations as well as considerable heterotic effects in F1 hybrids. However, in order to maximize genetic recombination and raise possible yields, choosing the appropriate parents for crossing is crucial.

When examining genetic diversity and choosing parents for breeding, multivariate approaches like as principal component analysis (PCA) and cluster analysis are essential. Cluster analysis uses D2 statistics to group genotypes, and principal component analysis (PCA) is a useful statistical technique for simplifying quantitative data. In order to categorize genotypes, PCA transforms a number of correlated factors into a fresh set of uncorrelated variables. The potential genetic variety found in the current genotypes of soybeans has not yet been investigated, despite the approaches' applicability [2]. Thus, in order to utilize genetic diversity in upcoming breeding operations, it is essential to obtain a comprehensive understanding of the genetic variance. To create the best recombinant lines or new kinds, it is essential to discover superior parents with desirable features. In order to aid in future breeding efforts and the creation of new variations, this study was created to look into the genetic diversity found among genotypes of soybeans.

Around the world, soybeans are known for being a cheap source of protein and edible oil that can be fed to humans and animals. Brazil is currently the world's top producer of soybeans, followed by the US, Argentina, China, and India. Dietary habits are changing as a result of globalization, with meat and other high-value agricultural goods like vegetable protein and oil gradually replacing traditional mainstays. Due to this change, there is an increasing need for high-protein grains—especially those derived from soybeans—for animal feed as well as vegetable oil for human use. Soybeans are a vital commodity in global trade and are regarded as a superior feed supplement for animals with a single stomach because of their high protein content and well-balanced amino acid profile [3].

In Pakistan, commercial soybean farming began in the early 1970s, and subsequently, extensive work has been conducted for variety evaluation. However, Pakistan continues to struggle to meet the domestic demand for edible oil and other soybean products, and its global share of soybean production remains insignificant. As a result, the country has to import soybean products and edible oil. In 2018–2019, the production of soybean was only 0.002 million tons, whereas >2225.08 thousand tons of soybean seeds (worth PKR 123.0623 billion) and 150.91 thousand tons of soybean oil (worth PKR 14.832 billion) were imported. Although

Pakistan's soil and climate are ideal for growing soybean, the lack of known cultivars adapted to Pakistan has prevented them from becoming widely used by farmers [4]. Numerous factors contribute to the low production of soybeans in Pakistan as follows: (1) a lack of high-yielding and well-adapted cultivars suitable for different agro-ecological zones of the nation, as well as the growing seasons; (2) a lack of diverse soybean germplasm; (3) a lack of photo-insensitive soybean germplasm; and (4) competition for cultivation with already well-established crops in many regions of the country. However, there are possible locations with favorable agro-climatic conditions in which soybean can be grown [5].

Abiotic stressors have a substantial effect on soybean production because they cause disruptions to the plant's physiology and general growth, which results in a noticeable drop in yield [6]. These pressures frequently happen at the same time in nature; this is referred to as ”multistress.” For instance, high temperatures often coincide with water scarcity, aggravating the consequences of drought [7]. Moreover, the amount of moisture in different places used to cultivate soybeans is getting less and less conducive to normal growth and pod filling [8]. Worldwide, soils with hardpans are frequently used to cultivate soybeans [9]. Sandy soils in the coastal plains of the United States' southeast usually have a hardpan that prevents roots from penetrating and exploring, restricting their availability to water and nutrients and lowering yields [10]. Because these hardpans prevent plants from accessing stored water, the plants become more susceptible to dry spells [11]. Farmers frequently employ deep tillage, an energy- and labor-intensive, and environmentally unsustainable technique, to counteract soil compaction [12].

Since the compacted soil layer tends to re-form, deep tillage only provides temporary benefits. As a result, it's critical to establish an ideal root system that can support plant growth in spite of difficult circumstances like low soil nutrient and water availability. Plants depend on their roots for structural support as well as for the uptake of nutrients and water. Strong root systems let plants better survive abiotic stresses and challenging climatic circumstances, which is essential for soybean growth and yield . It is commonly known that root systems greatly improve nutrient and water utilization efficiency, particularly in legume crops such as soybeans [13].

Plant productivity under both ideal and suboptimal conditions is largely dependent on its root morphological traits (RMTs) and root system architecture (RSA) [14, 15]. It is also helpful to interpret plant responses to environmental cues and create disease management plans [16] when one is familiar with these traits.

Because of its complex structure and the changing conditions in which it grows, the root system of a plant—often referred to as the ”hidden half” is difficult to study. Standardizing root research is challenging due to the phenotypic plasticity of roots, which enables them to change their shape and physiology in response to environmental factors. The difficulty of extracting roots without causing damage and the lack of economical, effective screening methods exacerbate this dilemma [17–19]. Important markers that shed light on a plant's nutritional condition and reactions to outside stimuli are root morphological traits (RMTs) and root system architecture (RSA) [20]. Consequently, it is crucial to investigate RMTs and RSA in depth.

Plant morphology, seedling attributes, seed quality, and seed morphological features are common phenotypic parameters used to assess genetic variability in soybeans. Many researchers have emphasized the significance of phenotypic characterisation in evaluating genetic diversity in soybeans [21].The number of grains, pods, leaves, and nodules per plant, as well as the nodule dry weight, 100-seed weight, plant dry weight (shoot dry weight), root dry weight, harvest index, and seed yield per plant were among the characteristics that varied among the 34 mutant lines that examined [22]. Four groups were created from these genetics.

Similar to this, [23] used phenotypic parameters such as days to 50% flowering, days to maturity, plant height, number of branches, pods, seeds per plant, weight of 100 seeds, and yield per plant to examine the genetic variation in 55 different kinds of soybeans. In order to determine genotype variability, [24] looked at 52 morphological and agronomic characteristics. [25] carried out phenotypic characterisation of 139 genotypes of soybeans in a different study and discovered considerable variations in a range of characteristics.

The genetic diversity of 92 soybean genotypes was evaluated by research done by the Asian Vegetable Research and Development Centre (AVRDC) in cooperation with the US and Pakistan. In leaf area (44.8%), number of branches per plant (31.7%), pods per plant (29.5%), 100-seed weight (39.0%), and grain yield per plant (46.6%), they reported high coefficients of variation (CVs) and wide ranges, showing significant variety among the genotypes.

Three distinct sets of genotypes were identified, with Pakistan's germplasm constituting a separate cluster. Comparably, [26] looked at 42 genotypes and found seven clusters, noting that the most significant diversity was shown by the 100-seed weight, number of pods per plant, pod yield per plant, and seed yield per plot accounting for the largest phenotypic variation.

Notwithstanding possible environmental influences, phenotypic features are essential for genetic diversity research and agricultural improvement. In agricultural research, selection based on these qualities is still common and important for measuring genotype diversity using ANOVA. Elevated curriculum vitae and noteworthy variations in outcomes suggest a considerable room for choice.

In addition, the phenotypic data-derived clustering patterns, which show relatedness and variety in terms of the number of clusters and genotypes inside each cluster, are helpful. By utilizing agro-morphological and nutritional quality characteristics, this study seeks to ascertain the genetic diversity among genotypes of soybeans [27].

1.1 Problem Statement

The variations among different soybean genotypes in terms of their agricultural and morphological characteristics not completely addressed.

1.2 Scope

Soybean (*Glycine max*), is one of the most important crops in the world which provides key protein and oil for animal feed as well as human nourishment. Use of the genetic variety among soybean populations is essential for enhancing qualities including stress resistance, yield, and adaptability. A variety of genetic characteristics can improve the nutritional value of soybeans for consumption by humans and animals.

1.3 Aim and Objectives

The aim of this research is to comprehensively characterize and analyze the diversity among diverse local soybean genotypes based on agro morphological parameters and to identify promising genotypes among huge population.

1.3.1 Objectives

1. To study agro-morphological based diversity in diverse soybean genotypes for important quantitative traits.
2. To identify promising soybean genotypes with desirable traits for further breeding and crop improvement programs.

Chapter 2

Literature Review

One of the most important crops in the world is soybean (*Glycine max*), which provides key protein and oil for animal feed as well as human nourishment. Comprehending the genetic variety among soybean populations is essential for enhancing qualities including stress resistance, yield, and adaptability. Key indications of this diversity are agro-morphological parameters, which encompass a variety of features pertaining to plant structure, growth, and production. Researchers can determine the genetic basis of significant features by analyzing the genotype diversity of soybeans using these agro-morphological criteria. This allows for the development of focused breeding methods that will lead to the development of improved cultivars [28].

2.1 Genetic Diversity in Soybean

Since soybeans have been cultivated and domesticated for millennia, the species has a rich genetic variety reservoir. Geographic isolation, human-mediated breeding programs, and natural selection all contribute to this variety. To evaluate genetic diversity in soybean populations, morphological, molecular, and agronomic techniques have all been used. The allelic variances, population dynamics, and evolutionary links between various genotypes are revealed by these approaches, which help breeding programs that attempt to use this variety for agricultural improvement.

2.1.1 Importance of Genetic Diversity

Disease and Pest Resistance: By developing soybeans with resistance to a variety of diseases and pests, genetic diversity can help minimize the need for chemical treatments.

Environmental Adaptability: Soybeans' genetic diversity allows them to adjust to a variety of climates, soil types, and stressors like salinity and drought.

Enhancement of Yield: Through the use of genetic variety, breeding programs can combine advantageous features to produce increased yield and quality.

Nutritious Value: A variety of genetic characteristics can improve the nutritional value of soybeans for consumption by humans and animals [29].

2.1.1.1 Sources of Genetic Diversity

Wild Relatives: The Glycine soja wild soybean species are a major source of genetic diversity. They have characteristics that are frequently lacking in domesticated cultivars.

Landraces: Conventional soybean cultivars, cultivated for generations by farmers, may yield distinctive genetic characteristics useful for breeding initiatives.

Mutagenesis: New genetic variants can be produced by inducing mutations by chemicals or radiation. Genetic engineering: With the use of contemporary biotechnology, desired features can be created by introducing particular genes from other organisms [29].

2.1.2 Conservation of Genetic Diversity

Wild Relatives: The Glycine soja wild soybean species are a major source of genetic diversity. They have characteristics that are frequently lacking in domesticated cultivars.

Landraces: Conventional soybean cultivars, cultivated for generations by farmers, may yield distinctive genetic characteristics useful for breeding initiatives.

Mutagenesis: New genetic variants can be produced by inducing mutations by chemicals or radiation. Genetic engineering: With the use of contemporary biotechnology, desired features can be created by introducing particular genes from other organisms [30].

2.2 Techniques for Assessing Genetic Diversity

Molecular markers: To measure genetic variation at the DNA level, methods such as amplified fragment length polymorphism (AFLP), single nucleotide polymorphism (SNP), and simple sequence repeats SSR are employed.

Whole-genome sequencing offers comprehensive details on the genetic distinctions and overlaps between different soybean cultivars.

Phenotypic Analysis: This method evaluates genetic diversity by observing and quantifying physical characteristics, such as plant height, pod count, and seed size.

2.3 Agro-morphological Parameters in Soybean

The broad range of characteristics known as agro-morphological parameters work together to define the phenotype of soybean plants. Plant height, branching pattern, leaf morphology, blooming period, pod and seed features, and yield-related factors are a few examples of these traits, but there are many more. Every one of these factors influences how well soybean genotypes function overall and how adaptable they are to changing environmental situations. These characteristics' quantitative measurements and qualitative observations function as useful markers of genetic diversity and prospective breeding objectives [31].

2.4 Methods for Evaluating Diversity

To assess the genetic diversity of soybean genotypes based on agro-morphological factors, researchers use a variety of approaches. Commonly used methods include germplasm collections, field trials, and genetic analysis such molecular marker techniques. Field trials yield important information about how traits manifest in actual environments, and germplasm banks give access to a wide range of genetic variety for breeding and screening. SNPs (single nucleotide polymorphisms) and SSRs (simple sequence repeats) are two examples of molecular marker approaches that make precise genetic characterization possible and make it easier to identify genetic markers linked to desired attributes [31].

2.4.1 Morphological Evaluation

One of the simplest and most established ways to assess diversity in soybean germplasm is through morphological features. In order to do this, physical traits including plant height, leaf form, seed size, pod color, and flowering period must be evaluated. Both hereditary and environmental factors frequently affect these features. Although morphological assessment is very simple and economical, phenotypic plasticity—the strong influence of environmental factors on trait expression—means that it may not always reliably represent genetic variation. Nevertheless, it continues to be a useful first step in the characterisation of germplasm, giving breeders vital information. One essential morphological characteristic assessed in soybean germplasm is plant height. It gives important information about the genotype's adaptation to various farming systems and growth habit. It is measured from the base of the stem to the peak of the plant. Whether a plant has a determinate or indeterminate growth habit, it influences the timing of flowering and pod set as well as how the plant branches. Deciduous kinds have a more concentrated flowering time, while indeterminate varieties grow and produce flowers all through the growing season. These characteristics affect light interception, plant architecture, and eventually yield potential.

An additional crucial component of morphological assessment is leaf morphology, which includes dimensions, form, and color. In addition to being measured for length, width, and area, leaves are also evaluated for how they are arranged on the stem (phyllotaxy). Plant vigor and production can be impacted by changes in leaf shape, which can also have an impact on water usage and photosynthetic efficiency. Furthermore, cuticle thickness and leaf pubescence (hairiness) are assessed since they may affect a plant's ability to withstand pests and environmental stresses like heat and drought.

Understanding the reproductive behavior of soybean germplasm requires an understanding of flowering features, such as the length of flowering and the time it takes to produce a flower. Flowers are measured for their color, size, and shape, which provides markers for genetic variety and possible hybridization. Since they are closely correlated with yield, pod parameters like as length, width, and quantity of seeds per pod are analyzed. It is also observed how the pods are distributed throughout the stem and branches, as this influences the total production efficiency and harvest ability [31].

Essential markers of soybean quality and marketability are seed characteristics. Assessing a seed's morphological characteristics includes determining its weight, shape, color, and texture. Bigger seeds typically yield better germination rates and a higher market value. Features of the seed coat, like color and thickness, are also significant because they affect the seed's ability to fend against pests and illnesses. Certain seed qualities are essential for breeding efforts aimed at niche markets since they are sometimes connected to particular uses, including oil production or direct consumption.

In order to comprehend how stem characteristics, such as diameter, color, and pubescence, assist plant structure and resistance to lodging, they are analyzed.

The ability of the plant to tolerate severe winds and downpours can be influenced by the strength and flexibility of the stem. Although they are not as often evaluated in morphological assessments, root properties are just as significant. The plant's capacity to absorb water and nutrients, as well as its resistance to soil-borne illnesses and drought conditions, are assessed by looking at its root length, depth, and branching patterns [32].

2.4.2 Biochemical Markers

Isomers and seed storage proteins are two examples of biochemical markers that provide an additional level of diversity assessment. Isozymes are variations of enzymes that catalyze the same chemical reactions but have different amino acid sequences. These isozymes may be separated by electrophoresis, which reveals patterns that aid in genotype differentiation in soybeans. Similar to this, polymorphisms across various soybean lines can be found by examining seed storage proteins using methods like SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis). These biochemical markers can shed light on genetic diversity at the protein level and are quite easy to employ [32].

