Isolation and functional characterisation of banana *phytoene synthase* genes as potential cisgenes

Bulukani Mlalazi [1*], Ralf Welsch [2], Priver Namanya [1¶], Harjeet Khanna [1], R. Jason Geijskes [1], Mark D. Harrison [1], Rob Harding [1], James L. Dale [1], Marion Bateson [1]

[1] *Centre for Tropical Crops and Biocommodities, Queensland University of Technology, Brisbane, Queensland 4001, Australia;* [2] *Faculty of Biology, University of Freiburg, D79104 Freiburg, Germany*

[¶] Present address: National Agricultural Research Laboratories-Kawanda, P.O. Box 7065, Kampala, Uganda

* To whom correspondence should be addressed: Bulukani Mlalazi, Centre for Tropical Crops and Biocommodities, Queensland University of Technology, 2 George Street, Brisbane, Queensland 4001, Australia Tel.: 61-7-3138-1619; Fax: 61-7-3138-4132; E-mail: b.mlalazi@qut.edu.au

Abstract

Carotenoids occur in all photosynthetic organisms where they protect photosystems from auto-oxidation, participate in photosynthetic energy-transfer and are secondary metabolites. Of the more than 600 known plant carotenoids, few can be converted into vitamin A by humans and so these pro-vitamin A carotenoids (pVAC) are important in human nutrition. Phytoene synthase (PSY) is a key enzyme in the biosynthetic pathway of pVACs and plays a central role in regulating pVAC accumulation in the edible portion of crop plants. Bananas are a major commercial crop and serve as a staple crop for more than 30 million people. There is natural variation in fruit pVAC content across different banana cultivars, but this is not well understood. Therefore, we isolated *PSY* genes from banana cultivars with relatively high (cv. Asupina) and low (cv. Cavendish) pVAC content. We provide evidence that PSY in banana is encoded by two paralogs (*PSY1* and *PSY2*), each with a similar gene structure to homologous genes in other monocots. Further, we demonstrate that *PSY2* is more highly expressed in fruit pulp compared to leaf. Functional analysis of *PSY1* and *PSY2* in rice callus and *E. coli* demonstrate that both genes encode functional enzymes, and that Asupina PSYs have approximately twice the enzymatic activity of the corresponding Cavendish PSYs. These results suggest that differences in PSY enzyme activity contribute significantly to the differences in Asupina and Cavendish fruit pVAC content. Importantly, Asupina *PSY* genes could potentially be used to generate new cisgenic or intragenic banana cultivars with enhanced pVAC content.

Keywords: Biofortification, Carotene, Cisgenic, Intragenic, Isoprenoid, Metabolic engineering

Abbreviations:

PSY	Phytoene synthase
pVAC	Pro-vitamin A carotenoid
VAD	Vitamin A deficiency

Introduction

Carotenoids are tetraterpenoid lipophilic pigments that occur naturally in all photosynthetic organisms (for review see Taylor and Ramsay 2005). They include the carotenes (e.g. α-carotene and β-carotene) and their oxygenated derivatives, the xanthophylls (e.g. zeaxanthin and violaxanthin). The first committed reaction of carotenoid biosynthesis is the condensation of two geranylgeranyl diphosphate molecules by phytoene synthase (PSY) to produce phytoene, which is then converted to lycopene via a series of desaturation steps, prior to the formation of ionone end groups through the action of lycopene cyclases to form cyclic carotenes such as β-carotene (Hirschberg 2001). In plants, carotenoids are primarily localised in the chloroplasts and chromoplasts (Hirschberg 2001). These compounds are essential for plant growth as they shield the photosynthetic apparatus from damage by light-induced stress (Demmig-Adams and Adams 2002). Pro-vitamin A carotenoids (pVAC) are the main dietary source of vitamin A for humans (West 2000) and are enzymatically cleaved into retinol, which is subsequently converted to other vitamin A compounds (West 2000). The essential naturally occurring pVAC are α-carotene, β-carotene, γ-carotene and β-cryptoxanthin, which possess an un-substituted β-ionone ring at one or both ends of the molecule (Bender 2003). β-carotene has β-ionone rings at both ends and is therefore most efficiently converted into retinol (West 2000).

Bananas are herbaceous, perennial, non-graminaceous monocots that are vegetatively propagated. Globally, bananas are a major crop and are the staple food for more than 30 million people living in Sub-Saharan Africa (IITA 2009). They are also a vital cash crop for subsistence farmers worldwide (IITA 2009). Four banana genomes (A, B, S, and T) have been identified (Pillay et al. 2004). A and B genomes are present in most cultivated bananas, including Cavendish

(AAA) and Ladyfinger (AAB). The genotypes of a small number of Polynesian Fe'i bananas, which contain the T-genome, have been characterised (Daniells et al. 2001). However, Asupina, a Fe'i banana from Papua New Guinea is believed to possess an ATT genome (Sharrock and Frison 1998). Bananas containing the S genome include *Musa schizocarpa* (SS) and interspecific hybrids such as Tonton Kepa (AS) and Ugota (AS) (Hribova et al. 2011). Unlike other crops, such as rice, there is considerable natural variation in the fruit pVAC content from different banana cultivars and fruit pVAC content is the major contributor to the colour of the fruit pulp, ranging from white to dark orange (Englberger et al. 2003). For example, banana cultivars with creamy-coloured fruit pulp have low β-carotene levels (e.g. Cavendish [AAA], 0.72 μg g^{-1} fresh weight (FW), Ladyfinger [AAB], 1.12 μg g^{-1} FW; Englberger et al. 2006), while cultivars with yellow-coloured fruit pulp contain higher levels of β-carotene (e.g. Lakatan [AA], 2.89 μg g^{-1} FW, Wain [Fe'i genome], 5.01 μg g^{-1} FW; Englberger et al. 2006). However, the highest β-carotene levels have been measured in bananas with orange fruit pulp, such as Asupina (Fe'i genome, 14.5 μg g^{-1} FW; Englberger et al. 2006) and Uht en yap (Fe'i genome, 27.8 μg g^{-1} FW; Englberger et al. 2003). There is no obvious correlation between genotype and fruit pulp pVAC content, and the molecular basis for the wide range of fruit pulp pVAC content is unknown.

Vitamin A deficiency (VAD) affects over half of the world's population and is particularly prevalent in developing countries, where it is the principal cause of preventable childhood blindness (West 2002). VAD is also a major contributor to morbidity and mortality in children and pregnant women (West 2002). Treatment of VAD through pharmaceutical supplementation and food fortification has achieved limited success, and alternative approaches are required (Mayer et al. 2008).

Biofortification is a strategy for broad-based treatment and prevention of micronutrient deficiency in large populations, and involves increasing the micronutrient content in the edible portion of staple food crops. Biofortification can be achieved through conventional breeding in crops such as tomato (Lecomte et al. 2004) and wheat (Perez-Jones et al. 2006), where desirable agronomic traits are introduced into cultivated species from related species through introgression. Although there are banana cultivars with elevated pVAC content (Englberger et al. 2003), improvement using conventional breeding is a major challenge because the majority of banana cultivars are sterile triploids. In contrast, biofortification of pVAC content in bananas using genetic modification is a promising approach that overcomes this limitation. Similarly, because of the absence of rice germplasm accessions with yellow endosperm, genetic engineering for rice pVAC biofortification was required, resulting in 'Golden Rice' (Ye et al. 2000). This cultivar accumulates elevated levels of pVAC in its endosperm and this biofortification was achieved through expression of key enzymes in the carotenoid biosynthetic pathway: PSY and a carotene desaturase (Ye et al. 2000). Golden Rice has the potential to alleviate VAD in rice-growing regions but is not an appropriate strategy to combat VAD in every country, particularly in many parts of Africa where banana is the staple crop for a large proportion of the population (IITA 2009). Most staple banana cultivars have low pVAC content, making banana an important target for pVAC biofortification with the most appropriate strategy being the genetic modification of already accepted cultivars. The creation of genetically modified crops using transgenic material has been met with environmental (Garcia and Altieri 2005) and public concerns, and high regulatory complexity (Rommens 2010). However, cisgenics and intragenics, which involve the genetic modification of a plant species with sequences originating from its

own genome or those of sexually compatible species (Molesini et al. 2012), may increase the acceptability of genetically modified crops. Such approaches avoid the use of transgenes, creating plant species with genetic properties within the range of those plants produced by conventional breeding (Molesini et al. 2012), which may expedite regulatory approval (Schouten et al. 2006; Rommens et al. 2007) and improve public acceptance (Molesini et al. 2012). Cisgenics requires the insertion of an entire natural gene, which includes the coding region flanked by its native promoter and terminator in a sense orientation (Schouten et al. 2006). However, the genes introduced into intragenic plants may be novel combinations of coding regions and regulatory elements derived from one or more species (Rommens et al. 2007). Cisgenic and intragenic approaches have been used to create disease resistant apple (Vanblaere et al. 2011) and potato (Cavatorta et al. 2011) cultivars, respectively. For this approach to be successful, it is necessary to identify suitable banana genes and promoters that can be used to generate cisgenic or intragenic cultivars.

The carotenoid biosynthetic pathway is well characterised in a wide range of plants (for review see: Hirschberg 2001; Cazzonelli 2011). Over-expression of *PSY* in tomato fruit (Fray et al. 1995), rice endosperm (Ye et al. 2000; Paine et al. 2005) and white carrot roots (Maass et al. 2009) demonstrates that PSY catalyses a key rate-limiting step in carotenoid biosynthesis. Higher plants possess a range of *PSY* genes; for example, there is a single *PSY* gene in *Arabidopsis* (Scolnik and Bartley 1994) and daffodil (Schledz et al. 1996), two *PSY* genes in tomato (Bartley and Scolnik 1993) and tobacco (Busch et al. 2002), and three in many of the Poaceae, including *Zea mays* and rice (Li et al. 2008). In tomato, both *PSY* genes (*PSY1* and *PSY2*) are expressed in leaves and fruit (Bartley and Scolnik 1993), although expression of *PSY1* is predominantly associated with

chromoplast-bearing tissues (Giorio et al. 2008). Both *PSY1* and *PSY2* are expressed in yellow endosperm maize but *PSY1* expression is absent in white endosperm maize mutants (Gallagher et al. 2004). Further, maize quantitative trait loci have been identified that strongly link *PSY1* with carotenoid accumulation (Chander et al. 2008). Characterisation of tomato PSY1 (Gady et al. 2012) and cassava PSY2 (Welsch et al. 2010) has demonstrated that single amino acid changes in PSY can significantly impact pVAC biosynthesis. In addition, alternative splicing has been identified as the cause of amino acid insertions and deletions in PSY1 from wheat (Howitt et al. 2009) and barley (Rodriguez-Suarez et al. 2011), respectively, resulting in reduced endosperm pVAC content.

We are working towards the improvement of pVAC accumulation in the fruit of agronomically important banana cultivars. Therefore, given the importance of PSY in carotenoid biosynthesis in a range of plants, we investigated whether PSY was a contributing factor to the wide variation in pVAC content in banana cultivars. Here, we describe the isolation and characterisation of *PSY* genes from Cavendish (white pulp) and Asupina (orange pulp) bananas and present a preliminary molecular profile of the *PSY* gene family in bananas as a basis for identifying banana genes that could be utilised to enhance pVAC content via genetic modification.

Materials and methods

Plant material

Leaf tissue was obtained from tissue cultured banana plants. Cavendish and Ladyfinger banana plants were obtained from Jennifer Kleidon (Centre for Tropical Crops and Biocommodities, Queensland University of Technology). Leaf tissue from Asupina, Lakatan and Wain banana plants was obtained from Sharon Hamill (Queensland State Government Department of Employment, Economic Development and Innovation (DEEDI), Maroochy Research Station). Green and ripe Cavendish fruit were purchased from local markets. Green Asupina fruit (provided by Jeff Daniells, DEEDI, South Johnstone Research Station) was either sampled immediately or after it had ripened naturally (i.e. without any ethylene treatment). All plant material was stored at –80 °C prior to use.

Isolation of nucleic acids

Total RNA was isolated from 10 g of either banana leaf or fruit pulp using the CTAB extraction method described by Chang et al. (1993) and modified to include the RNA extraction buffer and polysaccharide removal procedure described by Asif et al. (2000). Total mRNA was purified from 100 to 500 μg of total RNA using the GenElute ^{TM} mRNA Miniprep Kit (Sigma-Aldrich).

