Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrated GWAS and linkage mapping in a multi-parent population prioritizes candidate genes for kernel number per row in maize

Yuxiang Luo
Yunnan University
Fuyan Jiang
Yunnan Academy of Agricultural Sciences
Haoran Lyu
Yunnan University
Ranjan K. Shaw
Yunnan Academy of Agricultural Sciences
Xingfu Yin
Yunnan Academy of Agricultural Sciences
Guohong Wang
Yunnan Academy of Agricultural Sciences
Xingming Fan
xingmingfan@163.com
Yunnan Academy of Agricultural Sciences

Research Article

Keywords: Maize, KNPR, QTL, GWAS, Candidate gene, Integrated mapping

Posted Date: July 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10069220/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Genomics on October 8th, 2026. See the published version at https://doi.org/10.1186/s12864-026-13436-w.

Integrated GWAS and linkage mapping in a multi-parent population prioritizes candidate genes for kernel number per row in maize

Yuxiang Luo[1,2], Fuyan Jiang[2], Haoran Lyu[1,2], Ranjan K. Shaw[2], Xingfu Yin[2], Guohong Wang[2], Xingming Fan[2*]

1 College of Agriculture, Yunnan University, Kunming 650500, China;

2 Institute of Food Crops, Yunnan Academy of Agricultural Sciences, Kunming 650205, China;

*** Correspondence: xingmingfan@163.com (X.F.) Orcid-ID: 0009-0009-4742-9246**

Contributing authors: yuxiangluolyx@163.com (Y.L.); jiangfuyansxx@126.com (F.J.); haoranLv2025@163.com (H.L.); ranjanshaw@gmail.com (R.K.S.); xingfuyin626@163.com (X.Y.); 987543168@qq.com (G.W)

Abstract

Background

Kernel number per row (KNPR) is a core trait determining single-plant yield in maize. The molecular mechanisms and favorable allelic variants controlling KNPR diversity are poorly characterized in temperate non-Reid germplasm. This study aimed to boost mapping resolution and screen functional candidate loci for KNPR by developing a multi-parent genetic population.

Results

We established a multi-parent population containing 530 recombinant inbred lines via crosses between the key Reid inbred Ye107 and four genetically distinct elite lines. An integrated analytical pipeline combining genome-wide association study, linkage QTL mapping, haplotype comparison,

protein structure simulation, spatiotemporal expression detection and genomic estimated breeding value modeling was adopted to mine KNPR-associated genomic regions and candidate genes. Four target genes were highlighted: *ZmNMD3* (ribosomal nuclear export), *ZmPAPS1* (mRNA processing), *ZmNAC70* (NAC transcription factor) and *ZmF-box1* (ubiquitin-dependent protein degradation). These genes were backed by suggestive association signals and overlapping quantitative trait loci. The Hap3 haplotype of *ZmNMD3* displayed population-specific distribution and exclusively existed in the non-Reid subgroup Pop3.

Conclusions

Multi-parent populations effectively facilitate the excavation of valuable allelic variations hidden in temperate non-Reid maize germplasm. Following subsequent functional validation, directional gene introgression and multi-gene pyramiding strategies can serve as feasible routes for future precision maize breeding focused on yield improvement.

Keywords: Maize, KNPR, QTL, GWAS, Candidate gene, Integrated mapping

1 Introduction

As one of the world's most important cereal crops, maize (*Zea mays* L.) faces a persistent challenge of sustaining yield gains per unit area to ensure global food security [1, 2]. Maize grain yield per plant is a complex quantitative trait governed by multiple interrelated components, mainly ear row number, kernel number per row (KNPR), and hundred-kernel weight [3, 4]. Among these, KNPR is closely and intricately linked to ear length and final single-plant yield; even minor variation in KNPR can substantially alter total grain number per ear and thus overall single-plant yield [5]. Despite its agronomic

43 importance, the molecular mechanisms governing natural variation in KNPR remain poorly understood.
44 Elucidating these mechanisms is therefore critical for guiding high single-plant yield breeding in maize.
45 Phenotypic variation in KNPR arises from the combined effects of a few major genes (e.g., *fasciated*
46 *ear2* [6]), numerous minor polygenes, and genotype-by-environment ($G \times E$) interactions [7]. Dissecting
47 the environment-dependent genetic effects underlying KNPR is thus essential. Among the primary tools
48 available for uncovering the genetic basis of this trait, quantitative trait locus (QTL) mapping and
49 genome-wide association studies (GWAS) have proven especially valuable [8, 9]. QTL mapping draws
50 on molecular markers and trait segregation within linkage groups to delimit broad chromosomal intervals
51 linked to KNPR [8, 10], whereas GWAS exploits the dense recombination of natural populations to
52 resolve associations down to individual single-nucleotide polymorphisms (SNPs) [9].
53 Classical biparental QTL studies have mapped hundreds of KNPR-associated QTLs across all ten
54 maize chromosomes [11, 12]. On chromosome 1, the major-effect cluster *qEL1* acts pleiotropically on
55 both ear length and KNPR, explaining up to 29.5% of phenotypic variation [13], while the major QTL
56 *qKN* on chromosome 10 was linked to kernel number under variable phosphorus supply [14]. GWAS
57 has added complementary resolution, recovering numerous ear-trait loci in diverse panels [15].
58 Sequencing *TD1* in 165 inbred lines, for instance, uncovered seven SNPs significantly associated with
59 KNPR ($P < 0.05$), three of them nonsynonymous (SNP_1665, SNP_1753, SNP_1896) and potentially
60 relevant to inflorescence development [16], and a GWAS of two recombinant inbred lines (RILs)
61 populations identified 21 KNPR-associated loci [12]. These advances notwithstanding, real obstacles
62 remain. Biparental RIL or F_2 populations sample only a narrow band of diversity, so detection is largely
63 confined to parent-specific alleles and some advantageous variants are missed; their resolution is
64 typically at the Mb scale, which hampers direct candidate-gene identification [17, 18]. GWAS achieves

higher resolution in natural populations but is prone to false positives at small-effect loci, especially under strong population structure. Integrating linkage and association mapping circumvents several of these limitations and offers a powerful route into complex quantitative traits.

In the present study, we adopted an integrated mapping strategy using a multi-parental population (MPP). This MPP, built upon the widely used Reid × non-Reid heterotic framework in Chinese maize breeding, was tailored to reduce false-positive detection while retaining statistical resolution [19]. The population was developed by crossing the common Reid parent Ye107 with four genetically diverse inbred lines (Reid or non-Reid). By preserving family structure for linkage analysis, this crossing scheme captures wide-ranging genetic variation, including rare alleles, and can improve mapping resolution [17, 20]. In addition, the accumulated recombination within the MPP sharpens mapping precision, and the more balanced allele frequencies help curb GWAS false positives, two persistent weaknesses of conventional approaches [21]. We therefore conducted multi-environment phenotypic and genotypic profiling to prioritize candidate genes for KNPR. Through joint analysis of sequence variation, haplotype effects, and spatiotemporal expression data, our goal was to explore potential advantageous variations within temperate breeding pools and generate testable hypotheses about molecular pathways active during ear development. The findings presented here provide candidate genetic resources for subsequent validation and possible future use in maize germplasm improvement.

2 Materials and methods

2.1 Experimental materials

In this study, the parental line Ye107 with low KNPR was used as a common parent and crossed

with four temperate maize inbred lines (Q11, R-2-1-1, YML1218, and Chang7-2) from different heterotic groups to produce F_1 hybrids. The F_1 plants were self-pollinated for seven consecutive generations via the single-seed descent (SSD) strategy, generating four RIL subpopulations at the $F_{2:8}$ generation: Pop1 (Q11 × Ye107), Pop2 (R-2-1-1 × Ye107), Pop3 (YML1218 × Ye107), and Pop4 (Chang7-2 × Ye107) (Fig. 1). The common paternal parent Ye107 and female parents Q11 and R-2-1-1 exhibited typical Reid genetic backgrounds, whereas Chang7-2 and YML1218 belonged to non-Reid heterotic pools. Due to inbreeding depression during population development, the final population contained 530 RILs. The size of each RIL subpopulation varied, ranging from 110 lines in Pop2 to 149 lines in Pop4. Detailed information regarding parental pedigrees and population sizes is listed in Table 1.

Table 1 Details of parental lines and their sources

Parents	KNPR	Pedigree	Ecological type	Heterotic group	Subpopulations	Number of RILs
Ye107	15	Derived from US hybrid DeKalb XL80	Temperate	Reid	-	-
Q11	20	Derived from US hybrid	Temperate	Reid	Pop1 (Q11 × Ye107)	138
R-2-1-1	26	Derived from US hybrid	Temperate	Reid	Pop2 (R-2-1-1 × Ye107)	110
YML1218	25	Huang Zhao Shi × Wei Chun	Temperate	non-Reid	Pop3 (YML1218×Ye107)	133
Chang7-2	18	V59× Huang Zhao Shi	Temperate	non-Reid	Pop4 (Chang7-2 × Ye107)	149

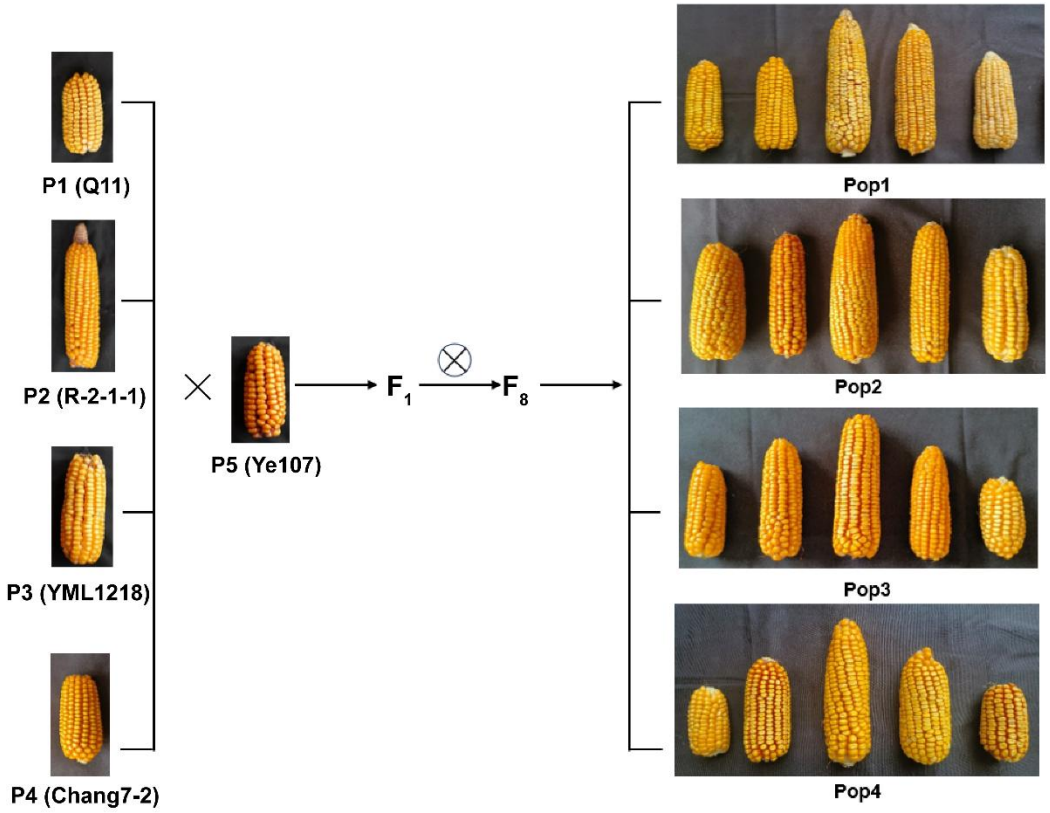


Fig. 1. Ear phenotypes of the parental inbred lines and their derived RIL subpopulations. Representative mature ears are displayed for the four female parental lines (Q11, R-2-1-1, YML1218, Chang7-2), the shared male parent

Ye107, and the four corresponding RIL subpopulations: Pop1 (Q11 × Ye107), Pop2 (R-2-1-1 × Ye107), Pop3 (YML1218 × Ye107), and Pop4 (Chang7-2 × Ye107). All subpopulations were constructed via seven successive self-pollination cycles following the single-seed descent (SSD) method, with Ye107 (Reid heterotic group) serving as the common male parent. Extensive phenotypic variation in ear length, ear diameter and kernel arrangement is observed both within each RIL subpopulation and between Reid-related pools (Pop1, Pop2) and non-Reid pools (Pop3, Pop4).

2.2 Field trials and phenotypic data analysis

Field trials spanned three environments: the materials were grown in 2021 at Jinghong City, Yunnan Province (21JH; 24°43′N, 98°58′E, 920 m) and in 2022 and 2023 at Yanshan County, Wenshan Prefecture, Yunnan Province (22YS and 23YS; 23°59′N, 104°33′E, 1830 m). Given the large number of RIL families, each family occupied a single row without three-row replication. Each experimental plot was 3.5 m in length, consisting of 14 plants per row, with a row spacing of 0.7 m and a plant spacing of 0.25 m. At physiological maturity, ten primary ears were randomly sampled from each inbred line, and their mean values were used as the phenotypic estimate for that genotype. KNPR was determined by counting normally developed grains from the ear base to the tip, excluding sterile grains [3]. Subsequently, Microsoft Excel 2021 and R version 4.3.2 [22] were used to compile the phenotypic data and perform descriptive statistical analyses, normality tests, and correlation analyses. Mixed linear models were implemented using the R (v4.3.2) package lme4 [23]. Best linear unbiased prediction (BLUP) was then performed on the raw KNPR phenotypic data of each RIL across three environments [24], which effectively reduced the influence of environmental variations on the phenotypes, using the Equation (1):

$$Y_{ij}=\mu+Line_i+Loc_j+\epsilon_{ij} \quad (1)$$

Where, Y_{ij} represents the observed phenotypic value, μ represents the intercept, $Line_i$ represents

129 the genetic effect of the *i*-th RIL genotype, Loc_j represents the fixed effect of the *j*-th environment, and
130 ϵ_{ij} represents the random effect (including random residual and genotype-by-environment interaction
131 effect). After normalizing the phenotypic data, the final BLUP values for each RIL were used for
132 subsequent GWAS and QTL mapping analyses. The ggpairs function (v2.2.1) from the GGally package
133 [25] was used to analyze the correlations of KNPR phenotypic data across multiple environments and
134 the four subpopulations. The correlation matrices were annotated with significance levels based on
135 Pearson's correlation coefficients. The lower triangle displays pairwise linear regressions with 95%
136 confidence intervals, while the diagonal panels show phenotypic density distributions with a
137 superimposed theoretical normal curve (red dashed line) to evaluate the data distribution. Since no
138 replicated plots were available within each single environment, the three geographic environments were
139 treated as macro-replicates, and a two-way analysis of variance (ANOVA) model incorporating genotype
140 and environment factors was applied to decompose variance components. The mean square for genotype
141 (MS_g) and the error mean square (MS_e) were obtained using the aov() function. Broad-sense heritability
142 (H^2) on an entry-mean basis was calculated according to the method [26], using Equation (2).

143 $$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_e^2}{n}} \times 100\ \% \quad (2)$$

144 Due to the unreplicated single-row design within each environment, the genotype-by-environment
145 ($G \times E$) interaction variance and the intra-environment plot residual error are entirely confounded within
146 the residual variance component (σ_e^2). Consequently, the derived values represent an empirical upper
147 bound rather than an absolute heritability. Here, σ_g^2 represents the genotypic variance among the 530
148 RIL families, and *n* denotes the number of test environments (n = 3). Variance components were extracted
149 via the restricted maximum likelihood (REML) method using the lme4 package, and the standard errors
150 of H^2 were computed using parametric bootstrapping (via the bootMer function with 500 resamples). To

avoid false precision stemming from the unreplicated plot design, heritability estimates are rounded to one decimal place.

2.3 DNA extraction and whole genome sequencing (WGS)

Genomic DNA from maize leaves was extracted using a modified cetyltrimethylammonium bromide (CTAB) method [27]. DNA integrity was assessed using 1% agarose gel electrophoresis, and DNA purity and concentration were measured using a NanoPhotometer (IMPLEN, Munich, Germany) and Qubit 2.0 (Life Technologies, Carlsbad, CA, USA). Eligible samples with OD260/280 in the range of 1.8 – 2.0 and a total DNA content ≥ 1.5 μg were used for library construction. Library construction was performed using the TruSeq Nano DNA HT protocol (Illumina, San Diego, CA, USA), in which DNA fragments were ultrasonically fragmented to approximately 350 bp and purified using the AMPure XP System (Beckman Coulter, Brea, CA, USA) after end repair, A-tailing, adaptor ligation, and PCR amplification. Library quality was assessed with Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), and effective concentrations were determined by real-time PCR. Qualified libraries were sequenced with 150 bp paired-end reads on the Illumina NovaSeq platform (Illumina, San Diego, CA, USA).