2.4.3 Molecular Markers

Molecular markers are the most accurate and thorough way to assess the diversity of soybean germplasm. These markers include DNA-based methods like SNPs (single nucleotide polymorphisms), amplified fragment length polymorphisms (AFLPs), microsatellites, restriction fragment length polymorphisms (RFLPs), and simple sequence repeats (SSRs). SSRs and SNPs are especially well-liked among them because of their high polymorphism and simplicity in automation. Basic Sequence Repetition SSRs, or short tandemly repeated DNA sequences, are found throughout the genome and are also referred to as microsatellites. They are helpful for determining genetic diversity and creating genetic linkage maps since they are highly polymorphic and co-dominant. SSR markers have been widely used in soybean breeding projects because they are comparatively simple to use.

SNPs are single base pair differences in the genome, whereas SSRs are short, repeating DNA sequences that differ in length among different genotypes. These molecular markers are crucial for genetic mapping and marker-assisted selection because they paint a clear and accurate picture of the genetic diversity and connections within soybean germplasm.

AFLPs use specific primers to amplify restriction fragments in order to identify DNA polymorphisms. This approach is very sensitive and can provide a large number of markers without any prior genome knowledge. When evaluating genetic variety, AFLPs are helpful, particularly in research involving sizable, diverse populations.

In RFLPs, genomic DNA is broken down using restriction enzymes, and then certain probes are hybridized with the DNA to identify polymorphisms. While more effective technologies have made RFLPs less frequent, they were among the first molecular markers employed in soybean research and were essential to early attempts at genetic mapping [32].

The advanced breeding methods known as marker-assisted backcrossing (MABC) and marker-assisted pyramiding mostly rely on molecular markers. In MABC, particular genes or quantitative trait loci (QTLs) are transferred from a donor variety to a superior recipient variety. Background selection for the recurrent parent genome is employed, and markers are used to monitor the presence of the targeted alleles. When adding features like drought tolerance or disease resistance to high-yielding types, this strategy works very well.

By using marker-assisted pyramiding, a single genotype is created by combining several genes or QTLs linked to distinct phenotypes. Breeders are able to create varieties with improved resilience and performance by employing markers to confirm the presence of all desirable alleles. For example, combining many resistance genes in a pyramid can produce soybean cultivars with long-lasting, broad-spectrum pest and disease.

Utilizing genome-wide marker data, genomic selection (GS) is a contemporary breeding strategy that forecasts individual performance within a breeding population. While MAS concentrates on a small number of important markers, GS takes into account the effects of every marker throughout the entire genome. This all-encompassing method enables more precise predictions of complex features influenced by multiple genes. By facilitating the identification of superior genotypes early in the breeding process, GS in soybean breeding can dramatically reduce the breeding cycle and boost genetic gain.

Breeding and genetics for soybeans have been transformed by the application of molecular markers. Markers facilitate more accurate and effective breeding techniques by offering comprehensive insights into the genetic basis of significant traits. Additionally, they aid in the preservation of genetic variety by guaranteeing the preservation and utilization of important alleles. The combination of molecular markers with other genetic methods holds the potential to further improve soybean crop resilience, productivity, and sustainability as genomic technologies progress[33].

2.4.4 Genomic and Transcriptomic Approaches

Advances in next-generation sequencing (NGS) technologies have made transcriptome analysis and whole-genome sequencing more practical for investigations of soybean diversity. The detection of novel alleles and structural variations is made possible by whole-genome sequencing, which offers comprehensive data on genetic diversity throughout the entire genome. The study of RNA sequences through transcriptome analysis aids in the comprehension of gene expression patterns and their role in phenotypic variability.

A significant accomplishment was the sequencing of the soybean genome, which produced a reference genome that is used as the basis for many genomic research projects. Researchers are now able to detect and define genes, regulatory elements, and genomic variants because to the availability of high-quality genome sequences. Thanks to developments in next-generation sequencing (NGS) technologies, the genetic diversity found in germplasm collections can now be shown through the sequencing of several genotypes of soybeans. Finding the alleles linked to desired qualities like yield, disease resistance, and stress tolerance requires this knowledge.

A useful method for establishing a connection between genetic variation and phenotypic features is genome-wide association studies, or GWAS. Using genomic analyses of vast populations of soybeans, GWAS is able to pinpoint single nucleotide polymorphisms (SNPs) and other markers linked to particular features. Using this method, many QTLs (quantitative trait loci) associated with significant agronomic qualities have been found, offering useful targets for marker-assisted selection (MAS) in breeding initiatives [34].

A high-throughput method for profiling gene expression throughout the entire genome is RNA sequencing, or RNA-Seq. To study the transcriptional responses to various biotic and abiotic stimuli, developmental stages, and environmental conditions in soybeans, RNA-Seq has been used. Through the identification of differentially expressed genes (DEGs) and the clarification of gene regulatory networks, this method illuminates the molecular processes that underlie growth, development, and stress responses.

Transcriptomics-based gene expression profiling makes it possible to identify potential genes linked to particular characteristics. Through transcriptome analysis of soybean varieties with differing phenotypes, scientists are able to identify the genes responsible for traits like drought resistance, disease resistance, and efficient utilization of nutrients. These putative genes are used as targets for genetic engineering and functional validation.

Transcriptomics data is used with other genomic data in functional genomics to uncover the functions of individual genes in biological processes. To confirm the roles of potential genes, methods like RNA interference (RNAi), CRISPR/Cas9 gene editing, and overexpression research are employed. Functional genomics has made it easier to characterize the genes in soybeans that are involved in pathways like lipid metabolism, symbiotic nitrogen fixation, and nodulation. These insights can be used to improve the crop.

The use of transcriptomics and genomics in soybean germplasm research has significant effects on crop development and breeding. By identifying genetic markers and potential genes for MAS, these methods expedite the creation of better soybean cultivars. Additionally, they aid in the identification of new genes and pathways that can be the focus of genetic engineering to improve qualities of nutrition, stress tolerance, and yield.

Moreover, by comprehending the transcriptional reactions to environmental stressors, breeders can create cultivars that are more adaptable to shifting climate circumstances. Through the application of genomics and transcriptomics expertise, breeders are able to develop soybean varieties that satisfy the needs of both global food security and sustainable agriculture [34].

2.4.5 Phenotypic and Genotypic Data Integration

The accuracy and efficiency of evaluating diversity are improved when phenotypic data, or morphological qualities, are integrated with genotypic data, or molecular markers, using methods like genome-wide association studies (GWAS) and genomic selection. Genes associated with significant agronomic qualities can be found more easily thanks to the connections that GWAS finds between genetic markers and desirable phenotypes. By predicting an individual's breeding value using genome-wide marker data, genomic selection speeds up the creation of better soybean varieties.

To identify the genetic architecture underlying complex traits, genotypic and phenotypic data integration combines genetic information with trait measures. Advanced statistical and computational techniques, including multi-trait analysis, genomic selection (GS), and genome-wide association studies (GWAS), enable this integrative approach.

GWAS is an effective method for connecting genetic markers in a broad population with phenotypic features. Through simultaneous analysis of genotypic and phenotypic data, GWAS is able to pinpoint SNPs and QTLs associated with particular traits. Genetic loci linked to characteristics like disease resistance, drought tolerance, and seed composition have been found in soybeans using GWAS. These results offer useful targets for functional genomics and MAS research [35].

Genome-wide markers are used in genomic selection (GS) to forecast an individual's breeding value for complex traits. Compared to conventional selection techniques, GS models provide more accurate predictions of trait performance by combining genotypic and phenotypic data to reflect the cumulative effects of numerous small-effect loci.

In order to hasten the production of high-yielding, stress-tolerant cultivars, GS has been effectively used in soybean breeding programs By taking into account the genetic connections between several qualities at once, multi-trait analysis offers a more thorough comprehension of trait interactions and trade-offs. By combining genotypic and phenotypic data for several traits, breeders can enhance genetic gain by selecting for the best possible trait combinations. Multi-trait analysis has been applied to soybeans to enhance characteristics like protein content, seed yield, and disease resistance [35].

Although there are many advantages to integrating genotypic and phenotypic data, there are drawbacks as well. Because of the intricacy of assessing some qualities and the diversity of the environment, accurate phenotyping is still a bottleneck. Additionally, powerful bioinformatics tools and computer resources are needed for analysis because to the massive number of data created.

Research on soybean germplasm will probably go in the following directions: enhancing multi-omics methods, creating more complex models for data integration, and improving phenotyping methods. The next generation of soybean improvement will be fueled by the integration of transcriptomic, proteomic, and metabolomic data with genotypic and phenotypic data to provide a more comprehensive understanding of the molecular mechanisms underpinning complex characteristics [35].

2.5 Factors Influencing Diversity

Numerous factors, including genetic inheritance, breeding history, environmental conditions, and human selection pressures, all have an impact on the diversity exhibited within soybean genotypes. The genetic variety seen in soybean populations is influenced by various genetic processes, including introgression, mutation, and gene pools. Furthermore, selected forces shaped by environmental factors including soil properties, climate, and management techniques influence how agro-morphological traits manifest [36]. Understanding soybean variety is made more difficult by the interaction between genetics and environment, which emphasizes the necessity for thorough and integrated study methods.

2.5.1 Evolutionary Processes

The genetic diversity of soybean populations is mostly shaped by natural evolutionary processes such genetic drift, gene flow, mutation, and natural selection. Through changes to DNA sequences, mutations bring about new genetic variations. While the majority of mutations are harmful or neutral, some can confer beneficial features that increase the fitness of a plant. Gene flow, or the transfer of genes between populations via seed and pollen distribution, encourages genetic mixing and keeps gene pools from becoming isolated. Genetic diversity can be greatly reduced in small populations due to genetic drift, which is the random variation of allele frequencies over generations. Natural selection promotes characteristics that increase the likelihood of survival and procreation, increasing the frequency of beneficial alleles in the population.

2.5.2 Mutation

The primary cause of genetic diversity is mutation. Nucleotide sequences in soybeans can alter as a result of spontaneous mutations that can happen during DNA replication. These mutations may be advantageous, inert, or harmful. While neutral mutations gradually increase the genetic pool without immediately affecting fitness, beneficial mutations can offer benefits like enhanced stress tolerance or disease resistance. Genetic diversity in soybean populations is maintained by the constant introduction of novel alleles brought about by mutations [36].

2.5.3 Gene Flow

Genetic diversity in soybeans is greatly influenced by gene flow, or the exchange of genetic material between populations. Seed movement and pollen dissemination can happen naturally or as a result of human intervention. A population's genetic diversity is increased through the introduction of new alleles through cross-pollination between several soybean types or wild cousins. In agricultural contexts, the transfer of germplasm between nations and regions fosters additional gene flow, which helps soybeans adapt to a variety of environmental circumstances and increases crop resilience.

2.5.4 Genetic Drift

Random variations in allele frequencies within a population, especially in small populations, are referred to as genetic drift. Genetic drift in soybeans, particularly in isolated or bottlenecked populations, can result in the loss of genetic variety over generations. The overall genetic variety may be impacted by this stochastic process, which can lead to some alleles becoming fixed and others being lost completely. Genetic drift, however, can also promote the emergence of distinct genetic variants that aid in local adaptation and population differentiation.

2.5.5 Natural Selection

Natural selection favors alleles that give adaptive advantages by acting on the genetic variety found in soybean populations. Different settings impose selected pressures on traits including drought tolerance, disease resistance, and high yield. The genetic makeup of soybean populations is shaped throughout time by the increasing frequency of genes linked to these beneficial features. In addition to increasing soybean fitness, natural selection promotes species diversification by allowing various populations to adapt to distinct ecological niches [37].

2.5.6 Breeding Practices

Soybean genetic diversity is greatly impacted by human-mediated breeding methods. In traditional breeding, plants with desirable qualities are generally chosen, and then they are cross-bred to create superior kinds. However, if the breeding pool is restricted to a small number of elite lines, this may result in a decrease in genetic diversity. Although more accurate and effective breeding is made possible by modern methods like genomic and marker-assisted selection, maintaining genetic variety still requires careful management. To preserve crop resilience and adaptation over the long term, breeders must strike a balance between the requirement for uniformity and specific qualities and the maintenance of a broad genetic foundation. By enabling early selection during the seedling stage, marker-assisted selection greatly accelerates the breeding process by identifying and choosing plants that carry advantageous genes. On the other hand, genetic modification entails making direct changes to the soybean plant's DNA in order to introduce particular features like resistance to herbicides or increased nutritional value.

After cross-pollination, the offspring are rigorously assessed over several generations to guarantee that the desirable features are stable and heritable. This entails outdoor testing in a range of environmental settings to evaluate adaptability and performance. Furthermore, genomic selection—which makes use of genome-wide genetic data to forecast a plant's breeding value—is included into sophisticated breeding techniques, which increases the effectiveness and accuracy of creating superior soybean varieties.

Additionally, breeding programs preserve genetic variety within the germplasm to mitigate the hazards that come with monocultures, such as pest and disease susceptibility. In order to ensure a broad genetic basis for next breeding efforts, this variety is maintained through the gathering and conservation of soybean genetic resources from various geographic regions and wild relatives.

2.5.7 Domestication and Cultivation

The genetic diversity of soybeans has been greatly influenced by their domestication from their wild progenitors. Choosing plants for domestication usually entails choosing ones with characteristics that are advantageous for farming, like bigger seeds, unbreakable pods, and consistent maturation. Even while these characteristics are useful for farming, domestication frequently results in a genetic bottleneck that lowers genetic diversity overall. The genetic base can also be further reduced by intensive farming methods like monoculture and the widespread use of a small number of high-yielding cultivars. Soybean crops become more susceptible to pests, diseases, and environmental changes as a result of this loss of diversity.

2.5.8 Geographical and Environmental Factors

The genetic diversity of soybeans has been greatly influenced by their domestication from their wild progenitors. Choosing plants for domestication usually entails choosing ones with characteristics that are advantageous for farming, like bigger seeds, unbreakable pods, and consistent maturation. Even while these characteristics are useful for farming, domestication frequently results in a genetic bottleneck that lowers genetic diversity overall. The genetic base can also be further reduced by intensive farming methods like monoculture and the widespread use of a small number of high-yielding cultivars. Soybean crops become more susceptible to pests, diseases, and environmental changes as a result of this loss of diversity.

The main environmental factor affecting the growth of soybeans is the climate. Important factors influencing soybean development phases like germination, flowering, and pod filling include temperature, rainfall, and photoperiod (day duration). Because soybeans are extremely susceptible to temperature fluctuations, breeders concentrate on creating cultivars that can tolerate cold temperatures in temperate zones and heat stress in tropical regions. Furthermore, sufficient irrigation or rainfall is necessary for the best possible soybean yield, which is why drought-resistant cultivars are a major priority in areas where water is scarce.

Soybean performance is also influenced by fertility and soil type. The ability of various soils to hold onto water and nutrients affects the growth and productivity of soybeans. Well-drained soils that are high in organic matter and vital nutrients like potassium, phosphate, and nitrogen are necessary for soybean growth. Improved nutrient uptake efficiency and resistance to soil-related stressors like salt and acidity are the goals of soybean breeders. This is especially crucial in regions with marginal soils, where conventional soybean cultivars could not yield satisfactory results.

Pests, illnesses, and weed competition are examples of biotic variables that soybean germplasm must overcome in the environment. In order to create new kinds of soybeans resistant to common pests and illnesses like rust, soybean cyst nematode, and different fungal infections, breeders of soybeans strive to achieve this. This incorporates both contemporary biotechnological methods like genetic engineering and marker-assisted selection as well as conventional breeding procedures. Another major challenge is weed competition, which is why herbicide-resistant soybean cultivars have been developed to help with successful weed management[37].

