

Journal Pre-proof

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

Salvatore Esposito, Nunzio D'Agostino, Francesca Taranto, Fabio Fania, Stefano Pavan, Ida Colella, Francesco Sestili, Domenico Lafiandra, Pasquale De Vita

PII: S2590-3462(25)00350-5

DOI: <https://doi.org/10.1016/j.xplc.2025.101588>

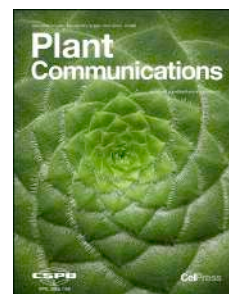
Reference: XPLC 101588

To appear in: *Plant Communications*

Received Date: 6 March 2025

Revised Date: 13 June 2025

Accepted Date: 3 November 2025



Please cite this article as: Esposito, S., D'Agostino, N., Taranto, F., Fania, F., Pavan, S., Colella, I., Sestili, F., Lafiandra, D., De Vita, P., Reverse BSA-QTLseq: a new genotype-driven bioinformatics approach for simultaneous trait mapping, *Plant Communications* (2025), doi: <https://doi.org/10.1016/j.xplc.2025.101588>.

This is a PDF of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability. This version will undergo additional copyediting, typesetting and review before it is published in its final form. As such, this version is no longer the Accepted Manuscript, but it is not yet the definitive Version of Record; we are providing this early version to give early visibility of the article. Please note that Elsevier's sharing policy for the Published Journal Article applies to this version, see: <https://www.elsevier.com/about/policies-and-standards/sharing#4-published-journal-article>. Please also note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2025 The Author(s). Published by Elsevier Inc. on behalf of CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Chinese Society for Plant Biology.

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

Salvatore Esposito^{1,2}, Nunzio D'Agostino³, Francesca Taranto⁴, Fabio Fania², Stefano Pavan⁵, Ida Colella^{2,6}, Francesco Sestili⁶, Domenico Lafiandra⁶, 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

* Corresponding author

Summary

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 leverages genotype-driven bulk reconstruction through bioinformatics, enabling the simultaneous mapping of multiple traits from the same genotypic dataset. The Reverse BSA-QTLseq approach employs a two-step strategy - low-resolution genotyping of the entire population followed by high-resolution sequencing of selected bulks - enabling cost-effective prioritization of genetically informative lines to enhance the discovery of quantitative trait loci (QTLs). Using a bread wheat recombinant inbred line (RIL) population as a case study, loci associated with heading date (HD) and plant height (PH) were mapped, confirming approximately 95% of known QTLs, including dwarfing genes (e.g., *Rht-B1*, *Rht-5*) and flowering-time regulators (e.g., *Vrn-A1*), as well as identifying 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 of the parental lines of the RIL population, from which the bulks were derived, aided in identifying candidate genes and regulatory networks underlying trait variation, including loci associated with photoperiod response, nutrient transport, and stress adaptation. The versatility and potential for data reuse of 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 improvements in crops.

Keywords: Bread wheat, Bulk segregant analysis, Exome capture, Single nucleotide polymorphisms, RNA-sequencing, multi-omics, Plant height, Heading time.

Short Summary

Reverse BSA-QTLseq combines molecular pre-selection of informative lines with in-silico assembly of contrasting bulks, enabling multi-trait mapping from a single sequencing dataset. This genotype-driven approach enhances QTL resolution and discovery efficiency, minimizing

experimental bias and costs while accelerating genetic analysis and marker-assisted breeding programs.

Introduction

Understanding the genetic basis of agronomic traits is crucial for advancing crop improvement programs, particularly in staple crops such as bread wheat (*Triticum aestivum* L.). One of the most effective tools in this domain is Bulk Segregant Analysis (BSA), a widely used genetic mapping technique for identifying genomic regions associated with specific traits. The conventional BSA approach 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 advancements not only reduce gaps in recombination mapping but also increase the resolution for identifying candidate regions with small-effect that might be overlooked using traditional marker systems (Zou et al., 2016; Itoh et al., 2019; Li et al., 2022; Raina et al., 2022; Wang et al., 2023).

Various NGS-enabled BSA derivatives have since emerged. Among them, QTL-seq -an approach that integrates BSA with whole-genome sequencing - has been applied across a wide range of crops, including rice (Arikiti et al., 2019; Lei et al., 2020), wheat (Xin et al., 2020; Komura et al., 2024a), tomato (Ruangrak et al., 2019), groundnut (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 (Takagi et al., 2013; Abe et al., 2012; Komura et al., 2024b). For transcriptome-level analysis, BSR-seq (Bulk Segregant RNA-seq) utilizes RNA-seq data to identify expression QTLs (eQTLs) by analyzing differentially expressed genes 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 (Ma et al., 2021; Qian et al., 2024; Yu et al., 2024; Zhu et al., 2020). Unlike DNA-based BSA methods, BSR-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 (Raina et al., 2022), and DeepBSA (Li et al., 2022) have further enhanced mapping resolution or broadened the applicability of BSA to complex populations and traits (Segawa et al., 2021; Raina et al., 2022; Chen et al., 2021; Li et al., 2022). The success of these approaches underscores the transformative potential of NGS-based BSA methods in accelerating genetic improvement in crops. Despite these advances, BSA approaches 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 for analyzing other traits, and phenotypic pre-selection remains the most labor-intensive step.

To overcome these constraints, we proposed 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 introduced BSA on the principle that contrasting phenotypes reflect contrasting alleles at the 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 primarily at trait-associated loci, while other genomic regions remain

largely unaffected due to random segregation. Initiating BSA with genetically divergent, homozygous individuals enhances the likelihood of capturing contrasting alleles at relevant loci, thereby increasing both allele frequency divergence and signal-to-noise ratio. Furthermore, even in the presence of phenotyping noise or residual heterozygosity, the strength of association remains proportional to allelic divergence between bulks, a principle that holds for both traditional BSA (based on phenotypic extremes BSA) and reverse BSA-QTLseq (based on genotypic divergence). Based on this principle, we employed principal component analysis (PCA) on low-density 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 applying the core principle of allelic homozygosity from traditional BSA in reverse, selecting genotypically divergent individuals to maximize phenotypic contrast detection.

Specifically, the Reverse BSA-QTLseq approach provides 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 take advantage of 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, using a subset of lines from a biparental RIL population for which conventional QTL mapping results from both $F_2:F_3$ and RIL generations were previously published (Vitale et al., 2021).

Using heading date and plant height as case studies, 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 a RIL population of pea (*Pisum sativum* L.), for which publicly available GBS data and phenotypic evaluations for broomrape (*Orobanche crenata* Forsk.) resistance were accessible (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.

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 reported by Vitale et al. (2021). PCA, performed on 6,795 high-quality SNPs (25 K iSelect Array), separated the parental genotypes along PC1 (23.4 % variance) and highlighted two groups representing the extremes of overall genomic divergence (Figure 1A). Twenty-three lines from each cluster, plus their respective parent, were chosen to constitute Bulk 1 and Bulk 2 (Supplementary 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 (Supplementary Table 3). Group averages were highly similar, indicating that Reverse BSA-QTLseq selected individuals

with comparable recombination landscapes to those in the phenotype-based bulks (Supplementary Table 3).

The BLUE values for HD and PH, measured for the 48 lines, demonstrated substantial variability. PH ranged from 46 cm to 107 cm, while HD spanned from 18 to 31 days from April 1st. As anticipated, the original bulks (Bulk 1 and Bulk 2) did not exhibit statistically significant differences for either trait (Figure 1B), validating the hypothesis that genetic distance-based sampling 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 ten 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 (Supplementary 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 (Supplementary Figure 1B). After identifying sequence variations, SNPs, and InDels located in annotated exons were classified according to their predicted biological effects (Supplementary Figure 1C). As expected, the majority of identified variants fall within exons and have minimal effects on protein function (Supplementary Figure 1B; Supplementary Figure 1C). Variants with deleterious effects on coding regions were also found (Supplementary Figure 1C). Among these, stop-gained variants were the most abundant, followed by stop-loss and start-loss variants.

Genomic regions putatively associated with PH

Twenty-eight QTLs for plant height were identified on eight chromosomes using both SNPs and InDels (Figure 2; Supplementary Table 2). On chromosome 2A, a region spanning 224 – 468 Mb was further divided into nine clusters (*qPH.2A.1* - *qPH.2A.9*) based on their Δ Variant index and confidence intervals at 0.01 (Figure 2; Supplementary Table 2). All 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 (Supplementary Table 2). Five QTLs were identified on chromosome 4D (*qPH.4D.1* - *qPH.4D.5*). Those detected by SNP variants (*qPH.4D.1*, *qPH.4D.3*, and *qPH.4D.4*) exhibited positive Δ Variant index values, while the two identified by InDels (*qPH.4D.2* and *qPH.4D.5*) showed negative values (Figure 2; Supplementary Table 2).

One QTL was also identified on each of chromosomes 2B, 3A, 7B, and 7D (Supplementary Table 2). In particular, *qPH.3A* and *qPH.7D* were identified by InDels and both showed a negative Δ Variant index value (Supplementary Table 2).

Genomic regions putatively associated with HD

A total of 30 QTLs across seven chromosomes were identified with both SNPs and InDels, with those on chromosomes 5A and 7D showing the lowest p-values (< 0.01) (Figure 3; Table 1). 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 Δ Variant 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*) spanned chromosome 3B, although all regions displayed low Δ Variant index values (Figure 3; Table 1). Four QTLs, all identified by SNPs, were located on chromosome 6A, each showing positive Δ Variant index values (Table 1). Three QTLs each were found on chromosomes 3A and 5B using both SNP 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; Table 1). Finally, a single region on chromosome 7D (*qHD.7D.1*), peaking at 459 Mb, showed a Δ Variant index of -0.54, suggesting the presence of major genes or regulatory loci strongly linked with HD (Figure 3; Table 1).

Transcriptome sequencing of tissues from parental lines

To investigate the molecular mechanisms underlying the phenotypic differences between the parental lines, twelve 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 data set totaling 37.82 Gb was obtained, which was used to identify 10,979 DEGs in leaves and 7,255 in roots (Figure 4A, Supplementary Table 4, and Supplementary Table 5). The comparison between leaves and roots revealed 3,213 common DEGs (21.4%), whereas 7,766 (51.7%) and 4,042 (26.9%) were leaf- and root-specific, respectively (Figure 4B). To understand the function of the DEGs, we performed Gene Ontology (GO) enrichment analysis, categorizing the top 50 enriched terms into molecular function (MF) and biological process (BP) domains (Figure 4C). The enriched GO terms confirmed differences in the genetic backgrounds of the parental lines, also 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, suggesting 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 (Supplementary Figure 2). In leaves, *Vrn-A1* and *Vrn-B1* were expressed at higher levels in LANKAODALI compared to REBELDE. Conversely, *Vrn-2* (*Vrn-2.1* and *Vrn-2.2*) transcripts were detected only in REBELDE, albeit at low levels (Supplementary Figure 2A). *Ppd-A1* and *Ppd-D1* showed a similar expression trend in both lines, whereas *Ppd-B1* was expressed only in LANKAODALI (Supplementary Figure 2A).

