CellPress
PARTNER JOURNAL

Molecular Plant
Resource Article

QTG-Seq Accelerates QTL Fine Mapping through QTL Partitioning and Whole-Genome Sequencing of Bulked Segregant Samples

Hongwei Zhang[1,2,5], Xi Wang[2,5], Qingchun Pan[2,5], Pei Li[3,5], Yunjun Liu[1], Xiaoduo Lu[4], Wanshun Zhong[2], Minqi Li[2], Linqian Han[2], Juan Li[2], Pingxi Wang[1], Dongdong Li[1], Yan Liu[1], Qing Li[2], Fang Yang[2], Yuan-Ming Zhang[3,*], Guoying Wang[1,*] and Lin Li[2,*]

[1]Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China

[2]National Key Laboratory of Crop Genetic Improvement, Crop Information Center, Huazhong Agricultural University, Wuhan 430070, China

[3]Crop Information Center, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China

[4]Institute of Molecular Breeding for Maize, Qilu Normal University, Jinan 250200, China

[5]These authors contributed equally to this article.

***Correspondence: Lin Li (hzaulilin@mail.hzau.edu.cn), Guoying Wang (Wangguoying@caas.cn), Yuan-Ming Zhang (soyzhang@mail.hzau.edu.cn)**

https://doi.org/10.1016/j.molp.2018.12.018

ABSTRACT

Deciphering the genetic mechanisms underlying agronomic traits is of great importance for crop improvement. Most of these traits are controlled by multiple quantitative trait loci (QTLs), and identifying the underlying genes by conventional QTL fine-mapping is time-consuming and labor-intensive. Here, we devised a new method, named quantitative trait gene sequencing (QTG-seq), to accelerate QTL fine-mapping. QTG-seq combines QTL partitioning to convert a quantitative trait into a near-qualitative trait, sequencing of bulked segregant pools from a large segregating population, and the use of a robust new algorithm for identifying candidate genes. Using QTG-seq, we fine-mapped a plant-height QTL in maize (*Zea mays* L.), *qPH7*, to a 300-kb genomic interval and verified that a gene encoding an NF-YC transcription factor was the functional gene. Functional analysis suggested that *qPH7*-encoding protein might influence plant height by interacting with a CO-like protein and an AP2 domain-containing protein. Selection footprint analysis indicated that *qPH7* was subject to strong selection during maize improvement. In summary, QTG-seq provides an efficient method for QTL fine-mapping in the era of "big data".

Key words: quantitative trait locus, QTL, QTL fine-mapping, whole genome sequencing, plant height

Zhang H., Wang X., Pan Q., Li P., Liu Y., Lu X., Zhong W., Li M., Han L., Li J., Wang P., Li D., Liu Y., Li Q., Yang F., Zhang Y.-M., Wang G., and Li L. (2019). QTG-Seq Accelerates QTL Fine Mapping through QTL Partitioning and Whole-Genome Sequencing of Bulked Segregant Samples. Mol. Plant. **12**, 426–437.

INTRODUCTION

Agronomic traits are directly targeted for improvement in crop breeding, and this process is crucial for global food security. Modern crop improvement relies on the dissection of the genetic and molecular mechanisms underlying agronomic traits. Most agronomic traits vary quantitatively and are subject to complex genetic regulation. Isolating the underlying quantitative trait genes (QTGs) has proved to be vital for crop improvement (Jiang et al., 2012). Map-based cloning is an effective way to isolate QTGs (Yan et al., 2003; Bortiri et al., 2006). However, this method usually requires several generations of backcrossing to eliminate the influence of the genetic background from the donor parent, screening of a large population for recombinants, and exhaustive field phenotyping (Zuo et al., 2015), which together make map-based cloning time consuming and costly. Although association mapping has been shown to be an efficient way to decipher the genetic mechanisms underlying agronomic traits, the ability of this method to detect rare alleles is inherently limited, and assembling a proper panel with a reasonable population size and representative diversity is expensive (Asimit and Zeggini, 2010; Huang et al., 2010).

An alternative method of genetic mapping, bulked segregant analysis (BSA), has been used to dissect the genetic basis of both qualitative and quantitative traits (Zou et al., 2016). The

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and IPPE, SIBS, CAS.

BSA strategy has been revolutionized by the big data made available through the advent and rapid development of high-throughput sequencing technology and shows great potential for high-resolution mapping (Schneeberger et al., 2009; Liu et al., 2012). The invention of methods such as MutMap, SHOREmap, and MMAPPR has demonstrated that combining BSA with high-throughput sequencing is an attractive alternative to traditional positional cloning for qualitative traits (Schneeberger et al., 2009; Abe et al., 2012; Hill et al., 2013). For quantitative traits, quantitative trait locus sequencing (QTL-seq) was devised to map QTLs, but the mapping resolution is too low to identify candidate genes, especially in species with a large genome size (Takagi et al., 2013). Therefore, a fine-mapping strategy for rapid identification of candidate genes underlying quantitative traits is still needed.

Maize, a widely studied genetic model, is one of the most widely grown crops in the world. During the domestication of maize from ancient teosinte around 9000 years ago (Matsuoka et al., 2002), a number of key traits, including flowering time, changed substantially, enabling the plants to adapt to different environments and variable climates. Along with flowering time, plant height, one of the most important targets for improvement through crop breeding, is quite variable as a consequence of this adaptive radiation. Although several flowering time and plant height loci have been cloned and shown to be subject to strong selection (Yang et al., 2013; Wills et al., 2018), the underlying genetic mechanisms are largely unknown.

To accelerate QTL fine-mapping, QTG sequencing (QTG-seq) was developed and used to map a QTL for plant height. Along with Euclidean distance (ED) and G′, a new statistic, smoothLOD, was used to delimit the candidate QTL to a small interval. The candidate gene was selected with the assistance of RNA sequencing (RNA-seq) analysis, and was validated using CRISPR. Selection footprint analysis also demonstrated that the candidate gene was a target of selection during maize improvement. Our method provides an efficient means of rapid QTL cloning, and will accelerate functional analysis of QTGs and development of molecular breeding tools.

RESULTS

Rationale Behind QTG-Seq

The QTG-seq strategy consists of the following steps (Figure 1): (1) F_1, F_2, and BC_1F_1 populations are developed using two inbred lines. (2) QTL position is determined using the F_2 population and confirmed using $F_{2:3}$ families if the researchers are not confident about the QTL mapping results. (3) QTL partitioning, in which BC_1F_1 plants heterozygous for the QTL of interest but homozygous for the other QTLs are selected, is performed using molecular markers tightly linked to each QTL. The desired BC_1F_1 plants are then self-pollinated to produce $BC_1F_{1:2}$ families. It is important to note that multiple BC_1F_1 individuals can and should be selected for a particular QTL to ensure a large population. (4) The derivative $BC_1F_{1:2}$ families are evaluated for the given trait, and the bottom and top 20% of plants showing extreme phenotypes are selected separately within each family (at least 1000 individuals in total for both phenotypic extremes). (5) DNA from the two groups of selected plants is extracted and bulked in equal amounts, forming two pools described as the "high pool" and "low pool". (6) The two large DNA pools are subjected to whole-genome sequencing with deep genome coverage. The sequence data are mapped to the reference genome, and genomic variants across the whole genome are identified. (7) Allele frequency is calculated for each pool. A new statistic named smoothLOD, the smoothed version of the LOD score, is introduced to accurately determine the position of the target locus. Meanwhile, other statistics, such as ED (Hill et al., 2013) and G′ (Magwene et al., 2011), are also calculated to determine the peak position of the target locus.

QTG-Seq Rapidly Fine-Mapped a Maize Plant-Height QTL, *qPH7*

As a proof of concept, we performed QTG-seq using plant height in maize as a model trait. We started with primary QTL mapping to demonstrate the utility of QTG-seq and were able to fine-map a plant-height QTL in four generations, which is in contrast to the many generations needed to construct an advanced backcross population for conventional QTL fine-mapping (Xing et al., 2015; Zuo et al., 2015). First, we constructed a segregating population from two maize inbred lines, HZS and 1462, that showed substantial differences in plant height (Supplemental Figure 1A). Plant height in the F_2 population ranged from 199 to 307 cm, with the heights of the two parents lying between these values. This transgressive segregation implied that plant height was controlled by multiple QTLs (Supplemental Figure 1B and 1C). Next, we performed primary QTL mapping using 1028 polymorphic markers. Using a LOD threshold of 3.7 we detected four QTLs located on chromosomes 1, 3, 6, and 7 that explained 11.22%, 7.49%, 9.39%, and 16.60% of the total phenotypic variance, respectively (Supplemental Figure 1C and 1D; Supplemental Table 1). To confirm the QTL positions, we evaluated the derivative $F_{2:3}$ families for plant height. The correlation coefficient between the heights of the F_2 and $F_{2:3}$ plants was 0.72, and the four QTLs detected in the F_2 population were also the top four QTLs with the largest effects in the $F_{2:3}$ population (Supplemental Figure 1E), confirming the reliability of the QTL mapping results. We chose the QTL with the highest LOD score (11.99 in the F_2 population), on chromosome 7 (hereafter referred to as *qPH7*), for QTL fine-mapping using the QTG-seq strategy. This locus showed a partial dominant effect, and the allele from 1462 increased plant height (Supplemental Table 1). This locus was also observed to be associated with flowering time (Supplemental Figure 2).

Screening of 813 BC_1F_1 individuals with 12 markers identified lines heterozygous at *qPH7* and homozygous at the other three detected QTLs. We selected 15 BC_1F_1 plants and self-pollinated them to obtain $BC_1F_{1:2}$ families, which we planted and evaluated for plant height. We measured the heights of 3120 $BC_1F_{1:2}$ plants from the 15 families, each containing 74–289 plants, and defined the two phenotypic extreme groups, "low" (short plants) and "high" (tall plants). There were 580 low plants and 567 high plants (Supplemental Table 2), and these plants were used to form the low and high pools. We then subjected the two pools to whole-genome sequencing up to >280× coverage, which identified 197 021 high-quality single-nucleotide polymorphisms (SNPs). Plots of smoothLOD (from a new maximum-likelihood method that we introduced to detect

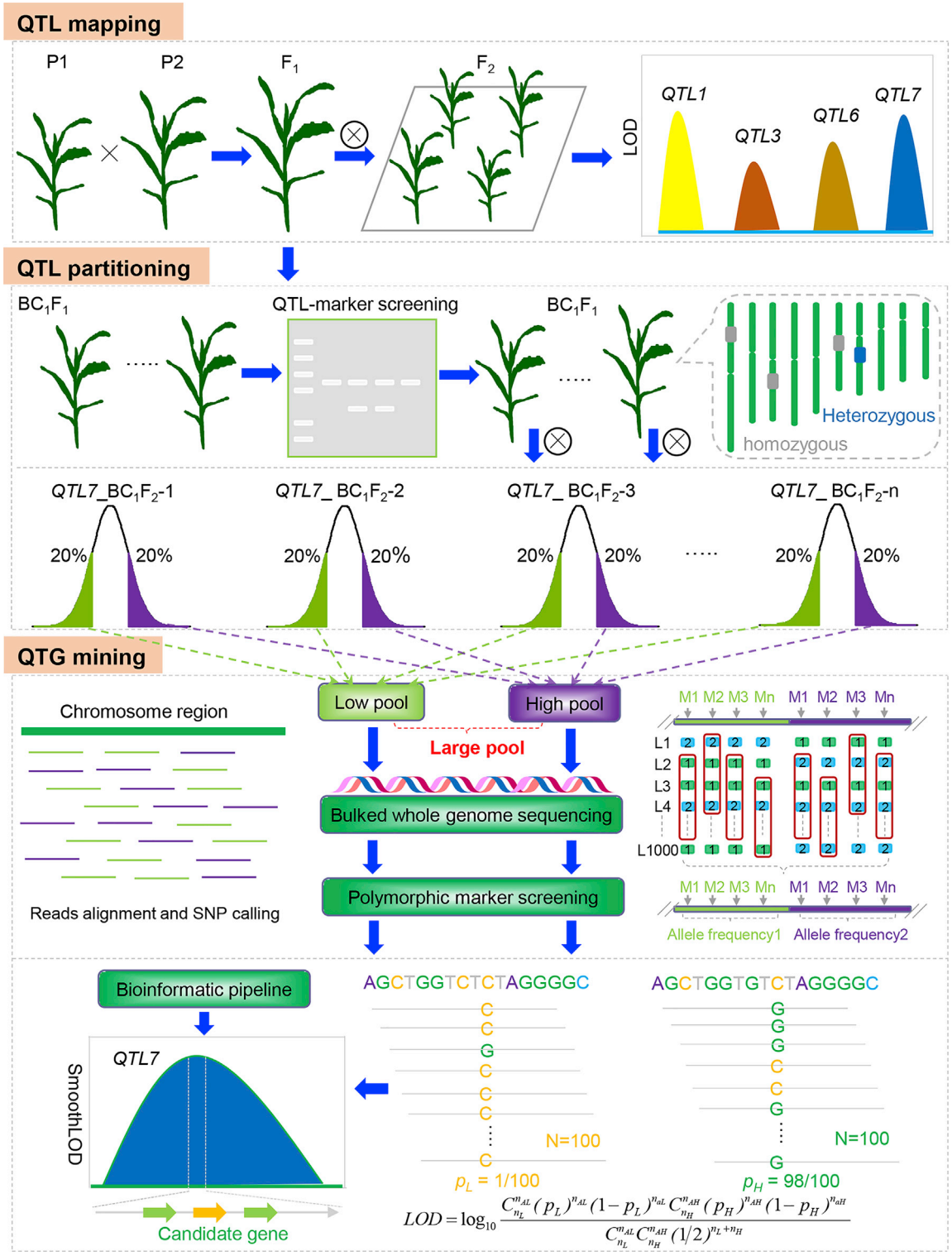

Figure 1. Rationale Behind QTG-Seq.

QTG-seq consists of three major steps: QTL mapping; QTL partitioning, in which a large number of segregating BC_1F_2 populations (>1000 individuals in total) are constructed from self-crosses of multiple BC_1F_1 individuals that are heterozygous for the target QTL and homozygous for the other QTLs; and QTG mining, which consists of high-throughput sequencing of bulked samples from the "high" and "low" pools derived from multiple segregating BC_1F_2 populations and a bioinformatic pipeline for rapid QTG fine-mapping. Specifically, the relatively deep high-throughput sequencing coverage compared with the large number of individuals in the bulked samples, in conjunction with the existence of recombination blocks, allows the frequencies of *A* and *a* alleles to be accurately quantified based on random sampling during the high-throughput sequencing process. SNP, single-nucleotide polymorphism.

the fine-mapping signal) and ED^4 statistics demonstrated that there was only one peak on chromosome 7 (Figure 2A–2C; Supplemental Figure 3A and 3B). In addition, a close-up view of the graph of the ED^4 statistic on chromosome 7 confirmed the peak position at 135.3 Mb (Figure 2C). KASP genotyping also confirmed that the peak position was between 135.1 and 135.2 Mb (Figure 2D). Intriguingly, using the new smoothLOD method we were able to fine-map the QTG of *qPH7* to around 135 216 475 Mb on chromosome 7 (Figure 2A), which fell right into the genic region for *Zm00001d020874*. *Zm00001d020874* encodes an NF-YC-domain containing protein, whose *Arabidopsis thaliana* homologs (Supplemental Table 3) have been reported to influence flowering time (Kumimoto et al., 2010).

Given the existence of recombination blocks in which all the markers segregate in an identical manner, we evaluated mapping signals in a sliding window of increasing size as a means to identify recombination blocks in the target region. Because the smoothLOD and G' statistics are continuous variables, they are not reliable for evaluating recombination blocks. Instead, we plotted ED per SNP (ED/SNP) around the position of peak ED while gradually increasing the window size by 10 kb at a time. ED/SNP reached a steady level below 150 kb, suggesting that the distance between the candidate gene and the peak position is <150 kb (Supplemental Figure 4A and 4B). Therefore, we expected the *qPH7* candidate gene to be located in the 300-kb interval surrounding the peak position, which contains 13 genes (Figure 2D), including the *Zm00001d020874* gene identified by smoothLOD.

A large pool size is crucial for guaranteeing high mapping resolution, because the number of recombinants and the mapping resolution of the target interval will increase with increasing pool size. However, in contrast to conventional methods in which saturated sequencing (the sequencing coverage is nearly equal to pool size) is conducted on each pool (Abe et al., 2012; Takagi et al., 2015), relatively low sequencing coverage (less than 100× coverage) used by QTG-seq was sufficient to detect 100% of the recombination blocks according to simulation of our real data (Supplemental Figure 5). The existence of recombination blocks might be the reason for the high detection resolution with a relatively low sequencing coverage; because each recombination block contained a set of SNPs, which should be detected in many sequencing reads, there was no need to sequence all reads for each recombination block. Therefore, a combination of large pool size and relatively low sequencing coverage is suggested when using the QTG-seq method to balance between high mapping resolution with a large pool size and the relatively low cost of sequencing.

Zm00001d020874 Was Demonstrated to be the Quantitative Trait Gene Underlying *qPH7*

We performed RNA-seq to assess whether there were significant differences in the expression of any of the candidate genes in the 300-kb interval between the two parents. Among the 13 genes (Supplemental Table 4), five were excluded owing to their low expression in both parents, and one was excluded because there were no significant differences in gene expression between the two parents and no non-synonymous sequence variants (Figure 2D and 2E). By contrast, we observed significant differences in *Zm00001d020874* expression between the two parents in both the shoot apical meristem (SAM) and juvenile internode tissues (Figure 3A). Based on the known function of its *Arabidopsis* homolog in controlling flowering time and the strong fine-mapping signal identified by the smoothLOD algorithm, we considered *Zm00001d020874* to be a strong candidate for *qPH7*.

To verify the function of *Zm00001d020874* at locus *qPH7*, we constructed an RNA-guided CRISPR-Cas9 expression vector targeting two regions in the *Zm00001d020874* exon (Supplemental Figure 6A) and used it to transform the wild-type maize inbred line CAL1. We identified two transformation-positive T_0 plants out of 20 T_0 plants, which we named T13 and T16 (Supplemental Figure 6B). The two PCR bands amplified from T13 DNA each contained deletions and were named the "a1" and "a2" alleles (Supplemental Figure 6B and 6C). The a1 allele contained a 1-bp deletion and an 8-bp deletion that caused a frameshift mutation, while the a2 allele contained a larger deletion of 217 bp. In the self-crossing progenies of T13, we identified six "a1a1" and "a2a2" homozygous plants. Plant heights of a transgene-negative line carrying unmodified *Zm00001d020874* and the two T13 homozygous transgenic lines (named T13_a1a1 and T13_a2a2) were compared, and *t*-test analysis revealed that the mean plant height of each of the two mutants was significantly lower than that of the transgene-negative line (Figure 3D and 3E). T16 was homozygous for the same 217-bp deletion found in the T13 a2 allele (Supplemental Figure 6C). The mean plant height of 10 T16 T_1 plants was also significantly lower than that of the transgene-negative line (Figure 3D and 3E). Furthermore, the total leaf number and flowering time of T13 and T16 progeny were significantly different from those of the transgene-negative line (Supplemental Figure 7). All of the transgenic CRISPR editing lines harbored mutations causing large amino acid changes representing loss-of-function of the target gene (Supplemental Figure 6C). These results support the hypothesis that *Zm00001d020874* is the gene underlying *qPH7*.

Returning to the segregating population we had constructed from the maize inbred lines HZS and 1462, we identified six differences in the nucleotide sequence of *Zm00001d020874* between the two parental strains: five synonymous SNPs within exons and a 6-bp insertion and deletion (indel) sequence (CTTCCT) in the 5′ UTR (Figure 2F). To investigate whether inbred lines with alleles differing in that 6-bp sequence had different expression levels, we conducted an RNA-seq analysis using 181 inbred lines from a Chinese association mapping panel (Yang et al., 2010). Twelve lines lacked the 6-bp sequence, 110 contained the 6-bp sequence (the same genotype as 1462), and 10 had two tandem repeats of the sequence (the same genotype as HZS); the remaining lines either did not yield PCR products or failed to produce phenotypic data. Analysis of variance revealed significant differences in the expression of *Zm00001d020874* among the three genotypes (Supplemental Figure 8). Correlation analysis showed a significant correlation between plant height and *Zm00001d020874* expression (Figure 3B), and a similar phenomenon was observed for other traits, including ear height and the developmental timing parameters pollen-shedding day, silking day, and heading day (Supplemental Figure 9). These

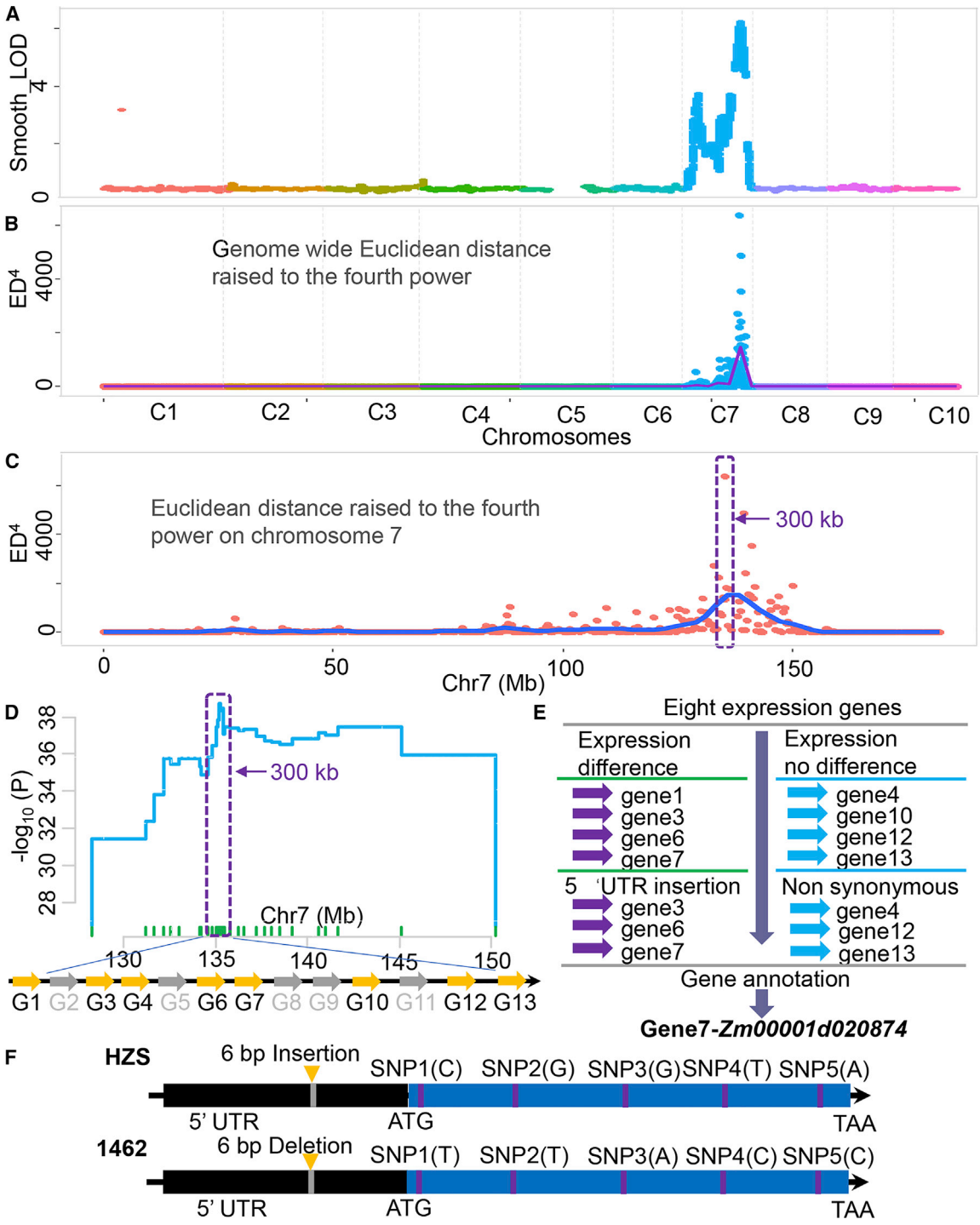

Figure 2. Fine-Mapping of *qPH7* and Candidate-Gene Mining.

(A) SmoothLOD profiling across the 10 maize chromosomes.

(B) Distribution of ED^4 across the 10 chromosomes.

(C) Close-up view of the distribution of ED^4 on chromosome 7.

(D) KASP genotyping confirmed the candidate interval and showed that it contains 13 genes; the genes in gray were excluded from consideration owing to their low expression.

(E) Gene 7 was selected as the candidate gene owing to its expression differences between the two parents and its known annotation.

(F) Comparison of the DNA sequence of *Zm00001d020874* between parental strains HZS and 1462; the 6-bp indel sequence is CTTCCT.

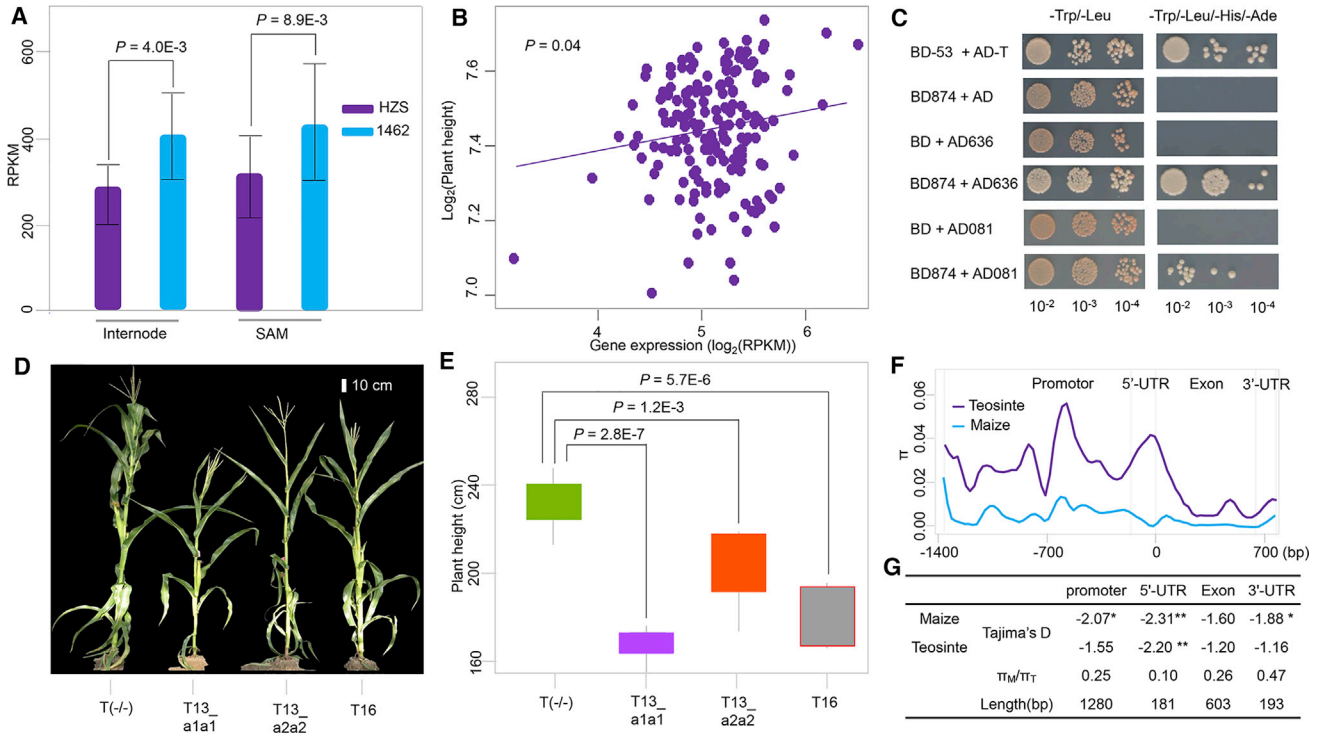

Figure 3. Verification of the Candidate Gene.

(A) Comparison of *Zm00001d020874* expression between HZS and 1462 in the juvenile internode and shoot apical meristem (SAM); error bars indicate SD, *P* values were calculated from *t*-tests.

(B) Correlation between *Zm00001d020874* expression and plant height was calculated using the expression data and plant heights of 170 inbred lines; *P* value was calculated from Pearson's correlation test.

(C) Yeast two-hybrid assay showing that the *Zm00001d020874* protein product interacts with the products of the CONSTANS (CO)-like gene *Zm00001d045636* (AD636) and the AP2 domain-containing gene *Zm00001d018081* (AD081). BD874 indicates the vector constructed by cloning the *Zm00001d020874* coding sequence into the bait vector pGBKT7. AD636 and AD081 indicate the vectors constructed by cloning the coding sequences of *Zm00001d045636* and *Zm00001d018081*, respectively, into the prey vector pGADT7. AD and BD indicate the prey vector pGADT7 and the bait vector pGBKT7, respectively.

(D) Examples of transgene-negative T(–/–) and homozygous transgenic T13 (a1a1 and a2a2) and T16 plants.

(E) Significant variation in plant height between a transgene-negative line and the homozygous T_1 transgenic lines. *P* values were calculated from *t*-tests.

(F) Nucleotide diversity across *Zm00001d020874* among 125 maize lines and 25 teosinte entries; nucleotide diversity (π) for teosinte (violet) and maize (cyan) was calculated using a 100-bp sliding window with a 25-bp step.

(G) The results from Tajima's *D* test and the π ratio, maize (π_M) to teosinte (π_T), are shown. $^*P < 0.05$; $^{**}P < 0.01$.

results indicate that the *Zm00001d020874* gene's influence on plant height might be related to reproductive development. This speculation was supported by our data indicating that the *Zm00001d020874* protein product interacted with the products of two putative maize flowering-time genes, *Zm00001d045636* (a *CO*-like gene) and *Zm00001d018081* (an AP2-domain-containing gene) (Figure 3C) (Yant et al., 2010; Nguyen et al., 2015).

Selection Footprint Analysis Revealed that *Zm00001d020874* Was Subject to Selection during Maize Improvement

To investigate the evolution of the *qPH7* locus, which controls both plant height and flowering time, we sequenced four regions (promoter, 5′ UTR, exon, and 3′ UTR) of the *Zm00001d020874* genes from 25 teosinte entries and 125 maize inbred lines. Dramatic evolutionary changes were identified across the entire *Zm00001d020874* gene. In the coding region and 3′ UTR, the diversity in both maize and teosinte is fairly low, suggesting that *Zm00001d020874* is evolving under functional constraint (Figure 3F). Conversely, significantly higher diversity was observed for teosinte than for maize in the promoter and 5′ UTR of *qPH7* (Figure 3F). Together with the results of a Tajima's *D* test, this indicates a possible selective sweep across *qPH7* in maize (Figure 3G). Moreover, this selective sweep spanned from a site 1.4 kb upstream of the coding start site through the exon and the 3′ UTR in maize, but spanned only the exon and 3′ UTR in teosinte (Figure 3F). High sequence diversity was also observed among teosinte and maize landraces using data from maize HapMap v.3 (Bukowski et al., 2017), suggesting that the selection in the promoter region might have occurred during maize improvement (Supplemental Figure 10; Supplemental Table 5).

Monte Carlo Simulation Studies Support the Robustness of QTG-Seq

To validate our maximum-likelihood method for QTG-seq, we performed four Monte Carlo simulation experiments. A small absolute bias for the estimated QTG position indicates the high

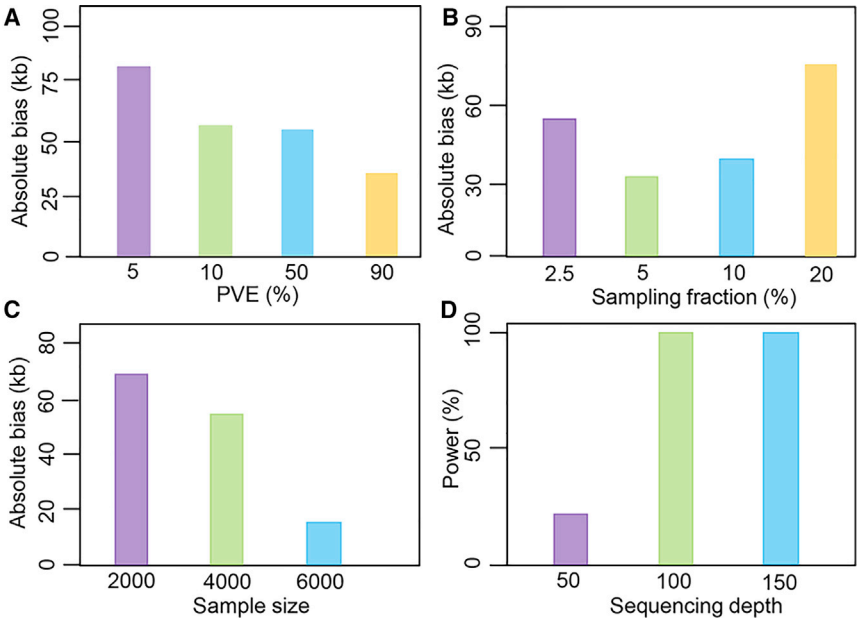

Figure 4. Influence of Various Factors on Mapping Accuracy Using SmoothLOD.
Monte Carlo simulations were conducted to evaluate the influence of PVE (phenotypic variance explained by each QTL) **(A)**, sampling fraction **(B)**, and F_2 sample size **(C)** on absolute bias determined using smoothLOD for QTG-seq. The detection power **(D)** for different sequencing depths was also calculated as described in Methods.

accuracy of the new algorithm (Figure 4). The absolute bias for the estimated QTG position was approximately 80 kb for the QTG with percentage of phenotypic variation explained (PVE) = 5% and 30–60 kb for the QTG with PVE $\geq$10% (Figure 4A). The absolute bias was also 30–60 kb when the sampling fraction was between 2.5% and 10% (Figure 4B). Although the absolute bias was slightly larger when the sampling fraction was 20%, a larger sampling fraction would increase population size and the number of effective recombinant events. The absolute bias was small (<60 kb) when the population size was larger than 4000 (Figure 4C). The statistical power for QTG detection was 100% when the sequencing depth was >100× for bulked samples from over 1000 individuals (Figure 4D). Compared with existing statistics, our maximum-likelihood method showed higher accuracy in fine-mapping QTGs (Supplemental Table 6). Therefore, the new method is reliable for QTG fine-mapping.

DISCUSSION

In this study, we devised a new strategy that we named QTG-seq (https://github.com/caulilin/QTG_Seq) and successfully implemented this strategy to clone the QTL *qPH7* more rapidly than could be done using a conventional map-based QTL cloning strategy (Xing et al., 2015). Our QTG-seq approach started with a cross of two maize inbred lines, HZS and 1462, followed by the use of an F_2 population to map QTLs underlying plant height variation. Then, we conducted QTL marker screening on BC_1F_1 populations to obtain BC_1F_1 individuals, which were further self-pollinated to construct $BC_1F_{1:2}$ families segregating only at the *qPH7* locus. In each $BC_1F_{1:2}$ family, we sampled 20% of the individuals with either extremely high or extremely low plant height, and bulked all the high and low samples for whole-genome sequencing with 280× coverage. We then implemented a series of bioinformatic analyses to fine-map *qPH7* to an ~300-kb genomic region with a mapping resolution of 150 kb. The functional gene *qPH7* was confirmed genetically and at the molecular level, and was shown to have been subject to strong selection during maize improvement. QTG-seq requires only four generations starting from the original cross of two parents for QTL fine-mapping and cloning. Thus, QTG-seq is a time- and cost-saving strategy for dissecting the genetic and molecular mechanisms underlying quantitative traits.

The CRISPR-Cas (CRISPR-associated) system has emerged as a powerful tool for functional gene validation in plants. Although it has been reported that a chance of off-target genome editing might exist, the possibility is pretty low even if the genome has off-target sequences highly homologous to the target sites (Feng et al., 2014; Tang et al., 2018). To avoid the problem of off-target deleterious mutations, we aligned the sequence of the target locus to the maize genome and found that the sequence did not significantly align with any other genomic regions in maize, indicative of a relatively low possibility of off-target editing. Moreover, our observation of T_1 plants showed that the targeted mutation co-segregated with phenotypic variation (Figure 3E, Supplemental Figure 7). Therefore, the function validation supports the hypothesis that *Zm00001d020874* is the functional gene *qPH7*.

Several methods for fine-mapping of genetic loci have been reported (Supplemental Table 7). However, QTG-seq has several notable benefits compared with other methods. First, QTG-seq is suitable for minor-effect QTLs, because when other QTLs are excluded through QTL partitioning, the effect of the target QTL increases (Zhang et al., 2012). This conclusion was also supported by our simulation analysis with real data (Figure 4A). Second, QTG-seq allows efficient QTG mining. Among the methods summarized in Supplemental Table 7, only two have been used to analyze quantitative traits. Whereas the mapping resolution of QTL-seq was several megabases (Takagi et al., 2013), that of QTG-seq can reach the gene level. Third, the large pool size and high sequencing coverage guarantee the high mapping resolution of QTG-seq. With a large pool size, a larger number of informative recombinants will be included in the pools, which is useful for narrowing down the target interval. With a high sequencing coverage, the number of SNPs identified should reach the maximum possible, thus SNP density will not be the limiting factor for mapping resolution. The most evident advantage is that QTG-seq requires only four generations from the first cross of any parent lines, a speed comparable to that for isolating a gene controlling a qualitative trait. Thus, in comparison with other methods, QTG-seq is a more powerful method for accelerating QTL fine-mapping.

To evaluate the best choice for genotyping using QTG-seq, we compared different techniques, including genome sequencing, exome capture sequencing, and KASP marker genotyping. In our study, genome sequencing was the best choice given the considerable size of the maize genome (similar to that of humans) and dramatic genomic variations because exome capture is biased towards certain sequences (King et al., 2015) and the number of KASP markers is limited. In other species, the situation may be different: for example, wheat (*Triticum aestivum)* and barley (*Hordeum vulgare*) have large genomes, and in this scenario exome capture may be the better choice, because it reduces sequencing costs without sacrificing accuracy in the quantification of allele frequency (Mascher et al., 2014). In rice (*Oryza sativa*), which has a relatively small genome, genome sequencing could be more affordable and allow robust fine-mapping of QTLs. Thus, different species with various genomic features may be best served using different genotyping techniques in combination with QTG-seq.

To improve the efficiency and optimize the parameters of QTG-seq, two aspects are particularly important: factors related to the BSA strategy and the characteristics of the QTL. In regard to the first aspect, pool size is the major factor to consider, starting with two questions: whether a larger pool size would increase mapping resolution, and whether the DNA of all the plants bulked into each pool could be sequenced. Our simulation studies indicated that mapping accuracy generally increased with increasing pool size (Figure 4B and 4C), which may be related to the fact that the number of recombinants increased with the increase in the size of the mapping population. When the pool size is large, it is preferable to increase sequencing coverage to capture as many recombinants as possible. However, rather than sequencing to obtain a coverage equal to the number of plants in the bulk pool multiplied by the genome size, a simulation using real QTG-seq data showed that $\sim$100× coverage should be sufficient for the identification of QTGs when bulking over 1000 individuals (Figure 4D; Supplemental Figure 5). As for the second aspect, first the influence of donor parent background should be effectively controlled so as to increase the phenotypic variation explained by the target QTL in the tested population (Yamamoto et al., 1998; Zhang et al., 2012). Second, the environment in which the population was evaluated should be carefully controlled to decrease unexpected phenotype variation (Lu et al., 2011). Third, it is preferable to select QTLs with a large additive effect to increase the genetic variance of the population (Xing et al., 2015). We believe that the efficiency of QTG-seq can be maximized by optimizing the factors mentioned above.

QTL fine-mapping is a complex process for cloning functional QTGs that is affected by many genetic factors, such as QTL effect size, population size, and recombination frequency. In our study, simulation analysis suggested that the mapping resolution of QTG-seq is influenced by QTL effect size, sampling fraction, sampling size, and sequencing depth. Consistent with our findings, it was reported that the mapping power increased with increasing QTL effect and population size, and was influenced by the sampling fraction (Sun et al., 2010). It was also found that sequencing depth influenced the density distribution of the SNP index at the causal SNP (Abe et al., 2012). However, the sequencing depth required could be substantially decreased due to the existence of recombination blocks. Recombination frequency is the key factor for functional gene cloning. Therefore, the mapping resolution should also be dependent on the chromosome region because the recombination frequency varies among different chromosome regions (Farkhari et al., 2011; Pan et al., 2016). In addition, genome complexity influences the prediction of candidate genes and the adoption of molecular techniques for genotyping (Sun et al., 2018). Collectively, the QTG-seq strategy should be carefully examined and improved to adapt to various influencing factors.

Although the concept of QTG-seq was demonstrated using a maize QTL, QTG-seq could be extended to other species. For instance, QTG-seq should be suitable for the model species rice and *Arabidopsis*. Because of their small genome size, long-range linkage disequilibrium, and moderate genome complexity, it might be easier to identify candidate genes in these species. QTG-seq is potentially applicable to most species for which no reference genome exists. In this case, it would be necessary to use all the next-generation sequencing data for *de novo* genome assembly, then use the assembled fragmented reference genome for genomic variant calling, and finally identify all the contigs completely linked to phenotypic variation. In addition, QTG-seq can potentially be used to assemble *de novo* functional genes that are not present in the reference genome. Various biparental and multiparental population types are suitable for conducting QTG-seq if the genetic background is properly controlled, including F_n, BC_n, and BC_nF_m populations (where n and m are integers ≥ 2), which can be screened from the by-product materials used to develop recombinant inbred lines, backcross inbred lines, residual heterozygous lines, nested association mapping populations, and multiparent advanced generation intercross populations. Thus, QTG-seq has great potential to be adapted to various plant species and populations.

METHODS

Plant Materials

The recurrent parent (HZS) used to construct the segregating population was an elite inbred line commonly used in Chinese maize hybrid development (Li, 1998), and the donor parent (1462) was a temperate inbred line; the two parents were chosen from an association mapping panel (Yang et al., 2014). HZS was crossed with 1462 in the winter of 2015 in Yuanjiang, Yunnan, China. In the spring of 2016, again in Yuanjiang, the F_1 plants were self-pollinated and backcrossed to HZS to produce F_2 and BC_1F_1 seeds, respectively. The F_2 population was sown in the summer of 2016 in Shunyi, Beijing, China. Sixty-five days after planting, plant height was measured from ground to tassel tip for each of 200 F_2 plants, which were self-pollinated to produce 200 $F_{2:3}$ families. In the winter of 2016, the 200 $F_{2:3}$ families were sown in two replicates in Sanya, Hainan, China. For each replicate, each family was sown in a one-row plot. Plant height was evaluated for the $F_{2:3}$ families. BC_1F_1 seeds were sown in the winter of 2016 in Sanya and thinned to 813 plants, and all of the BC_1F_1 plants were self-pollinated to produce 813 $BC_1F_{1:2}$ families. For all field trials, the row length and row space were 5 m and 60 cm, respectively. The space between plants in the same row was 25 cm.

Genotyping and QTL Mapping

DNA was extracted from fresh leaves from the 200 F_2 plants according to a published protocol with minor modifications (Murray and Thompson, 1980). DNA was sent to Beijing Compass Biotechnology for genotype characterization using 6K DNA chips. The chip contains 5179 SNPs evenly distributed across the whole genome (Supplemental Table 8).

Linkage analysis was conducted using QTL IciMapping version 4.0; the linkage groups were separated with the LOD score set at 3.0, and ordered using the nnTwoOpt algorithm (Meng et al., 2015). The genotype and phenotype data from the F_2 population were combined and used to perform QTL analysis using Windows QTL Cartographer 2.5 (Silva Lda et al., 2012). The LOD threshold for declaring a QTL was determined by conducting 1000 permutations. Composite interval mapping was employed with default settings to compute QTL information, including LOD score, PVE, and additive (a) and dominance (d) effects (Zeng, 1994). The absolute value of the ratio of dominance effect over additive effect (|d/a|) was used to measure the degree of dominance as described in a previous report (Stuber et al., 1987), where the genetic effect of the QTL was classified into additive ($0 < |d/a| \leq 0.2$), partial dominance ($0.2 < |d/a| \leq 0.8$), dominance ($0.8 < |d/a| \leq 1.2$), and overdominance ($|d/a| > 1.2$) based on the degree of dominance value (Stuber et al., 1987). To confirm the QTL mapping results obtained using the F_2 population, QTL mapping was also performed using the phenotype data for the 200 $F_{2:3}$ families.

QTL Partitioning and Selection of Plants Showing Extreme Height

Twelve indel markers (Supplemental Table 9) were developed based on a 1-LOD supporting interval for QTLs detected using the F_2 population (Settles et al., 2014). BC_1F_1 plants heterozygous at *qPH7* and homozygous at the other three QTLs were selected from the 813 BC_1F_1 plants. In the summer of 2017, 3120 plants belonging to 15 selected $BC_1F_{1:2}$ families were evaluated for plant height in Shunyi, Beijing, China. The 20% tallest (580) and 20% shortest (567) plants were selected from each $BC_1F_{1:2}$ family (Supplemental Table 2). The DNA of the 580 tallest and 567 shortest plants was bulked in equal ratios into two pools, which are termed the high pool and low pool, respectively.

Genome Sequencing, Read Mapping, and SNP Calling

Genome sequencing was performed using Illumina HiSeq Xten, generating a total of 700 Gb of sequence data from each pool. The sequence data containing 150-bp paired-end reads and 8-bp indexed reads were aligned to the maize B73 version 4 reference genome (AGP v.4) using BWA software (Jiao et al., 2017; Li and Durbin, 2009). SAMtools was used to convert sam files to bam files (Li et al., 2009). Picard software was employed to fix mates, sort read groups, and remove PCR duplicates from the bam files (http://broadinstitute.github.io/picard/). Sequence variants including SNPs and indels were called using HaplotypeCaller from GATK (McKenna et al., 2010; Jiao et al., 2017), generating a VCF file for both low and high pools. All the software and commands were run in a Linux server environment; the in-house Shell, R, and Perl scripts have been released and are available in GitHub (https://github.com/caulilin/QTG_Seq).

Calculation of Allele Frequency, ED, ED^4, and G′

Assuming that "A" is a SNP identified between HZS and 1462, the HZS and 1462 genotypes at this site are "aa" and "AA", respectively. For the low pool, n_{aL} and n_{AL} are the numbers of reads containing "a" and "A", respectively, and n_L is the sum of n_{aL} and n_{AL}. For the high pool, n_{aH} and n_{AH} are the numbers of reads containing "a" and "A", respectively, and n_H is the sum of n_{aH} and n_{AH} (Supplemental Table 10). For convenience, the allele frequencies of "a" in the low and high pools are described as p_{aL} and p_{aH}, respectively, and the allele frequencies of "A" in the low and high pools as p_{AL} and p_{AH}, respectively (Supplemental Table 10). Because lower read counts would lead to a bias in allele frequency, only sites with a minimum of 50 reads were retained.

First, the ED for each SNP, described as ED_{SNP}, was calculated according to a previous report (Hill et al., 2013):

$$ED_{SNP} = \sqrt{(p_{aL} - p_{aH})^2 + (p_{AL} - p_{AH})^2}$$

ED is the sum of 100 ED_{SNP} values within a window of 100 consecutive SNPs. ED^4 was then calculated by raising ED to the fourth power.

G′, the G value averaged across neighboring SNPs, was calculated as described by Magwene et al. (2011).

Maximum-Likelihood Method for Mapping QTGs

Let $p_L = n_{AL}/n_L$ and $p_H = n_{AH}/n_H$. To test for the absence of the QTG at the above locus, H_0:$p_L = p_H = 0.5$ and the LOD statistic is

$$LOD = \log_{10}\frac{C_{n_L}^{n_{AL}}(p_L)^{n_{AL}}(1-p_L)^{n_{aL}}C_{n_H}^{n_{AH}}(p_H)^{n_{AH}}(1-p_H)^{n_{aH}}}{C_{n_L}^{n_{AL}}C_{n_H}^{n_{AH}}(1/2)^{n_L+n_H}},$$

where $C_{n_L}^{n_{AL}} = n_L!/n_{AL}!n_{aL}!$ and $C_{n_H}^{n_{AH}} = n_H!/n_{AH}!n_{aH}!$.

To obtain a robust estimation of QTG position, a smoothed version of the standard *G* statistic proposed by Magwene et al. (2011) was used to obtain the smoothed LOD score statistic,

$$SmoothLOD = \sum_{j \text{ in } w} k_j LOD_j$$

where $k_j = (1 - D_j^3)^3/S_W$; D_j is standardized distance, with a value of 0 at the focal position and a value of 1 at the edge of the window; and S_W is the sum of $(1 - D_j^3)^3$ for all SNPs in *W* (Magwene et al., 2011). To calculate the probability of H_0:$p_L = p_H = 0.5$, an approximate χ^2 test was performed as shown below:

$$P(H_0) = \Pr\left(\chi^2_{\nu=df=2} > 4.605 \times SmoothLOD\right).$$

If $P(H_0) \leq 0.05$, one significant QTG exists.

KASP Genotyping and χ^2 Test

Thirty-one triplets of primers were designed for KASP genotyping around the peak position of ED (Khera et al., 2013). DNA from each of the 580 tallest (high) and 567 shortest (low) plants was used for PCR amplification. Bio-Rad CFXManager software was used to discriminate the genotypes of the PCR products. If the homozygous genotypes of HZS and 1462 for a given SNP were considered to be "aa" and "AA", the number of plants with the genotypes "aa" and "AA" was summarized to form a contingency table (Supplemental Table 10). The *P* value of the χ^2 test was calculated based the contingency table for each SNP marker separately.

RNA Extraction and RNA-Seq Analysis

Seeds of the two parents were sown in the autumn of 2017 in Haidian, Beijing, China. For two tissue types, SAM and juvenile internode, samples from six plants at the V6 stage were bulked to form one replicate. RNA was extracted from three replicates using an EasyPure Plant RNA Kit (TransGen Biotech). The 12 RNA samples were sequenced using the Illumina NovaSeq sequencing platform, generating 150-bp paired-end reads. For each sample, >10 Gb of data were produced for which the quality-control metric Q30 was >90%.

For RNA-seq data analysis, the following steps were used: (1) TopHat2 was used to align the sequencing reads to the maize B73 version 4 reference genome (Kim et al., 2013; Jiao et al., 2017); (2) Cufflinks was used to assemble the aligned reads into transcripts, which were deposited in a bam file (Trapnell et al., 2012); and (3) Cuffdiff was run to find the expression differences of all detected transcripts between HZS and 1462 (Trapnell et al., 2013).

Candidate-Gene Mining and Causal-Site Validation in a Natural Population

The 13 genes around the position of peak ED were screened as follows: (1) genes with a reads per kilobase per million mapped reads (RPKM) value of <1 in both parents were excluded; (2) genes with no significant differences between the two parents, and with no non-synonymous mutations identified between the two parents, were excluded; (3) the annotation information was used as a reference.

DNA from 181 inbred lines (Supplemental Table 11) was used for detection of the 6-bp indel in the promoter region of the candidate gene (*Zm00001d020874*); the forward and reverse primers used for amplification were TACCCGTTCCATCTCAAGTG and GGCAGCTAGGGTTGTG AGAC, respectively. The inbred lines were divided into three groups based on polymorphism for the 6-bp indel. Analysis of variance was conducted to detect whether the differences in candidate-gene expression among the three groups were significant.

One hundred seventy of the 181 inbred lines (Supplemental Table 11) were evaluated for plant height, ear height, and flowering-time traits in the summer of 2016 in Wuhan, Hubei, China. Leaf samples were harvested and subjected to RNA extraction. RNA-seq analysis was performed to generate the expression data for *Zm00001d020874*. Pearson's correlation coefficient between plant height and *Zm00001d020874* expression was calculated.

Yeast Two-Hybrid Assay

Yeast two-hybrid assays were carried out using the Matchmaker Two-Hybrid System Kit (TaKaRa Bio). The coding sequence of *Zm00001d020874* was cloned into the bait vector pGBKT7 (yielding a vector named BD874), and those of *Zm00001d045636* and *Zm00001d018081* were cloned into the prey vector pGADT7 (yielding vectors named AD636 and AD081, respectively). For the positive control, pGBKT7-53, encoding Gal4 BD fused with murine p53 (BD-53), and pGADT7-T, encoding Gal4 AD fused with the SV40 T-antigen (AD-T), were used. Because p53 and the large T-antigen interact with each other (Li and Fields, 1993), yeast cells containing both pGBKT7-53 and pGADT7-T can activate all reporter genes. Selection media without Trp and Leu enabled growth, and growth in the absence of Trp, Leu, Ade, and His required effective interaction.

Agrobacterium-Mediated Maize Transformation

Immature embryos from the maize inbred line CAL1 were transformed with *Agrobacterium tumefaciens* strain EHA105 containing the vector CPB-ZmUbi-hspCas9 following a previously described method (Zhang et al., 2013). The vector CPB-ZmUbi-hspCas9 was described by Li and colleagues (Li et al., 2017). Two guide RNA sequences were designed based on the exon sequence of *Zm00001d020874*; the target site of guide-19 in the exon was 220–242 bp, and the guide-19 sequence was GTGATTCTTGAAGTCGGTAGTGG. The target site of guide-3 in the exon was 437–459 bp, and the guide-3 sequence was GGAAGTCGAA TACCTCGGTGCGG. Twenty T_0 plants were obtained and transplanted to the field in Haidian, Beijing, in the spring of 2018. DNA extracted from fresh leaves of the 20 T_0 plants, and the Cas9-targeted sequences were amplified and sequenced to detect positive transgenic plants. The positive T_0 plants were self-pollinated to produce T_1 families, which were sown in Haidian in July 2018. The heights of the T_1 plants carrying homozygous edited sequences were evaluated and compared with those of CAL1.

Nucleotide Diversity and Tests for Neutrality

Zm00001d020874 was amplified from 25 teosinte entries and 125 maize inbred lines (Supplemental Table 12) using two primer pairs. The forward and reverse primer sequences of one primer pair were TACCC GTTCCATCTCAAGTGAAAGAG and CCAGGATTCTACCATCAAACAA CTCAA, respectively. The forward and reverse primer sequences of the other primer pair were GGGCGAAATGCTACGAATGAGGTTAT and GAGGAAGGTGCTTTTAGTGGGTGATT, respectively. PCR products were purified and sequenced by Shanghai Sangon Biotech. Nucleotide variants were identified using MEGA5 software (Tamura et al., 2011), and the output file was used to calculate nucleotide diversity (π) and Tajima's *D* using DnaSP version 6.10 (Rozas et al., 2017).

To further investigate whether the selection around *Zm00001d020874* occurred during maize domestication or improvement, the maize HapMap v.3 was downloaded from the link http://pan.baidu.com/s/1eRNGtxw (Bukowski et al., 2017). The maize lines consisted of 21 teosinte, 26 landrace, and 1163 other maize accessions; the variants around *Zm00001d020874* were identified and used to calculated nucleotide diversity (π) and Tajima's *D* as described above.

Monte Carlo Simulation Experiments

An F_2 population of 4000 individuals was simulated in the first simulation experiment (Wen et al., 2018). Each individual had one simulated chromosome with 20 000 markers. All interval lengths between adjacent markers were exactly the same as those on the seventh chromosome between 122.416752 and 151.993981 Mb in the real dataset. Here 1 Mb was approximately viewed as 1 cM. One additive QTG overlapping with the 10 137th marker was simulated, and the proportion of PVE by this QTG (r^2) was set to 5%, 10%, 50%, and 90%. The total average and residual variance were set at 100 and 10, respectively. The phenotypically smallest 100 individuals were defined as the low pool, while the phenotypically largest 100 individuals were defined as the high pool. The sequencing depth was 100×. The number of replicates was 1000. For each simulated QTG, samples in which $P < 0.05$ were counted. The ratio of the number of such samples to the total number of replicates represented the empirical power for this QTG. To measure the bias of the QTL position estimate, absolute bias was calculated.

In the second Monte Carlo simulation experiment, r^2 was fixed at 20%, and the sampling fractions were set at 2.5%, 5.0%, 10.0%, and 20.0%. The values of the other parameters were the same as those in the first experiment. In the third Monte Carlo simulation experiment, r^2 was fixed at 20% and the F_2 sample sizes were set at 2000, 4000, and 6000. The values of the other parameters were the same as those in the first experiment. In the last Monte Carlo simulation experiment, the number of replicates was 20 and the sequencing depths were set to 50×, 100×, and 150×. All reads were sampled from the real maize dataset.

Resource Distribution

Sequence data can be found in GenBank/EMBL under Bioproject no. PRJNA491393 or Study no. SRP178633.

Code Availability

All codes for the QTG-seq bioinformatic pipeline can be downloaded at GitHub (https://github.com/caulilin/QTG_Seq).

SUPPLEMENTAL INFORMATION

Supplemental Information is available at *Molecular Plant Online*.

AUTHOR CONTRIBUTIONS

L.L., G.W., Y.M.Z., H.Z., X.L., F.Y., and Q.L. discussed and designed the experiment; H.Z., X.W., Yu.L., J.L., W.Z., M.L., L.H., D.L., P.W., and Ya.L. performed the experiment; L.L., Y.M.Z., Q.P., and P.L. participated in the data analysis; L.L., Y.M.Z., H.Z., and X.W. drafted the manuscript. Specifically, the Li lab was in charge of the whole project, including initiation and design of the project, coordination, and bioinformatic pipeline and data analyses; the Wang lab was in charge of molecular validation and population construction; and the Zhang lab was in charge of algorithm development. All authors have read and approved the manuscript.

FUNDING

This research was supported by the National Key Research and Development Program of China (2016YFD0100404), the National Basic Research Program of China (2014CB138200), the National Natural Science Foundation of China (91735305, 31571268), the Fundamental Research Funds of the Central Non-profit Scientific Institution (Y2018LM04), the Xinjiang Key R&D Program (2018B01006-3), and the Huazhong Agricultural University Scientific & Technological Self-innovation Foundation (2662016PY096, 2014RC020). This research was also partly supported by the open funds of the National Key Laboratory of Crop Genetic Improvement.

ACKNOWLEDGMENTS

We thank Chunyi Zhang from the Chinese Academy of Agricultural Sciences for useful discussions on this project. The authors declare that they have no competing interests.

Received: November 20, 2018
Revised: December 20, 2018
Accepted: December 24, 2018
Published: December 27, 2018

REFERENCES

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., Matsumura, H., Yoshida, K., Mitsuoka, C., Tamiru, M., et al. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. Nat. Biotechnol. **30**:174–178.

Asimit, J., and Zeggini, E. (2010). Rare variant association analysis methods for complex traits. Annu. Rev. Genet. **44**:293–308.

Bortiri, E., Jackson, D., and Hake, S. (2006). Advances in maize genomics: the emergence of positional cloning. Curr. Opin. Plant Biol. **9**:164–171.

Bukowski, R., Guo, X.S., Lu, Y.L., Zou, C., He, B., Rong, Z.Q., Wang, B., Xu, D.W., Yang, B.C., Xie, C.X., et al. (2017). Construction of the third-generation *Zea mays* haplotype map. Gigascience **7**:1–12.

Farkhari, M., Lu, Y.L., Shah, T., Zhang, S.H., Naghavi, M.R., Rong, T.Z., and Xu, Y.B. (2011). Recombination frequency variation in maize as revealed by genomewide single-nucleotide polymorphisms. Plant Breed. **130**:533–539.

Feng, Z.Y., Mao, Y.F., Xu, N.F., Zhang, B.T., Wei, P.L., Yang, D.L., Wang, Z., Zhang, Z.J., Zheng, R., Yang, L., et al. (2014). Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in *Arabidopsis*. Proc. Natl. Acad. Sci. U S A **111**:4632–4637.

Hill, J.T., Demarest, B.L., Bisgrove, B.W., Gorsi, B., Su, Y.C., and Yost, H.J. (2013). MMAPPR: mutation mapping analysis pipeline for pooled RNA-seq. Genome Res. **23**:687–697.

Huang, X.H., Wei, X.H., Sang, T., Zhao, Q.A., Feng, Q., Zhao, Y., Li, C.Y., Zhu, C.R., Lu, T.T., Zhang, Z.W., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. Genet. **42**:961–U976.

Jiang, Y.H., Cai, Z.X., Xie, W.B., Long, T., Yu, H.H., and Zhang, Q.F. (2012). Rice functional genomics research: progress and implications for crop genetic improvement. Biotechnol. Adv. **30**:1059–1070.

Jiao, Y.P., Peluso, P., Shi, J.H., Liang, T., Stitzer, M.C., Wang, B., Campbell, M.S., Stein, J.C., Wei, X.H., Chin, C.S., et al. (2017). Improved maize reference genome with single-molecule technologies. Nature **546**:524–529.

Khera, P., Upadhyaya, H.D., Pandey, M.K., Roorkiwal, M., Sriswathi, M., Janila, P., Guo, Y.F., McKain, M.R., Nagy, E.D., Knapp, S.J., et al. (2013). Single nucleotide polymorphism-based genetic diversity in the reference set of peanut (*Arachis* spp.) by developing and applying cost-effective kompetitive allele specific polymerase chain reaction genotyping assays. Plant Genome https://doi.org/10.3835/plantgenome2013.06.0019.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. **14**:R36.

King, R., Bird, N., Ramirez-Gonzalez, R., Coghill, J.A., Patil, A., Hassani-Pak, K., Uauy, C., and Phillips, A.L. (2015). Mutation scanning in wheat by exon capture and next-generation sequencing. PLoS One **10**:e0137549.

Kumimoto, R.W., Zhang, Y., Siefers, N., and Holt, B.F. (2010). NF-YC3, NF-YC4 and NF-YC9 are required for CONSTANS-mediated, photoperiod-dependent flowering in *Arabidopsis thaliana*. Plant J. **63**:379–391.

Li, B., and Fields, S. (1993). Identification of mutations in p53 that affect its binding to SV40 large T antigen by using the yeast two-hybrid system. FASEB J. **7**:957–963.

Li, C., Liu, C., Qi, X., Wu, Y., Fei, X., Mao, L., Cheng, B., Li, X., and Xie, C. (2017). RNA-guided Cas9 as an in vivo desired-target mutator in maize. Plant Biotechnol. J. **15**:1566–1576.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25**:1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Proc, G.P.D. (2009). The sequence alignment/map format and SAMtools. Bioinformatics **25**:2078–2079.

Li, Y. (1998). Development and germplasm base of maize hybrids in China. Maydica **43**:259–269.

Liu, S.Z., Yeh, C.T., Tang, H.M., Nettleton, D., and Schnable, P.S. (2012). Gene mapping via bulked segregant RNA-seq (BSR-Seq). PLoS One **7**:e36406.

Lu, M., Xie, C.X., Li, X.H., Hao, Z.F., Li, M.S., Weng, J.F., Zhang, D.G., Bai, L., and Zhang, S.H. (2011). Mapping of quantitative trait loci for kernel row number in maize across seven environments. Mol. Breed. **28**:143–152.

Magwene, P.M., Willis, J.H., and Kelly, J.K. (2011). The statistics of bulk segregant analysis using next generation sequencing. PLoS Comput. Biol. **7**:e1002255.

Mascher, M., Jost, M., Kuon, J.E., Himmelbach, A., Assfalg, A., Beier, S., Scholz, U., Graner, A., and Stein, N. (2014). Mapping-by-sequencing accelerates forward genetics in barley. Genome Biol. **15**:R78.

Matsuoka, Y., Vigouroux, Y., Goodman, M.M., Sanchez, G.J., Buckler, E., and Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. Proc. Natl. Acad. Sci. U S A **99**:6080–6084.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res. **20**:1297–1303.

Meng, L., Li, H.H., Zhang, L.Y., and Wang, J.K. (2015). QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. Crop J. **3**:269–283.

Murray, M.G., and Thompson, W.F. (1980). Rapid isolation of high molecular-weight plant DNA. Nucleic Acids Res. **8**:4321–4325.

Nguyen, K.T., Park, J., Park, E., Lee, I., and Choi, G. (2015). The Arabidopsis RING domain protein BOI inhibits flowering via CO-dependent and CO-independent mechanisms. Mol. Plant **8**:1725–1736.

Pan, Q.C., Li, L., Yang, X.H., Tong, H., Xu, S.T., Li, Z.G., Li, W.Y., Muehlbauer, G.J., Li, J.S., and Yan, J.B. (2016). Genome-wide

recombination dynamics are associated with phenotypic variation in maize. New Phytol. **210**:1083–1094.

Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., and Sanchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol. Biol. Evol. **34**:3299–3302.

Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A.H., Nielsen, K.L., Jorgensen, J.E., Weigel, D., and Andersen, S.U. (2009). SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nat. Methods **6**:550–551.

Settles, A.M., Bagadion, A.M., Bai, F., Zhang, J.Y., Barron, B., Leach, K., Mudunkothge, J.S., Hoffner, C., Bihmidine, S., Finefield, E., et al. (2014). Efficient molecular marker design using the maizeGDB Mo17 SNPs and indels track. G3 (Bethesda) **4**:1143–1145.

Silva Lda, C., Wang, S., and Zeng, Z.B. (2012). Composite interval mapping and multiple interval mapping: procedures and guidelines for using Windows QTL Cartographer. Methods Mol. Biol. **871**:75–119.

Stuber, C.W., Edwards, M.D., and Wendel, J.F. (1987). Molecular marker-facilitated investigations of quantitative trait loci in maize. 2. Factors influencing yield and its component traits. Crop Sci. **27**:639–648.

Sun, S.L., Zhou, Y.S., Chen, J., Shi, J.P., Zhao, H.M., Zhao, H.N., Song, W.B., Zhang, M., Cui, Y., Dong, X.M., et al. (2018). Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. Nat. Genet. **50**:1289–1295.

Sun, Y.P., Wang, J.K., Crouch, J.H., and Xu, Y.B. (2010). Efficiency of selective genotyping for genetic analysis of complex traits and potential applications in crop improvement. Mol. Breed. **26**:493–511.

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., Uemura, A., Utsushi, H., Tamiru, M., Takuno, S., et al. (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J. **74**:174–183.

Takagi, H., Tamiru, M., Abe, A., Yoshida, K., Uemura, A., Yaegashi, H., Obara, T., Oikawa, K., Utsushi, H., Kanzaki, E., et al. (2015). MutMap accelerates breeding of a salt-tolerant rice cultivar. Nat. Biotechnol. **33**:445–449.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28**:2731–2739.

Tang, X., Liu, G.Q., Zhou, J.P., Ren, Q.R., You, Q., Tian, L., Xin, X.H., Zhong, Z.H., Liu, B.L., Zheng, X.L., et al. (2018). A large-scale whole-genome sequencing analysis reveals highly specific genome editing by both Cas9 and Cpf1 (Cas12a) nucleases in rice. Genome Biol. **19**:84.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. **31**:46–53.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. **7**:562–578.

Wen, Y.J., Zhang, Y.W., Zhang, J., Feng, J.Y., Dunwell, J.M., and Zhang, Y.M. (2018). An efficient multi-locus mixed model framework for the detection of small and linked QTLs in F2. Brief Bioinform. https://doi.org/10.1093/bib/bby058.

Wills, D.M., Fang, Z., York, A.M., Holland, J.B., and Doebley, J.F. (2018). Defining the role of the MADS-box gene, *Zea Agamous-like1*, a target of selection during maize domestication. J. Hered. **109**:333–338.

Xing, A.Q., Gao, Y.F., Ye, L.F., Zhang, W.P., Cai, L.C., Ching, A., Llaca, V., Johnson, B., Liu, L., Yang, X.H., et al. (2015). A rare SNP mutation in Brachytic2 moderately reduces plant height and increases yield potential in maize. J. Exp. Bot. **66**:3791–3802.

Yamamoto, T., Kuboki, Y., Lin, S.Y., Sasaki, T., and Yano, M. (1998). Fine mapping of quantitative trait loci Hd-1, Hd-2 and Hd-3, controlling heading date of rice, as single Mendelian factors. Theor. Appl. Genet. **97**:37–44.

Yan, L., Loukoianov, A., Tranquilli, G., Helguera, M., Fahima, T., and Dubcovsky, J. (2003). Positional cloning of the wheat vernalization gene VRN1. Proc. Natl. Acad. Sci. U S A **100**:6263–6268.

Yang, N., Lu, Y., Yang, X., Huang, J., Zhou, Y., Ali, F., Wen, W., Liu, J., Li, J., and Yan, J. (2014). Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. PLoS Genet. **10**:e1004573.

Yang, Q., Li, Z., Li, W.Q., Ku, L.X., Wang, C., Ye, J.R., Li, K., Yang, N., Li, Y.P., Zhong, T., et al. (2013). CACTA-like transposable element in *ZmCCT* attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. Proc. Natl. Acad. Sci. U S A **110**:16969–16974.

Yang, X.H., Yan, J.B., Shah, T., Warburton, M.L., Li, Q., Li, L., Gao, Y.F., Chai, Y.C., Fu, Z.Y., Zhou, Y., et al. (2010). Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. Theor. Appl. Genet. **121**:417–431.

Yant, L., Mathieu, J., Dinh, T.T., Ott, F., Lanz, C., Wollmann, H., Chen, X.M., and Schmid, M. (2010). Orchestration of the floral transition and floral development in *Arabidopsis* by the bifunctional transcription factor APETALA2. Plant Cell **22**:2156–2170.

Zeng, Z.B. (1994). Precision mapping of quantitative trait loci. Genetics **136**:1457–1468.

Zhang, H.W., Ma, P., Zhao, Z.N., Zhao, G.W., Tian, B.H., Wang, J.H., and Wang, G.Y. (2012). Mapping QTL controlling maize deep-seeding tolerance-related traits and confirmation of a major QTL for mesocotyl length. Theor. Appl. Genet. **124**:223–232.

Zhang, Y.W., Liu, Y.J., Ren, Y., Liu, Y., Liang, G.M., Song, F.P., Bai, S.X., Wang, J.H., and Wang, G.Y. (2013). Overexpression of a novel Cry1Ie gene confers resistance to Cry1Ac-resistant cotton bollworm in transgenic lines of maize. Plant Cell Tissue Organ Cult. **115**:151–158.

Zou, C., Wang, P.X., and Xu, Y.B. (2016). Bulked sample analysis in genetics, genomics and crop improvement. Plant Biotechnol. J. **14**:1941–1955.

Zuo, W.L., Chao, Q., Zhang, N., Ye, J.R., Tan, G.Q., Li, B.L., Xing, Y.X., Zhang, B.Q., Liu, H.J., Fengler, K.A., et al. (2015). A maize wall-associated kinase confers quantitative resistance to head smut. Nat. Genet. **47**:151–157.