GENOMIC PROFILING FOR ERUCIC ACID IN EUROPEAN GERMPLASM OF *BRASSICA JUNCEA* (L.)

By

Neha Bharti

(J-17-MB-31)

Thesis submitted to Faculty of Post Graduate Studies
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE
IN
BIOTECHNOLOGY

School of Biotechnology

Sher-e-Kashmir University of Agricultural Sciences & Technology of Jammu
Main Campus, Chatha, Jammu - 180009

2019

CERTIFICATE-I

This is to certify that the thesis entitled **"Genomic profiling for erucic acid in European germplasm of *Brassica Juncea* (L.)"** submitted in partial fulfillment of the requirements for the degree of **Master of Science in Biotechnology** to the Faculty of Post-Graduate Studies, Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu is a record of bonafide research carried out by **Ms. Neha Bharti,** Registration No. **J-17-MB-31** under my supervision and guidance. No part of the thesis has been submitted for any other degree or diploma. It is further certified that such help and assistance received during the course of investigation have been duly acknowledged.

Dr. Ravinder Singh
Major Advisor

Place: Jammu
Date: 24-07-2019

Endorsed

Coordinator

School of Biotechnology
SKUAST-J, Chatha

Date: 24-07-2019

CERTIFICATE-II

We, the members of the Advisory Committee of **Ms. Neha Bharti,** Registration No. **J-17-MB-31,** a candidate for the degree of **Master of Science in Biotechnology,** have gone through the manuscript of the thesis entitled **"Genomic profiling for erucic acid in European germplasm of *Brassica Juncea* (L.)"** and recommend that it may be submitted by the student in partial fulfillment of the requirements for the degree.

Dr. Ravinder Singh
Major Advisor

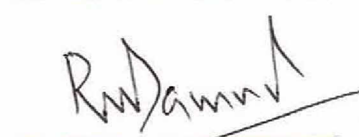
Place : Jammu
Date : 24-07-2019

Advisory Committee Members:

Dr. R. K. Salgotra
Professor and Coordinator
School of Biotechnology

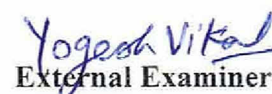
Dr. S.K. Gupta
Professor
Division of Plant Breeding and Genetics

Dr. R. K. Samnotra (Dean Nominee)
Professor
Division of Vegetable Sciences

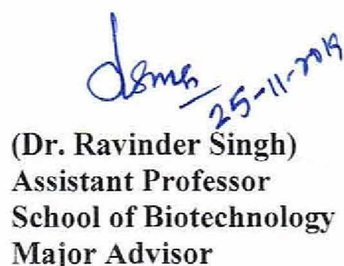


CERTIFICATE-III

This is to certify that thesis entitled **"Genomic profiling for erucic acid in European germplasm of *Brassica Juncea* (L.)"** submitted by **Ms. Neha Bharti,** Registration No. **J-17-MB-31** to the Faculty of Post Graduate Studies, Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu in partial fulfillment of the requirements for the degree of **Master of Science in Biotechnology** was examined and approved by the Advisory Committee and External Examiner(s) on ....25.11.19..............



External Examiner

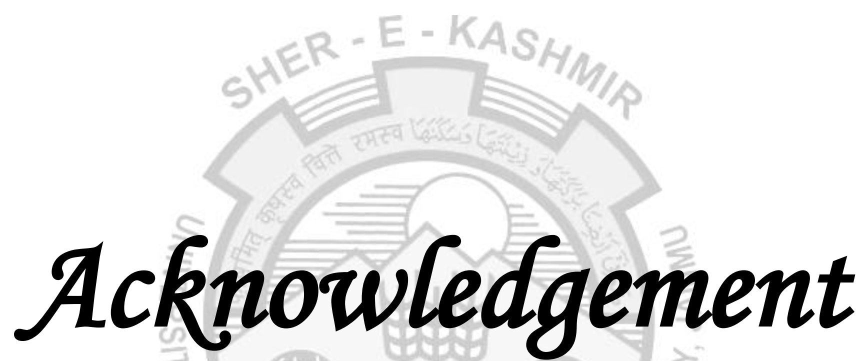


(Dr. Ravinder Singh)
Assistant Professor
School of Biotechnology
Major Advisor

Coordinator
School of Biotechnology

Dean, FoA

Acknowledgement



ACKNOWLEDGEMENT

Today, wherever I am standing is due to the tired less efforts of my parents for me, right from the day I took my first step. Diction is inadequate to express my feelings for the love, affection, care and blessings of my parents.

From the core of my heart, I thank Almighty for blessing me with enough patience endurance and strength in accepting the bitter truth of life and in accomplishment of my targets.

It is an honour and great privilege to pay in ineffable gratitude to my worthy, generous, venerated, unparalleled and innovative major advisor ***Dr. Ravinder Singh****, Assistant Professor, School of Biotechnology, for his perseverance, enthusiastic guidance, illuminating and affirmative attitude to accomplish this task successfully, I am tremendously fortunate to have advisory committee members* ***Dr. R.K. Salgotra****, (Professor & Coordinator, School of Biotechnology).* ***Dr. S.K. Gupta,*** *(Professor and Head, Division of Plant Breeding and Genetics) and Dean's Nominee* ***Dr. R.K. Samnotra,*** *(Professor, Division of Vegetable Sciences). I thank them for supporting this thesis and giving such thoughtful feedback, always aimed at moving me forward.*

It is my sole prerogative to place on records my everlasting gratitude and ecstatic thanks to ***Dr. R.K. Salgotra****, (Professor and Coordinator (School of Biotechnology),* ***Dr. A.K. Singh,*** *(Associate Professor, School of Biotechnology),* ***Dr. Manmohan Sharma,*** *(Associate Professor, School of Biotechnology),* ***Dr. Susheel Sharma,*** *(Assistant Professor, School of Biotechnology),* ***Dr. G.K. Rai*** *(Assistant Professor, School of Biotechnology)*

I express my veneration to Dean, PGS for providing necessary facilities during the pursuit of my research.

I do extend my thanks to the non- teaching staff of my Division for their help during the present study, especially ***Nitu Bhaiya****.*

I express Sincere thanks to my respected seniors ***Shazia mam, Meenakshi mam*** *and* ***Neeraj mam*** *for providing a friendly lab atmosphere and kind cooperation during my research.*

From my inner core of heart, I would like to thanks and respect to my heartiest friends- ***Wimpal*** *and* ***Jaspreet*** *for their everlasting care, support, help, cooperation and encouragement.*

Everything I am today is because of my beloved ***father*** *and* ***mother****. Their unbounded love, unfailing prayers, blessings, uncountable sacrifices, unbarred assistance and faith upon me made me strong to face the challenges which come on my way. Most importantly, I owe a lot to my ever supporting and caring husband* ***Mr. Deepak Kumar****, who provided me unending inspiration emotionally and mentally. I am grateful to him for being there with me at each and every time and supporting me to fulfil my dream.*

Place:-Jammu **Neha Bharti**

Date:- 28-11-2019

ABSTRACT

Title of Thesis/Dissertation	:	Genomic Profiling for Erucic Acid in European Germplasm of *Brassica Juncea* (L.)
Name of the Student	:	Neha Bharti
Registration No.	:	J-17-MB-31
Major Subject	:	Biotechnology
Major Advisor	:	Dr. Ravinder Singh (Asstt. Professor)
Degree to be Awarded	:	M.Sc. Biotechnology
Year of award of Degree	:	2019
Name of the University	:	Sher-e- Kashmir University of Agricultural Sciences and Technology of Jammu

Abstract

Rapeseed mustard (*Brassica juncea*), also known as Indian mustard, are grown world-wide including India as a major source of edible oil. Though grown widely, Indian mustard has high content of anti-nutritional compounds including erucic acid. Therefore, it is important to develop new varieties, containing low levels (<2%) of erucic acid. In the present study, biochemical profiling and genome-wide association analysis was carried out to identify genes for erucic acid. Biochemical analysis revealed three genotypes with least percent of erucic acid namely, Zem 1 (8.08 %), Burgonde (11.43%) and Volgogradskaja 189/191 (11.99%). In order to develop and genotype SNPs, the samples were sequenced using paired-end chemistry with a read length of 151 bases. In total, nearly 104.4 million reads amounting to more than 14.4Gbs of sequence data were generated. The analysis of ddRAD sequencing data revealed that more than 90% data had a Q score of greater than 30 for all the samples. The sequence data analysis led to identification of a total of 98,401 SNPs at a frequency of 9.40KB on all 18 chromosomes of *Brassica juncea*. Another 28,603 SNPs were identified on contigs which were not mapped to any of the chromosomes. The GLM-based association analysis (using TASSEL) led to the identification a total of 125 significant marker-trait associations (p-value < 0.001) distributed on different *Brassica juncea* chromosomes for erucic acid. The highest number of 22 associations were found on A03 followed by 15 and 12 on B07 and B03, respectively. In all, ten most significant associations (p-value < 1.0 x E-04) were found on chromosomes A01, A02, A03, A04, B01 and B05 chromosomes. Two SNPs found within 15nt (at positions 4034447 and 4034461) on A03 were the most significant ones with a p-value < 4.84 E-06. The significant associations found, would form a very important genomic resource for the development of new varieties and also for improvement of the existing varieties through marker-assisted selection (MAS) in future breeding programmes.

Key words: Single nucleotide polymorphisms, ddRAD-sequencing, association analysis, MAS, *Brassica juncea.*

Signature of Major Advisor

Signature of Student

TABLE OF CONTENTS

CHAPTER	TOPIC	PAGE NO.
1	INTRODUCTION	1-4
2	REVIEW OF LITERATURE	5-15
3	MATERIALS AND METHODS	16-22
4	RESULTS	23-29
5	DISCUSSION	30-32
6	SUMMARY AND CONCLUSIONS	33-34
	REFERENCES	35-42
	VITA	

LIST OF TABLES

Table No.	Particulars	Page No.
1.	Canola quality parameters (Oil content and composition)	9
2.	List of *Brassica juncea* genotypes used during present study	16
3.	List of chemical solution used for DNA extraction	19
4.	List of component used to make buffer	19
5.	Percent erucic acid among European genepool genotypes	23
6.	Estimates of erucic acid among 50 genotypes	25
7.	Summary of ddRAD sequence data of European genotypes	26
8.	Chromosome wise distribution of SNPs	28
9.	Ten most significant associations on the basis of P-value	29

LIST OF PLATES

Plate No.	Particulars	After Page No.
1.	Field trial at SKUAST-J for biochemical and molecular analysis of 50 diverse *Brassica juncea* genotypes	17
2.	DNA quality analysis of *Brassica juncea* genotypes	24
3.	Estimation of erucic acid by GC	24

LIST OF FIGURES

Figure No.	Particulars	After Page No.
1.	*Brassica* "U" triangle	**8**
2.	Estimation of erucic acid by GC analysis	**25**
3.	Microfludics-based analysis of a pool of five samples indicating fragment size distribution	**26**
		After Page No.
4.	Manhattan plot of log (P) values of marker-trait association for erucic acid	**29**

ABBREVIATIONS

A	Absorbance
bp	Base pair
CTAB	Cetyl trimethyl ammonium bromide
ddRAD	Double digest restriction-site associated DNA
DNA	Deoxyribonucleic acid
EDTA	Ethylene Diamine Tetra Acetic Acid
GBS	Genotyping-by-sequencing
GC	Gas Chromatography
MAS	Marker assisted selection
mg	Milligram
NaCl	Sodium chloride
NaoH	Sodium methoxide
ng	Nanogram
NGS	Next Generation Sequencing
No.	Number
OD	Optical density
RNase	Ribonuclease Enzyme
SNPs	Single nucleotide polymorphisms
TAE	Tris-acetate-EDTA
µl	Microliter

Introduction

CHAPTER-1

INTRODUCTION

Brassica juncea, commonly known as Indian mustard, is a dominant Brassica species grown on Indian subcontinent and regarded as of Asiatic origin (Gupta *et al.,* 2012). *Brassica juncea* (L.) Czern. belongs to the *Cruciferae* (*Brassicaceae*) plant family, commonly known as the mustard family. The name crucifer is derived from the shape of the flowers that have four diagonally opposed petals in the form of a cross. *B. juncea* has pale green foliage, with a few hairs on the first leaves and leaf blades that terminate well up the petiole.

