Scientia Horticulturae 257 (2019) 108727

Contents lists available at ScienceDirect

Scientia Horticulturae

journal homepage: www.elsevier.com/locate/scihorti

Transcriptional reprogramming during *Garcinia*-type recalcitrant seed germination of *Garcinia mangostana*

Hoe-Han Goh[a,*], Syuhaidah Abu Bakar[a], Nur Diyana Kamal Azlan[a], Zamri Zainal[a,b], Normah Mohd Noor[a]

[a] Institute of Systems Biology, Universiti Kebangsaan Malaysia, UKM Bangi, 43600 Selangor, Malaysia
[b] School of Bioscience and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, UKM Bangi, 43600 Selangor, Malaysia

ARTICLE INFO

Keywords:
Apomict
Germination
Mangosteen
Recalcitrant
RNA-seq
SeedNet

ABSTRACT

Mangosteen (*Garcinia mangostana* L.) is an apomictic tropical fruit tree species which asexually produces recalcitrant seeds that are sensitive to desiccation and not viable for long-term storage. It is important to understand seed germination to inform strategy for the propagation and conservation of important plants, especially for recalcitrant species. However, little is known about the molecular physiology of tropical seed germination. *Garcinia*-type seed germination is unique such that the primary root and shoot emerge from the opposite ends, which is akin to *in vivo* direct organogenesis. For the first time, time-course transcriptome-wide gene expression during mangosteen seed germination from day 0 to day 7 was performed through RNA-sequencing with RT-qPCR validation using independent samples in triplicates. The analyses of differentially expressed genes (DEGs) and expression profile identified D3 as a key turning point for transcriptional reprogramming with an indication of a transient shift to anaerobic respiration. Particularly, the activation of ethylene signalling through increased expression of genes involved in its biosynthesis and the upregulation of ethylene-responsive transcription factors lead to the induction of direct organogenesis concurrently with seed germination. Abscisic acid (ABA) signalling appeared to play a role in stress response while gibberellin (GA) provides growth potential during mangosteen seed germination. The antagonistic relationship between ABA and GA is conserved in mangosteen as manifested by global gene expression changes based on the Arabidopsis co-expression network analysis. The current transcriptomic study comprehensively describes various molecular aspects of *Garcinia*-type recalcitrant seed germination and highlights similarities to *in vitro* somatic embryogenesis but without a differentiated embryo. This study provides insights to improve seed germination for the mass propagation of mangosteen.

1. Introduction

Mangosteen (*Garcinia mangostana* L.) from the Clusiaceae family is known as the “Queen of fruits”, which is an evergreen tropical fruit tree native to Malaysia (Osman and Milan, 2006). Apart from its delectable white fruit pulp with a sweet unique taste, mangosteen pericarp contains pharmaceutically useful phytochemicals, especially the polyphenolic xanthones with various medicinal properties (Obolskiy et al., 2009). Due to the health benefits and commercial values from its pericarp extracts (Gutierrez-Orozco and Failla, 2013), mangosteen has attracted growers’ attention for mass propagation commonly through seeds. However, this effort is hindered by mangosteen being a slow-growing tree that bears fruits with only one or two seeds biennially after five to seven years (Osman and Milan, 2006).

Mangosteen is an obligate apomict, in which seeds with the identical genetic makeup to mother plant develop from integument cells through adventitious embryony instead of gamete cell fertilisation (Lim, 1984). Thus, crop improvement through a breeding programme is not feasible due to genetic homogeneity. Furthermore, mangosteen seeds cannot survive over two-week storage, low temperature (<7 °C) or desiccation below 30% (Normah et al., 1997) and fresh seed germination percentage is only around 75% (Noor et al., 2016). These factors make it difficult to secure and sustain planting materials for growers. Hence, *in vitro* propagation offer an alternative to seed propagation, such as shoot proliferation from seed segments (Goh et al., 1988, 1994; Huang et al., 2000; Normah et al., 1995) as well as embryogenic callus induction and somatic embryogenesis from leaf explants (Elviana et al., 2011; Maadon et al., 2016; Rohani et al., 2012; Te-Chato and Lim, 1999, 2000).

Mangosteen seed is unique with only a procambium structure and

* Corresponding author.
E-mail address: gohhh@ukm.edu.my (H.-H. Goh).

https://doi.org/10.1016/j.scienta.2019.108727
Received 3 July 2019; Received in revised form 25 July 2019; Accepted 26 July 2019
Available online 19 August 2019

no differentiated embryo upon maturation (Noor et al., 2016). This resembles globular somatic embryos of mangosteen that directly develop shoot organ without following typical embryogenesis stages (globular, heart, torpedo, and cotyledon) (Elviana et al., 2011). The seed is considered an elongated tuberculous hypocotyl with vasculature connecting the opposite ends of the seed that respectively germinates into the primary root and shoot (Noor et al., 2016). However, it is the adventitious root which formed at the base of the shoot that will eventually become the main root system. This is known as "*Garcinia*-type" seed germination, which is shared by many members of *Garcinia* genus, such as *G. indica*, *G. cambogia*, *G. xanthochymus* (Malik et al., 2005), *G. gummi-gutta* (Joshi et al., 2006), *G. intermedia* (Di Stefano et al., 2006), *G. kola* (Asomaning et al., 2011), *G. atroviridis*, *G. hombroniana*, and *G. prainiana* (Noor et al., 2016). Their seeds can have polyembryony characteristics (~12% occurrence in mangosteen) in which multiple seedlings can develop from a single seed, while seed cut into multiple segments containing procambium can independently develop shoot and root into individual seedlings (Noor et al., 2016). This has been proposed to be an evolutionary strategy of *Garcinia* species to exploit mammalian frugivory in seed dispersal (Joshi et al., 2006).

Functional genomics studies are still lacking a reference genome of mangosteen despite an on-going sequencing effort (Abu Bakar et al., 2016, 2017; Midin et al., 2017), perhaps hindered by the complex polyploidy nature of mangosteen genome of ~6 Gbp with 74–110 chromosomes (Midin et al., 2018). Hence, a *de novo* transcriptomics approach is required, such as the recent studies on mangosteen fruit ripening (Abdul-Rahman et al., 2017a; Matra et al., 2016) and somatic embryogenesis (Fadryin et al., 2018; Mahdavi-Darvari and Noor, 2017). Metabolomics approach was also taken to understand mangosteen seed development (Mazlan et al., 2018) and seed germination (Mazlan et al., 2019) by using mass spectrometry techniques coupled with gas-chromatography and liquid-chromatography.

Time-course transcriptome-wide studies of seed germination are largely limited to model species, such as Arabidopsis (Narsai et al., 2017a), rice (Narsai et al., 2017b), maize (Sekhon et al., 2013), barley (Lin et al., 2014), and wheat (Yu et al., 2014), which mostly produce orthodox seeds. Similar studies on recalcitrant tree species, including economically important coffee, cocoa, and rubber, are even scarcer. Literature available mainly reported analysis on targeted genes, such as *Quercus ilex* (Romero-Rodríguez et al., 2018) or focused on seed sensitivity towards desiccation, such as *Madhuca latifolia* (Chandra and Keshavkant, 2018). The current study represents the first time-course transcriptomic analysis of a recalcitrant species with unique *Garcinia*-type germination. This provides crucial information on the molecular mechanism of mangosteen seed germination as the start of its life cycle to maintain its survival and sustainability. Furthermore, comparative analysis against the orthodox seed of model plant Arabidopsis based on gene network (SeedNet) mapping allowed comprehensive comparison of global gene expression changes during early germination of mangosteen.

2. Materials and methods

2.1. Seed sample preparation

Seeds were harvested from multiple mature fruits of mangosteen trees at the experimental plot (2°55′09.0″N 101°47′04.8″E) of Universiti Kebangsaan Malaysia, Bangi. Fully developed seeds with uniform sizes (2.0 ± 0.2 cm) were chosen and cleaned by removing the flesh (aril) and rinsed with distilled water before germinated in autoclaved sand within a greenhouse (24–26 °C) with daily watering. Clean seeds before sowing were collected as the day 0 (D0) samples and subsequent samples were collected on day 3 (D3), day 5 (D5), and day 7 (D7) after sowing. Samples from all time points were immediately frozen in liquid nitrogen and stored at -80 °C before extraction.

2.2. RNA extraction, library preparation, and sequencing

Total RNA was extracted using modified CTAB method (Abdul-Rahman et al., 2017b) and combined with RNase-free QIAGEN RNeasy mini kit (QIAGEN, Germany) to remove any contaminating genomic DNA. The RNA quantity was estimated using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific., USA) to range from 1.80 to 2.20 for A_{260}/A_{280} ratio. The RNA quality was analysed by 1% agarose gel electrophoresis and RNA integrity (RIN > 7) was assessed by Bioanalyzer (Agilent, USA).

Total RNA samples of each time points of germination (D0, D3, D5, and D7) were selected to prepare four independent cDNA libraries with the TruSeq RNA Sample Prep kit (Illumina, USA) using a minimum of 4 µg total RNA according to the manufacturer's instructions. The four cDNA libraries were 100 bp paired-end sequenced using HiSeq 2000 Illumina platform at the Malaysia Genome Institute (MGI), Bangi, Selangor, Malaysia (Azlan et al., 2017). Raw sequencing reads were deposited to SRA public database: D0 (SRR5412331), D3 (SRR5412330), D5 (SRR5412329), and D7 (SRR5412328).

2.3. Transcriptome assembly and differentially expressed gene analysis

The mangosteen seed reference transcriptome was *de novo* assembled using Trinity analysis pipeline v2.1.1 (Haas et al., 2013) after Trimmomatic (Bolger et al., 2014) removal of adapter sequences and trimming of poor quality reads based on default settings. This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GGXM00000000. The version described in this paper is the first version, GGXM01000000.

Differentially expressed gene (DEG) analysis was performed between each time point (D0*vs*D3, D3*vs*D5, and D5*vs*D7) with edgeR (Robinson et al., 2009) based on the transcript abundance estimation using RSEM (Li and Dewey, 2011) from the read alignment results of individual samples against the reference transcriptome. Transcript abundance is expressed as Transcripts Per Kilobase Million (TPM), which not only normalised for gene length but also read depth for better comparison between samples. Statistically significant DEGs were defined by false discovery rate (FDR) < 0.001 and $|Log_2FC| > 2$.

2.4. Functional annotation and enrichment analysis

Assembled transcripts were annotated using Trinotate (Bryant et al., 2017) as part of Trinity pipeline based on BLAST searches with cut-off E-value of $1E^{-5}$ against the non-redundant (*nr*) and UniProt databases using BLASTx based on nucleotide sequences and BLASTp based on Transdecoder predicted peptide sequences respectively (Supplementary Material 1). Other annotation tools include RNAMMER, Pfam, SignalP, TmHMMM, eggnog, and gene ontology (GO) annotation. GO classification (Supplementary Material 1) was visualised using the online tool WEGO 2.0 (http://wego.genomics.org.cn/) based on 1/3/2013 GO file (Ye et al., 2006). KEGG Orthology (KO) annotation was performed using the online server (https://www.genome.jp/kegg/kaas/) with SBH method based on plant species selection (Moriya *et al.*, 2007). KEGG mapping (Supplementary Material 4) was performed by using an online server (https://www.genome.jp/kegg/mapper.html) with combined visualisation of DEGs using pathview (https://pathview.uncc.edu/home) (Luo *et al.*, 2017). Transcription factor was annotated using standalone software iTAK v1.7 with database v18.10 (Zheng et al., 2016) and combined with Arabidopsis annotation (Ilias et al., 2018).

Terms related to seed germination were semi-automatedly assigned to transcripts based on lists of AGI locus identifiers in combinations with a manual search of annotation description for specific keywords in a similar fashion to TAGGIT ontology (Supplementary Material 2) that was adapted from seed germination studies in Arabidopsis (Carrera et al., 2007; Nelson and Steber, 2017). The terms represent relevant biological processes during seed germination, which include

phytohormones, namely abscisic acid (ABA), gibberellin (GA), brassinosteroid (BR), auxin (IAA), ethylene (ET), cytokinin (CK), and jasmonic acid (JA).

Hierarchical clustering, profile plot, heatmap, and enrichment analysis (Fisher's exact test with Benjamini-Hochberg multiple test correction) were performed using Perseus software v1.6.2.2 (Tyanova et al., 2016) based on default settings. The hypergeometric test for overrepresentation analysis was performed in MS Excel. Enriched GO terms were visualised using the online tool REViGO (http://revigo.irb.hr/) (Supek *et al.*, 2011). Mapman v3.6.0RC1 (Thimm et al., 2004) visualisation of DEGs was generated based on Mercator mangosteen transcriptome sequence annotation (http://www.plabipd.de/portal/mercator-sequence-annotation) (Lohse et al., 2014).

2.5. Comparative DEG network analysis with Arabidopsis

Local BLASTn was performed against TAIR10 Arabidopsis sequence database (https://www.arabidopsis.org/) with a cut-off E-value of $1E^{-5}$ to identify homologous genes for comparative analysis between Arabidopsis and mangosteen seed germination based on SeedNet (http://netvis.ico2s.org/dev/seednet) using Cytoscape v3.4.0 (Smoot et al., 2011). SeedNet is generated based on co-expression analysis of data from seed germination studies through *Significance Analysis of Microarrays* (SAM) (Bassel et al., 2011). Mapping was done by matching mangosteen transcripts with Arabidopsis gene IDs and compared with *SAM_Classification* in the database (Supplementary Material 3). Hormonometer analysis (Volodarsky et al., 2009) was performed online (https://hormonometer.weizmann.ac.il/) based on FDR and Log_2FC values of Arabidopsis genes which matched significant DEGs. The list of 481 Arabidopsis essential genes was obtained from SeedGenes (http://seedgenes.org/GeneList.html) (Meinke et al., 2008) based on December, 2010 Release.

2.6. Reverse transcription quantitative PCR (RT-qPCR) analysis

First strand cDNA was synthesised using 250 ng of total RNA extracted from an independent set of seed samples in triplicates using Maxima H Minus First Strand cDNA synthesis kit (Thermo Scientific Inc., USA). The cDNA was adjusted to a 12.5 ng/μL concentration and diluted in a 1:10 ratio for standard curve analysis to evaluate primers' efficiencies. The primers were designed using Primer-BLAST (Ye et al., 2012) as listed in Table S1.