In roots, *Vrn-1* genes were barely expressed, with *Vrn-A1* being the only one detected (Supplementary 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 compared to leaves (Supplementary Figure 2B). *Ppd-D1* showed the highest expression in both lines, while *Ppd-B1* was expressed at a higher level in LANKAODALI (Supplementary Figure 2B).

Candidate genes identification and prioritization

Out of approximately 5,500 genes putatively associated with PH, 3,300 carried variants located within or near coding sequences, with the number of variants per gene ranging from one to 346. Most variants were predicted to have a modifier impact (e.g., intron variants, upstream or downstream) (N. = 27,401), followed by those with low-impact (N. = 6,504), moderate-impact (N. = 6,397), and high-impact (N. = 164) variants (Supplementary Table 6). In addition, among the 5,500 genes, 141 were DEGs between parental lines in both leaf and root tissues, while 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. Among them, notable genes included *TraesCS4B03G0290400* (PHD domain finger), *TraesCS2B03G0403600* (F-box protein), *TraesCS4B03G0117800* (ABC transporter), *TraesCS2B03G0441300* (Disease-resistance NBS/LRR type; a putative calcium-binding CML45), and *TraesCS4A03G1116300* (Sucrose Synthase) (Supplementary Table 6).

Similarly, among approximately 7,000 genes potentially associated with HD, 4,527 harbored variants predicted to have a modifier impact, 2,509 with low-impact, and 217 with high-impact. Of these, 265 genes were DEGs between parental lines in both leaf and root tissues, while 457 and 266 DEGs were specific to leaves and roots, respectively. Within the shared DEGs, 251

contained variants with modifier impact, 154 with moderate-impact, 184 with low-impact, and 17 with high-impact (Supplementary Table 6). Notable genes harboring high-impact included *TraesCS5A03G0634200* and *TraesCS5B03G1226800* (F-box proteins), *TraesCS4A03G1221200* (DEAD box helicase ATP-dependent), *TraesCS5A03G1012900* (Aldehyde Oxidase), *TraesCS4A03G1214600* (Dual specificity protein kinase zak2), as well as *TraesCS5B03G0863800* and *TraesCS5D03G1233100* (Disease resistance genes), all of which were differentially expressed in both tissues (Supplementary 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 [DAA]) (Borrill et al., 2019) (Supplementary Figure 3 and Supplementary Figure 4). Among the candidate genes identified, *TraesCS4A03G1116300* (Sucrose Synthase), *TraesCS3B03G0008500* (Subtilisin-like protease), *TraesCS4B03G0290400* (PHD domain-containing), and *TraesCS4D03G0456300* (EMBRYONIC FLOWER 2/VEFS) were expressed across all five tissues analyzed (root, shoot, leaf, spike, and grain) (Supplementary Figure 3). Notably, *TraesCS4B03G0290400* (OBERON-like) and *TraesCS4D03G0456300* (Polycomb group EMF2B), along with *TraesCS2B03G0441300* (Putative calcium-binding CML45), were expressed during leaf senescence stages, suggesting 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, while the latter displaying a slight decrease over time (Supplementary Figure 3).

Genomic variants located in the coding regions of candidate genes were tested using the 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 (N = 3) with significant effects was identified on chromosome 2B. These variants were located in the coding regions of *TraesCS2B03G0441300* (Putative calcium-binding CML45), *TraesCS2B03G0441600* (3 isopropyl malate dehydratase/Aconitase), and *TraesCS2B03G0442800* (NB-ARC domain-containing protein/NBS LRR disease resistance).

Two additional PH-associated variants were found on chromosome 3B, within the gene body regions of *TraesCS3B03G0099800* and *TraesCS3B03G0104900*, encoding an LRR receptor-like serine/threonine kinase and a deoxyhypusine synthase, respectively. Similarly, two significant variants were found on chromosome 7B, within *TraesCS7B03G1037100* (MYB transcription factor) and *TraesCS7B03G1031800* (F-box protein 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*, encoding homocysteine methyltransferase 2, while the variant on chromosome 4D was located in *TraesCS4D03G0561900*, which encodes the *NRT1/ PTR FAMILY 8.3* gene. Additionally, 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* (Carnitine transporter 7), *TraesCS5B03G0887900* (Serine/threonine-protein kinase SD1-8), *TraesCS5B03G1259700* (Heat shock 70), and *TraesCS6B03G0032400* (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 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 to wild-type Cadenza plants at maturity (Supplementary 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 conducted using a publicly available dataset from a different plant species, pea (*Pisum sativum* L.) (Delvento et al., 2023). The dataset included phenotypic evaluations for resistance to crenate broomrape (*Orobanche crenata* Forsk.). Six comparative analyses were performed, evaluating 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 (Δ Variant-index exceeding the 99% confidence interval) 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 increased 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, while larger bulks ($n = 25$) recovered two of the three major QTLs but showed reduced peak resolution and weaker Δ Variant-index signals. For comparison, conventional BSA-QTLseq applied to the same bulk sizes yielded overall weaker and less distinct signals, with only the QTLs on 2LG1 and 4LG4 surpassing the 95% confidence threshold when $n = 10$ (Figure 6). Three SNPs were then converted into KASP assays and used to screen the RIL population, confirming full concordance with the GBS data and validating the newly identified regions (Supplementary Figure 6).

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

Discussions

This study introduces the Reverse BSA-QTLseq approach that redefines the conventional BSA framework by utilizing *in silico* reconstruction of contrasting bulks (Figure 7). Unlike traditional phenotype-driven sampling (Figure 7A), this innovative method allows for the simultaneous analysis of multiple traits within the same initial population (Figure 7B). The new approach capitalized on genetic divergence within a RIL population as the first criterion for bulk sampling. This initial grouping, driven by SNP genotyping, was subsequently reorganized bioinformatically into trait-specific contrasting bulks, facilitating the concurrent study 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 the Reverse BSA-QTLseq approach 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, the Reverse BSA-QTLseq approach successfully identified loci associated with heading date (HD) and plant height (PH) in bread wheat, showcasing its utility across multiple traits.

Integration of genotyping and exome capture data

A phased methodological approach was employed: 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 (*QHD-5A*) and 7D (*QHD-7D*) -were detected in the late generations (F₆:F₇). Additionally, five QTLs were associated with PH. These findings were confirmed in the late generations, with an even larger number of associations compared to previous reports. The expanded QTL landscape demonstrates the effectiveness of the proposed approach in uncovering additional genomic regions associated with key agronomic traits. Most of the major QTLs reported in previous studies were successfully replicated, further validating the reliability of Reverse BSA-QTLseq (Supplementary Table 7). Notably, the results were aligned with the latest version of the bread wheat reference genome (IWGSC RefSeq v2.1; Zhu et al., 2021), allowing for meaningful comparisons, particularly with studies utilizing the same reference. A significant portion of the comparative analysis focused on the study by Jia et al. (2023), who mapped 332 QTLs, 270 GWAS loci, and 83 genes related to PH. As for QTLs associated with HD, comparisons were made with the studies by Pang et al. (2020) and Banaouda et al. (2022). Among the QTLs identified for PH, several were located in genomic regions previously associated with plant height in wheat. For example, the loci *qPH.3B.1* and *qPH.3B.2* on chromosome 3B coincide with those previously 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 plant height in several studies (Zhao et al., 2019; Shukla et al., 2015; Chen et al., 2020; Guan et al., 2018; Sannemann et al., 2018; Shirdelmoghanloo et al., 2016; Yuan et al., 2017; Wang et al., 2024) and are situated near the *Rht-B1* locus (Ellis et al., 2002; Lopes et al., 2014). Additionally, *qPH.7B* was mapped within the 655 Mb region where Abou-Elwafa et al. (2021) identified SNP IWB35241, further supporting its role in regulating plant height. 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*, located 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 plant height and heading date (Supplementary Table 7). This is particularly evident in *qHD.5A.4–8*, which spans 562–626 Mb on chromosome 5A and includes the *Vrn-A1* locus (*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, indicating that they may correspond to previously unreported genomic regions influencing these traits. This highlights the sensitivity of the Reverse BSA-QTLseq approach, even when applied to a limited number of lines in reconstructed bulks. However, further investigation is needed to confirm the functional role of these loci in trait determination.

Regarding the heading date, *qHD.3A.1* overlapped with *qHD3A.1* identified by Pang et al. (2020), whereas *qHD.5A.3* and *qHD.6A.2* were found nearby *qHD5A.1* and *qHD6A.1* (< 50 Mb), respectively (Pang et al., 2020). In contrast, *qHD.5A.2*, *qHD.5A.3*, *qHD.5A.4*, *qHD.5A.5*,

qHD.5A.6, *qHD.5A.7*, *qHD.5A.8*, and *qHD.5A.9* coincided with *TaHd102*, reported by Banaouda et al. (2022), while *qHD.5B.1* was located ~117 Kb from *TaHd125* (Banaouda et al., 2022). In particular, *qHD.5A.4.8*, spanning the region between 562 – 626 Mb on chromosome 5A, included the *Vrn-A1* locus (*TraesCS5A03G0935400*) at 589 Mb. This strong alignment with the existing literature provides robust support for the effectiveness of the proposed methodology in accurately identifying and mapping loci associated with plant height and heading date. Our aim was also to demonstrate that, by using the reverse BSA-QTLseq approach, it is possible to identify a significant number of associations even with a limited number of lines, comparable to, or exceeding, results from traditional QTL and QTL-seq mapping methods (Zheng et al., 2024). For instance, in wheat, Komura et al. (2024) applied QTL-seq using F3 progeny and detected three QTLs for HD. In rice, studies using the BSA QTL-seq approach are even more common (e.g., Ogiso-Tanaka et al., 2017; Nasution et al., 2024; Zhang et al., 2020; Wang et al., 2024). For example, Wang et al. (2024) utilized contrasting F2 and F3 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 (Wang et al., 2023; Zou et al., 2016).

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 and Dubcovsky 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 (Ghosh and Xu, 2014; Bostock et al., 2014; Trinidad et al., 2016). This variability in defense gene expression may influence resilience or susceptibility to soil-borne pathogens and abiotic stressors.