Brassica juncea (AABB; 2n = 36) is an amphidiploid species derived from interspecific crosses between *B. rapa* (AA; 2n = 20) and *B. nigra* (BB, 2n = 16). The cytogenetic relationship between the main species of the *Brassica* genus was first discovered by Morinaga (1934) and visualised by Dr. Woo Jang-choon in 1935 as U triangle (Nagaharu U, 1935). *Brassica juncea* has two distinct gene pools: the Indian, and the east European gene pool. The Indian gene pool has narrow genetic diversity. *Brassica juncea* has both oilseed and vegetable varieties that have different origins (Hemingway, 1976). Oilseed rape (*Brassica napus* L.) is an important oilseed crop in the world. *B. napus* is an amphidiploid (AACC genome, 2n = 38) and is believed to have arisen by interspecific hybridization between the diploid species *B. rapa* L. (syn. Campestris) (AA genome, 2n = 20) and *B. oleracea* L. (CC genome, 2n = 18) (Prakash and Hinata 1980).

Brassica juncea leaves are high in vitamin A, C and iron. A cupful (140gm) of mustard leaves provides an adult with 60% of his recommended daily vitamin A requirements, the entire vitamin C requirement and about one-fifth of the iron. 100 g of leaves contain 24 calories, 91.8 g H_2O, 2.4g protein, 0.4 g fat, 4.3 g total carbohydrate, 1.0 g fiber, 1.1 g ash, 160 mg calcium, 48 mg riboflavin, 0.8 mg niacin, and 73 mg ascorbic acid. Young tender leaves of mustard green are used in salads or mixed with other salad greens. Older leaves with stem may be eaten fresh, canned or frozen, for potherbs, and to a limited extent in salads. Mustard greens are often cooked with ham or salt pork, and may be used in soups and stew. Mustard is rubbed over meat before

roasting to enhance its flavour. It is also added to butter to give butter a pleasant flavour (Anuradha et al., 2012).

Brassica juncea is one such economically important plant, well known in India for centuries for its nutritive and medicinal values (Ram Manohar et al., 2009). The leaves as well as the seeds of this mustard variety are edible, and diverse medicinal uses of its seeds are also well known in other countries. Rapeseed mustard is the third leading source of vegetable oil in the world after soybean (*Glycine max*) and palm (*Elaeis guineensis*) and also the world's second leading source of protein meal. Mustard oil is rich in α-linolenic acid, contains high amount of mono-unsaturated fatty acids and a good ratio of polyunsaturated fatty acids. It also contains least amount of saturated fatty acids, making it safe for heart patients. The relatively high level of oleic acid and optimal balance between linolenic and linoleic acids makes mustard oil very healthy edible oil. In spite of its uses, rapeseed-mustard oil is not valued much in international market due to the presence of high amount of anti-nutritional compounds like glucosinolates and erucic acid. Tests on animals show that ingesting oils containing erucic acid over time can lead to a heart condition called myocardial lipidosis. Other potential effects observed in animals – including changes in the weight of the liver, kidney and skeletal muscle – occur at slightly higher doses. A lot of efforts are being put into development of genomic and molecular resources for the improvement of existing and development of new varieties with the desirable characters including fatty acids and glucosinolate content.

Mustard meal is a good source of phenolic compounds. These compounds were previously considered undesirable because the presence of phenolic compounds can cause bitterness and astringency and dark colors in protein products, but they are now emerging as value added products with antioxidant properties. *Brassica* has two major seed storage proteins such as napin (2S albumin), constituting about 45–50% of the total proteins, and cruciferin (12S globulin), constituting about 25% of the total proteins (Appelqvist et al., 1973). The *Brassicas* are served to be one of the most agronomical important oilseeds with a diverse range of species may be used as a variety of oilseed, vegetable, and fodder crops. The main outstanding trait for *Brassica* species is the high seeds oil content, varying 17% to 40% in wild relatives. The major cultivars belonged to

Brassica oilseed crops, i.e., *B. napus, B. juncea*, and *B. rapa*, with average oil content ranged 45 to 50%. *Brassica* species seed oil is characterized by significant amount of long-chain monounsaturated fatty acids, mainly erucic acid absent in any other commercial plant oil. High erucic acid contained oils are useful for industrial applications, but not for human consumption. Therefore, to develop varieties having both commercial free from erucic acid as well as those with high erucic acid is a promising objective for breeding programs in *Brassica* oilseed crops. Other important objectives are the increase of oleic acid and linoleic acid, and the reduction of linolenic acid content.

Association analysis also known as association mapping or linkage disequilibrium mapping, is a method that relies on linkage disequilibrium to study the relationship between phenotypic variation and genetic polymorphisms (Flint-Garcia et al., 2003). Linkage disequilibrium (LD) is the non-random combination of alleles at two genetic loci, which in random mating populations is mostly generated by mutation and genetic drift, and decays by recombination. Therefore, LD will be observed between two loci if they are in tight linkage or if the haplotype is recent (Hedrick, 2005). Association analysis is a method potentially useful for detection of marker-trait associations based on linkage disequilibrium. According to Gupta et al., (2005), Association analysis refers to the significant association of a marker locus with a phenotype trait. The methodology has been used in plants for the discovery of genes associated with a number of traits (Ravelombola et al., 2018, Khan et al 2018, Patishtan et al., 2018).

Genotyping by sequencing technology is used for identification and genotyping of SNP molecular markers from a genome. GBS technique is simple and has ability to reduce the complexity in a genome (Elshire et al., 2011). The multiplex libraries are developed for next generation sequencing by using of restriction endonuclease enzyme and with DNA barcoded adapters to detecting a section of genome from genotyping by sequencing (GBS). This technique has ability to generate huge number of markers (Elshire et al., 2011). GBS is a best technique for study of populations because it is help in genetic mapping through which genomic selection can be done widely on large scale (Poland and Trever, 2012).

In view of the above, the present study was conducted for erucic acid profiling and association analysis among genotypes of European genepool. In view of the above, the present study was undertaken in *Brassica juncea* with the following objectives:

1. Biochemical characterization of erucic acid in a diverse collection of European gene pool.
2. Molecular profiling of natural sequence variation through genotyping–by-sequencing.
3. Identification of SNP(s) associated with erucic acid.

Review Of Literature

CHAPTER-2

REVIEW OF LITERATURE

Mustard oil is one of the major edible oils in India and it is the third largest contributor to the world supply of vegetable oils. It has also got medicinal importance (Yousuf et al., 2013) and is utilized worldwide as an oilseed, condiment, vegetable, green manure, forage and fodder and cultivated primarily in tropical and sub- tropical countries (Gangapur et al., 2010). Production and productivity of mustard in India is low as compared to world average primarily due to lack of proficient varieties. In order to enhance the production by developing new cultivars, the knowledge about the genetic wealth of available germplasm is must. Availability of germplasm and understanding of relationship among genotypes provides an opportunity to develop improved crop varieties by more efficient sampling of genotypes for use in breeding programmes (Afiah et al., 2017). A variety of molecular markers have been used to study the extent of genetic variation among diverse group of important crops in the genus *Brassica.*

2.1 Importance of *Brassica juncea*

Brassica juncea is a herbaceous plant with an erect, branched stem up to 1.0 m tall, with a taproot reaching 60-80 cm in depth, lower leaves petioled, green, sometimes with a whitish bloom, ovate to obovate, variously lobed with toothed or frilled edges; upper leaves subentire, short and petioled, constricted at intervals, sessile. The flowers consist of four yellow petals arranged in a cruciform manner, four yellowish green sepals, a short green pistil with a knobby stigma, and tetradynamous stamens with yellow anthers. They are pollinated by bees that soon develop into sickle-shaped green seed pods. Seeds are sown in very early spring. Plants are generally harvested before fruits are 7 fully ripe to reduce shattering. The growing period is from 40–60 days depending on the variety and weather conditions. Indian mustard is a cool-season vegetable, growing well at monthly average temperatures ranging from 15 to 18°C. It can tolerate annual precipitation of about 500 to 4,200 mm, annual temperature of 6 to 27°C and pH of 4.3 to 8.3.

Brassica juncea is valued for its intense flavours and healing properties. This plant is cultivated mainly as an oil crop. It is a good bee plant. Mustard is used all over the world for its appetizing flavor and preservative value and the seeds are used largely for tempering food. It is available in the form of seeds, powders and oil. Recently, *B. juncea* has been explored for its biodiesel potential (Jham et al., 2009). Mustard leaves are high in vitamin A, C and iron. A cupful of mustard leaves provides an adult with 60% of his recommended daily vitamin A requirements, the entire Vitamin C requirement and about one-fifth of the iron. 100g of leaves contain 24 calories, 91.8g H_2O, 2.4g protein, 0.4g fat, 4.3g total carbohydrate, 1.0g fiber, 1.1g ash, 160mg Ca, 48mg riboflavin, 0.8 mg niacin, and 73 mg ascorbic acid. Young tender leaves of mustard green are used in salads or mixed with other salad greens. Older leaves with stem may be eaten fresh, canned or frozen, for potherbs, and to a limited extent in salads. Mustard green are often cooked with ham or salt pork, and may be used in soups and stew. Mustard is rubbed over meat before roasting. It is also added to butter to give butter a pleasant flavour (Anuradha et al., 2012).

The plant appears in some form in African, Indian, Chinese, Japanese, and continental food cuisine. The leaves, the seeds, and the stem of Indian mustard are edible. Seeds of *B. juncea* contain 25-30% fatty non-drying oil and glycoside sinigrine. The leaves of young plants are used as green vegetables as they supply enough sulphur and minerals in the diet. *B. juncea* is used to make the Indian pickle called “Achar”, and the Chinese pickle known as ‘Zha cai’ (Everitt, 2007). Young tender leaves of mustard greens are used in salads or mixed with other salad greens. Older leaves with stems may be eaten fresh, canned or frozen, for potherbs, and to a limited extent in salads. Its basal leaves are eaten raw and used in salads or cooked like spinach. Leaves and stems are also added to soups and stews. Mustard greens are often cooked with ham or salt pork, and may be used in soups and stews. Seed residue is used as cattle feed and in fertilizers. In Asia, some kinds of mustard are pickled (called hum choy and sajur asin). The seeds are very pungent and used to season meats and other dishes. Although widely and extensively grown as a vegetable, it is being grown more for its seeds which yield an essential oil and condiment. Oil is used to pickle foods in Kashimiri and Bengali cooking. It is used as cooking oil in parts of India and Bangladesh. Mustard oil is one of the

healthiest edible oils. Mustard oil is healthier than olive oil because it has no trans fats, has low saturated fats, high mono-unsaturated fats and polyunsaturated fatty acids such as omega-3. It is stable at high temperatures, which makes it ideal for Indian cooking and even deep frying. Mustard is also a cheaper alternative of edible oil and makes the food tastier. In very small amounts, it is often used by the food industry for flavoring. Mustard seed meal is good source of protein (28−36%) and phenolic 9 antioxidants such as sinapine and sinapic acid (Das et al., 2009). Mustard oil is also used as hair oil, lubricants and as a substitute for olive oil in Russia. Oilseed cake is used for cattle feed and manure.

Indian mustard is reported to be anodyne, apertif, emetic, diuretic, rubefacient, and stimulant, it is also a folk remedy for arthritis, footache and rheumatism (Duke and Wain, 1981). Seeds are used for curing tumors (China), and oil is used as a counter-irritant and stimulant. The seeds are used for abscesses, colds, rheumatism, lumbago and stomach disorders (Korea). Chinese eat the leaves in soups for bladder inflammation or hemorrhage. Mustard oil is used for skin eruptions and ulcers (Perry, 1980). The seeds, crushed in honey, are known to cure coryza. Swallowing mustard seeds soaked in mustard oil, cures stomachache. Massaging the body with mustard oil is very beneficial as it cures flatulence and makes the body strong. Massage with the oil is thought to improve blood circulation, muscular development and good texture to human skin. The oil is also antibacterial. In skin diseases, the local application of seed oil is beneficial as it is antiseptic and antiinflammatory. The oil, with salt is an effective gargle in dental infections and pyorrhea. Mustard also has antioxidant activity and pharmacological effects on cardiovascular disease, cancer and diabetes. The oil makes up about 30% of the mustard seeds. For North Indians, mustard oil is not just a cooking medium but it is very much intricately interwoven with their culture. Mustard oil is beneficial to human health because of its low content of saturated fats, ideal ratio of omega-3 and omega 6 fatty acids, content of antioxidants such as vitamin E.

