

Chapter 31

Coconut

☆ *M. Neema, M.K. Rajesh, S.V. Ramesh
and P. Chowdappa*

1. Introduction

Coconut is one of the versatile and indispensable food items for millions of inhabitants in South, South-East Asia, and Pacific islands. Coconut (*Cocos nucifera*) is a monospecific member of the monocotyledonous family Arecaceae. Several hundred varieties of the coconut palms are grown all over the tropics, possessing high morphological, physiological and biochemical variations. The variability extends from tall to dwarf, large to small sized nuts, susceptibility and resistance to drought and salinity, high to low oil and fiber content *etc.* The uses of different parts of coconut tree (both edible and non-edible) from the leaves down to roots for many ages on tropical countries have given it the name 'Tree of Life'. The antimicrobial property of coconut husk has also been proved (Jose *et al.*, 2014). Coconut water is rich in potassium and is recommended as a health drink (Manna *et al.*, 2014). Despite the importance of coconut for humans and tropical ecosystems, the potential of the crop has not been utilized to the full extent. A number of barriers like non-availability of quality planting material and true to type seed nuts, susceptibility to biotic and abiotic stresses and fluctuating prices have considerably affected the productivity of coconut palms (CPCRI, 2015). Apart from conventional breeding techniques which are constrained due to the long life cycle of the palm, non-conventional techniques of tissue culture, molecular markers and 'omics' approaches are gaining importance. The genome sequence information of coconut is relatively scarce compared to other economically important palms like oil palm and date palm, where whole genome sequence has been made available in the public domain (Al-Dous *et al.*, 2011; Al-Mssallem *et al.*, 2013; Singh *et al.*, 2013; Dussert *et al.*, 2013).

Genome sequencing provides perspectives to tackle genetic problems, discovering regulatory signals, determining coding regions and the novel genes involved in growth and functioning of the system. Non-model plants like coconut, where the species possesses characters such as long life cycle, difficulty to grow in the laboratory or poor fecundity, have been schemed out of sequencing projects earlier due to high running cost of Sanger sequencing; consequently, the information about their genomics and key biological processes are inadequate. However, the advent of fast and cost effective next generation sequencing (NGS) platforms in the recent past has enabled the unearthing of certain characteristic gene structures unique to these species (Unamba *et al.*, 2015). The chloroplast and mitochondrial genomes of coconut have been sequenced (Huang *et al.*, 2013; Aljohi *et al.*, 2016), but the whole genome sequence is not available till date in the public domain. Transcriptome analysis, using RNA-Seq and assembly, could be utilized to gain global overview of the transcriptome of a plant, thus providing a wider outlook into the genetic mechanisms of growth, varietal characterization, and resistance to biotic and abiotic stresses. It has also aided in gaining insights about mechanisms underlying processes of secondary metabolism as well as facilitated development of genomic resources for evolutionary analysis and marker assisted breeding even without prior availability of genomic sequence information (Unamba *et al.*, 2015). This chapter discusses the recent research advances in genomics and transcriptomics of coconut.

2. Coconut Genome

2.1. Genome Size

The C-value (haploid genome size) of a species corresponds to the DNA amount in its unreplicated haploid or gametic nucleus (pollen or sperm), regardless of its ploidy level (Greilhuber *et al.*, 2005) and it is measured in picograms (pg) or base pairs (bp). Determination of the genome sizes and ploidy levels of cultivated coconuts are essential prerequisite for the sequencing of the coconut genome as it would provide the precise information regarding optimal depth of reads required for the accurate assembly and coconut genome annotation (Gunn *et al.*, 2015). The genome size of a species is associated with growth, development and adaptation to environmental changes. Genome size variation among related palm species is of evolutionary significance because changes characterize clads and may be associated with polyploidy, transposon amplification, deletion or rearrangements. Gunn *et al.* (2015) have conducted analysis, through flow cytometry, to estimate the genome sizes of 23 coconut cultivars which included dwarf and tall cultivars and hybrids. They found that the coconut genome is large, averaging 5.757 Gbp. The value was larger than the report of Zonneveld *et al.* (2005), but was consistent with data published by Sandoval *et al.* (2003) but differs from the results of Röser *et al.* (1997). Their results also suggest that intraspecific variation was associated with domestication and the variation among tall coconut was significantly greater in comparison to dwarfs.

2.2. Chloroplast Genome

Chloroplast is the photosynthetic organelle that has a genome of its own. Huang *et al.* (2013) reported the chloroplast genome sequence of a dwarf coconut plant. Gene content and organization, inverted repeat fluctuations, repeated sequence structure, and occurrence of RNA editing in coconut chloroplast genome were described in detail. Analysis of the data obtained from high-throughput sequencing demonstrated that the chloroplast genome of coconut is a typical quadripartite molecule within which a pair of inverted repeats is separated by a large single copy region and a small single copy region. The chloroplastic genome was 154,731 bp in length and was predicted to encode 130 genes and four pseudogenes. Although the chloroplast genome of coconut was the smallest found so far among palms, it shared the same overall organization, gene content and repeat structures that have been observed with chloroplast DNA sequenced from other palm species (Huang *et al.*, 2013). Some of the unique features found in the coconut chloroplast genome included pseudogenization of *rps19*-like gene and an unusually high number of RNA editing sites.