The raw sequencing data were rigorously filtered to obtain high-quality clean reads by removing reads containing more than 10% N bases, reads with more than 50% low-quality bases (*Phred* < 5), and adaptor-contaminated sequences. Clean reads were aligned to the maize reference genome (B73 RefGen_v5) using BWA [28], and the reference genome was obtained from the maize genome assembly reported by Hufford et al. (2021) [29]. SAMtools [30] was used to convert the initial alignment SAM files into sorted BAM format. Variant detection, coordinate-based PCR duplicate marking, and

deduplication were performed using the Sentieon Genomics software package (Sentieon Inc., Mountain View, CA, USA, https://www.sentieon.com/), which provides optimized implementations of standard variant calling workflows. SNPs were filtered using the following filtering criteria: a minimum of four supporting reads, missing data < 0.2, minor allele frequency (*MAF*) > 0.05, mapping quality (*MQ*) ≥ 40, and genotype quality (*GQ*) ≥ 20, to reduce error rates. Finally, ANNOVAR [31] was used to functionally annotate SNPs and determine their genomic locations and mutation types. WGS covered all 530 RILs, and the raw sequence data were deposited in the Genome Sequence Archive (GSA) under accession CRA037261; these records are indexed and citable per GSA guidelines [32, 33].

2.4 Phylogenetic tree, PCA, population structure, and linkage disequilibrium analysis

To examine evolutionary relationships among the lines, we built a Neighbor-Joining (NJ) tree from a genetic-distance matrix in MEGA11 [34]. The genetic relationship matrix across the 530 RILs was estimated by the IBS method in GCTA (v1.94.1), and principal component analysis (PCA) extracted the top three components [35]. The first three principal components were used as coordinates to visualize the three-dimensional spatial distribution of the samples.

Population structure was inferred using the ADMIXTURE program [36], and the optimal number of clusters (*K*) was determined based on the trend of the cross-validation (*CV*) error profile. To further analyze genetic relationships within the MPP, a kinship-based correlation heatmap of the four subgroups was generated using the GAPIT package [37]. For linkage disequilibrium (LD) analysis, PopLDdecay (v3.04) was used to calculate the level of LD (r^2) between pairs of markers [38]. The built-in script Plot_OnePop.pl was then used to plot the LD decay curve. The LD decay distance, which represents the physical distance at which LD declines to background levels or plateaus, was used to determine the

effective range for selecting SNPs linked to candidate genes. The LD decay plot was generated using default parameters.

2.5 Genome-wide association study (GWAS)

In GWAS analysis, individual kinship and population stratification are the primary sources of false-positive associations. To effectively control for these confounding effects, a mixed linear model (MLM) implemented in EMMAX [39] was used for association analysis of KNPR, which is based on the unified mixed-model framework for association mapping [9]. The model incorporated population structure, represented by the top three principal components as fixed-effect covariates, and individual kinship as a random effect to control for both population stratification and cryptic relatedness. Since GWAS was performed separately for each environment, environmental factors were not included as covariates in these models. The MLM was implemented and estimated according to Equation (3):

$$y=X\beta+Zu+e \quad (3)$$

where y is the phenotype vector, X is the matrix of fixed effects (including population structure), β is the fixed-effect coefficients, Z is the genotype matrix, u is the random genetic effect following $u \sim N(0, K\sigma^2)$, and e is the residual error [9].

The genome-wide significance threshold was calculated as $-\log_{10}(1/N_{ind})$, corresponding to one expected false positive per genome scan [40], where N_{ind} represents the number of independent SNP markers estimated following Duggal et al. [41]. To obtain the set of independent SNPs required for N_{ind} calculation, we performed pairwise linkage disequilibrium (LD) pruning using PLINK v1.9 with the custom parameter --indep-pairwise 100 5 0.2. Briefly, the genome was split into sliding windows of 100 SNPs with a step size of 5 SNPs; within each window, only markers with pairwise LD $r^2 < 0.2$ were

retained as genetically independent variants.

To balance statistical detection power and the risk of false negatives for the complex quantitative trait KNPR, a dual-threshold screening strategy was implemented. A suggestive cutoff of $-\log_{10}(P) \geq 5.0$ was applied to retain potential minor-effect loci that might otherwise be missed by conservative adjustment methods. The stringent genome-wide threshold of $-\log_{10}(P) = 5.66$ was calculated from the final set of 455,274 LD-pruned independent SNPs and was used as a conservative reference for genome-wide significance. Therefore, loci supported only by the suggestive threshold were treated as exploratory candidates and were carried forward only when supported by QTL colocalization and biological annotation. The genomic inflation factor (λ) was calculated based on the median of the chi-square test statistics, and a Q-Q plot was generated to evaluate the deviation of observed P-values from the expected distribution.

2.6 Linkage map construction and QTL mapping

To identify QTLs controlling KNPR, genetic linkage maps were constructed and QTL mapping was performed separately for each of the four subpopulations (Pop1 − Pop4). Polymorphic SNP markers within each subpopulation were used to construct genetic linkage maps using JoinMap v4.0 [42]. Progeny genotypes were filtered based on marker completeness ≥ 0.8 and a segregation distortion significance threshold of $P \geq 0.001$ (via chi-square testing) to obtain high-quality population markers. Marker ordering was performed using the maximum likelihood method, and genetic distances were calculated using the Kosambi mapping function [43].

QTL mapping for KNPR was performed using composite interval mapping (CIM) implemented in Windows QTL Cartographer v2.0 [44]. A LOD threshold of 2.5 was applied as a suggestive criterion to

detect putative QTLs, particularly minor-effect loci. The interpretation of QTL effects followed the general framework proposed by Lander and Kruglyak [45]. While recognizing that LOD = 2.5 does not by itself indicate genome-wide significance. Loci with higher LOD values and/or repeated evidence across environments were prioritized for downstream comparison. The proportion of phenotypic variance explained (R^2) and additive effects were estimated for each QTL. Stable QTLs were defined based on consistent detection across environments and subpopulations. To minimize redundancy, QTLs detected in the same subpopulation across different environments (including BLUP values) that exhibited overlapping genetic intervals (in cM) were considered to represent the same QTL and were merged accordingly.

2.7 Candidate gene identification and functional annotation

In this study, a joint GWAS analysis was conducted across the entire MPP to capture common variants and improve resolution, while independent QTL mapping was performed within each sub-population to identify family-specific and rare alleles. We defined consistent loci as genomic regions in which a GWAS signal meeting at least the suggestive threshold overlapped with a QTL interval and was located within the LD-based candidate window. For broad QTL intervals, this colocalization was interpreted cautiously because the chance inclusion of nearby SNPs cannot be completely excluded. The search window for candidate genes was defined based on the genome-wide LD decay of the subpopulations. Specifically, a ± 20 kb flanking genomic window around the lead SNPs was defined for screening, and functional annotation of candidate genes was performed based on the maize B73 reference genome (B73 RefGen_v5) [29]. Given that multiple genes often exist within an LD region, we conducted functional annotation of these genes using public databases such as NCBI, MaizeGDB, InterPro, and

UniProt. Genes whose annotated functions are biologically related to plant growth and development, gene expression regulation, or key metabolic pathways were considered prioritized candidate genes in this study. We also focused on nonsynonymous mutations within the LD regions that could potentially lead to amino acid changes.

2.8 Haplotype analysis and allelic effect evaluation of candidate genes

To evaluate the potential contribution of allelic variation in candidate genes to the phenotype, a gene-centric haplotype association analysis was performed based on whole-genome sequencing data from 530 RILs. The target regions were anchored to the B73 reference genome (B73 RefGen_v5). Considering the reported LD decay in maize [46], regions 20 kb upstream and downstream of the gene coding regions were included to cover potential cis-regulatory elements, such as promoters and enhancers. Considering differences in locus-specific variation, tailored analytical strategies were adopted. For genes located within the LD blocks, after removing low-frequency variants with a $MAF \leq 0.05$ using bcftools [47], haplotype blocks within the target regions were subsequently constructed in Haploview v4.2 [48] using the Gabriel algorithm for haplotype block definition [49]. For genes harboring putative functional variations, such as non-synonymous mutations or candidate regulatory variants, associated SNPs were extracted and their allele effects were evaluated. Finally, the RILs were categorized based on their specific genotypes or major haplotypes (with a haplotype frequency > 5%). Phenotypic differences among the resulting genetic groups regarding the BLUP values of KNPR were statistically evaluated via one-way ANOVA followed by Tukey's Honestly Significant Difference (HSD) test in R.

2.9 Sequence alignment and variant analysis, promoter cis-element analysis, and protein structure modeling

In this study, the B73 reference genome RefGen_v5 was used as a reference, and the complete coding sequences of the target genes were extracted as alignment templates. The corresponding homologous regions were then retrieved from the resequencing data of the representative parental lines, and multiple sequence alignment was performed using DNAMAN software. The analysis primarily focused on nonsynonymous mutations within the coding regions that could lead to amino acid changes. Domain annotation was conducted using the Conserved Domain Database (CDD) from NCBI [50], and motif composition changes were predicted using the MEME suite [51]. For downstream target genes, the 2000 bp region upstream of the translation start site (ATG) was defined as the promoter region, and cis-regulatory elements were predicted using PlantCARE [52]. At the same time, the SWISS-MODEL platform [53] was used for homology-based modeling approaches to analyze the three-dimensional conformation of upstream regulatory factors. Using PyMOL version 2.5, key missense mutations identified in this study were mapped onto the corresponding predicted protein structures to enable visual comparison of conformational differences [54].

2.10 Expression analysis of candidate genes

To target the core determinant stages governing maize KNPR [55], immature ear tissue samples were collected from five parents of MPP at three key stages of female inflorescence development: stage I (1 – 2 mm), stage II (5 – 8 mm), and stage III (10 – 20 mm). Three biological replicates were used for each set of samples, and the maize tissues were immediately flash-frozen in liquid nitrogen and stored at

295 −80 ℃. Total RNA was extracted using the RNAprep Pure Plant Kit (Tiangen, Beijing, China) and its
296 purity and concentration were measured using a NanoPhotometer. One microgram (1 μg) of total RNA
297 was used as a template for cDNA synthesis by reverse transcription using the FastKing RT Kit (with
298 gDNase) (Tiangen, Beijing, China).

299 qRT-PCR assays were performed on a CFX Opus 96 Real-Time PCR System (Bio-Rad Laboratories,
300 Hercules, CA, USA) using the SuperReal PreMix Plus (SYBR Green) Kit. The 20 μL reaction mixture
301 contained 10 μL of 2 × SuperReal PreMix Plus, 0.6 μL of forward and reverse primers (10 μM), 2 μL of
302 diluted cDNA, and 6.8 μL of ddH_2O. The amplification conditions were as follows: pre-denaturation at
303 95 ℃ for 3 min; followed by 40 cycles of denaturation at 95 ℃ for 20 s and gene-specific
304 annealing/extension at optimal temperatures (ranging from 58 to 62 ℃, as detailed in Table S1) for 30 s.
305 *GAPDH* was used as the internal reference gene. The amplification performance of all qRT-PCR primers
306 was evaluated via standard curve analysis using serially diluted cDNA templates. For the reference gene
307 *GAPDH*, the amplification efficiency was 100.51% with a correlation coefficient (R^2) of 0.996, and all
308 candidate gene primers exhibited amplification efficiencies within the acceptable range of 90% − 110%
309 with $R^2 > 0.98$, in accordance with MIQE guidelines for reliable qRT-PCR quantification. Additionally,
310 the Ct values of *GAPDH* showed minimal variation across all sampled developmental stages and parental
311 lines (standard deviation < 0.5 cycles), indicating relatively stable expression under the experimental
312 conditions. All primers were designed using Primer-BLAST [56] (see Table S1 for details). Three
313 technical replicates were established for each sample. Relative expression levels of genes were calculated
314 using the $2^{-\Delta\Delta Ct}$ method [57], and statistical significance between groups was assessed using Student's *t*-
315 test.

2.11 Estimation of Genomic Estimated Breeding Values (GEBV)

In this study, the genomic best linear unbiased prediction (GBLUP)model was employed [58] to estimate the genomic estimated breeding values (GEBV) of 530 RIL individuals. The GCTA [35] software was used to iteratively estimate additive genetic variance and residual variance components based on the restricted maximum likelihood (REML) method. The core mixed linear model was calculated using Equation (4):

$$y=X\beta+g+e \quad (4)$$

where y represents the vector of individual phenotypic observations for ear KNPR; Xβ denotes the fixed effect term used to correct systematic confounding factors such as population structure (the first three principal components of PCA and environmental differences; g represents the vector of individual additive genomic genetic effects (i.e., the GEBV to be estimated), which follows a multivariate normal distribution: $g \sim N(0, G\sigma_g^2)$, where G is the kinship matrix constructed from genome-wide markers, and σ_g^2 represents the additive genetic variance explained by genome-wide SNP markers; e denotes the random residual vector. The GREML algorithm was iteratively performed until model convergence to obtain stable variance component parameters. Finally, the model was solved using the BLUP algorithm to obtain the optimal additive genetic effect values for all individuals, which served as the basis for evaluating genetic performance and conducting superior haplotype pyramiding breeding analysis.

To evaluate model fit and the exploratory breeding relevance of key haplotypes, the entire population was divided into four RIL subpopulations (Pop1 − Pop4) according to pedigree. All population statistical analyses and advanced visualizations were performed using the R software platform [22]. Pearson correlation analysis and linear regression fitting between the observed phenotypic values

of KNPR and the fitted GEBVs were conducted to evaluate the goodness-of-fit of the model, not out-of-sample validation. One-way ANOVA and Tukey's HSD post hoc test implemented via the built-in stats package of R were used to compare the overall GEBV differences among the four subpopulations and to explore the fitted genetic effects of different *ZmNMD3* haplotypes (Hap1, Hap2, and Hap3) in the entire population. Finally, ggplot2 and its extension packages [59] were used to visualize the correlation scatter plots, subgroup difference boxplots, and haplotype effect violin plots described above.

3 Results

3.1 Phenotypic evaluation

The frequency distribution of KNPR phenotypic values and BLUP estimates for the MPP comprising 530 RILs was analyzed across three environments (Fig. 2). KNPR exhibited extensive continuous variation across all environments, following a near-normal distribution. This pattern is consistent with the inheritance of complex quantitative traits controlled by multiple minor-effect polygenes. The phenotypic performance of the four subpopulations (Pop1 − Pop4) across environments is summarized in Table 2. Among them, Pop4 showed the lowest mean KNPR, whereas Pop3 had the highest mean. It is worth noting that within each environment the highest subpopulation mean was reached by Pop2 in 23YS (KNPR = 22.60; Table 2). The coefficient of variation (*CV*) ranged from 14.73% to 30.88% across populations and environments, pointing to considerable phenotypic diversity and confirming that KNPR varies widely enough among these populations to support genetic mapping. To evaluate phenotypic stability and environmental consistency, correlations among environments (21JH, 22YS and 23YS) were assessed for each population (Fig. 3). KNPR generally displayed continuous and

near-normal distribution across populations. Pop1 showed strong agreement between 21JH and 22YS, with a Pearson correlation coefficient (r) of 0.965 ($P < 0.001$) (Fig. 3A). Most inter-environment correlations were also significant, ranging from 0.540 to 0.945 ($P < 0.001$). In contrast, Pop3 showed a relatively low correlation between 21JH and 23YS environments ($r = 0.263$) (Fig. 3C), suggesting possible environment-dependent variation that requires further replicated testing. The empirical upper-bound estimates of broad-sense heritability (H^2) for KNPR ranged from 79.4% to 93.3% on an entry-mean basis (Table 2). Because genotype-by-environment interaction and within-environment plot error were confounded in the unreplicated design, these values should be interpreted as upper-bound estimates rather than absolute heritability. Accordingly, cross-environment Best Linear Unbiased Prediction (BLUP) values were used as integrated phenotypic estimates for downstream QTL linkage mapping and GWAS analyses.

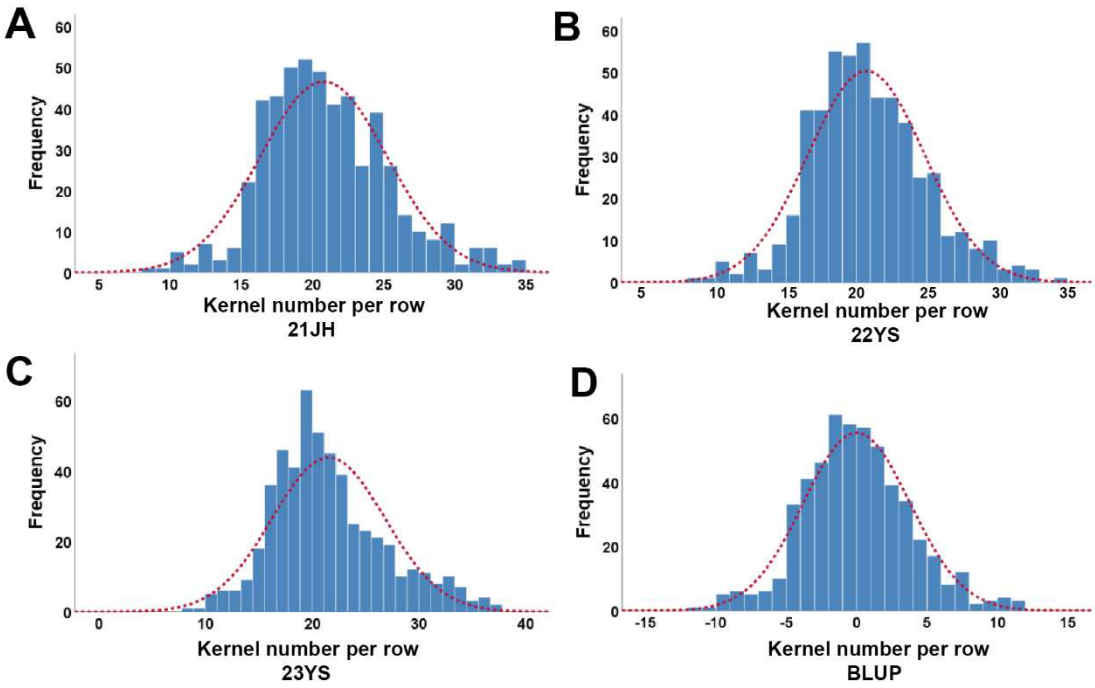


Fig. 2. Phenotypic distribution of KNPR in the MPP. (A – C) Frequency distribution of KNPR in the MPP across three environments (21JH, 22YS, and 23YS). (D) Frequency distribution of cross-environment best linear unbiased

prediction (BLUP) values. The *x*-axis represents KNPR or BLUP values, and the *y*-axis represents individual frequency. Red dotted lines indicate the fitted normal distribution curves.