2.2 Evolution history of *Brassica* species

Among the different species of *Brassica*, three allotetraploids and three diploids species are particularly important for agriculture production. Interrelationship among these species is known as “Triangle of U” a described by Nagaharu U (1935). Nagaharu

U was a korean botanist who was working in Japan and his former Korean name Woo Jang- choon was translated into the Japanese alphabet. Woo resynthesized hybrids from diploid and tetraplod species and studies the chromosomal pairing behaviour in resulted triploids. The “Triangle of U” theory explained the contribution of three diploid species for the development of allotetraploids species. According to the theory the three allotetraploids species namely *Brassica carinata* A. Braun. (BBCC,2n=34), *Brassica juncea* L.Czern (AABB,2n=36) and *Brassica napus* L. (AACC,2n=38) originated from three diploid *Brassica* species *Brassica nigra* L. (BB,2n=16), *Brassica oleracea* L. (CC, 2n=18) and *Brassica campestris* L. (syn. *Brassica rapa* L. AA, 2n=20). The amphidiploid *Brassica napus* L. (2n=38, AACC) originated from three diploid species *Brassica rapa* L. (2n=20, AA) and *Brassica oleracea* L. (2n=18,CC).

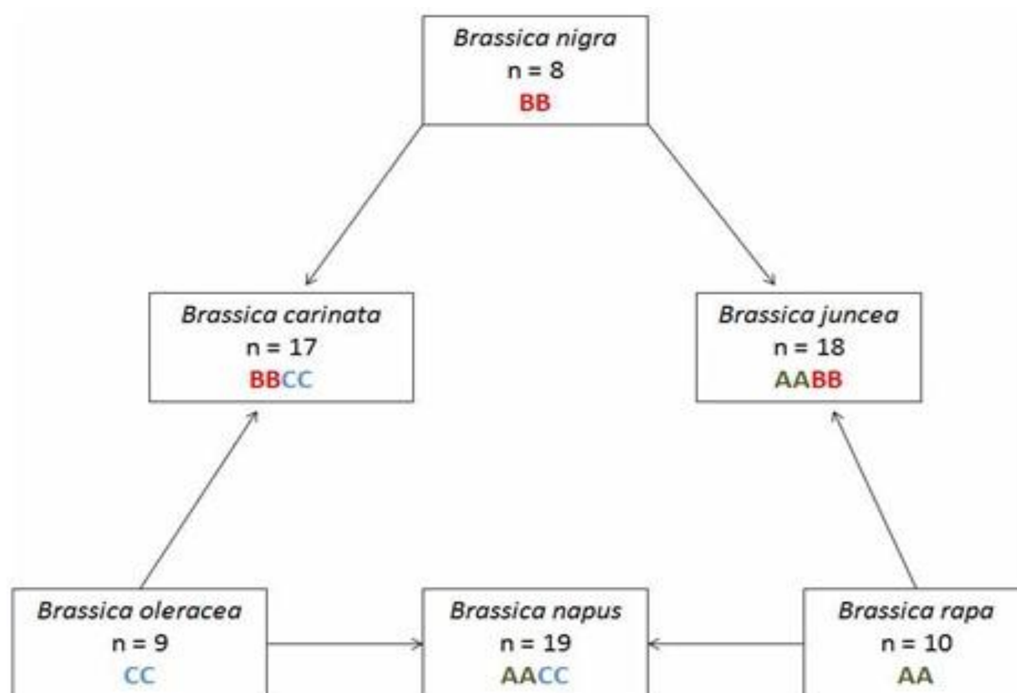


Figure 1: *Brassica* “U” triangle

2.3 Chemical compositions and nutritive value of *Brassica* oil

Oil content is expressed as a percentage of whole seed at 6% to 8.5% moisture (GRDC 2009; Mailer et al., 1999). Canola oil (both from *B. napus* and *B. juncea*) is high in unsaturated fats (92.1%), has no cholesterol or trans-fat, and has the lowest saturated

fat (7.9%) of any common edible oil. Because of this and the fact that it is low in low-density lipoproteins, the US Food and Drug Administration (FDA) now allows manufacturers to claim potential health benefits for canola oil due to reduced risk of coronary disease.

Table 1: Canola quality parameters, oil content and composition. (a) Average quality parameters of canola. Adapted from GRDC (2009).

Quality parameter	**Mean**
Oil content (% in whole seed, 6% moisture)	41.5
Protein content (% in oil-free meal, 10% moisture)	39.2
Total glucosinolates (µml/g of meal, 6% moisture)	20.0
Calories (per 100g of oil)	884
Saturated fats (% in oil)	7.9
Monounsaturated fats (% in oil)	63.7
Polyunsaturated fats (% in oil)	28.2
Erucic acid (% in oil)	0.1
Vitamin E (mg/100 g of oil)	17.46
Vitamin K (mg/100 g of oil)	71.3

2.4 Erucic acid and *Brassica juncea*

Erucic acid is one of the main constituent of oil composition of *Brassica juncea*. In Indian genepool, it is present to the extent of 50% of total oil composition. High concentration of erucic acid is considered anti-nutritional. Erucic acid is a 22-carbon monounsaturated fatty acid, with a single double bond at the omega 9 position. Erucic acid constitutes about 30-60% of the total fatty acids of rapeseed and mustard. It is synthetised in the cytosol by elongation of oleic acid, which is produced in plastids (Bao et al. 1998). Studies demonstrating a possible correlation between exposure to dietary erucic acid and number and severity of heart lesions in rats have led to human health concerns (Sauer & Kramer 1983). Myocardial lipidosis has also been described in pigs and monkeys following erucic acid consumption, indicating that this fatty acid is poorly

metabolised (Gopalan et al. 1974; Shenolikar & Tilak 1980). The consumption of high erucic acid-containing rapeseed oils (*B. napus*, *B. juncea* and *B. rapa*) since ancient times does not appear to have been associated with nutritional or health problems (Monsalve et al. 2001; Sauer & Kramer 1983). Because of physiological differences with humans, rats are not considered an appropriate model to study the effect of erucic acid (FSANZ 2003). It has been suggested that the incidence and severity of heart lesions in rats can be influenced by feeding of marine/vegetable oils but may not be specifically related to the erucic acid content of the oil (FSANZ 2003). Because of this and in the absence of adequate human data, FSANZ has set a no-observable effect level (NOEL) of 750 mg/kg bw/day, based on results obtained for nursling pigs. A provisional tolerable daily intake (PTDI) 22 was derived from it, using a safety risk factor of 100 (10 for extrapolating data from pigs to humans and 10 for variations within humans). The tolerable level for human exposure is thus 7.5 mg/kg bw/day (about 500 mg erucic acid per day for an average adult) (FSANZ 2003). For the average consumer, the dietary intake of erucic acid is 124 mg/day or 28% of the PTDI. The erucic acid content in *Brassica* species varied from 0.80- 49.40% in *B. juncea* varieties, 10.04 - 34.96% in *B. napus* varieties and 43.77-49.99% in *B. rapa* varieties. Higher erucic acid in cooking oil hampers the myocardial conductance in humans and leads to increased blood cholesterol levels (Bozzini et al., 2007; Sinha et al., 2007).

2.5 Importance of Fatty acid elongation gene (FAE1)

Presence of high erucic acid in edible oil makes it nutritionally undesirable for human consumption as it causes myocardial infarction and increased blood cholesterol (Mortuza *et al.* 2006). As per the international norms, erucic acid should be less than 2 per cent of the total fatty acids. It is therefore, imperative to reduce the erucic acid level to less than 2 per cent in seed oil for better human health. Studies have now confirmed that erucic acid content is controlled by *fatty acid elongase 1* (*FAE1*) gene that encodes the enzyme β-ketoacyl-CoA synthase (KCS) in erucic acid biosynthesis pathway and catalyzes the first four enzymatic reactions in synthesis of very long chain monounsaturated fatty acids (VLCMFAs) (Gupta *et al.* 2004, James *et al.* 1995, Millar and Kunst 1997). The mutation in *FAE1* gene leads to the loss of function in enzymatic activity and reduces the

accumulation of VLCMFAs in seeds (Katavic *et al.* 2002, Roscoe *et al.* 2001). Erucic acid trait is governed by two independent genes *FAE1.1* and *FAE1.2*. The diploid *Brassica* species (*B. rapa* and *B. nigra*) have one copy of *FAE1* gene while the amphidiploid species (*B. napus* and *B. juncea*) have two copies with additive effect (Gupta *et al.* 2004, Lühs *et al.* 1999). The *FAE1* gene is 1521 bp long without introns and encodes a protein of 507 amino acids (Xu *et al.* 2010). Several SNPs were reported between LEA and high erucic acid (HEA) genotypes in *B. napus* and *B. rapa* (Wang *et al.* 2010, Wu *et al.* 2008, Yan *et al.* 2015).Gupta *et al.* (2004) found substitution-type single-nucleotide polymorphisms (SNPs) in *FAE1.1* and *FAE1.2* to distinguish low erucic from high erucic types in *B. juncea*.

2.6 Genotyping-by-sequencing (GBS) and SNP identification

Single nucleotide polymorphisms (SNPs) are single base changes at a given locus in a population. In other word SNP is polymorphism which occurs within two DNA samples and difference between single bases by addition, deletion, transversion, transition of chromosome (Ayeh et al., 2008; Hearne et al., 2008). SNP markers are extremely reliable directly furnish phenotype (Batley and Edwards, 2007). They are the easiest type of markers as having minor heredity entity as alone base and can produce large number of markers. SNPs are commonly assigned and connected with morphological changes used as marker. SNPs play an important role in creations of phenotypic variation by DNA polymorphism in genome of plants, animals and humans. According to international working group of SNP 1.42 million of SNPs are found in human genome and average one SNP is equal to 1.9kb (Sachidanandam et al., 2001). In addition to plants SNP polymorphism are present high density in genome (Ching et al., 2002).

A number of methods are available for identification and genotyping including Genotyping by sequencing(GBS), next generation sequencing technology enzymatic and chemical mismatch assays, nucleotide amplification polymorphisms (SNAP), ligase chain reaction, Single Stranded confirmation polymorphism analysis (SSCP), Dideoxy fingerprinting, cleaved amplified polymorphic, sequence (CAPS) and derived CAPS (Gupta PK, 2001 and Lee.GA., 2009).

The use of SNP markers in breeding programs has been growing at a faster pace and so is the development of technologies and platforms for the discovery and screening of SNPs in many crops. SNP markers have become extremely popular in plant molecular genetics due to their genome-wide abundance and amenability for high to ultra high throughput detection platforms. Unlike earlier marker systems, SNPs made it possible to create saturated, if not, supersaturated genetic maps, thereby enabling genome-wide tracking, fine mapping of target regions, rapid association of markers with a trait, and accelerated cloning of gene QTLs of interest. SNP markers are important tool for linkage mapping, map based cloning and marker assistant selection due to the high level of polymorphism. SNPs dominate nature enable the markers to distinguish homozygous and heterozygous alleles (Shaheen et al., 2009). Due to high polymorphism rate SNP markers were to measure the gene mapping, genetic diversity, and construction of genetic map, and analysis for QTL in cotton (Michael et al., 2014 and Hulse Kemp et al., 2015). SNP markers were presented rapidly and efficiently genotyping of large scale by using a next generation sequencing technologies (NGS). SNPs have many advantages like easily accessible, data management, rapid, and flexibility cost. Biallic SNPs added data direct into group and make large database marker information because same type of data used from different genotyping places

Although sequence comparison between two or more individuals is central to the identification of SNPs, its discovery on a genome-wide scale is hampered sure to presence of repetitive sequences that are not aligned to the locations during assembling of a genome. In view of this, methods leading to reduction of genome complexity have been highly useful for discovery of genome-wide SNPs. All the methods of reduced library preparation can be grouped into three broad classes: (i) reduced-representation sequencing, including reduced- representation libraries (RRLs) and complexity reduction of polymorphic sequences (CROPS); (ii) restriction site-association DNA (RAD)-seq/ddRAD-seq; and (iii) low coverage genotyping, include multiplexed shotgun genotyping (MSG) and genotyping by sequencing (GBS). Depending on the need, one of the above classes can be selected. For instance, for the study of crop plants, where no reference genome is yet available, a large number of markers need to be scored

accurately in representation methods are most appropriate (Mir and Varshney, 2012). However. For genotyping applications in QTL mapping and MAS, where parental genotypes are well known and progenies with limited polymorphism are to be sequenced, low-coverage genotyping is considered sufficient for linkage to be inferred, provided that a reference genome is available.

