GENOMICS OF SEED OIL BIOSYNTHESIS AND DISEASE RESPONSE COMPONENTS IN *JATROPHA CURCAS* L.

A THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF DOCTOR OF PHILOSOPHY

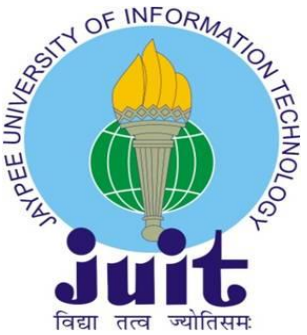
IN

BIOTECHNOLOGY

BY

ARCHIT SOOD

Enrollment No. 106568



JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

WAKNAGHAT

SEPTEMBER, 2016

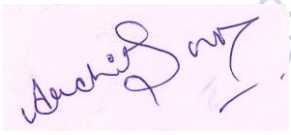
DEDICATED TO
JATROPHA GROWERS
ALL OVER THE WORLD

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY
(Established by H.P. State Legislative Vide Act no. 14 of 2002)
Waknaghat, P.O. Dumehar Bani, Kandaghat, Distt. Solan – 173215 (H.P.) INDIA
Website: www.juit.ac.in
Phone No. (91) 01792-257999(30 Lines).
Fax: (91) 01792 245362

DECLARATION

I certify that:

a. The work contained in this thesis is original and has been done by me under the guidance of my supervisor.
b. The work has not been submitted to any other organisation for any degree or diploma.
c. Wherever, I have used materials (data, analysis, figures or text), I have given due credit by citing them in the text of the thesis.



Archit Sood **Date:** 24.09.2016
Enrollment No. 106568
Department of Biotechnology & Bioinformatics
Jaypee University of Information Technology
Waknaghat, Solan, India

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY
(Established by H.P. State Legislative Vide Act no. 14 of 2002)
Waknaghat, P.O. Dumehar Bani, Kandaghat, Distt. Solan – 173215 (H.P.) INDIA
Website: www.juit.ac.in
Phone No. (91) 01792-257999(30 Lines).
Fax: (91) 01792 245362

CERTIFICATE

This is to certify that the thesis entitled, **"Genomics of Seed Oil Biosynthesis and Disease Response Components in *Jatropha curcas* L."** which is being submitted by **Archit Sood (Enrollment No. 106568)** in fulfillment for the award of degree of **Doctor of Philosophy** in **Biotechnology** at **Jaypee University of Information Technology, Waknaghat, India** is the record of candidate's own work carried out by him under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.

Prof. (Dr.) Rajinder Singh Chauhan (Supervisor) **Date:** 24.09.2016
Dean (Biotechnology) & Head
Department of Biotechnology & Bioinformatics
Jaypee University of Information Technology
Waknaghat, Solan, India

ACKNOWLEDGMENT

Its my privilege and honor to record my gratitude to the following without whose guidance and support I would not have been able to complete my Ph.D thesis.

The insuperable contribution and peer support rendered by my adorable, debonair, dynamic and revered supervisor, ***Prof (Dr) R.S. Chauhan, Dean and HOD, Department of Biotechnology, JUIT*** *is incomprehensible. He enthused inexhaustible inspiration, volition and constructive criticism for generating valuable and innovative ideas for the accomplishment of the present endeavor. I will remain indebted to him for pruning my personality and giving new dimension in the scientific field through his analytical scientific outlook. I salute his great wisdom and personality.*

*With generous humbleness, I thank "****Almighty God****" for bestowing me a supportive family, my parents,* ***Sh. Joginder Kumar Sood****,* ***Smt. Neena Sood*** *and my younger brother,* ***Er. Atul Sood*** *whose extreme support, love and sacrifice is inexplicable which enabled me to pursue and accomplish this Doctoral programme.*

I emphatically extend my sincere thanks to the worthy administration of JUIT, ***Prof (Dr) S.C. Saxena (Acting vice chancellor), Prof (Dr) S.D. Gupta (Director, Academics), Brig. (Retd.) Balbir Singh (Director, Administration), Prof (Dr) T.S. Lamba (Dean, Academics and Research) Prof (Dr) Y. Medury (Former COO, JES), Prof (Dr) Ravi Prakash (Former vice chancellor), Prof (Dr) S.K. Kak (Former vice chancellor)*** *for providing opportunity to pursue a Doctorate Degree, fellowship and lab infrastructure.*

I am also highly beholden to ***Dr (Mrs) Hemant Sood*** *for her valuable suggestions and tremendous cooperation during whole tenure. I am deeply obliged to all the faculty members of Department of Biotechnology and Bioinformatics, especially* ***Dr Suhir Syal and Dr Tiratha Raj Singh*** *and for their guidance and encouragement for easing the present research investigation.*

I owe my grateful thanks to technical staff of the department, ***Mrs. Somlata Sharma, Mrs. Mamta Mishra, Mr. Ravi kant, Mr. Baleshwar, Mr Kamlesh*** *for their cooperation and support, during the course of this study.*

I am also indebted to ***Dr. Sandeep Sharma, Scientist, Himalayan Forest Research Institute (HFRI), Shimla*** *for providing experimental materials or/and farm facilities.*

I am deeply obliged to ***Prof. (Dr) H.K. Chaudhary, Head, Department of Crop Improvement, CSK HP Agricultural University, Palampur*** *for his incessant guidance, whenever I needed.*

No expression of thanks will be sufficient without recognition of help and support rendered by my senior ***Dr. Pankaj Bhardwaj, Assistant Professor, Central University of Punjab, Bathinda****.*

I am also gratified to ***Department of Biotechnology (DBT), Ministry of Science and Technology, Government of India*** *for providing me fellowship during whole tenure of my study.*

This thesis could not have been completed without the support of my friend and roommate ***Mr. Arun Dhiman*** *whose constant help during writing made this a successful venture.*

My special thanks are also due to all my seniors, colleagues and juniors – ***Dr. Varun Jaiswal, Dr Sree Krishna Chanumolu, Nikhil, Jibesh, Tarun Pal, Kirti Shitiz, Anil, Raman, Rohit, Sanjay, Imran, Ashwani, Lalit, Arun Sharma, Vineet, Sita Sharan Patel, Amit, Ankush, Swapnil, Tamanna Sharma, Priya, Dr Manika Sehgal, Tamanna Dhiman, Ambika, Ira, Neha, Shivani Sood, Manali*** *and others whom I have not mentioned.*

Last but not least, hats off to my friends outside JUIT, ***Vipan Thakur (Moni), Vikas Mahant, Bhupinder Singh Sandhu (Sandhu), Nishu, Abhishek (Pandee), Kapil****, for their support.*

I also extend my sincere thanks to field assistants, ***Mr. Mohinder (JUIT)*** *and* ***Mr. Dalip Parasher (HFRI)*** *for their aid during field visits and samples collection.*

Needless to say, all errors and omissions are mine.

ARCHIT SOOD

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ACCase	Acetyl-CoA Carboxylase
ACP	Acyl carrier protein
AP2	Apetala 2
ATP	Adenosine tri phosphate
bHLH	Basic helix-loop-helix
BLAST	Basic local alignment search tool
bZIP	Basic leucine zipper
C	Carboxyl terminus
CBF	Core-binding factor
cDNA	Complementary Deoxy ribonucleic acid
CLK	Choline kinase
CML	Calmodulin like
CMV	Cucumber mosaic virus
CNL	Coiled-coil NBS-LRR
CoA	C*oenzyme A*
Cq	Quantitation cycle
Ct	Cycle threshold
DAG	Diacylglycerol
DAP	Days after pollination
DGAT	Diacylglycerol acyl transferase
DGK1	Diacylglycerol kinase 1
dNTP	Deoxynucleotide triphosphates
Dof	DNA-binding with one finger
ECH	Enoyl-CoA hydratase
Emb	Embryo
End	Endosperm
ER	Enoyl Reductase

ERF	Ethylene responsive factor
EST	Expressed sequence tag
FA	Fatty acid
FAME	Fatty acid methyl ester
FATA	Linoleoyl thioesterase
Fig	Figure
FPKM	Fragments per kilobase of transcripts per million mapped reads
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GC	Guanine cytosine
GO	Gene ontology
GPAT	Glycerol-3-phosphate
HAB1	Homology to ABI1
HD-Zip	Homeodomain-leucine zipper
HMM	Hidden markov model
JA	Jasmonic acid
*Jc*MD	*Jatropha curcas* mosaic disease
*Jc*TF	*Jatropha curcas* transcription factor
*Jc*NL	*Jatropha curcas* NBS-LRR
KASI	β-ketoacyl-ACP synthase I
KASII	β-ketoacyl-ACP synthase II
KASIII	β-ketoacyl-ACP synthase III
KCR2	3-ketoacyl-CoA reductase isoform 2
KCS	Ketoacyl-CoA synthase
KEGG	Kyoto encyclopedia of genes and genomes
LACS8	Long-chain acyl-CoA synthetase 8
LD	Linoleoyl desaturase
LEC1	Leafy cotyledon 1
LPAT	Lyso-phosphatidic acid acyltransferase
M	Mature

MAST	Motif alignment & search tool
Mb	Mega base pair
mRNA	Messenger ribonucleic acid
MT	Malonyl transferase
MYB	Myeloblastosis
N	Amino terminus
NB-ARC	Nucleotide-binding adaptor shared by APAF-1, R proteins, and CED-4
NBS-LRR	Nucleotide binding site-leucine rich repeat
NCBI	National center for biotechnology information
NGS	Next-generation sequencing
NH	Nalagarh high oil content genotype
NL	Nalagarh low oil content genotype
OAD	Oleoyl-ACP desaturase
OCD	Oleoyl-CoA desaturase
PAD	Palmitoyl-ACP desaturase
PAP	Phosphatidic acid phosphatase
PCA	Principal component analysis
PCR	Polymerase chain reaction
PDAT	Phospholipid diacylglycerol acyltransferase
PERL	Practical extraction and report language
pme_TPM	Posterior mean estimate transcripts per million
PT	Palmitoyl thioesterase
R	Ripened
RGA	Resistance gene analogue
R genes	Resistance genes
RNA	Ribonucleic acid
RPKM	Reads per kilobase of transcripts per million mapped reads
RPM1	Resistance to *Pseudomonas syringae* PV Maculicola 1
RPS2	Resistance to *Pseudomonas syringae* 2

RPW8	Resistance to powdery mildew 8
rRNA	Ribosomal ribonucleic acid
RSEM	RNA-Seq by expectation maximization
RT-qPCR	Reverse transcription-quantitative polymerase chain reaction
SA	Salicylic acid
SAD	Stearoyl-ACP desaturase
SBP	Squamosa binding promoter
SD	Sterol desaturase
SH	Sunni high oil content genotype
SL	Sunni low oil content genotype
SNP	Single nucleotide polymorphism
SORLIP	Sequences over represented in light induced promoters
SRA	Sequence read archive
ST	Stearoyl thioesterase
TAE	Tris acetate-EDTA
TAG	Triacylglycerol
Taq	*Thermus aquaticus*
TF	Transcription factor
TIR	Toll/interleukin-1 receptor
TIR2	Toll/interleukin-1 receptor 2
TNL	Toll/interleukin-1 receptor NBS-LRR
TPM	Transcripts per million
TSS	Translational start site
U	Unripened
VIGS	Virus-induced gene silencing
WEPA	World environmental protection agency

ABSTRACT

The rising demand for biofuels has raised concerns about selecting alternate and promising renewable energy crops which do not compete with food supply. Jatropha (*Jatropha curcas* L.), a non-edible energy crop of the family euphorbiaceae, has the potential of providing biodiesel feedstock due to the presence of high proportion of unsaturated fatty acids (75%) in seed oil which is mainly accumulated in endosperm and embryo. Virus causing mosaic disease is becoming prevalent in Jatropha plantations and is responsible for significant reduction in seed yield and quality and also affecting its oil quality and content. The molecular basis of seed oil biosynthesis machinery has been studied in *J. curcas*, however, what genetic differences contribute to differential oil biosynthesis and accumulation in genotypes varying for oil content is poorly understood. Also insights into the molecular mechanism in response to virus infection in *J. curcas* are lacking and no report exists as of today on molecular components associated with disease resistance in this potential bioenergy plant. The current study, therefore, investigated: (1) relative expression of FA and TAG biosynthesis pathway genes in high (42%) versus low (30%) oil content genotypes of *Jatropha curcas*; (2) deciphering molecular components of a viral disease response in *Jatropha curcas*; (3) detection of NBS-LRR genes and defense-related transcription factors in *Jatropha curcas*.

The expression profile of 18 genes encoding enzymes catalyzing FA and TAG biosynthetic pathway in different developmental stages of embryo and endosperm from high (42%) and low (30%) oil content genotypes grown at two geographical locations was investigated. Most of the genes showed higher expression in ripened and mature oil accumulating stages of high oil content genotype, implying genetic differences contributing towards variation in oil content among genotypes. Genes encoding rate limiting enzymes showing higher expression in oil accumulating stages at low altitude were identified, thereby implying that oil content increases with decrease in altitude. To understand transcriptional regulation of oil accumulation in Jatropha, promoter regions of key genes implicated in oil biosynthesis and accumulation were analyzed for regulatory elements specific to oil accumulation such as Dof, CBF (LEC1), SORLIP, GATA, Skn-1_motif etc. and also transcription factors regulating oil accumulation i.e *bZIP*, *Dof*, *MYB*, *bHLH*, *CBF*, *AP2* were identified. Furthermore, to better understand, the molecular mechanisms associated with virus infection response, RNA-seq based comprehensive transcriptome sequencing of symptomatic virus infected (JV) and healthy (JH) leaf tissues of *J. curcas* using NextSeq 500 platform of Illumina was performed. In order to identify genes linked to pathways

upregulated and downregulated during mosaic virus infection, differential expression analysis and functional annotation was performed, indicating that various metabolism-associated processes along with oxidative phosphorylation, endocytosis, terpenoid biosynthesis, hormone signal transduction were activated whereas photosynthesis, anthocyanin biosynthesis, plant-pathogen interaction and calcium signaling were repressed in response to virus infection. Also to get insights into molecular components associated with disease resistance in Jatropha, transcriptome mining approach was followed which identified 47 NBS-LRR genes, in addition to previously identified 92 genes and 122 defense response-related transcription factors.

The current study provides repertoire of key genes from oil biosynthesis and transcriptional regulators specific to oil deposition which will be useful not only in dissecting the molecular basis of high oil content but also improving seed oil content through transgenic or molecular breeding approaches. Further this study provides information on molecular components which have been affected in response to virus infection and their precise role can be further validated. The present study also provides a repertoire of NBS-LRR genes and transcription factors which can be used in dissecting the molecular basis of disease resistance phenotype and developing disease resistant genotypes in Jatropha through genetic interventions.

CHAPTER 1

INTRODUCTION

The overall increase in energy consumption and increasing concerns about the pollutants and carbon dioxide has necessitated alternative sources of fuel other than fossil fuels. Plants and algae are being considered as most promising sources of biofuels. Plant triacylglycerols (seed storage oils) are excellent precursors for biodiesel production due to their similarity to fossil oils upto a higher extent [1]. Biodiesel, majorly produced from plants, is an alternate to fossil fuel as it is non-toxic, biodegradable and emits lower amount of carbon monoxides and hydrocarbons than petro-diesel. Chemically, biodiesel is fatty acid methyl esters (FAMEs) and is produced by transesterification of plant triacylglycerols with methanol and in the presence of alkali [2]. Fuel properties of biodiesel depend upon the composition of fatty acids blend in the oil. Palmitic acid (16:0), stearic acid (18:0), oleic acid (18:1), linoleic acid (18:2) and linolenic acid (18:3) are the five common fatty acids present in most of the plant oils. Most of the crop plants being used for biodiesel production are edible in nature and are causing scarcity for the overall food supply. For sustainable biodiesel production, the whole focus should shift from edible to non-edible crops so that there is no competition with food security and effective management of agricultural wastelands could also be achieved [3]. Moreover, the influence of oil crops for biodiesel production on the prices of food commodities is an apprehension. There is a hurdle in the effective management of these energy crops as they have been plagued with diseases that thwart production and cause a global yearly average yield loss of upto 16 percent. Various biotic stresses are associated to these energy crops and are reducing their yield potential.

Of many energy crops, Jatropha (*Jatropha curcas* L.) is gaining an immense potential for biodiesel production due to merits like it is perennial, drought-resistant and have high oil content. Jatropha (*Jatropha curcas* L.), (2n=2x=22) (Figure 1.1) is a non-edible energy crop of the family euphorbiaceae. It is an economically important source of oil used for the production of various industrial products like lubricants, cosmetics, medicines, including high quality biodiesel due to the presence of large proportion of unsaturated fatty acids and high oil content (up to 50%) (Table 1.1). *J. curcas* is a perennial shrub or a small tree of height upto 6 meters with a life expectancy of up to 40-50 years. Depending on the quality of soil and rainfall, oil can be extracted from the its seeds after 2–5 years. It is well adapted to various soil and climate types [4]. Genus *Jatropha* is very diverse in terms of morphological features and comprises of more than 200 species dispersed mainly in the dry tropical areas of America. Its center of origin is supposedly

defined as central parts of America [5]. Later it has been introduced into many parts of Asia and Africa and now is being cultivated globally [6]. It has been now introduced into many parts of our country. *J. curcas* is listed as a fuel and fuel additive with the world environmental protection agency (WEPA) [7]. Before its consideration as a suitable and potential bio-energy crop, *J. curcas* was mainly used for medicinal products. In some parts, it was also used as fence around arable land as animals do not eat this plant owing to the presence of phorbol esters in its seeds which make it toxic in nature. However, there are some non-toxic genotypes also. Amongst various oil seed plants, *J. curcas* is anticipated as potential source of biodiesel production due to features like drought hardiness, easy propagation, wide adaptability, rapid growth and small gestation period. *J. curcas* is still considered as an undomesticated plant [8]. For many *J. curcas* germplasms, low productivity is inherent and raising large-scale plantations using such low yielding planting material can lead to wasteful ventures. Though numerous efforts have been made to develop *J. curcas* as an industrial crop, the scant information on its agronomic practices and lack of improved genotypes and cultivars are the major bottlenecks in its full exploitation as a potential bioenergy crop [9].

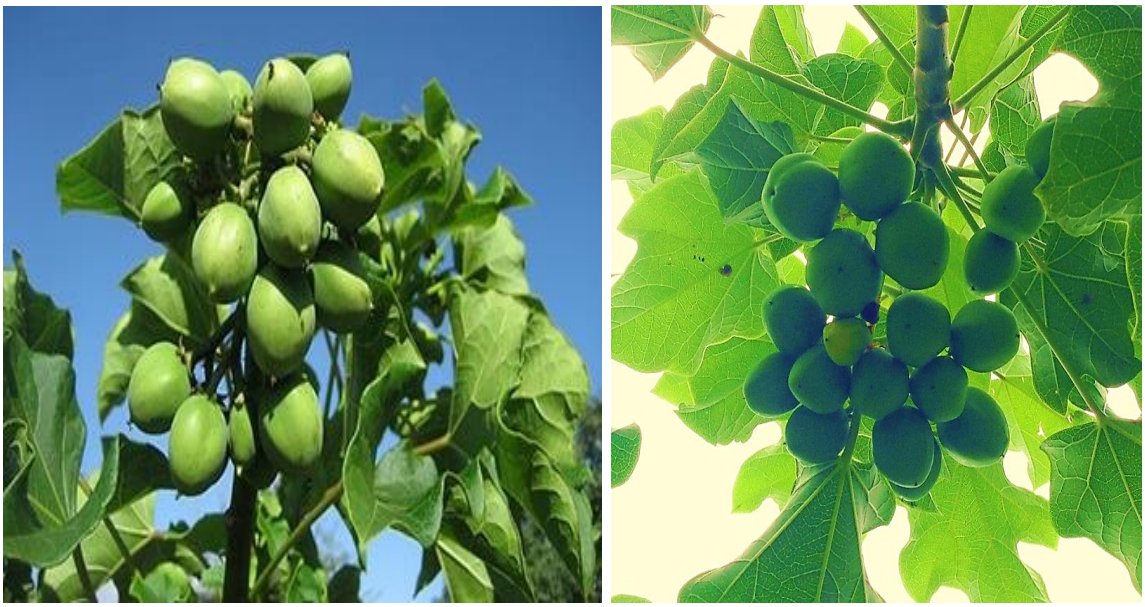


Figure 1.1 *Jatropha curcas* plants with fruits

Taxonomic classification of *Jatropha curcas* L.

Kingdom	Plantae
Subkingdom	Tracheobionta
Division	Magnoliophyta
Sub division	Spermatophytina
Class	Magnoliopsida
Subclass	Rosidae
Order	Euphorbiales
Family	Euphorbiaceae
Genus	*Jatropha* L.
Species	*Jatropha curcas* L.

J. curcas is an economically important plant as its oil is used for the production of good quality biodiesel. However some major constraints like low productivity, unreliable flowering and fruiting, non-availability of sufficient feedstock and susceptibility to biotic stresses are limitations in implementing this plant as a source of biodiesel.

Mature seeds of Jatropha are enriched in oil that accumulates mainly in the endosperm and embryo [10, 11] (Figure 1.2). Oil biosynthesis in Jatropha is composed of two main pathways, fatty acid (FA) and triacylglycerol (TAG) biosynthesis which occur in plastid and endoplasmic reticulum, respectively (Figure 1.3). A total of 18 enzymes catalyzing the whole oil biosynthetic pathway are present out of which 14 enzymes catalyzes the formation of specific fatty acids and 4 enzymes are involved in catalyzing the formation of triacylglycerols. After their synthesis, traicylglycerols move to small spherical structures known as oil bodies in endoplasmic reticulum. These oil bodies form the total mass of oil accumulated in seeds. The oil accumulation in developing seeds initiates at the early stage and reaches maximum at mid-later or later developmental stages. Generally, triacylglycerols are accumulated specifically in endosperm and embryo in oil seeds [10]. In Jatropha, oil accumulation also takes place in embryo along with endosperm but relatively in low amounts [12]. Cotyledons, from the developing embryos are the main storage tissues and the major location of triacylglycerol accumulation on an entire embryo level.

In oil plants, the seed oil content along with fatty acid composition varies in developmental stages as per various environmental cues like altitude, light, temperature and drought. Overall increase in total oil content and fatty acid composition has been reported with decrease in altitude level for plants. This is attributed to the fact that low partial pressure of carbon-dioxide (CO_2) at higher altitudes is responsible for reduced rate of photosynthesis and subsequently there is overall decrease in oil content at higher altitudes.

Jatropha has been extensively studied in terms of biotechnological interventions for understanding oil biosynthesis mechanisms in its seeds. As Jatropha has been domesticated recently, a number of interventions have been made to improve its yield traits like seed yield and oil content, either by conventional breeding approaches or gene level technologies.

Figure 1.2 Mature seeds of *J. curcas* (Harvesting stage)

Table 1.1 Fatty acid composition and oil content of major oil plants

Fatty acids (%)	Jatropha	Castor bean	Sunflower	Soybean
Palmitic acid	10	3	10	10
Stearic acid	10	2	5	5
Oleic acid	45	10	30	35
Linoleic acid	35	10	50	45
Linolenic acid	1	-	5	5
Ricinoleic acid	-	75	-	-
Total oil content (%)	25-50	40-45	25-35	20-25

Various approaches of modern day biotechnology like genomics, transcriptomics and bioinformatics are expected to fill gaps like lack of information on genetic factors contributing towards higher oil accumulation and transcriptional regulation of oil biosynthesis and accumulation in Jatropha. Overall elucidation of oil biosynthesis mechanism have been understood by Sato et al. [13] while analyzing Jatropha genome for the first time (Table 1.2) where genes related to oil biosynthesis were identified. There is vast information available on molecular components related to oil biosynthesis and accumulation in *J. curcas*. Previous studies have identified genes, ESTs, miRNAs associated with oil biosynthesis in *J. curcas* [11, 14, 15].

In spite of availability of molecular components associated with oil biosynthesis and accumulation in *J. curcas*, genetic factors contributing to the variations in seed oil content (25-45%) among different genotypes are not known. Most of the studies on understanding the molecular regulation of fatty acid and lipid biosynthesis were limited to correlating the expression profile of FA and TAG biosynthetic pathway genes with the developmental stages of seeds. The gene expression only reflected expression status of FA and TAG biosynthesis pathway genes at particular developmental stage vis-à-vis oil content rather pinpointing what genetic differences contribute to differential biosynthesis and accumulation of oil content in seeds of different genotypes of *J. curcas*. Also less information is available for molecular level understanding of oil accumulation in embryo in *J. curcas*. Despite a specific understanding of the overall biosynthesis of oil including genes coding for the enzymes involved, our knowledge regarding transcriptional

regulation of oil biosynthesis and accumulation remains partial in Jatropha. The information pertaining to identification of transcriptional regulators such as transcription factors and promoter regions associated to oil biosynthesis genes is scarce till date. The identification of key genes associated with oil biosynthesis and accumulation are expected to play a major role in engineering Jatropha with enhanced oil production. Further there are no reports as of today on molecular data related to effect of major environmental factors such as altitude on oil accumulation in *J. curcas*.

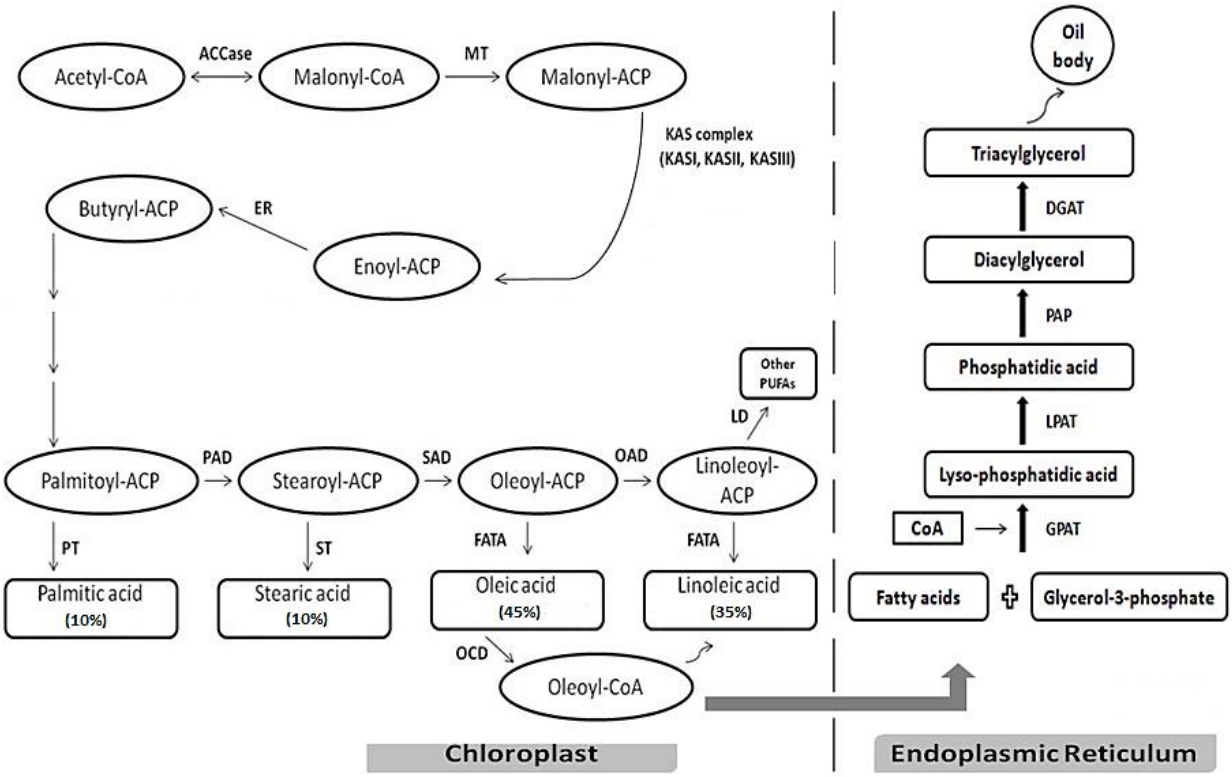


Figure 1.3 Schematic representation of fatty acid (FA) and tri acyl glycerol (TAG) biosynthesis pathway in *J. curcas* [16]. *ACCase*- Acetyl-CoA Carboxylase; *DGAT*- Diacylglycerol acyl transferase; *ER*- Enoyl Reductase; *FATA*- Linoleoyl Thioesterase; *GPAT*- Glycerol-3-phosphate acyl transferase; *KASI*- β-ketoacyl-ACP synthase I; *KASII*- β-ketoacyl-ACP synthase II; *KASIII*- β-ketoacyl-ACP synthase III; *LD*- Linoleoyl Desaturase; *LPAT*- Lyso-phosphatidic acid acyltransferase; *MT*- Malonyl Transferase; *OAD*- Oleoyl-ACP Desaturase; *OCD*- Oleoyl-CoA Desaturase; *PAP*- Phosphatidic acid Phosphatase; *PT*- Palmitoyl Thioesterase; *SAD*- Stearoyl-ACP Desaturase; *ST*- Stearoyl Thioesterase; *PAD*- Palmitoyl-ACP Desaturase

Table 1.2 Statistics of *J. curcas* genome

Feature	**Description**
Size (Mb)	370
Chromosome number	11
GC content (%)	33
Gene number	45,000 (app.)
Repetitive DNA (%)	52

The large-scale cultivation of selected genotypes of *J. curcas* have made it vulnerable to biotic stresses including diseases and pests, thereby affecting their oil yield potential. Fungal strains of *Alternaria alternate*, *Neoscytalidium dimidiatum*, *Botryosphaeria dothidea* and *Colletotrichum gloeosporioides* were reported to be responsible for infectious spots, root rot, black rot and anthracnose disease, respectively causing reduction in overall yield of Jatropha [17, 18]. Although, fungal and bacterial strains are responsible for many associated diseases in Jatropha, *Jatropha curcas* mosaic disease (*Jc*MD) caused by mosaic virus has been found to be prevalent in the plantations and is continuously reducing fruit yield and quality of plants in the field [19]. It is characterized by leaf curling, reduction of fruit size and distortion. Mosaic disease of Jatropha is a severe constraint for its full yield and restrictive factor in Jatropha cultivation around the globe. Several reports exist describing the outbreak of viruses in Jatropha in other parts of the world. Ramkat et al. [20] reported the incidence of Jatropha infection with cassava mosaic viruses in Europe. Begomovirus associated with mosaic disease of *J. curcas* has also been identified in Nigeria [21].

Apart from reduction in overall yield and seed oil content, various physiological processes get affected in plants in response to virus infection. These includes increase in respiration rate, decrease in photosynthesis rate, decrease in transpiration rate, decrease in relative water content, decrease in pigment content etc. [22, 23, 24, 25]. Changes in these general processes are attributed to the fact that during infection viruses use host machinery for multiplication.

In recent years, next-generation sequencing (NGS) has provided novel and advanced ways to fasten the identification of genes in many plant species, mainly those under biotic and abiotic stresses [26, 27]. RNA-Seq is a whole transcriptome sequencing

technique which sequences the overlapping short fragments of mRNA or cDNA to describe quantitatively the entire transcriptome. Analysis of transcriptome in any biological system can deliver major understanding into entities such as genes and transcription factors involved in biological processes. The high throughput transcriptome analysis technique has been studied to understand molecular aspects regarding various diseases including virus infection response in many plants.

In Jatropha, all the biotic stresses including mosaic disease are constantly increasing the risks associated with the yield potential and seed oil content as directly affecting the fruit and flower development. Although few reports pertaining to identification and characterization of mosaic virus responsible for causing mosaic disease in *J. curcas* exist [19, 28, 29], the molecular insights regarding mosaic virus infection are poorly understood till date. There is an immense need to understand the molecular mechanisms underlying virus (disease) response in *J. curcas* as it will facilitates the identification of genes and corresponding biological processes being affected in response to mosaic virus infection. Further it could lead to genetic interventions for broadening disease control methodologies, specifically virus control to maximize its potential to be an ideal bioenergy crop. Also new dimensions on understanding the molecular perspective of plant-pathogen interaction in Jatropha will be supplemented.

Plants have developed a range of mechanisms to identify and react to various possible pathogens [30]. Plants show resistance to many pathogens and pests due to the presence of resistance (R) genes. These R genes encode proteins which protect plants from various pathogenic organisms. Majority of plant R genes belong to nucleotide binding site-leucine rich repeat (NBS-LRR) class which provide resistance to a large number of pathogens. These NBS-LRR genes hold potential in the development of disease resistant transgenics. Further, the regulation of immunity and response to other stresses of plants in their natural habitat is enforced by a network of regulatory proteins or transcription factors [31]. Transcription factors normally bind to the promoters of resistance genes and thus regulate their expression. Identification of transcription factors related to defense response or disease resistance is also of great significance in predicting the pathogen responsive promoter elements.

As management of *Jatropha curcas* mosaic disease caused by mosaic virus is not economically viable through pesticides, the selection and development of disease resistant genotypes would be an alternative and sustainable strategy. Identification and

characterization of disease resistance genes, including NBS-LRRs and defense response related transcription factors is anticipated to accelerate the process of genetic improvement programmes and breeding for development of disease resistant varieties through transgenic or molecular breeding approaches [32]. As there is lack of information on molecular components associated with disease resistance in *J. curcas*, whole genome and transcriptome wide investigation of NBS-LRR resistance genes and defense related transcription factors can therefore, provide novel insights about the overall understanding of resistance architecture. Molecular components associated to defense response can be used in dissecting the molecular basis of disease resistance.

In order to understand the molecular mechanisms associated to maximize the yield and to gather information on genetic factors contributing to differences in oil content among genotypes in Jatropha, the expression profile of 18 genes encoding enzymes catalyzing the FA and TAG biosynthetic pathway in various developmental stages (fruits with days after pollination) of endosperms and embryos of high (42%) versus low (30%) oil content genotypes of *J. curcas* grown at two different geographical locations, was studied. To unravel transcriptional regulation associated with oil biosynthesis and accumulation in *J. curcas*, the promoter regions of key genes showing elevated expression in this study along with other genes encoding enzymes associated with overall lipid biosynthesis, reported in the previous studies by Xu et al. [33] and Gu et al. [11] were also analyzed to identify regulatory elements specific to oil or lipid accumulation in plants. The transcription factor families regulating oil accumulation in developmental stages of Jatropha endosperm were also identified, which provided first glimpse of regulatory control of oil accumulation in *J. curcas*. Furthermore, to provide insights into the molecular mechanism associated to disease response in Jatropha, comparative transcriptomic analysis of healthy and mosaic virus infected leaves was performed. This analysis provided the repertoire of genes associated with biological processes being upregulated and downregulated in response to virus infection. Additionally, the available transcriptome of Jatropha was analyzed to identify NBS-LRR genes, their genome location, and characterization into toll/interleukin A 1 receptor NBS-LRRs (TNLs) or coiled-coil NBS-LRRs (CNLs). Transcriptome mining approach was also followed to identify transcription factors related to defense response.

Keeping in view, the lack of information on genetic factors contributing to differences in oil content among oil contrasting genotypes and our partial knowledge towards

understanding of molecular mechanisms associated with disease response and disease resistance, the present study was carried out with following objectives:

1) Relative expression of FA and TAG biosynthesis pathway genes in high versus low oil content genotypes of *Jatropha curcas*
2) Deciphering molecular components of a viral disease response in *Jatropha curcas*
3) Detection of NBS-LRR genes and defense related transcription factors in *Jatropha curcas*

CHAPTER 2

REVIEW OF LITERATURE

The literature pertaining to the present study has been reviewed as under:

2.1 Origin and taxonomy of *J. curcas*

Jatropha curcas L. (Physic nut) originated in central parts of America and is now confined to tropical regions throughout the globe. Later it was introduced into Africa and Asia, and is grown as a hedge plant there [34, 35]. It is a member of family euphorbiaceae. Physic nut was named as *Jatropha curcas* L. firstly by Linnaeus [36]. The genus name *Jatropha* originated from the Greek word jatr´os means 'doctor' and troph´e which means 'food', suggests its uses in medicine. The genus *Jatropha* is related to tribe joannesieae of crotonoideae in the family euphorbiaceae with approximately 200 known species, as per Dehgan and Webster [37].

2.2 Morphological features of *J. curcas*

J. curcas is a perennial shrub or small tree of height upto 5-6 meters with a life expectancy of up to 40-50 years. *J. curcas* is characterized by a smooth bark, strong branches, and dense leaves. Its leaves are 8-18 cm wide, glossy and glabrous, supplemented with exiguous and pilose stipules. The petiole is 12-15 cm long. Flowers of *J. curcas* are unisexual, monoecious, greenish yellow in colour with peduncled paniculate cymes. The flowers are smaller in size with more number of male flowers as compared to female flowers and male and female flowers on the same inflorescence [38]. The flowering occurs at branch terminal [39]. The fruit is a capsule, with length and width of 4-5 cm and 2-3 cm, respectively. Initially, the fruit is green and then turns to yellow in the ripened stage and dark brown at the mature stage. A single fruit contains 3 seeds. The seeds are elliptical in shape and blackish in colour (Figures 1.2 and 2.1). Seeds are 1.5-2 cm long and 1-1.2 cm wide [39].

2.3 Seed oil biosynthesis

Oil biosynthesis in Jatropha is composed of two main pathways, fatty acid (FA) and triacyl glycerol (TAG) biosynthesis which occur in plastid and endoplasmic reticulum, respectively [41]. After the synthesis of fatty acids in plastid, those are activated (acyl-CoAs) and transported to the endoplasmic reticulum to synthesize the triacylglycerol or oil. Several reports are available now for the identification and characterization of genes encoding

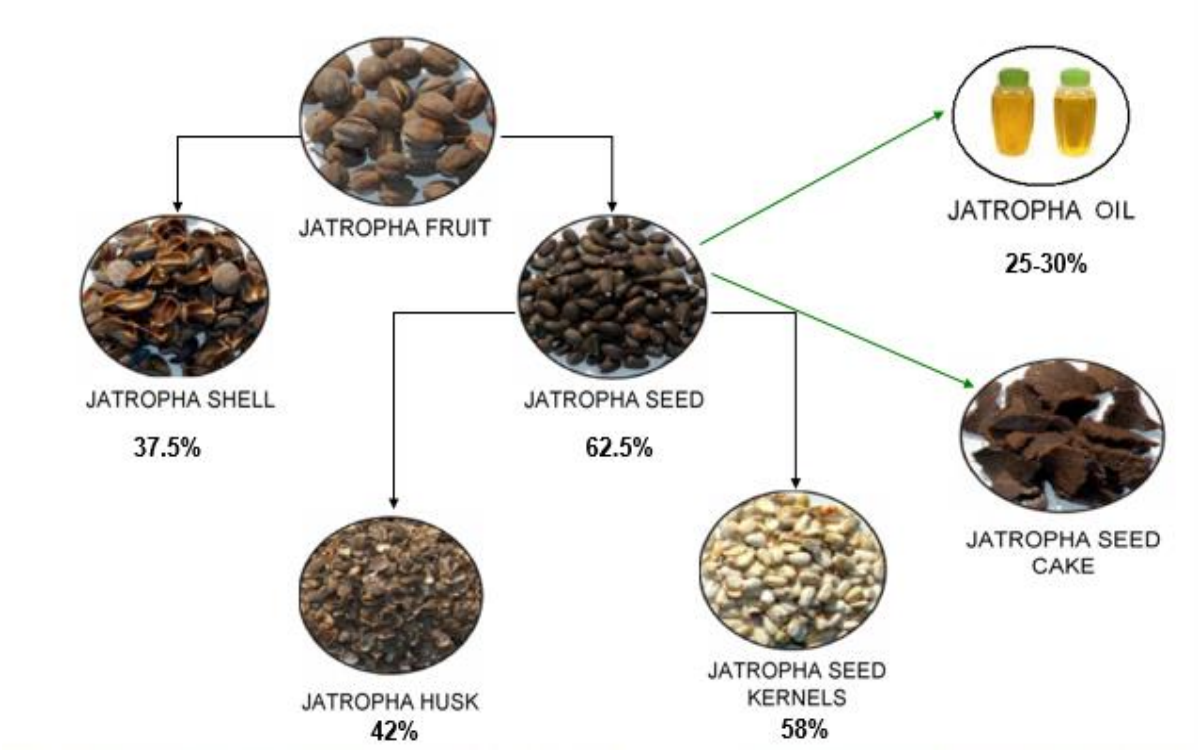


Figure 2.1 Percentage wise distribution of major parts of *J. curcas* fruits for industrial purposes [40]

enzymes regulating various steps in FA and TAG biosynthesis in Jatropha [13, 33, 41, 42]. The oil composition and content can be altered by altering the expression of genes regulating FA and TAG biosynthetic pathway, respectively. Biosynthesis of fatty acids is initiated by the abridgment of acetyl-CoA to malonyl-CoA and is catalyzed by heteromeric acetyl CoA carboxylase (*ACCase*).This step is considered as rate limiting. Acetyl-CoA is synthesized mainly from the processes like glycolysis. Malonyl-ACP (substrate of fatty acid synthase complex) is formed from malonyl-CoA by transferring of malonyl group to acyl carrier protein (*ACP*). The conversion of malonyl-CoA to malonyl-ACP is catalyzed by malonyl transferase. A repeated chain of condensation, reduction, and dehydration reactions directs the *de novo* fatty acid synthesis in plastid. A set of three substrate specific enzymes i.e. β-ketoacyl-ACP synthase I (*KASI*), β-ketoacyl-ACP synthase II (*KASII*), β-ketoacyl-ACP synthase III (*KASIII*) aid in the formation of 18 carbon-fatty acid with malonyl-ACP and aceto acteyl-ACP by a chain of condensation reactions. Additional condensation reactions are required to obtain saturated fatty acids with enzymes such as enoyl reductase (*ER*). Then series of fatty acid desaturation reactions occur in which desaturase (*FAD*) plays a major role [11]. Stearoyl-ACP desaturase converts stearoyl-ACP to Oleoyl-ACP, which is considered as rate-limiting step in fatty acid synthesis and is

responsible for conversion of saturated fatty acids to unsaturated fatty acids. Fatty acid elongation and synthesis is terminated by inter plastidial enzyme thioesterases (*FAT*), which are of two types in plants. One class i.e *FATA* eliminates oleate from ACP, whereas other one, *FATB* thioesterases are implicated in saturated and unsaturated acyl ACPs. It also exports acyl moieties to the endoplasmic reticulum for the production of glycerolipids [43] and involved mainly in saturated fatty acids biosynthesis, essential for plant development and growth [44]. Free fatty acids are then released from ACP, exported from the plastid and converted to acyl-CoAs. In endoplasmic reticulum, assembly of tri acyl glycerol sequentially consumes the acyl-CoA using substrate glycerol-3-phosphate. Lyso-phosphatidic acid and Phosphatidic acid catalyzed by enzymes glycerol-3-phosphate acyltransferase (*GPAT*) and lyso-phosphatidic acid acyltransferase (*LPAT*), respectively are produced. Following phosphate removal, phosphatidic acid is converted then to diacylglycerol, which is the precursor of tri acyl glycerols, catalyzed by phosphatidic acid phosphatase (*PAP*). Diacylglycerol acyltransferase (*DGAT*) with acyl-CoA as an acyl donor, converts diacylglycerol (DAG) to triacylglycerols (TAG). Then these triacylglycerols (TAG) are assembled in proteins like oleosins to form oil bodies in seeds. Generally, oleosins and caleosins are considered as two predominant seed oil storage proteins in plants. These reactions taking place in endoplasmic reticulum are termed collectively as Kennedy pathway and each step catalyzed by different enzymes is considered as rate-limiting step as nascent fatty acids are being incorporated into triacylglycerols (TAGs) in seeds [16, 45].

2.3.1 Biosynthesis and accumulation of oil in *J. curcas* seeds

Generally, triacylglycerols are synthesized and accumulated specifically in endosperm and embryo in oil seeds [11, 46, 47, 48]. Various studies have shown that the oil accumulation in developing seeds starts at the early stage and reaches maximum at mid-later or later developmental stages which correspond with dehydration of seeds [49]. In Jatropha, the storage lipids are mainly synthesized and accumulated in endosperm of seeds. The oil accumulation in endosperm increases with the developmental stages of seeds after fertilization as is evident from the formation of oil bodies [11]. For oil plants, many studies have reported in different systems where oil and lipid accumulation is developmentally regulated at various developmental stages of embryos along with endosperms [16, 46, 50, 51, 52]. In Jatropha, oil accumulation also takes place in embryo but relatively in low

amounts as compared to endosperm and seed coat [10, 12, 14]. Different embryo developmental stages with respect to days after flowering showed gradual increase in the lipid content in Jatropha [14]. Recently, another study by Kim et al. [52] showed that accumulation of oil is developmentally regulated at early, mid and desiccation stages of embryo development in Jatropha.

2.3.2 Molecular basis of oil biosynthesis and accumulation in *J. curcas*

Genomics, transcriptomics, proteomics and bioinformatics are anticipated to unravel gaps in the molecular understanding of oil biosynthesis and accumulation in Jatropha seeds (Table 2.1). Genome of Jatropha has been analyzed by Sato et al. [13] which shed light on overall oil biosynthesis mechanism. Various transcriptomic analysis have been done in order to identify transcripts regulating oil biosynthesis and accumulation in seeds. Natarajan and Parani [53] reported 56 transcripts that have direct role in oil biosynthesis from developing seeds and embryo transcriptomes. Pyrosequencing approach was applied to *J. curcas* developing seed transcriptome to identify transcripts involved in triacylglycerol accumulation [9]. Moreover recently miRNAs regulating lipid metabolism have been identified in Jatropha [15]. In another study, expressed sequence tags (ESTs) linked with seed development and lipid metabolism were identified at different embryo developmental stages [14]. Recently, studies have been carried out to identify key genes of oil and lipid biosynthesis up regulating among various developmental stages of seed and endosperm in *J. curcas*. Xu et al. [33] reported expression profiles of 21 lipid biosynthesis genes and observed that many genes were up regulated during developmental and lipid accumulating stages in seeds. Another study by Gu et al. [11] also checked the expression status of 68 genes from fatty acid and lipid biosynthesis and found that majority of the genes were up regulated in consistent with the lipid accumulation in endosperm. Both these studies on understanding the molecular basis of fatty acid and lipid biosynthesis were limited to correlating gene expression with seed developmental stages in one genotype only, no oil contrasting genotypes were used thereby limiting the identification of key genes which would be having higher transcript abundance in a high oil content genotype. Also endosperms and whole seed part were considered in these studies, restricting the identification of molecular switches regulation oil formation in embryo of *J. curcas*, as embryo along with endosperm contribute to overall oil content in *J. curcas* seeds [14, 52]. A few studies are there describing about differences in key genes contributing to oil biosynthesis in genotypes varying for oil content in other plant species.

Differentially expressed genes related to photosynthetic activity among high and low oil content genotypes of *B. napus* were directly responsible for oil content variation [54]. Furthermore in sesame, resequencing revealed high genetic diversity of lipid related genes in 29 different accessions from 12 countries which suggested possible association with the variation in oil content among accessions [55].

Table 2.1 Recent molecular interventions for enhancing seed yield and oil content in *J. curcas*

Yield trait	**Approach**	**Reference**
Seed size	Auxin signal transduction genes *JcARF19* & *JcIAA9* mapped to seed size QTLs	Ye et al. [56]
Fatty acid Biosynthesis	Transcriptome analysis for oil biosynthesis genes	Grover et al. [57]
Oleic acid content (>78%)	Seed-specific *JcFAD2-1* RNAi	Qu et al. [58]
~30% higher seed storage lipid	Seed-specific lipase *JcSDP-1* RNAi	Kim et al. [52]
Oil metabolism	miRNA regulators	Galli et al. [15]

2.3.3 Transcriptional regulation of oil biosynthesis and accumulation

Transcription factors (TFs) facilitate various cellular responses by recognizing specific cis-regulatory DNA sequences in the promoter regions of target genes. Identification of transcription factors is valuable for reviewing the transcriptional regulatory switches involved in development, reproduction and various responses to the changing environment in plants. Transcription factors may be regarded as molecular switches that connect signal transduction pathways to gene expression [59]. Multiple steps have been regulated by transcription factors concurrently. They provide a potential substitute to single-enzyme approaches for varying complex traits in plants [60].

Various studies have shown that seed oil content can be increased by altering the expression levels of individual genes encoding enzymes regulating oil metabolism pathway [52, 61, 62, 63]. The expression of transcription factors that direct the multiple enzymes regulation can also be manipulated to target individual enzymes [60, 64]. Transcription factors manipulation can regulate expression of genes in fatty acid biosynthesis and alter the fatty acid/oil levels [65].

For many oil plants, transcription factors governing fatty acid synthesis and overall lipid accumulation has been identified and characterized (Figure 2.2, Table 2.2). *LEC1* and *WRI1*, two transcription factors have been reported to enhance oil content in Arabidopsis and maize [66]. *WRI1*, *WRI3* and *WRI4*, members of APETALA2-ethylene-responsive element binding protein (*AP2-EREBP*) family also triggered fatty acid biosynthesis in seeds of Arabidopsis for triacylglycerol production [67]. Similarly, the oil production in *Brassica napus* was enhanced by conditional expression of *LEC1* and *LEC1-LIKE* transcription factors, which are the key regulators of fatty acid biosynthesis [65]. In Arabidopsis, *PHR1*, a member of *MYB* family regulated triacylglycerol accumulation [68]. Another transcription factor family, *Dof*, has been studied extensively to correlate its role in lipid accumulation and enhancement. Two soybean orthologs, *GmDof4* and *GmDof11* increased the fatty acid and lipid content in transgenic Arabidopsis seeds by upregulating the genes which are associated with fatty acid biosynthesis [69]. In a similar study, the overexpression of soybean transcription factor, *GmDof4* also enhanced the lipid content in microalgae *Chlorella ellipsoidea* [51]. Member of *bZIP* transcription factor family i.e. *bZIP123* from soybean was also responsible for enhanced lipid content in transgenic Arabidopsis seeds, when overexpressed [70]. Recently Hu et al. [71] reported identification of 11 different families of transcription factors (*NF-YC*, *bZIP*, *HB-other*, *HSF*, *C3H*, *E2F/DP*, *AP2*, *MYB_related*, *CPP*, *MYB* and *LFY*) associated with lipid synthesis in oleaginous microalgae *Nannochloropsis*.

Table 2.2 Transcription factor families involved oil biosynthesis and accumulation

TF Family	**Reference**
Dof	Wang et al. [69]
AP2	Ma et al. [72]

B3 domain	Palaniswamy et al. [73]
GATA	Fobert [74]
bHLH	Courchesne et al. [75]
MYB	Liu et al. [76]
bZIP	Song et al. [70]
HD-Zip	Chew et al. [77]
CBF	Fobert [74]

2.3.4 Effect of altitudinal variation on oil biosynthesis and accumulation

External environmental factors can also regulate concentration and amount of oil in plants [78]. At the time of developing stages, fatty acids and seed oil content in plants varies depending on various environmental conditions like altitude, radiations, temperature, drought and heavy metal content. During growth of seed development, rise in temperature may affect the composition of fatty acids which results in poor oil quality. Regardless of fact that identification of fatty acids composition is genetically determined, various environmental conditions also influence them. Less temperature and solar radiations negatively affected the linoleic acid content in sunflower [79]. Generally the oil content and fatty acid composition goes on increasing with decrease in altitude level due to the fact that low partial pressure of carbon-dioxide (CO_2) at higher altitudes reduces photosynthesis rate and ultimately reduction in oil content [80, 81]. There are various reports about the effect on total oil content with change in altitude along with other geographical parameters. For a medicinal plant, *Achillea wilhelmsii* the biosynthesis and accumulation of essential oils percentage was affected to a greater extent with varying altitudes [82]. For rapeseed, identification and expression analysis of the genes linked with oil synthesis has been reported in accessions grown at varying altitudes [83] which explained about the inverse relationship of oil accumulation

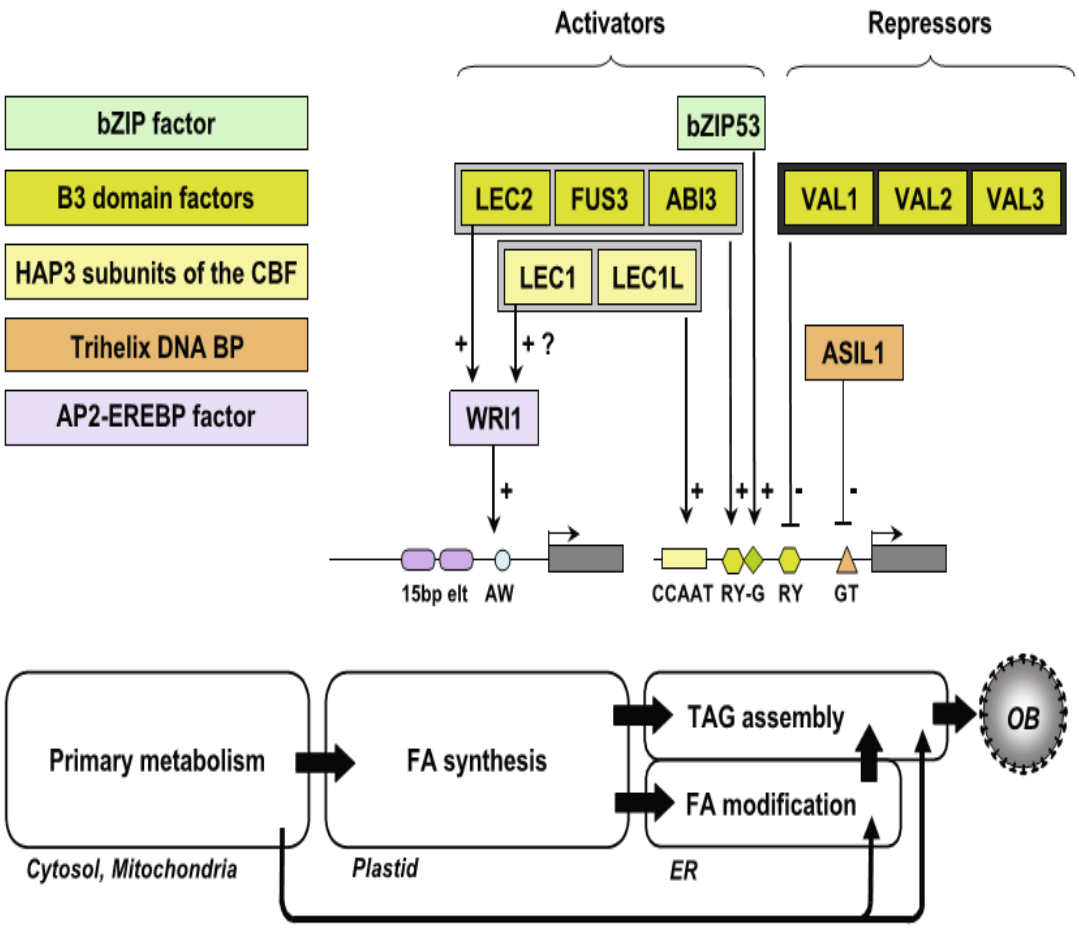


Figure 2.2 Model for transcriptional regulation of triacylglycerol formation in mature seeds of *Arabidopsis thaliana* (*ASIL1*-ARABIDOPSIS 6B-INTERACTING PROTEIN 1-LIKE1; *AP2*-EREBP-APETALA2-ethylene responsive element-binding protein; *bZIP*-basic leucine zipper; *CBF*-CAAT box-binding factor; ER-endoplasmic reticulum; FA-fatty acid; OB-oil body; *FUS3*- FUSCA3; *LEC1,2*-LEAFY COTYLEDON1,2; *L1L*-LEAFY COTYLEDON1-LIKE; TAG- triacylglycerol; Trihelix DNA BP-trihelix DNA binding protein; *VAL1, 2, 3*-VP1/ABSCISIC ACID INSENSITIVE3-LIKE1, 2, 3; *WRI1*-WRINKLED1) [47].

and altitude. Information regarding effect of environmental factors such as altitude variation on oil biosynthesis and accumulation is not available for *J. curcas* as of today.

2.4 Metabolic engineering for modification of fatty acid composition and oil content

Metabolic engineering approaches have been performed in many oil plants to alter the fatty acid composition and oil content in seeds by targeting genes encoding the key enzymes catalyzing oil biosynthesis (Figure 2.3, Table 2.3). *ACCase*, the rate limiting enzyme of fatty acid biosynthesis in chloroplast has been extensively studied to correlate its expression with enhanced fatty acid production. For plants like maize, tobacco and canola

the overexpression of gene encoding *ACCase* enzyme was associated with overall increase in fatty acids [65, 84, 85, 86]. By altering the expression level of genes encoding *KAS* complex, increase and decrease in the saturated and unsaturated fatty acid composition and increase in oil content was observed for Arabidopsis, *Brassica* spp. and oil palm [87, 88, 89]. Genes encoding desaturase enzymes are the imminent candidates for genetic engineering to modify polyunsaturated fatty acids in seed oil plants [11].

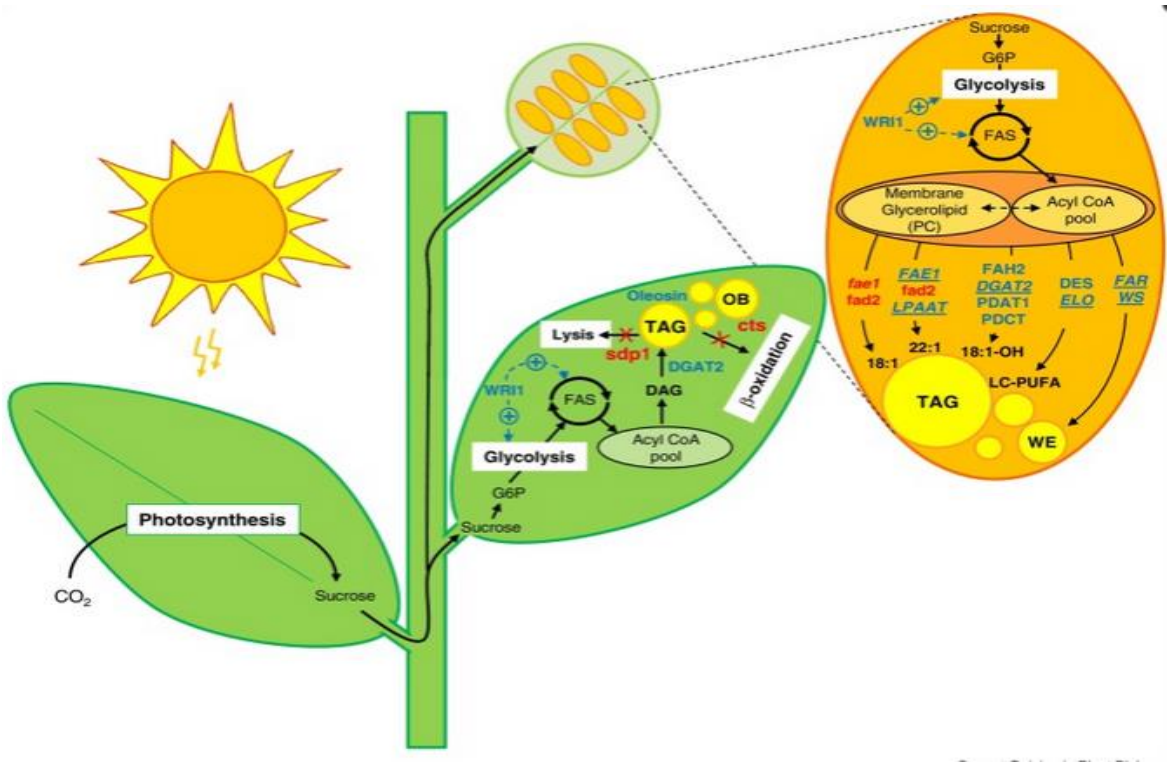


Figure 2.3 Diagrammatic representation of metabolic engineering strategies for manipulation of oil content and composition in leaves and seeds. (Blue: Target genes for overexpression; Red: Target genes for inactivation by mutation or RNAi. Genes encoding enzymes using acyl-CoA substrates are underlined [90]

Many studies correlated the expression of desaturases with increase in oleic acid composition and increase in seed oil content in many oil plants exists [65, 91, 92]. Acyltransferases were also targeted to increase the overall oil content in oil seed plants. In maize, a phenylalanine insertion in *DGAT* gene has resulted in increase of oleic acid and total oil content [93]. The overexpression of *DGAT2* from fungus *Umbelopsis ramanniana* in soybean lead to 1.5% increase in oil yield of seeds. Similarly Zhang et al. [94] demonstrated that overexpression of *DGAT* from oleaginous marine protist *Thraustochy triumaureum* led to overall increase in oleic acid content in Arabidopsis seeds.

Modification of oil quality has also been achieved by the over expression of *DGAT* gene in the seeds of *Euonymus alatus* and *Arabidopsis thaliana* [95]. In another study, co expression of *DGAT* along with *WRI1* resulted in significant increase in triacylglycerol level in *Nicotiana benthamiana* [96]. The upregulation of gene encoding *LPAT* from *Crambe abyssinica* has resulted in production of transgenic rapeseed plants with enhanced oil content [97]. For *A. thaliana* and *Ricinus communis* also, the overexpression of *LPAT* resulted in overall increase in triacylglycerol accumulation [98, 99]. Overexpression of another acyltransferase i.e. *GPAT* in *Brassica napus* and *A. thaliana* seeds raised oil accumulation and oil content [100, 101].

Table 2.3 Genetic modifications for increase in seed oil content

Modification	**Plant spp.**	**% increase in oil content**	**Reference**
Carthamus tinctorius GPAT over expression	*A. thaliana*	10-20%	Jain et al. [102]
Over expression of *LEC1* and *WRI1*	*Z. mays*	40%	Shen et al. [66]
Over expression of *Glycine max MYB73*	*A. thaliana*	5-15%	Liu et al. [76]
Over expression of multiple genes i.e. *BnGPDH*, *BnGPAT*, *BnDGAT*, *ScGPDH* and *ScLPAAT*	*B. napus*	10-15%	Liu et al. [103]
Over expression of *GmbZIP123*	*A. thaliana*	10%	Song et al. [70]
Over expression of transcription factors i.e. *GmDof4* and *GmDof11*	*A. thaliana*	6-10%	Wang et al. [69]
Synergistic effect of *WRI1* and *DGAT1*	*N. benthamiana*	59%	Vanhercke et al. [96]

Engineering of multiple genes of triacylglycerol metabolism (Over expression of *DGAT1* and *WRINKLED1* and suppression of *SDP1*)	*A. thaliana*	40%	van Erp et al. [64]
Over expression of *IKU2* (seed development gene)	*A. thaliana*	35%	Fatihi et al. [104]
Over expression of *WRI1*	*B. napus*	20%	Wu et al. [105]

2.5 Reduction in overall yield and oil content in response to biotic stresses

Off late the large-scale cultivation of selected genotypes of *J. curcas* have made it vulnerable to biotic stresses including diseases and pests [106]. Various pathogens are now becoming prevalent in *J. curcas* plantations, reducing its potential to be an ideal bioenergy crop. Biotic stresses are known to reduce yield and seed oil content in different oil plants, including *J. curcas*. Fungal strains of *Alternaria alternate*, *Neoscytalidium dimidiatum*, *Botryosphaeria dothidea* and *Colletotrichum gloeosporioides* were reported to be responsible for infectious spots, root rot, black rot and anthracnose disease, respectively causing reduction in overall yield of Jatropha [17, 18, 107]. Heteroptera, a major pest of *J. curcas* induced abortion of flowers and fruits, along with reduced size, weight and oil content in seeds. Due to anthracnose disease caused by *Colletotrichum gloeosporioides*, there was appearance of dark brown lesions in the fruits leading to malformation of seeds [108, 109]. The powdery mildew caused by the fungus *Pseudoidium jatrophae* is prevalent in various plantations in India, inducing the formation of abundant white or grey colored mycelia in flowers and fruits in *J. curcas* [110, 111].

Also, virus infection in *J. curcas* has major consequences as it reduces overall seed yield and oil content (Table 2.4). Raj et al. [19] reported enhanced sterility rate resulting in few number of flowers in response to cucumber mosaic virus infection in *J. curcas*. Number of fruits per plant and reduction in the size of fruits was observed due to

infestation of mosaic virus in *J. curcas* [112]. Another study reported overall reduction in seeds per capsule and 1000-seed weight due to mosaic disease in *J. curcas* [113]. Overall reduction of 78% and 42% in total seed yield and oil content, respectively was reported in *J. curcas* due to mosaic virus infection [112]. The mosaic virus infection, is therefore, a major concern in the cultivation of *J. curcas*.

