

# Timing and Biosynthetic Potential for Carotenoid Accumulation in Genetically Diverse Germplasm of Maize<sup>1</sup>[C][W][OA]

Ratnakar Vallabhaneni and Eleanore T. Wurtzel\*

Department of Biological Sciences, Lehman College, City University of New York, Bronx, New York 10468; and Graduate School and University Center-City University of New York, New York, New York 10016-4309

Enhancement of the carotenoid biosynthetic pathway in food crops benefits human health and adds commercial value of natural food colorants. However, predictable metabolic engineering or breeding is limited by the incomplete understanding of endogenous pathway regulation, including rate-controlling steps and timing of expression in carotenogenic tissues. The grass family (Poaceae) contains major crop staples, including maize (*Zea mays*), wheat (*Triticum aestivum*), rice (*Oryza sativa*), sorghum (*Sorghum bicolor*), and millet (*Pennisetum glaucum*). Maize carotenogenesis was investigated using a novel approach to discover genes encoding limiting biosynthetic steps in the nutritionally targeted seed endosperm. A combination of bioinformatics and cloning were first used to identify and map gene families encoding enzymes in maize and other grasses. These enzymes represented upstream pathways for isopentenyl diphosphate and geranylgeranyl diphosphate synthesis and the downstream carotenoid biosynthetic pathway, including conversion to abscisic acid. A maize germplasm collection was used for statistical testing of the correlation between carotenoid content and candidate gene transcript levels. Multiple pathway bottlenecks for isoprenoid biosynthesis and carotenoid biosynthesis were discovered in specific temporal windows of endosperm development. Transcript levels of paralogs encoding isoprenoid isopentenyl diphosphate and geranylgeranyl diphosphate-producing enzymes, *DXS3*, *DXR*, *HDR*, and *GGPPS1*, were found to positively correlate with endosperm carotenoid content. For carotenoid pathway enzymes, transcript levels for *CrtISO* inversely correlated with seed carotenoid content, as compared with positive correlation of *PSY1* transcripts. Since zeaxanthin epoxidase (*ZEP*) depletes the carotenoid pool in subsequent conversion to abscisic acid, *ZEP* transcripts were examined. Carotenoid accumulation was found to be inversely associated with *ZEP1* and *ZEP2* transcript levels. Extension of the maize results using phylogenetic analysis identified orthologs in other grass species that may serve as potential metabolic engineering targets.

Carotenoids are a complex class of isoprenoid pigments providing nutritional value as provitamin A and nonprovitamin A compounds; their varied colors provide added commercial value as colorants in foods (for review, see Matthews and Wurtzel, 2007). Metabolic engineering of the biosynthetic pathway has been of interest to specifically address vitamin A deficiency in millions of children worldwide (for review, see Giuliano et al., 2008); this serious global health problem could be alleviated by breeding cereal crop staples in the Poaceae (grass family) for elevated levels of provitamin A carotenoids. Engineering high levels of specific carot-

enoid structures requires controlled enhancement of total carotenoid levels (enhancing pathway flux, minimizing degradation, and optimizing sequestration) plus controlled composition for specific pathway end products. While most of the nuclear genes for the plastid-localized pathway are available (Li et al., 2007) and/or can be identified, questions remain about the rate-controlling steps that limit the predictability of metabolic engineering in plants.

The most important cereal crop staples worldwide, such as maize (*Zea mays*), sorghum (*Sorghum bicolor*), wheat (*Triticum aestivum*), and rice (*Oryza sativa*), are all contained within the grass family. We have taken advantage of the resources available for maize, an important crop staple in sub-Saharan Africa and Latin America, to explore those factors that control endosperm carotenogenesis. Ortholog targets in related grass species can be identified using comparative genomics and thereby potentially extend maize studies for predictive metabolic engineering throughout the cereal crops.

Predictable manipulation of the seed carotenoid biosynthetic pathway in diverse maize genotypes necessitates the elucidation of biosynthetic step(s) that control carotenoid accumulation in endosperm tissue (Fig. 1). Extensive studies have implicated phytoene synthase (*PSY*), the first committed enzyme, as rate

<sup>1</sup> This work was supported by the National Institutes of Health (grant nos. S06-GM08225, 1SC1GM081160-01, and 5SC1GM081160-02 to E.T.W.), Professional Staff Congress-City University of New York, and New York State.

\* Corresponding author; e-mail wurtzel@lehman.cuny.edu.

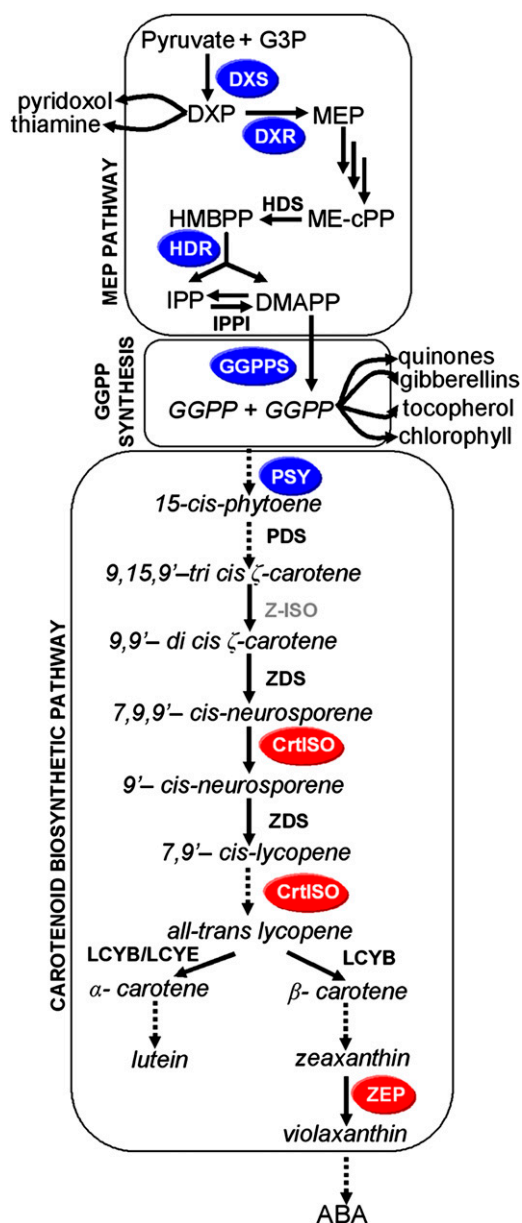
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors ([www.plantphysiol.org](http://www.plantphysiol.org)) is: Eleanore T. Wurtzel (wurtzel@lehman.cuny.edu).

[C] Some figures and tables in this article are displayed in color online but in black and white in the print edition.

[W] The online version of this article contains Web-only data.

[OA] Open Access articles can be viewed online without a subscription.

[www.plantphysiol.org/cgi/doi/10.1104/pp.109.137042](http://www.plantphysiol.org/cgi/doi/10.1104/pp.109.137042)



**Figure 1.** Simplified carotenoid biosynthetic pathway derived from isoprenoid precursors. Enzymatic reactions are represented by arrows. Genes encoding all enzymes shown, except Z-ISO, were tested for correlation with total accumulated carotenoids. Transcripts for enzymes in black letters had no effect; white letters on blue circles and white letters on red circles showed positive correlation and negative correlation, respectively. G3P, D-Glyceraldehyde-3-phosphate; ME-cPP, 2C-methyl-D-erythritol 2,4-cyclodiphosphate; MEP, 2C-methyl-D-erythritol-4-phosphate.

controlling for endosperm carotenoids (Randolph and Hand, 1940; Palaisa et al., 2003; Gallagher et al., 2004; Wong et al., 2004; Poznaniak et al., 2007; Li et al., 2008a, 2008b). However, upstream precursor pathways may also positively influence carotenoid accumulation (Matthews and Wurtzel, 2000; Mahmoud and Croteau, 2001), while downstream degradative pathways may deplete the carotenoid pool (Galpaz et al., 2008). We

decided to utilize natural genetic and biochemical variation in a maize germplasm collection to identify potential transcriptional control points affecting endosperm carotenoid accumulation, since transcriptional control plays a large role in carotenogenesis (Giuliano et al., 2008). An alternative approach to probe the pathway with transgenes was considered less desirable, as the resulting changes would only reflect and be predictive for the particular transformed genotype, and an ultimate goal is to derive regulatory/temporal information applicable to metabolic engineering/breeding of a broader range of genotypes by nontransgenic (Harjes et al., 2008) and/or transgenic means (Zhu et al., 2008). Association of candidate gene cis-located DNA variation with kernel carotenoid levels in a germplasm collection has been effective in identifying metabolic breeding tools for the carotenoid pathway (Harjes et al., 2008). However, association analysis may overlook a rate-controlling biosynthetic step if gene expression is regulated by another locus. Moreover, optimal timing

**Table 1.** Correlation of mRNA levels to total carotenoid content in mature seeds

Pearson correlation ( $r$ ) and statistical significance ( $P$ ), for comparison of transcript levels and kernel carotenoid content in maize inbred lines, was performed using JMP version 5.1.2 (SAS Institute) to test the statistical significance ( $P \leq 0.05$ ) of the relationship. Letters on a white background indicate no effect; letters on blue and red backgrounds indicate positive correlation and negative correlation, respectively. Inbred lines are as listed in Figure 6. Carotenoid content correlated to mRNA levels is shown on a gray background. [See online article for color version of this table.]

Gene	Chromosome Bin	Content $r$ ( $P$ ) Values	
		20 DAP	25 DAP
<i>DXS1</i>	6.05 <sup>a</sup>	-0.2 (0.5)	0.07 (0.84)
<i>DXS2</i>	9.03 <sup>a</sup>	0.45 (0.21)	0.34 (0.36)
<i>DXS3</i>	4.06 <sup>b</sup>	0.20 (0.5)	0.75 (0.01)
<i>DXR</i>	3.04 <sup>a</sup>	0.41 (0.23)	0.61 (0.05)
<i>HDS</i>	5.03 <sup>b</sup>	-0.43 (0.36)	-0.12 (0.73)
<i>HDR</i>	1.09 <sup>b</sup>	-0.17 (0.62)	0.64 (0.04)
<i>IPPI1</i>	7.03 <sup>b</sup>	-0.08 (0.81)	0.27 (0.43)
<i>IPPI2</i>	8.03 <sup>b</sup>	0.21 (0.54)	0.16 (0.66)
<i>IPPI3</i>	6.05 <sup>b</sup>	-0.12 (0.73)	-0.2 (0.49)
<i>GGPPS1</i>	2.08 <sup>b</sup>	0.52 (0.11)	0.69 (0.02)
<i>GGPPS2</i>	7.04 <sup>b</sup>	-0.27 (0.47)	-0.10 (0.76)
<i>GGPPS3</i>	8.01 <sup>b</sup>	NA	NA
<i>PSY1</i>	6.01 <sup>a</sup>	0.96 (0.001)	0.37 (0.28)
<i>PSY2</i>	8.07 <sup>a</sup>	-0.04 (0.89)	0.50 (0.13)
<i>PSY3</i>	7.03 <sup>a</sup>	0.03 (0.93)	0.59 (0.07)
<i>PDS</i>	1.02 <sup>a</sup>	0.12 (0.73)	-0.22 (0.53)
<i>ZDS</i>	7.02 <sup>a</sup>	-0.01 (0.95)	-0.103 (0.77)
<i>CrtISO1</i>	4.08 <sup>b</sup>	-0.76 (0.001)	-0.29 (0.57)
<i>CrtISO2</i> (pseudogene)	2.09 <sup>b</sup>	NA	NA
<i>LCYB</i>	5.04 <sup>c</sup>	0.19 (0.58)	0.60 (0.06)
<i>LCYE</i>	8.05 <sup>a</sup>	0.18 (0.61)	-0.31 (0.37)

<sup>a</sup>MaizeDatabase (www.maizegdb.org). <sup>b</sup>Vallabhaneni and Wurtzel (this article). <sup>c</sup>Singh et al. (2003).

of gene expression is not revealed. Therefore, to identify rate-controlling steps, we focused on transcript-level variation as a manifestation of DNA polymorphisms at the candidate gene or regulation by an unlinked locus. RNA profiling would also facilitate the optimization of temporal expression for transgene expression efforts.

## RESULTS

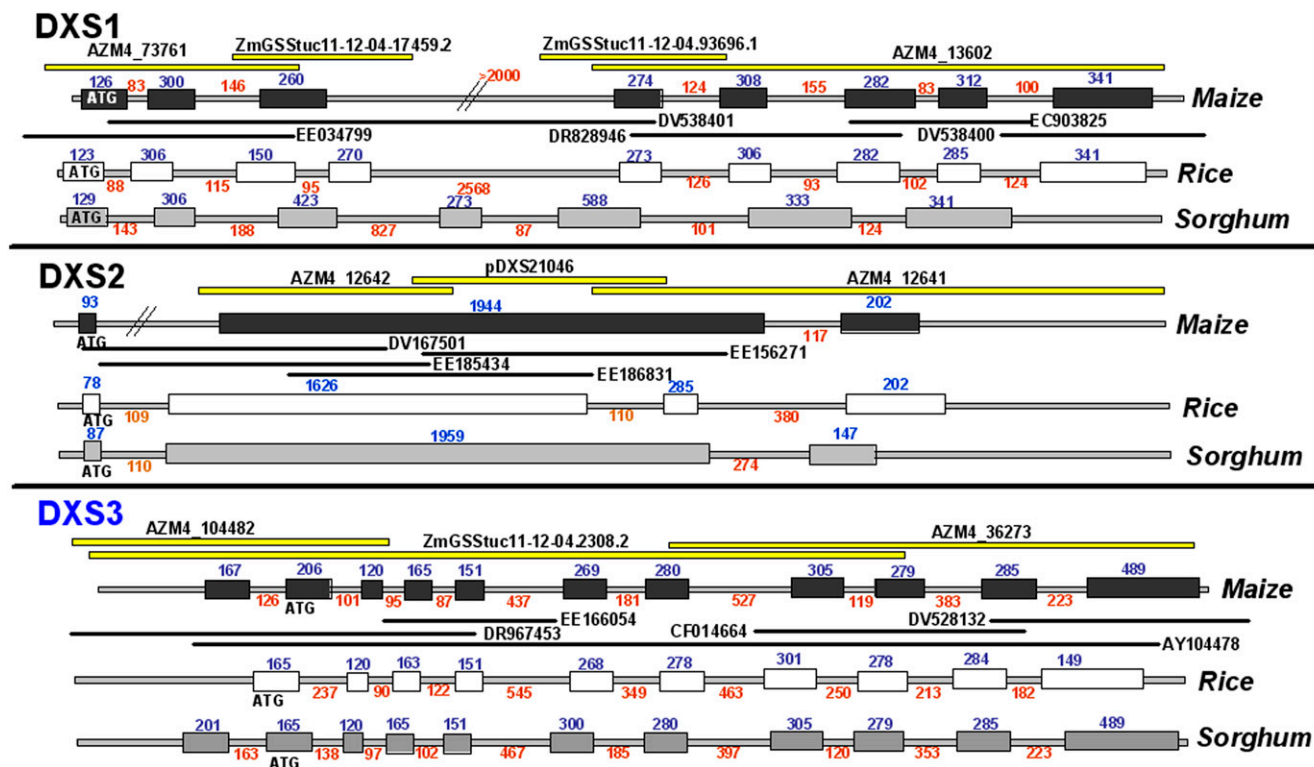
### Pathway Gene Families and Chromosome Mapping

To be able to examine transcriptional control of pathway flux, we assembled a collection of genes from maize and other grass species to represent upstream pathways for isopentenyl diphosphate (IPP) and geranylgeranyl diphosphate (GGPP) synthesis and the downstream carotenoid biosynthetic pathway, including conversion to abscisic acid (ABA). We used a combination of bioinformatics and cloning to identify gene families for enzymes shown in Figure 1. Genes were mapped in maize to chromosome positions with no apparent linkage seen among the pathway genes (Table I; Supplemental Table S1).

