Theoretical and Applied Genetics

SNP-based pool genotyping and haplotype analysis accelerates fine mapping of the wheat genomic region containing stripe rust resistance gene Yr26

--Manuscript Draft--

Manuscript Number:	TAAG-D-17-00473R1
Full Title:	SNP-based pool genotyping and haplotype analysis accelerates fine mapping of the wheat genomic region containing stripe rust resistance gene Yr26
Article Type:	Original Article
Funding Information:	International S&T Cooperation Program of China (2015DFG32340) — Mr. Zhensheng Kang National Natural Science Foundation of China (31371924) — Prof. Dejun Han the National Key Research and Development Program of China (2016YFE0108600) — Prof. Dejun Han the earmarked funds for Modern Agro-industry Technology Research System (CARS-3-1-11) — Mr. Zhensheng Kang National Natural Science Foundation for Young Scientists of China (31701421) — Dr. Qingdong Zeng
Abstract:	Conventional gene mapping methods to identify genes associated with important agronomic traits require significant amounts of financial support and time. Here, a single nucleotide polymorphism (SNP)-based mapping approach, RNA-Seq and SNP array assisted super pooling analysis, was used for rapid mining of a candidate genomic region for stripe rust resistance gene Yr26 that has been widely used in wheat breeding programs in China. Large DNA and RNA super-pools were genotyped by Wheat SNP Array and sequenced by Illumina HiSeq, respectively. Hundreds of thousands of SNPs were identified and then filtered by multiple filtering criteria. Among selected SNPs, over 900 were found within an overlapping interval of less than 30 Mb as the Yr26 candidate genomic region in the centromeric region of chromosome arm 1BL. The 235 chromosome-specific SNPs were converted into KASP assays to validate the Yr26 interval in different genetic populations. Using a high-resolution mapping population (>30,000 gametes) we confined Yr26 to a 0.003 cM interval. The Yr26 target region was anchored to the common wheat IWGSC RefSeq v1.0 and wild emmer WEWSeq v.1.0 sequences, from which 488 kb and 454 kb fragments were obtained. Several candidate genes were identified in the target genomic region, but there was no typical resistance gene in either genome region. Haplotype analysis identified specific SNPs linked to Yr26 and developed robust and breeder-friendly KASP markers. This integration strategy can be applied to accelerate generating many markers closely linked to target genes/QTL for a trait of interest in wheat and other polyploid species.
Corresponding Author:	Zhensheng Kang, Ph.D. Northwest A&F University Yangling, CHINA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Northwest A&F University
Corresponding Author's Secondary Institution:	
First Author:	Jianhui Wu
First Author Secondary Information:	
Order of Authors:	Jianhui Wu

	Qingdong Zeng
	Qilin Wang
	Shengjie Liu
	Shizhou Yu
	Jingmei Mu
	Shuo Huang
	Hanan Sela
	Assaf Distelfeld
	Lili Huang
	Dejun Han
	Zhensheng Kang, Ph.D.
Order of Authors Secondary Information:	
Author Comments:	Dear editor, Thank you for your email enclosing the reviewers' comments. We have carefully reviewed the comments and have revised the manuscript accordingly. We very much appreciate the time and effort put into comments. The advice about formatting and structure of the paper was most helpful. Our responses are given in a point-by-point manner below. Except for the mentioned above, we have checked it again for English editing and the changes to the manuscript are also highlighted in yellow color. We hope that the changes satisfy all concerns about the article. If necessary we are more than happy to make any further changes to improve our paper and/or to its facilitate publication. Thank you very much for your work concerning our paper and looking forward to hearing from you soon. Sincerely yours, Zhensheng Kang and Dejun Han

Dear editor,

Thank you for your email enclosing the reviewers' comments. We have carefully reviewed the comments and have revised the manuscript accordingly. We very much appreciate the time and effort put into comments. The advice about formatting and structure of the paper was most helpful. Our responses are given in a point-by-point manner below. Changes to the manuscript are highlighted in yellow color.

According to editor's instructions:

1. **Please prepare your revised manuscript as follows:**

- The text, including tables and figure legends, in one file (preferably as ~.doc or ~.docx).

- Each figure in a separate file without the text.

Response: Thanks for the kind reminder! We have corrected the format.

2. **I fully agree with the suggestion from reviewer #4 that you should change the title. Please write instead of "uncovering specific genomic segment around a trait-associated locus" -> fine mapping of the *Yr26* gene for stripe rust. This will give you a much greater audience.**

Response: Thank you for your comments! We have rephrased the tile.

