Development of Genome Resources and Their Utilization in Jatropha Through Comparative Genomics with Castor Bean

By

ARTI SHARMA

A THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

BIOTECHNOLOGY & BIOINFORMATICS

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY,

WAKNAGHAT-173234, H.P. (INDIA)

February 2011

CERTIFICATE

This is to certify that the thesis entitled **"Development of Genome Resources and Their Utilization in Jatropha Through Comparative Genomics with Castor Bean "** submitted by **Mrs. Arti Sharma** to the Jaypee University of Information Technology, Waknaghat in fulfillment of the requirement for the award of the degree of **Doctor of Philosophy in Biotechnology & Bioinformatics** is a record of bona fide research work carried out by her under my guidance and supervision and no part of this work has been submitted for any other degree or diploma.

(Dr. R. S. Chauhan)

Acknowledgement

The path of success is not straight, but riddled with temptations, impediments and pitfalls all along. All these difficulties make us realize the importance of success humbly.

I feel short of words to express my deep sense of gratitude towards the Almighty God with whose grace present study has been completed.

It has extreme pleasure in expressing my sincere gratitude to my worthy guide, Prof. R. S. Chauhan, Head, Department of Biotechnology & Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan, for introducing me to the enthusiastic field of 'Plant Biotechnology', his kind supervision, guidance and encouragement throughout the course of this study. I also appreciate his untiring efforts during the entire tenure of my research work and patience during writing of this thesis. I earnestly believe that without his right suggestions, this work would not have been completed. He is more than a supervisor for me. Thanks Sir, for believing in me and standing by me over the years.

I earnestly wish to express my deep venerable gratitude to Brig (Retd.) Balbir Singh, Registrar, Jaypee University of Information Technology for his constant support, not only for administrative support but also for constant inspiration.

Special thanks are due to Vice Chancellor, Dr. Ravi Parkash for extending his facilities to carry out research and critically reviewing my synopsis.

I am delighted to express my thanks to COO, Dr. Y. Medury, for providing all the necessary facilities, and creating research environment to accomplish this scientific venture of my life.

I wish to convey my sincere thanks to all the faculty member of Department of Biotechnology and Bioinformatics, particularly Dr. Harvinder Singh and Dr. Hemant Sood for their help and guidance at the various stages of this study.

I am extremely thankful to non-teaching staff particularly Somlata mam and Mamta mam of the department for their selfless assistance. Their boundless love shall be remembered forever.

I am grateful to my father Sh. D.P. Aghnihotri who has been a moving spirit behind all my achievements. I bow my head to my mother Smt. Sarita Aghnihotri for her generous support invaluable blessings, selfless but untold sacrifices for my better and bright future. Without the encouragement and moral support of my in-laws, I could never have been what I am today. I am also thankful to my brother Vikas for his support and encouragement.

It is high time to acknowledge all my fellow colleagues especially Nidhi, Saurabh, Chintu, Payal di, Krishna, Anu, Chinku, Jatin, Tamanna mam, Anshu, Mani, Varun, Jeet , Sneha, Asha, Arun, Mamta, Navneet and Priya for their problem solving attitude and healthy discussions during the course of my research work.

I highly appreciate for unsolicited cooperation and cheery assistance of my husband Dr. Jyoti Parkash. His unconditional support, strength and love helped me to make my dream come true.

I am thankful to the Department of Biotechnology, Ministry of Science and Technology, Govt. of India, New Delhi for awarding me the research fellowship.

(Arti Sharma)

Table of Contents

List of Tables

List of figures

INTRODUCTION

Jatropha (*Jatropha curcas)*

Castor bean (*Ricinus communis*)

Table 1.1: Taxonomical and biochemical similarities between Jatropha and Castor bean

Category	**Jatropha**	**Castor bean**
1. Family	Euphorbiaceae	Euphorbiaceae
2. White sap (Latex)	Present	Present
3. Toxic protein	Curcin	Ricin
4. Oil storing organ	Seed	Seed
5. Oil value	Biodiesel	Biodiesel
6. Seed Oil Content	45-50%	50-55%
7. Saponification Value of Oil	195.0	185.0
8. Iodine Value	101.7	85.0
9. Oil Viscosity	40.5	31.5
10. Fatty acid composition in seeds	Oleic and Linoleic acids (79%)	Ricinoleic acid (87%)
11. Oil Density	0.88 g/ml	0.95 g/ml

A major challenge mankind is facing in this century is the gradual exhaustion of the earth's fossil energy resources. The combustion of those fossil energy sources used as transportation fuel is one of the key factors responsible for global warming and environmental pollution due to large-scale carbon dioxide emissions. Thus, alternative energy sources based on sustainable and ecologically friendly processes are urgently needed. At present, about 90% of the biofuel market is captured by bioethanol and biodiesel, which are already applied as gasoline or diesel substitutes on a large scale and are referred to as first generation biofuels (Antoni *et al.*, 2007). Bio-fuel is a renewable fuel which can be an alternative to or an addition to fossil-derived fuel with multitude of environmental benefits. Off late, various oil seed plants suited to wide agro climatic conditions are being explored as the sources of future fuels. The need to reduce greenhouse gas emissions and provide fuel security has increased the demand for oil-rich plants as raw materials for biodiesel production. Plant seed storage oils, in the form of triacylglycerols (TAGs), are excellent sources for the generation of biodiesel due to their high chemical similarity to fossil oils (Dyer *et al.*, 2008; Durrett *et al.*, 2008). Biodiesel is produced by the transesterification of plant TAGs with methanol in the presence of acid or alkali to produce fatty acid methyl esters (FAMEs) (Fairless, 2007). Although vegetable oils have long been used for food, the ideal crop source for biodiesel products should consider other ecological, environmental and ethical concerns. Ideally, the entire process, from cultivation to fuel burning in engines, should favour carbon sequestration, reduce water needs and promote energy efficiency. Moreover, the impact of oil crops for biodiesel production on the prices of food commodities is a matter of concern. Ideally, such crops should be non-edible and grown on non-agricultural lands so that they do not compete for soil with food crops and do not affect the prices of food commodities (Costa *et al.*, 2010).

Much of the current interest in *J. curcas* oil is for its potential as a feedstock for biodiesel production. Jatropha is a perennial shrub to small evergreen tree of upto 6 m height, adapted to all kinds of soils and does not demand any special nutritive regime (Patil and Singh, 1991). The genus Jatropha is morphologically diverse encompassing more than 200 species, which are distributed chiefly in the dry tropical regions of America, and has been later introduced into Africa and Asia and is now cultivated worldwide (Sujatha and Prabakaran, 1997). In India, it is believed to be introduced by the Portuguese settlers during 16th century. Recently, it has also been introduced in the northern and southern states of India (Ginwal *et al.,* 2005). The plant is widely distributed and fits easily into agricultural system in the form of hedges, windbreak, and erosion barrier or as a source of firewood (Srivastava, 1999). It has less gestation period, rapid growth, not grazed by animals, which strengthen its promotion in wastelands. Even though the seeds of Jatropha are highly toxic due to the protein 'curcin' and phorbol esters (Gubitz *et al*., 1998), almost all parts of this plant have been utilized, either in insecticides, green manure, soap making, medicine, just to name a few (Lin *et al*., 2003). Jatropha seeds contain about 40% of oil enriched with both saturated [palmitic acid (16:0, 14.1%) and stearic acid (18:0, 6.7%)] and unsaturated [oleic acid (18:1Δ9, 47.0%) and linoleic acid (18:2Δ9, 12, 31.6%)] fatty acids (Augustus *et al*., 2002; Bringi, 1987; Makkar, 1997). Current estimates suggest that there are now 2.5 million hectares of *J. curcas* planted in India and China alone, with plans for an additional 9.3 million hectares by 2010 (Fairless, 2007). Interest in the cultivation of *J. curcas* is coming from both the public and private sectors, and a number of public companies are now involved in *J. curcas* cultivation including D1 Oils plc (www.d1plc.com), Viridas plc (www.viridasplc.com), Reliance Life Sciences (www.rellife.com), and Energem Resources Inc (www.energem.com).

The development of *J. curcas* genotypes (varieties) with high seed yield, high oil content and adapted to varied conditions is a priority (Divakara *et al*., 2010). In recent years, *J. curcas* germplasm has been collected and analyzed in Brazil, India, Indonesia, and China (Ou *et al*., 2009; Tatikonda *et al*., 2009). *J. curcas* is a diploid species with a 2n chromosome number of 22 (Dehgan, 1984). A recent study has estimated the genome size (1C) to be 416 Mbp (Carvalho *et al*., 2008). This is relatively small for a plant genome (Zonneveld *et al*., 2005). Genome of *J. curcas* has recently been sequenced in December 2010 by Sato *et al*. (2011), which was not available earlier for the perusal of this study. The genetic map of *J. curcas* is not well-developed and very limited information is available with respect to molecular markers, which are pre-requisites for genetic improvement of Jatropha for high oil content or other desirable traits. The traditional methods of developing molecular markers are time consuming and labor-intensive. However, comparative genomics has been successfully utilized in the development of molecular markers across different plant species, which are taxonomically related such as Cruciferae (Axelsson *et al*., 2001), Solanaceae (Causse *et al*., 2007), and grass genomes (Bennetzen and Ramakrishna, 2002). Previous studies based on Random Amplified Polymorphic DNA (RAPD), Simple Sequence Repeat (SSR) and Amplified Fragment Length Polymorphism (AFLP) analysis have indicated that the genetic base of *J. curcas* is very narrow. Basha and Sujatha (2009) demonstrated polymorphism of 61.8 and 35.5% with RAPD and Interstitial Simple Sequence Repeat (ISSR) primers, respectively. They identified 12 microsatellite primers differentiating the toxic and non-toxic Mexican accessions. Pamidimarri *et al.* (2009) identified RAPD, AFLP and one SSR marker to differentiate the toxic and non toxic varieties of *J. curcas*. Wen *et al.* (2010) identified 241 novel EST-SSRs and genomic SSR markers from cassava which showed polymorphism among *J. curcas* accessions.

Castor bean (*Ricinus communis*) is a plant species, which is taxonomically related to *J. curcas* such that both belong to family Euphorbiaceae, both produce white sap, the seeds of both contain a similar toxic protein, and more importantly the seed oil from both is used in the production of biodiesel and other industrial products. The high level of synteny at genomes level can be expected between these two plant species, which can be exploited to develop anchor markers by utilizing genome resources of castor bean. In addition, the castor genome project, which is currently available as four-times draft (http:// castorbean.jcvi.org), could serve as a useful resource for the identification of genes by homology to assist in creating genetic and physical maps. The availability of castor bean genome sequence in 2007 at the TIGR web site (http://www.tigr.org/) enabled us to take up the current study, which would be further strengthened with the recent publication of genome sequences of Jatropha and castor bean (Sato *et al.*, 2011; Chan *et al.*, 2010).

The fatty acid composition of seed oil varies considerably between and within species. The total seed oil contents of Arabidopsis, Jatropha, castor bean Brassica and soybean are 30-37%, 30-50%, 40-45%, 30-40% and 15-20%, respectively (Table 1.2) (Li *et al.*, 2006; Velasco and Becker, 1998; Ramos *et al.*, 1984, Sangwan *et al.*, 1986). Plant oils are mostly composed of five common fatty acids, namely palmitate (16:0), stearate (18:0), oleate (18:1), linoleate (18:2) and linolenate (18:3), although, depending on the particular species, longer or shorter fatty acids may also be major constituents. The variation of fatty acids occurs both in chain length and degrees of desaturation. Consequently, the fuel properties of biodiesel derived from a mixture of fatty acids are dependent on the composition of fatty acids in the seed oil. Altering the fatty acid profile can, therefore, improve fuel properties of biodiesel such as cold-temperature flow characteristics, oxidative stability and NOx emissions. Improving cold-temperature flow

characteristics requires a fuel with low saturated fatty acid levels, whereas increasing oxidative stability and reducing NOx emissions requires decreasing the amounts of unsaturated and polyunsaturated fatty acids. Additionally, the ignition quality (as measured by cetane number) is also adversely affected by increased unsaturation (Klopfenstein, 1985; Serdari *et al.*, 1999; Stournas *et al.*, 1995). However, a compromise can be reached by considering a fuel high in the mono-unsaturated fatty acids, such as oleate or palmitoleate (16:1Δ9), and low in both saturated and polyunsaturated fatty acids. The presence of a single double bond greatly enhances the cold-temperature flow properties of methyl oleate compared to methyl stereate (Serdari *et al.*, 1999; Stournas *et al.*, 1995). Various studies suggest that biodiesel with high levels of methyl oleate will have good characteristics with regard to ignition quality, NOx emissions and fuel stability. For example, while unsaturation tends to reduce the cetane number of biodiesel, that of methyl oleate is higher than the minimal biodiesel standard (Knothe, 2005). Additionally, it has been estimated that biodiesel fuels with an average of 1.5 double bonds per molecule will produce an equivalent amount of NOx to conventional diesel (McCormick *et al.*, 2001), thus a fuel high in oleates (one double bond per molecule) should not result in higher NOx emissions. Finally, given that polyunsaturated fatty acids have a disproportionably large effect on the auto-oxidation of biodiesel (Knothe and Dunn, 2003), reducing the polyunsaturated fatty acid content will improve the stability of the fuel (Durrett *et al.*, 2008).

Fatty acid biosynthetic pathway is highly conserved in plants; but there are significant variations in fatty acid contents and composition in plants. In spite of the availability of full genome sequences for oil seed rich plant species and the cloning and characterization of fatty acids biosynthesis genes, it is not known as yet, what determines differences in the contents and composition of fatty acids and subsequently the total oil yield in seeds. Comparative

computational analysis of fatty acid biosynthesis pathway genes across different oil seed plant species may be helpful in understanding the possible causes of differences in content and composition of oil content.

Candidate gene approach is becoming a widespread method for characterizing the Quantitative Trait Loci (QTLs) as well as Mendelian traits in both the animal and plant systems. The selection and validation of candidate genes relies on two non-exclusive approaches. A correlation between the trait under study and allelic polymorphism of the candidate gene, regardless of the genetic background, is a strong argument in favour of the candidate gene. Candidate genes provide a valuable resource for mapping in various plants and further germplasm analysis using association genetics. Association mapping uses diverse material to associate genetic markers with a phenotype of interest, taking advantage of lower levels of linkage disequilibrium than are present in linkage populations. Association mapping has been used to identify genes of interest in many plant species with varying degrees of success (Yang *et al*., 2010). Compared to the conventional linkage mapping, association mapping, using the nonrandom associations of loci in haplotypes, is a powerful high-resolution mapping tool for complex quantitative traits. Association mapping does not require special mapping populations but rely on the extensive history of mutation and recombination to dissect a trait. Association mapping evaluates whether certain alleles within a population are found with specific phenotypes more frequently than expected. The lack of genetic map and synteny of candidate gene markers associated with high oil content warrants that association mapping can be pursued in *J. curcas*.

Table 1.2: Fatty acid composition of five plant species

Fatty acid composition (%)	Arabidopsis	Jatropha	Castor bean	Brassica	Soybean
Palmitic acid	8.7	4.2	2.0	1.5	7-11
Stearic acid	3.6	6.9	1.0	0.4	2-6
Oleic acid	15.0	**43.1**	7.0	**22.0**	**22-34**
Linolenic acid	**29.0**	0.3	----	6.8	5-11
Linoleic acid	19.2	**34.3**	5.0	14.2	**43-56**
Ricinolic acid	-----	------	86-90	----	-----
Others	24.5	11.2	----	**47** (Erucic)	-----
Total oil content	30-37	30-50	45-50	33-40	15-20

The large-scale cultivation of selected genotypes of *J. curcas* across India has resulted in venerability of this plant species to biotic stresses such as fungal and viral pathogens. Recently, Jatropha curcas mosaic disease (JcMD) has been found to reduce fruit yield and quality of *J. curcas* plants in the field and similar pathological stresses are expected to emerge in future. Management of plant diseases through host resistance is considered as one of the best options available for crop protection. The DNA sequence analysis of cloned plant disease resistance genes effective against viral, bacterial or fungal pathogens has revealed that the majority of them contain similar sequences and structural motifs across wide range of plant taxa. The structural homology has made it possible to identify resistance gene analogues in different plant species (Van der Linden *et al*., 2004). *J. integerrima* has been found to withstand disease pressure in the field conditions compared to *J. curcas* (KT Parthivan unpublished). It is, therefore, prudent to undertake identification and characterization of resistance gene analogues from *J. integerrima* with the possibility of their utilization in molecular breeding of *J. curcas* for disease resistance.

PCR approaches using degenerate primers based on the conserved nucleotide binding site (NBS) domains of cloned R-genes can provide an attractive strategy to amplify multiple resistance-gene analogs sequences that can be developed either into molecular markers for use in marker-assisted selection (MAS) or even lead to molecular cloning of new disease resistance genes.

Keeping in view the economic importance of Jatropha as a potential source of biofuel, lack of sufficient genome resources and scant information about the fatty acid biosynthesis pathway genes, the current study was carried out with the following objectives:

1. Whole genome analysis of simple sequence repeats (SSRs) in castor bean genome and their utilization in *Jatropha curcas* and other Jatropha species
2. Comparative analysis of fatty acid biosynthesis genes from major oil producing plant species
3. Identification and annotation of castor bean genome sequence for candidate genes involved in fatty acid biosynthesis
4. Association analysis of candidate gene markers (SNPs and SSRs) with high oil content in *J. curcas*
5. Whole genome analysis of castor bean NBS-LRR genes and their utilization in cloning resistance gene analogues (RGAs) in *J. integerrima*

Chapter-II

REVIEW OF LITERATURE

The literature pertinent to the present study has been reviewed under the following heads:

2.1) Identification and utilization of simple sequence repeats (SSRs)

2.2) Fatty acids and TAG biosynthesis in plants

2.3) Comparative genomics for identification of candidate gene markers

2.4) Candidate gene markers in fatty acid biosynthesis genes

2.5) Disease resistance genes in plants

2.1) Identification and utilization of simple sequence repeats (SSRs)

Microsatellites or SSRs are tandemly arranged repeats of short DNA motifs that frequently exhibit variation in the number of repeats at a locus. SSRs have been used as genetic markers because of their presence in the genomes of all living organisms, high level of allelic variation; co-dominant inheritance and potential for cross species amplification. SSRs occur as frequently as 1 in approximately every 6 kb in plant genomes (Varshney *et al.,* 2005). SSR enrichment methods improve throughput and reduce development costs (Powell *et al.,* 1996; Kumpatla and Mukhopadhyay, 2005). More recently, the availability of enormous sequence data for a large number of plant genomes has accelerated research aimed at understanding the origin and functions of microsatellites and searching for new applications. In addition to molecular markers, the SSRs have many important functions in terms of development, gene regulation, and evolution ((Li *et al*., 2004; Robinson *et al.,* 2004; Varshney *et al.,* 2005; Zhang *et al.,* 2008). The

locations of SSRs appear to determine the types of functional role SSRs might play, and changes in SSRs in different genetic locations can lead to changes in the phenotypes of an organism (Gupta and Varshney, 2000; Poncet *et al.,* 2006). SSRs in coding regions can determine whether or not a gene gets activated or whether the protein product is truncated (Varshney *et al.,* 2005). The development of the nervous system in *Drosophila* appears to be associated with length variation of trinucleotide repeats in genes involved in developmental control (Kashi and King, 2006). SSRs in other genic regions can have large effects on organisms as well. For example, SSRs in 5'-untranslated regions (UTRs) have an effect on gene transcription and/or regulation (Li *et al.,* 2004). Variations in SSRs in 5'UTRs and 3'UTRs have been reported to effect gene expression (Decroocq *et al.,* 2003). The SSRs found in the 3'UTRs may cause transcription slippage and produce expanded mRNA (Li *et al.,* 2004). Large numbers of SSRs were detected and documented in transcribed regions of genomes, including coding DNA sequences (Largercrantz *et al*., 1993; Morgante *et al.,* 2002). It has been found that 12% of identified SSRs in Japanese pufferfish (Edwards *et al*., 1998), 10% in primate (Jurka and Pethiyagoda, 1995), 15% in rabbit (van Lith and van Zutphen, 1996), and 9.1% and 10.6%, respectively, in pig and chicken are located in the protein coding genes or open reading frames (ORFs) (Moran, 1993). In cereals (maize, wheat, barley, sorghum, and rice) 1.5%–7.5% of ESTs consist of SSRs (Kantety *et al.,* 2002; Thiel *et al.,* 2003). These ESTs have a range of functions such as metabolic enzymes, structural and storage proteins, disease signaling, and transcription factors, suggesting some role(s) of SSRs in plant metabolism and gene evolution. In protein-coding regions of all known proteins, 14% proved to contain repeated sequences, with a three times higher abundance of repeats in eukaryotes as in prokaryotes (Marcotte *et al*., 1999). SSRs are considered to be evolutionarily neutral DNA markers (Morgante *et al.,* 2002). Length polymorphism arises from

variations in the number of repeat units, probably due to DNA polymerase slippage during the replication of SSRs (Kantety *et al.,* 2002). SSRs have been used for population genetics and typing studies due to several advantages such as that they are highly polymorphic, multiallelic, highly reproducible, and detectable by PCR (Schlotterer, 2000; Rota *et al.,* 2005).

It has been found in other organisms that SSRs with longer repeat motifs are more informative for detection of polymorphisms (Moretzsohn *et al.,* 2005; Cardle *et al.,* 2000; Rota *et al.,* 2005; Morgante and Olivieri, 1993; Mortimer *et al.,* 2005; Riva *et al.,* 2000). For example Sharma and Chauhan (2008) identified a SSR with longer repeat motif $(TTC)_{31}$ in the iron transporter genes of maize which showed higher levels of polymorphisms among maize inbreds compared to other repeat motifs with lesser repeat length. Dubey *et al.* (2006) showed that intron 1 of the Epidermal Growth Factor Receptor (EGFR) gene in human contained a polymorphic simple sequence repeat (SSR) of 14 to 21 CA dinucleotides, the length of which correlated inversely with the level of EGFR transcription.

SSRs can be identified empirically by screening DNA libraries for repeat motifs via hybridization and sequencing candidate clones (Li, 2000; Taramino and Tingey, 1996), applying SSR primers from related species (Westman and Kresovich, 1998) or mining sequence databases (Chin *et al.,* 1996; Senior and Heun, 1993; Wang *et al.,* 1994). The cross species amplification of microsatellites was reported in many species (Lopes and Maciel, 2006; Rob and Durka, 2006) that can save time and resources as these markers can be utilized in marker-assisted breeding programs for the improvement of cross amplified taxa. Faba bean EST- SSRs were transferred to *Pisum sativum* (Gong *et al*., 2010). Li *et al*. (2010) used 14 EST-SSR primer pairs from *Prunus mume* to test their transferability on peach and plum. A high level of marker polymorphism was observed in the plum species (65%) and low in the peach (46%), and the clustering analysis of

the three species indicated that these SSR markers were useful in the evaluation of genetic relationships and diversity between and within the *Prunus* species.

The narrow genetic base in crops is the bottleneck of genetic improvement (Callow *et al.*, 1998; Bennetzen *et al.*, 1998). Utilization of wild species, local varieties and mutants would be an effective way to change this situation (Swanson, 1996; Marques *et al.*, 2002). Therefore, it is important for plant breeders to evaluate genetic diversity comprehensively not only in various cultivated plant varieties but also in related wild species. Previous studies based on RAPD analysis have indicated that the genetic base of *J. curcas* is very narrow or the RAPD markers were not sufficient to detect range of variation (Basha and Sujatha, 2007). Further studies based on RAPD, SSR and AFLP analysis have also indicated that the genetic base of *J. curcas* is narrow (Basha and Sujatha, 2009; Sun *et al.*, 2008; Basha *et al.*, 2009). Basha *et al.* (2009) demonstrated polymorphism of 61.8 and 35.5% with RAPD and ISSR primers, respectively. They identified 12 microsatellite primers differentiating the toxic and non-toxic Mexican accessions of *J. curcas*. Pamidimarri *et al.* (2009) identified RAPD, AFLP and SSR markers to differentiate toxic and non toxic varieties of *J. curcas*. Tatikonda *et al.* (2009) tested 7 AFLP primer combinations on 48 accessions of *J. curcas* that generated a total of 770 fragments with an average of 110 fragments per primer combination. They identified 680 (88%) polymorphic fragments in the germplasm analyzed, of which 59 (8.7%) fragments were unique (accession specific) and 108 (15.9%) fragments were rare (present in less than 10% accessions). Grativol *et al.* (2010) used ISSR markers to assess the genetic variability of *Jatropha curcas* accessions from eight states in Brazil. Seven ISSR primers amplified a total of 21,253 bands, of which 19,472 bands (91%) showed polymorphism. Wen *et al.* (2010) identified 187 expressed sequence tag (EST)-SSR and 54 genomic (G)-SSR markers from cassava which showed polymorphism

among the *J. curcas* accessions. The EST-SSR markers comprised 26.20% dinucleotide repeats, 57.75% trinucleotide repeats, 7.49% tetranucleotide repeats, and 8.56% pentanucleotide repeats; whereas the majority of G-SSR markers were dinucleotide repeats (62.96%).

2.2) Fatty acids and TAG biosynthesis in plants

The biosynthesis of fatty acids in plants begins with the formation of acetyl Co-A from pyruvate. The acetyl CoA produced in plastids is activated to malonyl CoA; the malonyl group is subsequently transferred to acyl carrier protein (ACP) giving rise to malonyl ACP, the primary substrate of the fatty acid synthase complex. The formation of malonyl CoA is the committed step in fatty acid synthesis and is catalyzed by the highly regulated plastidic acetyl CoA carboxylase complex (Nikolau *et al.*, 2003). *De novo* fatty acid synthesis in the plastids occurs through a repeated series of condensation, reduction and dehydration reactions that add two carbon units derived from malonyl ACP to the elongating fatty acid chain. A series of condensation reactions proceed with acetyl-CoA and malonyl-ACP, then acyl-ACP acceptors. Three separate condensing enzymes, or 3-ketoacyl-ACP synthases (KAS I– III) are necessary for the production of an 18-carbon fatty acid. Three additional condensation reactions are required; each condensation step to obtain a saturated fatty acid that is two carbons longer than at the start of the cycle. These reactions are catalyzed by 3-ketoacyl-ACP reductase (KAR), 3-hydroxyacyl-ACP dehydratase (HD), and enoyl-ACP reductase (ENR). The first desaturation step also occurs in the plastid; while the acyl chain is still conjugated to ACP a Δ 9-desaturase convert stearoyl ACP to oleoyl ACP. Termination of fatty acid elongation is catalyzed by acyl ACP thioesterases, which are mainly of two types in plants. The FatA class removes oleate from ACP, whereas FatB thioesterases are involved in saturated and unsaturated acyl ACPs, and, in some species, with shorter-chain-length acyl ACPs (Mayer and Shanklin, 2007; Pollard *et al.*, 1991; Salas and

Ohlrogge, 2002). After release from ACP, the free fatty acids are exported from the plastid and converted to acyl CoAs. Nascent fatty acids can be incorporated into TAGs (triacyglycerol acyltransferase) in developing seeds (Mayer and Shanklin, 2007). Oleic acid can be further desaturated to oleate acids by FAD2 (Okuley *et al.,* 1994) and FAD6 (Hitz *et al.,* 1994) in the cytosol and the plastid, respectively. Cytosolic and plastid ω-3 desaturations that result in the production of linolenic acids are catalyzed by FAD3 (Reed *et al.,* 2000) and FAD7 (Iba *et al.,* 1993), respectively. Fatty acids can be incorporated into TAGs in developing seeds in a number of ways. For example, a series of reactions known as the Kennedy pathway results in the esterification of two acyl chains from acyl CoA to glycerol-3-phosphate to form phosphatidic acid (PA) and, following phosphate removal, diacylglycerol (DAG). A diacyglycerol acyltransferase (DGAT), using acyl CoA as an acyl donor, convert DAG to TAG. Two classes of DGAT enzymes have been isolated (Cases *et al.,* 1998; Lardizabal *et al.,* 2001), and orthologs have been identified in numerous plant species. DAG and phosphatidylcholine (PC) are interchangeable via the action of cholinephosphotransferase, suggesting a route for the flux of fatty acids into and out of PC. Acyl chains from PC can be incorporated into TAG, either via conversion back to DAG or by the action of a phospholipid diacylglycerol acyltransferase (PDAT) that uses PC as an acyl donor to converts DAG to TAG. There are two predominant seed oil storage proteins in plants: caleosin and oleosin. TAG assembled in these storage proteins forms oil bodies in seeds (Fig. 2.1).

2.3) Comparative genomics for identification of candidate gene markers

Comparative genomics can be used to gain knowledge of gene organization, and is particularly helpful in examining genome evolution (Keller and Geuillet, 2000; Ellwood *et al.,* 2008). Closely related species have extensive regions of gene co-linearity, a phenomenon also

known as synteny (Passarge *et al.,* 1999; Zang *et al.,* 2009), but as the evolutionary distance between two species increases, the segments of co-linearity get shorter.

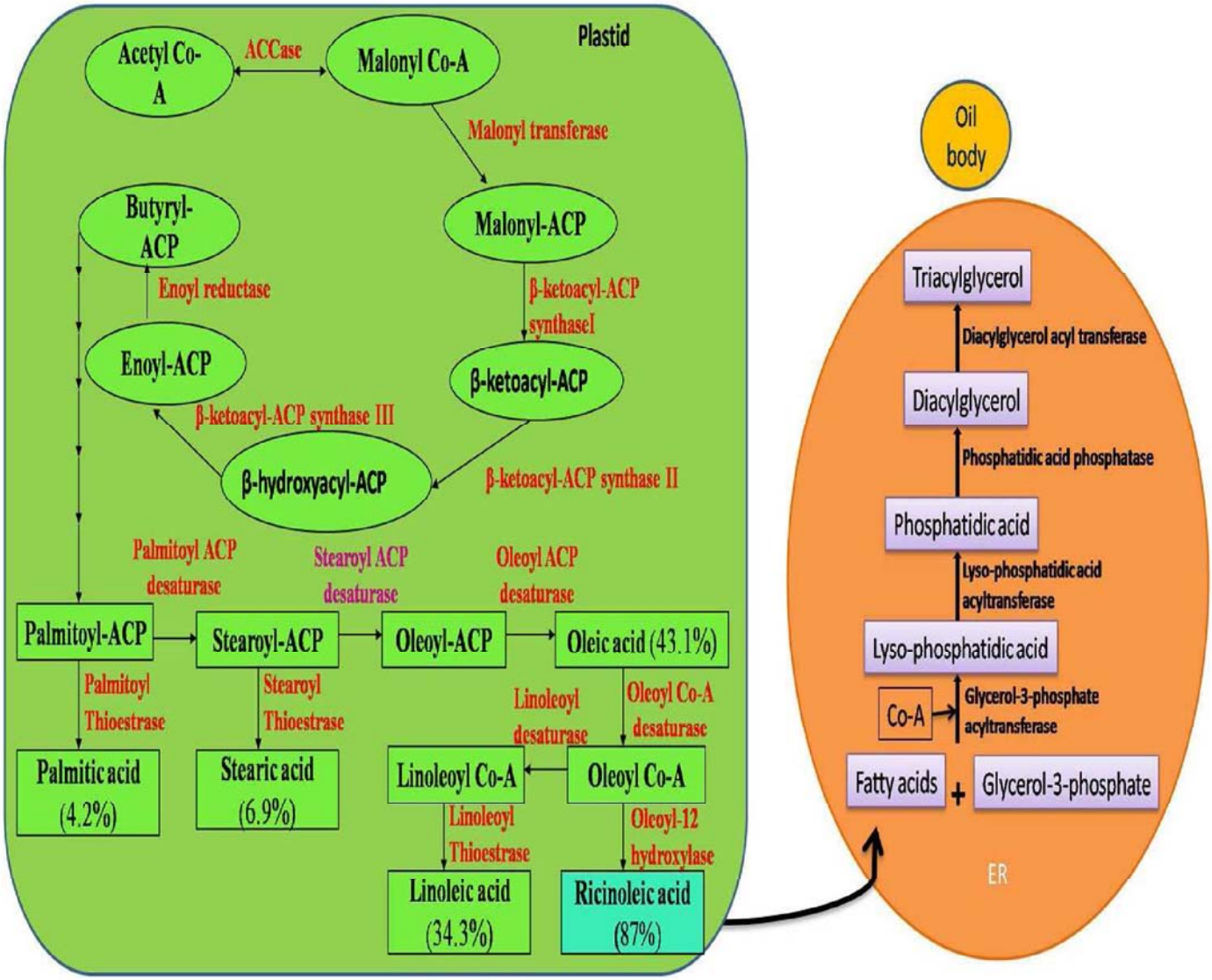

Fig. 2.1: Fatty acid and TAG biosynthesis pathway in Jatropha and castor; ricinoleic acid only present in castor bean; concentrations of each fatty acid are given for Jatropha

The recent availability of complete genomes of several model systems has sparked renewed interest in the study of co-linearity because of the phenomenon's potential for transferring useful information from well-studied small genomes to larger ones (Rubin *et al.,* 2000). Conservation in structure and function of genetic loci among species have been documented, even though genome sizes in plant species exhibit large variation, e.g., 125 Mb for *A. thaliana* to 125 Gb for *Fritillaria assyriaca* (Bennett and Smith, 1976). Extensive gene conservation, both in structure and function, has been reported in grass genomes (Bennetzen and Ramakrishna, 2002), Crucifers (Axelsson, 2001), and solanaceous plants (Causse, 2007). Comparative genomics is widely used in cereal crops to study various traits (yield, disease resistance etc.). Comparative genomics research has several goals: (1) to compare the organization of related genomes and infer the basic processes of genome evolution, (2) to transfer information from model species to related organisms, and (3) to integrate information on gene location and expression across species. Crop improvement programs can use comparative genetics to transfer information about genes from model species to their species of interest, to help identify the genes controlling traits of interest, and to assess within-species allelic diversity so that the best alleles can be identified and assembled in superior varieties (Sorrells *et al.,* 2003; Heesacker *et al.,* 2009). Shin *et al*. (2008) used comparative genomics to study lipoxygenase gene family in *Medicago truncatula* and soybean. They analyzed two *Lx* regions in *Medicago truncatula* showing synteny with soybean. Zang *et al.* (2009) studied glucosinolate biosynthesis in *Brassica rapa* through comparative genomics with *Arabidopsis thaliana* and identified total 56 putative biosynthetic and regulator genes. They found high level of co-linearity in the glucosinolate biosynthesis pathway between Arabidopsis and *B. rapa*. Glucosinolate genes in *B.*

rapa share 72–94% nucleotide sequence identity with the Arabidopsis orthologs and exist in different copy numbers.

Candidate gene approach is becoming a widespread method for characterizing QTLs as well as Mendelian traits in both the animal and plant systems. The selection and validation of candidate genes relies on two non-exclusive approaches (de Vienne *et al.,* 1999). The first one involves the 'functional' candidate gene approach which is based on a priori choice of gene(s) which may be functionally related to the trait. A correlation between the trait under study and allelic polymorphism of the candidate gene, regardless of the genetic background, is a strong argument in favour of the candidate gene. The second approach, the 'positional' candidate gene approach, relies on QTL mapping and on examination of known functional genes or mutations which map in the same region, the effect of which may be related to the trait. Thus, the first step for validation of candidate genes requires the search for correlation between trait value and allelic polymorphism. Candidate gene approach has been successfully used in association mapping of silage corn digestibility, kernel composition and starch production in maize (Wilson *et al.,* 2004). Candidate-gene association mapping requires the identification of SNPs or Insertions or Deletions (INDELs) between lines and within specific genes. Significant associations with SNPs and INDELs located in the vicinity of 12 candidate genes from seven loci known to be involved in the genetic determination of growth habit and inflorescence type were identified in 102 barley accessions through re-sequencing of candidate genes (Cuesta-Marcos *et al*., 2010). Comadran *et al.* (2010) carried out a genome-wide analysis of polymorphism (4,596 SNP loci across 190 elite cultivated accessions) chosen to represent the available genetic variation in current elite North West European and North American barley germplasm. They observed population sub-structure, patterns of diversity and linkage

disequilibrium varied considerably across seven barley chromosomes. The study found that candidate genes involved in biosynthesis and transport are co-localized in 23 of 25 identified QTL regions. Co-localizations of the QTL regions with candidate genes involved in metal transport support the putative functions of these genes. Deshmukh *et al.* (2010) used a combination of QTL mapping and transcriptome profiling to identify candidate genes for grain number. QTL mapping using 3 years phenotyping data on a set of recombinant inbred lines derived from a cross between *Pusa 1266* (high grain number) and *Pusa Basmati 1* (low grain number) identified one consistent QTL *qGN4-1* on the long arm of chromosome 4 with major effect on grain number. This QTL co-localized with major QTLs for primary and secondary branches per panicle and number of panicles per plant.

Association mapping does not require special mapping populations but rely on the extensive history of mutation and recombination to dissect a trait. Association mapping, based on linkage disequilibrium (LD), offers an alternative method for mapping QTLs. Association mapping evaluates whether certain alleles within a population are found with specific phenotypes more frequently than expected (Fig. 2.2). Association mapping is to detect the degree to which a gene and trait or genotype and phenotype vary together in a sampled population on the basis of linkage disequilibrium (Zondervan and Cardon, 2004). When a molecular marker is associated with a phenotypic trait, it should associate with others that highly correlate with this trait in theory.

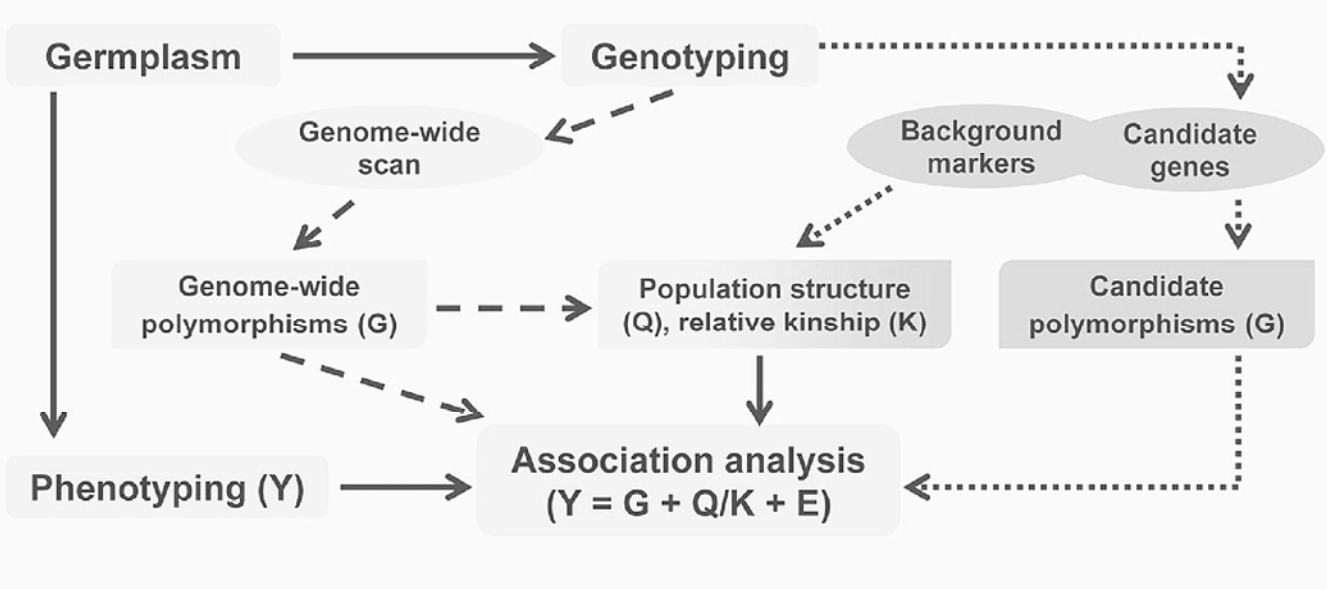

Fig. 2.2: Illustration of steps necessary to associate a particular genotype with phenotype. Association study employs techniques from molecular biology, field sampling/ breeding, bioinformatics and statistics (Zhu *et al.,* 2008); Y a vector of phenotypes, X a vector of single locus genotypes, Q a matrix of the $K - 1$ subpopulation ancestry estimates for each individual, Z is an identity matrix.

2.4) Candidate gene markers in fatty acid biosynthesis genes

Fatty acid biosynthetic pathway is highly conserved in plants; but there are significant variations in fatty acid contents and composition in plants (Table 2.1). Tanhuanpaa *et al.* (1998) developed an allele-specific PCR marker for oleic acid by comparing the wild-type and high-oleic allele of the Fatty Acid Desaturase 2 (FAD-2) gene locus in spring turnip rape (*B. rapa* ssp. *oleifera*). Knutzon *et al.* (1992) identified an indel in the exon of Oleoyl ACP carrier associated with oil content in safflower. SNPs were also identified in the exon of Oleoyl desaturase gene in Peanut, sunflower, soybean and Brassica (Lopez *et al.,* 2000; Lacombe and Bervillé, 2001; Hu *et al.,* 2006; Ha *et al*., 2010). Hu *et al*. (2006) sequenced FAD 2 genomic fragment from the mutant line DMS100 and wild-type line Quantum of *B. napus* and identified a single nucleotide

mutation, C to T, in the gene. This particular mutation created a stop codon (TAG) leading to premature termination of the peptide chain during translation. Bilyeu *et al.* (2005) identified an SNP in the intron of Linoleoyl desaturase in soybean. Aghoram *et al.* (2006) identified a SNP in one of the homologs of KAS II gene in soybean cultivar from C1727 which converted a tryptophan codon into a premature stop codon, a mutation that would be predicted to render the encoded enzyme nonfunctional. Pérez-Vich *et al.* (2006) identified SSRs and INDELs in Stearoyl–ACP desaturase gene associated with high stearic acid in sunflower. The *Brassica napus* mutant line DMS100 carrying a G-to-A base substitution at the 5' splice site of intron 6 in FAD 3 had reduced C18:3 content in oil seeds (Hu *et al.,* 2007). An SNP was also identified in the exon of Steroyl desaturase gene in Arabidopsis and soybean (Kachroo *et al.,* 2006; Zhang *et al.,* 2008). Zheng *et al.* (2008) characterized a maize high-oil QTL (qHO6) and found that this associated with Diacylglycerol Acyltranferase (DGAT1-2). In addition, the sequence of this particular DGAT isoform contained a single amino acid insertion (phenylalanine 469) that was present in ancestral varieties but was absent from more recent commercial lines. Ectopic expression of the F469 allele of DGAT in transgenic maize resulted in a significant increase in the levels of total oil and also oleic acid, compared with similar expression of the conventional allele of DGAT1-2. Burgal *et al.* (2008) demonstrated that co-expressing the castor bean DGAT2 gene with the castor FA 12 hydroxylase resulted in almost double the levels of hydroxylated fatty acids in neutral lipids (up to 30% of total, compared with 17% in the absence of DGAT2). Wu *et al.* (2008) sequenced the *FAE1* gene from eight high and zero erucic acid rapeseed cultivars (*Brassica napus* L.) and identified four base pair deletion between T1366 and G1369 in the *FAE1* gene leading to a frameshift mutation and a premature stop of the translation after the 466th amino acid residue. Singh *et al.* (2009) constructed genetic map using AFLP, RFLP and

SSR markers for oil palm. They detected quantitative trait loci (QTLs) controlling oil quality (measured in terms of iodine value and fatty acid composition) and identified significant QTLs associated with iodine value (IV), myristic acid (C14:0), palmitic acid (C16:0), palmitoleic acid (C16:1), stearic acid (C18:0), oleic acid (C18:1) and linoleic acid (C18:2) content. The FatA and FatB genes of castor bean were heterologously expressed in *Escherichia coli* for biochemical characterization after purification, resulting in high catalytic efficiency of RcFatA on oleoyl-ACP and palmitoleoyl-ACP and high efficiencies of RcFatB for oleoyl-ACP and palmitoyl-ACP. The expression profile of these genes displayed the highest levels in expanding tissues that typically are very active in lipid biosynthesis such as developing seed endosperm and young expanding leaves (Sánchez-García *et al.,* 2010). *Arabidopsis thaliana* gene diacylglycerol acyltransferase (DGAT) coding for a key enzyme in TAG biosynthesis was expressed in tobacco under the control of a strong ribulose-biphosphate carboxylase small subunit promoter. This modification led to up to a 20-fold increase in TAG accumulation in tobacco leaves and translated into an overall of about a twofold increase in extracted fatty acids up to 5.8% of dry biomass in *Nicotiana tabacum* (Andrianov *et al.,* 2010). Dimov and Moller (2010) tested genetic variation for saturated fatty acid content in two sets of modern winter oilseed rape cultivars (*Brassica napus* L.) in field experiments under typical German growing conditions. They observed highly significant genetic differences among the cultivars for total saturated fatty acid content, which ranged from 6.8% to 8.1%. Yang *et al.* (2010) identified INDELs, SNPs and SSRs in KAS III, ACCase, Stearoyl–ACP desaturase and DGAT associated with variation in composition and concentration of oil in maize.

These studies suggest the possibility of identifying candidate gene markers in fatty acid biosynthesis genes with their utilization in marker- assisted molecular breeding.

