Timing and Biosynthetic Potential for Carotenoid Accumulation in Genetically Diverse Germplasm of Maize$^{1[C]}[W][OA]$

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 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 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 carotenoid 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

---

$^1$ This work was supported by the National Institutes of Health (grant nos. 506-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) 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
controlling for endosperm carotenoids (Randolph and Hand, 1940; Palaisa et al., 2003; Gallagher et al., 2004; Wong et al., 2004; Pozniak 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

![Carotenoid Biosynthetic Pathway](image)

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, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate; MEP, 2-C-methyl-D-erythritol-4-phosphate.

| Table I. 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.]** |

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chromosome Bin</th>
<th>Content ($r$) Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dxs1</strong></td>
<td>6.95$^a$</td>
<td>-0.2 (0.5)</td>
</tr>
<tr>
<td><strong>Dxs2</strong></td>
<td>9.03$^a$</td>
<td>0.45 (0.21)</td>
</tr>
<tr>
<td><strong>Dxs3</strong></td>
<td>4.06$^a$</td>
<td>0.20 (0.5)</td>
</tr>
<tr>
<td><strong>Dxr</strong></td>
<td>3.04$^a$</td>
<td>0.41 (0.23)</td>
</tr>
<tr>
<td><strong>Hds</strong></td>
<td>5.03$^a$</td>
<td>-0.43 (0.36)</td>
</tr>
<tr>
<td><strong>Hol</strong></td>
<td>1.90$^a$</td>
<td>-0.37 (0.62)</td>
</tr>
<tr>
<td><strong>Ipp1</strong></td>
<td>7.03$^a$</td>
<td>-0.08 (0.81)</td>
</tr>
<tr>
<td><strong>Ipp2</strong></td>
<td>8.03$^a$</td>
<td>0.21 (0.54)</td>
</tr>
<tr>
<td><strong>Ipp3</strong></td>
<td>6.05$^a$</td>
<td>-0.12 (0.73)</td>
</tr>
<tr>
<td><strong>Cgpp1</strong></td>
<td>2.80$^a$</td>
<td>0.52 (0.11)</td>
</tr>
<tr>
<td><strong>Cgpp2</strong></td>
<td>7.04$^a$</td>
<td>-0.27 (0.47)</td>
</tr>
<tr>
<td><strong>Cgpp3</strong></td>
<td>8.01$^a$</td>
<td>NA</td>
</tr>
<tr>
<td><strong>Psid</strong></td>
<td>6.01$^a$</td>
<td>0.56 (0.003)</td>
</tr>
<tr>
<td><strong>Psid2</strong></td>
<td>8.07$^a$</td>
<td>-0.04 (0.89)</td>
</tr>
<tr>
<td><strong>Psid3</strong></td>
<td>7.03$^a$</td>
<td>0.03 (0.93)</td>
</tr>
<tr>
<td><strong>Pds</strong></td>
<td>1.02$^a$</td>
<td>0.12 (0.73)</td>
</tr>
<tr>
<td><strong>Zds</strong></td>
<td>7.02$^a$</td>
<td>-0.01 (0.95)</td>
</tr>
<tr>
<td><strong>Crtsso1</strong></td>
<td>4.00$^a$</td>
<td>-0.76 (0.001)</td>
</tr>
<tr>
<td><strong>Crtsso2</strong></td>
<td>2.09$^a$</td>
<td>NA</td>
</tr>
<tr>
<td><strong>Lcyb</strong></td>
<td>5.04$^a$</td>
<td>0.19 (0.58)</td>
</tr>
<tr>
<td><strong>LCYE</strong></td>
<td>8.05$^a$</td>
<td>0.18 (0.61)</td>
</tr>
</tbody>
</table>

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 S1, S1, 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 IPP (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.]
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 zetacarotenoid desaturase), LCYE (for lycopene ε-cyclase), and LCYB (for lycopene β-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
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...
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 \textit{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 \textit{DXS3}, \textit{DXR}, \textit{HDR}, and \textit{GGPPS1}, all at 25 DAP. Within the carotenoid pathway itself, \textit{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, \textit{PSY1} (\( r = 0.96, P = 0.001 \)), both at 20 DAP. The difference in \( r \) values for \textit{CrtISO1} and \textit{PSY1} is likely due to \textit{PSY1} having been under strong selection (Palaisa et al., 2003), whereas \textit{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 \textit{DXS3} and \textit{DXS1} both showed temporal increases during endosperm development, but only \textit{DXS3} showed correlative expression with carotenoids. Similarly, transcript levels of \textit{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 \textit{ZEP1} and \textit{ZEP2}

![Figure 5. Maize B73 inbred developing endosperm transcript profiles of genes encoding enzymes for isoprenoid and carotenoid biosynthesis. Means of three replicates ± sd are shown.](image-url)
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...
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 overexpression 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...
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 correlation 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 DXS 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) sequences (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 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 (Pampanwar et al., 2005) or at Maize GDB (www.maizegdb.org).

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 Tzadia, 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


Carretero-Paulet L, Cairó A, Botella-Pavia P, Besumbes O, Campos N, test function. These plasmids together with empty PET23a vector were transformed into Escherichia coli BL21 (DE3) cells containing pACCAR25ΔcrtE. Double transformants were grown in Luria-Bertani medium containing ampicillin and chloramphenicol and selected for their capacity to accumulate the yellow pigment zeaxanthin-β-d-glucoside. Pigments were extracted and analyzed by HFLC (Gallagher et al., 2004). The sense and antisense primers used to construct the respective plasmids are as follows: GGPPS1, 5′-GAATTCGTA- TAAAGCGCTGAACACTGTCG-3′ (#1000) and 5′-AACGTTCTCCGTGACAAA-GTCACACAGG-3′ (#1001); GGPPS2, 5′-GAATTCATGGCGTTTCACTTCCTCCTACCCCTGGTGC-3′ (#1005) and 5′-AACGTTAATGAGCAAGAAGACAAGAATAAG-3′ (#1004); and GGPPS3, 5′-GAATTCCTCATGAAACAGTTGGCAT-CCTGCTG-3′ (#1565) and 5′-AACGTTCTATGGCCTCATGTCGCGCATG-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 Tzadia, 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


Carretero-Paulet L, Cairó A, Botella-Pavia P, Besumbes O, Campos N,


Randolph LF, Hand DB (1940) Relation between carotenoid content and number of genes per cell in diploid and tetraploid corn. J Agric Res 60: 51–64


