Plant Communications
Resource article

CellPress
Partner Journal

Reverse BSA-QTLseq: A new genotype-driven bioinformatics approach for simultaneous trait mapping

Salvatore Esposito[1,2], Nunzio D'Agostino[3], Francesca Taranto[4], Fabio Fania[2], Stefano Pavan[5], Ida Colella[2,6], Francesco Sestili[6], Domenico Lafiandra[6] and Pasquale De Vita[2,*]

[1]Institute of Biosciences and Bioresources (CNR-IBBR), 80055 Portici, Naples, Italy

[2]Council for Agricultural Research and Economics (CREA), Research Centre for Cereal and Industrial Crops (CREA-CI), 71122 Foggia, Italy

[3]Department of Agricultural Sciences, University of Naples Federico II, 80055 Portici, Italy

[4]Institute of Biosciences and Bioresources (CNR-IBBR), 70126 Bari, Italy

[5]Department of Soil, Plant and Food Sciences, University of Bari Aldo Moro, 70126 Bari, Italy

[6]Department of Agriculture and Forest Sciences, University of Tuscia, 01100 Viterbo, Italy

***Correspondence: Pasquale De Vita (pasquale.devita@crea.gov.it)**

https://doi.org/10.1016/j.xplc.2025.101588

ABSTRACT

Bulked segregant analysis (BSA) is a widely used method for identifying genomic loci associated with traits of interest in crops. However, conventional BSA is limited by its reliance on phenotype-driven bulk sampling, which restricts its scalability and confines its applicability to single-trait analysis. This study introduces a novel method, reverse BSA-QTLseq, which uses genotype-driven bulk reconstruction through bioinformatics, enabling the simultaneous mapping of multiple traits from the same genotypic dataset. Reverse BSA-QTLseq uses a two-step strategy—low-resolution genotyping of the entire population followed by high-resolution sequencing of selected bulks—enabling cost-effective identification of genetically divergent lines to enhance the discovery of quantitative trait loci (QTLs). Using a bread wheat recombinant inbred line (RIL) population as a case study, we mapped loci associated with heading date and plant height , confirming approximately 95% of known QTLs, including both dwarfing genes (e.g., *Rht-B1* and *Rht-5*) and flowering-time regulators (e.g., *Vrn-A1*), and identified novel QTLs and candidate loci with strong phenotypic effects. The phased genotyping strategy maximized genetic distance in the initial sampling, facilitating the *in silico* reconstruction of trait-specific contrasting bulks. Integration of transcriptional profiles from the parental lines of the RIL population, from which the bulks were derived, aided in identifying candidate genes and regulatory networks underlying the variation of traits such as photoperiod response, nutrient transport, and stress adaptation. The versatility and potential for data reuse offered by the proposed method represent a significant advancement in QTL mapping, with broad implications for marker-assisted breeding and selection programs. Future integration of transcriptomic and epigenomic data is expected to further enhance the power of reverse BSA-QTLseq, accelerating genetic improvement in crops.

Key words: bread wheat, bulked segregant analysis, exome capture, single-nucleotide polymorphisms, RNA sequencing, multi-omics, plant height, heading time

Esposito S., D'Agostino N., Taranto F., Fania F., Pavan S., Colella I., Sestili F., Lafiandra D., and De Vita P. (2026). Reverse BSA-QTLseq: A new genotype-driven bioinformatics approach for simultaneous trait mapping. Plant Comm. **7**, 101588.

INTRODUCTION

Understanding the genetic basis of agronomic traits is crucial for advancing crop improvement programs, particularly for staple crops such as bread wheat (*Triticum aestivum* L.). One of the most effective tools in this domain is bulked segregant analysis (BSA), a widely used genetic mapping technique for identifying genomic regions associated with specific traits. Conventional

BSA involves pooling DNA from individuals with contrasting phenotypes and genotyping the pooled samples to detect genetic markers linked to the target trait (Michelmore et al., 1991; Sham et al., 2002; Zou et al., 2016). With the advent of next-generation sequencing (NGS), BSA has been significantly refined. These advances not only reduce gaps in recombination mapping but also increase resolution, enabling the identification of candidate regions with small effects that might be overlooked using traditional marker systems (Zou et al., 2016; Itoh et al., 2019; Li and Xu, 2022; Majeed et al., 2022; Wang et al., 2023).

Various NGS-enabled BSA derivatives have subsequently emerged. Among them, quantitative trait locus (QTL) sequencing (QTL-seq)—an approach that integrates BSA with whole-genome sequencing—has been used across a wide range of crops, including rice (Arikit et al., 2019; Lei et al., 2020), wheat (Xin et al., 2020; Komura et al., 2024a), tomato (Ruangrak et al., 2019), groundnut (peanut; Pandey et al., 2017), chickpea (Deokar et al., 2019), sunflower (Imerovski et al., 2019), squash (Ramos et al., 2020), and watermelon (Branham et al., 2018). Similarly, MutMap is a specialized form of BSA designed for mutant populations, enabling the identification of causal mutations through comparative sequencing of mutants and wild-type individuals (Abe et al., 2012; Komura et al., 2024b). For transcriptome-level analysis, bulked segregant RNA sequencing (RNA-seq) uses RNA-seq data to identify expression QTLs by analyzing differentially expressed genes (DEGs) in RNA pools from individuals with contrasting phenotypes (Huang et al., 2020). This approach has been used to map resistance genes for stripe rust (Wang et al., 2017) and powdery mildew (Zhu et al., 2020; Ma et al., 2021; Qian et al., 2024; Yu et al., 2024). Unlike DNA-based BSA methods, bulked segregant RNA-seq focuses on RNA, allowing researchers to identify genes or loci that are differentially expressed and potentially involved in phenotypic variation.

Recent innovations such as GradedPool-seq (Wang et al., 2019), OcBSA (Zhang et al., 2024), haplotagging BSA (Majeed et al., 2022), and DeepBSA (Li and Xu, 2022) have further enhanced mapping resolution or broadened the applicability of BSA to complex populations and traits (Chen et al., 2021; Segawa et al., 2021; Li and Xu, 2022; Majeed et al., 2022). The success of these approaches underscores the transformative potential of NGS-based BSA methods in accelerating genetic improvement in crops. However, despite these advances, BSA methods still rely on phenotype-driven bulk formation and remain limited in flexibility. Once individuals are pooled based on a specific phenotypic trait, the samples cannot be reused to analyze other traits, and phenotypic preselection remains the most labor-intensive step.

To overcome these constraints, we propose a reverse BSA-QTLseq approach that shifts the focus from phenotypic to genotypic variation. This method enables the *in silico* reconstruction of contrasting bulks post-sequencing, allowing for the analysis of multiple traits from the same genotypic dataset. Our rationale is based on Michelmore et al. (1991), who developed BSA using the principle that contrasting phenotypes reflect contrasting alleles in the corresponding QTL, with success depending on parental divergence in the target region. This concept has been further reinforced by computational models from Shen and Messer (2022) demonstrating that the power of BSA to detect causal loci is maximized when the bulks consist of individuals homozygous for opposite alleles at the QTL. According to Michelmore et al. (1991), pooling individuals with divergent phenotypes results in allele frequency differences that primarily occur at trait-associated loci, whereas other genomic regions remain largely unaffected due to random segregation. Initiating BSA with genetically divergent, homozygous individuals increases the likelihood of capturing contrasting alleles at relevant loci, thereby increasing both the allele frequency divergence and the signal-to-noise ratio. Furthermore, even in the presence of phenotyping noise or residual heterozygosity, the strength of association remains proportional to the allelic divergence between bulks, a relationship that holds for both traditional BSA (based on phenotypic extremes) and reverse BSA-QTLseq (based on genotypic divergence). Based on this relationship, we used principal component analysis (PCA) on low-density single-nucleotide polymorphism (SNP) data to guide the selection of genetically divergent lines for sequencing. This strategy enhances the likelihood of capturing informative variation across all contrasting traits of the original parents by reversing the core principle of allelic homozygosity from traditional BSA, instead selecting genotypically divergent individuals to maximize phenotypic contrast detection.

Specifically, reverse BSA-QTLseq uses a two-step strategy—low-resolution genotyping of the full population followed by high-resolution genotyping of selected core lines—that is both cost effective and scalable. This enables researchers to use inexpensive genotyping tools to narrow down a manageable set of samples for more intensive analysis. Consequently, initial mapping can be conducted without full-scale phenotyping, allowing us to prioritize a smaller subset of individuals (or bulks) for downstream phenotypic or transcriptomic analysis while still adhering to the theoretical principles of the original BSA framework.

To evaluate the performance and applicability of the proposed framework, we designed the study with two complementary validation strategies. First, we applied reverse BSA-QTLseq to two well-characterized traits in wheat, heading date (HD) and plant height (PH), using a subset of lines from a biparental recombinant inbred line (RIL) population for which conventional QTL mapping results from both F_2:F_3 and RIL generations were previously published (Vitale et al., 2021). This internal benchmark aimed to (i) test whether genotype-driven bulks could recover the major QTLs identified through full-population analysis, (ii) demonstrate the method's adaptability to multiple traits within a single genotyped population, (iii) identify chromosomal regions associated with the target traits, and (iv) confirm the findings by identifying diagnostic molecular markers, thereby enhancing the utility of this approach for QTL discovery. Second, we conducted an independent validation using an RIL population of pea (*Pisum sativum* L.) for which publicly available genotyping by sequencing (GBS) data and phenotypic evaluations for resistance to broomrape (*Orobanche crenata* Forsk.) were available (Delvento et al., 2023). This independent case study served to assess the generalizability of the method to a different species, genome structure, and phenotyping context while focusing on a single disease-resistance trait.

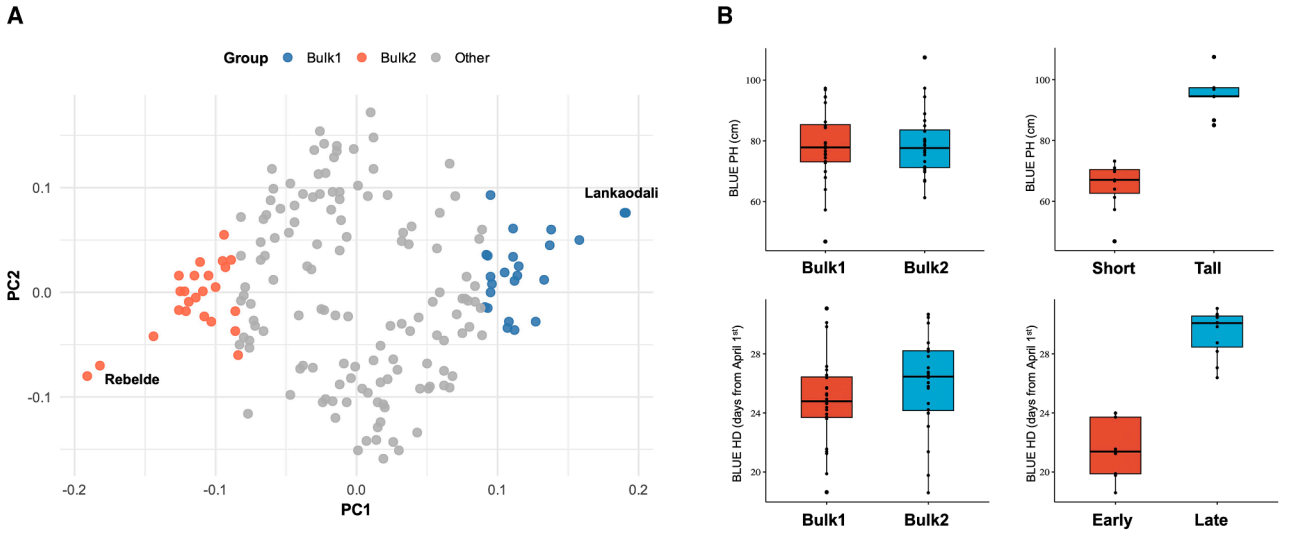

Figure 1. Genetic divergence and phenotypic distribution defining reverse BSA-QTLseq bulks

(A) Genetic distance analysis of the segregating population identified two distinct bulks, designated as Bulk 1 and Bulk 2, which capture the maximum genetic divergence within the population. **(B)** BLUE values for plant height (PH, cm; top) and heading date (HD, days from April 1; bottom) derived from phenotypic data across four growing seasons. Data were collected under rain-fed conditions for the first 2 years and under both rain-fed and well-irrigated conditions for the final 2 years. The left-hand panels display BLUE values for the PCA-based clustering bulks (Bulk 1 and Bulk 2), whereas the right-hand panels show BLUE values for phenotypically reordered bulks for PH (high and low) and HD (late and early).

RESULTS

Case study 1: Genetic divergence and trait mapping in bread wheat RILs

Phenotyping and bulk reconstruction

The initial genetic material consisted of 176 RILs derived from a cross between two phenotypically contrasting bread wheat varieties, Lankaodali and Rebelde, as previously described by Vitale et al. (2021). PCA, performed on 6795 high-quality SNPs (using a 25K iSelect Array), showed separation of the parental genotypes along PC1 (23.4% variance) and revealed two groups representing the extremes of overall genomic divergence (Figure 1A). Twenty-three lines from each cluster, plus their corresponding parents, were chosen to constitute Bulk 1 and Bulk 2 (Supplemental Table 1). This 25% sampling fraction is consistent with theoretical work showing near-optimal QTL detection power for bulks comprising 10%–30% of the population (Shen and Messer 2022). Recombination rates and runs of homozygosity were then estimated for the selected bulks and compared to those of the phenotypic tails of the full RIL population (Supplemental Table 3). Group averages were similar, indicating that reverse BSA-QTLseq selected individuals with recombination landscapes comparable to those in the phenotype-based bulks (Supplemental Table 3).