Table 2 Phenotypic variation of KNPR in four RIL subpopulation

Subpopulations	Environments	Min	Max	Means	Standard Deviation	Skewness	Kurtosis	Coefficient of Variation CV (%)	Heritability Upper-Bound (H^2 ± SE) [95% CI] (%)
Pop1	21JH	10.00	36.67	20.52	5.14	0.88	1.08	25.06	93.3 ± 1.0 [90.9, 94.9]
	22YS	10.00	36.67	20.54	4.90	0.81	1.39	23.86	
	23YS	10.00	36.67	21.44	5.71	0.60	0.11	26.63	
Pop2	21JH	9.00	38.00	21.24	4.95	0.49	1.12	23.31	88.5 ± 2.1 [83.2, 91.6]
	22YS	9.00	40.32	21.25	4.88	0.61	2.04	22.96	
	23YS	9.00	47.29	22.60	6.98	0.91	0.88	30.88	
Pop3	21JH	15.80	34.00	21.94	3.75	0.64	0.31	17.09	79.4 ± 3.1 [71.8, 84.3]
	22YS	15.80	30.63	21.79	3.21	0.56	0.21	14.73	
	23YS	15.56	35.75	22.48	4.14	0.84	0.42	18.42	
Pop4	21JH	8.50	36.67	20.24	4.65	0.76	1.22	22.97	89.1 ± 1.7 [85.1, 91.7]
	22YS	8.50	32.80	20.02	4.24	0.51	0.81	21.18	
	23YS	8.50	42.00	20.51	5.15	0.95	2.03	25.11	

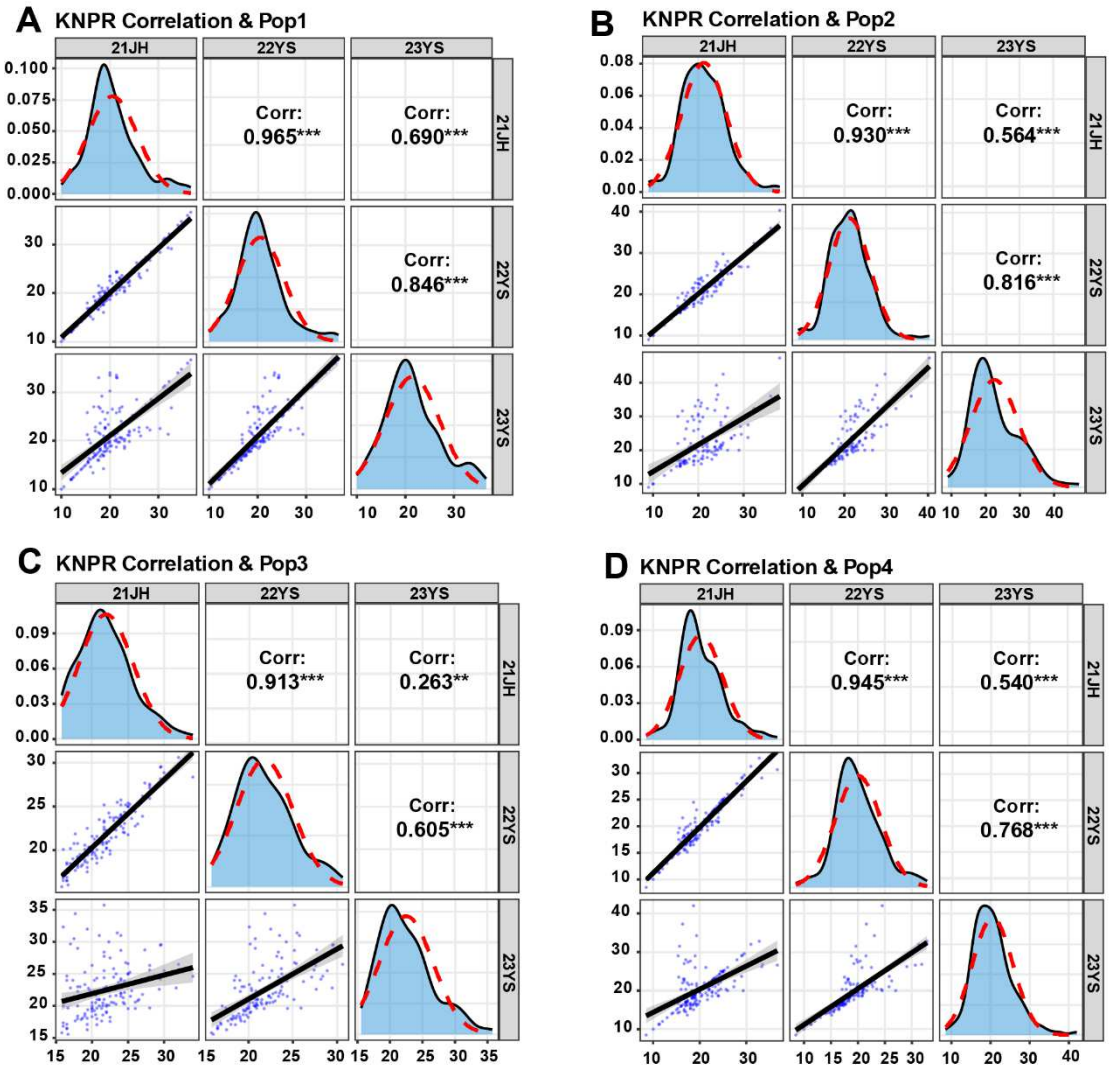


Fig. 3. Correlation analysis and distribution of KNPR across four subpopulations under different environments. The matrix plots display the phenotypic performance of (A) Pop1, (B) Pop2, (C) Pop3, and (D) Pop4 across three environments (21JH, 22YS, and 23YS). The diagonal panels show the density distribution of KNPR in each environment (blue shading), with a red dashed line indicating the fitted normal distribution based on the mean and standard deviation. The lower panels present pairwise scatter plots, together with fitted linear regression lines and 95% confidence intervals (gray shading). The Pearson correlation coefficient (r) is displayed in each matrix unit, where ** and *** indicate significance levels at $P < 0.01$ and $P < 0.001$, respectively.

3.2 SNP characterization, LD decay distance, and population structure

Quality control of the WGS data left more than 6.3 million high-quality SNPs across the MPP. These averaged 296.8 SNPs per 0.1 Mb and were distributed with high coverage and fair uniformity across all ten chromosomes (Fig. 4A): chromosome 1 carried the most markers (886,223; 287.31 SNPs per 0.1 Mb) and chromosome 10 the fewest (461,640; 302.84 SNPs per 0.1 Mb). For association analysis we then dropped SNPs with a missing rate > 0.2 (Fig. 4B) or *MAF* < 0.05 (Fig. 4C), leaving 903,693 high-quality SNPs.

These filtered 903,693 high-quality SNPs were further used to evaluate LD decay across the MPP. As shown in Fig. 4D, the distance at which LD decay tended to plateau for all four subpopulations was approximately 20 kb. This rapid LD decay (~ 20 kb) indicates that the MPP harbors abundant historical recombination events and enables high-resolution genetic mapping. Therefore, candidate genes were subsequently identified by screening the ± 20 kb flanking regions of significant SNPs.

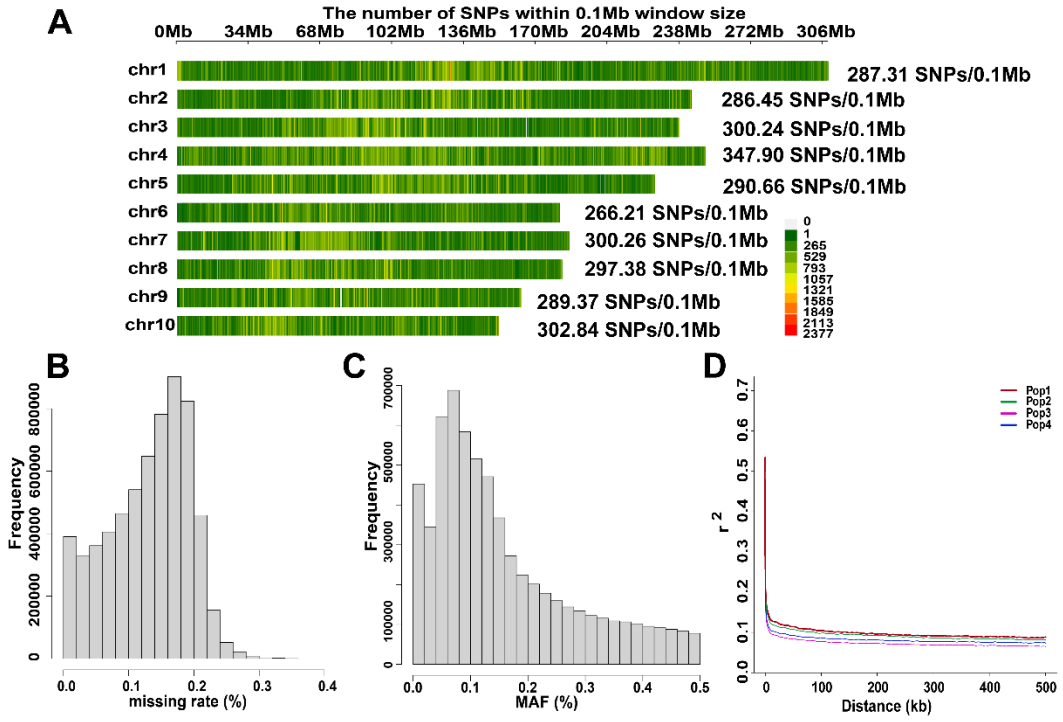
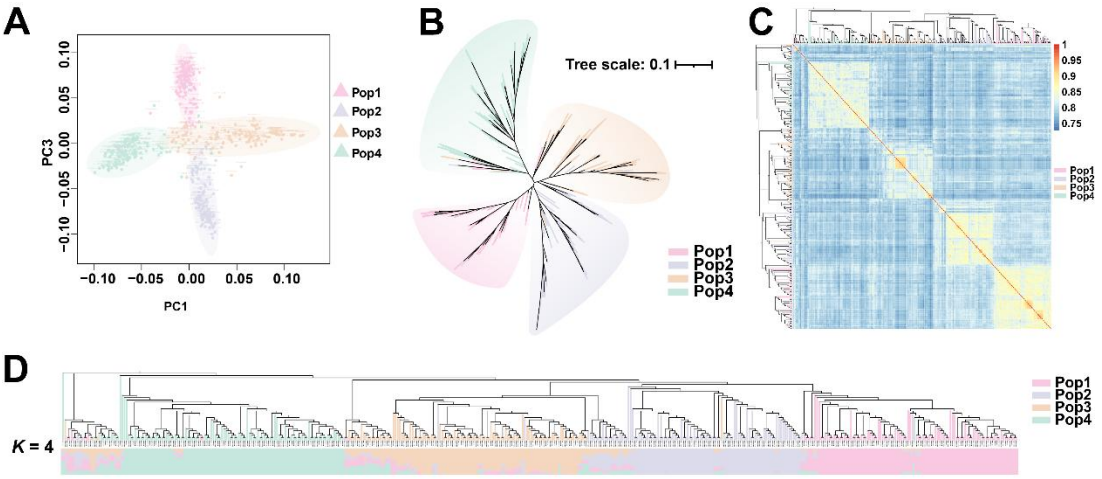


Fig. 4. SNP distribution characteristics across the ten maize chromosomes. (A) Genome-wide distribution of SNPs in the MPP, showing SNP density across all chromosomes. (B) Frequency distribution of SNP missing rates, reflecting the overall quality of genotyping data. (C) Distribution of minor allele frequency (*MAF*) across all qualified SNPs. (D) Linkage disequilibrium (LD) decay curves. The four distinct colors represent the LD decay trends with physical distance for the four subpopulations. r^2: linkage disequilibrium coefficient.

From the PCA plot, the four subpopulations exhibited a cross-shaped spatial distribution centered on the common parental background (Fig. 5A). Pop3 and Pop4 were clearly differentiated along the PC1 axis, which explained the largest proportion of genetic variation, effectively broadened the genetic base within the temperate heterotic framework of the MPP. The distribution patterns of Pop1 and Pop2 along the PC3 axis reflected subtle genetic differentiation within the Reid germplasm group.

The genetic differentiation patterns among and within subpopulations were further validated by the phylogenetic tree (Fig. 5B), kinship heatmap (Fig. 5C), and ADMIXTURE-based population structure analysis (K = 4; Fig. 5D). Collectively, these population structure analyses demonstrate that the MPP captures both inter-group heterotic patterns (Reid × non-Reid, Pop3/Pop4) and intra-group genetic improvement patterns (Reid × Reid, Pop1/Pop2), providing a robust genetic foundation for dissecting the molecular basis of single-plant yield traits across diverse genetic backgrounds.



411

412 **Fig. 5.** PCA, phylogeny, and population structure analyses. (A) Two-dimensional PCA scatter plot showing that

413 the MPP is divided into four clusters. (B) Phylogenetic tree showing that all lines clustered into four subgroups. (C)

414 Kinship heatmap showing higher relatedness among individuals within each subpopulation. (D) Population structure

415 (ADMIXTURE maximum-likelihood clustering) plot at $K = 4$, confirming that the MPP is composed of four

416 subpopulations.

417 **3.3 Joint genetic mapping using GWAS and linkage analysis**

418 We investigated KNPR-associated loci by combining a GWAS across the entire MPP with

419 subpopulation-specific QTL linkage mapping. Rather than relying only on a conservative genome-wide

420 significance threshold that could obscure minor-effect loci, we implemented a dual-threshold strategy.

421 The conservative threshold ($-\log_{10}(P) = 5.66$) was used as a reference for genome-wide significance,

422 whereas the suggestive threshold ($-\log_{10}(P) \geq 5.0$) was used to retain exploratory signals for subsequent

423 QTL colocalization. Adopting only the conservative threshold would have omitted a number of

424 intermediate-effect SNPs, some of which colocalized with QTL intervals (Table S2). However, because

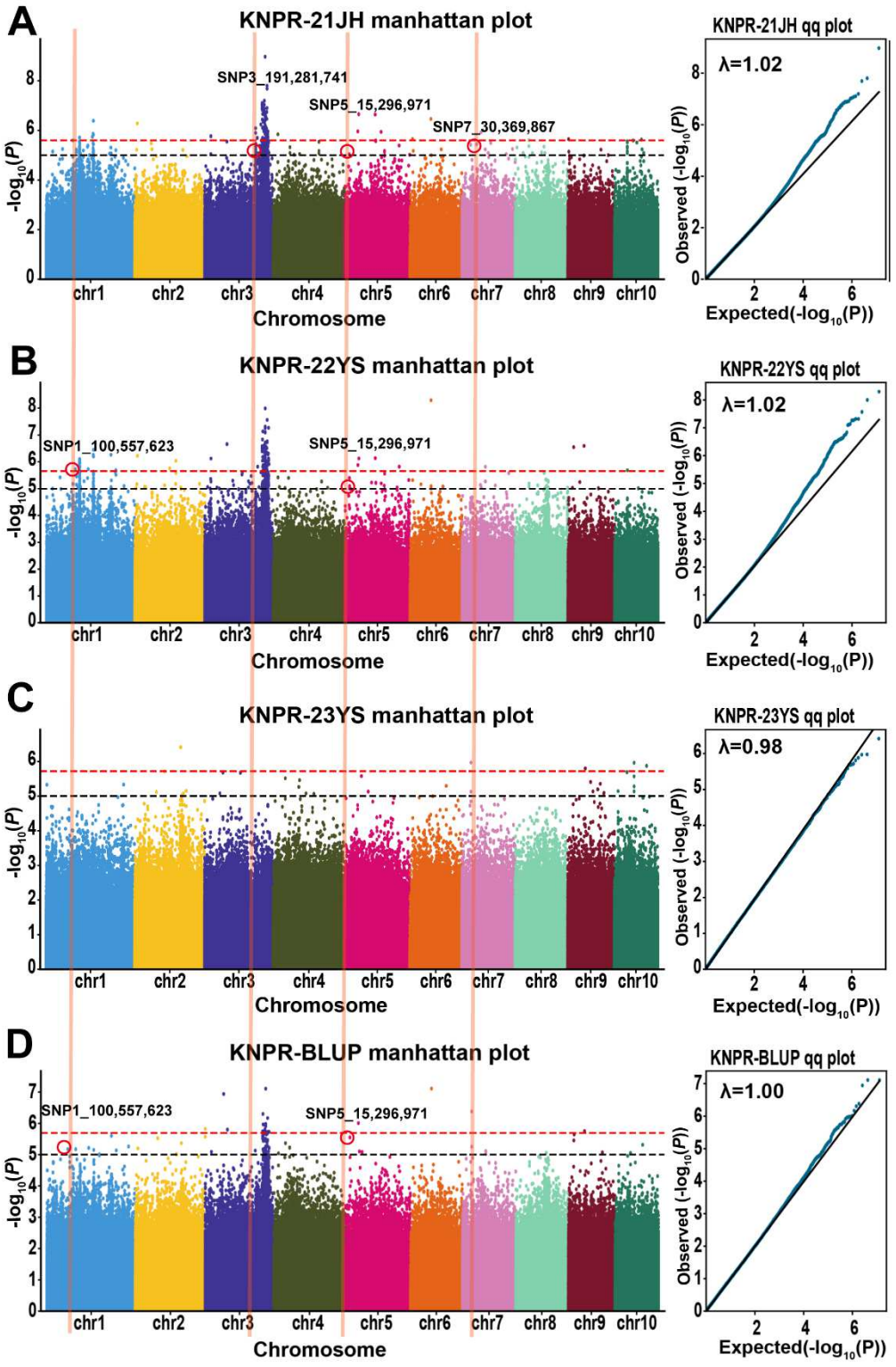
425 suggestive signals have a higher false-positive risk, all downstream candidate genes identified from these

426 signals were interpreted as prioritized hypotheses rather than confirmed causal genes. The MLM

effectively controlled for population stratification and cryptic relatedness, as evidenced by a genomic inflation factor (λ) that remained tightly bounded between 0.98 and 1.02 across all environments and the BLUP dataset.