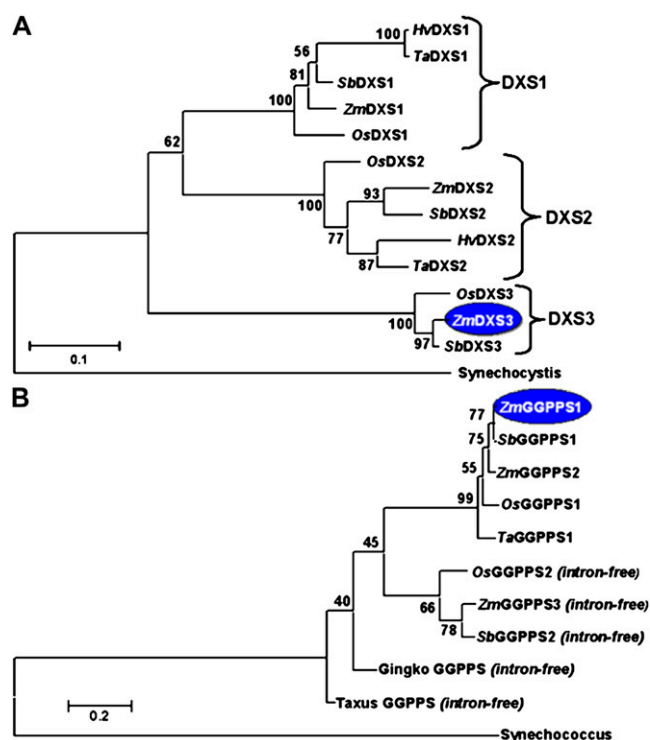
MEP pathway enzymes DXP reductoisomerase (DXR), 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (HDS), and 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (HDR) were found to be encoded

by single-copy genes in the maize genome. Three paralogs for maize DXS (for 1-deoxy-D-xylulose-5-phosphate synthase) were identified (Figs. 2 and 3A; Walter et al., 2002); *DXS3* is a newly discovered paralog in maize and is also present in sorghum and rice, species spanning two subfamilies of the Poaceae (Figs. 2 and 3A; Supplemental Table S1; Kim et al., 2005). The ChloroP program (Emanuelsson et al., 1999) predicted a plastid localization signal in the first 51, 31, and 50 amino acids of *ZmDXS1*, *ZmDXS2*, and *ZmDXS3*, yielding plastid-localized proteins of 71.7, 73.5, and 73.1 kD, respectively. The three maize enzymes are similar in size, yet they share only 52.5% identity/65.8% similarity between DXS1 and DXS3 and 47.6% identity/62.4% similarity between DXS2 and DXS3 as compared with 69.7% identity/79.1% similarity between the more closely related DXS1 and DXS2 (Supplemental Fig. S1).

The IPPI (for isopentenyl pyrophosphate isomerase) gene family contained three genes (Supplemental Table S1). GGPPS (for geranylgeranyl pyrophosphate synthase) was also encoded by a small gene family of three paralogs. The encoded proteins have the requisite GGPPS structural domains and chloroplast targeting signals of 52, 51 and 44 amino acids for GGPPS1, GGPPS2, and GGPPS3, yielding plastid-localized forms of 33.5, 32.9, and 32.7 kD, respectively (Supplemental Fig. S2; Supplemental Table S1; Okada et al., 2000). *GGPPS3* was the only intronless gene, and its



**Figure 2.** Gene family structure of *DXS* genes from the Poaceae. Boxes and lines indicate exons and introns, respectively; sizes are in base pairs. Bars above maize gene structures represent the genomic counterparts used to build the gene structure; dark bars below maize gene structures represent ESTs covering the entire coding region. [See online article for color version of this figure.]



**Figure 3.** Phylogenetic analysis based on translation products. A, *DXS* gene family. GenBank accession numbers are in parentheses: *Synechocystis* (sll1945, BAA17089.1), *Zea mays* (ZmDXS1, TC312721; ZmDXS2, TC293862; ZmDXS3, TC307233), *Oryza sativa* (OsDXS1, LOC\_Os05g33840; OsDXS2, LOC\_Os07g09190; OsDXS3, LOC\_Os06g05100), *Sorghum bicolor* (SbDXS1, Sb09g020140; SbDXS2, Sb02g005380; SbDXS3, Sb10g002960); *Hordeum vulgare* (HvDXS1, TC149768; HvDXS2, TC152494); *Triticum aestivum* (TaDXS1, TC238788; TaDXS2, TC258088). B, *GGPPS* gene family. GenBank accession numbers are in parentheses: *Synechococcus* (crtE, ABI45773); *Zea mays* (ZmGGPPS1, EF417573; ZmGGPPS2, EF417574; ZmGGPPS3, EF417575); *Oryza sativa* (OsGGPPS1, LOC\_Os07g39270; OsGGPPS2, LOC\_Os01g14630); *Sorghum bicolor* (SbGGPPS1, Sb02g037510; SbGGPPS2, Sb03g009380); *Ginkgo biloba* (AY371321); *Taxus media* (AY566309). Maize enzymes are highlighted in blue if transcripts correlated with carotenoid content. [See online article for color version of this figure.]

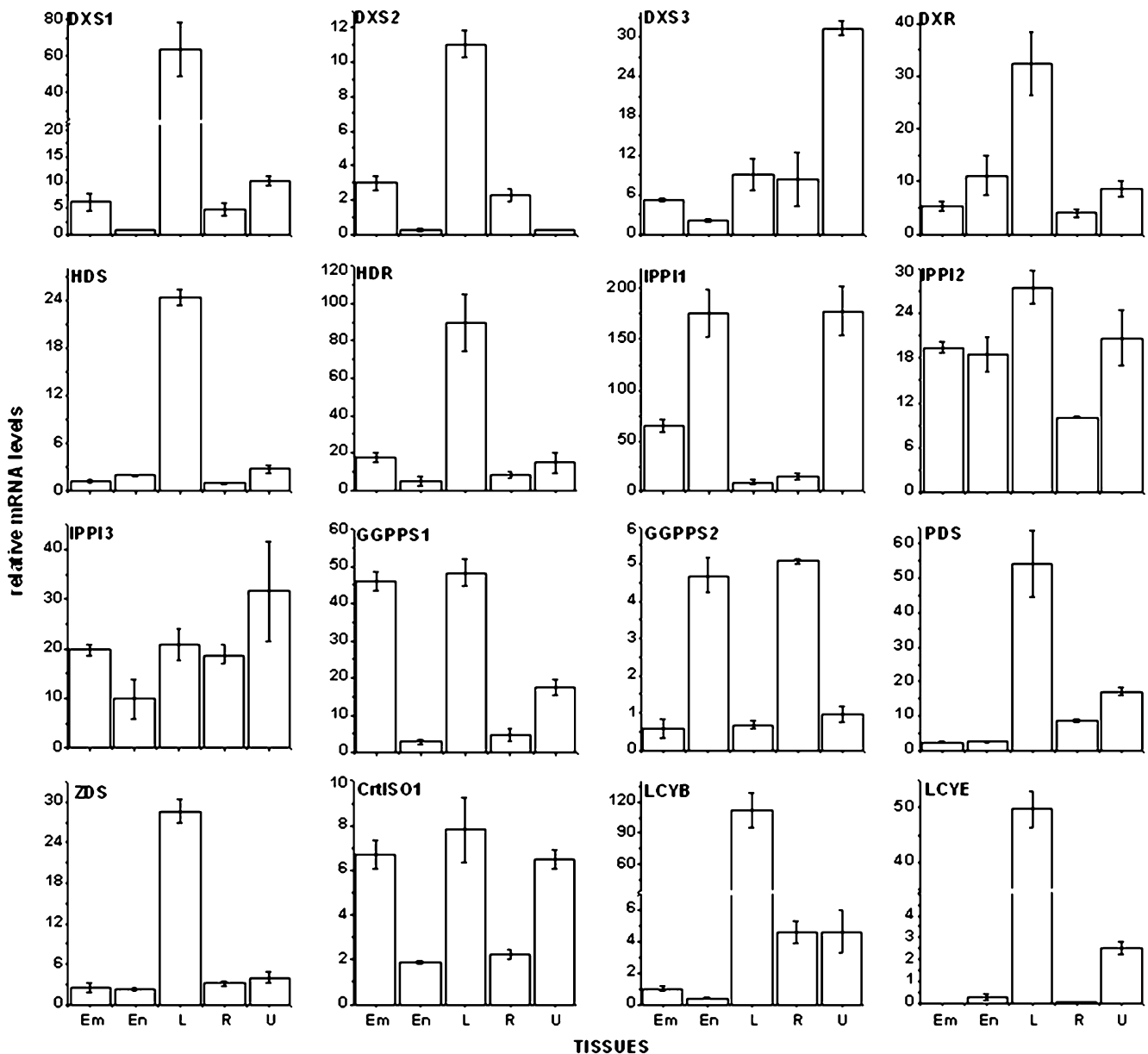
predicted peptide sequence had one variant region immediately downstream of domain III. Phylogenetic analysis showed that the intron-containing *GGPPS* genes and the intronless genes formed two distinct clades within the angiosperms (Fig. 3B). Maize *GGPPS1* and *GGPPS2* appear to be recent duplications arising after speciation. Function of the three *GGPPS* genes was tested by heterologous complementation in *Escherichia coli* (Supplemental Fig. S3) using methods described previously (Cervantes-Cervantes et al., 2006); *GGPPS1* and *GGPPS2* proved functional, but *GGPPS3* did not (data not shown). *GGPPS3* transcripts were barely detectable, as evidenced by reverse transcription (RT)-PCR and as reflected by low EST prevalence in maize and rice. Therefore, our studies limited evaluation to the two functional *GGPPS* genes.

For the carotenoid biosynthetic pathway, we identified two *CrtISO* (for carotene isomerase) genes in maize on chromosomes 2 and 4, adding to previously isolated pathway genes for PDS (for phytoene desaturase), ZDS (for zetacarotene desaturase), LCYE (for lycopene  $\epsilon$ -cyclase), and LCYB (for lycopene  $\beta$ -cyclase), all of which are single copy in maize and other grasses (Buckner et al., 1990; Li et al., 1996, 2008a, 2008b; Matthews et al., 2003; Singh et al., 2003; Harjes et al., 2008; Supplemental Table S1). In contrast to two maize *CrtISO* genes, only one was found in rice, sorghum, and Arabidopsis (*Arabidopsis thaliana*; Supplemental Table S1; Supplemental Fig. S4). However, maize *CrtISO2* proved to be a pseudogene; alignment of maize *CrtISO* sequences (Supplemental Fig. S4; verified from several DNA sources [see "Materials and Methods"]) revealed a stop codon in exon 1 of maize *CrtISO2* that is predicted to interrupt translation. The functional *CrtISO1* encodes an enzyme of 63.6 kD with a predicted plastid-targeting sequence of 43 residues, yielding a plastid-localized form of 59.3 kD. Maize *CrtISO1* is located on chromosome 4 in a region syntenous with rice *CrtISO* and linked to marker *csu704* on rice chromosome 11.

For examination of carotenoid metabolism to ABA, we identified genes for zeaxanthin epoxidase (*ZEP*), an enzyme catalyzing the conversion of zeaxanthin to violaxanthin, a precursor of ABA in endosperm (Fig. 1). Two *ZEP* genes were found in maize compared with only one in rice and sorghum (Supplemental Fig. S5; Supplemental Table S1); all showed gene and protein structural similarities, including chloroplast-targeting signals. Maize *ZEP1* and *ZEP2* have predicted transit peptides of 68 and 60 residues yielding 66.6-kD plastid-localized enzymes.

#### Transcript Profiling of Gene Family Members and Correlation with Carotenoid Content

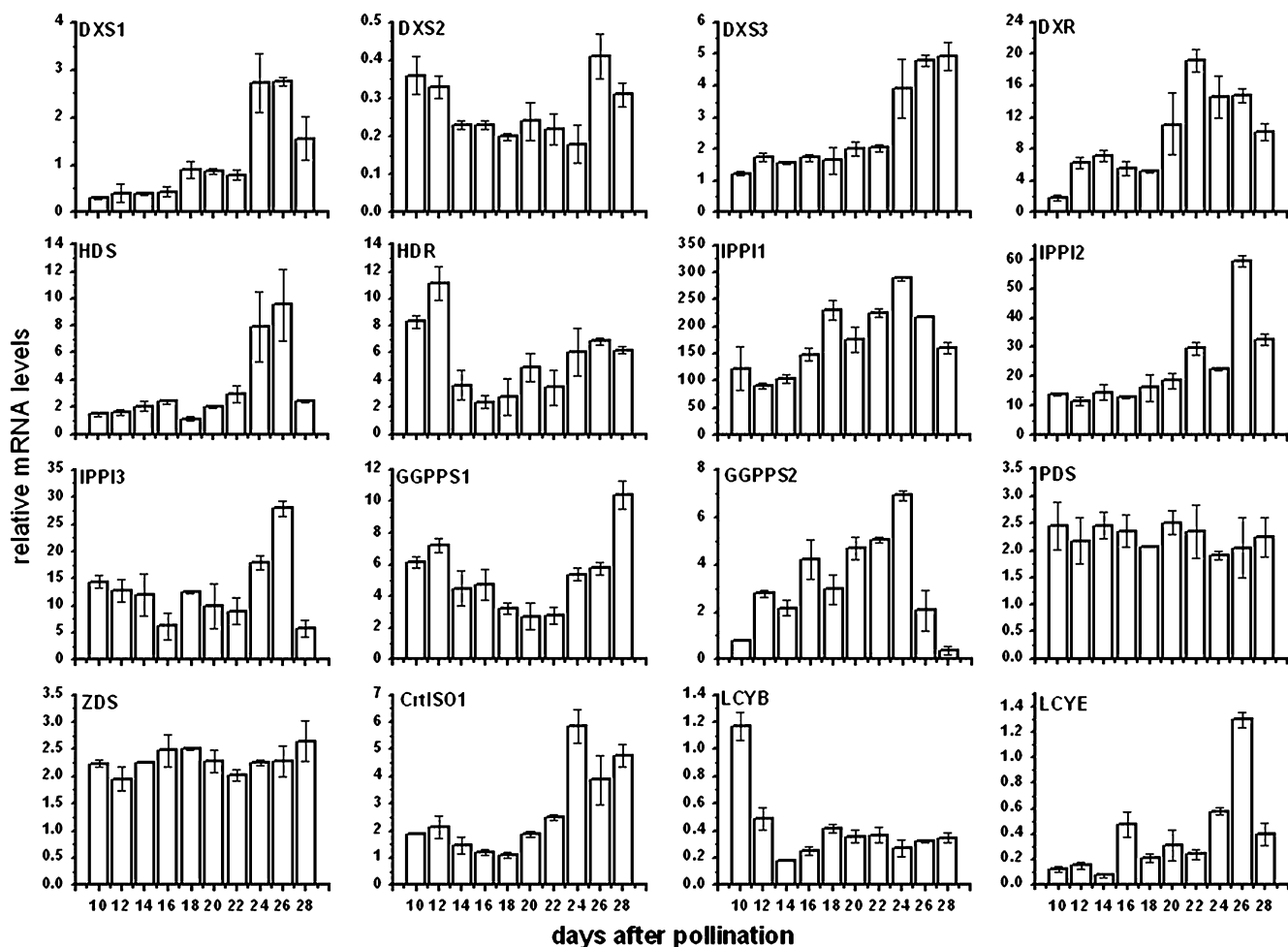
Predictive metabolic engineering/breeding of kernel carotenoid content is predicated on identifying genes and their temporal expression in developing endosperm that correlates with carotenoid content. Using the standard inbred line B73 (Li et al., 2008a), transcript levels were quantified across a panel of tissues and during endosperm development for all functional genes that were required for carotenoid accumulation (Figs. 4 and 5). Paralog-specific variation in transcript levels was observed in B73 maize. However, it is unknown how transcript levels correlate with accumulated endosperm carotenoids in genetically diverse inbreds varying in carotenoid content and composition (Islam, 2004; Harjes et al., 2008). It is also unknown how timing of gene expression during endosperm development is related to net accumulation of carotenoids in mature kernels. To investigate timing and correlation between gene expression and carotenoid content, transcript levels at specific endosperm developmental stages were compared with carotenoid content of mature kernels that varied in a genetically diverse germplasm collection. Quantitative transcript profiling was conducted in a core subset of