The best linear unbiased estimate (BLUE) values for HD and PH, measured for the 48 lines, demonstrated substantial variability. PH BLUE values ranged from 46 to 107 cm, whereas HD BLUE values ranged from 18 to 31 days from April 1. As expected, the original bulks (Bulk 1 and Bulk 2) did not exhibit statistically significant differences for either trait (Figure 1B), validating the hypothesis that sampling based on genetic distance maintained the variability of the initial RIL population (Vitale et al., 2021). In contrast, the reconstructed bulks for HD (Bulk A and Bulk B) and PH (Bulk C and Bulk D) displayed significant differences for both traits, consistent with the classical BSA requirement for distinct phenotypic contrasts between bulks. Consequently, two bulks, one for PH and one for HD, each comprising at least 10 contrasting lines, were formed and used for subsequent BSA-QTLseq analysis.

Whole-exome sequencing and variant annotation

Raw data from the 48 selected samples retrieved from Esposito et al. (2022) yielded 768 490 high-quality SNPs (Supplemental Figure 1A). As noted by Esposito et al. (2022), some variants originated from non-target regions, despite the bait design being based exclusively on annotated exons. This is likely due to the extended regions captured within each bait (Supplemental Figure 1B). After identification of sequence variations, SNPs and insertions or deletions (indels) located in annotated exons were classified according to their predicted biological effects (Supplemental Figure 1C). As expected, the majority of identified variants fell within exons and had minimal effects on protein function (Supplemental Figure 1B and 1C). Variants with deleterious effects on coding regions were also found (Supplemental Figure 1C). Among these, stop-gained variants were the most abundant, followed by stop-loss and start-loss variants.

Genomic regions putatively associated with plant height (PH)

Twenty-eight QTLs for PH were identified on eight chromosomes using both SNPs and indels (Figure 2 and Supplemental Table 2). On chromosome 2A, the region spanning 224–468 Mb was divided into nine clusters (*qPH.2A.1–qPH.2A.9*) based on Δ(Variant index) values and the 99% confidence interval (Figure 2 and Supplemental Table 2). All of these exhibited negative Δ(Variant index) values ranging from −0.38 to −0.42. Similarly, eight QTL clusters (*qPH.4B.1–qPH.4B.8*) were found on chromosome 4B, all showing positive Δ(Variant index) values ranging from 0.42 to 0.48 (Supplemental Table 2). Five QTLs were identified on chromosome 4D (*qPH.4D.1–qPH.4D.5*). Those

Figure 2. QTLs identified for plant height (PH) at maturity.

(A and B) The distribution of Δ(Variant index) values across chromosomes for single-nucleotide polymorphisms (SNPs) **(A)** and insertions and deletions (indels) **(B)**. The *x* axis represents chromosomes, labeled from 1A to 7D (Unk: unmapped regions). Blue and red horizontal lines denote the 95% and 99% confidence intervals, respectively, indicating significant QTL regions.

detected by SNPs (*qPH.4D.1*, *qPH.4D.3*, and *qPH.4D.4*) exhibited positive Δ(Variant index) values, whereas the two identified by indels (*qPH.4D.2* and *qPH.4D.5*) showed negative values (Figure 2 and Supplemental Table 2). One QTL was also identified on each of chromosomes 2B, 3A, 7B, and 7D (Supplemental Table 2). Among these, *qPH.3A* and *qPH.7D* were identified by indels and both exhibited a negative Δ(Variant index) value (Supplemental Table 2).

Genomic regions putatively associated with heading date (HD)

A total of 30 QTLs were identified across seven chromosomes using both SNPs and indels, with those on chromosomes 5A and 7D showing the lowest *p* values (< 0.01) (Figure 3 and Table S1). Specifically, on chromosome 5A, nine clusters (*qHD.5A.1–qHD.5A.9*) were annotated, among which *qHD.5A.5* and *qHD.5A.7* exhibited the highest positive average ΔVariant index values (0.45 and 0.46). In contrast, *qHD.5A.6* and *qHD.5A.8*, both identified by indels, showed negative Δ(Variant index) values (−0.39 and −0.43). Eight QTL clusters (*qHD.3B.1–qHD.3B.8*) were annotated on chromosome 3B, although all regions displayed low Δ(Variant index) values (Figure 3 and Table 1).

Four QTLs, all identified by SNPs, were located on chromosome 6A, each showing positive Δ(Variant index) values (Table 1). Three QTLs were found on each of chromosomes 3A and 5B using both SNPs and indels (Figure 3). Interestingly, QTLs identified by SNPs (*qHD.3A.1*, *qHD.3A.3*, *qHD.5B.1*, *qHD.5B.2*, and *qHD.5B.3*) showed positive Δ(Variant index) values, whereas *qHD.3A.2* exhibited a negative value (Figure 3 and Table 1). Finally, a

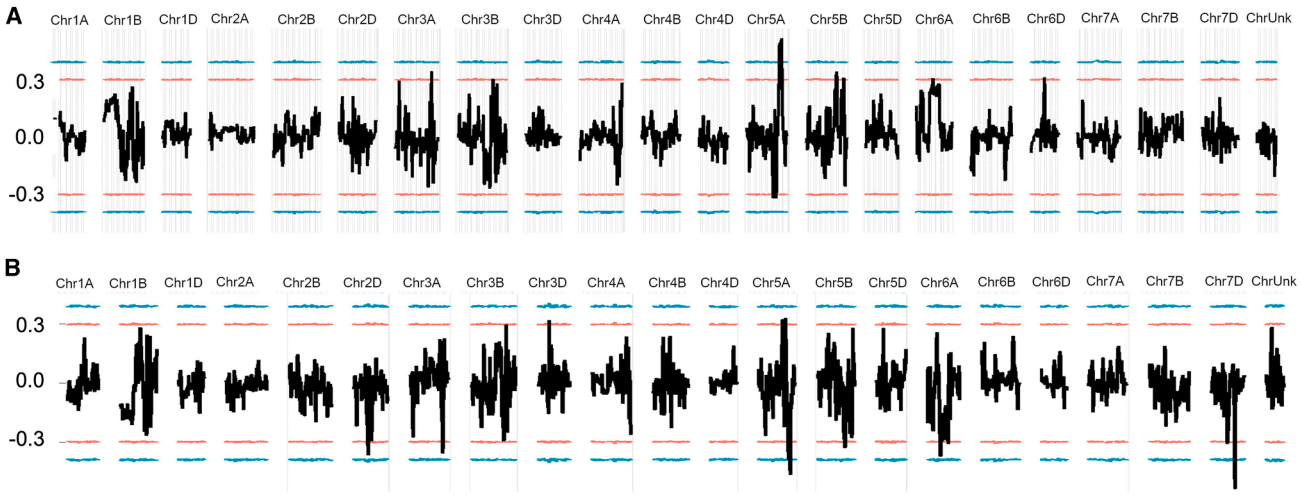

Figure 3. QTLs identified for heading date (HD) at maturity.

(A and B) The distribution of Δ(Variant index) values across chromosomes for single-nucleotide polymorphisms (SNPs) **(A)** and insertions and deletions (indels) **(B)**. The *x* axis represents chromosomes, labeled from 1A to 7D (Unk: unmapped regions). Blue and red horizontal lines denote the 95% and 99% confidence intervals, respectively, indicating significant QTL regions.

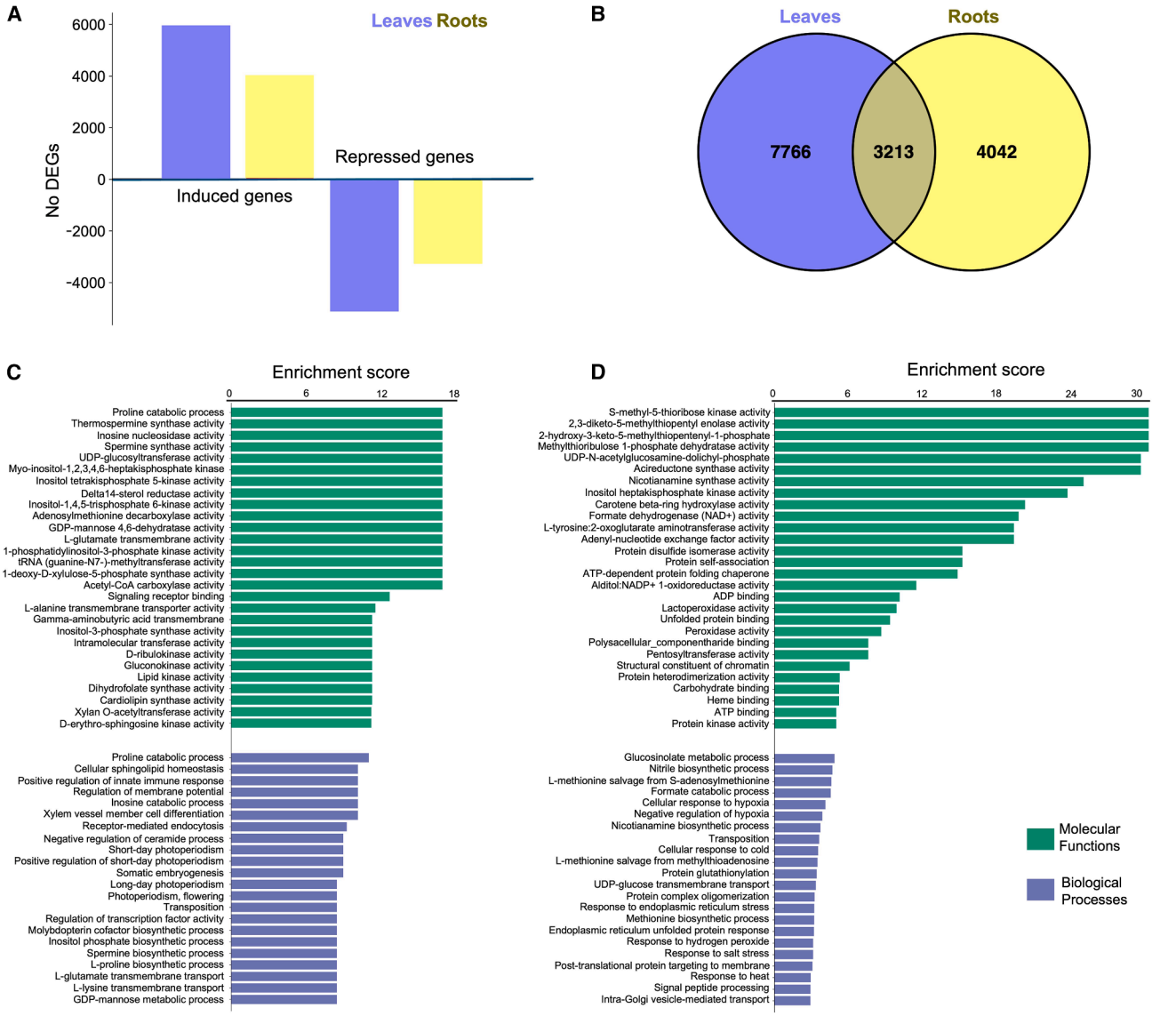

Figure 4. Overview of the transcriptional landscape in Lankaodali and Rebelde.

(A) Bar plots showing the number of differentially expressed genes (DEGs) in leaves and roots, with Lankaodali used as the reference.

(B) Venn diagram illustrating the overlap and tissue specificity of DEGs between leaves and roots.

(C and D) Enriched Gene Ontology (GO) terms associated with DEGs in leaves and roots. Colors indicate GO categories in the molecular function and biological process domains.

single region on chromosome 7D (*qHD.7D.1*), with a peak at 459 Mb, showed a Δ(Variant index) of −0.54, which suggests the presence of major genes or regulatory loci strongly linked with HD (Figure 3 and Table 1).

Transcriptome sequencing of tissues from parental lines

To investigate the molecular mechanisms underlying the phenotypic differences between the parental lines, 12 cDNA libraries from leaf and root tissues were sequenced, each with three biological replicates. After filtering out low-quality sequences and adapters, a high-quality 37.82-Gb dataset was obtained and used to identify 10 979 DEGs in leaves and 7255 in roots (Figure 4A and Supplemental Tables 4 and 5). Comparison between leaves and roots revealed 3213 shared DEGs (21.4%), whereas 7766 (51.7%) and 4042 (26.9%) were leaf- and root-specific, respectively (Figure 4B). To understand the function of the DEGs, we performed Gene Ontology (GO) enrichment analysis and categorized the top 50 enriched terms into molecular function and biological process domains (Figure 4C). The enriched GO terms confirmed differences in the genetic backgrounds of the parental lines at the transcriptomic level. Notably, several GO terms related to photoperiodism (e.g., short-day photoperiodism and flowering) were enriched in leaves (Figure 4C). In contrast, GO terms associated with plant defense responses (e.g., response to fungus and response to salt stress) were enriched in roots, which suggests that the parents may differ in resistance to abiotic and biotic stresses (Figure 4D). Given the importance of *Vrn* and *Ppd* genes, their expression levels were investigated (Supplemental Figure 2). In leaves, *Vrn-A1* and *Vrn-B1* were expressed at higher levels in Lankaodali compared to Rebelde. In contrast, *Vrn-2* (*Vrn-2.1* and *Vrn-2.2*)

transcripts were detected only in Rebelde, albeit at low levels (Supplemental Figure 2A). *Ppd-A1* and *Ppd-D1* showed similar expression trends in both lines, whereas *Ppd-B1* was expressed only in Lankaodali (Supplemental Figure 2A).

