Theor Appl Genet
DOI 10.1007/s00122-012-2026-0

ORIGINAL PAPER

Natural variation in the sequence of *PSY1* and frequency of favorable polymorphisms among tropical and temperate maize germplasm

Zhiyuan Fu · Yuchao Chai · Yi Zhou · Xiaohong Yang · Marilyn L. Warburton · Shutu Xu · Ye Cai · Dalong Zhang · Jiansheng Li · Jianbing Yan

Received: 22 July 2012 / Accepted: 28 November 2012

Abstract Provitamin A (Pro-VA) is necessary for human vision and immune system health, especially in growing children. The first committed step in the maize carotenoid biosynthesis pathway is catalyzed by phytoene synthase 1 (encoded by *PSY1*) which controls the flux of substrates into the pathway. The flow of these substrates could be directed into production of the β-branch carotenoids (the step controlled largely by the *lycopene epsilon cyclase* gene), but terminated after the production of β-carotene, rather than allowing it to be converted into the next metabolite (the step controlled largely by the *β-carotenoid hydroxylase* gene). In this study, *PSY1* was subjected to association mapping in two diverse maize populations, quantitative trait loci (QTL) mapping in one segregating population, and expression analysis of lines polymorphic for sites within *PSY1*. The results indicated that a 378-bp InDel upstream of the transcription start site and a SNP in the fifth exon resulting in a Thr to Asn substitution, explaining 7 and 8 % of the total carotenoid variation, respectively, may be functional sites associated with total carotenoid levels in maize grain. Analysis of the evolution of *PSY1* strongly suggests that there was positive selection for these polymorphic sites after the divergence of yellow maize from white maize.

Communicated by T. Luebberstedt.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-012-2026-0) contains supplementary material, which is available to authorized users.

Z. Fu · Y. Chai · Y. Zhou · X. Yang · S. Xu · Y. Cai · D. Zhang · J. Li
Key Laboratory of Crop Genomics and Genetic Improvement (Ministry of Agriculture), National Maize Improvement Center of China, China Agricultural University, Haidian 100193, Beijing, China

Z. Fu
Zhengzhou National Maize Improvement Sub-center of China, Henan Agricultural University, Zhengzhou 450002, China

M. L. Warburton
USDA-ARS Corn Host Plant Resistance Research Unit, 9555, Mississippi, MS 39762, USA

J. Yan (✉)
National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, Hubei, China
e-mail: yjianbing@mail.hzau.edu.cn

Introduction

In developing countries, 127 million pre-school-aged children suffer from vitamin A deficiency (VAD), and approximately 4.4 million people have developed eye damage due to VAD (http://www.harvestplus.org/vita.html). Dietary diversity, food biofortification, supplements tactics, and crop biofortification have been suggested to solve VAD problem. The first three solutions are expensive and inaccessible for the poverty in developing countries, reducing their efficiency and application (FAMOBIB FAO/Food Nutrition Division). Crop biofortification, a process of breeding crop rich in bioavailable micronutrients combining traditional breeding and molecular marker-assisted strategy, is therefore an economically and socially sound way to address the global challenge including VAD.

Provitamin A (Pro-VA) refers to the carotenoids that can be converted into physiologically activated vitamin A in the human body and includes α-carotene, β-carotene, and β-cryptoxanthin. Moreover, β-carotene is converted to Pro-VA in the human body with twice the efficiency of either

Published online: 13 December 2012

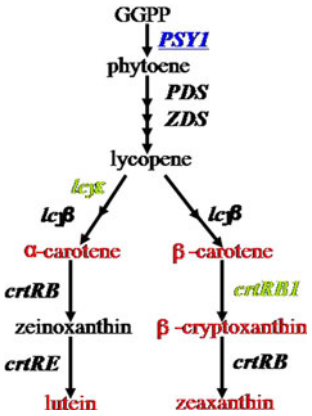

Fig. 1 Major genes and carotenoid components in carotenoid biosynthetic pathway of yellow maize kernels (Harjes et al. 2008; Yan et al. 2010; Buckner et al. 1996). Carotenoid intermediates highlighted in *red* are those detected by HPLC in this study. *GGPP* geranylgeranyl pyrophosphate, *PSY1* phytoene synthase 1, *PDS* phytoene desaturase, *ZDS* ζ-carotene desaturase, *lcyε* lycopene ε-cyclase, *lcyβ* lycopene β-cyclase, *crtRB1* β-carotene hydroxylase 1 (color figure online)

α-carotene or β-cryptoxanthin. Maize, which had a production of 817 million tons in 2009 (http://faostat.fao.org), is the only major cereal crop that can naturally accumulate appreciable levels of Pro-VA in the kernels (Gallagher et al. 2004). Maize has wide genetic variation and allelic diversity for carotenoid components and carotenoid biosynthesis genes (Chander et al. 2008a; Harjes et al. 2008; Yan et al. 2010). Pro-VA biofortification of maize has become a feasible approach to address the challenge of VAD in developing countries (Harjes et al. 2008; Yan et al. 2010).

The genes encoding key enzymes in the carotenoid biosynthesis pathway have been elucidated and cloned in plants (Fig. 1). Geranylgeranyl pyrophosphate (GGPP), which is the competitive substrate for several biosynthetic pathways including carotenoids (Chappell 1995), can be converted to phytoene by phytoene synthase, which is encoded by the phytoene synthase gene (*PSY1*) (Buckner et al. 1990, 1996). Subsequently, lycopene, the first colored component (by virtue of having a chromophore), is generated from phytoene via desaturation by phytoene desaturase (encoded by *PDS*) and ζ-carotene desaturase (encoded by *ZDS*) (Li et al. 1996; Hable et al. 1998; Luo et al. 1999; Matthews et al. 2003). Lycopene is further cyclized by the lycopene epsilon cyclase (encoded by *lcyε*) (Harjes et al. 2008) and/or lycopene beta cyclase (encoded by *lcyβ*) (Singh et al. 2003) to α-carotene and β-carotene. β-carotenoid hydroxylases (encoded by *crtRB*) then catalyze α-carotene and β-carotene to zeinoxanthin/lutein and β-cryptoxanthin/zeaxanthin, respectively.

Downregulation of *lcyε* reduces the ratio of the α-carotene branch to the β-carotene branch; four natural *lcyε* polymorphisms explained 58 % of the phenotypic variation of Pro-VA with a threefold increase compounds between the lines with the lowest and highest carotenoid contents were found in a study by Harjes et al. (2008); meanwhile, downregulation of *crtRB1* favors the accumulation of β-carotene over that of β-cryptoxanthin, leading to higher effective Pro-VA levels, since β-carotene is converted to Pro-VA in the human body with twice the efficiency of either α-carotene or β-cryptoxanthin. Haplotypes of favorable alleles from *crtRB1* and *lcyε* together may promote further accumulation of β-carotene (Yan et al. 2010). Although β-carotene accumulation via *crtRB1* and *lcyε* is currently the primary selection target for Pro-VA biofortification, any potential improvement is limited unless the total flux of GGPP into the carotenoid biosynthesis pathway can be increased.