Quantitative real-time PCR (RT-qPCR) was performed using Biorad IQ5 multicolour real time PCR detection system using Thermo Scientific Luminaris Colour HiGreen qPCR mastermix 2×. PCR reactions include 1 μL of cDNA template (12.5 ng), 5 μL of LuminarisColorHiGreen master mix, 0.4 μL of 10 μM of each forward and reverse primer and nuclease-free water to a 10 μL final volume. The amplification programme was as follows: 95 °C for 3 min and 40 cycles of 95 °C for 5 s, and 55–60 °C for 30 s. PCR specificity was confirmed by end-cycle melt curve analysis and no-template controls (NTCs). Primer efficiency was determined based on a standard curve with five serial dilutions. The RT-qPCR analysis was performed using three biological replicates, each with three technical replicates. Expression of target genes were normalised based on internal reference actin (*ACT*) (comp53280_c1_seq1) and elongation factor 1 (*EF1*) (comp56279_c0_seq7) using 2^{-ddCt} method (Livak and Schmittgen, 2001). Statistical analysis was performed using linear regression to obtain the correlation between relative expression data of qPCR and RNA-seq.

2.7. Data availability statement

Transcriptome annotation is provided as Supplementary Material 1 available at Figshare: https://doi.org/10.6084/m9.figshare.8299217.v1 together with other Supplementary Materials. The datasets generated in the current study are available in the public repositories through the NCBI BioProject accession PRJNA321150.

3. Results

3.1. Mangosteen early seed germination

In this study, early seed germination of mangosteen was investigated at D0, D3, D5, and D7 (Fig. 1). The seed radicle became noticeable on D5 (0.5–1 mm) and clearly emerged on D7 (1–2 mm) when plumule was also becoming visible.

The previous study reported six to seven days until the radicle becomes emerged (1–2 mm) from the seed (Noor et al., 2016). However, radicle can sometimes be noticeable even on D3 (Mazlan et al., 2019). We followed the definition of seed germination *sensu stricto* which started as imbibition takes place and completed when radicle protrusion appears through the covering seed layer (Weitbrecht et al., 2011). Therefore, D0 to D5 represented all three phases of germination in this study, namely water uptake-lag interval (D0–D3) and radicle protrusion (D3–D5), whereas D5 to D7 can be considered as seedling establishment phase. Nonetheless, seedling establishment until the adventitious root system fully developed can take up to 24 days (Noor et al., 2016).

3.2. Mangosteen seed transcriptome assembly and functional annotation

From four cDNA libraries of samples that were harvested on day 0, 3, 5, and 7 after sowing, a total of 160,461,955 raw reads were trimmed into 141,687,906 (88.3%) clean reads for *de novo* assembly of a reference mangosteen seed transcriptome (Table 1). In total, 279,525 transcripts were generated with 121,506 unigenes and 188,214 peptide sequences predicted by Transdecoder. Transcript lengths range from 201 to 13,872 bp with an N50 length of 2382 bp and a mean length of 1383 bp. This is considered a good quality assembly with the majority (64.5%) of transcripts greater than 500 bp and 26.1% transcripts >2 kbp long (Table 1).

BLASTx searches against nr protein database found hits to 124,234 (44.4%) transcripts compared to 87,567 (31.3%) transcripts from BLASTp searches of predicted peptide sequences, of which 82,618 (29.6%) transcripts found hits to Pfam. Local BLASTn against Arabidopsis genes found 129,462 (46.3%) transcripts with significant hits. A total of 11,061 transcripts were annotated as transcription factors. No annotation information was found for 126,744 (45.3%) transcripts. To date, there are only 535 *Garcinia mangostana* nucleotide and 322 protein sequences in NCBI database and 187 sequences with only one reviewed in UniProtKB. Compared to the recent transcriptome

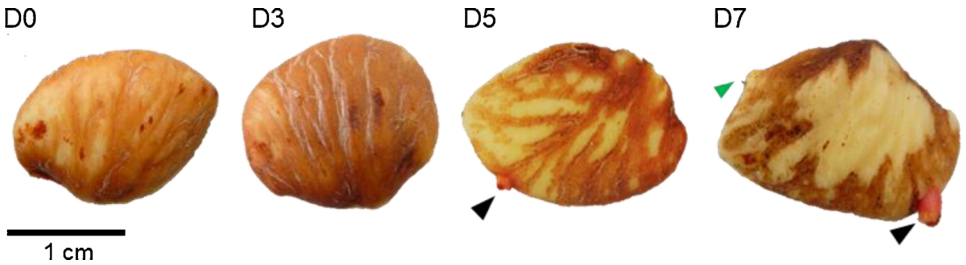

Fig. 1. Mangosteen seed germination at sowing (D0), 3 days (D3), 5 days (D5), and 7 days (D7) after sowing. There is a visible sign of protuberance after 5 days with radicle emergence by day 7 (black arrowheads). Plumule emergence at the opposite end of the radicle is indicated by a green arrowhead. Cross-section images can refer to Mazlan et al. (2019).

Table 1
Summary statistics on mangosteen seed transcriptome assembly and functional annotation.

Attribute	Number	%
Raw reads	160,461,955	
Clean reads	141,687,906	88.3
Percentage GC (%)		40.6
Number of unigenes	121,506	
Number of transcripts	279,525	
Number of predicted peptides (Transdecoder)	188,214	
Transcript length (bp)		
Total	386,610,882	
Range	201-13,872	
N50	2,382	
Mean	1,383	
Transcript length distribution (bp)		
200-299	53,724	19.2
300-499	45,552	16.3
500–999	47,276	16.9
1000–1999	59,878	21.4
>2000	73,095	26.1
Functional annotation		
nr (BLASTx)	124,234	44.4
Swiss-Prot (BLASTp)	87,567	31.3
AGI (BLASTn)	129,462	46.3
Pfam	82,618	29.6
eggNOG	62,001	22.2
KEGG (KAAS)	57,167	20.5
Transcription factor	11,061	4.0
Unannotated	126,744	45.3

analysis of mangosteen somatic embryos with 34.8% functionally annotated transcripts (Mahdavi-Darvari and Noor, 2017), the current dataset provides the most extensive functional annotation of mangosteen transcripts (Supplementary Material 1) for further studies.

3.3. Differentially expressed gene (DEG) analysis

Pairwise comparisons of transcripts expressed across the four sampling time points (D0, D3, D5, and D7) identified a total of 22,565 (8.1%) significant differentially expressed transcripts (DEGs) (Fig. 2a). Noteworthy, many of the DEGs (11,757) were shared among the three comparisons. There were more upregulated DEGs than downregulated DEGs across all three comparisons. The highest number of DEGs were found between D0 and D3 (0v3), in which more than half were shared with 3v5 comparison (Fig. 2b). The majority of the 0v3 upregulated transcripts were 3v5 downregulated; whereas many of the 0v3 downregulated transcripts were 3v5 upregulated (Fig. S1). However, this was not observed between 3v5 and 5v7. This indicates a major transcriptional reprogramming on D3 involving a common set of DEGs.

Based on Eukaryotic Orthologous Groups (KOG) classification (Fig. 3), DEGs were highly enriched in functional categories involving "cell cycle control, cell division, chromosome partitioning", "carbohydrate transport and metabolism", "transcription", and "signal transduction mechanisms". Conversely, "translation, ribosomal structure and biogenesis" and "posttranslational modification, protein turnover, chaperones" were highly underrepresented in DEGs relative to the transcriptome.

3.4. Transcriptome functional categories, expression profile, and enrichment analysis

To comprehend the biological processes throughout the time course of mangosteen seed germination, we functionally categorised transcripts according to processes related to seed germination based on a keyword search of annotation in a similar fashion to previous studies in Arabidopsis, known as a TAGGIT analysis (Carrera et al., 2007; Nelson and Steber, 2017). For functional enrichment analysis, we focused our analysis on 92,202 (33%) transcripts with Transcripts Per Kilobase Million (TPM) > 1 at any time point to reduce noise from transcripts with low abundance (TPM < 1). We further increased the statistical stringency of enrichment analysis, resulting in 27,464 transcripts by excluding "others", "unknown", and "unannotated" to avoid disproportionate overrepresentation of transcripts without categorised functions (Fig. 4; Supplementary Material 2).

Between D0 and D3, the upregulated DEGs were enriched in processes involving ethylene, glycolysis & gluconeogenesis, heat shock, and phosphatase, compared to the downregulated DEGs with cell-wall modification, cytokinin, photosynthesis/chloroplast related, and transporter (Fig. 4). Other regulated processes with both up- and down-regulation include kinase, lipid metabolism, reactive oxygen species (ROS)-related, seed storage/Late Embryogenesis Abundant (LEA) proteins, stress, and transcription factor. The opposite trend was observed for 3v5 with cell-wall modification, photosynthesis/chloroplast related, and transporter enriched in the upregulated DEGs, compared to ethylene, glycolysis & gluconeogenesis, phosphatase, and seed storage/LEA proteins in the downregulated DEGs. Cytoskeleton, pentose phosphate pathway, seed storage/LEA proteins, and translation-associated were also enriched in 3v5 upregulated DEGs. Lipid metabolism was enriched in both the up- and down-regulated 3v5 DEGs. For 5v7, the upregulated DEGs were enriched in cell-wall modification and RNA processing, compared to DNA repair and glycolysis & gluconeogenesis for the downregulated DEGs. Notably, dormancy-related, lipid metabolism, pentose phosphate pathway, kinase, and phosphatase were enriched in both the up- and down-regulated 5v7 DEGs.

Next, we grouped DEGs into eight clusters according to similar expression patterns and performed enrichment analysis (Fig. 5, Supplementary Material 3).

The largest cluster 4 consisting 4172 DEGs showed higher transcript

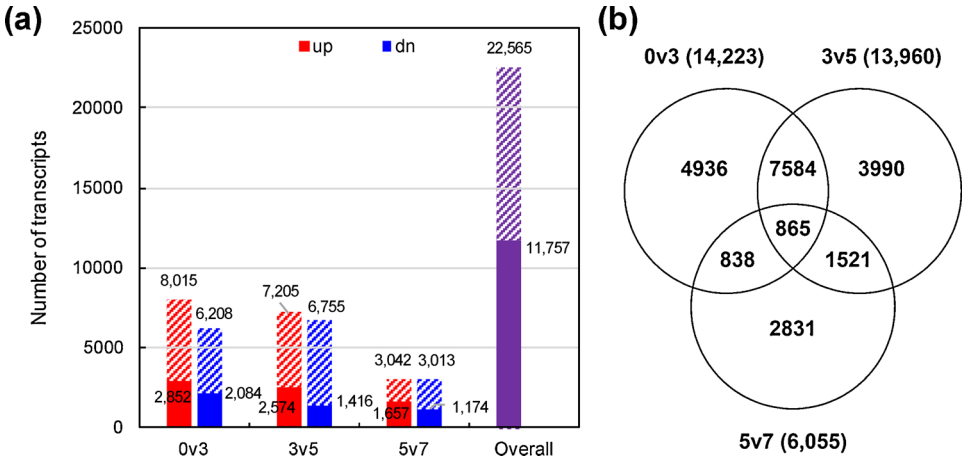

Fig. 2. Differentially expressed transcripts (DEGs). (a) Number of significantly (FDR < 0.001, $|Log_2FC| > 2$) upregulated (up) and downregulated (dn) transcripts for D0vsD3 (0v3), D3vsD5 (3v5), and D5vsD7 (5v7) comparisons. Solid fills represent the number of differentially expressed transcripts which are unique to the specific comparisons. (b) Venn analysis of DEGs from different comparisons. Number in parentheses showing the total number of transcripts.

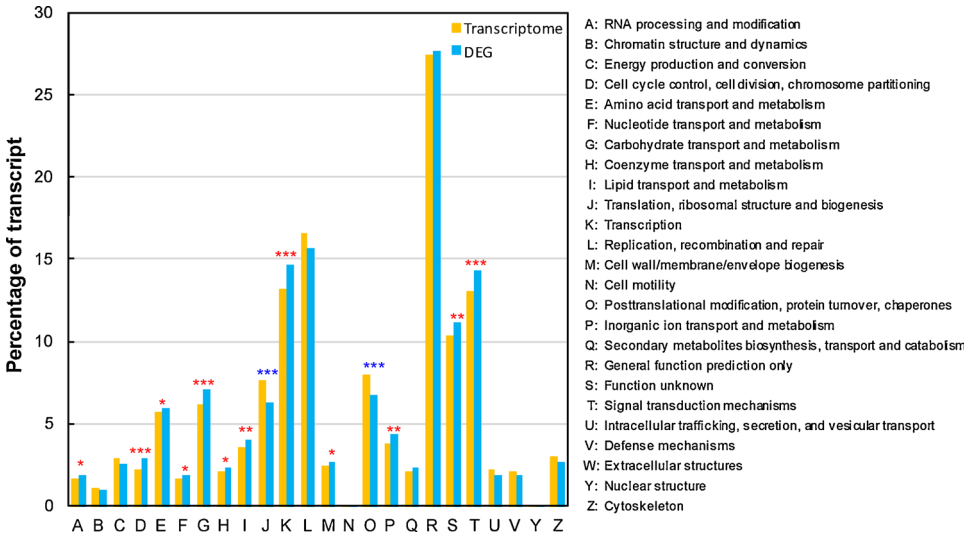

Fig. 3. Eukaryotic Orthologous Groups (KOG) classification of transcriptome and DEG based on eggNOGv3 classification. Fisher's exact test with Benjamini–Hochberg multiple test correction: *FDR < 0.05, **FDR < 0.01, ***FDR < 0.001. Red: overrepresentation; blue: underrepresentation.

expression on D3 with 0v3 upregulation and 3v5 downregulation, which was enriched in ethylene-activated signalling, fructose-6-phosphate (F6P) metabolism, and response to chitin, brassinosteroid (BR), salicylic acid (SA), and sucrose based on the gene ontology (GO) analysis. Cluster 7 showed the opposite pattern with lower expression on D3 and enriched with photosynthesis, response to karrikin and gibberellin (GA), secondary cell wall biogenesis, phloem sucrose loading, biosynthesis of tocopherol (vitamin E), cinnamic acid, flavonoid, and cytokinin (CK), carbohydrate metabolism, and catabolism of pectin and SA. Cluster 1 and cluster 2 showing higher expression on D0 and D3 with 3v5 downregulation were enriched in the regulation of transcription and RNA splicing respectively. This was contrasting to cluster 5 and cluster 6 with higher expression on D5, which were enriched in C-5 methylation of cytosine and cell wall pectin biosynthesis respectively. Cluster 3 and cluster 8 showed opposite expression patterns with lower and higher D7 expression, respectively, in which both were enriched in

