Short Scientific Report

doi:10.25081/jpc.2021.v49.i3.7458
Journal of Plantation Crops, 2021, 49(3): 225-230

Transcriptome assembly of coconut endosperm callus (Laccadive Micro Tall cv.) and its functional annotation

Geethu Venugopal[1,2], K.P. Gangaraj[1], K.S. Muralikrishna[1], T.S. Keshava Prasad[3*] and M.K. Rajesh[1*]

[1]*ICAR-Central Plantation Crops Research Institute, Kasaragod-671 124, Kerala, India*
[2]*Mangalore University, Mangalagangotri, Mangaluru-574 199, Karnataka, India*
[3]*Center for Systems Biology and Molecular Medicine, Yenepoya Research Centre, Yenepoya (Deemed to be University), Mangalore-575 018, Karnataka, India*

(Manuscript Received: 16-08-2021, Revised: 17-11-2021, Accepted: 20-11-2021)

Keywords: Coconut, endosperm, *in vitro* culture, transcriptome

Coconut (*Cocos nucifera* L.; 2n=32; Arecaceae), a perennial palm well adapted to tropical conditions, is one of the main sources of oil that has numerous edible industrial and medicinal uses (Deen *et al*., 2021). The coconut endosperm comprises both solid and liquid parts, with oil content in the solid endosperm ranging from 55-65 per cent depending on germplasm (Niral *et al*., 2009). The fats in coconut oil are predominantly saturated, comprising medium-chain triglycerides like lauric acid, caprylic acid and capric acid (Suryani *et al*., 2020).

With the constant increase in the demand for vegetable oils worldwide, the identification of genes that regulate the quantitative and qualitative variations with respect to the composition and quality of coconut oil assumes significance. In contrast to other important palms, *viz*., oil palm and date palm, till recently, the available genomic resources were relatively scarce for coconut. The recent availability of whole-genome sequences of Hainan Tall (Xiao *et al*., 2017), Catigan Green Dwarf (Lantican *et al*., 2019) and Chowghat Green Dwarf (Rajesh *et al*., 2020) cultivars have opened up new avenues for the identification of gene(s) governing important pathways in coconut. Apart from this, transcriptome sequences have been generated from many tissues (reviewed by Rajesh *et al*., 2021). More information is essential to bridge the knowledge gap in genomic and transcriptomic resources in this important palm, which would deliver new perceptions for the detection and characterization of novel and useful gene(s) linked to traits of agronomic importance (Huang *et al*., 2014).

In vitro culture of the immature endosperm of coconut for the production of friable callus has been reported earlier by Kumar *et al*. (1985). Also, the fatty acid profile of the endosperm calli was comparable with the normal immature endosperm (Ceniza *et al*., 1992). However, genes governing fatty acid biosynthesis in *in vitro* endosperm cultures have not been studied.

We report the transcriptome assembly from coconut endosperm callus derived from immature (7 months old) nuts of Laccadive Micro Tall (LMT). This variety possesses high oil content (Devakumar *et al*., 2010). To initiate *in vitro* culture, seven-month-old immature nuts were harvested from Laccadive Micro Tall (LMT) cultivar maintained at the Experimental Farm of ICAR-CPCRI, Kasaragod, Kerala State, India (12.30^{o} N, 75^{o} E; altitude of 10.7m above MSL). The nuts were surface sterilized using alcohol, and after minimal processing, the nuts were taken to a laminar airflow chamber. The husk in the micropylar region was further trimmed and nuts were flame sterilized. A layer of endosperm, along with a thin layer of the shell at the micropylar region, was inoculated in Y3

[*] *Corresponding Authors: rajesh.mk@icar.gov.in & keshav@yenepoya.edu.in*

medium (Eeuwens, 1976) supplemented with sucrose (30 g L^{-1}), 2,4-D (4 mg L^{-1}), NAA (4 mg L^{-1}), BAP (0.5 mg L^{-1}), agar (7 g L^{-1}) and activated charcoal (2 g L^{-1}). The pH was adjusted to 5.75 before autoclaving the medium. Cultures were incubated in the dark at 27 ±2ºC temperature and 80 per cent relative humidity. Sub-culturing was done in monthly intervals to the same basal medium with half the concentrations of growth regulators. After three months, cultures were sub-cultured in callus multiplication media consisting of 2,4-D (0.1 mg L^{-1}) alone. Different steps in sampling and inoculation of endosperm explants *in vitro* are given in Figure 1.