In roots, *Vrn-1* genes were barely expressed, with *Vrn-A1* being the only one detected (Supplemental Figure 2B). Similarly, among the *Vrn-2* genes, only *Vrn-2.2* was expressed exclusively in Lankaodali. Interestingly, *Ppd-1* genes were expressed at higher levels in roots than in leaves (Supplemental Figure 2B). *Ppd-D1* showed the highest expression in both lines, whereas *Ppd-B1* was expressed at a higher level in Lankaodali (Supplemental Figure 2B).

Candidate gene identification and prioritization

Out of approximately 5500 genes putatively associated with PH, 3300 carried variants located within or near coding sequences, with the number of variants per gene ranging from 1 to 346. Most variants were predicted to have a modifier impact (e.g., intronic, upstream, or downstream variants) (N = 27 401), followed by low-impact (N = 6504), moderate-impact (N = 6397), and high-impact (N = 164) variants (Supplemental Table 6). In addition, among the 5500 genes, 141 were differentially expressed between parental lines in both leaf and root tissues, whereas 321 and 170 were specific to leaves and roots, respectively. Among the shared DEGs, 93 genes harbored 529 low-impact variants, 80 genes carried 373 moderate-impact variants, and 12 contained high-impact deleterious mutations. Notable, shared DEGs included *TraesCS4B03G0290400* (a PHD domain–containing gene), *TraesCS2B03G0403600* (an F-box gene), *TraesCS4B03G0117800* (an ABC transporter gene), *TraesCS2B03G0441300* (a disease-resistance NBS-LRR–type gene; a putative calcium-binding *CML45* homolog), and *TraesCS4A03G1116300* (a sucrose synthase gene) (Supplemental Table 6).

Similarly, among approximately 7000 genes potentially associated with HD, 4527 harbored variants predicted to have a modifier impact, 2509 harbored variants with low impact, and 217 harbored variants with high impact. Of these, 265 genes were differentially expressed between parental lines in both leaf and root tissues, whereas 457 and 266 DEGs were specific to leaves and roots, respectively. Among the shared DEGs, 251 contained variants with modifier impact, 154 with moderate impact, 184 with low impact, and 17 with high impact (Supplemental Table 6). Notable genes harboring high-impact variants included *TraesCS5A03G0634200* and *TraesCS5B03G1226800* (F-box genes), *TraesCS4A03G1221200* (a DEAD-box ATP-dependent helicase gene), *TraesCS5A03G1012900* (an aldehyde oxidase gene), *TraesCS4A03G1214600* (the dual-specificity protein kinase *ZAK2*), and *TraesCS5B03G0863800* and *TraesCS5D03G1233100* (disease resistance genes), all of which were differentially expressed in both tissues (Supplemental Table 6). The expression of the candidate genes was further evaluated using 407 publicly available transcriptomic datasets related to spike development (from the floret meristem stage to the booting stage) (Chen et al., 2021), grain development (spanning the two-cell to mature stages [E1–E7], endosperm [E8–E9], and pericarp [E10]) (Xiang et al., 2019), and leaf senescence (from 3 to 10 days after anthesis) (Borrill et al., 2019) (Supplemental Figures 3 and 4). Among the candidate genes identified, *TraesCS4A03G1116300* (the sucrose synthase gene), *TraesCS3B03G0008500* (a subtilisin-like protease gene), *TraesCS4B03G0290400* (the PHD domain–containing gene), and *TraesCS4D03G0456300* (*EMBRYONIC FLOWER 2*, a VEFS-box gene) were expressed across all five tissues analyzed (root, shoot, leaf, spike, and grain) (Supplemental Figure 3). Notably, *TraesCS4B03G0290400* (the PHD domain–containing gene), *TraesCS4D03G0456300* (*EMF2B, a* Polycomb group gene), and *TraesCS2B03G0441300* (the putative calcium-binding *CML45* homolog), were expressed during leaf senescence stages, which suggests potential roles in regulating plant growth. In addition, *TraesCS4A03G1116300* and *TraesCS4D03G0456300* exhibited intriguing expression patterns, with the former showing increased expression from spike development stages B6 to B11 and the latter displaying a slight decrease over time (Supplemental Figure 3).

Genomic variants located in the coding regions of candidate genes were evaluated using *t*-tests (Figure 5). The population was divided into two groups based on their allelic profiles to determine whether the mean BLUE values differed significantly. In total, nine variants were found to have a significant effect on PH (Figure 5). The largest number of variants with significant effects (N = 3) was identified on chromosome 2B. These variants were located in the coding regions of *TraesCS2B03G0441300* (the putative calcium-binding *CML45* homolog), *TraesCS2B03G0441600* (a 3-isopropyl malate dehydratase/aconitase gene), and *TraesCS2B03G0442800* (an NB-ARC domain–containing, NBS-LRR disease resistance gene).

Two additional PH-associated variants were found on chromosome 3B, within the gene bodies of *TraesCS3B03G0099800* and *TraesCS3B03G0104900*, which encode an LRR receptor-like serine/threonine kinase and a deoxyhypusine synthase, respectively. Similarly, two significant variants were found on chromosome 7B, within *TraesCS7B03G1037100* (a MYB transcription factor gene) and *TraesCS7B03G1031800* (an F-box gene similar to *At5g07610*). The remaining PH-associated variants were located on chromosomes 4B and 4D (Figure 5). On chromosome 4B, the variant was found in *TraesCS4B03G0030000*, which encodes homocysteine methyltransferase 2, whereas the variant on chromosome 4D was located in *TraesCS4D03G0561900*, which encodes the NRT1/PTR FAMILY 8.3 transporter. In addition, four variants significantly associated with HD, all predicted to have deleterious effects, were identified on chromosomes 5A, 5B, and 6B (Figure 5). These were located within *TraesCS5A03G0903600* (encoding carnitine transporter 7), *TraesCS5B03G0887900* (the serine/threonine protein kinase *SD1-8*), *TraesCS5B03G1259700* (a heat shock protein 70 gene), and *TraesCS6B03G0032400* (which has an unknown function) (Figure 5).

To assess whether candidate genes influence the phenotypes of interest, we screened the Cadenza TILLING population for polymorphisms in closely related genes. We identified the line Cadenza1329, which carries a high-impact nonsense mutation (W/*) at the start of the second exon of the *TraesCS2B03G0403600* gene. Homozygous mutant plants exhibited an average reduction in height of 14 cm compared with wild-type Cadenza plants at maturity (Supplemental Figure 5).

Case study 2: Genetic divergence and trait mapping in pea RILs

To further assess the robustness and generalizability of the reverse BSA-QTLseq workflow, an independent validation was

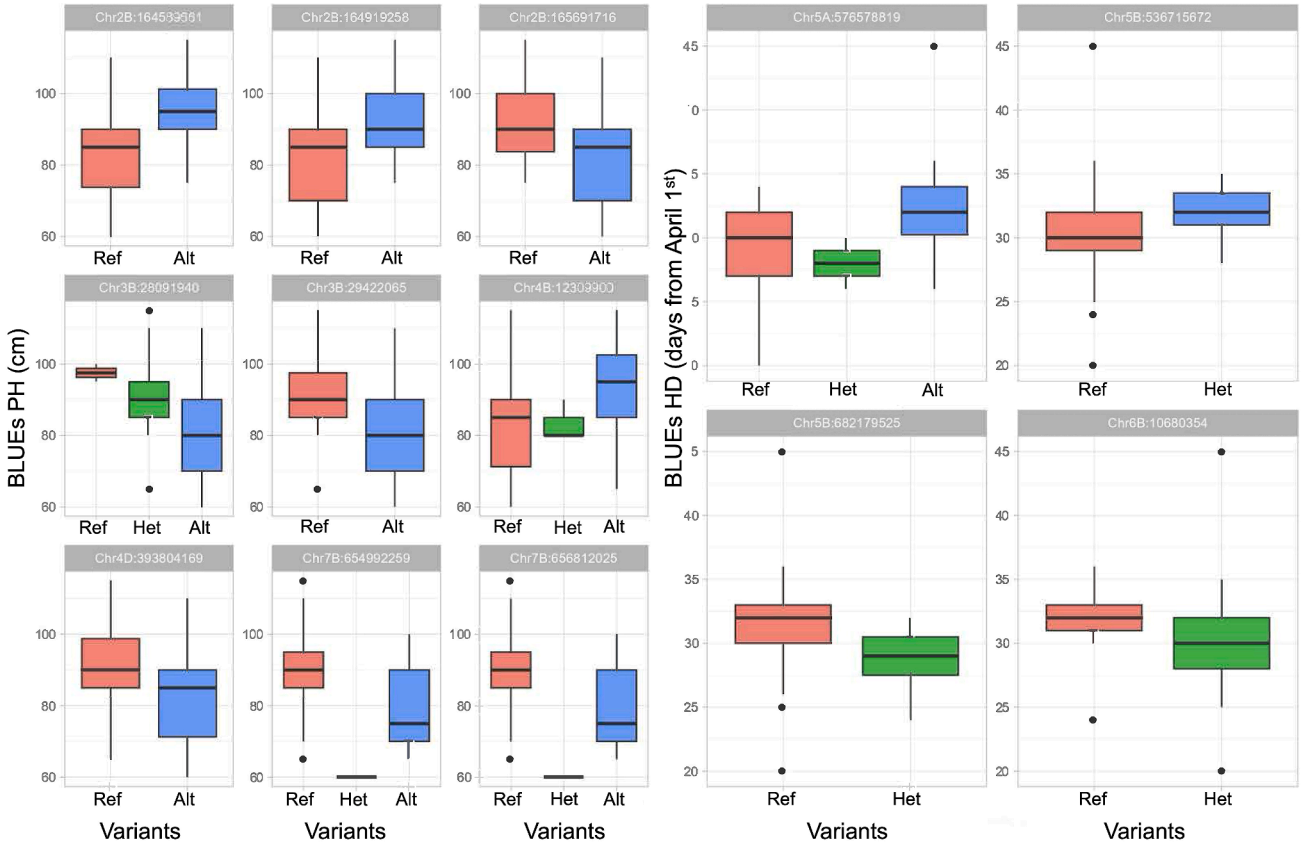

Figure 5. Boxplots of variants affecting plant height (PH) and heading date (HD).

Genotypes are grouped by allele type on the *x* axis, whereas the *y* axis represents the BLUE values for the corresponding traits. "Ref" (homozygous reference) indicates that both copies of the allele at a given genetic locus are the same as the reference sequence (the baseline sequence used for comparison). "Alt" (homozygous alternate) indicates that both copies of the allele at a genetic locus differ from the reference sequence. "Het" (heterozygous) refers to an individual having one reference allele and one alternate allele at a particular locus.
$p < 0.01$.

conducted using a publicly available dataset from a different plant species, pea (Pisum *sativum* L.) (Delvento et al., 2023). This dataset included phenotypic evaluations for resistance to crenate broomrape (*O. crenata* Forsk.). Six comparative analyses were performed to evaluate the performance of reverse BSA-QTLseq against conventional BSA-QTLseq across three bulk sizes (n = 5, 10, and 25).

Across all simulations, reverse BSA-QTLseq consistently identified major QTL regions associated with *O. crenata* resistance on chromosomes 1LG6, 4LG4, and 5LG3 (with ΔVariant index values exceeding the 99% confidence threshold) when bulk sizes of 10 and 25 individuals were used (Figure 6). Notably, with a bulk size of 10, additional candidate regions were detected on chromosomes 2LG1, 5LG3, 6LG2, and 7LG7, reflecting greater statistical power to uncover putative loci beyond those identified by conventional methods. In contrast, analyses with smaller bulks (n = 5) exhibited increased noise and failed to detect significant deviations, whereas larger bulks (n = 25) recovered two of the three major QTLs but showed reduced peak resolution and weaker ΔVariant index signals. For comparison, conventional BSA-QTLseq applied to the same bulk sizes yielded weaker and less distinct signals overall, with only the QTLs on 2LG1 and 4LG4 surpassing the 95% confidence threshold when n = 10 (Figure 6). Three SNPs were then converted into Kompetitive Allele-Specific PCR (KASP) assays and used to screen the RIL population, confirming full concordance with the GBS data and validating the newly identified regions (Supplemental Figure 6).

These results confirm that reverse BSA-QTLseq provides superior resolution and improved power for detecting genomic regions associated with *O. crenata* resistance, even when using relatively small sample sizes. Its successful application to a second species underscores the broader applicability and robustness of the method across diverse genetic architectures and crop systems.