On the other hand, 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 plant height and heading date. Notably, genes such as *TraesCS4B03G0290400* (*PHD domain-containing*), *TraesCS2B03G0403600* (*F-box protein*), *TraesCS4B03G0117800* (*ABC transporter*), and *NBS/LRR-type disease-resistance* genes appear to influence PH through mechanisms related to nutrient transport and plant growth/defense signaling (Krattinger et al., 2009; Gu et al., 2015; Sánchez-Martín et al., 2021; Wang et al., 2024b). Similarly, the identification of seventeen 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 like Sucrose Synthase (*TraesCS4A03G1116300*) and embryonic flower 2/vefs (*TraesCS4D03G0456300*) to spike and grain development (Stein et al., 2019; Hou et al., 2014; Roux et al., 2006). Furthermore, significant allelic effects in multiple genes (e.g., *CML45*, *MYB transcription factors*, *F-box proteins*) reinforce their potential role in PH and HD variation under field conditions (Zhang et al., 2019; Dubos et al., 2010). A TILLING mutant confirmed *TraesCS2B03G0403600* as a gene affecting plant height, 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.

Taken together, 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 utilizing advanced molecular tools such as gene editing, additional TILLING lines, or 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 the Reverse BSA-QTLseq approach for high-resolution mapping of *Orobanche crenata* resistance in pea

To validate the robustness and applicability of the Reverse BSA-QTLseq approach, we utilized the publicly available pea dataset from Delvento et al. (2023), a well-established benchmark for QTL mapping of resistance to *Orobanche 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 on a 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, corresponding precisely to the positions of *PsOcr-1*, *PsOcr-2*, and *PsOcr-3*, when applying the Reverse BSA-QTLseq approach with optimized bulk sizes. Importantly, the detection of additional candidate regions using moderate bulk sizes (N. = 10) indicates an enhanced sensitivity to identify 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 the conventional BSA-QTLseq strategy, which recovered only a subset of the major QTLs and at lower resolution, our findings demonstrate that the Reverse BSA-QTLseq approach maintains or improves mapping efficiency, as confirmed by the development of new KASP marker assays. This holds particular promise for future applications targeting complex traits controlled by multiple loci of small to moderate effect, where resource limitations often restrict extensive genotyping and phenotyping efforts. The successful validation of key loci using an additional, species-divergent dataset underscores the potential of Reverse BSA-QTLseq as a robust and scalable approach for QTL discovery across diverse crop species, delivering high resolution while 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 advancement over conventional BSA methods. By preserving individual-level genetic information, this strategy enables the redefinition of bulks tailored to different traits, providing a cost-effective and scalable solution for investigating complex polygenic traits. This capability is particularly valuable in crops like wheat, where genetic diversity, large genome size, and trait complexity, including genotype-by-environment-by-management (GxExM) interaction, 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 the Reverse BSA-QTLseq approach 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 contrasts at causal loci. Nonetheless, other factors, including the number and effect size 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 the Reverse BSA-QTLseq approach

Although the Reverse BSA-QTLseq strategy 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., 2017; Cobb et al., 2019). In contrast, Reverse BSA-QTLseq enables researchers to concentrate 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), this approach offers a more sustainable and scalable model for dissecting complex traits. Moreover, the reusability of the genotypic dataset across multiple traits amplifies these economic benefits over time, especially as genotyping costs continue to decline while field phenotyping remains the primary bottleneck in crop improvement pipelines (Scheben et al., 2017; Yang et al., 2020; Reynolds et al., 2021; Großkinsky et al., 2015).

Implications for wheat breeding and beyond

Heading date and plant height are key traits for adapting wheat to diverse agro-climatic conditions, optimizing resource use, and ensuring yield stability under both abiotic and biotic stresses (Zhao et al., 2019; Borrill et al., 2019). By validating previously reported QTLs and identifying novel candidate loci, this study highlights the practical value of the Reverse BSA-QTLseq approach for marker-assisted selection. The markers linked to these 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 primary sampling bulk, enabling the identification of eQTLs 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 gene expression levels.

Conclusion

This study provides compelling evidence that Reverse BSA-QTLseq is a powerful and versatile strategy for trait mapping in wheat and potentially 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.

Experimental procedures

Plant materials and field experiments

The initial genetic material consisted of a population of recombinant inbred lines (RILs) derived from a cross between two phenotypically contrasting bread wheat varieties, LANKAODALI and REBELDE, which differ in several morpho-physiological traits, as reported by Vitale et al. (2021). The population was advanced to the F7 generation using the single seed descent (SSD) method, starting from the F2 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 considered. 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 rainfed conditions. In contrast, in the final two growing seasons, two irrigation treatments were applied (rainfed and well-irrigated), resulting in six year-by-treatment combinations. The field trials were conducted on clay soil classified according to 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 per m² was employed. Experimental plots, each measuring 10.2 m² (1.36 m wide × 7.5 m long), were arranged in a randomized complete block design with three replications per line. Crop management followed standard agronomic practices.

Bulk sampling criteria

Contrasting bulks were initially identified by maximizing genetic distance through principal component analysis (PCA), using single nucleotide polymorphism (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), applying a filter for a call rate below 95% and a minor allele frequency (MAF) below 5%. A total of 6,795 SNPs distributed across all 21 chromosomes were used for PCA to explore genetic relationships within the population. The selection of contrasting bulks was identified based on the first principal component (PC1) according to the scattering of the two parental lines (Figure 1 and Supplementary 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, each assigned to one of the groups (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 (ROH) were identified using PLINK v1.9 (Purcell et al., 2007) with high-quality SNPs and the following parameters: homozyg-kb = 100 --homozyg-snp 5. Recombination rate was then estimated by counting recombination events: markers were sorted based on 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 the Reverse BSA-QTLseq approach 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 approach enabled the redefinition of bulks after genotyping based on target traits. Raw reads deposited by Esposito et al. (2022) were re-analyzed using the updated bread wheat genome version v2.1 (Zhou 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 score < 20; SLIDING WINDOW = 6). High-quality reads were aligned to the Chinese spring wheat genome v2.1 (Zhou 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 et al., 2009a). Duplicate reads were removed using the Picard Markduplicate 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 re-call variants within each bulk, following the best practice and hard filtering procedure (Van der Auwera & O'Connor, 2020). Row SNPs were filtered based on the depth of coverage (DP) > 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 and heading date

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

First, Best Linear Unbiased Estimate (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} are the observed values of the i^{th} genotype within the k^{th} replicate at the j^{th} environment, μ is the overall mean, g_i corresponds to the effect of the i^{th} genotype assuming it was a fixed effect, t_j is the j^{th} environment (all combination treatment-year-location) effect modelled as the random effect with $t_j \sim N(0, I\sigma_t^2)$, $r_{k(j)}$ is the effect of the k^{th} replicates within the trial considered as random factor with $r_{k(j)} \sim N(0, I\sigma_r^2)$, gt_{ij} represents the genotype by environment interaction modelled as random effect with $gt_{ij} \sim N(0, I\sigma_{gt}^2)$, and e_{ijk} is the residuals effect considered as random factor and assuming to have a normal distribution $e_{ijk} \sim N(0, I\sigma_e^2)$. This model was constructed within the R environment using the 'lmer' function from the 'lme4' package (Bates et al., 2016). Then, the resulting values were 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 then merged, creating 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 the best practice and hard filtering procedure (Van der Auwera & O'Connor, 2020). The function *FromGATKToTable* was used to extract the DP and AD values and prepare input files for BSA-QTLseq analysis, which was performed using the *QTL-seqR* package (Magwene et al., 2011) with the following criteria: minimal 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 $\text{SNP index}_i = \text{alternate allele depth} / \text{total read depth}$, where *i* represents the location of the SNP. In a given genomic interval, the distribution of Δ (SNP index) between the bulks was estimated using a 1-Mb sliding window. The regions where the Δ exceeded the 95% and 99% confidence intervals, 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 (60 million reads per sample) was performed on the original parents, REBELDE and LANKAODALI. RNA was extracted from three biological replicates per variety using 1 mL of TRIzol (Thermo Fisher Scientific, Wilmington, DE, United States), according to the manufacturer's protocol, yielding an average RNA concentration of 2,000 ng/ μ L with a RIN > 8.0. RNA quality and concentration were assessed using the Qubit (Thermo Fisher Scientific, Wilmington, DE, United States) and NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, United States). Sequencing was performed using the 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 (Zhou et al., 2021) was used as a reference.

The *HTSFilter* package (Rau et al., 2013) was used to remove lowly expressed genes based on the Jaccard similarity index, and the Trimmed Means of M-values (TMM) normalization method was then 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 the *T. aestivum* proteins was uploaded, setting the minimum query and hit coverage at 0.4 and the minimum alignment length at 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, leveraging *bedtools intersect* (Quinlan et al., 2010). To prioritize candidate genes potentially involved in PH and HD, a systematic SNP prioritization approach was undertaken. First, candidate genes located near significant QTLs and harboring genetic variants were identified and categorized using *bedtools intersect* (Quinlan et al., 2010). Then, 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 the 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 (in Transcript per Million, TPM) 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 to 26 days after anthesis, DAA). 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). Briefly, lines with high-impact mutations were identified using the JBrowse graphical interface on Ensemble Plants (<https://plants.ensembl.org/index.html>). The corresponding M4 seeds were obtained through the SeedStore repository (Horler et al., 2018) and sown in 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) (Supplementary Table 8).

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

The Reverse BSA-QTLseq approach was independently validated using publicly available data from Delvento et al. (2023). Genotyping data in Variant Calling Format (VCF) for 148 F₇ 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 addition, BLUP values calculated from two field trials and raw sequencing data (FASTQ file) 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 the 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) (Supplementary Figure 5).

Author contributions

Conceptualization P.D., N.D., F.T.; Methodology S.E., N.D., F.T., and F.F.; Data Curation, Formal Analysis and Visualization S.E.; Investigation L.D., F.S., F.F.; Writing-Original Draft S.E., P.D.; Supervision, Funding Acquisition and Resources P.D.; Validation S.P. and I.C.; Writing-Review & Editing ALL.

Acknowledgments

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).

Conflict of interest

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

Data availability statement

The datasets presented in this study can be found on 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>

Figure legends

Figure 1. Genetic distance analysis of the segregating population was conducted, leading to the identification of two distinct bulks, designated as Bulk 1 and Bulk 2, which capture the maximum genetic divergence within the population (left panel). BLUE values for plant height (PH, cm) (top panel) and heading date (HD, days from April 1st) (bottom panel) derived from phenotypic data across four growing seasons. Data were collected under rainfed conditions for the first two years and under both rainfed and well-irrigated conditions for the final two years. The left panels display BLUE values for the PCA-based clustering bulks (Bulk 1 and Bulk 2), while the right panels show BLUE values for phenotypically re-ordered bulks for PH (high and low) and HD (late and early).

