

A SNP-based genome-wide association study (GWAS) of seed yield related traits in *Psathyrostachys juncea* using wheat as a reference genome (#112251)

1

First revision

Guidance from your Editor

Please submit by **10 Apr 2025** for the benefit of the authors (and your token reward) .



Structure and Criteria

Please read the 'Structure and Criteria' page for guidance.



Custom checks

Make sure you include the custom checks shown below, in your review.



Author notes

Have you read the author notes on the [guidance page](#)?



Raw data check

Review the raw data.



Image check

Check that figures and images have not been inappropriately manipulated.

If this article is published your review will be made public. You can choose whether to sign your review. If uploading a PDF please remove any identifiable information (if you want to remain anonymous).

Files

Download and review all files from the [materials page](#).

1 Tracked changes manuscript(s)

1 Rebuttal letter(s)

9 Figure file(s)

4 Table file(s)

1 Other file(s)

! Custom checks

DNA data checks



Have you checked the authors [data deposition statement](#)?



Can you access the deposited data?



Has the data been deposited correctly?



Is the deposition information noted in the manuscript?



Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. BASIC REPORTING
2. EXPERIMENTAL DESIGN
3. VALIDITY OF THE FINDINGS
4. General comments
5. Confidential notes to the editor

 You can also annotate this PDF and upload it as part of your review

When ready [submit online](#).

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your [guidance page](#).




BASIC REPORTING

-  Clear, unambiguous, professional English language used throughout.
-  Intro & background to show context. Literature well referenced & relevant.
-  Structure conforms to [Peerj standards](#), discipline norm, or improved for clarity.
-  Figures are relevant, high quality, well labelled & described.
-  Raw data supplied (see [Peerj policy](#)).

EXPERIMENTAL DESIGN

-  Original primary research within [Scope of the journal](#).
-  Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
-  Rigorous investigation performed to a high technical & ethical standard.
-  Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

-  **Impact and novelty is not assessed.** Meaningful replication encouraged where rationale & benefit to literature is clearly stated.
-  All underlying data have been provided; they are robust, statistically sound, & controlled.
-  Conclusions are well stated, linked to original research question & limited to supporting results.



The best reviewers use these techniques

Tip

Example

Support criticisms with evidence from the text or from other sources

Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.

Give specific suggestions on how to improve the manuscript

Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).

Comment on language and grammar issues

The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult. I suggest you have a colleague who is proficient in English and familiar with the subject matter review your manuscript, or contact a professional editing service.

Organize by importance of the issues, and number your points

1. Your most important issue
2. The next most important item
3. ...
4. The least important points

Please provide constructive criticism, and avoid personal opinions

I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC

Comment on strengths (as well as weaknesses) of the manuscript

I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.

A SNP-based genome-wide association study (GWAS) of seed yield related traits in *Psathyrostachys juncea* using wheat as a reference genome

Zhen Li¹, Tian Wang², Xiaomin Ren³, Feng Han¹, Yingmei Ma⁴, Lan Yun^{Corresp. 1}

¹ College of Grassland Science, Inner Mongolia Agricultural University, Inner Mongolia, China

² Bayannur Agriculture and Animal Husbandry, Inner Mongolia, China

³ Inner Mongolia University, Inner Mongolia, China

⁴ College of Desert Control Science and Engineering, Inner Mongolia Agricultural University, Inner Mongolia, China

Corresponding Author: Lan Yun

Email address: nmg_yunlan@163.com

Background: *Psathyrostachys juncea* is an popular perennial grass for both mowing and grazing when used for pasture establishment in high latitude regions. *P. juncea* has the advantages of high yield, high quality, good palatability and high nutritional value. It is widely used in artificial grassland construction and ecological restoration in Inner Mongolia and other regions.

Methods : The current study aimed to identify genetic signals associated with seed yield in *P. juncea* germplasm through a genome-wide association study (GWAS) using wheat as a reference genome. 300 accessions of *P. juncea* germplasm from different countries were used as materials to assess eight seed yield-related traits in two environments for two consecutive years.

Results: All the trait values varied considerably across genotypes. Across different locations, the coefficient of variation among genotypes was the highest for seed yield per plant, which exceeded 70%. Further correlation analysis of seed yield factors showed that seed yield per plant had a significant positive correlation with reproductive tiller number. The population exhibited five population structures (Q) assessed by using 84,024 single-nucleotide polymorphisms (SNPs). After controlling Q and K, GWAS identified 121 SNPs significantly associated with eight traits. Among them, 19 SNPs were detected in multiple environments, and a total of 91 candidate genes were annotated, which involve the synthesis of cell wall polysaccharides and proteins, plant growth and development, photosynthesis, gibberellin regulation, hormone signal transduction, phenylalanine metabolism, and amino acid metabolism processes. The identification of SNP signals and related candidate genes could enrich the existing genomic resources and lay a foundation for the study of molecular breeding, mapping , and cloning of important genes in *P. juncea*

.

A SNP-based genome-wide association study (GWAS) of seed yield related traits in *Psathyrostachys juncea* using wheat as a reference genome

Zhen Li¹, Tian Wang², Xiaomin Ren³, Feng Han¹, Yingmei Ma⁴, Lan Yun^{1*}

¹ Key Laboratory of Grassland Resources of the Ministry of Education, College of Grassland Science, Inner Mongolia Agricultural University, Hohhot 010000, China.

² Bayannur Agriculture and Animal Husbandry Bureau, Bayannur, 015000, China

³ Inner Mongolia University, Hohhot 010000, China.

⁴ College of Desert Control Science and Engineering, Inner Mongolia Agricultural University, Hohhot, 010000, China

Correspondence

Corresponding Author: Lan Yun

E-mail: yunlan@imau.edu.cn

Abstract

Background: *Psathyrostachys juncea* is an popular perennial grass for both mowing and grazing when used for pasture establishment in high latitude regions. *P. juncea* has the advantages of high yield, high quality, good palatability and high nutritional value. It is widely used in artificial grassland construction and ecological restoration in Inner Mongolia and other regions.

Methods: The current study aimed to identify genetic signals associated with seed yield in *P. juncea* germplasm through a genome-wide association study (GWAS) using wheat as a reference genome. 300 accessions of *P. juncea* germplasm from different countries were used as materials to assess eight seed yield-related traits in two environments for two consecutive years.

Results: All the trait values varied considerably across genotypes. Across different locations, the coefficient of variation among genotypes was the highest for seed yield per plant, which exceeded 70%. Further correlation analysis of seed yield factors showed that seed yield per plant had a significant positive correlation with reproductive tiller number. The population exhibited five population structures (Q) assessed by using 84,024 single-nucleotide polymorphisms (SNPs). After controlling Q and K, GWAS identified 121 SNPs significantly associated with eight traits. Among them, 19 SNPs were detected in multiple environments, and a total of 91 candidate genes were annotated, which involve the synthesis of cell wall polysaccharides and proteins, plant growth and development, photosynthesis, gibberellin regulation, hormone signal transduction,

phenylalanine metabolism, and amino acid metabolism processes. The identification of SNP signals and related candidate genes could enrich the existing genomic resources and lay a foundation for the study of molecular breeding, mapping, and cloning of important genes in *P. juncea*.

Keywords: SNP, *Psathyrostachys juncea*, Genome-Wide Association Study, seed yield trait.

Introduction

Psathyrostachys is a small genus of Gramineae Triticeae with only about ten species. *Psathyrostachys* is originated in Eurasia, ranging from the Middle East to Russia, across central Asia to Mongolia and northern China. There are four wild ~~distribution~~ species in China, namely *P. huashanica* Keng ex P. C. Kuo, *P. lanuginosa* (Trin) Nevski, *P. juncea* (Fisch.) Nevski and *P. kronenburgii* (Hack.) Nevski. *P. huashanica* is an endemic species in China and its distribution is limited to Huashan Mountain in Shaanxi Province. Other species are mainly distributed in Tianshan Mountain and Altai Mountain of Xinjiang and Gansu Province and Qinghai-Tibet Plateau (Dewey 1984). Most of *Psathyrostachys* species are perennial cool- season grasses and well-adapted to cold uplands and semiarid deserts. After been introduced into North America *P. juncea* also known as Russian wildrye. It's a cross-pollinated, perennial bunch grass with dense clustered leaves, multiple tillers, and strong roots (Asay 2008). It is the only grass species with feeding value of *Psathyrostachys*. Due to its strong cold and drought tolerance, as well as maintains high nutritional value in summer and autumn, it is a valuable grass species for forage production and ecological restoration (Xiong et al. 2020). It is mainly used in arid and semi-arid areas of North America for grassland reseeding and dry pasture establishment, as well as for salinized and alkaline grassland improvement. There are multiple strong resistance genes of *P. juncea* that have been used for wheat improvement, including resistance to powdery mildew, sharp eyespot and yellow stunt virus (Han et al. 2020; Bai et al. 2020; Hu et al. 2018). Therefore, *P. juncea* is an excellent perennial grass with feeding, ecological and breeding value.

Compared with crops, most perennial forage grasses have strong vegetative reproduction ability, while their sexual reproduction ability is relatively weak (Garcia et al. 2019; Easson et al. 1993). The short blooming period and cross-pollination ability of *P. juncea* lead to erratic and relatively low seed yields (Foulkes et al. 2007). The breeders have to focused on the development of cultivars with higher seed yield and yield stability based on the reproductive characteristics of perennial cross-pollinated plants. Seed yield is controlled by numerous genes that interact with each other and with the environment. Seed yield is a complex trait that is determined by spike number, seed weight per spike, spike length, spike node number and thousand kernel weight, and each component trait is a quantitative trait controlled or affected by multiple loci (Shokat et al. 2020; Gao et al. 2017). These traits have a great influence on improving plant seed yield. Thus, there needs to be a detailed genetic dissection of the seed yield trait of cross-pollinated perennial plants and its component traits to manipulate the alleles at the relevant loci to the greatest advantage.

With the development of high-throughput sequencing technology, more and more plant genome sequences have been published, which is conducive to obtaining high density molecular markers covering the whole genome, facilitating genome-wide association analysis, and enabling the discovery of many agronomic traits related loci. As a critical breeding focus, the research on molecular mechanism of seed yield traits has application value in varia-type of plants (Nayak et al. 2022). As an effective tool for dissecting the genetic architecture of complex quantitative traits, genome-wide association studies provide a high-resolution approach for the identification of QTLs (Sukumaran et al. 2015) and have been widely used for QTL detection of different agronomic traits in crops under different environmental conditions including seed yield related traits (Zhang et al. 2020b; Xu et al. 2018; Xiao et al. 2016), leaf architecture (Tian et al. 2011), stalk lodging-related traits (Zhang et al. 2018), and seedling root development (Pace et al. 2015). Currently, GWAS analysis of yield-related traits in plants has been reported. Zhou et al. (2020) conducted GWAS on yield traits of 93 *Bromus inermis* Leyss, and a total of 95,708 effective SNPs were identified, and further analysis identified 247 core SNPs related to seed yield. Saba Akram et al. (2021) conducted GWAS on yield traits of bread wheat and found that, chromosomes 1B and 2A carried loci linked to yield in two different seasons, and an increase of up to 8.20% is possible in yield by positive allele mining. These studies give different perspectives for other gramineous plants, and provide references for genetic research of gramineous plants that have no completed whole genome sequencing. In recent years, many scholars have conducted GWAS analysis on yield traits of Triticeae crops, but there is little known about natural genetic variation of forage plant seed yield. Only by understanding the corresponding molecular mechanism can natural variation be effectively manipulated and applied to breeding work.

Therefore, in this study, 300 germplasms from different regions were collected for the main breeding target traits of *P. juncea*, and a large number of SNPs covering the whole genome of *P. juncea* were developed by using SLAF-seq (Specific-Locus Amplified Fragment Sequencing) simplified genome sequencing technology. Through the two years and two locations phenotypic identification of eight seed yield related traits, the interannual growth dynamics of perennial grasses were studied, and the effects of genotypic and environmental interaction on seed yield traits of *P. juncea* were analyzed. The genetic basis of these eight traits was analyzed by GWAS and significant allele variants and candidate genes of target traits were explored, which laid a foundation for the research on molecular breeding of *P. juncea*, localization and cloning of important genes.

Materials & Methods

Plant materials and experimental design

The germplasm of *P. juncea* used in this study was collected from the China National Medium-term Gene Bank for Forage Germplasm and the United States National Plant Germplasm Resources Conservation System (NPGS). A total of 300 genotypes comprised of 45 China, 41 U.S., 43 Mongolia, 67 Russian, 15 Canada, 15 Estonia, 43 Kazakhstan and 31 Former

Soviet Union origin were used in the GWAS analyses. The germplasm information including sample ID, sample number, origin and cultivation are shown in Additional file Table S1. The *P. juncea* accessions were grown and measured under four environments (2 locations×2 consecutive growth years), i.e., Hohhot (40°48N, 111°41E) and Baotou (40°39N, 109°49E) of Inner Mongolia China in 2021 and 2022 (Hohhot and Baotou referred to as E1 and E2 in 2021, Hohhot and Baotou referred to as E3 and E4 in 2022). A randomized block design was adopted for *P. juncea* planting in the field, with a row spacing of 60cm and a plant spacing of 50cm. Field management was consistent with local field production management. During the growth period, conventional water and fertilizer management measures were adopted, and fertilization was carried out at the tillering stage every year, with the amount of fertilization of 150 kg/hm² 18-18-18 (N-P₂O₅-K₂O). Watering after fertilization, weeding and other field management during the growth period. Soil samples were taken from two locations and repeated three times at each location to determine soil pH, soil organic carbon (g·kg⁻¹), total nitrogen (g·kg⁻¹), soil available phosphorus (mg·kg⁻¹) and rapidly available potassium (mg·kg⁻¹). Environmental and soil conditions for each location are listed (Additional file, Table S2).

Seed yield related traits measurements

The spikes of each line were harvested after full maturity. Five consistent-growth spikes were selected for seed yield related traits measurements after thorough drying in each replication. Seed yield traits including reproductive tiller length (RTL, cm), reproductive tiller number (RTN), spike length (SL, cm), spike width (SW, cm), spike node number (SNN), thousand kernel weight (TKW, g), seed weight per spike (SWS, g) and seed yield per plant (SY, g) were measured. All of the traits were represented by the mean values of the selected five spikes except RTL, RTN and SY. TKW was the average weight of three repeated measures of 1000 randomly selected kernels from the bulked kernels of each line.

