

SCIENTIFIC REPORTS



OPEN

A gene co-expression network model identifies yield-related vicinity networks in *Jatropha curcas* shoot system

Nisha Govender^{1,2}, Siju Senan¹, Zeti-Azura Mohamed-Hussein^{2,3} & Ratnam Wickneswari¹

The plant shoot system consists of reproductive organs such as inflorescences, buds and fruits, and the vegetative leaves and stems. In this study, the reproductive part of the *Jatropha curcas* shoot system, which includes the aerial shoots, shoots bearing the inflorescence and inflorescence were investigated in regard to gene-to-gene interactions underpinning yield-related biological processes. An RNA-seq based sequencing of shoot tissues performed on an Illumina HiSeq. 2500 platform generated 18 transcriptomes. Using the reference genome-based mapping approach, a total of 64 361 genes was identified in all samples and the data was annotated against the non-redundant database by the BLAST2GO Pro. Suite. After removing the outlier genes and samples, a total of 12 734 genes across 17 samples were subjected to gene co-expression network construction using petal, an R library. A gene co-expression network model built with scale-free and small-world properties extracted four vicinity networks (VNs) with putative involvement in yield-related biological processes as follow; heat stress tolerance, floral and shoot meristem differentiation, biosynthesis of chlorophyll molecules and laticifers, cell wall metabolism and epigenetic regulations. Our VNs revealed putative key players that could be adapted in breeding strategies for *J. curcas* shoot system improvements.

Jatropha curcas L. or the physic nut is an environmentally friendly and cost-effective feedstock for sustainable biofuel production. The tree-like shrub has a productive life span of 50 years and endows oil-rich (37–50%) seeds^{1–3}. The *Jatropha* crude oil (JCO) physiochemical characteristics are highly desired for biodiesel production; long chain fatty acids, low acidity, low viscosity, good stability and a relatively high cetane number. In addition, JCO-based biodiesel is able to perform better than the conventional petro-diesel in terms of engine maintenance and carbon dioxide emission^{4,5}. Edible vegetable oils are primary candidates for biofuel feedstock. Developed nations such as the US, UK and Canada are dependent on soybean, canola, sunflower, rapeseed and corn oil for biodiesel feedstock while others such as Malaysia and Indonesia have deployed the palm oil. The usage of edible vegetable oil for biofuel is feasible provided a surplus supply which exceeds the consumption demand is met^{6–8}. Under this scenario, the JCO also a toxic non-edible oil with lucrative properties for biofuel production, could potentially resolve the present food versus fuel pressure faced by the edible vegetable oils. At present, the commercialization of *Jatropha*-based biofuel is hampered by a lack in planting materials with consistent flowering, fruiting and seed yields throughout the planting cycles.

In higher plants, the post-embryonic developments of vegetative and reproductive tissues are inter-dependent^{9,10}. Yield component is affected by reproductive structures such as seed, flower and axillary meristem and the reproductive phase is only reached after a substantial period in vegetative phase. Nevertheless, both the vegetative and reproductive phase remain confounded to one another for the reproductive success in plants. In agronomic evaluation of *J. curcas*, numerous studies have shown strong association between *J. curcas* yield and reproductive structures such as flower number, pollen fertility, male to female flower ratio and seed number^{1,11,12}. In addition, breeding and management programs had utilized extensive number of genomics and molecular

¹School of Environmental and Natural Resource Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM, Bangi, Selangor, Malaysia. ²Center for Bioinformatics Research, Institute of Systems Biology (INBIOSIS), Universiti Kebangsaan Malaysia, 43600 UKM, Bangi, Selangor, Malaysia. ³School of Biosciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM, Bangi, Selangor, Malaysia. Correspondence and requests for materials should be addressed to N.G. (email: nishag@ukm.edu.my)

approaches for the examination of *J. curcas* reproductive structures^{13–20}. In a phenotypic evaluation of *J. curcas* accessions from South-East Asia, the yield-associated traits showed positive correlation to number of branches²¹. Likewise, in another study conducted in India, the seed yield of 2-year-old *J. curcas* showed significant correlation to number of secondary and tertiary branches per plant¹². Despite interesting agronomic findings which had indicated strong association between the shoot system and yield, the molecular studies of *J. curcas* shoot system is under-emphasized for yield enhancement breeding strategies.

The aerial shoot and its small population of mitotically dividing cells at the center, called the shoot apical meristem (SAM) are critical tissues required to start the vegetative and subsequent reproductive development in plants. The SAM maintains a pool of undifferentiated cells in the center while generating above-ground plant organs: the stems, leaves and flowers. The reiterative formation of lateral organs and basal regions from the SAM requires communication among cells within SAM and between the SAM and incipient organs^{22–24}. The indeterminate shoot growth (vegetative phase) is terminated by the determinate growth of flowers (reproductive phase) when SAM switches from inflorescence meristem into floral meristem, and the entire developmental phase changes are orchestrated by a complex gene network²⁴. Gene expression patterns obtained from the aerial shoot, together with shoot bearing inflorescence (basal region) and inflorescence provide essential information to elucidate gene associations among the reproductive-related shoot tissues. Therefore, a gene co-expression network which predicts potential functional relationship between genes and subsequently predicts a gene's function^{25,26} is employed to investigate the gene-to-gene relationship in *J. curcas* reproductive-related shoot system.

Biological networks are widely used to characterize the interactions among genes in order to understand the biological processes and functions underpinning traits of interest. Genes with similar expression patterns are likely to have similar functions, which may have been regulated by the same mechanism or a joint transcription factor^{27,28}. Expression data generated from transcriptomes are primarily used to build networks. High-throughput, genome sequencing technology offers an array of techniques for the measurement of gene expression specific to a given experimental condition, tissue or developmental stage. Gene co-expression is the expression profiles of a gene across the different conditions and that they are compared against other genes to acquire similar relationships between the genes. In a network, nodes represent genes and the interconnecting edges between the nodes reflect the degree of correlated expressions. A subset of nodes that are tightly connected to each other is a module. Within a module, the highly connected genes, also referred as 'hub genes' are likely to have important biological function^{25,26,29}. A gene co-expression network is widely used to infer the 'true' biological process which includes interactions and other combined activities at the cellular system.

In a previous paper³⁰, we identified differentially expressed (DE) genes in *J. curcas* inflorescence relative to shoot tissues and categorized them according to their ontologies term. A differential gene expression analysis scores for statistically significant DE genes between a pair of treatments in accordance to the absolute count value. Valuable information on relationship among the DEs (Guilt-by-Association) remains disentangled, despite their significant importance in gene function prediction and regulatory mechanisms. Thus, we constructed a gene co-expression network using petal, a publically available R library. The RNA-seq based transcriptomes generated from *J. curcas* shoot system were fed into petal to obtain a biologically meaningful gene co-expression network model which follows the scale-free and small-world properties. A list of DE genes³⁰ with yield-related terms (biological process) were selected to construct vicinity networks (VNs). Our VNs containing groups of functionally annotated genes underlying yield-related biological processes in *J. curcas* such as flowering, epigenetics, stress tolerance and the biosynthesis of chlorophyll molecules and laticifers are putatively characterized based on literature.

Material and Methods

Isolation of rna, assembly construction and transcriptome analysis. For RNA-sequencing library preparation, mature shoots subtending an inflorescence at 2.5 cm from the base of peduncle, aerial shoots at 2.5 cm from the tip of the apex and whole inflorescences were collected from six *Jatropha* plants. Total RNA was isolated using a combined modified method; CTAB + silica column³¹ and RNeasy Plant mini kit (Qiagen, Hilden, Germany) and the concentration was measured using a Nanodrop ND-10000 spectrophotometer (NanoDrop Technologies, USA). The quality and integrity of total RNA were determined at 260/280 nm ratio by NanoBioanalyzerRNA-6000 analysis (Agilent-Technologies). The cDNA library was constructed according to the SureSelect Strand-Specific RNA Library Prep for NGS Workflow protocol (Agilent Technologies, USA) and the corresponding fragmented genomic DNA was sequenced using Illumina HiSeq. 2500 (Yourgene Bioscience, Taiwan). The output, raw fastq reads were trimmed using the Trim Galore package and aligned to *J. curcas* genomic sequences obtained from Kazusa's *Jatropha* Genome Database version r4.5 (<http://ftp.kazusa.or.jp/pub/jatropha/>). Alignment was performed using the STAR aligner (<https://github.com/alexdobin/STAR>) and Cufflinks were used to produce a merged assembly. Assembly construction was performed using the CLC Genomics Workbench (CLC bio, USA). For a descriptive characterization, the *J. curcas* transcripts were annotated using BLAST2GO PRO 4.0 suite^{32,33}, BLAST, mapping, InterPro Scan and annotation.