Table 2.4 Consequences of viral infection on *Jatropha curcas* yield

Yield trait affected	**Reference**
Few flowers with enhanced sterility	Raj et al. [19]
Reduction in number of fruits per plant, size of the fruit	Jayanna [112]
Reduction in seeds per capsule and reduction in 1000-seed weight	Gao et al. [113]
Overall reduction in yield (upto 78%)	Aswatha Narayana et al. [28]
Overall reduction in oil content (upto 40%)	Jayanna [112]

2.5.1 *Jatropha curcas* mosaic disease

Although, fungal and bacterial strains are responsible for many associated diseases in Jatropha, mosaic disease caused by *Jatropha curcas* mosaic virus is prevalent in the field conditions and is reducing overall fruit yield and quality. It is characterized by leaf curling, blistering, distortion and reduction of fruit size. Mosaic disease of Jatropha is a severe constraint for its full yield and restrictive factor in Jatropha cultivation around the globe. Jatropha mosaic disease was firstly reported from Puerto Rico and subsequently from Jamaica and Cuba [114]. There are many reports describing the outbreak of viruses in *J. curcas* from other parts of world also. Viruses like begomovirus and cucumber mosaic virus (CMV) have been identified from large plantations in India a few years back [19, 28]. Mosaic disease severity ranges from 25% in northern parts of India and up to 47% in southern parts [19, 28]. In India, Jatropha mosaic disease was firstly reported from the state of Karnataka [115].

Jatropha curcas mosaic virus (*Jc*MV) responsible for causing mosaic disease belongs to genera begomovirus of class geminiviruses [19, 28, 113]. It has single stranded

DNA molecule as genetic material and is characterized by rod like appearance. The general mode of *Jatropha curcas* mosaic virus transmission is through whitefly (*Bemisia tabaci*) infestation [113]. *Jatropha curcas* mosaic virus encodes a coat protein (CP) that packages entire genomic and satellite molecules. For virus particle, this CP act as the coat and is required for the successful transmission of virus from infected plants to healthy plants. Begomoviruses originating from similar geographical area consist of highly conserved coat protein due to which it has been adapted to transmission by native vector inhabitants [116]. The core region of coat protein has been shown to be associated for diversity and classification purposes [117]. Therefore, coat protein is being considered as a vital component for existence of begomoviruses and has been used extensively to characterize and establish the connection among different begomoviruses [118].

2.5.2 Molecular basis of mosaic disease response in different plant species

In response to mosaic virus infection, many physiological process associated for overall growth and development of plant gets affected. Increase in the respiration rate was observed for cucumber (*Cucumis melo*) during mosaic virus infection [22]. Decrease in transpiration rate and relative water content have been reported for *Hordeum vulgare* and *Capsicum annuum* respectively in response to mosaic virus infection [23, 24]. Afreen et al. [119] reported increase in phosphorus content for *Daucus carota* whereas Mohamed [120] gave dimensions about increase in proline content for *Beta vulgaris* in response to mosaic virus infection. Decrease in pigment content was also associated with mosaic virus infestation in species like *Capsicum annuum* and *Solanum tuberosum* [23, 25].

Many reports underlying the molecular basis of mosaic disease response exists for plants. Various modern biotechnological approaches have been used in this regard out of which high throughput transcriptomic analysis technique has been studied extensively to gain molecular aspects regarding various diseases including virus infection response in plants. Lu et al. [121] performed deep sequencing of healthy and sequential mosaic virus infected plants of tobacco and found that biological processes, such as pigment metabolism, photosynthesis and plant-pathogen interaction, were linked with virus symptom development. In another study the transcriptome sequencing of African cassava mosaic virus infected cassava leaves gave novel insights into the upregulation of genes associated with degradation of chlorophyll and thus photosynthesis [122]. Recently Choi et al. [123] carried out the transcriptomic investigation of chrysanthemum in response to

tomato spotted wilt virus, cucumber mosaic virus and potato virus X. Genes involved in stress response such as ethylene mediated signaling pathway and chitin response were up-regulated. Tomato spotted wilt virus infection down-regulated genes related to DNA metabolic process such as DNA replication, cytokinesis, histone modification and chromatin organization. Genes related to photosynthesis and flowering were downregulated whereas genes linked to metabolic pathways, transcription and stress responses were found to be upregulated in response to stripe virus infection in rice as revealed by transcriptome profiling of infected seedlings [124]. Similarly, transcriptome sequencing approach has been employed to understand the molecular basis of disease response in plants like alfalfa, beet, capsicum, tomato and white pine [26, 125, 126, 127, 128]. For Jatropha also, many transcriptome based sequencing methodologies have been employed to elucidate molecular responses to important biological processes such as oil biosynthesis, flower formation, abiotic stress, waterlogging, etc. [9, 129, 130, 131, 132], however molecular basis of mosaic virus infection is not understood.

2.6 NBS-LRR genes

Plants have acquired resistance to many pathogens and pests due to the presence of disease resistance (R) genes that encode proteins which protect them from pathogenic organisms [133]. Plants have large number of resistance genes and other defense responsive elements to counter variety of pathogens. The research in the recent past on R-genes and downstream signal transduction mechanism has provided a strong base, thereby paving the way for their use in disease control [134, 135]. The bulk of R-genes in plants are from nucleotide binding site-leucine rich repeat (NBS-LRR) class, providing resistance to a large number of pathogens including parasites, fungi, bacteria, oomycetes, insects, and viruses [133, 136, 137, 138, 139]. NBS proteins are classified into two sub categories based primarily on domains and motifs (Figure 2.4). Those having N-terminal domain with resemblance to the Toll and interleukin-1 receptors are designated as TIR proteins, and those without a TIR domain are categorized as non-TIR proteins [140]. A few of non-TIR proteins encode an N-terminal coiled-coil (CC) domain that may be involved in signaling and interaction of proteins [139, 141]. The NBS domains linked with both TIR and non-TIR proteins consist of a P-loop (kinase-1), kinase-2, kinase-3, and some additional short motifs of unknown role [142]. The NBS domain functions by binding ATP [143], and the C-terminal leucine rich repeat (LRR) is implicated in pathogen binding and regulation of signal transduction [139, 140]. TIR domains are also involved in resistance specificity

determination and signaling [140, 144]. 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 *A. thaliana* code for TIR domains [145], members of this subclass appear to be absent in cereal species [146, 147]. This findings suggests that since divergence occurred >200 million years ago [148], TIR domain association with NBS-encoding genes was preserved by dicots but lost in monocots. NBS families are ancient, but gene duplication and gene loss events have changed the composition of these gene subfamilies over time [149].

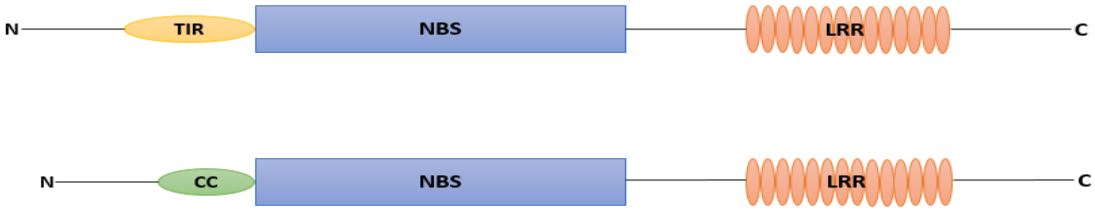


Figure 2.4 Common structure of NBS-LRR gene (NBS- Nucleotide binding site; LRR- Leucine rich repeat; TIR- Toll and interleukin-1 receptors domain; CC- Coiled coil domain; N- Amino terminus; C- Carboxyl terminus) [150]

In the recent past, many resistance genes, including NBS-LRR genes have been employed to produce genetically modified and transgenic disease resistant varieties. In case of tobacco, N gene encoding TIR-NBS-LRR was transferred to develop transgenic lines which showed resistance to the mosaic virus [151]. Similarly, transgenic tobacco lines were developed using common bean TIR-NBS-LRR gene, *RT4-4* exhibiting resistance towards mosaic virus from tomato or pepper [152]. In tomato, *Bs2* gene encoding NBS-LRR protein has been transferred to develop resistance against bacterial spot disease [153]. Another gene responsible for bacterial blight resistance, *Xa21* was introduced into Chinese rice varieties and the transgenic plants exhibited resistance to bacterial blight [154]. In case of wheat, the *Pm3b* gene has been introgressed which showed resistance against powdery mildew [155]. In another important study, *RPS4* and *RRS1*, two NBS-LRR type R genes

exhibited resistance to members of brassicaceae and solanaceae by providing immunity against various bacterial and fungal pathogens [156].

For Jatropha, 92 NBS-LRR genes have been identified previously which is quite a small number in comparison to other sequenced plant genomes with same range of genome sizes [13]. NBS-LRR genes number vary in different plant species, irrespective to the genome size. For example, Arabidopsis genome (125 Mb) contains 149 genes [157], rice genome (420 Mb) contains 535 genes [147], potato genome (840 Mb) has 438 NBS-LRR genes [158], soybean genome (1,115 Mb) comprise 319 NBS-LRR genes [159], populus genome (500 Mb) having about 400 NBS-LRR genes [160] and cucumber genome (350 Mb) contains 57 NBS-LRR genes [161].

2.7 Transcription factors related to defense response

In their natural habitats, plants regulate the immunity and response to other stresses by transcription factors which are considered as potential targets for engineering plant defense [162]. Association between activating and repressing transcription factors from many families regulate the defense response expression of the linked genes [163]. Many of the defense or disease resistance related transcription factors have been studied recently including the TGA family of basic domain-leucine zipper (*bZIP*) proteins [164, 165], the *MYB* proteins [166], the ethylene responsive element binding factors [*ERFs*, having a DNA binding domain also reside in the APETALA2 (*AP2*) protein family], the *WRKYs* [31], and the *Whirly* family [162] (Table 2.5).

Advances in sequencing technology has intensely extended genomic and transcriptomic information which is available publicly. To unravel the intricate mechanisms and transcriptional reprogramming that function in defense responses, genome, transcriptome and bioinformatics approaches in integration with experimental techniques should be incorporated (Figure 2.5). Whole genome analysis of defense-related transcription factors have been done in many plant species. In recent years, the genome and transcriptome wide identification and characterization of members of defense response associated largest transcription factor family i.e. *WRKY* have been done in many plant species. Genome wide identification and characterization of *WRKY* transcription factors conferring resistance to biotic stresses was performed in populus by Jiang et al. [167]. Genome wide expression analysis of *WRKY33* following *Botrytis cinerea* infection in *A. thaliana* revealed differential transcriptional reprogramming and indicated involvement of

WRKY33 in the defense mechanism [168]. Similarly, by genome wide identification, *WRKY45* was found to be a master regulator of the transcriptional cascade governing defense response in benzothiadiazole induced disease resistance in *A. thaliana* [169]. Transcriptome based identification and analysis of *WRKY* transcription factors regulating defense against leaf rust was performed in wheat recently by Satapathy et al. [170]. For *bZIP* transcription factors, the genome wide identification approach has been used [171]. In Chinese cabbage genome, 291 putative *AP2/ERF* transcription factors regulating resistance against disease and biotic stresses were identified [70]. In Arabidopsis, 118 transcription factors of families *APETALA2*/ ethylene responsive element binding proteins, *MYB* domain-containing proteins, *C2H2* zinc finger proteins and *WRKY* domain showing response to defense elicitor, Chitin were identified using affymetrix Arabidopsis whole-genome array [172]. In another example, for soybean, an important crop species, biotic stress response related *trihelix-GT* and *bHLH* transcription factors were identified and characterized using *in silico* approach [173].

In the past, many transgenic crop and model plants with improved disease resistance have been developed by over expressing the defense related transcription factors. Over expression of *WRKY* and *ERF* transcription factors have resulted in developing disease resistant varieties of many plants [174]. Over expression of the defense associated transcription factors can provide resistance to many dissimilar pathogens also. *AtMYB44*, a transcription factor of *MYB* family was found to regulate the plant defense response against aphid [175]. Arabidopsis transcription factor *RAP2.6* over expression has resulted in enhanced resistance to beet cyst nematode *Heterodera schachtii*. Recently, Sun et al. [176] used virus-induced gene silencing (VIGS) approach to study the role of *NAC* transcription factors in disease resistance against *Magnaporthe grisea.* Similarly, overexpression of *NAC* transcription factor from tomato positively regulates defense response against infection of pathogen, *Botrytis cinerea* [177].

2.8 Disease resistance in *J. curcas*

Some species of genus Jatropha such as *J. integerrima* has been found to withstand disease pressure in the field conditions compared to *J. curcas*. Commercial cultivation of selected genotypes of *J. curcas* has predisposed to a plethora of biotic stresses, including insect pests and fungal, viral and bacterial diseases. Despite prevalence of *J. curcas* towards

various biotic stresses, our knowledge regarding disease resistance and defense response mechanisms are still limited.

Table 2.5 Transcription factors involved in plant defense response

TF family	**Reference**
MYB	Katiyar et al. [166]
WRKY	Dong et al. [178]
AP2-EREBP	Ohme-Takagi et al. [179]
bZIP	Alves et al. [165]
CBF	Sakuma et al. [180]
TFIIA	Jiang et al. [181]
NAM	Collinge and Boller, Singh et al. [182, 183]
Whirly	Desveaux et al. [162]
SBP/SPL6	Padmanabhan et al. [184]
Homeo-domain	Luo et al. [185]

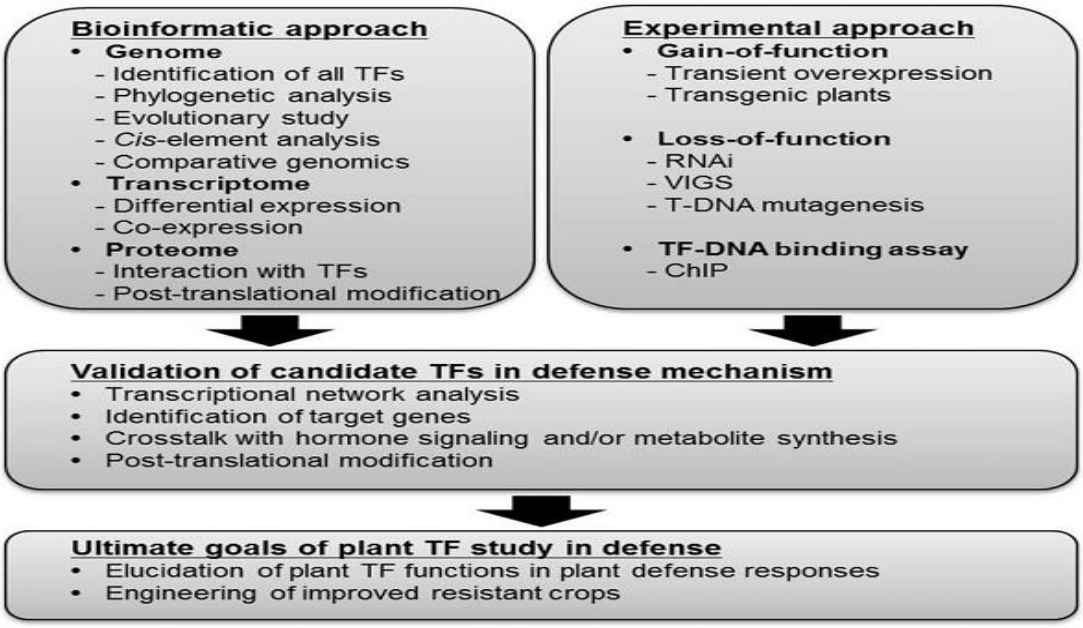


Figure 2.5 Scheme of functional studies to identify transcription factors involved in defense response [186]

The molecular mechanisms underlying biotic stresses in *J. curcas* is not well documented however, there are various reports describing the role of molecular entities governing various abiotic stresses such as salt stress, drought stress and cold stress. A novel *AP2/ERF* transcription factor from *J. curcas* has been identified which confers drought and salt tolerance to transgenic tobacco [187]. *NAC1*, another transcription factor responsible for multiple abiotic stresses was identified and characterized from *J. curcas* [188]. Also, Li et al. [189] showed the association of *MYB* transcription factor in response to salt stress. Although few NBS-LRR genes have been identified in *J. curcas*, this identification approach was applied through genome mining strategy which may contain pseudogenes [13]. There is no report exist as of today on identification of defense response related transcription factors in *J. curcas*, however over-expression of a transcription factor *NAC*, showed increased susceptibility to pathogens [188].

The comprehensive review of literature therefore pinpoints the following gaps in our understanding towards this bioenergy crop, *J. curcas*:

(1) No information is available on genetic factors contributing towards higher oil accumulation in high oil content genotypes.

(2) No information is available on molecular mechanisms underlying virus (disease) response.

(3) Scant information on molecular components associated to disease resistance/defense response.

CHAPTER 3

MATERIALS & METHODS

The present study was carried out at the Jaypee University of Information Technology, Waknaghat, Himachal Pradesh, India. The material used and the methodologies followed to accomplish objectives of the study are described here under:

3.1 Plant material

Fruit samples of unripened (U; 35 DAP), ripened (R; 60 DAP) and mature (M; 85 DAP) developmental stages of high oil content (42%, IC 561235; SH and NH) and low oil content (30%, IC 561227; SL and NL) genotypes of *J. curcas* were collected from the experimental farm of Himalayan Forest Research Institute, Shimla at Sunni (Shimla District, 1021 m altitude, 31 14' N, 77 70'E) and Nalagarh (Solan District, 521 m altitude, 30 50'N, 76 58'E), Himachal Pradesh, India (Figure 3.1). Healthy (JH) and symptomatic virus infected (JV) leaves were collected from mature tree of *Jatropha curcas* genotype IC 561235 from experimental farm of Himalayan Forest Research Institute at Jwalaji (508 m altitude, 31º 51' N, 76º 18' E), Himachal Pradesh, India (Figure 3.2). Tissues were immediately frozen in liquid nitrogen and stored at -80^{0}C for further use.

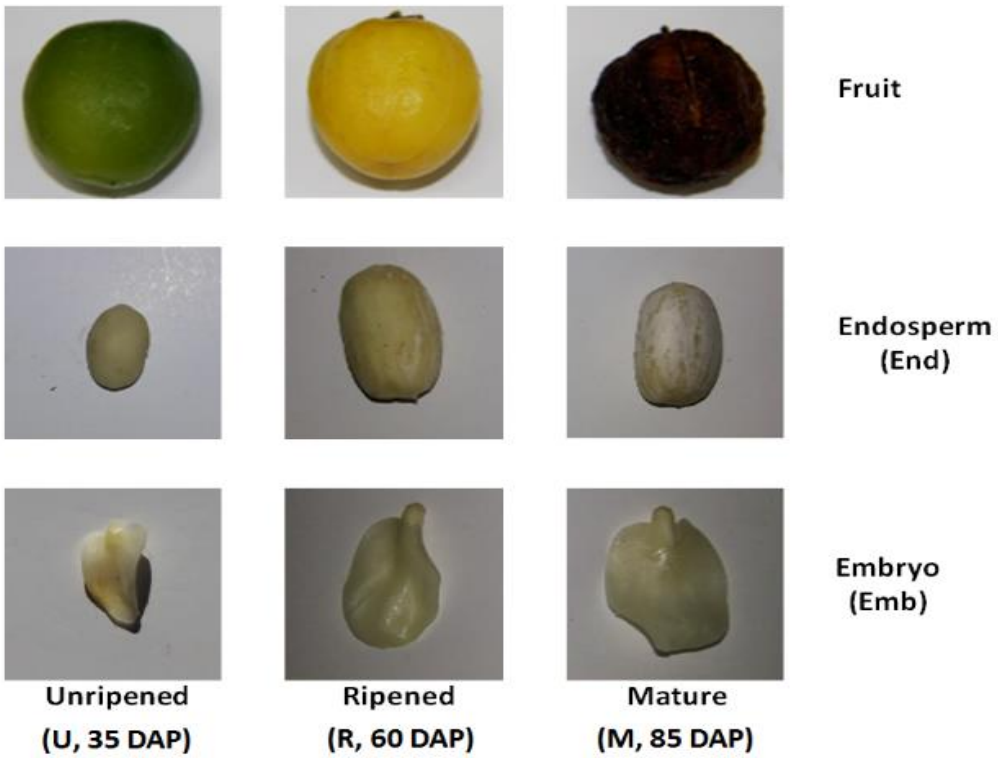


Figure 3.1 Developmental stages of endosperm and embryo used for expression analysis

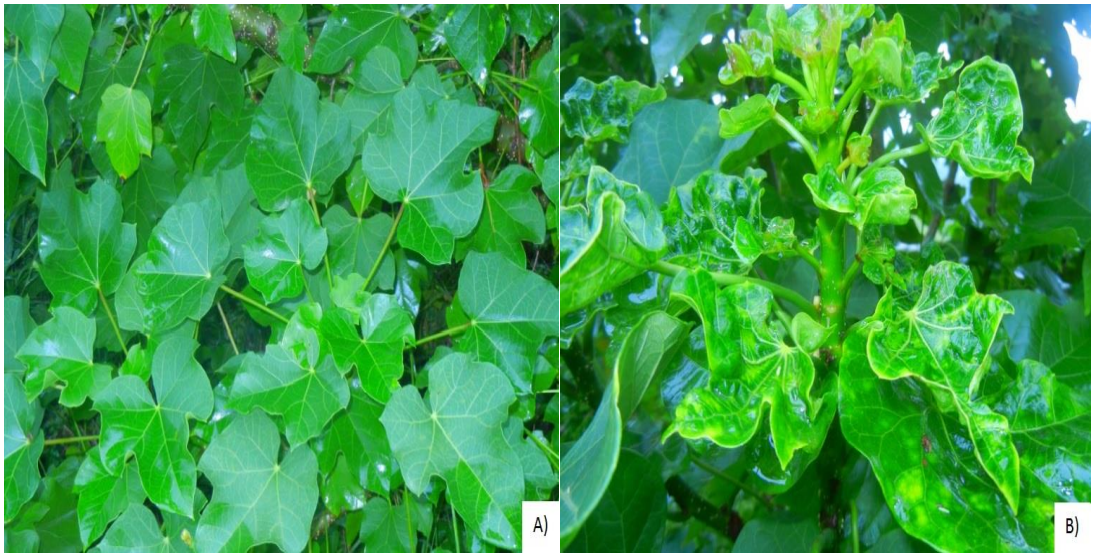


Figure 3.2 A) Healthy Jatropha plant (JH) **B)** Mosaic virus infected Jatropha plant (JV)

3.2 Oil extraction and content estimation

Seed samples from dried mature stage of genotypes i.e. IC 561235 and IC 561227 were used for oil extraction according to the protocol described by Kaushik and Bhardwaj [190] with some specific parameters taken into consideration (Table 3.1).

The seed oil content was calculated using the following formula:

Seed oil content (%) = Weight of the oil extracted (in g) x 100/ Weight of the seed powder taken for the extraction (in g)

3.2.1 Selection of high versus low oil content genotype

On the basis of oil content estimation, high (IC 561235; 42%) and low (IC 561227; 30%) oil content genotypes of *J. curcas* were selected and were further used in different experiments.

3.3 Primer designing

Specific primers of genes and transcription factors were designed using Primer 3 (http://bioinfo.ut.ee/primer3-0.4.0/)

Table 3.1 Parameters studied for oil content analysis

Parameter	**Description (all values are approx.)**
Weight of 15 healthy fruits (g)	45
Individual fruit weight (g)	3
Individual fruit length (mm)	28
Individual fruit diameter (mm)	20
Individual shell weight (g)	0.90
Number of seeds	3
Total seed weight (g)	2.4
Length of each seed (mm)	19
Diameter of each seed (mm)	10

(g: gram, mm: millimeter)

3.4 RNA extraction and cDNA synthesis

Endosperms and embryos were excised from the seeds of various developmental stages of both high and low oil content genotypes of both locations, Sunni and Nalagarh. Total RNA was isolated from embryo and endosperm at unripened, ripened and mature developmental stages of *J. curcas* seeds by using Qiagen RNeasy mini kit and digested with DNase I (Invitrogen). From virus infected (JV) and healthy (JH) leaves, total RNA was isolated using RaFlex Total RNA isolation Kit (Genei) as per manufacturer's instruction. The quality of total RNA was checked on 1% denatured agarose gel (1μg) for the presence of 28S and 18S bands, along with absorbance spectrum at 260 nm and 280 nm wavelengths. Further, total RNA was quantified using Qubit RNA BR kit (fluorometer). First-strand cDNA synthesis was done using Verso cDNA synthesis kit (Thermo Scientific, USA) from total RNA (2mg) template as per manufacturer's instructions.

3.5 DNA isolation and detection of virus

Total DNA was isolated from healthy (JH) and virus infected (JV) leaves using CTAB extraction method with minor modifications. Jatropha mosaic virus coat protein gene (NCBI accession no. 9247600) was used for polymerase chain reaction (PCR) amplification performed on thermocycler (Applied Biosystems) in 25 μl reaction volume (Table 3.2). PCR was performed on 30 ng of genomic DNA with primer pairs, Mg2+, dNTPs (Intron Technologies), and Taq DNA polymerase (Intron Technologies).

Amplification programs included 94°C for 5 min, 30 cycles of 94°C for 45 s, annealing temperature of 54.6°C for 45 s, 72°C for 2 min, and a final extension of 7 min at 72°C. Amplified products on 1% agarose gel were analyzed using the gel documentation system AlphaImager EP (Alpha Innotech Corp., USA).

Table 3.2 Primer sequences of Jatropha mosaic virus coat protein gene

Forward primer	**Reverse primer**	**Annealing temp. (°C)**
AACTTCGACAGTCCATTCAG	ATACAGGATTAGAGGCGTGA	54.6

3.6 Data collection

The nucleotide sequences of 18 genes of FA and TAG biosynthetic pathway (Figure 3.3) were retrieved from Jatropha Genome Database (http://www.kazusa.or.jp/jatropha/) and Genbank, NCBI (http://www.ncbi.nlm.nih.gov/genbank/). For transcription factors associated with oil biosynthesis and accumulation, the transcriptome data was downloaded from sequence read archive (SRA) module of NCBI with accession no. SRX809788 (Developing seeds with 45 days after pollination). For NBS-LRR genes and defense response related transcription factors, the transcriptome data of Jatropha and castor bean were downloaded from sequence read archive (SRA) module of NCBI with accession nos. SRR087417 and ERA047687 respectively. The whole genomes of Jatropha and castor bean were downloaded (ftp://ftp.kazusa.or.jp/pub/jatropha/; http://castorbean.jcvi.org/downloads.php). Velvet software [191] was downloaded from (http://www.ebi.ac.uk/*zerbino/velvet/) for assembly of transcriptome SRA files (NGS data). For similarity search, all the available NBS-LRR mRNA sequences were downloaded from the GenBank module of NCBI. Perl program, pfam_scan.pl and Pfam library of hidden Markov models (HMMs) of protein families were retrieved from Pfam website (http://pfam.janelia.org/) for domains prediction in protein sequences translated from transcripts.

3.7 Expression analysis of FA and TAG biosynthesis pathway genes through reverse transcription-quantitative real-time PCR (RT-qPCR)

The expression profiles of 18 genes involved in FA and TAG biosynthetic pathway were investigated by RT-qPCR. Using gene specific primers (Table 3.3), RT-qPCR was

performed on CFX96 system (BioRad, USA) with the iScript one step RT PCR kit (Bio-Rad, USA).The PCR protocol was as: denaturation for 5 min at 94°C, followed by 40 cycles each of denaturation for 20 s at 94°C, annealing for 30 s at 50-55°C, followed by one elongation step for 20 s at 72°C. All quantitative PCR experiments were repeated with three replicates. For calculating transcript abundances, *26S rRNA* and *GAPDH* were used as internal controls to normalize the expression data as previously described [192]. These candidate reference genes were considered for their expression across different tissues. Unripened stage was taken into consideration for calculating relative fold expression of genes in ripened and mature stage and was kept as calibrator. Relative fold changes were determined from the Cq values using the comparative Ct (DDCt) method described by Schmittgen and Livak [193]. Standard deviation with percentage error was used to statistically evaluate significant differences between treatments.

3.7.1 Statistical analysis

Principal component analysis (PCA) was performed using XLSTAT to correlate the expression of FA and TAG biosynthesis pathway genes and different experimental conditions [194]. Plot between variability, eigen-values and principal components were generated. To determine relative expression of FA and TAG biosynthetic pathway genes in endosperm and embryo of Jatropha, the heat map was made for RT-qPCR data using GenEx software (V 1.1).

3.7.2 *In-silico* promoter analysis

To identify putative regulatory elements associated with lipid accumulation, *in-silico* analysis of promoter regions of the genes was carried out. The genes showing elevated expression in oil accumulating developmental stages of high oil content genotype along with the genes previously identified [33, 11] showing consistently higher expression among developmental stages of oil accumulation were analyzed for promoter analysis (Table 3.6). The gene structure was elucidated by using FGENESH. For each gene, 2 kb upstream region (potential promoter region) of the translational start site (TSS) was selected from Jatropha genome database. TSSP was used for *in-silico* identification of promoterregions. Plant Care (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) and PLACE (http://www.dna.affrc.go.jp/PLACE/) were used for the identification of cis-regulatory elements.

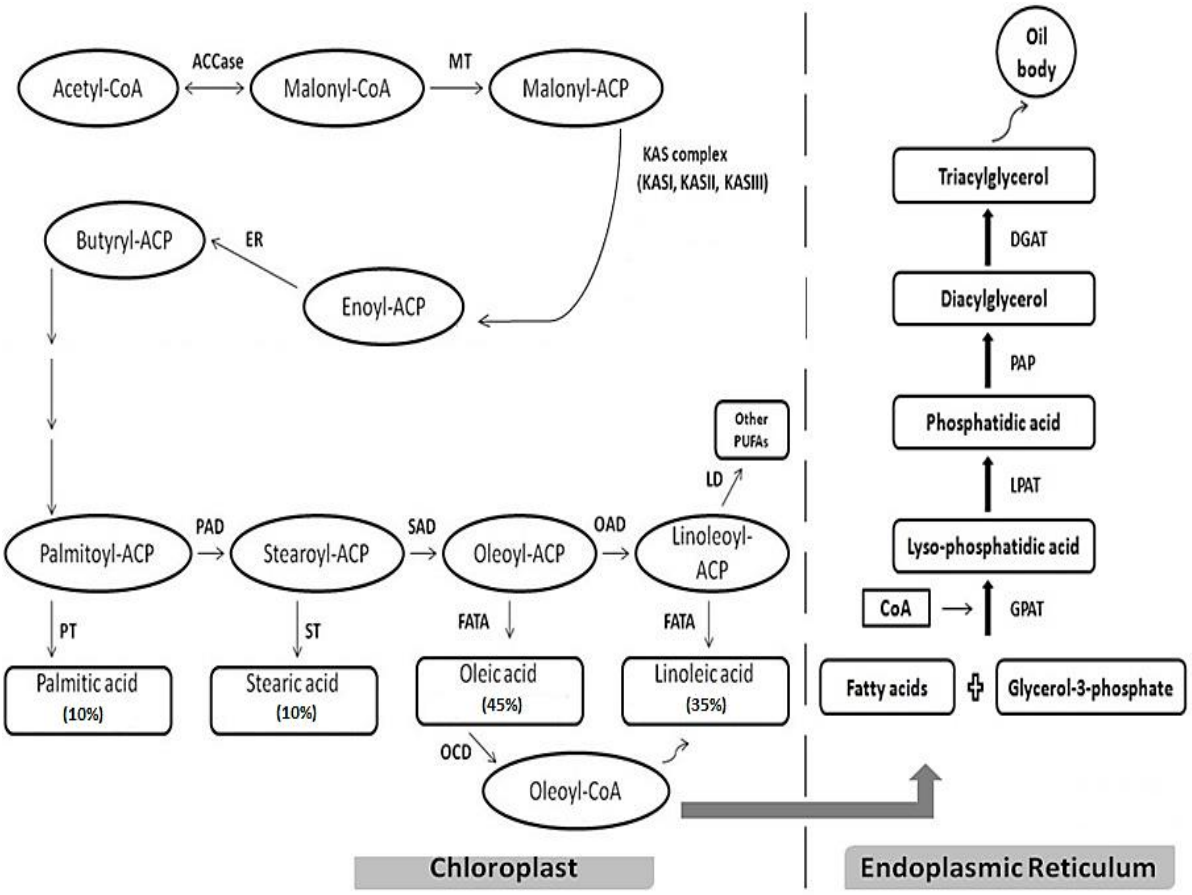


Figure 3.3 Schematic representation of fatty acid (FA) and tri acyl glycerol (TAG) biosynthesis pathway in Jatropha [16]. *ACCase*- Acetyl-CoA Carboxylase; *DGAT*- Diacylglycerol acyl transferase; *ER*- Enoyl Reductase; *FATA*- Linoleoyl Thioesterase; *GPAT*- Glycerol-3-phosphate acyl transferase; *KASI*- β-ketoacyl-ACP synthase I; *KASII*- β-ketoacyl-ACP synthase II; *KASIII*- β-ketoacyl-ACP synthase III; *LD*- Linoleoyl Desaturase; *LPAT*- Lyso-phosphatidic acid acyltransferase; *MT*- Malonyl Transferase; *OAD*- Oleoyl-ACP Desaturase; *OCD*- Oleoyl-CoA Desaturase; *PAP*- Phosphatidic acid Phosphatase; *PT*- Palmitoyl Thioesterase; *SAD*- Stearoyl-ACP Desaturase; *ST*- Stearoyl Thioesterase; *PAD*- Palmitoyl-ACP Desaturase

Table 3.3 Primers from FA and TAG biosynthetic pathway genes used for expression analysis through RT-qPCR

S.No.	Gene	Gene abbrev.	Forward primer	Reverse primer	Annealing temperature (°C)
1	Acetyl-CoA carboxylase	*ACCase*	GGAGAAGCAACACCACATAC	CCACAGGAACAAGAGGAGTA	52
2	Malonyl transferase	*MT*	CTACCCTTGCTTTTCTGCT	GGCTTATCACAGTTGAATCC	51
3	β-ketoacyl-ACP synthaseI	*KASI*	TGGGTGGTCTTACAGTCTTT	AGGTAGCACAAGCAGTTGAG	53
4	β-ketoacyl-ACP synthase II	*KASII*	AGAAAGACACTTCACGCTCA	CTACAAGGCTCAAACGCTAA	55
5	β-ketoacyl-ACP synthaseIII	*KASIII*	GTTGGTAGTGGTTCAGCAGT	CTCAGTGGCTAAGGAAATCA	52
6	Enoyl reductase	*ER*	AGTATTGTTGGGCAGTGTGT	GAGACCCTTGACATTCTTCA	52
7	Palmitoyl-ACP desaturase	*PAD*	CTTGATGTATGATGGTCGTG	GTAATCCTGAGCCTTTTGC	56
8	Stearoyl-ACP desaturase	*SAD*	CACCCCAGAAGATTGAGATA	AGTTCCCTGACTTGTTCATC	54
9	Oleoyl-ACP desaturase	*OAD*	TCCTACCTTCCAGTGTCGTA	TCTCCCTAAGTTCCCTGACT	51
10	Palmitoyl thioestrase	*PT*	CTAACGCACAACAGAAACG	CAAAGTCCAACAACAGCAG	55
11	Stearoyl thioestrase	*ST*	AGATTCCAAGTCCACCAAG	GTCACCCTCATTTTTCACAC	54
12	Linoleoyl thioestrase	*FATA*	TCTGAAGGACTATGCCACTG	CCTCTGGAAATGCTAATCTG	51
13	Linoleoyl desaturase	*LD*	GAGAATCAGCCACAGAACTC	CCCACAGATAGAAAGGGTAA	50

14	Oleoyl-CoA desaturase	*OCD*	GATGTTCTGGGCTATCTTTG	GTGGTGAATCCTATGGCTAA	50
15	Glycerol-3-phosphate acyl transferase	*GPAT*	CGGAAGCAGTCATTTACAAC	GGAACCTTTTTGAGACCTG	56
16	Lyso-phosphatidic acid acyltransferase	*LPAT*	CATATGCAACCACAGAAGTG	TCGAGGAAGAGGTATTCAGA	53
17	Phosphatidic acid phosphatase	*PAP*	ATTCAGTTGGGTTCTCACAC	TCTGTCATCATATCCCTTCC	52
18	Diacylglycerol acyl transferase	*DGAT*	TGCTGTCTTACCCTCCCTAT	GACCCACAACTGAGAATCAC	52
19	26S ribosomal RNA*	*26S rRNA*	CACAATGATAGGAAGAGCCGAC	CAAGGGAACGGGCTTGGCAGAATC	58
20	Glyceraldehyde-3-Phosphate dehydrogenase*	*GAPDH*	TTGCCATCAATGACCCCTTCA	CGCCCCACTTGATTTTGGA	56

*Reference genes (Internal controls)

3.7.3 Cloning of promoter region

For experimental validation, the promoter region of *SAD* (Rate limiting gene) was cloned in high oil content (IC 561235) and low oil content genotype (IC 561227). Using *SAD* promoter region specific primers, PCR was performed on 30 ng of genomic DNA with varying amounts of primer pairs, Mg2+, dNTPs and Taq DNA polymerase. The amplification programs were as: 1 cycle of 94°C for 3 min, 30 cycles of 94°C for 30 s, 30 cycles of annealing temperature (49°C) for 45 s, 30 cycles of 72°C for 2 min and a final extension of 1 cycle of 7 min at 72°C. 10µl of each PCR product was mixed with 2µl of 6x gel loading dye (0.2% xylene cyanol dye, 0.2% bromophenol blue, and 30% glycerol) and was electrophoresed in a 1.2% agarose gel prepared in 1x Tris acetate-EDTA (TAE) buffer. The gels were visualized using gel documentation system Alpha Imager EP (Alpha Innotech Corp., USA). For cloning of PCR products, pGEMT vector (Promega) was used and then were further sequenced. Basic local alignment search tool (BLAST) (http://www.ncbi.nlm.nih.gov/BLAST) was used for calculating sequence similarities.

3.8 *In-silico* identification of transcription factors (TFs) controlling oil biosynthesis

The transcriptome data was downloaded from sequence read archive (SRA) module of NCBI with accession no. SRX809788. Velvet software was downloaded from (http://www.ebi.ac.uk/*zerbino/velvet/) for assembly of transcriptome SRA files (NGS data). Perl program, pfam_scan.pl and Pfam library of hidden Markov models (HMMs) of protein families were retrieved from Pfam website (http://pfam.janelia.org/) for domains prediction in protein sequences translated from transcripts. Literature based mining approach was employed to identify TFs important in oil biosynthesis and accumulation (Figure 3.4). A total of 9 different TF families were found and a 'master list' was prepared (Table 3.4). All 9 TF families were examined in Pfam to find out domains associated with each family and significant role in oil accumulation. PlantTFcat, a plant based TF database (http://plantgrn.noble.org/PlantTFcat/familylist.do) was used for identification and classification of TFs involved in oil accumulation. Transcript abundancy estimation was carried out using RSEM [195]. The transcript abundance of the transcripts from the transcriptome were calculated using FPKM parameter of RSEM package. All the parameters were kept default in the query option.

Table 3.4 Transcription factors involved in oil biosynthesis and accumulation

TF Family	Pfam Ids	Reference
Dof	PF02701	Wang et al. [69]
AP2	PF00847	Ma et al. [72]
B3 domain	PF02362	Palaniswamy et al. [73]
GATA	PF00320	Fobert [74]
bHLH	PF00010	Courchesne et al. [75]
MYB	PF00249	Liu et al. [76]
bZIP	PF00170	Song et al. [70]
HD-Zip	PF11569	Chew et al. [77]
CBF	PF02045, PF03914	Fobert [74]

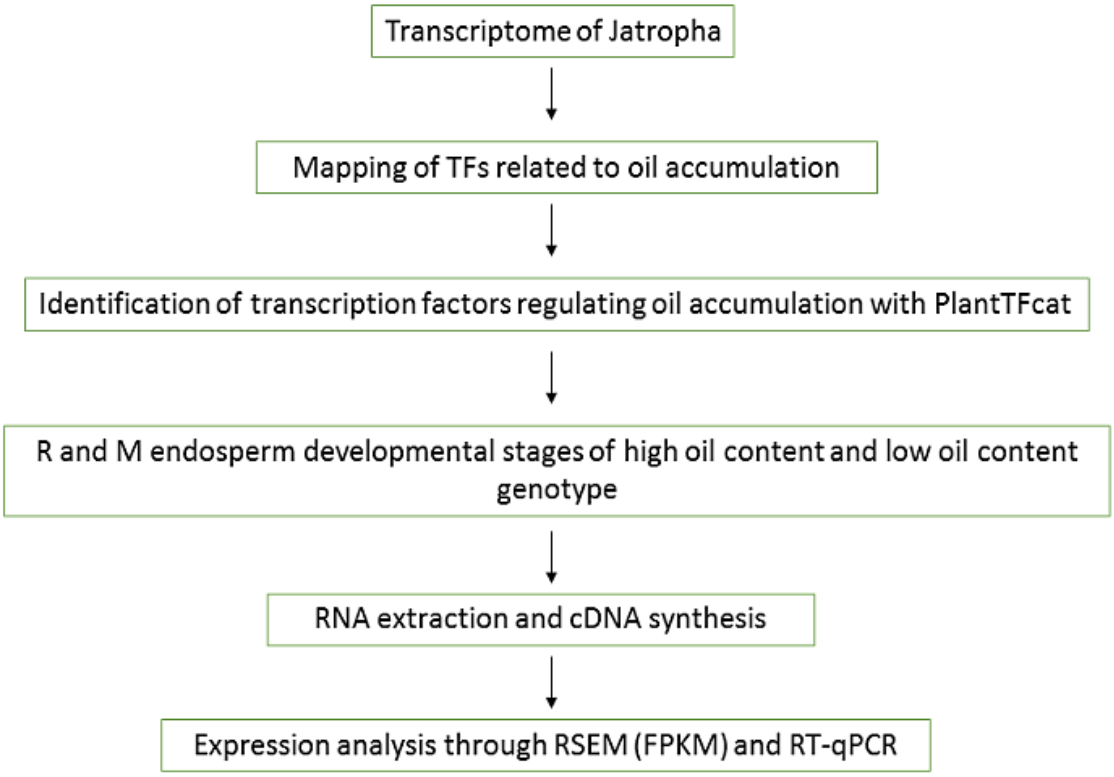


Figure 3.4 Methodology followed for identification of transcription factor families regulating oil biosynthesis and accumulation

3.8.1 Expression analysis of transcription factor genes through reverse transcription-quantitative real-time PCR (RT-qPCR)

The expression profile of genes encoding transcription factors identified as per *in silico* analysis was done by RT-qPCR. Ripened (R) and mature (M) developmental stages of endosperms of high oil content and low oil content genotypes of location Nalagarh (NH) were used for this analysis (Figure 3.4). Using specific primers (Table 3.5), RT-qPCR was performed on CFX96 system (Bio-Rad, USA) with the iScript one step RT PCR kit (Bio-Rad, USA).The PCR protocol was as: denaturation for 5 min at 94°C, followed by 40 cycles each of denaturation for 20 s at 94°C, annealing for 30 s at 50-55°C, followed by one elongation step for 20 s at 72°C. For calculating transcript abundances, *26S rRNA* and *GAPDH* were used as internal controls to normalize the expression data. Relative fold changes were determined from the Cq values using the comparative Ct (ΔΔCt) method [193]. Standard deviation with percentage error was used to statistically evaluate significant differences.

Table 3.5 Primers for genes encoding transcription factors regulating oil biosynthesis and accumulation used in RT-qPCR

S.No.	TF family	Forward primer	Reverse primer	Annealing temperature (ºC)
1	*bZIP*	TGCATATCAGAGAGTTGCAG	CTAAGAGTGTGGGATGCAAT	51
2	*Dof*	AGAGCTGGACTTAGGCTTTT	ATGGAGACTCGAGAAGAACA	53
3	*MYB*	AGGTGCCAGACTAGAAATCA	CTTCGTTTCCTCTCATCAAG	53
4	*bHLH*	GATGAGGAGTTAGCAGGTCA	ATCAGTTCTCTCGTCACCAC	54
5	*CBF*	CAAAACCATGGGTGAGTAGT	AGTATCTCGGGAGAGGATTC	55
6	*AP2*	CTCCATCAAAGAGCAAGAAG	GAATGTCCCTAACCAATGTC	50
7	*26S rRNA**	CACAATGATAGGAAGAGCCGAC	CAAGGGAACGGGCTTGGCAGAATC	58
8	*GAPDH**	TTGCCATCAATGACCCCTTCA	CGCCCCACTTGATTTTGGA	56

*Reference genes (Internal controls)

Table 3.6 Genes selected from Xu et al. [33] and Gu et al. [11] showing elevated transcript abundance vis-a-vis seed development and oil biosynthesis

Gene	Abbreviation	Function in lipid biosynthesis	Reference
Oleosin 1	*Oleosin 1*	Oil bodies regulation	Siloto et al. [196]
Oleosin 2	*Oleosin 2*	Oil bodies regulation	Siloto et al. [196]
Phospholipid diacylglycerol acyltransferase	*PDAT*	Acyl-CoA-independent triacylglycerol biosynthesis	Stahl et al. [197]
Choline kinase	*CLK*	phosphatidylcholine (Phospholipids) formation	Gibellini and Smith [198]
Diacylglycerol kinase 1	*DGK1*	Biosynthesis of Phosphatidic acid from diacylglycerol	Han et al. [199]
Enoyl-CoA hydratase	*ECH*	Degradation of saturated fatty acids	Allenbach and Poirier [200]
3-ketoacyl-CoA reductase isoform 2	*KCR2*	Elongation of fatty acids	Puyaubert et al. [201]
Ketoacyl-CoA synthase	*KCS*	Formation of very long chains of mono unsaturated fatty acids	Taylor et al. [202]
Long-chain acyl-CoA synthetase 8	*LACS8*	Esterification of free fatty acids to fo rm acyl-CoAs	Shockey et al. [203]
Lipase	*Lipase*	Hydrolysis of triacylglycerol to form free fatty acids and glycerol	Quettier and Eastmond [204]
Sterol desaturase	*SD*	Desaturation of fatty acids	Taton and Rahier [205]

3.9. Illumina NextSeq 2 x 150 PE library preparation

The 7 paired-end cDNA sequencing libraries were prepared from healthy (JH) and virus-infected (JV) leaf tissues using illumina TruSeq stranded total RNA library preparation kit and as per protocol described by manufacturer (Illumina). Briefly, rRNA was depleted from total RNA followed by fragmentation. The fragmented rRNA depleted RNA was converted into first-strand cDNA, followed by second-strand cDNA generation, A-tailing, adapter ligation and finally ended by index PCR amplification of adaptor-ligated library. Library quantification and validation was performed using Qubit dsDNA HS kit and high sensitivity assay kit, respectively. RNA-Seq libraries were prepared using Illumina TrueSeq stranded total RNA HT (with Ribo-Zero plant) kit using 1 μg of total RNA. The mean size of the fragment distribution is ranging from 550-700 bp.

3.9.1 Data generation and mapping of reads to genome

The raw data was generated on NextSeq. The raw reads were filtered using Trimmomatic (v 0.30) with quality value QV > 20 and other contaminants such as adapters were also trimmed. The reference genome of *J. curcas* was downloaded from Jatropha genome database (http://www.kazusa.or.jp/jatropha/). The Illumina NextSeq transcriptome data for both samples i.e. JH and JV were separately mapped to the Jatropha reference genome using BWA version 0.7.5a (http://bio-bwa.sourceforge.net/) with default settings. The software package SAMtools (http://samtools.sourceforge.net/) was used to convert sequence alignment/map (SAM) file to sorted binary alignment/map (BAM) file. Mapped reads ratio (MRR) to the reference in each dataset was calculated by applying flagstat command of SAMtools software to the BAM file (Figure 3.5).

3.9.2 Differential gene expression analysis

The CDSs from Jatropha GFF file were used to study the gene expression analysis. The expression analysis of these genes was carried out with R package DESeq. The expression of genes was calculated in terms of FPKM (Fragment per kilobase per million mapped reads). The FPKM values for each gene were calculated for healthy (JH) and virus infected (JV) samples with DESeq package. These FPKM values were further used to calculate the log fold change [$\log_2$ (FPKM_JV/FPKM_JH)]. The analysis was carried out to identify commonly expressed genes between JH and JV samples respectively. These genes were further divided on the basis of their statistical significance (depending on whether p value is less than 0.05 for their significant expression). These genes were further categorized as up-regulated genes and down-regulated genes in JV as compared to JH. Gene that exhibited an p-value<0.05 and estimated absolute log2fold were determined to be significantly expressed genes (Figure 3.5).

3.9.3 Heat map analysis

A complete linkage hierarchical cluster analysis was performed on top 100 differentially expressed genes (50 up regulated and 50 down regulated in JH or JV) obtained from DESeq using multiple experiment viewer (MEV v4.8.1). Heat map was constructed using the log-transformed and normalized value of genes based on Pearson's uncentered correlation distance as well as based on complete linkage method.

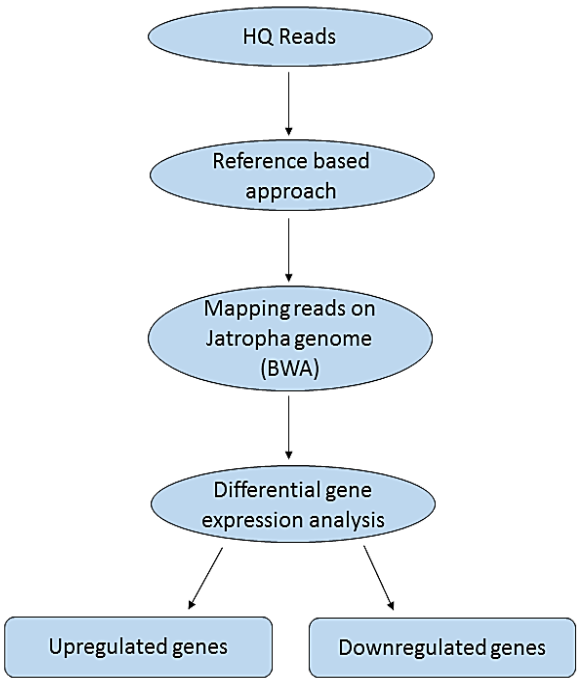


Figure 3.5 Flowchart depicting bioinformatics methodology followed for differential gene expression analysis in JH and JV

3.9.4 Gene ontology analysis

For functional annotation of the predicted CDS in both JV and JH, BLAST2GO program was used with default parameters to retrieve GO annotation which distinguishes on the basis of molecular function, biological process and cellular component ontologies [206]. Gene ontology analysis specifies all the annotated nodes comprising GO functional groups such as cellular component, biological process and molecular function. Main GO categories were determined after the genes were further analyzed for BLAST, gene mapping and annotation.

3.9.5 Pathway analysis

CDSs of JV and JH samples were functionally annotated by KAAS (KEGG automatic annotation server) with BLAST comparisons against KEGG GENES database. KEGG gene database has an advantage over other databases, as it is a single resource for cross species depiction by assigning KEGG orthology to all existing genomes. The BBH (Bi-directional best hit) option was used to assign KO terms. For pathway mapping, KEGG Orthology database (http://www.genome.jp/kegg/ko.html) was used.

3.9.6 SNP identification

The Illumina NextSeq transcriptome data for both samples were separately mapped to the Jatropha reference genome using BWA version 0.7.5a (http://bio-bwa.sourceforge.net/) with default settings. The software package SAMtools (http://samtools.sourceforge.net/) was used to convert sequence alignment/map (SAM) file to sorted binary alignment/map (BAM) file. The alignment file was used for SNP detection using SAMtools. The SNPs were detected in both the samples using the samtools mpileup pipeline. SNPs with quality score more than 20 and read depth over 5 with flanking of 100 bp were filtered as high quality SNPs.

3.9.7 Co-expression network analysis

A gene co-expression network (GCN) is an undirected graph, where each node corresponds to a gene, and a pair of nodes is connected with an edge if there is a significant co-expression relationship between them [207]. BLAST analysis was performed for transcriptome sequences of both JH and JV samples with NCBI non-redundant database. Annotated Jatropha scaffold ID were mapped with sequences and genes were extracted from annotated list on the basis of fold change (threshold 2 fold) using in house perl scripts (Figure 3.6). Constructing gene co-expression network involved combination of Pearson's correlation coefficient (PCC), gene ontology score, abundance score and codon score that was calculated as the geometric mean of the correlation rank of gene X to gene Y and of gene Y to gene X. Cytoscape network construction and network analyser plugin were used for co-expression network analysis.

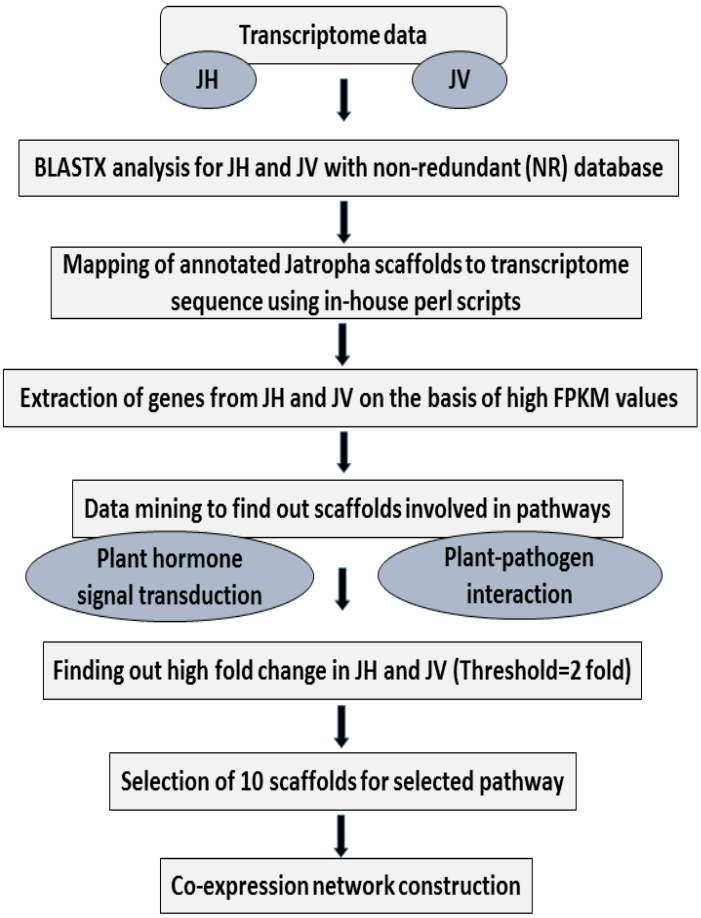


Figure 3.6 Flowchart depicting methodology followed to perform co-expression network analysis (Pre-processing of data and selection of reference genes for co-expression analysis)

3.9.8 RT-qPCR based experimental validation

For experimental validation, RT-qPCR approach was used to confirm the transcriptome data. cDNAs of JH and JV were used for this analysis and the expression profiles were investigated by RT-qPCR. Using gene specific primers (Table 3.7), RT-qPCR was performed on CFX96 system (Bio-Rad, USA) with the iScript one step RT PCR kit (Bio-Rad, USA).The PCR protocol was as: denaturation for 5 min at 94°C, followed by 40 cycles each of denaturation for 20 s at 94°C, annealing for 30 s at 50-55°C, followed by one elongation step for 20 s at 72°C. All quantitative PCR experiments were repeated with three replicates. For calculating transcript abundances, *26S rRNA* and *GAPDH* were used as internal controls. Standard deviation with percentage error was used to statistically evaluate significant differences between treatments.

Table 3.7 Primers for genes involved in 'Plant hormone signal transduction' used for experimental validation through RT-qPCR

S.No.	Gene	Abbreviation	Forward primer	Reverse primer	Annealing temperature (ºC)
1	Jasmonate ZIM domain-containing protein	*PHST_1*	TAGATTCCTCGATCTCATGG	CAGCTTCTCGCCAATACTA	53
2	Auxin responsive GH3 gene family	*PHST_2*	ATATCGCAGGAAGAGTCTGA	ATTGCTATCCACGGAGATG	53
3	Two-component response regulator ARR-A family	*PHST_3*	CGATGTTAAGTGAGGTGGAT	CTCCAAAAAGCTCTTCTGCT	54
4	Auxin-responsive protein IAA	*PHST_4*	ATCCTCTGTTGTGAGGTTTG	GCATCCTAGTGGATACCAGA	54
5	Abscisic acid receptor	*PHST_5*	GAGCTTGTGTTTAGCCACTT	AGATCCAATAACCCATCTCC	50

	PYR/PYL family				
6	Two-component response regulator ARR-B family	*PHST_6*	GCAGAAGAATTAAGGGAGGT	ATCACCATCCCTATCAACG	51
7	SAUR family protein	*PHST_7*	AGACGCTTATTGTGTCCACT	GACACCTAGAGAATGGCAAG	50
8	26S ribosomal RNA*	*26S rRNA*	CACAATGATAGGAAGAGCCGAC	CAAGGGAACGGGCTTGGCAG AATC	58
9	Glyceraldehyde-3-Phosphate dehydrogenase*	*GAPDH*	TTGCCATCAATGACCCCTTCA	CGCCCCACTTGATTTTGGA	56

*Reference genes (Internal controls)

3.10 Identification of Pfam domains/families associated with NBS-LRR genes and transcription factors related to disease resistance

The domains/families associated with NBS region were considered in the study due to the conserved nature of NBS region. Pfam keyword search with 'NBS-LRR' and associated key words (Table 3.8) was performed in the Pfam database. All hits of domains/ families from the keywords were manually checked for their role in plant defense response and included in 'Master list 1' (NBS-LRR). Transcription factors important in disease resistance were retrieved from the literature. It was found that 10 different transcription factor families were involved in disease resistance (Table 3.9). All 10 transcription factor (TF) families were searched in Pfam text search to find out domains associated with each family. All hits of domains/families from the keyword were checked for their role as transcription factors for plant defense response. Only those domains were incorporated in the 'Master list 2 (transcription factors)' which had significant functional role as TF in plant defense response.

Table 3.8 NBS-LRR domains and their respective Pfam Ids

NBS-LRR domain (keywords)	**Pfam Ids**
NB-ARC	PF00931
TIR	PF01582
TIR2	PF13676
RPW8	PF05659
Dirigent	PF03018

Table 3.9 Transcription factors involved in plant disease resistance

Transcription factor family	**Pfam Ids**	**Reference**
MYB	PF00249, PF13921, PF14379, PF13837, PF12776, PF14215	Katiyar et al. [166]
WRKY	PF03106	Dong et al. [178]
ERF-type/AP2-EREBP	PF00847	Ohme-Takagi et al. [179]
CBF	PF02312, PF00808, PF03914	Sakuma et al. [180]

bZIP	PF00170, PF03131, PF07716, PF12498	Alves et al. [165]
SBP/SPL6	PF03110	Padmanabhan et al. [184]
NAC domain/NAM	PF02365, PF14303	Collinge and Boller [182], Singh et al. [183]
TFIIA	PF03153, PF02268, PF02751	Jiang et al. [181]
Homeo-domain	PF00046, PF05920, PF00157, PF13384, PF13565	Luo et al. [185]
Whirly	PF08536	Desveaux et al. [162]

3.10.1 Identification of NBS-LRR genes and defense response associated transcription factors

The NBS-LRR mRNA sequences collected previously from NCBI were mapped on to the transcriptome of Jatropha using BLAST in order to identify all NBS-LRR containing transcripts in the transcriptome other than predicted 92 NBS-LRR genes in Jatropha [13]. The manually adopted Pfam IDs (domains/families) associated with NBS regions and transcription factors were also mapped using Pfam domain/ family search against the transcriptome of Jatropha in order to identify all genes and transcription factors having the domain IDs from the NBS regions and families of transcription factors, respectively (Figure 3.7).

Domain architecture of a protein can be explored through searching the sequence against the Pfam library of HMMs. NBS-LRR genes and the transcription factors were identified according to domain architecture. All transcripts and transcription factors of Jatropha were translated into proteins (using canonical codon table) according to reading frames and then proteins were subjected to Pfam domain/family search to find out presence of domains. Finally, proteins matching with Pfam domains/ families listed in the 'Master list 1' (refer section: Pfam NBS LRR domain identification) were selected as NBS proteins and corresponding transcript as NBS transcript. The proteins showing match to Pfam domains/families listed in the 'Master list 2' (refer section: Pfam NBS LRR domain identification) were selected as transcription factors associated with disease resistance. In-house PERL programs were used to translate transcripts to proteins, Pfam domains

prediction in translated proteins and comparison of predicted domains for their presence in Master list. Finally results were cross checked manually.

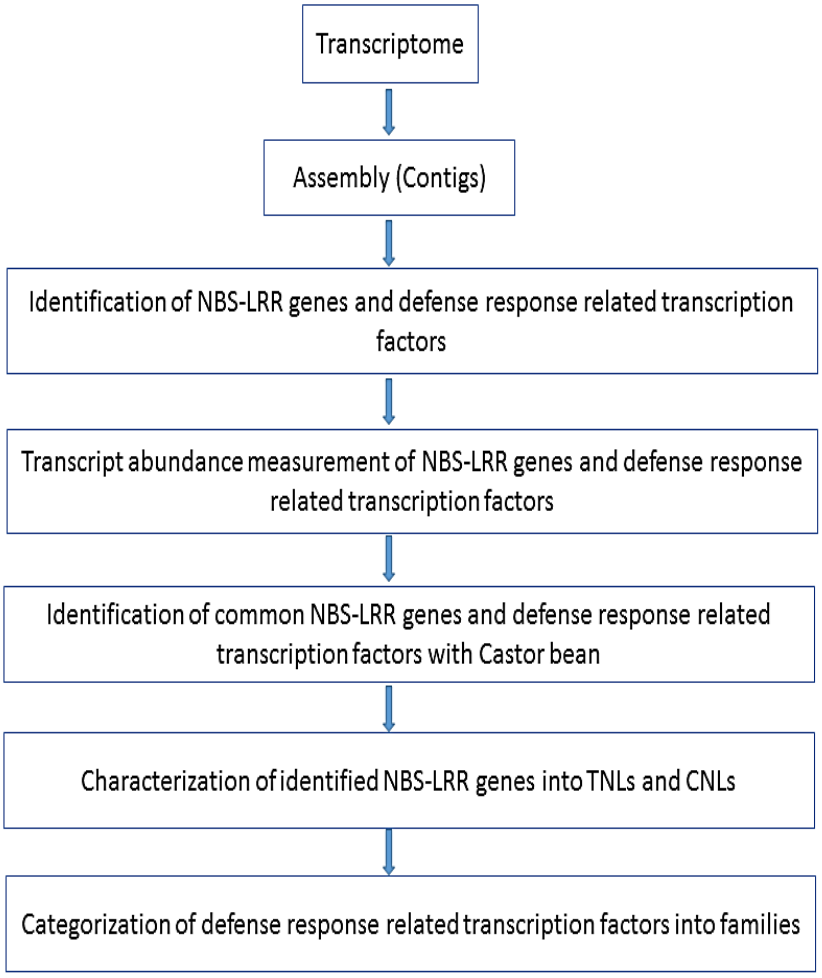


Figure 3.7 Flow diagram depicting methodology for identification of NBS-LRR genes and defense response related transcription factors

3.10.2 Location of NBS-LRR genes in sequence contigs

All predicted NBS-LRR genes were mapped to *J. curcas* genome sequence contigs. BLAST search was used to map contigs on whole genome with exact matching cut off. Position of contigs on genome was extracted from BLAST alignment output file. All analysis, BLAST search and extraction of contigs location were done through in-house developed PERL programs.

3.10.3 Identification of common and unique NBS-LRR genes and transcription factors in Jatropha and castor bean genomes

To identify common and unique NBS-LRR genes and defense response related transcription factors between Jatropha and castor bean, all predicted genes and transcription factors from both the species were used in similarity search. BLASTN was used for finding similarity among contigs of Jatropha and castor bean with cut off values of equal to or more than 70 % within at least a length of 100 nucleotides. In-house PERL program was used to perform BLASTN and to extract results within mentioned cut off, further results were also cross checked manually.

3.10.4 Expression analysis of identified NBS-LRR genes and transcription factors

Transcript abundancy/quantification was carried out using RSEM [195]. RSEM is an user oriented software for quantification of transcript abundances from RNASeq data. RSEM calculates abundance estimates and posterior mean estimates and 95 % credibility intervals for genes/isoforms. There are two measures which specify abundance estimates, one gives an estimate of the fragments number that can be derived from an isoform or gene [the expected counts (EC)], and the other is the probable part of transcripts within the sample represented by the specified isoform or gene. The expression profiles were obtained through pme_TPM (pme: posterior mean estimates; TPM: transcripts per million) values. The TPM value is considered best over other metrics such as FPKM (Fragments per kilobase of transcript per million mapped reads) [208] and RPKM (Reads per kilobase per million) [209] as it is not dependent on the mean expressed transcript length and so more comparable among diverse species and samples [195]. The transcript abundance of the contigs from the transcriptomes of Jatropha and castor bean were calculated using pme_TPM parameter of RSEM package. All the parameters were kept default in the query option.

3.10.5 Identification of CNLs and TNLs in predicted NBS-LRR genes

Using PCOILS (http://toolkit.tuebingen.mpg.de/pcoils), the predicted NBS-LRR genes were further characterized into CNLs and TNLs with default parameters. PCOILS compares a sequence to previously identified parallel two-stranded coiled-coils and determines a similarity score.

3.10.6 Retrieval of disease resistance gene sequences of Jatropha

The gene sequences of Jatropha were used as query against the Jatropha genome database where they were subjected to BLASTN (http://www.kazusa.or.jp/ jatropha/cgi-bin/blast.cgi). This analysis was performed to predict position of these genes in respective sequence contigs. The protein sequences of genes were also subjected to similar analysis using BLASTP.