Table 2.1: Sequence variations in fatty acids and TAG biosynthesis pathway genes associated with high oil content in different plant species

Targeted Genes	Descriptions of variations	Gene regions harboring variations	Plant/ organism	References
FAD 2, FAD 3	SNP for high oleic acid and low linolenic acid	Exon	Brassica	Hu *et al.*, 2006
Stearoyl–ACP desaturase	SNP for high stearic acid	Exon	Soybean	Zang *et al.*, 2008
FAD 2	SNPs for high oleic acid	Exon	Peanut	Lopez *et al.*, 2000
FAD 3	SNP for low linolenic acid	Intron-Exon junction	Soybean	Bilyeu *et al.*, 2005
KAS I	SNPs and Indel associated with oleic acid content	5'UTR, Exon, Intron	Soybean	Ha *et al.*, 2010
KAS III, ACCase, Stearoyl–ACP desaturase, DGAT	Indels, SNPs and SSRs associated with variation in composition and concentration of oil	-------	Maize	Yang *et al.*, 2010
FAD 2	3 base pair variation leads to change in amino acid which contribute to high oleate content in oil	Exon	Peanut	Bruner *et al.*, 2001
DGAT1	3 bp Insertion leads to high oleic acid content	Exon	Maize	Zeng *et al.*, 2008
FAD 2	SSR linked to oleic acid content		Soybean	Bachlava *et al.*, 2008
FAD 3	Deletion in soybean FAD 3 leads to reduced linolenate	Exon	Soybean	Anai *et al.*, 2005
KAS III	SNP associated with high palmitic acid content	Exon	Soybean	Aghoram *et al.*, 2006
Stearoyl–ACP desaturase	SSRs associated with high stearic acid	___	Soybean	Spencer *et al.*, 2003
Stearoyl–ACP desaturase	SSRs and INDELs associated with high stearic acid	___	Sunflower	Pérez-Vich *et al.*, 2006
FatB	Deletions associated with low palmitic acid content	Exons and Introns	Soybean	Cardinal *et al.*, 2007

2.5) Disease resistance genes in plants

Disease resistance genes (R) are specificity determinants of the plant immune response. This simple but sophisticated immune system involves an allele specific genetic interaction between a host R gene and a pathogen avirulence gene (avr). Most of the disease resistance genes in different plant species belong to NBS-LRR (nucleotide binding site-leucine-rich repeat) class of proteins (Dangl and Jones, 2001), which are characterized by an N-terminal NBS and C-terminal LRRs. Plant resistance to a range of pathogenic organisms including parasites, fungi, bacteria, insects, and viruses is conferred by a diverse group of disease resistance proteins belonging to NBS-LRR class of proteins (Hulbert *et al.,* 2001). NBS proteins are categorized into two subclasses (Fig. 2.3). Those that have an N-terminal domain with similarity to the Toll and interleukin-1 receptors are referred to as TIR proteins, and those without a TIR domain are broadly categorized as non-TIR proteins (DeYoung and Innes, 2006). Some non-TIR proteins encode an N-terminal coiled-coil (CC) domain that may be involved in protein–protein interaction and signaling (Martin *et al.,* 2003; van Ooijen *et al*., 2007). The NBS domains of both TIR and non-TIR proteins consist of a P-loop (kinase-1a), kinase-2, kinase-3a, and other short motifs of unknown function (van der Biezen and Jones, 1998; van Ooijen *et al*., 2007). The NBS domain functions by binding and/or hydrolyzing ATP (Tameling *et al.,* 2006), and the C-terminal leucine rich repeat (LRR) is implicated in pathogen binding and signal transduction regulation (DeYoung and Innes, 2006; Martin *et al.,* 2003). TIR domains have also been shown to be involved in resistance specificity determination and signaling (Luck *et al*., 2000; DeYoung and Innes, 2006). Together, the domains of NBS-LRR proteins function to directly or indirectly detect pathogen effectors and activate defense signal transduction in plants. All angiosperms evaluated to date contain NBS-LRR encoding genes, but differences exist between monocot and

dicot species. While more than half of the NBS-encoding genes identified in *Arabidopsis thaliana* code for TIR domains (Meyers *et al*., 2003), members of this subclass appear to be absent in cereal species (Monosi *et al*., 2004; Zhou *et al*., 2004). This finding suggests that since divergence >200 million years ago (Wolfe *et al*., 1989), TIR domain association with NBS-encoding genes was preserved by dicots but lost in monocots.

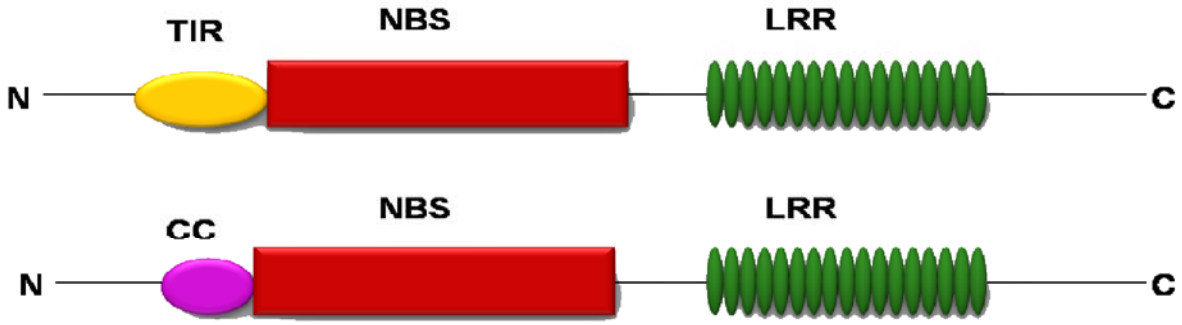

Fig. 2.3: General structure of NBS-LRR gene, showing TIR (Toll/ interleukin-1 receptor like domain) or CC (Coiled coil) at N terminal (Amino terminus); a conserved domain NBS and LRR domain at C terminal (Carboxyl terminus) (McHale *et al*., 2006)

NBS resistance gene (R-gene) families have been detected in numerous plants including *Medicago sativa* L. (Cordero and Skinner, 2002), *A. thaliana* (Meyers *et al.,* 2003; Tan *et al.,* 2007), *Prunus armeniaca* L. (Soriano *et al*., 2005), *Vicia faba* L. (Palomino *et al*., 2006), *Cicer arietinum* L. (Palomino *et al.,* 2006), *Ipomoea batatas* Lam. (Chen *et al*., 2007), *Populus trichocarpa* (Kohler *et al.,* 2008; Yang *et al.,* 2008a), *Rosa roxburghii* "Chestnut Rose" (Xu *et al.,* 2008), *Saccharum* spp. (Glynn *et al*., 2008), *Medicago truncatula* (Ameline-Torregrosa *et al.,* 2008), *Oryza sativa* (Monosi *et al*., 2004; Zhou *et al*., 2004; Yang *et al*., 2008b), *Helianthus annuus* L. (Radwan *et al.,* 2008), and *Vitis vinifera* (Yang *et al.,* 2008a).

In cereal genomes, it is estimated that >1% of all genes encode NBS domains (Monosi *et al*., 2004). Some dicot species contain large numbers of NBS-LRR genes as well (Table 2.2). The *M. truncatula* genome is estimated to contain approximately 400–500 NBS-LRR genes (Ameline- Torregrosa *et al*., 2008), and in sunflower, 630 NBS-LRR homologs were identified (Radwan *et al*., 2008).

Table 2.1: The number of predicted NBS-encoding R genes identified in sequenced plant genomes

Plant species	Genome size (Mb)	Total number of NBS-LRR genes	Reference
Arabidopsis thaliana	125	174 (0.68%)	Meyers *et al*., 2003
Oryza sativa	389	519 (1.38%)	Monosi *et al*., 2004
Populus trichocarpa	485	416 (0.91%)	Tuskan *et al*., 2006
Vitis vinifera	487	535 (1.76%)	Jaillon *et al*., 2007
Carica papaya	372	54 (0.20%)	Ming *et al*., 2008
Medicago trancatula	500	333 (1.63%)	Ameline-Torregrosa *et al*., 2008
Ricinus communis	450	121 (0.40%)	Chan *et al*., 2010

NBS profiling has been shown to be highly effective in detecting rapid polymorphisms for the NBS gene sequences so as to understand whether any two genotypes of plant species carry same or different NBS-LRR genes (Van der Linden *et al.,* 2004). Modified amplified fragment length polymorphisms (AFLP) and nucleotide binding site (NBS) profiling were proposed as new strategies by Hayes and Saghai-Maroof (2000) and by Van der Linden *et al.* (2004) to generate polymorphisms between NBS-LRR genes. The methodology is based on the simultaneous use of an adapter primer matching a restriction enzyme site and of a degenerate primer targeting the NBS encoding region. Conserved motifs in such resistance genes in different plants offer a way to isolate RGAs related to other resistance genes. Currently, RGAs isolated using this approach have been obtained extensively by PCR amplification, with degenerate

primers designed based on the conserved domain of NBS-LRR, such as potato (Leister *et al.*, 1996; Brugmans *et al.*, 2008), soybean (Kanazin *et al.*, 1996), lettuce (Shen *et al.*, 1998), barley (Leister *et al.*, 1998), coffee (Noir *et al.*, 2001), sunflower (Ayele-Gedil *et al.*, 2001), strawberry (Martínez-Zamora *et al.*, 2004), ginger (Nair and Thomas, 2007),sweet potato (Glynn *et al.*, 2008), tobacco (Gao *et al.*, 2010), rice (Yuan *et al.*, 2011), cucumber (Wan *et al.*, 2010) and turmeric (Joshi *et al.*, 2010).

Many cloned RGAs are either closely linked to known R gene loci or are arranged in clusters similar to functional R genes. These studies indicate that the PCR approaches using degenerate primers based on the conserved NBS domains of cloned R-genes can provide an attractive strategy to clone multiple resistance-gene analogs, which can be used either in marker-assisted selection (MAS) or in molecular cloning of new disease resistance genes.

Chapter-III

MATERIALS AND METHODS

The present investigation was carried out in the Jaypee University of Information Technology, Waknaghat, Himachal Pradesh. The material used and the methodologies adopted to achieve objectives of the investigation are described here under.

3.1) Annotation of castor bean genome for SSRs

The castor bean genome sequence version 3.0 (~ 450 Mb) consisting of 28,518 contigs (4X coverage), was downloaded from The Institute of Genomic Research (http://castorbean.jvci.org). The identification of SSRs in castor bean genome sequence was carried out using an in house designed Perl script. The Perl script uses regular expressions to locate SSR patterns in FASTA-formatted sequence files and reports the castor bean sequence contig ID, SSR motif, number of repeats and sequence coordinates for each SSR. The minimum repeat unit was defined as six for dinucleotides, and five for all the higher order motifs including tri-, tetra-, penta-, and hexanucleotides. The FASTA-formatted sequence file was allowed to search each sequence for all possible combinations of di-, tri-, tetra- and penta-nucleotide repeats. Castor bean genome sequence contigs harboring SSRs were annotated for open reading frames, including the 5'UTRs and 3'UTRs, using gene prediction algorithms of FGenesH (http://linux1.softberry.com/berry.phtml?topic=fgenesh&group= programs&subgroup=gfind) (Yu *et al.*, 2002; Goff *et al.*, 2002). SSR motifs were identified in exons, introns, 3'UTRs, 5'UTRs and non-genic regions of the castor bean genome. Primers were designed from the sequences flanking the microsatellite repeat motifs using Primer 3.0 (http://frodo.wi.mit.edu/cgi-

bin/primer3/primer 3_www.cgi) (Table 3.1). The target amplicon sizes were set as 300–400bp, the optimal annealing temperature as 60°C, and the optimal primer length as 20 bp. From the total SSRs identified in the castor bean genome, 302 primer pairs were designed with a repeat motif of >10 representing different genome regions such as 70 from 5'UTRs, 70 from exons, 42 from 3'UTRs, 57 from introns and 63 from the non-genic regions to test their transferability in Jatropha. Twelve primer pairs were also designed from flanking regions of SSRs from exons, introns, 3'UTR and 5'UTR regions of fatty acid biosynthesis genes from castor bean so as to test their amplification (Table 3.2).

Table 3.1: SSRs and their flanking primers

Contig number with corresponding SSR motif	Primer sequence
51460 (TA)14-F	TCATTTGTCTGAAGAGTGTACCTGG
51460 (TA)14-R	CGGAGGGAGAAGGAAGGAAA
52072 (TA)38-F	CACACCAGAAACTTGTTGATCTGTT
52072 (TA)38-R	AAAGCAGAGGAAGAAGAGCACG
54337 (AG)15-F	GTAAGAAGATGTGACCAGTTGGCTT
54337 (AG)15-R	GTTTAGGTGAAGCAAAGAAAGTGGA
57955 (AT)22-F	ACAACTAATCCTTTGGACACAGAGC
57955 (AT)22-R	CACATGATTAGGGTTAGAAACTGTCG
57729 (CT)23-F	CCGACTGACTCTTCTAGACACATGA
57729 (CT)23-R	AGTCCATGAATCTGAACCCACTTTA
36540 (TA)20-F	GTTATTGGAGATGGTGATGGTGAAG
36540 (TA)20-R	CAAGGATAGGATGCTTAGACAAGGA
28859 (AT)31-F	GCCTGTTAACGTTTGATTTGTTCAC
28859 (AT)31-R	ATGGACTGCCCATGTAACTGAGTAT
27934 (CT)27-F	AATGGGAAGGAGGGAGTTAAATAGA
27934 (CT)27-R	GGAGAGCTGCTTTAGTATCCAAAGAT
28962 (AT)26-F	TTATTTGGTCGCAAGCAGTTGTTAC
28962 (AT)26-R	CAAATCCGTCTCCTTTCAGAGTTTA
27956 (AT)22-F	GGCTTACTTACGGACTTAGCAATGA
27956 (AT)22-R	ACATAGATGCTGTGGCCCTGAT
28962 (AT)26-F	TTATTTGGTCGCAAGCAGTTGTTAC
28962 (AT)26-R	CAAATCCGTCTCCTTTCAGAGTTTA
29080 (AT)37-F	GCCATGAACTTTCAGGACCATT

29080 (AT)37-R	GCTTGGGTCCTATTTGTGGTTT
28152 (TA)26-F	GCATGACTCTGGTTAGGAAAGTGTT
28152 (TA)26-R	AACACAGGAAGATGCTCAGAAGTTT
28196 (TA)26-F	CAATTATGGTTAGTTCTCACGGCTT
28196 (TA)26-R	GTCAAATGTTCTACGAATGGTTGTG
29227 (TA)20-F	CTGAAGAGTGTACCTAGGAAACCGA
29227 (TA)20-R	ACTGTCGGAGGGAGAAGGAAG
29214 (TA)34-F	TAAATATCTCCGTATTTCGTGGTCG
29214 (TA)34-R	AGAGGGAGAAGGGAGAAAGCAGT
27394 (AT)12-F	GTCTTCCGAAATCTCTTGTGAACAT
27394 (AT)12-R	TCGAGTTCTGTTGTGTAGAGAGAGG
28297 (TA)33-F	GCCTTAGTTGTGCATTGCTCTATTT
28297 (TA)33-R	ACTCAAACTTATGTCCCAATCGTCT
38354 (TA)29-F	TTCTTCTGGATTGATGGAAGCTACT
38354 (TA)29-R	CGGAGGGAGAAGGAAGGAAA
42884 (AT)31-F	ACTCAACACCAATCATTTCCATCTC
42884 (AT)31-R	CGCTAGTATAGTTCCCTATCCACGA
43672 (AT)37-F	ACTCAACACCAATCATTTCCATCTC
43672 (AT)37-R	AACTTCTTACTCATTGCTCGTGTCA
28200 (TA)32-F	TAGCCTCTAAACCAATCAATCCAAC
28200 (TA)32-R	CCGGATACTCCTGCTAATACTTGAA
46308 (TA)22-F	ACTCAACACCAATCATTTCCATCTC
46308 (TA)22-R	GAAGAGGAGAGCACCCAGTGTT
28859 (AT)32-F	GAGATGGAAATGATTGGTGTTGAGT
28859 (AT)32-R	CGCCTCATCCTCACATTATACACTT
33589 (AG)20-F	GAAGAGGATGACCAATGAAGAAAGA
33589 (AG)20-R	ATTTAACCTCTCCCTTCCCTTCTCT
28842 (AT)28-F	ATGCTATCGGAATAGATCCTTCGAG
28842 (AT)28-R	TGGTAAACAAGAGTTGAGGGTTAGG
28880 (CT)21-F	CCATTTCCATTTCCTCTCTTTCTCT
28880 (CT)21-R	AAGAACCGGTCTTTAGGTAACTTGG
27894 (TA)37-F	AATTACGAGCCAAGTTCGAGTTCTA
27894 (TA)37-R	CAAGCTTGGATAATTTCTCTCTCTCTC
27895 (AT)22-F	GGTAGCAACAACTCTATTTAAGCTCAC
27895 (AT)22-R	TAGCGTTAGAATGTTGACTCATCCA
27945 (TA)39-F	CTTCTTGGTTCCAAGGGACAAC
27945 (TA)39-R	AGTGGACATGTCTAGGCCATATGAT
27956 (TA)27-F	GGAAATGCTCGACAATTAGCTACAA
27956 (TA)27-R	CCGACTCTTGCTAACCTACCCTAAT
27969 (TA)30-F	CCAAATTACAGATGCCTAACAGCTT
27969 (TA)30R	TTTCTGCTCTCACAATTTCCTGAAC
28035 (AT)32-F	GGTGATGAAGTTGGTAACATTGTGT
28035 (AT)32-R	GGAGTGATAAGGATGCACATAAACC
28040 (TA)25-F	CATTTGTGTGTGCCCATTTCATTAG
28040 (TA)25-R	CATTATGGATTAGAAGCATGTCACC
29044 (TA)22-F	AGAGGTACAAGAGAGAAAGCCAACA

29044 (TA)22-R	TCATAGACGCATGTAAATGAGATGG
28076 (TA)38-F	GACTGTTAGCAAATAACAACAGAGAGC
28076 (TA)38-R	GTGATGGCCAAATTTAATGTGATGG
46666 (GA)12-F	AGAAAGGAAAGAAGAGAGTGTTGGG
46666 (GA)12-R	CCACCAAGATCAATGGCTACAG
27375 (AG)12-F	GGCCTAATGTCGATATTGAAAGGAA
27375 (AG)12-R	CAGCGTCCCTCTCTCTCTCTTC
27446 (CT)24-F	GAAAGACCCTTCCTTCCTCATACTT
27446 (CT)24-R	CCAAGGATCATCTCGGCTTAAA
27934 (GA)21-F	TGAAATGTTGAGTTCAAGGTCTAGC
27934 (GA)21-R	TTAAAGATCCTCTGCCTCTTGATTG
27934 (GA)20-F	TGAAATGTTGAGTTCAAGGTCTAGC
27934 (GA)20-R	TTAAAGATCCTCTGCCTCTTGATTG
27956 (TA)19-F	AATGCGTGAATCTGGTGAAACTTAG
27956 (TA)19-R	CCCTTACACCATACGGGACTTACTA
28035 (AG)18-F	GGAACAAGGCACATGATATTAGGA
28035 (AG)18-R	GGGAACGGAATTGTATTAGTAAGGC
28062 (CT)22-F	GCACACCTTCTACATGGTAATTGAA
28062 (CT)22-R	CATAACCCATTTCACCTTGTTTCTC
28140 (AG)23-F	AATTACTAGACATGGGTGCATCTCC
28140 (AG)23-R	CTTATGCAAAGGGTGCACTAAAGAT
28152 (TA)16-F	GAAACGTGATGCACACAAGAAA
28152 (TA)16-R	TACATGTCAAACCGACAATCCTATG
28162 (AG)21-F	CTGAGAAACATACACATCAAGGTGA
28162 (AG)21-R	TAAAGACTATAACGCTCGTCGCTCT
51537 (AT)31-F	TTATGGGTCCTTTCAAGTAAGACGA
51537 (AT)31-R	CTTAGCTTGAGGATACTCATCTGCC
52814 (TA)34-F	TTTGGGCCTACCCTAAAGAGAA
52814 (TA)34-R	CCGTGAGAACAATAATGTCAAACCT
53489 (AT)48-F	TAAATTACCAATCTTGCTACCCGTG
53489 (AT)48-R	TGACAAATACACTGAACATCCTGCT
53508 (AT)46-F	CAGAACGTAATTGACAAGCTTCACT
53508 (AT)46-R	GTGTGTGTGTGTGTGTGTGTGTCTA
54566 (TA)27-F	AGTTGATTTCCTGTTGAACGTGAGT
54566 (TA)27-R	AGTTGATTTCCTGTTGAACGTGAGT
37566 (AT)37-F	GTGCCTAAATCCTGTTGGATAAGAA
37566 (AT)37-R	GCCCTTCAACTATTAGTGTTGGTGT
56672 (AT)36-F	AGATGCGTTGACATTTCCCTTT
56672 (AT)36-R	AACCCAGTTTCCAATCTATGATGTC
59676 (TA)46-F	ATGAACCTTTCGGCTCCATTATAC
59676 (TA)46-R	ATAACTTTGGGCATTTGTTCCTCAG
40553 (AT)26-F	GGATCCAAGCACTGATATATGTTGA
40553 (AT)26-R	CACCTTCATAGTTTCCGAGAAGAGT
42740 (AT)40-F	CAATATTAAGCAAACGAGGGAGAGA
42740 (AT)40-R	GCTGCCAATTTCCAATTTCA
27373 (AT)39-F	TCGAGGTGATATTAGTGTCTTGGTG

27373 (AT)39-R	TCCACTTCGGGATGTAACTAAATTG
27394 (CT)33-F	TATCACCGATTATCGAGATCCAGTC
27394 (CT)33-R	GAAGCTGCTAGAGGGTTTCCAAT
28037 (TA)38-F	ATTCAGTCCATGAATCTCGATCCTA
28037 (TA)38-R	AGGACAACAATTTATCACAAGGACC
28049 (AT)31-F	GGTCATACAACACATGGATAAACGA
28049 (AT)31-R	GAAAGTAACGAGCGTACCTTATCAA
29075 (AT)43-F	TAGGCCAGGTGGTATACCTATGAAA
29075 (AT)43-R	AACATGTCACTTTCTTGGAGGATCT
28108 (AT)35-F	TTGAAATTGCAAGTGAAGGGAG
28108 (AT)35-R	GGAATATGCTTCCGTGAACGTC
29168 (AT)34-F	GTTTGGCGATGAGCTAATTGAGATT
29168 (AT)34-R	GGCTCGAACCTTTCTGATCTAATGT
28327 (TA)42-F	TGTAGATAGCCTTAGCTGTGCATTG
28327 (TA)42-R	GTACTCTCGAGGGAGTTGATTGTGT
29235 (AT)28-F	GAACATGTTTGTGTCAAAGTCCAAG
29235 (AT)28-R	GCACCAATCAGATCCATTCAATTC
49727 (TA)22-F	AAGAGGGAGGAAGGAAAGAGAGAG
49727 (TA)22-R	TCATTTGTCTGAAGAGTGTACCTGG
36845 (TA)18-F	TCCGGAGCTTGACTAGAAGATAGAA
36845 (TA)18-R	GGCAGTTTCAGAGAGTCTTACAACC
37959 (AT)25-F	GTACATGAATGAATCGAATCGCAC
37959 (AT)25-R	TGTGTTATTGGAGATGGTGATGGT
54308 (TA)22-F	GCCTCATCCTCACATTACACACTTT
54308 (TA)22-R	CTTGGAGCATTAGTTAAACCAAACG
57713 (AT)35-F	TGTACCACAGCTCTCAGCTTTACTG
57713 (AT)35-R	GAGGGAGTGGTATATGTGGAGAAGA
41844 (TA)42-F	GCATACTTGTGGCTTCATCTCTCTT
41844 (TA)42-R	TAGTCATGTGGCTCTCATGGAATAG
43365 (AT)23-F	TACTTCCTGATATCATCACCTGCAA
43365 (AT)23-R	CATCTCCTTCCCATCTTTGCTTTAT
44956 (TA)22-F	GGAGGCACGACAAATTAAGAATAGA
44956 (TA)22-R	GAGCCAAAGATTGTGGTTATAATGG
27383 (AT)28-F	CTGAAGGCAAAGTTCTCGATTACA
27383 (AT)28-R	TCTCCACTCCTAACTCCAACAAATC
27394 (GA)28-F	GAAACCCAGAATCAACCACAAATAC
27394 (GA)28-R	GAAACCCAGAATCAACCACAAATAC
28830 (TA)40-F	ACCAGATAGGAGAAAGAATTGATCG
28830 (TA)40-R	AGTGGACACCAAGGAAATTGAGTAG
28842 (AT)22-F	GAGATCGGCTATTGGCTATTCCTAT
28842 (AT)22-R	GAAATAATCATGTAGGCTCAATGGG
27894 (TC)24-F	CTAAACAGATGACCGAGTACCGACT
27894 (TC)24-R	GTTTGCCTATTATCTGTTGATTGGG
29005 (TC)30-F	CTTCATCTTCATCTCCATCTCCATC
29005 (TC)30-R	CCCTTAACAATCATTGCCTTCATAG
28054 (AT)27-F	GTCAGTACCTACAAGCTGCCTTCAT

28054 (AT)27-R	GCCTTTGGATGAACCTATTCACATA
28293 (TC)29-F	CCGACAATACATAGTATTCCCAAGG
28293 (TC)29-R	TCCTGCAGCTTATAACAGATGTGAA
28324 (TA)25-F	TTGACATGTTTGCAACAGGGTT
28324 (TA)25-R	GAGACAAACACGCAATTACAAACTC
29222 (TA)39-F	AACTGGATGCAAAGTTGAGTGACTA
29222 (TA)39-R	AGCTATCTGTTCACACGCAGTGATA
28256 (TA)38-F	TGCTCTGACCACTTGAACTACAATC
28256 (TA)38-R	GCTTATATGCTCCTTGCGGAAA
28350 (AT)24-F	CCAATCGGAGAGTGAAATAGAACAT
28350 (AT)24-R	TCTCGAGTTTAAATCTTGGGTATGC
28329 (GA)13-F	TCGATATTGAAAGGAGTGAGATTGG
28329 (GA)13-R	GTCGCCTTCCTCTTCTTCTTCTT
29216 (AT)16-F	TGTATAGTTGTGCCATGTCAATGTG
29216 (AT)16-R	GTCTCCAATGGATTCTCAGGAACTA
28226 (CT)17-F	AGTGTCCTGTGATCTCCATTTCTTC
28226 (CT)17-R	CGAGGAGGATGACTTTGTCAGTAAT
28219 (TA)15-F	TTGGACCTTATCATCCTCGTCTATC
28219 (TA)15-R	AGTTTGATGTTCCCACCTATACACC
28132 (AT)18-F	GCGGAGTCGATTTCTCATATTAAAG
28132 (AT)18-R	TCATTTGTCTGAAGAGTGTACCTGG
28035 (CT)19-F	CTGACATATCTTATTGGGTGTGGAA
28035 (CT)19-R	TGTAAGAGTATCATCCATTTGCCAG
27985 (TA)18-F	GCAATTCAGCAGTAGCCTGATTAAA
27985 (TA)18-R	ACCAAGTCCTATCCAATTCCATCTT
28952 (AT)17-F	TTTGATGAATGTTTCACCGTCG
28952 (AT)17-R	CTCTCCTCTACTGCCTTTAATGACG
27967 (TA)19-F	GAAATGAGAAGCCTTTACCCTCATT
27967 (TA)19-R	AGAGGGAGAAGGGAGAAAGCAGT
28859 (AT)20-F	GCCTGTTAACGTTTGATTTGTTCAC
28859 (AT)20-R	ATGGACTGCCCATGTAACTGAGTAT
27894 (TC)18-F	TCAGTGAAGTTTCAGGACAGAAGTG
27894 (TC)18-R	CAGTACAGGAACCATCGAATTAACC
32482 (AT)14-F	AGAGGGAGAAGGGTGAAAGCAGT
32482 (AT)14-R	TCATTTGTCTCAAGAGTGTACCTGG
31700 (TC)16-F	AAGCCTGCTTCATACTAATGGACTG
31700 (TC)16-R	TGAAGCTTGTAGACATTAGCACAGG
42824 (TA)12-F	GCCTTCCTTGGAAGATAGGCTAGTA
42824 (TA)12-R	CCATGTTTCTGTATCAGACATTGGA
52072 (CT)15-F	CTGACTCAGCTTGACCTAGCAAA
52072 (CT)15-R	CATGTTTAATGTCGGAGAAACTTCC
27394 (AT)12-F	GTCTTCCGAAATCTCTTGTGAACAT
27394 (AT)12-R	TCGAGTTCTGTTGTGTAGAGAGAGG
28838 (TA)14-F	CTTGCCTTGATGGACAATAACTTTC
28838 (TA)14-R	AGCTTGAGATGGAGTGTTGAAGAAT
27964 (AT)19-F	TGATGGATTGAGAACTGAAGAGGAT

27964 (AT)19-R	ACTCTAATTAGGCCCAGATTCCAAC
36953 (CT)14-F	AATTCAGTGAAGACGGTGGTTCTAC
36953 (CT)14-R	GATTGTAAATGGAGTTTGGAACTGG
43672 (AT)21-F	TTAGATTCACTGCTTTAGGCTTGCT
43672 (AT)21-R	TTGAACTTCTTACTCATTGCTCGTG
47083 (TA)15-F	GGAAAGCCCAAGAACTTGGAAA
47083 (TA)15-R	CATGTATAATTTGAGGGATGGCTTC
27446 (GA)16-F	GTTAAGTCAAACGTTGTGTTTGTGG
27446 (GA)16-R	CTTAGCTCATTCCTTGGAGTTGAAG
33564 (AG)14-F	GAAAGGCTAAAGGAAAGTTTAGTGG
33564 (AG)14-R	CGGCTCCTCCTTCTTCTTCAC
27914 (CT)15-F	TTTCCAAGATATACTCTCGAGGTGC
27914 (CT)15-R	TGAGTGAGTGAGTGAAAGGTGGATA
28883 (GA)18-F	CAGATAAACGAAACCCTAAATTCCC
28883 (GA)18-R	GATACATTCACCTCTTCCCAGCTTA
27918 (TA)15-F	GAACGAGTGAAGAAGAAACTTTGGA
27918 (TA)15-R	TCTGTTGAATACCTTCGTTACAGCA
28859 (AT)32-F	GAGATGGAAATGATTGGTGTTGAGT
28859 (AT)32-R	CGCCTCATCCTCACATTATACACTT
27942 (TA)22-F	TTTGAGTGCTCCTATTTGGCTAGAA
27942 (TA)22-R	CCAAATGACAAGTAGGCAGAACTTT
27942 (TA)16-F	TTCCATGTCCTAGTTTCTTTCTTGG
27942 (TA)16-R	GCTTCCATATCTCTCTCTCTTTCCC
27956 (AT)26-F	GCAAGAAATAAGGTACAACCGAAAC
27956 (AT)26-R	GTGAGCAATTACCAAAGGAAACAAG
28934 (AT)16-F	GCTAGCAATTGGCCAAACTAAGAT
28934 (AT)16-R	CGTTTATCTGTGTGTTACACGTCTG
27964 (TA)21-F	AAACAGTGTTAAGCAGCAAGAATCC
27964 (TA)21-R	TACTCGAGTTGTCACATGGTTTGTC
27970 (CT)13-F	TTTAAGACTGAGATTGCAGTCAACG
27970 (CT)13-R	GAAAGGCAAAGAAGAAGTTTAAGGG
27973 (GA)17-F	GCTGCAAGAAGAATAGAGAAACAGG
27973 (GA)17-R	GTCGTCTCCAGAGAATTAAGGGAG
28035 (TG)13-F	AAATTGTGTGCATGCTTGGAGT
28035 (TG)13-R	CTCATTTCTTCTTCTCATTATCCCG
29010 (TC)13-F	GGTTAGCTTAATTGGCTCCTTCAAT
29010 (TC)13-R	AGTCATAGCACAAAGAAGCATGGAC
28040 (TC)18-F	GCCTTTAAGAGATTGATGGCAACTA
28040 (TC)18-R	AAGTATTCATATGCCCTAAGCCTCC
28077 (AT)18-F	CATAGTTTCGGCTTAATCAACGAGT
28077 (AT)18-R	CGGCTTGTGAGTAATAAGTTGGTCT
28093 (TA)24-F	TTGGAGGTTACAATCAATGGCA
28093 (TA)24-R	GCATGTGCCCGAATTGAATA
28109 (AT)27-F	CCATTTGGTGTTAATCACATGAGTC
28109 (AT)27-R	GACAATAGTGATGTTGGATTCCACT
28111 (AT)14-F	ATCTCGTAACCTCAGAAAGTTGTGC

28111 (AT)14-R	TCTGGGATGAGGATCGTAGTAGAAC
28147 (GA)20-F	TGATGGCTCAAGTACAAGTTCTAGG
28147 (GA)20-R	CCTTTGTGAGTACACAGAGAGAGAAA
28152 (TC)17-F	GAGTCACAACTCATAACCCTTTCGT
28152 (TC)17-R	GGGAGAAGGAGAAGAGTGATAAAGG
28161 (TA)23-F	AGATTTAGAAATGGTAATAGGGCGG
28161 (TA)23-R	GACCTATCCGTGTCGTGTAGATTT
28200 (TA)32-F	TAGCCTCTAAACCAATCAATCCAAC
28200 (TA)32-R	CCGGATACTCCTGCTAATACTTGAA
28202 (AT)25-F	ATTCAGGCCATCCACATAGTCTAAC
28202 (AT)25-R	GACCCTATTGATTGATTTAAGAGCC
28211 (AG)24-F	TTCTCCAAAGCCAGAGATAGACATT
28211 (AG)24-R	ACCTCTCAACAGGATTTGAACACTC
28226 (TA)31-F	CTTCTCAGCAACATCACATCAAACT
28226 (TA)31-R	CGCTAAGTTACATAGGACAAAGGGA
29222 (TA)23-F	AATTAACAGTACTGCCACATGGTCC
29222 (TA)23-R	CCAACCGTTACCCATTTGATATAGT
28350 (AT)24-F	CCAATCGGAGAGTGAAATAGAACAT
28350 (AT)24-R	TCTCGAGTTTAAATCTTGGGTATGC
28304 (TA)15-F	CAAGTGAGCATCAATGAAAGGAAAG
28304 (TA)15-R	GAGAGGTAGAAGGAAGGGAGGAAG
28301 (AT)25-F	GCATGGAAATTCAATTCTGTGCTAC
28301 (AT)25-R	TTTGGTGATGAGGATTGTTGCT
28281 (AT)19-F	TTGGGTTAAACCTACACTAGGTATCG
28281 (AT)19-R	ACAAGCTAAAGTCAAATAGCGAGGA
29227 (TA)20-F	ACCAGTCTCGTAGAAGGGCTAAACT
29227 (TA)20-R	GAGGATGGCTGCAACAACTG
29222 (TA)24-F	TGAGAGTGTTACAGAGAGTGTTGCTT
29222 (TA)24-R	TGTTACTGCTGAAACATGGAATGAC
29212 (GA)16-F	ATTGGACCAACAACTCTGAGGAC
29212 (GA)16-R	CCCTTAACCTTTGCTTTAATGAGGT
29212 (AT)27-F	CTGTTTGAGGATCAGACTTTGAAGC
29212 (AT)27-R	AAAGAAATAATGAGGAGGGAGGTTG
29195 (TG)19-F	GGTAAGGAATTACCAGACTGCTCAC
29195 (TG)19-R	GGCATGATGGGTGACAACTAAA
28226 (TA)21-F	GAACATTATCACTGTTGTCTTGGGA
28226 (TA)21-R	AGAATAAGGACAAGCATCTAACCCA
28208 (TA)15-F	GCGTCATCTCTGTAGGCAGTAGAAT
28208 (TA)15-R	GCTTGCTGCTGGTAAGGATTTATTT
29172 (AG)19-F	TTGTACTTTGGGATAAAGTGGAGGA
29172 (AG)19-R	AAATCCCAGTCTCTTAGCCAATACC
29171 (AG)25-F	AAAGTAGTACCGACAAGCAGTCGAG
29171 (AG)25-R	TGCATTTCTCCAGGAACTCAGTATT
28202 (AT)22-F	AGCAAGAAACCATACTTCGAGTGTC
28202 (AT)22-R	GAGATGCCAACCTTTGTGATTAGTT
29168 (CT)21-F	CACTCTGTAGGACGCTCCACTATTT

29168 (CT)21-R	TGGTGTGAATTAGAGTTTGGTGAGT
28200 (CT)16-F	GAATTGGCGTCTGCTATTATGAGTT
28200 (CT)16-R	TAGCGATTAAGGAAGAACAAAGACG
28192 (TA)13-F	GAGGTGGAGGATGATAGAGGGTAAT
28192 (TA)13-R	GACCAAAGCAAACTCGAATACTCAC
28152 (CT)19-F	AAATAGTGTTCCGTCCGTATTGTTG
28152 (CT)19-R	ATCCATAGAGCAGGACTGTGATAGG
29118 (GA)22-F	GGATCATCTATGCTTGTAGCGAACT
29118 (GA)22-R	TTAACAGAGTGAAGGATTTGCCTGT
29117 (TA)13-F	AAATGTATTGAGCCCAAGTCGTAAG
29117 (TA)13-R	TGATGTCTCATGCATATCTCTTTCC
28128 (TC)23-F	GAATTTAGAAGCCACATTTGAGACG
28128 (TC)23-R	CCTATGTAACCCAAGAAAGACGATG
29093 (CT)21-F	GGACTTCTAACAATCATTTAGGTGGG
29093 (CT)21-R	TGTGATAATTGCCTGAAGGATATGG
28102 (TA)19-F	AGGGCATATATATTTCCAGGTTGGT
28102 (TA)19-R	AACAACACACACCTAAAGTGACAGC
29080 (GA)20-F	TGAGAGTGGACTTGGATTAGGTTCT
29080 (GA)20-R	AGCCTGCAGTTCAATCTGTTTACTT
28035 (GA)21-F	TTCAATAACAGATTTGGCTAGGCTC
28035 (GA)21-R	GACAATTGAAAGGTGCAATCCTAGT
29005 (TA)19-F	TATTTGTGTGTGTATGTGTGCGTCT
29005 (TA)19-R	GATCTAAATGGTTCGAGCTTCACTC
28966 (AT)22-F	CCTTGTTGGAGGGAAGAGAAATAAT
28966 (AT)22-R	TTTACATCATCTAGCGTGCCATGTT
27985 (TC)23-R	AAATTACTTCCAGAGACGAATGCCT
27985 (TC)23-F	CTCACCAACAGGGTAATTTACATGC
27973 (AT)24-F	CAGACCCATCTGATCATCATTGTAG
27973 (AT)24-R	TCCTCAGGTAAATTGCTCATCTTTC
28350 (AT)14-F	AGCAGCATCAGGCAGTTGTATTT
28350 (AT)14-R	GACAATACGGTTTGCTACAACCATT
29227 (AG)14-F	GGGTGAAATTAGAATTATGCAGGCT
29227 (AG)14-R	CAGCGTCCCTCTCTCTCTCTTC
29220 (AG)15-F	GAGAGAGAGAGAGAGACAGAGTGCG
29220 (AG)15-R	CTACTTTCTCTCCACACCGGATCT
28154 (AT)13-F	AGACCAACGCCTTATTGACTACATC
28154 (AT)13-R	CTTGAAACCAACTTACTTGTTGCCT
28102 (TA)13-F	GTGTAGGAGGAGGAGTGAGACAATG
28102 (TA)13-R	TTGATACAGTTGAGGTATGCTCCTG
28092 (CT)15-F	CACACTTACGCATCCTGTTTAGTTG
28092 (CT)15-R	TAGTTTCCAGCATGTGAGATTACGA
28035 (CT)13-F	GTCGACTCTATTGGTGCCACTG
28035 (CT)13-R	GAAGGAAGAAAGAAAGAAGGAATGG
29005 (TC)13-F	GCTGTCCCATTCTAATTTGTTTGAC
29005 (TC)13-R	CACATATACCAGACATCCTCTTCCC
27985 (AT)13-F	GATGTAGATTTGGTTTCAATGAGGG

27985 (AT)13-R	TACATAATGCGTGATGGACTAAGGA
27985 (TC)15-F	GTTAGAGAAGGCCAAATTGAAACTG
27985 (TC)15-R	ACTTCATTACGTCGAGAGATATCGG
53467 (CT)12-F	CAGCGTCCCTCTCTCTCTTCC
53467 (CT)12-R	AGGAAGTTGAGGGACCAAATTGTA
52562 (AG)12-F	GTCGATATTGAAAGGAATGGAAGTC
52562 (AG)12-R	AAAGAAGCTTCGGTCATCTTCCT
28308 (TA)24-F	GATGTGATTCGACTTTATGTCGTTG
28308 (TA)24-R	GCACAAGCGTACTACTACAAACTAGGG
28333 (TA)21-F	TTCAGCTCTTCTTGTTGATTCCTCT
28333 (TA)21-R	GCTGGATTAAACAGATGGGTTACCT
28254 (AT)21-F	TGCTGCTATTATAGGATGGGAAGAG
28254 (AT)21-R	TATACGTAAATTTGCCAAGTACGGG
29217 (AT)20-F	ACAGTCAATCTCATCCACAGTCTTG
29217 (AT)20-R	TTCGGTTTAATTCACTGAACCCTAC
29212 (AT)29-F	GAAAGCTAGAAATCAATGAACGCAC
29212 (AT)29-R	TCATTTAGTACATTGACCGGAGACA
28238 (TA)16-F	TTTGTGGTTGTACAAGCTTTCTAGG
28238 (TA)16-R	CCTGAAATGTTGAGCAATGTGGTAT
28219 (TA)13-F	CAATCTAAATCCATTCCAGCTTCAG
28219 (TA)13-R	TCAAACTCAAGACATCTACCTCCAA
29169 (AT)20-F	GTTAAGCTAGGGTTCTTGAAGGAGG
29169 (AT)20-R	CTGCGATAAGTGTACGAGGAAATG
28200 (AG)28-F	AATACTGGATATGGAGGAAAGCACA
28200 (AG)28-R	AATGAGAAAGGACTGCATGTCAAAG
28184 (AT)15-F	AATTCTGGGTGTTACAATAACCTCG
28184 (AT)15-R	CCCATATTACGGTGTTGTCTTCTG
28161 (AT)20-F	CTAGAAGACTGGAAACGGAGCAATA
28161 (AT)20-R	CCTTTCTGAAGAAGATGAACACCAT
29117 (AT)20-F	TGAGTCTAAACAAGCCCAATTAAGG
29117 (AT)20-R	TAGCATCCTCCATTAATCCTAGCTG
28138 (TC)19-F	AGCTAACATTCTCATGAGCTTCCAT
28138 (TC)19-R	TTTCCTTGCCTGCTTCCTTC
28129 (TA)29-F	CTTACTCTCCAAGCATTCTCACACC
28129 (TA)29-R	GTCGGACGACCGCCATTGAAA
29044 (AT)18-F	CATGAGTACGATGCTTCTAGGGTTT
29044 (AT)18-R	CCAGAAGTCCTAACTACAAACGTCC
28923 (TA)22-F	GATGGAGACTCGAATGAAGATGACT
28923 (TA)22-R	GTATGCGCATGTGAATAGTGGATT
28909 (AT)26-F	GTCAGTACCTACAAGCTGCCTTCAT
28909 (AT)26-R	GCCTTTGGATGAACCTATTCACATA
27895 (TA)26-F	GCAACAAGTTGACAACAAATAAGGC
27895 (TA)26-R	GCCTAAGGAGTTCCCTTTAAGCTAT
27882 (AT)26-F	GTTTCTTCGTGTTTGCGAGGAGTAA
27882 (AT)26-R	CGACTCCCTAACAGATTTAGAAGCA
28830 (AT)24-F	ATGTTCCCTTTCTACGTCTCTTCCT

28830 (AT)24-R	TCTTCTCTCTAACGCACTCCTGAAT
28829 (TA)33-F	TTGGCTCATAATAACTCCTCAAAGC
28829 (TA)33-R	GCGAGTGCTGTCTAAAGCCTAATTT
27866 (TA)25-F	TAGCTATTCTCATCATTCCTCCACC
27866 (TA)25-R	GTCAAAGGTACGCTTCTTGACCTAA
32500 (AT)20-F	GGAGTTAAGCATCCCATAGTTTGAA
32500 (AT)20-R	ATTTGTCAAAGTGCATTGTAGTGGC
31721 (AT)21-F	AACAAATAGCTTTGCAGCAGTAACC
31721 (AT)21-R	ATAATATGGTGCAACGGTTCATCTC
47284 (AAT)11-F	TAGAGAGATGTCTAATCACACGCCA
47284 (AAT)11-R	AAAGTTCCTAATTATGCCATCGTCC
27914 (AAT)18-F	CACTCCCATTACGCTCTACCAAATA
27914 (AAT)18-R	GTATGGCGTAATTTGAAGCATGGAT
29206 (ATA)23-F	ACAGGCACAGAAGCAGAAATTATGT
29206 (ATA)23-R	CGTCTTTCCTTCTCTAGCCTTCTCT
28324 (TTA)10-F	ATGGTTTGGTCCTACTATGGAGTGA
28324 (TTA)10-R	GAGCAGTCTTAGAAGAAGATGGCAG
36262 (TTA)13-F	GTGAGTGTTGAGACCAATTCACATC
36262 (TTA)13-R	AGCCCAATTTGTATTTCTTACCTCC
41833 (TCT)12-F	ATTAGCTGCTGAATATCAATCAGGG
41833 (TCT)12-R	CGATCAACTTCAGACTTGTCTTGTG
27893 (ATT)18-F	GAGATTACTTAGAGTTGCAGGCATTG
27893 (ATT)18-R	AAAGGTTCAAGAGAGTGAAGGTTGA
27914 (AAT)13-F	AATGAGTCCACATAATCCCACTTTC
27914 (AAT)13-R	ACGAGTTTATGAGGATTGTCTTTGC
27894 (TTA)23-F	AGTTCGACTAGTTCCACCAGAATTG
27894 (TTA)23-R	ATTGCCTTCCTCGTCCTAGAGTAAT
27985 (ATA)22-F	TCATAATCGGTCTATTCTTCGTGCT
27985 (ATA)22-R	TTGGTCGTTGTATGGTTAGTCAAAG
29171 (TTA)22-F	GCAGAAACTCGGTAGAACTGTGAGT
29171 (TTA)22-R	GGCATAATCTACTGTTATCTCATCCC
28333 (TTA)12-F	TACCGGATCTCTTGTTCTCTTCTGA
28333 (TTA)12-R	CTTCGATTCTTGGTTTAGTGGAAGA
28226 (TTC)10-F	CACGTCACAACCAGACATCTTATTC
28226 (TTC)10-R	CTTTACCGTGTCTGGTAGCCATAGT
28211 (TTA)11-F	CTTGATCACCTTCTTTCTTGCTACC
28211 (TTA)11-R	TGGCTAGTAAAGTTGTTATGGCACA
56420 (TCT)11-F	GTCATGGCCGTTCACTACTGTC
56420 (TCT)11-R	ACCAACTTGCGATAGAGAAGAGAAA
57713 (CTT)8-F	GTAGCTAGGATTAATTGGAGGGACC
57713 (CTT)8-R	CAGCAAGAGGCTGTTCTAGAGGT
41929 (CCA)11-F	AGCAGAAGGCAAGAAATCAGAACTA
41929 (CCA)11-R	TGGAAAGAGAAGTTGGAGAAAGAGA
33978 (GAA)10-F	GAAGATGGCTTGAAGGACATGAT
33978 (GAA)10-R	CTTGGAGTTGTAGCCTTCCTTCTC
34878 (TCT)14-F	ATCCACCATTTGTGAGCCTTTAC