2.5.9 Conservation Efforts

The creation of germplasm banks, in situ conservation, and the utilization of wild relatives are among methods used to preserve the genetic variety of soybeans. In order to ensure the preservation of genetic resources for next breeding and research, germplasm banks preserve a broad variety of genetic material related to soybeans, including landraces, traditional varieties, and wild species. In situ conservation refers to preserving the diversity of soybeans in their natural environment so that populations can continue to evolve and adapt. By introducing useful genes into cultivated varieties, wild relatives of soybeans—who frequently have distinctive features and a higher genetic diversity—can increase the durability and adaptability of cultivated kinds.

2.5.10 Biotechnological Advances

Genetic engineering and genome editing are two recent biotechnological developments that provide new means of increasing soybean genetic variety. With the use of methods like CRISPR-Cas9, it is possible to precisely modify particular genes, introducing new features and removing unwanted ones. By preserving or even enhancing genetic diversity, these technologies help hasten the generation of soybean cultivars with better yield, stress tolerance, and nutritional value. However, careful consideration of ethical, ecological, and regulatory concerns is necessary for the appropriate use of these technologies.

2.6 Agro-morphological Diversity in Soybean Genotypes

There is a wealth of research on studies examining the genotype diversity of soybeans using agro-morphological factors. These studies show that different soybean genotypes differ significantly in terms of factors like plant height, branching pattern, leaf morphology, flowering time, pod and seed characteristics, and yield-related traits. A genotype may exhibit a higher frequency of a particular characteristic, such as early flowering time or resistance to a particular pest or disease, as a result of adaptations to the local environment or breeding goals. Geographic diversity patterns also demonstrate how human selection and environmental variables affect the genotype distributions of soybeans [10].

2.6.1 Growth Parameters

A key growth metric that indicates a soybean plant's general vigor is plant height. Plant height, which is measured from the base to the top of the main stem, can affect other parameters like lodging resistance, as taller plants may be more likely to topple over in severe weather or high winds. Important growth criteria include also the number of days till flowering and maturity. The duration required for seeds to fully develop and become ready for harvest (days to maturity) and the number of days from planting to the first flowers appearing (days to flowering) are important indicators of the growth cycle and adaptability of soybean varieties to various climatic conditions and growing seasons [38].

2.6.2 Yield Parameters

Soybean crop productivity is directly impacted by yield characteristics. Two important measures of prospective yield are the number of pods per plant and the number of seeds per pod. These factors impact total yield by dictating a plant's seed output.

Seed weight, which is commonly determined by weighing one hundred seeds, is a significant factor that influences both market choice and seed quality. The most comprehensive indicator of a variety's productivity is its total seed yield, which combines the effects of all yield characteristics and represents the total weight of seeds produced per unit area.

2.6.3 Morphological Parameters

In the growth and development of soybean plants, morphological traits such leaf size and form, stem thickness, and branching pattern are important. The dimensions of a leaf, such as its length, width, and area, have an impact on the photosynthetic efficiency and growth of the plant. The diameter of the main stem, or stem thickness, is a measure of the structural strength and lodging resistance of the plant.

The architecture of the plant and its capacity for pod development are influenced by the branching pattern, which specifies the quantity and configuration of branches on the main stem.

2.6.4 Physiological Parameters

Water utilization efficiency and chlorophyll content are two examples of physiological characteristics that are essential for evaluating the production and health of soybean plants. Using a spectrophotometer or chlorophyll meter, one may measure a plant's chlorophyll concentration, which provides an indication of its general health and photosynthetic capacity. The ratio of biomass generated to water consumed is known as water use efficiency, and it is an essential component of drought tolerance and effective water use, particularly in areas with limited water resources. One essential mechanism that directly affects plant development and yield is photosynthesis. Variations in photosynthetic efficiency across different germplasms of soybeans have been thoroughly researched. Studies have demonstrated that genotypes of soybeans can differ considerably in photosynthetic rates, which affects biomass production and yield potential. For instance, research by Sun et al. (2014) showed that improved light-harvesting capacity and higher chlorophyll content were linked to specific genotypes of soybeans with higher photosynthetic efficiency. These results imply that one practical method for increasing soybean output would be to select for characteristics linked to enhanced photosynthesis.

A vital physiological measure is water use efficiency (WUE), especially in areas where water scarcity is a possibility. The ratio of biomass produced to water consumed is known as WUE. There are genotypic variations in WUE in soybeans, with certain genotypes showing higher WUE during droughts. Root architecture, transpiration rates, and stomatal conductance are all strongly correlated with this characteristic. According to research by Devi et al. (2014), genotypes of soybeans with deeper root systems and lower stomatal conductance sustained higher yields and better WUE during drought stress. Therefore, breeding for improved WUE is crucial to creating soybean types that can withstand drought.

For the best possible development and output from soybeans, efficient nutrient uptake and use are essential. Given that soybean plants rely on symbiotic nitrogen fixation, nitrogen (N) is very crucial. Soybean germplasm variations in nutrient uptake efficiency can affect the total yield of the plants. For example, increased biomass and seed yield are typically found in genotypes with more effective nitrogen fixation and absorption (Van Berkum & Sloger, 2014).

Other vital elements that affect soybean growth are phosphorus (P) and potassium (K). Research has demonstrated that in P-deficient environments, some soybean genotypes exhibit improved root exudation features that boost P acquisition from the soil, resulting in greater growth and production [38].

Many physiological mechanisms, including as hormone regulation, leaf area development, and canopy design, affect the growth and development of soybean plants. The growth rates and yield potential of different soybean germplasms can be greatly impacted by variations in these characteristics. Larger leaf areas and optimized canopy structures in soybean genotypes resulted in higher light interception and photosynthetic rates, which raised biomass and seed output, according to studies by Purcell et al. (2014).

Furthermore, hormone modulation is essential for controlling the growth and development of soybean plants, especially when it comes to auxins, gibberellins, and cytokinins [38].

2.6.5 Biotic and Abiotic Stress Parameters

To sustain soybean health and yield, resistance to biotic and abiotic stressors is crucial. Assessing a plant's resistance to common soybean diseases such as soybean rust, Phytophthora root rot, and soybean cyst nematode is known as disease resistance.

The ability of a plant to withstand insect pests like soybean pod borer and aphids is measured as pest resistance. Furthermore, it is essential to have tolerance for abiotic stimuli including heat, cold, and drought. Heat and cold tolerance refers to a plant's capacity to tolerate extremes in temperature, whereas drought tolerance describes a plant's ability to continue growing and producing when faced with water constraints.

2.6.6 Agronomic Parameters

Practical cultivation and harvesting depend on agronomic factors such lodging resistance, harvest index, and plant architecture. The plant's capacity to stay upright, or lodging resistance, influences production and harvesting ease. The ratio of seed yield to total biomass, or harvest index, measures how well a plant allocates its resources to seed production. Plant architecture, which includes the plant's overall form and structure, affects airflow, light interception, and compatibility for various planting densities.

Plant height is an important and well-researched agronomic characteristic. Plant height in soybeans is regulated by a number of genes, and changes in this characteristic have a major effect on photosynthetic efficiency and biomass accumulation [39]. According to studies, choosing plants with the ideal height will enhance light interception and decrease lodging, which would boost crop performance. Through quantitative trait locus (QTL) mapping, the genetic basis of plant height has been investigated. Numerous loci have been found to be related with this attribute.

The most important economic characteristic, seed yield, has been thoroughly studied to identify the genetic and environmental factors that influence it. Seed production is a complicated attribute that is regulated by a multitude of factors, including genetics, environment, and their interactions. In breeding projects, the genetic variability in seed yield across soybean germplasms has been a key focus. To find the parts of the genome linked to high yield, scientists have employed QTL mapping and genome-wide association studies (GWAS) .These findings highlight how crucial it is to combine conventional breeding methods with molecular genetics in order to create high-yielding soybean cultivars.

Another crucial factor that affects how well soybeans adapt to various climate zones is their maturity date. Multiple genes and environmental interactions play a complicated role in the genetic control of maturity, as demonstrated by studies conducted by Jin et al. (2010). Varieties that mature early and late offer crop management flexibility and can lessen the dangers brought on by climatic unpredictability. Numerous maturity genes, including E1, E2, E3, and E4, have been found via research to be important regulators of flowering and maturation timeframes. These results are important for breeding initiatives that try to increase crop resilience and lengthen the growing season.

Numerous research have examined the relationship between yield and pod quantity and seed size. Different soybean germplasms differ significantly genetically in terms of seed size and pod number. These characteristics are polygenic and impacted by environmental and genetic variables. In order to improve these traits using marker-assisted selection (MAS), research has concentrated on identifying QTLs linked to pod quantity and seed size [40]. Research has also examined the physiological foundation of these characteristics, connecting them to the partitioning of nutrients and photosynthetic efficiency.

An essential agronomic factor that guarantees the health and productivity of soybeans is disease resistance. The genetic foundation of resistance to significant soybean diseases, including Asian soybean rust, Phytophthora root rot, and soybean cyst nematode, was examined by Hartman et al. (2015).

In order to reduce the need for chemical controls, resistant cultivars have been developed by utilizing the genetic variability in disease resistance among soybean germplasms. Resistance genes have been found and added to breeding programs using genomic technologies and molecular markers.

2.6.7 Implications for Breeding and Crop Improvement

Comprehending the heterogeneity in soybean genotypes' agro-morphology has significant consequences for breeding and agricultural development initiatives. Breeders can deliberately introduce genetic diversity into breeding programs to create cultivars that are hardy, high-yielding, and suited to particular environmental circumstances by identifying genotypes that exhibit desired qualities. To harness soybean variety for sustainable agriculture and food security, many breeding procedures are employed, such as trait introgression, hybridization, and genomic selection. Furthermore, the long-term availability of diverse genetic material for next breeding projects is ensured by the preservation and utilization of germplasm, which contributes to the conservation of genetic resources.

2.6.8 Targeted Breeding for Yield Improvement

Most breeding programs have yield as their main objective, and agro-morphological characteristics like the number of pods per plant, the weight of the seeds, and the number of seeds per pod are crucial in achieving this goal. Breeders can greatly increase overall production by choosing plants with higher pod and seed counts as well as favorable seed weights.

Because these characteristics are frequently quantitative and influenced by several genes, it is necessary to employ sophisticated breeding methods, including marker-assisted selection (MAS), in order to effectively discover and incorporate these desired features. High-yielding cultivars that meet the needs of both commercial and subsistence farming can be generated via strict selection and cross-breeding.

2.7 Improving Nutritional and Industrial Qualities

Since soybeans are an essential source of both oil and protein, breeding initiatives frequently aim to enhance these nutritional attributes. For both industrial and human consumption, parameters including the profile of amino acids, oil composition, and protein concentration in seeds are crucial.

Breeders can choose for soybean lines with improved nutritional profiles, which will help consumers and the food sector, by knowing the genetic basis of these qualities. Furthermore, soybeans with particular oil compositions can be created for industrial uses, such the manufacturing of biodiesel, increasing the crop's market value and adaptability.

2.7.1 Incorporating Diverse Germplasm

Successful breeding programs are built on genetic variety, which offers a pool of features that can be used to improve crops. Agromorphological analyses of a variety of germplasm, including landraces and wild relatives, are useful in identifying distinctive features that can be incorporated into contemporary cultivars. By improving the genetic foundation of grown soybeans, this method lessens the plants' susceptibility to pests, illnesses, and environmental challenges. Breeders can guarantee soybean crops' long-term adaptability and durability by protecting and exploiting genetic variety.

2.7.2 Integrating Advanced Technologies

Soybean breeding is being revolutionized by the combination of cutting-edge technologies including phenomics, genomics, and bioinformatics. Accurate assessment of agro-morphological parameters is made possible by high-throughput phenotyping platforms, and the identification of genes linked to desired traits is made possible by genomic techniques. Large datasets may be analyzed more easily thanks to bioinformatics, which also helps to identify intricate genetic relationships and direct selection processes. By improving breeding programs' precision and efficiency, these technologies hasten the creation of better soybean varieties.

2.7.3 Addressing Socio-Economic Needs

The socioeconomic requirements of farmers and consumers must be taken into account in breeding projects. Agro-morphological factors that lower labor costs and increase efficiency, like plant architecture and lodging resistance, are crucial for robotic harvesting. Farmers can increase their marketability and profitability by selecting varieties with characteristics, including seed size and color that cater to particular market preferences. Breeders can create commercially feasible and agronomically superior soybean varieties by coordinating breeding objectives with socioeconomic realities. Another important factor in the socioeconomic impact of soybean germplasm is consumer nutrition and health. Because they are high in important fatty acids, protein, and other vitamins and minerals, soybeans are a valuable part of diets all around the world. Better health outcomes for consumers can result from breeding efforts targeted at improving the nutritional profile of soybeans, such as raising protein content or balancing the necessary amino acid composition. Moreover, the development of soybeans with decreased allergic qualities and anti-nutritional components can increase their accessibility and attractiveness, encouraging healthier diets among various demographic segments.

Breeding projects can help the industrial sector, which depends largely on soybeans for goods like plastics, food additives, and biodiesel, as well as special socioeconomic demands. The efficiency and sustainability of these businesses can be improved by creating soybean varieties with characteristics specifically suited for industrial use, including a high oil content or particular fatty acid profiles. By encouraging the use of renewable resources and lowering the dependency on fossil fuels, this promotes economic growth while simultaneously contributing to environmental goals.

When developing new soybean germplasm, equity and inclusivity are crucial factors to take into account. Collaboration between extension agencies, legislators, and breeders is necessary to guarantee that the advantages of novel soybean varieties reach underprivileged and marginalized areas. This entails facilitating market integration, educating people about proper farming techniques, and granting access to improved seeds. Soybean breeding projects can help reduce poverty, improve gender equity, and empower disadvantaged groups through promoting inclusive agricultural growth [41].

2.8 Challenges and Future Directions

Studying the variety of soybean genotypes based on agro-morphological factors continues to present a number of difficulties despite improvements in research methodology and technologies. To maximize the use of genetic resources and enable comparative studies, standardization of measuring methodologies, data exchange procedures, and integration of genomic tools are required. Furthermore, multidisciplinary cooperation and comprehensive research methods are needed to address the intricate relationships between genetics and environment. Potential avenues for future study could include utilizing cutting-edge tools like gene editing, high-throughput phenotyping, and predictive modeling to improve our comprehension of the diversity of soybeans and how it affects sustainable agriculture [41].

Developing superior soybean varieties depends heavily on the genetic material utilized in breeding and research—soybean germplasm. Nevertheless, a number of obstacles make its efficient use and preservation more difficult. The limited genetic base of commercial soybean cultivars presents a major obstacle. Crops are more susceptible to pests, illnesses, and environmental pressures as a result of their low genetic diversity. A limited fraction of the germplasm is frequently used in breeding attempts, which can impede the creation of new varieties with desired characteristics including increased yield, better nutritional quality, and tolerance to biotic and abiotic challenges.