DISCUSSION

This study introduces reverse BSA-QTLseq, which redefines the conventional BSA framework using *in silico* reconstruction of contrasting bulks (Figure 7). Unlike traditional phenotype-driven sampling (Figure 7A), this innovative method enables the simultaneous analysis of multiple traits using the same initial population (Figure 7B). It capitalizes on genetic divergence within an RIL population as the primary criterion for bulk sampling. The initial grouping, based on SNP genotyping, is

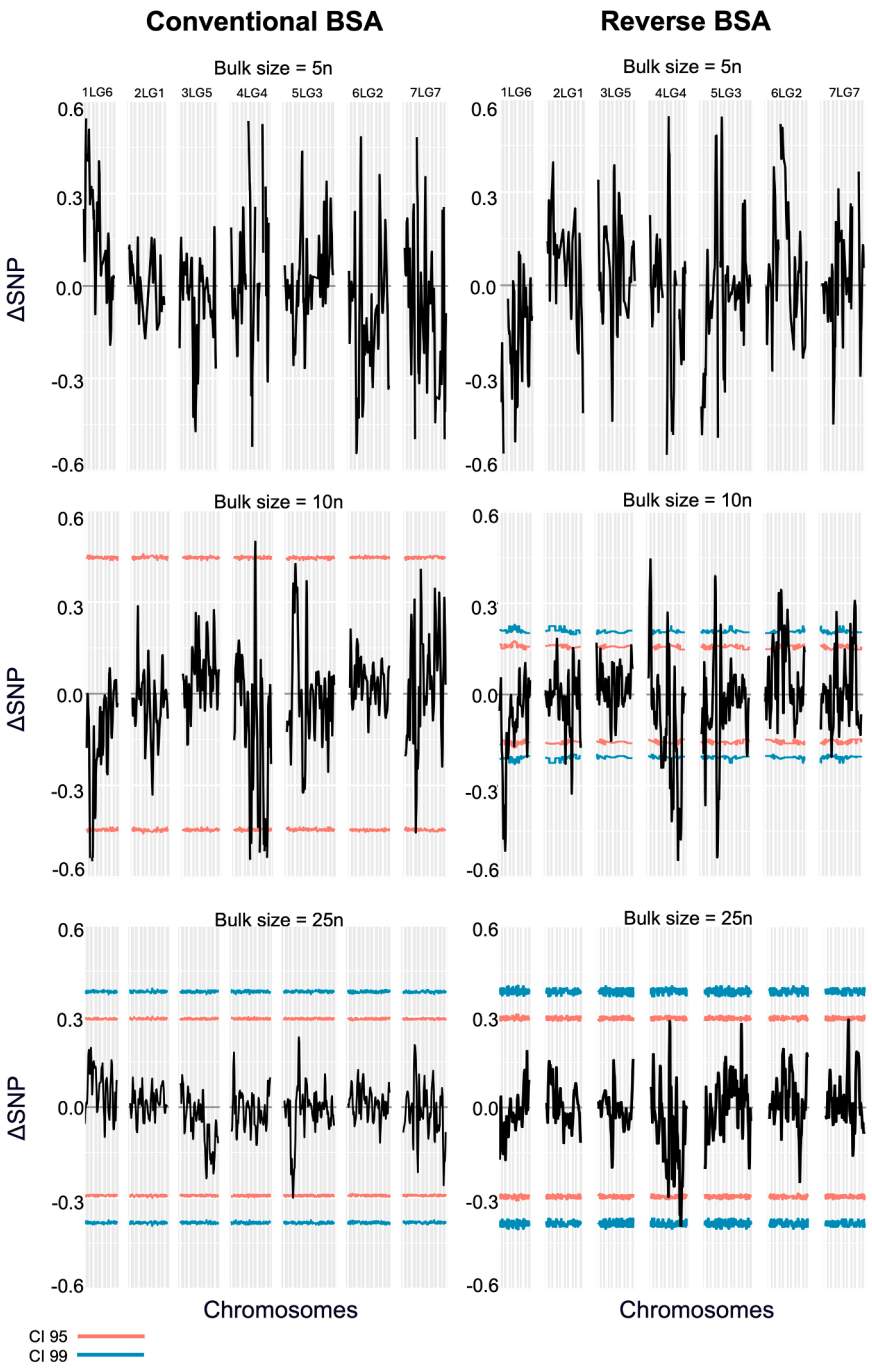

Figure 6. Validation of the reverse BSA-QTLseq.

The distribution of Δ(Variant index) values across pea chromosomes obtained by reverse BSA-QTLseq and conventional BSA-QTLseq under three different bulk sizes (n = 5, 10, and 25). The x axis represents chromosomes, labeled from 1LG6 to 7LG7. Blue and red horizontal lines denote the 95% and 99% confidence intervals, respectively, indicating significant QTL regions.

subsequently reorganized bioinformatically into trait-specific contrasting bulks, facilitating the concurrent analysis of multiple phenotypes from the same genotypic dataset.

The observed concordance in recombination rates and homozygosity patterns suggests that the genomic structure of the selected individuals does not introduce biases that could confound the mapping results. It also supports the robustness of reverse BSA-QTLseq in accurately capturing the genetic architecture underlying the trait of interest. Moreover, maintaining comparable recombination profiles between bulks enhances the resolution and interpretability of QTL detection, reducing the likelihood of false positives due to structural genomic differences.

This flexibility enhances mapping resolution and maximizes the reusability of genotypic data. As demonstrated in this study, reverse BSA-QTLseq successfully identified loci associated with HD and PH in bread wheat, showcasing its utility across multiple traits.

Integration of genotyping and exome capture data

A phased methodological approach was used: initial low-density genotyping was used to characterize the genetic diversity of a large RIL population, enabling the selection of genetically divergent lines and their division into two groups (Bulk 1 and Bulk 2). These selected lines then underwent whole-exome sequencing, increasing marker density from a few thousand to several hundred thousand variants (Esposito et al., 2022). This two-step strategy preserved the initial genetic variability while enhancing resolution, thus enabling trait-specific bulk reconstruction and QTL mapping of HD and PH.

The choice of HD and PH as case studies was strategic, given their well-characterized genetic basis in wheat, which provided a solid foundation for validating the proposed methodology. A comparison was first made with the results of Vitale et al. (2021), who performed QTL mapping on the entire RIL population from which our core set of lines was derived. In that study, six QTLs for HD were identified across early and late generations on chromosomes 2B, 2D, 4D, 5A, and 7D. Notably, two QTLs—on chromosomes 5A (designated *QHd-5A*) and 7D (designated *QHd-7D*)—were detected in the late generations (F_6:F_7). In addition, five QTLs were associated with PH. These findings were confirmed in the late generations, with an even larger number of associations than in previous reports. The expanded QTL landscape demonstrates the effectiveness of the proposed approach in uncovering additional genomic regions associated with key agronomic traits. Most major QTLs reported in previous studies were successfully replicated, further validating the reliability of reverse BSA-QTLseq (Supplemental Table 7). Notably, the results were aligned to the latest version of the bread wheat reference genome (IWGSC RefSeq v2.1; Zhu et al., 2021), allowing for meaningful comparisons,

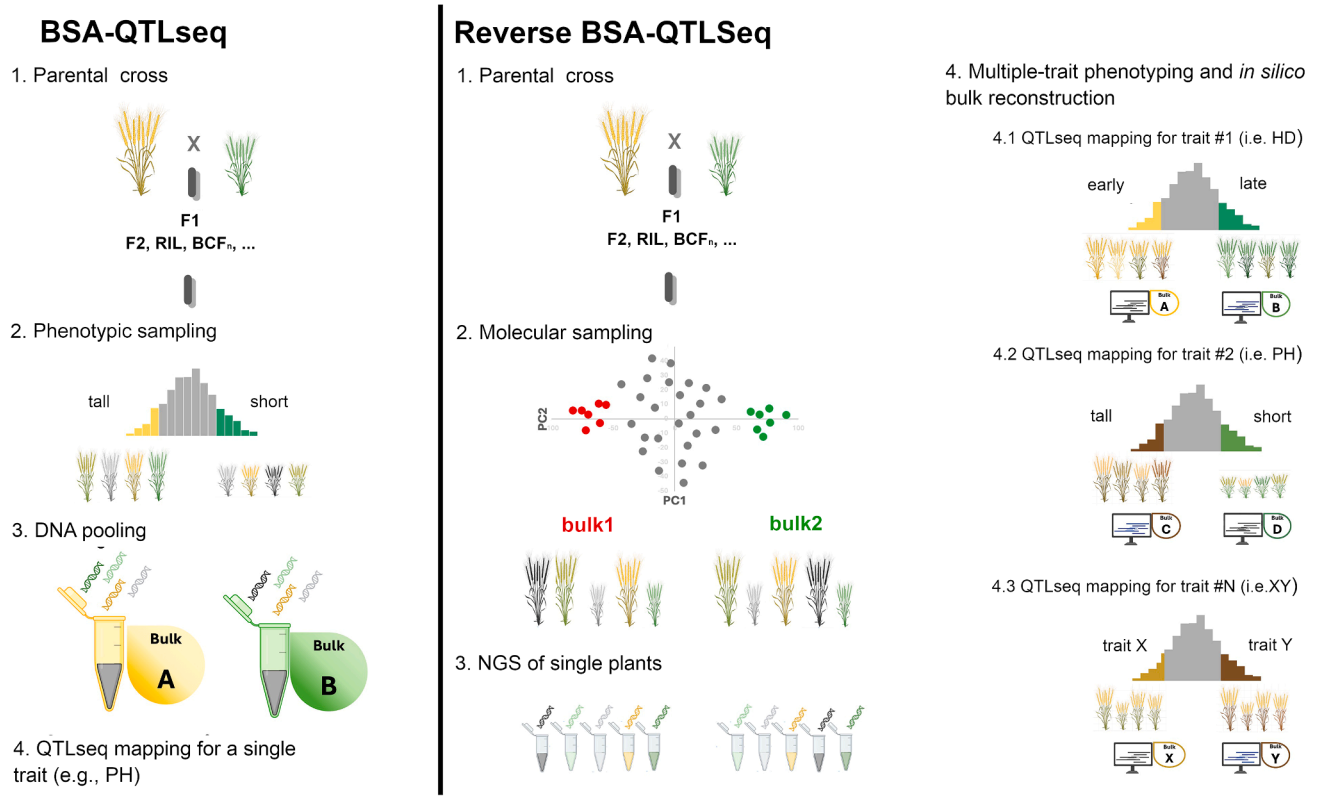

Figure 7. Working model.

This flowchart illustrates the methodological framework of two complementary approaches: classical BSA-QTLseq and reverse BSA-QTLseq. In classical BSA-QTLseq, a biparental cross is performed between genotypes with contrasting phenotypic traits (1), generating a segregating population. From this population, individuals exhibiting extreme phenotypes for a specific trait are selected to form two contrasting bulks (2). DNA from these bulks is pooled (3) and subjected to QTL-seq analysis (4) to identify genomic regions associated with the target trait. In reverse BSA-QTLseq, after a biparental cross (1), low-density SNP array sequencing is performed on the entire segregating population as well as the two parental lines. Next, PCA is conducted on the segregant population to analyze genetic distances derived from the SNP array. On this basis, two groups, each including a parent, are separated along PC1 and designated as Bulk 1 and Bulk 2, thereby capturing the maximum genetic divergence within the population (2). Next-generation sequencing (NGS), such as exome sequencing or other sequencing approaches, is performed on all individuals within these bulks (3). The genotyped individuals are phenotyped for multiple traits, and new bulks of individuals with contrasting phenotypes are reconstructed *in silico* (4) for each trait. These rearranged bulks are then used to perform QTL-seq analyses for each trait (4.1, 4.2, and 4.3).

particularly with studies using the same reference. A significant portion of the comparative analysis focused on the study by Xu et al. (2023), who mapped 332 QTLs, 270 genome-wide association study loci, and 83 genes related to PH. For QTLs associated with HD, comparisons were made with the studies by Pang et al. (2020) and Benaouda et al. (2022). Among the QTLs identified for PH, several were located in genomic regions previously associated with PH in wheat. For example, the loci *qPH.3B.1* and *qPH.3B.2* on chromosome 3B coincide with those mapped by Zhang et al. (2011) and Daoura et al. (2014). This region corresponds to the dwarfing gene *Rht-5* (*TraesCS3B02G025600*), which was fine-mapped by Cui et al. (2022) in the Chinese Spring reference genome (IWGSC RefSeq v2.1; Zhu et al., 2021). Similarly, *qPH.4B.1* and *qPH.4B.2*, located on chromosome 4B, have also been associated with PH in several studies (Shukla et al., 2015; Shirdelmoghanloo et al., 2016; Yuan et al., 2017; Guan et al., 2018; Sannemann et al., 2018; Zhao et al., 2019; Chen et al., 2020; Wang et al., 2024a) and are situated near the *Rht-B1* locus (Ellis et al., 2002; Lopes et al., 2014). In addition, *qPH.7B* was mapped within the 655 Mb region where Abou-Elwafa and Shehzad (2021) identified SNP IWB35241, further supporting its role in regulating PH. QTLs on chromosome 4D, including *qPH.4D.1*–*qPH.4D.5*, were found in regions closely linked to QPH-WMC473 (Gaynor, 2010) and Qph4D (Zhang et al., 2011), with physical distances ranging from 7 Mb (*qPH.4D.3*–QPH-WMC473) to 41 Mb (*qPH.4D.1*–QPH-WMC473). Similarly, *qHD.7D.1* on chromosome 7D was only 1.2 Mb away from *QPht.dms-7D.2*, previously reported by Semagn et al. (2021).

A notable result was the colocalization of several PH and HD QTLs, consistent with findings summarized by Xu et al. (2023) and indicative of a strong positive correlation between PH and HD (Supplemental Table 7). This was particularly evident in *qHD.5A.4*–*qHD.5A.8*, located from 562 to 626 Mb on chromosome 5A; this region includes the *Vrn-A1* gene (*TraesCS5A03G0935400*) at 589 Mb, a well-known gene controlling vernalization and heading time. Notably, some QTLs identified in our study did not overlap with any previously reported regions, suggesting the discovery of novel loci associated with PH and HD. For example, *qPH.2A.7*, *qPH.2A.8*, *qPH.2A.9*, *qHD.3B.7*, and *qHD.3B.8* showed no overlap with known QTLs and may

correspond to previously unreported genomic regions influencing these traits. This highlights the sensitivity of reverse BSA-QTLseq, even when applied to a limited number of lines in reconstructed bulks. However, further investigation is needed to confirm the functional roles of these loci in trait determination.