3.10.7 Protein characterization, motif distribution and domain prediction

By using PCOILS, the predicted disease resistance proteins were characterized into CNLs and TNLs (http://toolkit.tue bingen.mpg.de/pcoils). The distributions of motifs in these proteins were predicted using MAST (http://meme.sdsc.edu/ meme/cgi-bin/mast.cgi). Protein function domains of disease resistance genes were predicted using NCBI conserved domain search (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?INPUT_TYPE=live&SEQUENCE) and the HMM search using Pfam (http://pfam.jouy.inra.fr/hmmsearch.shtml).

CHAPTER 4

RESULTS

The findings of the present study are explained under the following main headings:

4.1 Oil extraction and oil content analysis

4.2 Expression analysis of FA and TAG biosynthetic pathway genes at different developmental stages of embryo and endosperm

4.3 Differential expression pattern in high versus low oil content genotypes vis-a-vis altitude variations

4.4 Relative transcript abundance in endosperm and embryo

4.5 Statistical analysis

4.6 *In–silico* analysis of promoter region

4.7 Cloning of promoter region of *SAD* gene from high and low oil content genotype

4.8 Identification of TFs regulating oil accumulation

4.9 *In-silico* transcript abundance of TFs regulating oil accumulation

4.10 Expression analysis of TFs regulating oil accumulation through RT-qPCR

4.11 Identification of virus

4.12 Reduction in fruits size, seed yield and oil content in response to virus infection

4.13 Transcriptome sequencing and data generation

4.14 Differential gene expression analysis

4.15 Gene ontology analysis based functional classification of JH versus JV transcripts

4.16 Pathway analysis and identification of pathways upregulated in response to viral infection (JV transcriptome)

4.17 Identification of pathways downregulated in response to viral infection

4.18 Identification of SNPs

4.19 Co-expression network analysis

4.20 RT-qPCR based validation of informative transcripts

4.21 Identification of NBS-LRR genes and defense response associated transcription factors in *J. curcas*

4.22 Location of NBS-LRR genes in genome sequence contigs

4.23 Transcript abundance of NBS-LRR genes and transcription factors associated with disease resistance

4.24 Identification of TNLs and CNLs in identified NBS-LRR genes

4.25 Distribution of identified transcription factors into families

4.26 Identification of common and unique NBS-LRR genes and transcription factors between Jatropha and castor bean genomes

4.27 Organization of disease resistance genes in castor bean and Jatropha genome

4.28 Data availability

4.1 Oil extraction and oil content analysis

The oil extraction was done as per protocol described by Kaushik and Bhardwaj [190] with some specific parameters taken into consideration. Genotypes IC 561227 and IC 561235 had 30% and 42% oil content, respectively. IC 561227 was considered as low oil content genotype whereas IC 561235 as high oil content genotype.

4.2 Expression analysis of FA and TAG biosynthetic pathway genes at different developmental stages of embryo and endosperm

Almost all eighteen genes of FA and TAG biosynthesis pathway showed relatively higher expression in high oil content genotype compared to low oil content genotype grown at both the locations (Figures 4.2, 4.3 and 4.4). The relative fold expression of genes in endosperm and embryo was calculated in ripened and mature stage with respect to unripened stage. The FA and TAG biosynthetic pathway genes were further divided into three main clusters i.e. cluster I, cluster II and cluster III (Figure 4.1, Table 4.1). Cluster I comprised of genes involved in the formation of common intermediates in FA synthesis pathway (*ACCase*, *KASI*, *KASII*, *KASIII*, *MT*, *ER*). Cluster II comprised of genes directly linked to the formation of specific fatty acids and their precursors (*SAD*, *OAD*, *FATA*, *PT*, *PAD*, *ST*, *OCD*, *LD*). The genes *DGAT*, *LPAT*, *GPAT* and *PAP* in cluster III are involved in the triacylglycerol biosynthesis. Based on relative expression fold in oil accumulating developmental stages of endosperm or embryo (R and M), the clusters were further divided into sub-clusters. The cluster I was divided into two sub-clusters, cluster Ia comprised genes having fold expression of more than 10 (*ACCase*, *KASI*, *KASII*, *KASIII*) whereas cluster Ib comprised of genes having less than 10 fold value (*MT*, *ER*). Similarly, genes *SAD*, *OAD* and *FATA* were grouped into sub cluster IIa (>10 fold) and *PT*, *PAD*, *ST*, *OCD*, *LD* were present in sub cluster IIb (<10

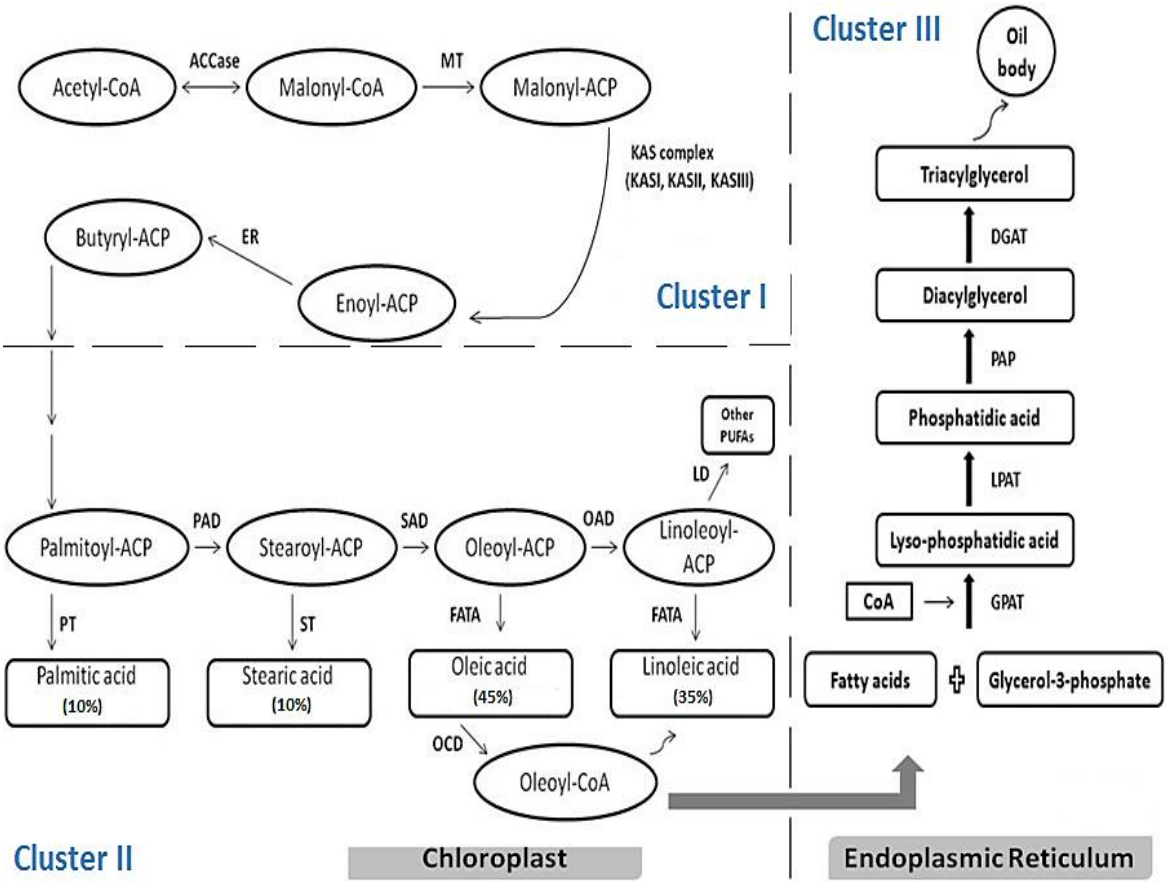


Figure 4.1 Schematic representation of fatty acid (FA) and triacylglycerol (TAG) biosynthesis pathway in Jatropha [16]. *ACCase*- Acetyl-CoA Carboxylase; *DGAT*- Diacylglycerol acyl transferase; *ER*- Enoyl Reductase; *FATA*- Linoleoyl Thioesterase; *GPAT*- Glycerol-3-phosphate acyl transferase; *KASI*- β-ketoacyl-ACP synthase I; *KASII*- β-ketoacyl-ACP synthase II; *KASIII*- β-ketoacyl-ACP synthase III; *LD*- Linoleoyl Desaturase; *LPAT*- Lyso-phosphatidic acid acyltransferase; *MT*- Malonyl Transferase; *OAD*- Oleoyl-ACP Desaturase; *OCD*- Oleoyl-CoA Desaturase; *PAP*- Phosphatidic acid Phosphatase; *PT*- Palmitoyl Thioesterase; *SAD*- Stearoyl-ACP Desaturase; *ST*- Stearoyl Thioesterase; *PAD*- Palmitoyl-ACP Desaturase (Cluster I: Enzymatic steps contributing to formation of common intermediates in FA biosynthesis pathway, Cluster II: Enzymatic steps contributing to formation of specific fatty acids and their direct precursors in FA biosynthesis pathway, Cluster III: Enzymatic steps contributing to formation of triacylglycerols)

fold). Genes *DGAT*, *LPAT* and *GPAT* were present in cluster IIIa (>10 fold) whereas *PAP* was in cluster IIIb (<10 fold). For clustering analysis, the high oil content genotype (SH and NH) selection was considered as all the genes showed low or negligible expression in endosperm or embryo developmental stages from low oil content genotypes (SL and NL). Specifically, *KASII* showed peak expression at mature stages of embryo (42.1 and 46.2 fold) whereas *KASI* and *KASIII* were maximally expressed at mature stages (42.5 and 41.7

fold; 78.9 and 86.1 fold) of endosperms of high oil content genotype. *FATA* showed transcript abundance of about 20 fold in ripened and mature stages of both embryo and endosperm of high oil content genotype. For *LPAT*, the expression was significantly high (55-80 fold) in mature stage of both embryo and endosperm for SH and NH. Interestingly, for *DGAT*, maximum expression of about 268 and 298 fold was only observed in mature stage of endosperm of SH and NH, however for embryo it was relatively low. Key genes linked with fatty acid and triacylglycerol biosynthesis were identified on the basis of expression fold.

Table 4.1 Clustering of FA and TAG biosynthesis pathway genes

Cluster I **(Common intermediates in fatty acid biosynthesis)**		**Cluster II** **(Direct precursors of fatty acid biosynthesis)**		**Cluster III** **(Triacylglycerol biosynthesis)**	
Cluster Ia	**Cluster Ib**	**Cluster IIa**	**Cluster IIb**	**Cluster IIIa**	**Cluster IIIb**
ACCase *KASI* *KASII* *KASIII*	*MT* *ER*	*SAD* *OAD* *FATA*	*PT* *PAD* *ST* *OCD* *LD*	*DGAT* *LPAT* *GPAT*	*PAP*

Table 4.2 Relative expression fold values for FA and TAG biosynthesis pathway genes in high and low oil content genotypes

Gene	Endosperm								Embryo							
	Ripened (R) stage				Mature (M) stage				Ripened (R) stage				Mature (M) stage			
	SH	**NH**	**SL**	**NL**	**SH**	**NH**	**SL**	**NL**	**SH**	**NH**	**SL**	**NL**	**SH**	**NH**	**SL**	**NL**
ACCase	10.49	10.24	1	2.96	20.06	26.85	3	2.3	10.88	10.41	2	2.75	11.09	10.74	3	1.06
MT	4.06	4.96	N	N	4.14	4.08	N	N	5.94	1.2	1.33	1.71	5.3	3.84	1.22	1.89
KASI	12.7	14	N	N	42.5	41.27	N	N	16.7	16.63	N	3.89	27.4	23.8	1.13	1.08
KASII	20.95	23.29	N	N	30.89	29.65	N	N	20.43	19.1	N	N	42.1	46.2	N	N
KASIII	20.58	23.34	1.18	N	78.9	86.17	1.75	3.53	19.3	24.3	1.56	N	78.4	69.41	1.21	N
ER	5.13	4.33	1	N	9.34	9.44	N	1.2	4.28	3.58	1.14	N	8.81	2.13	2.44	1.5
PAD	2.2	4	N	N	4.31	3.98	N	N	1.59	1.6	N	N	4	3.6	N	N
SAD	15.63	16.8	2.1	2.36	25.69	28.1	1.2	2.45	12	11.9	2.11	1.9	18.21	23.11	2.09	2.11
OAD	10.31	10.69	N	N	11.13	10.2	N	N	10.53	10.68	N	N	10.08	10.84	N	1.28
PT	4.46	6.77	N	N	3.31	5.03	N	N	1.45	4.2	N	N	4.88	4.36	N	N
ST	2	3.86	N	N	6.31	3.92	N	N	1.59	1.09	N	N	5.79	4.6	N	N
FATA	19.4	19.02	N	N	19.41	18.92	N	1.3	18.43	19.71	N	1.62	16.81	17.17	1.44	N

LD	8.26	8.56	3.26	N	9.59	9.09	2.68	N	6.28	6.33	3.27	1.97	9.71	8.25	2.69	N
OCD	1	3.21	N	3.59	2.55	3.92	N	3.89	1.92	3.56	N	1.4	2.85	3.48	N	3.94
GPAT	20.4	19.55	N	N	40.3	35.63	1.14	N	25.67	28.53	N	3.05	21.21	20	2.2	1.56
LPAT	30.81	32.22	1.54	2.22	61.3	83.3	2.88	2.94	26.1	28.88	3.68	2.36	58.87	55.6	3.78	3.01
PAP	5.06	6.4	1.16	N	8.47	6.85	N	1.23	6.45	1.43	N	N	7.26	3.58	2.6	1
DGAT	54.94	60.8	1.01	2.45	268.6	298.5	2.66	3.77	10.82	10.38	3.92	1.15	11.96	10.93	3.9	3.11

SH: High oil content genotype (Sunni); NH: High oil content genotype (Nalagarh); SL: Low oil content genotype (Sunni); NL: Low oil content genotype (Nalagarh); N-Negligible fold; Genes in bold letters are showing significant higher expression in oil accumulating developmental stage

4.3 Differential expression pattern in high versus low oil content genotypes vis-a-vis altitude variations

High oil content genotype i.e. IC 561235 and low oil content genotype i.e. IC 561227 from locations Sunni (SH, SL) and Nalagarh (NH, NL) were taken for expression analysis. It was observed that each gene of FA and TAG biosynthetic pathway showed significantly higher transcript abundance in ripened and mature developmental stages of endosperm and embryo in high oil content genotype compared to low oil content genotype (Figures 4.2, 4.3 and 4.4). The relative expression fold increase values for all the genes are described in Table 4.2. Out of total 18 genes, 8 genes (*MT*, *ER*, *PT*, *PAD*, *ST*, *OCD*, *LD*, *PAP*) exhibited low expression even in accumulating stages of high oil content genotype. Genes showing significantly higher transcript abundance (>10 fold) in high oil content genotype were further shortlisted on the basis of showing maximum expression in two varying altitudes (SH and NH). A slight difference for the oil content was observed for high oil content genotype between two varying altitude locations, however, the expression level of some genes differed significantly. Out of 10 genes, *KASI*, *OAD*, *FATA* and *GPAT* showed maximum expression for any oil accumulating developmental stage (R/M of endosperm/embryo) of Sunni (SH, High altitude). Genes *ACCase*, *KASII*, *KASIII*, *SAD*, *LPAT* and *DGAT* exhibited maximum abundance in any oil accumulating developmental stage (R/M of endosperm/embryo) of Nalagarh (NH, Low altitude).

4.4 Relative transcript abundance in endosperm and embryo

From the fold increase values of gene expression level, the relative transcript abundance of genes for endosperm to embryo was calculated. Inconsistency in the expression fold values of endosperm to embryo was observed in low oil content genotype at both the altitude locations (Figure 4.5). In case of high oil content genotype, the endosperm to embryo fold value was almost similar at both the locations i.e. SH and NH. For example in case of *ACCase* gene, the endosperm to embryo fold increase was 3.15, 1.84 and 2.1 in U, R, M stages respectively of SH. Similarly, fold increase of 2.14, 1.07, 1.13 was observed in U, R, M stages at NH location. On contrary, for low oil content genotype, 3.25, 1.67, 2.79 fold in U, R, M stages was observed at SL and 0.76, 0.81, 1.64 fold in U, R, M stages, respectively for NL. Genes *FATA*, *GPAT*, *OAD* and *PAP* showed maximum fold increase in the transcript abundance level in either or both the oil accumulating developmental stages, R and M. For low oil content genotype (SL and NL), there was negligible or zero fold increase in the transcript abundance of most of the FA and TAG

pathway genes in endosperm to embryo of oil accumulating and non-accumulating developmental stages.

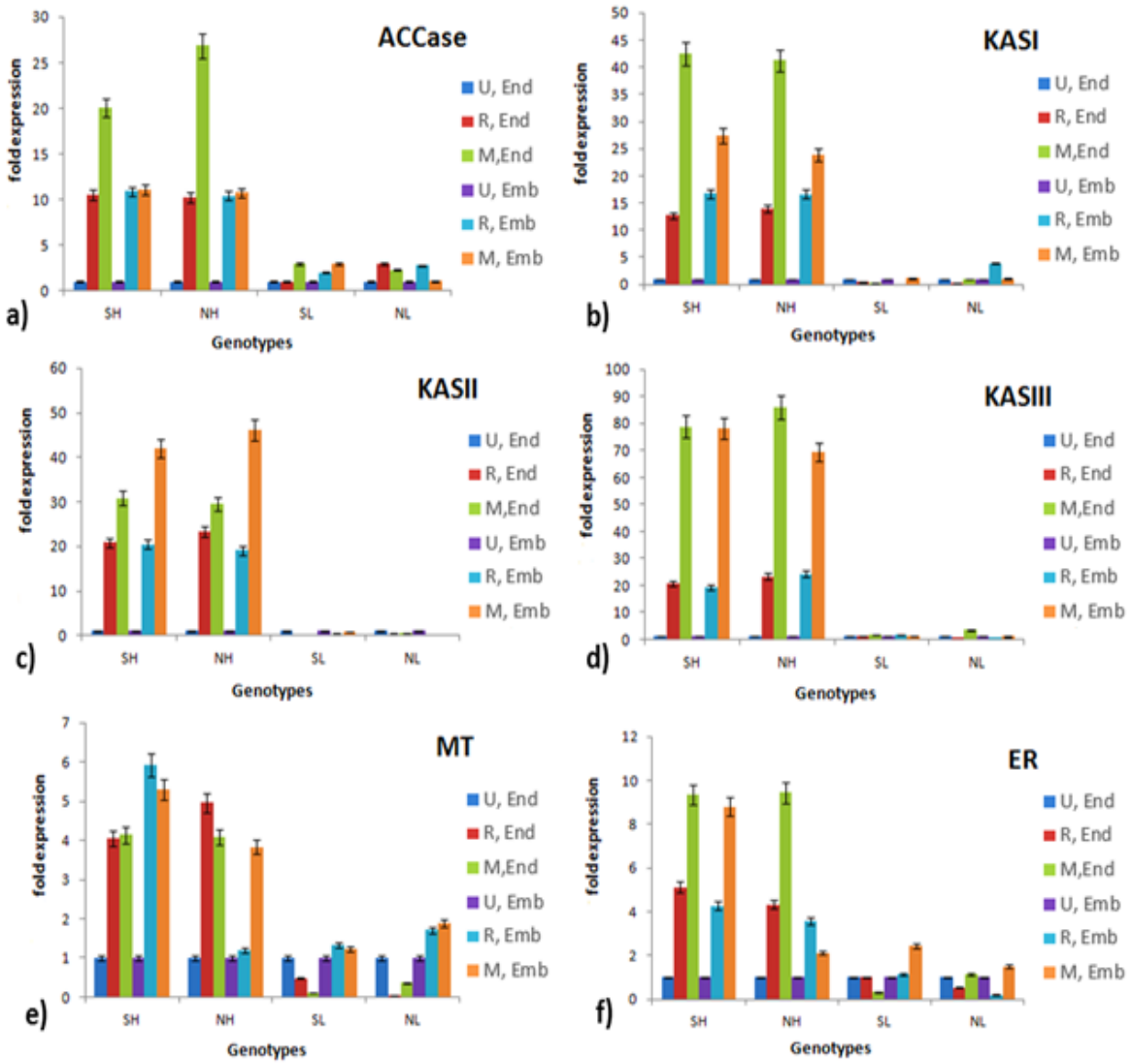


Figure 4.2 Relative expression pattern of Cluster I FA and TAG biosynthetic pathway genes in different developmental stages of endosperm and embryo. (a) *ACCase*- Acetyl-CoA Carboxylase (b) *KASI*- β-ketoacyl-ACP synthase I (c) *KASII*- β-ketoacyl-ACP synthase II (d) *KASIII*- β-ketoacyl-ACP synthase III (e) *MT*- Malonyl Transferase (f) *ER*- Enoyl Reductase. (SH- Sunni high oil content genotype, NH- Nalagarh high oil content genotype, SL- Sunni low oil content genotype, NL- Nalagarh low oil content genotype, U- Unripened, R- Ripened, M- Mature, End- Endosperm, Emb- Embryo)

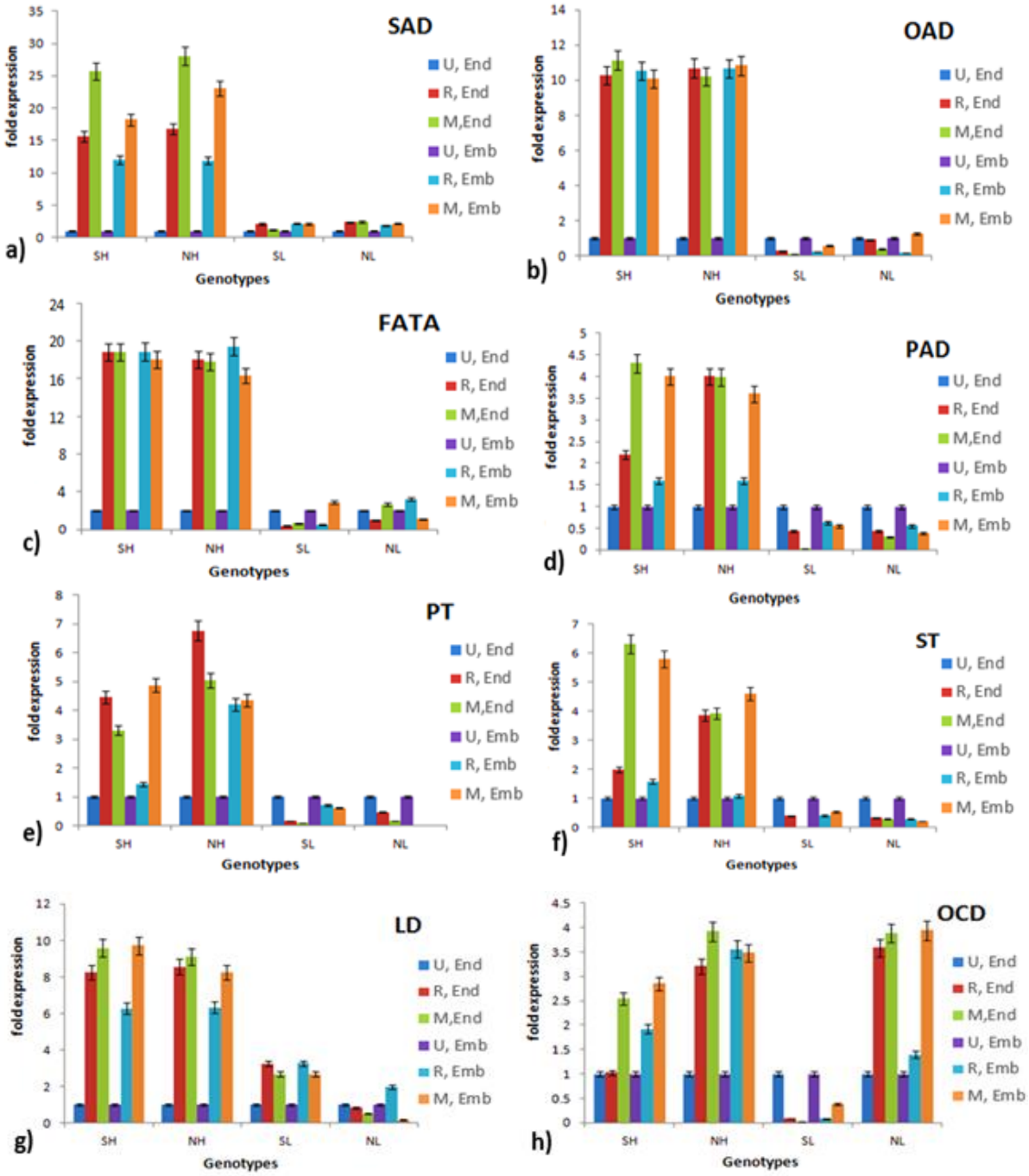


Figure 4.3 Relative expression pattern of Cluster II FA and TAG biosynthetic pathway genes in different developmental stages of endosperm and embryo. (a) *SAD*-Stearoyl-ACP Desaturase (b) *OAD*- Oleoyl-ACP Desaturase (c) *FATA*-Linoleoyl Thioesterase (d) *PAD*-Palmitoyl-ACP Desaturase (e) *PT*-Palmitoyl Thioesterase (f) *ST*-Stearoyl Thioesterase (g) *LD*-Linoleoyl Desaturase (h) *OCD*-Oleoyl-CoA Desaturase. (SH- Sunni high oil content genotype, NH- Nalagarh high oil content genotype, SL- Sunni low oil content genotype, NL- Nalagarh low oil content genotype, U- Unripened, R- Ripened, M- Mature, End- Endosperm, Emb- Embryo)

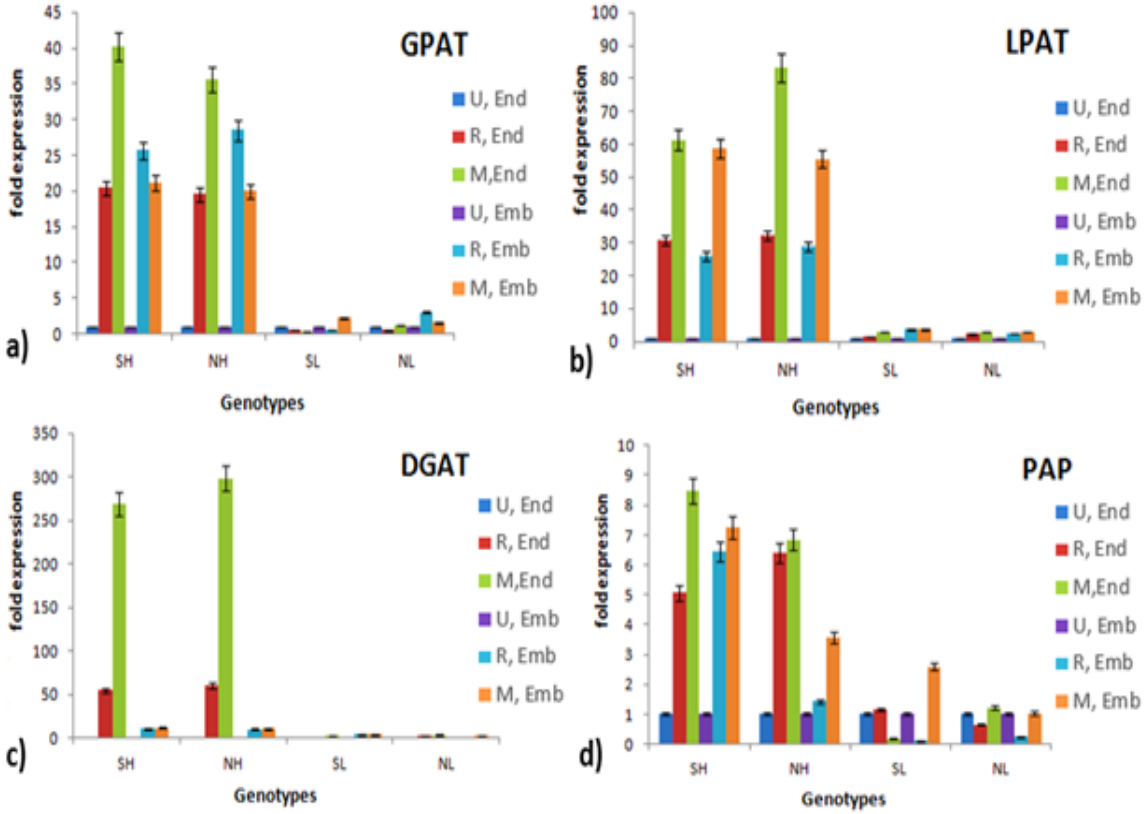
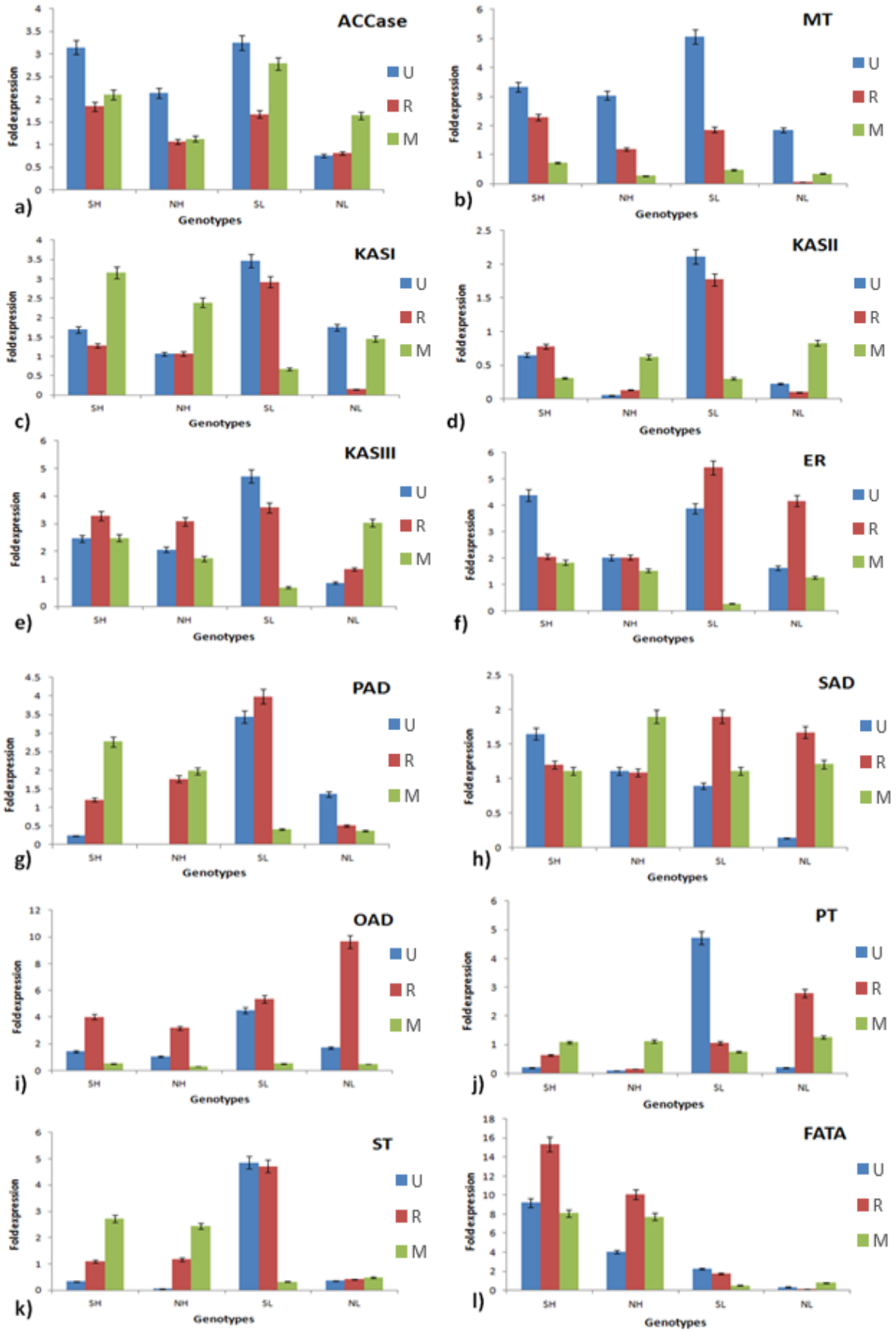


Figure 4.4 Relative expression pattern of Cluster III FA and TAG biosynthetic pathway genes in different developmental stages of endosperm and embryo. (a) *GPAT*- Glycerol-3-phosphate acyl transferase (b) *LPAT*- Lyso-phosphatidic acid acyltransferase (c) *DGAT*- Diacylglycerol acyl transferase (d) *PAP*- Phosphatidic acid phosphatase. (SH- Sunni high oil content genotype, NH- Nalagarh high oil content genotype, SL- Sunni low oil content genotype, NL- Nalagarh low oil content genotype, U- Unripened, R- Ripened, M- Mature, End- Endosperm, Emb- Embryo)


ACCase
a)
MT
b)
KASI
c)
KASII
d)
KASIII
e)
ER
f)
PAD
g)
SAD
h)
OAD
i)
PT
j)
ST
k)
FATA
l)
Foldexpression
Genotypes
SH
NH
SL
NL
U
R
M

Cont'd

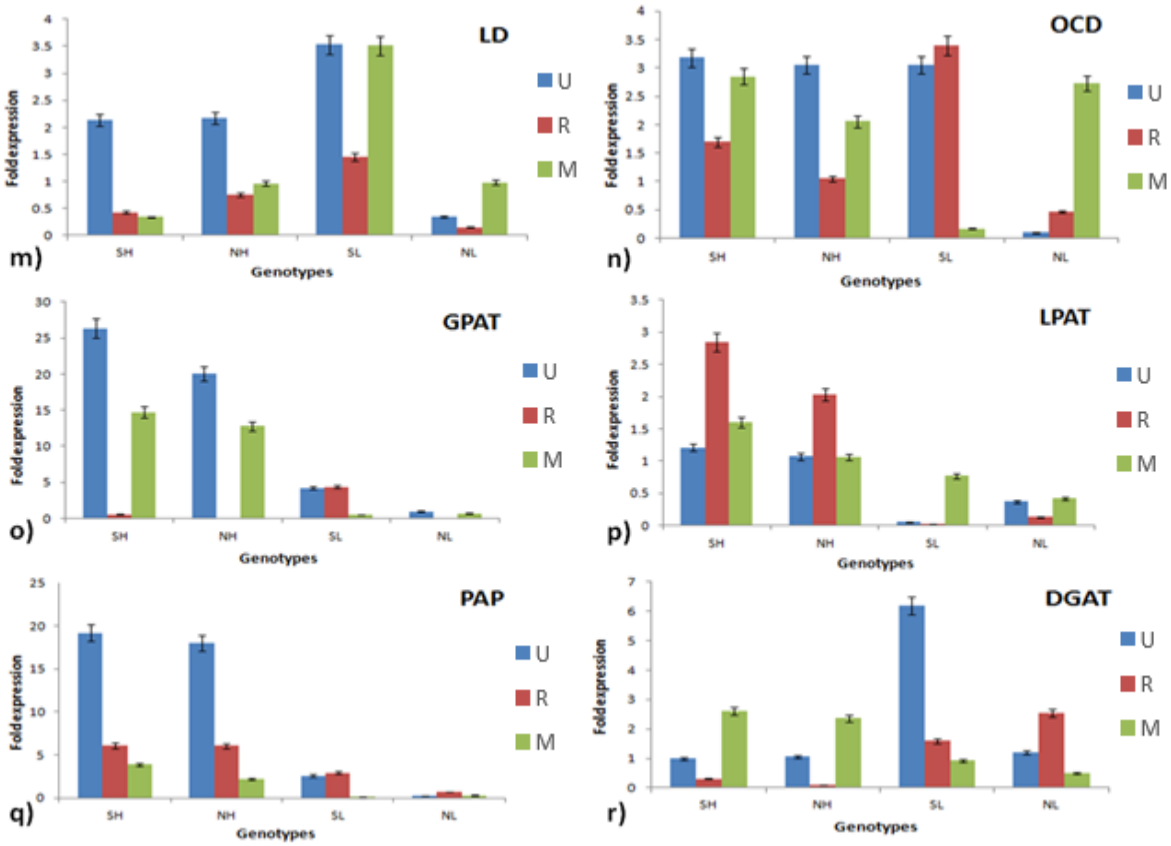


Figure 4.5 Expression pattern of FA and TAG biosynthetic pathway genes for endosperm to embryo ratio in different developmental stages (a) *ACCase*-Acetyl-CoA Carboxylase (b) *MT*-Malonyl Transferase (c) *KASI*-β-ketoacyl-ACP synthase I (d) *KASII*-β-ketoacyl-ACP synthase II (e) *KASIII*-β-ketoacyl-ACP synthase III (f) *ER*-Enoyl Reductase (g) *PAD*-Palmitoyl-ACP Desaturase (h) *SAD*-Stearoyl-ACP Desaturase (i) *OAD*-Oleoyl-ACP Desaturase (j) *PT*-Palmitoyl Thioestrase (k) *ST*-Stearoyl Thioestrase (l) *FATA*-Linoleoyl Thioestrase (m) *LD*-Linoleoyl Desaturase (n) *OCD*-Oleoyl-CoA Desaturase (o) *GPAT*-Glycerol-3-phosphate acyl transferase (p) *LPAT*-Lyso-phosphatidic acid acyltransferase (q) *PAP*-Phosphatidic acid phosphatase (r) *DGAT*-Diacylglycerol acyl transferase. (SH-Sunni high oil content genotype, NH- Nalagarh high oil content genotype, SL- Sunni low oil content genotype, NL- Nalagarh low oil content genotype, U- Unripened, R- Ripened, M- Mature)

4.5 Statistical analysis

PCA analysis showed the correlation of 18 FA and TAG biosynthesis pathway genes in oil accumulating stages (R, M) of endosperm and embryo exhibiting higher expression in high oil content genotypes. Based on the squared cosine values, genes *ACCase*, *KASI*, *KASII*, *KASIII*, *SAD*, *OAD*, *FATA*, *DGAT*, *LPAT* and *GPAT* were observed to be associated

with oil accumulating developmental stages of both embryo and endosperm (Figure 4.6). A representative heat map showing the expression pattern of genes of FA and TAG biosynthetic pathway from different oil accumulating developmental stages of endosperm and embryo was also generated using GenEx (Figure 4.7).

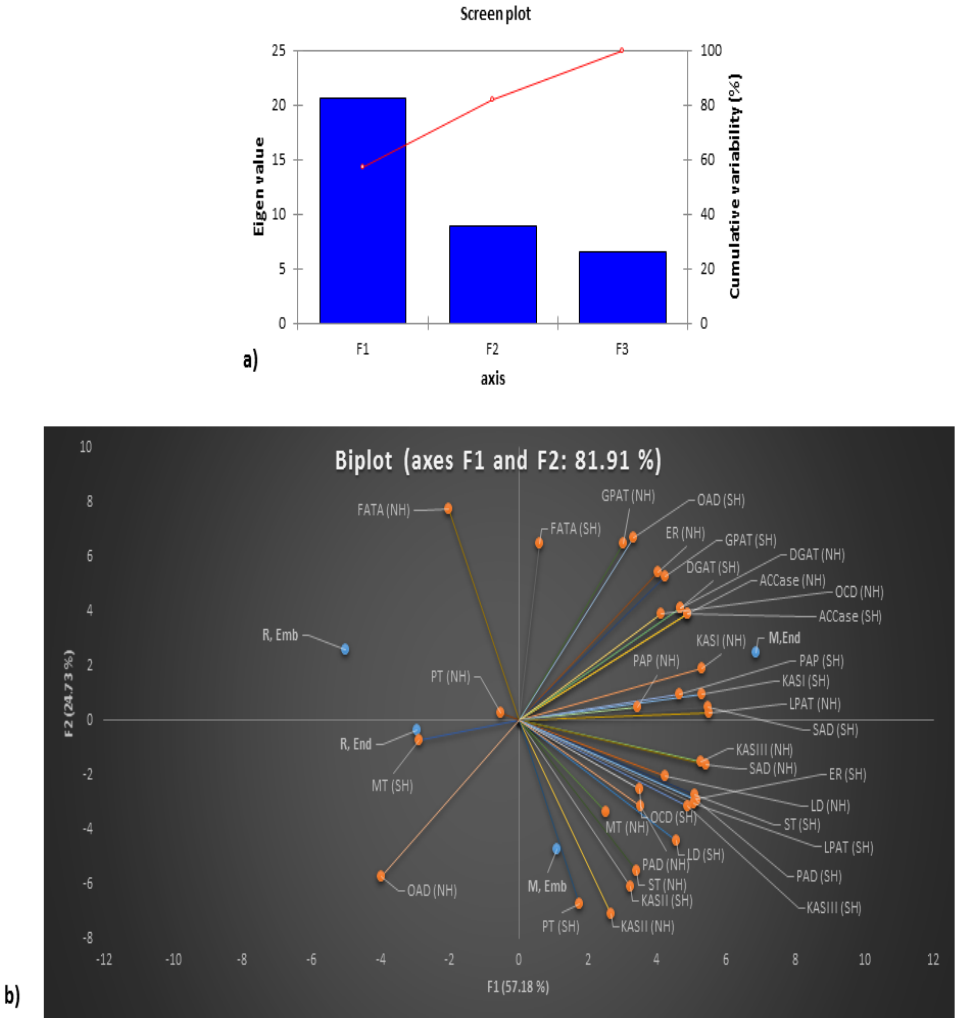


Figure 4.6 Principal Component Analysis (PCA) of FA and TAG biosynthesis pathway genes at developmental stages of high oil content genotype embryo and endosperm **(a)** Screen plot for principal components (F1-F3), eigen values, cumulative variability. Major variance was contributed by component F1 **(b)** Biplot for PCA (R, End-Ripened stage of endosperm; M, End-Mature stage of endosperm; R, Emb-Ripened stage of embryo; M, Emb-Mature stage of embryo)

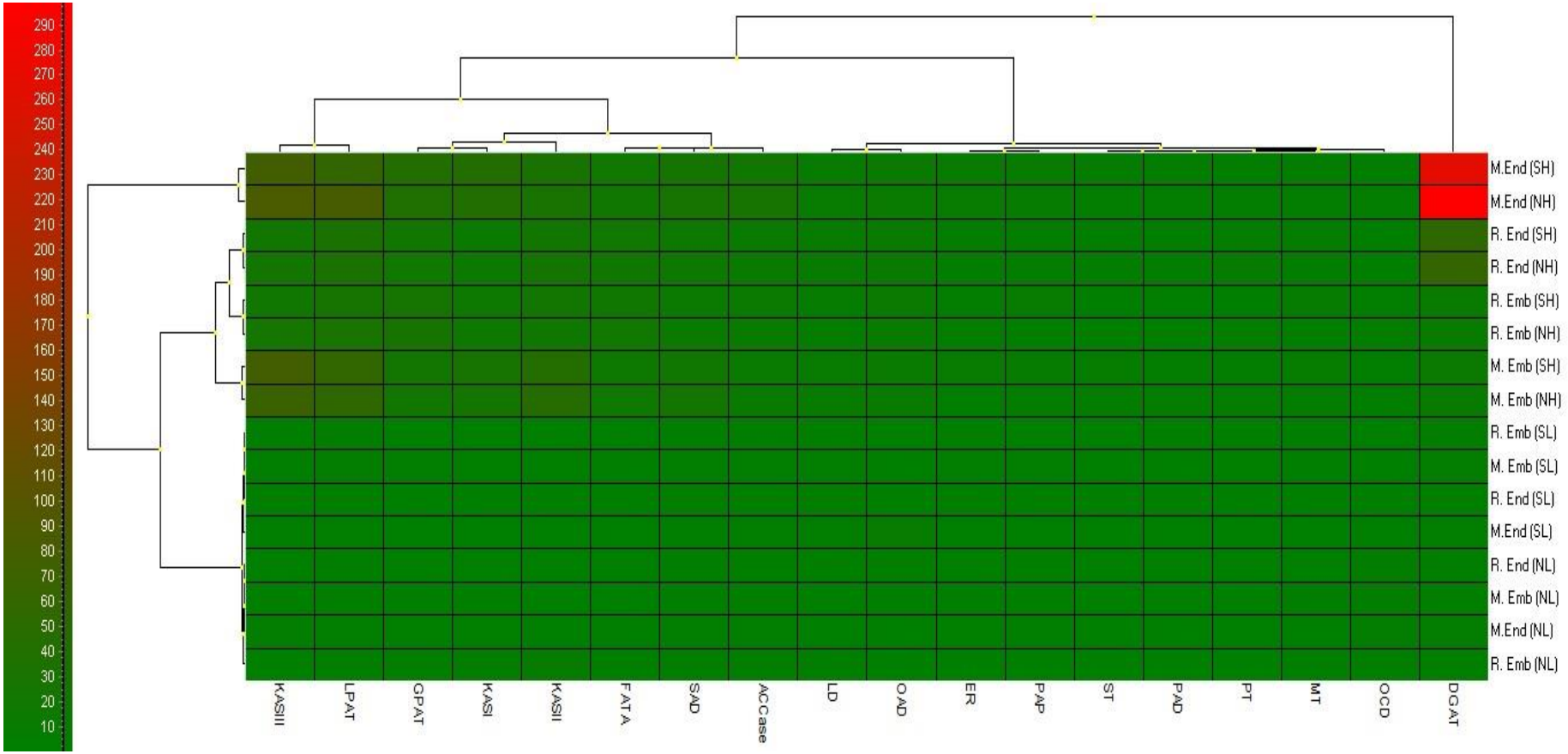


Figure 4.7 A representative heat map demonstrating differential expression pattern of genes of FA and TAG biosynthetic pathway from oil accumulating developmental stages of endosperm and embryo

4.6 *In–silico* analysis of promoter region

Total ten genes from clusters Ia, IIa and IIIa (*ACCase*, *KASI*, *KASII*, *KASIII*, *OAD*, *SAD*, *FATA*, *GPAT*, *LPAT*, *DGAT*) exhibited elevated levels of transcript abundance in oil accumulating developmental stages in the current study. Along with these genes, the genes showing developmental regulation with oil accumulation from previous studies by Xu et al. [33] and Gu et al. [11] were also chosen for *in-silico* promoter analysis. Nine genes (*DGAT*, *ACCase*, *KASI*, *KASII*, *KASIII*, *LPAT*, *Oleosin1*, *Oleosin2*, *PDAT*) from Xu et al. [33] and ten genes (*SD*, *KCR2*, *ECH*, *Oleosin1*, *LACS8*, *DGK1*, *LPAT*, *Lipase*, *KCS*, *CLK*) from Gu et al. [11] were selected along with ten genes from current study to have a set of total twenty nine genes. Out of 29 genes, 8 genes were common in either two or all three studies (Tables 4.3, 4.4). Upstream regions (1.5-2 kb) of 29 genes were analyzed for promoter elements through Plant Care and PLACE which showed the presence of oil deposition specific regulatory elements like Dof, CBF (LEC1), SORLIP, Skn-1_motif, GATA along with the common elements like TATA box, CAAT box etc (Tables 4.5, 4.6). Genes were further categorized on the basis of presence of these elements. Category I comprised nine genes (*KASI*, *KASII*, *FATA*, *LPAT*, *DGAT*, *CLK*, *KCR2*, *Lipase*, *OAD*) which were having all five specific elements in common. Genes from category II i.e. nine genes (*Oleosin1*, *Oleosin2*, *PDAT*, *DGK1*, *ECH*, *KCS*, *LACS8*, *SD* and *SAD*) showed the presence of four common elements whereas there were only three genes in category III (*ACCase*, *KASIII*, *GPAT*) on the basis of presence of 3 common elements in the promoter region. This analysis suggested common regulation of genes linked to oil biosynthesis and provided potential elements/transcription factors for oil enhancement in Jatropha. On correlation with previous clusters it was found genes showing higher expression (>10 fold) exhibited the presence of all five, four or three oil deposition specific elements. The promoter regions of genes showing more than 10 fold expression i.e. *KAS I* and *KAS II* from sub-cluster Ia (fatty acid biosynthesis), *OAD* and *FATA* from sub-cluster IIa (fatty acid biosynthesis), *LPAT* and *DGAT* from sub-cluster IIIa (triacylglycerol biosynthesis) exhibited the presence of all 5 sequence elements important to oil deposition. *SAD* gene showing more than 10 fold expression showed from sub-cluster IIa had four elements in the promoter region whereas *ACCase* and *KAS III* from sub-cluster Ia and *GPAT* from sub-cluster IIIa showed presence of three elements in the promoter regions. These predicted elements were associated to oil accumulating developmental stages (R/M) as they were found in the promoter region of the genes showing higher expression in R and M stages.

Table 4.3 Genes with elevated transcript abundance vis-a-vis seed development and oil biosynthesis

Genes	**Reference**
DGAT, ACCase, KASI, KASII, KASIII, LPAT, Oleosin 1, Oleosin 2, PDAT	Xu et al. [33]
SD, KCR2, ECH, Oleosin1, LACS8, DGK1, LPAT, lipase, KCS, CLK	Gu et al. [11]
ACCase, KASI, KASII, KASIII, OAD, SAD, FATA, GPAT, LPAT, DGAT	Current study
Common genes with elevated transcript abundance vis-a-vis seed development and oil biosynthesis	
LPAT and *Oleosin 1*	Xu et al. [33] and Gu et al. [11]
ACCase, DGAT, KASI, KASII, KASIII, LPAT	Xu et al. [33] and current study
LPAT	Gu et al. [11] and current study
LPAT	Xu et al. [33], Gu et al. [11] and current study

Table 4.4 Unique genes from Xu et al. [33] and Gu et al. [11] showing elevated transcript abundance vis-a-vis seed development and oil biosynthesis

Gene	**Function in lipid biosynthesis**	**Reference**
Oleosin 1	Oil bodies regulation	Siloto et al. [196]
Oleosin 2	Oil bodies regulation	Siloto et al. [196]
Phospholipid diacylglycerol acyltransferase (*PDAT*)	Acyl-CoA-independent triacylglycerol biosynthesis	Stahl et al. [197]
Choline kinase (*CLK*)	phosphatidylcholine (Phospholipids) formation	Gibellini and Smith [198]
Diacylglycerol kinase 1 (*DGK1*)	Biosynthesis of phosphatidic acid from diacylglycerol	Han et al. [199]
Enoyl-CoA hydratase	Degradation of saturated fatty acids	Allenbach and Poirier [200]

(*ECH*)		
3-ketoacyl-CoA reductase isoform 2 (*KCR2*)	Elongation of fatty acids	Puyaubert et al. [201]
Ketoacyl-CoA synthase (*KCS*)	Formation of very long chains of mono unsaturated fatty acids	Taylor et al. [202]
Long-chain acyl-CoA synthetase 8 (*LACS8*)	Esterification of free fatty acids to form acyl-CoAs	Shockey et al. [203]
Lipase	Hydrolysis of triacylglycerol to form free fatty acids and glycerol	Quettier and Eastmond [204]
Sterol desaturase (*SD*)	Desaturation of fatty acids	Taton and Rahier [205]

Table 4.5 Regulatory elements in the promoter regions of genes

Gene	**Regulatory elements**				
	Dof	**CBF (LEC1)**	**SORLIP**	**GATA**	**Skn-1_motif**
ACCase	+	-	-	+	+
DGAT	+	+	+	+	+
KASI	+	+	+	+	+
KASII	+	+	+	+	+
KASIII	+	-	-	+	+
LPAT	+	+	+	+	+
Oleosin 1	+	-	+	+	+
Oleosin 2	+	-	+	+	+
PDAT	+	-	+	+	+
CLK	+	+	+	+	+
DGK1	+	+	-	+	+
ECH	+	-	+	+	+
KCR2	+	+	+	+	+

KCS	+	-	+	+	+
LACS8	+	-	+	+	+
Lipase	+	+	+	+	+
SD	+	+	+	+	-
GPAT	+	-	-	+	+
SAD	+	-	+	+	+
OAD	+	+	+	+	+
FATA	+	+	+	+	+

Table 4.6 Oil deposition specific regulatory elements

Regulatory element	**Conserved sequence**	**Function**	**Plant spp.**	**Reference**
Dof	AAAG	Oil content enhancement	Soybean	Zhang et al. [51]
CBF (LEC1)	RYCGAC	Fatty acid composition and oil content increase	Maize	Shen et al. [66], Tan et al. [65]
SORLIP	GCCAC	Associated with fatty acid biosynthesis genes and seed storage reserve accumulation	Arabidopsis	Peng and Weselake [210]
GATA	GATA	Associated with fatty acid biosynthesis and accumulation	Arabidopsis	Fobert [74]
Skn-1_motif	GTCAT	cis-acting regulatory element required for endosperm expression	Soybean	Takaiwa et al. [211]

4.7 Cloning of promoter region of *SAD* gene from high and low oil content genotype

The promoter region of *SAD* gene was cloned in high as well as low oil content genotype. *SAD* was considered for cloning the promoter region as it is the rate limiting gene of FA and TAG biosynthesis. A band of 1.5 kb was observed in both high as well as low oil content genotypes. The band was eluted from the gel and further sequenced. Upon computational identification with Plant Care and PLACE, it was observed that promoter region sequence of high oil content genotype showed presence of more number of oil deposition specific elements like SORLIP (8), Dof (6), GATA (9) along with basic promoter elements like TATA box, CAAT box etc. whereas, less number of oil deposition specific element for the promoter region sequence in low oil content genotype was found (Table 4.7), i.e. SORLIP (2), Dof (1) and GATA (4). The experimental validation through cloning of the promoter region of *SAD* gene confirmed the role of specific regulatory elements in oil accumulation as they were more in the promoter region of high oil content genotype as compared to low oil content genotype.

4.8 Identification of TFs regulating oil accumulation

PlantTFcat was used for the identification of TFs regulating oil accumulation. BLASTX was performed for transcriptome data (developing seeds with 45 DAP) against whole TF families database in PlantTFcat. A total of 102 TF families were identified. Out of 9 families from 'Master list', TF families *Dof*, *MYB*, *bZIP*, *bHLH*, *CBF* and *AP2* were identified on the basis of significant FPKM values. The multiple sequence alignment of TF encoding transcripts was performed using clustalW and corresponding conserved domains were identified through conserved domain database available at NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Out of identified TF families through mapping of transcriptome to PlantTFcat, the TF families associated to oil accumulation through literature mining were surveyed and it was found all TF families i.e. *Dof*, *MYB*, *bZIP*, *bHLH*, *CBF*, *AP2*, *B3 domain*, *GATA* and *HD-ZIP* were present. *In-silico* transcript abundance of identified TFs revealed that expression of *Dof*, *MYB*, *bZIP*, *AP2*, *CBF* and *bHLH* was significantly higher.

Table 4.7 Regulatory elements in *SAD* promoter region of high versus low oil content genotypes

Gene	Promoter region			
	Length		Regulatory elements	
	High oil content genotype (IC561235)	**Low oil content genotype (IC561227)**	**High oil content genotype (IC561235)**	**Low oil content genotype (IC561227)**
Stearoyl ACP desaturase (*SAD*)	1.5 Kb	1.5 Kb	TATA Box (TATAA)	TATA Box (TATAA)
			CAAT Box (CCAAT)	CAAT Box (CCAAT)
			GATA Box (GATA) (9 number)	GATA Box (GATA) (4 number)
			SORLIP5 (GAGTGA) (8 number)	SORLIP5 (GAGTGA) (2 number)
			Dof (AAAG) (6 number)	Dof (AAAG) (1 number)

4.9 *In-silico* transcript abundance of TFs regulating oil accumulation

The transcript abundance for TFs was checked in transcriptome by RSEM through *in-silico* approach. The transcripts showing higher FPKM value were selected from each family. Higher FPKM value of a transcript encoding identified TF families involved in oil accumulation ranged from 12.6 to 77.6. Transcript encoding *Dof* showed highest transcript abundance of 77.6 followed by transcripts coding for *MYB* and *bZIP* with transcript abundance of 65 and 48.5, respectively. The FPKM based transcript abundance values were 40, 26 and 12.6 for transcripts coding for *AP2*, *CBF* and *bHLH* TF families, respectively. Whereas low abundance of 2.6, 3.2 and 1.9 was observed for families *B3 domain*, *GATA* and *HD-ZIP*, respectively [Figure 4.8].

4.10 Expression analysis of TFs regulating oil accumulation through RT-qPCR

The expression status of identified TFs in oil accumulating developmental stages of endosperm of high and low oil content genotypes was checked by RT-qPCR to validate their potential role in oil accumulation. Ripened (R) and mature (M) stages of endosperm were selected as these both have shown correlation with more oil accumulation, as identified previously. It was observed that 6

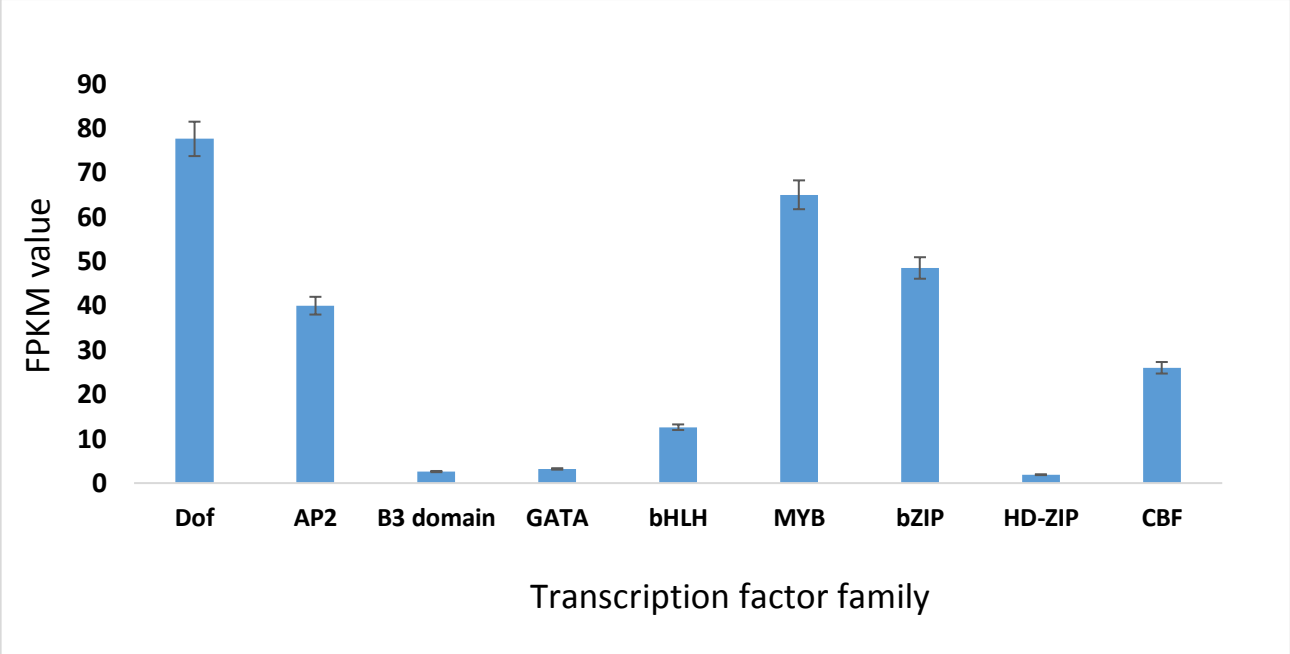


Figure 4.8 *In-silico* transcript abundance (FPKM value) of TFs regulating oil accumulation

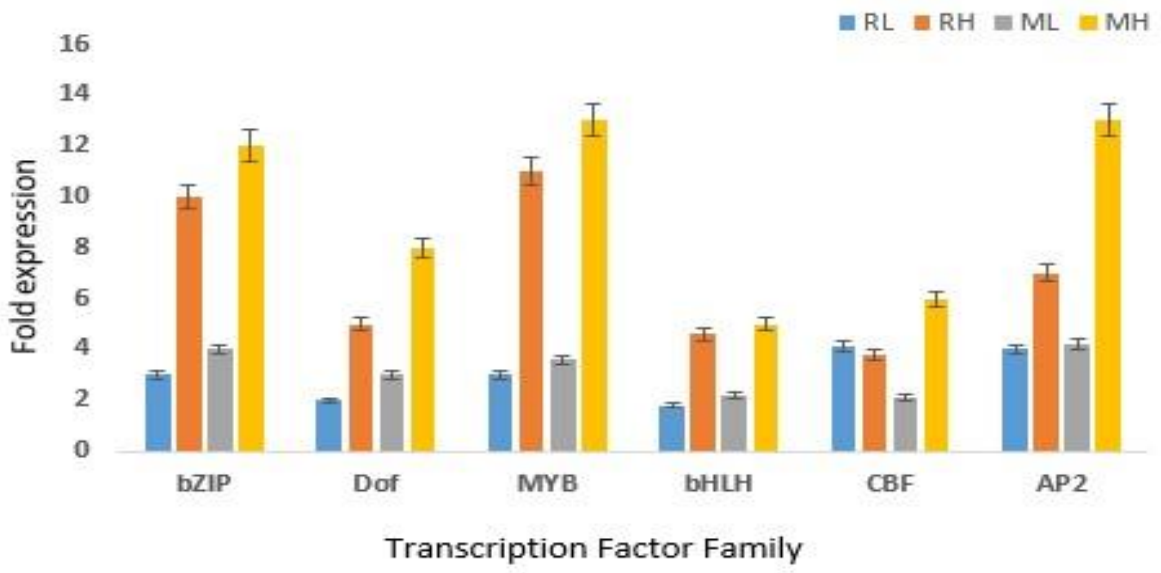


Figure 4.9 Expression pattern of TFs regulating oil accumulation in developmental stages of endosperm of high and low oil content genotypes (RL- Ripened stage of low oil content genotype; RH- Ripened stage of high oil content genotype; ML- Mature stage of low oil content genotype; MH- Mature stage of high oil content genotype)

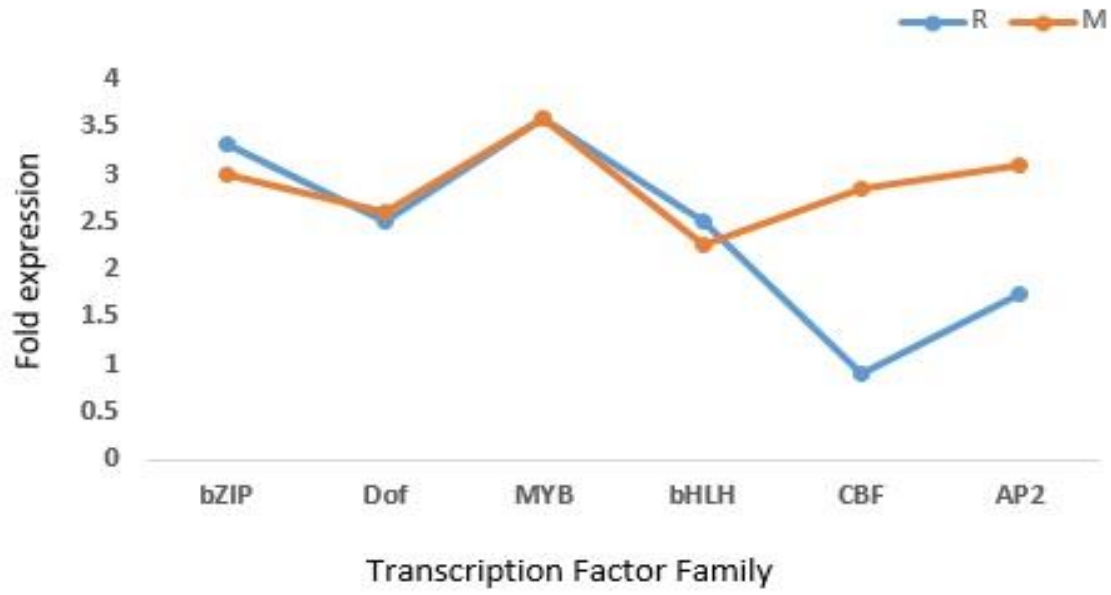


Figure 4.10 Fold expression pattern of TFs regulating oil accumulation in high oil content genotype as compared to low oil content genotype

TFs (*bZIP*, *Dof*, *MYB*, *bHLH*, *CBF* and *AP2*) showed higher expression in high oil content genotype as compared to low oil content genotype in both ripened (R) and mature (M) stages of endosperm, except for *CBF* where expression in R stage of high oil content genotype was slightly less as compared to low oil content genotype (Figure 4.9). On comparing expression of high oil content genotype to low oil content genotype, significant fold increase in both ripened and mature stages was observed. In case of R stage, fold increase of 3.3, 2.5, 3.6, 2.5 and 1.75 was observed for *bZIP*, *Dof*, *MYB*, *bHLH* and *AP2* respectively. For M stage, fold increase of 3, 2.6, 3.6, 2.27, 2.85 and 3.09 was observed for *bZIP*, *Dof*, *MYB*, *bHLH*, *CBF* and *AP2*, respectively. Overall higher expression in M stage as compared to R stage was also observed in both high and low oil content genotypes (Figure 4.10). RT-qPCR based transcript abundance pattern of TFs (*bZIP*, *Dof*, *MYB*, *bHLH*, *CBF* and *AP2*) showed positive correlation with oil accumulating R and M stages of endosperm in high oil content genotype which suggested their role in regulating and controlling the biosynthesis and accumulation in *J. curcas*.