2.3. Mitochondrial Genome

Mitochondria are semi-autonomous bodies that contribute to energy production, metabolism and cell homeostasis. The mitochondrial genome (mt genome) codes for important polypeptides that build up complexes for the oxidative phosphorylation chain, together with nuclear-encoded subunits. In most plant species, the mt DNA gene sequences evolve very slowly and point mutations are rare because plant mitochondria contain an active DNA recombination system that allows copy correction of mutations (Gualberto *et al.*, 2014). The plant mitochondrial DNA (mtDNA) is large and variable in size (200 to 2,500 kb), contains many introns and repeated elements (typically 90 per cent of the total sequence), and experiences frequent gene gain/loss/transfer/duplication, and genome rearrangements (Galtier, 2011).

The coconut mitochondrial genome of an Oman Local Tall cultivar was sequenced by Aljohi *et al.* (2016). This was the second whole genome mitochondria assembly from *Arecaceae* family after that of date palm (Fang *et al.*, 2012). The coconut mitochondrial genome is around 679 kb in length with GC content of 45.5 per cent. Within the genome, 72 proteins, nine pseudogenes, 23 tRNAs, and three ribosomal RNAs were encoded. The chloroplast (cp) derived regions accounted for 5.07 per cent of the total assembly length in contrast to the date palm mt genome, where the 93.5 per cent of the genome sequence is cp derived (Fang *et al.*, 2012). In coconut, the cp-derived regions included 13 proteins, two pseudogenes, and 11 tRNAs. The mt genome of coconut has a relatively large fraction of repeat content (17.26 per cent), including both forward (tandem) and inverted (palindromic) repeats whereas in date palm there are 0.33 per cent tandem and 2.3 per cent long repeats. In coconut, sequence variation analysis shows that the transition transversion ratio of 0.3 in the mt genome was much lower (2-2.1) when compared to that of the nuclear genome (Aljohi *et al.*, 2016).

2.4. Whole Genome Sequencing of Coconut

Alsaihati *et al.* (2014) had reported a draft genome of coconut by generating seven libraries for pair-end and mate-pair genome sequencing using HiSeq. The estimated size of coconut genome was around 2.6 Gb, with repetitive sequences between 50 to 70 per cent based on draft genome data analysis. Since the genome assembly using ALLPATHS-LG and SOAPdenovo2 was unsatisfactory (Table 31.1), a custom pipeline was developed using SOAP assembly and gap closing modules, SSPACE scaffolder and intermediate correction scripts to produce quality genome draft with coverage of 94.5 per cent.

Table 31.1: Comparison of Coconut Draft Genome using different Assembly Pipelines (Alsaihati *et al.*, 2014)

<i>Draft</i>	<i>AllPathsLG</i>	<i>SoapDenovo2</i>	<i>Custom Pipeline</i>
Scaffolds	80,145	212,688	88,400
Total size (est. coverage)	0.97Gb (37 per cent)	2.6Gb (~100 per cent)	2.47Gb (94.5 per cent)
Longest scaffold	771Kb	891Kb	1246Kb
N50	48Kb	66.6Kb	114Kb
N90	8Kb	4Kb	14.8Kb
Total gap (per cent of draft size)	255Mb (26 per cent)	749Mb (29 per cent)	251Mb (10 per cent)

It is imperative to obtain whole genome sequence of coconut which is essential for precise understanding of various cellular processes, signaling pathways, defense mechanisms and host pathogen interactions of the palm.

3. Coconut Transcriptome

The transcriptome represents the whole complement of RNA transcripts in cells or tissues and reflects the expressed genes at various life stages, tissue types, physiological states, and environmental conditions. In crops where whole genome sequence is not available, transcriptome analysis using RNA-Seq technology enables the precise measurement of level of transcripts. Knowledge of the transcripts would be advantageous in determining the various genetic factors involved in various morphological and physiological processes in the plant. Next-generation transcriptome sequencing (RNA-seq) has been validated as an efficient methodology for production of genomic sequences by assembling short- read sequences (Xia *et al.*, 2011) with higher sensitivity and broader dynamic range. Studies based on large scale transcriptome profiling have given a fundamental insights into the aspects of co-expressing genes and their roles in metabolic pathways (Wickramasuriya and Dunwell, 2015), candidate genes associated with tolerance/resistance to different stresses as well as genes associated with several agronomic traits. Until 2012, the coconut transcriptome data made available publicly were very scarce; only 774 sequences were available in National Centre for Biotechnology Information database (Fan *et al.*, 2013).