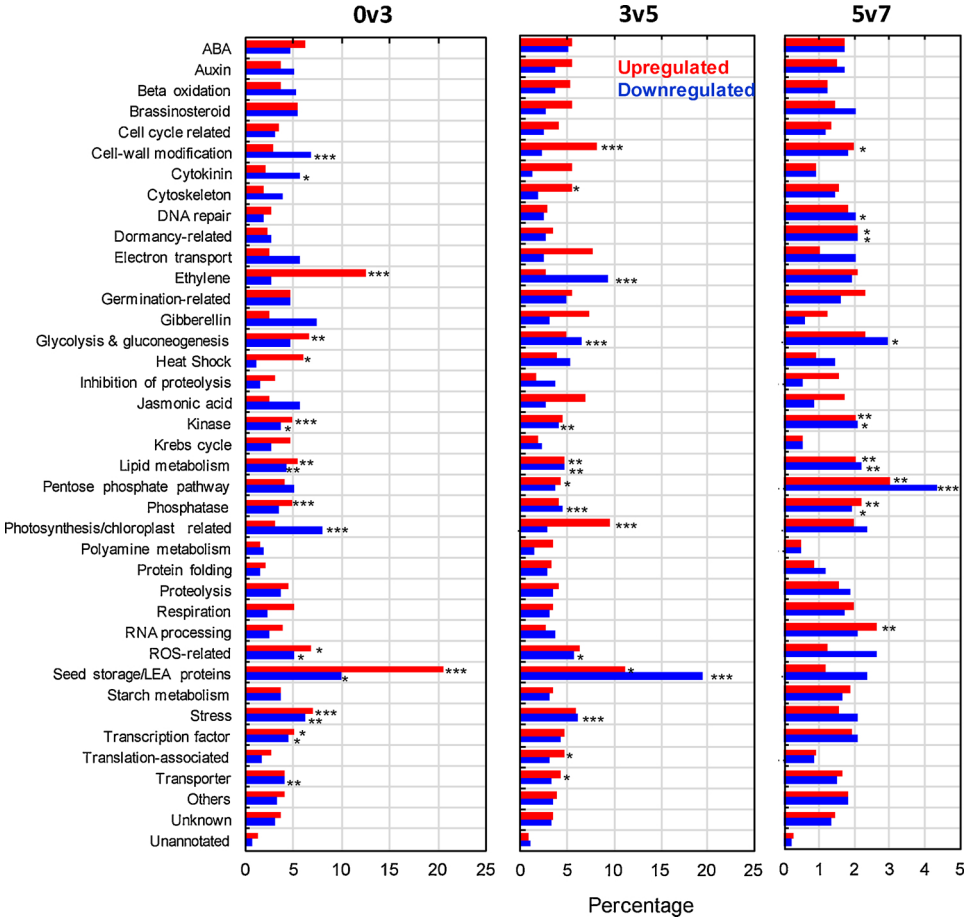

Fig. 4. TAGGIT ontology analysis of DEGs for D0vsD3 (0v3), D3vsD5 (3v5), and D5vsD7 (5v7) comparisons. The value on the *x*-axis shows the percentage number of DEGs against the total number of transcripts in each category. The *Y*-axis is sorted according to alphabetical order, except for others, unknown, and unannotated which were excluded in Fisher's exact test. Fisher's exact test with Benjamini-Hochberg multiple test correction: *FDR < 0.05, **FDR < 0.01, ***FDR < 0.001.

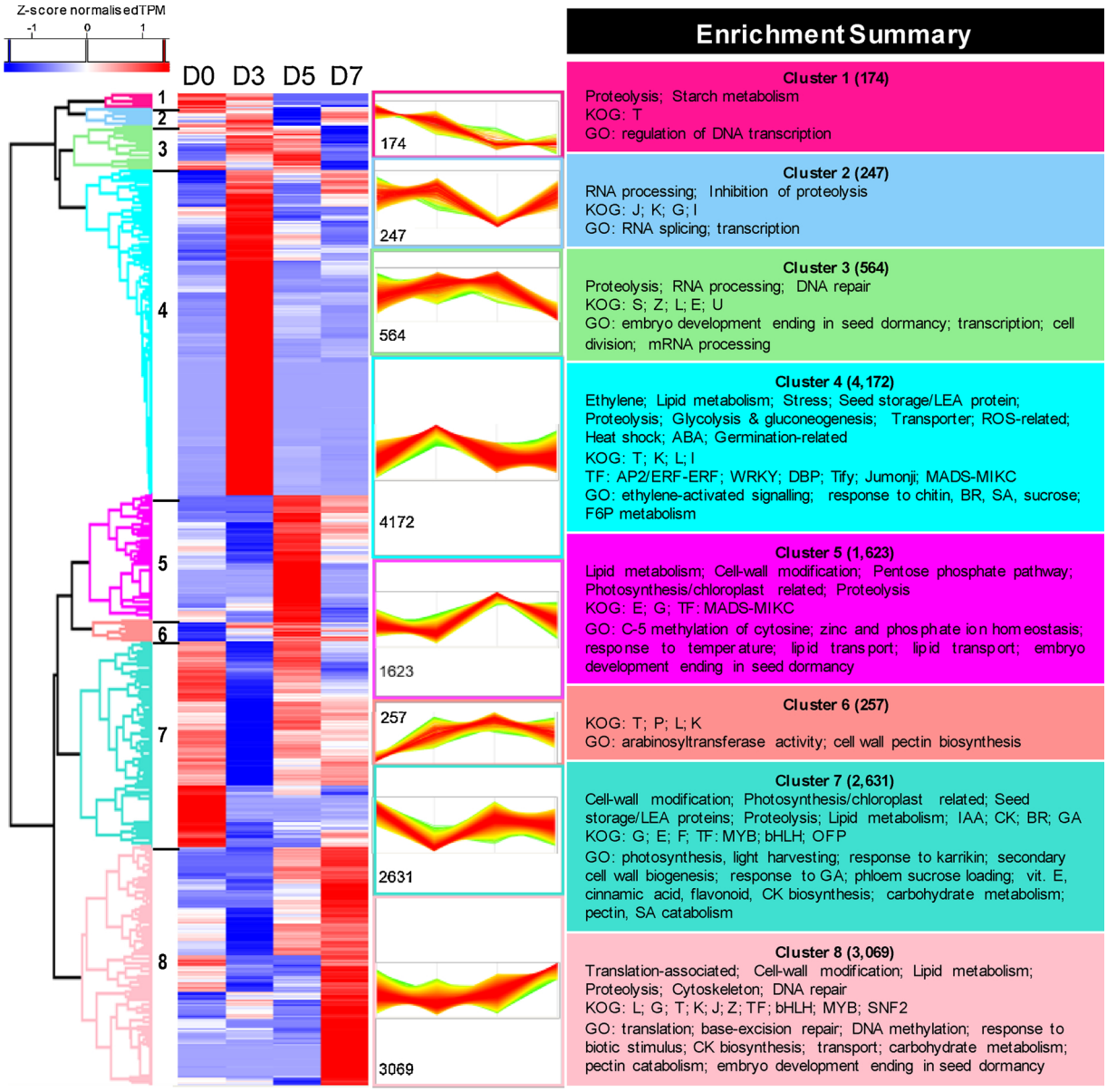

Fig. 5. Expression profile analysis and cluster enrichment summary. Heatmap and hierarchical clustering were performed based on 12,737 DEGs with TPM > 1 in at least one time point after excluding "unknown" and "unannotated". Hotter colour of expression profile plot represents closer distance to cluster average. Fisher's exact test with Benjamini-Hochberg multiple test correction (FDR < 0.001) for enrichment analysis was based on a background of 92,202 transcripts with TPM > 1 in at least one time point. For larger clusters (4, 5, 7 and 8), unannotated/unknown transcripts were excluded to increase statistical stringency. The list in the enrichment summary is not exhaustive with selected terms sorted according to decreasing statistical significance (Supplementary Material 3).

embryo development ending in seed dormancy. Together with cluster 5, this showed that transcripts involved in embryogenesis were highly regulated during mangosteen seed germination and seedling establishment.

3.5. Overview of DEG expression

To visualise the overall expression pattern of DEGs in the context of seed germination, we mapped the transcript expression results onto the Arabidopsis seed pathway map using MapMan based on the Mercator annotation of DEGs (Fig. 6).

Overall, the expression patterns were consistent with the results described above with the majority of transcripts appeared to be regulated on D3. Many transcripts related to heat and cold stresses were upregulated on D3 compared to the downregulation of redox-related transcripts. Ethylene signalling and its related AP2/EREBP transcription factor (TF) family were upregulated compared to the downregulation of cytokinin signalling on D3. Other TF families, such as C2C2(Zn) DOF, Trihelix, and WRKY were also upregulated on D3. Most cell wall-related transcripts were downregulated on D3 before upregulated on D5. For carbohydrate (CHO) metabolism, starch degradation was upregulated in contrast to the downregulation of sucrose degradation on D3. For lipid metabolism, fatty acid (FA) desaturation and triacylglycerol (TAG) synthesis were upregulated. Conversely, the secondary metabolism and photosynthesis were downregulated. In general, the degradation of other amino acids also appeared to be upregulated along with the downregulation of their synthesis on D3, which is consistent with previous report on their rapid reduction during early seed germination (Mazlan et al., 2019), possibly utilised for protein synthesis or energy supply. For energy metabolism, there was a transient switch to anaerobic respiration on day 3 with the upregulation of the fermentation pathway (Fig. S2). In relation to Perl's and γ-aminobutyric acid (GABA)

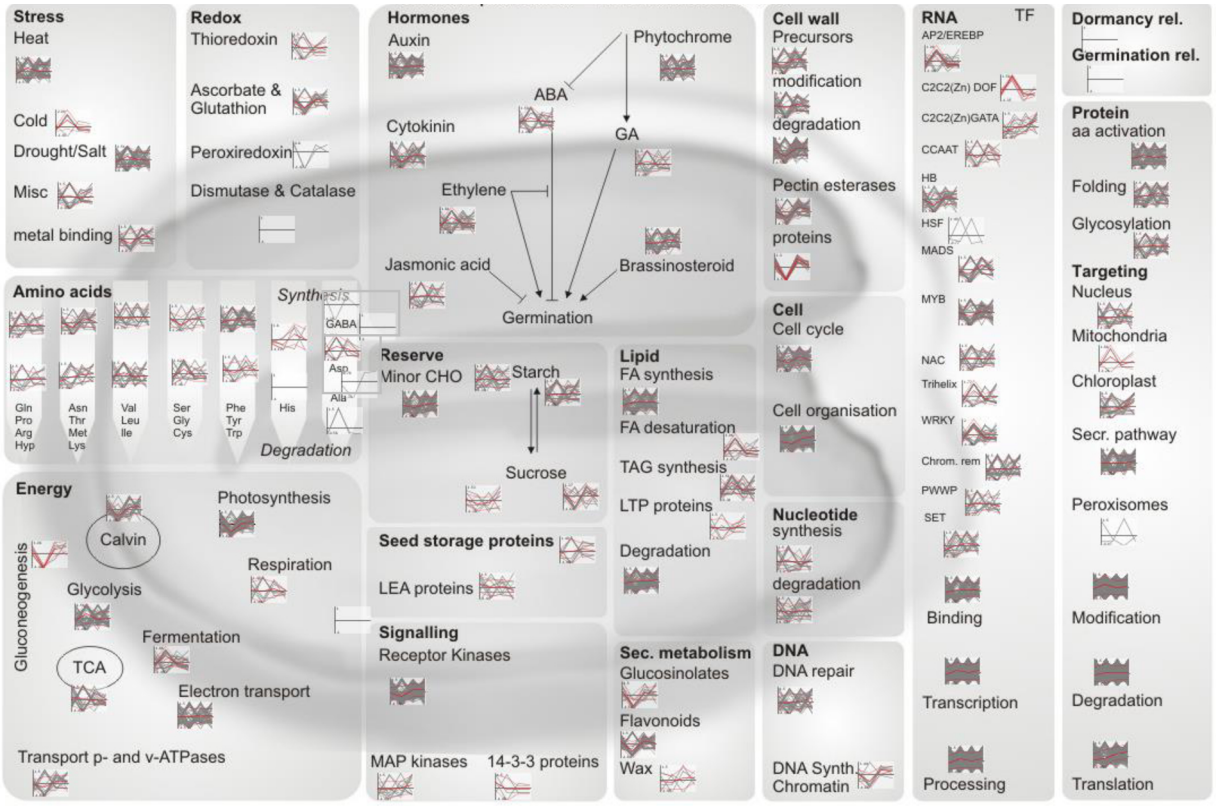

Fig. 6. Overview of DEG expression over the time course of mangosteen seed germination based on Mercator annotated DEG sequences with TPM > 1 in at least one time point. TPM values were *Z*-score normalised and mapped according to Arabidopsis seed pathway map template using MapMan. Red lines showing the median and 95% confidence interval of expression values of all transcript in the same classification.

shunt pathways, some of the aspartate (Asp) and alanine (Ala) aminotransferases (ATs) were upregulated while glutamate decarboxylase (GAD) was downregulated on D3, which indicate the conversion of amino acids into pyruvate for energy generation through anaerobic respiration and suppression of GABA shunt into the tricarboxylic acid (TCA) cycle. Interestingly, while all the mitochondrial isocitrate dehydrogenases (IDHs) showed reduced expression on D3, the cytosolic and peroxisomal IDHs which may play a role in redox homeostasis (Mhamdi et al., 2010) were upregulated (Supplementary Material 4).