According to reviewers' instructions:

Reviewer #3:

General Comment:

Wu et al conducted bulked segregant RNA-seq (BSR-seq) and array baseed SNP identification for Yr26 resistance in hexaploid wheat. The authors were able to further confine the Yr26 loci to a 0.003 cM interval, and developed breeder-friendly KASP markers for the target gene. The reviewer finds that the quantity and quality of the works are in high standards.

Response: Thank you for your comments!

Major Compulsory Revisions

1. The authors used DiscoSNP for variant calling. This seems to be a little weired as this method appears to be for species without reference genomes? Why not try TASSEL or BWA-SAMTOOLS-BCFTOOLs etc approaches which are more established?

Response: Thank you for your comments! Dr. Hanan used DiscoSNP for variant calling based on the wild emmer wheat reference genome.

Minor Essential Revisions

2. Line 90-93. This sentence does not read well. "Compared to more traditional marker systems SNP and their alleles and haplotypes have a unique advantage in high resolution of genetic diversity which can be easily involved in relevant genes and even their functions"

Response: We have rewrite this sentence “Compared to more traditional marker systems SNP have the unique advantage of high resolution of genetic diversity and enable identification of relevant genes and even their functions.”

3. Line 135, Yr26 should be removed from Yr24, YrCH42 etc.

Response: Thank you for pointing this out! We have removed it.

4. Line 176, Wang et al 2014 paper should be cited when you first mention 90K.

Response: Thanks for the kind reminder! We have added the reference.

5. Line 330-331 'Previous studies' be explicit and add references.

Response: We have added the references.

6. Line 376-380. The authors first mentioned "thirty-six plants were identified". then mentioned sth like "reduced (again) to 36". This does not seem to be logical.

Response: Thank you for pointing this out! Here we have to apologize that we made a mistake. In fact, we reduced the number of plants to five (figure 4C). We have corrected it.

7. Line 411. Guarantees better be modified to 'enabled' or some other words with soft tones. No method can guarantee a reliable result for gene localization. People and experiment execution, alignment software, genetic stocks etc. all are critical.

Response: Thanks for the valuable suggestion! We have modified it as “provided”.

8. In the discussion section, the authors mentioned that MutRenSeq was not able to provide a candidate NLR for Yr26. The authors might want to add that this negative result is consistent with the putative gene contents there? (ABC transporter, WAT protein, and E3 ligase). Have the authors considered MutChromSeq (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5087116/) since mutants were already developed... as a possible alternative to locate the R-gene?

Response: Thanks for your good comments! We have added the sentence. We have contacted with Dr. Jaroslav Doležel for help. Next we will try to use this method.

9. The language and writings need improvements. Better be reviewed by some native English speakers?

Response: Thanks for your suggestion! We have improved language and writings again.

Reviewer #4: The authors used SNP-based pool genotyping and haplotype analysis to uncover the genomic segment of Yr26 in common wheat. RNA-Seq and SNP array were used to analyze several populations. These experiments produced solid data for mapping Yr26 gene, and provided reliable markers for marker assisted selection. It is a good paper. I have no major comments.

Response: Thank you very much for taking time to review our manuscript!

Minor comments:

1) One of the major goals of this study is fine mapping of Yr26 gene. It would help readers find the paper if the authors include Yr26 in the title.

Response: Thanks for the valuable suggestion! We have rephrased the tile.

2) Three recombinant inbred line (RIL) populations were developed to build three pairs of super-pools and one of the RIL populations was subjected to BSA and RNA-Seq. These populations provide solid results. However, it takes effort, time, and cost to develop these populations, collect and analyze genotypic and phenotypic data. I am wondering how redundant the results from the three populations. If the authors can provide some data and discussion about the types and numbers of populations required for pool genotyping, it would make this paper more attractive to general readers.

Response: Thank you for your valuable suggestion! Indeed it takes a lot of time to develop the RIL population. Some wheat BSR-Seq studies also used $F_{2:3}$ population to construct extreme pools and they got good results for gene mapping (Ramirez-Gonzalez et al. 2015 Plant Biotechnology Journal; Trick et al. 2012 BMC Plant Biology; Wang et al. 2017 The Crop Journal; Wu et al. 2018; Frontiers in Plant Science). However, *Yr26* is located at near centromere, we think it is better to pool RILs than to pool $F_{2:3}$ lines as the crossover rate in RILs is more. We have shown the results of three pairs of super-pools in figure 2. Your suggestions are pertinent and next we will try to analyze it. In fact, we have provided some data and discussion about the types and numbers of populations required for pool genotyping in another manuscript which is also submitted to TAG.