3.1. Fatty Acid Biosynthesis and Metabolism

The first genome-wide study of coconut using transcriptome analysis was carried out by Fan *et al.* (2013). Total RNA was isolated from spear leaves, young leaves and fruit flesh of Hainan Tall cultivar. Illumina RNA-Seq technology was used to generate 54,931,406 short reads containing a total of 4,943,826,540 nucleotide bases. *De novo* assembly using the short read assembly programme 'Trinity' was applied to combine these short reads to 57,304 unigenes, with an average length of 752 bp. Unigene sequences were annotated against NCBI Nr database with E-value, 0.00001. They were also aligned by BLASTX to protein databases such as Swiss-Prot, KEGG and COG, in order to retrieve proteins with the highest sequence similarity. ESTScan, Blast2GO program, WEGO software and Blast All software's were used for annotations. Ten putative non-coding RNA genes were identified when the coconut transcriptome was compared to whole genome of date palm. A total of 347 unigenes involved in the biosynthesis and metabolism of fatty acid was also identified. These unigenes could be assigned to five steps of the fatty acid biosynthesis pathway, providing a means of elucidating the molecular mechanisms for fatty acid biosynthesis in coconut palm. Of these, 20 unigenes were predicted to be related to fatty acyl-ACP thioesterase, which is a crucial enzyme for terminating the elongation of carbon chains and therefore regulating the length of fatty acids. Therefore, the study suggested that the expression of fatty acyl-ACP thioesterase might be correlated with the observed accumulation of medium chain fatty acids (*i.e.* lauric acid) in coconut.

3.2. Embryogenesis

Somatic embryogenesis (SE), where a single or a group of somatic cells differentiate to form embryonic cells under suitable *in vitro* conditions (Zimmerman, 1993), is an ideal system to explore gene expression patterns associated with initial stages of embryo development. The formation of embryos from somatic cells closely resembles the developmental pathway of zygotic embryos (ZEs) and hence, the molecular information generated for the SE pathway could be used to explain the dynamic molecular interactions that take place during early embryogenesis (Zimmerman, 1993). Rajesh *et al.* (2015a) carried out *de novo* assembly and characterization of global transcriptome of coconut embryogenic calli using Illumina paired end sequencing. Here transcriptome analysis of coconut embryogenic calli derived from plumular explants of West Coast tall cultivar was undertaken on Illumina Hi Seq 2000 platform. The reads obtained (50,839,994 paired end reads; 7.73 Gb) were trimmed using a Perl script (46.2 million reads; 5.42GB), assembled using SOAPdenovo assembler and the trimmed reads were aligned using Bowtie programme. The assembled reads were subjected to annotation, classification and ontology analysis using BLASTx, BLAST2GO and KEGG programmes. Genes known to be involved in SE, namely protein kinases like receptor-like kinases [somatic embryogenesis receptor kinase (*SERK*) and *CLAVATA 1 (CLV1)*], mitogen-activated protein kinase (*MAPK*), transcription factors [*WUSCHEL (WUS)*, *APETALA2/Ethylene-responsive factor (AP2/ERF)*, *PICKLE (PKL)*, *AINTEGUMENTA (ANT)* and *WRKY*], extracellular proteins [arabinogalactan protein (*AGP*), Germin-like protein (*GLP*), embryogenic cell protein (*ECP*), and late embryogenesis-abundant

protein (*LEA*) and glutathione S-transferase (*GST*) were identified. Gene ontology (GO) annotations identified 8300 transcripts associated with biological processes majority having transcription and regulatory function, 13193 transcripts with molecular functions involved in ATP, zinc ion and metal ion binding processes and 6076 transcripts with cellular components; the highly expressed one were components integral to membrane followed by nucleus. Gene validation and expression analysis was carried out by real time quantitative PCR (RT-qPCR). For these, three stages of developmental pathway of somatic embryogenesis, namely initial calli, embryogenic calli and somatic embryos, were selected. The results indicated that *CLV* was upregulated in the initial stage of callogenesis. Transcripts of *GLP*, *GST*, *PKL*, *WUS* and *WRKY* were expressed more in somatic embryo stage. The expression of *SERK*, *MAPK*, *AP2*, *SAUR*, *ECP*, *AGP*, *LEA* and *ANT* were higher in embryogenic callus stage compared to initial culture and somatic embryo stages. This study was an effort to aid in the development of an efficient *in vitro* production protocol for coconut which is otherwise recalcitrant to *in vitro* culture.

Bandupriya *et al.* (2015) carried out a detailed analysis of ESTs by examining the transcriptome data of the different embryo tissue types together with one somatic tissue in order to identify important embryo-specific genes as well as other functional genes in different biochemical pathways in coconut. Four cDNA libraries were generated from immature embryo, mature embryo, microspore derived embryo and mature leaves and cDNA was sequenced by the Roche-454 GS-FLX system.