**Figure 4.** Maize B73 inbred transcript profiles of genes encoding enzymes for isoprenoid and carotenoid biosynthesis. Transcript levels were expressed relative to endosperm levels at 20 DAP. Means of three replicates  $\pm$  SD are shown. Embryo (Em) and endosperm (En) were collected at 20 DAP from field-grown plants; leaf (L) and root (R) samples were collected from seedlings at the six-leaf stage; U, unfertilized ovule.

maize lines chosen from a collection varying in kernel carotenoid content and spanning 80% of maize genetic diversity (Liu et al., 2003a; Islam, 2004; Harjes et al., 2008). The goal was to identify genes/paralogs with a statistically significant correlation between transcript levels at certain endosperm developmental stages and carotenoid content in the mature kernel. From a collection of 148 maize lines analyzed for carotenoid content (Islam, 2004), 10 lines were chosen that represented extremes of carotenoid content and composition (e.g. highest levels of zeaxanthin, lutein, etc.). These core lines spanned four major genetic diversity groups and eight

subgroups (Liu et al., 2003a). To validate the utility of the core subset to yield predictive results, we used the collection to test the *PSY* gene family, which contains one member, *PSY1*, known to control flux to endosperm carotenoids. Only *PSY1*, of three paralogs, showed the expected correlation of transcripts and carotenoid content in endosperm (Li et al., 2008b). The validated germplasm subset was then used to assess the influence of other genes in the pathway collection described above using paralog-specific primers (Supplemental Table S2). Transcript levels were quantified in developing endosperm 10 to 25 d after pollination (DAP; Supplemental



**Figure 5.** Maize B73 inbred developing endosperm transcript profiles of genes encoding enzymes for isoprenoid and carotenoid biosynthesis. Means of three replicates  $\pm$  SD are shown.

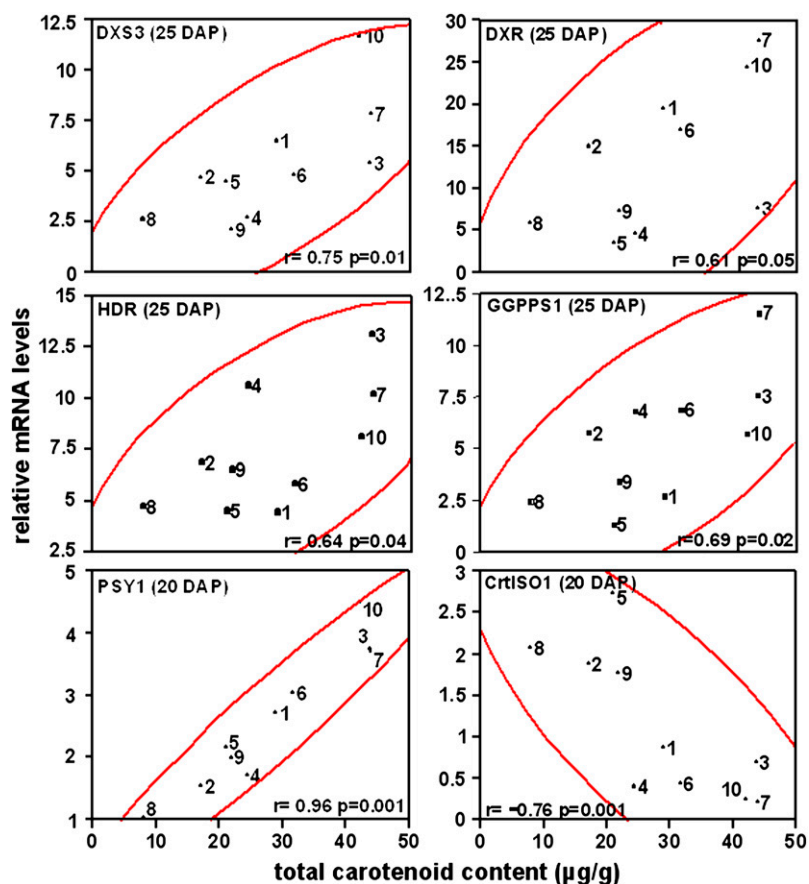
Table S3), stages when carotenoid accumulates (Li et al., 2008b). Transcript levels at each developmental stage were then statistically compared with carotenoid content of the mature kernel. Pearson correlation analysis (Table I) revealed that in addition to *PSY1*, several other genes showed a statistically significant correlation of transcripts with total carotenoid content (Fig. 6). Positively correlating genes upstream of the carotenoid pathway included *DXS3*, *DXR*, *HDR*, and *GGPPS1*, all at 25 DAP. Within the carotenoid pathway itself, *CrtISO1* transcripts negatively correlated with carotenoid content ( $r = -0.76$ ,  $P = 0.001$ ) as compared with the positive correlation seen with transcripts of the carotenoid pathway gene, *PSY1* ( $r = 0.96$ ,  $P = 0.001$ ), both at 20 DAP. The difference in  $r$  values for *CrtISO1* and *PSY1* is likely due to *PSY1* having been under strong selection (Palaisa et al., 2003), whereas *CrtISO1* has not been under any selection because it was unknown. A significant lesson drawn from these experiments is that trends in transcript accumulation patterns (Fig. 5) were not necessarily predictive of carotenoid accumulation. For example, transcripts of *DXS3* and *DXS1* both showed temporal

increases during endosperm development, but only *DXS3* showed correlative expression with carotenoids. Similarly, transcript levels of *IPPI* gene family members showed temporal increases during endosperm development but were not limiting as judged by the correlation analysis.

#### Role of ZEP in Depleting Endosperm Carotenoids

Conversion of carotenoids to ABA requires the activity of ZEP, an enzyme upstream of the rate-controlling cleavage step mediated by 9-cis-epoxycarotenoid dioxygenase (Fig. 1). To test whether ZEP expression was associated with reduced kernel carotenoids, gene-specific primers were developed for quantitative RT-PCR (Supplemental Table S1). In maize B73, ZEP genes showed similar tissue-specific transcript levels (Fig. 7A) but differed during endosperm development. ZEP transcript levels were next evaluated in developing endosperm of the 10 inbred line core subset (Supplemental Table S4) and tested for correlation with seed carotenoid composition (Table II). Both *ZEP1* and *ZEP2*

**Figure 6.** Correlation analysis showing that endosperm carotenoid content positively or negatively correlates with mRNA levels of specific pathway enzymes. Pearson correlation ( $r$ ) and statistical significance ( $P$ ) for comparison of transcript levels and kernel carotenoid content in maize inbred lines were determined using JMP version 5.1.2 (SAS Institute) to test the statistical significance ( $P \leq 0.05$ ) of the relationship at a 95% confidence interval (red lines). Inbred lines are as follows: 1, A619; 2, B73; 3, B37; 4, Cl.7; 5, C131A; 6, DE3; 7, KUI2007; 8, NC300; 9, SD44; 10, TZI18. PSY1 data were taken from Li et al. (2008b). [See online article for color version of this figure.]



transcript levels at 20 DAP showed statistically significant inverse correlation with seed zeaxanthin levels (Table II; *ZEP1*,  $r = -0.77$ ,  $P = 0.008$ ; *ZEP2*,  $r = -0.86$ ,  $P = 0.001$ ). This correlation matched the temporal window (20 DAP) of the earlier carotenoid pathway genes (encoding PSY and CrtISO). In contrast, no correlation was seen between ZEP transcripts and lutein levels, as predicted, since lutein is not a substrate of ZEP (Table II). The finding that ZEP expression negatively affects maize endosperm carotenoid accumulation is consistent with analysis of the tomato (*Solanum lycopersicum*) ZEP mutant (*hp3*), which showed a 30% increase in carotenoids compared with nonmutant fruits (Galpaz et al., 2008).

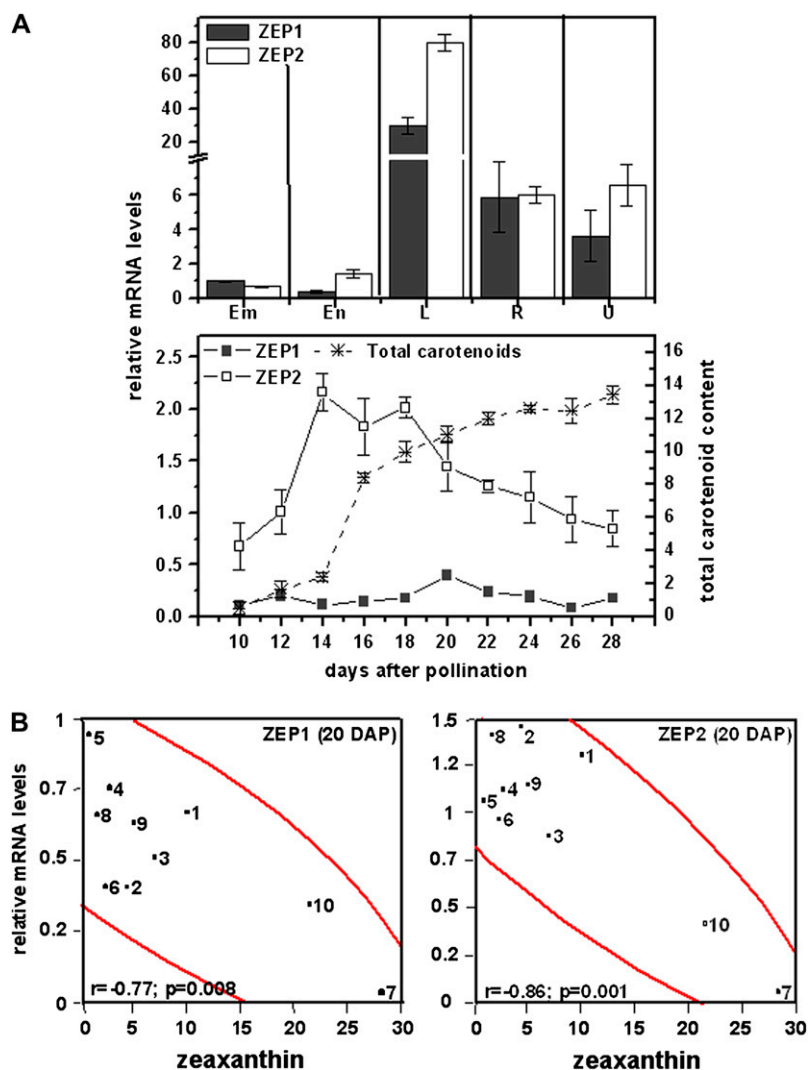
## DISCUSSION

Using bioinformatics and genome analysis, we produced a comprehensive identification of key genes and gene family members involved in the biosynthesis of carotenoids in maize and related grass species. Study of a genetically diverse maize germplasm helped elucidate the rate-controlling steps in seed carotenoid biosynthesis. We distinguished gene family members for which transcript levels statistically correlated with seed carotenoid content. We also provided temporal information on gene expression that will guide future breeding efforts whether transgenes or natural alleles

are utilized. These data will contribute to predictive strategies to engineer the maize endosperm carotenoid pathway, to stack traits that affect multiple biosynthetic pathways competing for isoprenoid precursors, or to divert isoprenoid precursors to alternate pathways. Lastly, phylogenetic analysis was used to identify putative targets in related grasses, extending these studies to other species that may lack the resources of maize.

### Upstream Isoprenoid Biosynthesis Controls Downstream Endosperm Carotenoid Content

The observation that several steps, including upstream isoprenoid biosynthesis pathways, influence flux to carotenoids indicates that PSY is not the only rate-controlling step in controlling maize endosperm carotenoid accumulation. The identification of genes encoding the isoprenoid IPP and GGPP-producing enzymes *DXS3*, *DXR*, *HDR*, and *GGPPS1* provides new targets for metabolic engineering of carotenoids in maize across genetically diverse germplasm. *DXS* and *GGPPS* both function at branch points of metabolism, whereby several pathways must compete for enzyme products. *HDR* produces a 5:1 mixture of IPP and dimethylallyl diphosphate isomers that must be combined in specific ratios for downstream prenyl transfer reactions of multiple pathways, including that for carotenoid biosynthesis (Rodriguez-Concepcion and Boronat, 2002). It is not surprising to have found these



**Figure 7.** Transcript profiling of *ZEP* gene family members. A, Top, tissue (see Fig. 4); bottom, developing endosperm (see Fig. 5). Transcript levels were expressed relative to endosperm levels at 20 DAP. Em, Embryo; En, endosperm; L, leaf; R, root; U, unfertilized ovule. B, Pearson correlation ( $r$ ) and statistical significance ( $P$ ) for comparison of transcript levels and kernel carotenoid content in maize inbred lines (see Fig. 6). Means of three replicates  $\pm$  SD are shown. [See online article for color version of this figure.]

particular enzymes to be influential, given evidence in other species that the enzymes may be rate controlling.

