Molecular analysis of anthocyanin biosynthesis pathway genes and their differential expression in mango peel

Journal:	<i>Genome</i>
Manuscript ID	gen-2017-0205.R1
Manuscript Type:	Article
Date Submitted by the Author:	30-Nov-2017
Complete List of Authors:	Bajpai, Anju; ICAR-Central Institute for Subtropical Horticulture, Division of Crop Improvement and Biotechnology Khan, Kasim; ICAR-Central Institute for Subtropical Horticulture, Division of Crop Improvement and Biotechnology Muthukumar, M; ICAR-Central Institute for Subtropical Horticulture, Division of Crop Improvement and Biotechnology Rajan, S; ICAR-Central Institute for Subtropical Horticulture, Director Singh, Nagendra; ICAR-National Research Centre for Plant Biotechnology, NRCPB
Is the invited manuscript for consideration in a Special Issue? :	N/A
Keyword:	Alternate splice variants, Anthocyanin, Gene expression, Peel color, Mango

SCHOLARONE™
Manuscripts

Molecular analysis of anthocyanin biosynthesis pathway genes and their differential expression in mango peel

Anju Bajpai^{1*}, Kasim Khan¹, Muthukumar, M.¹, S. Rajan¹ and N.K Singh²

¹ICAR-Central Institute for Subtropical Horticulture, Lucknow-226101

²ICAR-National Research Centre on Plant Biotechnology, Pusa Campus, New Delhi-110012

***Corresponding author**

Dr. Anju Bajpai, anju.bajpai@gmail.com

Coauthors

Dr Kasim Khan, kasimkhan123@yahoo.com

Dr.Muthukumar,M., muthukumarbt@gmail.com

Dr.S.Rajan, srajanlko@gmail.com

Dr.N.K Singh, nksingh4@gmail.com

Abstract

Mango fruit is cherished by masses for its taste and nutrition, contributed by color, flavor and aroma. Among these, peel color is an important trait contributing to fruit quality and market value. We attempted to elucidate the role of key genes of anthocyanin biosynthesis pathway related to fruit peel color from the leaf transcriptome of cv. Amrapali. A total of 108 mined transcript sequences were assigned to phenylpropanoid flavonoid pathway from which 15 contigs representing anthocyanin biosynthesis genes were annotated. Alternate splice variants were identified by mapping against genes of *Citrus clementina*/ *Vitis vinifera* (closest relatives) and determined the protein subcellular localization. Phylogenetic analysis of these pathway genes clustered them into distinct groups aligning with homologous genes of *Mangifera indica*, *C. clementina* and *Vitis vinifera*. Expression profiling revealed higher relative fold expressions in mature fruit peel of red colored varieties (Arunika, Ambika and Tommy Atkins) in comparison with green peeled Amrapali. *MiCHS*, *MiCHI* and *MiF3H* alternate splice variants revealed differential gene expression. Functionally divergent variants indicate availability of allelic pool programmed to play critical roles in peel color. This study gave insights into molecular genetic basis of peel color and offers scope for development of biomarkers in varietal improvement programs.

Keywords: Alternate splice variants, anthocyanin, gene expression, peel color, mango

Introduction

Mango (*Mangifera indica* L.), a native fruit tree species to Indian subcontinent has its evolutionary route in the tropical rainforests of South and South-East Asia (Kostermans and Bompard 1994). Being acclaimed as King of fruits, it is well-known for its taste, flavor, aroma and nutritive value. Moreover, extensive natural selections over years has evolved genetic variability in fruit size, shape, color, flavor, seed size and pulp related traits. Many mango varieties retain green peel color even at fully ripened stage, while some acquire red peel. Red coloration in flowers, fruits, and other plant tissues is associated with anthocyanin (red pigments), which has multifarious roles such as conferring plant disease resistance and protection against UV-radiation (Bieza and Lois 2001; Sivankalyani et al. 2016). Anthocyanins also provide human health care benefits against cancer, cardiovascular and other chronic diseases (Rao and Rao 2007; Butelli et al. 2008; Singh et al. 2008).

Attractive mango peel color is considered one of the most important factors for export markets (Nambi et al. 2016; Sivankalyani et al. 2017). The most predominant anthocyanins identified in mango are cyanidin-3-glucosides and 7-methylcyanidin-3-galactosides (Berardini et al. 2005). Accumulation of pigments, their concentration and the intensity determines the overall appearance of color in the epicarp portion of fruits and flowers in mango (Martin et al. 2017; Pervaiz et al. 2017). Besides pigments, metabolites like mangiferin, flavonoids and polyphenols are reported in mango peel (Nordey et al. 2014). Recent developments in NGS technologies, has facilitated mining RNA seq data for identifying gene families regulating metabolic pathways which governs phenotypic expression of specific traits. Transcriptome and proteome analysis by Wu et al. (2014) and Dautt-Castro et al. (2015), helped in identification of ripening related genes from mango fruit mesocarp of cv. Zill and Kent, respectively. Similarly, Luria et al. (2014) utilized transcriptome analysis to identify disease resistance genes associated with post harvest hot

water treatment in mango cv. Shelly. Even though genes associated with the physiology and biochemistry of fruit biology related traits have been studied extensively related to post harvest treatments and fruit ripening (Hoang et al. 2015; Pandit et al. 2010), no systematic study on peel color genes is existing. The anthocyanin biosynthetic pathway (Fig. 1) beginning with the precursor phenylalanine results in anthocyanin accumulation *via* a series of steps involving specific enzymes (Dixon and Steele, 1999). Utilizing this information, the present study was carried out exploring transcriptome data for elucidating the molecular genetic basis of peel coloration in mango in different genetic backgrounds.

Materials and methods

Genetic materials

Whole genome leaf transcriptome profiling through RNA sequencing was performed from mango variety 'Amrapali' (*Mangifera indica* L.) collected from the mango orchard at ICAR-Central Institute for Subtropical Horticulture, located in Lucknow (26.9168° N, 80.7076° E, 128.87 m above sea level). Fruit peel was collected from four varieties *viz.*, Amrapali, Tommy Atkins, Ambika and Arunika that are characterized by variations in peel color at fully ripened stage (Fig. 2).

Estimation of anthocyanin content in mango peel

One gram of fruit pericarp was extracted in 50ml of ethanol containing 1%HCL, as per standard protocol (Ranganna, 1997). Absorbance was measured at 535 nm using Chemito Spectrophotometer UV2100. Results were expressed as mg of total anthocyanin content per 100 g fresh weight.