3.6. Regulation of metabolism, oxidative stress, and embryogenesis during mangosteen seed germination

Based on the KEGG pathway analysis (Table 2, Supplementary Material 4), starch and sucrose metabolism appeared to be the most highly regulated pathway across all time points. Glycerolipid metabolism was upregulated for 0v3 and 5v7, while glycolysis/gluconeogenesis was upregulated for 3v5. A few pathways of secondary metabolite biosynthesis were downregulated on D3 followed by an upregulation on D5, including the biosynthesis of phenylpropanoid, stilbenoid, diarylheptanoid, and gingerol. These are consistent with the MapMan transcript expression results. Noteworthy, the upregulated 0v3 DEGs were enriched in plant hormone signal transduction, calcium signalling pathway, and iron-dependent programmed cell death (ferroptosis) related to ROS lipid peroxidation.

This led us to examine further the expression of transcripts related to ROS production in the context of hormonal crosstalk with amino acid, polyamine, and glutathione metabolism, which are involved in the regulation of oxidative stress in seed germination (Fig. S3). Overall, the main ROS, hydrogen peroxide which promotes seed germination (Wojtyla et al., 2016), appeared to be induced on D3 with the upregulation of superoxide dismutase coupled by the downregulation of ROS scavengers such as catalase, peroxidase, and peroxiredoxin, despite NADPH oxidases which also produce ROS were downregulated. Furthermore, glutathione metabolism that produces antioxidants for neutralising oxygen radicals was downregulated on D3. Ethylene signalling was activated on D3 perhaps through ROS-induced MAPK signalling with the upregulation of key biosynthesis genes leading to the production of ethylene, namely 1-aminocyclopropane 1-carboxylate (ACC) synthase and ACC oxidase. However, nitric-oxide synthase, which generates nitric oxide that promotes seed germination (Signorelli and Considine, 2018), was downregulated.

On the other hand, polyamine biosynthesis appeared to be upregulated on D3 (Fig. S3). Arginine and ornithine associated with the biosynthesis of polyamines, namely putrescine, spermidine, and spermine, are important regulators of somatic embryogenesis (De Oliveira et al., 2018). This suggests the induction of embryogenesis started upon water imbibition for the development of shoot and radicle. We further profiled all the genes related to embryogenesis based on Arabidopsis studies (Supplementary Material 5, Fig. S4). For DEGs related to embryogenesis (Fig. S4a), upregulated transcripts on D3 were related to the establishment of the root stem cell niche during embryogenesis (*PLT*), embryonic pattern formation (*RPK1*), lateral root initiation (*LBD16*), and BR-dependent growth (*BAK1*). Meanwhile, D3 downregulated transcripts were related to root quiescent centre cell specification and stem cell maintenance (*SHR*, *SCR*), cell differentiation (*ZLL*, *AS1*), vascular development (*ERL1*, *T5L1*), embryo patterning (*ESR2*, *ATHB15*), and BR signalling (*SERK2*). A repressor of embryonic pathways and sugar-inducible seed maturation genes (*VAL2*) was downregulated on D5. Meanwhile, *CUC2*, required for embryonic shoot apical meristem formation, was upregulated on D3; whereas *TPL* related to shoot differentiation and *MP/ARF5* for embryo axis formation and vascular development were upregulated on D7.

For genes related to chromatin modifications (Fig. S4b), transcripts

Table 2
Summary of KEGG pathway enrichment analysis. Bolded font indicates significant over-representation ($P < 0.01$). Numbers showing the values of enrichment factor in heatmap. Hypergeometric tests were performed according to up- and down (dn)-regulated DEGs for each time point comparisons.

	0v3		3v5		5v7	
KEGG Pathway	**up**	**dn**	**up**	**dn**	**up**	**dn**
00500 Starch and sucrose metabolism	**2.0**	**1.8**	**1.9**	**2.3**	**2.0**	1.9
00561 Glycerolipid metabolism	**2.2**	1.1	1.2	1.8	**2.3**	1.3
00010 Glycolysis / Gluconeogenesis	1.2	1.5	**1.6**	1.5	**1.1**	**2.0**
00195 Photosynthesis	0.0	1.5	**1.8**	0.0	0.2	0.3
00196 Photosynthesis - antenna proteins	0.3	**2.8**	**2.2**	0.0	0.5	0.0
00940 Phenylpropanoid biosynthesis	1.0	**2.0**	**2.1**	1.1	0.6	0.6
00941 Flavonoid biosynthesis	0.6	**2.4**	1.9	0.6	0.4	0.0
00945 Stilbenoid, diarylheptanoid and gingerol biosynthesis	0.8	**3.1**	**3.0**	0.9	0.0	0.0
00130 Ubiquinone and other terpenoid-quinone biosynthesis	1.8	**2.1**	1.7	1.7	0.9	1.6
04075 Plant hormone signal transduction	**1.5**	1.1	1.2	1.4	0.9	1.0
04020 Calcium signalling pathway	**2.2**	1.5	2.1	2.1	1.7	1.4
04216 Ferroptosis	**2.5**	1.7	1.5	1.5	2.4	2.4
03010 Ribosome	0.6	0.2	**1.3**	0.6	0.2	0.2
03018 RNA degradation	1.2	0.5	0.8	1.2	**1.8**	1.3
04626 Plant-pathogen interaction	1.4	1.3	1.2	1.4	**1.2**	1.8

encoding DNA methyltransferases (*DM*) for transcriptional activation were downregulated on D3 and highly expressed on D7, compared to the high expression on D5 for some of the DNA (cytosine-5) demethylases (*CMT/DNMT*) for gene silencing. We also studied the expression profile of essential genes based on Arabidopsis SeedGenes (Meinke et al., 2008), 352 out of 481 (73%) were identified to be specific in the embryo development (Muralla et al., 2011). Of the 352 true embryo defective genes (EMB), 324 genes matched to our mangosteen seed transcriptome, in which 310 (96%) transcripts were active (mean TPM > 1) and 120 (37%) were DEGs. This indicates active embryo development during mangosteen seed germination. Three main clusters can be observed based on DEG expression patterns (Fig. S4c). DEGs specifically downregulated on D3 were enriched in peroxisome, reproductive structure, and post-embryonic development; whereas DEGs highly expressed on D3 and D5 were enriched in metabolic regulation, especially amino acid and glycerolipid metabolism (Supplementary Material 5). DEGs highly expressed on D7 were enriched in embryonic axis specification, nitrogen compound metabolic process, and RNA degradation.

3.7. Regulation of hormone-related transcripts

To investigate the role of hormones during mangosteen seed germination, the expression patterns of 4440 transcripts related to different plant hormones known to play role in seed germination were profiled based on the KEGG Orthology (KO) and TAGGIT analysis (Fig. 7, Supplementary Material 6).

Based on the average transcript abundance (Fig. 7a), gibberellin (GA) showed the highest percentage of active transcripts (mean TPM > 1), followed by jasmonic acid (JA), abscisic acid (ABA), brassinosteroid (BR), ethylene (ET), auxin (IAA), and cytokinin (CK). However, ET showed the highest percentage of regulated transcripts (DEG) followed by ABA. Based on time-course analysis of TPM distribution (Fig. 7b), CK, GA, JA, BR, and IAA can be observed with a reduced number of active transcripts (TPM > 1) on D3, opposite to ET and ABA. This was followed by an increase of active transcripts for nearly all hormones, especially CK, GA, JA, and BR at the expense of ET downregulation on D5. On D7, active transcripts continued to greatly increase for GA with a shift to a greater proportion of lower abundance transcripts (1 < TPM < 10). Interestingly, ABA showed a gradual increase of active transcripts throughout the time-course of mangosteen seed germination and seedling establishment.

The trend depicted by the expression of key genes involved in the biosynthesis and signal transduction of hormones (Fig. 7c) was generally reflecting the overall expression pattern of hormone-related transcripts described above and result from hormonometer analysis (Supplementary Material 6). The biosynthesis of CK was highly regulated with 0v3 downregulation and 3v5 upregulation parallel to the changes in active transcripts. The biosynthesis of BR was downregulated on D3 and highly upregulated on D5, which was consistent with the expression of downstream target *Cyclin-D3* (*CYCD3*) for cell division but opposite to the expression of BR receptor kinases (*BAK1* and *BRI1*).

Despite the upregulation of certain genes for IAA and JA biosynthesis (*YUCCA*, *4CL*) and auxin receptor *TIR1*, IAA-amido synthetase *GH3* that inactivating IAA as conjugate and *JAZ* that inhibits JA responses were specifically active on D3. Meanwhile, auxin-responsive protein *IAA* and auxin response factor *ARF* were downregulated with mix regulation for auxin-induced protein *SAUR*. This indicates the repression of IAA and JA responses which is consistent with reduced active transcripts for IAA and JA on D3 (Fig. 7b). On the contrary, the upregulation of transcripts for ethylene biosynthesis (*1-aminocyclopropane-1-carboxylate oxidase*, *ACO*), receptor (*ETR*), and *Ethylene insensitive 3* (*EIN3*) with the downregulation of EIN3 negative regulator (*EBF*) support the activation of ethylene signalling on D3.

For ABA, *Zeaxanthin epoxidase* (*ZEP*) that produces ABA precursor was downregulated on D3 but later upregulated on D7. The downregulation of *ZEP* on D3 was accompanied by the upregulation of *Abscisic acid 8′-hydroxylases* (*CYP707A*) which inactivate ABA. However, ABA receptor (*PYL*) and *Protein phosphatase 2C* (*PP2C*) for ABA signalling were upregulated on D3. Surprisingly for GA, *Gibberellin 3-beta-dioxygenase* (*GA3ox*) which produces bioactive 3′OH-GA was also downregulated, while *Gibberellin 20 oxidase* (*GA20ox*) which produces inactive C19-GA precursor for GA3ox and gibberellin receptor (*GID1*) were upregulated on D3. This indicates increased sensitivity towards ABA and GA signalling despite reduced biosynthesis on D3. Meanwhile, *Gibberellin 2-beta-dioxygenase* (*GA2ox*), which converts bioactive 3′OH-GA to inactive 2′OH-GA, was gradually upregulated after D3 suggesting negative feedback against overall increasing GA-related active transcripts (Fig. 7b).

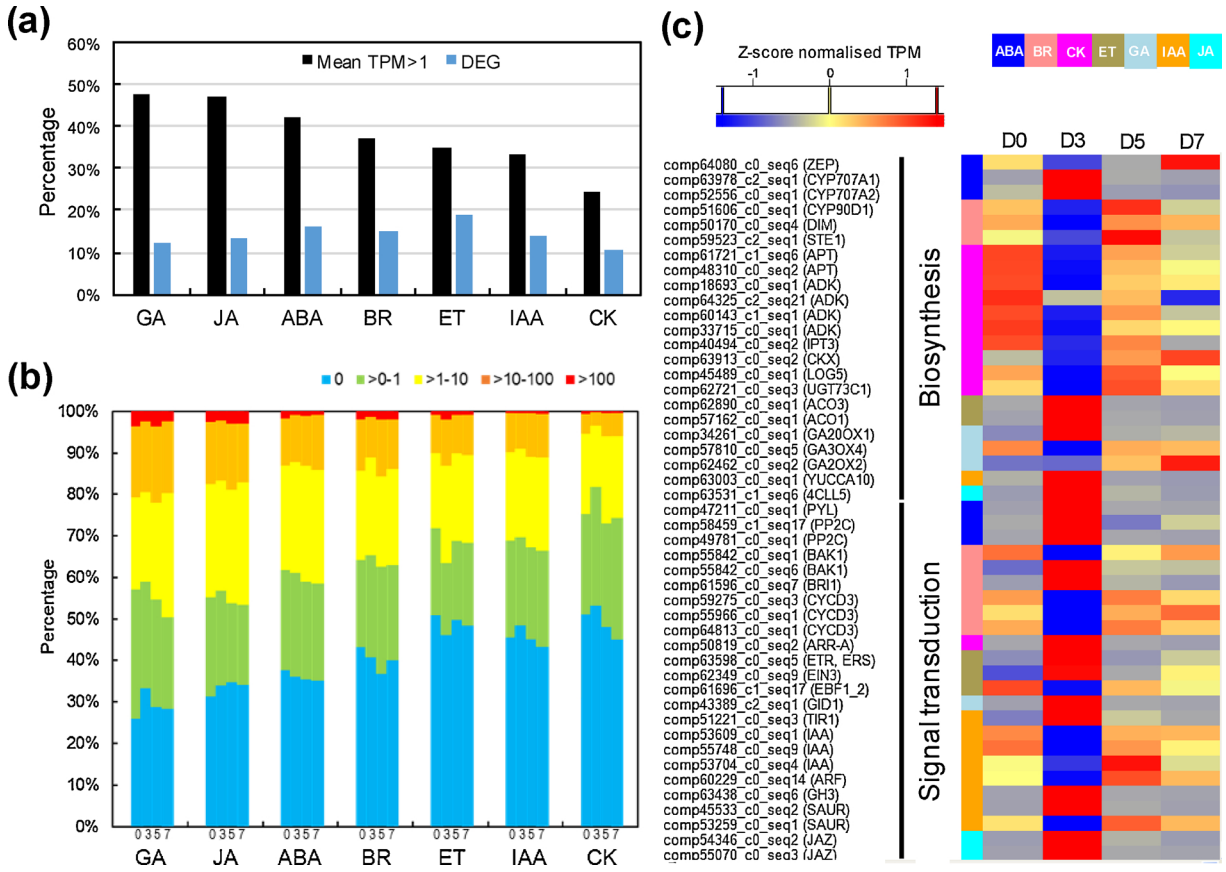

Fig. 7. Expression of hormone-related transcripts. (a) Proportion of transcripts with average TPM values greater than one (TPM > 1) and DEGs relative to total number of transcripts for each hormone. (b) Distribution of transcripts according to TPM values at different time points (D0, D3, D5, and D7). (c) Heatmap of representative DEGs related to hormonal biosynthesis and signal transduction pathways with mean TPM > 1. Colour scale bar represents *Z*-score normalised TPM values. Different hormones are indicated by different colour bars.