34878 (TCT)14-R	AGACCAACTTGCGATAGAGAAGAGA
28859 (AGA)12-F	GGTTAATTGGAATTTCTGCAGGATG
28859 (AGA)12-R	GGTTAATTGGAATTTCTGCAGGATG
27956 (TTC)11-F	ACCTGAATGGAGATAGAGAATGCAC
27956 (TTC)11-R	GCAGTCAACAAGAAGATGATCAAGA
27956 (TCT)10-F	CACAGCCTATTCCTTTCCTTTGTTA
27956 (TCT)10-R	GAGAGTGTTTCTTCCATCTCTCTGG
27985 (AGA)10-F	TCTCTTTCTCTCTCTGTTTCGCATT
27985 (AGA)10-R	GTAGGAATAGGTGATGATGAGGACG
28092 (ATA)10-F	CAATAGTGACACGTCATCAACTTCC
28092 (ATA)10-R	AGAACGCGGAATCTCTCTTAATTCT
28134 (ATA)20-F	GAAGAAGAATGAAAGGTCACTTGGA
28134 (ATA)20-R	TCAGTCCAATCACTATCAGTAAGCG
28162 (CAT)12-F	GCATGCAAACCCTGAATTATGTACT
28162 (CAT)12-R	GCTGCTGACTCTGTTTCTCCTTCTA
29222 (GAA)15-F	TGGAAGACGAATACTATGACGATGA
29222 (GAA)15-R	CAGGTGCTACTTCTTCTTCTTCAGG
50926 (AAT)20-F	AGGAAAGTTATAAGGACAAGTGGCA
50926 (AAT)20-R	AGGAAATGAAATGGTGATGGTAGAG
32351 (GAA)15-F	AGACCAACTTGCGATAGAGAAGAGA
32351 (GAA)15-R	CCTCGTAATTACACGACATTTACGG
33124 (GAG)8-F	GATGGTGAGATATAGGTAGGCATGG
33124 (GAG)8-R	ATGTTCAAGTAGCTAATGCCCACTC
59658 (TTA)23-F	GCCACAATAATCACAAGGGATAAAC
59658 (TTA)23-R	GTAGCTTCCATCAATCCAGAAGAAA
40845 (TTC)13-F	TTCTGACGTGGCACTAAAGAGTAAA
40845 (TTC)13-R	GCATATGTTGATGTCCGATTGTAAG
32351 (GAA)15-F	AGACCAACTTGCGATAGAGAAGAGA
32351 (GAA)15-R	CCTCGTAATTACACGACATTTACGG
27894 (TTA)23-F	AGTTCGACTAGTTCCACCAGAATTG
27894 (TTA)23-R	ATTGCCTTCCTCGTCCTAGAGTAAT
28863 (TTA)26-F	GCGGCTAAAGCTGTTAGAGATCAAT
28863 (TTA)26-R	AATACGAACGTAAATCCGGCTAGTT
28035 (TTA)21-F	CTAACTCATAATTCTGCCGAGTTCC
28035 (TTA)21-R	ACAACCTAATTGGAGTCAAATGCC
28327 (AGA)16-F	TGCAGAGCATGTCTGTTTATAGCTT
28327 (AGA)16-R	ACTTCATTGTCTCAGGGATCATAGC
41833 (CTT)10-F	ACCTTCCAGAATGAAGAAGATGATG
41833 (CTT)10-R	CCGGATAAATTTCTCTCCACAAGAG
44057 (AAT)14-F	ATTTCCATCCACCATTACCCTACTT
44057 (AAT)14-R	AGACCAAATAAGACCGAGTCGAATA
28842 (AAT)14-F	TTACTCTCATTAACGGAGCGTGTAG
28842 (AAT)14-R	CAGTCGTTCGTACCAAATAACCATT
27890 (AAT)13-F	AAACAGAAGACAGCTTGAAACAAGG
27890 (AAT)13-R	AAACAGAAGACAGCTTGAAACAAGG
28883 (ATT)14-F	TTATTCTCAGGGCACACAATTAAGG

28883 (ATT)14-R	AGCTGTTTATCCTCCTCAGCTAACA
29212 (TAT)17-F	CAGCCTTTCGGCAGATAATAGAATA
29212 (TAT)17-R	GCTTTAATATGGAGCAGGATTTGTG
29206 (AAT)13-F	ACAGGCACAGAAGCAGAAATTATGT
29206 (AAT)13-R	CGTCTTTCCTTCTCTAGCCTTCTCT
27942 (AAT)12-F	TTGCTTCTCATGGAGAGATATTACC
27942 (AAT)12-R	CCATACTTGAAAGGAATGGCTTATG
36262 (TAT)12-F	CTATCGACAATGAAACTTTCTCCGA
36262 (TAT)12-R	GAAAGTTATAAGGACAATTGGCACC
41177 (TTC)11-F	AGTTGATGAATCTAACGGTGCTGAT
41177 (TTC)11-R	TTGTAACTCGTTGGATAGTGAGCAA
27916 (ATA)10-F	GCTTGATTCATTTAAGAACCTCGTG
27916 (ATA)10-R	CATGAATAGTTCGAGCTTGCTTTGT
27940 (ATT)10-F	TAGGCCAGGTGGTATACCTATGAAA
27940 (ATT)10-R	AATCAAATTTGGTACGGCGTGT
27956 (ATA)12-F	AAGGTTGTTCCTGTTAACATCGGT
27956 (ATA)12-R	GTTTCGGTTGTACCTTATTTCTTGC
28962 (ATA)13-F	GTGATCATCGCCTTACTAGTCCATT
28962 (ATA)13-R	CGTTGATTGTGTACGTTAGATCTGC
28094 (GAT)12-F	GCTGTTATTTGGTACAAGGAAATGG
28094 (GAT)12-R	CATCATAGTTGTCCTCATCATCCTG
28196 (TAA)17-F	AGCTTGGTTTGTTCATAAAGGTCAC
28196 (TAA)17-R	AAAGCACACGAAGCTAATCTTTCAC
29171 (TTA)22-F	GCAGAAACTCGGTAGAACTGTGAGT
29171 (TTA)22-R	GGCATAATCTACTGTTATCTCATCCC
28324 (AGA)20-F	TTAAGAGTAAGTGAGATGGGTTGGC
28324 (AGA)20-R	TACATTTCCTGGCCTCAATTACACT
28226 (ATA)19-F	TGCTCCCAGTAAGCATAAGAAGAAG
28226 (ATA)19-R	CTTGCTCGGTTACCATTACCATTAC
28281 (TTA)13-F	CCCTGCTTATATCAAAGGGTAAGAA
28281 (TTA)13-R	GAAACAAAGGCCAATTTGATCCTC
28153 (AAT)16-F	AATCACATCAGTTGTAACGGCA
28153 (AAT)16-R	ATAATCTGATGGTTCAGTCAGCTCC
28094 (TAT)15-F	TGGTATAGTTGTTATGCCTCATTGC
28094 (TAT)15-R	CTCAAATAAGAGAGTGGTGAAACAGAG
51866 (AAAT)7-F	GACTATTGCAGTCTCCACAGGAAAT
51866 (AAAT)7-R	AAGAGAGCCGAAAGGCTACTAGAAT
53377 (TTTA)6-F	TTAAGTGACCAGACTGGAGGTAAGG
53377 (TTTA)6-R	CAACCATGCACGTAAATACAATCC
38055 (AATA)6-F	TGGGAACTATTGCATTCTTCTCAG
38055 (AATA)6-R	TTAAGTGACCAGACTGGAGGTAAGG
36229 (TTAT)6-F	CTCTGGGCTGAGTTTAGTTAATTGC
36229 (TTAT)6-R	CGAATCCATATATCAAACCCTGAAC
41262 (TTTA)8-F	TTAGGTGTAGGTTGCATAAAGAGCC
41262 (TTTA)8-R	TTGCAACCTCTTCCTTAAGAATCTG
47136 (ATTT)10-F	GATCTGATCTCCGCTATTGTGAAAT

47136 (ATTT)10-R	TATTACTTCATCCCGCGTTATCTGT
27964 (TTCT)7-F	AACGCGAAACTCTATGCATATCTTC
27964 (TTCT)7-R	ATTAAACCACGACAAGGACCAGTAA
31548 (TTTA)8-F	GGTTATTATGAATTTGTCTGCCCTG
31548 (TTTA)8-R	CGAATCCATATATCAAACCCTGAAC
37474 (AATA)6-F	CGAGAGGACTAACAGTCACGAATCT
37474 (AATA)6-R	TAAGAGAGCCGAAAGGCTATTAGAA
57368 (TTAT)6-F	AGGGAGTACTTCTATCCTTGCCTTC
57368 (TTAT)6-R	GCACTACCTAACCGCATAAACCTAA
46497 (TTTA)6-F	TAAGAGAGCCGAAAGGCTATTAGAA
46497 (TTTA)6-R	CAACCATGCACGTAAATACAATCC
55461 (AATAA)7-F	AACCATGCACGTAAATGCAATC
55461 (AATAA)7-R	AAGAGAGCCGAAAGGCTACTAGAAA
40189 (TTTAT)5-F	ACTAGTGAGGATACACCGTCAGCA
40189 (TTTAT)5-R	AAACTCAAGACAGGCACAACATC
41398 (AAAAG)6-F	CTTCTTCAAGGTCGAAAGTTTGCT
41398 (AAAAG)6-R	CAAACTTGATTTGCCTAAGTGTGTG
44109 (AGAAA)16-F	CCTAGAAATGGTAGCCCAAGGTTAT
44109 (AGAAA)16-R	AATCTCGTTCATCGAGTTCCTTCTT
31511 (TCTTT)9-F	CTTGTCTTTGGAGTCTTGTGGAGAT
31511 (TCTTT)9-R	ATCATCATCCTCTTCCAGTGCTAGT
34784 (TTTGC)22-F	TTCTATTGGGAAGTGATGCTGTGTA
34784 (TTTGC)22-R	CTTGAGGCGAGCTAACGAATC
28872 (AGTTG)5-F	AGGTACCAACTCGGACATAATTGAA
28872 (AGTTG)5-R	TTGAGACCAAAGCATACAGTGGTTA
57946 (TATTT)6-F	TAATTACAATCCAGCCAGGTAAAGC
57946 (TATTT)6-R	CCAACTACACCACAAACACAACAAC
46597 (TTTAT)8-F	AGCAGACTCCTAGCGGAAAGAAA
46597 (TTTAT)8-R	ACTCAAGATAGGCACAACATCGGTA
52899 (TTTAT)8-F	AGCAGACTCCTAGCGGAAAGAAA
52899 (TTTAT)8-R	AAGATAAGCACAACATCGGTAGGTC
41328 (TATTT)6-F	CCTATCTCACCCTTCTCACTTTCCT
41328 (TATTT)6-R	CCAACTACACCACAAACACAACAAC

Table 3.2: SSRs identified in the candidate genes for fatty acid biosynthesis in castor bean

Candidate gene	SSR motif	Gene Location of SSR	Primer sequence	Amplicon size (bp)
ACCase	(ag)13	Intron	GAAAGTTGTTTGATTACCATCCCAG AGAGTAACTGGACTTCCAAGTGTCG	268
Malonyl transferase	(tc)7	Intron	GATGAAATAGCAAGTGATGTTGCTG ACACAGAGCCTACCCTAATTAACCC	235

Malonyl transferase	(tat)6	Intron	GGTAGGAGATAAAGTGCAGAAAGGG AATAGAGGCTGCATTGAGAGAACAG	385
Dehydrase	(aac)4	Exon	ATTTAGGCCTCCTTCTTGTACGACT TCTGCTGGTTGTTTCCTAGATTGAT	281
Dehydrase	(tg)15	3'UTR	ATCAATCTAGGAAACAACCAGCAGA TTTCTGGTCATAGATCCCTGTCTTC	396
Dehydrase	(ctttt)6	Intron	ATCAATCTAGGAAACAACCAGCAGA TTTCTGGTCATAGATCCCTGTCTTC	396
Dehydrase	(ag)13	Intron	ACCAGATAGGAGAAAGAATTGATCG AGTGGACACCAAGGAAATTGAGTAG	396
Dehydrase	(taa)7	Intron	TTTACATGTGCGTCAGACAGTGAC TGTGAGAATTAGACTCAGGAAGAGC	396
Stearoyl ACP Desaturase	(ag)15	5'UTR	CCATCCATCCTGCTCTATACATTTC AGTTTAAGGTCTGTGGAAGGGAGTT	396
Linoeoyl desaturase	(gaa)4	3'UTR	TAGATTTCTAGGACATCCAAGCAGG GCATCATACAGCTCCTCATATTCCT	334
Palmitoyl ACP	(ata)22	3'UTR	TCATAATCGGTCTATTCTTCGTGCT TTGGTCGTTGTATGGTTAGTCAAAG	398
Palmitoyl ACP	(aga)10	Exon	TCTCTTTCTCTCTCTGTTTCGCATT GTAGGAATAGGTGATGATGAGGACG	131

3.2) Retrieval of fatty acid biosynthesis genes sequences for comparative genomics

Thirty-two genes involved in the biosynthesis and storage of fatty acids were retrieved from Arabidopsis database (http://www.arabidopsis.org/browse/genefamily/acyllipid.jsp) by referring to the comprehensive lipid gene catalog provided by Beisson *et al.* (2003). The selected genes covered all the major biochemical events in the biosynthesis and storage of fatty acids. The protein sequences of these genes were used as query against castor bean database in TIGR (http://blast.jcvi.org/erblast/index.cgi?project=rca1) and soybean database in soybase (http://soybase.org/GlycineBlastPages/). Full length coding sequences of Brassica were downloaded from GenBank (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Protein function domains were examined with 'CDD' from NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml).

3.3) Prediction of gene structures

Gene models for castor bean and soybean genomes were downloaded from Phytozome (http://www.phytozome.net/). The Arabidopsis gene models were downloaded from TAIR (http://www.arabidopsis.org/). Arabidopsis, castor bean, *B. rapa* and soybean sequences were further annotated for gene models (open reading frames, including the 5'UTRs and 3'UTRs) using gene prediction algorithms of FGenesH (http://linux1.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind) (Yu *et al*., 2002; Goff *et al*., 2002). The sequence identity among *B. rapa*, castor bean, soybean and Arabidopsis genes was confirmed using ClustalW in MegAlign in DNASTAR (DNASTAR Inc., Madison, WI, USA). The expression status of candidate genes belonging to different families was searched in ESTdb of NCBI (National Centre for Biotechnology Information) http://www.ncbi.nlm.nih.gov/BLAST/ and TIGR (http://compbio.dfci.harvard.edu/tgi/cgi-in/tgi/Blast/ index.cgi).

3.4) Identification of critical candidate genes involved in fatty acid biosynthesis in Castor bean

The candidate genes of fatty acid biosynthesis identified in castor bean were present in multiple copy numbers. The critical genes (genes involved in fatty acid biosynthesis in seeds) were identified through *in silico* expression and comparing with the functionally characterized genes of fatty acid biosynthesis pathway in Arabidopsis, Brassica, soybean and cotton. The *in silico* expression status of candidate genes was searched in ESTdb of NCBI, TIGR at http://compbio.dfci. harvard.edu/ tgi/cgiin/tgi/Blast/index.cgi. Annotation of candidate genes was also checked with the coding DNA Sequence (CDS) of fatty acid biosynthesis gene sequences for *J. curcas* available in GenBank. Forty nine primer pairs were designed from different regions of candidate genes such as intron, exon, 5' and 3'UTRs (Table 3.3). Twenty one primer pairs

were designed from the ends of exons to amplify intronic regions, 18 to amplify 5' & 3'UTR regions and 9 to amplify exons of the candidate genes (Fig. 3.1).

Table 3.3: Primer pairs used to amplify target candidate genes for fatty acid biosynthesis

Gene	Primer	Primer sequence	Location in gene	Product size (bp)
Acetyl-CoA Carboxylase	RfaI1-F RfaI1-R	TGTCAACTCCCACTCCCATT AATGGCCAATGCACCTCTAC	Exon	500
Acetyl-CoA Carboxylase	RfaI2-F RfaI2-R	GCTCAGCAAGTGAAATGCAG ATGGGTATCTTGCAGCCATT	5'UTR	700
Acetyl-CoA Carboxylase	RfaI3-F RfaI3-R	TCAATTAGCAGGGAGCCATT GGCATGCATTTGGCTTAGAT	Intron+ Exon	850
Malonyl transferase	RfaI4-F RfaI4-R	CTCCCCAATTTCTCACCAAA TGTCCACACTGACAGGCTTC	Intron+ Exon	500
Malonyl transferase	RfaI5-F RfaI5-R	ATTCAACAATGGCATCGACA TCATTCAGCTGTGCCTTCAC	Intron+ Exon	740
KAS I	RfaI6-F RfaI6-R	CTTCATCAATCGCCAATTCC AGATACTTCGCATGCCCAAA	Intron+ Exon	930
KAS I	RfaI7-F RfaI7-R	CAGACAACCGATTAACCCACA GCTCACTGGTATGCTGCTGA	Intron+ Exon	500
KAS I	RfaI8-F RfaI8-R	CTGTCTTCCACGCTTTCACA GCCTTGGGTGCATAAACATT	Intron+ Exon	630
KAS I	RfaI9-F RfaI9-R	GGTTCTCCAATTTGCCTGAA GGACTTCATACATGGCACTCAA	Intron+ Exon	450
KAS I	RfaI10-F RfaI10-R	GTCTGGGCTCCCTGTGAATA GCACCGCACTTAGCTTCTTC	Intron+ Exon	630
KAS II	RfaI11-F RfaI11-R	GGGAAGTAGAGGCCAGGAAG GAAGCTCCGAGCGAGATTTA	Intron+ Exon	650
KAS II	RfaI12-F RfaI12-R	GCGTGATCAGTCTTTCACCA CGTTTATTGTTTCCGGGTTG	Intron+Exon	780
KAS III	RfaI13-F RfaI13-R	GGTTATGTTGACGCTGCTGA AGCACCTTTCATTAGGATGGT	Intron+ Exon	640
KAS III	RfaI14-F RfaI14-R	ATCATGCAGCCATGTTTCAG CATCAAATGTGGCAGAGGTG	Intron+ Exon	500
Enoyl reductase	RfaI15-F RfaI15-R	TGAAGCTGATCACTATCAAGTTTGC CATCAGGAGGGATAATGTAATTTGC	Intron+ Exon	790
Enoyl reductase	RfaI16-F	CCCATAGTCAGCCACGAAAT	Intron+	820

	RfaI16-R	TGACAGCATCAAACACTGGA	Exon	
Enoyl reductase	RfaI17-F RfaI17-R	TCCAGTGTTTGATGCTGTCA TGGAATCAAGCTCGTCAAAA	Intron+ Exon	600
Palmitoyl ACP	RfaI18-F RfaI18-R	AATGAAGGTTGCCCTTTCCT TCGAGTCTTTTCTGCGGTTT	Intron+ Exon	500
Palmitoyl ACP Thioestrases	RfaI19-F RfaI19-R	CATGCAATTCATCCCATCAC TTAGGAAAGCAGGACCCAGA	Intron+ Exon	630
Stearoyl ACP desaturase	RfaI20-F RfaI20-R	TGGCAGAGGAACAAGGATTT TGCCAGTAAGCCCCATACAT	Intron+E xon	650
Stearoyl ACP Thioestrases	RfaI21-F RfaI21-R	GGACAAATGGATCTGCCACT TACCATGCCACTTCCAACAA	Intron+ Exon	700
Oleoyl- desaturase	RfaI22-F RfaI22-R	AGAGGGAGAAAGACGCACAA AAGGCAGGCAGTTAAGCAAA	Intron+ Exon	500
Oleoyl- desaturase	RfaI23-F RfaI23-R	TAGCGGCCTAAATTGTGCTT AAGGGAAGGCGTTTCTCAGT	Intron+ Exon	550
Stearoyl ACP	RfaU1-F RfaU1-R	GACAAGGATTGATTGTTGTCATCTG CATTACTCTCCTCTTCTCGAAGCAC	5'UTR	800
Acetyl-CoA Carboxylase	RfaU2-F RfaU1-R	TTCCCCACTTCGTCTGTTTC TGTTGGTTGACATTGCAGGT	5'UTR	490
Malonyl transferase	RfaU3-F RfaU3-R	CACAGGTGCAATGGTAGCAA CCTGCCCTTTAGGGTCTTGT	3'UTR	500
β-ketoacyl-ACP synthasc	RfaU4-F RfaU4-R	AGTGTGCTGCAAAATGCAAG GGCTGGCATTGTGATTTTCT	5'UTR	680
β-ketoacyl-ACP synthase	RfaU5-F RfaU5-R	GGCTAGAAAATTTAGGGGTCAA GTCGCAGACATGGAATTGAA	5'UTR	700
Oleoyl Co-A desaturase	RfaU6-F RfaU6-R	CGTAGGCAAGTTCCTTCCAA AGAAAGGCAAAAGGCAACAA	5'UTR	670
β-ketoacyl-ACP reductase	RfaU7-F RfaU7-R	TGCAACTTGTGGAGACCAAG TCAACCAGTCGCATTTCTGA	3'UTR	480
β-ketoacyl-ACP reductase	RfaU8-F RfaU8-R	GGTTCTCCAATTTGCCTGAA CGGCACTTGCTAACTCCTCT	Exon	350
β-ketoacyl-ACP reductase	RfaU9-F RfaU9-R	GAAGTTCCTGGTGGGACAAA GCCTCGTCAAAATCAAAAGG	Exon	490
KASIII	RfaU10-F RfaU10-R	CACCTCTGCCACATTTGATG CAGATATTCACGGCTGCTCA	5'UTR	680
KASIII	RfaU11-F RfaU11-R	TGCGTGCATTAAGGACTTTG GAAGCTCCGAGCGAGATTTA	3'UTR	490
Enoyl reductase	RfaU12-F RfaU12-R	CATGCCAATGTGCTAATGCT GGAACTGTTTTGGCTTTTGC	5'UTR	790

Enoyl reductase	RfaU13-F RfaU13-R	CACCTGTTTCATCCCGAACT GGATTTCTGGCATTTGTTGG	5'UTR + Exon	800
Oleoyl ACP desaturase	RfaU14-F RfaU14-R	GAGGAAAACAATCGCAGCTC CCATCAAATGCCTAAGCACA	3'UTR	500
Palmitoyl ACP	RfaU15-F RfaU15-R	GGACGATTGCCACTTGAGTA TCACAGCGTAGGTTCCACAG	5'UTR	630

3.5) Plant Material and genotyping

J. curcas genotypes used in the study were obtained from the Himalayan Forest Research Institute, Shimla and the National Bureau of Plant Genetic Resources (NBPGR), New Delhi, India (Table 3.4, 3.5). A network program implemented by the Department of Biotechnology, Govt. of India had identified a collection of Jatropha genotypes with high versus low oil content in their seeds. Total genomic DNA was isolated from leaves and seeds of all genotypes according to a modified CTAB-based procedure (Murray and Thompsan, 1980). The quality of DNA was checked on 1% agarose gels. PCR reactions for all the primer pairs were performed in 25 µl reaction volume. The PCR was performed on 30 ng of genomic DNA with primer pairs, $Mg^{2+,}$ dNTPs and *Taq* DNA polymerase. Amplification programs included 94°C for 5 min, 30 cycles of 94°C for 45 sec, annealing temperature (52–57°C) for 45 sec, 72°C for 2 min and a final extension of 7 min at 72°C. Ten µl of each PCR product was mixed with 2 µl of 10X gel loading dye (0.2% bromophenol blue, 0.2% xylene cyanol dye and 30% glycerol in a TA buffer) and electrophoresed in a 4% agarose gel prepared in 0.5X Tris borate-EDTA (TBE) buffer (0.05 M Tris, 0.05 M boric acid, 1mM, EDTA pH 8.0). The gel was run at a constant voltage of 80 volts for 1.5 to 2 h and stained with ethidium bromide and analyzed using the gel documentation system AlphaImager EP (Alpha Innotech Corp., USA).

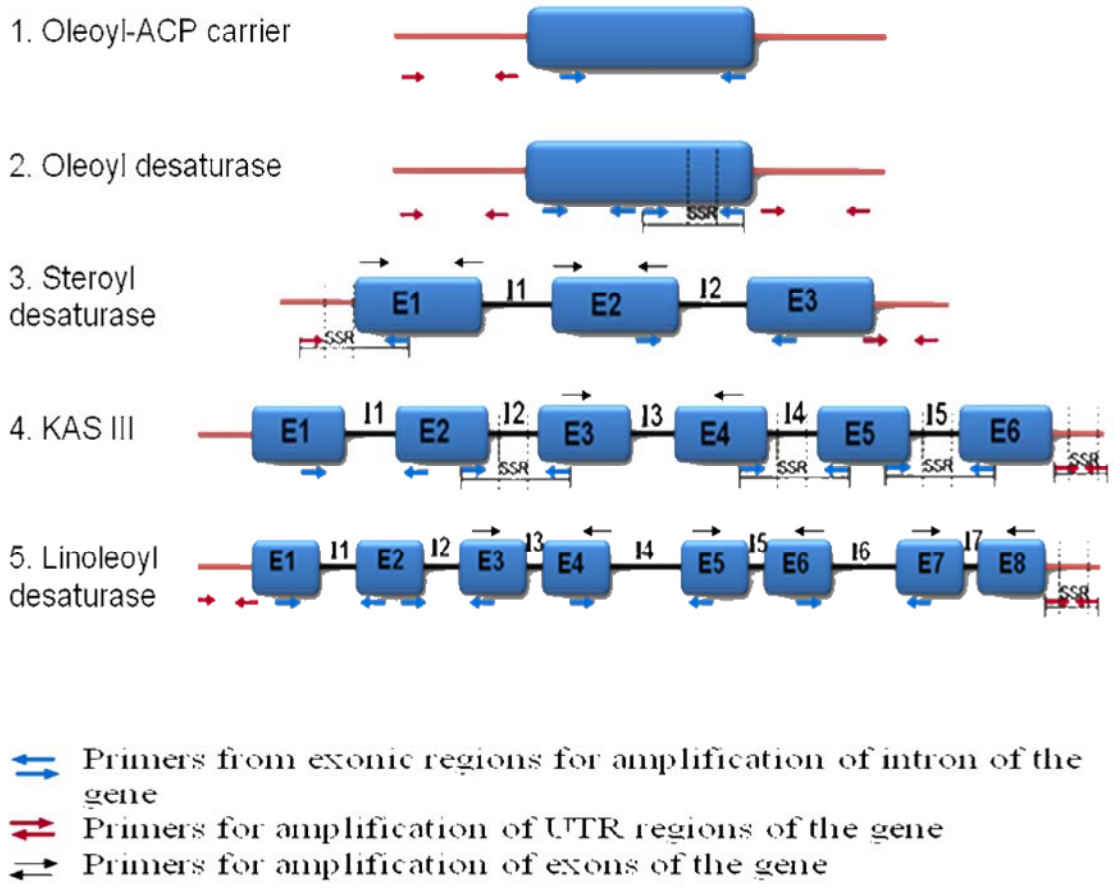

Fig. 3.1: Location of primer pairs used for amplifying different regions of fatty acid biosynthesis candidate genes

Table 3.4: *J. curcas* genotypes used in SSRs diversity analysis

S.No.	Genotypes
1	Urli-Kanchan
2	KcJK5
3	KcPBS19
4	KcPBS15
5	KcJK2
6	PBS SKS25
7	KcPBS16
8	KcPBS22
9	KcPBS26
10	KcJK8

11	KcJK1
12	KcJK6
13	KcJK9
14	KcJK3
15	KcPBS4
16	KcJK10
17	PBS SKS21
18	PBS SKS20
19	Chaterpatti
20	PBS SKS22
21	KcJK11
22	KcPBS2
23	FRI-EL-C5
24	FRI-EL-C2
25	PBS SKS24
26	PBS SKS26
27	KcJk7
28	CSNCRI-Guj-C3
29	SKN-Big
30	CSNCRI-Guj-C2
31	KcPBS17
32	Hissar local
33	SKNJ2
34	KcPBS13
35	PBS SKS27
36	Hansraj
37	PBS SKS19
38	CSNCRI-Guj-C4
39	CSNCRI-Guj-C1
40	KcPBS3
41	KcJk4
42	FRI-EL-C4
43	PBS SKS23
44	CSNCRI-Guj-C5
45	PBS SKS18
46	FRI-EL-C3
47	FRI-EL-C1
48	PBS SKS28
49	KcPBS10

Table 3.5: *J. curcas* genotypes used in the association analysis of candidate gene markers with high oil content

S.No.	Accession ID	Source	Known seed oil content
H1	IC550416	NBPGR, New Delhi	>35%
H2	IC550419	NBPGR, New Delhi	>35%
H3	IC550420	NBPGR, New Delhi	>35%
H4	IC550421	NBPGR, New Delhi	>35%
H5	IC550424	NBPGR, New Delhi	>35%
H6	IC550425	NBPGR, New Delhi	>35%
H7	IC550426	NBPGR, New Delhi	>35%
H8	IC550427	NBPGR, New Delhi	>35%
H9	IC550428	NBPGR, New Delhi	>35%
H10	IC550429	NBPGR, New Delhi	>35%
H11	IC550430	NBPGR, New Delhi	>35%
H12	IC550444	NBPGR, New Delhi	>35%
H13	IC550446	NBPGR, New Delhi	>35%
H14	IC550448	NBPGR, New Delhi	>35%
H15	IC550449	NBPGR, New Delhi	>35%
H16	IC550457	NBPGR, New Delhi	>35%
H17	IC550461	NBPGR, New Delhi	>35%
H18	IC566089	NBPGR, New Delhi	>35%
H19	IC566090	NBPGR, New Delhi	>35%
H20	IC550447	HFRI, Shimla	>35%
H21	IC565027	HFRI, Shimla	>35%
H22	HAP-IC550448	HFRI, Shimla	>35%
H23	IC553591	HFRI, Shimla	>35%
H24	IC558210	HFRI, Shimla	>35%
H25	IC558217	HFRI, Shimla	>35%
H26	IC565028	HFRI, Shimla	>35%
H27	IC565512	HFRI, Shimla	>35%
H28	IC565513	HFRI, Shimla	>35%
H29	IC555380	HFRI, Shimla	>35%
H30	IC550431	HFRI, Shimla	>35%
H31	IC555381	HFRI, Shimla	>35%
H32	IC468910	HFRI, Shimla	>35%
H33	IC565029	HFRI, Shimla	>35%
H34	IC555370	HFRI, Shimla	>35%
H35	IC555382	HFRI, Shimla	>35%
H36	IC560688	HFRI, Shimla	>35%
H37	IC555388	HFRI, Shimla	>35%
H38	IC555383	HFRI, Shimla	>35%
H39	Panjgain (T3)	HFRI, Shimla	>35%
H40	Panjgain (T4)	HFRI, Shimla	>35%

L1	IC545445	NBPGR, New Delhi	<30%
L2	IC545457	NBPGR, New Delhi	<30%
L3	IC545458	NBPGR, New Delhi	<30%
L4	IC545460	NBPGR, New Delhi	<30%
L5	IC545461	NBPGR, New Delhi	<30%
L6	IC545463	NBPGR, New Delhi	<30%
L7	IC555451	NBPGR, New Delhi	<30%
L8	IC555452	NBPGR, New Delhi	<30%
L9	IC555455	NBPGR, New Delhi	<30%
L10	IC559360	NBPGR, New Delhi	<30%
L11	IC559382	NBPGR, New Delhi	<30%
L12	IC560617	NBPGR, New Delhi	<30%
L13	IC560621	NBPGR, New Delhi	<30%
L14	IC560622	NBPGR, New Delhi	<30%

3.6) Statistical analysis for detection of linkage disequilibrium

PowerMarker version 3.25 (Liu and Muse, 2005) and Gen- AlEx version 6.2 (Peakall and Smouse, 2006) were used to measure the variability at each locus. Deviations from Hardy–Weinberg (HW) and tests for linkage disequilibrium were evaluated using Fisher's exact tests and sequential Bonferroni corrections. Principal components analysis (PCA) was conducted using 11 markers. GenAlex 6.2 (Peakall and Smouse 2006) software was used to generate a PC-matrix. FSTAT (Goudet, 1995) was used to estimate pairwise Fst values among the resulting structural populations. Analysis of molecular variances (AMOVA) was conducted using GenAlex 6.2 (Peakall and Smouse, 2006). Mixed model approach was used for conducting structured associations (Yu *et al.,* 2006, Zhao *et al.,* 2007). The model was $Y = X\alpha + Q\beta + Zu + \varepsilon$, with Y a vector of phenotypes, X a vector of single locus genotypes that were considered as fixed effects, 'α' a vector of fixed effects of the n – 1 genotype classes, Q a matrix of the $K - 1$ subpopulation ancestry estimates for each individual from STRUCTURE, 'β' a vector of the fixed effects for each of the $K - 1$ subpopulations, Z an identity matrix, 'u' a matrix of random deviates due to genome wide relatedness (as inferred from K), and ε a vector of residual errors.

For the linked markers, one-way AMOVA was conducted to test the association of markers with differences in oil content across the genotypes.

3.7) Statistical analysis for molecular diversity

PowerMarker version 3.25 (Liu and Muse, 2005) and Gen- AlEx version 6.1 (Peakall and Smouse, 2006) were used to measure the variability at each locus: the observed heterozygosity (HO), the expected heterozygosity (HE), the polymorphism information content (PIC), the deviation from Hardy–Weinberg equilibrium (HW). Deviations from Hardy–Weinberg (HW) and tests for linkage disequilibrium were evaluated using Fisher's exact tests and sequential Bonferroni corrections. The polymorphism information content (PIC) of each microsatellite locus was determined as described by Weir (1999): $PIC = 1-\Sigma Pi^2$; where Pi is the frequency of the ith allele in the genotypes examined. Pair-wise similarity matrices were generated by Jaccard's coefficient of similarity (Jaccard, 1908) by using the SIMQUAL format of NTSYS-pc (Rohlf, 2005). The presence or absence of amplicons in the genotypes was scored as 1 or 0, respectively. A dendrogram was constructed by using the unweighted pair group method with arithmetic average (UPGMA) with the SAHN module of NTSYS-pc to show a phenetic representation of genetic relationships as revealed by the similarity coefficient (Sneath and Sokal, 1973).

3.8) Identification of NBS-LRR genes in the castor bean genome

NBS proteins of plant genomes were downloaded from NCBI (http://www.ncbi.nlm.nih.gov/protein/). These protein sequences were used as query against castor bean database in TIGR (http://blast.jcvi.org/erblast/index.cgi?project=rca1). Protein function domains of the identified NBS-encoding genes in castor bean were predicted using

Hidden Markov Model (HMM) searches using Pfam (http://pfam.jouy.inra.fr/ hmmsearch.shtml). The identified NBS-LRR proteins were categorized into CNLs (Coiled coil NBS- LRR) and TNLs (TIR NBS- LRR) (http://toolkit.tuebingen.mpg.de/ pcoils) by using PCOILS. The expression status of candidate genes belonging to different families was searched in ESTdb of NCBI (http://www.ncbi.nlm.nih.gov/BLAST/) and TIGR (http://compbio.dfci.harvard.edu/tgi/cgi-in/tgi/Blast/ index.cgi).

3.9) Detection of polymorphisms at NBS loci between *J. curcas* versus *J. integerrima*

NBS-AFLP profiling of *J. curcas* versus *J. integerrima* (disease resistant) was performed as per the methodology described by Van der Linden *et al*. (2004). In brief, genomic DNA was digested with *Mse* I and *EcoR* I for 4 h using 400 ng DNA per sample. An adapter was ligated to the restriction fragments. Adapter primers with restriction enzyme sites were:

Adapter primer/ *EcoR* I: 5′- GAC TGC GTA CCA ATT C -3′

Adapter primer/ *Mse* I: 5′-ACTCGATTCTCAACCCGAAAG- 3′

Four different primer pairs were then used for a two-step PCR procedure. These primers were designed from part of the conserved P-loop and Kinase-2 motif, respectively, belonging to the NBS encoding region of NBS-LRR genes in castor bean and other plant species (Van der Linden *et al.*, 2004). The first PCR was linear, and only one primer was used. The second PCR was exponential and was performed with both the degenerate and adapter primers. The PCR products were separated in 6% (w/v) polyacrylamide gels for 3 h at 65 W, which were silver stained (Xiao *et al*., 2008) then dried at room temperature and photographed.

3.10) Cloning of resistance genes analogues (RGAs)

The genomic DNA was isolated from young leaf samples of *J. integerrima* by using standard methods (Murray and Thompsan, 1980). For amplification of RGAs, the PCR primers were designed from the P-loop and GLPLQ Motifs of the NBS regions of NBS-LRR genes identified in castor bean (Table 3.6). These primers were used in different combinations for PCR amplification of RGAs from *J. integerrima*. The amplified products were gel eluted, cloned in pGEM-T vector and transferred to the DH5α chemical-competent cells of *E. coli* for multiplication. The transformants were screened based on blue/white colonies observed on LA plates containing IPTG, X-gal and ampicillin. The identity of recombinants was confirmed by sequencing. The sequence data were subjected to homology search by using the default settings of BLASTn with nr database of GenBank (http://www.ncbi.nlm.nih.gov) to identify resistance gene analogues as well as other homologous sequences in the database.

Table 3.6: Primer pairs used for amplification of resistance gene analogues in *J. integerrima*

Primer name	Sequence	Product size (bp)
Lukinase2-[F]	CTTCTAGTTCTIGATGAYGI	500
Lukinase2-[R]	ACATCGTCIAGAATAGIAG	500
LuGLPL-[F]	GGACTTCCACTAGCICT	450
LuGLPL-[R]	GAGIGCIAGIGGIARICC	450
SP-PLOOP[F]2	GGDATGVSVGGHDYVGGKAARAC	500
SP-GLPL[R]2	GCMRCCARAGGMARYCC	500
SP-PLOOP[F]3	GGNGGNRTHGGNAARACCHAC	350
SP-GLPL[R]3	AGNGCHAGNGGYAANCC	350
Ha-1[F]	GGIGGIGTIGGIAAIACIAC	500
Ha-1[R]	YCTAGTTGTRAYDATDAYYYTRC	500
P1a- PLOOP[F]	GGIATGCCIGGIIIIGGIAARACIAC	450
P3a-GLPL[R]	AIITYIRIIRYIAGIGGYAAICC	450
RNBS-D[R]	GGRAAIARISHRCARTAIVIRAARC	500
PLOOP-4[F]	CCGGGITCAGGIAARACWAC	500
NBS-2[R]	GTWGTYTTICCYRAICCISSCAT	500
NBS-6[F]	YYTKRTHGTMITKGATGATATITGG	500

GLPL-4[R]	CCCGAAGGAAACCRISRACWARA	450
NBS-9[F]	TGTGGAGGRTTACCTCTAGC	450
GLPL-5[R]	CCKGARGGIRATCGKRRITTTCA	500
KIN-5[F]	CTTGTMATITTGGATGATGTWTGG	500

3.11) Statistical analysis for detection of polymorphisms in NBS-LRR genes

PowerMarker version 3.25 (Liu and Muse, 2005) and Gen- AlEx version 6.1 (Peakall and Smouse, 2006) were used to measure the variability at each locus: the observed heterozygosity (HO), the expected heterozygosity (HE), the polymorphism information content (PIC).The polymorphism information content (PIC) was determined as described by Weir (1999). $PIC = 1-\Sigma Pi^2$; where Pi is the frequency of the ith allele in the genotypes examined. The presence or absence of amplicons in the genotypes was scored as 1 or 0, respectively. Marker Index (MI) calculated as (Milbourne *et al*., 1997): MI=PIC X (Number of polymorphic loci/number of primers).

Chapter-IV

RESULTS

The results obtained in the present investigation are explained under the following heads:

4.9) Identification of SSRs in the castor bean genome

4.10) Transferability and polymorphism survey of castor bean SSRs in *Jatropha curcas*

4.11) Comparative genomics of fatty acid biosynthesis genes in major oil seed plant species

4.12) Identification of candidate gene markers in *J. curcas*

4.13) Association analysis of candidate gene markers with high oil content in Jatropha genotypes

4.14) Identification of NBS-LRR genes in castor bean genome

4.15) NBS profiling between *J. curcas* and *J. integerrima*

4.16) Cloning and analysis of resistance gene analogs (RGAs) in *J. integrrima*

4.1) Identification of SSRs in the castor bean genome

Genomic sequence of castor bean consisting of 25,828 contigs (4X coverage) was surveyed for simple sequence repeats (SSRs). A total of 5, 80,986 SSRs were identified in the castor bean genome with a frequency of 1 per 680 bp. The location of SSRs in exons, introns, 5' and 3' UTRs and non-genic regions of castor bean genome was inferred by annotation of 25,828 sequence contigs for open reading frames, including the 5'UTRs and 3'UTRs. A total of 31,221 genes were predicted in 25,828 sequence contigs of castor bean.

Whole genome SSRs analysis in castor bean genome revealed that 51% SSRs were dinucleotide repeats, 29% trinucleotide, 12% tetranucleotide and 8% pentanucleotide repeats (Table 4.1). Dinucleotide repeats were more frequent in the non-genic regions (genome), introns, 5'UTRs and 3'UTRs whereas trinucleotide repeats were more common in the exons. The tetra- and penta- nucleotide repeats were randomly distributed. Among dinucleotide repeats, SSRs with (AT)n repeat motif were common (43%) with a repeat motif ranging from 7 to 48. The frequency of repeat motifs differed in different genome regions such as (AT)n and (AG)n in 5'UTRs, (TA)n and (AATA)n in 3'UTRs, (AT)n and (TC)n in introns, and (AT)n and (TA)n in non-genic regions. Trinucleotide repeats are runs of particular amino acids. An analysis of the frequencies of trinucleotide repeats out of total SSRs indicated the predominance of TCT/GAA/CGC/TTC. The most frequent amino acid runs identified in the castor bean SSRs were serine (TCT)n (16.5%), glutamate (GAA)n (13.6%), arginine (CGC)n (12.3%) and phenylalanine (TTC)n (9.7%). Abundance of SSRs in different regions of the castor bean genome showed that 73% were located in the genic regions and 27% in the non-genic regions. Comparison of SSR densities in the genic regions showed that SSRs were more frequent in 5'UTRs (26%) and introns (25%) followed by 3'UTRs (16%) and exons (6%).