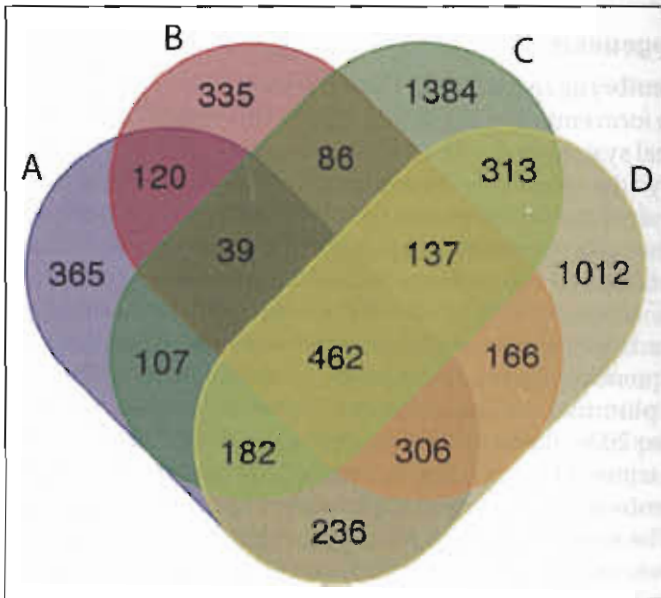


Figure 31.1: Venn Chart Showing Unique and Shared Unigenes found in Four Coconut Transcriptome Libraries.

(A) Immature embryos at the age of nine months after pollination, (B) Mature embryos at the age of 12 months after pollination, (C) Microspore-derived embryos and (D) Developing leaf.

The unigene sequences were annotated in BLAST; 462 unigenes were shared by all four libraries and the numbers of tissue specific unigenes were 365 in immature embryos at the age of nine months after pollination, 335 in mature embryos at the age of 12 months after pollination, 1012 in microspore-derived embryos and 1384 in developing leaf (Figure 31.1).

Comparative expression analysis demonstrated that the embryo tissue libraries shared certain degree of similarities while they were different from leaf abundant ESTs. The result emphasizes on the effectiveness of 454 sequencing approach for the identification of transcripts for a particular organ of coconut. In this study, a list of putative transcripts such as chitinase, beta-1,3-glucanase, ATP synthase CF0 subunit, thaumatin-like protein and metallothionein-like protein, which may be involved in various biological processes during embryogenesis were identified, which can further be utilized for future research.

3.3. Host-pathogen Interactions

The root (wilt) disease is a serious debilitating disease causing severe yield loss in coconut palms in Southern India (Figure 31.2). RNA-seq was utilized to characterize global transcriptome responsive to root (wilt) disease in coconut (Rajesh *et al.*, 2013). To explore the molecular mechanisms involved in compatible and incompatible interactions, transcriptome profiling of diseased and healthy Chowghat Green Dwarf (CGD) palms was conducted. The flow diagram of different steps followed is provided in Figure 31.3. RNA-Seq analysis generated more than 102 million 86 bp paired-end reads, which were assembled into 2,54,302 contigs, with a mean size of 488 bp. Based on sequence similarity searches, 1,66,429 (~ 65 per



Figure 31.2: Chowghat Green Dwarf Palm Affected by Root (Wilt) Disease.

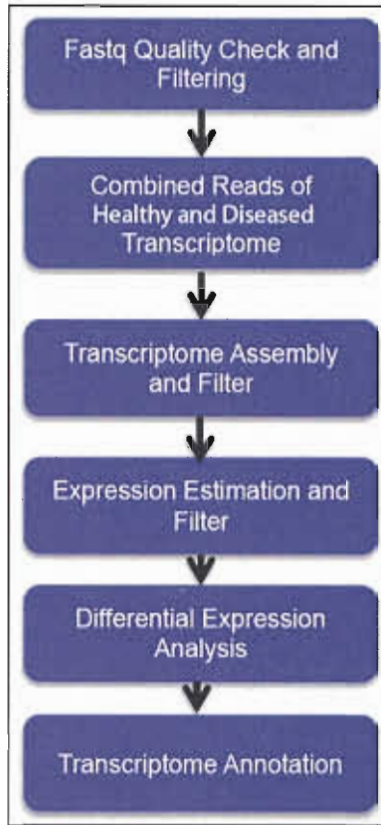


Figure 31.3: Flow Diagram of Whole Transcriptome Analysis from Healthy and Root (Wilt) Diseased Leaf Samples of Coconut (Rajesh *et al.*, 2013).

cent) of assembled transcripts had at least one significant hit (E^{-20}) in NCBI database. Differential gene expression analysis was performed statistically using DESeq program. Overall, 3,033 transcripts differentially expressed transcripts (p -value < 0.01 and at least 100 reads in one of the sample) could be detected between healthy and root (wilt) diseased samples. Of the total, 1266 transcripts were up-regulated in healthy samples. Many of these differentially expressed transcripts were primarily involved in defense responses, signalling pathways, cellular transport and other metabolic processes.

The gene ontology (GO) terms for transcripts were extracted wherever possible. The top 25 GO terms identified in molecular function, biological process and cellular component category are provided in Figures 31.4–31.6.