4.11 Identification of virus

The leaves of virus infected plants (JV) showing symptomatic conditions like mosaic, blistering and mottling were observed in the experimental farm. Leaves were also characterized by reduced size, chlorotic spots and rolling [19, 21] (Figure 3.2). Reduction

in fruit number and size in plants infected with mosaic virus was also observed as compared to healthy plants. Upon PCR amplification with mosaic virus coat protein gene, an amplicon of size 700 bp was clearly observed in JV sample as compared to JH which confirmed the presence of virus in the infected leaves (Figure 4.11).

4.12 Reduction in fruits size, seed yield and oil content in response to virus infection

The data on parameters like fruit size, seed yield and oil content was recorded consecutively for two years for healthy and virus infected plants of high oil content genotype IC 561235 at experimental farm of Himalayan Forest Research Institute, Jwalaji (Himachal Pradesh). It was observed that the plants infected with mosaic virus (JV) had reduced fruit size as compared to healthy plants (JH). The mature fruits showed the presence of 1-2 seeds in plants with virus infection as compared to healthy plants having 3 seeds on average. Further, the overall number and weight of seeds was less in virus infected plants compared to healthy plants (JH) (Table 4.8).

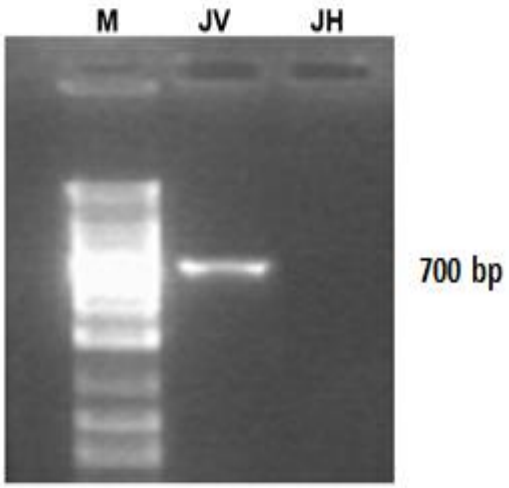


Figure 4.11 PCR based confirmation of *Jatropha curcas* mosaic virus (M: DNA ladder; JV: Virus infected leaf tissue; JH: Healthy leaf tissue)

Table 4.8 Effect of mosaic virus infection on Jatropha yield and oil content

Yield parameters	**Healthy plant (JH)**	**Virus infected plant (JV)**
Seeds per fruit	3	1-2
Number of seeds per plant	620	280
Weight of seeds per plant	400 g	190 g
Oil content	42%	37%

4.13 Transcriptome sequencing and data generation

The cDNA libraries of JH and JV leaves were sequenced on NextSeq 500 platform of Illumina using 2x150 PE Chemistry. The raw reads were filtered using Trimmomatic v0.30 to filter out the adaptor contamination and low quality (reads with QV<20) reads. The high quality reads (QV>20) were used for the RNA-Seq analysis. A total of 10,548,434 and 12,226,847 high quality reads were obtained in JH and JV, respectively (Table 4.9). High quality reads were mapped on to the reference genome using mapping software BWA with optimized parameters.

Table 4.9 Statistics of generated reads

Sample	Number of high quality reads	Number of reads mapped
JH	10,548,434	20,950,364
JV	12,226,847	19,958,750

4.14 Differential gene expression analysis

The expression analysis of genes was carried out with R package DESeq and was calculated in terms of FPKM (Fragment per kilobase per million mapped reads). Upon annotation, it was observed that a total of 55,755 transcripts associated to multiple pathways were expressed in both the samples i.e. JH and JV on the basis of FPKM values. 619 and 330 transcripts were expressed uniquely in JH and JV, respectively. Whereas, 685 and 2132 transcripts were upregulated and down regulated respectively in JV (Figure 4.12). Heat map of top 100 differentially expressed genes was also generated for JH and JV (Figure 4.13).

4.15 Gene ontology analysis based functional classification of JH versus JV transcripts

Gene ontology (GO) is a standardized gene functional cataloging system whose terms are derived from ontologies that can be used to describe the functions of genes and their products in any organism. Three ontologies i.e. molecular function, cellular component and biological process are linked to GO database which are the mainstay of any GO annotation. The predicted CDSs in response to virus infection (JV) and healthy leaf (JH) were annotated by BLAST2GO to characterize transcripts into functional classification. The analysis revealed that majority of transcripts were assigned to ‘biological process’

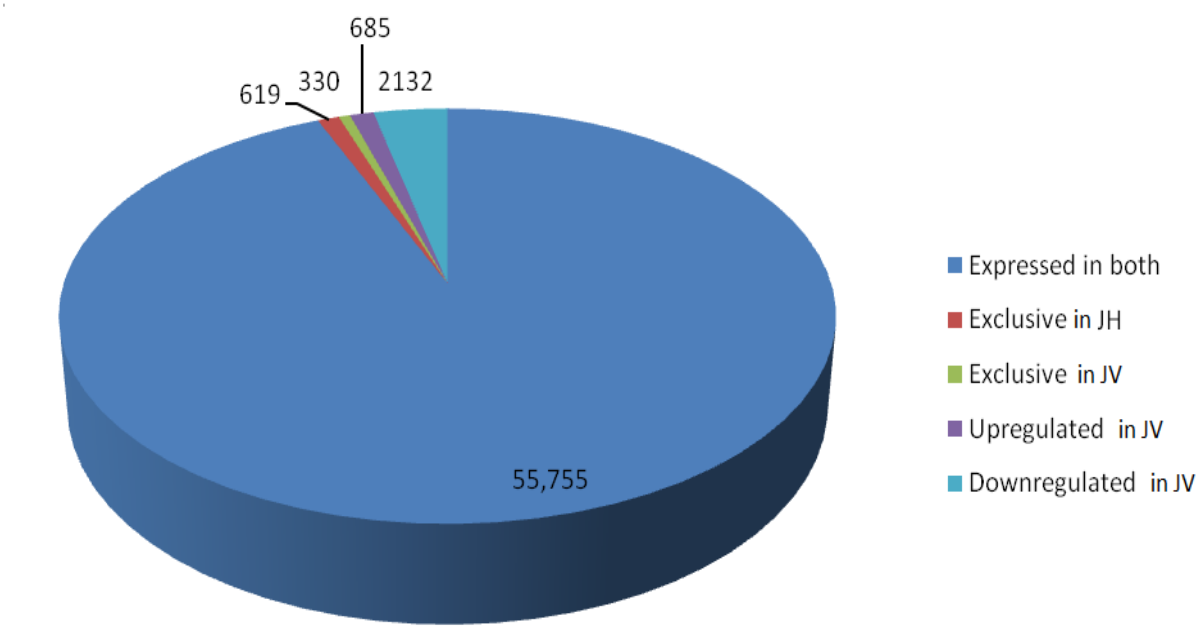


Figure 4.12 Distribution of genes expressed in healthy (JH) and virus infected (JV) leaves of *J. curcas*

followed by 'molecular function' and 'cellular component' categories in both JV and JH derived transcriptomes (Figure 4.14). Further comparison between JV and JH transcriptomes revealed that in JV classification of 'biological process', more percentage of transcripts belonged to terms i.e. response to stress (50%), transport (30%), catabolic process (20%), biosynthetic process (30%), immune system process (20%), signal transduction (15%) and cellular protein modification process (15%) whereas for 'molecular function', nucleic acid binding (40%) was the largest class followed by catalytic (30%), transporter (30%) and antioxidant (25%) activities. For JV, under category 'cellular component' the sub-category integral component of membrane dominated with 35% transcripts followed by membrane (25%) and organelle (25%). In case of JH, for category, 'biological process', photosynthesis (40%) dominated followed by small molecule metabolic process (30%), carbohydrate metabolic process (20%) and reproduction (15%). For category, 'molecular function', protein binding transcription factor activity dominated with 50% transcripts followed by nucleic acid binding, antioxidant, transporter and catalytic activity with less percentage of genes, in JH. Further

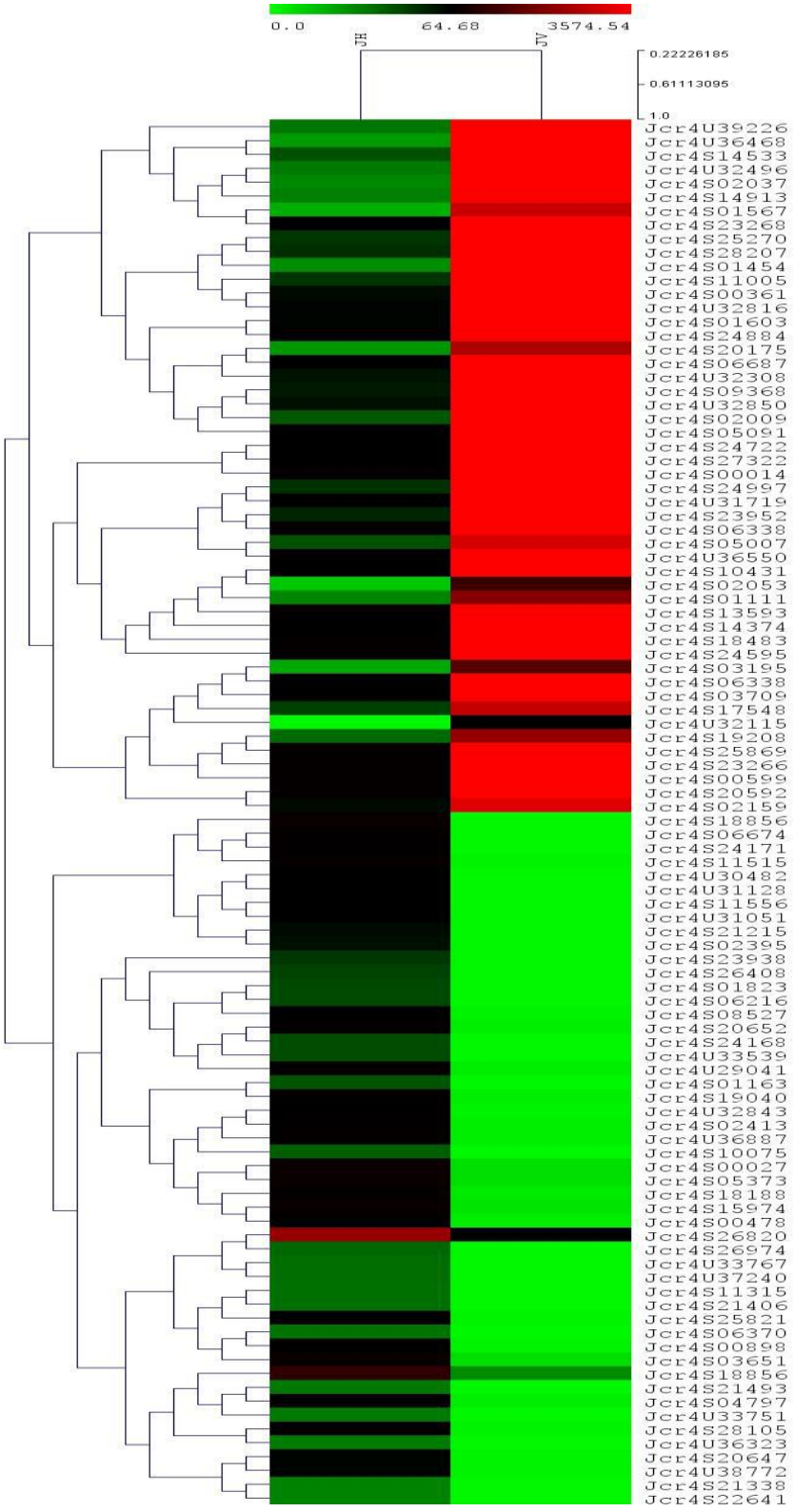


Figure 4.13 Heat map representing top 100 differentially expressed genes (50 up regulated and 50 down regulated in JH or JV)

for ‘cellular component’, sub-category integral component of membrane was with 30% transcripts followed by membrane (25%), organelle (20%) and extracellular region (20%), in JH.

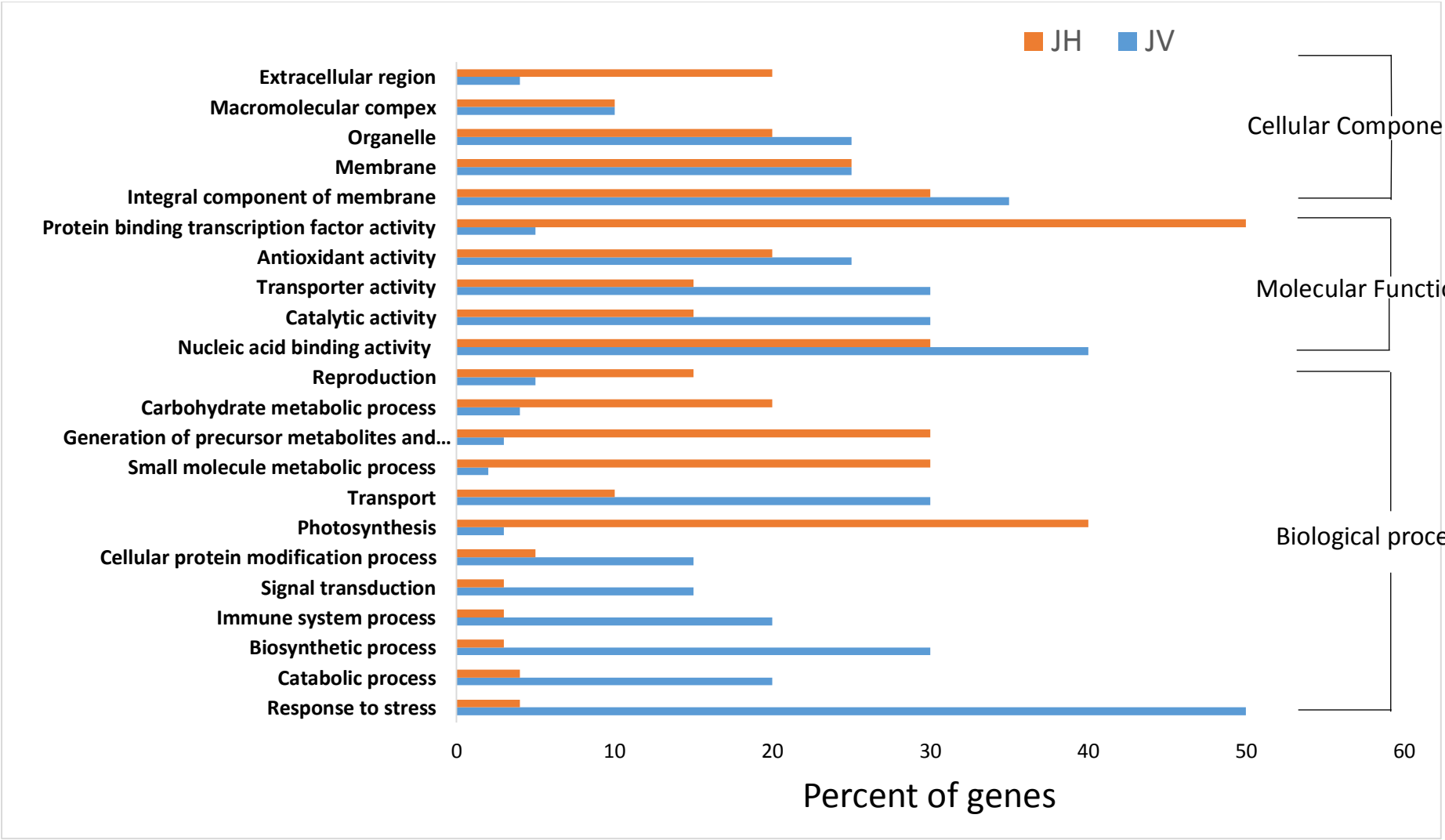


Figure 4.14 GO classification and distribution of GO annotated transcripts in JV- and JH-derived transcriptomes

4.16 Pathway analysis and identification of pathways upregulated in response to viral infection (JV transcriptome)

KAAS (KEGG automatic annotation server) was used for functional annotation in both the samples with BLAST comparisons against KEGG GENES database. KEGG Orthology database was used for pathway mapping. Upon annotation, it was observed that maximum number of transcripts i.e. 773 corresponded to the category 'Signal transduction' followed by 704 in 'Carbohydrate metabolism' and 591 in 'Translation'. The other associated pathways are provided in the figure 4.15. Under the pathway category 'Environmental adaptation', there is a sub-category 'Plant-pathogen interaction' which have 127 transcripts. Upon KEGG annotation, it was observed that a total of 23 different pathways were upregulated in response to virus infection in JV as revealed by the transcripts associated (Figure 4.15). Out of these, majority of upregulated genes mapped to pathways such as oxidative phosphorylation, endocytosis, arginine and proline metabolism, terpenoid biosynthesis, ascorbate metabolism, amino sugar and nucleotide sugar metabolism and lipid metabolism which implied that these were the main pathways upregulated in JV (Figure 4.16). The most affected process was the metabolism process, as most of the annotated genes were correlated to metabolism processes.

4.16.1 Oxidative phosphorylation

It was observed that 80 genes involved in 'Oxidative phosphorylation' were upregulated in response to viral infection (JV) (Table 4.11). Genes such as ATPase, oxidoreductase, oxidase and dehydrogenase were significantly overexpressed (Appendix Table A1).

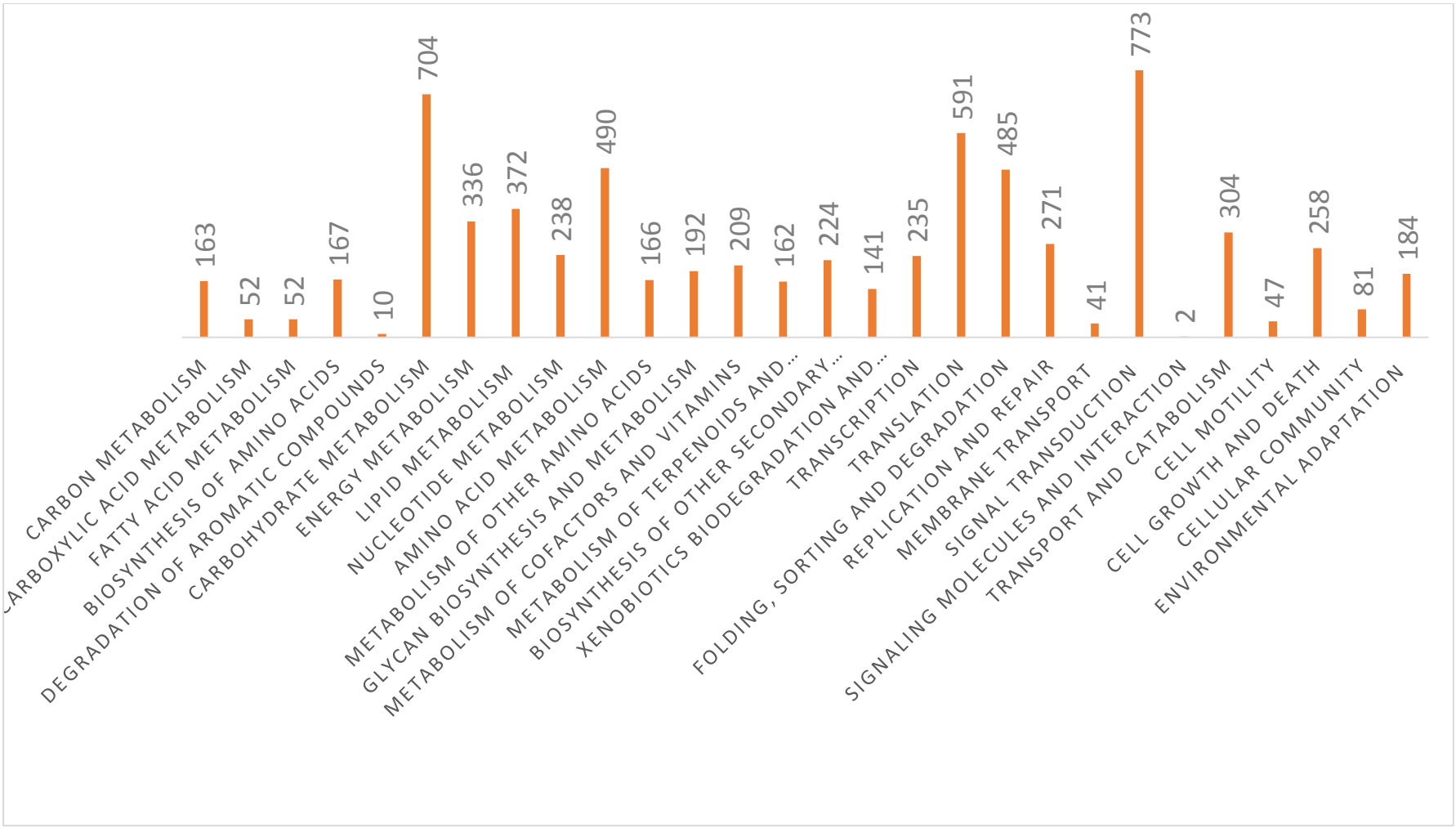


Figure 4.15 KEGG based classification of transcripts in different pathways for virus infected (JV) derived transcriptome

4.16.2 Endocytosis

A total of 40 genes associated with endocytosis process showed higher transcript abundance in response to viral infection (JV) (Table 4.11) out of which charged multivesicular body protein 5, Ras-related protein Rab-11A, epsin and DnaJ homolog subfamily C member were significantly overexpressed in response to viral infection (Appendix Table A2).

4.16.3 Metabolism of amino acids and vitamins

Genes involved in metabolism of arginine and proline (amino acids) and ascorbate (vitamins) were upregulated in response to viral infection (JV) as compared to healthy (JH). 14 genes linked to 'Arginine and proline metabolism' showed upregulation in response to viral infection where genes such as nitric-oxide synthase and prolyl 4-hydroxylase were significantly overexpressed (Table 4.11) (Appendix Table A3). For 'Ascorbate metabolism', 11 genes were upregulated in JV (Table 4.11). Genes such as GDP-L-galactose phosphorylase and L-ascorbate peroxidase were significantly upregulated in response to viral infection (Appendix Table A4).

4.16.4 Fatty acid and lipid catabolism

A total of 56 and 28 genes related to lipid catabolism and fatty acid degradation respectively were found to be upregulated in JV (Table 4.11). For lipid catabolism, the significantly overexpressed genes in JV were lipase and kinase (Appendix Table A5). In case of fatty acid catabolism, genes such as acetyl-CoA acyltransferase, long-chain acyl-CoA synthetase, very-long-chain enoyl-CoA reductase and acyl-carrier-protein desaturase showed significantly higher expression in response to viral infection (Appendix Table A6).

4.16.5 Amino sugar and nucleotide sugar metabolism

It was observed that 28 genes related to sugar metabolism (synthesis) were upregulated in JV as per transcript abundance (Table 4.11). In response to viral infection, the expression level of genes such as UDP-apiose/xylose synthase, and L-arabinokinase were significantly elevated (Appendix Table A7).

4.16.6 Terpenoid biosynthesis

Seventeen genes related to biosynthesis of monoterpenoid, diterpenoid and triterpenoid biosynthesis showed higher transcript abundance in response to viral infection (JV) (Table 4.11). Genes such as synthases, oxidases and dehydrogenases were significantly enriched in response to viral infection (Appendix Table A8).

4.16.7 Signal transduction of hormones

Further, upon KEGG based functional annotation, genes upregulated in 'Plant hormone signal transduction' were identified in response to virus infection. Total of 56 genes were upregulated in JV in response to virus infection (Tables 4.10, 4.11). The majority of genes upregulated belonged to the families involved in hormone signaling. This implied that signaling of various plant hormones such as salicylic acid (SA), ethylene, jasmonic acid (JA), abscisic acid (ABA), auxins, cytokinins and gibberellins is activated during viral infection. Further it was observed that 21 transcription factors linked to signal transduction of plant hormones were upregulated in JV. The upregulated transcription factors were of *MYC2* (3), *TGA* (7), *ABA responsive element binding factor* (9) and *ethylene-responsive transcription factor* (2) TF families (Appendix Table A9).

4.17 Identification of pathways downregulated in response to viral infection

The pathways which were downregulated in response to viral infection (JV) were photosynthesis, anthocyanin biosynthesis, plant-pathogen interaction and calcium signaling pathway (Figure 4.16).

4.17.1 Photosynthesis

Fifty genes involved in photosynthesis showed low transcript abundance in JV compared to JH (Table 4.12). Significantly repressed genes in response to viral infection were genes associated with PSI and PSII, genes related to light harvesting complex, LHC I and LHC II, ferredoxin and cytochrome b6 (Appendix Table A10).

Table 4.10 Gene families of 'Plant hormone signal transduction' activated in JV-derived transcriptome

Gene family	**Upregulated/Downregulated**
SAUR family protein	+
Auxin responsive GH3 gene family	+
Transport inhibitor response 1	-
Pathogenesis-related protein 1	-
Auxin-responsive protein IAA	+
DELLA protein	-
F-box protein GID2	-
Two-component response regulator ARR-B family	+
Auxin influx carrier (AUX1 LAX family)	+
Arabidopsis histidine kinase 2/3/4 (cytokinin receptor)	+
Abscisic acid receptor PYR/PYL family	+
Jasmonic acid-amino synthetase	+
Protein phosphatase 2C	-
BR-signaling kinase	-
Jasmonate ZIM domain-containing protein	+
Regulatory protein NPR1	-
Two-component response regulator ARR-A family	-
Ethylene-insensitive protein	+
BRI1 kinase inhibitor 1	-

(+ Upregulated; - Downregulated)

4.17.2 Anthocyanin biosynthesis

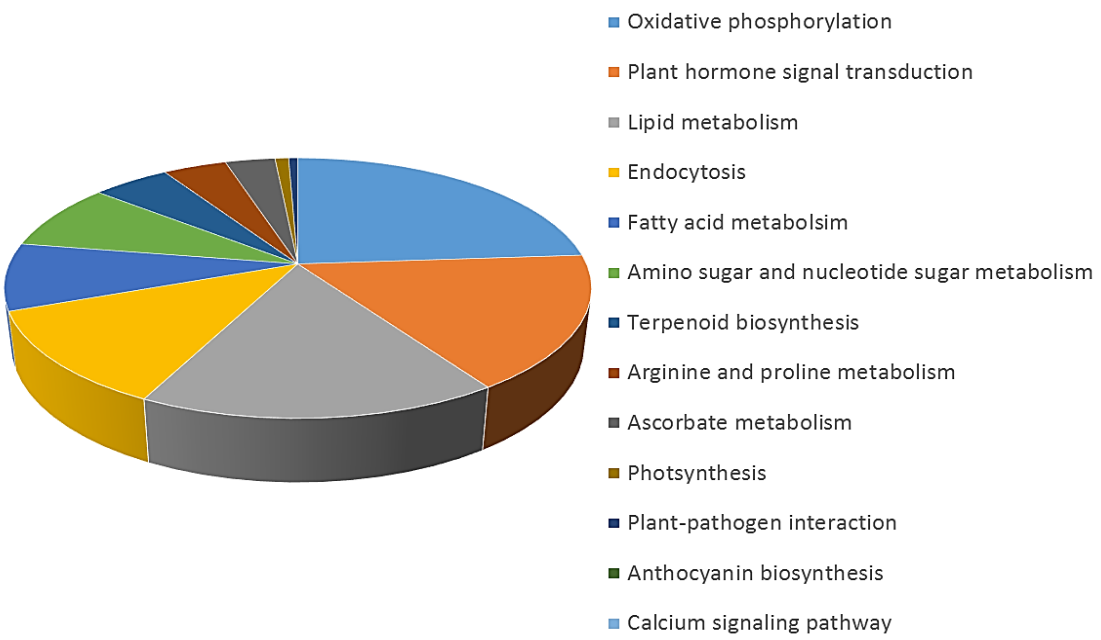
Only 3 genes linked to anthocyanin biosynthesis were identified on the basis of low transcript abundance in response to viral infection (Table 4.12). The genes were anthocyanidin 3-O-glucoside 2'''-O-xylosyltransferase and anthocyanidin 3-O-glucoside 5-O-glucosyltransferase (Appendix Table A11).

4.17.3 Plant-pathogen interaction

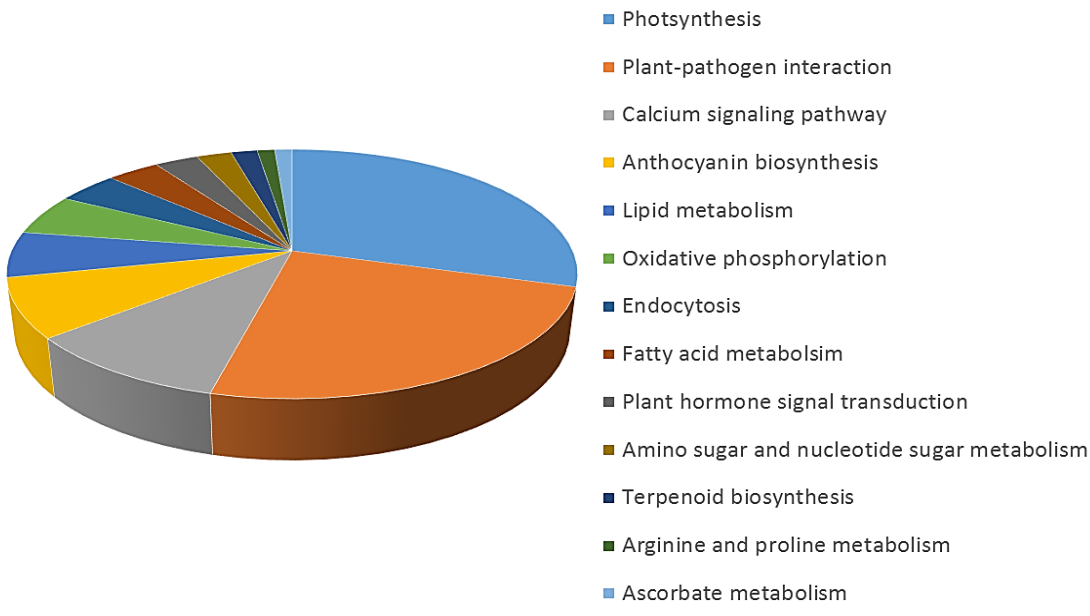
Based on KEGG pathway assignments, it was observed that total of 41 genes linked with 'Plant-pathogen interaction' were downregulated in JV (Table 4.12). Genes related to innate immunity such as cyclic nucleotide gated ion channels (CNGCs), calcium-binding protein (CML) disease resistance proteins i.e. *RPM1*, *RPS2*, kinases and defense related transcription factor genes such as *WRKY* were significantly repressed in JV as compared to JH. Furthermore, 4 transcription factors related to *WRKY* family linked with 'Plant-pathogen interaction' were significantly downregulated in JV (Appendix Table A12).

4.17.4 Calcium signaling pathway

Total of 7 genes linked to Calcium signaling pathway showed low transcript abundance in response to viral infection (JV) (Table 4.12). Genes such as solute carrier family 25 (mitochondrial adenine nucleotide translocator) and voltage-dependent anion channel protein 2 were significantly repressed in JV (Appendix Table A13).



A)



B)

Figure 4.16 Pathways upregulated **(A)** and downregulated **(B)** in response to virus infection

Table 4.11 Number of upregulated genes in JV and JH

Pathway/Process	**Genes upregulated in JV (Number)**	**Genes upregulated in JH (Number)**
Oxidative phosphorylation (Energy metabolism)	80	9
Lipid metabolism	56	8
Plant hormone signal transduction	56	6
Endocytosis	40	7
Fatty acid metabolism	28	4
Amino sugar and nucleotide sugar metabolism	28	3
Terpenoid biosynthesis	17	2
Arginine and proline metabolism	14	2
Ascorbate metabolism	11	2

Table 4.12 Number of downregulated genes in JV and JH

Pathway/Process	**Genes downregulated in JV (Number)**	**Genes downregulated in JH (Number)**
Photosynthesis	50	3
Plant-pathogen interaction	41	4
Calcium signaling pathway	7	0
Anthocyanin biosynthesis	3	0

4.18 Identification of SNPs

SNPs were identified in both the transcriptomes derived from JH and JV after mapping to *J. curcas* genome sequence. The SNPs were detected using the samtools mpileup pipeline. Quality score of more than 20 and read depth of over 5 with flanking of 100 bp were parameters applied to filter SNPs as high quality SNPs. A total of 41,774 SNPs were identified in JH, out of which 21,105 were in coding regions whereas 36,774 SNPs identified in JV, out of which 17,499 were in the coding region (Table 4.13).

Table 4.13 Statistics of SNP identification

Transcriptome	Number of SNPs	Number of SNPs in coding region
JH	41,774	21,105
JV	36,774	17,499

4.19 Co-expression network analysis

For co-expression network analysis, genes from one upregulated pathway and one downregulated pathway in response to viral infection were considered. The genes linked to 'Plant hormone signal transduction' and 'Plant-pathogen interaction' were analysed to identify genes co-expressed and showing interactions with identified reference genes on the basis of scoring function (Figure 4.17). From 'Plant hormone signal transduction' pathway, *HAB1* (HOMOLOGY TO ABI1) had interaction with *ORE14*, *RKP*, *REF4*, *ZIG4A*, *SIZ1*, *LRS1*, *UPL3*, *DegP7*, *KEA2*, *IBR3* and *RR1* whereas closely related to *ZIG4A* and *SIZ1*. Reference gene *CML13 (calmodulin-like protein)* from 'Plant-pathogen interaction' showed interaction with genes such as *SCN1*, *NTF2A*, *PRA1.E*, *RABA2c*, *PRF1*, *UBC22*, *FAH2*, *RAB6A*, *SBH2*, *WPP1*, *WPP2* and *VHA-G2*. *SCN1*, *NTF2A*, *PRF1* and *WPP2* were closely related to reference gene, *CML13*. The interaction of other identified reference genes from both the pathways to other co-expressed genes are shown in figure 4.17. The co-expression network diagram of the gene *CML38* involved in 'Plant-pathogen interaction' is shown in figure 4.18. Co-expression network diagrams of other genes involved in 'Plant hormone signal transduction' and 'Plant-pathogen interaction' are shown in appendix figures A1, A2, A3, A4, A5, A6, A7, A8, A9 and A10.

Reference Genes	Coexpressed Genes												
HAB1	ORE14	RKP	REF4	ZIGA4	SIZ1	LRS1	UPL3	DegP7	KEA2	IBR3	RR1		
IAA3	XCP1	FAF3	FLA11	TRX1	ATAO1	XYP2	SAH7	SBTI1.1	AtAGAL2	TBL3	820169		
CML42	TEM2	PRA8	EDL2	ZHD6									
CML30	UBC17	MCP1a	RTM1	PP2-A1	WDY								
IAA13	TSD1	WAT1	IAA4	ACL5	PAP2	IAA12	IAA9	BRH1	MYC4	SAMDC4	DOF2	SHY2	TMO5
BES1	MPK6	BON2	SBI1	RSW9	WAV2	NADK1	UBP14	GNTI	bHLH121	RFI2			
FCA	SCC2	RIK	FC1	PEX1	SAC3B	REV1	SET29	EDM2					
RR17	BGLU32	AGL78											
CML13	SCN1	NTF2A	PRA1.E	RABA2c	PRF1	UBC22	FAH2	RAB6A	SBH2	842627	WPP1	WPP2	VHA-G2
ABF1	HSFA8	STH3	GI	CYCT1.3	HSFA1B	PSD3	WRKY3	TAG1	PRR5	BET9			
RIN4	ECH2	GG1	ZAR1	SOBIR1	ERD6	PEN3	XBAT34	PMR2	BIR1	CRK19			
PBS1	SYP123	RSH3	CYP89A7	RHS10	RHS18	RHS8	IRE	EXO70C2	TET12				
CML41	SYP123	RSH3	CYP89A7	RHS10	RHS12	RHS18	RHS8	IRE	EXO70C2	TET12			
CPK1	DIS1	PME41	MLO8	CPK4	XI-6	ACA2	PEPR1	MUR3	AtKINUb				
CML38	ATAGP1	WRKY33	CML39	SCABP5	WRKY40	SRO5	ZAT10	ERF11	ERF2	DIC2	SZF1	AR781	PP2-A11

Plant Hormone Signal Transduction

Plant Pathogen Interaction

Figure 4.17 List of co-expressed genes in 'Plant-pathogen interaction' and 'Plant hormone signal transduction' pathways. Red colour genes showed close association with reference genes. Reference genes in bold letters representing homologs

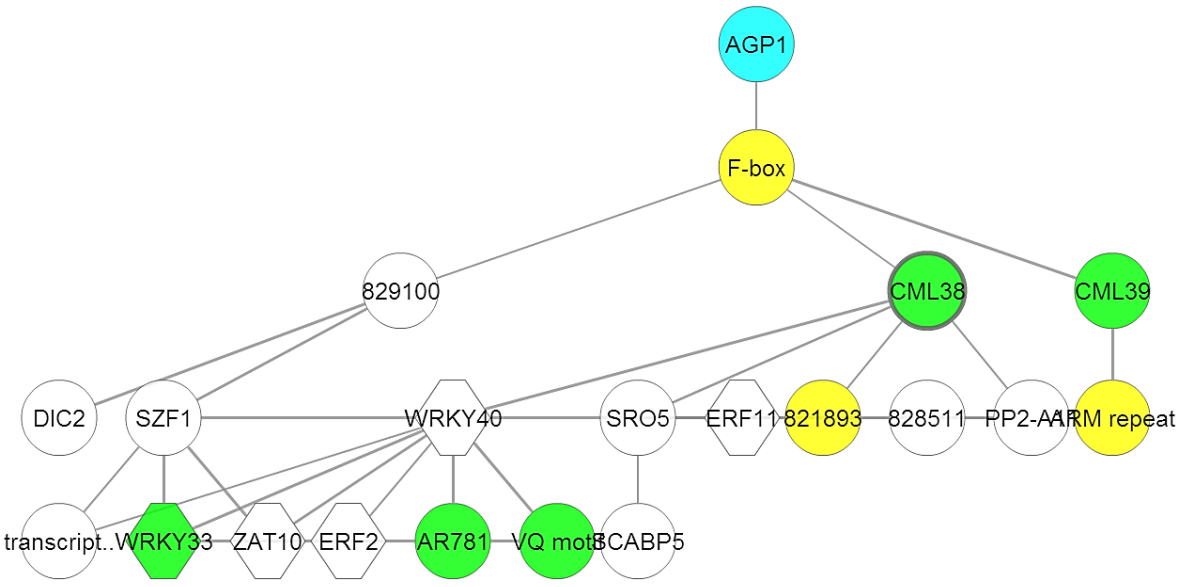


Figure 4.18 Co-expression network analysis of gene *CML38* in 'Plant-pathogen interaction'

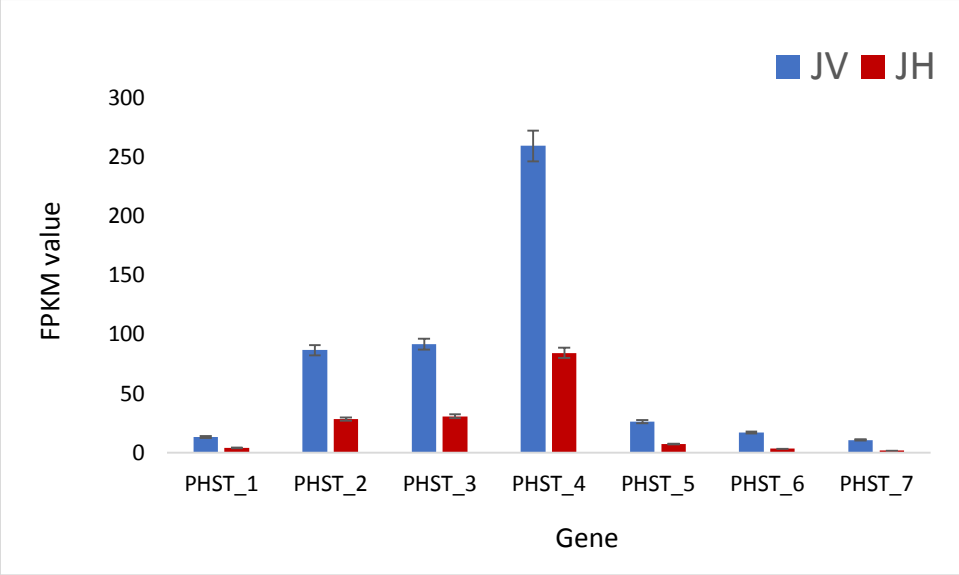
4.20 RT-qPCR based validation of informative transcripts

For experimental validation, a set of 7 genes involved in 'Plant hormone signal transduction' showing more than three fold FPKM values in JV as compared to JH, were examined. The genes from 'Plant hormone signal transduction' were selected for

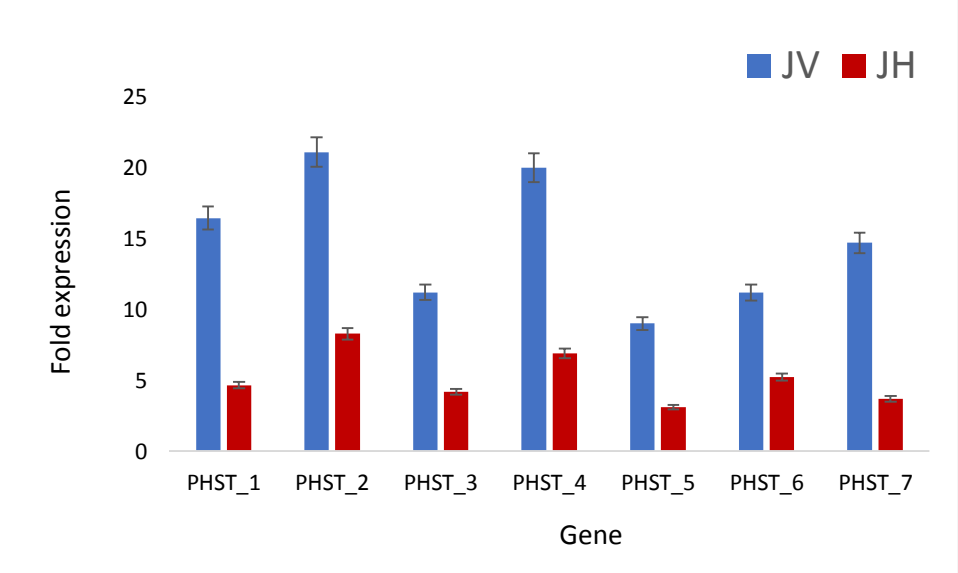
experimental validation as maximum identified genes belonged to this process. Identified transcription factors showing upregulation and downregulation were also analyzed using RT-qPCR to confirm the transcriptome data. All 7 genes showed higher expression in JV as compared to JH (Figure 4.19). For example, fold expression of 16 was observed in JV for gene *PHST_1* (Jasmonate ZIM domain-containing protein) as compared to 4 fold expression value in JH. Maximum fold expression value of 20 was observed for gene *PHST_2* (Auxin responsive GH3 gene family) in case of JV whereas the fold expression value was 7.5 in JH. For transcription factors, *MYC2* showed fold expression value of 3.5 in JV whereas 1.2 in JH. TGA exhibited fold expression of 4.5 in JV and 2 in JH. *ABA responsive element binding factor* showed fold expression of 3.3 in JV followed by 1.4 in JH. Similarly, *ethylene-responsive transcription factor* had fold expression of 2.2 in JV and 1 fold expression value in JH. On the other hand *WRKY* showed higher expression in JH (6.5) and low fold expression of 2.2 in JV (Figure 4.20). Computational identification of transcript abundance and validation with RT-qPCR are in conjunction with each other.

4.21 Identification of NBS-LRR genes and defense response associated transcription factors in *J. curcas*

45 NBS-LRR genes were identified by mapping Pfam domains and 7 by mapping NBS-LRR mRNA sequences with BLAST analysis out of which 5 showed common identity to both Pfam domain mapping and BLAST analysis (mRNA sequence mapping). A total of 47 new NBS-LRR genes were identified in *J. curcas* genome in addition to previously identified (92) NBS-LRR genes. All these newly identified NBS-LRR genes were confirmed for unique identity through similarity search with previously reported NBS-LRR genes [13] for their uniqueness. Similarity search (70% identity) provided that all identified NBS-LRR were new and not reported earlier. Similarly when Pfam domains specific to the transcription factors involved in disease resistance or defense response were mapped on to the transcriptome, 122 transcription factors were identified.



A)



B)

Figure 4.19 *In-silico* analysis based transcript abundance (FPKM) **(A)** and RT-qPCR based fold expression pattern **(B)** of genes involved in signal transduction of hormones showing about and more than two fold in JV as compared to JH (*PHST_1*: Jasmonate ZIM domain-containing protein, *PHST_2*: Auxin responsive GH3 gene family, *PHST_3*: Two-component response regulator ARR-A family, *PHST_4*: Auxin-responsive protein IAA, *PHST_5*: Abscisic acid receptor PYR/PYL family, *PHST_6*: Two-component response regulator ARR-B family, *PHST_7*: SAUR family protein)

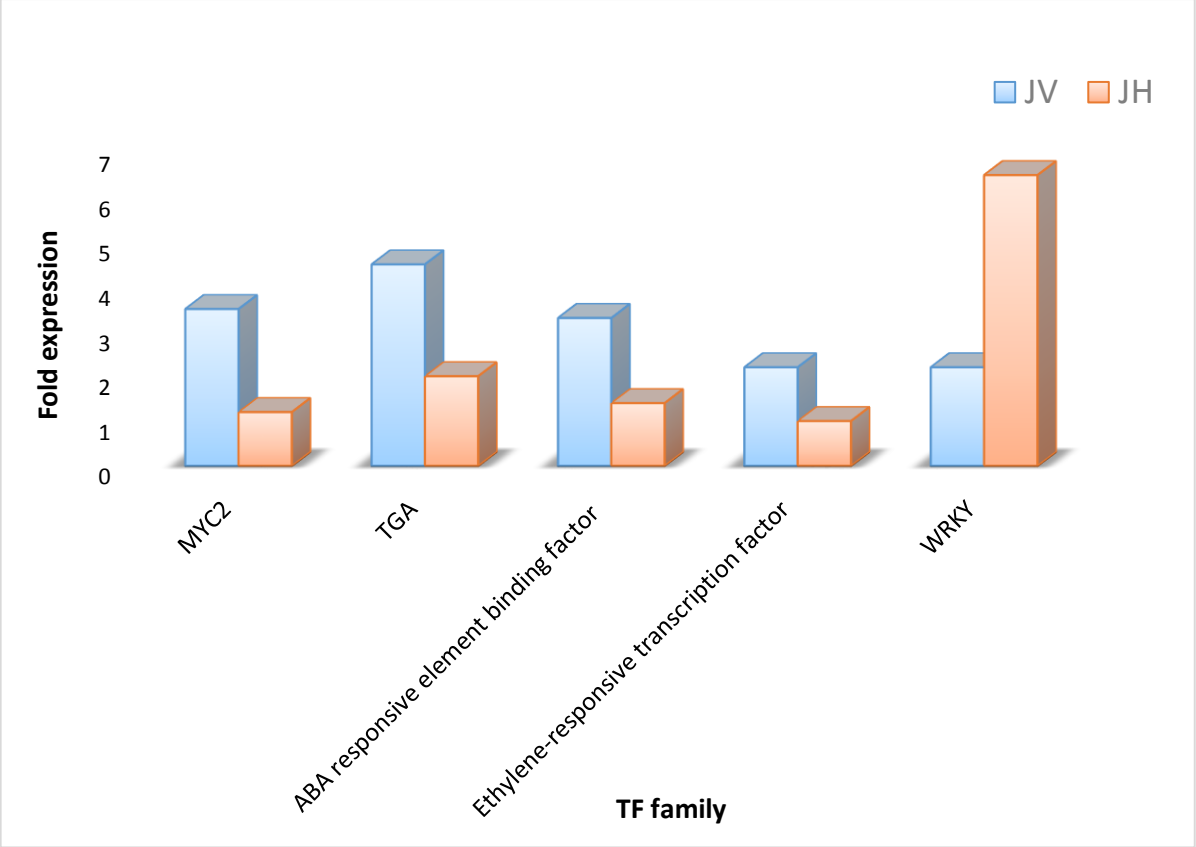


Figure 4.20 RT-qPCR based fold expression pattern of transcription factors upregulated and downregulated during virus infection

4.22 Location of NBS-LRR genes in genome sequence contigs

The identified NBS-LRR genes were mapped on to the respective *J. curcas* genome sequence contigs using BLAST search. The contigs showing multiple matches were manually curated to cover the entire query length. The query genes mapped on to the respective genome contigs showed identity in the range of 95–100%. In case of Jatropha according to location, NBS-LRR genes were classified into three categories. First category had 28 genes, each of them were located in a single contig without any disruption in the coding sequence, thereby suggesting that these genes lacked introns. Second category had 7 genes and each were mapped on to single contigs with one or more gaps indicating insertion which may correspond to an intron. Third category comprised of 12 genes and each of these genes had match in more than one genomic contigs which implied that these genes were transcribed from different location and may have introns. Out of 47 genes, 16 genes were found in cluster of two i.e. these 16 genes were present in 8 genomic contigs

(Jc476461637, Jc476470256, Jc476481852, Jc476483387, Jc476485273, Jc476487282, Jc476487650 and Jc476489371).

4.23 Transcript abundance of NBS-LRR genes and transcription factors associated with disease resistance

RSEM was used for the transcript abundancy measurements of identified set of NBS-LRR genes and transcription factors associated with the disease resistance mechanism. The expression profiles were obtained through pme_TPM (pme: posterior mean estimates; TPM: transcripts per million) values using RSEM software package. In RSEM, posterior mean estimate (pme) is computed for each gene and isoform abundance, with a maximum likelihood (ML) estimate [195]. pme_TPM for transcriptome samples were generated. Expression profile of 47 NBS-LRR genes and disease resistance specific transcription factors was mined. The pme_TPM values of genes ranged between 0.4–133.54 in Jatropha. In case of Jatropha out of 47 genes, gene with id JcNL_14680 showed highest pme_TPM value i.e. 133.54 (Table 4.14). When common genes between Jatropha and castor bean were analyzed, it was found that in Jatropha gene JcNL_10121 showed highest pme_TPM value of 9.28 while its corresponding gene in castor bean i.e. RcNL_57743 showed pme_TPM value of 3.12. In case of castor bean, gene RcNL_34103 showed the highest pme_TPM value i.e.10.46 as compared to Jatropha where the corresponding gene JcNL_00810 showed value of 0.64. The pme_TPM values for transcription factors ranged from 0.42 to 289.67 for Jatropha. Similarly, on analyzing the transcript abundancy of transcription factors it was found that in case of Jatropha the transcription factor JcTF_15319 showed highest pme_TPM value i.e. 289.67 (Table 4.14). On analyzing the common transcription factors between Jatropha and castor bean, it was found that in Jatropha the transcription factor JcTF_14789 showed highest pme_TPM value of 142.48 whereas its corresponding transcription factor in castor bean RcTF_20625 showed a value of 9.05. In case of castor bean, the transcription factor RcTF_32546 gave highest pme_TPM value of 237.15 while its counterpart in Jatropha i.e. JcTF_04420 showed a value of 1.93. Further it was observed that 4 transcription factors showed higher pme_TPM values in Jatropha (range from 100 to 300) i.e. JcTF_14789, JcTF_14930, JcTF_15218 and JcTF_15319. The pme_TPM values for identified NBS-LRR genes and defense response related transcription factors of *J. curcas* are shown in appendix tables A14 and A15.

Table 4.14 Top ten NBS-LRR genes and transcription factors related to defense response on the basis of transcript abundance (pme_TPM values)

Gene ID	**pme_TPM**	**TF ID**	**pme_TPM**
JcNL_10695	11.17	JcTF_12728	20.85
JcNL_11173	13.08	JcTF_12732	20.85
JcNL_11632	14.73	JcTF_12677	22.38
JcNL_12635	20.04	JcTF_13181	24.31
JcNL_12889	22.61	JcTF_13463	27.35
JcNL_13460	27.65	JcTF_13501	28.26
JcNL_13997	47.46	JcTF_14789	142.48
JcNL_14601	107.75	JcTF_14930	167.1
JcNL_14712	128.5	JcTF_15218	243.07
JcNL_14680	133.54	JcTF_15319	289.67

4.24 Identification of TNLs and CNLs in identified NBS-LRR genes

For the prediction of TNLs (Toll/interleukin-1 receptor NBS-LRR genes) and CNLs (Coiled-coil NBS-LRR genes) in the identified transcripts, PCOILS used sliding windows of 14 (green), 21 (blue), and 28 (red) and predictions were made based on coiled coil probability. Out of 47 identified NBS-LRR genes, 37 were predicted as TNLs and 10 were CNLs (Table 4.15).

Table 4.15 Categorization of NBS-LRR genes into TNLs and CNLs

Gene	**TNL/CNL**
JcNL_00073	TNL
JcNL_00090	TNL
JcNL_00096	TNL
JcNL_00134	TNL
JcNL_00259	TNL
JcNL_00553	TNL

JcNL_00565	TNL
JcNL_00587	TNL
JcNL_00618	CNL
JcNL_00810	TNL
JcNL_00860	TNL
JcNL_01339	TNL
JcNL_01427	TNL
JcNL_02388	TNL
JcNL_02545	TNL
JcNL_03063	CNL
JcNL_03152	CNL
JcNL_04261	CNL
JcNL_04456	TNL
JcNL_04489	TNL
JcNL_04766	TNL
JcNL_04778	TNL
JcNL_04926	TNL
JcNL_06396	TNL
JcNL_06992	TNL
JcNL_07289	TNL
JcNL_08981	TNL
JcNL_09160	CNL
JcNL_09317	TNL
JcNL_09428	TNL
JcNL_09636	TNL
JcNL_10049	CNL
JcNL_10121	CNL
JcNL_10302	TNL
JcNL_10695	TNL
JcNL_11173	TNL
JcNL_11632	TNL
JcNL_12635	TNL

JcNL_12889	TNL
JcNL_13460	TNL
JcNL_13997	TNL
JcNL_14601	TNL
JcNL_14680	TNL
JcNL_14712	CNL
JcNL_16532	CNL
JcNL_16899	CNL
JcNL_17188	TNL

4.25 Distribution of identified transcription factors into families

The identified transcription factors related to defense response were further classified into families on the basis of BLASTn analysis. Upon similarity search with BLAST it was observed that families *NAM*, *WRKY*, *MYB* and *Homeo-domain* contributed to large number of transcription factors followed by families like *ERF-type/AP2-EREBP*, *bZIP*, *TFIIA*, *CBF*, *SBP* and *Whirly* (Table 4.16).

4.26 Identification of common and unique NBS-LRR genes and transcription factors between Jatropha and castor bean genomes

The identified 47 genes in Jatropha were aligned to genes of castor bean in order to analyze the common and unique NBS-LRR genes between Jatropha and castor bean. This analysis was performed using BLASTN with cut off values i.e. identity 70 % and length of 100 bp. In order to identify common and unique genes, the genes from castor bean were taken as database whereas the genes from Jatropha were taken as query. In case of Jatropha, 7 genes showed identity to castor bean genes whereas in castor bean 8 genes showed identity to Jatropha genes implying that 7 and

Table 4.16 Distribution of identified defense related transcription factors into families

TF family	**Number**
NAC domain/NAM	25
WRKY	18
MYB	18
Homeo-domain	17
ERF-type/AP2-EREBP	13

bZIP	11
TFIIA	8
CBF	5
SBP/SPL6	5
Whirly	2

8 NBS-LRR genes are common between Jatropha and castor bean, respectively (Table 4.17). It was found that in case of castor bean, out of 8 genes, 6 showed identity to each and specific gene from Jatropha whereas 2 genes from castor bean showed similarity to 1 common (same) gene from Jatropha. Further 40 and 39 genes were uniquely present in Jatropha and castor bean genomes, respectively. In case of transcription factors (TF), castor bean TFs were taken as database whereas the TFs from Jatropha were taken as query. 70 transcription factors were found common in both Jatropha and castor bean (Table 4.18). Further, 52 and 255 transcription factors were found to be uniquely present in Jatropha and castor bean, respectively.

Table 4.17 Common NBS-LRR genes between Jatropha and castor bean genomes

Transcript IDs of Jatropha	**pme_TPM value (Transcript abundance)**	**Transcript IDs of castor bean**	**pme_TPM value (Transcript abundance)**	**Sequence similarity (%)**
JcNL_10049	9.16	RcNL_61167	2.27	78
JcNL_10121	9.28	RcNL_57743	3.12	80
JcNL_06992	3.91	RcNL_44733	2.65	86
JcNL_00860	0.66	RcNL_50409	3.77	81
JcNL_00810	0.64	RcNL_34103	10.46	80
JcNL_00096	0.41	RcNL_44257	3.11	79
JcNL_00073	0.40	RcNL_33311	5.99	82
JcNL_00073	0.40	RcNL_25810	4.37	73

Table 4.18 Common defense-related transcription factors between Jatropha and castor bean

Transcription factor IDs of Jatropha	**pme_TPM value (Transcript abundance)**	**Transcription factor IDs of castor bean**	**pme_TPM value (Transcript abundance)**	**Sequence similarity (%)**
JcTF_00115	0.42	RcTF_54884	8.11	81
JcTF_00439	0.54	RcTF_19852	6.74	82
JcTF_00764	1.26	RcTF_53730	3.03	84
JcTF_00785	0.63	RcTF_44818	6.96	88
JcTF_00918	1.33	RcTF_20813	21.38	87
JcTF_00978	0.68	RcTF_20336	10.4	88
JcTF_01151	1.45	RcTF_58632	4.72	86
JcTF_01250	0.75	RcTF_16473	10.98	85
JcTF_01393	0.78	RcTF_41238	10.91	82
JcTF_01461	0.8	RcTF_24119	10.32	82
JcTF_01506	0.81	RcTF_69743	2.19	86
JcTF_01687	0.86	RcTF_49729	5.06	83
JcTF_01737	0.88	RcTF_24287	17.16	88
JcTF_01778	1.78	RcTF_28835	47.82	77
JcTF_01856	0.91	RcTF_42528	6.62	87
JcTF_01856	0.94	RcTF_26203	14.65	84
JcTF_02046	0.96	RcTF_32079	36.22	91
JcTF_02082	0.98	RcTF_39006	11.54	83
JcTF_02191	1	RcTF_2255	124.52	78
JcTF_02233	1.01	RcTF_29374	10.63	83
JcTF_02499	1.09	RcTF_37931	5.65	79
JcTF_02772	1.17	RcTF_22987	11.23	84

JcTF_02905	1.23	RcTF_21956	10.01	85
JcTF_03033	1.28	RcTF_14008	29.54	80
JcTF_03086	1.3	RcTF_16180	7.4	81
JcTF_03225	1.35	RcTF_6142	6.89	82
JcTF_03268	1.37	RcTF_36148	13.09	84
JcTF_03301	1.38	RcTF_35550	8.79	85
JcTF_03314	1.39	RcTF_49024	8.88	83
JcTF_03366	1.41	RcTF_1841	9.47	87
JcTF_03375	1.41	RcTF_19375	47.55	80
JcTF_03662	1.54	RcTF_92723	2.31	87
JcTF_03682	1.57	RcTF_49327	45.18	91
JcTF_04051	1.74	RcTF_47525	7.51	86
JcTF_04119	1.77	RcTF_45058	9.88	88
JcTF_04123	1.78	RcTF_30372	14.28	80
JcTF_04169	1.81	RcTF_20813	21.38	87
JcTF_04237	1.83	RcTF_59197	11.76	92
JcTF_04330	1.88	RcTF_30440	4.42	82
JcTF_04386	1.91	RcTF_58524	8.6	82
JcTF_04420	1.93	RcTF_32546	237.15	90
JcTF_04493	1.98	RcTF_43799	3.67	86
JcTF_04775	2.14	RcTF_53152	6.76	79
JcTF_04781	2.15	RcTF_2430	13.91	85
JcTF_04854	2.17	RcTF_70181	7.5	85
JcTF_05008	2.27	RcTF_68484	2.1	83
JcTF_05026	2.27	RcTF_41658	8.3	81
JcTF_05255	2.43	RcTF_12179	15.78	86

JcTF_05708	2.75	RcTF_30440	4.42	89
JcTF_05776	2.79	RcTF_48430	12.96	89
JcTF_05940	2.93	RcTF_31599	19.53	86
JcTF_05953	2.93	RcTF_11012	19.15	89
JcTF_06110	3.08	RcTF_61146	3.66	90
JcTF_06232	3.16	RcTF_35772	5.8	85
JcTF_06247	3.18	RcTF_76068	10.84	82
JcTF_06299	3.21	RcTF_40708	11.05	79
JcTF_06326	3.23	RcTF_38145	11.11	93
JcTF_06768	3.58	RcTF_35340	5.07	86
JcTF_07041	3.88	RcTF_55941	2.16	83
JcTF_08297	5.55	RcTF_30401	2.24	83
JcTF_08810	6.37	RcTF_49389	16.86	92
JcTF_08950	6.79	RcTF_40676	11.13	84
JcTF_09429	7.7	RcTF_47609	4.89	75
JcTF_10466	10.77	RcTF_51270	5.93	89
JcTF_11496	14.23	RcTF_24005	18.95	92
JcTF_11676	15.4	RcTF_41259	10.75	85
JcTF_12223	17.84	RcTF_64484	5.14	84
JcTF_12486	19.08	RcTF_18048	14.93	86
JcTF_12728	20.85	RcTF_30480	13.19	89
JcTF_14789	142.48	RcTF_20625	9.05	87

4.27 Organization of disease resistance genes in castor bean and Jatropha genome

In Jatropha, out of 92 genes, 82 genes were distributed among 82 contigs and no significant result was found for 10 other genes. Further analysis revealed that 7 of the disease resistance genes present in castor bean genome, viz. XM_002517562.1,

XM_002517561.1, XM_00 2518665.1, XM_002517526.1, XM_002517548.1, XM_0025 21759.1, and XM_002529578.1 showed similarity to Jatropha genome, corresponding to genes XP_002517608.1, XP_0025 17607.1, XP_002518711.1, XP_002517572.1, XP_0025175 94.1, XP_002521805.1, and XP_002529624.1, respectively. Due to the presence of same domains and motifs, the genes were further clustered in varying sizes, comprising 2–4 genes in most clusters. Although both Jatropha and castor bean showed almost similar type of motifs (Kinase 1, Kinase 2, Kinase 3, GLPL, MHDL, and AAA+) which were found to be conserved in their disease resistance genes, certain differences were also observed with respect to the presence of conserved domains, which included presence of dirigent domain/superfamily along with protein kinase domain in castor bean genome, and RPW8 domain/ superfamily which was found to be unique to Jatropha genome (Table 4.19) (Appendix Table A16).

4.28 Data availability

The sequences of identified NBS-LRR genes and defense response related transcription factors are available at the link:

http://sites.google.com/site/combiogroup/datadownload.

Table 4.19 Comparative distribution of NBS-LRR disease resistance genes between Jatropha and castor bean genomes

Characteristics	**Jatropha**	**Castor bean**
Genome size	~410 Mb	~380 Mb
Number of NBS LRR disease resistance genes present	92	121
Unique domain/Superfamily	RPW8	Dirigent, protein kinases
Occurrence of NBS domain	~0.3 %	~0.4 %

CHAPTER 5

DISCUSSION

Overall enhancement of oil production is the long-term objective for plant-based biodiesel production. *J. curcas* due to its various merits is considered as a potential source of biodiesel. However, some major constraints like low productivity in terms of yield, variation in oil content among genotypes, non-availability of sufficient feedstock and susceptibility to various biotic stresses have been limitating this plant species as a viable alternative for biodiesel production. As Jatropha is becoming vulnerable to various diseases, specifically viral diseases which in turn reduce its seed yield and oil content, identification of molecular insights to comprehend disease response was crucial. Also, there was a requirement to identify molecular components associated with defense response in *J. curcas* since no systematic breeding efforts have been made towards the development of disease resistant genotypes in this species. Therefore, present study was carried out with an aim of elucidating molecular basis of oil biosynthesis and accumulation, oil content variation among genotypes and understanding molecular mechanisms and components underlying disease response and disease resistance in *J. curcas*. This research work has provided leads which can be taken forward to carry out genetic improvement for enhancement of oil content and disease resistance in *J. curcas*. The results obtained are discussed as under:

5.1 Variation in oil content among high and low oil content genotypes of *J. curcas*

Oil content analysis indicated that there is significant variation for oil content in two genotypes i.e. 30% (low oil content) and 42% (high oil content) in IC 561227 and IC 561235, respectively in accordance to Kaushik and Bhardwaj [190]. In Jatropha, the mature stage of fruit is generally considered suitable for harvesting as it has high oil accumulation. Endosperm has been reported to accumulate more oil content compared to embryo in Jatropha seeds due to more number of oil bodies. The lipid accumulation increases from ripened to mature stages of endosperm and embryo as the oil bodies formation and deposition are developmentally regulated in these stages. The mature (Brown) and ripened (yellow) stages of endosperm and embryo have high oil accumulation followed by unripened stage in Jatropha.