Table 4.1: Genome/ gene location of SSRs in castor bean

SSR motif	Contig no.	SSRs start position in the contig	Genome/ gene location
(AAT)12	50176	5329	Non genic
(AT)31	51537	9	Non genic
(AT)20	50870	951	Non genic
(TA)14	50872	1962	Non genic
(AATA)7	49564	1705	Non genic
(AAT)20	50926	2249	Non genic
(TA)20	51068	677	Non genic
(TTTA)6	49861	324	Non genic

(AAAAT)5	52127	1089	Non genic
(AT)12	52177	395	Non genic
(GA)18	52298	7123	Non genic
(AT)12	52302	671	Non genic
(TA)34	52814	141	Non genic
(TATT)6	52814	207	Non genic
(TA)22	51641	1580	Non genic
(AT)14	51768	900	Non genic
(TA)13	51775	848	Non genic
(AATA)6	53087	436	Non genic
(TA)35	53808	850	Non genic
(GCCCG)5	53969	714	Non genic
(AT)13	53243	542	Non genic
(AT)12	37096	929	Non genic
(AT)48	53489	185	Non genic
(AT)46	53508	1050	Non genic
(AT)15	53508	1279	Non genic
(AC)17	53508	1479	Non genic
(TA)27	54566	1243	Non genic
(AT)14	36612	693	Non genic
(AT)25	36709	221	Non genic
(TG)18	36709	207	Non genic
(TTTA)6	36958	357	Non genic
(TA)14	54065	1354	Non genic
(TTTA)6	37986	638	Non genic
(ATTTA)5	38048	657	Non genic
(TTA)12	54298	799	Non genic
(TTTA)6	36958	357	Non genic
(TA)14	54065	1354	Non genic
(TTTA)6	37986	638	Non genic
(ATTTA)5	38048	657	Non genic
(TTA)12	54298	799	Non genic
(AATA)6	37474	272	Non genic
(TA)28	55402	1270	Non genic
(AT)37	37566	159	Non genic
(TTTA)6	37744	863	Non genic
(TC)24	37844	934	Non genic
(AATA)6	39574	4420	Non genic
(TC)23	55062	320	Non genic
(TTA)8	56487	2644	Non genic
(TA)21	57055	503	Non genic
(AT)14	57060	1002	Non genic
(TTTA)6	57125	42	Non genic
(CT)12	57169	849	Non genic
(AATA)4	56610	195	Non genic

(AT)21	56672	251	Non genic
(AT)36	56672	369	Non genic
(TATTT)6	57946	54	Non genic
(TA)31	58112	772	Non genic
(TTTA)6	57492	786	Non genic
(CT)13	57604	41	Non genic
(TTTTA)6	58818	475	Non genic
(CT)17	58307	1052	Non genic
(AATA)6	58326	797	Non genic
(TCTT)6	59502	938	Non genic
(AGA)10	59651	3636	Non genic
(TTA)23	59658	26560	Non genic
(TA)46	59676	565	Non genic
(AAT)9	59800	530	Non genic
(GAA)10	59009	748	Non genic
(AATA)6	59034	228	Non genic
(TA)26	59075	441	Non genic
(AT)15	59102	1099	Non genic
(AT)19	59140	713	Non genic
(TTA)8	59147	949	Non genic
(AATA)6	59289	506	Non genic
(AG)14	60404	1142	Non genic
(TTTA)6	60175	209	Non genic
(TA)16	36228	1	Non genic
(AT)12	38715	1322	Non genic
(AT)22	38902	967	Non genic
(AATA)6	38938	36	Non genic
(AAT)13	38982	63	Non genic
(ATAA)6	38998	436	Non genic
(TTC)9	38391	462	Non genic
(TCTT)6	39464	1256	Non genic
(TTTA)6	40037	2071	Non genic
(AT)20	40040	145	Non genic
(ATTTT)5	40049	92	Non genic
(AT)15	40215	584	Non genic
(AATA)7	40749	727	Non genic
(ATA)11	40779	4495	Non genic
(TAA)8	40779	5232	Non genic
(TA)26	40821	472	Non genic
(TTC)13	40845	3042	Non genic
(TAT)9	40925	2180	Non genic
(TA)14	41014	14	Non genic
(TTA)8	41132	6316	Non genic
(AATA)6	41133	390	Non genic
(AT)26	40553	593	Non genic

(AT)13	40677	1	Non genic
(TTTA)7	41623	944	Non genic
(TA)15	41152	69	Non genic
(AT)24	41423	258	Non genic
(AT)24	41423	790	Non genic
(AT)13	41436	3	Non genic
(TA)13	42517	727	Non genic
(AG)12	42608	755	Non genic
(AT)40	42740	38	Non genic
(CT)14	42760	662	Non genic
(AT)18	41991	299	Non genic
(TATTT)6	43234	31	Non genic
(AATA)6	43306	365	Non genic
(TAT)8	43357	564	Non genic
(TTTA)6	43025	38	Non genic
(TA)22	43045	129	Non genic
(ATT)9	27401	45589	Non genic
(TA)24	27401	47523	Non genic
(TC)33	27401	58617	Non genic
(AT)15	27401	59267	Non genic
(AT)27	27404	44594	Non genic
(TTATT)5	27415	16918	Non genic
(TA)13	27448	10470	Non genic
(TA)16	27451	10092	Non genic
(AAT)8	27451	11926	Non genic
(TC)13	27451	158584	Non genic
(TTCTC)5	27455	2087	Non genic
(TC)14	27455	2148	Non genic
(TA)12	27461	7300	Non genic
(TA)37	27461	27159	Non genic
(TTTA)6	31225	740	Non genic
(TA)12	31266	868	Non genic
(TTTA)6	44098	309	Non genic
(AT)16	44417	1108	Non genic
(TA)16	43858	5	Non genic
(TTC)9	43970	1029	Non genic
(TTTA)6	43980	968	Non genic
(ATA)13	45107	583	Non genic
(TA)38	45702	693	Non genic
(AT)36	45728	725	Non genic
(TTTA)6	45737	823	Non genic
(TA)22	45953	113	Non genic
(AT)18	45953	1125	Non genic
(AAATA)5	46558	730	Non genic
(TTTAT)8	46597	98	Non genic

(TA)14	46855	880	Non genic
(AATA)6	46862	530	Non genic
(AATA)6	47535	448	Non genic
(AATAA)5	46907	756	Non genic
(AT)14	46923	845	Non genic
(AATT)7	47283	540	Non genic
(AT)39	27373	27087	Non genic
(TTTA)6	27373	57009	Non genic
(AAAAG)5	27373	99237	Non genic
(ATT)12	27383	1977	Non genic
(TA)19	27389	7510	Non genic
(AT)30	27394	29340	Non genic
(TAA)14	27394	36609	Non genic
(GT)21	27394	44318	Non genic
(TA)15	27394	49202	Non genic
(CT)33	27394	83102	Non genic
(AATA)6	27394	153652	Non genic
(AG)14	31566	553	Non genic
(AT)21	31721	648	Non genic
(CT)13	31959	760	Non genic
(TA)13	32355	310	Non genic
(TAT)9	32385	1748	Non genic
(TA)14	32428	397	Non genic
(AT)20	32500	825	Non genic
(TTATT)5	32520	117	Non genic
(AT)12	33260	871	Non genic
(TA)13	33293	1361	Non genic
(AATA)6	33415	1046	Non genic
(AATA)6	33663	446	Non genic
(AT)12	33760	1129	Non genic
(TTTCT)5	34115	92	Non genic
(TG)26	34425	865	Non genic
(AATA)6	34465	256	Non genic
(TTTA)6	34511	320	Non genic
(TA)14	34711	939	Non genic
(TCAC)8	34922	1905	Non genic
(CT)14	35021	888	Non genic
(AG)14	35265	1128	Non genic
(TA)30	28806	36831	Non genic
(ATT)10	28806	20419	Non genic
(TA)20	27865	3544	Non genic
(CT)14	27866	108377	Non genic
(TA)25	27866	108404	Non genic
(TA)33	28829	35305	Non genic
(TA)19	28830	54215	Non genic

(AT)14	28830	132496	Non genic
(AT)24	28830	3542	Non genic
(TA)15	28833	2535	Non genic
(AG)18	28833	23667	Non genic
(TA)18	28838	133486	Non genic
(AATA)9	28838	74723	Non genic
(AT)26	27882	4407	Non genic
(CT)13	28842	27218	Non genic
(AT)13	28842	398137	Non genic
(TA)30	27889	48242	Non genic
(TC)16	27893	2626	Non genic
(AT)26	27893	40579	Non genic
(TTC)11	27893	3077	Non genic
(TTA)14	27893	51782	Non genic
(TTA)15	27893	94353	Non genic
(ATATA)5	27893	36069	Non genic
(AT)17	27894	59570	Non genic
(TA)37	27894	405540	Non genic
(ATT)9	27894	22679	Non genic
(GAA)9	27894	432886	Non genic
(TTA)23	27894	485951	Non genic
(TA)17	28859	158318	Non genic
(TAT)9	27897	965	Non genic
(AG)15	28863	76990	Non genic
(TTA)26	28863	37196	Non genic
(ATAA)8	28863	78595	Non genic
(ATT)13	28863	78625	Non genic
(TA)24	27904	96907	Non genic
(TA)28	28869	916	Non genic
(AT)21	27905	41590	Non genic
(AT)31	28879	15964	Non genic
(TA)15	27914	47808	Non genic
(ATT)9	28883	9385	Non genic
(AT)13	28885	45907	Non genic
(ATA)10	28885	45419	Non genic
(TA)15	27917	11829	Non genic
(ATA)11	27917	12743	Non genic
(CT)19	27934	93001	Non genic
(CT)27	27934	100470	Non genic
(AT)17	27894	59570	Non genic
(AG)14	27895	7971	Non genic
(TA)26	27895	11899	Non genic
(AT)26	28909	10569	Non genic
(TA)13	28909	22900	Non genic
(AG)13	28909	22925	Non genic

(TTACA)5	28909	50527	Non genic
(AT)30	27940	138802	Non genic
(AT)23	27945	5568	Non genic
(TA)22	28923	6948	Non genic
(AG)13	27955	124814	Non genic
(TA)18	27956	52648	Non genic
(TA)25	27957	4	Non genic
(TTAA)7	27961	90165	Non genic
(AT)14	27964	38529	Non genic
(AT)16	27964	197064	Non genic
(AT)17	27965	13214	Non genic
(GA)17	27965	26071	Non genic
(TC)15	27965	79202	Non genic
(TA)23	27968	119108	Non genic
(AT)13	27973	9720	Non genic
(AG)20	27973	9746	Non genic
(TA)21	27973	13164	Non genic
(TA)15	27985	74756	Non genic
(TA)14	27985	155377	Non genic
(TA)20	27985	353935	Non genic
(CT)13	28962	104638	Non genic
(TTC)9	28962	60322	Non genic
(ATA)11	28962	99605	Non genic
(AG)16	28966	134858	Non genic
(ATT)9	28966	31022	Non genic
(AG)19	29003	45569	Non genic
(TA)18	29003	46988	Non genic
(TA)17	29003	10590	Non genic
(AT)15	29003	46489	Non genic
(AT)28	29005	97606	Non genic
(CT)20	28035	185735	Non genic
(TA)14	28035	277521	Non genic
(TTA)21	28035	16081	Non genic
(TA)38	28037	47186	Non genic
(TA)17	28040	29130	Non genic
(AT)24	28049	45014	Non genic
(ATA)13	28049	106094	Non genic
(AT)31	28049	105322	Non genic
(TA)19	28052	40351	Non genic
(TA)15	29037	56047	Non genic
(TA)18	29042	8598	Non genic
(AG)17	29044	15072	Non genic
(TA)18	29044	16943	Non genic
(AT)18	29044	34981	Non genic
(AT)14	29044	38686	Non genic

(TTA)11	28067	31188	Non genic
(GA)14	29048	7018	Non genic
(TA)13	28076	113257	Non genic
(TA)14	28076	167819	Non genic
(TAT)15	28094	26200	Non genic
(TAT)9	28094	33915	Non genic
(AT)43	29075	12	Non genic
(CT)16	28098	1959	Non genic
(GA)16	29080	30720	Non genic
(TAA)12	29080	39997	Non genic
(AG)16	28102	13081	Non genic
(TA)13	28102	44268	Non genic
(AT)19	28102	72554	Non genic
(AG)13	28108	20216	Non genic
(AT)35	28108	115640	Non genic
(TA)24	28108	116471	Non genic
(TAA)9	28108	71490	Non genic
(AT)13	29092	241745	Non genic
(TCT)14	28128	5892	Non genic
(TA)29	28129	41351	Non genic
(ATA)15	28134	109864	Non genic
(GA)13	28138	99273	Non genic
(TC)19	28138	100219	Non genic
(CT)13	28140	57121	Non genic
(AT)20	29117	9386	Non genic
(AT)13	28152	100389	Non genic
(AG)18	28152	192296	Non genic
(AAT)9	28152	100099	Non genic
(TAT)11	28152	204748	Non genic
(AAT)16	28153	11503	Non genic
(ATT)11	29130	4064	Non genic
(AAG)11	28154	19858	Non genic
(TA)18	28161	26023	Non genic
(AT)20	28161	64404	Non genic
(TTA)10	28161	43950	Non genic
(CT)15	28162	21022	Non genic
(AT)18	28162	24755	Non genic
(TATT)7	28180	31131	Non genic
(TAT)11	28180	129303	Non genic
(AT)15	28184	3343	Non genic
(TAT)13	28196	37293	Non genic
(AAT)12	28196	38786	Non genic
(ATT)11	28197	5408	Non genic
(AT)14	28200	8334	Non genic
(AG)28	28200	9088	Non genic

(TTA)13	28200	36299	Non genic
(AT)13	29168	45015	Non genic
(AT)34	29168	138226	Non genic
(AT)20	29169	10576	Non genic
(AT)24	28205	3052	Non genic
(TA)13	28219	6199	Non genic
(AAT)9	29188	21181	Non genic
(TA)16	28238	1481	Non genic
(TA)14	29206	83996	Non genic
(AT)13	29212	535	Non genic
(AT)29	29212	79991	Non genic
(TAA)9	29212	79078	Non genic
(AT)20	29217	33594	Non genic
(AT)21	28254	11980	Non genic
(AT)28	29235	125	Non genic
(TTA)13	28281	93009	Non genic
(TTA)11	28282	3597	Non genic
(CT)14	28324	12354	Non genic
(CA)16	28324	12382	Non genic
(TA)13	28327	38009	Non genic
(AT)13	28327	43015	Non genic
(AT)16	28327	64227	Non genic
(AT)24	28327	77390	Non genic
(TC)17	28327	100535	Non genic
(TC)14	28327	102249	Non genic
(TA)42	28327	121213	Non genic
(TA)14	28327	138898	Non genic
(AGA)16	28327	145379	Non genic
(TA)26	28327	222464	Non genic
(TTA)10	28327	38802	Non genic
(AAT)10	28327	182074	Non genic
(TAT)12	28327	213201	Non genic
(TTA)14	28288	2066	Non genic
(TA)21	28333	158592	Non genic
(ATT)12	28333	264311	Non genic
(CT)14	28342	9049	Non genic
(AT)19	28349	160	Non genic
(TA)24	28308	20483	Non genic
(CT)14	36953	385	5'UTR
(TAT)9	28348	48728	5'UTR
(AT)29	28304	20967	5'UTR
(AT)19	28297	11825	5'UTR
(TA)33	28297	12794	5'UTR
(AT)13	28333	202594	5'UTR
(TTA)12	28333	247844	5'UTR

(AT)14	28333	136094	5'UTR
(AT)19	28329	23575	5'UTR
(CT)26	28324	192053	5'UTR
(ATA)14	28324	195531	5'UTR
(AGA)20	28324	134658	5'UTR
(AT)13	28281	107569	5'UTR
(TA)16	29235	7263	5'UTR
(CT)13	29235	10320	5'UTR
(AT)15	29235	14213	5'UTR
(AT)13	29235	54186	5'UTR
(TCT)10	29235	61551	5'UTR
(TA)14	28266	68639	5'UTR
(AT)20	28266	70154	5'UTR
(TA)23	29222	42024	5'UTR
(TA)23	28256	43163	5'UTR
(TAT)12	28256	33931	5'UTR
(AT)23	28256	2956	5'UTR
(AAG)9	29217	641	5'UTR
(AAT)9	29216	4590	5'UTR
(CTT)11	29216	5213	5'UTR
(TC)13	28246	57620	5'UTR
(CT)14	29200	35220	5'UTR
(CT)17	29200	53145	5'UTR
(ATT)11	28229	755	5'UTR
(TA)31	28226	117104	5'UTR
(TA)15	28226	117572	5'UTR
(AG)20	28226	117601	5'UTR
(AG)21	28226	194069	5'UTR
(ATT)14	28223	10452	5'UTR
(TCT)11	28223	59346	5'UTR
(AT)24	28223	4811	5'UTR
(AG)9	28212	110475	5'UTR
(AT)16	28211	8967	5'UTR
(AG)24	28211	241	5'UTR
(TAA)11	28206	65770	5'UTR
(ATT)13	28206	65802	5'UTR
(AT)16	28207	18238	5'UTR
(AT)18	28207	23095	5'UTR
(AG)16	28207	23131	5'UTR
(AT)16	28207	62550	5'UTR
(TTA)22	29171	76487	5'UTR
(TC)14	29169	15060	5'UTR
(GA)16	29171	162067	5'UTR
(AT)25	28202	64657	5'UTR
(TCT)10	29168	117756	5'UTR

(TA)32	28200	22704	5'UTR
(AT)14	28199	21190	5'UTR
(TAA)17	28196	37242	5'UTR
(CT)15	28193	28193	5'UTR
(TA)13	28192	17568	5'UTR
(AT)17	28180	21377	5'UTR
(TA)16	28162	37326	5'UTR
(CT)17	28162	38191	5'UTR
(TA)23	28161	8546	5'UTR
(TTA)11	28154	9766	5'UTR
(AT)14	28153	5593	5'UTR
(AT)13	28153	49993	5'UTR
(CTT)10	28152	2180	5'UTR
(AG)18	28152	24900	5'UTR
(TA)16	28152	233622	5'UTR
(TC)17	28152	128866	5'UTR
(AT)14	28152	26560	5'UTR
(GA)20	28147	29200	5'UTR
(AG)18	28140	30773	5'UTR
(CT)17	29108	1212	5'UTR
(ATA)11	29105	1027	5'UTR
(TCT)10	28134	9643	5'UTR
(AAT)9	28128	73855	5'UTR
(AT)13	28109	21739	5'UTR
(AT)27	28109	21986	5'UTR
(AT)14	28111	36307	5'UTR
(AT)15	29092	20135	5'UTR
(AT)20	29092	25740	5'UTR
(AT)27	29092	178970	5'UTR
(AT)17	29092	205082	5'UTR
(AT)18	29092	224895	5'UTR
(AT)27	29092	239564	5'UTR
(GAT)12	28094	102218	5'UTR
(TA)24	28093	21087	5'UTR
(AT)38	28093	21608	5'UTR
(AATTA)5	28093	45315	5'UTR
(AT)18	28077	3436	5'UTR
(AT)17	28077	4035	5'UTR
(TTA)9	28077	3217	5'UTR
(TTTGG)8	28077	31095	5'UTR
(TA)18	29070	26441	5'UTR
(TA)38	28076	16173	5'UTR
(TA)14	28076	19352	5'UTR
(TA)21	28076	24534	5'UTR
(TA)22	29044	82104	5'UTR

(TC)13	28049	99678	5'UTR
(TC)18	28040	32058	5'UTR
(TA)25	28040	3001	5'UTR
(TC)13	29010	94626	5'UTR
(TG)13	28035	236079	5'UTR
(AT)32	28035	189557	5'UTR
(TTA)9	29005	99907	5'UTR
(ATA)13	28962	25440	5'UTR
(ATA)22	27985	239798	5'UTR
(TTA)10	27985	253713	5'UTR
(TTA)10	27981	33277	5'UTR
(GA)17	27973	12822	5'UTR
(CT)13	27970	33455	5'UTR
(TA)30	27969	4966	5'UTR
(TA)21	27964	199210	5'UTR
(AAAG)7	27964	23902	5'UTR
(TTCT)7	27964	80036	5'UTR
(AT)16	28934	30612	5'UTR
(AAGA)7	27956	112793	5'UTR
(ATA)12	27956	128735	5'UTR
(AT)26	27956	129005	5'UTR
(TA)27	27956	143073	5'UTR
(TA)14	27955	137035	5'UTR
(TA)39	27945	29894	5'UTR
(AGA)10	27953	26267	5'UTR
(TA)14	27942	2430	5'UTR
(TA)14	27942	55321	5'UTR
(TA)22	27942	55369	5'UTR
(AT)19	27942	83555	5'UTR
(AT)13	27942	100775	5'UTR
(TA)16	27942	109278	5'UTR
(TAA)9	27940	45225	5'UTR
(ATT)10	27940	22	5'UTR
(GA)13	27940	130600	5'UTR
(AAT)9	28859	109631	5'UTR
(TA)16	28859	158318	5'UTR
(AT)32	28859	93158	5'UTR
(TC)15	28859	95999	5'UTR
(AT)22	27895	46741	5'UTR
(TTA)23	27894	485951	5'UTR
(GAA)9	27894	432886	5'UTR
(TA)37	27894	405540	5'UTR
(TA)15	27918	20788	5'UTR
(ATA)10	27916	80859	5'UTR
(AAT)13	27914	187761	5'UTR

(GA)18	28883	80045	5'UTR
(CT)21	28880	10231	5'UTR
(CT)15	27914	40474	5'UTR
(ATT)18	27893	34048	5'UTR
(TTA)9	28842	328842	5'UTR
(AT)28	28842	281901	5'UTR
(AG)14	33564	573	5'UTR
(AG)20	33589	817	5'UTR
(CTT)8	31781	195	5'UTR
(GA)16	27446	168793	5'UTR
(AGA)8	27394	74462	5'UTR
(TA)15	47083	4085	5'UTR
(ATTT)10	47136	998	5'UTR
(TA)22	46308	49	5'UTR
(TTA)9	46783	785	5'UTR
(AT)12	44525	243	5'UTR
(AT)17	43672	1	5'UTR
(AT)37	43672	70	5'UTR
(AT)21	43672	108	5'UTR
(AT)31	42884	28	5'UTR
(TTC)11	41177	1091	5'UTR
(TTTA)8	41262	6646	5'UTR
(TCT)12	41833	309	5'UTR
(TTTA)6	40723	251	5'UTR
(TA)29	38354	37	5'UTR
(TTA)13	36262	1466	5'UTR
(TAT)12	36262	5426	5'UTR
(CTTT)6	36347	1446	5'UTR
(TTTA)6	60482	716	5'UTR
(TA)7	51460	816	3'UTR
(TA)38	52072	3674	3'UTR
(CT)15	52072	3971	3'UTR
(AAAT)7	51866	290	3'UTR
(TTTA)6	53377	570	3'UTR
(AATA)6	38055	268	3'UTR
(AATAA)7	55461	3931	3'UTR
(TTTA)6	38103	713	3'UTR
(AG)15	54337	641	3'UTR
(TTTA)6	38163	937	3'UTR
(AATA)6	57415	431	3'UTR
(AT)22	57955	2035	3'UTR
(CT)23	57729	691	3'UTR
(AATA)6	59531	186	3'UTR
(TTAT)6	36229	400	3'UTR
(TA)20	36540	559	3'UTR

(AATA)6	38570	208	3'UTR
(TTTAT)5	40189	708	3'UTR
(TTAT)7	40517	659	3'UTR
(AAAAG)6	41398	3031	3'UTR
(AATA)6	43358	158	3'UTR
(TA)12	42824	793	3'UTR
(AATA)6	43083	258	3'UTR
(AGAAA)16	44109	861	3'UTR
(TTC)8	45930	1058	3'UTR
(TTTA)6	45933	119	3"UTR
(TTTA)6	46066	257	3'UTR
(AATA)6	46197	503	3'UTR
(AAT)11	47284	20840	3'UTR
(AATA)6	27373	54226	3'UTR
(AT)12	27394	161851	3'UTR
(TCTTT)7	31414	838	3'UTR
(TCTTT)9	31511	484	3'UTR
(TC)16	31700	947	3'UTR
(TTTA)6	31733	1407	3'UTR
(TTTA)6	32008	489	3'UTR
(TTTA)6	32063	505	3'UTR
(AT)14	32482	1460	3'UTR
(TTTAT)5	33249	1075	3'UTR
(TTTGC)22	34784	981	3'UTR
(TA)14	28838	111859	3'UTR
(TC)18	27894	126453	3'UTR
(AT)31	28859	168385	3'UTR
(CT)13	27904	13536	3'UTR
(AAT)18	27914	184882	3'UTR
(TC)18	27894	126453	3'UTR
(AT)20	28859	168385	3'UTR
(AAT)9	28859	118937	3'UTR
(CT)27	27934	100470	3'UTR
(AAT)12	27942	113277	3'UTR
(AT)22	27956	174045	3'UTR
(AT)19	27964	100528	3'UTR
(TA)19	27967	31046	3'UTR
(AT)17	28952	13490	3'UTR
(TC)13	28952	14191	3'UTR
(AT)17	28952	13490	3'UTR
(TA)18	27985	64925	3'UTR
(AT)26	28962	14504	3'UTR
(CT)19	28035	222611	3'UTR
(AT)37	29080	34638	3'UTR
(AGA)9	29080	34078	3'UTR

(AT)18	28132	2460	3'UTR
(TA)26	28152	300061	3'UTR
(TA)26	28196	69014	3'UTR
(TA)15	28219	17326	3'UTR
(CT)17	28226	236462	3'UTR
(AAT)13	29206	6638	3'UTR
(ATA)23	29206	6593	3'UTR
(TA)34	29214	1108	3'UTR
(AT)16	29216	80218	3'UTR
(TA)20	29227	20205	3'UTR
(TTA)10	28324	133144	3'UTR
(GA)13	28329	40343	3'UTR
(TTAT)6	57368	597	Intron
(AT)22	49727	728	Intron
(CTT)10	41833	9481	Intron
(TA)13	52298	3253	Intron
(TTTAT)8	52899	677	Intron
(TA)18	36845	972	Intron
(AT)25	37959	408	Intron
(TA)22	54308	632	Intron
(AT)35	57713	1709	Intron
(AT)13	39340	538	Intron
(TA)13	40322	4421	Intron
(TA)42	41844	134	Intron
(TATTT)6	41328	433	Intron
(AT)23	43365	340	Intron
(AAT)14	44057	504	Intron
(TA)22	44956	489	Intron
(TA)13	45715	390	Intron
(TTTA)6	46497	342	Intron
(AT)14	27375	96745	Intron
(AT)28	27383	34283	Intron
(TAT)8	27394	21099	Intron
(GA)28	27394	128536	Intron
(AT)16	27394	165617	Intron
(AG)12	27401	103806	Intron
(CGC)8	27403	800	Intron
(AT)12	27428	11493	Intron
(AATA)6	27448	41080	Intron
(AT)22	34943	727	Intron
(TTTA)6	35428	856	Intron
(TA)17	28828	14166	Intron
(TA)40	28830	91652	Intron
(TAT)11	28830	28952	Intron
(CT)13	28838	92004	Intron

(CT)18	28842	86309	Intron
(TA)17	28842	239455	Intron
(AT)22	28842	245584	Intron
(TG)22	28842	320869	Intron
(TC)14	28842	436085	Intron
(AAT)14	28842	132488	Intron
(TCC)9	28842	169523	Intron
(AAT)14	28842	132488	Intron
(TAA)9	28842	542370	Intron
(AAT)13	27890	14287	Intron
(GT)23	27894	265454	Intron
(TC)24	27894	392751	Intron
(CT)17	28859	126267	Intron
(TA)17	28859	126300	Intron
(AT)21	27904	5061	Intron
(TTA)13	27914	43529	Intron
(TA)18	28883	171641	Intron
(ATT)14	28883	82113	Intron
(AT)22	27916	49240	Intron
(AT)14	28885	28830	Intron
(TC)24	27894	392751	Intron
(ATT)16	27894	22679	Intron
(GT)23	27894	265454	Intron
(TA)21	28859	89263	Intron
(CT)17	28859	126267	Intron
(TA)17	28859	126300	Intron
(AGA)10	28859	133152	Intron
(AGA)11	27955	8493	Intron
(TTC)9	27955	68550	Intron
(TTCT)7	27955	9235	Intron
(TA)17	28927	12275	Intron
(TA)20	28934	23434	Intron
(AT)24	27973	10721	Intron
(AAG)10	28952	17030	Intron
(TC)23	27985	7133	Intron
(CT)19	27985	344673	Intron
(CT)14	27985	363618	Intron
(AG)18	28966	85672	Intron
(TC)13	28966	227952	Intron
(AT)22	28966	19463	Intron
(AG)17	29005	32810	Intron
(TC)19	29005	40213	Intron
(TG)13	29005	47793	Intron
(TA)19	29005	54842	Intron
(TC)30	29005	114255	Intron

(TTA)11	29005	89438	Intron
(CT)15	28035	25374	Intron
(GA)21	28035	167033	Intron
(AT)27	28054	16863	Intron
(TAT)9	29070	34555	Intron
(GA)20	29080	13988	Intron
(AG)17	28102	8108	Intron
(TA)19	28102	55618	Intron
(CT)21	29093	68626	Intron
(TC)23	28128	32036	Intron
(TAA)12	28128	13202	Intron
(TA)13	29117	3549	Intron
(GA)22	29118	27234	Intron
(CT)19	28152	17569	Intron
(AC)14	28152	17612	Intron
(CT)13	28152	31087	Intron
(TC)18	28152	114430	Intron
(TA)23	28152	152854	Intron
(TA)13	28192	63548	Intron
(AAT)9	28193	81563	Intron
(CT)16	28200	48277	Intron
(AT)13	29168	125861	Intron
(CT)21	29168	131479	Intron
(AT)16	29168	131672	Intron
(CT)16	28202	12419	Intron
(AT)22	28202	38122	Intron
(CT)15	28202	66738	Intron
(AG)25	29171	237145	Intron
(ATA)9	29171	142016	Intron
(AG)19	29172	5504	Intron
(TTA)9	29172	50725	Intron
(TA)15	28208	18054	Intron
(GA)10	29176	41165	Intron
(TAA)9	29176	50513	Intron
(AC)14	28211	7843	Intron
(AT)13	28211	7871	Intron
(AT)16	28223	46016	Intron
(TA)21	28226	46275	Intron
(ATA)19	28226	176339	Intron
(TG)19	29195	20240	Intron
(TA)16	29200	30564	Intron
(AT)21	29212	12888	Intron
(AT)27	29212	48531	Intron
(TC)13	29212	94840	Intron
(GA)16	29212	127815	Intron

(TAT)17	29212	55936	Intron
(TA)38	28256	19649	Intron
(AT)20	29222	30874	Intron
(AT)14	29222	59541	Intron
(TA)24	29222	108971	Intron
(AT)39	29222	215390	Intron
(GAA)9	29222	28004	Intron
(TA)20	29227	12498	Intron
(TTC)9	29235	101789	Intron
(TC)14	28281	66764	Intron
(TA)16	28281	90824	Intron
(AT)19	28281	96079	Intron
(AT)25	28324	166692	Intron
(TC)29	28293	20515	Intron
(AT)25	28301	7089	Intron
(TA)15	28304	81435	Intron
(GAA)10	28348	11321	Intron
(AT)24	28350	38434	Intron
(AT)22	49727	728	Intron
(TA)13	52298	3253	Intron
(TTTAT)8	52899	677	Intron
(TA)18	36845	972	Intron
(AT)25	37959	408	Intron
(TA)22	54308	632	Intron
(AT)35	57713	1709	Intron
(AT)13	39340	538	Intron
(TA)13	40322	4421	Intron
(TA)42	41844	134	Intron
(TATTT)6	41328	433	Intron
(AT)23	43365	340	Intron
(AAT)14	44057	504	Intron
(TA)22	44956	489	Intron
(TA)13	45715	390	Intron
(TTTA)6	46497	342	Intron
(AT)14	27375	96745	Intron
(AT)28	27383	34283	Intron
(TAT)8	27394	21099	Intron
(GA)28	27394	128536	Intron
(AT)16	27394	165617	Intron
(AG)12	27401	103806	Intron
(CGC)8	27403	800	Intron
(AT)12	27428	11493	Intron
(AATA)6	27448	41080	Intron
(AT)22	34943	727	Intron
(TTTA)6	35428	856	Intron

(TA)17	28828	14166	Intron
(TA)40	28830	91652	Intron
(TAT)11	28830	28952	Intron
(CT)13	28838	92004	Intron
(CT)18	28842	86309	Intron
(TA)17	28842	239455	Intron
(AT)22	28842	245584	Intron
(TG)22	28842	320869	Intron
(TC)14	28842	436085	Intron
(AAT)14	28842	132488	Intron
(TCC)9	28842	169523	Intron
(AAT)14	28842	132488	Intron
(TAA)9	28842	542370	Intron
(AAT)13	27890	14287	Intron
(GT)23	27894	265454	Intron
(TC)24	27894	392751	Intron
(CT)17	28859	126267	Intron
(TA)17	28859	126300	Intron
(AT)21	27904	5061	Intron
(TTA)13	27914	43529	Intron
(TA)18	28883	171641	Intron
(ATT)14	28883	82113	Intron
(AT)22	27916	49240	Intron
(AT)14	28885	28830	Intron
(TC)24	27894	392751	Intron
(ATT)16	27894	22679	Intron
(GT)23	27894	265454	Intron
(TA)21	28859	89263	Intron
(CT)17	28859	126267	Intron
(TA)17	28859	126300	Intron
(AGA)10	28859	133152	Intron
(AGA)11	27955	8493	Intron
(TTC)9	27955	68550	Intron
(TTCT)7	27955	9235	Intron
(TA)17	28927	12275	Intron
(TA)20	28934	23434	Intron
(AT)24	27973	10721	Intron
(AAG)10	28952	17030	Intron
(TC)23	27985	7133	Intron
(CT)19	27985	344673	Intron
(CT)14	27985	363618	Intron
(AG)18	28966	85672	Intron
(TC)13	28966	227952	Intron
(AT)22	28966	19463	Intron
(AG)17	29005	32810	Intron

(TC)19	29005	40213	Intron
(TG)13	29005	47793	Intron
(TA)19	29005	54842	Intron
(TC)30	29005	114255	Intron
(TTA)11	29005	89438	Intron
(CT)15	28035	25374	Intron
(GA)21	28035	167033	Intron
(AT)27	28054	16863	Intron
(TAT)9	29070	34555	Intron
(GA)20	29080	13988	Intron
(AG)17	28102	8108	Intron
(TA)19	28102	55618	Intron
(CT)21	29093	68626	Intron
(TC)23	28128	32036	Intron
(TAA)12	28128	13202	Intron
(TA)13	29117	3549	Intron
(GA)22	29118	27234	Intron
(CT)19	28152	17569	Intron
(AC)14	28152	17612	Intron
(CT)13	28152	31087	Intron
(TC)18	28152	114430	Intron
(TA)23	28152	152854	Intron
(TA)13	28192	63548	Intron
(AAT)9	28193	81563	Intron
(CT)16	28200	48277	Intron
(AT)13	29168	125861	Intron
(CT)21	29168	131479	Intron
(AT)16	29168	131672	Intron
(CT)16	28202	12419	Intron
(AT)22	28202	38122	Intron
(CT)15	28202	66738	Intron
(AG)25	29171	237145	Intron
(ATA)9	29171	142016	Intron
(AG)19	29172	5504	Intron
(TTA)9	29172	50725	Intron
(TA)15	28208	18054	Intron
(GA)10	29176	41165	Intron
(TAA)9	29176	50513	Intron
(AC)14	28211	7843	Intron
(AT)13	28211	7871	Intron
(AT)16	28223	46016	Intron
(TA)21	28226	46275	Intron
(ATA)19	28226	176339	Intron
(TG)19	29195	20240	Intron
(TA)16	29200	30564	Intron

(AT)21	29212	12888	Intron
(AT)27	29212	48531	Intron
(TC)13	29212	94840	Intron
(GA)16	29212	127815	Intron
(TAT)17	29212	55936	Intron
(TA)38	28256	19649	Intron
(AT)20	29222	30874	Intron
(AT)14	29222	59541	Intron
(TA)24	29222	108971	Intron
(AT)39	29222	215390	Intron
(GAA)9	29222	28004	Intron
(TA)20	29227	12498	Intron
(TTC)9	29235	101789	Intron
(TC)14	28281	66764	Intron
(TA)16	28281	90824	Intron
(AT)19	28281	96079	Intron
(AT)25	28324	166692	Intron
(TC)29	28293	20515	Intron
(AT)25	28301	7089	Intron
(TA)15	28304	81435	Intron
(GAA)10	28348	11321	Intron
(AT)24	28350	38434	Intron
(AT)22	49727	728	Intron
(TA)13	52298	3253	Intron
(TTTAT)8	52899	677	Intron
(TA)18	36845	972	Intron
(AT)25	37959	408	Intron
(TA)22	54308	632	Intron
(AT)35	57713	1709	Intron
(AT)13	39340	538	Intron
(TA)13	40322	4421	Intron
(TA)42	41844	134	Intron
(TATTT)6	41328	433	Intron
(AT)23	43365	340	Intron
(AAT)14	44057	504	Intron
(TA)22	44956	489	Intron
(TA)13	45715	390	Intron
(TTTA)6	46497	342	Intron
(AT)14	27375	96745	Intron
(AT)28	27383	34283	Intron
(TAT)8	27394	21099	Intron
(GA)28	27394	128536	Intron
(AT)16	27394	165617	Intron
(AG)12	27401	103806	Intron
(CGC)8	27403	800	Intron

(AT)12	27428	11493	Intron
(AATA)6	27448	41080	Intron
(AT)22	34943	727	Intron
(TTTA)6	35428	856	Intron
(TA)17	28828	14166	Intron
(TA)40	28830	91652	Intron
(TAT)11	28830	28952	Intron
(CT)13	28838	92004	Intron
(CT)18	28842	86309	Intron
(TA)17	28842	239455	Intron
(AT)22	28842	245584	Intron
(TG)22	28842	320869	Intron
(TC)14	28842	436085	Intron
(AAT)14	28842	132488	Intron
(TCC)9	28842	169523	Intron
(AAT)14	28842	132488	Intron
(TAA)9	28842	542370	Intron
(AAT)13	27890	14287	Intron
(GT)23	27894	265454	Intron
(TC)24	27894	392751	Intron
(CT)17	28859	126267	Intron
(TA)17	28859	126300	Intron
(AT)21	27904	5061	Intron
(TTA)13	27914	43529	Intron
(TA)18	28883	171641	Intron
(ATT)14	28883	82113	Intron
(AT)22	27916	49240	Intron
(AT)14	28885	28830	Intron
(TC)24	27894	392751	Intron
(ATT)16	27894	22679	Intron
(GT)23	27894	265454	Intron
(TA)21	28859	89263	Intron
(CT)17	28859	126267	Intron
(TA)17	28859	126300	Intron
(AGA)10	28859	133152	Intron
(AGA)11	27955	8493	Intron
(TTC)9	27955	68550	Intron
(TTCT)7	27955	9235	Intron
(TA)17	28927	12275	Intron
(TA)20	28934	23434	Intron
(AT)24	27973	10721	Intron
(AAG)10	28952	17030	Intron
(TC)23	27985	7133	Intron
(CT)19	27985	344673	Intron
(CT)14	27985	363618	Intron

(AG)18	28966	85672	Intron
(TC)13	28966	227952	Intron
(AT)22	28966	19463	Intron
(AG)17	29005	32810	Intron
(TC)19	29005	40213	Intron
(TG)13	29005	47793	Intron
(TA)19	29005	54842	Intron
(TC)30	29005	114255	Intron
(TTA)11	29005	89438	Intron
(CT)15	28035	25374	Intron
(GA)21	28035	167033	Intron
(AT)27	28054	16863	Intron
(TAT)9	29070	34555	Intron
(GA)20	29080	13988	Intron
(AG)17	28102	8108	Intron
(TA)19	28102	55618	Intron
(CT)21	29093	68626	Intron
(TC)23	28128	32036	Intron
(TAA)12	28128	13202	Intron
(TA)13	29117	3549	Intron
(GA)22	29118	27234	Intron
(CT)19	28152	17569	Intron
(AC)14	28152	17612	Intron
(CT)13	28152	31087	Intron
(TC)18	28152	114430	Intron
(TA)23	28152	152854	Intron
(TA)13	28192	63548	Intron
(AAT)9	28193	81563	Intron
(CT)16	28200	48277	Intron
(AT)13	29168	125861	Intron
(CT)21	29168	131479	Intron
(AT)16	29168	131672	Intron
(CT)16	28202	12419	Intron
(AT)22	28202	38122	Intron
(CT)15	28202	66738	Intron
(AG)25	29171	237145	Intron
(ATA)9	29171	142016	Intron
(AG)19	29172	5504	Intron
(TTA)9	29172	50725	Intron
(TA)15	28208	18054	Intron
(GA)10	29176	41165	Intron
(TAA)9	29176	50513	Intron
(AC)14	28211	7843	Intron
(AT)13	28211	7871	Intron
(AT)16	28223	46016	Intron

(TA)21	28226	46275	Intron
(ATA)19	28226	176339	Intron
(TG)19	29195	20240	Intron
(TA)16	29200	30564	Intron
(AT)21	29212	12888	Intron
(AT)27	29212	48531	Intron
(TC)13	29212	94840	Intron
(GA)16	29212	127815	Intron
(TAT)17	29212	55936	Intron
(TA)38	28256	19649	Intron
(AT)20	29222	30874	Intron
(AT)14	29222	59541	Intron
(TA)24	29222	108971	Intron
(AT)39	29222	215390	Intron
(GAA)9	29222	28004	Intron
(TA)20	29227	12498	Intron
(TTC)9	29235	101789	Intron
(TC)14	28281	66764	Intron
(TA)16	28281	90824	Intron
(AT)19	28281	96079	Intron
(AT)25	28324	166692	Intron
(TC)29	28293	20515	Intron
(AT)25	28301	7089	Intron
(TA)15	28304	81435	Intron
(GAA)10	28348	11321	Intron
(AT)24	28350	38434	Intron
(AG)12	52562	705	Exon
(CT)12	53467	625	Exon
(TTC)9	56356	795	Exon
(TCT)11	56420	703	Exon
(CTT)8	57713	1908	Exon
(CCA)11	41929	614	Exon
(GA)12	46666	567	Exon
(AG)12	27375	57986	Exon
(CT)24	27446	66135	Exon
(TTTA)8	31548	415	Exon
(ATA)9	31781	778	Exon
(GAA)15	32351	363	Exon
(GAG)8	33124	291	Exon
(GAA)10	33978	1155	Exon
(TCT)14	34878	906	Exon
(AGTTG)5	28872	74253	Exon
(GA)21	27934	101685	Exon
(AGA)12	28859	184014	Exon
(GA)20	27934	101685	Exon

(AGA)9	27940	64094	Exon
(TCT)9	27942	37672	Exon
(TA)19	27956	165721	Exon
(TTC)11	27956	110013	Exon
(TCT)10	27956	139185	Exon
(TC)15	27985	25011	Exon
(AT)13	27985	361749	Exon
(AGA)10	27985	240139	Exon
(GAA)9	28966	69434	Exon
(TC)13	29005	12261	Exon
(AG)18	28035	65768	Exon
(CT)13	28035	118480	Exon
(CTT)9	28035	259835	Exon
(CT)22	28062	1047	Exon
(CT)15	28092	8389	Exon
(ATA)10	28092	39896	Exon
(TA)13	28102	38190	Exon
(ATA)20	28134	37830	Exon
(AG)23	28140	53602	Exon
(TA)16	28152	39736	Exon
(AT)13	28154	12198	Exon
(AG)21	28162	7145	Exon
(AG)15	28162	7284	Exon
(CAT)12	28162	38367	Exon
(TTA)11	28211	8303	Exon
(TTC)10	28226	82476	Exon
(AG)15	29220	500	Exon
(AGA)9	29222	186906	Exon
(GAA)15	29222	188153	Exon
(AG)14	29227	6328	Exon
(AT)14	28350	28350	Exon

4.2) Transferability and polymorphism survey of castor bean SSRs in *Jatropha curcas*

The transferability of SSRs from castor bean to Jatropha and the extent of polymorphism among *J. curcas* genotypes were tested for 302 SSRs (87 from exons, 78 from non- genic regions, 71 from introns and 66 from 5' & 3' UTRs) on 49 genotypes of *J. curcas.* Out of 302 SSRs, 273 amplified on castor bean DNA out of which 211 amplified on 43 genotypes of *J. curcas*. The amplification failure in *J. curcas* genotypes, Urli-Kanchan, KcJK5, Hissar local, SKN-Big and Hansraj amplification was 2.6%, 2.6%, 4.8%, 6.23%, 7.1% and 7.6%, respectively

in comparison to 43 genotypes. Six Jatropha species produced amplicons with 211 primer pairs except for *J. mahotwani*, *J. multifida, J. glandulifera* where the percent failure was 4.74%, 6.23% and 8.2%, respectively. Ten percent of the SSRs failed to amplify on castor bean DNA which was attributed to primer mismatches. Out of 211 SSRs, 36.01% from exons, 21.8% from introns, 16.6% from 5' and 3' UTRs, and 25.6% from non-genic regions showed amplicon in Jatropha. Sixteen SSRs, including 5 from 5'UTRs, 5 from non- genic regions, 3 from introns and 3 from exons showed polymorphism in *J. curcas* genotypes (Table 4.2). The number of alleles per SSR locus ranged from 2 to 6 with a total of 43 alleles ranging in sizes from 200 bp to 600 bp in *J. curcas* genotypes (Fig. 4.1). Twenty six SSRs, 12 from 5' & 3' UTRs, 7 from introns, 4 from non genic regions and 3 from exons were polymorphic on 9 Jatropha species (*J. maheshwarii*, *J. multifida*, *J. gossypifolia*, *J. podagrica*, *J. glandulifera*, *J. curcas, J. tanjorensis, J. villosa, J. integerrima*) (Table 4.3). The number of alleles per SSR locus ranged from 2 to 7 in *Jatropha* species (Fig. 4.2). Five SSRs (JM8, JM10, JM11, JM15 and JM16) showed polymorphisms on *J. curcas* genotypes as well as Jatropha species. The transferability of SSRs from castor bean was higher to *J. curcas* (~70%) and lowest to *J. glandulifera* (58%), with 63-68% transferability for other species. The level of polymorphism was higher (37.8%) in SSRs from 5'UTRs followed by 24.3% from introns, 18.9% from non-genic regions, 10.8% from exons and 8.10% from 3'UTRs of the genome. The SSRs with dinucleotide repeat motifs showed higher levels of polymorphisms than trinucleotide repeats. Out of 37 SSRs, which were polymorphic in Jatropha, 35 (94.5%) were dinucleotide repeats. Tetra- and penta- nucleotide repeats did not show any polymorphisms in Jatropha.

Out of all SSRs, which were successfully transferred to Jatropha, 50% contained 15 to 30 repeat units whereas 20% of the SSRs had repeat unit of > 30. The majority of SSRs with

successful amplification and polymorphisms contained more than 15 repeat units (Table 4.4). The PIC values for polymorphic SSRs in *J. curcas* genotypes and Jatropha species varied from 0.2 to 0.5 with an average of 0.2 and 0.3 to 0.7 with an average of 0.5, respectively. The SSRs with dinucleotide repeat motifs showed higher allele numbers (average 2.7 per locus) followed by trinucleotide (average alleles 2.3 per locus). To understand the possible relationship between polymorphism of SSR markers with repeat unit length of corresponding SSRs in *J. curcas* genotypes and Jatropha species, a line graph was plotted between repeat unit length and numbers of alleles detected (Fig. 4.3). The analysis showed wide variation in the number of alleles detected for SSRs with 16 and 25 repeat motifs compared to SSRs with low or high numbers of repeat motifs. An exception to this observation was found for SSR, JM15, which contained maximum number of repeat units (TA)42 with only two alleles, whereas, SSR, JM20 with lower repeat motifs (TA)23 showed highest number of alleles (7).

4.2.1) Genetic diversity analysis in Jatropha genotypes and species with SSRs

The major allele frequency (MAF) for a SSR ranged from 0.4 to 0.9 for *J. curcas* genotypes and 0.1 to 0.5 for Jatropha species. The observed heterozygosity (HO) ranged from 0.1 to 0.5 (mean = 0.2) in *J. curcas* genotypes and 0.4 to 0.7 (mean = 0.6) in Jatropha species, and expected (HE) heterozygosities ranged from 0.1 to 0.5 (mean = 0.2) in *J. curcas* genotypes and 0.4 to 0.7 (mean = 0.6) in Jatropha species. Hardy–Weinberg probability tests revealed no significant deviations from expected genotype proportions ($P > 0.004$). There was no evidence of linkage disequilibrium among loci ($P > 0.001$) after corrections for multiple tests.

Phylogenetic relationships among different genotypes of *J. curcas* and 9 species of Jatropha were inferred based on SSRs analysis. Jaccard's genetic co-efficient for *J. curcas* genotypes varied from 1.078 to 9.016. The highest genetic dissimilarity co-efficient (9.016) was

observed between 16 polymorphic SSRs in *J. curcas* genotypes while the lowest value of (1.078) was measured between eight combinations. In case of Jatropha species, highest value of dissimilarity co-efficient was 1.42 between eight pair-wise species combinations and lowest value was 1.11 between *J. integerrima* and *J. gossypifolia.* UPGMA cluster analysis of the Jaccard's co-efficient generated a dendrogram for *J. curcas* genotypes (Fig. 4.4) and Jatropha species (Fig. 4.5) which illustrated the overall genetic relationship among the genotypes and species surveyed. Cluster analysis indicated four distinct clusters one comprising all the species of Jatropha, whereas in *J. curcas* genotypes there were six different clusters. The *J. curcas* genotype 1 (Urli-Kanchan) and 32 (Hissar local) remained as outliers and formed the first and sixth clusters respectively.