Transcriptome analysis to identify complex host pathogen interactions, with respect to coconut yellow decline disease, was carried out by Nejat *et al.* (2015). Whole transcriptome profiles of naturally infected leaves of Malayan Red Dwarf in response to yellow decline phytoplasma, and healthy leaves were generated through RNA-Seq technique. Illumina sequencing yielded a total of 72,019,264 and 70,935,896 reads from healthy and infected leaves transcriptome respectively which

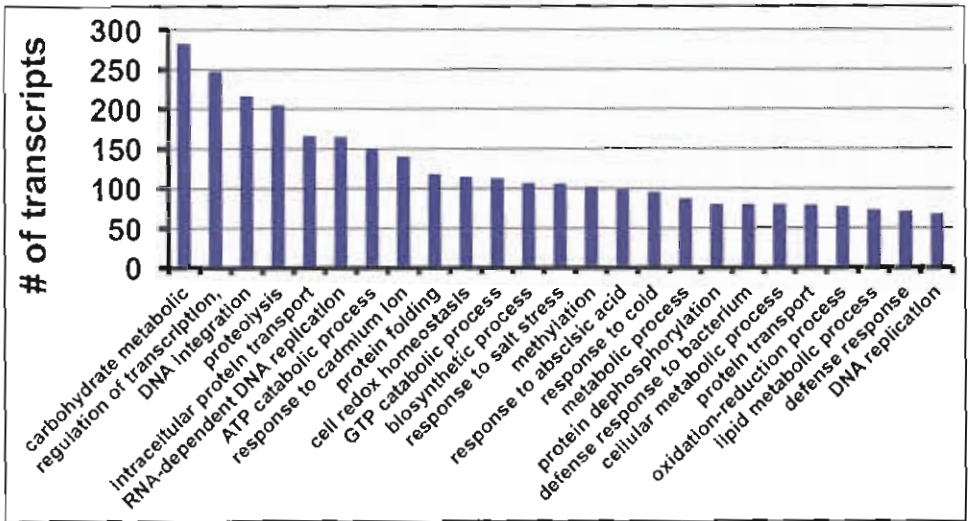


Figure 31.4: Top 25 Terms in Biological Process Category from GO Annotation.

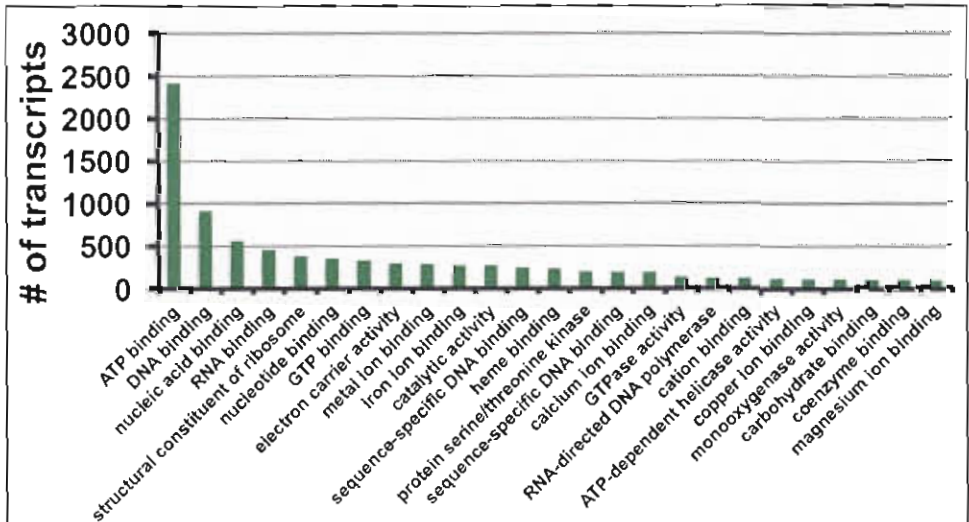


Figure 31.5: Top 25 Terms in Molecular Function Category from GO Annotation.

were assembled into 108,994 contigs for healthy and 148,264 contigs for diseased coconut palm. An analysis to determine the differentially expressed genes in the healthy as well as infected revealed that 18,013 transcripts were upregulated and 21,860 transcripts were down-regulated in infected leaves compared to healthy ones. This transcriptome analysis showed that many of the biological and cellular processes were reprogrammed in coconut due to phytoplasma infection. A number of genes associated with the production of defense related proteins, reactive oxygen species, ABC transport protein family, hydrolase and kinases, flavanol synthesis, auxin induced protein, no apical meristem (NAM) gene family, and ethylene were

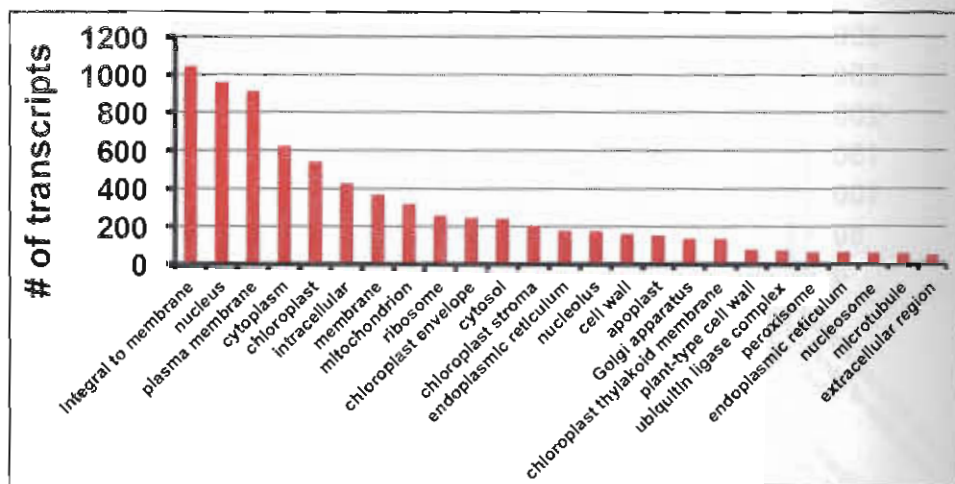


Figure 31.6: Top 25 Terms in Cellular Component Category from GO Annotation.

