Transcriptional transitions in Alphonso mango (*Mangifera indica* L.) during fruit development and ripening explain its distinct aroma and shelf life characteristics

Ashish B. Deshpande, Krishanpal Anamika, Vineet Jha, Hemangi G. Chidley, Pranjali S. Oak, Narendra Y. Kadoo, Keshav H. Pujari, Ashok P. Giri & Vidya S. Gupta

Alphonso is known as the “King of mangos” due to its unique flavor, attractive color, low fiber pulp and long shelf life. We analyzed the transcriptome of Alphonso mango through Illumina sequencing from seven stages of fruit development and ripening as well as flower. Total transcriptome data from these stages ranged between 65 and 143 Mb. Importantly, 20,755 unique transcripts were annotated and 4,611 were assigned enzyme commission numbers, which encoded 142 biological pathways. These included ethylene and flavor related secondary metabolite biosynthesis pathways, as well as those involved in metabolism of starch, sucrose, amino acids and fatty acids. Differential regulation (p-value ≤ 0.05) of thousands of transcripts was evident in various stages of fruit development and ripening. Novel transcripts for biosynthesis of mono-terpenes, sesqui-terpenes, di-terpenes, lactones and furanones involved in flavor formation were identified. Large number of transcripts encoding cell wall modifying enzymes was found to be steady in their expression, while few were differentially regulated through these stages. Novel 79 transcripts of inhibitors of cell wall modifying enzymes were simultaneously detected throughout Alphonso fruit development and ripening, suggesting controlled activity of these enzymes involved in fruit softening.

Mango (*Mangifera indica* L.) is one of the most popular and highly favored fruit. Global mango production was reported to be 43.3 million metric tons in 2013 preceding banana, apple, grape and orange. There are thousands of mango cultivars worldwide, among which Alphonso, Keitt, Kent, Lilli, Zill, Osteen, Haden, Kesar, Pairi, Dashehari, Langra and Banganapalli are well known. These varieties vary in their fruit color, size, shape, flavor, taste and ripening period and pattern. To understand the composition and biosynthesis of their unique flavor and complex ripening process various studies have been carried out at metabolic, proteomic, genetic and post-harvest processing levels. Whole genome sequencing and RNA sequencing (RNAseq) are the two important high throughput technologies recently adopted to understand the complex cellular and physiological processes in fruits such as citrus, tomato and strawberry as well as domestication and diseases tolerance in citrus. Although mango genome sequence is yet not available, few recent studies have described the transcriptomic analysis from various tissues of few mango cultivars. The first report from Zill mango provided extensive transcriptomic and proteomic profiling from pulp and skin tissues of four fruit developing stages using pooled RNA but not stage specific and differentially expressed transcripts. Another study of leaf transcriptome and chloroplast genome sequencing from cultivar Langra provided information about the production of several bioactive compounds. Transcriptome analysis from two (raw and ripe) and three (raw, mid ripe and ripe) stages

1Plant Molecular Biology Group, Biochemical Sciences Division, CSIR-National Chemical Laboratory, Pune, 411008, Maharashtra, India. 2Labs, Persistent Systems Limited, Pingala-Aryabhata, Erandwane, Pune, 411004, India. 3Dr. Balasaheb Sawant Konkan Agriculture University, Dapoli, 415712, Maharashtra, India. Correspondence and requests for materials should be addressed to V.S.G. (email: vs.gupta@ncl.res.in)
of Kent\textsuperscript{27} and Dashehari fruit pulp\textsuperscript{28}, respectively provided important insights into the ripening process and flavor biogenesis in these mango cultivars.

India is the largest producer and exporter of mango with 40.6% share in international mango market (http://www.fao.org). Among the Indian mango cultivars Alphonso is globally favored and highly exported mango due to its unique and attractive flavor, low fiber containing pulp and high carotene content\textsuperscript{26,30}. The ripening duration of Alphonso mango is 15 days from harvest, which is the highest among all mango cultivars; for example, the ripening duration for Kent and Dashehari mango fruit is 10 and 6 days, respectively\textsuperscript{27,28}. Fruit ripening in Alphonso mango progresses from skin towards the stone leading to attractive skin color and easy monitoring of ripening progress. On the other hand various mango varieties, viz. Haden, Keitt, Kent, Tommy Atkins (National Mango Board, USA; http://www.mango.org) and Dashehari\textsuperscript{29} show polarity of their ripening from fruit stone to skin, making it difficult to identify ripened fruits. Longer ripening duration and shelf life of Alphonso mango provides sufficient time for its transportation globally. The mechanisms underlying these unique properties of Alphonso mango need to be explored in depth at spatial and temporal level of fruit development and ripening using transcriptomic, proteomic and metabolomic approaches as they can be correlated to the specific phenotype.

In the present study, we performed transcriptome analysis of Alphonso mango through eight different tissues such as flower, whole fruit at 30 and 60 DAP (Days After Pollination), pulp and skin of 90 DAP fruit (mature raw fruit) and pulp from three fruit ripening stages i.e. 5 DAH (Days After Harvest): table green stage; 10 DAH: mid ripe stage and 15 DAH: ripe stage to analyze various fruit developing and ripening processes in Alphonso mango.

Results

Alphonso mango transcriptome. Alphonso mango transcriptome was screened through eight tissue samples. To map differentially expressed transcripts a merged assembly was generated from the reads of all the tissues, which reflected upon overall Alphonso mango flower and fruit transcriptome. For each tissue read numbers were more than 100 million, which were assembled using k-mers 67, 75 and 83 separately and then merged for individual tissue. Average number of unique transcripts post assembly was 76,043 and with N50 and N80 values as 1,835 and 1,008 bp, respectively (Table 1). The minimum and maximum lengths of transcript from these assemblies were 100 and 17,342 bp, respectively (Table 1). Average number of transcripts was 11,925 upon filtering for redundancy and identifying candidate coding regions with maximum 90% identity and minimum 70% coverage (Table 2).

Unique transcripts from each assembled and filtered set were subjected to BLASTx against the non-redundant dataset from NCBI (http://www.ncbi.nlm.nih.gov). From a total of 20,755 unique transcripts from the merged assembl, 92.22% transcripts were annotated while 954 transcripts (4.59%) encoded hypothetical proteins and assembly, 92.22% transcripts were annotated while 954 transcripts (4.59%) encoded hypothetical proteins and

Differentially expressed transcripts. Comparison between adjacent tissue stages was carried out to identify differentially expressed transcripts at each stage of fruit development and ripening (Table 3, Supplementary Table ST1). Flower to fruit of 30DAP, 524 transcripts were down regulated while 181 were up-regulated. Among the down-regulated transcripts alpha-amylase and subtilisin inhibitor-like (contig_6593), carbonic anhydrase (contig_5377), chitinase (contig_4907) and maternal effect embryo arrest 59 (contig_4949) exhibited the highest fold change (\texttimes 2-fold); while homeodomain-like protein (contig_12912 and 12913), cytochrome p450 - cyp72a219-like (contig_4083), heat shock cognate 70kDa protein (contig_9620) and inositol-3-phosphate synthase (contig_9620) were highly up-regulated (\texttimes 2-fold).