For a single environment, the mean value from five replications of each line was calculated as phenotypic data for descriptive statistics. Origin 2019b (MicroCal, USA) was used for analysis of variance, and the significance level was 0.05. SAS 9.4 (USA) was used to analyze the mean, standard deviation (SD), minimum, maximum, coefficient of variance (CV), skewness and kurtosis. Pearson correlation coefficients (*r*) for each pair of traits was calculated using the “Hmisc” and “corrplot” packages in R software. The BLUP (Best linear unbiased prediction) value of eight traits under different environments was calculated and optimized by IciMapping V4.0 (<http://www.isbreeding.net>) (Li et al. 2007). R software package "lme4" was used to calculate the generalized heritability (*h*²) of seed yield traits in multiple environments. The *h*² was calculated as follows: $h^2 = V_g / (V_g + V_e)$, where *V_g* represents genotype variance and *V_e* represents environment variance (Smith et al. 1998).

DNA extraction and SNP tag analysis

A total of 300 individual plants were selected from *P. juncea* materials. DNA was extracted from the *P. juncea* leaves of 2-3 young two-week-old seedlings using Plant Genomic DNA Kit

(TIANGEN Biotech Co., Ltd, Beijing, China) following the manufacturer's instructions. The NanoDrop™2000 spectrophotometer (Thermo Fisher, Waltham, MA, USA) was used to evaluate the quantity and quality of DNA on 1% agarose gel.

The *RsaI* enzyme was selected to digest the genomic DNA of each sample. The enzymatic fragments (SLAF tags) were end-repaired, ploy A tails were added, sequencing connectors were added, purification was performed, PCR amplification was performed, and the target fragments were selected by mixing and gel cutting to construct libraries. Then, the libraries were tested for insert size and sequenced on the Illumina Hi Seq TM2500 sequencing platform after passing the library quality check. The raw data obtained from sequencing were identified using Dual-index, and after filtering the junctions of sequenced reads, sequencing quality and data volume were evaluated (Li and Durbin 2010). The SLAF-seq sequencing technology was used to simplify the genome sequencing for all materials of *P. juncea*. The sequence type with the deepest sequence in each SLAF tag was used as the reference sequence. The sequences obtained from sequencing were compared to the reference genome using BWA software (Li et al. 2009). Bread wheat (*Triticum aestivum*, 17 Gb) was selected as the reference genome for SNP localization (<http://dx.doi.org/10.1126/science.1251788>). SNP markers were developed using GATK and SAMtools methods (Takeda and Matsuoka 2008; Heffner et al. 2009). The SNP marker intersections obtained by the methods were used as the final reliable SNP marker dataset, and then, after screening based on $MAF > 0.05$ and integrity > 0.85 , the screened high-quality SNP loci were used for population polymorphism analysis, genetic evolution analysis, and genome-wide association analysis.

Linkage disequilibrium (LD), population genetic evolution, and kinship analysis

Genetic evolution can reveal genetic evolutionary relationships and environmental adaptation mechanisms among different populations. The genetic evolution was analyzed by SNPs filtering. The “CM plot package” in R was used for SNP marker density mapping. The polymorphic information of SNP was evaluated using the Power-Marker V 3.25 software, the genetic diversity was calculated. In this study, Gemma software (Zhou and Stephens 2012) was used to assess the genetic relationships of populations, and then principal component analysis (PCA) was performed by Genome-wide Complex Trait (GCTA) software (Yang et al. 2011). After that, the population structure of the *P. juncea* natural population was analyzed by cross-validation method using Admixture 1.23 software (Montana and Hoggart 2007), and a variety of materials in the natural population were clustered into corresponding subgroups. For the research population, the range of subgroups (K value) was set from 1 to 10, and K for iterative operations was calculated from two. The number of runs and repetitions was set to 10,000, and the number of optimal clusters was determined according to the K value with the lowest cross-validation error rate. The PopLDdecay software (Zhang et al. 2019) was used to calculate the LD of the population. The degree of linkage disequilibrium (r^2) between two SNPs in the 200kb segment of the genome was calculated. The r^2 of all the significant loci combinations was selected according to the selection criteria $r^2 > 0.1$ for the LD decay plot.

GWAS and haplotype analysis to screen candidate genes

Using the mixed linear model (MLM) approach of TASSEL V5.0 software (Bradbury et al. 2007), BLUPs for eight seed yield-related trait phenotypic data were analyzed for GWAS using population structure and kinship as covariates, and the selection thresholds for significance $p \leq 1.27 \times 10^{-6}$ ($-\log_{10}(p) = 5.89$) and $p \leq 1.27 \times 10^{-7}$ ($-\log_{10}(p) = 6.89$) (Li et al. 2018b). Manhattan plots were plotted using the R package "CMplot". Significant differences in phenotypic traits corresponding to alleles of significantly associated markers were tested in the R program using t-tests. The screened significant markers were physical map by MapChart software to determine the physical location of the genes on the chromosomes, and the Haploview V4.2 software (Barrett 2009) was used for haplotype analysis of loci closely associated with the target traits. Finally, the candidate genes that might be associated with the target traits were analyzed based on the reference information of typical.

Results

Phenotypic analysis of eight seed yield related traits

Statistical analysis of seed yield revealed that there were significant phenotypic variations among 300 *P. juncea* germplasm in different environments and years. From 2021 to 2022, the coefficient of variation of SY was the largest, with an average of 72.41%; and the coefficient of variation of RTL was the smallest, at 14.27%. By observing the data for two consecutive years, it was found that except for SL, SW and SNN, the coefficient of variation of the other five traits in 2022 was higher than that in 2021, which indicated that these traits were greatly affected by growth years (Table 1). Further analysis of variance on seed yield showed that all traits except RTN in 2022 were significantly differences between the two locations in the same year. SL, SNN, TKW and SWS were significantly different between different locations and years (Fig. 1). The joint ANOVA analysis of seed yield-related traits of *P. juncea* at the two experimental sites in two years showed that significant environment and genotype effects, and significant year by environment interactions were noted in RTN, SW, SNN, TKW, SWS and SY traits ($P < 0.001$). We calculated h^2 for 300 genotypes at both locations, ranging from 0.57 (SY) to 0.81 (TKW), with an average of 0.68 (Table 2). Because of the high heritability in each environment, the least squares mean of individual traits can be calculated and used to analyze the association between markers and traits.

The correlation coefficients between eight seed yield related traits were analyzed. These eight indicators were normally distributed or skewed. Pearson's correlation coefficients indicated that there were five agronomic traits with extremely significant correlation with SY in Hohhot ($P < 0.001$), seven agronomic traits with extremely significant correlation with SY in Baotou, and the correlation coefficient of RTN was the largest. SNN was correlated with the six agronomic traits, among which SL had the strongest correlation (Fig. 2). This suggests that the traits are genetically linked or that the traits are affected by genes with pleiotropic effects. It also indicates that there is a significant interaction between seed-related traits, and the improvement of RTN has a significant effect on seed yield.

Development of SNP markers, population structure and linkage disequilibrium analysis of *P. juncea*.

The SLAF-seq technology was used to simplify the genome sequencing of 300 *P. juncea* germplasm, and ~~across~~ genome-wide SNP markers of *P. juncea* were developed and obtained. Using the 1,176,240 SLAF tags obtained, a total of 84,024 SNPs' accurate positioning on the bread wheat genome was established based on the criteria integrity > 0.85 and MAF > 0.05, including 19,913 on subgenome A, 24,760 on subgenome B, and 39,351 on subgenome D, accounting for 23.7%, 29.5% and 46.8% of the total number of SNPs, respectively. The chromosomal level distribution of the 84,024 SNPs on the three subgenomes showed the maximum SNPs in Chr2D (6717), while the minimum SNPs in Chr6A (2141). In general, SNP markers are evenly distributed on the genome of *P. juncea*, which provides an important guarantee for genetic diversity analysis and genome-wide association analysis of *P. juncea*. We calculated the genetic diversity of each chromosome based on the 84,024 SNP markers obtained by screening. The average genetic diversity of the three subgenomes were 0.3081, 0.2991 and 0.3120, respectively. The above results showed that the genetic diversity of *P. juncea* germplasm was relatively high (Table 3).

Population structure indicates the degree of genetic variation between individuals or populations. Population structure was analyzed using 84,024 SNPs, and the combination trends of the likelihood values of LnP(D) and ΔK calculated for each K showed that the 300 genotypes could be assigned into five subgroups. Based on the Q value, it can be found that five subgroups were clustered according to geographical sources, including 44 (Former Soviet Union and Estonia), 72 (Russia), 84 (China and Mongolia), 41 (Kazakhstan) and 59 (U.S. and Canada) lines, respectively (Fig. 3a, 3d). Three hundred *P. juncea* genotypes were evaluated using PCA, and PC scatter plots showed that the first and second PC were composed of subgroups of different geographical sources (Fig. 3b). The seed yield traits of the five subgroups were analyzed. Group 1 had the highest values for all traits, while Group 4 had the lowest values, and the seed yield traits of the other three groups were not significantly different (Additional file, Table S3).

Chromosome wise LD plots were generated using 84,024 selected SNP markers (Fig. 3c). The LD (r^2) across 21 wheat chromosomes ranged from 0.131 (Chr5D) to 0.526 (Chr5A). There were certain differences in LD estimates at the subgenome level, 0.445 for subgenome A, 0.363 for subgenome B, and 0.194 for subgenome D. The complete statistics of the 84,024 markers, including SNP numbers, chromosome LD and markers with perfect LD are presented in Table 3. The results indicated that there are different levels of LD on different chromosomes of each subgenome. We found that LD decays rate is different in different subgroups, with the fastest decay rate is Former Soviet Union and Estonia, and the slowest is Kazakhstan.

GWAS, pleiotropism SNPs screening and favorable allele variation

We further performed GWAS on the differences in eight seed yield-related traits to identify significantly associated genes. A total of 121 SNPs significantly associated with the target traits were detected on all 21 chromosomes of wheat using 84,024 SNPs associated with eight seed

yield phenotype traits and BLUP values in four environments, respectively. And explained (R^2) between 4.15% and 15.64% of the variation. Notably, 7, 24, 48, 7, 6, 9, 14 and 6 significantly associated SNPs were obtained for RTL, RTN, SL, SW, SNN, TKW, SWS and SY, respectively (Fig. 4). The maximum number of SNPs was found on subgenome D (47 SNPs) followed by subgenomes A (39 SNPs) and B (35 SNPs). Chromosome wise, the highest number of SNPs was detected on Chr2D (12 SNPs), followed by Chr5A, 5D and 7D (9 SNPs). 19 SNPs were found to be associated with the target trait in four different environments, with the lowest P -value (Additional file, Table S4).

We observed that fifteen SNPs were associated with more than one seed yield related traits. Among them, Chr7A_675095747 was related to four traits of RTN, SL, SNN, and TKW, which was an important locus for regulating seed yield. At the same time, this locus was detected in multiple environments, and the P -value showed extremely significant. SL, RTN and SY were associated with 6, 5 and 5 pleiotropism loci respectively. Besides multi-trait SNPs, some genomic regions contained multi-QTLs for more than one trait. The genomic regions harboring multi-SNPs for more than one trait included Chr2A (72.6-88.7 Mb), which had three SNPs associated with seed yield related traits; Chr2D (111.6-142.4Mb, 637.7-650.9Mb), with eight SNPs, four associated with SL; Chr7D (6.8-32.6Mb, 110.6-146.4Mb) had four SNPs associated with RTN, SL, SNN and SY (Fig. 5).

In order to study the phenotype effects of significantly associated SNP locus allele variations, eight pleiotropism loci that could be detected in four different environments were selected, and the phenotype traits corresponding to genotype were analyzed. The allele variation of each SNP locus was Chr2A_72607643 (G/A), Chr2D_640215049 (G/T), Chr2D_650894588 (G/A), Chr4A_722894581 (C/T), Chr5A_657544633 (C/G), Chr5D_405156269(C/T), Chr6A_17541092 (G/A) and Chr7A_675095747 (G/T). Further analysis showed that Chr2A_72607643-AA, Chr2D_640215049-TT, Chr2D_650894588-AA, Chr4A_722894581-CC, Chr5A_657544633-GG Chr5D_405156269-TT, Chr6A_17541092-AG and Chr7A_675095747-GG were favorable allele variations for seed yield-related traits of *P. juncea* (Additional file, Figure S1). To further determine whether the favorable allele variation significant associated with seed yield-related traits have a polymerization effect, we compared and analyzed phenotypic data of *P. juncea* with different numbers of favorable allele variation loci. The results showed that among the eight significantly related SNPs, each germplasm contained up to five favorable allele variation and at least one favorable allele variation. Compared with materials with less favorable allele variation loci, the plant with more favorable allele variation loci showed excellent traits. For example, the average SY value of *P. juncea* containing one favorable allele variation is 14.91g, while the average SY value of *P. juncea* containing five favorable allele variation is 19.47g. These results indicate that favorable allele variation loci have a significant polymerization effect on eight seed yield traits (Additional file, Figure S2).