Total RNA was extracted from 150 mg of endosperm callus [90 days after inoculation (DAI)] using the NucleoSpin® RNA Plant Kit (Macherey-Nagel). The isolated RNA samples were pooled together from three biological replicates and used to build an RNA-seq library (TruSeq Stranded mRNA LT Sample Prep Kit, Illumina). After assessing the library quality, sequencing of the libraries was undertaken on an Illumina HiSeq 4000 platform using paired-end sequencing.

The quality of raw reads (.fastq) files was analyzed initially by using FastQC (Andrews, 2010) to search the raw reads for ambiguous bases, Phred score >Q20, read length, nucleotide base content, and other parameters. Trimmomatic (Bolger *et al.*, 2014) was then used to filter poor-quality sequences. The high-quality reads obtained were mapped with

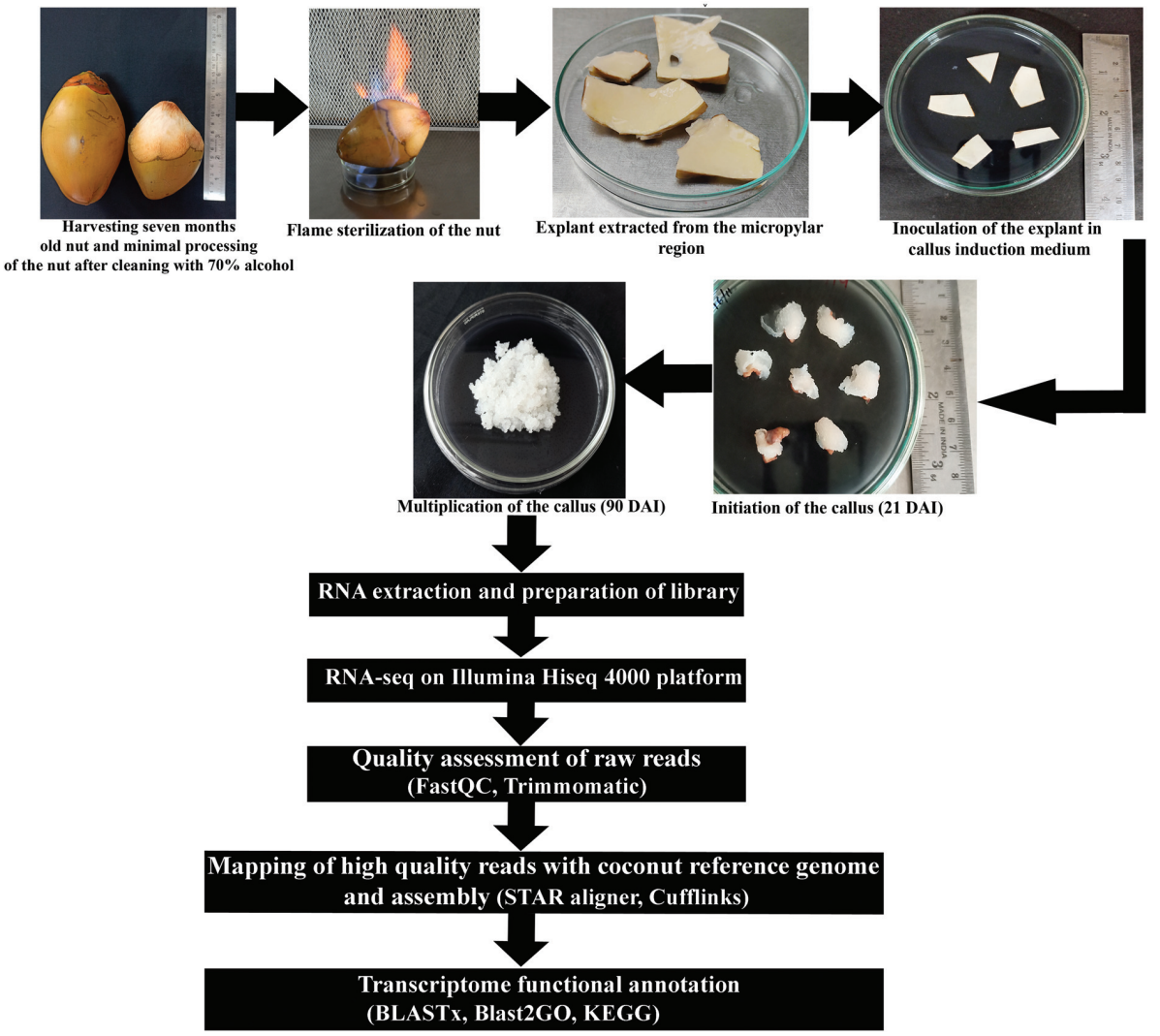

Fig. 1. Different steps in sampling, inoculation of endosperm explants from seven-month-old nuts, callus initiation and multiplication, RNA-sequencing and data analysis

CGD reference genome (PRJNA413280) (Rajesh *et al*., 2020) genome using STAR aligner (Dobin *et al.,* 2013) and assembled with Cufflinks (Trapnell *et al*., 2012). Local protein databases were created from *Arabidopsis thaliana* (Taxonomy ID: 3702), *Triticum aestivum* (Taxonomy ID: 4565), *Zea mays* (Taxonomy ID: 4577), *Oryza sativa* (Taxonomy ID: 4530) and Arecaceae (Taxonomy ID: 4710) families. The transcripts were annotated by BLASTx search against these databases, keeping a 1×10^{-3} maximum E-value threshold. Omics Box (Blast2GO) (Conesa *et al*., 2005) was used to perform GO annotation and KEGG pathway analysis of the assembled transcripts from endosperm callus from coconut.

