nature genetics

Rare genetic variation at *Zea mays crtRB1* increases β-carotene in maize grain

Jianbing Yan[1–3,13], Catherine Bermudez Kandianis[4,5,13], Carlos E Harjes[3,12], Ling Bai[6], Eun-Ha Kim[7], Xiaohong Yang[2], Debra J Skinner[4,5], Zhiyuan Fu[2], Sharon Mitchell[3], Qing Li[2], Maria G Salas Fernandez[3,8], Maria Zaharieva[1], Raman Babu[1], Yang Fu[2,4,5], Natalia Palacios[1], Jiansheng Li[2], Dean DellaPenna[7], Thomas Brutnell[6], Edward S Buckler[3,9,10], Marilyn L Warburton[1,11] & Torbert Rocheford[4,5]

Breeding to increase β-carotene levels in cereal grains, termed provitamin A biofortification, is an economical approach to address dietary vitamin A deficiency in the developing world. Experimental evidence from association and linkage populations in maize (*Zea mays* L.) demonstrate that the gene encoding β-carotene hydroxylase 1 (*crtRB1*) underlies a principal quantitative trait locus associated with β-carotene concentration and conversion in maize kernels. *crtRB1* alleles associated with reduced transcript expression correlate with higher β-carotene concentrations. Genetic variation at *crtRB1* also affects hydroxylation efficiency among encoded allozymes, as observed by resultant carotenoid profiles in recombinant expression assays. The most favorable *crtRB1* alleles, rare in frequency and unique to temperate germplasm, are being introgressed via inexpensive PCR marker-assisted selection into tropical maize germplasm adapted to developing countries, where it is most needed for human health.

Vitamin A deficiency (VAD) leads to blindness in 250,000–500,000 children each year, with half dying from VAD-related illness within 12 months (data from the World Health Organization (WHO); see URL list in Online Methods section). Improving the micronutrient balance of staple crops such as maize through biofortification is therefore an economically and socially sound way to address micronutrient malnutrition, including VAD, on a global scale[1]. Considering factors of portion size, bioavailability and bioconversion, HarvestPlus (HP), a Consultative Group on International Agricultural Research Challenge program designed to use breeding for crop biofortification, projects that daily dietary intake of maize with 15 μg g^{-1} provitamin A carotenoids could greatly alleviate VAD[2].

A limited number of carotenoids, including α-carotene, β-carotene (βC) and β-cryptoxanthin (βCX), can be converted to vitamin A through animal metabolism[3]. A recent study combining information about carotenoid pathways from model organisms with natural variation for carotenoids in maize grain identified several haplotypes of the gene encoding lycopene epsilon cyclase (*lcyε*; also known as *LOC100280448* and *lyce1*) that substantially increase the ratio of β- to α-carotenoids in grain[4] (**Fig. 1a**). The favorable *lcyε* alleles increase the proportion of βC, but a large amount is hydroxylated to βCX and zeaxanthin (Z), which have 50% and 0% of the provitamin A activity of βC, respectively. Extrapolation from studies of *Arabidopsis* mutants defective in one or more of four carotenoid hydroxylase genes[5,6] indicated that natural genetic variation affecting corresponding biochemical reactions in maize might positively alter the βC/βCX + Z ratio (**Fig. 1a**), thus further increasing the provitamin A activity of maize grain. Toward this end, we identified maize orthologs of *Arabidopsis* β-carotene hydroxylases, and we surveyed and tested naturally occurring allelic variation to establish the molecular basis for reduced βC conversion.

RESULTS

crtRB1 is associated with β-carotene concentration

Three polymorphisms in *Zea mays crtRB1* (also known as *HYD3*) were significantly associated with carotenoid variation in association panel P1 (**Table 1**), identified as 5′TE, InDel4, and 3′TE (**Fig. 1b**; see Online Methods for details). PCR markers for these polymorphisms were assayed in association panels P1, P2 and P3 (**Supplementary Fig. 1**), which contain different sets of inbred lines with differing *crtRB1* allele frequencies. The 5′TE is significantly correlated with β-carotene concentrations accounting for 32% of phenotypic variation in P1, as well

[1]International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico. [2]National Maize Improvement Center of China, China Agricultural University, Beijing, China. [3]Institute for Genomic Diversity, Cornell University, Ithaca, New York, USA. [4]Department of Crop Sciences, University of Illinois, Urbana, Illinois, USA. [5]Department of Agronomy, Purdue University, West Lafayette, Indiana, USA. [6]Boyce Thompson Institute, Ithaca, New York, USA. [7]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan, USA. [8]Agronomy Department, Iowa State University, Ames, Iowa, USA. [9]United States Department of Agriculture–Agricultural Research Service: Plant, Soil and Nutrition Research Unit, Ithaca, New York, USA. [10]Department of Plant Breeding and Genetics, Cornell University, Ithaca, New York, USA. [11]United States Department of Agriculture–Agricultural Research Service: Corn Host Plant Resistance Research Unit, Starkville, Mississippi, USA. [12]Monsanto, Leesburg, Georgia, USA. [13]These authors contributed equally to this work. Correspondence should be addressed to T.R. (torbert@purdue.edu) and M.W. (marilyn.warburton@ars.usda.gov).

Received 23 September 2009; accepted 19 February 2010; published online 21 March 2010; doi:10.1038/ng.551