Genotyping by sequencing technology is used for identification and genotyping of SNP molecular markers from a genome. GBS technique development is simple and has ability to reduce the intrication in genome (Elshire et al., 2011). The multiplex libraries are developed for next generation sequencing by using of restriction endonuclease enzyme and with DNA barcoded adapters to detecting a section of genome from genotyping by sequencing (GBS). In GBS one enzyme strategy, methylation sensitive restriction enzyme which does not cut frequently in the repetitive regions. Two different types of adapters were used (Barcode and common). Barcode adaptors contain 4-8 bp unique barcode sequence and complementary sequence for sticky ends generated by restriction enzyme ApekI. The common adapter has only an ApekI compatible sticky end. Adapters were designed so that the ApekI recognition site did not occur in any adapter sequence and are not regenerated after ligation to genome DNA. This technique has shown firstly in varieties and has ability to evolving huge markers (Elshire et al., 2011; Poland et al., 2012). GBS firstly evolve large sequence variants and after that a complete genome sequencing (Poland and Trever, 2012). GBS is a best technique for study of populations because it is help in genetic mapping through which genomic selection can be done widely on large scale (Poland and Trever, 2012). Two enzyme system include one “rare-cutter” and one “common-cutter” enzyme for suitable and uniform complexity reduction.

RAD or ddRAD method provides a reliable means for genome complexity reduction and is based on obtaining the sequence adjacent to a set of particular restriction enzymes recognition sites. The application of high throughput sequencing technology has allowed restriction site-association genome DNA (i.e., RAD tags) for high-density SNP discovery and genotyping was first demonstration by Baird et al., (2008). This involves

digestion of genomic DNA with six to eight base-cutter restriction enzyme, and a barcode adapter is ligated to compatible sticky ends. Before sequencing, DNA sample each with a different barcode are pooled, randomly sheared to a length suitable for the sequencing platform(300-700 bp), and a second adapter is ligated after polishing and filling ends. RAD-tag sequencing has been found very effective for the rapid and large discovery of molecular markers, even in a species with a low polymorphism. For instance, in case of eggplant (*Solanum melongena),* RAD-tag sequencing has resulted in the development of >10,000 SNPs, 1600 indels and 1800 putative SSRs (Brachi et al.,2011). These markers will prove useful for rapid saturation of the best available intraspecific genetic map in eggplant and for the study of comparative genomics analysis within the *Solanaceae* family. In addition, Rad-seq has also been used for the construction of linkage maps in barley (Chutimanitsakun et al.,2011) and ryegrass (Pfender et al.,2011).

Traditional RAD-seq uses restriction enzyme and random shearing to generate fragments for the genomic DNA. However, these are high DNA loss steps and offer little control over fragments that are sequenced. For organisms without a reference genome, a significant portion of the RAD-seq short-comings. In ddRAD-seq, genomic DNA is digested with two restriction enzymes, and the resulting fragments undergo adaptor ligations and precise size selection before sequencing (Peterson et al., 2012). Only a very small fraction of the fragments will be sequenced. These fragments are naturally selected to be from the same genomic regions across individuals. Further, ddRAD requires half as many reads to achieve high confidence SNP calling, because the chance of obtaining duplicate read from the same restriction site are very low. Due to these modifications, ddRAD has become a more economical method used for SNP discovery for species that have no reference sequence.

2.7 Association analysis and gene identifications

The development of molecular markers for the detection and exploitation of DNA polymorphisms in plant systems is one of the most significant developments in the field of molecular biology and biotechnology. Linkage mapping has been a key tool for identifying the genetic basis of quantitative traits in plants. However, for linkage studies,

suitable crosses, sometimes limited by low polymorphism or small population size, are required. In addition, only two alleles per locus and few recombination events are considered to estimate the genetic distances between marker loci and to identify the causative genomic regions for quantitative trait loci (QTL), thereby limiting the mapping resolution. To circumvent these limitations, linkage disequilibrium (LD) mapping or association mapping (AM) has been used extensively (Slatkin, 2008).

AM has the potential to identify a single polymorphism within a gene that is responsible for phenotypic differences. AM involves searching for genotype-phenotype correlations among unrelated individuals. Its high resolution is accounted for by the historical recombination accumulated in natural populations and collections of landraces, breeding materials and varieties. By exploiting broader genetic diversity, AM offers three main advantages over linkage mapping: mapping resolution, allele number and time saving in establishing a marker-trait association and its application in a breeding program (Flint-Garcia et al., 2003). Association analysis has been successfully applied to identify marker–trait associations in different crops. Additionally, association analysis uses preexisting germplasm such as landraces, modern cultivars, and advanced breeding lines to detect potential associations between molecular markers and traits of interest (Zhu et al., 2008). In wheat, an AM approach has been used to map agronomic and quality traits such as kernel size and milling quality (Breseghello & Sorrels 2006). Grain yield, high-molecular-weight glutenin, resistance to foliar diseases, and Fusarium head blight (FHB) resistance. In an investigation, twelve cadmium-tolerant genotypes were selected among 472 worldwide collection of natural rapeseeds based on the criterion of relative radicle length (RRL) >60%. Nine associated SNPs localized in four QTLs were identified by genome-wide association study (GWAS) in which 60 K Brassica Infinium SNP array was used. Seven candidate genes for Cd tolerance, including HIPP27, EXPB4, EMB1793 and CDSP32 orthologues, were obtained in these loci. (Zhang et al., 2018).

Materials And Methods

CHAPTER-3

MATERIAL AND METHODS

This chapter covers the experimental details regarding materials, protocols and techniques used during the course of present investigation. The present studies were carried out in the School of Biotechnology, Sher-e-Kashmir University of Agriculture Sciences & Technology-Jammu during 2018-2019.

3.1 Materials

3.1.1 Plant Material and Experimental Site

A total of 50 *Brassica juncea* genotypes (Table 2) of European genepool were used for biochemical and molecular characterization. To collect seed material for biochemical analysis, the genotypes were sown in alpha-lattice design (two replications) at research farm of School of Biotechnology, Sher-e-Kashmir University of Agricultural Sciences and Technology, Chatha, Jammu. Each genotype was grown in a plot of 3 mt sq. (four rows of 2 mtr each with a row to row spacing of 50cm). The plot to plot distance was maintained at 75cm. Plots were kept free from weeds, diseases, and insect/pests throughout the cropping cycle by following standard agronomic practices. The location of field trial was research Farm, School of Biotechnology, Sher-e-Kashmir University of Agricultural Sciences and Technology, Main Campus Chatha, Jammu.

Table 2: List of *Brassica* genotypes used during present study.

S.No.	Genotype
1.	Neosypajuscajasla 2
2.	Volgogradskaja 189/191
3.	Skorospelka
4.	VNIIMK 351
5.	VNIIMK 405
6.	Zeltosemiannaja 230
7.	Zaria
8.	Donskaja 4
9.	PGR 3330
10.	PGR 3383

11.	Commercial Brown Mustard
12.	Lethbridge 22A
13.	Blaze
14.	Domo
15.	AC Vulcan
16.	Yanagawa Shirokuki
17.	Miike Akachirimen
18.	Sendai Bashouna
19.	Burgonde
20.	Ekla
21.	Primus
22.	Stoke
23.	Jubilejnaja
24.	Skorospelka 2
25.	PGR 12568
26.	74/5
27.	PGR 12573
28.	PGR 12574
29.	Stepniacka
30.	J/807/1/6
31.	J/807/12/1
32.	J/817/2
33.	J/824
34.	J/824/6
35.	PGR 12586
36.	Cutlass
37.	Scimitar
38.	SRS 319
39.	I-49-24
40.	M.Br.4
41.	63-0134-68
42.	64-1398-69
43.	Bass
44.	R 871
45.	Kafiav V Zagora
46.	German accession No. 114
47.	Zem 1
48.	Skorospieka II
49.	EC 187711
50.	EC 491584

PLATE-1

Field Trials at SKUAST-Jammu for biochemical and molecular analysis of 50 diverse *Brassica juncea* genotypes.

3.1.2 Important chemicals used for biochemical and molecular analysis

A number of reagents and buffers used for various applications were prepared as following:

(i) **1M Tris pH 8.0**: A total of 121.1g Tris base was dissolved in 800ml of water. pH was adjusted to 8.0 by adding 42ml of concentration HCL. The solution was allowed to cool to room temperature before making the final adjustment of the pH. The final volume was make up to 1 L with water and sterilize using an autoclave.

(ii) **0.5 EDTA**: A total of 186.1g EDTA was dissolved to 800ml of double distilled water. Sodium hydroxide were used to bring the pH to 8.0 followed by filtering with 0.5 micron filter and autoclave). (**Note:** EDTA would not completely dissolve until the pH is around 8).

(iii) **5M NaCl**: About 292.2g of NaCl was dissolved in exactly 880 ml ultrapure water. The volume was adjusted to 1L with ultrapure water and sterilized by heat.

(iv) **10X TE Buffer:** About 100ml of Tris-Cl (pH 7.5) and 20 ml of 0.5 EDTA was mixed and final volume was adjusted to 1L with distilled water.

(v) **Choloroform: Isoamyl alcohol (24:1):** About 24 volumes of chloroform using a measuring cylinder was added and transferred to a dark bottle and then measure and add volumes of isoamyl alcohol into it. Mix the contents properly and store.

(vi) **Phenol: Chloroform: Isoamyl alcohol (25:24:1):** About 25 volumes of phenol, 24 volumes of chloroform and 1 volume of isoamyl alcohol was added and transferred in a dark bottle. The content were mixed properly and stored.

(vii) **DNA loading dye:** About 30% (v/v) Glycerol, 0.25% (w/w) Bromophenol blue and 0.25% (w/v) Xylene cyanol were stirred and filtered.

Table 3:- List of chemical solution used for DNA extraction		
Chemicals/ Stock	**Final concentration**	**Amount to be added for 100ml**
B-mercaptoethanol	0.2(w/v)	200.00µl
CTAB powder	2.0%(W/V)	2.00g
NaCl	1.4 M	8.12g
1M Tris- Cl, pH -8.0	100Mm	10.00ml
0.5 M EDTA	20mM	4.00ml

(ix) **DNA extraction buffer:** The DNA extraction buffer was prepared using ingredients

(x) **Electrophoresis Buffer:** An electrophoresis buffer of 20x TAE was prepared by using reagents listed in Table 4. The volume was made up to 100ml by double distilled water, autoclaved and stored at room temperature.

Table 4:- List of components used to make buffer	
Components	**Volume**
Tris base	24.2g
Glacial Acetic Acid	5.7g
0.5 M EDTA	10ml

(xi) **Ethidium Bromide:** For visualization of genomic DNA on agarose gels, the ethidium bromide (EtBr) dye was prepared by dissolving 10 mg of EtBr in one ml of distilled water.

3.2 Methods

3.2.1 Seed Sowing

The crop was sown at the Biotechnology Experimental Farm, SKUAST-Jammu. The experimental material for the study comprised of 50 genotypes. The genotypes were sown in the second fortnight October 2018 in field.

3.2.2 Genomic DNA extraction:

The young and fresh leaves were collected, washed and then used for genomic DNA extraction. For DNA extraction leaf tissue (1g) was grounded to fine powder in liquid nitrogen in a pre chilled pestle and mortar. The frozen powder was transferred to a 50 ml autoclaved oakridge tube. CTAB extraction buffer (7ml) was added to each tube and the tubes were inverted several times with gentle shaking and leaf powder was completely suspended in CTAB buffer. The tubes were incubated at 65◦c in a water bath for 60 minutes. The tube contents were mixed gently by inverting the tubes after 10-15 minutes. Following this, the tubes were cooled to room temperature and 7ml of chloroform; isoamyl; alcohol, in the ratio of 24:1, was added to each tubes; and the contents were mixed gently by inverting the tubes several times. The tubes were centrifuged at 10,000 rpm for 10 minutes at 20°C and the upper clear aqueous supernatant was removed and transferred into fresh Oakridge tubes. The above two were repeated twice to ensure that there no more protein contamination left. The absolute ethanol was added to each tube and the contents were mixed by gently inverting the tubes several times and the tubes were incubated overnight at -20°C. After incubation the samples were centrifuge at 10,000 for 15 minutes at 4°C. The supernatant was decanted off without disturbing the DNA pellet at the bottom of the tubes. One ml of 70% ethanol was added to wash the pellet and the pellet was air dried for 15-20 minutes, in a laminar flow / hot air oven (30°C) to hasten the air drying. About 200µl of TE buffer (pH-8.0) was added to the pellet and pellet was dissolved completely.