Using the suggestive threshold ($-\log_{10}(P) \geq 5.0$), 787 association signals were retained from the 903,693 high-quality SNPs evaluated across 530 RILs (Fig. 6). These associated markers explained between 3.19% and 9.40% of the phenotypic variance per locus and comprised 316 SNPs in 21JH, 343 in 22YS, 37 in 23YS, and 91 within the cross-environment BLUP dataset (Table S3). Some peak associations exceeded the conservative genome-wide threshold, whereas the final downstream candidate-gene prioritization relied on the combined evidence of suggestive GWAS signals, QTL overlap, LD-based gene proximity, and biological annotation.

Cross-environmental colocalization of peak markers was clearly discernible in the manhattan plots (Fig. 6). On chromosome 1, the lead marker SNP 1_100,557,623 was repeatedly detected in both the 22YS and BLUP datasets. Furthermore, SNP 3_191,281,741 was mapped on chromosome 3 in the 21JH environment, while SNP 5_15,296,971 on chromosome 5 demonstrated reliable cross-environment significance across the 21JH, 22YS, and BLUP datasets. Finally, SNP 7_30,369,867 on chromosome 7 was successfully resolved within the 21JH environment.



443

444 **Fig. 6.** Manhattan plots (left) and Q-Q plots (right) for (A) 21JH, (B) 22YS, (C) 23YS, and (D) BLUP. The figure

445 displays SNPs associated with KNPR. In the manhattan plots, each dot represents a SNP; the black dashed line

446 indicates the suggestive threshold ($-\log_{10}(P) > 5.0$), and the red dashed line denotes the conservative genome-wide

447 significance threshold ($-\log_{10}(P) = 5.66$). The right-side Q-Q plots compare observed $-\log_{10}(P)$ values against
448 theoretical expected values. The dark green solid line stands for the ideal fitting trend line. The genomic inflation
449 factor (λ) of each group is marked at the upper left corner of the respective Q-Q plot. Four semi-transparent orange
450 vertical lines and red circles highlight SNPs prioritized for colocalization analysis.

451 To identify the genetic loci linked to maize KNPR, four linkage maps were constructed for the four
452 subpopulations. For Pop1, a total of 2,172 SNP markers were mapped across the ten chromosomes,
453 spanning a genetic distance of 1706.93 cM with an average inter-marker distance of 0.79 cM. The largest
454 inter-marker distance was 18.35 cM on LG 10, with 98.39% of inter-marker intervals less than 5 cM,
455 indicating high map density (Table S4). The genetic map of Pop2 was constructed using 2,837 SNPs
456 covering 2,487.83 cM with an average inter-marker distance of 0.88 cM. The maximum inter-marker
457 distance was 18.42 cM on LG 9, and 97.92% of inter-marker intervals were below 5 cM (Table S4). For
458 Pop3 and Pop4, 2,400 and 3,486 SNP markers were mapped across the ten chromosomes respectively to
459 construct the genetic maps. The total genetic distances were 1,501.77 cM (Pop3) and 2,981.64 cM (Pop4),
460 with average inter-marker intervals of 0.63 cM and 0.86 cM. The largest inter-marker interval was 23.74
461 cM (Pop3, LG 9) and 29.56 cM (Pop4, LG 7), with 99.25% and 97.93% of inter-marker intervals shorter
462 than 5 cM, respectively (Table S4).

463 Using these four high-density genetic maps, a total of 34 putative QTLs (LOD $\geq$ 2.5) linked to
464 KNPR were identified via CIM (Table 3). These QTLs were distributed across eight chromosomes, with
465 chromosomes 2 and 6 excluded. The number of QTLs detected in different subpopulations varied
466 considerably; Pop2 contained the largest number (15), whereas Pop3 harbored the fewest (4). The
467 phenotypic variation explained (R^2) by individual QTLs varied widely, spanning from 5.04% (*qKNPR5-*
468 *4*) to 25.73% (*qKNPR4-4*). The largest-effect locus was *qKNPR4-4*, detected in Pop2 under 23YS, where

LOD and R^2 values reached 9.18% and 25.73%, respectively. Its additive effect was −3.32, suggesting that the allele derived from the common parent Ye107 promoted KNPR formation, while the allele inherited from parent R-2-1-1 reduced this trait value. Additionally, *qKNPR4-1* (22YS, LOD = 5.35, additive effect = −2.03) in Pop1 and *qKNPR3-1* (23YS, LOD = 6.00, additive effect = −2.49) in Pop4 represented relatively strong QTL signals (Table 3). Because small mapping populations can inflate estimated QTL effects, these additive and PVE estimates should be treated as preliminary and validated in larger or replicated populations before causal interpretation.

Table 3 Putative QTLs for KNPR identified by QTL mapping across four environments

Chromosome	QTL	Mapping Interval (bp)	Subpopulations	Environment	Position (cM)	LOD	Additive effect	R^2 (%)
3	*qKNPR1-1*	77,728,888 – 113,337,916	Pop2	BLUP	166.59	3.42	1.35	9.11
		96,267,183 – 113,337,916	Pop2	23YS	167.13	6.07	2.6	15.84
	qKNPR1-2	76,741,694 – 83,776,612	Pop2	22YS	177.34	2.76	1.32	7.33
	qKNPR1-3	54,130,735 – 55,627,953	Pop2	22YS	228.69	2.68	1.24	7.72
	qKNPR1-4	449,456 – 454,293	Pop2	23YS	247.2	2.51	-1.51	5.41
	qKNPR 1-5	285,151,119 – 306,022,015	Pop2	BLUP	6.9	4.67	-1.74	12.78
	qKNPR1-6	83,514,874 – 84,178,680	Pop4	23YS	269.49	2.82	1.48	7.86
	qKNPR3-1	187,150,174 – 189,692,576	Pop4	21JH	109.26	4.02	-1.63	13.05
		187,150,174 – 196,353,448	Pop4	22YS	101.75	5.13	-1.66	15.07
		187,149,947 – 196,353,448	Pop4	23YS	104.26	6	-2.49	19.37
		187,149,947 – 196,353,448	Pop4	BLUP	101.75	5.74	-1.54	17.52
	qKNPR3-2	195,892,631 – 199,010,917	Pop1	23YS	58.54	3.61	-2.19	11.91
4	*qKNPR4-1*	74,935,236 – 105,973,706	Pop1	21JH	180.39	5.25	-2.21	16.92
		74,935,236 – 101,249,396	Pop1	22YS	180.39	5.35	-2.03	17.04
	qKNPR4-2	3,697,819 – 5,187,608	Pop1	23YS	268.07	2.77	-1.75	9.41
	qKNPR4-3	88,252,108 – 100,945,325	Pop1	BLUP	182.42	3.64	-1.53	11.77
		116,943,005 – 134,435,864	Pop2	23YS	150.36	9.18	-3.32	25.73
	qKNPR4-4	116,943,005 – 139,382,947	Pop2	BLUP	146.84	4.44	-1.5	12.03
		122,187,908 – 122,446,482	Pop2	22YS	146.58	2.51	-1.27	7.5
	qKNPR4-5	106,931,321 – 114,989,963	Pop2	23YS	181.43	5.41	2.45	13.89
5	*qKNPR5-1*	14,227,093 – 17,472,157	Pop1	21JH	107.21	3.52	1.77	10.92

Chromosome	QTL	Mapping Interval (bp)	Subpopulations	Environment	Position (cM)	LOD	Additive effect	R^2 (%)
		14,227,093 – 17,472,157	Pop1	22YS	170.21	3.38	1.56	10.29
	qKNPR5-2	75,417,145 – 122,154,086	Pop2	21JH	126.38	3.67	1.69	12.21
		114,546,783 – 114,942,726	Pop2	22YS	126.74	3.1	1.46	9.37
	qKNPR5-3	114,426,515 – 131,964,054	Pop4	23YS	173.94	4.56	2.04	13.19
	qKNPR5-4	198,526,905 – 198,717,432	Pop4	BLUP	288	2.62	1.01	5.04
7	*qKNPR7-1*	19,520,737 – 34,857,649	Pop2	23YS	186.15	3.74	-1.98	9.25
	qKNPR8-1	174,639,777 – 175,200,298	Pop1	BLUP	9.99	2.76	1.36	8.75
8	*qKNPR8-2*	106,793,803 – 123,154,830	Pop3	23YS	38.45	4.1	1.54	13.16
		117,268,222 – 127,889,918	Pop3	BLUP	39.46	2.6	0.81	7.66
9	*qKNPR9-1*	143,934,601 – 147,846,511	Pop2	22YS	267.87	3.09	1.63	9.32
		143,510,502 – 150,732,740	Pop2	BLUP	267.87	4.83	1.78	13.25
10	*qKNPR10-1*	117,758,508 – 118,754,658	Pop3	23YS	49.44	3.07	-1.3	9.67
		117,758,508 – 118,391,886	Pop3	BLUP	49.44	2.56	-0.81	7.55

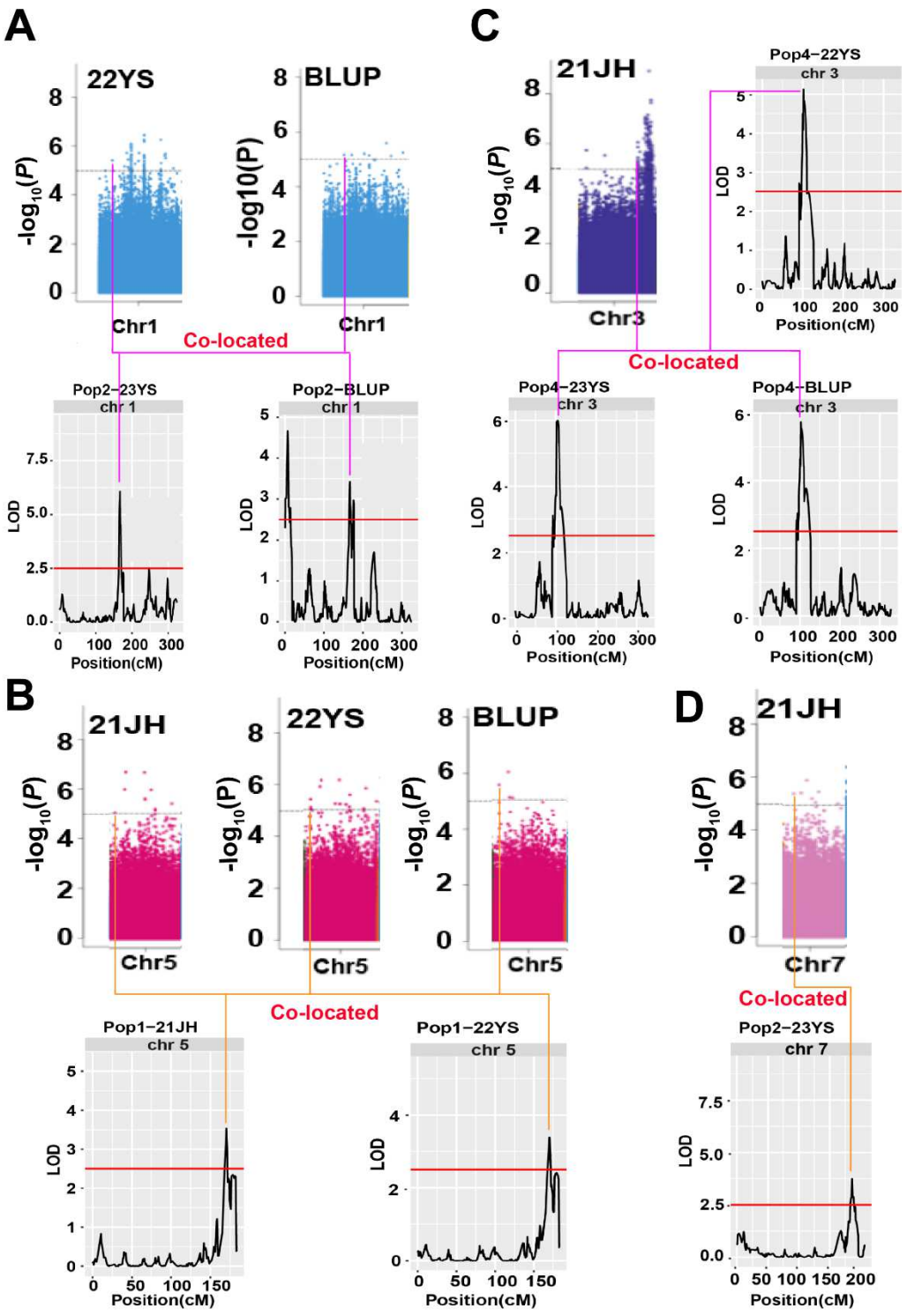
Note: LOD, Logarithm of odds score; a threshold of LOD ≥ 2.5 was used to define putative QTLs. R^2 (%), Percentage of phenotypic variance explained (PVE) by the individual QTL.

3.4 Candidate gene identification and functional annotation

To prioritize candidate genes associated with KNPR, colocalization analysis was conducted to determine whether SNPs retained from GWAS ($-\log_{10}(P) > 5.0$) were located within the 34 putative QTL intervals (LOD ≥ 2.5). Four overlapping SNPs were obtained on chromosomes 1, 3, 5 and 7 (Table 4). The SNP 1_100,557,623 (22YS, BLUP) on chromosome 1 overlaps with the QTL intervals *qKNPR1-1* (23YS and BLUP) identified in the Pop2 subpopulation (Fig. 7A). The SNP 5_15,296,971 (21JH, 22YS and BLUP) on chromosome 5 overlaps with the QTL intervals *qKNPR5-1* (21JH and 22YS) identified in the Pop1 subpopulation (Fig. 7B). The SNP 3_191,281,741 (21JH) on chromosome 3 overlaps with the QTL intervals *qKNPR3-1* (22YS, 23YS and BLUP) identified in the Pop4 subpopulation (Fig. 7C). As shown in Fig. 7D, the SNP 7_30,369,867 (21JH) on chromosome 7 was located within the QTL

interval *qKNPR7-1* (23YS) of the Pop2 subpopulation. Because several QTL intervals are broad, the possibility of chance overlap cannot be fully excluded.

According to the LD decay distance, candidate genes were screened within the ±20 kb flanking regions of each core SNP. All annotated genes in these four loci were retrieved, with detailed gene information listed in Table S5. A total of 11 genes were annotated within these intervals. Following the functional screening criteria described in Section 2.7, one candidate gene was prioritized from each locus. Functional annotation showed that *Zm00001eb024450* encodes 60S ribosomal export protein NMD3, designated *ZmNMD3*; *Zm00001eb217900* encodes nuclear poly(A) polymerase, designated *ZmPAPS1*; *Zm00001eb149810* encodes NAC transcription factor, designated *ZmNAC70*; and *Zm00001eb305010* encodes F-box protein, designated *ZmF-box1* (Table 4). These genes are prioritized candidates for follow-up validation rather than confirmed causal genes.



500

501 **Fig. 7.** Colocalization of SNPs associated with four prioritized candidate genes and their corresponding QTL

502 intervals. Panels A-D correspond to the colocalization plots of *ZmNMD3*, *ZmPAPS1*, *ZmNAC70* and *ZmF-box1*,

503 respectively. The red horizontal line indicates the suggestive QTL threshold (LOD = 2.5), while pink and orange

504 vertical lines denote physical overlaps between QTL peak positions and GWAS SNPs retained for candidate-gene

505 prioritization.