Regarding HD, *qHD.3A.1* overlapped with *qHD3A.1* from Pang et al. (2020), whereas *qHD.5A.3* and *qHD.6A.2* were found near *qHD5A.1* and *qHD6A.1* (< 50 Mb), respectively (Pang et al., 2020). In contrast, *qHD.5A.2–qHD.5A.9* coincided with *TaHd102* from Benaouda et al. (2022), whereas *qHD.5B.1* was located about 117 kb from *TaHd125* (Benaouda et al., 2022). Notably, *qHD.5A.4.8*, located from 562 to 626 Mb on chromosome 5A, included the *Vrn-A1* gene (TraesCS5A03G0935400) at 589 Mb. This alignment with the existing literature strongly supports the effectiveness of reverse BSA-QTLseq in accurately identifying and mapping loci associated with PH and HD. We also aimed to demonstrate that reverse BSA-QTLseq can identify a substantial number of associations even with a limited number of lines, comparable to or exceeding the results from traditional QTL and QTL-seq mapping methods (Zheng et al., 2024). For instance, in wheat, Komura et al., (2024a) used QTL-seq with F_3 progeny and detected three QTLs for HD. In rice, studies using BSA-QTLseq are even more common (e.g., Ogiso-Tanaka et al., 2017; Zhang et al., 2020; Nasution et al., 2025; Wang et al., 2024a). For example, Wang et al. (2024a) used contrasting F_2 and F_3 bulks and identified five QTLs for PH and four for HD. The outcomes of such studies are also influenced by the genetic characteristics of the parental lines used to generate the segregating populations (Zou et al., 2016; Wang et al., 2023).

Multi-omics analyses reveal key regulatory networks underpinning wheat growth and flowering

The combined transcriptomic and QTL-based analysis underscores the intricate molecular mechanisms governing major agronomic traits in wheat. Transcriptional profiles of leaf and root tissues from the parental lines of the RIL population reveal differences in gene expression linked to the phenotypic divergence between the two lines. In leaves, the enrichment of genes involved in photoperiod regulation, flowering, and transcription factor activity highlights variation in environmental perception, a critical factor in determining flowering time in cereals (Shi et al., 2019; Kiss et al., 2024). Supporting these findings, contrasting expression patterns of *Vrn* genes in leaves suggest differing vernalization and photoperiod response programs. These differences potentially explain the spring growth habit of Lankaodali (*Vrn-A1*, *Vrn-B1*) compared to the winter habit of Rebelde (*Vrn-2*) (Distelfeld et al., 2009; Shi et al., 2019). Moreover, root-specific enrichment of defense-related GO terms, including responses to salt stress, hypoxia, and fungal pathogens, suggests divergent stress adaptation strategies (Bostock et al., 2014; Ghosh and Xu, 2014; Trinidad et al., 2016). This variability in defense gene expression may influence resilience or susceptibility to soil-borne pathogens and abiotic stressors.

In contrast, the integrative pipeline for candidate gene identification—combining QTL mapping, transcriptome profiling, and the detection of deleterious variants—has identified a targeted set of loci critical for both PH and HD. Notably, genes such as *TraesCS4B03G0290400* (the PHD domain–containing gene), *TraesCS2B03G0403600* (the F-box gene), *TraesCS4B03G0117800* (the ABC transporter gene), and NBS-LRR–type disease-resistance genes appear to influence PH through mechanisms related to nutrient transport and to plant growth and defense signaling (Krattinger et al., 2009; Gu et al., 2015; Sánchez-Martín and Keller, 2021; Wang et al., 2024b). Similarly, the identification of 17 candidate genes for HD, including F-box and kinase-related genes, highlights the diverse molecular pathways that regulate floral initiation (Stefanowicz et al., 2015; Abd-Hamid et al., 2020). The functional relevance of these loci is further supported by their expression patterns in large transcriptomic datasets, linking genes such as *TraesCS4A03G1116300* (the sucrose synthase gene) and *TraesCS4D03G0456300* (*EMBRYONIC FLOWER 2, the VEFS-box gene*) to spike and grain development (Roux et al., 2006; Hou et al., 2014; Stein and Granot, 2019). Furthermore, significant allelic effects in multiple genes (e.g., *CML45* homologs, MYB transcription factor genes, *F-box* genes) reinforce their potential roles in PH and HD variation under field conditions (Dubos et al., 2010; Zhang et al., 2019). A TILLING mutant confirmed *TraesCS2B03G0403600* as a gene affecting PH, with a nonsense mutation in Cadenza1329 causing reduced stature. This validates the mapping approach and illustrates the value of TILLING for linking genotype to phenotype.

Overall, the application of this multi-omics approach demonstrates how gene expression profiles and candidate gene prioritization can converge to reveal interconnected regulatory networks governing growth, flowering, and stress responses. Future work using advanced molecular tools such as gene editing, TILLING, and transgenic complementation will be essential for validating the mechanistic roles of other candidate genes. This will ultimately facilitate the breeding of wheat varieties with optimized plant architecture, adaptive flowering schedules, and improved resilience in changing environments (Langridge and Reynolds, 2021).

Validation of reverse BSA-QTLseq for high-resolution mapping of *O. crenata* resistance in pea

To validate the robustness and applicability of reverse BSA-QTLseq, we used the publicly available pea dataset from Delvento et al. (2023), a well-established benchmark for QTL mapping of resistance to *O. crenata*. In their study, Delvento et al. (2023) identified three major QTLs (*PsOcr-1*, *PsOcr-2*, and *PsOcr-3*) using a high-density GBS-based linkage map in an RIL population derived from a cross between the resistant line "ROR12" and the susceptible cultivar "Sprinter." Notably, *PsOcr-1* accounted for 69.3% of the genotypic variance, highlighting its role as a major-effect locus for *O. crenata* resistance.

Our analyses confirmed the presence of these major QTLs on linkage groups 1LG6, 4LG4, and 5LG3, exactly corresponding to the positions of *PsOcr-1*, *PsOcr-2*, and *PsOcr-3*, when using reverse BSA-QTLseq with optimized bulk sizes. Importantly, the detection of additional candidate regions using moderate bulk sizes ($N = 10$) indicates enhanced sensitivity for identifying loci that may fall below the detection threshold in conventional BSA approaches. These results align with previously reported additive and epistatic interactions among *PsOcr* QTLs (Delvento et al., 2023), highlighting the capacity of reverse BSA-QTLseq to

capture both major- and minor-effect loci, even with reduced sample sizes and without relying on a full-population linkage map.

Compared to conventional BSA-QTLseq, which recovered only a subset of the major QTLs and at lower resolution, our findings demonstrate that reverse BSA-QTLseq maintains or improves mapping efficiency, as confirmed by the development of new KASP marker assays. This holds particular promise for targeting complex traits controlled by multiple loci of small to moderate effect, a context in which resource limitations often restrict extensive genotyping and phenotyping. The successful validation of key loci using an additional dataset from a different species underscores the potential of reverse BSA-QTLseq as a robust and scalable approach for QTL discovery across diverse crop species, delivering high resolution and minimizing experimental demands.

Methodological advantages and limitations

The flexibility of reverse BSA-QTLseq is demonstrated by its ability to analyze multiple traits using the same genotyped plant material, marking a significant advance over conventional BSA methods. By preserving individual-level genetic information, this strategy enables the reconstruction of bulks for different traits, providing a cost-effective and scalable solution for investigating complex polygenic traits. This capability is particularly valuable in crops like wheat, in which genetic diversity, large genome size, and trait complexity, including genotype-by-environment-by-management interactions, pose substantial challenges for trait mapping across diverse environments and growing conditions.

Although small bulk sizes are theoretically expected to reduce statistical power, particularly for loci with minor effects, our results provide empirical evidence that reverse BSA-QTLseq robustly detects both major and minor QTLs under these constraints.

This suggests that strategic selection of genetically divergent individuals, guided by genome-wide data, can compensate for smaller sample sizes by enhancing allele frequency differences at causal loci. Nonetheless, other factors, including the number and effect sizes of target loci, their interactions, marker density, and genotyping costs, remain critical considerations when optimizing experimental design (Yan et al., 2009).

Moreover, the integration of high-throughput sequencing technologies has greatly improved the resolution and robustness of this approach (Majeed et al., 2022; Wang et al., 2023), as reflected by the strong concordance between our findings and those reported in the literature. Future studies involving larger populations and broader trait panels may further increase the power and applicability of this methodology for high-resolution mapping and gene cloning efforts.

Cost effectiveness of reverse BSA-QTLseq

Although reverse BSA-QTLseq involves a two-step genotyping process (initial low-density genotyping followed by high-density genotyping of selected individuals), this upfront investment is offset by substantial cost savings during the phenotyping phase. Conventional phenotype-driven BSA approaches typically require extensive phenotyping of the entire population, which becomes increasingly expensive when multiple traits, environments, and growing seasons are involved (Hickey et al., 2019; Cobb et al., 2013). In contrast, reverse BSA-QTLseq enables researchers to focus phenotyping efforts on a much smaller, strategically selected subset of individuals, drastically reducing labor, time, and resource demands.

Given that phenotyping often surpasses genotyping in cost, especially in large-scale and multi-environment trials (Araus and Cairns, 2014; Reynolds et al., 2019), reverse BSA-QTLseq offers a more sustainable and scalable model for dissecting complex traits. Moreover, the reusability of genotypic data across multiple traits amplifies these economic benefits over time, especially because genotyping costs continue to decline while field phenotyping remains the primary bottleneck in crop improvement pipelines (Großkinsky et al., 2015; Scheben et al., 2017; Yang et al., 2020; Reynolds et al., 2021).

Implications for wheat breeding and beyond

HD and PH are key traits for adapting wheat to diverse agroclimatic conditions, optimizing resource use, and ensuring yield stability under both abiotic and biotic stresses (Borrill et al., 2019; Zhao et al., 2019). By validating previously reported QTLs and identifying novel candidate loci, this study highlights the practical value of reverse BSA-QTLseq for marker-assisted selection. The markers linked to such loci enable precise selection of favorable traits, thereby streamlining breeding pipelines and accelerating the development of varieties with improved phenology and plant architecture. Furthermore, the core principle of reusable genetic data introduces significant potential for breeding programs aiming to assess multiple traits concurrently. Researchers could perform an initial round of NGS on a diverse set of genotypes and reuse the data to define new contrasting bulk sets for various phenotypes, ranging from abiotic stress tolerance to quality traits, without the need for additional sequencing.

Future work could extend this framework to include the entire initial sampling population, enabling the identification of expression QTLs through this novel bulked-segregant strategy. Integrating these data would provide deeper insights into genotype-to-phenotype relationships, capturing regulatory variation at both the sequence and the gene expression levels.

Concluding remarks

This study provides compelling evidence that reverse BSA-QTLseq is a powerful and versatile strategy for trait mapping in wheat and potentially in other crop species. By shifting from phenotype-driven to genotype-driven bulk formation, this approach maximizes the reusability of NGS data, accelerates QTL discovery, and enables efficient validation of both previously known and novel loci. These advantages position reverse BSA-QTLseq as a valuable tool for modern breeding programs, with the potential to drive genetic gains across multiple agronomic traits simultaneously. Future research integrating larger populations and complementary omics datasets will further enhance the applicability and transformative potential of this innovative framework.

METHODS

Plant materials and field experiments

The initial genetic material consisted of a population of RILs derived from a cross between two phenotypically contrasting bread wheat varieties,

Lankaodali and Rebelde, which differ in several morphophysiological traits, as reported by Vitale et al. (2021). The population was advanced to the F_7 generation using the single-seed descent method, starting from the F_2 generation. After the final selfing, seeds from each line were collected and used for molecular analyses and subsequent field experiments, as described by Vitale et al. (2021).

In this study, 48 genotypes, including 46 RILs and the two parental lines, were evaluated. The plant material was grown across four consecutive growing seasons (2018–2019, 2019–2020, 2020–2021, and 2021–2022) at the CREA Research Centre for Cereal and Industrial Crops (CREA-CI), Foggia, Italy (41°27′40.2″N, 15°30′04.5″E). During the first two years, the plant material was grown under rain-fed conditions. In contrast, in the final two years, two irrigation treatments were applied (rain-fed and well-irrigated), resulting in six year-by-treatment combinations. The field trials were conducted on soil classified as clay under USDA standards (https://www.nrcs.usda.gov/sites/default/files/2025-01/Field-Book-for-Describing-and-Sampling-Soils-v4.pdf). A sowing density of approximately 350 seeds/m^2 was used. Experimental plots, each measuring 10.2 m^2 (1.36 m wide × 7.5 m long), were arranged in a randomized complete block design with three replicates per line. Crop management followed standard agronomic practices.

Bulk sampling criteria

Contrasting bulks were initially identified by maximizing genetic distance through PCA, using SNP markers from a larger population of RILs previously studied by Vitale et al. (2021). Genotyping was performed using the Illumina Wheat 25K iSelect Array (developed by TraitGenetics GmbH, Gatersleben, Germany; www.traitgenetics.com), as described by Vitale et al. (2021). Raw genotyping data were processed using PLINK v1.9 (Purcell et al., 2007); variants with a call rate below 95% or a minor allele frequency below 5% were excluded. A total of 6795 SNPs across all 21 chromosomes were used for PCA to explore genetic relationships within the population. The selection of contrasting bulks was based on the first principal component, reflecting the distributions of the two parental genotypes (Figure 1 and Supplemental Table 1). Lines positioned far apart in the PCA plot were considered genetically distant and were grouped accordingly, providing a solid framework for selecting two distinct, contrasting groups, named Bulk 1 and Bulk 2. The two bulks included 23 genotypes each, and the two parental lines were each assigned to their corresponding group (Figure 1). The selection of contrasting bulks based on genetic distance ensured maximum genetic divergence between groups, thereby increasing the power to detect associations between genetic markers and phenotypic traits. To assess the recombination rate within the selected bulks, two complementary approaches were used. First, runs of homozygosity were identified using PLINK v1.9 (Purcell et al., 2007) with high-quality SNPs and the following parameters: –homozyg-kb 100 and –homozyg-snp 5. Recombination rates were then estimated by counting recombination events: markers were sorted by their positions in the linkage maps, and only adjacent markers showing genotype changes were considered recombination breakpoints. Missing genotypes were excluded from the analysis. For both methods, mean values were calculated for individuals grouped either by phenotypic data or by reverse BSA-QTLseq and subsequently compared.

