

Genetic basis of kernel nutritional traits during maize domestication and improvement

Hui Fang^{1,†}, Xiuyi Fu^{1,2,†}, Yuebin Wang^{3,†}, Jing Xu¹, Haiying Feng¹, Weiya Li¹, Jieting Xu³, Orawan Jittham¹, Xuan Zhang¹, Lili Zhang¹, Ning Yang³, Gen Xu¹, Min Wang¹, Xiaowei Li¹, Jiansheng Li¹, Jianbing Yan³  and Xiaohong Yang^{1,*} 

¹State Key Laboratory of Plant Physiology and Biochemistry, National Maize Improvement Center of China, MOA Key Laboratory of Maize Biology, Beijing Key Laboratory of Crop Genetic Improvement, China Agricultural University, Beijing 100193, China,

²Maize Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China, and

³National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China

Received 7 July 2019; revised 1 September 2019; accepted 9 September 2019; published online 16 September 2019.

*For correspondence (e-mail: yxiaohong@cau.edu.cn).

[†]These authors contribute equally to this work.

SUMMARY

The nutritional traits of maize kernels are important for human and animal nutrition, and these traits have undergone selection to meet the diverse nutritional needs of humans. However, our knowledge of the genetic basis of selecting for kernel nutritional traits is limited. Here, we identified both single and epistatic quantitative trait loci (QTLs) that contributed to the differences of oil and carotenoid traits between maize and teosinte. Over half of teosinte alleles of single QTLs increased the values of the detected oil and carotenoid traits. Based on the pleiotropism or linkage information of the identified single QTLs, we constructed a trait–locus network to help clarify the genetic basis of correlations among oil and carotenoid traits. Furthermore, the selection features and evolutionary trajectories of the genes or loci underlying variations in oil and carotenoid traits revealed that these nutritional traits produced diverse selection events during maize domestication and improvement. To illustrate more, a mutator distance–relative transposable element (TE) in intron 1 of *DXS2*, which encoded a rate-limiting enzyme in the methylerythritol phosphate pathway, was identified to increase carotenoid biosynthesis by enhancing *DXS2* expression. This TE occurs in the grass teosinte, and has been found to have undergone selection during maize domestication and improvement, and is almost fixed in yellow maize. Our findings not only provide important insights into evolutionary changes in nutritional traits, but also highlight the feasibility of reintroducing back into commercial agricultural germplasm those nutritionally important genes hidden in wild relatives.

Keywords: Oil, Carotenoid, Domestication and improvement, Genetic basis, Selection.

INTRODUCTION

Maize (*Zea mays* ssp. *mays*) is the most abundantly cultivated and highly valued food commodity in the world. Genetic and archeological evidence has suggested that maize was domesticated from its wild progenitor, teosinte (*Zea mays* ssp. *parviglumis*, referred hereafter as *parviglumis*), ~9000 years ago in the Balsas River Basin of southwestern Mexico (Matsuoka *et al.*, 2002; Piperno *et al.*, 2009; Van Heerwaarden *et al.*, 2011). Following initial domestication, maize experienced an improvement process that involved both conscious and unconscious selection to adapt it to different agro-ecological and cultural environments and to meet human demands (Meyer and

Purugganan, 2013). Together, these two processes have led to the formation and diversification of modern maize, which differs substantially from teosinte in terms of ear, kernel, and general plant characteristics (Doebley, 2004; Flint-Garcia, 2017). Therefore, it is critical to understand which genes or loci control these important traits and to determine their selection features to ensure continued improvement of maize yield and quality.

Early genome-scale surveys of maize suggested that up to 2–4% of the genome has undergone positive selection (Wright *et al.*, 2005), but recent work indicated that a much larger percentage (~7.6%) has been affected by domestication and subsequent improvement (Hufford *et al.*, 2012).

Although, in many cases, there is no evidence that the phenotypes controlled by the selected genes arose as a result of selection during domestication and improvement, the genes that have undergone selection can be categorized as domestication or improvement genes, i.e. genes with domestication features when gene diversity is greatly reduced in landraces and inbreds; genes with improvement features when gene diversity is severely reduced only in inbreds; and genes with both domestication and improvement features when gene diversity is initially reduced in landraces and then continued to be reduced in inbreds. Among these genes, two canonical domestication genes have been shown to have large effects on maize morphology: *teosinte branched1 (tb1)*, which affects overall plant architecture and branching patterns (Doebley *et al.*, 1995, 1997; Studer *et al.*, 2011), and *teosinte glume architecture1 (tga1)*, which controls the formation of the stony fruit case that encapsulates teosinte seeds (Doebley and Stec, 1993; Wang *et al.*, 2005). In addition, several variants having improvement features were recently shown to contribute to the adaptability of maize to different growing conditions (Yang *et al.*, 2013; Guo *et al.*, 2018; Huang *et al.*, 2018). For example, insertions of the CACTA-like transposon at *ZmCCT10* and of the Harbinger-like transposon at *ZmCCT9* arose following domestication, and these transposons were selected for maize adaptation to temperate regions with long-day conditions (Yang *et al.*, 2013; Huang *et al.*, 2018).

The nutritional traits of maize kernels are important for both human and animal nutrition. Previous studies have noted that kernel nutritional traits underwent selection during maize domestication and improvement, and divergent selection tended to occur owing to diverse desired human needs (Hanson *et al.*, 1996; Palaisa *et al.*, 2003; Zheng *et al.*, 2008; Fu *et al.*, 2010; Chai *et al.*, 2012). For example, yellow kernels predominate in commodity corn because of its great carotenoid content that benefits animals, whereas white kernels are preferred for human consumption in many regions around the world (Poneleit, 2000; Flint-Garcia, 2017). A population-genetics analysis based on candidate genes revealed that certain genes were subjected to selection during maize improvement, for example, *SU1* and *AE1* in the starch pathway, *PSY1* in the carotenoid pathway, and *C1* in the anthocyanin pathway (Hanson *et al.*, 1996; Whitt *et al.*, 2002; Palaisa *et al.*, 2003; Fu *et al.*, 2010). *PSY1* shows classic selection signatures during the divergence of yellow maize from white maize (Palaisa *et al.*, 2003; Fu *et al.*, 2010). Moreover, the causative variant of *DGAT1-2* in the oil pathway, which confers a high oil content, is ancestral in teosinte, was lost during the regular maize breeding program, and re-selected in a high-oil maize breeding program (Zheng *et al.*, 2008; Chai *et al.*, 2012). These results not only indicate the complex selection patterns of the genes underlying maize kernel

nutritional traits but also reveal the value of certain teosinte alleles for improving nutritional traits.

To understand the selection features of maize kernel nutritional traits and evaluate the potential of teosinte alleles for improving the nutritional traits of modern maize germplasm, we dissected the genetic architecture of oil and carotenoid traits in a teosinte-maize (TM) population. The artificial selection of both kinds of nutritional traits went through similar stages as the development of high-oil maize was initiated ~120 years ago (Hopkins, 1899; Dudley and Lambert, 2004; Yang and Li, 2018), whereas strong selection for carotenoids to form yellow color started ~100 years ago (Benz *et al.*, 2007). We subsequently constructed a trait-QTL network to assess the genetic relationship among 33 oil and carotenoid-related traits. Based on the cross-population composite likelihood ratio (XP-CLR) and nucleotide diversity previously calculated using inbreds, landraces, and teosintes (Hufford *et al.*, 2012), the selection features of the identified QTLs for oil and carotenoid traits were investigated, some of which were confirmed by nine causal variants of five genes in the oil and carotenoid biosynthesis pathways.

RESULTS

Phenotypic variation of oil and carotenoid traits in a TM population

All 33 oil and carotenoid traits exhibited abundant diversity, with most of the traits being continuous and having approximately normal distribution (Supporting Information Figure S2). For the oil traits, the averaged values in the TM population were similar to those of Mo17; for most carotenoid traits, however, the averaged values were smaller than for Mo17 (Table S2). Consequently, the carotenoid traits had a greater range of fold change (1.94–11.2) compared with the oil traits (1.24–2.44) in the TM population (Figure 1a,b), indicating higher genetic diversity. Analysis of variance (ANOVA) indicated that both genotype and environment had significant effects on each trait (Table S2), whereas the phenotypic variations were mainly controlled by genetic factors, with the broad-sense heritability ranging from 0.52 to 0.98 (Table S2). The pairwise correlations among the metabolites in the same metabolic pathway were stronger than those across different metabolic pathways (Figure 1c), suggesting that a different genetic basis underlies oil and carotenoid traits although carotenoids have been reported to be lipid-soluble metabolites (Muzhingi *et al.*, 2017).

Genetic architecture of the oil and carotenoid traits

In total, 98 single QTLs, with the QTL number per trait ranging from one to six, were detected for the 33 oil and carotenoid traits in the TM population (Figure 2a and Data S2). The 1.5-LOD (logarithm of the odds) support confidence

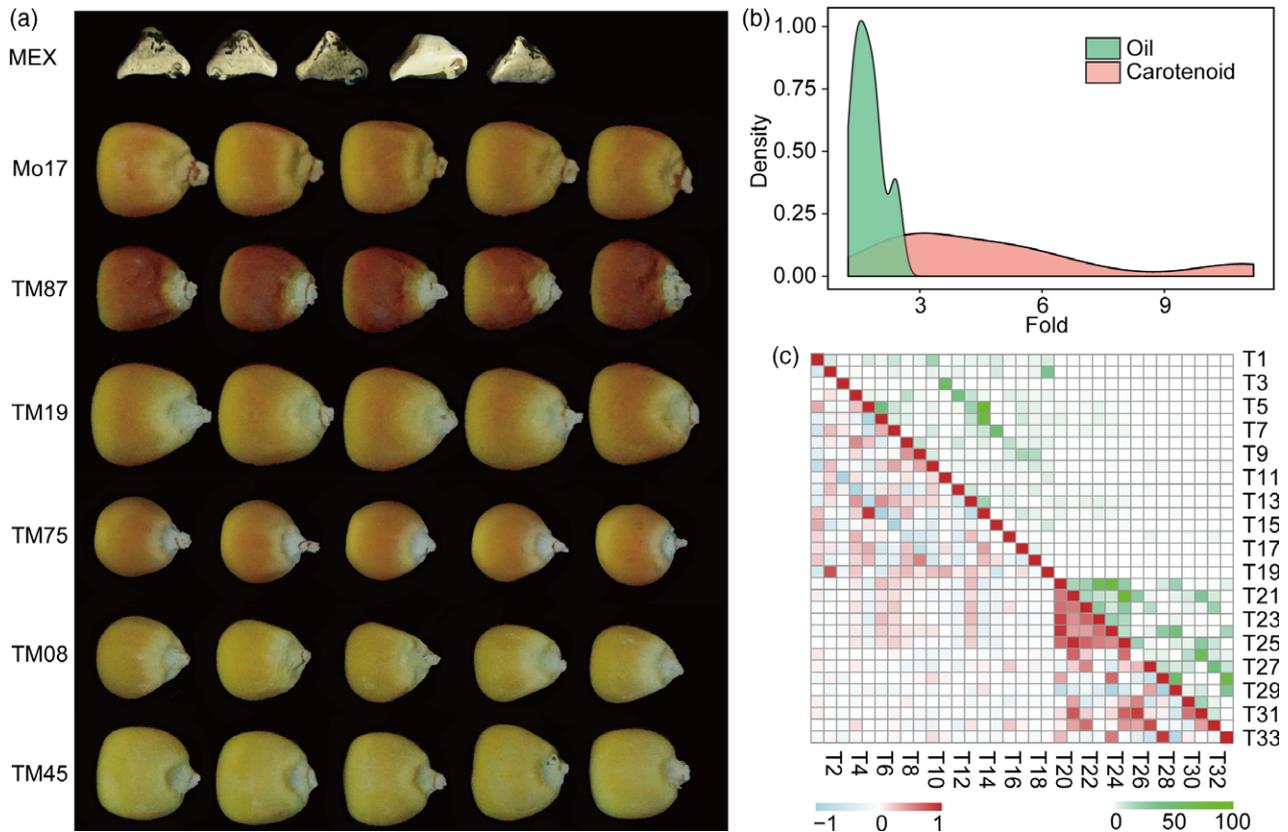


Figure 1. Phenotypic variation of 33 oil and carotenoid traits in the TM population. (a) Kernels of parents and representative lines. Mo17 and MEX are parents of the TM population, whereas TM87, TM19, TM75, TM08 and TM45 are lines in the TM population. (b) Fold-change distribution of oil and carotenoid traits. (c) Pearson correlation coefficients (bottom left) for the oil and carotenoid traits and $-\log_{10}$ (P -value) of the Pearson correlation (upper right). Detailed information for trait names is given in Table S2.

