GENETIC STUDIES OF SOYBEAN [*Glycine max* (L.) Merrill] GENOTYPES UNDER NAGALAND CONDITIONS

Thesis

Submitted to

NAGALAND UNIVERSITY

In partial fulfilment of requirements for the degree

Of

Doctor of Philosophy

In

GENETICS AND PLANT BREEDING

By

ASHNA AKBAR

Admin. No. Ph-178/15; Reg. No. Ph-821/2018

Department of Genetics and Plant Breeding
School of Agricultural Sciences
Nagaland University, Medziphema Campus -797106
Nagaland

2022

DECLARATION

I, Mrs. **ASHNA AKBAR** hereby declare that the subject matter of this thesis is the record of work done by me, that the contents of this thesis did not form the basis of the award of any previous degree to me or to the best of my knowledge to anybody else, and that the thesis has not been submitted by me for any research degree to any other university/institute.

This is submitted to SAS, Nagaland University for the degree of Doctor of Philosophy in **Genetics and Plant Breeding**.

Date: 7/12/2022

Place: SAS (Medziphema)

.....

(**ASHNA AKBAR**)

Supervisor
(Dr. H. P. Chaturvedi)

NAGALAND UNIVERSITY
MEDZIPHEMA CAMPUS
SCHOOL OF AGRICULTURAL SCIENCES
MEDZIPHEMA- 797106, NAGALAND

Dr. H.P. Chaturvedi
Assistant Professor
Department of Genetics and Plant Breeding

CERTIFICATE – I

This is to certify that the Thesis entitled "**Genetic studies of soybean [*Glycine max* (L.) Merrill] genotypes under Nagaland conditions** submitted to Nagaland University in partial fulfilment of the requirements for the degree of Doctor of philosophy in the discipline of Genetics and Plant Breeding is a record of research work carried out by Mrs. Ashna Akbar, Registration No. 821/2018 under my personal supervision and guidance.

The results of the investigation reported in the thesis have not been submitted for any other degree or diploma. The assistance of all kinds received by the student has been duly acknowledged.

Dated: 7/12/2022

Place: SAS (Medziphema)

.....

(Dr. H.P. Chaturvedi)
Supervisor

NAGALAND UNIVERSITY
MEDZIPHEMA CAMPUS
SCHOOL OF AGRICULTURAL SCIENCES
MEDZIPHEMA- 797106, NAGALAND

CERTIFICATE – II

VIVA VOCE ON THESIS OF DOCTOR OF PHILOSOPHY IN GENETICS AND PLANT BREEDING

This is to certify that the thesis entitled "**Genetic studies of soybean [*Glycine max* (L.) Merrill] genotypes under Nagaland conditions**" submitted by Mrs. Ashna Akbar Admission No. Ph-178/15, Registration No. 821/2018 to the NAGALAND UNIVERSITY in partial fulfilment of the requirements for the award of degree of Doctor of Philosophy in Genetics and Plant Breeding has been examined by the Advisory Board and External examiner on 12-09-2023

The performance of the student has been found Satisfactory/~~Unsatisfactory~~.

Members:	**Signature**
1. Dr. H.P. Chaturvedi (Supervisor)	12.09.2023
2. Dr. Akshay Talukdar. (External examiner)	12/9/23
3. Pro-Vice Chancellor Nominee (Dean, SAS)	12.09.23
4. Dr. Kigwe Seyie (member)	12/9/23
5. Dr. Pankaj K. Shah (member)	12.09.23
6. Dr. L. Daiho (member)	12/9/23
7. Dr. S.P. Kanaujia (member)	12.09.23

Head
(Dept. of Genetics & Plant Breeding)

Dean
(School of Agricultural Sciences)

ACKNOWLEDGEMENTS

First and foremost, I offer my obeisance to the 'Almighty' for his boundless blessing, which accompanied me in all the endeavors. I would like to extend my sincere gratitude to many people, who so generously contributed to the work presented in this thesis.

Fervently and modestly, I extol my profound sense of gratitude and heartfelt devotion to my supervisor Dr. H.P. Chaturvedi, Assistant Professor, Department of Genetics and Plant Breeding , for his expert guidance, genuine cooperation, inspiration, valuable advice, sustained help, keen interest and concrete suggestions, right from the commencement of my research work to ship-shaping of the manuscript. His expedient recommendations, constructive criticism and assiduous guidance will be forever cherished.

I wish to express my thanks to my advisory board members, Dr. Kigwe Seyie, Associate Professor, Department of Genetics and Plant Breeding, Dr. Pankaj Shah, Assistant Professor, Department of Genetics and Plant breeding, Dr. L. Daiho, Professor, Department of Plant Pathology, and Dr. S. P. Kanaujia Associate Professor, Department of Horticulture for their valuable suggestion and practical approach that has resulted in the work presented. I would also like to convey my sincere thanks to Dr. Malini B. Sharma, Professor and head, Department of Genetics and Plant Breeding.

I am thankful to all the technical staffs, Department of Genetics and Plant breeding for their valuable help in every possible way during my PhD research work. I deem it my privilege in expressing my fidelity to Indian Institute of Soybean Research (IISR) fortheir valuable help and cooperation with my

molecular work. My special thanks to my husband Firdous Zia, who supported me at every turn giving me strength to proceed on the bad days and cheering me ahead on the good. My heartfelt gratitude to my parents, grandparents, in-laws, brother, sisters and friends for their constant prayer, moral support throughout my research work. I owe my success to them. Finally, I wish my humble thanks to one and all who have directly or indirectly contributed to the conduct of the study.

Place: (SAS), Medziphema (Ashna Akbar)

Date: 7/ 12/ 2022

CONTENTS

CHAPTER	TITLE	PAGE NO.
1	**INTRODUCTION**	**1-6**
2	**REVIEW OF LITERATURE**	**7-38**
	2.1 Genetic variability in soybean	7-13
	2.2 Heritability and genetic advance	13-16
	2.3 Correlation and path study for seed yield and its components	16-25
	2.4 Genetic divergence	25-32
	2.5 Application of SSR markers to evaluate the genetic diversity in soybean germplasm	32-38
3	**MATERIALS AND METHODS**	**39-76**
	3.1 Experimental details	41-49
	3.2 Observations recorded	49-52
	3.3 Statistical analysis	52-76
4	**RESULTS AND DISCUSSION**	**77-130**
	4.1 Analysis of variance	78
	4.2 Mean performance of genotypes	78-85
	4.3 Genotypic and phenotypic coefficient of variation (GCV and PCV)	90-92
	4.4 Heritability (h^2) and genetic advance (GA)	92-95
	4.5 Correlation analysis	97-100
	4.6 Path coefficient analysis	102-108
	4.7 Genetic divergence	110-116
	4.8 Evaluation of genetic diversity using SSR molecular markers	118-130

DISCUSSION **131-158**

Analysis of variance 131-132

Mean performance of genotypes 132-135

Genotypic and phenotypic coefficient of variation (GCV and PCV) 136-138

Heritability (h^2) and genetic advance (GA) 138-140

Correlation analysis 140-144

Path coefficient analysis 144-147

Genetic divergence 147-153

Evaluation of genetic diversity using SSR (Simple Sequence Repeat) 154-158

SUMMARY AND CONCLUSIONS **159-166**

REFERENCES **i-xix**

APPENDICES **A-C**

LIST OF TABLES

TABLE NO.	TITLE	PAGE NO.
3.1	Meteorological data during the crop investigation period from June-December (2017 and 2018)	40
3.2	Layout plan details	41
3.3	Details of the soybean genotypes and their place of collection with source	46-47
3.4	Pooled analysis of variance (ANOVA) for RBD	53
3.5	SSR primers used in the study	68-69
3.6	Concentration of the reagents which were used in PCR	73
3.7	Thermo-cycler profiling for amplification of SSR markers	73
4.1	Analysis of variance data for traits in soybean genotypes that contribute to seed yield	79
4.2	Fourteen morphological traits' mean, range, and coefficient of variance	82
4.3	Mean performance for fourteen characters of soybean genotypes	86
4.4	Genetic parameters for various morphological traits in soybean	93
4.5	Correlation between seed yield and its contributing traits in soybean at both Genotypic and phenotypic level	101
4.6	Direct (diagonal) and Indirect (above and below diagonal) path effects on seed yield in soybean through different characters both at genotypic and phenotypic level	109
4.7	Soybean genotypes in various clusters	111
4.8	Percent contribution of 14 characters for divergence	113
4.9	Average D^2 and D (parenthesis) values within and between clusters	115

4.10	Cluster mean values for soybean genotypes under study four fourteen characters	117
4.11	Consolidated genomic DNA for 20 genotypes of soybean	118
4.12	Specifics of 25 SSR loci used in the study for twenty soybean treatments	121
4.13	SSR amplicons/bands generated by primers in 20 genotypes of soybean	122
4.14	Genetic diversity parameters obtained at each locus across 20 soybean genotypes	123
4.15	Jaccard Similarity coefficient values in soybean under study	124
4.16	Average similarity Index of 20 soybean genotypes	128
4.17	Genotypes under different clusters based on UPGMA clustering	130
4.18	Diverse genotypes based on inter cluster distances and superior *per se* performance for the traits under investigation	153

LIST OF FIGURES

FIGURE NO.	CAPTION	PAGE NO.
1.1	Global soybean production scenario	3
3.1	Temperatures (minimum and maximum) during crop growing season-2017 & 2018	40
3.2	Rainfall recorded over crop growing season-2017 & 2018	41
3.3	Layout of the Experiment in Randomized Block Design	42
4.1	Scatter diagram representing relation between number of pods per plant (NPo/P), days to 50% flowering (DFF), biological yield per plant (BY/P), plant height (PH), days to maturity (DM), and seed yield/plant (SY/P)	87
4.2	Scatter diagram representing relation between number of pods per cluster (NPo/C), number of primary branches per plant (NPB/P), number of clusters/plant (NC/P) and seed yield/plant (SY/P)	88
4.3	Scatter diagram representing relation between hundred seed weight (HSW), oil% and seed yield/plant (SY/P)	89
4.4	Graphical representation of GCV and PCV for various associated traits of seed yield	91
4.5	Heritability and genetic advance as a % of mean presented in bar diagram	96
4.6	Clustering of soybean genotypes by Tocher's method	112
4.7	Dendrogram showing genetic relationship among 20 soybeans based on UPGMA clustering.	129

LIST OF PLATES

PLATE NO.	CAPTION	PAGE NO.
1	Genotypes under study	43-46
2	SSR profiles with primer Satt 155	125
3	SSR profiles with primer Satt 411	125
4	SSR profiles with primer Satt 588	126
5	SSR profiles with primer Satt 196	126
6	Layout of experimental farm	A
7	Field view of Initial germination stage	A
8	Crop at flowering stage - (a) purple and (b) white flowers	B
9	Pod formation stage	B-C
10	Crop at physiological maturity	C

LIST OF ABBREVIATIONS

°C	:	Degree Celsius
%	:	Percent
≤	:	Less than or equal to
≥	:	Greater than or equal to
>	:	more
<	:	less
pH	:	Potential of hydrogen
atm	:	Atmospheric pressure
g	:	gram
Sq. m	:	Square meter
cm	:	centimetre
mM	:	millimolar
nm	:	nanometer
O.D	:	Optical density
µl	:	microlitre
etc.	:	Et cetera
viz.	:	Videlicet (Namely)
et al.	:	et alli (and others)
per se	:	By itself
i.e.	:	That is or in other words
d.f	:	Degree of freedom
No.	:	number of observation
S.E	:	Standard Error
TE	:	Tris-EDTA
CAU	:	Central Agricultural University, Imphal
JNKVV	:	Jawaharlal Nehru Krishi Vishwa Vidyalaya
EDTA	:	Ethylene diamine tetra acetic acid
NTSYS	:	Numerical taxonomy system
UPGMA	:	Unweighted Pair Group Method with Arithmetic mean

ABSTRACT

One of the most important oilseed crops farmed for its edible oil and protein in India and throughout the world is soybean [*Glycine max* (L.) Merrill]. Seed yield in soybean is a quantitative character which is regulated by a number of yield-contributing traits. The choice of a suitable type of soybean should be made based on yield in addition to other yield contributing factors. In light of this, the current study was carried out during the *kharif* seasons of 2017 and 2018.In order to assess genetic variability of soybean germplasm based on agro-morphological traits and molecular markers a set of 20 distinct soybean genotypes, including indigenous genotypes obtained from various states in the north-east India and check variety JS-9752 were used in the field experiment, following randomized block design with three replications. According to the pooled analysis of variance, there were remarkably significant differences among all the genotypes for the analyzed fourteen characters. Based on mean values for majority of the traits, the five most prevalent genotypes were G1 (Assam), G10 (Nagaland) G11 (Nagaland) G12 (Nagaland) and G9 (Nagaland).Estimates of phenotypic and genotypic coefficients of variation showed that, PCV values were higher than GCV values. The number of clusters per plant, number of pods per plant, hundred seed weight, biological yield per plant, and seed yield per plant showed high genotypic and phenotypic coefficients of variation in the current study. Days to 50% flowering, plant height, number of clusters per plant, number of pods per plant, and 100 seed weight showed high heritability values with high GA%.The results of investigations using correlation and path analysis suggested that number of pods per plant, number of pods per cluster, and number of days to maturity would improve seed yield per plant. Using D^2 statistics, five groups were created. Clusters I and II had the maximum inter-cluster distance and showed high genetic diversity. Inter varietal hybridization programmes can be employed with the genotypes from these two clusters to produce recombinants with high yields. Genotypes from clusters III and IV were most preferred because they exhibited values higher than average mean for desired agronomic traits and quality attributes in addition to higher seed yields per plant.In order to ascertain the genetic diversity and relatedness among 20 soybeans, SSR analysis was used. 18 of the 25 SSR primer pairs employed were able to amplify polymorphic SSRs from each of these genotypes. With an average of 1.77 alleles per locus, a total of 32 polymorphic

alleles were found.The PIC value of the markers ranged from 0.180 to 0.882, with an average of 0.587. Furthermore, 13 out of 25 SSR markers had PIC values above 0.5 and were very informative, making them suitable for DNA fingerprinting. Sat_409, Satt 055 and Satt 588were the top three primers with PIC values above 0.5 and 100% polymorphism. The range of pair-wise similarity coefficients for all genotypes was 0.25 to 0.71, with an average of 0.56. The gene diversity/expected heterozygosity (He) ratio, which ranged from 0.14 to 0.65 with an average of 0.42, showed that there was a sizable amount of genetic variation among the genotypes. Using the UPGMA analysis, the genotypes were divided into two large clusters with 10 and 9 genotypes each. Additionally, one outlier was found. Comparative analysis of diversity based on morphological and molecular features showed that genotypes from different eco-geographical regions clustered together, while genotypes from the same eco-geographical region distributed into different clusters, indicating that geographic diversity does not always correspond to genetic diversity. The data generated and the potential genotypes identified in this study would be useful in developing high yielding soybean varieties for Nagaland.

Key words: Genetic diversity, genetic variability, SSR primer, soybean and yield.

CHAPTER-I

INTRODUCTION

INTRODUCTION

One of the most profitable, adaptable, and crucial legumes in the world is the soybean. It can be cultivated in a wide range of agro climatic conditions, with a wide range of management techniques and for a range of end users. *Glycine max* (L.) Merrill is the scientific name for soybean (2n = 40) also called soja bean is a member of the family Fabaceae (Leguminaceae), order Fabales, and subfamily Faboideae (Papilionoideae). It is an erect, branched plant that has self pollination as a mode of reproduction. Two colours of flowers either white or purple are common. Each pod bears one to four seeds, which might be brown, black, yellow, green, or bicoloured.

Originally a tropical crop, soybeans are now grown in subtropical and temperate climates. Soybean may be grown on a broad variety of well-drained soils, although it does best on clay loams, where the lowest growing temperature is around 10°C. Soybean is moderately tolerance to salinity, the ideal pH range for its cultivation is between 6 and 6.5. Shallow water tables can have a negative impact on the yield, particularly in the early stages of development. The crop is vulnerable to water logging, especially in the early stages. Under ideal conditions, soybean seeds germinate epigeally and start to grow into seedlings in 4-5 days. If the soil moisture stress reaches 6.5 atm, germination is impossible. The ideal germination temperature ranges from 30°C to 35°C. The root system starts to form and the rate of root penetration is highest during early blooming. Being a legume, soybeans have the facility to fix nitrogen in a symbiotic relationship with *Rhizobium spp*. Nodule formation starts 10 days after emergence andnitrogen fixation begins two weeks later. Nodulation only occurs in spaces between the little emerging root hair at the root tip, not on the adult root walls.

Due to its high productivity, financial success and considerable impact on soil fertility, soybeans play a large part in the world's oilseed agriculture. Of all oilseed crops, soybeans occupy the largest area contributing 37% to total oilseeds production and 25% of the edible oil production (Lakshmy *et al*. 2020). It is cultivated on 130.43 million hectares of land worldwide, producing 352.74 metric tonnes with an average productivity of around 2.70 metric tonnes per hectare (USDA 2022). With a total yield of around 11.9 metric tonnes and an average productivity of about 0.95 metric tonnes per hectare, soybean agriculture in India has reached approximately 12.5 million hectares (USDA 2022). Vegetable oil consumption per person is anticipated to expand quickly, reaching 16.44 and 19.16 kg per year by 2020 and 2050 respectively, with India ranking as the world's fifth-largest consumer for the commodity (Lakshmy *et al*. 2020). Hence, soybean has a large scope in the oilseed scenario of the country.

The total area under cultivation of soybean is limited in North Eastern states, although the productivity is higher. Practically all of Nagaland's districts farm soybeans (Anonymous 2016). In Nagaland, there are 25,040 hectares of soybeans planted, yielding a total of 31,520 tonnes (Anonymous 2017). Compared to other cultivated crops, it yields more usable protein per hectare-at least three times more than rice, wheat, or maize (Lakshmy *et al.* 2020). The contribution of soybean production by different countries across the globe is shown in figure 1.1.

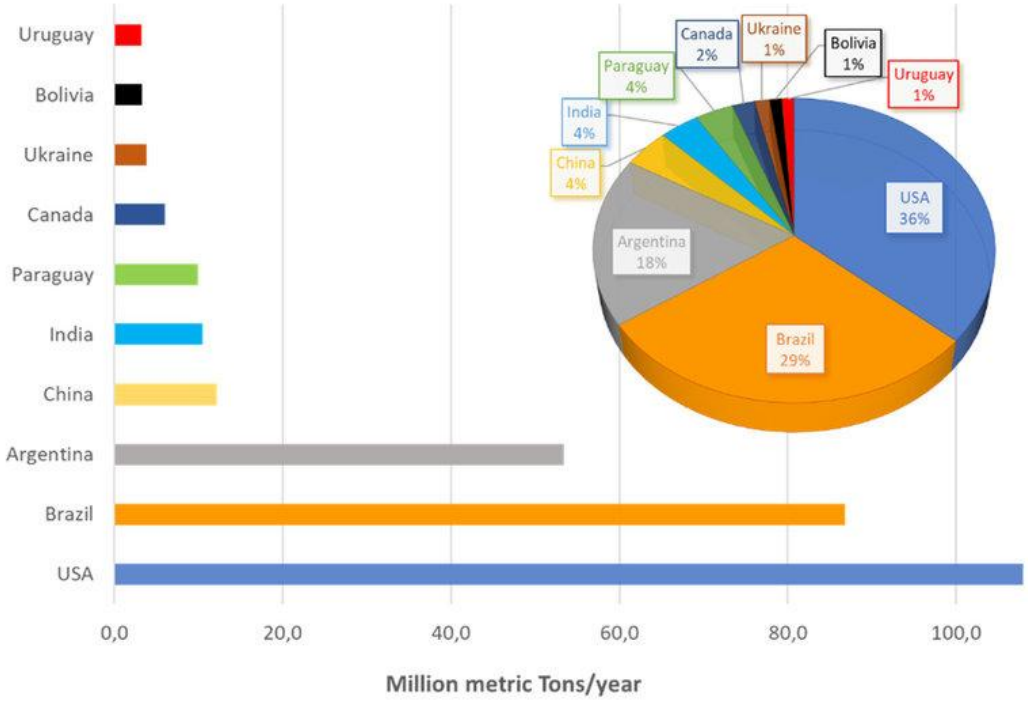


Figure 1.1: Global soybean production scenario

While soybeans have long been a staple food in Asian nations (Liu 1997 and Jo *et al.* 2021), soybean cultivation in Western nations is primarily focused on producing high-protein livestock feed and vegetable oils. Products made from soybeans include soy sprouts, soymilk, fermented soy dishes, soy pulp, miso, tempeh, soy sauce, and tofu. They are used to create spreads and pastes as well as dairy product substitutes (cheese, soy milk), industrial additives (like cosmetics, plastics, and colours), food additives (like soy lecithin), and meat substitutes (Chen *et al.* 2012 and Modgil *et al.* 2021). There are many soy products available right now, and demand for them is always rising.

It is a significant source of high-quality protein (37–42%) and oil (18–22%), and it has 85% of the fats which are unsaturated, with 55% of them being poly- unsaturated fatty acids, and with two important fatty acids linoleic and

linolenic acid usually generated by the human body (Balasubramaniyan and Palaniappan 2004). For this reason, it is also referred to as Wonder Seed, Miracle Crop, and Golden Bean. It contains 25–30% sugars with no starch (useful for diabetic patients), 4%–5% vitamins, and antioxidants, such as ascorbic acid (9–10 mg/100 g of sprouted soybean), beta-carotene (0.2 mg/100 g of sprouted soybean) and 0.3% of isoflavones (daidzein and genistein). According to Garcia *et al*. (1998), among plant-based protein sources, soybean protein is believed to have the highest biological value. When it comes to exogenous amino acids like methionine, phenylalanine, valine, leucine, isoleucine, lysine, threonine, and tryptophan, soy proteins and animal proteins are comparable in their amino acid profiles. The corresponding percent rates of their respective contents in soybean grain are reported in the reference literature by Chen *et al*. 2012 and Modgil *et al.* 2021: Leucine ca. makes up 8 g in every 100 g of protein; valine ca. 5 g/100 g of protein; lysine ca. 6.5 g/100 g of protein; and phenylalanine ca. makes up 4 g per 100g of protein. Compared to animal proteins, soybean protein contains less sulphur containing amino acids (Kudelka *et al.* 2021).

Soybean is considered to be originated in China and several centuries ago, it is likely that traders from Indonesia imported soybeans into India through Myanmar after crossing the Himalayas from China (Shurtleff and Aoyagi 2010). Two native soybean varieties, "yellow cultivar" and "dark brown cultivar," are planted and harvested between May and June. To generate a range of fermented and non-fermented foods, the dried soybean seeds are naturally used in eastern Nepal, the Darjeeling highlands, Sikkim, north-eastern India, and southern Bhutan near the Mongolian races (Tamang 2009). Since the beginning of time, every household in Nagaland has consumed Akhuni, a traditional dish made of fermented soybeans, especially in the Zunheboto district. *Akhuni* is a fermented

soy bean product with excellent gastronomic and health properties that is used as a Naga food addition.

Global programmes for *Glycine max* (L.) Merrill hybridization resulted in hundreds of premium cultivars and thousands of breeding lines each year. The creation of these breeding lines improved genetic potential in soybean. Improvement of the crop and its survival in nature depends on the diversity and genetic variability. Plant breeders use the genetic diversity in the plants to create new and improved cultivars with desirable features including those attributes favoured by farmers and other stakeholders. Since the beginning of agriculture, natural genetic variability within the crop species has been used to provide enough food for sustenance, however with the increased demand for food along with the increased population pressure, it needed the development of genotypes with higher productivity per unit area. The present focus is on both productivity and quality components of major food crops in order to provide humans with a balanced diet (Bhandari *et al.* 2017).

Genetic variability can be studied using agronomic and biochemical variables as well as molecular marker polymorphisms. The knowledge of genetic variability, genetic parameters, and their application, which aids breeders in precise selection processes, is essential for the ongoing progress of genetic breeding of soybeans. Heritability, genetic advance, and the phenotypic and genotypic coefficient of variation are crucial genetic indicators that help plant breeders decide on the optimum breeding approach.An index that aids in forecasting yield responses to changes associated with a given character is the degree of relatedness between significant plant traits (Malek *et al.*2014).Therefore, In order to increase yields for soybean, it is crucial to identify essential traits linked to yield and other contributing factors.

The presence of genetic diversity in wild species, allied species, breeding stocks, mutant species, lines, etc., can be the sources of desirable alleles that can help bean breeders develop varieties which are tolerant to changing climatic conditions.Several genes in cultivated and cultivable crops need to be reserved in the form of germplasm tools for ever-changing breeding objectives. Prerequisites for identifying significant genotypes and selection criteria for soybean improvement include taking into account the genetic diversity and relationships among soybean genotypes based on their morphological traits and molecular profile. Breeding plants with agronomically and economically superior traits is the ultimate goal of the plant breeders. The presence of genetic variability between certain plant species helps breeders in selecting superior genotypes. The best use of soybean germplasm requires an understanding of genetic variability and diversity studies. The current inquiry was conducted with the aforementioned perspective in mind to achieve the following goals:

- To assess genetic variability of soybean germplasm based on agro-morphological traits.
- To evaluate genetic diversity in soybean germplasm by using molecular markers.

CHAPTER II
REVIEW OF LITERATURE

REVIEW OF LITERATURE

In order to meet the requirements of ever-increasing population, it is continuously necessary to increase the production of soybean. One approach to increase the oil production is the quantitative improvement in oil content of oilseed crops. Soybean has become a major oilseed crop in India to be followed by groundnut and Indian mustard. Moreover, due to export of de-oiled cake, it has also gained importance in foreign trade. In the present study an attempt was made to study variability, diversity and association among seed yield and dependent component characters in some genotypes of soybean commonly found in north-east India. The literatures pertaining to the present study have been reviewed as follows:

2.1Genetic variability in soybean:

Khumukcham*et al*. (2022) concluded results of fifty soybean genotypes including five checks which were evaluated for fourteen quantitative characters. Analysis of variance indicated that mean sum of square due to genotypes were found significant for all the traits. The phenotypic coefficient of variation (PCV) was found higher than genotypic coefficient of variation (GCV) indicating the influence of environment in the expression of the traits under study. The high values of PCV and GCV were observed for number of seeds per plant followed by seed yield per plant, biological yield, pod bearing length, number of pods per plant and plant height. Days to maturity exhibited low GCV and PCV.

Sahoo (2022) conducted experiment with 9 crosses of soybean in compact family block design for 14 quantitative traits during kharif 2019 at, Pantnagar, Uttarakhand to assess the genetic variability, heritability and genetic advance.

Genetic variability was found among the experimental materials for all traits. Phenotypic coefficient of variation was higher than genotypic coefficient of variation for most of the traits. Seed yield per plant showed highest genotypic coefficient of variation (29.1%).

Goonde and Ayana (2021) conducted an experiment to evaluate the variability on 100 soybean genotypes. The results showed significant variance among the genotypes. High genotypic and phenotypic coefficient of variation was recorded for number of pods per plant, number of primary branches per plant, biological yield and seed yield.

Kuswantoro *et al*. (2021) identified the useful traits that may be used as selection criteria in soybean breeding. From April to July 2020, 100 different soybean genotypes were planted in the Muneng Agricultural Technology Research and Assessment Installation. The findings demonstrated that there was significant genetic variation in the number of days to maturity, number of branches/plant, the number of nodes/plant, the weight of 100 seeds and the seed production.

Jandong *et al*. (2020) studied 20 soybean genotypes and revealed significant variation among the genotypes for 7 characters namely days to 50% flowering, plant height (cm), number of leaves, number of branches, number of pods, pod weight and seed yield. The estimate of PCV was higher than the GCV. Moderate PCV and GCV were recorded for days to 50 % flowering and high PCV and GCV recorded for rest of 6 characters.

Kumar *et al*. (2020) studied 10 characters on 307 soybean germplasm lines and found that the investigation had significant level of variability present. High values of PCV and GCV were recorded for characters namely days to 50% flowering, plant height, number of branches per plant, number pods per plant, 100

seed weight, seed yield per plant, biological yield per plant and harvest index recorded. Whereas high PCV coupled with moderate GCV values were noted for harvest index.

Sharma and Lal (2020) studied genetic variability among 40 genotypes for 11 quantitative traits. For all the investigated traits, analysis of variance revealed substantial differences across the 40 soybean genotypes. The phenotypic coefficient of variation was greater than the genotypic coefficient of variation, and the difference between PCV and GCV was small for the majority of the characters. This suggested that the environment had a lesser impact on how these characters are expressed in soybean germplasm. The number of pods per plant and seed yield per plant both had high values of GCV and PCV.

Krisnawati and Adie (2019) studied 16 soybean genotypes and found high level of significant variability for days to maturity, 100-seed weight (g) and seed yield (t/ha).

Verma (2019) looked at 100 soybean genotypes with 12 characters and found that the characters had highly significant genotype variability. Number of seed per plant had the highest genotypic coefficient of variation (GCV), followed by seed yield per plant, number of pods per plant, pod bearing length and plant height, with days to maturity having the lowest value. The GCV for the number of seeds per plant was high, indicating that there is a lot of room for yield improvement in the current soybean gene pool.

Ibrahim *et al*. (2018) conducted experiment on eight soybean varieties, which were distributed in a randomized complete block design and duplicated four times for the assessment of 15 traits in order to determine variability, heritability and correlation. The results of the analysis of variance showed that there were four

characters with a significant difference and seven characters with a highly significant difference. The high genotypic and phenotypic coefficient of variation along with low environmental coefficient of variation was observed in this study, indicated the existence of variability.

Joshi *et al*. (2018) studied variance for days to 50 percent flowering, number of nodes per plant, number of pods per plant and 100-seed weight and found significant amount of variability for future study.

Kuswantoro *et al*. (2018) reported variability among 16 varieties of soybean for traits such as, plant height, days to flowering, days to maturity, number of reproductive nodes, number of branches per plant, number of pods per plant, weight of 100 seeds, number of unfilled pods per plant, and grain yield and indicated broad and narrow GCV for these traits.

Neelima *et al*. (2018) experimented on genetic variability in 124 soybean genotypes for 13 characters. All characters show wide range of variation. The characters like days to maturity, oil content and protein content shows the lowest genotypic coefficient of variation (GCV) and rest of the other characters days to initial flowering, days to 50% flowering, plant height, number of primary branches/ plant, number of nodes/plant, number of cluster/plant, number of pods/plant, 100-seed weight, seed yield per plant and seed yield per row showed high level of GCV.

Akkamahadevi and Basavaraja (2017) assessed 17 vegetable soybean genotypes for 11 traits at the Main Agricultural Research Station, University of Agricultural Sciences, Dharwad, in kharif 2015. The results of the analysis of variance showed that all of the genotypes for each character differed significantly. For 100 seed weight, the largest genotypic coefficient of variation was observed.

Jain *et al*. (2017) studied the genetic variability, phenotypic, genotypic, and environmental coefficient of variation in 24 genotypes of soybean for 9 traits. All the traits have shown significant variation indicated by analysis of variance. Test weight, pod number per plant, height of plant, and harvest index exhibited the maximum genotypic and phonotypic coefficient of variation.

Akram *et al*. (2016) estimated genetic variability in eleven soybean genotypes and found significant variation among the genotypes. The highest phenotypic and genotypic variation was reported for number of seeds per plant and the lowest for pod length.

Dubey *et al*. (2015) compared the genetic variability of 50 genotypes of soybean and found that phenotypic coefficient of variation was greater than those of the genotypic coefficient of variation for all the traits such as plant height, occurrence of nodes and branches per plant, pod number per nodes as well as per plant number of seeds per pod and per plant, weight of 100 seeds, harvest index, biological yield per plant, and seed yield per plant.

Jain *et al*. (2015) reported wide range of variability in 41 soybean genotypes for 9 characters namely days to 50% flowering, days to maturity, plant height, number of primary branches per plant, number of pods per plant, number of seeds per pod, 100-seed weight (g), biological yield and seed yield per plant (g).

Pagde *et al*. (2015) examined 30 soybean strains for genetic variability. Plant height, seed yield per plant, secondary branches per plant, test weight and seeds per pod showed high GCV and PCV.

Malek *et al*. (2014) studied genetic variability, genetic diversity, and character association between 27 mutants of soybean mutants with four mother

genotypes. For nine physical traits, analysis of variance revealed considerable differences between mutants and mothers. Numerous traits with slight differences between phenotypic and genotypic coefficients of variation suggested less environmental effect on their expression. High GCV for branch number, pod number, height of plant, and seed weight can be considered as favourable traits for soybean development by phenotypic selection, and high predicted genetic gain can be attained.

Reni and Rao (2013) experimented on forty five genotypes of soybean (Glycine max (L.) Merrill) of diverse origin in randomized block design with three replications for variability, heritability and genetic advance and concluded that analysis of variance revealed highly significant differences among the genotypes for the all the characters. High PCV coupled with high GCV was observed for branches per plant, pods per plant, biological yield, harvest index and yield per plant indicated the presence of wider adaptability for these traits in the genotypes studied, suggested less influence of environment in the expression of characters.

Athoni and Basavaraja (2012) concluded that analysis of variance revealed the prevalence of significant difference among the genotypes for all the 11 characters studied. Plant height was the only character which showed high phenotypic and genotypic co-efficient of variation while days to maturity, number of nodes per plant and oil content recorded a low phenotypic and genotypic coefficient of variation and rest of characters recorded moderate phenotypic and genotypic coefficient of variation in soybean

Patil *et al*. (2011) measured the variability of 11 characters in soybeans. Plant height followed by seed yield per plant and number pods per plant had the highest genotypic and phenotypic coefficient of variation while it was lower for days to 50 percent flowering, days to maturity and protein content.

2.2 Heritability and genetic advance:

Adewusi (2022) examined the genetics of seed yield and the traits that contribute to it in the F2 population of soybean genotypes. A randomised complete block design with three replications was used to set up the field experiment. Number of seeds and number of pods had heritability in the broad sense ranging from 40.00 % to 99.97 % respectively. For days to maturity and seed yield, the genetic advance mean varied between 10.03 and 130.17, respectively.

Dutta *et al*. (2021) assessed forty soybean genotypes for yield and other yield contributing variables for two successive years, kharif 2018 and 2019 to determine genetic diversity, heritability as well as genetic advance. For all of the traits tested, plant seed yield, seeds number per pod, oil content, days to 50% flowering, number of branches, pods number per plant, plant height of plant, and weight of 100 seeds, had higher heritability values along with high genetic advance.

Goonde and Ayana (2021) revealed that high heritability coupled with high genetic advance as % mean was recorded for days to 50% emergence, grain filling period, biological yield and seed yield per plant. Such character exhibited additive gene effects and direct selection would be rewarding.

Azevedo *et al*. (2020) determined the genetic characteristics of soybean populations produced through crossings between various food and grain genotypes and to identify progenies with desirable agronomic and commercial qualities. Plant height at maturity, first pod insertion height, lodging, agronomic value, number of pods per plant, number of days until maturity, number of branches, number of nodes, 100-seed weight, and grain yield per plant were the traits that

were evaluated. Most of the traits had strong heritability, which suggested good potential for selecting superior genotypes.

El-Mouhamady and El-Metwally (2020) used imported soybean lines for the improvement of local varieties and observed highly significant values for heritability and genetic advance in most studied traits.

Pawar *et al.* (2020) studied 11 characters of 30 soybean genotypes and noted highest heritability value for number of pods per plant and days to 50% flowering. The traits like seed yield per plant, number of pods per plant and plant height also observed high values of genetic advance.

Verma (2019) found that seed yield per plant, pods per plant, pod bearing length (cm), plant height (cm), number of primary branches and hundred seed weight had high heritability with high genetic advance as percent of mean.

Joshi *et al*. (2018) conducted experiment on soybean genotypes and found highest heritability for 100-seed weight followed by other characters. High genetic advance as percent of mean for number of nodes per plant, number of pods per plant, seed weight, number of branches per plant, plant height, seed yield per plant and days to 50% flowering exhibited additive genetic effects.

Neelima *et al*. (2018) observed variability among 124 diverse soybean accessions for 13 different characters and found that the most of the characters had high heritability. High heritability was coupled with high genetic advance as percent of mean found for days to initial flowering, days to maturity, plant height, number of pods per plant and seed yield per row.

Chandrawat *et al*. (2017) observed that the traits pods per plant, plant height, yield per plant, branches per plant and 100-seed weight had high

heritability and high genetic advance, indicating the existence of additive gene action and the need for population improvement by selection in soybean genotypes.

Ghiday *et al*. (2017) determined the presence of variability for desirable features among 22 promising new genotypes of soybean and three checks. Estimates of broad sense heritability for different traits ranged from 74.62 to 99.73 %.

Jain *et al*. (2017) studied 24 genotypes on 9 traits of soybean. In this study high heritability coupled with high genetic advance were found for number of pods per plant, harvest index and plant height.

Irshad *et al.* (2016) reported high broad sense heritability estimation in the range of 73.3 to 97.2% for days to maturity and 100 seed weight respectively.

Dubey *et al*. (2015) observed high heritability as well as high genetic advance as percent of mean for harvest index, number of seeds per plant, biological yield per plant, number of pods per plant, and seed yield per plant in soybean genotypes.

Baraskar *et al*. (2014) revealed high heritability coupled with high genetic advance in soybean for plant height, number of clusters per plant, number of primary branches per plant, seed yield per plant, biological yield per plant and number of pods per plant.

Kumar *et al.* (2013) observed that high heritability for biological yield per plant and seed yield per plant. Moderate values of genetic advance were observed for plant height followed by days to maturity.

Osekita and Ajayi. (2013) reported that heritability was highest in five characters *viz.,* days to maturity, days to 50% flowering, seed yield, seed dry weight and 100 seed weight.

Zinaw *et al.* (2013) studied forty-eight soybean genotypes for heritability and genetic advance for nine important traits. High broad sense heritability and genetic advance was found for plant height whereas, high heritability and moderate genetic advance for days to 50 % flowering.

Aditya *et al*. (2011) observed the highest heritability for three characters namely 100 seed weight, number of primary branches per plant and days to 50% flowering. On the other hand, high heritability coupled with high genetic advance (GA) noted for number of pods per plant and dry matter weight per plant.

2.3Correlation and path study for seed yield and its components:

Saharia and Sarma (2022) carried out investigation with 38 soybean genotypes. The analysis was done using 10 quantitative and 11 qualitative characters reported that correlation and path coefficient analyses identified plant height, the branches per plant and the pods per plant as important traits for yield improvement in soybean.

Berhanu *et al*. (2021) reported that seed yield in soybean genotypes had highly significant positive genotypic and phenotypic correlation with primary number of branches/plant, number of pods/plant, number of seeds/pod and plant height, indicating that simultaneous improvement of grain yields with the associated traits is favourable. Plant height exerted the highest genotypic and phenotypic direct effect on seed yield, followed by 100 seeds weight and number

of pods/plant. This suggested that attention should be given for these traits mainly for direct and indirect selection for variety development.

Kuswantoro *et al*. (2021) reported that except for the correlations between seed yield and days until maturity, plant height, number of branches, and number of productive nodes, all other phenotypic associations were statistically significant. All of the reported agronomic parameters did not have a genotypic association with the seed yield. As a result, either directly using the seed yield parameter or indirectly using the 100-seed weight, selection can be done to improve seed production.

Amogne *et al*. (2020) studied 81 genotypes of soybean during 2018/2019 cropping season to assess their association of traits. Plant height and number of pods were shown to be positively and significantly correlated with grain yield. Number of pods/plant and number of nodules/plant showed a positive and substantial direct effect at the genotypic level. Number of pods/plant had a high phenotypic and favourable direct effect on grain yield, while number of branches had a negative direct effect. In order to boost grain output through direct selection and as parental material for future breeding projects, genotypes with features like number of pods and number of nodules per plant with high and positive relationship along with high direct effects should be evaluated.

Geetanjali (2020) concluded that correlation study showed number of pods per plant, number of seeds per pod, number of seeds per plant, 100-seed weight, oil content and protein content showed significant and positive correlation with seed yield at genotypic level. Whereas, some characters like days to 50% flowering and number of primary branches per plant showed negative but significant correlation with seed yield at genotypic level.

Kumar *et al*. (2020) found highly significant differences for characters under his study showing a lot of variation. At both phenotypic and genotypic stages, number of branches per plant, number of seeds per pod, biological yield per plant and harvest index showed significant positive correlation with seed yield per plant.

Li *et al*. (2020) did correlation studies and found that the yield showed very significant positive correlations with plant height, nodes on main stem, branches, pods, grains, 100-grain weight, and growth periods.

Parihar *et al*. (2020) revealed that in path analysis seed yield is directly affected by days to 50% flowering, 100 seed weight, through a low magnitude of direct effect, while plant height, pod bearing length and harvest index contributes through a moderate magnitude of direct effect on seed yield. Filled pods and biological yield contribute to seed yield by the highest magnitude of direct effect.

