J. Trop. Agric. and Fd. Sc. 41(2)(2013): 283 – 308

Identification of phytochemicals and the associated genes in Eksotika papaya at ripening index 5 using functional genomics

(Pengenalpastian fitokimia dan gen-gen berkaitan pada indeks pemasakan 5 buah betik Eksotika menggunakan genomik berfungsi)

S. Sanimah*, V. Maheswary**, J. Sarip***, O. Nur Qistina* and S. Vasanthi*

Keywords: fruit metabolomics and genomics, ripening index 5, bioinformatics, phytonutrients

Abstract

Phytochemicals and the associated genes involved in primary and secondary metabolite pathways at ripening index 5 of *Carica papaya* var. Eksotika were investigated using GC-MS based metabolomics approach and mRNA paired-end sequencing. A total of 54 metabolites were identified for polar and nonpolar compounds belonging to amino and organic acids, saturated/unsaturated fatty acids, sugars/sugar alcohols, simple phenolics, vitamins and sterols and alkanes. Ascorbic acid (vitamin C), α- and γ-tocopherols (vitamin E) and plant sterols such as stigmasterol, campesterol and β-sitosterol were among the metabolites identified in this study. A total of 48,765 putative unigenes were obtained from mRNA paired-end sequencing of which 59% (28,736) showed significant (E-value $\leq 10^{-5}$) expression to enzymes in the non-redundant (nr) database. Of these, only 4.9% (1,413) were predicted to be potentially involved in the biosynthesis and metabolism of primary and secondary metabolites based on KEGG pathway analysis. These include carotenoids, fruit volatiles including mono- and diterpenoids, stillbenoids, flavonoids, flavones and flavonols, anthocyanins, phenylpropanoids, betalains, isothiocyanates, steroids, saturated and unsaturated fatty acids, sugars and sugar alcohols, brassinosteroids, amino- and organic acids, abscisic acid, laticifers and proteases. GO analysis identified 67 of the unigenes in the molecular function category to be involved in antioxidant activity. By mapping to the KEGG pathway, 29 metabolites were identified to have direct association with several of the predicted papaya unigenes including metabolites belonging to amino and organic acids, sugars, fatty acids and plant sterols.

Introduction

Papaya (*Carica papaya* L.) is a productive climacteric fruit crop cultivated in tropical and sub-tropical regions worldwide for its delicious and nutritious fresh fruit. In Malaysia, it occupies about 3,000 ha contributing to about 45,000 metric tonnes of production and RM27 million of export value in 2011 (FAOSTAT 2013). Papaya belongs to the small family of Caricacea and is placed in the order Brassicales (Yu et al. 2009). It is one of the few

*Biotechnology Research Centre, MARDI Headquarters, Serdang, P.O. Box 12301, 50774 Kuala Lumpur, Malaysia
**Technical Services Centre, MARDI Headquarters, Serdang, P.O. Box 12301, 50774 Kuala Lumpur, Malaysia
***Horticultural Research Centre, MARDI Headquarters, Serdang, P.O. Box 12301, 50774 Kuala Lumpur, Malaysia
Authors' full names: Sanimah Simoh, Maheswary Vellupillai, Johari Sarip, Nur Qistina Othman and Vasanthi Sivarajah
E-mail: sanimah@mardi.gov.my

fruit tree crops that can flower and fruit throughout the year. It has a small genome of 372 Mbp (Ming et al. 2008) with nine pairs of chromosomes, a short juvenile phase of 3 – 8 months, and a short generation time of 9 – 15 months (Yu et al. 2009). Fruits are ready to be harvested 5 – 6 months after flowering, which can occur within 8 months of seed germination. The short generation time, continuous year-long flowering and relatively small genome makes papaya a good model for fruit-tree functional genomics studies among the perennials.

The draft of the sequenced papaya genome (Ming et al. 2008) provides a foundation for revealing the basis of *Carica*'s unique medicinal and nutritional properties. Although the genome size is three times the size of the *Arabidopsis* genome, the assembled ESTs unigene set was found to contain only 16,362 unigenes and the transcribed sequences matched only 3.6% (13.4 Mbp) of the whole genome due to lack of whole genome duplication (Ming et al. 2008; Paull et al. 2008). Due to this reduction in most gene families and biosynthetic pathways, papaya contains fewer genes predicted for the impact on sugar accumulation, ethylene synthesis and response, respiration, chlorophyll degradation and carotenoid synthesis compared to the *Arabidopsis* genome.

Like all climacteric fruits, papaya demonstrates a peak in respiration and ethylene production during fruit ripening. The timing of ethylene peak and other ripening events such as skin colour changes, carotenoid synthesis, flavour development and softening varies widely between species and cultivars (Gussman et al. 1993). Papaya fruit flesh colour, in particular, is indicative of the ripening process and is the result of the accumulation of carotenoid pigments in the chromoplasts of fruit cells, lycopene in the red fleshed fruit and β-carotene and β-cryptoxanthin in yellow fleshed fruits (Blas et al. 2010). Red flesh papaya, often called "strawberry papaya" in the market, is often preferred by some consumers even though it softens faster and has a shorter shelf life (Blas et al. 2010).

According to the reference values for nutrition labelling in the U.S. Food and Drug Administration, papaya is an excellent source of vitamin C, followed by folate, potassium, dietary fibre and vitamins A, E and K. Consumption of the fruit is recommended to prevent vitamin A deficiency, a cause of childhood blindness in tropical and sub-tropical developing countries (Rodriguez-Amaya 2003). Vitamin C as well as vitamins E, A and K in papaya also act as very powerful antioxidants through their concentration of pro-vitamin A carotenoid phytonutrients. These nutrients help prevent the oxidation of cholesterol which otherwise would build up in blood vessel walls, forming dangerous plaques that can eventually cause heart attacks or strokes (Palozza et al. 2008). Baybutt et al. (2000) showed in their animal studies that foods rich in vitamin A, such as papaya, can counter this effect. Eating lycopene-rich fruits such as papaya may also greatly reduce the development of prostate cancer (Jian et al. 2007). One way in which dietary vitamins E and C may exert this effect, may be through their association with a compound called paraoxonase, an enzyme that inhibits LDL cholesterol and HDL cholesterol oxidation (Durrington et al. 2001).

In tropical fruits like papaya, fruit volatiles also contribute to components of flavour and aroma. The biosynthesis of monoterpenes, diterpenes and sesquiterpenes fruit flavour volatiles occurs via the isoprenoid pathway derived substrate, geranyl diphosphate (GPP) (Pino et al. 2003). The accumulation of fatty acid derived volatiles also contributes to papaya's unique flavour profile. In cut fruits, fatty acid volatiles are formed via the oxylipin pathway while in intact fruits, they are formed via the β-oxidation pathway (Dixon and Hewett 2000). Methyl and ethyl ester

derivatives of lipid catabolism have also been identified as strong contributors of aroma (Pino et al. 2003).

In this paper, we report the phytochemicals and the associated genes involved in primary and secondary metabolite pathways at ripening index 5 in our local papaya variety, Eksotika using Gas Chromatography (GC-MS) based metabolomics and mRNA paired-end sequencing. These findings can contribute to the identification of potential chemical biomarkers for the development of standardised extracts and to delineate the pathways involved in the production of these phytochemicals in our local papaya variety. Identification of potential candidate genes can contribute to future genetic improvements through marker-assisted breeding or genetic engineering.

Materials and methods

Chemicals and reagents

Methanol and chloroform were purchased from Fisher Scientific (Leicester, UK) while ribitol and methylnonadecanoate were purchased from Sigma Aldrich (St. Louis, USA). Methoxyamine hydrochloride (HCL) and N-methyl-N-trimethylsilyl trifluroacetamide (MSTFA) were purchased from Acros Organics (New Jersey, USA). Pyridine was purchased from Merck (Darmstadt, Germany).

Fruit sampling

Three papaya plants from variety Eksotika grown at MARDI's papaya germplasm, Serdang, were randomly tagged for this study. Papaya fruits were harvested from the tagged plants at ripening index 2 and stored in the laboratory until the fruits reached ripening index 5 (Lam 1994). Ripening index 5 was selected based on the optimum stage for consumption (Lam 1994) (*Figure 1*). The whole fruit including the peel, was divided equally into 2 halves and cut into 1 – 2 cm cubes. One half was ground into powder using liquid nitrogen, freeze-dried and stored at –70 °C for metabolite profiling using GC-MS while the other half was immediately snap-frozen in liquid nitrogen and stored at –70 °C for total RNA extraction.

Metabolite extraction

Metabolite extraction and derivatisation steps prior to GC-MS analysis were performed using the two-phase methanol/chloroform method (Roessner et al. 2000) with slight modifications. Methanol (1.4 ml) containing ribitol and methylnonadecanoate as polar and non-polar internal standards respectively, was added to 500 mg of freeze-dried papaya samples and incubated for 15 min at 70 °C. The extracts were then immediately mixed vigorously with 1 volume of distilled deionised water. Finally, 750 µl of chloroform was added to the mixture before phase separation by centrifugation at 10,000 rpm for 10 min. Aliquots of polar and non-polar supernatant (250 µl) were dried in vacuo for 2 – 3 h.

Derivatisation

The dried extracts were re-dissolved in 50 µl of pyridine and sonicated for 10 min before methoximation step by adding 40 µl of 20 mg/ml methoxyamine HCL (in pyridine)

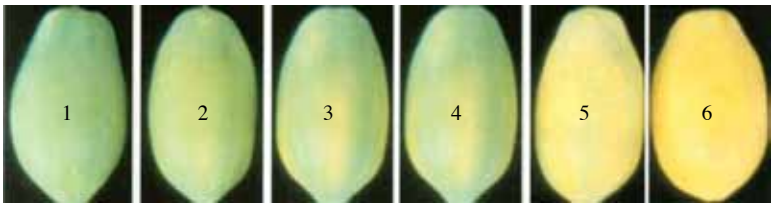

Figure 1. Ripening indices of Eksotika papaya. 1 - full green; 2 - green with trace of yellow; 3 - more green than yellow; 4 - more yellow than green; 5 - yellow with trace of green and 6 - fully yellow (Source: Lam 1994)

(Roessner et al. 2000). The mixtures were then sonicated again for 5 min and incubated with constant agitation for another 90 min at 37 °C. The trimethylsilylation (TMS) step was performed by adding 250 µl MSTFA to the extracts, followed by incubation for 1 h at 37 °C. The extracts were then cooled down to room temperature for at least 1 h prior to GC-MS analysis. For the non polar extract, only the TMS step with MSTFA was performed.

Analysis of polar and non polar metabolites by GC-MS

The polar and non polar samples were analysed using Agilent's GC-MS system. A sample volume of 1 µl was injected with a splitless mode into a GC-MS system consisting of an Agilent 6890 gas chromatograph and a HP5973 mass selective detector. The GC column used for the analysis was DB-5MS 5% phenyl methyl siloxane with an inner diameter of 0.25 µm and 0.25 µm film thickness. The initial oven temperature was set at 50 °C for 3 min, and then raised to a target temperature of 315 °C in 10 min at a rate of 10 °C/min. Helium was used as the carrier gas at a rate of 1ml/min. The injector and ion source temperatures were set at 330 °C and 250 °C respectively. Mass spectra was acquired using full scan monitoring mode with a mass scan range of 50 – 550 m/z after a solvent delay of 7 min. The method used for the GC-MS analysis was the optimised protocol which has been previously carried out specifically for papaya samples (unpublished data). The spectra of all chromatogram peaks were compared with the database library of NIST08 and WILEY05 and the retention time index of common primary and secondary metabolites. The chromatogram and mass spectra were evaluated using the Agilent Chemstation, Automated Mass Spectral Deconvolution and Identification System (AMDIS) and Deconvoluted Reporting Software (DRS), Agilent. After GC-MS analysis, all the peak areas were integrated and then normalised with respective internal standards (IS) to polar and non polar metabolites, to obtain the peak area ratios for polar and non polar metabolites in each chromatogram. Six replicates were used in this study.