Data: Quality control and detection for outliers. The transcripts were normalized into counts per million (cpm) and lowly expressed genes were filtered out as follows: retain genes that are expressed at cpm > 5 in all samples. The shoot and inflorescence treatments were clustered to detect presence/absence of outliers. A cluster of dendrogram for each group of samples was generated using the Jaccard distance (vegan package). To confirm dissimilarities between the treatment groups, a multidimensional (MDS) plot was constructed using limma and edgeR packages.

Gene co-expression network algorithm and parameters. The gene co-expression network was constructed according to petal, a novel co-expression modeling system²⁵. This system is designed for construction

| Attributes | Values | |
|-------------------------------|-----------------------|--------------|
| Read length | 2 × 100 bp | |
| Sequence analyzed (pair) | 22 961 194–35 876 223 | |
| Filtered | 3.49–4.75% | |
| Aligner reads (pair) input | 22 053 488–34 343 944 | |
| Uniquely mapped reads | 81.02–88.22% | |
| Reads mapped to multiple loci | 8.15–11.25% | |
| Shoot Biosample ID | SAMN05827448 | SAMN05827459 |
| | SAMN05827450 | SAMN05827460 |
| | SAMN05827452 | SAMN05827462 |
| | SAMN05827454 | SAMN05827463 |
| | SAMN05827455 | SAMN05827464 |
| | SAMN05827456 | |
| Inflorescence Biosample ID | SAMN05827449 | SAMN05827457 |
| | SAMN05827451 | SAMN05827461 |
| | SAMN05827453 | SAMN05827465 |
| Reads mapped to gene model | 16 886 160–26 833 093 | |
| With BLASTX hits | 21 188 | |
| With mapping | 5163 | |
| With GO annotation | 32 299 | |
| With IPS | 29 923 | |
| IPS with GOs | 16 615 | |
| Non-coding RNAs (RFAM) | 82 | |
| KEGG pathways | 102 | |

Table 1. The *Jatropha curcas* shoot system transcriptome data.

of biologically meaningful networks with a small-world and scale free structure. Under this approach, Spearman (SP) was selected to generate an association matrix. The SP applies for transcriptome count data with a non-normal distribution (Supp. 1). The association matrix was converted to adjacency matrices using a series of thresholds generated by petal's default function (Add. 2) and the "best threshold" was selected after a customized refining-threshold procedure. Criteria for the network construction were as follows: Coefficient of determination (R^2) >0.9, average cluster coefficient (meanCC) <0.5, meanPath <5, average path length: percentage of original dataset utilized in the network model (%used) >95, and percentage of vertices that are in the biggest component of the network model (%bigComp) >95. From the gene co-expression network (network model), we extracted vicinity networks (VNs) and expression profiles of samples using nine yield-related genes, previously described as differentially expressed genes in inflorescence³⁰. The gene co-expression network and VNs were drawn, visualized and annotated using the Cytoscape 3.0 software. Topological measures were generated using Network analyzer³⁴.

Pathway analysis. Each vicinity network (VN) was subjected to pathway enrichment analysis using the KaPPA-View4 *Jatropha* (<http://kpv.kazusa.or.jp/kpv4>) metabolic pathway database³⁵. All genes in each VN were categorized according to KEGG pathways as follow: metabolism, genetic information processing, environmental information processing and cellular process. Correlations (>0.9) of gene to gene co-expression are indicated as red curves and each gene is represented as square box on the pathway maps.

Data availability. The transcriptome sequence data utilized in this study is found in National Center for Biotechnology (NCBI)'s Sequence Read Archive (SRA) database with the accession number, SRP090662 (<https://www.ncbi.nlm.nih.gov/sra/?term=SRP090662>).

Results and Discussion

Transcriptome analysis. The number of clean and mapped reads obtained for 18 RNA libraries sequenced, the descriptive statistics of the transcriptomes and the sample descriptions are presented in Table 1. The number and percentage of reads mapped onto genomic loci and multiple loci ranged from 81.02–88.22 and 8.15–11.25, respectively. After removal of low-quality reads, the total number of reads in the 18 samples ranged from 22 961 194–35 876 223. Reads were mapped onto *J. curcas* reference genome and a total of 64 361 transcript sequences were obtained from the 18 transcriptomes. The average assembly size for each transcriptome ranged from 16 886 160 to 26 833 093 transcripts. From the 21 188 sequences with BLASTX hits, only 5163 and 32 299 sequences acquired mapping and annotation respectively. The InterPro Scan (IPS) revealed 34 438 sequences without an IPS match, while 29 923 and 16 615 sequences scored match with IPS and with Gene Ontology (GO), respectively. The transcripts with GO terms showed 673 ECs that are associated with 102 KEGG pathways. Sequence search against the Rfam database indicated presence of 82 non-coding RNAs (Table 1).

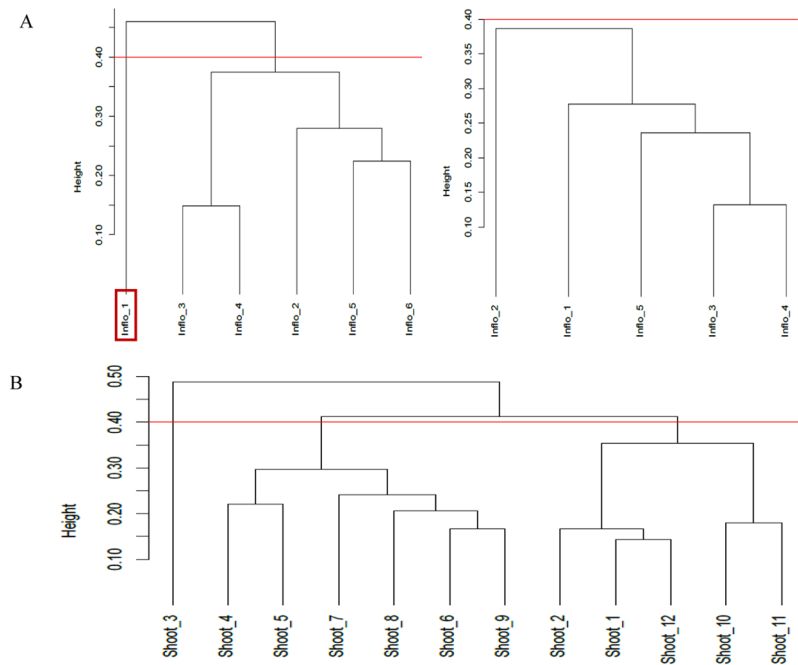


Figure 1. Jaccard-based similarity dendrogram applied to *Jatropha curcas* inflorescence (A) and shoot (B) transcriptome data. Read line indicates cut-off height at 0.40. (A) Cluster of dendrogram obtained before (left) and after (right) sample editing; the original data set of six samples (left) indicates presence of an outlier (boxed).

Detection for outliers within and between the treatments. The Jaccard similarity dendrogram of six inflorescence samples indicated presence of an outlier at height > 0.4 . The outlier was removed from the present study and the remaining 5 replicates were re-clustered. Biological replicates of each shoot and inflorescence treatment showed absence of outliers at cut-off height = 0.4 (Fig. 1). A multidimensional-scaling (MDS) plot showed apparent dissimilarity between the shoot and inflorescence treatments. Twelve shoot replicates (circles) found on the right panel were distinguished from five inflorescence replicates (crosses) scattered on the left region of dimension 1 at leading $\log_{FC} = 0$ (Fig. 2).