Pawar *et al*. (2020) conducted study on 30 different soybean genotypes for correlation and path and found that dependent variable (seed yield) showed positive and significant correlation with traits namely number of pods per plant, number of seeds per plant, hundred seed weight, plant height, days to 50% flowering, oil content and number of primary branches per plant at genotypic level. According to path coefficient analysis based on genotypic correlation, number of pods per plant registered highest positive direct effect on seed yield per plant followed by the weight of 100 seeds, number of seeds per pod, height of the plant, and days to 50% flowering. The highest negative direct effect was observed for days to maturity. The number of primary branches per plant has a direct effect with low magnitude.

Hang Vu *et al*. (2019) concluded highly positive direct effects on grain yield for the total number of pods, total number of seeds and 100 seed weight. Whereas 100 seed weight noted negative effect with plant height.

Mishra (2019) carried out association analysis to know the performance behavior in parent, F1 and F2 populations and their components in soybean.The traits days to 50% flowering had positive correlation with days to maturity. Number of seeds per plant had positive significant correlation with seed yield and number of pods per plant at the phenotypic level. Also studies on path coefficient analysis revealed existence of positive direct effect of 100-seed weight, number seeds per plant and number of pods per plant on seed yield. Though, 100-seed weight had substantial positive direct effect, but indirect effect for all the characters were found to be negative.

Painkra *et al*. (2018) studied that the seed yield per plant (g) had a highly significant and positive relationship with the number of pods per plant, number of seeds per pod, number of seeds per plant and 100 seed weight, Seed yield had a negative relationship with the number of pod bearing nodes, as well as days to maturity, protein content (%) and oil content (%).

Shree *et al*. (2018) studied correlation analysis on ninety genotypes of soybean and characters like number of seeds per pod and 100 seed weight found highly significantly association with yield as a dependent character.

Akkamahadevi and Basavaraja (2017) reported significant positive genotypic and phenotypic correlations between green pod production and number of branches/plant, number of pods/plant, hundred seed weight and oil content. Plant height, branch number, pod number, hundred seed weight, protein and oil content all had a positive direct impact on green pod production in the path

analysis. In order to increase the output of green soybean pods, consideration should be given to plant height, number of branches per plant, number of pods per plant, and hundred seed weight during selection.

Baig *et al*. (2017) have recorded strong and important correlation between seed yield and plant branch numbers, plant weight, seed weight, harvest, seed yield per plant and plant growth which indicated that the yield rise was primarily due to the increase in one or more of these characteristics.

Balla and Ibrahim (2017) identified that grain yield was found to have highly significant positive genotypic associations with days to 50% flowering, days to maturity, plant height, number of pods per plant. They also showed that the highest positive direct effect on grain yield was due to fodder yield, plant height and maturity days.

Chandel *et al.* (2017) showed that biological yield per plant followed by harvest index and days to 50 percent flowering had the highest positive direct effect on the seed yield. The characters like pods per plant, pods per cluster and primary branches per plant had seen direct negative effect on seed yield.

Akram *et al*. (2016) studied that number of seeds per plant had the highest positive direct effect on seed yield per plant followed by number of pods per plant, days to first flowering, number of branches per plant, 100-seed weight, plant height and pod length. On the other hand, number of seeds per pod, days to 50% flowering and days to maturity had negative direct effects on seed yield as a dependent character.

Chavan *et al*. (2016) studied 30 soybean genotypes and found that seed yield per plant as a dependent character showed positive correlation as well as

highly significant association with 100 seed weight, followed by number of pods per plant while branches per plant and seeds per pod showed non-significant and negative correlation with seed yield. They also concluded that 100 seed weight, number of pods per plant, days to 50% flowering and days to maturity had high positive direct effect on seed yield per plant as a dependent character.

Deshmukh (2016) concluded that traits like 100-seed weight, number of pods per plant, days to 50% flowering and days to maturity had recorded high positive direct effect on seed yield whereas number of pods per plant had noted the highest positive and significant direct effect on seed yield.

Ekka and Lal (2016) recorded positive phenotypic correlation between grain yield per plant and pods per plant, seed index, plant cluster, plant height and pod weight in soybean genotypes.

Mahbub and Shirazy (2016) studied 28 soybean genotypes and the traits like plant height, pod length, number of seeds per pod, number of pods per plant, hundred seed weight, branches per plant and number of seeds per pod recorded significant and positive correlation with the dependent character.

Alpna *et al*. (2015) reported that the number of pods per plant, 100 seed weight, harvest index and dry matter weight per plant had major contribution in determining seed yield per plant in soybean. And these characters show significant correlation with seed yield as a dependent character.

Chandrawat *et al*. (2015) showed that harvest index had maximum positive direct influence followed by days 50 percent flowering, days to maturity and number of pods per plant on the quantity and quality of the seed per plant in soybean.

Dubey *et al*. (2015) carried out experiment on fifty genotypes of soybean and observed significant positive correlation between the traits like biological yield per plant, number of seeds per plant, number of pods per node, number of nodes per plant and harvest index with seed yield per plant.

Mahbub *et al.* (2015) reported significant positive genotypic and phenotypic correlation with seed yield and plant, plant height, pod length, number of plant seeds per pod, number of pods per plant, 100 seed weight, branches per plant and number of seeds per pod. They also found that the character number of seeds per pod shows the highest positive direct effect on dependent character and other character like 100 seed weight, days to maturity and plant height showed positive direct effect on dependent character. Whereas the trait number of pods per plant noted the maximum negative direct effects on seed yield.

Mishra *et al.* (2015) studied the maximum positive direct effect for seed yield per plant was shown by number of seeds per plant followed by 100 seed weight and number of pods per plant while negative direct effect on seed yield per plant was recorded for plant height.

Silva *et al*. (2015) identified the relative contribution and correlation in the final yield of six genotypes of soybean using path analysis of yield components. A significant positive correlation was found for the number of pods per plant with the productivity; however the correlation was highly negative for the 100 seed weight.

Yahaya and Ankrumah (2015) studied the associations of some growth characters and grain yield in Soybean. The findings of this study revealed a positive correlation between number of branches and grain yield. According to the path analysis, number of branches has the biggest direct impact on grain yield.

Based on these findings, improvements in the number of branches per plant and the number of leaves can be made as criterion for choosing soybeans for higher grain production.

Badkul *et al*. (2014) studied that the seed yield showed positive and highly significant association with number of seeds per pod, biological yield, number of seeds per plant, plant height, number pods per plant and harvest index. It was also found that the independent variable like biological yield per plant, number of pods per plant, number of seeds per plant and harvest index showed considerable phenotypic and genotypic direct effects on seed yield as a dependable variable.

Malek *et al*. (2014) stated that the character number of seeds per plant showed highest positive and significant direct effects on yield per plant followed by 100-seed weight, number of pods per plant and days to maturity. The characters like day to flowering, plant height and number of branches per plant exhibited negative direct effects on yield per plant in soybean.

Abady *et al*. (2013) reported that days to maturity, harvest index and number of pods per plant had positive direct effect on seed yield per plot. However, plant height and number of branches per plant exerted positive indirect effect on seed yield per plot through number of pods per plant.

Ghodrati (2013) found significant correlation between crop yield and plant height and concluded that the simultaneous selection to improve seed yield by growing plant nodes, plant number and plant height would be a successful approach to increasing seed yield and protein yield.

Valencia-Ramírez and Ligarreto-Moreno (2012) determined the relationships between soybean seed yield and agronomic characteristics as well

direct and indirect effects of various yield components on seed yield through the analysis of path coefficients. Data from all locations combined showed a substantial positive association between seed yield and number of pods per plant, number of nodes per plant, number of seeds per pod, and the weight of the seed. Coefficient analysis showed number of pods per plant had the most direct beneficial impact on seed yield, followed by number of nodes per plant. However, there was a strong correlation between seed yield and the number of pods with three seeds. As indirect selection criteria for genetically improving soybean seed production, more focus should be placed on these yield components (number of pods per plant, number of nudes per plant, and number of pods with three seeds).

Aditya *et al*. (2011) examined 31 soybean genotypes for genetic parameters and correlations for eight quantitative variables, including grain yield. Dry matter weight/plant, number of primary branches/plant, number of pods/plant and harvest index all showed highly significant and favourable genetic correlations with grain yield/plant.

Datt *et al*. (2011) reported a strong positive and significant association of phenotypes and genotypes with flowering days, plant height, primary branches/plant and seed yield.

Machikowa and Laosuwan (2011) used a randomised complete block design with three replications for 14 soybean lines/varieties at the Suranaree University of Technology, Thailand. Eight characters were examined. Days till flowering, branches per plant, nodes per plant, pods per plant, seeds per plant, weighing 100 seeds, days until flowering to maturity, and yield were all taken into consideration. Between seed yield and days to flowering, a positive phenotypic correlation was found. Furthermore, genotypic correlation revealed that all parameters except for the weight of 100 seeds were positively connected with seed

yield. Pods per plant and branches per plant had the greatest direct positive effects on seed yield.

Iqbal *et al*. (2010) stated that seed yield was positive and strongly associated with all the characteristics analyzed except plant height. The oil content displayed a strong and favorable association between seed production and 100-seed weight while harvest index had a significant negative correlation between days of maturity, plant height and number of plant branches.

2.4 Genetic divergence:

Upadhyay *et al*. (2022) used genetic diversity of 50 exotic lines of soybean (including two checks, JS 20-98 and JS 20-34) on seed yield during the 2019 Kharif season. Soybean genotypes were grouped into five clusters. The maximum percentage of contribution towards genetic divergence was shown by the number of seeds per plant and the minimum contribution was shown by the number of primary branches per plant. Cluster I showed a maximum intra-cluster D^2 value of 231.14 while a maximum inter-cluster distance was observed between cluster IV and cluster II.

Nag and Sarawgi (2021) grouped hundred genotypes into six clusters. The cluster VI showed the maximum intra cluster distance. The maximum inter-cluster distance was recorded between clusters III and VI, followed by cluster I and VI and clusters I and V.

Shilpashree *et al*. (2021) divided 28 genotypes of soybean into eight clusters using Trocher's method described by Rao (1952).They found that in yield attributing traits, the genotypes GM-6 and GM-27 (cluster VIII) were agronomically superior. As a result, these genotypes could be employed for

commercial production as well as genetic improvement. Additionally, they could be used in a variety of parental crosses to aid in the development of even more diverse lines.

Sharma and Lal (2020) studied genetic divergence for eleven quantitative traits of 40 soybean genotypes using Mahalanobis's D^2 statistics. The genotypes were divided into nine groups based on the relative size of the D^2 values. With 19 genotypes, cluster III had the most genotypes overall, followed by cluster I. Cluster III and Cluster II had the greatest intra-cluster distance, respectively. Cluster III and cluster V had the greatest inter-cluster distance (D^2) and the greatest contribution to genetic difference came from test weight. The genotypes TNAU 20051, MAUS 128 and KB 17 with test weight attributes should be given top priority for the selection of the genetically divergent parents in the upcoming breeding programme.

Singh *et al*. (2020) worked on sixteen different soybean genotypes. Four groups of sixteen soybean genotypes were created. The genotypes fell into Cluster I, which had the maximum days to maturity value. The genotypes in cluster II exhibited the highest grain yield and plant height values.

Getnet (2019) categorized 49 soybean genotypes into three separate clusters in their study of genetic divergence, showing that the genotypes were moderately diverse.

Mishra (2019) divided thirty-three soybean genotypes into six clusters based on degree of divergence studies, with cluster II having the maximum number of genotypes in group (eleven) followed by clusters I, IV, III, V and VI, which had eight, six, five, two and one genotypes respectively.

Kumar *et al*. (2018) studied genetic diversity among 31 soybean genotypes grown in randomised block design with three replications. The genotypes could be divided into 10 clusters suggested that the tested genotypes had enough variation. With thirteen genotypes, cluster II was the largest. The distance between clusters III and IX was found to be the greatest, followed by the distances between clusters II and III and clusters V and IX. The maximum number of pods per plant was seen in Cluster IV, which indicated that the genotype belonging to this cluster may be chosen directly and employed in a hybridization procedure. Maximum genetic divergence was influenced by days to 75% maturity, then by days to 50% blooming and 100-seed weight. Based on cluster means, the genotypes belonging to clusters V and VII can be employed as a source population for early flowering and improved production.

Painkra *et al*. (2018) reported genetic diversity in 273 soybean germplasm and cluster analysis was performed for quantitative traits. The genotypes were divided into seven clusters based on cluster analysis, with Cluster IV being the largest with 67 genotypes and Cluster VI being the smallest with only five genotypes. Cluster VI (5.42), which contains 67 genotypes, had the highest intra-cluster distance. Cluster II had the smallest intra-cluster distance (2. 49). Cluster VI and Cluster III had the highest inter-cluster distance values (8.10). The distance between clusters ranged from 2.49 to 5.41. Cluster IV and cluster III had the lowest inter-cluster D^2 value (2.47).

Mahesh *et al.* (2017) classified 40 soybean genotypes into six groups. With the thirteen genotypes Cluster III was the largest cluster. Cluster I and Cluster VI had the greatest inter cluster distance, followed by V and VI respectively.

Nag *et al*. (2017) studied 100 different soybean accessions which were divided into four clusters. Cluster I had most genotypes (43), followed by Cluster

III (33 genotypes), Cluster II (18 genotypes) and Cluster IV (only six genotypes). The pattern of group constellation suggested that there was a lot of variation. The distances between and within four clusters were calculated. Cluster distances ranged from 15.11 (cluster II) to 17.67 (cluster III). Cluster I and Cluster II had the highest inter-cluster distance, while Cluster II and Cluster III had the smallest inter-cluster distance.

Marconato *et al*. (2016) evaluated the genetic diversity among 93 soybean accessions from different continents. The generated dendrogram revealed similarities between 11 national genotypes and eight subgroups, indicating the genetic diversity among the accessions.

Pushpendra *et al*. (2016) grouped the genotypes in ten clusters, where Cluster I represented the maximum number of genotypes (34 genotypes) followed by Cluster II (7 genotypes) and Cluster III (6 genotypes). Clusters VII, VIII, IX and X each contained only one genotype. There highest genetic diversity was between clusters IX and X.

Ghiday and Sentayehu (2015) evaluated the diversity of yield and yield-related variables among 49 soybean genotypes. The divergent genotypes have been divided into two groups using D-square statistics and five clusters by the cluster analysis. The genotypes for 13 traits were examined, and they revealed moderate variability for the components under investigation.

Thakur *et al*. (2015) studied that 40 genotypes were grouped into six clusters. Clusters III, VI, IV and I, respectively had 12, 8, 5and 3 genotypes. Clusters II and V each had six genotypes. Cluster VI had the highest intra-cluster distance, followed by III, II, IV, I and V while I and IV had the highest inter-cluster distance, followed by I and VI.

Barh *et al*. (2014) reported soybean genotypes in 12 clusters. In Cluster I, five genotypes were included, while thirteen genotypes were included in Cluster II. There were four genotypes in Cluster III, nineteen in Cluster IV, while each was mono genotypic in Cluster V, VI, VIII, IX, X, XI and XII, but two genotypes were included in Cluster VII. The average intra-cluster value varied from 0.00 to 10.004.

Kachhadia *et al*. (2014) created 11 separate clusters using D^2 statistic developed by Mahalanobis, from 61 genotypes of soybean. The largest inter-cluster distance, was found between clusters II and IX, followed by clusters II and XI, II and VIII, X and XI, IV and IX, and IV and XI, which showed that these genotype groups were very different from one another. The genotypes in the aforementioned clusters showed a significant difference in the means for key yield-contributing traits, indicating that the genotypes from clusters II, IX, XI and VIII should be chosen as parents in a hybridization programme for soybean improvement. Plant height, oil content and number of clusters per plant were the traits with the greatest genetic divergence.

Shinde *et al.* (2013) conducted genetic diversity studies for 41 genotypes of soybean collected from different geographical areas. These genotypes were grouped into seven clusters. Cluster II, I, V, VI, and III comprised 17, 10, 7,3 and 2 genotypes, respectively. The clusters IV and VII were mono-genotypic indicating wide divergence from other clusters. The highest inter-cluster distance was observed between clusters II and VII followed by IV and VII suggesting the use of genotypes from these clusters to serve as potential parents for hybridization. The characters iron content(70.12%) contributed maximum towards divergence followed by plant height (11.72%),days to physiological maturity (7.07%) and days to 50% flowering (5.49%).

Athoni and Basavaraja (2012) conducted an experiment with 84 soybean genotypes which, most of the cultivars were released in India, along with some indigenous and exotic lines. There was not much amount of diversity obtained in the material, representing diverse eco-geographical regions of the country hence, revealed no relationship between geographic diversity and genetic diversity.

Sharma *et al.* (2012) conducted genetic diversity among 35 soybean genotypes for yield attributing traits by using D^2 analysis. The 35 genotypes were grouped into 5 clusters and clustering pattern revealed that genetic diversity may not necessarily be related to geographical diversity. The average inter-cluster distance was maximum between cluster IV and V (35.04) followed by cluster I and IV (29.32), cluster II and V (24.92) and cluster II and IV (24.85) indicating the presence of greater diversity between genotypes belonging to these groups. Days to 50 per cent flowering, plant height, days to maturity and pod breadth together contributed for 87.88 per cent of total divergence. Based on inter-cluster distance values and per se performance, the cross combination between Bragg and Gaurav, JS 80-21 and Bragg, TS 148 and Gaurav, KB 230 and Gaurav, and MAUS 144 and Bragg are expected to give better heterosis and desirable recombinants in order to achieve better yield levels in soybean under agro climatic condition of Manipur.

Tyagi *et al*. (2012) studied genetic divergence through the D^2 statistics on 16 characters of 40 soybean genotypes and these were grouped into six clusters. Cluster III noted biggest cluster which were 12 number of genotypes after that cluster V, I, VI, II and IV respectively.

Anuradha *et al.* (2011) studied genetic divergence for yield and different yield contributing traits in 282 black soybean accessions and grouped into 9clusters. The first four principle component axes (PCA) accounted for 70.3% of

total variance. Minimum mean value for days to flowering (45.5), plant height (82.8 cm) and daysto maturity (121.3) were obtained in cluster I, indicating that this cluster could be useful to develop early maturing genotypes. Cluster IX contained eight accessions that showed the maximum mean value for pod length (4.3 cm), 100 seed weight (14.4 g) and seed yield per plant (8.1 gm). From a yield point of view, this cluster can be used to develop high yielding as well as high grain weight genotypes. The maximum inter-cluster distance was to be between cluster IV and IX (6.4). Hence genotypes from these clusters could be used in hybridization to obtain desirable recombinants. Accessions VBS 25, VBS 48 from cluster VII and VBS161, VBS 152 from cluster VIII, were found to be exceptional donors who could be used in multiple crossing programmes to get transgressive segregants for desirable traits.

Dhapke *et al.* (2011) studied genetic divergence in high yielding 66 elite soybean genotypes selected from 401 germplasm accessions. These genotypes fell into 10 clusters out of which one was monogenotypic. Among the seven characters, number of pods per plant contributed maximum to the genetic divergence, followed by number of branches per plant. This indicated that these characters were mainly responsible for genetic divergence. The highest divergence was observed between clusters I and VIII followed by clusters I and VII, clusters III and VII and clusters III and VIII which may serve as potential parents for hybridization programme. The potential combination based on the D^2 statistics was found to be AMS-MB-5-19 x AMS-248, AMS-MB-5-19 x H5P23, AMS-MB-5-28 x H5P23, AMSMB-5-28 x AMS-248, AMS-MB-5-19 x H6P5, H6P5 x IC-118482 and AMS-MB-5-19 xIC-118429. These combinations may result in maximum hybrid vigor and highest number of useful segregants.

Patil *et al*. (2011) experimented with 36 soybean genotypes and recorded that there was a large genetic diversity across D^2-value genotypes varying from 33.64 to 379.08. Genotypes were grouped into 6 clusters. The trend of clustering showed that genetic variation in this crop was not inherently synonymous with geographical diversity.

2.5 Application of SSR markers to evaluate the genetic diversity in soybean germplasm:

Kumar *et al*. (2022) carried out an investigation on genetic divergence between 29 soybean cultivars using 35 SSR primers. There were 14 polymorphic primer pairs among them, resulting in a total of 34 polymorphic alleles; the number of alleles per locus ranged from two to four, with 2.43 alleles per primer pair on average. The study discovered eight unique and two rare alleles that might be used for the analysis and identification of genetic purity and cultivar.

Saharia and Nath Sarma (2022) carried out investigation with 38 soybean genotypes during *Kharif*, 2020. The variability analysis was done using 10 quantitative and 11 qualitative characters. Twenty four SSR markers were used to study genetic relationships among the genotypes based on Jaccard's coefficient of similarity out of which 19 were found to be polymorphic. The number of SSR allele per locus ranged from one to three with an average of 1.4 alleles per locus. DNA marker analysis revealed a range of diversity in the experimental materials with few potential markers for diversity analysis due to their high PIC values.

Jo *et al.* (2021) analysed the variability of 470 soybeans accessions with black seed coats and green cotyledons in Korean germplasm using 6K single nucleotide polymorphic loci and provided 36 accessions having 99.5% of the genetic diversity.

Karikari *et al*. (2020) used 68 trait-linked SSR markers to observe variability from three continents *viz*., Africa, America and Asia. Phylo-genetic analysis effectively separated genotypes from Africa from those of the other two continents, demonstrating that geographic difference plays an important role in genetic variability. The findings revealed that soybean germplasm had migrated from Asia to America and then to Africa.

Mukuze *et al*. (2020) in a study on genetic diversity of soybean genotypes in Uganda analysed about 21 polymorphic SSR markers and reported the existence of 59 alleles with the frequency of average 2.85 average alleles per locus. A total of 21 SSR markers showed significant connection with days to flowering and 100-seed weight based on association mapping.

Kujane *et al*. (2019) observed the genetic diversity and polymorphism among 30 soybean genotypes using 20 SSR markers and detected total of 216 alleles with 10.8 alleles per locus on average. It was reported that B 66 S 31, 69S 7, and R5-4-2 M had the most diverse genotypes, demonstrating the effectiveness of SSR markers in detecting genetic variety.

Moniruzzaman *et al*. (2019) observed salt tolerance ability and measured genetic diversity and relatedness in five soybean genotypes, GC840, Asset, Binasoybean-1, Binasoybean-3, and Binasoybean-5, grown in hydroponic culture under control and varying salt stressed conditions using 10 SSR markers. According to the findings of morphological and genetic studies, GC840 and Binasoybean-3 are fairly salt tolerant.

Tiwari *et al*. (2019) studied and analysed the genetic diversity and population structure of 148 Indian soybeans (*Glycine max* (L.) Merrill) genotypes

using 26 SSR markers showed distinctive polymorphism among 148 lines with an average of 2.8 alleles per SSR locus.

Hipparagi *et al*. (2017) used twenty one SSR markers for the genetic analysis of 75 genotypes collected from several areas in Uttarakhand. A total of 60 alleles were amplified, with 2.85 alleles per locus on average. According to the genetic diversity indices used, Kala-bhat genotypes can be a suitable source for the soybean breeding programme since Kala-bhat genotypes were more diverse than brown seed coat and yellow seed coat colour genotypes

Bisen *et al.* (2015) used sixteen polymorphic SSR markers to detect the genetic diversity and varietal identification of 38 soybean genotypes. A total of 51 alleles were found, with an average of 2.22 alleles per locus. Twelve of the 38 soybean genotypes were effectively identified using these 16 SSR markers. These findings imply that SSR markers can be used to measure genetic diversity and relatedness as well as identify soybean varieties.

Kumawat *et al*. (2015) used 44 SSR markers for the determination of genetic diversity and molecular characterisation of 82 soybean accessions. Forty of the 44 SSR markers tested were polymorphic. These 40 polymorphic markers resulted in 119 alleles, five of which were unique and four of which were rare and revealed the substantial genetic similarity among Indian soybean germplasm collection.

Wang *et al*. (2015) used the SSR marker for evaluation of genetic diversity and population genetic structure in *Glycine soja* (wild soybean) and observed a high degree of variability in the population collected from Dongying, China.

Dong *et al.* (2013) studied a total of 100 vegetable soybeans using 53 SSR markers that were evenly distributed in the well-established soybean linkage groups. All markers provided unambiguous bands and gave a total of 296 alleles across all accessions, with an average of 5.6 alleles per locus. The most polymorphic marker satt_005 and the least polymorphic marker satt_588 amplified 11 and 2 alleles, respectively.

Hosamani *et al.* (2013) studied 33 genotypes of soybean varying in storability (good and poor) and seed coat colour (black and yellow) were characterized with 53 SSR and 51 RAPD markers. Polymorphisms detected by SSR and RAPD markers were 62.26 and 68.62%, respectively. Genotypes with black seed coat colour showed better storability (89.85%) than the yellow seed coated genotypes (71.15%). Genetic similarity coefficients obtained through SSR data analysis grouped the genotypes into two major clusters representing black and yellow seeded genotypes. SSR markers SatG371, SatG453and SatG618 produced specific allelic bands making them candidate markers for linkage with seed storability and testa colour.

Rani e*t* a*l.* (2013) studied simple sequencer to lipoxygenase-1 gene in soybean. Parental polymorphism was surveyed using SSR markers Sat-074 andSatG522 reported to be linked with Lox2 locus and the SSR markers in its proximity. F2:3 seeds were used for assaying lipoxygenase-1 to identify the genotype of the F2 individuals. SSR marker SatG656 was found to be tightly linked with Lox1locus at distance of 3.6 and 4.8 cm in the mapping population of LSb1 × PI408251 and JS335 × PI408251, respectively. SSR marker SatG656can be useful for marker assisted selection for transferring recessive allele oflipoxygenase-1 in the background of high yielding soybean genotypes.

Velusamy *et al.* (2013) analyzed 78 wild soybean accessions collected from different Korean provinces using 9 SSR markers. The number of alleles investigated ranged from 6 to 11, with mean value of 9.11 per locus and 82 alleles were detected. The most variable locus was found to be Satt 155and Satt 203 with 11 alleles per locus. SatG423 showed least variability with six alleles. The mean genetic diversity and PIC value was 0.824 and 0.804. The wild soybeans from different regions were included in the same groups by cluster analysis. The high genetic diversity observed in this study suggested that South Korea might be the major center for genetic diversity of wild soybean.

Zhang *et al*. (2013) used "Expressed Sequence Tag-derived Simple Sequence Repeats (EST-SSRs)" to explore the genetic diversity among more than 45 vegetable soybean accessions and indicated that these new markers could be useful in the field of molecular breeding, taxonomy, and relative mapping of soybean.

Sayama *et al.* (2011) showed that amongst commonly applied molecular markers, simple sequence repeats (SSRs, or microsatellites) possess advantages such as a high level of polymorphism and co dominant pattern of inheritance at individual loci. To facilitate systematic and rapid genetic mapping in soybean, they designed a genotyping panel comprised of 304 SSR markers selected for allelic diversity and chromosomal location so as to provide wide coverage. Those 80 loci showed an average allele number and polymorphic information content value of 14.8 and 0.78 respectively. High level of polymorphism, ease of analysis, and high accuracy of the SSR genotyping panel rendered it widely applicable to soybean genetics and breeding.

Tantasawat *et al.* (2011) found 53 alleles with an average of 4.82 alleles per locus among 25 soybean genotypes consisting of 15 certified varieties, 8breeding

lines and 2 plant introductions in Thailand involving 11SSR markers. The polymorphic information content (PIC) among genotypes varied from 0.13 (Soy satt 285) to 0.88 (Soy satt 173) with an average of 0.60.

Guan *et al.* (2010) investigated the genetic relationship between 205 Chinese soybean accessions and 39 Japanese soybean accessions from various regions using 46 SSR loci. Cluster analysis with UPGMA separated the Chinese accessions from Japanese accessions, suggested that soybean in these two countries were form different gene pools and accessions from China have more genetic diversity than those from Japan. This study provided interesting insights into further utilization of Japanese soybean in Chinese soybean breeding.

Min *et al*. (2010) used 40 SSR primer pairs and established disparities in genetic diversity levels across 40 soybean accessions of wild soybeans, cultivars, and land races collected at the Shanxi Agricultural University. When three varieties of soybeans were compared, wild soybeans and landraces had more genetic diversity and allelic diversity than cultivars.

Mulato *et al*. (2010) observed the genetic variation in 79 soybean accessions from around the world, cluster them based on the similarity of accessions, and check if the two types of markers *i.e*. SSR and EST-SSR employed are correlated. 20 genomic and 10 EST-SSR primer pairs were chosen based on their distribution throughout the 20 soybean genetic linkage groups, their tri-nucleotide repetition unit, and their polymorphism information richness. The observed genetic variety was considerable, allowing the creation of five distinct groupings and subgroups. There was a moderate link between genetic divergence and geographic origin of the accessions.

Shi *et al*. (2010) used 65 SSR markers dispersed on 20 soybean chromosomes, genetic diversity and association analyses were done among 105 food-grade soybean genotypes. The 105 soybean genotypes were classified into four clusters with six sub-groups based on SSR marker data. Protein and oil content were found to have a negative correlation. The resulted data will help breeders to choose parents for crossover, apply marker-assisted selection in food-grade soybean breeding, and map QTL for soybean protein and oil content.

Mimura and Coyne (2007) utilised SSRs for the examination of the genetic diversity of 130 accessions of soybean, including edamame cultivars and landraces from China, Japan, and the United States, as well as novel breeding lines in the US. Despite the fact that superior edamame cultivars are thought to have limited genetic variation, 17 SSRs were able to identify 99 out of the 130 accessions. According to cluster analysis, the patterns of SSR variability in edamame may generally separate maturity classes and testa colour. They concluded that Japanese edamame have a distinct and narrow genetic base from others and those SSRs can be used to define genetic diversity patterns in the elite vegetable soybean.

CHAPTER III

MATERIAL AND METHODS

MATERIALS AND METHODS

The present investigation entitled "Genetic studies of Soybean [*Glycine max* (L.) Merrill] genotypes under Nagaland conditions" was carried out to know the genetic variability of soybean genotypes based on various morphological characters and molecular diversity. This chapter includes all the materials used and methods employed during the course of investigation. All the techniques used are detailed under respective headings and their original references quoted.

The experiment was carried out during *kharif* season for two consecutive years 2017 and 2018 at the experimental farm of School of Agricultural Sciences (SAS), Medziphema Campus, Nagaland, which is geographically located at 25.75^{o}N latitude and 93.86^{o}E longitude at an altitude of 360.0 meters above mean sea level. The experimental field had fairly levelled topography and good drainage system.

The experimental locations fall under subtropical climate with high humidity, moderate temperatures, and moderate to high rainfall. It is neither a hill nor a valley and it gently slopes down towards the southern region from the north-eastern side of the town. This town actually represents an interface of the hilly Nagaland and the valleys as the actual hill region. The temperature during crop season ranged between 12.3-32°C in 2017 and from 11 to 33.6°C in 2018 whereas total rainfall varied from 1574.8 mm (2017) to 1140.5 mm (2018). The meteorological data during the experiment period regarding distribution of rainfall maximum and minimum temperature and relative humidity was obtained from ICAR regional centre Jharnapani, is shown in table 3.1 and depicted graphically in Figure 3.1 and 3.2.

Table 3.1: Meteorological data during the crop investigation period from June-December (2017 and 2018)

Month	Temperature (°C)				Relative humidity (%)				Total Rainfall (mm)	
	2017		2018		2017		2018		2017	2018
	Max.	Min.	Max.	Min.	Max.	Min.	Max.	Min.		
June	31.6	23.8	33.4	24.2	93	72	94	73	278.7	354.7
July	31.4	24.4	33.2	24.9	94	75	92	72	485.6	240
August	32.0	24.7	33.5	24.9	93	72	94	71	492.5	302.8
September	31.6	24.7	33.6	23.9	95	74	94	67	235.9	115.7
October	30.7	23.5	29.9	20.1	95	72	96	67	33.9	64
November	28.1	16.3	28.2	14.1	96	63	97	54	16.4	13.3
December	25.5	12.3	24.6	11.0	96	66	96	56	31.8	50

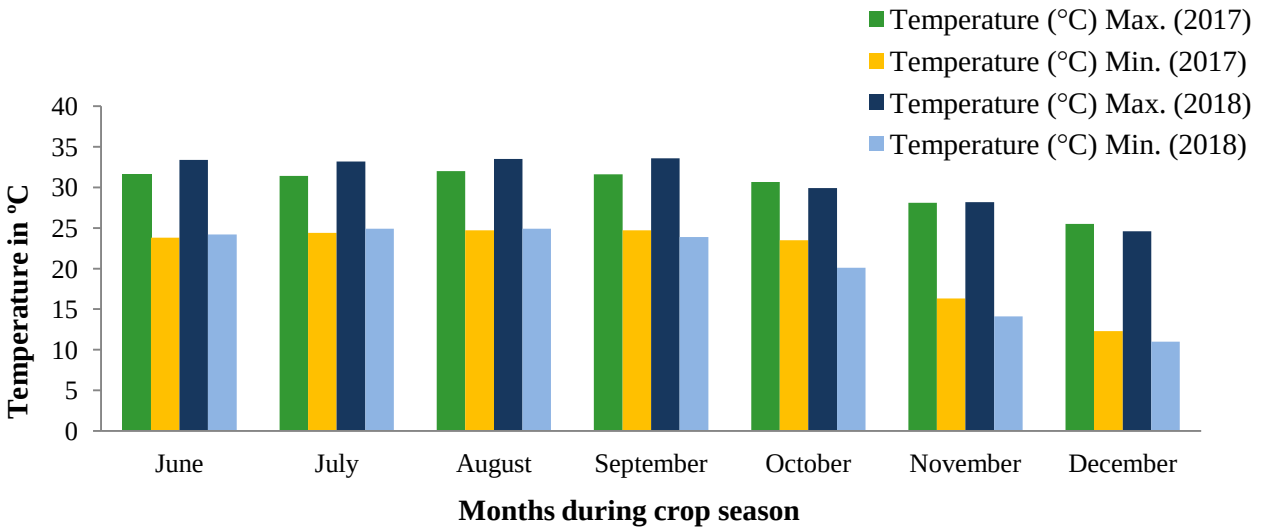


Figure 3.1: Temperatures (minimum and maximum) during crop growing season-2017 & 2018

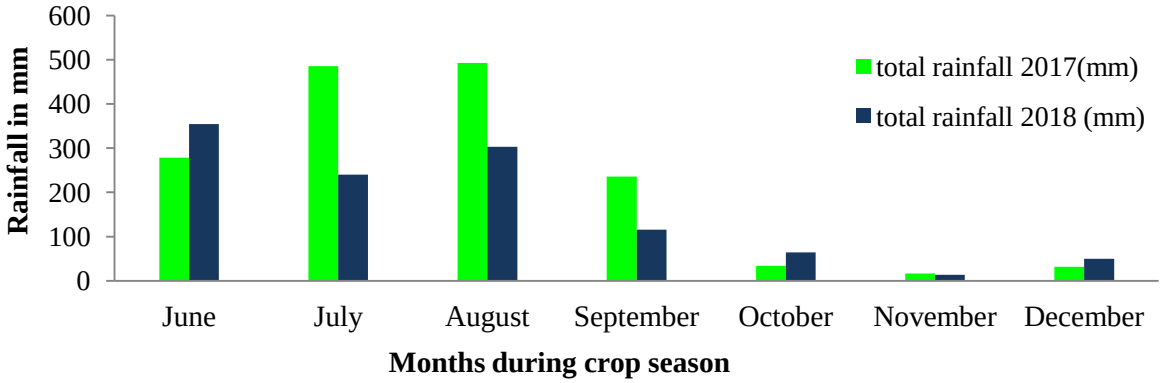


Figure 3.2: Rainfall recorded over crop growing season-2017 & 2018

3.1. Experimental details:

3.1.1 Particulars of the experiment and its layout

The experiment was set up in a Randomized Block Design (RBD) format in both the seasons. The layout plan details are given as below:-

Table 3.2: Layout plan details

Season	:	*Kharif* (rain fed) - 2017 and 2018
Date of sowing	:	14 July (2017 & 2018)
Design	:	Randomized Block Design
Replication	:	03
Genotypes	:	Twenty
No. of checks	:	01 (G19: JS-9752)
Total number of plots	:	60
Plot size	:	1.5 x 1.0 sq. m
Spacing	:	50 x 10 cm
Number of rows in each plot	:	4

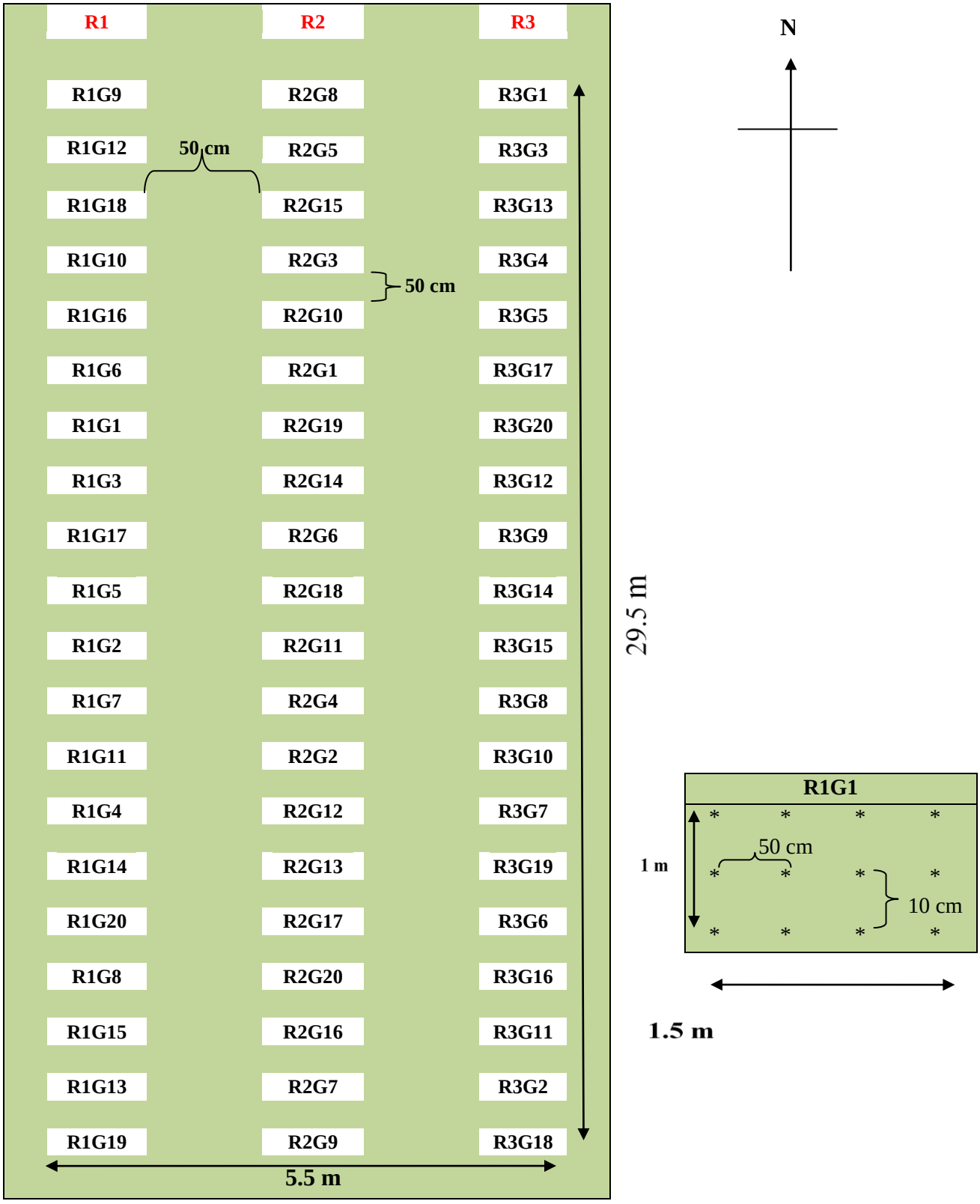
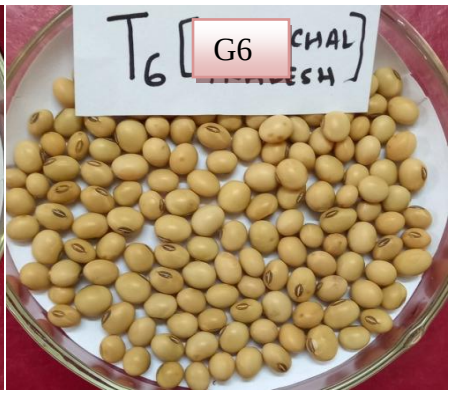


Figure 3.3: Layout of the Experiment in Randomized Block Design

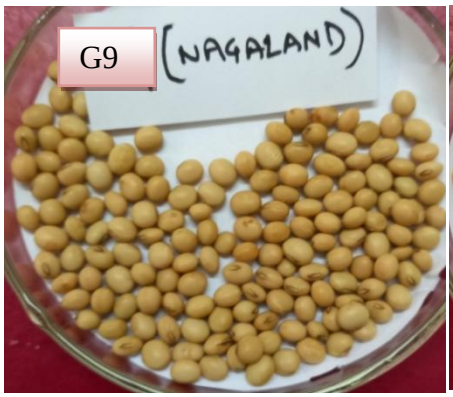
3.1.2Genotypes under study and their source

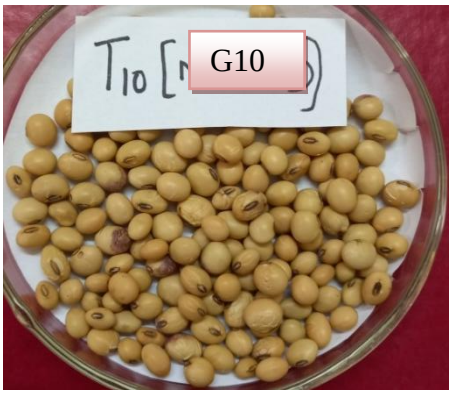
In the present study, the experimental material included twenty genotypes of soybean that were collected from various states (Table 3.3).