Carotenoid levels are undetectable in white maize endosperm, but overexpression of the *PSY1* gene in white kernels leads to significant carotenoid accumulation, confirming the essential role of *PSY1* for carotenoid biosynthesis in maize (Zhu et al. 2008). *PSY1* was strongly targeted for selection during maize domestication and improvement from white to yellow kernels (Palaisa et al. 2004), and following this selection bottleneck, there is very little sequence variation in *PSY1* within the yellow maize germplasm. The extent of linkage disequilibrium (LD) in *PSY1* in yellow maize of more than 4 kb is considerably longer than the maize average of about 2 kb (Palaisa et al. 2004; Fu et al. 2010). A quantitative trait locus (QTL) cluster with large phenotypic effects on total carotenoid and individual compounds was mapped to the same genomic region as *PSY1* using segregating populations generated by crossing two yellow maize lines (Chander et al. 2008a; Wong et al. 2004). This implies that allelic variation exists within the sequence of *PSY1* and that this can be exploited to achieve greater flux into the carotenoid pathway.

In this study, we conducted an association analysis using a large number of diverse yellow maize inbred lines consisting of temperate and tropical/subtropical backgrounds. Our objectives were as follows: (1) to explore the natural sequence variation of *PSY1* within yellow maize and determine whether identified variants were associated with total carotenoid levels in maize grain, (2) to detect the feasibility of integrating *PSY1* with *crtRB1* and *lcyε* in marker-assisted improvement of Pro-VA, and (3) to independently validate the most favorable haplotype of these three genes for Pro-VA biofortification in yellow maize.

Materials and methods

Field design and germplasm evaluation

We used two maize association mapping panels with 155 temperate inbred lines (Yang et al. 2010) and 527 diverse

lines (Yang et al. 2011) in this study. The 155 temperate inbred lines were grown in one-row plots with two replications in a randomized complete block design at the Agronomy Farm, China Agricultural University (CAU) during the summers of 2005, 2006, and 2007 (Beijing, China), and the winter of 2007 (Hainan, China). The 527 diverse lines were grown at the Sichuan Agricultural University experimental station at Ya'an, Sichuan, during the summer of 2009 with one replication. Each plot was 4-m long. Rows were spaced 0.67 m apart and plants were grown at a planting density of 45,000 plants/ha; seeds were produced via self-pollination of each plant in the plot. After harvest, we phenotyped 129 yellow temperate lines and 507 yellow diverse lines. The 186 teosinte accessions from the CIMMYT genebank represented a major portion of the diversity present in the teosinte collection. Leaf material from one individual of each accession was used for DNA extraction. Detailed passport information is listed in Table S1. Fresh leaves of 2-week-old plants were harvested and frozen at −70 °C. Leaf material was ground in liquid nitrogen using a mortar and pestle and DNA was extracted using the CTAB method (Murray and Thompson 1980).

Quantification of carotenoids

Multiple self-pollinated ears of each line for each replication were combined for high performance liquid chromatography (HPLC) analysis. Measured metabolites included α-carotene, lutein, β-carotene, β-cryptoxanthin, and zeaxanthin. Carotenoids were extracted as described in Chander et al. (2008b) and quantified by standard regression against external standards (Kurilich and Juvik 1999). External standards were extracted by the method recommended by the HarvestPlus program (http://www.harvestplus.org). Besides the five carotenoid components, we also screened for two derived traits, total carotenoids (lutein + zeaxanthin + β-cryptoxanthin + α-carotene + β-carotene) and Pro-VA (β-carotene + 1/2 (β-cryptoxanthin + α-carotene)) in the two association mapping panels.

Genotyping

Based on previous data (Fu et al. 2010), two regions containing a 5′ InDel and single nucleotide polymorphisms (SNP) within exon regions of the *PSY1* gene were re-sequenced in 129 temperate yellow lines. 30 µl PCR product of each line was sequenced directly using an ABI 3730 sequencer. Alignments were performed by means of the Muscle (Edgar 2004) software package and improved manually in BioEdit (http://www.mbio.ncsu.edu/BioEdit/bioedit) using the B73 reference sequence as a control. The seven polymorphic sites of InDel1, InDel2, InDel4, InDel8, InDel9, SSR7, and SNP7 were converted into simple PCR-based markers to verify the variations in the two association panels.

Transcript profiling

We assayed endosperm tissue at 20 days after pollination (DAP) in 2006 (30 yellow lines) and 2007 (17 yellow lines) for transcript profiling based on previous data (Vallabhaneni and Wurtzel 2009; Li et al. 2008). Maize inbred lines used for *PSY1* expression profiles were grown on the Agronomy Farm, China Agricultural University, Beijing, China. Endosperm collected 20 DAP was dissected and stored at −70 °C until use. Total RNA was isolated using the Plant Total RNA Extraction kit (Bioteke Corporation, China) according to the manufacturer's protocol. The absence of genomic DNA contamination was confirmed by PCR, using primers designed to flank the intron sequences of a control gene (Actin: TGGCATTGTCAACAACTGG, CTCCTTGCTCATACGATCGG). First-strand cDNA was synthesized using oligo (dT)15 as a primer and M-MLV Reverse Transcriptase (Promega, USA) in a 25-µl reaction volume. Approximately 2 µg total RNA and 100 µM oligo (dT)15 were mixed and RNase-free water was added to bring the reaction volume to 10 µl. The mixture was incubated at 70 °C for 5 min and then cooled on ice before adding 15 µl cDNA synthesis mix (containing 5 µl of 10× RT buffer, 5 µl of 2.5 mM dNTP mix, 0.5 µl of rRNasin Ribonuclease Inhibitor (50 U/µl), and 1 µl of MMLV RT (200 U/µl)). The mixture was then incubated for 1 h at 42 °C. Samples were collected by brief centrifugation and diluted eight-fold. Quantitative real-time PCR was performed using SYBR® Premix Ex Taq™ (Takara, Dalian, China) on a DNA Engine Opticon® 2 Continuous Fluorescence detection system (Bio-Rad, Waltham, USA). Reactions contained 1 µl diluted cDNA, 5 µl SYBR® Premix Ex Taq™ (2×), 0.2 µl (10 µM) of each gene-specific primer, and 3.6 µl nuclease-free water in a final volume of 10 µl. Thermal cycling conditions included an initial incubation at 95 °C for 10 s, followed by 40 cycles of 95 °C for 5 s, 60 °C for 20 s, and 72 °C for 30 s. Melt curve analysis was performed to verify primer specificity, and PCR products were confirmed by sequencing. The relative quantity of the transcripts was calculated by the comparative threshold cycle (CT) method. An Actin control was used to normalize data across samples. Each sample was analyzed in triplicate for *PSY1* and Actin, and the average value for each was used in analyses. The sequences of primers for real-time PCR were as follows: *PSY1*, AGCCAACCTAGATGGATACGG, AGTCGCAG GGGGAGCACAAT, Act, CGATTGAGCATGGCATTGT CA, CCCACTAGCGTACAACGAA.