interval averaged 9.0 cM (13.4 Mb), with a range from 1.0 to 24.9 cM (0.5 to 100.0 Mb). The explained phenotypic variation (PVE) of each QTL ranged from 2.56% (*qOLIO8*) to 67.57% (*qOLIO6-3*), and the total PVE of all identified QTLs for a trait ranged from 4.4% (BCRY/ZEA) to 67.8% (C18:2) (Figure 2d). Of these single QTLs, 19.4% (19/98) had large effects with PVE \geq 15%. Most of these QTLs affect the composition of oil or carotenoid traits, which are likely to be controlled by a few large-effect QTLs owing to their simplicity as intermediate products in a particular metabolic pathway. Additionally, the teosinte alleles at 55.1% (54/98) of the QTLs had additive effects for increasing the values of the detected traits (Figure 2b). This feature indicates that teosinte possesses abundant favorable alleles that can be applied for improvement of both the quantity and quality of oil and carotenoid traits during maize breeding.

In addition to single QTLs, in total 11 pairs of epistatic QTLs involved in 12 loci were detected for eight oil and carotenoid traits (Figure 2c and Data S2). The number of epistatic QTL pairs was small (< 4) for each trait, and the PVE of each epistatic QTL pair was also small (PVE = 1.5–

7.9%). These results indicated that epistatic interactions between two QTLs with additive effects contributed less than additive effects to the genetic basis of oil and carotenoid variation in the TM population. Taken together, the single and epistatic QTLs could explain 4.4–72.9% of the phenotypic variation for each trait, some of which were comparable to the broad-sense heritability (Figure 2d).

Genetic basis of the correlations among oil and carotenoid traits

Traits underlying pleiotropic or closely linked genetic factors are often correlated and have tended to be selected together during maize domestication and improvement. For the 33 traits measured in the TM population, we observed both strong and weak as well as positive and negative correlations between trait pairs (Figure 1c). To assess the genetic relationship among traits, we constructed a trait-QTL network based on the information about locus-locus linkages and locus-trait associations (Figure 3). In this network, all traits had at least one QTL to bridge other traits. As expected, traits in the same metabolic pathway tended to be closely located in the network

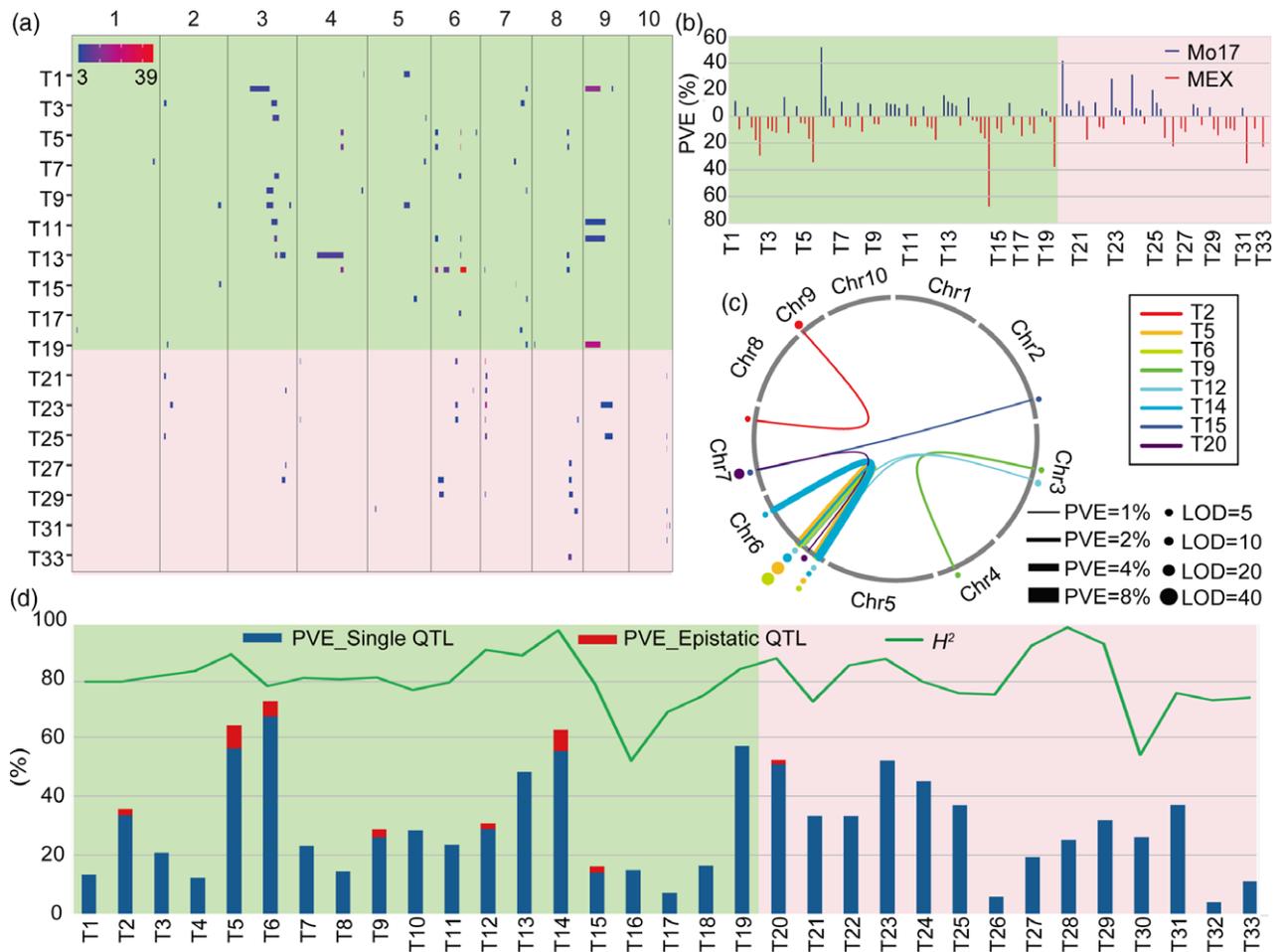


Figure 2. The distribution and effects of single and epistatic QTLs identified for 33 oil and carotenoid traits in the TM population. (a) Distribution of single QTLs on chromosomes. Light-green background, oil QTLs; pink background, carotenoid QTLs. QTL regions across the maize genome are represented by confidence intervals, and LOD values are scaled by color. (b) Effect size (represented by PVE) and the origin of the increasing alleles of the identified single QTLs. Red and blue bars indicate that increasing alleles come from Mo17 and MEX, respectively. (c) Epistatic interaction among single QTLs for eight traits. Traits are designated by color. Dot size indicates the magnitude of the LOD value for each QTL. QTLs that have an epistatic interaction are connected by a colored line, and the line thickness represents the PVE of the epistatic QTLs. (d) The total PVE of single (blue bars) and epistatic (red bars) QTLs for each trait. The green line indicates broad-sense heritability, H^2 (%). Detailed information for trait names is given in Table S2.

and have a high level of connectivity, indicating that they shared a large number of common loci. This is reasonable, given that the alteration of genes encoding key enzymes in a particular metabolic pathway often leads to changes in the production of metabolites downstream or upstream of the enzyme-catalyzed reaction owing to feedforward and feedback regulation. For example, C18:1 (T5) and its directly linked loci on chromosome 6 were associated with two additional fatty acid compositions and the ratios of their related traits (Figure 3 and Data S2). This result was further supported by the previous cloning of this QTL cluster as *DGAT1-2*, which encodes acyl-CoA:diacylglycerol acyltransferase and catalyzes the final step in the Kennedy pathway for triacylglycerol biosynthesis (Zheng *et al.*, 2008). Similar patterns were observed and approved for traits that were directly linked to loci encoding *SAD1* and

FAD2 in the oil biosynthesis pathway and for *PSY1*, *IcyE* and *crtRB1* in the carotenoid biosynthesis pathway (Figure 3 and Data S2).

Compared with traits in the same metabolic pathway, those in different metabolic pathways had few connections, consistent with the correlation patterns. The existence of connectivity provides a potential means for predicting whether there is a common genetic basis for oil and carotenoid biosynthesis, whereas close linkage might explain the co-localization of loci for both oil and carotenoid traits. For instance, L3, L9 and L39 were found to be the loci that affect both oil and carotenoid traits. This trait-locus network provides useful information for understanding the genetic basis of correlations among oil and carotenoid traits, which could facilitate the improvement of multiple traits by pyramiding desirable QTLs.

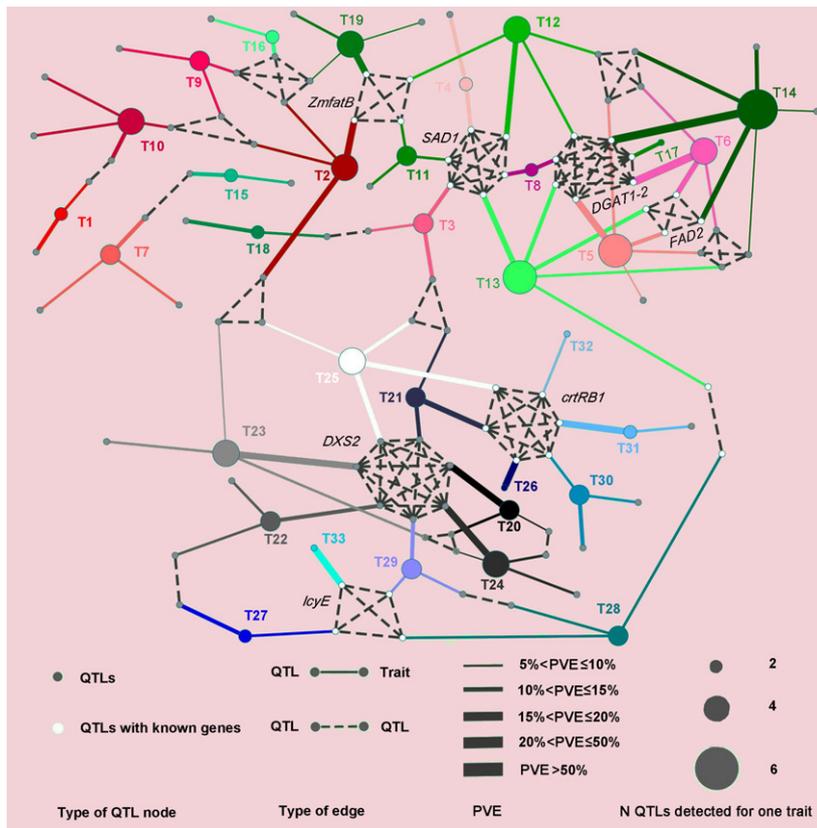


Figure 3. Trait-QTL network for 33 traits and 98 single QTLs. The trait-QTL network was constructed based on the QTL results for 33 traits and the co-localization information for all 98 QTLs. Traits and loci are connected by the solid lines in the network if the loci were associated with the traits. Loci are connected by the dashed lines in the network if two QTLs co-localized at the same loci. Detailed information for trait names is given in Table S2.