3.8. Regulation of transcription factors

Transcription factors (TFs) play an important role in the regulation of gene expression as effectors of hormonal signalling and crosstalk. We classified 11,061 TFs into 70 families of TFs and 24 families of transcription regulators (TRs) in which 1692 (15.3%) are DEGs from mangosteen seed transcriptome (Supplementary Material 7). Fisher's exact test found eight TF families significantly overrepresented, namely bHLH, AP2/ERF-ERF, MYB, WRKY, MADS-MIKC, Trihelix, Tify, and DBP (protein phosphatase 2C) in descending order of abundance (Fig. S5a). HSF (heat shock factor) was only significantly enriched in the overall number of DEG (Fig. S5a), while Jumonji (chromatin regulation), OFP (transcription repressor related to GA), and IWS1 (RNA polymerase II elongation factor) were significantly enriched at specific time point comparisons (Fig. S5b) but not in the overall number of DEG. Furthermore, AP2/ERF-ERF, WRKY, and MYB also accounted for the majority of differentially regulated TFs with high abundance (TPM > 20) (Fig. S5c). These results were consistent with the expression profile enrichment analysis (Fig. 4). Furthermore, some of these overrepresented TF families (bHLH, MYB, and MADS) were previously implicated in cell differentiation, embryogenic patterning, and embryo maturation processes in mangosteen somatic embryogenesis (Mahdavi-Darvari and Noor, 2017), while other TF families such as NAC (meristem maintenance or identity), FAR1 (FAR-RED-IMPAIRED RESPONSE1) and bZIP involved in light and ABA signalling, were also found in high abundance among the DEGs (Fig. S5a). Trihelix TFs which repress seedling establishment (Barr et al., 2012) were downregulated towards D7 of mangosteen seed germination (Fig. S5b).

3.9. RT-qPCR analysis

To validate the transcriptomic analysis, transcripts related to seed germination and embryogenesis (Table S1) were selected to determine their expression levels by RT-qPCR in an independent experiment with three biological replicates (Fig. 8).

Transcription factor *ABA insensitive 5* (*ABI5*), which mediates ABA repression of growth during germination (Lopez-Molina et al., 2002), showed high expression on D0 and decreased during seed germination. Ethylene-responsive transcription factor *ERF105* involved in stress signal transduction pathway showed upregulation on D3 and D7. Three GA biosynthesis genes *Gibberellin 20-oxidase 1* (*GA20ox1*) and *Gibberellin 3-beta-dioxygenases* (*GA3ox1* and *GA3ox4*) were downregulated on D3, in which the latter showed upregulation from D3 onwards. Meanwhile, *Gibberellin 2-beta-dioxygenase 2* (*GA2ox2*) which inactivates GA was gradually upregulated from D0 to D7. *Late embryogenesis abundant protein LEA14-A* with higher expression early seed germination was downregulated on D5. A transcription repressor CCCH-type zinc finger protein *SOMNUS*, which inhibits seed germination by regulating GA and ABA metabolic genes (Dong et al., 2008), was gradually downregulated. Both *Somatic embryogenesis receptor kinase 2* (*SERK2*) and *Brassinosteroid-regulated protein BRU1* related to BR showed similar expression patterns with drastic upregulation on D3 followed by decreasing expression levels. *BRU1* encodes a xyloglucan endotransglucosylase/hydrolase (XTH) which cleaves and relegates xyloglucan polymers for cell wall modification of growing tissue (Campbell and Braam, 1999). Another XTH gene (*XTH9*) for cell elongation (Hyodo et al., 2003) showed increased expression on D5 and D7. *Expansin A1* (*EXPA1*), a cell wall loosening protein, was downregulated

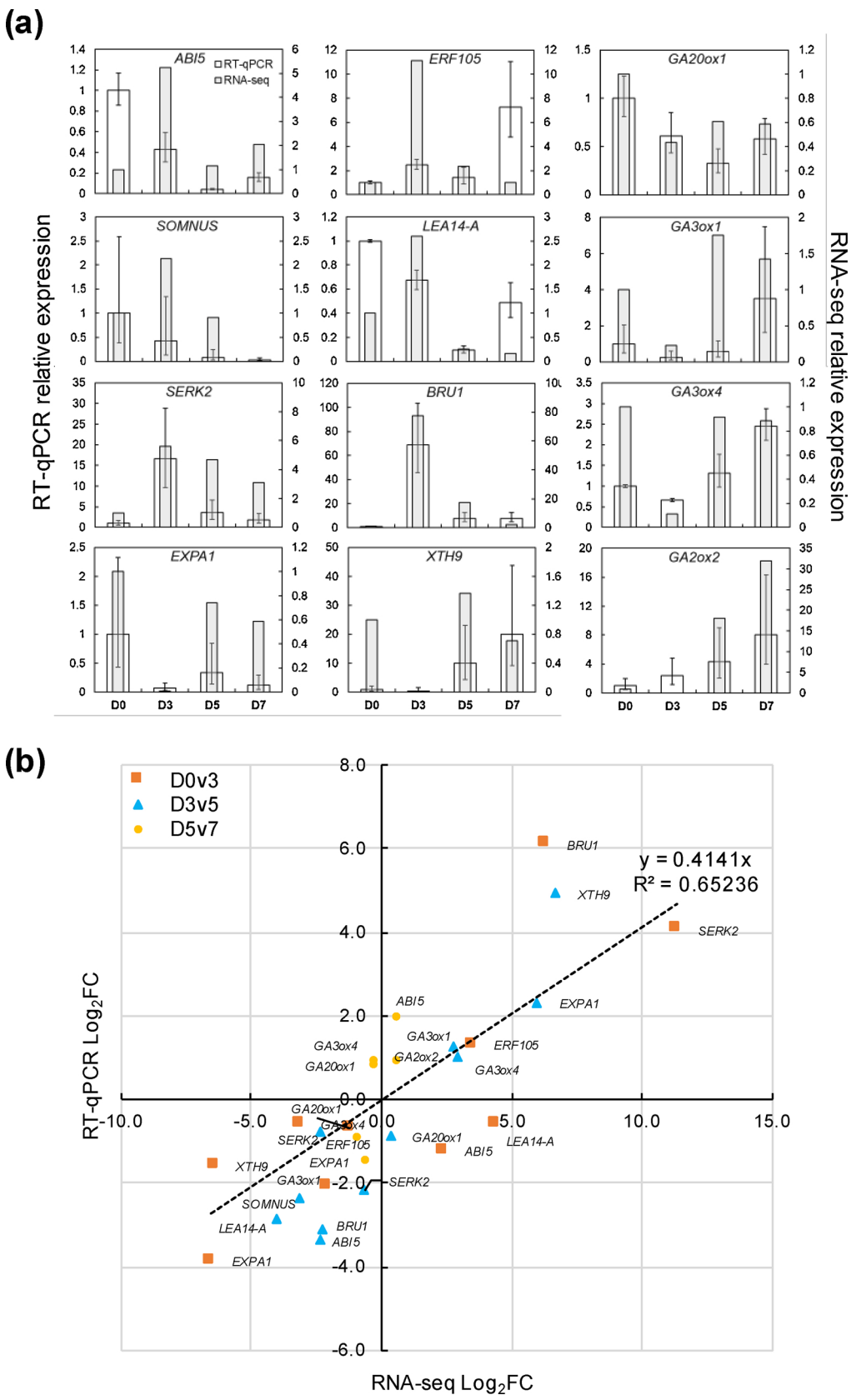

Fig. 8. RT-qPCR analysis. (a) RT-qPCR analysis of selected genes in comparison to RNA-seq analysis based on relative expression with D0 as baseline. Error bars represent standard deviations from three biological replicates. (b) Correlation analysis between RT-qPCR and RNA-seq expression fold-change (Log_2 FC) for comparisons between different time points.

specifically on D3 but highly expressed on other time points.

Apart from some discrepancies between RT-qPCR and RNA-seq relative expressions, such as the higher expression on D7 of *ERF105* and *LEA14-A* (Fig. 8a), overall results based on Log_2 fold-change comparisons showed good correlation ($R^2 > 0.65$) between the expression patterns obtained from both analyses (Fig. 8b). This showed that transcriptomic analysis is reliable and reproducible despite based on only a single sample, considering that the RT-qPCR experiment was performed using an independent set of three biological replicates.

3.10. Comparative DEG network analysis with Arabidopsis

For comparison of global gene expression changes during mangosteen seed germination with that of an orthodox seed model, we mapped differentially expressed transcripts (DEGs) at different time points against the Arabidopsis co-expression gene network, SeedNet (Fig. 9).

The closest resemblance of mangosteen DEGs with that of the Arabidopsis gene clustering pattern (Fig. 9a) is between D3 and D5 comparison (Fig. 9c), with more upregulated genes related to germination and downregulated genes related to dormancy. From mapped 2,368 of 8,621 nodes in SeedNet, the upregulated genes related to germination were significantly overrepresented in 0v3_dn and 3v5_up with genes upregulated by GA (Supplementary material 8). Conversely, the genes upregulated by ABA were overrepresented in 0v3_up and 3v5_dn, which is consistent with the antagonistic relationship between GA and ABA. Additionally, genes upregulated by ABA were overrepresented in both 5v7_up and 5v7_dn, reflecting the role of ABA signalling pathway throughout mangosteen seed germination and seedling establishment.

4. Discussion

4.1. Molecular physiology at the beginning of mangosteen seed germination

At the end of fruit development, mature mangosteen seeds, which are considerably large (~2 cm), contain high (61%) moisture content (Noor et al., 2016) compared to that of <5% in typical orthodox Arabidopsis seeds (Weitbrecht et al., 2011). High moisture content and nutrient reserves are common characteristics of recalcitrant seeds in tropical tree species proposed to be an ancestral adaptation to conducive humid tropics for non-dormant seeds to germinate rapidly after shedding (Barbedo et al., 2013). This contrasts to orthodox seeds with dormancy in many temperate species which have evolved to survive seasonal variation for germination only under favourable environments (Pammenter and Berjak, 2000). However, Arabidopsis germination *sensu stricto* only takes around 20–55 h (Weitbrecht et al., 2011) compared to 3–5 days for mangosteen.

To understand the mechanism of recalcitrant seed germination, we

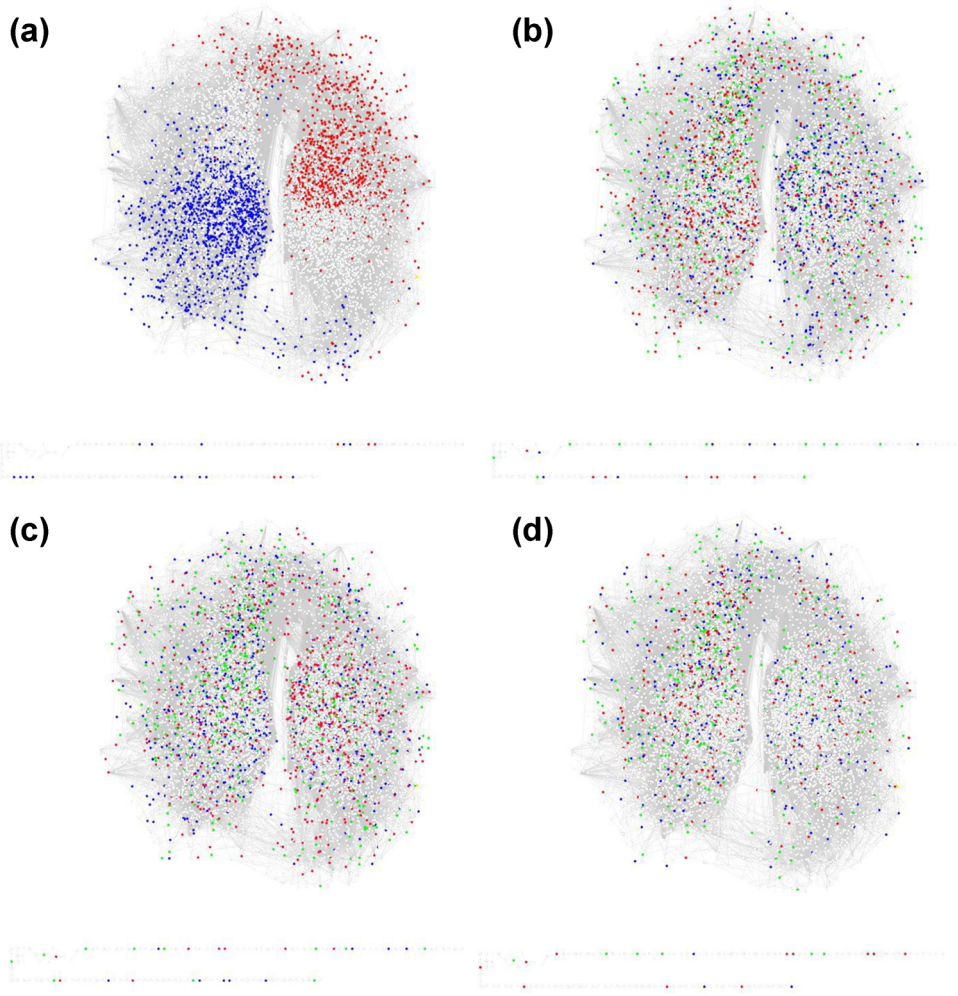

Fig. 9. Comparative DEG network analysis with Arabidopsis SeedNet. (a) Arabidopsis SeedNet. Red: upregulated during germination; blue: upregulated during dormancy. (b) D0vsD3 (c) D3vsD5 (d) D5vsD7. Red: upregulated; blue: downregulated; green: mix regulation.

can relate to the acquisition of desiccation tolerance during orthodox seed development, which mainly involves three phases of histodifferentiation, reserve accumulation, and maturation. During histodifferentiation with endosperm and embryo formation, orthodox seeds are very sensitive to desiccation, which is analogous to mature recalcitrant seeds. Therefore, it has been proposed that the recalcitrant and orthodox seeds form a continuum such that orthodox seeds have evolved to complete the full developmental programme with desiccation while recalcitrant seeds are halted at different stages during seed development resulting in intermediate seeds with different levels of sensitivity to desiccation (Barbedo et al., 2013). The acquisition of desiccation tolerance begins with the accumulation of reserves, insoluble compounds, non-reducing sugars/oligosaccharides (*e.g.*, sucrose, raffinose), late embryogenic abundant proteins (LEAs), and heat shock proteins (HSPs) to confer mechanical support for cellular integrity during desiccation (Berjak and Pammenter, 2013). It is assumed that such a protective mechanism is absent or insufficient for recalcitrant species to repair damages of cellular constituents from drying, thus producing desiccation intolerant seeds with high moisture content.