Candidate gene annotation and haplotype analysis

In addition, 19 significant SNPs with the smallest P -values for each seed yield trait and detected in four different environments were selected, and the predicted candidate genes were

identified in the range of 100 kb upstream and downstream of the SNP. A total of 91 candidate genes were obtained from 19 SNPs within the candidate regions (Table 4). Two candidate genes were detected for RTL, including *TraesCS2A02G122600.1* encoding a cellulose synthase-like protein (*CSLD*) on chromosome 2A within 83.85 kb of an SNP, *TraesCS2A02G122700.1* encoding a momilactone A synthase. A total of three SNP significantly associated with RTN were detected and 13 candidate genes were annotated. *TraesCS2D02G577000.1* encoding a serine/threonine-protein kinase (*STK*) on chromosome 2D within 82.43 kb of target SNP. Four candidate genes, *TraesCS4A02G457400.1*, *TraesCS4A02G457800.1*, *TraesCS4A02G457500.1* and *TraesCS4A02G457600.3* on chromosome 4A, which were about 54.79 kb, 10.49 kb, 54.79kb and 7.2kb away from the target SNP, respectively. *TraesCS7A02G483300.1* encoding an F-box protein within 72.13 kb of target SNP. A total of six SNP significantly associated with SL were detected and 30 candidate genes were annotated. *TraesCS2D02G599600.1* encoding mevalonate kinase (*MK*), *TraesCS2D02G599800.1* encoding LRR receptor-like *STK*, *TraesCS2D02G599900.1* encoding CBS domain-containing protein (*CDCP*), *TraesCS2D02G600000.2* encoding callose synthase, *TraesCS2D02G600100.1* encoding stem-specific protein (*TSJT1*) on chromosome 2D. These genes were located approximately 30.87-, 11.06-, 32.99-, 37.26- and 61.02- from the target SNPs, respectively. Four candidate genes, *TraesCS5A02G486900.1*, *TraesCS5A02G487000.1*, *TraesCS5A02G487300.1* and *TraesCS5A02G487500.1* on chromosome 5A, encoding photosystem II reaction center protein, *STK*, cytochrome P450 and F-box protein, respectively. Three candidate genes, on chromosome 5D, annotation information is RNA polymerase II and putative pentatricopeptide repeat-containing protein (*PPR*). Only one of the four candidate genes of Chr6A_17541092 was annotated to transcription repressor OFP14. One candidate gene *TraesCS7B02G250800.1* encoding a GDSL esterase/lipase on chromosome 7B within 1.6 kb of an SNP. A total of 16 candidate genes related to SW were annotated, among which, three genes encoding membrane protein of ER body-like protein (*MEBL*), three genes encoding *STK*. One candidate gene, *TraesCS7A02G483300.1* on chromosome 7A was detected for SNN and TKW, encoding an F-box protein at 72.13 kb from the target SNP. In addition, three candidate genes of Chr2D_640215049, which were significantly associated with TKW, were detected. Including *TraesCS2D02G576800.1* encoding a DDB1- and CUL4-associated factor homolog (*DCAF*) and *TraesCS2D02G576900.1* encoding uncharacterized acetyltransferase. A total of four SNP were significantly associated with SWS and SY, and 17 candidate genes were annotated. These four SNPs were all pleiotropism loci, and the significant associated SNPs of SL and RTL are the same loci with the same annotation information.

To further study the role of candidate genes, we conducted haplotype analyses on the 300 *P. juncea* genotypes to identify the elite haplotypes. We analyzed the four most significant SNPs among the nineteen SNPs that were consistently present in four environments. Chr2D_640215049, which was significantly associated with RTN and TKW, formed a 0 kb linkage region with the surrounding six SNPs (Fig. 6a). A total of five haplotypes were found in this region. RTN and TKW containing Hap.4 were higher than other haplotypes, indicating that Hap.4 was an excellent haplotype (Fig. 6b, 6c). Chr2D_650894588, which was significantly

345 associated with SL and SY, formed a 46 kb linkage region with the surrounding two SNPs (Fig.
 346 7a). This linkage region contains five haplotypes, and it was found that SL and SY containing
 347 Hap.4 material were higher than other haplotype materials (Fig. 7b, 7c). Similarly, the
 348 Chr6A_17541092 significant association with SL and SY was detected to form a 0 kb linkage
 349 region with four SNPs (Fig. 8a). This linkage region contains four haplotypes, and it was found
 350 that SL and SY containing Hap.3 material were higher than other haplotype materials (Fig. 8b,
 351 8c). The Chr7A_675095747 significant association with RTN, SL, SNN and TKW was detected
 352 to form a 0 kb linkage region with two SNPs (Fig. 9a). This linkage region contains five
 353 haplotypes, and it was found that the average value of each trait of containing the Hap.4 was
 354 higher than other haplotype. The haplotype analysis suggests that Chr7A_675095747^{Hap. 4} is
 355 strongly associated with seed yield in *P. juncea* (Fig. 9b, 9c).

Discussion

Trait variations and correlations in *P. juncea*

P. juncea is one of the most important perennial grasses in the high latitude areas of the Northern Hemisphere, with economic, feeding and breeding research value. Breeding *P. juncea* germplasm that can adapt to different regions is particularly important for efficient utilization and production (Cuevas and Prom 2020). From previous research results, we selected eight quantitative traits related to seed yield for measurement and analysis in two environments for two consecutive years. The results showed that from 2021 to 2022, the coefficient of variation of SY was the largest, ranging from 71.21% to 74.09%. A similar result was also found by Zhang et al. (2018) in their study on seed yield traits of *P. juncea* germplasm. This indicates that there was extensive phenotypic variation and rich genetic diversity in *P. juncea*. The variance analysis of seed yield traits of two locations showed that the seed traits indicators in Baotou were higher than in Hohhot over the years. This could be attributed to differences in the climate conditions between the two locations. The annual effective sunshine hours, daily mean temperature and relative humidity in Baotou were significantly higher than in Hohhot. Studies have shown that these factors have a positive effect on the seed yield per plant, and the relative humidity of the air has a promoting effect on plant growth (Liang et al., 2023). The relatively high generalized heritability of all traits showed that seed traits of *P. juncea* are high heritability traits, and the natural population has rich genetic diversity in them. Although affected by both genotype and environmental factor, seed traits are mainly controlled by genotype due to high heritability.

Correlation analysis further revealed ~~the correlation between traits, and~~ the majority of seed yield-related traits in the associated population were significantly positively correlated. SY had a significant correlation with multiple traits, and a highly significant positive correlation with RTN. Zhang et al. (2018) found that although the correlation between seed setting rate and seed yield per plant and its direct effects were not significant, the seed setting rate mainly played a role through the indirect effect on the number of reproductive branches and the number of spikelets per panicle. and this result was consistent with the current study. In this research, SW was significantly positively correlated with TKW, and similar finding was reported in *Secale cereale* subsp. *Segetale* (Mu et al. 2019). Therefore, we speculated that the increase of SW would increase kernels per spikelet and TKW. Spike is the primary condition for seed yield formation (Bian et al. 2020). In this research, we also found a large number of traits significantly related to SL and SW, which indirectly indicated that these traits co-regulate seed yield. At the same time, there were some differences in the adaptability to the environment, which produced genetic variations and showed different traits (Xie et al. 2015). The above results showed that there were significant genetic variations in the seed yield traits of *P. juncea* among different materials, and the seed traits data of *P. juncea* from multi-years and locations could provide reliable phenotypic data for GWAS.

Genetic diversity, population structure and LD decay of *P. juncea*

Genetic diversity is an important basis for plant improvement. A large number of SNPs were obtained by SLAF-seq sequencing, and genetic diversity of 300 *P. juncea* germplasm was analyzed at the whole wheat genome level. The average genetic diversity of the natural population was 0.3065, with the highest genetic diversity in the D subgenome. Li et al. (2022) found that the genetic diversity of 21 *P. juncea* materials based on SSR molecular markers, ranged from 0.0912 to 0.4142, with an average of 0.3728. The genetic diversity of SSR markers was higher than that of SNP markers, mostly attribute to the fact that SSR are multiallelic markers, and SNP are biallelic markers. In general, the level of genetic diversity estimated by a limited number of SNP markers is lower than that estimated by SSR markers (Van Inghelandt et al. 2010). The following views are generally accepted that in the analysis of genetic diversity, the number of markers provides much more genetic information than the number of alleles. Therefore, a large number of SNP markers obtained based on high-throughput sequencing technology can accurately reflect the genetic diversity of the population (Lu et al. 2009). Overall, the relatively high genetic diversity of *P. juncea* population mostly attribute to the strict cross-pollination features.

In GWAS, The accuracy of GWAS is influenced by population structure, kinship, and LD decay distance. Thus, the MLM method was used to take these factors into account for GWAS analysis (Zhang et al. 2010). In this research, through the analysis of population genetic structure of *P. juncea* individual lines from different countries and regions, it was found that the population structure of 300 *P. juncea* individual lines had a certain correlation with geographical distribution, which was also verified through PCA. This may be attributed to the genetic background of the genotype diversity in this studied *P. juncea* materials, indicating that SNP markers can effectively group genotypes according to gene composition. Through different statistical methods, it was found that although both methods clustered 300 *P. juncea* lines into five groups, there were some individual lines was clustered differently. The main reason for this result may be due to breeding domestication, which has a significant impact on the diversity structure. In addition, different environments can also cause genetic changes, which can affect the division of population structure.

LD determines the marker density required for GWAS analysis, and is one of the important indicators to measure whether there is correlation between molecular markers. In general, mutation and recombination will affect the change of LD size. So far, LD analysis of perennial forage grass population are rarely reported. Li et al. (2023) used SSR markers to estimate a wider distribution of LD, which is difficult to estimate accurately. In this study, a large number of SNP markers were used throughout the wheat genome. The LD level of the *P. juncea* population is relatively higher than annul crop. As cross-pollinated plant, individuals in the population are more heterozygous and leading to the increase in LD (Albrechtsen et al. 1966). LD decay rate is different in subgroups, with the fastest decay rate in the population from the Former Soviet Union and Estonia, indicating the high genetic diversity of the *P. juncea* population in this region. The above analysis indicates that the sample population in this research has high genetic diversity, suitable population structure and LD decay level, indicating that the population is

434 suitable for GWAS.

435 Significant SNPs and potential candidate genes

436 Seed yield is a super complex quantitative trait controlled by multiple genes and susceptible
437 to environmental influences. It is the underlying reason of multiple seed yield-related traits be
438 analyzed in the research. So far, there are few genome information of *P. juncea* on seed yield-
439 related trait. Therefore, we used GWAS on the seed yield-related traits of *P. juncea*, and
440 comprehensively compared with the results of previous studies on wheat genome association
441 analysis to further determine the accuracy of SNP loci significantly associated with seed yield
442 traits in the previous analysis.

443 Genome-wide association analysis showed that a total of 121 SNPs were significantly
444 associated with eight seed yield-related traits. Among them, seven loci were significantly
445 associated with RTL. A SNP (Chr2A_72607643) that regulates RTL was found at the position of
446 72.6 Mb on the short arm of Chr2A in wheat, which is consistent with the reported by Kuang
447 (Kuang et al. 2020), confirming that the gene region was involved in the regulation of RTL and
448 two candidate genes were annotated. The annotation information of *TraesCS2A02G122600.1* is
449 *CSLD*. Previous studies have shown that *OsCSLD4* plays an important role in cell wall
450 polysaccharide synthesis, participates in the regulation of growth and differentiation of rice cells,
451 and has an important effect on rice plant height and leaf type (Yoshikawa et al. 2013). *OsCSLD4*
452 also plays an important role in rice grain development. Overexpression of *OsCSLD4* promotes
453 rice seed growth and significantly widens the grains, indicating that *OsCSLD4* plays an
454 important role in synergistically promoting rice growth and increasing seed yield (Zhao et al.
455 2022).

456 Thousand grain weight, grain number per spike and number of reproductive branches
457 significantly contribute to seed yield of *P. juncea*. In this study, 24 SNPs were found to be
458 significantly associated with RTN, distributed on 14 chromosomes, and a single locus could
459 explain 4.33%-12.44% of the phenotypic variation. Cui et al. (2014) reported a QTL (*QSnpp-*
460 *4A.3*) located on chromosome 4A controlling RTN in wheat, was detected as the locus
461 Chr4A_722894581 in this study, and the locus was detected in all environments. the locus could
462 annotated to seven candidate genes, *TraesCS4A02G457400.1* and *TraesCS4A02G457800.1*
463 annotation information was 30S ribosomal protein (*RPS*). In recent years, a large number of
464 research reports have reported that the function of *RPS* in plants is mainly manifested in the
465 change of phenotype of mutant plants. The abnormal development of female gametophyte in
466 *Arabidopsis thaliana* *RPL24B* mutant, leading to seed abortion (Nishimura et al. 2005); The
467 *RPL23aA* mutation causes slow root growth and lateral root deformity in *Arabidopsis*
468 (Degenhardt and Bonham-Smith 2008). Therefore, we speculate that the gene will regulate the
469 reproductive development of plants, thereby affecting the number of reproductive branches of *P.*
470 *juncea*. The SNP located at the 595.3Mb on chromosome 5A is detected as the SNP
471 (S5A_594322620) significantly associated with RTN reported by Tian (2023). Cao et al. (2020)
472 located the QTL of RTN to the position of 675 Mb on chromosome 7A, which aligned to
473 Chr7A_675095747 in this study, and annotated to five candidate genes. *TraesCS7A02G483300.1*

gene encoded F-box protein. F-box protein is reported to be involved in signal transduction of plant hormones (ethylene, auxin, gibberellin, jasmonic acid), flower organ development and other biological processes (Xu et al. 2021), the haplotype of Chr7A_675095747 was analyzed and Hap.4 was found the elite haplotype. The elite haplotypes can be selected in molecular assisted selection in the breeding of *P. juncea*. Although the physical location of Chr2D_640215049 is different from the reported QTLs or genes related to RTN, it is detected in multiple environments, and the RTN of haplotype-AACTCTT (Hap.4) of this locus is significantly higher than that of other haplotypes, which is speculated to be a potential RTN related new locus.