Except for the mentioned above, we have checked it again for English editing and the changes to the manuscript are also highlighted in yellow color. We hope that the changes satisfy all concerns about the article. If necessary we are more than happy to make any further changes to improve our paper and/or to its facilitate publication.

Thank you very much for your work concerning our paper and looking forward to hearing from you soon.

Sincerely yours,

Zhensheng Kang and Dejun Han

SNP-based pool genotyping and haplotype analysis accelerates fine mapping of the wheat genomic region containing stripe rust resistance gene *Yr26*

Jianhui Wu[1]†, Qingdong Zeng[1]†, Qilin Wang[1]†, Shengjie Liu[1], Shizhou Yu[1], Jingmei Mu[1], Shuo Huang[1], Hanan Sela[2], Assaf Distelfeld[3,4], Lili Huang[1], Dejun Han[1]*, and Zhensheng Kang[1]*

[1] State Key Laboratory of Crop Stress Biology for Arid Areas, Northwest A&F University, Yangling, Shaanxi 712100, P. R. China

[2] The Institute for Cereal Crops Improvement, Tel-Aviv University, Tel Aviv, Israel

[3] School of Plant Sciences and Food Security, Tel Aviv University, Tel Aviv, Israel. NRGene Ltd., Ness Ziona, Israel.

[4] Helmholtz Zentrum München, Plant Genome and Systems Biology, Neuherberg, Germany.

*Corresponding authors: (Han DJ, handj@nwsuaf.edu.cn; Kang ZS, kangzs@nwsuaf.edu.cn)

†Authors contributed equally to this work.

Author contributions statement

JH Wu designed and conducted the experiments, analyzed the data, and wrote the manuscript. QD Zeng analyzed the data, prepared the figures for the manuscript and contributed to writing the RNA-Seq sections; QL Wang participated in creating the genetic populations and analyzed the SNP array data. SJ Liu, JM Mu and S Huang participated in greenhouse and field experiments and contributed to the genotyping experiment. SZ Yu assisted in analyzing the data and prepared the figures for the manuscript. H Sela and A Distelfeld analyzed the data with the wild emmer genome. LL Huang participated in revising the manuscript. DJ Han and ZS Kang conceived and directed the project and revised the manuscript.

Compliance with ethical standards

Conflict of interest The authors have declared that no competing interests exist.

Ethical standard I declare on behalf of my co-authors that the work described is original, previously unpublished research, and not under consideration for publication elsewhere. The experiments in this study comply with the current laws of China.

Key message NGS-assisted super pooling emerging as powerful tool to accelerate gene mapping and haplotype association analysis within target region uncovering specific linkage SNPs or alleles for marker-assisted gene pyramiding.

Abstract

Conventional gene mapping methods to identify genes associated with important agronomic traits require significant amounts of financial support and time. Here, a single nucleotide polymorphism (SNP)-based mapping approach, RNA-Seq and SNP array assisted super pooling analysis, was used for rapid mining of a candidate genomic region for stripe rust resistance gene *Yr26* that has been widely used in wheat breeding programs in China. Large DNA and RNA super-pools were genotyped by Wheat SNP Array and sequenced by Illumina HiSeq, respectively. Hundreds of thousands of SNPs were identified and then filtered by multiple filtering criteria. Among selected SNPs, over 900 were found within an overlapping interval of less than 30 Mb as the *Yr26* candidate genomic region in the centromeric region of chromosome arm 1BL. The 235 chromosome-specific SNPs were converted into KASP assays to validate the *Yr26* interval in different genetic populations. Using a high-resolution mapping population (>30,000 gametes) we confined *Yr26* to a 0.003 cM interval. The *Yr26* target region was anchored to the common wheat IWGSC RefSeq v1.0 and wild emmer WEWSeq v.1.0 sequences, from which 488 kb and 454 kb fragments were obtained. Several candidate genes were identified in the target genomic region, but there was no typical resistance gene in either genome region. Haplotype analysis identified specific SNPs linked to *Yr26* and developed robust and breeder-friendly KASP markers. This integration strategy can be applied to accelerate generating many markers closely linked to target genes/QTL for a trait of interest in wheat and other polyploid species.

