Comparative Transcriptomics of Mango (*Mangifera indica* L.) Cultivars Provide Insights of Biochemical Pathways Involved in Flavor and Color

Waqasuddin Khan[1], Safina Abdul Razzak[1], M. Kamran Azim[1,2*]

[1]Jamil-ur-Rahman Center for Genome Research, Dr. Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi, Karachi-75270, Pakistan.

[2]Mohammad Ali Jinnah University, Karachi 75400, Pakistan.

Short running title: Transcriptome analysis of mango cultivars.

Keywords: RNA-seq, fruit color, fruit flavor, fruit terpenoids, fruit volatiles.

***Corresponding author:**

Prof. Dr. M. Kamran Azim,

Mohammad Ali Jinnah University,

Block-6 P.E.C.H.S.,

22-E, Shahrah-e-Faisal Service Road South,

Karachi-75400, Pakistan.

Email: kamran@jinnah.edu.pk

Phone: +92-336-2154268 **Fax:** +92-213-4819018

Abstract

Mango is an economically important fruit crop of many tropical and subtropical countries. Recently, leaf and fruit transcriptomes of mango cultivars grown in different geographical regions have characterized. Here, we presented comparative transcriptome analysis of four mango cultivars i.e. cv. *Langra*, cv. *Zill,* cv. *Shelly* and cv. *Kent* from Pakistan, China, Israel and Mexico respectively. De-novo sequence assembly generated 30,953-85,036 unigenes from RNA-Seq datasets of mango cultivars. KEGG pathway mapping of mango unigenes identified terpenoids, flavonoids and carotenoids biosynthetic pathways involved in flavor and color. The analysis revealed linalool as major monoterpenoid found in all cultivars studied whereas, monoterpene α-terpineol was specifically found in cv. *Shelly*. Ditepene gibberellin biosynthesis pathway was found in all cultivars whereas, homoterpene synthase involved in biosynthesis of 4,8,12-trimethyltrideca-1,3,7,11-tetraene (TMTT; an insect induced diterpene) was found in cv. *Kent*. Among sesquiterpenes and triterpenes, biosynthetic pathway of Germacrene-D, an anti-bacterial and anti-insecticidal metabolite was found in cv. *Zill* and cv. *Shelly*. Two bioactive triterpenes, lupeol and β-amyrin were found in cv. *Langra* and cv. *Zill*. Unigenes involved in biosynthesis of carotenoids, β-carotene and lycopene, were found in cultivars studied. Many unigenes involved in flavonoid biosynthesis were also found. Comparative transcriptomics revealed naringenin (an anti-inflammatory and antioxidant metabolite) as ‘central’ flavanone responsible for biosynthesis of an array of flavonoids. The present study provided insights on genetic resources responsible for flavor and color of mango fruit.

Introduction

As a member of the family Anacardiaceae, mango (*Mangifera indica* Linn.) ranks second among tropical fruit crops after banana due to its rich sensational taste, color, aroma and huge economics significance (Srivastava et al. 2016; Litz 2009). Many mango varieties (i.e. cultivar abbreviated as cv.) are commercially grown in tropical and subtropical countries worldwide (Mukherjee and Litz 2009). According to Food and Agriculture organization of the United Nations (FAO), India holds the 1st position in mango production followed by China, whereas Pakistan and Mexico rank 5th and 6th position respectively (FAOSTAT-2014; www.faostat.fao.org). Mango fruit is a rich source of bioactive phytochemicals including antioxidants and other health–promoting compounds (Lauricella et al. 2017; Fessard et al. 2017; Shah et al. 2010; Masibo and He 2009). This fruit is known for attractive colors, cherishing aroma, delightful taste and high nutritional value, due to its high content of vitamin C, carotenoids, flavones, terpenoids and minerals **(**Lauricella et al. 2017**)**. The biochemical composition of mango pulp obtained from different mango cultivars varies with location of cultivation, variety, and stage of maturity (Dautt-Castro et al. 2015). Previous studies on mango were focused on the ripening process, volatile composition, antioxidant capacity, postharvest treatment and fruit quality (Srivastava et al. 2016; White et al. 2016; El-Hadi et al. 2013; Litz et al. 2009; Pino and Mesa 2006; Pino et al. 2005). Recently, Kuhn et al (2017) reported a consensus genetic map of mango using seven mapping population.

Recent transcriptome analysis of leaves and fruits of several mango cultivars by RNA–Seq provided insights of fundamental molecular biology of this plant. We first of all reported the mango leaf transcriptome of cv. *Langra* variety in 2014 (Azim et al. 2014). Mango fruit

transcriptomes of cv. *Zill* (Wu et al. 2014), cv. *Shelly* (Luria et al. 2014), cv. *Kent* (Dautt-Castro et al. 2015), cv. *Dashehari* (Srivastava et al. 2016) and more recently cv. *Keitt* (Tafolla-Arellano et al. 2017) have been reported from China, Israel, Mexico, India and USA respectively. In another study, a leaf transcriptome identified genic-SSR markers and SNP heterozygosity in cv. *Armpali* (Mahoto et al. 2016). Simple sequence repeats (SSR) and SNP identification have proved to be informative DNA-based markers in plant molecular genetics (Mahoto et al. 2016; Khan and Azim 2011). Identification of SSR and SNP markers by transcriptome sequence datasets has potential to be utilized in mango breeding programs.

Genome-wide association, genetic mapping and identification of trait specific markers help to deploy important genes involved in flavor, aroma and pulp consistency for which mango is popularly consumed (Srivastava et al. 2016). Transcriptomic sequence datasets obtained from RNA-Seq of different mango cultivars resulted in 30,000–85,000 unigenes (Azim et al. 2014; Wu et al. 2014; Luria et al. 2014; Dautt-Castro et al. 2015; Srivastava et al. 2016). The transcriptome sequencing described in these reports, has been carried out using mango cultivars grown in different geographical regions. Hence, a comparative transcriptomic analysis was needed, in order to characterize common as well as cultivar and/or tissue-specific transcripts. Here, we present comparative analysis of transcriptomic datasets of available mango cultivars. This bioinformatics study identified genetic characteristics of mango responsible for its color, aroma, flavor and other agronomic traits at the systemic level.