506 **Table 4** Loci prioritized by colocalization analysis using two mapping approaches

QTL	QTL Interval(bp)	SNP	Maximum $-\log_{10}(P)$	PVE of SNP (%)				Candidate Gene	Functional annotation
				21JH	22YS	23YS	BLUP		
qKNPR1-1	96,267,183–113,337,916 77,728,888–113,337,916	1_100,557,623	5.72	-	6.39	-	5.48	*ZmNMD3*	60S ribosomal export protein NMD3
qKNPR5-1	14,227,093–17,472,157	5_15,296,971	5.54	4.52	5.28	-	5.28	*ZmPAPS1*	Nuclear poly(A) polymerase
qKNPR3-1	187,150,174–196,353,448	3_191,281,741	5.37	5.31	-	-	-	*ZmNAC70*	NAC domain-containing protein 70
qKNPR7-1	19,520,737–34,857,649	7_30,369,867	5.41	5.86	-	-	-	*ZmF-box1*	F-box/FBD/LRR-repeat protein

507 **3.5 Haplotype analysis and allelic effect evaluation of candidate genes**

508 We resolved haplotypes at the *ZmNMD3* locus to weigh candidate KNPR-associated variants within
509 the MPP. Rather than define broad blocks, we used five high-quality polymorphic SNPs spanning the
510 *ZmNMD3* promoter and gene body: SNP 1_100,557,623, 4,065 bp upstream of the start codon, anchored
511 the classification (Fig. 8A, 8B), but the final calls integrated all five sites. Three major homozygous
512 haplotypes emerged: Hap1 (GCGTC; n = 65), Hap2 (GCGTT; n = 80), and Hap3 (GTGTC; n = 24) (Fig.
513 8C). These 169 lines make up the main segregating subpopulation from the four representative founders
514 (Chang7-2, Q11, R-2-1-1, YML1218), which had numbered 173 RILs before four lines with low-
515 frequency recombinant haplotypes (local frequency below 5%) were removed. Hap3 carriers showed
516 higher KNPR BLUP values than Hap1 or Hap2 carriers (Fig. 8C), pointing to a possible link between
517 Hap3 and elevated KNPR. Hap3 was uncommon overall and confined entirely to the non-Reid
518 subpopulation Pop3 (YML1218) (Fig. 8D), which makes it a candidate allelic resource that still needs
519 validation.

520 For the other three candidate genes (*ZmPAPS1*, *ZmNAC70* and *ZmF-box1*), single-SNP allelic effect

analysis was performed because several putative functional variants occurred at low frequency. SNP 5_15,296,971 on chromosome 5 is situated 13,566 bp downstream of the 3' untranslated region (UTR) of *ZmPAPS1* (Fig. S1A, B). The low-frequency allele A (AA = 12) corresponded to higher KNPR than the G allele (GG = 205) in the current population (Fig. S1C). Allele A was detected only in Pop2 and Pop3, and its apparent effect should be validated in larger populations before being interpreted as a stable favorable allele (Fig. S1D).

At the *ZmNAC70*-associated interval on chromosome 3, the core SNP 3_191,281,741 locates 3,115 bp upstream of the gene start codon (Fig. S2A, B). Allele distribution was highly uneven across the population. Most lines, including materials sharing the genetic background of common parent Ye107, carried reference allele C, whereas only four lines harbored alternative allele G. Because this rare allele was represented by very few lines, credible comparison between homozygous genotypes was difficult to implement (Fig. S2C). Although the few lines harboring the rare G allele displayed elevated trait values (Fig. S2D), this pattern may reflect sampling noise or background effects and should be interpreted only as a preliminary hypothesis.

Unlike the above loci, *ZmF-box1* on chromosome 7 presented a relatively complex genetic pattern. SNP 7_30,369,867 falls inside the LD block and belongs to intragenic variation within the coding region (Fig. S3A, B). Such a change could plausibly modify the protein's three-dimensional fold or its ability to recognize substrates during ubiquitin-mediated degradation. The variant allele C showed an extremely low frequency in parental and derived populations, which hindered credible comparison between homozygous genotypes. Allele A dominated the whole population with 339 homozygous individuals (Fig. S3C). Comparative analysis under the AA genetic background revealed that lines from Pop3 derived from donor YML1218 had higher KNPR_BLUP values than those from Pop1 and Pop4 (Fig. S3D).

However, with limited representation of the variant allele, genetic background dependence cannot be distinguished from sampling noise, so this locus should be treated as a tentative candidate.

All current findings remain preliminary. Further verification needs larger natural panels or specific populations such as near-isogenic lines (NILs) to increase the frequency of low-frequency alleles and achieve accurate phenotypic effect assessment.

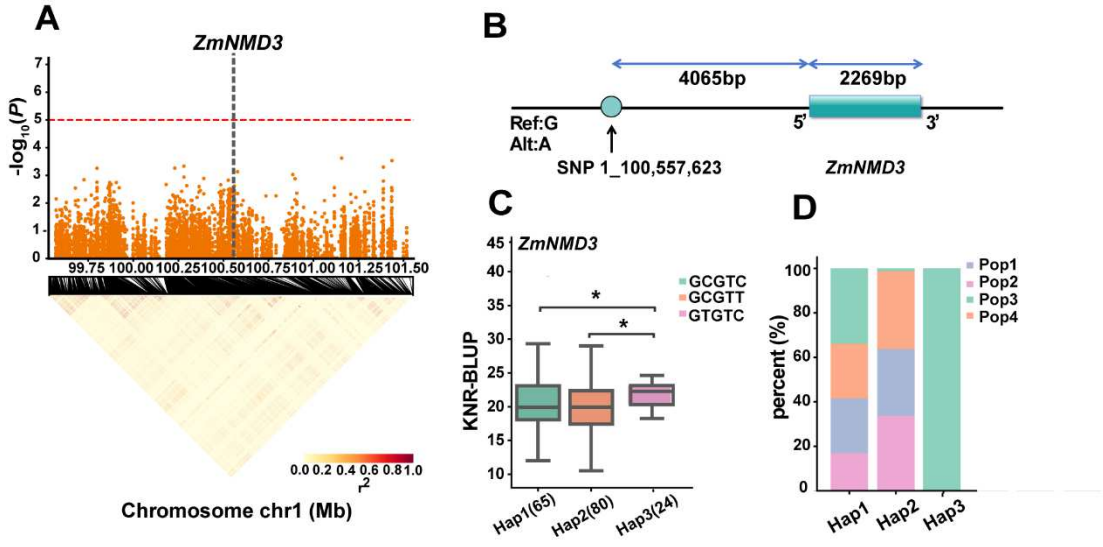


Fig. 8. Haplotype analysis of *ZmNMD3* (A) LD block of *ZmNMD3* (B) Relative physical positions of *ZmNMD3* and the significant SNP 1_100,557,623. (C) phenotypic variation among the three haplotypes based on BLUP values. (D) Distribution of haplotypes across the four subpopulations. Asterisks indicate significance levels: * indicate $P < 0.05$; ** indicate $P < 0.01$; *** indicate $P < 0.001$.

3.6 Nucleotide and amino acid variations of candidate genes, promoter cis-element analysis, and protein structure modeling

Comparing the coding sequences of the four candidates revealed contrasting patterns. *ZmNMD3* carried only synonymous changes, with no amino-acid substitutions, which suggests that any phenotypic effect at this locus would act through non-coding regulatory variation, though direct evidence tying

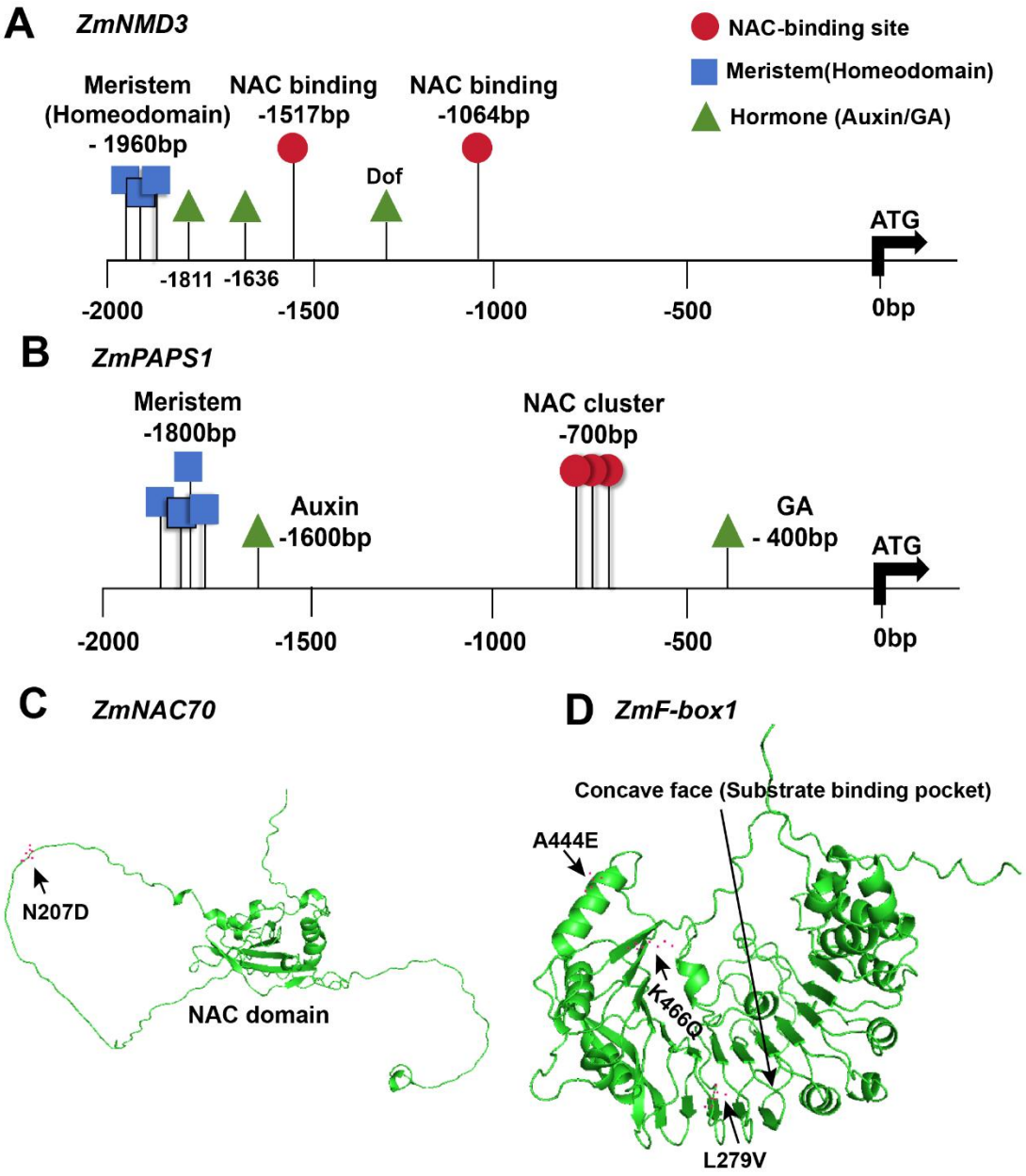
promoter variants to KNPR is still lacking. The other three genes each carried missense mutations of potential functional relevance.

Detailed variant screening (Table 5) identified three missense mutations within the 11th exon of *ZmPAPS1*. One prominent mutation occurred at CDS position 1620, where a T-to-G base transition led to an aspartate-to-glutamate amino acid replacement at residue 540 (p.D540E). Two additional missense variants, p.S588C and p.C598S, were also detected at this locus. Importantly, all three mutations were exclusively present in the high-KNPR parental line R-2-1-1. By comparison, *ZmNAC70* exhibited the densest variant distribution among candidate genes, with six independent nonsynonymous SNPs located within its third exon. Intriguingly, the low-KNPR parent Ye107 and high-KNPR parent YML1218 carried an identical *ZmNAC70* coding haplotype. This consistent haplotype background between phenotypically distinct parents suggests that these exon variants alone are insufficient to drive the observed KNPR differences. Accordingly, *ZmNAC70* was retained as a tentative low-frequency candidate, pending further genetic validation and functional characterization. Three unique missense substitutions were also detected in *ZmF-box1*: a G-to-C transition at position 835 resulting in a leucine-to-valine shift (p.L279V), a G-to-T mutation at position 1331 causing p.A444E, and a T-to-G substitution at position 1396 corresponding to p.K466Q. All three variations were shared by the parental inbreds Ye107 and Chang7-2.

Given the limited explanatory power of coding variants for phenotypic divergence, we further characterized sequence variations in the promoter regions of *ZmNMD3* and *ZmPAPS1*. Both gene promoters harbored cis-regulatory elements functionally associated with meristem growth and hormonal signaling, including canonical homeodomain binding motifs. Bioinformatic prediction also detected conserved NAC-binding sites in both promoter sequences. For *ZmNMD3*, two NAC-binding elements

were located at −1517 bp and −1064 bp upstream of the translation initiation site (Fig. 9A), whereas *ZmPAPS1* possessed a cluster of NAC-binding motifs approximately 700 bp upstream of its start codon (Fig. 9B). It should be noted that these regulatory motifs are prevalent in plant promoter sequences and were identified through in silico prediction; thus, their potential roles in NAC-mediated transcriptional regulation remain preliminary and require experimental verification.

To infer the structural consequences of detected missense mutations, we constructed three-dimensional protein homology models for *ZmNAC70* and *ZmF-box1*. In *ZmNAC70*, the N207D substitution mapped to the surface of the conserved NAC functional domain (Fig. 9C). This replacement of neutral asparagine with negatively charged aspartate potentially alters local electrostatic properties, though its actual functional impact remains unclear. This uncertainty is further supported by the fact that Ye107 and YML1218 harbor identical mutant haplotypes but exhibit contrasting KNPR phenotypes. The *ZmF-box1* protein adopts a canonical LRR-solenoid folding pattern, with its three missense mutations distributed across structurally distinct domains. The L279V variant resides within the substrate recognition interface, while the A444E mutation is positioned along the convex structural backbone (Fig. 9D). Although these structural alterations imply possible changes in substrate affinity and protein stability, current predictive evidence cannot confirm their direct causal effects on KNPR phenotypic variation.



596

597 **Fig. 9.** Structural characterization of putative variations in candidate genes and cis-regulatory analysis. (A, B)

598 Schematic diagrams of predicted cis-regulatory elements within the 2 kb promoter region upstream of the start codon

599 ATG for *ZmNMD3* (A) and *ZmPAPS1* (B). Different symbols mark predicted functional motifs: red circles for NAC

600 binding sites, blue squares for meristem-specific elements including Homeodomain and WOX, and green triangles

601 for hormone-responsive elements such as auxin, gibberellin and Dof binding sites. (C, D) Predicted 3D protein

602 structures. (C) *ZmNAC70* modeled using AlphaFold template K7W3P2; (D) *ZmF-box1* constructed based on

603 template A0A1D6HX57. Green ribbons depict the conserved NAC domain and typical horseshoe-shaped leucine-

604 rich repeat (LRR) domain, and key nonsynonymous mutation sites are labeled with red spheres.