Table 4.2: Polymorphism analysis with SSRs in *Jatropha curcas* genotypes

SSR locus	Repeat motif	Location of repeat motif	No. alleles detected	HO	HE	PIC
JM1	(TA)22	5'UTR	3	0.1832	0.1833	0.2
JM2	(TA)20	5'UTR	2	0.5114	0.5115	0.4
JM3	(AT)20	5'UTR	2	0.2832	0.2832	0.3
JM4	(AT)29	5'UTR	2	0.1166	0.1166	0.1
JM5	(TTA)22	5'UTR	3	0.2998	0.2999	0.2
JM6	(AT)24	Intron	6	0.2149	0.2149	0.2
JM7	(TA)24	Intron	4	0.1499	0.1499	0.1
JM8	(TC)23	Intron	3	0.1499	0.1499	0.1
JM9	(TC)15	Exon	2	0.1832	0.1833	0.2
JM10	(GAA)15	Exon	2	0.1149	0.1150	0.1
JM11	(CAT)12	Exon	2	0.2149	0.2149	0.2
JM12	(TA)33	Non genic	2	0.1149	0.1150	0.1
JM13	(AT)25	Non genic	5	0.1832	0.1833	0.2
JM14	(AT)34	Non genic	2	0.2732	0.2732	0.2
JM15	(TA)42	Non genic	2	0.5397	0.5398	0.5
JM16	(TA)38	Non genic	2	0.4481	0.4481	0.4

JM: Jatropha Microsatellite; F: Forward; R: Reverse; HO: Observed heterzygosity; HE: Expected heterozygosity; PIC: Polymorphic information content

Table 4.3: Polymorphisms analysis with SSRs in Jatropha species

SSR locus	Repeat motif	Location of repeat motif	No. of alleles detected	HO	HE	PIC
JM8	(TC)23	Intron	3	0.7654	0.7654	0.7
JM10	(GAA)15	Exon	2	0.5925	0.5926	0.5
JM11	(CAT)12	Exon	2	0.5679	0.5679	0.5
JM15	(TA)42	Non genic	2	0.4938	0.4938	0.4
JM16	(TA)38	Non genic	2	0.6913	0.6914	0.6
JM17	(TA)31	5'UTR	3	0.6172	0.6173	0.6
JM18	(AT)25	5'UTR	2	0.5925	0.5926	0.5
JM19	(TA)33	5'UTR	2	0.7407	0.7407	0.7
JM20	(TA)23	5'UTR	7	0.6419	0.6420	0.6
JM21	(AT)25	5'UTR	2	0.6913	0.6914	0.6
JM22	(AT)29	5'UTR	3	0.4444	0.4444	0.3
JM23	(AT)32	5'UTR	3	0.7654	0.7654	0.7
JM24	(AT)20	5'UTR	4	0.6913	0.6914	0.6
JM25	(CT)20	5'UTR	2	0.6913	0.6914	0.6
JM26	(TC)18	3'UTR	2	0.7160	0.7160	0.7
JM27	(TA)24	3'UTR	3	0.7654	0.7654	0.7
JM28	(AT)27	3'UTR	2	0.7160	0.7160	0.7
JM29	(AT)25	Intron	2	0.7160	0.7160	0.7
JM30	(AT)25	Intron	2	0.7160	0.7160	0.7
JM31	(ATA)19	Intron	2	0.7160	0.7160	0.7
JM32	(AT)22	Intron	2	0.6666	0.6667	0.6
JM33	(GA)21	Intron	3	0.4938	0.4938	0.4
JM34	(AT)24	Intron	3	0.5679	0.5679	0.5
JM35	(CT)12	Exon	2	0.6419	0.6420	0.6
JM36	(AT)29	Non genic	3	0.6419	0.6420	0.6
JM37	(AAT)16	Non genic	3	0.7407	0.7407	0.7

Table 4.4: Extent of amplification and polymorphism among SSRs of varying repeat units in *J. curcas*

Repeat units	SSRs tested (%)	SSRs with amplification (%)	Polymorphic SSRs (%)
10-15	11	3.00	1.89
16-20	13	12.26	3.31
21-25	14	18.20	6.16
26-30	26	20.20	3.31
31-35	22	9.00	1.89
36-40	8	3.20	0.47
>40	6	4.00	0.47

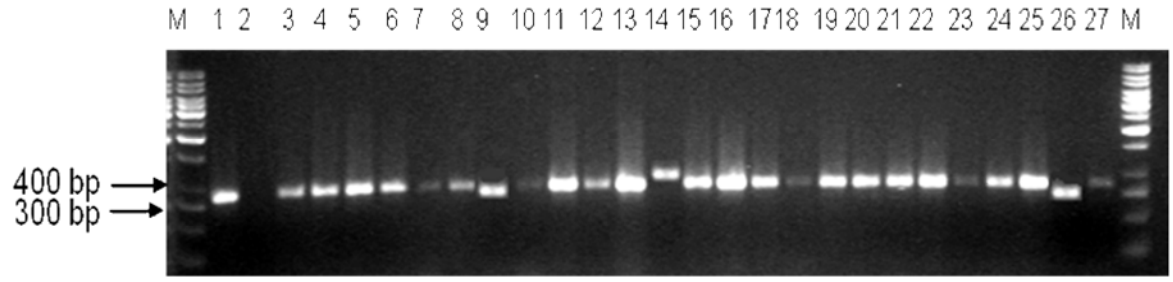

Fig. 4.1: Polymorphisms in amplicon size for intronic SSR (TC)23 at locus JM8 in Jatropha genotypes (1-27); M =Molecular weight marker of 100 bp

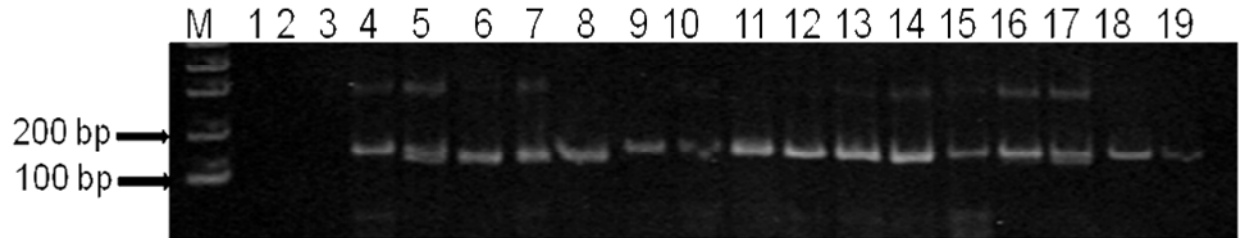

Fig. 4.1.1: Polymorphisms in amplicon size for 5'UTR SSR (AT)20 at locus JM3 in 19 Jatropha genotypes; M =Molecular weight marker of 100 bp

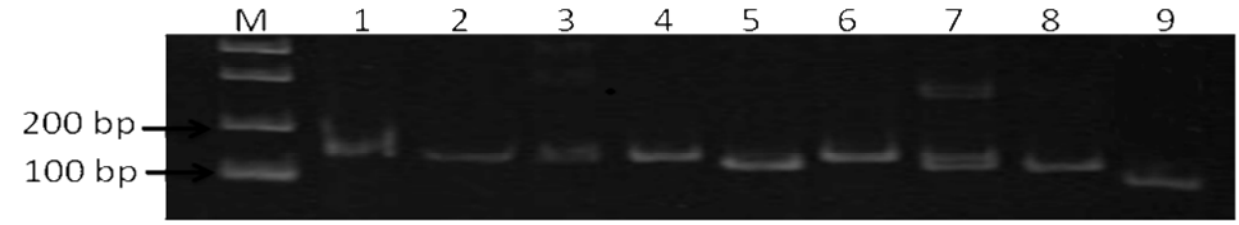

Fig. 4.2: Polymorphism at SSR locus JM22 (5'UTR) (AT)29 on Jatropha species (1-*J.maheshwarii*, 2- *J.multifida*, 3- *J.gossypifolia*, 4- *J.podagrica*, 5- *J.glandulifera*, 6- *J.curcas,* 7-*J.tanjorensis,* 8- *J.villosa,* 9- *J.integerrima)*; M =Molecular weight marker of 100 bp

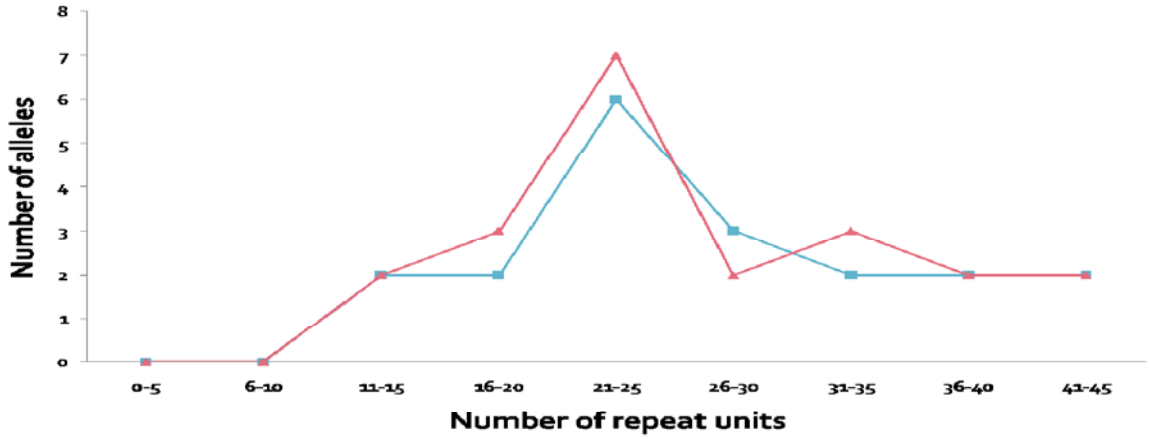

Fig. 4.3: Number of alleles per locus for SSR markers of different repeat units in *J. curcas* genotypes (■) and Jatropha species (▲)

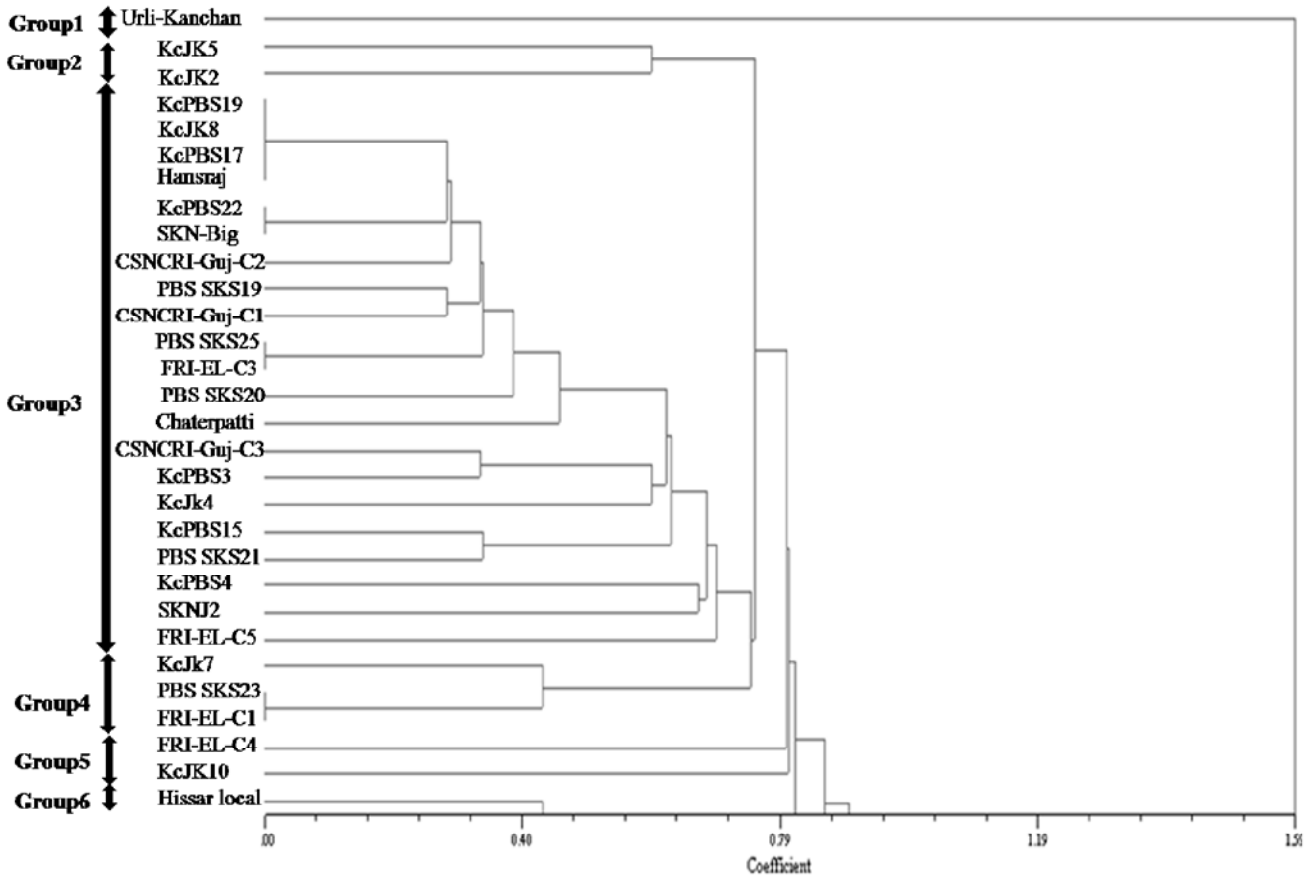

Fig. 4.4: Dendrogram based on allele sharing genetic distances of 49 genotypes of *J. curcas* on the basis Jaccard's similarity coefficient

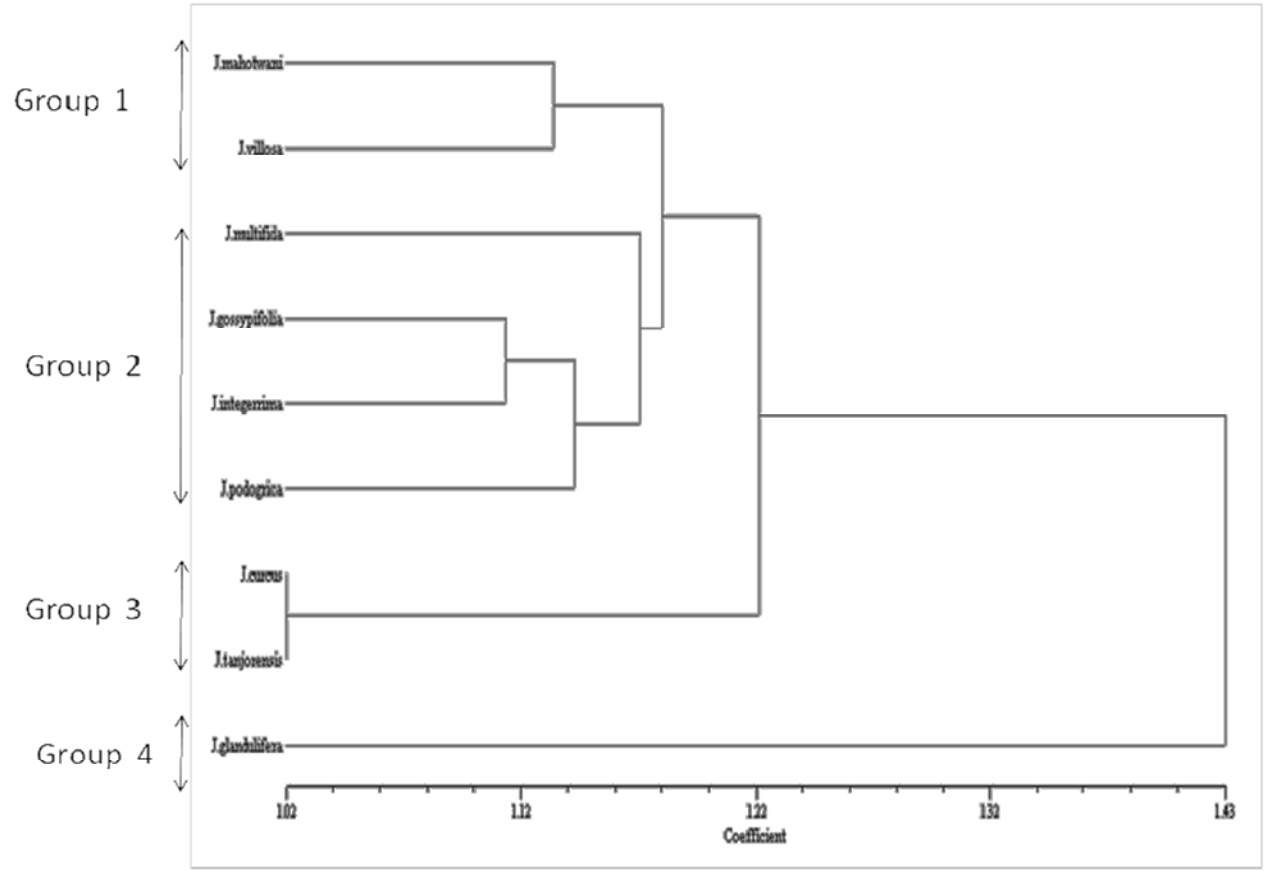

Fig. 4.5: Dendrogram based on allele sharing genetic distances of 9 species of Jatropha on the basis Jaccard's similarity coefficient

4.3) Comparative genomics of fatty acid biosynthesis genes in major oil seed plant species

The fatty acid biosynthesis pathway includes 32 gene families involved in the conversion of acetyl Co-A into different fatty acids and their storage in oil bodies. A total of 68 protein sequences were retrieved for 32 gene families from the comprehensive lipid gene catalog of Arabidopsis (Beisson *et al.*, 2003) and functional domains were identified for each gene family. The 68 protein sequences from Arabidopsis were queried for fatty acid biosynthesis genes in *B. rapa*, soybean and castor bean databases. A total of 261 genes belonging to 32 gene families were identified and retrieved from four plant species, out of which, 68 were from Arabidopsis,

62 from *B. rapa*, 55 from castor bean and 76 from soybean (Table 4.5). Detailed gene structures, exon- intron coordinates of each gene are given in Table 4.6.

4.3.1) Expression status of fatty acid biosynthesis genes

In silico expression analysis revealed that a total of 32 gene families were encoded by 68 genes in Arabidopsis, 62 genes in Brassica, 49 genes in castor bean and 76 genes in soybean with identities to ESTs (Fig. 4.6). Thirteen genes of Arabidopsis, 15 from castor bean, 8 from soybean and 2 from Brassica showed tissue preferential expression patterns. Twenty two genes from four plant species were expressed in seeds, 4 in leaves, 3 in flower, and 1 in roots (Table 4.7). FAD 2 and one homolog of Stearoyl desaturase gene had maximum seed ESTs in castor bean. The transcripts of FatB and DGAT were highly abundant in soybean and Brassica seeds, whereas Oleosin transcript was highly abundant in the seeds of the four plant species on the basis of matching ESTs.

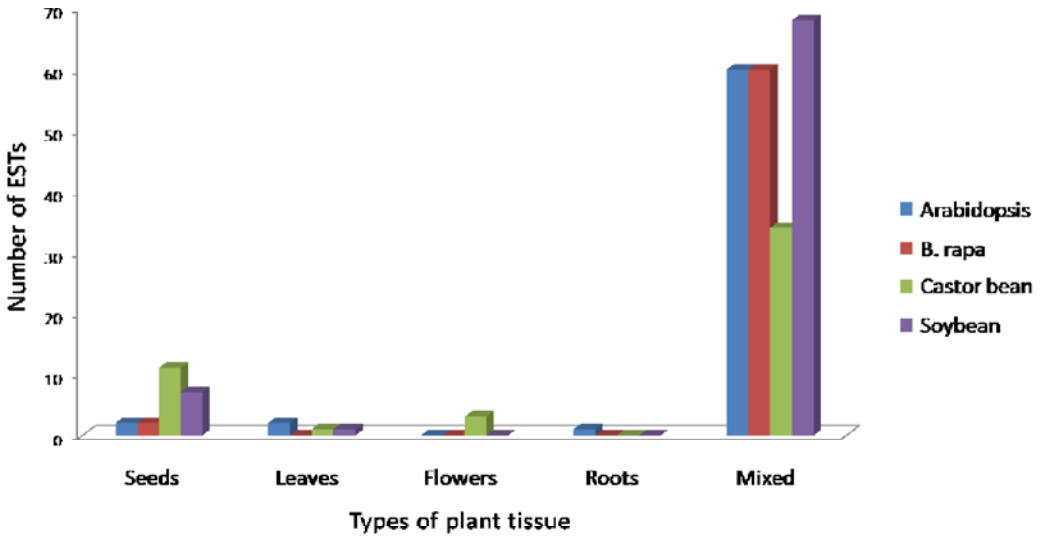

Fig.4.6: Transcript abundance (based on matching ESTs) of oil biosynthesis and accumulation genes in different tissues

Table 4.5: Gene families and genes involved in oil biosynthesis in Arabidopsis, *B. rapa*, castor bean and soybean

Gene Family	Gene Name	GenBank Accession number				CDS length (bp)			
		Arabidopsis	*B. rapa*	Castor bean	Soybean	Arabidopsis	*B. rapa*	Castor bean	Soybean
ACCase	ACCase	At1g36180	X77382	29908.m005991	Glyma04g11550	5997	2193	6723	6834
	Alpha- carboxyl transferase	At2g38040	AY538675	27798.m000585	Glyma18g42280	2346	2295	2355	2130
	Beta- carboxyl transferase	ATCG00500	Z50868	28890.m000006	_____	636	984	528	------
	Biotin carboxylase	At5g35360	AY034410	30185.m000954	Glyma05g36450	1683	1608	1935	1731
	Biotin carrier	At5g15530	AY538674	29630.m000809 29929.m004560	Glyma08g03120	726	783	780 924	1611
Elongase	Malonyl Co-A transacylase	At2g30200	AJ007046	30113.m001448	Glyma18g06500	1104	993	987	1113
	Beta-Ketoacyl ACP synthase I	At5g46290	AF244519	29693.m002034	Glyma10g04680	1422	1380	1437	1452
	Beta-Ketoacyl ACP synthase II	At1g74960	AF244520	29739.m003711	Glyma13g19010	1770	1302	2013	1305
	Beta-Ketoacyl ACP synthase III	At1g62640	AF179854	28455.m000368	Glyma09g41380 Glyma15g00550 Glyma18g44350	1215	186	1233	1194 831 1254
	3-Ketoacyl- acp-dehydrase	At1g62610 At3g46170 At3g55290 At3g55310	AF382146	30147.m013777	Glyma08g10760 Glyma18g01280	924 867 822 822		735	924 963
	3-Ketoacyl- Co-A reductase (KAR)	At1g24360	AY196197	29929.m004732	Glyma11g37320	927	960	987	963
	Enoyl- ACP reductase (ENR)	At2g05990	AJ243087 AJ243088 AJ243089 AJ243090 x95462	27843.m000160 29650.m000277	Glyma11g10770 Glyma12g03060 Glyma18g31780	11155	1158 1161 1161 1164 1158	108310 83	1176 1203 1533

	Hydroxyacyl ACP Dehydrase (HD)	At2g22230	AF382146	30200.m000354	Glyma05g24650 Glyma08g07870 Glyma08g19200 Glyma15g05800	663	672	534	417 513 219 822
	Plastidial 1 acylglycerol phosphate acyltransferase	At4g30580	----------	29687.m000572	Glyma06g28540 Glyma12g28470	1071	------	987	2406 1776
	Plastidial Glycerol phosphate acyltransferase	At1g32200	----------	30068.m002660	Glyma09g34110	1380	------	1236	1413
Desaturase	Monoglactoacsyldiacylglycerol desaturase (FAD 5)	At3g15850		29841.m002863	Glyma07g03370 Glyma08g22730	1116		1161	1101 1173
	Stearoyl –ACP desaturase	At1g43800 At2g43710 At3g02610 At3g02630 At3g02620 At5g16230 At5g16240	X63364 X74782 AY642537	27985.m000877 28470.m000428 29929.m004515 30020.m000203	Glyma02g15600 Glyma07g32850 Glyma13g08970 Glyma14g27990	1176 1206 984 1191 984 1206 1185	1200 1206 1200	1176 336 960 1191	1176 1176 1185 1014
	Oleate desaturase (FAD 6)	At4g30950	AY642535 AY642540	29696.m000105	Glyma02g36460	1347	1332 1293	825	1287
	Linoleate desaturase (FAD 7)	At3g11170	AY592974 AY599884 FJ985689 FJ985690 FJ985691 L01418 L22962	28176.m000273 29681.m001360 29814.m000719	Glyma01g29630 Glyma03g07570 Glyma07g18350	1467	1320 1152 1134 1299 1335 1152 1134	1359 1131 1383	1359 1362 1362
	Linoleate desaturase (FAD 8)	At5g05580	AY592974 AY599884 FJ985691	28176.m000273 29681.m001360 29814.m000719	Glyma01g29630 Glyma03g07570 Glyma07g18350	1308 1152	1320 1152 1335	1359 1131 1383	1359 1362 1362
	ER-Oleate desaturase (FAD 2)	At3g12120	AY577313 DQ518276	28035.m000362	Glyma03g30070 Glyma09g17170	1152	1155 780	1164	1152 1161

			DQ518277 DQ518278 FJ907397 FJ907398 FJ907399 FJ907400 FJ907401 FJ952144		Glyma10g42470 Glyma19g32940 Glyma20g24530		780 780 1155 1155 1155 1155 1155 1155		1140 1152 1140
	ER-Linoleate desaturase (FAD 3)	At2g29980	AY592974 AY599884 FJ985689 L01418 L22962	29681.m001360	Glyma01g29630 Glyma07g18350	1161	1320 1152 1134 1152 1134	1131	1359 1362
Thioesterase	Acyl- ACP thioesterase (FatA)	At3g25110 At4g13050	X87842	30217.m000262	Glyma08g46360 Glyma18g36130	1089 1104	1176	1269	1191 1125
	Palmitoyl- ACP thioesterase (FatB)	At1g08510	DQ847275 Fj715952	29842.m003515 29848.m004677	Glyma0421910 Glyma05g08060 Glyma06g23560 Glyma17g12940	1239	1245 1239	753 1260	1332 1251 1422 1251
TAG synthesis	Diacylglycerol Acyltranferase (DGAT 1)	At2g19450	AF164434	29912.m005373	Glyma13g16560	1593	1512	1830	1347
	Diacylglycerol Acyltranferase (DGAT 2)	At3g51520	AF155224	29682.m000581	Glyma09g32790	945	1056	768	987
	Lysophosphosphatidic acid acyltransferase (LPAAT)	At1g01610 At1g51260 At1g78690 At1g80950 At2g27090 At2g38110 At3g05510 At3g11430 At3g18850 At3g57650	AF111161 Gu045434 GU04535 Gu045436 Z49860	27810.m000646 29851.m002448 30169.m006433 30170.m013990 29736.m002070 29822.m003441 29969.m000267 30174.m008615	Glyma01g27900 Glyma03g01070 Glyma03g14180 Glyma07g07580 Glyma07g17720 Glyma10g23560 Glyma18g42580	1512 1119 873 1140 2232 1506 1347 1509 1146 1170	1035 1173 1176 1173 936 1188	1188 1050 852 738 1515 1539 594 1506	1536 678 858 675 1491 1428 1620

		At4g00400 At5g06090				1503 1503			
	Diacylglycerol cholinephosphotransferase	At3g25585	AY179560	30138.m003845	Glyma12g08720 Glyma0214210	1170	1449	1449	771 1188
	Digalactosyldiacylglycerol synthase (DGD1)	At3g11670	------	28726.m000069	Glyma03g36050 Glyma19g38720	2388	-------	2538	2352 2361
	ER Phosphatidate Phosphatase	At1g15080	-------	29586.m000620 29660.m000760 29660.m000759 29660.m000759	Glyma09g18450 Glyma10g41580 Glyma20g25650	813	-------	954 564 930 945	957 969 909
Oil body protein	Caleosin	At1g23240 At1g23250 At1g70670 At1g70680 At2g33380 At4g26740 At5g55240	AY966447 DQ140380	29673.m000932 30008.m000820	Glyma3g41030 Glyma09g22310 Glyma09g22330 Glyma09g22580 Glyma09g25350 Glyma10g33350 Glyma19g43680 Glyma20g34300	669 663 588 552 711 738 732	717 705	597 702	723 615 606 384 384 570 723 402
	Oleosin	At1g48990 At2g25890 At3g01570 At3g18570 At3g27660 At4g25140 At5g40420 At5g51210	DQ328612 S37032	29794.m003372 30147.m013891 30147.m014333	Glyma05g07880 Glyma14g15020 Glyma17g13120	510 450 552 501 576 522 600 426	699 426	564 489 495	492 492 471

Table 4.6: Gene structure of oil biosynthesis genes in different plant species

GenBank Accession number			Exon / Intron co-ordinates in sequence contigs		
Arabidopsis	**Castor bean**	**Soybean**	**Arabidopsis**	**Castor bean**	**Soybean**
At1g36180	29908.m005991	Glyma04g11550	4023 - 4127, 4317 -4490, 4572-4664, 4752 - 4937, 5021 -5071, 5155 – 5193, 5283 – 5405, 5487 - 5603, 5679 -5753, 5831 - 5887, 6013 -6234, 6412-6603, 6689 – 6877, 6959 -7069, 7169 - 7258, 7345 -7521, 7611 -7688, 7772 –7894, 7969 -8238, 8344 - 8670, 8769 -9005, 9092 -9442, 9530 -9742, 9830 -10006, 10093 -12243, 12312 - 12368	1 -351, 835 -1041, 1121 -1282, 1380 - 1472, 1573 -1668, 1814 - 1987, 2420 -2512, 2607 -2789, 2878 -2928, 3025 -3063, 4045-4167, 4287 -4403, 4543 - 4617, 4697 -4753, 4896 -4985, 5155 - 5400, 5491 -5679, 5856 -5966, 6052 - 6138, 6230 -6406, 7030 - 7107, 7216 - 7338, 7623 -7892, 7980 -8318, 8606 - 8842, 8956 - 9306, 9393 -9605, 10099 -10275, 10430 -12583, 12671 - 12715,	1 -348, 732 -938, 1017 -1178, 1397 -1489, 1578 -1673, 1755 - 1928, 2329 -2421, 2783 -2968, 3049 -3099, 3175 -3213, 4569 – 4691, 4775 -4891, 4987 -5061, 5157 -5213, 5340 -5429, 5486 – 5533, 6173 -6226, 6304 -6495, 6575 -6763, 6920 -7030, 7117 - 7203, 7289 -7465, 7866 -7943, 8044 -8166, 8380 -8649, 8768 - 9106, 9387 -9623, 9722 -10075, 10168 -10380, 11014 -11190, 11328 – 13481, 13590 -13634, 13814 -13870
At2g38040	27798.m000585	Glyma18g42280	520 -858, 960 -1034, 1169 - 1276, 1369 -1527, 1624 -1737, 1819 -1971, 2024 -2110, 2192 - 2260, 2349 -2426, 2500 - 3657	764 -1915,2100 -2177,2425 -2547, 2627 -2677,3258 -3410, 3491-3604, 3741 -3899, 4034 -4141, 5006 -5080, 5187 -5522	1 -318, 463 -537, 1202 -1309, 1438-1596, 1679 -1792, 1877-2029, 2716 -2766, 2848 -2916, 3181 -3258, 3348 -4346
AtCG00500	28890.m000006	____	730 -1365	115 - 642	
At5g35360	30185.m000954	Glyma05g36450	67 -189, 297 -629, 710 -850, 992-1090, 1182 -1238, 1442 - 1513, 1877 -1966, 2103 -2162, 2468 -2653, 2738 -2830, 2909 - 2989, 3081 -3185, 3401 -3475, 3633 -3677, 3766 -3828, 4064 - 4111	193 -246, 332 -679, 792 -932, 2133 - 2231, 2942 -2998, 4560 - 4631, 4995 -5084, 5689 -5814, 6091 -6150, 7302 -7487, 7584 - 7676, 7771 - 7851, 8689 -8793, 9588 -9662, 9928 - 9972, 10031 - 10150, 10396 -10569	91 -147, 247 -573, 688 -828, 1515 -1613, 1715 -1771, 2671 - 2742, 3437 -3526, 4260 -4319, 4886 5071, 5173 -5265, 5331 - 5411, 5565 -5669, 6250 -6324, 6643-6687, 6810 -6872, 7347 - 7457, 7861 -7920
At5g15530	29630.m000809 29929.m004560	Glyma08g03120	205 -336, 669 -785, 881 -1192, 1338 -1409, 1512 -1577, 1664 - 1687	465 – 578,1165 -1218,1316 -1465, 1563 -1862, 2470 -2541, 2692 -2757, 3013 -3033 162 -308,1702 -2001,2083 -2418, 2521 -2592, 2990 - 3055	213-272, 363-689,806 -946,1651 -1749, 1852 -1908, 2313 -2384, 3158 -3247, 3950 -4009, 4627-4812,4915 -5007,5073 - 5153, 5304 -5408, 5963 -6037, 6362 - 6406, 6529 -6591, 6759 -6806
At2g30200	30113.m001448	Glyma18g06500	131 -343, 592 -675,756 -821, 982-1122, 1293 -1340, 1486 -	4630 - 4698, 4940 -5119, 5359 - 5406, 5510 -5647, 5739 -5819, 5916 -	1 -186, 284 -367, 455 -520, 1490 -1666, 2231 -2278, 2379 - 2516,

			1623, 1708 -1788, 1874 -2050, 2137 -2223, 2311 - 2373	6092, 6166 -6252, 6365 - 6418, 6483 -6626	2618 -2698, 2776 -2952, 3030 - 3116, 3460 -3522
At5g46290	29693.m002034	Glyma10g04680	107 -571, 663 -1094, 1364 - 1537, 1647 -1760, 1860 -1961, 2081 -2146, 2232 -2297	63 -515, 1394 -1825, 2050 -2223, 3088 -3201, 3308 -3409, 3544 -3615, 3737 -3826	96 -212, 355 -582, 1112 -1219, 1304 -1375, 1534 -1620, 2123 - 2272, 2345 -2422, 2551 -2583, 2698 -2763, 2854 -2928, 3172 - 3201, 3804 -3884, 4038 -4121, 4170 -4226, 4274 -4444
At1g74960	29739.m003711	Glyma13g19010	280 -597, 991 -1191,1300- 1503, 1703 -1813, 1904 -1999, 2094 - 2138,2232-2309,2545-2613, 2751 -2834, 2947 -3099, 3175 - 3255, 3379 -3438, 3600 -3716, 3955 -4098	256 -609, 1080 -1283, 1370 -1573, 1744 -1854, 1956 -2051, 2158 -2373, 2857 - 2925, 3458 -3556, 3589 - 3693,3994 -4125,4451 -4603, 4681 - 4761, 5013 -5072, 5191 - 5307	166 -387,1104 -1211,1300 -1371, 1529 -1615, 2195 -2344, 2418 - 2495, 2632 -2664, 2775 -2840, 2932 -3006, 3740 -3802, 3911 - 4006, 4164 -4235, 4388 - 4558
At1g62640	28455.m000368	Glyma09g41380 Glyma15g00550 Glyma18g44350	275 -463, 718 -864, 952 -1086, 1176 -1280, 1365 -1481, 1676 - 1738, 1835 -2059, 2289 -2510	418 -597, 2310 -2456, 3062 -3196, 3899 - 4003, 4088 -4186, 4930 - 5037, 5124 -5345, 5501 -5722	261 -431, 655 -801, 1746 -1880, 2141 -2245, 2327 -2443, 2838 - 2900, 2989 - 3210, 3761 -3982 1 -102, 135 -212, 302 -406, 496 - 612, 712 -774, 858 -992, 1557 – 1778 1 -171, 393 -539, 1475 -1609, 1882 -1986, 2068 -2184, 2573 - 2635, 2724 -2945, 3221 -3280, 3502 -3723
At1g62610 At3g46170 At3g55290 At3g55310	30147.m013777	Glyma08g10760 Glyma18g01280	774 - 1697 1 – 867 298 - 1119 315 -1136	241 - 372, 484 -537, 1384 -1464, 1551 -1595, 1668 -1718, 1851 - 1883, 2117 -2251, 2406 -2498, 2907 -2987, 3085 -3108	226 - 471, 559 -612, 1336 -1416, 1509 -1553, 1653 -1745, 1789 - 1845, 2145 -2279, 2362 -2454, 2570 - 2650, 2839 -2871. 178-342,44-587,1230-1283, 1734 -1814,1888 -1932,2029 - 2121, 2448 -2480, 2790 -2924, 3018 - 3110, 3193 -3273, 3771 - 3803
At1g24360	29929.m004732	Glyma11g37320	239 -400,490 -630,870 - 923,1364 -1444, 1525 -1569, 1641 -1733, 2045 -2179, 2276 - 2368, 2498 -2578, 2670 - 2702	438 -581, 686 -739, 1170 -1250, 1334 -1378, 1455 -1547, 1630 -1662, 2088 -2222, 2293 -2385, 2608 -2688, 2791 -2823	343 -507, 606 -746, 1391 -1444, 1892 -1972, 2046 -2090, 2172 - 2264,2595-2627,2927-3061,3155 - 3247, 3330 -3410, 3909 -3941
At2g05990	27843.m000160	Glyma11g10770	619 -897, 976 -1083, 1163 - 1216, 1293 -1424, 1502 -1606, 1665 -1778, 2005 -2055, 2136 -	1019 -1294,1604 -1654,1965-2072, 2173 -2226, 2322 -2405, 2491 - 2538, 3855 -3911, 3972- 4058, 4159 - 4215,	791 -1102, 1582 -1701, 1869 - 1922, 2005 - 2088, 2175 -2222, 3071 -3145, 3228 -3287, 3433 -

	29650.m000277	Glyma12g03060 Glyma18g31780	2192, 2288 -2383, 2467 -2610	4453 -4548, 4810 – 4959 1 -279, 989 -1096, 1194 -1247, 1336 - 1419, 3871 -3945, 4020 - 4079, 4225 - 4272, 4364 - 4414, 4488 -4544, 4687 - 4782, 4879 -5031	3480, 3572 - 3622, 3718 -3774, 4425 - 4520, 4618 – 4770. 981 - 1256, 1740 - 1847, 2017 - 2070, 2152 - 2235, 2319 -2366, 3271 -3345, 3431 -3565, 3636 - 3683, 3774 - 3824, 3946 -4002, 4653 -4748, 4846 -4998 26 -307, 803 -910, 1034 -1087, 1180 - 1263, 1355 -1402, 1952 - 2272, 3913 -3987, 4063 - 4176, 4256 - 4306, 4371 - 4445, 4581 - 4637, 5197 -5292, 5388 - 5540
At2g22230	30200.m000354	Glyma05g24650 Glyma08g07870 Glyma08g19200 Glyma15g05800	59 -259, 340 -396, 513 -656, 1190 -1444	52 -306, 394 -450, 598 -741, 1175 - 1246	193 - 249, 377 - 514, 1669 - 1887 121 -234, 336 -479, 1507 -1758 1486 – 1704 188- 400,483 -539,628-771, 1676 -1861, 2449 -2601, 2641 - 2703
At4g30580	29687.m000572	Glyma06g28540 Glyma12g28470	40 - 81, 173 -214, 490 -561, 807 -995, 1131 -1535, 1624 -1725, 1805 -2014	278 -349, 641 - 817, 919 - 1323, 1447 - 1548, 1960 - 2175	71 -259, 695 - 802, 940 -1707, 1791 -1949, 2002 -2319, 3230 - 3331, 3663 - 3746, 4021 - 4104, 4358 - 4705, 5036 -5263 351 - 431, 570 -656, 949 - 1107, 1209 -1613, 1765 -1866, 2497 - 2736, 3088 - 3498, 3998 - 4063, 4942 -5154
At1g32200	30068.m002660	Glyma09g34110	60 -374, 772 -867, 989 -1075, 1173 -1268, 1425 -1481, 1585 - 1686, 1759 -1818, 1919 -2026, 2132 - 2209, 2285 -2404, 2500 - 2613, 2704 -2838	1 -303, 798 -893, 1685 -1771, 1965 - 2048, 2959 -3015, 4623 -4724, 4818 - 4877, 4957 - 5064, 6553 -6630, 8337 - 8456, 8560 -8688	20 -367,1318 -1413, 2257 -2343, 2511 - 2606, 4565 - 4621, 5326 - 5427, 5524 - 5583, 5694 -5801, 6193 - 6270, 7904 -8023, 8096 - 8209, 8931 - 9065
At3g15850	29841.m002863	Glyma07g03370 Glyma08g22730	79 -603, 694 -834, 922 -1062, 1590 -1694, 1790 -1990	1 -573, 726 -866, 958 -1098, 1816-1920, 2019 - 2216	1 -570, 654 -794, 883 -1023, 1193 -1225, 1397 -1609. 49 -618, 713 -853, 952 -1092, 1251 -1355, 1448 -1660
At1g43800 At2g43710 At3g02610 At3g02630 At3g02620	27985.m000877 28470.m000428 29929.m004515 30020.m000203	Glyma02g15600 Glyma07g32850 Glyma13g08970	53 -364, 547 -840, 1371 -1940 146 -283, 1394 -1897,1974-2534 758 -1165, 1258 -1833 135 -254, 847 -1350,1759 2199, 2373 -2495 407 - 814, 930 -1505	68 - 676, 772 -1338 1 - 99, 331 – 564 1 -84, 174 -677, 836 -1204 27 -149, 3445 - 3948, 4628 - 5188	162 -269, 2014 -2517,2939-3499 90 -197,1996 -2499, 2923 -3483 105 -233, 554 -748, 1055 -1087, 1324 - 1434, 1729 -1803, 1884 - 1955, 2259 -2333, 2374 - 2505, 2725 - 3081

At5g16230 At5g16240		Glyma14g27990	47 -172,1445 -1951, 2028 -2597 175 -294, 1426 -1923, 2023 -2463, 2711 - 2833		496 - 939, 1787 -2356
At4g30950	29696.m000105	Glyma02g36460	122 -160, 287 -349, 625 - 870, 949 -1092, 1166 -1315, 1401 - 1520, 1622 - 1798, 1892 -2008, 2098 - 2211, 2353 - 2517	149 - 202, 769 -912, 980 - 1063, 1664 - 1807, 3287 - 3403, 3901 -4014, 4288 - 4449	105 -143, 241 -294, 781 -1017, 1473 -1616, 1772 -1921, 2729 - 2848, 3018 - 3194, 4040 -4156, 4441 - 4554, 4889 -5050
At3g11170	28176.m000273 29681.m001360 29814.m000719	Glyma01g29630 Glyma03g07570 Glyma07g18350	159 - 650, 954 -1040, 1125 - 1190, 1281 -1373, 1453 -1638, 1732 -1812, 1871 - 2326	1 -519, 627 -713, 803 -868, 981 - 1073, 1162 -1347, 1453 -1533, 1694 - 1831, 1966 -2148 260 - 544, 633 -719, 822 -881, 977 - 1069, 1196 -1381, 1469 -1549, 1674 - 1811, 1884 – 2078 345 -878, 1073 -1159, 1286 -1351, 1460 -1552, 1635 -1820, 2041 -2121, 2214 - 2351, 2447 -2638	651 -1166, 1318 -1404, 1536 - 1601, 1710 -1802, 1906 -2091, 2186 - 2266, 2648 -2785, 2885 - 3070 688 -1206, 1371 -1457, 1594 - 1659, 1770 -1862, 1965 - 2150, 2266 -2346, 2739 -2876, 2982 - 3167 443 -961,1094-1180,1494 -1559, 1672 -1764, 1859 -2044, 2137 - 2217,2323 -2460,2557 - 2742
At5g05580	28176.m000273 29681.m001360 29814.m000719	Glyma01g29630 Glyma03g07570 Glyma07g18350	198 -668, 766 -852, 970 -1035, 1137 -1229, 1315 -1500, 1610 - 1690, 1788 -1925,2033 - 2212	1 -519, 627 -713, 803 -868, 981 - 1073, 1162 -1347, 1453 -1533, 1694 - 1831, 1966 – 2148 260 - 544, 633 -719, 822 -881, 977 - 1069, 1196 -1381, 1469 -1549, 1674 - 1811, 1884 – 2078 345 -878, 1073 -1159, 1286 -1351, 1460 -1552, 1635 -1820, 2041 -2121, 2214 - 2351, 2447 -2638	651 -1166, 1318 -1404, 1536 - 1601, 1710 -1802, 1906 -2091, 2186 - 2266, 2648 -2785, 2885 - 3070 688 -1206, 1371 -1457, 1594 - 1659, 1770 -1862, 1965 - 2150, 2266 -2346, 2739 -2876, 2982 - 3167 443 -961,1094 -1180,1494-1559, 1672 -1764, 1859 -2044, 2137 - 2217, 2323 -2460, 2557 - 2742
At3g12120	28035.m000362	Glyma03g30070 Glyma09g17170 Glyma10g42470 Glyma19g32940 Glyma20g24530	1308 - 2459	2529 - 3692	4696 - 5847 1 - 1161 426 - 1565 3204 - 4355 1213 - 2352
At2g29980	29681.m001360	Glyma01g29630 Glyma07g18350	108 -413, 1040 -1126, 1676 - 1741, 2221 -2313, 2401 -2586, 2675 - 2755, 2834 -2971, 3070 - 3267	260 - 544, 633 -719, 822 -881, 977 - 1069, 1196 -1381, 1469 - 1549, 1674 - 1811, 1884 - 2078	651 -1166, 1318 -1404, 1536 - 1601, 1710 -1802, 1906 -2091, 2186 -2266, 2648 -2785, 2885 - 3070 443-961,1094-1180,1494- 1559, 1672-1764,1859-2044,2137-