Materials and Methods

Retrieval of Mango RNA-seq Data

For *de novo* assembly of RNA-Seq reads, the NGS reads of cv. *Langra* (SRA ID: SRR947746) (Azim et al. 2014), cv. *Zill* (SRA ID: SRP035450) (Wu et al. 2014), cv. *Shelly* (SRA ID: SRX375390) (Luria et al. 2014) and cv. *Kent* (SRA ID: SRP045880) (Dautt-Castro et al. 2015) were retrieved from NCBI Sequence Read Archive (SRA) (Leinonen et al. 2010). The cv. *Langra* sequence reads were from mango leaves where as other three were from mango fruits. In case of cv. *Shelly*, sequence reads of samples at different time intervals were pooled and used for subsequent analysis. For cv. *Kent*, sequence reads of mature mango RNA–Seq data were processed.

Preprocessing of Mango RNA-seq Reads

Reads were converted from SRA format to fastq format using SRA Toolkit (https://www.ncbi.nlm.nih.gov/sra/docs/toolkitsoft/). Each SRA file provided two paired–end fastq files. The sequence reads were examined for quality by FASTQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and were filtered by FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) to obtain high–quality reads (reads with Q score ≥27). The processed forward (F1) and reverse (R2) read files were then paired using Pairfq script (https://github.com/sestaton/Pairfq). The headers of F1 and R2 files were configured according to CASAVA 1.8 format for Trinity (Grabherr et al. 2011) software using Fastool, a Trinity plugin.

De novo Transcriptomic Assembly of Processed Reads

Transcriptome *de novo* assembly of clean reads was performed using Trinity (Grabherr et al. 2011) which uses three independent software modules – Inchworm, Chrysalis, and Butterfly – applied sequentially to process the sequencing data of RNA–seq reads. Bowtie aligner with some Perl scripts is part of Trinity pipeline. In brief, Trinity assembles the reads into unique sequences of transcripts, known as contigs. These contigs were clustered by constructing the complete *de Bruijn*graph for each cluster, and then partitioned the full read set among these disjoint graphs. Finally, Trinity processed the individual graphs in parallel, tracing the path that reads and pairs of reads take within the graph, ultimately reporting full–length transcripts for alternatively spliced isoforms, and teasing apart transcripts that corresponds to paralogous genes (Grabherret al. 2011). Consequently, we obtained four unigenes datasets corresponding to four mango cultivars.

BLAST Analysis of Unigenes

Multiple BLAST strategies were used for sequence comparisons of unigenes against different sequences databases. (a) All unigenes from four assemblies were aligned against each other by BLASTN (E value cutoff $\leq 1e^{-5}$) to find the common and unique transcripts. The stringent alignment was defined by an E-value threshold of $<1e^{-10}$, and a percent alignment and percent coverage length of ≥75%. (b) Unigenes in four datasets were aligned using BLASTN against the coding sequences (CDS) of *Citrus sinensis* (https://www.citrusgenomedb.org/), *Citrus Clementina* (https://www.citrusgenomedb.org/), *Vitis Vinifera* (http://www.plantgdb.org/VvGDB/), *Ricinuscommunis* (http://www.plantgdb.org/RcGDB/) *and Populous tricocarpa*

(http://www.plantgdb.org/PtGDB/) as well as against the plant–specific sequence (nucleotide and protein) databases of NCBI, UniProtKB, and Swiss–Prot with an E-value cut–off $\leq 1e^{-5}$. (c) The assembled datasets were also BLASTed against NCBI non–redundant nucleotide (NT) and non–redundant (NR) protein sequence databases with E-value cut–off $\leq 1e^{-5}$.

Functional Annotations of Unigenes

The assembled unigenes datasets were also filtered for redundant sequences using CD–HIT (Cluster Database at High Identity with Tolerance) (Fu et al. 2012) at 90% identity threshold. The non-redundant unigenes were further analyzed for coding regions using TransDecoder (http://transdecoder.github.io/). Obtained coding sequences (CDS) were then subjected to InterProScan v.5.15.54.0 (Jones et al. 2014) using IPRLOOKUP service for functional annotation and Gene Ontology (GO) assignments. Translated protein sequences were also scanned against the following InterPro signature databases: Hamap (201502.04), ProDom (2006.1), PRISF (3.01), SMART (6.2), TIGRFAM (15.0), PRINTS (42.0) and SUPERFAMILY (1.75). CateGOrizer (http://www.animalgenome.org/tools/catego/) was used to analyze GO term datasets into three GO classes' i.e. biological process, cellular component and molecular function. InterProScan's XML output along with their corresponding BLASTX result was loaded into Blast2GO java application (https://www.blast2go.com/start–blast2go–2–8) for gene annotations. The active biochemical pathway analysis was done by KAAS (Moriya et al. 2007) using KEGG database (Kanehisa 2002).

Results

The transcriptomic *de novo* assembly resulted from Trinity (Grabherr et al. 2011) generated 30,953, 57,544, 58,797 and 85,036 of unigenes from cv. *Langra*, cv. *Zill*, cv. *Shelly* and cv. *Kent* mango cultivars. These unigenes were further characterized for functional annotations using BLAST and InterProScan (Jones et al. 2014). BLAST homology search showed 83-96% similarity of unigenes with sequences in Nt and Nr databases. The cultivar-specific consensus search among the four datasets of mango unigenes showed that 80–98% unigenes sequences were matched with each other with a similarity index of ≥75%. However, InterProScan identified 12,388, 29,303, 25,878, and 18,793 protein coding sequences in cv. *Langra*, cv. *Zill*, cv. *Shelly* and cv. *Kent* unigenes datasets respectively. Biochemical pathway analysis using KEGG-KASS (Kanehisa 2002; Moriya et al. 2007) identified numerous unigenes involved in the formation of genes that are involved for the production of biomolecules responsible for color and flavor of mango fruit.