Statistical analyses

Heritability for each trait was calculated as σ_g^2/σ_p^2 where σ_g^2 is the genotypic variance and σ_p^2 is the phenotypic variance (Allard 1960). The genetic parameters were determined based on the following models: $MSe = \sigma_e^2$, $MSgy = \sigma_e^2 + r\sigma_{gy}^2$, $MSg = \sigma_e^2 + r\sigma_{gy}^2 + ry\sigma_g^2$, $\sigma_p^2 = \sigma_g^2 + \sigma_{gy}^2/y + \sigma_e^2/yr$, where MSe is the error mean square; σ_e^2 is the error variance; MSgy is the mean square of genotype by environment; MSg is the mean square of genotype; r is the number of replications in each year; and y is the number of years. The TASSEL software package (Bradbury et al. 2007) was used to determine associations between changes in the DNA sequence of the *PSY1* gene and changes in the measured phenotypes. Population structure and kinship for the two maize populations were estimated using 926 and 1067 SNP markers, respectively (Yang et al. 2011). The software package STRUCTURE 2.1 (Pritchard et al. 2000) was used to estimate the population structure and final K was set to two. Lines with membership probabilities ≥ 0.6 were assigned to subpopulations; others were assigned to admixed groups. The relative kinship matrix was calculated by means of the software package SPAGeDi (Hardy and Vekemans 2002); negative values between individuals were set to zero. TASSEL applies a logistic regression ratio test to calculate the likelihood that either (1) the candidate gene distribution is associated with population structure and phenotypic variation or (2) the candidate gene distribution is associated with population structure only. We used the mixture linear model (MLM) analysis in TASSEL (Bradbury et al. 2007; Yu et al. 2006) to identify associations between polymorphisms and markers, taking population structure and kinship (Yang et al. 2011) into account. The r^2 (Hill and Robertson 1968) linkage disequilibrium values were also calculated. Polymorphisms with a minor allele frequency >2 % were used in the association tests. This threshold was chosen because sample-size requirements increase dramatically when allele frequencies fall below 2 %. To assist in the selection of assigning haplotypes by eye, the polymorphic sites of the whole panel were entered into Excel and sorted based on similarity. Based on these groupings of inferred haplotype, the two-way ANOVA analysis in Statistica Software (StatSoft 1993) and one-way ANOVA analysis in Excel were used to test the significance of the interactions and combinations between genes. ANOVA was conducted based on the model as follows. $y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ij}$, where y_{ij} is the trait performance under allele i of gene A and allele j of gene B (i, $j = 1$ and 2 corresponding to two alleles of target genes in this study); μ is the overall mean of the model; α_i is the effect of gene A under allele i; β_j is the effect of gene B under allele j; γ_{ij} is the interaction effect of gene A under allele i and gene B under allele j; and e_{ij} is the residual random error which is assumed to be normally distributed. P values showed in Table S2 are the significance of γ_{ij} for specific pairs of genes. Where necessary, we used power transformations to transform phenotypic data of lutein, zeaxanthin, β-cryptoxanthin, β-cryptoxanthin, β-carotene, total-, and Pro-VA to a more normal distribution. Statistical inferences are drawn from transformed data, and back-transformed estimated effects are reported.

Results

Gene structure and variation of *PSY1*

The gene structure and sequence of *PSY1* are shown in Fig. 2 and Fig. S1. Two regions covering about 63 % of the full reference *PSY1* gene sequence in B73 were sequenced in 129 temperate yellow lines. We found four major polymorphisms (InDel1, 378 bp; InDel2, 644 bp; InDel4, 390 bp; and SSR7, a CCA trinucleotide repeat) in the 5′UTR, and two (InDel8, 345 bp and InDel9, 18 bp) in the 3′UTR. In addition, three SNPs causing amino acid changes reported in our previous study (Fu et al. 2010) were also identified in the current study. These were located between the fourth to sixth exons; SNP3 (Thr to Ser), SNP5 (Gly to Arg), and SNP7 (Thr to Asn). All sites found to be polymorphic in yellow maize genotypes are shown in Table S3.

We developed simple PCR-based markers for the five InDels, one SSR, and three non-synonymous SNPs described above. These markers were used to validate their associations with carotenoid phenotypes in a larger, diverse population consisting of 527 lines (Yang et al. 2011). The reference marker map is shown in Fig. S2. In yellow lines (507 of the 527 lines), the large InDels 1, 2, and 4 and SNPs3–6 were all in complete LD, and there were only two InDel haplotypes: haplotype 1 (378 bp insertion +644 bp insertion +390 bp deletion), and haplotype 2 (378 bp deletion +644 bp deletion +390 bp insertion). The 20 white endosperm lines did not contain insertions at any of the three InDel sites. Similarly, SNP7, a C to A variant, segregated in yellow lines but was monomorphic for the C state in all white lines. However, SNP3 and SNP5 were polymorphic in both yellow and white lines. There were no differences in haplotype state or frequency of InDel7, InDel8, or InDel9 between yellow and white lines. Based on the LD level (Table 1), InDel1, SSR7, SNP7, InDel8, and InDel9 were identified as the most informative polymorphisms for association mapping analysis.

Fig. 2 Maize *PSY1* gene. Translated exons are depicted as *black boxes* and putative start of transcription (TSS) and polyA sites are indicated. *Gray rectangles* show regions sequenced in 129 temperate *yellow lines*, *red triangles* indicate InDels. Positions of polymorphic sites relative to sequence alignment are indicated numerically with polymorphisms according to *yellow maize line* B73. Polymorphisms found by aligning sequences of 129 temperate *yellow lines* are marked in diagram and those that are significantly associated with changes in measured phenotypes are labeled with *asterisks*. The 5′ insertion/deletion variation includes InDel1, InDel2, and InDel4 (color figure online)

Table 1 r^2 value of informative polymorphic sites on *PSY1* gene in 129 temperate yellow inbreds

r^2	InDel1	SSR7	SNP7	InDel8	InDel9
InDel1					
SSR7	0.31				
SNP7	0.42	0.71			
InDel8	0.02	0.32	0.45		
InDel9	0.33	0.10	0.14	0.01	

Association analysis of *PSY1*

Although there were significant differences among environments and significant genotype × environment interactions for all traits (Table 2), the high broad sense heritability of temperate germplasm (>0.85 for all traits except α-carotene, which was 0.74) suggested that much of the variation associated with the phenotypes of these traits was caused by genotypic differences among the lines. The extensive range of total carotenoid levels (from 1.09 to 31.89 μg/g) and significant genotypic variation further suggested that it is feasible to identify genetic factors and to genetically improve carotenoid content in this temperate germplasm panel.

