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

Jianbing Yan
CIMMYT

Catherine Bermudez Kandianis
University of Illinois at Urbana-Champaign

Carlos E. Harjes
Cornell University

Ling Bai
Boyce Thompson Institute, Ithaca

Eun-Ha Kim
Michigan State University

Follow this and additional works at: https://lib.dr.iastate.edu/agron_pubs
Part of the Agricultural Economics Commons, Agronomy and Crop Sciences Commons, Nutrition Commons, and the Plant Breeding and Genetics Commons

The complete bibliographic information for this item can be found at https://lib.dr.iastate.edu/agron_pubs/463. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.
Rare genetic variation at Zea mays crtRB1 increases β-carotene in maize grain

Abstract
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.

Disciplines
Agricultural Economics | Agronomy and Crop Sciences | Nutrition | Plant Breeding and Genetics

Comments
This article is published as Yan, Jianbing, Catherine Bermudez Kandianis, Carlos E. Harjes, Ling Bai, Eun-Ha Kim, Xiaohong Yang, Debra J. Skinner et al. “Rare genetic variation at Zea mays crtRB1 increases β-carotene in maize grain.” Nature genetics 42, no. 4 (2010): 322. doi: 10.1038/ng.551.

Rights
Works produced by employees of the U.S. Government as part of their official duties are not copyrighted within the U.S. The content of this document is not copyrighted.

Authors
Jianbing Yan, Catherine Bermudez Kandianis, Carlos E. Harjes, Ling Bai, Eun-Ha Kim, Xiaohong Yang, Debra J. Skinner, Zhiyuan Fu, Sharon Mitchell, Qing Li, Maria G. Salas Fernandez, Maria Zaharieva, Raman Babu, Yang Fu, Natalia Palacios, Jiasheng Li, Dean DellaPenna, Thomas Brutnell, Edward S. Buckler, Marilyn L. Warburton, and Torbert Rocheford

This article is available at Iowa State University Digital Repository: https://lib.dr.iastate.edu/agron_pubs/463
Rare genetic variation at Zea mays crtRB1 increases β-carotene in maize grain

Jianbing Yan1–3,13, Catherine Bermudez Kandianis4,5,13, Carlos E Harjes3,12, Ling Bai6, Eun-Ha Kim7, Xiaohong Yang2, Debra J Skinner4,5, Zhiyuan Fu2, Sharon Mitchell1, Qing Li2, Maria G Salas Fernandez3,8, Maria Zaharieva1, Raman Babu1, Yang Fu2,4,5, Natalia Palacios1, Jiasheng Li2, Dean DellaPenna7, Thomas Brutnell6, Edward S Buckler3,9,10, Marilyn L Warburton1,11 & Torbert Rocheford4,5

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 lycε1) that substantially increase the ratio of βC to α-carotenoids in grain4 (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 genes5,6 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 as...
as derived traits $\beta C/\beta C X (60\%), \beta C/Z (42\%)$ and $\beta C/\beta C X (i.e., \beta C + \beta C X + Z + \alpha$-carotene + lutein) (42%) in 2003. InDel4 and 3TE 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 Table 1, as well as in P2 and P3 (Supplementary Tables 2 and 3). The 5TE-InDel4 and the InDel4-3TE polymorphisms were in linkage disequilibrium (LD) ($r^2 = 0.23$ and 0.12, respectively), but the 5TE-3TE 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 $\beta C$ to $Z$, but the 5TE has the most pronounced effect (Supplementary Table 4), leading to an average increase of 6.50 $\mu$g g$^{-1}$ $\beta C$ above the average effect of the unfavorable allelic class (1.50 $\mu$g g$^{-1}$). The significant associations between the three polymorphisms with $\beta C$ and $\beta C/\beta C X$ show that the $crtRB1$ allelic state influences both the absolute amount of $\beta C$ 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 5TE (allele 2; Fig. 1b), leading to higher $\beta C$ 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 3TE 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 5TE, InDel4 and 3TE allelic states on carotenoid phenotypes. Of the seven observed haplotype classes in P1, $\beta C$ was highest when the favorable alleles of 5TE and 3TE (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 $\beta C$, 80% for $\beta C/\beta C X$, 68% for $\beta C/Z$ and 50% for $\beta C/\beta C X$ in P1. In P3, the