Spike length is closely related to yield traits such as spikelet number, grain number per spike and thousand grain weight. Among 48 loci significantly associated with SL detected on 19 wheat chromosomes, The SNP (Chr7A_675095747) regulating RTN is also in the column, and is consistent with the SNP found in winter wheat by Gill et al. (2022). Yao (2017) found that *OsFBX76*-OE reduced thousand grain weight, seed setting rate and yield per plant, while *OsFBX76*-RNAi increased seed yield per plant. Therefore, *OsFBX76* affects the yield per plant of rice by negatively regulating grain size and seed setting rate. Li et al. (2020) found the SNP regulating SL on the chromosome 7B of wild Emmer wheat was consistent with Chr7B_463032072 found in this study, and the annotated gene *TracesCS7B02G250800.1* encodes the GDSL lipase gene family, which was essential for the development of anther and pollen as well as hormone signal transduction (Zhang 2020a). Therefore, we speculate that these genes affect the morphological changes of plants and spike by regulating hormone changes in plants. We further analyzed the other two significantly associated SNPs that regulate SL. Chr2D_650894588 annotated to the *TraesCS2D02G599600.1* encodes the *MK*. At present, the research on *MK* mainly focuses on important agronomic traits such as plant overexpression and secondary metabolites, grain size and grain weight (Champany and Tourte 1998). *TraesCS2D02G599900.1* encodes *CDCP*. *CDCP* can improve the utilization efficiency of nitrogen in plants, so as to achieve the purpose of increasing yield (Hao et al. 2016). In this study, the significantly associated SNPs on the short arm of chromosome 6A did not overlap with the physical positions of previous QTLs of SL. However, this locus was detected in multiple environments. Therefore, it may be potential locus to be discovered. Chr6A_17541092 annotates the *TraesCS6A02G035700.1* encoding the transcriptional repressor *OFPI4*, which found regulating the grains shape of rice (Zhao et al. 2018). Meanwhile, Chr6A_17541092 Hap.3 is an elite haplotype which have advantages on SL and SY traits.

A total of seven loci were significantly associated with SW in this study. A stable major QTL associated with SW was reported on chromosome 5A (*qSW-5A.1*, 646.3-657.5Mb) (Shui et al. 2020). The SNP Chr5A_657544633 was also associated with the same position, and was annotated to nine candidate genes. The *TraesCS5A02G487000.1* encoded *STK*. It has been found that *ZmSTK1* is a protein kinase specifically expressed in maize pollen, which plays a role in the development of maize pollen or the elongation of pollen tubes, and pollination and fertilization are closely related to the growth and fruiting of maize (Fan et al. 2018). Therefore, we infer that

the *STK* affects seed growth and compactness by affecting the pollination ability of *P. juncea* during flowering stage, which leads to the change of SW. The results laid the foundation for further research on spike width traits.

The number of grains per spike is the main factor affecting the seed yield of Triticeae crops. The number of grains per panicle is mainly determined by the number of spike rachis nodes. Fan et al. (2024) cloned a gene *HvSRN1* that controls the number of spike rachis nodes in barley, and clarified the regulation of this gene on traits such as spike length and yield per plant in barley. QTLs regulating SNN were also found to be distributed in Chr7A (*qNRN-7A.h2*, 674Mb) in wheat (Voss-Fels et al. 2019). The candidate gene *TraesCS7A01G481600* encodes F-box protein. In this study, Chr7A_675095747 was similar to the locus located on chromosome 7A in previous studies and was detected in multiple environments. Therefore, it can be determined that this locus controls SNN. Further analysis identified a candidate gene *TraesCS7A02G483300.1* on chromosome 7A encoding F-box protein, which is the same as the annotation information in the above study, and is homologous to the rice ABBERANT PANICLE ORGANIZATION1 (*AP01*) gene, which has a significant effect on spike morphological changes.

So far, extensive research has been conducted of thousand grain weight, which is essential for understanding its genetic mechanism. In this study, a total of nine SNPs related to TKW were detected. Cao et al. (2020) located the QTL of TKW in wheat at the position of 634.5-647.5 Mb (*QTKW.ndsu.2D*). Guan et al. (2018) QTL of TKW in wheat at the position of 675 Mb position (*QTgw.cau-7A.4*). They are aligned to the SNPs on chromosome 2D and 7A in this study. *TraesCS2D02G576800.1* annotated by Chr2D_640215049 encode *DCAF*. The *DCAF* gene is associated with abortion and blighted grain in rice (Luo et al. 2011), leading to a decrease in seed yield. A total of fourteen SNPs were associated with SWS. The two SNPs identified on Chr5D and Chr6D aligned to reported QTLs (Liu et al. 2023). *Chr5D_405156269* has annotated three candidate genes, with annotated information for RNA polymerase II and *PPR*, respectively. Research has shown that RNA polymerase II is involved in many biological processes, including gene regulation, cell differentiation, cell proliferation, and so on (Zhang et al. 2022). Liu (2016) cloned the *TaPPR4* in wheat and found that *TaPPR4*-OE wheat lines tiller number decreased, the spike length increased, the number of grains per spike decreased, and the seeds were shriveled. This result is of great significance for the study of *PPR* protein in other plants. Six SNPs were significantly associated with SY. SNP loci (AX_109941480, 643.0-650.7Mb) controlling seed yield have been found on wheat 2D chromosome (Li et al. 2018a), which are aligned to the same genomic region (Chr2D_650894588) in this study. Further analysis annotated a candidate gene *TraesCS2D02G600000.2* on Chr2D, which encodes callose synthase and plays an important role in endosperm development (Wang et al. 2004). The Chr2D_650894588 Hap.3 was an elite haplotype on seed yield trait. Fei et al. (2022) located the yield regulating SNP of wheat during the late filling stage to the positions of 15.94-18.67Mb on wheat chromosome 6A, which overlapped with the Chr6A_17541092 locus. These two loci could be identified as the yield-controlling SNPs.

In this study, 15 pleiotropic loci associated with two or more seed related traits were found

by genome-wide association analysis. Combined with phenotype, it is obviously that the function of the same locus gene can simultaneously regulate multiple traits, such as increased spike length and seed yield. Therefore, through technological innovation, the full development and utilization of forage resources, and the precise improvement of forage germplasm through molecular design are revolutionary technologies for cultivating super forage varieties.

Conclusion

Seed yield-related traits in *P. juncea*, a key forage species for grassland restoration and livestock production, are polygenic quantitative traits critical for breeding improvement. GWAS was conducted on a natural population of *P. juncea* using the wheat reference genome to dissect the genetic basis of eight seed yield traits. We identified 121 significant SNPs, including 19 stable loci detected across two environments, a total of 91 candidate genes were annotated, which involve the synthesis of cell wall polysaccharides and proteins, plant growth and development, photosynthesis, gibberellin regulation, hormone signal transduction, phenylalanine metabolism, and amino acid metabolism processes. The objectives of this study, QTL and the genomic information of seed traits may integrate marker-assisted selection (MAS) in the development of new cultivars.

Acknowledgment

We thank the test support Key Laboratory of Grassland Resources and Key Laboratory of Forage Cultivation, Processing and High Efficient Utilization of the Ministry of Agriculture and Grassland Germplasm Innovation and Sustainable Utilization of Grassland Resources in Inner Mongolia Autonomous Region, Inner Mongolia Agricultural University.

Funding

This study was financially supported by the National Natural Science Foundation of China, and the Key Project of the Natural Science Foundation of Inner Mongolia of China. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

The National Natural Science Foundation of China: 32371762.

The Key Project of Natural Science Foundation Inner Mongolia of China: 2023ZD07.

Competing interests

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

Author Contributions

ZL: wrote the manuscript. TW and XMR: conducted experiment, analyzed data. FH and YMM: analyzed data. LY: conceived and designed research. All authors have read and approved the manuscript.

Data Availability

The raw sequencing reads in this study are available in the Sequence Read Archive database (www.ncbi.nlm.nih.gov/sra) at NCBI (National Center for Biotechnology Information), accession number PRJNA1014568 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1014568/>).

Supplemental Information

Supplemental information for this article can be found at Supplemental Files.

References

- Albrechtsen RS, Davis RL, Keim WF (1966) Components of Seed Yield and Associated Characteristics in *Lotus corniculatus* L. ¹. Crop Science 6:355–358.
<https://doi.org/10.2135/cropsci1966.0011183X000600040018x>
- Asay KH (2008) Breeding potentials in perennial Triticeae grasses. Hereditas 116:167–173.
<https://doi.org/10.1111/j.1601-5223.1992.tb00223.x>
- Bai S, Yuan F, Zhang H, et al (2020) Characterization of the Wheat-Psathyrostachys huashania Keng 2Ns/2D Substitution Line H139: A Novel Germplasm With Enhanced Resistance to Wheat Take-All. Front Plant Sci 11:233. <https://doi.org/10.3389/fpls.2020.00233>
- Barrett JC (2009) Haploview: Visualization and Analysis of SNP Genotype Data. Cold Spring Harb Protoc 2009:pdb.ip71. <https://doi.org/10.1101/pdb.ip71>

609 Bian J, Ren G, Han C, et al (2020) Comparative analysis on grain quality and yield of different
 610 panicle weight indica-japonica hybrid rice (*Oryza sativa* L.) cultivars. *Journal of Integrative*
 611 *Agriculture* 19:999–1009. [https://doi.org/10.1016/S2095-3119\(19\)62798-X](https://doi.org/10.1016/S2095-3119(19)62798-X)

612 Bradbury PJ, Zhang Z, Kroon DE, et al (2007) TASSEL: software for association mapping of
 613 complex traits in diverse samples. *Bioinformatics* 23:2633–2635.
 614 <https://doi.org/10.1093/bioinformatics/btm308>

615 Champenoy S, Tourte M (1998) Expression of the yeast mevalonate kinase gene in transgenic
 616 tobacco. *Molecular Breeding* 4:291–300. <https://doi.org/10.1023/A:1009686514215>

617 Cuevas HE, Prom LK (2020) Evaluation of genetic diversity, agronomic traits, and anthracnose
 618 resistance in the NPGS Sudan Sorghum Core collection. *BMC Genomics* 21:88.
 619 <https://doi.org/10.1186/s12864-020-6489-0>

620 Degenhardt RF, Bonham-Smith PC (2008) Arabidopsis Ribosomal Proteins RPL23aA and
 621 RPL23aB Are Differentially Targeted to the Nucleolus and Are Disparately Required for Normal
 622 Development. *Plant Physiology* 147:128–142. <https://doi.org/10.1104/pp.107.111799>

623 Dewey DR (1984) THE GENOMIC SYSTEM OF CLASSIFICATION AS A GUIDE TO
 624 INTERGENERIC HYBRIDIZATION WITH THE PERENNIAL TRITICEAE. Plenum Press,
 625 New York 209–279. https://doi.org/10.1007/978-1-4613-2429-4_9

626 Easson DL, White EM, Pickles SJ (1993) The effects of weather, seed rate and cultivar on
 627 lodging and yield in winter wheat. *J Agric Sci* 121:145–156.
 628 <https://doi.org/10.1017/S0021859600077005>

629 Fan M, Zhang C, Shi L, et al (2018) *Zm STK 1* and *Zm STK 2*, encoding receptor-like
 630 cytoplasmic kinase, are involved in maize pollen development with additive effect. *Plant*
 631 *Biotechnology Journal* 16:1402–1414. <https://doi.org/10.1111/pbi.12880>

632 Foulkes MJ, Sylvester-Bradley R, Weightman R, Snape JW (2007) Identifying physiological
 633 traits associated with improved drought resistance in winter wheat. *Field Crops Research*

103:11–24. <https://doi.org/10.1016/j.fcr.2007.04.007>

Gao F, Ma D, Yin G, et al (2017) Genetic Progress in Grain Yield and Physiological Traits in Chinese Wheat Cultivars of Southern Yellow and Huai Valley since 1950. *Crop Science* 57:760–773. <https://doi.org/10.2135/cropsci2016.05.0362>

Garcia M, Eckermann P, Haefele S, et al (2019) Genome-wide association mapping of grain yield in a diverse collection of spring wheat (*Triticum aestivum* L.) evaluated in southern Australia. *PLoS ONE* 14:e0211730. <https://doi.org/10.1371/journal.pone.0211730>

Han J, Liu Y, Hou C, et al (2020) A 1Ns Disomic Addition from *Psathyrostachys Huashanica* Keng Confers Resistance to Powdery Mildew in Wheat. *Agronomy* 10:312. <https://doi.org/10.3390/agronomy10020312>

Hao Q, Shang W, Zhang C, et al (2016) Identification and Comparative Analysis of CBS Domain-Containing Proteins in Soybean (*Glycine max*) and the Primary Function of GmCBS21 in Enhanced Tolerance to Low Nitrogen Stress. *IJMS* 17:620. <https://doi.org/10.3390/ijms17050620>

Heffner EL, Sorrells ME, Jannink J (2009) Genomic Selection for Crop Improvement. *Crop Science* 49:1–12. <https://doi.org/10.2135/cropsci2008.08.0512>

Hu XK, Dai SF, Ouellet T, et al (2018) Characterization of novel D-hordeins from *Psathyrostachys juncea*. *Biologia plant* 62:369–378. <https://doi.org/10.1007/s10535-018-0775-6>

Li F, Wen W, He Z, et al (2018a) Genome-wide linkage mapping of yield-related traits in three Chinese bread wheat populations using high-density SNP markers. *Theor Appl Genet* 131:1903–1924. <https://doi.org/10.1007/s00122-018-3122-6>

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26:589–595. <https://doi.org/10.1093/bioinformatics/btp698>

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

659 Li H, Ye G, Wang J (2007) A Modified Algorithm for the Improvement of Composite Interval
660 Mapping. *Genetics* 175:361–374. <https://doi.org/10.1534/genetics.106.066811>

661 Li J, Tang W, Zhang Y-W, et al (2018b) Genome-Wide Association Studies for Five Forage
662 Quality-Related Traits in Sorghum (*Sorghum bicolor* L.). *Front Plant Sci* 9:1146.
663 <https://doi.org/10.3389/fpls.2018.01146>

664 Li Z, Wang T, Yun L, et al (2023) Association Analysis of Tiller-Related Traits with EST-SSR
665 Markers in *Psathyrostachys juncea*. *Genes* 14:1970. <https://doi.org/10.3390/genes14101970>

666 Liang J, Li H, Li N, et al (2023) Analysis and Prediction of the Impact of Socio-Economic and
667 Meteorological Factors on Rapeseed Yield Based on Machine Learning. *Agronomy* 13:1867.
668 <https://doi.org/10.3390/agronomy13071867>

669 Liu J-M, Xu Z-S, Lu P-P, et al (2016) Genome-wide investigation and expression analyses of the
670 pentatricopeptide repeat protein gene family in foxtail millet. *BMC Genomics* 17:840.
671 <https://doi.org/10.1186/s12864-016-3184-2>