Gene co-expression network construction and analysis. After filtration, a total of 12 734 genes across 17 samples were applied in the construction of gene co-expression network. Since the transcriptome data utilized in this study showed a non-normal distribution (Supp. 1), the Spearman Correlation Coefficient (SP) was used to establish association between pairs of genes among the samples. Petal's default function produced a series of thresholds with corresponding mathematical models. We selected the range between the "best" and "alternative" thresholds (0.845–0.862) to perform a customized refining procedure. A threshold of 0.855 owed the best description for a scale-free architecture and small-world model and was therefore used in all subsequent analyses. The topological properties of the network at a threshold of 0.855 were as follows: Coefficient of determination (R^2): 0.81, average cluster of co-efficient (meanCC): 0.4146, meanPath: 4.1603, percentage of original dataset utilized in the network model (%used): 96.5%, percentage of vertices that are in the biggest component of the network model (%bigComp): 99.4 (Table 2). The gene co-expression network from 718599 measures contains 12 290 nodes with an average number of neighbors of 117. (Supp. 2). A total of nine vicinity networks (VNs) were obtained from a list of ten gene identifiers. Using a list of gene identifiers², a total of nine VNs were obtained. Number of neighbors and density score for each VN ranged from 10–617 and 0.4–0.63, respectively. Only four VNs (described as VN 1–4) with a density greater than 0.5 were subjected to further characterizations (BLAST description and node's degree). The node size drawn in each VN indicates their corresponding degree; large node size applies a large degree (Fig. 3).

Putative involvement of VN1 in epigenetic event, signal transduction and cell wall metabolism. VN1 constructed based on Jcr4S04537.10 (yellow node), described as *DNA polymerase delta catalytic subunit* is depicted in Fig. 3A. It contains 36 nodes connected by 616 edges and the clustering coefficient is 0.766. The average number of neighbors is 16.8 with a density of 0.51. The sequence IDs and the corresponding descriptions of VN1 genes are given in Supp. Dataset 1. VN1 identified 4 nodes that could be implicated to epigenetic events: *DNA polymerase alpha catalytic subunit*, *DNA polymerase delta catalytic subunit*, *DNA mismatch repair MSH6* and *ABC transporter G family member 3*. DNA polymerases are responsible for the replication of genetic materials, maintenance of genome stability and transmission of genetic information from one generation to the other. Integrity of a genome is constantly affected by endogenous and environmental damaging agents, and thus, a number of DNA repair pathways are constantly recruited for protection. The B family DNA polymerase delta

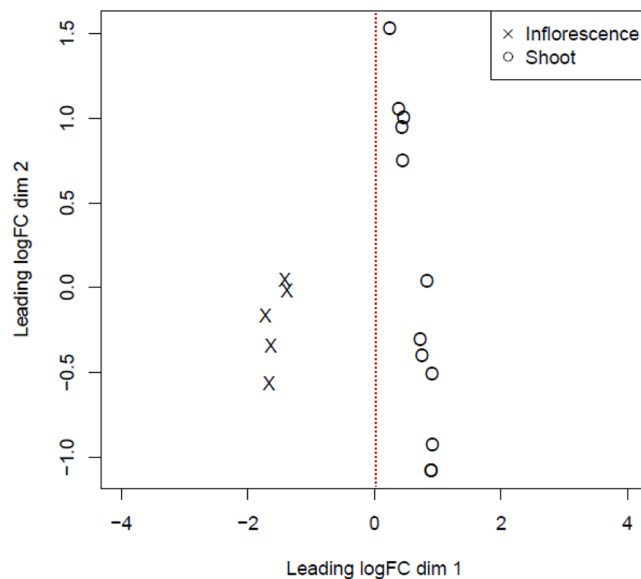


Figure 2. A multidimensional scaling (MDS) plot shows the dissimilarities between the *Jatropha curcas* shoot and inflorescence samples at dimension (dim) 1 with leading logFC = 0. Twelve circles (shoot) scattered at the right panel separates the five crosses (inflorescence) clustered on the left panel from the logFC dim 1 = 0 (dotted red line).

| Threshold | R ² | slope/power | meanCC | meanPath | %used | %bigComp |
|-----------|----------------|-------------|--------|----------|---------|----------|
| 0.862 | 0.82 | -1.3305 | 0.4069 | 4.3394 | 95.7515 | 99.3275 |
| 0.86 | 0.82 | -1.2965 | 0.4099 | 4.2793 | 95.9714 | 99.4027 |
| 0.858 | 0.81 | -1.2914 | 0.4123 | 4.2296 | 96.1992 | 99.4367 |
| 0.855 | 0.81 | -1.2877 | 0.4146 | 4.1603 | 96.5133 | 99.406 |
| 0.85 | 0.80 | -1.2481 | 0.4184 | 4.0528 | 96.9609 | 99.4898 |
| 0.845 | 0.79 | -1.2125 | 0.422 | 3.9596 | 97.4242 | 99.6292 |

Table 2. Network refined-threshold table. R²: Coefficient of determination, meanCC: average cluster coefficient, meanPath: average. path length, %used: percentage of original dataset utilized in the network model, %bigComp: percentage of vertices that are in the biggest component of the network model.

is composed of at least two subunits needed in DNA replication and repair; replisome system, excision repair and recombination repair. The MSH6 gene product forms a complex to recognize base pair mismatches, small insertion or deletions in DNA and subsequently initiate the DNA repair. The mismatch repair is an ATP-driven process³⁶. Concomitantly, the ATP consumption in DNA replication and repair was parallel with the presence of *ABC transporter G family member 3* node, which hydrolyzes ATP for transportation of substrate required for DNA repair and RNA translocation^{37,38}.

In VN1, several nodes from the repeat protein gene families were identified as follow: *armadillo repeat-containing kinesin 1 isoform X1* node, *MORC family CW-type zinc finger isoform 1* node, and the *TBC1 domain family member 8B* node. Repeat proteins with an ability to bind to peptides and ligands are an integrated component of the protein complexes. Thus, the high binding range of these repeat sequences may facilitate interactions between proteins whilst playing a regulatory role in essential pathways such as plant growth, development and stress tolerance, as previously described in literature³⁹. The presence of regulatory nodes in VN1 may also implicate occurrence of the signal transduction, cell division, and floral development processes. The calcium transporting ATPase plasma membrane-type node imparts possible regulation of calcium ions in signal transduction pathway. In Arabidopsis, the stomatal cytokinesis-defective 1-1 (*scd1-1*), encodes a protein which regulates intracellular protein transport and/or mitogen-activated protein kinase signaling and is also involved in flower morphogenesis⁴⁰. Therefore, the homologue present in VN1 may suggest probable occurrence of floral morphogenesis.

Plant cell wall is continually subjected to modification especially during growth and differentiation. In VN1, two nodes namely *Alpha-L-fucosidase 2-like* and the *Alpha-galactosidase-like isoform X1* may have involved in cell wall metabolism. Alpha-L-fucosidase 2-like releases the t-fucosyl residue from xyloglucan oligosaccharides side chain and is a contributing member of the xyloglucan metabolism in Arabidopsis⁴¹. The xyloglucans are cross-linkers of cellulose microfibrils in plant cell wall and thus, impacts the rate of plant cell expansion tremendously. The *galactosidase-like isoform X1* node from the glycosyl hydrolases family are involved in raffinose and galactolipid metabolisms. Cell wall loosening mediates plant cell growth for further expansion^{40-42,45}. The