Genomic DNA (gDNA) was isolated from 1 g of leaf tissue using a protocol adapted from Stewart and Via (1993). Briefly, tissue was ground in liquid nitrogen, extracted in 3 mL pre-warmed DNA extraction buffer, and

incubated for 60 min at 65 °C. The aqueous phase was extracted twice with an equal volume of chloroform and then treated with 10 µg RNase A (Sigma-Aldrich) at 37 °C for 60 min. The aqueous phase was re-extracted with chloroform before precipitation of the gDNA with an equal volume of isopropanol at –80 °C for 30 min. For gDNA of higher purity, leaf tissue (0.1 g) was extracted using a DNeasy ™ Plant Mini Kit (Qiagen) according to the manufacturer's instructions.

Cloning and analysis of banana *PSY* sequences

Conserved PSY amino acid motifs, WAIYVW and IEANDYNNF, were identified from a multiple sequence alignment of available plant PSY proteins (Table S1) and used to design degenerate primers (Primers P1 and P2, Table S2). These primers were used to amplify a 680 base pair (bp) fragment of the *PSY* coding region from total banana leaf mRNA (Titan ™ One Tube RT-PCR System, Roche Applied Science). RT-PCR products were ligated into pGEM-T® Easy (Promega) and sequenced (BigDye® Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems). Putative cDNA clones were analysed using the SeqMan (Lasergene8) DNA analysis program to create consensus sequences by clustering individual sequences into groups on the basis of 80 – 95% minimum identity. BLAST (Altschul et al. 1990) tools for both nucleotide and protein comparison at NCBI (www.ncbi.nlm.nih.gov/) were used to confirm the identity of the consensus sequences as *PSY*. For phylogenetic analysis, banana *PSY* consensus sequences generated by SeqMan (Lasergene8) were cropped to 600 bp (encoding 200 amino acids) and compared to the corresponding nucleotide sequences of selected *PSY* GenBank accessions (Table S3). Bootstrapped phylogenetic trees

and protein sequence alignments were produced using Clustal X2 v2.0.6 (Larkin et al. 2007), and Clustal W2 v2.0.1.0 (Larkin et al. 2007).

The 5' and 3' ends of selected *PSY* transcripts were isolated from leaf and fruit mRNA (FirstChoice® RLM-RACE kit, Ambion), with SuperScript® III (Invitrogen) replacing the supplied reverse transcriptase. Primer pairs for outer and nested PCR (Tables S2 and S4) were designed to ensure significant overlap (167 – 218 bp) between new and existing partial sequences. Where sequences could not be obtained using 5' RLM-RACE, genome-walking libraries were constructed from Cavendish gDNA (0.3 to 1 µg) digested with *Eco*RI, *Eco*RV, *Nae*I, *Pvu*II, *Stu*I, *Sty*I or *Swa*I (GenomeWalker ™ Universal Kit, Clontech Laboratories Inc.). Genomic fragments were obtained following three sequential nested PCR reactions (Expand ™ Long Template PCR system, Roche Applied Science) using the primers described in Tables S2 and S4.

Cloning of the full-length *PSY* coding regions from banana mRNA and gDNA

Gene-specific primers (Tables S2 and S4) were used to amplify full-length PSY coding regions from banana leaf and fruit mRNA (Titan ™ One Tube RT-PCR, Roche Applied Science) or gDNA using either GoTaq Green® PCR (Promega) or Expand ™ Long Template PCR (Roche Applied Science). PCR products were cloned into pGEM-T® Easy (Promega) and sequenced. Exons and introns were identified by alignment with the respective mRNA coding regions using Spidey (Wheelan et al. 2001).

Detection of *PSY* transcripts in banana leaves and fruit pulp

Total mRNA from Asupina and Cavendish leaf tissue and fruit pulp was treated with Turbo DNAfree ™ (Ambion) to remove contaminating gDNA, and quantified (Quant-it ™ RiboGreen RNA assay kit, Invitrogen). Primers complimentary to *Musa acuminata actin1* mRNA (GenBank Accession No. AF246288) (P28/P29, Tables S2 and S4) were used in PCR with GoTaq Green® (Promega) to detect gDNA contamination in mRNA templates and as the RT-PCR control. RT-PCR (Titan ™ One Tube RT-PCR system, Roche Applied Science) was used to amplify representative PCR fragments from 3 ng mRNA with primer pairs P30/P31 (Asupina and Cavendish *PSY1*), P32/P34 (Asupina *PSY2a*), and P33/P34 (Cavendish *PSY2a*) (Tables S2 and S4). PCR products were resolved by electrophoresis on 2% (w/v) TAE agarose gels and visualised using SYBR® safe DNA gel stain (Invitrogen).

Construction of *PSY* plant expression vectors

Asupina and Cavendish *PSY1* and *PSY2a* coding regions were excised from pGEM-T® Easy and sub-cloned into the pGEM®-4Z vector between a maize polyubiquitin1 (*ubi1*) promoter (Christensen and Quail 1996) and *Agrobacterium tumefaciens* nopaline synthase (*nos*) terminator (Depicker et al. 1982) to generate *pGEM-4Z-ubi1-PSY-nos* vectors. The PSY expression cassettes (*ubi1-PSY-nos*) were then excised from the pGEM®-4Z backbone and sub-cloned into the pCAMBIA-1302 binary vector (GenBank Accession No. AF234298.1), to create four *pCAMBIA-1302-ubi-PSY-nos* vectors for expression of Asupina or Cavendish *PSY1* or *PSY2a* coding regions. The maize *PSY1* expression cassette (*ubi1-Zm-*

PSY1-nos) was excised from *pBIN-ubi1-Zm-PSY1-nos* and sub-cloned into *pCAMBIA-1302*. DNA from all five *pCAMBIA-1302-ubi1-PSY-nos* vectors were purified from *E. coli* (UltraClean ™ 6 Minute Mini Plasmid Prep Kit ™, Mo Bio Laboratories, Inc.) and sequenced prior to transformation into *Agrobacterium*.

Agrobacterium-mediated transformation of rice callus

Dehusked *Oryza sativa* L. spp. *japonica* cv. Nipponbare seeds (Russell Reinke, Yanco Agricultural Institute, New South Wales Department of Primary Industries) were surface-sterilised with 70% (v/v) ethanol for one minute followed by vigorous shaking for 30 minutes in 80 % commercial bleach (4% (w/v) hypochlorite) containing 0.1% (v/v) Tween® 20. Seeds were rinsed 4 times in sterile distilled water, blotted on sterile Whatman ™ filter paper and cultured on solidified 2N6 callus induction media (Chu et al. 1975) at 27 °C for 3 weeks in the dark. Embryogenic callus was multiplied on fresh 2N6 solid media for a further 2 weeks and then again for four days before transformation.

The *pCAMBIA-1302* vector control and five *pCAMBIA-1302-ubi1-PSY-nos* binary vectors were electroporated into *A. tumefaciens* (AGL1 strain) (Dower et al. 1988). Transformed colonies were selected on solid YM media (Malik 1992) supplemented with 25 mg L^{-1} rifampicin and 50 mg L^{-1} kanamycin. Transformed *Agrobacterium* colonies were cultured in LB media supplemented with 50 mg L^{-1} kanamycin at 28 °C for 72 h and used immediately for transformation of rice callus.

Rice callus was transformed with *Agrobacterium* using centrifugation assisted *Agrobacterium*-mediated transformation (CAAT) (Khanna et al. 2004) and transformed calli were selected on 2N6 media containing 50 mg L^{-1}

hygromycin and 200 mg L^{-1} Timentin® at 27 °C in the dark. Controls included untransformed rice callus cultured on the same media with or without hygromycin. In the experiment, each vector was transformed in four replicates consisting of 30 calli of approximately 3 mm in diameter.

Carotenoid extraction and quantification

Hygromycin resistant rice calli proliferating on selection media were collected two, four and six weeks post-transformation and photographed using a SZ-CTV stereo microscope (Olympus) with CAMEDIA C-5060 camera (Olympus). Callus samples (300 to 400 mg) were freeze dried overnight (Flexi-Dry MP freeze drying system, FTS Systems), weighed, and ground to a powder (Mini-Beadbeater-8, BioSpec Products). Powdered tissue was mixed thoroughly with 2 mL acetone (HPLC grade), clarified by centrifugation at 4,000 *g* for 5 min under low light to prevent carotenoid degradation, and the supernatant collected. The pellet was extracted twice more with 2 mL acetone (HPLC grade) and the pooled acetone supernatants mixed with 2 mL petroleum ether (40 – 60 °C fraction): diethyl ether (PE:DE) (2:1; v/v) before addition of 1% (w/v) NaCl to a final volume of 14 mL. The upper organic phase was collected after centrifugation at 4000 *g* for 5 min. Extraction with 1 ml PE:DE (2:1 v/v) was repeated, the PE:DE (2:1 v/v) fractions were pooled and dried under vacuum and then re-suspended in 2 mL chloroform. Total carotenoid content (μg g^{-1} dry weight) was determined by absorption spectrophotometry at 465 nm against a β-carotene (Sigma-Aldrich) standard curve.

In-situ detection of β-carotene in rice callus by Raman spectroscopy

Raman spectra were obtained using a Spectrum 2000 Near Infrared Fourier Transform Raman spectrometer (Perkin-Elmer) fitted with a Nd:YAG laser emitting at 1064 nm with a power output of 320 mW. A reference spectrum for β-carotene was generated using 1 mg L^{-1} β-carotene standard (Sigma-Aldrich) dissolved in carbon tetrachloride (CCl_4). Representative proliferating callus clumps were selected from single transgenic callus lines constitutively expressing Asupina *PSY2a* (*A-PSY2a*) or *Zm-PSY1*. Callus expressing a bacterial desaturase (*CrtI*) enzyme from *P. ananatis* and untransformed wild type callus were used as controls. Spectra were collected using a single laser beam by performing 64 scans with a spectral resolution of 8 cm^{-1} in the Raman shift range 200 – 3,800 cm^{-1}. Peaks arising from the CCl_4 solvent were observed in the 200 – 800 cm^{-1} range.

Heterologous expression of banana PSY enzymes in *E. coli*

Banana PSY1 and PSY2a enzymes were co-expressed with geranylgeranyl diphosphate synthase from *Sinapis alba* (GGPS) and *CrtI* from *P. ananatis* in *E. coli* BL21-CodonPlus-RIL using the DUET vector suite (Novagen-Merck). *pET-GGPS* was obtained by sub-cloning a BamHI/SalI fragment from *pGGPS* (Welsch et al. 2000) encoding the mature GGPS protein into *pETDuet-1*. To generate *pCDF-CrtI*, the *CrtI* coding region was amplified from *pBaal2* (Ye et al. 2000) with primers P35 and P36 (Table S2), the PCR fragment was digested with *Eco*RI/*Hind*III, and the resulting DNA fragment was sub-cloned into *pCDFDuet-1*. The *pRSF-PSY* vectors containing the mature banana PSY coding regions were generated by amplifying the coding regions from their corresponding *pCAMBIA-*

1302 vectors using primer pairs P37/P38 (*Ma-PSY1*), P37/P39 (*A-PSY1*), P40/P41 (*A-PSY2a*), and P42/P43 (*Ma-PSY2a*) (Tables S2 and S4). PCR fragments were digested with *Eco*RI/*Hind*III and ligated into vector *pRSF*.

Overnight cultures of *E. coli* co-transformed with *pET-GGPS*, *pCDF-CrtI* and one of the *pRSF-PSY* vectors were sub-cultured in 50 ml fresh culture medium and grown to $OD_{600\ nm}$ = 0.5 at 37 °C. Expression was induced with 0.5 mM isopropyl-β-D-thio-galactoside and cultures were incubated at 28 °C. The $OD_{600\ nm}$ was recorded three and four hours after induction and 4.5 ml culture spun down and extracted twice with chloroform/methanol (2:1, v/v). Lycopene concentration was determined spectrophotometrically in petroleum benzene ($\varepsilon_{470\ nm}$ = 185,230 L mol^{-1} cm^{-1}). Western analysis was carried out on total protein extracts from the co-transformed *E. coli* cultures to confirm the expression of similar levels of GGPS and PSY proteins. *E. coli* cells in phosphate buffered saline were disrupted with a French Press (Amicon). After centrifugation (3,100 x g) the protein concentration was determined in the supernatant. Proteins were separated by SDS-PAGE, blotted onto PVDF membranes (Carl Roth, Karlsruhe, Germany) and His6-GGPS and His6-PSY were detected with an anti-6*His-Tag antibody (Sigma-Aldrich).