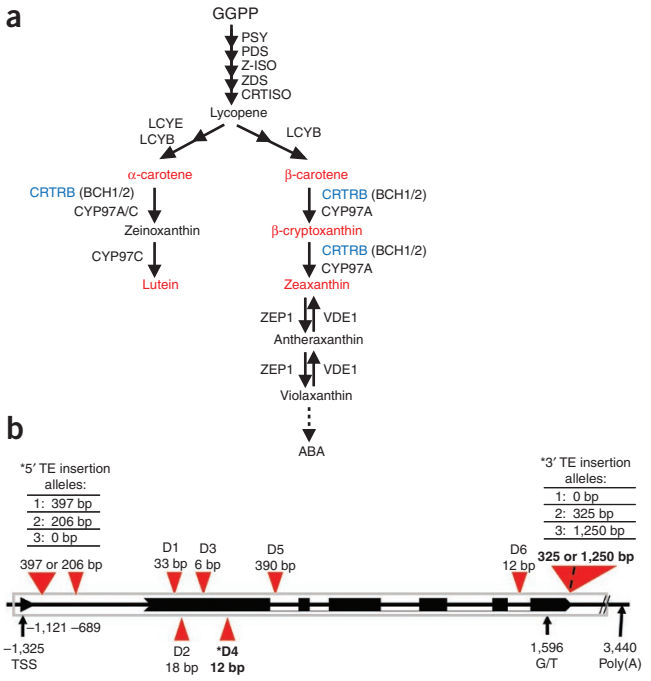


Figure 1 Carotenoid biosynthetic pathway and *Zea mays crtRB1* gene structure. (**a**) Simplified carotenoid biosynthetic pathway in maize and *Arabidopsis*[5,6,12,13] CRTRB, in blue, represents the nonheme di-iron β-carotene hydroxylase (BCH) family in maize, which has at least five members (**Supplementary Table 1**); the orthologous family in *Arabidopsis* has two members (BCH1 and BCH2). Carotenoid intermediates highlighted in red are compounds detected by HPLC in this study. (**b**) *Zea mays crtRB1* is the target gene in the present study. The sequenced region is framed in gray, translated exons are depicted as black boxes and the putative start of transcription (TSS) and poly(A) sites are indicated. Polymorphisms found in original P1 sequence alignments are marked in the diagram, and those that are significantly associated with changes in βC, βC/βCX, βC/Z and βC/ALL are labeled with asterisks. GGPP, geranylgeranyl pyrophosphate; PSY, phytoene synthase; PDS, phytoene desaturase; Z-ISO, ζ-carotene isomerase; ZDS, ζ-carotene desaturase; CRTISO, carotenoid isomerase; LCYE, lycopene ε-cyclase; LCYB, lycopene β-cyclase; CRTRB, β-carotene hydroxylase family; CYP97A, β-carotene hydroxylase (P450); CYP97C, ε-carotene hydroxylase (P450); ZEP1, zeaxanthin epoxidase; VDE1, violaxanthin de-epoxidase; ABA, abscisic acid.

as derived traits βC/βCX (60%), βC/Z (42%) and βC/ALL (i.e., βC + βCX + Z + α-carotene + lutein) (42%) in 2003. InDel4 and 3′TE explain 7–27% of phenotypic variation in these four traits (**Table 1** and **Supplementary Table 1**). Similar significant associations were found between these three polymorphisms and nearly all of these traits in each of the years tested in P1 (**Table 1**), as well as in P2 and P3 (**Supplementary Tables 2** and **3**). The 5′TE-InDel4 and the InDel4-3′TE polymorphisms were in linkage disequilibrium (LD) ($r^2 = 0.23$ and 0.12, respectively), but the 5′TE-3′TE polymorphisms were not ($r^2 = 0.02$). Variation in significance levels across years for each panel is partly due to differences in allele frequency and environmental effects during each growing season.

All three *crtRB1* polymorphisms affect conversion of βC to Z, but the 5′TE has the most pronounced effect (**Supplementary Table 4**), leading to an average increase of 6.50 μg g^{-1} βC above the average effect of the unfavorable allelic class (1.50 μg g^{-1}). The significant associations between the three polymorphisms with βC and βC/ALL show that the *crtRB1* allelic state influences both the absolute amount of βC and its proportion relative to total carotenoids, indicating that the effect of *crtRB1* is likely not attributable to an overall increase in carotenoid accumulation. The 206-bp insertion allele of 5′TE (allele 2; **Fig. 1b**), leading to higher βC concentrations, was detected only in temperate germplasm at low frequency (2.9%, absent and 1.9% in P1, P2 and P3, respectively). The frequencies of the three 3′TE alleles were more evenly distributed within populations, with the most favorable (allele 1, no insertion) present at 20.1%, 4.6% and 18.0% in P1, P2 and P3, respectively (**Supplementary Table 5**). *crtRB1* haplotypes were used to determine the joint effect of 5′TE, InDel4 and 3′TE allelic states on carotenoid phenotypes. Of the seven observed haplotype classes in P1, βC was highest when the favorable alleles of 5′TE (allele 2) and 3′TE (allele 1) were combined (haplotypes 2, 12, 1 and 2, 0, 1), shown in **Table 2**. The combination of three *crtRB1* polymorphisms accounted for 40% of the phenotypic variation for βC, 80% for βC/βCX, 68% for βC/Z and 50% for βC/ALL in P1 in 2003. In P3, the

Table 1 ***crtRB1*** **polymorphisms associated with carotenoid traits in the P1 diversity panel**

		Environment:	2002[c]	2003[c,d]	2004[c]	2005[c]	R^2 (2003)[e]	P (2003)[f]
Polymorphic site[a]	Alleles in series[b]	Average observation no.:	42	168	166	153	170	
5′TE	1/**2**/3	βC	n.s.	3.67×10^{-14}	3.25×10^{-7}	n.s.	32%	1.26×10^{-14}
		βC/βCX	0.0417	1.74×10^{-33}	n.s.	1.46×10^{-6}	60%	6.53×10^{-34}
		βC/Z	n.s.	4.72×10^{-21}	0.0163	3.53×10^{-4}	42%	1.27×10^{-20}
		βC/ALL	0.0653	3.99×10^{-20}	2.20×10^{-7}	0.0064	42%	2.73×10^{-20}
InDel4	**12**/0	βC	0.0498	2.40×10^{-4}	2.00×10^{-5}	n.s.	7%	5.65×10^{-4}
		βC/βCX	8.64×10^{-5}	4.90×10^{-13}	0.0248	0.0035	27%	1.12×10^{-13}
		βC/Z	0.0202	9.10×10^{-10}	1.52×10^{-4}	n.s.	21%	1.33×10^{-10}
		βC/ALL	0.0163	1.46×10^{-08}	5.0610^{-8}	0.0121	16%	3.30×10^{-8}
3′TE	**1**/2/3	βC	0.0034	6.11×10^{-6}	1.56×10^{-6}	0.0421.	10%	1.46×10^{-4}
		βC/βCX	4.94×10^{-4}	2.54×10^{-10}	n.s.	1.01×10^{-12}	26%	5.88×10^{-12}
		βC/Z	0.0828	8.16×10^{-7}	0.0017	3.53×10^{-6}	17%	1.77×10^{-7}
		βC/ALL	0.0215	7.86×10^{-10}	5.15×10^{-8}	8.94×10^{-5}	18%	6.06×10^{-8}

[a]Only significant polymorphic sites are shown. [b]Alleles in series are listed for each polymorphism, and favorable alleles (higher β-carotene content) are in boldface type. 5′TE allelic series: 1, 397-bp insertion; 2, 206-bp insertion; 3, 0-bp insertion. InDel4 allelic series: 12-bp or 0-bp insertion. 3′TE allelic series: 1, no insertion; 2, 325-bp insertion; 3, 1,250-bp insertion. [c]P value from association analysis carried out using the mixed model incorporating population structure and kinship, using data from 4 different years. [d]An outlier was excluded in all 2003 analyses. [e]R^2 values from analysis of variance (ANOVA) of 2003 data showing percentage phenotypic variation explained. [f]P value from ANOVA analysis of 2003 data.
n.s., not significant at $\alpha = 0.05$; βC, β-carotene μg g^{-1}; βC/βCX, β-carotene over β-cryptoxanthin; βC/Z, β-carotene over zeaxanthin; βC/ALL, β-carotene over total carotenoids (β-carotene, β-cryptoxanthin, α-carotene, zeaxanthin and lutein).