Total RNA extraction and mRNA paired-end sequencing

Samples were ground to a fine powder using a mixture mill and liquid nitrogen. Total RNA was extracted from these frozen samples using the CTAB method according to Chang et al. (1993). The integrity of the extracted samples was first confirmed by gel-electrophoresis on a 1% formaldehyde agarose-gel before quantifying them using a nanodrop ND-1000 spectrophotometer. The samples were then concentrated to 200 ng/µl before outsourcing 20 ug of total RNA to the Beijing Genome Institute (BGI) for performing mRNA paired-end sequencing.

Blast analysis for identification of enzymes

A total of 10,641,022 raw sequence reads (91-bp long) with an average GC content of 46.8% were generated from mRNA paired-end sequencing. Clustering of the high quality, paired-end sequences revealed a total of 48,765 putative unigenes which were queried against the publicly available non-redundant (nr) protein database by Beijing Genome Institute (BGI) using the Basic Local Alignment Search Tool (BLAST) x (Altschul et al. 1997) search algorithm from the National Centre for Biotechnology Information (NCBI) website. After BLAST analysis, a total of 337,372 hits (including the top 10 or < hits for each sequence) were obtained. Selection of the best hit was determined as those unigenes containing ≥100 bp of sequence and a significant E-value of $\leq 10^{-5}$. Each selected unigene was then assigned an ID representing the protein with which the papaya cDNA had the closest similarity with.

Functional analysis using gene ontology (GO)

Functional analysis was performed on sequences that had a cut-off E value of $\leq 10^{-5}$ from BLAST analysis using Blast2GO software (http://www.blast2go.de/) (Conesa et al. 2005). Enzyme commission (EC) numbers were then assigned to sequences mapped to the KEGG pathway.

KEGG pathway assignments

All putative unigenes which showed significant (E-value $\leq 10^{-5}$) similarity to enzymes in the nr protein database were also mapped to the Kyoto Encyclopaedia of Genes and Genomes (KEGG) biochemical pathways according to the EC distribution in the pathway database. By mapping to the KEGG pathway, genes were associated with the biosynthesis of phytochemicals identified using GC-MS.

Results and discussion

Metabolomic profiling of Eksotika papaya using GC-MS

Separation of plant extracts to polar and nonpolar fractions is one of the established and common approaches in plant metabolomics, as plants have a complex and diverse array of metabolites which belong to different classes and have different chemical structures and concentrations. By separation of the extracts to polar and nonpolar fractions, better resolution of the metabolites can be achieved in a single GC-MS analysis (Ossipov et al. 2008). In this study, a total of 110 and 120 peaks were detected in polar and nonpolar fractions of the samples respectively.

A total of 54 metabolites from the total numbers of peaks were assigned metabolite identification by matching retention index and mass spectra with chemical standards or with database libraries from WILEY05 and NIST08. Most of the metabolites identified belonged to the group of amino/organic acids, saturated/unsaturated fatty acids, sugars/sugar alcohols, alkanes, simple phenolics, vitamins and plant sterols. Several metabolites such as fructose, glucose and glucopyranoside belonging to the sugars/sugar alcohols group and asparagine and malic acid belonging to the organic acid group produced multiple peaks (2 – 3 peaks). In this study, only one peak was chosen since the multiple peaks were highly correlated. A similar finding was also reported by Dobson et al. (2008). The retention time and peak area ratios for all the 54 metabolites are shown in *Table 1*.

Most of the amino acids and organic acids as well as sugars appeared in the polar fraction whereas the fatty acids, simple phenolics, plant sterols and alkanes appeared in the nonpolar fraction. The low percentage of identified peaks (45%) in both polar and nonpolar fractions as compared to the total peaks in the chromatogram could be due to several reasons such as a very low level of metabolites or no spectra information available in the database libraries (Mahdi et al. 2010).

Peak area ratios were calculated after normalisation with respective internal standards for polar and nonpolar metabolites. Based on the peak area ratios, the results indicated that metabolites in the sugars group including fructose, glucopyranoside and glucose had the highest intensity levels (peak area ratios of 3.94 – 12.39) followed by metabolites in the amino acid/organic acid groups including alanine, serine, threonine, glycine, asparagine, malic acid , gluconic acid and D-glucuronic acid (peak area ratios of 1 – 5) and hexadecanoic acid in the fatty acid group (peak area ratio 3.45) as compared to the other metabolites (*Table 1*). Ascorbic acid (vitamin C), α and γ-tocopherols (vitamin E) and plant sterols such as stigmasterol, campesterol and β-sitosterol were also identified in this study. These metabolites including fatty acids were among the metabolites that were consistently found in all replicates.

Table 1. Metabolites identified in polar and nonpolar fractions of *Carica papaya* var Eksotika at ripening index 5

Metabolite	RT ± SD	PA ratio ± SD	Metabolite	RT ± SD	PA ratio ± SD	Metabolite	RT ± SD	PA ratio ± SD
Amino acids			**Sugar/sugar alcohols**			**Phenolics/sterols/ vitamins**		
1. Alanine	10.61 ± 1.54	1.03 ± 0.82	22. D-Glucose	14.01 ± 0.30	3.94 ± 3.17	41 3-Bromo-5-ethoxy-benzyldehyde	15.14 ± 2.35	1.06 ± 1.03
2. Glycine	9.88 ± 0.02	5.25 ± 1.08	23. D-Galactose	17.45 ± 2.49	0.71 ± 1.16	42. Benzoic acid	13.29 ± 4.10	0.09 ± 0.03
3. Isoleucine	9.18 ± 0.04	0.21 ± 0.30	24. Fructose	13.83 ± 2.47	12.39 ± 7.37	43. Diacetylamino benzoic acid	18.53 ± 0.00	0.29 ± 0.20
4. Leucine	9.05 ± 0.04	0.06 ± 0.05	25. Ribose	14.41 ± 2.04	1.16 ± 1.12	44. Alpha tocopherol	19.37 ± 0.00	0.11 ± 0.05
5. Proline	9.55 ± 0.90	0.22 ± 0.39	26. D-Arabinose	13.54 ± 0.00	1.91 ± 1.33	45. Gamma tocopherol	18.40 ± 0.00	0.13 ± 0.00
6. Serine	9.69 ± 0.25	1.31 ± 0.73	27. D-Xylapyranose	15.64 ± 3.32	3.67 ± 3.84	46. Cholesterol	19.72 ± 0.32	0.07 ± 0.02
7. Threonine	9.78 ± 0.02	1.07 ± 0.33	28. Glucopyranoside	18.57 ± 2.75	5.21 ± 1.99	47. Campesterol	20.48 ± 0.01	1.04 ± 0.30
8. Valine	9.21 ± 0.29	0.08 ± 0.08	29. Myo-Inositol	14.42 ± 0.04	2.08 ± 1.54	48. Stigmasterol	20.67 ± 0 .00	0.43 ± 0.10
			30. Mellibiose	17.48 ± 0.00	0.05 ± 0.12	49. Beta sitosterol	21.28 ± 0.02	0.99 ± 0.30
			31. Xylose	14.63 ± 2.18	2.20 ± 2.12	50. Ascorbic acid	13.46 ± 0.02	0.26 ± 0.24
Organic acids			**Fatty acids**			**Alkanes**		
9. Lactic acid	8.18 ± 0.08	0.14 ± 0.19	32. Decanoic acid	16.22 ± 0.65	0.10 ± 0.06	51. Dodecane	11.58 ± 0.90	0.11 ± 0.08
10. Butanoic acid	10.89 ± 0.80	0.68 ± 0.55	33. Hexadecanoic acid	14.01 ± 0.00	3.45 ± 1.24	52. Eicosane	12.74 ± 1.50	0.06 ± 0.01
11. Fumaric acid	11.45 ± 2.42	0.37 ± 0.74	34. Octadecanoic acid	14.94 ± 0.00	1.14 ± 0.78	53. Tetracosane	15.37 ± 0.00	0.04 ± 0.00
12. Succinic acid	10.08 ± 0.00	0.03 ± 0.04	35. Tetradecanoic acid	12.99 ± 0.00	0.60 ± 0.22	54. Heneicosane	12.77 ± 1.59	0.07 ± 0.05
13. Aspartic acid	10.54 ± 0.01	0.34 ± 0.13	36. Docosanoic acid	17.18 ± 0.51	0.05 ± 0.01			
14. L Asparagine	12.15 ± 0.26	3.95 ± 1.43	37. Heptadecanoic acid	14.48 ± 0.00	0.05 ± 0.00			
15. DL-Malic acid	10.89 ± 0.00	2.17 ± 0.63	38. Alpha-linolenic acid	14.84 ± 0.00	1.89 ± 0.63			
16. D-Gluconic acid	12.56 ± 3.43	2.49 ± 0.90	39. Dodecanoic acid	11.94 ± 0.11	0.12 ± 0.03			

(cont.)

Table 1. *Cont.*

Metabolite	RT ± SD	PA ratio ± SD	Metabolite	RT ± SD	PA ratio ± SD	Metabolite	RT ± SD	PA ratio ± SD
17. D-Glucuronic acid	15.14 ± 2.35	1.80 ± 2.18	40. Myristic acid propyl ester	15.57 ± 0.00	0.15 ± 0.08			
18. Galacturonic acid	16.69 ± 1.11	0.44 ± 0.43						
19. Mannonic acid	12.78 ± 0.11	0.13 ± 0.07						
20. Gulonic acid	13.46 ±1.95	0.81 ± 0.88						
21. Propanoic acid	9.94 ± 0.00	0.84 ± 1.16						

RT = retention time; PA ratio = peak area ratio relative to respective internal standard for polar and non polar metabolites; SD = standard deviation value from mean of at least 3 replicate extractions of 500 mg freeze-dried samples

Identification of genes associated with primary and secondary metabolite pathways

Out of a total of 48,765 unigenes queried, and after selection of the best hit from the 10 or < hits, we obtained 28,736 putative unigenes (58.9%) which showed significant (E-value $\leq 10^{-5}$) similarity to enzymes in the nr protein database ranging from 100 bp – 7.8 Kb in size. *Table 2* shows the eighteen most abundant enzymes represented by 15 or more unigenes. These enzymes were found to be associated with the biosynthesis and or metabolism of phenylpropanoids, laticifers and proteases, stilbenoids, diarylheptanoids and gingerols, abscisic acids, fatty acids, amino-acids, sugar/sugar alcohols and steroids.