Introduction

It is predicted that the Earth's population will reach approximate 10 billion people by 2050 (United Nations, 2015). As a staple diet wheat provides about 20% of total grain production and significant growth in production of wheat plays an important role in global demand for food (www.fao.org/faostat). However, its productivity is often reduced by biotic stresses such as the rusts and its potential yield is rarely achieved (Hovmøller et al. 2010). Of the rusts, stripe rust or yellow rust (YR) caused by *Puccinia striiformis* f. sp. *tritici* (*Pst*) is a constant threat and leads the list of diseases capable of causing 5 to 25% yield losses in almost all wheat-growing regions (Chen 2005; Wellings 2011). China is one of the largest stripe rust epidemic regions in the world (Stubbs 1985) and widespread stripe rust epidemics have occurred with losses in some cases amounting to several million tons of grain (Chen et al. 2009; Li and Zeng 2002). Fungicide applications are often effectively used to control stripe rust; however, there are long-term environmental concerns (Chen 2014). On the contrary, utilization of cultivars carrying stacked effective resistance (R) genes is the most sustainable control

strategy.

Thus far, there are currently 80 permanently designated *Yr* genes for resistance to stripe rust in wheat. Most of them have been mapped using different types of molecular markers (McIntosh et al. 2016, 2017, pers. comm. 2018) and some linked markers have been successfully applied in molecular breeding. However, fine mapping and map-based cloning of *Yr* genes is still an enormous challenge and only four *Yr* genes have been cloned in wheat, namely, *Yr10*, *Yr36*, *Lr34/Yr18* and *Lr67*/*Yr46* (Fu et al. 2009; Krattinger et al. 2009; Liu et al. 2014; Moore et al. 2015). One limiting factor in gene cloning in wheat is the ~17Gb allohexaploid genome size with extensive stretches of repetitive DNA (>80%). Most markers are not effective for mapping genes in chromosomal regions with low gene density or in regions with low recombination. A recent milestone in wheat genomes has been completion of the entire *T. aestivum* cv. Chinese Spring RefSeq v.1.0 and subsequent high-quality gene models, (the International Wheat Genome Consortium (IWGSC), http://www.wheatgenome.org/; Clavijo et al. 2017) and *T. turgidum ssp. dicoccoides* cv. Zavitan WEWSeq v.1.0 (International Wild Emmer Wheat genome sequencing consortium, http://wewseq.wixsite.com/consortium; Avni et al. 2017). Moreover, the rapid development of next-generation sequencing (NGS) technologies has ignited an explosion in genome sequencing of its progenitor species and other bread wheat and durum wheat cultivars subsequently (Chapman et al. 2015; Jia et al. 2013; Ling et al. 2013; Uauy 2017a). NGS enables efficient high throughput discovery of DNA variants in wheat (Allen et al. 2011; Allen et al. 2013; van Poecke et al. 2013), and single nucleotide polymorphisms (SNP) are now the preferred landmarks for genetic analysis based on sequencing. Compared to more traditional marker systems SNP have the unique advantage of high resolution of genetic diversity and enable identification of relevant genes and even their functions (Varshney et al. 2014; Xu et al. 2017). Current SNP assay platforms including Illumina Bead Chip™, Affymetrix Gene Chip™ and Kompetitive Allele Specific PCR (KASP™, http://www.Lgcgenomics.com), have been widely adopted for mapping and marker-assisted selection (MAS) (Rasheed et al. 2017).

Another significant technique is pooling analysis or bulked segregant analysis (BSA), (Giovannoni et al. 1991; Michelmore et al. 1991), which involves the use of selected and pooled DNA samples from sets of individuals exhibiting contrasting extreme phenotypes to provide a simple and rapid way to obtain markers linked to target traits. This selective genotyping approach reduces costs and time compared to more conventional analyses

involving entire populations. It is very flexible and has been improved, optimized and extended in recent research (Zou et al. 2016). BSA combined with NGS is also being used for gene mapping and marker development in several crops (Abe et al. 2012; Garcia et al. 2016; Liu et al. 2012; Ramirez-Gonzalez et al. 2015; Singh et al. 2016; Takagi et al. 2013). BSA coupled with transcriptional profiles from RNA sequences (RNA-seq) for mapping genes of interest is a rapid and cost-efficient method based on NGS sequencing. BSR-Seq not only enables SNP discovery for markers linked to target genes but also provides patterns of differential gene expression genes between pools (Schlötterer et al. 2014). The strength of BSR-Seq for rapid gene mapping has been demonstrated in wheat (Ramirez-Gonzalez et al. 2015; Trick et al. 2012; Wang et al. 2017).