Table 2 ***crtRB1*** **haplotype estimated effects in P1 (2003 phenotypic data)**

	crtRB1				Traits[b]							
Haplotype[a]	5′TE	InDel4	3′TE	*N*	βC	s.e.m.	βC/βCX	s.e.m.	βC/Z[b]	s.e.m.	βC/ALL	s.e.m.
1,0,**1**	1	0	**1**	28	2.15	1.56	2.32	1.16	0.41	0.27	0.1	0.06
1,0,2	1	0	2	15	0.88	0.71	0.64	0.32	0.10	0.07	0.03	0.02
1,0,3	1	0	3	109	1.40	1.42	0.95	0.61	0.18	0.29	0.06	0.05
1,**12**,**1**	1	**12**	**1**	6	1.91	0.95	4.09	2.48	1.24	1.19	0.10	0.04
2,0,**1**	**2**	0	**1**	1	13.34	–	19.06	–	3.79	–	0.42	–
2,**12**,**1**	**2**	**12**	**1**	5	6.79	4.77	10.18	4.00	2.18	1.83	0.39	0.22
3,0,3	3	0	3	5	2.64	1.21	2.92	1.10	0.40	0.11	0.11	0.04
R^2					40%		80%		68%		50%	
P (ANOVA)					5.23×10^{-16}		1.68×10^{-53}		1.57×10^{-38}		1.02×10^{-21}	
Avg. change between "**2**,**12**,**1**" and "1,0,2"[c]					7.6		15.2		18.3		10.5	

[a]Haplotype is shown as linear combination of 5′TE allele (1, 397-bp insertion; 2, 206-bp insertion; 3, 0-bp insertion), InDel4 allele (12-bp or 0-bp insertion), 3′TE allele (1, no insertion; 2, 325-bp insertion; 3, 1,250-bp insertion). Favorable alleles are in boldface type, and only observed haplotypes are listed. [b]Traits are described in **Table 1**. [c]Comparison between best and worst haplotypes as predicted by component allelic effects.

six haplotype classes resulting from combinations of 5′TE, InDel4 and 3′TE explained 62–95% of the variation for the same derived traits (**Supplementary Table 6**).

We observed a 7.6-fold change in the average βC values between the most and least favorable haplotype classes in P1 in 2003 (**Table 2**) and a 10.5-fold change for the same comparison in P3 in 2006 (**Supplementary Table 6**). Haplotypes containing the favorable 5′TE and 3′TE alleles ($n = 6$) had an average βC concentration of 7.88 (±5.03) μg g^{-1}, in contrast to all other haplotypes ($n = 165$) averaging 1.53 (±1.41) μg g^{-1}. This 6.34 μg g^{-1} difference represents 42% of the HP target goal and indicates that *crtRB1*-specific allelic variation yields a marked increase in maize grain βC concentration. The derived traits of βC/βCX, βC/Z and βC/ALL, respectively, showed 15.2-, 18.3- and 10.5-fold change between the most and least favorable haplotype classes in P1 (**Table 2**), and 43.3-, 94.7- and 14.5-fold changes in P3 (**Supplementary Table 6**). Derived traits were not calculated for P2, as the frequency of 5′TE and InDel4 were too low to be analyzed.

crtRB1 maps to a principal QTL for β-carotene

We used five populations consisting of recombinant inbred lines (RILs) or $F_{2:3}$ progenies (see Online Methods) that segregate for *crtRB1* polymorphisms (parental genotypes listed in **Supplementary Table 7**) for quantitative trait locus (QTL) mapping (**Supplementary Table 8**). In the B73 × BY804 RIL population, *crtRB1* mapped to a genetic interval containing a principal QTL for βC concentration and βC/ALL explaining 16.3% and 32.8% of the trait variation, respectively (**Supplementary Fig. 2** and **Supplementary Table 8**). Principal QTLs for βC and other β-branch traits also mapped to an interval containing *crtRB1* in DEexp × CI7 and A619 × SC55 $F_{2:3}$ populations (**Supplementary Table 8**). The effect of *crtRB1* on βC concentration was also further confirmed through single-factor analysis in the KI3 × SC55 and KI3 × B77 $F_{2:3}$ populations (analysis of variance (ANOVA), $\alpha = 0.01$, $P < 0.001$; **Supplementary Table 8**). QTL for variation in βC did not map to the four other BCH maize orthologs.

Three populations (A619 × SC55, KI3 × SC55 and KI3 × B77) segregate for at least one significant polymorphism in both *crtRB1* and *lcyε*. To estimate *crtRB1* effects with a single-gene model (**Table 3**), we adjusted for *lcyε* effects and found *crtRB1* to explain significant variation in β-branch traits including βC (R^2 range 8–44%), βC/βCX (R^2 range 37–69%) and βC/ALL (R^2 range 15–60%). Favorable homozygous genotypes lead to β-carotene increases of 2.78 μg g^{-1} in KI3 × SC55, 1.61 μg g^{-1} in KI3 × B77 and 0.53 μg g^{-1} in A619 × SC55 above βC concentrations of the respective unfavorable homozygotes, which averaged 2.79, 0.76 and 4.62 μg g^{-1} (**Table 3**). Additive effects for *crtRB1* (ANOVA, $\alpha = 0.05$, $P < 0.001$) in all three populations were associated with maximal per-copy increases in βC of 1.25 μg g^{-1}, as estimated through linear regression (**Supplementary Table 9**). Collectively, these results indicate that favorable *crtRB1* alleles can lead to higher βC concentrations across a range of segregating genetic backgrounds.

crtRB1 genotypes associated with higher βC were also associated with a decrease in all other tested carotenoids and total carotenoid concentration (**Fig. 2**). Comparing favorable homozygous genotypes with unfavorable homozygous genotypes at *crtRB1* showed that favorable *crtRB1* homozygotes resulted in an 8.9–48% reduction in total carotenoids ($P < 0.0001$). Reduction in total carotenoids is possibly conferred by the favorable 5′TE allele, in that significant variation in total carotenoids was most often observed in populations polymorphic at 5′TE as compared with InDel4 or 3′TE (**Supplementary Table 7**).

Table 3 ***crtRB1*** **genetic effects in three segregating populations**

Copies of favorable haplotype	*N*			Traits[b] βC[a]			βC/ALL[a]			βC/Z[a]			βC/βCX[a]		
	A	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*
2	32	33	12	5.15†	2.37	5.57	0.47	0.17	0.46†	8.98	0.33	9.65	12.33	2.25	5.80
1	114	42	26	5.02†	1.49	3.39†	0.41	0.07	0.28†	4.83	0.10	1.50	7.28	0.63	1.63
0	50	35	28	4.62	0.76	2.79†	0.38	0.03	0.17	3.60	0.04	0.62	5.4	0.26	0.77
R^2 (%)				7.8	43.5	25.4	17.0	59.6	49.0	15.0	63.1	46.8	36.8	58.1	69.3
Actual difference between homozygotes				0.53	1.61	2.78	0.09	0.14	0.29	5.38	0.29	9.03	6.93	1.99	5.03
Relative change between homozygotes				1.11	3.11	2.00	1.23	5.67	2.70	2.49	8.25	15.56	2.28	8.65	7.53