A total of 10 putative enzymes showed significant similarity to enzymes directly associated with the isoprenoid biosynthetic pathway leading to carotenoid biosynthesis as shown in *Table 3*. Carotenoids are C40 isoprenoids, synthesized in plastids from the universal 5-carbon (C_5) isoprenoid precursors into hydrocarbons (carotenes) and their oxygenated derivatives (xanthopylls) from the mevalonate MEP pathway (Fraser and Bramley 2004) (*Figure 2*). Four of these, *phytoene synthase* (*PSY*), *phytoene desaturase* (*PDS*), *zeta-carotene desaturase* (*ZDS*) and *lycopene* β-*cyclase* (*β-LCY*) showed similarity to genes previously identified in papaya. *Phytoene synthase* is a key regulator in carotenoid biosynthesis and has been found to be the rate-limiting enzyme in ripening tomato fruits (Fraser et al. 1994). The presence of the *β-LCY* enzyme in our local variety indicated that the b-ring cyclisation pathway was active for the synthesis of the yellow pigmented carotenoids, cryptoxanthins. Yamamoto (1964) also showed that the profile of yellow-fleshed fruits showed mostly β-cryptoxanthin and β-carotene derivatives which made up 75% of the total carotenoids content. The remaining 6 including β-*carotene hydroxylase* (*β-CHX*) which catalyses the production of β-cryptoxanthin, the major xanthophyll identified in papaya

Table 2. Most abundant putative enzymes associated with primary and secondary metabolite pathways with significant similarity (E-value ≤$10^{-5)}$ to proteins in nr database in *Carica papaya* var. Eksotika at ripening index 5

Biosynthesis and metabolism pathways/ putative genes identified (including isoforms) (EC*)	No. of unigenes	Blast X protein ID	Organism
Phenylpropanoids			
4-coumarate-CoA ligase (*4CL*) (EC 6.2.1.12)	24	XP_002329323	*Populus trichocarpa*
Cinnamoyl-CoA reductase (*CCR*) (EC 1.2.1.44)	16	ACE95172	*Populus tormentosa*
Laticifers and proteases			
Glycosyltransferase (EC 2.4.1.43)	77	XP_002320745	*Populus trichocarpa*
Invertase (*INV*) (EC 3.2.1.26)	21	dbjBAC21161	*Nicotiana tabacum*
Stilbenoids, diarylheptanoids and gingerols			
Cytochrome P450 (-)	71	AAZ05071	*Citrus sinensis*
Abscisic acids			
Aldehyde oxidase (-)	20	XP_002313633	*Populus trichocarpa*
Fatty acids			
Lipoxygenase (*LOX*) (EC 1.13.11.12)	20	AAR84664	*Carica papaya*
Carboxylesterase (CXE) (EC 3.1.1)	22	AAF26738	*Malus pumila*
Enoyl-CoA hydratase (EC 4.2.1.17)	16	NP_172142	*Arabidopsis thaliana*
Phospholipase (EC 3.1.4.4)	33	ADA72022	*Jatropha curcas*
Amino acids			
Serine/threonine protein phosphatase (-)	21	spQ9LU89	*Arabidopsis thaliana*
Serine-type peptidase (-)	21	NP_568577	*Arabidopsis thaliana*
Sugars/Sugar alcohols			
Alpha-glucosidase (*AGLU*) (EC 3.2.1.20)	32	ACC78255	*Carica papaya*
Alpha-amylase (EC 3.2.1.1)	16	dbjBAA33879	*Phaseolus vulgaris*
Beta-amylase (EC 3.2.1.2)	19	dbjBAF34362	*Citrus sinensis*
Beta-glucosidase (EC 3.2.1.23)	17	ADD17684	*Vitis vinifera*
Beta-1,3-glucanase (EC 3.2.1.39)	19	ADB24764	*Gossypium hirsutum*
Steroids			
Cycloartenol synthase (-)	20	dbjBAB83085	*Betula platyphylla*

*Enzyme commission number in KEGG pathway

fruit (Yamamoto 1964; Gamage et al. 2003), is shown in *Table 3* and *Figure 2*. A study related to fruit ripening genes in papaya by Devitt et al. (2006) also reported the presence of *ZDS* and *β-CHX*.

Fruit aroma compounds are generally derived from lipids, sugars and amino acids such as alcohols, aldehydes, ketones, sesquiterpenes, polypropanoids and esters (Schaffer et al. 2007). Blast analysis revealed a total of 11 putative enzymes which showed significant similarity to enymes associated with fruit volatiles in the terpenoid biosynthesis pathways (*Table 4*). A total of 5 enzymes were identified in the monoterpene biosynthesis pathway including *myrcene synthase*, *limonoid UDP glucosyltransferase*, *terpene cyclase* and *terpene synthase* (*Table 4* and *Figure 3*). In contrast to the findings by Pino et al. (2003) who reported that *S-linalool synthase* is a common fruit volatile in all papaya fruits and a major fruit volatile in the Hawaiian papaya cultivars, we identified a different

Table 3. Putative enzymes associated with carotenoid biosynthesis pathway with significant similarity (E-value <10-5) to proteins in nr database in *Carica papaya* var. Eksotika at ripening index 5

Biosynthesis pathway/putative genes identified (including isoforms) (EC*)	No. of unigenes	Blast X protein ID	Organism
Carotenoids biosynthesis pathway			
Phytoene synthase (*PSY*) (EC 2.5.1.32)	7	ABG72805	*Carica papaya*
Phytoene desaturase (*PDS*) (EC 1.14.99.-)	6	ABG77271	*Carica papaya*
Zeta-carotene desaturase (*ZDS)* (EC 1.14.99.30)	7	ACO40527	*Carica papaya*
Lycopene beta-cyclase (*β-LCY*) (EC 1.14.-.-)	1	ABD91578	*Carica papaya*
Beta-carotene hydroxylase (*β-CHX*) (EC 1.14.13.-)	2	ABA43899	*Coffea canephora*
Zeaxanthin epoxidase (*ZEP)* (EC 1.14.13.90)	4	BAI79260	*Citrus sinensis*
9-cis-epoxycarotenoid dioxygenase (*NCED*) (EC 1.13.11.51)	4	ABC26013	*Citrus clementina*
Carotenoid isomerase (*CRTISO*) (EC 5.-.-.-)	3	ACI12955	*Manihot esculenta*
Neoxanthin synthase (*NSY*) (-)	2	(-)	*Citrus sinensis*
Violaxanthin de-epoxidase (*VDE*) (EC 1.10.99.3)	1	ADH82116	*Citrus sinensis*

*Enzyme commission number in KEGG pathway

isoform of the gene, *linalool synthase 2* (*LIS2*), in this study. The remaining 6 including *(-)-germacrene D synthase*, *ent-kaurenoic acid oxidase/cytochrome P450, ent-kaurenoic acid hydroxylase*, *ent-kaurene oxidase*, *gibberellin 2-oxidase* and *ent-kaurene synthase* belonged to the sesquiterpene and diterpenoid biosynthesis pathways respectively.

A total of 30 putative enzymes were identified with similarity to the phenylpropanoid biosynthesis enzymes derived from the shikimic acid pathway (*Table 5* and *Figure 4*). *4-coumarate:CoA ligase* (*4CL)* had the highest number of unigenes (24) in this group followed by *cinnamoyl-CoA reductase* (*CCR*) (16) and *hydroxycinnamoyl CoA shikimate* (14). Three putative enzymes including *cinnamate 4-hydroxylase* (*C4H*), *S-adenosyl-L-methionine:salicylic acid carboxyl methyltransferase* (*SAMT*) and *caffeic acid 3-O methyltransferase* (*COMT*) were also reported by Devitt et al. (2006). *C4H* is the second enzyme of the phenylpropanoid pathway from phenylalanine to (hydroxyl) cinnamic acid while *COMT* catalyses the conversion of caffeic acid to ferulic acid, a precursor of monolignols required for lignification (Rogers and Campbell 2004). *SAMT* has been reported to catalyse the formation of the important floral scent compound, methylsalicylate (MSA). In contrast to previous findings where only 3 isoforms of the *phenylalanine ammonia lyas*e (*PAL*) gene were predicted in the papaya genome (Raes et al. 2003), we identified a total of 4 *PAL* unigenes in our papaya variety (*Table 5*). Of these, one was similar to a previously identified *PAL* gene in *C. papaya* fruit (GeneBank accession no. ACV30588). In the flavonoids/flavones and flavonols/anthocyanin biosynthesis pathways (*Table 5* and *Figure 4*), a total of 14 enzymes were identified with very low frequency of unigenes.

Lipoxygenase (*LOX*) genes which catalyse the deoxygenation of free fatty acid substrates, linoleic and linolenic acids to hydroperoxides in the oxylipin fatty acids biosynthesis pathway (Dixon and Hewett 2000), was relatively highly represented (20 unigenes) (*Table 6*). One *LOX* sequence was identified to be most similar to a previously identified cDNA from ripening *C. papaya* fruit (GeneBank accession no. AAR84664). *Alcohol dehydrogenase* (*ADH*), which contributes to the variety of C_6 and C_{10} alcohols in papaya or may be associated with the formation of phenylpropenes, another important group of flavour volatiles, was also identified