Common wheat line 92R137 (and several sister lines) was developed by the Cytogenetics Institute at Nanjing Agricultural University from a wheat × *Dasypyrum villosum* (*Haynaldia villosa*) cross and was selected for a powdery mildew resistance gene (*Pm21*) that resides in a chromosome 6A/6VS translocation (Cao et al. 2011). The lines also carry *Yr26* located on chromosome 1B (Ma et al. 2001; Wang et al. 2008) and presumably derived from a *T. durum* line that was used as a bridging parent in initial crosses. Over the past two decades 92R137 and sib lines were widely used in wheat breeding programs and varieties with *Yr26* as well as *Pm21* have been grown on more than 3.4 million hectares (Wang et al. 2008). *Pm21* is still effective against currently predominant races of the powdery mildew pathogen, but *Yr26* has been overcome by V26 *Pst* races that have increased in prevalence over the past two years (Han et al. 2015; Liu et al. 2017). Cloning *Yr26* will contribute to an understanding of the molecular mechanisms of stripe rust resistance and provide basis for wisely using this gene and other genes to achieve long lasting, and high level resistances to stripe rust.

Yr26 probably came from *Triticum turgidum* cv. γ80-1 that was used in a bridging cross in the derivation of the Nanjing material. The gene was located on chromosome arm 1BS and showed linkage to SSR markers *Xgwm11*, *Xgwm18* and *Xgwm413* (Ma et al. 2001). Wang et al. (2008) subsequently localized it to deletion bin C-1BL6-0-0.32 between markers *WE173* and *Xbarc181* with genetic distances of 1.4 cM and 6.7 cM. The genomic region was more recently narrowed to a 0.25 cM interval flanked by EST-STS markers *CON-4* and *CON-12* (Zhang et al. 2013). However, due to the notoriously low levels of genetic recombination in the centromeric regions of wheat chromosomes the actual physical distance is more than 10 Mb and existing markers are not suitable to screen BACs for construction of a physical map. Hence, closer markers are needed for further fine mapping and eventual map-based cloning of

Yr26. Genes named as *Yr24, YrCH42* and *YrGn22* derived from durum wheat sources and independently mapped to the same location on chromosome 1B (McIntosh and Lagudah 2000; Li et al. 2006; Li et al. 2016, respectively) are likely the same as *Yr26* (McIntosh et al. 2018).

The purpose of the present work was to fine-map and eventually clone *Yr26* in the knowledge of its near-centromeric location. We therefore attempted to combine super-pool with high-throughput genotyping and sequencing. Three recombinant inbred line (RIL) populations were developed to build three pairs of super-pools by phenotypic selection of individuals with extremes of rust resistance and susceptibility and genotyping by SNP arrays. One of the RIL populations was subjected to BSA and RNA-Seq. We subsequently i) analysed the sequencing data of the extreme pools to identify SNPs linked to *Yr26*; ii) performed KASP assays to validate the linked SNPs and fine map *Yr26*; iii) identified positional candidate genes in the region linked to rust resistance; and iv) refined unique SNPs linked to *Yr26* and developed robust and breeder-friendly KASP markers based on haplotype analysis.

Materials and methods

Population establishment and phenotyping

The parental lines used in this study were the susceptible lines Yangmai 5 (YM5), Yangmai 158 (Y158) and Avocet S, and resistant lines 92R137 and Y158/6*Yr26. Yangmai 5 and Yangmai 158 are elite wheat cultivars but both are highly susceptible to stripe rust in China. Avocet S (AvS) is an Australian spring wheat selection that is highly susceptible to most *Pst* races in China. A pair of near-isogenic lines (NIL-R/NIL-S) differing in presence and absence of *Yr26* selected from a Y158 backcross line (BC_6) was kindly provided by Prof. Peidu Chen, Cytogenetics Institute, Nanjing Agricultural University. These materials were used to construct genetic mapping populations and identify SNPs linked with *Yr26*. The first population consisting of 2,341 F_2 individuals and 156 $F_{2:7}$ RILs was developed from the cross AvS/92R137 (Zhang et al., 2013). The second was an $F_{2:8}$ population of 273 RILs derived from the cross YM5/92R137 (Wang et al., 2008). The third population containing 13,128 F_2 plants, 240 $F_{2:3}$ lines and 1,034 $F_{2:6}$ RILs was constructed by a cross between near-isogenic lines NIL-S and NIL-R (Fig. 1A). F_2, $F_{2:3}$, $F_{2:6}$, $F_{2:7}$, $F_{2:8}$ populations and their parents were inoculated with *Pst* race CYR32, which is avirulent for *Yr26* and virulent to YM5, Y158 and AvS (Zeng et al. 2014), in the greenhouse during 2012-2015. In seedling tests 15−20 plants of parents and their corresponding progenies were grown in 12 × 12 × 12 cm pots. The inoculation procedure was as described by Wu et al. (2016). Infection type (IT) data were

recorded based on a 0–9 scale (0–6 considered is resistant, 7–9 considered is susceptible) as described by Line and Qayoum (1992).