Up-regulation of the upstream isoprenoid and downstream carotenoid pathways is needed to force flux to carotenoids. For example, overexpression of *DXS* or *GGPPS* in combination with a carotenoid cluster significantly enhanced carotenoids produced in an *E. coli* bacterial platform (Wang et al., 1999; Matthews and Wurtzel, 2000). Similar examples in other plants have shown that overexpression of genes for certain upstream precursor pathway enzymes (*DXS*, *DXR*, or *HDR*) plus a downstream pathway enzyme forced pathway flux to the downstream pathway (Botella-Pavia et al., 2004; Carretero-Paulet et al., 2006; Munoz-Bertomeu et al., 2006). In contrast, overexpression of *HDS* had no effect on enhancing carotenoid accumulation in *E. coli* or *Arabidopsis* (Flores-Perez et al., 2008), consistent with the results shown here for maize.

Manipulation of a downstream pathway, in the absence of elevating the isoprenoid pool, may have negative consequences. For example, early attempts to increase tomato fruit carotenoids by *PSY* overexpres-

sion caused a dwarf phenotype due to redirection of GGPP away from gibberellins (Fray et al., 1995). Similar problems in improvement of tomato volatiles by overexpression of geraniol synthase to divert isoprenoids caused reduced carotenoid accumulation (Davidovich-Rikanati et al., 2007), and overexpression of taxadiene synthase redirected GGPP from GGPPS to taxadiene but also depleted the carotenoid pool (Besumbes et al., 2004).

Given that certain isoprenoid biosynthetic enzymes play such a critical role in controlling endosperm carotenoids, breeders might consider new approaches to breeding that incorporate surveys for “biosynthetic potential.” That is, maize carrying a mutation in *PSY1* is blocked in the downstream carotenoid pathway and appears as “white maize.” Such an ear would not be considered suitable for breeding for high (yellow) carotenoid content. However, this line may possess high levels of transcripts for one of the “rate-controlling” isoprenoid enzymes and hence have biosynthetic potential, once the carotenoid block is released. Conversely, ears that have high levels of *ZEP* transcripts may appear

**Table II.** Correlation of mRNA levels to carotenoid composition in mature seeds

Pearson correlation ( $r$ ) and statistical significance ( $P$ ), for comparison of transcript levels and kernel lutein and zeaxanthin content in maize inbred lines, was performed using JMP version 5.1.2 (SAS Institute) to test the statistical significance ( $P \leq 0.05$ ) of the relationship. Letters on a red background indicate negative correlation. Inbred lines are as listed in Figure 6. Carotenoid content correlated to mRNA levels is shown on a gray background. L, Lutein; Z, zeaxanthin. [See online article for color version of this table.]

Gene	Chromosome Bin	Content $r(\beta)$ Values			
		L	Z	L	Z
		20 DAP		25 DAP	
ZEP1	2.04 <sup>a</sup>	0.06 (0.84)	-0.77 (0.008)	-0.19 (0.59)	-0.35 (0.31)
ZEP2	10.04 <sup>a</sup>	-0.14 (0.68)	-0.86 (0.001)	-0.33 (0.35)	-0.42 (0.22)

<sup>a</sup>Vallabhaneni and Wurtzel (this article).

light yellow, masking optimal expression of other required biosynthetic steps.

### Isomerase Expression Negatively Affects Endosperm Carotenoid Content

The negative correlation seen for *CrtISO* was surprising given its placement in the pathway. *PDS* and *ZDS* were not correlative and the *Z-ISO* (for 15-cis zetacarotene isomerase) gene has yet to be cloned and therefore could not be tested. The isomerase *CrtISO* is one of several enzymes, including the desaturases *PDS* and *ZDS*, and a second isomerase, *Z-ISO*, unique to plant carotenoid biosynthesis and required to mediate desaturation of 15-cis phytoene to all-trans lycopene (Li et al., 2007). In contrast, bacterial pathways carry out these conversions with only a single enzyme, *CrtI*, which is why bacterial *CrtI* has been used as a simple tool in plant carotenoid metabolic engineering (Matthews and Wurtzel, 2007; Giuliano et al., 2008). At this point, it is a mystery why the *CrtISO* negative regulation is operative in this part of the pathway. Given that *CrtISO* expression controls pathway flux in endosperm, one should question whether engineering with a bacterial gene interferes with endogenous regulatory mechanisms of isomerization, as hinted by recent evidence of epigenetic control of *CrtISO* in *Arabidopsis* (Cazzonelli et al., 2009). Evidence that an isomerase controls flux partitioning in photosynthetic tissue is suggested by reduced lutein observed in *CrtISO* mutants and in plants overexpressing bacterial *CrtI* (Misawa et al., 1994; Park et al., 2002).

### Temporal Windows of Gene Expression That Influence Endosperm Carotenoid Content

Predictive metabolic breeding also requires an understanding of the timing of expression for pathway-controlling genes. Carotenoid accumulation in maize endosperm occurs continuously from 10 DAP onward as the seed matures (Li et al., 2008b). The correlations observed for upstream and downstream steps fell into two temporal categories: all genes controlling steps upstream of the carotenoid pathway showed correla-

tion at 25 DAP, whereas the carotenoid pathway genes (*PSY1* and *CrtISO*, *ZEP1*, and *ZEP2*) showed correlation at 20 DAP. One possible explanation for the temporal difference is that several pathways, including the carotenoid pathway, compete for isoprenoid precursors during the later phases of endosperm development.

### Using Maize to Identify Putative Targets in Other Grass Species for Improvement of Seed Carotenoids

Extension of the maize results using phylogenetic analysis identified orthologs in other grass species that may serve as potential engineering targets. The *DXS* phylogenetic tree (Fig. 3A) showed that the *DXS* triplication occurred prior to evolution of the grasses, suggesting that grass orthologs may have similar sub-functionalized roles, as shown for the *PSY* gene family (Li et al., 2008a, 2009; Welsch et al., 2008). Therefore *DXS3* orthologs in other grass species would be an attractive target for initiating investigations of rate-controlling steps for endosperm isoprenoid-derived metabolites. Similarly, the correlative single-copy genes (*DXR*, *HDR*, and *CrtISO*) are potential endosperm targets for other grass species having single-copy genes.

### CONCLUSION

In summary, transcript levels of seven genes were found to positively or negatively correlate with endosperm carotenoid content in genetically diverse germplasm; the critical timing of expression of these seven genes fell into two temporal windows of endosperm development. Manipulation of the pathway across the diversity of maize, in cultivars worldwide, therefore, will be somewhat complicated, with timing adding another dimension to this complexity. Incorporating a *DXS* transgene will not overcome suboptimal alleles that limit pathway flux when the trait is introgressed into various genotypes. However, these data now allow for the development of additional tools to canvas suitable alleles for the downstream steps or to use other transgenes as necessary.

In this study, the influence of candidate gene expression on endosperm carotenoid content was examined from the perspective of synthesis. This approach can also be used to dissect those factors that control endosperm composition for specific carotenoids (e.g. provitamin A carotenoids). Future study of the degradation genes/enzymes is also needed, since carotenoid degradation (Galpaz et al., 2008; Vogel et al., 2008) may potentially reduce carotenoid pools achieved by targeting the genes reported here. Beyond candidate genes, there are likely other factors, as evidenced from genetic (mutant and quantitative trait locus) studies in maize and other plants (Liu et al., 2003b; Wurtzel, 2004).

The utilization of a genetically and biochemically diverse maize germplasm collection to infer pathway regulation is an important resource in identifying potential gene targets for controlling endosperm carotenoid content. Germplasm collections in many plant species hold untapped potential in elucidating underlying regulatory mechanisms that will guide breeding and facilitate predictive metabolic engineering. Future investigations of pathway regulation, including timing of gene expression, which underlies the global network of plant metabolism, could be accomplished by deep transcript profiling of tissues from core collections chosen by metabolite subsorting and correlations made between mRNA levels and metabolite profiles.

## MATERIALS AND METHODS

### Sequence Analysis and Chromosome Mapping

Rice (*Oryza sativa*) genes (www.gramene.org) were used as a query to identify orthologs from maize (*Zea mays*; www.tigr.org; www.plantgdb.org) and to decipher gene families. Sequence analysis was performed using Vector NTI Suite 9.0 (Invitrogen). ESTs and conceptual translation were used to distinguish gene paralogs. Chromosomal positions of genes in the maize B73 inbred line were mapped either by utilizing tools available in the WebAGCoL package (Pampanwar et al., 2005) or at Maize GDB (www.maizegdb.org). Prediction of the chloroplast targeting site was made using ChloroP software (Emanuelsson et al., 1999).

The presence of a stop codon in *CrtISO2* was confirmed from a bacterial artificial chromosome clone (98301-103300; AC183901), EST (EE045563), genomic sequence from a methyl filtered library (AZM4\_69491), and GSS contig sequence (ZmGSSTUC11-12-04.150.3) from Plant GDB (www.plantgdb.org).

### Quantitative RT-PCR and Carotenoid Measurements

Plants and tissues were collected as described (Li et al., 2008b). RNA extraction and quantitative RT-PCR were performed using gene-specific primers (Supplemental Table S2) and normalized to actin, as described previously (Li et al., 2008b). Values are expressed as means of three RT-PCR replicates  $\pm$  SD. The carotenoid extraction procedure was based on Kurilich and Juvik (1999) and as described previously (Li et al., 2008b). Values are expressed as means of three extraction replicates  $\pm$  SD.

### Plasmids and Functional Complementation

Full-length cDNAs encoding *GGPPS1* and *GGPPS2* were amplified using approximately 100 ng of cDNA prepared from frozen maize endosperm (20 DAP), whereas *GGPPS3* was amplified using 0.5  $\mu$ g of maize B73 genomic DNA by PCR. Primers were designed on available genomic sequence adapted to the pET23 series of vectors (Novagen) with *EcoRI* and *HindIII* restriction sites. The resultant PCR products were cloned in the respective sites, designated as pET23aZmGGPPS1, pET23aZmGGPPS2, and pET23cZmGGPPS3, and used to

test function. These plasmids together with empty pET23a vector were transformed into *Escherichia coli* BL21 (DE3) cells containing pACCAR25 $\Delta$ crTE. Double transformants were grown in Luria-Bertani medium containing ampicillin and chloramphenicol and selected for their capacity to accumulate the yellow pigment zeaxanthin- $\beta$ -D-glucoside. Pigments were extracted and analyzed by HPLC (Gallagher et al., 2004). The sense and antisense primers used to construct the respective plasmids are as follows: GGPPS1, 5'-GAATTCGATAAGGAGGCCGTAACCTGTC-3' (#1000) and 5'-AAGCTTGCCTGACAAA-GTCACAACAGG-3' (#1001); GGPPS2, 5'-GAATTCATGGCGTTTCACTTC-CACCCGCTGGTC-3' (#1005) and 5'-AAGCTTAGTGAGCAAGAACAAGAGATAAAG-3' (#1004); and GGPPS3, 5'-GAATTCATGAACAAGTTGGCAT-CCTGCT-3' (#1565) and 5'-AAGCTTCAATGCTGTCTGTGCGCCATGA-3' (#1474).

### Statistical Analyses

Pearson correlation analysis of transcript and carotenoid composition from different maize inbred lines was performed using JMP version 5.1.2 (SAS Institute) to test the statistical significance ( $P \leq 0.05$ ) of the relationship.

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers EF417573 (ZmGGPPS1), EF417574 (ZmGGPPS2), and EF417575 (ZmGGPPS3).

### Supplemental Data

The following materials are available in the online version of this article.

**Supplemental Figure S1.** Multiple alignment of DXS proteins.

**Supplemental Figure S2.** GGPPS proteins of maize.

**Supplemental Figure S3.** Functional complementation of GGPPS1 and GGPPS2 from maize.

**Supplemental Figure S4.** Multiple alignment of CrtISO protein sequences.

**Supplemental Figure S5.** *ZEP* gene family structure.

**Supplemental Table S1.** Gene family sizes in three species of the Poaceae.

**Supplemental Table S2.** Primers used in the study.

**Supplemental Table S3.** Transcript levels in developing endosperm from diverse maize lines.

**Supplemental Table S4.** Transcript levels of *ZEP* genes in developing endosperm from diverse maize lines.

## ACKNOWLEDGMENTS

We thank Dr. Dwight Kincaid for advice on statistical analysis, Christina Murillo for technical assistance, and members of the Wurtzel laboratory (Dr. Faqiang Li, Rena Quinlan, Oren Tzfadia, Dr. Louis Bradbury, Dr. Abby Cuttriss, Dr. Maria Shumskaya, and Dr. Yu Chen) for helpful discussions.

Received February 17, 2009; accepted March 27, 2009; published April 3, 2009.

## LITERATURE CITED

- Besumbes O, Sauret-Gueto S, Phillips MA, Imperial S, Rodriguez-Concepcion M, Boronat A (2004) Metabolic engineering of isoprenoid biosynthesis in Arabidopsis for the production of taxadiene, the first committed precursor of taxol. *Biotechnol Bioeng* 88: 168-175
- Botella-Pavía P, Besumbes O, Phillips MA, Carretero-Paulet L, Boronat A, Rodriguez-Concepcion M (2004) Regulation of carotenoid biosynthesis in plants: evidence for a key role of hydroxymethylbutenyl diphosphate reductase in controlling the supply of plastidial isoprenoid precursors. *Plant J* 40: 188-199
- 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
- Carretero-Paulet L, Cairó A, Botella-Pavía P, Besumbes O, Campos N,