Transition from 30 DAP to 60 DAP resulted in up-regulation of 73 and down-regulation of seven transcripts. Important down-regulated transcripts were n-acetyltransferase (contig_1261), 9-cis-epoxyoctenolid dioxygenase (contig_4907), protein reversion-to-ethylene sensitivity (contig_9164) and ethylene receptor 2 (contig_6147). While important up-regulated transcripts were nucleotide sugar transporter family protein (contig_6337), beta-xylosyltransferase (contig_4150), various cellulose synthase catalytic subunits and laccases. Comparison of 60 DAP fruit tissue with 90 DAP pulp and skin tissue, respectively revealed down-regulation of beta-xylosyltransferase, beta-d-xylosidase, cellulose synthase and galacturonosyl transferase; whereas spx and exs domain-containing protein was up-regulated in both 90 DAP pulp and skin. Up-regulation of homeobox
Table 1. Alphonso de novo transcriptome assembly statistics.

| Assembly | No of NR transcripts |
|----------|----------------------|
| Flower   | 15778                |
| 30 DAP   | 14622                |
| 60 DAP   | 14090                |
| 90 DAP pulp | 9382               |
| 90 DAP skin | 12664              |
| 5 DAH    | 10084                |
| 10 DAH   | 9748                 |
| 15 DAH   | 9032                 |

Table 2. Number of non redundant (NR) transcripts/sample.

| Protein | Frequency |
|---------|-----------|
| protein sbh1, gds1 esterase lipase, caffeoyl shikimate esterase, hydroperoxide lyase, udp-rhamnose:rhamnosyl transferase and pectinesterase inhibitor was evicted from 90 DAP pulp tissue compared to that in 60 DAP fruit tissue. Evaluation of differentially expressed transcripts between pulp and skin of 90 DAP revealed very few |
transcripts down-regulated in skin with none of them showing change more than 2-fold, while 54 transcripts were up-regulated in 90 DAP skin tissue including the important ones as DNA mismatch repair protein msh5, acyl carrier protein, amino-acid permease, calcium-transporting ATPase, isoflavone reductase and various disease resistance proteins.

During ripening of Alphonso pulp from 90 DAP to 5 DAH, 42 transcripts were up-regulated (>2-fold) including methyltransferase, amino-acid permease, chloroplastic 9-cis-epoxycarotenoid dioxygenase,

Figure 1. Blast statistics showing (a) distribution of top hit species, (b) distribution of top gene ontologies from BP: biological processes, MF: molecular functions and CC: cellular components and (c) Enzyme code distribution: number of transcripts (Y-axis) encoding six classes of enzymes through various stages of Alphonso mango fruit development and ripening (X-axis).
beta-galactosidase and protein phosphatase mainly, whereas 52 transcripts were down-regulated, important ones being few disease resistance proteins, peroxidase, sucrose synthase and glycerol-3-phosphate dehydrogenase. The highest level of differential expression was evinced through transition from 5 DAH to 10 DAH among all the ripening tissues wherein 418 transcripts were down and 191 were up-regulated. These stages are known for non-climacteric to climacteric transition during Alphonso fruit ripening based on metabolite analysis. Prominently down regulated transcripts were phospholipase-A LCAT3, sugar phosphate exchanger, auxin-responsive protein IAA9, abscisate β-glucosyltransferase and membrane-associated kinase regulator with more than 3-fold change, while up-regulated transcripts were aspartic proteinase nepenthesin, UDP-glucose 6-dehydrogenase, 1-aminocyclopropane-1-carboxylate oxidase, peroxidase, bidirectional sugar transporter sweet1-like, omega-6 fatty acid desaturase and squalene monooxygenase. During the transition from 10 DAH to 15 DAH stage, increased phosphate metabolism was evident. Total 30 transcripts were up-regulated mainly including phospholipase-D, inorganic pyrophosphatase, phosphate transporter, transcription factor glk2 and DNA translocase, whereas 12 transcripts were down-regulated and important ones were aspartic protease, plastocyanin-like domain protein, annexin d2-like and glucuronosyltransferase (>2-fold).

**Idiosyncratic transcripts for the stage.** Various transcripts unique to each of these stages (Fig. 2, Table 4, Supplementary Table ST2) representing multiple stage specific processes were identified during this comparative analysis. A total of 388 transcripts were identified as unique to flower tissue, which mostly included.

| Comparison                  | Up regulated | Down regulated |
|-----------------------------|--------------|----------------|
| 30DAP vs Flower             | 181          | 524            |
| 60DAP vs 30DAP              | 73           | 7              |
| 90DAP pulp vs 60DAP         | 31           | 158            |
| 90DAP skin vs 60DAP         | 4            | 93             |
| 90DAP skin vs 90DAP pulp    | 54           | 4              |
| 5DAH vs 90DAP pulp          | 42           | 52             |
| 10DAH vs 5DAH               | 191          | 418            |
| 15DAH vs 10DAH              | 30           | 12             |

Table 3. Number of differentially expressed transcripts.

**Figure 2.** Venn diagrams representing common and distinct transcripts through various comparisons (a) Flower, 30 DAP, 60 DAP and 90 DAP skin; (b) 90 DAP pulp, 5 DAH, 10 DAH and 15 DAH; (c) 90 DAP pulp and 90 DAP skin and (d) fruit developing and ripening stages.
various transcription and translation factors, late embryogenesis abundant proteins, stress-sensitive and dehydration-responsive element-binding protein-1b and various ribosomal proteins.

Similarly, distinct transcripts specific to fruit developing and ripening stages were identified, wherein 1,090 and 27 transcripts were idiosyncratic to the fruit developing and ripening stages, respectively (Supplementary Table ST2). Various auxin and gibberellin induced and regulated proteins as well as several proteins responsible for various vacuole activities, multiple disease resistance proteins and various terpene synthases were distinct to the developing stages. Transcripts encoding for multiple transcription and translation factors during Alphonso fruit development were also detected. Interestingly multiple ethylene responsive transcription factors along with the protein reversion-to-ethylene sensitivity were uniquely revealed in the developing stages. Whereas, NIN-like protein, respiratory burst oxidase homolog protein d-like, rhamnogalacturonate lyase b-like, lectin receptor kinase, gag protein and methionine \( \Upsilon \)-lyase were exclusive to the Alphonso ripening stages. Similarly, WRKY transcription factor 43, b3 domain-containing val3 and ap2 ERF domain-containing transcription factors were uniquely identified from the ripening stages. Several hypothetical and uncharacterized proteins were also found to be Alphonso ripening specific and their characterization might help to reveal the ripening process in Alphonso.

Gene ontology (GO) enrichment. Fisher’s exact test was performed to understand over- and down-represented GOs (p-value \( \leq 0.001 \)) during the transition, which gave an overall picture of the Alphonso mango development and ripening (Fig. 3). During the flower to 30 DAP fruit transition certain GOs were over-expressed such as post-embryonic developmental and anatomical structure morphogenesis process, response to abiotic stimulus and various plastid and thylakoid processes. At the same time, hydrolysis of O-glycosyl compounds and UDP-glycosyltransferase activity along with the methyltransferase, sucrose metabolic and lipid biosynthetic activities coding GOs were down-represented. 30 DAP to 60 DAP transition described enriched GOs for starch and sucrose metabolic process, fatty acid, cellulose and vitamin biosynthetic processes, protein glycosylation, response to oxidative stress and heat along with HSP binding. During the same event cell differentiation, growth and cell death in addition to the anatomical structure morphogenesis, secondary metabolic process and response to abiotic stimulus coding GOs were observed to be decreased. Comparison between 60 DAP fruit and 90 DAP pulp revealed over-represented GOs for cell differentiation and cell growth, response to biotic stimulus, secondary metabolic process, tropism and regulation of gene expression. While, starch and sucrose metabolic processes, O-glycosyl hydrolase UDP-glucosyltransferase activities, fatty acid biosynthesis and catabolic processes along with the lipid transport, various amino acid and terpenoid biosynthesis and cell wall biogenesis were down-represented. 90 DAP pulp and skin tissues showed over-expression of various ion, ADP and sequence-specific DNA binding activities along with the response to hormone, chemical and organic substances in the skin tissue; whereas only cytoplasm and cytoplasmic part related GOs were found to be down-expressed in the skin compared to that in the pulp at 90 DAP.