Selection features of QTLs for oil and carotenoid traits

To determine whether the identified QTLs for oil and carotenoid traits underwent domestication and improvement, we defined one QTL cluster as a locus and, in total, 42 loci were observed (Data S2). Combined with evidence obtained from the previous analysis of the XP-CLR and π ratio (Hufford *et al.*, 2012), we defined 17 loci with features that reflected domestication or improvement (Figure 4a). Of these loci, five underwent selection only during maize domestication, e.g. L10 with a high XP-CLR value and greatly reduced nucleotide diversity only in landraces compared with teosintes (Figure 4b), whereas eight loci underwent selection only during maize improvement, e.g. L3 with a high XP-CLR value and greatly reduced nucleotide diversity only in maize compared with the landraces (Figure 4c). Interestingly, we identified four loci (e.g., L37) that underwent selection during both maize domestication and improvement (Figure 4d). Unfortunately, the selection features of 45.2% (19/42) of the loci that had presumably undergone selection could not be exactly determined owing to the complex underpinnings of selection, such as L25 (Figure S3c) and L29 (Figure S4). Collectively, three kinds of selection features could be inferred for loci that affect nutritional traits in maize kernels. Note, however, that our prediction of the selection fate of the detected

QTLs was based on a computational analysis, so whether these QTLs actually experienced such selection during maize domestication and improvement will require cloning of the identified QTLs.

Diverse selection patterns of favorable alleles for oil and carotenoid traits

The QTL-gene co-localization identified six known genes falling within six loci for oil and carotenoid traits, but none of the genes was located within or in close proximity to the selection signals (Figure S3 and Data S2). To investigate whether these genes experienced selection during maize domestication and improvement, we carried out an in-depth frequency analysis of the favorable alleles that were previously identified, which are polymorphic between the parents of the TM population, in teosintes, landraces, and maize inbreds. These favorable alleles of four genes contained the following: a 3-bp insertion in *DGAT1-2* (F469; Zheng *et al.*, 2008; Chai *et al.*, 2012); a 11-bp insertion in *ZmfatB* (S_4294, Li *et al.*, 2011); three functional sites in *lcyE* including a 933-bp insertion in the promoter (5'TE), a SNP in exon one (SNP216), and a 8-bp deletion in the 3'UTR (3'InDel) (Harjes *et al.*, 2008); and, finally, three functional sites in *crtRB1* including a 206-bp insertion in the 5'UTR (5'TE), a 12-bp insertion in exon one (InDel4), and a

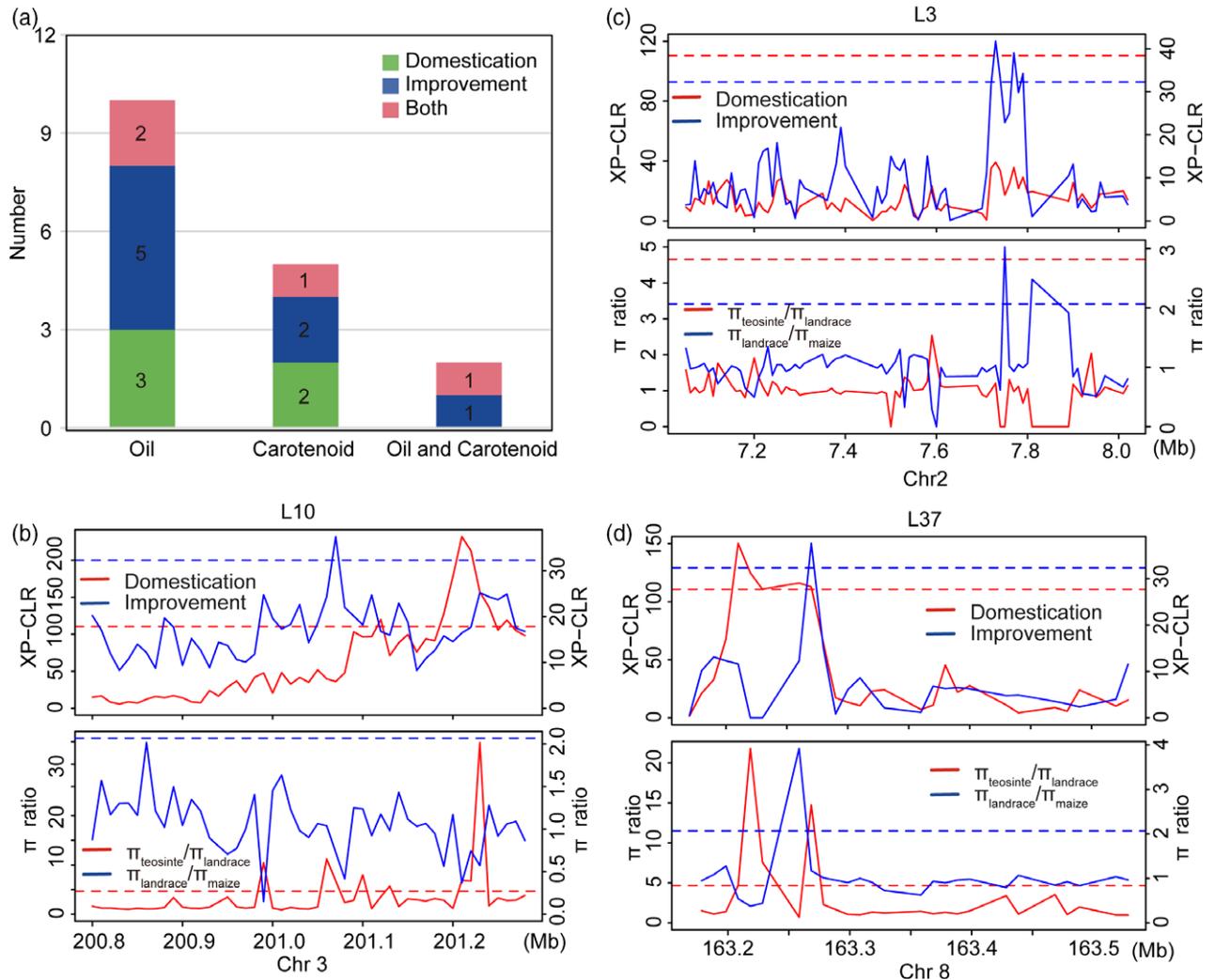


Figure 4. Selection features of detected QTLs for oil and carotenoid traits identified in the TM population. (a) Distribution of QTL with features that reflect domestication or improvement. The numbers of QTLs with domestication or improvement features are shown in the histogram. (b–d) XP-CLR values for selection during domestication and improvement (upper) and nucleotide diversity (lower) across L10 with domestication feature (b), L3 with improvement feature (c) and L37 with both domestication and improvement features (d). The red and blue dashed lines represent the threshold of XP-CLR and π ratio, respectively. The left y-axis indicates the XP-CLR (upper graph) and $\pi_{\text{teosinte}}/\pi_{\text{landrace}}$ (lower graph) values reflecting domestication, whereas the right y-axis indicates the XP-CLR (upper) and $\pi_{\text{landrace}}/\pi_{\text{maize}}$ (lower graph) values reflecting improvement.

1250-bp deletion in the coding sequence and 3'UTR (3'TE) (Yan *et al.*, 2010) (Figure 5a–d). Notably, all of the favorable alleles were detected in teosinte, indicating the teosinte alleles have potential value for improving oil and carotenoid traits. The frequency profiles of favorable alleles revealed diverse selection events during maize domestication and improvement (Figure 5e–h). The frequency of the *DGAT1-2* favorable allele decreased dramatically from 0.99 in teosinte to 0.18 in landraces, was consistently low at 0.09 in regular maize inbreds, but was high (0.97) in high-oil inbreds (Figure 5f). The loss of this favorable allele in landraces and regular maize and its reselection in high-oil maize suggests that *DGAT1-2* underwent selection throughout maize domestication and improvement.

Similarly, the frequency of two *lcyE* favorable alleles (SNP216 and 3'InDel) and two *crtRB1* favorable alleles (3'TE and InDel4) consistently reduced in landraces, and three favorable alleles (SNP216, 3'InDel and 3'TE) underwent divergent selection between white and yellow maize (Figure 5g–h). Interestingly, the remaining favorable alleles of both *lcyE* and *crtRB1* (5'TE) were considerably rare in teosinte, landraces and maize (Figure 5g–h), indicating that no selection underwent for these two sites. By contrast, only one teosinte entry (*parviglumis*) was found to harbor the favorable allele of *ZmfatB*. The 11-bp insertion at the S_4239 site in teosinte may have resulted from post-domestication gene flow between maize and *parviglumis*, which was confirmed by the