Exome capture and SNP identification

To preserve individual-level genetic information, the lines identified in each group (Bulk 1 and Bulk 2) were analyzed individually using exome sequencing, rather than being physically pooled as described in many BSA-QTLseq analyses (Zhang et al., 2020; Zhan et al., 2023; Zheng et al., 2024). This enabled the redefinition of bulks after genotyping based on target traits. Raw reads deposited by Esposito et al. (2022) were reanalyzed using the updated bread wheat genome version 2.1 (Zhu et al., 2021).

Read quality control was assessed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and Trimmomatic (Bolger et al., 2014) was used to remove reads with poor quality ($Q < 20$; SLIDINGWINDOW = 6). High-quality reads were aligned to the Chinese Spring wheat genome v2.1 (Zhu et al., 2021) using BWA-MEM (Li et al., 2009) (MINIMUMSEEDLENGTH = 19; BANDWIDTH = 100). The resulting SAM files were converted, sorted, and indexed using SAMtools (Li and Durbin, 2009). Duplicate reads were removed using the Picard MarkDuplicates tool (http://broadinstitute.github.io/picard). Only properly paired reads, defined as correctly oriented paired reads with a mapping distance less than the average insert size, were retained. GATK4 was subsequently used to call variants within each bulk, following best practices and using a hard-filtering procedure (Van der Auwera and O'Connor, 2024). Raw SNPs were filtered based on depth of coverage > 5 and allele coverage > 80%. High-quality variants were further annotated using the online tool Variant Effect Predictor (VEP; McLaren et al., 2016).

Bulk reconstruction and QTL-seq for plant height (PH) and heading date (HD)

As part of the validation process, HD and PH were measured for the 48 lines. HD was recorded as the number of days from April 1 to the date when 50% of the plants in a plot reached the GS59 stage on the Zadoks scale (Zadoks et al., 1974). PH (in cm) was measured at maturity on 10 randomly selected plants from each plot.

First, BLUE values for each individual were calculated as follows:

$$y_{ijk} = \mu + g_i + t_j + r_{k(j)} + gt_{ij} + e_{ijk}$$

where y_{ijk} is the observed value for the *i*th genotype in the *j*th environment and the *kth replicate*; μ is the overall mean; g_i is the effect of the *i*th genotype, assumed to be a fixed effect; t_j is the effect of the *j*th environment (defined as the treatment–year–location combination), modeled as a random effect with $t_j \sim N(0, \boldsymbol{I}\sigma_t^2)$; $r_{k(j)}$is the effect of the *k*th replicate within the trial, considered to be a random factor with $r_{k(j)} \sim N(0, \boldsymbol{I}\sigma_r^2)$; gt_{ij} is the genotype-by-environment interaction, modeled as a random effect with $gt_{ij} \sim N\left(0, \boldsymbol{I}\sigma_{gt}^2\right)$; and e_{ijk} is the residual effect, considered to be a random factor and assumed to follow a normal distribution $e_{ijk} \sim N(0, \boldsymbol{I}\sigma_e^2)$. This model was constructed in R using the *lmer* function from the "lme4" package (Bates et al., 2015). The resulting values were then used to rank the 48 individuals for each phenotypic trait and to divide them into two contrasting groups (Bulk A and Bulk B for PH and Bulk C and Bulk D for HD) (Figure 2). To maximize the phenotypic differences between groups, 10 contrasting lines were selected for each trait, and the BLUE means were compared using a *t*-test. Furthermore, boxplots illustrating the distribution of BLUE values for HD and PH across the different bulks were generated using the "ggplot2" package (Wickham, 2016). The BAM files for each group were merged to create an *in silico* pool within each group, in contrast to the DNA pooling used in typical BSA analysis. GATK was subsequently used to re-call variants within each bulk, following best practices and using a hard-filtering procedure (Van der Auwera and O'Connor, 2024). The function *FromGATKToTable* was used to extract the depth of coverage and Allelic Depth values and to prepare input files for BSA-QTLseq analysis, which was performed using the "QTL-seqR" package (Mansfeld and Grumet, 2018) with the following criteria: minimum total depth = 100, maximum total depth = 400, depth difference between bulks = 60, minimum sample depth = 40, and minimum genotype quality = 99. The SNP index was calculated as SNP $index_i$ = alternate allele depth/total read depth, where *i* represents the SNP's location. In each genomic interval, the distribution of Δ(SNP index) between the bulks was estimated using a 1-Mb sliding window. Regions where Δ(SNP index) exceeded the 95% or 99% confidence thresholds, as determined by statistical simulations, were considered to harbor major QTLs responsible for the differences between the bulks (Takagi et al., 2013; Itoh et al., 2019).

RNA-seq

An RNA-seq experiment with 60 million reads per sample was performed on the parental lines (Rebelde and Lankaodali). RNA was extracted from three biological replicates per line using 1 ml of TRIzol (Thermo Fisher

Scientific, Wilmington, DE, USA) according to the manufacturer's protocol, yielding an average RNA concentration of 2000 ng/μl with an RIN > 8.0. RNA quality and concentration were assessed using a Qubit (Thermo Fisher Scientific, Wilmington, DE, USA) and a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). Sequencing was performed using an Illumina HiSeq 2500 (WSU Genomics Core) and read quality control was performed using FastQC. Pseudo-alignment and quantification of RNA-seq reads were carried out with Kallisto-0.42.3 (Bray et al., 2016) using the *quant* command with options –b 100 and –t 8. The wheat RefSeq v2.1 genome assembly (Zhu et al., 2021) was used as the reference.

The "HTSFilter" package (Rau et al., 2013) was used to remove scarcely expressed genes based on the Jaccard similarity index, then the trimmed mean of *M* values normalization method was applied. Differential expression analysis was carried out using the "EdgeR" package in R (Pimentel et al., 2017), with Lankaodali used as the reference condition. To obtain information about the gene ontologies and gene descriptions associated with the *T. aestivum* genes, the Pannzer2 web tool was used (http://ekhidna2.biocenter.helsinki.fi/sanspanz/). A FASTA file containing all *T. aestivum* proteins was uploaded, with a minimum query and hit coverage of 0.4 and a minimum alignment length of 50 amino acids. The output was then used to extract gene descriptions and associated GO terms.

Identification of candidate genes associated with significant QTLs

Candidate genes were identified within the flanking regions of significant QTLs. Gene annotations were retrieved using the wheat RefSeq v2.1 genome assembly (Zhu et al., 2021) and high-confidence gene models, then intersected using *bedtools intersect* (Quinlan and Hall, 2010). To prioritize candidate genes potentially involved in PH and HD, a systematic SNP prioritization approach was used. First, candidate genes located near significant QTLs and harboring genetic variants were identified and categorized using *bedtools intersect* (Quinlan and Hall, 2010). Next, candidate genes near significant QTLs that showed differential expression between the parental lines in at least one analyzed tissue and were associated with regulatory variants were prioritized for further investigation. Variants identified in the selected genes were tested on the bulk samples using a *t*-test ($p \leq 0.01$). The population was divided into two groups based on allele type, and genes showing significant differences between the two groups were reported. Expression levels of the candidate genes (measured in transcripts per million) were analyzed in various tissues (leaves, roots, shoots, spikes, and grains), across developmental stages of the spike and grain (booting stages B1 to B11 and E1 to E10), and during leaf senescence (3–26 days after anthesis). Data for these analyses were obtained from the online database wGRN (http://wheat.cau.edu.cn/wGRN/). To validate our findings, Cadenza TILLING lines carrying mutations in the candidate genes were selected as previously described (Krasileva et al., 2017). In brief, lines with high-impact mutations were identified using the JBrowse graphical interface from Ensembl Plants (https://plants.ensembl.org/index.html). The corresponding M4 seeds were obtained through the SeedStore repository (Horler et al., 2018) and sown under open-field conditions at CREA-CI, Foggia, Italy (41°27′40.2″N, 15°30′04.5″E). The presence of the target mutations was confirmed by Sanger sequencing conducted in collaboration with Eurofins Genomics (Ebersberg, Germany). Primers for genotyping were designed using Primer3 (Untergasser et al., 2012) (Supplemental Table 8).

Validation of reverse BSA-QTLseq with publicly available pea RIL data

Reverse BSA-QTLseq was independently validated using publicly available data from Delvento et al. (2023). Genotyping data for 148 F_7 RILs derived from a cross between the resistant line "ROR12" and the susceptible cultivar "Sprinter" were downloaded from the figshare repository (https://doi.org/10.6084/m9.figshare.22741481.v2) in variant calling format. In addition, best linear unbiased prediction (BLUP) values calculated from two field trials and raw sequencing data (in FASTQ format) were kindly provided by the original authors. A subset of 48 accessions was selected based on PCA results, and three bulk sizes (5, 10, and 25 individuals) were constructed using BLUP data. For comparison, three corresponding bulk groups (5, 10, and 25 individuals) were also created solely based on phenotypic data from the entire RIL population. All six groups were then analyzed using the reverse BSA-QTLseq pipeline, and the results were systematically compared to assess the approach's robustness and resolution.

To validate the novel regions identified through the reverse BSA-QTLseq pipeline, three SNPs were selected and converted into KASP assays. Assay design was performed using LGC's KASP-by-Design service, and genotyping results were visualized with SNPviewer (https://www.biosearchtech.com/support/tools/genotyping-software/snpviewer). The KASP markers were then used to genotype the same segregating populations described by Delvento et al. (2023) (Supplemental Figure 5).

Data and code availability

The datasets presented in this study can be found in the NCBI repository (https://www.ncbi.nlm.nih.gov/), accession number PRJNA821683.

Publicly available data were downloaded from Delvento et al. (2023). All commands and scripts used in this study have been deposited in GitHub (https://github.com/salvatoreesposito01/ReverseBSA).

FUNDING

This work was supported by the Italian Ministry of Enterprises and Made in Italy – Mimit (MIMIT) in the framework of a project entitled "GRANO.IT Nr. F/360091/01-02/X75" (*CUP B19J24000140005*) and within the Agritech National Research Center, funded by MIUR–European Union PNRR – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4–SPOKE 1 INVESTIMENTO 1.4 (D.D. 1032 17/06/2022, CN00000022).

ACKNOWLEDGMENTS

No conflict of interest is declared.

AUTHOR CONTRIBUTIONS

Conceptualization, P.D.V., N.D., and F.T.; methodology, S.E., N.D., F.T., and F.F.; data curation, formal analysis, and visualization, S.E.; investigation, D.L., F.S., and F.F.; writing – original draft, S.E. and P.D.V.; supervision, funding acquisition, and resources, P.D.V.; validation, S.P. and I.C.; writing – review & editing, all authors.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

Received: March 6, 2025
Revised: June 13, 2025
Accepted: November 3, 2025
Published: November 7, 2025

REFERENCES

Abd-Hamid, N.A., Ahmad-Fauzi, M.I., Zainal, Z., and Ismail, I. (2020). Diverse and dynamic roles of F-box proteins in plant biology. Planta **251**:68. https://doi.org/10.1007/s00425-020-03356-8.

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., Matsumura, H., Yoshida, K., Mitsuoka, C., Tamiru, M., et al. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. Nat. Biotechnol. **30**:174–178. https://doi.org/10.1038/nbt.2095.

Abou-Elwafa, S.F., and Shehzad, T. (2021). Genetic diversity, GWAS and prediction for drought and terminal heat stress tolerance in

bread wheat (*Triticum aestivum* L.). Genet. Resour. Crop Evol. **68**:711–728. https://doi.org/10.1007/s10722-020-01018-y.

Araus, J.L., and Cairns, J.E. (2014). Field high-throughput phenotyping: the new crop breeding frontier. Trends Plant Sci. **19**:52–61. https://doi.org/10.1016/j.tplants.2013.09.008.

Arikit, S., Wanchana, S., Khanthong, S., Saensuk, C., Thianthavon, T., Vanavichit, A., and Toojinda, T. (2019). QTL-seq identifies cooked grain elongation QTLs near soluble starch synthase and starch branching enzymes in rice (*Oryza sativa* L.). Sci. Rep. **9**:8328. https://doi.org/10.1038/s41598-019-44856-2.

Branham, S.E., Patrick Wechter, W., Lambel, S., Massey, L., Ma, M., Fauve, J., Farnham, M.W., and Levi, A. (2018). QTL-seq and marker development for resistance to *Fusarium oxysporum f. sp. niveum* race 1 in cultivated watermelon. Mol. Breeding **38**:139. https://doi.org/10.1007/s11032-018-0896-9.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw. **67**:1–48. https://doi.org/10.18637/jss.v067.i01.

Benaouda, S., Dadshani, S., Koua, P., Léon, J., and Ballvora, A. (2022). Identification of QTLs for wheat heading time across multiple-environments. Theor. Appl. Genet. **135**:2833–2848. https://doi.org/10.1007/s00122-022-04152-6.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Borrill, P., Harrington, S.A., Simmonds, J., and Uauy, C. (2019). Identification of Transcription Factors Regulating Senescence in Wheat through Gene Regulatory Network Modelling. Plant Physiol. **180**:1740–1755. https://doi.org/10.1104/pp.19.00380.

Bostock, R.M., Pye, M.F., and Roubtsova, T.V. (2014). Predisposition in plant disease: exploiting the nexus in abiotic and biotic stress perception and response. Annu. Rev. Phytopathol. **52**:517–549. https://doi.org/10.1146/annurev-phyto-081211-172902.