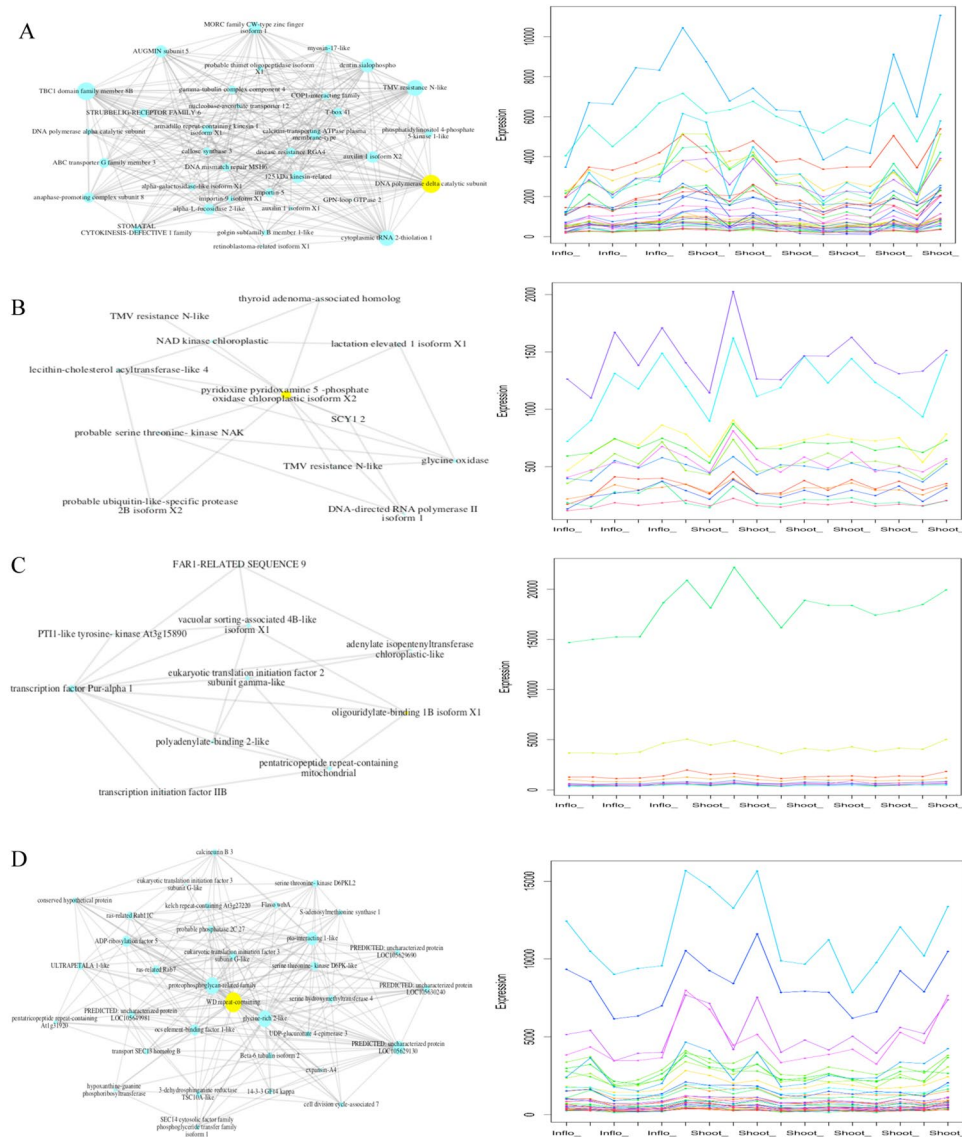


Figure 3. Vicinity networks (VNs) (left) and their corresponding expression profiles (right) extracted based on a gene identifier (yellow node) from *J. curcas* shoot-inflorescence network model. Right: Nodes (blue circles) represent each individual gene. Yellow node in each VCN their corresponding gene identifier: VN1; Jcr4S04537.10, VN2; Jcr4S01867.10, VN3; Jcr4S03942.20 and VN4; Jcr4S03725.10. Edge line (grey) indicates interaction between the nodes. Left: Each coloured line represent a gene and the corresponding expression measures are shown in inflorescence (Inflo_) and shoot (Shoot_) samples (A–D represents VN1–VN4).

presence *Gamma-tubulin complex component 4* node may suggest the re-organization of cytoskeletal microtubules as described in *Arabidopsis*. In addition, the presence of motor proteins, the *125 kDa kinesin related* and *Myosin 17-like* nodes may have either directly or indirectly affected the movement of the related proteins involved in cell wall metabolism⁴⁶ (Fig. 4A).

Putative involvement of VN2 in the biosynthesis of chlorophyll molecules and laticifers. VN2, constructed based on Jcr4S01867.10 (yellow node) described as *pyridoxine pyridoxamine 5-phosphate oxidase chloroplastic isoform X2* is depicted in Fig. 3B. It contains 12 nodes connected by 11 edges and the clustering coefficient is 0.632. The average number of neighbours is 3.5 with a density of 0.32. The sequence IDs and the corresponding descriptions of VN2 are given in Supp. Dataset 2. In higher plants, the salvage pathway synthesizes nucleotides (purine and pyrimidine) by making use of the intermediates produced by the nucleotide degradative pathway. Bases and nucleotides that were formed during degradation of RNA and DNA are recovered. Pyridoxine (pyridoxamine) of the vitamin B6 group is a potent cofactor for a wide range of biochemical reactions in plants; regulatory molecules in signal transduction and membrane ion transporters, potential antioxidant, facilitates post embryonic root development and confers protection against oxidative stress⁴⁷. The gene identifier (yellow node) and its associated gene members in VN2 may have regulated the formation of laticifer cells

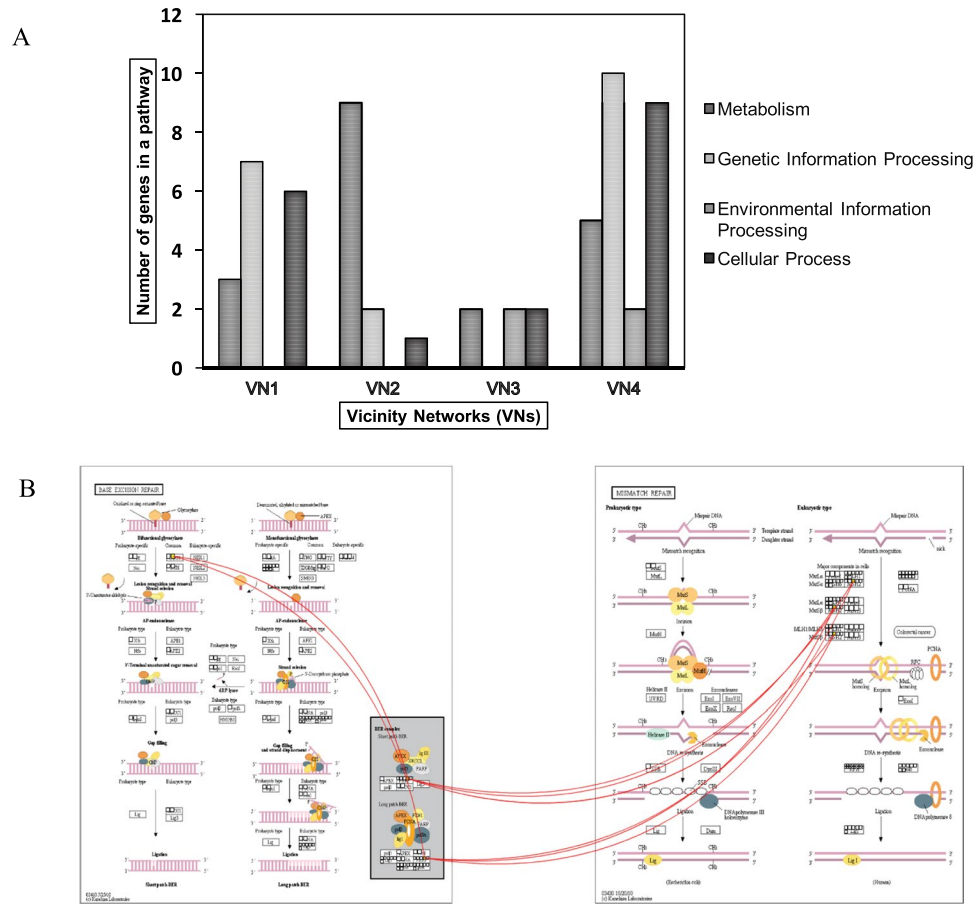


Figure 4. Pathway enrichment analysis: (A) Classification of genes in vicinity networks (VNs) into four major maps; metabolism, genetic information processing, environmental information processing and cellular process. (B) Correlations (>0.9) between genes in the base excision repair and mismatch repair maps in VN1 are indicated with red curved lines.

and chlorophyll molecules in *Jatropha* shoot system. Among the VN2 neighbours, the *NAD kinase chloroplast* node may suggest an essential role in chlorophyll synthesis and chloroplast protection against stress. In chloroplast, the NAD kinase catalyzes NAD and ATP into NADP through phosphorylation in presence of sunlight and thus, a number of NADP-dependent biosynthetic pathways are likely to be regulated. For instance, NADPH may have been consumed in reducing reactions in latex biosynthesis. The regulation of NAD kinase activity indicates an increase in metabolic demand which may have occurred due to developmental changes and environmental signals. The presence of *TMV resistance N-like* and *thyroid adenoma-associated homolog* nodes may infer the occurrence of resistance against stressors, whereas the *lactation elevated 1 isoform X1* may explain possibilities for the biosynthesis of laticifers, a group of latex containing cells found ubiquitously permeating the *J. curcas* plant body⁴⁸. The presence of *probable serine threonine-kinase NAK* node further explains probable occurrence of signal transduction. The general role of protein kinases is to perceive environmental stimuli from sensor and receptor proteins, process and finally establish the corresponding responses. In line with signal transduction, the *probable ubiquitin-like-specific-protease 2B isoform X2* may have played a role in the developmental process such as activation of stress response and organelle biogenesis. The ubiquitin-specific protease UBP14 affected early embryo development in *Arabidopsis*⁴⁹ (Fig. 4B).