Figure 2. Quantitative trait locus (QTL) identified for plant height (PH) at maturity. (A) The plots show the distribution of Δ Variant index values across chromosomes for single nucleotide polymorphisms (SNPs) and insertions and deletions (InDels) (B). The x-axis represents chromosomes, labeled from 1A to 7D and Unk (unmapped regions). Blue and red horizontal lines denote 95% and 99% confidence intervals, respectively, indicating significant QTL regions.

Figure 3. Quantitative trait locus (QTL) identified for heading date (HD) at maturity. (A) The plots show the distribution of Δ Variant index values across chromosomes for single nucleotide polymorphisms (SNPs) and insertions and deletions (InDels) (B). The x-axis represents chromosomes, labeled from 1A to 7D and Unk (unmapped regions). Blue and red horizontal lines denote 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-D) Enriched Gene Ontology (GO) terms associated with DEGs in leaves and roots. Different colors indicate GO categories in the molecular function and biological process domains.

Figure 5. Boxplot of variants affecting plant height (PH) and heading date (HD) ($p < 0.01$). Genotypes are grouped by allele type on the x-axis, while the y-axis represents the BLUE values for the corresponding traits. Homo 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). Homo Alt (Homozygous Alternate): indicates that both copies of the allele at a genetic locus differ from the reference sequence. Het (Heterozygous): This refers to an individual having one reference allele and one alternate allele at a particular locus.

Figure 6. Validation of the Reverse BSA-QTLseq. The plots show the distribution of Δ Variant index values across pea chromosomes obtained by Reverse BSA-QTLseq and conventional BSA-QTLseq under three different bulk sizes ($n = 5, 10, \text{ and } 25$). The x-axis represents chromosomes, labeled from 1LG6 to 7LG7. Blue and red horizontal lines denote 95% and 99% confidence intervals, respectively, indicating significant QTL regions.

Figure 7. Working model. The flowchart illustrates the methodological framework of two complementary approaches: Classical BSA-QTLseq and Reverse BSA-QTLseq. In the 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 the Reverse BSA-QTLseq, after a biparental cross (1), a high-density SNP array sequencing is performed on the entire segregating population as well as the two parental lines. Then, a PCA is conducted on the segregant population and used 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, which capture 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 utilized to perform QTL-seq analyses for each trait (4.1, 4.2 and 4.3).

Short Legends for Supporting Information

Supplementary Figure 1. Variant calling summary and annotation. A) Number of variants per chromosome. B) Top ten most frequent category variants. C) Most abundant coding consequences.

Supplementary Figure 2. Expression pattern of *Vrn-1*, *Vrn-2*, and *Ppd-1*. Expression levels of *Vrn-1* (*Vrn-A1*, *Vrn-B1*, and *Vrn-D1*), *Vrn-2* (*Vrn-2.1* and *Vrn2.2*), and *Ppd-1* (*Ppd-A1*, *Ppd-B1*, and *Ppd-D1*) in LANKAODALI and REBELDE detected in leaves (A) and roots (B). Transcript per million (TPM) values were used.

Supplementary Figure 3. Expression pattern of candidate genes for PH. Expression levels of candidate genes for plant height (PH) analyzed using public RNAseq data from the wGRN database (<http://wheat.cau.edu.cn/wGRN/>). Specifically, five tissues (root, shoot, leaf, spike, and grain), ten grain developmental stages (two-cell embryo = E1; pre-embryo = E2, transition = E3, leaf early = E4, leaf middle = E5, leaf late = E6, mature embryo = E7, transition stage endosperm = E8, leaf late stage endosperm = E9, and the leaf early stage pericarp = E10), spikes taken from 0 day post-floret meristem (stage B1) to 34-day post-floret meristem stage (stage B11) with an interval of 4 days (stage B1 to B5) or 3 days (stage B6 to B11) and during leaf senescence at different days after anthesis (DAA). References for each dataset are reported in Materials and Methods.

Supplementary Figure 4. Expression pattern of candidate genes for HD. Expression levels of candidate genes for heading date (HD) analyzed using a public RNAseq dataset from the wGRN database (<http://wheat.cau.edu.cn/wGRN/>). Specifically, five tissues (root, shoot, leaf, spike, and grain), ten grain developmental stages (two-cell embryo = E1; pre-embryo = E2, transition = E3, leaf early = E4, leaf middle = E5, leaf late = E6, mature embryo = E7, transition stage endosperm = E8, leaf late stage endosperm = E9, and the leaf early stage pericarp = E10), spikes taken from 0 day post-floret meristem (stage B1) to 34-day post-floret meristem stage (stage B11) with an interval of 4 days (stage B1 to B5) or 3 days (stage B6 to B11) and during leaf senescence at different days after anthesis (DAA). References for each dataset are reported in the Materials and Methods.

Supplementary Figure 5. Phenotypic comparison between the Cadenza1329 TILLING line (WCAD1329), carrying a high-impact nonsense mutation (W/*) at the start of the second exon of the candidate gene *TraesCS2B03G0403600*, and the wild-type (WT) Cadenza1329. Homozygous mutants were confirmed by Sanger sequencing performed on both strands (forward and reverse) to ensure accuracy and confirm the mutation reliably.

Supplementary Figure 6. Validation of QTL using Kompetitive Allele-Specific PCR (KASP) assays in *Pisum*. The scatter plots display genotyping results for three SNP markers (A=5LG3:169428426; B=6LG2:168321201; C=7LG7:360833885) where red dots represent individuals homozygous for the HEX-type allele; blue dots indicate homozygous individuals for the FAM-type allele, and green dots correspond to heterozygous individuals. Black dots denote negative controls, while pink dots indicate uncallable genotypes.

Supplementary Table 1. List of 48 genotypes used in the present study (**in bold**), selected from the initial RIL population (Vitale et al., 2021) based on divergent values of the first two components obtained from PCA analysis.

Supplementary Table 2. QTLs associated with plant height (PH) and heading date (HD) identified using QTL-seq and three statistics (G' , Δ Variant, and p.value). For each region, the peak Δ Variant value, along with its physical position and the related variant, is also reported.

Supplementary Table 3. Summary of crossover (CO) counts and runs of homozygosity (ROH) in bulks selected using the Reverse BSA-QTLseq and phenotype-based approaches. For each group, the average recombination rate and total ROH length (in kilobases) are also reported.

Supplementary Table 4. List of DEGs in leaves.

Supplementary Table 5. List of DEGs in roots.

Supplementary Table 6. List of candidate genes identified under significant QTLs, harboring high impact variants and being differentially expressed among parental lines.

Supplementary Table 7. QTL comparison with the literature. The comparison was carried out using the physical positions reported by Xu et al. (2023). The common QTL name, along with the start and end position for each region, is reported.

878 **Supplementary Table 8.** Primer sequences used for Sanger sequencing to confirm the target mutations in the
879 CadenzaTILLING lines.
880

Journal Pre-proof

References

- Abd-Hamid, N. A., Ahmad-Fauzi, M. I., Zainal, Z., & Ismail, I.** (2020). Diverse and dynamic roles of F-box proteins in plant biology. *Planta* 251, 1-31. doi: 10.1007/s00425-020-03356-8
- Abe, A., Kosugi, S., Yoshida, K., et al.** (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nature Biotechnology* 30(2), 174–178. <https://doi.org/10.1038/nbt.2095>
- Abou-Elwafa, S.F., Shehzad, T.** (2021). Genetic diversity, GWAS and prediction for drought and terminal heat stress tolerance in bread wheat (*Triticum aestivum* L.). *Genetic Resources and Crop Evolution* 68, 711–728. <https://doi.org/10.1007/s10722-020-01018-y>
- Araus, J.L., Cairns, J.E.** (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* 19(1), 52-61. doi: 10.1016/j.tplants.2013.09.008.
- Arikat, S., Wanchana, S., Khanthong, S., Saensuk, C., Thianthavon, T., et al.** (2019). QTL-seq identifies cooked grain elongation QTLs near soluble starch synthase and starch branching enzymes in rice (*Oryza sativa* L.). *Scientific Reports* 9,1–10. doi:10.1038/s41598-019-44856-2.
- Branham SE, Patrick W, Lambel S, Massey L, Ma M, et al.** (2018). QTL-seq and marker development for resistance to *Fusarium oxysporum* f. sp. *niveum* race 1 in cultivated watermelon. *Molecular Breeding* 38, 139. doi:10.1007/s11032-018-0896-9.
- Benaouda, S., Dadshani, S., Koua, P. et al.** (2022). Identification of QTLs for wheat heading time across multiple-environments. *Theoretical and Applied Genetics* 135, 2833–2848. <https://doi.org/10.1007/s00122-022-04152-6>
- Borrill, P., Harrington, S.A., Simmonds, J., Uauy, C.** (2019). Identification of Transcription Factors Regulating Senescence in Wheat through Gene Regulatory Network Modelling. *Plant Physiology* 180(3), 1740–1755, <https://doi.org/10.1104/pp.19.00380>
- Bostock, R. M., Pye, M. F., & Roubtsova, T. V.** (2014). Predisposition in plant disease: exploiting the nexus in abiotic and biotic stress perception and response. *Annual review of phytopathology* 52(1), 517-549. 10.1146/annurev-phyto-081211-172902
- Chen, H., Bemister, D. H., Iqbal, M., Strelkov, S. E., & Spaner, D. M.** (2020). Mapping genomic regions controlling agronomic traits in spring wheat under conventional and organic managements. *Crop Science* 60(4), 2038-2052. <https://doi.org/10.1002/csc2.20157>
- Chen, M., Fan, W., Ji, F., Hua, H., Liu, J., Yan, M., & Huang, X.** (2021). Genome-wide identification of agronomically important genes in outcrossing crops using OutcrossSeq. *Molecular Plant* 14(4), 556-570. <https://doi.org/10.1016/j.molp.2021.01.003>
- Cobb, J.N., DeClerck, G., Greenberg, A. et al.** (2013). Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. *Theoretical and Applied Genetics* 126, 867–887. doi: 10.1007/s00122-013-2066-0
- Cui, C., Lu, Q., Zhao, Z., Lu, S., Duan, S., Yang, Y., Qiao, Y., Chen, L., 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(6), 114. doi: 10.1007/s00425-022-03888-1.

