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 limiting 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 (CO2) 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 (*JcMV*) 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 (JcMD) 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 (*JcMD*), 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 CO₂ 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 homologues 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, $B_s2$ 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 $J. \text{curcas}$ 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.