We measured the associations between the seven informative polymorphic sites and the six traits in the panel of 129 temperate yellow lines over four growing seasons and in a Best Linear Unbiased Predictor (BLUP) analysis (Table 3). Only InDel1 and SNP7 were significantly associated with total carotenoid content in most environments. In the BLUP analysis, InDel1 and SNP7 explained 7 and 8 % of total carotenoid phenotypic variation, respectively (Table 3). These two polymorphisms were also associated with most of the other five carotenoid components. However, the magnitude and significance of their contributions to the traits varied among the different growing seasons, possibly due to environmental effects and/or to different population sizes phenotyped in each growing season (Table 3). SNP7 was more strongly associated with total carotenoid content than was InDel1 in all tests when population structure and kinship were controlled (Table 3). Further functional validation of the two polymorphic sites was determined in another association mapping panel consisting of 507 diverse yellow lines, which included genetically and geographically distinct tropical/sub-tropical and temperate germplasm. There were significant associations between total carotenoid content and InDel1 and SNP7 ($p = 7.78E{-}04$ and 0.0044 in MLM,

Table 2 Combined analysis of variance for major carotenoid components measured in 129 temperate yellow inbreds over four growing seasons

Traits[a] (μg/g)	Joint mean square (df)			Mean (±SE)	Range	σ_g^2	σ_p^2	h_B^2 (%)
	y (3)	*g* (128)	*y***g* (349)					
Lutein	44.99***	53.52***	3.96***	5.34 ± 0.10	0.23–23.57	6.20	6.69	92.6
Zeaxanthin	104.39***	31.84***	1.79***	3.44 ± 0.08	0.13–23.57	3.76	3.98	94.4
β-cryptoxanthin	13.88***	2.38***	0.32***	0.75 ± 0.02	0.04–5.404	0.26	0.30	86.8
α-carotene	0.21***	0.07***	0.02***	0.10 ± 0.00	0.01–1.31	0.01	0.01	73.9
β-carotene	20.55***	4.31***	0.47***	0.76 ± 0.03	0.04–12.90	0.48	0.54	89.2
Total carotenoid	288.61***	122.31***	11.26***	10.39 ± 0.16	1.09–31.89	13.88	15.29	90.8
Pro-VA	33.73***	5.86***	0.69***	1.19 ± 0.04	0.08–13.34	0.65	0.73	88.3

Pro-VA = β-carotene + 1/2 (β-cryptoxanthin + α-carotene)

df degrees of freedom, *g* genotype, *y* season, *SE* standard error, σ_p^2 phenotypic variance, σ_g^2 genotypic variance, h_B^2 (%), heritability in broad sense

*** $p = 0.0001$ significance level

[a] Total carotenoid = lutein + zeaxanthin + β-cryptoxanthin + α-carotene + β-carotene

Table 3 *PSY1* polymorphisms associated with carotenoid traits measured in 129 temperate yellow inbreds over four growing seasons

Polymorphic site[a]	Alleles in series[b]	Traits[c]	BLUP[d] 127	2005[d] 115	2006[d] 122	2007[d] 121	2007H[d] 114	R^{2e}	P^{f}	Average[g] 127	Allelic effect[h] 127
InDel1	**0**/378	Lutein	N.S	N.S	N.S	N.S	N.S	6 %	0.0051	5.442/3.477	0.389/−0.389
		Zeaxanthin	0.0045	0.0065	0.0265	5.83E−04	N.S	N/A	N.S	3.506/2.695	1.037/−1.037
		β-cryptoxanthin	0.0045	0.0486	N.S	0.0285	0.0026	N/A	N.S	0.752/0.620	0.259/−0.259
		β-carotene	N.S	N.S	N.S	N.S	N.S	N/A	N.S	0.842/0.523	0.177/−0.177
		Total carotenoid	0.0087	0.0216	0.0118	0.0052	N.S	7 %	0.0023	10.646/7.463	1.798/−1.798
		Pro-VA	N.S	N.S	N.S	N.S	N.S	N/A	N.S	1.269/0.885	0.295/−0.295
SNP7	**A**/C	Lutein	0.0218	0.0090	N.S	0.0258	N.S	N/A	N.S	5.404/4.621	0.746/−0.746
		Zeaxanthin	3.74E−04	1.56E−04	0.0053	3.74E−05	N.S	5 %	0.0120	3.660/2.601	0.940/−0.940
		β-cryptoxanthin	0.005	0.0188	N.S	0.0166	0.0376	5 %	0.0102	0.802/0.521	0.187/−0.187
		β-carotene	0.0454	0.0032	N.S	N.S	N.S	6 %	0.0048	0.915/0.444	0.239/−0.239
		Total carotenoid	5.80E−06	1.42E−04	5.93E−04	6.15E−06	0.0194	8 %	0.0009	10.885/8.313	2.210/−2.210
		Pro-VA	0.0104	0.0070	N.S	0.0300	N.S	9 %	0.0008	1.367/0.756	0.327/−0.327

[a] Only polymorphic sites significantly associated over at least three of four seasons and for BLUP phenotype of the four seasons are shown

[b] Alleles in series are listed for each polymorphism, and favorable alleles (higher total carotenoids) are in bold: InDel1 insertion = 378 bp

[c] Total carotenoid = β-carotene + β-cryptoxanthin + α-carotene + zeaxanthin + lutein

Pro-VA = β-carotene + 1/2(α-carotene + β-cryptoxanthin)

[d] *p*-value from association analysis conducted using the mixture model incorporating population structure and kinship, using data from 4 different seasons. BLUP is the integration of the phenotype in all four seasons. 2007H = Hainan 2007; all other environments are in Beijing

[e] *p*-value from ANOVA analysis using BLUP data

[f] R^2 values from ANOVA showing % phenotypic variation explained using BLUP data

[g] Average of each genotype class for different trait using BLUP data

[h] Allelic effect of the polymorphic sites using BLUP data

respectively), each explaining 5 % of the total phenotypic variation in carotenoid content in this panel. These data support the argument that these polymorphisms are associated with total carotenoid accumulation (Table 4).

Linkage and expression analysis of *PSY1*

In a previous study, a linkage mapping population was created by crossing two yellow parental inbred lines, B73 and By804. Phenotyping of the carotenoid traits in this population identified one QTL cluster that co-located with the *PSY1* gene on chromosome 6 near SSR7 (Chander et al. 2008a). We used this linkage mapping population to further validate the association mapping results described above. Of the two possible functional polymorphic sites of InDel1 and SNP7 suggested by association results, only SNP7 was polymorphic between B73 and By804 (Fig. S3). The SNP7 allele from By804 was found to be the favorable allele, suggesting that the change from C to A, and thus, the protein change from Thr to Asn at this amino acid, might have a positive effect on increasing the total flux of substrates into the pathway. The favorable allele of SNP7 identified by linkage mapping was the same as that identified in association mapping. QTL and association results thus support the probable role of SNP7 in total carotenoid accumulation; however, these analyses cannot prove it to be the only causal mutation since other undetected SNPs or InDels in other genes within the QTL region may also affect carotenoid accumulation.