- 925 **Daoura, B. G., Chen, L., Du, Y., & Hu, Y. G.** (2014). Genetic effects of dwarfing gene *Rht-5*
 926 on agronomic traits in common wheat (*Triticum aestivum* L.) and QTL analysis on its
 927 linked traits. *Field Crops Research* 156, 22-29. <https://doi.org/10.1016/j.fcr.2013.10.007>
- 928 **Delvento, C., Arcieri, F., Marcotrigiano, A.R., et al.** (2023). High-density linkage mapping
 929 and genetic dissection of resistance to broomrape (*Orobanche crenata* Forsk.) in pea
 930 (*Pisum sativum* L.). *Frontiers in Plant Science* 14,1216297.
 931 <https://doi.org/10.3389/fpls.2023.1216297>
- 932 **Deokar, A., Sagi, M., Daba, K., Tar'an, B.** (2019). QTL sequencing strategy to map genomic
 933 regions associated with resistance to ascochyta blight in chickpea. *Plant Biotechnology*
 934 *Journal* 17, 275–288. doi:10.1111/pbi.12964.
- 935 **Distelfeld, A., Li, C., & Dubcovsky, J.** (2009). Regulation of flowering in temperate
 936 cereals. *Current opinion in plant biology* 12(2), 178-184. doi: 10.1016/j.pbi.2008.12.010
- 937 **Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., & Lepiniec, L.** (2010).
 938 MYB transcription factors in *Arabidopsis*. *Trends in Plant Science* 15(10), 573-581. doi:
 939 10.1016/j.tplants.2010.06.005.
- 940 **Ellis, H., Spielmeyer, W., Gale, R., Rebetzke, J., Richards, A.** (2002). "Perfect" markers for
 941 the *Rht-B1b* and *Rht-D1b* dwarfing genes in wheat. *Theoretical and Applied Genetics*
 942 105(6-7) 1038-1042. doi: 10.1007/s00122-002-1048-4.
- 943 **Esposito, S., D'Agostino, N., Taranto, F., Sonnante, G., Sestili, F., Lafiandra, D. and De**
 944 **Vita, P.** (2022). Whole-exome sequencing of selected bread wheat recombinant inbred
 945 lines as a useful resource for allele mining and bulked segregant analysis. *Frontiers*
 946 *Genetics* 13:1058471. doi: 10.3389/fgene.2022.1058471
- 947 **Ghosh, D., & Xu, J.** (2014). Abiotic stress responses in plant roots: a proteomics
 948 perspective. *Frontiers in Plant Science* 5, 6. <https://doi.org/10.3389/fpls.2014.00006>
- 949 **Griffiths, S., Wingen, L., Pietragalla, J., et al.** (2015). Genetic dissection of grain size and
 950 grain number trade-offs in CIMMYT wheat germplasm. *PloS one* 10(3), e0118847. doi:
 951 10.1371/journal.pone.0118847
- 952 **Großkinsky, D.K., Svensgaard, J., Christensen, S., Roitsch, T.** (2015). Plant phenomics and
 953 the need for physiological phenotyping across scales to narrow the genotype-to-
 954 phenotype knowledge gap. *Journal of Experimental Botany* 66(18),5429-40. doi:
 955 10.1093/jxb/erv345
- 956 **Gu, L., Si, W., Zhao, L., Yang, S., & Zhang, X.** (2015). Dynamic evolution of NBS–LRR genes
 957 in bread wheat and its progenitors. *Molecular Genetics and Genomics* 290, 727-738. doi:
 958 10.1007/s00438-014-0948-8
- 959 **Guan, P., Lu, L., Jia, L., et al.** (2018). Global QTL analysis identifies genomic regions on
 960 chromosomes 4A and 4B harboring stable loci for yield-related traits across different
 961 environments in wheat (*Triticum aestivum* L.). *Frontiers in Plant Science* 9, 529. doi:
 962 10.3389/fpls.2018.00529
- 963 **Hickey, L.T.N., Hafeez, A., Robinson, H., et al.** (2019). Breeding crops to feed 10 billion.
 964 *Nature Biotechnology* 37(7):744-754. <https://doi.org/10.1038/s41587-019-0152-9>
- 965 **Horler, R.S.P., Turner, A.S., Fretter, P., Ambrose, M.** (2018). SeedStor: A Germplasm
 966 Information Management System and Public Database. *Plant and Cell Physiology* 59(1),
 967 e5. doi: 10.1093/pcp/pcx195

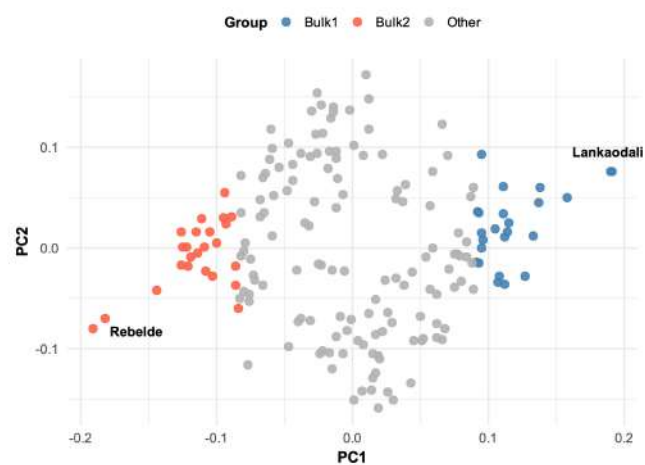
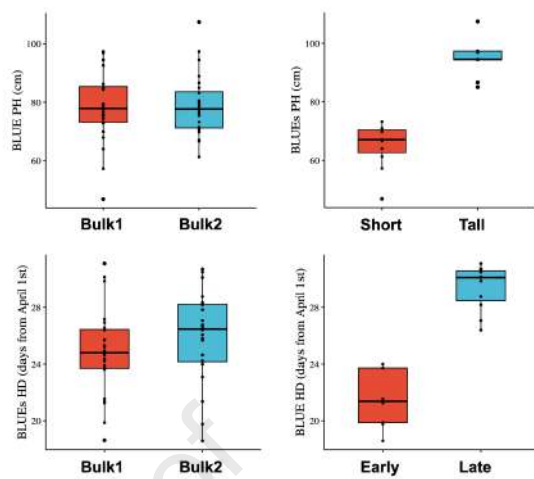
- 968 **Hou, J., Jiang, Q., Hao, C., Wang, Y., Zhang, H., & Zhang, X.** (2014). Global selection on
 969 sucrose synthase haplotypes during a century of wheat breeding. *Plant*
 970 *Physiology* 164(4), 1918-1929. doi:10.1104/pp.113.232454
- 971 **Huang, L., Tang, W., Bu, S., Wu, W.** (2020). BRM: a statistical method for QTL mapping
 972 based on bulked segregant analysis by deep sequencing. *Bioinformatics* 36, 2150–2156.
 973 doi:10.1093/bioinformatics/btz861.
- 974 **Imerovski, I., Dedić, B., Cvejić, S., Miladinović, D., Jocić, S., et al.** (2019). BSA-seq mapping
 975 reveals major QTL for broomrape resistance in four sunflower lines. *Molecular Breeding*
 976 39, 41. doi:10.1007/s11032-019-0948-9
- 977 **Itoh, N., Segawa, T., Tamiru, M., et al.** (2019). Next-generation sequencing-based bulked
 978 segregant analysis for QTL mapping in heterozygous species *Brassica rapa*. *Theoretical*
 979 *and Applied Genetics* 132(2), 455–467. <https://doi.org/10.1007/s00122-019-03396-z>
- 980 **Jia, C., Lyu, X., Yang, T., Qin, H., Wang, Y., Hao, Q., ... & Ma, W.** (2023). In silico curation
 981 of QTL-rich clusters and candidate gene identification for plant height of bread wheat.
 982 *The Crop Journal* 11(5), 1480-1490. <https://doi.org/10.1016/j.cj.2023.05.007>
- 983 **Kiss, T., Horváth, Á. D., Cseh, A., Berki, Z., Balla, K., & Karsai, I.** (2024). Molecular genetic
 984 regulation of the vegetative–generative transition in wheat from an environmental
 985 perspective. *Annals of Botany* mcae174. doi: 10.1093/aob/mcae174.
- 986 **Komura, S., Kobayashi, F., Oono, Y., Handa, H., Inoue, Y., & Yoshida, K.** (2024).
 987 Identification of three QTLs that additively affect heading time in bread wheat (*Triticum*
 988 *aestivum* L.) by QTL-seq approach. *Euphytica* 220(12), 183.
 989 <https://doi.org/10.1007/s10681-024-03441-z>
- 990 **Komura, S., Yoshida, K., Jinno, H., Oono, Y., Handa, H., Takumi, S., & Kobayashi, F.**
 991 (2024b). Identification of the causal mutation in early heading mutant of bread wheat
 992 (*Triticum aestivum* L.) using MutMap approach. *Molecular Breeding* 44(6), 1-20.
 993 <https://doi.org/10.1007/s11032-024-01478-5>
- 994 **Krasileva, K.V., Vasquez-Gross, H.A., Howell, T., et al.** (2017). Uncovering hidden variation
 995 in polyploid wheat. *Proc Natl Acad Sci USA*. 2017 Feb 7;114(6):E913-E921. doi:
 996 10.1073/pnas.1619268114. Epub 2017 Jan 17. Erratum in: *Proc Natl Acad Sci U S A*.
 997 Sep 12;120(37):e2313894120. doi: 10.1073/pnas.2313894120
- 998
- 999 **Krattinger, S. G., Lagudah, E. S., Spielmeier, W., et al.** (2009). A putative ABC transporter
 1000 confers durable resistance to multiple fungal pathogens in wheat. *Science* 323(5919),
 1001 1360-1363. doi: 10.1126/science.1166453
- 1002 **Langridge, P., Reynolds, M.** (2021). Breeding for drought and heat tolerance in
 1003 wheat. *Theoretical and Applied Genetics* 134, 1753–1769.
 1004 <https://doi.org/10.1007/s00122-021-03795-1>
- 1005 **Li, Z., and Yuhui X.** (2022). Bulk segregation analysis in the NGS era: a review of its teenage
 1006 years. *The Plant Journal* 109(6), 1355-1374. doi: 10.1111/tpj.15646
- 1007 **Lei, L., Zheng, H., Bi, Y., Yang, L., Liu, H., et al.** (2020). Identification of a major QTL and
 1008 candidate gene analysis of salt tolerance at the bud burst stage in rice (*Oryza sativa* L.)
 1009 using QTL-seq and RNA-seq. *Rice*, 13,14. doi:10.1186/s12284-020-00416-1.
- 1010 **Lopes MS, Dreisigacker S, Peña RJ, Sukumaran S, Reynolds MP.** (2014). Genetic
 1011 characterization of the wheat association mapping initiative (WAMI) panel for dissection
 1012 of complex traits in spring wheat. *Theoretical and Applied Genetics* 128(3), 453-64. doi:
 1013 10.1007/s00122-014-2444-2.