Results

Identification and analysis of partial *PSY* sequences in banana

Messenger RNA from Asupina, Cavendish, Ladyfinger, Lakatan and Wain banana leaves, and from Cavendish and Asupina fruit, were used as templates to amplify fragments of *PSY* coding sequences using degenerate primers (P1 and P2, Table S2). PCR products of the expected size of 680 bp were obtained from each of the five cultivars and sequenced. A total of 6, 7, 9, 8 and 11 clones were isolated from Asupina, Cavendish, Ladyfinger, Lakatan and Wain leaf tissues, respectively, while 13 and 10 clones were isolated from ripe fruit of Asupina and Cavendish, respectively. When compared to monocot *PSY* mRNAs, the sequence of these fragments segregated into two groups each with greater than 76% nucleotide sequence identity to either *PSY1* or *PSY2* sequences (results not shown), confirming the presence of two *PSY* paralogs in banana. Interestingly, all of the fruit-derived sequences showed greatest similarity to *PSY2*. Further analysis of the *PSY2* nucleotide sequences from Cavendish revealed evidence of two subgroups, designated *PSY2a* and *PSY2b*, which shared 85% nucleotide sequence identity (results not shown). This was supported by alignment of the banana PSY partial protein sequences (Fig. S1). Cavendish PSY2b was 98% identical to Ladyfinger PSY2, but only 90% identical to other banana PSY2 partial protein sequences including Cavendish PSY2a (results not shown). Comparatively, all banana PSY1 partial protein sequences from each cultivar (Fig. S1) were 83% identical to their respective PSY2 sequence (results not shown). Of the 13 Cavendish *PSY2* sequences isolated, 9 formed the *PSY2a* group with the remaining Cavendish sequences grouping as *PSY2b*. However, further characterisation is required to

confirm the presence of *PSY2b* in banana as initial cloning found this gene only in Cavendish. Consequently, subsequent analyses focused on *PSY1* and *PSY2a* genes.

The banana *PSY* partial sequences were compared to *PSY* sequences from a broad range of plants (Table S3, Fig. 1). With the exception of carrot *PSY1*, all dicot and monocot *PSY* sequences formed distinct clades, and the banana *PSY* sequences clustered with other monocots. As expected, banana *PSY1* and *PSY2* sequences clustered separately. Within these banana subclades, *PSY* sequences from Asupina and Wain cultivars (containing Fe'i genomes) clustered together, as did those from Cavendish, Ladyfinger and Lakatan, which all possess A genomes.

Isolation of full-length Asupina and Cavendish *PSY1* and *PSY2a* transcripts

To further characterise and compare *PSY* sequences from banana cultivars with relatively low (<1 μg g^{-1} FW) and relatively high (>10 μg g^{-1} FW) fruit pVAC content, the complete mRNA sequences of *PSY1* and *PSY2a* from Asupina and Cavendish leaves and fruit, respectively, were obtained. Full-length consensus sequences of Asupina *PSY1* (*A-PSY1*), Cavendish (*Musa acuminata* var Cavendish) *PSY1* (*Ma-PSY1*) and *A-PSY2a* were generated from overlapping 5' and 3' RLM-RACE clones, and partial coding sequences. Interestingly, 5' and 3' RLM-RACE clones of different sizes were obtained for all three banana *PSY* sequences. As 5' RLM-RACE can only amplify mRNAs that have intact 7-methylguanosine caps at their 5' termini, this suggested the presence of multiple transcription start sites, while different clones from 3' RLM-RACE indicated the presence of multiple polyadenylation signals (Fig. S2a). The *Ma-PSY1*, *A-PSY1* and *A-PSY2a* transcripts were predicted to be up to 1,854, 1,890 and 1,551

nucleotides (nt) in length, respectively, due to the variation in both transcription start sites and polyadenylation signals (Fig. S2a). *A-PSY1* and *Ma-PSY1* transcripts both included open reading frames of 1,290 nt, while that of *A-PSY2a* transcripts was 1,182 nt.

No full-length *Ma-PSY2a* sequences were obtained from Cavendish ripe fruit RNA using 5' RLM-RACE despite extensive modification of cDNA synthesis and PCR amplification conditions (results not shown). One full-length clone was obtained using 3'RLM-RACE and included 390 bp encoding the C terminal 130 amino acids of the Cavendish PSY2a (Ma-PSY2a) enzyme and 110 bp that was 97% identical to the 3' UTR of the longest *A-PSY2a* clone. Interestingly, this clone also contained a duplication corresponding to 162 bp of the 3' end of the *Ma-PSY2a* coding region and 9 bp of 3' UTR (Fig. S2a). To obtain the sequence of the putative *Ma-PSY2a* mRNA, a genome walking strategy was used to identify the corresponding DNA sequence from Cavendish gDNA. Following three rounds of nested PCR, a consensus sequence was generated that included 246 bp with 92% identity to the *A-PSY2a* 5' UTR and the first three exons/introns of the *Ma-PSY2a* coding region (results not shown).

Comparison of Asupina and Cavendish *PSY1* and *PSY2a* coding region structure

The genomic sequences of the Asupina and Cavendish *PSY1* and *PSY2a* genes were compared to determine if differences in gene structure or coding sequences could account for the differences in fruit pVAC content between these banana cultivars. Genomic regions were amplified from Asupina and Cavendish gDNA and the predicted intron and exon organisation compared to equivalent genes from maize. Like maize, the organisation of banana *PSY1* and *PSY2* genes was highly

conserved, consisting of six exons separated by five introns (Fig. S2b). The sizes of exons 2, 3, 4 and 5 (which encode the PSY catalytic domain) were highly conserved in *PSY1* and *PSY2* in both banana and maize. In contrast, exon 1, which encodes the chloroplast transit peptide (TP), and exon 6, which encodes the C terminus of the enzyme, varied significantly in length (Fig. S2b).

The amino acid sequences of A-PSY1 and Ma-PSY1 (Fig. 2a) differed by 5 and 10 amino acids in the putative TP and mature protein, respectively. A-PSY2a and Ma-PSY2a (Fig. 2b) differed by 7 and 9 amino acids in the putative TP and mature protein, respectively. In addition, there was a three amino acid insert (AVE) in the N terminus of the mature protein of Ma-PSY2a. Protein motifs that have previously been associated with PSY activity according to the Conserved Domain Database (Marchler-Bauer et al. 2011), including the active site lid, Aspartate-Rich Motif (ARM), SQS-PSY1 and SQS-PSY2 motifs, were present in all PSY amino acid sequences and were identical between PSY1 and PSY2a (Fig. 2). There were no amino acid differences between the Cavendish and Asupina PSY proteins in these conserved motifs, suggesting that the PSYs from both cultivars would be functional.

Analysis of tissue specificity of *PSY1* and *PSY2a* in Asupina and Cavendish

Differential *PSY* expression has been linked to variable pVAC levels in carrot (Clotault et al. 2008), tomato (Giorio et al. 2008) and in maize mutants (Gallagher et al. 2004). The absence of *PSY1* partial sequence mRNA in banana fruit suggested that differential expression of *PSY* genes may occur in banana tissues. To provide further insight into *PSY* gene expression in banana, RT-PCR was used to compare the relative abundance of *PSY1* and *PSY2a* mRNA in leaf, green fruit

and ripe fruit pulp of Asupina and Cavendish cultivars. The absence of gDNA contamination was confirmed using standard PCR without the inclusion of a reverse transcription step (results not shown) and control (*ACT1*) amplicons of the expected size were present in all three tissues of both Asupina and Cavendish cultivars at a similar intensity, confirming the integrity of the mRNA (Fig. 3). *PSY1* and *PSY2a* amplicons were detected in leaf and green fruit of Asupina and Cavendish cultivars, but were not detected in ripe fruit of either cultivar (Fig. 3). In both banana cultivars, *PSY1* amplicons were more abundant in leaf tissues, while *PSY2a* appeared to be more abundant in green fruit than in leaf, suggesting some tissue specificity of *PSY* expression in banana.

Functional analysis of Asupina and Cavendish PSY enzymes

To confirm that the *PSY1* and *PSY2a* genes from both banana cultivars encode functional proteins, banana *PSY1* and *PSY2* were expressed in rice callus under the control of the maize polyubiquitin1 promoter. Rice callus was selected because it has low carotenoid content and callus tissue has previously been used to compare the activity of PSY enzymes from different plants (Paine et al. 2005). *Zm-PSY1*, the maize *PSY* gene used to generate 'Golden Rice 2' (Paine et al. 2005), was selected for comparison. Rice callus expressing banana *PSY1* or *PSY2a* varied in colour from pale yellow to dark orange, compared to pale negative control callus (Fig. 4). Callus proliferated rapidly regardless of colour, with the exception of some dark orange lines which grew very slowly (results not shown). Extracts from three of the most intensely coloured callus clumps transformed with each *PSY* gene were analysed for carotenoid content. Carotenoid content in all callus expressing *PSY* was significantly higher than that measured in

negative control callus (Fig. 5a), confirming that all of the banana PSYs tested were functional. In addition, the carotenoid content in callus expressing *PSY* ranged from 600 to 960 μg g^{-1} DW and did not vary significantly (Fig. 5a).

In similar *PSY* expression experiments performed in *Arabidopsis* callus, large increases in total carotenoid levels were determined to result from increases in β-carotene, while other carotenes and xanthophylls were only slightly elevated compared to non-transformed callus tissues (Maass et al. 2009). In order to determine if β-carotene was the main carotenoid in rice callus expressing *PSY*, rice callus expressing either *A-PSY2a* or *Zm-PSY1* was analysed by *in situ* Raman spectroscopy, which can be used to identify carotenoids and discriminate between major constituents (Baranska et al. 2006). The Raman spectrum of β-carotene contains three characteristic peaks at 1,523, 1,156 and 1,003 cm^{-1} (Fig. 5b). Identical peaks (1,521 ± 2, 1,155 ± 2 and 1,005 ± 2 cm^{-1}) were observed in the Raman spectrum obtained from rice callus transformed with *A-PSY2a* and *Zm-PSY1* but not in spectra obtained from untransformed rice callus extracts (Fig. 5c). The absence of marked peaks from other carotenoids suggested that β-carotene is the major carotenoid accumulating in rice callus expressing *PSY*.

While the rice callus system confirmed functionality of the banana PSYs, enzyme activities could not be accurately compared due to variations between *PSY* expression levels in different transformed calli. Therefore, banana *PSY1* and *PSY2a* were co-expressed with geranylgeranyl diphosphate synthase (GGPS) and a bacterial desaturase (CrtI) in *E. coli*. Co-expression with GGPS and CrtI facilitates lycopene accumulation in *E. coli* when complemented with functional PSY (Welsch et al. 2010; Fig. 6). While there was no significant difference in lycopene accumulation detected between Cavendish PSY enzymes, lycopene accumulation was significantly higher (~2 fold) in *E. coli* expressing Asupina

PSY1 and PSY2a, compared to their respective Cavendish PSY enzymes (Fig. 6a). Further, A-PSY2a activity was significantly higher than that of both A-PSY1 and Ma-PSY1 (Fig. 6a). Immunoblots (Fig. 6b) confirmed that the lycopene accumulation observed in bacterial lysates of *E. coli* cells expressing *PSY1* and *PSY2a* from both Asupina and Cavendish (Fig. 6a) was catalysed by similar amounts of PSY enzyme.