Discussion

Mango is an important fruit crop of many countries located in tropical and subtropical regions. In the absence of genome sequence information, transcriptomic sequences of different mango cultivars provided a wealth of data related to protein coding sequences. This study resulted in a 'consensus transcriptome sequence reference' obtained from four mango cultivars grown in Pakistan, China, Israel and Mexico. Initially, we retrieved 12.1, 68.4, 83.2, and 22.0 million paired-end RNA-Seq reads of cv. *Langra* (Pakistan), cv. *Zill* (China), cv. *Shelly* (Israel) and cv. *Kent* (Mexico) respectively from Sequence Read Archive (SRA). These transcriptome sequences were obtained from RNA-Seq experiments using Illumina NGS technology. After filtering by

FASTX toolkit, the high quality clean reads were used for *de novo* assembly by Trinity using uniform parameters. Collectively, all cleaned high quality sequence read datasets contained 102.9 million reads, with more than 6.5 billion nucleotides (Table 1).

The four RNA-Seq datasets were assembled individually resulting in four datasets of unigenes (Table 2). The N50 of the assembled transcripts datasets were in the range of 525 – 1598 nucleotides (Table 2). The number of unigenes were as follows; cv. *Langra* = 30,953; cv. *Zill* = 58,797; cv. *Shelly* = 57,544; and cv. *Kent* = 85,036. The number of unigenes obtained as Trinity outputs were comparable as previously reported in respective publications (Azim et al. 2014; Wu et al. 2014; Luria et al. 2014; Dautt-Castro et al. 2014). This observation provided confidence for comparative transcriptome analysis.

Functional Annotation of Assembled Unigenes

To characterize the putative functions of mango unigenes, three different BLAST sequence similarity search strategies were adopted.

(1) All unigenes datasets were aligned against NT (non-redundant nucleotide sequence database), NR (non-redundant translated sequence database), Plant NT/NR, SwissProt and UniProt databases using BLAST. Sequence similarity searching showed homology of 83-96% unigenes with sequences in Nt and Nr databases.

(2) The unigene sequences were also compared with genomic sequence datasets of different Viridiplantae which revealed considerable sequence similarity with *Cirus sinensus*, *Citrus clementina*, *Populus tricarpa*, *Vitis vinifera* and *Riccinus communis* (Figure 1).

(3) To find out cultivar-specific and 'consensus' sequences (i.e. unigenes common in four mango datasets studied); unigenes of four cultivars were compared to each other using BLAST. At every instance, one unigenes dataset was selected as query while other three datasets were considered as database. BLAST searches showed that 80–98% unigenes sequences of four mango cultivars have ≥75% identity with each other.

InterProScan provides a systematic language to describe the attributes of genes and gene products, which includes the functional characterization and annotation in combination with different protein signature recognition methods into one resource (Jones et al. 2014). InterProScan identified 12,388, 29,303, 25,878, and 18,793 protein coding sequences in cv. *Langra*, cv. *Zill*, cv. *Shelly* and cv. *Kent* unigenes datasets respectively. Functional characterization by Gene Ontology (GO) annotated an array of expressed genes in these mango varieties. GO analysis identified 17,704 (cv. *Langra)*, 18,846 (cv. *Zill*), 18,325 (cv. *Shelly*) and 12,119 (cv. *Kent*) GO terms, assigned to one of the three biological domains (i.e. Biological, Cellular and Molecular functions).

Genes Related to Mango Flavor

Flavor of mango i.e. taste and aroma is constituted by a complex mixture of natural products. More than 500 volatile compounds have been reported to contribute in mango aroma and taste (Singh et al. 2013). The mango cultivars compared during present study have characteristic flavor (taste and aroma), color and consistency of pulp which is supposed to be due to different terpenoids, flavonoids and carotenoids biosynthetic pathways. Analysis of RNA-seq datasets of mango cultivars revealed unigenes involved in biosynthesis of oxygenated volatile compounds including esters, furanones and lactones. These secondary metabolites contribute as determinants

of the characteristic aroma (Quijano et al. 2007). Amount and type of volatile compounds in mango often depend on area of production. Asian mangoes have more oxygenated volatile compounds such as esters, furanones, and lactones which give pineapple- or peach-like aromas to some varieties (Moshonas and Shaw 1994), while western mangoes that are hybrids of Asian stock have higher levels of certain hydrocarbons (Moshonas and Shaw 1994; MacLeod and de Troconis 1982). KEGG pathway analysis of four mango unigenes datasets identified active biochemical pathways involved in mango flavor, color and antioxidant activity.

Terpenoids: Terpene hydrocarbons are considered to be important contributors to flavor in most of the mango varieties (Quijano et al. 2007; El-Hadi et al. 2013)**.** Many monoterpenes (C10) and sesquiterpenes (C20) comprise the most abundant group of compounds present in the aroma profile (Lichtenthaler 1999). The terpenoids are synthesized from two universal precursors isopentenyldiphosphate (IPP) and dimethylallyldiphosphate (DMAPP) which are products of two independent pathways: the cytosolic mevalonate (MVA) pathway; and plastidic1-deoxy-d-xylulose-5-phosphate (DOXP) pathway (Nagegowda 2010). IPP and DMAPP are metabolized by a series of synthases (FDPS; farnesyldiphosphate synthase [EC 2.5.1.1], FPPS; farnesyldiphosphate synthase [EC 2.5.1.10] and GGPS1; geranylgeranyldiphosphate synthase, type III [EC 2.5.1.29]) into Geranyldiphosphate (GPP), Geranylgeranyldiphosphate (GGPP), and Farnesyldiphosphate (FPP). GPP processed to form monoterpenoids, whereas FPP enters in sesquiterpenoid/triterpenoid, carotenoid and N-glycan biosynthesis pathways (Figure 2). The analysis of unigenes of mango cultivars showed that GGPP acts as precursor of phytyldiphospahte (PyPP) and nona-prenyldiphosphate (NoPP) by all-trans-nonaprenyl-diphosphate synthases [EC 2.5.1.85; EC 2.5.1.82]. PyPP and NoPP enter in Ubiquinone and other terpenoid-quinone biosynthesis pathways. Interestingly, in cv. *Shelly* dataset, two

additional precursors i.e. hexaprenyldiphosphate and decaprenyldiphosphate are formed as indicated by unigenes encoding hexaprenyldiphosphate synthase [EC 2.5.1.83] and decaprenyldiphosphate synthase [EC 2.5.1.91]. These cv. *Shelly* specific precursors enter in Ubiquinone and other terpenoid-quinone biosynthesis pathway (Figure 2).