Putative involvement of VN3 in heat stress tolerance. VN3, constructed based on Jcr4S03942.20 (yellow node) is depicted in Fig. 3C. It contains 10 nodes connected by 9 edges and the clustering coefficient is 0.67. The average number of neighbors is 4.2 with a density of 0.47. The sequence IDs and the corresponding descriptions of VN3 are given in Supp. Dataset 3. *Jatropha curcas* is a drought and stress tolerant plant. In VN3, the *Oligouridylate-binding 1B isoform X1* node, also a protein component of the stress granules in plants^{50,51} was fixed as the gene identifier. Under heat stress, the mRNA content degrades, however, plant establishes protection via a post-transcriptional process³⁷. The mRNA protective mechanism could have incurred in VN3 as evident by the following nodes: *Transcription factor Pur-alpha 1*, *Transcription initiation factor IIB* and *Eukaryotic translation initiation factor 2 subunit gamma-like* and *FAR-RELATED SEQUENCE 9*. In *Arabidopsis*, FAR1 proteins activates the transcription of light unresponsive genes⁵². The *Polyadenylate-binding 2-like* node may suggest probable stabilization of the nearing degradation mRNAs found within the cell. The primary factor for mRNA degradation is

the RNases. They are characterized as relatively non-specific general RNA binding-proteins. On the other hand, the vacuolar RNases make up the bulk of the total RNase activity in plant cells and have been associated to mRNA decay⁵³. In VN3, the *Vacuolar-sorting associated 4B-like isoform X1* node may imply occurrence of mRNA degradation while other corresponding neighbors may have established the protection response. The *Pt11-LIKE TYROSINE KINASE Atg15890* node may implicate signal transduction in favor of heat stress. The *Adenylate isopentenyltransferase (IPT) chloroplastic like* node is involved in the cytokinin biosynthetic pathway, where, the isoprenyl group is transferred to form either an adenylate-IPT or tRNA-IPT⁵⁴ (Fig. 3C).

Putative involvement of VN4 in flowering mechanism. VN4 constructed based on Jcr4S03725.10, the *WD repeat-containing* node (yellow circle) is depicted in Fig. 3D. It contains 34 nodes connected by 33 edges and the clustering coefficient is 0.69. The average number of neighbors is 13.4 with a density of 0.41. The sequence IDs and the corresponding descriptions of VN4 are given in Supp. Dataset 4. The WD repeat-containing complexes are involved in signal transduction, transcriptional regulation⁵⁵ and other plant processes such as meristem organization, cell division and floral development. This module is made of conserved amino acids (44–60) repeat sequence that ends with tryptophan and aspartate (WD) residues³⁴. Generally, it serves as rigid scaffolds for protein-protein and protein-DNA interactions⁵⁵. In VN4, the *WD repeat-containing* node, a member of the regulatory protein group which participates in a wide array of biochemical mechanisms and cellular processes⁵⁶ was selected as the gene identifier. There are four predicted uncharacterized proteins and one conserved hypothetical protein in VN4 that could be implicated with probable protein-protein interactions. Other repeat protein nodes present in VN4 are the *Kelch repeat containing At3g27220* and *Pentatricopeptide repeat-containing At1g31920*. The kelch repeat and WD forms a beta-propeller⁵⁷ while the pentatricopeptide repeat encodes the folding of a quadrilateral beta-helix⁵⁸. The repeat families have been characterized in signal transduction, cell division, light signaling, floral development and other transcriptional and signaling mechanisms^{59,60}. In addition, the presence of *Ras-related Rab11C* and *Ras-related Rab7* nodes, from the group of small GTPase binding proteins may implicate the event of signal transduction. The GTP-binding proteins are molecular switches that activates and inactivates in the presence of GTP and GDP, respectively^{58,60}.

In plants, the shoot apical meristem (SAM) homeostasis enables continuous and reiterative growth of plant organs until termination takes place for the function of floral meristem and flowering. The SAM homeostasis is governed by a feedback loop signaling which involves WUSCHEL (WUS), a homeobox transcription factor encoding stem cell specification at the organizing centre (OC) or the middle region of SAM. The presence of *ULTRAPETALA (ULT) 1-like* node in VN4, a plant specific protein suggests putative regulation of shoot and floral meristem size, floral organ regulation and indeterminate phase change during *J. curcas* development⁶¹. In *A. thaliana*, the *ULT1* has been described as a negative regulator of mitotic division taking place in the shoot apical meristems (SAM), where the gene restricts the expression of WUS at OC. A *ULT1* loss-of-function mutants showed enlarged inflorescence and flower meristem size, thereby confirmed the *ULT1* gene function for maintenance of shoot and floral meristem. The *ULT1* is composed of a SAND (Sp100, AIRE-1, NucP41/75 and DEAF-1) putative DNA-interacting domain and a B-box, among others that are actively involved in protein-protein interaction^{23,61,62}.

Three nodes associated to protein kinase namely the *serine threonine-kinase D6PK-like*, *serine threonine-kinase D6PKL2* and *serine hydroxymethyltransferase 4* were identified in VN4. Protein kinases are cellular regulatory components of signal transduction in plants. The presence of regulatory nodes such as *protein kinases* and *calcineurin B3* may indicate probable occurrence of cell differentiation, growth and development in the plant system⁶³. The calcineurin B like (CBL) proteins are calcium binding proteins that decodes Ca²⁺ signals for signaling and regulatory processes in plants whereas protein serine/threonine kinases have been reported to act as the central processing unit which accepts (environmental conditions) and converts information into appropriate outputs/responses during plant growth and development⁶⁴. A typical plant genome encodes more than 200 putative WDR-containing proteins^{39,60,65}. Probable shoot and floral meristem regulations, depicted by the *ULT 1-like* node together with other regulatory nodes further enhanced the role of VN4 in *Jatropha* shoot system homeostasis. Further studies especially mutant analyses of novel proteins in VN4 could potentially confirmed the putative roles of these candidate genes in the production of flowers and inflorescences (Fig. 3D).

Pathway enrichment analysis. Each VN were described according to the number of genes present in the metabolism, genetic information processing, environmental information processing and cellular process pathways (Fig. 4A). The VN1 showed active participation in genetic information (7 genes) processing; folding sorting and degradation and replication and repair. In addition, the genes under the base excision repair showed strong correlation to mismatch repair map (Fig. 4B). Others include participation in metabolism (3 genes) and cellular processes (6 genes). VN2 was highly involved in metabolism (9); energy (oxidative phosphorylation and nitrogen metabolism), lipid and amino acid. Number of genes in genetic information processing and cell communication were 2 and 1, respectively and genes with putative role in environmental information processing pathway were absent. The genetic information processes include transcription, folding, sorting and degradation and no correlations between the pathway maps were found. Both plant hormone signal transduction and mTOR signaling pathway were recruited under the environmental information processing in VN2. VN3 showed participation in metabolism (amino acid and cofactors and vitamin) environmental information processing and cellular process with 2 genes in each pathway. Environmental processing includes membrane transport and signal transduction. No correlation was found between the pathway maps. Pathway enrichment analysis of VN4 indicated involvement in metabolism (5 genes), genetic (10 genes) and environmental (2 genes) information processing and cellular process (9 genes) (Fig. 4).

The impact of vicinity networks (VNs) on *J. curcas* yield. Although the *J. curcas* plant has been extensively characterized as a promising biofuel feedstock, an unreliable and inconsistent yield during large-scale cultivation has severely affected the production. Experimental yields ranging from 10–12 tonnes of oil per hectare per year is about 20-fold far exceeding the average yield obtained at field scale; 0.5 tonnes of oil per hectare per year^{1–5}. The seed-oil yield is highly dependent on the plant shoot system as the yield index is influenced by various biological processes as followings; vegetative-floral meristem transition, secondary and tertiary branching, floral meristem to floral identity transition, pollen fertility, sex specification and female flower seed setting efficiency^{11,12,23}. Nevertheless, the plant shoot system affects the seed yield harvesting index at both vegetative and reproductive states. In plant, the shoot apical meristem (SAM) established during embryogenesis forms all the other above-ground organs and a switch from SAM to inflorescence meristem initiates the plant's reproductive biology^{24,61,62}. Therefore, SAM plays a crucial function in the development of plant shoot system and subsequent seed yield harvesting index. In this study, we report the first gene co-expression network model constructed to study gene-to-gene association among the reproductive-related shoot tissues in *J. curcas* which includes shoot apex bearing SAM, shoot bearing the inflorescence and the inflorescence.