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5TE</td>
<td>1/2/3</td>
<td>n.s.</td>
<td>3.67 $\times 10^{-14}$</td>
<td>3.25 $\times 10^{-7}$</td>
<td>n.s.</td>
<td>32%</td>
<td>1.26 $\times 10^{-14}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta C/\beta C X$</td>
<td>0.0147</td>
<td>1.74 $\times 10^{-3}$</td>
<td>n.s.</td>
<td>1.46 $\times 10^{-6}$</td>
<td>60%</td>
<td>6.53 $\times 10^{-34}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta C/Z$</td>
<td>0.0163</td>
<td>4.72 $\times 10^{-21}$</td>
<td>3.0163</td>
<td>3.53 $\times 10^{-4}$</td>
<td>42%</td>
<td>1.27 $\times 10^{-20}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta C/\beta C X$</td>
<td>0.0653</td>
<td>3.99 $\times 10^{-20}$</td>
<td>2.20 $\times 10^{-7}$</td>
<td>0.0064</td>
<td>42%</td>
<td>2.73 $\times 10^{-20}$</td>
<td></td>
</tr>
<tr>
<td>InDel4</td>
<td>1/2/0</td>
<td>n.s.</td>
<td>2.40 $\times 10^{-4}$</td>
<td>2.00 $\times 10^{-6}$</td>
<td>n.s.</td>
<td>7%</td>
<td>5.65 $\times 10^{-4}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta C/\beta C X$</td>
<td>0.0498</td>
<td>8.64 $\times 10^{-5}$</td>
<td>4.90 $\times 10^{-13}$</td>
<td>0.0248</td>
<td>0.0035</td>
<td>27%</td>
<td>1.12 $\times 10^{-13}$</td>
</tr>
<tr>
<td></td>
<td>$\beta C/Z$</td>
<td>0.0202</td>
<td>9.10 $\times 10^{-10}$</td>
<td>1.52 $\times 10^{-4}$</td>
<td>n.s.</td>
<td>21%</td>
<td>1.33 $\times 10^{-10}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta C/\beta C X$</td>
<td>0.0163</td>
<td>1.46 $\times 10^{-8}$</td>
<td>5.0610 $\times 8$</td>
<td>0.0121</td>
<td>16%</td>
<td>3.30 $\times 10^{-8}$</td>
<td></td>
</tr>
<tr>
<td>3TE</td>
<td>1/2/3</td>
<td>n.s.</td>
<td>6.11 $\times 10^{-6}$</td>
<td>1.56 $\times 10^{-6}$</td>
<td>n.s.</td>
<td>10%</td>
<td>1.46 $\times 10^{-6}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta C/\beta C X$</td>
<td>0.0034</td>
<td>4.94 $\times 10^{-4}$</td>
<td>2.54 $\times 10^{-10}$</td>
<td>n.s.</td>
<td>26%</td>
<td>5.88 $\times 10^{-12}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta C/Z$</td>
<td>0.0828</td>
<td>8.16 $\times 10^{-7}$</td>
<td>0.0017</td>
<td>3.53 $\times 10^{-6}$</td>
<td>17%</td>
<td>1.77 $\times 10^{-7}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta C/\beta C X$</td>
<td>0.0215</td>
<td>7.86 $\times 10^{-10}$</td>
<td>5.15 $\times 10^{-8}$</td>
<td>8.94 $\times 10^{-5}$</td>
<td>18%</td>
<td>6.06 $\times 10^{-8}$</td>
<td></td>
</tr>
</tbody>
</table>