- Boronat A, Rodríguez-Concepción M (2006) Enhanced flux through the methylerythritol 4-phosphate pathway in Arabidopsis plants overexpressing deoxyxylulose 5-phosphate reductoisomerase. *Plant Mol Biol* **62**: 683–695
- Cazzonelli CI, Cuttriss AJ, Cossetto SB, Pye W, Crisp P, Whelan J, Finnegan EJ, Turnbull C, Pogson BJ (2009) Regulation of carotenoid composition and shoot branching in *Arabidopsis* by a chromatin modifying histone methyltransferase, SDG8. *Plant Cell* **21**: 39–53
- Cervantes-Cervantes M, Gallagher CE, Zhu C, Wurtzel ET (2006) Maize cDNAs expressed in endosperm encode functional farnesyl diphosphate synthase with geranylgeranyl diphosphate synthase activity. *Plant Physiol* **141**: 220–231
- Davidovich-Rikanati R, Sitrit Y, Tadmor Y, Iijima Y, Bilenko N, Bar E, Carmona B, Fallik E, Dudai N, Simon JE, et al (2007) Enrichment of tomato flavor by diversion of the early plastidial terpenoid pathway. *Nat Biotechnol* **25**: 899–901
- Emanuelsson O, Nielsen H, von Heijne G (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci* **8**: 978–984
- Flores-Perez U, Perez-Gil J, Rodriguez-Villalon A, Gil MJ, Vera P, Rodriguez-Concepcion M (2008) Contribution of hydroxymethylbutenyl diphosphate synthase to carotenoid biosynthesis in bacteria and plants. *Biochem Biophys Res Commun* **371**: 510–514
- Fray R, Wallace A, Fraser P, Valero D, Hedden P, Bramley P, Grierson D (1995) Constitutive expression of a fruit phytoene synthase gene in transgenic tomatoes causes dwarfism by redirecting metabolites from gibberellin pathway. *Plant J* **8**: 693–701
- Gallagher CE, Matthews PD, Li F, Wurtzel ET (2004) Gene duplication in the carotenoid biosynthetic pathway preceded evolution of the grasses (Poaceae). *Plant Physiol* **135**: 1776–1783
- Galpaz N, Wang Q, Menda N, Zamir D, Hirschberg J (2008) Abscisic acid deficiency in the tomato mutant *high-pigment 3* leading to increased plastid number and higher fruit lycopene content. *Plant J* **53**: 717–730
- Giuliano G, Tavazza R, Diretto G, Beyer P, Taylor MA (2008) Metabolic engineering of carotenoid biosynthesis in plants. *Trends Biotechnol* **26**: 139–145
- Harjes CE, Rocheford TR, Bai L, Brutnell TP, Kandianis CB, Sowinski SG, Stapleton AE, Vallabhaneni R, Williams M, Wurtzel ET, et al (2008) Natural genetic variation in *lycopene epsilon cyclase* tapped for maize biofortification. *Science* **319**: 330–333
- Islam SN (2004) Survey of carotenoid variation and quantitative trait loci mapping for carotenoid and tocopherol variation in maize. Masters thesis. University of Illinois, Urbana-Champaign, IL
- Kim BR, Kim SU, Chang YJ (2005) Differential expression of three 1-deoxy-D-xylulose-5-phosphate synthase genes in rice. *Biotechnol Lett* **27**: 997–1001
- Kurilich A, Juvik J (1999) Quantification of carotenoid and tocopherol antioxidants in *Zea mays*. *J Agric Food Chem* **47**: 1948–1955
- Li F, Murillo C, Wurtzel ET (2007) Maize Y9 encodes a product essential for 15-cis zeta-carotene isomerization. *Plant Physiol* **144**: 1181–1189
- Li F, Tzfadia O, Wurtzel ET (2009) The *Phytoene Synthase* gene family in the grasses: Subfunctionalization provides tissue-specific control of carotenogenesis. *Plant Signal Behav* **4**: 208–211
- Li F, Vallabhaneni R, Wurtzel ET (2008a) *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
- Li F, Vallabhaneni R, Yu J, Rocheford T, Wurtzel ET (2008b) The maize phytoene synthase gene family: overlapping roles for carotenogenesis in endosperm, photomorphogenesis, and thermal stress-tolerance. *Plant Physiol* **147**: 1334–1346
- 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
- Liu K, Goodman M, Muse S, Smith JS, Buckler E, Doebley J (2003a) Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. *Genetics* **165**: 2117–2128
- Liu YS, Gur A, Ronen G, Causse M, Damidaux R, Buret M, Hirschberg J, Zamir D (2003b) There is more to tomato fruit colour than candidate carotenoid genes. *Plant Biotechnol J* **1**: 195–207
- Mahmoud SS, Croteau RB (2001) Metabolic engineering of essential oil yield and composition in mint by altering expression of deoxyxylulose phosphate reductoisomerase and menthofuran synthase. *Proc Natl Acad Sci USA* **98**: 8915–8920
- Matthews PD, Luo R, Wurtzel ET (2003) Maize phytoene desaturase and zeta-carotene desaturase catalyze a poly-Z desaturation pathway: implications for genetic engineering of carotenoid content among cereal crops. *J Exp Bot* **54**: 2215–2230
- Matthews PD, Wurtzel ET (2000) Metabolic engineering of carotenoid accumulation in *Escherichia coli* by modulation of the isoprenoid precursor pool with expression of deoxyxylulose phosphate synthase. *Appl Microbiol Biotechnol* **53**: 396–400
- Matthews PD, Wurtzel ET (2007) Biotechnology of food colorant production. In C Socaciu, ed, *Food Colorants: Chemical and Functional Properties*. CRC Press, Boca Raton, FL, pp 347–398
- Misawa N, Masamoto K, Hori T, Ohtani T, Boger P, Sandmann G (1994) Expression of an Erwinia phytoene desaturase gene not only confers multiple resistance to herbicides interfering with carotenoid biosynthesis but also alters xanthophyll metabolism in transgenic plants. *Plant J* **6**: 481–489
- Munoz-Bertomeu J, Arrillaga I, Ros R, Segura J (2006) Up-regulation of 1-deoxy-D-xylulose-5-phosphate synthase enhances production of essential oils in transgenic spike lavender. *Plant Physiol* **142**: 890–900
- Okada K, Saito T, Nakagawa T, Kawamukai M, Kamiya Y (2000) Five geranylgeranyl diphosphate synthases expressed in different organs are localized into three subcellular compartments in Arabidopsis. *Plant Physiol* **122**: 1045–1056
- Palaisa KA, Morgante M, Williams M, Rafalski A (2003) Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. *Plant Cell* **15**: 1795–1806
- Pampanwar V, Engler F, Hatfield J, Blundy S, Gupta G, Soderlund C (2005) FPC Web tools for rice, maize, and distribution. *Plant Physiol* **138**: 116–126
- Park H, Kreunen SS, Cuttriss AJ, DellaPenna D, Pogson B (2002) Identification of the carotenoid isomerase provides insight into carotenoid biosynthesis, prolamellar body formation, and photomorphogenesis. *Plant Cell* **14**: 321–332
- Pozniak CJ, Knox RE, Clarke FR, Clarke JM (2007) Identification of QTL and association of a phytoene synthase gene with endosperm colour in durum wheat. *Theor Appl Genet* **114**: 525–537
- Randolph LE, Hand DB (1940) Relation between carotenoid content and number of genes per cell in diploid and tetraploid corn. *J Agric Res* **60**: 51–64
- Rodriguez-Concepcion M, Boronat A (2002) Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids: a metabolic milestone achieved through genomics. *Plant Physiol* **130**: 1079–1089
- 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
- Vogel JT, Tan BC, McCarty DR, Klee HJ (2008) The carotenoid cleavage dioxygenase 1 enzyme has broad substrate specificity, cleaving multiple carotenoids at two different bond positions. *J Biol Chem* **283**: 11364–11373
- Walter M, Hans J, Strack D (2002) Two distantly related genes encoding 1-deoxy-D-xylulose 5-phosphate synthases: differential regulation in shoots and apocarotenoid-accumulating mycorrhizal roots. *Plant J* **31**: 243–254
- Wang CW, Oh MK, Liao JC (1999) Engineered isoprenoid pathway enhances astaxanthin production in *Escherichia coli*. *Biotechnol Bioeng* **62**: 235–241
- Welsch R, Wust F, Bar C, Al-Babili S, Beyer P (2008) A third phytoene synthase is devoted to abiotic stress-induced abscisic acid formation in rice and defines functional diversification of phytoene synthase genes. *Plant Physiol* **147**: 367–380
- Wong JC, Lambert RJ, Wurtzel ET, Rocheford TR (2004) QTL and candidate genes phytoene synthase and zeta-carotene desaturase associated with the accumulation of carotenoids in maize. *Theor Appl Genet* **108**: 349–359
- Wurtzel ET (2004) Genomics, genetics, and biochemistry of maize carotenoid biosynthesis. *Recent Adv Phytochem* **38**: 85–110
- Zhu C, Naqvi S, Breitenbach J, Sandmann G, Christou P, Capell T (2008) Combinatorial genetic transformation generates a library of metabolic phenotypes for the carotenoid pathway in maize. *Proc Natl Acad Sci USA* **105**: 18232–18237

Enzyme	Gene family size		
	<i>Zea mays</i> GenBank ID	<i>Oryza sativa</i> Gramene ID	<i>Sorghum bicolor</i> Phytozome ID
<b>DXS1</b>	AY946271	LOC_Os05g33840	Sb09g020140
<b>DXS2</b>	AY946270	LOC_Os07g09190	Sb02g005380
<b>DXS3</b>	DR967452	LOC_Os06g05100	Sb10g002960
<b>DXR</b>	AJ297566	LOC_Os01g01710	Sb03g008650
<b>HDS</b>	AY562489	LOC_Os02g39160	Sb04g025290
<b>HDR</b>	DR789385	LOC_OS03g52170	Sb01g009140
<b>IPPI1</b>	AF330034	LOC_Os07g36190	Sb02g035700
<b>IPPI2</b>	DN215017	LOC_Os05g34180	Sb09g020370
<b>IPPI3</b>	CO530856	NA	NA
<b>GGPPS1</b>	EF417573	LOC_Os07g39270	Sb02g037510
<b>GGPPS2</b>	EF417574	NA	NA
<b>GGPPS3</b>	EF417575	LOC_Os01g14630	Sb03g009380
<b>PSY1</b>	AY324431	LOC_Os06g51290	Sb10g031020
<b>PSY2</b>	AY325302	LOC_Os12g43130	Sb08g022310
<b>PSY3</b>	DQ372936	LOC_Os09g38320	Sb02g032370
<b>PDS</b>	L39266	LOC_Os03g08570	Sb06g030030
<b>ZDS</b>	AF047490	LOC_Os07g10490	Sb02g006100
<b>CRTISO1</b>	DR812825	LOC_Os11g36440	Sb05g022240
<b>CRTISO2</b>	AC183901	NA	NA
<b>LCYB</b>	AY206862	LOC_Os02g09750	Sb04g006120
<b>LCYE</b>	EB674039	LOC_Os01g39960	Sb03g026020
<b>ZEP1</b>	DR820114	LOC_Os04g37619	Sb06g018220
<b>ZEP2</b>	CO532283	NA	NA

**Table S1. Gene family sizes in three species of the Poaceae.** Enzymes highlighted in blue or red are encoded by a maize gene family member showing positive or negative correlation, respectively, between transcripts and total carotenoid content in maize endosperm. Accession numbers are obtained as stated: *Zea mays*, GenBank; *Oryza sativa*, Gramene; *Sorghum bicolor*, Phytozome

Gene	Primer No.	Sequence	Gen Bank Acc #
DXS1	1206	CAGAACAGGGAGGCTCTTG	AY946271
	1207	GCGACACTTTCATCCAAAGA	
DXS2	1212	GCTGAACTACTTCCAGAAGCG	AY946270
	1213	CTGCAGGAACGACGAGTAGA	
DXS3	1194	CTGTGCCAGCAGCATAGTTT	DR967452
	1195	TTCGACCGTCAAGTAGACCA	
DXR	1164	TCCAATGTCACGCTTCTAGC	AJ297566
	1165	TGGCGAGCAACTTCTATGAC	
HDS	1176	CAGTCCAGGGTAGGAAGGAA	AY562489
	1177	TAGGGCTACCGTAGGAGCAA	
HDR	1218	GCCCAACTCGTTACATCCTT	DR789385
	1219	TGAGATCAAGCGTCAGGAAA	
IPPI1	1159A	TGAGCTTATCCAGGAGAACTA	AF330034
	1163A	ATGGTGCCTTGTAAGCATT	
IPPI2	1333	TCAGGCTGGTGGTGGACAACCTT	DN215017
	1335	TGATGGCTAAAACAAAGGCA	
IPPI3	1333	TCAGGCTGGTGGTGGACAACCTT	CO530856
	1334	AGCACCAAACAGCAGGACAA	
GGPPS1	1475	CGGAGGAGTTGCTCTCTGAT	EF417573
	1476	AGTCACAACAGGTCCGACAA	
GGPPS2	1477	CAGAGAAGTTGGTCTCTGAC	EF417574
	1478	AGATCGACAGTGAGCAAGAA	
GGPPS3	1426	ATCCATGTGCACAAGACTGC	EF417575
	1427	CACGTCCAGAACGTCGTC	
PSY1	505	CATCTTCAAAGGGGTCGTC	AY324431
	254	CAGGATCTGCCTGTACAACA	
PSY2	503	TCACCCATCTCGACTCTGCTA	AY325302
	676	GATGTGATCTACGGATGGTTCAT	
PSY3	996	GCCAGAGCCTTCTTCAGGCAGG	DQ372936
	997	GTCTTCGGAACGTAGGCCCTC	
PDS	933	GAAATCATCGATGCAACTATGGAA	L39266
	934	CTTCGATAGGTGACCTTTGGA	
ZDS	281	GTGTGGTAAAGATCGGACAA	AF047490
	917	AGAGAGTTGCTCCTTCCAT	
CrtISO1	1325	GGTTGTCAAATGGGAAGGAG	DR812825
	1326	GAATGATGGTGCCTTGACAT	
LCYB	1095	CATCGTAAGGTTCTCGACA	AY206862
	1096	ATGCCGAAGCAGAAGAACTC	
LCYE	1037	TTT ACG TGC AAA TGC AGT CAA	EB674039
	1038	TGA CTC TGA AGC TAG AGA AAG	
ZEP1	1664	TTGGTGGCAACAGCTCAAAA	DR820114
	1665	CCTTCACTTGTTGCGATGAG	
ZEP2	1400	TTGGTGGCAATAGCTCAAAG	CO532283
	1401	GCTTCACTTGTTGCGAAGAG	
Actin	1134	CGATTGAGCATGGCATTGTCA	J01238
	1135	CCCCTAGCGTACAACGAA	

**Table S2.** Primers used in the study

Relative transcript levels of DXS3, DXR, HDR, GGPPS1, PSY1, CrtISO1 in diverse maize lines						
20 DAP						
inbred	DXS3	DXR	HDR	GGPPS1	PSY1	CrtISO1
A619	1.37 (0.12)	7.02 (0.97)	2.89 (0.42)	6.52 (1.26)	2.69 (0.52)	0.86 (0.01)
B73	2.03 (0.21)	11.14 (3.86)	4.89 (0.64)	2.73 (0.83)	1.53 (0.05)	1.88 (0.08)
B37	3.59 (0.44)	7.47 (0.85)	3.17 (0.86)	12.05 (1.05)	3.74 (0.52)	0.7 (0.03)
CI7	1.80 (0.60)	3.32 (0.93)	3.58 (0.37)	6.98 (2.00)	1.71 (0.11)	0.41 (0.08)
C131A	5.14 (0.34)	3.21 (0.35)	2.47 (0.69)	0.81 (0.12)	2.15 (0.49)	2.8 (0.35)
DE3	4.16 (0.19)	8.23 (0.46)	4.26 (0.65)	2.18 (0.48)	3.04 (0.06)	0.42 (0.04)
KUI2007	4.80 (0.14)	20.93 (4.59)	3.39 (0.69)	4.86 (0.43)	3.65 (0.17)	0.21 (0.02)
NC300	2.55 (0.53)	4.26 (0.59)	2.9 (0.94)	2.51 (0.52)	1.01 (0.31)	2.08 (0.54)
SD44	6.43 (0.38)	10.96 (2.53)	4.36 (0.63)	5.15 (0.40)	1.98 (0.35)	1.77 (0.30)
TZI18	4.51 (0.18)	3.56 (0.37)	2.84 (0.43)	4.38 (0.52)	4.47 (0.96)	0.25 (0.05)
25 DAP						
inbred	DXS3	DXR	HDR	GGPPS1	PSY1	CrtISO1
A619	6.56 (0.21)	19.43 (0.95)	4.42 (0.72)	2.69 (0.46)	9.88 (1.76)	1.02 (0.34)
B73	4.79 (0.19)	14.87 (0.88)	6.84 (0.19)	5.76 (0.41)	1.15 (0.25)	3.87 (0.88)
B37	5.43 (0.66)	7.57 (1.33)	13.11 (1.81)	7.55 (1.08)	16.21 (0.29)	1.01 (0.29)
CI7	2.73 (0.21)	4.58 (0.46)	10.62 (1.88)	6.81 (1.40)	1.52 (0.42)	0.93 (0.27)
C131A	4.51 (0.28)	3.49 (0.11)	4.52 (0.14)	1.32 (0.35)	0.74 (0.23)	5.26 (0.06)
DE3	4.80 (0.19)	16.84 (0.88)	5.81 (1.29)	6.86 (0.97)	2.03 (0.74)	1.55 (0.01)
KUI2007	7.80 (0.11)	27.68 (2.66)	10.16 (1.48)	11.48 (0.40)	1.20 (0.24)	3.75 (0.29)
NC300	2.60 (0.50)	5.69 (0.38)	4.7 (0.02)	2.42 (0.40)	0.94 (0.19)	2.17 (0.12)
SD44	2.16 (0.11)	7.30 (0.11)	6.52 (1.18)	3.37 (0.50)	2.37 (0.46)	1.46 (0.12)
TZI18	11.72 (0.48)	24.43 (3.30)	8.11 (1.13)	5.69 (0.10)	3.98 (0.40)	0.59 (0.13)

**Table S3. Transcript levels in developing endosperm from diverse maize lines.** Transcripts were measured quantitatively from endosperm collected at the indicated DAP. Parentheses show SD, standard deviation. Quantitative RT-PCR was performed using gene specific primers (Table S2) and normalized to actin, as previously described (Li et al., 2008a). Values are expressed as the mean of three RT-PCR replicates +/- standard deviation.