Chen, H., Bemister, D.H., Iqbal, M., Strelkov, S.E., and Spaner, D.M. (2020). Mapping genomic regions controlling agronomic traits in spring wheat under conventional and organic managements. Crop Sci. **60**:2038–2052. https://doi.org/10.1002/csc2.20157.

Chen, M., Fan, W., Ji, F., Hua, H., Liu, J., Yan, M., Ma, Q., Fan, J., Wang, Q., Zhang, S., et al. (2021). Genome-wide identification of agronomically important genes in outcrossing crops using OutcrossSeq. Mol. Plant **14**:556–570. https://doi.org/10.1016/j.molp.2021.01.003.

Cobb, J.N., DeClerck, G., Greenberg, A., Clark, R., and McCouch, S. (2013). Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. Theor. Appl. Genet. **126**:867–887. https://doi.org/10.1007/s00122-013-2066-0.

Cui, C., Lu, Q., Zhao, Z., Lu, S., Duan, S., Yang, Y., Qiao, Y., Chen, L., and Hu, Y.G. (2022). The fine mapping of dwarf gene *Rht5* in bread wheat and its effects on plant height and main agronomic traits. Planta **255**:114. https://doi.org/10.1007/s00425-022-03888-1.

Daoura, B.G., Chen, L., Du, Y., and Hu, Y.G. (2014). Genetic effects of dwarfing gene *Rht-5* on agronomic traits in common wheat (*Triticum aestivum* L.) and QTL analysis on its linked traits. Field Crops Res. **156**:22–29. https://doi.org/10.1016/j.fcr.2013.10.007.

Delvento, C., Arcieri, F., Marcotrigiano, A.R., Guerriero, M., Fanelli, V., Dellino, M., Curci, P.L., Bouwmeester, H., Lotti, C., Ricciardi, L., et al. (2023). High-density linkage mapping and genetic dissection of resistance to broomrape (*Orobanche crenata* Forsk.) in pea (*Pisum sativum* L.). Front. Plant Sci. **14**:1216297. https://doi.org/10.3389/fpls.2023.1216297.

Deokar, A., Sagi, M., Daba, K., and Tar'an, B. (2019). QTL sequencing strategy to map genomic regions associated with resistance to ascochyta blight in chickpea. Plant Biotechnol. J. **17**:275–288. https://doi.org/10.1111/pbi.12964.

Distelfeld, A., Li, C., and Dubcovsky, J. (2009). Regulation of flowering in temperate cereals. Curr. Opin. Plant Biol. **12**:178–184. https://doi.org/10.1016/j.pbi.2008.12.010.

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). *MYB* transcription factors in *Arabidopsis*. Trends Plant Sci. **15**:573–581. https://doi.org/10.1016/j.tplants.2010.06.005.

Ellis, H., Spielmeyer, W., Gale, R., Rebetzke, J., and Richards, A. (2002). "Perfect" markers for the *Rht-B1b* and *Rht-D1b* dwarfing genes in wheat. Theor. Appl. Genet. **105**:1038–1042. https://doi.org/10.1007/s00122-002-1048-4.

Esposito, S., D'Agostino, N., Taranto, F., Sonnante, G., Sestili, F., Lafiandra, D., and De Vita, P. (2022). Whole-exome sequencing of selected bread wheat recombinant inbred lines as a useful resource for allele mining and bulked segregant analysis. Front. Genet. **13**:1058471. https://doi.org/10.3389/fgene.2022.1058471.

Gaynor, R.C. (2010). Quantitative trait loci mapping of yield, its related traits, and spike morphology factors in winter wheat (Triticum aestivum L.). M.S. Thesis. Oregon State University.

Ghosh, D., and Xu, J. (2014). Abiotic stress responses in plant roots: a proteomics perspective. Front. Plant Sci. **5**:6. https://doi.org/10.3389/fpls.2014.00006.

Großkinsky, D.K., Svensgaard, J., Christensen, S., and Roitsch, T. (2015). Plant phenomics and the need for physiological phenotyping across scales to narrow the genotype-to-phenotype knowledge gap. J. Exp. Bot. **66**:5429–5440. https://doi.org/10.1093/jxb/erv345.

Gu, L., Si, W., Zhao, L., Yang, S., and Zhang, X. (2015). Dynamic evolution of NBS–LRR genes in bread wheat and its progenitors. Mol. Genet. Genomics. **290**:727–738. https://doi.org/10.1007/s00438-014-0948-8.

Guan, P., Lu, L., Jia, L., Kabir, M.R., Zhang, J., Lan, T., Zhao, Y., Xin, M., Hu, Z., Yao, Y., et al. (2018). Global QTL analysis identifies genomic regions on chromosomes 4A and 4B harboring stable loci for yield-related traits across different environments in wheat (*Triticum aestivum* L.). Front. Plant Sci. **9**:529. https://doi.org/10.3389/fpls.2018.00529.

Hickey, L.T., N Hafeez, A., Robinson, H., Jackson, S.A., Leal-Bertioli, S.C.M., Tester, M., Gao, C., Godwin, I.D., Hayes, B.J., and Wulff, B.B.H. (2019). Breeding crops to feed 10 billion. Nat. Biotechnol. **37**:744–754. https://doi.org/10.1038/s41587-019-0152-9.

Horler, R.S.P., Turner, A.S., Fretter, P., and Ambrose, M. (2018). SeedStor: A Germplasm Information Management System and Public Database. Plant Cell Physiol. **59**:e5. https://doi.org/10.1093/pcp/pcx195.

Hou, J., Jiang, Q., Hao, C., Wang, Y., Zhang, H., and Zhang, X. (2014). Global selection on sucrose synthase haplotypes during a century of wheat breeding. Plant Physiol. **164**:1918–1929. https://doi.org/10.1104/pp.113.232454.

Huang, L., Tang, W., Bu, S., and Wu, W. (2020). BRM: a statistical method for QTL mapping based on bulked segregant analysis by deep sequencing. Bioinformatics **36**:2150–2156. https://doi.org/10.1093/bioinformatics/btz861.

Imerovski, I., Dedić, B., Cvejić, S., Miladinović, D., Jocić, S., Owens, G.L., Tubić, N.K., and Rieseberg, L.H. (2019). BSA-seq mapping reveals major QTL for broomrape resistance in four sunflower lines. Mol. Breeding **39**:41. https://doi.org/10.1007/s11032-019-0948-9.

Itoh, N., Segawa, T., Tamiru, M., Abe, A., Sakamoto, S., Uemura, A., Oikawa, K., Kutsuzawa, H., Koga, H., Imamura, T., et al. (2019). Next-generation sequencing-based bulked segregant analysis for QTL mapping in heterozygous species *Brassica rapa*. Theor. Appl. Genetics **132**:455–467. https://doi.org/10.1007/s00122-019-03396-z.

Kiss, T., Horváth, Á.D., Cseh, A., Berki, Z., Balla, K., and Karsai, I. (2024). Molecular genetic regulation of the vegetative–generative transition in wheat from an environmental perspective. Ann. Bot. **2024**:mcae174. https://doi.org/10.1093/aob/mcae174.

Komura, S., Kobayashi, F., Oono, Y., Handa, H., Inoue, Y., and Yoshida, K. (2024a). Identification of three QTLs that additively affect heading time in bread wheat (*Triticum aestivum* L.) by QTL-seq approach. Euphytica **220**:183. https://doi.org/10.1007/s10681-024-03441-z.

Komura, S., Yoshida, K., Jinno, H., Oono, Y., Handa, H., Takumi, S., and Kobayashi, F. (2024b). Identification of the causal mutation in early heading mutant of bread wheat (*Triticum aestivum* L.) using MutMap approach. Mol. Breed. **44**:41. https://doi.org/10.1007/s11032-024-01478-5.

Krasileva, K.V., Vasquez-Gross, H.A., Howell, T., Bailey, P., Paraiso, F., Clissold, L., Simmonds, J., Ramirez-Gonzalez, R.H., Wang, X., Borrill, P., et al. (2017). Uncovering hidden variation in polyploid wheat. Proc. Natl. Acad. Sci. USA **114**:E913–E921. https://doi.org/10.1073/pnas.1619268114.

Krattinger, S.G., Lagudah, E.S., Spielmeyer, W., Singh, R.P., Huerta-Espino, J., McFadden, H., Bossolini, E., Selter, L.L., and Keller, B. (2009). A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. Science **323**:1360–1363. https://doi.org/10.1126/science.1166453.

Langridge, P., and Reynolds, M. (2021). Breeding for drought and heat tolerance in wheat. Theor. Appl. Genet. **134**:1753–1769. https://doi.org/10.1007/s00122-021-03795-1.

Li, H., Handsaker, B., Wysoker, A., et al. (2009). 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics **25**:1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Li, Z., and Xu, Y. (2022). Bulk segregation analysis in the NGS era: a review of its teenage years. Plant J. **109**:1355–1374. https://doi.org/10.1111/tpj.15646.

Lei, L., Zheng, H., Bi, Y., Yang, L., Liu, H., Wang, J., Sun, J., Zhao, H., Li, X., Li, J., et al. (2020). Identification of a major QTL and candidate gene analysis of salt tolerance at the bud burst stage in rice (*Oryza sativa* L.) using QTL-seq and RNA-seq. Rice **13**:55. https://doi.org/10.1186/s12284-020-00416-1.

Lopes, M.S., Dreisigacker, S., Peña, R.J., Sukumaran, S., and Reynolds, M.P. (2014). Genetic characterization of the wheat association mapping initiative (*WAMI*) panel for dissection of complex traits in spring wheat. Theor. Appl. Genet. **128**:453–464. https://doi.org/10.1007/s00122-014-2444-2.

Ma, P., Wu, L., Xu, Y., Xu, H., Zhang, X., Wang, W., Liu, C., and Wang, B. (2021). Bulked segregant RNA-Seq provides distinctive expression profile against powdery mildew in the wheat genotype YD588. Front. Plant Sci. **12**:764978. https://doi.org/10.3389/fpls.2021.764978.

Majeed, A., Johar, P., Raina, A., Salgotra, R.K., Feng, X., and Bhat, J.A. (2022). Harnessing the potential of bulk segregant analysis sequencing and its related approaches in crop breeding. Front. Genet. **13**:944501. https://doi.org/10.3389/fgene.2022.944501.

Mansfeld, B.N., and Grumet, R. (2018). QTLseqr: An R package for bulk segregant analysis with next-generation sequencing. Plant Genome **11**. https://doi.org/10.3835/plantgenome2018.01.0006.

McLaren, W., Gil, L., Hunt, S.E., et al. (2016). The Ensembl Variant Effect Predictor. Genome Biol. **17**:122. https://doi.org/10.1186/s13059-016-0974-4.

Michelmore, R.W., Paran, I., and Kesseli, R.V. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc. Natl. Acad. Sci. USA **88**:9828–9832. https://doi.org/10.1073/pnas.88.21.9828.

Nasution, K.Y., Satyawan, D., Yunus, M., Dewi, A.K., Melati, P., Maryono, M.Y., Dwimahyani, I., Enggarini, W., and Sobrizal, S. (2025). Detection of genomic loci associated with days to heading in tropical japonica rice through QTL-seq. Czech J. Genet. Plant Breed. **61**:23–30. https://doi.org/10.17221/66/2024-CJGPB.

Ogiso-Tanaka, E., Tanaka, T., Tanaka, K., Nonoue, Y., Sasaki, T., Fushimi, E., Koide, Y., Okumoto, Y., Yano, M., and Saito, H. (2017). Detection of novel QTLs qDTH4. 5 and qDTH6. 3, which confer late heading under short-day conditions, by SSR marker-based and QTL-seq analysis. Breed. Sci. **67**:101–109. https://doi.org/10.1270/jsbbs.16096.

Pandey, M.K., Khan, A.W., Singh, V.K., Vishwakarma, M.K., Shasidhar, Y., Kumar, V., Garg, V., Bhat, R.S., Chitikineni, A., Janila, P., et al. (2017). QTL-seq approach identified genomic regions and diagnostic markers for rust and late leaf spot resistance in groundnut (*Arachis hypogaea* L.). Plant Biotechnol. J. **15**:927–941. https://doi.org/10.1111/pbi.12686.

Pang, Y., Liu, C., Wang, D., St Amand, P., Bernardo, A., Li, W., He, F., Li, L., Wang, L., Yuan, X., et al. (2020). High-resolution genome-wide association study identifies genomic regions and candidate genes for important agronomic traits in wheat. Mol. Plant **13**:1311–1327. https://doi.org/10.1016/j.molp.2020.07.008.

Purcell, S., Neale, B., Todd-Brown, K., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. **81**:559–575. https://doi.org/10.1086/519795.

Qian, Z., Liu, R., Liu, X., Qie, Y., Wang, J., Yin, Y., Xin, Q., Yu, N., Zhang, J., Li, Y., et al. (2024). Bulked segregant RNA-seq reveals complex resistance expression profile to powdery mildew in wild emmer wheat W762. Front. Plant Sci. **15**:1387427. https://doi.org/10.3389/fpls.2024.1387427.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26**:841–842. https://doi.org/10.1093/bioinformatics/btq033.

Ramos, A., Fu, Y., Michael, V., and Meru, G. (2020). QTL-seq for identification of loci associated with resistance to Phytophthora crown rot in squash. Sci. Rep. **10**:5326–5328. https://doi.org/10.1038/s41598-020–62228-z.

Rau, A., Gallopin, M., Celeux, G., Jaffrézic, F., and Rau, M.A. (2013). Data-based filtering for replicated high-throughput transcriptome sequencing experiments. Bioinformatics **29**:2146–2152. https://doi.org/10.1093/bioinformatics/btt350.