Leaf transcriptome

Total RNA was isolated from Amrapali young leaves through Spectrum™ Plant Total RNA Kit (Sigma USA) and quality of total RNA was confirmed by agarose gel and quantified using spectrophotometer, QIAxpert (Qiagen Germany). Total RNA (4 µg) was

subjected to library preparation using Illumina® TruSeq® RNA Sample Preparation V2 kit as per protocol. The mRNA was enriched and fragmented enzymatically and used for first and second strand cDNA conversion, followed by end repair, A-tailing and adapter ligation, and finally by index PCR amplification of adaptor-ligated library. Library quantification and quality analysis was performed using HT DNA High Sensitivity Assay Kit.

cDNA libraries were sequenced using an Illumina NextSeq 500/MiSeq platform using 2 x 150 PE chemistry. These reads were subjected to quality filtration at QV 20 (mean quality score ≥ 20), adapter trimming using Trimmomatic version v0.30 and high quality reads were assembled using CLC genomics workbench v6.0). These final assembled contigs were annotated using CANoPI (Contig Annotator Pipeline) for *de novo* transcriptome assembly. These transcripts were annotated using BLASTX program v2.2.24 to identify the probable genes based on e-value $\leq 1e-5$ and similarity score $\geq 40\%$, used for Uniprot protein annotation and gene ontology classification into classes of molecular function, biological process and cellular component categories. The completely annotated transcriptome sequence data was deposited in the NCBI database under SRP070908 (Bioproject PRJNA313340).

Computational analysis for alternate splice site based structural predictions, transcript variants and protein localization

The contigs identified in the leaf transcriptome were examined for mining the genes involved in anthocyanin biosynthetic pathway (Table S1). Computational analysis of the selected contigs was performed to predict transcript variants based on alternate splicing sites and the protein functional predictions based on sequence conversion from transcript sequences to amino acid/protein sequences (<http://web.expasy.org/translate/>). The transcript sequences were annotated using BLAST analysis to predict genes coding sequences. These transcript sequences were mapped onto the genomic regions using the top score hit orthologous gene sequences of either *Citrus clementina* or *Vitis vinifera* employing

EST2genome tool (<http://www.hpa-bioinfotools.org.uk/pise/est2genome.html>) for prediction of the alternate splicing events and identifying transcript variants. Multiple sequence alignment of the protein sequences as described by Edgar (2004) were carried out using MUSCLE (<http://www.ebi.ac.uk/Tools/msa/muscle/>) for phylogenetic analysis using phylogeny tool (<http://www.phylogeny.fr/>) as per the protocol described by Dereeper et al. (2008). WoLF-PSORT (<https://www.genscript.com/wolf-psort.html>) was used for protein subcellular localization which classified proteins into more than 10 localization sites including proteins which shuttle between the cytosol and nucleus. It was reconfirmed using Multiloc2-high resolution prediction tool accessible at <https://abi.inf.uni-tuebingen.de/Services/MultiLoc2> (Blum et al. 2009) and Sherloc2 tool (Version 2) accessible from <https://abi.inf.uni-tuebingen.de/Services/SherLoc2> (Briesemeister et al. 2009).

Expression profiling of genes involved in anthocyanin biosynthetic pathway through Real-time (qRT-PCR) analysis

RNA from the fruit peel samples of four varieties *viz.*, Amrapali, Ambika, Arunika and Tommy Atkins were isolated through protocol 2 of Spectrum™ Plant Total RNA Kit (Sigma USA) involving DNase step. Total RNA (3 µg template) was converted into first strand cDNA using Maxima First Strand cDNA Synthesis Kit (Thermo Scientific USA). Primers for the selected genes were designed using NCBI Primer designing tool (Table S2). Real Time (qRT-PCR) assay was performed using the peel first strand cDNA samples of four mango varieties as template (10 ng), with standard conditions of forward primer (0.5 µM), reverse primer (0.5 µM) and SYBR Green mix (6 µl) and running conditions of pre PCR at 60 °C for 30s, holding at 95°C for 10 min and cycling at 95°C for 15s followed by 60 °C for 1 min in the real time instrument (ABI Fast 7500, NY, USA). *MiActin*, a housekeeping gene was considered as an endogenous reference gene - for data normalization. Data normalization was done using C_t values of the genes of experimental samples (varieties) with their

respective reference gene (ΔC_t) and comparison of expressed genes with respect to each other was calculated as $\Delta\Delta C_t$ and the overall gene expression levels in terms of fold change was calculated using the formula; $\text{Fold Change} = 2^{(-\Delta\Delta C_t)}$.

Results and Discussion

Whole genome leaf transcriptome profiling in mango variety Amrapali

Mango genome size is about 439 Mbp and till date its draft genome sequence only has been reported (Singh et al. 2016). Leaf transcriptome profiling of Amrapali produced 54,207,725 raw reads (15.03 GB data), which were assembled into 50,945 contigs with an average length of 1004.51 bp, (Fig. S 1, A & B). The functional annotation of the assembled contigs using BLASTX predicted 43,037 CDS with their genomic annotations for only 31,834 contigs with e-value less than $1e-5$. Most of the top hits were matching with the genomic sequences of *Citrus clementina*, *Theobroma cocoa*, *Populus trichocarpa*, *Rhus chinensis*, and *Vitis vinifera* (Fig. S1, C). A total of 54% (approx. 23,239 out of total 44,472) genes were annotated with *Citrus clementina*. These results of transcriptome data statistics corroborates with the earlier reports where nearly 25000 expressed genes could be annotated in mesocarp of 'Dashehari' (Srivastava et al. 2016) and 'Kent' (Dautt-Castro et al. 2015). Similar results have been reported in fruit of 'Zill' (Wu et al. 2014), peel of 'Shelly' (Luria et al. 2014), leaf of 'Langra' (Azim et al. 2014) and leaves of Neelam, Dashehari and Amrapali (Mahato et al. 2016). More recently, Srivastava et al. (2016) also confirmed maximum match of Dashehari transcriptome with that of *Citrus* sp.