G7
[MIZORAM]


G8


G9
(NAGALAND)


G10


G11
(NAGALAND)


G12


G13
JS-9305


G14


G15
NAGALAND


G16


G17
NAGALAND


G18

Plate 1: Genotypes under study

Table 3.3: Details of the soybean genotypes and their place of collection with source

S. No.	Genotypes	Place of collection			Source
		State	District	Institute or town	
1.	**G1**	Assam	Cachar	Jirighat	Farmer
2.	**G2 (small)**	Manipur	Imphal	CAU	CAU
3.	**G3 (large)**	Manipur	Imphal	CAU	CAU
4.	**G4**	Arunachal Pradesh	Lower Dibang valley	Jiali	Farmer
5.	**G5**	Mizoram	Lawngtlai	Lawngtlai	Farmer
6.	**G6**	Arunachal Pradesh	Lower Siang	Basar	Farmer
7.	**G7**	Mizoram	Serchip	East Lungdar	Farmer

8.	**G8**	Mizoram	Lunglei	Hnahthial	Farmer
9.	**G9**	Nagaland	Dimapur	Dimapur	Farmer
10.	**G10**	Nagaland	Kohima	Chiephobozu	Farmer
11.	**G11**	Nagaland	Kiphere	Mimi	Farmer
12.	**G12**	Nagaland	Tuensang	Tuensang	Farmer
13.	**G13 (JS-9305)**	MP	Jabalpur	JNKVV	JNKVV
14.	**G14**	Nagaland	Peren	Samjuiram	Farmer
15.	**G15**	Nagaland	Tuensang	Tuensang	Farmer
16.	**G16 (JS-9560)**	MP	Jabalpur	JNKVV	JNKVV
17.	**G17**	Nagaland	Wokha	Yimkha	Farmer
18.	**G18**	Nagaland	Zunhebuto	Mishilimi	Farmer
19.	**G19 (JS-9752)** **Check variety**	MP	Betul	Sirkhed	Farmer
20.	**G20 (JS-335)**	MP	Jabalpur	Jabalpur	JNKVV

3.1.3Cultural practices

Field preparation

During the both the seasons, the ploughing was done in the month of March, followed by two harrowing using a tractor-drawn disc harrow. Plots for solarisation were preserved.

Fertilizer application

The nature of the soil was acidic, with a pH ranging from 4.5 to 6.2. Farm yard manure was applied @ 5t/ha. The crop was provided with 25:100:50:50 ($N:P_2O_5:K_2O$:S kg/ha,). The entire P, K and S and 50% of total nitrogen was applied as basal, while remaining 50% of nitrogen was applied as top dressing at pod formation stage (Dupare and Billore, 2016).

Sowing

Prior to sowing seeds were treated with Carbendezim @ 2g/kg of seed 24 hours before sowing to protect it from soil borne diseases. Sowing of seeds was done manually. The field was laid out in raised-bed and furrow system. This has an advantage over flat sowing system in draining off excess water. At the time of sowing, the seeds were placed at 5-6 cm depth in the soil and covered.

Intercultural operations and Harvesting

Weed control was prioritized to reduce weed interference with crops. Two hand weeding at 20-25 days after sowing and 40-45 days after sowing were done. In the months that followed, there was very little weeding and only mild hoeing.

The crop was harvested when all the leaves became yellow and started dropping and the stalk stood only with pods. Harvesting was done by cutting the

plant manually with the help of sickle. The harvested crop was dried for few days and when the crop was fully dried, was thrashed with a stick. The soybean grain was dried properly before storage to ensure that the moisture content of seeds does not exceed 10%. Seeds were stored in dry bins or polythene bags in airtight condition. Irrigation was not done since crop was grown under rain fed conditions.

3.2Observations recorded:

For data collection, five random plants were selected from each plot and replication for the following fourteen characters to be studied.

3.2.1 Days to 50% flowering

It was recorded as the number of days from date of sowing to the date when 50 percent of plants flowered on a plot basis.

3.2.2 Days to maturity

On a plot basis, the number of days from sowing to physiological maturity of the crop was recorded as days to maturity.

3.2.3 Plant height (cm)

Plant height was measured in centimetre (cm) at maturity, from the base to the tip of the plant.

3.2.4 Number of primary branches per plant

The first order of branches emerged from main shoot were counted and considered as number of primary branches. The number of primary branches arising from the main shoot were counted in five randomly selected plants in a plot and averaged.

3.2.5 Number of clusters per plant

The number of clusters (more than two pods) per plant was counted at maturity and averaged.

3.2.6 Number of pods per cluster

Total numbers of the pods per cluster were counted at maturity.

3.2.7 Number of pods per plant

The total number of filled pods per plant was recorded at the time of harvest and average was taken.

3.2.8 Number of seeds per pod

The total number of seeds per pod (all pods in a plant) counted at maturity and averaged.

3.2.9 Pod length (cm)

In each of the five observational plants, the length of five randomly selected pods was measured in centimetres and an average was determined.

3.2.10 Hundred Seed weight (g)

A total of one hundred seeds were selected from the five randomly selected plants for each genotype, weighed in grams and an average was determined.

3.2.11 Oil percentage

The seeds of selected genotypes from each plot and replication were collected after harvest and oil was extracted through Soxhlet extraction method

using acetone. For oil calculation, 0.5 g of the sample was taken and put in an extraction thimble. After that, the sample was put in a pre-weighed extraction flask (A). The solvent was poured into the beaker making volume of 80 ml and after loading the beakers in the system the process is incubated at 80^{o}C for 60 minutes followed by increase in temperature to 120^{0}C for 30 minutes then rinsing is done about 2-3 times in order to collect the remaining fat present in the sample. The residual solvent was extracted by heating the flask in an oven at 80°C kept overnight. The flask was weighed (B) after cooling in desiccators and the oil content was estimated as follows. The percent oil content in the sample was calculated as:

$$\textbf{Oil content (\%)} = \frac{\textbf{Weight of flask (B)} - \textbf{Weight of flask (A)}}{\textbf{Weight of sample (gm)}} \textbf{ X 100}$$

3.2.12 Biological yield per plant (g)

Plants were harvested when they were physiologically mature; sun dried and weighed. The average weight of the selected plants after drying was recorded (including root weight). The crop biological yield refers to the total dry matter accumulation of a plant system which included both shoot and root dry matter.

3.2.13 Seed yield per plant (g)

The average yield per plant was calculated after the seeds from five randomly selected plants were dried and weighed. This weight was recorded as seed yield per plant.

3.2.14 Harvest index

The following formula was used to calculate the harvest index (HI)

$$\text{HI (\%)} = \frac{\text{Grain yield per plant (g)}}{\text{Biological yield per plant (g)}} \text{X } 100$$

3.3Statistical analysis:

The experimental data collected on fourteen characters were compiled by taking the mean values over selected plants for each replication. It was then analyzed for various statistical parameters as follows:

3.3.1 Analysis of variance

The genotypic differences between the entries were examined before moving on to the biometrical genetic analysis of the data. Further analysis was done, only when the mean squares attributable to genotypes were significant. As a result, the data for distinct characters were statistically examined for significance using pooled analysis of variance and coefficients of variance computed according to formulae given by Lush (1940) and Chaudhary and Prasad (1968). The chosen design was a three-fold replication of the Randomized Block Design (RBD). Analysis of variance was done under the fixed effective model given below:

To test the hypothesis

H_0 : G1 = G2 ……= $G2_0$, the fixed effect model for the analysis of variance in RBD is as follows:

$$\mathbf{Y_{ij} = \mu + g_i + r_j + e_{ij}}$$

Where,

Y_{ij} = phenotypic observation in the i^{th} treatment and j^{th} replication

μ = Overall mean

g_i = effect of i^{th} treatment

r_j = effect of j^{th} replication

e_{ij} = Random error associated with i^{th} treatment and j^{th} replication

i = No. of treatments

j = No. of replications

Table 3.4:Pooled analysis of variance (ANOVA) for RBD

Sources of variation	Degree of freedom (d.f)	Sum of Square (SS)	Mean square (MS)	Variance ratio
Year (Y)	(Y-1)	Y SS	Y MS	Y MS/EMS
Replication within year	Y (R-1)	R SS	R MS	R MS/EMS
Treatment (T)	(T-1)	T SS	T MS	T MS/EMS
Year x genotype	(Y-1)(T-1)	Y SS x T SS	Y x T MS	Y x T MS/EMS
Pooled error	Y(R-1)(T-1)	E SS	EMS	-
Total	(YRT-1)			

Where,

Y = No. of years (season)

R = No. of replications

T = No. of treatments

Y SS = sum of square due to year

R SS = sum of squares due to replications within year

T SS = sum of squares due to genotypes

E SS = sum of squares due to pooled error

TSS = Treatment sum of squares

Y MS = Mean sum of square due to year

R MS = Mean sum of square due to replication within year

T MS = Mean sum of squares due to treatments

EMS = Error mean sum of squares

Critical difference

Critical difference was calculated by following formula:

$$\mathbf{CD} = \sqrt{\frac{2EMS}{r}} \times t - value$$

t-value = table value of error d.f at 5% level of significance

Where,

r = number of replications

EMS = error mean sum of squares

Significant "F" value indicates that, there is significant difference among the treatments. But, to compare the difference between any two particular treatments, it is tested against CD value.

3.3.2Variability parameters

(i) Genotypic variance

The genotypic variance (σ_g^2) is the variance due to the genotypes present in the population. The formula used for calculation of genotypic variance was as follows:

$$\text{Genotypic variance } \left(\sigma_g^2\right) = \frac{MS_g - EMS}{r}$$

(ii) Environmental or Error variance

Environmental variance (σ_e^2) is the variance due to environmental deviation.

$$\sigma_e^2 = EMS$$

(iii) Phenotypic variance

Phenotypic variance (σ_p^2)denotes the total variance present in a Population for particular character and is calculated by following formula:

$$\sigma_p^2 = \sigma_g^2 + \sigma_e^2$$

Where,

σ_g^2= Genotypic variance

σ_e^2= Error variance

3.3.3 Coefficient of variation

It is the measure of variability observed. Coefficient of variation is the ratio of standard deviation of a sample to its mean and expressed in percentage.

$$CV\ (\%) = \frac{\text{Standard deviation}}{\text{Mean}} \text{ X } 100$$

In the present investigation, three types of coefficients of variation were estimated *viz*., phenotypic coefficient of variation (PCV), genotypic coefficient of variation (GCV) and error/environmental coefficient of variation (ECV). The

formulae used to calculate PCV, GCV and ECV were given by Burton and Devane (1953):

Phenotypic coefficient of variation (P.C.V):

$$\mathbf{PCV} = \frac{\sqrt{\boldsymbol{\sigma_p^2}}}{\bar{\mathbf{X}}} \mathbf{X\ 100}$$

Genotypic coefficient of variation (G.C.V):

$$\mathbf{GCV} = \frac{\sqrt{\boldsymbol{\sigma_g^2}}}{\bar{\mathbf{X}}} \mathbf{X\ 100}$$

Environmental coefficient of variation (E.C.V):

$$\mathbf{ECV} = \frac{\sqrt{\boldsymbol{\sigma_e^2}}}{\bar{\mathbf{X}}} \mathbf{X\ 100}$$

Where,

$\sqrt{\sigma_p^2}$ = Phenotypic standard deviation

$\sqrt{\sigma_g^2}$ = Genotypic standard deviation

$\sqrt{\sigma_e^2}$ = Error standard deviation

$\bar{X}$ = General mean of the character

σ_p^2= Phenotypic variance

σ_g^2= Genotypic variance

σ_e^2= Environmental variance

GCV and PCV values were categorized as low, moderate and high as indicated by Sivasubrananian and Menon (1973).It is as follows:

0-10% = Low
10-20% = Moderate
>20% = High

3.3.4 Heritability

Heritability is the ratio of genotypic variance to the total phenotypic variance. Broadly, it was estimated according to the formula given by Allard (1960).

$$h^2(\text{broad sense}) = \frac{\sigma_g^2}{\sigma_p^2} \text{ X } 100$$

Where,

h^2 = Heritability in broad sense

σ_g^2 = Genotypic variance

σ_p^2 = Phenotypic variance

Heritability values are ranked as low, moderate and high according to Robinson *et al*. (1949)

0-30% = Low
30-60% = Moderate
>60% = High

3.3.5 Genetic advance

Genetic advance is defined as an increase in the mean genotypic value of selected plants over the parental population. The estimates of genetic advance

were obtained by the formula given by Lush (1949), Johnson *et al.* (1955) and Allard (1960):

$$\mathbf{GA = k.\sigma_p.h^2}$$

Where,

GA = Expected genetic advance

k = Constant (Standard selection differential) having the value of 2.06 at 5% level of selection intensity

σ_p = Phenotypic standard deviation

h^2 = Heritability in broad sense

In order to visualize the relative utility of genetic advance among the characters, genetic advance as percent of mean was computed as follows:

$$\textbf{Genetic advance as percent of mean} = \frac{\textbf{Genetic advance}}{\textbf{Grand mean}} \textbf{ X 100}$$

The range of genetic advance is classified as suggested by Johonson *et al.* (1955):

< 10 % = low
10-20 % = moderate
> 20 % = high

3.3.6 Correlation coefficient

Correlation coefficient is the mutual association between variables without implying any cause and effect relationship. Simple correlation coefficients were computed at genotypic and phenotypic levels between pair of characters adopting following formula given by Al-Jibouri *etal.* (1958)as well as Panse and Sukhatme (1967).

Phenotypic correlation coefficients

Phenotypic correlation coefficient between character x and y

$$\mathbf{r_{xy}\,(p)} = \frac{\boldsymbol{\sigma^2_p\,(xy)}}{\sqrt{\boldsymbol{\sigma^2_p(x)\ .\ \sigma^2_p\,(y)}}}$$

Where,

rxy (p) = Phenotypic correlation between x and y

σ^2_p(xy) = Phenotypic covariance between traits x and y

σ^2_p (x) = Phenotypic variance for x

σ^2_p (y) = Phenotypic variance for y

Genotypic correlation coefficients

Genotypic correlation coefficient between character x and y

$$\mathbf{r_{xy}\,(g)} = \frac{\boldsymbol{\sigma^2_g(xy)}}{\sqrt{\boldsymbol{\sigma^2_g(x)\ X\ \sigma^2_g\,(y)}}}$$

Where,

r_{xy} (g) = Genotypic correlation between x and y

σ^2_g (x y) = Genotypic covariance between traits x and y

σ^2_g (x) = Genotypic variance for x

σ^2_g (y) = Genotypic variance for y

Test of significance

The calculated values were compared with the table value of the correlation coefficient recommended by Fisher and Yates (1938), at (n-2) treatment degree of

freedom at 5% and 1% level of significance in order to determine the significance of the correlation coefficient. It is considered to be significant if the calculated value of correlation coefficient is higher than the tabular value.

3.3.7 Path coefficient analysis

The use of path coefficient analysis explains cause and effect of relationship among the variables. It is a standardized partial regression coefficient and as such measures the direct influence of one variable upon another and permits the separation of the correlation coefficients into components of direct and indirect effects (Dewey and Lu 1959). This method permits breeder to identify relatively important components of a variable, on the basis of their direct and indirect influences.

The direct and indirect effects both at genotypic and phenotypic level were estimated with grain yield per plant as dependent variable using path coefficient analysis suggested by Wright (1921) and Dewey and Lu (1959). The following set of simultaneous equations were formed and solved for estimating various direct and indirect effects.

$$r_{1y} = P_{1y} r_{11} + P_{2y} r_{12} + P_{3y} r_{13} \ldots\ldots\ldots\ldots + P_{ny} r_{1n}$$

$$r_{2y} = P_{1y} r_{21} + P_{2y} r_{22} + P_{3y} r_{23} \ldots\ldots\ldots\ldots + P_{ny} r_{2n}$$

$$\begin{array}{ccccc} . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \end{array}$$

$$r_{ny} = P_{1y} r_{n1} + P_{2y} r_{n2} + P_{3y} r_{n3} \ldots\ldots\ldots\ldots + P_{ny} r_{nn}$$

Where,

1, 2 . . . . . n = Independent variable

y = Dependent variable (yield per plant)

r1y r2y . . . rny = Coefficient of correlation between causal factors‘1’ to ‘n’ on dependent character 1

P1y P2y . . . . Pny = Direct effect of characters ‘1’ to ‘n’ on character Y

The above equations can be written in matrix form as:

A C B

$$\begin{pmatrix} r_{1y} \\ r_{2y} \\ : \\ : \\ r_{ny} \end{pmatrix} \begin{pmatrix} 1 & r_{12} & r_{13} & \dots\dots r_{1n} \\ r_{21} & 1 & r_{23} & \dots\dots . r_{2n} \\ : & : & : & : \\ : & : & : & : \\ r_{n1} & r_{n2} & r_{n3} & \dots\dots 1 \end{pmatrix} \begin{pmatrix} \mathrm{P}_{1y} \\ \mathrm{P}_{24} \\ : \\ : \\ P_{ny} \end{pmatrix}$$

A and B vector values are known. Hence, to calculate C vector

$B = [C]^{-1} A$

Where,

$$C^{-1} = \begin{pmatrix} \mathrm{C}_{11} & \mathrm{C}_{12} & \mathrm{C}_{13} & \dots\dots \mathrm{C}_{1n} \\ \mathrm{C}_{21} & \mathrm{C}_{22} & \mathrm{C}_{23} & \dots\dots \mathrm{C}_{2n} \\ : & : & : & : \\ : & :: & : & : \\ \mathrm{C}_{n1} & \mathrm{C}_{n2} & \mathrm{C}_{n3} & \dots\dots \mathrm{C}_{nn} \end{pmatrix}$$

Direct effects were as follows:

$P_{1y} = \sum_{i=1}^{k} \mathrm{C}_{1i}\mathrm{r}_{iy}$

$P_{2y} = \sum_{i=1}^{k} \mathrm{C}_{2i}\mathrm{r}_{iy}$

$P_{ny} = \sum_{i=1}^{k} \mathrm{C}_{ni}\mathrm{r}_{iy}$

Residual Effect

In plant breeding, it is very difficult to have complete knowledge of all component traits of yield. The residual effect permits precise explanation about the pattern of interaction of other possible components of yield. In other words,

residual effects measure the role of other possible independent variables which were not included in the study on the dependent variable. The residual effect is estimated with the help of direct effects and simple correlation coefficients. It was calculated by using following formulae.

$$\mathbf{Pry} = \sqrt{\mathbf{1} - (\mathbf{p_{1y}r_{iy}} + \mathbf{p_{2y}r_{iy}} + \cdots + \mathbf{p_{ny}r_{ny}})}$$

Where,

p_{ny} = direct effect of X_n on Y

r_{iy} = correlation coefficient of X_n on Y

The direct and indirect effects are rated as follows by Lenka and Mishra (1973).

0.00-0.09 – Negligible
0.10-0.19 – Low
0.20-0.29 – Moderate
0.30-0.99 – High
>1.00 −Very high significant and vice-versa

3.3.8 Estimation of Genetic Divergence

Usually to assess the diversity in population of diverse origin, important method *i.e.* Mahalanobis D^2 Statistics is employed.

Mahalanobis' D^2 analysis

The data collected on different characters were analysed through Mahalanobis' D^2 analysis to determine the genetic divergence among the genotypes. D^2 value between i^{th} and j^{th} genotypes for 'P' characters was calculated as:

$$\mathbf{D_{ij}^2} = \sum_{\mathbf{t=1}}^{\mathbf{P}} \left(\mathbf{Y_i^t} - \mathbf{Y_j^t}\right)^{\mathbf{2}}$$

Where,

Y_i^t = Uncorrelated mean value of i^{th} genotype for 't' characters

Y_j^t = Uncorrelated mean value of j^{th} genotype for 't' characters

D^2_{ij} = D^2 value between i^{th} and j^{th} genotypes.

The various steps involved in estimation of D^2 values are given below:

i) Test of significance

Variances were calculated for all the characters investigated and test of significance was done. Analysis of covariance for the character pairs was estimated on the basis of mean values (Panse and Sukhatme 1967). After testing difference between genotypes for each of the characters, a simultaneous test of significance for the differences in the mean values of a number of correlated variables with regard to the pooled effect of characters was carried out using V statistic, which in turn utilizes Wilk's criterion. The sum of squares and sum of products of error and error + variety, variance and covariance matrix were used for this purpose. The estimation of Wilk's criterion was done using following relationship.

$$\text{‹}\Lambda\text{›} = \frac{|\boldsymbol{E}|}{|\boldsymbol{E+V}|}$$

Where,

Λ = Wilk's criterion

$|E|$ = Determinant of error matrix and

$|E + V|$ = Determinant of error + variety matrix

The significance of '^' was tested by:

V (Stat) = -m log e ^ = - [n - (P + Q + 1) / 2] log e '^'

Where, m = n-(P + Q +1) / 2

P = Number of variables or characters i.e.14

Q = Number of varieties -1 (or d.f. for populations) i.e., 20-1 = 19

n = degree of freedom for error + varieties

Log e '^' = 2.3026 log 10 '^'

V (Stat) is distributed as χ^2 with PQ degrees of freedom i.e., (14 x 19) = 266 in the present study.

ii) Transformation of correlated variables

In the present model, computation of D^2 values were reduced to simple summation of the differences in mean values of various characters of the two genotypes *i.e.* Σd^2_i. Therefore, transformation of correlated variables into uncorrelated ones was done before working out the D^2 values. Transformation was done using pivotal condensation method.

iii) Computation of D^2 values

For the given combination of i and j genotype, the mean deviation *i.e.* Yit – Yjt, where t = 1, 2…p variables are computed and the D^2 values were calculated.

iv) Testing the significance of D^2 values

The D^2 value obtained for a pair of population is taken as calculated value of χ^2 and is tested against the tabulated value of χ^2 for P degree of freedom where P is the number of characters considered. In the present study P is 14.

v) Contribution of individual characters towards divergence

In all combinations each character was ranked based on their contribution towards divergence between two entries ($d_i = Y_i^t - Y_j^t$). Rank 1 is given to the highest mean difference and rank P to the lowest difference, where, P is the total number of characters. Percentage contribution towards genetic divergence was calculated using the following formula.

$$\textbf{Percentage contribution of a character X} = \frac{N}{M} \textbf{ x 100}$$

Where,

X = Percent contribution of character

N = Number of genotype combinations where the character was ranked first.

M = All possible combinations of number of genotypes considered in the present study.

vi) Grouping of genotypes into various clusters

The grouping of genotypes into different clusters was done using the Tocher's method as described by Rao (1952). The criterion was that the two varieties belonging to the same cluster should at least on an average show a smaller D^2 value than those belonging to different clusters. For this purpose D^2 values of all combinations of each genotype were arranged in ascending order of magnitude in tabular form as described by Singh and Chaudhary (1977). To start with, two populations having the smallest D^2 value from the first two populations was added. Similarly, the next nearest fourth population was considered and this procedure was continued. At certain stage when it was felt that after adding a particular population there was an increase in the average D^2, that population was not considered for including in that cluster. The genotypes of the first cluster were

then eliminated and the rest were treated in a similar way. This procedure was continued until all the genotypes were included into one or other clusters.

vii) Average intra-cluster distance

The average intra cluster distances were calculated by formula given by Singh and Chaudhary (1977).

$$\textbf{Square of the intra cluster distances} = \frac{\sum D_i^2}{n}$$

Where,

$\Sigma D_i^{2=}$ sum of distances between all possible combinations (n) of the populations included in a cluster.

n = number of possible combinations

viii) Average inter-cluster distance

Clusters were taken one by one and the distances from other clusters were calculated. The distance between two clusters was the sum of D^2 values between the genotypes of one cluster to each of the genotypes of the other cluster divided by the product of number of genotypes in both the clusters under consideration. The square root of the average D^2 value gave the genetic distance between the clusters. Based on D^2 values (inter cluster distance) the scale given by Rao (1952) for rating of the disease was adopted and the cluster diagram was prepared.

$$\textbf{Average-inter cluster distance} = \frac{\sum D_i^2}{(n_1 X n_2)}$$

Where,

D^2i = Sum distances between all possible combinations (n1, n2) of the entries included in the cluster study.

n_1 and n_2 = number of genotypes of two clusters.

Category 'D' Value

Closely related : Below 22
Moderately divergent : Between 22 and 30
Highly divergent : Above 30

ix) Cluster Diagram

The clusters and their mutual relationship were presented diagrammatically. The square root of average D^2, which is an approximate measure of divergence between groups, had been used to denote the distance.

3.3.9 Molecular Characterisation by SSR Marker

A total of twenty five different SSR markers from all twenty linkage groups of soybean genome were chosen at Indian Institute of Soybean Research, Indore and further characterization of genotypes was performed there itself. A brief description of SSR primers used in the present investigation is given in Table 3.5 below-

Table 3.5: SSR primers used in the study

Sl. no	Linkage group	SSR Name	Chromosome No.	Forward primer Sequence (5'→ 3')	Reverse primer Sequence (5'→ 3')
1	A1	Satt 155	5	AGATCCAACACCTGGCCTAAT	GCTGCACAATTCATTCCATTT
2	A2	Sat_409	8	CCTTAGACCATGAATGTCTCGAAGATA	CTTAAGGACACGTGGAAGATGACTAC
3	B1	Satt 484	11	GCGTTTAATAAAACTAATTTAATTGTACT	GCGTTCCCTTTCTCTCCTTTCTTTCTT
4	B2	Satt 126	14	GCTTGGTAGCTGTAGGAA	ATAAAACAAATTCGCTGATAT
5	B2	Satt 687	14	ACCGCAACTCACTCACCTT	GCGCCCAATTAACAGAAAC
6	C1	Satt 164	4	CACCAATGGCTAAAGGTACATAT	AGGAGAAGAAAAAATCACATAAAATATC
7	C1	Satt 396	4	GCGAAAAGGGATAAGTTTAAAAAT	GCGGGCCTGTAAAGGGATTCC
8	C2	Satt 557	6	GCGGGATCCACCATGTAATATGTG	GCGCACTAACCCTTTATTGAA
9	D1a	Satt077	1	GATCTAAAGTCTGATATTTTTAACTA	AAAAGGAGAAGGAATGC
10	D1b	Sat_227	2	GCGCAAAATGATTTGGGAAAATAACTTACA	GCGTTATATACTTTTGGCGAGTTATCC
11	D2	Satt 310	17	GCGAGTTTTTATCTCATGACTTTT	GCGGGGGTATGGGACCTAAAGAAAC
12	E	Satt 230	15	CCGTCACCGTTAATAAAATAGCAT	CTCCCCCAAATTTAACCTTAAAGA
13	E	Satt 411	15	TGGCCATGTCAAACCATAACAACA	GCGTTGAAGCCGCCTACAAATATAAT
14	F	Satt 362	13	GCGTTGTTGTTTCAAATGTATTTTAGTT	GCGGACGGATCATCAAACCAATCAAGAC
15	G	Satt 163	18	AATAGCACGAGAAAAGGAGAGA	GTGTATGTGAAGGGGAAAAACTA

16	H	Sat_218	12	GCGCACGTTAAATGAACTGGTATGATA	GCGGGCCAAAGAGGAAGATTGTAAT
17	H	Satt 666	12	TGGCTTGTCATCTCTACTTTTATTAG	TCATGCATCTAATTTGTTTTATCTATCA
18	I	Satt 270	20	TGTGATGCCCCTTTTCT	GCGCAGTGCATGGTTTTCTCA
19	J	Sat_393	16	GCGGTCCTGCATGTTAATGTTGATT	GCGGGTCCCTACAATGTGAGTGG
20	K	Satt 055	9	AGTTAAGGAAGAATTTATTGTTAT	AACATTTTATTTGAGTATTTAGAAT
21	K	Satt 588	9	GCTGCATATCCACTCTCATTGACT	GAGCCAAAACCAAAGTGAAGAAC
22	L	Sat_286	19	GCGTTGCTTGCTAAGTAGTGTTTTTAATCCT	GCGTCTCCCATCATGCAACTTCAATA
23	M	Sat_316	7	GCGCAACGTCTAAAGCACAAGGATT	GCGCGACTACGTTACAGTTCCAA
24	N	Satt 022	3	GGGGGATCTGATTGTATTTTACCT	CGGGTTTCAAAAAACCATCCTTAC
25	O	Sat_196	10	GCGAAACGAGATACTAGGATTTTGACTT	GCGAGCCTTAGGAGTAGTTAATGATGA

i) Genomic DNA Isolation

Total genomic DNA was isolated from young leaves of 20 genotypes of soybean following CTAB (Cetyl trimethyl ammonium bromide) extraction method as given by Murray and Thompson (1980) and modified by Saghai-Maroof *et al.* (1984) and Xu *et al.* (1994).

Reagents:

1) Extraction Buffer: (For 100 ml)
   i. 100 mM Tris-HCl (pH 7.5) : 10 ml of 1M stock
   ii. 25 mM EDTA : 5 ml of 0.5M stock
   iii. 1.5 M NaCl : 30 ml of 5M stock
   iv. 2% (w/v) CTAB : 20 ml of 10% stock
   v. Polyvinylpyrrolidone (PVP) (1%) : 1 gm
   vi. 0.3% (v/v) β-mercaptoethanol : 300 ul-added immediately before use
2) RNAse A stock solution (10 mg/mL)
3) Chloroform: Isoamyl alcohol 24:1
4) Isopropanol
5) 70% ethanol (v/v)
6) TE Buffer: 10 mM Tris-HCl (pH 7.6), 0.1 mM EDTA

ii) Procedure

Leaf samples were taken from 3-4 week old plants. Approximately, 5 g of the leaf tissue was hand homogenized to fine powder in liquid nitrogen using sterilized pre- chilled mortar and pestle. Then 100 mg of powder was transferred

to polypropylene tube and was mixed with 5ml extraction buffer into it. The samples were thoroughly mixed with the extraction buffer by gently inverting the tubes several times and were incubated in water bath at 65^{o}C for 50 minutes. After incubation, the samples were cooled to room temperature followed by adding equal volume of Chloroform: Isoamyl alcohol (24:1) and mixed gently.

Samples were then centrifuged for 10 minutes at 10,000 rpm in centrifuge (REMI C-24). After centrifugation, the aqueous phase was transferred to a pre-sterilized centrifuge tube followed by addition of 0.66 volumes (330µl) of chilled Isopropanol & precipitate at -20^{o}C for 30 minutes. Again the sample was centrifuged at 10,000 rpm at 4^{o}C for 10 min, and then the supernatant was discarded carefully keeping DNA pellet. Further the DNA pellet was washed with 70% ethanol (500 µl) at 10,000 rpm for 10 minutes then the DNA pellet was dried at room temperature and dissolve in 50 µl TE buffer. After that RNase was added (30µl) and incubated at 37^{0}C for 30 minutes and followed by checking the quality and quantity of DNA by electrophoresis in 0.8% agarose gel. Finally the DNA was stored at 4^{o}C for further use.

iii) Qualitative and Quantitative Estimation of DNA

Quality and quantity of DNA was estimated by UV spectroscopy and agarose gel electrophoresis. For UV spectroscopy, an aliquot of DNA samples was suitably diluted and absorbance (A) was determined at 260 nm and 280 nm wavelength in spectrophotometer. Using the relationship of O.D. unit of 1.0 at 260 nm equivalent to 50 g DNA per ml, the quantity of DNA was estimated from the following formula:

Concentration of DNA (g/ml) = A260 x 50 x dilution factor

The DNA concentrations were also checked by visual assessment of band intensity in comparison with Lambda DNA of known concentration in 0.8% agarose gel.

The quality of DNA samples was checked both by UV-spectrophotometer and on agarose gel electrophoresis in comparison with a λ marker of concentration of 50 ng/ µl. A total of 1µl of each DNA sample as well as marker was loaded in the gel. Using spectrophotometer, the ratio of the absorbance at 260 nm and 280 nm was noted. Samples with a ratio of 1.8 to 2.0 were considered of good quality.

A260/A280 = 1.8 (pure DNA)

Quality of DNA preparation was also tested in comparison with λ DNA standards of known concentrations on ethidium bromide stained gels by submerged horizontal agarose (0.8%) gel electrophoresis.

iv) Polymerase Chain Reaction (PCR) Optimization and Amplification

PCR amplification was carried out in programmable thermal cycler from Biorad-G100* thermal cycler.

PCR Reaction mix

The reaction volume used during the experiment and their reagents are given below in table 3.6.

Table 3.6: Concentration of the reagents which were used in PCR

S.no	Reagent	Reaction volume used for PCR
1.	*Taq* buffer	1µl
2.	dNTPs	1 µl
3.	$MgCl_2$	1.0µl
4.	*Taq* DNA polymerase	0.1µl (Bio-rad)
5.	Primers	0.5 µl (each)
6.	DNA	0.5 µl
7.	Sterile water	According to reaction volume to make up final volume 10µl

The following protocol was used for thermal profiling for PCR

Table 3.7: Thermo-cycler profiling for amplification of SSR markers

Step	Temperature	Duration	cycle	Activity
1.	95 ºC	3 min.	1	Initial Denaturation
2.	94º C	1 min.	35	Denaturation
	55^0C	1 min.		Annealing
	72º C	1 min.		Extension
3.	72 ºC	7 min.	1	Final extension
4.	4 ºC	∞		Storage

v) Electrophoretic separation of amplified product

Agarose Gel Electrophoresis Reagents used to perform the analysis are given below:

Composition of TAE buffer (1000ml):	
1) Tris base	: 242.03 gm
Glacial acetic acid	: 57.1ml
EDTA(0.5M)	: 100ml
Final volume	: 1000 m l
2) 6X Loading dye Sucrose	: 4.0 g
3) Bromophenol blue	: 0.025 g
4) Xylene cyanol	: 0.025 g
Volume	: 10 ml

Procedure:

First of all, stoppers were fitted to the edges of the casting tray and appropriate combs were placed for the well formation. Then the metaphor Agarose gel (0.8%) which was prepared by dissolving 0.8 gm agarose in 100 ml 1x TAE buffer and 3.5% metaphor was boiled in a microwave oven with intermittent stirring and after complete dissolving of agarose, gel is allowed to cool at 50-55°C, then 2 µl of Goodview stain is added and mixed properly. After proper mixing of stain, the solution was poured into the casting tray and allowed to solidify. Upon solidification, gel was kept into electrophoresis chamber having 1x TAE buffer and then combs were removed from the gel. Then 1µl 6x loading dye was mixed with 5µl of each sample DNA and samples were loaded in the gel wells. Separation of DNA is done for 1 hour at 100 V in 1x TBE running buffer. Finally Agarose gel was kept in the gel documentation system, exposed with UV light and photograph was taken for interpretation of results on UV trans-illuminator (SynGene, UK). Polymorphism was assessed by visual examination.

The band profiles of each gel were scored visually. Each amplified product was considered as a DNA allele locus and was scored across all samples. These bands were transferred into a binary matrix with '1' for the presence and '0' for the

absence of a band at a particular position. Bands within genotypes were scored as missing data if they resolved poorly or if template did not amplify well. Only bright and distinguishable bands were used for the genetic analysis. Molecular weight of the bands was estimated by using gene ruler 100 bp plus DNA ladder as standard.

vi) Data analysis

Similarity coefficient

The data set of cultivars and reproducible bands were used to calculate pair-wise similarity coefficient following Jaccard (1908). It represents frequency of presence and absence of the band in i^{th} and j^{th} genotypes

Dendrogram

The matrix of similarity coefficient was subjected to unweighted pair group method for arithmetic mean (UPGMA) to generate a dendrogram using average linkage procedure. The standard data matrix was used to calculate correlations using average among variables. The relation between genetic similarity identified by SSR markers and taxonomic distance measured by mean genetic distance was analyzed using Jaccard's similarity index. The computations were carried out using NTSYS-pc version 2.1 (Rohlf 2000).

Analysis of marker based polymorphism

The molecular genetic diversity based on recorded data on size of base pair of polymorphic markers was assessed using GeneAlEx version 6.5 (Peakall and Smouse 2012). The parameters such as- allele frequency, number of allele, gene

diversity, observed heterozygosity, polymorphism information content, were computed.

Polymorphism information content (PIC) values

The PIC values described by Botstein *et al.* (1980)were used to refer to the relative value of each marker with respect to the amount of polymorphism exhibited. PIC values for each primer were estimated using formula:

$$\mathbf{PIC} = \mathbf{1} - \sum \mathbf{P_i^2}$$

Where, Pi is the frequency of i^{th} allele in the set of genotypes analyzed, calculated for each SSR locus. PIC is synonymous with the term 'gene diversity' as described by Weir(1990). The PIC takes into account not only the number of alleles that are expressed but also the relative frequencies of those alleles (Smith *et al.* 1997).

CHAPTER IV

RESULTS AND DISCUSSION

RESULTS AND DISCUSSION

The purpose of the present work was to better understand how agronomic and qualitative traits of soybean varied, their diversity and their association between and among the parameters. The study entitled “Genetic studies of soybean [*Glycine max* (L.) Merrill] genotypes under Nagaland conditions” was carried out at School of Agricultural Sciences (SAS), Medziphema Campus, Nagaland, experimental farm (Genetics and Plant Breeding) during *kharif* season 2017 and 2018.

The data were collected on 14 different characters *viz*., days to 50 per cent flowering, days to maturity, plant height (cm), number of primary branches per plant, number of clusters per plant, number of pods per cluster, number of pods per plant, number of seeds per pod, pod length (cm), 100 seed weight (g), oil%, biological yield per plant (g), harvest index (%) and seed yield per plant (g). Each trait's data was examined independently. The outcomes are displayed under the following headings:

4.1 Analysis of variance

4.2 Mean performance of genotypes

4.3 Genotypic and phenotypic coefficient of variation (GCV and PCV)

4.4 Heritability and genetic advance

4.5 Correlation analysis

4.6 Path coefficient analysis

4.7 Genetic divergence

4.8 Evaluation of genetic diversity using SSR molecular markers

4.1 Analysis of variance

Variation describes apparent differences between individuals for specific characteristics. The analysis of variance for all the characters in the present investigation was carried out to partition the total variance due to genotypes and other sources. For all the traits listed in Table 4.1 including days to 50 per cent flowering, days to maturity, plant height (cm), number of primary branches per plant, number of clusters per plant, number of pods per cluster, number of pods per plant, number of seeds per pod, pod length, 100 seed weight (g), oil%, biological yield per plant (g), harvest index and seed yield per plant (g),analysis of variance revealed highly significant differences ($p \leq 0.01$) among genotypes.

These characteristics indicated that genotypes had inherent genotypic differences among themselves along with more opportunities for selection-based improvement. With the exception of days to maturity, number of seeds per pod, and oil percentage, the mean sum of squares due to genotype x year was significant for all the characters, indicating that genotypes and environments interacted significantly.

4.2 Mean performance of genotypes

Table 4.2 shows the mean, range, and coefficient of variation for the various characters found in the soybean population. Table 4.3 shows the average performance and degree of variability across twenty genotypes tested for fourteen characteristics. The majority of the traits among the genotypes showed a broad range of variation, according to the mean performance. The range of differences in mean value that were examined revealed significant levels of variability among the genotypes for these characters. The following is a description of how each character was performed:

Table 4.1: Analysis of variance data for traits in soybean genotypes that contribute to seed yield

Source of variation	d.f	Mean sum of squares for the characters under study												
		DFF	DM	PH	NPB/P	NC/P	NPo/C	NPo/P	NS/Po	PoL	HSW	Oil%	BY/P	HI	SY/P
Year	1	2.41	72.07	336.51	18.93	230.56	31.22	15262.59	0.02	0.08	6.96	3.25	14209.93	6858.70	197.39
Rep(Year)	4	41.13*	28.17	9.25	2.71*	0.52	0.45	48.69	0.02	0.02	0.05	1.21	65.98	13.02	4.80
Gen	19	1402.90**	2216.85**	1257.53**	3.74**	85.60**	3.52**	4701.34**	0.25**	0.92**	69.70**	40.37**	761.95**	159.00**	53.98**
Year * Gen	19	8.43**	7.74	88.96**	1.73**	11.83**	0.51*	802.82**	0.04	0.04**	2.59**	0.67	219.60**	113.35**	15.61**
Pooled Error	76	3.56	4.89	12.17	0.24	2.39	0.25	105.32	0.02	0.02	0.09	0.98	38.19	7.48	3.14

*** Significant at 5 % probability level and ** Significant at 1% probability level**

Abbreviations:-DFF: Days to 50% flowering, **DM:** Days to maturity, **PH**: Plant height (cm), **NPB/P**: Number of primary branches per plant, **NC/P**: Number of clusters per plant, **NPo/C**: Number of pods per cluster, **NPo/P**: Number of pods per plant, **NS/Po**: Number of seeds per pod, **PoL**: Pod length (cm), **HSW**: Hundred seed weight, **BY/P**: Biological yield per plant (g), **HI**: Harvest index and **SY/P**: Seed yield per plant (g).