To validate the effect of InDel1 in *PSY1* on total carotenoid accumulation, we conducted transcription profiling using two subpopulations, one of which comprised 30 yellow lines monomorphic for no insertion at locus InDel1 (Table S4) and the other comprising 17 yellow lines polymorphic at InDel1 (Table S5). The Pearson correlation coefficients between *PSY1* transcription and total carotenoid levels in mature temperate maize kernels are shown in Fig. 3a, b. The Pearson correlation coefficient was higher in the polymorphic lines ($r = 0.68$, $p = 0.003$) than in the monomorphic lines ($r = 0.51$, $p = 0.004$). The significant correlation in the monomorphic lines indicates that there might be another sequence change that was not detected in this study influences carotenoid levels as well. In addition, the absence of the insertion (378 bp) at this locus resulted in a nearly 6-fold change in *PSY1* expression (Fig. 3c; Table S5) and a 2.6-fold change in total carotenoid content (Table S5). To a certain extent, these results validate an upregulating function of *PSY1* on carotenoid accumulation with which the 378-bp deletion appears to interfere.

Analysis of the evolution of polymorphic sites in *PSY1*

To test the change in the function of *PSY1* associated with InDel1, InDel2, InDel4, and SNP7 during maize evolution, we genotyped one individual from each of 186 teosinte accessions in a global collection (Table S1) to determine the original state prior to the mutations causing the polymorphisms seen in this study (data not shown). The same haplotypes found in white maize lines were found in the teosintes, but we did not find any haplotypes in the teosintes that were associated only with yellow maize lines. This suggests that selection at these sites did not occur during domestication, as there were no variants in the diverse teosinte sequences tested here that could have been selected by ancient farmers. Rather, the mutation and subsequent selection probably happened during improvement of maize landraces. According to the variation in haplotypes of InDel1, InDel2, InDel4, and SNP7 seen between white and yellow lines, as well as among yellow lines, the large insertion upstream of the translational start site might have been the critical mutation for carotenoid accumulation in yellow lines.

Frequency of favorable alleles for *PSY1*, *lcyε*, and *crtRB1* in current yellow germplasm

Favorable alleles of *PSY1* were defined as the polymorphisms associated with higher total carotenoid content. The 378-bp deletion and "A" variant represented the favorable alleles for InDel1 and SNP7, respectively. In our genotype analyses, we observed very high frequencies of these alleles in the 278 tropical/subtropical yellow lines (100 % for InDel1 and 97 % for SNP7) and the 226 temperate yellow lines (92 % for InDel1 and 77 % for SNP7) (Table 5). The relatively lower frequency of SNP7 in the temperate germplasm (77 %) leaves room for further improvement of carotenoid levels in this population. Frequencies of *lcyε* and *crtRB1* in this germplasm showed a different pattern, as the positive alleles of both these genes were rare (4 % or less) in different yellow germplasm lines (Table 5). In addition, the best combination of haplotypes of the three genes was not identified in any of the 507 yellow lines surveyed, suggesting that there is still large potential to increase the Pro-VA content in maize grain by combining the best alleles of these genes (Table 5).

Combined effects of *PSY1*, *lcyε*, and *crtRB1* on carotenoids

We conducted two-way ANOVA to detect epistatic interactions between *PSY1*, *lcyε*, and *crtRB1*, analyzing two traits at a time. Since the dominant effects derive from heterogenous alleles that are absent for these three genes in the association population, we surveyed only the additive × additive interaction in these analyses. We detected an additive × additive interaction between *PSY1* and *lcyε* for β-cryptoxanthin ($p = 0.040$), ProVA ($p = 0.001$), and

Table 4 Association analysis results of polymorphisms in *PSY1* with carotenoid traits in 507 diverse yellow inbreds grown in one environment

Polymorphic site[a]	Alleles in series[b]	Traits[c]	2009[d] 381	R^{2e}	Average[f] 381	Allelic effect[g] 381
InDel1	**0**/378	Lutein	0.0179	1 %	7.755/5.531	1.527/−1.527
		Zeaxanthin	0.0036	3 %	5.110/2.194	1.444/−1.444
		β-cryptoxanthin	N.S	N/A	2.602/0.852	0.442/−0.442
		Total carotenoid	7.78E−04	5 %	16.981/9.110	3.611/−3.611
		Pro-VA	N.S	N/A	2.713/0.873	0.405/−0.405
SNP7	**A**/C	Lutein	N.S	N/A	7.641/8.059	0.568/−0.568
		Zeaxanthin	0.0033	5 %	5.279/3.088	0.891/−0.891
		β-cryptoxanthin	N.S	N/A	2.746/1.032	0.291/−0.291
		Total carotenoid	0.0044	5 %	17.278/12.732	1.871/−1.871
		Pro-VA	0.0148	14 %	2.893/0.967	0.367/−0.367

[a] Only polymorphic sites significantly associated with the trait are shown

[b] Alleles in series are listed for each polymorphism, and favorable alleles (higher total carotenoid) are in bold: InDel1 insertion = 378 bp

[c] Total carotenoid = β-carotene + β-cryptoxanthin + α-carotene + zeaxanthin + lutein

Pro-VA = β-carotene + 1/2(α-carotene + β-cryptoxanthin)

[d] *p*-value from association analysis conducted using the MLM model

[e] R^2 values from ANOVA showing % phenotypic variation explained

[f] Average of each genotype class for different trait

[g] Allelic effect of the polymorphic sites using BLUP data

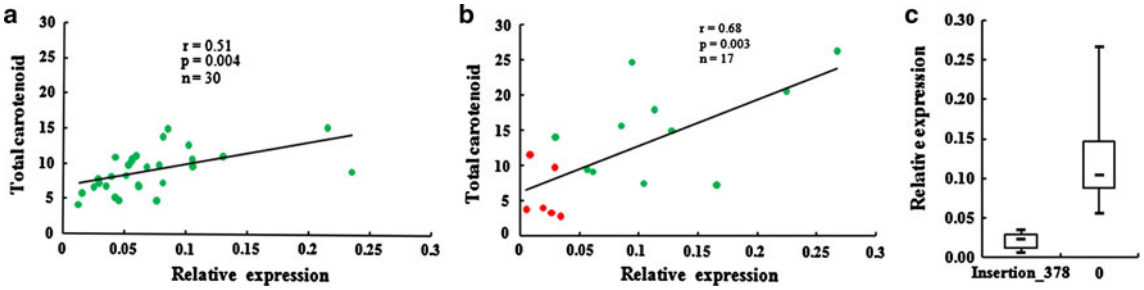

Fig. 3 Relative expression profile of *PSY1* in endosperm of 20 DAP. **a** Correlation between *PSY1* expression level and total carotenoid content (μg/g) in mature kernels of 30 *yellow lines* monomorphic for deletion in InDel1. **b** Correlation between *PSY1* expression level and total carotenoid content (μg/g) in mature kernels of 17 *yellow lines* carrying different alleles for InDel1. *Green rhombus* indicates *lines* with insertion of InDel. **c** Differential expression of 17 *yellow lines* between favorable allele (0 bp insertion) and unfavorable allele (378 bp insertion) of InDel1 showing 5.9-fold change in expression level of *PSY1*. *Error bars* show maximum and minimum values (color figure online)

total carotenoids ($p = 0.010$) (Table S2), suggesting that the function of either of these genes may be enhanced in the presence of the other to increase Pro-VA and total carotenoid levels. We detected a significant additive × additive interaction between *lcyε* and *crtRB1* only for zeaxanthin ($p = 7.96E{-}05$). No significant interactions were identified between *PSY1* and *crtRB1*. Yan et al. (2010) found no epistatic interaction between *lcyε* and *crtRB1,* only a simple additive accumulation, which is consistent with our current results.