Discussion

The *PSY* gene has been previously characterised as a multi-gene family consisting of two or more members in several monocots and dicots (Busch et al. 2002; Clotault et al. 2008; Giorio et al. 2008; Li et al. 2008). Phylogenetic comparison of *PSY* partial sequences from five different banana cultivars against *PSY* sequences from a broad range of plants revealed two paralogs in the banana *PSY* gene family (Fig. 1). Interestingly, these two paralogs exist in all five banana cultivars, which have different genotypes (containing the A, B and T genomes) and fruit pVAC content. Asupina (Fe'i genome), Wain (Fe'i genome) and Lakatan (AA genome) accumulate high levels of pVAC in fruit with β-carotene levels of 14.5 μg g^{-1} FW, 5.01 μg g^{-1} FW and 2.89 μg g^{-1} FW (Englberger et al. 2006), respectively, while Cavendish (AAA genome) and Ladyfinger (AAB genome) have low fruit pVAC content of 0.72 μg g^{-1} FW and 1.12 μg g^{-1} FW, respectively (Englberger et al. 2006). Analysis of gene structure (Fig. S2b) indicated that both *PSY* genes were intact in the genome of Asupina and Cavendish. In addition, the exon/intron structure of the Asupina and Cavendish *PSY1* and *PSY2a* genes (Fig. S2b) was consistent with those of *PSY* genes in other monocots, such as rice, maize (Li et al. 2008) and wheat (Singh et al. 2009). Interestingly, in Poaceae, a third *PSY* gene (*PSY3*) is present and is exclusively induced in abiotic stress conditions (Li et al. 2008). Although we isolated clones corresponding to only two distinct paralogs in Asupina, we identified a partial sequence from Cavendish fruit mRNA putatively named *PSY2b* (results not shown), which appeared to be a subgroup of PSY2 (Fig. S1). Therefore, it is possible that bananas possess more than two *PSY* gene family members in their genomes. However, further characterization of this sequence would be required to validate this hypothesis.

The Asupina and Cavendish cultivars were selected for preliminary investigation of the expression of *PSY1* and *PSY2a* in the leaf, green fruit and ripe fruit (Fig. 3) because of contrasting levels of pVAC in ripe fruit tissue. In both of these cultivars, *PSY1* and *PSY2a* mRNA was detected in leaf and green fruit tissue but not in ripe fruit. This suggested that the mRNA abundance profile of both genes was similar in these cultivars. Further, our expression analysis indicated that carotenoid biosynthesis is highly active in Asupina and Cavendish green fruits and is strongly reduced in ripe fruits, similar to what is observed during tomato fruit ripening (Giorio et al. 2008). Maize *PSY1* expression was found to correlate with carotenoid production in the seed endosperm, while *PSY2* mRNA was only detected in leaf (Gallagher et al. 2004). In contrast, in banana fruits, the *PSY* paralogs which had higher DNA sequence homology to maize *PSY2* (*A-PSY2a* and *Ma-PSY2a*) were more abundant compared to the banana *PSY1* homolog. This provides support for the hypothesis that the roles of *PSY1* and *PSY2* are likewise segregated in banana. Further, tomato, which is also a climacteric fruit, exhibits differential *PSY* gene expression; *PSY1* expression was observed in fruit and petals while *PSY2* was expressed primarily in leaves and petals (Giorio et al. 2008). The possession of multiple *PSY* genes may enable the use of critical photosynthetic pigments in non-photosynthetic organs and subcellular compartments without detriment to essential photosynthetic functions (Li et al. 2008).The absence of *PSY1* and *PSY2a* amplicons in RT-PCRs performed using ripe fruit mRNA as templates (Fig. 3) suggests an increasing degradation of these transcripts at this stage of fruit development. We suggest that this affected the quality of the *PSY2a* transcripts in ripe fruit mRNA preparations used as templates for RLM-RACE and explains why we were unable to isolate full length *Ma-PSY2a* using 5' and 3' RLM-RACE from fruit mRNA.

Degradation of plant mRNAs may commence from either the 5' or 3' end of the molecule (reviewed in Belostotsky and Sieburth 2009). Therefore, the position of the *Ma-PSY2a* RT-PCR primers in the 5' UTR, compared to the degenerate primers located in the middle of the coding region, may also explain why we could obtain partial sequences from ripe fruit while RT-PCR yielded no amplicons.

Differences in the regulation of post-transcriptional mRNA processing may explain the inability to isolate the 5' region of *PSY2a* from Cavendish fruit mRNA using RLM-RACE (Fig. S2a). Tandem repetition of the 3' coding region in the *Ma-PSY2a* 3' UTR (Fig. S2a) and the isolation of a large number of prematurely truncated *Ma-PSY2a* transcripts (results not shown) provide evidence for this hypothesis. The length of the 3' UTR has been reported to affect mRNA stability by altering secondary structure (Belostotsky and Sieburth 2009). Therefore, differences in 3' UTR structure of *Ma-PSY2a*, as compared to *A-PSY2a*, may influence the regulation of these transcripts, and the tandem repeat in the *Ma-PSY2a* 3' UTR may reduce mRNA stability. This insertion may also result in the production of non-functional Ma-PSY2a polypeptides. The production of alternately spliced transcripts in wheat *PSY1* due to a sequence duplication in the 3' end of exon 2 resulted in non-functional PSY1 isoforms (Howitt et al. 2009). These variants included a transcript encoding the complete PSY1 polypeptide with a 35 amino acid insertion in the catalytic site. Recently, the role of alternative splicing of *PSY1* transcripts in barley endosperm carotenoid biosynthesis was investigated (Rodriguez-Suarez et al. 2011). This study identified a large number of mRNA variants, many of which were endosperm specific. Functional analysis in *E. coli* showed that only 5 of the 16 identified barley PSY1 isoforms directed carotenoid accumulation (Rodriguez-Suarez et al. 2011). Therefore, if alternative

splicing produces non-functional *Ma-PSY2a* mRNAs in fruit, this would likely cause production of low levels of active MaPSY2a, despite comparable mRNA levels to those in Asupina fruit (Fig. 3). This may be linked to the insertion of the 3' end of the *Ma-PSY2a* coding region in the 3' UTR. In conjunction with the differences observed in enzyme activity (Fig. 6), high levels of non-functional *Ma-PSY2a* mRNA, due to alternative splicing, may play a significant role in the differences in fruit pVAC content between Asupina and Cavendish.

Qualitative analysis in rice callus revealed that all four banana *PSY* sequences encoded active enzymes (Fig. 4). Further, the absence of peaks other than those corresponding to β-carotene in the Raman spectra obtained from rice callus (Fig. 5b and Fig. 5c) suggests that carotenoid accumulation in transgenic rice callus (Fig. 4 and Fig. 5a) was largely due to accumulation of β-carotene. This is similar to what was observed in transgenic *Arabidopsis* callus over-expressing endogenous *PSY* (Maass et al. 2009).

Carotenoid content in banana fruit from Asupina and Cavendish cultivars differ by at least an order of magnitude. However, in green fruits, the transcript levels of both *PSY* genes were of similar abundance in both cultivars and thus mRNA abundance does not explain the difference in pVAC content. Since the functional analysis performed with rice callus (Fig. 5a) did not provide comparative enzyme activity data, we compared Asupina and Cavendish PSY activity using heterologous expression in a bacterial system. Previously, this system was used to determine that a single amino acid substitution in cassava *PSY2* resulted in a three-fold increase in enzymatic activity, which was associated with a 7 – 10 fold increase in root carotenoid accumulation (Welsch et al. 2010). This relationship between increased enzyme activity and carotenoid accumulation correlates well with the data that we obtained; we observed 2 fold higher

enzymatic activity with Asupina PSYs, compared to their Cavendish counterparts (Fig. 6), although the mature proteins differed in only 10 amino acids (Fig. 2), while fruit β-carotene content was approximately 20 fold higher in Asupina compared to Cavendish (Englberger et al. 2006).

Thus, we propose that the differences in the amino acid sequences of Cavendish and Asupina PSY2a are significant contributors to the differences in pVAC content in these cultivars. Further, we suggest that increased metabolic flux arising from these modest changes in PSY enzymatic activity may counteract carotenoid degradation processes during fruit ripening. Differences in PSY protein sequences have been shown to affect carotenoid accumulation in other systems. A 35 amino acid insertion in the wheat PSY1 active site was identified in cultivars with low endosperm carotenoid content (Howitt et al. 2009). Further, significantly reduced carotenoid accumulation during tomato fruit ripening is observed in lines where point mutations in *PSY1* produced proteins with either a single amino acid substitution or a truncated polypeptide (Gady et al. 2012). It is worth noting, however, that none of the important PSY mutations previously identified in cassava (Welsch et al. 2010), wheat (Howitt et al. 2009) and tomato (Gady et al. 2012) were observed when the amino acid sequences of Asupina and Cavendish PSYs were aligned with the amino acid sequences of the plant *PSY* genes listed in Table S3 (results not shown). Therefore, any amino acid substitutions that may be responsible for differences in Asupina and Cavendish PSY enzyme activity have not been previously characterised.

This study is the first to report the characterisation of *PSY* genes in banana and has demonstrated that at least two *PSY* paralogs, *PSY1* and *PSY2a*, are present in both Cavendish and Asupina banana cultivars. We have also shown that PSY1 and PSY2a enzymes from both banana cultivars were functional but that the

Asupina enzymes exhibited approximately 2–fold higher enzymatic activity compared to the Cavendish enzymes in a heterologous system. These results are similar to those obtained previously in characterization of a *PSY2* SNP variant in cassava (Welsch et al. 2010), where this mutation was also responsible for a significant difference in pVAC content *in planta*. We therefore propose that differences in PSY enzyme activity contribute significantly to the variation in pVAC content in Asupina and Cavendish fruit. The genetic improvement of pVAC levels in rice endosperm to produce "Golden Rice 1" was first achieved using PSY from the monocot species *N. pseudonarcissus* (daffodil) (Ye et al. 2000). Subsequently, the accumulation of pVAC in rice endosperm was significantly enhanced in rice lines transformed with PSY1 from maize, which is a more closely related monocot to rice than daffodil (Paine et al. 2005). Similarly, the *A-PSY2a* gene that we have identified, isolated from Asupina, is a good candidate for the genetic improvement of fruit pVAC content in banana. The use of this banana gene to create genetically modified banana cultivars could potentially reduce the complexity of regulatory approval and increase their acceptance by consumers. We are also characterising promoter elements from suitable banana genes to be used in conjunction with *A-PSY2a*, which may lead to the production of cisgenic or intragenic banana cultivars with high pVAC content from existing agronomically important cultivars. We have a very efficient banana transformation system within our group (Khanna et al. 2004), and creation of cisgenic banana cultivars is possible using appropriate transformation vector and selection systems. The potential impact of such genetically enhanced cultivars to alleviate VAD in banana dependent communities may be significant in terms of nutritional health, regulatory compliance and public acceptance.