Monoterpenoids: Linalool and α-Terpineol are monoterpene alcohols, mainly found in flowers and spice plants having important commercial applications. These alcoholic compounds are responsible for pleasant aroma in fruits (Pino and Mesa 2006). The unigenes encoding S-Linalool synthase [EC 4.2.3.25] responsible for synthesis of Linalool was found in all mango cultivars studied (Figure 2). This observation indicated that Linalool is the main monoterpene found in mango; while transcript encoding α-Terpineol synthase [EC 4.2.3.111] for α-Terpineol biosynthesis from Geranyl PP was found specifically in cv. *Shelly*. Among other monoterpenes, myrcene is found to be the major compound in most New World mango cultivars, along with sesquiterpene hydrocarbons which present in amounts as high as 10% (Lewinsohn et al. 2001).

Diterpenoids: Two diterpene biosynthetic pathways were found active in mango transcriptome datasets. (i) We found unigenes encoding enzymes involved in Gibberellins hormone biosynthesis produced from geranylgeranyldiphosphate (GGPP) via ent-copalyldiphosphate by the bifunctionalent-copalyldiphosphate/ent-kaurene synthase (CPS/KS). Gibberellins are tetracyclic diterpenes (C20) which stimulate wide variety of responses during plant growth (Phinney and Spray 1987). (ii) A unigene encoding homoterpene synthase [EC 1.14.13.B14] (a cytochrome P450 enzyme) was also found except in cv. *Kent* dataset. This enzyme catalyses the conversion of secondary metabolite geranyl linalool to the homoterpene 4,8,12-trimethyltrideca-1,3,7,11-tetraene (TMTT) (C16). TMTT is an insect-induced volatile compound involved in

plant defense response. Terpene volatiles play a vital role in plant-organism interactions as attractants of pollinators or as defense compounds against herbivores (Lee et al. 2010).

Sesquiterpene and Triterpenoids: A number of unigenes encoding enzymes responsible for biosynthesis of sesquiterpenes and triterpenes were found in mango transcriptome datasets (Figure 2). (i) The unigenes encoding germacradienol synthase [EC 4.2.3.22] which catalyzes the sesquiterpenoid Germacrene D biosynthesis, was found in datasets of cv. *Zill* and cv. *Shelly*. This sesquiterpene is reported to have antimicrobial and anti-insecticidal activities (Noge and Becerra 2009)**.** (ii) The unigenes encoding lupeol synthases [EC 5.4.99.-] and β-amyrin synthase [EC 5.4.99.39] for biosynthesis of pharmacologically active triterpenoids, i.e. Lupeol and β-amyrin were found in cv. *Langra*, cv. *Zill* and cv. *Shelly* datasets (Saleem 2009; Siddique and Saleem 2011; Santos et al. 2012).

Besides terpenoids, other compounds also contribute in mango flavor and fragrance. Mango possesses a very attractive sweet essence characteristic due to the presence of different sugars. The major sugars found in mango are glucose, fructose, and sucrose (MacLeod and de Troconis 1982). In accordance with the literature, unigenes encoding enzymes responsible for the synthesis of sucrose, fructose and glucose were found in all four cultivars. In addition, the cv. *Langra* dataset have transcripts encoding enzymes of mannose and galactose pathways; whereas 17 transcripts encoding enzymes of pentose sugar pathway were detected in cv. *Zill* and cv. *Shelly* datasets.

Among the carbonyls, 14 unigenes encoding series of oxidoreductases for biosynthesis of (E)-2-hexenal and hexanal were found in fruit transcriptome datasets (i.e. cv. *Kent*, cv. *Shelly* and cv. *Zill*) while absent in leaf dataset (cv. *Langra*). These compounds have fatty–grassy and

green–fruity notes that could make a minor contribution to mango aroma and reported to be bactericidal (Pino and Mesa 2006; Lanciotti et al. 2003). Several unigenes encoding esterases which hydrolyse fruit lactone as intramolecular esters of 4- and 5-hydroxy acids were also found. These were previously characterized in mangos (Moshonas and Shaw 1994; Macleod and de Troconis 1982), and are considered to be important contributors to the flavor and aroma of this fruit (Pino et al. 2005; Fahlbusch et al. 2007).

Genes Involved in Carotenoids and Flavonoids Biosynthesis

Plant carotenoids are tetraterpenes and the most vital colored phytochemicals which occurs as all-trans and cis-isomers (Khoo et al. 2011). This group of natural products referred as pigment and nutraceuticals (Botella-Pavía et al. 2004) accounting for the brilliant colors in fruits and vegetables (Khoo et al. 2011). Carotenoids also act as a precursor for the production of apocarotenoid hormones such as abscisic acid which regulate development of plant and its interaction with their environment (Nambara and Marion-Poll 2005)**.** Carotenoids are derived from the 40-carbon isoprenoid phytoene that participate in light harvesting and essential for photoprotection against excess light (Ruiz-Sola and Rodríguez-Concepción 2012; Moran and Jarvik 2010). Plant carotenoids are red, orange, and yellow lipid-soluble color pigments embedded in the membranes of chloroplasts and chromoplasts (Bartley and Scolnik 1995). The most studied carotenoids include β-carotene, lycopene, lutein and zeaxanthin. However, the intensity of color in fruits and vegetables depends on the concentration of carotenoids and their growth maturity (Khoo et al. 2011).