The significantly associated polymorphic sites in *lcyε* (Harjes et al. 2008) and *crtRB1* (Yan et al. 2010) were also screened in the present study in the 507 yellow lines of the diverse association mapping population (Table S6, S7). Significant association was verified for all polymorphic sites of *lcyε* and *crtRB1* so we performed a haplotype combination analysis with *PSY1*, *lcyε*, and *crtRB1* simultaneously. Because of LD within each of the genes (Harjes et al. 2008; Yan et al. 2010; Fu et al. 2010), only a limited number of all possible haplotype classes were identified when characterizing all significantly associated polymorphisms for each gene at the same time. In addition, because of the frequency of favorable alleles (Table 5), intragenic

Table 5 Comparison of favorable allele frequencies in different populations (%)

Panel	N	*PSY1*[a]		*lcyε*[a]			*crtRB1*[a]			Observed best favorable haplotype[b]			Best haplotype in theory for three genes[c]
		InDel1	SNP7	5′InDel	SNP216	3′InDel	5′TE	InDel4	3′TE	*PSY1*	*lcyε*	*crtRB1*	
Temperate	226	92.1	77.0	5.2	48.1	12.5	1.4	3.0	13.9	77.0	4.0	0.9	0.0
Tropical/sub-tropical	278	100.0	97.0	3.0	84.0	45.0	0.0	8.4	7.7	97.0	1.5	0.0	0.0

[a] The most favorable allele for each polymorphism is described as high total carotenoid for *PSY1*, low (α-carotene + lutein) over (β-carotene + β-cryptoxanthin + zeaxanthin) for *lcyε*, and high carotene for *crtRB1*

[b] Denotes population percentage of best theoretic haplotype consisting entirely of favorable alleles for each gene

[c] Denotes population percentage of best theoretic haplotype combination for *PSY1*, *lcyε* and *crtRB1*

combination effects of *PSY1* (Table 6), *lcyε*, and *crtRB1* (data not shown), and significant associations (Table 5; S6, S7), we could only use three polymorphic sites (SNP7 in *PSY1*, SNP216 in *lcyε*, and 3′TE in *crtRB1*) for haplotype analysis in the 507 diverse yellow lines (Tables 4; S6, S7). The combined effects were associated with a 10.5-fold change in β-carotene levels ($p = 1.29E{-}23$) in the six lines with the most favorable haplotypes, compared with those lacking them (Table 7).

Discussion

The importance of the *PSY1* gene in carotenoid biosynthesis has been validated in previous studies (Zhu et al. 2008; Paine et al. 2005). However, important questions about its mode of action and its interaction with other genes in the carotenoid biosynthetic pathway are still unanswered. In this study, we used association analysis to identify two significant polymorphic sites in *PSY1* (InDel1 and SNP7) with predicted large effects on total carotenoid levels. We validated the association results of *PSY1* polymorphisms using linkage-based QTL ananlysis (Fig. S3) and the functional basis by expression analysis (Fig. 3). The upregulation of *PSY1* might be mediated through haplotypes of InDels at the 5′UTR, as allelic differences in these InDels showed a 5.9-fold change in transcript expression levels, corresponding to a 2.6-fold difference in total carotenoid levels. We further confirmed the positive correlation between expression of *PSY1* and total carotenoid levels in mature kernels reported in a previous study (Vallabhaneni and Wurtzel 2009). The significant positive correlation between *PSY1* expression and total carotenoid content in yellow lines lacking the polymorphic InDel1 suggests that there are additional functional variations outside of *PSY1* that have yet to be uncovered (Fig. 3a).

A QTL with large effects on total carotenoid and carotenoid component levels was identified that co-located with *PSY1* in a mapping population segregating for SNP7. One possible sequence of evolution of these functional polymorphisms is presented here, as suggested by *PSY1* sequence data from maize compared with that of its ancestral species, teosinte. Teosinte contains only the same haplotypes found in white maize lines at SNP7 and InDels 1, 2, and 4; these are all the unfavorable alleles for carotenoid accumulation. The InDels and SNP7 are all polymorphic in yellow maize lines, supporting the hypothesis that mutations created functional variation at these sites following the divergence of maize from teosinte, and most likely occurred rather late in the selection history of maize. Similar results have also been reported in other studies (Palaisa et al. 2004; Mangelsdorf and Fraps 1931). In addition, the data reported here and in the study of Palasia et al. (2004) indicate that the selective sweep that occurred in *PSY1* during the development of yellow maize races was fairly weak, suggesting that sequence variation for this trait remains to be exploited to obtain even higher levels of Pro-VA carotenoids in yellow maize.

Several factors, including genetic drift, foundation effects, bottleneck effects, mutation, non-random mating, and selection, will influence the frequency of favorable alleles of different genes in distinct populations. However, we suggest that these three genes influencing carotenoid components are important for human health. Because selection for each one could lead to colored maize, we further suggest that different selection strengths for each gene during maize improvement can account for the different frequencies of the favorable alleles in different populations. It is difficult to detect higher carotenoid levels beyond what is visible in the change from white to yellow kernels. Therefore, it would have been difficult to pyramid favorable alleles of all three genes based on phenotype. With the development of PCR-based markers for these genes and an understanding of the different frequencies of favorable alleles for each gene among different germplasm, breeders can now use targeted strategies to improve the trait in any germplasm pool. In tropical/subtropical germplasm, strong selection for *PSY1* produced a very high

Table 6 *PSY1* haplotype estimated effects in 507 diverse yellow lines

Haplotype[a]	*PSY1*			Traits[b] (μg/g)										
InDel1, SNP7	InDel1	SNP7	N	Lutein	SD	Zeaxanthin	SD	β-cryptoxanthin	SD	β-carotene	SD	Total carotenoid	SD	Pro-VA	SD
378, C	378	C	15	5.53	3.22	2.19	1.53	0.85	0.43	0.36	0.23	9.11	3.95	0.87	0.34
0, C	**0**	C	32	9.24	3.6	3.51	1.92	1.12	0.95	0.34	0.28	14.43	5.11	1.01	0.71
0, **A**	**0**	**A**	333	7.64	4.53	5.28	3.18	2.75	1.87	1.39	1.08	17.28	6.72	2.89	1.72
R^2				2 %		6 %		9 %		10 %		7 %		14 %	
P(ANOVA)[c]			380	0.02/0.053		1.28E−05/0.002		9.98E−09/1.61E−6		1.64E−09/1.01E−7		2.39E−06/0.020		2.14E−12/2.16E−9	
Maximum change[*]				1.4		2.4		3.2		3.8		1.9		3.3	