Using Illumina Hiseq 4000 sequencing platform, 8.789 Gb (PE) raw sequences were generated. The details of the RNA-seq statistics of raw and clean reads are provided in Table 1. A total of 39,151 transcripts could be generated from the endosperm callus of LMT by reference-based assembly with the genome of the Chowghat Green Dwarf cultivar. Mapping statistics and assembly details of high-quality reads with the reference coconut genome are given in Table 2. Figure 2 represents the nucleotide length distribution of transcripts in coconut endosperm callus. Sequence similarity analysis and annotation of these transcripts against the Gene Ontology (GO) database identified 97,001 GO annotations. A total of 87,015,828 raw reads (around 8.789 Gb paired-end data) could be generated by RNA-seq of

Table 1. RNA-seq statistics of raw read and clean reads obtained from endosperm callus of LMT

Raw reads	87,015,828
Total bases (Gb)	8.789
Read length (bp)	101
G + C percentage	50.58
High quality reads (filtered)	78,572,878 (90.29%)

Table 2. Mapping, assembly and annotation statistics

Raw reads	87,015,828
Number of inputs reads for mapping	78,572,878
Uniquely mapped reads number	65,553,054
Uniquely mapped reads (%)	83.43
Total transcripts generated	39,151
With Blast Hits	33,042
With Mapping	27,478
With Annotation	26,232
without IPS	6,829
with IPS	32,322
with GOs	19,253