605 **Table 5** Non-synonymous variations in prioritized candidate genes associated with KNPR in parental

606 lines

Gene ID	Chromosome	Ref. Allele	Alt. Allele	Amino acid Substitution	Location (CDS)	Parent carrying the variant
ZmPAPS1	5	T	G	D540E	exon11:1620	R-2-1-1
	5	A	T	S588C	exon11:1762	
	5	G	C	C598S	exon11:1793	
ZmNAC70	3	A	G	N207D	exon3:619	Ye107 YML1218
	3	C	T	S214L	exon3:641	
	3	T	C	V219A	exon3:656	
	3	T	C	F220S	exon3:659	
	3	C	G	P238A	exon3:712	
	3	G	A	A284T	exon3:850	
ZmF-box1	7	G	C	L279V	exon1:835	Ye107 Chang7-2
	7	G	T	A444E	exon2:1331	
	7	T	G	K466Q	exon2:1396	

607 **3.7 Expression pattern analysis of candidate genes in parental lines**

608 qRT-PCR results revealed that the four candidate genes exhibited pronounced stage-specific

609 expression patterns during inflorescence development, and their expression trends were broadly

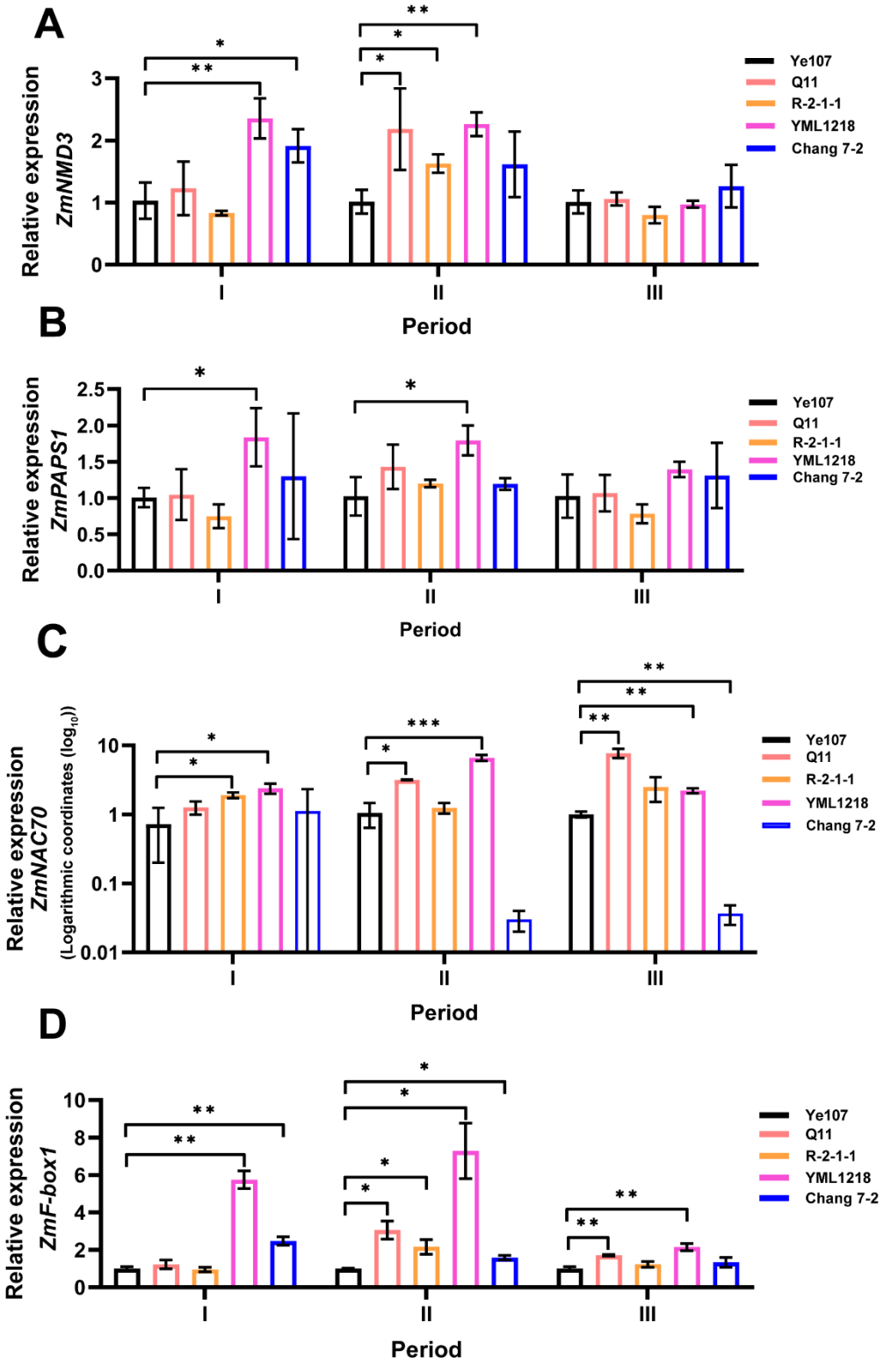
610 consistent with parental KNPR differences (Fig. 10). Because only five parental lines were evaluated,

611 these expression-pattern comparisons should be treated as qualitative support rather than strong

612 quantitative correlation evidence. The mean relative expression values, standard errors of the mean
613 (SEM), and corresponding statistical *P*-values across developmental stages are summarized in Table S6.
614 The detailed expression profiles of the four candidate genes are described as follows.

615 The expression pattern of *ZmNMD3* was broadly consistent with superior KNPR performance.
616 During the rapid differentiation stage of female ear primordia (Stage I), the relative expression levels of
617 *ZmNMD3* in the high-KNPR parent YML1218 (KNPR = 25) and medium-KNPR parent Chang7-2
618 (KNPR = 18) were higher than those in the low-KNPR parent Ye107 (KNPR = 15). By Stage II, both
619 YML1218 and R-2-1-1 (KNPR = 26) continued to show markedly elevated *ZmNMD3* expression than
620 Ye107 (Fig. 10A). For *ZmPAPS1*, the high-KNPR parent YML1218 showed higher transcript abundance
621 than Ye107 at both critical developmental stages (Stage I and Stage II) (Fig. 10B).

622 For *ZmNAC70*, elevated expression was detected in the high-KNPR parents YML1218 and R-2-1-
623 1 at Stage I. YML1218 also maintained higher expression than Ye107 at Stage II and Stage III (Fig. 10C),
624 broadly matching the observed KNPR trend. *ZmF-box1* exhibited development-dependent expression
625 characteristics. At Stage I, *ZmF-box1* expression was transiently upregulated in the highest-KNPR parent
626 YML1218, showing higher levels than all other parental lines, and this expression trend was also
627 observed at Stage II (Fig. 10D).



628

629 **Fig. 10.** Schematic diagram of gene expression profiles. (A – D) Relative expression levels of *ZmNMD3*,

630 *ZmPAPS1*, *ZmNAC70* and *ZmF-box1* across five parental lines. Bar heights indicate mean expression values from

three biological replicates, while error bars represent the standard error of the mean (SEM). **Asterisks mark significant differences compared with the low-KNPR parent Ye107 at the identical developmental stage (Student's t-test: *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$). The *x*-axis presents three developmental stages: I, ear primordium length of 1 – 2 mm; II, ear primordium length of 5 – 8 mm; III, ear primordium length of 10 – 20 mm. The *y*-axis of panel (c) adopts a $\log_{10}$ scale.

3.8 GEBV Model Fit and Exploratory Evaluation of Candidate Haplotypes

We calculated GEBVs using genome-wide markers to evaluate genome-wide additive genetic effects and to explore the potential relevance of four candidate genes (*ZmNAC70*, *ZmPAPS1*, *ZmNMD3*, and *ZmF-box1*), alongside their haplotypes and favorable variants, for future molecular breeding. Individual GEBVs across the RIL lines were estimated directly via a genome-wide GBLUP framework implemented in GCTA, utilizing the complete set of genome-wide markers without cross-validation partitioning. Because no independent validation or cross-validation partitioning was performed, the correlations reported here represent model fit rather than independent validation. To evaluate the goodness-of-fit of the GBLUP model across distinct genetic backgrounds, we performed a correlation analysis between the multi-environment BLUP phenotypic values and the fitted GEBVs (Fig. 11A). Pearson correlation coefficients (r) reached 0.51, 0.65, 0.49, and 0.54 for Pop1, Pop2, Pop3, and Pop4, respectively, with all associations being highly significant ($P < 0.0001$). These results indicate that the genome-wide GBLUP model fits a substantial component of additive genetic variation for KNPR across diverse germplasm backgrounds, providing a descriptive baseline for exploratory haplotype comparison. Subpopulation comparison of fitted GEBV distribution revealed differences in genome-wide additive genetic values among groups (Fig. 11B), which provides a population-level reference for exploratory

comparison of germplasm groups. Pop3 presented higher fitted GEBV levels than the other three subpopulations, a finding consistent with its abundant non-Reid germplasm composition. These observations imply that temperate non-Reid germplasm may possess useful genetic variation for KNPR improvement, but this conclusion requires validation in replicated breeding populations.

To trace the basis of Pop3's higher fitted GEBVs, we compared GEBV levels among the *ZmNMD3* haplotypes across the whole population. The violin plots showed Hap3, enriched specifically in the non-Reid Pop3, at higher fitted GEBVs than the widely distributed Hap1 and Hap2 (Fig. 11C). The pattern again suggests a positive link between *ZmNMD3*-Hap3 and improved KNPR, and marks it as a candidate allelic resource and marker target for future validation.

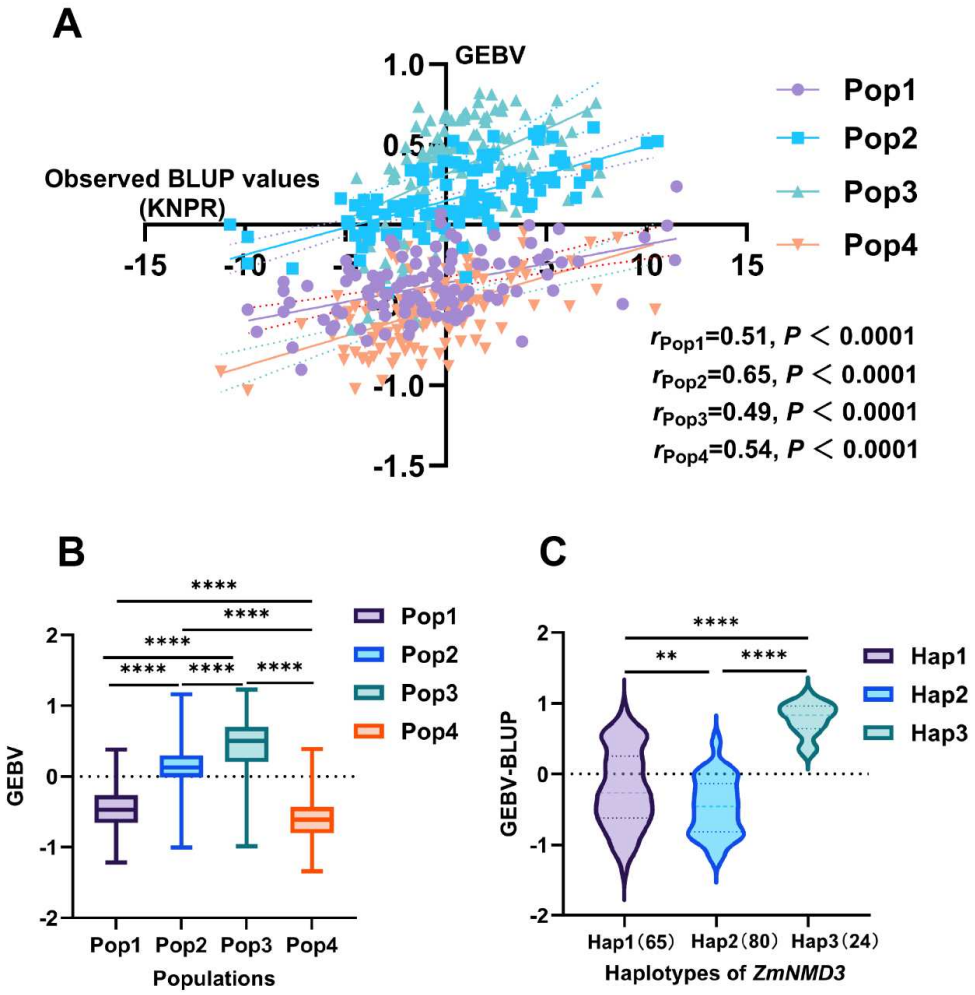


Fig. 11. GEBV model-fit analysis and exploratory haplotype comparison across four recombinant inbred line

populations. (A) Scatter plots showing correlations between observed BLUP values and fitted GEBVs in different subpopulations. The *x*-axis represents multi-environment KNPR BLUP values, and the *y*-axis represents the corresponding fitted GEBVs. Purple circles, blue squares, green triangles, and orange inverted triangles represent Pop1, Pop2, Pop3, and Pop4, respectively. The Pearson correlation coefficient (*r*) and significance level are shown in the lower-right corner for each population (all $P < 0.0001$). Dashed lines flanking the regression line indicate the 95% confidence interval. (B) Boxplot comparison of genome-wide fitted breeding value distributions among the four subpopulations, illustrating population differences across distinct genetic backgrounds. Each box denotes the interquartile range, with the internal horizontal line representing the median. (C) Violin plots showing fitted GEBV distributions of three major *ZmNMD3* haplotypes. Internal dashed lines indicate the median and upper/lower quartiles. Numbers in parentheses indicate the sample size of each haplotype group. Asterisks indicate significant differences based on one-way ANOVA followed by Tukey's HSD post-hoc test: ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

4 Discussion

4.1 Efficacy of integrated mapping strategies in deciphering the genetic architecture of complex traits

The ability to prioritize KNPR-associated loci in this work owes much to the integrated use of GWAS and QTL mapping of the MPP. Based on the classical heterotic pattern (Reid x non-Reid) widely adopted in Chinese maize breeding, we constructed an MPP comprising 530 RILs. Compared with traditional biparental populations, this design captures historical recombination and diversity, which are reflected in the improved mapping resolution observed in this study. MPPs offer a balance between the

683 limited diversity of biparental crosses and the complexity of natural populations. They retain the
684 statistical power associated with controlled crosses while introducing broader allelic variation from
685 multiple founders, thus facilitating fine mapping [17, 60]. In this study, QTL mapping identified putative
686 loci contributing to phenotypic variation [61, 62], whereas GWAS improved resolution and helped
687 identify loci with minor effects [63]. The joint analysis provided a candidate-locus framework for KNPR
688 and helped prioritize genomic regions on chromosomes 1, 3, 5, and 7 for further validation.

689 Notably, the MPP constructed in this study based on the classical heterosis pattern helped reduce
690 population stratification, improve mapping resolution of complex traits, and explore advantageous allelic
691 variations. The abundant genome-wide historical recombination events in the MPP enabled us to identify
692 the candidate *ZmNMD3* haplotype Hap3 that exists exclusively within the non-Reid background. This
693 methodological advantage is consistent with recent landmark studies showing that multi-parent nested
694 designs, such as NAM or MPP, can improve the detection power of rare and minor-effect variants and
695 advance molecular mapping of complex quantitative traits [64]. Nevertheless, the candidate genes
696 prioritized here were screened mainly from statistical association, QTL colocalization, and annotation
697 evidence. Their biological functions and causal effects remain to be experimentally validated through
698 approaches such as gene knockout and advanced-generation NILs.

699 **4.2 Comparative analysis of KNPR-associated SNPs and QTLs with previous studies**

700 In this study, 34 QTLs and 787 SNPs associated with KNPR in maize were identified through QTL
701 mapping and GWAS. To improve the reliability of these findings, an integrated analysis combining both
702 approaches was performed, and four SNPs located within four QTL intervals were finally identified.
703 Comparison with previously reported loci revealed several overlapping key genomic regions.

On chromosome 1, the physical interval of *qKNPR*1-1 (77,728,888 – 113,337,916 bp), based on BLUP values, partially overlaps with the major KNPR QTL (95,695,823 – 95,701,008 bp) reported by Yang et al [65]. Additionally, our *qKNPR*1-1 (BLUP) partially overlaps with two QTL intervals reported by Wang et al., namely *qKNR1-2* (87,932,948 – 87,933,008 bp) and *qKNR*1-3 (82,649,009 – 87,932,982 bp) [12]. The KNPR QTL intervals *qKNPR1-1* detected in both the 23YS and BLUP environments (96,267,183 – 113,337,916 bp and 77,728,888 – 113,337,916 bp, respectively) encompass three SNP loci (chr1: 94,489,467, 94,684,126, and 100,991,594 bp) previously identified through A-D testing [65]. This overlap supports the relevance of this genomic region, although the broad QTL interval means that the specific causal gene requires further fine mapping.

On chromosome 5, the interval spanning *qKNPR5-1* (14,227,093 – 17,472,157 bp) partially overlaps with *KRN5b* (15,330,470 – 15,338,225 bp). The gene *Zm00001eb217930* (v4 ortholog: *Zm00001d013603*), located in this region, has been implicated in panicle development and kernel row number [66], indicating that this region may represent a candidate hotspot for single-plant yield traits.

For chromosome 3, the bin 3.04 – 3.05 region identified by Chen et al. in a multiparent population has been recognized as a pleiotropic locus controlling maize grain traits. This region includes *qKL3-1*, *qHKW3-1* and *qKNR3-1* corresponding to a physical interval of 187.16 – 195.90 Mb (B73 RefGen_v5). This interval partially overlaps with *qKNPR3-1* (187,150,174 – 189,692,576 and 187,150,174 – 196,353,448 bp) identified in this study, highlighting a genomic region potentially associated with trade-offs between grain size and number. Additionally, pleiotropic QTLs in the Bin 3.08 region (chr3: 180 – 190 Mb) partially overlap with the *qKNPR3-1* QTL interval identified on chromosome 3 in this study, suggesting a candidate hotspot for maize ear and grain traits [67].

On chromosome 7, *qKNPR7-1* (19,520,737 – 34,857,649 bp) partially overlaps with a previously reported QTL for hundred-kernel weight (HKW) (2,630,972 – 21,162,260) [68]. The consistency between our results and earlier studies suggests that this region may harbor candidate genes involved in KNPR regulation, although additional validation is needed.

4.3 Functional annotation of candidate genes

4.3.1 *ZmNMD3*: Constraints of ribosomal export efficiency on ear developmental potential

ZmNMD3 on chromosome 1 encodes a highly conserved eukaryotic 60S ribosomal export adaptor critical for nuclear subunit export and maturation rather than nonsense-mediated mRNA decay [69]. In rice, the ortholog *OsNMD3* exports the 60S subunit via the CRM1/XPO1 pathway; disrupting it or trapping it in the cytoplasm impairs ribosome maturation, which curbs translation and, in turn, causes dwarfism, less internode cell proliferation, and smaller, lighter grain [70]. Taken together, this work casts NMD3 as a support for efficient protein synthesis through ribosomal export, with knock-on effects on internode growth, ear development, and grain-yield traits.

In keeping with this conserved role across grasses, we prioritized *ZmNMD3* as a maize candidate that may connect transcriptional variation to KNPR. Its coding sequence is highly conserved among the parents, but its promoter holds predicted NAC-binding motifs and meristem-specific elements, and its upregulation during the early differentiation of female ear primordia (Stages I/II) in high-KNPR lines is compatible with a cis-regulatory mechanism. Even so, no direct evidence yet ties specific promoter SNPs, NAC binding, and KNPR variation together, so this mechanism remains speculative.