[*] Maximum change for total carotenoid is between the best haplotype "0A" and the worst haplotype "378C"

[a] Haplotype is shown as linear combination of InDel1 allele and SNP7 allele. Favorable alleles are in bold, and only observed haplotypes are listed

[b] Total carotenoid = β-carotene + β-cryptoxanthin + α-carotene + zeaxanthin + lutein Pro-VA = β-carotene + 1/2(α-carotene + β-cryptoxanthin)

[c] *P* value of haplotype "378C", "0C", and "0A"/"0A" and "0C"

frequency of alleles favoring high total carotenoid levels. Favorable alleles of *lcyε* and *crtRB1*, however, are very rare in this germplasm and may be targeted for immediate Pro-VA improvement. In temperate germplasm, there is the potential to increase the frequency of favorable alleles at all three loci to increase Pro-VA content.

The *PSY1*, *lcyε*, and *crtRB1* genes tested in this study affect Pro-VA levels and composition differently. The high level of natural variation for Pro-VA components largely relates to the upregulation of *PSY1* (Fu et al. 2010), but downregulation of *lcyε* and *crtRB1* (Yan et al. 2010), which coincide with the biologic functions of these genes. Combining favorable alleles for *lcyε* and *crtRB1* will increase the content of β-carotene at the expense of other carotenoid components, whereas *PSY1* can increase β-carotene content by increasing the amount of substrate flowing into the carotenoid biosynthesis pathway. The results presented here and in previous studies (Harjes et al. 2008; Yan et al. 2010) highlight the potential to integrate these three genes in a Pro-VA biofortification program by selecting maize varieties with favorable *PSY1* alleles (associated with InDel1 and SNP7) to increase the flux of metabolites into the pathway (increasing potential carotenoid levels) in temperate germplasms; followed by shunting more of the metabolites toward lycopene and the β-carotenoid branch by selecting *lcyε*-favorable alleles (5′InDel, SNP216, and 3′InDel); finally, decreasing transformation of β-carotenoid to downstream metabolites by selecting favorable alleles of *crtRB1* (5′TE, InDel4, and 3′TE). However, the favorable alleles of *PSY1* are almost fixed in the tropical background used in this study that implied we should focus on the other two genes (*lcyε* and *crtRB1*) in the future for high Pro-vitamin A maize breeding the tropical region.

The identification of donor lines in this study (Table S3) and the design of user-friendly PCR-based markers (Fig. S2) for *PSY1*, *lcyε*, and *crtRB1* will be useful for allele-specific marker-assisted selection in breeding programs targeting developing countries, the populations of which are at risk of VAD. Knowledge of the entire pathway and an understanding of the key genes at each step in the pathway, gained via physiologic and genetic studies such as this one, now allow manipulation of the pathway to create maize grain with higher levels of the desired components. In the case of VAD prevention, the target is to increase Pro-VA content. In many developed countries, however, macular degeneration of the retina is a bigger concern, especially among the older population. Macular degeneration is exacerbated by a deficiency of lutein and zeaxanthin (Bone et al. 1988; Handelman et al. 1988). Further research on the genes and allelic diversity in the carotenoid pathway will determine the best alleles for high production of these compounds as well, which may be used

Table 7 Estimated effects of combined haplotypes of *PSY1*, *lcyε,* and *crtRB1* in 507 diverse yellow lines

Haplotype[a]	*PSY1*	*lcyε*	*CrtR-B1*	N	Traits[b] (μg/g)															
PSY1, lcyε, crtRB1	SNP7	SNP216	3'TE		Lutein	SD	Zeaxanthin	SD	β-carotene	SD	Total carotenoid	SD	Pro-VA	SD	βC/βCX	SD	βC/ALL	SD	Ln(Ratio)	SD
A, G, 1	**A**	**G**	**1**	6	5.06	1.89	3.08	1.63	2.46	0.77	11.76	1.07	3.43	0.81	2.05	2.12	0.23	0.13	−0.32	0.61
A, **G** 2	**A**	**G**	2	25	7.27	2.83	5.73	2.51	0.99	0.82	16.86	5.18	2.43	1.63	0.41	0.22	0.05	0.03	−0.17	0.71
A, G, 3	**A**	**G**	3	28	6.17	3.36	5.96	3.45	1.57	0.91	17.27	6.72	3.35	1.64	0.51	0.31	0.10	0.06	−0.57	0.64
C, **G, 1**	C	**G**	**1**	3	7.12	2.21	1.78	0.30	0.65	0.22	10.24	1.73	0.99	0.29	1.39	0.35	0.07	0.03	0.92	0.47
C**, G**, 2	C	**G**	2	8	6.17	3.09	3.36	2.00	0.26	0.11	10.86	4.29	0.79	0.42	0.36	0.16	0.03	0.01	0.34	0.71
C, **G**, 3	C	**G**	3	12	5.29	3.49	2.95	2.91	0.35	0.22	9.99	6.64	1.05	0.67	0.31	0.12	0.05	0.04	0.17	0.63
A, T, **1**	**A**	T	**1**	14	10.04	3.87	1.90	0.80	1.38	1.30	15.21	5.36	2.32	1.26	1.92	3.19	0.09	0.07	0.92	0.37
A, T, 2	**A**	T	2	27	12.97	5.74	4.51	2.15	0.56	0.53	20.03	7.80	1.55	1.08	0.34	0.17	0.03	0.02	0.76	0.57
A, T, 3	**A**	T	3	29	11.20	5.36	4.61	2.34	0.81	0.61	18.96	7.35	1.98	1.35	0.42	0.18	0.04	0.03	0.46	0.73
C, T, 2	C	T	2	18	10.48	2.63	3.32	1.16	0.23	0.17	15.17	3.52	0.80	0.47	0.30	0.15	0.02	0.01	0.93	0.45
C, T, 3/**1**	C	T	3/**1**	4	10.23	5.42	2.72	1.09	0.55	0.25	14.73	6.73	1.16	0.39	0.71	0.37	0.04	0.02	0.92	0.31
R^2					30		17		31		12		28		23		37		46	
P(ANOVA)				371	5.27E−23		8.57E−11		1.29E−23		1.01E−06		4.76E−21		1.07E−15		7.24E−31		4.88E−43	
Maximum change*					2.6		3.3		10.5		2.0		4.3		6.6		15.3		4.9	

* Maximum change for each carotenoid component is between the best haplotype and the worst haplotype

[a] Haplotype is shown as linear combination of three alleles. Favorable alleles are in bold, and only observed haplotypes are listed

[b] Total carotenoid = β-carotene + β-cryptoxanthin + α-carotene + zeaxanthin + lutein Pro-VA = β-carotene + 1/2(α-carotene + β-cryptoxanthin) βC/βCX = β-carotene over β-cryptoxanthin βC/ALL = β-carotene over total carotenoid ln(Ratio) = ln((α-carotene + lutein) over (β-carotene + β-cryptoxanthin + zeaxanthin))

to develop maize specifically targeted to slow down macular degeneration.