The effect of climate change on the production of soybeans presents another difficulty. Soybean cultivation is threatened by changing climatic circumstances, such as temperature swings, modified precipitation patterns, and an increase in the frequency of extreme weather events. To create cultivars that can withstand these shifting environments, a thorough comprehension of the genetic processes underpinning stress tolerance is necessary. This is a difficult undertaking, though, because of how intricately these features interact with the environment. Stress-resilient genes can be found and incorporated into soybean varieties with the use of sophisticated breeding techniques, such as genomics-assisted breeding, but this needs a significant investment in infrastructure and research.

There are financial and practical obstacles to the preservation of soybean germplasm. Significant resources are needed to maintain a varied and easily available germplasm collection, including appropriate storage facilities, ongoing observation, and classification. Funding and infrastructural issues plague many genebanks worldwide, which can result in the loss of priceless genetic material. Additionally, the global pool of genetic material available for breeding and study might be reduced by geopolitical difficulties and legislative barriers that impede the flow of germplasm between nations[41].

Future directions for soybean germplasm center on utilizing technology breakthroughs and promoting global cooperation. Technologies for high-throughput genotyping and phenotyping can hasten the discovery of desirable features and their underlying genetic foundation. Technological developments in biotechnology, like the CRISPR-Cas9 gene editing system, present an opportunity to apply precise genetic alterations, hence improving desirable qualities without requiring conventional cross-breeding. Furthermore, better administration and sharing of germplasm information can be facilitated by digital tools and databases, which will make it simpler for researchers and breeders to access and use these resources.

To overcome the difficulties with soybean germplasm, international cooperation is essential. The genetic diversity available for breeding programs can be increased by initiatives that facilitate the cross-border exchange of germplasm, knowledge, and technologies. Additionally, collaborative research projects can pool resources and expertise to more efficiently address complex concerns like disease resistance and climate resilience. To guarantee the sustainable development of soybean varieties that can satisfy the rising demand for this significant commodity on a worldwide scale, it will be imperative to reinforce frameworks and regulations that encourage the exchange and conservation of germplasm.

2.9 Soybean Imports and Exports in Pakistan

2.9.1 Soybean Imports in Pakistan

Following three years of expansion, soy bean imports fell by 34% to 1.6 million tons in 2022. Overall, imports experienced a sharp decline. With a 15% increase, 2021 saw the highest growth rate on record. As a result, imports peaked at 2.5 million tons and then sharply decreased the following year. Imports of soy beans fell to $1 billion in value in 2022. Total imports showed strong growth from 2019 to 2022 over the time under study; during the previous three years, its value grew at an average annual rate of +10.7%. On the other hand, the trend pattern showed several distinct variations that were noted over the course of the analysis. Compared to 2019 indexes, imports rose by +35.5% based on 2022 statistics. The greatest notable growth rate was seen in 2021, when imports rose by 62% over the prior year.

Top Suppliers of Soya Bean to Pakistan in 2022:

- Brazil (1064.3K tons)
- United States (556.3K tons) [41]

2.9.2 Soybean Exports in Pakistan

Pakistan's soybean exports fell sharply to 186 tons in 2022, a decrease of -44.7% from the previous year. Exports experienced a severe decline overall. The year 2021 saw the fastest growth rate, with exports rising by 363% over the previous year. As a result, exports peaked at 336 tons and then drastically decreased the next year. Soybean exports fell dramatically in value to $127K in 2022. There was a significant decline in exports over the time under consideration. The year 2021 saw the fastest growth rate, with exports rising by 401% [42].

Top Export Markets for Soya Bean from Pakistan in 2022:

- Sri Lanka (100.0 tons)
- Japan (70.8 tons)
- Afghanistan (14.5 tons) [42]

2.10 Worldwide Production of Soybean

Only five countries worldwide produce more than 90% of the world's soybeans, with South American nations accounting for more than 51% of global soybean production.

Notably, 86% of the continent's total soybean production in 2021 came from the five largest producers of soybeans in Europe, which includes the Russian part of Europe.

Global soybean production is expected to increase from 359.80 metric tons (Mt) in 2021–2022 to 369.64 Mt in 2022–2023, according to the US Department of Agriculture.

Because soybean production is geographically concentrated in a small number of places, harvest losses can have a substantial impact on the world supply of soybeans. In addition, soybeans are particularly vulnerable to hot, dry weather during their reproductive stage of growth in the summer, which may have an impact on output. In recent years, Brazil has outproduced the US in terms of soybean production. More than 80% of the world's soybean supply is produced in the US, central Brazil, and southeast South America put together. In addition to these major players, soybean production is also expanding in other regions, such as India, Paraguay, and parts of Africa. These regions are increasingly adopting soybean cultivation to enhance food security and tap into the growing demand for soy products. While their production volumes are currently lower than the top producers, these areas hold potential for future growth, especially with investments in agricultural infrastructure and technology.

Overall, the worldwide production of soybeans continues to evolve, driven by technological advancements, expanding markets, and the need for sustainable practices. As global demand for protein and renewable resources grows, soybeans are likely to remain a critical component of the agricultural landscape, influencing economic, environmental, and social outcomes across the globe[42].

2.10.1 Brazil

Brazil is anticipated to produce an all-time high of 154.0 million Mt of soybeans for the marketing year 2022–2023. This was 130.50 MT in 2021–2022, the year the nation surpassed the US to become the world's top producer of soybeans.

2.10.2 United States

The US produced 114.75 Mt of soybeans in 2020–2021. More than 121 million tons were recorded in 2021–2022, and 116.38 million tons are anticipated to be produced in 2022–2023.

2.10.3 Argentina

For the marketing year 2022–2023, the nation is expected to produce 27.0 million Mt of soybeans, a 38% decrease from the previous year. The nation produced over 44.90 million tons of soybeans in 2021–2022, compared to 46.20 million tons the year before. The production has dropped month after month in March 2023 as a result of the hot, dry weather as the soybean crop was unable to benefit from the rains that occurred in the second part of the month.

2.10.4 China

China's soybean production is predicted to reach 20.28 million tons in 2022–2023, down from 16.4 million tons in the previous year and roughly three million tons less than in 2020–2021, when it reached 19.6 million tons. However, according to the most recent data, production has increased significantly. Despite being one of the countries that produces the most soybeans, China is one of those that imports a lot of soybeans to meet domestic demand.

2.10.5 India

India is predicted to produce 12 million tonnes of soybeans, almost the same amount as the previous year (11.89 million tonnes), and more than 1.5 million tonnes greater than the 10.45 million tonne production in 2020–2021 [42].

Chapter 3

Materials and Methods

3.1 Plant Materials

Thirty one local genotypes and one check genotype of soybean were collected from different areas of Pakistan and the detail of these genotypes is given in the Table 3.1.The Oilseeds Research Program, National Agricultural Research Centre (NARC), Islamabad, Pakistan, and several growing regions in Pakistan provided the experimental seed samples.

TABLE 3.1: The detail of collected information of Soybean (*Glycine max*) Germplasm used

Serial No	Accession	Source
1	PJB-2005	Punjab
2	PJB-2006	Punjab
3	PJB-2009	Punjab
4	PJB-2016	Punjab
5	PJB-2017	Punjab
6	PJB-2021	Punjab
7	PJB-2027	Punjab
8	PJB-2028	Punjab
9	PJB-2030	Punjab
10	PJB-2034	Punjab

Table 3.1 continued from previous page

Serial No	Accession	Source
11	PJB-2037	Punjab
12	PJB-2043	Punjab
13	PJB-2047	Punjab
14	PJB-2049	Punjab
15	PJB-2057	Punjab
16	PJB-2068	Punjab
17	PJB-2078	Punjab
18	PJB-2083	Punjab
19	PJB-2090	Punjab
20	PJB-2094	Punjab
21	PJB-2101	Punjab
22	PJB-2102	Punjab
23	PJB-2110	Punjab
24	GB-195	Gilgit Baltistan
25	GB-201	Gilgit Baltistan
26	GB-207	Gilgit Baltistan
27	GB-210	Gilgit Baltistan
28	GB-215	Gilgit Baltistan
29	GB-220	Gilgit Baltistan
30	GB-221	Gilgit Baltistan
31	GB-227	Gilgit Baltistan
32	NARC-2	Islamabad

3.2 Agro-morphological Characterization & Crop Management

A field experiment was designed by using augmented design. The row to row distance of 50cm was maintained while the path of 25 cm gap was kept between rows. For every genotype, there were two rows that were four meters long. Before and after the first week of seeding, the field was irrigated to provide the maximum amount of moisture possible. Withholding water allowed the field to remain in a rain-fed state. A single-row hand drill machine was used to plant seeds deeply in the ground. Weeding and thinning were done after germination. To manage pest insects, pre- and post-germination insecticides were applied throughout the field. Three chosen plants' essential morphological characteristics were used to score the data.

A single-row hand drill machine was used to plant seeds deeply in the ground. Weeding and thinning were done after germination. To manage pest insects, pre- and post-germination insecticides were applied throughout the field. Three chosen plants' essential morphological characteristics were used to score the data.

The mean morphological data for nine quantitative traits, such as days to flower initiation (DFI), days to 50% flower completeness (DF50%), number of pods/-plant, number of branches/plants, plant height, 100s seed weight, days to flower completion (DFC), days to maturity (DM), and plant yield/plant was recorded.

3.3 Data Analysis

For data analysis, the average data from three plants were used. Statistica, version 7.0, was used to examine the data from the descriptive statistics [43].

Principal component analysis (PCA) was used as a multivariate technique to study the taxonomic association between genotypes for all agro-morphologically significant variables using Statistica version 7.0 software and to identify the diverse genotypes [43].

Using Statistica, version 7.0 software, scatter plots of all principal components were created to determine the graphical depiction of the pattern of genetic variation among the genotypes of Soybean germplasm [43].

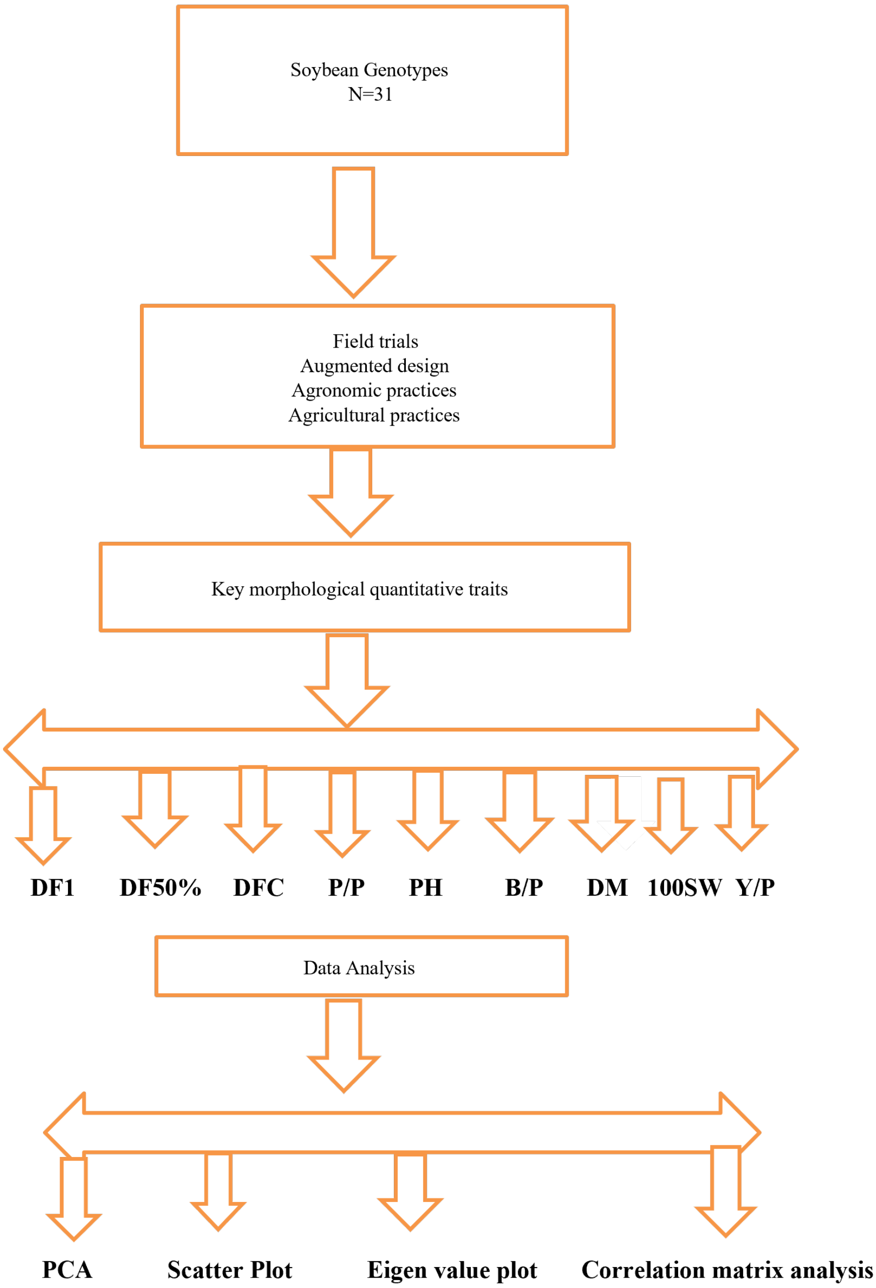


FIGURE 3.1: Graphical representation of the methodology used for exploring the agro morphological parameters of Soybean germplasm used.

Chapter 4

Results and Discussion

Agro morphological based variation is important to screen best genotypes under field conditions. The agromorphological based diverse genotypes are useful for further biochemical and molecular evaluation. The genetic diversity study is used for efficient utilization and for the development of improved cultivar/varities. Morphological based screening of different crop species/ sub-species is therefore so important for all plant breeders. Proper strategies and planning is needed to evaluate local and exotic germplasm and to select best genotypes against both qualitative and quantitative characters.

4.1 Evaluation of Morphological and Agronomic Traits in Diverse Accessions

The present data includes a comprehensive set of morphological traits measured within diverse accessions. The traits include days to flower initiation (DFI), days to 50% flowering (DF50%), days to flower completion (DFC), pods per plant (P/P), plant height (PH), branches per plant (B/P), days to maturity (DM), 100 seed weight (100 SW), and yield per plant (Y/P). Analyzing these traits provides valuable insights into the performance and potential of each accession.

According to Li et al, [45] soybean varieties that exhibit traits like longer growth periods, more nodes on the main stem, more branches, more pods, heavier grains, and greater plant height typically provide higher yields. This is consistent with research by Achina et al. which found that the number of pods per plant, the weight of the seeds, and the number of seeds per pod were the main determinants of seed yield.

In particular, the NARC-2 and PJB -2006 variety had the best yields per plant as well as the most pods per plant. That soybean cultivars with these advantageous features typically yield higher is supported by this, as reported by Li et al, [45].

According to the study, the number of branches per plant, the number of pods per plant, the weight of the pods prior to threshing, the number of seeds per plant, the yield per plant, the protein content, and the oleic acid content were the most important factors influencing yield variance

4.1.1 Flowering Time

Days to Flower Initiation (DFI): The accessions exhibit variability in DFI, ranging from 40 days (PJB-2006, PJB-2030, PJB-2110, GB-210, NARC-2) to 51 days (GB-221). Early flowering accessions can be advantageous in regions with shorter growing seasons.