	Relative transcript levels of <i>ZEP1</i> and <i>ZEP2</i> (SD)			
	20 DAP		25 DAP	
	<i>ZEP 1</i>	<i>ZEP2</i>	<i>ZEP1</i>	<i>ZEP2</i>
<b>A619</b>	0.67 (0.10)	1.30 (0.29)	0.32 (0.01)	0.53 (0.07)
<b>B73</b>	0.40 (0.03)	1.45 (0.24)	0.09 (0.03)	0.94 (0.22)
<b>B37</b>	0.51 (0.21)	0.88 (0.22)	0.22 (0.03)	0.27 (0.01)
<b>CI7</b>	0.75 (0.00)	1.12 (0.01)	0.09 (0.02)	0.18 (0.05)
<b>C131A</b>	0.94 (0.37)	1.06 (0.37)	0.46 (0.05)	0.89 (0.09)
<b>DE3</b>	0.40 (0.06)	0.96 (0.17)	0.02 (0.01)	0.62 (0.00)
<b>KUI2007</b>	0.04 (0.02)	0.05 (0.00)	0.14 (0.00)	0.56 (0.04)
<b>NC300</b>	0.66 (0.26)	1.41 (0.21)	0.84 (0.05)	1.11 (0.07)
<b>SD44</b>	0.63 (0.16)	1.15 (0.14)	0.79 (0.19)	0.97 (0.25)
<b>TZI18</b>	0.34 (0.08)	0.41(0.13)	0.07 (0.01)	0.16 (0.02)

**Table S4. Transcript levels of *ZEP* genes in developing endosperm from diverse maize lines.** Transcripts were measured quantitatively from endosperm collected at the indicated DAP. Parentheses show SD, standard deviation. Quantitative RT-PCR was performed using gene specific primers (Table S2) and normalized to actin, as previously described (Li et al., 2008a). Values are expressed as the mean of three RT-PCR replicates +/- standard deviation.



```

OsDXS3 : MGGRYLHSPVANLICEESTFVRSTSCCLHSLCLCRCKTSLNLCPCRTLVLVFNMPSSRRFNRRVAAALPD---VDD---FFWEKDPPIPLDTIDAPIHLKLNLSKSELKQLADEVRSEIAFTMSRKCQPCN-TGHSVVELLTAIH
SbDXS3 : MDTAFLRPLARNLVYDEFVLHPTSYPFHTLRLYLRNCPMYSRPLLTSPASPSRGLTQRMAALPD---VDD---FFWEKDPPIPLDTIDAPIHLKLNLSKSELKQLADEVRSEIAFTMSRKCQPCG-PGRSVVELLTAIH
ZmDXS3 : MDTAFLRPLARNLVYDEFVLHPTSYPFHTLRLYLRNCPMYSRPLLTSPASPSRGLTQRMAALPD---VDD---FFWEKDPPIPLDTIDAPIHLKLNLSKSELKQLADEVRSEIAFTMSRKCQPCG-ADRSVVELLTAIH
OsDXS2 : MALQASS-PSMFRAPITNTNASCRRLQVRASAAAAAN----GGDGVMMRK-----EAS---GAWKIDYS-GEKBATPLLDVTVYVPHMKNLSTPELEQLAAELRAEIVHTVSKTGGHLS-SSLGVELLVALH
SbDXS2 : MALQASS-PSAFRAVATANASCRRQFQVRAQVAGSSSSSVGADGAGKMMFKAEPAAAAATSSGSSGPKIDFTSGEKPATPLLDVTVYVPHMKNLSTPELEQLAAELRAEIVHTVSKTGGHLS-SSLGVELLVALH
ZmDXS2 : MALQASSPSAFRAVATANASCRRQFQVRAQVAGSSSSSISGADGG-KMMVSK-EPAAAAATSS---GWKIDFS-GEKPPPLLDVTVYVPHMKNLSTPELEQLAAELRAEIVHTVSKTGGHLS-SSLGVELLVALH
OsDXS1 : -----MALTTFSISRGGFVGLPQ-EGHFAPAABELSHLKLQS-RPHKAR-RSSSSISASLS---TEREAAYHSQREPPPLLDVTVYVPHMKNLSTPELEQLADELRSDVIFHVSKTGGHLG-SSLGVELLVALH
SbDXS1 : -----MALWTFESVPRG-FLAVFAHDSHFASAVEQLHGHKLLQARPLRPRRHRPACVSASLS---PERE-AEYYSQREPPPLLDVTVYVPHMKNLSTPELEQLADELRSDVIFHVSKTGGHLG-SSLGVELLVALH
ZmDXS1 : -----MALSTFSVPRG-FLGVPAQ-DSHFASAVE-LHVNKLLQARPINLKPRRRPACVSASLS---SERE-AEYYSQREPPPLLDVTVYVPHMKNLSTPELEQLADELRSDVIFHVSKTGGHLG-SSLGVELLVALH

OsDXS3 : YVFNAPMDKILWDACQHTYAHKILTGRRSLFHTIKRKNLSGFTSRSESEYDPPFCAGHGCNLSLACLGMAVARDLGRKKNRIVTVISNWTMTAGQVYEAAMGHAGFLDSNMVVLNDSRHHTLTPKADSG-SKMSINALSSA
SbDXS3 : YVFNAPMDKILWDACQHAYAHKILTGRRSQFHTIKRKNLSGFTSRSESEYDPPFCAGHGCNLSLACLGMAVARDINGRKNRIVTVISNWTMTAGQVYEAAMGHAGFLDSNMVVLNDSCHTLLPKADGR-EKMSVNAEASSA
ZmDXS3 : YVFNAPMDKILWDACQHAYAHKILTGRRSLFHTIKRKNLSGFTSRSESEYDPPFCAGHGCNLSLACLGMAVARDINGRKNRIVTVISNWTMTAGQVYEAAMGHAGFLDSNMVVLNDSCHTLLPKADGG-AKMSVNAEASSA
OsDXS2 : HVFTPEBDKI IWDVGHQAYPHKILTGRSRMHTI RQTSGLAGPKRDESAHDAFGVGHSSSTISAAALGMAVARDLLGRKNHVTISVIGDGMTAGQAYEAAMNNSGYLDSNMIVLNDNKQVSLPTATLDGSPVPVCAALSSA
SbDXS2 : HVFTPEBDKI IWDVGHQAYPHKILTGRSRMHTI RQTSGLAGPKRDESAHDAFGVGHSSSTISAAALGMAVARDLLGRKNHVTISVIGDGMTAGQAYEAAMNNSGYLDSNMIVLNDNKQVSLPTATLDGSPVPVCAALSSA
ZmDXS2 : HVFTPEBDKI IWDVGHQAYPHKILTGRSRMHTI RQTSGLAGPKRDESAHDAFGVGHSSSTISAAALGMAVARDLLGRKNHVTISVIGDGMTAGQAYEAAMNNSGYLDSNMIVLNDNKQVSLPTATLDGSPVPVCAALSSA
OsDXS1 : YVFNAPMDRILWDVGHQSYPHKILTGRDKMPTMRCQTNGLAGTKRAESEYDSFGTGHSSSTISAAALGMAVARDLKGKKNVAVIGDGMTAGQAYEAAMNAGYLDSDMIVLNDNKQVSLPTATLDGSPVPVCAALSSA
SbDXS1 : YVFNAPMDRILWDVGHQSYPHKILTGRDKMPTMRCQTNGLAGTKRAESEYDSFGTGHSSSTISAAALGMAVARDLKGKKNVAVIGDGMTAGQAYEAAMNAGYLDSDMIVLNDNKQVSLPTATLDGSPVPVCAALSSA
ZmDXS1 : YVFNAPMDRILWDVGHQSYPHKILTGRDKMPTMRCQTNGLAGTKRAESEYDSFGTGHSSSTISAAALGMAVARDLKGKKNVAVIGDGMTAGQAYEAAMNAGYLDSDMIVLNDNKQVSLPTATLDGSPVPVCAALSSA

OsDXS3 : LSKVQSSKGRFRFREAAGLAKWFCCKGMHEFAAKIDEYARGMIGPHCATLFEELGLYYIGPIDGNNDLLICVWKEVSTLDSTGPVLVHVITNEKDSGGEFNSEITPDEECPDSSQDIKFLENG-LRSTYNDCEVES
SbDXS3 : LSKIQSSKGRFRFREAAGLAKWFCCKGMHEFAAKVDEYARGMIGPHCATLFEELGLYYIGPIDGNNDLLICVWKEVSTLDSTGPVLVHVITGTESDTDNIRSEITASEECPNSNSNDLLKPLETG-LRSTYNDCEVES
ZmDXS3 : LSKIQSSKGRFRFREAAGLAKWFCCKGMHEFAAKVDEYARGMIGPHCATLFEELGLYYIGPIDGNNDLLICVWKEVSTLDSTGPVLVHVITGTENDTCGIIGSEINAVEECPNSNSPDLKFLGTG-LRSTYNDCEVES
OsDXS2 : LTKLQSSTKLRRFREAAGLAKWFCCKGMHEFAAKVDEYARGMIGPHCATLFEELGLYYIGPIDGNNDLLICVWKEVSTLDSTGPVLVHVITNEKDSGGEFNSEITPDEECPDSSQDIKFLENG-LRSTYNDCEVES
SbDXS2 : LTKLQSSTKLRRFREAAGLAKWFCCKGMHEFAAKVDEYARGMIGPHCATLFEELGLYYIGPIDGNNDLLICVWKEVSTLDSTGPVLVHVITGTESDTDNIRSEITASEECPNSNSNDLLKPLETG-LRSTYNDCEVES
ZmDXS2 : LTKLQSSTKLRRFREAAGLAKWFCCKGMHEFAAKVDEYARGMIGPHCATLFEELGLYYIGPIDGNNDLLICVWKEVSTLDSTGPVLVHVITGTESDTDNIRSEITASEECPNSNSNDLLKPLETG-LRSTYNDCEVES
OsDXS1 : LSKLQSSRPLRELEVAAGLAKWFCCKGMHEFAAKVDEYARGMIGPHCATLFEELGLYYIGPIDGNNDLLICVWKEVSTLDSTGPVLVHVITNEKDSGGEFNSEITPDEECPDSSQDIKFLENG-LRSTYNDCEVES
SbDXS1 : LSKLQSSRPLRELEVAAGLAKWFCCKGMHEFAAKVDEYARGMIGPHCATLFEELGLYYIGPIDGNNDLLICVWKEVSTLDSTGPVLVHVITGTESDTDNIRSEITASEECPNSNSNDLLKPLETG-LRSTYNDCEVES
ZmDXS1 : LSKLQSSRPLRELEVAAGLAKWFCCKGMHEFAAKVDEYARGMIGPHCATLFEELGLYYIGPIDGNNDLLICVWKEVSTLDSTGPVLVHVITGTESDTDNIRSEITASEECPNSNSNDLLKPLETG-LRSTYNDCEVES

OsDXS3 : LIAEAEADKRIIVVHGGMGIDRSIRLQSRFPDRFDFDLGIAEQHAVTFSAGLACGGKLPFCITPSTFLQRAYDQIVEDVDVMQKIPVRFATITNAGLVGSEGPTNSGPFDIIFMSGLPNMIVMSPSNEDELIDMVATAAMVE
SbDXS3 : LIAEAEADKRIIVVHGGMGIDRSIRLQSRFPDRFDFDLGIAEQHAVTFSAGLACGGKLPFCITPSTFLQRAYDQIVEDVDVMQKIPVRFATITNAGLVGSEGPTNSGPFDIIFMSGLPNMIVMSPSNEDELIDMVATAAMVE
ZmDXS3 : LIAEAEADKRIIVVHGGMGIDRSIRLQSRFPDRFDFDLGIAEQHAVTFSAGLACGGKLPFCITPSTFLQRAYDQIVEDVDVMQKIPVRFATITNAGLVGSEGPTNSGPFDIIFMSGLPNMIVMSPSNEDELIDMVATAAMVE
OsDXS2 : LIAEAEADKRIIVVHGGMGIDRSIRLQSRFPDRFDFDLGIAEQHAVTFSAGLACGGKLPFCITPSTFLQRAYDQIVEDVDVMQKIPVRFATITNAGLVGSEGPTNSGPFDIIFMSGLPNMIVMSPSNEDELIDMVATAAMVE
SbDXS2 : LIAEAEADKRIIVVHGGMGIDRSIRLQSRFPDRFDFDLGIAEQHAVTFSAGLACGGKLPFCITPSTFLQRAYDQIVEDVDVMQKIPVRFATITNAGLVGSEGPTNSGPFDIIFMSGLPNMIVMSPSNEDELIDMVATAAMVE
ZmDXS2 : LIAEAEADKRIIVVHGGMGIDRSIRLQSRFPDRFDFDLGIAEQHAVTFSAGLACGGKLPFCITPSTFLQRAYDQIVEDVDVMQKIPVRFATITNAGLVGSEGPTNSGPFDIIFMSGLPNMIVMSPSNEDELIDMVATAAMVE
OsDXS1 : LIAEAEADKRIIVVHGGMGIDRSIRLQSRFPDRFDFDLGIAEQHAVTFSAGLACGGKLPFCITPSTFLQRAYDQIVEDVDVMQKIPVRFATITNAGLVGSEGPTNSGPFDIIFMSGLPNMIVMSPSNEDELIDMVATAAMVE
SbDXS1 : LIAEAEADKRIIVVHGGMGIDRSIRLQSRFPDRFDFDLGIAEQHAVTFSAGLACGGKLPFCITPSTFLQRAYDQIVEDVDVMQKIPVRFATITNAGLVGSEGPTNSGPFDIIFMSGLPNMIVMSPSNEDELIDMVATAAMVE
ZmDXS1 : LIAEAEADKRIIVVHGGMGIDRSIRLQSRFPDRFDFDLGIAEQHAVTFSAGLACGGKLPFCITPSTFLQRAYDQIVEDVDVMQKIPVRFATITNAGLVGSEGPTNSGPFDIIFMSGLPNMIVMSPSNEDELIDMVATAAMVE

OsDXS3 : DRPFCFRYPKRGAVIGTSG-TLAYGNPLEIGKGEILVBC-KEIAEFLGYGVVQRCILARSILSNFGIQATVANARFCKPLDIDLIRLTCQCHSFLITVEEGTVGGFGSHVSOFTSLDGLLDGRLKWRPVLVLDPRYIDHGASL
SbDXS3 : DRPFCFRYPKRGAVIGTSG-TLYYGNPLEIGKGEILVBC-KEIAEFLGYGVVQRCILARSILSNFGIQATVANARFCKPLDIDLIRLTCQCHSFLITVEEGTVGGFGSHVSOFTSLDGLLDGRLKWRPVLVLDPRYIDHGASL
ZmDXS3 : DRPFCFRYPKRGAVIGTSG-SVTYGNPLEIGKGEILVBC-KEIAEFLGYGVVQRCILARSILSNFGIQATVANARFCKPLDIDLIRLTCQCHSFLITVEEGTVGGFGSHVSOFTSLDGLLDGRLKWRPVLVLDPRYIDHGASL
OsDXS2 : DRPFCFRYPKRGAVIGTSG-TLAYGNPLEIGKGEILVBC-KEIAEFLGYGVVQRCILARSILSNFGIQATVANARFCKPLDIDLIRLTCQCHSFLITVEEGTVGGFGSHVSOFTSLDGLLDGRLKWRPVLVLDPRYIDHGASL
SbDXS2 : DRPFCFRYPKRGAVIGTSG-TLYYGNPLEIGKGEILVBC-KEIAEFLGYGVVQRCILARSILSNFGIQATVANARFCKPLDIDLIRLTCQCHSFLITVEEGTVGGFGSHVSOFTSLDGLLDGRLKWRPVLVLDPRYIDHGASL
ZmDXS2 : DRPFCFRYPKRGAVIGTSG-SVTYGNPLEIGKGEILVBC-KEIAEFLGYGVVQRCILARSILSNFGIQATVANARFCKPLDIDLIRLTCQCHSFLITVEEGTVGGFGSHVSOFTSLDGLLDGRLKWRPVLVLDPRYIDHGASL
OsDXS1 : DRPFCFRYPKRGAVIGTSG-TLAYGNPLEIGKGEILVBC-KEIAEFLGYGVVQRCILARSILSNFGIQATVANARFCKPLDIDLIRLTCQCHSFLITVEEGTVGGFGSHVSOFTSLDGLLDGRLKWRPVLVLDPRYIDHGASL
SbDXS1 : DRPFCFRYPKRGAVIGTSG-TLYYGNPLEIGKGEILVBC-KEIAEFLGYGVVQRCILARSILSNFGIQATVANARFCKPLDIDLIRLTCQCHSFLITVEEGTVGGFGSHVSOFTSLDGLLDGRLKWRPVLVLDPRYIDHGASL
ZmDXS1 : DRPFCFRYPKRGAVIGTSG-SVTYGNPLEIGKGEILVBC-KEIAEFLGYGVVQRCILARSILSNFGIQATVANARFCKPLDIDLIRLTCQCHSFLITVEEGTVGGFGSHVSOFTSLDGLLDGRLKWRPVLVLDPRYIDHGASL

OsDXS3 : TEQIDMAGLTAHHIAATALTLLCRHRDAILLMMK---- (722)
SbDXS3 : AEQIDLAGLTAHHIAATALTLLCRHRDAILLMMK---- (722)
ZmDXS3 : AEQIDLAGLTAHHIAATALTLLCRHRDAILLMMK---- (722)
OsDXS2 : VDQIEEAGLTPRHIAATVLSLLCRPLEAOLK---- (713)
SbDXS2 : QDQMEEAGLTPRHIAATVLSLLCRPLEAOLK---- (727)
ZmDXS2 : QDQIEEAGLTPRHIAATVLSLLCRPLEAOLK---- (723)
OsDXS1 : ADQIAEAGLTPSHIAATVFNVLQAREALAMTVPNA (720)
SbDXS1 : ADQIAEAGLTPSHIAATVFNVLQAREALAMTVPNA (722)
ZmDXS1 : ADQIAEAGLTPSHIAATVFNVLQAREALAMTVPNA (719)