Acknowledgements

The authors are grateful to: Doug Becker at QUT for coordinating the supply of plant material from Queensland Department of Employment, Economic Development and Innovation (DEEDI); Jeff Daniells at DEEDI (South Johnstone) for supplying Asupina green fruit tissue; Jennifer Kleidon at QUT for supplying Cavendish and Ladyfinger leaf tissue, maintenance of Asupina, Lakatan and Wain banana plants at QUT, and for training in rice tissue culture and transformation techniques; Sharon Hamill at DEEDI (Nambour) for supplying Asupina, Lakatan and Wain banana plants; Russell Reinke (Yanco Agricultural Institute) for supplying Nipponbare rice seeds; Queensland University of Technology International Doctorate Scholarship program for postgraduate scholarship support and funding. The research was funded by a grant from the Bill and Melinda Gates Foundation Grand Challenges in Global Health Initiative. The authors declare that they have no conflict of interest.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410

Asif MH, Dhawan P, Nath P (2000) A simple procedure for the isolation of high quality RNA from ripening banana fruit. Plant Mol Biol Rep 18:109-115

Baranska M, Baranski R, Schulz H, Nothnagel T (2006) Tissue-specific accumulation of carotenoids in carrot roots. Planta 224:1028-1037

Bartley GE, Scolnik PA (1993) cDNA cloning, expression during development, and genome mapping of PSY2, a second tomato gene encoding phytoene synthase. J Biol Chem 268:25718-25721

Belostotsky DA, Sieburth LE (2009) Kill the messenger: mRNA decay and plant development. Current Opin Plant Biol 12:96-102

Bender DA (2003) Nutritional biochemistry of the vitamins. 2nd edn. Cambridge University Press, Cambridge

Busch M, Seuter A, Hain R (2002) Functional analysis of the early steps of carotenoid biosynthesis in tobacco. Plant Physiol 128:439-453

Cavatorta J, Perez KW, Gray SM, Van Eck J, Yeam I, Jahn M (2011) Engineering virus resistance using a modified potato gene. Plant Biotechnol J 9:1014-1021

Cazzonelli CI (2011) Carotenoids in nature: insights from plants and beyond. Funct Plant Biol 38:833-847

Chander S, Guo YQ, Yang XH, Zhang J, Lu XQ, Yan JB, Song TM, Rocheford TR, Li JS (2008) Using molecular markers to identify two major loci controlling carotenoid contents in maize grain. Theor Appl Genet 116:223-233

Chang S, Puryear J, Cairney J (1993) A simple and efficient method for isolating RNA from pine trees. Plant Mol Biol Rep 11:113-116

Christensen AH, Quail PH (1996) Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. Transgenic Res 5:213-218

Chu CC, Wang CS, Sun CS, Hsu C, Yin KC, Chu CY, Bi FY (1975) Establishment of an efficient medium for anther culture of rice through comparative experiments on the nitrogen sources. Sci Sinica 18:659-668

Clotault J, Peltier D, Berruyer R, Thomas M, Briard M, Geoffriau E (2008) Expression of carotenoid biosynthesis genes during carrot root development. J Exp Bot 59:3563-3573

Daniells J, Jenny C, Karamura D, Tomekpe K (2001) Diversity in the genus *Musa*. Musalogue: a catalogue of *Musa* germplasm. International Network for the Improvement of Banana and Plantain, Montpellier

Demmig-Adams B, Adams WW, 3rd (2002) Antioxidants in photosynthesis and human nutrition. Science 298:2149-2153

Depicker A, Stachel S, Dhaese P, Zambryski P, Goodman HM (1982) Nopaline synthase: transcript mapping and DNA sequence. J Mol Appl Genet 1:561-573

Dower WJ, Miller JF, Ragsdale CW (1988) High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res 16:6127-6145

Englberger L, Schierle J, Marks GC, Fitzgerald MH (2003) Micronesian banana, taro, and other foods: newly recognized sources of provitamin A and other carotenoids. J Food Compos Anal 16:3-19

Englberger L, Wills RB, Blades B, Dufficy L, Daniells JW, Coyne T (2006) Carotenoid content and flesh color of selected banana cultivars growing in Australia. Food Nutr Bull 27:281-291

Fray RG, Wallace A, Fraser PD, Valero D, Hedden P, Bramley PM, Grierson D (1995) Constitutive expression of a fruit phytoene synthase gene in transgenic tomatoes causes dwarfism by redirecting metabolites from the gibberellin pathway. Plant J 8:693-701

Gady A, Vriezen W, Van de Wal M, Huang P, Bovy A, Visser R, Bachem C (2012) Induced point mutations in the phytoene synthase 1 gene cause differences in carotenoid content during tomato fruit ripening. Mol Breed 29:801-812

Gallagher CE, Matthews PD, Li F, Wurtzel ET (2004) Gene duplication in the carotenoid biosynthetic pathway preceded evolution of the grasses. Plant Physiol 135:1776-1783

Garcia MA, Altieri MA (2005) Transgenic Crops: Implications for biodiversity and sustainable agriculture. Bull Sci Technol Soc 25:335-353

Giorio G, Stigliani AL, D'Ambrosio C (2008) Phytoene synthase genes in tomato (*Solanum lycopersicum* L.) - new data on the structures, the deduced amino acid sequences and the expression patterns. FEBS J 275:527-535

Hirschberg J (2001) Carotenoid biosynthesis in flowering plants. Curr Opin Plant Biol 4:210-218

Howitt CA, Cavanagh CR, Bowerman AF, Cazzonelli C, Rampling L, Mimica JL, Pogson BJ (2009) Alternative splicing, activation of cryptic exons and amino acid substitutions in carotenoid biosynthetic genes are associated with lutein accumulation in wheat endosperm. Funct Integr Genomics 9:363-376
Hribova E, Cizkova J, Christelova P, Taudien S, de Langhe E, Dolezel J (2011) The ITS1-5.8S-ITS2 sequence region in the Musaceae: structure, diversity and use in molecular phylogeny. PLoS One 6:e17863
IITA (2009) Banana and Plantain Systems. http://www.iita.org/c/document_library/get_file?p_l_id=25410&folderId=27021&name=DLFE-172.pdf
Khanna H, Becker D, Kleidon J, Dale JL (2004) Centrifugation assisted agrobacterium tumefaciens-mediated transformation (CAAT) of embryogenic cell suspensions of banana (*Musa* spp. Cavendish AAA and Lady finger AAB). Mol Breed 14:239-252
Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics (Oxford, England) 23:2947-2948
Lecomte L, Duffe P, Buret M, Servin B, Hospital F, Causse M (2004) Marker-assisted introgression of five QTLs controlling fruit quality traits into three tomato lines revealed interactions between QTLs and genetic backgrounds. Theor Appl Genet 109:658-668
Li F, Vallabhaneni R, Wurtzel ET (2008) *PSY3*, a new member of the phytoene synthase gene family conserved in the Poaceae and regulator of abiotic stress-induced root carotenogenesis. Plant Physiol 146:1333-1345
Maass D, Arango J, Wüst F, Beyer P, Welsch R (2009) Carotenoid crystal formation in *Arabidopsis* and carrot roots caused by increased phytoene synthase protein levels. PloS One 4:e6373
Malik KA (1992) Some universal media for the isolation, growth and purity checking of a broad-spectrum of microorganisms. World J Microb Biot 8:453-456
Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH (2011) CDD: a conserved domain database for the functional annotation of proteins. Nucleic Acids Res 39: D225-229
Mayer, J. E., Pfeiffer, W. H., Beyer, P. (2008) Biofortified crops to alleviate micronutrient malnutrition. Curr Opin Plant Biol 11: 166-170
Molesini B, Pii Y, Pandolfini T (2012) Fruit improvement using intragenesis and artificial microRNA. Trends Biotechnol 30:80-88
Paine JA, Shipton CA, Chaggar S, Howells RM, Kennedy MJ, Vernon G, Wright SY, Hinchliffe E, Adams JL, Silverstone AL, Drake R (2005) Improving the nutritional value of Golden Rice through increased pro-vitamin A content. Nat Biotechnol 23:482-487
Perez-Jones A, Mallory-Smith CA, Hansen JL, Zemetra RS (2006) Introgression of an imidazolinone-resistance gene from winter wheat (*Triticum aestivum* L.) into jointed goatgrass (*Aegilops cylindrica* Host). Theor Appl Genet 114:177-186
Pillay M, Tenkouano A, Ude G, Ortiz R (2004) Molecular characterization of genomes in *Musa* and its applications. In: Jain SM, Swennen R (eds) Banana improvement: cellular, molecular biology, and induced mutations. Science Publishers, Enfield
Rodriguez-Suarez C, Atienza SG, Piston F (2011) Allelic variation, alternative splicing and expression analysis of *Psy1* gene in *Hordeum chilense* Roem. et Schult. PloS One 6:e19885
Rommens CM (2010) Barriers and paths to market for genetically engineered crops. Plant Biotechnol J 8:101-111
Rommens CM, Haring MA, Swords K, Davies HV, Belknap WR (2007) The intragenic approach as a new extension to traditional plant breeding. Trends Plant Sci 12:397-403
Schledz M, Al Babili S, Von Lintig J, Haubruck H, Rabbani S, Kleinig H, Beyer P (1996) Phytoene synthase from *Narcissus pseudonarcissus*: functional expression, galactolipid requirement, topological distribution in chromoplasts and induction during flowering. Plant J 10:781-792
Schouten HJ, Krens FA, Jacobsen E (2006) Cisgenic plants are similar to traditionally bred plants. EMBO Rep 7:750-753
Scolnik PA, Bartley GE (1994) Nucleotide sequence of an *Arabidopsis* cDNA for phytoene synthase. Plant Physiol 104:1471-1472
Sharrock S, Frison E (1998) *Musa* production around the world – trends, varieties and regional importance. Networking Banana and Plantain: Annual Report 1998:41-46

Singh A, Reimer S, Pozniak CJ, Clarke FR, Clarke JM, Knox RE, Singh AK (2009) Allelic variation at *Psy1-A1* and association with yellow pigment in durum wheat grain. Theor Appl Genet 118:1539-1548
Stewart CN, Jr., Via LE (1993) A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. Biotechniques 14:748-750
Taylor M, Ramsay G (2005) Carotenoid biosynthesis in plant storage organs: recent advances and prospects for improving plant food quality. Physiologia Plantarum 124:143-151
Vanblaere T, Szankowski I, Schaart J, Schouten H, Flachowsky H, Broggini GAL, Gessler C (2011) The development of a cisgenic apple plant. J Biotechnol 154:304-311
Welsch R, Arango J, Bar C, Salazar B, Al-Babili S, Beltran J, Chavarriaga P, Ceballos H, Tohme J, Beyer P (2010) Provitamin A accumulation in cassava (*Manihot esculenta*) roots driven by a single nucleotide polymorphism in a phytoene synthase gene. Plant Cell 22:3348-3356
Welsch R, Beyer P, Hugueney P, Kleinig H, von Lintig J (2000) Regulation and activation of phytoene synthase, a key enzyme in carotenoid biosynthesis, during photomorphogenesis. Planta 211:846-854
West CE (2000) Meeting requirements for vitamin A. Nutr Rev 58:341-345
West KP, Jr. (2002) Extent of vitamin A deficiency among preschool children and women of reproductive age. J Nutr 132:2857S-2866S
Wheelan SJ, Church DM, Ostell JM (2001) Spidey: a tool for mRNA-to-genomic alignments. Genome Res 11: 1952-1957
Ye X, Al-Babili S, Kloti A, Zhang J, Lucca P, Beyer P, Potrykus I (2000) Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science 287:303-305

Tables

Table S1 PSY sequences used for design of degenerate primers

Source	Accession Number
Rice (PSY1)	AAS18307
Maize (PSY1)	NP_001108124.2
Maize (PSY2)	NP_001108117.1
Arabidopsis	NP_001031895
Grapefruit	AAD38051.2
Orange	AAF33237.1
Melon	CAA85775.1
Tomato	ABM45873
Capsicum	EU753855
Wolfberry	AAW88383
Marigold	AAM45379
Sunflower	AJ304825.1
Daffodil	CAA55391
Carrot	BAA84763.1

Table S2 Sequence of primers used in this study. Restriction enzyme sites are indicated in the primer sequence in lowercase

Primer	Sequence (5'-3')
P1	RTNTGGGCNATHTAYGTNTGGTG
P2	RAARTYRTARTCRTTNGCYTCDAT
P3	GCTGATGGCGATGAATGAACACTG
P4	CGACTTCTTCAAGTCCATTCTCATTCC
P5	TTGTACCTCGATTTCCGCAGGTC
P6	CGCGGATCCGAACACTGCGTTTGCTGGCTTTGATG
P7	ATTCTCATTCCTTCGATCATGTCCCTG
P8	TTCCGCAGGTCCATTCTCATTCC
P9	GCGAGCACAGAATTAATACGACT
P10	GCGGCAAGGTGACTGAGAAATGG
P11	CATTAGCTCTTGGCATCGCGAATC
P12	CGCGGATCCGAATTAATACGACTCACTATAGG
P13	AATGGAGGAGCTTCATGAAG
P14	TCGCGAATCAACTCACCAACATAC
P15	GTAATACGACTCACTATAGGGC
P16	ACTATAGGGCACGCGTGGT
P17	GTACGCCGCGTCGAGTAG
P18	CAAGTAGAAAGTCTTGGCATACTC
P19	CCTCCACCATCTCCTCTTCA
P20	AACTGGCTGAACCCAGAGCTC
P21	ggatccATGGCGTGCCTGTTGCTACGG
P22	tctagaTCATGTTTTAGCTAAACTTTGG
P23	tctagaTCATGTTTTTGCTAAGTTTGACTGGCTC
P24	ggatccATGTCTGGCTCTATTGTTTGG
P25	tctagaTTATAGTGTTCCTGCAAATTTGG
P26	ccatggATGTCTGGCTCTGTTGTTTGGGTTGT
P27	tctagaTTATAGTGTTCCTGCAAATCTTGAAGG
P28	TGGCTGACACTGACGACATTC
P29	CAACAATACTTGGGAAAACGG
P30	GCCTGTTGCTACGGATGATTGC
P31	ACTCCATCTTCGCCTCTTCCTC
P32	CAAGTGGACGCCTGTCTCGG
P33	CAAGTGGATGCCTGTCTCGG
P34	CGCTGCTTAGTTTCGTCTTCCG
P35	GAGTAGAATTCATGAAACCAACTACGG
P36	CCTGCAAGCTTTCAAATCAGATCCTCC
P37	GTCTGAATTCACCCCATGCTGATTCC
P38	GAACGTAAGCTTCATGTTTTTGCTAAG
P39	GAACGTAAGCTTCATGTTTTAGCTAAAC
P40	GTGCGGGGAATTCGTGTGCCCGTTCGGCG
P41	TTTGAACGAAGCTTTTATAGTGTTCCTGC
P42	GTGCGGGGAATTCGTGCGCCCGGTCGGCG
P43	TTTGAACGAAGCTTTTATAGTGTTCCTGC