Days to 50% Flowering (DF50%) and Days to Flower Completion (DFC): These traits also show variation among accessions. For example, PJB-2030 has the shortest DF50% at 45 days, while GB-201 has the longest at 57 days. Similarly, DFC ranges from 52 days (PJB-2006) to 67 days (GB-201), indicating different rates of flowering progression.

4.1.2 Yield Components

Pods per Plant (P/P): There is significant variation in P/P, with PJB-2006 and PJB-2110 having the highest number of pods per plant (98), and PJB-2094 the lowest (16). Accessions with higher P/P are likely to produce greater yields.

4.1.3 Plant Height and Yield

While some accessions with taller plants (PH) like GB-220 and PJB-2005 show good yields, there are exceptions like PJB-2034 with moderate height but low yield, indicating that plant height alone does not determine yield.

4.1.4 Branching

Accessions with more branches per plant (B/P), such as PJB-2006 and GB-220, tend to show higher yields, suggesting that more branching might contribute to higher pod production and yield.

4.1.5 Maturity Time

Accessions with earlier maturity (lower DM) do not necessarily have lower yields, as seen with NARC-2 which has a relatively low DM (100 days) but the highest yield.

4.2 Principal Components and Their Constituent Traits Contribution To Overall Variability Among Studied Germplasm

According to Shinwari et al. [44], Agro morphological trait variation is essential for determining optimal genotypes in field settings. These different genotypes that were found by agromorphological screening are useful for further analyses at the molecular and biochemical levels. Research on genetic diversity is crucial for developing better cultivars or variations and for their effective application. For plant breeders, morphological screening of various crop species and sub-species is therefore essential. To assess both domestic and foreign germplasm and choose the best genotypes based on qualitative and quantitative features, appropriate planning and techniques are required.

In the present study, morphological data of 9 quantitative traits recorded for 32 Soybean accessions were subjected to multivariate procedures of Principal Component Analysis (PCA). The first five principal components (PCs) with eigen value unity or more (>1) were considered important which accounted for 89.16% of overall phenotypic variation in population. The first principal component (PC1) had 4.42 Eigen values and depicted 49.12 % of variability among accessions. PC2 has Eigen value 1.37 representing 15.33 % of total variability. PC3 has Eigen value 0.9 representing 10.56% of total variability. PC4 has Eigen value 0.7 representing 8.32 % of total variability. PC5 has Eigen value 0.5 representing 5.83 % of total variability (Table 4.1).

Table 4.1: Principal components and their constituent traits contribution to overall variability among studied germplasm.

Value	**Eigen Value**	**% Total Varience**	**Cummulative %**
1	4.42	49.12	49.12
2	1.37	15.33	64.45
3	0.9	10.56	75.01
4	0.7	8.32	83.33
5	0.5	5.83	89.16

s

Table 4.2: Various quantitative traits their mean values and standard deviation

Var	**Mean**	**Min**	**Max**	**Std. Dev.**
DFI	44.3438	40.00000	51.0000	3.23897
DF50%	52.4688	45.00000	57.0000	3.67190
DFC	59.1875	52.00000	67.0000	4.03563
P/P	59.6563	16.00000	116.0000	23.18176
PH	37.6031	19.40000	54.0000	8.94061
B/P	3.2584	2.00000	4.0000	0.66797

Table 4.2 continued from previous page

Var	Mean	Min	Max	Std. Dev.
DM	105.1563	95.00000	110.0000	2.94147
100 SW	10.1550	6.90000	12.5000	1.41186
Y/P	104.2416	20.50000	238.5000	45.23701

According to Li et al. [45], soybean varieties that exhibit traits like longer growth periods, more nodes on the main stem, more branches, more pods, heavier grains, and greater plant height typically provide higher yields. This is consistent with research by Achina et al. [45], which found that the number of pods per plant, the weight of the seeds, and the number of seeds per pod were the main determinants of seed yield.

The average value of DFI is 44.3438, with observed values ranging from 40 to 51. The standard deviation of 3.23897 indicates that most values are within approximately 3.24 units of the mean, suggesting moderate variability. The average value of DF50% is 52.4688, with a range of 45 to 57. The standard deviation of 3.67190 shows moderate variability around the mean. The average value of DFC is 59.1875, with values ranging from 52 to 67. The standard deviation of 4.03563 indicates a moderate spread of values around the mean. The average value of P/P is 59.6563, with a wide range from 16 to 116. The large standard deviation of 23.18176 indicates high variability, suggesting significant differences in the data. The average value of PH is 37.6031, with values ranging from 19.4 to 54. The standard deviation of 8.94061 indicates substantial variability around the mean. The average value of B/P is 3.2584, with values ranging from 2 to 4. The standard deviation of 0.66797 suggests relatively low variability around the mean. The average value of DM is 105.1563, with values ranging from 95 to 110. The standard deviation of 2.94147 indicates moderate variability around the mean. The average value of 100 SW is 10.1550, with values ranging from 6.9 to 12.5. The standard deviation of 1.41186 indicates relatively low variability around the mean.

The average value of Y/P is 104.2416, with a wide range from 20.5 to 238.5. The large standard deviation of 45.23701 indicates high variability, suggesting significant differences in the data. Low Variability: Variables like B/P and 100

SW have lower standard deviations, indicating more consistent values around the mean.

High Variability: Variables like P/P and Y/P have high standard deviations, suggesting a wide spread of values and less consistency.

Range of Values: The range between the minimum and maximum values provides insight into the spread and possible outliers within the dataset.

4.3 Principal Component Analysis of Agronomic and Morphological Traits in Diverse Accessions

According to Mofokeng and Mashingaidze [46], this analysis aids in assessing each variable that affects the variation among the genotypes under study. By classifying them correctly, PCA helps differentiate between significant and less significant features. Principal Component Analysis (PCA) is a statistical technique used to reduce the dimensionality of data by transforming original variables into a new set of uncorrelated variables called principal components (PCs). These PCs capture the maximum variance in the data set. The table provided shows the loadings of various agronomic and morphological traits on the first six principal components. This analysis helps to identify which traits contribute most to the variability observed among the different accessions.

TABLE 4.3: Principal components and their constituent traits contribution to overall variability among studied germplasm

Variables	PC1	PC2	PC3	PC4	PC5
DFI	0.701587	0.565664	-0.171917	0.128335	-0.057669
DF50%	0.847685	0.389236	-0.076581	0.053653	-0.018681
DFC	0.887074	0.239452	-0.148879	-0.074573	-0.093992
P/P	-0.803658	0.199690	-0.305373	0.014502	-0.352393
PH	-0.507596	0.602625	-0.112106	0.409899	0.299109
B/P	-0.362501	0.343804	0.794346	0.203171	-0.250150

Table 4.3 continued from previous page

Variables	PC1	PC2	PC3	PC4	PC5
DM	0.765923	0.014238	0.101847	-0.129747	-0.334120
100 SW	-0.392059	0.544676	0.147473	-0.702765	0.180248
Y/P	-0.804706	0.180913	-0.351731	-0.059463	-0.303458

1 **DFI (Days to Flower Initiation)**

- **PC1**: 0.701587
- **PC2**: 0.565664
- **PC3**: -0.171917
- **PC4**: 0.128335
- **PC5**: -0.057669

DFI has high positive loadings on PC1 and PC2, indicating it significantly contributes to the variability captured by these components. Lower negative contributions to PC3 and PC5 suggest minor influences on those axes.

2 **DF50% (Days to 50% Flowering)**

- **PC1**: 0.847685
- **PC2**: 0.389236
- **PC3**: -0.076581
- **PC4**: 0.053653
- **PC5**: -0.018681

DF50% shows strong positive loading on PC1, emphasizing its importance in explaining the primary variation. It also contributes positively to PC2, while its contributions to other PCs are relatively minor.

3 **DFC (Days to Flower Completion)**

- **PC1**: 0.887074

- **PC2**: 0.239452
- **PC3**: -0.148879
- **PC4**: -0.074573
- **PC5**: -0.093992

DFC has the highest loading on PC1 among all traits, indicating it is a major driver of variation in the first principal component. It also has a moderate positive loading on PC2.

4 **P/P (Pods per Plant)**

- **PC1**: -0.803658
- **PC2**: 0.199690
- **PC3**: -0.305373
- **PC4**: 0.014502
- **PC5**: -0.352393

P/P shows a strong negative loading on PC1, suggesting it varies inversely with traits like DFI, DF50%, and DFC. It also has moderate negative and positive loadings on PC3 and PC5, respectively.

5 **PH (Plant Height)**

- **PC1**: -0.507596
- **PC2**: 0.602625
- **PC3**: -0.112106
- **PC4**: 0.409899
- **PC5**: 0.299109

PH has a strong positive loading on PC2 and PC4, indicating it is a significant factor in explaining the variability in these components. Negative loading on PC1 suggests an inverse relationship with the traits dominating that component.

6 B/P (Branches per Plant)

- **PC1**: -0.362501
- **PC2**: 0.343804
- **PC3**: 0.794346
- **PC4**: 0.203171
- **PC5**: -0.250150

B/P has the highest positive loading on PC3, indicating it is a key contributor to this component. It also positively contributes to PC2 and PC4.

7 DM (Days to Maturity)

- **PC1**: 0.765923
- **PC2**: 0.014238
- **PC3**: 0.101847
- **PC4**: -0.129747
- **PC5**: -0.334120

DM is an important trait for PC1 with a strong positive loading. Its contributions to other PCs are relatively minor.

8 100 SW (100-Seed Weight)

- **PC1**: -0.392059
- **PC2**: 0.544676
- **PC3**: 0.147473
- **PC4**: -0.702765
- **PC5**: 0.180248

100 SW shows a strong positive loading on PC2 and a strong negative loading on PC4, indicating it significantly influences these components.

9 Y/P (Yield per Plant)

- **PC1**: -0.804706
- **PC2**: 0.180913
- **PC3**: -0.351731
- **PC4**: -0.059463
- **PC5**: -0.303458

Y/P has strong negative loadings on PC1 and PC3, suggesting it inversely varies with traits dominant in these components. Its positive loading on PC2 is moderate.

The PCA reveals that traits such as DFI, DF50%, and DFC heavily influence the first principal component, which accounts for the majority of the variance in the data. Traits like PH and 100 SW contribute significantly to the second and fourth components. B/P is the key contributor to the third component. These insights can be utilized in breeding programs to select accessions with desired trait combinations for improved agronomic performance.

According to Li et al. [45], soybean varieties that exhibit traits like longer growth periods, more nodes on the main stem, more branches, more pods, heavier grains, and greater plant height typically provide higher yields. This is consistent with research by Achina et al. [46], which found that the number of pods per plant, the weight of the seeds, and the number of seeds per pod were the main determinants of seed yield.

In particular, the PJB-2006 variety had the best yields per plant as well as the most pods per plant and some other varieties e.g. GB-220 and PJB-2030 also considered as best genotypes. That soybean cultivars with these advantageous features typically yield higher is supported by this, as reported by Li et al., [45].

According to the study, the number of branches per plant, the number of pods per plant, the weight of the pods prior to threshing, the number of seeds per plant, the yield per plant, the protein content, and the oleic acid content were the most important factors influencing yield variance. Additionally, seed yield has a tendency to positively correlate with other desirable yield features, according to Anand and Torrie [46], indicating that an increase in one trait would probably result in an increase in another. As a result, high-yielding soybean genotypes are usually easily recognized because of the significant positive connection that exists between seed yield and other features.

4.4 Eigenvalues and Variance Explained by Principal Components

Principal Component Analysis (PCA) is a powerful technique used to reduce the dimensionality of a dataset while preserving as much variance as possible. Each principal component (PC) is associated with an eigenvalue that indicates the amount of variance it explains. The proportion of total variance explained by each PC helps in understanding the significance of each component in capturing the underlying data structure. The following table provides the eigenvalues and the percentage of total variance explained by the first two principal components, as well as their cumulative contributions.

PC1 has a high eigenvalue of 4.420521, indicating that it accounts for a significant portion of the total variance in the dataset. Specifically, it explains 49.11690% of the total variance. This means that nearly half of the variability in the data can be attributed to the first principal component alone. The high percentage of variance explained by PC1 suggests that this component captures the most critical patterns in the data, making it a crucial factor in understanding the overall structure of the dataset.

PC2 has an eigenvalue of 1.379656, explaining 15.32951% of the total variance. When combined with PC1, the cumulative variance explained increases to 64.441%. Although PC2 contributes less to the total variance compared to PC1, it still plays a significant role in capturing additional patterns that are not accounted for by

PC1. The inclusion of PC2 increases the explained variance to over 64%, indicating that these two components together provide a comprehensive understanding of the data's structure.

The eigenvalues and the percentage of variance explained by the first two principal components highlight the importance of these components in capturing the variability within the dataset. PC1, with its high eigenvalue and substantial variance explanation, is the primary component driving the data's structure. PC2, while contributing less variance individually, adds significant value by accounting for additional patterns. Together, these components explain 64.44641% of the total variance, indicating that they effectively summarize the key features of the dataset.

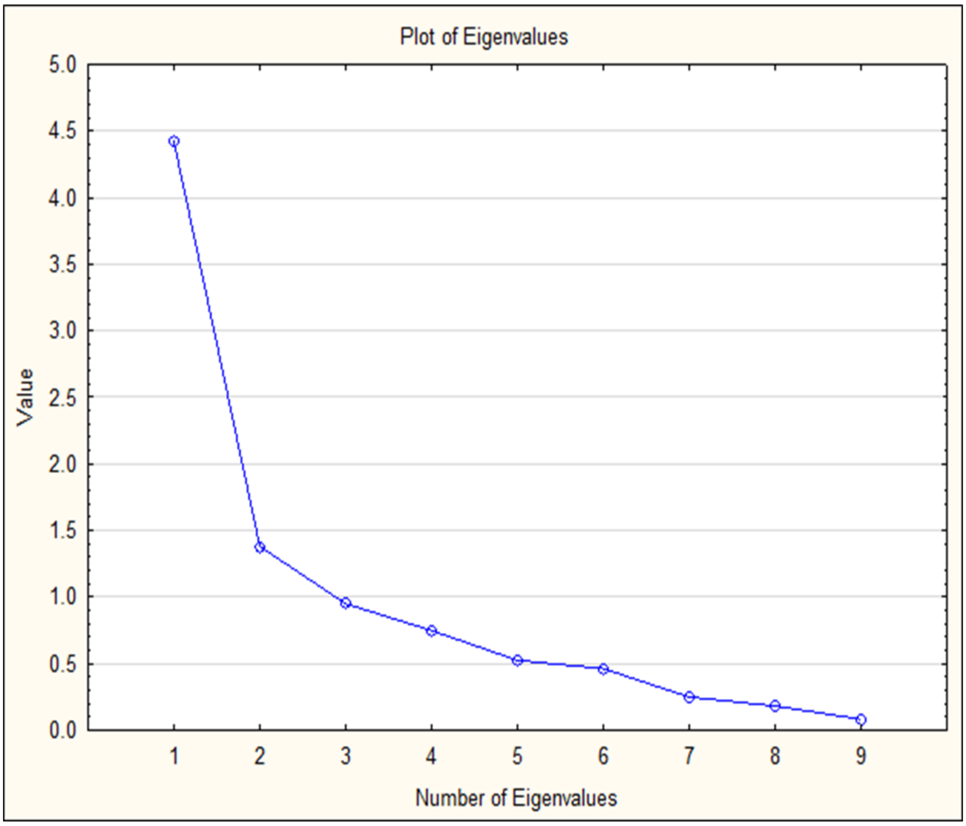


FIGURE 4.1: Eigen value plot

PC1 captures the most variance in the data, indicating it is the most significant principal component. PC2 captures a substantial amount of variance but significantly less than PC1.