Transition of 90 DAP pulp to 5 DAH pulp revealed only over-representation of GOs such as biosynthesis of RNA, aromatic compounds, and nucleobase containing compounds; binding of various ions; regulation of primary, cellular, nitrogenous compounds, RNA biosynthesis processes and gene expression and response to hormone, chemical and organic substance. Surprisingly, none of the GOs showed significant (P-value \( \leq 0.001 \)) enrichment during the transition from 5 DAH to 10 DAH and reflected as the stationary phase of Alphonso mango ripening. The 10 DAH (mid ripe stage) to 15 DAH (ripe stage) transition showed over-expression of many GOs such as binding of various ions; phospholipid, glycolipid, amino acid, co-enzyme and water soluble vitamin.

Table 4. Distinct transcripts.

| Comparison           | Stage/s          | No. of unique transcripts |
|----------------------|------------------|---------------------------|
| 30DAP vs Flower      | Flower           | 383                       |
|                      | 30DAP            | 44                        |
| 60DAP vs 30DAP       | 30DAP            | 134                       |
|                      | 60DAP            | 195                       |
| 90DAP pulp vs 60DAP  | 60DAP            | 1306                      |
|                      | 90DAP pulp       | 36                        |
| 90DAP skin vs 60DAP  | 60DAP            | 388                       |
|                      | 90DAP skin       | 79                        |
| 90DAP skin vs 90DAP pulp | 90DAP pulp | 82                        |
|                      | 90DAP skin       | 1043                      |
| 5DAH vs 90DAP pulp   | 90DAP pulp       | 489                       |
|                      | 5DAH             | 390                       |
| 10DAH vs 5DAH        | 5DAH             | 1027                      |
|                      | 10DAH            | 343                       |
| 15DAH vs 10DAH       | 10DAH            | 563                       |
|                      | 15DAH            | 473                       |

| Development vs Ripening | No. of unique transcripts |
|-------------------------|---------------------------|
| Development             | 1090                      |
| Ripening                | 27                        |
biosynthetic processes; methyltransferase, UDP-glycosyltransferase and O-glycosyl hydrolase activities; starch, sucrose metabolic processes, gluconeogenesis and glycolytic process; protein phosphorylation, dephosphorylation, folding, protein modification and small protein conjugation along with the regulation and response to oxidative stress, oxidoreductase activity and redox homeostasis. While the GOs related to plastid and cytosol; post embryonic development; response to abiotic and endogenous stimulus; anatomical structure morphogenesis, plasma membrane, cell differentiation and communication were down-represented.

Spatial changes in transcriptome at 90 DAP. In case of Alphonso mango, 90 DAP stage is a mature raw stage of the fruit and is considered as the right stage of fruit harvest (0 DAH) for further artificial ripening of the fruit. Hence transcriptome analysis of skin and pulp were separately carried out at this stage. Overall 90 DAP skin was found to be metabolically more active compared to the 90 DAP pulp with respect to differentially
expressed genes (Supplementary Table ST1), unique genes (Supplementary Table ST2) and enriched GOs (Fig. 3). In the skin, 54 transcripts were up-regulated whereas only four were down-regulated compared to the pulp. Among the up-regulated transcripts (>2-fold), isoflavone reductase, transcription and translation regulatory proteins, hydrolases, methyltransferase etc. were more prominent; whereas four down-regulated transcripts were membrane and cytoplasm related GOs. Unique transcripts upon comparison between 90 DAP pulp and skin showed carotene and xanthophylls biosynthesis related contigs, beta-carotene hydroxylase (contig_5406) and anthocyanidin 3-o-glucosyltransferase, flavor related various terpene synthases and ripening related contigs such as ethylene-responsive transcription factors, pectate lyase, pectin esterase and cellulase as unique to skin compared to the pulp. Interestingly, in comparison of 5DAH stage with 90 DAP pulp few of these transcripts such as isoflavone reductase (contig_5337) and beta-carotene hydroxylase (contig_5406) were observed to be idiosyncratic to 5 DAH stage as seen in 90 DAP skin stage. Similarly, transcripts encoding enzymes involved in pectin degradation (contigs_9578, 9579, 18096 and 3603) were found to be upregulated (>2-fold) in the comparison between 10DAH vs 90 DAP pulp and 15 DAH vs 90 DAP pulp. These finding suggest progression of ripening related molecular processes in the pulp of 5 DAH and onward stages which were initiated at 90 DAP skin and highlight initiation of Alphonso ripening process from skin and its probable progress towards fruit stone.

Figure 4. Cladogram representing phylogenetic analysis of proteins encoded by various terpene synthases identified in the present study (blue color) along with the terpene synthases from other Angiosperm plants (black color) using the neighbor-joining method. The node label is composed of two letters representing botanical name of the plant followed by name of the enzyme. Details of the plant species, enzyme and NCBI accession numbers in parenthesis are as follows Vitis vinifera_ (E)-beta-caryophyllene synthase (ADR74192.1), Vitis vinifera_germacrene D synthase (ADR74198.1), Vitis vinifera_nerolidol synthase (ADR74211.1), Vitis vinifera_(E,E)-geranyl linalool synthase (ADR74219.1), Malus domestica_ent-kaurene synthase (AFG18184.1), Solanum lycopersicum_ent-kaurene synthase (AEP82778.1), Vitis vinifera_(E)-beta-ocimene synthase (ADR74204.1), Vitis vinifera_Alphaterpeneol synthase (ADR74202.1), Vitis vinifera_myrcene synthase (NP_001268009), Citrus sinensis_alpha-farnesene synthase (XP_006467948.1) Populus trichocarpa_alpha-farnesene synthase (XP_002317269.2), Ricinus communis_alpha-farnesene synthase (XP_015574261.1), Theobroma cacao_alpha-farnesene synthase (EOY28527.1) and Vitis vinifera alpha_farnesene synthase (NP_001268183.1).
Figure 5. Quantitative reverse-transcriptase PCR validation of various transcripts obtained through RNAseq from carbohydrate metabolism through various tissues (flower and fruit developing and ripening stages). Vertical bars at each data point represent standard error in the relative quantification among the three biological replicates. X-axis represents fruit development and ripening stages and Y-axis represents relative transcript abundance. The horizontal bar above each histogram represents the expression level of the same transcript as obtained through RNAseq analysis, wherein dark-red color indicates higher expression and light-green color indicates lower expression. Gene names are as referred in result section "Transcriptome validation through qRT PCR".
Genes involved in the flavor biogenesis in Alphonso mango. Quantitatively mono-terpenes are abundant in Alphonso followed by sesqui-terpenes. The present data revealed six contigs encoding mono-terpene synthases (limonene synthase1, limonene synthase2, beta ocimine synthase1, beta ocimine synthase2, isoprene synthase1 and isoprene synthase2), five contigs encoding sesquiterpene synthases (germacreneD synthase1, germacreneD synthase2, nerolidol synthase1, nerolidol synthase2 and alpha farnesene synthase) and three contigs coding for di-terpene synthases (ent-kaurene synthase and casbene synthase (E,E)-geranillinalool synthase). Phylogenetic analysis (Fig. 4) of these genes along with other plant terpene synthases (TPS) showed distribution of these genes in to the TPS-a, TPS-b, TPS-e and TPS-f clades, respectively.