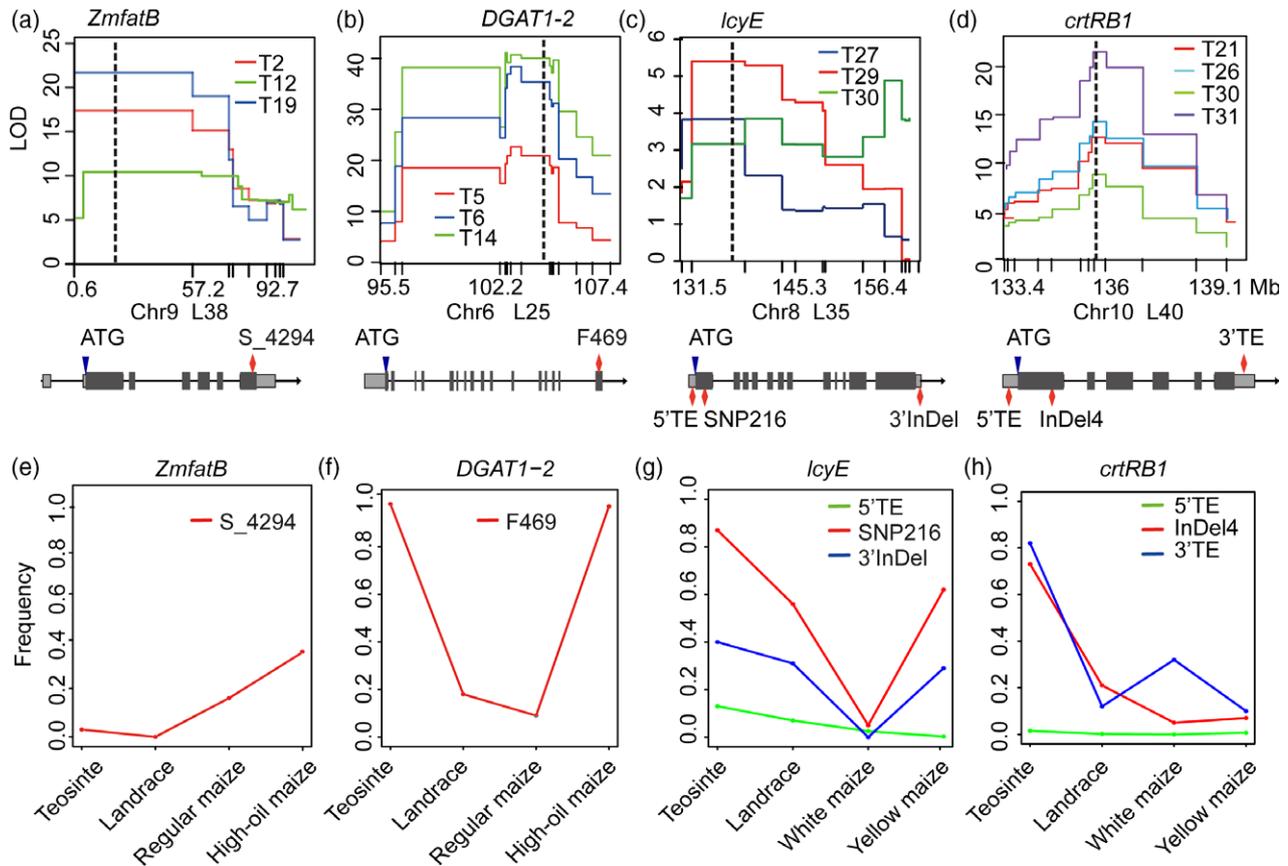


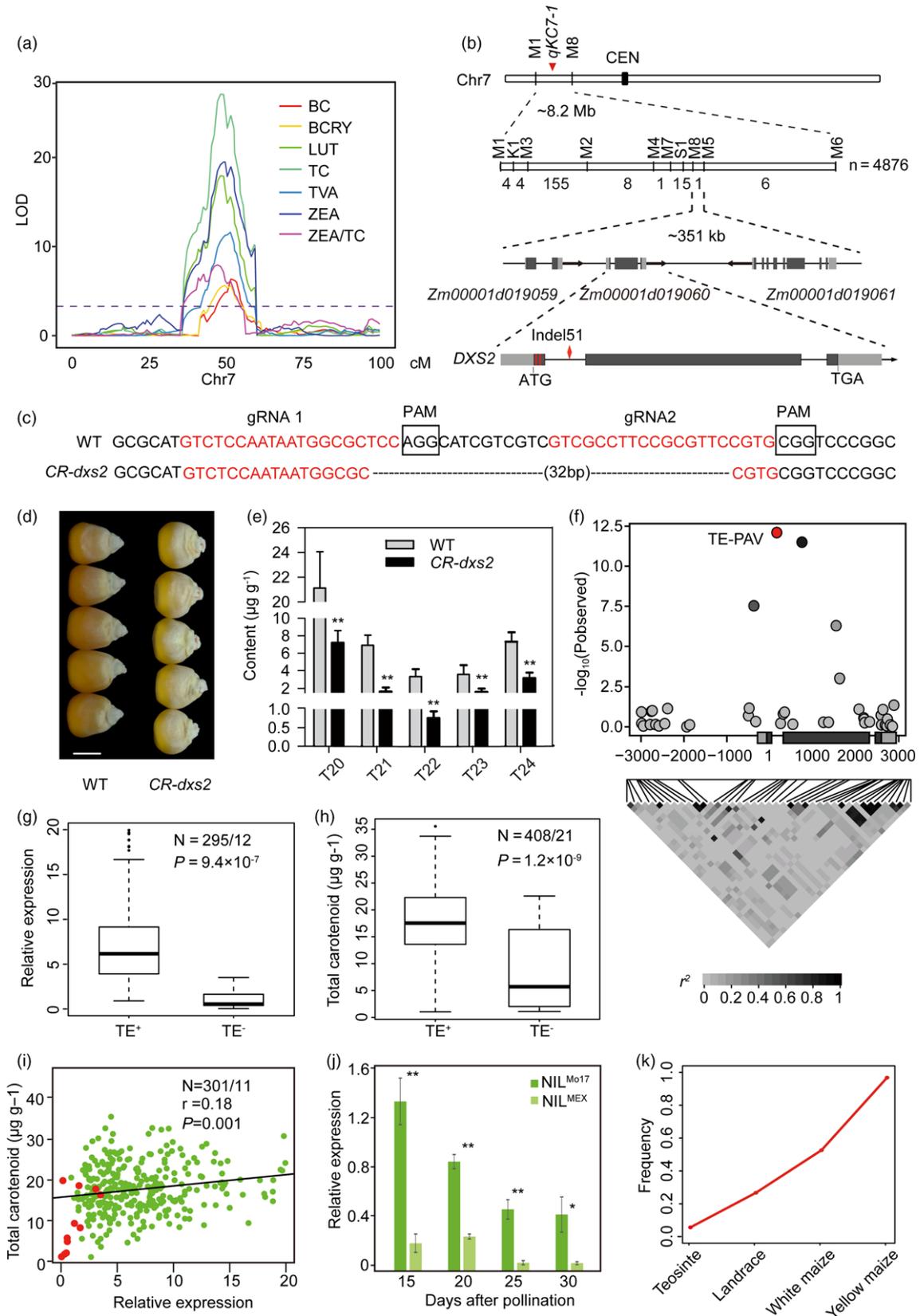
Figure 5. Selection patterns of known favorable alleles for oil and carotenoid traits during maize domestication and improvement. (a–d) LOD profiles for the identified QTLs that co-localized with previously cloned genes. The gene structure and the position of their functional sites are also shown. The vertical dashed lines indicate the position of four genes, and the red diamonds indicate the positions of functional sites. Dark-gray and light-gray shading represents exons and UTRs, respectively. (e–h) Frequency changes for favorable alleles of *ZmfatB* (e), *DGAT1-2* (f), *lcyE* (g), and *crtRB1* (h) during maize domestication and improvement.

finding that the landraces lacked this insertion, whereas its frequency was higher in regular maize (18.6%) and high-oil maize (34.5%), indicating that this insertion might occur during maize improvement (Figure 5e). In summary, these findings again revealed that three distinct selection patterns were identified for genes that affect kernel nutritional traits. In addition, the selection patterns of the genes or loci for oil and carotenoid traits are complex.

A natural variant undergone selection had pleiotropic effects on carotenoid traits

A large-effect QTL, namely L29 (hereafter, *qKC7-1*) on chromosome seven, was detected for seven carotenoid-related traits including TC, BC, BCRY, ZEA, LUT, TVA and ZEA/TC, which explained 10.7–42.0% of phenotypic variation (Figure 6a). There were multiple domestication and improvement signals in the QTL interval (Figure S4 and Data S2). To elucidate the selection feature of *qKC7-1*, we carried out

Figure 6. *DXS2* alters carotenoid biosynthesis in maize kernels, and changes in allele frequency of the favorable *DXS2* allele during maize domestication and improvement. (a) LOD profiles of the QTL cluster *qKC7-1* for carotenoid-related traits. (b) Fine mapping of *qKC7-1* and gene structure of *DXS2*. Numbers near markers denote the number of recombinants. Dark-gray and light-gray shading represents exons and UTRs, respectively. The red diamond indicates the 51-bp InDel in exon 1 of *DXS2*. The red bars in exon 1 indicate the targeting sites for sgRNA. (c) Sequences of *DXS2* in a *CR-dxs2* plant and its wide type (WT). Nucleotides in red denote the sgRNA targets, and boxed nucleotides indicate the protospacer-adjacent motif (PAM, i.e. NGG). (d) Kernel performance for *CR-dxs2* and WT. Scale bar, 0.5 cm. (e) Comparison of kernel carotenoids between *CR-dxs2* and WT. Error bars indicate SE based on five biological replicates. * $P < 0.05$, ** $P < 0.01$, Student's *t*-test. (f) *DXS2*-based candidate-gene association analysis for kernel color and the linkage disequilibrium (LD) patterns for all identified variants ($MAF \geq 0.05$) at the *DXS2* locus. The most significant TE-PAV is shown in red. The intensity of gray shading indicates the extent of LD (r^2) between TE-PAV and the other variants identified in this region. The gene structure is shown on the x-axis. Dark-gray and light-gray shading represents exons and UTRs, respectively. (g) Comparison of *DXS2* expression in TE⁺ and TE⁻ genotypes. (h) The effect of TE-PAV in an association mapping panel. (i) Correlation of kernel total carotenoid content with the relative expression of *DXS2* in kernels at 15 DAP. Dot color: red, TE⁻; green, TE⁺. (j) Expression pattern of *DXS2* in endosperms at different developmental stages. NIL, near-isogenic line. (k) Changes in the frequency of the TE during maize domestication and improvement.



fine mapping of *qKC7-1* using a near-isogenic line (NIL) population that was developed by twice backcrossing Mo17 with a line having the homologous teosinte allele at the *qKC7-1* locus (Figure S1). Using 10 markers developed at the *qKC7-1* locus and 4,876 plants in the NIL population, *qKC7-1* was narrowed to a 351-kb region flanked by markers S1 and M5, which harbor three annotated genes: *Zm00001d019059*, *Zm00001d019060* and *Zm00001d019061* (Figures 6b and S5). *Zm00001d019060* encodes 1-deoxy-D-xylulose-5-phosphate synthase 2 (*DXS2*), which catalyzes glyceraldehyde 3-phosphate and pyruvate to produce 1-deoxy-D-xylulose-5-phosphate in the methylerythritol phosphate (MEP) pathway (Cordoba *et al.*, 2011). The Arabidopsis homolog of *DXS2* is reportedly associated with the regulation of carotenoid biosynthesis (Estévez *et al.*, 2001). Consequently, *DXS2* was considered as the candidate gene for the *qKC7-1* locus. Therefore, we engineered mutations in *DXS2* in the maize inbred line C01 using the CRISPR/Cas9 technology, which yielded one null mutant referred to here as *CR-dxs2* (Figure 6c–e). Genome sequencing identified a 32-bp deletion in exon one of *DXS2*, resulting in a frame shift (Figure 6c). Along with our results for the pale-yellow kernels of *CR-dxs2* and the significant changes in four primary carotenoid traits (T21–24) as well as total carotenoid content (T20) (Figure 6d–e), these results confirmed that the *qKC7-1* locus encodes *DXS2*.

Sequence analysis revealed that a set of 83 SNPs and 16 InDels in the promoter region and full-length of *DXS2* differed between Mo17 and MEX (Figure S6). To identify the potential causative variant of *DXS2*, we extracted 39 SNPs with $MAF \geq 0.05$ across full-length *DXS2* from a dataset of 456 maize inbreds. In addition, four InDels were genotyped in the maize panel, including InDel863 and InDel292 in the promoter and InDel51 in intron one and InDel3 in exon two (Figure S6 and Data S1). No polymorphisms were detected for the InDel863 and InDel292 sites in the maize association panel, whereas three and five alleles were identified for the InDel51 and InDel3 sites, respectively. Interestingly, a 78-bp insertion harboring a mutator distance–relative (MuDR) TE as well as the 51-bp insertion (Figure S6) were identified for the InDel51 site in the maize association panel, but total carotenoid content did not differ between plants harboring these two different insertions (Figure S7). Thus, the InDel51 site was subsequently referred to as a TE-presence/absence variant (TE-PAV). Marker-trait association analysis identified seven variants that were significantly associated with kernel color or total carotenoid content ($P \leq 2.5 \times 10^{-4}$; Figures 6f and S8 and Table S3). Among these variants, only TE-PAV was significantly associated with both kernel color ($P = 7.29 \times 10^{-13}$) and total carotenoid content ($P = 4.5 \times 10^{-5}$), indicating that TE-PAV is a potential functional site of *DXS2*. Analysis of *DXS2* expression in 15-days after pollination (DAP) kernels from 368 inbreds (Fu *et al.*, 2013a) indicated that the lines harboring

the TE insertion expressed *DXS2* at high levels and that the expression level was correlated with total carotenoid content (Figure 6g–i). In addition, *DXS2* expression was upregulated in endosperms at different developmental stages of the NIL carrying the Mo17 allele (Figure 6j), indicating that the TE insertion might upregulate *DXS2* expression and consequently enhance the production of carotenoids. Collectively, these data suggested that TE-PAV is the causal variant for *DXS2*.