[a]*crtRB1* genotypic class means for βC (μg g^{-1}), βC/ALL, βC/Z and βC/βCX traits. All effects are adjusted by *lcyε* covariates. Adjusted means are all significantly different at $\alpha = 0.05$, except where noted (†). [b]Traits are described in **Table 1**. *A*, population A619 × SC55; *B*, population KI3 × B77; *C*, population KI3 × SC55.

crtRB1 and *lcyε* show additive effects

Carotenoid traits associated with the β-branch are affected by allelic variation at both *lcyε* (ref. 4) and *crtRB1*. Because *lcyε* also explained a significant amount of variation in the proportion of βC concentration in P1 (ref. 4), P2 and P3 (**Supplementary Tables 10** and **11**), we evaluated the combined effects on βC concentrations of one main functional polymorphism of *lcyε* (5′TE) and two of *crtRB1* (5′TE and/or 3′TE) in three association panels; polymorphisms with highest estimated effects and lowest LD were selected. A greater proportion of the phenotypic variation in βC and βC/ALL (52% and 65%, respectively; **Table 4**) was explained by the combined *crtRB1*/*lcyε* polymorphisms than by those of *crtRB1* alone (40% and 50%; **Table 2**) in P1 in 2003. A similar pattern was seen for βC, βC/βCX and βC/ALL in P2 (**Supplementary Tables 2** and **12**). Combined *lcyε* and *crtRB1* effects at each polymorphism were largely additive (data not shown). No inbred haplotype in the association panels combines the most favorable haplotypes for *crtRB1* 5′TE, *crtRB1* 3′TE and *lcyε* 5′TE. However, between the existing common haplotypes predicted to yield the most and least βC, we observed a 12.1-fold difference in βC concentration and a 19.3-fold difference in βC/ALL in P1 (**Table 4**) as well as 12.9- and 17.8-fold differences in βC concentration and βC/ALL, respectively, in P3 (**Supplementary Table 13**). Increases in βC and provitamin A derived from the more favorable haplotype in P1 averaged 5.91 μg g^{-1} (39% of HP target goal) and 5.65 μg g^{-1} (38% of HP target goal), respectively. Notably, no significant epistasis for βC was detected in association panels or mapping populations (data not shown).

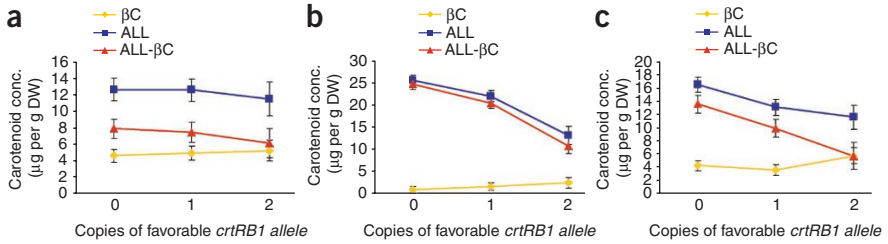


Figure 2 Mean carotenoid concentration (μg g^{-1} dry weight, DW) for *crtRB1* allele classes across three $F_{2:3}$ populations. Series represent genotype class average concentrations of βC, ALL and ALL-βC (that is, total carotenoids minus βC). (**a**) A619 × SC55; (**b**) KI3 × B77; (**c**) KI3 × SC55. Error bars, s.e.m.

Expression analysis

Previous studies have shown the importance of transcript abundance in the control of carotenoid profiles[4,7,8]. We examined *crtRB1* transcripts from kernels harvested at 15 days after pollination (DAP) in six inbred lines (A619, B77, CI7, Hi27, NC320 and NC356), which varied by an order of magnitude in their seed βC concentration[4]. Peak expression of carotenoid biosynthetic genes and crtRB1 was noted to occur within 15–20 DAP in other studies[7,9]. In lines containing relatively high concentrations of βC (CI7 and B77; identical *crtRB1* haplotypes), *crtRB1* transcripts accumulated to only 1/70 of the level of *crtRB1* transcripts in lines with low βC (NC356, Hi27 and NC320; **Fig. 3a**). Relative to all tested lines, transcript and βC concentrations were both intermediate in A619, which carries only the 3′TE insertion. This implies that polymorphisms associated with 5′TE, 3′TE and InDel4 result in substantial transcript reduction in kernels and probably decrease β-hydroxylase activity through reduction of CRTRB1 protein levels.

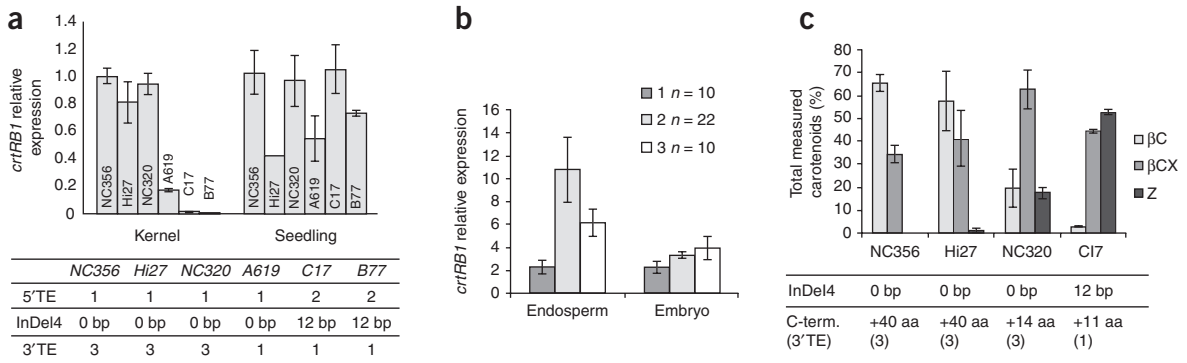
To test the independent effect of 3′TE alleles on *crtRB1* expression, we determined embryo and endosperm profiles at 20 DAP for 42 additional lines monomorphic at the 5′TE and InDel4 but polymorphic for the 3′TE (**Fig. 3b** and **Supplementary Table 14**). In endosperm, *crtRB1* allele-specific expression levels inversely matched the allelic effect on βC concentration (ANOVA, $\alpha = 0.01$): lines with the most favorable 3′TE allele had the lowest expression, whereas lines with the least favorable allele had the highest expression (**Fig. 3b**). This trend was less striking in embryos, in which only the most favorable allele showed significantly lower expression than the other two. Seedling leaf expression showed no significant difference between 3′TE allele classes (**Fig. 3a**) and no correlation with kernel carotenoid composition, which suggests tissue-specific regulation of *crtRB1*.