3.2.3 DNA Purificaion:

To the dissolved DNA obtained after extraction, two µl of RNase-A (10mg/ml) added to the tube containing 200µl of extracted DNA dissolved in TE buffer. The tubes were incubated for 1hr at 37°C in a water bath for RNAse treatment. After incubation, 100µl of phenol: chloroform: isoamyl alcohol in the ratio of 25:24:1 was added to the tubes and mixed by inverting the tubes. The samples were centrifuged at 10,000 rpm for 15 minutes at 4°C. The upper phase was transferred to new tubes. Nearly double volume of ethanol was added to each tube and tubes were kept overnight at -20°C. The suspension was again centrifuged at 10,000 rpm for 15 minutes at 4°C. The supernatant

was discarded, Pellet was washed with 70% ethanol and air dried. The DNA pellet was re-suspended in 100µl of TE buffer (pH 8.0).

3.2.4 Quantification and quality check of DNA

The DNA samples were quantified using a small samples bio spectrophotometer (Eppendorf Germany). The absorbance was read in spectrophotometer by setting the blank against Tris-EDTA buffer. The absorbance was taken at 260nm and 280nm. The ratio of Abs260/Abs280 provides an estimate of the purity of nucleic acids. Pure preparation of DNA has Abs260/Abs280 ratio of 1.8-2.0. The quality of DNA samples was checked by running the samples on 0.8% agarose gels and looking for any signs of fragmentation as DNA degradation. Following stock solutions were used for agarose gel electrophoresis.

3.3 Extraction of Fatty Acids and estimation of Erucic acid

A total of 20-25 seeds were grounded in a 5ml culture tube to a fine powder using pestle and mortar. To it, 0.5 ml of hexane was added and incubate at room temperature for 16 hours (overnight incubation). After incubation, transfer hexane supernatant to a fresh tube and 0.5 ml of sodium methoxide was added (prepared by adding 80mg of NaOH to 100ml of methanol). Then incubate the mixture at room temperature for 45 minutes. After incubation, 750ml of NaCl solution was added (prepared by adding 8gms of NaCl to 100ml of distilled water) and vortex. After vortexing, the mixture was left on the table for approximately 30 minutes, and the upper phase was collected and used for GC injection. The peak area was used for calculation of percent erucic acid by the inbuilt software of GC system.

3.4 ddRAD libraries for Next Generation Sequencing and SNP Identification

High quality DNA was sent to the service provider for the development of ddRAD libraries to be put on next generation sequencing. The libraries were prepared as per manufacturer (Illumina Inc., USA) protocol by adding two barcoded adapters to the each sample followed by pooling of sample libraries. Following sequencing, the service provider analysed sequences using DOCENT sequence analysis pipeline (Puritz, J.B., et al. 2014). Only high quality sequence data was for sequence assembly and SNP

identification. The SNPs were reported in the form a SNP sample for further use in downstream analysis.

3.5 Association mapping

Association mapping was performed by using TASSEL software.

Results

CHAPTER-4

RESULTS

The results of the various experiments carried out for the current research work are presented here. The experiments included biochemical analysis of erucic acid, development and genotyping of SNPs for the identification of genes responsible for erucic acid in *Brassica juncea.*

4.1 Molecular Analysis

4.1.1 DNA isolation:

The genomic DNA was isolated using standard protocol (by Doyle and Doyle method). And quality was check by agarose gel electrophoresis and quantification by Bio spectrophotometer.

4.2 Estimation of Erucic Acid

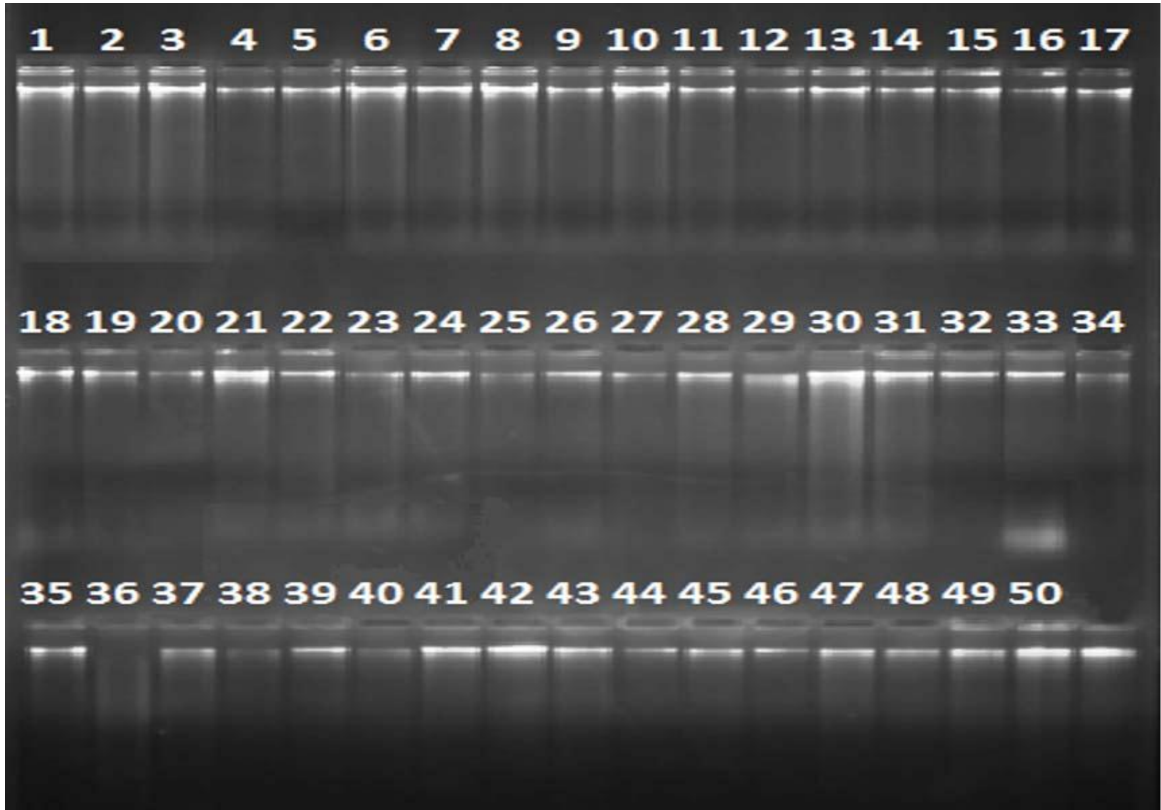
A total of fifty (50) genotypes of European genepool were used for biochemical estimation of erucic acid, however, the data could be generated for 49 genotypes only. The estimation was carried out using gas chromatography (GC) based method. The samples were injected into GC at ICAR-DRMR, Bharatpur and the retention time for different fractions of oil including erucic acid was extrapolated into percent erucic acid present in the oil samples. The results of fifty genotypes are given in the (Table 5). Based on the results of GC analysis, the samples were found to have erucic acid in different percent ranging as given in (Table 6). Three genotypes with least percent erucic acid were Zem I (8.08 %), Burgonde (11.43%) and Volgogradskaja 189/191 (11.99%).

Table 5: Percent erucic acid among European genepool genotypes

S.No.	Sample Name	Erucic acid
1.	Neosypajuscajasla 2	21.98
2.	Volgogradskaja 189/191	**11.99**
3.	Skorospelka	26.44
4.	VNIIMK 351	17.81
5.	VNIIMK 405	49.57
6.	Zeltosemiannaja 230	19.41
7.	Zaria	22.29

8.	Donskaja 4	21.2
9.	PGR 3330	24.29
10.	PGR 3383	33.31
11.	Commercial Brown Mustard	25.27
12.	Lethbridge 22A	16.86
13.	Blaze	25.28
14.	Domo	15.6
15.	AC Vulcan	20.82
16.	Yanagawa Shirokuki	18.94
17.	Miike Akachirimen	24.23
18.	Sendai Bashouna	29.39
19.	Burgonde	**11.43**
20.	Ekla	17.79
21.	Primus	12.6
22.	Stoke	20.61
23.	Jubilejnaja	41.49
24.	Skorospelka 2	31.28
25.	PGR 12568	16.99
26.	74/5	27.04
27.	PGR 12573	15.93
28.	PGR 12574	27.37
29.	Stepniacka	17.97
30.	J807/1/6	19.48
31.	J/807/12/1	14.76
32.	J/817/2	30.2
33.	J/824	33.67
34.	J/824/6	36.94
35.	PGR 12586	24.88
36.	Cutlass	23.65
37.	Scimitar	16.96
38.	SRS 319	19.29
39.	I-49-24	16.49
40.	M.Br.4	17.51
41.	63-0134-68	27.14
42.	64-1398-69	40.59
43.	Bass	22.38
44.	R 871	24.24
45.	Kafiav N Zagora	18.75
46.	German accession No. 114	20.11
47.	Zem 1	**8.08**
48.	Skorospieka	40.01
49.	EC 187711	41.11

PLATE-2



DNA quality analysis of *Brassica juncea* genotype

PLATE-3

(Here the sample was injected)

Estimation of erucic acid by GC analysis

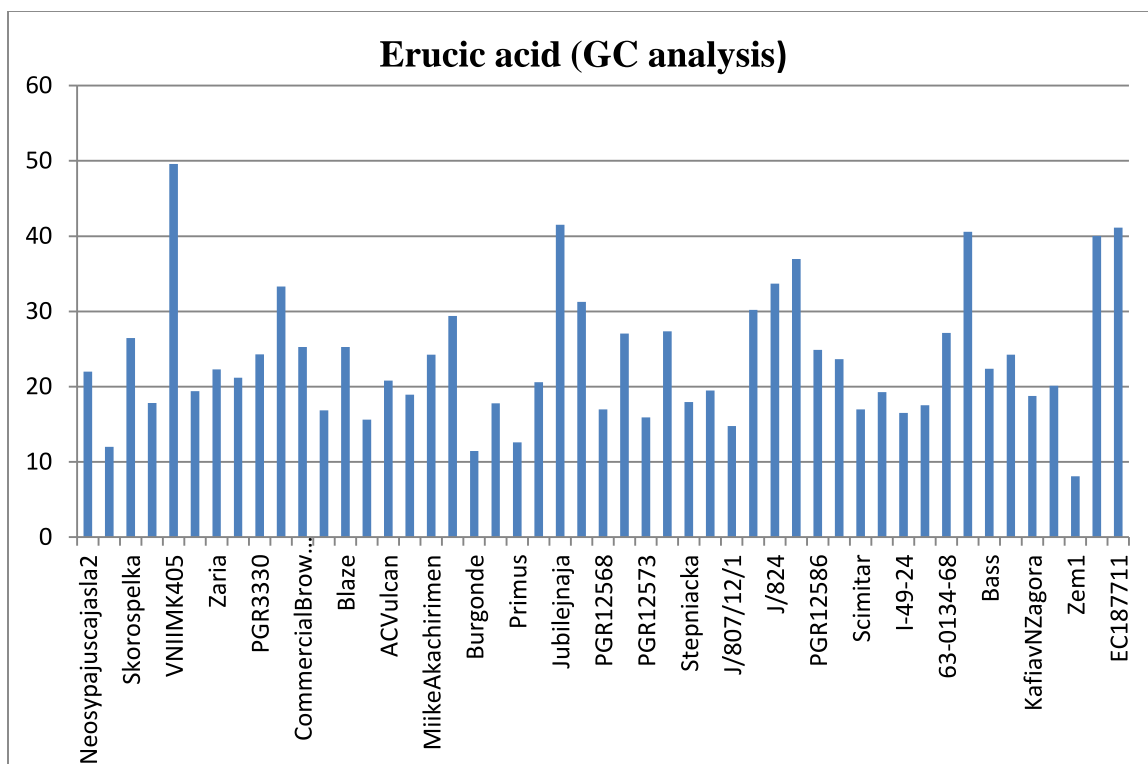


Figure 2: Estimation of erucic acid by GC analysis

Table 6: Estimates of erucic acid among 50 genotypes		
S.No.	**Erucic acid (%age)**	**Percent genotypes (Number of genotypes)**
1	Less than 10	2.04 (1)
2	Between 10 and 19.99	38.78 (19)
3	Between 20 and 29.99	38.78 (19)
4	Between 30 and 39.99	10.20 (5)
5	More than 40	10.20 (5)

4.3 ddRAD-sequencing and SNP development

4.3.1 NGS library preparation:

The samples of 50 genotypes were processed on high-throughput next-generation sequencing (NGS) platform through a service provider. The high quality one microgram DNA of each sample was digested with two restriction enzyme namely, *Sph*l and *Mluc*1 to generate smaller sized fragments. The digested products, after purifications, were

ligated with proprietary Illumina adapter (P1 and P2) sequences. The ligated products for individual samples were electrophoresed on 2% agarose for size selection of adapter ligated fragments in the range of 300-400bp and same were validated using microfluidics-based Agilent Bioanalyzer electrophoresis system (Fig. 3).