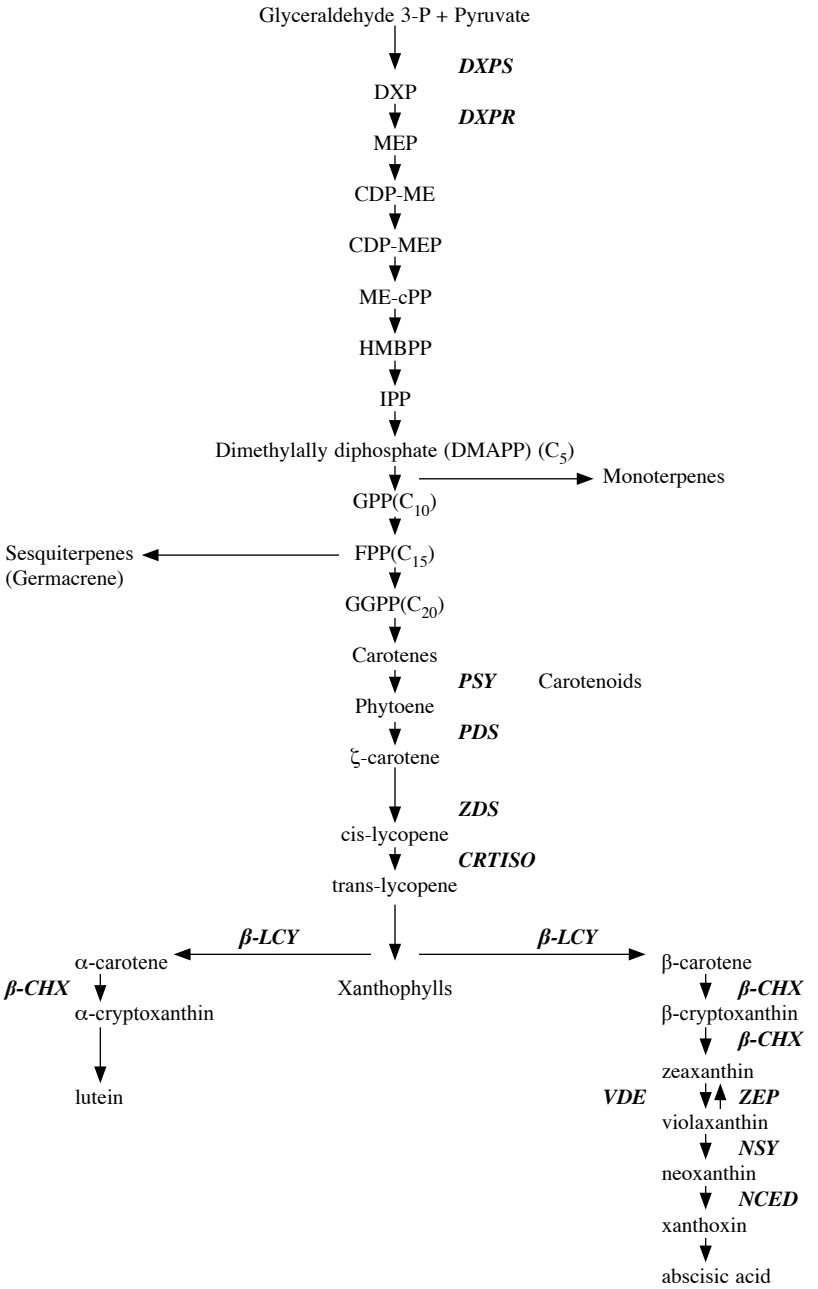

Figure 2. MEP pathway enzymes (in italics) and products towards carotenoid biosynthesis in Carica papaya var. Eksotika. DXPS:1-deoxy-D-xylulose 5-phosphate synthase; DXPR:1-deoxy-D-xylulose 5-phosphate reductase; PSY: phytoene synthase; PDS: phytoene desaturase; ZDS: ζ-carotene desaturase; CRTISO: carotenoid isomerase; β-LCY: lycopene-β-cyclase; β-CHX: β-carotene hydroxylase; ZEP: zeaxanthin epoxidase; VDE: violaxanthin de-epoxidase; NSY: neoxanthin synthase; NCED: 9-cis-epoxycarotenoid dioxygenase [Adapted from Devitt et al. 2006]

Table 4. Putative enzymes associated with terpenoid biosynthesis pathways with significant similarity (E-value ≤10^{-5}) to proteins in nr database in *Carica papaya* var. Eksotika at ripening index 5

Biosynthesis pathways/putative genes identified (including isoforms) (EC*)	No. of unigenes	Blast X protein ID	Organism
Monoterpene biosynthesis pathway			
Myrcene synthase (*TPS10*) (EC 4.2.3.15)	3	spQ93X23	–
Linalool synthase 2 (*LIS2*) (–)	1	AAD19840	*Clarkia breweri*
Limonoid UDP glucosyltransferase (–)	5	ABX46255	*Citrus maxima*
Terpene cyclase/mutase-related (–)	2	NP_001119109	*Arabidopsis thaliana*
Terpene synthase (–)	8	ACO40485	*Actinidia deliciosa*
Sesquiterpene biosynthesis pathway			
(–) *germacrene D synthase* (EC 4.2.3.8)	2	AAS66357	*Vitis vinifera*
Diterpenes biosynthesis pathway			
Ent-kaurenoic acid oxidase/cytochrome P450 (–)	4	AAO23063	*Pisum sativum*
Ent-kaurene oxidase (EC 1.14.13.78)	2	ADE61678	*Pyrus pyrifolia*
Ent-kaurenoic acid hydroxylase (*KAO2*)	6	AAK11564	*Arabidopsis thaliana*
(EC 1.14.13.79)	4	XP_002300430	*Populus trichocarpa*
Gibberellin 2-oxidase (EC 1.14.11.13)	5	XP_002311286	*Populus trichocarpa*
Ent-kaurene synthase (EC: 4.2.3.19)			

*Enzyme commission number in KEGG pathway

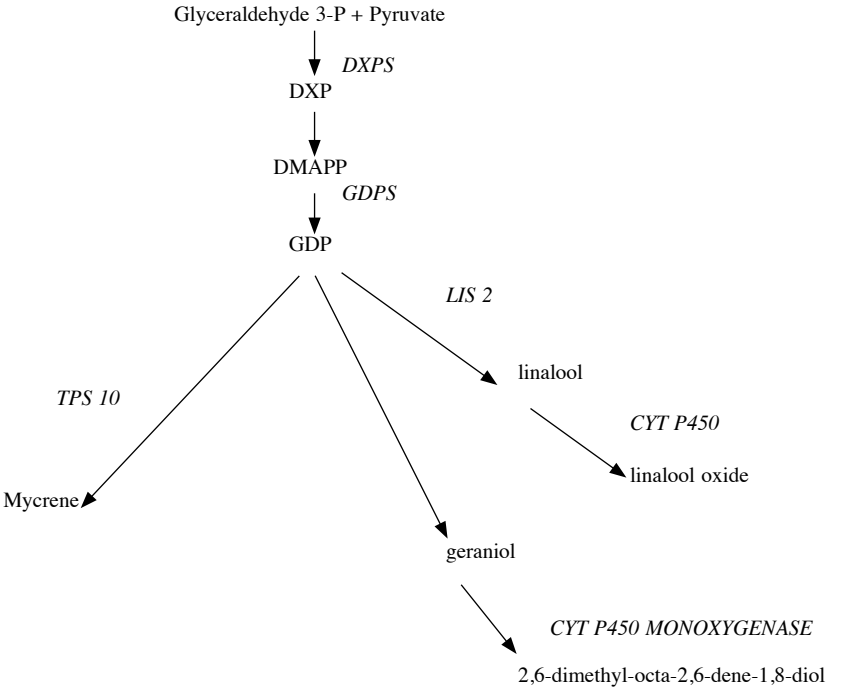

Figure 3. Putative metabolic pathway from pyruvate and glyceraldehyde-3-phosphate leading to monoterpene scent biosynthesis, and its related enzymes (in italics) in Carica papaya var. Eksotika. DXPS: deoxylylulose-5-phosphate synthase; GDPS: Geranyl diphosphate synthase; LIS 2: linalool synthase 2; CYT P450: cytochrome P450; TPS10: myrcene synthase (Adapted from Maheswary et al. 2008)

Table 5. Putative enzymes associated with biosynthesis pathways derived from shikimic acid pathway with significant similarity (E-value $\leq 10^{-5)}$ to proteins in nr database in *Carica papaya* var. Eksotika at ripening index 5

Biosynthesis and metabolism pathways/ putative genes identified (including isoforms) (EC*)	No. of unigenes	Blast X protein ID	Organism
Phenylpropanoid biosynthesis pathway			
Cinnamate 4-hydroxylase (*C4H*) (EC 1.14.13.11)	2	AAT39513	*Camptotheca acuminata*
S-adenosyl-L-methionine:salicylic acid carboxylmethyltransferase (*SAMT*) (EC 4.1.1.50)	4	ACH88356	*Nicotiana tabacum*
S-adenosylmethionine synthetase (EC 2.5.1.6)	2	XP_002319463	*Populus trichocarpa*
Caffeic acid 3-O methyltransferase (*COMT*) (EC 2.1.1.68)	2	spQ8GU25	*Rosa chinensis*
4-coumarate:CoA ligase (*4CL*) (EC 6.2.1.12)	24	XP_002329323	*Populus trichocarpa*
Cinnamoyl- CoA reductase (*CCR*) (EC 1.2.1.44)	16	ACE95172	*Populus tormentosa*
Phenylalanine ammonia-lyase (*PAL*) (EC 4.3.1.24)	4	ACV30588	*Carica papaya*
Cinnamyl alcohol dehydrogenase (EC 1.1.1.195)	7	ABM67695	*Citrus sinensis*
Ferulate 5- hydroxylase (*F5H*) (EC 1.14.-.-)	1	ABG73616	*Brassica napus*
Laccase (-)	5	XP_002312187	*Populus trichocarpa*
Polyphenol oxidase (EC 1.10.3.1)	2	AAB41022	*Vitis vinifera*
Acetyl-CoA synthetase (EC 6.2.1.1)	3	embCAA67130	*Solanum tuberosum*
3-dehydroquinate dehydratase (EC:1.1.1.25)	2	AAW65140	*Juglans regia*
3-dehydroquinate synthase (EC 4.2.3.4)	2	ACM45081	*Vitis vinifera*
Shikimate kinase (EC 2.7.1.12)	5	ACM45083	*Vitis vinifera*
Hydroxycinnamoyl CoA shikimate (EC:2.3.1.133)	14	XP_002303858	*Populus trichocarpa*
Prephenate dehydratase (EC 4.2.1.51)	8	dbjBAH10647	*Hevea brasiliensis*
Resveratrol (-)	2	ABH03018	*Vitis labrusca*
Caffeoyl-CoA O/3-O-methyltransferase (EC 2.1.1.104)	2	AAW55668	*Betula platyphylla*
p-coumarate:CoA-ligase (*pCL*) (-)	1	ADE96996	*Sorbus aucuparia*
Sinapoylglucose:choline sinapoyltransferase (EC 2.3.1.91)	1	embCAM91991	*Brassica napus* var. *napus*
Lignan glucosyltransferase (-)	1	dbjBAF96583	*Sesamum radiatum*
Coniferyl-alcohol glucosyltransferase (EC 2.4.1.111)	2	NP_566938	*Arabidopsis thaliana*
Chorismate mutase (EC:5.4.99.5)	5	NP_196648	*Arabidopsis thaliana*
Isochorismate synthase (EC 5.4.4.2)	3	ACX46383	*Populus trichocarpa*
Naringenin 3-dioxygenase (EC 1.14.11.9)	1	AAM12973	*Arabidopsis thaliana*
Acetylornithine aminotransferase (EC 2.6.1.11)	2	spO04866	*Alnus glutinosa*
Histidinol-phosphate aminotransferase (EC 2.6.1.9)	1	spQ9FEW2	*Nicotiana plumbaginifolia*
Tyrosine aminotransferase (EC 2.6.1.5)	1	NP_180058	*Arabidopsis thaliana*
Hydroxyphenylpyruvate dioxygenase (EC 1.13.11.27)	1	spO23920	*Daucus carota*
Flavonoid/Flavone and flavonol/Anthocyanin biosynthesis pathways			
Chalcone isomerase (*CHI*) (EC 5.5.1.6)	5	NP_180199	*Arabidopsis thaliana*
Chalcone synthase (*CHS*) (EC 2.3.1.74)	1	XP_002889339	Arabidopsis lyrata
Flavonoid 3'- hydroxylase (*F3H*) (EC 1.14.13.21)	5	ABC86840	*Vitis vinifera*
Flavonoid 5-glucosyltransferase (-)	1	dbjBAF96592	*Rosa hybrid cultivar*
Dihydrokaempferol 4-reductase (EC 1.1.1.219 1.1.1.234)	2	NP_194455	*Arabidopsis thaliana*
Flavonol synthase (*FLS*) (EC 1.14.11.23)	1	XP_002513987	*Ricinus communis*
Catechol o-methyltransferase (-)	1	XP_002320237	*Populus trichocarpa*
Coniferylalcohol 5-hydroxylase (-)	1	AAS92625	*Centaurium erythraea*
Dihydroflavonol 4-reductase (*DFR*) (EC 1.1.1.219)	1	spP52581	*Lupinus albus*
Isoflavone reductase (*IFR*) (-)	2	NP_194455	*Arabidopsis thaliana*
Leucoanthocyanidin dioxygenase (*LDOX*) (EC 1.14.11.19)	2	dbjBAA97487	*Arabidopsis thaliana*
Anthocyanin 5-O glucosyltransferase (EC 2.4.1.-)	1	dbjBAA89009	*Petunia x hybrida*
Anthocyanidin reductase (*ANR*) (EC 1.3.1.77)	1	ADD51354	*Theobroma cacao*
1-O-acylglucose:anthocyanin-O-acyltransferase (-)	3	dbjBAF99694	*Clitoria ternatea*