Reynolds, D., Baret, F., Welcker, C., Bostrom, A., Ball, J., Cellini, F., Lorence, A., Chawade, A., Khafif, M., Noshita, K., et al. (2019). What is cost-efficient phenotyping? Optimizing costs for different scenarios. Plant Sci. **282**:14–22. https://doi.org/10.1016/j.plantsci.2018.06.015.

Reynolds, M., Atkin, O.K., Bennett, M., Cooper, M., Dodd, I.C., Foulkes, M.J., Frohberg, C., Hammer, G., Henderson, I.R., Huang, B., et al. (2021). Addressing Research Bottlenecks to Crop Productivity. Trends Plant Sci. **26**:607–630. https://doi.org/10.1016/j.tplants.2021.03.011.

Roux, F., Touzet, P., Cuguen, J., and Le Corre, V. (2006). How to be early flowering: an evolutionary perspective. Trends Plant Sci. **11**:375–381. https://doi.org/10.1016/j.tplants.2006.06.006.

Ruangrak, E., Du, Y., Htwe, N.M.P.S., Pimorat, P., and Gao, J. (2019). Identification of early tomato fruit ripening loci by QTL-seq. J. Agric. Sci. **11**:51. https://doi.org/10.5539/jas.v11n2p51.

Sánchez-Martín, J., and Keller, B. (2021). NLR immune receptors and diverse types of non-NLR proteins control race-specific resistance in Triticeae. Curr. Opin. Plant Biol. **62**:102053. https://doi.org/10.1016/j.pbi.2021.102053.

Sannemann, W., Lisker, A., Maurer, A., Léon, J., Kazman, E., Cöster, H., Holzapfel, J., Kempf, H., Korzun, V., Ebmeyer, E., et al. (2018). Adaptive selection of founder segments and epistatic control of plant height in the MAGIC winter wheat population WM-800. BMC Genom. **19**:559. https://doi.org/10.1186/s12864-018-4915-3.

Scheben, A., Batley, J., and Edwards, D. (2017). Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. Plant Biotechnol. J. **15**:149–161. https://doi.org/10.1111/pbi.12645.

Segawa, T., Nishiyama, C., Tamiru-Oli, M., Sugihara, Y., Abe, A., Sone, H., Itoh, N., Asukai, M., Uemura, A., Oikawa, K., et al. (2021). Sat-BSA: an NGS-based method using local de novo assembly of long reads for rapid identification of genomic structural variations associated with agronomic traits. Breed. Sci. **71**:299–312. https://doi.org/10.1270/jsbbs.20148.

Semagn, K., Iqbal, M., Chen, H., Perez-Lara, E., Bemister, D.H., Xiang, R., Zou, J., Asif, M., Kamran, A., N'Diaye, A., et al. (2021). Physical Mapping of QTL in Four Spring Wheat Populations under Conventional and Organic Management Systems. I. Earliness. Plants **10**:853. https://doi.org/10.3390/plants10050853.

Sham, P., Bader, J.S., Craig, I., O'Donovan, M., and Owen, M. (2002). DNA pooling: a tool for large-scale association studies. Nat. Rev. Genet. **3**:862–871. https://doi.org/10.1038/nrg930.

Shen, R., and Messer, P.W. (2022). Predicting the genomic resolution of bulk segregant analysis. *G3: Genes, Genomes*. Genetics **12**:jkac012. https://doi.org/10.1093/g3journal/jkac012.

Shi, C., Zhao, L., Zhang, X., Lv, G., Pan, Y., and Chen, F. (2019). Gene regulatory network and abundant genetic variation play critical roles in heading stage of polyploidy wheat. BMC Plant Biol. **19**:6–16. https://doi.org/10.1186/s12870-018-1591-z.

Shirdelmoghanloo, H., Taylor, J.D., Lohraseb, I., Rabie, H., Brien, C., Timmins, A., Martin, P., Mather, D.E., Emebiri, L., and Collins, N.C. (2016). A QTL on the short arm of wheat (*Triticum aestivum* L.) chromosome 3B affects the stability of grain weight in plants exposed to a brief heat shock early in grain filling. BMC Plant Biol. **16**:100–115. https://doi.org/10.1186/s12870-016-0784-6.

Stefanowicz, K., Lannoo, N., and Van Damme, E.J.M. (2015). Plant F-box proteins–judges between life and death. Crit. Rev. Plant Sci. **34**:523–552. https://doi.org/10.1080/07352689.2015.1024566.

Stein, O., and Granot, D. (2019). An overview of sucrose synthases in plants. Front. Plant Sci. **10**:95. https://doi.org/10.3389/fpls.2019.00095.

Shukla, S., Singh, K., Patil, R.V., Kadam, S., Bharti, S., Prasad, P., Singh, N.K., and Khanna-Chopra, R. (2015). Genomic regions associated with grain yield under drought stress in wheat (*Triticum aestivum* L.). Euphytica **203**:449–467. https://doi.org/10.1007/s10681-014-1314-y.

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., Uemura, A., Utsushi, H., Tamiru, M., Takuno, S., et al. (2013). QTL–seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J. **74**:174–183. https://doi.org/10.1111/tpj.12105.

Trinidad, J.L., Grajo, H.L., Abucay, J.B., and Kohli, A. (2016). Cereal root proteomics for complementing the mechanistic understanding of plant abiotic stress tolerance. In Agricultural Proteomics Volume 2, G. Salekdeh, ed. (Springer). https://doi.org/10.1007/978-3-319-43278-6_2.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3–new capabilities and interfaces. Nucleic Acids Res. **40**:e115. https://doi.org/10.1093/nar/gks596.

Van der Auwera, G., and D O'Connor, B. (2024). Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. First edition. (Sebastopol, CA: O'Reilly Media).

Vitale, P., Fania, F., Esposito, S., Pecorella, I., Pecchioni, N., Palombieri, S., Sestili, F., Lafiandra, D., Taranto, F., and De Vita, P. (2021). QTL Analysis of Five Morpho-Physiological Traits in Bread Wheat Using Two Mapping Populations Derived from Common Parents. Genes **12**:604. https://doi.org/10.3390/genes12040604.

Wang, G., Gu, J., Long, D., Zhang, X., Zhao, C., Zhang, H., Chen, C., and Ji, W. (2024a). Genome-wide identification of wheat ABC gene family and expression in response to fungal stress treatment. Plant Biotechnol. Rep. **18**:401–413. https://doi.org/10.1007/s11816-023-00881-2.

Wang, R., Li, K., Zhang, W., Liu, H., Tao, Y., Liu, Y., Ding, G., Yang, G., Zhou, Y., Wang, J., et al. (2024b). QTL-seq analysis identified the genomic regions of plant height and days to heading in high-latitude rice. Front. Genet. **15**:1305681. https://doi.org/10.3389/fgene.2024.1305681.

Wang, X., Han, L., Li, J., Shang, X., Liu, Q., Li, L., and Zhang, H. (2023). Next-generation bulked segregant analysis for Breeding 4.0. Cell Rep. **42**:113039. https://doi.org/10.1016/j.celrep.2023.113039.

Wang, Y., Xie, J., Zhang, H., Guo, B., Ning, S., Chen, Y., Lu, P., Wu, Q., Li, M., Zhang, D., et al. (2017). Mapping stripe rust resistance gene YrZH22 in Chinese wheat cultivar Zhoumai 22 by bulked segregant RNA-Seq (BSR-Seq) and comparative genomics analyses. Theor. Appl. Genet. **130**:2191–2201. https://doi.org/10.1007/s00122-017-2950-0.

Wang, C., Tang, S., Zhan, Q., Hou, Q., Zhao, Y., Zhao, Q., Feng, Q., Zhou, C., Lyu, D., Cui, L., et al. (2019). Dissecting a heterotic gene through GradedPool-Seq mapping informs a rice-improvement strategy. Nat. Commun. **10**:2982. https://doi.org/10.1038/s41467-019-11017-y.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (New York: Springer-Verlag), https://doi.org/10.1007/978-3-319-24277-4.

Xu, D., Jia, C., Lyu, X., Yang, T., Qin, H., Wang, Y., Hao, Q., Liu, W., Dai, X., Zeng, J., et al. (2023). In silico curation of QTL-rich clusters and candidate gene identification for plant height of bread wheat. Crop J. **11**:1480–1490. https://doi.org/10.1016/j.cj.2023.05.007.

Xin, F., Zhu, T., Wei, S., Han, Y., Zhao, Y., Zhang, D., Ma, L., and Ding, Q. (2020). QTL Mapping of kernel traits and validation of a major QTL for kernel length-width ratio using SNP and bulked segregant analysis in wheat. Sci. Rep. **10**:25. https://doi.org/10.1038/s41598-019-56979-7.

Xiang, D., Quilichini, T.D., Liu, Z., Gao, P., Pan, Y., Li, Q., Nilsen, K.T., Venglat, P., Esteban, E., Pasha, A., et al. (2019). The Transcriptional Landscape of Polyploid Wheats and Their Diploid Ancestors during Embryogenesis and Grain Development. Plant Cell **31**:2888–2911. https://doi.org/10.1105/tpc.19.00397.

Yan, A.O., Hu, Z.Q., Tang, Z.X., Wang, X.F., and Xu, C.W. (2009). A general method for QTL mapping in multiple related populations derived from multiple parents of rice Rice Science **16**:45–50. https://doi.org/10.1016/S1672-6308(08)60055-4.

Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J.H., Batchelor, W.D., Xiong, L., and Yan, J. (2020). Crop Phenomics and High-Throughput Phenotyping: Past Decades, Current Challenges, and Future Perspectives. Mol. Plant **13**:187–214. https://doi.org/10.1016/j.molp.2020.01.008.

Yuan, Y., Gao, M., Zhang, M., Zheng, H., Zhou, X., Guo, Y., Zhao, Y., Kong, F., and Li, S. (2017). QTL mapping for phosphorus efficiency and morphological traits at seedling and maturity stages in wheat. Front. Plant Sci. **8**:614. https://doi.org/10.3389/fpls.2017.00614.

Yu, T., Cao, S., Jin, Y., Xu, C., Liu, R., Wang, B., Lv, Y., Meng, T., and Ma, P. (2024). Unveiling resistance expression profile to powdery mildew in wheat via Bulked Segregant RNA-Seq. BMC Plant Biol. **24**:1061. https://doi.org/10.1186/s12870-024-05789-9.

Zadoks, J.C., Chang, T.T., and Konzac, C.F. (1974). A decimal code for the growth stages of cereals. Weed Res. **14**:415–421. https://doi.org/10.1111/j.1365-3180.1974.tb01084.x.

Zhan, Y., Hu, W., He, H., Dang, X., Chen, S., and Bie, Z. (2023). A major QTL identification and candidate gene analysis of watermelon fruit cracking using QTL-seq and RNA-seq. Front. Plant Sci. **14**:1166008. https://doi.org/10.3389/fpls.2023.1166008.

Zhang, J., Hao, C., Ren, Q., Chang, X., Liu, G., and Jing, R. (2011). Association mapping of dynamic developmental plant height in common wheat. Planta **234**:891–902. https://doi.org/10.1007/s00425-011-1434-8.

Zhang, L., Duan, Y., Zhang, Z., Zhang, L., Chen, S., Cai, C., Duan, S., Zhang, K., Li, G., and Cheng, F. (2024). OcBSA: An NGS-based bulk segregant analysis tool for outcross populations. Mol. Plant **17**:648–657. https://doi.org/10.1016/j.molp.2024.02.011.

Zhang, X., Gonzalez-Carranza, Z.H., Zhang, S., Miao, Y., Liu, C.J., and Roberts, J.A. (2019). F-Box proteins in plants. Annu. Rev. Plant Biol. **2**:307–328. https://doi.org/10.1002/9781119312994.apr0701.

Zhang, X., Zhang, K., Wu, J., Guo, N., Liang, J., Wang, X., and Cheng, F. (2020). QTL-Seq and Sequence Assembly Rapidly Mapped the Gene *BrMYBL2.1* for the Purple Trait in *Brassica rapa*. Sci. Rep. **10**:2328. https://doi.org/10.1038/s41598-020-58916-5.

Zhao, C., Zhang, N., Wu, Y., Sun, H., Liu, C., Fan, X., Yan, X., Xu, H., Ji, J., and Cui, F. (2019). QTL for spike-layer uniformity and their influence on yield-related traits in wheat. BMC Genet. **20**:23. https://doi.org/10.1186/s12863-019-0730-3.

Zheng, Y., Ei Khine, E., Mar Thi, K., Ei Nyein, E., Huang, L., Lin, L., Xie, X., Htay Wai Lin, M., Than Oo, K., Myat Moe, M., et al. (2024). Multi-environment BSA-seq using large F3 populations is able to achieve reliable QTL mapping with high power and resolution: An experimental demonstration in rice. Crop J. **12**:549–557. https://doi.org/10.1016/j.cj.2024.01.009.

Zhu, T., Wu, L., He, H., Song, J., Jia, M., Liu, L., Wang, X., Han, R., Niu, L., Du, W., et al. (2020). Bulked Segregant RNA-Seq Reveals Distinct Expression Profiling in Chinese Wheat Cultivar Jimai 23 Responding to Powdery Mildew. Front. Genet. **11**:474. https://doi.org/10.3389/fgene.2020.00474.

Zhu, T., Wang, L., Rimbert, H., Rodriguez, J.C., Deal, K.R., De Oliveira, R., Choulet, F., Keeble-Gagnère, G., Tibbits, J., Rogers, J., et al. (2021). Optical maps refine the bread wheat Triticum aestivum cv Chinese Spring genome assembly. Plant J. **107**:303–314. https://doi.org/10.1111/tpj.15289.

Zou, C., Wang, P., and Xu, Y. (2016). Bulked sample analysis in genetics, genomics and crop improvement. Plant Biotechnol. J. **14**:1941–1955. https://doi.org/10.1111/pbi.12559.