					2217, 2323 -2460, 2557 - 2742
At3g25110 At4g13050	30217.m000262	Glyma08g46360 Glyma18g36130	214 -546, 647 -778, 886 -996, 1144 -1236, 1318 -1404, 1489 - 1557, 1644 -1898 179 -508, 630 -761, 838 -948, 1109 -1201, 1279 -1434, 1523 - 1795	64 -336, 988 -1119, 1214 -1324, 1686 - 1778, 2550 - 2705, 3143 -3286, 3374 - 3442, 3523 -3801	64 -408, 1236 -1373,1412- 1555, 2374 -2466, 2814 -2843, 4173 - 4256, 4432 - 4500, 4852 -5133. 114 - 455, 1304 -1435, 1513 - 1623, 2248 - 2340, 4437 - 4523, 4684 - 4752, 5976 - 6257
At1g08510	29842.m003515 29848.m004677	Glyma04g21910 Glyma05g08060 Glyma06g23560 Glyma17g12940	969 -1466, 1548 -1679, 1910 - 2020, 2251 -2421, 2512 -2832	1306 -1806, 1928 -2059, 3016 - 3126, 3218 - 3388, 3472 -3540, 4684 – 4953 1306 -1806, 1928 -2059, 3016 -3126, 3218 - 3388, 3472 -3540, 4684 - 4953	807-1331,1670-1801,2446-2556, 2932-3099,3184-3252,4041-4361 1022 -1513, 1683 -1814, 2343 - 2453, 2853 -3023, 3130 - 3198, 3280 – 3549 945 -1436, 1849 -1980, 2463 - 2573, 2944 -3111, 3207 - 3275, 3345 - 3506, 4163 -4444 1356-1847,2021-2152,2673 - 2783, 3176 -3346, 3453 - 3521, 3604 -3873
At2g19450	29912.m005373	Glyma13g16560	231 -620, 705 -785, 877 -948, 1054 -1128, 1175 -1291, 1376 - 1498, 1579 -1656, 1735 -1836, 1937 -2071, 2219 -2299, 2509 - 2571, 2655 -2702, 2826 -2888, 2970 -3020, 3146 -3250	237 -614, 934 -1014, 1145 -1216, 1307 -1381, 1504 -1569, 2002 -2124, 2446 - 2541, 3174 -3275, 3376 -3510, 4017 - 4097, 4922 - 4984, 5953 - 6240, 6322 - 6363, 6916 - 6978, 7131 -7184, 7425 - 7523	28 - 327, 674 -760, 1292 -1357, 1755 -1877, 2143 -2220, 2378 - 2422, 2594 -2731, 3660 - 3761, 3897 - 4031, 4698 - 4778, 4975 - 5022, 5242 -5301, 5373 -5444
At3g51520	29682.m000581	Glyma09g32790	144 -290, 379 -537, 628 -732, 817 -972, 1177 -1263, 1421 - 1516, 1614 -1700, 1825 -1923	413 -460, 591 -674, 1156 -1260, 1357 - 1512, 2244 -2330, 2589 -2684, 2893 - 2979, 3052 -3147	234 -446, 557 -628, 773 -856, 2230 -2334, 2418 -2573, 3627 - 3782, 5363 - 5449, 5866 -5967
At1g01610 At1g51260 At1g78690 At1g80950 At2g27090	27810.m000646 29851.m002448 30169.m006433 30170.m013990 29736.m002070 29822.m003441	Glyma01g27900 Glyma03g01070 Glyma03g14180 Glyma07g07580 Glyma07g17720 Glyma10g23560 Glyma18g42580	86 -394, 476 -790, 1245 - 1718, 1984 – 2391 130 -210,387 -509,592 -630, 711 -782, 862 -921,1009 -1086,1176 -1334, 1418 -1543, 1650 -1730, 1813 -1899, 1973 -2182. 330 -554, 663 -791, 973 -1056, 1138 -1275, 1465 -1755 179 -694,1032 -1157,1293-1328, 1467 -1529, 1735 -1800, 1883 - 1957, 2116 -2256, 2354 -2461. 1512 -2558, 2721 -2789, 2870 - 3082, 3176 -3397, 3495 -4166.	267 -347, 585 -707, 1628 -1666, 1802 - 1873, 2111 -2170, 2581 -2658, 2809 - 2967, 4289 - 4414, 5324 - 5416, 5499 - 5585, 5707 -5973 1 -570, 727 -948, 1120 -1155, 1362 - 1577 1 -81,389 -511,1013 -1051,1190 - 1261, 1744 -1803, 2034 -2111, 2198 - 2356, 2667 - 2792, 3274 -3366, 3456 - 3473. 1 -351, 532 -660, 1556 -1807 1 -615, 806 -1705	1 -615, 1588 -1638, 2657 -3523 71 -376, 454 – 822 2915 – 3772 68 -373, 441 -752, 872 -925 70 -693, 2526 – 3392 37 - 216, 891 -1271, 2009 -2875 11 -763, 2559 -3425

At2g38110 At3g05510 At3g11430 At3g18850 At3g57650 At4g00400 At5g06090	29969.m000267 30174.m008615		131 -766, 1810 – 2679 183 - 407, 698 - 814, 901 - 966, 1067 -1264, 1366 -1482, 1584 - 1688, 2169 -2681 147 - 734, 994 – 1914 382 -489, 613 -1071,1155 -1283, 1391 -1831 245 -322, 690 -812, 922 - 960, 1078 -1149, 1445 -1504, 1725 - 1802, 1963 - 2121, 2328 -2453, 2697 - 2789, 2922 -3008, 3082 - 3333 120 - 428, 504 -815, 1318 -1791, 2141 -2542 1 -582, 697 -1617	1 -639, 1799 -2698 62 – 655 47 - 355, 508 -819, 2296 - 3177	
At3g25585	30138.m003845	Glyma12g08720 Glyma0214210	484 - 618, 723 -731, 822 - 872, 961 -1038, 1140 -1337, 1426 - 1518, 1602 -1667, 1772 -1837, 1951 -2013, 2111 -2149, 2231 - 2314, 2387 - 2449, 2528 -2584, 2660 - 2782, 2879 -2899	617 -763,1163 -1201,1398 -1448, 1554 -1631, 1742 -1816, 1916 -1936, 2014 - 2112, 2308 - 2400, 2563 - 2628, 3465 - 3530, 4049 - 4111, 4192 - 4230, 4519 - 4563, 4748 - 4867, 5466 -5513, 5955 - 6017, 6491 - 6556, 6878 - 6949, 7052 - 7105, 7463 -7504, 7712 -7783	10916 -10918, 11025 -11075, 11373 -11450, 12068 -12142, 12238 -12258, 12341 -12439, 13118 -13246, 13386 -13442, 14179 -14244, 14457 -14567, 15148 -15216 2309 -2311, 2418 -2468, 2789 - 2866, 3444 - 3518, 3614 - 3634, 3712 -3810, 4024 - 4116, 4218 - 4283, 4698 - 4751, 5414 -5464, 5604 - 5672, 7426 -7542, 7618 - 7656, 7760 -7843, 8571 -8633, 9393 - 9449, 9545 - 9595, 9724 - 9792, 10003 -10023
At3g11670	28726.m000069	Glyma03g36050 Glyma19g38720	103 -759, 955 -987, 1102 -1188, 1579 -2058, 2159 - 2425, 2527 - 2874, 2960 - 3469	1 -639, 1330 -1359, 2004 -2042, 2222 - 2311, 2378 -2473, 3017 -3499, 3607 - 3873, 4054 - 4401, 4794 - 5327	65 - 658,1375 -1407, 1973-2059, 2924 - 3406, 3499 - 3765, 4132 - 4479, 4829 -5362 72 - 677,1376 -1408,1941- 2027, 2927 - 3409, 3494 -3760, 4123 - 4470, 4813 -5319, 5503 -5526
At1g15080	29586.m000620 29660.m000760 29660.m000759	Glyma09g18450 Glyma10g41580	455 -670, 759 - 845, 937 -1044, 1123 -1215, 1300 - 1485, 1738 - 1851	1 -216, 317 -403, 519 -626, 698 -787, 1002 - 1187, 1306 -1362, 2631 -2828. 6754 -6780, 7063 -7155, 7219 -7404, 7495 - 7551, 8141 -8332. 1 -60,671 -895,1002 -1088,1182 -	859 -1074, 1184 -1270, 1366 - 1473, 1572 -1664, 2413 - 2598, 2689 - 2745, 3841 - 3885, 5352 - 5501. 1797 -2012, 2139 -2225, 2314 -

	29660.m000759	Glyma20g25650		1289, 1420 -1509, 1715 -1900, 2200 - 2256, 2378 – 2482 1 -255, 374 -460, 571 -678, 801 -893, 1087 - 1260, 1385 -1441, 1606 -1764	2421, 2543 -2635, 3441 -3626, 3711 - 3767, 4896 -5105. 1926 -2141, 2269 - 2355, 2439 - 2546, 2667 - 2759, 3573 -3758, 4745 – 4954.
At1g23240	29673.m000932	Glyma3g41030	96 -185, 476 -574, 725 - 808, 1078 -1170, 1304 -1426, 1565 - 1732.	1 -39, 128 -223, 323 -406, 2037 - 2129, 2363 - 2485, 2598 -2747	235 -345, 536 -682, 856 -939, 1038 -1130, 1213 -1335, 1431 - 1583.
At1g23250	30008.m000820	Glyma09g22310	155 - 247, 460 -561, 671 -757, 939 -1025, 1129 -1251, 1374 - 1538.	1 -105,550 - 696,796 -879,1264 - 1356, 1461 -1583, 1674 -1811	91 -129, 396 -512, 1362 -1445, 2579 - 2671, 2822 -2944, 3060 - 3206.
At1g70670		Glyma09g22330	249 -269, 355 -453, 649 -732, 831 - 923, 1004 -1126, 1234 - 1389		1 -30, 401 -517, 787 -870, 1550 - 1642, 1724 -1846, 1976 -2122.
At1g70680		Glyma09g22580	238 - 336, 577 -660, 745 - 837, 949 -1071, 1179 -1322.		4232 - 4234, 4316 -4357, 4771 - 4863, 5379 -5468, 5587 -5733.
At2g33380		Glyma09g25350	98 -199, 477 -623, 710 -793, 890 - 982, 1078 -1200, 1339 -1488.		4232 - 4234, 4316 -4357, 4771 - 4863, 5379 -5468, 5587 -5733.
At4g26740		Glyma10g33350	66 -182, 505 -651, 744 - 827, 952 -1044, 1138 -1260, 1390 - 1551		72 -179, 730 -876, 993 - 1076, 1376 -1468, 1571 -1699.
At5g55240		Glyma19g43680	38 -151, 243 -389, 495 -578, 663 - 755, 855 -977, 1075 -1233		85 -195, 410 -556, 738 -821, 919 - 1011, 1087 -1209, 1316 -1468.
		Glyma20g34300			621 -728, 824 -907, 1441 -1533, 1860 -1970
At1g48990	29794.m003372	Glyma05g07880	37 – 546	1 -699	241 - 732
At2g25890	30147.m013891	Glyma14g15020	56 – 505	1 -426	241 - 732
At3g01570	30147.m014333	Glyma17g13120	50 -157, 233 -676	87 -581	49 - 519
At3g18570			35 – 535		
At3g27660			97 -249, 645 -1067		
At4g25140			69 - 419, 663 -830		
At5g40420			44 - 202, 722 – 1162		
At5g51210			71 - 373, 487 -606		

Table 4.7: *In silico* expression status of fatty acids biosynthesis genes

Tissue	Gene	GenBank Accession no.			
		Arabidopsis	*B.rapa*	Soybean	Castor Bean
Seeds	Alpha carboxyltranferase				27798.m000585
	Enoyl ACP reductase				27843.m000160
	Stearoyl desaturase		X74782	Glyma13g08970	27985.m000877
	FAD-2			Glyma10g42470	28035.m000362
	ERPhosphatidate Phosphatase				29660.m000760
	DGAT 2			Glyma17g06120	29682.m000581
	FatB			Glyma17g12940	29842.m003515
	Oleosin Oleosin Oleosin	At5g40420	S37032	Glyma14g15020 Glyma17g13120	30147.m014333 30147.m013891 29794.m003372
	Hydroxyacyl ACP dehydrase				30200.m000354
	Caleosin	At5g55240		Glyma20g34300	
Leaves	FatB				29848.m004677
	LPAAT			Glyma07g07580	
	3-Ketoacyl- acp-dehydrase	At3g55290 At3g55310			
Flowers	DGD1				28726.m000069
	Beta- carboxyl transferase				28890.m000006
	ACCase				29908.m005991
Roots	Stearoyl desaturase	At3g02620			
Roots+ flowers	Stearoyl desaturase	At3g02610			
Seed+ flowers	Oleosin	At1g48990			
Leaves+ flowers	FAD 7	At3g11170			
	Oleosin	At2g25890 At3g18570			
	Caleosin	At1g23240 At1g23250 At4g26740			

4.3.2) Comparative analysis of gene structures in different plant species

Comparative genomics of fatty acid biosynthesis genes was done to understand as what determines differences, if any, for variations in contents and compositions of fatty acids in different plant species. The gene structure analysis revealed that the exon-intron structure of fatty acid biosynthesis genes in Brassica was almost identical to Arabidopsis homologs, whereas castor bean genes shared structure similarity with the soybean fatty acid biosynthesis genes. However insertion, deletion and intron size variations were found in castor bean, soybean and Brassica genes with reference to Arabidopsis.

Conversion of acetyl Co-A to malonyl Co-A by acetyl carboxylase (ACCase) is the most committed step in fatty acid biosynthesis. Exon/intron number and CDS length for ACCase gene was almost same between castor bean (31 exons) and soybean (33 exons), whereas slightly less in Arabidopsis (26 exons). Comparative structural analysis revealed that homomeric ACCase gene from Arabidopsis (1-26 exons) showed microsynteny with castor bean (6-31 exons) and soybean (6-33 exons), with a 3bp deletion in 8^{th} and 26^{th} exons of castor bean, 3 bp deletion and 3 bp insertion in 31^{st} and 29^{th} exons of soybean and a 2bp insertion in 24^{th} and 26^{th} exons of castor bean and soybean, respectively. First five exons of homomeric ACCase in castor bean and soybean (missing in Arabidopsis) showed colinearity for exon size, with the exception of a 3 bp insertion in the first exon of castor bean gene. Sixteenth exon of ACCase showed sequence identity to 3 exons (16^{th}, 17^{th} and 18^{th}) of soybean (Fig. 4.7).

Two distinct classes of thioesterases, FatA and FatB are responsible for release of fatty acids from ACP by thioesterases. FatA gene structure was most diverse with exons number varying from 5 to 11 among four plant species. Two homologs of FatA gene were present in Arabidopsis and soybean whereas FatB gene had 2 homologs in castor bean and 4 homologs in

soybean. The first exon of FatB gene had an insertion of 3 bp in castor bean and 27 bp insertion in one of soybean homologs (Glyma0421910) and other three homologs of soybean had 6 bp deletion compared to Arabidopsis (Fig. 4.8). An insertion of one exon of 69 bp was present in FatB gene of castor bean and soybean which was absent in Arabidopsis. The last exon of FatB (5^{th} exon) in Arabidopsis showed homology to the last exon (6^{th} exon) of one of the homologs of soybean (Glyma04g21910); whereas last exon of castor bean showed homology to the last exon of other two homologs of soybean (Glyma05g08060 and Glyma17g12940).

Stearoyl ACP desaturase gene had maximum number of homologs (6 in Arabidopsis, 3 in Brassica, 4 in soybean, and 4 in castor bean) in fatty acid desaturase category of enzymes. Oleoyl deasturase (Fad2) and Linoleate desaturase (Fad3) genes showed more relatedness in relation to number and sizes of exons and introns in each homolog among four plant species. Oleoyl desaturase (FAD 2) had only one exon in Arabidopsis, castor bean and soybean with an insertion of 12 bp in the exon of castor bean and 9 bp insertion in the exon of one homolog of soybean (Glyma09g17170). FAD 3 gene structure was conserved with respect to exon–intron number and size between Arabidopsis, castor bean and soybean except for first and last exons. A 21 bp deletion in the first exon of castor bean (29681.m001360) and an insertion of 210 and 213 bp was observed in the first exon of castor bean and two homologs of soybean (Glyma01g29630 and Glyma07g18350), respectively. Two deletions of 3 and 12 bp were observed in the last exon (8^{th} exon) of castor bean and soybean, respectively. A deletion of 6 bp was observed in the 3^{rd} exon of FAD 3 of castor bean. An SNP (G→A) was also identified at the exon- intron junction of FAD 3 gene in the 3^{rd} exon of one homolog of soybean (Glyma01g29630) w.r.t. castor bean, Arabidopsis and other homologs of soybean (Fig. 4.9).

The DGAT gene involved in TAG (Tri-acyl Glyceride) synthesis has two isoforms, DGAT-1 and DGAT-2. These two genes showed variation in number and sizes of exons and introns. DGAT-1 gene had 15 exons in Arabidopsis, 13 exons in castor bean and 16 exons in soybean. DGAT-2 had 8 exons in Arabidopsis & castor bean and 7 exons in soybean. The detailed comparative genomics of fattty acids biosynthesis genes in 4 oil seed plant species provided insights to undertake identification and utilization of castor bean fatty acid biosynthesis genes and sequence variations for the development of candidate gene markers in Jatropha.

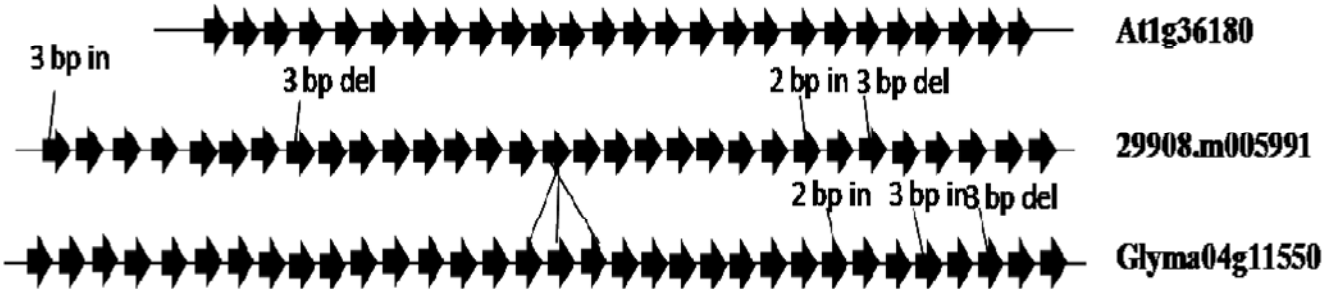

Fig. 4.7: Structure of ACCase gene in Arabidopsis (26 exons), castor bean (31 exons) and soybean (33 exons); thick arrows and thin lines represented exons and introns, respectively. Arabidopsis 1-26 exons showed identity to 6 to 31st and 6 to 33rd exons of castor bean and soybean, respectively; 16th exon of castor bean showed identity to three exons of soybean (16th, 17th and 18th). A 3 bp deletion (del) in the 8th and 26th exons of castor bean, 3 bp deletion and 3 bp insertion (in) in the 31st and 29th exons of soybean and a 2 bp insertion in the 24th and 26th exons of castor bean and soybean, respectively. At1g36180: Arabidopsis ACCase gene; 29908.m005991: Castor bean ACCase gene; Glyma04g11550: Soybean ACCase gene

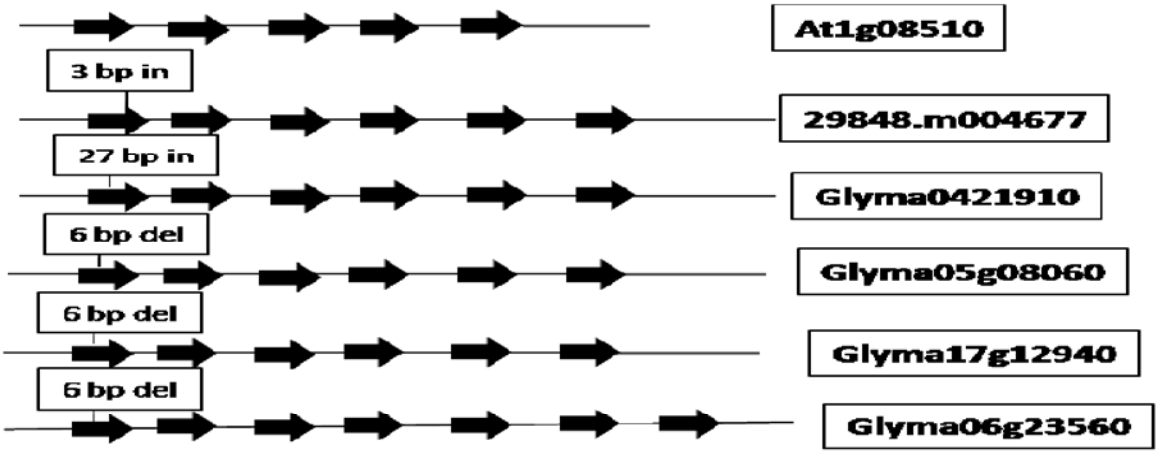

Fig. 4.8: Structure of FatB (Palmitoyl thioesterase) gene in Arabidopsis (At1g08510), castor bean (29848.m004677) and four soybean homologs (Glyma0421910, Glyma05g08060, Glyma17g12940, Glyma06g23560). The 5th exon of FatB in Arabidopsis showed homology to the 6th exon of one of the homologs of soybean (Glyma04g21910); whereas 6th exon of castor bean showed homology to the 6th exons of other two homologs of soybean (Glyma05g08060 and Glyma17g12940).

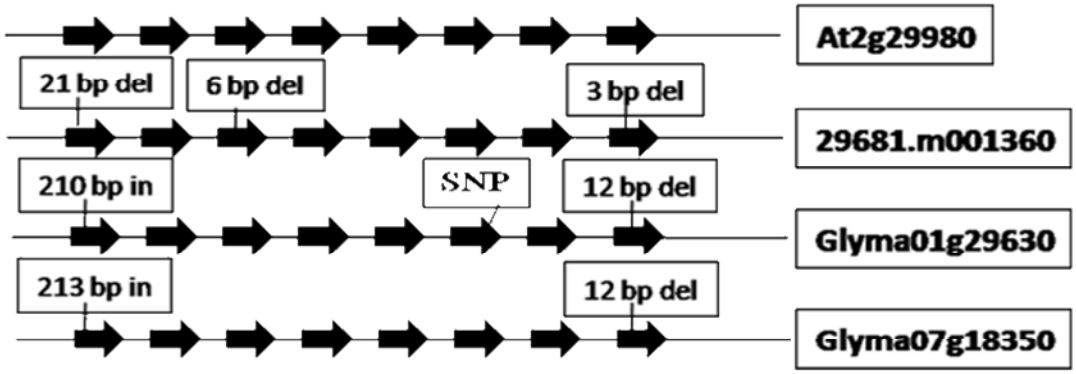

Fig. 4.9: Structure of FAD 3 (Linoleoyl desaturase) gene in Arabidopsis (At2g29980), castor bean (29681.m001360) and two soybean homologs (Glyma01g29630, Glyma07g18350). Exon/ intron numbers are conserved in FAD 3 while variation in sizes was observed in the first and last exons. SNP identified in the 6th exon of soybean homolog (Glyma01g29630) was reported to be associated with low linolenic acid content (Bilyeu *et al.*, 2005).

4.4) Identification of candidate gene markers in *J. curcas*

Thirty two critical genes involved in fatty acid biosynthesis were identified in castor bean. The coding DNA sequence (CDS) length for candidate genes varied from 326 bp (one homolog of Stearoyl –ACP desaturase) to 6723 bp (ACCase). The exon number varied from 1 to 31 (ACCase). The complete FGenesH annotation, including gene length, gene CDS, castor bean genome sequence contigs, intron-exon coordinates, etc. are given in Table 4.6. There was ~90% identity for critical candidate genes with the available fatty acid biosynthesis genes sequences of *J. curcas*. Out of 48 primer pairs designed to check the transferability of exons, introns and 5'& 3'UTR regions of these candidate genes, 32 amplified both on castor bean and Jatropha whereas 10 primer pairs amplified only in castor bean. All the candidate gene regions amplified successfully in *J. curcas* with maximum amplification (~ 77%) from coding regions (exons) followed by intronic (~ 76%), 5' and 3' UTR regions (~ 44%). Eleven primer pairs (7 from exon-intron junctions, 2 from UTRs and 2 from exonic regions) showed polymorphisms in 54 genotypes of *J. curcas*. Seven primer pairs from introns of ACCase (6^{th} and 11^{th} exon-intron junctions), Linoleoyl desaturase (2^{nd} exon-intron junction), FatA (1^{st} exon-intron junction), KAS III (2^{nd} exon-intron junction), Stearoyl desaturase (2^{nd} exon-intron junction), KAS II (1^{st} exon-intron junction) genes were polymorphic among *J. curcas* genotypes (Fig. 4.10). Second exon of Linoleoyl desaturase and DGAT 2 showed polymorphisms in *J. curcas* genotypes. Out of 11 polymorphic primer pairs from candidate genes, one amplicon from exon-intron junction of Stearoyl desaturase gene (Fig. 4.11) named as JJM1 (Jaypee Jatropha Marker1) showed amplification only in 36 high oil content genotypes, whereas no amplicons in 11 low oil content genotypes of *Jatropha curcas* (Fig. 4.12). Sequencing of Stearoyl desaturase amplicon (1.3 kb) encompassing the JJM1 (700 bp fragment) showed 2 SNPs (TT→AA) in the exonic region and an insertion of 5 bp (AAAAC) in the intron of low oil content genotypes (Fig 4.13). Comparison

of sequence from low and high oil content genotypes with the Stearoyl desaturase DNA sequence of *J. curcas* in the GenBank (DQ084491) showed functional polymorphism in the low-oil content genotypes. SNPs identified in the exonic region lead to a substitution of leucine with glutamine in Stearoyl desaturase polypepetide of low oil content genotypes (Fig. 4.14).

Twelve primer pairs were designed to amplify SSRs identified in the candidate genes, out of which 7 amplified in castor bean and *J. curcas* whereas 3 amplified only in castor bean. One SSR primer pair with (AG)15 repeat from 5'UTR of Stearyol desaturase showed polymorphism in Jatropha genotypes (Fig. 4.15), but there was no association with the high oil content genotypes.

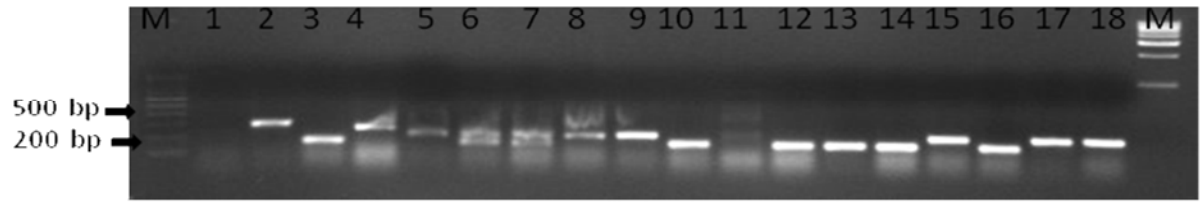

Fig. 4.10: Polymorphisms in amplicon size for 2nd intron of Linoleoyl desaturase in Jatropha genotypes. Numerical represent 18 Jatropha genotypes; M: Molecular weight marker of 100 bp

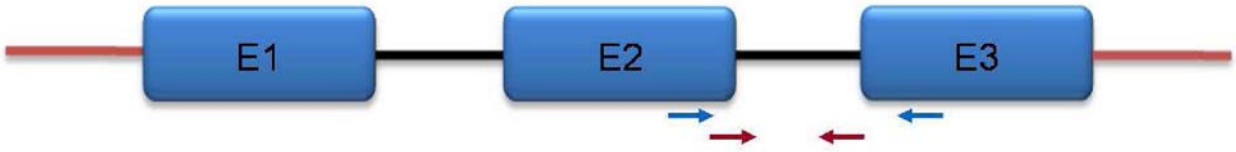

Fig. 4.11: Location of primers designed to amplify (red arrows; 700 bp) (JJM1) and sequence (blue arrows; 1.3 kb) parts of Stearoyl desaturase gene associated with high oil content in Jatropha

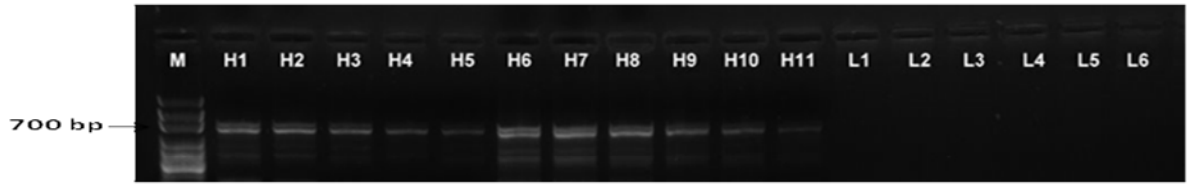

Fig. 4.12: Association of JJM1 marker from exon- intron junction of Steaoryl desaturse gene with high oil content in *J. curcas*; PCR amplification only in high oil content (>35%) genotypes (H1-H11) whereas, no amplification in low oil content (< 30%) genotypes (L1-L6); M: Molecular weight marker of 100 bp

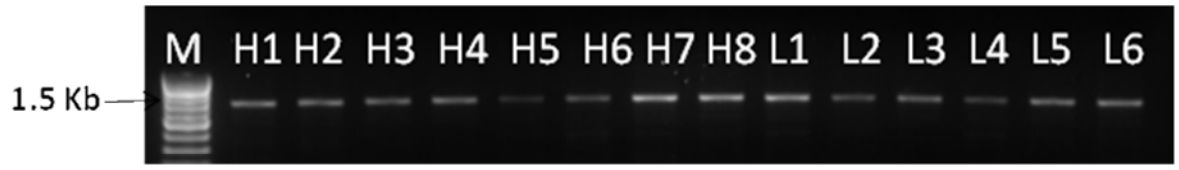

Fig. 4.13: Monomorphic amplification for JJM1 encompassing region (1.3 kb) of Steaoryl desaturse gene in high (H1-H8) and low oil content genotypes (L1-L6); M: Molecular weight marker of 100 bp

```
DQ084491:  714 GAAACACTGCCAGACTTGCCAAAGAACATGGAGACATAAAGTTGGCTCAAA
J. curcas_H: 633 GAAACACTGCCAGACTTGCCAAAGAACATGGAGACATAAAGTTGGCTCAAA
J. curcas_L: 633 GAAACACTGCCAGACAAGCCAAAGAACATGGAGACATAAAGTTGGCTCAAA
```

Fig. 4.14: DNA sequence variations in JJM1 region of high versus low oil content genotypes of *J. curcas*; DQ084491: GenBank accession of *J. curcas* Stearoyl desaturase; *J. curcas*_H: High oil content (> 35 %); *J. curcas*_L: Low oil content (< 30%)

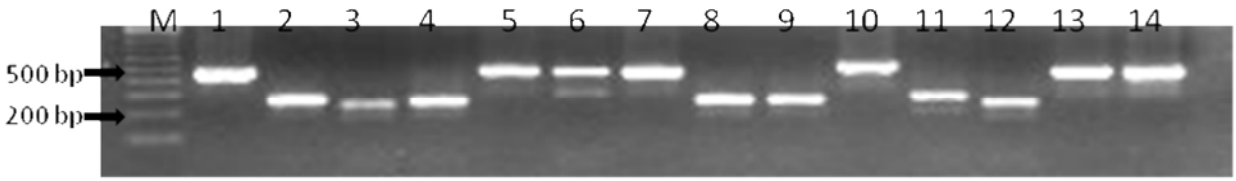

Fig 4.15: Polymorphisms among 14 Jatropha genotypes in amplicon size for SSR (AG)15 identified in the 5'UTR of Steaoryl desaturse gene; M: Molecular weight marker of 100 bp

4.4.1) Genetic structure

We used the model-based method approach of Pritchard *et al.* (2000) for the gene specific marker sets ability to detect population structure and assign individuals to populations. Analysis of population structure using STRUCTURE software provides correct number of subpopulations for association mapping. For all data sets, likelihood increased most when k (the number of populations in the model) was increased from two to three; results were very consistent across runs with k= 3 but became less consistent at higher values of k. Interestingly, the percent of individuals assigned to populations did not continue to increase with k, as might be expected: maximal assignment occurred at k =3 and then decreased steadily, suggesting the genetic diversity should be structured into subpopulations in this sampled collection. The three Ks structuration corresponded to the genetic populations revealed by Principal components analysis (PCA) (Fig. 4.16). PCA analysis of 12 marker loci suggested that 11 principal components have been extracted (Table. 4.8). The high oil content genotypes were structured in K1 and low oil content genotypes were structured in K2, and K3. This validated the results from PCA including population structure and PC- matrix for following association mapping. For these K populations Fst of K1 was 0.383 with Fst of K2, 0.221 with K3 and others.

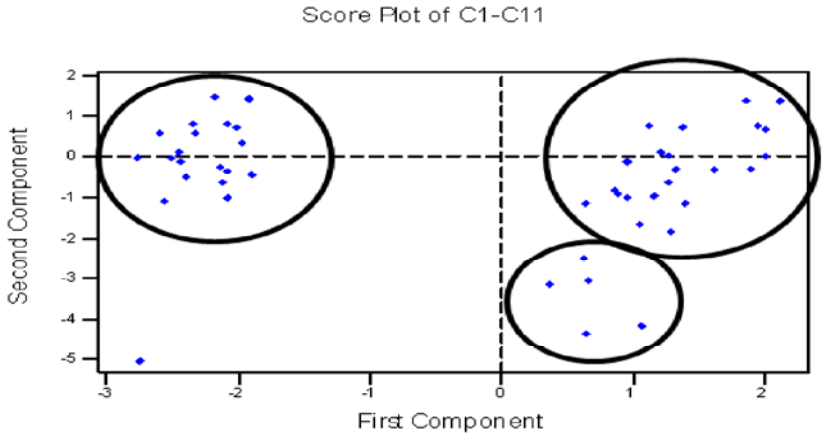

Fig. 4.16: PCA for 54 *J. curcas* accessions genotyped with 11 molecular markers

Table 4.8: Factor loadings from PCA based data

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	mean	St Dev	r^2
Marker 1	0.539	-0.112	-0.012	-0.116	0.038	-0.045	-0.019	-0.042	0.106	0.707	-0.41	0.6154	0.4892	0.299
Marker 2	0.539	-0.112	-0.012	-0.116	0.038	-0.045	-0.019	-0.042	0.106	-0.71	-0.41	0.6154	0.4892	0.004
Marker 3	0.539	-0.112	-0.012	-0.116	0.038	-0.045	-0.019	-0.042	0.106	0.056	0.816	0.6154	0.4892	0.564
Marker 4	0.106	-0.008	0.319	0.677	0.344	-0.01	0.45	-0.325	0.048	0.322	0	0.8022	0.4005	0.663
Marker 5	0.167	0.34	0.008	0.5	0.02	-0.312	-0.397	0.587	0.079	0.068	0	0.8022	0.4005	0.748
Marker 6	-0.005	0.497	-0.271	0.01	0.233	-0.018	-0.46	-0.634	0.107	-0.019	0	0.9231	0.2679	0.823
Marker 7	0.029	0.364	-0.12	-0.336	0.703	0.097	0.313	0.331	-0.168	0.241	0	0.9231	0.2679	0.891
Marker 8	-0.283	-0.241	0.181	-0.229	0.302	-0.591	-0.038	-0.015	0.578	0.04	0	0.8242	0.3828	0.955
Marker 9	0.04	0.272	-0.532	0.027	-0.36	-0.463	0.052	-0.061	0.056	-0.21	0	0.8132	0.3919	0.91
Marker 10	0.29	0.499	0.376	-0.149	-0.29	0.357	0.185	0.073	0.577	0.131	0	0.6703	0.4727	0.92
Marker 11	0.08	0.294	0.594	-0.25	-0.16	-0.442	-0.005	-0.153	-0.497	0.105	0	0.6593	0.4766	0.92

4.5) Association analysis of candidate gene marker with high oil content in Jatropha genotypes

Extent of genome-wide linkage disequilibrium (LD) was evaluated through pairwise comparisons among 12 marker loci. The r^2 averaged 0.4952 ± 0.0013 ranging from 0.0004 to 0.76. One marker locus maintained a strong LD index with each other (D' > 0.78 and r^2 > 0.76). The r^2 value corresponding to the 95th percentile of the distribution was 0.0152 and this value was used as a population specific threshold for r^2 as evidence of genetic linkage. Based on the criterion P-value less than 0.0001 with 12 markers for high and low oil traits, we identified one significant marker-trait association for JJM1 (p=0.00013) with high oil content in *J. curcas*.

4.6) Identification of NBS-LRR genes in castor bean genome

For identification and characterization of NBS-LRR genes in castor bean genome, 4,655 NBS proteins were downloaded from the NCBI and used as query sequences against 31,221 castor bean proteins sequences available at the TIGR. Three hundred and sixty five proteins were selected from castor bean on the basis of 50% identity and subjected to BlastP against NCBI database out of which 220 true NBS- LRR proteins were identified. The classification of 220 NBS-LRR proteins into domains was done by using Hidden Markov Model (HMM) searches

using Pfam. The NBS-LRR proteins were categorized into CNLs (N-terminal coiled-coil motif) and TNLs (N-terminal domain with Toll/Interleukin-1 Receptor homology) by using PCOILS. From 220 NBS proteins, 113 were CNLs and 107 were TNLs. Details of motifs of CNLs and TNLs in the 220 NBS-LRR proteins are given in Table 4.10. The *in silico* expression status of identified NBS-LRR genes revealed that 51 showed identities to castor bean ESTs.

4.6.1) Organization of NBS-LRR genes in the castor bean genome

Two hundred and twenty NBS-LRR genes identified in the castor bean genome were distributed among 136 sequence contigs. One hundred and thirty six NBS-LRR genes (~61.4 %) were in clusters of varying sizes (2-11 genes). Most of the clusters comprised 2 to 3 genes except for sequence contigs 29758, 30147, 29838, 30131, 29662, 29666, 30190, 29757, , 29841, 29910, 30074, consisting of clusters of 4, 4, 5, 6, 6, 6, 6, 8, 8, 10, 11, NBS-LRR genes, respectively. Most of the genes present in a cluster consisted of the same domains except for contigs 29841, 30131, 29666 and 30190 where NBS- LRR genes belonged to different classes (either TNL or CNL) in the same contig.

Table 4.10: Distribution of motifs in the castor bean NBS-LRR genes

Castor bean Contig No.	Motifs					
	P-loop	kinase-2	kinase-3	GLPL	MHDL	TIR-2
299222	VGMGGIGKTTLAQLVFN					
299222	GMGGIGKTTLAQLVFN					
29161	VGMGVLGKTTLVKKVYDSQ		IFTTRSSNV	GLPLAL		
27436			IITTRDER			
27483	GCGKTTIHQK		IITTRIENLA			
27467			VVTTRNEKIA			
28589	GCGKTTIHQK		IITTRLSKLC			
28657	GCGKTTIHQK		IITTRLPKVM			
29716	WGMGGVGKTTLLTYIYN				MHDLIRHMAI	
29726	YGMGGVGKTTLATYIYN		VTTRSEEV		MHDLVHDLAL	
29706	YGMGGVGKTTLATYIYN				MHDLIRDMAH	
29736	YGMGGVGKTTLATYIYN		VTTRHENV		MHDLIRDMAH	
29732	YGMGGVGKTTLATYIYN		VTTRHENV			
29757				AKDIVKYCHGLPL		

29757			AKDIVKYCHGLPL	
29757			AKDIVKYCHGLPL	
29841	YGMGGIGKTTLAKKIYNQ	LRRKKFVLLLDDV		
29841	YGMGGIGKTTLAK	RKKFLIVLLLDDI		
29838	YGMGGIGKTTLAKKIYNQ	RKKFLIVLDDIW		
29838	GMGGIGKTTLAKKIYNQ	RKKFLIVLDDIW		
29805	YGMGGIGKTTLAKKIYNQ	RKKFLIVLDDIW		
29841	YGMGGVGKTTLLKQINR			
29841	YGMGGVGKTTLLKQINR			
29841	YGMGGVGKTTLLKQINR			
29841	YGMGGVGKTTLLKQINR			
29841	YGMGGVGKTTLLKQINR			
29841	VGMIGIYGMGGIGKTT			MHDLIHDLA
29872	VGMIGIYGMGGIGKTT			
29904	VGMIGIYGMGGIGKTT			
29872	VGMIGIYGMGGIGKTT			MHDIIHDMA
29910	VGMIGIYGMGGIGKTT			
29948	VGMGGVGKTTLAQYVYN	RFLLVLDDVWND		

29950	VGMGGIGKTTLAQLVYN	RFLLVLDDVWND	VRKCGGLPLA
29937	VGMGGIGKTTLAQLVYN	RFLLVLDDVWND	VRKCGGLPLA
29948	VGMGGIGKTTLAQLVYN	RFLLVLDDVWND	VRKCGGLPLA
29912	VGMGGIGKTTLAQL	RFLLVLDDVWND	CGGLPLA
20178	VGMGGLGKTTLAQLVYND	RYLIVLDDIWNE	GSRITTRSKKVA
30061	VGMGGLGKTTLAQLVYND	RYLIVLDDIWNE	GSRITTRSKKVA
29990	VGMGGLGKTTLAQLVYND	RYLIVLDDIWNE	GSRITTRSKKVA
29994	VGMGGLGKTTLAQLVYND	RYLIVLDDIWNE	GSRITTRSKKVA
30040	VGMGGLGKTTLAQLVYND	RYLIVLDDIWNE	
30063	WGMGGIGKTTLAQLVY		
30063	WGMGGIGKTTLAQLVY		
30074	WGMGGIGKTTLAQLVY		
30074	WGMGGIGKTTLAQLVY		
29588	WGMGGIGKTTLAQLVY		
30128	YGMGGVGKTTLAKSVY		
30115			GSRITTRNPSVA
30115	YGMGGVGKTTLAKSVY		GSRITTRNPSVA
30110	YGMGGVGKTTLAKSVY		GSRITTRNPSVA

30074			GSRITTRNP	
30131	VGMAGIGKTTLAQ			VRKCRGLPLA
30131	VGMAGIGKTTLAQ			VRKCRGLPLA
30131	VGMAGIGKTTLAQ			VRKCRGLPLA
30131	VGMAGIGKTTLAQ			VRKCRGLPLA
30131	VGMAGIGKTTLAQ			VRKCRGLPLA
30143		VLDDIWN		VRKCGGLPLA
30143		VLDDIWN		VRKCGGLPLA
30143		VLDDIWN		VRKCGGLPLA
30131		VLDDIWN		VRKCGGLPLA
30110				VRKCGGLPLA
29577	VGMGGIGKTTVA			
29577	VGMGGIGKTTVA			
30148	VGMGGIGKTTVA			
30148	VGMGGIGKTTVA			
29634	VGMGGIGKTTLA			VRKCGGLPLA
29609	VGMGGIGKTTLA			VRKCGGLPLA
29585	VGMGGIGKTTLA			VRKCGGLPLA

29577	VGMGGIGKTTLA			VRKCGGLPLA	
29579	VGMGGIGKTTLA			VRKCGGLPLA	
29666	GMGGVGKTTL	ILDDVWGY	GSRITTRLQHVC		
29666	GMGGVGKTTL	ILDDVWGY	GSRITTRLQHVC		
29666	GMGGVGKTTL	ILDDVWGY	GSRITTRLQHVC		
29666	GMGGVGKTTL	ILDDVWGY			MHDMVRDF
29666	GMGGVGKTTL	ILDDVWGY	GSRITTRLQHV		
29676	GPGGCGKTTL	VLDDVWHDAHW		VVKCKGLPLA	
29690	GPGGCGKTTL	VLDDVWHDAHW		VVKCKGLPLA	
30170	GPGGCGKTTL	VLDDVWHDAHW		VVKCKGLPLA	
30710	GPGGCGKTTL	VLDDVWHDAHW		VVKCKGLPLA	
30205	GMGGLGKTTL			VNKCKGLPLA	
30226	GMGGLGKTTL			VNKCKGLPLA	
28966	GMGGLGKTTL			GLPLAIKA	
30170	GMGGLGKTTL			VNKCKGLPLAIKA	
27904	GMGGLGKTTL				
28660	GMGGIGKTTI				MHDLLQEMG
28490	GMGGIGKTTI				MHDLLQEMG

28525	GMGGIGKTT			MHDLLQEMG
29801	GMGGIGKTT			MHDLLQEMG
29726	GMGGIGKTT			MHDLLQEMG
29910	GMGGIGKTT	GSRIIITSRDK		
30190	GMGGIGKTTTA	GSRIIITSRDK		
30169	GMGGIGKTTTA	GSRIIITSRDK		
29838	GMGGIGKTTTA	GSRIIITSRDK		
30074	GMGGIGKTTTA	GSRIIITSRDK		
29801	GMGGIGKTTIA			
29801	GMGGIGKTTIA			
29747	GMGGIGKTTIA			
29757	GMGGIGKTTIA			
29736	GMGGIGKTTIA			
29838	GMGGIGKTTIA	GSRIIITSRDQ		MHDLIQQMG
29910	GMGGIGKTTIA	GSRIIITSRDQ		MHDLIQQMG
29910	GMGGIGKTTIA	GSRIIITSRDQ	AIRYANGLPL	MHDLIQQMG
29838	GMGGIGKTTIA	GSRIIITSRDQ		MHDLIQQMG
29841	GMGGIGKTTIA	GSRIIITSRDQ		MHDLIQQMG

29929	GMGGIGKTTIA			MHDLLQEMG
29929	GMGGIGKTTIA			MHDLLQEMG
29983	GMGGIGKTTIA			MHDLLQEMG
30074	GMGGIGKTTIA			MHDLLQEMG
30063	GMGGIGKTTIA			MHDLLQEMG
29662	GMGGIGKTTIA			MHDLLQDMG
30074	GMGGIGKTTIA			MHDLLQDMG
29841	GMGGIGKTTIA			MHDLLQDMG
30190	GMGGIGKTTIA			MHDLLQDMG
30128	GMGGIGKTTIA			MHDLLQDMG
29579		GSRIIITTRDEH		
29910		GSRIIITTRDEH		MHDLIQEMG
27904	GMGGWGKTTLA		EKCEGLPLA	
27904	GMGGWGKTTLA		EKCEGLPLA	
28152	GMGGWGKTTLA		EKCEGLPLA	MHDLMR
28824	GMGGWGKTTLA			MHDLMR
28108	GMGGWGKTTLA			
29773	MGGVGKTTLA			

29773	MGGVGKTTLA		
30074		FEAIGLPL	MHDLIR
30074		FEAIGLPL	MHDMVR
30074			MHDMVR
29910			PWCLDELVKIL
29910			PWCLDELVKIL
29910			PWCLDELVKIL
29662			PWCLDELVKIL
29666			PWCLDELVKIL

4.7) NBS profiling between *J. curcas* and *J. integerrima*

After the identification of NBS-LRR gene in the castor bean genome, the NBS domain information was used to design primers from the conserved motifs for utilization in *J. integerrima*. The NBS-AFLP analysis between *J. curcas* and *J. integerrima* resulted in the amplification of 199 fragments out of which 44 were polymorphic between both the species with an allele frequency ranging from 0.2 to 0.6. Mean number of polymorphism per enzymes- primer combination was 21.2 +/- 3.8 with a mean PIC value ranging from 0.1 to 0.33. The NBS polymorphisms were characterized by a mean marker index (MI) equal to 6.35 +/- 1.3 (Table 4.11).

4.8) Cloning and analysis of resistance gene analogues (RGAs) in *J. integerrima*

The resistance gene analogues were amplified from *J. integerrima* using primer pairs designed from the NBS and membrane-spanning motifs of castor bean NBS-LRR gene sequences representing both the TIR-NBS-LRR and non-TIR-NBS-LRR classes (Fig 4.17). The PCR amplicons were cloned into a pGEM-T vector and sequenced (Fig. 4.18). To estimate the abundance and divergence of NBS-LRR encoding RGAs and their evolution in the genome of *J. integerrima*, 30 clones were randomly selected from the library, sequenced and analyzed. Seventeen RGAs showed significantly high similarities (e-value < 0.001) at the amino acid level to the cloned plant NBS-LRR-encoding *R* genes. Multiple sequence alignments of 17 RGAs of *J. integerrima* with the 220 NBS-LRR sequences of castor bean showed that these RGAs clustered with the NBS-LRR genes of castor bean having Kinase 1, Kinase 2, P loop and GLPL motifs (Fig 4.19).

Table 4.11: Detection of Polymorphism in *J. curcas* vs. *J. integerrima* for NBS-LRR genes

Enzyme-Primer combination	Amplified bands	Polymorphic bands	Mean PIC	Marker index
Mse I/ NBS 1	23	7	0.33	9.58
EcoR I/ NBS 1	26	4	0.28	6.65
Mse I/ NBS 2	31	9	0.28	6.07
EcoR I/ NBS 2	28	6	0.31	8.67
Mse I/ NBS 3	16	3	0.23	5.32
EcoR I/ NBS 3	16	5	0.22	6.20
Mse I/ NBS 4	33	4	0.17	2.58
EcoR I/ NBS 4	26	6	0.32	5.74
Mean			0.27	6.35

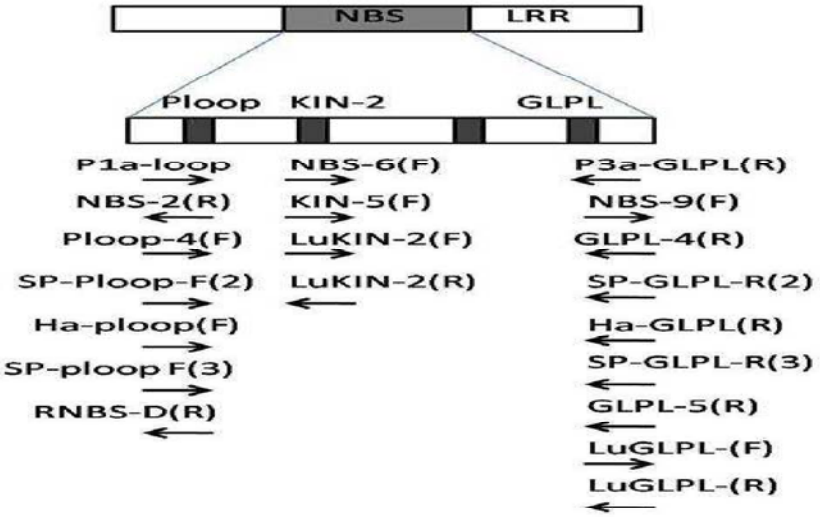

Fig. 4.17: Relative position and orientation of primers (from conserved motifs of NBS region) to amplify resistance gene analogues

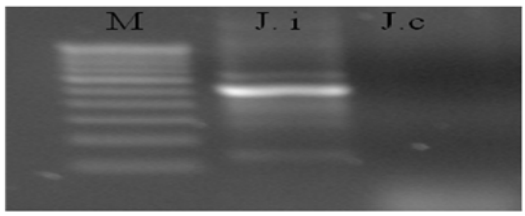

Fig. 4.18: Resistance gene analouge amplicon (~500 bp) from *J. integerrima*; Primer pair represents SPloop-F2, SP GLPL-R2 regions from conserved NBS region; J.i: *J. integerrima*; J.c: *J. curcas;* M: Molecular weight marker 100 bp

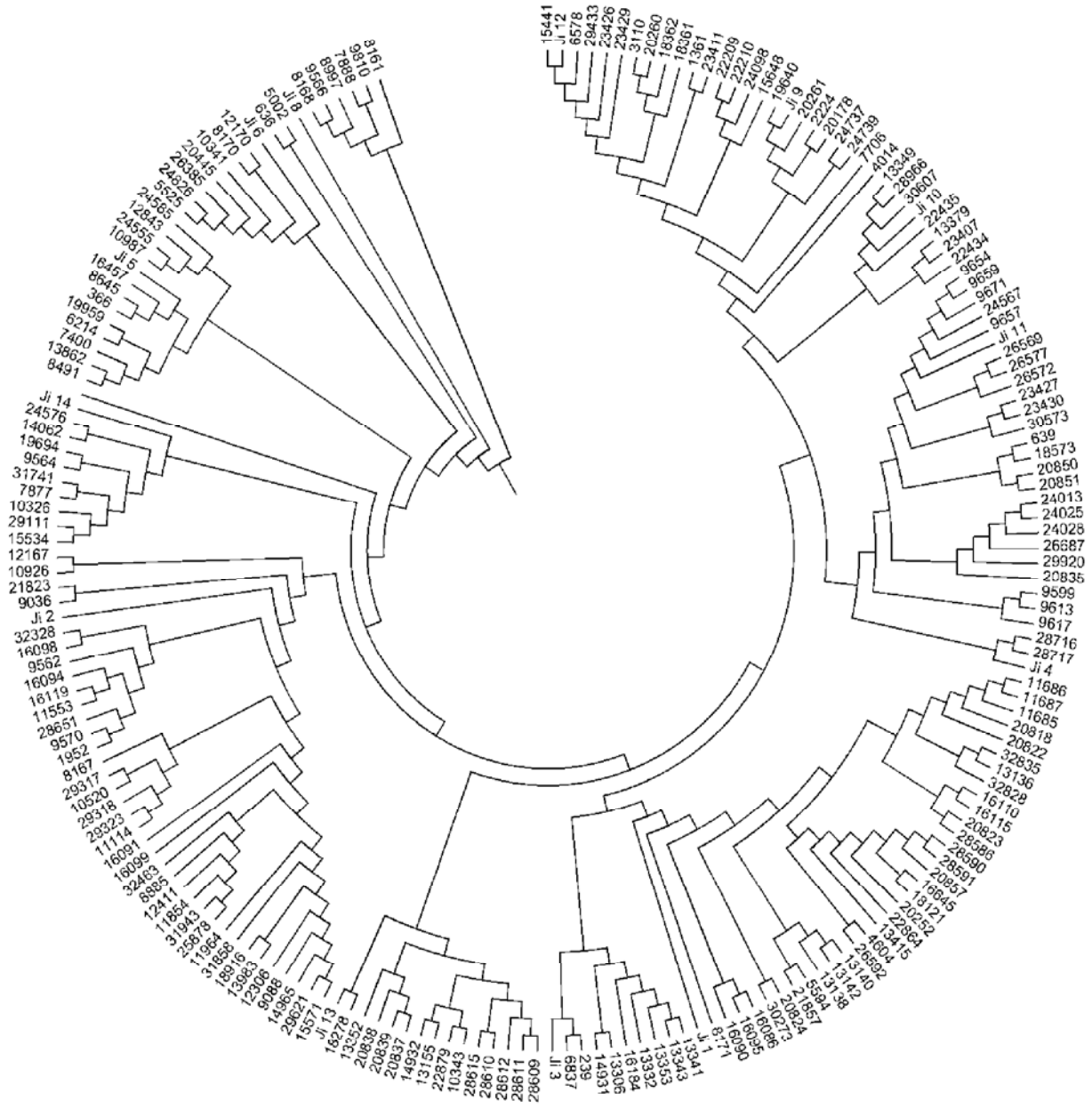

Fig. 4.19: Consensus tree of resistance gene analogs (RGAs) of *J. integrrima* and castor bean NBS-LRRs constructed by using Mega 4

Chapter-V

DISCUSSION

Bio-diesel, considered as an alternative to petro-diesel (with 5% less efficiency), can be made after transesterification from virgin or used vegetable oils (both edible and non-edible). Seed oil from *Jatropha curcas* has been projected as a major source of bio-diesel in India. Large scale cultivation of *J. curcas* in tropical and sub- tropical regions of India has warranted that systematic efforts be undertaken towards its genetic improvement. It requires little or no engine modification up to 20% blend and minor modification at higher percentage blends. The current study was taken up with the advantage of available whole genome sequence of a related plant species castor bean to develop molecular markers for *J. curcas*. Comparative genomics was successfully utilized in developing molecular markers (SSRs) including gene markers in *J. curcas* and the outcome of experiments undertaken has been described with possible explanations in the following sections:

Identification and polymorphism survey of SSRs

Genomic resources of castor bean were used for the development of SSR markers and assessing their transferability across *J. curcas* and related Jatropha species, thereby, establishing that SSR markers are a valuable genetic resource for investigating relationships and comparative mapping in Euphorbiaceae. The availability of whole genome sequences and comparative genomics have opened up several avenues for the identification of anchor makers through computational approaches, thus avoiding tedious, costly and time consuming techniques of genomic or EST library construction for the identification of SSRs (Durand *et al.*, 2010). Wen *et*

al. (2010) observed a high level of transferability of SSRs from cassava to *J. curcas*; the level of transferability was higher for EST-SSRs (44.63%) than for genomic SSRs (29.67%). Computational approaches permit rapid discovery of SSRs from the ever-increasing repertoire of publicly available sequences from many genomes. SSRs were identified in the 25,828 contigs of castor bean genome sequence with a success rate of ~70% transferability from castor bean to Jatropha, thus showing a high level of sequence identity between two plant species. High level of synteny has been observed for other loci between castor bean and Jatropha such as for genes involved in the biosynthesis of fatty acid biosynthesis (Sato *et al.*, 2011).

The distribution of SSRs in different genic regions of the castor bean revealed that SSRs were more prevalent in the 5'UTRs with a frequency of 42% compared to intronic (25%) and exonic regions (6%), which was analogous to genomic distributions of SSRs in *A. thaliana*, *B. rapa* and *O. sativa*, all of which have higher prevalence of SSRs in the 5′UTR regions (Li *et al.*, 2002; Morgante *et al.*, 2002; Fujimori *et al.*, 2003; Zhang *et al.*, 2004; Mortimer *et al.*, 2005; Lawson and Zhang, 2006; Chang *et al.*, 2007). SSRs in the 5'UTR regions might be involved in the regulation of expression of nearby genes as shown by a GT repeat in the *Tilapia prolactin* 1 gene in fish, in response to a salt-challenged environment (Streelman and Kocher, 2002). Castor bean genome contained higher number of (AT)n dinucleotide repeats which are usually abundant in many taxonomic groups, such as human, plants, fungi (Tóth *et al.*, 2000; Zhang *et al.*, 2004; Lawson and Zhang, 2006; Chang *et al.*, 2007), shrimps (Kong & Gao, 2005), insects (Prasad *et al.*, 2005; Subirana and Messeguer, 2008). Tóth *et al.* (2000) suggested that the poly(A/T) tails of densely scattered retroposed sequences, like Alu, LINE-1, and pseudogenes are responsible for this higher proportion of (A+T)-rich repeats. Trinucleotide SSRs were more frequent in exons of castor bean genome analogous to other genomes. Majority of trinucleotide repeats identified in

this study were prevalent in coding regions of the castor bean genome. Trinucleotide SSRs within exons may encode expressed amino acid runs. The majority of trinucleotide SSRs in exonic regions can be attributed to the suppression of non-trinucleotide SSRs in coding regions due to the risk of frameshift mutations, which might occur with non-trinucleotide microsatellites (Metzgar *et al*., 2000; Katti *et al*., 2001; Li *et al*., 2004; Zhang *et al*., 2004). Although the biased distribution of codon repeats has been demonstrated in several eukaryotic genomes (Tóth *et al*., 2000; Katti *et al*., 2001; Li *et al*., 2004; Zhang *et al*., 2004; Lawson and Zhang, 2006; Cavagnaro *et al*., 2010), the over-representation of specific amino acid runs varies. The most frequent amino acid runs in *A. thaliana* are serine, proline, glycine, glutamate, glutamine, and aspartate, and those in *O. sativa* are alanine, glycine, proline, serine, arginine, and glutamate (Lawson and Zhang, 2006).The most frequent amino acid runs in *Brassica rapa* are serine, glutamic acid, aspartic acid, glycine, lysine, and asparagines (Chang *et al*., 2007). The most frequent amino acid runs identified in castor bean SSRs were serine, glutamate, arginine and phenylalanine which are most frequent amino acid runs in SSRs encoded by TCT/GAA/CGC/TTC in most of the plant genomes (Sosinski *et al*., 2000; Katti *et al*., 2001; Hong *et al*., 2006).

It has been found in other plants (Arabidodpsis, Brassica, Peanut, cucumber and Medicago) that SSRs with longer repeat motifs are more informative for detection of polymorphisms (Morgante and Olivieri, 1993; Cardle *et al*., 2000; Riva *et al*., 2000; Rota *et al*., 2005; Mortimer *et al*., 2005; Moretzsohn *et al*., 2005; Cavagnaro *et al*., 2010). Sharma and Chauhan (2008) identified a SSR with longer repeat motif (TTC)31 in the iron transporter genes of maize which showed higher level of polymorphisms among maize inbreds compared to other repeat motifs with lesser repeat length. On the contrary, we found that the repeat motifs of 16-25 repeat length showed higher polymorphisms than longer repeat motifs of >25 repeat units. Out of

all SSRs, which were successfully transferred to Jatropha, 50% contained 15 to 30 repeat units whereas 20% of the SSRs had repeat unit of > 30. The majority of SSRs, which were successfully amplified and showed polymorphisms, contained more than 15 repeat units. The analysis showed wide variation in the number of alleles detected for SSRs with 16 and 25 repeat motifs compared to lower number of alleles in the SSRs with low number or very high number of repeat motifs. However, exception to this observation was observed for SSR marker, JM15 containing maximum repeat units (TA)42 showed only two alleles whereas SSR marker JM20 with relatively lower repeat motifs (TA)23 showed highest number of alleles (seven). The SSRs with longer repeat number (> 25) were less frequent than those with shorter repeat number (< 25). Similar observations were recorded in rice (Temnykh *et al*., 2001) and cereals (Varshnay *et al*., 2005). According to Lai and Sun (2003), when slippage mutations happen, the expansion of short microsatellite repeats occurs more frequently than long repeats. Thus, the variability of shorter SSRs in genomes would be more than that of longer SSRs. Other studies have also found no relationship or weak correlation between SSR polymorphism and repeat unit length (Castillo *et al*., 2008; Ferguson *et al*., 2004; Love *et al*., 1990). Dubey *et al.* (2006) showed that intron 1 of the epidermal growth factor receptor (EGFR) gene in human contained a polymorphic simple sequence repeat (SSR) of 14 to 21 CA dinucleotides, the length of which correlates inversely with the level of EGFR transcription. The higher levels of polymorphisms observed in SSRs with dinucleotide repeat motifs was also reported in SSR genotyping in Pearl millet (Senthilvel *et al*., 2008) and White clover (Zhang *et al*., 2008). Dinucleotide repeat motifs from 5'UTR regions with repeat unit length of 16-25 showed higher polymorphisms, which suggested that additional primers can be synthesized from those SSRs (16-25 repeat unit) with a higher probability of detecting polymorphisms and can be tested on castor bean and Jatropha. The decrease in SSRs

with longer repeat units was attributed to loci with longer repeat units undergo stronger selection against the difference in size (Samadi *et al*., 1998). Berube *et al*. (2007) hypothesized that the absence of longer repeat units with more than 30-50 nucleotides represent a physical limit to SSR expansion.

Overall low level of genetic diversity was detected among Jatropha genotypes compared to Jatropha species in our SSR analysis which was also reported by Ganesh Ram *et al.* (2007). The low level of molecular diversity among Indian accessions of *J. curcas* germplasm indicates a narrow genetic base (Basha and Sujatha, 2007; Sun *et al.* 2008). The SSR markers developed in this study would be very useful for germplasm analysis, population genetic structure and taxonomic relationship in related taxa.

Comparative genomics of oil biosynthesis genes in major oil producing plant species

In general, plant oil biosynthesis mostly follows the common biosynthetic pathways for fatty acids in the plastid as well as TAG in the endoplasmic reticulum (ER) and the oil further accumulates in oil bodies. However, there are significant differences for content and composition of seed oil in different plant species. Using comparative genomics, we tried to infer the effect of change in gene structure differences on oil content in different plant species. In this study, 261 genes involved in biosynthesis and accumulation of seed oil were identified in four oil seed plant species, Arabidopsis, Brassica, castor bean and soybean. The genes corresponded to six different categories (ACCase, deasturase, elongase, thioesterase, TAG synthesis and oil body proteins). Gene families corresponding to these six categories of enzymes had multiple copies in plant species with the exception of homomeric ACCase.

In higher plants, many proteins and enzymes are encoded by gene families, and in Arabidopsis, it has been estimated that 20% of genes are members of gene families (Bevan *et al.*, 1998). The existence of gene families can sometimes reflect additional levels of genetic control or isoforms of proteins with specific functions. Therefore, it is of interest to detect potential gene families involved in the fatty acid biosynthesis pathway. There is a possibility that different copies of fatty acid biosynthesis genes are present in low oil content genotypes which gives leaky phenotypes as in the case of starch biosynthesis pathway where different copies of genes were responsible for low, medium and high amylase contents in rice (Tian *et al.*, 2009).

The oil biosynthesis may be limited by the production of fatty acids (Bao and Ohlrogge, 1999), which is regulated by acetyl CoA carboxylase (ACCase). Reduction of ACCase activity lowered (1.5-16%) the fatty acid content in transgenic seeds (Thelen and Ohlrogge, 2002). Conversion of acetyl Co-A to malonyl Co-A by acetyl carboxylase (ACCase) is the most committed step in fatty acid biosynthesis. ACCase of castor bean and soybean showed microsynteny to Arabidopsis, with a 3 bp deletion in 8^{th} and 26^{th} exons in castor bean, 3 bp deletion and 3 bp insertion in 29^{th} and 31^{st} exons in soybean and a 2 bp insertion in 24^{th} and 26^{th} exons of castor bean and soybean, respectively w.r.t. Arabidopsis. These sequence variations in ACCase genes may be possibly influencing the variations in fatty acid composition and content in seed oil among Arabidopsis, castor bean and soybean; as fatty acid content and composition was altered in many plant species with the variations in sequences or expression of ACCase gene (Roesler *et al.*, 1997; Yang *et al.*, 2010). Yang *et al.* (2010) identified two SNPs (T→G, G→A) in ACCase gene which lead to increase (1.3%) in oleic acid, lenolenic acid and lenoleic acid content in maize. Addition of a plastid transit sequence targeted the introduced ACCase protein to chloroplasts, ultimately resulting in a 5% increase in seed oil of rapeseed (Roesler *et al.*,

1997). The insertions and deletions identified in our analysis between Arabidopsis, castor bean and soybean might be responsible in reduction or enhancement of ACCase activity, which is associated with the variations in total fatty acid composition in seed oil among these plant species.

Studies in transgenic plants have demonstrated that thioesterases contribute to the regulation of fatty acid chain length (Jones *et al.,* 1995). Typically, FatB accepts saturated acyl-ACP substrates of varying length, while FatA is specific to unsaturated fatty acids and acts on C18:1, oleic, acyl-ACPs (Jones *et al.,* 1995). In *Brassica napus* and Arabidopsis, genetic engineering of Acyl-ACP thioesterase (FatB) resulted in maximum increase of 58% in palmitic acid content (Voelker *et al.,* 1996; Domann *et al.,* 2000). Preventing the release of saturated fatty acids from ACP by down-regulating FatB, which encodes a palmitoyl ACP thioesterase, lowered the levels of saturated fatty acids (Buhr *et al.,* 2002). Variations in palmitate content in seed oil in plant species can be related to the variations in FatB gene (Voelker *et al.,* 1996; Domann *et al.,* 2000; Cardinal *et al*., 2007). Cardinal *et al*. (2007) identified deletion in exon-inton junction in one homolg of FatB gene which was associated with low palmitic acid content in soybean cultivar Century (N79-2077 and N93-2008). Palmitate content was upto ~8% in Arabidopsis (Bonaventure *et al.,* 2003), ~2% in castor bean (Akbar *et al.,* 2009) and 7-11% in soybean (Kinney, 1997). Variations in the amount of palmitic acid in the seeds of Arabidopsis, castor bean and soybean might be attributed to the deletions in first exon of FatB gene, which can be further utilized for identification of markers associated with high level of palmitate (saturated fatty acid) in total seed oil in different plant species which is desirable for biodiesel purpose.

Soybean lines with high levels of oleic acid (85%) and low levels of saturated fatty acids (6%) have been developed using a transgenic strategy that results in down regulation of two

genes, FAD 2 and FatB involved in fatty acid synthesis. Down-regulation of the FAD 2 gene, encoding a Δ12 fatty acid desaturase, prevented the conversion of oleic acid to polyunsaturated fatty acids, resulting in increased levels of oleic acid. Additionally, preventing the release of saturated fatty acids from ACP (acyl carrier protein) by down-regulating FatB gene, which encodes a palmitoyl ACP (acyl carrier protein) thioesterase, lowered the levels of saturated fatty acids (Buhr *et al.,* 2002). Hu *et al.* (2006) sequenced the FAD 2 genomic fragment from the mutant line DMS100 and wild-type line Quantum of *Brassica napus*, and identified a single nucleotide mutation (C→T) in the FAD 2 gene. This particular mutation created a stop codon (TAG) leading to premature termination of the peptide chain during translation which leads to high oleic acid content in mutant line DMS100. *B. napus* mutant line DMS100 carrying a G-to-A substitution at the 5' splice site of intron 6 in FAD 3 had reduced lenolenic acid content in seed oil (Hu *et al.,* 2007). In our analysis insertions or deletions in FAD 2 and FAD 3 genes of soybean might be the possible causes of higher oleate and linoleate content in high oil yielding soybean genotypes. Higher amount of recinolic acid in castor bean can be related to insertion in FAD 2 gene which might be related to the higher level of oleic acid which is further utilized in castor bean as a substrate for fatty acid hydroxylase (FAH) which converts oleate to recinoleate. Low level of linoleate in castor bean oil can be related to a deletion in the 3rd exon of FAD 3 gene, which is otherwise conserved with each copy of Arabidopsis and soybean.

In our analysis, the acyl-CoA:diacylglycerol acyltransferases (DGAT) gene was highly diverse; which might be involved in the overall variation in triacylglycerols in the oil among the plant species as it is a key enzyme in determining the levels of triacylglycerols in seed oils (Burgal *et al.,* 2008; Venegas *et al.,* 2009). Burgal *et al.* (2008) demonstrated that co-expressing the castor bean DGAT2 gene with the castor FA 12 hydroxylase resulted in almost double the

levels of hydroxylated fatty acids in neutral lipids (up to 30% of total, compared with 17% in the absence of DGAT2). In our study, most of the variations observed in the coding regions are either insertion or deletion of 3 bp or multiple of three which leads to functional mutation that are expected to be related with oil content. Thus the sequence variations identified in fatty acid biosynthesis genes in this study can be tested for their functional role in altering content and composition of seed oil in Jatropha.

Candidate gene markers and their association with high oil content in *J. curcas*

The development of molecular markers from candidate genes encoding a particular phenotype is being pursued vigorously in human genetics as well as in plants with the long term goals of using candidate gene markers in molecular breeding and for genetic manipulation of crop plants for economically important traits. Candidate genes often present low levels of polymorphism because they are often relatively conserved and because it is generally difficult to map these regions using markers based on size polymorphism. Genomic resources of castor bean were successfully used for identification of candidate genes involved in fatty acid biosynthesis in *J. curcas* and establishing that the candidate gene markers are a valuable genetic resource for investigating relationships and comparative mapping in Euphorbiaceae. Thirty two candidate genes for fatty acid biosynthesis were identified in castor bean and their *in silico* expression status was checked. Candidate genes corresponding to seed ESTs were considered for candidate gene marker analysis. The transferability of castor bean gene markers was higher in exonic regions (~77%). Higher transferability of coding sequence and EST- derived markers has been reported (Fulton *et al.,* 2002; Eujayl *et al.,* 2004; Heesacker *et al.,* 2009).

Out of 11 candidate gene markers from different regions of fatty acid biosynthesis genes, one marker-trait association for high oil trait was identified in Jatropha. A marker from exon-intron junction (JJM1) of Steroyl desaturase showed association with high oil content in Jatropha genotypes. Association of a particular region of the gene with high oil content might be due to functional mutation in the corresponding region in low oil content genotypes. We identified one gene specific marker (JJM1) which showed association with high oil content genotypes with a r^2 value of ~0.76. For this marker no amplification was observed in low oil content genotypes. Association mapping has been used to detect the degree to which a gene and trait or genotype and phenotype vary together in a sampled population on the basis of linkage disequilibrium (Zondervan and Cardon, 2004). When a molecular marker is associated with a phenotypic trait, it should associate with others that highly correlate with this trait in theory. Sequencing of a 1.3 kb amplicon, including the 700 bp fragment (JJM1) of Stearoyl desaturase, associated with the high oil content, revealed SNPs in the exonic region. Comparison of Stearoyl desaturase sequences from low and high oil content genotypes, including the GenBank sequence (DQ084491) showed functional polymorphism (non polar amino acid to polar amino acid) in low oil content genotypes. The SNPs (TT→AA) in the exonic region of Stearoyl desaturase gene resulted in substitution of leucine (non polar) with glutamine (polar) in the open reading frame of Stearoyl desaturase in low oil content genotypes of *J. curcas*. The low oil content in Jatropha genotypes might be because of lower amount of oleic acid in the seed oil (low oil yielding genotypes) which is otherwise present in high amount (~ 43%) in Jatropha genotypes (high oil yielding genotypes). The *in- silico* sequence comparison of Stearoyl desaturase genes from low and high oil content Jatropha genotypes with the sequences of Stearoyl desaturase gene from soybean, castor bean and Arabidopsis revealed SNPs (TT→AA) between castor bean and Arabidopsis.

The same SNPs were identified in the low oil content Jatropha genotypes. Both plant species, castor bean and Arabidopsis has low oleic acid content of 15% and 7%, respectively (Akbar *et al.,* 2009; Bonaventure *et al.,* 2003). Variations in expression of Stearoyl desaturase genes at various stages from fruit formation to ripening were also observed between high versus low oil content genotypes of *J. curcas* (Chauhan *et al*., unpublished data).

There is precedence in the literature where SNPs associated with oil content have been identified in the fatty acid pathway genes. Knutzon *et al.* (1992) identified indel in exon, associated with Oleoyl ACP carrier content in safflower. SNP were also identified in the exon of Oleoyl desaturase gene in Peanut, sunflower and Brassica (Lopez *et al.,* 2000; Lacombe and Bervillé, 2001; Hu *et al.,* 2006). SNPs were also identified in exon of Steroyl desaturase gene in Arabidopsis and soybean associated with high steric acid content (Zhang *et al.,* 2008; Kachroo *et al.,* 2007). Bilyeu *et al.* (2005) identified a SNP in the intron of Linoleoyl desaturase in soybean associated with low linolenic acid content in seed oil. Pérez-Vich *et al.* (2006) identified SSRs and INDELs in Stearoyl–ACP desaturase associated with high stearic acid in sunflower.

Analysis of castor bean NBS-LRR genes and their utilization in cloning RGAs in *J. integerrima*

Commercial cultivation of selected genotypes of castor bean and Jatropha has predisposed to a plethora of biotic stresses, including insect pests and fungal, viral and bacterial diseases. No systematic breeding efforts have been made till date towards the development of disease resistant genotypes in both the plant species. Since a large number of pest and disease resistance genes in various plant species belong to NBS-LRR family of proteins, which is highly conserved across kingdoms, a need was necessitated to identify and characterize NBS-LRR genes in the castor bean genome. Two hundred and twenty CNL- and TNL encoding genes were

identified in the current version of castor bean genome representing ~ 0.4% of all predicted ORFs, which is relatively small in number compared to other sequenced plant genomes with the same range of genome sizes (Chan *et al.*, 2010). For example the genomes of *Arabidopsis thaliana* and *Vitis vinifera* contain relatively higher numbers of NBS-LRR genes, 174 and 535 with the corresponding genome sizes of 125 and 487 Mb, respectively (Yang *et al.,* 2008).

The primers designed from the conserved regions of the castor bean NBS-LRR genes resulted in successful amplification and cloning of resistance gene analogues (RGAs) from a wild species, *J. integerrima*, which has been found to possess tolerance to viral and fungal diseases of Jatropha under field conditions. Multiple sequence alignments of selected Jatropha resistance gene analogs and castor bean NBS- LRR genes showed significant homology in the kinase-1a, kinase-2 and GLPL motifs of the NBS domain. *J. integerrima* RGAs were classified into the TIR and non-TIR-NBS R-genes. The NBS-LRR genes are classified into two subfamilies: The TIR-NBS-LRR subfamily, which is characterized by the presence of a highly conserved aspartic acid (D) or aspartate (N) as the last residue of the kinase-2 domain, and the non-TIR-NBS-LRR subfamily with a highly conserved tryptophan (W) as the last residue of the kinase-2 domain (Pan *et al.*, 2000). Moreover, both TIR and non-TIR-NBS-LRR subclasses are present in dicots, but in monocots only the non-TIR subclass is present and the other one is completely absent (Meyers *et al.*, 1999; Cannon *et al.,* 2002). This clearly supports the hypothesis that Jatropha, being dicotyledonous, has resistance genes of both TIR and non-TIR-NBS-LRR class. The NBS-LRR genes were in clusters of 2 to 11 genes in the castor bean genome. The cluster analysis of the NBS analogs identified their groups based on similarity. Occurrence of a gene cluster on a sequence contig harboring the same domain suggests that these genes might have arisen from a recent duplication event or have been maintained without

substantial divergence. The phylogenetic tree based on neighbor joining using % identity of the deduced amino acid sequences of *J. integerrima* RGAs identified their relatedness with each other as well with castor bean NBS genes identified in this study. Whether these RGAs from *J. integerrima* contribute to the disease resistance phenotype remains to be seen through mapping of these RGAs in a mapping population segregating for resistance or susceptibility. It has been observed in various plant species that the RGAs map to locations of known disease resistance genes (Porter *et al*., 2009), thereby, reinforcing that the RGAs cloned from *J. integerrima* can be of practical importance in breeding for disease resistance in Jatropha.

The study has not only resulted in the identification of RGAs in *J. integerrima* but also identification and characterization of whole repertoire of NBS-LRR genes in the castor bean genome which opens up avenues for their functional characterization through genetic mapping or other functional genomics approaches. The outcome of the study is, therefore, of great practical importance in two major oilseed crops of industrial value. The NBS-LRR gene sequences of castor bean and RGAs from *J. integerrima* can be used in the development of candidate gene markers for molecular breeding of disease resistance in castor bean and Jatropha.

Summary

Jatropha (*Jatropha curcas*) and castor bean (*Ricinus communis*) possess several taxonomical similarities and their seeds contain high proportion of oil (up to 50%) which has been used in various industrial products, including biodiesel. Genetic improvement of both the plant species largely depends on the availability of genotypes with varying amounts of seed oil contents coupled with molecular markers and genetic maps. However, both the species lack basic genome resources, therefore, taxonomical similarities between them are expected to accelerate the process of developing genome resources. In the current study, comparative genomics was utilized for the identification of SSRs in the whole genome sequence of castor bean and their transferability and polymorphism analysis in *Jatropha curcas*. Whole genome analysis of castor bean revealed the presence of 5, 80,986 SSRs with a frequency of 1 per 680 bp. Genomic distribution of SSRs revealed that 27% were present in non-genic regions whereas 73% in the genic regions comprising 26% in 5'UTRs, 25% in introns, 16% in 3'UTRs and 6% in exons of castor bean. Dinucleotide repeats were more frequent in introns, 5'UTRs and 3'UTRs whereas trinucleotide repeats were predominantly present in the exons. The (AT)n repeat motif was more common (43%) among all SSRs identified. A total of 302 primer pairs were designed from selected SSRs representing different genome regions and of varying repeat motifs and lengths. The transferability of 302 SSRs from castor bean to 49 *J. curcas* genotypes and 9 Jatropha species including *J. curcas* showed that 211 (~70%) amplified successfully on Jatropha out of which 17.03% showed polymorphisms in Jatropha genotypes and species. Sixteen SSRs from 5'UTRs (5), intronic (3), exonic (3) and non-genic regions (5) showed polymorphisms in *J. curcas* genotypes with allele diversity of 17.5%. On the other hand 26 SSRs (12 from UTRs, 7 intronic, 3 exonic and 4 from non-genic regions) were polymorphic on Jatopha species. Five

SSRs showed polymorphisms on *J. curcas* genotypes as well as Jatropha species. The study showed observed (HO) and expected (HE) heterozygosities ranging from 0.11 to 0.54 (mean 0.25) and 0.13 to 0.54 (mean 0.25), respectively in *J. curcas* genotypes and 0.44 to 0.76 (mean 0.66) and 0.44 to 0.77 (mean 0.66), respectively in Jatropha species. The higher rate of transferability of SSR markers from castor bean to Jatropha coupled with a good level of PIC (polymorphic information content) value (0.23 in *J. curcas* genotypes and 0.50 in Jatropha species) suggested that SSRs would be useful in germplasm analysis, linkage mapping, diversity studies and phylogenetic relationships, etc. in *J. curcas* as well as related Jatropha species. A large number of SSRs identified in the castor bean genome can be further tested on *J. curcas* or other related plant species.

Since castor bean and Jatropha are both industrially important due to high oil contents in their seeds, the identification of candidate genes involved in fatty acid biosynthesis and their utilization in the development of candidate gene markers was undertaken. Genes involved in lipid metabolism have been extensively studied in model plant species such as Arabidopsis. Genes involved in fatty acids biosynthesis, modification and oil body formation are expected to be conserved in structure and function in different plant species. However, significant differences in the composition of fatty acids and total oil contents in seeds have been observed in different plant species. Molecular bases of these differences, other than branch points in the biosynthetic pathways resulting in the formation of a specific fatty acid, are not known. Comparative genomics, for gene structures and coding sequence variations, was performed on 261 genes involved in fatty acids biosynthesis, TAG synthesis, and oil bodies formation in four oil seed plant species, Arabidopsis, *B. rapa*, castor bean and soybean to understand whether differences in gene structures or coding sequence determine preferential biosynthesis of higher amounts of

particular fatty acids and their contents in the seeds of different plant species. The total seed oil contents of Arabidopsis, Jatropha, castor bean, Brassica and soybean are 30-37%, 30-50%, 40-45%, 30-40% and 15-20%, respectively (Li *et al.,* 2006; Akbar *et al.,* 2009; Velasco and Becker, 1998; Ramos *et al*., 1984; Sangwan *et al*., 1986). *In silico* expression analysis of 261 genes revealed that the transcripts of Stearoyl desaturase, FatB, FAD 2, Oleosin and DGAT are abundant in the seeds. Fatty acid biosynthetic pathway genes shared 72–94% nucleotide sequence identity among four plant species. The exon/intron split pattern of *B. rapa*, castor bean and soybean genes was almost identical to that of Arabidopsis, although insertion, deletion and intron size variations were common. ACCase is a key enzyme in fatty acid biosynthesis. Reduction of ACCase activity lowers (1.5-16%) the fatty acid content in transgenic seeds (Thelen and Ohlrogge, 2002). ACCase of castor bean and soybean showed microsynteny to Arabidopsis, with a 3 bp deletion in 8^{th} and 26^{th} exons of castor bean, 3 bp deletion and 3 bp insertion in 29^{th} and 31^{st} exons of soybean and a 2 bp insertion in 24^{th} and 26^{th} exons of castor bean and soybean, respectively. These sequence variations in ACCase gene may be one of the possible reasons to influence variation in fatty acid composition and content in seed oil among Arabidopsis, castor bean and soybean. Yang *et al*. (2010) identified two SNPs (T→G, G→A) in ACCase gene which lead to increase (1.3%) in oleic acid, lenolenic acid and lenoleic acid content in maize. FatB gene (palmitoyl ACP thioesterase) releases saturated fatty acid (palmitic acid) from ACP. The first exon of FatB gene had an insertion of 3 bp in castor bean and 27 bp insertion in one of soybean homologs (Glyma0421910) and other three homologs of soybean had 6 bp deletion in FatB gene w.r.t. Arabidopsis. Cardinal *et al*. (2007) identified deletion in exon-inton junction in one homolg of FatB gene which was associated with low palmitic acid content in soybean cultivar Century (N79-2077 and N93-2008). Palmitate content is upto ~8% in

Arabidopsis (Bonaventure *et al.,* 2003), ~2% in castor bean (Akbar *et al.,* 2009) and 7-11% in soybean (Kinney, 1997).Variations in the amount of palmitic acid in the seeds of Arabidopsis, castor bean and soybean may be due to the deletions in the first exon of FatB gene, which can be further tested and utilized for the identification of markers associated with high levels of palmitate (saturated fatty acid) in seed oil contents of different plant species. Sequencing FAD 2 gene from the mutant line DMS100 and wild-type line Quantum of *Brassica napus* identified a single nucleotide mutation (C→T) in the FAD 2 gene (Hu *et al*., 2006). This particular mutation created a stop codon (TAG) leading to premature termination of the peptide chain during translation which lead to high oleic acid content in mutant line DMS100. *B. napus* mutant line DMS100 carrying a G-to-A substitution at the 5' splice site of intron 6 in FAD 3 had reduced lenolenic acid content in seed oil (Hu *et al.,* 2007). In our analysis insertions or deletions in FAD 2 and FAD 3 genes of soybean may be the possible causes of higher oleate and linoleate contents in high oil yielding soybean genotypes. Higher amount of recinolic acid (87%) in the castor bean seed oil may be due to an insertion in the FAD 2 (Oleoyl desaturae) gene which probably results in the higher amount of oleic acid which is further utilized as a substrate by fatty acid hydroxylase (FAH) to convert oleate to recinoleate. Low level of linoleate in the castor bean seed oil may be due to a 6 bp deletion in the 3rd exon of FAD 3 gene (Linoleoyl desaturase), which is absent in the FAD 3 gene of Arabidopsis and soybean. In our study, most of the variations observed in the coding regions are either insertion or deletion of 3 bp or multiples of three which lead to functional mutations. The sequence variations identified in fatty acid biosynthesis genes in this study can be tested for their functional role in altering content and composition of seed oil in Jatropha.

Thirty two critical candidate genes responsible for fatty acid biosynthesis were identified in the castor bean genome. Testing of 48 primer pairs designed from different regions (16 for exons, 20 for introns and 12 for 5' & 3' UTRs) of the candidate genes on 54 genotypes of *J. curcas* showed that 43 (~71%) amplified successfully on Jatropha out of which 25.5% showed polymorphisms in size. A network program of several research institutions in India has categorized *J. curcas* genotypes into high (> 35%) versus low (< 30%) oil content, which were utilized in association analysis of candidate gene regions with high oil content. A mixed linear model combining PC-matrix and K-matrix was adapted for marker-trait associations. Steroyl desaturase amplicon (700 bp) marker named as JJM1 (Jaypee Jatropha Marker1) consisting of exon-intron junction ($p=0.00013$) showed association with high oil content in Jatropha genotypes. Sequencing of 1.3 kb amplicon, including the 700 bp fragment of JJM1, from the exon- intron junction of Stearoyl desaturase, revealed SNPs in the exonic region. SNPs (TT→AA) resulted in substitution of leucine with glutamine in the open reading frame of Stearoyl desaturase of low oil content *J. curcas* genotypes. The candidate gene marker, JJM1 holds promise in marker-assisted selection and breeding of high seed oil content in Jatropha.

The large-scale cultivation of selected genotypes of *J. curcas* across India has resulted in venerability of this plant species to biotic stresses such as fungal and viral pathogens. Recently, a mosaic disease has been found to reduce fruit yield and quality of *J. curcas* plants in the field conditions and similar pathological stresses are expected to emerge in future. It is, therefore, prudent to undertake identification and characterization of resistance gene analogues from disease resistant species such as *J. integerrima* with the possibility of their utilization in molecular breeding of *J. curcas* for disease resistance. The largest class of known disease resistance proteins contains a nucleotide binding site -leucine rich repeats (NBS-LRR proteins).

The sequencing of castor bean genome has provided a plethora of genes, including NBS-LRRs, which need to be characterized and utilized in other plant species. A total of 220 NBS-LRR proteins were identified in the castor bean genome and categorized into 113 CNLs (N-terminal coiled-coil motif) and 107 TNLs (N-terminal domain with Toll/Interleukin-1 Receptor homology). Castor bean genome had less NBS-LRR genes compared to other sequenced plant genomes. Fifty one NBS-LRR genes showed significant similarity to ESTs. Out of 220 NBS-LRR genes, 136 (~ 61.4%) were present in clusters. Most of the NBS-LRR genes were in clusters of 2 to 3 genes, except 11 clusters of 4 to 11 genes. Castor bean NBS-LRRs provided a resource for cloning resistance gene analogues from a disease resistant species, *J. integerrima*. NBS-AFLP profiling was done between *J. curcas* and *J. integerrima* to estimate genetic diversity at NBS-LRR loci between two plant spcies. Mean polymorphism for per enzyme-primer combination was equal to 21.2 +/- 3.8. Primers designed from the conserved motifs of NBS domain resulted in the cloning of 17 resistance gene analogs (RGA) from *J. integerrima*. The identification of RGAs in *J. integerrima* and characterization of whole repertoire of NBS-LRR genes in the castor bean genome opens up avenues for their functional characterization and utilization in resistance genes identification in Jatropha species.