*Only significant polymorphic sites are shown. Alleles in series are listed for each polymorphism, and favorable alleles (higher $\beta$-carotene content) are in bold face type. 5TE allelic series: 1, 397-bp insertion; 2, 206-bp insertion; 3, no insertion. InDel4 allelic series: 12-bp or 0-bp insertion. 3TE allelic series: 1, insertion; 2, 325-bp insertion; 3, 1,250-bp insertion. $^1$P value from association analysis carried out using the mixed model incorporating population structure and kinship, using data from 4 different years. $^2$An outlier was excluded in all 2003 analyses. $^aR^2$ values from analysis of variance (ANOVA) of 2003 data showing percentage phenotypic variation explained. $^bP$ value from ANOVA analysis of 2003 data. n.s., not significant at $\alpha = 0.05$; $\beta C$, $\beta$-carotene $\mu$g g$^{-1}$; $\beta C/\beta C X$, $\beta$-carotene over $\beta$-cryptoxanthin; $\beta C/Z$, $\beta$-carotene over zeaxanthin; $\beta C/\beta C X$, $\beta$-carotene over total carotenoids ($\beta$-carotene, $\beta$-cryptoxanthin, $\alpha$-carotene, zeaxanthin and lutein).
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 BC/βCX, BC/Z and BC/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 F2:3 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 BC/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 × CI17 and A619 × SC55 F2:3 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 F2:3 populations (analysis of variance (ANOVA), α = 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 LcyE. To estimate crtRB1 effects with a single-gene model (Table 3), we adjusted for LcyE effects and found crtRB1 to explain significant variation in β-branch traits including BC (R² range 15–60%), βC/BCX (R² range 37–69%) and BC/ALL (R² 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, α = 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 2: crtRB1 haplotype estimated effects in P1 (2003 phenotypic data)

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>5'TE</th>
<th>InDel4</th>
<th>3'TE</th>
<th>N</th>
<th>βC</th>
<th>s.e.m.</th>
<th>C/CX</th>
<th>s.e.m.</th>
<th>βC/Z</th>
<th>s.e.m.</th>
<th>βC/ALL</th>
<th>s.e.m.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,0,1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>28</td>
<td>2.15</td>
<td>1.56</td>
<td>2.32</td>
<td>1.16</td>
<td>0.41</td>
<td>0.27</td>
<td>0.1</td>
<td>0.06</td>
</tr>
<tr>
<td>1,0,2</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>15</td>
<td>0.88</td>
<td>0.71</td>
<td>0.64</td>
<td>0.32</td>
<td>0.10</td>
<td>0.07</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>1,0,3</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>109</td>
<td>1.40</td>
<td>1.42</td>
<td>0.95</td>
<td>0.61</td>
<td>0.18</td>
<td>0.29</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>1,12,1</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>1.91</td>
<td>0.95</td>
<td>4.09</td>
<td>2.48</td>
<td>1.24</td>
<td>1.19</td>
<td>0.10</td>
<td>0.04</td>
</tr>
<tr>
<td>2,0,1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>13.34</td>
<td>19.06</td>
<td>3.79</td>
<td>0.42</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2,12,1</td>
<td>2</td>
<td>12</td>
<td>1</td>
<td>5</td>
<td>6.79</td>
<td>4.77</td>
<td>10.18</td>
<td>4.00</td>
<td>2.18</td>
<td>1.83</td>
<td>0.39</td>
<td>0.22</td>
</tr>
<tr>
<td>3,0,3</td>
<td>3</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td>2.64</td>
<td>1.21</td>
<td>2.92</td>
<td>1.10</td>
<td>0.40</td>
<td>0.11</td>
<td>0.11</td>
<td>0.04</td>
</tr>
</tbody>
</table>

R² (ANOVA) 40% 80% 68% 50%

Avg. change between “1,0,2” and “1,0,1” 5.23 × 10⁻¹⁶ 1.68 × 10⁻⁵³ 1.57 × 10⁻³⁸ 1.02 × 10⁻²¹