Table S3 Sequences used for phylogenetic analysis

Abbrev.	Scientific name	Common name	Accession No.
		Monocot	
A	*Musa sp.*	Asupina	JX195658 JX195659
Ma	*Musa acuminata*	Cavendish	JX195664 JX195665
	Musa sp.	Ladyfinger	JX195671 JX195672
	Musa sp.	Lakatan	JX195669 JX195670
	Musa sp.	Wain	JX195667 JX195668
Os	*Oryza sativa*	rice	NM_001065182 AK073290 NM_001070427
Sb	*Sorghum bicolor*	sorghum	AY705389 XM_002442533 AY705390
Tt	*Triticum turgidum*	durum wheat	DQ642443 DQ642445
Zm	*Zea mays*	maize	AY324431 AY325302 DQ372936
Np	*Narcissus pseudonarcissus*	daffodil	X78814
		Dicot	
Dc	*Daucus carota*	carrot	DQ192186 DQ192187
Sl	*Solanum lycopersicum*	tomato	EF534739 EF534738
Me	*Manihot esculenta*	cassava	GU111719 GU111720
Ad	*Actinidia deliciosa*	kiwifruit	FJ797304
At	*Arabidopsis thaliana*	thale cress	NM_121729
Br	*Brassica rapa*	field mustard	GQ200740
Cp	*Carica papaya*	papaya	DQ666828
Cu	*Citrus unshiu*	satsuma orange	AF220218
Cm	*Cucumis melo*	muskmelon	GU361622
Ha	*Helianthus annuus*	sunflower	AJ304825
Lb	*Lycium barbarum*	Chinese wolfberry	AY920918
Nt	*Nicotiana tabacum*	tobacco	HM345582
Pm	*Prunus mume*	Japanese apricot	AB253628
Ds	*Dunaliella salina*	Green algae	AY601075

Table S4 Primer pairs

Application	**Type of PCR**	**Primer pair**
psy1 5'RACE	outer PCR	P3/P4
	nested PCR	P6/P7
psy1 3'RACE	outer PCR	P9/P10
	nested PCR	P12/P13
psy2 5'RACE	outer PCR	P3/P5
	nested PCR	P6/P8
psy2 3'RACE	outer PCR	P9/P11
	nested PCR	P12/P14
Genome Walking Fragment 1	outer PCR	P15/P5
	nested PCR	P16/P8
Genome Walking Fragment 2	outer PCR	P15/P17
	nested PCR	P16/P18
Genome Walking Fragment 3	outer PCR	P15/P19
	nested PCR	P16/P20
Asupina *psy1*coding region		P21/P22
Cavendish *psy1* coding region		P21/P23
Asupina *psy2a* coding region		P24/P25
Cavendish *psy2a* coding region		P26/P27

Figure Legends

Fig. 1 Phylogenetic analysis of banana *PSY* sequences. Neighbour Joining tree showing the relationship between the consensus nucleotide sequences of the partial coding regions of *PSY* genes isolated from Asupina (Fe'i genome, high pVAC), Cavendish (AAA genome, low pVAC), Ladyfinger (AAB genome, low pVAC), Lakatan (AA genome, high pVAC) and Wain (Fe'i genome, high pVAC) banana cultivars compared to a selection of monocot and dicot *PSY* sequences (Table S3): rice (*Os*), sorghum (*Sb*), durum wheat (*Tt*), maize (*Zm*), daffodil (*Ns*), carrot (*Dc*), tomato (*Sl*), cassava (*Me*), kiwifruit (*Ad*), thale cress (*At*), field mustard (*Br*), papaya (*Cp*), satsuma orange (*Cu*), muskmelon (*Cm*), sunflower (*Ha*), tobacco (*Nt*), Japanese apricot (*Pm*), and the green algae *Dunaliella salina* (*Ds*) which served as the outgroup. The tree was bootstrapped to 10,000 trials, and bootstrapping values are shown at each node of the rooted rectangular cladogram

Fig. 2 Analysis of banana PSY protein sequences. The Asupina (**a**) PSY1 and (**b**) PSY2a amino acid sequences were compared to their respective Cavendish PSY homologues. The position of amino acids is numbered according to the Asupina PSY polypeptides. Insertion of an AVE tripeptide in the Cavendish PSY2a protein is marked with an arrow (position in sequence marked in brackets), while amino acid substitutions are indicated on the sequence as a line. The putative transit peptide (TP) and mature protein (MP) are indicated above each sequence diagram with putative lengths in brackets

Fig. 3 RT-PCR detection of *PSY* mRNA in leaf (L), green fruit (GF) pulp, and ripe fruit (RF) pulp of Asupina and Cavendish banana plants. Primers were designed to amplify approximately 100 bp of the *actin1* 5' coding region, the *PSY1* 5' coding region or the *PSY2a* 5' UTR from mRNA. The same number of cycles were used in all RT-PCR determinations. Amplicons were resolved by electrophoresis on 2% (w/v) agarose gels using SYBR® safe DNA gel stain (Invitrogen)

Fig. 4 Visual analysis of carotenoid accumulation in rice callus transformed with *PSY* coding regions. Each panel is a close-up [8× magnification] of the darkest orange callus visible at 7 weeks post transformation and shows: Untransformed wild type (WT) callus without selection; WT callus on selection; *pCAMBIA-1302* empty vector (EV) control; *Zm-PSY1* ; *A-PSY1*; *A-PSY2a*; *Ma-PSY1* and *Ma-PSY2a*

Fig. 5 Analysis of carotenoid accumulation in transgenic rice calli. (**a**) Carotenoid amounts [μg g^{-1} dry Weight (dW)] of the three most intensively coloured calli transformed with the following *PSY* genes: maize *PSY1* (*Zm-PSY1*); Asupina *PSY1* (*A-PSY1*) and *PSY2a* (*A-PSY2a*); Cavendish *PSY1* (*Ma-PSY1*) and *PSY2a* (*Ma-PSY2a*) including untransformed (WT) and empty vector (EV) controls. Error bars represent SD. (**b**) Raman spectrum at 320 nm excitation for β-carotene standard with peaks for C=C at 1,523 cm^{-1}, C–C at 1,156 cm^{-1} and $–CH_3$ at 1,003 cm^{-1} Raman shift (peaks labelled a, b and c, respectively). (**c**) Raman spectra for rice callus expressing (1) *Zm-PSY1*, (2) *A-PSY2a*, and (3) *CrtI* all under the control of maize polyubiquitin1 promoter and (4) untransformed callus. β-carotene Raman shift peaks were observed at approximately 1,521 ± 2, 1,155 ± 2 and 1,005 ± 2 cm^{-1} (peaks labelled a, b and c, respectively)

Fig. 6 Heterologous expression of Asupina and Cavendish PSY proteins in *E. coli*. (**a**) Production of lycopene in recombinant *E. coli* cells. PSY1 and PSY2a from Asupina (A) and Cavendish (Ma) were expressed separately in an *E. coli* strain expressing GGPS and the bacterial phytoene desaturase (*CrtI*), thus enabling lycopene production. Error bars in (**a**) represent the standard error of seven biological replicates. One-way ANOVA with least significant difference post-hoc analysis was used to determine the statistical significance of lycopene accumulation with a 95% confidence level. Letters above each bar (a, b and c) in the graph indicate treatments where lycopene levels were not statistically different. (**b**) Immunoblots were conducted using anti-6*His-Tag antibodies and cell lysate corresponding to 10 μg (PSY) and 500 ng (GGPS) protein per lane, respectively

Fig. S1 Alignment of banana PSY partial protein sequences. Neighbour Joining multiple sequence alignment of banana PSY1 and PSY2 partial protein sequences translated from the partial nucleotide sequences used in Fig. 1 (Table S3) and Cavendish *PSY2b* (*Ma-PSY2b;* GenBank Accession No. JX195666)

Fig. S2 Comparative structure of *PSY* mRNA and genomic DNA sequences from Asupina and Cavendish cultivars. (**a**) Structure of putative Asupina (A) and Cavendish (Ma) *PSY* mRNA transcripts showing the coding region (black rectangle) between the start (AUG) and stop (UGA or UAA) codons of *A-PSY1* (GenBank Accession No. JX195658), *A-PSY2a* (GenBank Accession No. JX195659), *Ma-PSY1* (GenBank Accession No. JX195664) and *Ma-PSY2a* (GenBank Accession No. JX195665). Putative transcription initiation (t) and polyadenylation (*) sites are marked on the 5' and 3' UTRs (grey arrows), respectively. (**b**) Analysis of genomic sequences of Asupina and Cavendish *PSY* genes. Diagrammatic representation of the structure of the genomic sequences corresponding to the coding region (between the start and stop codons) of Asupina and Cavendish *PSY* genes *A-PSY1* (GenBank Accession No. JX195661), *A-PSY2a* (GenBank Accession No. JX195660), *Ma-PSY1* (GenBank Accession No. JX195663) and *Ma-PSY2a* (GenBank Accession No. JX195662) compared to that of maize (*Zm*) *PSY1* (GenBank Accession No. AY324431.1) and *PSY2* (GenBank Accession No. AY325302.1) showing the positions of the exons (black arrows) and introns (grey lines). The sizes (bp) of the exons (bold numbers) and introns (plain numbers) are shown above and below the coding region of each gene