To examine the evolutionary origin of the TE at the *DXS2* locus, we genotyped the TE-PAV in a larger set of teosintes, landraces and maize inbreds. We found that 5.7% (9.5/168) of the teosinte entries, 22.2% (62/279) of the landraces, 52.6% (10/19) of the white maize inbreds, and 96.8% (422/436) of the yellow maize inbreds harbor the TE allele (Figure 6k). This result suggested the TE exists in teosinte, underwent selection during maize domestication and improvement, and is nearly fixed in yellow maize. To determine whether selection acted on the TE, we analyzed nucleotide diversity across the promoter region and full-length *DXS2* in 67 maize inbreds with the TE insertion, 14 maize inbreds without the TE insertion, and 40 teosinte entries. Nucleotide diversity was substantially reduced in the A (promoter, 5'UTR and exon one), B (intron one), and C (exon two) regions of *DXS2* in maize inbreds harboring the TE, especially in the B region, which retained only 2.2% of the nucleotide diversity of teosinte (Figure S9). A coalescence simulation incorporating the maize domestication bottleneck revealed that this severe loss of genetic diversity in the A, B, C regions of *DXS2* cannot be explained by the maize domestication bottleneck alone, indicating strong selection near the TE (Figure S9b).

DISCUSSION

Kernel nutritional traits such as oil and carotenoids in maize have been the target of domestication and improvement. Owing to the diverse application of maize in human food and animal feed, divergent selection for oil and carotenoids has occurred between regular and high-oil maize and between white and yellow maize, respectively. Compared with the oil content of teosinte (5.6%), that of regular maize and landraces is lower (~4%) although it is higher (> 6%) in high-oil maize (Flint-Garcia *et al.*, 2009; Yang and Li, 2018). Yellow maize (maize or landraces) has a higher carotenoid content than teosinte or white maize, which has white endosperms (Steenbock and Boutwell, 1920). Our results demonstrated that similar features of genetic architecture underlie the oil and carotenoid variations between maize and teosinte, although the oil and carotenoid traits share a few common loci. For each examined trait, the additive effects rather than epistatic effects were found to play a main role in the simple genetic basis of kernel oil and carotenoid variation. In addition, nearly half of detected traits (15/33) appeared to be controlled by one or

a few large-effect QTL (1–2 QTL with PVE \geq 15%) plus a few small-effect QTLs (0–4 with PVE < 15%). Because of the short amount of time required to observe variations, traits that came under strong and recent selection during and after domestication should have architectures with fewer genes and/or larger effect sizes (Wallace *et al.*, 2014; Xu *et al.*, 2017). Therefore, oil and carotenoid traits underwent strong and recent selection during and after domestication, consistent with their evolutionary history. In addition, any large-effect alleles for such traits would presumably have been fixed rapidly (Orr, 1998). For example, the 3-bp insertion of *DGAT1-2*, which was almost lost in landraces and regular maize, became rapidly fixed to 0.97 in high-oil maize after artificial selection for a short period (Zheng *et al.*, 2008).

Of the 42 loci for oil and carotenoid traits detected in the TM population, 13 loci were specific for teosinte and were not identified in maize–maize linkage populations (Data S2) (Song *et al.*, 2004; Wong *et al.*, 2004; Chander *et al.*, 2008; Zhang *et al.*, 2008; Wassom *et al.*, 2008a,b; Yang *et al.*, 2010; Cook *et al.*, 2012; Kandianis *et al.*, 2013; Jittham *et al.*, 2017; Venado *et al.*, 2017). These loci were the most likely to have experienced selection during maize domestication, whereas only L11 and L23 have domestication features inferred by selection analysis using genomic data (Data S2). This situation might be a consequence of the limited maize–maize and maize–teosinte populations used for QTL mapping or the limited sample size of maize, landrace, and teosinte used for selection analysis using genomic data. In addition, the identified QTL intervals were, on average, 13.4 Mb, often making it difficult to identify the exact target of selection. Therefore, the cloning of QTLs and subsequent selection analysis in a large sample will clarify the selection features of the QTLs we identified, such as *DXS2* in the current study.

Through fine mapping, CRISPR/Cas9-mediated gene editing and candidate-gene association mapping, we found that *DXS2* is the causal gene for the QTL cluster that contributes to seven carotenoid traits, and the MuDR TE in intron one of *DXS2* appeared to represent the causal variant. *DXS2* is the rate-limiting enzyme in the MEP pathway, which produces the metabolic precursors for plastid carotenoid biosynthesis (Carretero-Paulet *et al.*, 2006; Cordoba *et al.*, 2011). The loss of *DXS2* function weakened the ability of maize kernels to synthesize carotenoids, and this is also true for *SCD* in the penultimate step of the MEP pathway (Zhang *et al.*, 2019). In addition, a previous pathway-level analysis of 281 maize inbred lines revealed that *DXS2* could affect carotenoid traits (Owens *et al.*, 2014). The positive correlation between *DXS2* expression in developing kernels and total carotenoid content in 368 maize inbreds revealed that an enhancement of *DXS2* expression would increase kernel carotenoid content, which we in fact observed in the NILs. Similarly, overexpression of the

orthologs of *DXS2* in *Arabidopsis*, potato, and soybean forced a metabolic flux to the downstream pathway, consequently affecting the carotenoid accumulation (Estévez *et al.*, 2001; Morris *et al.*, 2006; Zhang *et al.*, 2009). Taken together with *DXS2* expression pattern, the MuDR TE seems to increase carotenoid accumulation in maize kernels by enhancing the expression of *DXS2*, which might be a consequence of the presence of the predicted CAAAT motif in the TE sequence (Figure S6). The previous discovery of the enhancer element in the introns acting as a *cis* regulator partly supports our results (Gidekel *et al.*, 1996; Hauck *et al.*, 1999; Hural *et al.*, 2000; Le Hir *et al.*, 2003). However, further work is required to explain how the MuDR TE in intron one enhances *DXS2* expression. These results confirm the association between *DXS2* and carotenoid accumulation.

Crop domestication and improvement is a dynamic and continuous process that strongly reflects human preference and styles of crop production. The dynamic frequency changes in nine favorable alleles from five genes for oil and carotenoid traits revealed that the strength and occurrence of selection varied for alleles associated with desirable nutritional traits. It is consistent with diverse human demands for food sources during different periods of history. Striking examples are genes underlying the divergence between regular and high-oil maize, e.g., *DGAT1-2* (Zheng *et al.*, 2008; Chai *et al.*, 2012), and between white and yellow maize, e.g., *DXS2*, *lcyE*, *crtRB1* and *PSY1* (Palaisa *et al.*, 2003; Fu *et al.*, 2010). In comparison with nutritional traits, the traits like yield, plant architecture, seed shattering, and dormancy have tended to undergo divergent selection between teosinte and maize. Selection for higher yield with increased seed size, decreased branching and tillering, reduced grain shattering, and reduction or loss of dormancy in maize often resulted in the near fixation of the key genes associated with these traits, e.g. *tb1* (Doebly *et al.*, 1997; Studer *et al.*, 2011) and *tga1* (Wang *et al.*, 2005). Interestingly, teosinte harbors all of the nine favorable alleles of the five genes for oil and carotenoid traits, suggesting that teosinte can be exploited for the improvement of kernel nutritional traits in modern maize germplasm. In addition, 77.8% (7/9) favorable alleles are not fixed in high-oil maize and yellow maize, underscoring the feasibility of introducing these alleles to breed maize with high nutritional traits.

Selection of quantitative traits associated with the domestication and improvement syndromes may result in inadvertent trade-offs when the genes or loci underlying the traits are either pleiotropic or tightly linked. For instance, *tb1*, a typical domestication gene, controls differences in shoot architecture between maize and teosinte (Doebly *et al.*, 1997). When *TB1* underwent selection, multiple traits including plant architecture, inflorescence, and seed architecture were simultaneously changed. In the

current study, three loci for both oil and carotenoid traits, namely L3, L9 and L39, were mapped to the same region, indicating the possibility that selection could simultaneously act on oil and carotenoid traits during maize domestication and improvement. Therefore, the construction of a trait–locus network provided potential genetic information to understand the common genetic basis of trait selection before the genes underlying the QTL for oil and carotenoid traits were well characterized. In addition, the construction of a trait–locus network allows us to pyramid desirable alleles to breed elite lines with target traits.

EXPERIMENTAL PROCEDURES

Plant materials and field trials

A BC₂F₅ population (hereafter, TM population), containing 191 lines, was derived from a single cross between Mo17 and *Zea mays* ssp. *mexicana* (PI566686; hereafter, MEX). Detailed information about the development of the TM population has been published (Pan *et al.*, 2016; Yang *et al.*, 2017). The TM population, together with their parents, was previously genotyped using the Illumina MaizeSNP50 array, and a high-density bin map with the total length of 1757 cM was constructed using 1282 recombinant bins derived from 12 390 high-quality SNPs (Pan *et al.*, 2016). The TM population was planted in a randomized complete block design with one replicate in three environments including Beijing in 2013, Hainan in 2013, and Inner Mongolia in 2014. Each family line was grown in a single-row plot (2.5-m rows, 0.67 m between rows), and planting density was 45 000 plants/ha. All plants were self-pollinated and harvested at maturity. For chemical analysis, 20 kernels were randomly selected from the middle of each of five well grown ears (100 kernels in total).

Measurement of fatty acids and carotenoids in maize kernels

To assess fatty acid compositions, 50 kernels were dried at 45°C for 60 h and then ground into powder. Lipids were extracted using methods modified from Sukhija and Palmquist (1988). Gas chromatography (GC) was carried out with a HP7890A GC system (Agilent Technologies, USA) with a HP-INNOWAX polyethylene glycol capillary column (30 m × 320 μm × 0.5 μm, Agilent Technologies). Two replicates were measured for each sample. Oil content was calculated as the sum of all identified fatty acid concentrations as a percentage of kernel weight, and individual fatty acids are expressed as a percentage of oil content. The detailed protocol is described in Yang *et al.* (2010). Nine distinct fatty acids were detected: palmitic (C16:0), palmitoleic (C16:1), stearic (C18:0), oleic (C18:1), linoleic (C18:2), linolenic (C18:3), arachidic (C20:0), behenic (C22:0), and lignoceric (C24:0). In addition, nine ratio traits were derived from nine fatty acids: C16:0/C16:1, C16:1/C18:0, C18:0/C18:1, C18:1/C18:2, C18:2/C18:3, C18:0/C20:0, C20:0/C22:0, C22:0/C24:0, SFA/USFA (saturated fatty acid = C16:0 + C18:0 + C20:0 + C22:0 + C24:0; unsaturated fatty acid = C16:1 + C18:1 + C18:2 + C18:3).