upregulated as a result of phytoplasmal infection. A total of 35 *WRKY* genes were induced, whereas 21 were suppressed in infected coconut leaves. Transcripts of cytokinin dehydrogenase were down regulated. Up-regulation of gibberellin-2-oxidase was presumed to down-regulate levels of gibberellins in infected coconut palms and result in stunting, inflorescence necrosis and premature nut fall, which are the characteristics of phytoplasma infection. Wide range of genes associated with photosynthesis was down regulated, which is part of the defense mechanism in response to biotic stress. It was also suggested that phytoplasma might alter the expression of genes involved in carbohydrate metabolism in infected palms to import sugars, which are their main source of energy, through ABC transporter system, which might explain the up-regulation of ABC protein family transcripts in phytoplasma infection.

Rajesh *et al.* (2015b) carried out comprehensive bioinformatics analysis of transcriptome data of leaf samples of coconut root (wilt) disease-resistant cultivar Chowghat Green Dwarf, generated through RNA-Seq, and identified 243 resistance gene analog (RGA) sequences, comprising six classes of RGAs. Domain and conserved motif predictions of clusters were performed to analyze the architectural diversity. Phylogenetic analysis of deduced amino acid sequences revealed that coconut NBS-LRR type RGAs were classified into distinct groups based on the presence of TIR or CC motifs in the N-terminal regions.

3.4. RNA Directed DNA Methylation

Huang *et al.* (2014) carried out transcriptome analysis to classify gene expression in seed tissues and leaves of dwarf coconut palm with a focus on the identification of factors involved in RNA-directed DNA methylation. In the study, RNA-Seq data were collected from maturing gelatinous endosperm, mature embryo and young leaf of fragrant dwarf green coconut. Of the three tissues, the leaf transcriptome had the maximum total number of sequencing reads of 121,151,552 with total unigenes

of 33,446 and embryo transcriptome had total sequencing reads 81,128,552 with total unigenes of 86254. Total transcripts in endosperm were 229,866. The study was focused on factors involved in RNA directed DNA methylation (RdDM) and suggests that the relative abundance of most of these factors is highest in endosperm tissue compared with leaves and embryos. The research results suggest that small RNA-mediated silencing pathways were active in coconut seeds, particularly maturing endosperm.

3.5. Aroma

The liquid endosperm of Aromatic Green Dwarf coconut of Thailand has a peculiar 'pandan-like' aroma due to the presence of 2-acetyl-1-pyrroline (2AP) (Saensuk *et al.*, 2016), which is also present in rice accessions (Grimm *et al.*, 2011; Arikrit *et al.*, 2011). *De novo* assembly of transcriptome from the aromatic coconut endosperm was undertaken by Saensuk *et al.* (2016) to identify the gene(s) responsible for 2AP biosynthesis. The gene, *CnAMADH2*, which is an ortholog of rice aromatic gene, was present in all Aromatic Green Dwarf accessions. But compared to other aromatic plants, in the case of coconut, a G-to-C substitution found in exon 14 and was associated exclusively with 2AP content. This base substitution resulted in an amino acid change of alanine to proline at position 442, which might have resulted in unstabilized dimer conformation that could lower AMADH enzyme activity. The authors confirmed the substitution by PCR based on this sequence variation.

4. Conclusion

The precise understanding of key processes and pathways are important for the genetic manipulations of the plant, thereby making them more productive and tolerant to biotic and abiotic stresses. The whole genome sequence of coconut is not yet published except for a draft genome. Due to the reduced size and complexity of the transcriptome relative to the genome, transcriptome sequencing provides a rapid, inexpensive approach to access gene sequences, gene expression abundances, and gene expression patterns in any species, including those that lack a reference genome sequence. To date, successful applications of RNA sequencing in conjunction with *de novo* transcriptome assembly has enabled identification of new genes in an array of biochemical pathways in plants (Gongora-Castillo *et al.*, 2013). Transcriptome analysis till date in coconut has provided insight into the genetic mechanism and factors influencing host pathogen interaction, somatic embryogenesis, RNA directed DNA methylation and other novel genes. These studies have increased transcriptomics resources for coconut and provide a foundation for further functional and molecular studies that will improve coconut through molecular breeding and genetic engineering technologies. Still there is a need for improving the sequencing platforms to obtain long and high quality output reads and better algorithms and bioinformatic softwares to handle this palm which possess a large genome. Finally, unraveling of genome sequence of coconut and large scale transcriptome studies would provide molecular insights and help provide break-through in successful coconut regeneration and genetic transformation that are essential for accelerating crop improvement programmes.