I. Days to 50% flowering

This character varied from 32.83 to 79.50 days, with a mean of 56.34 ± 1.09 days. G2 (79.50 days) was the latest in flowering along withG1 (78.83) and G14 (77.83) at par, while G13 (32.83 days) was earliest in blooming upon 50%.Only genotypeG16 (34.50) was at par with G13. Most of the local genotypes exhibited medium to late flowering. A total of six genotypes specifically G3 (48.83), G5 (48.67), G9 (49.50), G10 (49.83), G11 (49.83) and G17 (49.67) were at par with check G19 (JS-9752) with medium flowering period of 49.50 days.

II. Days to maturity

The average number of days to maturity was 118.24 ± 1.28days, with mean values ranging from 89.33 (G16) to 148.00 (G2, G14).Among all 15 genotypes *viz*., G1 (137.33), G2 (148.00), G3 (107.17), G4 (136.83), G5 (114.50), G7 (113.67), G8 (143.33), G9 (106.00), G10 (108.00), G11 (109.17), G12 (133.50), G14 (148.00) G15 (133.00), G17 (113.33), and G18 (135.00) exhibited significantly late maturity as compared to the check G19 (100.00 days). G6 (98.50) and G20 (98.83) were at par.G13 (91.33) was significantly earlier in maturity and at par with G16.

III. Plant height (cm)

With a general mean value of 46.27 ± 2.01 cm, the average data for plant height (cm) for all the genotypes under trial ranged from 26.71 cm (G16) to 77.46 cm (G2).Most of the genotypes (*i.e.* fourteen) under investigation were significantly taller (*i.e*. >36.92 cm) in height than check G19-JS-9752 (32.91) whereas, only G3 (30.53), G13 (29.90) and G20 (31.60) were observed at par with check variety and two treatments were below the check range (<28.90). However, genotypes G3, G6 and G13 were at par to G16 while G2 was tallest.

IV. Number of primary branches per plant

The mean data ranged between 1.06 (G16) to 4.64 (G1) with average of 3.39 ± 0.28 branches per plant. GenotypesG8 (4.28) and G12 (4.11) were at par withG1. However six genotypes were at par (2.89 ± 0.57) with the check G19 (2.89) while twelve genotypes *viz*., G1 (4.64), G2 (3.64), G4 (3.53), G5 (3.69), G7 (3.92), G8 (4.28), G10 (3.61), G11 (3.89), G12 (4.11), G14 (3.50), G15 (3.97) and G18 (3.53) had more primary branches per plant.

V. Number of clusters per plant

The experimental data under study varied from 0.31 (G13) to 15.72 (G8) clusters per plant with an average mean value of 5.03 ± 0.89. Among all the genotypes, nine (G2, G6, G7, G9, G11, G14, G15, G17, and G20) were at par (3.72 ± 1.78) with check G19 (JS-9752, 3.72) while G1 (12.22), G4 (9.69), G5 (5.97), G8 (15.72), G10 (5.97) and G12 (6.11) had more clusters respectively. Also varieties G3, G16 (JS-9560) and G18 showed negligible clusters formation and were at par with G13 (JS-9305).

VI. Number of pods per cluster

The range of data for this character was 0.58 (G13) to 3.24 (G1),with average mean value of 2.34 ± 0.29. G3 (1.15) and G16 (0.82) were at par with G13 whereas; G4, G6, G8, G9, G10, G12 and G19 were at par to G1 having higher number of pods/cluster. However nearly 50% of the genotypes *viz*., G2, G3, G5, G7, G11, G13, G14, G16, G17, G18 and G20 had lesser number of pods in a cluster (i.e. < 2.54) when compared to check JS-9752 (3.12). The remaining eight genotypes (*viz*., G1, G4, G6, G8, G9, G10, G12 and G15) were at par (3.12 ± 0.58) to the check variety.

Table 4.2: Fourteen morphological traits- mean, range, and coefficient of variance

Sl.No	Character	Mean ± SEM	Range of variation		CV (%)
			Min	Max	
1	**Days to 50% flowering**	56.34 ± 1.09	32.83	79.50	3.35
2	**Days to maturity**	118.24 ± 1.28	89.33	148.00	1.87
3	**Plant height (cm)**	46.27 ± 2.01	26.71	77.46	7.54
4	**No. of pr branches per plant**	3.39 ± 0.28	1.06	4.64	14.53
5	**No. of clusters per plant**	5.03 ± 0.89	0.31	15.72	30.80
6	**No. of pods per cluster**	2.34 ± 0.29	0.58	3.24	21.55
7	**No. of pods per plant**	54.82 ± 5.93	10.92	115.00	18.72
8	**No. of seeds per pod**	2.16 ± 0.09	1.64	2.52	7.11
9	**Pod length (cm)**	3.30 ± 0.07	2.81	4.09	3.81
10	**Hundred seed weight (g)**	8.32 ± 0.18	2.82	13.47	3.66
11	**Oil %**	20.04 ± 0.57	14.17	22.76	4.93
12	**Biological yield per plant (g)**	27.28 ± 3.57	9.60	63.09	22.65
13	**Harvest index**	29.76 ± 1.58	23.17	38.90	9.19
14	**Seed yield per plant (g)**	7.33 ± 1.02	2.98	16.33	24.17

VII. Number of pods per plant

The mean data for number of pods per plant obtained was 54.82 ± 5.93. The range varied between 10.92 (G13) – 115.00 (G1). Ten genotypes *viz.*,G1 (115.00), G2 (73.64), G4 (68.47), G5 (57.00), G7 (64.08), G8 (111.44), G9 (55.22), G11 (60.33), G12 (69.47) and G14 (82.33) were found superior to check JS-9752 (38.47). Rest six genotypes were at par (38.47 ±11.80) .However G3 and JS-9560 (G16) were statistically inferior with mean value of 22.33 and 11.03 and with par to G13 respectively. Also G8 (111.44) was found to be at par with G1.

VIII. Number of seeds per pod

The mean experimental data of seeds per pod for genotypes under study ranged between 1.64 (G3) to 2.52 (G13) with an average mean of 2.16 ± 0.09. G10 (1.85) was at par with G3 while G14 (2.34) and G18 (2.34) were in range with G13 respectively. Among all nine genotypes were at par (2.08 ± 0.18) with the check JS-9752 (2.08) while G1 (2.30), G2 (2.32), G4 (2.30), G11 (2.32), G13 (2.52), G14 (2.34), G16 (2.33) and G18 (2.34) was significantly superior over the check and rest *i.e*. G3 and G10 had lesser no. of seeds per pod (*i.e*. <1.9).

IX. Pod length (cm)

The length of soybean pods in the study varied from 2.81 (G2) to 4.09 (G13). Fourteen genotypes *viz*.,G1, G3, G4, G5, G6, G9, G10, G11, G12, G13, G15, G16, G18 and G20 had lengthier (*i.e*. >3.11 cm) pods compared to check JS-9752 (2.97) while four genotypes were found to be at par (2.97 ± 0.14) and only G2 (2.81) had least pod length. General mean value obtained was 3.30 ± 0.07.

X. Hundred seed weight (g)

The average 100-seed weight was 8.32 ± 0.18, with the range being 2.82 (G14) to 13.47 (G3). G2 (3.10) and G10 (13.39) were at par with G14 and G3 respectively. Genotypes G3 (13.47), G6 (12.00), G10 (13.39), G13 (12.21), G16 (13.04) and G20 (12.23) were significantly superior to check JS-9752 (9.24). However rest (thirteen) other genotypes exhibited significantly lesser (<8.89) 100-seed weight than check.

XI. Oil content (%)

Oil content expressed in percentage for all genotypes under observation varied from 14.17 (G8) to 22.76 (G3) with general mean 20.04 ± 0.57%. G6, G10, G16, G17 and G20 were at par with G3 having higher oil percent. Almost 50% (eleven) of the genotypes were at par (21.28 ± 1.14) with the check G19 (21.28) while G2 (14.90), G5 (18.93), G7 (19.20), G8 (14.17) and G14 (15.50) were inferior. Only three genotypes videlicet, G3 (22.76), G6 (22.73) and G20 (22.70) possessed more oil % than check.

XII. Biological yield per plant (g)

The mean data for biological yield for all the genotypes under trial ranged from 9.60 (G13) to 63.09 (G1) with general mean value of 27.28 ± 3.57. Eight genotypes (G1, G4, G8, G9, G10, G11, G12 and G18) under investigation yielded significantly more (>26.14) than check JS-9752 (19.03)and remaining nine (G2, G3, G5, G6, G7, G14, G15, G17 and G20) were at par (19.03 ±7.11). Only G13 and G16 weighed lesser than check.

XIII. Harvest index (%)

It varied between 23.17 (G4) and 38.90 (G6) with general mean of 29.76 ± 1.58. Genotypes G2 (24.79), G3 (26.06), G5 (25.71), G8 (24.03), G14 (24.29), G15 (26.00) and G18 (24.06) were at par with G2 while G16 (38.19) and G19 (36.71) were in range with G6 respectively. None of the genotypes were superior to check JS-9752 (36.71) for this trait but seven were at par (36.71 ± 3.15) *viz*. G6, G7, G10, G13, G16, G17, and G20. Rest twelve genotypes had lesser (<33.56) harvest index percent compared to check.

XIV. Seed yield per plant (g)

Mean values for seed yield per plant ranged from 2.98 g (G13) to 16.33 g (G1) with a general mean of 7.33 ± 1.02g. genotypes G2 and G16 were at par with G13. Six genotypes viz. G1 (16.33), G8 (8.48), G9 (8.80), G10 (12.24), G11 (9.12) and G12 (9.02) yielded more (*i.e*. >8.41) per plant compared to check variety JS-9752 (6.37). Rest nearly 50% of the tested genotypes (eleven) was at par (6.37 ± 2.04)) and G13 and G16 was only genotype with minimum yield w.r.t. check (<4.33g).

Table 4.3: Mean performance for fourteen charactersof soybean genotypes

Sl.No.	Genotypes	DFF	DM	PH	NPB/P	NC/P	NPo/C	NPo/P	NS/Po	PoL	HSW	Oil %	BY/P	HI	SY/P
1	**G1**	78.83	137.33	61.97	4.64	12.22	3.24	115.00	2.30	3.47	6.38	17.73	63.09	28.38	16.33
2	**G2**	79.50	148.00	77.46	3.64	2.56	2.24	73.64	2.32	2.81	3.10	14.90	20.85	24.79	4.86
3	**G3**	48.83	107.17	30.53	2.42	1.42	1.15	22.33	1.64	4.05	13.47	22.76	24.08	26.06	5.76
4	**G4**	75.50	136.83	44.86	3.53	9.69	2.83	68.47	2.30	3.34	5.53	20.52	32.79	23.17	7.28
5	**G5**	48.67	114.50	52.75	3.69	5.97	2.19	57.00	2.00	3.17	7.03	18.93	27.78	25.71	6.84
6	**G6**	39.00	98.50	28.27	2.97	4.44	3.14	36.78	2.08	3.34	12.00	22.73	23.90	38.90	7.99
7	**G7**	52.17	113.67	39.67	3.92	4.97	2.49	64.08	2.07	2.86	6.20	19.20	23.18	31.67	7.02
8	**G8**	71.17	143.33	58.90	4.28	15.72	3.12	111.44	2.17	2.94	3.79	14.17	34.94	24.03	8.48
9	**G9**	49.50	106.00	55.11	3.28	5.17	2.79	55.22	2.03	3.30	8.80	20.80	31.74	29.28	8.80
10	**G10**	49.83	108.00	48.94	3.61	5.97	2.79	51.03	1.85	3.26	13.39	21.70	37.54	35.48	12.24
11	**G11**	49.83	109.17	44.62	3.89	4.03	2.39	60.33	2.32	3.41	7.30	20.50	33.38	28.59	9.12
12	**G12**	72.00	133.50	56.62	4.11	6.11	2.92	69.47	2.25	3.20	6.87	20.80	32.02	28.84	9.02
13	**G13**	32.83	91.33	29.90	2.42	0.31	0.58	10.92	2.52	4.09	12.21	21.47	9.60	33.30	2.98
14	**G14**	77.83	148.00	71.99	3.50	3.92	2.05	82.33	2.34	2.85	2.82	15.50	21.86	24.29	5.12
15	**G15**	64.67	133.00	45.73	3.97	4.69	2.59	47.94	2.17	3.13	7.66	20.33	25.30	26.00	6.55
16	**G16**	34.50	89.33	26.71	1.06	0.56	0.82	11.03	2.33	3.98	13.04	22.25	10.06	38.19	3.05
17	**G17**	49.67	113.33	39.79	3.17	3.72	2.35	50.06	1.96	2.89	7.00	22.10	22.69	32.36	6.44
18	**G18**	63.83	135.00	47.13	3.53	1.83	1.63	42.22	2.34	3.54	8.40	20.47	30.39	24.06	6.37
19	**G19 (check)**	49.50	100.00	32.91	2.89	3.72	3.12	38.47	2.08	2.97	9.24	21.28	19.03	36.71	6.37
20	**G20**	39.17	98.83	31.60	3.28	3.47	2.40	28.56	2.21	3.51	12.23	22.70	21.36	35.48	6.06
Grand mean		**56.34**	**118.24**	**46.27**	**3.39**	**5.03**	**2.34**	**54.82**	**2.16**	**3.30**	**8.32**	**20.04**	**27.28**	**29.76**	**7.33**
CD at 5%		**2.17**	**2.54**	**4.01**	**0.57**	**1.78**	**0.58**	**11.80**	**0.18**	**0.14**	**0.35**	**1.14**	**7.11**	**3.15**	**2.04**

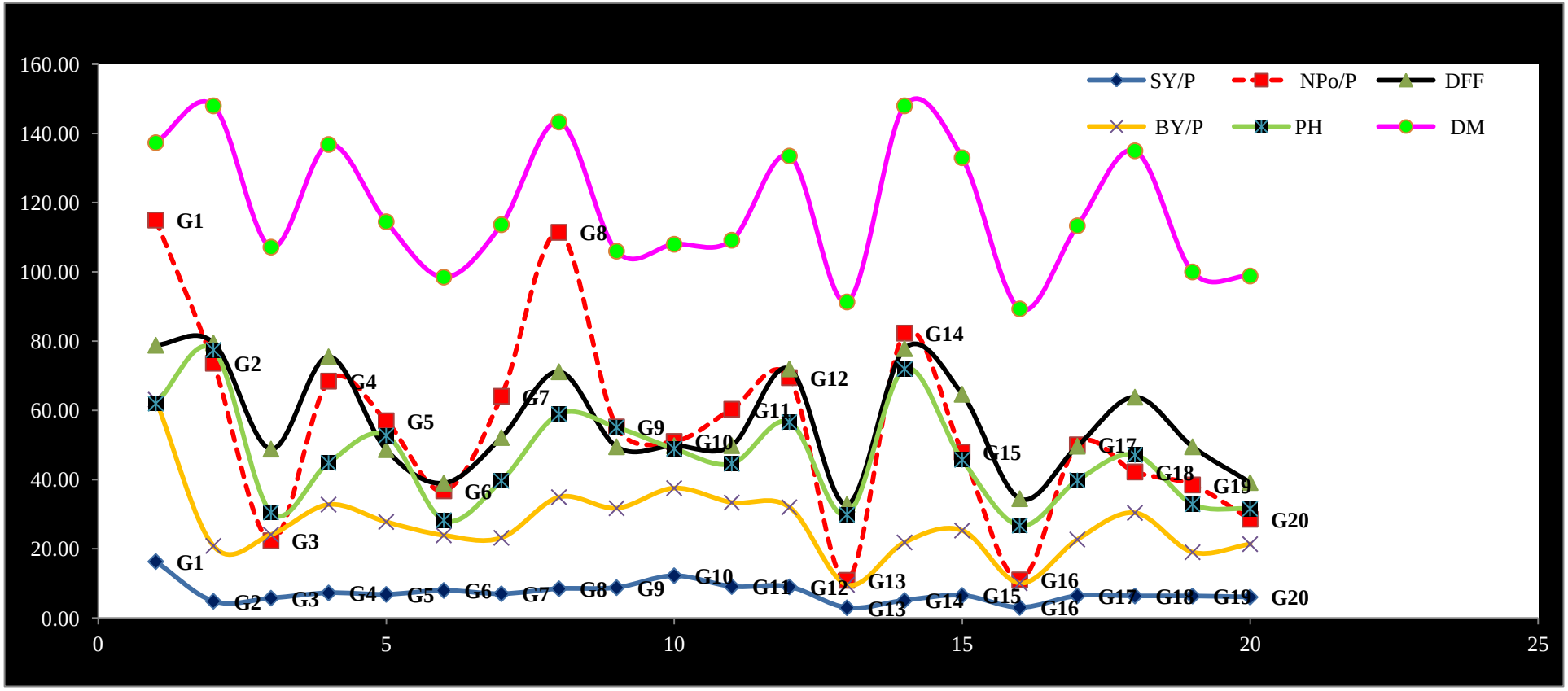


Figure 4.1: Scatter diagram representing relation between number of pods per plant (NPo/P), days to 50% flowering (DFF), biological yield per plant (BY/P), plant height (PH), days to maturity (DM), and seed yield/plant (SY/P)

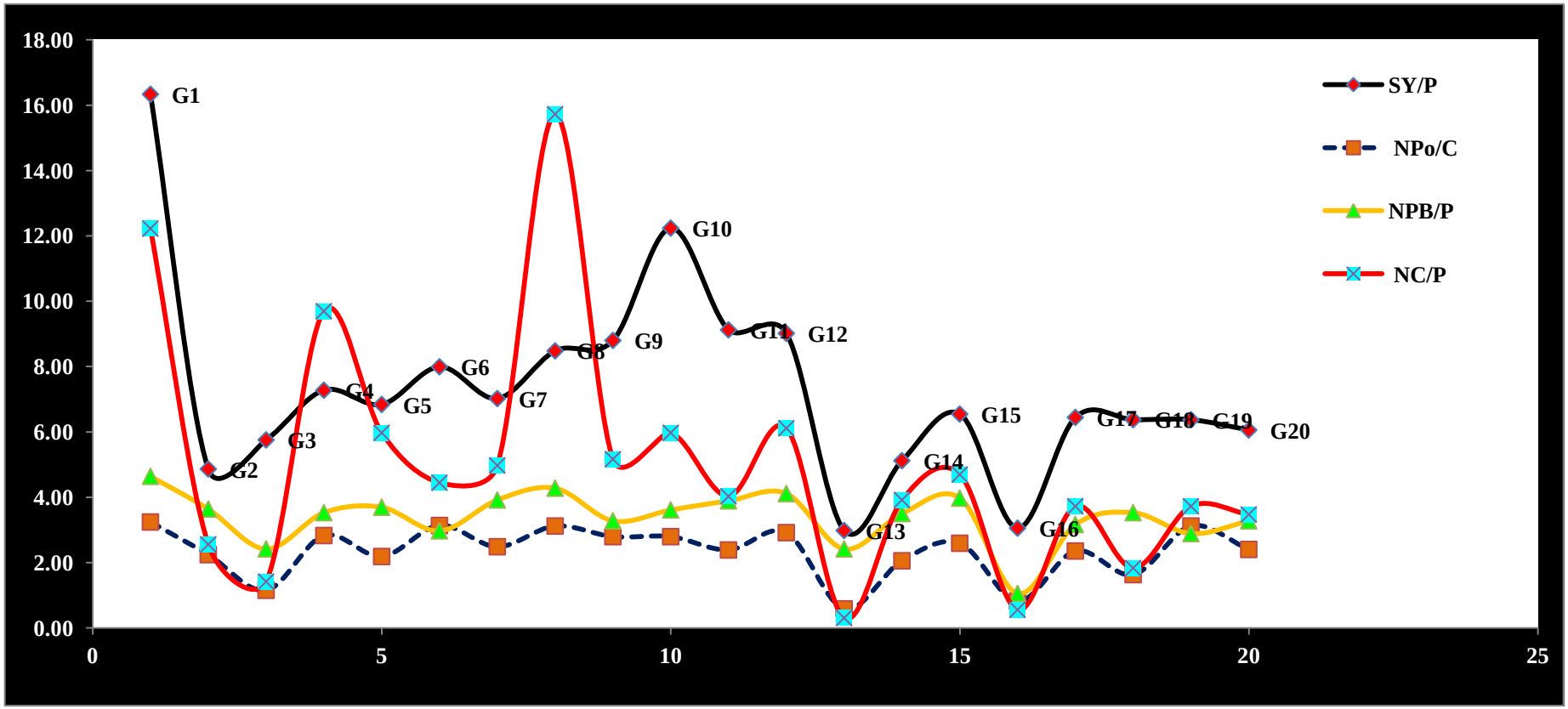


Figure 4.2: Scatter diagram representing relation between number of pods per cluster (NPo/C), number of primary branches per plant (NPB/P), number of clusters/plant (NC/P) and seed yield/plant (SY/P)

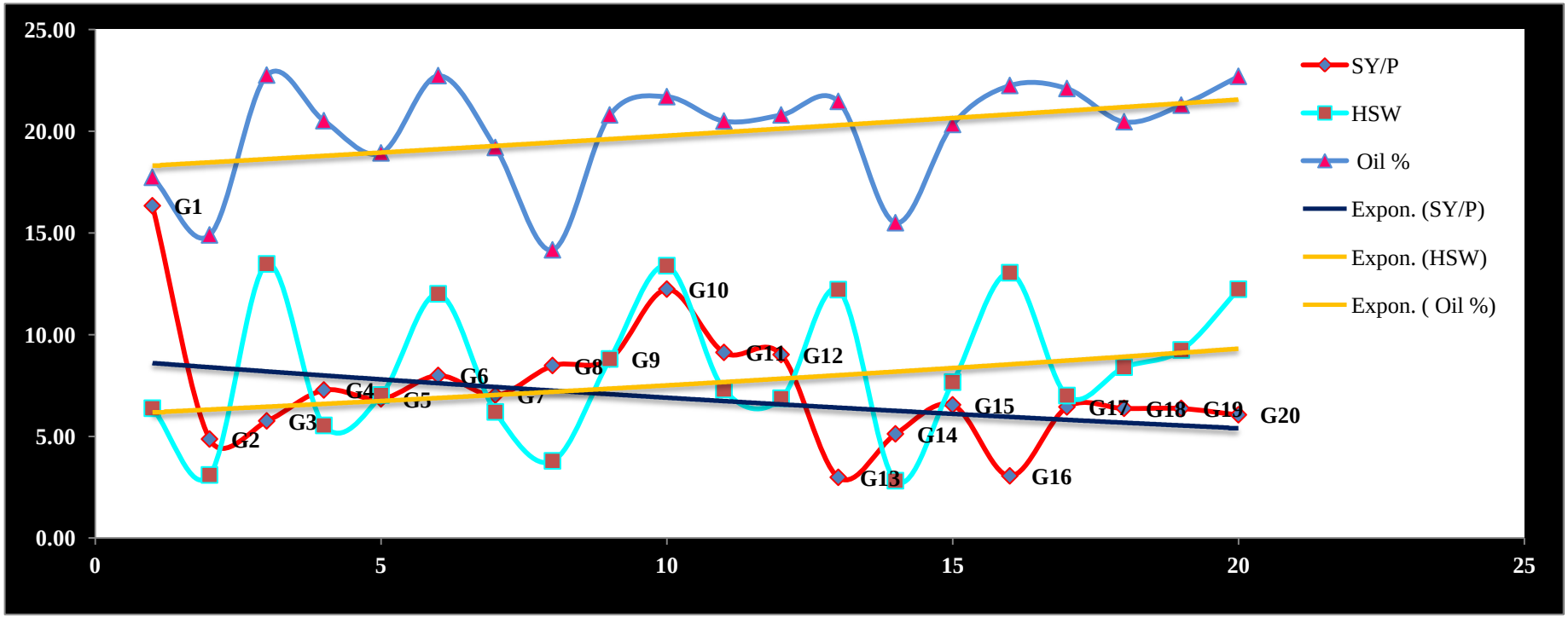


Figure 4.3: Scatter diagram representing relation between hundred seed weight (HSW), oil% and seed yield/plant (SY/P)

Abbreviations:-

DFF: Days to 50% flowering, **DM:** Days to maturity, **PH**: Plant height (cm), **NPB/P**: Number of primary branches per plant, **NC/P**: Number of clusters per plant, **NPo/C**: Number of pods per cluster, **NPo/P**: Number of pods per plant, **NS/Po**: Number of seeds per pod, **PoL**: Pod length (cm), **HSW**: Hundred seed weight, **BY/P**: Biological yield per plant (g), **HI**: Harvest index and **SY/P**: Seed yield per plant (g)

4.3 Genotypic and phenotypic coefficient of variation (GCV and PCV)

Plant breeders must choose superior individuals based on their phenotypic expression to improve any character. This may occasionally be misleading because both heritable and non-heritable factors contribute to how a character develops. Under such circumstances, the coefficient of variation is the best tool for assessing the relative magnitude of character variation and also predicting the degree of variability in the sample population. Thus, for isolating high yielding genotypes, it is prerequisite to know the degree of population variability. This indicates that dividing the total variability into its heritable and non-heritable components is essential. Table 4.4 presents the phenotypic and genotypic coefficients of variation (PCV and GCV) for various traits in the research material to understand the real facts about variability.

According to the data, PCV was just barely greater than the corresponding GCV. GCV and PCV were categorized as low (<10 %), moderate (10-20 %) and high (>20 %) as suggested by Sivasubramanian and Madhavamenon (1973). Table 4.4 shows the estimates of the genotypic and phenotypic coefficients of variation for yield and its components are discussed below:

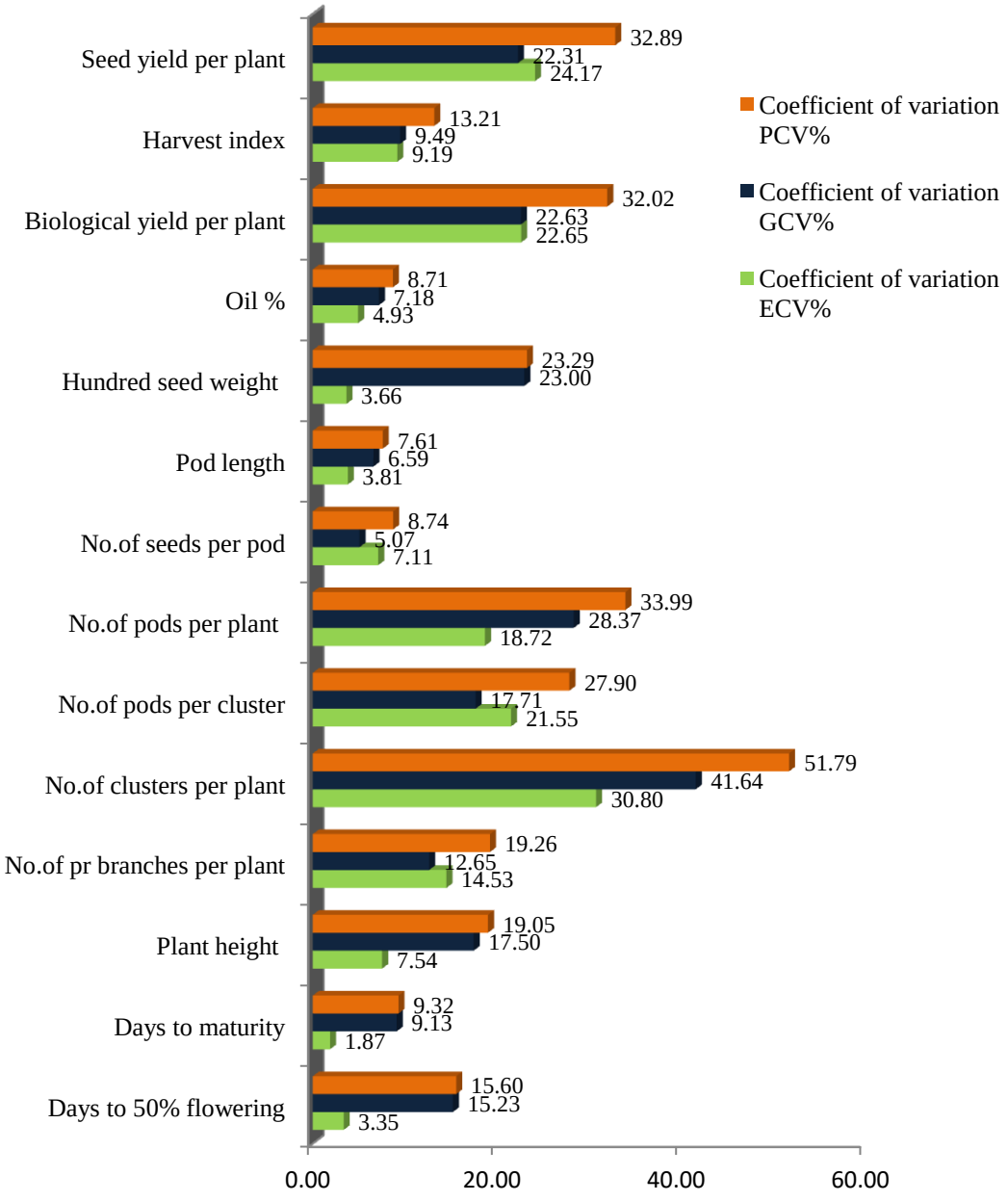


Figure 4.4: Graphical representation of GCV and PCV for various associated traits of seed yield

The number of clusters per plant (41.64%) had the highest magnitude of GCV followed by number of pods per plant (28.37%), hundred seed weight (23.00%), biological yield per plant (22.63%) and seed yield per plant (22.31%).

The number of pods per cluster (17.71%), plant height (17.50%), days to 50% flowering (15.23%), and number of primary branches per plant (12.65%) all showed moderate magnitudes of GCV.

Harvest index (9.49%), days to maturity (9.13%), oil content (7.18%), pod length (6.59%) and number of seeds per pod (5.07%) observed low magnitude of GCV. The range of GCV varied between 5.07-41.64%.

The PCV values ranged from 7.61 to 51.79%. The magnitude for PCV was highest for number of clusters per plant (51.79%) followed subsequently by number of pods per plant (33.99%), seed yield per plant (32.89%), biological yield per plant (32.02%), number of pods per cluster (27.90%) and hundred seed weight (23.29%).

The moderate magnitude of PCV was recorded by number of primary branches per plant (19.26%), plant height (19.05%), days to 50 per cent flowering (15.60%) and harvest index (13.21%).

Days to maturity (9.32%), number of seeds per pod (8.74%), oil (8.71%) and pod length (7.61%) resulted with low values of PCV %.

4.4 Heritability (h^2) and genetic advance (GA)

An important factor that defines the range of any crop species is the nature and extent of inherent capacity of a genotype for a character. Without sufficient heritability, genetic advance and genetic variability, genetic enhancement of any

character is challenging. Hence, heritability and genetic advance are crucial factors in choosing a genotype because they enable better selection efficiency by separating out the environmental influence from total variability.

Table 4.4: Genetic parameters for various morphological traits in soybean

Sl.No	Character	Coefficient of variation			h^2 (%) (broadsense)	Genetic Advance as percent of mean
		PCV%	GCV%	ECV%		
1	**DFF**	15.60	15.23	3.35	95.39	30.65
2	**DM**	9.32	9.13	1.87	95.97	18.41
3	**PH**	19.05	17.50	7.54	84.34	33.10
4	**NPB/P**	19.26	12.65	14.53	43.15	17.12
5	**NC/P**	51.79	41.64	30.80	64.65	68.98
6	**NPo/C**	27.90	17.71	21.55	40.31	23.16
7	**NPo/P**	33.99	28.37	18.72	69.67	48.78
8	**NS/Po**	8.74	5.07	7.11	33.72	6.07
9	**PoL**	7.61	6.59	3.81	74.99	11.76
10	**HSW**	23.29	23.00	3.66	97.53	46.78
11	**Oil %**	8.71	7.18	4.93	68.01	12.21
12	**BY/P**	32.02	22.63	22.65	49.94	32.94
13	**HI**	13.21	9.49	9.19	51.59	14.04
14	**SY/P**	32.89	22.31	24.17	46.00	31.16

Abbreviations: DFF: Days to 50% flowering, **DM:** Days to maturity, **PH**: Plant height (cm), **NPB/P**: Number of primary branches per plant, **NC/P**: Number of clusters per plant, **NPo/C**: Number of pods per cluster, **NPo/P**: Number of pods per plant, **NS/Po**: Number of seeds per pod, **PoL**: Pod length (cm), **HSW**: Hundred seed weight, **BY/P**: Biological yield per plant (g), **HI**: Harvest index and **SY/P**: Seed yield per plant (g).

Heritability estimates provide some insight into the gene activity responsible for the expression of various polygenic characteristics. As

recommended by Robinson(1966), the heritability in the current study was studied throughout a wide spectrum and categorized as follows: >60% indicates high, 30% to 60% indicates moderate, and 30% indicates low. Genetic variability, heritability, and selection intensity all have a role in how well genetic advance works. The experiment's heritability and genetic advance are shown in Table 4.4 and described below.

On the basis of this categorization, it was noted from the table that the heritability (broad sense) ranged from 33.72 to 97.53 per cent. High heritability were found in the present study for the traits *viz.*, 100 seed weight (97.53%), days to maturity (95.97%), days to 50 per cent flowering (95.39%), plant height (84.34%), pod length (74.99%), number of pods per plant (69.67%), oil content (68.01%) and number of clusters per plant (64.65%).

The harvest index (51.59%), biological yield per plant (49.94%), seed yield per plant (46.00%), number of primary branches per plant (43.15%), number of pods per cluster (40.31%) and number of seeds per pod (33.72%) showed moderate heritability.

However, the heritability value alone does not indicate the genetic improvement resulting from the selection of superior genotypes. When expressed in terms of genetic advance, heritability estimates are more beneficial. According to Hanson (1961), the concepts of heritability and genetic advance are mutually supportive. However, the existence of high genetic advance in a character does not necessarily imply high heritability (Johnson *et al*. 1955).The average percentage was calculated for the comparison of development on various traits of distinctive genetic advance.

In terms of genetic advances, the categories were divided as follows: >20% indicates high GAM, 10% to 20% indicates moderate GAM and <10% indicates

low GAM. Number of clusters per plant (68.98%), number of pods per plant (48.78%), 100 seed weight (46.78%), plant height (33.10%), biological yield per plant (32.94%), seed yield per plant (31.16%), days to 50 per cent flowering (30.65%) and number of pods per cluster (23.16%)were the traits with high genetic advance as a percentage of the mean.

Moderate values were recorded in descending order for days to maturity (18.41%) and number of primary branches per plant (17.12%), harvest index (14.04%), oil (12.21%), pod length (11.76%), and low for number of seeds per pod (6.07%) respectively.

Days to 50% flowering, plant height, number of clusters per plant, number of pods per plant and 100 seed weight all showed high heritability in combination with high GAM.

High heritability coupled with moderate GAM was recorded for days to maturity, pod length and oil content.

Moderate heritability coupled with high GAM were obtained for traits such number of pods per cluster, biological yield per plant and seed yield per plant.

Moderate heritability coupled with moderate GAM resulted for as number of primary branches per plant and harvest index.

Moderate heritability coupled with low GAM was observed for number of seeds per pod.

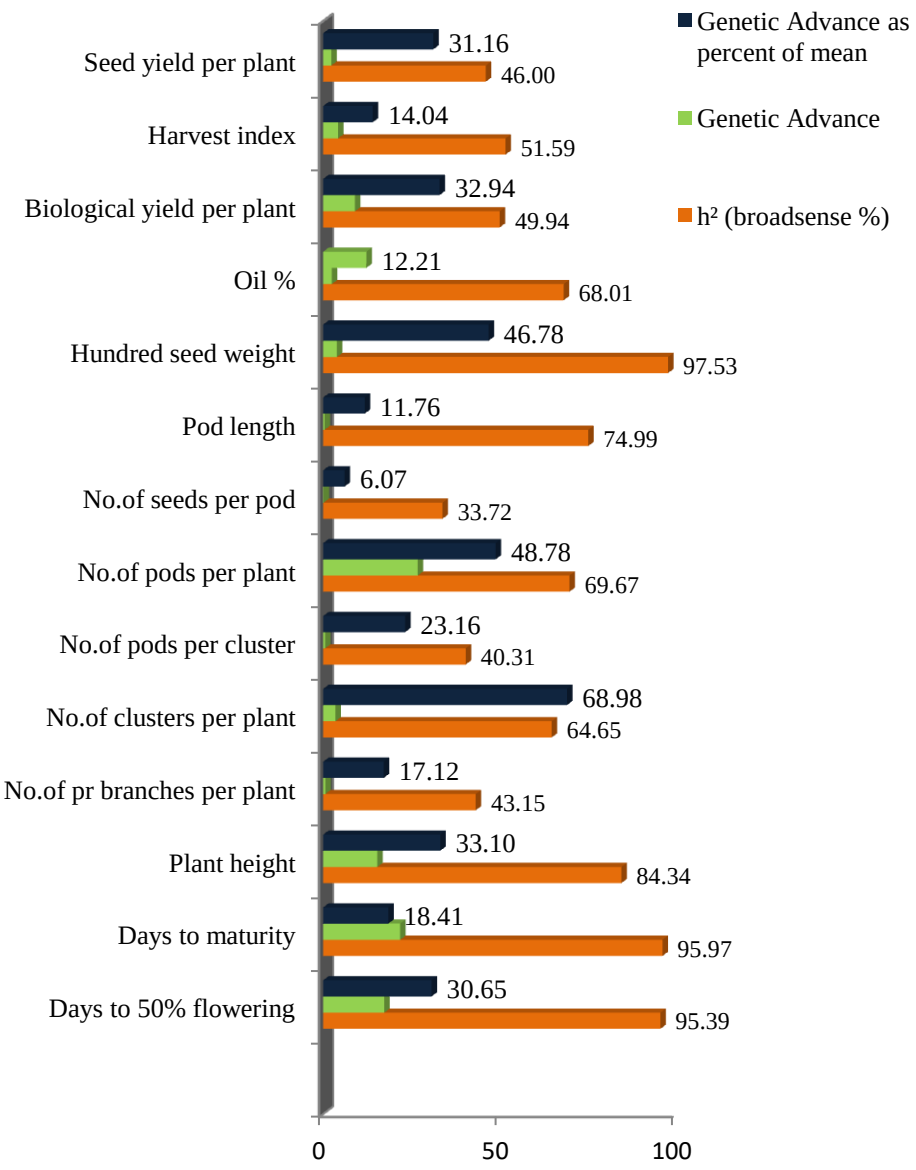


Fig. 4.5: Heritability and genetic advance as a % of mean presented in bar diagram

4.5 Correlation analysis

Direct selection for yield is inefficient because this quantitative traits is complex and highly environment-dependent or influenced. If selection is made solely on the basis of yield, improvement will be constrained by high genotype-environment interaction. Thus, selection based on yield component characters can effectively boost yield. Since selection on any particular trait may result in undesirable changes in other associated traits, the primary goal of correlation studies is to determine the suitability of multiple characters for indirect selection (Singh, 1999).

For the majority of the character pairs studied in the current investigation, the genotypic correlation coefficients were generally higher in magnitude when compared to the phenotypic correlation coefficients values, indicating a strong inherent association between the characters under study; while expression of their association is dimmed by the influence of the environment. This may be because of the genotypes' relative stability, given that the majority of genotypes underwent certain level of selection Johnson *et al.*,1955. These outcomes support the conclusions reached by Singh *et al.*, 2000, Sultana *et a.*,2005 and Malik *et al.*,2007.

To assess the relationship between two traits, the correlation coefficient at the phenotypic and genotypic levels was constructed for all potential combinations between yield components (Table 4.5).In the present study 91associations were obtained. Among them 36associations were found positive and significant($p \leq 0.01$) and twenty four associations were found to be negative but significant. However thirty one were found to be non-significant.

Seed yield per plant (g)

Number of primary branches per plant (0.71, 0.53), number of clusters per plant (0.67, 0.64), number of pods per cluster (0.73, 0.57), number of pods per plant (0.62, 0.67), and biological yield per plant (0.95, 0.94) showed highly significant and positive correlation with seed yield at the genotypic and phenotypic levels.

At both, genotypic and phenotypic levels, traits including days to 50% flowering, days to maturity, plant height, number of seeds per pod,pod length,100 seed weight (g), oil percentage, and harvest index were non-significant.

I. Days to 50% flowering

This trait showed highly significant positive association both at genotypic and phenotypic level with traits *viz*., days to maturity (0.97, 0.95) plant height (0.84, 0.78), number of pods per plant(0.84, 0.71), number of primary branches per plant (0.74, 0.51), number of clusters per plant (0.56, 0.48), biological yield per plant (0.57, 0.45) and number of pods per cluster (0.45). While, pod length (-0.52, -0.49), oil% (-0.74, -0.69), 100 seed weight (-0.82, -0.78), and harvest index (-0.83, -0.55) were significant but negatively correlated.