Our vicinity networks (VNs) predicted candidate genes underpinning yield-related biological processes in *J. curcas*. The pathway enrichment analysis of VN1 indicated that a large number of genes were involved in genetic information processes such as folding, sorting and degradation and replication and repair. These events are hallmarks of epigenetic modifications; a genetic process that affects the organism's phenotype by changing the gene expressions level and chromosome biology without altering the DNA sequence⁶⁶. In several studies conducted across different continents, the occurrence of stable epigenetic events during *J. curcas* development have shown to cause significant epigenetic diversity within and among the accessions^{66–69}. On the other hand, others particularly molecular marker studies demonstrated a high phenotypic variation in parallel to low level of genetic diversity^{70–73}. These findings suggest that phenotypic variances in *J. curcas* may had been caused by epigenetic polymorphisms as evident by VN1 pathway enrichment analysis. Here, VN1 could be exploited to confirm possible involvement of epigenetics in *J. curcas* development and to understand the effect of epigenetic events on yield-related biological processes.

Belonging to the Euphorbiaceae or the latex bearing family, the *J. curcas* plant is composed of laticifers or latex producing cells. The articulated (more than one cell is placed on the other) laticifers contain latex in a suspension emulsion state and are found ubiquitously permeating various tissues in *J. curcas* plant body. Articulate laticifers originate in both primary and secondary tissues and are composed of predominantly phenolic and others such as starch grains, lipid and mucilage⁴⁸. In VN2, a distinct relationship between the laticifers and chlorophyll molecule biosynthesis was observed in presence of the vitamin B6 cofactor node. Interestingly, the nodes in VN2 matched the content of latex cells; production of phenolics may have been facilitated by vitamin B6 cofactor whereas starch production by photosynthesis was aided by chlorophyll molecule. The chlorophyll molecules are green pigments that capture light energy during photosynthesis, and the end-product, glucose is either taken up for immediate use by plant or is stored as starch. The *J. curcas* total chlorophyll content had showed positive implication on oil yield⁷⁴. Latex cells indicate production of lipid and concomitantly, lipid metabolism is seen enriched within the VN2. Lipid metabolism is an energy consuming processes, both chlorophyll and latex biosynthesis were found correlated within the VN2 and the result matched with the pathway enrichment of VN2 which indicated highest activity in the metabolism map; energy (oxidative phosphorylation and nitrogen metabolism), lipid and nucleotide.

Despite being a drought-endurable hardy plant, the *J. curcas* suffers from adverse environmental conditions, affecting its productivity tremendously. Heat caused by extreme temperature damages the chlorophyll molecules found on leaves. As a result, a protective response is triggered to encounter or to minimize any further damage that is likely to incur. In parallel with the environmental stress factors, the VN3 was able to explain heats stress mechanisms and potential protective mechanisms which might had taken place in the chlorophyll molecules; the signal transduction and membrane transport processes from the environmental information processes showed the highest activity in the VN3 pathway maps. In addition, the involvement of amino acid metabolism together with cofactors and vitamin metabolism may impart the occurrence protective mechanisms. Therefore, the VN3 may correspond to environmental stresses such as drought and heat conditions.

Seed set and flowering are critical components for seed yield improvement. According to the Arabidopsis flowering model, the LEAFY (LFY) gene plays the central role to regulate flowering; LFY interacts with TERMINAL FLOWER1 (TFL1) and AGAMOUS (AG) for floral signals. In addition, LFY together with APETALA1 (AP1) genes are responsible for the vegetative to reproductive phase switch⁷⁵. Others, with potent role in floral development include the flowering time regulator (SVP) and flowering regulators such as FLOWERING LOCUS D and F, SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1 AND CONSTANS⁷⁶. In a study conducted in China, the JcLFY, an orthologue of the Arabidopsis LFY was overexpressed in *J. curcas* and the subsequent functional analysis significantly affected the inflorescence structure, floral organs and fruit shape¹⁹. The same study reported expression of LFY transcripts in inflorescence buds, carpels and flower buds (highest). Although numerous findings in seed plants have reported the expression of LFY in inflorescence and flower primordia, our data showed absence of LFY in the *J. curcas* shoot transcriptome data. These results were in agreement with that previously reported in *J. curcas* pedicels and inflorescence. In another study conducted on AP1, a floral meristem and organ identity gene in higher plants, the over-expressed plant showed no apparent effect on the flowering time and floral organs in *Jatropha*⁷⁷.

Here, VN4 postulated candidate genes that may have involved in *J. curcas* flowering mechanism, although none of these genes have been reported in Arabidopsis flowering model. The presence of ULTRAPETALA (ULT) in VN4 may explain the termination of shoot and floral meristems in *J. curcas*. Under the vegetative state, the shoot apical meristem (SAM) continuously reiterates to produce SAM with the production of leaves and axillary meristems. Flower development is only observed after the differentiation of SAM into floral meristems (FMs)⁵⁵.

The pathway analysis of VN4 indicated highest activity in both genetic information processing and cellular processes. Cellular processes which include transport and catabolism, cell growth and death and cell communication further corroborated the putative role of VN4 in tissue differentiation. Therefore, the VN4 may be employed in flowering enhancement strategies such improved rate of SAM to inflorescence and/or floral meristem switch and increased number of inflorescence in *J. curcas* plant.

Conclusion

This is the first gene expression profile-based gene co-expression network that investigated the gene-to-gene associations in *J. curcas* reproductive-related shoot system. An integration of present results and previous data available in literature elucidated the putative gene functions of vicinity networks involved in key biological processes associated to yield. We presented possible models underpinning the biosynthesis of chlorophyll molecules and laticifer cells, cell wall metabolism, flowering, heat stress tolerance and epigenetic events during *J. curcas* growth and development. The putative candidate genes modelled in the vicinity networks are of great importance in breeding strategies and subsequent production of superior and efficient *J. curcas* varieties. However, the candidate genes presented in each VN along with their expression profile requires functional validation prior to downstream application in *J. curcas* breeding programs.