### Table 3: crtRB1 genetic effects in three segregating populations

<table>
<thead>
<tr>
<th>Copies of favorable haplotype</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>N</th>
<th>BC</th>
<th>s.e.m.</th>
<th>BC/ALL</th>
<th>s.e.m.</th>
<th>BC/Z</th>
<th>s.e.m.</th>
<th>BC/ALL</th>
<th>s.e.m.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>32</td>
<td>33</td>
<td>12</td>
<td>2</td>
<td>5.15</td>
<td>2.37</td>
<td>5.57</td>
<td>0.47</td>
<td>0.17</td>
<td>0.46</td>
<td>9.89</td>
<td>0.33</td>
</tr>
<tr>
<td>1</td>
<td>114</td>
<td>42</td>
<td>26</td>
<td>5.02</td>
<td>1.49</td>
<td>3.39</td>
<td>4.17</td>
<td>0.07</td>
<td>0.28</td>
<td>4.83</td>
<td>0.10</td>
<td>1.50</td>
</tr>
<tr>
<td>0</td>
<td>50</td>
<td>35</td>
<td>28</td>
<td>4.62</td>
<td>0.76</td>
<td>2.79</td>
<td>0.38</td>
<td>0.03</td>
<td>0.17</td>
<td>3.60</td>
<td>0.04</td>
<td>0.62</td>
</tr>
</tbody>
</table>

R² (%) 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

°Traits described in Table 1: A, population A619 × SC55; B, population KI3 × B77; C, population KI3 × SC55.

α = 0.01, P < 0.001; Supplementary Table 8.

βC refers to the concentration of β-carotene. All effects are adjusted by covariates. Adjusted means are all significantly different at α = 0.05, except where noted (†).
**Table 4** Estimated effects of combined haplotypes of *crtRB1* and *lcy* genes in P1 (2003 phenotypic data)

<table>
<thead>
<tr>
<th>Haplotype*</th>
<th>5TE</th>
<th>3TE</th>
<th>N</th>
<th>(\beta)C</th>
<th>s.e.m.</th>
<th>(\beta)C/(\beta)C/X</th>
<th>s.e.m.</th>
<th>(\beta)C/Z</th>
<th>s.e.m.</th>
<th>(\beta)C/ALL</th>
<th>s.e.m.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1,2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td>3.31</td>
<td>1.98</td>
<td>2.44</td>
<td>1.46</td>
<td>0.87</td>
<td>1.04</td>
<td>0.15</td>
</tr>
<tr>
<td>1.1,3</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>24</td>
<td>1.65</td>
<td>0.96</td>
<td>2.72</td>
<td>1.67</td>
<td>0.45</td>
<td>0.34</td>
<td>0.07</td>
</tr>
<tr>
<td>1.2,2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>1.25</td>
<td>0.85</td>
<td>0.72</td>
<td>0.35</td>
<td>0.09</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>1.2,3b</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>8</td>
<td>0.56</td>
<td>0.38</td>
<td>0.58</td>
<td>0.31</td>
<td>0.11</td>
<td>0.09</td>
<td>0.02</td>
</tr>
<tr>
<td>1.3,2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>47</td>
<td>1.65</td>
<td>1.22</td>
<td>1.06</td>
<td>0.49</td>
<td>0.27</td>
<td>0.41</td>
<td>0.07</td>
</tr>
<tr>
<td>1.3,3</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>48</td>
<td>0.81</td>
<td>1.10</td>
<td>0.88</td>
<td>0.76</td>
<td>0.11</td>
<td>0.07</td>
<td>0.03</td>
</tr>
<tr>
<td>1.3,1/4</td>
<td>1</td>
<td>1</td>
<td>1/4b</td>
<td>11</td>
<td>2.71</td>
<td>2.12</td>
<td>0.76</td>
<td>0.23</td>
<td>0.15</td>
<td>0.08</td>
<td>0.04</td>
</tr>
<tr>
<td>2.1,2b</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>6.79</td>
<td>4.77</td>
<td>10.18</td>
<td>4.00</td>
<td>2.18</td>
<td>1.83</td>
<td>0.39</td>
</tr>
<tr>
<td>2.1,3</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>13.34</td>
<td>–</td>
<td>19.06</td>
<td>–</td>
<td>5.56</td>
<td>–</td>
<td>0.42</td>
</tr>
<tr>
<td>3,3,2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>2.79</td>
<td>1.34</td>
<td>3.03</td>
<td>1.24</td>
<td>0.38</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>3,3,3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2.05</td>
<td>–</td>
<td>2.49</td>
<td>–</td>
<td>0.48</td>
<td>–</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Maximum change 165 12.1 17.6 19.8 19.3

**Notes:**

1. Haplotype is shown as linear combination of *crtRB1* 5TE allele (1, 397 bp insertion; 2, 206 bp insertion; 3, 0 bp insertion), *lcy* 5TE 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. 2. Haplotypes predicted to yield most (2.1,2) and least (1.2,3) \(\beta\) carotene from available classes. 3. Allelic classes of 1 and 4 (1/4) at *lcy* 5TE allele were combined because both classes yield a similar effect on carotenoid traits. 4. Traits are described in Table 1.