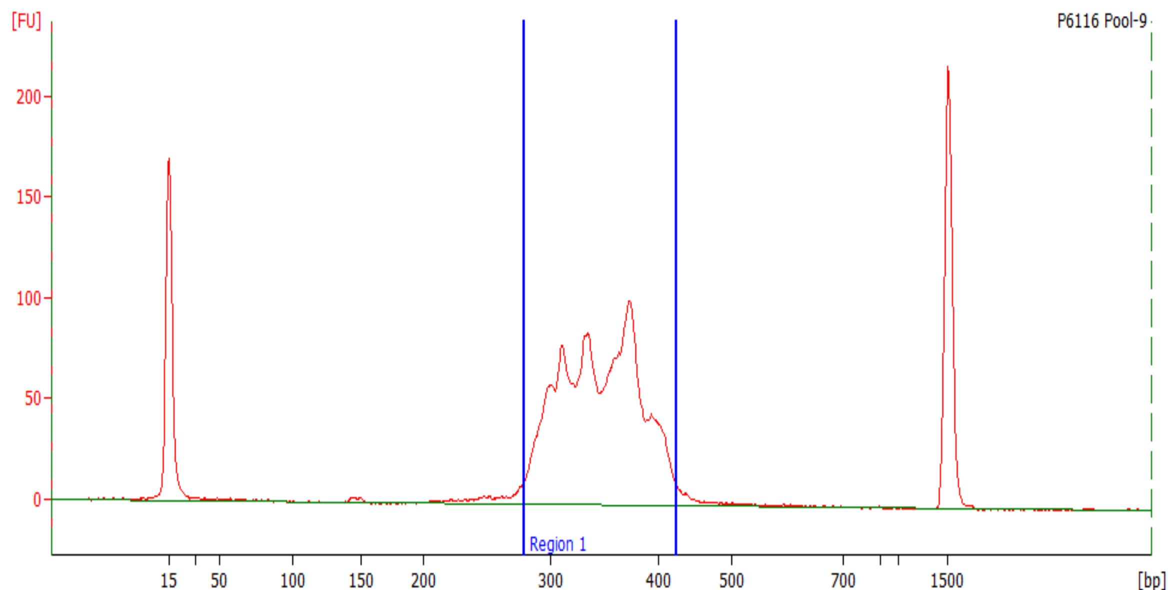


Figure 3: Microfluidics-based analysis of a pool of five samples indicating fragment size distribution

4.3.2 Sequence data QC analysis:

The sample libraries were processed for sequencing using Illumina TrueSeq chemistry on Illumina HiSeq X platform at a service provider facility. The samples were sequenced using paired-end chemistry with a read length of 151 bases. In total, nearly 104.4 million reads amounting to more than 14.4Gbs of sequence data. For all the samples more than 90% data had a Q score of greater than 30. Out of 50 samples, the data for 43 samples was found to be more than 100Mb.

Table 7: Summary of ddRAD-sequence data of European genotypes

S.No.	Sample Name	Number of reads	Number of bases (Mb)	GC (%)	Percent reads (Q>30)
1	Neosypajuscajasla 2	1886458	262.37	41.82	93.25
2	Volgogradskaja 189/191	1606336	220.59	41.84	92.55

3	Skorospelka	2569516	356.80	42.85	93.03
4	VNIIMK 351	3390022	471.50	41.50	93.01
5	VNIIMK 405	3049290	411.42	41.80	92.82
6	Zeltosemiannaja 230	1862110	256.72	41.96	92.4
7	Zaria	1914224	259.01	41.44	92.85
8	Donskaja 4	1928970	267.02	41.97	92.33
9	PGR 3330	2366240	334.06	41.53	93.10
10	PGR 3383	1632050	227.86	41.99	92.74
11	Commercial Brown Mustard	2839800	399.66	41.48	92.95
12	Lethbridge 22A	3033468	427.07	41.26	93.07
13	Blaze	1118604	155.98	42.04	92.65
14	Domo	2894826	398.72	41.71	93.13
15	AC Vulcan	3218882	448.99	41.53	93.08
16	Yanagawa Shirokuki	1741582	237.49	40.72	93.16
17	Miike Akachirimen	2332424	328.97	42.86	92.56
18	Sendai Bashouna	1057852	145.55	42.38	93.26
19	Burgonde	746776	100.22	41.94	93.36
20	Ekla	2813822	400.25	41.16	93.15
21	Primus	1698704	232.02	41.94	92.71
22	Stoke	3353408	474.93	41.15	92.93
23	Jubilejnaja	2872798	406.31	41.19	93.00
24	Skorospelka 2	3967748	545.41	41.03	92.77
25	PGR 12568	118398	16.15	42.22	93.59
26	74/5	2518362	353.47	41.50	93.34
27	PGR 12573	6280	0.83	42.06	93.74
28	PGR 12574	115076	16.16	37.80	93.21
29	Stepniacka	2908934	406.89	41.64	93.28
30	J807/1/6	2220442	310.76	41.66	93.34
31	J/807/12/1	2326718	322.62	41.53	93.48
32	J/817/2	1259710	176.94	42.26	92.80
33	J/824	1854832	261.12	41.37	93.55
34	J/824/6	2113928	295.41	41.40	93.48
35	PGR 12586	488820	63.11	41.60	93.55
36	Cutlass	1525790	216.27	41.73	92.735
37	Scimitar	2012876	278.08	41.685	93.265

38	SRS 319	699462	97.57	42.81	92.575
39	I-49-24	2034816	278.75	41.65	93.875
40	M.Br.4	2542760	353.14	42.015	92.835
41	63-0134-68	175232	23.48	42.74	93.995
42	64-1398-69	4635666	641.01	40.405	93.005
43	Bass	2900456	402.45	41.88	92.73
44	R 871	2848634	383.66	42.015	92.485
45	Kafiav N Zagora	3894624	547.99	41.68	95.275
46	German accession No. 114	1469372	204.86	41.865	93.93
47	Zem 1	2891310	401.44	41.865	92.775
48	Skorospieka	2378968	320.7	41.11	92.98
49	EC 187711	698682	95.79	42.605	94.065
50	EC 491584	1928318	261.81	41.915	93.84

4.4 SNP discovery among European genepool samples:

The sequence reads were aligned to the reference genome (ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001/687/265/GCA_001687265.1_*Brassica_juncea*_var_tumida_T84-66_v1/GCA_001687265.1_*Brassica_juncea*_var_tumida_T8466_v1_genomic.fna.gz) using dDocent pipeline by the sequence service provider to identify genome-wide SNPs. A total of 127,004 SNPs were identified among 50 European genotypes. Out of these, a total of 98401 SNPs were distributed on all 18 chromosomes and the remaining 28603 SNPs were found on contig sequences of the reference genome. The distributiuon of SNPs across the genome is given in (Table 8).

Table 8: Chromosome-wise distribution of SNPS

S.No.	Chromosome/ Contig	Reference Assembly Chromosomes	Number of SNPs
1	A01	CM007185.1	5947
2	A02	CM007186.1	3574
3	A03	CM007187.1	5988
4	A04	CM007188.1	4482
5	A05	CM007189.1	5341
6	A06	CM007190.1	5236
7	A07	CM007191.1	3864
8	A08	CM007192.1	4779
9	A09	CM007193.1	8845
10	A10	CM007194.1	3837

11	B01	CM007195.1	3598
12	B02	CM007196.1	7554
13	B03	CM007197.1	9262
14	B04	CM007198.1	4418
15	B05	CM007199.1	6154
16	B06	CM007200.1	2692
17	B07	CM007201.1	4621
18	B08	CM007202.1	8209
19	Contigs		28603

4.5 Marker-Trait association analysis for erucic acid

The biochemical data for erucic acid estimated using gas chromatography and SNP data generated using ddRAD-sequencing approach was used for the identification of marker trait association. For association analysis, only SNPs localised on chromosomes were used and those present on unlocalised contig sequences were excluded for further analysis. Based on analysis, a total of 125 significant marker-trait associations were found on various chromosomes. Highest number of 22 associations with erucic acid was found on A03 followed by 15 and 12 on B07 and B03, respectively. Overall, 73 associations were found on A-genome chromosomes and 52 on B-genome chromosomes. Ten most significant associations (p-value < 1.0 x E-04) were found on chromosomes listed in (Table 9, Fig. 4). Two SNPs found within 15nt (at positions 4034447 and 4034461) on A03 were the most significant ones with a p-value < 4.84 E-06.

Table 9: Ten most significant associations on the basis of P-value

S.No.	Chromosome	Position	Marker_F	P-value	Marker_Rsq
1	A03	4034447	12.59029	4.84E-06	0.46231
2	A03	4034461	12.59029	4.84E-06	0.46231
3	A02	13452513	12.69328	4.45E-05	0.36171
4	A03	35070676	12.6242	4.64E-05	0.36046
5	B01	4623114	9.80155	4.77E-05	0.4015
6	A01	22955715	12.19666	6.10E-05	0.3526
7	A04	13586835	11.90379	7.38E-05	0.34711
8	A01	17529004	9.0879	8.91E-05	0.38361
9	B05	11474465	9.0885	8.91E-05	0.38362
10	B05	39675514	11.47104	9.79E-05	0.33882

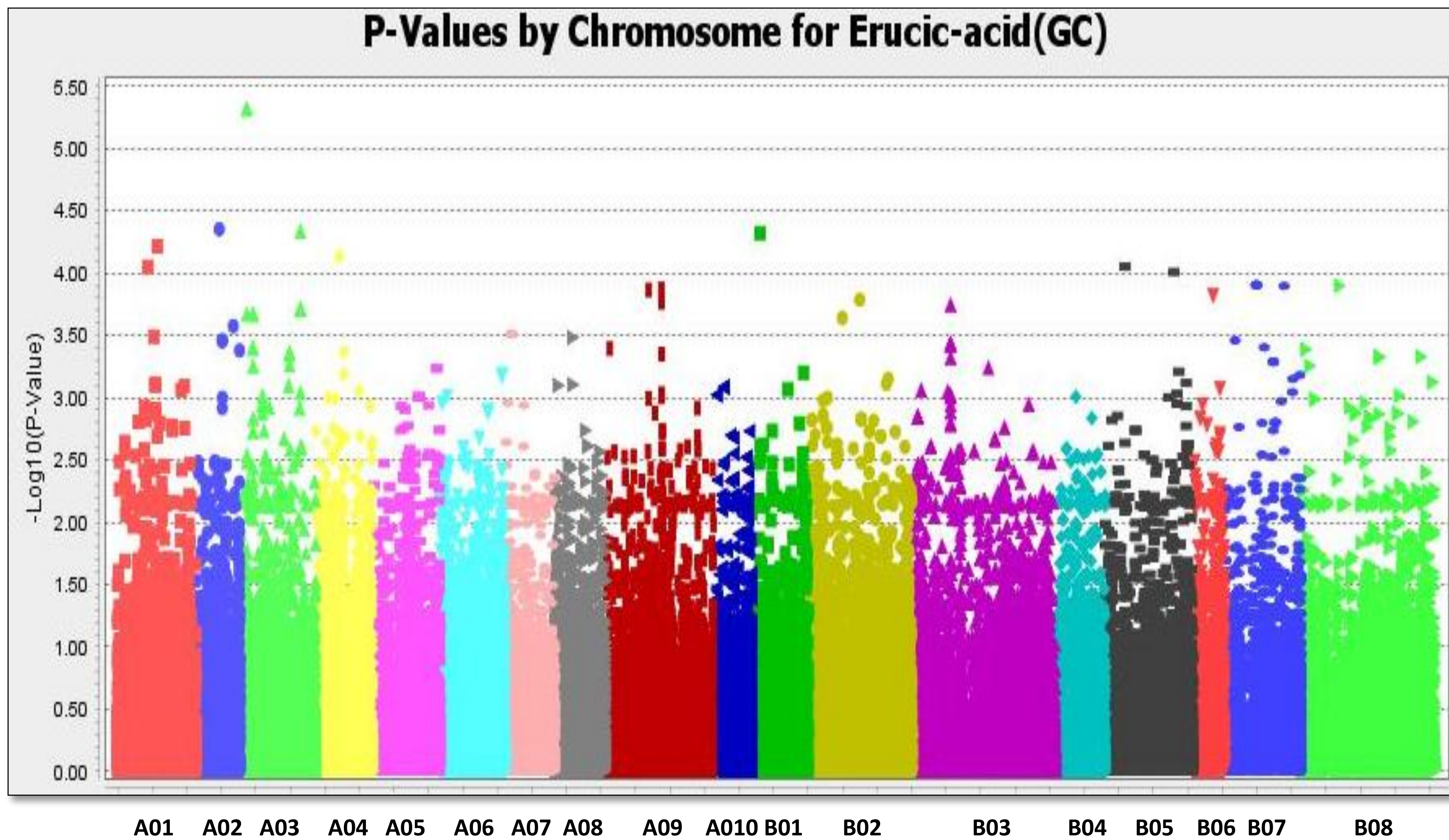


Figure 4: Manhattan plot of log (P) values of marker-trait association for erucic acid

Discussion

CHAPTER-5

DISCUSSION

Development of low erucic acid containing varieties is extremely important as the animals fed with high erucic acid diet have been reported to have developed lipidosis. For this, identification of genes controlling trait of interest is a pre-requisite. The research was carried out to identify SNPs and genomic regions of *Brassica juncea* associated with erucic acid. The analysis was carried out using association analysis, a method based on the associations of genomic regions with the trait of interest in diverse unrelated individuals. The above mentioned results of the present study are discussed in the following context.