PC3 captures a moderate amount of variance. From PC4 onwards, each principal component captures progressively less variance, and the eigenvalues become quite small.

The plot shows a noticeable ”elbow” after the second principal component. This point suggests that the first two or three principal components capture most of the variance in the dataset. PC1 and PC2 together explain a large portion of the variance. Subsequent components (PC4 and beyond) contribute less significantly to the total variance, indicating that they add relatively little new information to the data structure. Based on the scree plot, it would be reasonable to retain the first three principal components (PC1, PC2, and PC3) for further analysis, as they capture the most significant amount of variance.

Retaining more components beyond this point would add less explanatory power and could introduce noise rather than meaningful variation

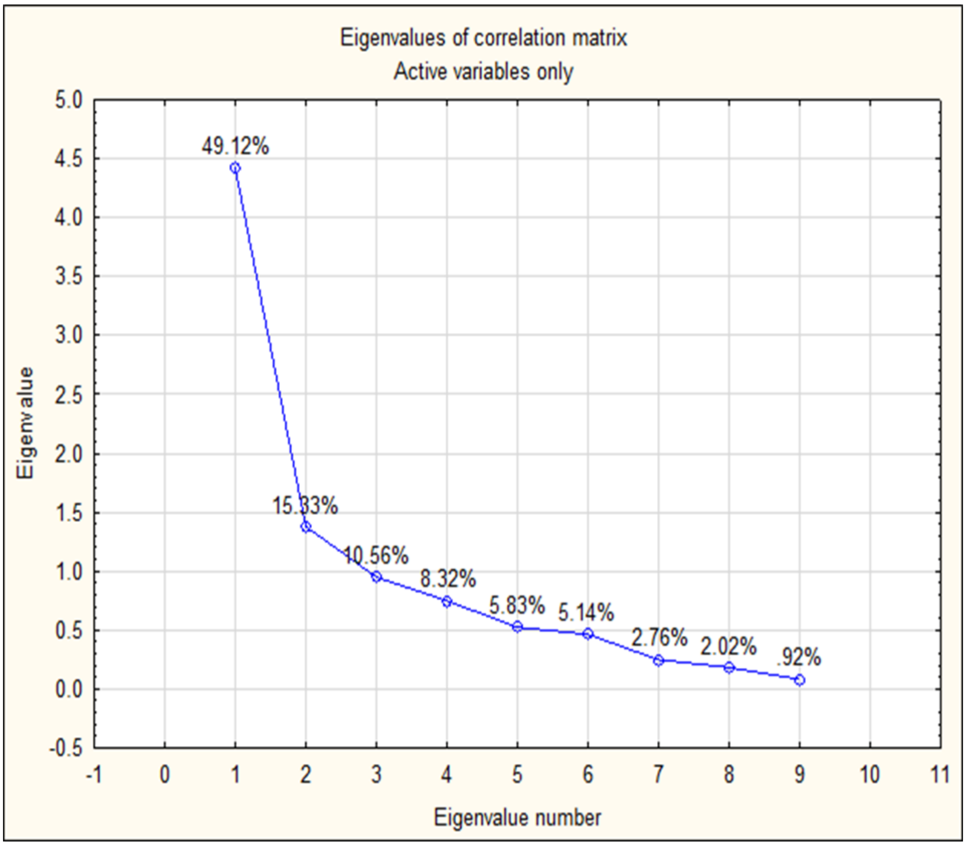


FIGURE 4.2: Eigen values of correlation matrix

This plot is a scree plot, which is often used in Principal Component Analysis (PCA) to show the eigenvalues of the correlation matrix. Eigenvalues indicate the amount of variance explained by each principal component. The first principal component (PC1) has the highest eigenvalue, meaning it explains the most variance in the data. As the component number increases, the eigenvalues decrease, showing that each successive principal component explains less variance.

Eigenvalues indicate the amount of variance explained by each principal component. The first principal component (PC1) has the highest eigenvalue, meaning it explains the most variance in the data.

As the component number increases, the eigenvalues decrease, showing that each successive principal component explains less variance.

4.5 Scree Plot Analysis

Elbow Point: In a scree plot, the 'elbow' point is a key feature. It is the point where the plot bends, indicating a natural cut-off for the number of principal components to retain.

In this plot, the elbow appears to be around PC3 or PC4, suggesting that the first 3 or 4 components capture most of the variance in the data.

Significant Components: Components before the elbow point are typically considered significant because they explain a substantial amount of the variance. In this plot, PC1, PC2, and possibly PC3 are significant.

Noise Components: Components after the elbow point usually contribute less to the variance and may be considered as noise.

From PC4 onwards, the eigenvalues and their corresponding variance proportions are much lower, indicating these components might not add much explanatory power to the model.

4.6 Principal Component Analysis (PCA) Biplot of Various Variables

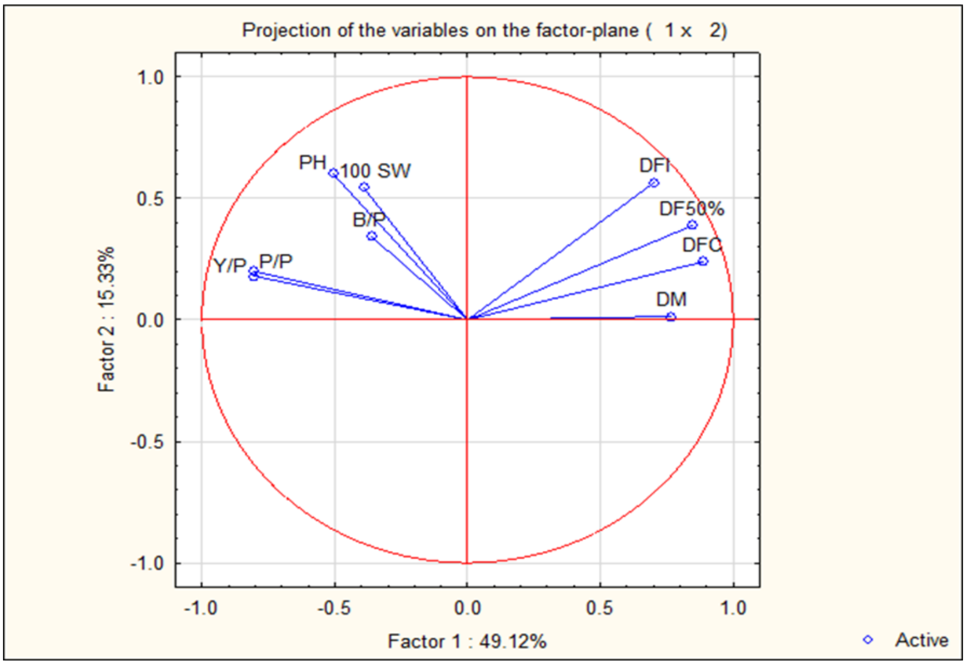


FIGURE 4.3: Principal Component Analysis Biplot. Factor 1 and Factor 2

This PCA biplot provides a clear visualization of how each variable contributes to the underlying factors in the dataset, helping to understand the structure and relationships within the data.

1. **Factor 1 (49.12%)**: This axis explains 49.12% of the variance in the data. It is the principal component that captures the maximum variability in the dataset.

2. **Factor 2 (15.33%)**: This axis explains an additional 15.33% of the variance. Together with Factor 1, these two factors explain a significant portion of the total variance in the data (64.45%).

The variables are represented as vectors (arrows) pointing from the origin (0,0) in the biplot. The length and direction of these vectors provide insights into the importance and relationship of the variables with the factors.

Positive correlation: Variables with vectors that are close together are positively correlated. For example, DF1 and DF50 % are closely aligned, suggesting a strong positive correlation.

Negative Correlation: Variables with vectors pointing in opposite directions are negatively correlated. For instance if there were vectors directly opposite to DF1 and DF50, they would indicate negative correlation.

No Correlation: Vectors that are orthogonal indicate no correlation between those variables.

1. **PH:** This variable has a moderate positive loading on Factor 1 and Factor 2, indicating that it contributes to the variance explained by both factors.

2. **100 SW:** Similar to PH, it has a moderate positive loading on both Factor 1 and Factor 2.

3. **BR:** This variable shows a small positive loading on Factor 1 and Factor 2, indicating a weak but positive contribution to the variance.

4. **DF1, DF50%, DFC:** These variables have a strong positive loading on Factor 1 and moderate positive loading on Factor 2, suggesting they are significant contributors to the variance explained by Factor 1.

5. **DM:** This variable has a positive loading on Factor 1 but a near-zero loading on Factor 2, indicating it contributes mainly to the variance explained by Factor 1.

6. **P/P, Y/P:** These variables have moderate positive loadings on Factor 1 but negative loadings on Factor 2, suggesting they contribute to the variance explained by Factor 1 and negatively to Factor 2.

    - **Factor 1** is primarily influenced by variables like DF1, DF50%, and DFC.

    - **Factor 2** is influenced to a lesser extent by variables like PH and 100 SW.

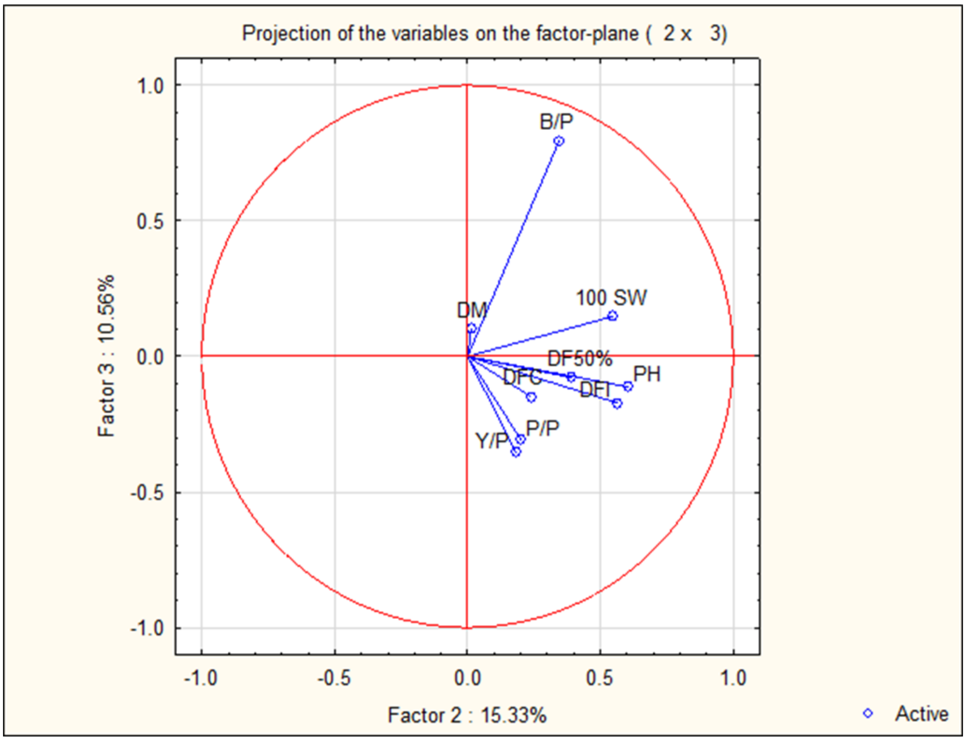


FIGURE 4.4: Principal Component Analysis Biplot. Factor 2 and Factor 3

- Variables like BR and DM have a relatively minor influence on both factors.
- Factor 2 is primarily influenced by variables like DF1, DF50%, and DFC.
- Factor 3 is influenced by variables like DM and BR.
- Variables like BR and PH have a relatively minor influence on both factors.

4.6.1 For instance

DF1, DF50%, and DFC are close to the edge and thus are significant contributors. BR and P/P are closer to the origin, indicating a lower contribution

This PCA biplot provides a clear visualization of how each variable contributes to the underlying factors in the dataset, helping to understand the structure and relationships within the data.

4.7 Factor Analysis Projection of Cases on Factor-Plane (1x2)

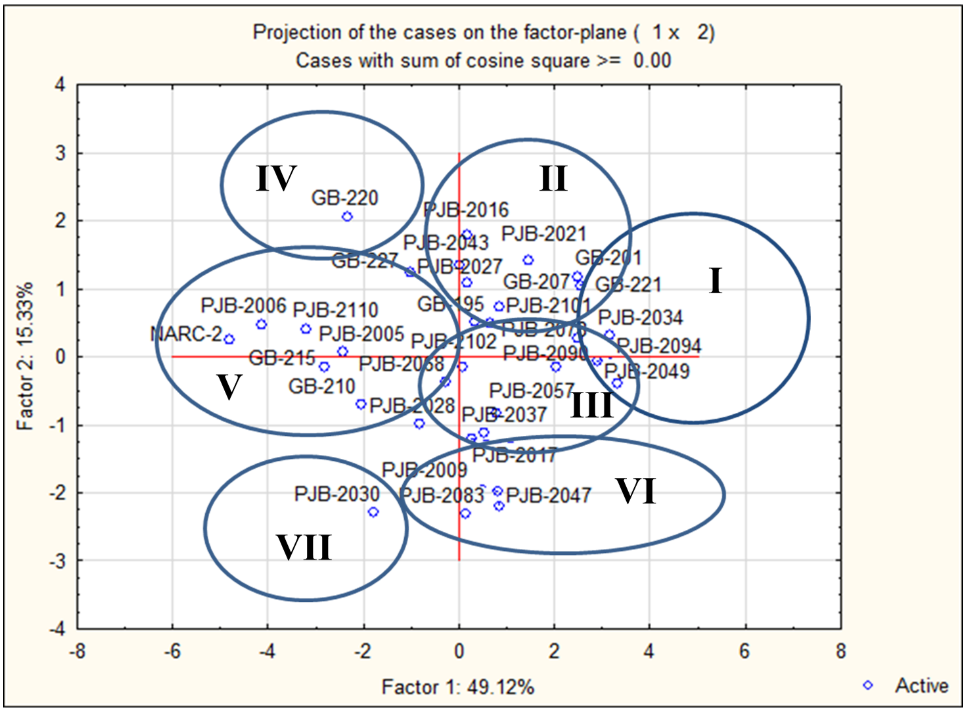


FIGURE 4.5: PCA scatter plot

The provided graph is a scatter plot that represents the projection of various cases on the factor-plane defined by Factor 1 and Factor 2. This type of plot is commonly used in factor analysis, principal component analysis (PCA), or correspondence analysis to visualize the relationships between cases (or variables) in a reduced dimensional space.

1. Axes and Factors:
    - Factor 1 (X-axis): This axis explains 49.12% of the total variance in the data. It represents the most significant dimension in which the data can be projected.
    - Factor 2 (Y-axis): This axis explains 15.33% of the total variance in the data. It represents the second most significant dimension in which the data can be projected.