4.3.2 *ZmPAPS1*: Mediating mRNA processing to influence ear development

ZmPAPS1 on chromosome 5 encodes a nuclear poly(A) polymerase essential for precursor mRNA 3'-end processing, a modification governing transcript stability, nuclear export, and translation efficiency [71]. In Arabidopsis thaliana, the PAPS1 homolog controls leaf and floral organ morphology by modulating cell proliferation and expansion, potentially via the preferential polyadenylation of auxin-responsive transcripts within the SAUR family [72]. Accordingly, PAPS proteins operate as regulatory intermediaries integrating upstream developmental and environmental stress signals into alternative polyadenylation (APA) responses; by adjusting 3' UTR lengths, these enzymes generate distinct mRNA isoforms to dynamically modulate target gene expression [72, 73].

In our study, *ZmPAPS1* from the high-KNPR parent R-2-1-1 harbors three missense mutations (p.D540E, p.S588C, and p.C598S) in exon 11, coupled with a promoter containing predicted NAC-binding elements. Transcript upregulation in high-KNPR lines is consistent with the hypothesis that *ZmPAPS1* may function as a post-transcriptional node. However, the combined effects of protein structural variations and cis-regulatory differences on mRNA processing and KNPR remain to be directly tested.

4.3.3 *ZmNAC70*: A low-frequency candidate requiring validation

ZmNAC70 encodes a typical NAC transcription factor featuring a conserved N-terminal DNA-binding domain and a C-terminal transcriptional regulatory region involved in stress and developmental signaling [74, 75]. Its rice ortholog *OsNAC4* governs programmed cell death (PCD) [76], which raises the possibility that conserved NAC-mediated cell homeostasis also shapes floret differentiation and ovary survival during maize female-ear development, and thereby final yield components. NAC transcription factors broadly direct seed maturation and grain-development programs across plants [77], and several

maize NAC homologs act in abiotic-stress adaptation; *ZmNAC080308*, for instance, helps protect single-plant grain yield under drought [78, 79]. Specific NAC-EXPANSIN modules further enhance single-plant yield potential by guiding spatial tissue development during kernel formation [80].

Here, the low-frequency variant introduces an N207D substitution on the surface of the conserved NAC domain, which we hypothesize could shift local electrostatics and perhaps DNA-binding affinity. Yet Ye107 and YML1218 share this *ZmNAC70* mutation haplotype despite their opposing KNPR values, so the haplotype alone cannot explain the difference between them. *ZmNAC70* is therefore best read as a tentative candidate whose effect may hinge on genetic background or linked variation, and which needs validation in larger segregating populations.

4.3.4 *ZmF-box1*: A candidate locus potentially related to protein degradation pathways

ZmF-box1 on chromosome 7 encodes an F-box/FBD/LRR-repeat protein acting as the substrate-recognition module of the SCF (Skp1-Cullin-F-box) E3 ubiquitin ligase complex, which directs cell division, organogenesis, and phytohormone response via the ubiquitin-proteasome pathway [81]. By mediating the targeted degradation of key transcriptional repressors (such as Aux/IAA or DELLA inhibitors), F-box proteins release downstream developmental programs required for reproductive meristem regulation and seed formation [82, 83]. Since ear development and final KNPR rely heavily on precise hormone-mediated floret differentiation and ovary establishment [84-86], post-translational protein turnover provides a critical layer of control.

Sequence analysis revealed three missense mutations within *ZmF-box1*, including p.L279V localized inside the substrate-recognition interface of the LRR domain, potentially modifying substrate-binding affinity. Combined with a stage-specific expression profile during early inflorescence

development, *ZmF-box1* is hypothesized to participate in ear development through protein turnover. Given the extremely low frequency of the variant allele and the difficulty of separating background effects from sampling noise, this model remains speculative and requires further experimental validation.

4.4 Comparative analysis of candidate genes with previous studies

KNPR, a major determinant of maize single-plant yield, is controlled by a complex regulatory network involving meristem maintenance, hormone signaling, and conserved peptide-receptor pathways. Genetic studies in both biparental and natural populations have identified multiple loci underlying variation in ear architecture. A key regulatory module is the CLAVATA signaling pathway, which controls inflorescence meristem size. In maize, the LRR receptor-like protein *FEA2* has been shown to regulate kernel row number by controlling stem cell proliferation in the shoot apical meristem [87]. Variation in *FEA2* leads to increased meristem size and enhanced kernel row number through modulation of CLAVATA signaling. Downstream of this pathway, the pseudokinase *ZmCRN* interacts with *FEA2* and contributes to the regulation of meristem activity; loss-of-function mutations result in fasciated ears and increased kernel number [88, 89]. In addition to receptor-mediated signaling components, transcriptional and cis-regulatory mechanisms also play central roles in KNPR variation. The *KRN4* locus acts as a distal cis-regulatory element of the SBP-box transcription factor *UB3*, thereby influencing inflorescence architecture and kernel row number through hormone-related developmental regulation [90]. Another key regulator, *KRN2*, encodes a WD40-repeat protein that negatively regulates grain number. Independent studies in maize and rice demonstrated that knockout of *KRN2* orthologs increases grain yield by approximately 8 – 10%, supporting its conserved role in yield regulation under convergent selection [91]. These findings suggest that KNPR is regulated by interconnected signaling, transcriptional,

and regulatory layers, highlighting both conserved developmental pathways and species-specific regulatory elements underlying maize yield formation.

Based on functional annotation and sequence variation, we propose a hypothetical regulatory framework in which *ZmNAC70*, a transcription factor with a DNA-binding domain mutation (N207D), may act as an upstream regulatory component potentially involved in stress- or hormone-responsive pathways. In this model, *ZmNMD3* and *ZmPAPS1*, which are annotated as genes involved in ribosome export and RNA processing, respectively, may represent potential downstream targets of transcriptional regulation; however, direct regulatory relationships have not yet been experimentally confirmed.

Furthermore, *ZmF-box1*, which encodes a putative F-box protein, may be involved in ubiquitin-mediated protein degradation pathways. Given the known role of F-box proteins in SCF E3 ubiquitin ligase complexes, it is possible that *ZmF-box1* participates in protein turnover processes that indirectly influence hormone signaling and developmental regulation. Nevertheless, these mechanistic hypotheses remain speculative and require further experimental validation. Importantly, the proposed interactions among *ZmNAC70*, *ZmNMD3*, *ZmPAPS1*, and *ZmF-box1* are currently based primarily on genomic colocalization, sequence variation, and in silico functional annotation. Therefore, these relationships should be interpreted as preliminary models rather than confirmed regulatory pathways. Future studies using functional assays such as ChIP-qPCR, yeast one-hybrid, electrophoretic mobility shift assays, and dual-luciferase reporter systems will be necessary to validate these predicted regulatory connections.

Taken together, our results extend previous knowledge of KNPR regulation by suggesting that, in addition to well-characterized signaling and cis-regulatory modules, ribosome-related processes, transcription factor variation, and ubiquitin-mediated protein turnover may contribute to KNPR variation. These findings provide working hypotheses for future functional dissection rather than definitive

evidence for causal regulatory pathways.

4.5 Analysis of environmental sensitivity and yield stability of non-Reid germplasm resources

Across the MPP, the empirical upper-bound estimates of broad-sense heritability for KNPR ranged from 79.4% to 93.3%. Because each environment lacked within-environment replication, these values should not be interpreted as absolute evidence that KNPR is predominantly genetically controlled. The relatively weak correlation observed for Pop3 between 21JH and 23YS environments suggests that environment-dependent variation may also be important, but this inference is based on a limited comparison and should be tested in replicated multi-environment trials. At the same time, temperate non-Reid germplasm carries potentially useful alleles, including *ZmNMD3*-Hap3, which was associated with improved KNPR in this study. These observations highlight the need to balance genetic novelty with stability and to validate candidate alleles in appropriate breeding backgrounds before practical deployment.

4.6 Post-validation breeding prospects based on candidate haplotypes

An important finding of this study is that the candidate haplotype Hap3 of *ZmNMD3* exists only in the non-Reid parent YML1218 and its derived population (Pop3). This distribution may have implications for breeding if the effect of Hap3 is confirmed. While Reid germplasm has been extensively utilized because of its overall performance, its extensive utilization has reduced genetic diversity, limiting further improvement in KNPR and ultimately restraining single-plant grain yield. In contrast, low-frequency haplotypes such as Hap3 may provide an opportunity to introduce novel variation into the elite

850 temperate breeding pool. If functional validation confirms a stable favorable effect, systematic
851 introgression of Hap3 into elite Reid backgrounds using marker-assisted backcrossing could be
852 considered. The GEBV model-fit analysis (Fig. 11C) showed that Hap3 tended to have higher fitted
853 breeding values in the current population; however, this pattern should be regarded as preliminary and
854 not as evidence of immediate breeding applicability (Fig. 12).
855 (1) Foreground Selection: If Hap3 and other candidate variants are functionally validated, KASP or
856 CAPS markers can be developed based on key variants, such as SNPs in promoter regions associated
857 with predicted NAC binding sites. Plants carrying the target allele during successive backcross
858 generations (BC_1 to BC_4) could then be selected.
859 (2) Background Selection: Genome-wide markers (e.g., 150 – 200 SSR or SNPs) could be used to
860 accelerate recovery of the recurrent parent genome, aiming for > 95% recovery by BC_2F_2 or BC_3F_1
861 generation after validation of the target allele. This would potentially shorten the breeding cycle (from
862 traditional 6 – 8 generations to 3 – 4 generations).
863 (3) Simulation-based evaluation: Phenotypic data should be used to estimate the expected gain in KNPR
864 and overall single-plant yield following introgression of the target haplotype.
865 In addition, a multi-gene stacking strategy can be considered as a post-validation breeding prospect
866 for the concurrent improvement of grain number and single-plant yield. Under this scenario, validated
867 favorable alleles of *ZmNMD3*, *ZmPAPS1*, *ZmF-box1*, and *ZmNAC70* could be combined to test whether
868 their effects are additive, synergistic, or background-dependent. NILs carrying validated favorable alleles
869 of each gene could be developed using MABC with Ye107 or other elite lines as recurrent parents.
870 Subsequent crossing among NILs combined with marker-assisted selection would help identify lines
871 carrying multiple alleles. Because the current GBLUP analysis reflects model fit rather than independent

validation, and because the four loci have not yet been functionally validated, the strategy of 'single-gene precise improvement + multi-gene pyramiding' should be regarded as a testable breeding framework rather than an immediately deployable scheme.

Overall, our results suggest that, after functional validation, combining targeted introgression of low-frequency candidate variations with multi-gene stacking may serve as a potential framework for improving maize single-plant yield. This approach should be evaluated in replicated NILs, multi-gene pyramiding populations, and multi-environment field trials before translation into breeding applications.

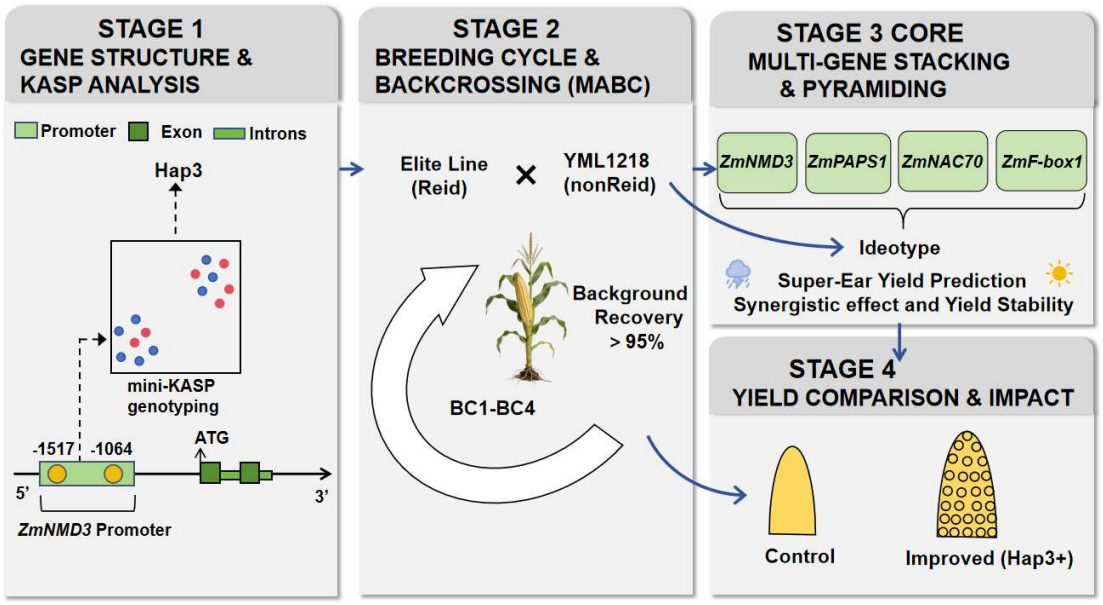


Fig. 12. Schematic representation of a post-validation molecular breeding design and gene stacking strategy based on the candidate haplotype Hap3 of *ZmNMD3*. Stage 1: If functional validation supports the effect of Hap3, high-throughput KASP markers could be developed targeting key SNPs located in predicted NAC-binding sites within the *ZmNMD3* promoter (−1517 bp and −1064 bp upstream of the ATG start codon). Stage 2: Marker-assisted selection (MAS) and early screening could be used to identify Hap3 carriers within backcross populations. Stage 3: Precision introgression (marker-assisted backcrossing, MABC) and gene pyramiding could be tested by introgressing the validated haplotype from non-Reid donor lines (e.g., YML1218) into elite Reid backgrounds,

887 followed by pyramiding with validated alleles of *ZmPAPS1*, *ZmNAC70*, and *ZmF-box1*. Stage 4: Multi-year and
888 multi-location field validation would be required before any commercialization assessment.

889 **4.7 Future research perspectives**

890 This study offers a first look at the genetic basis of KNPR, but moving from statistical association
891 to biological causality will demand rigorous experimental validation. We plan to use CRISPR/Cas9
892 genome editing to build targeted mutants and overexpression lines for the four candidate genes and to
893 quantify their contributions to KNPR across environments and years. At the molecular level, we intend
894 to apply chromatin immunoprecipitation (ChIP-seq) and dual-luciferase reporter assays to test the
895 proposed transcriptional cascade by which *ZmNAC70* may regulate its downstream genes *ZmNMD3* and
896 *ZmPAPS1*, and yeast two-hybrid (Y2H) assays to identify the substrate proteins that *ZmF-box1* targets
897 for degradation, thereby probing how those substrates might shape ear development. Beyond mechanism,
898 future work should evaluate low-frequency haplotypes with possible yield effects, such as Hap3, in
899 replicated backgrounds. One clear limitation is that our MPP used a single-row, non-replicated design
900 within each environment, which constrains the separation of true genetic variance from
901 microenvironmental error, including soil-fertility heterogeneity and border effects. Cross-environment
902 joint BLUP values reduced that noise to a degree, but future validation should adopt a randomized
903 complete block design (RCBD) with multi-row plots and at least three biological replicates per
904 environment to test the genetic effects of the core candidates, *ZmNMD3* especially, rigorously.

905 For the NILs and multi-gene pyramiding populations proposed here, follow-up work will run multi-
906 year, multi-location evaluations under high-density planting and marginal-stress conditions. Replicated
907 trials of this kind are what make it possible to disentangle genetic variance from environmental variance

and their interaction — the prerequisite for confirming that favorable variants deliver consistent yield gains before they enter commercial pipelines. Testing the stability of these variants across diverse backgrounds will be equally important, to ensure that gains in KNPR carry through to reliable single-plant grain yield and developmental robustness under variable conditions.

5 Conclusion

By integrating GWAS with linkage mapping in a maize multi-parent population (MPP), this study sharpened both the efficiency and the resolution of candidate-locus prioritization for KNPR. Because the MPP framework drew genetic variation from both Reid and non-Reid germplasm, it allowed us to prioritize four KNPR-associated genomic regions, on chromosomes 1, 3, 5, and 7.

Combined mapping and colocalization analyses then narrowed the list to four candidate genes — *ZmNMD3*, *ZmPAPS1*, *ZmNAC70,* and *ZmF-box1* — none previously linked to KNPR. Since the final candidates rest mainly on suggestive association signals, QTL colocalization, and functional annotation, they are best read as testable hypotheses rather than confirmed causal genes. Haplotype analysis showed that the *ZmNMD3* haplotype Hap3 was low in frequency and restricted to the non-Reid-derived subpopulation, underscoring the value of MPPs for uncovering rare alleles that conventional populations tend to miss. These resources should aid maize improvement: *ZmNMD3*-Hap3 and the regions around it, in particular, are promising targets for functional validation, marker development, and the eventual transfer of beneficial non-Reid alleles into elite breeding materials.

Supplementary Information

The online version contains supplementary material available for this article. The following supplementary tables are provided as separate files:

Table S1 Primer sequences used for qRT-PCR analysis.

Table S2 Key candidate loci missed by the strict genome-wide significance threshold ($-\log_{10}(P) = 5.66$).

Table S3 Significant SNP loci identified in the genome-wide association study (GWAS).

Table S4 Characteristics of the high-density genetic linkage maps for the four subpopulations.

Table S5 Comprehensive list of all annotated genes within the identified consistent loci.

Table S6 Statistical summary of relative expression analysis for candidate genes.

Fig. S1. Detailed haplotype analysis of candidate gene *ZmPAPS1*.

Fig. S2. Detailed haplotype analysis of candidate gene *ZmNAC70*.

Fig. S3. Detailed haplotype analysis of candidate gene *ZmF-box1*.