Computational analysis of anthocyanin biosynthetic genes mined from transcriptome data

Phenylpropanoid and flavonoid pathway is a complex network which includes the pathway leading to anthocyanin biosynthesis. Amrapali leaf transcriptome data revealed a series of CDS (Coding Sequence) encoding the genes of anthocyanin biosynthetic pathway

and a few distinctly annotated in the BLASTX analysis were shortlisted (Table 1). These structural genes are classified into early and late anthocyanin biosynthetic genes; former including *CHS*, *CHI*, *F3H* and *FLS* while latter encompass *DFR*, *ANS*, *ANR* etc. (Pelletier et al. 1997; 1999). A total of 15 candidate genes encoding for phenylalanine ammonia lyase (*PAL*), cinnamate 4-mono-oxygenase (*C4'H*), chalcone synthase (*CHS1*, 2 & 3), chalcone isomerase (*CHI* 1,2, 3, 4, 5 & 6), flavonoid 3-hydroxylase (*F3'H1*, *F3'H2*), flavonoid 3'5'-hydroxylase (*F3',5'H*), and dihydroxy-flavone reductase (*DFR*) were identified. These transcripts were used for structural predictions of transcript variants and probable protein subcellular localization. Transcriptome data set had only single transcripts for *MiPAL* and *MiC4H*, the starting genes of anthocyanin biosynthetic pathway. In the following steps of the pathway, a set of 3 transcripts encoding chalcone synthase (*MiCHS1*, *MiCHS2*, *MiCHS3*), 6 transcripts encoding chalcone isomerase (*MiCHI1*, *MiCHI2*, *MiCHI3*, *MiCHI4*, *MiCHI5*, *MiCHI6/FAP*), 3 encoding flavanone 3-hydroxylase (*MiF3'H1*, *MiF3'H2*, *MiF3',5'H*) and 1 encoding dihydro-flavone reductase (*MiDFR*) were identified (Table 2). Phylogenetic analysis of these identified genes further confirmed the families of these structural genes (Fig. 3). Among these enzymes, PAL is the initial enzyme of the pathway mediating carbon flux into the phenylpropanoid pathway to produce cinnamic acid, the substrate for the next step mediated by *C4'H*. In this study, a partial *MiPAL* gene sequence of 369 bp showed 94% identity at nucleotide level and 98% at amino acid level with *Rhus chinensis*. Earlier reports have also shown that the full length *MiPAL* encodes a protein of 707 amino acids with maximum 92 and 87% identity with *Rhus* and *Populus* PAL, respectively (Hoang et al. 2015). *C4'H*, and *4CL* are downstream genes to PAL activity, which provide the basis for metabolite synthesis leading to phenylpropanoid monomers. In this study, *MiC4H* (505 amino acids) shared 97% identity at nucleotide level with *Populus trichocarpa* x *Populus deltoides* (AAG50231) and at protein level shared 85% homology with *Citrus sinensis*.

Earlier sequence analysis of the cloned C4'H ortholog in *Hibiscus* revealed its length to be 1,518 bp ORF encoding 505 amino acids similar to *MiC4'H* (Kim et al. 2013).

Cluster analysis of chalcone synthase transcripts revealed that *MiCHS2* (CDS149) and *MiCHS3* (CDS41329) was aligned with *M.indica CHS1* (KF929407.1). *Citrus clementina* CHS (10028604m), *Mangifera indica CHS2* (KF929408.1) and *MiCHS1* (CDS38624) were found to be outgrouped (Fig. S2) even at protein level (Fig. S3). This was also evident from the alternate splice variants predictions where additional exonic segment was obtained in 5'-end. Gene level predictions also revealed that *MiCHS3* was a truncated exonic portion (686bp) representing exon 2 of *MiCHS2* (969 bp) (Fig. S3). The genes encoding CHS constitute a multigene family (type III polyketide synthase superfamily), catalyzing the first committed step in the PF pathway of plants for biosynthesis of flavonoids, isoflavonoids, and anthocyanins (Ferrer et al. 1999). Two CHS cDNAs in Citrus, sharing 86.6% identity at the amino-acid level, with unknown function have earlier been identified (Moriguchi et al. 1999; Lu et al. 2009). These results indicate that *CHS* gene family in fruit crops is tightly regulated at transcriptional level.

Chalcone isomerase (CHI, EC 5.5.1.6), is another multigene family involved in the initial phase of the phenylpropanoid/flavonoid biosynthesis pathway. It catalyzes the cyclization of chalcone (4,2',4',6'-tetrahydroxychalcone) and 6'-deoxychalcone (4,2',4'-trihydroxychalcone) into (2S)-flavanones, (naringenin and liquiritigenin) operating downstream to CHS. *MiCHI1* (CDS10544) and *MiCHI2* (CDS2356) in protein BLAST analysis matched with *Rhus chinensis* CHI (AGH13331) with 93 and 90% identity, respectively. Even in the alternate splice variations prediction, both these transcripts were matching maximum with *Citrus clementina* CHI (10032697) which was also evident from the phylogenetic analysis of *CHI* transcripts (Fig. 4 & Fig. S2). Previous studies have established that the CHI family comprises four subclasses: Type I CHI proteins are mainly involved in

the formation of 5,7,4'-trihydroxyflavanone and are ubiquitous in vascular plants. While Type II CHIs are mostly found in leguminous plants and catalyse the formation of 5-deoxy (iso) flavonoids (Shimada et al. 2003; Ralston et al. 2005). Type III CHIs are fatty acid binding proteins (FAPs) and widely distributed in land plants and green algae (Ngaki et al. 2012). Type IV CHIs are known as CHI-like proteins (CHILs) which are only found in land plants (Ralston et al. 2005; Ngaki et al. 2012). Interestingly, *MiCHI4* (CDS1818) and *MiCHI5* (CDS1784) matching with *Citrus clementina* CHI (ESR49159) could not be mapped to the genomic regions because of large introns. Instead, the next closest genomic sequence of *Vitis vinifera* CHI (GSVIVG01032685001) was able to predict that maximum sequence identity and alternate splice variants at the signal peptide regions (N-terminal) which targeted the two proteins to mitochondria and cytoplasm, respectively (Fig. 4, C). *MiCHI6/FAP* matched with *Arabidopsis thaliana* FAP (AT1G53520). All these transcript splice variants were grouped into separate subgroups in the phylogenetic analysis (Fig. 4, A; Fig. S2). The present study could identify two distinct classes of CHI, i.e., Type I and Type III. Cluster analysis at cDNA and protein level also grouped 6 *MiCHI* transcripts into three distinct classes.

Flavonoid 3'-hydroxylase (F3'H) and flavonoid 3',5'-hydroxylase (F3'5'H), further downstream enzymes control the hydroxylation at the 3' or the 3' and 5' position in the B ring of flavonoids, respectively (Wang et al. 2014; Zhou et al. 2016). They belong to cytochrome P450 superfamily with subfamilies of CYP75B and CYP75A (Seitz et al. 2007) are well characterized in several plants *Antirrhinum majus*, *Dianthus caryophyllus* and *Vinca major* (Bogs et al. 2006; Mori et al. 2004). *MiF3H* (Fig. S4) and *MiDFR* also were grouped into separate groups as per sequence homology and function (Fig. 3). In subcellular protein localization predictions, most of the proteins pertained to -cytoplasmic location while few were identified for their sorted location in endoplasmic reticulum/secretory protein (*MiPAL*, *MiC4H*), mitochondria (*MiF3'H1*, *MiF3'5H*) and chloroplast (*MiF3'H2*). Several reports

have shown PAL and C4H proteins trafficked to the endoplasmic reticulum (Ro and Douglas, 2000) and periplasm through secretory pathway (Hooper et al. 2016; Yang et al. 2016).