**Expression analysis**

Previous studies have shown the importance of transcript abundance in the control of carotenoid profiles.6,7,8 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 \(\beta\)C concentration.4 Peak expression of carotenoid biosynthetic genes and *crtRB1* was noted to occur within 15–20 DAP in other studies.5,9 In lines containing relatively high concentrations of \(\beta\)C (CI7 and B77; identical *crtRB1* haplotypes), *crtRB1* transcripts accumulated to only 1/70 of the level of *crtRB1* transcripts in lines with low \(\beta\)C (NC356, HI27 and NC320; Fig. 3a). Relative to all tested lines, transcript and \(\beta\)C concentrations were both intermediate in A619, which carries only the 3TE insertion. This implies that polymorphisms associated with 5TE, 3TE and InDel4 result in substantial transcript reduction in kernels and probably decrease \(\beta\)-hydroxylase activity through reduction of CRTRB1 protein levels.

To test the independent effect of 3TE alleles on *crtRB1* expression, we determined embryo and endosperm profiles at 20 DAP for 42 additional lines monomorphic at the 5TE and InDel4 but polymorphic for the 3TE (Fig. 3b and Supplementary Table 14). In endosperm, *crtRB1* allele-specific expression levels inversely matched the allelic effect on \(\beta\)C concentration (ANOVA, \(\alpha = 0.01\)): lines with the most favorable 3TE 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 3TE 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), 5TE/InDel4 (with complete LD between the two polymorphisms in the lines tested) and 3TE (Supplementary Fig. 3a). The Hi27 allozyme (identical in protein

**Fig. 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 \(\beta\)C, ALL and ALL-\(\beta\)C (that is, total carotenoids minus \(\beta\)C). (a) A619 × SC55; (b) KI3 × B77; (c) KI3 × SC55. Error bars, s.e.m.
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 \( \beta C \)) 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 alleles were expressed in \( E. coli \) engineered to accumulate \( \beta C \) as a substrate\(^{10} \), and the resulting hydroxylation products were analyzed by HPLC. All four alleles showed \( \beta C \)-carotene hydroxylase activity but differed in their product profiles and specificity for each \( \beta \)-ring (Fig. 3c and Supplementary Fig. 4).

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

**DISCUSSION**

Three polymorphisms in \( \text{crtrB1} \) (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 \( \beta C \) concentrations of 2.91–8.00 \( \mu \)g g\(^{-1} \) in germplasm with the favorable allelic composition at individual polymorphisms, which translates into \( \text{crtrB1} \)-specific increases of 1.45–6.50 \( \mu \)g g\(^{-1} \) provitamin A meeting up to 43% of the target HP goal. The functional basis of \( \text{crtrB1} \) 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 \( \beta C \) concentration. Altering the C terminus of the protein was shown in \( E. coli \) to enhance hydroxylation of \( \beta C \) and broaden the product profile to include \( \beta Z \). Notably, the \( \text{CI7} \) 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 \( \beta C \) concentrations in kernels, which correlated with the extremely reduced expression of \( \text{crtrB1} \) in CI7 kernels. This implies that \( \beta C \) concentrations in CI7 are high as a result of weak kernel expression, despite the potential of the CI7 allozyme for considerable \( \beta C \) 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 \( \beta C \) to downstream products and contributes to the mechanistic basis for favorable allelic states of \( \text{crtrB1} \). 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 \( \text{PSY} \) (ref. 11), \( \text{lcyE} \) (ref. 4), \( \text{lcyB} \) (ref. 7) and \( \text{HYD3}/\text{crtrB1} \) (ref. 9).