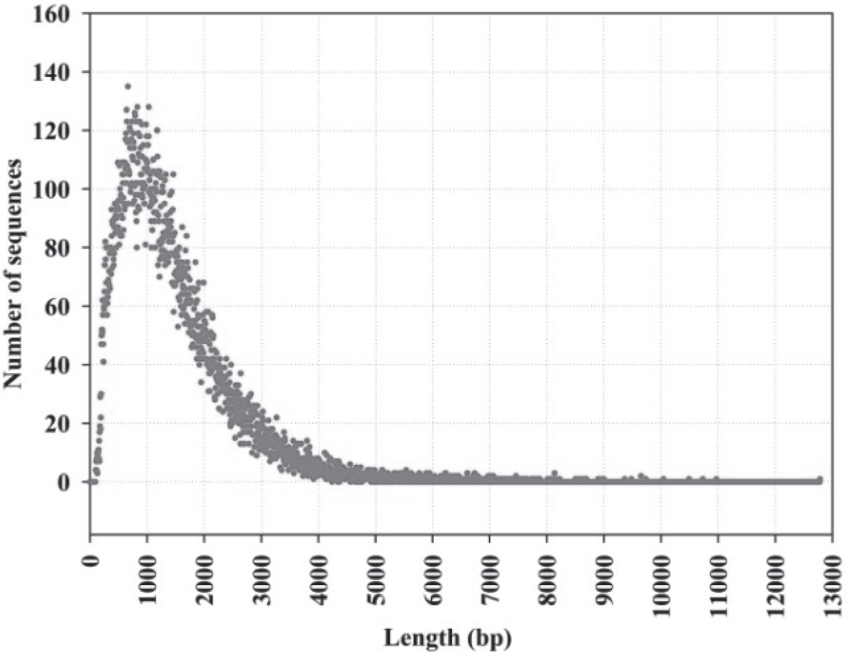

Fig. 2. Nucleotide length distribution of coconut endosperm callus transcripts. The transcript length ranged from 101-12785 bp