Variability of the varieties for peel color and anthocyanin content

The varieties used in the study possess characteristic peel trait under open canopy conditions, viz. Amrapali (popular Indian hybrid variety with green peel that turns yellow on bagging), Tommy Atkins (popular American variety from Florida with red peel), Ambika (newly developed Indian hybrid that is purple green at maturity and turns bright red on ripening), Arunika (hybrid variety from India with attractive red fruit color) (Fig.2, A). Anthocyanin profiling of the peel from these 4 varieties revealed that anthocyanin content was higher in the the colored varieties with maximum recorded in Arunika, followed by Ambika and Tommy Atkins while green colored Amrapali recorded the minimum (Fig.2,B). This proved our hypothesis of the anthocyanin accumulation in the red peel colored mango varieties. Most recently, Nambi et al. (2016) has described color kinetics to differentiate colored mango varieties to study the peel color and Sivankalyani et al. (2016a) has demonstrated increased anthocyanin and flavonoid accumulation in mango associated with cold and pathogen resistance. It is speculated that the variation in the pigmentation pattern and the fruit peel coloration of the colored varieties could also be attributed to the co-expression of other metabolites (Nordey et al. 2014) but predominantly contributed by anthocyanin accumulation.

Expression profiling of anthocyanin biosynthetic genes and elucidation of their role in peel coloration

Even though, red peel is desirable attribute, the mechanisms of its pigmentation and distribution pattern along with its physiological functions are poorly understood in mango. Therefore, expression profiling involving 1 green (Amrapali) and 3 red (Tommy Atkins, Ambika and Arunika) mango varieties to elucidate the mechanism of peel coloration was

performed using Real time PCR primers for all the transcript variants/genes related to the anthocyanin biosynthetic pathway. Recent studies by Hoang et al. (2015) have also demonstrated that phenylpropanoid flavonoid (PF) pathway genes are critical for extent of peel color variation and are greatly influenced by the genetic diversity of downstream genes of the pathway as well as their expression levels.

In the present study, overall trend in the gene expression profiling indicated that the most of the genes of anthocyanin biosynthetic pathways were up-regulated in higher folds in the colored mango varieties and to the highest extent in Arunika, followed by Ambika and Tommy Atkins (Fig.5). *MiPAL* is the first gene of the anthocyanin biosynthetic pathway and among the three red colored varieties, Arunika showed higher fold expression (200 folds) in comparison with the Tommy Atkins and Ambika, with only 40 and 5 fold expression, respectively. Earlier, Hoang et al. (2015) has reported transcript abundance of PAL gene in terms of relative fold expression varying between 12 - 45 folds in the peels of varieties viz., Kensington Pride, Irwin and NamDocMai. *MiPAL* activity was reported to be under transcriptional and post transcriptional control in mango variety 'Ataulfo' (Palafox-Carlos et al. 2014). Similar expression pattern was obtained with *MiC4H*, the next enzyme in the pathway. Interestingly, the expression levels of three chalcone synthase transcripts viz., *MiCHS1* (CDS38264), *MiCHS2* (CDS149) and *MiCHS3* (CDS41329) showed highest fold expression in Ambika followed by Arunika and Tommy Atkins. Already, computational analysis of transcript structures have evinced that *MiCHS3* is truncated form of *MiCHS2* and hence the same expression pattern is expected. Earlier two members of *CHS* have been reported in Kensington Pride, Irwin and Nam Doc Mai (Australian mango varieties) using EST library (Hoang et al. 2015). Previous studies have indicated that *CHS* is up-regulated by light (Feng et al. 2010), low temperature (Crifo et al. 2011; Piero et al. 2005), and UV-B (Ubi et al. 2006). Thus, the *CHS* in fruit tree crops seems to be environmentally influenced and

under different transcriptional controls. Such regulation is also evident from our study in mango. Transcript abundance of *CHI* was reported to be higher in red apple cultivars Fuji and Jonathan than yellow cv. Orin (Honda et al. 2002).

Five alternate splice variants of *MiCHI* showed differential expression patterns. Among these, three variants *i.e.*, *MiCHI* 1, 3, and 5 exhibited higher fold expression levels in Arunika demonstrating its rate-determining role in the production of flavonols similar to red color peel reported in tomato (Bovy et al. 2002). Furthermore, downstream genes of the pathway *viz.*, *F3'H*, *F3',5'H* and *DFR* displayed high expression in peel of red colored varieties. Identification and high activity of flavonoid 3'hydroxylase (*F3'H*) and flavonoid 3'5'hydroxylase (*F3',5'H*) in red peel is indicative of cyanindin and delphinidin precursors in mango peel. High upregulation of *MiF3',5'H* (CDS9124) and *MiF3'H2* (CDS4342) in Arunika, while *MiF3'H1* (CDS795) in Ambika was seen. Transcript abundance for *MiCHS1* (CDS38624), *MiCHI2* (CDS2356) and *MiF3'H1* (CDS795) genes in Ambika indicated key regulatory role of these genes for peel color. Additionally, Tommy Atkins displayed higher accumulation of naringen chalcone which was evident from higher expression profiles of *PAL*, *C4H* and *CHS* and relatively lesser levels in the downstream genes from *CHI* (Fig. 5). These gene families have been found to be differentially regulated in a systematic and coordinated manner there by exhibiting variant expression patterns in different genetic backgrounds in ripened mango peel supporting the results previously reported by Yamazaki et al. (2008) in red and green leaved perilla. Higher *MiDFR* and *MiANS* transcripts in red color varieties was also registered in mature fruit peel confirming up-regulation of these structural genes in red colored varieties.

Conclusion

The availability of coloration related genetic framework of anthocyanin pathway genes offers use of genomics for signal transduction systems particularly with reference to

environmental conditions in perennial fruit species like mango. The present study confirmed role of coordinated expression of anthocyanin structural genes in indigenous red peel varieties under North Indian conditions. The expression patterns of alternate splice variants in mango suggests presence of functionally diverse isoforms, most of which were up-regulated in red fruited varieties and strongly associated with high anthocyanin content. Availability of functionally divergent isoforms can be assigned to species like mango that synthesizes different classes of polyphenols and indicates availability of pool of alleles programmed to play critical roles in plant secondary metabolism besides peel color in mango. As anthocyanins have bioactive properties important for human health, high fold expression of pathway genes responsible for anthocyanin accumulation adds value to nutrition benefits of mango fruits, peel of which can be utilized as functional foods.

