Genome-Wide Association Studies and Whole-Genome Prediction Reveal the Genetic Architecture of KRN in Maize

Yixin An
Chinese Academy of Agricultural Sciences Institute of Crop Sciences
Lin Chen
Chinese Academy of Agricultural Sciences Institute of Crop Sciences
Yongxiang Li
Chinese Academy of Agricultural Sciences Institute of Crop Sciences
Chunhui Li
Chinese Academy of Agricultural Sciences Institute of Crop Sciences
Yunsu Shi
Chinese Academy of Agricultural Sciences Institute of Crop Sciences
Dengfeng Zhang
Chinese Academy of Agricultural Sciences Institute of Crop Sciences
Yu Li (liyu03@caas.cn)
Institute of Crop Sciences
Tianyu Wang
Chinese Academy of Agricultural Sciences Institute of Crop Sciences

Research article

Keywords: maize, kernel row number, genome-wide association study, quantitative trait nucleotide, whole-genome prediction

Posted Date: June 10th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-32231/v1

Version of Record: A version of this preprint was published on October 27th, 2020. See the published version at https://doi.org/10.1186/s12870-020-02676-x.

Abstract

Background: Kernel row number (KRN) is an important trait for the domestication and improvement of maize. To explore the genetic basis of KRN has great research significance and can provide the valuable information for molecular assisted selection.

Results: In this study, one single-locus method (MLM) and six multi-locus methods (mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, pKWmEB and ISIS EM-BLASSO) of genome-wide association studies (GWASs) were used to identify significant quantitative trait nucleotides (QTNs) for KRN in an association panel including 639 maize inbred lines that were genotyped by the MaizeSNP50 BeadChip. In three phenotyping environments and with best linear unbiased prediction (BLUP) values, seven GWAS methods revealed different numbers of KRN-associated QTNs, ranging from 11 to 177. Based on these results, seven important regions for KRN located on chromosomes 1, 2, 3, 5, 9, and 10 were identified by at least three methods and in at least two environments. Moreover, 49 genes from the seven regions were expressed in different maize tissues. Among the 49 genes, ARF29 (Zm00001d026540, encoding auxin response factor 29) and CKO4 (Zm00001d043293, encoding cytokinin oxidase protein) were significantly related to KRN based on expression analysis and candidate gene association mapping. Whole-genome prediction (WGP) for KRN was also performed, and we found that the KRN-associated tagSNPs achieved a high prediction accuracy. The best strategy was to integrate the total KRN-associated tagSNPs identified by all GWAS models.

Conclusions: These results aid in our understanding of the genetic architecture of KRN and provide useful information for genomic selection for KRN in maize breeding.

Background

Maize (*Zea mays* L.) arose from a single domestication from its wild progenitor, teosinte, in southern Mexico approximately 9,000 years ago and is now one of the most important cereal crops worldwide [1]. During domestication, its morphological characteristics, especially inflorescence architectures, differed profoundly [2, 3]. The shift from the small ears in teosinte to a larger ear in modern maize was accompanied by a dramatic increase in kernel row number (KRN) [4]. Thus, constant efforts have been made to explore the genetic basis underlying the striking diversities in inflorescence architecture and KRN in maize.

KRN is an important ear trait and is formed by multiple meristem types during female inflorescence development, including inflorescence meristems (IMs), spikelet pair meristems (SPMs), spikelet meristems (SMs) and floral meristems (FMs) [5]. To date, some genes have been cloned and found to be involved in complex regulatory networks responsible for meristem development and KRN modification by studying mutants [6-10]. However, these classical mutants show negative pleiotropy for other traits of plant architecture and are difficult to directly use in maize breeding [11]. Therefore, linkage and

association mapping have been performed in naturally varying populations with the aim of identifying more elite natural alleles controlling KRN.

Although many quantitative trait loci (QTLs) related to KRN were identified by linkage mapping in biparental segregation populations, few have been successfully cloned due to their small genetic effects, except for *KRN4* [12] and *KRN1* [13]. GWASs of KRN have also been conducted and revealed many quantitative trait nucleotides (QTNs) [14-16]. At the same time, GWAS results can be easily influenced by the population structure and rare variants in natural populations [17]. Therefore, many statistical models have been developed to improve the power of identifying genome-phenotype associations when using GWAS, such as the single-locus method mixed linear model (MLM) [18, 19] and the multi-locus methods mrMLM [20], ISIS EM-BLASSO [21], pLARmEB [22], FASTmrEMMA [23], pKWmEB [24], and FASTmrMLM [25]. The MLM is a single-locus fixed-SNP-effect approach used under polygenic background and population structure controls [18, 19]. To reduce the false positive rate (FPR), stringent Bonferroni correction in multiple test correction is used in MLM [26]. The multi-locus method is an alternative GWAS procedure that is based on a random-SNP-effect model, and no multiple test correction is needed [26]. There are two steps in this model. First, the reduced number of SNPs is selected through different algorithms, and the SNPs are then used in the multi-locus model to detect the true signals [20-26]. Recently, there have been a few studies focusing on the above GWAS methods to detect important loci controlling different traits in rice [27], maize [28], flax [29], bread wheat [30] and upland cotton [31, 32].

Previous studies have revealed that KRN is quantitatively inherited and that the effects of a single genetic locus are generally small, which poses challenges for genetic improvement in maize breeding. Therefore, the best approach is to improve the ability to predict KRN by integrated analysis of more markers distributed throughout the whole genome. Genomic selection (GS), or whole-genome prediction (WGP), has the capacity to use full-genome data to increase breeding efficiency [33]. In previous studies, WGPs of KRN were performed in F_1 hybrids between recombinant inbred lines [34], interconnected biparental maize populations [35] and 339 maize inbred lines [36], all of which showed that KRN was a trait suitable for genome-wide prediction. Liu et al. [15] showed that approximately 300 top KRN-associated tagSNPs were sufficient for predicting the KRN of inbred lines and hybrids using ridge regression best linear unbiased prediction (rr-BLUP). Based on these analyses, we are faced with how to select fewer markers to accurately predict KRN. Several studies have reported that selecting association markers from the results of GWASs and included these as fixed effects into WGP models resulted in better performance than that achieved with other single WGP models [37-39]. This might provide a way to simultaneously model different aspects of the genetic architecture and is especially accessible to breeders [39].

In this study, we performed a GWAS of an association panel including 639 maize inbred lines based on the MaizeSNP50 BeadChip by using one single-locus method, MLM, and six multi-locus methods, mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, pKWmEB and ISIS EM-BLASSO. The common significant QTNs codetected by different methods and across different environments were analyzed, and the candidate genes related to KRN were further predicted. WGP was also performed using various KRN-related tagSNPs to dissect the genetic architecture of KRN.

Results

Natural variation in KRN within the association panel

KRN was measured within our association panel, which included 639 maize inbred lines, in XX, BJ and GZL in 2011. The results showed that KRN was normally distributed under each environment, and the KRNs among environments were highly positively correlated, ranging from 0.73 between XX and BJ to 0.79 between XX and GZL (Figure 1a). The KRN exhibited high broad-sense heritability (H^2 = 0.90, Table 1), which was similar to the results of previous studies [14, 16]. Comparing the KRN among the different environments, we found that the KRN showed the smallest average (13.69), minimum (8.60) and maximum (20.60) values in XX, where all accessions were summer-planted in June. With increasing latitude, where the accessions were spring-planted in May, the average KRN increased (14.65 in BJ and 14.59 in GZL). The largest range (max - min) in KRN appeared in GZL (12.60), which had the most hours of daylight (Table 1). Based on the previous results, our association panel could be divided into five subgroups: Reid, tangsipingtou (TSPT), lvdahonggu (LRC), Lancaster and P. The KRN statistical analysis results of various subgroups are shown in Table S1. There was no significant difference in KRN among the five subgroups (Figure 1b). We found that the Reid subgroup had the highest value (14.65) for KRN, and the Lancaster subgroup had the lowest value (14.26) for KRN. The LRC subgroup showed the largest variation, ranging from 8.8 to 19.51 (Table S1). These results indicated that KRN was a quantitative trait and that the phenotypic variation among the tested inbred lines in the association panel was beneficial for dissecting the genetic architecture of KRN.

Table 1. Phenotypic variance in KRN for 639 maize inbred lines in three environments.

Env	Mean	Min	Max	*SD*	*CV* (%)	H^2
XX	13.69	8.60	20.60	2.02	14.76	0.90
BJ	14.65	9.20	21.00	1.69	11.56	
GZL	14.59	8.60	21.20	2.00	13.69	
BLUP	14.31	9.17	20.01	1.61	11.27	

Env, Environment; XX, Xinxiang; BJ, Beijing; GZL, Gongzhuling; Max, maximum; Min, minimum; *SD*, standard deviation; *CV*, coefficient of variation; H^2, broad-sense heritability.

QTNs for KRN identified by different methods

Single-locus analysis for KRN (MLM)

Based on the MaizeSNP50 BeadChip, we obtained 42,667 high-quality SNPs distributed on 10 maize chromosomes. Under the $P < 0.0001$ and $P < 0.001$ thresholds, 3/55, 3/45, 1/24, and 3/51 KRN-associated QTNs were found in XX (Figure 2a), BJ (Figure 2b), GZL (Figure 2c) and BLUP (Figure 2d), respectively. To account for overcorrection in this model, the $P < 0.001$ thresholds were selected to identify KRN-associated QTNs. Finally, 177 QTNs were found to be associated with KRN, and the proportion of phenotypic variance explained (PVE) by these individual QTNs ranged from 1.84 to 4.01% (Table S2).

Multiple-locus analysis for KRN

Using different multiple-locus models, we identified different numbers of significant QTNs for KRN in XX, BJ, GZL and together with BLUP across all locations. These QTNs were unevenly distributed on 10 chromosomes, with the most QTNs on Chr. 1 and the fewest on Chr. 8. Specifically, 15 (FASTmrEMMA)-177 (mrMLM) QTNs for XX, 11 (FASTmrEMMA)-30 (ISIS EM-BLASSO) QTNs for BJ, 12 (FASTmrEMMA)-55 (mrMLM) QTNs for GZL and 11 (FASTmrEMMA)-106 (mrMLM) QTNs for BLUP were identified by the six different methods (Table S3). Comparative analysis of the GWAS results among different statistical approaches showed that FASTmrEMMA detected the fewest QTNs in all the environments, while mrMLM detected the most QTNs in all the environments, except for BJ (Table S3). The QTN overlap analysis among the seven methods indicated that the majority of the association QTNs were specifically identified by only one method, followed by the QTNs codetected by two and three methods (Figure S1a and Table S4). Overall, ISIS EM-BLASSO, which detected the third largest number of QTNs, identified the most codetected QTNs, followed by FASTmrMLM (Table S4). The comparative analysis of the GWAS results among the different environments showed that the majority of the significant QTNs were specifically identified in only one environment (Figure S1b). The low repeatability of associated QTNs among various environments suggested that the genetic loci controlling KRN were environmentally sensitive, although the KRN was similar and consistently showed high heritability across all locations.

Annotation and expression of candidate genes for KRN

To obtain reliable significant QTNs and predict the candidate genes for KRN, only the QTNs simultaneously identified by at least three methods (either single-locus or multi-locus) and in at least two environments were used for the next analysis. Finally, seven QTNs controlling KRN were obtained (Table 2). The seven QTNs were located on chromosomes 1, 2, 3, 5, 9, and 10, and the PVE by these QTNs ranged from 1.06% to 5.21%. Based on the LD of the association panel (Figure S2), 49 genes around the QTNs (200 kb upstream and downstream) were obtained, and their expression varied widely in different maize tissues (Figure 3a and Table S5). For example, Zm00001d016760, which encodes the abscisic acid stress ripening6 protein, is highly expressed in the roots, Zm00001d031426, which encodes serine/threonine-protein kinase, and Zm00001d043298, which encodes a P-loop containing nucleoside triphosphate hydrolase superfamily protein, are highly expressed in tassels and anthers. Among the 49

genes, 22 genes were differentially expressed in different spike development mutants (Table S6), which suggested that these 22 genes might be involved in ear development in maize.

Table 2. Significant KRN-associated QTNs codetected in at least two environments and by at least three models.

SNP	Chr	Pos	Single-locus GWAS (MLM)		Multi-locus GWAS		
			LOD	PVE (%)	LOD	PVE (%)	Methods[1]
01124566	1	156580056	3.44	3.00	4.60-11.63	1.91-3.02	1, 2, 3, 4, 5, 6, 7
01144585	1	187526525	3.13	2.00	4.39-5.95	1.84-3.51	1, 3, 4, 5, 7
02176259	2	219023013	3.32	3.00	3.41-4.17	1.06-2.04	1, 2, 3, 4, 7
110967306-138	3	191981941	3.28	2.56	8.17-11.77	1.62-3.37	1, 2, 5, 6
05114980	5	171187130	/	/	4.35-8.20	1.15-2.29	2, 3, 5,6,7
09047930	9	79941271	4.61	4.00	5.73-10.40	2.43-5.21	1, 2, 3, 5, 6, 7
10106563	10	146944098	3.61	3.00	3.65-5.25	1.18-2.38	1, 2, 3, 4, 5, 6, 7

Methods[1]: Numbers 1 to 7 represent different GWAS methods: 1: MLM; 2: mrMLM; 3:

FASTmrMLM; 4: FASTmrEMMA; 5: pLARmEB; 6: pKWmEB; 7: ISIS EM-BLASSO.

Interestingly, we found that Zm00001d026540, which encodes auxin response factor29 (ARF29), had a higher expression in SAM and ears than in other tissues. Candidate gene association mapping showed that five SNPs (two SNPs in the gene and three SNPs in the gene upstream) around *ARF29* were significantly related to KRN (Figure 3b and Table 3). ARF29 can bind the *Bif1* (which is related to SAM development and final KRN) promoter by recognizing the TTTCGG motif [40, 41]. The S10_147122969 SNP, located within the gene body, was significantly associated with KRN. Two alleles for this SNP (A/T) were present in this panel, with the A allele having a higher KRN. Cytokinins also play an important role in the development of immature spikes and the formation of final KRN [42]. For example, *UB3* regulates KRN by the cytokinin pathway and CLAVATA-WUSCHEL pathway [42]. In this study, *CKO4* (Zm00001d043293, encodes cytokinin oxidase protein) was detected, and candidate gene association mapping for *CKO4* was also conducted. The results showed that two SNPs located upstream of *CKO4* were significantly associated with KRN (Figure 3c and Table 3). The S3_191837578 SNP had two alleles (T/G), and the T allele was associated with a higher KRN but had a lower allele frequency. This suggested that this allele may not be widely used in maize breeding.

Table 3. Candidate gene association analysis.

Gene ID	SNP	Chr	Pos	LOD	PVE	Allele	Frequency
ARF29	S10_147122969	10	147122969	4.57	8.97%	A/T	127/99
	S10_147121954	10	147121954	4.44	8.98%	G/A	94/90
	S10_147126021	10	147126021	3.88	7.58%	T/A	161/27
	S10_147123193	10	147123193	3.33	5.30%	A/C	119/110
	S10_147141311	10	147141311	3.17	4.92%	C/G	211/21
CKO4	S3_191837578	3	191837578	4.64	7.85%	G/T	177/45
	S3_191841761	3	191841761	4.67	6.99%	T/G	236/16

Whole-genomic prediction of KRN

We first analyzed the LD blocks of all markers using the threshold value $r^2 > 0.2$ and obtained 27,688 tagSNPs in our association panel. Then, we randomly selected different numbers of tagSNPs, from 5 to 27,000, in the whole genome to calculate the prediction accuracies for the KRN of the inbred lines. The results showed that the prediction accuracies increased as the number of tagSNPs increased (Figure 4a and Table S7). More specifically, the prediction accuracies sharply increased when the tagSNPs increased from 5 to 500 and then slowly grew when the tagSNPs increased from 400 to 2000. After 2000, the prediction accuracies maintained a consistently high level. Although a large number of tagSNPs was used to predict the KRN, the prediction accuracies were still less than 0.5. The effects of training population size on the prediction accuracy were also conducted based on the marker number 14000 (approximately 50% of the total tagSNPs). In the association panel, the prediction accuracies improved with increasing training population size. When the training population size increased from 50% to 90%, a slight increase was observed in the prediction accuracy (Figure 4b and Table S7).

To better understand the genetic architecture of KRN and improve its predictive ability, we ranked the 27,688 tagSNPs according to their significance to KRN obtained by MLM to obtain the top tagSNPs. We found that these top tagSNPs had a higher prediction accuracy (ranging from 0.60 for the top 100-tagSNPs to 0.74 for the top 700-tagSNPs) than these randomly selected tagSNPs (ranging from 0.22 for 100 random tagSNPs to 0.33 for 700 random tagSNPs) (Figure 4c and Table S7).

The tagSNPs representing the significant QTNs detected by different models based on the BLUP were collected and used to calculate the prediction accuracies for KRN in our association panel. The results showed that these tagSNPs identified by different methods had different prediction accuracies ranging from 0.43 (FASTmrEMMA) to 0.60 (ISIS EM-BLASSO) (Figure 4d and Table S8). We also found that the tagSNPs associated with KRN identified by the same method showed different prediction accuracies in diverse environments (Table S8). To explore whether using the codetected QTNs in different GWAS methods could increase the prediction accuracies for KRN, we selected the common QTNs identified by at least two, three, four, five and six methods to conduct the predictions. The results showed that only the common QTNs identified by at least two methods (common ≥ 2) could maintain predictability at a high level; other common QTNs had no advantage in predicting KRN, which may be due to the fewer QTN numbers (Table S8).

Additionally, to improve the prediction ability, we put the KRN-related tagSNPs detected by seven methods together in a single environment (204 in XX, 87 in BJ, 118 in GZL and 167 in BLUP), namely, M-total tagSNPs, to conduct the KRN prediction. As a result, we found that the prediction accuracies were improved sharply and reached 0.74 in XX, 0.66 in BJ, 0.75 in GZL and 0.75 in BLUP (Figure 4d and Table S8). These predictabilities were much higher than those of the single method in each environment (Table S8). Then, we collected the tagSNPs associated with KRN from all methods and all environments,

namely, E-M-total tagSNPs, and obtained 439 tagSNPs in total. However, there was only a slight increase in prediction accuracies (ranging from 0.68 in BJ to 0.79 in BLUP for the 439 tagSNPs) when we used the much higher number of E-M-total tagSNPs compared to the fewer M-total tagSNPs (Figure 4d and Table S8).

Discussion

To date, the GWAS approach has been widely used to investigate the genetic foundation of important traits in many species by calculating the association between genotypic and corresponding phenotypic variations [43]. To separate the true association signals, many statistical methods based on different algorithms have been established. In this study, we selected one single-locus method, MLM, and six multi-locus methods, mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, pKWmEB and ISIS EM-BLASSO, to perform comprehensive GWAS mapping for KRN in our association panel. Among the seven methods, mrMLM identified the largest number of QTNs, FASTmrEMMA identified the fewest number of QTNs, and ISIS EM-BLASSO identified the most codetected QTNs, which was consistent with the results identified by Cui et al. [27] for salt-tolerance loci in rice. Therefore, multi-locus models are valuable alternative methods for GWASs of KRN in maize. Additionally, the small number of common QTNs codetected by different methods was also observed in the study of Peng et al. [30] for free amino acid levels in bread wheat.

Comparing our GWAS results with those of previous studies, we found that some important genes controlling inflorescence architecture in maize were located within ± 200 kb of the significant QTNs (Table S9), including *CT2* (Zm00001d027886), *FEA3* (Zm00001d040130), *BAD1* (Zm00001d005737), *RA1* (Zm00001d020430), *VT2* (Zm00001d008700) and so on. Among these genes, *CT2* [7] and *FEA3* [10] function in CLAVATA-WUSCHEL feedback signaling, and their mutations result in enlarged and fascinated ear primordia and increased KRN. *BAD1* [44] and *RA1* [45], which both encode transcription factors, are involved in the genetic regulation of the floral branch system by the ROMASO pathway in maize. *VT2* [46] functions in auxin biosynthesis and has dramatic effects on vegetative and reproductive development, and the mutant ears showed obvious defects. Additionally, approximately 60% of the significant QTNs were codetected with previous GWAS mapping of inflorescence development within the LD regions, and some of these loci were pleiotropic [14, 15].

Additionally, WGP is also an effective method in animal breeding and plant improvement [47]. Because KRN is mainly controlled by additive loci, we selected the additive model rrBLUP to conduct WGP [48]. As expected, the prediction accuracy increased as the number of randomly selected tagSNPs increased, which was consistent with Liu et al. [15] and determined by the influence of marker density on the WGP [47]. However, the randomly selected tagSNPs showed a low predictive ability, and thus we decided to combine the GWAS results with WGP to explore the best marker dataset for KRN prediction. As a result, higher prediction levels were easily reached when using the significant tagSNPs, and the moderate to high values were consistent with those reported by Liu et al. [15], Guo et al. [34], Riedelsheimer et al. [35] and Xu et al. [36]. This result suggested that integrating significant signals from GWASs into WGP models as fixed effects was effective for enhancing the prediction of KRN. A similar conclusion was

drawn by Liu et al. [15] for KRN, by Bian and Holland [49] for resistance to southern leaf blight (SLB) and gray leaf spot (GLS) as well as for plant height (PHT) in maize and by Spindel et al. [39] for tropical rice improvement. Although there were different evaluations of WGP models incorporating peak GWAS signals in maize and sorghum [50], our research indicated that the use of QTNs passing a certain threshold in the above GWAS methods as fixed effects in the rrBLUP model was a powerful tool for KRN prediction, which was a trait-specific consideration in the given population in this study.

Based on the results of this study, we suggest that KRN is controlled by many additive loci and that the rrBLUP model can be used to conduct KRN prediction in maize inbred lines. The combined utilization of different GWAS methods is helpful for predicting candidate genes and KRN for maize breeding.

Conclusions

In this study, multiple methodsof genome-wide association studies (GWASs) were used to identify significant QTNs for KRN in maize. The seven GWAS methods revealed different numbers of KRN-associated QTNs, ranging from 11 to 177. Based on these results, seven important regions for KRN located on chromosomes 1, 2, 3, 5, 9, and 10 were identified by at least three methods and in at least two environments. Moreover, 49 genes from the seven regions were expressed in different maize tissues. Among the 49 genes, *ARF29* (Zm00001d026540, encoding auxin response factor 29) and *CKO4* (Zm00001d043293, encoding cytokinin oxidase protein) were significantly related to KRN based on expression analysis and candidate gene association mapping. Whole-genome prediction (WGP) for KRN was also performed, and we found that the KRN-associated tagSNPs achieved a high prediction accuracy. The best strategy was to integrate the total KRN-associated tagSNPs identified by all GWAS models. These results will facilitate our understanding of genetic basis of KRN and provide important candidate genes for further study on KRN.

Methods

Plant materials and phenotyping

An association panel of 639 maize inbred lines, representing a wide range of genetic diversity of temperate inbred lines in China [51], was collected for GWASs in this study. These plant materials in this study were conserved in our lab and we declare that all plant materials used in this study comply with the ‘Convention on the Trade in Endangered Species of Wild Fauna and Flora’.

All the accessions were planted with a randomized block design of three replicates under three environments in 2011: Gongzhuling in Jilin Province (43.50°N, 124.82°E), Xinxiang in Henan Province (35.19°N, 113.53°E) and Beijing (39.48°N, 116.28°E) in 2011. For descriptive purposes, the three environments were designated GZL, XX and BJ, respectively. At each location, the field experiments were arranged in a single row 3 m in length, with 0.6 m between adjacent rows and 12 individual plants per row. The institute of crop science belonging to the Chinese Academy of Agricultural Sciences has set up

experimental field bases at all the above locations. The institute of crop science was approved for field experiments, and the field management followed local maize management practices.

Five ears were harvested from each line, and the KRN was evaluated from the middle part of the ears [51]. The BLUPs were calculated using the SAS PROC MIXED model, with genotype, environment and replication as random effects [14, 52]. The broad-sense heritability (H^2) for KRN was calculated according to Wu et al. [53]. The coefficient of variation was calculated as CV (%) = SD/mean, where SD and mean refer to the standard deviation and mean of the KRN in each environment [52].

DNA extraction and genotyping

Genomic DNA was extracted from young maize leaves from five plants of each line according to the cetyltrimethylammonium bromide (CTAB) method [54]. All samples were quality checked and genotyped using the MaizeSNP50 BeadChip, which is an Illumina BeadChip array of 56,110 maize SNPs developed from the B73 reference sequence [55]. Then, the successfully called SNPs with a missing rate of more than 20% and minor allele frequency (MAF) of <0.05 were excluded from the genotyping dataset [56]. After that, 42,667 high-quality SNPs were used in further analysis.

GWAS mapping

One single-locus method, MLM, and six multi-locus methods, including mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, pKWmEB, and ISIS EM-BLASSO, were used in this study. Alleles of each polymorphism with a minor frequency > 0.05 were used for further analysis. Both the kinship matrix and the principal component analysis (PCA) were performed with the Tassel 5.2 program [57]. An MLM controlling for population structure (Q) and kinship (K) (MLM Q+K) was also conducted in Tassel 5.2 [18, 19]. Six multi-locus GWAS mapping methods were used along with the software package mrMLM. GUI v3.2 in the R environment (http://127.0.0.1:5846/) [26]. All parameters were set at default values, the critical thresholds of significant association for the MLM were set at $-\log_{10}{}^{P} \geq 3$, and the six multi-locus methods were set at LOD ≥3 [26].

Candidate gene analysis

The candidate genes around the 200-kb region of the significant QTNs detected by at least three models and two environments were collected based on the B73 reference genome V4 from MaizeGDB (https://www.maizegdb.org/). The expression data of these genes were collected from previous studies [58, 59]. The genome fragments of the SNPs within the selected genes, including a 10-kb promoter region, the gene bodies and a 10-kb region downstream of the genes, were obtained from the maize HapMap3 dataset [60]. The candidate gene mapping analyses were conducted on a global maize association mapping panel of 282 diverse lines. The phenotypes of this association panel were provided in our previous report [53]. The association analysis was conducted by the MLM in TASSEL 5.2, controlling for the population structure (Q) and kinship (K). The first three principal components (PCs), which were analyzed in a previous study [53], were used as the covariant variables to control the existing population

structure in the 282-line association mapping panel. Significant marker-trait associations were declared at $-\log_{10}{}^{P} > 3$.

Genomic prediction of KRN

To predict the KRN of the inbred lines, we estimated the predictability by WGP. We grouped the linkage disequilibrium (LD) blocks in PLINK software [61] using the threshold value $r^2 > 0.2$ and found the tagSNPs according to the LD blocks. The ridge regression best linear unbiased prediction (rrBLUP) package was used to perform the genomic prediction in R [62]. We randomly selected half of the lines of our association panel as the training population (320 inbred lines) and the remaining 319 inbred lines as the validation population [15]. We used the KRN-related tagSNPs identified by different methods to perform the genomic prediction of KRN for the inbred lines under four conditions (XX, BJ, GZL and BLUP). Simultaneously, 5 to 27,000 randomly selected tagSNPs, the total tagSNPs related to KRN identified by the seven methods in a single environment (M-total tagSNPs), the total tagSNPs for KRN from all methods and environments (E-M-total tagSNPs) and the common tagSNPs for KRN detected by at least two, three, four, five, and six methods were also used for the same procedure. The random sampling of tagSNP numbers, the training and validation populations and the predictions were all repeated 100 times.

Abbreviations

BJ: Beijing

FMs: floral meristems

GWAS: genome-wide association study

GZL: Gongzhuling

IM: inflorescence meristem

KRN: kernel row number

LD: linkage disequilibrium

MLM: mixed linear model

QTL: quantitative trait locus

QTN: quantitative trait nucleotide

SMs: spikelet meristems

SNP: single nucleotide polymorphism

SPMs: spikelet pair meristems

XX: Xinxiang

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

Thanks to the the National Natural Science Foundation (91735306, 31801373) we were able to conduct the genotyping of the association mapping panel. And, thanks to the Ministry of Science and Technology of China (2016YFD0100303, 2016YFD0100103), and the CAAS Innovation Program we were able to conduct the phenotype identification of the association mapping panel. None of these funding bodies have any relationship with the publication of this manuscript.

Authors' contributions

Y.A. and L. C. performed the GWAS and WGP and drafted the manuscript; Y-x. L. and C. L. conceived the study and helped to discuss the results. Y. S. and D. Z. lead the planning of this study. T.W. and Y. L. designed the research and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Authors' information

Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China

References

1. Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez GJ, Buckler E, Doebley J. A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci U S A. 2002;99:6080-4.
2. Iltis HH. From teosinte to maize: the catastrophic sexual transmutation. Science. 1983;222:886-94.
3. Iltis HH. Homeotic sexual translocations and the origin of maize (*Zea Mays*, *Poaceae*): a new look at an old problem. Econ Bot. 2000;54:7-42.
4. Doebley J. The genetics of maize evolution. Annu Rev Genet. 2004;38:37-59.
5. Thompson BE, Hake S. Translational biology: from Arabidopsis flowers to grass inflorescence architecture. Plant Physiol. 2009;149:38-45.
6. Taguchi-Shiobara F, Yuan Z, Hake S, Jackson D. The fasciated ear2 gene encodes a leucine-rich repeat receptor-like protein that regulates shoot meristem proliferation in maize. Genes Dev. 2001;15:2755-66.
7. Bommert P, Je BI, Goldshmidt A, Jackson D. The maize Galpha gene COMPACT PLANT2 functions in CLAVATA signalling to control shoot meristem size. Nature. 2013;502:555-8.
8. Bommert P, Nagasawa NS, Jackson D. Quantitative variation in maize kernel row number is controlled by the FASCIATED EAR2 locus. Nat Genet. 2013;45:334-7.
9. Chuck GS, Brown PJ, Meeley R, Hake S. Maize SBP-box transcription factors unbranched2 and unbranched3 affect yield traits by regulating the rate of lateral primordia initiation. Proc Natl Acad Sci U S A. 2014;111:18775-80.
10. Je BI, Gruel J, Lee YK, Bommert P, Arevalo ED, Eveland AL, et al. Signaling from maize organ primordia via FASCIATED EAR3 regulates stem cell proliferation and yield traits. Nat Genet. 2016;48:785-91.
11. Li M, Zhong W, Yang F, Zhang Z. Genetic and molecular mechanisms of quantitative trait loci controlling maize inflorescence architecture. Plant Cell Physiol. 2018;59:448-57.
12. Liu L, Du Y, Shen X, Li M, Sun W, Huang J, et al. KRN4 controls quantitative variation in maize kernel row number. PLoS Genet. 2015;11:e1005670.
13. Wang J, Lin Z, Zhang X, Liu H, Zhou L, Zhong S, et al. KRN1, a major quantitative trait locus for kernel row number in maize. New Phytol. 2019;223:1634-46.
14. Brown PJ, Upadyayula N, Mahone GS, Tian F, Bradbury PJ, Myles S, et al. Distinct genetic architectures for male and female inflorescence traits of maize. PLoS Genet. 2011;7:e1002383.
15. Liu L, Du Y, Huo D, Wang M, Shen X, Yue B, et al. Genetic architecture of maize kernel row number and whole genome prediction. Theor Appl Genet. 2015;128:2243-54.
16. Xiao Y, Tong H, Yang X, Xu S, Pan Q, Qiao F, et al. Genome-wide dissection of the maize ear genetic architecture using multiple populations. New Phytol. 2016;210:1095-106.
17. Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, Mitchell SE, et al. Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J. 2005;44:1054-64.

18. Zhang Y-M, Mao Y, Xie C, Smith H, Luo L, Xu S. Mapping quantitative trait loci using naturally occurring genetic variance among commercial inbred lines of maize (*Zea mays* L.). Genetics. 2005;169:2267-75.
19. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet. 2006;38:203-8.
20. Wang S-B, Feng J-Y, Ren W-L, Huang B, Zhou L, Wen Y-J, et al. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. Sci Rep. 2016;6:19444.
21. Tamba CL, Ni Y-L, Zhang Y-M. Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. PLoS Comput Biol. 2017;13:e1005357.
22. Zhang J, Feng JY, Ni YL, Wen YJ, Niu Y, Tamba CL, et al. pLARmEB: integration of least angle regression with empirical Bayes for multilocus genome-wide association studies. Heredity (Edinb). 2017;118:517-24.
23. Wen YJ, Zhang H, Ni YL, Huang B, Zhang J, Feng JY, et al. Methodological implementation of mixed linear models in multi-locus genome-wide association studies. Brief Bioinform. 2018;19:700-12.
24. Ren W-L, Wen Y-J, Dunwell JM, Zhang Y-M. pKWmEB: integration of Kruskal–Wallis test with empirical Bayes under polygenic background control for multi-locus genome-wide association study. Heredity. 2018;120:208-18.
25. Tamba CL, Zhang YM. A fast mrMLM algorithm for multi-locus genome-wide association studies. bioRxiv. 2018; doi:10.1101/341784.
26. Zhang Y-M, Jia Z, Dunwell JM. The applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits. Front Plant Sci. 2019;10:100.
27. Cui Y, Zhang F, Zhou Y. The application of multi-locus GWAS for the detection of salt-tolerance loci in rice. Front Plant Sci. 2018;9:1464.
28. Xu Y, Yang T, Zhou Y, Yin S, Li P, Liu J, et al. Genome-wide association mapping of starch pasting properties in maize using single-locus and multi-locus models. Front Plant Sci. 2018;9:1311.
29. He L, Xiao J, Rashid KY, Yao Z, Li P, Jia G, et al. Genome-wide association studies for pasmo resistance in flax (*Linum usitatissimum* L.). Front Plant Sci. 2019;9:1982.
30. Peng Y, Liu H, Chen J, Shi T, Zhang C, Sun D, et al. Genome-wide association studies of free amino acid levels by six multi-locus models in bread wheat. Front Plant Sci. 2018;9:1196.
31. Li C, Fu Y, Sun R, Wang Y, Wang Q. Single-locus and multi-locus genome-wide association studies in the genetic dissection of fiber quality traits in upland cotton (*Gossypium hirsutum* L.). Front Plant Sci. 2018;9:1083.
32. Su J, Ma Q, Li M, Hao F, Wang C. Multi-locus genome-wide association studies of fiber-quality related traits in chinese early-maturity upland cotton. Front Plant Sci. 2018;9:1169.
33. Heffner E, Sorrells M, Jannink J-L. Genomic selection for crop improvement. Crop Sci. 2009;49:1-12.

34. Guo T, Li H, Yan J, Tang J, Li J, Zhang Z, et al. Performance prediction of F1 hybrids between recombinant inbred lines derived from two elite maize inbred lines. Theor Appl Genet. 2013;126:189-201.
35. Riedelsheimer C, Endelman JB, Stange M, Sorrells ME, Jannink JL, Melchinger AE. Genomic predictability of interconnected biparental maize populations. Genetics. 2013;194:493-503.
36. Xu Y, Xu C, Xu S. Prediction and association mapping of agronomic traits in maize using multiple omic data. Heredity (Edinb). 2017;119:174-84.
37. Zhang Z, Ober U, Erbe M, Zhang H, Gao N, He J, et al. Improving the accuracy of whole genome prediction for complex traits using the results of genome wide association studies. PLoS One. 2014;9:e93017.
38. Arruda MP, Lipka AE, Brown PJ, Krill AM, Thurber C, Brown-Guedira G, et al. Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (*Triticum aestivum* L.). Mol Breed. 2016;36:84.
39. Spindel JE, Begum H, Akdemir D, Collard B, Redona E, Jannink JL, et al. Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. Heredity (Edinb). 2016;116:395-408.
40. Barazesh S, McSteen P. Barren inflorescence1 functions in organogenesis during vegetative and inflorescence development in maize. Genetics. 2008;179:389-401.
41. Galli M, Khakhar A, Lu Z, Chen Z, Sen S, Joshi T, et al. The DNA binding landscape of the maize AUXIN RESPONSE FACTOR family. Nat Commun. 2018;9:4526.
42. Du Y, Liu L, Li M, Fang S, Shen X, Chu J, et al. UNBRANCHED3 regulates branching by modulating cytokinin biosynthesis and signaling in maize and rice. New Phytol. 2017;214:721-33.
43. Xiao Y, Liu H, Wu L, Warburton M, Yan J. Genome-wide association studies in maize: praise and stargaze. Mol Plant. 2017;10:359-74.
44. Bai F, Reinheimer R, Durantini D, Kellogg EA, Schmidt RJ. TCP transcription factor, BRANCH ANGLE DEFECTIVE 1 (BAD1), is required for normal tassel branch angle formation in maize. Proc Natl Acad Sci U S A. 2012;109:12225-30.
45. Vollbrecht E, Springer PS, Goh L, Buckler ES, Martienssen R. Architecture of floral branch systems in maize and related grasses. Nature. 2005;436:1119-26.
46. Phillips KA, Skirpan AL, Liu X, Christensen A, Slewinski TL, Hudson C, et al. vanishing tassel2 encodes a grass-specific tryptophan aminotransferase required for vegetative and reproductive development in maize. Plant Cell. 2011;23:550-66.
47. Desta ZA, Ortiz R. Genomic selection: genome-wide prediction in plant improvement. Trends Plant Sci. 2014;19:592-601.
48. Zhang Z, Erbe M, He J, Ober U, Gao N, Zhang H, et al. Accuracy of whole-genome prediction using a genetic architecture-enhanced variance-covariance matrix. G3 (Bethesda). 2015;5:615-27.

49. Bian Y, Holland JB. Enhancing genomic prediction with genome-wide association studies in multiparental maize populations. Heredity (Edinb). 2017;118:585-93.
50. Rice B, Lipka AE. Evaluation of RR-BLUP genomic selection models that incorporate peak genome-wide association study signals in maize and sorghum. Plant Genome. 2019;12:180052.
51. Li X, Li Y, Wu X, Bai N, Song Y, Zhang D, et al. Characteristic analysis of kernel related traits in important maize inbred lines of China. J Plant Genet Resour. 2016;17:1-5.
52. Li YX, Li C, Bradbury PJ, Liu X, Lu F, Romay CM, et al. Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population. Plant J. 2016;86:391-402.
53. Wu X, Li Y, Shi Y, Song Y, Zhang D, Li C, et al. Joint-linkage mapping and GWAS reveal extensive genetic loci that regulate male inflorescence size in maize. Plant Biotechnol J. 2016;14:1551-62.
54. Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. 1980;8:4321-5.
55. Ganal MW, Durstewitz G, Polley A, Berard A, Buckler ES, Charcosset A, et al. A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS One. 2011;6:e28334.
56. Wu X, Li Y, Shi Y, Song Y, Wang T, Huang Y, et al. Fine genetic characterization of elite maize germplasm using high-throughput SNP genotyping. Theor Appl Genet. 2014;127:621-31.
57. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007;23:2633-5.
58. Bolduc N, Yilmaz A, Mejia-Guerra MK, Morohashi K, O'Connor D, Grotewold E, et al. Unraveling the KNOTTED1 regulatory network in maize meristems. Genes Dev. 2012;26:1685-90.
59. Yi F, Gu W, Chen J, Song N, Gao X, Zhang X, et al. High temporal-resolution transcriptome landscape of early maize seed development. Plant Cell. 2019;31:974-92.
60. Bukowski R, Guo X, Lu Y, Zou C, He B, Rong Z, et al. Construction of the third-generation *Zea mays* haplotype map. Gigascience. 2017;7:1-12.
61. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559-75.
62. Endelman J. Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome. 2011;4:250-5.

Additional Files

Figure S1. Common QTNs codetected with different models and environments. a, The common QTNs codetected by different methods. The X-axis represents different environments. The Y-axis represents the corresponding number of significant QTNs detected by only one method and at least by two, three, four, five, six and seven methods. b, The common QTNs codetected across different locations. The Venn diagrams for each GWAS method were drawn based on the colocalization among four environments.

Figure S2. LD decay with physical distance in our association panel.

Figure S3. Whole-genome prediction of KRN in the inbred lines. The bars with different colors represent prediction accuracies for the KRN when using tagSNPs identified by different models. *P*-values were estimated based on the two-tailed Student's t-test. ***: *P*-value < 0.0001; NS: *P*-value > 0.05.

Table S1. Descriptive statistics of KRN from the subgroups in the association panel.

Table S2. The significant QTNs for KRN identified by MLM.

Table S3. The significant QTNs for KRN identified by six multi-locus methods.

Table S4. The number of QTNs codetected by different methods.

Table S5. The expression of the candidate genes in different maize tissues.

Table S6. Genes related to spike mutation in maize.

Table S7. The prediction accuracies of KRN for the inbred lines using the tagSNPs.

Table S8. The prediction accuracies for the KRN of the inbred lines using the tagSNPs representing the significant QTNs identified by different methods.

Table S9. Comparison of our GWAS results with QTNs detected in previous studies.

Figures

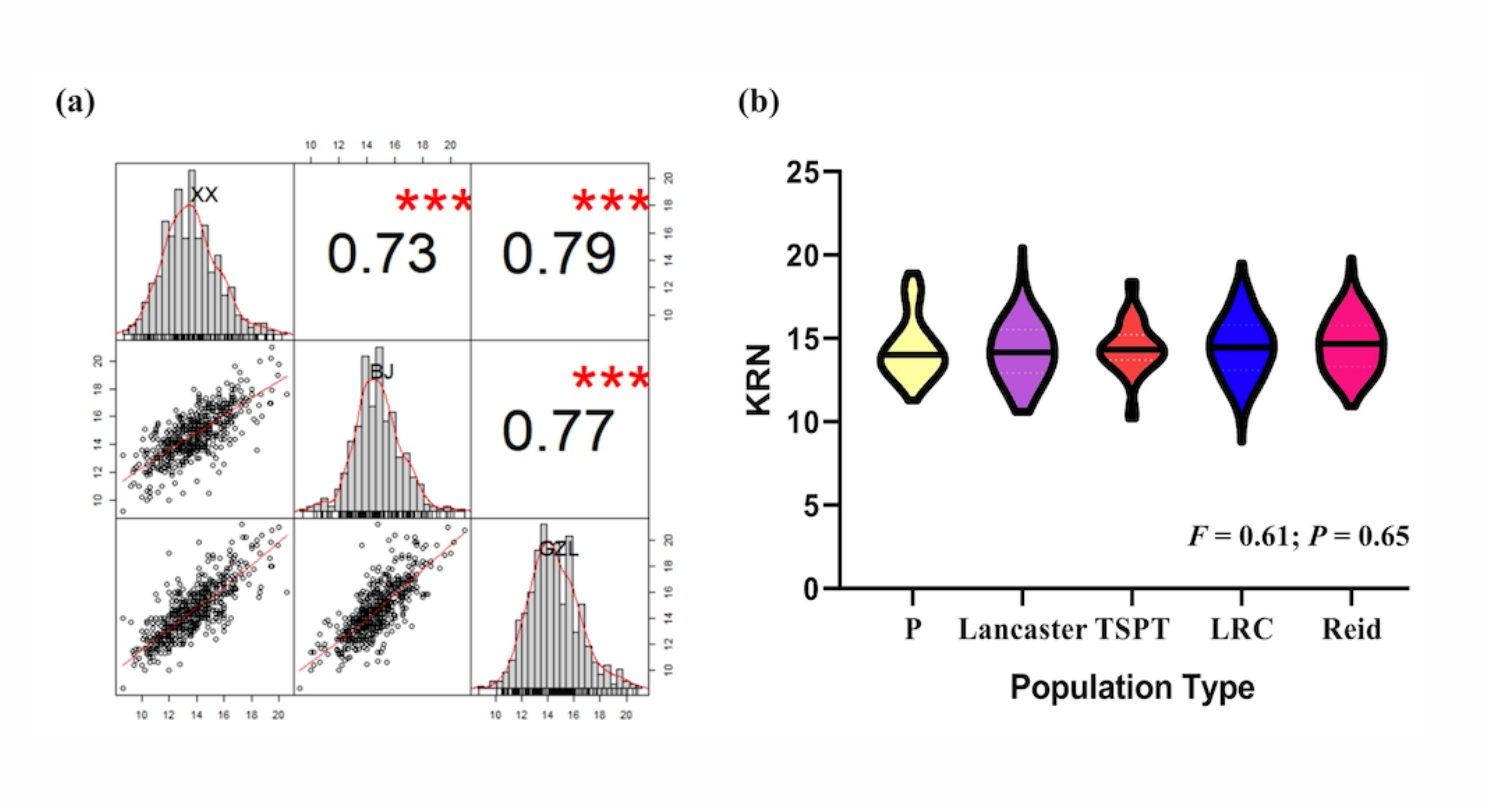

Figure 1

Phenotypic analysis. a, Correlation analysis of the KRN phenotype among XX, BJ and GZL. The frequency distribution diagrams of KRN under three environments were plotted, and the correlation coefficients between each environment were calculated. b, Violin plots of the KRN in the subgroups (P, Lancaster, TSPT, LRC, and Reid) in this association mapping panel.

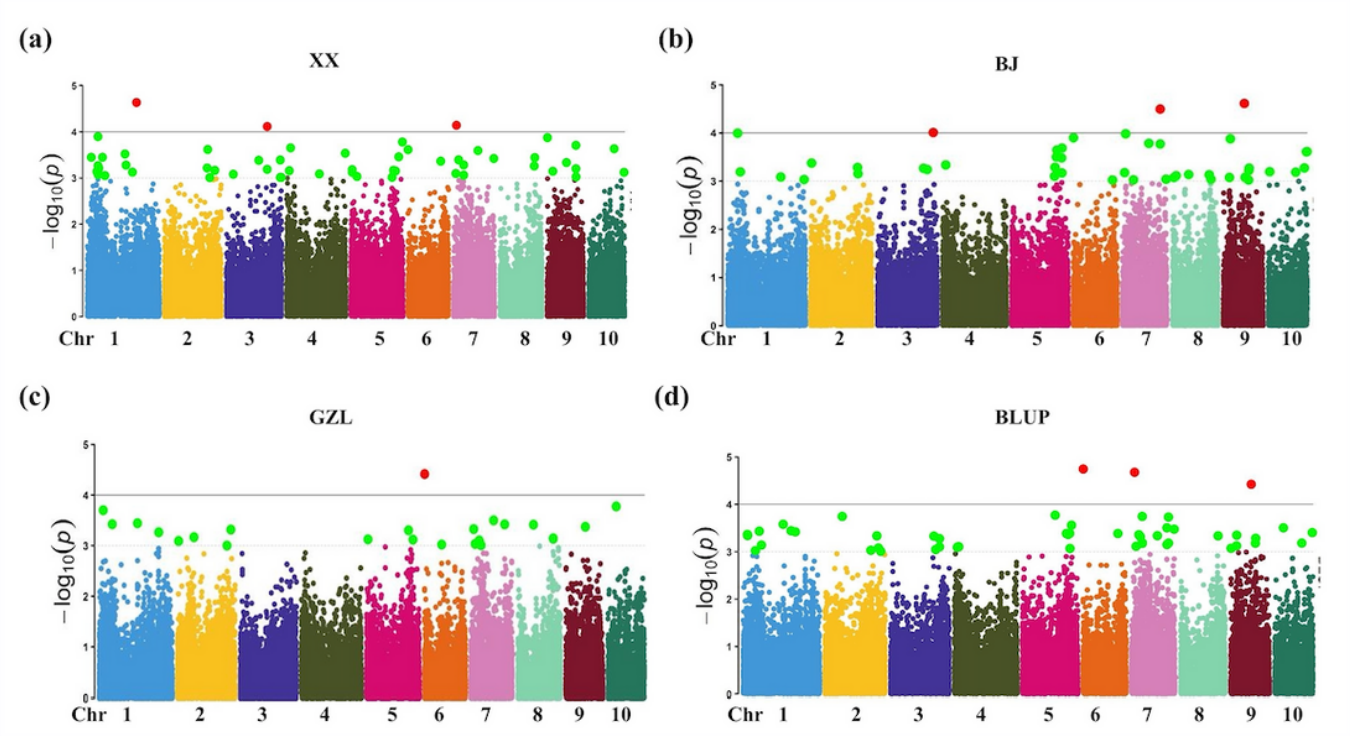

Figure 2

Genome-wide distribution of significant QTNs detected by the MLM model under four conditions. a, XinXiang (XX), Henan Province; b, Beijing (BJ); c, Gongzhuling (GZL), Jilin Province; d, BLUP across the three environments.

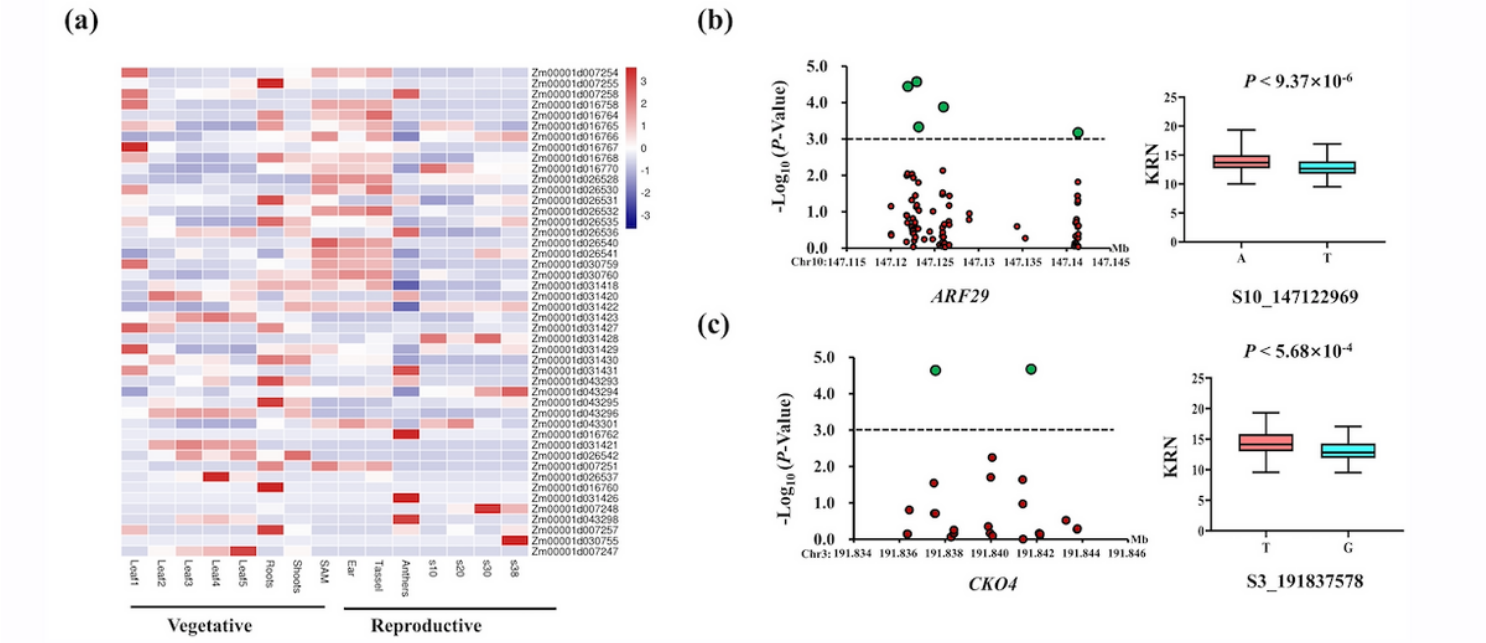

Figure 3

Candidate gene analysis for KRN. a, The expression heatmap of the genes located in the codetected regions. All expression data were collected from inbred B73. Leaf 1 means the leaf base; leaf 2 means the 1-cm leaf; leaf 3 means the 4-cm leaf; leaf 4 means the leaf tip; leaf 5 means the leaf at 20 DAP (days after pollination); S10 means the kernel at 10 DAP. b, ARF29 (Zm00001d026540) gene association mapping using the Ames 228 panel. c, CKO4 (Zm00001d043293) gene association mapping.

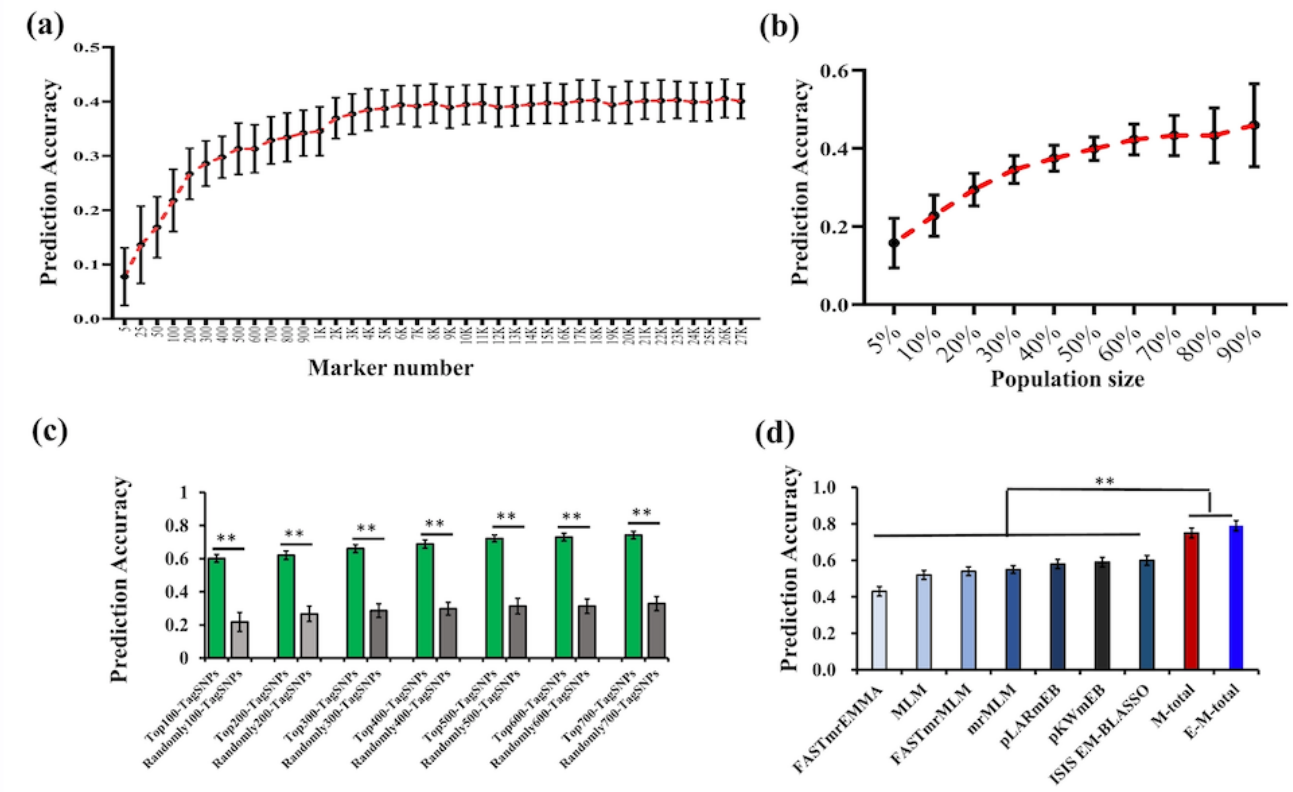

Figure 4

Whole-genome prediction of KRN in the inbred lines. a, The KRN prediction accuracy of randomly selected tagSNPs from 5 to 27,000 based on the BLUP value by using the rrBLUP model. b, The KRN prediction accuracy of different training population sizes. c, Comparison of the prediction accuracy between the top tagSNPs and random tagSNPs. **, P < 0.01. d, Comparison of the prediction accuracy of different tagSNPs identified by different models. **, P < 0.01.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.Tiff
- FigureS2.tif
- FigureS3.tif
- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx

- TableS5.xlsx
- TableS6.xlsx
- TableS7.xlsx
- TableS8.xlsx
- TableS9.xlsx