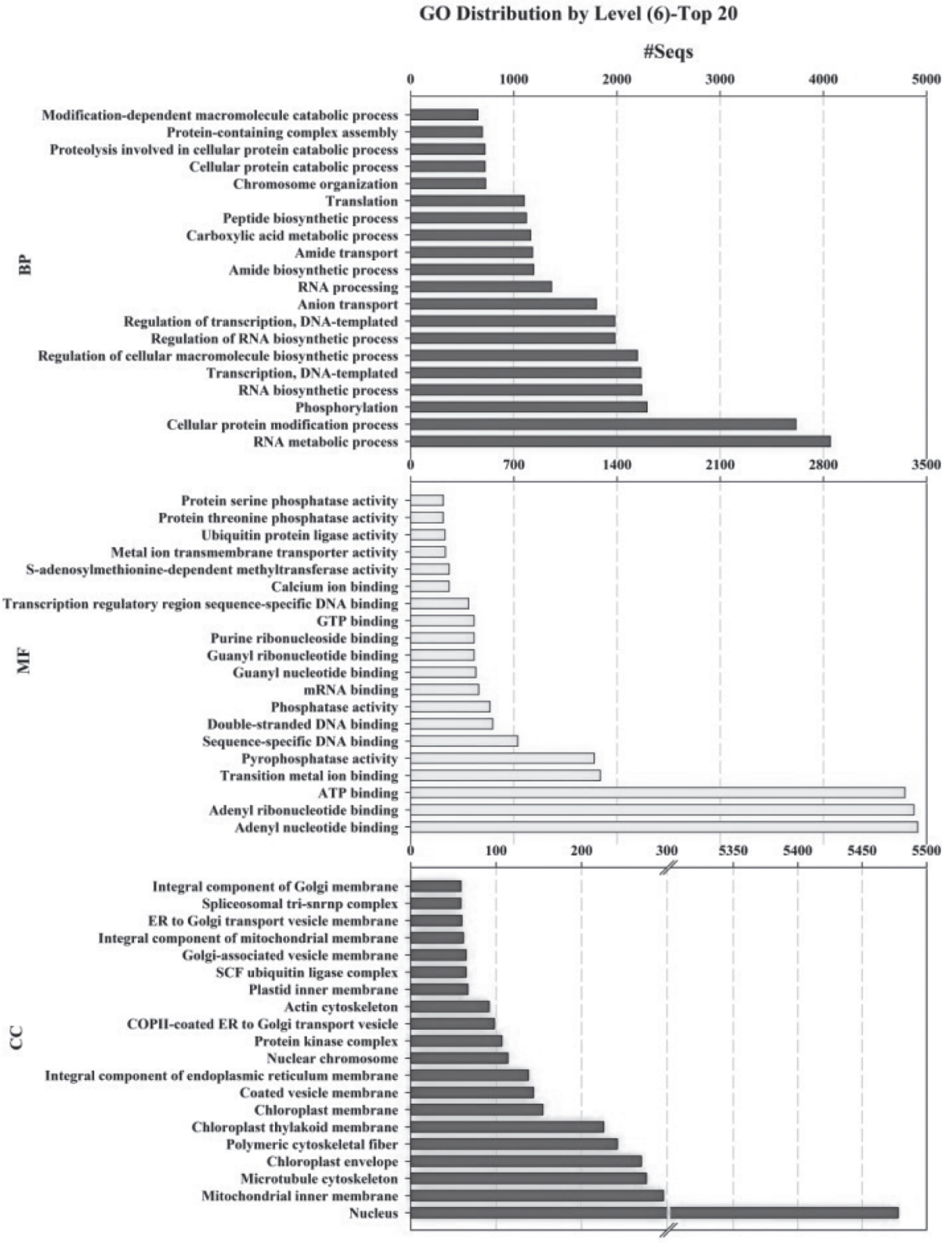

Fig. 3. A graphical representation of Gene Ontology (GO) distribution levels of annotated coconut endosperm callus transcripts- biological process (BP), molecular function (MF), and cellular component (CC)

endosperm callus of LMT cultivar. By mapping the 78,572,878 high-quality reads obtained after filtering to the CGD reference genome, 39,151 transcripts could be generated. Of these, 33,042 coconut transcripts could be annotated via the BLASTx program (84.4%). The functional analysis revealed that 19,253 transcripts were characterized with 97,001 GO annotation. Pathway analysis revealed 3370 transcripts to be active in different metabolic pathways (Table 2).

Functional classification of the transcripts into GO categories *viz*., biological process (BP), molecular function (MF), and cellular component (CC). Among the GO level annotations, the mean GO level was 6,859, with 59,854 annotations available at GO level 6 (biological processes = 33187; molecular functions = 18594 and cellular components =8073) (Fig. 3). A total of 3370 transcripts involved in 276 different KEGG pathways were mapped (Supplementary file S2). The KEGG pathway analysis indicated the presence of starch and sucrose metabolism, fructose and mannose metabolism, glycerolipid metabolism, oxidative phosphorylation, phosphatidylinositol signalling system, inositol phosphate metabolism, glycerophospholipid metabolism, sphingolipid metabolism, fatty acid degradation, galactose metabolism, pantothenate and CoA biosynthesis, glutathione metabolism, glycolysis/gluconeogenesis, fatty acid elongation apart from fatty acid biosynthesis pathways (Fig. 4). Twenty-seven transcripts with putative roles in fatty acid biosynthesis were identified (Fig. 4, Supplementary File S2).

The dataset provides information on the expression pattern of various genes involved in fatty acid biosynthesis in coconut endosperm cultures. It will be a valuable resource for research on oil metabolism in oil-yielding crops. This data resource would also help researchers further understand the metabolic mechanisms that connect functional gene networks with respect to oil biosynthesis. It would also allow the identification and comparison of orthologous genes and their corresponding proteins, with roles in oil biosynthesis, in different oilseed crops. The transcripts could be the basis for generating probes and primers for functional validation of key oil biosynthesis genes and transgenic studies in coconut.