Furaneol and mesifuran are the two furanones from Alphonso mango and their synthesis by enone oxidoreductase (EO) and O-methyltransferase (OMTS), respectively have been described earlier. Multiple contigs coding for quinone oxidoreductase and O-methyltransferases were detected in the present analysis. Phylogenetic

Figure 6. Quantitative reverse-transcriptase PCR validation of various transcripts obtained through RNAseq from lipid metabolism through various tissues (flower and fruit developing and ripening stages). Vertical bars at each data point represent standard error in the relative quantification among the three biological replicates. X-axis represents fruit development and ripening stages and Y-axis represents relative transcript abundance. The horizontal bar above each histogram represents the expression level of the same transcript as obtained through RNAseq analysis, wherein dark-red color indicates higher expression and light-green color indicates lower expression. Gene names are as referred in result section "Transcriptome validation through qRT PCR".
Figure 7. Quantitative reverse-transcriptase PCR validation of various transcripts obtained through RNAseq from ethylene responsive factors (ERF) and disease resistance proteins (DRP) through various tissues (flower and fruit developing and ripening stages). Vertical bars at each data point represent standard error in the relative quantification among the three biological replicates. X-axis represents fruit development and ripening stages and Y-axis represents relative transcript abundance. The horizontal bar above each histogram represents the expression level of the same transcript as obtained through RNAseq analysis, wherein dark-red color indicates higher expression and light-green color indicates lower expression.
analysis of these contigs with the characterized genes revealed another similar transcript variant for the MiEO (Supplementary Figure SF3), whereas none of the contigs showed similarity to the MiOMTS (Supplementary Figure SF4). Green grassy aroma of unripe fruits is due to the C6 volatiles formed during the lipoxygenase and hydroperoxide lyase (HPL) pathways. In the present study a single contig encoding the hydroperoxide lyase and six

Figure 8. Quantitative reverse-transcriptase PCR validation of various transcripts obtained through RNAseq from mono-terpene metabolism through various tissues (flower and fruit developing and ripening stages). Vertical bars at each data point represent standard error in the relative quantification among the three biological replicates. X-axis represents fruit development and ripening stages and Y-axis represents relative transcript abundance. The horizontal bar above each histogram represents the expression level of the same transcript as obtained through RNAseq analysis, wherein dark-red color indicates higher expression and light-green color indicates lower expression. Gene names are as referred in result section “Transcriptome validation through qRT PCR”.
Figure 9. Quantitative reverse-transcriptase PCR validation of various transcripts obtained through RNAseq from sesqui-terpene and di-terpene metabolism through various tissues (flower and fruit developing and ripening stages). Vertical bars at each data point represent standard error in the relative quantification among the three biological replicates. X-axis represents fruit development and ripening stages and Y-axis represents relative transcript abundance. The horizontal bar above each histogram represents the expression level of the same transcript as obtained through RNAseq analysis, wherein dark-red color indicates higher expression and light-green color indicates lower expression. Gene names are as referred in result section “Transcriptome validation through qRT PCR”.
contigs encoding 13-lipoxygenase (Mi9LOX) and epoxide hydrolase 2 (MiEH2) in the biogenesis of lactones from Alphonso mango was confirmed in our recent study\textsuperscript{35}. One more transcript encoding 9-lipoxygenase similar to that of the characterized Mi9LOX (Supplementary Figure SF5) and three contigs coding for epoxide hydrolase 2 grouping with the MiEH2 (Supplementary Figure SF6) were additionally detected in the present study. Three more novel contigs coding for epoxide hydrolase were also detected but neither grouped with the other EH1 or EH2 from different plant species (Supplementary Figure SF6).

We further analyzed the differential expression of all these flavor related genes in terms of transcript abundance in flower and fruit developing and ripening stages (Supplementary Figure SF7). Among these the contigs encoding various TPS were abundant in flower and during early developing stages. Interestingly, many contigs encoding EH (contigs 8280, 3904, 3901, 14123 and 8281), LOX (contigs 18105, 12748, 12747 and 12385) and EO (contigs 8026, 5618, 15594, 14137 and 13600) were found to be ripening specific and might be playing crucial role in generating unique aroma volatiles during Alphonso fruit ripening.

**Glycosidases and cell wall degrading enzymes from Alphonso mango.** Glycosidases are involved in a variety of functions such as hydrolysis of complex carbohydrates (storage and structural) to mono-saccharides, removal of sugars from various glycans including glycosidically bound aroma volatiles, which serve as storage pool for the aroma compounds. In the present study, many glycosidases acting on various sugars i.e. glucose, galactose, mannose, fructose, xylose, fucose and rhamnose, were detected. Among these, class glucosidase with the highest number of contigs (51 contigs) was observed to contain 28 and 21 contigs coding for glucan $\beta$-glucosidase and general $\beta$-glucosidase, respectively. Two contigs coding for $\alpha$-glucosidases were identified of which one coded for glucan $\alpha$-glucosidase and the other for general $\alpha$-glucosidase. Among these, at 30 DAP stage contig_7442 and contig_7857 were found to be down-regulated (>3-fold) compared to those in flower stage. Contig_16888 was down-regulated whereas contig_17138 was up-regulated at 10 DAH than those at 5 DAH. At 15 DAH contig_9072 was down-regulated (1.5-fold) compared to that at 10 DAH. Among the galactosidase class, nine contigs encoding $\alpha$-galactosidase and 21 contigs encoding $\beta$-galactosidase were detected. Two of these were down-regulated (contig_1095 and contig_1096) at 30 DAP compared to those in flower. Contig_3844 was down regulated in 90 DAP pulp compared to 60 DAP.

Transition from 90 DAP to 5 DAH reflected in to up-regulation of contig_3844 (2.72-fold) whereas, contig_1525 coding for $\alpha$-galactosidase was down-regulated in 10 DAH compared to that at 5 DAH. In the mannosidase class, five $\alpha$-mannosidase and eight endo-$\beta$-mannosidase coding contigs were identified out of which two encoding for endo-beta-mannosidase (contig_15554 and contig_15558) were down-regulated at 10 DAH compared to 5 DAH, rest didn’t show differential regulation. Among two $\alpha$-xylosidase and five $\beta$-xylosidase from Alphonso mango only contig_1633 showed differential regulation which was up-regulated at 60 DAP compared to 30 DAH and was further down-regulated in both the 90 DAH tissues. Two transcripts coding for the acid beta-fructofuranosidase (contig_12501 and contig_12502) were detected but did not show any differential regulation in various tissues analyzed. Also, seven contigs coding for the alpha-l-fucosidase were identified from Alphonso mango but did not show differential regulation.

Degradation of plant cell wall components, namely cellulose, hemi-cellulose and pectin by cellulases and glucanases in ripening fruits is responsible for the fruit softness. Cellulases encoding five contigs were identified among which one coded for acidic cellulase and found to be abundant during flower and early fruit developing stages. Total 18 contigs encoding glucanase were detected of which contig_4148 was up-regulated in flower compared to 30 DAH, whereas, four transcripts (contig_9268, contig_19283, contig_9267 and contig_17145) were down-regulated (>2-fold) at 10 DAH compared to 5 DAH stage. Pectin is another component of fruit cell
wall and is degraded by a set of enzymes viz. pectate lyase (PL), pectin esterase (PE), polygalacturonase (PG) and rhamnogalacturonate lyase. In the present study 18 PL, 24 PG and 10 PE coding transcripts were detected. However, only few were differentially expressed, for example, only PL contig. 7696 was down-regulated in 30 DAP fruit compared to flower and PL contig. 9578 was up-regulated in 10 DAH fruit than 5 DAH fruit. PG non-catalytic subunit ip650 coding contig. 9895 was down-regulated in 5 DAH pulp than in 90 DAP pulp, whereas PG contig. 3471 was down regulated and PG contig. 1614 was up-regulated in 10 DAH compared to 5 DAH stage. Likewise, PE contig. 7997 was down-regulated and contig. 9162 was up-regulated at 30 DAP stage compared to flower. Similarly, eight transcripts encoding rhamnogalacturonate lyase were detected of which two (contig. 7745 and contig. 7746) were distinct to the ripening stages (Supplementary Table ST2).