```

**Figure S1 Multiple-alignment of DXS proteins.** Amino acid sequences across available grass species (*Zea mays*, *Zm*; *Oryza sativa*, *Os*; *Sorghum bicolor*, *Sb*) were used to show the sequence signatures of transketolase and DXS enzymes (Rodriguez-Concepcion et al., 2003). These include thiamine di-phosphate binding domain (underlined red) and two transketolase motifs (underlined blue), with HIS residue for activity (blue triangle). Total number of aminoacids were marked at the end of each sequence in blue. The variant N-terminal amino acid sequence includes the transit peptide region. For sequence accessions, see Table S1.

```

ZmGGPPS1 : ...MAMSFHFHPLAASRVHFSPLPFAAAAAGAPSPSSVAIAAHQHGRRRFSAIVATAAASAAATTFDFKAYMGERAV
ZmGGPPS2 : .....MAFHFHPLIVASRVHFSPLPFAAATGAPSPSSVAIAAHYHHGRRRFSAIVATAT.....ITTFNFKAYMERAV
ZmGGPPS3 : MNKLASCFLOHGAPHTQIFKSYHVQRSPSLQLENRSVSMTRHRAADRAARCTIVDVAVDSG...TSFDEESYLSAKAR

ZmGGPPS1 : AVNRALDAAVPAGPEPPAALHDMRYALLAGGKVRPALCLAACAVVGGPEPWAMPAAAAVEMVHTMSLVHDDLPCMDDD
ZmGGPPS2 : AVNRALDAAI PAGGLPAALHDMRYALLAGGKVRPALCLAACAVVGGPEPWAMPAAAAVEMVHTMSLVHDDLPCMDDD
ZmGGPPS3 : AVHDALDLTLQGLRCEVLSSESMRYSVLGGKRLRPVLAIAACELVGGTAAAVPVACAVEMIHTASLIHDDMPCMDDDA
                                         (I)

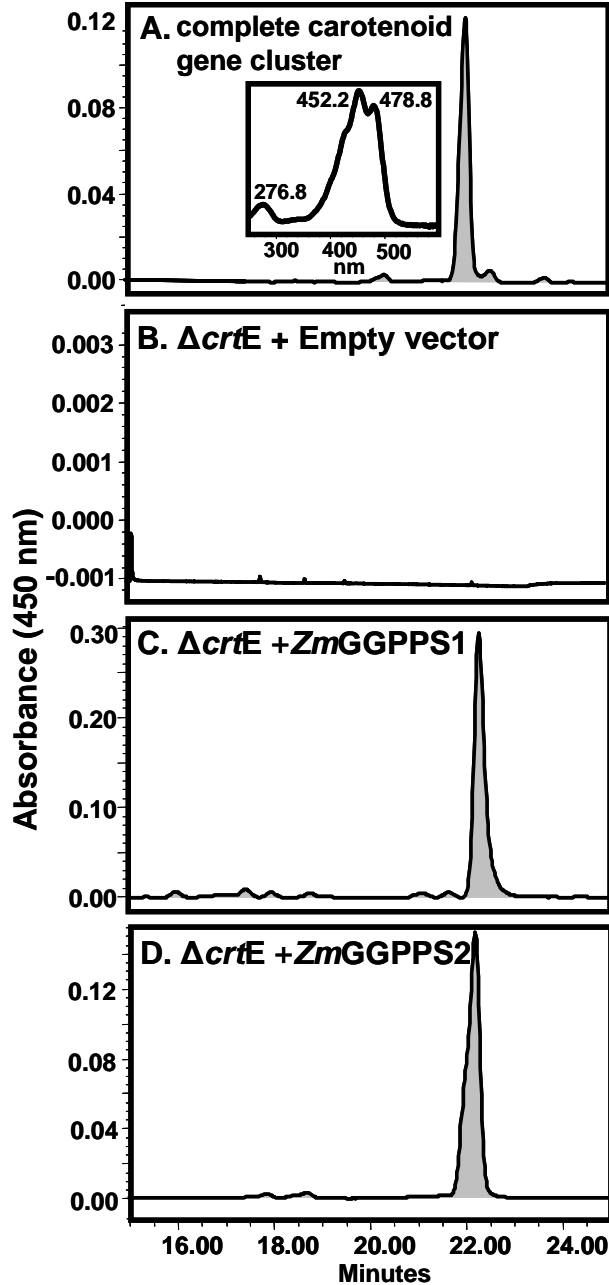
ZmGGPPS1 : LRRGKPTCHVYVGEPTAVLAGDALLSLSFHHMASVGSYPPDVDEKHPARVVRAIGELARCIGSEGLVAGQVVDLEMTGT
ZmGGPPS2 : LRRGKPTCHVYVGEPTAVLAGDALLSLSFHHMASVGSYPPAVDPEKHPARVVRAIGELARCIGSEGLVAGQVVDLEMTGT
ZmGGPPS3 : LRRGRPSNHVAFGEPTALLAGDALLALAFEHVARG.....SAGAGVPADRALRAVVELGSVAGVGGIAGOVADMASEGA
                                         (II)          (III)          (IV)

ZmGGPPS1 : .SETVPLEERLEYIHLHKTAALLEASVVI GAIIGGGTDEQIEERLKYARSIGLLFQVVDLDVTKSSEELGKTAGKDLAS
ZmGGPPS2 : .SOTVPLELLEYIHLHKTAALLEASVVI GAIIGGGTDEQIEERLKYARSIGLLFQVVDLDVTKSSEELGKTAGKDLAS
ZmGGPPS3 : PSGSVSLGCALEYIHVHKTARLVEAAAVSGAVVGGGDEGEVEVRRYAHFLGLLGQVVDVLDVDTGTSEQLGKTAGKDVA
                                         (V)          (VI)

ZmGGPPS1 : DKITYPKLLGLLEKSRFAEELSDAVTEQLACFDKEKAAPLHLANYIAHRQN
ZmGGPPS2 : DKITYPKLLGLAKSRFAEKLVS DATTEQLACFDEEKAAPLHLANYIAHRQN
ZmGGPPS3 : GKATYPRLMGLKGARAYMGELLAKAEALDGLDAARTAPLRHLARFMAHRQH
                                         (VII)

```

**Figure S2. GGPPS proteins of maize.** Blue lines indicate seven conserved domains among GGPPS gene family proteins that are highly conserved among prenyl diphosphate synthases with two Asp-rich motifs for substrate binding (domains II and VI) (Okada et al., 2000). Accession numbers: ZmGGPPS1, EF417573; ZmGGPPS2, EF417574; ZmGGPPS3, EF417575.