*Enzyme commission number in KEGG pathway

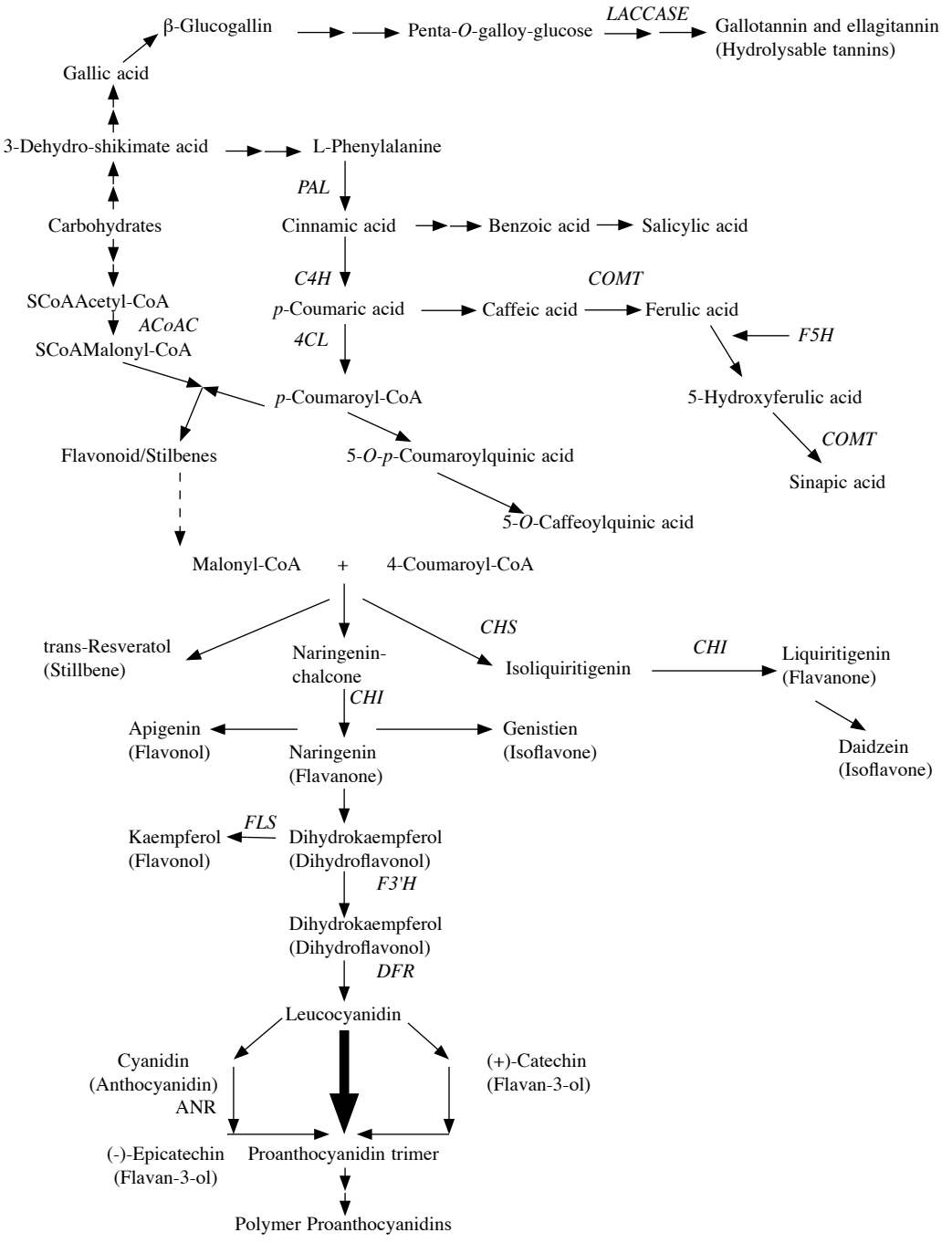

Figure 4. Putative metabolic pathway in Carica papaya var. Eksotika from shikimic acid pathway leading to enzymes (in italics) involved in (a) phenylpropanoid biosynthesis pathway {PAL: phenylalanine ammonia-lyase; C4H: cinnamate 4-hydroxylase; COMT: caffeic acid O-methyltransferase; 4CL: p-coumarate:CoA ligase; F5H: ferulate hydroxylase; ACoAC: acetylCoA carboxylase}; and (b) flavonoids, flavones and flavonol, and anthocyanin biosynthesis pathways {CHS: chalcone synthase; CHI: chalcone isomerase; FLS: flavonol synthase; DFR: dihydroflavonol 4-reductase; F3'H: flavonoid 3'-hydroxylase; ANR: anthocyanidin reductase}. [Source: Crozier et al. 2006]

Table 6. Putative enzymes associated with fatty acids biosynthesis pathway with significant similarity (E-value $\leq 10^{-5}$) to proteins in nr database in *Carica papaya* var. Eksotika at ripening index 5

Biosynthesis and metabolism pathways/putative genes identified (including isoforms) (EC)*	No. of unigenes	Blast X protein ID	Organism
Fatty acids biosynthesis pathway			
Lipoxygenase (*LOX*) (EC 1.13.11.12)	20	AAR84664	*Carica papaya*
Alcohol dehydrogenase (*ADH*) (EC 1.1.1.1)	8	YP_001671692	*Boechera lyallii*
Acetyl-CoA carboxylase (-)	4	embCAA47926	*Carica papaya*
3-ketoacyl-CoA thiolase (EC 2.3.1.16)	3	AAQ93070	*Glycine max*
Carboxylesterase (*CXE*) (EC 3.1.1)	22	AAF26738	*Malus pumila*
Beta-ketoacyl-ACP synthase (*KAS III*) (-)	4	ACY78677	*Elaeis guineensis*
Fatty acid elongase (-)	2	ABX82799	*Pistacia chinensis*
Acyl-ACP thioesterase (-)	4	AAW88320	*Jatropha curcas*
Acetylajmalan acetylesterase (-)	1	ABU96743	*Rauvolfia serpentina*
Omega-3 fatty acid desaturase (-)	6	NP_178516	*Jatropha curcas*
Long-chain-fatty-acid--CoA ligase (-)	5	-	*Arabidopsis thaliana*
Glycerol-3-phosphate dehydrogenase (EC 1.1.1.8)	9	NP_198877	*Arabidopsis thaliana*
Enoyl-CoA hydratase (EC 4.2.1.17)	16	NP_172142	*Arabidopsis thaliana*
Acyl-CoA thioesterase (EC 3.1.2.14)	2	XP_002892048	*Arabidopsis lyrata*
3-ketoacyl-CoA reductase (-)	5	AAY23354	*Gossypium hirsutum*
Enoyl-CoA isomerase (EC 5.3.3.8)	1	spO49809	*Brassica napus*
3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35)	2	NP_187342	*Arabidopsis thaliana*
Alpha-dioxygenase 2 (EC 1.-.-.-)	3	embCAH64542	*Solanum lycopersicum*
Acyl-CoA dehydrogenase (EC 1.3.99.3)	4	NP_187337	*Arabidopsis thaliana*
Phospholipase (EC 3.1.4.4)	33	ADA72022	*Jatropha curcas*
Digalactosyldiacylglycerol synthase (EC 2.4.1.241)	4	ABA55727	*Vigna unguiculata*
Ubiquitin thioesterase (EC 3.1.2.-)	12	NP_566666	*Arabidopsis thaliana*

*Enzyme commission number in KEGG pathway

(Devitt et al. 2006). However, *acetyl-CoA carboxylase* and *3-ketoacyl-CoA thiolase*, which had similarity to putative enzymes of the peroxisomal fatty acid β-oxidation pathway and encoding the final 2 steps of fatty acid degradation to acetyl-coenzyme A were present in very low frequency (*Table 6*). *Phospholipase* was the most highly represented in this group with 33 unigenes followed by carboxylesterase (*CXE*) (22 unigenes). *Acyl-ACP thioesterase*, an important enzyme substrate reported in plastids and *β-ketoacyl-ACP synthase* (*KAS III*), proposed as one control point within fatty acid synthesis (Schmid 2004), were identified in very low frequency.

With regards to a previous study on fruit ripening-related genes in the same papaya variety Eksotika (Sew et al. 2011), we also identified several unigenes with putative biochemical functions associated with fruit ripening such as *S-adenosylmethionine synthetase* and *polygalacturonase*, and fruit softening processes such as *β-fructofuranosidase* and *β-1,3-glucanase*. *Polygalacturonase* has been suggested to play an essential role in cell wall softening, particularly at the later stage of fruit ripening (Fabi et al. 2009). The number of unigenes (19) which matched to *β-1,3-glucanase* enzyme (*Table 2*) in the sugars/sugar alcohols biosynthesis pathway was comparable to that predicted by El Moussaoui et al. (2001) (at least 27 *beta-1,3-glucanase* proteins) in papaya latex activity.

Fruits like papaya depend upon a continuous supply of sucrose for sweetness when ripe. Two previously identified enzymes in papaya, *sucrose synthase* (*SUS*) and *sucrose phosphate synthase*, (*SPS*), thought to be involved in the transport of sucrose through the symplast pathway during early fruit growth (Zhang et al. 2006) were also identified in the starch and sucrose metabolism pathway in our papaya variety. *Glycosyltransferase*, also reported to be involved with sucrose transport, was the most highly represented (77 unigenes) in our study (*Table 2*).

After fruit growth stops, sucrose accumulation is thought to involve a *cell wall invertase* gene cleaving the sucrose arriving via the phloem to hexoses. Such a *cell wall invertase,* previously isolated from papaya, has been reported to be required for glucose and fructose accumulation (Devitt et al. 2006; Zhou et al. 2003). In our study, we found a total of 21 unigenes encoding *invertase* (*INV*) (*Table 2*). The hexoses are then taken up into the fruit cells and vacuoles via hexose transporters (Caspari et al. 1994). Papaya has been reported to contain at least 4 hexose transporter genes (Paull et al. 2008) and in our variety, we found 10 unigenes that had significant similarity to hexose transporters.

Functional analysis using GO

Of the total 28,736 putative unigenes which showed significant (E-value $\leq 10^{-5}$) similarity to enzymes in the nr protein database, only 11,086 (38.6%) were grouped by GO analysis into the broad functional categories of biological process (BP), molecular function (MF) and cellular component (CC) as shown in *Table 7*. Many of these unigenes were represented in more than one category as seen by the total number in each category. The remaining 17,650 (61.4%) either showed insufficient similarities to any proteins (NA or no hits) or hit proteins without a GO identifier (unclassified).

Among the biological process category, the largest proportion of functionally assigned sequences fell into the metabolic (49.5%) and cellular (43.3%) processes. Among the molecular function (MF) category, the most highly represented group was binding (57.2%) followed by catalytic activity (51.0%). In the cellular component category, the most highly represented group was cell and cell part (69.8%) followed by the organelle group (51.1%).