672 Liu Y, Chen J, Yin C, et al (2023) A high-resolution genotype–phenotype map identifies the
673 TaSPL17 controlling grain number and size in wheat. *Genome Biol* 24:196.
674 <https://doi.org/10.1186/s13059-023-03044-2>

675 Lu Y, Yan J, Guimarães CT, et al (2009) Molecular characterization of global maize breeding
676 germplasm based on genome-wide single nucleotide polymorphisms. *Theor Appl Genet* 120:93–
677 115. <https://doi.org/10.1007/s00122-009-1162-7>

678 Luo X, Niu X, Huang W, Liu Y (2011) Expression Characteristics and Functional Analysis of
679 Rice DDB1 Gene. *Chin J Appl Environ Biol* 17:001–006.
680 <https://doi.org/10.3724/SP.J.1145.2011.00001>

681 Montana G, Hoggart C (2007) Statistical software for gene mapping by admixture linkage
682 disequilibrium. *Briefings in Bioinformatics* 8:393–395. <https://doi.org/10.1093/bib/bbm035>

683 Mu Y, Wen X, Yang Y, et al (2019) Evaluation of Basic Agronomic Traits of *Secale cereale*

684 subsp. Segetale. Crops 20:42–48. <https://doi.org/10.16035/j.issn.1001-7283.2019.04.007>

685 Nayak AK, C A, Behera S, et al (2022) Genetic Dissection of Grain Size Traits Through
686 Genome-Wide Association Study Based on Genic Markers in Rice. Rice Science 29:462–472.
687 <https://doi.org/10.1016/j.rsci.2022.07.006>

688 Nishimura T, Wada T, Yamamoto KT, Okada K (2005) The *Arabidopsis* STV1 Protein,
689 Responsible for Translation Reinitiation, Is Required for Auxin-Mediated Gynoecium Patterning.
690 The Plant Cell 17:2940–2953. <https://doi.org/10.1105/tpc.105.036533>

691 Pace J, Gardner C, Romay C, et al (2015) Genome-wide association analysis of seedling root
692 development in maize (*Zea mays* L.). BMC Genomics 16:47. [https://doi.org/10.1186/s12864-](https://doi.org/10.1186/s12864-015-1226-9)
693 015-1226-9

694 Shokat S, Sehgal D, Vikram P, et al (2020) Molecular Markers Associated with Agro-
695 Physiological Traits under Terminal Drought Conditions in Bread Wheat. IJMS 21:3156.
696 <https://doi.org/10.3390/ijms21093156>

697 Shui Z, An P, Liu T, et al (2020) QTL analysis of spike length and width using RIL population of
698 synthetic wheat. Journal of Triticeae crops 40:656–664. [https://doi.org/10.7606/j.issn.1009-](https://doi.org/10.7606/j.issn.1009-1041.2020.06.02)
699 1041.2020.06.02

700 Smith SE, Kuehl RO, Ray IM, et al (1998) Evaluation of Simple Methods for Estimating
701 Broad-Sense Heritability in Stands of Randomly Planted Genotypes. Crop Science 38:1125–
702 1129. <https://doi.org/10.2135/cropsci1998.0011183X003800050003x>

703 Sukumaran S, Dreisigacker S, Lopes M, et al (2015) Genome-wide association study for grain
704 yield and related traits in an elite spring wheat population grown in temperate irrigated
705 environments. Theor Appl Genet 128:353–363. <https://doi.org/10.1007/s00122-014-2435-3>

706 Takeda S, Matsuoka M (2008) Genetic approaches to crop improvement: responding to
707 environmental and population changes. Nat Rev Genet 9:444–457.
708 <https://doi.org/10.1038/nrg2342>

- 709 Tian F, Bradbury PJ, Brown PJ, et al (2011) Genome-wide association study of leaf architecture
710 in the maize nested association mapping population. *Nat Genet* 43:159–162.
711 <https://doi.org/10.1038/ng.746>
- 712 Van Inghelandt D, Melchinger AE, Lebreton C, Stich B (2010) Population structure and genetic
713 diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theor*
714 *Appl Genet* 120:1289–1299. <https://doi.org/10.1007/s00122-009-1256-2>
- 715 Voss-Fels KP, Keeble-Gagnère G, Hickey LT, et al (2019) High-resolution mapping of rachis
716 nodes per rachis, a critical determinant of grain yield components in wheat. *Theor Appl Genet*
717 132:2707–2719. <https://doi.org/10.1007/s00122-019-03383-4>
- 718 Wang L, Liu X, Lu Y, et al (2004) Endosperm Development in Autotetraploid Rice: The
719 Development of the Cellulose Wall of Aleuronic Layer Cell □ Starch Accumulation of
720 Endosperm and Formation of a Callose “Sheath-like” Structure. *Chinese J Rice Sci* 18:507–514.
721 <https://doi.org/10.16819/j.1001-7216.2004.06.007>
- 722 Xiao Y, Tong H, Yang X, et al (2016) Genome-wide dissection of the maize ear genetic
723 architecture using multiple populations. *New Phytologist* 210:1095–1106.
724 <https://doi.org/10.1111/nph.13814>
- 725 Xie W, Zhao X, Zhang J, et al (2015) Assessment of genetic diversity of Siberian wild rye
726 (*Elymus sibiricus* L.) germplasms with variation of seed shattering and implication for future
727 genetic improvement. *Biochemical Systematics and Ecology* 58:211–218.
728 <https://doi.org/10.1016/j.bse.2014.12.006>
- 729 Xiong Y, Xiong Y, Yu Q, et al (2020) Genetic variability and structure of an important wild
730 steppe grass *Psathyrostachys juncea* (Triticeae: Poaceae) germplasm collection from north and
731 central Asia. *PeerJ* 8:e9033. <https://doi.org/10.7717/peerj.9033>
- 732 Xu C, Zhang H, Sun J, et al (2018) Genome-wide association study dissects yield components
733 associated with low-phosphorus stress tolerance in maize. *Theor Appl Genet* 131:1699–1714.
734 <https://doi.org/10.1007/s00122-018-3108-4>

735 Xu K, Wu N, Yao W, et al (2021) The Biological Function and Roles in Phytohormone
736 Signaling of the F-Box Protein in Plants. *Agronomy* 11:2360.
737 <https://doi.org/10.3390/agronomy11112360>

738 Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: A Tool for Genome-wide Complex
739 Trait Analysis. *The American Journal of Human Genetics* 88:76–82.
740 <https://doi.org/10.1016/j.ajhg.2010.11.011>

741 Yoshikawa T, Eiguchi M, Hibara K-I, et al (2013) Rice SLENDER LEAF 1 gene encodes
742 cellulose synthase-like D4 and is specifically expressed in M-phase cells to regulate cell
743 proliferation. *Journal of Experimental Botany* 64:2049–2061. <https://doi.org/10.1093/jxb/ert060>

744 Zhang C, Dong S-S, Xu J-Y, et al (2019) PopLDdecay: a fast and effective tool for linkage
745 disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35:1786–1788.
746 <https://doi.org/10.1093/bioinformatics/bty875>

747 Zhang H, Li X, Song R, et al (2022) Cap-binding complex assists RNA polymerase II
748 transcription in plant salt stress response. *Plant Cell & Environment* 45:2780–2793.
749 <https://doi.org/10.1111/pce.14388>

750 Zhang W (2020a) A GDSSL Lipase Is Required for Anther and Pollen Development. *Plant*
751 *Physiol* 182:1810–1811. <https://doi.org/10.1104/pp.20.00278>

752 Zhang X, Guan Z, Li Z, et al (2020b) A combination of linkage mapping and GWAS brings new
753 elements on the genetic basis of yield-related traits in maize across multiple environments.
754 *Theoretical and Applied Genetics* 133, 2881–2895. <https://doi.org/10.1007/s00122-020-03639-4>

755 Zhang Y, Liu P, Zhang X, et al (2018) Multi-Locus Genome-Wide Association Study Reveals
756 the Genetic Architecture of Stalk Lodging Resistance-Related Traits in Maize. *Front Plant Sci*
757 9:611. <https://doi.org/10.3389/fpls.2018.00611>

758 Zhang Z, Ersoz E, Lai C-Q, et al (2010) Mixed linear model approach adapted for genome-wide
759 association studies. *Nat Genet* 42:355–360. <https://doi.org/10.1038/ng.546>

760 Zhao D-S, Li Q-F, Zhang C-Q, et al (2018) GS9 acts as a transcriptional activator to regulate rice
761 grain shape and appearance quality. *Nat Commun* 9:1240. [https://doi.org/10.1038/s41467-018-](https://doi.org/10.1038/s41467-018-03616-y)
762 03616-y

763 Zhao H, Li Z, Wang Y, et al (2022) Cellulose synthase-like protein OsCSLD4 plays an important
764 role in the response of rice to salt stress by mediating abscisic acid biosynthesis to regulate
765 osmotic stress tolerance. *Plant Biotechnology Journal* 20:468–484.
766 <https://doi.org/10.1111/pbi.13729>

767 Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies.
768 *Nat Genet* 44:821–824. <https://doi.org/10.1038/ng.2310>

769

Figure 1

Figure 1 Boxplots of the mean amounts of eight seed yield-related traits of different locations during 2021-2022.

* $p < 0.05$; Different capital letters indicate the difference significance between the same place in different years.

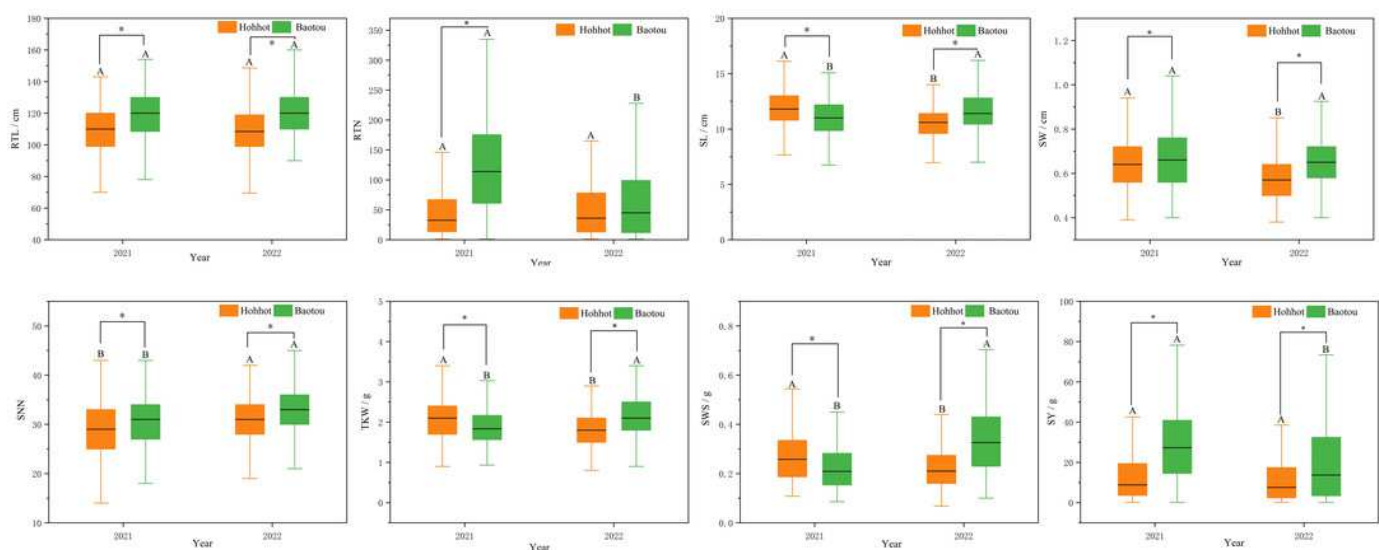


Figure 2

Figure 2 Pearson correlation coefficients for the eight seed yield-related traits evaluated in the two locations.

(a) Hohhot, (b) Baotou. Statistical significance of coefficients labeled as ***, **, * for $P < 0.0001$, <0.01 , <0.05 .

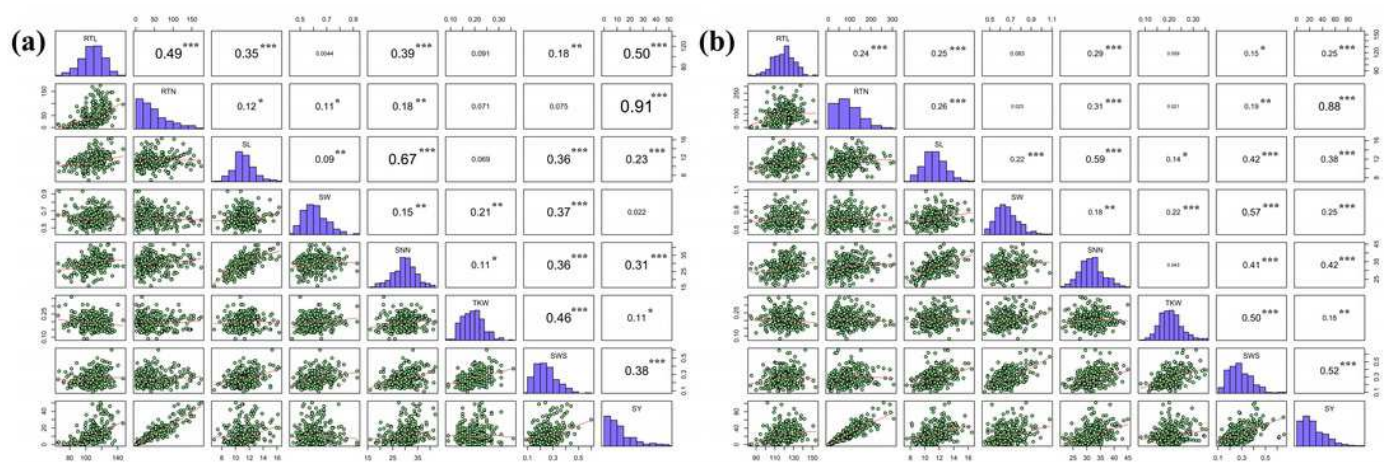


Figure 3

Figure 3 LD decay, PCA, and population structure of *P. juncea* .