II. Days to maturity

Plant height (0.85, 0.79), number of pods per plant (0.81, 0.70), number of primary branches per plant (0.74, 0.52), number of clusters per plant (0.53, 0.45)and biological yield per plant (0.48, 0.37) all had positive significant correlations at both the genotypic and phenotypic levels. However, a highly significant negative association was found with pod length (-0.55, -0.51), oil% (-0.79, -0.74), 100 seed weight (-0.84, -0.81) and harvest index (-0.91, -0.57).

III. Plant height (cm)

Plant height exhibited significant and positive association with number of pods per plant (0.81, 0.69), number of primary branches per plant (0.69, 0.52), biological yield per plant (0.47)and number of clusters per plant (0.43)whereas, negative significant relationship with pod length (-0.59, -0.53), harvest index (-0.73, -0.50), 100 seed weight (-0.81, -0.73) and oil% (-0.86, -0.75).

IV. Number of primary branches per plant

Number of pods per plant (0.90, 0.59), biological yield per plant (0.80, 0.54), number of pods per cluster (0.76, 0.56) and Number of clusters per plant (0.73, 0.54) were found highly significant and positively related whereas harvest index (-0.58), oil% (-0.59), pod length (-0.64, -0.48) and 100 seed weight (-0.70, -0.48) are associated negatively with significant values to number of primary branches per plant.

V. Number of clusters per plant

This character recorded highly significant positive correlation with number of pods per plant (0.84, 0.82), biological yield per plant (0.74, 0.68) and number of pods per cluster (0.73, 0.61) but significant negative correlation with, 100 seed weight (-0.50, -0.44) and oil% (-0.54, -0.44) at both level. However number of seeds per pod and harvest index was negative and non-significant with the trait.

VI. Number of pods per cluster

The trait number of pods per cluster showed significantly high positive correlation with number of pods per plant (0.69, 0.52) and biological yield per plant (0.65, 0.50) while, significant negative correlation with pod length (-0.70, -0.53).

VII. Number of pods per plant

The trait showed high significant positive correlation with biological yield per plant (0.73, 0.74). Whereas, significant negative correlation with pod length (-0.64,-0.47), harvest index (-0.69), oil% (-0.82, -0.68), and hundred seed weight (-0.82, -0.71) at both levels.

VIII. Number of seeds per pod

The character number of seeds per pod showed non-significant association with all other traits.

IX. Pod length (cm)

Pod length exhibited positive significant correlation with 100 seed weight (0.75, 0.69) and oil% (0.58, 0.50). However, biological yield per plant (g) and harvest index were non-significant both at genotypic and phenotypic level.

X. Hundred Seed Weight

This trait is positively correlated to oil% (0.85, 0.77) and harvest index (0.78, 0.47).

XI. Oil%

At both the genotypic and phenotypic levels, Oil% showed significant positive association with harvest index (0.70, 0.42).

XII. Biological yield per plant (g)

This trait showed negative association at genotypic level with harvest index (-0.46) and non significant association (-0.04) at phenotypic level.

Table 4.5:Correlation between seed yield and its contributing traits in soybeanat both Genotypic and phenotypic level

Characters		DFF	DM	PH	NPB/P	NC/P	NPo/C	NPo/P	NS/Po	PoL	HSW	Oil %	BY/P	HI	SY/P
DFF	**G**	**1.00**	0.97**	0.84**	0.74**	0.56**	0.45*	0.84**	0.25 ^{NS}	-0.52*	-0.82**	-0.74**	0.57**	-0.83**	0.37 ^{NS}
	P	**1.00**	0.95**	0.78**	0.51*	0.48*	0.36^{NS}	0.71**	0.19 ^{NS}	-0.49*	-0.78**	-0.69**	0.45*	-0.55**	0.28 ^{NS}
DM	**G**		**1.00**	0.85**	0.74**	0.53*	0.37^{NS}	0.81**	0.26 ^{NS}	-0.55**	-0.84**	-0.79**	0.48*	-0.91**	0.26 ^{NS}
	P		**1.00**	0.79**	0.52*	0.45*	0.29^{NS}	0.70**	0.21 ^{NS}	-0.51*	-0.81**	-0.74**	0.37 ^{NS}	-0.57**	0.20 ^{NS}
PH	**G**			**1.00**	0.69**	0.43*	0.36^{NS}	0.81**	0.24 ^{NS}	-0.59**	-0.81**	-0.86**	0.47*	-0.73**	0.33 ^{NS}
	P			**1.00**	0.52*	0.36^{NS}	0.30^{NS}	0.69**	0.13^{NS}	-0.53*	-0.73**	-0.75**	0.40 ^{NS}	-0.50*	0.31 ^{NS}
NPB/P	**G**				**1.00**	0.73**	0.76**	0.90**	0.07^{NS}	-0.64**	-0.70**	-0.59**	0.80**	-0.58**	0.71**
	P				**1.00**	0.54*	0.56**	0.59**	0.02 ^{NS}	-0.48*	-0.48*	-0.39 ^{NS}	0.54*	-0.40 ^{NS}	0.53**
NC/P	**G**					**1.00**	0.73**	0.84**	-0.02^{NS}	-0.40 ^{NS}	-0.50*	-0.54*	0.74**	-0.42 ^{NS}	0.67**
	P					**1.00**	0.61**	0.82**	0.01 ^{NS}	-0.31 ^{NS}	-0.44*	-0.44*	0.68**	-0.16 ^{NS}	0.64**
NPo/C	**G**						**1.00**	0.69**	-0.19 ^{NS}	-0.70**	-0.41 ^{NS}	-0.26 ^{NS}	0.65**	-0.05^{NS}	0.73**
	P						**1.00**	0.52**	-0.11 ^{NS}	-0.53*	-0.32 ^{NS}	-0.18 ^{NS}	0.50*	0.00 ^{NS}	0.57**
NPo/P	**G**							**1.00**	0.13^{NS}	-0.64**	-0.82**	-0.82**	0.73**	-0.69**	0.62**
	P							**1.00**	0.16 ^{NS}	-0.47*	-0.71**	-0.68**	0.74**	-0.20 ^{NS}	0.67**
NS/Po	**G**								**1.00**	0.07^{NS}	-0.35 ^{NS}	-0.34 ^{NS}	-0.13^{NS}	-0.22 ^{NS}	-0.25 ^{NS}
	P								**1.00**	0.10 ^{NS}	-0.28 ^{NS}	-0.24 ^{NS}	0.07 ^{NS}	0.13 ^{NS}	0.01 ^{NS}
PoL	**G**									**1.00**	0.75**	0.58**	-0.18 ^{NS}	0.24 ^{NS}	-0.20 ^{NS}
	P									**1.00**	0.69**	0.50*	-0.08 ^{NS}	0.24 ^{NS}	-0.08 ^{NS}
HSW	**G**										**1.00**	0.85**	-0.30 ^{NS}	0.78**	-0.12^{NS}
	P										**1.00**	0.77**	-0.24 ^{NS}	0.47*	-0.08 ^{NS}
Oil %	**G**											**1.00**	-0.30 ^{NS}	0.70**	-0.13^{NS}
	P											**1.00**	-0.21 ^{NS}	0.42 ^{NS}	-0.09 ^{NS}
BY/P	**G**												**1.00**	-0.46*	0.95**
	P												**1.00**	-0.04 ^{NS}	0.94**
HI	**G**													**1.00**	-0.14^{NS}
	P													**1.00**	0.15 ^{NS}

* and ** Significant at 5 %and 1%probability level

Abbreviations:-

DFF: Days to 50% flowering, **DM:** Days to maturity, **PH**: Plant height (cm), **NPB/P**: Number of primary branches per plant, **NC/P**: Number of clusters per plant, **NPo/C**: Number of pods per cluster, **NPo/P**: Number of pods per plant, **NS/Po**: Number of seeds per pod, **PoL**: Pod length (cm), **HSW**: Hundred seed weight, **BY/P**: Biological yield per plant (g), **HI**: Harvest index and **SY/P**: Seed yield per plant (g).

4.6Path coefficient analysis

The direct effect of the yield component and indirect effect through other yield attributing characters combine to produce the observed correlation between yield and yield attributing characters. The overall relationship between grain yield and its component characters can occasionally be deceptive. Since estimations of its relationship with other characters may be overestimated or understated. Therefore, direct selection based solely on correlated response may not be beneficial.

As multiple characters have an effect on a single trait, the correlation coefficient must be divided into direct and indirect effects using path coefficient analysis. Thus, when used together, correlation and path analysis can provide a deeper understanding of cause and effect relationship between various character pairs. The following general ideas may be kept in mind when interpreting the findings of path analysis: (Singh and Chaudhary 1977).

- Correlation explains the true relationship through the character if the correlation coefficient between a causal factor and the effect is almost equivalent to its direct effects.

- The indirect effects appear to be the cause of a positive correlation when the correlation coefficient is positive but the direct effect is negligible or negative. In these cases, the indirect causal factors are to be considered simultaneously for selection.
- The direct effect is positive and high even though the correlation coefficient may be negative. A restricted simultaneous selection model should be used in these cases, meaning restrictions should be imposed in order to eliminate any unfavourable indirect effects so that the direct effect can be utilized.
- If both the correlation coefficient and the direct effects are negative, we must discard the character-based selection.
- The residual effect determines how well the causal factors can explain the dependent factor's variability. To completely account for the difference in yield, other factors that have not been taken into consideration in this analysis must be added if residual impact is high.

Based on the aforementioned criteria, path coefficient analyses for 14 characters with seed yield were conducted in order to gather data on the direct and indirect contributions of different yield components to yield and to create a foundation for selection in soybean. Lenka and Mishra (1973) ratings for the direct and indirect effects are: 0.00-0.09 – Negligible; 0.10-0.19 – Low; 0.20-0.29 – Moderate; 0.30-0.99 – High; and >1.00 - Very high significant and vice-versa given by. Character-wise, the genotypic level path coefficient analysis results are provided along with a discussion.

Direct effect

The genotypic path coefficient revealed that number of pods per cluster (3.12), number of pods per plant (1.86), pod length (1.71), days to maturity (1.25) and 100 seed weight (0.61) recorded high significant positive direct effect on seed yield per plant.

However, high negative direct effect was contributed by days to 50 per cent flowering (-2.01), number of clusters per plant (-1.92), harvest index (-1.00), number of primary branches per plant (-0.69), oil% (-0.76); moderate effect of plant height (-0.19) and negligible values of number of seeds per pod (-0.03) and biological yield per plant (-0.01).

Indirect effect

I. Days to 50% flowering

It had a negative direct effect on yield but contributed through high indirect effects of number of pods/plant (1.56), number of pods/cluster (1.41), days to maturity (1.22), harvest index (0.83) and oil% (0.56) which counter balanced the indirect negative values of number of clusters/plant (-1.09), pod length (-0.89), number of primary branches/plant (-0.51), hundred seed weight(-0.50), plant height (-0.16), number of seeds/pod (-0.01) and bio yield/plant (-0.01).

II. Days to maturity

This trait being positive for direct effect contributes towards yield *via* number of pods/plant (1.50), number of pods/cluster (1.14), harvest index (0.90) and oil% (0.60) suppressing negative indirect effect of days to 50% flowering(-1.98), number of clusters/plant (-1.01), pod length (-0.93), number of primary

branches/plant (-0.51), 100 seed weight (-0.51), plant height (-0.16), number of seeds/pod (-0.01) and bio yield/plant (-0.01).

III. Plant height (cm)

Plant height had a negative direct impact on seed yield. Negative indirect effects through 50% flowering (-1.72), pod length (-1.01), number of clusters/plant (-0.84), 100 seed weight (-0.49), number of primary branches/plant (-0.48), number of seeds per pod (-0.01) and biological yield/plant (-0.01) were counter balanced by positive indirect effects manifested *via* number of pods per plant (1.51), number of pos/cluster (1.12), days to maturity (1.06), harvest index (0.73) and oil % (0.65).

IV. Number of primary branches per plant

It exhibited positive association with seed yield (0.71**) mainly at genotypic level due to its high positive indirect effect *via* number of pos/cluster (2.36), number of pods/plant (1.67), days to maturity (0.93), harvest index (0.58) and oil (0.45) which nullified negative indirect traits such days to 50 percent flowering (-1.50), number of clusters/plant (-1.40), pod length (-1.09), 100 seed weight (-0.43), as plant height (-0.13) and bio yield/plant (-0.01).

V. Number of clusters per plant

Number of clusters/plant showed negative direct effect on seed yield per plant. Positive indirect effects were recorded through number of pods/cluster (2.28), number of pods/plant (1.56), days to maturity (0.66), harvest index (0.42), oil content (0.41) and seeds per pod (0.00) and negative indirect effect throughdays to 50 % flowering (-1.15), pod length (-0.68), number of branches per plant (-0.51), hundred seed weight (-0.30), plant height (-0.08) and biological

yield/plant (-0.01) which finally resulted in significant positive genotypic correlation (0.67**) with seed yield per plant.

VI. Number of pods per cluster

The character had recorded indirect positive effects on the seed yield *via* number of pods/plant (1.28), days to maturity (0.46), oil (0.19), harvest index (0.05) number of seeds/pod (0.01, negligible) nullifying negative effects of characters such as number of clusters/plant (-1.40), pod length (-1.20), days to 50% flowering (-0.92), number of primary branches/plant (-0.52), hundred seed weight (-0.25), plant height (-0.07) and biological yield/plant (-0.01) and giving significant positive relationship with yield (0.73**).

VII. Number of pods per plant

Number of pods/cluster (2.14), days to maturity (1.01), harvest index (0.69) oil% (0.62), and negligible contribution of number of seeds/pod (0.00), were the factors responsible for indirect effect compensating the negative effects on seed yield due to traits such as days to flowering (-1.71), number of clusters/plant (-1.61), pod length (-1.09), number of primary branches/plant (-0.62), hundred seed weight (-0.50), plant height (-0.16) and biological yield/plant (-0.01) resulting in significant association (0.62**). Also this trait has positive direct effect on seed yield.

VIII. Number of seeds per pod

This trait doesn't support much for its selection, in contributing increased yield since the traits obtained are with low and moderate direct and indirect effects finally resulting in non-significant genotypic values (-0.25). Days to maturity (0.33), oil% (0.25), number of pods/plant (0.23), harvest index (0.22), pod length

(0.12) and number of cluster/plant (0.03) contributed towards positive indirect effect to this trait.

IX. Pod length (cm)

Pod length resulted in non-significant (-0.20) association with seed yield/plant but with positive direct effects. Characters such as days to 50% flowering (1.07), number of cluster/plant (0.76) and 100 seed weight (0.46) number of primary branches/plant (0.44) and plant height (0.11), even with positive indirect path values could not nullify the values resulting from traits such as days to maturity (-0.68), number of pods /cluster (-2.20), number of pods/plant (-1.18) , oil (-0.44) and harvest index (-0.24).

X. Hundred Seed Weight

Negative direct and indirect path values of the characters resulted in negative non-significant genotypic correlation values (-0.12). However positive and negative indirect values contributing to 100 seed weight were days to 50% flowering (1.67), pod length (1.27), number of clusters/plant (0.96), number of primary branches/plant (0.49) plant height (0.16), number of seeds/pods (0.01), number of pods/plant (-1.53), number of pods/cluster (-1.29), days to maturity (-1.05), harvest index (-0.77) and oil (-0.65) respectively which resulted in positive direct effect.

XI. Oil%

At genotypic level correlation values oil content showed negative non-significant association (-0.13) with seed yield due to its high negative direct effect and negative indirect effect through number of pods/plant (-1.53), days to maturity (-0.99), number of pods/cluster (-0.80), and harvest index (-0.70). The positive and

indirect effect for characters at genotypic level *viz.* days to 50% flowering (1.52), number of clusters/plant (1.03), pod length (0.98), hundred seed weight (0.52), number of primary branches/plant (0.41), plant height (0.17), number of seeds/pod (0.01) and biological yield/plant (0.00).

XII. Biological yield per plant (g)

Biological yield per plant exhibited negligible negative (-0.01) direct effect on dependent variable but with high 0.95** significant genotypic correlation values. These results were obtained due to positive indirect effects of number of pods/cluster (2.04), number of pods/plant (1.37), days to maturity (0.61), harvest index (0.45) and oil (0.23) which counter balances the negative indirect effects of number of clusters per plant (-1.42),days to 50% flowering (-1.17), number of primary branches per plant (-0.56), pod length (-0.30), hundred seed weight (-0.18) and plant height (-0.09).

XIII. Harvest index

This trait exhibited negative direct (-1.00) effect and negative non-significant (-0.14) correlation on seed yield per plant. The positive indirect effects were manifested through days to 50% flowering (1.70), number of clusters/plant (0.82), hundred seed weight (0.47), pod length (0.42), number of primary branches/plant (0.40), plant height (0.14), biological yield/plant (0.01) and number of seeds/pod (0.01) which was cancelled by negative indirect effects through number of pods per plant (-1.28), days to maturity (-1.13), oil% (-0.53) and number of pods/ cluster (-0.16).

Table 4.6: Direct (diagonal) and Indirect (above and below diagonal) path effects on seed yield in soybean through different characters both at genotypic and phenotypic level

Characters		DFF	DM	PH	NPB/P	NC/P	NPo/C	NPo/P	NS/Po	PoL	HSW	Oil %	BY/P	HI	G and P Correlation (SY/P)
DFF	**G**	**-2.04**	1.22	-0.16	-0.51	-1.09	1.41	1.56	-0.01	-0.89	-0.50	0.56	-0.01	0.83	0.37
	P	**0.07**	-0.16	0.09	0.07	-0.04	0.03	0.29	0.00	0.03	-0.29	-0.05	0.30	-0.06	0.28
DM	**G**	-1.98	**1.25**	-0.16	-0.51	-1.01	1.14	1.50	-0.01	-0.93	-0.51	0.60	-0.01	0.90	0.26
	P	0.07	**-0.17**	0.09	0.07	-0.04	0.02	0.28	0.00	0.03	-0.30	-0.05	0.24	-0.06	0.20
PH	**G**	-1.72	1.06	**-0.19**	-0.48	-0.84	1.12	1.51	-0.01	-1.01	-0.49	0.65	-0.01	0.73	0.33
	P	0.05	-0.13	**0.11**	0.07	-0.03	0.03	0.28	0.00	0.03	-0.27	-0.05	0.27	-0.05	0.31
NPB/P	**G**	-1.50	0.93	-0.13	**-0.69**	-1.40	2.36	1.67	0.00	-1.09	-0.43	0.45	-0.01	0.58	0.71
	P	0.04	-0.09	0.06	**0.13**	-0.04	0.05	0.24	0.00	0.03	-0.18	-0.03	0.36	-0.04	0.53
NC/P	**G**	-1.15	0.66	-0.08	-0.51	**-1.92**	2.28	1.56	0.00	-0.68	-0.30	0.41	-0.01	0.42	0.67
	P	0.03	-0.08	0.04	0.07	**-0.08**	0.05	0.33	0.00	0.02	-0.16	-0.03	0.45	-0.02	0.64
NPo/C	**G**	-0.92	0.46	-0.07	-0.52	-1.40	**3.12**	1.28	0.01	-1.20	-0.25	0.19	-0.01	0.05	0.73
	P	0.02	-0.05	0.03	0.07	-0.05	**0.09**	0.21	0.00	0.03	-0.12	-0.01	0.33	0.00	0.57
NPo/P	**G**	-1.71	1.01	-0.16	-0.62	-1.61	2.14	**1.86**	0.00	-1.09	-0.50	0.62	-0.01	0.69	0.62
	P	0.05	-0.12	0.08	0.08	-0.07	0.04	**0.41**	0.00	0.03	-0.26	-0.05	0.49	-0.02	0.67
NS/Po	**G**	-0.51	0.33	-0.05	-0.05	0.03	-0.59	0.23	**-0.03**	0.12	-0.21	0.25	0.00	0.22	-0.25
	P	0.01	-0.03	0.01	0.00	0.00	-0.01	0.07	**0.02**	-0.01	-0.10	-0.02	0.04	0.01	0.01
PoL	**G**	1.07	-0.68	0.11	0.44	0.76	-2.20	-1.18	0.00	**1.71**	0.46	-0.44	0.00	-0.24	-0.20
	P	-0.03	0.08	-0.06	-0.06	0.03	-0.05	-0.19	0.00	**-0.06**	0.25	0.04	-0.05	0.02	-0.08
HSW	**G**	1.67	-1.05	0.16	0.49	0.96	-1.29	-1.53	0.01	1.27	**0.61**	-0.65	0.00	-0.77	-0.12
	P	-0.05	0.14	-0.08	-0.06	0.04	-0.03	-0.29	-0.01	-0.04	**0.37**	0.05	-0.16	0.05	-0.08
Oil %	**G**	1.52	-0.99	0.17	0.41	1.03	-0.80	-1.53	0.01	0.98	0.52	**-0.76**	0.00	-0.70	-0.13
	P	-0.05	0.12	-0.08	-0.05	0.04	-0.02	-0.28	-0.01	-0.03	0.28	**0.07**	-0.14	0.04	-0.09
BY/P	**G**	-1.17	0.61	-0.09	-0.56	-1.42	2.04	1.37	0.00	-0.30	-0.18	0.23	**-0.01**	0.45	0.95
	P	0.03	-0.06	0.04	0.07	-0.06	0.04	0.30	0.00	0.00	-0.09	-0.01	**0.66**	0.00	0.94
HI	**G**	1.70	-1.13	0.14	0.40	0.82	-0.16	-1.28	0.01	0.42	0.47	-0.53	0.01	**-1.00**	-0.14
	P	-0.04	0.10	-0.05	-0.05	0.01	0.00	-0.08	0.00	-0.01	0.17	0.03	-0.03	**0.10**	0.15
Residual effect	**G**	0.227	**P**	0.180											

G = Genotypic, P = phenotypic

Abbreviations:-

DFF: Days to 50% flowering, **DM:** Days to maturity, **PH**: Plant height (cm), **NPB/P**: Number of primary branches per plant, **NC/P**: Number of clusters per plant, **NPo/C**: Number of pods per cluster, **NPo/P**: Number of pods per plant, **NS/Po**: Number of seeds per pod, **PoL**: Pod length (cm), **HSW**: Hundred seed weight, **BY/P**: Biological yield per plant (g), **HI**: Harvest index and **SY/P**: Seed yield per plant (g)

4.7Genetic divergence

The most crucial tool in the plant breeder's arsenal for selecting the most suitable kind of parents for a hybridization programme is genetic diversity. The degree of divergence between genotypic and phenotypic populations is measured using the multivariate analysis of Mahalanobis D^2, which may then be used to examine the corresponding role of a variable factor in the total amount of genetic variation existing in a population or species. The D^2 analysis divides genotypes into nearly identical groups using a methodology that maximizes diversity between clusters while minimizing variability within clusters. The corresponding genotypes from various clusters may be used in the breeding programme depending on the breeding objectives.

A collection of 20 soybean genotypes was subjected to D^2 analysis for 14 traits. The findings led to the formation of 5 groups based on D^2 values (Table 4.7).These demonstrated considerable variation of the soybean gene pool that was accessible.

Test with Wilk's Criterion

First, statistically significant differences in the genotypes for individual characters were determined, and then using Wilk's criterion 'Λ'statistical significant differences between the genotypes of all the characters were carried

out. The Wilk's criterion thus obtained was used in calculations of 'V' statistic. When all characters were taken into account simultaneously genotypes differed significantly, as shown by the highly significant 'V' calculated statistic (1902.57) at 266 d.f., which was much more than the table value.

Grouping of genotypes into various clusters

The genotypes were divided into various clusters using Tocher's method (Singh and Chaudhary, 1977). The 20 genotypes were divided into 05 groups using the calculated D^2 values as the squares of the generalized distance. Table 4.7 shows the genotype distribution within these groups. The average intra-cluster and inter-cluster D^2 values showed that the soybean accessible gene pool had significant variety, and mutual relationship between the clusters is represented diagrammatically (figure 4.6).

Table 4.7: Soybean genotypes in various clusters

Cluster no.	No. of genotypes	Names of the Genotypes	
Cluster I	2	G2- Manipur (CAU)	G14- Nagaland (Peren)
Cluster II	4	G6- A.P (Lower Siang)	G13- (JS-9305) M.P
		G16- (JS-9560) M.P	G20- (JS-335) M.P
Cluster III	7	G5- Mizoram (Lawngtlai)	G7- Mizoram (Serchip)
		G9- Nagaland (Dimapur)	G11- Nagaland (Kiphere)
		G10- Nagaland (Kohima)	G17- Nagaland (Wokha)
		G19- (JS-9752) M.P	
Cluster IV	6	G1- Assam (Cachar)	G4- A.P (Lower dibang valley)
		G8-Mizoram (Lunglei)	G12- Nagaland (Tuensang)
		G15- Nagaland (Tuensang)	G18- Nagaland (Zunhebuto)
Cluster V	1	G3- Manipur (Imphal)	

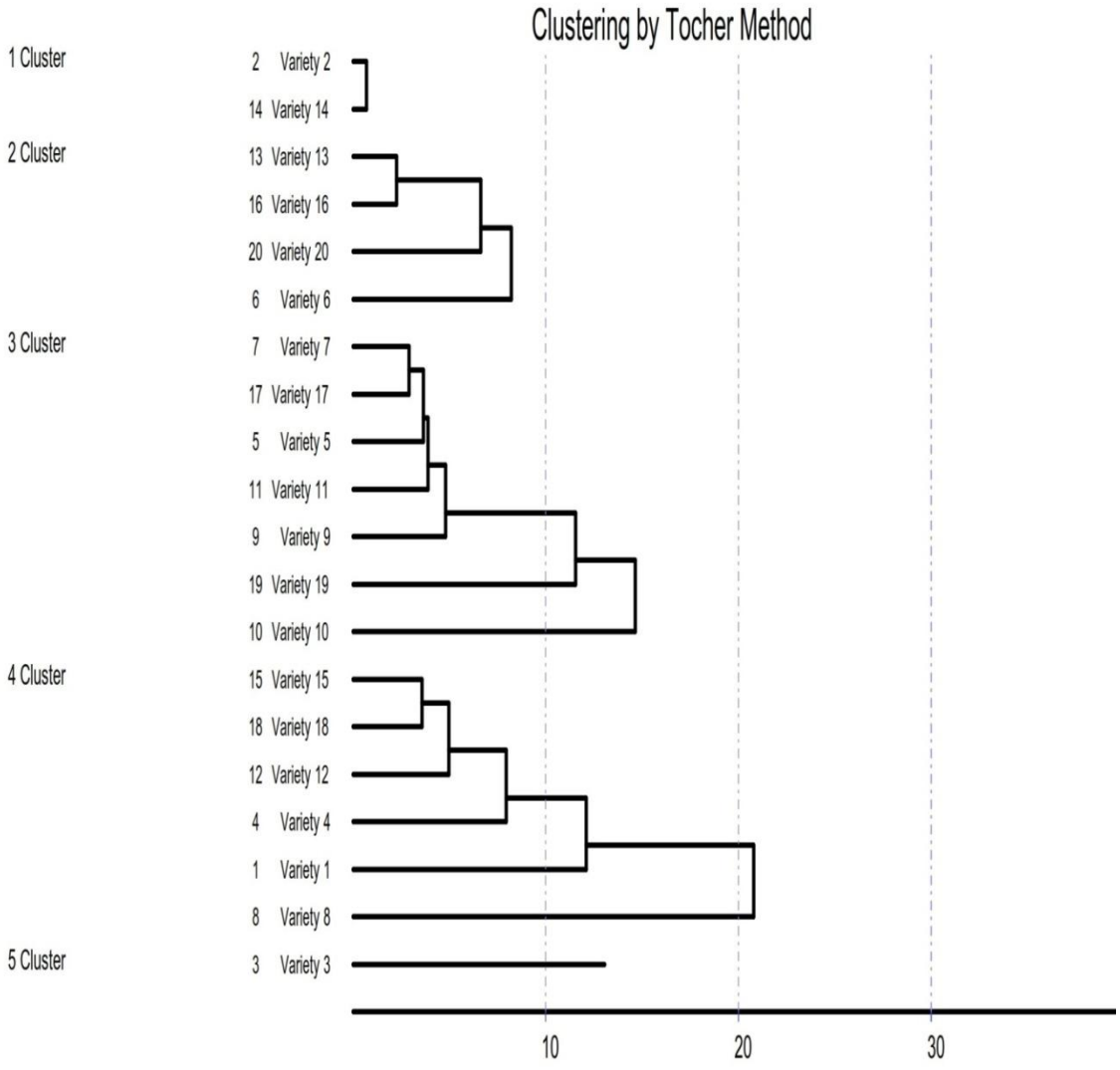


Figure 4.6: Clustering of soybean genotypes by Tocher's method

Cluster analysis recorded the largest Cluster-III (7 genotypes) followed by Cluster-IV (6 genotypes), Cluster-II (4 genotypes), Cluster-I (2 genotypes) and Cluster-V (1 genotype).Soybean genotypes have been found to be genetically diverse based on the clustering pattern. Therefore, the examined genotypes are sufficiently robust for hybridization and selection.

Table 4.7 provides the genotype constellations into various clusters. Genotypes distribution into various clusters was at random and there no correlation between geographic origin and genetic diversity observed since genotypes from various geographic regions were included in the same cluster. The clustering pattern showed that genetic diversity in the current study was not always linked to geographic diversity.

Table 4.8: Per cent contribution of 14 characters for divergence

Sl.No	Source	Times Ranked 1st	Contribution %
1	**Days to 50% flowering**	14	7.37%
2	**Days to maturity**	84	44.21%
3	**Plant height**	6	3.16%
4	**No.of primary branches per plant**	0	0.00%
5	**No.of clusters per plant**	5	2.63%
6	**No.of pods per cluster**	0	0.00%
7	**No.of pods per plant**	0	0.00%
8	**No.of seeds per pod**	0	0.00%
9	**Pod length**	25	13.16%
10	**Hundred seed weight**	49	25.79%
11	**Oil %**	6	3.16%
12	**Biological yield per plant**	0	0.00%
13	**Harvest index**	0	0.00%
14	**Seed yield per plant**	1	0.53%
		Total	100.00%

Percent contribution towards divergence

The sum of square differences between pairs of corresponding uncorrelated values of any two genotypes was taken into consideration to determine the statistical distance (D^2) between a pair of genotypes. There are 190 possible D^2 values since each genotype can make 19 other combinations with every other genotype. Table 4.8 shows the percentage contribution of various characters towards genetic divergence based on these D^2 values. Days to maturity showed maximum contribution (49.47 %) followed by hundred seed weight (24.21%), pod length (8.955), days to 50% flowering (7.89%), equally contributed by plant height and oil content (3.16%); number of clusters per plant and harvest index (1.05%) and number of primary branches per plant and seed yield per plant (0.53%).

Average Intra and inter-cluster D^2 Values

Table 4.9 shows the average intra-cluster and inter-cluster D^2 values calculated using the method described by Singh and Choudhary (1977). The twenty genotypes were categorized into five groups using D^2 analysis. The range of the average intra-cluster distance (mean D^2 value) was 0.00 to 15.71. Cluster IV had the highest intra-cluster distance (D^2=15.71), while cluster V had the least intra-cluster distance (D^2=0.00).

The average inter-cluster distance (average D^2 value) values varied from 26.06 (between cluster II & V) to 340.79 (between cluster I and II). Cluster I and II had the maximum inter-cluster distance (340.79), followed by cluster I and V (243.79), cluster II and IV (194.72), cluster I and III (165.01), and cluster IV and V (117.43), all of which indicated that the genotypes in these groupings were more diverse.

Cluster I in the current study had two genotypes and was the nearest to cluster IV (37.41) and the farthest from cluster II (340.79).

Cluster II, which had four genotypes, was the closest to cluster V (26.06) and extreme to cluster I (340.79). The minimum and maximum values of the distance between the groups were present in this cluster.

With a maximum of seven genotypes, Cluster III was closest to Cluster V (27.79) and farthest from Cluster I. (165.01).

Cluster IV comprised second largest group of six genotypes and was nearest to cluster I (37.41) and furthermost from cluster II (194.72).

Cluster V comprised of single genotype and was near to cluster II (26.06) and furthest to cluster I (243.79).

Table 4.9: Average D^2 and D (parenthesis) values within and between clusters

CLUSTER DISTANCE	Cluster I	Cluster II	Cluster III	Cluster IV	Cluster V
Cluster I	**0.70 (0.84)**	340.79 (18.46)	165.01 (12.85)	37.41 (6.12)	243.79 (15.61)
Cluster II		**9.17 (3.03)**	45.43 (6.74)	194.72 (13.95)	26.06 (5.10)
Cluster III			**10.74 (3.28)**	75.80 (8.71)	27.79 (5.27)
Cluster IV				**15.71 (3.96)**	117.43 (10.84)
Cluster V					**0.00 (0.00)**

Cluster Means

Cluster means for the 14 characters under study are reported in table 4.10 and are as follows:

The recorded mean values for days to 50% blooming ranged from 36.38 for cluster II to 78.67 for cluster I. The range of days to maturity was 94.50 for cluster II and 148.00 for cluster I. The range of plant heights measured was 29.12 cm for cluster II and 74.73 cm for cluster I. A range of 2.42 to 4.01fornumber of primary branches per plant were found for clusters V and IV.

While the number of pods per cluster varied from mean values of 1.15 in cluster V to 2.72 in cluster IV, the number of clusters per plant ranged from 1.42 for cluster V to 8.38 for cluster IV. The range of number of pods per plant was 21.82 (cluster II) to 77.99 (cluster I). Clusters V and I showed a range of 1.64 seeds per pod to 2.33 seeds per pod.

The range of pod lengths was 2.83 cm (cluster I) to 4.05 cm (cluster V). The range of hundred seed weights was 2.96 g for cluster I to 13.47 g for cluster V. Cluster I had an average oil content value of 15.20%, while cluster V had an average oil content value of 22.76. The value of the harvest index was 24.54% for cluster I and 36.47% for cluster II.

Clusters II and IV showed a biological yield/plant range of 16.23 g to 36.42 g. Data on seed yield by cluster analysis ranged from a minimum mean yield of 4.99g/plant for cluster I genotypes to a maximum value of 9.00g/plant for genotypes in cluster IV.

Table 4.10: Cluster mean values for soybean genotypes under study for fourteen characters

Cluster No.	No. of genotypes	DFF	DM	PH	NPB/P	NC/P	NPo/C	NPo/P	NS/Po	PoL	HSW	Oil %	BY/P	HI	SY/P
Cluster I	2	**78.67**	**148.00**	**74.73**	**3.57**	3.24	**2.15**	**77.99**	**2.33**	2.83	2.96	15.20	21.36	24.54	4.99
Cluster II	4	36.38	94.50	29.12	2.43	2.19	1.73	21.82	**2.29**	**3.73**	**12.37**	**22.29**	16.23	**36.47**	5.02
Cluster III	7	49.88	109.24	44.83	**3.49**	**4.79**	**2.59**	**53.74**	2.04	3.12	8.42	**20.65**	**27.91**	**31.40**	**8.12**
Cluster IV	6	**71.00**	**136.50**	**52.54**	**4.01**	**8.38**	**2.72**	**75.76**	**2.26**	3.27	6.44	19.00	36.42	25.74	9.00
Cluster V	1	48.83	107.17	30.53	2.42	1.42	1.15	22.33	1.64	**4.05**	**13.47**	**22.76**	24.08	26.06	5.76

Abbreviations:-

DFF: Days to 50% flowering, **DM:** Days to maturity, **PH**: Plant height (cm), **NPB/P**: Number of primary branches per plant, **NC/P**: Number of clusters per plant, **NPo/C**: Number of pods per cluster, **NPo/P**: Number of pods per plant, **NS/Po**: Number of seeds per pod, **PoL**: Pod length (cm), **HSW**: Hundred seed weight, **BY/P**: Biological yield per plant (g), **HI**: Harvest index and **SY/P**: Seed yield per plant (g).

4.8Evaluation of genetic diversity using SSR molecular markers

In order to distinguish between germplasm with similar morphological traits, DNA fingerprinting techniques based on polymerase chain reaction have emerged as the preferred techniques for diversity studies. One of the most popular markers used to analyze crop genetic diversity and its architecture is SSR, which is a co-dominant marker.SSR markers are now more widely used in the analysis of genetic diversity in soybean due to the availability of suitable software. Under the following headings, the findings of the investigation are discussed:

Qualitative and Quantitative analysis of DNA

DNA was found to be present in average concentrations of 428.27 ng/µl as per the values obtained in spectro-photometric analysis. In the current study, genotype JS 9305 had the highest DNA concentration (769.7 ng/µl), while genotype G2 had the lowest concentration (122.5 ng/µl, Table 4.11).

Table 4.11: Consolidated genomic DNA for 20 genotypes of soybean

Sl. No.	Genotypes	concentration (ng/µl)	A_{260}/A_{280}	Sl. No.	Genotypes	concentration (ng/µl)	A_{260}/A_{280}
1	**G1**	611.9	1.87	11	**G11**	329.2	1.82
2	**G2**	122.5	1.79	12	**G12**	491.8	1.89
3	**G3**	576.2	1.86	13	**G13**	769.7	1.82
4	**G4**	640.2	1.9	14	**G14**	130.5	1.8
5	**G5**	574.2	1.94	15	**G15**	342.7	1.8
6	**G6**	208.2	1.88	16	**G16**	555.7	1.91
7	**G7**	122.6	1.79	17	**G17**	150.6	1.9
8	**G8**	436.9	1.8	18	**G18**	424.2	1.78
9	**G9**	280.9	1.78	19	**G19**	491.8	1.89
10	**G10**	579.3	1.8	20	**G20**	726.4	1.8

This was done solely to determine the amount of DNA present in the sample that would be taken for a subsequent SSR analysis. The mean value of the absorbance ratio of DNA at A_{260}/A_{280}, which varied from 1.77 (T 9) to 1.94 (T 5), was 1.84 (Table 4.11). A_{260}/A_{280} ratio for pure DNA ranges from 1.7 to 1.9. Protein contamination is indicated by a ratio of less than 1.7 and RNA contamination by a ratio greater than 1.9.

Evaluation of SSR primers in discriminating the genotypes studied

In the current study, the genetic diversity of soybean genotypes was analyzed using a total of 25 SSR markers distributed across 20 linkage groups of soybean. It was found that only 18 of the 25 markers undertaken were amplified for scorable loci. A total of 48 alleles were found in the 20 accession of soybeans analyzed, of which 32 were polymorphic and 16 were monomorphic. The allele number for each SSR locus varied from two to three with an average of 1.77 (only polymorphic primers). These 32 alleles had fragment sizes ranging from 70 to 340 base pairs (bp), with Satt 126 having the smallest fragment size and Sat 393 having the largest.

If one of a gene's allele frequencies is less than or equal to 0.95 or 0.99, the gene is said to be polymorphic ($Pj = q \leq 0.95$ or 0.99, where Pj is rate of polymorphism and q = allele frequency). With an average of 0.68 per locus, the major allele frequencies of the polymorphic markers under study varied from 0.45 (Satt 055 and Satt 155) to 0.93 (Satt 557).

The value of polymorphism information content (PIC) reflects the diversity and frequency of alleles among genotypes. In the current study, 20 soybean genotypes with a high rate of polymorphic SSR loci (72.0%) revealed that 18 of the total 25 SSRs analyzed were polymorphic. Table 4.12 contains information on

SSR loci, including allele number, polymorphism percentage, PIC values, allele size range, and main allele frequency. With a mean of 0.587, the PIC value varied between 0.180 (Sat_393) and 0.882 (Satt 055). PIC values ≥ 0.5 are often indicative of informative primers, and primers with higher PIC values are the prime choices for use as molecular markers. Conversely, lower allele counts and PIC values suggest poor allelic diversity. Ten primers were discovered to be 100% polymorphic in the current investigation, including Satt 155,Sat_409, Satt 126, Satt 411, Satt 666, Satt 270, Satt 055, Satt 588, Sat_316 and Sat_196 which were found to be 100% polymorphic.

According on the PCR amplification data, 25 SSR primers produced 608 bands/amplicons. For the 20 genotypes being studied, 290 of the 608 amplified bands were polymorphic and 318 were monomorphic (Table 4.13). Out of 25 primers studied, 18 were polymorphic, contributing an average of 47.70 percent polymorphism for 290 bands, according to the data.

The SSR primers Satt 155,Sat_409, Satt 126, Satt 411, Satt 666, Satt 270, Satt 055, Satt 588, Sat_316 and Sat_196had the highest percentage of polymorphism for amplicons, with 100% polymorphism, while Satt 557 had the lowest percentage of polymorphism for amplicons/polymorphic bands, with 13.04%. In 25 SSR primers analyzed, the average number of polymorphic amplicons was 11.6 while the average number of monomorphic amplicons per SSR locus was 12.72. The high polymorphism result showed that soybean genotypes have a broad genetic base, and this genetic diversity may help explain the variation of morphological and physiological traits.