Acknowledgments We thank Drs. Brent Buckner and Sherry Flint-Garcia for their useful comments and suggestions. This research was supported by the National Natural Science Foundation of China (30821140352 and 31000715), the National Hi-Tech Research and Development Program of China (2012AA10A307), and the Harvest Plus Challenge Program. The funders had no role in study design, data collection and analysis, decision to publish, or the preparation of the manuscript.

References

Allard RW (1960) Principles of plant breeding, 1st edn. Willey, New York, p 36

Bone RA, Landrum JT, Fernandez L, Tarsist SL (1988) Analysis of the macular pigment by HPLC: retinal distribution and age study. Invest Ophthalmol Vis Sci 29:843–849

Bradbury PJ, Zhang ZW, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

Buckner B, Kelson TL, Robertson DS (1990) Cloning of the y1 locus of maize, a gene involved in the biosynthesis of carotenoids. Plant Cell 2:867–876

Buckner B, San Miguel P, Janick-Buckner D, Bennetzen JL (1996) The y1 gene of maize codes for phytoene synthase. Genetics 143:479–488

Chander S, Guo YQ, Yang XH, Zhang J, Lu XQ, Yan JB, Song TM, Rocheford TR, Li JS (2008a) Using molecular markers to identify two major loci controlling carotenoid contents in maize grain. Theor Appl Genet 16:223–233

Chander S, Meng YJ, Zhang YR, Yan JB, Li JS (2008b) Comparison of nutritional traits variability in selected eighty-seven inbreds from Chinese maize (*Zea mays* L.) germplasm. J Agr Food Chem 56:6506–6511

Chappell J (1995) Biochemistry and molecular biology of the isoprenoid biosynthetic pathway in plants. Ann Rev Plant Physiol Plant Mol Bio 46:521–547

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl Acid Res 32:1792–1797

Fu ZY, Yan JB, Zheng YP, Warburton ML, Crouch JH, Li JS (2010) Nucleotide diversity and molecular evolution of the PSY1 gene in Zea mays compared to some other grass species. Theor Appl Genet 120:709–720

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

Hable WE, Oishi KK, Schumaker KS (1998) Viviparous-5 encodes phytoene desaturase, an enzyme essential for abscisic acid (aba) accumulation and seed development in maize. Mol Gen Genet 257:167–176

Handelman GJ, Dratz EA, Reay CC, Van Kujik FJGM (1988) Carotenoids in the human macula and whole retina. Invest Ophthalmol Vis Sci 29:850–855

Hardy OJ, Vekemans X (2002) SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol Notes 2:618–620

Harjes CE, Rocheford T, Bai L, Brutnell T, Kandianis CB, Sowinski SG, Stapleton AE, Vallabhaneni R, Williams M, Wurtzel ET, Yan JB, Buckler ES (2008) Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. Science 319:330–333

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38:226–231

Kurilich AC, Juvik JA (1999) Simultaneous quantification of carotenoids and tocopherols in corn kernel extracts by HPLC. J Liq Chromatogr Relat Technol l22:2925–2934

Li ZH, Matthews PD, Burr B, Wurtzel ET (1996) Cloning and characterization of a maize cDNA encoding phytoene desaturase, an enzyme of the carotenoid biosynthetic pathway. Plant Mol Biol 30:269–279

Li FQ, Vallabhaneni R, Yu J, Rocheford T, Wurtzel ET (2008) The maize phytoene synthase gene family: overlapping roles for carotenogenesis in endosperm, photomorphogenesis, and thermal stress-tolerance. Plant Physiol 147:1334–1346

Luo R, Wurtzel ET (1999) A maize cDNA encoding zeta carotene desaturase. Plant Physiol 120:1206

Mangelsdorf PC, Fraps GS (1931) A direct quantitative relationship between vitamin A in corn and the number of genes for yellow pigmentation. Science 73:241–242

Matthews PD, Luo R, Wurtzel ET (2003) Maize phytoene desaturase and zetacarotene desaturase catalyze a poly-Z desaturation pathway: implications for genetic engineering of carotenoid content among cereal crops. J Exp Botany 54:2215–2230

Murray MG, Thompson WF (1980) Rapid isolation of high-molecular weight plant DNA. Nucl Acid Res 8:4321–4325

Paine JA, Shipton CA, Chaggar S, Howells RM, Kennedy MJ, ernon 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

Palaisa K, Morgante M, Tingey S, Rafalski A (2004) Long-range patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective sweep. Proc Natl Acad Sci USA 101:9885–9890

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945–959

Singh M, Lewis PE, Hardeman K, Bai L, Rose JKC, Mazourek M, Chomet P, Brutnell TP (2003) Activator mutagenesis of the pink scutellum1/viviparous7 locus of maize. Plant Cell 15:874–884

StatSoft (1993) STATISTICA (data analysis software system)

Vallabhaneni R, Wurtzel ET (2009) Timing and biosynthetic potential for carotenoid accumulation in genetically diverse germplasm of maize. Plant Physiol Prev 150:562–572

Wong JC, Lambert RJ, Wurtzel ET (2004) QTL and candidate genes phytoene synthase and carotene desaturase associated with the accumulation of carotenoids in maize. Theor Appl Genet 108:349–359

Yan JB, Kandianis CB, Harjes CE, Bai L, Kim EH, Yang XH, Skinner DJ, Fu ZY, Mitchell S, Li Q, Salas Fernandez MG, Zaharieva M, Babu R, Fu Y, Palacios N, Li JS, DellaPenna D, Brutnell T, Buckler ES, Warburton ML, Rocheford T (2010) Rare genetic variation at *Zea mays* crtRB1 increases β-carotene in maize grain. Nat Genet 42:322–327

Yang XH, Yan JB, Shah T, Warburton M, Li Q, Li L, Gao YF, Chai YC, Fu ZY, Zhou Y, Xu ST, Bai GH, Meng YJ, Zheng YP, Li JS (2010) Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. Theo Appl Genet 121:417–431

Yang XH, Gao SB, Xu ST, Zhang ZX, Prasanna BM, Li L, Li JS, Yan JB (2011) Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. Mol Breeding 28:511–526

Yu JM, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixture-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Zhu CF, Naqvi S, Breitenbach J, Sandmann G, Christou P, Capell T (2008) Combinational genetic transformation generates a library of metabolic phenotypes for the carotenoid pathway in maize. Proc Natl Acad Sci USA 105:18232–18237