2. Projection of Cases:
   - Each point on the plot represents a case (denoted by labels such as PJB-2016, GB-220, NARC-2, etc.).
   - The position of each case on the plot is determined by its score on Factor 1 and Factor 2.
   - The plot helps in visualizing how cases are related to each other in the context of these two factors.
3. Interpretation of Clusters and Outliers:
   - Clusters: Cases that are close to each other on the plot are similar in terms of the characteristics that Factor 1 and Factor 2 represent

The first two principal components were depicted on a scatter plot to observe the relationship among all 32 accessions. The pattern was constructed by distributing genotypes in 7 discrete groups. Group 1 had 5 accessions of which 70% were from Punjab Pakistan. The second group contain 5 accessions including 4 were from Punjab and 1 from Gilgit Baltistan. Group three had 6 accessions from Punjab. Group four had only one accession from Gilgit Baltistan. Group five had 9 accessions from which 5 were from Punjab, 3 were from Gilgit Baltistan and one from NARC Islamabad. Group 6 had 5 accessions from which all were from Punjab and Group 7 had only 1 accession that were from Punjab. Out of these 3 genotypes showed some unique performances (Table 4.4, Fig 4.5).

Our findings diverge from those of Worekeneh et al. [47], who found that the first five groups accounted for about 90% of the variability in Soybean genotypes. However, their additional results align closely with ours, indicating that the PC1 group contributed the most (49.12%) to the overall PC groups, followed by PC2 (15.33%). They identified plant height, the number of primary branches per plant, and 1000 seed weight as having the highest variability among quantitative traits. Similarly, Ilyas et al. [48] observed the highest eigenvalues in the first four PC groups in a study on 32 soybean germplasm accessions, with the first group contributing the most, followed by the second. Yu et al. [49] corroborated these findings in their study on European and Chinese soybean genotypes, noting that plant height and the number of siliquae per terminal raceme had the highest contribution in PC1. Ali et al. [50] also found that PC1 contributed 22.21% of the variability, followed by PC2 (14.65%) and PC3 (11.41%), with the number of leaves per plant and siliquae per main raceme showing the highest contribution in the first two groups. Their study further confirmed maximum variability in all PC groups in Pakistani soybean genotypes.

Ali et al. [50] also found that PC1 contributed 22.21% of the variability, followed by PC2 (14.65%) and PC3 (11.41%), with the number of leaves per plant and siliquae per main raceme showing the highest contribution in the first two groups. Their study further confirmed maximum variability in all PC groups in Pakistani soybean genotypes.

TABLE 4.4: Description of 3 agromorphologically diverse Soybean germplasm polulations revealed by PCA scatter plot.

Population	**Characteristics**	**No. of Accessions**
Population 1	Early maturing, high pod shattering, minimum leaf length and seed yield per plant.	5
Population 2	Average performing genotypes for traits like plant height, pri	5
Population 3	Earlier Flowering initiation, days to 50% flowering, days to flowering completion and maximum 100-seed weight	6
Population 4	No pod shattering at later stages, shortest main raceme and minimum 100 seed weight but highest seed yield/plant.	1
Population 5	Late flowering and maturity, largest leaf length, minimum pod length, minimum number of branches/plant and seed/pod.	9
Population 6	Maximum stem thickness and performed above average for most of the agronomic traits.	5
Population 7	Minimum days to flowering initiation, days to 50% flowering and days to flowering completion, Tallest plants with highest number of branches per plants, maximum leaf and pod width.	1

4.8 Factor Analysis Projection of Cases on Factor-Plane (2x3)

The factor analysis projection plot is a powerful tool for visualizing high-dimensional data in a lower-dimensional space. By analyzing the positions and clusters of cases on the factor-plane, we can gain valuable insights into the relationships and similarities between them. This information is crucial for exploratory data analysis, data reduction, and identifying key patterns in the dataset.

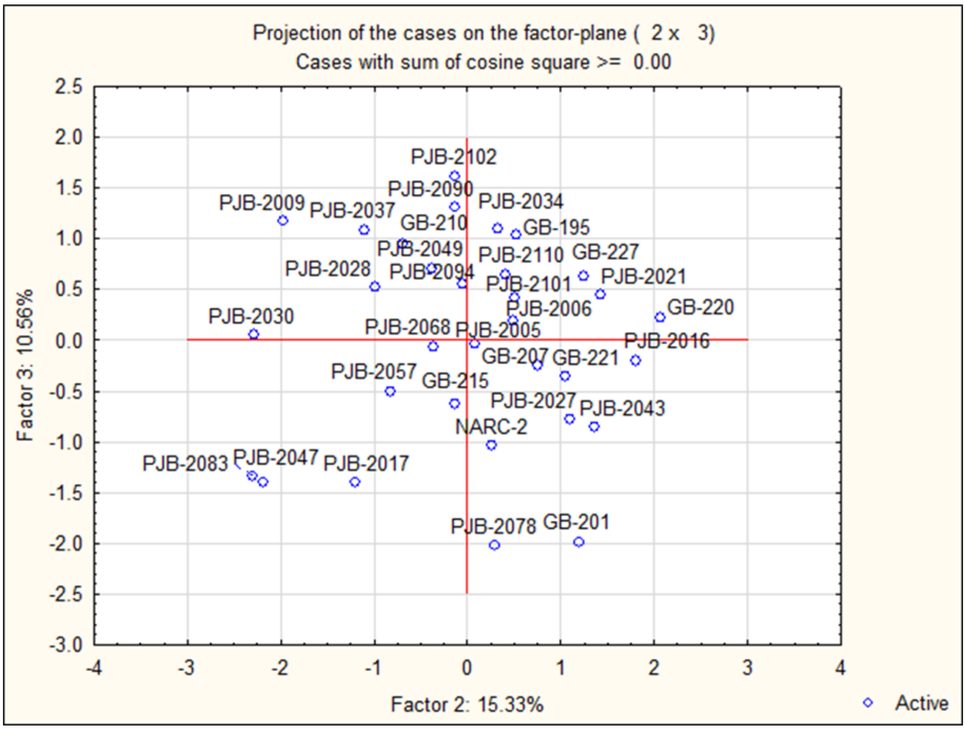


FIGURE 4.6: PCA scatter plot

The provided graph is a scatter plot that represents the projection of various cases on the factor-plane defined by Factor 2 and Factor 3. This type of plot is commonly used in factor analysis, principal component analysis (PCA), or correspondence analysis to visualize the relationships between cases (or variables) in a reduced dimensional space.

1. **Axes and Factors**:
   - **Factor 2 (X-axis):** This axis explains 15.33% of the total variance in the data. It represents the second most significant dimension in which

the data can be projected. **Factor 3 (Y-axis):** This axis explains 10.56% of the total variance in the data. It represents the third most significant dimension in which the data can be projected.

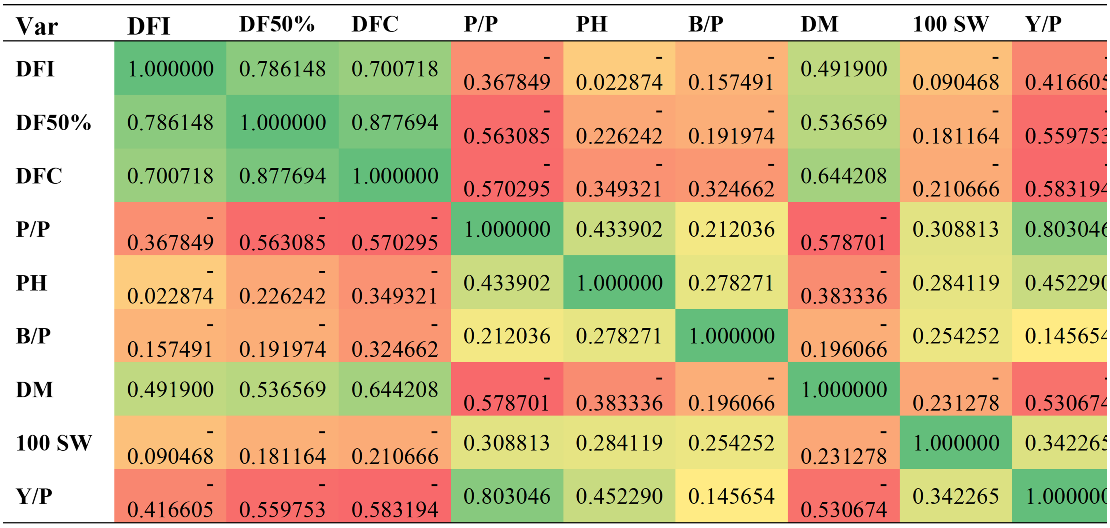
2. **Projection of Cases:**
   - Each point on the plot represents a case (denoted by labels such as PJB-2012, GB-195, NARC-2, etc.).
   - The position of each case on the plot is determined by its score on Factor 2 and Factor 3.
   - The plot helps in visualizing how cases are related to each other in the context of these two factors.
3. **Interpretation of Clusters and Outliers:**
   - **Clusters:** Cases that are close to each other on the plot are similar in terms of the characteristics that Factor 2 and Factor 3 represent. For example, cases like PJB-2102, PJB-2090, and GB-210 are clustered together, indicating similarity. **Outliers:** Cases that are far from the main cluster (e.g., PJB-2083, PJB-2078) are different from most other cases based on the factors considered.
4. **Understanding Factor Loadings:**
   - **Cosine Square:** The plot mentions ”Cases with sum of cosine square >= 0.00”. This suggests that all cases shown have a non-negative sum of squared cosines, indicating their contributions to the factors.
   - The cosine square of an angle between a case’s vector and a factor’s axis represents the quality of representation of that case by the factor. Higher values indicate better representation.
5. **Applications and Implications:**
   - **Dimensionality Reduction:** This plot is useful for reducing the complexity of the data by representing it in fewer dimensions while retaining most of the variability.

- **Exploratory Data Analysis:** Helps in identifying patterns, clusters, and outliers, providing insights into the structure of the data.
- **Hypothesis Generation:** Can be used to generate hypotheses about the relationships between cases based on their positions in the reduced-dimensional space.

4.9 Correlations Before Key Morphological Traits of Soil

The correlation matrix is a crucial tool in data analysis that helps identify the strength and direction of relationships between variables. This analysis provides valuable insights for further statistical modeling and decision-making processes.

Table 4.5: Correlation Matrix Analysis of Dataset Variables



Var	DFI	DF50%	DFC	P/P	PH	B/P	DM	100 SW	Y/P
DFI	1.000000	0.786148	0.700718	-0.367849	-0.022874	-0.157491	0.491900	-0.090468	-0.416605
DF50%	0.786148	1.000000	0.877694	-0.563085	-0.226242	-0.191974	0.536569	-0.181164	-0.559753
DFC	0.700718	0.877694	1.000000	-0.570295	-0.349321	-0.324662	0.644208	-0.210666	-0.583194
P/P	-0.367849	-0.563085	-0.570295	1.000000	0.433902	0.212036	-0.578701	0.308813	0.803046
PH	-0.022874	-0.226242	-0.349321	0.433902	1.000000	0.278271	-0.383336	0.284119	0.452290
B/P	-0.157491	-0.191974	-0.324662	0.212036	0.278271	1.000000	-0.196066	0.254252	0.145654
DM	0.491900	0.536569	0.644208	-0.578701	-0.383336	-0.196066	1.000000	-0.231278	-0.530674
100 SW	-0.090468	-0.181164	-0.210666	0.308813	0.284119	0.254252	-0.231278	1.000000	0.342265
Y/P	-0.416605	-0.559753	-0.583194	0.803046	0.452290	0.145654	-0.530674	0.342265	1.000000

1. **Significant Correlations:**
   - **Positive Correlations:**
     - Variables 1 and 2: =0.786148 = 0.786148=0.786148
     - Variables 2 and 3: =0.877694 = 0.877694=0.877694
     - Variables 3 and 4: =0.644208 = 0.644208=0.644208

    - Variables 4 and 5: =0.433902 = 0.433902=0.433902

  - **Negative Correlations:**
    - Variables 1 and 4: =-0.367849 = -0.367849=-0.367849
    - Variables 2 and 4: =-0.563085 = -0.563085=-0.563085
    - Variables 3 and 4: =-0.570295 = -0.570295=-0.570295
    - Variables 4 and 6: =-0.578701 = -0.578701=-0.578701

2. **Key Observations:**
   - **High Positive Correlations:**
     - There is a very high positive correlation between variables 2 and 3 (=0.877694r = 0.877694=0.877694).
     - Variables 4 and 5 have a moderately high positive correlation (=0.433902 = 0.433902=0.433902).
   - **High Negative Correlations:**
     - Variable 4 shows a significant negative correlation with multiple variables, especially with variable 6 (=-0.578701 = -0.578701=-0.578701).
     - Variables 1 and 6 also exhibit a moderately negative correlation (=-0.416605 = -0.416605=-0.416605).

3. **Implications:**
   - **Multicollinearity:**
     - High correlation coefficients (either positive or negative) between independent variables may indicate multicollinearity, which can affect regression analysis and the interpretation of model coefficients.
   - **Variable Relationships:**
     - Understanding these correlations helps in feature selection for predictive modeling, where highly correlated variables may be redundant and can be dropped to simplify models.
   - **Data Interpretation:**

(a) The correlations provide insights into the relationships between variables, which can be useful in hypothesis testing, exploratory data analysis, and identifying potential causal relationships.

Agromorphological trait variation is essential for determining optimal genotypes in field settings. These different genotypes that were found by agromorphological screening are useful for further analyses at the molecular and biochemical levels. Research on genetic diversity is crucial for developing better cultivars or variations and for their effective application. For plant breeders, morphological screening of various crop species and sub-species is therefore essential. To assess both domestic and foreign germplasm and choose the best genotypes based on qualitative and quantitative features, appropriate planning and techniques are required.

Traditional cluster analysis is a simple and effective method for assessing genetic variation within germplasm collections. By assembling core subsets and homogenizing accessions with comparable traits, this method analyzes genetic diversity. By grouping accessions according to commonalities, their relationships are made simpler for easier comprehension and interpretation. For hybridization initiatives, assessing the genetic variety of germplasm is essential because it enhances the genes of the most varied lines and encourages the use of genetic variations. In order to efficiently manage and maintain germplasm supplies and improve crop efficiency, breeders require knowledge on genetic diversity and the linkages among breeding materials.

In the present study, morphological data of 9 quantitative traits recorded for 32 Soybean accessions were subjected to multivariate procedures of Principal Component Analysis (PCA). The first five principal components (PCs) with eigen value unity or more (> 1) were considered important which accounted for 89.16% of overall phenotypic variation in population. The first principal component (PC1) had 4.42 Eigen values and depicted 49.12% of variability among accessions. PC2 has Eigen value 1.37 representing 15.33% of total variability. PC3 has Eigen value 0.9 representing 10.56% of total variability. PC4 has Eigen value 0.7 representing 8.32% of total variability. PC5 has Eigen value 0.5 representing 5.83% of total variability (Table 4.1)

The first two principal components were depicted on a scatter plot to observe the relationship among all 32 accessions. The patteren was constructed by distributing genotypes in 7 discrete groups. Group 1 had 5 accessions of which 70% were from Punjab Pakistan. The second group contain 5 accessions including 4 were from Punjab and 1 from Gilgit Baltistan. Group three had 6 accessions from Punjab. Group four had only one accession from Gilgit Baltistan. Group five had 9 accessions from which 5 were from Punjab, 3 were from Gilgit Baltistan and one from NARC Islamabad. Group 6 had 5 accessions from which all were from Punjab and Group 7 had only 1 accession that were from Punjab. Out of these 3 genotypes showed some unique performances.