Table 4.12: Specifics of 25 SSR loci used in the study for twenty soybean treatments

Sl. No.	Linkage group	SSR name	No. of alleles	No. of polymorphic alleles	Polymorhic %	PIC value	Allele size range (bp)	Major Allele frequency
1	A1	Satt 155	3	3	100	0.731	171-200	0.45
2	A2	Sat_409	2	2	100	0.75	160-180	0.55
3	B1	Satt 484	2	1	50	0.424	300-330	0.71
4	B2	Satt 126	2	2	100	0.675	70-100	0.65
5	B2	Satt 687	1	-	-	Monomorphic	167	-
6	C1	Satt 164	2	-	-	Monomorphic	230 and 250	-
7	C1	Satt 396	1	-	-	Monomorphic	200	-
8	C2	Satt 557	2	1	50	0.489	183-233	0.93
9	D1a	Satt 077	1	-	-	Monomorphic	115	-
10	D1b	Sat_227	1	-	-	Monomorphic	250	-
11	D2	Satt 310	3	2	66.67	0.652	200-240	0.85
12	E	Satt 230	1	-	-	Monomorphic	175	-
13	E	Satt 411	2	2	100	0.748	100-140	0.55
14	F	Satt 362	2	1	50	0.48	266-300	0.9
15	G	Satt 163	1	-	-	Monomorphic	235	-
16	H	Sat_218	2	1	50	0.399	250-300	0.8
17	H	Satt 666	2	2	100	0.59	240-260	0.9
18	I	Satt 270	2	2	100	0.688	130-150	0.75
19	J	Sat_393	2	1	50	0.18	300-340	0.6
20	K	Satt 055	3	3	100	0.882	90-160	0.45
21	K	Satt 588	3	3	100	0.875	115-150	0.5
22	L	Sat_286	2	1	50	0.398	157-186	0.71
23	M	Sat_316	2	2	100	0.728	240-270	0.65
24	N	Satt 022	2	1	50	0.289	220-260	0.68
25	O	Sat_196	2	2	100	0.594	200-250	0.63
-	**Total**		48	32	**Range**	0.180-0.882	70-340	0.45-0.93
-	**Average**		-	1.77	78.70%	0.587	-	0.68

Table 4.13: SSR amplicons/bands generated by primers in 20 genotypes of soybean

Sl. No.	SSR name	Total no. of amplicons	Total no. of monomorphic amplicons	Total no. of polymorphic amplicons	Polymorphism %
1	Satt 155	31	-	31	100.00
2	Sat_409	20	-	20	100.00
3	Satt 484	29	19	10	34.48
4	Satt 126	22	-	22	100.00
5	Satt 687	20	20	-	0.00
6	Satt 164	40	40	-	0.00
7	Satt 396	20	20	-	0.00
8	Satt 557	23	20	3	13.04
9	Satt 077	20	20	-	0.00
10	Sat_227	20	20	-	0.00
11	Satt 310	26	20	-	23.08
12	Satt 230	20	20	-	0.00
13	Satt 411	20	-	20	100.00
14	Satt 362	24	20	4	16.67
15	Satt 163	20	20	-	0.00
16	Sat_218	29	20	9	31.03
17	Satt 666	20	-	20	100.00
18	Satt 270	20	-	20	100.00
19	Sat_393	36	20	16	44.44
20	Satt 055	20	-	20	100.00
21	Satt 588	20	-	20	100.00
22	Sat_286	30	19	11	36.67
23	Sat_316	20	-	20	100.00
24	Satt 022	33	20	13	39.39
25	Sat_196	25	-	25	100.00
Total		608	318	290	-
%		-	52.30	47.70	49.55
Average amplicons		-	12.72	11.6	-

Table 4.14: Genetic diversity parameters obtained at each polymorphic locus across 20 soybean genotypes

Sl. No	Locus	Sample size (N)	No. of different Alleles (Na)	No. of Effective Alleles (Ne)	Shannon's Information Index (I)	Observed Heterozygosity (Ho)	Expected Heterozygosity (He)	Fixation Index (F)
1	Satt 155	20	3.00	2.83	1.07	0.55	0.65	0.15
2	Sat_409	20	2.00	1.98	0.69	0.00	0.50	1.00
3	Satt 484	19	2.00	1.70	0.60	0.47	0.41	-0.15
4	Satt 126	20	2.00	1.83	0.65	0.10	0.46	0.78
5	Satt 557	20	2.00	1.16	0.27	0.15	0.14	-0.08
6	Satt 310	20	3.00	1.36	0.53	0.30	0.27	-0.13
7	Satt 411	20	2.00	1.98	0.69	0.00	0.50	1.00
8	Satt 362	20	2.00	1.22	0.33	0.20	0.18	-0.11
9	Sat_218	20	2.00	1.47	0.50	0.40	0.32	-0.25
10	Satt 666	20	2.00	1.22	0.33	0.00	0.18	1.00
11	Satt 270	20	2.00	1.60	0.56	0.00	0.38	1.00
12	Sat_393	20	2.00	1.92	0.67	0.80	0.48	-0.67
13	Satt 055	20	3.00	2.82	1.07	0.00	0.65	1.00
14	Satt 588	20	3.00	2.67	1.04	0.00	0.63	1.00
15	Sat_286	19	2.00	1.70	0.60	0.58	0.41	-0.41
16	Sat_316	20	2.00	1.83	0.65	0.00	0.46	1.00
17	Satt 022	20	2.00	1.78	0.63	0.65	0.44	-0.48
18	Sat_196	20	2.00	1.88	0.66	0.25	0.47	0.47
Total		**358**	**40.00**	**32.96**	**11.52**	**4.45**	**7.49**	**6.12**
Mean		**19.89**	**2.22**	**1.83**	**0.64**	**0.25**	**0.42**	**0.34**
SE		**0.08**	**0.10**	**0.12**	**0.05**	**0.06**	**0.04**	**0.15**

Table 4.15: Jaccard Similarity coefficient values in soybean under study

Genotypes	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G16	G17	G18	G19	G20
G1	1.00																			
G2	0.50	1.00																		
G3	0.52	0.43	1.00																	
G4	0.60	0.56	0.50	1.00																
G5	0.52	0.43	0.54	0.54	1.00															
G6	0.58	0.61	0.63	0.56	0.52	1.00														
G7	0.43	0.48	0.46	0.50	0.46	0.56	1.00													
G8	0.50	0.46	0.52	0.52	0.48	0.54	0.43	1.00												
G9	0.52	0.48	0.61	0.58	0.54	0.52	0.41	0.43	1.00											
G10	0.54	0.58	0.63	0.63	0.56	0.46	0.43	0.58	0.43	1.00										
G11	0.52	0.52	0.61	0.61	0.50	0.66	0.46	0.56	0.54	0.52	1.00									
G12	0.50	0.54	0.52	0.56	0.52	0.61	0.25	0.50	0.48	0.41	0.43	1.00								
G13	0.52	0.60	0.65	0.65	0.50	0.43	0.61	0.63	0.61	0.48	0.58	0.60	1.00							
G14	0.54	0.50	0.56	0.60	0.56	0.61	0.60	0.58	0.60	0.46	0.60	0.58	0.43	1.00						
G15	0.58	0.50	0.48	0.56	0.52	0.58	0.52	0.61	0.60	0.41	0.56	0.50	0.48	0.29	1.00					
G16	0.56	0.56	0.54	0.54	0.46	0.52	0.54	0.63	0.61	0.56	0.58	0.52	0.46	0.56	0.48	1.00				
G17	0.63	0.56	0.41	0.54	0.50	0.63	0.58	0.66	0.71	0.60	0.61	0.56	0.54	0.52	0.43	0.41	1.00			
G18	0.48	0.60	0.54	0.61	0.46	0.56	0.54	0.60	0.65	0.52	0.58	0.52	0.41	0.52	0.56	0.46	0.41	1.00		
G19	0.58	0.58	0.60	0.63	0.48	0.46	0.60	0.61	0.63	0.50	0.60	0.58	0.43	0.58	0.58	0.48	0.48	0.38	1.00	
G20	0.61	0.61	0.63	0.66	0.56	0.50	0.66	0.68	0.63	0.54	0.63	0.65	0.43	0.58	0.58	0.52	0.48	0.48	0.29	1.00

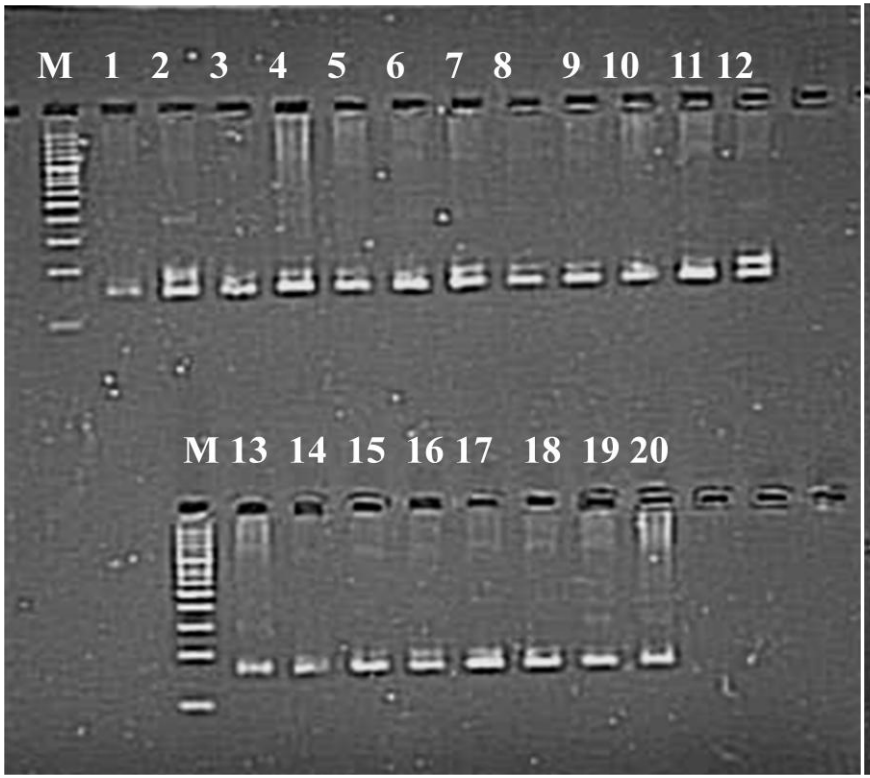


PLATE 2: SSR profiles withprimer Satt 155

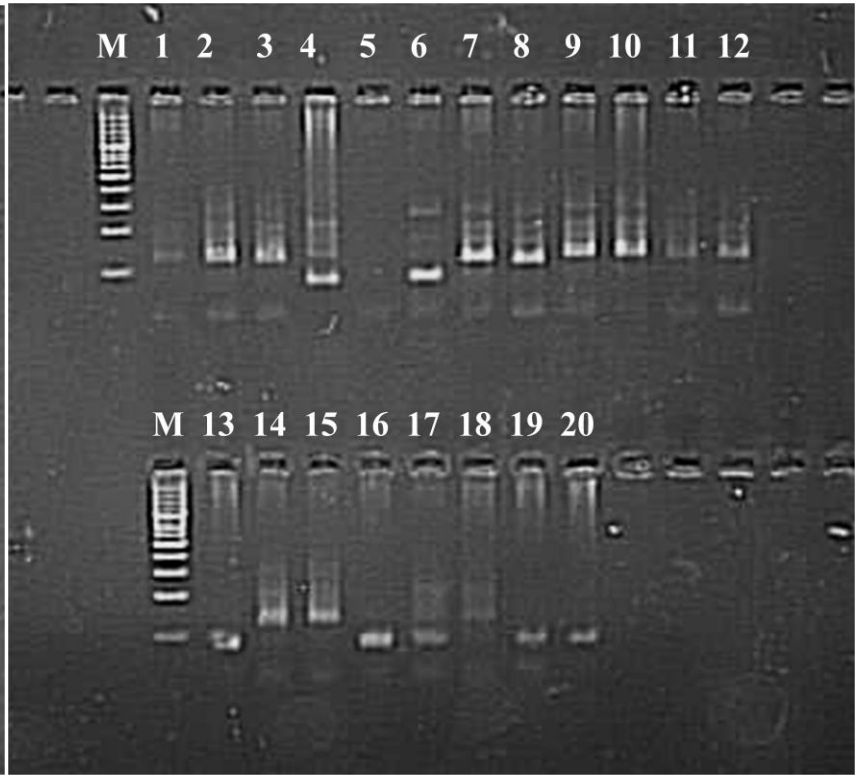


PLATE 3: SSR profiles with primer Satt 411

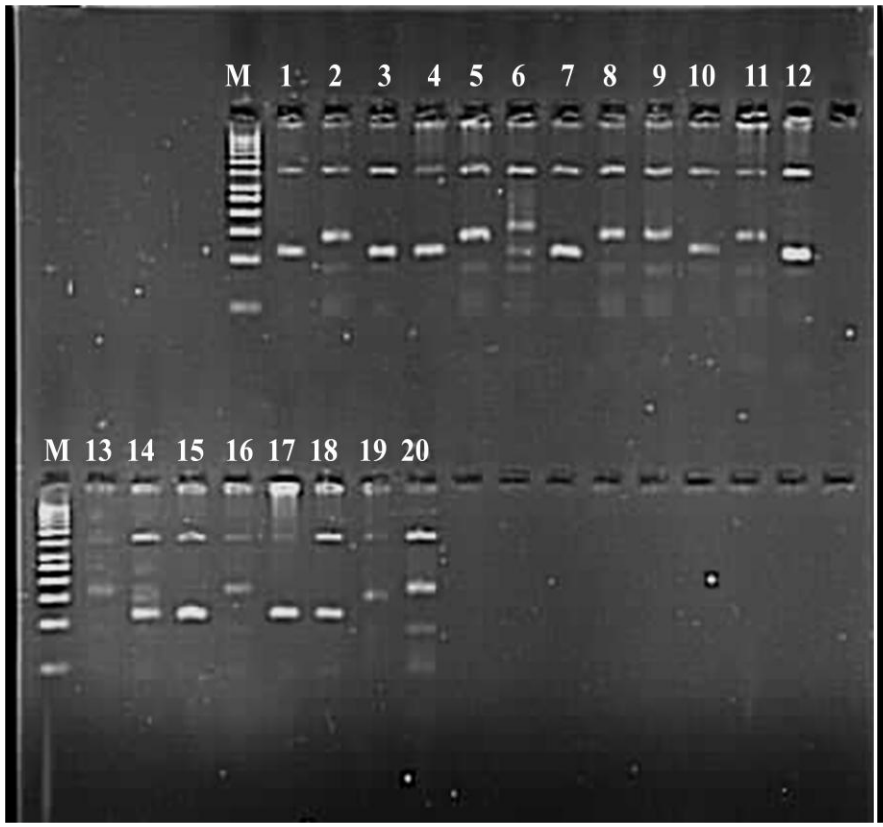


PLATE 4: SSR profiles with primer Satt 588

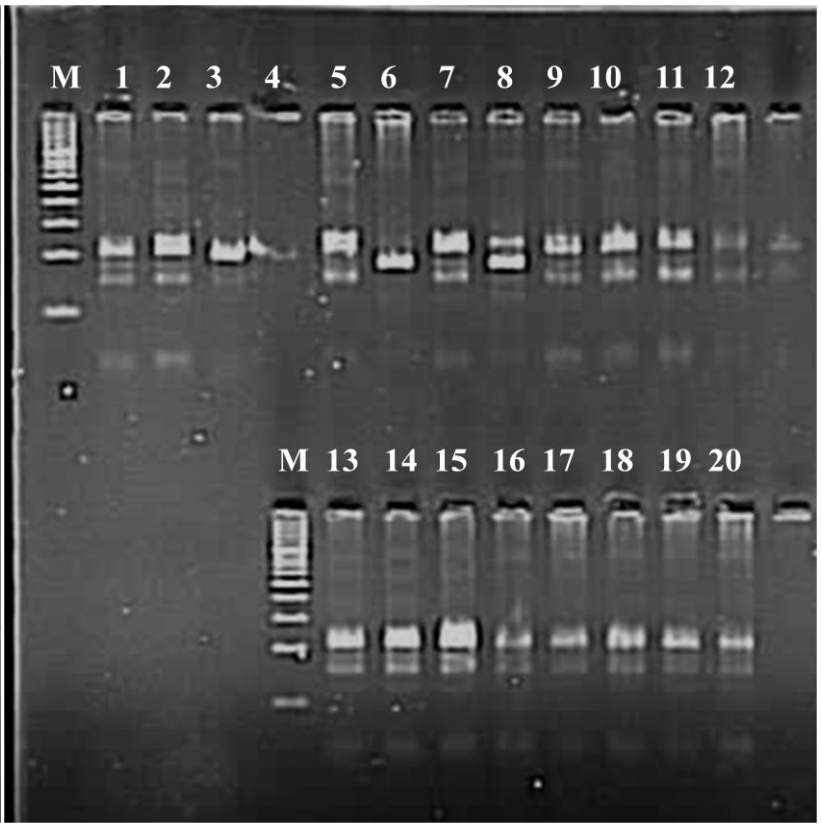


PLATE 5: SSR profiles with primer Satt 196

Genetic diversity parameters revealed by SSR primers

Table 4.14 summarizes the analysis and presentation of the genetic parameters, including No. of different Alleles (Na), No. of Effective Alleles (Ne), Shannon's Information Index (I), Observed Heterozygosity (Ho), Expected Heterozygosity (gene diversity = He) and Fixation Index (F)**.** With an average of 1.83, the effective allele count ranged from 1.16 (Satt 557) to 2.83 (Satt 155). With an average of 0.64 for all the primers used, Shannon's information index ranged from 0.27 (Satt 557) to 1.07 (Satt 155 and Satt 055). The expected heterozygosity (gene diversity) ranged from 0.14 (Satt 557) to 0.65 (Satt 155 and Satt 055), with an average of 0.42, whereas the observed heterozygosity varied from 0.000 to 0.80 with an average of 0.25.Additionally, with an average of 0.34, the fixation index ranged from -0.67 to 1.000. The gel electrophoresis DNA bands amplified by primers Satt 155, Satt 411, Satt 588 and Satt 196 for each of the 20 soybean genotypes are shown in plate 1, 2, 3 and respectively.

Jaccard's similarity coefficient and cluster analysis

For each SSR allele discovered among the 20 accessions, Jaccard's similarity coefficients were calculated in order to evaluate the genetic similarity between the genotypes. The pair wise genetic similarity among genotypes varied from 0.25 (between G7 and T 12) to 0.71 (between G9 and G17).The similarity coefficient has a value between 0 and 1.Table 4.16 contains the average similarity coefficient value for each variety that was determined. The average similarity coefficient value for genotype G4 was 0.60 (maximum), whereas genotypes G5 and G7 had the lowest average similarity coefficient values (0.53).The overall mean genetic similarity coefficient value was 0.56.

Table 4.16: Average similarity Index of 13 soybean genotypes

Genotypes	Average Similarity Value	Genotypes	Average Similarity Value
G1	0.56	**G11**	0.58
G2	0.55	**G12**	0.54
G3	0.57	**G13**	0.55
G4	0.60	**G14**	0.56
G5	0.53	**G15**	0.54
G6	0.58	**G16**	0.55
G7	0.53	**G17**	0.56
G8	0.58	**G18**	0.54
G9	0.58	**G19**	0.55
G10	0.54	**G20**	0.59
Mean value : 0.56			

UPGMA cluster analysis was conducted using the similarity coefficients matrix. Pair wise genetic similarity among 20 soybean accessions ranged from 0.25 to 0.58, according to the SM (simple matching) similarity coefficients, which were used to evaluate the genotypes' genetic similarity. The 20 genotypes formed 2 primary clusters, A and B, according to the dendrogram created based on genetic similarity between genotypes, and an out cluster with a single genotype, G4 (monogenotypic), was obtained. Both the clusters share a similarity of 56% between.

Ten genotypes, designated as G3, G17, G16, G14, G15, G6, G13, G18, G19, and G20, make up Cluster A. This cluster is partitioned further into sub clusters A1 and A2.Sub cluster A1 and A2 shares 53% similarity. A1 comprises of 5 genotypes (includes G6, G13, G18, G19 and G20.) whereas A2 has 5 genotypes (includes G3, G17, G16, G14, G15). This category includes the genotypes G13, G16, G19, G20, which are released varieties. In A1 G6 is out grouped while G13, 18, 19 and 20 are one group.

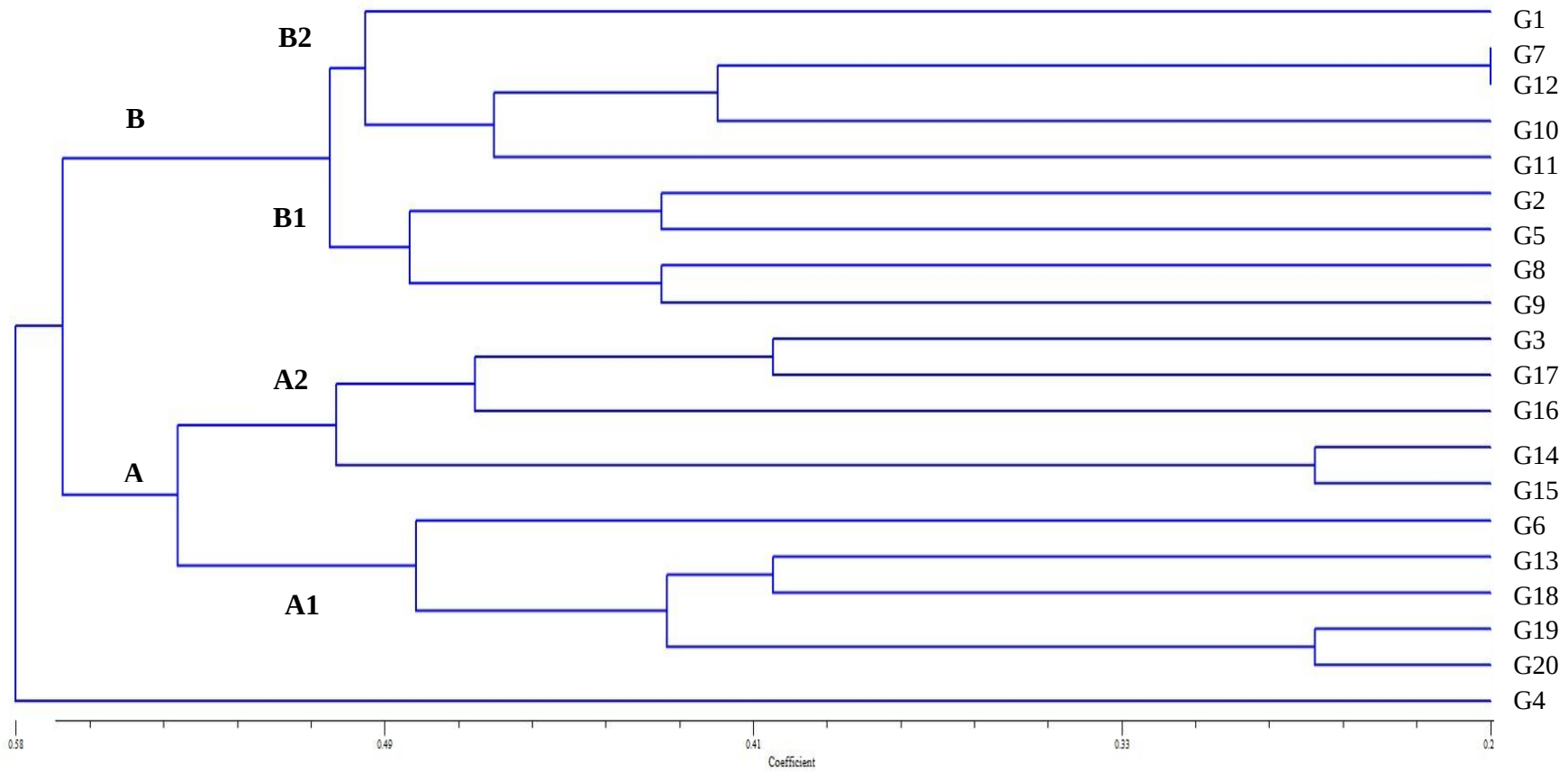


Figure 4.7: Dendrogram showing genetic relationship among 20 soybean genotypes based on UPGMA clustering.

In A2,G-3, G17 and G16 are one group while G-14 and G15 are another.

Cluster B comprises of nine genotypes *viz.*, G1, G7, G12, G10, G11, G2, G5, G8, and G9. This is further sub divided into sub cluster B1 and B2 that shared 50% similarity. In B1 G2 and G5 is one group and G8 and G9 is another. In B2 Ti out groups the cluster G7, G12, G10 and G11 forms one group. The genotypes G7 and G12 were most closely related. Clustering of genotypes based on UPGMA clustering is represented in table 4.17 and figure 4.5 for references in the present study.

Table 4.17: Genotypes under different clusters based on UPGMA clustering

CLUSTER	NO. OF GENOTYPES	SUBCLUSTERS	
		A1 (5 genotypes)	**A2 (5 genotypes)**
A	**10**	**G6**- A.P (Lower Siang) **G13**- (JS-9305) M.P **G18**- Nagaland (Zunhebuto) **G19**- (JS-9752) M.P **G20**- (JS-335) M.P	**G3**- Manipur (Imphal) **G17**- Nagaland (Wokha) **G16**- (JS-9560) M.P **G14**- Nagaland (Peren) **G15**- Nagaland (Tuensang)
		B1 (4 genotypes)	**B2 (5 genotypes)**
B	**9**	**G2**- Manipur (CAU) **G5**- Mizoram (Lawngtlai) **G8**- Mizoram (Lunglei) **G9**- Nagaland (Dimapur)	**G1**- Assam (Cachar) **G7**- Mizoram (Serchip) **G12**- Nagaland (Tuensang) **G10**-Nagaland (Kohima) **G11**- Nagaland (Kiphere)
Monogenotype	**1**	**G4**- A.P (Lower dibang valley)	

DISCUSSION

Yield and characters that contribute to yield are commonly targeted traits in worldwide soybean improvement programmes. The nature and extent of the genetic variability that is accessible, heritability, and the transfer of desired traits into new varieties are the main factors that determine how well a crop can be improved. Additionally, the diversity of plant genetic resources gives plant breeders the chance to create new and improved cultivars with desired qualities, including both farmer and breeder favoured attributes.

Analysis of variance

The existence of genetic variation and the inheritance of desired traits are key factors in crop breeding programmes effectiveness. The breeder can choose suitable strategy and selection criteria to employ for enhancing the target qualities with the help of genetic variation analysis. For all the traits, including days to 50 per cent flowering, days to maturity, plant height (cm), number of primary branches per plant, number of clusters per plant, number of pods per cluster, number of pods per plant, number of seeds per pod, pod length (cm), 100 seed weight (g), oil%, biological yield per plant (g), harvest index (%) and seed yield per plant (g),Pooled analysis of variance showed highly significant differences *i.e.* ($p \leq 0.01$); this indicated large phenotypic variability and as expected, there were inherent genetic differences among the genotypes used in the current study. With the exception of days to maturity, number of seeds per pod, and oil percentage, all the features had a significant mean sum of squares due to genotype x year. The lack of significance in the interaction effect for some measures indicated that the genotypes performance with regard to these traits was constant over the course of the year (Dutta *et al*. 2021).

The results of Baraskar *et al*. (2014), showed significant variability in the soybean crop, support the observations made above. For characters like days to 50 per cent flowering, days to maturity, plant height (cm), number of primary branches per plant,100 seed weight (g), oil% and seed yield per plant (g) Khumukcham *et al.* (2022), Painkra *et al*. (2018) and Chandrawat *et al*. (2017) corroborated the results above .Pawar *et al.* (2020) and Reni and Rao (2013)showed similar significant variability for number of pods per plant and number of seeds per pod. Further studies which supports the aforementioned findings for pod length, biological yield per plant (g) and harvest index were Khumukcham *et al.* (2022) and Reni and Rao (2013).

Mean performance of genotypes

Days to 50 per cent flowering, days to maturity, plant height (cm), number of clusters per plant, , number of pods per plant, 100 seed weight (g), oil%, biological yield per plant (g), harvest index (%) and seed yield per plant (g) all showed a rather broad range of differences in mean values.

Characters with a broad range of variation for majority of genotypes have a good potential for development through simple selection. The present investigation also showed that the majority of the characters had such a broad range. While the range of differences for the remaining characters, including the number of primary branches per plant, the number of pods per cluster, the number of seeds per pod, and the pod length (cm), was rather narrow, indicating less genetic variability for these traits.

Days to 50% flowering can be divided into early (<35 days), medium (35-45 days) and late (>45 days) flowering (Anonymous 2009). Out of twenty genotypes under study, G13 (JS-9305) and G16 (JS-9560) were early, G6 and G20

(JS-335) were medium and rest sixteen including check (JS-9752, 49.50 days) were late blooming genotypes. Gorad (2018) and Shankar (2014) supported above findings for JS-335 with medium flowering (38.67, 36.33 days) while Painkra (2018) supported for JS-9752 (check) with late flowering of 48 days. Also the genotypes ranged between 30-80 days for 50% flowering.

The trait days to maturity could categorize the genotypes into early (<95 days), medium (96-105 days) and late (>105 days), Anonymous (2009). The difference in days to maturity among soybean genotypes is closely related to day length and temperature. Long days generally result in soybean plants with long days to maturity (Liu *et al*. 2017). The range for days under maturity varied from 89 to 148 days. Genotypes under study such as G13 (JS-9305) and G16 (JS-9560) were early, G6, G19 (JS-9752) and G20 (JS-335) were medium and remaining fifteen were late maturing, which included the local genotypes collected from respective areas. The present finding is confirmed with Shankar (2014) for JS-9305 (early); Gorad (2018) for JS-335 (medium) and Painkra (2018) for JS-9752 (medium) respectively.

The tested soybean genotypes had plant heights from short to very tall (27-77cm) and was grouped into eight short (<40 cm) *viz.*, G3, G6, G7, G13, G16, G17, G19 and G20; nine medium (41-60 cm) viz., G4, G5, G8, G9, G10, G11, G12, G15 and G18; and three tall heighted (>60 cm) *viz*., G1, G2 and G14 plants respectively according to DUS guidelines (Anonymous, 2009). The above data obtained was in accordance with Shankar (2014) for JS-9305 i.e. G13 and JS-335 i.e. G20 but was contradicted for JS-9752 i.e. G19 as tall by Painkra (2018).

Number of primary branches per plant ranged between 1-5 branches/plant where twelve genotypes were above grand mean value of 3.39 whereas eight were below. The formation of branches in soybean is also influenced by population

density per unit area, thus the number of branches per plant is not a stable character. A genotype had various branches when it was planted at different population densities (Agudamu *et al*. 2016).

For number of clusters per plant, only seven out of twenty genotypes obtained more values than average mean of 5.03 clusters per plant. However, for G13 (JS-9305), G16 (JS-9560) and G3 (Manipur) pods rarely appear in clusters irrespective of common occurrences in cluster form. Number of pods per cluster had thirteen genotypes above average mean of 2.34 pods.

The trait number of pods per plant varied from 11-115 pods with wide variation. The number of filled pods is a character that determines seed yield per plant. Kuswantoro *et al*. (2019) reported that number of filled pods associated with seed yield, while Machado *et al.* (2017) stated that the number of seeds per pod directly affected seed yield. Top five genotypes with higher number pods were G1>G8>G14>G2>G12 which ranged 115-69 pods which is more than average mean of 55 pods/plant. However, 50% of genotypes were below and 50% were above the mean.

Number of seeds per pod was at a constant value of two seeds per pod with an exception for JS-9305 which had 3seeds/pods. Also the mean value was nearly the same with an average of two seeds per pod. The above results are in contradiction to Shankar (2014) and Gorad (2018) for G20 (JS-335) with three seeds/pods and Painkra *et al*. (2018) for JS-9752 (G19) with three seeds/pod. For pod length nine genotypes showed more value than mean of 3.30 cm while eleven were below.

Hundred seed weight divided the genotypes into three i.e. <10g (low); 10.1-13g (medium) and >13g was high (Anonymous, 2009). Fourteen genotypes were

categorized under low HSW while three (G6, JS-9305 and JS-335) fall under medium. Remaining three viz., G3, G10 and G16 (JS-9560) had high HSW. Gorad (2018) and Painkra*et al.* (2018) resulted similar values for JS-335 (medium) and JS-9752 (low) hundred seed weight. The HSW ranged between 2.82-13.47g.

Fourteen genotypes had more oil content% than mean value of 20.04% while six had less oil%. Generally, seeds with high HSW also contain more oil or vice-versa. Hence G3 with highest HSW also contain maximum oil%. The genotypes varied between 14.17-22.76%.

The data obtained for biological yield per plant showed nine genotypes above grand mean (27.28g) and eleven below it. However, plants with more number of pods have higher biological yield and vice-versa. The genotypes for this character had wide variation of 9.60-63.09g. The harvest index for the study was between 23.17-38.90%. Eight genotypes showed HI% above grand mean and twelve below. Top three genotypes *viz*., G6>G18>G19 (JS-9752) had higher harvest index.

The average amount of seed yield per plant ranged from 2.98 to 16.33g. This useful characteristic aids in the identification of desirable genotypes with desired *per se* performance for yield components and can be employed as potential parents in crop improvement programmes. Top five genotypes were G1 (16.33g)>G10 (12.24g)>G11 (9.12g) >G12 (9.02g) and G9 (8.80g) which were above average mean value of 7.33g/plant. Out of these G10 is also under topmost five genotypes for hundred seed weight. G1 was also highest for number of primary branches/plant, number of pods/cluster, number of pods/plant and biological yield. The oil % of these five genotypes ranged between 17.73-21.70 %. G12 is among top five for number of pods per plant.

Genotypic and phenotypic coefficient of variation (GCV and PCV)

The analysis of variance by itself is not enough and conclusive to explain all the inherent genotypic variance in the collection but can be improved when variability within the accessible germplasm is high, and allows the plant breeder to more rapidly produce new varieties or improve existing ones. Hence, knowledge of key genetic parameters is crucial for any crop improvement program, providing precise information for selection.

Genetic parameters like the genotypic coefficient of variation (GCV), phenotypic coefficient of variation (PCV), environmental coefficient of variation (ECV), heritability and genetic advance as percent of mean (GAM) are valuable biometric tools for measuring genetic variability (Aditya *et al.* 2011). Hence, characterizing the background in terms of genetic parameters of soybean and determining their breeding values should be done before carrying out any improvement programme.

The estimates of the genotypic and phenotypic coefficients of variation obtained demonstrated that the values of PCV were greater than those of GCV, but there was a closer difference between these two estimates in the majority of cases. It showed that there was sufficient genetic variation for these morphological traits and that the manifestation of the studied characters was less influenced by environmental factors that might aid in selection. Similar findings were previously reported by Karnwal and Singh (2009). Therefore, selection based on these characters's phenotypic performance would be a promising approach to significantly improve these parameters.

Few traits such as number of clusters/plant, number of pods/cluster and biological yield/plant had wide differences between PCV and GCV. This result

could be due to the fact that the traits measured in this study were influenced by environmental and other non-genetic factors which would have played some important role in the manifestation of these characters.

For number of clusters per plant, number of pods per plant, hundred seed weight, biological yield per plant and seed yield per plant the high magnitude of GCV and PCV was noted. While Khumukcham*et al*. (2022) and Goonde and Ayana (2021)reported comparable results for number of pods per plant, biological yield per plant, and seed yield per plant, Kumar *et al*. (2020) corroborated the preceding findings for all the traits with the exception of number of clusters per plant. Similar outcomes were observed for the number of clusters per plant by Baraskar *et al.* (2014). According to this, additive gene action plays a significant influence in the expression of these traits. Therefore, one might rely on such character and practise simple selection for subsequent development.

Moderate values for GCV and PCV resulted for days to 50% flowering, plant height and number of primary branches per plant. Tigga (2021) and Baraskar *et al*. (2014) agreed with this finding for number of primary branches per plant. However, Jandong *et al*. (2020) supported above results for days to 50% flowering.

Traits observed low magnitude of GCV and PCV for days to maturity, oil content, pod length and number of seeds per pod. Baraskar *et al*. (2014) supported above results for low GCV for pod length whereas, Khumukcham*et al*. (2022)and Bairwa *et al*. 2020 reported similar findings for days to maturity.

High PCV combined with moderate GCV was observed for number of pods per cluster. Baraskar *et al*. (2014) contradicts the results with moderate PCV and

low GCV. Moderate PCV with low GCV for harvest index was corroborated with findings of Baraskar *et al*. (2014).

Heritability and genetic advance

The amount of variation that is heritable cannot be ascertained just by the genotypic coefficient of variation. Heritability is a measure of how much genetic variation is passed down from parents to all offspring (Lush, 1940).Thus, by knowing a character's heritability, a plant breeder can estimate the genetic advance of any quantitative traits and help implement the required selection technique. In contrast to the heritability value alone, Burton (1952) proposed that the genotypic coefficient of variation in conjunction with the heritability estimate would provide the best overall picture anticipated for selection.

The increase in the mean genotypic value of a chosen plant over its parental population is referred to as genetic advance. Heritability estimates combined with GAM are typically more accurate at forecasting the gain than heritability alone. Majority of the traits examined in the current study had high heritability, which suggests that a sizable portion of the overall variance is under genetic control and that selection based on phenotypic levels would be beneficial for the improvement of these features.

Days to 50% flowering, plant height, number of clusters per plant, number of pods per plant, and 100 seed weight all showed high heritability in combination with high GAM. The results above, except the number of clusters per plant, were supported by Dutta *et al*. (2021) and Joshi *et al*. (2018).Sharma and Lal (2020) reported similar findings for above traits with exception of moderate GAM for days to 50% flowering.Due to the predominance of additive gene action, high heritability and high genetic advance may be caused, and this suggests that

selection in the early generation may be useful for the improvement of these traits. It also suggests that phenotypic level selection might be beneficial. A simple approach, such as mass selection without progeny testing, might easily improve such traits.

High heritability coupled with moderate GAM was recorded for days to maturity, pod length and oil content whereas Dubey *et al*. 2015 supports the above data for days to maturity. In above case where high heritability coupled with moderate genetic advance as percent of mean indicates that gene governing this character is under the influence of dominant effect so one can go for the progeny test or heterosis breeding for the improvement of this character.

Moderate heritability coupled with high GAM was obtained for traits such as number of pods per cluster, biological yield per plant and seed yield per plant. Baraskar *et al*. (2014) supports above findings but contradicts number of pods/cluster with low heritability and GAM. It reveals that the character is governed by additive gene effects and moderate heritability is being exhibited due to environmental effects.

For harvest index and number of primary branches per plant, moderate heritability and moderate GAM were found. The results for both the characters were consistent with those of Baraskar *et al.* (2014).An additive and non-additive gene effect predominates in characteristics with moderate heritability and moderate genetic advance, and such traits may benefit from heterosis breeding.

Moderate heritability of number of seeds per pod as well as low genetic advance as a percentage of mean could be attributed to a greater proportion of non-genetic effects. Selection for such features may not be profitable because the moderate heritability is being exhibited due to the positive influence of

environment rather than genes. This outcome was observed to be consistent with Baraskar *et al.* (2014). Selection in advance generations may be beneficial for improving traits with moderate heritability but low genetic advance.

Correlation analysis

Understanding the nature and degree of relationships between yield and its components is crucial for the simultaneous development of characters, which is likewise essential for efficient yield improvement. To rationally improve the desired traits, one must comprehend the relationship between contributing traits and their proportionate contribution to yield. When examining the genetic basis of association between two traits, Falconer (1960) proposed that complete linkage or pleiotropy may be the cause of the linear association. The overall impact of these genes on both traits is correlation due to linkage or pleiotropy. While some genes may increase both characters simultaneously (positive correlation), others may increase one character while decreasing another (negative correlation).

Seed yield is a complex trait since it is the result of the interaction of numerous parameters known as contributing components. Correlation evaluations thus reveal the nature and degree of link between any two sets of metric characteristics. From this, it could be possible to bring about genetic gradation in one character by selection of other pair. Breeders would be able to select the breeding techniques for improving the genotypes and modifying the unwanted ones by producing new variability by utilizing estimates of valuable correlations. The results of the current investigation showed that genotypic correlation coefficients were higher than their corresponding phenotypic correlation coefficients, showing that the observed correlations were brought about by genetic reasons, such as linkage or pleiotropic effect. It also showed that even though

there was a significant genotypic association between two variables, the impact of environmental factors on phenotypic expression dampened it.

In present investigation, it is evident that seed yield per plant exhibited significant and positive association with number of primary branches per plant (Dubey *et al*. 2015 and Akkamahadevi and Basavaraja, 2017), number of clusters per plant (Neelima *et al.* 2017 supported the findings), number of pods per cluster (Chandel *et al.* 2014), number of pods per plant (Dubey *et al*. 2015;Neelima *et al*. 2017 and Tigga, 2021reported similar association) and biological yield per plant (Baraskar *et al*. 2015 and Dubey *et al*. 2015) both at genotypic and phenotypic level. Such findings indicates, the selection for any one of the above characters would bring in simultaneous improvement of other characters and ultimately improve the seed yield since these characters are mutually correlated among themselves and can be effectively utilized.

The study found that biological yield per plant, number of pods per cluster, number of primary branches per plant, number of clusters per plant and number of pods per plant, in that order, all had positive associations. Additionally, Bhuva *et al.* (2020) showed a highest association between biological yield per plant and yield.