Transcriptome analysis identified novel enzyme inhibitors from Alphonso mango. We identified various classes of enzyme inhibitors (Supplementary Table ST3) such as α-amylase inhibitor (2 contigs), inhibitor of proliferation pd5, apoptosis inhibitor (2 contigs), bax inhibitor (4 contigs), lipid transfer protein inhibitor (6 contigs), various kinase inhibitors (16 contigs), cysteine proteinases inhibitor (4 contigs), inter-alpha-trypsin inhibitor (2 contigs), serine protease inhibitor, kunitz family trypsin and protease inhibitor, guanosine nucleotide diphosphate dissociation inhibitor, nf-kappa-b inhibitor (2 contigs), pectinesterase inhibitor (12 contigs), polygalacturonase inhibitor, proteasome inhibitor, protein transport inhibitor (4 contigs), protein phosphatase inhibitor and rho gdp-dissociation protein inhibitor (4 contigs). Along with these, contigs coding for protein reversion-to-ethylene sensitivity (3 contigs), cell wall and vascular inhibitor of beta-fructosidase (2 contigs) and macrophage migration inhibitory factor were also identified from Alphonso transcriptome. These inhibitors showed differential regulation during fruit development and ripening (Supplementary Figure SF8). Most of the inhibitors were found to be expressing throughout all the stages except for the group 4 (coding for bax inhibitor), which were abundant only during fruit ripening stages of Alphonso mango and probably played important role in ripening physiology of Alphonso mango.

Transcriptome validation through qRT PCR. Results of the transcriptomic analysis were validated by qRT PCR of 38 representative genes from various metabolic pathways such as carbohydrate metabolism (cellulose synthase: CS, chitinase: CTN, and pectate lyase: PL), fatty acid metabolism (omega 3 fatty acyl desaturase: O3FAD, omega 6 fatty acyl desaturase: O6FAD, glyceraldehyde-3-phosphate acyl transferase: G3PAT, alcohol dehydrogenase: ADH and long chain fatty acyl CoA ligase: LCFACL), terpene metabolism (mono-terpene synthases: MTPS, sesqui-terpene synthases: STPS and di-terpene synthases: DTPS) and proteins such as ethylene responsive factors: ERF and disease resistance proteins: DRP. Various transcript variants of these genes were selected wherever available to confirm the accuracy of assembly. qRT PCR analysis revealed similar differential expression pattern for almost all the transcripts through all the eight stages with correlation coefficient R ranging between 0.8 and 0.99. Few transcripts showed variation in the fold change of RNAseq and qRT PCR data at a few stages but still had good correlation (R ≥ 0.7) (Figs 5–9 and Supplementary Table ST5). Transcript variants confirmed the accurate assembly and showed differential expression of these variants from each other through various tissues analyzed (Figs 5–9).

Discussion

Recently transcriptome studies on mango cultivars namely, Zill25, Langra26, Kent27 and Dashehari28 have put forth important information regarding the fruit and leaf physiology of mango. These studies have identified genes encoding multiple enzymes involved in various pathways of primary and secondary metabolism such as, citrate cycle, glycolysis and gluconeogenesis from carbohydrate metabolism; fatty acid biosynthesis, beta oxidation and salicylic acid biosynthesis from fatty acid metabolism; biosynthesis and degradation of various amino acids as well as ethylene biosynthesis from methionine. Genes involved in the flavonoid biosynthesis, vitamin biosynthesis (β-carotene and α-tocopherols) as well as terpenoid backbone synthesis (mevalonate pathway) have also been well explored. In the present study genes involved in all these pathways were identified and their differential expression was also evident through various stages of the Alphonso mango fruit development and ripening (Supplementary Table ST1). In addition the present study revealed some novel findings highlighting better understanding of various processes involved in mango fruit development and ripening and some were unique to the most favoured Alphonso mango fruit, which are discussed below.

Quantitative abundance of terpenes in mangos is well known2, 35. Transcriptome and gene expression studies in mango have explored the terpene biosynthesis pathway till GPP, FPP and GGPP synthesis37–39. In this study, we identified six, five and three genes encoding mono-terpene synthases (MTPS), sesqui-terpene synthases (STPS) and di-terpene synthases (DTPS), respectively that are involved in biosynthesis of specific terpene molecules (Supplementary Figure SF9). These genes were abundant in the flower tissue followed by 30 DAP stage. Further, the transcript abundance of many of these terpene synthase genes in the present study has been depicted to be idiosyncratic to the developing stages, leading to their least expression in the ripening stages of Alphonso fruit (Figs 8 and 9, Supplementary Figure SF7 and Supplementary Table ST2). Previous aroma volatile analysis from Alphonso mango supports this observation wherein flower had the highest concentration of mono-terpenes, oxygenated mono-terpenes and sesqui-terpenes which decreased through the fruit development15.

Another flavor related pathway is lipoxigenase (LOX) followed by HPL pathway37–39, which produces C6 GLVs and lactones through peroxigenase pathway15. The transcriptomes analyzed from Kent and Dashehari mangos reported the presence of six and five genes coding for the LOX family, respectively27, 28. Here, we report detailed annotation of Alphonso mango LOX genes wherein, two genes encode the 9-LOX and six genes encode the 13-LOX. Likewise, Peroxynasenase and epoxide hydrolase (EH) genes have been well-studied for biosynthesis of cutin biopolymer14 and defense related compounds41, 42; while our recent study has shown involvement of these genes in the production of lactones15. In spite of their biological significance, none of the previous
transcriptome studies on mango could determine the presence of peroxygenase and epoxide hydrolase genes. On the other hand in the current study various contigs encoding novel EH (three contigs), EH2 (four contigs) and peroxygenase (three contigs) were detected. Similarly, multiple transcripts encoding enone oxidoreductase and O-methyltransferase having role in furanone biosynthesis were also identified and expression profiles of many of them were shown to be ripening specific (Supplementary Figure SF7). Abundance and ripening related expression of large number of these unique flavor related genes in Alphonso mango (Supplementary Figure SF7) signifies their role in synthesis of diverse aroma volatiles and their unique blend giving sweet and fruity flavor in Alphonso as shown by our previous studies.

Fruit ripening is a complex physiological process and can be characterized by means of fruit softening due to changes in the cell wall structure, increased sugar content by polysaccharide hydrolysis and changes in the aroma volatiles. Starch and pectin are the major storage and structural polysaccharides in the mango fruit, respectively. Various hydrolyases and lyases are known to carry out polysaccharide and cell wall hydrolysis, while amylases degrade the starch in to soluble sugars during ripening. We identified four α-amyrase, three isoamylase and 13 β-amylose coding transcripts. Only one transcript coding for α-amyrase (contig_1439) was down-regulated through flower to 30 DAP fruit transition while others were expressed throughout the fruit developing and ripening stages. On the other hand, transcriptome analysis of Kent mango revealed identification of four β-amylose and three α-amyrase transcripts out of which only two β-amylose coding transcripts were found to be up-regulated in ripe tissue, none of the other showed differential expression.

In addition, we also detected large number of PL, PG and PE encoding transcripts known to be responsible for degradation of complex hetero-polysaccharide, pectin were detected (Supplementary Table ST4). Most of these had stable expression and only few were differentially expressed in Alphonso mango. These results are in contrast to the observations from Kent and Dashehari mango transcriptomic data in terms of the number of unigenes had stable expression and only few were differentially expressed in Alphonso mango. These results signify controlled and steady activity of pectin degradation leading to slow and balanced transitions during Alphonso fruit ripening and this might be one of the reasons for its longer shelf life.