DNA extraction and SNP array genotyping

Wheat leaf samples from F_2 plants, RILs, parental lines, and tested entries were harvested to extract genomic DNA following the sodium dodecyl sulfonate (SDS) method (Song et al. 1994). Based on phenotypic evaluations, equivalent DNA mixtures from homozygous resistant (IT 1) and homozygous susceptible (IT 9) RILs in different hybrid populations were constituted as resistant and susceptible pools (Fig. 1B). The DNA pools along with parental lines were genotyped by both the 90K (Wang et al. 2014) and 660K (Jizeng Jia, personal communication) SNP arrays at CapitalBio Corporation (Beijing; http://www.capitalbio.com). SNP genotype calling and clustering was processed with the Illumina Genome Studio Polyploid Clustering v1.0 and Affymetrix Genotyping Console™ (GTC) softwares, respectively. SNP filtering criteria were as follows: monomorphic and poor quality SNP loci with more than 10% missing values, ambiguous SNP calling, or minor allele frequencies below 5%, were excluded from further analysis (Fig. 1D). Polymorphic SNPs detected across all three pairs of contrasting pools and corresponding parents were assumed to be linked to *Yr26*, and homozygous genotypes were localized to chromosomes based on the high-density 90K genetic map (Wang et al. 2014) and 660K genetic map (Jizeng Jia, personal communication).

RNA sample preparation and sequencing

Seedling leaf samples of F_8 RILs from YM5/92R137 were collected to construct two extreme pools at the two-leaf stage. Each pool contained 3 replicates (+*Yr26* or -*Yr26*) (Fig. 1B). Total RNA was isolated using the TRIzol protocol (Invitrogen, Carlsbad, CA). After quality test, a single RNA library was constructed for each sample (i.e., 6 libraries in total) and the library preparations were sequenced in an Illumina HiSeq2500 paired-end (PE) lane (2 × 125 bp).

Trimming and alignment of sequencing reads

Prior to alignment, each raw read was assessed for quality control by Trimmomatic v0.32 software (Bolger et al. 2014). Bases with PHRED quality values < 20 (out of 40) (Ewing and Green 1998), i.e., error rates of ≤ 1%, were removed by our trimming pipeline. Each read was examined in two phases. In the first phase reads were scanned starting at each end and nucleotides with quality values lower than the threshold were removed. The remaining

nucleotides were then scanned using overlapping windows of 10 bp and sequences beyond the last window with average quality values less than the specified threshold were truncated. The trimming parameters used were in reference to the trimming software Lucy (Li and Chou 2004). The trimmed reads were aligned to the hexaploid wheat cv. Chinese Spring genome assembly using GSNAP software (Wu and Nacu 2010). High-quality reads were also aligned to the wild emmer wheat cv. Zavitan genome reference sequence (Avni et al., 2017).

SNP discovery and filtering by allele difference ratio (ADR)

Confident and unique alignments of reads from each of the 6 samples to the genome assembly were extracted for SNP discovery. Variant discovery was performed between resistant and susceptible samples using homozygous and heterozygous variant discovery models, which were modified based on the method described by Liu et al. (2012). The first model is based on the assumption that a polymorphism will be homozygous within each pool (e.g., A/A within one pool and T/T within the other pool). The second model does not make this assumption. Homozygous SNP criteria were: 1) the first and last 3 aligned bases of each read were discarded, 2) each polymorphic base must have at least a PHRED base quality value of ≥20 (≤1% error rate), 3) at least 5 unique reads must support the base-pair call, and 4) the sum of reads of the two most common alleles must account for at least 80% of all aligned reads covering that nucleotide position. The heterozygous SNPs criteria were: 1) the first and last 3 aligned bases of each read were discarded, 2) each polymorphic base must have at least a PHRED base quality value of ≥20 (≤1% error rate), 3) at least 5 unique reads must support the base-pair call, 4) the two most common alleles must be supported by at least 30% of all aligned reads covering that position, and 5) the sum of reads of the two most common alleles must account for at least 80% of all aligned reads covering that nucleotide position. Polymorphisms called using the homozygous model are expected to have a lower probability of false-positive calls, but given the non-inbred nature of the mapping population both models were used in this analysis.