5.1 Profiling of erucic acid among European genepool

The biochemical analysis revealed that the germplasm included in the present study was highly diverse as a range of percent erucic acid was estimated among the studied genotypes. Previous studies for erucic acid have estimated that most Indian cultivars have high erucic acid (50%) to the total oil content (Pandey et al. 2013). High erucic acid in oil is considered anti-nutritional leading to the development of lipidosis. The European gene pool has been known to possess genes limiting total erucic acid concentration in edible oils. Therefore, the current study involved use of European genepool for profiling for erucic acid. The results have shown that nearly 40% genotypes had total erucic acid content to less than 20 percent. Although, the desirable erucic acid concentration for the development of 'zero' canola varieties is less than 2 percent. In the present study, three genotypes with low erucic acid can be an important genetic stock for the development and improvement of varieties for erucic acid through traditional and mutational breeding.

5.2 Use of ddRAD for SNP development in *Brassica juncea*

The development of SNPs requires comparison of sequence from the same locus across different individuals to identify nucleotide variations. While, this PCR based amplification of different individuals using common set of primers used to be the most

preferred method; the advent of next generation sequencing technology has completely revolutionised the process of SNPs identification in not only referenced genomes but also in genomes of economically lesser important species. The ddRAD first reported its use for simultaneous discovery and genotyping of diverse genomes (Peterson et al 2014). The method has found applications for SNPs development in species from both plants and animals (Laila et al. 2019; Jaiswal 2019).

The current study in *Brassica juncea* might well be the first report of application of ddRAD-sequencing for SNP development in *Brassica juncea*. A total of 98401 SNPs identified in the current study were placed on the *Brassica juncea* genome at a frequency of 9.40KB (assuming genome size of 925MB), thereby giving a very dense landscape of the *Brassica juncea* genome. The results also revealed that another 28,603 SNPs identified on contigs would be available to for mapping on the individual chromosomes after these contigs are assigned to the individual chromosomes.

5.3 Significant marker-trait associations and localization of genes for erucic acid

The association analysis is based on linkage disequilibrium (LD) between alleles of adjoining gnomic regions and their association with traits controlled by genes present in those genomic regions. In the present study, the association analysis revealed that presence of 125 highly significant SNPs associated (p value $< E^{-04}$) with erucic acid on all chromosomes of *Brassica* juncea indicating that genes with major and minor effects for erucic acid are present on all the chromosomes. An association analysis study in Brassica napus (AACC) has identified significant associations on A08, A09 and A10 chromosomes (Qu et al., 2017). We have also found significant associations on two genomic regions on A08 with a p-value ranging from 3.29E-04 to 7.98E-04 and phenotypic variance of approximately 27 percent. The genomic regions identified by Qu et al. (2017) localised to 5.21Mb and 5.93 Mb on A08 chromosome; and the current study has localised the regions for erucic acid to 2.2Mb and 10.55Mb on A08 chromosome. The differences in genomic regions between the two studies could be due to difference in the number of samples used for association analysis. The most number of significant associations were localised on A03, B07 and B03 indicating that these chromosomes are likely harboring maximum number of genes for erucic acid in *Brassica juncea*.

The significant associations found in the current study would form a very important genomic resource for the development of new varieties and also for improvement of the existing varieties through marker-assisted selection (MAS) in future breeding programmes.

Summary And Conclusions

CHAPTER-6

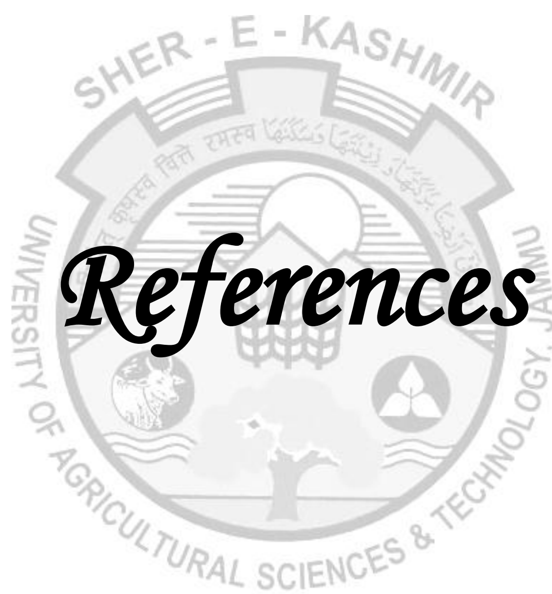
SUMMARY AND CONCLUSIONS

The salient features of the present study entitled **"Genomic profiling for erucic acid in European germplasm of *Brassica juncea* (L)."** are summarized below:

- A collection of 50 diverse *Brassica* juncea genotypes of European genepool were used in the present study.
- These 50 diverse genotypes were subjected to biochemical characterization of erucic acid using gas chromatography and molecular characterisation through ddRAD sequencing for SNP developments.
- The samples were prepared at the School of Biotechnology for biochemical and molecular characterisation.
- Based on GC analysis, three genotypes namely Zem I (8.08%), Burgonde (11.43%) and Volgogradskaja 189/191 (11.99%) were found to have lower erucic acid content among the European gene pool genotypes used in the current study.
- High-resolution ddRAD-sequencing revealed a total of 78,328 SNPs among different genotypes at a read depth of 10 with 90% sequence data having $Q > 30$.
- A total of 125 significant marker-trait associations ($p < 0.001$) were identified in the association analysis.
- Overall, A-genome carried significant association on two genomic regions on A08 with a p-value ranging from 3.29E-04 to 7.98E-04 and phenotypic variance of approximately 27 percent.
- Most significant associations were localised on A03, B07 and B03 indicating that these chromosomes are likely harbouring maximum number of genes for erucic acid in *Brassica juncea*.
- The significant SNPs would form a valuable resource for MAS in future breeding programmes not only for erucic acid, but for other traits as well.

From the results of this study, it can be concluded that the associations found would form a very important genomic resource for the development of new varieties and also for improvement of the existing varieties through marker-assisted selection (MAS) in future breeding programmes.

References



REFERENCES

Afiah, S. A., Abdelsalam, A.Z.E., Kamel, E.A., Dowidar A.E. and Ahmed, S.M. 2007. Molecular genetic studies on canola crosses under maryout conditions, *African crop Science Conference Proceeding*, **8**: 633-642.

Anuradha, M., Pragyandip, D., Murthy, P.N. and Poonam, K. 2012. A classical review on Rajika (*Brassica napus* L.). *Canadian Journal of Plant Science*, **61**:199-203.

Appelqvist, L.A., Ohlson, R. and Sprague, M.A. 1973. Rapeseed cultivation, composition, processing and utilization. *Soil science*, **116**(6):453.

Ayeh, K.O., 2008. Expressed sequence tags (ESTs) and single nucleotide polymorphisms (SNPs). Emerging molecular marker tools for improving agronomic traits in plant biotechnology, *African Journal of Biotechnology*, **7**(4):331-341.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. and Johnson, E.A. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS One*, **3**(10):e3376.

Bao, X., Pollard, M. and Ohlrogge, J. 1998. The biosynthesis of erucic acid in developing embryos of *Brassica rapa*. Journal of *Plant Physiology,* **118**: 183-190.

Barchi, L., Lanteri, S., Portis, E., Acquadro, A., Vale, G., Toppino, L. and Rotino, G.L. 2011. Identification of SNP and SSR markers in eggplant using RAD tag sequencing. *BMC Genomics*, **12**(1):304.

Batley, J., Jewell, E. and Edward, D. 2007. Automated discovery of single nucleotide polymorphism (SNP) and simple sequence repeat (SSR) molecular genetic markers. *Plant bioinformatics, Methods in Molecular Biology*, **56**:473-494.

Bozzini, A., Calcagno, F. and Soare, T. 2007. ‘‘Sincron’’, a new *Brassica carinata* cultivar for biodiesel production. *Helia,* **46**:207–214.

Breseghello, F. and Sorrells, M. (2006). Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics*, Vol. 172, No**.** 2, 1165-1177.

Ching, A.D.A., Caldwell, K.S., Jung, M., Dolan, M., Smith, O., Tingey, S., Morgante, M. and Rafalski, A.J. 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines, *BMC Genetics,* **3**(1): 1.

Chutimanitsakum, Y., Nipper, R.W., Cuesta-Marcos, A., Cistue, L., Corey, A., Filichkina, T., Johnson, E.A. and Hayes, P.M. 2011. Construction and application for QTL analysis of a Restriction Site Association DNA (RAD) Linkage map in barley. *BMC Genomics*, **12**(1):4.

Das, R., Bhattacherjee, C. and Ghosh, S. 2009. Preparation of mustard (*Brassica juncea* L.) Protein isolate and recovery of phenolic compounds by ultrafiltration. *Industrial & Engineering Chemistry Research,* **48**(10):4939-4947.

Duke, J.A. and Wain, K.K. (1981). Medicinal plants of the world. *Computer index with more than 85,000 entries*.3 vols.

Elshire, R.J., Glaubitz, J.C. and Sun, J.A. 2011. A robust, simple genotyping by-sequencing (GBS) approach for high diversity species. *Public Library of Sciences*, **6**:e19379.

Everitt, J.H., Lonard, R.L. and Little, C.R. 2007. Weeds in south texas and northern Mexico. Lubbock: Texas Tech University Press. ISBN 0-89672-614-2. Evidence and mechanistic basis. *Pharmacological Research*, **55**(3):224-236.

Flint-Garcia, S., Thornsberry, J. and Bukler, E. 2003. Structure of linkage disequilibrium in plants. *Annual Review Plant Biology*, Vol. 54, 357-374

Fsanz. 2003. Erucic acid in food: A toxicological review and risk assessment. Report No: *Technical report series no*. 21.

Gangapur, D.R., Prakash, B.G. and Hiremath, C.P. 2010. Genetic diversity analysis of Indian mustard (*Brassica juncea* L.), *Electronic journal of Plant Breeding*, **1**: 407-413.

Gopalan, C., Krishnamurthi, D., Shenolikar, I.S., and Krishnamachari, K.A.V.R. 1974 Mycocardial changes in monkeys fed mustard oil. *Annals of Nutrition and Metabolism,* **16**: 352-365.

GRDC 2009 Canola best practice management guide for south-eastern Australia. *Grains Research & Development Corporation.*

Gupta, P., Rustgi, S. and Kulwal, P. 2005. Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Molecular Biology*, Vol. 57, No. 4, 461-485.

Gupta, P.K., Roy, J.K. and Prasad, M. 2001. Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Journal of Current Science*, **80**:524 –35.

Gupta, S., Sangha, M.K., Kaur, G., Atwal, A.K., Kaur, P., Kumar, H., Banga, S. and Banga, S.S. 2012. Varition in glucosinolate and tocopherol concentration in a germplasm collection of *Brassica juncea* (L.) *Czern.* *Journal of Analytical Biochemistry*, **1**-7.

Gupta, V., A. Mukhopadhyay, N. Arumugam, Y.S. Sodhi, D. Pental and A.K. Pradhan 2004 Molecular tagging of erucic acid trait in oilseed mustard (*Brassica juncea*) by QTL mapping and single nucleotide polymorphisms in *FAE1* gene. *Theoretical and Applied Genetics*, **108**: 743–749.

Hearne, S.J., Lorenzen. J., Town, C. and Zhuang, E. 2008. EST derived genomic resources for Musa, IITA Report.

Hedrick, P.W. 2005. Genetics of populations. 3 ed. Jones and Bartlett, Sudbury, England.

Hemingway, J.S. 1976. Muatard: *Brassica* spp. and Sinapusalba (Cruciferae). In N.W. Simmonds (ed.) *Evolution of crop plants. Longman. London pp*. **56**-9.