To further explore the molecular basis of *crtRB1* variation on carotenoid profiles, we tested four *crtRB1* alleles encoding differing allozymes for hydroxylation activity in an *Escherichia coli* assay[10]. The selected genotypes differed in allelic states at InDel2 (not associated with carotenoid concentration), 5′TE/InDel4 (with complete LD between the two polymorphisms in the lines tested) and 3′TE (**Supplementary Fig. 3a**). The Hi27 allozyme (identical in protein

Table 4 Estimated effects of combined haplotypes of *crtRB1* and *lcyε* genes in P1 (2003 phenotypic data)

	crtRB1		*lcyε*		Traits[d]							
Haplotype[a]	5′TE	3′TE	5′TE	*N*	βC	s.e.m.	βC/βCX	s.e.m.	βC/Z	s.e.m.	βC/ALL	s.e.m.
1,**1**,2	1	**1**	2	9	3.31	1.98	2.44	1.46	0.87	1.04	0.15	0.08
1,**1**,3	1	**1**	3	24	1.65	0.96	2.72	1.67	0.45	0.34	0.07	0.03
1,2,2	1	2	2	7	1.25	0.85	0.72	0.35	0.09	0.05	0.05	0.02
1,2,3[b]	1	2	3	8	0.56	0.38	0.58	0.31	0.11	0.09	0.02	0.01
1,3,2	1	3	2	47	1.65	1.22	1.06	0.49	0.27	0.41	0.07	0.05
1,3,3	1	3	3	48	0.81	1.10	0.88	0.76	0.11	0.07	0.03	0.02
1,3,**1/4**	1	3	**1/4**[c]	11	2.71	2.12	0.76	0.23	0.15	0.08	0.08	0.04
2,**1**,2[b]	**2**	**1**	2	5	6.79	4.77	10.18	4.00	2.18	1.83	0.39	0.22
2,**1**,3	**2**	**1**	3	1	13.34	–	19.06	–	5.56	–	0.42	–
3,3,2	3	3	2	4	2.79	1.34	3.03	1.24	0.38	0.12	0.12	0.03
3,3,3	3	3	3	1	2.05	–	2.49	–	0.48	–	0.06	–
Maximum change				165	12.1		17.6		19.8		19.3	
R^2 (ANOVA)					52%		80%		61%		65%	
P					2.77×10^{-20}		3.15×10^{-49}		5.05×10^{-27}		5.61×10^{-30}	

[a]Haplotype is shown as linear combination of *crtRB1* 5′TE allele (1, 397-bp insertion; 2, 206-bp insertion; 3, 0-bp insertion), *crtRB1* 3′TE allele (1, no insertion; 2, 325-bp insertion; 3, 1,250-bp insertion), *lcyε* 5′TE allele (1, 150 + 280-bp insertion; 2, 250-bp insertion; 3, 250 + 380-bp insertion; 4, 993-bp insertion. Favorable alleles are in boldface type, and only observed haplotypes are listed. [b]Haplotypes predicted to yield most (**2**,**1**,2) and least (1,2,3) β-carotene from available classes. [c]Allelic classes of 1 and 4 (1/4) at *lcyε* 5′TE allele were combined because both classes yield a similar effect on carotenoid traits. [d]Traits are described in **Table 1**.



	NC356	Hi27	NC320	A619	C17	B77
5′TE	1	1	1	1	2	2
InDel4	0 bp	0 bp	0 bp	0 bp	12 bp	12 bp
3′TE	3	3	3	1	1	1

	NC356	Hi27	NC320	CI7
InDel4	0 bp	0 bp	0 bp	12 bp
C-term. (3′TE)	+40 aa (3)	+40 aa (3)	+14 aa (3)	+11 aa (1)

Figure 3 Allele-specific *crtRB1* effects on biochemical activity and transcriptional expression. (**a**) *crtRB1* quantitative RT-PCR from whole kernel at 15 d after pollination (DAP) and seedling leaf messenger RNA for the six indicated lines of *Zea mays*. Values are given as expression levels relative to kernel and seedling values for line NC356, most highly expressed in this set. Error bars are s.e.m.; all measurements are based on three replications except Hi27, which was unreplicated. Genetic variation for each inbred line is listed below according to 5′TE, InDel4 and 3′TE differences; allele codes can be found in **Table 2** and **Supplementary Figure 2**. (**b**) *crtRB1* quantitative RT-PCR for endosperm and embryo at 20 DAP from 42 lines differing in the 3′TE allele. Values are given as expression levels relative to endosperm and embryo values for line BY804 because it was the highest expresser in this panel. Error bars are s.e.m. 1, lines with no 3′TE insertion; 2, lines with 325-bp 3′TE insertion; 3, lines with 1,250-bp 3′TE insertion. (**c**) β-carotene hydroxylase product profiles for the four CRTRB1 allozymes expressed in a recombinant *E. coli* assay system producing β-carotene. Accumulated carotenoids are expressed as percentage of total carotenoids. βC, β-carotene; βCX, β-cryptoxanthin; Z, zeaxanthin. Error bars are s.d. Genetic variation for each allozyme is listed below according to InDel4 and C-terminal (3′TE) differences; allele codes can be found in **Table 2** and **Supplementary Figure 2**.

sequence to the B73 reference, **Supplementary Fig. 3b**) has no insertions in the coding region but maintains the extended, TE-encoded C terminus, whereas NC356 has the same C terminus and a 6-amino-acid insertion (InDel2) in the N terminus of the protein. NC320 also carries the 6-amino-acid insertion (InDel2) but varies from NC356 by a stop codon truncating 26 amino acids from the C terminus. The CI7 allele (which is associated with high concentrations of βC) carries the 6-amino-acid InDel2 insertion, the 4-amino-acid InDel4 insertion and an alternative C terminus encoded by the 3′TE polymorphism. These four allozymes were expressed in *E. coli* engineered to accumulate βC as a substrate[10], and the resulting hydroxylation products were analyzed by HPLC. All four allozymes showed β-carotene hydroxylase activity but differed in their product profiles and specificity for each β-ring (**Fig. 3c** and **Supplementary Fig. 4**).

The Hi27 and NC356 allozymes produced βCX (monohydroxy) as their chief hydroxylation product (34–40%) and little or no Z (dihydroxy), indicating that these CRTRB1 variants preferentially introduce a single hydroxyl group into βC. The NC320 allozyme accumulated βCX as the main product (63%) but also substantial amounts of Z (18%), suggesting that truncation of the C terminus in this allozyme positively affects Z production. Unexpectedly, the favorable allele for βC accumulation in maize grain, CI7, encodes an allozyme that converted nearly all βC to Z (53%) and βCX (45%).

DISCUSSION

Three polymorphisms in *crtRB1* (5′TE in the 5′ UTR, InDel4 in the coding region and 3′TE spanning the sixth exon and 3′ UTR) were shown to associate with βC concentration and conversion to βCX and Z in maize kernels. These effects led to average βC concentrations of 2.91–8.00 μg g^{-1} in germplasm with the favorable allelic composition at individual polymorphisms, which translates into *crtRB1*-specific increases of 1.45–6.50 μg g^{-1} provitamin A meeting up to 43% of the target HP goal. The functional basis of *crtRB1* allelic effects was investigated at the transcript and enzymatic activity levels. One level of transcript regulation is probably mediated through the 5′ and 3′ UTR polymorphisms, in that allelic differences at both the 5′TE and 3′TE polymorphisms conditioned significant changes in messenger RNA levels inversely proportional to βC concentration. Altering the C terminus of the protein was shown in *E. coli* to enhance hydroxylation of βC and broaden the product profile to include Z. Notably, the CI7 allozyme, which carries the favorable 5′TE and InDel4 insertions and the shorter C terminus, showed the strongest hydroxylation activity when assayed in *E. coli* but conditioned high βC concentrations in kernels, which correlated with the extremely reduced expression of *crtRB1* in CI7 kernels. This implies that βC concentrations in CI7 are high as a result of weak kernel expression, despite the potential of the CI7 allozyme for considerable βC hydroxylation activity.