Despite the absence of a desiccation phase, many transcripts important for protective mechanisms were found to be expressed at very high levels (mean TPM > 1000) (Fig. S6, Supplementary Material 9). These include stress-related protein, glutathione S-transferase (*GST*), catalase (*CAT*), dehydration-responsive/induces protein, abscisic stress-ripening protein in response to water deprivation, ethylene-responsive TF *RAP2-2* important for hypoxia survival *via* N-end rule pathway (Hinz et al., 2010; Licausi et al., 2011), MLP-like protein for defence, and thiamine thiazole synthase for stress adaptation to DNA damage (Machado et al., 1996). Transcripts present with TPM > 1500 throughout all time points include 2S sulfur-rich seed storage protein for nutrient reservoir, glyceraldehyde-3-phosphate dehydrogenase, malate synthase and fructose-bisphosphate aldolase 2 for respiration, auxin-repressed protein (homolog of Arabidopsis dormancy-associated protein 1), glycine-rich RNA-binding protein for RNA processing during stress (Kwak et al., 2011), chaperone protein DnaJ, major allergen Mal d in plant defence, polyubiquitin, protein translation factor, and metallothionein-like protein that binds heavy metals, especially copper important for seed development (Benatti et al., 2014). Other highly expressed transcripts on D0 include aquaporin *TIP1* essential for water and hydrogen peroxide transport (Bienert et al., 2007), peptidyl-prolyl cis-trans isomerase that accelerates protein folding, and 3-ketoacyl-CoA thiolase for long-chain fatty-acid beta-oxidation before gluconeogenesis during germination (Germain et al., 2001) or JA biosynthesis (Afitlhile et al., 2005). Despite the high expression of many stress factors since D0, transcripts for *HSP*, *LEA* (also *ARG2*), and oleosin (Miquel et al., 2014), which are important for desiccation tolerance in orthodox seeds, were only greatly induced on D3 (Fig. S6) with TPM > 8000. As such, D0 to D3 gene expression pattern was akin to the late maturation phase of Arabidopsis seed. Meanwhile, a blue copper protein, *PAR1* for post-translational modification of histidine, a lipid transfer protein, an isocitrate lyase for storage lipid mobilisation through glyoxylate cycle, a laccase for lignin synthesis, and *XTH9* (described above) highly expressed on D0 were specifically downregulated on D3. Transcripts with high expressions on D0, which represent stored transcripts at the end of seed maturation, are not necessarily translated into proteins and might be subjected to post-transcriptional regulation. Nonetheless, the translation of stored transcripts is critical for seed germination compared to *de novo* transcription (Rajjou et al., 2004).

Despite the presence of many desiccation stress-related transcripts at high abundance, mangosteen seeds do not undergo desiccation during the maturation phase. This shows that seed desiccation requires concerted actions of many processes at the correct timing beyond simple accumulation of stress factors with independent regulation of metabolite reserves. It can also be inferred that a species-specific genetic checkpoint possibly exists during seed development that determines whether a seed enters desiccation during the maturation phase. A summary of active biological processes based on top abundance transcripts is shown in Fig. S7. The overall expression pattern of highly abundant transcripts suggests active metabolism and defence mechanism from the start of mangosteen seed germination.

4.2. Mangosteen seed metabolism and hormonal regulation during germination

Mature seeds of mangosteen readily germinate upon imbibition without a phase of dormancy breaking/release as required by many orthodox seeds. This suggests that the seed metabolism is maintained at an optimal level to allow rapid germination, which is supported by highly abundant transcripts related to primary metabolism since D0, especially those related to energy production (Fig. S2). Hence, while metabolism reactivation is required for metabolically quiescent Arabidopsis seed germination (Weitbrecht et al., 2011), mature mangosteen seed metabolism might remain relatively active. Most of these transcripts, including glycolysis, TCA, oxidative pentose phosphate, glycerol shunt, and GABA shunt pathways, were transiently downregulated on D3 when alcohol fermentation became dominant before upregulated again on D5, perhaps due to oxygen deprivation during early seed imbibition with an active metabolism. This coincided with the upregulation of ROS and ethylene signalling (Fig. S3), known to promote seed germination (Wojtyla et al., 2016), which might have triggered the onset of transcriptional reprogramming as evident by the drastic D3 to D5 transition. On the other hand, COMATOSE (CTS), a peroxisomal ATP-binding cassette (ABC) transporter important for β-oxidation of storage lipids during gluconeogenesis and biosynthesis of IAA, JA and flavonoids, as well as a germination potential gene for radicle emergence (Carrera et al., 2007), was maintained at high levels (TPM > 20) from D0 to D7 (Fig. S2).

Energy production in mangosteen seed is greatly dependent on the starch and lipid reserves (Noor et al., 2016) as supported by the highly regulated starch and sucrose metabolism and glycerolipid metabolism throughout seed germination and seedling establishment (Figs. 3, 4, and Fig. 6, Table 2). In maturing legume seeds, embryo photosynthesis *via* photoheterotrophic plastids contributes to oxygen supply and metabolic assimilation of CO_2 released during reserve accumulation (Rolletschek et al., 2003). Despite the absence of a differentiated embryo in mangosteen seed, photosynthesis/chloroplast-related transcripts were found to be downregulated on D3 before upregulated again on D5 (Fig. 4, Table 2), which might be responsible for the shift from anaerobic back to aerobic respiration to cater for greater energy demand during embryogenesis and seedling establishment.

Apart from starch and lipid, protein reserves are also very important for hydrolytic activities and energy production through pyruvate conversion from amino acids, such as alanine, glycine, serine, threonine, and aspartate, derived from protein degradation/proteolysis. Based on the opposite trend of compositional changes in amino acids with that of Arabidopsis (Fait et al., 2006), it was proposed that recalcitrant seeds, such as mangosteen, employ a different mechanism for energy supply and amino acid metabolism during seed germination (Mazlan et al., 2019). Furthermore, proteolysis appeared to be a tightly regulated process during mangosteen seed germination as indicated by varied expression patterns of transcripts related to proteolysis and its inhibition (Fig. 5). This remained to be explored further on the type of protein turnover at each time point.

On the other hand, secondary metabolism also plays an important role during seed germination by protecting against pathogens and oxidative stress. Antioxidant compounds such as glutathione, tocopherol, flavonoids, and xanthones (Mazlan et al., 2019) may scavenge ROS from reoxygenating of germinating tissues to protect against damages of DNA, proteins, and membrane. Indeed, Arabidopsis mutants with impaired vitamin E and flavonoid biosynthesis exhibit severe growth defects or inhibited germination (Carrera et al., 2007; Sattler et al., 2004). Transcripts related to their biosynthesis were found to be

downregulated on D3 before upregulated and maintained on D5 onwards (Fig. 5, Table 2), opposite to that of ROS-related transcripts (Fig. S3). Other DNA repair mechanisms involving DNA ligase (Waterworth et al., 2010) or protein repair by protein-L-isoaspartate O-methyltransferase (Ogé et al., 2008) also appeared to be important during mangosteen seed germination (Fig. 5, Supplementary Material 3 & 9).

Phytohormones are important for signal transduction in response to stresses as well as regulating seed germination and seedling establishment. ABA is known to maintain seed dormancy and delay germination, while GA breaks dormancy and induces germination in Arabidopsis (Weitbrecht et al., 2011). Hence, the ABA:GA ratio is a central determinant of orthodox seed germination. Contrasting to the high level of ABA in mature Arabidopsis seeds, which is downregulated during seed germination, ABA was enriched in D3 upregulated DEGs (Fig. 5) with a trend of increasing active transcripts (Fig. 7) during mangosteen seed germination. Conversely, GA showed the highest proportion of active transcripts among all phytohormones with a drastic dip on D3 (Fig. 7). This indicates active transcripts related to GA that provide germination potential were suspended on D3 before the onset of proper germination between D3 and D5. It appeared that the antagonistic relationship between ABA and GA is conserved in mangosteen as evident from the opposite effect on the global expression pattern of Arabidopsis genes (Fig. 9). However, the role of ABA during mangosteen seed germination is more likely to confer stress response in the absence of dormancy. Additionally, transcripts in response to salicylic acid (SA) were found to be induced on D3 similarly to ABA followed by SA catabolism on D5 (Fig. 5) and possibly play roles in protection against pathogens or modulating oxidative stress (Fig. S3) (Lee et al., 2010).

IAA, CK, BR, and GA were found to be enriched for transcripts upregulated on D5 (Fig. 5) which coincided with radicle protrusion. This is consistent with their roles in cell proliferation, elongation, and differentiation for seedling growth and development, alongside nutrient mobilisation (phloem sucrose loading), cell-wall modification, and chloroplast biogenesis (Fig. 5). Among hormone-related TFs, AUX/IAA and B3-ARF families showed the most diverse expression pattern (Supplementary Material 6) indicating the complex role of IAA in mangosteen seed germination including vascular patterning as suggested during somatic embryogenesis (Mahdavi-Darvari and Noor, 2017). Meanwhile, cell division is likely to be regulated by BR signalling through the induction of cycle cell gene *CYCD3* with specific downregulation on D3 prior to D5 upregulation (Fig. 7), which shared similar pattern to a CRE family receptor *AHK4/WOL* (*Wooden leg*) (Supplementary Material 6) involved in CK signalling for procambial maintenance and/or proliferation (Mähönen et al., 2006). Additionally, TF family OFP which is a repressor of GA biosynthesis (Wang et al., 2016) was found to be enriched in 0v3_dn DEGs before the upregulation of GA response on D5 (Fig. S5).

Transcripts related to JA and ET were both induced on D3 (Fig. 7) with their respective TF family of ERF and Tify, which shared similar expression patterns with WRKY and DBP (protein phosphatase 2C) (Fig. S5) that are key to ABA signalling in stress response (Rushton et al., 2012). It is unclear how JA regulates germination, but the germination-inhibiting effects of JA have been implicated in contradictory relationships with ABA (Linkies and Leubner-Metzger, 2012). There was no clear enrichment of JA among DEGs at different time points (Fig. 4) which suggests minimal role played by JA during mangosteen seed germination. The role of ET will be further discussed below.

In addition to phytohormones, the involvement of other chemical compounds such as NO and ROS was also inferred through transcriptomic analysis (Fig. S3), which indicates the induction of H_2O_2 on D3 during mangosteen seed germination. Apart from well-known plant growth regulators, we also discovered the possible involvement of smoke-derived karrikin (KAR) with a similar expression pattern to transcripts upregulated on D5 in response to GA, CK biosynthesis, and SA catabolism (Fig. 5). KAR has been shown to stimulate germination in Arabidopsis (Waters and Smith, 2013) and maize (Soós et al., 2010) but delays soybean germination under shaded condition (Meng et al., 2016), possibly through crosstalk with other phytohormones, especially ABA, GA, and IAA. However, the actual involvement, role, and mechanism of KAR in recalcitrant seeds are yet to be elucidated.

4.3. Embryogenesis regulation during mangosteen seed germination

Mangosteen seed germination characteristic is unique because it involves *in vivo* asexual direct somatic embryogenesis in mature seed upon fruit shedding (Noor et al., 2016). In typical seeds, embryogenesis occurs after the double fertilisation resulting in endosperm and a zygotic embryo that enters a quiescent state at the end of seed maturation. Zygotic embryos typically accumulate a higher abundance of storage proteins and stress-related HSPs during seed maturation in preparation for dehydration. In comparison, somatic embryos have higher levels of proteins involved in energy production (glycolysis and TCA) which suggests a greater energy demand (Winkelmann, 2016). Sucrose synthase (*SUS*) which showed a dip on D3 (Fig. S2) is thought to be important for the transition between storage and metabolic sink in the embryo (Konrádová et al., 2002).

In mature mangosteen seeds, a differentiated embryo is absent with only a procambium ring structure. Based on histological observations, the procambium ring with two to three cell layers elongates towards the opposite ends where radicle and plumule emerge (Noor et al., 2016). Hence, it is conceivable that procambium cells divide rapidly resulting in elongating procambial ring and develop into vascular tissues for water transport and nutrient mobilisation during early seed germination. Along with the downregulation of cell-wall modification transcripts, cationic peroxidases which showed mix-regulation on D3 (Fig. S3) might play a role in preventing cell expansion during rapid cell division like early zygotic embryogenesis (Hoenemann et al., 2010). Notably, TF family bHLH involved in the development of vascular tissues and root was highly upregulated on D5 (Fig. S5), which coincided with radicle protrusion.

A high level of GABA on D0 (Mazlan et al., 2019) might play a role together with calcium signalling in plant organogenesis (Michaeli and Fromm, 2015) and cellular differentiation (Santa-Catarina et al., 2006), as reported previously in mangosteen ovule protuberance outgrowth (Yapwattanaphun et al., 2014; Yonemori et al., 2014). However, GABA content reduces rapidly during seed germination (Mazlan et al., 2019), which is consistent with *GAD* downregulation on D3 (Fig. S2). However, GAD upregulation on D5 was not followed by an increase of GABA which indicates rapid GABA shunt into the TCA cycle for energy production. Other factors important for somatic embryogenesis include polyamines (De Oliveira et al., 2018) with an indication of upregulation on D3, especially spermidine synthase. This is consistent with the higher expression of polyamine biosynthetic pathway transcripts in mangosteen embryogenic callus (Fadryin et al., 2018; Mahdavi-Darvari and Noor, 2017). A cytosolic glutamine synthetase localised to procambial tissue, which is a marker for the onset of germination with chloroplast differentiation found to be absent in zygotic but present in somatic pine embryos (Rodríguez et al., 2006), was expressed at a high level (TPM > 45) throughout mangosteen seed germination (Supplementary Material 4).

Stress is an important trigger of somatic embryogenesis at the right level and developmental time frame for the acquisition of embryogenic competence by inducing severe changes in cellular metabolism which involved ABA (Fehér, 2015) and ET especially for *in vitro* culture of recalcitrant species (Piyatrakul et al., 2012). During maturation, ABA content declines in zygotic embryos but increases in somatic embryos (Winkelmann, 2016), which is akin to the gradual increase of active ABA-related transcripts during mangosteen seed germination (Fig. 7). ABA treatment has been shown to increase storage protein accumulation in cocoa somatic embryos (Alemanno et al., 1997). Meanwhile,

endogenous ET was reported to be essential for the proliferation and differentiation of somatic embryogenesis but not its induction in *Medicago sativa* (Kępczyńska et al., 2009). Therefore, ET signalling through ERF TFs on D3 could play a critical role in the activation of *in vivo* somatic embryogenesis in mangosteen seed. Whether or not the application or suppression of ET will affect mangosteen seed germination rate is yet to be tested in the future.