5.2 Identification of genetic factors responsible for high oil content and key genes associated with oil biosynthesis and accumulation in *J. curcas*

The unripened stage in both endosperm and embryo was considered for calculating relative expression fold in ripened and mature stages [11, 14, 33]. As the oil formation in plant

seeds depends upon FA and TAG biosynthesis and accumulation of triacylglycerol in oil bodies [1, 16], all the genes corresponding to FA and TAG biosynthetic pathway showed higher expression in high oil content genotype as compared to low oil content genotype at both the locations. Genes *ACCase*, *KASI*, *KASII*, *KASIII*, *SAD*, *OAD*, *FATA*, *DGAT*, *LPAT* and *GPAT* from three clusters showed significantly higher expression in oil accumulating developmental stages whereas *MT*, *ER*, *PT*, *PAD*, *ST*, *OCD*, *LD* and *PAP* showed low expression. The genes showing higher expression might be contributing towards differential oil biosynthesis and accumulation in high versus low oil content genotypes of Jatropha. These results are in agreement with the previous reports where oil content was developmentally regulated and correlated with the expression pattern of FA and TAG biosynthetic pathway genes in other oil plants such as castor bean [212, 213], oil palm [214], *Brassica napus* [215] and sesame [55]. Except for genes *OAD* and *GPAT*, developmental regulation of all genes was consistent with previous studies in Jatropha endosperm and seed [11, 33]. Variation in seed oil content between high and low oil content Jatropha genotypes might be attributed to differences in transcript accumulation of oil biosynthesis genes in endosperm and embryo [216] as shown by expression variation.

Higher transcript abundance for *ACCase* was observed in oil accumulating developmental stages of embryo and endosperm which is linked with enhancement of overall fatty acid production as has been reported for maize, tobacco, Jatropha, canola [33, 65, 84, 86]. All three subunits of *KAS* i.e. *KASI*, *KASII*, *KASIII* showed high level of transcript abundance in oil accumulating stages of embryo as well as endosperm consistent with Xu et al. [33] but slightly different from Gu et al. [11] where late accumulating stages of endosperm did not exhibit peak expression. However, slight differences w.r.t. subunits expression were also observed and it was concluded that out of *KAS* enzyme complex, *KAS II* may be a potential target for oil engineering at embryo level in Jatropha [89]. *SAD* being a rate limiting enzyme is involved in the conversion of stearoyl-ACP to oleoyl-ACP, which is the precursor for oleic acid, the major fatty acid of Jatropha seed oil [217]. Gene encoding *SAD* showed significantly higher level of expression in ripened and mature (lipid accumulation) stages of both embryo and endosperm, being higher in mature stage in both the cases which is in positive correlation with the amount of oleic acid and is consistent with previous study in Jatropha [11]. In the past, a number of studies have been performed in oil plants to correlate the expression of *SAD* and other desaturase (*FAD*) genes with increase in oleic acid composition to alter seed oil content [65, 91, 92, 218,]. *FATA*, an

important enzyme of class thioesterase directly synthesize principal unsaturated fatty acids such as oleic acid and linoleic acid from their precursors in Jatropha seed oil. Higher transcript abundance of *FATA* in ripened and mature stages of both embryo and endosperm of high oil content genotypes supports the fact that oleic acid and linoleic acid are major fatty acids in Jatropha. For *LPAT*, consistent increase in the expression level at ripened and mature developmental stages was observed which is in accordance with previous study for endosperm [11]. However, it was also observed that for mature stage, *LPAT* gene showed almost 2x fold increase in transcript abundance in both endosperm and embryo. It is hereby suggested that *LPAT* gene can be over expressed mainly at mature stage of embryo and endosperm to enhance storage lipid production in Jatropha [97, 98, 99]. Expression level of *DGAT* is associated with lipid accumulation as it is directly linked with the formation of triacylglycerol. *DGAT* showed a gradual increase in transcript abundance at mid accumulating stages, however slight decline at late development stage of endosperm and this observation differed from Gu et al. [11]. Metabolic engineering approaches for *DGAT* gene have also been performed to alter the oil content and quality in many plants like maize, soybean, Arabidopsis, tobacco [93, 96].

Interestingly, it was also observed that genes encoding *OAD* and *GPAT* enzymes showed higher level of transcript abundance with significant fold increase (10x) suggesting that these genes might have important role in oil biosynthesis. These genes were not identified in previous studies [11, 33]. *OAD* belongs to the desaturase class that converts oleoyl-ACP to linoleoyl-ACP, the precursor of linoleic acid, which is second most abundant unsaturated fatty acid in Jatropha seed oil. The direct role of *OAD* in linoleic acid formation was inferred due to higher expression in oil accumulating stages of embryo and endosperm. For oil plants, a number of reports highlighting overexpression of desaturase genes for increase in linoleic acid and oil content exists [58, 91, 219, 220, 221]. *GPAT* initiates the biosynthesis of triacylglycerols with the help of free fatty acids and glycerol-3- phosphate. Gene encoding *GPAT* exhibited 25-40 fold increase in expression in mature embryo and endosperm stage of high oil content genotype at SH and NH locations. These findings support previous studies for *GPAT* in castor bean [213]. Over expression of *GPAT* in *Brassica napus* and Arabidopsis seeds has shown increase in oil accumulation and content [100, 101, 222]. The *OAD* and *GPAT* genes can thus be suitable targets in genetic improvement to enhance the overall oil production in the accumulating stages of seeds, especially in endosperm and embryo in Jatropha.

Principal component analysis provided a comprehensive correlation of FA and TAG biosynthesis pathway genes with oil accumulating developmental stages in conjunction to expression profiling. PCA of 3 components showed that influence of F1 component was more than other two components.

5.2.1 Molecular basis of oil accumulation vis-à-vis altitudinal variations

Out of 10 genes, six genes of FA and TAG biosynthesis i.e. *ACCase*, *KASII*, *KASIII*, *SAD*, *LPAT* and *DGAT* showed higher transcript abundance in any oil accumulating developmental stage of embryo/endosperm at low altitude location (Nalagarh, NH) whereas four genes (*KASI*, *OAD*, *FATA* and *GPAT*) showed higher expression in high altitude location (Sunni, SH). Among six genes showing higher expression in low altitude, most genes encode rate-limiting enzymes of FA and TAG biosynthesis such as *ACCase*, *SAD*, *DGAT*, supporting the fact that oil content and fatty acid content increase with a decrease in altitude [81, 223]. The reduction in oil content at higher altitudes might be due to the fact that low partial pressure of carbon-dioxide (CO_2) at higher altitude locations reduces the rate of photosynthesis and therefore decrease in oil content [80, 81]. Also it may be possible that the difference in gene expression between altitudinal variations despite similar oil content is due to environmental cues like altitude, temperature, solar radiation, etc. affecting the biosynthesis but not oil deposition [224, 225]. Genes *ACCase*, *KASII*, *KASIII*, *SAD*, *LPAT* and *DGAT* are hereby referred to as the potential candidates for genetic engineering to alter fatty acid composition and oil content in seeds of *J. curcas* as it is confined mainly to the lower altitude regions.

5.2.2 Molecular basis of high oil accumulation in endosperm as compared to embryo in *J. curcas*

Endosperm has been reported to accumulate more oil (65-70%) as compared to embryo part (8-10%) of seeds in Jatropha [14]. To provide molecular insights for discriminating endosperm and embryo with respect to oil accumulation, it was observed that the genes from FA and TAG biosynthetic pathway showed considerable fold increase in oil accumulating stages for endosperm to embryo ratio, indicating low oil content in embryo compared to endosperm [10, 14]. As the transcript abundance for endosperm to embryo ratio was consistent and similar in high oil content genotype, it is inferred that the endosperm of high oil content genotype might be contributing to variations in oil content between two genotypes. Further on comparing expression of endosperms and embryos

between high and low oil content genotypes, it was observed that except *OCD*, all genes showed more transcript abundance in high oil content genotypes. This analysis provided molecular cues for more oil accumulation in endosperm of *J. curcas* seeds rather than embryo.

Based on overall observations, genes *ACCase*, *KASI*, *KASII*, *KASIII*, *SAD*, *OAD*, *FATA* regulating fatty acid biosynthesis and *LPAT*, *GPAT*, *DGAT* from triacylglycerol biosynthetic pathway are hereby suggested for genetic interventions to increase desired fatty acid composition and overall oil content in Jatropha.

5.3 Transcriptional regulation of oil biosynthesis and accumulation in *J. curcas*

Transcriptional regulation of oil biosynthesis and accumulation has been not studied in *J. curcas* till date. In the present study, the regulatory elements and transcription factors governing oil biosynthesis and accumulation were identified in the genes associated with overall oil biosynthesis.

5.3.1 Regulatory elements in the promoter regions of oil biosynthesis genes

To date, very little information is available on identification of regulatory elements in oil biosynthesis genes in Jatropha. The biosynthetic pathways for fatty acids and TAGs are regulated at the level of transcription [65, 226]. Variation in oil accumulation in high and low oil content genotypes and even in endosperm and embryo tissues of seed in Jatropha could be primarily because of the differences in cis-regulatory elements in the promoters of highly expressed genes of FA and TAG biosynthetic pathway. Computational analysis of the promoter regions revealed the presence of oil deposition specific transcription factor binding elements like Dof, CBF (LEC1), SORLIP, GATA and Skn-1_motif along with other common promoter elements in the genes identified in current study and reported in previous studies [11, 33]. *Dof* family of transcription factors have been found to be associated with the enhancement of overall oil content in soybean as well as in algal systems like *Chlorella* and *Chlamydomonas* [51, 69, 227]. CBF (CCAAT-box binding factor, motif binds to LEC1) is a transcription factor class having the most studied and characterized factor, LEC1, in oil plants with major contribution in fatty acid and oil accumulation. Increase in fatty acid level and oil content was achieved by over expression of LEC1 factor in Arabidopsis and maize [65, 66]. SORLIP (Sequences over-represented in light induced promoters) are generally residing in promoter regions of fatty acid biosynthesis genes and are associated with the seed storage accumulation [210]. GATA is

a common element confined to the promoter regions and also linked with the fatty acid biosynthesis and accumulation [74, 210]. Skn-1_motif (GTCAT) is an element required for the endosperm expression, which is prerequisite for seed reserve and oil accumulation [211, 228]. FA and TAG biosynthetic pathway genes showed common regulation as these were further categorized on the basis of presence of common specific elements in the promoter regions to shortlist the potential target genes. Genes *KASI*, *KASII*, *LPAT*, *DGAT*, *CLK*, *KCR2*, *Lipase*, *OAD*, *FATA* from category I and genes *Oleosin1*, *Oleosin2*, *PDAT*, *DGK1*, *ECH*, *KCS*, *LACS8*, *SD* and *SAD* from category II can be targeted primarily for enhancement of seed oil content and fatty acid composition in Jatropha. It is inferred that all the five common elements (Dof, CBF (LEC1), SORLIP, GATA and Skn-1_motif) might be playing a role in governing transcriptional regulation of oil biosynthesis in Jatropha as the genes from cluster I, II (fatty acid biosynthesis) and cluster III (triacylglycerol biosynthesis) showed their presence in the promoter regions. Further, the experimental validation through cloning of the promoter region of *SAD* gene, the rate limiting gene associated with FA and TAG biosynthesis, confirmed the role of these specific regulatory elements in oil accumulation as they were more in number in the promoter region of high oil content genotype as compared to low oil content genotype. These identified regulatory elements can be targeted to distinguish high and low oil content genotypes and to develop high oil content lines of *J. curcas* as these are confined to the genes showing higher abundance in high oil content genotypes.

5.3.2 Transcription factors regulating oil biosynthesis and accumulation in *J. curcas*

Transcription factors (TFs) are proteins which along with other transcriptional regulators, activate or inhibit RNA polymerases to the DNA template [229]. Many cellular responses are being facilitated by transcription factors (TFs) by identifying specific cis-regulatory DNA sequences at the promoter region of targets genes. Transcription factors regulate various processes like synthesis of metabolites, abiotic and biotic stresses, lipid biosynthesis and accumulation, adaptation to environment, disease resistance, floral regulation, etc. and thus are involved in overall growth and development of plants. Transcription factors governing overall lipid biosynthesis and accumulation have been identified and characterized for many oil plants.

Transcription Factor families i.e. *Dof*, *MYB*, *bZIP*, *bHLH*, *CBF* and *AP2* regulating oil biosynthesis and accumulation were identified on the basis of computational expression

analysis i.e. significant FPKM values. Higher expression in mature stage of endosperm was observed as compared to ripened stage of endosperm in both high and low oil content genotypes, as the former has been reported for more oil accumulation [11, 33]. RT-qPCR based transcript abundance pattern of TFs (*bZIP*, *Dof*, *MYB*, *bHLH*, *CBF* and *AP2*) showed positive correlation with oil accumulation at R and M stages of endosperm in high oil content genotype which suggested their role in regulating and controlling the oil biosynthesis and accumulation in *J. curcas* [47, 66]. Further, validation and characterization of these identified transcription factors needs to be done so that genes regulating oil biosynthesis can be manipulated to enhance oil production in Jatropha. These observations are the initial leads towards transcription regulation of oil biosynthesis and accumulation in *J. curcas* and will further aid in better understanding of molecular basis of oil formation in *J. curcas*.

5.4 Understanding molecular mechanisms associated with mosaic disease in *J. curcas*

Off late mosaic disease caused by *Jatropha curcas* mosaic virus (*Jc*MV) has become prevalent in Jatropha plantations in India. This disease has become a major concern as it is reducing overall seed yield and also oil content in *J. curcas*. Therefore, a comparative transcriptomic analysis was performed between healthy and mosaic disease affected plants to get insights into molecular mechanisms associated with virus infection response in *J. curcas*.

5.4.1 Reduction in seed yield and oil content due to mosaic disease in *J. curcas*

Various biotic stresses confined to plants affect processes associated with general growth and development. The reduction in seed yield and oil content has been reported for many oil plants like sunflower, soybean, Brassica spp, maize in response to viral infections [230, 231, 232, 233]. Data pertaining to seed yield and oil content parameters was recorded consecutively for two years in mosaic virus infected (JV) and healthy (JH) Jatropha plants. It was observed that different parameters related to seed yield i.e. seeds per fruit, number of seeds per plant and weight of seeds per plant showed overall reduction in plants infected with virus as compared to healthy ones. On oil content estimation there was reduction of 5-6% in the total oil content of virus infected plants in comparison to healthy plants. These results are in line with previous studies for effect of virus infection on yield and oil content in *J. curcas* [112, 113]. These observations suggested that mosaic disease in *J. curcas* is a major biotic stress associated with overall yield reduction.

5.4.2 Gene ontology based functional annotation

For functional annotation, gene ontology (GO) analysis was performed which revealed association of identified genes with terms, biological process, molecular function and cellular component accompanying disease response mechanisms. The analysis revealed that major processes activated during viral infection are response to stress, catabolic process, transport, biosynthetic process, immune system process, signal transduction, and cellular protein modification process whereas processes like photosynthesis, small molecule metabolic process, carbohydrate metabolic process and reproduction were repressed. Observations of gene ontology analysis are consistent with the supposition that biotic stresses in plants mark a change from growth and reproduction to physiological and metabolic processes designed for defense related responses [234].

5.4.3 Enhanced energy metabolism during viral infection in *J. curcas*

Upon functional annotation via pathway mapping with KEGG, it was observed that the metabolism processes are affected the most in response to viral infection as most of the genes related to metabolism processes were upregulated or downregulated. In the current study, it was found that energy metabolism (oxidative phosphorylation) was upregulated in response to viral infection. In photosynthesis, light energy is captured and converted into ATP and NADPH. These ATP and NADPH act as energy sources for various physiological processes. Up regulation of genes such as NAD(P)H-quinone oxidoreductase, ubiquinol-cytochrome c reductase, NAD(P)H-quinone oxidoreductase, F-type H+-transporting ATPase, NADH dehydrogenase, cytochrome c oxidase during virus infection suggested that organelles like mitochondria produce energy to drive cellular processes. These results are in accordance to previous reports for virus infection in rice, tobacco, etc. [121, 235]. However, the present observations are deviated from the assumption which implied termination of processes for the production of plant energy in response to disease and infection [236].

5.4.4 Endocytosis is activated in response to viral infection in *J. curcas*

Endocytosis was found to be significantly enriched in response to viral infection. Endocytosis is a cellular process in which cells internalize extracellular material or foreign particles for recycling or degradation [237]. During virus infection, the host cells also destroy pathogens by engulfing them [238] which has been evidenced by the upregulation

of genes involved in endocytosis in current study. Genes regulating endocytosis such as charged multivesicular body protein 5, Ras-related protein Rab-11A, epsin and DnaJ homolog subfamily C member 6 were upregulated in response to mosaic viral infection [239, 240, 241].

5.4.5 Metabolism of amino acids and vitamins is induced in response to viral infection

Synthesis of amino acids i.e. arginine and proline was induced in response to viral infection as genes linked to 'Arginine and proline metabolism' were significantly over expressed. On exposure to specific infection or pathogen, plants produce reactive oxygen species (ROS) to programmed cell death and to terminate the disease process [242]. Proline and arginine act as potential scavengers of ROS and thus prevent the function of ROS. The genes involved in the synthesis of these amino acids were upregulated in response to mosaic virus infection which indicates potential role of these amino acids in infection. In response to viral infection, genes nitric-oxide synthase and prolyl 4-hydroxylase are significantly overexpressed. Nitric-oxide synthase has been reported to catalyze the production of arginine whereas prolyl 4-hydroxylase has been associated to synthesize proline [243, 244]. The genes related to 'Arginine and proline metabolism' have been previously shown to be linked with biotic and abiotic stresses in various plant species like Arabidopsis to tobacco etch virus infection, cotton to aphid infestation and eucalyptus to cold stress [245, 246, 247]. Biosynthesis of vitamins such as ascorbic acid was also induced in response to mosaic viral infection as evidenced by the upregulation of genes involved in 'Ascorbate metabolism'. Deficiency of ascorbic acid leads to the activation of cell death and disease resistance response in plants [248]. Genes involved in biosynthesis of ascorbic acid such as GDP-L-galactose phosphorylase and L-ascorbate peroxidase were significantly over expressed in response to mosaic virus infection suggesting potential role of ascorbic acid in disease induction in plants.

5.4.6 Catabolism of fatty acids and lipids is associated to sugar biosynthesis in response to viral infection

Lipids and fatty acids regulate plant defense response against various pathogens as they act as signaling molecules [249, 250]. Upregulation for fatty acids and lipids catabolism upon mosaic viral infection was observed as the genes involved in 'Lipid metabolism' and 'Fatty acid metabolism' showed higher transcript abundance. These results are supported by the fact that the *Jatropha curcas* mosaic virus disease (*Jc*MD) reduces the overall oil

content in *J. curcas* [19]. Genes involved in fatty acid catabolism such as acetyl-CoA acyltransferase, long-chain acyl-CoA synthetase, acyl-carrier-protein desaturase and in lipid catabolism i.e. lipase were significantly upregulated in response to mosaic viral infection. These results are in line with previous report by Freitas-Astúa et al. [251] suggesting that fatty acid and lipid metabolism is important for the susceptibility of virus infection and diseases. In plants, starch (sugar) is accumulated during the day and used in dark hours to provide energy for key metabolic processes [252]. There is a reciprocal relationship between sugar biosynthesis and oil biosynthesis [253]. Genes regulating sugar metabolism were found to be upregulated in response to mosaic virus infection where UDP-apiose/xylose synthase and L-arabinokinase were mainly over expressed. Both UDP-apiose/xylose synthase and L-arabinokinase are involved in sugar biosynthesis [254, 255].

5.4.7 Terpenoids function as plant growth regulators during viral infection

Apart from primary metabolites, various secondary metabolites also got affected during virus infection. In plant kingdom, terpenoids function in defense mechanisms against a broad range of pathogens [256]. Plants interact with pathogens through some signaling molecules such as terpenoid metabolites. It was observed that there is higher transcript abundance of diterpenoid biosynthesis genes such as gibberellin 2-oxidase and gibberellin 3-beta-dioxygenase in response to mosaic virus infection. These observations are in agreement with assumption that during pathogen infection, formation of some diterpenoid and sesquiterpene metabolites is induced as plant growth regulators (gibberellins) and phytoalexins [256, 257]. Genes encoding synthase enzymes catalyzing the formation of monoterpenoid, sesquiterpenoid and triterpenoid biosynthesis were upregulated in mosaic viral infection. The present observations are similar to what has been reported previously for virus infection in tobacco and cassava [121, 258] suggesting that biosynthesis of secondary metabolites is affected during mosaic disease in *J. curcas*.

5.4.8 Hormones signaling is enhanced during virus infection

During *Jatropha curcas* mosaic disease (*Jc*MD), various physiological abnormalities occurs such as leaves get curled, blistered and mottled [19]. Various phytohormones are present in basal amounts in plants and regulate plant growth and development. Any variations from normal levels of phytohormones due to infection with virus can alter physiological processes and morphology resulting in abnormal symptoms, as was

observed in this study. During the course of virus infection, some cytopathic effects occur which are supposed to alter the normal plant growth [259], which may be due to alterations in plant hormone metabolism. During mosaic virus infection, higher transcript abundance of genes i.e. SAUR family, auxin responsive GH3 gene and auxin-responsive protein IAA, which regulate signaling of auxins was observed. For auxins the current results supported the fact that substantial rise in activity of auxins sometimes causes severe symptomatology during infection [260]. Further genes related to cytokinin signaling were significantly upregulated during virus infection. Genes i.e. two-component response regulator ARR-B family and Arabidopsis histidine kinase 2/3/4 (cytokinin receptor) acts as positive regulator of cytokinin signaling, were upregulated in virus infection whereas genes from two-component response regulator ARR-A family, negative regulator of the cytokinin signaling were downregulated in response to infection. These observations supported the fact that cytokinins contribute to stress and pathogen responses in plants [261, 262]. Gibberellins have a negative role in plant defense which was supported by the current observations. Genes linked to family DELLA protein were downregulated during virus infection as this family of genes is intracellular repressor of gibberellin response [262]. Members of other gene family i.e. gibberellin receptor GID1 was upregulated which confirmed their role in positive regulation of gibberellin signaling during infection. Abscisic acid gets more accumulated during infestation with viruses [263]. The upregulation of genes such as abscisic acid receptor PYR/PYL family involved in the signaling of abscisic acid, in response to mosaic viral infection was observed. Ethylene-insensitive protein 3, a gene involved in the ethylene signaling was also overexpressed in response to mosaic virus infection suggesting possible role of ethylene signaling in virus accumulation and infection [264]. Another plant hormone, jasmonic acid was observed to be a negative regulator of infection and positive regulator of resistance against mosaic virus as evidenced by the upregulation of gene i.e. jasmonate ZIM domain-containing protein repressing the signaling of jasmonic acid [265]. RT-qPCR based experimental validation of the identified genes further confirmed the computational results. Thus, these results implied that signaling of various plant hormones such as auxins, cytokinins, gibberellins, abscisic acid (ABA) and ethylene is activated during viral infection.

5.4.9 Photosynthesis is affected during virus infection

Downregulation of major pathways related to overall growth and development in *J. curcas* due to mosaic virus infection was also observed. Photosynthesis, a major physiological

process was significantly repressed in response to mosaic virus infection in *J. curcas*. In photosynthesis, light energy is captured and converted into ATP and reducing power (NADPH). During mosaic virus infection, the overall chlorophyll in the leaves gets degraded due to the induction of chlorophyllase activity [122]. Also there is reduction in leaf area followed by chlorotic spots which also correlate with the degradation of chlorophyll. Protein complexes regulating photosynthesis have photosystems (PSI and PSII) as functional and structural units. In photosynthesis, these photosystems help to regulate the primary photochemistry of photosynthesis, absorption of light and energy transfer. PSI and PSII absorb photons of a wavelength of 700 nm and 680 nm, respectively. Electrons flow from PSII to PSI through cytochrome b6 (a membrane bound protein). Genes related to PSI, PSII and cytochrome b6 were downregulated which indicates less rate of photosynthesis during viral infection in *J. curcas*. Along with chlorophyll degradation, there is also deficiency of light harvesting complex (LHC) [122]. Light-harvesting complex (LHC) gathers light energy and transfer this energy to the photosynthetic reaction centers [266]. LHC is composed of LHC proteins that bind light harvesting pigments. During mosaic virus infection, less transcript abundance of light-harvesting complex I chlorophyll a/b binding protein and light-harvesting complex II chlorophyll a/b binding protein genes was observed. Also it was observed that downregulation of ferredoxin as this gene functions principally in photosynthesis. Electrons are transferred from photoreduced PS I to ferredoxin NADP(+) reductase by ferredoxin [267]. These results are in accordance with previous study on cassava where the genes related to chlorophyll degradation were upregulated and genes encoding the apoproteins in light-harvesting complex II were downregulated in response to African cassava mosaic virus [122]. Further repression of photosynthesis in virus infection condition was correlated with the reduction in fruit size, seed yield and oil content as per observations. ATP and NADPH, the photosynthesis products are utilized in CO_2 fixation that provides carbon skeletons for all cellular reactions [268]. Light reactions in photosynthetic reactions feed ATP, NADPH to carbon fixation. Since photosynthesis and carbon fixation are repressed there is less partitioning of carbon towards lipid accumulation which might be responsible for reduction in oil content during virus infection. The reduction in fruit size and seed yield is due to the fact that regulation of photosynthetic reactions is essential for the metabolic reactions. Further nitrogen can improve photosynthetic parameters, increase maximal photochemical efficiency and reduce fluorescent and non-photochemical quenching co-efficiency and as a result increase

fruit and seed yield with higher seed filling rate [269]. Also, solar radiation is a major factor which affects the uptake of nutrient solution and growth processes during photosynthesis.

5.4.10 Degradation of anthocyanin in viral infection

Anthocyanins (flavonoids) are water-soluble pigments which are synthesized in the cytosol and localized in the vacuoles. During mosaic virus infection, the leaves get curled, reduced in size and become chlorotic which lead to a significant degradation of pigmentation. Genes, anthocyanidin 3-O-glucoside 2'''-O-xylosyltransferase and anthocyanidin 3-O-glucoside 5-O-glucosyltransferase involved in anthocyanin biosynthesis showed less abundance supporting the degradation of anthocyanin during infection. These observations are in accordance with previous studies for tobacco and grapes [121, 270]. These results further supported the fact that anthocyanin also regulate defense response in plants [271].

5.4.11 Repression of defense mechanisms during viral infection

Further it was observed that various pathways related to defense response were significantly repressed during mosaic virus infection in *J. curcas*. Genes linked to 'Plant-pathogen interaction', a process related to basal defense response, showed more expression in healthy leaf tissue as compared to mosaic virus infected leaf tissue which directed its repression during infection. The interaction between plants and pathogens is a major factor towards understanding overall mechanisms associated with defense response. Various genes in the 'Plant-pathogen interaction' pathway were down-regulated during mosaic virus infection. Genes such as cyclic nucleotide gated ion channels (CNGCs), calcium-binding proteins (CML), disease resistance proteins i.e. *RPM1*, *RPS2*, *Kinases* and *WRKY* transcription factor genes were significantly repressed in viral infection, which have been implicated in defense response and innate immunity in other plants [121]. Cyclic nucleotide gated ion channels (CNGCs) have been reported for their role in immunity and has been characterized in plant species [272, 273, 274]. Calcium-binding proteins (CML) are also involved in providing immunity against various biotic stresses [275, 276]. Role of various kinases in providing innate immunity has also been described in plants [277, 278, 279]. Also various reports exist for *WRKY* transcription factors providing innate immunity against biotic stresses [31, 163]. Genes involved in calcium

signaling pathway also showed low transcript abundance in response to infection clearly indicating the role of calcium signaling in basal defense response [280]. Various genetic and functional studies have shown the role of calcium signaling as positive regulator in the establishment of defense response. Also calcium signaling could be controlled by other signaling systems like ubiquitin-proteasome system to mark immunity in plants against pathogens [281].

5.4.12 Host factors contributing towards replication and multiplication of virus

Plant viruses manipulate and use metabolites of the compatible host for translation and replication of their genomes. In host cells, virus infection overexpress or repress various pathways, which cause physiological and phenotypic changes in the host [282, 283, 284]. Disease formation is successful completion of genome replication of the viruses [284]. In order to complete life cycle, viruses are evolutionarily able to capture and manipulate cellular pathways and cellular components. The genes regulating auxins signaling like SAUR protein and auxin responsive protein were induced and showed upregulation during virus infection supported that fact that geminiviruses replicate in the apical leaves by regulating auxin signaling pathway to create a favorable cellular environment for their replication [285]. In addition, current observations about upregulation of genes linked to auxin signaling also support the fact that auxin may stimulate virus entry into the S-phase, geminiviruses operate the core cell cycle genes to provide an environment for efficient replication [286]. Further on comparison of JV and JH derived transcriptomes, it was found that genes like histone 3 K4-specific methyltransferase and putative transcriptional activator with NAC domain were upregulated in response to viral infection as they interact with proteins of monopartite geminivirus, TrAP/C2 and C3, respectively and thus promote replication [285]. They are also associated with repression of systemic host defenses, facilitating systemic accumulation of virus [287]. TIR-NBS-LRR proteins associated with disease resistance were downregulated in JV and upregulated in JH which supports the fact that mosaic virus suppresses these genes associated to disease resistance in order to replicate and spread. More proportion of genes belonged to generation of energy process in JH as compared to JV. The upregulation of genes involved in energy metabolism was observed during virus infection. These results are in accordance with the fact that the energy produced by host is being used by viruses for polymerization of their proteins and n-RNA synthesis as nucleoside triphosphate. Further, more number of genes related to transport mechanism were upregulated in JV as compared to JH which support the

assumption that during local movement virus initially moves from one infected cell to adjoining cells and move through vascular tissue to cause a systemic infection in the plant [288]. On the basis of sub cellular location, it was observed that percent of genes associated to membrane part was higher in JV as compared to JH. These observations are in line with the fact that various viral encoded proteins are involved in membrane targeting of the replication components during replication of viruses [289, 290].

5.4.13 Identification of transcription factors regulating genes associated with biological processes

Transcriptional regulation is an integral component towards overall understanding of disease response in plants. Immunity regulation and response to stresses are generally driven by transcription factors. The current analysis indicated that along with genes, these regulatory proteins are also activated in response to virus infection. There are many reports describing the activation of various transcription factors in response to virus infection in plants [121, 123]. During virus infection, the upregulation and downregulation of transcription factors in 'Plant hormone signal transduction' and 'Plant-pathogen interaction, respectively was observed. Upregulation of transcription factors involved in signaling of phytohormones during virus infection was observed. TF families i.e. *MYC2*, *TGA*, *ABA responsive element binding factor* and *ethylene-responsive transcription factor* were associated with hormone signaling in response to virus infection. In 'Plant-pathogen interaction', transcription factors related to *WRKY* family were downregulated during infection as those are involved in defense response mechanism [31].

5.4.14 Identification of SNPs in JV and JH transcriptomes of *J. curcas*

SNP markers are of wide choice for their application as they are abundant in genome, ubiquitous, and amenable to high-throughput automation [291, 292]. In the present analysis, many SNPs were identified in the loci of genes upregulated and downregulated in response to mosaic virus infection. For example, the identified SNP markers associated to different resistance genes have the potential to be used in the breeding programs for developing resistant varieties in Jatropha [293, 294]. SNPs can occur in coding as well as noncoding regions of genes and might be responsible for consequences in gene transcription or function. These consequences are the biological reason for the association of SNPs with various agronomic traits. On comparing, JH and JV, it was found that more number of SNPs were present in the coding region in JH as compared to JV [295].

5.4.15 Identification of genes co-expressed with genes involved in 'Plant hormone signal transduction' and 'Plant-pathogen interaction'

Gene co-expression networks (GCNs) are graphic illustrations that represent the synchronized transcription of genes in response to a particular stimuli. Gene co-expression analysis revealed the presence of various genes which are co-expressed with the target or reference genes. Few genes showed close association or interaction with the identified reference gene. These genes can be primarily targeted with identified disease inducing or resistance genes for understanding the molecular responses to develop resistant genotypes in Jatropha. For example, it was observed that most of the co-expressed genes identified with defense response pathways are involved in secretory pathways which are responsible for providing immune response to plants [296]. It was inferred that the genes which are co-expressed with the identified reference genes indicated that there is some degree of conservation of their function. Similar approach of constructing gene co-expression networks in response to disease resistance and immune responses have been applied for immunity expression data of *A. thaliana*, rice, soybean, tomato and cassava which shed light on global patterns of events activated throughout plant immune responses [297]. Co-expression networks have also been developed for resistance genes identified from transcriptome data in a number of plant species [298, 299].

5.4.16 Experimental validation of the transcriptome data

For experimental validation, selected genes and transcription factors on the basis of more transcript abundance in virus infection from transcriptome data were examined using RT-qPCR approach. The genes from 'Plant hormone signal transduction' were chosen for experimental validation as most genes were associated to this process. Also the foremost symptoms during mosaic virus infection are linked with changes in metabolism of major plant hormones. Transcription factors showing upregulation and downregulation in response to virus infection were also studied using RT-qPCR which showed results similar to observed through FPKM expression.

Overall, the transcriptome based characterization and comparative analysis of healthy and virus-infected leaves lead to new dimensions in understanding the molecular perspective of plant-pathogen interaction in *J. curcas*.

5.5 Identification of disease resistance (NBS-LRR) genes in *J. curcas*

Jatropha curcas mosaic disease caused by mosaic virus cannot be directly controlled by the application of pesticides or chemicals. There are some alternative strategies for the effective control of this disease such as

a) Biological or chemical control of the vector of this disease i.e. whitefly (*Bemisia tabaci*)
b) Growing varieties with enhanced resistance
c) Use of virus free plant material
d) Exclusion (Prevention in those areas where disease has not yet appeared)

Out of the above listed alternative strategies, the development of *J. curcas* varieties with resistance to virus infection is the appropriate alternative that has to be undertaken for the effective management of *J. curcas* in field conditions. To achieve this goal, there is need to identify molecular components associated with disease resistance or defense response which can be used to develop disease resistant genotypes.

Since the previously predicted NBS-LRR genes in Jatropha are quite small in number in comparison to other sequenced plant genomes with the same range of genome sizes (For example, the genomes *of A. thaliana* and *V. vinifera* contain relatively higher number of NBS-LRR genes (ranging from 174 to 535), even though their genome sizes are in the order of 125 and 487 Mb, respectively [300]. 47 new NBS-LRR genes in *J. curcas* from publicly available transcriptome were identified, while earlier identification of NBS-LRR genes was done through genome mining which may contain pseudogenes [13]. The identified NBS-LRR genes were mapped on to the genome sequence contigs to have a clue about their physical location [301]. Some of the NBS-LRR genes can be frequently clustered in the genome due to segmental and tandem duplication [157, 302]. Consistent with these findings, presence of NBS-LRR genes in clusters have been identified. 16 genes were found in clusters of two genes. The current results indicate that there is more clustering in case of Jatropha which may support the concept of novel resistance specificities through recombination or gene conversion and also rapid R gene evolution in Jatropha [32]. Moreover the NBS-LRR genes present in clusters can be primarily targeted for breeding to develop disease resistant varieties. Several NBS-LRR genes were mapped with gaps which represent the presence of intronic region in these genes and is in consonance with the fact that most of the eukaryotic genes comprised of introns [303].

Further these intronic regions can be explored in spliced site studies for disease [304, 305, 306].

5.5.1 Characterization of identified NBS-LRR genes into TNLs and CNLs

The identified NBS-LRR genes were further characterized into TNLs and CNLs and the number of TNLs were more compared to CNLs, as the TNLs were confined only to dicots [159] which further supports the motifs prediction as *J. curcas* is a dicotyledonous species. Further, these N terminal domains i.e. TIR (TNLs) and CC (CNLs) were responsible for pathogen recognition which supports the resistance potential of the associated genes [150].

5.5.2 Identification of transcription factors related to defense response in *J. curcas*

From transcriptome mining, a total of 122 transcription factors related to defense response in *J. curcas* were identified. These investigations are the first attempt to identify transcription factors related to disease resistance or defense response in *J. curcas*. Many of the transcription factors have been implicated in maintaining transcriptional reprogramming linked with plant defense and resistance response. An association among activating and repressing transcription factors from many families control the defense response expression of the target genes [163]. Transcription factors such as *WRKY*, *bZIP*, *ERF*, *MYB* and *Whirly* families bind to the promoters of resistance genes and regulate expression level [31, 162, 165, 166, 307]. In comparison to conventional screening of cDNA libraries or EST sequencing, the computational transcription factors discovery approach provides fast, simple, consistent and precise methods to reveal the transcription factor families specific to disease resistance and defense response at both the whole genome and transcriptome levels.

In the past, many transgenic crop and model plants with improved disease resistance have been developed [308] by over expressing the defense related transcription factors. Over expression of *WRKY* and *ERF* transcription factors have resulted in developing disease resistant varieties of many plants [174]. Over expression of the defense associated transcription factors can provide resistance to many dissimilar pathogens also. Arabidopsis transcription factor *MYB30* over expression has resulted in enhanced resistance to pathogenic bacteria and fungus in transgenic Arabidopsis and tobacco [309]. Identification of transcription factors related to defense response or disease resistance is of great significance in predicting the pathogen responsive promoter elements. Only a few pathogen responsive elements in the promoter regions have been identified. One of the

most cited example is the presence of W-box in the promoter region of various genes activated by *WRKY* transcription factors [310, 311, 312]. Out of identified transcription factors, some showed the higher transcript abundance which signifies their role as potential targets for achieving or providing disease resistance. However, these observations need to be fully validated through functional analysis.

5.5.3 Distribution of defense response related transcription factors into families

After the identification of transcription factors related to defense response, all transcription factors were checked for their family distribution on the basis of similarity search with BLASTn analysis. The identified 122 transcription factors were further distributed into families and it was observed that the majority of identified transcription factors belong to *NAM*, *WRKY*, *MYB* and *Homeo-domain* families followed by families like *ERF-type/AP2-EREBP*, *bZIP*, *TFIIA*, *CBF*, *SBP* and *Whirly*. The major TF families identified here (i.e. *NAM*, *WRKY*, *MYB* and *Homeo-domain*) have been previously characterized and validated for providing resistance to various biotic stresses in many plant species [166, 307, 313, 314, 315].

5.5.4 Comparative analysis between Jatropha and castor bean identifies potential NBS-LRR genes and transcription factors related to defense response

A comparative analysis between Jatropha and its closest family member, castor bean showed common and unique NBS-LRR genes because these genes have been successfully used in developing disease resistant plants within same family. It was found that 7 and 8 NBS-LRR genes were common between Jatropha and castor bean, respectively. In castor bean, out of 8 genes, 6 showed identity to each specific gene from Jatropha whereas 2 genes from castor bean showed similarity to 1 common gene from Jatropha. The results are in line with the previous analysis of NBS homolgues in two important members of family solanaceae, tomato and potato which suggested the conservation of synteny supporting the fact that these have earlier origin. Further, all syntenic tomato and potato loci confer resistance to dissimilar disease, suggested different pathogen specialization of NBS-LRR resistance genes [316]. Common transcripts/genes can be targeted in a cross generic or cross specific manner for enhancing the disease resistance potential of Jatropha and castor bean [32, 70, 269]. The common genes identified from both organisms implies that these are conserved in nature and may be responsible for providing resistance to general disease conditions not specific to any particular pathogen. The over-expression of

an NBS-LRR gene, *Bs2* from pepper conferred increased resistance against bacterial spot disease in transgenic tomato [153]. The transcript abundancy was measured for newly identified set of NBS-LRR genes with the help of *in-silico* expression analysis in order to support the identification of transcripts and their expression levels. A high variation was found in the expression values of identified genes. The genes showing higher values of expression with more transcript abundance can be used to design and conduct experiments for providing enhanced resistance to disease and pest conditions in *J. curcas* and other economically important plants of same family such as castor bean, rubber tree, cassava, etc. [32, 269]. Further, comparison between identified transcription factors was made between Jatropha and castor bean in order to identify common and unique numbers of transcription factors which showed that a large number of transcription factors [305] are common between Jatropha and castor bean, thereby suggesting a conserved defense response mechanism in Jatropha and castor bean [317, 318, 319].

Comparative study of varying expression profiles or variations in transcript abundance measurements of NBS-LRR genes and transcription factors associated to disease resistance between Jatropha and castor bean transcriptomes revealed that some NBS-LRR genes and transcription factors can be good candidates for enhancing the resistance potential. By using comparative analysis, the exploration of evolutionary fate of the NBS-LRR genes and transcription factors in the euphorbiaceae family and the understanding of disease resistance between the important family members is anticipated [320].

5.5.5 Comparative analysis between Jatropha and castor bean revealed the concept of duplication and synteny

The comparative analysis of previously identified NBS-LRR genes in Jatropha [13] and castor bean [321] revealed that 7 of the disease resistance genes present in castor bean genome showed similarity to Jatropha genome, signifying that these genes emerged from the recent duplication or have been conserved devoid of significant divergence, as was found for NBS-LRR genes and RGAs in sweet potato and Arabidopsis earlier [322, 323]. Furthermore, 60 % gene clustering was observed in both these plant species and the genes which were present in clusters consisted of same domains and motifs. Similar kind of motif patterns were observed in both these plants which also corroborates the concept of synteny [324], but certain differences with respect to the presence of conserved domains were also observed between two plant species, which included presence of dirigent

domain/superfamily along with protein kinase domain in castor bean genome, and RPW8 domain/superfamily in Jatropha genome.

5.5.6 Characterization of NBS-LRR genes predicted by Sato et al. [13] in *J. curcas*

A total of 92 NBS-LRR disease resistance genes were predicted in Jatropha genome by Sato et al. [13]. Out of 92 genes, 54 genes have been predicted as TNLs and 28 as CNLs whereas 10 genes did not fall into either class. Since the CNLs and TNLs are both involved in pathogen recognition [150], the prediction and classification of NBS-LRR proteins into CNLs and TNLs further support the disease resistance potential. The presence of TNLs is known exclusively only for dicots not for monocots [325] which further support the motifs prediction as Jatropha is a dicotyledonous species. These results are in accordance with the previous classification of TNLs and CNLs for the identified NBS-LRR genes. Also the NBS-LRR genes represent ~0.3 % of all predicted ORFs.

Based on overall observations it is anticipated that NBS-LRR genes, the defense related transcription factors predicted in this study and domain architecture of previously identified NBS-LRR genes will supplement the disease resistance knowledge pool in *J. curcas* so that better breeding and genomics-based interventions can be made for developing disease resistant varieties. Further, the *in-silico* based analysis and comparison of NBS-LRR genes and transcription factors between Jatropha and castor bean will reveal specific insights on the function, organization, conservation and evolution of the NBS-LRR resistance genes and defense response related transcription factors in related members of family euphorbiaceae. However, the current observations need to be fully authenticated through functional validation analysis of identified NBS-LRR genes and defense related transcription factors in context of enhanced resistance to various biotic stresses in *J. curcas*. Further, the identified NBS-LRR genes and transcription factors have been made publicly available over the internet to be used by scientific community.

SUMMARY

Current study has undertaken investigations on genomics-based understanding of seed oil biosynthesis, disease response and disease resistance components in a biodiesel plant, *J. curcas*. Due to economic importance of *J. curcas* as a potential biodiesel plant, there was need to look into genetic differences contributing towards variation in oil content between different genotypes and identifying key genes associated with oil biosynthesis. Also understanding of molecular mechanisms underlying response to a viral disease and repertoire of NBS-LRR genes and defense related transcription factors was carried out due to vulnerability of Jatropha to various biotic stresses, which reduce seed yield and oil content.

Oil content estimation among genotypes indicated that there is significant variation for oil content in two genotypes i.e. 30% (low oil content) and 42% (high oil content) in IC 561227 and IC 561235, respectively. As of today the expression of FA and TAG biosynthetic pathway genes was studied and reported at different seed developmental stages but not between genotypes varying in seed oil content. Genes *ACCase, KASI, KASII, KASIII*, from cluster I encoding enzymes involved in the formation of common intermediates in FA biosynthesis pathway and genes *SAD, OAD, FATA* from cluster II encoding enzymes catalyzing formation of specific fatty acids and their precursors showed higher expression (>10 fold) in high oil content genotype suggesting their role in more oil content. From cluster III, genes *DGAT, LPAT* and *GPAT* encoding enzymes catalyzing formation of triacylglycerols showed higher expression (>10 fold) in high oil content genotype, implying role of these genes also in contribution towards more oil content. Three genes of rate limiting steps i.e. *ACCase, SAD* and *DGAT* showed higher expression in oil accumulating stages of seeds collected from plants grown at low altitude suggesting that oil content increases with decrease in altitude. As endosperm has more oil accumulation (65-70%) as compared to embryo (8-10%), each gene of FA and TAG biosynthetic pathway showed considerable fold increase in oil accumulating stages for endosperm to embryo ratio. Further on comparing expression among endosperms of high and low oil content genotypes, it was observed that except *OCD*, all genes showed more transcript abundance in high oil content genotypes. Same observations were inferred for expression comparison among embryos concluding role of these genes for more oil accumulation in high oil content genotypes. PCA analysis proved useful in correlating FA and TAG biosynthetic pathway genes and oil accumulating ripened and mature stages of endosperm and embryo. The current study provided a repertoire of key genes encoding enzymes of

FA and TAG biosynthesis which can be further considered as suitable candidates either for engineering oil biosynthetic machinery and/or developing gene markers for seed oil enhancement. *In-silico* analysis of promoter regions of oil biosynthesis genes revealed the presence of putative regulatory elements associated with oil biosynthesis such as Dof, CBF (LEC1), SORLIP, GATA, Skn-1_motif etc. Transcription factors regulating oil biosynthesis i.e *bZIP*, *Dof*, *MYB*, *bHLH*, *CBF* and *AP2* were identified and showed positive correlation with mature stages of high oil content genotype. Overall, these findings together with the previous information provided a more comprehensive understanding of oil biosynthesis mechanisms in *J. curcas*.

Comparative transcriptome-based analysis of healthy versus virus-infected leaves provided details on molecular components associated with plant-pathogen interaction in Jatropha. The current study revealed that mosaic virus symptom expression in *J. curcas* is a complex process due to up-regulation and down-regulation of genes and transcription factors belonging to different biological processes. GO and KEGG based functional annotation showed that processes like metabolism processes, oxidative phosphorylation, endocytosis, terpenoid biosynthesis, hormone signal transduction were upregulated whereas photosynthesis, anthocyanin biosynthesis, plant-pathogen interaction and calcium signaling were downregulated in response to virus infection. Co-expression network analysis identified genes which were co-expressed with reference genes and associated with processes such as 'Plant hormone signal transduction' and 'Plant-pathogen interaction'. RT-qPCR analysis further validated the association of genes with 'Plant hormone signal transduction'.

The current study also led to the identification of 47 new, previously unidentified NBS-LRR genes, in addition to 92 genes, and 122 defense response related transcription factors which provided a rich resource of genes associated with disease resistance in *J. curcas*. The identified NBS-LRR genes were further characterized into TNLs and CNLs which showed that more number of genes belonged to TNLs class. The identified defense response-related transcription factors were distributed into families such as *NAM*, *WRKY*, *MYB*, *Homeo-domain*, *ERF-type/AP2-EREBP* and *bZIP*. This study also supplemented the comparative analysis of identified NBS-LRR genes and transcription factors with its closest family member, castor bean to identify common and unique genes and transcription factors. The sequences of identified NBS-LRR genes and transcription factors has been made publicly available (http://sites.google.com/site/combiogroup/datadownload).

FUTURE PROSPECTS

- ✓ It is anticipated that the identified key genes associated with fatty acid (FA) and triacylglycerol (TAG) biosynthesis and transcription factors regulating oil biosynthesis and accumulation can be used in genetic improvement of oil content in *J. curcas*.
- ✓ Identified NBS-LRR genes and transcription factors related to defense response can be further genetically analyzed to associate with resistance phenotypes in any disease resistant variety or species of Jatropha.
- ✓ Sequences of key genes of fatty acid (FA) and triacylglycerol (TAG) biosynthesis can be used to scan for SNPs for developing gene markers associated with high seed oil content in *J. curcas*
- ✓ Functional validation of genes associated with molecular responses to viral infection can help in understanding which particular biological processes are severely affected in response to viral infection.

APPENDICES

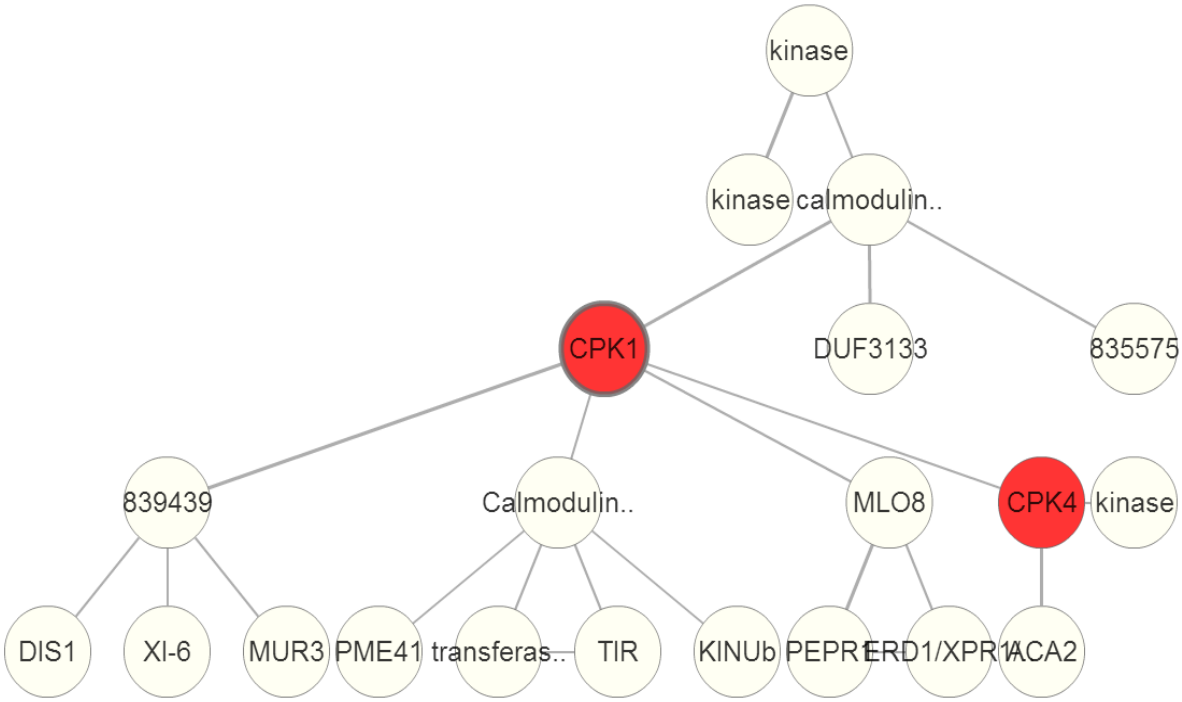


Figure A1 Co-expression network analysis of gene *CPK1* in 'Plant pathogen interaction'

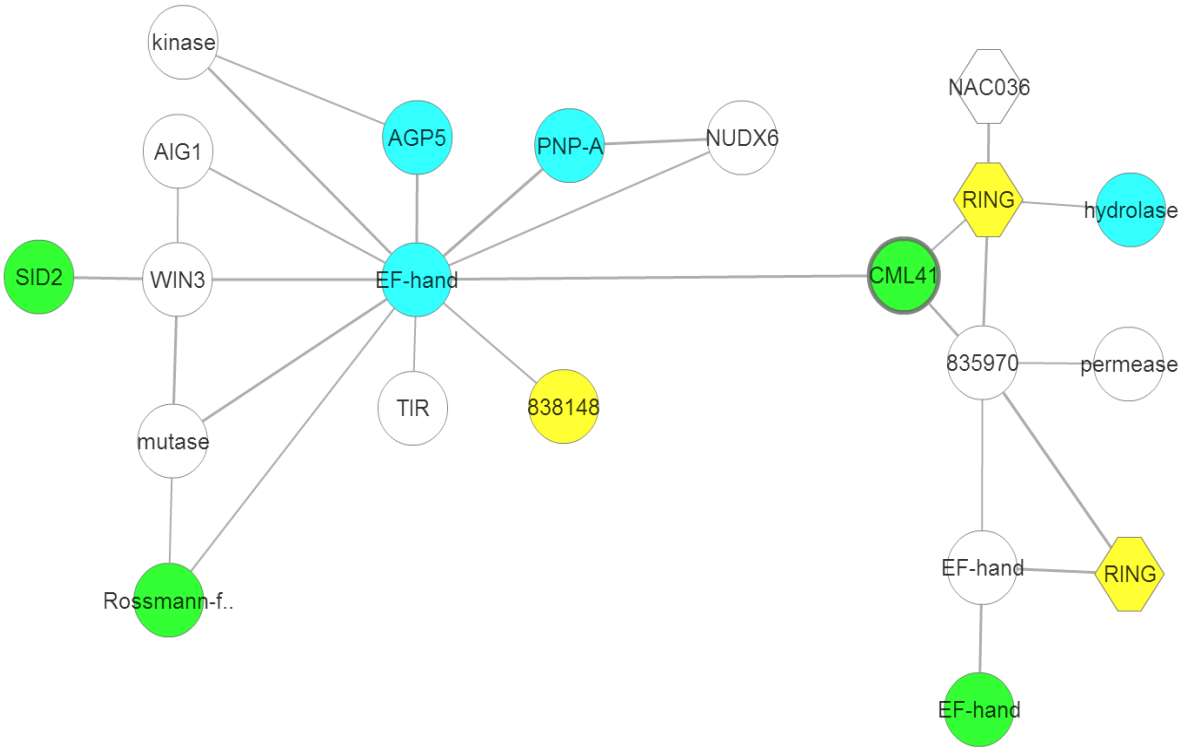


Figure A2 Co-expression network analysis of gene *CML41* in 'Plant pathogen interaction'

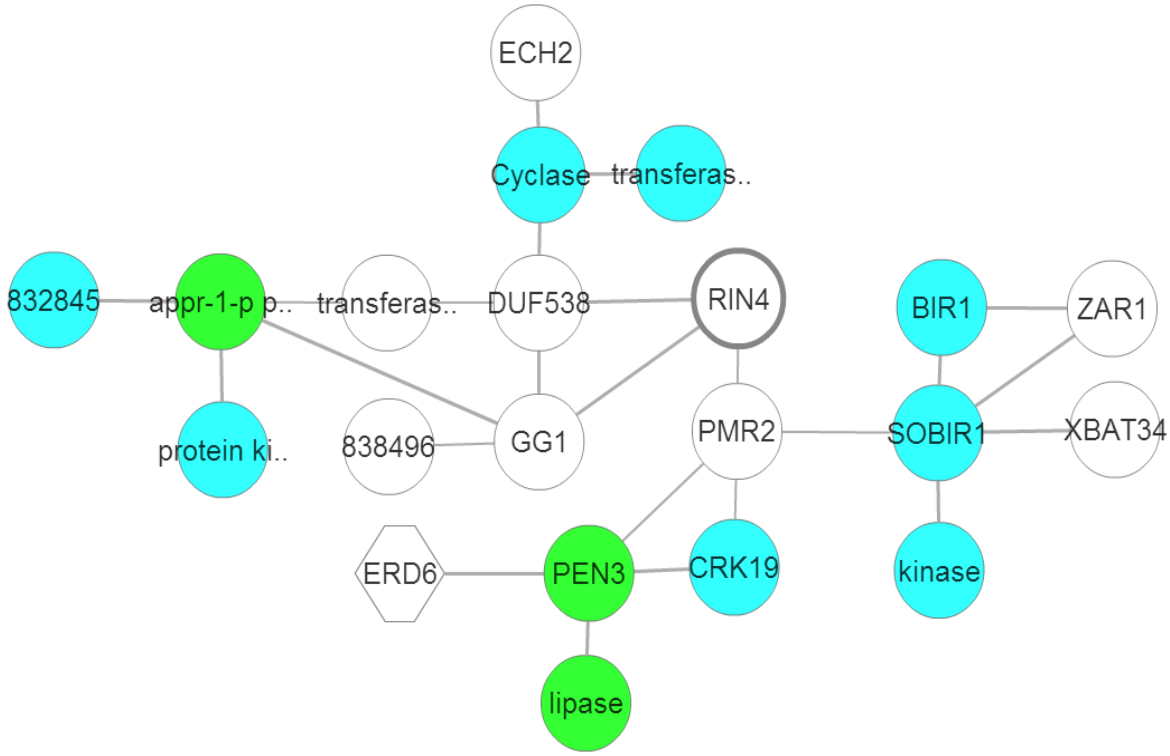


Figure A3 Co-expression network analysis of gene *RIN4* in 'Plant pathogen interaction'

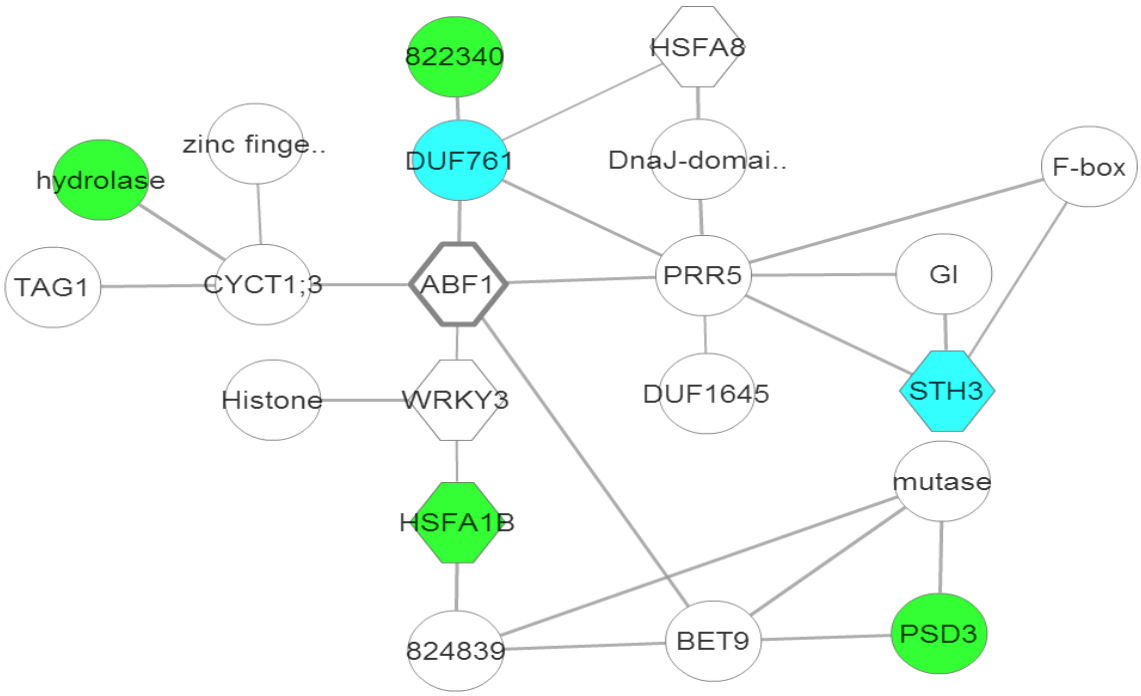


Figure A4 Co-expression network analysis of gene *ABF1* in 'Plant pathogen interaction'

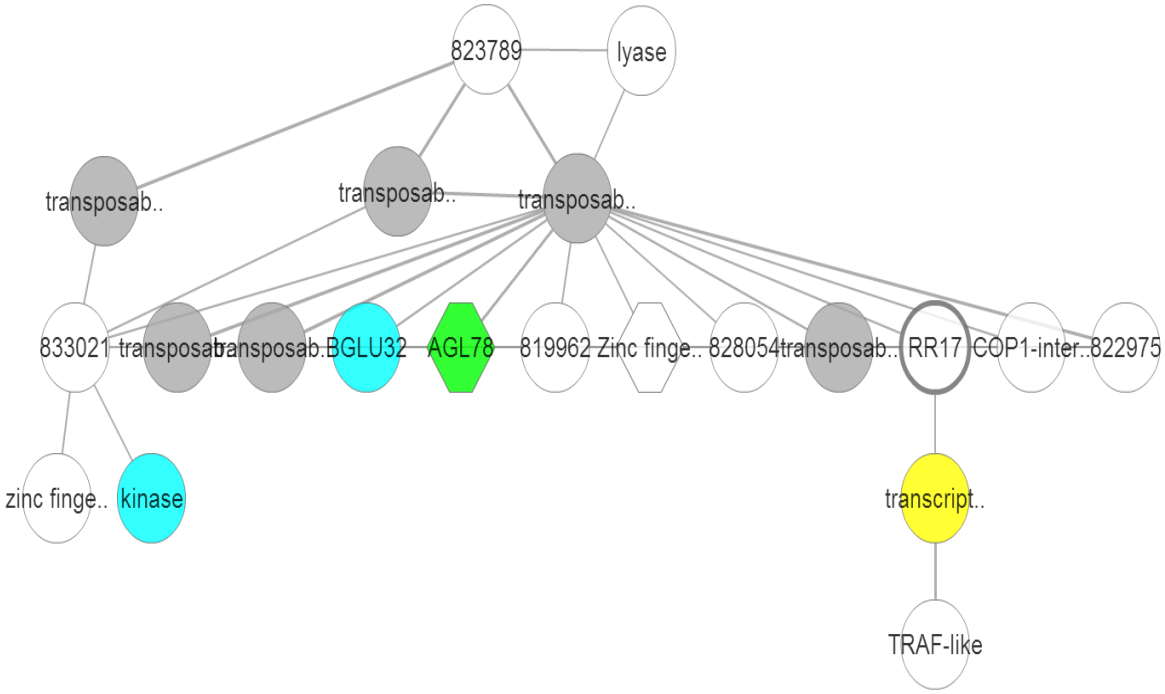


Figure A5 Co-expression network analysis of gene *RR17* in 'Plant hormone signal transduction'

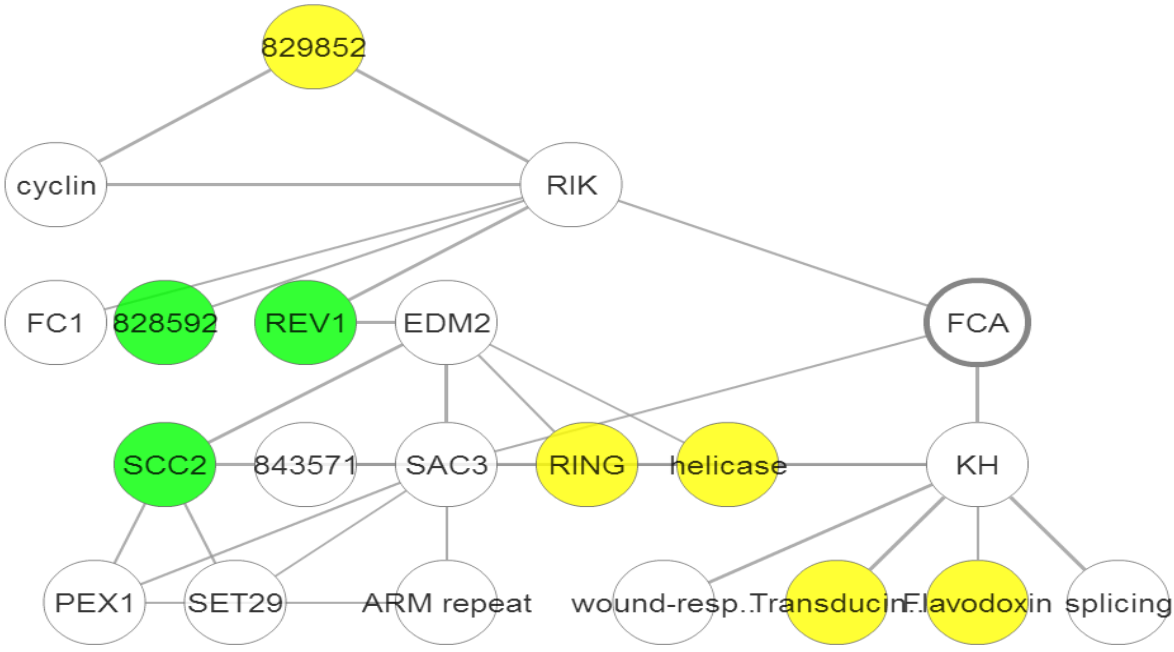


Figure A6 Co-expression network analysis of gene *FCA* in 'Plant hormone signal transduction'