However, number of seeds per pod and pod length showed negative non-significant correlation at genotypic level. Chandel *et al.* 2014 for number of seeds per pod and Shree et *al.* 2018 for pod length reported similar results. Traits such as 100 seed weight (Shekhar *et al*. 2018), oil % (Shree*et al*. 2018) and Bhuva *et al*. (2020) and harvest index in accordance with Yao (1989) were also negative and non-significant both at genotypic and phenotypic level. Alternatively, characters with negative associations and non-significant correlations could be disregarded

when selecting for crop improvement (Henry and Krishna, 1990; Akinyele and Osekita, 2006).

The selection of component traits may be negatively or positively impacted by the interactions between yield-contributing characters. Thus, understanding the relationships between the traits that make up a yield component may make it easier for breeders to choose the degree and direction of selection pressure to apply to related traits in order to simultaneously improve these characters.

Days to 50%flowering (DFF), days to maturity (DM) and plant height (PH) all displayed a significant and positive genotypic correlation with number of primary branches per plant (NPB/P), number of clusters per plant (NC/P), number of pods per plant (NPo/P), and biological yield per plant (BY/P) and also among themselves. This often implies that early maturation follows early blooming and *vice- versa*. Genotypes of soybean with late maturity usually have tall plants. The tallest genotype *i.e*. G2 was last to mature.

The above three traits *viz.*, DFF, DM and PH were negative but significantly correlated at genotypic level with pod length, 100 seed weight, oil content and harvest index. Dubey *et al*. (2015) supported the above outcomes for positive significant correlation between DFF, DM, PH and also with number of primary branches per plant, number of pods per plant, biological yield per plant and negative association with harvest index but contradicts the results with number of seeds per pod. Akkamahadevi and Basavaraja (2017) supports the results of negative association with pod length, 100 seed weight and oil%. In the present study, medium to late maturing genotypes have performed well in comparison to early maturing genotypes, which show an important part played by traits DFF and DM. soybean being *kharif* crop in Nagaland it received 1574.8 mm (2017) and 1140.5 mm (2018, table 3.2), during the study which is much more in

contrast to general water requirement of the crop *i.e*. 450-700 mm (FAO.org). Therefore delayed maturity is preferred in order to complete full life cycle of the crop.

Number of primary branches per plant (NPB/P), number of clusters per plant (NC/P), number of pods per cluster (NPo/C) and number of pods per plant (NPo/P) were positively and significantly correlated at genotypic level with each other and also to biological yield/plant (BY/P). This means that increasing in one character would ultimately increase another one and thereby increase in seed yield. However, NPB/P and NPo/P were negative and correlated significantly to pod length (PoL), hundred seed weight (HSW), oil% and harvest index (HI).Dubey *et al*. (2015) supports above findings for number of primary branches per plant and number of pods per plant positively associated to each other and to biological yield/plant and negatively to harvest index at genotypic level. Study on positive correlation between number of clusters per plant (NC/P) and number of pods per cluster (NPo/C) and also to biological yield/plant along with negatively related to hundred seed weight and oil % was found to be in agreement with Bhuva *et al*. (2020).

Number of seeds/pod (NS/Po) is non-significant to number of pods/plant (Bhuva *et al*. 2020).Negative or non-significant association in the present study indicates that amount of seed doesn't affect pod yields and seed weight. However in case of increment in number of seed per pod might reduce seed size which indirectly effects HSW and oil content.

Pod length had negative association with all the traits under study with an exception to hundred seed weight, oil% and harvest index which summarize to results negative association with seed yield. Chandel *et al.* (2014) reported similar findings for oil% and harvest index but contradicts positive association for

hundred seed weight. This observation in present data obtained revealed that increase in pod length may helps in increase in seed size and indirectly HSW but not necessarily seed number which is non-significant to pod length.

Hundred seed weight and oil% were negatively correlated to all the characters with an exception to pod length and harvest index. Carvalho *et al.* (2002) and Nogueira *et al.* (2012) point out that soybean often promotes compensation in grain size as a function to the number of pods. Thus, it can be explained the lack of correlation between total seed weight and seed yield.

HSW and oil% are positively related to each other. Interestingly, phenotypic and genotypic correlations of 100-seed weight in this study were significantly and negatively correlated with majority of agronomic characters, except for pod length and harvest index. It means that soybean genotypes with large seed sizes had early maturity, short plant, fewer branches, and pods (such as G13 and G16). Bekele and Alemahu (2011) also reported a negative genotypic correlation between 100-seed weight with number of branches. The trend of negative values on phenotypic and genotypic correlations in 100- seed weight was also observed by Machikowa and Laosuwan (2011). The linear correlation also shows a significant negative correlation between 100-seed weight and other agronomic characters (Krisnawati and Adie, 2016; Kuswantoro, 2017).

Biological yield per plant is negatively correlated to harvest index which predicts that increment in one will decrease the other and *vice-versa*. This trend is being in accordance to findings of Dubey *et al*. (2015).

Path coefficient analysis

Through the use of path analysis, breeders can determine whether the relationship between a causal variable and its outcome (seed yield) results from a

direct cause-and-effect relationship or from an indirect one caused by one or more additional characters. As a result, dividing total correlation into direct and indirect effects of cause using the Dewey and Lu, (1959) statistical tool of path coefficient analysis will provide a more meaningful interpretation of the reason of relationship between the variables like yield and independent variables like yield contributing characters. The genotypic path coefficient analysis was used to divide the genotypic correlation coefficients computed for various character pairings into their direct and indirect effects and the results are discussed below:

Direct effect

Number of pods per cluster, followed by number of pods per plant, pod length days to maturity, and 100 seed weight, exhibited the highest positive direct effect in path coefficient analysis at the genotypic level when seed yield per plant was taken into account as the dependent character.

Similar result has been reported for number of pods per plant by Gohil *et al.* (2003); Datta *et al.* (2005); Kumar *et al.* (2005); Gaikwad *et al.* (2007); Malik *et al.* (2007); Baraskar *et al.* (2015); Jain *et al.* (2015); Silva *et al.* (2015) and Dubey *et al.* (2018). The findings were in agreement with Bhuva *et al*. (2020) for number of pods per plant, pod length and days to maturity. Baraskar *et al.* (2015) supported the results for number of pods per cluster whereas Akkamahadevi and Basavaraja (2017) corroborate for 100 seed weight.

However, negative direct effect was contributed by days to 50 per cent flowering (Bhuva *et al*. 2020), number of clusters per plant, harvest index (Narne *et al.* 2002), plant height (Shrivastava *et al.* (2001) ; Chavan *et al.* (2016) and Bhuva *et al*. 2020), number of primary branches per plant (Baraskar *et al.* 2015),

oil%, and negligible values of number of seeds per pod (Baraskar *et al.* 2015) and biological yield per plant (Narne *et al.* 2002).

An overall perusal of the genotypic positive correlation analysis between seed yield/plant and characters such as days to maturity, number of pods per cluster and number of pods per plant was due to direct positive effect of a character, which reveals true relationship between them and direct selection for these traits will be rewarding for yield improvement. The residual effect (0.227) was of moderate magnitude suggesting that besides the characters studied, there are some other attributes which contribute for yield.

Indirect effect

Days to fifty per cent flowering, plant height, number of primary branches per plant and number of clusters/plant showed all had a negative directimpact on seed yield but correlation was manifested *via* number of pods/plant, number of pods/cluster, days to maturity, harvest index and oil% which counteract the indirect negative values of other traits for the four traits mentioned above, namely DFF, PH, NPB/P and NC/P respectively.

The findings agreed with those of Bhuva *et al.* (2020) for indirect effect on the number of primary branches through the number of pods/plant, days to maturity, and harvest index. Indirect effects on days to 50% flowering *via* number of pods/plant, number of pods/cluster, and oil content are supported by Baraskar *et al.* (2015). Number of pods per cluster, days to maturity, and harvest index all had an indirect impact on the number of clusters per plant that was in accordance with the findings of Chandel *et al.* (2014). Dubey *et al.* (2015) reported a positive indirect effect on plant height*via* days to maturity and number of pods per plant. Under the aforementioned conditions, the positive indirect impacts *via*

numerous characters on the direct effects appear to be the cause of the positive association. In such cases, the indirect causal parameters should be simultaneously taking into account for selection.

Given that pod length was non-significant but negatively associated with seed yield and had a direct positive path effect, a restricted simultaneous selection model could be used in this situation. Restrictions would be put in place to eliminate any unfavourable indirect effects so that the direct effect could be utilized. Even though there was no statistically significant correlation between100 seed weight and yield, but direct effect was positive and high therefore, direct selection for this trait should be used to minimize any unfavourable indirect effects.

Traits such as number of seeds per pod, oil% and harvest index were negative and non-significantly correlated to yield along with negative direct path effects. Thus, selections based on above characters will not be fruitful.

Genetic divergence

A significant concern for plant breeders is the selection of suitable parents for use in crop improvement programmes. In order to achieve a wide range of variability in segregating generations and heterosis response in F_1, genetic diversity is thought to be crucial (Arunachalam 1981). In an attempt to identify genetically diverse genotypes for use in breeding programmes, the D^2 analysis aids in characterizing the nature of diversity. More diversified parents within a suitable range increase the likelihood of enhancing the economic characteristics of the resulting offspring under consideration. The Mahalnobis D^2 statistic is a special technique for categorizing parents with diverse genetic backgrounds based on

quantitative traits that could be most effectively used in hybridization programmes.

The genetic divergence analysis was done for all the fourteen characters and the hierarchical cluster analysis of 20 genotypes yielded five clusters at Mahalnobis D^2 analysis and Ward's minimum variance dendrogram with variable number of genotypes. The statistics show that there is significant genetic diversity present among the genotypes. According to Das *et al.* (2000), there is no correlation between the geographical distribution of genotypes and their genetic diversity in terms of grouping patterns of diverse genotypes.

Clustering pattern

The genotypes within each cluster were closer to each other than the genotypes in different clusters. Maximum numbers of genotypes were clubbed in Cluster-III (7genotypes) followed by Cluster-IV (6genotypes), Cluster-II (4genotypes), Cluster-I (2genotypes) and Cluster-V (1genotype). Neelima *et al.* (2017) reported similar findings for cluster V while Jain *et al.* (2017) supported for with maximum genotypes in cluster III followed by cluster IV. The clustering pattern revealed that genotypes from various geographic locations were clustered into a single group, and that genotypes from the same geographic area were sorted into both distinct and the same clusters. JS-335 and JS-9752 were grouped into different clusters which were also verified with the results of Mishra *et al.* (2018) and Dubey *et al.* (2018) who also found out grouping of JS-335 and JS-9305 into same cluster. It shows that there was no apparent relationship between geographic diversity and genetic diversity, indicating that they were not entirely interconnected. This was consistent with earlier findings by Jeethava *et al.* (2000), Jeena and Arora (2002), Reddy *et al.* (2004), Patil *et al.* (2011), Sharma *et al.*

(2012), and Meena *et al*. (2017), which also found that the genetic divergence was not influenced by geographic regions.

Intra and inter-cluster distance

The genotypes which have greater morphological similarity were grouped in clusters (Ghatge and Kadu 1993).Maximum intra-cluster distance was recorded for cluster IV (D^2=15.71) while, minimum intra-cluster distance was observed in clusters V having D^2 value of 0.00 indicating mono-genotypic. It is also valuable considering genotypes within cluster with respect to a trait of interest as suggested by Chahal and Gosal (2002) and Keneni *et al.* (2005).

The inter-cluster distance (D^2) varied from 26.06 to 340.79. The highest inter-cluster distance was recorded between clusters cluster I and II (340.79) followed by cluster I and V (243.79), cluster II and IV (194.72), cluster I and III (165.01) and cluster IV and V (117.43) indicating the presence of greater diversity meriting their consideration in selection for hybridization. It is true that larger the divergence between genotypes, higher would be the heterosis when hybrid development programme is planned to develop yield superior varieties (Bekele *et al.* 2012).

Therefore, in the present study, based upon large inter cluster distances, it is advisable to attempt crossing of the genotypes from clusters II, III, and V with the genotypes of clusters I which may lead to broad spectrum of favourable genetic variability for seed yield improvement in soybean. The lowest inter-cluster distance was recorded between cluster II and V (26.06) indicating that genotypes of these clusters are genetically less diverse and were almost with the same genetic makeup and similar with regards to the characters studied for most of the genotypes in the two clusters. These findings indicate that there was significant

genetic diversity among the genotype under study, as the average inter-cluster distances were higher than the average intra-cluster distances.

Percentage contribution of characters towards total divergence

The key factor influencing the choice of parents is how the characters contribute to the divergence (Bose *et al.* 2011). Days to maturity (49.47%) contributed the most to the expression of genetic divergence, followed by hundred seed weight (24.21%), pod length (8.95), days to 50% flowering (7.89%), plant height and oil content (3.16%), number of clusters per plant and harvest index (1.05%), number of primary branches per plant and seed yield per plant (0.53%).

Maximum diversity was determined by Days to maturity and hundred seed weight. The enhanced diversity in the current materials is a result of these characters, which will provide an excellent opportunity for yield improvement through thoughtful parental genotype selection. These traits can be utilized in the hybridization programme to select parents. The greatest contribution for days to maturity, days to 50% blooming, and hundred seed weight were similarly supported by Kumar *et al*. (2018).The other remaining traits contributed less than 10% to the overall genetic divergence present in the soybean germplasm lines studied. Dubey *et al.* (2018) corroborate the above findings for days to 50% flowering, plant height, number of seeds/pod, oil content, harvest index and seed yield per plant which resulted in less than 10% contribution.

Cluster mean performance

For all the characters included in the study, the cluster mean revealed a wide range of variability. For various characters under consideration, different clusters included distinctive characteristics. According to cluster mean values, clusters I and IV had the following five traits with the maximum cluster means:

days to 50% flowering, days to maturity, plant height (cm), number of pods per plant and number of seeds per pod for cluster I while, number of primary branches per plant, number of clusters per plant, number of pods per cluster, biological yield per plant (g), and seed yield per plant (g) for cluster IV respectively. However, cluster V with single genotype recorded maximum data value for pod length (cm), hundred seed weight and oil% while cluster II observed for harvest index. On the other hand cluster III with maximum genotypes didn't ranged for maximum or minimum values of any traits. Also, early flowering and early maturity and short-heighted plants were observed for genotypes in cluster II. The details are mentioned in table 4.10.

Clusters with desired mean value may be used in hybridization programme to achieve desired yield. It can be inferred that it is not the genetic diversity alone, which decides choice of suitable parents but cluster mean also plays significant role in it. The results of this present investigation reveals that the genotypes of cluster III, were most desirable since their most of the mean values for desired agronomic characters and quality traits are higher than average mean along with good seed yield/plant. This cluster had average cluster mean of 109.24 days for maturity, mean hundred seed weight of 8.12g, average of 49.88 days to 50% flowering, 20.65% oil (>20.04, grand mean value), harvest index mean value of 31.4% and seed yield of 8.12g (>7.33g, grand mean value) as these characters are main contributors towards total genetic divergence. Early flowering and maturity are not considered under this category for north-east (Nagaland) region falls under heavy rainfall over the crop growing season as compared other states with high productivity of soybean such as Madhya Pradesh, Maharashtra and Rajasthan with scanty rainfall in soybean growing areas.

Choice of parents based on cluster mean and *per se* performance

In the current study, 20 diverse genotypes were clustered into different groups, and suitable diverse genotypes were chosen based on their superiority in cluster means and *per se* performance for various characters, as shown in table 4.18.These findings revealed that none of the clusters had genotypes that possessed all the desirable traits and could be directly preferred and utilized. Therefore, the development of desirable genotypes requires the hybridization between genotypes from various clusters.

Choice of parents based on cluster means along with *per se* performance of parents is one of the simplest selection criteria for identifying superior genotypes. The genotypes with high *per se* performance would be much useful as parents for introducing better offspring in any breeding programme. Based on cluster mean and *per se* performance the genotype, G13 (JS-9305) and G16 (JS-9560) grouped in cluster II exhibited earliness in flowering and maturity and are short heighted. Similarly, genotype G1 from Assam in cluster IV was selected for number of primary branches per plant, number of pods per cluster, number of pods per plant, biological yield per plant and seed yield per plant. For high oil and hundred seed weight G3 can be preferred from cluster V.With a significant degree of genetic diversity, clusters I and II had the maximum inter-cluster distance. To develop recombinants with high yields, inter-varietal hybridization programmes (transgressive breeding) may be used with the genotypes from these two clusters. Additionally, while choosing the genetically divergent parents for a future breeding programme, primary consideration ought to be given to the genotypes mentioned above for different traits.

Table 4.18: Diverse genotypes based on inter cluster distances and superior *per se* performance for the traits under investigation

Sl. No.	Characters		Cluster	Suitable genotype in cluster	*Per se* performance
1	**DFF**	Early	II	G13 (JS-9305) and G16 (JS-9560)	32.83 and 34.50
		Late	I	G2	79.50
2	**DM**	Early	II	G16 (JS-9560) and G13 (JS-9305)	89.33 and 91.33
		Late	I	G2, G14	148.00
3	**PH**	short	II	G16 (JS-9560) and G13 (JS-9305)	26.71 and 28.27
		tall	I	G2	77.46
4	**NPB/P**	Min.	II	G16 (JS-9560)	1.06
		Max.	IV	G1	4.64
5	**NC/P**	Min.	II	G13 (JS-9305)	0.31
		Max.	IV	G8	15.72
6	**NPo/C**	Min.	II	G13 (JS-9305)	0.58
		Max.	IV	G1	3.24
7	**NPo/P**	Min.	II	G13 (JS-9305)	10.92
		Max.	IV	G1	115.00
8	**NS/Po**	Min.	V	G3	1.64
		Max.	II	G13 (JS-9305)	2.52
9	**PoL**	Min.	I	G2	2.81
		Max.	II	G13 (JS-9305)	4.09
10	**HSW**	Min.	I	G14	2.82
		Max.	V	G3	13.47
11	**Oil %**	Min.	IV	G8	14.17
		Max.	V	G3	22.76
12	**BY/P**	Min.	II	G13 (JS-9305)	9.6
		Max.	IV	G1	63.09
13	**HI**	Min.	IV	G4	23.17
		Max.	II	G6	38.9
14	**SY/P**	Min.	II	G13 (JS-9305)	2.98
		Max.	IV	G1	16.33

Evaluation of genetic diversity using SSR (Simple Sequence Repeat)

The identification of parental genotypes for the generation of segregating populations and development of varieties is made easier with knowledge of genetic diversity of the germplasm. Genetic diversity based on morphological features is susceptible to environmental fluctuations, and the use of these traits in genetic diversity research has been constrained by the availability of a limited set of morphological markers. On the other hand, genetic diversity based on molecular markers is not affected by environmental variables, thereby making it highly reproducible and explicitly disperse across the genome. An understanding of molecular diversity is required to effectively widen the genetic base of contemporary soybean cultivars.

SSR markers are frequently used in genetic diversity research because to their reliability, reproducibility, and authentic results. Only 18 of the 25 SSR primer pairs used (72%), distributed across 16 linkage groups of soybean (Cregan*et al.* 1999), and produced scorable bands, resulting in the detection of 32 polymorphic alleles in total, with an average of 2–3 alleles per locus. Low allelic diversity/richness was indicated by the lower allele number in the currently assessed set of soybean genotypes.The average number of alleles per locus, or allelic richness, is an useful index for assessing diversity, but depends largely on the sample size (Hipparagi *et al*. 2017),suggesting that more genotypes need to be added to the breeding programme in order to increase genetic diversity or to boost allelic richness (Widaningsih *et al*. 2014).

The average number of alleles obtained was 1.77 in this study. However, the data in 82 indigenous and exotic soybean accessions from different maturity groups and sources in India that was reported by Kumawat *et al*. (2015)

discovered 2.97 alleles per locus with an average polymorphic information content (PIC) value of 0.477.Similar to this, Hipparagi *et al*. (2017) reported 2.61 alleles across 75 soybean genotypes measured by 21 SSR markers in India, with an average PIC value of 0.36. On 38 soybean varieties in an active seed multiplication chain, Bisen *et al*. (2015) used 16 SSR markers to detect 2.22 alleles per locus with an average PIC value of 0.199.

This study's average polymorphic information content (PIC) was 0.587, which was comparable to values from earlier research (Widaningsih *et al*., 2014; Kumawat *et al*., 2015; Ghosh *et al*., 2014;Wang *et al*., 2010).PIC values more than 0.5 were found for a total of 13 markers on distinct linkage groups, demonstrating both the high informativeness of these markers for identifying genotypes and also the dispersal of molecular polymorphism across the genome (Song *et al*. 2010).The SSR primer Satt 055 had the highest PIC (0.882), while the PIC for the SSR primer Sat_393 was the lowest (0.182), showing that primer Satt 055 was a highly useful tool for determining the genetic differences between the released varieties and local soybean genotypes as well as for examining phylogenetic relationships. In a prior work, Kumawat *et al*. (2015) found that the PIC values for the SSR primer Satt 055 and Sat_393 were 0.310 and 0.515, respectively.

A significant amount of genetic variation existed among genotypes, as indicated by the gene diversity/expected heterozygosity (He) ratio, which ranged from 0.14 (Satt 557) to 0.65 (Satt 155 and Satt 055) with an average of 0.42.The study's moderate genetic variation among soybean genotypes indicated the need to expand genetic diversity through selective cross-pollination of foreign with elite locally adapted germplasm.

The gene diversity found in this study was less than that found in earlier studies reported by Widaningsih *et al.* (2014) (0.66), Song *et al*. (2013) (0.65),

Zhao *et al*. (2018) (0.88) and Wang *et al*. (2015) (0.80), further demonstrating the need to introduce more germplasm in the soybean breeding programme.

This study's observed heterozygosity (Ho) was 0.25, which was less than the expected heterozygosity (He). Soybean being predominantly a self-pollinating crop is projected to have lower heterozygosity than other cross-breeding crops; is the reason for low heterozygosity (Zhang *et al*. 2013). According to comparable investigations, Zhao *et al*. (2018) and Hipparagi *et al*. (2017) reported heterozygosity with a value of 0.11 each.

The results obtained by Zhao *et al*. (2018) reported a Shannon's information index (I) of 2.53 and a fixation index of 0.99, which was significantly higher than the one found in this investigation. The Shannon's information index of 0.64 and fixation index of 0.34 in the current study was lower than these results. The results of Shannon's information index and fixation index showed a deviation from Hardy-Weinberg equilibrium, once more showing the presence of moderate genetic variation among the genotypes, which might be explained by the possibility that these genotypes share common paternal lines.

Jaccard’s similarity coefficient ranged from 0.25-0.71 with an average value of 0.56 indicating variations among the individual genotypes under study. This was in dichotomy with the result of Priolli *et al.* (2002), in which12 SSR loci were used to distinguish morphologically similar groups, depicting a mean similarity coefficient of 0.46.

Cluster analysis in the study grouped the genotypes into two major clusters A and B. cluster A comprised of ten genotypes whereas nine were clubbed in cluster B and a genotype G4 out clustered both groups. The genotypes used in this study showed genetic variation because soybean accessions based on the

dendrogram correlated with their place of collection; hence, they belonged to different or same clusters such as genotypes from Manipur (G2 and G3) belonged to different groups while genotypes from Mizoram (G5, G7 and G8) were in group B. Also genotypes from different places of collection grouped into one may be due to duplicacy in germplasm as farmers movement leads to travel of respective seeds. Furthermore, using 21 SSR markers and 75 genotypes, Hipparagi *et al.* (2017) reported three separate clusters, and Hirota *et al.* (2012) identified two distinct clusters. In addition, Tantasawat *et al.* (2011) used SSR markers to identify four large clusters in 25 soybean genotypes, in contrast to two major clusters detected by Wang *et al.* (2006) and Ghosh *et al.* (2014).The results of the present study are consistent with the findings of Wang *et al.*(2006), Wen *et al. (*2008), Hirota *et al*. (2012), Ghosh *et al.* (2014), and Bisen *et al*. (2015) as their findings classified genotypes into two major clusters.

A comparative study of diversity based on morphological and molecular analysis revealed that there is some similarity between the two. Genotypes such as G6, G13, G16, G20; similarly G5, G7, G9, G10, G11 and G1, G8, G12 which are clustered in similar groups in D^2 Mahalanobis method are also found in similar cluster when analyzed using SSR molecular markers. However, genotypes such as G2 and G14 found to be in one group in D^2 were found to be in different clusters based on marker analysis. Also G7 and G12 being most closely related as per SSR data were in different groups when compared with D^2 statistics. The study was carried out to aid breeders and farmers in selecting for desirable traits of the crops while omitting the undesirable ones in a breeding program. Assessment of genetic diversity is important for efficient management and protection of available genetic variability, as well as for crop improvement. The preferred method for breeding is molecular profiling because this method is authentic, reliable, and less affected by

environmental changes. Diversity studies play a major role in categorizing the population into diverse groups, which results in the development of gene pool.

CHAPTER V

SUMMARY AND CONCLUSIONS

SUMMARY AND CONCLUSIONS

The current study entitled “Genetic studies of soybean [*Glycine max* (L.) Merrill] genotypes under Nagaland conditions” which was conducted between kharif seasons 2017 and 2018 at the School of Agricultural Sciences (SAS), Medziphema Campus, Nagaland, experimental farm (Genetics and Plant Breeding) with the subsequent objectives:

- To assess genetic variability of soybean germplasm based on agro-morphological traits.
- To evaluate genetic diversity in soybean germplasm using molecular markers.

To achieve these objectives, it is important to know the variability in the genotypes, heritability of various traits, their patterns of inheritance, the character associations and how they relate to yield. Further, in order to undertake an effective hybridization programme the selection of diverse parents is very important based on the diversity study.In light of the aforementioned perspectives, the current inquiry was created to assess the yield and yield-attributing features of twenty genotypes of soybean. Therefore, the experiment was carried out in Randomized Block Design (RBD) with 3 replications using twenty different soybean genotypes, including the check JS-9752.For morphological and quality traits, observations were made on fourteen distinct characters, *viz*., days to 50% flowering, days to maturity, plant height (cm), number of primary branches per plant, number of clusters per plant, number of pods per cluster, number of pods per plant, number of seeds per pod, pod length, hundred seed weight (g), oil%, biological yield per plant (g), harvest index (%), and seed yield per plant (g).

Five randomly chosen plants from each entry were the basis of data collection for all the traits. Twenty genotypes were evaluated using biometrical analysis, including analysis of variance, coefficient of variation (Burton and Devane, 1953)-phenotypic, genotypic, and environmental coefficient of variation, heritability (broad sense), genetic advance as a percentage of mean, correlation coefficient analysis (phenotypic and genotypic), path coefficient analysis (phenotypic and genotypic), cause-and-effect relationship, and genetic divergence using Tocher's D^2 statistics and evaluation of genetic diversity in soybean germplasm using SSR molecular markers. The key conclusions from this project are summarized below: -

There were very significant differences among all 20 genotypes of soybean for the investigated characters, according to the pooled analysis of variance or estimates of mean sum of square. As a result, it demonstrated that the material (genotypes) analyzed included a significant amount of variability that may be used in future breeding programmes.

With the exception of days to maturity, number of seeds/pods, and oil content, the genotype x year interactions showed significant results for all the characteristics, demonstrating that treatments and genotypes behave independently over years. The small difference between the genotypic and phenotypic coefficients of variance demonstrated that the influence of the environment on the expression of all morphological indices was minimal.

Almost all morphological traits had a relatively wide range of mean value variations, but those that showed the highest genotype-to-genotype variation included number of pods per plant followed by days to maturity, biological yield per plant, plant height, days to 50% flowering, harvest index, number of clusters

per plant, seed yield per plant, hundred seed weight andoil percentage. The remaining four traits had negligible range (<5).

The five top genotypes based on seed yield/plant were G1 (Assam), G10 (Nagaland, kohima), G11 (Ngaland, Kiphere), G12 (Nagaland, Tuensang) and G9 (Nagaland, Dimapur). For the majority of the other traits, such as the number of primary branches per plant, the number of clusters per plant, the number of pods per cluster, the number of pods per plant, and the biological yield, G1 and G12 outperformed others under five best; however, they were medium to tall heighted plants with late blooming and delayed maturity.

Even though early flowering or maturing genotypes may have high oil content, it is not possible to promote them because to their low yield in North-east regions with considerable rainfall. However, in future breeding programmes crossing with early maturing genotypes like G13 (JS-9305) and G16 (Js-9560) as parents could lessen days in late maturity and late flowering genotypes.

However, genotypes G10 and G12 were effective producers with more than 20% oil content, taking soybeans as an important oilseed crop into consideration. The genotype with the best performance and maximum hundred seed weight was G3 (Manipur, CAU). Unusually, G17 (Nagaland, Wokha) had more oil content while having less hundred seed weight. Future breeding programmes can successfully use these genotypes as parents to produce the desired segregants.

Estimates of phenotypic and genotypic coefficients of variation showed that, for majority of the traits, PCV values were higher than GCV values; however the difference was not particularly considerable. This finding may be attributable to the fact that environmental influences did not have a significant impact on the traits assessed in this study. Therefore, selection based on these characters'

phenotypic performance would be an effective approach to significantly improve these features.

The number of clusters per plant, number of pods per plant, hundred seed weight, biological yield per plant, and seed yield per plant showed high genotypic and phenotypic coefficients of variation in the current study, indicating a high degree of genetic variability for these characters and develop future opportunity to choose suitable genotypes.

The presence of variation for traits like days to 50% flowering and plant height is indicated by moderate GCV and PCV values, which coupled with high heritability and genetic advance would further enable improvement through the selection of individual traits.

Heritability when paired with estimates of genetic advance (GA), seem to have more significance. Days to 50% flowering, plant height, number of clusters per plant, number of pods per plant, and 100 seed weight all showed high heritability values with high GA% indicating significant soybean traits that contribute to yield. This suggests that a significant fraction of the total variation is under genetic control and that selection based on phenotypic levels would be beneficial for optimizing these traits.

Moderate heritability and high GAM was obtained for traits such as number of pods per cluster, biological yield per plant and seed yield per plant.This result implies that these traits were mostly governed by additive gene effects, suggesting that selection pressure could profitably be applied for these traits where hybridization followed by progeny selection were likely to be successful. However, where low heritability and genetic advance were noted, there is a need to assemble and acquire a large collection of germplasm.

The best indicator of the degree of improvement that might be anticipated by phenotypic selection could be provided by heritability, GA, and GCV. In order to boost soybean yield, morphological features with high heritability and GA%, as well as high GCV for characters like number of clusters per plant, number of pods per plant, and 100 seed weight, can be regarded to be beneficial. Because there is enough genetic variation and the ability to take advantage of additive gene effects, transgressive segregation, and heterosis to increase yield production, the results of the present study suggested that there is room for improvement of soybean grain yield through effective phenotypic selection.

The results of the correlation investigations showed that genotypic correlation coefficients were higher than their corresponding phenotypic correlation coefficients, showing that the observed associations were brought about by genetic factors like linkage or pleotropic effect.

Seed yield per plant was significantly and positively correlated with traits including biological yield per plant, number of pods per cluster, number of primary branches per plant, number of clusters per plant and number of pods per plant and these characters also had significant and positive inter-correlation among themselves. In order to separate the superior plant genotypes and increase seed yield, selection can therefore be applied for these traits.

Hundred seed weight and oil% were negatively correlated to all the characters with an exception to pod length and harvest index. Therefore, from above study it could be suggested that bold seeds resulted in higher oil content.

By effectively dividing correlation coefficients into unidirectional and alternative pathways, path coefficient analysis enables a critical analysis of the particular components that contribute to a given correlation. The results of

genotypic path analysis indicated that direct relationships between the number of pods per cluster, number of pods per plant, and the number of days to maturity will probably increase seed yield per plant. These parameters have a positive correlation with yield as well.

Five clusters were formed from twenty genotypes; however the distribution of genotypes within each cluster varied. Cluster-III (7genotypes) had the highest number of genotypes followed by Cluster-IV (6genotypes), Cluster-II (4genotypes), Cluster-I (2genotypes) and Cluster-V (1genotype) respectively. It was found that genotypes from different eco-geographical regions were included in clusters with more than one genotype, and that genotypes from the same eco-geographical region were included in different clusters, indicating that geographic diversity does not always correspond to genetic diversity.

Cluster IV had the maximum intra cluster distance, whereas cluster V showed the minimum intra cluster distance (Monogenotype). From 26.06 to 340.79 was the range of the inter-cluster distance (D^2). With a significant degree of genetic diversity, clusters I and II had the maximum inter-cluster distance. To produce recombinants with high yields, intervarietal hybridization programmes (transgressive breeding) may be used with the genotypes from these two clusters.

Based on the cluster mean, cluster I showed maximum mean values for days to 50% flowering, the days to maturity, height of the plant, number of pods per plant, and the number of seeds per pod. Cluster II recorded the highest harvest index mean. The optimal cluster mean for Cluster IV included five traits: biological yield per plant, seed yield per plant, number of primary branches per plant, number of clusters per plant, and number of pods per cluster. However for pod length (cm), hundred seed weight, and oil percentage, clusterV with a single genotype achieved the highest data value. These clusters can be used further in a

hybridization programme to produce desired segregants based on the cluster mean and genotype *per se* performance.

The findings of this analysis showed that the genotypes of clusters III and IV are the most preferred since they have high mean values (more than average mean) for desired agronomic traits and quality features in addition to good seed yields per plant.

Days to maturity had the highest contribution to the manifestation of genetic difference among entries, followed by 100 seed weight, pod length, days to 50% flowering, and plant height. It is important to emphasize this while choosing possible parents for hybridization because these features accounted for 95% of the overall divergence.

The results from molecular diversity study revealed that 13 out of 25 SSRmarkers were highly informative with a PIC value above 0.5. Therefore, they can be widely used in future soybean genomic studies. Best three primers with values of PIC above 0.5 were Sat_409, Satt 055 and Satt 588 which showed 100% polymorphism indicating highly informative to be used for diversity studies in future programmes for characterizing genotypes.

The average number of alleles per locus and gene diversity has indicated the existence of broad genetic base in this collection and significant amount of genetic variation existed among the tested genotypes.

The genotypes in the study were grouped into two major clusters A and B and an out cluster G4 (monogenotype). Cluster A had ten genotypes while cluster B had nine. These were further classified into A1, A2, B1 and B2 sub-clusters respectively. The clusters A and B had 56% genetic similarity.

A comparative study of diversity based on morphological and molecular analysis revealed that there is some similarity between the data of two methods. Genotypes such as G6, G13, G16, G20; G5, G7, G9, G10, G11; and G1, G8, G12, which are clustered together in similar groups in D^2 are also found clubbed together when analyzed using SSR molecular markers. However, genotypes such as G2 and G14 from different places of collection found to be in one group in D^2 were found to be in different clusters based on marker analysis. Also G7 and G12 being most closely related as per SSR data were in different groups when compared with D^2 statistics.

The results obtained from molecular grouping of genotypes, along with field study aids, can be utilized to define heterotic groups, select divergent parents for hybrid breeding as well as in other breeding programmes, in order to exploit the genetic variation that exists in this population.

These findings might be crucial in furthering our understanding of soybean germplasm genetic differentiation. It will also serve as fundamental knowledge by giving breeders alternatives for developing new and more productive varieties that are adaptable to changing environment through selection and breeding. This germplasm might potentially be used in mapping studies and developing genotypes suitable for Nagaland.

REFERENCES

REFERENCES

Abady, S. F., Merkeb and Dilnesaw, Z. 2013. Heritability and path coefficient analysis in soybean genotypes. *Journal Environmental Science and Water Resources*. **2** (8): 270-276.

Adewusi, O. F. 2022. Genetic analysis of seed yield and its contributing characters in F2 population of soybean genotypes, *Glycine max* (L) Merrill using morphological and SNP Markers. *International Journal of Agriculture Technology*. **2** (1): 1-7.

Aditya, J. P., Bhartiya, P. and Bhartiya, A. 2011. Genetic variability, heritability andcharacter association for yield and component characters in soybean (*G. max* (L.)Merrill). *Journal Central European Agriculture*. **12** (1): 27-34.

Agudamu, Yoshihira, T. and Shiraiwa, T. 2016. Branch development responses to planting density and yield stability in soybean cultivars. *Plant Production Science*.**19** (3): 331-339.

Akinyele, B. O. and Osekita, O. S. 2006. Correlation and path coefficient analyses of seed yield attributes in okra [*Abelmoschus esculentus* (L.) Moench]. *African Journal of Biotechnology*.**5** (14): 1330-1338.

Akkamahadevi, S. and Basavaraja, G. T. 2017.Genetic variability and character association among vegetable soybean genotypes for yield and components traits. *Journal of Farm Sciences.* **30** (3): 315-319.

Akram, S., Al Bari, M. A., Hussain N. and Burritt, D. 2016. Genetic variability and association analysis of soybean (*Glycine max* (L.) Merrill) for yield and yield attributing traits. *Plant Gene and Trait*. **7** (13): 1-11.

Al-jibouri, A., Miller, P. A. and Robinson, H. F. (1958). Genotype and environmental Variation and correlation in an upland cotton crop of the interspecific origin. *Agronomy Journal.* **50**: 626-636.

Allard, R.W. 1960. Principles of plant breeding. *John Wiley and Sons Incorporation*, New York.

Alpna, P., Singh, K., Gupta, M. K., Bhareti, P. and Kumar, P. 2015. Morphological characterization and genetic divergence study in soybean. *Soybean Research*. **13** (2): 01-08.

Amogne, A., Atnaf, M. and Bantayehu, M. 2020. Association of traits in soybean (*Glycine max* (L.) Merrill) genotypes in northwestern Ethiopia. *International Journal of Scientific Engineering and Science*. **4 (**10): 8-14.

Anonymous. 2009. Guidelines for the conduct of test for distinctiveness, uniformity and stability (DUS) on soybean [*Glycine max* (L.) Merrill]. *Plant Variety Journal of India*. **3** (10): 289-98.

Anonymous. 2016. Top five soybean growing states of India. Ministry of Agriculture and Farmer's welfare. Government of India. Accessed on 23rd May 2020.

Anonymous. 2017. Statistical Handbook of Nagaland. Directorate of Economic Statistics, Government of Nagaland, Kohima.

Anuradha, B., Aditya, J. P., Singh, G., Gupta, A., Agarwal, P. K. and Bhat, J. C.2011. Assessment of genetic diversity in indigenous and exotic collections ofblack soybean (*Glycine max* (L.) Merrill.). *Journal of Breeding and Genetics.***43** (1): 81-90.

Arunachalam, V. 1981. Genetic distance in plant breeding. *Indian Journal Genetics*. **41**: 226-236.

Athoni, B. K and Basavaraja, G. T. 2012. Association analysis, genetic variability andgenetic diversity studies in soybean [*Glycine max* (L.) Merrill]. *Asian Journal ofBio Science.***7** (2): 114-118.

Azevedo, C. V. G., Val, B. H. P., Araújo, L. C. A. D., Juhász, A. C. P., Mauro, A. O. D. and Unêda-Trevisoli, S. H. 2020. Genetic parameters of soybean populations obtained from crosses between grain and food genotypes. *Acta Scientiarum. Agronomy*.**43**: e46968.

Badkul, A., Shrivastava, A. N., Bisen, R. and Mishra, S. 2014. Study of principal components analyses for yield contributing traits in fixed advanced generations of soybean [*Glycine max (*L.) Merrill]. *Soybean Research*. **2**: 44-50.

Baig, K. S., Kaware, P. P., Sarang, D. H. and Chandrawat, K. S. 2017. Association analysis for yield contributing and morpho physiological traits in soybean. *Soybean Research*. **15** (1): 18-24.

Bairwa, A. K., Shukla, P. S., Singh, K. and Dhaka, N. S. 2020. Study of variability and genetic parameters in soybean germplasm. *International Journal of Current Microbiology and Applied Science.* **9** (03): 978-985.

Balasubramaniyan, P. and Palaniappan, S.P. 2004. Principles and practices of agronomy. Second edition.Agrobios, Jodhpur, India. pp 45-46.

Balla, M. Y. and Ibrahim, S. E. 2017. Genotypic correlation and path coefficient analysis of soybean [*Glycine max* (L.) Merrill] for yield and its components. *Agricultural Research and Technology, Open Access Journal*, pp 2471-6774.

Baraskar, V. V., Kachhadia, V. H., Vachhanl, J. H., Barad, H. R., Patel, M. B. and Darwankar, M. S. 2014. Genetic variability, heritability and genetic advance in soybean (*Glycine max* (L.) Merrill). *Electronic Journal of Plant Breeding*. **5** (4): 802-806.

Baraskar, V. V., Vachhani, J. H., Kachhadia, V. H., Patel, M. B. and Barad, H. R. 2015. Correlation and path analysis for seed yield in soybean [*Glycine max* (L.) Merrill]. *Electronic Journal of Plant Breeding.* **6** (2): 570-573.

Barh, A., Pushpendra, Khulbe, R. K. and Joshi, M. 2014. Bhat: A new source of genetic divergence for soybean improvement. *African Journal of Agricultural Research*. **9** (1): 119-124.