**Figure S3. Functional complementation of GGPPS1 and GGPPS2 from maize.** *E. coli* cells were transformed with A, pACCAR25 (for zeaxanthin accumulation); B, pACCAR25 $\Delta crtE$  (missing bacterial GGPPS)+ pET23a (empty vector); C, pACCAR25 $\Delta crtE$  + maize GGPPS1; D, pACCAR25 $\Delta crtE$  + maize GGPPS2. Chromatograms show HPLC separation of extracted pigments; inset in A shows spectral fine structure for the pathway end product, zeaxanthin diglucoside.

```

AtCrtISO : MDLCFQNPVKCGDR[F]SALNTSTYYKLGTSNLGFNGFVLENR[KKKKKLP]MVTVK[SVSSV]WASTVQGTKRD[GG]---[ESL]
OsCrtISO : ----MPL[LLLLAAARPCAPL]LAPSAAGAVVGR[ESTARPLGR]TRRGAC[AAAAAAVAAE]KTVVKAEE[EEEG]-----[GE]
SbCrtISO : -----MPPLAARV[HAPILLAP]---VAP-HEATRLCAAS[CFRGGVGGFRRGALASEKPPQAVAVA]KAGGEV[GGGEGEGP]
ZmCrtISO1 : -----MPPLAARL[HAPLLVLP]---AASPR[FEATRLCAAS]CFRG-VGGFRRGALASEKPPV[VAVA]KAGGE[GGGGQEGP]
ZmCrtISO2 : -----MPPLAARL[YAPLN]LAPAVPAASL[FEATRLCAAS]CFRN-VGGFRRGALA[*]EKAAAVAVA[KAGGEGGGEGDEGP]

AtCrtISO : YDAIVIGSGIGGLVAATQLA[VKEARVLVLEKYLIPGGSSGFYER]DGYTFDVGSSVMFGFSDKGNLNLITQAL[KAVGRKME]
OsCrtISO : YDAIVVSGIGGMVAATQLAAK[ARVLVLEKYVIPGSSGYRRDGF]TFDVGSSVMFGFSDKGNLNLITQALEAV[GHKME]
SbCrtISO : YDAIVIGSGIGGLVAATQLAAK[ARVLVLEKYLIPGGSSGYRRDGF]TFDVGSSVMFGFSDKGNLNLITQALEAV[GRKME]
ZmCrtISO1 : YDAIVIGSGIGGLVAATQLAAK[ARVLVLEKYLIPGGSSGYRRDGF]TFDVGSSVMFGFSDKGNLNLITQALEAV[GRKME]
ZmCrtISO2 : YDAIVIGSGIGGLVAATQLAAK[ARVLVLEKYLIPGGSSGYRRDGF]TFDVGSSVMFGFSDKGNLNLITQALEAV[GRKMR]

AtCrtISO : VIPDPSTVHFHLP[NNLSVRIHREYD]DFIAEL[ITSKFPHEKEGILKFYGD]C[KW]IFNSLNSLELKSLEEP[LYLF]Q[OFFOKPLE]
OsCrtISO : VIPDPSTVHFHLP[GDLSVLVHREYD]DFVTE[LVN]KFPHEKEGILKFYGT[CW]KIFNSLNSLELKSLEEP[LYLF]Q[OFFOKPLE]
SbCrtISO : VLPDPSTVHFHLP[GDLSVLVHRKYED]FINEL[ISKFPHEKEGILKFYD]T[CW]KIFNSLNSLELKSLEEP[LYLF]Q[OFFOKPLE]
ZmCrtISO1 : VLPDPSTVHFHLP[GDLSVLVHRKYED]FINEL[ISKFPHEKEGILKFYGT]C[W]KIFNSLNSLELKSLEEP[LYLF]Q[OFFOKPLE]
ZmCrtISO2 : VLPDPSTVHFHLP[GDLSVLVHRKYED]FINEL[ISKFPHEKEGILKFYD]T[CW]TIFNSLNSLELKSLEEP[LYLF]R[OFFOKPRE]

AtCrtISO : CLTLAYYLPQNAG[AIARKYIKD]Q[ELLSFIDAECFIVSTVNALQTPMINAS]MVLCDRH[YGGINYPVGGVGGI]AKSLA[E]GLV
OsCrtISO : CLTLAYYLPQNAG[DIARKFIK]DQ[ELLSFIDAECFIVSTVNALQTPMINAS]MVLCDRH[FGGINYPVGGVGGI]AMSLA[DG]LV
SbCrtISO : CLTLAYYLPQNAG[DIARKFIK]DQ[ELLSFIDAECFIVSTVNALQTPMINAS]MVLCDRH[FGGINYPVGGVGGI]IASLA[DG]LV
ZmCrtISO1 : CLTLAYYLPQNAG[DIARKFIK]DQ[ELLSFIDAECFIVSTVNALQTPMINAS]MVLCDRH[FGGINYPVGGVGGI]IALSLA[DG]LV
ZmCrtISO2 : CLTLAYYLPQNAG[DIARKFIK]DQ[ELLSFIDAECFIVSTVNALQTPMINAS]MVLCDRH[FGGINYPVGGVGGI]IASLA[DG]LV

AtCrtISO : DQ[GSEI]QYKANVKS[ITLDHGKAVGVRLAD]GREFFAKT[II]SNATRWDTFGKLLK[EK]LPKEE[ENFQKV]YVKAPSFLSIHMG
OsCrtISO : D[KGSEI]RYKANVTNVILENGKAVGVRLSNGKEFFAKT[VISNATRWDTFGKLLKVE]ELPEE[KNFQKN]YVKAPSFLSIHMG
SbCrtISO : E[KGSEI]RYKANVTNVILENGKAVGVRLSNGKEFFAKT[VISNATRWDTFGKLLKEK]ELPEE[KNFQKN]YVKAPSFLSIHMG
ZmCrtISO1 : E[KGSEI]RYKANVTNVILENGKAVGVRLSNGKEFFAKT[VISNATRWDTFGKLLKEK]ELPEE[KNFQKN]YVKAPSFLSIHMG
ZmCrtISO2 : E[KGSEI]RYKANVTNVILENGKAVGVRLSNGKEFFAKT[VISNATRWDTFGKLVK-]E[ELPEE]KNFQKNYVKAPSFLSIHMG

AtCrtISO : VKAEVLEP[PD]DCHHFVLEDDW[KNLE]E[YGSIFLSIPTVLDS]SLAPD[GRHILHIFTT]SSIE[DWEGI]PP[KEYE]AKKE[D]VAAR
OsCrtISO : VKASVLPADT[DCHHFVLEDDW]ANLEK[YGSIFLSIPTVLDS]SLAP[RGHHILHIFTTAGIE]DWEGLSRKDY[EKKKE]LVATE
SbCrtISO : VKASVLPAGT[DCHHFVLEDNWN]NLEK[YGSIFLSIPTVLDS]SLAP[RGHHILHVFTTAGIE]DWEGLSRKEYE[EKKE]EVANE
ZmCrtISO1 : VKASVLPAGT[DCHHFVLEDNWN]NLEK[YGSIFLSIPTVLDS]SLAP[RGHHILHIFTTAGIE]DWEGLSRKEYE[EKKE]EVANE
ZmCrtISO2 : VKASVLPAGT[DCHHFVLEDNWN]NLEK[YGSIFLSIPTVLDS]SLAP[RGHHILHIFTTAGIE]DWEGLSRKEYE[EKKE]EVANE

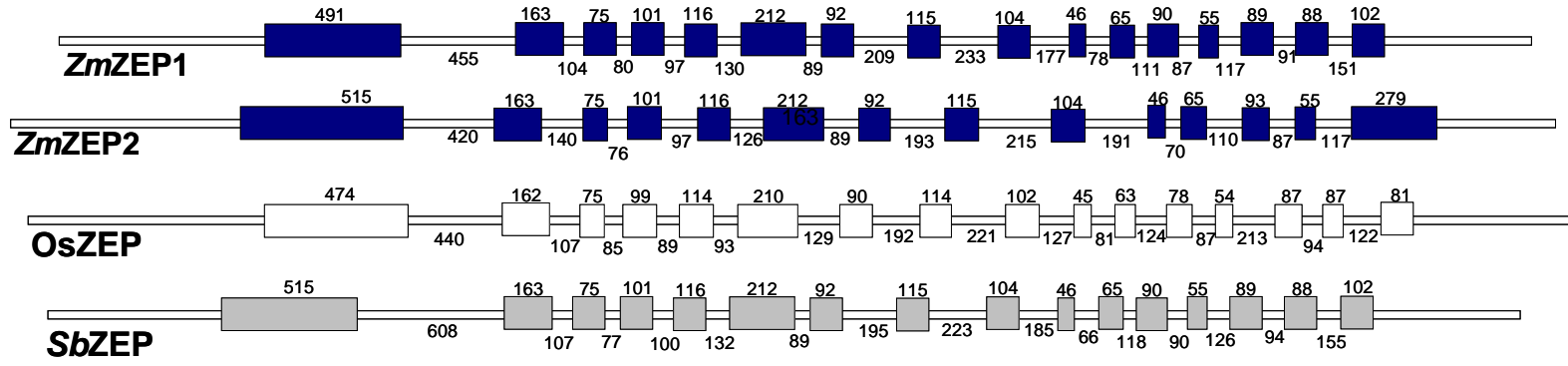
AtCrtISO : IIRLEK[KLFPGL]SS[ITFKEVGT]PRTHRRFLARND[GTYGMP]RGT[PKGL]L[MPFN]T[ADGLYCVGDSCFP]GQGVIAVA
OsCrtISO : IIRLEK[KLFPGL]QDSIVLKEVGS[PKTHRRFLARNDGTYGMP]RGT[PKGL]LAMPFN[TTSIDGLYCVGDSCFP]GQGVIAVA
SbCrtISO : IIRLEK[KLFPGL]QDSIVLKEVGS[PKTHRRFLARNDGTYGMP]RGT[PKGL]LAMPFN[TTSIDGLYCVGDSCFP]GQGVIAVA
ZmCrtISO1 : IIRLEK[KLFPGL]QDSIVLKEVGS[PKTHRRFLARNDGTYGMP]RGT[PKGL]LAMPFN[TTSIDGLYCVGDSCFP]GQGVIAVA
ZmCrtISO2 : IIRLEK[KLFPGL]QDSIVLKEVGS[PKTHRRFLARNDGTYGMP]RGT[PKGL]LAMPFN[TTSIDGLYCVGDSCFP]GQGVIAVA

AtCrtISO : FSGV[MC]AHRVAADIGLEK[KS]R[VLDV]GLLGLL[GW]LRTLA (595)
OsCrtISO : FSGI[MC]AHRVAADIGLEQ[RS]P[VLD]AGLLGLL[RW]LRTLA (586)
SbCrtISO : FSGI[MC]AHRVAADIGLEQ[KS]P[VLD]AGLLGLL[RW]LRTLA (588)
ZmCrtISO1 : FSGI[MC]AHRVAADIGLEQ[KS]P[VLD]AGLLGLL[RW]LRTLA (587)
ZmCrtISO2 : FSGI[MC]AHRVAADIGLEQ[KS]P[VLD]AGLLGLL[RW]LRTLA (588)

```

**Figure S4. Multiple-alignment of CrtISO protein sequences.** Amino acid sequences across available grass species (*Zea mays*, *Zm*; *Oryza sativa*, *Os*; *Sorghum bicolor*, *Sb*) and *Arabidopsis thaliana* (*At*) were used. Total number of amino acids were marked at the end of each sequence in blue. Accession numbers: **AtCrtISO**, At1g06820; **OsCrtISO**, LOC\_Os11g36440; **SbCrtISO**, Sb05g022240; **ZmCrtISO1**, AC218991; **ZmCrtISO2**, AC183901.

A



B

```

OsZEP : -----MAPLPSSSS-TRATSLVAIPGASCPHGQAQRLLAFR---PPRRGSQWRGLCVSRGRHATVAMAMPAAQAAG-RRARVLVAGGGIGGLVFALAARRKGFVVLVLERDMSAVRGEKGYRGPQLQSN
ZmZEP1 : MPTVLSTRRAASFETISLPSRS-SHSRTASPLHEVSRGN-RALRLLAALRSPSAPRPLRLRRPGLPPARAGLAAAAMPAPET----KARLLVAGGGIGGLVFALAARRKGFVVLVLERDMSAVRGEKGYRGPQLQSN
SbZEP : MPATLSTRAP-PRATSLPSSR---ASQAILLRHEVSRRN-RALRLLALP-SPSAPRRRSRPGLLPPARAGLVATAAMPAPPEP---KARVLVAGGGIGGLVFALAARRKGFVVLVLERDMSAVRGEKGYRGPQLQSN
ZmZEP2 : ---MLSTTWAS-PRKATPLPSSP---ASQAILLRHEVSSN-RTLRLVLP-SPSAPRRRR-RPGLPPARAGLVATAAMPAPPEP---KARVLVAGGGIGGLVFALAARRKGFVVLVLERDMSAVRGEKGYRGPQLQSN

OsZEP : ALAVLEAVDAGAADQVMDAGCITGNRVNGIVDGVSGSWYIKFDFTFPAERGLPVTRVISRMTLQQILARAVGDDAIIINGSHVVDFIDDGSKVTAILEDGRKFEGLLVGADGIWSKVRKTLFGHSDATYSYCYTGTGIA
ZmZEP1 : ALAALEAVDAAADEVMDAGCVTGDVRVNGIVDGVSGSWYIKFDFTFPAERGLPVTRVISRMTLQQILARAVGDDAIIINGSHVVDFIDDGSKVTAILEDGRKFEGLLVGADGIWSKVRKTLFGHSDATYSYCYTGTGIA
SbZEP : ALAALEAVDAAADEIMDAGCVTGDVRVNGIVDGVSGSWYIKFDFTFPAERGLPVTRVISRMTLQQILARAVGDDAIIINGSHVVDFIDDGSKVTAILEDGRKFEGLLVGADGIWSKVRKTLFGHSDATYSYCYTGTGIA
ZmZEP2 : ALAALEAVDAAADEIMDSGCVTGDVRVNGIVDGVSGSWYIKFDFTFPAERGLPVTRVISRMTLQQILARAVGDDAIIINGSHVVDFIDDGSKVTAILEDGRKFEGLLVGADGIWSKVRKTLFGHSDATYSYCYTGTGIA

OsZEP : DFVPPDIDTVGYRVFLGHKQYFVSSDVGAGKMQWYAFHKEPAGGIDPENGKKRLLLEIFDGWCDNVVDLINATDEEAILRRDIYDRPPTINWGKGRVTLTGDSVHAMQPNLQGGCMAIEDGYQLAVELENAWQESVKTG
ZmZEP1 : DFVPPDIDTVGYRVFLGHKQYFVSSDVGAGKMQWYAFHKEPAGGIDPENGKKRLLLEIFDGWCDNVVDLINATDEEAILRRDIYDRPPTINWGKGRVTLTGDSVHAMQPNLQGGCMAIEDGYQLAVELENAWQESVKTG
SbZEP : DFVPPDIDTVGYRVFLGHKQYFVSSDVGAGKMQWYAFHNEPAGGIDPENGKKRLLLEIFDGWCDNVVDLINATDEEAVLRRDIYDRPPTINWGKGRVTLTGDSVHAMQPNLQGGCMAIEDGYQLAVELENAWQESVKTG
ZmZEP2 : DFVPPDIDTVGYRVFLGHKQYFVSSDVGAGKMQWYAFHNEPAGGIDPENGKKRLLLEIFDGWCDNVVDLINATDEEAVLRRDIYDRPPTINWGKGRVTLTGDSVHAMQPNLQGGCMAIEDGYQLAVELENAWQESVKTG

OsZEP : TPMDIVSSLRRYEKERRLRVVIHGLARMAAIMATTYREYLVGLGPLSFLTKLRIPHPGRVGGRRFFIKYGMPLMSWVLGGNSKLEGRPLSCLSDKANDQLRRWFEDDDALEQAMGGEWYLFATSEANNNSLQOPIR
ZmZEP1 : TPMDIVSSLRRYEKERRVRVVIHGLARMAAIMATTYREYLVGLGPLSFLTKLRIPHPGRVGGRRFFIKYGMPLMSWVLGGNSKLEGRPLSCLSDKANDQLRRWFEDDDALEQAMGGEWYLFATSEANNNSLQOPIR
SbZEP : TPMDIVSSLRRYEKERRLRVVIHGLARMAAIMATTYREYLVGLGPLSFLTKLRIPHPGRVGGRRFFIKYGMPLMSWVLGGNSKLEGRPLSCLSDKANDQLRRWFEDDDALEQAMGGEWYLFATSEANNNSLQOPIR
ZmZEP2 : TPMDIVSSLRRYEKERRLRVVIHGLARMAAIMATTYREYLVGLGPLSFLTKLRIPHPGRVGGRRFFIKYGMPLMSWVLGGNSKLEGRPLSCLSDKANDQLRRWFEDDDALEQAMGGEWYLFATSEANNNSLQOPIR

OsZEP : LLRDEQRSLSVGSRSDANDSASSLSLPEQISERHATITCKNKAFYLDLHGEHGTWITDNEGRYRVPNFPVRFHPSDVEIFGSDKKAMFRVKVLTLPYESARSCKQOGLQOQEVLQAA
ZmZEP1 : LLRDEQRSLSVGSRSDANDSASSLSLPEQISERHATITCKNKAFYLDLHGEHGTWITDNEGRYRVPNFPVRFHPSDVEIFGSDKKAMFRVKVLTLPYESARSCKQOGLQOQEVLQAA
SbZEP : LLRDEQRSLSVGSRSDANDSASSLSLPEQISERHATITCKNKAFYLDLHGEHGTWITDNEGRYRVPNFPVRFHPSDVEIFGSDKKAMFRVKVLTLPYESARSCKQOGLQOQEVLQAA
ZmZEP2 : LLRDEQRSLSVGSRSDANDSASSLSLPEQISERHATITCKNKAFYLDLHGEHGTWITDNEGRYRVPNFPVRFHPSDVEIFGSDKKAMFRVKVLTLPYESARSCKQOGLQOQEVLQAA

```

**Figure S5. ZEP gene family structure.** (A) Gene organizational structure showing paralogs and orthologs of the ZEP gene family in *Zea mays* (*ZmZEP1*, AZM5\_13315; *ZmZEP2*, AZM5\_13312); *Oryza sativa* (*OsZEP*, LOC\_Os04g37619) and *Sorghum bicolor* (*SbZEP*, Sb06g018220) species representing two subfamilies of the Poaceae (grasses). Boxes and lines indicate exons and introns respectively, sizes for which are in base pairs (B) Multiple-alignment of ZEP paralogs in maize and orthologs from *Oryza sativa* (*Os*) and *Sorghum bicolor* (*Sb*).