For quantifying carotenoids, 50 air-dried kernels were ground into powder for analysis by high-performance liquid chromatography (HPLC). Carotenoids were extracted from kernels according to Kurilich and Juvik (1999) and analyzed using a Prominence HPLC unit (Agilent Technologies) equipped with a YMC[®] HPLC C30 Column (250 mm × 4.6 mm, 5 μm; water). Quantification was achieved by standard regression with external standards (Sigma,

St. Louis, MO, USA). The detailed protocol is described in Chander *et al.* (2008). Measured carotenoids include β-carotenoid (BC), β-cryptoxanthin (BCRY), zeaxanthin (ZEA) and lutein (LUT). Sums and ratios of carotenoids were also calculated, containing TC (total carotenoids = BC + BCRY + ZEA + LUT); VA (provitamin A = BC + 0.5 × BCRY), BC/TC, BCRY/TC, ZEA/TC, LUT/TC, BC/BCRY, BC/ZEA, BCRY/ZEA; ratio (α-to-β branch ratio = LUT/(BC + BCRY + ZEA)).

Analysis of phenotypic data

R version 3.1.1 (www.R-project.org) was used to carry out all statistical analyses. ANOVA was performed for each trait in the R function 'aov' to evaluate the genotype and environment effect. The model for ANOVA was $y = \mu + \alpha_g + \beta_e + \varepsilon$, where α_g is the effect of the g th genotype, β_e is the effect of the e th environment, and ε is error. The broad-sense heritability of each trait was calculated as: $H^2 = \frac{\sigma^2_g}{\sigma^2_g + \sigma^2_e}$ (Knapp *et al.*, 1985), where σ^2_g is the genetic variance, σ^2_e is the residual variance, and e is the number of environments. All variances were obtained from ANOVA, treating genotyping and environment as random effects. To eliminate the influence of environmental effects, the best linear unbiased predictor (BLUP) value for each line was calculated using the linear mixed model considering both genotype and environment as random effects in the R function 'lme4'. The BLUP values for each line were used to perform phenotype description statistics, calculate Pearson correlation coefficients, and carry out QTL mapping.

QTL mapping

QTL mapping was performed using composite interval mapping implemented in Windows QTL Cartographer 2.5 (Wang, 2007). Model six of the Zmapqtl module was used to detect QTLs throughout the genome by scanning with a 2.0-cM interval between markers with a 10-cM window size. Forward-backward stepwise regression with five controlling markers was used to control background from flanking markers. The threshold logarithm of odds (LOD) value for putative QTLs was determined after 1000 permutations at a significance level of $P < 0.05$. The confidence interval for QTL position was estimated with the 1.5-LOD support interval method (Lander and Botstein, 1989). The R function 'lm' was performed to determine total phenotypic variation explained (PVE) by significant QTLs (Li *et al.*, 2013).

Epistasis analysis

The peak bin markers were used in the epistatic interaction analysis. For simplicity, all the heterozygous genotypes were assigned as missing values to ensure that only homozygous allelic interactions were estimated and tested. Two-way ANOVA was used to estimate the pairwise additive-by-additive epistatic interactions for all identified QTLs for each trait ($P < 0.05$) (Yu *et al.*, 1997; Wen *et al.*, 2016). Epistatic effects were estimated by comparing the residual of the full model containing all single-locus effects and two-locus interaction effects with that of the reduced model excluding two-locus interaction effects.

Construction of a trait–locus network

The trait–locus network based on the results of QTL mapping was constructed using Cytoscape 3.2.0 (Shannon *et al.*, 2003). Traits and their corresponding QTLs were taken as nodes and the link between traits and QTLs (overlap based on physical position) as edges. The thickness of the links between traits and QTLs reflects

the PVE value of QTLs, and the size of nodes indicates the number of QTLs.

Analysis of selected features of the detected QTL

We defined the selection features of the identified QTL according to the values of the 10-kb window XP-CLR and π ratio in the QTL interval, which were previously estimated using ~21 million high-quality SNPs in 35 maize inbreds, 23 landraces, and 17 wild relatives (Hufford *et al.*, 2012). Evidence for selection during domestication and improvement processes was evaluated in two comparisons: teosinte versus landraces for domestication, and landraces versus maize inbreds for improvement. The top 5% of XP-CLR and the reduced nucleotide diversity ($\pi_{\text{teosinte}}/\pi_{\text{landrace}} \geq 4.66$; $\pi_{\text{landrace}}/\pi_{\text{maize}} \geq 2.07$) were set as the threshold for putative selection signals. When the values of both XP-CLR and π ratio of the same genomic region within the QTL interval exceeded the threshold, we defined the detected QTLs as 'domestication' or 'improvement'.

Molecular evolution of four known oil and carotenoid genes

To elucidate the molecular evolution of the four genes, *DGAT1-2*, *ZmfatB*, *lcyE*, and *crtRB1*, during maize domestication and improvement, one or three functional sites of each gene were genotyped in 539 maize inbred lines including 35 high-oil lines and 22 white maize inbreds, 279 landraces, and 168 teosinte entries (Table S1; Data S1). For *DGAT1-2*, the function site was amplified and sequenced using primers DGAT-1F and DGAT-2R (Table S1).

Fine mapping of *qKC7-1*

To fine map *qKC7-1*, line TM75 containing the teosinte allele at this locus (Figure S1) was used to generate the BC₂-BC₅ population by backcrossing with Mo17. Ten markers were developed to screen the recombinants using 4876 plants of the BC₂-BC₅ population (Table S1). The selfed recombinants with pale-yellow kernels contained homologous MEX alleles, those with yellow kernels contained homologous Mo17 alleles, and those with segregating color contained heterozygous alleles. Therefore, *qKC7-1* was fine mapped using the overlapped recombinants.

Generation and analysis of mutants generated by CRISPR/Cas9 technology

The CRISPR/Cas9 binary vector CPB-ZmUbi-hspCas9-*DXS2* was constructed to produce defined deletions in *DXS2* exon one using one sgRNA (single guide RNA) alongside the Cas9 endonuclease gene (Table S1) (Li *et al.*, 2017). The construct was transformed into *Agrobacterium tumefaciens* strain EHA105, and immature embryos of maize inbred line C01 were subjected to *Agrobacterium*-mediated transformation. T0 transgenic plants were genotyped using vector-specific primers and target-specific primers (Table S1). PCR products amplified using target-specific primers were cloned by using the pEASY T5 ZERO Cloning kit (TransGen Biotech, Beijing, China), and a minimum of six clones per PCR product were sequenced. Homozygous mutations in edited target genes, together with C01, were further planted for carotenoid quantification.

RNA extraction and reverse transcription quantitative PCR (RT-qPCR)

Total RNA was extracted from various tissues using the RNAprep Pure Plant kit (Tiangen Biotech, Beijing, China). First-strand cDNA was synthesized using the PrimeScript II 1st strand cDNA synthesis

kit (TaKaRa, Otsu, Japan). RT-qPCR was carried out in triplicate for each sample using the SYBR Green kit (Mei5 Biotech, Beijing, China) with the CFX CONNET real-time PCR system (Bio-Rad, CA, USA). Maize *ACT1N* was used as a control for normalization between samples (Table S1). Relative transcript levels were calculated using the comparative threshold cycle method (Livak and Schmittgen, 2001).

Candidate-gene association mapping

DXS2-based association mapping was performed using a subset of 435 inbreds with yellow kernels and 21 inbreds with white kernels (Data S1) (Yang *et al.*, 2011). SNPs in the promoter region as well as full-length *DXS2* were extracted from resequencing data for the 456 inbreds (Data S1). The promoter region and the full-length *DXS2* was sequenced in Mo17 and MEX using three pairs of primers (Table S1). A 863-bp InDel (InDel863) and a 292-bp InDel (InDel292) in the promoter, a 51-bp InDel (InDel51) in intron one, and a 3-bp InDel (InDel3) in exon 2 were genotyped in the whole maize panel using specific primers for each region (Table S1). The associations between all polymorphic sites including SNPs and InDels with minor allele frequency (MAF) ≥ 0.05 and kernel color and total carotenoid content (Fu *et al.*, 2013b) were analyzed using a mixed linear model in TASSEL 3.0 (Bradbury *et al.*, 2007) considering population structure and the kinship matrix which was conducted with ADMIXTURE (Alexander *et al.*, 2009) and TASSEL5.0 (Bradbury *et al.*, 2007), respectively. A Bonferroni adjusted significance threshold ($P \leq 0.01/40 = 2.5 \times 10^{-4}$) was used to identify significant associations.

Correlation between *DXS2* expression and total carotenoid content

The analysis of the correlation between *DXS2* expression in developing kernels at 15 DAP (Fu *et al.*, 2013a) and total carotenoid content in mature kernels (Fu *et al.*, 2013b) was performed in the R function 'cor.test'.

Nucleotide diversity and molecular evolution of *DXS2*

The promoter region and full-length *DXS2* was resequenced in 67 maize inbreds having a transposable element (TE) insertion, 14 maize inbreds without a TE insertion, and 40 teosinte entries (Data S1) using three primer pairs (Table S1). PCR products from maize were directly sequenced, whereas those from teosinte were cloned by using the pEASY T5 ZERO Cloning kit (TransGen Biotech), and one randomly chosen clone per PCR product was sequenced. The sequences were assembled using ContigExpress in Vector NTI Advance 10 (Invitrogen) and aligned and manually corrected using BioEdit (Hall, 1999). The nucleotide diversity was calculated using DNAsP5.0 software (Rozas *et al.*, 2003).

To further test whether the observed loss of genetic diversity in maize relative to that in teosinte could be explained by a domestication bottleneck alone, coalescent simulations that incorporated the domestication bottleneck (Eyre-Walker *et al.*, 1998; Tenaillon *et al.*, 2004; Wright *et al.*, 2005; Tian *et al.*, 2009) were performed for the sequenced regions of *DXS2*. The parameters $4Nc$ (population recombination) and θ (population mutation) were estimated from the teosinte data using DNAsP version 5.0 (Rozas *et al.*, 2003). In total, 10 000 coalescent simulations were performed using Hudson's ms program (Suárez-López *et al.*, 2001).

ACCESSION NUMBERS

Sequence data from this study can be found in the GenBank database under accession number MN026742–MN026862 for *DXS2*.

AUTHOR CONTRIBUTIONS

XY designed the research. HF, XF, JX and LZ conducted field trials and phenotyping. HF, XF, JO, MW and XL performed QTL mapping. XF, JX and WL performed fine mapping of *qKC7-1*. YW and J-TX generated the CRISPR/Cas9-mediated mutant. HF, YW, XF and XZ performed candidate-gene association mapping. HF, YW, NY and GX performed selection analysis. JL developed the TM population. JY assisted with the experiments. HF, XF, YW and XY wrote the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

This research was supported by the National Key Research and Development Program of China (2016YFD0100503) and the National Natural Science Foundation of China (31421005, 31722039, 31361140362).

COMPETING INTERESTS

The authors declare that they have no conflict of competing interest.

Data S1. Plant materials and genotypes used for candidate-gene association analysis and molecular evolution in this study.

Data S2. Summary for single and epistatic QTLs identified in the TM population.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Genomic composition of line TM75.

Figure S2. Phenotypic distribution of 33 oil and carotenoid traits in the TM population.

Figure S3. Selection features for six loci that co-localized with six known genes identified in the TM population.

Figure S4. Selection features of *qKC7-1* (L29) for seven carotenoid traits identified in the TM population.