Another \( \text{crtrB1} \) 5′UTR polymorphism was recently reported to correlate with \( \beta C \) 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 \( \beta C \). 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 \( \text{crtrB1} \) on \( \beta C \) concentration. However, joint additive effects for \( \text{crtrB1} \) and \( \text{lcyE} \) on \( \beta C \) were detected only by association mapping, whereas detection of \( \text{crtrB1} \) 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
biallytic 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 crrB1 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 crrB1. Specific haplotypes of lcyε have been shown to yield the most favorable α-carotenoid/β-carotenoid ratios and βC proportions of all measured classes, contributing to the joint additive effects of crrB1 and lcyε on βC concentrations observed in this study. The rarity of certain genetic variants is such that the most favorable haplotypes of crrB1 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 crrB1 alleles, indicates that HP targets are attainable in the presence of favorable crrB1 variation. Through the use of donor lines (Supplementary Table 17) and codominant, user-friendly PCR-based marker systems (Supplementary Fig. 1), introgression of favorable crrB1 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.R.), TRIAD Foundation (T.R.), China NSF-30821140352 (J.L., J.Y.), China Scholarship Fund (Y.F.) and JBT Fellowship-UlUIC (C.B.K.).

AUTHOR CONTRIBUTIONS


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/.

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\textsuperscript{4}. Measured metabolites include α-carotene (+ Z + β-carotene + lutein). β-Cryptoxanthin (βCX) and zeaxanthin (Z). Carotenoids were extracted and quantified for P3 as described in Eldredge et al.\textsuperscript{15}. Carotenoid profiles for P1 were obtained by Harjes et al.\textsuperscript{4}. Ratios and sums of carotenoid concentrations were also calculated as derived traits including βCX/βCX, βCX/Z, βCX/total carotenoids and the total sum of carotenoids (ALL = BC + 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.\textsuperscript{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 50,000, 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\textsuperscript{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\textsuperscript{5}, five maize homologs were identified (Supplementary Table 18). A single amplification 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 MAGI1_129349 from the MAGI maize genomic islands database\textsuperscript{19}, identified as the strongest association, was obtained from published maize BAC sequence data (BAC AC194430), and gene structure was predicted by Fgeneset (Softberry, Inc.) and confirmed by EST contig BG4_018115_01. Six primer pairs covered the full gene for sequencing across P1 and 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), InDel31–4 in exon 1, InDel6–8 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 lcy described by Harjes et al.\textsuperscript{6} 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\textsuperscript{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 lcy in F\textsubscript{2} population mapping populations A619 × SC55 and DEExp × CI7 was accomplished using the crtRB1 InDel4 polymorphism (primers listed in Supplementary Table 3) and the lcy 3′TE polymorphism (primers listed in Harjes et al.\textsuperscript{7}). JoinMap software\textsuperscript{23} and respective linkage maps\textsuperscript{22}, QTL analyses for these populations were carried out using R/qtl\textsuperscript{23}. Two additional F\textsubscript{2} populations (KX × B77 and KX × SC55) were genotyped with allele-specific primers for crtRB1 InDel4 and lcy 3′TE (primers listed in Harjes et al.\textsuperscript{7}). 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 α = 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, H27, NC320, C17, B77 and A619 maize seed harvested at 15 DAP and 2-week old seedlings as described in Bai et al.\textsuperscript{7} SYBR Green Real-Time PCR was carried out in kernel and seedling tissues as described in Bai et al.\textsuperscript{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 H27 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 C17 cDNA library using primers crtRB1_BamHI_F and crtRB1_HindIII_R. The full-length NC356 clone was used as a template for site-directed mutagenesis as described in Bai et al.\textsuperscript{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. Cells engineered to express genes necessary for BC production\textsuperscript{9} were cultured at 37 °C to a density of 1.2 A\textsubscript{600} with 0.4 mM isopropyl β-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\textsuperscript{2}. 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 (Biotek Corporation). First-strand cDNA was synthesized using oligo (dT)\textsubscript{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_ E5R. 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\textsuperscript{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.
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 \textit{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 \textit{crtRB1} main factor was carried out in SAS (Proc GLM), selecting \textit{lcy}\textsubscript{ε} haplotype as the covariate and using type 1 error rate of $\alpha = 0.05$.