Excessive stress such as under *in vitro* conditions might lead to cell death while programmed cell death (PCD) could be a part of a normal developmental programme. Iron-dependent PCD/ferroptosis appeared to be important during early mangosteen embryogenesis (Table 2). In *Cyclamen persicum*, non-embryogenic cells surrounding the differentiating somatic embryos undergo PCD (Schwenkel and Winkelmann, 1998). Moreover, epigenetic regulation of embryogenesis by methylation/demethylation and histone modifications in mangosteen seed germination, especially after radicle protrusion on D5 (Fig. S4b) will need further investigation.

5. Conclusion

In summary, transcriptomic analysis supported that mangosteen seed metabolism is maintained at an optimal level, continuous from the seed maturation stage, which is typical of a recalcitrant seed metabolism with starch and lipid as the main reserves. Based on comparative analysis with Arabidopsis, transcriptional changes showed the conserved antagonistic relationship between ABA and GA in mangosteen seeds. This study also revealed mangosteen seed germination as a fascinating process encompassing a developmental pathway of *in vivo* direct somatic embryogenesis/organogenesis through transcriptional reprogramming of a common set of genes with the activation of ethylene signalling pathway concurring anaerobic respiration on D3. As summarised in the descriptive Fig. S7, mangosteen seed germination is a complex process that involved concerted actions of many transcriptional and hormonal regulatory factors through concomitant coordination of metabolic and cellular processes with signal transduction. This first-time transcriptomic analysis of recalcitrant seed germination serves as a starting point for a more in-depth analysis of the many postulations generated from this study. Proteomic profiling will be necessary to ascertain protein abundance with transcript expression while the quantification of phytohormones will be meaningful to understand the role of each plant growth regulators and their effects on mangosteen seed germination. It will also be interesting to perform comparative transcriptomic analysis of other recalcitrant seeds with differentiated/developed embryos in the future. The ability to improve seed germination has great agronomic potential and is important for the propagation of recalcitrant species.

Author contributions

Conceived, designed and performed the experiments: SAB, NDKA, HHG. Analysed the data: SAB, NDKA, HHG. Contributed reagents/materials/analysis tools: HHG, ZZ, NMN. Wrote the manuscript: SAB, HHG. Review the manuscript: ZZ, NMN.

Competing interests

The authors declare no competing interests.

Acknowledgments

SAB was supported by NMN through Universiti Kebangsaan Malaysia Research University Grant AP-2012-018. HHG was funded by DIP-2014-008 and GUP-2017-057. We thank technical support and advice from Kok-Keong Loke, and Mohd Noor Mat Isa from Malaysia Genome Institute for access to the computational facility.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.scienta.2019.108727.

References

Abdul-Rahman, A., Goh, H.H., Loke, K.K., Noor, N.M., Aizat, W.M., 2017a. RNA-seq analysis of mangosteen (*Garcinia mangostana* L.) fruit ripening. Genomics Data 12, 159–160.
Abdul-Rahman, A., Suleman, N.I., Zakaria, W.A., Goh, H.H., Noor, N.M., Aizat, W.M., 2017b. RNA extractions of mangosteen (*Garcinia mangostana* L.) pericarps for sequencing. Sains Malaysiana 46, 1231–1240.
Abu Bakar, S., Kumar, S., Loke, K.K., Goh, H.H., Mohd Noor, N., 2017. DNA shotgun sequencing analysis of *Garcinia mangostana* L. variety Mesta. Genomics Data 12, 118–119.
Abu Bakar, S., Sampathrajan, S., Loke, K.K., Goh, H.H., Mohd Noor, N., 2016. DNA-seq analysis of Garcinia mangostana. Genomics Data 7, 62–63.
Afitlhile, M.M., Fukushige, H., Nishimura, M., Hildebrand, D.F., 2005. A defect in glyoxysomal fatty acid β-oxidation reduces jasmonic acid accumulation in Arabidopsis. Plant Physiol. Biochem. 43, 603–609.
Alemanno, L., Berthouly, M., Michaux-Ferriere, N., 1997. A comparison between *Theobroma cacao* L. Zygotic embryogenesis and somatic embryogenesis from floral explants. In Vitro Cell Develop Biol - Plant 33, 163–172.
Asomaning, J.M., Olympio, N.S., Sacande, M., 2011. Desiccation sensitivity and germination of recalcitrant Garcinia kola heckel seeds. Res. J. Seed Sci. 4, 15–27.
Azlan, N.D.K., Isa, M.N.M., Zainal, Z., 2017. RNA-seq assembly and analysis of *Garcinia mangostana* transcriptome during seed germination. Data Brief 14, 548–550.
Barbedo, C.J., Centeno, D., Ribeiro, d.C., R.d.C.L.F, 2013. Do recalcitrant seeds really exist? Hoehnea 40, 583–593.
Barr, M.S., Willmann, M.R., Jenik, P.D., 2012. Is there a role for trihelix transcription factors in embryo maturation? Plant Signal. Behav. 7, 205–209.
Bassel, G.W., Lan, H., Glaab, E., Gibbs, D.J., Gerjets, T., Krasnogor, N., Bonner, A.J., Holdsworth, M.J., Provart, N.J., 2011. Genome-wide network model capturing seed germination reveals coordinated regulation of plant cellular phase transitions. Proc. Natl. Acad. Sci. U. S. A. 108, 9709–9714.
Benatti, M.R., Yookongkaew, N., Meetam, M., Guo, W.J., Punyasuk, N., Abuqamar, S., Goldsbrough, P., 2014. Metallothionein deficiency impacts copper accumulation and redistribution in leaves and seeds of Arabidopsis. New Phytologist 202, 940–951.
Berjak, P., Pammenter, N.W., 2013. Implications of the lack of desiccation tolerance in recalcitrant seeds. Front. Plant Sci. 4.
Bienert, G.P., Møller, A.L.B., Kristiansen, K.A., Schulz, A., Møller, I.M., Schjoerring, J.K., Jahn, T.P., 2007. Specific aquaporins facilitate the diffusion of hydrogen peroxide across membranes. J. Biol. Chem. 282, 1183–1192.
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.
Bryant, D.M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M.B., Payzin-Dogru, D., Lee, T.J., Leigh, N.D., Kuo, T.H., Davis, F.G., Bateman, J., Bryant, S., Guzikowski, A.R., Tsai, S.L., Coyne, S., Ye, W.W., Freeman, R.M., Peshkin, L., Tabin, C.J., Regev, A., Haas, B.J., Whited, J.L., 2017. A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. Cell Reports 18, 762–776.
Campbell, P., Braam, J., 1999. In vitro activities of four xyloglucan endotransglycosylases from Arabidopsis. Plant J. 18, 371–382.
Carrera, E., Holman, T., Medhurst, A., Peer, W., Schmuths, H., Footitt, S., Theodoulou, F.L., Holdsworth, M.J., 2007. Gene expression profiling reveals defined functions of the ATP-binding cassette transporter COMATOSE late in phase II of germination. Plant Physiol. 143, 1669–1679.
Chandra, J., Keshavkant, S., 2018. Desiccation-induced ROS accumulation and lipid catabolism in recalcitrant Madhuca latifolia seeds. Physiol. Mol. Biol. Plants 24, 75–87.
De Oliveira, L.F., Navarro, B.V., Cerruti, G.V., Elbl, P., Minocha, R., Minocha, S.C., Dos Santos, A.L.W., Floh, E.I.S., 2018. Polyamine- and amino acid-related metabolism: the roles of arginine and ornithine are associated with the embryogenic potential. Plant Cell Physiol. 59, 1084–1098.
Di Stefano, J.F., Marín, W.A., Díaz, M.A., 2006. Secondary embryonary root associated with seed germination in Garcinia intermedia (Clusiaceae) and its possible role in seedling survival. Revista de Biologia Tropical 54, 927–934.
Dong, H.K., Yamaguchi, S., Lim, S., Oh, E., Park, J., Hanada, A., Kamiya, Y., Choi, G., 2008. SOMNUS, a CCCH-type zinc finger protein in Arabidopsis, negatively regulates light-dependent seed germination downstream of PIL5. Plant Cell 20, 1260–1277.
Elviana, M., Rohani, E.R., Ismanizan, I., Normah, M.N., 2011. Morphological and histological changes during the somatic embryogenesis of mangosteen. Biol. Plant. 55, 731–736.
Fadryin, N., Rohani, E.R., Muhammed-Hussein, Z.A., Noor, N.M., 2018. Somatic embryogenesis-related gene expression and functional genomics in mangosteen. Plant Gene 15, 51–66.
Fait, A., Angelovici, R., Less, H., Ohad, I., Urbanczyk-Wochniak, E., Fernie, A.R., Galili, G., 2006. Arabidopsis seed development and germination is associated with temporally distinct metabolic switches. Plant Physiol. 142, 839–854.
Fehér, A., 2015. Somatic embryogenesis - stress-induced remodeling of plant cell fate. Biochim. Biophys. Acta - Gene Regulatory Mechanisms 1849, 385–402.
Germain, V., Rylott, E.L., Larson, T.R., Sherson, S.M., Bechtold, N., Carde, J.P., Bryce, J.H., Graham, I.A., Smith, S.M., 2001. Requirement for 3-ketoacyl-CoA thiolase-2 in peroxisome development, fatty acid β-oxidation and breakdown of triacylglycerol in