Interestingly, 67 of the unigenes (0.6%) in the MF category were identified to be involved in antioxidant activity as shown in *Table 8*. The highest number of unigenes in the reactive oxygen scavengers (ROS) group was represented by *peroxidase* (*POD*) (27) followed by *glutathione peroxidase* (10), *manganese superoxide dismutase* (7), *respiratory burst oxidase* (6) and *chorismate mutase* (5). *Polyphenol oxidases* (*PPO*) and *peroxidases* (*POD*) are the 2 major groups of enzymes involved in oxidation (Paull et al. 2008). Other ROS identified include *glutathione S-transferase*, *cytosolic ascorbate peroxidase* and *germin-like protein 2*.

KEGG pathway assignment of predicted papaya genes related to identified metabolites

After selection of the best hit based on E-value $\leq 10^{-5}$ and removing duplicates, a total of 12,878 unigenes (44.8%) were mapped to the KEGG pathways. Of these, the largest number of unigenes (319) (2.5%) belonged to the starch and sucrose metabolism pathway, followed by phenylpropanoid biosynthesis (219) (1.7%), amino sugar and nucleotide sugar metabolism (179) (1.4%), glycerophospholipid metabolism (167) (1.3%) and carbon fixation in photosynthetic organisms (133) (1.03%). All the other pathways had less than 1% of the unigenes.

The integration between putative genes and the metabolites identified using GC-MS are shown in *Tables 9 – 13*. By mapping the KEGG pathway, 29 metabolites were identified to have association directly with predicted papaya unigenes. These included

Table 7. Functional classification of *Carica papaya* var. Eksotika genes at ripening index 5 using Blast2GO level 2 (E-value $\leq 10^{-5}$)

Category	Putative functions	No. of unigenes	% (out of 11,086 sequences)
Biological process	Anatomical structure formation	118	1.1
	Biological adhesion	3	0.03
	Biological regulation	1,068	9.6
	Cell killing	1	0.01
	Cellular component biogenesis	215	1.9
	Cellular component organization	605	5.5
	Cellular process	4,795	43.3
	Death	49	0.4
	Development process	929	8.4
	Establishment of localization	1,083	9.8
	Growth	103	0.9
	Immune system process	65	0.6
	Localization	1,110	10.0
	Locomotion	3	0.03
	Metabolic process	5,488	49.5
	Multi-organism process	210	1.9
	Multicellular organismal process	682	6.2
	Pigmentation	814	7.3
	Reproduction	513	4.6
	Reproductive process	509	4.6
	Response to stimulus	1,698	15.3
	Rhythmic process	29	0.3
	Viral reproduction	4	0.04
Molecular function	Antioxidant activity	67	0.6
	Binding	6,338	57.2
	Catalytic activity	5,648	51.0
	Electron carrier activity	6	0.05
	Enzyme regulator activity	99	0.9
	Molecular transducer activity	304	2.7
	Structural molecule activity	254	2.3
	Transcription regulator activity	114	1.0
	Translation regulator activity	154	1.4
	Transporter activity	655	5.9
Cellular component	Cell and cell part	7,740	69.8
	Envelope	449	4.1
	Extracellular region	147	1.3
	Extracellular region part	10	0.1
	Macromolecular complex	1,021	9.2
	Membrane-enclosed lumen	270	2.4
	Organelle	5,663	51.1
	Organelle part	1,391	12.6
	Virion and virion part	5	0.05
No hits or unclassified	NA or unknown, unnamed, hypothetical and others	17,650	

Table 8. *Carica papaya* var. Eksotika unigene sequences with significant similarity (E-value ≤10^{-5}) to proteins associated with antioxidant activity in the molecular function category from GO analysis

Predicted genes and gene functions (including isoforms) (EC)*	No. of unigenes	Blast X protein ID	Organism
Glutathione S-transferase (EC 2.5.1.18)	3	ADB85090	*Jatropha curcas*
Glutathione peroxidase (EC 1.11.1.9)	10	XP_002299535	*Populus trichocarpa*
Cytosolic ascorbate peroxidase (EC 1.11.1.5)	3	ABX79340	*Vitis vinifera*
2-cys peroxiredoxin/Peroxidase (EC 1.11.1.15)	2	ACZ56426	*Vigna radiata*
Chorismate mutase (*CM*) (EC 5.4.99.5)	5	XP_002324083	*Populus trichocarpa*
Alpha-dioxygenase (EC 1.-.-.-)	2	AAG52078	*Arabidopsis thaliana*
Manganese superoxide dismutase (EC 1.15.1.1)	7	AAT68778	*Camellia sinensis*
Respiratory burst oxidase (EC 1.6.3.- 1.11.1.-)	6	Q948U0	*Solanum tuberosum*
Ascorbate peroxidase (EC 1.11.1.11)	2	AAN60795	*Brassica juncea*
Peroxidase (*POD*) (EC 1.11.1.7)	27	ACN97180	*Populus trichocarpa*
Acid phosphatase (–)	1	ABN06091	*Medicago truncatula*
Oxygen binding/ steroid hydroxylase (EC 1.14.-.-)	1	NP_180239	*Arabidopsis thaliana*
GCN5L1 family protein (–)	1	NP_180592	*Arabidopsis thaliana*
Annexin 3 (–)	1	ACQ65866	*Brassica juncea*
Cationic amino acid transporter (–)	2	NP_187022	*Arabidopsis thaliana*
Fiber protein Fb12 (–)	1	AAN77151	*Gossypium barbadense*
Copia-like retrotransposable element (–)	1	BAB01972	*Arabidopsis thaliana*
CarD-like transcriptional regulator family protein (–)	1	ABA94062.2	*Oryza sativa*
Putative gag-pol polyprotein, identical (–)	1	AAT38758	*Solanum demissum*
OSIGBa0134J07.9 (–)	1	CAH66391	*Oryza sativa*
Germin-like protein 2 (–)	3	XP_002278736	*Vitis vinifera*
Nucleolar complex protein 2 homolog (–)	1	*NOC2L_ARATH*	–
RNA/lariat debranching enzyme (–)	2	CAA19746	*Arabidopsis thaliana*
Cytochrome P450 (–)	2	XP_002308312	*Populus trichocarpa*
Total	67		

*Enzyme commission number in KEGG pathway

Table 9. Association of amino acid metabolites with papaya unigenes in KEGG and nr databases

Metabolites	Putative enzymes identified (including isoforms) (EC)*	No. of unigenes in KEGG db	No. of unigenes in nr db	Organism/BLASTx protein ID	KEGG ID	KEGG pathway (KEGG pathway ID)
Glycine	*Glycine hydroxymethyltransferase* (EC 2.1.2.1)	5	1	*Arabidopsis thaliana*/ NP_564473	ath:AT1G36370	Glycine, serine and threonine metabolism (ko00260)
Serine	*Serine racemase* (EC 5.1.1.18)	2	1	*Oryza sativa* Japonica Group/NP_001053521	ath:AT4G11640	Glycine, serine and threonine metabolism (ko00260)
Threonine	*Threonine synthase* (EC 4.2.3.1)	2	2	*Arabidopsis thaliana*/ NP_194713	ath:AT4G29840	Glycine, serine and threonine metabolism (ko00260)
L-Valine/ L-Isoleucine/ L-Leucine	*Branched-chain amino acid aminotransferase* (EC 2.6.1.42)	8	3	*Oryza sativa* Japonica Group/NP_001042537	ath:AT1G10070	Valine, leucine and isoleucine biosynthesis (ko00290)
Alanine	*Beta-ureidopropionase* (EC 3.5.1.6)	2	1	*A. lyrata* subsp. Lyrata/ XP_002864896	ath:AT5G64370	Beta-alanine metabolism (ko00410)

*Enzyme commission number in KEGG pathway

Table 10. Association of organic acid metabolites with papaya unigenes in KEGG and nr databases

Metabolites	Putative enzymes identified (including isoforms) (EC)*	No. of unigenes in KEGG db	No. of unigenes in nr db	Organism/BLASTx protein ID	KEGG ID	KEGG pathway (KEGG pathway ID)
Succinic acid	*Succinate dehydrogenase* (EC 1.3.5.1)	5	5	*Arabidopsis lyrata* subsp. *lyrata*/ XP_002866741	ath:AT5G40650	Citrate cycle (TCA cycle) (ko00020)
Aspartic acid	*Aspartate transaminase* (EC 2.6.1.1)	7	–	–	pop:POPTR_835881	Carbon fixation in photosynthetic organisms (ko00710)
DL-Malic acid (Malate)	*Malate dehydrogenase* (EC 1.1.1.37)	15	15	*Glycine max*/AAC37464	pop:POPTR_707785	Carbon fixation in photosynthetic organisms (ko00710)
D-Glucuronic acid	*Inositol oxygenase* (EC 1.13.99.1)	1	1	*Eucalyptus grandis*/ ACF04280	ath:AT4G26260	Ascorbate and aldarate metabolism (ko00053)
D-Galacturonic acid	*Polygalacturonase* (*PGA*) (EC 3.2.1.15)	6	5	*Carica papaya*/ ACV85695	ath:AT3G15720	Pentose and glucuronate interconversions (ko00040)

*Enzyme commission number in KEGG pathway

Table 11. Association of fatty acid metabolites with papaya unigenes in KEGG and nr databases

Metabolites	Putative enzymes identified (including isoforms) (EC)*	No. of unigenes in KEGG db	No. of unigenes in nr db	Organism/BLASTx protein ID	KEGG ID	KEGG pathway (KEGG pathway ID)
Decanoic acid/ Octadecanoic acid/ Tetradecanoic acid/ Dodecanoic acid	*Acyl-CoA thioesterase* (EC 3.1.2.14)	2	2	*Arabidopsis lyrata*/ XP_002892048	ath:AT1G01710	Fatty acid biosynthesis (ko00061)
Hexadecanoic acid	*Palmitoyl-CoA hydrolase* (EC 3.1.2.2)	7	–	–	ath:AT3G25110	Unsaturated fatty acids biosynthesis (ko01040)
Docosanoic acid	*Ubiquitin thioesterase* (EC 3.1.2.-)	4	12	*Arabidopsis thaliana*/ NP_566666	ath:AT1G50670	Unsaturated fatty acids biosynthesis (ko01040)
Alpha-linolenic acid/ linoleic acid	*3-hydroxyacyl-CoA dehydrogenase* (EC 1.1.1.35)	5	2	*Arabidopsis thaliana*/ NP_187342	ath:AT3G06860	Alpha-linolenic acid/ linoleic acid metabolism (ko00592/ ko00591)
	Lipoxygenase (EC 1.13.11.12)	20	20	*Carica papaya*/ AAR84664	ath:AT3G45140	Alpha-linolenic acid/ linoleic acid metabolism (ko00592/ ko00591)
	Steroid hydroxylase (EC 1.14.-.-)	1	1	*Arabidopsis thaliana*/ NP_180239	ath:AT4G36380	Brassinosteroid biosynthesis (ko00905)
	3-ketoacyl-CoA thiolase (EC 2.3.1.16)	4	3	*Glycine max*/ AAQ93070	ath:AT2G33150	Biosynthesis of unsaturated fatty acids (ko01040)
	Alpha-dioxygenase (EC 1.-.-.-)	4	3	*Solanum lycopersicum*/ embCAH64542	ath:AT3G01420	Alpha-Linolenic acid metabolism (ko00592)
	3,2-trans-enoyl-CoA isomerase (EC 5.3.3.8)	1	1	*Brassica* napus/ spO49809	AFT_183312	Alpha-linolenic acid metabolism (ko00592)