The rarity of the favorable 5′TE and InDel4 insertion alleles, and their complete linkage with the favorable 3′TE allele, necessitates further experimentation to isolate the effects associated with each polymorphism. Combined results for the three polymorphisms validate our hypothesis that reduced enzyme presence either through amount or activity would diminish conversion of βC to downstream products and contributes to the mechanistic basis for favorable allelic states of *crtRB1*. Our study further supports the emerging concept that natural variation for carotenoid concentrations in maize kernels is largely regulated at the level of gene expression, as seen in studies of *PSY* (ref. 11), *lcyε* (ref. 4), *lcyβ* (ref. 7) and *HYD3/crtRB1* (ref. 9).

Another *crtRB1* 5′ UTR polymorphism was recently reported to correlate with βC concentrations[9], but the analysis involved only a small subset of P1 ($n = 51$). When we assessed this polymorphism in complete P1 and P3 panels (**Supplementary Table 15**), we found that it was only weakly associated with βC. This implies that the association reported by Vallabhaneni *et al.*[9] was probably attributed to considerable LD between their 5′ UTR polymorphism and the 5′TE and InDel4 polymorphisms because the LD with these polymorphisms was substantially reduced with increasing panel sample size (**Supplementary Table 16**).

Association and segregation mapping evidence showed a consistent effect of *crtRB1* on βC concentration. However, joint additive effects for *crtRB1* and *lcyε* on βC were detected only by association mapping, whereas detection of *crtRB1* effects on total carotenoids was only observed in segregating populations but not association panels. Statistical detection of these genetic effects probably results from several factors, including better phenotypic resolution in association panels and reduced genetic background complexity in

biallelic segregating populations. This study affirms that association and segregating populations can successfully cross-validate QTL effects and highlights the usefulness of leveraging the genetic complexity of diverse populations to further elucidate the genetic basis of phenotypes. Breeders can draw upon results from both association and segregating populations to make more informed decisions.

Selection for favorable *crtRB1* alleles at the three polymorphisms (5′TE, InDel4 and 3′TE) is effective, increasing average βC concentrations in segregating populations regardless of genetic background or dosage. Haplotypes containing favorable variation at both 5′TE and 3′TE polymorphisms yielded average increases in βC and provitamin A of 6.34 and 6.12 μg g^{-1}, respectively, above the average of all other haplotype classes in P1. Because the provitamin A mean of the most favorable combined class (8.57 ± 0.89 μg g^{-1}) represents 57% of the 15 μg g^{-1} provitamin A target value set by HP for enhanced human health, it is evident that substantial nutritional enhancement can be achieved by allelic variation at *crtRB1*. Specific haplotypes of *lcyε* have been shown to yield the most favorable α-carotenoid/β-carotenoid ratios and βC proportions of all measured classes[4], contributing to the joint additive effects of *crtRB1*and *lcyε* on βC concentrations observed in this study. The rarity of certain genetic variants is such that the most favorable haplotypes of *crtRB1* and *lcyε* do not naturally occur together. Experiments to combine the best haplotypes for both loci to evaluate the combined genetic effects in breeding crosses are continuing.

The temperate hybrid CI7 × DEexp, used in human and animal nutrition studies (S. Tanumihardjo, personal communication), produces a confirmed level of 15 μg g^{-1} provitamin A (T.R.R., unpublished data). This hybrid, which contains the most favorable *crtRB1*alleles, indicates that HP targets are attainable in the presence of favorable *crtRB1* variation. Through the use of donor lines (**Supplementary Table 17**) and codominant, user-friendly PCR-based marker systems (**Supplementary Fig. 1**), introgression of favorable *crtRB1* and *lcyε* alleles into tropical germplasm has also achieved HP provitamin A target concentrations in preliminary evaluations of breeding lines (CIMMYT; K. Pixley, personal communication). Education and outreach programs are underway to facilitate consumer acceptance of biofortified orange grain and to help address vitamin A deficiencies throughout the developing world (HP).

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Accession codes. GenBank sequences have been deposited under the accession numbers GQ889501–GQ889872.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank K. Pixley and J. Chandler for seed and field experiment coordination and W. White, P. Bermudez, S. Islam, C. Paul, W. Liu, S. Xu and Y. Zhou for carotenoid HPLC profiling. Helpful comments on the manuscript from J. Wilkinson and J. Yu are appreciated. This research was supported by the US Agency for International Development (T.R.), HarvestPlus (T.R., D.D.P., J.Y., J.L., M.L.W.), National Science Foundation (NSF) DBI-0321467, DBI-0820619 (E.S.B.), NSF DBI-0604923 (T.R.), USDA-ARS (E.S.B.), NSF DBI-0501713 (T.B.), TRIAD Foundation (T.B.), China NSF-30821140352 (J.L., J.Y.), China Scholarship Fund (Y.F.) and JBT Fellowship-UIUC (C.B.K.).