Draft

Author contributions

Conceived and designed the experiments: AB.

Performed the experiments: KK

Analyzed the data: AB, KK, MM

Contributed reagents/ resources/analysis tools: AB, SR

Wrote the paper: KK, MM and AB.

Overall review and supervision: SR and NKS

Funding: NKS and AB

Acknowledgments :

Authors are thankful to ICAR and Director ICAR-CISH for funding the project and providing necessary support. Thanks are also due to Director NBFGR, Lucknow for sharing their lab facilities to carry out RT- PCR experiments.

References

- Azim, M.K., Khan, I.A., and Zhang, Y. 2014. Characterization of mango (*Mangifera indica* L.) transcriptome and chloroplast genome. *Plant Mol. Biol.*, **85**:193-208. doi: 10.1007/s11103-014-0179-8.
- Berardini, N., Knodler, M., Schieber, A., and Carle, R. 2005. Utilization of mango peels as a source of pectin and polyphenolics. *Innovative Food Science & Emerging Technologies*, **6**:442-452. <https://doi.org/10.1016/j.ifset.2005.06.004>.
- Bieza, K., and Lois, R. 2001. An *Arabidopsis* mutant tolerant to lethal ultraviolet-B levels shows constitutively elevated accumulation of flavonoids and other phenolics. *Plant Physiol.*, **126**: 1105-1115.
- Blum, T., Briesemeister, S., and Kohlbacher, O. 2009. MultiLoc2: integrating phylogeny and Gene Ontology terms improves subcellular protein localization prediction. *BMC Bioinformatics*, **10**:274. doi: 10.1186/1471-2105-10-274.
- Bogs, J., Downey, M.O., Harvey, J.S., Ashton, A.R., Tanner, G.J., and Robinson, S.P. 2005. Proanthocyanidin synthesis and expression of genes encoding leucoanthocyanidin reductase and anthocyanidin reductase in developing grape berries and grapevine leaves. *Plant Physiol.*, **139**:652-663. doi:10.1104/pp.105.064238.
- Bovy, A., Vos, R.D., Kemper, M., Schijlen, E., Pertejo, M.A., Muir, S., Collins, G., Robinson, S., Verhoeyen, M., and Hughes, S. 2002. High-flavonol tomatoes resulting from the heterologous expression of the maize transcription factor genes LC and C1. *The Plant Cell*, **14**:2509-2526.
- Briesemeister, S., Blum, T., Brady, S., Lam, Y., Kohlbacher, O., and H. Shatkay. 2009. SherLoc2: a high-accuracy hybrid method for predicting subcellular localization of proteins. *J. Proteome Res.*, **8(11)**: 5363–5366. doi: 10.1021/pr900665y.

- Butelli, E., Titta, L., Giorgio, M., Mock, H.P., Matros, A., Peterek, S., Schijlen, E.G.W.M., Hall, R.D., Bovy, A.G., and Luo, J. 2008. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat. Biotechnol.*, **26**:1301-1308. doi:10.1038/nbt.1506.
- Crifo, T., Puglisi, I., Petrone, G., Recupero, G.R., and Piero, A.R.L. 2011. Expression analysis in response to low temperature stress in blood oranges: implication of the flavonoid biosynthetic pathway. *Gene*, **476**:1-9. doi: 10.1016/j.gene.2011.02.005.
- Dautt-Castro, M., Ochoa-Leyva, A., Contreras-Vergara, C.A., Pacheco-Sanchez, M.A., Casas-Flores, S., Sanchez-Flores, A., Kuhn, D.N., and Islas-Osuna, M.A. 2015. Mango (*Mangifera indica* L.) cv. Kent fruit mesocarp *de novo* transcriptome assembly identifies gene families important for ripening. *Front. Plant Sci.*, **6**: 62. doi:10.3389/fpls.2015.00062.
- Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.M., and Gascuel, O. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucl. Acids Res.*, **1**: 36 (Web Server issue):W465-9. Epub 2008 Apr 19. doi: 10.1093/nar/gkn180.
- Dixon, R.A., and Steele, C.L. 1999. Flavonoids and isoflavonoids- a gold mine for metabolic engineering. *Trends Plant Sci.*, **4**:394-400.
- Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.*, **32**(5): 1792–1797. doi: 10.1093/nar/gkh340.
- Ferrer, J.L., Jez, J.M., Bowman, M.E., Dixon, R.A., and Noel, J.P. 1999. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nat. Struct. Mol. Biol.*, **6**:775-784.
- Hoang, V.L.T., Innes, D.J., Shaw, P.N., Monteith, G.R., Gidley, M.J., and Dietzgen, R.G. 2015. Sequence diversity and differential expression of major phenylpropanoid-flavonoid

biosynthetic genes among three mango varieties. BMC Genomics, **16**:1. doi: 10.1186/s12864-015-1784-x.

Honda, C., Kotoda, N., Wada, M., Kondo, S., Kobayashi, S., Soejima, J., Zhang, Z., Tsuda, T., and Moriguchi, T. 2002. Anthocyanin biosynthetic genes are coordinately expressed during red coloration in apple skin. Plant Physiol. Biochem., **40**: 955-962.

Hooper C. M., Castleden, I.R., Aryamanesh, N., Jacoby, R.P., and Millar, A.H. 2016. Finding the subcellular location of barley, wheat, rice and maize proteins: The compendium of crop Proteins with Annotated Locations (cropPAL). Plant Cell Physiol., **57**(1): e9. doi: 10.1093/pcp/pcv170.

Kim, J., Choi, B., Natarajan, S., and Bae, H. 2013. Expression analysis of kenaf cinnamate 4-hydroxylase (*C4H*) ortholog during developmental and stress responses. Plant Omics, **6**:65. <http://search.informit.com.au/documentSummary;dn=226840348112203;res=IELHSS> ISSN: 1836-0661.

Kostermans, A.J.G.H., and Bompard, J.M. 1993. The Mangoes: Their Botany, Nomenclature, Horticulture and Utilization. Academic Press, Waltham.

Lu, X., Zhou, W., and Gao, F. 2009. Cloning, characterization and localization of CHS gene from blood orange, *Citrus sinensis* (L.) Osbeck cv. Ruby. Mol. Biol. Rep., **36**:1983-1990. doi: 10.1007/s11033-008-9408-z.

Luria, N., Sela, N., Yaari, M., Feygenberg, O., Kobiler, I., Lers, A., and Prusky, D. 2014. *De-novo* assembly of mango fruit peel transcriptome reveals mechanisms of mango response to hot water treatment. BMC Genomics, **15**:1. <https://doi.org/10.1186/1471-2164-15-957>.