*Enzyme commission number in KEGG pathway

Table 12. Association of sterols and vitamins with papaya unigenes in KEGG and nr databases

Metabolites	Putative enzymes identified (including isoforms) (EC)*	No. of unigenes in KEGG db	No. of unigenes in nr db	Organism/BLASTx protein ID	KEGG ID	KEGG pathway (KEGG pathway ID)
Beta sitosterol/ Campesterol	*Delta24-sterol reductase* (*DWARF*) (EC 1.3.1.72)	3	1	*Arabidopsis thaliana*/ NP_188801	ath:AT3G19820	Steroid biosynthesis (ko00100)
Stigmasterol	*C-22 sterol desaturase* (*CYP710*A) (-)	3	1	*Arabidopsis thaliana*/ NP_180996	ath:AT2G34500	Steroid biosynthesis (ko00100)
Vitamin E (Alpha tocopherol)	*VITAMIN E DEFICIENT 1 (VTE1)*; *Tocopherol cyclase* (-)	6	1	*Eucalyptus gunnii*/ AAP97931	ath:AT4G32770	Ubiquinone and other terpenoid-quinone biosynthesis (ko00130)
Gama tocopherol	*Gamma-tocopherol O-methyltransferase* (*G-TMT)* (EC 2.1.1.95)	1	1	*Gossypium hirsutum*/ ABE41798	ath:AT1G64970	Ubiquinone and other terpenoid-quinone biosynthesis (ko00130)
Cholesterol	*Sterol delta7 reductase* (*DWARF*) (EC 1.3.1.21)	1	1	*Tropaeolum majus*/ AAR29980	ath:AT1G50430	Steroid biosynthesis (ko00100)
Ascorbic acid (Vitamin C)	*Ascorbate peroxidase* (EC 1.11.1.5)	4	10	*Carica papaya*/ ABS01350	vvi:100233013	Glutathione metabolism (ko00480)/Ascorbate and aldarate metabolism (ko00053)

*Enzyme commission number in KEGG pathway

Table 13. Association of sugars/sugar alcohols with papaya unigenes in KEGG and nr databases

Metabolites	Putative enzymes identified (including isoforms) (EC)*	No. of unigenes in KEGG db	No. of unigenes in nr db	Organism/BLASTx protein ID	KEGG ID	KEGG pathway (KEGG pathway ID)
D-Galactose	*α-galactosidase* (EC 3.2.1.22)	7	8	*Carica papaya*/ AAP04002	pop:POPTR_249183	Galactose metabolism (ko00052)
D-Glucose	*α-glucosidase* (*AGLU*) (EC 3.2.1.20)	4	3	*Arabidopsis thaliana*/ NP_001031247	ath:AT5G11720	Starch and sucrose metabolism (ko00500)
1D-Myo-inositol	*Myo-inositol monophosphatase* (EC 3.1.3.25)	2	1	*Arabidopsis thaliana*/ NP_564376	pop:POPTR_586756	Inositol Phosphate metabolism (ko00562)
Mellibiose	*β-fructofuranosidase* (EC 3.2.1.26)	10	16	*Citrus sinensis*/ dbjBAF34362	ath:AT1G12240	Galactose metabolism (ko00052)
Xylose	*Xylose isomerase* (EC 5.3.1.5)	3	6	*Arabidopsis thaliana*/ dbjBAE98492	ath:AT5G57655	Pentose and glucuronate interconversions (ko00040)
Fructose	*Fructose biphosphatase* (EC 3.1.3.11)	4	–	-	ath:AT1G07110	Fructose and mannose metabolism (ko00051)
	Fructose biphosphate aldose (EC 4.1.2.13)	8	6	*Plantago major*/ embCAL34034	pop:POPTR578574	
Ribose	*Ribose 5-phosphate isomerase* (EC 5.3.1.6)	3	1	*Arabidopsis thaliana*/ NP_178238	pop:POPTR640727	Pentose phosphate pathway (ko00030)/ Carbon fixation in photosynthetic organisms (ko00710)
D-arabinose	*Arabinose kinase* (EC 2.7.1.46)	13	7	*Gossypium hirsutum*/ ACJ11758	ath:AT4G16130	Amino sugar and nucleotide metabolism (ko00520)
Glucopyranoside	*Beta-glucosidase* (EC 3.2.1.21)	9	13	*Glycine max*/ AAD09291	ath:AT3G18080	Starch and sucrose metabolism (ko00500)/ Phenylpropanoid biosynthesis (ko00940)

*Enzyme commission number in KEGG pathway

the metabolites belonging to the amino acids (*Table 9*), organic acids (*Table 10*), fatty acids (*Table 11*), plant sterols and vitamins (*Table 12*) and sugars/sugar alcohols (*Table 13*). Most of the predicted unigenes were involved in the biosynthesis and metabolic pathways in KEGG database while very few were involved in the oxidative phosphorylation/TCA/citrate cycle, carbon fixation during photosynthesis and pentose phosphate pathway. As most of the secondary metabolites such as carotenoids, phenylpropanoids and flavonoids could not be detected by our GC-MS analysis, the correlation between these phytochemicals and the identified unigenes could not be performed. However, our mRNA paired-end sequencing results showed that the unigenes involved in these secondary metabolites were present in Eksotika papaya (*Tables 3* and *5*) and we also managed to map these genes in the appropriate pathways (*Figures 2* and *4*). Rivera-Pastrana et al. (2009) reported several phenolic acids from mature fruit flesh of *C. papaya* L. Maradol including caffeic acid, protocatechuic acid and gallic acid hexosides. In addition, carotenoids such as lycopene, β-cryptoxanthin and β-carotene found in the flesh have been identified using LC/DAD-MS-APCI (Rivera-Pastrana et al. 2009). Different phytochemicals observed in our papaya samples may be attributable to the different cultivar, extraction solvents and analytical techniques that have been used in our study (Perez-Gutierrez et al. 2011).

Some of the important phytochemicals identified by GC-MS in this study include vitamin C (ascorbic acid), vitamin E (tocopherols) and plant sterols (*Table 12*). In addition to their primary role in plants, these phytochemicals which are considered as antioxidant nutrients have been subjected to numerous studies focusing on the potential health benefits. Studies have shown that vitamin C-rich foods, such as papaya, provide humans with protection against inflammatory polyarthritis, a form of rheumatoid arthritis involving two or more joints (Pattison et al. 2004). Khaw et al. (2001) also reported that the existence of ascorbic acids in blood plasma was inversely correlated with the mortality of cardiovascular disease (CVD) and ischaemic patients. Vitamins C and A, which are made in the body from beta-carotene, are both needed for the proper functioning of a healthy immune system. Papaya may therefore, be the healthy fruit choice for preventing colds and flu.

Results from several studies have shown that the antioxidant properties of both α- and γ-tocopherols (vitamin E) identified in our study, play important roles in exerting positive effects on human health. Previous studies on animal models and human colon cancer cell lines have shown that vitamin E may help in preventing colon cancer by reducing the mutagen as a result of free radical oxidation activities (Stone et al. 2004).

The benefits of plant sterols for human health have already been established. The United States National Cholesterol Education Programme suggests consumption of 2 g/day of phytosterols as a supplement in order to reduce LDL cholesterol accumulation in the body by 10% (Ostlund 2002). Beta-sitosterol, campesterol and stigmasterol found in this study are the most abundant sterols in plants including papaya. These bioactive compounds are known to have the properties that can reduce the absorption of dietary cholesterol from the gut as well as help in assisting the cholesterol elimination from the body (Law 2000).

Unsaturated fatty acids found in ripening index 5 of Eksotika papaya are known to have beneficial effects on health, specifically related to cardiovascular disease (Lorgeril and Salen, 2004; Pan et al. 2012). For example, alpha-linolenic acid is a polyunsaturated fatty acid (omega 3 fatty acid) while decanoic acid (capric acid) and dodecanoic acid (lauric acid) belong to the group of medium-chain triglycerides (MTC). Other unsaturated fatty acids like

arachidonic and oleic acids were present inconsistently throughout the experiment (replicates). The saturated fatty acids commonly found in plant oils were also found in this study.

Conclusion

Metabolite profiling revealed 29 phytochemicals that can be correlated with the identified enzymes in the primary and secondary metabolite KEGG pathways. These include the metabolites belonging to the group of amino acids, organic acids, fatty acids, plant sterols, vitamins and sugars/sugar alcohols. The results of this study may contribute to the identification of potential chemical biomarkers or candidate genes to genetically improve our local papaya varieties for nutritional properties through marker-assisted breeding or genetic engineering.

Acknowledgements

The authors are grateful to Mr Fazri Abdul Rahim and Ms Noor Faimah Ghazali for their technical support in metabolite profiling, Mr. Lee Weng Wah and Dr. Khairun Hisam Nasir for assisting us in the bioinformatics analysis of the genomics data. This study was supported by the Ministry of Agriculture (MOA) under the Grant No. 05-03-08 SF1051.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucl. Acids Res.* 25: 3389 – 3402

Baybutt, R.C., Hu, L. and Molteni, A. (2000). Vitamin A deficiency injures lung and liver parenchyma and impairs function of rat type II pneumocytes. *J. Nutr.* 130(5): 1159 – 1165

Blas, A.L., Ming, R., Liu, Z.,Veatch, O.J., Paull, R.E., Moore, P.H. and Yu, Q. (2010). Cloning of the papaya chromoplast-specific *lycopene* β*-cyclase, CpCYC-b,* controlling fruit flesh colour reveals conserved microsynteny and a recombination hot spot. *Plant Physiology* 152: 2013 – 2022

Caspari, T., Will, A., Opekarová, M., Sauer, N. and Tanner, W. (1994). Hexose/h+ symporters in lower and higher plants. *J. Exp. Biol.* 196: 483 – 491

Chang, S.J., Puryear, J. and Cairney, J. (1993): A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology* 11(2): 113 – 116

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M. and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21(21): 3674 – 3676

Crozier, A., Jaganath, I.B. and Clifford, M.N. (2006). Phenols, polyphenols and tannins: An overview. In: *Plant secondary metabolites and the human diet*, (Crozier, A., Ashihara, H. and Clifford, M.N., eds.). p. 1–31. Oxford: Blackwell Publishing

Devitt, L.C., Sawbridge, T., Holton, T.A., Mitchelson, K. and Dietzgen, R.G. (2006). Discovery of genes associated with fruit ripening in *Carica papaya* using expressed sequence tags. *Plant Sci.* 170: 356 – 363

Dixon, J. and Hewett, E.W. (2000). Factors affecting apple aroma/flavour volatile concentration: a review. *N.Z. J. Crop Hort. Sci.* 28: 155 – 173

Dobson, G., Shepherd T., Verral, S.R., Conner, S., McNicol, J.W., Ramsay, G., Shepherd, L.V.T., Davies, H.V. and Stewart, D. (2008). Phytochemical diversity in tubers of potato cultivars and landraces using GC-MS metabolomics approach. *J. Agric. Food Chem.* 56: 10280 – 10291

Durrington, I.N., Mackness, B. and Mackness, M.I. (2001). Paraoxonase and atherosclerosis. *Arterioscler. Thromb. Vacs. Biol.* 21: 473 – 480