lipid bodies of Arabidopsis seedlings. Plant J. 28, 1–12.
Goh, C.J., Lakshmanan, P., Loh, C.S., 1994. High frequency direct shoot bud regeneration from excised leaves of mangosteen (*Garcinia mangostana* L.). Plant Sci. 101, 173–180.
Goh, H.K.L., Rao, A.N., Loh, C.S., 1988. In vitro plantlet formation in mangosteen (*Garcinia mangostana* L.). Ann. Bot. 62, 87–93.
Gutierrez-Orozco, F., Failla, M.L., 2013. Biological activities and bioavailability of mangosteen xanthones: A critical review of the current evidence. Nutrients 5, 3163–3183.
Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 8, 1494–1512.
Hinz, M., Wilson, I.W., Yang, J., Buerstenbinder, K., Llewellyn, D., Dennis, E.S., Sauter, M., Dolferus, R., 2010. Arabidopsis RAP2.2: An ethylene response transcription factor that is important for hypoxia survival. Plant Physiol. 153, 757–772.
Hoenemann, C., Richardt, S., Krüger, K., Zimmer, A.D., Hohe, A., Rensing, S.A., 2010. Large impact of the apoplast on somatic embryogenesis in Cyclamen persicum offers possibilities for improved developmental control in vitro. BMC Plant Biol. 10.
Huang, L.C., Huang, B.L., Wang, C.H., Kuo, C.I., Murashige, T., 2000. Developing an improved in vitro propagation system for slow-growing species using *Garcinia mangostana* L. (Mangosteen). In Vitro Cell. Develop. Biol. - Plant 36, 501–504.
Hyodo, H., Yamakawa, S., Takeda, Y., Tsuduki, M., Yokota, A., Nishitani, K., Kohchi, T., 2003. Active gene expression of a xyloglucan endotransglucosylase/hydrolase gene, XTH9, in inflorescence apices is related to cell elongation in *Arabidopsis thaliana*. Plant Mol. Biol. 52, 473–482.
Ilias, I.A., Negishi, K., Yasue, K., Jomura, N., Morohashi, K., Baharum, S.N., Goh, H.H., 2018. Transcriptome-wide effects of expansin gene manipulation in etiolated Arabidopsis seedling. J. Plant Res.
Joshi, G., Arun Kumar, A.N., Gowda, B., Srinivasa, Y.B., 2006. Production of supernumerary plants from seed fragments in Garcinia gummi-gutta: Evolutionary implications of mammalian frugivory. Curr. Sci. 91, 372–376.
Kępczyńska, E., Ruduś, I., Kępczyński, J., 2009. Endogenous ethylene in indirect somatic embryogenesis of *Medicago sativa* L. Plant Growth Regul. 59, 63–73.
Konrádová, H., Lipavská, H., Albrechtová, J., Vreugdenhil, D., 2002. Sucrose metabolism during somatic and zygotic embryogeneses in Norway spruce: Content of soluble saccharides and localisation of key enzyme activities. J. Plant Physiol. 159, 387–396.
Kwak, K.J., Park, S.J., Han, J.H., Kim, M.K., Oh, S.H., Han, Y.S., Kang, H., 2011. Structural determinants crucial to the RNA chaperone activity of glycine-rich RNA-binding proteins 4 and 7 in *Arabidopsis thaliana* during the cold adaptation process. J. Exp. Bot. 62, 4003–4011.
Lee, S., Kim, S.G., Park, C.M., 2010. Salicylic acid promotes seed germination under high salinity by modulating antioxidant activity in Arabidopsis. New Phytologist 188, 626–637.
Li, B., Dewey, C.N., 2011. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinform. 12.
Licausi, F., Kosmacz, M., Weits, D.A., Giuntoli, B., Giorgi, F.M., Voesenek, L.A.C.J., Perata, P., Van Dongen, J.T., 2011. Oxygen sensing in plants is mediated by an N-end rule pathway for protein destabilization. Nature 479, 419–422.
Lim, A.L., 1984. The embryology of *Garcinia mangostana* L. (Clusiaceae). Gard Bull Sing 37, 93–103.
Lin, L., Tian, S., Kaeppler, S., Liu, Z., An, Y.Q., 2014. Conserved transcriptional regulatory programs underlying rice and barley germination. PLoS ONE 9.
Linkies, A., Leubner-Metzger, G., 2012. Beyond gibberellins and abscisic acid: how ethylene and jasmonates control seed germination. Plant Cell Rep. 31, 253–270.
Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods 25, 402–408.
Lohse, M., Nagel, A., Herter, T., May, P., Schroda, M., Zrenner, R., Tohge, T., Fernie, A.R., Stitt, M., Usadel, B., 2014. Mercator: A fast and simple web server for genome scale functional annotation of plant sequence data. Plant Cell Environ. 37, 1250–1258.
Lopez-Molina, L., Mongrand, S., McLachlin, D.T., Chait, B.T., Chua, N.H., 2002. ABI5 acts downstream of ABI3 to execute an ABA-dependent growth arrest during germination. Plant J. 32, 317–328.
Luo, W., Pant, G., Bhavnasi, Y.K., Blanchard, S.G., Brouwer, C., 2017. Pathview Web: user friendly pathway visualization and data integration. Nucl. Acids Res. 45, W501–W508.
Maadon, S.N., Rohani, E.R., Ismail, I., Baharum, S.N., Normah, M.N., 2016. Somatic embryogenesis and metabolic differences between embryogenic and non-embryogenic structures in mangosteen. Plant Cell Tissue Organ Culture 127, 443–459.
Machado, C.R., Costa De Oliveira, R.L., Boiteux, S., Praekelt, U.M., Meacock, P.A., Menck, C.F.M., 1996. Thi1, a thiamine biosynthetic gene in *Arabidopsis thaliana*, complements bacterial defects in DNA repair. Plant Mol. Biol. 31, 585–593.
Mahdavi-Darvari, F., Noor, N.M., 2017. New insight into early somatic embryogenesis of mangosteen (*Garcinia mangostana*) through de novo and comparative transcriptome analyses. Trop. Plant Biol. 10, 30–44.
Mähönen, A.P., Bishopp, A., Higuchi, M., Nieminen, K.M., Kinoshita, K., Törmäkangas, K., Ikeda, Y., Oka, A., Kakimoto, T., Helariutta, Y., 2006. Cytokinin signaling and its inhibitor AHP6 regulate cell fate during vascular development. Science 311, 94–98.
Malik, S.K., Chaudhury, R., Abraham, Z., 2005. Seed morphology and germination characteristics in three Garcinia species. Seed Sci. Technol. 33, 595–604.
Matra, D.D., Kozaki, T., Ishii, K., Poerwanto, R., Inoue, E., 2016. De novo transcriptome assembly of mangosteen (*Garcinia mangostana* L.) fruit. Genomics Data 10, 35–37.
Mazlan, O., Aizat, W.M., Aziz Zuddin, N.S., Baharum, S.N., Noor, N.M., 2019. Metabolite profiling of mangosteen seed germination highlights metabolic changes related to carbon utilization and seed protection. Sci. Horticult. 243, 226–234.
Mazlan, O., Aizat, W.M., Baharum, S.N., Azizan, K.A., Noor, N.M., 2018. Metabolomics analysis of developing *Garcinia mangostana* seed reveals modulated levels of sugars, organic acids and phenylpropanoid compounds. Sci. Horticult. 233, 323–330.
Meinke, D., Muralla, R., Sweeney, C., Dickerman, A., 2008. Identifying essential genes in *Arabidopsis thaliana*. Trends Plant Sci. 13, 483–491.
Meng, Y., Chen, F., Shuai, H., Luo, X., Ding, J., Tang, S., Xu, S., Liu, J., Liu, W., Du, J., Liu, J., Yang, F., Sun, X., Yong, T., Wang, X., Feng, Y., Shu, K., Yang, W., 2016. Karrikins delay soybean seed germination by mediating abscisic acid and gibberellin biogenesis under shaded conditions. Sci. Rep. 6.
Mhamdi, A., Mauve, C., Gouia, H., Saindrenan, P., Hodges, M., Noctor, G., 2010. Cytosolic NADP-dependent isocitrate dehydrogenase contributes to redox homeostasis and the regulation of pathogen responses in Arabidopsis leaves. Plant Cell Environ. 33, 1112–1123.
Michaeli, S., Fromm, H., 2015. Closing the loop on the GABA shunt in plants: Are GABA metabolism and signaling entwined? Front. Plant Sci. 6, 419.
Midin, M.R., Loke, K.K., Madon, M., Nordin, M.S., Goh, H.H., Mohd Noor, N., 2017. SMRT sequencing data for *Garcinia mangostana* L. variety Mesta. Genomics Data 12, 134–135.
Midin, M.R., Nordin, M.S., Madon, M., Saleh, M.N., Goh, H.H., Mohd Noor, N., 2018. Determination of the chromosome number and genome size of *Garcinia mangostana* L. via cytogenetics, flow cytometry and k-mer analyses. Caryologia 71, 35–44.
Miquel, M., Trigui, G., d'Andréa, S., Kelemen, Z., Baud, S., Berger, A., Deruyffelaere, C., Trubuil, A., Lepiniec, L., Dubreucq, B., 2014. Specialization of oleosins in oil body dynamics during seed development in Arabidopsis seeds. Plant Physiol. 164, 1866–1878.
Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., Kanehisa, M., 2007. KAAS: An automatic genome annotation and pathway reconstruction server. Nucl. Acids Res. 35, W182–W185.
Muralla, R., Lloyd, J., Meinke, D., 2011. Molecular foundations of reproductive lethality in *Arabidopsis thaliana*. PLOS ONE 6, e28398.
Narsai, R., Gouil, Q., Secco, D., Srivastava, A., Karpievitch, Y.V., Liew, L.C., Lister, R., Lewsey, M.G., Whelan, J., 2017a. Extensive transcriptomic and epigenomic remodelling occurs during *Arabidopsis thaliana* germination. Genome Biol. 18 172-172.
Narsai, R., Secco, D., Schultz, M.D., Ecker, J.R., Lister, R., Whelan, J., 2017b. Dynamic and rapid changes in the transcriptome and epigenome during germination and in developing rice (*Oryza sativa*) coleoptiles under anoxia and re-oxygenation. Plant J. 89, 805–824.
Nelson, S.K., Steber, C.M., 2017. Transcriptional mechanisms associated with seed dormancy and dormancy loss in the gibberellin-insensitive sly1-2 mutant of *Arabidopsis thaliana*. PLoS ONE 12.
Noor, N.M., Aizat, W.M., Hussin, K., Rohani, E.R., 2016. Seed characteristics and germination properties of four Garcinia (Clusiaceae) fruit species. Fruits 71, 199–207.
Normah, M.N., Nor-Azza, A.B., Aliudin, R., 1995. Factors affecting in vitro shoot proliferation and ex vitro establishment of mangosteen. Plant Cell Tissue Organ Culture 43, 291–294.
Normah, M.N., Ramiya, S.D., Gintangga, M., 1997. Desiccation sensitivity of recalcitrant seeds-a study on tropical fruit species. Seed Sci. Res. 7, 179–183.
Obolskiy, D., Pischel, I., Siriwatanametanon, N., Heinrich, M., 2009. *Garcinia mangostana* L.: A phytochemical and pharmacological review. Phytotherapy Res. 23, 1047–1065.
Ogé, L., Bourdais, G., Bove, J., BorisCollet, Godin, B., Granier, F., Boutin, J.P., Job, D., Jullien, M., Grappin, P., 2008. Protein repair L-Isoaspartyl methyltransferasel is involved in both seed longevity and germination vigor in arabidopsis. Plant Cell 20, 3022–3037.
Osman, M.B., Milan, A.R., 2006. Mangosteen - Garcinia Mangostana L. Mangosteen - Garcinia Mangostana. Southampton Centre for Underutilised Crops.
Pammenter, N.W., Berjak, P., 2000. Evolutionary and ecological aspects of recalcitrant seed biology. Seed Sci. Res. 10, 301–306.
Piyatrakul, P., Putranto, R.A., Martin, F., Rio, M., Dessailly, F., Leclercq, J., Dufayard, J.F., Lardet, L., Montoro, P., 2012. Some ethylene biosynthesis and AP2/ERF genes reveal a specific pattern of expression during somatic embryogenesis in *Hevea brasiliensis*. BMC Plant Biol. 12.
Rajjou, L., Gallardo, K., Debeaujon, I., Vandekerckhove, J., Job, C., Job, D., 2004. The effect of α-amanitin on the arabidopsis seed proteome highlights the distinct roles of stored and neosynthesized mRNAs during germination. Plant Physiol. 134, 1598–1613.
Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2009. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.
Rodríguez, M.J.P., Suárez, M.F., Heredia, R., Ávila, C., Breton, D., Trontin, J.F., Filonova, L., Bozhkov, P., Von Arnold, S., Harvengt, L., Cánovas, F.M., 2006. Expression patterns of two glutamine synthetase genes in zygotic and somatic pine embryos support specific roles in nitrogen metabolism during embryogenesis. New Phytologist 169, 35–44.
Rohani, E.R., Ismanizan, I., Noor, N.M., 2012. Somatic embryogenesis of mangosteen. Plant Cell Tissue Organ Culture 110, 251–259.
Rolletschek, H., Weber, H., Borisjuk, L., 2003. Energy status and its control on embryogenesis of legumes. Embryo photosynthesis contributes to oxygen supply and is coupled to biosynthetic fluxes. Plant Physiol. 132, 1196–1206.
Romero-Rodríguez, M.C., Archidona-Yuste, A., Abril, N., Gil-Serrano, A.M., Meijón, M., Jorrín-Novo, J.V., 2018. Germination and early seedling development in *Quercus ilex* recalcitrant and non-dormant seeds: targeted transcriptional, hormonal, and sugar analysis. Front. Plant Sci. 9, 1508.
Rushton, D.L., Tripathi, P., Rabara, R.C., Lin, J., Ringler, P., Boken, A.K., Langum, T.J., Smidt, L., Boomsma, D.D., Emme, N.J., Chen, X., Finer, J.J., Shen, Q.J., Rushton, P.J., 2012. WRKY transcription factors: Key components in abscisic acid signalling. Plant Biotechnol. J. 10, 2–11.
Santa-Catarina, C., Silveira, V., Balbuena, T.S., Viana, A.M., Estelita, M.E.M., Handro, W.,

Floh, E.I.S., 2006. IAA, ABA, polyamines and free amino acids associated with zygotic embryo development of *Ocotea catharinensis*. Plant Growth Regul. 49, 237–247.
Sattler, S.E., Gilliland, L.U., Magallanes-Lundback, M., Pollard, M., DellaPenna, D., 2004. Vitamin E is essential for seed longevity and for preventing lipid peroxidation during germination. Plant Cell 16, 1419–1432.
Schwenkel, H.G., Winkelmann, T., 1998. Plant regeneration via somatic embryogenesis from ovules of *Cyclamen persicum* Mill. Plant Tissue Culture Biotechnol. 4, 28–34.
Sekhon, R.S., Briskine, R., Hirsch, C.N., Myers, C.L., Springer, N.M., Buell, C.R., de Leon, N., Kaeppler, S.M., 2013. Maize gene atlas developed by RNA sequencing and comparative evaluation of transcriptomes based on RNA sequencing and microarrays. PLOS ONE 8, e61005.
Signorelli, S., Considine, M.J., 2018. Nitric oxide enables germination by a four-pronged attack on ABA-induced seed dormancy. Front. Plant Sci. 9, 296.
Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L., Ideker, T., 2011. Cytoscape 2.8: New features for data integration and network visualization. Bioinformatics 27, 431–432.
Soós, V., Sebestyén, E., Juhász, A., Light, M.E., Kohout, L., Szalai, G., Tandori, J., Van Staden, J., Balázs, E., 2010. Transcriptome analysis of germinating maize kernels exposed to smoke-water and the active compound KAR1. BMC Plant Biol. 10, 236.
Supek, F., Bošnjak, M., Škunca, N., Šmuc, T., 2011. Revigo summarizes and visualizes long lists of gene ontology terms. PLoS ONE 6.
Te-Chato, S., Lim, M., 1999. Plant regeneration of mangosteen via nodular callus formation. Plant Cell Tissue Organ Culture 59, 89–93.
Te-chato, S., Lim, M., 2000. Improvement of mangosteen micropropagation through meristematic nodular callus formation from in vitro-derived leaf explants. Sci. Horticult. 86, 291–298.
Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., Selbig, J., Müller, L.A., Rhee, S.Y., Stitt, M., 2004. MAPMAN: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J. 37, 914–939.
Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., Cox, J., 2016. The Perseus computational platform for comprehensive analysis of (prote) omics data. Nature Methods 13, 731–740.
Volodarsky, D., Leviatan, N., Otcheretianski, A., Fluhr, R., 2009. HORMONOMETER: A tool for discerning transcript signatures of hormone action in the Arabidopsis transcriptome. Plant Physiol. 150, 1796–1805.
Wang, S., Chang, Y., Ellis, B., 2016. Overview of OVATE FAMILY PROTEINS, a novel class of plant-specific growth regulators. Front. Plant Sci. 7, 417.
Waters, M.T., Smith, S.M., 2013. KAI2- and MAX2-mediated responses to karrikins and strigolactones are largely independent of HY5 in arabidopsis seedlings. Mol. Plant 6, 63–75.
Waterworth, W.M., Masnavi, G., Bhardwaj, R.M., Jiang, Q., Bray, C.M., West, C.E., 2010. A plant DNA ligase is an important determinant of seed longevity. Plant J. 63, 848–860.
Weitbrecht, K., Müller, K., Leubner-Metzger, G., 2011. First off the mark: Early seed germination. J. Exp. Bot. 62, 3289–3309.
Winkelmann, T., 2016. Somatic versus zygotic embryogenesis: learning from seeds. In: Germana, M.A., Lambardi, M. (Eds.), In Vitro Embryogenesis in Higher Plants. Springer, New York, New York, NY, pp. 25–46.
Wojtyla, Ł., Lechowska, K., Kubala, S., Garnczarska, M., 2016. Different modes of hydrogen peroxide action during seed germination. Front. Plant Sci. 7, 66.
Yapwattanaphun, C., Kobayashi, S., Yonemori, K., Ueda, J., 2014. Hormone analysis in the locule of mangosteen fruit during apomictic seed development. Acta Horticult. 1024, 141–146.
Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., Madden, T.L., 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinform. 13, 134.
Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R., Bolund, L., Wang, J., 2006. WEGO: A web tool for plotting GO annotations. Nucleic Acids Res. 34, W293–W297.
Yonemori, K., Nishiyama, S., Yapwattanaphun, C., Ueda, J., 2014. Identification of plant hormones in endosperm liquid of mangosteen fruits at young developmental stages. Acta Horticult. 1042, 89–95.
Yu, Y., Guo, G., Lv, D., Hu, Y., Li, J., Li, X., Yan, Y., 2014. Transcriptome analysis during seed germination of elite Chinese bread wheat cultivar Jimai 20. BMC Plant Biol. 14.
Zheng, Y., Jiao, C., Sun, H., Rosli, H.G., Pombo, M.A., Zhang, P., Banf, M., Dai, X., Martin, G.B., Giovannoni, J.J., Zhao, P.X., Rhee, S.Y., Fei, Z., 2016. iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. Mol. Plant 9, 1667–1670.