Mahato, A.K., Sharma, N., Singh, A., Srivastav, M., Singh, S.K., Singh, A.K., Sharma, T.R. and Singh, N.K. 2016. Leaf transcriptome sequencing for identifying genic-SSR markers and SNP heterozygosity in crossbred mango variety Amrapali (*Mangifera indica* L.). PloS One, **11**:e0164325. <https://doi.org/10.1371/journal.pone.0164325>.

- Martin, C., Prescott, A., Mackay, S., Bartlett, J., and Vrijlandt, E. 1991. Control of anthocyanin biosynthesis in flowers of *Antirrhinum majus*. *Plant J.*, **1**:37-49. doi: 10.1111/j.1365-313X.1991.00037.x.
- Mori, S., Kobayashi, H., Hoshi, Y., Kondo, M., and Nakano, M. 2004. Heterologous expression of the flavonoid 3',5'-hydroxylase gene of *Vinca major* alters flower color in transgenic *Petunia hybrida*. *Plant Cell Rep.*, **22**:415-421. doi 10.1007/s00299-003-0709-3.
- Moriguchi, T., Kita, M., Tomono, Y., Endo-Inagaki, T., and Omura, M. 1999. One type of chalcone synthase gene expressed during embryogenesis regulates the flavonoid accumulation in citrus cell cultures. *Plant Cell Physiol.*, **40**:651-655.
- Nambi, V. E., Thangavel, K., Shahir, S., and Chandrasekar, V. 2016. Color kinetics during ripening of Indian mangoes. *Int. J. Food Prop.*, **19(10)**: 2147-2155. doi: 10.1080/10942912.2015.1089281.
- Ngaki, M.N., Louie, G.V., Philippe, R.N., Manning, G., Pojer, F., Bowman, M.E., Li, L., Larsen, E., Wurtele, E.S., and Noel, J.P. 2012. Evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis. *Nature*, **485**:530-533. doi 10.1038/nature11009.
- Nordey, T., Joas, J., Davrieux, F., Genard, M., and Lechaudel, M. 2014. Non destructive prediction of color and pigment contents in mango peel. *Sci. Hort.*, **171**: 37-44. doi: 10.1016/j.scienta.2014.01.025.
- Palafox-Carlos, H., Contreras-Vergara, C.A., Muhlia-Almazan, A., Islas-Osuna, M.A. and Gonzalez-Aguilar, G.A. 2014. Expression and enzymatic activity of phenylalanine ammonia-lyase and p-coumarate 3-hydroxylase in mango (*Mangifera indica* 'Ataulfo') during ripening. *Genet. Mol. Res.*, **13**:3850-8. doi: 10.4238/2014.May.16.10.
- Pandit, S.S., Kulkarni, R.S., Giri, A.P., Kollner, T.G., Degenhardt, J.R., Gershenzon, J., and Gupta, V.S. 2010. Expression profiling of various genes during the fruit development and ripening of mango. *Plant Physiol. Biochem.*, **48**:426-433. doi:10.1016/j.plaphy.2010.02.012.

- Pelletier, M.K., Burbulis, I.E., and Winkel-Shirley, B. 1999. Disruption of specific flavonoid genes enhances the accumulation of flavonoid enzymes and end-products in *Arabidopsis* seedlings. *Plant Mol. Biol.*, **40**:45-54.
- Pelletier, M.K., Murrell, J.R., and Shirley, B.W. 1997. Characterization of flavonol synthase and leucoanthocyanidin dioxygenase genes in *Arabidopsis* (Further evidence for differential regulation of early and late genes). *Plant Physiol.*, **113**:1437-1445.
- Pervaiz, T., Songtao, J., Faghihi, F., Haider, M.S., and Fang, J. 2017. Naturally occurring anthocyanin structure, functions and biosynthetic pathway in fruit plants. *J. Plant Biochem. Physiol.*, **5(2)**: 187. doi:10.4172/2329-9029.1000187.
- Piero, A.R.L., Puglisi, I., Rapisarda, P., and Petrone, G. 2005. Anthocyanins accumulation and related gene expression in red orange fruit induced by low temperature storage. *J. Agr. Food Chem.*, **53**:9083-9088. doi: 10.1021/jf051609s.
- Ralston, L., Subramanian, S., Matsuno, M., and Yu, O. 2005. Partial reconstruction of flavonoid and isoflavonoid biosynthesis in yeast using soybean type I and type II chalcone isomerases. *Plant Physiol.*, **137**:1375-1388. doi: 10.1104/pp.104.054502.
- Ranganna, S. 1999. Hand Book of analysis and quality control for fruits and vegetables products, Tata-McGraw Hill Publications Limited, New Delhi p.112.
- Rao, A.V., and Rao, L.G. 2007. Carotenoids and human health. *Pharmacological research.* **55**:207-216. doi: 10.1016/j.phrs.2007.01.012.
- Ro, D., and Douglas, C. J. 2000. Reconstitution of general phenylpropanoid metabolism in yeast. In: 6th International Congress of Plant Molecular Biology (Quebec, Canada), p. S 06-39.
- Seitz, C., Ameres, S., and Forkmann, G. 2007. Identification of the molecular basis for the functional difference between flavonoid 3'-hydroxylase and flavonoid 3',5'-hydroxylase. *FEBS Lett.*, **581(18)**:3429-34. doi:10.1016/j.febslet.2007.06.045.

- Shimada, N., Aoki, T., Sato, S., Nakamura, Y., Tabata, S., and Ayabe, S.I. 2003. A cluster of genes encodes the two types of chalcone isomerase involved in the biosynthesis of general flavonoids and legume-specific 5-deoxy (iso) flavonoids in *Lotus japonicus*. *Plant Physiol.*, **131**:941-951.
- Singh, M., Arseneault, M., Sanderson, T., Murthy, V., and Ramassamy, C. 2008. Challenges for research on polyphenols from foods in Alzheimer disease: bioavailability, metabolism, and cellular and molecular mechanisms. *J. Agr. Food Chem.*, **56**: 4855-4873. doi: 10.1021/jf0735073.
- Sivankalyani, V., Sela, N., Feygenberg, O., Zemach, H., Maurer, D., and Alkan, N. 2016. Transcriptome dynamics in mango fruit peel reveals mechanisms of chilling stress. *Front. Plant Sci.*, **7**: 1579. <http://doi.org/10.3389/fpls.2016.01579>.
- Sivankalyani, V., Feygenberg, O., Diskin, S., Wright, B., and Alkan, N. 2016a. Increased anthocyanin and flavonoids in mango fruit peel are associated with cold and pathogen resistance. *Postharvest Biol. Tec.*, **111**(C):132-139. <https://doi.org/10.1016/j.postharvbio.2015.08.001>.
- Srivastava, S., Singh, R.K., Pathak, G., Goel, R., Asif, M.H., Sane, A.P., and Sane, V.A. 2016. Comparative transcriptome analysis of unripe and mid-ripe fruit of *Mangifera indica* (var. 'Dashehari') unravels ripening associated genes. *Sci. Rep.*, **6**: 32557. doi: 10.1016/j.plaphy.2010.02.012.
- Ubi, B.E., Honda, C., Bessho, H., Kondo, S., Wada, M., Kobayashi, S., and Moriguchi, T. 2006. Expression analysis of anthocyanin biosynthetic genes in apple skin: effect of UV-B and temperature. *Plant Sci.*, **170**:571-578. <https://doi.org/10.1016/j.plantsci.2005.10.009>.
- Wang, Y. S., Xu, Y.J., Gao, L.P., Yu, O., Wang, X.Z., He, X.J., Jiang, X.L., Liu, Y.J., and Xia, T. 2014. Functional analysis of Flavonoid 3',5'-hydroxylase from tea plant (*Camellia*