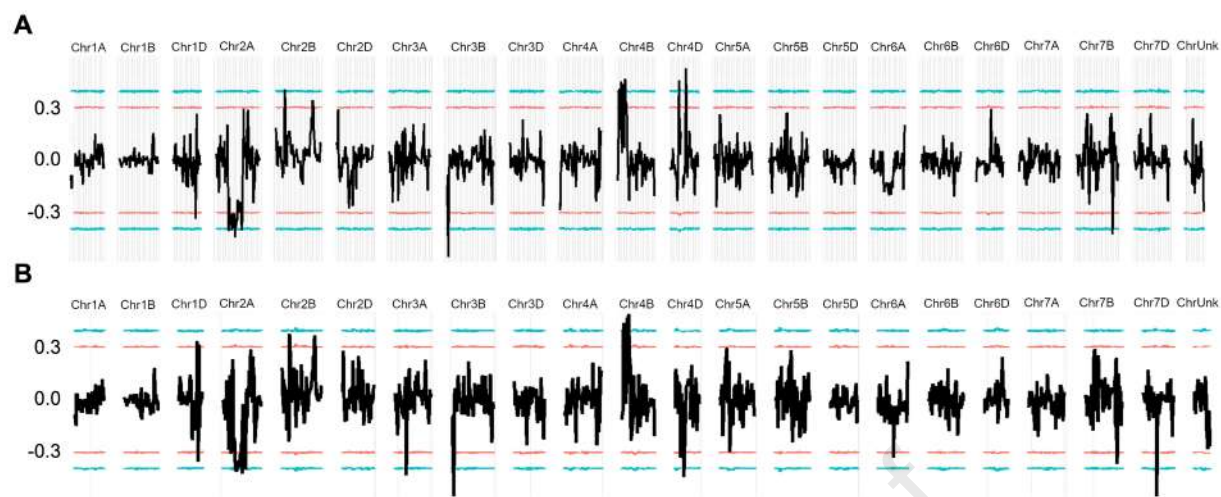
- 1014 **Ma, P., Wu, L., Xu, Y., et al.** (2021). Bulk segregant RNA-Seq provides distinctive
 1015 expression profile against powdery mildew in the wheat genotype YD588. *Frontiers in*
 1016 *Plant Science* 12, 764978. <https://doi.org/10.3389/fpls.2021.764978>
- 1017 **Majeed, A., Johar, P., Raina, A., Salgotra, R. K., Feng, X., & Bhat, J. A.** (2022). Harnessing
 1018 the potential of bulk segregant analysis sequencing and its related approaches in crop
 1019 breeding. *Frontiers in Genetics* 13, 944501. <https://doi.org/10.3389/fgene.2022.944501>
- 1020 **Michelmore, R.W., Paran, I., Kesseli, R.V.** (1991). Identification of markers linked to disease-
 1021 resistance genes by bulked segregant analysis: a rapid method to detect markers in
 1022 specific genomic regions by using segregating populations. *Proc Natl Acad Sci USA*
 1023 88,9828–9832. doi: 10.1073/pnas.88.21.9828.
- 1024 **Nasution, K. Y., Satyawati, D., Yunus, M., et al.** (2024). Detection of genomic loci associated
 1025 with days to heading in tropical japonica rice through QTL-seq. *Czech Journal of*
 1026 *Genetics and Plant Breeding* 61(1),23-30. doi: 10.17221/66/2024-CJGPB
- 1027 **Ogiso-Tanaka, E., Tanaka, T., Tanaka, K., et al.** (2017). Detection of novel QTLs qDTH4. 5
 1028 and qDTH6. 3, which confer late heading under short-day conditions, by SSR marker-
 1029 based and QTL-seq analysis. *Breeding Science* 67(2), 101-109. doi: 10.1270/jsbbs.16096
- 1030 **Pandey, M.K., Khan, A.W., Singh, V.K., Vishwakarma, M.K., Shasidhar, Y. et al.** (2017).
 1031 QTL-seq approach identified genomic regions and diagnostic markers for rust and late
 1032 leaf spot resistance in groundnut (*Arachis hypogaea* L.). *Plant Biotechnol Journal* 15,
 1033 927–941. doi:10.1111/pbi.12686.
- 1034 **Pang, Y., Liu, C., Wang, D., et al.** (2020). High-resolution genome-wide association study
 1035 identifies genomic regions and candidate genes for important agronomic traits in
 1036 wheat. *Molecular Plant* 13(9), 1311-1327. doi: [10.1016/j.molp.2020.07.008](https://doi.org/10.1016/j.molp.2020.07.008)
- 1037 **Raina, A., Salgotra, R. K., Feng, X., et al.** (2022). Harnessing the potential of bulk segregant
 1038 analysis sequencing and its related approaches in crop breeding. *Frontiers in Genetics* 13,
 1039 944501. <https://doi.org/10.3389/fgene.2022.944501>
- 1040 **Ramos, A., Fu, Y., Michael, V., Meru, G.** (2020). QTL-seq for identification of loci associated
 1041 with resistance to Phytophthora crown rot in squash. *Scientific Reports* 10, 1–8.
 1042 doi:10.1038/s41598-020-62228-z.
- 1043 **Rau, A., Gallopin, M., Celeux, G., Jaffrezic, F., Rau, M.A.** (2013). Data-based filtering for
 1044 replicated high-throughput transcriptome sequencing experiments. *Bioinformatics*
 1045 29(17), 2146-2152. <https://doi.org/10.1093/bioinformatics/btt350>
- 1046 **Reynolds, D., Baret, F., Welcker, C., et al.** (2019). What is cost-efficient phenotyping?
 1047 Optimizing costs for different scenarios. *Plant Science* 282,14-22.
 1048 <https://doi.org/10.1016/j.plantsci.2018.06.015>
- 1049 **Reynolds, M., Atkin, O.K., Bennett, M., et al.** (2021). Addressing Research Bottlenecks to
 1050 Crop Productivity. *Trends Plant Science* 26(6),607-630.
 1051 <https://doi.org/10.1016/j.tplants.2021.03.011>
- 1052 **Roux, F., Touzet, P., Cuguen, J., & Le Corre, V.** (2006). How to be early flowering: an
 1053 evolutionary perspective. *Trends in Plant Science* 11(8), 375-381. doi:
 1054 [10.1016/j.tplants.2006.06.006](https://doi.org/10.1016/j.tplants.2006.06.006)
- 1055 **Ruangrak, E., Du, Y., Htwe, N.M.P.S., Pimorat, P., Gao, J.** (2019). Identification of early
 1056 tomato fruit ripening loci by QTL-seq. *Journal of Agricultural Science*. 11,
 1057 51.doi:10.5539/jas.v11n2p51.

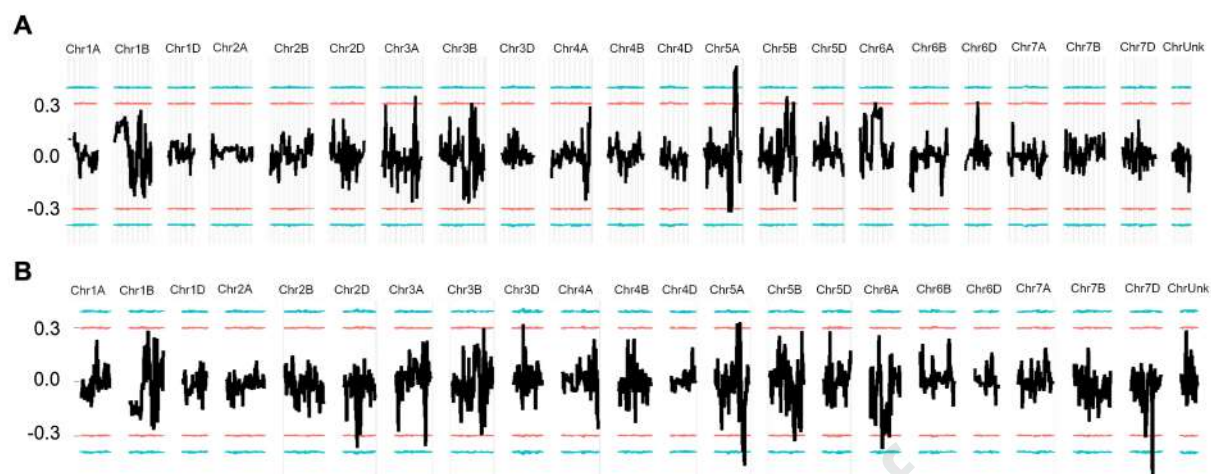
- 1058 **Sánchez-Martín, J., & Keller, B.** (2021). NLR immune receptors and diverse types of non-NLR
 1059 proteins control race-specific resistance in Triticeae. *Current Opinion in Plant*
 1060 *Biology* 62, 102053. <https://doi.org/10.1016/j.pbi.2021.102053>
- 1061 **Sannemann, W., Lisker, A., Maurer, A., et al.** (2018). Adaptive selection of founder segments
 1062 and epistatic control of plant height in the MAGIC winter wheat population WM-
 1063 800. *BMC genomics* 19, 1-16. <https://doi.org/10.1186/s12864-018-4915-3>
- 1064 **Scheben, A., Batley, J., Edwards, D.** (2017). Genotyping-by-sequencing approaches to
 1065 characterize crop genomes: choosing the right tool for the right application. *Plant*
 1066 *Biotechnology Journal* 15(2),149-161. doi: 10.1111/pbi.12645.
- 1067 **Segawa, T., Nishiyama, C., Tamiru-Oli, M., et al.** (2021). Sat-BSA: an NGS-based method
 1068 using local de novo assembly of long reads for rapid identification of genomic structural
 1069 variations associated with agronomic traits. *Breeding Science* 71(3), 299-312. doi:
 1070 10.1270/jsbbs.20148
- 1071 **Semagn, K., Iqbal, M., Chen, H. et al.** (2021). Physical Mapping of QTL in Four Spring Wheat
 1072 Populations under Conventional and Organic Management Systems. I. Earliness. *Plants*
 1073 10, 853. <https://doi.org/10.3390/plants10050853>
- 1074 **Sham P, Bader JS, Craig I, O'Donovan M, Owen M.** (2002). DNA pooling: a tool for large-
 1075 scale association studies. *Nature Reviews Genetics* 3, 862–871. doi: 10.1038/nrg930.
- 1076 **Shen, R., & Messer, P. W.** (2022). Predicting the genomic resolution of bulk segregant analysis.
 1077 *G3: Genes, Genomes, Genetics* 12(3), jkac012.
 1078 <https://doi.org/10.1093/g3journal/jkac012>
- 1079 **Shi, C., Zhao, L., Zhang, X., Lv, G., Pan, Y., & Chen, F.** (2019). Gene regulatory network and
 1080 abundant genetic variation play critical roles in heading stage of polyploidy wheat. *BMC*
 1081 *Plant Biology* 19, 1-16. <https://doi.org/10.1186/s12870-018-1591-z>
- 1082 **Shirdelmoghanloo, H., Taylor, J. D., Lohraseb, I., et al.** (2016). A QTL on the short arm of
 1083 wheat (*Triticum aestivum* L.) chromosome 3B affects the stability of grain weight in
 1084 plants exposed to a brief heat shock early in grain filling. *BMC Plant Biology* 16, 1-15.
 1085 <https://doi.org/10.1186/s12870-016-0784-6>
- 1086 **Stefanowicz, K., Lannoo, N., & Van Damme, E. J.** (2015). Plant F-box proteins–judges
 1087 between life and death. *Critical Reviews in Plant Sciences* 34(6), 523-552.
 1088 <https://doi.org/10.1080/07352689.2015.1024566>
- 1089 **Stein, O., & Granot, D.** (2019). An overview of sucrose synthases in plants. *Frontiers in Plant*
 1090 *Science* 10, 95. <https://doi.org/10.3389/fpls.2019.00095>
- 1091 **Shukla, S., Singh, K., Patil, R. V., et al.** (2015). Genomic regions associated with grain yield
 1092 under drought stress in wheat (*Triticum aestivum* L.). *Euphytica* 203(2), 449-467.
 1093 <https://doi.org/10.1007/s10681-014-1314-y>
- 1094 **Takagi, H., Uemura, A., Yaegashi, H., et al.** (2013). MutMap: Genetic mapping of agronomic
 1095 traits in rice by whole genome sequencing. *Plant Biotechnology Journal* 11(6), 747–755.
- 1096 **Trinidad, J. L., Grajo, H. L., Abucay, J. B., & Kohli, A.** (2016). Cereal root proteomics for
 1097 complementing the mechanistic understanding of plant abiotic stress
 1098 tolerance. *Agricultural proteomics volume 2: Environmental stresses* 19-51.
 1099 doi:10.1007/978-3-319-43278-6_2
- 1100 **Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen,**
 1101 **S.G.** (2012) Primer3--new capabilities and interfaces. *Nucleic Acids Research*
 1102 40(15):e115. doi: 10.1093/nar/gks596.