References

- Al-Dous, E.K., George, B., Al-Mahmoud, M.E., Al-Jaber, M.Y., Wang, H., Salameh, Y.M., Al-Azwani, E.K., Chaluvadi, S., Pontaroli, A.C., DeBarry, J. and Arondel, V. (2011). *De novo* genome sequencing and comparative genomics of date palm (*Phoenix dactylifera* L.). *Nature Biotechnology*, **29**: 521-527.
- Aljohi, H. A., Liu, W., Lin, Q., Zhao, Y., Zeng, J., Alamer, A., Alanazi, I.O., Alawad, A.O., Al-Sadi, A.M., Hu, S. and Yu, J. (2016). Complete sequence and analysis of coconut palm (*Cocos nucifera*) mitochondrial genome. *PLoS ONE*, **11**: 0163990. doi: 10.1371/journal.pone.0163990.
- Al-Mssallem, I.S., Hu, S., Zhang, X., Lin, Q., Liu, W., Tan, J., Yu, X., Liu, J., Pan, L., Zhang, T. and Yin, Y. (2013). Genome sequence of the date palm (*Phoenix dactylifera*). *Nature Communications*, **6**: 4.
- Alsaihati, B., W. Liu, Q. Lin, and Al-Mssallem, I. S. (2014). Coconut genome *de novo* sequencing. Genome sequencing and ESTs session, January 13, 2014. *Plant and Animal Genome XXII (International Plant and Animal Genome Conference)*, San Diego, California, USA. Poster PO46.
- Arikit, S., Yoshihashi, T., Wanchana, S., Uyen, T. T., Huong, N.T., Wongpornchai, S., and Vanavichit, A. (2011). Deficiency in the amino aldehyde dehydrogenase encoded by *GmAMADH2*, the homologue of rice *Os2AP*, enhances 2acetylpyrroline biosynthesis in soybeans (*Glycine max* L.). *Plant Biotechnology Journal*, **9**: 75-87.
- Bandupriya, H.D. and Dunwell, J.M. (2015). Transcriptome analysis for discovering candidate genes involve in embryogenesis in coconut (*Cocos nucifera* L.) through 454 pyrosequencing. *Journal of the National Science Foundation of Sri Lanka*, **43**: 319-336.
- CPCRI (2015). Vision 2050. *Indian Council of Agricultural Research*, New Delhi. 71p.
- Dussert, S., Guerin, C., Andersson, M., Joët, T., Tranbarger, T.J., Pizot, M., Sarah, G., Omore, A., Durand-Gasselien, T. and Morcillo, F. (2013). Comparative transcriptome analysis of three oil palm fruit and seed tissues that differ in oil content and fatty acid composition. *Plant Physiology*, **162**: 1337-1358.
- Fan, H., Xiao, Y., Yang, Y., Xia, W., Mason, A.S., Xia, Z., Qiao, F., Zhao, S. and Tang, H. (2013). RNA-Seq analysis of *Cocos nucifera*: Transcriptome sequencing and *de novo* assembly for subsequent functional genomics approaches. *PLoS ONE*, **8**: 59997. doi: 10.1371/journal.pone.0059997.
- Fang, Y., Wu, H., Zhang, T., Yang, M., Yin, Y., Pan, L., Yu, X., Zhang, X., Hu, S., Al-Mssallem, I.S. and Yu, J. (2012). A complete sequence and transcriptomic analyses of date palm (*Phoenix dactylifera* L.) mitochondrial genome. *PloS one*, **7**: 37164.
- Galtier, N. (2011). The intriguing evolutionary dynamics of plant mitochondrial DNA. *BMC Biology*, **9**: 1. doi: 10.1186/1741-7007-9-61.