Figure S5. *qKC7-1* was fine mapped to a 351-kb region on chromosome 7.

Figure S6. DNA sequence alignment of *DXS2* between NIL^{Mo17} and NIL^{MEX} (NIL, near-isogenic line).

Figure S7. Effects of the 51- and 78-bp insertion in an association mapping panel.

Figure S8. *DXS2*-based candidate-gene association analysis for total carotenoid content in maize kernels.

Figure S9. Molecular evolution of *DXS2*.

Table S1. Primers used in this study.

Table S2. Statistic summary, broad-sense heritability, and variances of 33 oil- and carotenoid-related traits in the TM population.

Table S3. Associations between *DXS2* polymorphisms and kernel color and total carotenoid content in 456 maize inbred lines.

REFERENCES

- Alexander, D.H., Novembre, J. and Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664. <https://doi.org/10.1101/gr.094052.109>.
- Benz, B., Perales, H. and Brush, S. (2007) Tzeltal and Tzotzil farmer knowledge and maize diversity in Chiapas, Mexico. *Cur Anthropol.* **48**, 289–300. <https://doi.org/10.1086/512986>.
- Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, **23**, 2633–2635. <https://doi.org/10.1093/bioinformatics/btm308>.
- Carretero-Paulet, L., Cairo, A., Botella-Pavía, P., Besumbes, O., Campos, N., Boronat, A. and Rodríguez-Concepción, M. (2006) Enhanced flux through the methylerythritol 4-phosphate pathway in Arabidopsis plants overexpressing deoxyxylulose 5-phosphate reductoisomerase. *Plant Mol. Biol.* **62**, 683–695. <https://doi.org/10.1007/s11103-006-9051-9>.
- Chai, Y.C., Hao, X.M., Yang, X.H., Allen, W.B., Li, J.M., Yan, J.B., Shen, B. and Li, J.S. (2012) Validation of DGAT1-2 polymorphisms associated with oil content and development of functional markers for molecular breeding of high-oil maize. *Mol. Breeding*, **29**, 939–949. <https://doi.org/10.1007/s11032-011-9644-0>.
- Chander, S., Guo, Y.Q., Yang, X., Zhang, J., Lu, X.Q., Yan, J.B., Song, T.M., Rocheford, T.R. and Li, J.S. (2008) Using molecular markers to identify two major loci controlling carotenoid contents in maize grain. *Theor. Appl. Genet.* **116**, 223–233. <https://doi.org/10.1007/s00122-007-0661-7>.
- Cook, J.P., McMullen, M.D., Holland, J.B., Tian, F., Bradbury, P., Ross-Ibarra, J., Buckler, E.S. and Flint-Garcia, S.A. (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. *Plant Physiol.* **158**, 824–834. <https://doi.org/10.1104/pp.111.185033>.
- Cordoba, E., Porta, H., Arroyo, A., Román, C.S., Medina, L., Rodríguez-Concepción, M. and León, P. (2011) Functional characterization of the three genes encoding 1-deoxy-D-xylulose 5-phosphate synthase in maize. *J. Exp. Bot.* **62**, 2023–2038. <https://doi.org/10.1104/pp.111.185033>.
- Doebley, J. (2004) The genetics of maize evolution. *Annu. Rev. Genet.* **38**, 37–59. <https://doi.org/10.1146/annurev.genet.38.072902.092425>.
- Doebley, J. and Stec, A. (1993) Inheritance of the morphological differences between maize and teosinte: comparison of results for two F2 populations. *Genetics*, **134**, 559–570.
- Doebley, J., Stec, A. and Gustus, C. (1995) Teosinte branched1 and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics*, **141**, 333–346.
- Doebley, J., Stec, A. and Hubbard, L. (1997) The evolution of apical dominance in maize. *Nature*, **386**, 485–488. <https://doi.org/10.1038/386485a0>.
- Dudley, J.W. and Lambert, R.J. (2004) 100 generations of selection for oil and protein in corn. *Plant Breed. Rev.* **24**, 79–110. <https://doi.org/10.1002/9780470650240.ch5>.
- Estévez, J.M., Cantero, A., Reindl, A., Reichler, S. and León, P. (2001) 1-Deoxy-D-xylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants. *J. Biol. Chem.* **276**, 22901–22909. <https://doi.org/10.1074/jbc.M100854200>.
- Eyre-Walker, A., Gaut, R.L., Hilton, H., Feldman, D.L. and Gaut, B.S. (1998) Investigation of the bottleneck leading to the domestication of maize. *Proc. Natl Acad. Sci. USA*, **95**, 4441–4446. <https://doi.org/10.1073/pnas.95.8.4441>.
- Flint-Garcia, S.A. (2017) Kernel evolution: From teosinte to maize. In *Maize kernel development* (Brian, A., ed). Wallingford CABI, pp. 1–15.
- Flint-Garcia, S.A., Bodnar, A.L. and Scott, M.P. (2009) Wide variability in kernel composition, seed characteristics, and zein profiles among diverse maize inbreds, landraces, and teosinte. *Theor. Appl. Genet.* **119**, 1129–1142. <https://doi.org/10.1007/s00122-009-1115-1>.
- Fu, Z.Y., Yan, J.B., Zheng, Y.P., Warburton, M.L., Crouch, J.H. and Li, J.S. (2010) Nucleotide diversity and molecular evolution of the PSY1 gene in Zea mays compared with some other grass species. *Theor. Appl. Genet.* **120**, 709–720. <https://doi.org/10.1007/s00122-009-1188-x>.
- Fu, J.J., Cheng, Y.B., Linghu, J.J. et al. (2013a) RNA sequencing reveals the complex regulatory network in the maize kernel. *Nat. Commun.* **4**, 2832–2843. <https://doi.org/10.1038/ncomms3832>.
- Fu, Z.Y., Chai, Y.C., Zhou, Y., Yang, X.H., Warburton, M.L., Xu, S.T., Cai, Y., Zhang, D.L., Li, J.S. and Yan, J.B. (2013b) Natural variation in the sequence of PSY1 and frequency of favorable polymorphisms among tropical and temperate maize germplasm. *Theor. Appl. Genet.* **126**, 923–935. <https://doi.org/10.1007/s00122-012-2026-0>.
- Gidekel, M., Jimenez, B. and Herrera-Estrella, L. (1996) The first intron of the *Arabidopsis thaliana* gene coding for elongation factor 1β contains

- an enhancer-like element. *Gene*, **170**, 201–206. [https://doi.org/10.1016/0378-1119\(95\)00837-3](https://doi.org/10.1016/0378-1119(95)00837-3).
- Guo, L., Wang, X.H., Zhao, M. *et al.* (2018) Stepwise cis-regulatory changes in ZCN8 contribute to maize flowering-time adaptation. *Curr. Biol.* **28**, 3005–3015. <https://doi.org/10.1016/j.cub.2018.07.029>.
- Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98.
- Hanson, M.A., Gaut, B.S., Stec, A.O., Fuerstenberg, S.I., Goodman, M.M., Coe, E.H. and Doebley, J.F. (1996) Evolution of anthocyanin biosynthesis in maize kernels: the role of regulatory and enzymatic loci. *Genetics*, **143**, 1395–1407.
- Harjes, C.E., Rocheford, T.R., Bai, L. *et al.* (2008) Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science*, **319**, 330–333. <https://doi.org/10.1126/science.1150255>.
- Hauck, B., Gehring, W.J. and Walldorf, U. (1999) Functional analysis of an eye specific enhancer of the eyeless gene in *Drosophila*. *Proc. Natl Acad. Sci. USA*, **96**, 564–569. <https://doi.org/10.1073/pnas.96.2.564>.
- Van Heerwaarden, J., Doebley, J., Briggs, W.H., Glaubitz, J.C., Goodman, M.M., De Gonzalez, J. S. and Ross-Ibarra, J. (2011) Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proc. Natl Acad. Sci. USA*, **108**, 1088–1092. <https://doi.org/10.1073/pnas.1013011108>.
- Le Hir, H., Nott, A. and Moore, M.J. (2003) How introns influence and enhance eukaryotic gene expression. *Trends Biochem. Sci.* **28**, 215–220. [https://doi.org/10.1016/S0968-0004\(03\)00052-5](https://doi.org/10.1016/S0968-0004(03)00052-5).
- Hopkins, C.G. (1899) Improvement in the chemical composition of the corn kernel. *J. Am. Chem. Soc.* **21**, 1039–1057. <https://doi.org/10.2135/1974>.
- Huang, C., Sun, H.Y., Xu, D.Y. *et al.* (2018) ZmCCT9 enhances maize adaptation to higher latitudes. *Proc. Natl Acad. Sci. USA*, **115**, 334–341. <https://doi.org/10.1073/pnas.1718058115>.
- Hufford, M.B., Xu, X., Van Heerwaarden, J. *et al.* (2012) Comparative population genomics of maize domestication and improvement. *Nat. Genet.* **44**, 808–811. <https://doi.org/10.1038/ng.2309>.
- Hural, J.A., Kwan, M., Henkel, G., Hock, M.B. and Brown, M.A. (2000) An intron transcriptional enhancer element regulates IL-4 gene locus accessibility in mast cells. *J. Immunol.* **165**, 3239–3249. <https://doi.org/10.4049/jimmunol.165.6.3239>.
- Jittham, O., Fu, X.Y., Xu, J., Chander, S., Li, J.S. and Yang, X.H. (2017) Genetic dissection of carotenoids in maize kernels using high-density single nucleotide polymorphism markers in a recombinant inbred line population. *Crop J.* **5**, 63–72. <https://doi.org/10.1016/j.cj.2016.06.006>.
- Kandianis, C.B., Stevens, R., Liu, W.P., Palacios, N., Montgomery, K., Pixley, K., White, W.S. and Rocheford, T. (2013) Genetic architecture controlling variation in grain carotenoid composition and concentrations in two maize populations. *Theor. Appl. Genet.* **126**, 2879–2895. <https://doi.org/10.1007/s00122-013-2179-5>.
- Knapp, S.J., Stroup, W.W. and Ross, W.M. (1985) Exact confidence intervals for heritability on a progeny mean basis 1. *Crop Sci.* **25**, 192–194. <https://doi.org/10.2135/cropsci1985.0011183X002500010046x>.
- Kurilich, A.C. and Juvik, J.A. (1999) Simultaneous quantification of carotenoids and tocopherols in corn kernel extracts by HPLC. *J. Liq. Chromatogr. R. T.* **22**, 2925–2934. <https://doi.org/10.1081/JLC-100102068>.
- Lander, E.S. and Botstein, D. (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **121**, 185–199.
- Li, L., Li, H., Li, Q. *et al.* (2011) An 11-bp insertion in *Zea mays fatb* reduces the palmitic acid content of fatty acids in maize grain. *PLoS ONE*, **6**, e24699. <https://doi.org/10.1371/journal.pone.0024699>.
- Li, H., Peng, Z.Y., Yang, X.H. *et al.* (2013) Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat. Genet.* **45**, 43–50. <https://doi.org/10.1038/ng.2484>.
- Li, C.X., Liu, C.L., Qi, X.T., Wu, Y.C., Fei, X.H., Mao, L., Cheng, B.J., Li, X. and Xie, C.X. (2017) RNA-guided Cas9 as an in vivo desired-target mutator in maize. *Plant Biol.* **15**, 1566–1576. <https://doi.org/10.1111/pbi.12739>.
- Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods*, **25**, 402–408. <https://doi.org/10.1006/meth.2001.1262>.
- Matsuoka, Y., Vigouroux, Y., Goodman, M.M., Sanchez, G.J., Buckler, E. and Doebley, J. (2002) A single domestication for maize shown by multi-locus microsatellite genotyping. *Proc. Natl Acad. Sci. USA*, **99**, 6080–6084. <https://doi.org/10.1073/pnas.052125199>.
- Meyer, R.S. and Purugganan, M.D. (2013) Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* **14**, 840–852. <https://doi.org/10.1038/nrg3605>.
- Morris, W.L., Ducreux, L.J.M., Hedden, P., Millam, S. and Taylor, M.A. (2006) Overexpression of a bacterial 1-deoxy-D-xylulose 5-phosphate synthase gene in potato tubers perturbs the isoprenoid metabolic network: implications for the control of the tuber life cycle. *J. Exp. Bot.* **57**, 3007–3018. <https://doi.org/10.1093/jxb/erl061>.
- Muzhingi, T., Palacios-Rojas, N., Miranda, A., Cabrera, M.L., Yeum, K.J. and Tang, G. (2017) Genetic variation of carotenoids, vitamin E and phenolic compounds in Provitamin A biofortified maize. *J. Sci. Food Agr.* **97**, 793–801. <https://doi.org/10.1002/jsfa.7798>.
- Orr, H.A. (1998) The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution*, **52**, 935–949. <https://doi.org/10.1111/j.1558-5646.1998.tb01823.x>.
- Owens, B.F., Lipka, A.E., Magallanes-Lundback, M., Tiede, T., Diepenbrock, C.H., Kandianis, C.B., Kim, E., Cepela, J., Mateos-Hernandez, M. and Buell, C.R. (2014) A foundation for provitamin A biofortification of maize: genome-wide association and genomic prediction models of carotenoid levels. *Genetics*, **198**, 1699–1716. <https://doi.org/10.1534/genetics.114.169979>.
- Palaisa, K.A., Morgante, M., Williams, M. and Rafalski, A. (2003) Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. *Plant Cell*, **15**, 1795–1806. <https://doi.org/10.1105/tpc.012526>.
- Pan, Q.C., Li, L., Yang, X.H., Tong, H., Xu, S.T., Li, Z.G., Li, W.Y., Muehlbauer, G.J., Li, J. and Yan, J.B. (2016) Genome-wide recombination dynamics are associated with phenotypic variation in maize. *New Phytol.* **210**, 1083–1094. <https://doi.org/10.1111/nph.13810>.
- Piperno, D.R., Ranere, A.J., Holst, I., Iriarte, J. and Dickau, R. (2009) Starch grain and phytolith evidence for early ninth millennium BP maize from the Central Balsas River Valley. *Mexico. Proc. Natl Acad. Sci. USA*, **106**, 5019–5024. <https://doi.org/10.1073/pnas.0812525106>.
- Poneleit, C.G. (2000) Breeding white endosperm corn. In *Specialty Corns*, 2nd edn. (Hallauer, A.R., ed). Florida, Boca Raton: CRC Press, pp. 235–273.
- Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X. and Rozas, R. (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497. <https://doi.org/10.1093/bioinformatics/btg359>.
- Shannon, P., Markiel, A., Ozier, O., Baliga, O.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504. <https://doi.org/10.1101/gr.1239303>.
- Song, X.F., Song, T.M., Dai, J.R., Rocheford, T. and Li, J.S. (2004) QTL mapping of kernel oil concentration with high-oil maize by SSR markers. *Maydica*, **49**, 41–48.
- Steenbock, H. and Boutwell, P.W. (1920) Fat-soluble vitamins. III. The comparative nutritive value of hite and yellow maizes. *J. Biol. Chem.* **41**, 81–96.
- Studer, A., Zhao, Q., Ross-Ibarra, J. and Doebley, J. (2011) Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nat. Genet.* **43**, 1160–1163. <https://doi.org/10.1038/ng.942>.
- Suárez-López, P., Wheatley, K., Robson, F., Onouchi, H., Valverde, F. and Coupland, G. (2001) CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis. *Nature*, **410**, 1116–1120. <https://doi.org/10.1038/35074138>.
- Sukhija, P.S. and Palmquist, D.L. (1988) Rapid method for determination of total fatty acid content and composition of feedstuffs and feces. *J. Agr. Food Chem.* **36**, 1202–1206. <https://doi.org/10.1021/jf00084a019>.
- Tenaillon, M.I. and U'ren, J., Tenaillon, O. and Gaut, B.S. (2004) Selection versus demography: a multilocus investigation of the domestication process in maize. *Mol. Bio. Evo.* **21**, 1214–1225. <https://doi.org/10.1093/molbev/msh102>.
- Tian, F., Stevens, N.M. and Buckler, E.S. (2009) Tracking footprints of maize domestication and evidence for a massive selective sweep on chromosome 10. *Proc. Natl Acad. Sci. USA*, **106**, 9979–9986. <https://doi.org/10.1073/pnas.0901122106>.
- Venado, R.E., Owens, B.F., Ortiz, D., Lawson, T., Mateos-Hernandez, M., Ferruzzi, M.G. and Rocheford, T.R. (2017) Genetic analysis of provitamin A carotenoid β -cryptoxanthin concentration and relationship with other