Another interesting observation in our study was, the identification of 79 contigs encoding 20 different inhibitor classes from Alphonso mango (Supplementary Figure SF8). Previously, only three unigenes coding for the cysteine proteinase inhibitors were reported from Langra leaf transcriptome, while no such inhibitors were reported from Zill, Kent and Dashehari mango transcriptomes. The present study suggests Alphonso mango transcriptome to be rich in these inhibitors throughout the fruit development and ripening stages.

Thus, the presence of large numbers of amylase, PL, PE and PG transcripts with very few of them differentially regulated, perpetual expression of most of the starch and cell wall hydrolyzing enzymes along with the persistent presence of inhibitors for amylase, pectinesterase, polygalacturonase and ethylene sensitivity can be cumulatively suggested to play a crucial role in controlled and slow ripening and longer fruit shelf life of Alphonso mango. As a representative example the balance between pectinesterases and their inhibitors during Alphonso fruit development and ripening has been demonstrated in Fig. 10 as heatmap and area chart of expression of PE and PEI. We further observed that phosphate metabolism related gene ontologies were enriched at complete ripe stage (15 DAH) of Alphonso mango. Among these some transcripts were uniquely found at 15 DAH stage (Supplementary Table ST2) suggesting accelerated primary metabolism. Moreover, contig 6946 encoding cytokinin riboside 5'-monophosphate phosphoribohydrolase was uniquely observed at 15 DAH stage. This enzyme plays important role in converting endogenous inactive cytokinin nucleotides to the biologically active free cytokinin responsible for delayed ripening. On the other hand contig_6770 encoding phospholipase-D (PLD) involved in degradation of important cell membrane component i.e. phospholipids was observed to be upregulated (3.4fold).

A key role of PLD in softening was well explained in tomato plants transformed with antisense PLD cDNA which showed delayed ripening and increased firmness in tomato fruits. This counter play at molecular level might be responsible for fine tuning of ripening process in Alphonso mango providing longer shelf life to the fruits.

Additionally oxidative burst, oxidoreductase activities and oxidative stress related gene ontologies were observed during ripening. These factors can generate reactive oxygen species and lead to cell death and fruit damage. Such reactive oxygen species induced cell death is suppressed by Bax inhibitor, which was well studied in Arabidopsis thaliana. Interestingly, four contigs coding for Bax inhibitor in Alphonso mango showed ripening specific expression probably responsible for preventing cell death in ripening fruits and thus longer shelf life. Further detailed study on these inhibitors might help to understand jelly formation in Dashehari mangoes due to excessive ripening and spongy tissue formation in Alphonso mangoes due to uneven ripening.

It is known that ripened fruits are prone to attack by various pathogens. A well distinct defense mechanism was observed in Alphonso mango wherein various defense related proteins (227 contigs) and chitinases (19 contigs) acting on fungal cell wall were differentially regulated (Supplementary Table ST1). Chitinases were found to be accumulated in the flower and in the early fruit developing stages. Insect driven pollination has the risk of fungal infection to flower and further spore accumulation around ovary causing internal infection to the fruit. This might be restricted by the presence of various chitinases in Alphonso mango. Similarly, various disease resistance proteins might play role during fruit development and ripening process to defend infections.

Conclusions

The transcriptome of Alphonso mango analyzed through eight stages of flower to fruit development and ripening transitions revealed various differentially regulated and stage specific genes. Unique transcript profiles probably responsible for distinct and favorable characteristics of Alphonso mango fruit such as flavor, color, ripening duration, skin to stone ripening pattern and longer shelf life were identified and analyzed. This study provides large data sets for further functional validation of fruit ripening process.
Methods

Plant material. Flower and fruits of cv. Alphonso were collected from three independent trees as biological replicates from the Mango Research Sub Centre, Deogad (16.52836 N, 73.344790 E) affiliated to Dr. Balasaheb Sawant Konkan Agricultural University, Dapoli, Maharashtra, India. Flowers from inflorescence were collected and snap frozen. Fruits from developing stages were collected at 30, 60 and 90 days after pollination (DAP). Fruits from 30 and 60 DAP were analyzed as whole fruit in the present study after removing fruit stone; whereas at mature raw stage (90 DAP) pulp (mesocarp) and skin (exocarp) were separated, snap frozen in liquid nitrogen and stored at −80 °C until further analysis. A set of fruits were additionally harvested at their mature raw stage and kept in the hay containing boxes at ambient temperature for ripening and only pulp tissue for ripening stages as table green, mid ripe and ripe were collected at 27 °C for 30 sec followed by a dissociation curve analysis of transcripts. Relative quantification (Applied Biosystem, Carlsbad, CA, USA).

Total RNA isolation was carried out for all the tissues sampled for current study using RNaseq Plus mini kit (Quiagen, Hilden, Germany). RNA quality as 260 nm/280 nm ratio was evaluated using Nanodrop 1000 (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and RNA integrity was checked using Bioanalyzer 2100 (Agilent Technologies, Santa Clara, USA). Two microgram of total RNA was used to carry out reverse transcription for synthesis of cDNA using High Capacity cDNA reverse transcription kit (Applied Biosystem, Carlsbad, CA, USA).

Library preparation and sequencing. High quality RNAs from seven progressive stages of fruit development and ripening as well as flower tissue from single representative biological replicate were sent to the Next Generation Genomics Facility (NGGF) at the Centre for Cellular and Molecular Platforms (C-CAMP), Bangalore for performing transcriptomic sequencing. Briefly, 1 ug of total RNA from each sample was used to prepare eight individual libraries and mRNA was purified using polY oligo beads. The purified mRNA was fragmented in the range of 100 to 140 bases with optimum at around 120 bases and the cDNA was synthesized. End repair, A-Tailing, Adapter ligation and the library preparation were performed using Tru Seq RNA sample preparation kit v2 (Illumina, San Diego, USA) as per manufacturer's instructions. PCR enrichment was performed for 15 cycles and the sample was validated on the Bioanalyzer 2100. Libraries were sequenced in a Paired End 100 base run, using TruSeq SBS Kit v3-HS (Catalog No.: FC-401-3001) for sequencing on the Illumina HiSeq. 1000 platform according to the manufacturer’s recommended protocols. (http://www.illumina.com/systems/hiseq_systems/hiseq_2000_1000/kits.illm).

Bioinformatics data analysis. Paired end RNA sequencing was performed using Illumina Hiseq2000. Quality check on raw data was performed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adapter free, good quality reads (Q >= 30; min read length = 85) were obtained using Cutadapt59. Alphonso transcriptome for each stage was assembled using Velvet-Oase60 with K-mers 67, 75, 83 and merging them at 27 k-mer. Additionally, a merged transcriptome was also generated using k-mer 55. Further, for all the merged assembled transcripts, we used Transdecoder (https://transdecoder.github.io/) to extract potential candidate coding regions within transcripts. Partial cds were discarded and only those transcripts with start and stop codon were made non redundant based upon sequence identity cut-off of 90% using CD-HIT-est61 and used for downstream analyses.

Further, merged full length transcripts (from merged assembly) were used as reference to map back all the reads from each stage using default parameters of Bowtie62. DESeq2 was used to identify differentially expressed transcripts63 and were filtered based upon p-value ≤ 0.05 and expression value >0. All those transcripts having mapping count zero were excluded from further analysis such as in identification of uniquely expressed transcripts in a particular stage or a specific set of stages (e.g. developing and ripening stages). Unique and common list of transcripts were represented using Venny (http://bioinfogp.cnb.cisc.es/tools/venny/) and provided in Supplementary Table ST2. Full length transcripts (from merged assembly) were used as reference for differential expression analysis using DESeq263.