(a) $\ln P(D)$, the log probability of the data. K means subpopulations. (b) Principal component analysis (PCA) of 300 accessions. (c) LD decay for five subgroup. The values on the Y-axis represent the squared correlation coefficient (r^2) and the values at X-axis represent physical distance (kb). (d) Population structure. Bar graphs for five subpopulations are indicated by different colors. The vertical coordinates of each subpopulation indicate the membership coefficient for each individual, and the digits on the horizontal coordinates represent the corresponding genotypes corresponding to the table. In each subpopulation, each vertical bar represents one genotype.

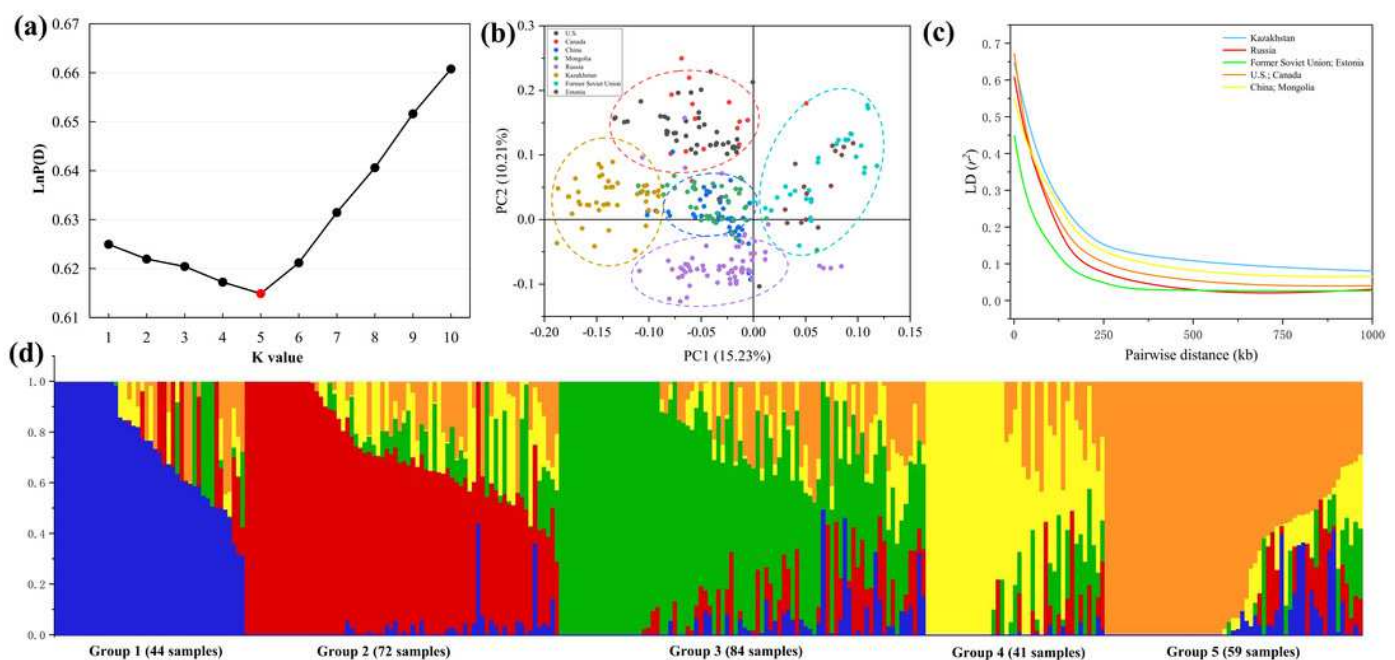


Figure 4

Figure 4 Manhattan plots for eight different traits of *P. juncea* .

Ring from inside to outside denotes RTL, RTN, SL, SW, SNN, TKW, SWS and SY respectively. Red horizontal line represents the draw P value threshold (P value $< 1.27E-06$). Blue dash line indicates Bonferroni corrected P value (P value $< 1.27E-07$). The outermost circle is the SNP density map.

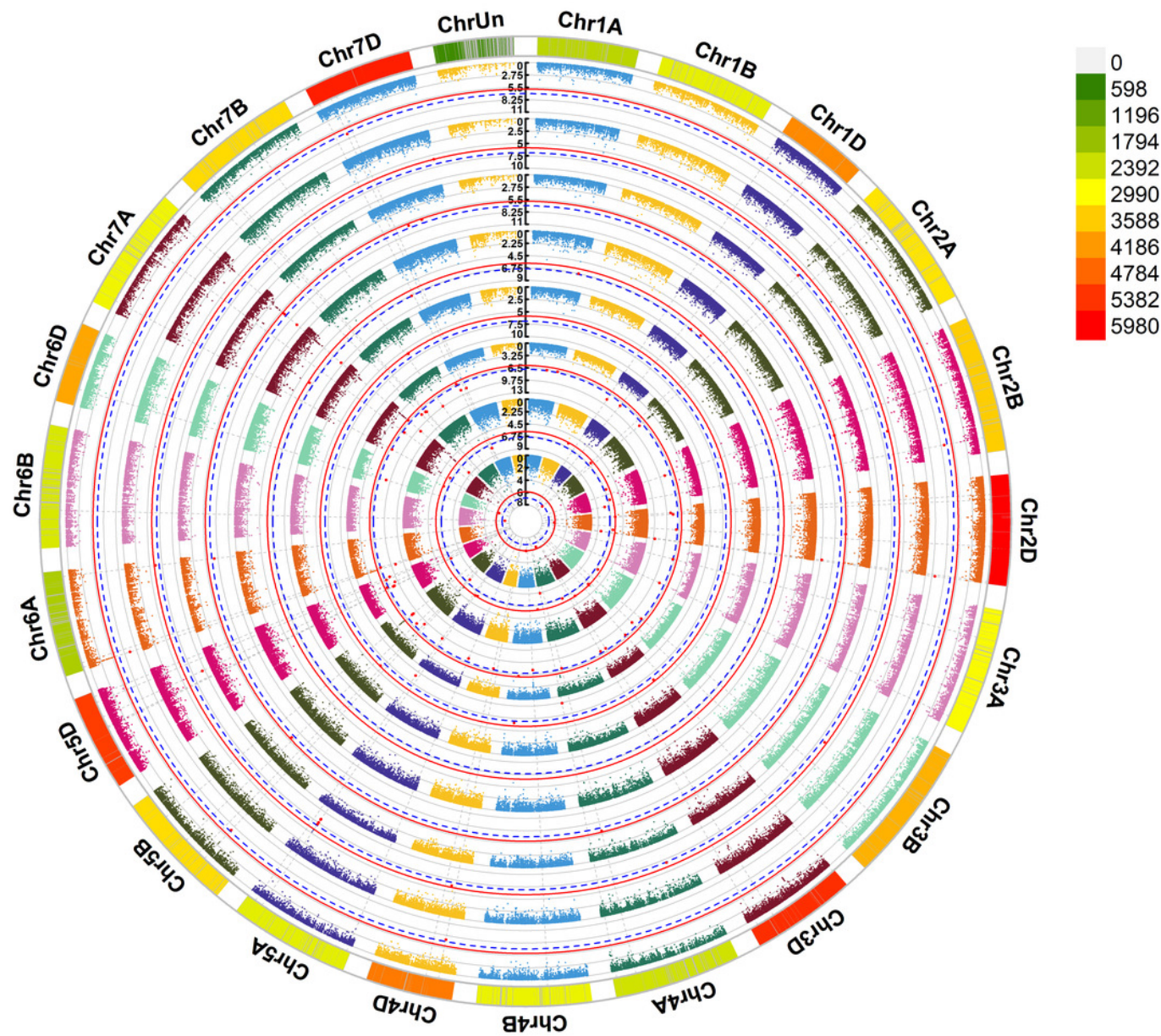


Figure 5

Figure 5 Distribution of 121 significantly associated SNPs on 21 chromosomes based on the physical distance from the eight seed yield agronomic trait differences.

The numbers on the right side of each column represent the physical location (Mb) of each lead SNP. Letters to the left of each column represent the SNP .

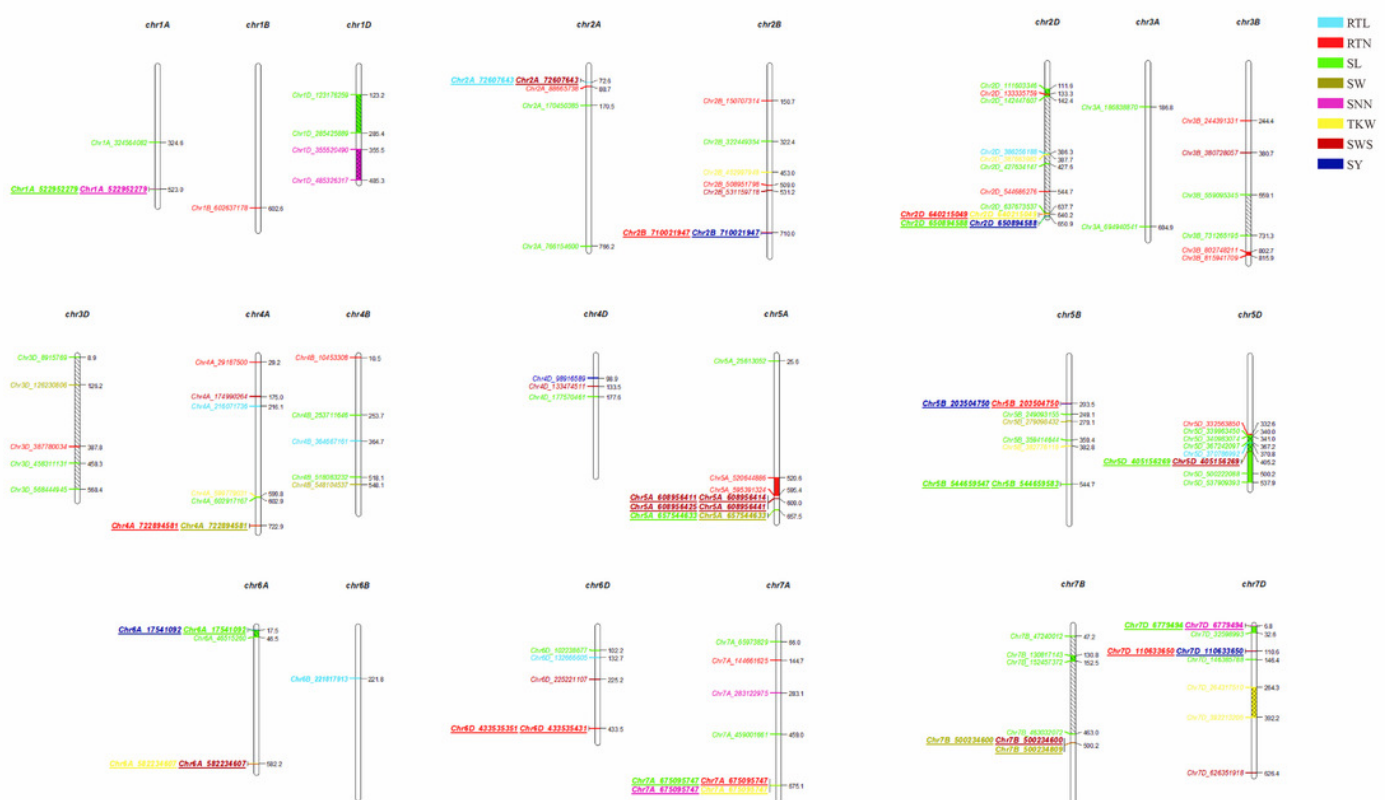


Figure 6

Figure 6 Haplotype analysis of Chr2D_640215049 and phenotypes of RTN and TKW among different haplotypes.

(a) Associated linkage regions. (b) Haplotype. (c) Phenotype. Num.: number of accessions, the same below.

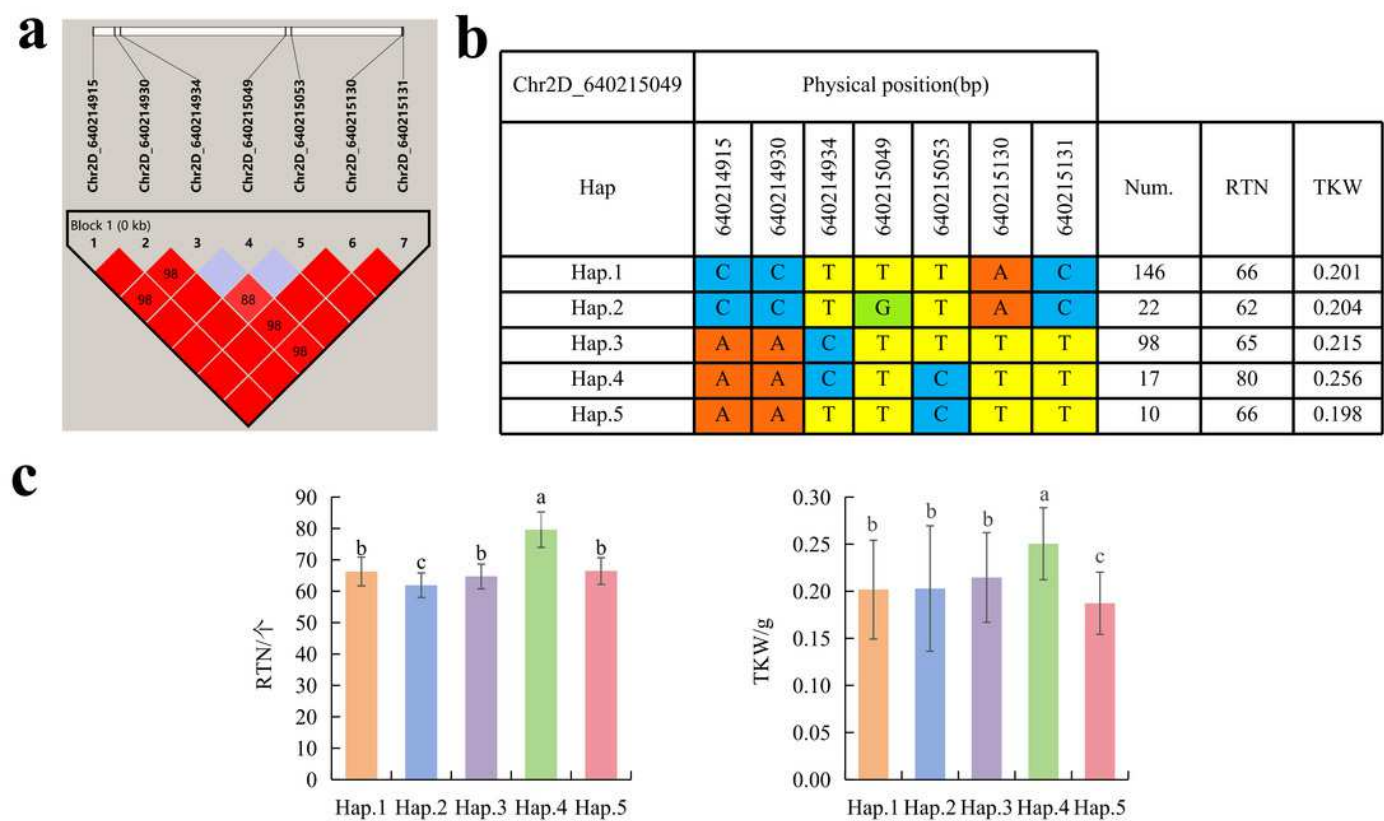


Figure 7

Figure 7 Haplotype analysis of Chr2D_650894588 and phenotypes of SL and SY among different haplotypes