AUTHOR CONTRIBUTIONS

C.E.H. and J.Y. identified the gene. X.Y., Z.F., Y.F., R.B., C.B.K., J.Y., M.G.S.F., M.Z. and S.M. carried out the sequencing and genotyping. L.B., E.-H.K. and X.Y. carried out the transcript expression and biochemical assays. J.Y. and D.J.S. developed the *crtRB1* molecular markers. R.B. and J.Y. supervised the field testing. C.B.K., Z.F., Q.L. and N.P. carried out the carotenoid profiling. C.B.K. and X.Y. completed the genetic mapping and QTL analyses. J.Y. and C.B.K. carried out the association and genetic analyses. The study was designed and supervised by J.Y., J.L., D.D.P., T.B., E.S.B., M.L.W. and T.R. The manuscript was prepared by J.Y., C.B.K., D.J.S., M.L.W. and T.R. and was edited by D.D.P., T.B. and E.S.B.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Tanumihardjo, S.A., Bouis, H., Hotz, C., Meenakshi, J.V. & McClafferty, B. Biofortification of staple crops: an emerging strategy to combat hidden hunger. *Compr. Rev. Food Sci. Food Safety* **7**, 329–334 (2008).
2. Pfeiffer, W.H. & McClafferty, B. HarvestPlus: breeding crops for better nutrition. *Crop Sci.* **47**, S88–S105 (2007).
3. Fierce, Y. *et al. In vitro* and *in vivo* characterization of retinoid synthesis from β-carotene. *Arch. Biochem. Biophys.* **472**, 126–138 (2008).
4. Harjes, C.E. *et al.* Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science* **319**, 330–333 (2008).
5. Sun, Z.R., Gantt, E. & Cunningham, F.X. Cloning and functional analysis of the β-carotene hydroxylase of *Arabidopsis thaliana*. *J. Biol. Chem.* **271**, 24349–24352 (1996).
6. Tian, L. & DellaPenna, D. Characterization of a second carotenoid β-hydroxylase gene from *Arabidopsis* and its relationship to the LUT1 locus. *Plant Mol. Biol.* **47**, 379–388 (2001).
7. Bai, L., Kim, E., DellaPenna, D. & Brutnell, T.P. Novel lycopene epsilon cyclase activities in maize revealed through perturbation of carotenoid biosynthesis. *Plant J.* **59**, 588–599 (2009).
8. Aluru, M. *et al.* Generation of transgenic maize with enhanced provitamin A content. *J. Exp. Bot.* **59**, 3551–3562 (2008).
9. Vallabhaneni, R. *et al.* Metabolite sorting of a germplasm collection reveals the hydroxylase3 locus as a new target for maize provitamin A biofortification. *Plant Physiol.* **151**, 1635–1645 (2009).
10. Cunningham, F.X. Jr., Lee, H. & Gantt, E. Carotenoid biosynthesis in the primitive red alga *Cyanidioschyzon merolae*. *Eukaryot. Cell* **6**, 533–545 (2007).
11. Vallabhaneni, R. & Wurtzel, E.T. Timing and biosynthetic potential for carotenoid accumulation in genetically diverse germplasm of maize. *Plant Physiol.* **150**, 562–572 (2009).
12. Li, F., Murillo, C. & Wurtzel, E.T. Maize Y9 encodes a product essential for 15-cis-zeta-carotene isomerization. *Plant Physiol.* **144**, 1181–1189 (2007).
13. Matthews, P.D. & Wurtzel, E.T in *Food Colorants: Chemical and Functional Properties* (ed. Sociaciu, C.) 361–368 (CRC Press, Boca Raton, Florida, USA, 2007).

ONLINE METHODS

Germplasm evaluation. Three independent association panels were used in this study.

Panel 1 (P1). 281 maize inbred lines grown in Urbana, Illinois (USA) in 2002–2005, as described[4].

Panel 2 (P2). 245 diverse maize inbred lines predominantly derived from tropical and subtropical adapted maize germplasm. Inbred lines were grown in one-row plots (4.5 m, 25-cm spacing) with two replications in a randomized complete-block design at CIMMYT experimental station at Tlaltizapan, Morelos (Mexico) during the summer of 2006. Fertilizer, irrigation water and herbicides were added according to standard agronomic practices.

Panel 3 (P3). 155 diverse maize inbred lines derived from temperate-adapted maize germplasm. Inbred lines were grown in one-row plots (4 m, 67-cm spacing) with two replications in a randomized complete-block design on the agronomy farm of the China Agricultural University (Beijing, China) during the summers of 2005, 2006 and 2007. Fertilizer, irrigation water and herbicides were added according to standard agronomic practices.

Carotenoid quantification. A balanced bulk of self-pollinated ears from each replication per genotype was used for HPLC analysis. Extraction of carotenoids for P2 and all segregating mapping populations was carried out according to Kurilich and Juvik[14]. Quantification of carotenoids was accomplished by standard regression with external standards according to Harjes *et al.*[4]. Measured metabolites include α-carotene (not measured in P2), lutein, β-carotene (βC), β-cryptoxanthin (βCX) and zeaxanthin (Z). Carotenoids were extracted and quantified for P3 as described in Chander *et al.*[15]. Carotenoid profiles for P1 were obtained by Harjes *et al.*[4]. Ratios and sums of carotenoid concentrations were also calculated as derived traits including βC/βCX, βCX/Z, βC/total carotenoids and the total sum of carotenoids (ALL = βC + βCX + Z + α-carotene + lutein).

Population structure and kinship analyses. Population structure and kinship for P1 was estimated using 89 simple sequence repeat (SSR) markers and 553 SNP markers, respectively, according to Yu *et al.*[16]. STRUCTURE 2.1 (ref. 17) was used to estimate the population structure of P2 and P3 using 46 and 86 SSRs, respectively. Five independent runs were carried out using the following: number of populations (*K*) from 1 to 10, burn-in time and Markov-chain Monte Carlo replication number both set to 500,000, model of admixture and correlated allele frequencies. The *K* value was determined by LnP(D) in STRUCTURE output, and the population structure at $K = 2$ was used for association analysis both in P2 and P3. The relative kinship matrix was calculated using SPAGeDi[18], using the same 46 SSRs for P2, and using 884 SNP markers for P3. Negative kinship values between individuals were set to 0 (ref. 16).

Sequencing and genotyping. Using the *A. thaliana BCH1* sequence[5], five maize homologs were identified (**Supplementary Table 18**). A single amplicon from each homolog was sequenced from P1. Sequenced polymorphisms were tested for association with β-carotene concentration as described in the section below on statistical analysis. The full sequence of MAGI4_129349 from the MAGI (maize assembled genomic islands) database[19], identified as the strongest association, was obtained from published maize BAC sequence data (BAC AC194430), and gene structure was predicted by Fgenesh (Softberry, Inc.) and confirmed by EST contig GB4_018115_01. Six primer pairs covered the full gene for sequencing across P1and P3 (**Supplementary Tables 2** and **3**). The entire *crtRB1* coding region was sequenced across 281 lines from the P1 association mapping panel including B73. In addition, 1,501 bp 5′ and 1,287 bp 3′ of the gene were sequenced in 54 and 46 lines, respectively, from panel P3. Sequences were submitted to GenBank (NCBI); see accession codes section in main article. Several insertions and deletions (indels) and SNPs in coding and noncoding regions (**Fig. 1b** and **Supplementary Table 19**) were found, including a variable 5′InDel (397 or 206 bp, termed 5′TE), InDels1–4 in exon 1, InDels5–6 in introns, a SNP that causes a premature stop codon in the sixth exon (causing a truncation of 26 amino acids), and transposable element(s) (3′TE) causing a large, variable InDel (1,250 or 325 bp) in coding sequence and 3′ UTR. Both 3′TE insertion alleles (with the B73 reference allele being representative of the 1,250-bp insertion) result in displacement of sequence encoding 11 amino acids, and addition of 40 previously unknown amino acids encoded by sequence within the insertion. LD decayed rapidly across *crtRB1*, within 1,000 bp (**Supplementary Fig. 5**), which allows intragenic association analysis within a few hundred base pairs. Subsequently, three markers for *lcyε* described by Harjes *et al.*[4] were designed and used to score P2 and P3; three markers for *crtRB1* (described in **Supplementary Fig. 2** and **Supplementary Table 20**) were used to score P1, P2 and P3. Markers found to be monomorphic in a population were not used in subsequent analyses.