According to Li et al, soybean varieties that exhibit traits like longer growth periods, more nodes on the main stem, more branches, more pods, heavier grains, and greater plant height typically provide higher yields. This is consistent with research by Achina et al. which found that the number of pods per plant, the weight of the seeds, and the number of seeds per pod were the main determinants of seed yield.

In particular, the NARC-2 variety had the best yields per plant as well as the most pods per plant. That soybean cultivars with these advantageous features typically yield higher is supported by this, as reported by Li et al.

According to the study, the number of branches per plant, the number of pods per plant, the weight of the pods prior to threshing, the number of seeds per plant, the yield per plant, the protein content, and the oleic acid content were the most important factors influencing yield variance. Additionally, seed yield has a tendency to positively correlate with other desirable yield features, according to Anand and Torrie [49], indicating that an increase in one trait would probably result in an increase in another. As a result, high-yielding soybean genotypes are usually easily recognized because of the significant positive connection that exists between seed yield and other features.

In population genetics, principal component analysis (PCA) is a statistical technique that is widely used to determine genetic distribution and organization among various geographic and ethnic groups [49]. According to Mofokeng and Mashingaidze [50], this analysis aids in assessing each variable that affects the variation among the genotypes under study. By classifying them correctly, PCA helps differentiate between significant and less significant features.

Chapter 5

Conclusion

Using agro-morphological and quality traits, some of the most important soybean cultivars were observed. NARC-2 had the highest number of pods per plant and NARC-2 also had the highest yields and PJB-2030, GB-220 and PJB 2006 were also considered best and can be used for further breeding programs. There were positive correlations showed between most of the assessed traits in relation to one another and this will be very helpful in assisting in the combined improvement of these traits by selecting ones that were found to have a positive and high correlation, as well as easily measurable phenotypic traits, although most were found to have highly significant and positive correlation with seed yield per plant.

Soybean is one of the most important leguminous crops grown globally. Understanding the genetic diversity and its interaction with the environment is of paramount importance in developing cultivars considering farmer's preferred traits. The multivariate analysis was able to reveal the relationship between the genetic diversity, agronomic and nutritional composition of selected soybean genotypes. For a successful breeding programme to function a complete understanding of the genetic diversity of the crop is required. Better knowledge of the genetic similarities and dissimilarities of breeding material could aid breeders and curators in maintaining genetic diversity, sustain long-term selection gain and conserve the germplasm. Monitoring the genetic diversity among genetic resources of elite breeding material could make crop improvement more efficient by the directed accumulation of favored alleles thus reducing the amount of material to be screened.

Further biochemical characterization is needed for saturated and unsaturated fatty acids. Genome wide associated studies, Single nucleotide polymorphism and Simple sequence repeats can be employed for further characterization of these genotypes.

Bibliography

[1] Asfaw, T. Abush, and A. Sentayehu, ”Soybean genetic improvement in Ethiopia,” in Food and forage legumes of Ethiopia: progress and prospects, K. Ali et al., Eds. Ghion Hotel, Ethiopia: Workshop on Food and Forage Legume, 2006, pp. 22-26.

[2] A. Bekele, A. Getint, and Z. Habtamu, ”Genetic divergence among soybean (*Glycine max* (L) Merrill) introductions in Ethiopia based on agronomic traits,” Journal of Biology, Agriculture and Healthcare, vol. 2, no. 6, pp. 1-12, 2012.

[3] H. R. Bhandari, A. N. Bhanu, K. Srivastava, and A. Hemantaranjan, ”Assessment of genetic diversity in crop plants: An overview,” Advances in Plants and Agricultural Research, vol. 7, no. 3, pp. 279-286, 2017.

[4] G. M. Bhatt, ”Multivariate analysis approach to the selection of parents for hybridization aiming at yield improvement in self-pollinated crops,” Australian Journal of Agricultural Research, vol. 20, pp. 1-7, 1970.

[5] G. S. Chahal and S. S. Gosal, ”Principles and procedures of plant breeding: Biotechnology and conventional approaches,” Narosa Publishing House, New Delhi, 2002.

[6] Central Statistical Agency (CSA), ”Agricultural sample survey 2018/2019: Report on area and production of major crops (private peasant holdings, Meher Season). Volume I, Statistical Bulletin 584,” Addis Ababa, Ethiopia, 2019. [Online]. Available: https://www.statsethiopia.gov.et/work-with-us/

[7] J. I. da Silva Rodrigues, K. M. Antunes Arruda, N. Deniz Piovesan, E. G. de Barros, and M. Moreira, ”Selection of progenitors for increase in oil content in soybean,” Revista. Ceres, vol. 63, no. 5, pp. 661-667, 2016.

[8] J. Sun, H. Mooney, W. Wu, H. Tang, Y. Tong, Z. Xu, B. Huang, Y. Cheng, X. Yang, D. Wei, et al., ”Importing food damages domestic environment: Evidence from global soybean trade,” Proc. Natl. Acad. Sci. USA, vol. 115, pp. 5415–5419, 2018.

[9] M. Staniak, E. Szpunar-Krok, and A. Kocira, ”Responses of soybean to Selected Abiotic Stresses—Photoperiod, Temperature and Water,” Agriculture, vol. 13, p. 146, 2023.

[10] A. Kotecki and S. Lewandowska, Studia na Uprawa Soi Zwyczajnej (*Glycine max* (L.) Merrill) w Poludniowo-Zachodniej Polsce, Wroclaw, Poland: Wydawnictwo Uniwersytetu Przyrodniczego we Wroclawiu, 2020, pp. 54–79.

[11] W. swiecicki, M. Chudy, and K. Zuk-Golaszewska, ”Rosliny straczkowe w projektach badawczych Unii Europejskiej (Grain legumes in the UE framework programmes),” Zesz. Probl. Postepow Nauk. Rol., vol. 522, pp. 55–65, 2007.

[12] N. Bellaloui, H.A. Bruns, H.K. Abbas, A. Mengistu, D.K. Fisher, and K.N. Reddy, ”Agricultural practices altered soybean seed protein, oil, fatty acids, sugars, and minerals in the Midsouth USA,” Front. Plant Sci., vol. 6, p. 31, 2015.

[13] S.M. Adeyemo and A.A. Onilude, ”Enzymatic Reduction of Anti-nutritional Factors in Fermenting Soybeans by Lactobacillus plantarum Isolates from Fermenting Cereals,” Niger. Food J., vol. 31, pp. 84–90, 2013.

[14] E.J. Anderson, M.L. Ali, W.D. Beavis, P. Chen, T.E. Clemente, B.W. Diers, G.L. Graef, P. Grassini, D.L. Hyten, and L.K. McHale, ”Soybean [*Glycine max* (L.) Merr.] breeding: History, improvement, production and future opportunities,” in Advances in Plant Breeding Strategies: Legumes, Cham, Switzerland: Springer International Publishing, 2019, vol. 7, pp. 431–516.

[15] S. Michałek and E. Borowski, ”Seed germination and seedling growth of the Polish soybean (*Glycine max* (L.) Merr.) cultivars in drought conditions,” Biul. IHAR, pp. 195–202, 2002.

[16] 16. G. Hou, G.R. Ablett, K.P. Pauls, and I. Rajcan, ”Environmental effects on fatty acid levels in soybean seed oil,” J. Am. Oil. Chem. Soc., vol. 83, pp. 759–763, 2006.

[17] 17. G.R. Cramer, K. Urano, S. Delrot, M. Pezzotti, and K. Shinozaki, ”Effects of abiotic stress on plants: A systems biology perspective,” BMC Plant Biol., vol. 11, p. 163, 2011.

[18] L. Wang, L. Liu, Y. Ma, S. Li, S. Dong, and W. Zu, ”Transcriptome profiling analysis characterized the gene expression patterns responded to combined drought and heat stresses in soybean,” Comput. Biol. Chem., vol. 77, pp. 413–429, 2018.

[19] K. Jumrani and V.S. Bhatia, ”Interactive effect of temperature and water stress on physiological and biochemical processes in soybean,” Physiol. Mol. Biol. Plants, vol. 25, pp. 667–681, 2019.

[20] P. Gupta, R. Sharma, M.K. Sharma, M.P. Sharma, G.K. Satpute, S. Garg, S.L. Singla-Pareek, and A. Pareek, ”Signaling cross talk between biotic and abiotic stress responses in soybean,” in Abiotic and Biotic Stresses in Soybean Production, Oxford, UK: Oxford Academic Press, 2016, vol. 1, pp. 27–52.

[21] 21. C. Wijewardana, K. Reddy, F. Alsajri, J. Irby, J. Krutz, and B. Golden, ”Quantifying soil moisture deficit effects on soybean yield and yield component distribution patterns,” Irrig. Sci., vol. 36, pp. 241–255, 2018.

[22] A.G. Bengough and C.E. Mullins, ”Mechanical impedance to root growth: A review of experimental techniques and root growth responses,” J. Soil Sci., vol. 41, pp. 341–358, 1990.

[23] J. Tu and C. Tan, ”Effect of soil compaction on growth, yield and root rots of white beans in clay loam and sandy loam soil,” Soil Biol. Biochem., vol. 23, pp. 233–238, 1991.

[24] L.X. Yu, J.D. Ray, J.C. O'Toole, and H.T. Nguyen, ”Use of wax_petrolatum layers for screening rice root penetration,” Crop Sci., vol. 35, pp. 684–687, 1995.

[25] P. Barraclough and A. Weir, ”Effects of a compacted subsoil layer on root and shoot growth, water use and nutrient uptake of winter wheat,” J. Agric. Sci., vol. 110, pp. 207–216, 1988.

[26] W. Busscher, R. Sojka, and C. Doty, ”Residual effects of tillage on coastal plain soil strength,” Soil Sci., vol. 141, pp. 144–148, 1986.

[27] S.H. Kim, P. Subramanian, B.S. Hahn, and B.K. Ha, ”High-Throughput Phenotypic Characterization and Diversity Analysis of soybean Roots (*Glycine max* L.),” Plants, vol. 11, p. 2017, 2022.

[28] H.G. Fried, S. Narayanan, and B. Fallen, ”Characterization of a soybean (*Glycine max* L. Merr.) germplasm collection for root traits,” PLoS ONE, vol. 13, p. e0200463, 2018.

[29] T. Carter, ”Breeding for drought tolerance in soybean-where do we stand,” in World Soybean Research Conference IV, Rosario, Argentina: Asociacion Agrentina de la Soja, 1989.

[30] P.M. Gaur, L. Krishnamurthy, and J. Kashiwagi, ”Improving drought-avoidance root traits in chickpea (Cicer arietinum L.)-current status of research at ICRISAT,” Plant Prod. Sci., vol. 11, pp. 3–11, 2008.

[31] J. Ao, J. Fu, J. Tian, X. Yan, and H. Liao, ”Genetic variability for root morph-architecture traits and root growth dynamics as related to phosphorus efficiency in soybean,” Funct. Plant Biol., vol. 37, pp. 304–312, 2010.

[32] Q. Liang, X. Cheng, M. Mei, X. Yan, and H. Liao, ”QTL analysis of root traits as related to phosphorus efficiency in soybean,” Ann. Bot., vol. 106, pp. 223–234, 2010.

[33] A.G. Bengough, B.M. McKenzie, P.D. Hallett, and T.A. Valentine, ”Root elongation, water stress, and mechanical impedance: A review of limiting stresses and beneficial root tip traits,” J. Exp. Bot., vol. 62, pp. 59–68, 2011.

[34] J. Lynch, ”Root Architecture and Plant Productivity,” Plant Physiol., vol. 109, pp. 7–13, 1995.

[35] C. Zhao, Y. Zhang, J. Du, X. Guo, W. Wen, S. Gu, J. Wang, and J. Fan, ”Crop Phenomics: Current Status and Perspectives,” Front. Plant Sci., vol. 10, p. 714, 2019.

[36] Y. Waisel, A. Eshel, and U. Kafkafi, Plant-Roots–The Hidden Half, 3rd ed., New York, NY, USA: Marcel Decker, 2002, pp. 775–776.

[37] H. Poorter and O.W. Nagel, ”The role of biomass allocation in the growth response of plants to different levels of light, CO2, nutrients and water: A quantitative review,” Funct. Plant Biol., vol. 27.

[38] J. A. Davila, M. P. S. Hoz, Y. Loarce, and E. Ferrer, ”The use of random amplified microsatellite polymorphic DNA and coefficients of parentage to determine genetic relationships in barley,” Genome, vol. 41, pp. 477-486, 1998.

[39] N. N. Denwar, F. J. Awuku, B. Diers, F. Addae-Frimpomaah, G. Chigeza, R. Oteng-Frimpong, S. K. Puozaa, and M. T. Barnor, ”Genetic diversity, population structure, and key phenotypic traits driving variation within soyabean (*Glycine max*) collection in Ghana,” Plant Breeding, vol. 138, no. 5, pp. 577-587, 2019.

[40] A. Ghaderi, M. W. Adams, and A. M. Nassib, ”Relationship between genetic distance and heterosis for yield and morphological traits in dry edible bean and faba bean,” Crop Science, vol. 24, pp. 37-42, 1984.

[41] A. A. Ghafoor, M. A. Ahmad, M. A. Rabbani, Z. Sharif, and M. A. Zahid, ”Genetic diversity in black gram (Vigna mungo (L.) Hepper),” Field Crops Research, vol. 69, pp. 183-190, 2001.

[42] E. F. Hashash, ”Genetic diversity of soybean yield based on cluster and principal component analyses,” Journal of Advanced Biology, vol. 10, no. 3, pp. 1-4, 2016.

[43] IndexBox. (n.d.). IndexBox - Market Research Reports and Analytics Platform. Retrieved from www.indexbox.io”

[44] United States Department of Agriculture. (2022). Agricultural Research Service. Retrieved from https://www.usda.gov/our-agency/agricultural-research-service

[45] M. Li, Y. Liu, W. Jiang, C. Wang, X. Yang, D. Li, X. Zhang, C. Xu, Y. Zhang, W. Li, and L. Zhao, ”Identification of traits contributing to high and stable yields in different soybean varieties across three different Chinese latitudes,” Front. Plant Sci., vol. 10, p. 1642, 2020.

[46] I. A. Khatab, A. R. Morsy, W. M. Fares, and T. Kumamaru, ”Genetic diversity of soybean genotypes revealed by agro-morphological and SSR markers,” J. Fac. Agric., Kyushu Univ., vol. 61, pp. 79-87, 2016.

[47] Marques de Sa, J. 2007. Estimating Data Parameters. Applied Statistics Using SPSS, STATISTICA, MATLAB and R, 81-109.

[48] Phylogenetic analysis of selected medicinal plant species from Pakistan: DNA barcoding approach,” Pak. J. Bot., vol. 50, no. 2, pp. 553-560, 2018.

[49] S. Kato, T. Sayama, F. Taguchi-Shiobara, A. Kikuchi, M. Ishimoto, and M. Cober, ”Effect of change from a determinate to a semi-determinate growth habit on the yield and lodging resistance of soybeans in the northeast region of Japan,” Breeding Science, vol. 69, pp. 151-159, 2019, doi: 10.1270/jsbbs.18112.

[50] I. A. Khatab, A. R. Morsy, W. M. Fares, and T. Kumamaru, ”Genetic diversity of soybean genotypes revealed by agro-morphological and SSR markers,” Journal of the Faculty of Agriculture, Kyushu University, vol. 61, pp. 79-87, 2016.