Figures

Fig. 1

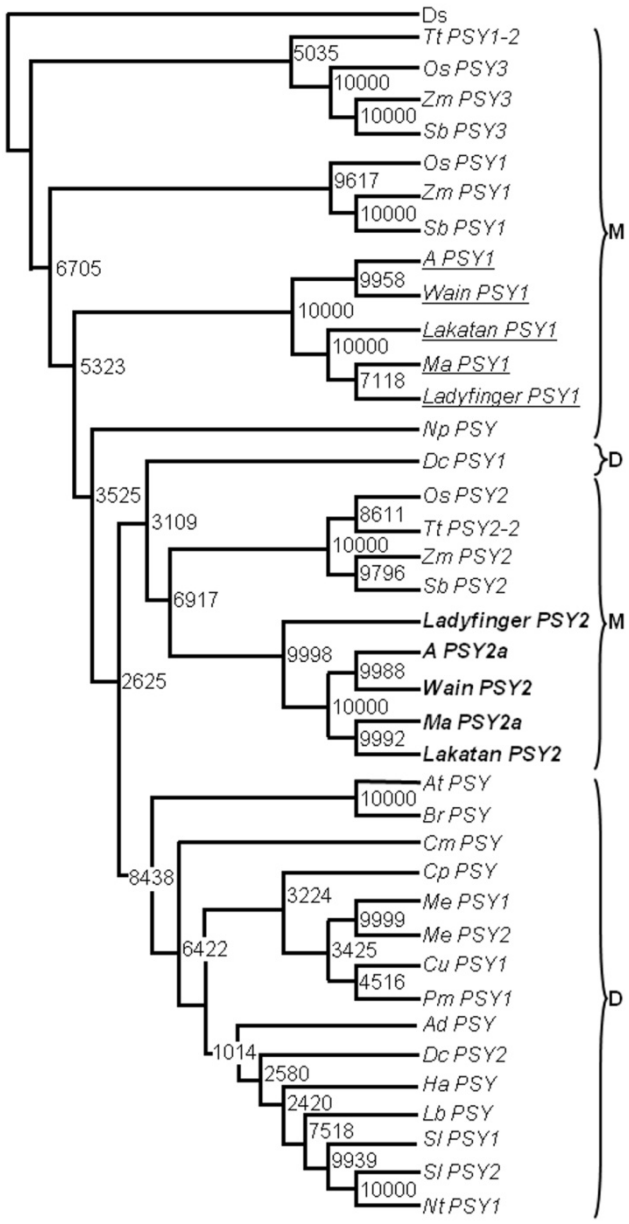


Fig. 2

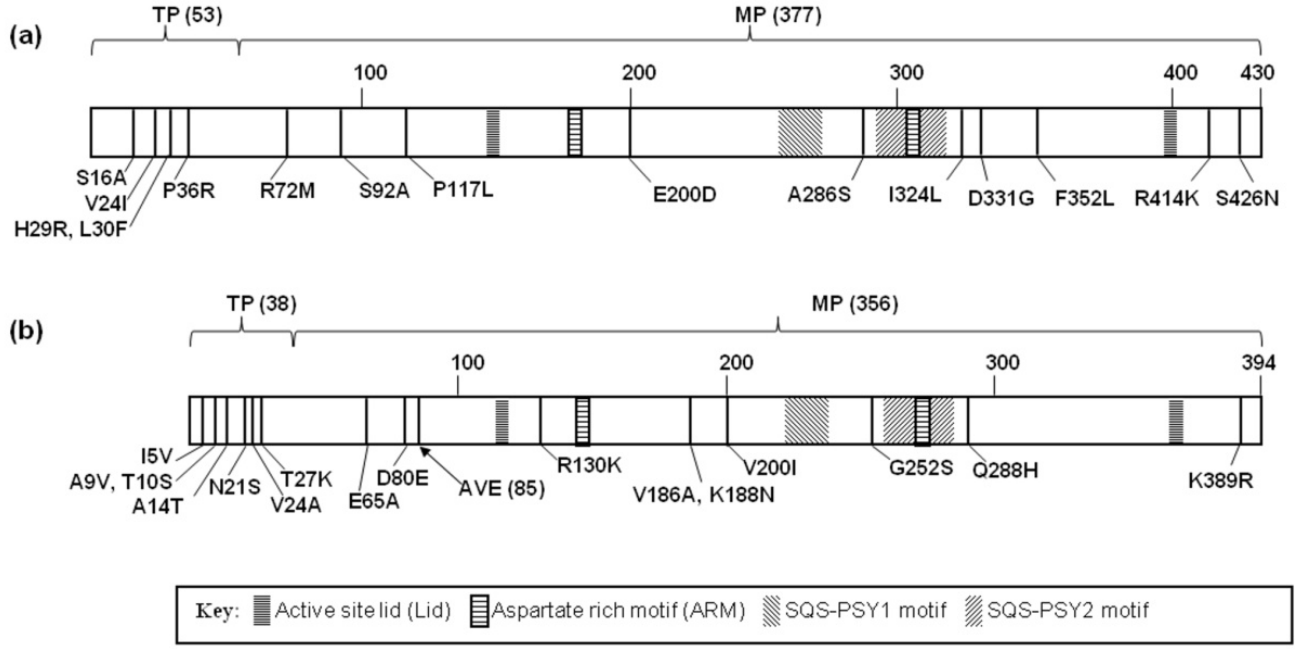

(a)
TP (53)
MP (377)
100
200
300
400
430
S16A
V24I
H29R, L30F
P36R
R72M
S92A
P117L
E200D
A286S
I324L
D331G
F352L
R414K
S426N
(b)
TP (38)
MP (356)
100
200
300
394
I5V
A9V, T10S
A14T
N21S
T27K
V24A
D80E
E65A
AVE (85)
R130K
V186A, K188N
V200I
G252S
Q288H
K389R
Key: Active site lid (Lid) Aspartate rich motif (ARM) SQS-PSY1 motif SQS-PSY2 motif

Fig. 3

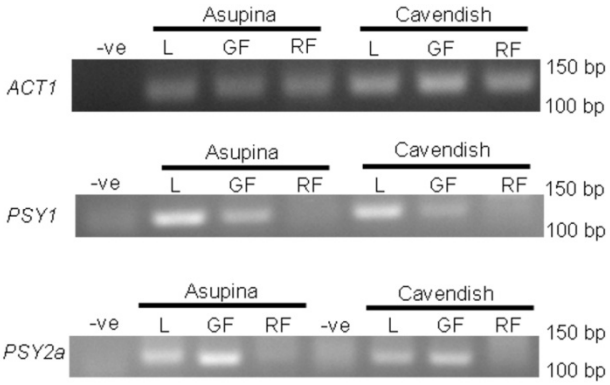

Asupina
Cavendish
-ve
L
GF
RF
L
GF
RF
ACT1
150 bp
100 bp
Asupina
Cavendish
-ve
L
GF
RF
L
GF
RF
PSY1
150 bp
100 bp
Asupina
Cavendish
-ve
L
GF
RF
-ve
L
GF
RF
PSY2a
150 bp
100 bp

Fig. 4

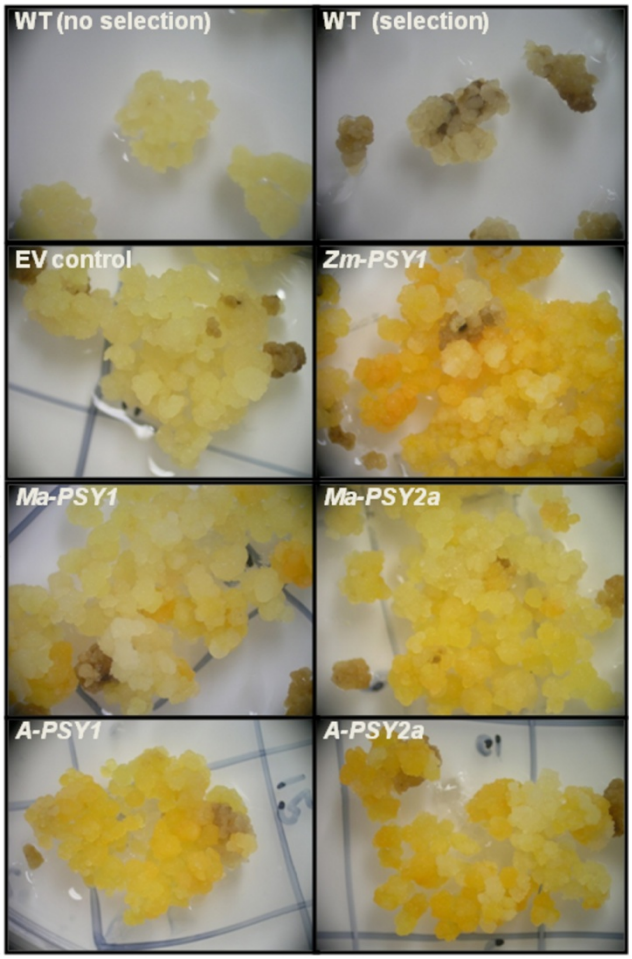

WT (no selection)
WT (selection)
EV control
Zm-PSY1
Ma-PSY1
Ma-PSY2a
A-PSY1
A-PSY2a

Fig. 5

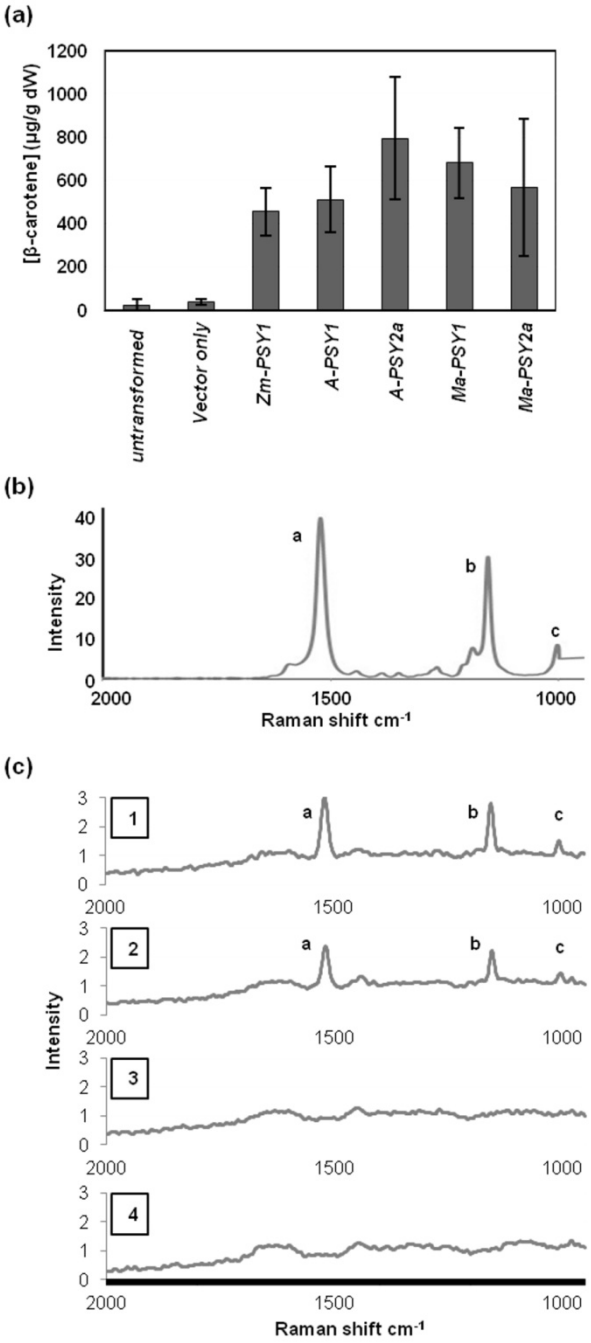

(a)
[β-carotene] (µg/g dW)
1200
1000
800
600
400
200
0
untransformed
Vector only
Zm-PSY1
A-PSY1
A-PSY2a
Ma-PSY1
Ma-PSY2a
(b)
Intensity
40
30
20
10
0
a
b
c
2000
1500
1000
Raman shift cm-1
(c)
Intensity
1
2
3
4
a
b
c
2000
1500
1000
Raman shift cm-1

Fig. 6

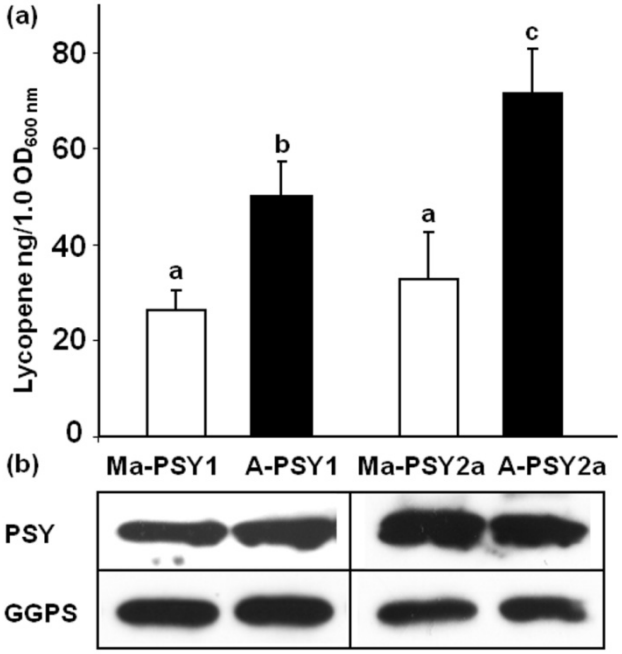

(a)
Lycopene ng/1.0 OD600 nm
0
20
40
60
80
a
b
a
c
(b)
Ma-PSY1
A-PSY1
Ma-PSY2a
A-PSY2a
PSY
GGPS