Linkage mapping and QTL mapping. *crtRB1* was mapped via genetic linkage mapping in a RIL population derived from B73 and BY804[15], using the *crtRB1* 3′TE polymorphism (primers listed in **Supplementary Table 3**). QTL analysis in this population was done using QTL Cartographer 2.5 (ref. 20). Gene mapping of *crtRB1* and *lcyε* in $F_{2:3}$ mapping populations A619 × SC55 and DEexp × CI7 was accomplished using the *crtRB1* InDel4 polymorphism (primers listed in **Supplementary Table 3**) and the *lcyε* 3′TE polymorphism (primers listed in Harjes *et al.*[4]), JoinMap software[21] and respective linkage maps[22]. QTL analyses for these populations were carried out using PlabQTL[23]. Two additional $F_{2:3}$ populations (KI3 × B77 and KI3 × SC55) were genotyped with allele-specific primers for *crtRB1* InDel4 and *lcyε* 5′TE (primers listed in Harjes *et al.*[4]). The effect of *crtRB1* in these populations was evaluated by single-factor analysis using one-way ANOVA in SAS version 9.2 (ref. 24), with type 1 error rate $\alpha = 0.01$.

Transcript profiling and allozyme analysis in *E. coli*. All primer information is listed in **Supplementary Table 20**.

Complementary DNA (cDNA) libraries were generated from NC356, Hi27, NC320, CI7, B77 and A619 maize seed harvested at 15 DAP and 2-week old seedlings as described in Bai *et al.*[7] SYBR Green Real-Time PCR was carried out in kernel and seedling tissues as described in Bai *et al.*[7] Primers crtRB1_E1F and crtRB1_E2R, designed to span an intron-exon junction of *crtRB1*, were used for relative quantification. The full-length *crtRB1* coding region was amplified from NC356 and Hi27 cDNA libraries using High-Fidelity Platinum Taq DNA polymerase (Invitrogen) in PCR reactions with primers crtRB1_BamHI_F and crtRB1_HindIII_R. Similarly, the full-length *crtRB1* ORF was amplified from a CI7 cDNA library using primers crtRB1_BamHI_F and crtRB1_CI7_HindIII_R. The full-length NC356 clone was used as a template for site-directed mutagenesis as described in Bai *et al.*[7] to recreate the naturally occurring substitution found in NC320 (G300STOP). This point mutation was recreated by amplifying DNA from the NC320 clone with primers crtRB1_E3F and crtRB1_E4R. All amplified sequences were inserted in-frame into the BamH1 and HindIII sites of the expression vector pTrcHisB (Invitrogen) and transformed into *E. coli* lines engineered to express genes necessary for βC production[6]. Cell cultures were grown at 37 °C to a density of 1.2 A_{600}, induced with 0.4 mM isopropyl β-D-1-thiogalactopyranoside and then transferred to 28 °C. At time points ranging from zero hour (preinduction) to 43 h after induction, 1 ml of culture was pelleted and total lipids were extracted and assayed for carotenoid composition as described[7]. Protein gels were fractionated on equivalent amounts of induced culture for all allozymes and did not show any obvious differences in extent of CRTRB1 recombinant expression.

Forty-two maize inbred lines from P3 grown at China Agricultural University, Beijing, China (**Supplementary Table 17**), were used for *crtRB1* endosperm and embryo expression profiles at 20 DAP. Total RNA was isolated using the Plant Total RNA Extraction Kit (Bioteke Corporation). First-strand cDNA was synthesized using oligo $(dT)_{15}$ as a primer and M-MLV Reverse Transcriptase (Promega). Quantitative real-time PCR was carried out in triplicate for each sample with SYBR Premix Ex TaqTM Kit (Takara, Dalian, China) on a DNA Engine Option2 Continuous Fluorescence Detection System (Bio-Rad Laboratories) using target gene primers crtRB1_E5F and crtRB1_E6R. Relative quantity of the transcripts was calculated by using the comparative threshold cycle method. A ubiquitin (*UBI*) control, amplified by primers UBI_E1F and UBI_E2R, was used for normalization between samples.

Statistical analysis. Association analysis was carried out using a mixed model incorporating kinship and population structure[16] as implemented in TASSEL2.1 (ref. 25). This approach accounts for relatedness based on population structure and kinship as estimated with random genetic markers.

doi:10.1038/ng.551

Associations for individual carotenoids and derived traits were carried out on lines in panels P1, P2 and P3 that were orange or yellow (colored); only significant associations were reported. LD analysis was carried out using TASSEL2.1 (ref. 25) with the entire sequence of *crtRB1*; a window size of 50 bp was used to plot the average r^2 against the distance (base pairs). Association panel-specific allele frequencies were calculated from all lines in each population, which included white (noncolored) lines. Two-way ANOVA was used to detect additive and nonadditive interactions between selected pairs of markers using Statistica (DataSoft) and SAS version 9.2 (ref. 24) software packages. Power transformations of phenotypic data for segregating populations were used to satisfy normality assumptions with the Kolmogorov-Smirnov test; statistical inferences are drawn on transformed data, and back-transformed estimated effects are reported. ANOVA for the *crtRB1* main factor was carried out in SAS (Proc GLM), selecting *lcyε* haplotype as the covariate and using type 1 error rate of $\alpha = 0.05$.

URLs. WHO, http://www.who.int/nutrition/topics/vad/en/; HP, http://www.harvestplus.org; NCBI, http://www.ncbi.nlm.nih.gov/sites/entrez?db=nuccore; CIMMYT, http://www.cimmyt.org.

14. Kurilich, A.C. & Juvik, J.A. Quantification of carotenoid and tocopherol antioxidants in *Zea mays*. *J. Agric. Food Chem.* **47**, 1948–1955 (1999).
15. Chander, S. *et al.* Using molecular markers to identify two major loci controlling carotenoid contents in maize grain. *Theor. Appl. Genet.* **116**, 223–233 (2008).
16. Yu, J.M. *et al.* A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **38**, 203–208 (2006).
17. Pritchard, J.K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
18. Hardy, O.J. & Vekemans, X. SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes* **2**, 618–620 (2002).
19. Fu, Y. *et al.* Quality assessment of maize assembled genomic islands (MAGIs) and large-scale experimental verification of predicted genes. *Proc. Natl. Acad. Sci. USA* **102**, 12282–12287 (2005).
20. Wang, S., Basten, C.J. & Zeng, Z. *Windows QTL Cartographer 2.5* (North Carolina State University, Raleigh, North Carolina, USA, 2005).
21. Ooijen, J.W.V. & Voorrips, R.E. *JOINMAP version 3.0, Software for the Calculation of Genetic Linkage Maps* (Plant Research International, Wageningen, The Netherlands, 2001).
22. Kandianis, C. *Genetic Dissection of Carotenoid Concentration and Compositional Traits in Maize Grain* Doctoral dissertation, University of Illinois, Urbana-Champaign. ⟨http://hdl.handle.net/2142/14699⟩ (December 2009).
23. Utz, H.F. & Melchinger, A.E. *PLABQTL: A Computer Program to Map QTL, Version 1.2* (1993). <https://www.uni-hohenheim.de/plantbreeding/software/>
24. SAS Institute. *SAS/STAT User's Guide Version 9.2* (SAS Institute, Cary, North Carolina, USA, 2008).
25. Bradbury, P.J. *et al.* TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).