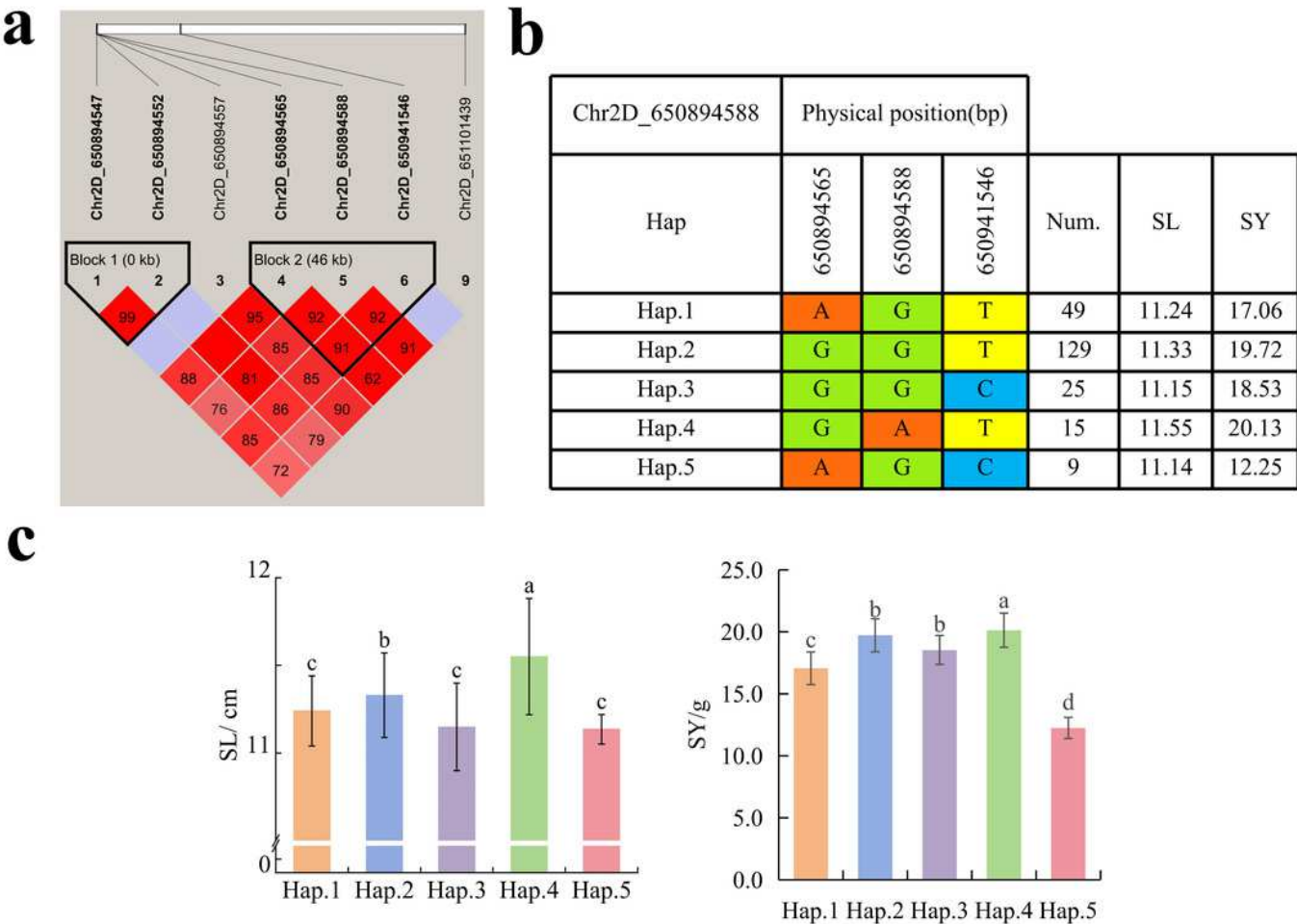


Figure 8

Figure 8 Haplotype analysis of Chr6A_17541092 and phenotypes of SL and SY among different haplotypes.

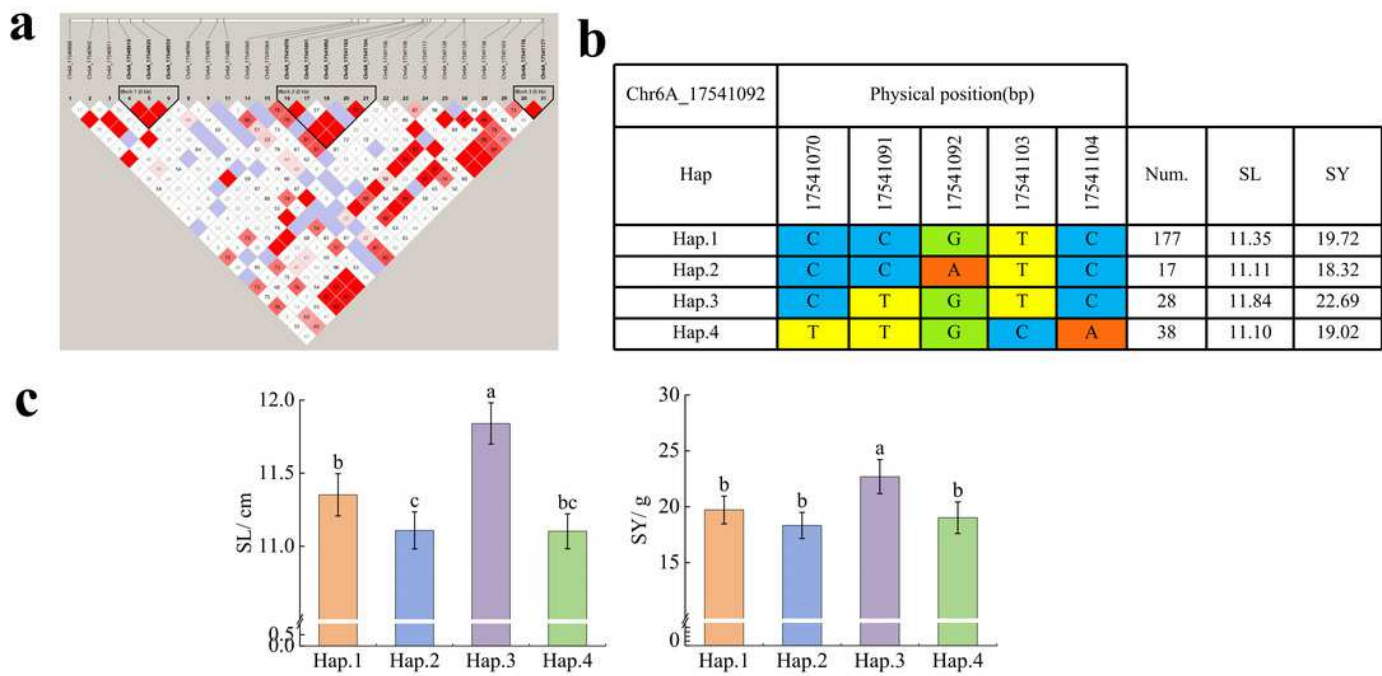


Figure 9

Figure 9 Haplotype analysis of Chr7A_675095747 and phenotypes of RTN-SL SNN and TKW among different haplotypes.

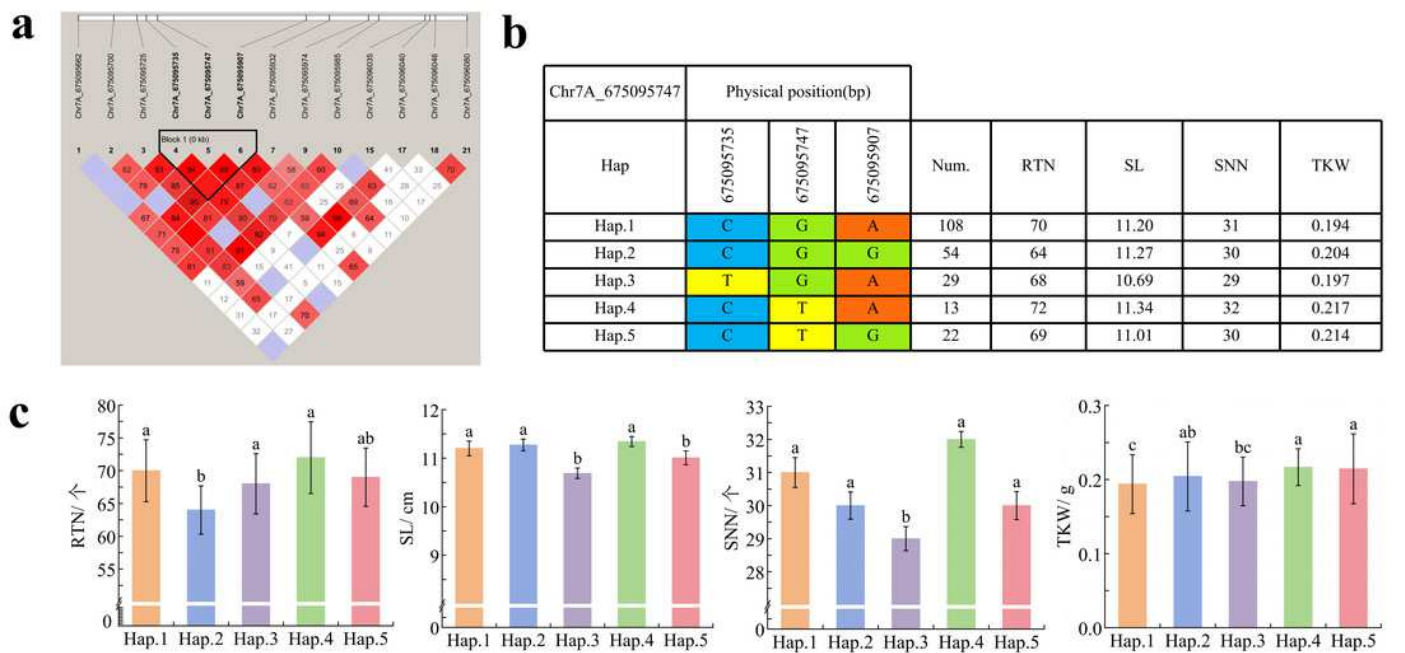


Table 1 (on next page)

Table 1 Statistical descriptions of eight seed -related traits in the 300 *P. juncea* accessions evaluated in four environments

Note: The eight traits included reproductive tiller length (RTL), reproductive tiller number (RTN), spike length (SL), spike width (SW), spike node number (SNN), thousand kernel weight (TKW), seed weight per spike (SWS), seed yield per plant (SY). Hohhot and Baotou were the two locations used for field evaluation of the traits.

Table 1 Statistical descriptions of eight seed-related traits in the 300 *P. juncea* accessions evaluated in four environments

Trait	Environment	Mean±SD	Min	Max	CV/%
RTL	2021-Hohhot	108.80±16.44	59.00	143.00	15.11
	2022-Hohhot	107.94±17.48	57.00	155.00	16.20
	2021-Baotou	118.44±16.59	50.00	154.00	11.01
	2022-Baotou	120.47±14.21	90.00	170.00	14.79
RTN	2021-Hohhot	44.61±39.64	1.00	209.00	58.87
	2022-Hohhot	52.18±50.48	1.00	207.00	66.73
	2021-Baotou	123.29±80.5	1.00	335.00	65.29
	2022-Baotou	61.12±58.49	1.00	293.00	65.71
SL	2021-Hohhot	11.93±2.03	6.12	17.86	16.99
	2022-Hohhot	10.92±5.89	5.80	15.40	13.93
	2021-Baotou	11.02±1.66	6.74	16.60	16.10
	2022-Baotou	11.66±1.97	6.75	18.00	15.92
SW	2021-Hohhot	0.65±0.12	0.39	1.08	19.17
	2022-Hohhot	0.58±0.09	0.38	0.85	15.24
	2021-Baotou	0.68±0.16	0.40	1.18	23.20
	2022-Baotou	0.66±0.11	0.40	1.02	17.13
SNN	2021-Hohhot	29.02±5.97	14.00	46.00	20.56
	2022-Hohhot	30.62±4.83	15.00	44.00	15.76
	2021-Baotou	30.58±5.14	18.00	48.00	16.82
	2022-Baotou	32.98±5.06	21.00	46.00	15.33
TKW	2021-Hohhot	2.23±0.25	0.94	3.93	26.58
	2022-Hohhot	1.89±0.15	0.83	3.54	28.48
	2021-Baotou	1.95±0.15	0.65	3.44	25.22
	2022-Baotou	2.29±0.26	0.61	4.42	27.45
SWS	2021-Hohhot	0.28±0.11	0.11	0.69	39.81
	2022-Hohhot	0.22±0.09	0.07	0.53	40.85
	2021-Baotou	0.23±0.09	0.09	0.56	40.42
	2022-Baotou	0.34±0.15	0.10	0.96	42.60
SY	2021-Hohhot	13.38±13.58	0.18	77.56	72.51
	2022-Hohhot	11.77±12.25	0.13	74.57	74.09
	2021-Baotou	30.69±22.04	0.12	122.12	71.21
	2022-Baotou	21.11±21.43	0.13	97.89	71.81

Note: The eight traits included reproductive tiller length (RTL), reproductive tiller number (RTN), spike length (SL), spike width (SW), spike node number (SNN), thousand kernel weight (TKW), seed weight per spike (SWS), seed yield per plant (SY). Hohhot and Baotou were the two locations used for field evaluation of the traits.