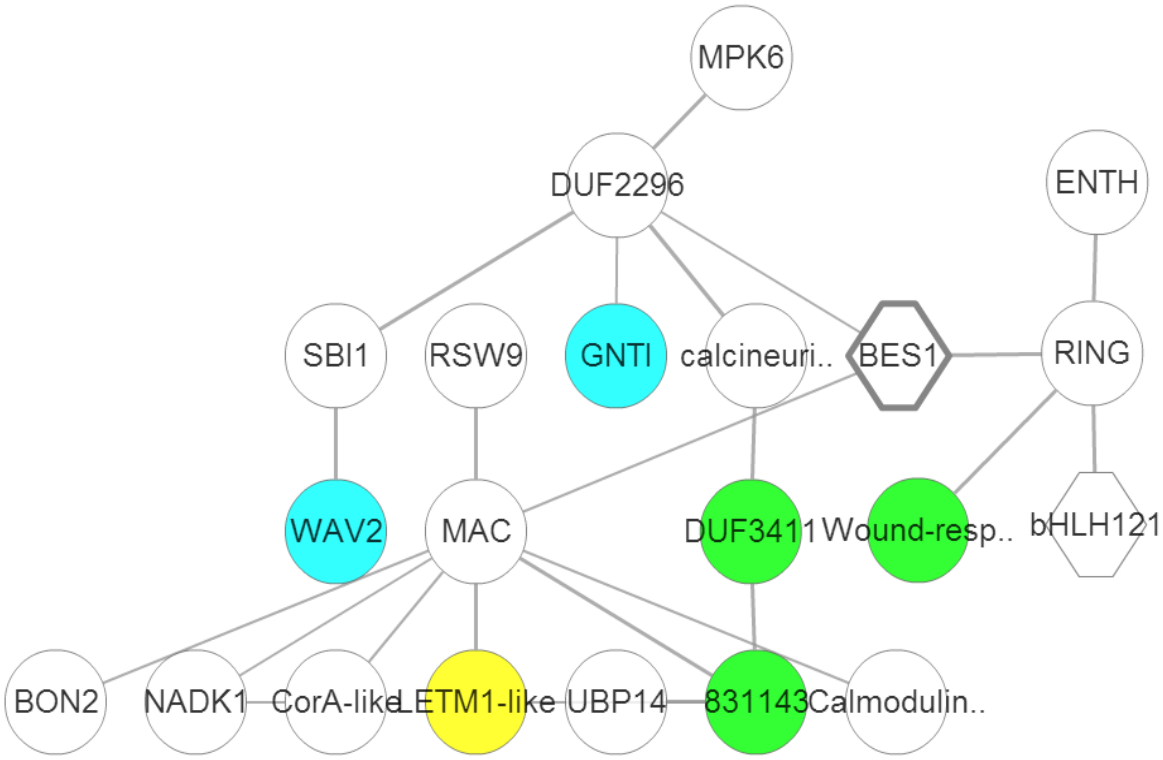


Figure A7 Co-expression network analysis of gene *BES1* in 'Plant hormone signal transduction'

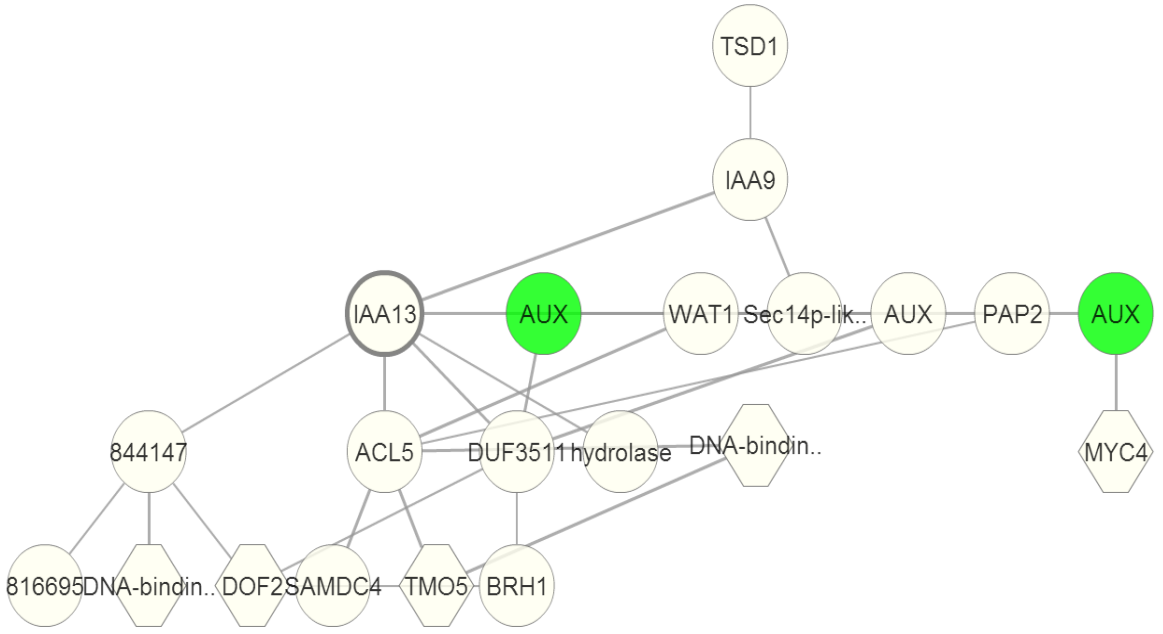


Figure A8 Co-expression network analysis of gene *IAA13* in 'Plant hormone signal transduction'

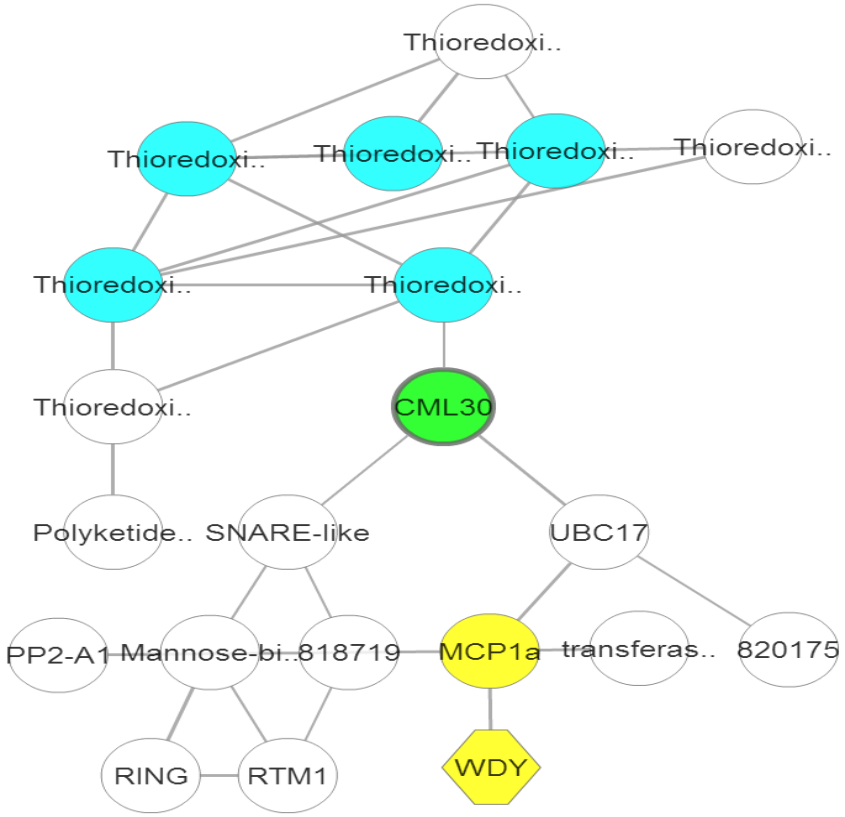


Figure A9 Co-expression network analysis of gene *CML30* in 'Plant hormone signal transduction'

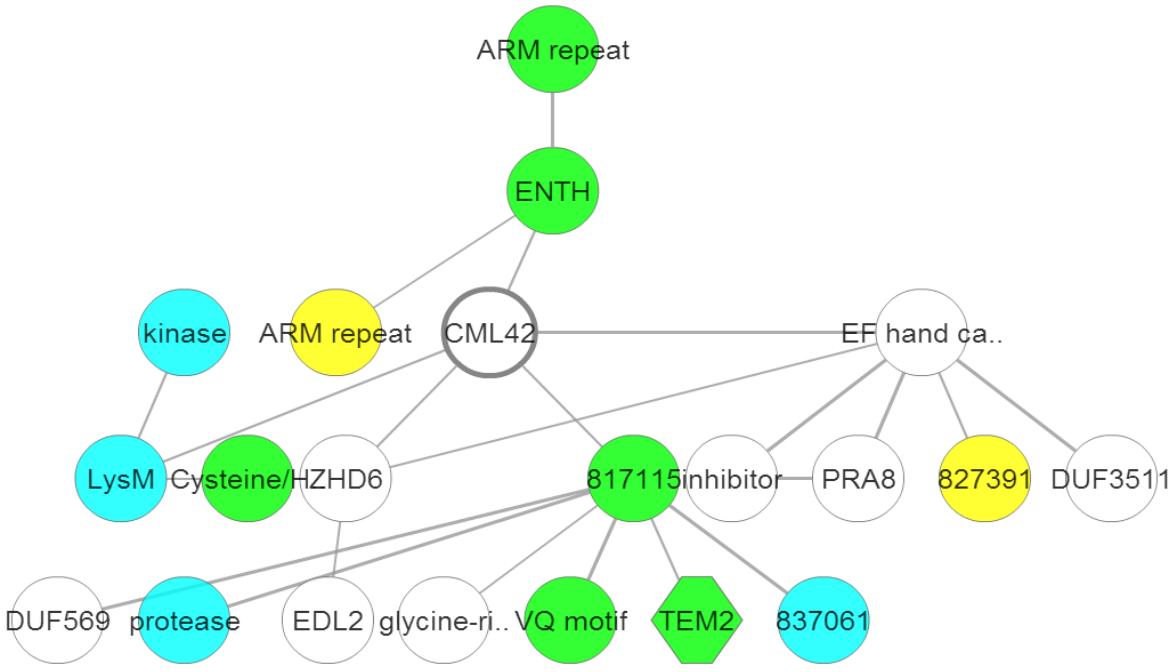


Figure A10 Co-expression network analysis of gene *CML42* in 'Plant hormone signal transduction'

Table A1 Genes associated to 'Oxidative phosphorylation' upregulated in JV

Jatropha Scaffold Ids	CDS Start Position	CDS End Position	KO IDS	Enzyme Name	JH FPKM	JV FPKM	log2 Fold Change
Jcr4S01020	6283	13142	K00411 UQCRFS1, RIP1, petA	ubiquinol-cytochrome c reductase iron-sulfur subunit [EC:1.10.2.2]	162.27	241.7	0.57
Jcr4S10290	4401	4754	K02134 ATPeF1D, ATP5D, ATP16	F-type H+-transporting ATPase subunit delta	8.42	11.76	0.48
Jcr4S01070	21212	21905	K02266 COX6A	cytochrome c oxidase subunit 6a	20.67	31.36	0.6
Jcr4S11126	850	3244	K02155 ATPeV0C, ATP6L	V-type H+-transporting ATPase 16kDa proteolipid subunit	26.02	48.34	0.89
Jcr4S11198	1	533	K05577 ndhF	NAD(P)H-quinone oxidoreductase subunit 5 [EC:1.6.5.3]	851.91	5039.1	2.56
Jcr4S11344	4035	4358	K05573 ndhB	NAD(P)H-quinone oxidoreductase subunit 2 [EC:1.6.5.3]	773.07	3861.96	2.32
Jcr4S11884	4457	8515	K02258 COX11	cytochrome c oxidase assembly protein subunit 11	120.17	159.39	0.41
Jcr4S01215	55453	57814	K00417 QCR7, UQCRB	ubiquinol-cytochrome c reductase subunit 7	35.21	47.03	0.42
Jcr4S01232	54663	57773	K03955 NDUFAB1	NADH dehydrogenase (ubiquinone) 1 alpha/beta subcomplex 1	141.6	257.38	0.86
Jcr4S12411	687	1204	K02155 ATPeV0C, ATP6L	V-type H+-transporting ATPase 16kDa proteolipid subunit	5.36	10.45	0.96
Jcr4S12718	2684	3081	K02261 COX2	cytochrome c oxidase subunit 2	357.45	16986.89	5.57
Jcr4S12859	1066	2492	K03950 NDUFA6	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex subunit 6	28.32	28.74	0.02
Jcr4S01331	10546	12857	K00413 CYC1, CYT1, petC	ubiquinol-cytochrome c reductase cytochrome c1 subunit	33.68	67.94	1.01
Jcr4S01338	783	1106	K02111 ATPF1A, atpA	F-type H+-transporting ATPase subunit alpha [EC:3.6.3.14]	2232.71	2709.64	0.28
Jcr4S01351	25928	26083	K00419 QCR9, UCRC	ubiquinol-cytochrome c reductase subunit 9	1.53	3.92	1.36
Jcr4S14025	744	3159	K02257 COX10	protoheme IX farnesyltransferase [EC:2.5.1.-]	40.57	64.02	0.66
Jcr4S15558	3065	3199	K05575 ndhD	NAD(P)H-quinone oxidoreductase subunit 4 [EC:1.6.5.3]	318.41	2185.75	2.78
Jcr4S01673	1992	8832	K01535 E3.6.3.6	H+-transporting ATPase [EC:3.6.3.6]	181.4	232.55	0.36

Jcr4S00169	55484	58425	K03937 NDUFS4	NADH dehydrogenase (ubiquinone) Fe-S protein 4	54.34	83.61	0.62
Jcr4S17248	9217	10815	K02152 ATPeV1G, ATP6G	V-type H+-transporting ATPase subunit G	39.8	54.87	0.46
Jcr4S17345	1112	1261	K03935 NDUFS2	NADH dehydrogenase (ubiquinone) Fe-S protein 2 [EC:1.6.5.3 1.6.99.3]	141.6	4212.1	4.89
Jcr4S17940	1131	1437	K03879 ND2	NADH-ubiquinone oxidoreductase chain 2 [EC:1.6.5.3]	385	12265.26	4.99
Jcr4S17940	132	749	K03879 ND2	NADH-ubiquinone oxidoreductase chain 2 [EC:1.6.5.3]	299.28	18515.47	5.95
Jcr4S00198	61057	62917	K03941 NDUFS8	NADH dehydrogenase (ubiquinone) Fe-S protein 8 [EC:1.6.5.3 1.6.99.3]	47.46	78.39	0.72
Jcr4S01979	16250	18492	K02138 ATPeF0D, ATP5H, ATP7	F-type H+-transporting ATPase subunit d	54.34	120.2	1.15
Jcr4S00202	25627	27992	K03938 NDUFS5	NADH dehydrogenase (ubiquinone) Fe-S protein 5	64.29	84.92	0.4
Jcr4S02130	14194	14979	K02149 ATPeV1D, ATP6M	V-type H+-transporting ATPase subunit D	26.02	31.36	0.27
Jcr4S02245	18324	18515	K03952 NDUFA8	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex subunit 8	3.06	9.15	1.58
Jcr4S02414	15412	18266	K11353 NDUFA13	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex subunit 13	54.34	54.87	0.01
Jcr4S24336	58	312	K02128 ATPeF0C, ATP5G, ATP9	F-type H+-transporting ATPase subunit c	339.84	22018.15	6.02
Jcr4S02623	1849	2366	K02140 ATPeFG, ATP5L, ATP20	F-type H+-transporting ATPase subunit g	32.91	33.97	0.05
Jcr4S25955	1003	1398	K02111 ATPF1A, atpA	F-type H+-transporting ATPase subunit alpha [EC:3.6.3.14]	189.82	2477.09	3.71
Jcr4S02700	14652	17508	K01507 ppa	inorganic pyrophosphatase [EC:3.6.1.1]	91.85	165.92	0.85
Jcr4S02815	2246	4034	K02151 ATPeV1F, ATP6S14	V-type H+-transporting ATPase subunit F	65.83	95.37	0.53
Jcr4S27284	1789	2278	K05580 ndhI	NAD(P)H-quinone oxidoreductase subunit I [EC:1.6.5.3]	531.96	5271.66	3.31
Jcr4S27322	264	760	K02109 ATPF0B, atpF	F-type H+-transporting ATPase subunit b	1880.62	3777.04	1.01
Jcr4S27372	213	447	K03940 NDUFS7	NADH dehydrogenase (ubiquinone) Fe-S protein 7 [EC:1.6.5.3 1.6.99.3]	1.53	2.61	0.77

Jcr4S02850	26970	37194	K02154 ATPeV0A, ATP6N	V-type H+-transporting ATPase subunit a	383.47	1030.81	1.43
Jcr4S28292	4140	4382	K02110 ATPF0C, atpE	F-type H+-transporting ATPase subunit c	438.58	2538.5	2.53
Jcr4S03174	4431	8177	K02138 ATPeF0D, ATP5H, ATP7	F-type H+-transporting ATPase subunit d	91.08	105.83	0.22
Jcr4S03476	27428	27870	K03966 NDUFB10	NADH dehydrogenase (ubiquinone) 1 beta subcomplex subunit 10	8.42	18.29	1.12
Jcr4S00377	43141	49642	K02154 ATPeV0A, ATP6N	V-type H+-transporting ATPase subunit a	147.72	222.1	0.59
Jcr4S03949	27465	33601	K02144 ATPeV1H	V-type H+-transporting ATPase subunit H	204.37	218.18	0.09
Jcr4S04308	1727	3368	K02155 ATPeV0C, ATP6L	V-type H+-transporting ATPase 16kDa proteolipid subunit	42.1	61.4	0.54
Jcr4S04398	14373	15942	K02155 ATPeV0C, ATP6L	V-type H+-transporting ATPase 16kDa proteolipid subunit	114.81	209.04	0.86
Jcr4S00045	47945	49777	K02153 ATPeV0E, ATP6H	V-type H+-transporting ATPase subunit e	28.32	43.11	0.61
Jcr4S04536	15700	18333	K03965 NDUFB9	NADH dehydrogenase (ubiquinone) 1 beta subcomplex subunit 9	38.27	94.07	1.3
Jcr4S00049	59951	67112	K02145 ATPeV1A, ATP6A	V-type H+-transporting ATPase subunit A [EC:3.6.3.14]	121.7	212.96	0.81
Jcr4S05002	1090	7914	K02154 ATPeV0A, ATP6N	V-type H+-transporting ATPase subunit a	153.08	206.42	0.43
Jcr4S05047	1210	2581	K02267 COX6B	cytochrome c oxidase subunit 6b	39.04	62.71	0.68
Jcr4S05057	4436	8808	K02146 ATPeV0D, ATP6D	V-type H+-transporting ATPase subunit d	161.5	182.91	0.18
Jcr4S00559	5842	7233	K02112 ATPF1B, atpD	F-type H+-transporting ATPase subunit beta [EC:3.6.3.14]	2530.46	10999.28	2.12
Jcr4S05584	12588	13370	K02149 ATPeV1D, ATP6M	V-type H+-transporting ATPase subunit D	26.02	28.74	0.14
Jcr4S05696	15521	20426	K01507 ppa	inorganic pyrophosphatase [EC:3.6.1.1]	117.11	155.47	0.41
Jcr4S05712	6721	10114	K11352 NDUFA12	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex subunit 12	96.44	126.73	0.39
Jcr4S05722	1717	5810	K01535 E3.6.3.6	H+-transporting ATPase [EC:3.6.3.6]	74.25	131.95	0.83
Jcr4S05733	8334	8791	K02155 ATPeV0C, ATP6L	V-type H+-transporting ATPase 16kDa proteolipid subunit	6.12	16.98	1.47

Jcr4S00580	39568	41310	K02135 ATPeF1E, ATP5E, ATP15	F-type H+-transporting ATPase subunit epsilon	63.53	75.78	0.25
Jcr4S00059	85242	85580	K02134 ATPeF1D, ATP5D, ATP16	F-type H+-transporting ATPase subunit delta	12.25	13.06	0.09
Jcr4S00588	31441	32325	K02152 ATPeV1G, ATP6G	V-type H+-transporting ATPase subunit G	93.38	97.99	0.07
Jcr4S00609	17304	25164	K02267 COX6B	cytochrome c oxidase subunit 6b	275.55	3178.67	3.53
Jcr4S00626	18589	18963	K05577 ndhF	NAD(P)H-quinone oxidoreductase subunit 5 [EC:1.6.5.3]	322.24	1508.99	2.23
Jcr4S06338	7171	7509	K03879 ND2	NADH-ubiquinone oxidoreductase chain 2 [EC:1.6.5.3]	108.69	7829.75	6.17
Jcr4S06447	14461	16855	K02155 ATPeV0C, ATP6L	V-type H+-transporting ATPase 16kDa proteolipid subunit	27.55	40.5	0.56
Jcr4S00067	40117	42797	K03937 NDUFS4	NADH dehydrogenase (ubiquinone) Fe-S protein 4	76.54	82.31	0.1
Jcr4S00752	66768	68242	K02140 ATPeFG, ATP5L, ATP20	F-type H+-transporting ATPase subunit g	41.33	66.63	0.69
Jcr4S00770	45323	47044	K03949 NDUFA5	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex subunit 5	62.76	95.37	0.6
Jcr4S07738	12924	16221	K02147 ATPeV1B, ATP6B	V-type H+-transporting ATPase subunit B	274.78	312.25	0.18
Jcr4S00785	3706	10081	K01535 E3.6.3.6	H+-transporting ATPase [EC:3.6.3.6]	146.19	207.73	0.51
Jcr4S07846	3156	3549	K03935 NDUFS2	NADH dehydrogenase (ubiquinone) Fe-S protein 2 [EC:1.6.5.3 1.6.99.3]	150.02	11538.85	6.27
Jcr4S00087	83282	84277	K02140 ATPeFG, ATP5L, ATP20	F-type H+-transporting ATPase subunit g	39.04	49.65	0.35
Jcr4S00087	92804	94332	K03940 NDUFS7	NADH dehydrogenase (ubiquinone) Fe-S protein 7 [EC:1.6.5.3 1.6.99.3]	39.8	54.87	0.46
Jcr4S00868	4940	10137	K02136 ATPeF1G, ATP5C1, ATP3	F-type H+-transporting ATPase subunit gamma	133.95	169.84	0.34
Jcr4S08734	53	196	K05573 ndhB	NAD(P)H-quinone oxidoreductase subunit 2 [EC:1.6.5.3]	557.22	2034.19	1.87
Jcr4S00914	32001	34845	K02150 ATPeV1E, ATP6E	V-type H+-transporting ATPase subunit E	108.69	125.42	0.21
Jcr4S00095	68085	68393	K03963 NDUFB7	NADH dehydrogenase (ubiquinone) 1 beta subcomplex subunit 7	6.89	11.76	0.77

Jcr4U28949	2104	2406	K05576 ndhE	NAD(P)H-quinone oxidoreductase subunit 4L [EC:1.6.5.3]	907.02	3935.13	2.12
Jcr4U30797	1358	1756	K02111 ATPF1A, atpA	F-type H+-transporting ATPase subunit alpha [EC:3.6.3.14]	65.06	1601.75	4.62
Jcr4U30977	1	2899	K11353 NDUFA13	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex subunit 13	52.05	75.78	0.54
Jcr4U39226	332	558	K02256 COX1	cytochrome c oxidase subunit 1 [EC:1.9.3.1]	34.44	8178.58	7.89

Table A2 Genes associated to 'Endocytosis' upregulated in JV

Jatropha Scaffold Ids	**CDS Start Position**	**CDS End Position**	**KO IDS**	**Enzyme Name**	**JH FPKM**	**JV FPKM**	**log2 Fold Change**
Jcr4S01057	14583	14945	K12198 CHMP5, VPS60	charged multivesicular body protein 5	10.72	27.44	1.36
Jcr4S00110	58633	63582	K18443 GBF1	golgi-specific brefeldin A-resistance guanine nucleotide exchange factor 1	149.26	156.78	0.07
Jcr4S00112	15877	21031	K11866 STAMBP, AMSH	STAM-binding protein [EC:3.1.2.15]	105.63	158.08	0.58
Jcr4S01124	7520	9148	K07904 RAB11A	Ras-related protein Rab-11A	35.97	47.03	0.39
Jcr4S01302	24626	27410	K15053 CHMP7	charged multivesicular body protein 7	60.47	88.84	0.56
Jcr4S01365	103	7280	K11824 AP2A	AP-2 complex subunit alpha	103.33	159.39	0.63
Jcr4S01429	37996	45174	K12489 ACAP	Arf-GAP with coiled-coil, ANK repeat and PH domain-containing protein	133.95	228.63	0.77
Jcr4S01456	24965	26546	K07937 ARF1	ADP-ribosylation factor 1	32.15	37.89	0.24
Jcr4S01468	20501	21788	K07904 RAB11A	Ras-related protein Rab-11A	19.14	66.63	1.8
Jcr4S01631	30938	32787	K07904 RAB11A	Ras-related protein Rab-11A	52.81	71.86	0.44
Jcr4S01751	3754	14019	K04646 CLTC	clathrin heavy chain	200.54	276.97	0.47
Jcr4S00192	29777	38100	K12489 ACAP	Arf-GAP with coiled-coil, ANK repeat and PH domain-containing protein	258.71	262.6	0.02
Jcr4S01932	13198	15133	K07904 RAB11A	Ras-related protein Rab-11A	28.32	75.78	1.42
Jcr4S19808	2796	4554	K07897 RAB7A	Ras-related protein Rab-7A	40.57	48.34	0.25
Jcr4S00021	77270	77844	K07897 RAB7A	Ras-related protein Rab-7A	13.01	13.06	0.01
Jcr4S02074	11582	20042	K12471 EPN	epsin	294.68	296.57	0.01
Jcr4S02087	27483	27824	K07904 RAB11A	Ras-related protein Rab-11A	22.2	27.44	0.31
Jcr4S21368	6908	13879	K11826 AP2M1	AP-2 complex subunit mu-1	140.07	248.23	0.83

Jcr4S02307	314	2324	K07904 RAB11A	Ras-related protein Rab-11A	65.06	70.55	0.12
Jcr4S02696	13212	15253	K07937 ARF1	ADP-ribosylation factor 1	35.21	35.28	0
Jcr4S00272	13993	14438	K12198 CHMP5, VPS60	charged multivesicular body protein 5	36.74	100.6	1.45
Jcr4S00287	42639	43696	K12197 CHMP1, VPS46, DID2	charged multivesicular body protein 1	26.79	43.11	0.69
Jcr4S02870	10151	21020	K01528 DNM	dynamin GTPase [EC:3.6.5.5]	332.96	397.17	0.25
Jcr4S00291	45398	46027	K12184 VPS28	ESCRT-I complex subunit VPS28	13.78	22.21	0.69
Jcr4S00365	39866	50592	K12489 ACAP	Arf-GAP with coiled-coil, ANK repeat and PH domain-containing protein	309.23	423.3	0.45
Jcr4S04380	23630	31290	K12200 PDCD6IP, ALIX, RIM20	programmed cell death 6-interacting protein	260.24	275.67	0.08
Jcr4S04837	2656	10302	K12479 VPS45	vacuolar protein sorting-associated protein 45	283.2	506.92	0.84
Jcr4S00549	5033	6157	K12198 CHMP5, VPS60	charged multivesicular body protein 5	30.62	79.7	1.38
Jcr4S00057	154545	156700	K07904 RAB11A	Ras-related protein Rab-11A	46.69	70.55	0.6
Jcr4S00562	45303	46896	K07904 RAB11A	Ras-related protein Rab-11A	52.05	60.1	0.21
Jcr4S05672	7961	11926	K12192 CHMP2B	charged multivesicular body protein 2B	88.02	159.39	0.86
Jcr4S05717	4429	6968	K07904 RAB11A	Ras-related protein Rab-11A	41.33	83.61	1.02
Jcr4S00587	42257	49580	K11866 STAMBP, AMSH	STAM-binding protein [EC:3.1.2.15]	175.28	207.73	0.25
Jcr4S08055	5746	7319	K12194 CHMP4, SNF7, VPS32	charged multivesicular body protein 4	37.51	44.42	0.24
Jcr4S00822	32187	40609	K01528 DNM	dynamin GTPase [EC:3.6.5.5]	198.24	357.98	0.85
Jcr4S00832	40545	42818	K07937 ARF1	ADP-ribosylation factor 1	57.41	77.08	0.43
Jcr4S08472	1117	3434	K07904 RAB11A	Ras-related protein Rab-11A	60.47	551.34	3.19
Jcr4S00087	1099	5003	K12471 EPN	epsin	86.49	406.32	2.23
Jcr4S00962	24440	29973	K12196 VPS4	vacuolar protein-sorting-associated protein 4	251.06	282.2	0.17
Jcr4S09961	409	5596	K09526 DNAJC6	DnaJ homolog subfamily C member 6	156.14	3404.69	4.45

Table A3 Genes associated to 'Arginine and proline metabolism' upregulated in JV

Jatropha Scaffold Ids	**CDS Start Position**	**CDS End Position**	**KO IDS**	**Enzyme Name**	**JH FPKM**	**JV FPKM**	**log2 Fold Change**
Jcr4S01015	7237	12088	K13427 NOA1	nitric-oxide synthase, plant [EC:1.14.13.39]	103.33	193.36	0.9

Jcr4S01142	13334	15557	K00472 E1.14.11.2	prolyl 4-hydroxylase [EC:1.14.11.2]	92.62	563.09	2.6
Jcr4S01861	20045	27028	K13366 MPAO, PAO1	polyamine oxidase [EC:1.5.3.14 1.5.3.16 1.5.3.-]	158.44	239.09	0.59
Jcr4S00221	23448	24704	K00318 PRODH	proline dehydrogenase [EC:1.5.-.-]	19.9	27.44	0.46
Jcr4S00225	2644	9744	K00819 E2.6.1.13, rocD	ornithine--oxo-acid transaminase [EC:2.6.1.13]	246.46	295.27	0.26
Jcr4S00004	28579	30944	K00472 E1.14.11.2	prolyl 4-hydroxylase [EC:1.14.11.2]	54.34	70.55	0.38
Jcr4S00004	63856	65103	K01581 E4.1.1.17, ODC1, speC, speF	ornithine decarboxylase [EC:4.1.1.17]	28.32	35.28	0.32
Jcr4S04145	3259	4347	K01581 E4.1.1.17, ODC1, speC, speF	ornithine decarboxylase [EC:4.1.1.17]	44.39	57.49	0.37
Jcr4S00056	67627	73486	K01259 pip	proline iminopeptidase [EC:3.4.11.5]	195.18	206.42	0.08
Jcr4S00007	180442	183801	K00472 E1.14.11.2	prolyl 4-hydroxylase [EC:1.14.11.2]	172.22	3228.32	4.23
Jcr4S06167	12417	14677	K13366 MPAO, PAO1	polyamine oxidase [EC:1.5.3.14 1.5.3.16 1.5.3.-]	51.28	88.84	0.79
Jcr4S07719	7885	9448	K17839 PAO4, PAO3, PAO2	polyamine oxidase [EC:1.5.3.17 1.5.3.-]	44.39	56.18	0.34
Jcr4S00799	21523	23199	K12259 SMOX, PAO5	spermine oxidase [EC:1.5.3.16 1.5.3.-]	38.27	45.73	0.26
Jcr4S00089	1021	2931	K13366 MPAO, PAO1	polyamine oxidase [EC:1.5.3.14 1.5.3.16 1.5.3.-]	52.05	56.18	0.11

Table A4 Genes associated to 'Ascorbate metabolism' upregulated in JV

Jatropha Scaffold Ids	**CDS Start Position**	**CDS End Position**	**KO IDS**	**Enzyme Name**	**JH FPKM**	**JV FPKM**	**log2 Fold Change**
Jcr4S00174	77934	80615	K00225 GLDH	L-galactono-1,4-lactone dehydrogenase [EC:1.3.2.3]	106.39	881.88	3.05

Jcr4S02228	3656	8958	K10047 VTC4	inositol-phosphate phosphatase / L-galactose 1-phosphate phosphatase [EC:3.1.3.25 3.1.3.93]	137.01	253.46	0.89
Jcr4S02264	23502	25224	K14190 VTC2_5	GDP-L-galactose phosphorylase [EC:2.7.7.69]	33.68	74.47	1.14
Jcr4S02642	5021	7928	K00469 MIOX	inositol oxygenase [EC:1.13.99.1]	86.49	96.68	0.16
Jcr4S03260	1821	5025	K17744 GalDH	L-galactose dehydrogenase [EC:1.1.1.316]	95.68	118.89	0.31
Jcr4S00512	17499	21740	K00434 E1.11.1.11	L-ascorbate peroxidase [EC:1.11.1.11]	103.33	114.97	0.15
Jcr4S00535	38221	47325	K08232 E1.6.5.4	monodehydroascorbate reductase (NADH) [EC:1.6.5.4]	271.72	420.69	0.63
Jcr4S00609	32873	35567	K00434 E1.11.1.11	L-ascorbate peroxidase [EC:1.11.1.11]	50.52	60.1	0.25
Jcr4S06417	4626	7354	K00434 E1.11.1.11	L-ascorbate peroxidase [EC:1.11.1.11]	75.01	146.33	0.96
Jcr4S00717	47054	48241	K00434 E1.11.1.11	L-ascorbate peroxidase [EC:1.11.1.11]	23.73	361.9	3.93
Jcr4S00820	14045	18639	K08232 E1.6.5.4	monodehydroascorbate reductase (NADH) [EC:1.6.5.4]	116.34	156.78	0.43

Table A5 Genes associated to 'Lipid metabolism' upregulated in JV

Jatropha Scaffold Ids	**CDS Start Position**	**CDS End Position**	**KO IDS**	**Enzyme Name**	**JH FPKM**	**JV FPKM**	**log2 Fold Change**
Jcr4S01030	59212	64955	K03715 E2.4.1.46	1,2-diacylglycerol 3-beta-galactosyltransferase [EC:2.4.1.46]	163.03	455.96	1.48
Jcr4S10925	4226	12780	K13506 GPAT3_4, AGPAT9, AGPAT6	glycerol-3-phosphate O-acyltransferase 3/4 [EC:2.3.1.15]	267.13	330.54	0.31
Jcr4S01121	45761	48621	K06130 LYPLA2	lysophospholipase II [EC:3.1.1.5]	42.86	70.55	0.72
Jcr4S11690	331	1626	K16818 K16818, DAD1	phospholipase A1 [EC:3.1.1.32]	20.67	49.65	1.26
Jcr4S00118	76399	78520	K01634 E4.1.2.27	sphinganine-1-phosphate aldolase [EC:4.1.2.27]	52.81	67.94	0.36
Jcr4S12073	2975	5640	K14156 CHK	choline/ethanolamine kinase [EC:2.7.1.32 2.7.1.82]	79.6	108.44	0.45
Jcr4S01215	16690	20063	K06130 LYPLA2	lysophospholipase II [EC:3.1.1.5]	64.29	99.29	0.63
Jcr4S12284	1345	4502	K00981 E2.7.7.41, CDS1, CDS2, cdsA	phosphatidate cytidylyltransferase [EC:2.7.7.41]	78.84	112.36	0.51

Jcr4S12769	2581	4051	K01094 GEP4	phosphatidylglycerophosphatase GEP4 [EC:3.1.3.27]	62	70.55	0.19
Jcr4S13939	246	741	K01047 PLA2G, SPLA2	secretory phospholipase A2 [EC:3.1.1.4]	19.14	20.9	0.13
Jcr4S14422	3030	5950	K01613 psd, PISD	phosphatidylserine decarboxylase [EC:4.1.1.65]	74.25	94.07	0.34
Jcr4S01662	7436	13190	K00894 ETNK, EKI	ethanolamine kinase [EC:2.7.1.82]	109.45	226.02	1.05
Jcr4S17175	6201	6929	K04713 SUR2	sphinganine C4-monooxygenase [EC:1.14.13.169]	6.89	11.76	0.77
Jcr4S00177	95013	97540	K01115 PLD1_2	phospholipase D1/2 [EC:3.1.4.4]	98.74	165.92	0.75
Jcr4S00184	67119	73293	K13510 LPCAT1_2	lysophosphatidylcholine acyltransferase / lyso-PAF acetyltransferase [EC:2.3.1.23 2.3.1.67]	206.66	235.17	0.19
Jcr4S00189	50559	53589	K00901 E2.7.1.107, DGK, dgkA	diacylglycerol kinase (ATP dependent) [EC:2.7.1.107]	93.38	128.04	0.46
Jcr4S01903	29748	34049	K00993 EPT1	ethanolaminephosphotransferase [EC:2.7.8.1]	98.74	165.92	0.75
Jcr4S01943	12234	17460	K00654 E2.3.1.50	serine palmitoyltransferase [EC:2.3.1.50]	172.98	182.91	0.08
Jcr4S20593	471	2555	K00968 PCYT1	choline-phosphate cytidylyltransferase [EC:2.7.7.15]	58.17	99.29	0.77
Jcr4S20632	3240	9620	K00967 PCYT2	ethanolamine-phosphate cytidylyltransferase [EC:2.7.7.14]	90.32	101.91	0.17
Jcr4S02133	6334	10258	K13519 LPT1, ALE1	lysophospholipid acyltransferase [EC:2.3.1.51 2.3.1.23 2.3.1.-]	120.17	126.73	0.08
Jcr4S02216	30955	36521	K01115 PLD1_2	phospholipase D1/2 [EC:3.1.4.4]	155.38	198.59	0.35
Jcr4S00233	118697	122835	K00981 E2.7.7.41, CDS1, CDS2, cdsA	phosphatidate cytidylyltransferase [EC:2.7.7.41]	75.78	104.52	0.46
Jcr4S26021	1798	4201	K13508 GPAT	glycerol-3-phosphate acyltransferase [EC:2.3.1.15]	47.46	69.24	0.55
Jcr4S02881	17539	19802	K09480 E2.4.1.241	digalactosyldiacylglycerol synthase [EC:2.4.1.241]	52.81	73.16	0.47
Jcr4S00290	63689	66955	K01115 PLD1_2	phospholipase D1/2 [EC:3.1.4.4]	76.54	95.37	0.32
Jcr4S00291	60983	66313	K04718 SPHK	sphingosine kinase [EC:2.7.1.91]	115.58	155.47	0.43
Jcr4S03010	7498	10828	K13508 GPAT	glycerol-3-phosphate acyltransferase [EC:2.3.1.15]	95.68	143.71	0.59
Jcr4S03186	28430	32705	K00901 E2.7.1.107, DGK, dgkA	diacylglycerol kinase (ATP dependent) [EC:2.7.1.107]	143.9	158.08	0.14
Jcr4S00343	7094	8164	K00655 plsC	1-acyl-sn-glycerol-3-phosphate acyltransferase [EC:2.3.1.51]	18.37	41.81	1.19
Jcr4S00359	37760	39143	K06129 LYPLA3	lysophospholipase III [EC:3.1.1.5]	25.26	36.58	0.53
Jcr4S03605	38885	43543	K06130 LYPLA2	lysophospholipase II [EC:3.1.1.5]	160.74	188.13	0.23
Jcr4S03642	12560	14150	K13508 GPAT	glycerol-3-phosphate acyltransferase [EC:2.3.1.15]	55.11	65.32	0.25
Jcr4S03882	265	4382	K05929 E2.1.1.103, NMT	phosphoethanolamine N-methyltransferase [EC:2.1.1.103]	114.05	152.86	0.42
Jcr4S00441	18008	23994	K04715 E2.7.1.138, CERK	ceramide kinase [EC:2.7.1.138]	149.26	158.08	0.08
Jcr4S04717	5181	12114	K01115 PLD1_2	phospholipase D1/2 [EC:3.1.4.4]	143.13	189.44	0.4

Jcr4S00489	40622	46111	K01613 psd, PISD	phosphatidylserine decarboxylase [EC:4.1.1.65]	287.8	296.57	0.04
Jcr4S00052	54013	57136	K16860 PLD3_4	phospholipase D3/4 [EC:3.1.4.4]	132.42	743.39	2.49
Jcr4S05261	10800	12193	K13508 GPAT	glycerol-3-phosphate acyltransferase [EC:2.3.1.15]	58.17	62.71	0.11
Jcr4S00537	26857	28360	K00550 OPI3	methylene-fatty-acyl-phospholipid synthase [EC:2.1.1.16]	18.37	26.13	0.51
Jcr4S05451	8483	10956	K08744 CRLS	cardiolipin synthase [EC:2.7.8.-]	92.62	96.68	0.06
Jcr4S00594	46937	48574	K04711 YDC1	dihydroceramidase [EC:3.5.1.-]	23.73	35.28	0.57
Jcr4S05976	11031	11962	K00981 E2.7.7.41, CDS1, CDS2, cdsA	phosphatidate cytidylyltransferase [EC:2.7.7.41]	63.53	343.6	2.44
Jcr4S05976	8471	9933	K04711 YDC1	dihydroceramidase [EC:3.5.1.-]	19.9	37.89	0.93
Jcr4S05986	25313	30274	K05929 E2.1.1.103, NMT	phosphoethanolamine N-methyltransferase [EC:2.1.1.103]	137.01	222.1	0.7
Jcr4S06277	24699	25232	K04713 SUR2	sphinganine C4-monooxygenase [EC:1.14.13.169]	3.06	6.53	1.09
Jcr4S06284	21175	25072	K05929 E2.1.1.103, NMT	phosphoethanolamine N-methyltransferase [EC:2.1.1.103]	79.6	103.21	0.37
Jcr4S00065	3640	10149	K15728 LPIN	phosphatidate phosphatase LPIN [EC:3.1.3.4]	173.75	192.05	0.14
Jcr4S06785	20881	22767	K04710 CERS	ceramide synthetase [EC:2.3.1.24]	44.39	58.79	0.41
Jcr4S06819	14718	19714	K04713 SUR2	sphinganine C4-monooxygenase [EC:1.14.13.169]	88.79	120.2	0.44
Jcr4S07665	4960	9019	K00630 ATS1	glycerol-3-phosphate O-acyltransferase [EC:2.3.1.15]	98.74	180.29	0.87
Jcr4S00843	20132	26425	K00654 E2.3.1.50	serine palmitoyltransferase [EC:2.3.1.50]	78.84	176.38	1.16
Jcr4S00866	23556	27785	K01115 PLD1_2	phospholipase D1/2 [EC:3.1.4.4]	111.75	194.67	0.8
Jcr4S00868	43509	47405	K01126 E3.1.4.46, glpQ, ugpQ	glycerophosphoryl diester phosphodiesterase [EC:3.1.4.46]	104.1	134.57	0.37
Jcr4S08851	2590	5721	K00679 E2.3.1.158	phospholipid:diacylglycerol acyltransferase [EC:2.3.1.158]	78.84	91.45	0.21
Jcr4U30985	4353	6038	K06130 LYPLA2	lysophospholipase II [EC:3.1.1.5]	7.65	15.68	1.03

Table A6 Genes associated to 'Fatty acid metabolism' upregulated in JV

Jatropha Scaffold Ids	**CDS Start Position**	**CDS End Position**	**KO IDS**	**Enzyme Name**	**JH FPKM**	**JV FPKM**	**log2 Fold Change**
Jcr4S01110	19435	25174	K01897 ACSL, fadD	long-chain acyl-CoA synthetase [EC:6.2.1.3]	227.33	246.93	0.12
Jcr4S11223	1987	3150	K10256 FAD2	omega-6 fatty acid desaturase (delta-12 desaturase) [EC:1.14.19.-]	28.32	49.65	0.81

Jcr4S01280	48335	49353	K00232 E1.3.3.6, ACOX1, ACOX3	acyl-CoA oxidase [EC:1.3.3.6]	13.01	16.98	0.38
Jcr4S01370	14660	16448	K03921 DESA1	acyl-[acyl-carrier-protein] desaturase [EC:1.14.19.2]	44.39	78.39	0.82
Jcr4S15816	3772	6990	K07513 ACAA1	acetyl-CoA acyltransferase 1 [EC:2.3.1.16]	62.76	137.18	1.13
Jcr4S01960	13949	17954	K00208 fabI	enoyl-[acyl-carrier protein] reductase I [EC:1.3.1.9 1.3.1.10]	142.37	145.02	0.03
Jcr4S19645	6999	9238	K10257 FAD8, desB	omega-3 fatty acid desaturase (delta-15 desaturase) [EC:1.14.19.-]	26.02	36.58	0.49
Jcr4S02200	26764	39836	K11262 ACAC	acetyl-CoA carboxylase / biotin carboxylase [EC:6.4.1.2 6.3.4.14]	440.11	441.59	0
Jcr4S00244	25069	29600	K01074 PPT	palmitoyl-protein thioesterase [EC:3.1.2.22]	88.02	210.34	1.26
Jcr4S00031	117994	118812	K00059 fabG	3-oxoacyl-[acyl-carrier protein] reductase [EC:1.1.1.100]	30.62	36.58	0.26
Jcr4S03143	20868	22845	K01074 PPT	palmitoyl-protein thioesterase [EC:3.1.2.22]	50.52	58.79	0.22
Jcr4S03253	3672	8413	K00208 fabI	enoyl-[acyl-carrier protein] reductase I [EC:1.3.1.9 1.3.1.10]	172.22	182.91	0.09
Jcr4S03305	1811	2059	K02372 fabZ	3-hydroxyacyl-[acyl-carrier-protein] dehydratase [EC:4.2.1.59]	10.72	14.37	0.42
Jcr4S03307	3328	5378	K10257 FAD8, desB	omega-3 fatty acid desaturase (delta-15 desaturase) [EC:1.14.19.-]	48.22	67.94	0.49
Jcr4S03452	2436	7410	K10255 FAD6, desA	omega-6 fatty acid desaturase (delta-12 desaturase) [EC:1.14.19.-]	127.06	160.7	0.34
Jcr4S03522	4345	6148	K03921 DESA1	acyl-[acyl-carrier-protein] desaturase [EC:1.14.19.2]	52.05	101.91	0.97
Jcr4S03974	4354	8026	K10781 FATB	fatty acyl-ACP thioesterase B [EC:3.1.2.14 3.1.2.21]	132.42	155.47	0.23
Jcr4S04095	21673	23912	K10257 FAD8, desB	omega-3 fatty acid desaturase (delta-15 desaturase) [EC:1.14.19.-]	22.2	30.05	0.44
Jcr4S04503	16425	20639	K07512 MECR, NRBF1	mitochondrial trans-2-enoyl-CoA reductase [EC:1.3.1.38]	129.35	509.53	1.98
Jcr4S04875	19590	24384	K09458 fabF	3-oxoacyl-[acyl-carrier-protein] synthase II [EC:2.3.1.179]	138.54	176.38	0.35
Jcr4S00505	3451	4473	K10258 TER, TSC13, CER10	very-long-chain enoyl-CoA reductase [EC:1.3.1.93]	68.89	176.38	1.36
Jcr4S05261	14446	18185	K01897 ACSL, fadD	long-chain acyl-CoA synthetase [EC:6.2.1.3]	63.53	81	0.35
Jcr4S00616	35175	40805	K10527 MFP2	enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase [EC:4.2.1.17 1.1.1.35 1.1.1.211]	183.7	199.89	0.12

Jcr4S00066	127095	128235	K10251 HSD17B12, KAR, IFA38	17beta-estradiol 17-dehydrogenase / very-long-chain 3-oxoacyl-CoA reductase [EC:1.1.1.62 1.1.1.330]	15.31	26.13	0.77
Jcr4S00733	19113	23630	K01897 ACSL, fadD	long-chain acyl-CoA synthetase [EC:6.2.1.3]	85.73	163.31	0.93
Jcr4S00893	35578	43224	K01897 ACSL, fadD	long-chain acyl-CoA synthetase [EC:6.2.1.3]	163.8	334.46	1.03
Jcr4S00100	80695	83102	K00059 fabG	3-oxoacyl-[acyl-carrier protein] reductase [EC:1.1.1.100]	78.84	120.2	0.61

Table A7 Genes associated to 'Amino sugar and nucleotide sugar metabolism' upregulated in JV

Jatropha Scaffold Ids	**CDS Start Position**	**CDS End Position**	**KO IDS**	**Enzyme Name**	**JH FPKM**	**JV FPKM**	**log2 Fold Change**
Jcr4S00001	90127	91017	K01183 E3.2.1.14	chitinase [EC:3.2.1.14]	8.42	22.21	1.4
Jcr4S00001	92393	93286	K01183 E3.2.1.14	chitinase [EC:3.2.1.14]	16.84	18.29	0.12
Jcr4S11439	60	4275	K01209 E3.2.1.55, abfA	alpha-N-arabinofuranosidase [EC:3.2.1.55]	52.81	64.02	0.28
Jcr4S01202	7216	8755	K01183 E3.2.1.14	chitinase [EC:3.2.1.14]	12.25	32.66	1.42
Jcr4S01297	25691	34587	K00972 UAP1	UDP-N-acetylglucosamine/UDP-N-acetylgalactosamine diphosphorylase [EC:2.7.7.23 2.7.7.83]	287.03	313.56	0.13
Jcr4S13269	73	479	K01183 E3.2.1.14	chitinase [EC:3.2.1.14]	3.83	7.84	1.03
Jcr4S01670	17946	20004	K12450 RHM	UDP-glucose 4,6-dehydratase [EC:4.2.1.76]	80.37	314.86	1.97
Jcr4S19079	21	312	K01183 E3.2.1.14	chitinase [EC:3.2.1.14]	0.77	5.23	2.77
Jcr4S00208	29077	39805	K12446 E2.7.1.46	L-arabinokinase [EC:2.7.1.46]	196.71	252.15	0.36
Jcr4S00022	23878	24768	K01183 E3.2.1.14	chitinase [EC:3.2.1.14]	13.01	16.98	0.38
Jcr4S02113	8316	9413	K01183 E3.2.1.14	chitinase [EC:3.2.1.14]	22.96	43.11	0.91
Jcr4S25840	267	4300	K12446 E2.7.1.46	L-arabinokinase [EC:2.7.1.46]	81.13	111.05	0.45
Jcr4S26960	3518	5056	K01183 E3.2.1.14	chitinase [EC:3.2.1.14]	20.67	40.5	0.97
Jcr4S03110	10440	13334	K12448 UXE	UDP-arabinose 4-epimerase [EC:5.1.3.5]	43.63	134.57	1.62
Jcr4S00405	22605	30394	K12373 HEXA_B	hexosaminidase [EC:3.2.1.52]	618.45	1297.34	1.07
Jcr4S04076	24545	28310	K01836 E5.4.2.3	phosphoacetylglucosamine mutase [EC:5.4.2.3]	123.23	129.34	0.07
Jcr4S00428	9376	13290	K01209 E3.2.1.55, abfA	alpha-N-arabinofuranosidase [EC:3.2.1.55]	71.18	117.58	0.72
Jcr4S04441	576	5026	K12448 UXE	UDP-arabinose 4-epimerase [EC:5.1.3.5]	113.28	3253.14	4.84

Jcr4S04532	4926	9340	K12373 HEXA_B	hexosaminidase [EC:3.2.1.52]	78.07	189.44	1.28
Jcr4S00622	6407	7627	K01183 E3.2.1.14	chitinase [EC:3.2.1.14]	17.6	36.58	1.06
Jcr4S06366	14816	15656	K12449 AXS	UDP-apiose/xylose synthase	21.43	28.74	0.42
Jcr4S07028	1878	3590	K12451 UER1	3,5-epimerase/4-reductase [EC:5.1.3.- 1.1.1.-]	30.62	86.23	1.49
Jcr4S07198	18189	19783	K01183 E3.2.1.14	chitinase [EC:3.2.1.14]	32.15	50.95	0.66
Jcr4S07468	2158	5587	K01183 E3.2.1.14	chitinase [EC:3.2.1.14]	496.75	751.23	0.6
Jcr4S09708	4204	6964	K18677 GALAK	galacturonokinase [EC:2.7.1.44]	49.75	92.76	0.9
Jcr4S00975	35345	36959	K01183 E3.2.1.14	chitinase [EC:3.2.1.14]	75.01	152.86	1.03
Jcr4S09881	626	1032	K01183 E3.2.1.14	chitinase [EC:3.2.1.14]	0.77	13.06	4.09
Jcr4U35651	4217	5107	K01183 E3.2.1.14	chitinase [EC:3.2.1.14]	4.59	5.23	0.19

Table A8 Genes associated to 'Terpenoid biosynthesis' upregulated in JV

Jatropha Scaffold Ids	**CDS Start Position**	**CDS End Position**	**KO IDS**	**Enzyme Name**	**JH FPKM**	**JV FPKM**	**log2 Fold Change**
Jcr4S10676	1279	3486	K15813 LUP4	beta-amyrin synthase [EC:5.4.99.39]	35.21	83.61	1.25
Jcr4S01327	18126	20561	K15086 TPS14	(3S)-linalool synthase [EC:4.2.3.25]	19.9	121.5	2.61
Jcr4S14240	142	1362	K07385 TPS-Cin	1,8-cineole synthase [EC:4.2.3.108]	23.73	39.19	0.72
Jcr4S01621	23594	28560	K17982 TPS04, GES	geranyllinalool synthase [EC:4.2.3.144]	116.34	162	0.48
Jcr4S17398	2304	3425	K04123 KAO	ent-kaurenoic acid hydroxylase [EC:1.14.13.79]	22.2	40.5	0.87
Jcr4S01951	1086	5370	K15813 LUP4	beta-amyrin synthase [EC:5.4.99.39]	406.43	523.9	0.37
Jcr4S22350	4503	5946	K04125 E1.14.11.13	gibberellin 2-oxidase [EC:1.14.11.13]	13.78	31.36	1.19
Jcr4S02602	10015	12814	K18108 E4.2.3.111	(-)-alpha-terpineol synthase [EC:4.2.3.111]	133.18	190.75	0.52
Jcr4S00030	25404	28246	K04122 GA3, CYP701	ent-kaurene oxidase [EC:1.14.13.78]	75.01	92.76	0.31
Jcr4S03333	6165	7831	K15803 GERD	(-)-germacrene D synthase [EC:4.2.3.22 4.2.3.75]	68.89	142.41	1.05
Jcr4S03817	23506	25538	K04124 E1.14.11.15	gibberellin 3-beta-dioxygenase [EC:1.14.11.15]	31.38	81	1.37
Jcr4S03853	509	3341	K15803 GERD	(-)-germacrene D synthase [EC:4.2.3.22 4.2.3.75]	97.97	231.25	1.24
Jcr4S00574	28574	33679	K04120 E5.5.1.13	ent-copalyl diphosphate synthase [EC:5.5.1.13]	97.21	231.25	1.25

Jcr4S06166	7879	9178	K05282 E1.14.11.12	gibberellin 20-oxidase [EC:1.14.11.12]	13.78	77.08	2.48
Jcr4S06457	4841	8995	K04121 E4.2.3.19	ent-kaurene synthase [EC:4.2.3.19]	75.78	121.5	0.68
Jcr4S00075	23895	34277	K15095 E1.1.1.208	(+)-neomenthol dehydrogenase [EC:1.1.1.208]	547.27	1502.45	1.46
Jcr4S00092	37049	39293	K15095 E1.1.1.208	(+)-neomenthol dehydrogenase [EC:1.1.1.208]	30.62	78.39	1.36

Table A9 Genes associated to 'Plant hormone signal transduction' upregulated in JV

Jatropha Scaffold Ids	**CDS Start Position**	**CDS End Position**	**KO IDS**	**Enzyme Name**	**JV FPKM**	**JH FPKM**	**log2 Fold Change**
Jcr4S00002	192943	193347	K14488 SAUR	SAUR family protein	18.37	11.76	0.64
Jcr4S00105	47941	52341	K14485 TIR1	transport inhibitor response 1	150.79	111.05	0.44
Jcr4S10927	12900	14690	K14484 IAA	auxin-responsive protein IAA	81.13	35.28	1.2
Jcr4S11683	4775	6022	K14490 AHP	histidine-containing phosphotransfer protein	35.21	19.6	0.85
Jcr4S01181	8805	10040	K14484 IAA	auxin-responsive protein IAA	57.41	52.26	0.14
Jcr4S01183	2107	2925	K13463 COI-1	coronatine-insensitive protein 1	58.17	26.13	1.15
Jcr4S11945	8454	8819	K14488 SAUR	SAUR family protein	3.06	1.31	1.23
Jcr4S13192	8754	11388	K14496 PYL	abscisic acid receptor PYR/PYL family	189.06	40.5	2.22
Jcr4S13195	35137	40366	K14506 JAR1	jasmonic acid-amino synthetase	133.95	126.73	0.08
Jcr4S01346	18442	18744	K14488 SAUR	SAUR family protein	13.01	6.53	0.99
Jcr4S01346	24098	24406	K14488 SAUR	SAUR family protein	13.78	6.53	1.08
Jcr4S00136	46109	46627	K14488 SAUR	SAUR family protein	23.73	6.53	1.86
Jcr4S01352	4239	5938	K14497 PP2C	protein phosphatase 2C [EC:3.1.3.16]	69.65	32.66	1.09
Jcr4S13611	405	1388	K14490 AHP	histidine-containing phosphotransfer peotein	48.99	41.81	0.23
Jcr4S14162	1718	2158	K14488 SAUR	SAUR family protein	24.49	15.68	0.64
Jcr4S01426	36407	36982	K14496 PYL	abscisic acid receptor PYR/PYL family	16.07	6.53	1.3
Jcr4S01477	53915	55411	K14503 BZR1_2	brassinosteroid resistant 1/2	159.21	47.03	1.76
Jcr4S01542	3778	7695	K14491 ARR-B	two-component response regulator ARR-B family	129.35	109.74	0.24
Jcr4S01562	33606	36118	K13946 AUX1, LAX	auxin influx carrier (AUX1 LAX family)	30.05	19.14	0.65
Jcr4S00167	113458	113775	K14488 SAUR	SAUR family protein	13.01	11.76	0.15
Jcr4S01667	10448	10876	K14488 SAUR	SAUR family protein	19.14	7.84	1.29
Jcr4S00168	90949	91494	K13464 JAZ	jasmonate ZIM domain-containing protein	13.83	3.06	1.77
Jcr4S00174	67604	70052	K14508 NPR1	regulatory protein NPR1	178.34	91.45	0.96

Jcr4S01759	11162	18682	K14509 ETR, ERS	ethylene receptor [EC:2.7.13.-]	395.72	254.76	0.64
Jcr4S00177	36655	37526	K14484 IAA	auxin-responsive protein IAA	18.37	16.98	0.11
Jcr4S00177	82998	86435	K14515 EBF1_2	EIN3-binding F-box protein	131.65	101.91	0.37
Jcr4S00179	67930	77140	K14489 AHK2_3_4	arabidopsis histidine kinase 2/3/4 (cytokinin receptor) [EC:2.7.13.3]	385	333.15	0.21
Jcr4S00208	51596	52102	K14488 SAUR	SAUR family protein	15.31	13.06	0.23
Jcr4S00216	17036	22561	K14489 AHK2_3_4	arabidopsis histidine kinase 2/3/4 (cytokinin receptor) [EC:2.7.13.3]	182.93	154.16	0.25
Jcr4S02294	15966	19115	K14497 PP2C	protein phosphatase 2C [EC:3.1.3.16]	127.06	88.84	0.52
Jcr4S02314	11018	17172	K14486 K14486, ARF	auxin response factor	229.62	141.1	0.7
Jcr4S02323	6624	8483	K14514 EIN3	ethylene-insensitive protein 3	180.64	62.71	1.53
Jcr4S00235	29197	30457	K14493 GID1	gibberellin receptor GID1 [EC:3.-.-.-]	45.92	31.36	0.55
Jcr4S02357	942	3048	K14487 GH3	auxin responsive GH3 gene family	86.63	28.32	1.23
Jcr4S02550	1115	4019	K14497 PP2C	protein phosphatase 2C [EC:3.1.3.16]	62.76	39.19	0.68
Jcr4S00256	18847	20755	K14484 IAA	auxin-responsive protein IAA	110.99	58.79	0.92
Jcr4S00260	2871	6315	K14485 TIR1	transport inhibitor response 1	133.95	104.52	0.36
Jcr4S02762	33868	34206	K14488 SAUR	SAUR family protein	21.43	5.23	2.04
Jcr4S00299	24169	25017	K14484 IAA	auxin-responsive protein IAA	45.16	44.42	0.02
Jcr4S03127	3745	7782	K14508 NPR1	regulatory protein NPR1	108.69	107.13	0.02
Jcr4S03391	19543	24227	K14491 ARR-B	two-component response regulator ARR-B family	146.19	133.26	0.13
Jcr4S00036	96162	98037	K14487 GH3	auxin responsive GH3 gene family	78.84	52.26	0.59
Jcr4S00036	123748	126386	K14490 AHP	histidine-containing phosphotransfer peotein	105.63	69.24	0.61
Jcr4S03535	199	468	K14488 SAUR	SAUR family protein	13.78	6.53	1.08
Jcr4S03599	20390	26310	K14513 EIN2	ethylene-insensitive protein 2	144.66	111.05	0.38
Jcr4S03691	22349	22847	K14488 SAUR	SAUR family protein	24.49	13.06	0.91
Jcr4S03704	890	2365	K14506 JAR1	jasmonic acid-amino synthetase	10.72	10.45	0.04
Jcr4S00005	128307	130238	K14497 PP2C	protein phosphatase 2C [EC:3.1.3.16]	68.89	33.97	1.02
Jcr4S00409	31399	31893	K14488 SAUR	SAUR family protein	6.12	5.23	0.23
Jcr4S04185	13121	13489	K14488 SAUR	SAUR family protein	35.97	22.21	0.7
Jcr4S00458	50619	53558	K14484 IAA	auxin-responsive protein IAA	259.54	84.2	1.57
Jcr4S00497	63734	67994	K14486 K14486, ARF	auxin response factor	257.18	122.81	1.07
Jcr4S05248	12359	12964	K14496 PYL	abscisic acid receptor PYR/PYL family	26.89	6.13	1.92
Jcr4S05493	10994	12797	K14514 EIN3	ethylene-insensitive protein 3	65.06	44.42	0.55

Jcr4S00007	186049	187501	K14490 AHP	histidine-containing phosphotransfer peotein	79.6	75.78	0.07
Jcr4S00615	32965	34015	K14488 SAUR	SAUR family protein	46.69	24.82	0.91
Jcr4S06304	5727	6023	K14488 SAUR	SAUR family protein	9.18	6.53	0.49
Jcr4S06403	1897	2651	K14510 CTR1	serine/threonine-protein kinase CTR1 [EC:2.7.11.1]	39.04	20.9	0.9
Jcr4S00649	22766	25885	K13464 JAZ	jasmonate ZIM domain-containing protein	177.58	133.26	0.41
Jcr4S00066	166630	168832	K14491 ARR-B	two-component response regulator ARR-B family	167.63	99.29	0.76
Jcr4S06653	27541	28435	K14484 IAA	auxin-responsive protein IAA	18.37	2.61	2.81
Jcr4S06840	11939	13656	K14493 GID1	gibberellin receptor GID1 [EC:3.-.-.-]	49.75	37.89	0.39
Jcr4S00712	6555	13008	K14486 K14486, ARF	auxin response factor	292.39	253.46	0.21
Jcr4S07198	384	1978	K14491 ARR-B	two-component response regulator ARR-B family	16.98	3.06	2.47
Jcr4S00074	3733	4638	K14484 IAA	auxin-responsive protein IAA	37.51	26.13	0.52
Jcr4S07634	15017	16619	K14497 PP2C	protein phosphatase 2C [EC:3.1.3.16]	116.34	32.66	1.83
Jcr4S00768	42788	45503	K14486 K14486, ARF	auxin response factor	88.02	70.55	0.32
Jcr4S08276	13861	14166	K14488 SAUR	SAUR family protein	10.45	1.53	2.77
Jcr4S00865	13209	14328	K14503 BZR1_2	brassinosteroid resistant 1/2	233.45	50.95	2.2
Jcr4S00867	39217	42121	K14497 PP2C	protein phosphatase 2C [EC:3.1.3.16]	71.18	64.02	0.15
Jcr4S00872	4092	10968	K14497 PP2C	protein phosphatase 2C [EC:3.1.3.16]	242.64	232.55	0.06
Jcr4S09250	5011	6286	K14484 IAA	auxin-responsive protein IAA	44.39	41.81	0.09
Jcr4S09423	1048	3561	K13946 AUX1, LAX	auxin influx carrier (AUX1 LAX family)	52.05	40.5	0.36
Jcr4S00944	381	6059	K14512 MPK6	mitogen-activated protein kinase 6 [EC:2.7.11.24]	203.6	181.6	0.16
Jcr4S09468	3632	7007	K14509 ETR, ERS	ethylene receptor [EC:2.7.13.-]	192.12	79.7	1.27
Jcr4S00096	63904	66483	K14496 PYL	abscisic acid receptor PYR/PYL family	114.05	105.83	0.11
Jcr4S00096	92641	94097	K14497 PP2C	protein phosphatase 2C [EC:3.1.3.16]	135.48	33.97	2
Jcr4S01003	6958	8406	K13422 MYC2	transcription factor MYC2	39.8	23.52	0.76
Jcr4S01122	4162	5610	K13422 MYC2	transcription factor MYC2	33.68	35.28	-0.07
Jcr4S01136	46674	49869	K14431 TGA	transcription factor TGA	89.55	109.74	-0.29
Jcr4S00121	67036	70857	K14431 TGA	transcription factor TGA	84.2	150.25	-0.84
Jcr4S00121	79733	80997	K14432 ABF	ABA responsive element binding factor	42.1	47.03	-0.16
Jcr4S01221	63527	66553	K14431 TGA	transcription factor TGA	122.47	266.52	-1.12
Jcr4S00145	8269	11499	K14431 TGA	transcription factor TGA	566.41	489.93	0.21
Jcr4S14409	6054	7588	K14432 ABF	ABA responsive element binding factor	67.36	39.19	0.78
Jcr4S00146	93078	95162	K14432 ABF	ABA responsive element binding factor	124.76	67.94	0.88