Hulse-Kemp, A.M., Jana, L., Joerg, P., Hamid, A., Ramesh, B., David, D.F., James, F., Marc, G., Steve Hague, L.L., Hinze, K.J., Kochan, P.K., Riggs, J.A.S., Joshua, A.U., Mauricio, U., Shirley, S.W., Qian-Hao, Z., Sumit, K.B., Archana, B., John, J.B., Robert, L.B., Michel, C., Michael, A.G., David, B.H., Md, S. I., Johnie, N., Jenkins, D.C.J., Jean-Marc, L., Danny, J. L., Richard, G.P., Alan, E., Pepper, J.A., Poland, Krishan, M.R., Samir, V.S., Sunil Kumar, S., Andrew, S., Jen, M., Taylor, F.W., Scott, M.Y., Xiuting, Z., Cindy, T., Lawley, M.W., Ganal, A.V.W. and David, M.S. 2015. Development of a 63K SNP array for cotton and high-density mapping of intra and inter-specific populations of Gossypium spp. G3: Genes, Genomes, *Genetics. doi*, **10**. 1534/g3. 115. 018416.

Jaiswal, V., Bandyopadhyay, T., Gahlaut, V., Gupta, S., Dhaka, A., Ramchiary, N. and Prasad, M., 2019. Genome-wide association study (GWAS) delineates genomic loci for ten nutritional elements in foxtail millet (Setaria italica L.). *Journal of cereal science*, *85*, pp.48-55.

James, D.W., E. Lim, J. Keller, I. Plooy, E. Ralston and H.K. Dooner 1995 Directed tagging of the *Arabidopsis fatty acid elongation 1* (*FAE1*) gene with the maize transposon activator. *Plant Cell* **7**: 309−319.

Jham, G.N., Moser, B.R., Shah, S.N., Holser, R.A., Dhingra, O.D., Vaughn, S.F., Berhow, M.A, Winkler-Moser, J.K., Isbell, T.A, Holloway, R.K., Walter, E.L., Natalino, R., Anderson, J.C. and Stelly, D.M. 2009.Wild Brazilian Mustard (*Brassica juncea* L.) Seed oil methyl esters as biodiesel fuel. *Journal of American Oil Chemists Society*, **86**:917-926.

Katavic, V., E. Mietkiewska, D.L. Barton, E.M. Giblin, D.W. Reed and D.C. Taylor (2002) Restoring enzyme activity in nonfunctional low erucic acid *Brassica napus fatty acid elongase 1* by a single amino acid substitution. *European Journal of Biochemistry*. **269**: 5625–5631.

Khan, M.A., Tong, F., Wang, W., He, J., Zhao, T. and Gai, J., 2018. Analysis of QTL–allele system conferring drought tolerance at seedling stage in a nested association

mapping population of soybean [Glycine max (L.) Merr.] using a novel GWAS procedure. *Planta*, *248*(4), pp.947-962.

Laila, R., Park, J.I., Robin, A.H.K., Natarajan, S., Vijayakumar, H., Shirasawa, K., Isobe, S., Kim, H.T. and Nou, I.S., 2019. Mapping of a novel clubroot resistance QTL using ddRAD-seq in Chinese cabbage (*Brassica rapa* L.). *BMC plant biology*, *19*(1), p.13.

Lee, G.A., Koh, H.J., Chung, H.K., Dixit, A., Chung, J.W., Ma, K.H., Lee, S.Y., Lee, J.R., Lee, G.S., Gwag, J.G., Kim, T.S. and Park, Y.J. 2009. Development of SNP-based CAPS and dCAPS markers in eight different genes involved in starch biosynthesis in rice. *Molecular Breeding, **24**:93– 101. doi*: 10.1007/s11032-009-9278-7.

Lühs, W.W., A. Voss, F. Seyis and W. Friedt 1999. Molecular genetics of erucic acid content in the genus Brassica. *In*: *Proc. 10th Intern. Rapeseed Congress*, 26–29.

Mailer, R., Product quality. Salisbury, P.A., Potter, T.D., McDonald, G. and Green, A.G. 1999. eds. *Organising Committee of the 10th International Rapeseed Congress*, **71**-74.

Michael, A., Gore., Fang, D.D., Jesse, A., Poland., Zhang, J., Richard, G., Percy., Roy, G., Cantrell., Gregory, T. and Alexander, E.L., 2014. Linkage map construction and quantitative trait locus analysis of agronomic and fiber quality traits in cotton. *Pl. Genome*, **7**(1): 213-215.

Millar, A.A. and L. Kunst (1997). Very-long-chain fatty acid biosynthesis is controlled through the expression and specificity of the condensing enzyme. *The Plant Journal,* **12**: 121–131.

Mir, R.R, and Varshney, R.K. (2012). Future prospects of molecular markers in plants. Molecular Marker in Plants. **22**:23.

Monsalve, R.I., Villalba, M. and Rodríguez, R. 2001. Allergy to mustard seeds: The importance of 2S albumins as food allergens. In *Internet Symposium on Food Allergens* (Vol. **3**, No. 2, pp. 57-69).

Morinaga, T. 1934. Interspecific hybridization in *Brassica*. VI. The cytology of F1 hybrids of *B. juncea* and *B. nigra*. *Cytologia,* **6**: 62-67.

Mortuza, M.G., P.C. Dutta and M.L. Das. 2006. Erucic acid content in some rapeseed/mustard cultivars developed in Bangladesh. *Journal of Sciences of Food Agriculture*, **86**: 135–139.

Nagaharu, U. 1935. Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Japan. J. Biot*.**7**: 389-452

Pandey, S., Kabdal, M. and Tripathi, M.K., 2013. Study of inheritance of erucic acid in Indian mustard (Brassica juncea L. Czern & Coss). *Octa Journal of Biosciences*, ***1***(1).

Patishtan, J., Hartley, T.N., Fonseca de Carvalho, R. and Maathuis, F.J., 2018. Genome-wide association studies to identify rice salt-tolerance markers. *Plant, cell & environment*, 41(5), pp.970-982.

Perry, L.M. and Metzger, J., (1980). *Medicinal plants of East and Southeast Asia: attributed properties and uses*. MIT press.

Peterson, B.K., Weber, J.N., Kay, E.H. and Hoekstra H.E. 2012. Double digest RADseq: An Inexpensive method for de novo SNP discovery and genotyping in model and non- model species. *Public Library of Science One*, **7**(5):e37135.

Pfender, W.F., Saha, M.C., Johnson, E.A. and Slabaugh, M.B. 2011. Mapping with RAD (restriction-site association DNA markers to rapidly identify QTL. For stem rust resistance in Loloiumperenne. *Theoretical and Applied Genetics*, **122**(8):1467-1480.

Poland, J.A. and Rife, T.W. (2012). Genotyping-by-sequencing for plant breeding and genetics, *The Plant Genome*, **5**(3): 92-10.

Prakash, S. and Hinata, K. (1980). Taxonomy, cytogenetics, and origin of crop *Brassica*, a review. *Opera Botanica,* **55**:1–59.

Puritz, J.B., Matz, M.V., Toonen, R.J., Weber, J.N., Bolnick, D.I. and Bird, C.E., 2014. Demystifying the RAD fad. *Molecular ecology*, **23**(24), pp.5937-5942.

Qu, C., Jia, L., Fu, F., Zhao, H., Lu, K., Wei, L., Xu, X., Liang, Y., Li, S., Wang, R. and Li, J., 2017. Genome-wide association mapping and Identification of candidate genes for fatty acid composition in *Brassica napus* L. using SNP markers. *BMC genomics*, **18**(1), p.232.

Ram Manohar, P., Reshmi., Pushpan. and Rohini, S. 2009. Mustard and its uses in ayurveda. *Indian Journal of Traditional Knowledge*, **8**(3):400-404.

Ravelombola, W., Shi, A., Weng, Y., Mou, B., Motes, D., Clark, J., Chen, P., Srivastava, V., Qin, J., Dong, L. and Yang, W., 2018. Association analysis of salt tolerance in cowpea (Vigna unguiculata (L.) Walp) at germination and seedling stages. *Theoretical and applied genetics*, **131**(1), pp.79-91.

Sachidanandam, R., Wiessman, D., Schmidt, S.C. and Kakol, J.M. 2001. A map of genome sequence variation containing 1.42 million single nucleotide polymorphism. *Nature*, **409**: 928-933.

Sauer, F.D., Pigden W.J. and Kramer, J.K.G. 1983. The problems associated with feeding high erucic acid rapeseed oils and some fish oils to experimental animals. In: *High and Low Erucic Acid Rapeseed Oils*, eds. *Academic Press Toronto, Canada*. 253-292.

Shaheen, T., Asif, M. and Zafar, Y. 2009. Single nucleotide polymorphism analysis of MTSHSP gene of Gossypium arboreum and its relationship with other diploid cotton genomes, G. Hirsutum and Arabidopsisthaliana. *Pakistan Jornal of Botany*, **41**(1): 177-183.

Shenolikar, I.S. and Tilak, T.B.G. (1980). Effect of feeding different levels of mustard oil in monkeys. *Annals of Nutrition and Metabolism* **24**: 199-208.

Sinha, S., Jha, J.K., Maiti, M.K., Basu, A., Mukhopadhyay, U.K. and Sen, S.K., 2007. Metabolic engineering of fatty acid biosynthesis in Indian mustard (*Brassica juncea*) improves nutritional quality of seed oil. *Plant Biotechnology Reports*, 1(4), pp.185-197.

Slatkin, M. 2008. Linkage disequilibrium: Understanding the evolutionary past and mapping the medical future. *Nature Review Genetics*, Vol. 9, No. 6, 477-485.

Wang, N., L. Shi, F. Tian, H. Ning, X. WU, Y. Long and J. Meng 2010. Assessment of FAE1 polymorphisms in three *Brassica* species using Eco Tilling and their association with differences in seed erucic acid contents. *BMC Plant Biology*. **10**:137.

Wu, G., Y. Wu, L. Xiao, X. Li. and C. Lu 2008 Zero erucic acid trait of rapeseed (Brassica napus L.) results from a deletion of four base pairs in the fatty acid elongase 1 gene. *Theoretical and applied genetics.* **116**:491-499

Xu, A., Z. Huang, C. Ma, E. Xiao, X. Zhang, J. Tu and t. Fu 2010 FAE1 sequence characteristics and its relationship with erucic acid content in *Brassica juncea*. *Acta. Agronomica Sinica*, **36**:794-800

Yousuf, M., Bhat, T.M. and Kudesia, R. 2013. Comparative genetic diversity studies in mustard (Brassica juncea) varieties using randomly amplified polymorphic DNA (RAPD) analysis. *African J. of Biotech*, **12**(22): 3430-3434.

Zhang, Y. 2010. Allyl isothiocyanate as a cancer chemopreventive phytochemical. *Molecular Nutrition and Food Resources*, **54**(1):127-135.

Zhu, C., Gore, M., Buckler, E. and Yu, J. 2008. Status and prospects of association mapping in plants. *The Plant Genome*, Vol. 1, No. 1, 5-20.

Vita

CURRICULUM VITA

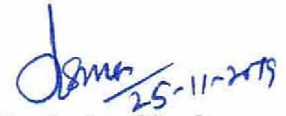
Name of the Student	**:**	Ms. Neha Bharti
Father's Name	**:**	Mr. Sethi Ram
Mother's Name	**:**	Mrs. Santosh Kumari
Nationality	**:**	Indian
Date of Birth	**:**	29-04-1994
Address	**:**	H. No. 106, Ward No. 6 Bishnah, Jammu
Email Id	**:**	Nehabhartiks52@gmail.com

EDUCATIONAL QUALIFICATION

Bachelor's Degree	**:**	B.Sc. (Hons.) Biotechnology
University and Year of Award	**:**	SKUAST-J (2013)
Title of Master's Thesis	**:**	Genomic Profiling for Erucic Acid in European Germplasm of *Brassica Juncea* (L.)
OGPA	**:**	6.69/10
Master's Degree	**:**	M.Sc. Biotechnology
OGPA	**:**	7.03/10

CERTIFICATE-IV

Certified that all the necessary corrections as suggested by the external examiner and the Advisory Committee have been duly incorporated in the thesis entitled **"Genomic profiling for erucic acid in European germplasm of *Brassica Juncea* (L.)"** submitted by **Ms. Neha Bharti**, Regd. No. **J-17-MB-31.**



Dr. Ravinder Singh
Major Advisor

Place: Jammu
Date: 25-11-2019

Coordinator

School of Biotechnology