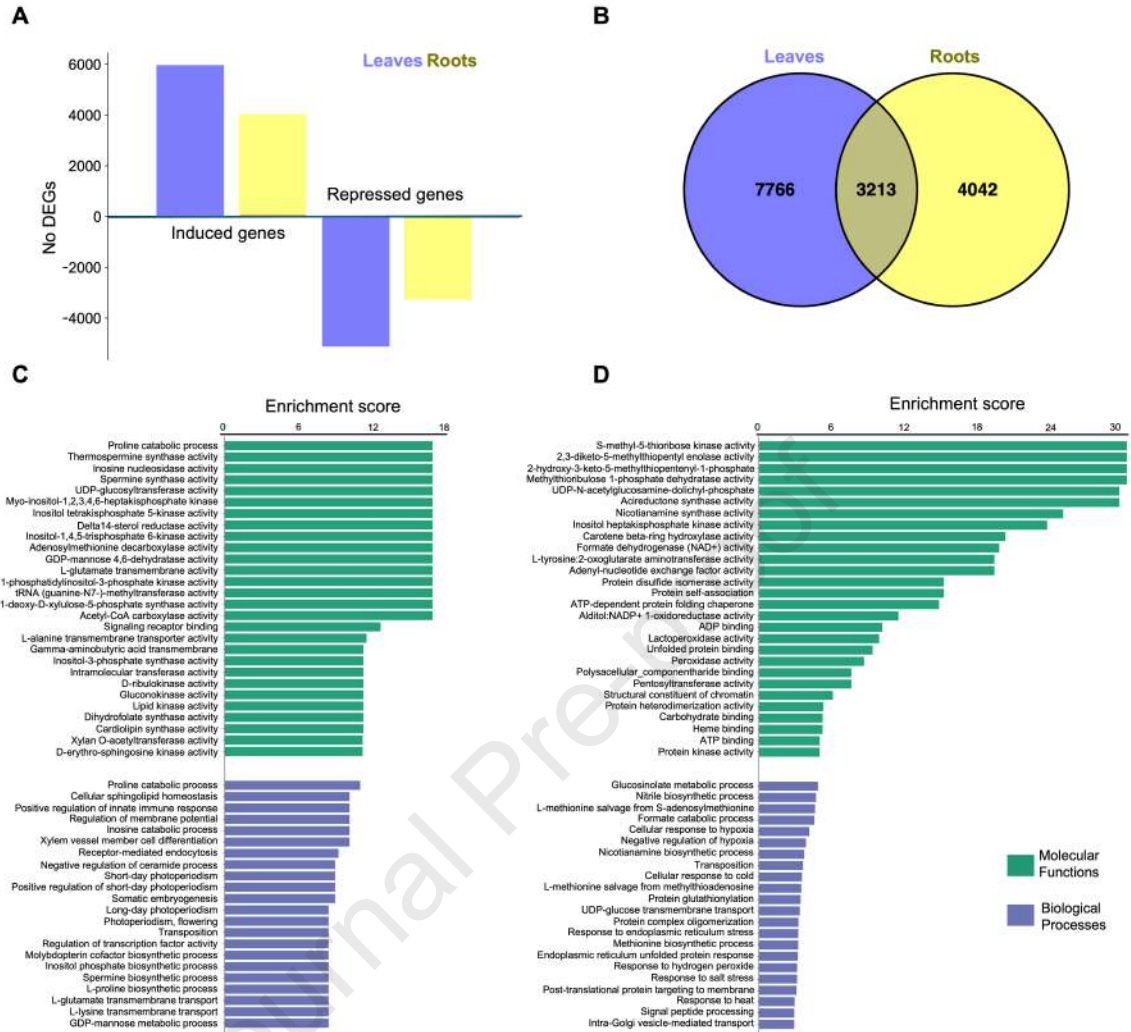
- 1103 **Vitale, P., Fania, F., Esposito, S., et al.** (2021). QTL Analysis of Five Morpho-Physiological
 1104 Traits in Bread Wheat Using Two Mapping Populations Derived from Common Parents.
 1105 *Genes* 12, 604. <https://doi.org/10.3390/genes12040604>
- 1106 **Wang, G., Gu, J., Long, D., et al.** (2024). Genome-wide identification of wheat ABC gene
 1107 family and expression in response to fungal stress treatment. *Plant Biotechnology*
 1108 *Reports* 18(3), 401-413. <https://doi.org/10.1007/s11816-023-00881-2>
- 1109 **Wang, R., Li, K., Zhang, W., et al.** (2024b). QTL-seq analysis identified the genomic regions
 1110 of plant height and days to heading in high-latitude rice. *Frontiers in*
 1111 *Genetics* 15,1305681. <https://doi.org/10.3389/fgene.2024.1305681>
- 1112 **Wang, X., Han, L., Li, J., Shang, X., Liu, Q., Li, L., & Zhang, H.** (2023). Next-generation
 1113 bulked segregant analysis for Breeding 4.0. *Cell Reports* 42(9).
 1114 <https://doi.org/10.1016/j.celrep.2023.113039>
- 1115 **Wang, Y., Xie, J., Zhang, H., et al.** (2017). Mapping stripe rust resistance gene YrZH22 in
 1116 Chinese wheat cultivar Zhoumai 22 by bulked segregant RNA-Seq (BSR-Seq) and
 1117 comparative genomics analyses. *Theoretical and Applied Genetics* 130, 2191-2201. doi:
 1118 10.1007/s00122-017-2950-0
- 1119 **Wang, C., Tang, S., Zhan, Q. et al.** (2019). Dissecting a heterotic gene through GradedPool-
 1120 Seq mapping informs a rice-improvement strategy. *Nature Communications* 10, 2982.
 1121 <https://doi.org/10.1038/s41467-019-11017-y>
- 1122
- 1123 **Yang, W., Feng, H., Zhang, X., et al.** (2020). Crop Phenomics and High-Throughput
 1124 Phenotyping: Past Decades, Current Challenges, and Future Perspectives. *Molecular*
 1125 *Plant* 13(2),187-214. doi: 10.1016/j.molp.2020.01.008.
- 1126 **Yuan, Y., Gao, M., Zhang, M., Zheng, H., Zhou, X., Guo, Y., & Li, S.** (2017). QTL mapping
 1127 for phosphorus efficiency and morphological traits at seedling and maturity stages in
 1128 wheat. *Frontiers in Plant Science* 8, 614. <https://doi.org/10.3389/fpls.2017.00614>
- 1129 **Yu, T., Cao, S., Jin, Y., Xu, C., Liu, R., Wang, B., Lv, Y., Meng, T., Ma, P.** (2024) Unveiling
 1130 resistance expression profile to powdery mildew in wheat via Bulk Segregant RNA-
 1131 Seq. *BMC Plant Biology* 11, 24(1), 1061. doi: 10.1186/s12870-024-05789-9.
- 1132 **Zhan Y, Hu W, He H, Dang X, Chen S, and Bie Z** (2023). A major QTL identification and
 1133 candidate gene analysis of watermelon fruit cracking using QTL-seq and RNA-seq.
 1134 *Frontiers Plant Science* 14:1166008. doi: 10.3389/fpls.2023.1166008
- 1135 **Zhang, B., Qi, F., Hu, G., Yang, Y., Zhang, L., Meng, J., & Xing, Y.** (2021). BSA-seq-based
 1136 identification of a major additive plant height QTL with an effect equivalent to that of
 1137 Semi-dwarf 1 in a large rice F2 population. *The Crop Journal* 9(6), 1428-1437.
 1138 <https://doi.org/10.1016/j.cj.2020.11.011>
- 1139 **Zhang, J., Hao, C., Ren, Q., Chang, X., Liu, G., Jing, R.** (2011). Association mapping of
 1140 dynamic developmental plant height in common wheat. *Planta* 234(5):891–902. doi:
 1141 10.1007/s00425-011-1434-8
- 1142 **Zhang, L., Duan, Y., Zhang, Z., et al.** (2024). OcBSA: An NGS-based bulk segregant analysis
 1143 tool for outcross populations. *Molecular Plant* 17(4), 648-657. doi:
 1144 10.1016/j.molp.2024.02.011.
- 1145 **Zhang, X., Gonzalez-Carranza, Z. H., Zhang, S., Miao, Y., Liu, C. J., & Roberts, J. A.**
 1146 (2019). F-Box proteins in plants. *Annual Review of Plant Biology* 2,307-328. doi:
 1147 [10.1002/9781119312994.apr0701](https://doi.org/10.1002/9781119312994.apr0701)