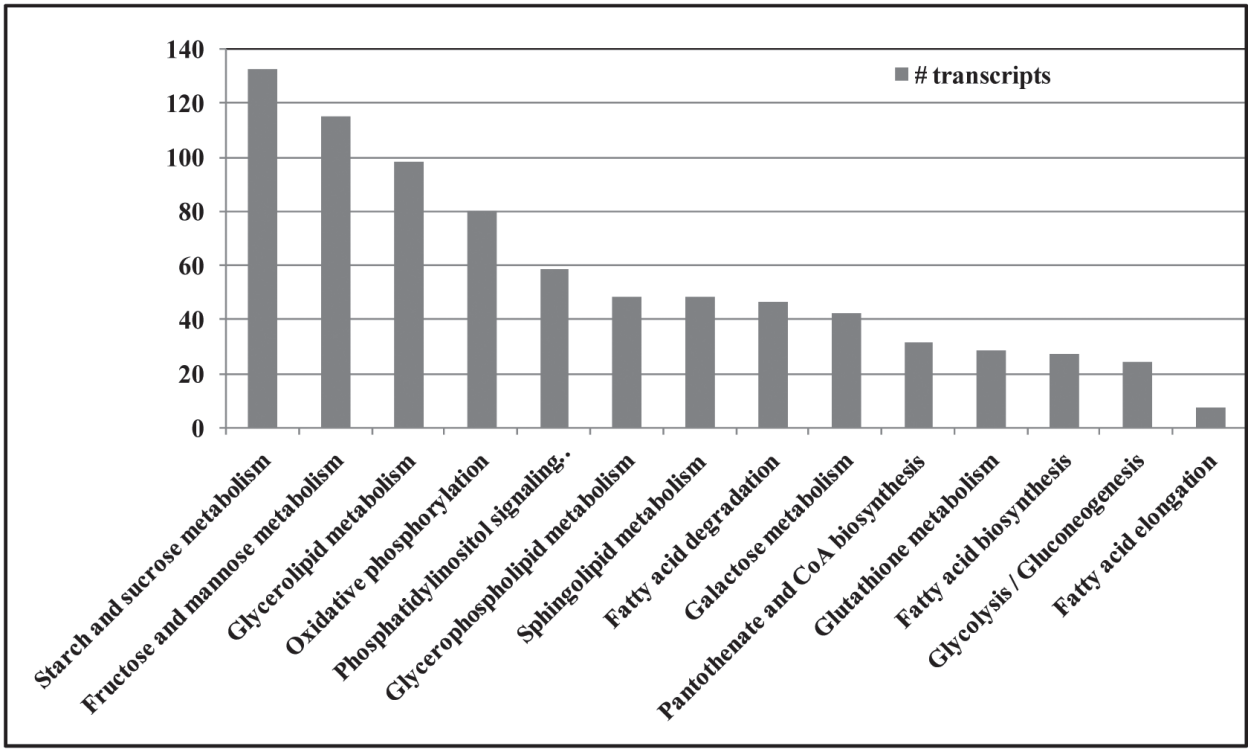

Fig. 4. Selected KEGG pathways in endosperm calli of LMT cultivar of coconut

Supplementary files are available in the online version of the article. Supplementary File S1: An overview of functional annotation and GO of each transcript; Supplementary File S2: Details of the KEGG pathway analysis; Supplementary File S3: Transcripts related to fatty acid biosynthesis pathway.

Acknowledgements

The authors would like to thank the Indian Council of Agricultural Research (ICAR) [ICAR-CPCRI Project No. 1000761031] and Karnataka Biotechnology and Information Technology Services (KBITS), Government of Karnataka [for support to the Center for Systems Biology and Molecular Medicine at Yenepoya (Deemed to be University), Mangalore, under the Biotechnology Skill Enhancement Programme in Multiomics Technology (BiSEP GO ITD 02MDA2017)]. The first author is a recipient of the Junior Research Fellowship (JRF) from the Council of Scientific and Industrial Research-University Grants Commission (CSIR-UGC), Government of India, New Delhi, India. Authors are thankful to Dr. Niral V, Principal Scientist, ICAR-CPCRI for providing the resources.