Lycopene and β-carotene are important carotenoids of mango (Khoo et al. 2011)**.** Unigenes encoding enzymes for lycopene and β-carotene were found in cv. *Zill* and cv. *Shelly*

fruits datasets giving a reddish-orange color to the fruit (Figure 3) (Wu et al. 2014; Luria et al. 2014). On the other hand, absence of β-carotene and presence of lycopene in cv. *Kent* is predicted to be responsible for orange-yellow color of the fruit (Dautt-Castro et al. 2015).

Many unigenes encoding enzymes and related proteins involved in biosynthesis of flavonoids, the group of pigments that color most flowers, fruits, and seeds were present in four mango cultivars datasets (Figure 4). These flavonoids included naringenin, pinobanksin, afzelechin, apiferol, eriodictoyl, luteolin, catechins (epicatechin, gallocatechin, epigallocatechin) myricetin, and quercetin. Flavonoids are phenylpropanoid-derived plant metabolites and ubiquitous in nature (Hoang et al. 2015). According to chemical structure, these secondary metabolites are classified into flavonols, flavones, flavanones, isoflavones, catechins, anthocyanidins and chalcones. Flavonoids are known to perform diverse functions including color-based attractants to pollinators and symbionts (Dixon and Pasinetti 2012). In higher order plants, flavonoids are also required for UV filtration, nitrogen fixation, cell cycle inhibition, and as chemical messengers. These compounds also act as allelochemicals, antimicrobial, antiherbivore, antiallergic, antiplatelet, anti-inflammatory, anti-tumor and antioxidant agents (Falcone Ferreyra et al. 2012).

Genes involved in biosynthesis of antioxidants

Tocopherols and tocotrienols (vitamin E), ascorbic acid (vitamin C) and carotenoids react with free radicals and reactive oxygen species, which is the basis for their function as antioxidants (Sies and Stahl 1995). The presence of active pathways for biosynthesis of Vitamin E and C and carotenoids in all four cultivars of mango signifies presence of the antioxidant activity.

The present comparative transcriptomic analysis of four mango cultivars from different countries provided insight of genes encoding for enzymes for biosynthesis of terpenoids, carotenoids, flavonoids and other natural products. These volatile and nonvolatile metabolites are determinants of flavor (aroma and taste) and color.

References

1. Azim MK, Khan IA, Zhang Y (2014) Characterization of mango (Mangifera indica L.) transcriptome and chloroplast genome. Plant Mol. Biol. 85(1-2):193-208. doi:10.1007/s11103-014-0179-8.

2. Bartley GE, Scolnik PA (1995) Plant carotenoids: pigments for photoprotection, visual attraction, and human health. The Plant Cell 7(7):1027-1038. doi: 10.1105/tpc.7.7.1027.

3. Botella-Pavía P, Besumbes O, Phillips MA, Carretero-Paulet L, Boronat A, Rodríguez-Concepción M (2004) Regulation of carotenoid biosynthesis in plants: evidence for a key role of hydroxymethylbutenyl diphosphate reductase in controlling the supply of plastidialisoprenoid precursors. Plant J. 40(2):188-199. doi:10.1111/j.1365-313X.2004.02198.

4. Dautt-Castro M, Leyva OA, Vergara CA, Sanchez MA, Flores CS, Flores SA, Kuhn DN, Osuna MA (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.

5. Dixon RA, Pasinetti GM (2012) Flavonoids and isoflavonoids: from plant biology to agriculture and neuroscience. Plant Physiol. 154(2):453-457. doi: doi.org/10.1104/pp.110.161430.

6. El-Hadi MA, Zhang FJ, Wu FF, Zhou CH, Tao J (2013) Advances in fruit aroma volatile research. Molecules 18(7):8200-8229. doi:10.3390/molecules18078200.

7. Fahlbusch KG, Hammerschmidt FJ, Panten J, Pickenhagen W, Schatkowski D, Bauer K, Garbe D, Surburg H (2007) Flavors and fragrances Wiley, pp. 74□78. doi:10.1002/14356007.a11_141.

8. Falcone Ferreyra ML, Rius S, Casati P (2012) Flavonoids: biosynthesis, biological functions, and biotechnological applications. Front. Plant Sci. 3:222. doi:10.3389/fpls.2012.00222.

9. Fessard A, Kapoor A, Patche J, Assemat S, Hoarau M, Bourdon E, Bahorun T, Remize F (2017) Lactic Fermentation as an Efficient Tool to Enhance the Antioxidant Activity of Tropical Fruit Juices and Teas. Microorganisms 5(2):23. doi:10.3390/microorganisms5020023.

10. Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28(23):3150-3152. doi:10.1093/bioinformatics/bts565.

11. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat. Biotechnology 29(7):644. doi:10.1038/nbt.1883.

12. Hoang VL, Innes DJ, Shaw PN, Monteith GR, Gidley MJ, Dietzgen RG (2015) Sequence diversity and differential expression of major phenylpropanoid-flavonoid biosynthetic genes among three mango varieties. BMC Genomics 16(1):561. doi:10.1186/s12864-015-1784-x.

13. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9):1236-1240. doi:0.1093/bioinformatics/btu031.

14. Kanehisa M (2002) The KEGG database. Novartis Found Symp. 247:91-101.

15. Khan IA, Azim MK (2011) Variations in intergenic spacer rpl20-rps12 of Mango (*Mangifera indica)* chloroplast DNA: implications in cultivar identification. Plant Syst. Evol. 292(3-4):249-255.doi:10.1007/s00606-011-0424-4.

16. Khoo HE, Prasad KN, Kong KW, Jiang Y, Ismail A (2011) Carotenoids and their isomers: color pigments in fruits and vegetables. Molecules 16(2):1710-1738.

17. Kuhn DN, Bally IS, Dillon NL, Innes D, Groh AM, Rahaman J, Ophir R, Cohen Y, Sherman A(2017) Genetic map of mango: A tool for mango breeding. Front. Plant Sci. 20(8):577. doi:10.3389/fpls.2017.00577.

18. Lanciotti R, Belletti N, Patrignani F, Gianotti A, Gardini F, Guerzoni ME (2003) Application of hexanal, (E)-2-hexenal, and hexyl acetate to improve the safety of fresh-sliced apples. J. Agric. Food Chem. 51(10):2958-2963.