Bekele, A. and Alemahu, G. 2011. Desirable traits influencing grain yield in soybean.*Innovative Systems Design and Engineering*. **2** (3): ISSN 2222-2871.

Bekele, A., Alema, G. W. and Zeleke, H. 2012. Genetic divergence among soybean introductions in Ethiopia based on agronomic traits. *Journal of Biology, Agriculture and Healthcare*. **2** (6): 6–12.

Berhanu, H., Tesso, B. and Lule, D. 2021. Correlation and path coefficient analysis for seed yield and yield related traits in soybean (*Glycine max* (L.)) genotypes. *Plant*. **9** (4): 106-110.

Bhandari, H. R., Bhanu, A. N., Srivastava, K., Singh, M. N., Shreya and Hemantharanjan A. 2017. Assessment of genetic diversity in crop plants - An overview. *Advances in Plantsand Agriculture Research*. **7** (3): 279-286.

Bhuva, R. B., Babariya, C. A., Movaliya, H. M., Gadhiya, J. A., and Balar, V. S. 2020. Correlation and Path Analysis for Seed Yield in Soybean [*Glycine max* (L.) Merrill]. *Indian Journal of Pure and Applied Bioscience.* **8** (4): 375-380.

Bisen, A., Khare, D., Nair, P. and Tripathi, N. 2015. SSR analysis of 38 genotypes of soybean (*Glycin max* (L.) Merrill) genetic diversity in India. *Physiology and Molecular Biology Plants*. **21**:109-15.

Bose, L. K., Singh, O. N., Subudhi, H. N. and Rao, G. J. N. 2011. Genetic diversity in direct seeded Aerobic rice. *International Journal of agriculture Sciences*. **7** (2): 321-324

Botstein, D., White, R. L., Skolnick, M. and Davis, R. W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *The American Journal of Human Genetics*. **32**: 314-331.

Burton, G. W. 1952. Quantitative inheritance in grasses. Proceedings of theSixth International Congress.**1**: 277-293.

Burton, G.W. and Devane. 1953. Estimating heritability in tall fescue (*Festucaarundinacea*) from replicated clonal material. *Agronomical Journal*. **45**: 478-481.

Carvalho, C. G. P., Arias, C. C. A., Toledo, J. F. F., Oliveira, M. F. and Vello, N. A. 2002. Correlations and path analysis in soybean lines sowed at different sowing dates. *Brazilian Journal of Agricultural Research.* **37** (3): 311-320.

Chahal, G. S. and Gosal, S. S. 2002. Principle and procedures of plant breeding biotechnological and conventional approaches. Narosa Publishing House, New Delhi.

Chandel, K .K., Patel, N. B., Sharma, L. K. and Gali, S. 2017. Genetic variability, correlation coefficient and path analysis for yield and yield attributing characters in soybean (*Glycine max* L. Merrill). *Green Farming*. **8** (3): 547-551.

Chandel, K. K., Patel, N. B. and Patel, J. B. 2014. Correlation coefficients and path analysis in soybean (*Glycine max* L. Merrill). *AGRES –International e-Journal.* **3** (1): 25-31.

Chandrawat, K. S., Baig, K. S., Hashmi, S., Sarang, D. H., Kumar, A. and Dumai K. 2017. Study on genetic variability, heritability and genetic advance in soybean. *International Journal of Pure and Applied Bioscience*. **5** (1): 57-63.

Chandrawat, K. S., Baig, K. S., Sarang, D. H., Dumai, P. H. K., Dhone, P. U. and Kumar, A. 2015. Association analysis for yield contributing and quality parameters in soybean. *International Journal of Environmental Science*. **7**: 113-118.

Chaudhary, L. B. and Prasad, B. 1968. Genetic variation and heritability of quantitative characters in Indian mustard (*Brassica juncea*). *Indian journal of Agricultural science*.**124** (1): 48-55.

Chavan, B. H., Dahat, D. V., Rajput, H. J., Deshmukh, M. P. and Diwane, S .L. 2016. Correlation and path analysis in soybean. *International Multidisciplinary Research Journal.***2** (9): 1-5.

Chen, K. I., Erh, M. H., Su, N. W., Liu, W. H., Chou, C. C. and Cheng, K. C. 2012. Soy foods and soybean products: From traditional use to modern applications. *Applied microbiology and biotechnology*. **96**: 9–22.

Cregan, P.B., Jarvik, T. and Bush, A. L. 1999. An integrated genetic linkage map of the soybean genome. *Crop Science*. **39** (5): 1464-1490.

Das, S. P., Harer, P.N and Biradar, A. B. 2000. Genetic divergence and selection of genotypes in soybean. *Journal of Maharshtra Agricultural Universities.***25** (3): 250-252.

Datt, S., Sharma, P. R., Kumar, M. and Gupta, A. K. 2011. Genetic variability and trait relationships among yield and other quantitative traits in soybean (*Glycine max* (L.) Merrill). *Vegetos*. **24** (2): 117-120.

Datta, A. K., Hossain, M. A. and Rahman, L. 2005. Genetic parameters, inter-relationship and path coefficient analysis in soybean [*Glycine max* (L.) Merrill]. *Journal of Bangladesh Agrilcutral University.* **3** (2): 225-230.

Deshmukh, M. 2016. Correlation and path analysis in soybean. *International Multidisciplinary Research Journal.***2**: 1-5.

Dewey, D. R. and Lu, K. H. 1959. A correlation and path coefficient analysis of components of crested wheat, grass, seed production. *Agronomy Journal.* **47**: 477-483.

Dhapke, S. K., Patil, S. P. and Wankhade, R. S. 2011. Studies on genetic diversity inelite soybean genotypes. *Crop Research, Hisar*. **42** (3): 255-258.

Dong, D., Fu, X., Yuan, F., Chen, P., Zhu, S., Li, B., Yang, Q., Yu, X. and Zhu, D. 2013. Genetic diversity and population structure of vegetable soybean (*Glycine max* (L.) Merrill) in China as revealed by SSR markers. *Genetic Resources and Crop Evolution.* **61**: 173-183.

Dubey, N., Avinashe, H. A. and Shrivasthava, A. N. 2018. Genetic parameters and character association studies of soybean genotypes in Madhya Pradesh region. *Annals of Biology.* **34** (2): 207-211.

Dubey, N., Shrivasthava, A. N., Avinashe, H. A. and Jaiwar, S. 2015. Genetic variability, correlation and path analysis for yield and yield contributing characters in soybean (*Glycine max* L.). *Electronic Journal of Plant Breeding*. **6** (1): 318-325.

Dupare, B. U. and Billore, S. D. 2016. Soybean: Package of practices for crop management. Extension bulletin no: 13. ICAR- Indian Institute of Soybean Research. *Edition* 2016. Pp:1-54.

Dutta, P., Goswami P. K. and Borah, M. 2021. Assessment of genetic variability, heritability and genetic advance in soybean genotypes. *Electronic Journal of PlantBreeding.* **12** (4): 1461-1465.

Ekka, N. P. and Lal, G. M. 2016. Study on genetic variability and character association in soybean [*Glycine max* (L.) Merrill] germplasm at vindhyan zone of Uttar Pradesh. *Agricultural Science Digest- A Research Journal*. **36** (1): 69-71.

El-Mouhamady, A. B. A. and El-Metwally, M. A. 2020. Appreciation of genetic parameters and molecular characterization in some promising accessions of soybean (*Glycine max* L.). *Pakistan Journal of Biological Science*. **23** (4): 425-438.

Falconer, D. S. (1960). Introduction to quantitative genetic. Oilver and Boyd Ltd. Edinburgh and London.

Fisher, R. A. and Yates, A. H. 1938. Statistical tables for biological, agricultural and medical research. Olivar and Boyd. Edinburgh. 134-135.

Food and agriculture organisation: https://www.fao.org/3/t0532e/t0532e02.htm

Gaikwad, S. R. Bangar, N. D and Chavan, B. H. 2007. Correlation and path coefficientanalysis in soybean. *Journal of Maharashtra Agricultural Universities.* **32** (2):276-277.

García, M.C., Marina, M., Laborda, F. and Torre, M. 1998. Chemical characterization of commercial soybean products. *Food Chemistry*. **62** (3): 325–331.

Geetanjali. 2020. Genetic study in recombinant inbred lines population for earliness in soybean (*Glycine max* (L.) Merrill.). M.Sc. Thesis, IndiraGandhi Krishi Vishwavidyalaya, Raipur, pp 42.

Getnet, B. E. 2019. Genetic variability, heritability and expected genetic advance as indices for selection in soybean [*Glycine max* (L.) Merrill] varieties. *American Journal of Life Sciences.***6** (4): 52-56.

Ghatge, R. D., and Kadu, R. N. 1993. Genetic diversity in soybean. *Annals of Agricultural Research*. **14** (2): 143–148.

Ghiday, T., Amogne, A., Tefera, G. and Malede, M. 2017. Heritability, genetic advance and path coefficient analysis for grain yield and its component characters in soybean (*Glycine max* L. Merrill.). *International Journal of Research Studies in Agricultural Sciences.***3** (5): 1-11.

Ghiday, T., and Sentayehu, A. 2015. Genetic divergence analysis on some soybean (*Glycine max* L. Merrill) genotypes grown in Pawe, Ethiopia. *American Eurasian Journal of Agriculture and Environmental Science*. **15** (10): 1927-1933.

Ghodrati, G. H. 2013. Study of genetic variation and broad sense heritability for some qualitative and quantitative traits in soybean (*Glycine max* L.) genotypes. *Current Opinion in Agriculture.***2** (1): 31–35.

Ghosh, J., Ghosh, P. D., Choudhury, P. R. 2014. An assessment of genetic relatedness between soybean [*Glycine max* (L.) Merrill] cultivars using SSR markers. *American Journal of Plant Sciences*. **05** (20): 3089-3096.

Gohil, V. N., Pandya, H. M. and Mehta, D. R. 2003. Genetic association and path analysis for seed yield, its components and quality attributes in soybean. *Gujarat Agricultural University Research Journal.* **28** (1-20): 9-11.

Goonde, D. B. and Ayana, N. G. 2021. Genetic diversity and character association for yield and yield related traits in soybean (*Glycine max* L.) genotypes. *Journal of Agricultural Science and Food Research*. **12** (1): 280.

Gorad, T. C. 2018. Evaluation of genotypes to study quantitative traits in soybean [*Glycine max* (L.)] during summer. M.Sc. Thesis Mahatma Phule Krishi Vidyapeeth, Rahuri, pp 23-48.

Guan, R., Chang, R., Li, Y., Wang, L., Liu, Z. and Qiu, L. 2010. Genetic diversity comparison between Chinese and Japanese soybeans (*Glycine max* (L.) Merrill) revealed by nuclear SSRs. *Genetic Resources and Crop Evolution.* **57**: 229-242.

Hang Vu, T. T., Cham Le, T. T., Vu, D. H., Nguyen, T. T. and Ngoc, T. 2019. Correlations and path coefficients for yield related traits in soybean progenies. *Asian Journal of Crop Science*. **11**: 32-39.

Hanson, W. D. 1961. Statistical genetics and plant breeding. *National Academy of Science, National Research Council, Washington,* pp.125-140.

Henry, A. and Krishna, G. V. 1990. Correlation and path coefficient analysis in pigeon pea. *The Madras Agricultural Journal.***77** (9-12): 443-446.

Hipparagi, Y., Singh, R., Choudhary, D. R. and Gupta, V. 2017. Genetic diversity and population structure analysis of Kala bhat (*Glycine max* (L.) Merrill) genotypes using SSR markers. *Hereditas*. **154** (9): 1-11.

Hirota, T., Sayama, T., Yamasaki, M., Sasama, H., Sugimoto, T., Ishimoto, M. and Yoshida, S. 2012. Diversity and population structure of black soybean landraces

originating from Tanba and neighboring regions. *Breeding Science*. **61** (5): 593-601.

Hosamani, J., Kumar, M. A., Talukdar, A., Lal, S. K. and Dadlani, M. 2013. Molecular characterization and identification of candidate markers for seed longevity in soybean [*Glycine max* (L.) Merill]. *Indian Journal of Genetics and Plant Breeding*. **73** (1): 64-71.

Ibrahim, A. K., Dawaki, K. D., and Hassan, S. M. 2018. Genetic variability, heritability and correlation among soybean [*Glycine max*. (L.) Merrill] varieties. *Bayero Journal of Pure and Applied Sciences.***11** (2): 72 – 80.

Iqbal, Z., Arshad, M., Ashraf, M., Naeem, R., Malik, M. F. and Waheed, A. 2010. Genetic divergence and correlation studies of soybean (*Glycine max* (L.) Merrill) genotypes. *Pakistan Journal of Botany*. **42** (2): 971-976.

Irshad, M., Farhatullah, A., Khan, N., and Begum, S. 2016. Genetic variability, heritability and genotype x year interaction for morpho-yield traits in soybean. *Pakistan Journal of Botany*. **48** (4): 1601-1607.

Jaccard, P. 1908. Nouvelles researches sur la distribution florale. *Bulletin de la Société vaudoise des sciences naturelles*. **44**: 223-270.

Jain, R. K., Joshi, A., Chaudhary, H. M., Dashora, A. and Khatik, C. L. 2017. Study on genetic variability, heritability and genetic advance in soybean [*Glycine max* (L.) Merrill], *Legume Research*. **41** (4): 532-536.

Jain, S., Srivastava, S. C., Sing, S. K., Indapurkar, Y. M. and Singh, B. K. 2015. Studies on genetic variability character association and path analysis for yield and its contributing traits in soybean (*Glycine max* (L.) Merrill). *Legume Research*. **38** (2): 182-184.

Jandong, E. A., Uguru, M. I. and Okechukwu, E. C. 2020. Estimates of genetic variability, heritability and genetic advance for agronomic and yield traits in soybean (*Glycine max* L.). *African journal of biotechnology*. **19** (4): 201-206.

Jeena, A. S. and Arora, P. P. 2002. Multivariate technique in chickpea. *Agriculture Science Digest.* **22**: 57-58.

Jeethava, A. S., Yasufzal, A. S., Poshiya, V.K. and Vaddoria, M.A. 2000. Divergence analysis in chickpea. *Gujrat Agriculture University Research Journal.* **22**: 23-28.

Jo, H., Lee, J. Y., Cho, H., Choi, H. J., Son, C. K., Bae, J. S., and Lee, J. D. 2021. Genetic diversity of soybeans (*Glycine max* (L.) Merrill.) with black seed coats and green cotyledons in Korean germplasm. *Agronomy*. **11**(3): 581.

Johnson, H. W., Robinson, H. F and Comstock, R. E. 1955. Estimates of genetic and environmental variability in soybean. *Agronomy Journal*. **47** (2): 314-318.

Joshi, D., Singh, P. K. and Adhikari, S. 2018. Study of genetic parameters in soybean germplasm based on yield and yield contributing traits. *International Journal of Current Microbiology and Applied Sciences*. **7** (1): 700-709.

Kachhadia, V. H., Baraskar, V. V., Vachhani, J. H., Patel, M. B. and Barad, H. R. 2014. Genetic divergence in soybean [*Glycine max* (L.) Merrill.]. *Electronic Journal of Plant Breeding*. **5** (3): 563-566.

Karikari, B., Bhat, J. A., Denwar, N. N. and Zhao, T. 2020. Exploring the genetic base of the soybean germplasm from Africa, America and Asia as well as mining of beneficial allele for flowering and seed weight. *Biotechnology*. **10** (5): 195.

Karnwal, M. K. and Singh K. 2009. Studies on genetic variability, character association and path coefficient for seed yield and its contributing traits in soybean[*Glycine max* (L.) Merrill], *Legume Research.***32**: 70-73.

Keneni, G., Jarso, M., Wolabu, T., and Dino, G. 2005. Extent and pattern of genetic diversity from morpho-agronomic traits in Ethiopian Filed pea landraces. *Genetic Resources and Crop Evolution.* **25**: 539-549.

Khumukcham, J., Devi, K. T., Nag, S. and Shrivastava, R. 2022. Genetic variability analysis in soybean genotypes based on yield and yield contributing traits. *PharmaInnovation*. **11** (7): 3233-3236.

Krisnawati, A. and Adie, M. M. 2016. Hubungan antarkomponen morfologi dengan karakter hasil biji kedelai. *Buletin Palawija* **14** (2): 49-54.

Krisnawati, A. and Adie, M. M. 2019. Genetic variability of soybean (*Glycine max* (L.) Merrill) genotypes for pod shattering resistance. *Conference Series Earth Environmental Science*. **293** (01): 2003.

Kudełka, W., Kowalska, M. and Popis, M. 2021. Quality of soybean products in terms of essential amino acids composition. *Molecules*.**26**: 5071.

Kujane, K., Sedibe, M. M. and Mofokeng, A. 2019. Genetic diversity analysis of soybean (*Glycine max* (L.) Merrill) genotypes making use of SSR markers. *Australian Journal of Crop Science*. **13** (07): 1113-1119.

Kumar, A. P., Ramana, M. V., Sultana, R. and Rao, A. S. 2005. Character association and path analysis in soybean. *The Andhra Agicultural. Journal.* **52** (1-2): 48-51.

Kumar, A., Lal G. M. and Mishra P. K. 2013. Genetic variability and character association for yield and its components in soybean. *Annals of Plant and Soil Research.***16** (1): 48-52.

Kumar, S. P. J., Susmita, C. and Sripathy, K. V. 2022. Molecular characterization and genetic diversity studies of Indian soybean (*Glycine max* (L.) Merrill) cultivars using SSR markers. *Molecular Biology Reports.* **49**: 2129–2140.

Kumar, S., Kumari, V. and Kumar, V. 2018. Assessment of genetic diversity in soybean[*Glycine max* (L.) Merrill] germplasm under North-Western Himalayas.*Journal of Pharmacognosy and Phytochemistry*.**7** (2): 2567-2570.

Kumar, S., Kumari, V., and Kumar, V. 2020. Genetic variability and character association studies for seed yield and component characters in soybean [*Glycinemax* (L.) Merrill] under North-western Himalayas. *Legume Research-An International Journal*. **43** (4): 507-511.

Kumawat, G., Singh, G., Gireesh, C., Shivakumar, M., Arya, M., Agarwal, D. K. and Husain, S. M. 2015. Molecular characterization and genetic diversity analysis of soybean (*Glycine max* (L.) Merrill) germplasm accessions in India. *Physiology and Molecular Biology of Plants.* **21** (1): 101–107.

Kuswantoro, H. 2017. The role of heritability and genetic variability in estimated selection response of soybean lines on tidal swamp land. *Pertanika Journal of Tropical Agricultural Sciences.* **40** (2): 319-328.

Kuswantoro, H. 2019. Performance, similarity and genetic parameters of agronomical characters of soybean [*Glycine max* (L) Merrill.] germplasms. *Agriculture and Natural Resources*. **53** (3) 228–236.

Kuswantoro, H., Artari, R., Rahajeng, W., Ginting, E. and Supeno, A. 2018. Genetic variability, heritability, and correlation of some agronomical characters of soybean varieties. *Biosaintifika: Journal of Biology and Biology Education*. **10** (1): 9-15.

Kuswantoro, H., Muchlish, M. A. and Haksiwi, P. P. 2021. Genetic variability, heritability, and genotypic correlation of soybean agronomic characters. *Buletin Palawija*. **19** (2): 117-125.

Lakshmy, M. P., Engrala, A., Yabi, G., and Singh, A. K. 2020. Response of soybean (*Glycine max* L. Merrill) to foliar application of nutrients. *International Journal of Current Microbiology and Applied Sciences.* **9** (04): 494-500.

Lenka, D. and Mishra, B. 1973. Path coefficient analysis of yield in rice varieties. *Indian Journal of Agricultural Science.* **43**: 376-379.

Li, M., Liu, Y., Wang, C., Yang, X., Li, D., Zhang, X., Xu, C., Zhang, Y., Li, W. and Zhao, L. 2020. Identification of traits contributing to high and stable yield in different soybean varieties across three Chinese latitudes. *Frontiers in Plant Science*. **10**: 1642.

Liu, K. 1997. Chemistry and Nutritional Value of Soybean Components. In: Soybeans. Springer, Boston, MA. https://doi.org/10.1007/978-1-4615-1763-4_2.

Liu, X., Wu, J., Ren, H., Qi, Y., Li, C., Cao, J., Zhang, X., Zhang, Z., Cai, Z. and Gai, J. 2017. Genetic variation of world soybean maturity date and geographic distribution of maturity groups. *Breeding Science Preview*. **67** (3): 221–232.

Lush, J. L. 1940. Inter size correlation and regression of offspring on dams as a method of estimating heritability of characters. *Proceedings of American Society of Animal Production.* **33**: 293-301.

Machado, B. Q. V., Nogueira, A. P. O., Hamawaki, O. T., Rezende, G. F., Jorge, G. L., Silveira, I. C., Medeiros, L. A., Hamawaki, R. L. and Hamawaki, C. D. L. 2017. Phenotypic and genotypic correlations between soybean agronomic traits and path analysis. *Genetics and Molecular Research*. **16** (2): gmr16029696.

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

Mahbub, M. M. and Shirazy, B. J. 2016. Evaluation of genetic diversity in different genotypes of soybean [*Glycine max* (L.) Merrill]. *American Journal of PlantScience*. **1** (1): 24-29.

Mahbub, M. M., Mamunur, M., Rahman, H. M. S., Mahmud, F. and Mir Kabir, M. M. 2015. Genetic variability, correlation and path analysis for yield and yield contributing components in soybean. *American-Eurasian Journal of Agriculture and Environmental Science*. **15** (2): 231-236.

Mahesh, M., Sudhanshu, J. and Bhadoria, V. S. 2017. Genetic divergence studies in soybean (*Glycine max* (L.) Merrill). *International Journal of Agriculture Sciences*. **9** (19): 4196-4197.

Malek, M.A., Rafii, M.Y., Nath, U.K. and Mondal, M. 2014. Morphological characterization and assessment of genetic variability, character association, and divergence in soybean mutants. *The Scientific World Journal*. **1**:1–12.

Malik, M. F. A., Muhammad, A., Qureshi, A. S. and Abdul, G. 2007.Assessment of genetic variability, correlation and path analyses for yield and itscomponents in soybean. *Pakistan Journal of Botany.* **39** (2): 405-413.

Marconato, M. B., Pereira, E. M., Silva, F. M., Bizari, E. H., Pinheiro, J. B., Mauro, A. O. and Unêda-Trevisoli, S. H. 2016. Genetic divergence in a soybean (*Glycine max*) diversity panel based on agro-morphological traits. *Genetics and molecular research*. **15:** 1-10.

Meena, M., Sudhanshu, J.and Bhadoria, V.S. 2017. Genetic divergence studies in soybean (*Glycine max* L. Merrill). *International Journal of Agriculture Sciences*. **9** (19): 4196-4197.

Mimura and Coyne. 2007. SSR diversity of vegetable Soybean (*Glycine max* L. Merrill).*Genetic Resources and Crop Evolution.* **54**: 97-508.

Min, W., Run-zhi, L., Wan-ming, Y., and Wei-jun, D. 2010. Assessing the genetic diversity of cultivars and wild soybeans using SSR markers. *African Journal of Biotechnology*. **9** (31): 4857-4866.

Mishra, A. K. 2019. Association analysis in diverse populations of soybean. *Soybean Research*. **17** (1 and 2): 30-39.

Mishra, S., Jha, A., Panchshwar, D. K. and Shrivastava, A. N. 2018. Study of genetic divergence in advance breeding lines of soybean [*Glycine max* (L.) Merrill] for yield attributing traits. *International Journal of Bio-resource and Stress Management.* **9** (1): 103-107.

Mishra, S., Shrivastava, A. N., Jha, A. and Tantaway, S. 2015. Correlation and path coefficient analysis of advance breeding lines of soybean. *Annals of plant soil research*. **17**: 34-37.

Modgil, R., Kumar, V., Tanwar, A. and Goyal, A. 2021. Soybean (*Glycine max*). In oilseeds: Health attributes and food applications. Editions-Springer. Singapore, pp1-46. ISBN: 978-981-15-4194-0.

Moniruzzaman, M., Saiem, R. M., Emon, R. M., Haque, M. S., Saha, N. R., Malek, M. A. and Khatun, K. 2019. Genetic diversity analysis of soybean genotypes using SSR markers for salinity tolerance. *Progressive Agriculture*. **3030** (1): 1-9.

Mukuze, C., Tukamuhabwa, P., Maphosa, M., Dari, S., Dramadri, I. O., Obua, T., Kongai, H. and Rubaihayo, P. 2020. Genetic diversity analysis among soybean genotypes using SSR markers in Uganda. *African Journal of Biotechnology*. **19** (7): 439–448.

Mulato, B. M., Möller, M., Zucchi, M. I., Quecini, V. and Pinheiro, J. B. 2010. Genetic diversity in soybean germplasm identified by SSR and EST-SSR markers. *Pesquisa Agropecuaria Brasileira*. **45**: 276-283.

Murray, M. G. and Thompson, W. F. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acid Research*. **8**: 4321-4325.

Nag, S. K. and Sarawgi, A. K. 2021. Genetic divergence in soybean (*Glycine max* (L.) Merrill). *The Pharma Innovation Journal*. **10** (6): 475-478.

Nag, S. K., Sarawgi, A. K., Mehta, N., Verulkar, S. B. and Saxena, R. R. 2017. Study of genetic divergence in soybean (*Glycine max* L. Merrill) using principle component analysis. *Bulletin of Environment, Pharmacology and Life Sciences*. **6** (4): 45-50.

Narne, Aher, P., Dahat, V. and Aher, A. R. 2002. Selection of protein rich genotypes in soybean. *Crop Research.* **24** (1): 106-112.

Neelima, G., Mehtre, S. P. and Narkhede, G. W. 2017. Correlation coefficient and path analysis studies in soybean(*Glycine max* (L.) Merrill.)*Multilogic in Science*. **7** (23): 96-99.

Neelima, G., Mehtre, S. P. and Narkhede, G. W. 2018. Genetic variability, heritability and genetic advance in soybean. *International Journal of Pure and Applied Bioscience*. **6** (2): 1011-1017.

Nogueira, A. P. O., Sediyama, T., Souza, L. B., Hamawaki, O. T., Cruz, C. D., Pereira, D. G. and Matsuo, E. 2012. Path analysis and correlations between characters in soybean cultivated at two sowing seasons. *Bioscience Journal*.**28** (6): 877-888.

Osekita, O. S. and Ajayi, A. T. 2013. Character expression and selection differential for yield and its components in soybean [*Glycine max* (L.) Merrill]. *Academia Journal of Agricultural Research*. **1** (9): 167-171.

Pagde, L. S., Aubakkar, S. and Ingole, D.G. 2015. Study of genetic variability for yield and yield contribution traits in soybean (*Glycine max* (L.) Merrill). *Bioinfolet*. **12**: 256-258.

Painkra, P., Shrivastava, R., Nag, S. K. and Kute, I. 2018. Correlation analysis for seed yield and its attributing traits in soybean (*Glycine max* L. Merrill). *International Journal of Current Microbiology and Applied Science*. **7** (4): 2034-2040.

Panse, V. G. and Sukhatme, P. V. 1967. Statistical methods for agricultural workers. ICAR, New Delhi.

Parihar, R., Nag, S. and Shrivastava, R. 2020. Association analysis studies for seed yield and quality characters of soybean (*Glycine max*) at Bilaspur plain region of Chhattisgarh. *Crop Research*. **55** (5 and 6): 276-280.

Patil, S. S., Naik, M. R., Patil, P. P. and Shinde, D. A. 2011. Genetic variability,correlation and path analysis in soybean. *Legume Research.* **34** (1): 36-40.

Pawar, M. G., Chaudhary, S. B., Pawar, V. S. and Chavan, S. B. 2020. Study of genetic variability in different soybean genotypes based on yield and yield contributing traits. *International Journal of Current Microbiology and Applied Science*. **9** (9): 426-433.

Peakall, R. and Smouse, P. E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics*. **28**: 2537-2539.

Priolli, R. H. G., Mendes, C. T. Jr., Arantes, N. E. and Contel, E. P. B. 2002. Characterization of Brazilian soybean cultivars using microsatellite markers. *Genetics and Molecular Biology*.**25**: 185–193.

Pushpendra, A., Singh, K., Gupta, M. K. and Bhareti, P. 2016. Morphological characterization and genetic divergence study in soybean. *Soybean Research*. **13** (2): 1-8.

Rani, A., Kumar, V. and Rawal, R. 2013. Identification of simple sequence repeat markers linked to lipoxygenase-1 gene in soybean. *Journal of Plant Biochemistry andBiotechnology*. **22** (4): 488–491.

Rao, C. R. 1952. Advanced statistical methods in biometrical research. John Wiley and Sons, Incorporation, New York, pp 357-363.

Reddy, D. M., Rao, Y. K., Murthy, S. S. N. and Reddy, M. V. 2004. Genetic variability and divergence in mungbean. *Indian Journal Pulse Research.* **17** (1): 77-79.

Reni, Y. P. and Rao, Y. K. 2013. Genetic variability in soybean *(Glycine max* (L.)Merrill). *International Journal of Plant, Animal and Environment Science.* **3** (4): 2231-4490.

Robinson, H. F., Comstock, R. E. and Harvey, P. H. 1949. Estimates of heritability and degree of dominance in corn. *Journal of Agronomy*. **41**: 353-359.

Rohlf, F. J. 2000. NTSYS-pc: Numerical taxonomy and multivariate analysis system, version 2.1. Exeter software, Setauket, New York, USA.

Saghai-Maroof, M. A., Soliman, K. M., Jorgensen, R. A. and Allad, R. W. 1984. Ribosomal DNA spacer-length polymorphism in barley: Mendelian inheritance, Chromosomal location and Population dynamics. *Proceedings of National Academy Science*, U.S.A. **81**: 8014-8019.

Saharia, N. and Sarma, N. R. 2022. Characterization of soybean genotypes based on morphological and molecular markers. *Electronic Journal of Plant Breeding*. **13** (1): 182-192.

Sahoo, S. 2022. Assessment of genetic variability of parameters for yield and its components in soybean [Glycine max (L.) Merrill]. *International Journal of Plant and Soil Science*. **34** (19): 140-148.

Sayama, T., Hwang, T. Y., Komatsu, K., Takada, Y., Takahashi, M., Kato, S., Sasama, H., Higashi, A., Nakamoto, Y., Funatsuki, H. and Ishimoto, M. 2011. Development and application of a whole genome simple sequence repeat panel for high throughput genotyping in soybean. *DNA Research Advance Access Publication.* **18**: 107-115.

Shankar, K. 2014. Genetic diversity studies in soybean [*Glycine max* (L.) Merrill] for various seed and yield contributing characters. M.Sc. Thesis Acharya N. G. Ranga Agricultural University Hyderabad.

Sharma, K. D., Kumar, M., Datt, S., Singh, N. K. and Sharma, R. 2012. Estimation of genetic divergence and selection of promising genotypes for hybridization in soybean under foothill condition of Manipur.*Soybean Research* **10**: 11-17.

Sharma, R. and Lal, G. M. 2020. Study of Genetic Divergence in Soyabean (*Glycine max. (L.)* Merrill) Germplasm. *International Journal of Current Microbiology and Applied Sciences.***9** (09): 2532-2537.

Shekar, G. C., Pushpendra, M., Prasanth, H., Lokesha, M., Swamy, M., Lokesh, K., Shrotia, P. K. and Singh, K. 2018. Correlation and Path Analysis in Soybean [*Glycine max* (L.) Merrill]. *International Journal of Current Microbiology Applied Science.* **7** (09): 1232-1239.

Shi, A., Chen, P., Zhang, B. and Hou, A. 2010. Genetic diversity and association analysis of protein and oil content in food-grade soybeans from Asia and the United States. *Plant Breeding.***129** (3): 250–256.

Shilpashree, N., Devi, S. N., Manjunathagowda, D. C., Muddappa, A., Abdelmohsen, S. A. M., Tamam, N., Elansary, H. O., El-Abedin, T. K. Z., Abdelbacki, A. M. M. and Janhavi, V. 2021. Morphological characterization, variability and diversity among vegetable soybean (*Glycine max* L.) genotypes. *Plants*. **10*:*** 671.

Shinde, S. R., Desai, S. V and Pawar, R. M. 2013. Genetic diversity pattern in soybean[*Glycine max* (L.) Merrill]. *Electronic Journal of Plant Breeding*. **4** (3): 1242-1245.

Shree, Y., Ram, S., Bhushan, S., Verma, N., Ahmad, E. and Kumar, S. 2018. Correlation between yield and yield attributing traits in soybean (*Glycine max* (L.) Merrill). *Journal of Pharmacognosy and Phytochemistry*. SPI: 298-301.

Shrivastava, M. K., Shukla, R. S and Singh, C. B. 2001. Genetic divergence in soybean.*JNKVV Research Journal publication*. **34** (1-2): 25-28.

Shurtleff ,W. and Aoyagi, A. 2010. History of Soybeans and Soy foods in South Asia/Indian Subcontinent (1656-2010). ISBN: 978-1-928914-31-0.

Silva, A. F., Sediyama, T., Silva, F. C. S., Bezerra, A. R. G.and Ferreira, L. V. 2015. Correlationand path analysis of soybean yield components.*International Journal of Plant, Animal andEnvironmental Sciences*. **5** (1): 177-17.

Singh, I. 1999. Genetic variability and correlation studies in soybean [*Glycine max* (L.) Merrill]. *Journal of Oilseeds Research*. **16**: 118-120.

Singh, K., Singh, S. and Kler, D. S. 2000. Correlation and regression studies among various growth and yield parameters of soybean [*Glycine max* (L.) Merrill] under Punjab conditions. *Crop Research*.**19**: 287-297.

Singh, P. K., Shrestha, J. and Kushwaha, U. K. S. 2020. Multivariate analysis of soybean genotypes. *The Journal of Agriculture and Natural Resources*. **3** (1): 69-76.

Singh, R. K and Chaudhary, B. D. 1977. Biometrical methods in quantitative genetic analysis. Kalyan Publishers, New Delhi.

Sirohi, S. P. S., Malik, S., Singh, S. P., Yadav, R. and Meenakshi. 2007. Genetic variability, correlations and path coefficient analysis for seed yield and itscomponents in soybean [*Glycine max* (L.) Merrill]. *Progressive Agriculture*. **7**: 119-123.

Sivasubramanian, S. and Menon, M. 1973. Heterosis and inbreeding depression in rice. *Madras Agricultural Journal*. **60**: 1139-1149.

Smith, J. S. C., Chin, E. C. L., Shu, H., Smith, O. S., Wall, S. J., Senoir, M. L., Mitchell, S. E., Kresovich, S. and Ziegle, J. 1997. An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPs and pedigree. *Theoretical and Applied Genetics*. **95**: 163-173.

Song, J. Y., Piao, X., Choi, Y., Lee, G., Chung, J., Lee, J. and Lee, M. C. 2013. Evaluation of genetic diversity and comparison of biochemical traits of soybean (*Glycine max* L.) germplasm collections. *Journal of Plant Breeding and* Biotechnology.**1** (4): 374-384.

Sultana, S., Rahman, L. and Habib, S. H. 2005. Genetic variability, correlation and path analysis in soybean (*Glycine max* (L.) Merrill). *International Journal ofSustainable Agricultural Technology*. **1** (5): 16-20.

Tamang, J.P., 2009. Himalayan Fermented Foods: Microbiology, Nutrition, and Ethnic values. CRC Press, Taylor & Francis Group, New York, pp- 320.

Tantasawat, P., Trongchuen, J., Prajongjai, T., Jenweerawat, S. and Chaowiset, W. 2011. SSR analysis of soybean (*Glycine max* L.) megenetic relationship and variety identification in Thailand. *Australian journal of crop science*.**5** (3): 283 290.

Thakur, D. K., Nag, S. K., Sharma, A. and Chandrakar, S. 2015. Analysis of genetic divergence in soybean (*Glycine max* (L.) Merrill.). *Trends in Biosciences*. **8** (6): 1480-1483.

Tigga, A. 2021. Genetic variability and diversity analysis for yield and its contributing characters in soybean (*Glycine max* (L.) Merrill). M.Sc. Thesis, Indira Gandhi Krishi Vishwavidyalaya, Raipur.

Tiwari, S., Tripathi, N., Tsuji, K. and Tantwai, K. 2019. Genetic diversity and population structure of Indian soybean (*Glycine max* (L.) Merrill) as revealed by microsatellite markers.*Physiology and Molecular Biology of Plants: an International Journal of Functional Plant Biology*. **25** (4): 953-964.

Tyagi, S. D., Khan, M. H. and Tyagi, V. 2012. Genetic divergence in soybean (*Glycinemax* (L.) Merrill) grown in over environments. *Indian Journal of Ecology*. **39** (2): 272-275.

Upadhyay, P., Shrivastava, M. K., Amrate, P. K., Sharma, S., Thakur, S. and Anand, K. J. 2022. Assessing genetic diversity of exotic lines of soybean based on D^2 and principal component analysis.*The Pharma Innovation Journal*. **11** (5): 89-93.

USDA. 2022. World agricultural production. United States Department of Agriculture (USDA) Foreign Agricultural Service/Global Market Analysis. https://apps.fas.usda.gov/psdonline/circulars/production.pdf.

Valencia-Ramirez, R. A. and Ligarreto-Moreno, G. A. 2012. Phenotypic correlation and path analysis for yield in soybean (Glycine max (L.) Merrill). *Acta Agronomy.* **61** (4): 353-363.

Velusamy, V., Lee, K. J., Ha, B. K., Kim, J. B., Kim, S. H., Ahn, J. W., Kang, S. Y. and Kim, D. S. 2013. Evaluation of genetic diversity in Korean soybean landraces by protein banding patterns using high-throughput screening. *Journal of Crop Science and Biotechnology*. **16**: 189–195.

Verma, S. 2019. Genetic diversity and molecular characterization of soybean (*Glycine max* L. Merrill). M.Sc. Thesis, Indira Gandhi Krishi Vishwavidyalaya, Raipur, pp 1-82.

Wang, L., Guan, R., Zhangxiong, L., Chang, R. and Qiu, L. 2006. Genetic diversity of Chinese cultivated soybean revealed by SSR markers. *Crop Science*. **46** (3): 1032-1038.

Wang, M., Li, R. Z., Yang, W. M. and Du, W. J. 2010. Assessing the genetic diversity of cultivars and wild soybeans using SSR markers. African Journal of Biotechnology. **9** (31): 4857-4866.

Wang, Y. H., Zhang, X. J. and Fan, S. J. 2015. Genetic diversity of wild soybean populations in Dongying, China, by simple sequence repeats analysis. *Genetics and Molecular Research*. **14**: 11613–23.

Weir, B. S. 1990. Genetic Data Analysis: Methods for Discreate Population Genetic Data. *Sinauer Associates Incorporated, Science*, pp-377.

Wen, Z. X., Zhao, T. J., Zheng, Y. Z., Liu, S. H., Wang, C. E., Wang, F. and Gai, J. Y. 2008. Association analysis of agronomic and quality traits with SSR markers in *Glycine max* and *Glycine soja* in China: I. Population structure and associated markers. *Acta Agronomica Sinica*. **34** (7): 1169-1178.

Widaningsih, N. A., Purwanto, E., Nandariyah, N. and Reflinur, R. 2014. The Use of DNA Microsatellite markers for genetic diversity identification of soybean (*Glycine max* (L) Merrill) as a supplementary method in reference collections management. *Indonesian Journal of Biotechnology*. **19** (2): 136-145.

Xu, Y., Shimoro, X. and Hofstra, H. 1994. Plant DNA isolation protocol. *Nucleic Acid Research*. **22**: 2399-2403.

Yahaya, S. U. and Ankrumah, E. 2015 Character association and path coefficient analysis between some growth characters and grain yield in soybean (*Glycine max* (L.) Merril.) *Bayero Journal of Pure and Applied Sciences*. **8** (2): 210 – 215.

Yao, T. Z. 1989. Study on harvest index of soybean in high yield breeding. *Journal ofJilin Agricultural Universities.* **2**: 8-22.

Zhang, G., Xu, S., Mao, W., Hu, Q. and Gong, Y. 2013. Determination of the genetic diversity of vegetable soybean [*Glycine max* (L.) Merrill] using EST-SSR markers. *Journal of Zhejiang University- Science B*. **14** (4): 279–88.

Zhao, H., Wang, Y., Xing, F., Liu, X., Yuan, C., Qi, G. and Dong, Y. 2018. The genetic diversity and geographic differentiation of the wild soybean [*Glycine soja* (L.) Merrill.] in Northeast China based on nuclear microsatellite variation. *International Journal of Genomics*. **2018**: 1-9.

Zinaw, D., Seltene, A. and Addisu, G. 2013. Genetic variability and heritability of soybean (*Glycine max* (L.) Merrill) genotypes in Pawe district, Metekel zone, Benishangule Gumuz regional state, northwestern Ethiopia. *Wudpecker Journal of Agricultural Research*. **2** (9): 240-245.

APPENDICES

Plate 6: Layout of experimental farm

Plate 7: Field view of Initial germination stage

(a) **(b)**

Plate 8: Crop at flowering stage -(a) purple and (b) white flowers

Plate 9: Pod formation stage

Plate 10: Crop at physiological maturity