Table 2 (on next page)

Table 2 ANOVA of seed yield-related traits evaluated at two locations for two years

Note: NS, Non-Significant; DF, Degree of freedom. *, ** and *** indicate 0.05, 0.01, and 0.001 levels of significance.

1

Table 2 ANOVA of seed yield-related traits evaluated at two locations for two years

Source	DF	RTL	RTN	SL	SW	SNN	TKW	SWS	SY
Environment (E)	1	180.47 ***	50068 ***	1.5 NS	653.44 ***	411.83 ***	282.53 ***	444.17 ***	30429.3 ***
Year (Y)	1	1.33 NS	2075 ***	6.93 **	276.85 ***	401.48 ***	85.13 ***	325.33 ***	5414.58 ***
Genotype (G)	299	2.52 ***	684.65 ***	6.75 ***	15.34 ***	23.47 ***	48.7 ***	26.79 ***	473.77 ***
E×Y	1	1.26 NS	32874 ***	197.71 ***	127.71 ***	14.87 ***	1686.57 ***	1982.99 ***	12863 ***
G×E	299	1.13 NS	315.72 ***	2.8 ***	7.99 ***	5.73 ***	16.65 ***	11.41 ***	247.02 ***
G×Y	299	1.3 **	195.08 ***	1.94 ***	7.6 ***	5.79 ***	14.71 ***	8.88 ***	194.4 ***
G×E×Y	299	0.88 NS	177.27 ***	1.91 ***	3.19 ***	5.95 ***	16.34 ***	9.08 ***	4.76 ***
Error	1069	3.78	12.40	1.23	8.57	4.21	3.54	4.43	10.32
Heritability (h^2)	-	0.75	0.65	0.68	0.62	0.69	0.81	0.64	0.57

2 Note: NS, Non-Significant; DF, Degree of freedom. *, ** and *** indicate 0.05, 0.01, and 0.001 levels of significance.

3

Table 3(on next page)

Table 3 . Chromosome wise - distribution and genetic diversity of 84 , 024 SNPs and the intra-chromosomal estimated LD among 300 *P. juncea* genotypes.

Table 3. Chromosome wise-distribution and genetic diversity of 84,024 SNPs and the intra-chromosomal estimated LD among 300 *P. juncea* genotypes.

Chr	No. of SNP	Genetic diversity	Chr. LD (r^2)	D prime	No of marker pairs in perfect LD
1A	2381	0.2854	0.510	0.848	5523
1B	2994	0.3123	0.297	0.594	7083
1D	4858	0.3223	0.203	0.537	16833
2A	3613	0.3126	0.267	0.560	12542
2B	3961	0.2975	0.229	0.541	8465
2D	6717	0.2583	0.176	0.367	36193
3A	3234	0.3287	0.507	0.844	8384
3B	4282	0.3839	0.203	0.531	11137
3D	6122	0.2921	0.142	0.380	23018
4A	2604	0.3035	0.512	0.857	5275
4B	3149	0.2512	0.521	0.891	8341
4D	5051	0.3363	0.395	0.625	20224
5A	2855	0.3065	0.526	0.899	6034
5B	3753	0.2764	0.270	0.571	9690
5D	5868	0.3183	0.131	0.328	25089
6A	2141	0.3192	0.500	0.810	5035
6B	2794	0.2092	0.517	0.881	6476
6D	4330	0.3012	0.163	0.429	18403
7A	3085	0.3011	0.296	0.591	8463
7B	3827	0.3635	0.502	0.826	9329
7D	6405	0.3252	0.151	0.397	23211
A subgenome	19913	0.3081	0.445	0.773	7322
B subgenome	24760	0.2991	0.363	0.691	8646
D subgenome	39351	0.3120	0.194	0.438	23282

Table 4(on next page)

Table 4 The potential candidate genes identified corresponding to the significant SNP associated with eight seed yield traits in the population.

1 **Table 4.** The potential candidate genes identified corresponding to the significant SNP associated with eight seed yield traits in the population.

Traits	SNP	Gene ID	Start	End	Distance to peak SNP	Encoding Protein
RTL	Chr2A_72607643	TraesCS2A02G122600 .1	72691495	72695638	5'_83852	Cellulose synthase-like protein
		TraesCS2A02G122700 .1	72696518	72697360	3'_88875	Momilactone A synthase
RTN	Chr2D_64021504 9	TraesCS2D02G577000 .1	640297481	640299386	3'_82432	Serine/threonine-protein kinase SAPK5
		TraesCS4A02G457400 .1	722837844	722839786	3'_54795	30S ribosomal protein S13, RPS13
	Chr4A_72289458 1	TraesCS4A02G457500 .1	722841516	722845182	5'_49399	Glutaminyl-peptide cyclotransferase
		TraesCS4A02G457600 .3	722882391	722887381	3'_7200	Membrane protein of ER body-like protein, MEBL
		TraesCS4A02G457700 .2	722899708	722905412	5'_5127	Membrane protein of ER body-like protein, MEBL
		TraesCS4A02G457700 .1	722899708	722905534	5'_5127	Membrane protein of ER body-like protein, MEBL
		TraesCS4A02G457800 .1	722905079	722906903	3'_10498	30S ribosomal protein S13, RPS13
		TraesCS4A02G457900 .1	722907610	722916249	3'_13029	Disease resistance protein RPM1
	Chr7A_67509574 7	TraesCS7A02G483300 .1	675021874	675023612	3'_72135	F-box protein
		TraesCS7A02G483400 .1	675025492	675027295	3'_68452	--
		TraesCS7A02G483400 .2	675025492	675027309	3'_68438	--
		TraesCS7A02G483500	675046750	675047270	3'_48477	--

SL	Chr2D_65089458 8	.1	TraesCS7A02G483600	675112899	675115741	3'_17152	--
		.1	TraesCS2D02G599500	650831742	650837613	5'_56975	--
		.1	TraesCS2D02G599600	650860731	650863713	5'_30875	Mevalonate kinase
		.1	TraesCS2D02G599700	650869233	650871862	5'_22726	--
		.1	TraesCS2D02G599800	650879896	650883522	5'_11066	Probable LRR receptor-like serine/threonine-protein kinase
		.1	TraesCS2D02G599900	650927579	650929016	5'_32991	CBS domain-containing protein
		.2	TraesCS2D02G600000	650931852	650947979	5'_37264	Callose synthase
		.1	TraesCS2D02G600100	650955614	650957060	5'_61026	Stem-specific protein TSJT1
		.1	TraesCS2D02G600200	650962449	650964288	5'_67861	E3 ubiquitin-protein ligase SINA-like 2
	Chr5A_65754463 3	.1	TraesCS5A02G486900	657452611	657452823	3'_91810	Photosystem II reaction center protein
		.1	TraesCS5A02G487000	657481959	657486595	3'_58038	Probable serine/threonine-protein kinase
		.2	TraesCS5A02G487000	657481959	657486706	3'_57927	Probable serine/threonine-protein kinase
		.3	TraesCS5A02G487000	657483607	657486595	3'_58038	Probable serine/threonine-protein kinase
		.1	TraesCS5A02G487100	657490411	657493777	3'_50856	--

	TraesCS5A02G487200 .1	657496902	657498825	5'_45808	--
	TraesCS5A02G487300 .1	657537556	657539589	3'_5044	Cytochrome P450
	TraesCS5A02G487400 .1	657540736	657548288	--	--
	TraesCS5A02G487500 .1	657550034	657554787	5'_5401	F-box protein
Chr5D_40515626 9	TraesCS5D02G308200 .1	405154704	405157124	--	RNA polymerase II transcriptional coactivator KIWI
	TraesCS5D02G308300 .1	405209676	405211820	3'_53407	Putative pentatricopeptide repeat-containing protein
	TraesCS5D02G308400 .1	405251408	405255247	3'_95139	--
Chr6A_17541092	TraesCS6A02G035500 .1	17528410	17531735	3'_9357	--
	TraesCS6A02G035600 .1	17540023	17543295	--	--
	TraesCS6A02G035700 .1	17601604	17602335	5'_60512	Transcription repressor OFP14
	TraesCS6A02G035800 .1	17638201	17641019	3'_97109	--
Chr7A_67509574 7	TraesCS7A02G483300 .1	675021874	675023612	3'_72135	F-box protein
	TraesCS7A02G483400 .1	675025492	675027295	3'_68452	--
	TraesCS7A02G483400 .2	675025492	675027309	3'_68438	--
	TraesCS7A02G483500	675046750	675047270	3'_48477	--

		.1				
		TraesCS7A02G483600	675112899	675115741	3'_17152	--
Chr7B_46303207	2	.1				
		TraesCS7B02G250800	463033674	463037960	3'_1602	GDSL esterase/lipase EXL1
SW	Chr4A_72289458	1				
		TraesCS4A02G457400	722837844	722839786	3'_54795	30S ribosomal protein S13, RPS13
		.1				
		TraesCS4A02G457500	722841516	722845182	5'_49399	Glutaminyl-peptide cyclotransferase
		.1				
		TraesCS4A02G457600	722882391	722887381	3'_7200	Membrane protein of ER body-like protein, MEBL
		.3				
		TraesCS4A02G457700	722899708	722905412	5'_5127	Membrane protein of ER body-like protein, MEBL
		.2				
		TraesCS4A02G457700	722899708	722905534	5'_5127	Membrane protein of ER body-like protein, MEBL
		.1				
		TraesCS4A02G457800	722905079	722906903	3'_10498	30S ribosomal protein S13, RPS13
		.1				
	Chr5A_65754463	TraesCS4A02G457900	722907610	722916249	3'_13029	Disease resistance protein RPM1
		.1				
		TraesCS5A02G486900	657452611	657452823	3'_91810	Photosystem II reaction center protein
		.1				
		TraesCS5A02G487000	657481959	657486595	3'_58038	Probable serine/threonine-protein kinase
		.1				
		TraesCS5A02G487000	657481959	657486706	3'_57927	Probable serine/threonine-protein kinase
	3	.2				
		TraesCS5A02G487000	657483607	657486595	3'_58038	Probable serine/threonine-protein kinase
		.3				
		TraesCS5A02G487100	657490411	657493777	3'_50856	--
		.1				

		TraesCS5A02G487200 .1	657496902	657498825	5'_45808	--
		TraesCS5A02G487300 .1	657537556	657539589	3'_5044	Cytochrome P450
		TraesCS5A02G487400 .1	657540736	657548288	--	--
		TraesCS5A02G487500 .1	657550034	657554787	5'_5401	F-box protein
SNN	Chr7A_67509574 7	TraesCS7A02G483300 .1	675021874	675023612	3'_72135	F-box protein
		TraesCS7A02G483400 .1	675025492	675027295	3'_68452	--
		TraesCS7A02G483400 .2	675025492	675027309	3'_68438	--
		TraesCS7A02G483500 .1	675046750	675047270	3'_48477	--
		TraesCS7A02G483600 .1	675112899	675115741	3'_17152	--
TKW	Chr2D_64021504 9	TraesCS2D02G576700 .1	640174317	640176964	3'_38085	--
		TraesCS2D02G576800 .1	640206750	640216632	--	DDB1- and CUL4-associated factor homolog 1 GN=DCAF1
		TraesCS2D02G576900 .1	640240670	640242328	5'_25621	Uncharacterized acetyltransferase
	Chr7A_67509574 7	TraesCS7A02G483300 .1	675021874	675023612	3'_72135	F-box protein
		TraesCS7A02G483400 .1	675025492	675027295	3'_68452	--
		TraesCS7A02G483400	675025492	675027309	3'_68438	--

		.2				
		TraesCS7A02G483500	675046750	675047270	3'_48477	--
		.1				
		TraesCS7A02G483600	675112899	675115741	3'_17152	--
		.1				
		TraesCS2A02G122600	72691495	72695638	5'_83852	Cellulose synthase-like protein D3
		.1				
	Chr2A_72607643	TraesCS2A02G122700	72696518	72697360	3'_88875	Momilactone A synthase
		.1				
SWS		TraesCS5D02G308200	405154704	405157124	--	RNA polymerase II transcriptional coactivator KIWI
		.1				
	Chr5D_40515626	TraesCS5D02G308300	405209676	405211820	3'_53407	Putative pentatricopeptide repeat-containing protein
	9	.1				
		TraesCS5D02G308400	405251408	405255247	3'_95139	--
		.1				
		TraesCS2D02G599500	650831742	650837613	5'_56975	--
		.1				
		TraesCS2D02G599600	650860731	650863713	5'_30875	Mevalonate kinase
		.1				
		TraesCS2D02G599700	650869233	650871862	5'_22726	--
		.1				
	Chr2D_65089458	TraesCS2D02G599800	650879896	650883522	5'_11066	Probable LRR receptor-like serine/threonine-protein kinase
SY	8	.1				
		TraesCS2D02G599900	650927579	650929016	5'_32991	CBS domain-containing protein CBSX5
		.1				
		TraesCS2D02G600000	650931852	650947979	5'_37264	Callose synthase 7
		.2				
		TraesCS2D02G600100	650955614	650957060	5'_61026	Stem-specific protein TSJT1
		.1				

Chr6A_17541092	TraesCS2D02G600200 .1	650962449	650964288	5'_67861	E3 ubiquitin-protein ligase SINA-like 2
	TraesCS6A02G035500 .1	17528410	17531735	3'_9357	--
	TraesCS6A02G035600 .1	17540023	17543295	--	--
	TraesCS6A02G035700 .1	17601604	17602335	5'_60512	Transcription repressor OFP14
	TraesCS6A02G035800 .1	17638201	17641019	3'_97109	--