For each SNP, a Pearson's χ^2 test was performed using read counts of two alleles (Ref and Alt) in both pools. The χ^2 test was designed to test the null hypothesis that the tested SNP locus is not associated with a phenotypic difference. A multiple testing adjustment *p*-value, which is called a *q*-value, was obtained to control false discovery (Benjamini and Hochberg 1995). A SNP with a *q*-value ≤0.01 and exhibiting different major alleles in both pools and the product of major allele coverage ratio in both pools ≥90%, as well as the minor allele in each

pool having ≤3 reads were defined as linked SNPs (Fig. 1C). The allele difference ratio (ADR) of each linked SNP was calculated based on the formula:

$$\text{ADR} = \frac{\text{Reads of susceptible [major allele] in S pools}}{\text{Total reads of this site in S pools}} \times \frac{\text{Reads of resistant [major allele] in R pools}}{\text{Total reads of this site in R pools}}$$

SNP variant calling and filtering by bulk frequency ratios

Resistant and susceptible pooled reads were compared to discover SNP variants using DiscoSNP software (Uricaru et al. 2015). DiscoSNP also mapped the SNPs to the wild emmer reference genome (Avni et al., 2017). SNPs with bulk frequency ratio (BFR) >6 (Ramirez-Gonzalez et al. 2014) and with >50 read coverage in each pool were selected for further analyses.

KASP assays design

KASP assays were designed to validate the putative SNPs between R&S pools and generate a genetic map of the *Yr26* locus. Initial KASP markers were derived from the 90K and 660K SNP arrays and BSR-Seq. SNPs were converted into KASP markers using a similar approach to that described in (Ramirez-Gonzalez et al. 2015). Only chromosome-specific KASP markers were selected according to their physical positions compared to previous flanking markers linked to *Yr26* in the wheat reference genome. The procedure of selective KASP-SNP assays was described in Wu et al. (2018a). To saturate the genetic map further marker development was based on regions surrounding flanking KASP markers linked to *Yr26.*

Recombinants and high-density genetic map

A recombination screen was carried out using F_2 plants and RILs from the three crosses. Initial SSR markers, viz., *Xgwm11* and *Xwmc419* were selected from a previous study as flanking markers to identify recombinant chromosomes. KASP markers *WRS435* and *WRS312* were subsequently used to screen recombinants. Chi squared (χ^2) tests for goodness of fit were performed to determine agreement of observed segregation ratios with theoretically expected ratios. Linkage analysis and genetic map construction was established using JoinMap version 4.0 (Van Ooijen 2006) with default parameters. Linkage to *Yr26* was determined using the Kosambi mapping function (Kosambi 1943) and a LOD score of 3.0 as a threshold. The genetic linkage map was drawn with the software Mapchart V2.3 (Voorrips 2002).

Comparative genomics analysis

The sequences of polymorphic KASP-SNPs located in the genetic map were blasted on the

Ensemble Plants website (http://plants.ensembl.org/index.html) to find TGACv1 gene models. The corresponding wheat gene sequences to which the SNPs best hit were also analyzed by BLASTn against coding sequences (CDSs) of rice, barley, *B. distachyon*, maize and sorghum to identify orthologous gene pairs. All CDSs were downloaded from the Ensemble Plants website. An expectation value (*E*) of 1e-10 was used as the significance threshold. Synteny analyses with the common wheat, rice, barley, *Brachypodium*, maize and sorghum genomes were performed based on the SNP orders in the genetic map and on the corresponding CDSs in the genome sequences of the respective species.

Physical mapping and prediction of candidate genes

To obtain physical positions of polymorphic SNPs, the SNP probes were aligned with respect to the newly released Chinese Spring sequence through a BLAST search (IWGSC RefSeq v1.0). We calculated the number of SNPs per megabase (Mb) by perl script. In addition, the SNP probes were also blasted to the Zavitan genome sequence (WEWseq v1.0). Recently, IWGSC RefSeq v1.0 and WEWSeq v1.0 with gene annotations became available on websites https://wheat-urgi.versailles.inra.fr/Seq-Repository/Annotations and http://wewseq.wixsite.com/consortium, respectively. Annotated genes in the target region were extracted for analyzing genes involved in plant disease resistance.