- sinensis*): critical role in the accumulation of catechins. BMC Plant Biol., **14**:1. doi: 10.1186/s12870-014-0347-7.
- Wu, H.X., Jia H.M., Ma X.W., Wang, S.B., Yao, Q.S., Xu, W.T., Zhou, Y.G., Gao, Z.S., and Zhan, R.L. 2014. Transcriptome and proteomic analysis of mango (*Mangifera indica* Linn) fruits. J. Proteomics, **105**: 19-30. <https://doi.org/10.1016/j.jprot.2014.03.030>.
- Yamazaki, M., Shibata M., Nishiyama Y., Springob K., Kitayama M., Shimada N., Aoki T., Ayabe, S.I., and Saito, K. 2008. Differential gene expression profiles of red and green forms of *Perilla frutescens* leading to comprehensive identification of anthocyanin biosynthetic genes. FEBS J., **275**:3494-3502. doi: 10.1111/j.1742-4658.2008.06496.x.
- Yang, H.Y. , Dong, T., Li, J.F., and Wang, M.Y. 2016. Molecular cloning, expression, and subcellular localization of a *PAL* gene from *Citrus reticulata* under iron deficiency. Biol. Plantarum, **60**(3): 482–488. doi: 10.1007/s10535-016-0625-3.
- Zhou, T.S., Zhou, R., Yu, Y.B., Xiao, Y., Li, D.H., Xiao, B., Yu, O., and Yang, Y.J. 2016. Cloning and characterization of a flavonoid 3'-hydroxylase gene from tea plant (*Camellia sinensis*). Int. J. Mol. Sci., **17**: 261. doi:10.3390/ijms17020261.

Table 1: Anthocyanin pathway genes identified by BLAST analysis of Amrapali leaf RNA-Seq data

Gene id	Length (amino acids)	Orthologous/Homologous genes (Top score hit match, Accession No)	Identity (%)
MiPAL	123	<i>Citrus clementina</i> PAL (ESR52831)	92
MiC4H	505	<i>Populus trichocarpa</i> x <i>Populus deltoides</i> C4H (AAG50231)	97
MiCHS1	390	<i>Citrus clementina</i> CHS (ESR58785)	79
MiCHS2	394	<i>Rhus chinensis</i> CHS (AGH13332)	94
MiCHS3	252	<i>Rhus chinensis</i> CHS (ESR39326)	74
MiCHI1	216	<i>Rhus chinensis</i> CHI3 (AGH13331)	93
MiCHI2	245	<i>Rhus chinensis</i> CHI (AGH13331)	90
MiCHI3	250	<i>Prunus persica</i> CHI (AEJ88218)	78
MiCHI4	244	<i>Citrus clementina</i> CHS3 (ESR49159)	91
MiCHI5	208	<i>Citrus clementina</i> CHI (ESR49159)	89
MiCHI6/FAP	167	<i>Citrus clementina</i> (ESR55559)	87
MiF3H1	520	<i>Citrus clementina</i> F3H (ESR39920)	91
MiF3H2	512	<i>Citrus clementina</i> F3H (ESR58067)	93
MiF3',5'H	507	<i>Cyclamen persicum</i> F3',5'H (ACX37698)	90
MiDFR	299	<i>Citrus clementina</i> DFR (ESR36537)	89

1 **Table 2: Bioinformatics analysis depicting the transcript variants, alternate splicing, predictions of exons and subcellular protein**
 2 **localization**
 3

S. No.	Gene	Protein	Transcript variants (CDS id, size)	Homologs/ Orthologs (Phytozome sequence id)	Alternate splice variants (Predicted exons with sizes)	Protein localization (Prediction significance)
1	<i>MiPAL</i>	Phenylalanine ammonia lyase	<i>MiPAL</i> (CDS6204, 369 bp)	<i>Citrus clementina PAL</i> (10011134m)	exon 1 (369 bp)	Endoplasmic reticulum (0.34)/ Secretory (0.93)
2	<i>MiC4H</i>	Cinnamate-4-hydroxylase	<i>MiC4H</i> (CDS4127, 1496 bp)	<i>Citrus clementina C4H</i> (10008125m)	unpredicted (25 bp), exon 1 (760 bp), exon 2 (709 bp)	Endoplasmic reticulum (0.61)/ Secretory (0.83)
3	<i>MiCHS</i>	Chalcone synthase	<i>MiCHS1</i> (CDS38624, 1119 bp)	<i>Citrus clementina CHS</i> (10028604m)	Unpredicted (184 bp), exon 1 (934bp)	Cytoplasm (0.94)
			<i>MiCHS2</i> (CDS149, 1148 bp)	<i>Citrus clementina CHS</i> (10028604m)	Unpredicted (40bp), exon 1 (137 bp), exon 2 (969bp)	Cytoplasm (0.92)
			<i>MiCHS3</i> (CDS41329, 686bp)	<i>Citrus clementina CHS</i> (10028604m)	Truncated exon 2 (686bp)	Cytoplasm (0.81)
4	<i>MiCHI</i>	Chalcone isomerase	<i>MiCHI1</i> (CDS10544, 666 bp)	<i>Citrus clementina CHI</i> (10032697m)	exon 1 (106 bp), exon 2 (158 bp), exon 3 (223bp), exon 4 (176bp)	Cytoplasm (0.66)
			<i>MiCHI2</i> (CDS2356, 669 bp)	<i>Citrus clementina CHI</i> (10032697m)	exon 1 (98 bp), exon 2 (158 bp), exon 3 (223bp), exon 4 (197bp)	Cytoplasm (0.64)
			<i>MiCHI3</i> (CDS12079, 1069 bp)	<i>Citrus clementina CHI</i> (10032697m)	unpredicted (648bp), exon 3 (223 bp), exon 4 (197bp)	Cytoplasm (0.86)
			<i>MiCHI4</i> (CDS1818, 732bp)	<i>Vitis vinifera CHI</i> (GSVIVG01032685001)	unpredicted (102 bp), exon 1 (90bp), exon 2 (155 bp), exon 3 (226 bp), exon 4 (155 bp)	Mitochondria (0.93)
			<i>MiCHI5</i> (CDS1784, 624 bp)	<i>Vitis vinifera CHI</i> (GSVIVG01032685001)	unpredicted (7bp), exon 1 (77 bp), exon 2 (155 bp), exon 3 (226 bp), exon 4 (155 bp)	Cytoplasm (0.85)