Annotation, enzyme code distribution and GO mapping and InterProScan analyses were carried using BLAST2GO 3.1.3 workbench (Biobam Bioinformatics S.L., Valencia, Spain) as described in the user manual64. GO enrichment analysis was carried out in given test and reference sets by Fisher's exact test in BLAST2GO with p-value (0.001) and FDR filters.

Quantitative reverse-transcriptase PCR. Quantitative reverse-transcriptase PCR was performed using the Fast Start Universal SYBR Green master mix (Roche Inc. Indianapolis, Indiana, USA) and elongation factor 1α (EF1α) as an endogenous reference gene employing the primers reported earlier7. Various transcripts selected from transcriptome data were amplified using gene specific primers (Supplementary Table ST5) and quantification was performed using 7900HT Fast Real-Time PCR System (Applied Biosystems, California, USA) having thermal cycle program of initial denaturation at 95 °C for 10 min with subsequent 40 cycles of 95 °C for 3 sec and 60 °C for 30 sec followed by a dissociation curve analysis of transcripts. Relative quantification (ΔΔCT method) and statistical analysis was carried out using DataAssist™ v3.01 software (Applied Biosystems, California, USA). Eight tissue samples as described above from three independent biological replicates were used for this analysis. For individual transcript the highest expression at a particular fruit development or ripening stage was considered as 1 and expression in other tissues including flower was normalized to that for better graphical presentation.

Data availability. The raw data files generated during the current study will be available in the NCBI repository under the BioProject - PRJNA391381.
References

1. Gil, A. M. et al. Study of the compositional changes of mango during ripening by use of nuclear magnetic resonance spectroscopy. Journal of Agricultural and Food Chemistry 48, 1524 (2000).

2. Pandit, S. S. et al. Cultivar relationships in mango based on fruit volatile profiles. Food Chemistry 114, 363, doi:10.1016/j.foodchem.2008.09.107 (2009).

3. Deshpande, A. B. et al. Data on changes in the fatty acid composition during fruit development and ripening of three mango cultivars (Alphonso, Pairi and Kent) varying in lactone content. Data in Brief 3, 480, doi:10.1016/j.dib.2016.09.018 (2016).

4. de Magalhães Andrade, J. et al. 2D-DIGE analysis of mango (Mangifera indica L.) fruit reveals major proteomic changes associated with ripening. Journal of proteomics 75, 3331 (2012).

5. Fasoli, E. & Righetti, P. G. The peel and pulp of mango fruit: A proteomic samba. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics 1834, 2539 (2013).

6. Renuse, S. et al. Protein profiling of an unsequenced plant-Mangifera indica. Journal of proteomics 75, 5793 (2012).

7. Pandit, S. S. et al. Expression profiling of various genes during the fruit development and ripening of mango. Plant Physiology and Biochemistry 48, 426, doi:10.1016/j.plaphy.2010.02.012 (2010).

8. Pandit, S. S. et al. Genetic diversity analysis of mango cultivars using inter simple sequence repeat markers. Current science 93, 1135 (2007).

9. Kulkarni, R. et al. Characterization of three novel isoprenyl diphosphate synthases from the terpenoid rich mango fruit. Plant physiology and biochemistry 71, 121 (2013).

10. Kulkarni, R. et al. An oxidoreductase from Alphonso mango catalyzing biosynthesis of furaneol and reduction of reactive carbonyls. SpringerPlus 2, 492 (2013).

11. Chidley, H. G. et al. Molecular cloning and characterization of 9-methyltransferase from mango fruit (Mangifera indica cv. Alphonso). Molecular Biotechnology 58, 340, doi:10.1007/s12033-016-9933-2 (2016).

12. Martínez, P. G., Gómez, R. L. & Gómez-Limb, M. A. Identification of an ETR1-homologue from mango fruit expressing during fruit ripening and wound. Journal of Plant Physiology 158, 101 (2001).

13. Sane, V. A., Chourasia, A. & Nath, P. Softening in mango (Mangifera indica cv. Dashehari) is correlated with the expression of an early ethylene responsive, ripening related expansin gene, MIEExpA. Postharvest biology and technology 38, 223 (2005).

14. Singh, R. K., Ali, S. A., Nath, P. & Sane, V. A. Activation of ethylene-responsive p-hydroxyphenylpyruvate dioxygenase leads to increased tocopherol levels during ripening in mango. Journal of experimental botany 62, 3375 (2011).

15. Deshpande, A. R. et al. Isolation and characterization of 9-lipoxygenase and epoxide hydrolase 2 genes: Insight into lactone biosynthesis in mango fruit (Mangifera indica L.). Phytochemistry 138, 65, doi:10.1016/j.phytochem.2017.03.002 (2017).

16. Chidley, H. G. et al. Effect of postharvest ethylene treatment on sugar content, glycosidase activity and its gene expression in mango fruit. Journal of the Science of Food and Agriculture 97, 1624 (2017).

17. Chidley, H. G., Kulkarni, R. S., Pujari, K. H., Giri, A. P. & Gupta, V. S. Spatial and temporal changes in the volatile profile of Alphonso mango upon exogenous ethylene treatment. Food Chemistry 136, 585, doi:10.1016/j.foodchem.2012.08.029 (2013).

18. González-Aguilar, G., Zavaleta-Gatica, R. & Tiznado-Hernández, M. Improving postharvest quality of mango ‘Haden’by UV-C treatment. Postharvest Biology and Technology 45, 108 (2007).

19. Hofman, P. J., Smith, L. G., Joyce, D. C., Johnson, G. I. & Meiburg, G. F. Bagging of mango (Mangifera indica cv. Kent) fruit influences fruit quality and mineral composition. Postharvest Biology and Technology 12, 83 (1997).

20. Jhang, Y. & Joyce, D. Effects of 1-methylcyclopropene alone and in combination with polyethylene bags on the postharvest life of mango fruit. Annals of Applied Biology 137, 321 (2000).

21. Wu, G. et al. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. Nature biotechnology 32, 656 (2014).

22. Consortium, T. G. The tomato genome sequence provides insights into fleshly fruit evolution. Nature 485, 635 (2012).

23. Shulaev, V. et al. The genome of woodland strawberry ( Fragaria vesca ). Nature genetics 43, 109 (2011).

24. Martellini, L. et al. Transcriptome profiling of citrus fruit response to huanglongbing disease. Plos one 7, 1 (2012).

25. Wu, H.-x et al. Transcriptome and proteomic analysis of mango ( Mangifera indica Linn ) fruits. Journal of proteomics 105, 19 (2014).

26. Azim, M. K., Khan, I. A. & Zhang, Y. Characterization of mango ( Mangifera indica L. ) transcriptome and chloroplast genome. Plant molecular biology 85, 193 (2014).

27. Duarte-Castro, M. et al. Mango ( Mangifera indica L. ) cv. Kent fruit mesocarp de novo transcriptome assembly identifies gene families important for ripening. Frontiers in plant science 6, 62 (2015).

28. Srivastava, S. et al. Comparative transcriptome analysis of unripe and mid-ripe fruit of Mangifera indica (var.‘Dashehari’) unravels ripening associated genes. Scientific Reports 6, 1 (2016).

29. Veda, S., Platel, K. & Sinrivasan, K. Varietal differences in the bioaccessibility of β-carotene from mango ( Mangifera indica ) and papaya ( Carica papaya ) fruits. Journal of agricutural and food chemistry 55, 731 (2007).

30. Tharanathan, R., Yashoda, H. & Prabhra, T. Mango ( Mangifera indica L. ), “The king of fruits”—An overview. Food Reviews International 22, 95 (2006).

31. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic acids research 45, D353, doi:10.1093/nar/gkw1092 (2017).

32. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research 28, 27 (2000).

33. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. Nucleic acids research 44, D457, doi:10.1093/nar/gkt1070 (2016).

34. Kulkarni, R. S., Chidley, H. G., Pujari, K. H., Giri, A. P. & Gupta, V. S. Geographic variation in the flavour volatiles of Alphonso mango. Food Chemistry 130, 58, doi:10.1016/j.foodchem.2011.06.053 (2012).

35. Pandit, S. S. et al. Changes in volatile composition during fruit development and ripening of ‘Alphonso’ mango. Journal of the Science of Food and Agriculture 89, 2071, doi:10.1002/jsfa.3692 (2009).

36. Sallaud, C. et al. A novel pathway for sesquiterpene biosynthesis from Z, Z-farnesyl pyrophosphate in the wild tomato Solanum habrochaites. The Plant Cell 21, 301 (2009).

37. Huang, F.-C. & Schwab, W. Cloning and characterization of a 9-lipoxygenase gene induced by pathogen attack from Nicotiana benthamiana for biotechnological application. BMC Biotechnology 11, 1, doi:10.1186/1472-6750-11-30 (2011).

38. Huang, F.-C. & Schwab, W. Overexpression of hydroperoxide lyase, peroxigenase and epoxide hydrolase in tobacco for the biotechnological production of flavours and polymer precursors. Plant Biotechnology Journal 10, 1099, doi:10.1111/j.1476-7652.2012.00739.x (2012).

39. Baysal, T. & Demirdöven, A. Lipoxygenase in fruits and vegetables: A review. Enzyme and Microbial Technology 40, 491, doi:10.1016/j.enzmictec.2006.11.025 (2007).

40. Blee, E. & Schuber, F. Biosynthesis of cutin monomers: involvement of a lipoxygenase/peroxigenase pathway. The Plant Journal 4, 113 (1993).

41. Masui, H., Kondo, T. & Kojima, M. An antifungal compound, 9, 12, 13-trihydroxy-(E)-10-octadecenoic acid, from Colocasia antiquorum inoculated with Ceratocystis fimbriata. Phytochemistry 28, 2613 (1989).

42. Ohita, H. et al. The occurrence of lipid hydroperoxide-decomposing activities in rice and the relationship of such activities to the formation of antifungal substances. Plant and cell physiology 31, 1117 (1990).

43. Atkinson, R. G. et al. Down-regulation of polygalacturonase1 alters firmness, tensile strength and water loss in apple (Malus x domestica) fruit. BMC plant biology 12, 129 (2012).
44. Gross, K. C. & Sams, C. E. Changes in cell wall neutral sugar composition during fruit ripening: a species survey. *Phytochemistry* **23**, 2457 (1984).
45. Lunn, D., Phan, T. D., Tucker, G. A. & Lyckett, G. W. Cell wall composition of tomato fruit changes during development and inhibition of vesicle trafficking is associated with reduced pectin levels and reduced softening. *Plant Physiology and Biochemistry* **66**, 91 (2013).
46. Peroni, F. H. Ga et al. Mango starch degradation. II. The binding of α-amylase and β-amylase to the starch granule. *Journal of agricultural and food chemistry* **56**, 7416 (2008).
47. Lizada, C. In *Biochemistry of fruit ripening* (eds G. Seymour, J. Taylor & G. Tucker) 255 (Chapman and Hall, 1993).
48. Ali, Z. M., Chin, L.-H. & Lazan, H. A comparative study on wall degrading enzymes, pectin modifications and softening during ripening of selected tropical fruits. *Plant Science* **167**, 317 (2004).
49. Chourasia, A., Sane, V. A. & Nath, P. Differential expression of pectate lyase during ethylene-induced postharvest softening of mango (*Mangifera indica* var. Dashehari). *Physiologia Plantarum* **128**, 546 (2006).
50. Fischer, R. L. & Bennett, A. B. Role of cell wall hydrolases in fruit ripening. *Annual review of plant biology* **42**, 675 (1991).
51. Kurakawa, T. et al. Direct control of shoot meristem activity by a cytokinin-activating enzyme. *Nature* **445**, 652, doi:10.1038/nature05504 (2007).
52. Ludford, P. In *Postharvest Physiology and Pathology of Vegetables* (CRC Press, 2002).
53. Grittle Pinhero, R., Almoquist, K. C., Novotna, Z. & Paliyath, G. Developmental regulation of phospholipase D in tomato fruits. *Plant Physiology and Biochemistry* **41**, 223, doi:10.1016/S0981-9428(03)00014-7 (2003).
54. Kawai-Yamada, M., Ohori, Y. & Uchimiy, H. Dissection of Arabidopsis Bax inhibitor-1 suppressing Bax–, hydrogen peroxide–, and salicylic acid–induced cell death. *The Plant Cell* **16**, 21 (2004).
55. Shivashankar, S., Ravindra, V. & Louis, H. Biochemical changes in seed and mesocarp of mango (*Mangifera indica* L.) cv. Alphonso and their significance during the development of spongy tissue. *The Journal of Horticultural Science and Biotechnology* **82**, 35 (2007).
56. Srivastav, M., Singh, S. & Ajang, M. Evaluation of mango genotypes for jelly seed disorder. *Indian Journal of Horticulture* **72**, 408 (2015).
57. Coates, L. & Johnson, G. Postharvest diseases of fruit and vegetables. *Plant pathogens and plant diseases* 533 (1997).
58. Daulagala, P. Expression of chitinase with antifungal activities in ripening banana fruit. *Tropical Plant Research* **1**, 72 (2014).
59. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBO journal* **17**, 10 (2011).
60. Schulz, M. H., Zerbino, D. R., Vingron, M. & Birney, E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**, 1086 (2012).
61. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658 (2006).
62. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**, R25.1 (2009).
63. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq. *2*. *Genome biology* **15**, 1 (2014).
64. Conesa, A. et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674 (2005).

**Acknowledgements**

The authors would like to thank Dr. Aarti Desai for helpful discussion and Dr. Abhay Jere (Persistent Systems Ltd., Pune, India) for his critical suggestions, guidance and support. The authors acknowledge the Next Generation Genomics Facility (NGGF) at the Centre for Cellular and Molecular Platforms (C-CAMP), Bangalore (Project ID: BT/PR3481/INF/22/140/2011) for performing transcriptome sequencing of the eight mango tissues. Authors would like to thank Dr. Sanjiv Galande, Dr. Ram Kulkarni and Mr. Saurabh Pradhan (Indian Institute of Science Education and Research, Pune, India) for their help in evaluation of RNA quality by Bioanalyzer 2100. This research was funded by the Council of Scientific and Industrial Research, New Delhi, India under the project CSC0133 (FUNHEALTH) to CSIR-NCL and 21(0997)/16/EMR-II grant as Emeritus Scientist scheme to VSG.

**Author Contributions**

A.B.D. designed, standardized and performed majority of the experiments, analyzed the data and wrote the manuscript. K.A. and V.Z. carried out bioinformatics analysis. H.G.C. and P.S.O. helped with high quality RNA isolation, cDNA synthesis and qRT PCR studies. K.H.P. contributed by providing the experimental tissue. A.P.G. contributed with suggestions for the work and for manuscript planning, while NYK contributed in bioinformatic analysis and manuscript writing. V.S.G. designed the study, participated in all parts of the work and contributed to the preparation of the manuscript. All authors read and approved the final manuscript.

**Additional Information**

**Supplementary information** accompanies this paper at doi:10.1038/s41598-017-08499-5

**Competing Interests:** The authors declare that they have no competing interests.

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