19. Lauricella M, Emanuele S, Calvaruso G, Giuliano M, D'Anneo A (2017) Multifaceted Health Benefits of Mangifera indica L.(Mango): The Inestimable Value of Orchards Recently Planted in Sicilian Rural Areas. Nutrients 9(5):525. doi:10.3390/nu9050525.

20. Lee S, Badieyan S, Bevan DR, Herde M, Gatz C, Tholl D (2010) Herbivore-induced and floral homoterpene volatiles are biosynthesized by a single P450 enzyme (CYP82G1) in Arabidopsis. Proc. Natl. Acad. Sci. U.S.A. 107(49):21205-21210. doi:10.1073/pnas.1009975107.

21. Leinonen R, Sugawara H, Shumway M (2010) The sequence read archive (SRA). Nucleic Acids Res. 39(Database issue):D19-D21.

22. Lewinsohn E, Schalechet F, Wilkinson J, Matsui K, Tadmor Y, Nam KH, Amar O, Lastochkin E, Larkov O, Ravid U, Hiatt W (2001) Enhanced levels of the aroma and flavor compound S-linalool by metabolic engineering of the terpenoid pathway in tomato fruits. Plant Physiol. 127(3):1256-1265. doi.org/10.1104/pp.010293.

23. Lichtenthaler HK (1999) The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. Annu. Rev. Plant Biol. 50 (1):47-65. doi: 10.1146/annurev.arplant.50.1.47.

24. Litz RE, editor (2009) The mango: botany, production and uses. CABI pg14. doi: 10.1079/9781845934897.0000.

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

26. MacLeod AJ, de Troconis NG (1982) Volatile flavour components of mango fruit. Phytochemistry 21(10):2523-2526. doi:10.1016/0031-9422(82)85249-7.

27. Mahato AK, Sharma N, Singh A, Srivastav M, Jaiprakash, Singh SK, Singh AK, Sharma TR, Singh NK (2016) Leaf Transcriptome Sequencing for Identifying Genic-SSR Markers and SNP Heterozygosity in Crossbred Mango Variety 'Amrapali' (Mangifera indica L.). PLoS One 11(10):e0164325. doi:10.1371/journal.pone.0164325.

28. Masibo M, He Q (2009) Mango bioactive compounds and related nutraceutical properties: A review. Food Rev. Int. 25(4):346-370. doi:10.1080/87559120903153524.

29. Moran NA, Jarvik T (2010) Lateral transfer of genes from fungi underlies carotenoid production in aphids. Science 328(5978):624-627. doi:10.1126/science.1187113.

30. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35(Web Server issue):W182-W185.

31. Moshonas MG, Shaw PE (1994) Quantitative determination of 46 volatile constituents in fresh, unpasteurized orange juices using dynamic headspace gas chromatography. J. Agric. Food Chem. 42(7):1525-1528.

32. Mukherjee S, Litz RE (2009) Introduction: botany and importance in The mango: botany, production and uses. CABI pg 1-18. doi:10.1079/9781845934897.0000.

33. Nagegowda DA (2010) Plant volatile terpenoid metabolism: biosynthetic genes, transcriptional regulation and subcellular compartmentation. FEBS lett. 584(14):2965-2973.doi:10.1016/j.febslet.2010.05.045.

34. Nambara E, Marion-Poll A (2005) Abscisic acid biosynthesis and catabolism. Annu. Rev. Plant Biol. 56:165-185. doi:10.1146/annurev.arplant.56.032604.144046.

35. Noge K, Becerra JX (2009) Germacrene D, A Common Sesquiterpene in the Genus Bursera (Burseraceae). Molecules 14(12): 5289-5297. doi:10.3390/molecules14125289.

36. Phinney BO, Spray CR (1987) Diterpenes-the gibberellin biosynthetic pathway in Zea mays. The metabolism, structure, and function of plant lipids. Springer New York, pp. 19-27. doi: 10.1007/978-1-4684-5263-1_3.

37. Pino JA, Mesa J (2006) Contribution of volatile compounds to mango (Mangifera indica L.) aroma. Flavour and Frag. J. 21(2):207-213. doi:10.1002/ffj.1703.

38. Pino JA, Mesa J, Muñoz Y, Martí MP, Marbot R (2005) Volatile components from mango (Mangifera indica L.) cultivars. J. Agric. Food Chem. 53(6):2213-2223. doi:10.1021/jf0402633.

39. Quijano CE, Salamanca G, Pino JA (2007) Aroma volatile constituents of Colombian varieties of mango (Mangifera indica L.). Flavour and Frag. J. 22(5):401-406. doi:0.1002/ffj.1812.

40. Ruiz-Sola MÁ, Rodríguez-Concepción M (2012) Carotenoid biosynthesis in Arabidopsis: a colorful pathway. The Arabidopsis Book 10:e0158. doi:10.1199/tab.0158.

41. Saleem M (2009) Lupeol, a novel anti-inflammatory and anti-cancer dietary triterpene. Cancer Lett. 285(2):109-115. doi:10.1016/j.canlet.2009.04.033.

42. Santos FA, Frota JT, Arruda BR, de Melo TS, de Castro Brito GA, Chaves MH, Rao VS 2012 Antihyperglycemic and hypolipidemic effects of α, β-amyrin, a triterpenoid mixture from Protiumheptaphyllum in mice. Lipids Health Dis. 11(1):98. doi:10.1186/1476-511X-11-98.

43. Shah KA, Patel MB, Patel RJ, Parmar PK (2010) Mangifera indica (mango). Pharmacogn. Rev. 4(7):42–48. doi:10.4103/0973-7847.65325.

44. Siddique HR, Saleem M (2011) Beneficial health effects of lupeol triterpene: a review of preclinical studies. Life Sci. 88(7):285-293. doi:10.1016/j.lfs.2010.11.020.

45. Sies H, Stahl W (1995) Vitamins E and C, beta-carotene, and other carotenoids as antioxidants. Am. J. Clin. Nutr. 62(6Suppl):1315S-1321S.