El Moussaoui, A., Nijs, M., Paul, C., Wintjens, R., Vincentelli, J., Azarkan, M. and Looze,Y. (2001). Revisiting the enzymes stored in the laticifers of *Carica papaya* in the context of their possible participation in the plant defence mechanism. *Cell and Molecular Life Sciences* 58: 556 – 570

Fabi, J.P., Cordenunsi, B.R., Seymour, G.B., Lajolo, F.M. and do Nascimento, J.R. (2009). Molecular cloning and characterization of a ripening-induced polygalacturonase related to papaya fruit softening. *Plant Physiol. Biochem.* 47(11): 1075 – 1081

FAOSTAT (2013). FAO Statistical Databases (United Nations). Retrieved on 10 November 2013 from http://faostat.fao.org/site/291/default.aspx

Fraser, P.D. and Bramley, P.M. (2004). The biosynthesis and nutritional uses of carotenoids. *Prog. Lipid Res*. 43: 228 – 265

Fraser, P.D., Truesdale, M.R., Bird, C.R., Schuch, W. and Bramley, P.M**. (**1994). Carotenoid biosynthesis during tomato fruit-development. *Plant Physiology* 105**:** 405 – 413

Gamage, C.U., Jansz, E.R., Wickramasinghe, N.S.M.D. and Warnasuriya, N.D. (2003). Carotenoids in yellow- and red-fleshed papaya (*Carica papaya* L.). *J. Sci. Food Agric.* 83: 1279 – 1282

Gussman, C.D., Goffredz, J.C. and Gianfagna, T.J. (1993). Ethylene production and fruit-softening rates in several apple fruit ripening variants. *HortScience* 28: 135 – 137

Jian, L., Lee, A.H. and Binns, C.W. (2007). Tea and lycopene protect against prostate cancer. *Asia Pac. J. Clin.Nutr.* 16(Suppl 1): 453 – 457

Khaw, K.T., Bingham. S. and Welch, A. (2001). Relation between plasma ascorbic acid and mortality in men and women in EPIC-Norfolk prospective study: A prospective population study. European prospective investigation into cancer and nutrition. *Lancet* 357: 657 – 63

Lam, P.F. (1994). Pengendalian lepas tuai. In: *Penanaman betik. Insititut Penyelidikan dan Kemajuan Pertanian Malaysia*, (Chan, Y.K., Raveendranathan, P., Raziah, M.L. and Choo, S.T., eds.). *p*. 38 – 42

Law, M.R. (2000). Plant sterol and stanol margarines and health. *West J. Med*. 173: 43 – 47

Lorgeril, D. and Salen, P. (2004). Alpha linolenic acid and coronary heart disease. *Nutr. Metab. Cardiovasc. Dics.*14: 162 – 169

Maheswary, V., Yong, S.H., Nurul Aishah, Y., Sew, Y.S., Khairun, H.N. and Hassan, M.D. (2008). Isolation of transcripts related to floral scent biosynthesis from cempaka putih (*Michelia alba*) flower using substractive hybridization approach. *J. Trop. Agric. Sci*. 31: 107 – 116

Mahdi, H.J., Andayani, R. and Ishak, A. (2010). Metabolic fingerprinting of three Malaysian ginger (*Zingiber officinalis Roscoe*) using gas chromatography-mass spectrometry. *American J. Appl. Sci.* 7: 17 – 23

Ming, R., Hou, S., Feng, Y.,Yu, Q., Dionne-Laporte, A., Saw, J.H., Senin, P., Wang, W., Ly, B.V. and Lewis, K.L. (2008). The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* 452: 991 – 997

Ossipov, V., Ossipova, S., Bykov, V., Oksanen, E., Koricheva, J. and Haukioja, E. (2008). Application of metabolomics to genotype and phenotype discrimination of birch trees grown in a long-term open-field experiment. *Metabolomics* 4: 39 – 51

Ostlund, R.E. (2002). Phytosterols in human nutrition. *Annual Review of Nutrition*, 22: 533 – 549

Palozza, P., Barone, E., Mancuso, C. and Picci, N. (2008). The protective role of carotenoids against 7-keto-cholesterol formation in solution. *Molecular and Cellular Biochemistry* 309: 61– 68

Pan A., Chen, M., Chowdhury , R., Wu, J.H., Sun,Q., Campos, H., Mozaffarian, D. and Hu, F.B. (2012). Linolenic acid and risk of cardiovascular disease: a systematic review and meta-analysis. *Am.J. Clin.Nutr.* 96: 1262 – 1273

Pattison, D.J., Silman, A.J., Goodson, N.J., Lunt, M., Bunn, D., Luben, R., Welch, A., Bingham, S., Khaw, K.T., Day, N. and Symmons, D.P. (2004). Vitamin C and the risk of developing inflammatory polyarthritis: prospective nested case-control study. *Ann. Rheum. Dis.* 63(7): 843 – 847

Paull, R.E., Irikura, B., Wu, P., Turano, H., Chen, N.J., Blas, A., Fellman, J.K., Gschwend, A.R., Wai, C.M., Yu, Q., Presting, G., Alam, M. and Ming, R. (2008). Fruit development, ripening and quality related genes in the papaya genome. *Tropical Plant Biol.* 1: 246 – 277

Perez-Gutierrez, S., Zavala-Sanchez, M.A., Gonzalez-Chavez, M.M., Cardena-Ortega, N.C. and Ramos-Lopez, M.A. (2011). Bioactivity of *Carica papaya* (Caricaceae) against *Spodoptera frugiperda* (Lepidoptera: Noctuidae). *Molecules* 16: 7502 – 7509

Pino, J.A., Almora, K. and Marbot, R. (2003). Volatile components of papaya (*Carica papaya* L., Maradol variety) fruit. *Flavour Frag. J.* 18: 492 – 496

Raes, J., Rohde, A., Christensen, J.H., Van de Peer, Y. and Boerjan, W. (2003). Genome-wide characterization of the lignification toolbox in *Arabidopsis*. *Plant Physiol*. 133: 1051 – 1071

Rivera-Pastrana, D.M., Yahia, E.M. and Gonzalez-Aguilar, G.A (2009). Identification of phenolic and carotenoid compounds in *Carica papaya* L. maradol using liquid chromatography-mass spectrometry. *Acta Horticulture* 877: 34 – 35

Rodriguez-Amaya, D.B. (2003). Enhancing the carotenoid levels of food through agriculture and food technology. Food Africa. Retrieved on 4th March 2010 from http://foodafrica.nri.org/nutrition/internetpapers/DeliaBRodriguez.pdf

Roessner, U., Wagner, C., Kopka , J., Trethewey, R.N. and Willmitzer, L. (2000). Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. *Plant J*. 23: 131 – 142

Rogers, L.A. and Campbell, M.M. (2004). The genetic control of lignin deposition during plant growth and development. *New Phytologist* 162: 17 – 30

Schaffer, R.J., Friel, E.N., Souleyre, E.J.F., Bolitho, K., Thodey, K., Ledger, S., Bowen, J.H., Ma, J.H., Nain, B., Cohen, D., Gleave, A.P., Crowhurst, R.N., Janssen, B.J., Yao, J.L. and Newcomb, R.D. (2007). A genomics approach reveals that aroma production in apple is controlled by ethylene predominantly at the final step in each biosynthetic pathway. *Plant Physiology* 144(4): 1899 – 1912

Schmid, K.M. (2004). Lipid metabolism. In: *Handbook of Plant Biotechnology, Vol. 2*, (Christou, P. and Klee, H., eds.). p. 472 – 524. John Wiley and Sons Ltd.

Sew, Y.S., Johari, S., Maheswary, V. and Umi Kalsom, A.B. (2011). Isolation of fruit ripening genes from *Carica papaya* var. Eksotika cDNA libraries. *J. Trop. Agric. and Fd. Sc*. 39(2): 203 – 211

Stone, W.L., Krishnan. K., Campbell, S.E., Qui. M.,Whaley. S.G. and Yang, H. (2004). Tocopherols and the treatment of colon cancer. *Annal. N. Y. Acad. Science* 1031: 223 – 233

Yamamoto, H.Y. (1964). Comparison of the carotenoids in yellow and red-fleshed *Carica papaya*. *Nature* 201: 1049 – 1050

Yu, Q., Tong, E., Skelton, R.L., Bowers, J.E., Jones, M.R., Murray, J.E., Hou, S., Guan, P. Acob, R.A., Luo, M.C., Moore, P.H., Alam, M., Paterson, A.H. and Ming, R. (2009). A physical map of the papaya genome with integrated genetic map and genome sequence. BMC genomics 10: 371 doi:10.1186/1471-2164-10-371

Zhang, X.Y., Wang, X.L., Wang, X.F., Xia, G.H., Pan, Q.H., Fan, R.C., Wu, F.Q., Yu, X.C. and Zhang, D.P. (2006). A shift of phloem unloading from symplasmic to apoplasmic pathway is involved in developmental onset of ripening in grape berry. *Plant Physiol*. 142: 220 – 32

Zhou, L., Chen, C.C., Ming, R., Christopher, D.A. and Paull, R.P. (2003). Apoplastic invertase and its enhanced expression and post-translational control during papaya fruit maturation and ripening. *J. Amer. Soc. Hort. Sci*. 128: 628 – 635

Abstrak

Bahan fitokimia yang terlibat dalam laluan metabolit primer dan sekunder dan gen-gen yang mempunyai pertalian dengannya dalam indeks pemasakan 5 buah betik Eksotika telah dikaji dengan menggunakan pendekatan metabolomik berasaskan GC-MS dan penjujukan mRNA secara pasangan hujung (paired-end). Sejumlah 54 metabolit telah berjaya dikenal pasti untuk sebatian polar dan tidak polar yang terdiri daripada asid amino dan asid organik, asid lemak tepu/tak tepu, gula dan gula beralkohol, fenol ringkas, vitamin dan sterol serta alkana. Asid askorbik (vitamin C), α-tokoferol (vitamin E), dan sterol tumbuhan termasuk stigmasterol, β-sitosterol dan kampesterol juga telah berjaya dikenal pasti dalam kajian ini. Sejumlah 48,765 unigen anggapan telah dikenal pasti daripada penjujukan mRNA secara pasangan hujung di mana 59% daripadanya (28,736) menunjukkan ekspresi yang ketara ((E-value $\leq 10^{-5}$) terhadap enzim di dalam pangkalan data tidak lewah (nr). Daripada 28,736 unigen anggapan tersebut, hanya 4.9% (1,413) telah dikenal pasti berpotensi untuk terlibat dengan biosintesis dan metabolisme metabolit sekunder berdasarkan laluan KEGG. Ini termasuk metabolit karotenoid, mono and diterpenoid, stillbenoid, flavonoid, flavon, flavonol, antosianin, fenilpropanid, betalain, isothiosianat, steroid, asid lemak tepu dan tidak tepu, gula, gula beralkohol, brassinosteroid, asid amino dan organik, asid absisik, latisifer dan protease. Analisis Ontologi Gen (GO) pula mengenalpasti 67 unigen dalam kategori fungsi molekul terlibat dalam aktiviti antioksidan. Pemetaan pada laluan KEGG mengenal pasti 29 metabolit mempunyai pertalian rapat dengan unigen anggapan yang telah diramal. Ini termasuk metabolit dari kumpulan asid amino/organik, gula, asid lemak dan sterol tumbuhan.

Accepted for publication on 9 November 2013