Jcr4S00154	64235	64879	K14516 ERF1	ethylene-responsive transcription factor 1	29.09	14.37	1.02
Jcr4S15311	6364	7453	K14432 ABF	ABA responsive element binding factor	38.27	11.76	1.7
Jcr4S02741	20442	26799	K14431 TGA	transcription factor TGA	212.02	205.12	0.05
Jcr4S03002	5880	6984	K14432 ABF	ABA responsive element binding factor	65.83	57.49	0.2
Jcr4S03023	22492	23151	K14516 ERF1	ethylene-responsive transcription factor 1	30.62	18.29	0.74
Jcr4S03571	6568	8595	K13422 MYC2	transcription factor MYC2	147.72	49.65	1.57
Jcr4S04929	9847	12712	K14431 TGA	transcription factor TGA	45.92	116.28	-1.34
Jcr4S05391	3741	4631	K14432 ABF	ABA responsive element binding factor	63.53	41.81	0.6
Jcr4S05882	6152	7347	K14432 ABF	ABA responsive element binding factor	51.28	27.44	0.9
Jcr4S06409	7418	9034	K14432 ABF	ABA responsive element binding factor	26.02	78.39	-1.59
Jcr4S08050	93	4590	K14431 TGA	transcription factor TGA	76.54	91.45	-0.26
Jcr4U29916	5286	6396	K14432 ABF	ABA responsive element binding factor	61.23	22.21	1.46

Table A10 Genes associated with 'Photosynthesis' downregulated in JV

Jatropha Scaffold Ids	**CDS Start Position**	**CDS End Position**	**KO IDS**	**Enzyme Name**	**JV FPKM**	**JH FPKM**	**log2 Fold Change**
Jcr4S00002	84410	87459	K14172 LHCB7	light-harvesting complex II chlorophyll a/b binding protein 7	56.64	84.92	-0.58
Jcr4S00013	121663	122920	K02716 psbO	photosystem II oxygen-evolving enhancer protein 1	23.73	32.66	-0.46
Jcr4S00132	96270	97760	K08908 LHCA2	light-harvesting complex I chlorophyll a/b binding protein 2	23.73	39.19	-0.72
Jcr4S01334	19013	19681	K02693 psaE	photosystem I subunit IV	13.78	32.66	-1.25
Jcr4S13852	2500	2970	K02721 psbW	photosystem II PsbW protein	4.59	13.06	-1.51
Jcr4S14329	6495	7338	K08917 LHCB6	light-harvesting complex II chlorophyll a/b binding protein 6	11.48	13.06	-0.19
Jcr4S14384	3851	4790	K14332 psaO	photosystem I subunit PsaO	19.9	27.44	-0.46
Jcr4S01478	30103	31246	K02695 psaH	photosystem I subunit VI	26.79	32.66	-0.29
Jcr4S00152	27207	28163	K08915 LHCB4	light-harvesting complex II chlorophyll a/b binding protein 4	20.67	28.74	-0.48
Jcr4S15558	2673	2915	K02691 psaC	photosystem I subunit VII	328.36	3323.69	-3.34
Jcr4S16042	1	493	K02705 psbC	photosystem II CP43 chlorophyll apoprotein	2874.89	8390.23	-1.55

Jcr4S17180	1394	3298	K02636 petC	cytochrome b6-f complex iron-sulfur subunit [EC:1.10.9.1]	24.49	88.84	-1.86
Jcr4S17467	6002	8224	K02717 psbP	photosystem II oxygen-evolving enhancer protein 2	30.62	66.63	-1.12
Jcr4S01794	19047	20251	K08914 LHCB3	light-harvesting complex II chlorophyll a/b binding protein 3	21.43	43.11	-1.01
Jcr4S01878	4131	5206	K08907 LHCA1	light-harvesting complex I chlorophyll a/b binding protein 1	38.27	48.34	-0.34
Jcr4S19524	587	931	K02634 petA	apocytochrome f	56.64	57.49	-0.02
Jcr4S20322	1	350	K02705 psbC	photosystem II CP43 chlorophyll apoprotein	512.06	2680.9	-2.39
Jcr4S00217	73371	74542	K08909 LHCA3	light-harvesting complex I chlorophyll a/b binding protein 3	21.43	27.44	-0.36
Jcr4S22203	875	1174	K02639 petF	ferredoxin	3.83	26.13	-2.77
Jcr4S26036	1238	2201	K08906 petJ	cytochrome c6	25.26	52.26	-1.05
Jcr4S26836	123	2347	K08913 LHCB2	light-harvesting complex II chlorophyll a/b binding protein 2	22.2	65.32	-1.56
Jcr4S02812	132	9944	K02712 psbK	photosystem II PsbK protein	580.95	2403.93	-2.05
Jcr4S28596	451	826	K02689 psaA	photosystem I P700 chlorophyll a apoprotein A1	841.19	8382.39	-3.32
Jcr4S00004	59679	60185	K02638 petE	plastocyanin	7.65	31.36	-2.03
Jcr4S00313	61165	62078	K08910 LHCA4	light-harvesting complex I chlorophyll a/b binding protein 4	15.31	33.97	-1.15
Jcr4S00033	37885	38319	K02639 petF	ferredoxin	13.01	13.06	-0.01
Jcr4S03380	9268	13828	K02639 petF	ferredoxin	391.89	476.87	-0.28
Jcr4S03698	1040	2765	K03542 psbS	photosystem II 22kDa protein	58.94	94.07	-0.67
Jcr4S04672	11961	14176	K03541 psbR	photosystem II 10kDa protein	69.65	113.66	-0.71
Jcr4S00006	31263	33359	K02641 petH	ferredoxin--NADP+ reductase [EC:1.18.1.2]	54.34	91.45	-0.75
Jcr4S05338	5245	8028	K02717 psbP	photosystem II oxygen-evolving enhancer protein 2	39.04	88.84	-1.19
Jcr4S05614	8793	9507	K02701 psaN	photosystem I subunit PsaN	9.95	22.21	-1.16
Jcr4S05689	1097	2882	K08915 LHCB4	light-harvesting complex II chlorophyll a/b binding protein 4	23.73	58.79	-1.31
Jcr4S00600	39881	40402	K08902 psb27	photosystem II Psb27 protein	11.48	11.76	-0.03
Jcr4S00072	51797	52989	K02698 psaK	photosystem I subunit X	17.6	57.49	-1.71

Jcr4S07335	4710	8417	K02641 petH	ferredoxin--NADP+ reductase [EC:1.18.1.2]	114.05	125.42	-0.14
Jcr4S07377	23712	24402	K02721 psbW	photosystem II PsbW protein	15.31	20.9	-0.45
Jcr4S07828	13202	15418	K08913 LHCB2	light-harvesting complex II chlorophyll a/b binding protein 2	38.27	62.71	-0.71
Jcr4S08028	3523	4270	K02721 psbW	photosystem II PsbW protein	30.62	57.49	-0.91
Jcr4S00866	40924	41361	K02639 petF	ferredoxin	9.18	22.21	-1.27
Jcr4S09374	3850	5346	K02636 petC	cytochrome b6-f complex iron-sulfur subunit [EC:1.10.9.1]	176.05	224.71	-0.35
Jcr4S09385	6575	6685	K02696 psaI	photosystem I subunit VIII	96.44	1741.54	-4.17
Jcr4S09385	9252	9483	K02634 petA	apocytochrome f	515.12	2471.86	-2.26
Jcr4U29440	6161	7100	K14332 psaO	photosystem I subunit PsaO	18.37	23.52	-0.36
Jcr4U29703	360	626	K02635 petB	cytochrome b6	623.81	1988.47	-1.67
Jcr4U30584	53	526	K08905 psaG	photosystem I subunit V	3.83	15.68	-2.03
Jcr4U31389	1974	2360	K02635 petB	cytochrome b6	1491.03	2970.94	-0.99
Jcr4U31990	869	1291	K02704 psbB	photosystem II CP47 chlorophyll apoprotein	3046.35	3215.25	-0.08
Jcr4U33013	640	849	K02637 petD	cytochrome b6-f complex subunit 4	261.01	2586.84	-3.31
Jcr4U38315	390	1280	K02704 psbB	photosystem II CP47 chlorophyll apoprotein	1697.69	2534.58	-0.58

Table A11 Genes associated with 'Anthocyanin biosynthesis' downregulated in JV

Jatropha Scaffold Ids	CDS Start Position	CDS End Position	KO IDS	Enzyme Name	JV FPKM	JH FPKM	log2 Fold Change
Jcr4S13230	4929	6311	K17193 UGT79B1	anthocyanidin 3-O-glucoside 2'''-O-xylosyltransferase [EC:2.4.2.51]	43.63	56.18	-0.36
Jcr4S16590	4007	5374	K12338 UGT75C1	anthocyanidin 3-O-glucoside 5-O-glucosyltransferase [EC:2.4.1.298]	22.2	137.18	-2.63
Jcr4S19550	3735	5102	K12338 UGT75C1	anthocyanidin 3-O-glucoside 5-O-glucosyltransferase [EC:2.4.1.298]	19.14	147.63	-2.95

Table A12 Genes associated to 'Plant-pathogen interaction' downregulated in JV

Jatropha scaffold Id	CDS start position	CDS end position	Name	JH FPKM	JV FPKM	Log2 fold change
Jcr4S00002	157545	158126	calcium-binding protein CML	18.37	10.45	-0.81
Jcr4S00101	58922	59497	calcium-binding protein CML	15.31	3.92	-1.97
Jcr4S10091	1	2205	cyclic nucleotide gated channel, other eukaryote	39.8	31.36	-0.34
Jcr4S01083	7460	10267	disease resistance protein RPS2	151.55	100.6	-0.59
Jcr4S00111	47202	47627	calcium-binding protein CML	32.15	24.82	-0.37
Jcr4S11691	2676	3317	calcium-binding protein CML	22.2	11.76	-0.92
Jcr4S01180	24764	27674	calcium-dependent protein kinase [EC:2.7.11.1]	119.4	101.91	-0.23
Jcr4S01203	15702	16127	calcium-binding protein CML	55.11	26.13	-1.08
Jcr4S12069	183	1699	disease resistance protein RPS2	111.75	73.16	-0.61
Jcr4S01285	38654	39229	calcium-binding protein CML	36.74	9.15	-2.01
Jcr4S14092	12433	15033	cyclic nucleotide gated channel, other eukaryote	80.37	74.47	-0.11
Jcr4S01445	7158	8903	calcium-dependent protein kinase [EC:2.7.11.1]	117.87	54.87	-1.1
Jcr4S01640	8133	11755	cyclic nucleotide gated channel, other eukaryote	130.89	91.45	-0.52
Jcr4S00171	75142	76959	disease resistance protein RPS2	83.43	53.57	-0.64
Jcr4S17275	189	2686	chitin elicitor receptor kinase 1	68.12	67.94	0
Jcr4S00174	26608	27267	calcium-binding protein CML	29.85	14.37	-1.05
Jcr4S01747	996	1550	calcium-binding protein CML	27.55	10.45	-1.4
Jcr4S00003	194254	198274	calcium-dependent protein kinase [EC:2.7.11.1]	159.21	108.44	-0.55

Jcr4S02024	5487	6712	mitogen-activated protein kinase kinase kinase 1, plant [EC:2.7.11.25]	24.49	15.68	-0.64
Jcr4S21798	64	2275	disease resistance protein RPS2	91.85	60.1	-0.61
Jcr4S02271	7258	8883	serine/threonine-protein kinase PBS1 [EC:2.7.11.1]	39.8	15.68	-1.34
Jcr4S22561	416	3262	disease resistance protein RPM1	67.36	40.5	-0.73
Jcr4S02470	40255	43035	disease resistance protein RPM1	122.47	88.84	-0.46
Jcr4S02486	15611	17230	serine/threonine-protein kinase PBS1 [EC:2.7.11.1]	10.72	3.92	-1.45
Jcr4S02582	22199	25829	calcium-dependent protein kinase [EC:2.7.11.1]	135.48	103.21	-0.39
Jcr4S27090	4320	7166	disease resistance protein RPM1	53.58	48.34	-0.15
Jcr4S03361	11574	14429	disease resistance protein RPS2	157.68	95.37	-0.73
Jcr4S03605	17984	18614	RPM1-interacting protein 4	49.75	16.98	-1.55
Jcr4S00005	46802	49042	disease resistance protein RPS2	127.06	104.52	-0.43
Jcr4S00433	525	1151	chitin elicitor receptor kinase 1	13.01	6.53	-0.99
Jcr4S04750	7258	7812	calcium-binding protein CML	26.02	16.98	-0.62
Jcr4S04907	27089	31426	calcium-dependent protein kinase [EC:2.7.11.1]	171.45	150.25	-0.19
Jcr4S00578	57	6985	mitogen-activated protein kinase kinase kinase 1, plant [EC:2.7.11.25]	270.96	244.31	-0.15
Jcr4S06013	1409	10232	calcium-dependent protein kinase [EC:2.7.11.1]	352.09	314.86	-0.16
Jcr4S00670	2683	6338	cyclic nucleotide gated channel, other eukaryote	133.95	92.76	-0.53
Jcr4S07200	289	5187	calcium-dependent protein kinase [EC:2.7.11.1]	117.11	66.63	-0.81
Jcr4S07424	13716	13970	calcium-binding protein CML	15.31	10.45	-0.55
Jcr4S08110	15382	18387	disease resistance protein RPM1	38.27	31.36	-0.29
Jcr4S00083	41069	41512	calcium-binding protein CML	26.02	7.84	-1.73
Jcr4S09427	5674	9961	cyclic nucleotide gated channel, other eukaryote	173.75	164.62	-0.08
Jcr4S09723	4570	5211	calcium-binding protein CML	26.79	13.06	-1.04
Jcr4S01613	1080	1829	WRKY transcription factor 29	167.63	28.74	-2.54
Jcr4S04113	4066	6404	WRKY transcription factor 33	104.52	78.84	-0.41
Jcr4S04574	11524	11940	WRKY transcription factor 22	9.18	5.23	-0.81
Jcr4S00770	50527	51950	WRKY transcription factor 22	91.08	56.18	-0.7

Table A13 Genes associated to 'Calcium signaling pathway' downregulated in JV

Jatropha Scaffold Ids	CDS Start Position	CDS End Position	KO IDS	Enzyme Name	JV FPKM	JH FPKM	log2 Fold Change
Jcr4S00017	84189	85385	K05863 SLC25A4S, ANT	solute carrier family 25 (mitochondrial adenine nucleotide translocator), member 4/5/6/31	3.83	31.36	-3.03
Jcr4S03161	17843	19209	K15040 VDAC2	voltage-dependent anion channel protein 2	37.51	45.73	-0.29
Jcr4S00335	36801	37409	K05863 SLC25A4S, ANT	solute carrier family 25 (mitochondrial adenine nucleotide translocator), member 4/5/6/31	14.54	19.6	-0.43
Jcr4S03986	1858	5130	K15040 VDAC2	voltage-dependent anion channel protein 2	100.6	104.86	-0.06
Jcr4S05688	1177	4787	K15040 VDAC2	voltage-dependent anion channel protein 2	80.37	103.21	-0.36
Jcr4S07364	12688	15415	K15040 VDAC2	voltage-dependent anion channel protein 2	64.02	64.29	-0.01
Jcr4S00084	62244	65279	K15040 VDAC2	voltage-dependent anion channel protein 2	122.47	147.63	-0.27

Table A14 *In-silico* (RSEM based) transcript abundance of identified NBS-LRR genes of *J. curcas*

Gene_id	Length	Effective_length	pme_TPM	IsoPct_from_pme_TPM	TPM_ci_lower_bound	TPM_ci_upper_bound
JcNL_14680_length=307_numreads=5	307	10.54	133.54	100	0.00580055	403.791
JcNL_14712_length=310_numreads=19	310	10.95	128.5	100	0.0105725	382.973
JcNL_14601_length=324_numreads=13	324	13.06	107.75	100	0.00129076	323.54
JcNL_13997_length=393_numreads=7	393	29.65	47.46	100	0.00508882	142.696
JcNL_13460_length=441_numreads=3	441	50.89	27.65	100	2.91818e-05	82.0624
JcNL_12889_length=460_numreads=4	460	62.23	22.61	100	2.03423e-05	67.7356
JcNL_12635_length=472_numreads=5	472	70.23	20.04	100	0.000377925	59.8527
JcNL_11632_length=505_numreads=4	505	95.5	14.73	100	0.000319989	44.3365
JcNL_11173_length=519_numreads=5	519	107.6	13.08	100	0.000162594	39.6457
JcNL_10695_length=539_numreads=7	539	125.92	11.17	100	2.00393e-05	34.0056
JcNL_10302_length=553_numreads=17	553	139.19	10.11	100	7.09448e-05	30.4698
JcNL_10121_length=566_numreads=12	566	151.67	9.28	100	5.57474e-05	27.5898
JcNL_10049_length=568_numreads=7	568	153.6	9.16	100	4.94181e-05	27.6008
JcNL_09636_length=587_numreads=8	587	172.04	8.18	100	0.00013121	24.692
JcNL_09428_length=596_numreads=5	596	180.81	7.78	100	1.69391e-05	23.1942
JcNL_09317_length=603_numreads=9	603	187.65	7.5	100	2.31247e-06	22.442

JcNL_09160_length=612_numreads=24	612	196.46	7.16	100	2.08537e-05	21.5462
JcNL_08981_length=627_numreads=17	627	211.19	6.66	100	0.000118978	19.9194
JcNL_07289_length=754_numreads=11	754	337.12	4.17	100	0.00011408	12.3718
JcNL_06992_length=777_numreads=5	777	360.07	3.91	100	0.000157816	11.7091
JcNL_06396_length=846_numreads=8	846	428.98	3.28	100	1.23318e-05	9.79747
JcNL_04926_length=1050_numreads=9	1050	632.95	2.22	100	6.58012e-05	6.67745
JcNL_04778_length=1073_numreads=22	1073	655.95	2.15	100	1.74392e-05	6.38345
JcNL_04766_length=1079_numreads=68	1079	661.95	2.13	100	2.4907e-05	6.38538
JcNL_04489_length=1129_numreads=16	1129	711.95	1.98	100	8.75788e-06	5.95732
JcNL_04456_length=1133_numreads=19	1133	715.95	1.97	100	4.57821e-08	5.90379
JcNL_04261_length=1181_numreads=47	1181	763.95	1.84	100	2.12905e-05	5.50927
JcNL_03152_length=1472_numreads=26	1472	1054.95	1.33	100	4.63661e-05	4.04954
JcNL_03063_length=1508_numreads=19	1508	1090.95	1.29	100	1.44142e-05	3.9085
JcNL_00587_length=2840_numreads=69	2840	2422.95	1.16	100	0.0369321	2.7684
JcNL_02545_length=1691_numreads=26	1691	1273.95	1.1	100	1.64358e-05	3.30141
JcNL_02388_length=1752_numreads=76	1752	1334.95	1.05	100	2.86513e-05	3.14167
JcNL_01427_length=2172_numreads=25	2172	1754.95	0.8	100	2.59394e-05	2.39852
JcNL_01339_length=2256_numreads=36	2256	1838.95	0.77	100	7.22727e-06	2.28769
JcNL_00860_length=2564_numreads=35	2564	2146.95	0.66	100	3.82679e-05	1.96691
JcNL_00810_length=2620_numreads=86	2620	2202.95	0.64	100	1.98976e-05	1.88884
JcNL_00618_length=2812_numreads=39	2812	2394.95	0.59	100	9.2815e-07	1.76707
JcNL_00553_length=2890_numreads=173	2890	2472.95	0.57	100	1.98547e-06	1.70797
JcNL_00565_length=2878_numreads=61	2878	2460.95	0.57	100	6.56487e-05	1.73699
JcNL_00259_length=3355_numreads=62	3355	2937.95	0.48	100	7.96725e-07	1.43408
JcNL_00134_length=3683_numreads=98	3683	3265.95	0.43	100	2.45022e-06	1.29085
JcNL_00090_length=3850_numreads=68	3850	3432.95	0.41	100	6.80228e-07	1.23454
JcNL_00096_length=3826_numreads=144	3826	3408.95	0.41	100	1.65211e-05	1.2473
JcNL_00073_length=3933_numreads=192	3933	3515.95	0.4	100	1.32675e-05	1.20332
JcNL_16532_length=155_numreads=6	155	0	0	0	0	0
JcNL_16899_length=125_numreads=11	125	0	0	0	0	0
JcNL_17188_length=117_numreads=11	117	0	0	0	0	0

Table A15 *In-silico* (RSEM based) transcript abundance of identified defense response related transcription factors of *J. curcas*

TF_id	Length	Effective_length	pme_TPM	IsoPct_from_pme_TPM
JcTF_00115_length=3742_numreads=67	3742	3324.95	0.42	100
JcTF_00439_length=3045_numreads=90	3045	2627.95	0.54	100
JcTF_00764_length=2650_numreads=47	2650	2232.95	1.26	100
JcTF_00785_length=2637_numreads=70	2637	2219.95	0.63	100
JcTF_00880_length=2558_numreads=51	2558	2140.95	0.66	100
JcTF_00918_length=2534_numreads=95	2534	2116.95	1.33	100
JcTF_00925_length=2522_numreads=67	2522	2104.95	0.67	100
JcTF_00978_length=2480_numreads=29	2480	2062.95	0.68	100
JcTF_01151_length=2360_numreads=44	2360	1942.95	1.45	100
JcTF_01250_length=2304_numreads=46	2304	1886.95	0.75	100
JcTF_01393_length=2210_numreads=25	2210	1792.95	0.78	100
JcTF_01461_length=2167_numreads=46	2167	1749.95	0.8	100
JcTF_01506_length=2144_numreads=31	2144	1726.95	0.81	100
JcTF_01687_length=2044_numreads=35	2044	1626.95	0.86	100
JcTF_01737_length=2020_numreads=65	2020	1602.95	0.88	100
JcTF_01778_length=2000_numreads=72	2000	1582.95	1.78	100
JcTF_01856_length=1963_numreads=62	1963	1545.95	0.91	100
JcTF_01981_length=1907_numreads=35	1907	1489.95	0.94	100
JcTF_01982_length=1914_numreads=70	1914	1496.95	0.94	100
JcTF_02046_length=1878_numreads=42	1878	1460.95	0.96	100
JcTF_02082_length=1857_numreads=38	1857	1439.95	0.98	100
JcTF_02191_length=1823_numreads=34	1823	1405.95	1	100
JcTF_02233_length=1806_numreads=24	1806	1388.95	1.01	100
JcTF_02341_length=1761_numreads=29	1761	1343.95	1.05	100
JcTF_02499_length=1707_numreads=31	1707	1289.95	1.09	100
JcTF_02772_length=1615_numreads=46	1615	1197.95	1.17	100
JcTF_02905_length=1561_numreads=26	1561	1143.95	1.23	100
JcTF_02936_length=1548_numreads=68	1548	1130.95	2.49	100
JcTF_03033_length=1520_numreads=45	1520	1102.95	1.28	100
JcTF_03086_length=1501_numreads=51	1501	1083.95	1.3	100
JcTF_03225_length=1458_numreads=48	1458	1040.95	1.35	100
JcTF_03268_length=1445_numreads=22	1445	1027.95	1.37	100

JcTF_03301_length=1434_numreads=20	1434	1016.95	1.38	100
JcTF_03314_length=1432_numreads=23	1432	1014.95	1.39	100
JcTF_03326_length=1427_numreads=31	1427	1009.95	1.39	100
JcTF_03366_length=1415_numreads=22	1415	997.95	1.41	100
JcTF_03375_length=1415_numreads=32	1415	997.95	1.41	100
JcTF_03662_length=1331_numreads=13	1331	913.95	1.54	100
JcTF_03682_length=1313_numreads=21	1313	895.95	1.57	100
JcTF_03739_length=1303_numreads=24	1303	885.95	1.59	100
JcTF_04051_length=1227_numreads=14	1227	809.95	1.74	100
JcTF_04119_length=1213_numreads=47	1213	795.95	1.77	100
JcTF_04123_length=1209_numreads=21	1209	791.95	1.78	100
JcTF_04163_length=1201_numreads=18	1201	783.95	1.79	100
JcTF_04169_length=1196_numreads=23	1196	778.95	1.81	100
JcTF_04237_length=1188_numreads=17	1188	770.95	1.83	100
JcTF_04311_length=1170_numreads=20	1170	752.95	1.87	100
JcTF_04330_length=1164_numreads=43	1164	746.95	1.88	100
JcTF_04381_length=1155_numreads=15	1155	737.95	1.91	100
JcTF_04386_length=1155_numreads=18	1155	737.95	1.91	100
JcTF_04420_length=1148_numreads=20	1148	730.95	1.93	100
JcTF_04429_length=1142_numreads=15	1142	724.95	1.94	100
JcTF_04493_length=1127_numreads=21	1127	709.95	1.98	100
JcTF_04559_length=1114_numreads=23	1114	696.95	2.02	100
JcTF_04744_length=1076_numreads=14	1076	658.95	2.14	100
JcTF_04775_length=1076_numreads=19	1076	658.95	2.14	100
JcTF_04781_length=1072_numreads=13	1072	654.95	2.15	100
JcTF_04854_length=1066_numreads=19	1066	648.95	2.17	100
JcTF_05008_length=1037_numreads=15	1037	619.95	2.27	100
JcTF_05024_length=1035_numreads=9	1035	617.95	2.28	100
JcTF_05026_length=1036_numreads=14	1036	618.95	2.27	100
JcTF_05149_length=1017_numreads=27	1017	599.95	2.35	100
JcTF_05255_length=997_numreads=16	997	579.95	2.43	100
JcTF_05708_length=929_numreads=17	929	511.95	2.75	100
JcTF_05776_length=921_numreads=28	921	503.95	2.79	100
JcTF_05837_length=914_numreads=10	914	496.95	2.83	100
JcTF_05940_length=898_numreads=24	898	480.96	2.93	100

JcTF_05953_length=897_numreads=21	897	479.96	2.93	100
JcTF_06110_length=874_numreads=15	874	456.96	3.08	100
JcTF_06232_length=862_numreads=9	862	444.97	3.16	100
JcTF_06247_length=859_numreads=9	859	441.97	3.18	100
JcTF_06299_length=855_numreads=15	855	437.97	3.21	100
JcTF_06326_length=852_numreads=8	852	434.97	3.23	100
JcTF_06488_length=840_numreads=18	840	422.98	3.33	100
JcTF_06682_length=817_numreads=16	817	400.01	3.52	100
JcTF_06768_length=810_numreads=8	810	393.02	3.58	100
JcTF_06868_length=797_numreads=12	797	380.03	3.7	100
JcTF_06920_length=791_numreads=9	791	374.04	3.76	100
JcTF_07041_length=780_numreads=11	780	363.06	3.88	100
JcTF_07136_length=771_numreads=9	771	354.08	3.97	100
JcTF_07165_length=769_numreads=16	769	352.08	4	100
JcTF_07449_length=743_numreads=9	743	326.15	4.31	100
JcTF_07616_length=725_numreads=9	725	308.22	4.57	100
JcTF_07950_length=685_numreads=5	685	268.47	5.24	100
JcTF_08075_length=689_numreads=9	689	272.43	5.16	100
JcTF_08297_length=670_numreads=8	670	253.61	5.55	100
JcTF_08498_length=649_numreads=6	649	232.85	6.04	100
JcTF_08810_length=637_numreads=6	637	221.02	6.37	100
JcTF_08950_length=623_numreads=5	623	207.26	6.79	100
JcTF_09148_length=613_numreads=11	613	197.44	7.13	100
JcTF_09242_length=608_numreads=5	608	192.54	7.31	100
JcTF_09345_length=605_numreads=13	605	189.61	7.42	100
JcTF_09429_length=598_numreads=6	598	182.76	7.7	100
JcTF_09745_length=582_numreads=10	582	167.18	8.42	100
JcTF_09830_length=578_numreads=6	578	163.29	8.62	100
JcTF_10040_length=569_numreads=9	569	154.56	9.1	100
JcTF_10318_length=554_numreads=10	554	140.14	10.04	100
JcTF_10466_length=544_numreads=4	544	130.63	10.77	100
JcTF_10501_length=548_numreads=10	548	134.42	10.47	100
JcTF_10538_length=543_numreads=11	543	129.69	10.85	100
JcTF_10663_length=539_numreads=11	539	125.92	11.17	100
JcTF_11285_length=515_numreads=9	515	104.07	13.52	100

JcTF_11470_length=509_numreads=11	509	98.88	14.23	100
JcTF_11496_length=509_numreads=8	509	98.88	14.23	100
JcTF_11596_length=504_numreads=4	504	94.67	14.86	100
JcTF_11676_length=500_numreads=3	500	91.37	15.4	100
JcTF_11993_length=491_numreads=5	491	84.19	16.71	100
JcTF_12223_length=484_numreads=5	484	78.86	17.84	100
JcTF_12401_length=480_numreads=5	480	75.91	18.54	100
JcTF_12486_length=477_numreads=6	477	73.75	19.08	100
JcTF_12555_length=474_numreads=3	474	71.63	19.64	100
JcTF_12662_length=472_numreads=10	472	70.23	20.04	100
JcTF_12677_length=461_numreads=5	461	62.87	22.38	100
JcTF_12728_length=468_numreads=5	468	67.49	20.85	100
JcTF_12732_length=468_numreads=4	468	67.49	20.85	100
JcTF_13181_length=453_numreads=10	453	57.87	24.31	100
JcTF_13463_length=442_numreads=7	442	51.45	27.35	100
JcTF_13501_length=439_numreads=4	439	49.79	28.26	100
JcTF_14789_length=302_numreads=3	302	9.88	142.48	100
JcTF_14930_length=290_numreads=4	290	8.42	167.1	100
JcTF_15218_length=263_numreads=12	263	5.79	243.07	100
JcTF_15319_length=251_numreads=3	251	4.86	289.67	100

Table A16 Position of disease resistance genes in contigs of *J .curcas*

Accession no.	Gene length	Contig	Contig length	Identity	Subgroup	Domains
JHL06P13.14	6835	JcCB0017361	6974	2968/2968 (100%)	CNL	RPW8/ LRRs/ NB-ARC/ P-loop NTPase/ TIRs
JHL06P13.15	3140	JcCB0017361	6974	3140/3140 (100%)	TNL	
JHL25H03.3	3264	JcCA0063341	5033	3209/3271 (98%)	CNL	
JHL25H03.4	2429	JcCA0009611	6833	1926/1926 (100%)	CNL	
JHL25H03.6	959	JcCB0795021	2325	920/929 (99%)	CNL	
JHL25P11.3	2832	JcCA0312791	12484	2553/2558 (99%)	TNL	
JHL25P11.7	2557	JcCA0312791	12484	2539/2557 (99%)	TNL	
JMS10C05.7	2422	JcCA0150781	13611	2422/2422 (100%)	CNL	
JHS03A10.2	3688	JcCA0076681	8352	3669/3689 (99%)	TNL	
XP 002867519.1	2868	JcCB0142551	3139	21/21 (100%)	TNL	RPW8/ LRRs/ NB-ARC/ P-loop NTPase/ TIRs/ RT-LTRs/ AAA+/ NACHT/ AMN1
XP 002865044.1	2433	JcCB0496021	2792	21/21 (100%)	CNL	
XP 002871778.1	3846	JcCA0021521	5942	25/25 (100%)	TNL	
BAB08845.1	2703	JcCA0009611	6833	35/40 (87%)	TNL	
NP 192816.1	2679	JcCB0139241	7305	21/21 (100%)	TNL	
AAK96709.1	2670	JcCA0079961	14578	21/21 (100%)	TNL	
XP 002884622.1	2736	JcCB0095731	3289	20/20 (100%)	TNL	
AAC31552.1*						
CAJ26369.1	2817	JcCB0059401	7830	24/25 (96%)	TNL	
ACP30565.1	3030	JcCB0094831	4421	22/22 (100%)	TNL	
ACP30573.1	3117	JcCB0521921	4485	20/20 (100%)	CNL	
ACP30609.1	8181	JcCA0046251	20074	44/51 (86%)	TNL	
ACP30637.1	2667	JcCB0182971	4157	23/24 (95%)	TNL	
ACP30557.1	2661	JcCB0048231	3367	21/21 (100%)	TNL	
ACP30621.1	2442	JcCB0078131	3735	21/21 (100%)	TNL	
AAC99466.1	2781	JcCB0000491	19087	22/22 (100%)	TNL	
AAN62353.1	2673	JcCB0123681	3177	22/22 (100%)	TNL	
AAN62350.1	2670	JcCA0153621	8895	24/24 (100%)	TNL	
AAO45748.1	3279	JcCA0312851	8889	30/32 (93%)	TNL	
ACN78965.1	9168	JcCA0030091	10362	25/25 (100%)	TNL	
ACM89637.1	2805	JcCA0312791	12484	50/59 (84%)	TNL	
ACJ37419.1	2730	JcCA0315571	9082	30/32 (93%)	CNL	
AAB96976.1*						
CAD45036.1	2824	JcCA0306451	7583	23/24 (95%)	CNL	
ABO15685.1	2667	JcCB0078311	5034	20/20 (100%)	CNL	
AAD03671.1	5474	JcCB0064381	6817	63/65 (96%)	CNL	
AAK28803.1	3603	JcCB0087091	4092	27/28 (96%)	TNL	
ACG70794.1	2442	JcCB0005211	7867	24/25 (96%)	TNL	
ADL36726.1	2730	JcCA0152071	8836	28/30 (93%)	TNL	
CAJ44364.1	2079	JcCB0228541	3269	32/35 (91%)	CNL	
ABD32877.1*						
AAN62760.1	2115	JcCA0006542	5033	20/20 (100%)	TNL	

ABN05946.1	3693	JcCB0495471	7361	33/37 (89%)	CNL	
ABD32335.1	2421	JcCB0048231	3367	27/29 (93%)	TNL	
ABN08495.1	2562	JcCA0300101	5605	24/25 (96%)	TNL	
AAY54606.1	2553	JcCB0423811	3013	21/21 (100%)	CNL	
EAY80918.1*						
EAY76104.1*						
EEE69427.1*						
NP 001045443.2	3021	JcCA0044491	14390	25/26 (96%)	CNL	
NP 001061834.1	2724	JcCB0483581	3402	20/20 (100%)	TNL	
EEE68711.1*						
EEE68252.1*						
ABH07384.1	3402	JcCB0039671	8353	39/44 (88%)	TNL	
ADB85254.1	2544	JcCB0001791	16667	23/24 (95%)	CNL	
ACN40007.1	2181	JcCA0063341	5033	24/25 (96%)	CNL	
XP 002332952.1	3948	JcCA0136661	5709	35/39 (89%)	CNL	
XP 002310744.1	2415	JcCB0033951	7029	25/26 (96%)	CNL	
XP 002329169.1	3378	JcCB0495471	7361	37/39 (94%)	TNL	
XP 002298700.1	2643	JcCB0182971	4157	27/28 (96%)	CNL	
XP 002318943.1	2463	JcCA0063341	5033	24/25 (96%)	CNL	
XP 002326562.1	2847	JcCA0028981	11361	32/35 (91%)	TNL	
XP 002333530.1	2526	JcCB0113161	6841	33/36 (91%)	CNL	
AAT09451.1	2754	JcCA0312791	12484	23/23 (100%)	TNL	
ACR19031.1	3162	JcCA0288161	5378	32/35 (91%)	TNL	
XP 002517608.1**	4977	JcCB0055521	5952	21/21 (100%)	TNL	
XP 002517607.1**	2421	JcCB0028271	8578	28/30 (93%)	TNL	
XP 002518711.1**	3330	JcCA0267961	6949	27/28 (96%)	TNL	
XP 002517572.1**	2439	JcCA0076691	3941	21/21 (100%)	TNL	
XP 002517594.1**	2733	JcCA0009611	6833	31/34 (91%)	TNL	
XP 002521805.1**	2832	JcCA0317171	12462	20/20 (100%)	TNL	
XP 002529624.1**	2583	JcCB0008231	9369	28/30 (93%)	TNL	
ABM30222.2	2661	JcCA0312791	12484	25/26 (96%)	TNL	
AAP44390.1	3411	JcCB0039671	8353	23/23 (100%)	CNL	
AAP45164.2	2823	JcCA0020751	7774	27/28 (96%)	TNL	
XP 002458619.1	2724	JcPR03ENPT2	2924	21/21 (100%)	TNL	
XP 002442418.1	2616	JcCB0033951	7029	20/20 (100%)	CNL	
XP 002448182.1	2841	JcCA0077731	14390	20/20 (100%)	TNL	
XP 002437863.1	2865	JcCB0144411	3416	20/20 (100%)	TNL	
XP 002438943.1	2790	JcCB0138841	4395	31/34 (91%)	TNL	
XP 002438921.1	2493	JcCB0083141	5901	22/22 (100%)	TNL	
AAK20742.1	2781	JcCB0025091	7569	21/21 (100%)	CNL	
CAN83232.1	5085	JcCA0133581	5304	22/22 (100%)	CNL	
XP 002276590.1	2472	JcCB0017361	6974	28/30 (93%)	TNL	
XP 002264046.1	5463	JcCB0039671	8353	39/41 (95%)	CNL	
XP 002263674.1	2712	JcCB0142741	4257	21/21 (100%)	TNL	
XP 002281592.1	2682	JcCA0009611	6833	39/44 (88%)	TNL	
CAN74463.1	2796	JcCB0059401	7830	36/41 (87%)	TNL	

XP 002278041.1	2550	JcCB0086751	4747	35/39 (89%)	TNL	
AAX31149.1	2718	JcCA0213171	3666	21/21 (100%)	TNL	
NP 001105809.1	2730	JcCA0312791	12484	25/26 (96%)	CNL	

*No significant hits; **Similarity to castor bean disease resistance genes

REFERENCES

[1] J. M. Dyer, S. Stymne, A. G. Green, and A. S. Carlsson, "High-value oils from plants," *The Plant Journal*, vol. 54, pp. 640–655, 2008.

[2] D. Fairless, "Biofuel: the little shrub that could – maybe," *Nature*, vol. 449, pp. 652–655, 2007.

[3] M. Ricci, P. Bellaby, and R. Flynn, "Stakeholders' and publics' perceptions of hydrogen energy technologies," in *Risk and the Public Acceptance of New Technologies*, R. Flynn and P. Bellaby, Eds. New York: Palgrave Macmillan, 2007, pp. 175-197.

[4] V. Patil and K. Singh, "Oil gloom to oil book – *Jatropha curcas*," Agroforestry Federation, Maharashtra and National Wastelands Development Mission. MoEF, 1991, p. 81.

[5] J. Heller, "Physic nut. *Jatropha curcas* L," in "Promoting the conservation and use of underutilized and neglected crops. (Prom Underused Crops)," International Plant Genetic Resources Institute (IPGRI), vol. 1, pp. 1–66, 1996.

[6] M. Sujatha and A. J. Prabakaran, "New ornamental Jatropha hybrids through interspecific hybridization," *Genetic Resources Crop Evolution*, vol. 50, pp. 75–82, 2003.

[7] M. Debnath and P. S. Bisen, "*Jatropha curcas* L., a multipurpose stress resistant plant with a potential for ethnomedicine and renewable energy," *Current Pharmaceutical Biotechnology*, vol. 9, pp. 288–306, 2008.

[8] R. E. E. Jongschaap, W. J. Corré, P. S. Bindraban, and W. A. Brandenburg, "Claims and facts on *Jatropha curcas* L.," in "Global *Jatropha curcas* evaluation, breeding and propagation programme," Plant Research International B.V, Wageningen, The Netherlands, vol. 42, 2007.

[9] A. J. King, Y. Li, and I. A. Graham, "Profiling the developing *Jatropha curcas* L. seed transcriptome by pyrosequencing," *Bioenergy Research*, vol. 4, pp. 211-21, 2011.

[10] H. Liu, Y. J. Liu, M. F. Yang, and S. H. Shen, "A comparative analysis of embryo and endosperm proteome from seeds of *Jatropha curcas*," *Journal of Integrative Plant Biology*, vol.51, pp. 850–857, 2009.

[11] K. Gu, C. Yi, D. Tian, J.S. Sangha, Y. Hong, and Z. Chin, "Expression of fatty acid and lipid biosynthetic genes in developing endosperm of *Jatropha curcas*," *Biotechnology for Biofuels*, vol. 5, pp. 1–15, 2012.

[12] L.de S. Lopes, M. I. Gallao, and C. H. C. de M. Bertini, "Mobilization of reserves during germination of Jatropha seeds," *Revista Ciência Agronômica*, vol. 44 (no. 2), pp. 371-378, 2013.

[13] S. Sato, H. Hirakawa, S. Isobe, E. Fukai, A. Watanabe, M. Kato, K. Kawashima, C. Minami, A. Muraki, N. Nakazaki, C. Takahashi, S. Nakayama, Y. Kishida, M. Kohara, M. Yamada, H. Tsuruoka, S. Sasamoto, S. Tabata, T. Aizu, A. Toyoda, T. Shin-i, Y. Minakuchi, Y. Kohara, A. Fujiyama, S. Tsuchimoto, S. i. Kajiyama, E. Makigano, N. Ohmido, N. Shibagaki, J. A. Cartagena, N. Wada, T. Kohinata, A. Atefeh, S. Yuasa, S. Matsunaga, and K. Fukui, "Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L*," DNA Research*, vol. 18, pp. 65-76, 2011.

[14] M. S. Chen, G. J. Wang, R. L. Wang, J. Wang, S. Q. Song, Z. F. Xu, "Analysis of expressed sequence tags from biodiesel plant *Jatropha curcas* embryos at different developmental stages," Plant Science, vol. 181, pp. 696–700, 2011.

[15] V. Galli, F. Guzman, L. F. de Oliveira, G. Loss-Morais, A. P. Korbes, S. D. Silva, M. M. Margis-Pinheiro, and R. Margis, "Identifying microRNAs and transcript targets in Jatropha Seeds," *PLoS ONE* e83727, 2014.

[16] T. P. Durrett, C. Benning, and J. Ohlrogge, "Plant triacylglycerols as feedstocks for the production of biofuels," *The Plant Journal*, vol. 54, pp. 593–607, 2008.

[17] R. Kumar, A. Sinha, S. R. Singh, and D. Kamil, "Incidence of a leaf spot disease in J*atropha curcas* L. from Eastern Uttar Pradesh," *Journal of Mycology and Plant Pathology*, vol. 39 (no. 3), pp. 536–538, 2009.

[18] J. H. Kwon, O. Choi, J. Kim, and Y. S. Kwak, “First report of anthracnose disease on *Jatropha curcas L*," caused by *Colletotrichum gloeosporioides* in Korea,” *Journal of Phytopathology*, vol. 160 (no. 5), pp. 255-257, 2012.

[19] S. K. Raj, S. Kumar, S. K. Snehi, and U. Pathre, “First report of Cucumber mosaic virus on *Jatropha curcas* L. in India,” *Plant Disease,* vol. 92, pp. 171, 2008.

[20] R. C. Ramkat, A. Calari, F. Maghuly, and M. Laimer, “Biotechnological approaches to determine the impact of viruses in the energy crop plant *Jatropha curcas*,” *Virology Journal*, vol. 8, pp. 386, 2011.

[21] B. D. Kashina, M. D. Alegbejo, O. O. Banwo, S. L. Nielsen, and M. Nicolaisen, “Molecular identification of a new begomovirus associated with mosaic disease of *Jatropha curcas* L. in Nigeria,” *Archives of Virology*, vol. 158, pp. 511-514, 2013.

[22] D. Shalitin and S. Wolf, “Cucumber mosaic virus infection affects sugar transport in melon plants,” *Plant Physiology*, vol. 123, pp. 597-604, 2000.

[23] S. Pazarlar, M. Gumus, and G. B. Oztekin, “The effects of *Tobacco mosaic* virus infection on growth and physiological parameters in some pepper varieties (*Capsicum annuum* L.),” *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*, vol. 41, pp. 427-433, 2013.

[24] G. Erion and W. Riedell, “Barley yellow dwarf virus effects on cereal plant growth and transpiration,” *Crop Science*, vol. 52 (no. 6), pp. 2794-2799, 2012.

[25] M. Milavec, M. Ravnikar, and M. Kovač, “Peroxidases and photosynthetic pigments in susceptible potato infected with potato virus Y^{NTN},” *Plant Physiology and Biochemistry*, vol. 39, pp. 891–898, 2001.

[26] E. Góngora-Castillo, E. Ibarra-Laclette, D. L. Trejo-Saavedra, and R. F. Rivera-Bustamante, “Transcriptome analysis of symptomatic and recovered leaves of geminivirus-infected pepper (*Capsicum annuum*),” *Virology Journal*, vol. 9, pp. 295, 2012.

[27] F. Allie, E. J. Pierce, M. J. Okoniewski, and M. E. C. Rey, “Transcriptional analysis of South African cassava mosaic virus-infected susceptible and tolerant landraces

of cassava highlights differences in resistance, basal defense and cell wall associated genes during infection," *BMC Genomics*, vol. 15, pp. 1006, 2006.

[28] D. S. Aswatha Narayana, K. S. Rangaswamy, M. N. Shankarappa, M. N. Maruthi, C. N. L. Reddy, A. R. Rekha, and K. V. K. Murthy, "Distinct Begomoviruses closely related to cassava mosaic viruses causes Indian Jatropha Mosaic Disease," *International Journal of Virology*, vol. 3, pp. 1–11, 2007.

[29] J. E. Polston, M. A. Londoño, and H. Capobianco, "The complete genome sequence of new world jatropha mosaic virus," *Archives of Virology*, vol. 159, pp. 3131-3136, 2014.

[30] J. Cohn, G. Sessa, and G. B. Martin, "Innate immunity in plants," *Current Opinion in Immunology*, vol. 13, pp. 55–62, 2001.

[31] S. P. Pandey and I. E. Somssich, "The role of WRKY transcription factors in plant immunity," *Plant Physiology*, vol. 150, pp. 1648–1655, 2009.

[32] D. Marone, M. A. Russo, G. Laido`, A. M. D. Leonardis, and A. M. Mastrangelo, "Plant nucleotide binding site–leucine-rich repeat (NBS-LRR) genes: active guardians in host defense responses," *International Journal of Molecular Science*, vol. 14, pp. 7302–7326, 2013.

[33] R. Xu, R. Wang, and A. Liu, "Expression profiles of genes involved in fatty acid and triacylglycerol synthesis in developing seeds of Jatropha (*Jatropha curcas* L.)," *Biomass and Bioenergy*, vol. 35, pp. 1683-1692, 2011.

[34] S. D. Basha, F. George, H. P. S. Makkar, K. Becher, and M. Sujatha, "A comparative study of biochemical traits and molecular markers for assessment of genetic relationships between *Jatropha curcas* L. germplasm from different countries," *Plant Science*, vol. 176 pp. 812 – 823, 2009.

[35] K. Subramanyam, D. Muralidhararao, and N. Devanna, "Genetic diversity assessment of wild and cultivated varieties of *Jatropha curcas* (L.) in India by RAPD analysis," *African Journal of Biotechnology*, vol. 8 (no. 9), pp. 1900-1910, 2009.

[36] C. Linnaeus, "*Species plantarum*, exhibentes plantas rite cognitas, ad genera relatas, cum differentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus, secundum systema sexuale digestas," Impensis Laurentii Salvii, Holmiae [Stockholm], vol. 2, pp. [i], 561-1200, [1-30, index], [i, err.], 1753.

[37] B. Dehgan and G. L. Webster, "Morphology and infrageneric relationship of the genus Jatropha (Euphorbiaceae)," *University of California Publications in Botany*, vol. 74, pp. 957-981, 1979.

[38] A. J. King, W. He, J. A. Cuevas, M. Freudenberger, D. Ramiaramanana, and I. A. Graham, "Potential of *Jatropha curcas* as a source of renewable oil and animal feed," Journal of Experimental Botany, vol. 60, pp. 2897–2905, 2009.

[39] P. Srinophakun, A. Saimaneerat, I. Sooksathan, N. Visarathanon, S. Malaipan, K. Charernsom, and W. Chongrattanameteekul, "Integrated research on *Jatropha curcas* plantation management," *World Renewable Energy Congress*, Linköping, Sweden, 8-13 May, 2011.

[40] F. Abreu, "Alternative By-Products from *Jatropha*," in *International Consultation on Pro-Poor Jatropha development*, Rome, Italy, 2008.

[41] P. Natarajan, D. Kanagasabapathy, G. Gunadayalan, J. Panchalingam, N. Shree, P. A. Sugantham, K. K. Singh, and P. Madasamy, "Gene discovery from *Jatropha curcas* by sequencing of ESTs from normalized and full-length enriched cDNA library from developing seeds," *BMC Genomics*, vol. 11, pp. 606, 2010.

[42] J. Li, M. R. Li, P. Z. Wu, C. E. Tian, H. W. Jiang, and G. J. Wu, "Molecular cloning and expression analysis of a gene encoding a putative beta-ketoacyl-acyl carrier protein (ACP) synthase III (KAS III) from *Jatropha curcas*. Tree Physiology, vol. 28, pp. 921–927, 2008.

[43] S. Moreno-Perez, A. Sanchez-Garcia, J. J. Salas, R. Garces, and E. Martinez-Force, "Acyl-ACP thioesterases from macadamia (*Macadamia tetraphylla*) nuts: Cloning, characterization and their impact on oil composition," *Plant Physiology and Biochemistry*, vol. 49, pp. 82-87, 2011.

[44] G. Bonaventure, J. J. Salas, M. R. Pollard, and J. B. Ohlrogge, "Disruption of the FATB gene in Arabidopsis demonstrates an essential role of saturated fatty acids in plant growth," *The Plant Cell*, vol. 15, pp. 1020–1033, 2003.

[45] A. Sharma and R. S. Chauhan, "Identification and association analysis of castor bean orthologous candidate gene-based markers for high oil content in *Jatropha curcas*," *Plant Molecular Biology Reporter*, vol. 30, pp. 1025–1031, 2012.

[46] I. Cummins, M. J. Hills, J. H. E. Ross, D. H. Hobbs, M. D. Watson, and D. J. Murphy, "Differential, temporal and spatial expression of genes involved in storage oil and oleosin accumulation in developing rapeseed embryos: implications for the role of oleosins and the mechanisms of oil-body formation," *Plant Molecular Biology*, vol. 23, pp. 1015–27, 1993.

[47] S. Baud and L. Lepiniec, "Physiological and developmental regulation of seed oil production," *Progress in Lipid Research*, vol. 49, pp. 235–249, 2010.

[48] Z. Y. Hu, W. Hua, L. Zhang, L. B. Deng, X. F. Wang, G. H. Liu, W. J. Hao, and H. Z. Wang, "Seed structure characteristics to form ultrahigh oil content in rapeseed," *PLoS One*, vol. 8, pp. e62099, 2013.

[49] T. L. Shimada, T. Shimada, and I. Hara-Nishimura, "A rapid and non-destructive screenable marker, FAST, for identifying transformed seeds of *Arabidopsis thaliana*," *The Plant Journal*, vol. 61, pp. 519–528, 2010.

[50] S. Baud, J. P Boutin., M. Miquel, L. Lepiniec, and C. Rochat, "An integrated overview of seed development in *Arabidopsis thaliana* ecotype WS," *Plant Physiology and Biochemistry*, vol. 40, pp. 151–160, 2002.

[51] J. Zhang, Q. Hao, L. Bai, J. Xu, W. Yin, L. Song, L. Xu, X. Guo, C. Fan, Y. Chen, J. Ruan, S. Hao, Y. Li, R. R. C. Wang, and Z. Hu, "Overexpression of the soybean transcription factor GmDof4 significantly enhances the lipid content of *Chlorella ellipsoidea*," *Biotechnology for Biofuels*, vol. 7, pp. 128, 2014.

[52] M. J. Kim, S. W. Yang, H-Z. Mao, S. P. Veena, J-L. Yin, N-H. Chua, "Gene silencing of Sugar-dependent 1 (JcSDP1), encoding a patatin-domain triacylglycerol lipase, enhances seed oil accumulation in *Jatropha curcas*," *Biotechnology for Biofuels*, vol. 7, pp. 36, 2014.

[53] P. Natarajan and M. Parani, "De novo assembly and transcriptome analysis of five major tissues of *Jatropha curcas* L. using GS FLX titanium platform of 454 pyrosequencing," *BMC Genomics*, vol. 12, pp. 191, 2011

[54] W. Hua, R. J. Li, G. M. Zhan, J. Liu, J. Li, X. F. Wang, G. H. Liu, and H. Z. Wang, "Maternal control of seed oil content in *Brassica napus*: the role of silique wall photosynthesis," *The Plant Journal*, vol. 69, pp. 432-444, 2012.

[55] L. Wang, S. Yu, C. Tong, Y. Zhao, Y. Liu, C. Song, Y. Zhang, X. Zhang, Y. Wang, et al., "Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis," *Genome Biology*, vol. 15, pp. 39, 2014.

[56] J. Ye, P. Liu, C. S. Zhu, J. Qu, X. Wang, Y. Sun, F. Sun, Y. L. Jiang, G. H. Yue, and C. M. Wang, "Identification of candidate genes JcARF19 and JcIAA9 associated with seed size traits in Jatropha," *Functional & Integrative Genomics*, vol. 14, pp. 757–766, 2014.

[57] A. Grover, M. Kumari, S. Singh, S. S. Rathode, S. M. Gupta, P. Pandey, S. Gilotra, D. Kumar, M. Arif, and Z. Ahmed, "Analysis of Jatropha curcas transcriptome for oil enhancement and genic markers," *Physiology and Molecular Biology of Plants*, vol. 20 (no. 1), pp. 139-42, 2014.

[58] J. Qu, H. Z. Mao, W. Chen, S. Q. Gao, Y. N. Bai, Y. W. Sun, Y. F. Geng, and J. Ye, "Development of marker-free transgenic Jatropha plants with increased levels of seed oleic acid," *Biotechnology for Biofuels*, vol. 5, pp. 10, 2012.

[59] N. Sharma, P. L. Bhalla, and M. B. Singh, "Transcriptome-wide profiling and expression analysis of transcription factor families in a liverwort, *Marchantia polymorpha*," *BMC Genomics*, vol. 14, pp. 915, 2013.

[60] E. Grotewold, “Transcription factors for predictive plant metabolic engineering: are we there yet?,” *Current Opinion in Biotechnology*, vol. 19, pp. 138–144, 2008.

[61] J. Zou, V. Katavic, E. M. Giblin, D. L. Barton, S. L. MacKenzie, W. A. Keller, X. Hu, and D. C. Taylor, “Modification of seed oil content and acyl composition in the Brassicaceae by expression of a yeast sn-2 acyltransferase gene,” *The Plant Cell*, vol. 9, pp. 909–923, 1997.

[62] H. Vigeolas, P. Waldeck, T. Zank, and P. Geigenberger, “Increasing seed oil content in oil-seed rape (*Brassica napus* L.) by over-expression of a yeast glycerol-3-phosphate dehydrogenase under the control of a seed-specific promoter,” *Plant Biotechnology Journal*, vol. 5, pp. 431–441, 2007.

[63] A. A. Kelly, E. Shaw, S. J. Powers, S. Kurup, and P. J. Eastmond, “Suppression of the SUGAR-DEPENDENT1 triacylglycerol lipase family during seed development enhances oil yield in oilseed rape (*Brassica napus* L.),” *Plant Biotechnology Journal*, vol. 11, pp. 355–361, 2013.

[64] H. van Erp, A. A. Kelly, G. Menard, and. P. J. Eastmond, “Multigene engineering of triacylglycerol metabolism boosts seed oil content in Arabidopsis,” *Plant Physiology*, vol. 165 (no. 1), pp. 30-36, 2014.

[65] H. Tan, X. Yang, F. Zhang, X. Zheng, C. Qu, J. Mu, F. Fu, J. Li, R. Guan, H. Zhang, G. Wang, and J. Zuo, “Enhanced seed oil production in canola by conditional expression of *Brassica napus* LEAFY COTYLEDON1 (BnLEC1) and LEC1-LIKE (BnL1L) in developing seeds,” *Plant Physiology*, vol. 156, pp. 1577–1588, 2011.

[66] B. Shen, W. B. Allen, P. Zheng, C. Li, K. Glassman, J. Ranch, D. Nubel, and M. C. Tarczynski, “Expression of ZmLEC1 and ZmWRI1 increases seed oil production in maize,” *Plant Physiology*, vol. 153, pp. 980–987, 2010.

[67] A. To, J. Joubès, G. Barthole, A. Lécureuil, A. Scagnelli, S. Jasinski, L. Lepiniec, and S. Baud, “WRINKLED transcription factors orchestrate tissue specific

regulation of fatty acid biosynthesis in Arabidopsis," *The Plant Cell*, vol. 24, pp. 5007–5023, 2012.

[68] B. D. Pant, A. Burgos, P. Pant, A. Cuadros-Inostroza, L. Willmitze, and W. R. Scheible, "The transcription factor PHR1 regulates lipid remodeling and triacylglycerol accumulation in *Arabidopsis thaliana* during phosphorus starvation," *Journal of Experimental Botany*, vol. 66, pp. 1907–1918, 2015.

[69] H. W. Wang, B. Zhang, Y. J. Hao, J. Huang, A. G. Tian, Y. Liao, J. S. Zhang, and S. Y. Chen, "The soybean Dof-type transcription factor genes, GmDof4 and GmDof11, enhance lipid content in the seeds of transgenic Arabidopsis plants," *The Plant Journal*, vol. 52, pp. 716-729, 2007.

[70] X. Song, Y. Li, and X. Hou, "Genome-wide analysis of the AP2/ERF transcription factor superfamily in Chinese cabbage (*Brassica rapa* ssp. pekinensis)," *BMC Genomics*, vol. 14, p. 573, 2013.

[71] J. Hu, D. Wang, J. Li, et al., "Genome-wide identification of transcription factors and transcription-factor binding sites in oleaginous microalgae *Nannochloropsis*," *Scientific Reports*, vol. 4, pp. 5454, 2014.

[72] W. Ma, Q. Kong, V. Arondel, A. Kilaru, P. D. Bates, N. A. Thrower, C. Benning, and J. B. Ohlrogge, "WRINKLED1, A ubiquitous regulator in oil accumulating tissues from arabidopsis embryos to oil palm mesocarp," *PloS One*, vol. 8 (no. 7), pp. e68887, 2013.

[73] S. K. Palaniswamy, S. James, H. Sun, R. S. Lamb, R. V. Davuluri, and E. Grotewold, "AGRIS and AtRegNet. A platform to link cis-regulatory elements and transcription factors into regulatory networks," *Plant Physiology*, vol. 140, pp. 818-829, 2006.

[74] P. R. Fobert, (2012) *Role of transcription factors in storage lipid accumulation in plants* [Online]. Available http: //lipidlibrary.aocs.org/plantbio/transfactors/index.htm

[75] N. M. D. Courchesne, A. Parisien, B. Wang, and C. Q. Lan, "Enhancement of lipid production using biochemical, genetic and transcription factor engineering approaches," *Journal of Biotechnology*, vol. 141, pp. 31-41, 2009.

[76] Y. F. Liu, Q. T. Li, X. Lu, et al., "Soybean GmMYB73promotes lipid accumulation in transgenic plants," *BMC Plant Biology*, vol. 14, pp. 73, 2014.

[77] W. Chew, M. Hrmova, and S. Lopato, "Role of homeodomain leucine zipper (HD-Zip) IV transcription factors in plant development and plant protection from deleterious environmental factors," *International Journal of Molecular Sciences*, vol. 14, pp. 8122–8147, 2013.

[78] N. A. Abdelmajeed, E. N. Danial, and H.S. Ayad, "The effect of environmental stress on qualitative and quantitative essential oil of aromatic and medicinal plants," *Archives Des Sciences*, vol. 66 (no. 4), pp. 100-120, 2013.

[79] G. J. Seiler, "Analysis of the relationships of environmental factors with seed oil and fatty acid concentrations of wild annual sunflower," *Field Crops Research*, vol. 15, pp. 57-72, 1986.

[80] B. D. Oomah, M. Busson, D. V. Godfrey, and J. C. G. Drover, "Characteristics of hemp (*Cannabis sativa* L.) seed oil," *Food Chemistry*, vol. 76, pp. 33-43, 2002.

[81] B. Behtari, F. Gholami, K. A. Khalid, G. D. Tilaki, and R. Bahari, "Effect of Growth Stages and Altitude on *Artemisia herba-alba* Asso Essential Oil Growing in Iran," *Journal of Essential Oil Bearing Plants*, vol. 15 (no. 2), pp. 307-313, 2012.

[82] E. Yazdanpanah, A. Moradshahi, and V. Rowshan, "The effects of growth habitat of medicinal plant yarrow (*Achillea wilhelmsii* c.koch) on the essential oil constituents," *Indian Journal of Fundamental and Applied Life Sciences*, vol. 4 (no. 4), pp. 36-42, 2014.

[83] S. X. Fu, H. Cheng, and C. Qi, "Microarray analysis of gene expression in seeds of *Brassica napus* planted in Nanjing (altitude: 8.9 m), Xining (altitude: 2261.2 m)

and Lhasa (altitude: 3658 m) with different oil content," *Molecular Biology Reports*, vol. 36, pp. 2375–2386, 2009.

[84] D.A. Somers, R. A. Keith, M. A. Egli, L. C. Marshall, B. G. Gengenbach, J. W. Gronwald, and D. L. Wyse, "Expression of the Acc1 gene-encoded acetyl-coenzyme A carboxylase in developing maize (*Zea mays* L.) kernels," *Plant Physiology*, vol. 101, pp. 1097-1101, 1993.

[85] Y. Madoka, K. I. Tomizawa, J. Mizoi, I. Nishida, Y. Nagano, *and* Y. Sasaki, "Chloroplast transformation with modified *accD* operon increases acetyl-CoA carboxylase and causes extension of leaf longevity and increase in seed yield in tobacco," *Plant and Cell Physiology*, vol. 43, pp. 1518–1525, 2002.

[86] W. W. Xie, S. Gao, S. H. Wang, J. Q. Zhu, Y. Xu, L. Tang, and F. Chen, "Cloning and expression analysis of carboxyltransferase of acetyl-coa carboxylase from *Jatropha curcas*," *Zeitschrift für Naturforschung*, vol. 65, pp. 103–108, 2010.

[87] K. Dehesh, H. Tai, P. Edwards, J. Byrne, and J. G. Jaworski, "Over-expression of 3-ketoacyl-acyl-carrier protein synthase IIIs in plants reduces the rate of lipid synthesis," *Plant Physiology*, vol. 125, pp. 1103–1114, 2001.

[88] I. I. G. S. Verwoert, K. H. vd. Linden, M. C. Walsh, H. J. J. Nijkamp, and A. R. Stuitje, "Modification of *Brassica napus* seed oil by expression of the *Esherichia coli* fabH gene, encoding 3-ketoacyl-acyl carrier protein synthase III," *Plant Molecular Biology*, vol. 27, pp. 875–886, 1995.

[89] M. S. Pidkowich, H. T. Nguyen, I. Heilmann, T. Ischebeck, and J. Shanklin, "Modulating seed beta-ketoacyl-acyl carrier protein synthase II level converts the composition of a temperate seed oil to that of a palm-like tropical oil," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, pp. 4742–4747, 2007.

[90] J. A. Napier, R. P. Haslam, F. Beaudoin, and E. B. Cahoon, "Understanding and manipulating lipid composition: metabolic engineering leads the way," *Current Opinion in Plant Biology*, vol. 19, pp. 68–75, 2014.

[91] X. Wu, X. Zou, Q. Chang, Y. Zhang, Y. Li, L. Zhang, J. Huang, and B. Liang, "The evolutionary pattern and the regulation of stearoyl-CoA desaturase genes," *Biomed Research International*, Article ID 856521, 12 pages. doi:10.1155/2013/856521, 2013.

[92] S. Okada, X. R. Zhou, K. Damcevski, N. Gibb, C. Wood, M. Hamberg, and V. S. Haritos, "Diversity of Δ12 Fatty acid desaturases in santalaceae and their role in production of seed oil acetylenic fatty acids," *Journal of Biological Chemistry*, vol. 288 (no. 45), pp. 32405–32413, 2013.

[93] P. Zheng, W. B. Allen, K. Roesler, M. E. Williams, S. Zhang, J. Li, K. Glassman, J. Ranch, D. Nubel, W. Solawetz, D. Bhattramakki, V. Llaca, S. Deschamps, G. Y. Zhong, M. C. Tarczynski, and B. Shen, "A phenylalanine in DGAT is a key determinant of oil content and composition in maize," *Nature Genetics*, vol. 40, pp. 367–372, 2008.

[94] C. Y. Zhang, U. Iskandarov, E. T. Klotz, R. L. Stevens, R. E. Cahoon, T. J. Nazarenus, S. L. Pereira, and E. B. Cahoon, "A thraustochytrid diacylglycerol acyltransferase 2 with broad substrate specificity strongly increases oleic acid content in engineered *Arabidopsis thaliana* seeds," *Journal of Experimental Botany*, vol. 64, pp. 3189–3200, 2013.

[95] T. P. Durrett, D. D. McClosky, A. W. Tumaney, D. A. Elzinga, J. Ohlrogge, and M. Pollard, "A distinct DGAT with sn-3 acetyltransferase activity that synthesizes unusual, reduced-viscosity oils in Euonymus and transgenic seeds," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, pp. 9464–9469, 2010.

[96] T. Vanhercke, E. T. A, P. Shrestha, X. R. Zhou, S. P. Singh, and J. R. Petrie, "Synergistic effect of WRI1 and DGAT1 co expression on triacylglycerol biosynthesis in plants," *FEBS Letters*, vol. 587, pp. 364–369, 2013.