46. Singh Z, Singh RK, Sane VA, Nath P (2013) Mango-postharvest biology and biotechnology. CRC Crit. Rev. Plant Sci. 2(4):217-236. doi:10.1080/07352689.2012.743399.

47. Srivastava S, Singh RK, Pathak G, Goel R, Asif MH, Sane AP, Sane VA (2016) Comparative transcriptome analysis of unripe and mid-ripe fruit of Mangifera indica (var.“Dashehari”) unravels ripening associated genes. Sci Rep. 6:32557 doi:10.1038/srep32557.

48. Tafolla-Arellano JC, Zheng Y, Sun H, Jiao C, Ruiz-May E, Hernández-Oñate MA, González-León A, Báez-Sañudo R, Fei Z, Domozych D, Rose JK (2017) Transcriptome Analysis of Mango (Mangifera indica L.) Fruit Epidermal Peel to Identify Putative Cuticle-Associated Genes. Sci. Rep. 7:46163. doi:10.1038/srep46163.

49. White IR, Blake RS, Taylor AJ, Monks PS (2016) Metabolite profiling of the ripening of Mangoes Mangifera indica L. cv.‘Tommy Atkins’ by real-time measurement of volatile organic compounds. Metabolomics 12:57. doi: 10.1007/s11306-016-0973-1.

50. Wu HX, Jia HM, Ma XW, Wang SB, Yao QS, Xu WT, Zhou YG, Gao ZS, Zhan RL (2014) Transcriptome and proteomic analysis of mango (Mangifera indica Linn) fruits. J. Proteomics 105:19-30. doi:10.1371/journal.pone.0078644.

Figures legends:

Figure 1: Comparative species distribution of mango cultivars transcriptomes.

Figure 2: Terpenoid backbone biosynthesis in mango cultivars.

Figure 3: Carotenoid biosynthesis in mango cultivars.

Figure 4: Flavonoid biosynthesis in mango cultivars.

Table 1: Statistics of RNA-Seq experiments of mango cultivars.

RNA–Seq read datasets	References	Number of reads before filtering	Number of reads after filtering	Read length	Total bases
cv. *Langra*	Azim et al. 2014	12,153,196	5,314,486	80	425,158,880
cv. *Zill*	Wu et al. 2014	68,419,722	44,578,030	80	3,566,242,400
cv. Shelly	Luria et al. 2014	83,251,214	38,820,608	90	1,552,572,180
cv. Kent	Dautt-Castro 2015	22,018,116	14,210,006	72	1,023,120,432

Table 2: Statistics of unigenes data of mango cultivars assembled by Trinity program in this study.

RNA–Seq read datasets	N50 of unigenes (bases)	Total length of unigenes (bases)	Total number of unigenes	Length of largest unigenes (bases)	No. of unigenes reported in the respective paper
cv. *Langra*	525	14,216,135	30,953	5,821	30,509 (Azim et al. 2014)
cv. *Zill*	1,050	40,388,175	58,797	11,243	54,207 (Wu et al. 2014)
cv. Shelly	1,598	49,681,799	57,544	8,723	57,544 (Luria et al. 2014)
cv. Kent	1,017	58,412,829	85,036	8,305	80,969 (Dautt-Castro et al. 2015)

Figure 1:

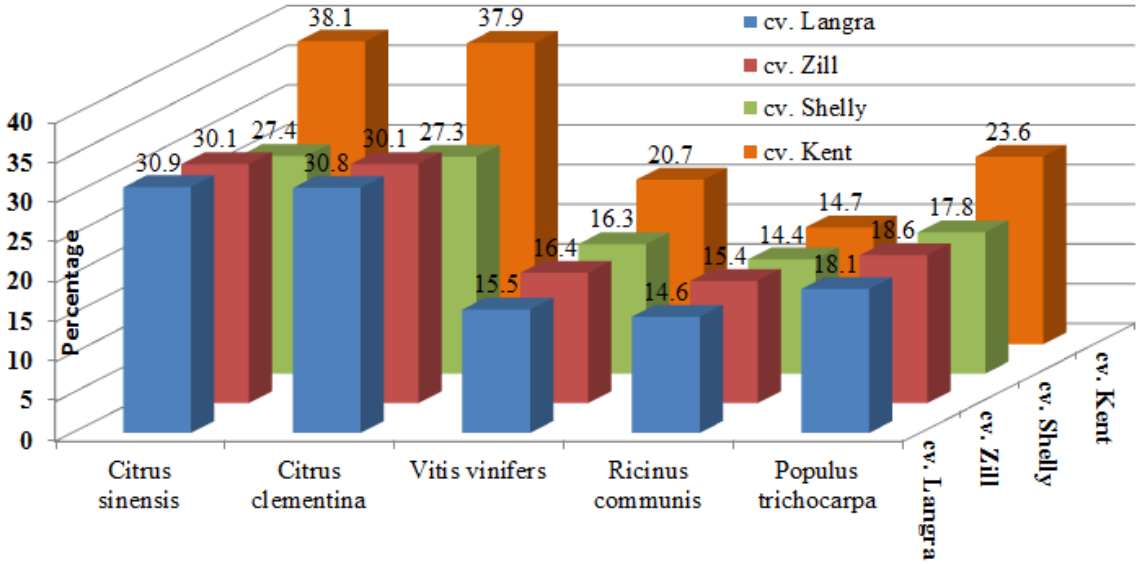

Figure 2:

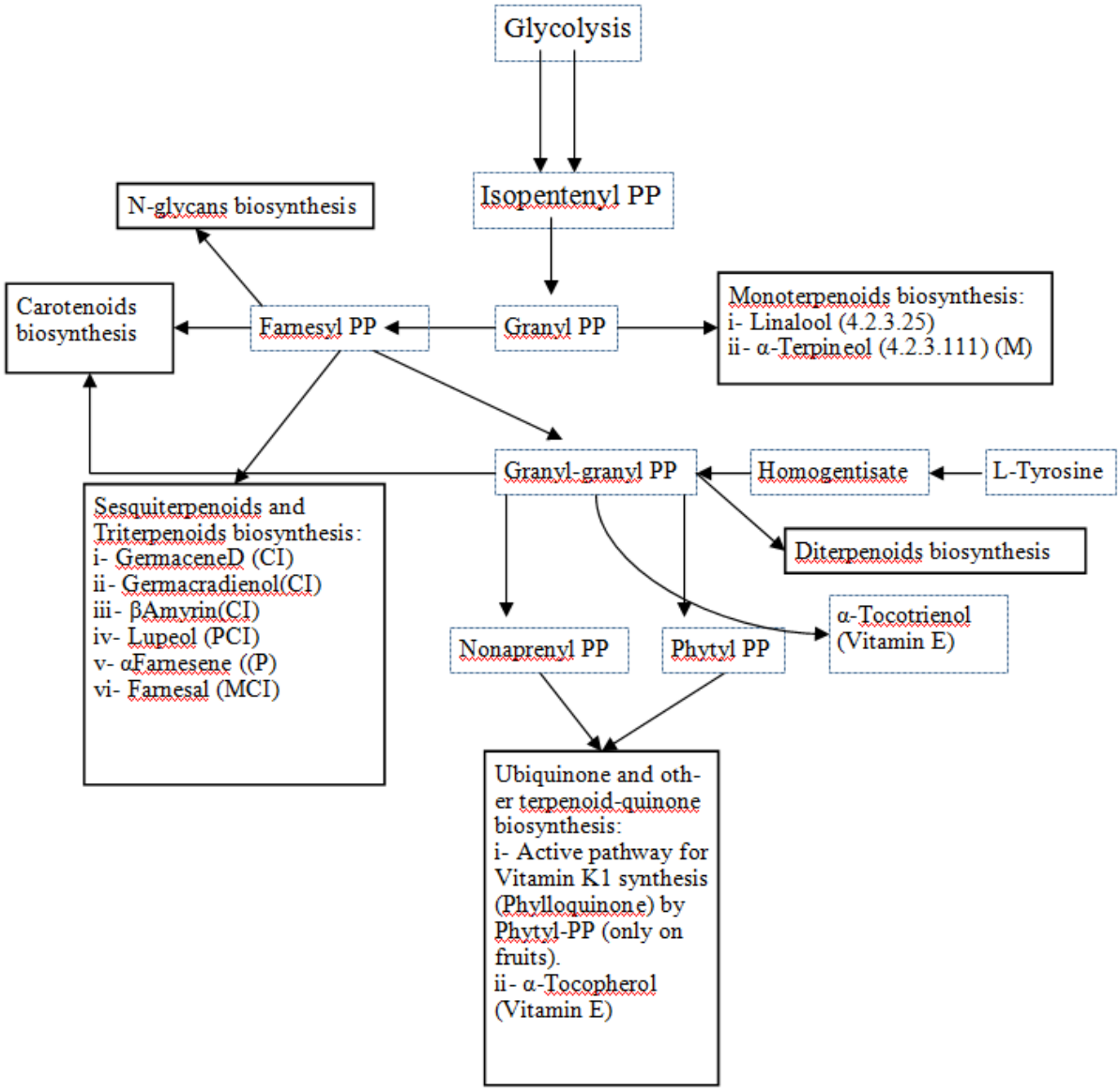

Figure 3:

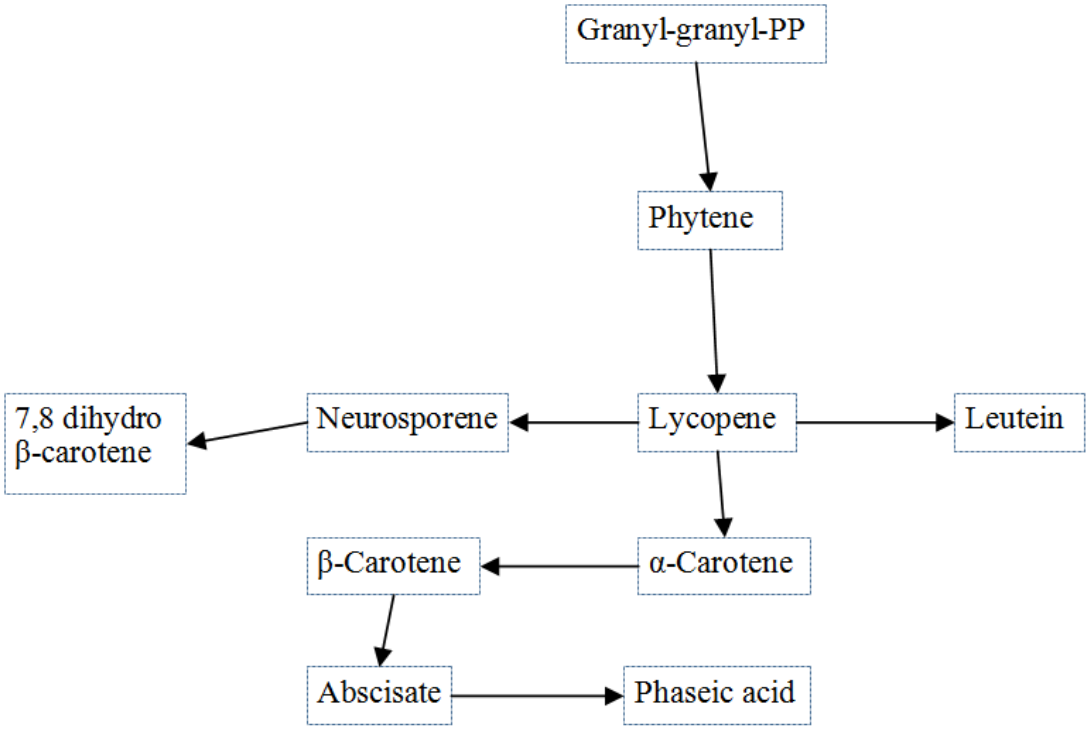

Figure 4: