International Journal of
Molecular Sciences

MDPI

Article

CQD: A Rapid and Accurate Tool for Identifying High-Confidence QTLs and Candidate Genes from GWAS Summary Statistics

Xiangjian Gou [1], Shaohong Fu [2], Jiakun Li [3], Wanzhuo Gong [2], Qiaobo Wu [2], Yun Li [2], Jin Yang [2], Qiong Zou [2], Qin Yu [2], Lanrong Tao [2] and Haoran Shi [2,*]

[1] Hubei Hongshan Laboratory, Huazhong Agricultural University, Wuhan 430070, China
[2] Chengdu Academy of Agricultural and Forestry Sciences, Chengdu 611130, China
[3] Triticeae Research Institute, Sichuan Agricultural University, Chengdu 611130, China
* Correspondence: sicau3339@outlook.com

Abstract

Genome-wide association studies (GWAS) are a powerful approach for elucidating the genetic architecture of complex traits. The widespread application of GWAS across diverse crop species has generated vast amounts of GWAS summary statistics, creating a pressing need for effective tools to identify high-confidence quantitative trait loci (QTLs) from these data. Here, we present Candidate QTL Detector (CQD), a tool that enables the rapid and accurate identification of confident QTL regions from GWAS summary statistics. To evaluate the performance of CQD, we reanalyzed 64 phenotypic traits from a previously published maize diversity panel. CQD identified 108 high-confidence QTLs from 4179 significant GWAS signals, and these QTLs harbored several well-characterized genes, including *NS2* (a regulator of tassel branch number), *DRL1* (a key regulator of plant architecture), and *ZmACS6* (an ACC synthase involved in ethylene biosynthesis), demonstrating the reliability of the QTLs detected by CQD. Overall, CQD provides an efficient and flexible framework for extracting robust QTLs from GWAS summary statistics, thereby facilitating the genetic improvement of complex traits and advancing crop functional genomics.

Keywords: GWAS summary statistics; QTL; CQD; functional genomics

Check for updates

Academic Editor: Kwon-Kyoo Kang

Received: 18 January 2026
Revised: 11 February 2026
Accepted: 15 February 2026
Published: 25 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Understanding the genetic architecture of complex traits is fundamental to crop genetic improvement [1]. Since their introduction in the 1990s, genome-wide association studies (GWAS) have emerged as a powerful approach for dissecting the genetic basis of complex traits [2,3]. By associating genetic variants with phenotypic traits, GWAS enables the identification of genomic loci and candidate genes underlying trait variation. Advances in high-throughput sequencing technologies and the increasing availability of reference genome sequences have greatly facilitated the generation of genome-wide high-resolution single-nucleotide polymorphism (SNP) variation data, thereby promoting the widespread application of GWAS in crop genetics and breeding [4]. For example, GWAS has been extensively used to improve crop yield [5–7] and quality [8–10], identify genes conferring drought tolerance [11,12], and elucidate the genetic basis of climate adaptation [13,14]. Consequently, a large volume of GWAS summary statistics has been accumulated [15–17]. Meanwhile, the rapid development of omics technologies, including transcriptomics and metabolomics, has further expanded the availability of GWAS summary statistics related to

biochemical and molecular traits [18,19]. Therefore, accurately identifying high-confidence quantitative trait loci (QTLs) from large-scale GWAS summary statistics is crucial for advancing our understanding of the genetic architecture of complex traits in crop species.

To date, a variety of state-of-the-art statistical models have been developed to account for population structure and kinship in GWAS, thereby improving statistical power and enhancing the reliability of detected association signals, including mixed linear model [20], multiple loci mixed model [21], and fixed and random model circulating probability unification [22]. Several studies have systematically compared these models in terms of statistical power and control of false positives [1,23,24]. Nevertheless, despite these methodological advances, the large volume of GWAS summary statistics generated inevitably contains many spurious association signals. Therefore, there is a clear need for a lightweight downstream tool that can directly analyze GWAS summary statistics and further extract robust QTLs from these data.

In this study, to address the aforementioned challenges, we developed the Candidate QTL Detector (CQD), a lightweight tool implemented in Perl5. CQD enables the rapid and accurate identification of confident QTL regions from GWAS summary statistics, thereby facilitating the genetic dissection of complex traits and the discovery of candidate genes. Moreover, CQD is compatible with GWAS results generated by different software packages and allows users to flexibly define QTL identification criteria to meet diverse analytical needs. Using previously published phenotypic and genotypic data from a maize (*Zea mays* L.) diversity panel, we further validated the performance of CQD. The successful identification of several well-characterized genes supports the reliability of the QTLs detected by CQD.

2. Results

2.1. CQD Was Developed to Facilitate the Identification of QTLs and Candidate Genes from GWAS Summary Statistics

To facilitate the rapid identification of reliable QTLs from aggregated GWAS results, we developed CQD, a tool implemented in the Perl programming language. CQD does not rely on any external modules and can be executed on any platform with Perl5 installed. The options -t and -m allow CQD to flexibly parse GWAS summary statistics generated by different software packages (Table 1). The combination of the -*p* and -s options defines the significance threshold (*p*/s) for identifying significant SNPs, while the -d and -c options jointly determine the criteria used to define a QTL. To further identify potential candidate genes associated with each QTL, CQD provides the -a and -r options to specify the criteria for candidate gene identification. In addition, the -np and -ns options ensure consistency of chromosome identifiers between the GWAS summary statistics and the genome annotation file. Overall, CQD is a portable, flexible, and broadly applicable tool that enables the rapid and accurate identification of confident QTL regions from GWAS summary statistics.

Table 1. Parameter settings inside CQD.

Options	Description	Default Value
-i	provide a GWAS file	no default value
-t	column (snp_id,chrom,pos,*p*value) index in GWAS file	1,2,3,4
-m	delimiter (comma/tab/space) among column in GWAS file	comma
-o	set the output file name	candidate_qtls.txt
-*p*	*p* value for Bonferroni correction	0.05
-s	number of SNPs for Bonferroni correction	1,000,000
-d	max distance between adjacent significant SNPs within a QTL	100,000 bp
-c	min number of significant SNPs within a QTL	3
-a	a GFF3 file, if provided, will list candidate genes for each QTL	no default value
-r	up/downstream range of QTL to list candidate genes	100,000 bp
-np	add a prefix to the chromosome in GWAS file	no prefix
-ns	add a suffix to the chromosome in GWAS file	no suffix
-v	show the version information	no default value
-h	show the help information	no default value

2.2. Sixty-Four Phenotypic Traits and 9,146,558 SNPs from a Maize Diversity Panel Were Used to Evaluate the Performance of CQD

We employed phenotypic and genotypic data from previous publications to evaluate the performance of CQD [17]. A maize diversity panel consisting of 350 elite inbred lines, representing germplasm from multiple breeding eras in China (CN1960&70s, CN1980&90s, and CN2000&10s) and the United States (Public-US and Ex-PVP), was included in the current analysis. Specifically, we analyzed 64 phenotypic traits, which primarily characterized maize plant architecture and flowering time, such as plant height (PH), lower leaf angle (LAL), and days to anthesis (DTA). Correlation-based cluster analysis indicated that phenotypes of the same category clustered together, confirming the reliability of the phenotypic data (Figure 1A). By performing variant calling on resequencing data from previous publications, we obtained 9,146,558 high-quality SNPs with a minor allele frequency greater than 5% and a missing rate lower than 20%. Principal component analysis revealed clear population structure stratification between Chinese and US inbred lines (Figure 1B,C). Furthermore, genome-wide linkage disequilibrium (LD) decay analysis indicated that when the r^2 decayed to 0.2, the LD decay distance reached approximately 90 kb (Figure 1D). This value was subsequently used as a criterion for QTL identification in CQD.

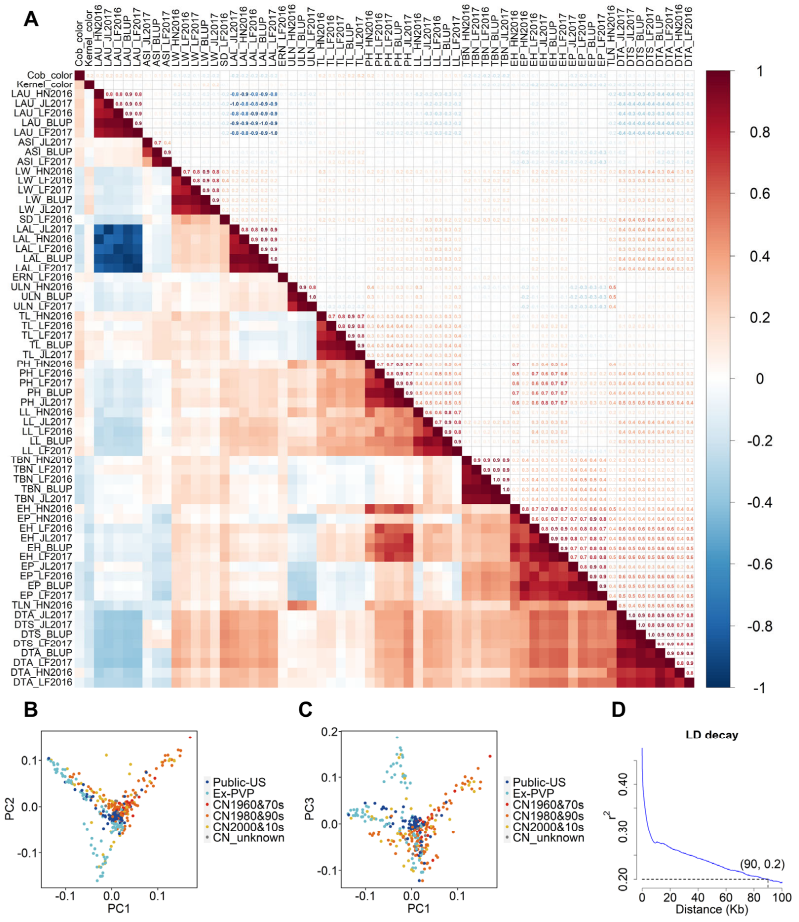


Figure 1. Phenotypic and genotypic analysis of the maize diversity panel. (**A**) Correlation analysis of 64 phenotypic traits. Each item in the figure represents a phenotype, and the nomenclature follows that used in the previous publication. Except for Cob_color and Kernel_color, all other phenotype names are formatted as an abbreviation of the trait name followed by an underscore and the environmental code. The correlations were clustered using the hierarchical clustering method. (**B**) Principal component analysis plot of principal component 1 (PC1) and principal component 2 (PC2). Each color represents a subpopulation. (**C**) Principal component analysis plot of PC1 and principal component 3 (PC3). (**D**) Genome-wide linkage disequilibrium decay analysis. r^2, squared Pearson correlation coefficient.

2.3. CQD Successfully Identified 108 High-Confidence QTLs from GWAS Summary Statistics

We first performed GWAS on 64 phenotypic traits. In total, 4179 significant SNPs were detected, which were associated with 53 traits. Notably, a large number of significant SNPs were associated with the traits Cob_color (2312 SNPs) and Kernel_color (1276 SNPs), suggesting that a substantial proportion of spurious associations may be present among these signals. Therefore, we further applied CQD to the corresponding GWAS summary statistics to identify reliable QTLs by setting the parameters (the -d and -c options; see Table 1) such that each QTL contained at least three significant SNPs and the distance between adjacent significant SNPs did not exceed 90 kb. As a result, CQD identified a total of 108 high-confidence QTLs, which were significantly associated with 28 phenotypic traits (Table S1).

Taking kernel color as an example, the original GWAS results identified several well-characterized quality-related genes in maize, including *PSY1* (Zm00001d036345) [25] and *Zmccr3* (Zm00001d015956) [26]. However, the original GWAS results still contained a large number of individual signals, which substantially hindered the identification of reliable candidate genes (Figure 2A). Application of CQD removed 126 low-confidence individual signals, thereby highlighting its ability to effectively filter GWAS results and retain high-confidence association signals (Figure 2B).

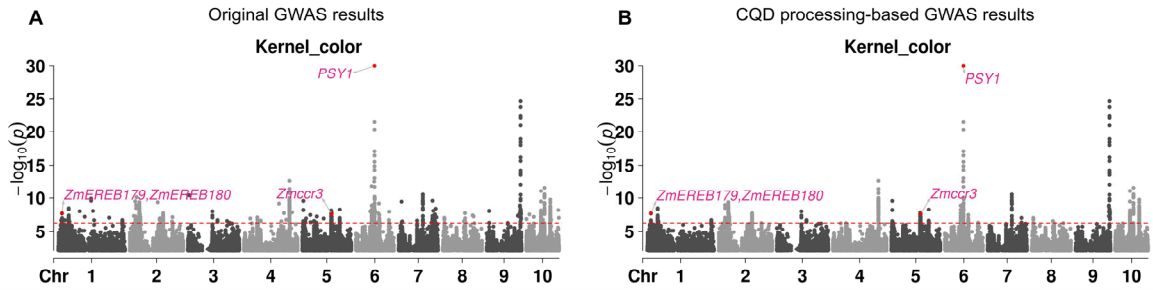


Figure 2. GWAS results of kernel color. (**A**) The original GWAS results output by the GEMMA software. (**B**) GWAS results after CQD processing, with low-confidence signals removed. Red lines indicate the significance threshold set by Bonferroni correction. The *p*-values of the eight SNPs on chromosome 6 were far below 1×10^{-30}; therefore, they were uniformly set to 1×10^{-30}.

To further demonstrate the reliability of the QTLs identified by CQD, we examined the association mapping results for tassel branch number (TBN), relative ear height (EP), and ear row number (ERN). The genomic inflation factors (λ) for the three phenotypic traits ranged from 0.994 to 1.033, indicating that the GWAS models effectively controlled potential confounding effects and produced well-calibrated association statistics (Figure 3). For the TBN_JL2017 phenotype, three QTLs (QTL1, QTL2, and QTL3) were identified, each containing 4, 3, and 5 significant SNPs, respectively (Figure 3A). Among them, QTL2, located on chromosome 4, harbored three candidate genes. One of these genes, *NS2* (Zm00001d052598), is located 29 kb downstream of QTL2 and has been cloned and functionally validated to regulate tassel branch number in maize [27] (Figure 3B,C). For the EP_BLUP phenotype, four QTLs (QTL4, QTL5, QTL6, and QTL7) showed strong associations (Figure 3A). QTL6, located on chromosome 1, contained 50 significant SNPs. A known gene, *DRL1* (Zm00001d028216), was identified 72 kb downstream of QTL6 (Figure 3D,E). *DRL1* has been previously characterized for its role in regulating plant architecture in maize [28]. Similarly, for the ERN_LF2016 phenotype, one QTL (QTL8) was detected (Figure 3A). QTL8 is located on chromosome 1 and contains 45 significant SNPs. We identified *ZmACS6* (Zm00001d033862), located 25 kb downstream of QTL8 (Figure 3F,G). *ZmACS6* encodes an ACC synthase, a key enzyme catalyzing the first step of ethylene biosynthesis. Previous studies have demonstrated the important role of *ZmACS6* in regulating drought tolerance and grain yield in maize [29,30]. Overall, the

identification of these well-characterized genes provides strong support for the reliability of the QTLs identified by CQD.

A

QTL_id	Chr	Lead_SNP_id	Lead_SNP_pos	Lead_SNP_p.value	QTL_start	QTL_end	SNP_num_in_QTL
QTL1	3	S3_174410423	174410423	2.28E-07	174371618	174410704	4
QTL2	4	S4_194285721	194285721	3.31E-07	194284722	194285721	3
QTL3	5	S5_204942824	204942824	6.83E-11	204942824	204943415	5
QTL4	1	S1_25779655	25779655	5.77E-09	25775085	25799504	5
QTL5	1	S1_26086469	26086469	7.50E-08	26029608	26202983	10
QTL6	1	S1_26362219	26362219	7.97E-09	26334751	26699348	50
QTL7	3	S3_190535336	190535336	2.09E-07	190535336	190538518	6
QTL8	1	S1_276634571	276634571	2.21E-08	276547447	276656650	45

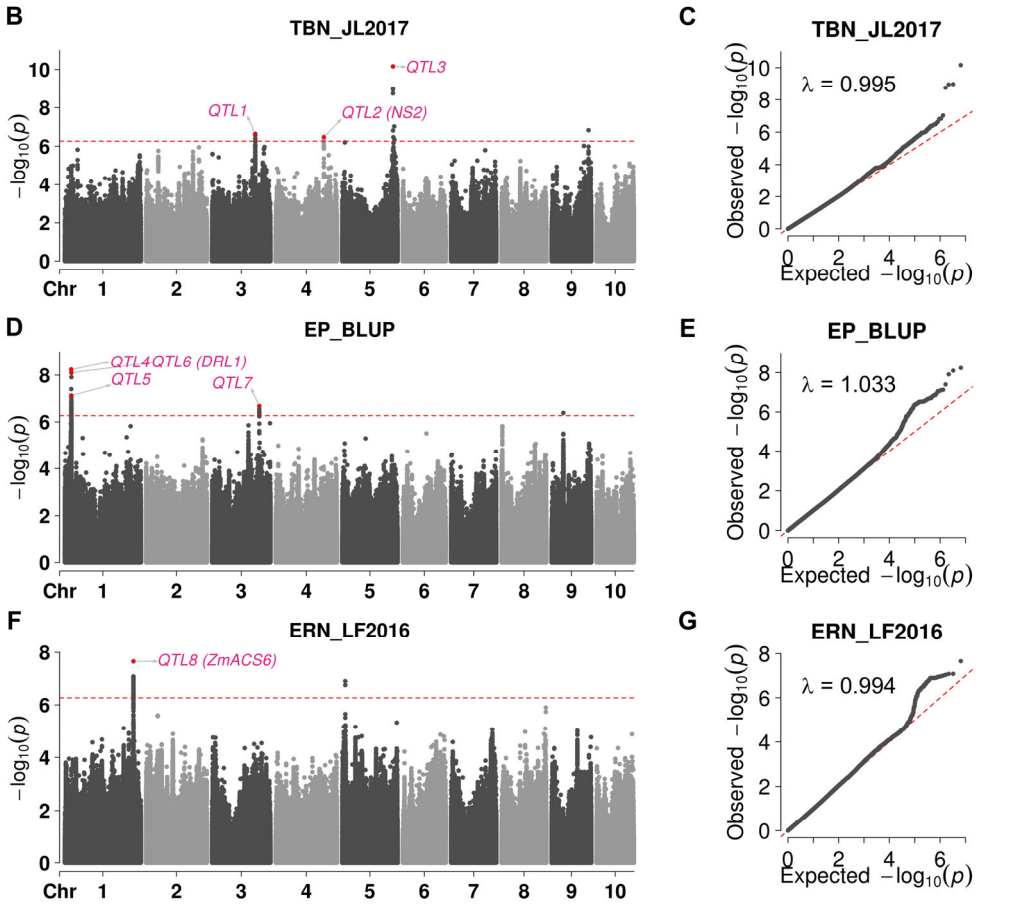


Figure 3. GWAS results of tassel branch number (TBN), relative ear height (EP), and ear row number (ERN). (**A**) A total of eight QTLs identified by CQD. (**B**) Manhattan plot for TBN_JL2017. (**C**) Q-Q plot for TBN_JL2017. (**D**) Manhattan plot for EP_BLUP. (**E**) Q-Q plot for EP_BLUP. (**F**) Manhattan plot for ERN_LF2016. (**G**) Q-Q plot for ERN_LF2016. Red lines in Manhattan plot indicate the significance threshold by Bonferroni correction. The genomic inflation factor (λ) is indicated for each trait-specific Q-Q plot, and λ values close to 1 suggest effective control of population structure and relatedness in the GWAS analysis.

3. Discussion

Accurately identifying causal variants and cloning key genes are essential for improving crop yield, stress resistance, and climate adaptability, thereby helping to mitigate the challenges posed by population growth and climate change [31,32]. However, the large number of spurious associations present in the rapidly expanding body of GWAS summary statistics severely hampers progress in gene cloning and functional genomics. Although

state-of-the-art statistical models have substantially improved the statistical power of GWAS [33], the exponential accumulation of data inevitably leads to an increased burden of false-positive signals [34,35]. In this study, we developed CQD to accurately identify robust QTL regions from large-scale GWAS summary statistics, thereby minimizing the impact of spurious associations and enabling the reliable discovery of causal genes underlying trait variation. Moreover, CQD can be readily integrated into the downstream analysis pipelines of diverse GWAS software, making it a highly versatile tool.

Notably, CQD does not perform statistical hypothesis testing and therefore does not modify the false discovery rate control implemented in GWAS. Instead, CQD consolidates significant SNPs into QTL regions to improve interpretability and reduce signal fragmentation. CQD is designed to complement existing LD-based SNP clustering approaches by providing a region-based consolidation strategy for significant GWAS signals and candidate gene prioritization. The recovery of previously cloned genes further supports that CQD does not increase the false-positive association burden.

In addition to the approaches evaluated here, other computational strategies have been proposed to enhance the interpretation of GWAS and related association data. For example, network-based gene prioritization methods have been developed to integrate GWAS signals with functional interaction networks, thereby improving candidate gene identification from association signals [36]. Similarly, tools originating in biomedical association analysis have addressed large-scale summary statistics processing and modular gene prediction using hybrid clustering approaches [37]. These studies share conceptual parallels with the CQD strategy of clustering significant SNPs and prioritizing candidate genes within consolidated QTL regions and reinforce the value of developing downstream analytical frameworks to extract biologically meaningful signals from dense GWAS datasets.

Tool assessment is necessary to verify practical applicability. In this study, the performance of CQD was assessed using a maize diversity panel. CQD successfully condensed thousands of phenotype-associated SNPs into 108 high-confidence QTLs. Furthermore, several well-characterized genes, including *PSY1*, *Zmccr3*, *NS2*, *DRL1*, and *ZmACS6*, were recovered within these QTLs, supporting the reliability of the detected loci. Together, these results demonstrate that CQD can effectively refine GWAS results by filtering low-confidence signals and retaining only robust and biologically meaningful associations.

Although CQD was validated using maize in this study, the algorithm relies on general genomic features, including spatial clustering of significant SNPs, genome annotation, and linkage disequilibrium or physical distance parameters. Therefore, CQD can be readily applied to other crop species where GWAS summary statistics and genome annotations are available. Moreover, CQD is not restricted to specific trait categories and can also be applied to GWAS analyses of root system architecture and other belowground traits, which play critical roles in nutrient acquisition, stress adaptation, and crop productivity. We hope that future studies will integrate CQD into GWAS pipelines, thereby promoting crop genetic improvement and helping to alleviate food security challenges arising from rapid population growth and worsening environmental conditions.

4. Materials and Methods

4.1. Plant Materials and Phenotypic Trait Acquisition

A maize diversity panel was employed to evaluate the performance of CQD. This panel consisted of 350 elite inbred lines representing germplasm from multiple breeding eras in China and the United States. Phenotypic data for this panel were obtained from previous publications, in which multiple traits were measured across four environments (LF2016, LF2017, HN2016, and JL2017). For example, JL2017 represents a field trial conducted in Jilin Province, China, in 2017. Detailed descriptions of the remaining environments

are available in the original publication. All field trials were conducted under standard agronomic management without intentional abiotic or biotic stress treatments. These traits included ear row number (ERN), tassel branch number (TBN), tassel length (TL), lower leaf angle (LAL), upper leaf angle (LAU), leaf length (LL), leaf width (LW), plant height (PH), ear height (EH), relative ear height (EP), total leaf number (TLN), upper leaf number (ULN), stem diameter (SD), days to anthesis (DTA), days to silking (DTS), anthesis-silking interval (ASI), cob color, and kernel color. In total, 64 phenotypic traits were compiled from previous studies and used in the present analysis [17].

4.2. Variant Calling and Genotypic Data Acquisition

Genome sequencing reads for all maize lines were obtained from previous publications under the accession number PRJNA609577 [17]. Variant calling was performed following the GATK v4.1.2.0 best-practice workflow [38]. Specifically, raw sequencing reads were subjected to quality control using NGSQC v2.3.3 [39] with the parameters "-l 70 -s 20". Clean reads were then aligned to the B73 RefGen_v4 maize reference genome [40] using BWA v0.7.13 [41] with default settings (MEM algorithm). The aligned reads were subsequently sorted using SortSam, and PCR duplicates were marked with MarkDuplicates. HaplotypeCaller was then applied to generate a GVCF file for each sample. Next, GVCF files from the 350 samples were merged using GenomicsDBImport, followed by joint genotyping with GenotypeGVCFs. SNPs were further filtered using VariantFiltration with the following criteria: QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < −12.5 || ReadPosRankSum < −8.0 || DP < 8.0 || QUAL < 20. Finally, non-biallelic variants and variants with a minor allele frequency <5% or a missing rate >20% were removed, resulting in a final set of 9,146,558 high-quality SNPs retained for subsequent analyses.

4.3. Population Structure and Linkage Disequilibrium Decay Analysis

Principal component analysis was performed on the filtered SNP set using PLINK v1.9 [42] to characterize the population structure of the maize diversity panel. The first three principal components were used to visualize population stratification. Genome-wide linkage disequilibrium decay was estimated using PopLDdecay v3.41 [43]. Linkage disequilibrium was quantified using the squared Pearson correlation coefficient (r^2), and the physical distance at which r^2 decayed to 0.2 was subsequently used as a criterion for QTL identification and candidate gene determination.

4.4. Genome-Wide Association Study

Using the 9,146,558 high-quality SNPs as the genotype matrix, GWAS of 64 phenotypic traits were conducted using GEMMA v0.98.5 [44] under a linear mixed model [20]. The first five principal components and an intercept term consisting entirely of ones were included as covariates, and the kinship matrix was calculated using GEMMA with default parameters. The effective number (EN = 1,850,299) of independent SNPs was estimated using GEC v0.2 [45], and the significance threshold ($1/EN = 5.4 \times 10^{-7}$) was determined based on the modified Bonferroni correction. Manhattan plots were generated using the CMplot package [46]. To evaluate the calibration of the GWAS test statistics, genomic inflation factors (λ) were calculated as the ratio of the median observed chi-square statistic to the expected median chi-square statistic under the null hypothesis.

4.5. Quantitative Trait Locus Identification

The CQD tool (https://github.com/GouXiangJian/CQD, accessed on 14 February 2026) developed in this study was applied to identify QTLs from GWAS summary statistics. Based on the previously determined genome-wide linkage disequilibrium decay distance and the effective number of independent SNPs, the key parameters for running CQD were set as

follows: *-p* 1 -s 1,850,299 -d 90,000 -c 3 -r 90,000 (Table 1). In addition, CQD generated a list of candidate genes flanking each QTL. These candidate genes were extracted from the B73 RefGen_v4 maize reference genome annotations obtained from the MaizeGDB database.

5. Conclusions

In this study, we developed a portable and flexible tool, CQD, that can be applied downstream of diverse GWAS pipelines to identify reliable QTLs and candidate genes from GWAS summary statistics. CQD provides a region-based framework for consolidating GWAS signals, improving result interpretability, and prioritizing biologically meaningful candidate loci. The performance of CQD was validated using 64 phenotypic traits and 9,146,558 SNPs from a maize diversity panel. Although CQD was evaluated using maize data, the method is species-independent and can be readily applied to other crop species and diverse trait categories where GWAS summary statistics are available. We anticipate that CQD will accelerate research in crop functional genomics and facilitate the discovery and cloning of key genes in crop species.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms27052124/s1.

Author Contributions: Methodology, X.G.; data curation, X.G., S.F., W.G., Q.W. and Y.L.; formal analysis, X.G., J.L., J.Y., Q.Z., Q.Y. and L.T.; writing—original draft, X.G.; conceptualization, H.S.; funding acquisition, H.S.; writing—review and editing, H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was financially supported by the Sichuan Innovation Team Program (SCCXTD-2026-3) and the Major Science and Technology Program in Chengdu (2024-YF09-00036-SN).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in the article.

Acknowledgments: The authors thank the editors and reviewers for their constructive and helpful comments on the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Tibbs Cortes, L.; Zhang, Z.; Yu, J. Status and prospects of genome-wide association studies in plants. *Plant Genome* **2021**, *14*, e20077. [CrossRef]
2. Risch, N.; Merikangas, K. The future of genetic studies of complex human diseases. *Science* **1996**, *273*, 1516–1517. [CrossRef] [PubMed]
3. Liu, H.J.; Yan, J. Crop genome-wide association study: A harvest of biological relevance. *Plant J.* **2019**, *97*, 8–18. [CrossRef] [PubMed]
4. Purugganan, M.D.; Jackson, S.A. Advancing crop genomics from lab to field. *Nat. Genet.* **2021**, *53*, 595–601. [CrossRef] [PubMed]
5. Luo, Y.; Zhang, M.; Liu, Y.; Liu, J.; Li, W.; Chen, G.; Peng, Y.; Jin, M.; Wei, W.; Jian, L. Genetic variation in YIGE1 contributes to ear length and grain yield in maize. *New Phytol.* **2022**, *234*, 513–526. [CrossRef]
6. Zeng, T.; Meng, Z.; Yue, R.; Lu, S.; Li, W.; Li, W.; Meng, H.; Sun, Q. Genome wide association analysis for yield related traits in maize. *BMC Plant Biol.* **2022**, *22*, 449. [CrossRef]
7. Zhong, H.; Liu, S.; Sun, T.; Kong, W.; Deng, X.; Peng, Z.; Li, Y. Multi-locus genome-wide association studies for five yield-related traits in rice. *BMC Plant Biol.* **2021**, *21*, 364. [CrossRef]
8. Li, K.; Yu, Y.; Yan, S.; Li, W.; Xu, J.; Li, G.; Li, W.; Liu, J.; Qi, X.; Huang, W. Genetic basis of flavor complexity in sweet corn. *Nat. Genet.* **2025**, *57*, 2842–2851. [CrossRef]
9. Lu, X.; Lei, Y.; Xu, Z.; Cheng, Z.; Liu, M.; Tai, Y.; Han, X.; Hao, Z.; Li, M.; Zhang, D. Natural variations in the promoter of ZmDeSI2 encoding a deSUMOylating isopeptidase controls kernel methionine content in maize. *Mol. Plant* **2025**, *18*, 872–891. [CrossRef]

10. Lian, T.; Guo, W.; Wang, Y.; Wen, W.; Wang, W.; Jiang, L.; Liang, Q.; Liu, J.a.; Liu, H.; Xue, Y. An Allele of Glutamate Formiminotransferase Triggers 5-Methyl-Tetrahydrofolate-to-MeFox Conversion and Facilitates Folate Biofortification in Maize. *Adv. Sci.* **2025**, *12*, e15082. [CrossRef]
11. Wang, X.; Wang, H.; Liu, S.; Ferjani, A.; Li, J.; Yan, J.; Yang, X.; Qin, F. Genetic variation in ZmVPP1 contributes to drought tolerance in maize seedlings. *Nat. Genet.* **2016**, *48*, 1233–1241. [CrossRef] [PubMed]
12. Lian, Y.; Yang, S.; Tian, T.; Yang, Z.; Liu, S.; Fu, X.; Liu, C.; Zhu, T.; Wang, Y.; Bai, Y. Natural variation in ZmDapF1 enhances maize drought resilience. *Nat. Plants* **2025**, *11*, 2381–2394. [CrossRef] [PubMed]
13. Gou, X.; Shao, Y.; Wang, X.; Shi, H.; Yu, J.; Li, X.; Guo, T. Evolutionary patterns of DNA base composition at polymorphic sites highlight the role of the environment in shaping barley and rice genomes. *Plant Genome* **2024**, *17*, e20456. [CrossRef] [PubMed]
14. Zhang, F.; Long, R.; Ma, Z.; Xiao, H.; Xu, X.; Liu, Z.; Wei, C.; Wang, Y.; Peng, Y.; Yang, X. Evolutionary genomics of climatic adaptation and resilience to climate change in alfalfa. *Mol. Plant* **2024**, *17*, 867–883. [CrossRef]
15. Huang, X.; Wei, X.; Sang, T.; Zhao, Q.; Feng, Q.; Zhao, Y.; Li, C.; Zhu, C.; Lu, T.; Zhang, Z. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **2010**, *42*, 961–967. [CrossRef]
16. Wang, Q.J.; Yuan, Y.; Liao, Z.; Jiang, Y.; Wang, Q.; Zhang, L.; Gao, S.; Wu, F.; Li, M.; Xie, W. Genome-wide association study of 13 traits in maize seedlings under low phosphorus stress. *Plant Genome* **2019**, *12*, 190039. [CrossRef]
17. Wang, B.; Lin, Z.; Li, X.; Zhao, Y.; Zhao, B.; Wu, G.; Ma, X.; Wang, H.; Xie, Y.; Li, Q. Genome-wide selection and genetic improvement during modern maize breeding. *Nat. Genet.* **2020**, *52*, 565–571. [CrossRef]
18. Zhang, F.; Wu, J.; Sade, N.; Wu, S.; Egbaria, A.; Fernie, A.R.; Yan, J.; Qin, F.; Chen, W.; Brotman, Y. Genomic basis underlying the metabolome-mediated drought adaptation of maize. *Genome Biol.* **2021**, *22*, 260. [CrossRef]
19. Sun, G.; Yu, H.; Wang, P.; Lopez-Guerrero, M.; Mural, R.V.; Mizero, O.N.; Grzybowski, M.; Song, B.; Van Dijk, K.; Schachtman, D.P. A role for heritable transcriptomic variation in maize adaptation to temperate environments. *Genome Biol.* **2023**, *24*, 55. [CrossRef]
20. Yu, J.; Pressoir, G.; Briggs, W.H.; Vroh Bi, I.; Yamasaki, M.; Doebley, J.F.; McMullen, M.D.; Gaut, B.S.; Nielsen, D.M.; Holland, J.B. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **2006**, *38*, 203–208. [CrossRef]
21. Segura, V.; Vilhjálmsson, B.J.; Platt, A.; Korte, A.; Seren, Ü.; Long, Q.; Nordborg, M. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.* **2012**, *44*, 825–830. [CrossRef] [PubMed]
22. Liu, X.; Huang, M.; Fan, B.; Buckler, E.S.; Zhang, Z. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* **2016**, *12*, e1005767. Erratum in *PLoS Genet.* **2016**, *12*, e1005957. [CrossRef] [PubMed]
23. Kaler, A.S.; Gillman, J.D.; Beissinger, T.; Purcell, L.C. Comparing different statistical models and multiple testing corrections for association mapping in soybean and maize. *Front. Plant Sci.* **2020**, *10*, 1794. [CrossRef] [PubMed]
24. Saini, D.K.; Chopra, Y.; Singh, J.; Sandhu, K.S.; Kumar, A.; Bazzer, S.; Srivastava, P. Comprehensive evaluation of mapping complex traits in wheat using genome-wide association studies. *Mol. Breed.* **2022**, *42*, 1. [CrossRef]
25. Buckner, B.; Kelson, T.L.; Robertson, D.S. Cloning of the y1 Locus of Maize, a Gene Involved in the Biosynthesis of Carotenoids. *Plant Cell* **1990**, *2*, 867–876. [CrossRef]
26. Feng, L.; Zhou, M.; Tao, A.; Ma, X.; Wang, N.; Zhang, H.; Duan, H.; Tao, Y. Map-based cloning of Zmccr3 and its network construction and validation for regulating maize seed germination. *Theor. Appl. Genet.* **2025**, *138*, 105. [CrossRef]
27. Wang, X.; Li, J.; Han, L.; Liang, C.; Li, J.; Shang, X.; Miao, X.; Luo, Z.; Zhu, W.; Li, Z. QTG-Miner aids rapid dissection of the genetic base of tassel branch number in maize. *Nat. Commun.* **2023**, *14*, 5232. [CrossRef]
28. Strable, J.; Wallace, J.G.; Unger-Wallace, E.; Briggs, S.; Bradbury, P.J.; Buckler, E.S.; Vollbrecht, E. Maize YABBY genes drooping leaf1 and drooping leaf2 regulate plant architecture. *Plant Cell* **2017**, *29*, 1622–1641. [CrossRef]
29. Habben, J.E.; Bao, X.; Bate, N.J.; DeBruin, J.L.; Dolan, D.; Hasegawa, D.; Helentjaris, T.G.; Lafitte, R.H.; Lovan, N.; Mo, H. Transgenic alteration of ethylene biosynthesis increases grain yield in maize under field drought-stress conditions. *Plant Biotechnol. J.* **2014**, *12*, 685–693. [CrossRef]
30. Young, T.E.; Meeley, R.B.; Gallie, D.R. ACC synthase expression regulates leaf performance and drought tolerance in maize. *Plant J.* **2004**, *40*, 813–825. [CrossRef]
31. Ray, D.K.; Mueller, N.D.; West, P.C.; Foley, J.A. Yield trends are insufficient to double global crop production by 2050. *PLoS ONE* **2013**, *8*, e66428. [CrossRef]
32. Tigchelaar, M.; Battisti, D.S.; Naylor, R.L.; Ray, D.K. Future warming increases probability of globally synchronized maize production shocks. *Proc. Natl. Acad. Sci.* **2018**, *115*, 6644–6649. [CrossRef]
33. Shin, J.; Lee, C. A mixed model reduces spurious genetic associations produced by population stratification in genome-wide association studies. *Genomics* **2015**, *105*, 191–196. [CrossRef] [PubMed]
34. Kremling, K.A.; Chen, S.-Y.; Su, M.-H.; Lepak, N.K.; Romay, M.C.; Swarts, K.L.; Lu, F.; Lorant, A.; Bradbury, P.J.; Buckler, E.S. Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. *Nature* **2018**, *555*, 520–523. [CrossRef] [PubMed]

35. Yang, S.; Wang, Y.; Huang, Q.; Wang, M.; Wang, S.; Fu, X.; Zhu, C.; Cheng, J.; Liu, S.; Yang, Z. A pangenome of maize provides genetic insights into drought resistance. *Nat. Genet.* **2025**, *57*, 2831–2841. [CrossRef] [PubMed]
36. He, B.; Wang, K.; Xiang, J.; Bing, P.; Tang, M.; Tian, G.; Guo, C.; Xu, M.; Yang, J. DGHNE: Network enhancement-based method in identifying disease-causing genes through a heterogeneous biomedical network. *Brief. Bioinform.* **2022**, *23*, bbac405. [CrossRef]
37. Xiang, J.; Zhang, J.; Zhao, Y.; Wu, F.-X.; Li, M. Biomedical data, computational methods and tools for evaluating disease–disease associations. *Brief. Bioinform.* **2022**, *23*, bbac006. [CrossRef]
38. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [CrossRef]
39. Patel, R.K.; Jain, M. NGS QC Toolkit: A toolkit for quality control of next generation sequencing data. *PLoS ONE* **2012**, *7*, e30619. [CrossRef]
40. Jiao, Y.; Peluso, P.; Shi, J.; Liang, T.; Stitzer, M.C.; Wang, B.; Campbell, M.S.; Stein, J.C.; Wei, X.; Chin, C.-S. Improved maize reference genome with single-molecule technologies. *Nature* **2017**, *546*, 524–527. [CrossRef]
41. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef]
42. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.; Bender, D.; Maller, J.; Sklar, P.; De Bakker, P.I.; Daly, M.J. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [CrossRef]
43. Zhang, C.; Dong, S.-S.; Xu, J.-Y.; He, W.-M.; Yang, T.-L. PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **2019**, *35*, 1786–1788. [CrossRef]
44. Zhou, X.; Stephens, M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* **2014**, *11*, 407–409. [CrossRef]
45. Li, M.-X.; Yeung, J.M.; Cherny, S.S.; Sham, P.C. Evaluating the effective numbers of independent tests and significant *p*-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Human. Genet.* **2012**, *131*, 747–756. [CrossRef]
46. Yin, L.; Zhang, H.; Tang, Z.; Xu, J.; Yin, D.; Zhang, Z.; Yuan, X.; Zhu, M.; Zhao, S.; Li, X. rMVP: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genom. Proteom. Bioinform.* **2021**, *19*, 619–628. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.