List of publications (presented in conferences, consideration/communicated) from of the Ph.D. research work

A) Patent filed

- Chauhan RS, Sharma A and Sood P. Patent Application No. 292/DEL/2010 filed in India on "**Gene markers for selection and development of high oil content Jatropha**"

B) Papers communicated in International journals

1. Sharma A. & Chauhan RS. 2010. Repertoire of SSRs in the castor bean genome and their utilization in genetic diversity analysis in *Jatropha* species. Comparative and Functional Genomics (Manuscript ID: CFG/286089; Revised Submitted)
2. Sharma A. & Chauhan RS. 2010. Computational identification of candidate gene markers through *in-silico* analysis of genes involved in biosynthesis and accumulation of seed oil in plants. Journal of Structural and Functional Genomics (Manuscript ID: JSFG120)
3. Sharma A. & Chauhan RS. 2010. Development of candidate gene markers and their association with high oil content in *Jatropha curcas* through comparative genomics with castor bean (*Ricinus communis*). Genome (Manuscript ID: 10-318)
4. Sharma A, Sood P, Pandit S and Chauhan RS. 2010. Genome wide identification of castor bean NBS-LRR genes and their utilization in cloning Resistance Gene Analogues (RGAs) in *Jatropha integerrima.* Journal of Plant Research (Manuscript ID: JPR-11-0006)

C) International Conferences Papers

1. Sharma A. Sood P. and Chauhan R.S. 2008. Development of candidate gene markers in Jatropha through comparative genomics. International Conference on Molecular Biology and Biotechnology (ICMBB 19-21Oct.), Banasthali Vidyapith (Rajasthan), India.
2. Sharma A. & Chauhan RS. 2010. Comparative analysis of genes involved in oil synthesis and accumulation in plants for candidate gene markers identification. International Conference Genomics Sciences- Recent Trends (ICGS 12-14 Nov.), Madurai Kamraj University (TN), India.
3. Chauhan RS, Sharma A and Sood P. 2011. Comparative and functional genomics to decipher fatty acid biosynthesis genes in Jatropha (*Jatropha curcas*) and their utilization in molecular breeding. Plant & Animal Genome XI, Jan. 15-19, San Diego, California, USA

REFERENCES

- Aghoram K, Wilson RF, Burton JW and Dewey RE (2006) A mutation in a 3-keto-ACP synthase II gene is associated with elevated palmitic acid levels in soybean seeds. Crop Science 46: 2453–2459
- Akbar E, Yaakob J, Kamarudin SK, Ismail M and Salimon J (2009) Characteristic and Composition of *Jatropha Curcas* Oil Seed from Malaysia and its Potential as Biodiesel Feedstock. European Journal of Scientific Research 29: 396-403
- Ameline-Torregrosa C, Wang BB, O'Bleness MS, Deshpande S, Zhu H, Roe B, Young ND and Cannon SB (2008) Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. Plant Physiology 146: 5–21
- Anai T, Yamada T, Kinoshita T, Rahman SM and Takagi Y (2005) Identification of corresponding genes for three low- α-linolenic acid mutants and elucidation of their contribution to fatty acid biosynthesis in soybean seed. Plant Science 168:1615–1623.
- Andrianov V, Borisjuk N, Pogrebnyak N, Brinker A, Dixon J, Spitsin S, Flynn J, Matyszczuk P, Andryszak K, Laurelli M, Golovkin M and Koprowski H (2010) Tobacco as a production platform for biofuel: overexpression of Arabidopsis DGAT and LEC2 genes increases accumulation and shifts the composition of lipids in green biomass. Plant Biotechnology Journal 8(3): 277-287
- Antoni D, Zverlov VV and Schwarz WH (2007) Biofuels from microbes. Applied Microbiology and Biotechnology 77: 23–35
- Augustus GDPS, Jayabalan M and Seiler GJ (2002) Evaluation and bioinduction of energy components of *Jatropha curcas*. Biomass and Bioenergy 23: 161-164

- Axelsson T, Shavorskaya O and Lagercrantz U (2001) Multiple flowering time QTLs within several *Brassica* species could be the result of duplicated copies of one ancestral gene. Genome 44: 856–864
- Ayele-Gedil M, Slabaugh MB, Berry S, Johnson R, Michelmore R, Miller J, Gulya T and Knapp S (2001) Candidate disease resistance genes in sunflower cloned using conserved nucleotide-binding site motifs: genetic mapping and linkage to the downy mildew resistance gene Pl1. Genome 44: 205-212
- Bachlava E, Dewey RE, Auclair J, Wang S, Burton JW and Cardinal AJ (2008) Mapping genes encoding microsomal omega-6 desaturase enzymes and their cosegregation with QTL affecting oleate content in soybean. Crop Science 48: 640–650
- Bao XM and Ohlrogge J (1999) Supply of fatty acid is one limiting factor in the accumulation of triacylglycerol in developing embryos. Plant Physiology 120: 1057–1062
- Basha SD and Sujatha M (2007) Inter and intra-population variability of *Jatropha curcas* (L.) characterized by RAPD and ISSR markers and development of population-specific SCAR markers. Euphytica 156: 375–386
- Basha SD and Sujatha M (2009) Genetic analysis of *Jatropha* species and interspecific hybrids of *Jatropha curcas* using nuclear and organelle specific markers. Euphytica 168: 197-214
- Basha SD, Francis G, Makkar HPS, Becker K and Sujatha M (2009) A comparative study of biochemical traits and molecular markers for assessment of genetic relationships between *Jatropha curcas* L. germplasm from different countries. Plant Science 176: 812–823

- Beisson F, Koo A, Ruuska S, Schwender J, Pollard M, Thelen J, Paddock T, Salas J, Savage L, Milcamps A, *et al.* (2003) Arabidopsis genes involved in acyl lipid metabolism. A 2003 census of the candidates, a study of the distribution of expressed sequence tags in organs, and a Web-based database. Plant Physiology 132: 681–697
- Bennett MD and Smith JB (1976) Nuclear DNA amounts in angiosperms. Proceedings of Royal Society London B 274: 227–274
- Bennetzen JL and Ramakrishna W (2002) Numerous small rearrangements of gene content, order, and orientation differentiate grass genomes. Plant Molecular Biology 48: 821–827
- Bennetzen JL, SanMiguel P, Chen MS, Tikhonov A, Francki M and Avramova Z (1998) Grass genomes. Proceeding of National Academy of Sciences USA 95: 1975-1978
- Berube Y, Zhuang J, Rungis D, Ralph S, Bohlmann J and Ritland K (2007) Charecterization of EST- SSRs in loblolly pine and spruce. Tree Genetics and Genomics 3: 251-259
- Bevan M, Bancroft I, Bent E, Love K, Goodman H, Dean C, Bergkamp R, Dirkse W, Van Staveren M, Stiekema W, Drost L, Ridley P, Hudson SA, Patel K, Murphy G, Piffanelli P, Wedler H, Wedler E, Wambutt R, Weitzenegger T, Pohl TM, Terryn N, Gielen J, Villarroel R, De Clerck R, Van Montagu M, Lecharny A, Aubog S, Gy I, Kreis M, Lao N, Kavanagh T, Hempel S, Kotter P, Entian KD, Rieger M, Schaeffer M, Funk B, Mueller- Auer S, Silvey M, James R, Montfort A, Pons A, Puigdomenech P, Douka A, Voukelatou E, Milioni D, Hatzopoulos P, Piravandi E, Obermaier B, Hilbert H, Du¨ sterho¨ ft A, Moores T, Jones JDG, Eneva T, Palme K, Benes V, Rechman S, Ansorge W, Cooke R, Berger C, Delseny M, Voet M, Volckaert G, Mewes HW, Klosterman S,

Schueller C and Chalwatzis N (1998) Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. Nature 391: 485–488

- Bilyeu KD, Palavalli L, Sleper DA and Beuselinck PR (2005) Mutations in soybean microsomal omega-3 fatty acid desaturase genes reduces linolenic acid concentration in soybean seeds. Crop Science 45: 1830–1836
- Bonaventure G, Salas JJ, Pollard MR and Ohlrogge JB (2003) Disruption of the *FATB* Gene in Arabidopsis Demonstrates an Essential Role of Saturated Fatty Acids in Plant Growth. The Plant Cell 15: 1020-1033
- Bringi NY (1987) Non-traditional oilseeds and oils of India. New Delhi, Oxford & IBH Publishing Co. Pvt. Ltd.
- Brugmans B, Wouters D, van Os H, Hutten R, van der Linden G, Visser RG, van Eck HJ and van der Vossen EA (2008) Genetic mapping and transcription analyses of resistance gene loci in potato using NBS profiling. Theoretical and Applied Genetics 117(8): 1379-1388
- Bruner AC, Jung S, Abbott AG and Powell GL (2001) The naturally occurring high oleate oil character in some Peanut varieties fesults from reduced oleoyl desaturase activity from mutation of Aspartate 150 to Asparagine. Crop Science 41: 522–526
- Buhr T, Sato S, Ebrahim F, Xing AQ, Zhou Y, Mathiesen M, Schweiger B, Kinney A, Staswick P and Clemente T (2002) Ribozyme termination of RNA transcripts down-regulate seed fatty acid genes in transgenic soybean. Plant Journal 30: 155–163
- Burgal J, Shockey J, Lu C, Dyer J, Larson T, Graham I and Browse J (2008) Metabolic engineering of hydroxy fatty acid production in plants: RcDGAT2 drives dramatic increases in ricinoleate levels in seed oil. Plant Biotechnology Journal 6: 819-831

- Callow JA, Ford-Lloyd BV and Newbury HJ (1998) Biotechnology and plant genetic resources: conservation and use. Euphytica 101: 49–76
- Cannon SB, Zhu H, Baumgarten AM, Spangler R, May G, Cook DR and Young ND (2002) Diversity, distribution, and ancient taxonomic relationships within the TIR and non-TIR NBS-LRR resistance gene subfamilies. Journal of Molecular Evolution 54: 548–562
- Cardinal AJ, Burton JW, Camacho-Roger AM, Yang JH, Wilson RF and Dewey RE (2007) Molecular analysis of Soybean lines with low palmitic acid content in the seed oil. Crop Science 47: 304–310
- Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D and Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156: 847−854
- Carvalho CR, Clarindo WR, Pracxa MM, Arau´jo FS and Carels N (2008) Genome size, base composition and karyotype of *Jatropha curcas* L., an important biofuel plant. Plant Science 174: 613–617
- Cases S, Smith SJ, Zheng YW, Myers HM, Lear SR, Sande E, Novak S, Collins C, Welch CB, Lusis AJ, *et al.* (1998) Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. Proceeding of National Academy of Sciences USA 95: 13018–13023
- Castillo A, Budak H, Varshney RK, Dorado G, Graner A and Hernandez P (2008) Transferability and polymorphism of barley EST-SSR markers used for phylogenetic analysis in *Hordeum chilense.* BMC Plant Biology 8: 97-105

- Causse M, Damidaux R and Rousselle P (2007) Traditional and enhanced breeding for quality traits in tomato. In: Genetic improvement of Solanaceous crops; tomato Enfield, NH: Science Publishers 2: 153-192
- Cavagnaro PF, Senalik DA, Yang L, Simon PW, Harkins TT, Kodira CD, Huang S and Weng Y (2010) Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). BMC Genomics 15(11): 569-587
- Chan AP**,** Crabtree J, Zhao Q**,** Lorenzi H, Orvis J, Puiu D, Melake-Berhan A, Jones KM, Redman J, Chen G, *et al*. (2010) Draft genome sequence of the oilseed species *Ricinus communis*. Nature Biotechnology 28(9): 951-959
- Chang PH, Zhong YP, Tae WK, Jacqueline B, Tae-Jin Y, Yoon-Kang H, Jong B, Beom-Seok P, David E and Yong PL (2007) Genomic Distribution of Simple Sequence Repeats in *Brassica rapa*. Molecular Cells 23: 349-356
- Chen G, Pan D, Zhou Y, Lin S and Ke X (2007) Diversity and evolutionary relationship of nucleotide binding site-encoding disease resistance gene analogues in sweet potato (*Ipomoea batatas* Lam.). Journal of Biosciences 32: 713–721
- Chin ECL, Senior ML, Shu H and Smith JSC (1996) Maize simple repetitive DNA sequences: abundance and allelic variation. Genome 39: 866–873
- Comadran J, Ramsay L, MacKenzie K, Hayes P, Close TJ, Muehlbauer G, Stein N and Waugh R (2010) Patterns of polymorphism and linkage disequilibrium in cultivated barley. Theoretical and Applied Genetics 122(3): 523-531
- Cordero JC and Skinner DZ (2002) Isolation from alfalfa of resistance gene analogues containing nucleotide binding sites. Theoretical and Applied Genetics 104: 1283–1289

- Costa GGL, Cardoso KC, Del Bem LEV, Lima AC, Cunha MAS, Campos-Leite lD, Vicentini R, Papes F, Moreira RC, Yunes JA, Campos FAP and Da Silva MJ (2010) Transcriptome analysis of the oil-rich seed of the bioenergy crop *Jatropha curcas* L. BMC Genomics 11: 462-478
- Cuesta-Marcos A, Szőcs P, Close TJ, Filichkin T, Muehlbauer GJ, Smith KP and Hayes PM (2010) Genome-wide SNPs and re-sequencing of growth habit and inflorescence genes in barley: implications for association mapping in germplasm arrays varying in size and structure. BMC Genomics 11: 707-716
- Dangl JL and Jones JD (2001) Plant pathogens and integrated defense responses to infection. Nature 411: 826–833
- de Vienne D, Leonardi A, Damerval C and Zivy M (1999) Genetics of proteome variation for QTL characterization: application to drought-stress responses in maize. Journal of Experimental Botany 50: 303-309
- Decroocq V, Fave MG, Hagen L, Bordenave L and Decroocq S (2003) Development and transferability of apricot and grape microsatellite markers across taxa. Theoretical and Applied Genetics 106: 912–922
- Dehgan B (1984) Phylogenetic significance of interspecific hybridization in *Jatropha* (Euphorbiaceae). Systematic Botany 9: 467-478
- Deshmukh R, Singh A, Jain N, Anand S, Gacche R, Singh A, Gaikwad K, Sharma T, Mohapatra T and Singh N (2010) Identification of candidate genes for grain number in rice (*Oryza sativa* L.). Functional and Integrative Genomics 10(3): 339-347
- DeYoung BJ and Innes RW (2006) Plant NBS-LRR proteins in pathogen sensing and host defense. Nature Immunology 7: 1243–1249

- Dimov Z and Moller C (2010) Genetic variation for saturated fatty acid content in a collection of European winter oilseed rape material (*Brassica napus*). Plant Breeding 129: 82-86
- Divakara BN, Upadhyaya HD, Wani SP and Laxmipathi Gowda CL (2010) Biology and genetic improvement of *Jatropha curcas* L.: A review. Applied Energy 87: 732-742
- Domann P, Voelker TA and Ohlrogge JB (2000) Accumulation of palmitate in *Arabidopsis* mediated by the acyl-acyl carrier protein thioesterase *FATB1*. Plant Physiology 123: 637–643
- Dubey S, Stephenson P, Levy DE, Miller JA, Keller SM, Schiller JH, Johnson DH and Kolesar JM (2006) EGFR dinucleotide repeat polymorphism as a prognostic indicator in non-small cell lung cancer. Journal of Thoracic Oncology 5: 395-397
- Durand J, Bodénès C, Chancerel E, Frigerio JM, Vendramin G, Sebastiani F, Buonamici A, Gailing O, Koelewijn HP, Villani F, Mattioni C, Cherubini M, Goicoechea PG, Herrán A, Ikaran Z, Cabané C, Ueno S, Alberto F, Dumoulin PY, Guichoux E, de Daruvar A, Kremer A and Plomion C (2010) A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. BMC Genomics 11: 570-583
- Durrett TP, Benning C and Ohlrogge J (2008) Plant triacylglycerols as feedstocks for the production of biofuels. Plant Journal 54: 593-607
- Dyer JM, Stymne S, Green AG and Carlsson AS (2008) High-value oils from plants. Plant Journal 54: 640-655
- Edwards YJ, Elgar G, Clark MS and Bishop MJ (1998) The identification and characterization of microsatellites in the compact genome of the Japanese pufferfish,

Fugu rubripes: perspectives in functional and comparative genomic analyses. Journal of Molecular Biology 278: 843–854

- Ellwood SR, Phan HTT, Jordan M, Hane J, Torres AM, Avila CM, Serafin Cruz-Izquierdo S and Oliver RP (2008) Construction of a comparative genetic map in faba bean (*Vicia faba* L.); conservation of genome structure with *Lens culinaris*. BMC Genomics 9: 380-391
- Eujayl I, Sledge MK, Wang L, May GD, Chekhovskiy K, Zwonitzer JC and Mian MAR (2004) *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. Theoretical and Applied Genetics 108: 414–422
- Fairless D (2007) Biofuel: the little shrub that could: maybe. Nature 499: 652–655
- Ferguson ME, Burow MD, Schultz SR, Bramel PJ, Paterson AH, Kresovich S and Mitchell S (2004) Microsatellite identification and characterization in peanut (*A. hypogaea* L.). Theoretical and Applied Genetics 108: 1064-1070
- Fujimori S, Washio T, Higo K, Ohtomo Y, Murakami K, Matsubara K, Kawai J, Carninci P, Hayashizaki Y, Kikuchi S and Tomita M (2003) A novel feature of microsatellites in plants: a distribution gradient along the direction of transcription. FEBS Letters 554: 17–22
- Fulton TM, Van der Hoeven R, Eannetta NT and Tanksley SD (2002) Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14: 1457–1467
- Ganesh Ram S, Parthiban KT, Kumar RS, Thiruvengadam V and Paramathma M (2007) Genetic diversity among Jatropha species as revealed by RAPD markers. Genetic Resources and Crop Evolution 55: 803–809

- Gao Y, Xu Z, Jiao F, Yu H, Xiao B, Li Y and Lu X (2010) Cloning, structural features, and expression analysis of resistance gene analogs in tobacco. Molecular Biology Reporters 37(1): 345-354
- Ginwal HS, Phartyal SS, Rawat PS and Srivastava RL (2005) Seed source variation in morphology, germination and seedling growth of *Jatropha crucas* Linn. in Central India. Silvae Genetics 54(2): 76–80
- Glynn NC, Comstock JC, Sood SG, Dang PM and Chaparro JX (2008) Isolation of nucleotide binding site-leucine rich repeat and kinase resistance gene analogues from sugarcane (*Saccharum* spp.). Pest Managment Science 64: 48–56
- Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, *et al.* (2002) A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *japonica*). Science 296: 92-100
- Gong YM, Xu SC, Mao WH, Hu QZ, Zhang GW, Ding J and Li YD (2010) Developing new SSR markers from ESTs of pea (*Pisum sativum* L.). Journal of Zhejiang University Science Bulletin 11(9): 702-707
- Goudet J (1995) FSTAT (Version 1.2): a computer program to calculate F-statistics. Journal of Heredity 86: 485-489
- Grativol C, Lira-Medeiros CdF, Hemerly AS and Ferreira PCG (2010) High efficiency and reliability of inter-simple sequence repeats (ISSR) markers for evaluation of genetic diversity in Brazilian cultivated *Jatropha curcas* L. accessions. Molecular Biology Reporters 36: 547-554
- Gubitz GM, Mittelbach M and Trabi M (1998) Exploitation of the tropical oil seed plant *Jatropha curcas* L. Bioresources Technology 67: 73-79

- Gupta PK and Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113: 163-185
- Ha BK, Monteros MJ and Boerma HR (2010) Development of SNP assays associated with oleic acid QTLs in N00-3350 soybean. Euphytica 176: 403–415
- Hayes AJ and Saghai Maroof MA (2000) Targeted resistance gene mapping in soybean using modified AFLPs. Theoretical and Applied Genetics 100: 1279-1283
- Heesacker A, Venkata KK, Gao W, Tang S, Kolkman JM, Gingle A, Matvienko M, Kozik A, Michelmore RM, Lai Z, Rieseberg LH and Knapp SJ (2009) SSRs and INDELs mined from the sunflower EST database: abundance, polymorphisms, and cross-taxa utility. Theoretical and Applied Genetics 117: 1021-1029
- Hitz WD, Carlson TJ, Booth JR, Kinney AJ, Stecca KL and Yadav NS (1994) Cloning of a higher-plant plastid omega-6 fatty acid desaturase cDNA and its expression in a cyanobacterium. Plant Physiology 105(2): 635-641
- Hong CP, Plaha P, Koo DH, Yang TJ, Choi SR, *et al.* (2006) A survery of the *Brassica rapa* Genome by BAC-end sequence analysis and comparison with *Arabidopsis thaliana*. Molecular Cells 22: 300−307
- Hu X, Gilbert MS, Gupta M and Thompson SA (2006) Mapping of the loci controlling oleic and linolenic acid contents and development of *fad2* and *fad3* allele-specific markers in canola (*Brassica napus* L.). Theoretical and Applied Genetics 113: 497–507
- Hu X, Gilbert MS, Gupta M and Thompson SA (2007) G-to-A mutation at a 5_ splice site of *fad3c* caused impaired splicing in a low linolenic mutant of canola (*Brassica napus* L.). Plant Biotechnology 24: 397–400

- Hulbert SH, Webb CA, Smith SM and Sun Q (2001) Resistance gene complexes: evolution and utilization. Annual Review of Phytopathology 39: 285–312
- Iba K, Gibson S, Nishiuchi T, Fuse T, Nishimura M, Arondel V, Hugly S and Somerville C (1993) A gene encoding a chloroplast omega-3 fatty acid desaturase complements alterations in fatty acid desaturation and chloroplast copy number of the fad7 mutant of *Arabidopsis thaliana*. Journal of Biological Chemistry 268(32): 24099-24105
- Jaccard P (1908) Nouvelles recherches sur la distribution florale. Société Vaudoise des Sciences Naturelles 44: 223–270
- Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, *et al.* (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467
- Jones A, Davies HM and Voelker TA (1995) Palmitoyl-acyl carrier protein (ACP) thioesterase and the evolutionary origin of plant acyl-ACP thioesterases. The Plant Cell 7: 359-371
- Joshi RK, Mohanty S, Subudhi E and Nayak S (2010) Isolation and characterization of NBS-LRR- resistance gene candidates in turmeric (*Curcuma longa* cv. surama). Genetics and Molecular Research 9(3):1796-1806
- Jurka J and Pethiyagoda C (1995) Simple repetitive DNA sequences from primates: compilation and analysis. Journal of Molecular Evolution 40: 120–126
- Kachroo A, Shanklin J, Whittle E, Lapchyk L, Hildebrand D and Kachroo P (2006) The *Arabidopsis* stearoyl-acyl carrier protein-desaturase family and the contribution of leaf isoforms to oleic acid synthesis. Plant Molecular Biology 63: 257-271

- Kanazin V, Mareck L and Shoemaker P (1996) Resistance gene analogs are conserved and clustered in soybean. Proceedings of National Academy of Science USA 93: 11746-11750
- Kantety RV, La Rota M, Matthews DE and Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Molecular Biology 48: 501–510
- Kashi Y and King DG (2006) Simple sequence repeats as advantageous mutators in evolution. Trends in Genetics 22(5): 253-259
- Katti MV, Ranjekar PK and Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. Molecular Biology and Evolution 18: 1161−1167
- Keller B and Geuillet C (2000) Collinearity and gene density in grass genomes. Trends Plant Science 5: 246–251
- Kinney AJ (1997) Genetic engineering of oilseeds for desired traits. In J.K. Setlow (Ed.), Genetic engineering: Vol. 19 (pp. 149-166). New York: Plenum Press
- Klopfenstein WE (1985) Effect of molecular-weights of fatty-acid esters on cetane numbers as diesel fuels. Journal of American Oil Chemists' Society 62: 1029–1031
- Knothe G (2005) Dependence of biodiesel fuel properties on the structure of fatty acid alkyl esters. Fuel Process Technology 86: 1059–1061
- Knothe G and Dunn RO (2003) Dependence of oil stability index of fatty compounds on their structure and concentration and presence of metals. Journal of American Oil Chemists' Society 80: 1021–1026

- Knutzon DS, Bleibaum JL, Nelsen J, Kridl JC and Thompson GA (1992) Isolation and characterization of two safflower oleoyl-acyl carrier protein thioesterase cDNA clones. Plant Physiology 100: 1751-1758
- Kohler A, Rinaldi C, Duplessis S, Baucher M, Geelen D, Duchaussoy F, Meyers BC, Boerjan W and Martin F (2008) Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. Plant Molecular Biology 66: 619–636
- Kong J and Gao H (2005) Analysis of tandem repeats in the genome of Chinese shrimp *Fenneropenaeus chinensis*. Chinese Science Bulletin 50(14): 1462-1469
- Kumpatla SP and Mukhopadhyay S (2005) Miming and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. Genome 48: 985–998
- Lacombe S and Bervillé A (2001) A dominant mutation for high oleic acid content in sunflower (*Helianthus annuus* L.) seed oil is genetically linked to a single oleate-desaturase RFLP locus. Molecular Breeding 8: 129-137
- Lai Y and Sun F (2003) The relationship between microsatellite slippage mutation rate and the number of repeat units. Molecular Biology Evolution 20(12): 2123-2131
- Lardizabal KD, Mai JT, Wagner NW, Wyrick A, Voelker T and Hawkins DJ (2001) DGAT2 is a new diacylglycerol acyltransferase gene family: purification, cloning, and expression in insect cells of two polypeptides from *Mortierella ramanniana* with diacylglycerol acyltransferase activity. Journal of Biological Chemistry 276: 38862–38869
- Largercrantz U, Ellegren H and Andersson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Research 21: 1111-1115

- Lawson MJ and Zhang L (2006) Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. Genome Biology 7: 1-11
- Leister D, Ballvora A, Salamini F and Gebhardt C (1996) A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. Nature Genetics 14: 421-429
- Leister D, Kurth J, Laurie D, Yano M, Sasaki T, Devos K, Graner A and Schulze- Lefert P (1998) Rapid organization of resistance gene homologues in cereal genomes. Proceeding of National Academy of Sciences USA 95: 370-375
- Li X, Shangguan L, Song C, Wang C, Gao Z, Yu H and Fang J (2010) Analysis of expressed sequence tags from *Prunus mume* flower and fruit and development of simple sequence repeat markers. BMC Genetics 11: 66-78
- Li Y, Beisson F, Pollard M and Ohlrogge J (2006) Oil content of Arabidopsis seeds: the influence of seed anatomy, light and plant-to-plant variation. *Phytochemistry* 67: 904–915
- Li YC, Korol AB, Fahima T and Nevo E (2004) Microsatellites within genes: structure, function, and evolution. Molecular Biology and Evolution 21: 991−1007
- Li YC, Korol AB, Fahima T, Beiles A and Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Molecular Ecology 11: 2453−2465
- Li ZK (2000) QTL mapping in rice: a few critical considerations. In Proc. Fourth Int. Rice Genetics Symp. (eds G. S. Khush, D. S. Brar & B. Hardy), pp. 153–171. Los Ban˜os, The Philippines: International Rice Research Institute

- Lin J, Chen Y, Xu Y, Yan F, Tang L and Chen F (2003) Cloning and expression of curcin, a ribosome-inactivating protein from the seeds of *Jatropha curcas*. Acta Botanica Sinica 45: 858-863
- Liu K and Muse S (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21: 2128–2134
- Lopes MS and Maciel GB (2006) Isolation and characterization of simple sequence repeat loci in *Rubus hochstetterorum* and their use in other species from the Rosaceae family. Molecular Ecology 6(3): 750–752
- López Y, Nadaf HL, Smith OD, Connell JP, Reddy AS and Fritz AK (2000) Isolation and characterization of the Δ12-fatty acid desaturase in peanut (*Arachis hypogaea* L.) and search for polymorphisms for the high oleate trait in Spanish market-type lines. Theoretical and Applied Genetics 101: 1131–1138
- Love J, Knight A, McAleer M and Todd J (1990) Towards construction of a high-resolution map of the mouse genome using PCR analyzed microsatellites. Nucleic Acids Research 18: 4123-4130
- Luck JE, Lawrence GJ, Dodds PN, Shepherd KW and Ellis JG (2000) Regions outside of the leucine-rich repeats of flax rust resistance proteins play a role in specificity determination. Plant Cell 12: 1367–1377
- Makkar HPS, Becker K, Sporen F and Wink M (1997) Studies on nutritive potential and toxic constituents of different provenances of *Jatropha curcas*. Journal of Agricultural Food Chemistry 45: 3152-3157
- Marcotte EM, Pellegrini M, Yeates TO and Eisenberg D (1999) A census of protein repeats. Journal of Molecular Biology 293: 151–160

- Marques CM, Brondani RPV, Grattapaglia D and Sederoff R (2002) Conservation and synteny of SSR loci and QTLs for vegetative propagation in four *Eucalyptus* species. Theoretical and Applied Genetics 105: 474–478
- Martin GB, Bogdanove AJ and Sessa G (2003) Understanding the functions of plant disease resistance proteins. Annual Review of Plant Biology 54: 23–61
- Martínez-Zamora MG, Castagnaro AP and Díaz-Ricci JC (2004) Isolation and diversity analysis of resistance gene analogues (RGAs) from cultivated and wild strawberries. Molecular Genetics and Genomics 272: 480-487
- Mayer KM and Shanklin J (2007) Identification of amino acid residues involved in substrate specificity of plant acyl-ACP thioesterases using a bioinformatics-guided approach. BMC Plant Biology 7: 1-15
- McCormick RL, Graboski MS, Alleman TL and Herring AM (2001) Impact of biodiesel source material and chemical structure on emissions of criteria pollutants from a heavy-duty engine. Environmental Science and Technology 35: 1742–1747
- McHale L, Tan X, Koehl P and Michelmore RW (2006) Plant NBS-LRR proteins: adaptable guards. Genome Biology 7: 212-222
- Metzgar D, Bytof J and Wills C (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Research 10:72-80
- Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW and Young ND (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant Journal 20: 317–332

- Meyers BC, Kozik A, Griego A, Kuang H and Michelmore RW (2003) Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. Plant Cell 15(4): 809-834
- Milbourne D, Meyer R, Bradshaw JE, Baird E, Provan J, Powell W and Waugh R (1997) Comparison of PCR- based marker systems for the analysis of genetic relationship in cultivated potato. Molecular breeding 3: 127-136
- Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, *et al.* (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). Nature 452: 991–996
- Monosi B, Wisser RJ, Pennill L and Hulbert SH (2004) Full-genome analysis of resistance gene homologues in rice. Theoretical and Applied Genetics 109: 1434–1447
- Moran C (1993) Microsatellite repeats in pig (*Sus domestica*) and chicken (*Gallus domesticus*) genomes. Journal of Heredity 84: 274–280
- Moretzsohn MC, Leoi L, Proite K, Guimaraes PM, Leal-Bertioli SCM, Gimanes MA, Martin WS, Valls JFM, Grattapaglia D and Bertioli D (2005) A microsatellite – based, gene-rich linkage map for the AA genome of *Arachis* (Fabaceae). Theoretical and Applied Genetics 111: 1060-1071
- Morgante M and Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. Plant Journal 3: 175-182
- Morgante M, Hanafey M and Powell W (2002) Microsatellites are preferentially associated with non repetitive DNA in plant genomes. Nature Genetics 30: 194−200
- Mortimer JC, Batley J, Love CG, Logan E and Edwards D (2005) Simple sequence repeat (SSR) and GC distribution in the *Arabidopsis thaliana* genome. Journal of Plant Biotechnology 7: 17−25

- Murray MG and Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Research 8: 4321–4325
- Nair RA and Thomas G (2007) Isolation, characterization and expression studies of resistance gene candidates (RGCs) from zingiber spp. Theoretical and Applied Genetics 116: 123-134
- Nikolau BJ, Ohlrogge JB and Wurtele ES (2003) Plant biotin containing carboxylases. Archives of Biochemistry and Biophysics 414: 211–222
- Noir S, Combes MC, Anthony F and Lashermes P (2001) Origin, diversity and evolution of NBS-type disease-resistance gene homologues in coffee trees (*Coffea* L.). Molecular Genetics and Genomics 265: 654-662
- Okuley J, Lightner J, Feldmann K, Yadav N, Lark E and Browse J (1994) *Arabidopsis* FAD2 gene encodes the enzyme that is essential for polyunsaturated lipid synthesis. Plant Cell 6: 147-158
- Ou WJ, Wang WQ and Li KM (2009) Molecular Genetic Diversity Analysis of 120 Accessions *Jatropha curcas* L. Germplasm. Chinese Journal of Tropical Crops 30: 287-292
- Palomino C, Satovic Z, Cubero JI and Torres AM (2006) Identification and characterization of NBS-LRR class resistance gene analogs in faba bean (*Vicia faba* L.) and chickpea (*Cicer arietinum* L.). Genome 49: 1227–1237
- Pamidimarri D, Singh S, Mastan SG, Patel J and Reddy MP (2009) Molecular characterization and identification of markers for toxic and non-toxic varieties of *Jatropha curcas* L. using RAPD, AFLP and SSR markers. Molecular Biology Reports 36(6): 1357-1364

- Pan Q, Wendel J and Fluhr R (2000) Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. Journal of Molecular Evolution 50: 203–213
- Passarge E, Horsthemke B and Farber RA (1999) Incorrect use of the term synteny. Nature Genetics 23: 387-389
- Patil V and Singh K (1991) Oil gloom to oil boom – *Jatropha curcas* a promising agro-forestry crop. Shree Offset Press, Nashik
- Peakall R and Smouse PE (2006) GENALEX 6: genetic analysis in excel population genetic software for teaching and research. Molecular Ecology Notes 6: 288–295
- Pérez-Vich B, Leon AJ, Grondona M, Velasco L and Fernández-Martínez JM (2006) Molecular analysis of the high stearic acid content in sunflower mutant CAS-14. Theoretical and Applied Genetics 112(5): 867-875
- Pollard MR, Anderson L, Fan C, Hawkins DJ and Davies HM (1991) A specific acyl-ACP thioesterase implicated in medium chain fatty-acid production in immature cotyledons of *Umbellularia californica*. Archives of Biochemistry and Biophysics 284: 306–312
- Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, de Kochko A and Hamon P (2006) SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. Molecular Genetics and Genomics 276: 436–449
- Porter BW, Paidi M, Ming R, Alam M, Nishijima WT and Zhu YJ (2009) Genome-wide analysis of *Carica papaya* reveals a small NBS resistance gene family. Molecular Genetics and Genomics 281: 609–626

- Powell W, Machray GC and Provan J (1996) Polymorphisms revealved by simple sequence repeats. Trends in Plant Science 1: 215-222
- Prasad MD, Muthulakshmi M, Madhu M, Archak S, Mita K and Nagaraju J (2005) Survey and analysis of microsatellites in the silkworm, *Bombyx mori*: frequency, distribution, mutations, marker potential and their conservation in heterologous species. Genetics 169(1): 197-214
- Pritchard JK, Stephens M and Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945–959
- Radwan O, Gandhi S, Heesacker A, Whitaker B, Taylor C, Plocik A, Kesseli R, Kozik A, Michelmore RW and Knapp SJ (2008) Genetic diversity and genomic distribution of homologs encoding NBS-LRR disease resistance proteins in sunflower. Molecular Genetics and Genomics 280: 111–125
- Ramos LCDS, Tango JS and Savi A (1984) Variability for oil and fatty acid composition in castor bean varieties. Journal of the American Oil Chemists' Society 16: 1841-1843
- Reed DW, Schafer UA and Covello PS (2000) Characterization of the *Brassica napus* extraplastidial linoleate desaturase by expression in *Saccharomyces cerevisiae*. Plant Physiology 122(3): 715-720
- Riva GA, Cyril JC, Yuval E, Shelef L, Hallerman EM and Kashi Y (2000) Simple sequence repeats in *Escherichia coli*: abundance, distribution, composition, and polymorphism. Genome Research 10: 62-71
- Rob C and Durka W (2006) Isolation and characterization of microsatellite markers in the invasive shrub *Mahonia aquifolium* (Berberidaceae) and their applicability in related species. Molecular Ecology 6: 948–950

- Robinson AJ, Love CG, Batley J, Barker G and Edwards D (2004) Simple sequence repeat marker loci discovery using SSR primer. Bioinformatics 20: 1475–1476
- Roesler K, Shintani D, Savage L, Boddupalli S and Ohlrogge J (1997) Targeting of the Arabidopsis homomeric acetyl-coenzyme A carboxylase to plastids of rapeseeds. Plant Physiology 113: 75–81
- Rohlf FJ (2005) NTSYS-pc numerical taxonomy and multivariate analysis system, version 2.2. Exeter Software, New York, USA: Exeter Software
- Rota L M, Kantety RV, Yu JK and Sorrells ME (2005) Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. BMC Genomics 6: 23-31
- Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GR, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W, *et al.* (2000) Comparative genomics of the eukaryotes. Science 287: 2204–2215
- Salas JJ and Ohlrogge JB (2002) Characterization of substrate specificity of plant FatA and FatB acyl-ACP thioesterases. Archives of Biochemistry and Biophysics 403: 25–34
- Samadi S, Artiguebielle E, Estoup A, Pointier JP, Silvain JF, Heller J, Cariou ML and Jarne P (1998) Density and variability of dinucleotide microsatellites in the pathogenetic polylpoid snail *Melanoides tuberculata*. Molecular Ecology 7: 1233-1236
- Sánchez-García A, Moreno-Pérez AJ, Muro-Pastor AM, Salas JJ, Garcés R and Martínez-Force E (2010) Acyl-ACP thioesterases from castor (*Ricinus communis* L.): An enzymatic system appropriate for high rates of oil synthesis and accumulation. Phytochemistry 71: 860- 869

- Sangwan NK, Gupta K and Dhidsa KS (1986) Fatty acid composition of developing soybeans. Journal of Agricultural Food Chemistry 34: 415-417
- Sato S, Hirakawa H, Isobe S, Fukai E, Watanabe A, Kato M, Kawashima K, Minami C, Muraki A, Nakazaki N, Takahashi C, Nakayama S, Kishida Y, Kohara M,Yamada M, Tsuruoka H, Sasamoto S, Tabata S, Aizu T, Toyoda A, Shin-I T, Minakuchi Y, Kohara Y, Fujiyama A, Tsuchimoto S, Kajiyama S, Makigano E, Ohmido N, Shibagaki N, Cartagena JA, Wada N, Kohinata T, Atefeh A, Yuasa S, Matsunaga S and Fukui K (2011) Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. DNA Research 18: 65–76
- Schlotterer C (2000) Evolutionary dynamics of microsatellite DNA. Chromosoma 109: 365−371
- Senior ML and Heun M (1993) Mapping maize microsatellites and polymerase chain reaction confirmation of the targeted repeats using a CT primer. Genome 36: 884–889
- Senthilvel S, Jayashree B, Mahalakshmi V,Sathish Kumar P, Nakka S, Nepolean T and Hash CT (2008) Development and mapping of Simple Sequence Repeat markers for pearl millet from data mining of Expressed Sequence Tags. BMC Plant Biology 8: 119-128
- Serdari A, Lois E and Stournas S (1999) Impact of esters of mono- and dicarboxylic acids on diesel fuel quality. Industrial & Engineering Chemistry Research 38: 3543–3548
- Sharma A and Chauhan RS (2008) Identification of candidate gene-based markers (SNPs and SSRs) in the zinc and iron transporter sequences of maize (*Zea mays* L.). Current Science 95: 1051-1059
- Shen K, Meyers B, Islam-Faridi M, Chin D, Stelly D and Michelmore R (1998) Resistance gene candidates identified by PCR with degenerate oligonucleotide primers

map to clusters of resistance genes in lettuce. Molecular Plant Microbe Interaction 11: 815-823

- Shin JH, Van K, Kim DH, Kim KD, Jang YE, Choi BS, Kim MY and Lee SH (2008) The lipoxygenase gene family: a genomic fossil of shared polyploidy between *Glycine max* and *Medicago truncatula*. BMC Plant Biology 8: 133-144
- Singh R, Tan SG, Panandam JM, Rahman RA, Ooi LCL, Low ETL, Sharma M, Jansen J and Cheah SC (2009) Mapping quantitative trait loci (QTLs) for fatty acid composition in an interspecific cross of oil palm. BMC Plant Biology 9: 114-133
- Sneath PHA and Sokal RR (1973) Numerical taxonomy- The principles and practice of numerical classification. Freeman Press, San Francisco, California, USA, 573
- Soriano JM, Vilanova S, Romero C, Llácer G and Badenes ML (2005) Characterization and mapping of NBS-LRR resistance gene analogs in apricot (*Prunus armeniaca* L.). Theoretical and Applied Genetics 110: 980–989
- Sorrells ME, Rota ML, Bermudez-Kandianis CE, *et al.* (2003) Comparative DNA sequence analysis of wheat and rice genomes. Genome Research 13: 1818-1827
- Sosinski B, Gannavarapu M, Hager LD, Beck LE, King GJ, Ryder CD, Rajapakse S, Baird WV, Ballard RE and Abbott AG (2000) Characterization of microsatellite markers in peach [*Prunus persica* (L.) Batsch]. Theoretical and Applied Genetics 101: 421-428
- Spencer MM, Pantalone VR, Meyer EJ, Landau-Ellis D and Hyten DL (2003) Mapping the *Fas* locus controlling stearic acid content in soybean. Theoretical and Applied Genetics 106(4): 615-624

- Srivastava R (1999) Study in variation in morpho-physiological parameters with reference to oil yield and quality in *Jatropha curcas* Linn. Ph.D. Thesis, Forest Research Institute, Dehra Dun, India
- Stournas S, Lois E and Serdari A (1995) Effects of fatty-acid derivatives on the ignition quality and cold flow of diesel fuel. Journal of American Oil Chemists' Society 72: 433–437
- Streelman JT and Kocher TD (2002) Microsatellite variation associated with prolactin expression and growth of salt-challenged tilapia. Physiological Genomics 9: 1–4
- Subirana JA and Messeguer X (2008) Structural families of genomic microsatellites. Gene 408(1-2): 124-132
- Sujatha M and Prabakaran AJ (1997) Characterization and utilization of Indian Jatropha. Indian Journal of Plant Genetic Resources 10(1): 123–128
- Sun QB, Li LF, Li Y, Wu GJ and Ge XJ (2008) SSR and AFLP markers reveal low genetic diversity in the biofuel plant *Jatropha curcas* in China. Crop Science 48(5): 1865–1871
- Swanson T (1996) Global values of biological diversity: the public interest in the conservation of plant genetic resources for agriculture. Plant Genetics Resources Newsletter 105: 1–3
- Tameling WI, Vossen JH, Albrecht M, Lengauer T, Berden JA, Haring MA, Cornelissen BJ and Takken FL (2006) Mutations in the NBARC domain of I-2 that impair ATP hydrolysis cause autoactivation. Plant Physiology 140: 1233–1245

- Tan X, Meyers BC, Kozik A, West MA, Morgante M, St Clair DA, Bent AF and Michelmore RW (2007) Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in Arabidopsis. BMC Plant Biology 7: 56-62
- Tanhuanpää P, Vilkki J and Vihinen M (1998) Mapping and cloning of FAD2 gene to develop allele-specific PCR for oleic acid in spring turnip rape (*Brassica rapa* ssp. *oleifera*). Molecular Breeding 4: 543–550
- Taramino G and Tingey S (1996) Simple sequence repeats for germplasm analysis and mapping in maize. Genome 39: 277–287
- Tatikonda L, Wani SP, Kannan S, Beerelli N, Sreedevi TK, Hoisington DA, Devi P and Varshney RK (2009). AFLP-based molecular characterization of an elite germplasm collection of *Jatropha curcas* L., a biofuel plant. Plant Science 176: 505–513
- Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S and McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. Genome Research 11: 1441–1452
- Thelen JJ and Ohlrogge JB (2002) Metabolic engineering of fatty acid biosynthesis in plants. Metabolic Engineering 4: 12–21
- Thiel T, Michalek W, Varshney RK and Graner J (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theoretical and Applied Genetics 106: 411-422
- Tian Z, Qian Q, Liu Q, Yan M, Liu X, Yan C, Liu G, Gao Z, Tang S, Zeng D, Wang Y, Yu J, Gu M and Li J (2009) Allelic diversities in rice starch biosynthesis lead to a diverse

array of rice eating and cooking qualities. Proceeding of National Academy of Sciences USA 106: 21760-21765

- Tóth G, Gáspári Z and Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. Genome Research 10: 967-981
- Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, *et al.* (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313: 1596–1604
- van der Biezen EA and Jones JD (1998) The NB-ARC domain: a novel signaling motif shared by plant resistance gene products and regulators of cell death in animals. Current Biology 8: 226–229
- Van der Linden CG, Wouters DCAE, Mihalka V, Kochieva EZ, Smulders MJM and Vosman B (2004) Efficient targeting of plant disease resistance loci using NBS profiling. Theoretical and Applied Genetics 109: 384-393
- van Lith HA and van Zutphen LF (1996) Characterization of rabbit DNA microsatellites extracted from the EMBL nucleotide sequence database. Animal Genetics 27: 387–395
- van Ooijen G, van den Burg HA, Cornelissen BJ and Takken FL (2007) Structure and function of resistance proteins in solanaceous plants. Annual Review of Phytopathology 45: 43–72
- Varshney RK, Graner A and Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends in Biotechnology 23: 48-55
- Velasco L and Becker HC (1998) Estimating the fatty acid composition of the oil in intact seed-rapeseed (*Brassica napus* L.) by near-infrared reflectance spectroscopy. Euphytica 101: 221-230

- Venegas-Calero´ n M, Sayanova O and Napier JA (2009) An alternative to fish oils: metabolic engineering of oil-seed crops to produce omega-3 long chain polyunsaturated fatty acids. Progess in Lipid Research 49 (2): 108-119
- Voelker TA, Hayes TR, Cranmer AM, *et al.* (1996) Genetic engineering of a quantitative trait: Metabolic and genetic parameters influencing the accumulation of laurate in rapeseed. Plant Journal 9: 229–241
- Wan H, Zhao Z, Malik AA, Qian C and Chen J (2010) Identification and characterization of potential NBS-encoding resistance genes and induction kinetics of a putative candidate gene associated with downy mildew resistance in *Cucumis*. BMC Plant Biology 10: 186-192
- Wang Z, Weber JL, Zhong G and Tanksley (1994) SD Survey of plant short tandem DNA repeats. Theoretical and Applied Genetics 88: 1-6
- Weir BS (1999) Genetic data analysis. Methods for discrete genetic data. Sunderland Massachusetts: Sinauer Associates, 377
- Wen M, Wang H, Xia Z, Zou M, Lu C and Wang W (2010) Development of EST-SSR and genomic-SSR markers to assess genetic diversity in *Jatropha curcas* L. BMC Research Notes 3: 42-49
- Westman AL and Kresovich S (1998) The potential for cross-taxa simple-sequence repeat (SSR) amplification between *Arabidopsis thaliana* L. and crop *Brassica*s. Theoretical and Applied Genetics 96: 272-281
- Wilson LM, Whitt SR, Ibanez AM, Rocheford TR, Goodman MM and Buckler ESI (2004) Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell 16: 2719–2733

- Wolfe KH, Gouy M, Yang YW, Sharp PM and Li WH (1989) Date of the monocot–dicot divergence estimated from chloroplast DNA sequence data. Proceeding of National Academy of Sciences USA 86: 6201–6205
- Wu G, Wu Y, Xiao L, Li X and Lu C (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
- Xiao X, Li H and Tang C (2008) A Silver-Staining cDNA-AFLP Protocol Suitable for Transcript Profiling in the Latex of *Hevea brasiliensis* (Para Rubber Tree). Molecular Biotechnology 42(1): 91-99
- Xu Q, Wen X and Deng X (2008) Genomic organization, rapid evolution and meiotic instability of nucleotide-binding-site-encoding genes in a new fruit crop, "chestnut rose". Genetics 178: 2081– 2091
- Yang X, Guo Y, Yan J, Zhang J, Song T, Rocheford T and Li JS (2010) Major and minor QTL and epistasis contribute to fatty acid compositions and oil concentration in high-oil maize. Theoretical and Applied Genetics 120: 665–678
- Yang S, Gu T, Pan C, Feng Z, Ding J, Hang Y, Chen JQ and Tian D (2008a) Genetic variation of NBS-LRR class resistance genes in rice lines. Theoretical and Applied Genetics 116: 165–177
- Yang S, Zhang X, Yue JX, Tian D and Chen JQ (2008b) Recent duplications dominate NBS-encoding gene expansion in two woody species. Molecular Genetics and Genomics 280: 187–198
- Yu J, Hu S, Wang J, *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296: 79-92

- Yu J, Pressoir G, Briggs W, Vroh BI, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland J, Kresovich S and Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38: 203–208
- Yuan B, Zhai C, Wang W, Zeng X, Xu X, Hu H, Lin F, Wang L and Pan Q (2011) The Pik-p resistance to *Magnaporthe oryzae* in rice is mediated by a pair of closely linked CC-NBS-LRR genes. Theoretical and Applied Genetics 122(5): 1017-1028
- Zang YX, Kim HU, Kim JA, Lim MH, Jin M, Lee SC, Kwon SJ, Lee SI, Hong JK, Park TH, Mun JH, Seol YJ, Hong SB and Park BS (2009) Genome-wide identification of glucosinolate synthesis genes in *Brassica rapa*. FEBS Journal 276: 3559–3574
- Zhang L, Yuan D, Yu S, Li Z, Cao Y, *et al.* (2004) Preference of simple sequence repeats in coding and non-coding regions of *Arabidopsis thaliana*. Bioinformatics 20: 1081−1086
- Zhang P, Burton JW, Upchurch RG, Whittle E, Shanklin J and Dewey RE (2008) Mutations in a Δ^9–Stearoyl-ACP-Desaturase Gene Are Associated with Enhanced Stearic Acid Levels in Soybean Seeds. Crop Science 48: 2305-2313
- Zhang Y, Ji H, Zhao PX, Bouton JH and Monteros MJ (2008) Genome-wide identification of microsatellites in white clover (*Trifolium repens* L.) using FIASCO and phpSSRMiner. Plant Methods 4: 19-44
- Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P and Nordborg M (2007) An Arabidopsis example of association mapping in structured samples. PLoS Genetics 3(1): 31-34

- Zheng P, Allen WB, Roesler K, Williams ME, Zhang S, Li J, Glassman K, Ranch J, Nubel D, Solawetz W, *et al*. (2008) A phenylalanine in DGAT is a key determinant of oil content and composition in maize. Nature Genetics 40: 367-372
- Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, Shen J and Tian D (2004) Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. Molecular Genetics and Genomics 271: 402–415
- Zhu C, Gore M, Buckler ES and Yu J (2008) Status and Prospects of Association Mapping in Plants. The Plant Genome 1: 5–20
- Zondervan KT and Cardon LR (2004) The complex interplay among factors that influence allelic association. Nature Review Genetics 5: 89–100
- Zonneveld BJM, Leitch IJ and Bennett MD (2005) First nuclear DNA amounts in more than 300 angiosperms. Annals of Botany 96: 229–244