References

Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Bolger, A.M., Lohse, M. and Usadel, M. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114-2120.

Ceniza, M.S., Ueda, S. and Sugimura, Y. 1992. *In vitro* culture of coconut endosperm: callus induction and its fatty acids. *Plant Cell Reports* **11**(11):546-549.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M. and Robles, M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**:3674–3676.

Deen, A., Visvanathan, R., Wickramarachchi, D., Marikkar, N., Nammi, S., Jayawardana, B.C. and Liyanage, R. 2021. Chemical composition and health benefits of coconut oil: An overview. *Journal of the Science of Food and Agriculture* **101**:2182-2193.

Devakumar, K., Niral, V., Jerard, B.A., Jayabose, C., Chandramohanan, R. and Jacob, P.M., 2010. Microsatellite analysis of distinct coconut accessions from Agatti and Kavaratti Islands, Lakshadweep, India. *Scientia Horticulturae* **125**:309-315.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15-21.

Eeuwens, C.J. 1976. Mineral requirements for growth and callus initiation of tissue explants excised from mature coconut (*Cocos nucifera*) and date (*Phoenix dactylifera*) palms cultured *in vitro*. *Physiologia Plantarum* **36**:23-28.

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**(11):2147-2157.

Kumar, P.P., Raju, C.R., Chandramohan, M. and Iyer, R.D. 1985. Induction and maintenance of friable callus from the cellular endosperm of *Cocos nucifera* L. *Plant Science* **40**:203-207.

Lantican, D.V., Strickler, S.R., Canama, A.O., Gardoce, R.R., Mueller, L.A. and Galvez, H.F. 2019. *De novo* genome sequence assembly of dwarf coconut (Cocos nucifera L.'Catigan Green Dwarf') provides insights into genomic variation between coconut types and related palm species. *G3: Genes, Genomes, Genetics* **9**(8):2377-2393.

Niral, V., Nair, R.V., Jerard, B.A., Samsudeen, K. and Ratnambal, M.J. 2009. Evaluation of coconut germplasm for fruit component traits and oil yield. *Journal of Oilseeds Research* **26**(Special Issue):668-670.

Rajesh, M.K., Ramesh, S.V., Karun, A. and Chowdappa, P. 2021. Genome Sequencing, Transcriptomics, Proteomics and Metabolomics. In: The Coconut Genome. Compendium of Plant Genomes (Eds.) Rajesh, M.K., Ramesh, S.V., Perera, L. and Kole, C. Springer, Cham. pp.119-132. https://doi.org/10.1007/978-3-030-76649-8_8.

Rajesh, M.K., Chowdappa, P., Behera, S.K., Kasaragod, S., Gangaraj, K.P., Kotimoole, C.N., Nekrakalaya, B., Mohanty, V., Sampgod, R.B., Banerjee, G., Das, A.J, Niral, V., Karun, A., Mahato, A.K., Gaikwad, K., Singh, N.K. and Prasad, T.S.K. 2020. Assembly and annotation of the nuclear and organellar genomes of a dwarf coconut (Chowghat Green Dwarf) possessing enhanced disease resistance. *OMICS: A Journal of Integrative Biology* **24**:726-742.

Suryani, S., Sariani, S., Earnestly, F., Marganof, M., Rahmawati, R., Sevindrajuta, S., Mahlia, T.M.I. and Fudholi, A. 2020. A comparative study of virgin coconut oil, coconut oil and palm oil in terms of their active ingredients. *Processes* **8:**402. (https://doi-org/10.3390/pr8040402).

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L. and Pachter, L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocol* **7**:562-578.

Xiao, Y., Xu, P., Fan, H., Baudouin, L., Xia, W., Bocs, S., Xu, J., Li, Q., Guo, A., Zhou, L. and Li, J. 2017. The genome draft of coconut (*Cocos nucifera*). *Gigascience* **6**(11):doi: 10-1093/gigascience/gix095.