- Gongora-Castillo, E. and Buell, C.R. (2013). Bioinformatics challenges in *de novo* transcriptome assembly using short read sequences in the absence of a reference genome sequence. *Natural Product Reports*, **30**: 490-500.
- Greilhuber, J., Dolezel, J., Lysak, M.A. and Bennett, M.D. (2005). The origin, evolution and proposed stabilization of the terms 'genome size' and 'C-value' to describe nuclear DNA contents. *Annals of Botany*, **95**: 255-260.
- Grimm, C.C., Champagne, E.T., Lloyd, S.W., Easson, M., Condon, B. and McClung, A. (2011). Analysis of 2-acetyl-1-pyrroline in rice by HSSE/GC/MS. *Cereal Chemistry*, **88**: 271-277.
- Gualberto, J.M., Mileshina, D., Wallet, C., Niazi, A.K., Weber-Lotfi, F. and Dietrich, A. (2014). *The plant mitochondrial genome: Dynamics and maintenance*. *Biochimie*, **100**: 107-120.
- Gunn, B.F., Baudouin, L., Beulé, T., Ilbert, P., Duperray, C., Crisp, M., Issali, A., Konan, J.L. and Rival, A. (2015). Ploidy and domestication are associated with genome size variation in Palms. *American Journal of Botany*, **102**: 1625-1633.
- Huang, Y.Y., Lee, C.P., Fu, J.L., Chang, B.C.H., Matzke, A.J. and Matzke, M. (2014). *De novo* transcriptome sequence assembly from coconut leaves and seeds with a focus on factors involved in RNA-directed DNA methylation. *G3: Genes Genomes Genetics*, **4**: 2147-2157.
- Huang, Y.Y., Matzke, A.J. and Matzke, M. (2013). Complete sequence and comparative analysis of the chloroplast genome of coconut palm (*Cocos nucifera*). *PLoS ONE*, **8**: 74736. doi: 10.1371/journal.pone.0074736.
- Jose, M., Cyriac, M.B., Pai, V., Varghese, I. and Shantaram, M. (2014). Anti-microbial properties of *Cocos nucifera* (coconut) husk: An extrapolation to oral health. *Journal of Natural Science Biology and Medicine*, **5**: 359-364.
- Manna, K., Khan, A., Das, D.K., Kesh, S.B., Das, U., Ghosh, S., Dey, R.S., Saha, K.D., Chakraborty, A., Chattopadhyay, S. and Dey, S. (2014). Protective effect of coconut water concentrate and its active component shikimic acid against hydroperoxide mediated oxidative stress through suppression of NF- κ B and activation of Nrf2 pathway. *Journal of Ethnopharmacology*, **155**: 132-146.
- Nejat, N., Cahill, D.M., Vadamalai, G., Ziemann, M., Rookes, J. and Naderali, N. (2015). Transcriptomics-based analysis using RNA-Seq of the coconut (*Cocos nucifera*) leaf in response to yellow decline phytoplasma infection. *Molecular Genetics and Genomics*, **290**: 1899-1910.
- Rajesh, M. K., Rachana, K. E., Babu, M., Thomas, R. J. and Anitha Karun (2013). Characterization of the global transcriptome responsive to root (wilt) disease in coconut using RNA-seq. In: *National Symposium on 'Pathogenomics for Diagnosis and Management of Plant Diseases*. CTCRI, Thiruvananthapuram, India, 24-25 October, 2013.
- Rajesh, M.K., Fayas, T.P., Naganeeswaran, S., Rachana, K.E., Bhavyashree, U., Sajini, K.K. and Karun, A. (2015a). *De novo* assembly and characterization of global transcriptome of coconut palm (*Cocos nucifera* L.) embryogenic calli using Illumina paired-end sequencing. *Protoplasma*, **253**: 913-928.

- Rajesh, M.K., Rachana, K.E., Naganeeswaran, S., Shafeeq, R., Thomas, R. J., Shareefa, M., Merin, B. and Anitha Karun (2015b). Identification of expressed resistance gene analog sequences in coconut leaf transcriptome and their evolutionary analysis. *Turkish Journal of Agriculture and Forestry*, **39**: 489-502.
- Röser, M., Johnson, M.A.T. and Hanson, L. (1997). Nuclear DNA amounts in palms (Arecaceae). *Acta Botanica*, **110**: 79-89.
- Saensuk, C., Wanchana, S., Choowongkomon, K., Wongpornchai, S., Kraithong, T., Imsabai, W., Chaichoompu, E., Ruanjaichon, V., Toojinda, T., Vanavichit, A. and Arikrit, S. (2016). De novo transcriptome assembly and identification of the gene conferring a "pandan-like" aroma in coconut (*Cocos nucifera* L.). *Plant Science*, **252**: 324-334.
- Sandoval, A., Hoher, V. and Verdeil, J.L. (2003). Flow cytometric analysis of the cell cycle in different coconut palm (*Cocos nucifera* L.) tissues cultured *in vitro*. *Plant Cell Reports*, **22**: 25-31.
- Singh, R., Ong-Abdullah, M., Low, E.T.L., Manaf, M.A.A., Rosli, R., Nookiah, R., Ooi, L.C.L., Ooi, S.E., Chan, K.L., Halim, M.A. and Azizi, N. (2013). Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature*, **500**: 335-339.
- Unamba, C.I.N., Nag, A. and Sharma, R.K. (2015). Next generation sequencing technologies: The doorway to the unexplored genomics of non-model plants. *Frontiers in Plant Science*, **6**: 1074. doi: 10.3389/fpls.2015.01074.
- Wickramasuriya, A.M. and Dunwell, J.M. (2015). Global scale transcriptome analysis of *Arabidopsis* embryogenesis *in vitro*. *BMC Genomics*, **16**: 301.
- Xia, Z., Xu, H., Zhai, J., Li, D., Luo, H., He, C. and Huang, X. (2011). RNA-Seq analysis and *de novo* transcriptome assembly of *Hevea brasiliensis*. *Plant Molecular Biology*, **77**: 299-308.
- Zimmerman, J.L. (1993). Somatic embryogenesis: A model for early development in higher plants. *The Plant Cell*, **5**: 1411.
- Zonneveld, B.J.M., Leitch, I.J. and Bennett, M.D. (2005). First nuclear DNA amounts in more than 300 angiosperms. *Annals of Botany*, **96**: 229-244.