Fig. S1

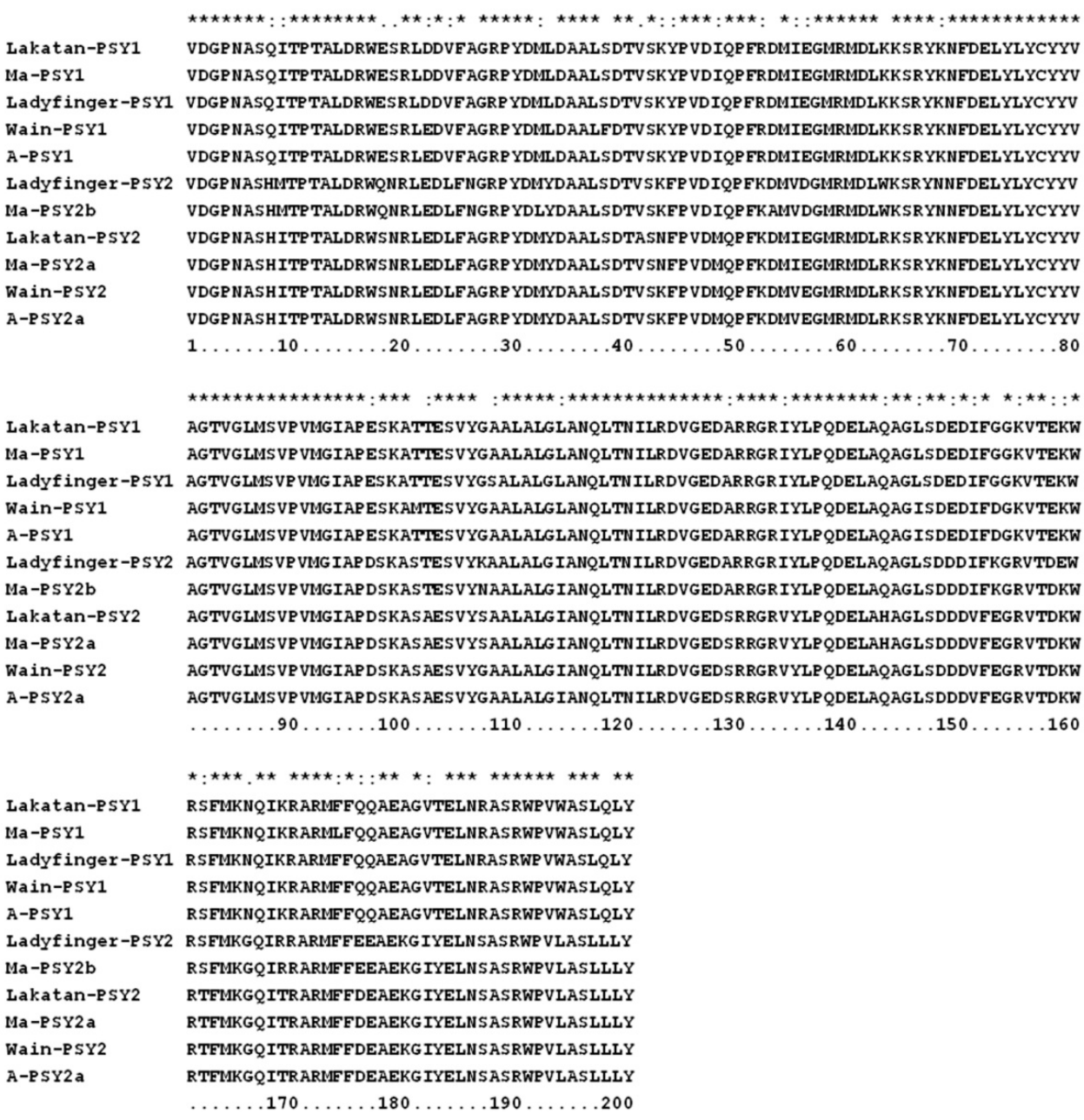
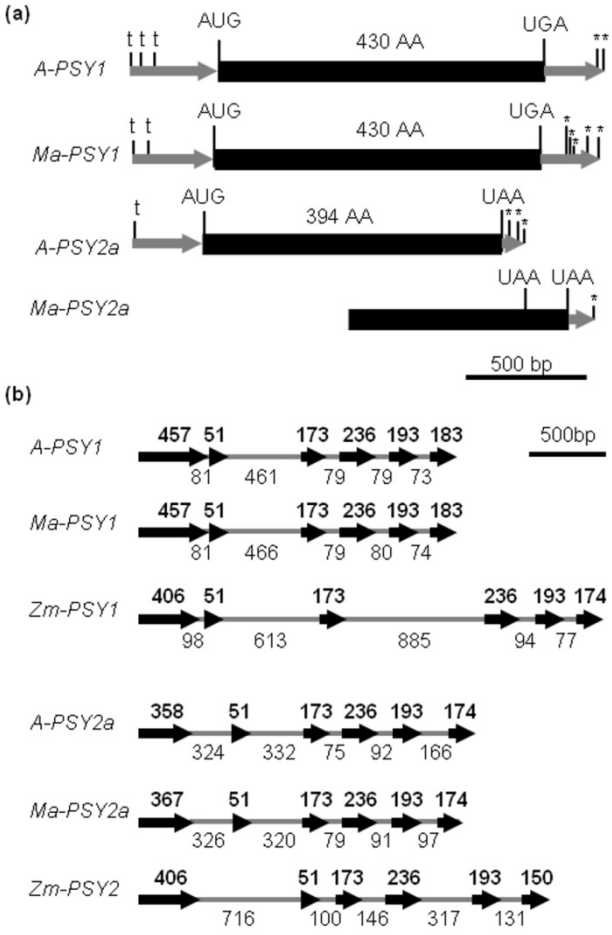

Lakatan-PSY1 VDGPNASQITPTALDRWESRLDDVFAGRPYDMLDAALSDTVSKYPVDIQPFRDMIEGMRMDLKKSRYKNFDELYLYCYYV
Ma-PSY1 VDGPNASQITPTALDRWESRLDDVFAGRPYDMLDAALSDTVSKYPVDIQPFRDMIEGMRMDLKKSRYKNFDELYLYCYYV
Ladyfinger-PSY1 VDGPNASQITPTALDRWESRLDDVFAGRPYDMLDAALSDTVSKYPVDIQPFRDMIEGMRMDLKKSRYKNFDELYLYCYYV
Wain-PSY1 VDGPNASQITPTALDRWESRLEDVFAGRPYDMLDAALFDTVSKYPVDIQPFRDMIEGMRMDLKKSRYKNFDELYLYCYYV
A-PSY1 VDGPNASQITPTALDRWESRLEDVFAGRPYDMLDAALSDTVSKYPVDIQPFRDMIEGMRMDLKKSRYKNFDELYLYCYYV
Ladyfinger-PSY2 VDGPNASHMTPTALDRWQNRLEDLFNGRPYDMYDAALSDTVSKFPVDIQPFKDMVDGMRMDLWKSRYNNFDELYLYCYYV
Ma-PSY2b VDGPNASHMTPTALDRWQNRLEDLFNGRPYDLYDAALSDTVSKFPVDIQPFKAMVDGMRMDLWKSRYNNFDELYLYCYYV
Lakatan-PSY2 VDGPNASHITPTALDRWSNRLEDLFAGRPYDMYDAALSDTASNFPVDMQPFKDMIEGMRMDLRKSRYKNFDELYLYCYYV
Ma-PSY2a VDGPNASHITPTALDRWSNRLEDLFAGRPYDMYDAALSDTVSNFPVDMQPFKDMIEGMRMDLRKSRYKNFDELYLYCYYV
Wain-PSY2 VDGPNASHITPTALDRWSNRLEDLFAGRPYDMYDAALSDTVSKFPVDMQPFKDMVEGMRMDLRKSRYKNFDELYLYCYYV
A-PSY2a VDGPNASHITPTALDRWSNRLEDLFAGRPYDMYDAALSDTVSKFPVDMQPFKDMVEGMRMDLRKSRYKNFDELYLYCYYV
1.....10.....20.....30.....40.....50.....60.....70.....80
Lakatan-PSY1 AGTVGLMSVPVMGIAPESKATTESVYGAALALGLANQLTNILRDVGEDARRGRIYLPQDELAQAGLSDEDIFGGKVTEKW
Ma-PSY1 AGTVGLMSVPVMGIAPESKATTESVYGAALALGLANQLTNILRDVGEDARRGRIYLPQDELAQAGLSDEDIFGGKVTEKW
Ladyfinger-PSY1 AGTVGLMSVPVMGIAPESKATTESVYGSALALGLANQLTNILRDVGEDARRGRIYLPQDELAQAGLSDEDIFGGKVTEKW
Wain-PSY1 AGTVGLMSVPVMGIAPESKAMTESVYGAALALGLANQLTNILRDVGEDARRGRIYLPQDELAQAGISDEDIFDGKVTEKW
A-PSY1 AGTVGLMSVPVMGIAPESKATTESVYGAALALGLANQLTNILRDVGEDARRGRIYLPQDELAQAGISDEDIFDGKVTEKW
Ladyfinger-PSY2 AGTVGLMSVPVMGIAPDSKASTESVYKAALALGIANQLTNILRDVGEDARRGRIYLPQDELAQAGLSDDDIFKGRVTDEW
Ma-PSY2b AGTVGLMSVPVMGIAPDSKASTESVYNAALALGIANQLTNILRDVGEDARRGRIYLPQDELAQAGLSDDDIFKGRVTDKW
Lakatan-PSY2 AGTVGLMSVPVMGIAPDSKASAESVYSAALALGIANQLTNILRDVGEDSRRGRVYLPQDELAHAGLSDDDVFEGRVTDKW
Ma-PSY2a AGTVGLMSVPVMGIAPDSKASAESVYSAALALGIANQLTNILRDVGEDSRRGRVYLPQDELAHAGLSDDDVFEGRVTDKW
Wain-PSY2 AGTVGLMSVPVMGIAPDSKASAESVYGAALALGIANQLTNILRDVGEDSRRGRVYLPQDELAQAGLSDDDVFEGRVTDKW
A-PSY2a AGTVGLMSVPVMGIAPDSKASAESVYGAALALGIANQLTNILRDVGEDSRRGRVYLPQDELAQAGLSDDDVFEGRVTDKW
.....90.....100.....110.....120.....130.....140.....150.....160
Lakatan-PSY1 RSFMKNQIKRARMFFQQAEAGVTELNRASRWPVWASLQLY
Ma-PSY1 RSFMKNQIKRARMLFQQAEAGVTELNRASRWPVWASLQLY
Ladyfinger-PSY1 RSFMKNQIKRARMFFQQAEAGVTELNRASRWPVWASLQLY
Wain-PSY1 RSFMKNQIKRARMFFQQAEAGVTELNRASRWPVWASLQLY
A-PSY1 RSFMKNQIKRARMFFQQAEAGVTELNRASRWPVWASLQLY
Ladyfinger-PSY2 RSFMKGQIRRARMFFEEAEKGIYELNSASRWPVLASLLLY
Ma-PSY2b RSFMKGQIRRARMFFEEAEKGIYELNSASRWPVLASLLLY
Lakatan-PSY2 RTFMKGQITRARMFFDEAEKGIYELNSASRWPVLASLLLY
Ma-PSY2a RTFMKGQITRARMFFDEAEKGIYELNSASRWPVLASLLLY
Wain-PSY2 RTFMKGQITRARMFFDEAEKGIYELNSASRWPVLASLLLY
A-PSY2a RTFMKGQITRARMFFDEAEKGIYELNSASRWPVLASLLLY
.....170.....180.....190.....200

Fig. S2


(a)
A-PSY1
AUG
UGA
430 AA
Ma-PSY1
AUG
UGA
430 AA
A-PSY2a
AUG
UAA
394 AA
Ma-PSY2a
UAA UAA
500 bp
(b)
A-PSY1
457 51 173 236 193 183
81 461 79 79 73
500bp
Ma-PSY1
457 51 173 236 193 183
81 466 79 80 74
Zm-PSY1
406 51 173 236 193 174
98 613 885 94 77
A-PSY2a
358 51 173 236 193 174
324 332 75 92 166
Ma-PSY2a
367 51 173 236 193 174
326 320 79 91 97
Zm-PSY2
406 51 173 236 193 150
716 100 146 317 131