References

- Dias, L. A., Missio, R. F. & Dias, D. C. Antiquity, botany, origin and domestication of *Jatropha curcas* Euphorbiaceae), a plant species with potential for biodiesel production. *Genetics and Molecular Research* 2719–2728, <https://doi.org/10.4238/2012.June.25.6> (2012).
- Fairless, D. Biofuel: the little shrub that could-maybe. *Nature* 449, 652–655, <https://doi.org/10.1038/449652a> (2007).
- Ghosh, A. *et al.* Prospects for *Jatropha* methyl ester (biodiesel) in India. *International Journal of Environmental Studies* 64, 659–674 (2007).
- Jain, S. & Sharma, M. P. Biodiesel production from *Jatropha curcas* oil. *Renewable & Sustainable Energy Reviews* 14, 3140–3147 (2010).
- Koh, M. Y., Idaty, T. & Ghazi, M. A review of biodiesel production from *Jatropha curcas* L. oil. *Renewable & Sustainable Energy Reviews* 15, 2240–2251 (2011).
- Mandpe, S., Kadlaskar, S., Degen, W. & Keppeler, S. On road testing of advanced common rail diesel vehicles with biodiesel from the *Jatropha curcas* plant. *Soc. Automot. Eng.* 26, 356–364 (2005).
- Openshaw. A review of *Jatropha curcas*: an oil plant of unfulfilled promise. *Biomass Bioenergy* 9, 1–15 [https://doi.org/10.1016/S0961-9534\(00\)00019-2](https://doi.org/10.1016/S0961-9534(00)00019-2) (2000).
- Prakash, A. R., Singh, S., Prakash, C. R., Ghosh, A. & Agarwal, P. K. Development of *Jatropha* hybrids with enhanced growth, yield and oil attributes suitable for semi-arid wastelands. *Agroforestry Systems* 90, 541–553 (2016).
- Alvarez-Buylla, E. R. *et al.* Flower development. *Arabidopsis* 8, e0127, <https://doi.org/10.1199/tab.0127> (2010).
- Schmitz, G. & Theres, K. Shoot and inflorescence branching. *Current Opinion in Plant Biology* 8, 506–511, <https://doi.org/10.1016/j.pbi.2005.07.010> (2005).
- Bhattacharya, A., Kalayani, D. & Subodh, K. D. Floral biology, floral resource constraints and pollination limitation in *Jatropha curcas* L. *Pakistan J Biol Sci* 8, 456–460, <https://doi.org/10.3923/pjbs.2005.456.460> (2005).
- Bhuva, H. *et al.* Variability in growth and yield parameters in different provenances of *Jatropha curcas* on marginal land of Gujarat. In: Satish Kumar M (ed.) Bio-fuel plants, pp 139–146 (2007).
- Azevedo Peixoto, L. D., Laviola, B. G., Alves, A. A., Rosado, T. B. & Bhering, L. L. Breeding *Jatropha curcas* by genomic selection: A pilot assessment of the accuracy of predictive models. *PLoS ONE* 12, e0173368, <https://doi.org/10.1371/journal.pone.0173368> (2017).
- Costa, G. G. L. *et al.* Transcriptome analysis of the oil-rich seed of the bioenergy crop *Jatropha curcas* L. *BMC Genomics* 11, 462, <https://doi.org/10.1186/1471-2164-11-462> (2010).
- Dhakshanamoorthy, D., Selvaraj, R. & Chidambaram, A. L. Induced mutagenesis in *Jatropha curcas* L. using gamma rays and detection of DNA polymorphism through RAPD marker. *Comptes Rendus. Biology* 334, 24–30, <https://doi.org/10.1016/j.crv.2010.11.004> (2011).
- Gangwar, M., Sood, H. & Chauhan, R. S. Genomics and relative expression analysis identifies key genes associated with high female to male flower ratio in *Jatropha curcas* L. *Molecular Biology Reports* 43, 305–322, <https://doi.org/10.1007/s11033-016-3953-7> (2016).
- King, A. J., Li, Y. & Graham, I. A. Profiling the Developing *Jatropha curcas* L. Seed Transcriptome by Pyrosequencing. *Bioenergy. Research* 4, 211–221 (2011).
- Maghuly, F. & Laimer, M. *Jatropha curcas*, a biofuel crop: Functional genomics for understanding metabolic pathways and genetic improvement. *Biotechnology Journal* 8, 1172–1182, <https://doi.org/10.1002/biot.201300231> (2013).
- Tang, M. *et al.* An ortholog of LEAFY in *Jatropha curcas* regulates flowering time and floral organ development. *Scientific Reports* 6, 37306, <https://doi.org/10.1038/srep37306> (2016).
- Xu, G., Huang, J., Yang, Y. & Yao, Y. Transcriptome analysis of flower sex differentiation in *Jatropha curcas* L. using RNA Sequencing. *PLoS ONE* 11, e0145613, <https://doi.org/10.1371/journal.pone.0145613> (2016).
- Che-Mat, N. H., Bhuiyan, M. A. R., Senan, S., Yaakob, Z. & Ratnam, W. Selection of high yielding *Jatropha curcas* L. accessions for elite hybrid seed production. *Sains Malaysiana* 44, 1567–1572 (2015).
- Barton, M. K. & Poethig, R. S. Formation of the shoot apical meristem in *Arabidopsis thaliana*: an analysis of development in the wild type and in the shoot meristemless mutant. *Development* 119, 823–831 (1993).
- Benlloch, R., Berbel, A., Serrano-Mislata, A. & Madueno, F. Floral initiation and inflorescence architecture: A comparative view. *Ann. Bot.* 100, 1609–1609, <https://doi.org/10.1093/aob/mcm293> (2007).
- Poethig, R. S. Phase change and the regulation of shoot morphogenesis in plants. *Science* 250, 923–930 (1990).
- Petereit, J., Smith, S. Jr., F. C. H. & Schlauch, K. A. petal: Co-expression network modelling in R. *BMC Systems Biology* 10, 51, <https://doi.org/10.1186/s12918-016-0298-8> (2016).
- Schaefer, R. J., Michno, J. M. & Myers, C. L. Unraveling gene function in agricultural species using gene co-expression networks. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1860, 53–63, <https://doi.org/10.1016/j.bbagr.2016.07.016> (2016).
- Allocco, D. J., Kohane, I. S. & Butte, A. J. Quantifying the relationship between co-expression, co-regulation and gene function. *BMC Bioinformatics* 5, 18, <https://doi.org/10.1186/1471-2105-5-18> (2004).
- Ouyang, Y., Huang, X., Lu, Z. & Yao, J. Genomic survey, expression profile and co-expression network analysis of OsWD40 family in rice. *BMC Genomics* 13, 100, <https://doi.org/10.1186/1471-2164-13-100> (2012).
- Usadel, B. *et al.* Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. *Plant Cell Environment* 32, 1633–1651, <https://doi.org/10.1111/j.1365-3040.2009.02040.x> (2009).
- Govender, N., Senan, S., Mohamed-Hussein, Z. A. & Ratnam, W. Genomics Data Transcriptome analysis of reproductive tissue differentiation in *Jatropha curcas* L. *Genomics. Data* 13, 11–14, <https://doi.org/10.1016/j.gdata.2017.05.008> (2017).