Acknowledgements

We thank all contributors to this research, especially the field and technical staff for their assistance with trials and data collection. We also thank the germplasm providers and gratefully acknowledge the funding agencies for their generous financial support.

Author contributions

X.F and F.J conceptualized and designed the research. Y.L, F.J, H.L, and X.Y performed the experiments and conducted the fieldwork. H.L and HW performed the formal analysis and curated the data. Y.L prepared the original draft of the manuscript. X.F and R.K.S provided critical revisions and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding

The research was supported by the Xingdian Talent Support Program of Yunnan Province, Yunnan Seed Laboratory (No. 202205AR070001), the High-level Scientific and Technological Talents and

Innovation Team Program of Yunnan Province (No. 202405AS350030), the Yunnan Key Laboratory of Tropical and Subtropical Maize Breeding (No. 20254916CE340055).

Data Availability

The raw whole-genome sequencing (WGS) data of the maize recombinant inbred lines (RILs) generated in this study have been deposited in the Genome Sequence Archive (GSA: CRA037261) at the National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences, and are publicly accessible at https://ngdc.cncb.ac.cn/gsa.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Ray D, Mueller N, West P, Foley J: Yield Trends Are Insufficient to Double Global Crop Production by 2050. PloS one 2013, 8:e66428.
2. Foley J, Ramankutty N, Brauman K, Cassidy E, Gerber J, Johnston M, Mueller N, O'Connell C, Ray D, West P *et al*: Solutions for a Cultivated Planet. Nature 2011, 478:337-342.
3. Andrade F, Vega C, Uhart SA, Cirilo A, Cantarero M, Valentinuz O: Kernel number determination in maize. Crop Sci 1999, 39:305–311.
4. Upadyayula N, da Silva HS, Bohn MO, Rocheford TR: Genetic and QTL analysis of maize tassel and ear inflorescence architecture. Theor Appl Genet 2006, 112(4):592–606.
5. Ma J, Cao Y: Genetic dissection of grain yield of maize and yield-related traits through association mapping and genomic prediction. Front Plant Sci 2021, 12:690059.

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(20):2755–2766.
7. Peng B, Li Y, Wang Y, Liu C, Liu Z, Tan W, Zhang Y, Wang D, Shi Y, Sun B *et al*: QTL analysis for yield components and kernel-related traits in maize across multi-environments. Theor Appl Genet 2011, 122(7):1305–1320.
8. Collard B, Jahufer Z, Brouwer JB, Pang E: An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 2005, 142:169-196.
9. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB *et al*: A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 2006, 38(2):203–208.
10. Lander ES, Botstein D: Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 1989, 121(1):185–199.
11. Liu R, Jia H, Cao X, Huang J, Li F, Tao Y, Qiu F, Zheng Y, Zhang Z: Fine mapping and candidate gene prediction of a pleiotropic quantitative trait locus for yield-related trait in *Zea mays*. PLoS One 2012, 7(11):e49836.
12. Wang Y, Bi Y, Jiang F, Shaw RK, Sun J, Hu C, Guo R, Fan X: Mapping and Functional Analysis of QTL for Kernel Number per Row in Tropical and Temperate-Tropical Introgression Lines of Maize (Zea mays L.). Curr Issues Mol Biol 2023, 45(5):4416-4430.
13. Huo D, Ning Q, Shen X, Liu L, Zhang Z: QTL Mapping of Kernel Number-Related Traits and Validation of One Major QTL for Ear Length in Maize. PLoS One 2016, 11(5):e0155506.
14. Zhang G, Wang X, Wang B, Tian Y, Li M, Nie Y, Peng Q, Wang Z: Fine mapping a major QTL for kernel number per row under different phosphorus regimes in maize (*Zea mays* L.). Theor Appl Genet 2013, 126(6):1545–1553.
15. An Y, Chen L, Li YX, Li C, Shi Y, Song Y, Zhang D, Li Y, Wang T: Candidate loci for the kernel row number in maize revealed by a combination of transcriptome analysis and regional association mapping. BMC Plant Biol 2019, 19(1):201.
16. Liu C, Zhou Y, Zhang X, Zhang J, Zhou Z, Weng J, Li X, Wang Z: Natural variation in the *THICK TASSEL DWARF1* (*TD1*) gene in the regulation of maize (*Zea mays* L.) ear-related traits. Breed Sci 2019, 69(2):323–331.
17. Scott MF, Ladejobi O, Amer S, Bentley AR, Biernaskie J, Boden SA, Clark M, Dell'Acqua M, Dixon LE, Filippi CV *et al*: Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. Heredity 2020, 125(6):396-416.
18. Xu Y, Li P, Yang Z, Xu C: Genetic mapping of quantitative trait loci in crops. The Crop Journal 2017, 5(2):175–184.
19. Shu G, Cao G, Li N, Wang A, Wei F, Li T, Yi L, Xu Y, Wang Y: Genetic variation and population structure in China summer maize germplasm. Sci Rep 2021, 11(1):8012.
20. Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R: A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis* thaliana. PLoS Genet 2009, 5(7):e1000551.

21. Li X, Zhou Z, Ding J, Wu Y, Zhou B, Wang R, Ma J, Wang S, Zhang X, Xia Z *et al*: Combined linkage and association mapping reveals QTL and candidate genes for plant and ear height in maize. Front Plant Sci 2016, 7:833.
22. R Core Team: R: A language and environment for statistical computing. In.: R Foundation for Statistical Computing, Vienna, Austria; 2023.
23. Bates D, Mächler M, Bolker B, Walker S: Fitting linear mixed-effects models using lme4. J Stat Softw 2015, 67(1):1–48.
24. Henderson CR: Best linear unbiased estimation and prediction under a selection model. Biometrics 1975, 31(2):423–447.
25. Schloerke B, Cook D, Larmarange J, Briatte F, Marbach M, Thoen E, Elberg A, Toomet O, Crowley J: GGally: Extension to 'ggplot2'. 2025.
26. Holland JB, Nyquist WE, Cervantes-Martinez CT: Estimating and interpreting heritability for plant breeding: an update. Plant Breed Rev 2003, 22:9–112.
27. Doyle J: A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 1987, 19:11–15.
28. Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009, 25(14):1754–1760.
29. Hufford M, Seetharam A, Woodhouse M, Chougule K, Ou S, Liu J, Ricci W, Guo T, Olson A, Qiu Y *et al*: De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. Science 2021, 373(6555):655-662.
30. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009, 25(16):2078–2079.
31. Wang K, Li M, Hakonarson H: ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010, 38(16):e164.
32. CNCB-NGDC Members: Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2025. Nucleic Acids Res 2025, 53(D1):D30-D44.
33. GSA Family Members: The GSA Family in 2025: A Broadened Sharing Platform for Multi-Omics and Multimodal Data. Genom Proteom Bioinform 2025, 23(4):qzaf072.
34. Tamura K, Stecher G, Kumar S: MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol 2021, 38(7):3022-3027.
35. Yang J, Lee SH, Goddard ME, Visscher PM: GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 2011, 88(1):76–82.
36. Alexander DH, Novembre J, Lange K: Fast model-based estimation of ancestry in unrelated individuals. Genome Res 2009, 19(9):1655–1664.
37. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z: GAPIT: genome association and prediction integrated tool. Bioinformatics 2012, 28(18):2397–2399.
38. Zhang C, Dong SS, Xu JY, He WM, Yang TL: PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics 2019, 35(10):1786–1788.

39. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S-y, Freimer NB, Sabatti C, Eskin E: Variance component model to account for sample structure in genome-wide association studies. Nat Genet 2010, 42(4):348–354.
40. Lander ES, Kruglyak L: Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 1995, 11:241-247.
41. Duggal P, Gillanders EM, Holmes TN, Bailey-Wilson JE: Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. BMC Genom 2008, 9:516.
42. Van Ooijen JW: JoinMap 4.0: software for the calculation of genetic linkage maps in experimental populations. Wageningen, the Netherlands.: Kyazma BV; 2006.
43. Kosambi DD: The Estimation of Map Distances from Recombination Values. In.; 1943: 125-130.
44. Zeng Z-B: Precision mapping of quantitative trait loci. Genetics 1994, 136(4):1457.
45. Lander E, Kruglyak L: Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 1995, 11(3):241-247.
46. Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ESt: Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci U S A 2001, 98(20):11479-11484.
47. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM *et al*: Twelve years of SAMtools and BCFtools. GigaScience 2021, 10(2):giab008.
48. Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005, 21(2):263–265.
49. Gabriel S, Schaffner S, Nguyen H, Moore J, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M *et al*: The Structure of Haplotype Blocks in the Human Genome. Science (New York, NY) 2002, 296:2225-2229.
50. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR *et al*: CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res 2017, 45(D1):D200–D203.
51. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS: MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 2009, 37(Web Server issue):W202–W208.
52. Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé P, Rombauts S: PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res 2002, 30(1):325–327.
53. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L *et al*: SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 2018, 46(W1):W296–W303.
54. Schrödinger L: The PyMOL Molecular Graphics System, Version 2.5. In. New York: Schrödinger, LLC; 2023.
55. Dong Z, Wang Y, Bao J, Li Yn, Yin Z, Long Y, Wan X: The genetic structures and molecular mechanisms underlying ear traits in maize (*Zea mays* L.). Cells 2023, 12:1900.

56. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL: Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 2012, 13:134.
57. Livak KJ, Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ Method. Methods 2001, 25(4):402–408.
58. VanRaden P: Efficient Methods to Compute Genomic Predictions. J Dairy Sci 2008, 91:4414-4423.
59. Wickham H: ggplot2: elegant Graphics for data analysis, 2nd edn. Cham: Springer; 2016.
60. Li W, Boer MP, Joosen RVL, Zheng C, Percival-Alwyn L, Cockram J, Van Eeuwijk FA: Modeling QTL-by-environment interactions for multi-parent populations. Front Plant Sci 2024, 15:1410851.
61. Ma F, Xu Y, Wang R, Tong Y, Zhang A, Liu D, An D: Identification of major QTLs for yield-related traits with improved genetic map in wheat. Front Plant Sci 2023, 14:1138696.
62. Manjunath KK, Krishna H, Devate NB, Sunilkumar VP, Patil SP, Chauhan D, Singh S, Kumar S, Jain N, Singh GP *et al*: QTL mapping: insights into genomic regions governing component traits of yield under combined heat and drought stress in wheat. Front Genet 2024, 14:1282240.
63. Kontou PI, Bagos PG: The goldmine of GWAS summary statistics: a systematic review of methods and tools. BioData Min 2024, 17(1):31.
64. Wei X, Chen M, Huang X: Genomic investigation of 18,421 lines reveals the genetic architecture of rice. Science 2024, 385:eadm8762.
65. Yang N, Lu Y, Yang X, Huang J, Zhou Y, Ali F, Wen W, Liu J, Li J, Yan J: 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 2014, 10(9):e1004573.
66. Shen X, Liu L, Tran T, Ning Q, Li M, Huang L, Zhao R, Li Y, Qing X, Jackson D *et al*: KRN5b regulates maize kernel row number through mediating phosphoinositol signalling. Plant Biotechnol J 2024, 22(12):3427-3441.
67. Chen J, Zhang L, Liu S, Li Z, Huang R, Li Y, Cheng H, Li X, Zhou B, Wu S *et al*: The Genetic Basis of Natural Variation in Kernel Size and Related Traits Using a Four-Way Cross Population in Maize. PLoS One 2016, 11(4):e0153428.
68. Zhang X, Sun J, Zhang Y, Li J, Liu M, Li L, Li S, Wang T, Shaw RK, Jiang F *et al*: Hotspot regions of quantitative trait loci and candidate genes for ear-related traits in maize: a literature review. Genes 2023, 15(1):15.
69. Sengupta J, Bussiere C, Pallesen J, West M, Johnson A, Frank J: Characterization of the nuclear export adaptor protein Nmd3 in association with the 60S ribosomal subunit. J Cell Biol 2010, 189:1079-1086.
70. Shi Y, Liu X, Li R, Gao Y, Xu Z, Zhang B, Zhou Y: Retention of OsNMD3 in the cytoplasm disturbs protein synthesis efficiency and affects plant development in rice. J Exp Bot 2014, 65.
71. Laishram R: Poly(A) polymerase (PAP) diversity in gene expression – Star-PAP vs canonical PAP. FEBS Lett 2014, 588.
72. Vi SL, Trost G, Lange P, Czesnick H, Rao N, Lieber D, Laux T, Gray WM, Manley JL, Groth D *et al*: Target specificity among canonical nuclear poly(A) polymerases in plants modulates organ growth and pathogen response. Proc Natl Acad Sci USA 2013, 110(34):13994–13999.

73. Li W, Li W, Laishram R, Hoque M, Ji Z, Tian B, Anderson R: Distinct regulation of alternative polyadenylation and gene expression by nuclear poly(A) polymerases. Nucleic Acids Res 2017, 45:8930-8942.
74. Fan K, Wang M, Miao Y, Ni M, Bibi N: Molecular Evolution and Expansion Analysis of the NAC Transcription Factor in Zea mays. PLoS One 2014, 9(11):-.
75. Puranik S, Sahu PP, Srivastava PS, Prasad M: NAC proteins: regulation and role in stress tolerance. Trends Plant Sci 2012, 17(6):369–381.
76. Kaneda T, Taga Y, Takai R, Iwano M, Matsui H, Takayama S, Isogai A, Che FS: The transcription factor OsNAC4 is a key positive regulator of plant hypersensitive cell death. EMBO J 2009, 28(7):926-936.
77. Aguilar J, Matilla A: Transcriptional Control of Seed Life: New Insights into the Role of the NAC Family. Int J Mol Sci 2024, 25:5369.
78. Hu Y, Li C, Zhou R, Song Y, Lv Z, Wang Q, Dong X, Liu S, Feng C, Zhou Y *et al*: The Transcription Factor ZmNAC89 Gene Is Involved in Salt Tolerance in Maize (Zea mays L.). Int J Mol Sci 2023, 24:15099.
79. Wang N, Cheng M, Chen Y, Liu B, Wang X, Li G, Zhou Y, Luo P, Xi Z, Yong H *et al*: Natural variations in the non-coding region of ZmNAC080308 contributes maintaining grain yield under drought stress in maize. BMC Plant Biol 2021, 21:305.
80. Sun Q, Li Y, Gong D, Hu A, Zhong W, Zhao H, Ning Q, Tan Z, Liang K, Mu L *et al*: A NAC-EXPANSIN module enhances maize kernel size by controlling nucellus elimination. Nat Commun 2022, 13(1):5708.
81. Yu H, Wu J, Xu N, Peng M: Roles of F-box Proteins in Plant Hormone Responses. Acta Biochim Biophys Sin 2008, 39:915-922.
82. Ito T, Okada K, Fukazawa J, Takahashi Y: DELLA-dependent and -independent gibberellin signaling. Plant Signal Behav 2018, 13(3):e1445933.
83. Kuroda H, Yanagawa Y, Takahashi N, Horii Y, Matsui M: A comprehensive analysis of interaction and localization of Arabidopsis SKP1-like (ASK) and F-box (FBX) proteins. PLoS One 2012, 7:e50009.
84. Li M, Zheng Y, Cui D, Du Y, Zhang D, Sun W, Du H, Zhang Z: GIF1 controls ear inflorescence architecture and floral development by regulating key genes in hormone biosynthesis and meristem determinacy in maize. BMC Plant Biol 2022, 22.
85. Wang Y, Li Y, Zhang W, Yang Y, Ma Y, Li X, Meng D, Luo H, Xue W, Lv X *et al*: BSA-Seq and Transcriptomic Analysis Provide Candidate Genes Associated with Inflorescence Architecture and Kernel Orientation by Phytohormone Homeostasis in Maize. Int J Mol Sci 2023, 24:10728.
86. Chaudhry A, Chen Z, Gallavotti A: Hormonal influence on maize inflorescence development and reproduction. Plant Reprod 2024, 37:393-407.
87. Bommert P, Lunde C, Nardmann J, Vollbrecht E, Running M, Jackson D, Hake S, Werr W: thick tassel dwarf1 encodes a putative maize ortholog of the Arabidopsis CLAVATA1 leucine-rich repeat receptor-like kinase. Development 2005, 132(6):1235-1245.
88. Je BI, Xu F, Wu Q, Liu L, Meeley R, Gallagher J, Corcilius L, Payne R, Bartlett M, Jackson D: The CLAVATA receptor FASCIATED EAR2 responds to distinct CLE peptides by signaling through two downstream effectors. eLife 2018, 7.

89. Bommert P, Satoh-Nagasawa N, Jackson D: Quantitative variation in maize kernel row number is controlled by the FASCIATED EAR2 locus. Nat Genet 2013, 45:334–337.
90. Du Y, Liu L, Peng Y, Li M, Li Y, Liu D, Li X, Zhang Z: UNBRANCHED3 Expression and Inflorescence Development is Mediated by UNBRANCHED2 and the Distal Enhancer, KRN4, in Maize. PLoS Genet 2020, 16(4):e1008764.
91. Chen W, Chen L, Zhang X, Yang N, Guo J, Wang M, Ji S, Zhao X, Yin P, Cai L *et al*: Convergent selection of a WD40 protein that enhances grain yield in maize and rice. Science (New York, NY) 2022, 375:eabg7985.

Supplementary Files

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

- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx
- TableS5.xlsx
- TableS6.xlsx
- Fig.S1.pdf
- Fig.S2.pdf
- Fig.S3.pdf