			<i>MiCHI6/FAP</i> (CDS17832, 497 bp)	<i>Arabidopsis thaliana</i> <i>FAP</i> (AT1G53520)	exon 1 (130 bp), exon 2 (214 bp), exon 3 (151 bp)	Cytoplasm (0.62)
5	<i>MiF3H</i>	Flavonoid hydroxylase	3 <i>MiF3H1</i> (CDS795, 1536 bp)	<i>Citrus clementina F3H</i> (10031266m)	signal peptide (53 bp), exon 1 (855bp), exon 2 (626 bp)	Mitochondria (0.52)
			<i>MiF3H2</i> (CDS4342, 1502 bp)	<i>Citrus clementina F3H</i> (10031266m)	unpredicted (919 bp includes signal peptide 16 bp), exon 2 (586 bp)	Chloroplast (0.7)
			<i>MiF3',5'H</i> (CDS9124, 1521 bp)	<i>Citrus clementina F3H</i> (10031266m)	unpredicted (890 bp includes signal peptide 96 bp), exon 2 (630 bp)	Mitochondria (0.38)
6	<i>MiDFR</i>	Dihydro-flavone reductase	<i>MiDFR</i> (CDS37350, 897 bp)	<i>Citrus clementina 2HFR</i> (10028526m)	exon 1 (183 bp), exon 2 (134 bp), exon 3 (197 bp), exon 4 (203 bp), exon 5 (136 bp)	Cytoplasm (0.76)

4

Draft

5 Figure captions

6 **Fig. 1: Schematic representation of the anthocyanin biosynthesis pathway.** *PAL*,
 7 phenylalanine ammonia lyase ; *C4H*, cinnamate-4-hydroxylase gene; *4CL*, 4-coumarateCoA
 8 ligase ; *CHS*, chalcone synthase ; *CHI*, chalconeisomerase-; *F3H*, flavanone 3-hydroxylase;
 9 *F3'H*, flavonoid 3'-hydroxylase; *F3',5'H*, flavonoid 3',5'-hydroxylase; *DFR*, dihydroflavonol
 10 4-reductase; *ANS*, anthocyanidin synthase; UFGT, UDP-glucose: flavonoid
 11 glucosyltransferase.

12 **Fig. 2: Variation in peel color and corresponding total anthocyanin contents in four**
 13 **mango varieties. A)** Fruit peel color variations in Amrapali, Tommy Atkins, Ambika and
 14 Arunika. **B)** Total anthocyanin content in fruit peel of green and red mango varieties. Data
 15 represents mean $\pm$ SD (n=3).

16 **Fig. 3: Phylogenetic analysis of genes involved in anthocyanin biosynthesis from**
 17 **transcript data and gene orthologs.** Neighbor-joining tree constructed from *PAL*, *C4H*,
 18 *CHSs*, *CHIs*, *F3Hs*, *F3',5'H*, *DFR*, and *ANS* amino acid sequences and their orthologs with
 19 1000 Bootstrap values.

20 **Fig. 4: Phylogenetic classification and comparative sequence variations of *CHI***
 21 **transcripts.** A) Phylogenetic analysis of *MiCHI* transcripts (amino acid sequences) and
 22 orthologous genes. B) Amino acid sequence variation between *MiCHI*1, 2 and 3. C) Amino
 23 acid sequence variation between *MiCHI*4 (mitochondrial) and 5 (cytoplasmic).

24 **Fig. 5: Expression profiling of anthocyanin biosynthetic genes in four mango varieties.**
 25 Transcript abundance pattern in Amrapali, Tommy Atkins, Ambika and Arunika for
 26 anthocyanin biosynthesis genes by q-RT-PCR after data normalization with actin. Data
 27 represents mean $\pm$ SD of three biological replicates.

Fig. 1: Schematic representation of the anthocyanin biosynthesis pathway. PAL, phenylalanine ammonia lyase ; C4H, cinnamate-4-hydroxylase gene; 4CL, 4-coumarateCoA ligase ; CHS, chalcone synthase ; CHI, chalcone isomerase-; F3H, flavanone 3-hydroxylase; F3'H, flavonoid 3'-hydroxylase; F3',5'H, flavonoid 3',5'-hydroxylase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase; UFGT, UDP-glucose: flavonoid glucosyltransferase.

341x337mm (96 x 96 DPI)

Fig. 2: Variation in peel color and corresponding total anthocyanin contents in four mango varieties. A) Fruit peel color variations in Amrapali, Tommy Atkins, Ambika and Arunika. B) Total anthocyanin content in fruit peel of green and red mango varieties. Data represents mean $\pm$ SD (n=3).

229x210mm (96 x 96 DPI)

Fig. 3: Phylogenetic analysis of genes involved in anthocyanin biosynthesis from transcript data and gene orthologs. Neighbor-joining tree constructed from PAL, C4H, CHSs, CHIs, F3Hs, F3',5'H, DFR, and ANS amino acid sequences and their orthologs with 1000 Bootstrap values.

298x225mm (96 x 96 DPI)

Fig. 4: Phylogenetic classification and comparative sequence variations of CHI transcripts. A) Phylogenetic analysis of MiCHI transcripts (amino acid sequences) and orthologous genes. B) Amino acid sequence variation between MiCHI1, 2 and 3. C) Amino acid sequence variation between MiCHI4 (mitochondrial) and 5 (cytoplasmic).

384x486mm (96 x 96 DPI)

Fig. 5: Expression profiling of anthocyanin biosynthetic genes in four mango varieties. Transcript abundance pattern inAmrapali, Tommy Atkins, Ambika and Arunika for anthocyanin biosynthesis genes by q-RT-PCR after data normalization with actin. Data represents mean $\pm$ SD of three biological replicates.

300x155mm (150 x 150 DPI)