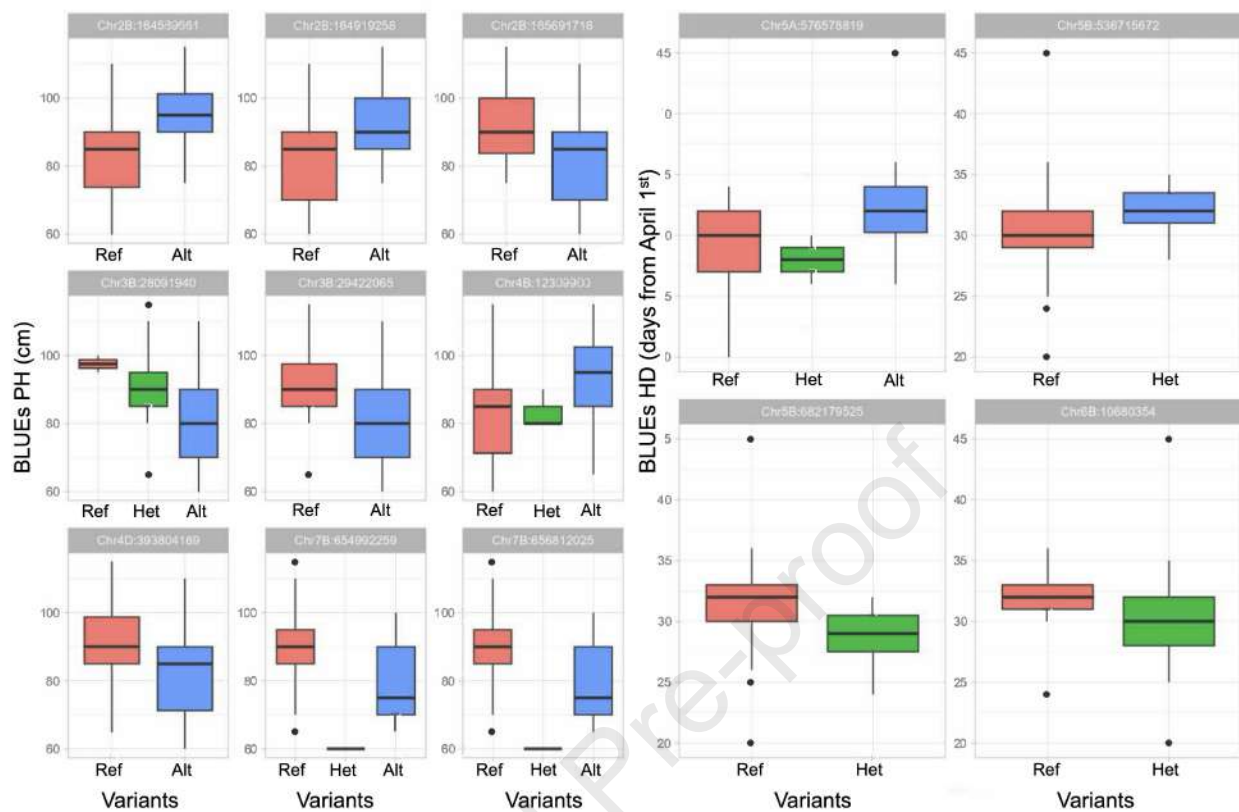
- Zhang, X., Zhang, K., Wu, J. et al.** (2020). QTL-Seq and Sequence Assembly Rapidly Mapped the Gene *BrMYBL2.1* for the Purple Trait in *Brassica rapa*. *Scientific Reports* 10,2328. <https://doi.org/10.1038/s41598-020-58916-5>
- Zhao, C., Zhang, N., Wu, Y., et al.** (2019). QTL for spike-layer uniformity and their influence on yield-related traits in wheat. *BMC Genetics* 20, 1-11. <https://doi.org/10.1186/s12863-019-0730-3>
- Zheng, Y., Ei Khine, E., Mar Thi, K., 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. *The Crop Journal* 12(2), 549-557, <https://doi.org/10.1016/j.cj.2024.01.009>.
- Zheng, Y., Khine, E. E., Thi, K. 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. *The Crop Journal* 12(2), 549-557 <https://doi.org/10.1016/j.cj.2024.01.009>
- Zhu, T., Wu, L., He, H., et al.** (2020). Bulk Segregant RNA-Seq Reveals Distinct Expression Profiling in Chinese Wheat Cultivar Jimai 23 Responding to Powdery Mildew. *Frontiers Genetics* 27, 11, 474. doi:10.3389/fgene.2020.00474.
- Zhu, T., Wang, L., Rimbart, H., et al.** (2021). Optical maps refine the bread wheat *Triticum aestivum* cv Chinese Spring genome assembly. *The Plant Journal* <https://doi.org/10.1111/tpj.15289>
- Zou, C., Wang, P., & Xu, Y.** (2016). Bulk sample analysis in genetics, genomics and crop improvement. *Plant biotechnology journal* 14(10), 1941-1955. doi:10.1111/pbi.12559
- Xin, F., Zhu, T., Wei, S., Han, Y., Zhao, Y., et al.** (2020). QTL Mapping of kernel traits and validation of a major QTL for kernel length-width ratio using SNP and bulk segregant analysis in wheat. *Scientific Reports*, 10,1–12. doi:10.1038/s41598-019-56979-7.
- Xiang, D., Quilichini, T.D., Liu, Z., et al.** (2019). The Transcriptional Landscape of Polyploid Wheats and Their Diploid Ancestors during Embryogenesis and Grain Development. *Plant Cell* 31(12), 2888-2911. doi:10.1105/tpc.19.00397.
- Qian, Z., Liu, R., Liu, X., et al.** (2024). Bulk segregant RNA-seq reveals complex resistance expression profile to powdery mildew in wild emmer wheat W762. *Frontiers Plant Science* 15, 1387427. doi:10.3389/fpls.2024.1387427
- Xu, D., Jia, C., Lyu, X., Yang, T., et al.** (2023). In silico curation of QTL-rich clusters and candidate gene identification for plant height of bread wheat. *The Crop Journal* 11(5), 1480-1490. <https://doi.org/10.1016/j.cj.2023.05.007>

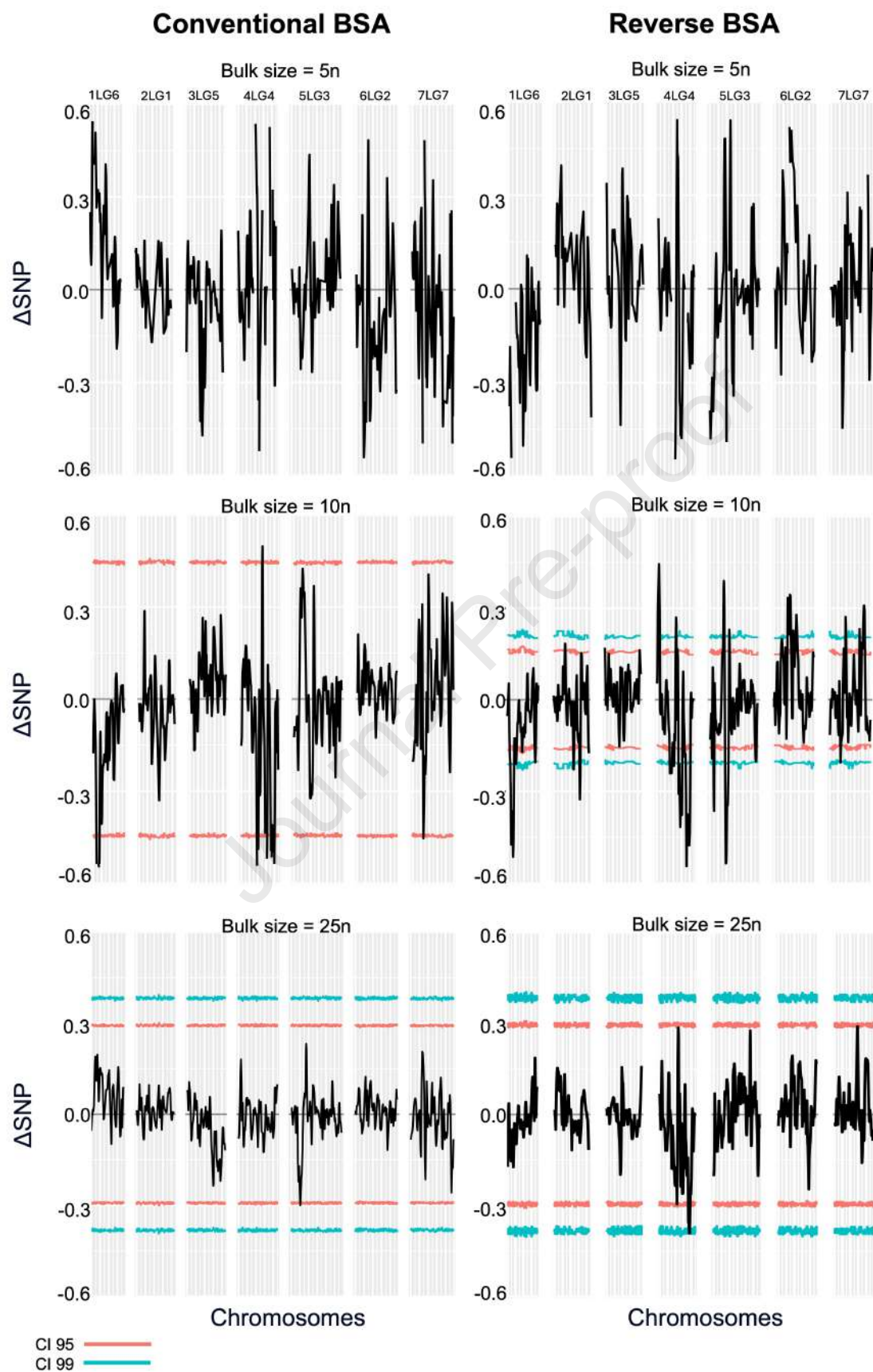
A**B**





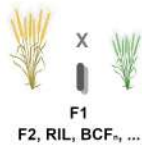




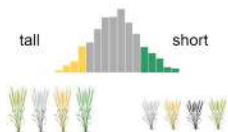


BSA-QTLseq

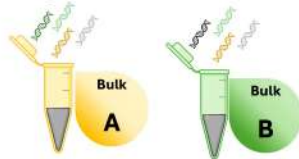
1. Parental cross



2. Phenotypic sampling



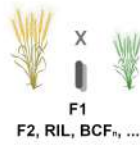
3. DNA pooling



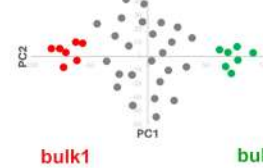
4. QTLseq mapping for a single trait (e.g., PH)

Reverse BSA-QTLseq

1. Parental cross



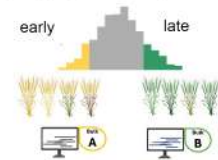
2. Molecular sampling



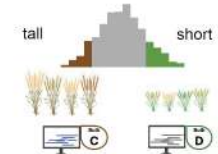
3. NGS of single plants

4. Multiple-trait phenotyping and *in silico* bulk reconstruction

4.1 QTLseq mapping for trait #1 (i.e. HD)



4.2 QTLseq mapping for trait #2 (i.e. PH)



4.3 QTLseq mapping for trait #N (i.e. XY)