[97] K. Xu, Y. Yang, and X. Li, "Ectopic expression of *Crambe abyssinica* lysophosphatidic acid acyltransferase in transgenic rapeseed increases its oil content," *African Journal of Biotechnology*, vol. 9, pp. 3904-3910, 2010.

[98] S. Maisonneuve, J. J. Bessoule, R. Lessire, M. Delseny, and T. J. Roscoe, "Expression of rapeseed microsomal lysophosphatidic acid acyltransferase isozymes enhances seed oil content in Arabidopsis," *Plant Physiology*, vol. 152, pp. 670–684, 2010.

[99] J. M. Arroyo-Caro, T. Chileh, M. Kazachkov, and J. Zou, "The multigene family of lysophosphatidate acyltransferase (LPAT)-related enzymes in *Ricinus communis*: cloning and molecular characterization of two LPAT genes that are expressed in castor seeds," *Plant Science*, vol. 199–200, pp. 29-40, 2013.

[100] R. J. Weselake, D. C. Taylor, M. H. Rahman, S. Shah, A. Laroche, P. B. McVetty, and J. L. Harwood, "Increasing the flow of carbon into seed oil," *Biotechnology Advances*, vol. 27, pp. 866-878, 2009.

[101] X. Chen, C. L. Snyder, M. Truksa, S. Shah, and R. J. Weselake, "sn-Glycerol 3-phosphate acyltransferase in plants," *Plant Signaling & Behavior*, vol. 6, pp. 1695–1699, 2011.

[102] R. K. Jain, M. Coffey, K. Lai, A. Kumar, and S. L. MacKenzie, "Enhancement of seed oil content by expression of glycerol-3-phosphateacyltransferases genes," *Biochemical Society Transactions*, vol. 28, pp. 958–961, 2000.

[103] F. Liu, Y. Xia, L. Wu, D. Fu, A. Hayward, J. Luo, X. Yan, X. Xiong, P. Fu, G. Wu, and C. Lu, "Enhanced seed oil content by overexpressing genes related to triacylglyceride synthesis," *Gene*, vol. 557 (no. 2), pp. 163-171, 2015.

[104] A. Fatihi, A. M. Zbierzak, and P. Dörmann, "Alterations in seed development gene expression affect size and oil content of Arabidopsis seeds," *Plant Physiology*, vol. 163, pp. 973–985, 2013.

[105] X. L. Wu, Z. H. Liu, Z. H. Hu, and R. Z. Huang, "BnWRI1 coordinates fatty acid biosynthesis and photosynthesis pathways during oil accumulation in rapeseed," *Journal of Integrative Plant Biology*, vol. 56 (no. 6), 582-593, 2014.

[106] J. P. Tewari, H. D. Dwivedi, M. Pathak, and S. K. Srivastava, "Incidence of a mosaic disease in *Jatropha curcas* L. from eastern Uttar Pradesh," *Current Science*, vol. 93 (no. 8), pp. 1048–1049, 2007.

[107] C. H. Srinivasa Rao, M. P. Kumari, S. P. Wani, and S. Marimuthu, "Occurrence of black rot in *Jatropha curcas* L. plantations in India caused by *Botryosphaeria dothidea*," *Current Science*, vol. 100 (no. 10), pp. 1547–1549, 2011.

[108] F. C. O. Freire and G. B. Parente, "As doenças das Jatrofas (*Jatropha curcas* L. e J. podagri-caHook.) no estado do Ceará. C. omunicado Técnico," 120. Embrapa, 4., 2006.

[109] D. A. C. Sá, G. R. S. Santos, G. Q. Furtado, E. A. L. Erasmo, and I. R. Nascimento, "Transporte, patogenicidade e transmissibilidade de fungos associados às sementes de pinhão Manso," *Revista Brasileira de Sementes*, vol. 33 (no. 4), pp. 663-670, 2011.

[110] A. C. Dianese, and A. Cargnin, "Ocorrência de Oidiumsp. em pinhão manso(*Jatropha curcas* L.) em Planaltina," DF.Documentos 231. Planaltina, DF: Embrapa Cer-rados, 15p, 2008.

[111] U. Braun and R. T. A. Cook, "Taxonomic Manual of the Erysiphales (Powdery Mildews)," *CBS Biodiversity Series*, vol. 11 (no. 11), pp. 1-707, 2012.

[112] K. Jayanna, "Studies on Jatropha mosaic virus disease", M.Sc. Dissertation, University of Agricultural Sciences, Dharwad, Karnataka, India, 2006.

[113] S. Gao, J. Qu, N.-H. Chua, and J. Ye, "A new strain of Indian cassava mosaic virus causes a mosaic disease in the biodiesel crop *Jatropha curcas*," *Archives of Virology*, vol. 155, pp. 607-612, 2010.

[114] J. Bird, "A whitefly transmitted mosaic of *Jatropha gossypifolia*. Technical paper 22. Agricultural Experimental station, University of Puerto Rico, 1957.

[115] K. T. Rangaswamy, N. Raghavendra, K. S. Shankarappa, D. S. Aswatha Narayana, H. A. Prameela, M. R. Govindappa, and M. R. Gururaj Rao, "Natural occurrence of a mosaic virus disease on Jatropha," *National Symposium on Crop Disease*

Management in Dry Land Agriculture and 57th IPS, New Delhi, India, January 12-14, 2005, pp. 156,

[116] M. N. Maruthi, J. Colvin, S. Seal, G. Gibbson, and J. Cooper, "Co-adaptation between cassava mosaic Gemini viruses and their local vector populations," *Virus Research*, vol. 86, pp. 71-85, 2002.

[117] J. K. Brown, "Molecular markers for the identification and global tracking of whitefly vector-begomovirus complexes," *Virus Research*, vol. 71, pp. 233–260, 2000.

[118] B. D. Harrison, M. M. Swanson, and D. Fargette, "Begomovirus coat protein: Serology, variation and functions," *Physiological and Molecular Plant Pathology*, vol. 60, pp. 257-271, 2002.

[119] B. Afreen, M. Gulfishan, G. Baghel, M. Fatma, A. A. Khan, and Q.A. Naqvi, "Molecular detection of a virus infecting carrot and its effect on some cytological and physiological parameters," *African Journal of Plant Sciences*, vol. 5, pp. 407-411, 2011.

[120] E. F. Mohamed, "Changes in protein, amino acids composition and leaf cells of beet plants (*Beta vulgaris* L.) due to Beet mosaic virus (*Bt*MV) infection," *Journal of American Science*, vol. 7, pp. 245-254, 2011.

[121] J. Lu, Z. X. Du, J. Kong, L. N. Chen, Y. H. Qiu, et al., "Transcriptome analysis of *Nicotiana tabacum* infected by cucumber mosaic virus during systemic symptom development," *PLoS One*, vol. 7, pp. e43447, 2012.

[122] J. Liu, J. Yang, H. Bi, and P. Zhang, "Why mosaic? Gene expression profiling of African cassava mosaic virus infected cassava reveals the effect of chlorophyll degradation on symptom development," *Journal of Integrative Plant Biology*, vol. 56 (no. 2), pp. 122-132, 2014.

[123] H. Choi, Y. Jo, S. Lian, K. M. Jo, H. Chu, J. Y. Yoon, S. K. Choi, K. H. Kim, and W. K. Cho, "Comparative analysis of chrysanthemum transcriptome in response to three RNA viruses: Cucumber mosaic virus, Tomato spotted wilt virus and Potato virus X," *Plant Molecular Biology*, vol. 88 (no. 3), pp. 233-248, 2015.

[124] W. K. Cho, S. Lian, S. M. Kim, B. Y. Seo, J. K. Jung, and K. H. Kim, "Time-course RNA-seq analysis reveals transcriptional changes in rice plants triggered by rice stripe virus infection," *PLoS One*, vol. 10 (no. 8), pp. e0136736, 2015.

[125] O. A. Postnikova, M. Hult, J. Shao, A. Skantar, and L. G. Nemchinov, "Transcriptome analysis of resistant and susceptible Alfalfa cultivars infected with root-knot nematode *Meloidogyne incognita*," *PLoS One*, vol. 10 (no. 2), pp. e0118269, 2015.

[126] H. Fan, Y. Zhang, H. Sun, J. Liu, Y. Wang, X. Wang, et al., "Transcriptome analysis of *Beta macrocarpa* and identification of differentially expressed transcripts in response to beet necrotic yellow vein virus infection," *PLoS ONE*, vol. 10 (no. 7), pp. e0132277, 2015.

[127] T. Ishihara, I. Mitsuhara, H. Takahashi, and K. Nakaho, "Transcriptome analysis of quantitative resistance-specific response upon *Ralstonia solanacearum* infection in tomato," *PLoS One*, vol. 7, pp. e46763, 2012.

[128] J. J. Liu, R. Sturrock, and R. Benton, "Transcriptome analysis of *Pinus monticola* primary needles by RNA-seq provides novel insight into host resistance to *Cronartium ribicola*," *BMC Genomics*, vol. 14, pp. 884, 2013.

[129] G. G. L. Costa, K. C. Cardoso, L. E. V. Del Bem, A. C. Lima, M. A. S. Cunha, L. D. Campos-Leite, R. Vicentini, F. Papes, R. C. Moreira, J. A. Yunes, F. A. P. Campos, and M. J. D. Silva, "Transcriptome analysis of the oil-rich seed of the bioenergy crop *Jatropha curcas* L.," *BMC Genomics*, vol. 11, pp. 462, 2010.

[130] H. Wang, Z. Zou, S. Wang, and M. Gong, "Global analysis of transcriptome responses and gene expression profiles to cold stress of *Jatropha curcas* L," *PLoS ONE*, vol. 8 (no. 12), pp. e82817, 2013.

[131] P. Juntawong, A. Sirikhachornkit, R. Pimjan, et al., "Elucidation of the molecular responses to waterlogging in Jatropha roots by transcriptome profiling," *Frontiers in Plant Science*, vol. 5, pp. 658, 2014.

[132] B. Z. Pan, M. S. Chen, J. Ni, and Z. F. Xu, “Transcriptome of the inflorescence meristems of the biofuel plant *Jatropha curcas* treated with cytokinin,” *BMC Genomics*, vol. 15, pp. 974, 2014.

[133] Y.-L. Guo, J. Fitz, K. Schneeberger, S. Ossowski, J. Cao, and D. Weigel, "Genome-wide comparison of nucleotide-binding site-leucine-rich repeat-encoding genes in Arabidopsis," *Plant Physiology*, vol. 157, pp. 757-769, 2011.

[134] D. F. Klessig, J. Durner, J. Shah, and Y. Yang, “Salicylic acid mediated signal transduction in plant disease resistance," in *Phytochemical signals and plant–microbe interaction: recent advances in phytochemistry*, vol. 32, J. T. Romeo, K. R. Downum, and R. Verpoorte, Eds. New York: Plenum Press, 1998, pp. 119-137.

[135] Y. Zhang, S. Goritschnig, X. Dong, and X. Li, "A gain-of-function mutation in a plant disease resistance gene leads to constitutive activation of downstream signal transduction pathways in suppressor of npr1-1, constitutive 1," *The Plant Cell*, vol. 15, pp. 2636-2646, 2003.

[136] J. L. Dangl and J. D. Jones, “Plant pathogens and integrated defence responses to infection,” *Nature*, vol. 411 (no. 6839), pp. 826-833, 2001.

[137] J. M. Lorang, T. A. Sweat, and T. J. Wolpert, "Plant disease susceptibility conferred by a "resistance" gene," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, pp. 14861-14866, 2007.

[138] D. E. K. Tarr and H. M. Alexander, "TIR-NBS-LRR genes are rare in monocots: evidence from diverse monocot orders," *BMC Research Notes*, vol. 2, pp. 197, 2009.

[139] G. B. Martin, A. J. Bogdanove, and G. Sessa, "Understanding the functions of plant disease resistance proteins," *Annual Review of Plant Biology*, vol. 54, pp. 23-61, 2003.

[140] B. J. DeYoung and R. W. Innes, "Plant NBS-LRR proteins in pathogen sensing and host defense," *Nature Immunology*, vol. 7, pp. 1243-9, 2006.

[141] G. van Ooijen, H. A. van den Burg, B. J. C. Cornelissen, and F. L. W. Takken, "Structure and function of resistance proteins in solanaceous plants," *Annual Review of Phytopathology*, vol. 45, pp. 43-72, 2007.

[142] E. A. van der Biezen and J. D. Jones, "The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals," *Current Biology*, vol. 8 (no. 7), pp. 226-227, 1998.

[143] W. I. Tameling, J. H. Vossen, M. Albrecht, T. Lengauer, J. A. Berden, M. A. Haring, B. J. Cornelissen, and F. L. Takken, "Mutations in the NB-ARC domain of I-2 that impair ATP hydrolysis cause auto activation," *Plant Physiology*, vol. 140, pp. 1233–1245, 2006.

[144] J. E. Luck, G. J. Lawrence, P. N. Dodds, K. W. Shepherd, and J. G. Ellis, "Regions outside of the leucine-rich repeats of flax rust resistance proteins play a role in specificity determination," *The Plant Cell*, vol. 12, pp. 1367-1377, 2000.

[145] B. C. Meyers, A. W. Dickerman, R. W. Michelmore, S. Sivaramakrishnan, B. W. Sobral, and N. D. Young, "Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily," *The Plant Journal*, vol. 20, pp. 317-332, 1999.

[146] B. Monosi, R. J. Wisser, L. Pennill, and S. H. Hulbert, "Full-genome analysis of resistance gene homologues in rice," *Theoretical and Applied Genetics*, vol. 109, pp. 1434-1447, 2004.

[147] T. Zhou, Y. Wang, J. Q. Chen, H. Araki, Z. Jing, K. Jiang, J. Shen, and D. Tian, "Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes," *Molecular Genetics and Genomics*, vol. 271, pp. 402-415, 2004.

[148] K. H. Wolfe, M. Gouy, Y. W. Yang, P. M. Sharp, and W. H. Li, "Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 86, pp. 6201-6205, 1989.

[149] A. Plocik, J. Layden, and R. Kesseli, "Comparative analysis of NBS domain sequences of NBS-LRR disease resistance genes from sunflower, lettuce, and chicory," *Molecular Phylogenetics and Evolution*, vol. 31, pp. 153-163, 2004.

[150] L. McHale, X. Tan, P. Koehl, and R. W. Michelmore, "Plant NBS-LRR proteins: adaptable guards," *Genome Biology*, vol. 7, pp. 212, 2006.

[151] Y. Liu, M. Schiff, R. Marathe, and S. P. Dinesh-Kumar, "Tobacco Rar1, EDS1 and NPR1/NIM1 like genes are required for N-mediated resistance to tobacco mosaic virus," *The Plant Journal*, vol. 30, pp. 415-29, 2002.

[152] Y. S. Seo, M. R. Rojas, J.-Y.Lee, S.-W.Lee, J.-S.Jeon, P. Ronald, W. J. Lucas, and R. L. Gilbertson, "A viral resistance gene from common bean functions across plant families and is up-regulated in a non-virus-specific manner," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, pp. 11856-11861, 2006.

[153] T. H. Tai, D. Dahlbeck, E. T. Clark, P. Gajiwala, R. Pasion, M. C. Whalen, R. E. Stall, and B. J. Staskawicz, "Expression of the Bs2 pepper gene confers resistance to bacterial spot disease in tomato," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, pp. 14153-14158, 1999.

[154] L. Gao, Y. Cao, Z. Xia, G. Jiang, G. Liu, W. Zhang, and W. Zhai, "Do transgenesis and marker-assisted backcross breeding produce substantially equivalent plants? A comparative study of transgenic and backcross rice carrying bacterial blight resistant gene Xa21," *BMC Genomics*, vol. 14, p. 738, 2013.

[155] F. Mascher, C. Matasci, Y. Kneubuehler, S. Brunner, A. Schori, B., and B. Keller, “Evaluation of disease resistance in wheat supplemented with Pm3b," *Tagung der Vereinigung der Pflanzenzüchter und Saatgutkaufleute Österreichs*, vol. 2012, pp. 15-18, 2012.

[156] Y. Narusaka, T. Shinya, M. Narusaka, N. Motoyama, H. Shimada, K. Murakami, and N. Shibuya, “Presence of LYM2 dependent but CERK1 independent disease resistance in Arabidopsis,” *Plant Signaling & Behavior*, vol. 8 (no. 9), pp. e25345, 2013.

[157] B. C. Meyers, A. Kozik, A. Griego, H. Kuang, and R. W. Michelmore, "Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis," *The Plant Cell*, vol. 15, pp. 809-34, 2003.

[158] F. Jupe, L. Pritchard, G. J. Etherington, K. Mackenzie, P. J. A. Cock, F. Wright, S. K. Sharma, D. Bolser, G. J. Bryan, J. D. G. Jones, and I. Hein, "Identification and localisation of the NB-LRR gene family within the potato genome," *BMC Genomics*, vol. 13, pp. 75, 2012.

[159] Y. J. Kang, K. H. Kim, S. Shim, M. Y. Yoon, S. Sun, M. Y. Kim, K. Van, and S. H. Lee, "Genome-wide mapping of NBS-LRR genes and their association with disease resistance in soybean," *BMC Plant Biology*, vol. 12, pp. 139, 2012.

[160] A. Kohler, C. Rinaldi, S. Duplessis, M. Baucher, D. Geelen, F. Duchaussoy, B. C. Meyers, W. Boerjan, and F. Martin, "Genome-wide identification of NBS resistance genes in *Populus trichocarpa*," *Plant Molecular Biology*, vol. 66, pp. 619-636, 2008.

[161] X. H. Wan, S. X.Chen, C. Y.Wang, R. R.Zhang, S. Q.Cheng, H. W.Meng, and X. Q. Shen, "Isolation, expression, and characterization of a hydroperoxide lyase gene from cucumber," *International Journal of Molecular Sciences*, vol. 14, pp. 22082-22101, 2013.

[162] D. Desveaux, R. Subramaniam, C. Després, J.-N. Mess, C. Lévesque, P. R. Fobert, J. L. Dangl, and N. Brisson, "A "Whirly" transcription factor is required for salicylic acid-dependent disease resistance in Arabidopsis," *Developmental Cell*, vol. 6, pp. 229-40, 2004.

[163] T. Eulgem, "Regulation of the Arabidopsis defense transcriptome," *Trends in Plant Science*, vol. 10 (no. 2), pp. 71-78, 2005.

[164] W. Fan and X. Dong, "In vivo interaction between NPR1 and transcription factor TGA2 leads to salicylic acid-mediated gene activation in Arabidopsis," *The Plant Cell*, vol. 14, pp. 1377-89, 2002.

[165] M. S. Alves, S. P. Dadalto, A. B. Goncalves, G. B. de Souza, V. A. Barros, and L. G. Fietto, "Plant bZIP transcription factors responsive to pathogens: a review," *International Journal of Molecular Sciences*, vol. 14, pp. 7815–7828, 2013.

[166] A. Katiyar, S. Smita, S. K. Lenka, R. Rajwanshi, V. Chinnusamy, and K. C. Bansal, “Genome-wide classification and expression analysis of MYB transcription factor families in rice and Arabidopsis, *BMC Genomics*, vol. 13, pp. 544, 2012.

[167] Y. Jiang, Y. Duan, J. Yin, S. Ye, J. Zhu, F. Zhang, W. Lu, D. Fan, and K. Luo, “Genome-wide identification and characterization of the Populus WRKY transcription factor family and analysis of their expression in response to biotic and abiotic stresses,” *Journal of Exerimental Botany*, vol. 65, pp. 6629-6644, 2014.

[168] R. P. Birkenbihl, C. Diezel, and I. E. Somssich, “*Arabidopsis* WRKY33 is a key transcriptional regulator of hormonal and metabolic responses toward *Botrytis cinerea* infection,” *Plant Physiology*, vol. 159, pp. 266–285, 2012.

[169] A. Nakayama, S. Fukushima, S. Goto, A. Matsushita, M. Shimono, S. Sugano, C. J. Jiang, A. Akagi, M. Yamazaki, H. Inoue, and H. Takatsuji, “Genome-wide identification of WRKY45-regulated genes that mediate benzothiadiazole-induced defense responses in rice,” *BMC Plant Biology*, vol. 13, pp. 150, 2013.

[170] L. Satapathy, D. Singh, P. Ranjan, D. Kumar, M. Kumar, K. V. Prabhu, et al., “Transcriptome-wide analysis of WRKY transcription factors in wheat and their leaf rust responsive expression profiling,” *Molecular Genetics and Genomics*, vol. 289, pp. 1289–1306, 2014.

[171] K. Wei, J. Chen, Y. Wang, Y. Chen, S. Chen, Y. Lin, S. Pan, X. Zhong, and D. Xie, “Genome-wide analysis of bZIP-encoding genes in maize.,” *DNA Researc*h, vol. 19 (no. 6), pp. 463-476, 2012.

[172] M. Libault, J. Wan, T. Czechowski, M. Udvardi, and G. Stacey, "Identification of 118 Arabidopsis transcription factor and 30 ubiquitin-ligase genes responding to chitin, a plant-defense elicitor,*" Molecular Plant-Microbe Interactions*, vol. 20, pp. 900-911, 2007

[173] M. B. Osorio, L. Bücker-Neto, G. Castilhos, A. C. Turchetto-Zolet, B. Wiebke-Strohm, M. H. Bodanese-Zanettini, and M. Margis-Pinheiro, "Identification and *in silico* characterization of soybean trihelix-GT and bHLH transcription factors involved in stress responses," *Genetics and Molecular Biology*, vol. 35, pp. 233-246, 2012.

[174] D. Salomon and G. Sessa, “Biotechnological strategies for engineering plants with durable resistance to fungal and bacterial pathogens,” Tel Aviv University, Israel, 2012, 2012, pp. 329–339.

[175] R. Liu, B. Lü, X. Wang, C. Zhang, S. Zhang, J. Qian, L. Chen, H. Shi, and H. Dong, “Thirty-seven transcription factor genes differentially respond to a harpin protein and affect resistance to the green peach aphid in *Arabidopsis*,” *Journal of Biosciences*, vol. 35, pp. 435-450, 2010.

[176] L. Sun, H. Zhang, D. Li, L. Huang, Y. Hong, X. S. Ding, et al., “Functions of rice NAC transcriptional factors, ONAC122 and ONAC131, in defense responses against *Magnaporthe grisea*,” *Plant Molecular Biology*, vol. 81, pp. 41–56, 2013.

[177] B. Liu, Z. Ouyang, Y. Zhang, X. Li, Y. Hong, L. Huang, et al., “Tomato NAC transcript tion factor SlSRN1 positively regulates defense response against biotic stress but negatively regulates abiotic stress response,” *PLoS One*, vol. 9, pp. e102067, 2014.

[178] J. Dong, C. H. Chen, and Chen ZX, “Expression profiles of the Arabidopsis WRKY gene superfamily during plant defense response,” *Plant Molecular Biology*, vol. 5, pp. 21-37, 2003.

[179] M. Ohme-Takagi, K. Suzuki, and H. Shinshi, “Regulation of ethylene-induced transcription of defense genes,” *Plant and Cell Physiology*, vol. 41, pp. 1187–1192, 2000.

[180] Y. Sakuma, K. Maruyama, Y. Osakabe, F. Qin, M. Seki, K. Shinozaki, and K. Yamaguchi-Shinozaki, “Functional analysis of an Arabidopsis transcription factor, DREB2A, involved in drought-responsive gene expression,” *The Plant Cell*, vol. 18, pp. 1292–1309, 2006.

[181] G. H. Jiang, Z. H. Xia, Y. L. Zhou, J. Wan, D. Y. Li, R. S. Chen, W. X. Zhai, and L. H. Zhu, “Testifying the rice bacterial blight resistance gene xa5 by genetic complementation and further analyzing xa5 (Xa5) in comparison with its homolog TFIIAgamma1,” *Molecular Genetics and Genomics*, vol. 275, pp. 354–366, 2006.

[182] M. Collinge and T. Boller, "Differential induction of two potato genes, *Stprx2* and *StNAC*, in response to infection by *Phytophthora infestans* and to wounding," *Plant Molecular Biology*, vol. 46, pp. 521–529, 2001.

[183] A. K. Singh, V. Sharma, A. K. Pal, V. Acharya, and P. S. Ahuja, "Genome-wide organization and expression profiling of the NAC transcription factor family in potato (*Solanum tuberosum* L.)," *DNA Research*, vol. 20, pp. 403–423, 2013.

[184] M. S. Padmanabhan, S. Ma, T. M. Burch-Smith, K. Czymmek, P. Huijser, et al., "Novel positive regulatory role for the SPL6 transcription factor in the N TIR-NB-LRR receptor-mediated plant innate immunity," *PLoS Pathogens*, vol. 9, pp. e1003235, 2013.

[185] H. Luo, F. Song, R. M. Goodman, and Z. Zheng, "Up-regulation of OsBIHD1, a rice gene encoding BELL homeodomain transcriptional factor, in disease resistance responses," *Plant Biology*, vol. 7, pp. 459-468, 2005.

[186] E. Seo and D. C. Choi, "Functional studies of transcription factors involved in plant defenses in the genomics era," *Briefings in Functional Genomics*, vol. 14 (no. 4), pp. 260-267, 2015.

[187] X. Wang, H. Han, J. Yan, F. Chen, and W. Wei, "A new AP2/ERF transcription factor from the oil plant *Jatropha curcas* confers salt and drought tolerance to transgenic tobacco," *Applied Biochemistry and Biotechnology*, vol. 176, pp. 582–597, 2015.

[188] X. Qin, X. Zheng, X. Huang, Y. Lii, C. Shao, Y. Xu, and F. Chen, "A novel transcription factor *Jc*NAC1 response to stress in new model woody plant *Jatropha curcas*," *Planta*, vol. 239 (no. 2), pp. 511–520, 2014.

[189] H. L. Li, D. Guo, and S. Q. Peng, "Molecular characterization of the *Jatropha curcas JcR1MYB1* gene encoding a putative R1-MYB transcription factor," *Genetics and Molecular Biology*, vol. 37 (no. 3), pp. 549–555, 2014.

[190] N. Kaushik and D. Bhardwaj, "Screening of *Jatropha curcas* germplasm for oil content and fatty acid composition," *Biomass and Bioenergy*, vol. 58, pp. 210–218, 2013.

[191] D. R. Zerbino and E. Birney, "Velvet: Algorithms for de novo short read assembly using de Bruijn graphs," *Genome Research*, vol. 18, pp. 821–829, 2008.

[192] J. Vandesompele, K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, and F. Speleman, "Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes," *Genome Biology*, vol. 3, pp. research0034, 2002.

[193] T. D. Schmittgen and K. J. Livak, "Analyzing real-time PCR data by the comparative CT method," *Nature Protocols*, vol. 3, pp. 1101-1108, 2008.

[194] H. Hotelling, "Analysis of a Complex of Statistical Variables into Principal Components," *Journal of Educational Psychology*, vol. 24 (no. 6), pp. 417–441, 1933.

[195] B. Li and C. N. Dewey, "RSEM: Accurate transcript quantification from RNA-seq data with or without a reference genome," *BMC Bioinformatics*, vol. 12, pp. 323, 2011.

[196] R. M. Siloto, K. Findlay, A. Lopez-Villalobos, E. C. Yeung, C. L. Nykiforuk, and M. M. Moloney, "The accumulation of oleosins determines the size of seed oil bodies in Arabidopsis," *The Plant Cell*, vol. 18, pp. 1961–1974, 2006.

[197] U. Stahl, A. S. Carlsson, M. Lenman, A. Dahlqvist, B. Q. Huang, W. Banas, A. Banas, and S. Stymne, "Cloning and functional characterization of a phospholipid: diacylglycerol acyltransferase from Arabidopsis," *Plant Physiology*, vol. 135, pp. 1324–1335, 2004.

[198] F. Gibellini and T. K. Smith, "The Kennedy pathway-de novo synthesis of phosphatidylethanolamine and phosphatidylcholine," *IUBMB Life*, vol. 62 (no. 6), pp. 414–428, 2010.

[199] G. S. Han, L. O'Hara, S. Siniossoglou, and G. M. Carman, "Characterization of the yeast DGK1-encoded CTP-dependent diacylglycerol kinase," *Journal of Biological Chemistry*, vol. 283, pp. 20443–20453, 2008.

[200] L. Allenbach and Y. Poirier, "Analysis of the alternative pathways for the beta-oxidation of unsaturated fatty acids using transgenic plants synthesizing

polyhydroxyalkanoates in peroxisomes," *Plant Physiology*, vol. 124, pp. 1159–1168, 2000.

[201] J. Puyaubert, B. Garbay, P. Costaglioli, W. Dieryck, T. J. Roscoe, M. Renard, C. Cassagne, and R. Lessire, "Acyl-CoA elongase expression during seed development in *Brassica napus*," *BBA Molecular and Cell Biology of Lipids*, vol. 1533, pp. 141-152, 2001.

[202] D. C. Taylor, T. Francis, Y. Guo, J. M. Brost, V. Katavic, E. Mietkiewska, E. Michael Giblin, S. Lozinsky, and T. Hoffman, "Molecular cloning and characterization of a KCS gene from *Cardamine graeca* and its heterologous expression in Brassica oilseeds to engineer high nervonic acid oils for potential medical and industrial use," *Plant Biotechnology Journal*, vol. 7, pp. 925-938, 2009.

[203] J. M. Shockey, M. S. Fulda, and J. A. Browse, "Arabidopsis contains nine long chain acyl-coenzyme A synthetase genes that participate in fatty acid and glycerol lipid metabolism," *Plant Physiology*, vol. 129, pp. 1710–1722, 2002.

[204] A. L. Quettier and P. J. Eastmond, "Storage oil hydrolysis during early seedling growth," *Plant Physiology and Biochemistry*, vol. 47, pp. 485–490, 2009.

[205] M. Taton and A. Rahier, "Plant sterol biosynthesis: identification and characterization of higher plant delta 7-sterol C5(6)-desaturase," *Archives of Biochemistry* and *Biophysics*, vol. 325 (no. 2), pp. 279–288, 1996.

[206] A. Conesa, S. Gotz, J. M. Garcia-Gomez, J. Terol, M. Talon, and M. Robles, "Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research," *Bioinformatics*, vol. 21 (no. 18), pp. 3674-3676, 2005.

[207] J. M. Stuart, E. Segal, D. Koller, and S. K. Kim, "A gene co-expression network for global discovery of conserved genetic modules," *Science*, vol. 302, pp. 249–255, 2003.

[208] C. Trapnell, B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van Baren, S. L. Salzberg, B. J. Wold, and L. Pachter, "Transcript assembly and quantification

by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation," *Nature Biotechnology*, vol. 28, pp. 511-515, 2010.

[209] A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, and B. Wold, "Mapping and quantifying mammalian transcriptomes by RNA-Seq," *Nature Methods*, vol. 5, pp. 621-628, 2008.

[210] F. Y. Peng and R. J. Weselake, "Gene co expression clusters and putative regulatory elements underlying seed storage reserve accumulation in Arabidopsis," *BMC Genomics*, vol. 12, pp. 286, 2011.

[211] F. Takaiwa, K. Oono, and A. Kato, "Analysis of the 50 flanking region responsible for the endosperm-specific expression of a rice glutelin chimeric gene in transgenic tobacco," *Plant Molecular Biology*, vol. 16, pp. 49-58, 1991.

[212] G. Q. Chen, C. Turner, X. He, T. Nguyen, T. A. McKeon, and D. Laudencia-Chingcuanco, "Expression profiles of genes involved in fatty acid and triacylglycerol synthesis in Castor bean (*Ricinus communis* L)," *Lipids*, vol. 42, pp. 263-274, 2007.

[213] U. Chandrasekaran and A. Liu, "Differential expression patterns of lipid genes in developing seeds of Castor bean (*Ricinus communis* L.)," *Bangladesh Journal of Botany*, vol. 43 (no. 1), pp. 37-43, 2014.

[214] Y. C. Wong, Q. B. Kwong, H. L. Lee, C. K. Ong, S. Mayes, F. T. Chew, D. R. Appleton, and H. Kulaveerasingam, "Expression comparison of oil biosynthesis genes in oil palm mesocarp tissue using custom array," *Microarrays*, vol. 3 (no. 4), pp. 263-281, 2014.

[215] M. A. Troncoso-Ponce, A. Kilaru, X. Cao, T. P. Durrett, J. Fan, J. K. Jensen, N. A. Thrower, M. Pauly, C. Wilkerson, and J. B. Ohlrogge, "Comparative deep transcriptional profiling of four developing oilseeds," *The Plant Journal*, vol. 68, pp. 1014–1027, 2011.

[216] W. Goettel, E. Xia, R. Upchurch, M. L. Wang, P. Chen, and Y.Q.C. An, "Identification and characterization of transcript polymorphisms in soybean lines varying in oil composition and content," *BMC Genomics*, vol. 15, pp. 299, 2014.

[217] D. S. Knutzon, J. L. Bleibaum, J. Nelsen, J. C Kridl, and G.A. Thompson, "Isolation and characterization of two safflower oleoyl-acyl carrier protein thioesterase cDNA clones," *Plant Physiology*, vol. 100, pp. 1751–1758, 1992.

[218] K. Haralampidis, D. Milioni, J. Sanchez, M. Baltrusch, E. Heinz, and P. Hatzopoulos, "Temporal and transient expression of stearoyl-ACP carrier protein desaturase gene during olive fruit development," *Journal of Experimental Botany*, vol. 49, pp. 1661–1669, 1998.

[219] T. Dominguez, M. L. Hernandez, J. C. Pennycooke, P. Jimenez, J. M. Martinez-Rivas, C. Sanz, E. J. Stockinger, J. J. Sasnchez-Serrano, and M. Sanmartín, "Increasing omega-3 desaturase expression in tomato results in altered aroma profile and enhanced resistance to cold stress," *Plant Physiology*, vol. 153, pp. 655–665, 2010.

[220] G. S. Suresha and I. M. Santha, "Molecular cloning and *in silico* analysis of novel oleate desaturase gene homologues from *Brassica juncea* through sub-genomic library approach," *Plant Omics Journal*, vol. 6, pp. 55-64, 2013.

[221] G. Chen, S. Qu, Q. Wang, F. Bian, Z. Peng, Y. Zhang, H. Ge, J. Yu, N. Xuan, Y. Bi, and Q. He, "Transgenic expression of delta-6 and delta-15 fatty acid desaturases enhances omega-3 polyunsaturated fatty acid accumulation in Synechocystis sp. PCC6803," *Biotechnology for Biofuels*, vol. 7, pp. 32, 2014.

[222] Z. Zheng, Q. Xia, M. Dauk, W. Shen, G. Selvaraj, and J. Zou, "Arabidopsis AtGPAT1, a member of the membrane-bound glycerol-3-phosphate acyltransferase gene family, is essential for tapetum differentiation and male fertility," *The Plant Cell*, vol. 15, pp. 1872-1887, 2003.

[223] Y. M. Mousa, D. Gerasopoulos, I. Metzidakis, and A. Kiritsakis, "Effect of altitude on fruit and oil quality characteristics of Mastoides olives," *Journal of the Science of Food and Agriculture*, vol. 71, pp. 345-350, 1996.

[224] R. J. Abbott and A. C. Brennan, "Altitudinal gradients, plant hybrid zones and evolutionary novelty," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 369, pp. 20130346, 2014.

[225] L. Suter, M. Ruegg, N. Zemp, L. Hennig, and A. Widmer, "Gene Regulatory Variation Mediates Flowering Responses to Vernalization along an Altitudinal Gradient in Arabidopsis," *Plant Physiology*, vol. 166 (no. 4), pp. 1928–1942, 2014.

[226] P. O'Hara, A. R. Slabas, and T. Fawcett, "Fatty acid and lipid biosynthetic genes are expressed at constant molar ratios but different absolute levels during embryogenesis," *Plant Physiology*, vol. 129, pp. 310–20, 2002.

[227] A. Ibá˜nez-Salazar, S. Rosales-Mendoza, A. Rocha-Uribe, J. I. Ramirez-Alonso, I. Lara-Hernandez, A. Hernandez-Torres, L. M. Paz-Maldonado, A. S. Silva-Ramirez, B. Banuelos-Hernandez, J. L. Martinez-Salgado, and R. E. Soria-Guerra, "Over-expression of Dof-type transcription factor increases lipid production in *Chlamydomonas reinhardtii*," *Journal of Biotechnology*, vol. 184, pp. 27-38, 2014.

[228] L. Zhou, Y. Liu, Z. Liu, D. Kong, M. Duan, and L. Luo, "Genome-wide identification and analysis of drought-responsive microRNAs in *Oryza sativa*," *Journal of Experimental Botany*, vol. 61, pp. 4157-4168, 2010.

[229] M. K. Udvardi, K. Kakar, M. Wandrey, O. Montanari, J. Murray, A. Andriankaja, J. Y. Zhang, V. Benedito, J. M. I. Hofer, F. Chueng, et al., "Legume transcription factors: global regulators of plant development and response to the environment," *Plant Physiology*, vol. 144, pp. 538–549, 2007.

[230] M. C. Arias, C. Luna, M. Rodríguez, S. L. Lenardon, and E. Taleisnik, "Sunflower chlorotic mottle virus in compatible interactions with sunflower: ROS generation and antioxidant response," *European Journal of Plant Pathology*, vol. 113, pp. 223-232, 2005.

[231] A. D. Ziems, L. J. Giesler, G. L. Graef, M. G. Redinbaugh, J. L. Vacha, S.Berry, L. V. Madden, and A. E. Dorrance, "Response of soybean cultivars to bean pod mottle virus infection," *Plant Disease*, vol. 91, pp. 719–726, 2007.

[232] O. Arogundade, S. O. Balogun, and T. H. Aliyu, "Effects of cowpea mottle virus and cucumber mosaic virus on six soybean (*Glycine max* L.) cultivars," *Virology Journal*, vol. 6, pp. 220-225, 2009.

[233] C. N. Jay, S. Rossall, and H. G. Smith, "Effects of beet western yellows virus on growth and yield of oilseed rape (*Brassica napus*)," *The Journal of Agricultural Science*, vol. 133, pp. 131-139, 1999.

[234] D. Bilgin, J. A. Zavala, J. Zhu, S. J. Clough, D. R. Ort, and E.H. DeLucia, "Biotic stress globally down regulates photosynthesis genes," *Plant, Cell & Environment*, vol. 33, pp. 1597–1613, 2010.

[235] K. Satoh, H. Kondoh, T. Sasaya, T. Shimizu, I. R. Choi, et al., "Selective modification of rice (*Oryza sativa*) gene expression by *Rice stripe virus* infection," *Journal of General Virology*, vol. 91, pp. 294–305, 2010.

[236] M. Mardi, L. Karimi Farsad, J. Gharechahi, and G. H. Salekdeh, "In-depth transcriptome sequencing of mexican lime trees infected with *Candidatus Phytoplasma aurantifolia*," *PLoS One*, vol. 10 (no. 7), pp. e0130425, 2015.

[237] O. Šamajová, T. Takáč, D. von Wangenheim, et al., "Update on methods and techniques to study endocytosis in plants," in *Endocytosis in plants*, J. Šamaj, Ed. Berlin: Springer, 2012, pp 1–36.

[238] B. J. Cassone, S. Wijeratne, A. P. Michel, L. R. Stewart, Y. Chen, et al., "Virus-independent and common transcriptome responses of leafhopper vectors feeding on maize infected with semi-persistently and persistent propagatively transmitted viruses," *BMC Genomics*, vol. 15, pp. 133, 2014.

[239] E. Nielsen, A. Y. Cheung, and T. Ueda, "The regulatory RAB and ARF GTPases for vesicular trafficking," *Plant Physiology*, vol. 147, pp. 1516–1526, 2008.

[240] S. Haupt, G. H. Cowan, A. Ziegler, A. G. Roberts, K. J. Oparka, and L. Torrance, "Two plant-viral movement proteins traffic in the endocytic recycling pathway," *Plant Cell*, vol. 17, pp. 164–181, 2005.

[241] A. Sen, K. Madhivanan, D. Mukherjee, and R. C. Aguilar, "The epsin protein family: coordinators of endocytosis and signaling," *Biomolecular Concepts*, vol. 3 (no. 2), pp. 117–126, 2012.

[242] S. Signorelli, E. L. Coitino, O. Borsani, and J. Monza, "Molecular mechanisms for the reaction between OH radicals and proline: insights on the role as reactive oxygen species scavenger in plant stress," *The Journal of Physical Chemistry B*, vol. 118, pp. 337–347, 2014.

[243] F. Q. Guo and N. M. Crawford, "Arabidopsis nitric oxide synthase1 is targeted to mitochondria and protects against oxidative damage and dark-induced senescence," *The Plant Cell*, vol. 17, pp. 3436–3450, 2005.

[244] F. Vlad, T. Spano, D. Vlad, F. B. Daher, A. Ouelhadj, S. Fragkostefanakis, et al., "Involvement of Arabidopsis prolyl 4 hydroxylases in hypoxia, anoxia and mechanical wounding," *Plant Signaling & Behavior*, vol. 2, pp. 368–369, 2007.

[245] P. Agudelo-Romero, P. Carbonell, F. de la Iglesia, J. Carrera, G. Rodrigo, et al., "Changes in the gene expression profile of *Arabidopsis thaliana* after infection with Tobacco etch virus," *Virology Journal*, vol. 5, pp. 92, 2008.

[246] N. K. Dubey, R. Goel, A. Ranjan, P. K. Singh, S. V. Sawant, et al., "Comparative transcriptome analysis of *Gossypium hirsutum L.* in response to sap sucking insects: aphid and whitefly," *BMC Genomics*, vol. 14, pp. 241, 2013.

[247] Y. Liu, Y. Jiang, J. Lan, Y. Zou, and J. Gao, "Comparative transcriptomic analysis of the response to cold acclimation in *Eucalyptus dunnii*," *PLoS ONE*, vol. 9 (no. 11), pp. e113091, 2014.

[248] V. Pavet, E. Olmos, G. Kiddle, S. Mowla, S. Kumar, J. Antoniw, M. E. Alvarez, and C. H. Foyer, "Ascorbic acid deficiency activates cell death and disease resistance responses in Arabidopsis," *Plant Physiology*, vol. 139, pp. 1291–1303, 2005.

[249] A. M. Laxalt and T. Munnik, "Phospholipid signaling in plant defense," *Current Opinion in Plant Biology*, vol. 5, pp. 332–338, 2002.

[250] B. Thomma, I. Penninckx, W. F. Broekaert, and B. P. A. Cammue, "The complexity of disease signaling in Arabidopsis," *Current Opinion in Immunology*, vol. 13, pp. 63–68, 2001.

[251] J. Freitas-Astúa, M. Bastianel, E. C. Locali-Fabris, V. M. Novelli, A. C. Silva-Pinhati, A. C. Basílio-Palmieri, M. L. P. N. Targon, and M. A. Machado, "Differentially expressed stress-related genes in the compatible citrus-Citrus leprosis virus interaction," *Genetics* and *Molecular Biology*, vol. 30, pp. 980-990, 2007.

[252] G. Breuer, L. de Jaeger, V. P. G. Artus, D. E. Martens, J. Springer, R. B. Draaisma, G. Eggink, R. H. Wijffels, and P. P. Lamers, "Superior triacylglycerol (TAG) accumulation in starchless mutants of *Scenedesmus obliquus*: (II) evaluation of TAG yield and productivity in controlled photobioreactors," *Biotechnology for Biofuels*, vol. 7, pp. 70, 2014.

[253] J. Fan, C. Yan, C. Andre, J. Shanklin, J. Schwender, and C. Xu, "Oil accumulation is controlled by carbon precursor supply for fatty acid synthesis in *Chlamydomonas reinhardtii*," *Plant* and *Cell Physiology*, vol. 53, pp. 1380–1390, 2012.

[254] M. Molhoj, R. Verma, and W. D. Reiter, "The biosynthesis of D-galacturonate in plants: functional cloning and characterization of a membrane-anchored UDP-D-glucuronate 4-epimerase from Arabidopsis*," Plant Physiology,* vol. 135, pp. 1221–1230, 2004.

[255] E. G. Burget, R. Verma, M. Molhoj, and W. D. Reiter, "The biosynthesis of L-arabinose in plants: molecular cloning and characterization of a Golgi-localized

UDP-D-xylose 4-epimerase encoded by the MUR4 gene of Arabidopsis," *The Plant Cell*, vol. 15, pp. 523-531, 2003.

[256] R. Li, C. S. Tee, Y. L. Jiang, X. Y. Jiang, P. N. Venkatesh, R. Sarojam, and J. Ye, "A terpenoid phytoalexin plays a role in basal defense of *Nicotiana benthamiana* against Potato virus X," *Scientific Reports*, vol. 5, pp. 9682, 2015.

[257] V. Pallas and J. A. Garcia, "How do plant viruses induce disease? Interactions and interference with host components," *Journal of General Virology*, vol. 92, pp. 2691-2705, 2011.

[258] M. N. Maruthi, S. Bouvaine, H. A. Tufan, I. U. Mohammed, and R. J. Hillocks, "Transcriptional response of virus-infected cassava and identification of putative sources of resistance for cassava brown streak disease," *PLoS One*, vol. 9, pp. e96642, 2014.

[259] R. J. Whenham and R. S. S. Fraser, "Stimulation by abscisic acid of RNA synthesis in discs from healthy and tobacco mosaic virus-infected tobacco leaves," *Planta*, vol. 150, pp. 349–353, 1990.

[260] S. Pennazio and P. Roggero, "Plant hormones and plant virus diseases. The auxins," *New Microbiologica*, vol. 19 (no. 4), pp. 369–378, 1996.

[261] P. E. Jameson, "Cytokinins and auxins in plant-pathogen interactions-an overview," *Journal of Plant Growth Regulation*, vol. 32, pp. 369-380, 2000.

[262] M. Alazem and N. S. Lin, "Roles of plant hormones in the regulation of host–virus interactions," *Molecular Plant Pathology*, vol. 16, pp. 529–540, 2014.

[263] M. Alazem, K.Y. Lin, and N.S. Lin, "The abscisic acid pathway has multifaceted effects on the accumulation of bamboo mosaic virus," *Molecular Plant–Microbe Interactions*, vol. 27, pp. 177–189, 2014.

[264] C. L. Casteel, M. De Alwis, A. Bak, H. Dong, S. A. Whitham, and G. Jander, "Disruption of ethylene responses by turnip mosaic virus mediates suppression of

plant defense against the green peach aphid vector," *Plant Physiology*, vol. 169, pp. 209–218, 2015.

[265] T. Qi, S. Song, Q. Ren, D. Wu, H. Huang, Y. Chen, et al., "The Jasmonate-ZIM-domain proteins interact with the WD-Repeat/bHLH/MYB complexes to regulate Jasmonate-mediated anthocyanin accumulation and trichome initiation in *Arabidopsis thaliana*," *The Plant Cell*, vol. 23, pp. 1795–1814, 2011.

[266] R. Tokutsu, H. Teramoto, Y. Takahashi, T. A. Ono, and J. Minagawa, "The light-harvesting complex of photosystem I in *Chlamydomonas reinhardtii*: protein composition, gene structures and phylogenic implications," *Plant and Cell Physiology*, vol. 45, pp. 138–145, 2004.

[267] K. Fukuyama, "Structure and function of plant-type ferredoxins," *Photosynthesis Research*, vol. 81, pp. 289–301, 2004.

[268] X. Johnson and J. Alric, "Central carbon metabolism and electron transport in *Chlamydomonas reinhardtii*: metabolic constraints for carbon partitioning between oil and starch," *Eukaryotic Cell*, vol. 12 (no. 6), pp. 776–793, 2013.

[269] H. Z. Wang, P. J. Zhao, J. C. Xu, H. Zhao, and H. S. Zhang, "Virus resistance in transgenic watermelon plants containing a WMV-2 coat protein gene," *Acta Genetica Sinica*, vol. 30, pp. 70-75, 2003.

[270] K. Weng, Z. Q. Li, R. Q. Liu, L. Wang, Y. J. Wang, and Y. Xu, "Transcriptome of *Erysiphe necator*-infected *Vitis pseudoreticulata* leaves provides insight into grapevine resistance to powdery mildew," *Horticultural Research*, vol. 1, article 14049, 2014.

[271] S. Lev-Yadun and K. S. Gould, "Role of anthocyanins in plant defense," in *Life's colorful solutions: the biosynthesis, functions, and applications of anthocyanins*, K. S. Gould, K. M. Davies, C. Winefield, Eds. Berlin, Germant: Springer-Verlag 2009, pp. 21-48.

[272] B. Kaplan, T. Sherman, and H. Fromm, "Cyclic nucleotide-gated channels in plants," *FEBS Letters*, vol. 581 (no. 12), pp. 2237–2246, 2007.

[273] W. Moeder, W. Urquhart, H. Ung, and K. Yoshioka, "The role of cyclic nucleotide gated ion channels in plant immunity," *Molecular Plant*, vol. 4, pp. 442–452, 2011.

[274] K. Chin, T. A. DeFalco, W. Moeder, and K. Yoshioka, "The Arabidopsis cyclic nucleotide-gated ion channels AtCNGC2 and AtCNGC4 work in the same signaling pathway to regulate pathogen defense and floral transition," *Plant Physiology*, vol. 163, pp. 611–624, 2013.

[275] A. S. N. Reddy, G. S. Ali, H. Celesnik, and I. S. Day, "Coping with stresses: roles of calcium- and calcium/calmodulin-regulated gene expression," *The Plant Cell*, vol. 23, pp. 2010–2032, 2011.

[276] C. Cheval, D. Aldon, J. P. Galaud, and B. Ranty, "Calcium/calmodulin-mediated regulation of plant immunity," *Biochimica et Biophysica Acta*, vol. 1833, pp. 1766–1771, 2013.

[277] K. F. Pedley and G .B. Martin, "Role of mitogen-activated protein kinases in plant immunity," *Current Opinion in Plant Biology*, vol. 8, pp. 541–547, 2005.

[278] G. Tena, M. Boudsocq, and J. Sheen, "Protein kinase signaling networks in plant innate immunity," *Current Opinion in Plant Biology*, vol. 14, pp. 519–529, 2011.

[279] P. Singh and L. Zimmerli, "Lectin receptor kinases in plant innate immunity," *Frontiers in Plant Science*, vol. 4, pp. 1–4, 2013.

[280] H. Seybold, F. Trempel, and S. Ranf, "Ca^{2+} signalling in plant immune response: from pattern recognition receptors to Ca^{2+} decoding mechanisms," *New Phytologist*, vol. 204, pp. 782–790, 2014.

[281] L. Zhang, L. Du, and B.W. Poovaiah, "Calcium signaling and biotic defense responses in plants," *Plant Signaling and Behavior*, vol. 9 (no. 11), pp. e973818, 2014.

[282] S. A. Whitham, S. Quan, H. S. Chang, B. Cooper, B. Estes, T. Zhu, X. Wang, and Y. M. Hou, "Diverse RNA viruses elicit the expression of common sets of genes

in susceptible *Arabidopsis thaliana* plants," *The Plant Journal*, vol. 33, pp. 271–283, 2003.

[283] M. Babu, A. G. Gagarinova, J. E. Brandle, and A. Wang, "Association of the transcriptional response of soybean plants with soybean mosaic virus systemic infection," *Journal of General Virology*, vol. 89, pp. 1069–1080, 2008.

[284] P. Agudelo-Romero, P. Carbonell, M. A. Perez-Amador, and S. F. Elena, "Virus adaptation by manipulation of host's gene expression," *PLoS One*, vol. 3, pp. e2397, 2008.

[285] E. J. Pierce and M. E. Rey, "Assessing global transcriptome changes in response to South African cassava mosaic virus [ZA-99] infection in susceptible *Arabidopsis thaliana*," *PLoS One*, vol. 8 (no. 6), pp. e67534, 2013.

[286] J. T. Ascencio-Ibanez, R. Sozzani, T. J. Lee, T. M. Chu, R. D. Wolfinger, R. Cella, and L. Hanley-Bowdoin, "Global analysis of Arabidopsis gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection," *Plant Physiology*, vol. 148, pp. 436–454, 2008.

[287] X. Wang, P. Sameer, P. Goregaoker, and J. N. Culver, "Interaction of the Tobacco mosaic virus replicase protein with a NAC domain transcription factor is associated with the suppression of systemic host defenses," *Journal of Virology*, vol. 83, pp. 9720–9730, 2009.

[288] M. Agbeci, R. Grangeon, R. S. Nelson, H. Zheng, and J. F. Laliberté, "Contribution of host intracellular transport machineries to intercellular movement of turnip mosaic virus," *PLoS Pathogens*, vol. 9, pp. e1003683, 2013.

[289] A. Jakubiec, J. Notaise, V. Tournier, F. Hericourt, M. A. Block, G. Drugeon, L. van Aelst, and I. Jupin, "Assembly of turnip yellow mosaic virus replication complexes: Interaction between the proteinase and polymerase domains of the replication proteins," *Journal of Virology*, vol. 78 (no. 15), pp. 7945-7957, 2004.

[290] T. Panavas, C. M. Hawkins, Z. Panaviene, and P. D. Nagy, "The role of the p33: p33/p92 interaction domain in RNA replication and intracellular localization of

p33 and p92 proteins of Cucumber necrosis tombusvirus," *Virology*, vol. 338 (no. 1), pp. 81-95, 2005.

[291] J. Mammadov, R. Aggarwal, R. Buyyarapu, and S. Kumpatla, "SNP markers and their impact on plant breeding. *International Journal of Plant Genomics*, vol. 2012, article ID 728398, 2012.

[292] R. H. Ramirez-Gonzalez, C. Uauy, and M. Caccamo, "PolyMarker: a fast polyploid primer design pipeline," *Bioinformatics*, vol. 31 (no. 12), pp. 2038-2039, 2015.

[293] S. K. Parida, M. Mukerji, A. K. Singh, N. K. Singh, and T. Mohapatra, "SNPs in stress-responsive rice genes: validation, genotyping, functional relevance and population structure," *BMC Genomics*, vol. 13 (no. 1), pp. 426, 2012.

[294] Y. Lee, N. Jeong, J. H. Kim, K. J. Lee, and K. H. Kim, "Development, validation and genetic analysis of a large soybean SNP genotyping array," *The Plant Journal*, vol. 81, pp. 625-636, 2015.

[295] E. Zuriaga, J. M. Soriano, T. Zhebentyayeva, C. Romero, C. Dardick, J. Cañizares, and M. L. Badenes, "Genomic analysis reveals MATH gene (s) as candidate for Plum pox virus (PPV) resistance in apricot (*Prunus armeniaca* L.)," *Molecular Plant Pathology*, vol. 13, pp. 663–677, 2013.

[296] A. Kwon, P. Bednarek, and P. Schulze-Lefert, "Secretory pathways in plant immune responses," *Plant Physiology*, vol. 147, pp. 1575–1583, 2008.

[297] L. G. Leal, C. L´opez, and L. L´opez-Kleine, "Construction and comparison of gene co-expression networks shows complex plant immune responses," *PeerJ*, vol. 2, pp. e610, 2014.

[298] Z. L. Zheng and Y. Zhao, "Transcriptome comparison and gene coexpression network analysis provide a systems view of citrus response to *'Candidatus* Liberibacter asiaticus' infection," *BMC Genomics*, vol. 14, pp. 27, 2013.

[299] J. P. Tully, A. E. Hill, H. M. Ahmed, R. Whitley, A. Skjellum, and M. S. Mukhtar, "Expression-based network biology identifies immune-related functional modules involved in plant defense," *BMC Genomics*, vol. 15, pp. 421, 2014.

[300] S. Yang, X. Zhang, J. X. Yue, D. Tian, and J. Q. Chen, “Recent duplications dominate NBS-encoding gene expansion in two woody species,” *Molecular Genetics and Genomics*, vol. 280, pp. 187-198, 2008.

[301] P. L. Morrell, E. S. Buckler, and J. Ross-Ibarra, “Crop genomics: advances and applications," *Nature Reviews Genetics*, vol. 13, pp. 85-96, 2012.

[302] S. H. Hulbert, C. A. Webb, S. M. Smith, and Q. Sun, "Resistance gene complexes: evolution and utilization," *Annual Review of Phytopathology*, vol. 39, pp. 285-312, 2001.

[303] M. Deutsch and M. Long, “Intron-exon structures of eukaryotic model organisms,” *Nucleic Acids Research*, vol. 27, pp. 3219–3228, 1999.

[304] S. P. Dinesh-Kumar and B. J. Baker, "Alternatively spliced N resistance gene transcripts: their possible role in tobacco mosaic virus resistance," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, pp. 1908-1913, 2000.

[305] A. S. Reddy, “Alternative splicing of pre-messenger RNAs in plants in the genomic era,” *Annual Review of Plant Biology*, vol. 58, pp. 267-294, 2007.

[306] S. Tang and A. Riva, “PASTA: splice junction identification from RNA-sequencing data,” *BMC Bioinformatics*, vol. 14, pp. 116, 2013.

[307] P. J. Rushton and I. E. Somssich, “Transcriptional control of plant genes responsive to pathogens,” *Current Opinion in Plant Biology*, vol. 1, pp. 311–315, 1998.

[308] R. Shin, J. H. Han, G. J.Lee, and K. H. Peak, "The potential use of a viral coat protein gene as a transgene screening marker and multiple virus resistance of pepper plants coexpressing coat proteins of cucumber mosaic virus and tomato mosaic virus," *Transgenic Research*, vol. 11, pp. 215-219, 2002.

[309] F. Vailleau, X. Daniel, M. Tronchet, J.-L.Montillet, C. Triantaphylidès, and D. Roby, "A R2R3-MYB gene, AtMYB30, acts as a positive regulator of the hypersensitive cell death program in plants in response to pathogen attack," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, pp. 10179-10184, 2002.

[310] C. Kirsch, E. Logemann, B. Lippok, E. Schmelzer, and K. Hahlbrock, "A highly specific pathogen-responsive promoter element from the immediate-early activated CMPG1 gene in *Petroselinum crispum*," *The Plant Journal*, vol. 26, pp. 217-227, 2001.

[311] A. Heise, B. Lippok, C. Kirsch, and K. Hahlbrock, "Two immediate-early pathogen-responsive members of the AtCMPG gene family in *Arabidopsis thaliana* and the W-box-containing elicitor-response element of AtCMPG1," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, pp. 9049-9054, 2002.

[312] P. S. Choi, Y. D. Kim, K. M. Choi, H. J. Chung, D. W. Choi, and J. R. Liu, "Plant regeneration from hairy-root cultures transformed by infection with Agrobacterium rhizogenes in *Catharanthus roseus*" *Plant Cell Reports*, vol. 22, pp. 828-831, 2004.

[313] M. Nuruzzaman, A. M. Sharoni, and S. Kikuchi, "Roles of NAC transcription factors in the regulation of biotic and abiotic stress responses in plants," *Frontiers in Microbiology*, vol. 4, pp. 248, 2013.

[314] Z. Zhang, J. Chen, Y. Su, H. Liu, Y. Chen, P. Luo, et al., "*TaLHY*, a 1R-MYB transcription factor, plays an important role in disease resistance against stripe rust fungus and ear heading in wheat," *PLoS One*, vol. 10 (no. 5), pp. e0127723, 2015.

[315] A. Coego, V. Ramirez, M. J. Gil, V. Flors, B. Mauch-Mani, and P. Vera, "An Arabidopsis homeodomain transcription factor, OVEREXPRESSOR OF CATIONIC PEROXIDASE 3, mediates resistance to infection by necrotrophic pathogens," *The Plant Cell*, vol. 17, pp. 2123–2137, 2005.

[316] Q. Pan, J. Wendel, and R. Fluhr, "Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes," *Journal of Molecular Evolution*, vol. 50, pp. 203–213, 2000.

[317] V. Haake, D. Cook, J. L. Riechmann, O. Pineda, M. F. Thomashow, and J. Z. Zhang, "Transcription factor CBF4 is a regulator of drought adaptation in Arabidopsis," *Plant Physiology*, vol. 130, pp. 639-648, 2002.

[318] A. M. Voitsik, S. Muench, H. B. Deising, and L. M. Voll, “Two recently duplicated maize NAC transcription factor paralogs are induced in response to *Colletotrichum graminicola* infection,” *BMC Plant Biology*, vol. 13, pp. 85, 2013.

[319] L. Chen, L. Zhang, D. Li, F. Wang, and D. Yu, "WRKY8 transcription factor functions in the TMV-cg defense response by mediating both abscisic acid and ethylene signaling in Arabidopsis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, pp. 1963-1971, 2013.

[320] J. Yu, S. Tehrim, F. Zhang, et al., “Genome-wide comparative analysis of NBS-encoding genes between *Brassica* species and *Arabidopsis thaliana*,” *BMC Genomics*, vol. 15, pp. 3, 2014.

[321] A. P. Chan, J. Crabtree, Q. Zhao, et al., “Draft genome sequence of the oilseed species *Ricinus communis*, *Nature Biotechnology*, vol. 28, pp. 951–956, 2010.

[322] G. Chen, D. Pan, Y. Zhou, S. Lin, and X. Ke, "Diversity and evolutionary relationship of nucleotide binding site-encoding disease-resistance gene analogues in sweet potato (*Ipomoea batatas Lam*.)," *Journal of Biosciences*, vol. 32, pp. 713-721, 2007.

[323] Q. Chen, Z. Han, H. Jiang, D. Tian, and S. Yang, "Strong positive selection drives rapid diversification of R-genes in Arabidopsis relatives," *Journal of Molecular Evolution*, vol. 70, pp. 137-148, 2010.

[324] H. C. Rawal, N. K. Singh, and T. R. Sharma, "Conservation, divergence, and genome wide distribution of PAL and POX A gene families in plants," *International Journal of Genomics*, vol. 2013, pp. 678969, 2013.

[325] T. H. Kim, H. H. Kunz, S. Bhattacharjee, et al., “Natural variation in small molecule-induced TIR-NB-LRR signalling induces root growth arrest via EDS1- and PAD4- complexed R protein VICTR in Arabidopsis,” *The Plant Cell*, vol. 24, pp. 5177–5192, 2012.

PUBLICATIONS

Research papers:

- **Sood A**, Jaiswal V, Chanumolu SK, Malhotra N, Pal T, Chauhan RS (2014) Mining whole genomes and transcriptomes of Jatropha (*Jatropha curcas*) and Castor bean (*Ricinus communis*) for NBS-LRR genes and defense response associated transcription factors. Molecular Biology Reports. 41(11), 7683-7695. [ISSN: 0301-4851, IF: 2.5]

- **Sood A**, Chauhan RS (2015) Regulation of FA and TAG biosynthesis pathway genes in endosperms and embryos of high and low oil content genotypes of *Jatropha curcas* L. Plant Physiology & Biochemistry. 94, 253–267. [ISSN: 0981-9428, IF: 2.75]

- **Sood A**, Chauhan RS (2016) Comparative NGS transcriptomics unravels molecular components associated with mosaic virus infection in a bioenergy plant species, *Jatropha curcas* L. Bioenergy Research. 2016, 1-17. [IF – 3.30]

- **Sood A**, Jaiswal V, Chauhan RS. Identification of transcription factors regulating oil accumulation in endosperm of *Jatropha curcas* through transcriptome mining. (Communicated in 'Protoplasma', 2016)

Research papers (other than PhD work):

- Kumar P, Saini M, Bhushan S, Warghat AR, Pal T, Malhotra N, **Sood A** (2014) Effect of salicylic acid on the activity of PAL and PHB geranyltransferase and shikonin derivatives production in cell suspension cultures of *Arnebia euchroma* (Royle) Johnst—a medicinally important plant species. Applied Biochemistry and Biotechnology, 173(1), 248-258. [IF – 1.60]
- Gangwar M*, **Sood A***, Bansal A and Chauhan RS. Comparative transcriptomics reveals a reduction in carbon capture and flux between source and sink in cytokinin treated inflorescences of *Jatropha curcas* L. (* Equal contribution) (Under revision in 'Journal of Plant Physiology', 2016)

Presentations in National and International Conferences:

- **Sood A**, Sharma A and Chauhan RS (2013) Genomics-assisted genetic improvement of a bioenergy crop (*Jatropha curcas*). World Biotechnology Congress, Boston: June 3-6, 2013]

- **Sood A**, Chauhan RS (2014) Expression analysis of fatty acid biosynthesis genes in high versus low oil content genotypes of *Jatropha curcas*. Proceedings of the National Symposium on Advances in Biotechnology for Crop Improvement [Eternal University, Badu Sahib, HP : July 12, 2014]

- Chauhan RS, **Sood A** and Gangwar M (2014) Genomics-Assisted Genetic Improvement of *Jatropha curcas*. Jatropha Updates 2014. [TERI, New Delhi: October 13, 2014]

Book chapter:

- Chauhan RS and **Sood A** (2013) Comparative Genomics in Euphorbiaceae. In: Jatropha: Challenges for a New Energy Crop. Vol 2. Eds. B. Bahadur, M. Sujatha and N. Carels. Springer, New York pp 351-374.

Workshops attended:

- International workshop on "Computational Aspects of working with Genomes" at Institute of Microbial Technology (IMTECH, CSIR), Chandigarh, India from 27-29, March 2011.
- Training-cum-workshop on "Genome Analysis" at Bioinformatics Centre, Indian Institute of Technology, Delhi (IIT-D), India from 29-31, October 2012.
- "IInd National Workshop on TILLING in Crop Plants" at RTGR, Department of Plant Sciences, University of Hyderabad, Hyderabad, India from 7-18 December, 2015 sponsored by DBT, Govt. of India.