- carotenoids in maize grain (*Zea mays* L.). *Mol. Breeding*, **37**, 127. <https://doi.org/10.1007/s11032-017-0723-8>.
- Wallace, J.G., Larsson, S.J. and Buckler, E.S. (2014) Entering the second century of maize quantitative genetics. *Heredity*, **112**, 30–38. <https://doi.org/10.1038/hdy.2013.6>.
- Wang, S. (2007) *Windows QTL cartographer 2.5*. [WWW document]. <http://statgen.ncsu.edu/qtlcart/WQTLCart>.
- Wang, H., Nussbaum-Wagler, T., Li, B., Zhao, Q., Vigouroux, Y., Faller, M., Bomblies, K., Lukens, L. and Doebley, J.F. (2005) The origin of the naked grains of maize. *Nature*, **436**, 714–719. <https://doi.org/10.1038/nature03863>.
- Wassom, J.J., Mikkilineni, V., Bohn, M.O. and Rocheford, T.R. (2008a) QTL for fatty acid composition of maize kernel oil in Illinois High Oil × B73 backcross-derived lines. *Crop Sci.* **48**, 69–78. <https://doi.org/10.2135/cropsci2007.04.0208>.
- Wassom, J.J., Wong, J.C., Martinez, E., King, J.J., DeBaene, J., Hotchkiss, J.R., Mikkilineni, V., Bohn, M.O. and Rocheford, T.R. (2008b) QTL associated with maize kernel oil, protein, and starch concentrations; kernel mass; and grain yield in Illinois high oil × B73 backcross-derived lines. *Crop Sci.* **48**, 243–252. <https://doi.org/10.2135/cropsci2007.04.0205>.
- Wen, W.W., Liu, H.J., Zhou, Y. et al. (2016) Combining quantitative genetics approaches with regulatory network analysis to dissect the complex metabolism of the maize kernel. *Plant Physiol.* **170**, 136–146. <https://doi.org/10.1104/pp.15.01444>.
- Whitt, S.R., Wilson, L.M., Tenailon, M.I., Gaut, B.S. and Buckler, E.S. (2002) Genetic diversity and selection in the maize starch pathway. *Proc. Natl Acad. Sci. USA*, **99**, 12959–12962. <https://doi.org/10.1073/pnas.202476999>.
- Wong, J.C., Lambert, R.J., Wurtzel, E.T. and Rocheford, T.R. (2004) QTL and candidate genes phytoene synthase and ζ -carotene desaturase associated with the accumulation of carotenoids in maize. *Theor. Appl. Genet.* **108**, 349–359. <https://doi.org/10.1007/s00122-003-1436-4>.
- Wright, S.I., Bi, I.V., Schroeder, S.G., Yamasaki, M., Doebley, J.F., McMullen, M.D. and Gaut, B.S. (2005) The effects of artificial selection on the maize genome. *Science*, **308**, 1310–1314. <https://doi.org/10.1126/science.1107891>.
- Xu, G.H., Wang, X.F., Huang, C., et al. (2017) Complex genetic architecture underlies maize tassel domestication. *New Phytol.* **214**, 852–864. <https://doi.org/10.1111/nph.14400>.
- Yan, J.B., Kandianis, C.B., Harjes, C.E., et al. (2010) Rare genetic variation at *Zea mays crtRB1* increases β -carotene in maize grain. *Nat. Genet.* **42**, 322–327. <https://doi.org/10.1038/ng.551>.
- Yang, X.H. and Li, J.S. (2018) High-Oil Maize Genomics. In *The Maize Genome* (Benetzen, J., Flint-Garcia, S., Hirsch, C. and Tuberosa, R., eds). Cham: Springer, pp. 305–317.
- Yang, X.H., Guo, Y.Q., Yan, J.B., Zhang, J., Song, T.M., Rocheford, T. and Li, J.S. (2010) Major and minor QTL and epistasis contribute to fatty acid compositions and oil concentration in high-oil maize. *Theor. Appl. Genet.* **120**, 665–678. <https://doi.org/10.1007/s00122-009-1184-1>.
- Yang, X.H., Gao, S.B., Xu, S.T., Zhang, Z.X., Prasanna, B.M., Li, L., Li, J.S. and Yan, J.B. (2011) Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol. Breeding*, **28**, 511–526. <https://doi.org/10.1007/s11032-010-9500-7>.
- Yang, Q., Li, Z., Li, W.Q. et al. (2013) CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. *Proc. Natl Acad. Sci. USA*, **110**, 16969–16974. <https://doi.org/10.1073/pnas.1310949110>.
- Yang, N., Xu, X.W., Wang, R.R. et al. (2017) Contributions of *Zea mays* subspecies mexicana haplotypes to modern maize. *Nature Commun.* **8**, 1874–1883. <https://doi.org/10.1038/s41467-017-02063-5>.
- Yu, S.B., Li, J.X., Xu, C.G., Tan, Y., Gao, Y.F., Li, X.H., Zhang, Q.F. and Maroof, M.A.S. (1997) Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proc. Natl Acad. Sci. USA*, **94**, 9226–9231. <https://doi.org/10.1073/pnas.94.17.9226>.
- Zhang, J., Lu, X.Q., Song, X.F., Yan, J.B., Song, T.M., Dai, J.R., Rocheford, T. and Li, J.S. (2008) Mapping quantitative trait loci for oil, starch, and protein concentrations in grain with high-oil maize by SSR markers. *Euphytica*, **162**, 335–34. <https://doi.org/10.1007/s10681-007-9500-9>.
- Zhang, M., Li, K., Zhang, C.H., Gai, J.Y. and Yu, D.Y. (2009) Identification and characterization of class 1 DXS gene encoding 1-deoxy-D-xylulose-5-phosphate synthase, the first committed enzyme of the MEP pathway from soybean. *Mol. Bio. Rep.* **36**, 879–887. <https://doi.org/10.1007/s11033-008-9258-8>.
- Zhang, L.L., Zhang, X., Wang, X.J. et al. (2019) SEED CAROTENOID DEFICIENT Functions in Isoprenoid Biosynthesis via the Plastid MEP Pathway. *Plant Physiol.* **179**, 1723–1738. <https://doi.org/10.1104/pp.18.01148>.
- Zheng, P.Z., Allen, W.B., Roesler, K. et al. (2008) A phenylalanine in DGAT is a key determinant of oil content and composition in maize. *Nat. Genet.* **40**, 367–372. <https://doi.org/10.1038/ng.85>.