31. Sangha, J. S., Gu, K., Kaur, J. & Yin, Z. An improved method for RNA isolation and cDNA library construction from immature seeds of *Jatropha curcas* L. *BMC Research Notes* **3**, 126, <https://doi.org/10.1186/1756-0500-3-126> (2010).
32. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–6, <https://doi.org/10.1093/bioinformatics/bti610> (2005).
33. Gotz, S. *et al.* High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* **36**, 3420–35, <https://doi.org/10.1093/nar/gkn176> (2008).
34. Neer, E. J., Schmidt, C. J., Nambudripad, R. & Smith, T. F. The ancient regulatory-protein family of WD-repeat proteins. *Nature* **371**, 812, <https://doi.org/10.1038/371297a0> (1994).
35. Sakurai, N. *et al.* KaPPA-View4: a metabolic pathway database for representation and analysis of correlation networks of gene co-expression and metabolite co-accumulation and omics data. *Nucleic Acids Research* **39**, D677–D684, <https://doi.org/10.1093/nar/gkq989> (2011).
36. Antony, E. & Hingorani, M. M. Mismatch recognition-coupled stabilization of Msh2-Msh6 in an ATP-bound state at the initiation of DNA repair. *Biochemistry* **42**, 7682–7693, <https://doi.org/10.1021/bi034602h> (2003).
37. Chau, N. C. *et al.* Oligouridylylate binding protein 1b plays an integral role in plant heat stress tolerance. *Frontiers in Plant Science*, <https://doi.org/10.3389/fpls.2016.00853> (2016).
38. Davidson, A. L., Dassa, E., Orelle, C. & Chen, J. Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiology & Molecular Biology Reviews* **72**, 317–64, <https://doi.org/10.1128/MMBR.00031-07> (2008).
39. Sharma, M. & Pandey, G. K. Expansion and function of repeat domains proteins during stress and development in plants. *Frontiers in Plant Science* **6**, 1218, <https://doi.org/10.3389/fpls.2015.01218> (2015).
40. Falbel, T. G. *et al.* SCD1 is required for cell cytokinesis and polarized cell expansion in *Arabidopsis thaliana*. *Development* **130**, 4011–4024 (2003).
41. Kurth, E. G. *et al.* Myosin-driven transport network in plants. *PNAS* **114**, E1385–E1394 (2017).
42. Long, X., He, B., Fang, Y. & Tang, C. Identification and characterization of the glucose phosphate dehydrogenase gene family in the para rubber tree, *Hevea brasiliensis*. *Frontiers in Plant Science* **7**, 215, <https://doi.org/10.3389/fpls.2016.00215> (2016).
43. Mangeon, A., Junqueira, R. M. & Sachetto-martins, G. Functional diversity of the plant glycine-rich proteins superfamily. *Plant signaling & behavior* **5**, 99–104 (2010).
44. Marowa, P., Ding, A. & Kong, Y. Expansins: roles in plant growth and potential applications in crop improvement. *Plant Cell Reports* **35**, 949–965, <https://doi.org/10.1007/s00299-016-1948-4> (2016).
45. Tenhaken, R. Cell wall remodeling under abiotic stress. *Frontiers in Plant Science* **5**, 1–9, <https://doi.org/10.3389/fpls.2014.00771> (2015).
46. Turner, W. L., Waller, J. C. & Snedden, W. A. Identification, molecular cloning and functional characterization of a novel NADH kinase from *Arabidopsis thaliana*. *Biochemical Journal* **385**, 217–23, <https://doi.org/10.1042/BJ20040292> (2004).
47. Martino, L., Safo, M. K., Musayev, F. N., Bossa, F. & Schirch, V. Structure and mechanism of *Escherichia coli* pyridoxine 5 V -phosphate oxidase. **1647**, 76–82 (2003).
48. Krishnamurthy *et al.* Chapter 1: Laticifers of *Jatropha*. Bahadur B. *et al.* (eds), In *Jatropha*, Challenges for a new energy crop: Vol 2: Genetic improvement and Biotechnology, https://doi.org/10.1007/978-1-4614-4915-7_1 (2013).
49. Doelling, J. H., Yan, N., Kurepa, J., Walker, J. & Vierstra, R. D. The ubiquitin-specific protease UBPI4 is essential for early embryo development in *Arabidopsis thaliana*. *Plant Journal* **27**, 393–405, <https://doi.org/10.1046/j.1365-313X.2001.01106.x> (2001).
50. Lambermon, M. H. *et al.* UBPI, a novel hnRNP-like protein that functions at multiple steps of higher plant nuclear pre-mRNA maturation. *EMBO Journal* **19**, 1638–1649 (2000).
51. Weber, C., Nover, L. & Fauth, M. Plant stress granules and mRNA processing bodies are distinct from heat stress granules. *Plant Journal* **56**, 517–530, <https://doi.org/10.1111/j.1365-313X.2008.03623.x> (2008).
52. Hudson, M. E., Lisch, D. R. & Quail, P. H. The *FHY3* and *FAR1* genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway. *The Plant Journal* **34**, 453–471 (2003).
53. Green, P. J. Control of mRNA stability in higher plants. *Plant Physiology* **102**, 1065–1070 (1993).
54. Lindner, A. C. *et al.* Isopentenyltransferase-1 (IPT1) knockout in *Physcomitrella* together with phylogenetic analyses of IPTs provide insights into evolution of plant cytokinin biosynthesis. *Journal of Experimental Botany* **65**, 2533–2543, <https://doi.org/10.1093/jxb/eru142> (2014).
55. Miller, J. C., Chezem, W. R. & Clay, N. K. Ternary WD40 repeat-containing protein complexes: Evolution, composition and roles in plant immunity. *Frontiers in Plant Science* **6**, 1–16, <https://doi.org/10.3389/fpls.2015.01108> (2016).
56. van Nocker, S., & Ludwig, P. (2003). The WD-repeat protein super family in *Arabidopsis*: conservation and divergence in structure and function. *BMC Genomics* **4**, <https://doi.org/10.1186/1471-2164-4-50> (2003).
57. Adams, J., Kelso, R. & Cooley, L. The kelch repeat superfamily of proteins: propellers of cell function. *Trends Cell Biol.* **10**, 17–24, [https://doi.org/10.1016/S0962-8924\(99\)01673-6](https://doi.org/10.1016/S0962-8924(99)01673-6) (2000).
58. Vetting, M. W. *et al.* Pentapeptide repeat proteins. *Biochemistry* **45**, 1–10, <https://doi.org/10.1021/bi052130w> (2006).
59. Saito, R. *et al.* A travel guide to Cytoscape plugins. *Nature Methods* **9**, 1069–1076, <https://doi.org/10.1038/nmeth.2212> (2012).
60. Villanueva, M. A., Islas-flores, T. & Ullah, H. Editorial: Signaling through WD-Repeat Proteins in Plants. *Frontiers in Plant Science* **7**, 1–2, <https://doi.org/10.3389/fpls.2016.01157> (2016).
61. Moreau, F. *et al.* The Myb-domain protein ULTRAPETALA1 INTERACTING FACTOR 1 controls floral meristem activities in *Arabidopsis*. *Development* **143**, 1108–1119, <https://doi.org/10.1242/dev.127365> (2016).
62. Coen, E. S. *et al.* FLORICAULA a homeotic gene required for flower development in *Antirrhinum-majus*. *Cell* **63**, 1311–1322, [https://doi.org/10.1016/0092-8674\(90\)90426-f](https://doi.org/10.1016/0092-8674(90)90426-f) (1990).
63. Stress, B. Functional roles of plant protein kinases in signal transduction pathways during sbiotic and biotic stress. *Journal of Biodiversity, Bioprospecting and Development* **2**, 147, <https://doi.org/10.4172/2376-0214.1000147> (2015).
64. Hardie, D. G. PLANT PROTEIN SERINE/THREONINE KINASES: Classification and functions. *Annual Review of Plant Physiology and Plant Molecular Biology* **50**, 97–131 (1999).
65. Mesarich, C. H., Bowen, J. K., Hamiaux, C. & Templeton, M. D. Repeat-containing protein effectors of plant-associated organisms. *Frontiers in Plant Science* **6**, 872, <https://doi.org/10.3389/fpls.2015.00872> (2015).
66. Yi, C., Zhang, S., Liu, X., Bui, H. T. N. & Hong, Y. Does epigenetic polymorphism contribute to phenotypic variances in *Jatropha curcas* L? *BMC Biology* <https://doi.org/10.1186/1471-2229-10-259> (2010).
67. Li, H. *et al.* Genetic tracing of *Jatropha curcas* L. from its mesoamerican origin to the worldfront. *Frontiers in Plant Science* **8**, 1539, <https://doi.org/10.3389/fpls.2017.01539> (2017).
68. Montes Osorio, L. R. *et al.* High level of molecular and phenotypic biodiversity in *Jatropha curcas* from Central America compared to Africa, Asia and South America. *BMC Plant Biology* **14**, <https://doi.org/10.1186/1471-2229-14-77> (2014).
69. Popluechai, S. *et al.* Narrow genetic and apparent phenetic diversity in *Jatropha curcas*: initial success with generating low phorbol ester interspecific hybrids. *Nature Proceedings* (2009).
70. Senan, S., Ismail, I. & Ratnam, W. Genetic homogeneity in *Jatropha curcas* L. individuals as revealed by microsatellite markers: Implication to breeding strategies. *Brazilian Journal of Botany* **39**, 3, <https://doi.org/10.1007/s40415-014-0117-7> (2016).
71. Shabanimofrad, M., Rafii, M. Y., Megat, W., Biabani, A. R. & Latif, M. A. Phenotypic, genotypic and genetic divergence found in 48 newly collected Malaysian accessions of *Jatropha curcas* L. *Industrial Crops and Products* **42**, 543–551, <https://doi.org/10.1016/j.indcrop.2012.06.023> (2013).

72. Sudheer, D. V. N. P., Mastan, S. G., Rahman, H. & Reddy, M. P. Molecular characterization and genetic diversity analysis of *Jatropha curcas* L. in India using RAPD and AFLP analysis. *Molecular Biology Reports* **37**, 2249–2257 (2010).
73. Yue, G. H. *et al.* No variation at 29 microsatellites in the genome of *Jatropha curcas*. *Journal of Genomics* **2**, 59–63, <https://doi.org/10.7150/jgen.7435> (2013).
74. Singh, S. *et al.* Trait selection by path and principal component analysis in *Jatropha curcas* for enhanced oil yield. *Industrial Crops and Products* **86**, 173–17, <https://doi.org/10.1016/j.indcrop.2016.03.047> (2016).
75. Weigel, D., Alvarez, J., Smyth, D. R., Yanofsky, M. F. & Meyerowitz, E. M. LEAFY controls floral meristem identity in *Arabidopsis*. *Cell* **69**, 843–859, [https://doi.org/10.1016/0092-8674\(92\)90295-n](https://doi.org/10.1016/0092-8674(92)90295-n) (1992).
76. Liljegren, S. J., Gustafson-Brown, C., Pinyopich, A., Ditta, G. S. & Yanofsky, M. F. Interactions among APETALA1, LEAFY, and TERMINAL FLOWER1 specify meristem fate. *Plant Cell* **11**, 1007–1018 (1999).
77. Torre, F., De, Sampedro, J., Zarra, I. & Revilla, G. AtFXG1, an Arabidopsis gene Encoding L-fucosidase active against fucosylated xyloglucan oligosaccharides 1. **128**, 247–255, <https://doi.org/10.1104/pp.010508> (2002).

Acknowledgements

This research was funded by the Research University Grants (AP-2012-008 and DIP-2012-019) from Universiti Kebangsaan Malaysia (UKM) and Ministry of Higher Education Malaysia (MOHE).

Author Contributions

N.G. conducted the bioinformatics analyses and wrote the manuscript S.S. performed the RNA isolation Z.A.M.H. and R.W. modified the analyses, conceived the project and secured funding.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-27493-z>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018