Phylogenetic analysis and validation of KASP markers in marker assisted selection

SNP array genotyping of a set of 384 Chinese elite winter wheat cultivars was performed for phylogenetic analysis using the Wheat660 array. SNPs extracted from a defined genomic region were used to infer a population structure by the Bayesian model-based clustering method in the software Structure v 2.3.4 (Pritchard et al. 2000). The parameter settings of STRUCTURE were: admixture model of population structure, ten replicates at each K value and 20,000 length of burn-in period followed by 100,000 Markov Chain Monte Carlo (MCMC) iterations, hypothetical subpopulations *K* settings from 2 to 7. The output data were collated by Structure Harvester (Earl and VonHoldt 2012) to detect the optimum number of groups to represent the population (Evanno et al. 2005). A phylogenetic neighbor-joining (NJ) tree was constructed using MEGA 5.0 (Tamura et al. 2011). Another group of 1,322 wheat accessions from the China Agriculture Research System (CARS) wheat germplasm collection was used to validate polymorphisms of KASP markers flanking the resistance gene (Fig. S1).

Results

Inheritance of *Yr26* and construction of pools

As shown in Fig. 1A and Table S1, Yangmai 5, Yangmai 158, Avocet S and NIL-S were susceptible (IT = 9), whereas 92R137 and NIL-R were resistant (IT = 1). Segregation of rust resistance in all RIL populations and $F_{2:3}$ lines complied with that expected for a single locus was and as expected from previous results. Equivalent DNA mixtures from 102 homozygous resistant (IT 1) and 102 homozygous susceptible (IT 9) F_6 RILs were made as resistant and susceptible pools. Two other pairs of pools were prepared from 115 YM5/92R137 F_8 plants and 55 AvS/92R137 F_7 plants (Fig. 1B). For RNA-pools, each pool (±*Yr26*) contained three replicates and each replicate comprised 92 homozygous individuals selected from 273 F_8 RILs from cross YM5/92R137 (Fig. 1B).

Different sets of SNPs identified in pool from SNP arrays and BSR-Seq

Three hundred and thirty two SNPs were polymorphic between the three pairs DNA pools following genotyping by the 90K SNP array (Fig. 2A); 252 of them were placed on chromosome (chr) 1B and 57 SNPs were not (Fig. 2B). Based on the number of SNPs per Mb on 1B, most of the SNPs were within a chr 1B interval of 300-330 Mb (Fig. 2C, Table S2A). Approximately 2,000 SNPs showed polymorphisms in three pools with the 660K SNP array (Fig. 2D). A total of 1,745 SNPs were located on chr 1B (Fig. 2E) and most of the linked SNPs were within an interval 297 to 343 Mb (Fig. 2F, Table S2A). These overlapping regions were most likely linked with the resistance locus.

RNA-Seq generated 60.5 Gb of raw data and after quality control, 59.6 Gb (98.4%) of clean data remained. Approximately 80.3% of the trimmed reads from each sample could be aligned to at least one position and 65.8% of them could be uniquely aligned to the reference genome (Table S3, Fig. S2). After using the coordinates of uniquely aligned reads from BSR-Seq to discover SNP variants, a total 462,943 putative SNP loci were identified, of which 338,731 (73.2%) could be positioned on a single chromosome (Table S4, Fig. 3A). After triple filtering criteria (Fisher's Exact Test *p*-value was <0.01; multiple testing adjustment *q*-value <0.01; and ADR ≥0.90 (Fig. 3B)) 530 of 602 (88.0%) linked SNPs were blasted to chr 1B (Table S5, Fig. 3C), accounting for most of the total linked SNPs, indicating the causal locus was consistent with *Yr26* in previous studies. After plotting each linked SNP (x-axis) against its ADR (y-axis) based on their physical positions, the linked SNPs were enriched in three conserved intervals of chr 1B, i.e., 120-195 Mb, 277-349 Mb and 408-488 Mb, respectively. These results combined with the SNP array data indicated that the

overlapping interval likely harbored *Yr26*.

To identify and verify candidates for SNP variation we screened pools with bulk frequency ratios (BFR) >6 for polymorphism using the wild emmer wheat genome as a reference (Avni et al., 2017; Ramirez-Gonzalez et al. 2014; Trick et al. 2012). Two hundred and seventy three loci were identified and three sets of clusters with high BFR were near the chr 1B centromere (Fig. S3). This was consistent with the result obtained by ADR. A list of SNPs with BFR >6 is presented in Table S6.

To further refine this location wheat EST sequences corresponding to genetic markers linked to *Yr26* in Wang et al. (2008) and Zhang et al. (2013) were used to blast the IWGSC RefSeq v1.0. The order of genetic markers showed good synteny in physical position (Fig. 3C, Table S2A). These results were consistent with the previous studies that located *Yr26* in the long arm of chr 1B, near the centromere (Table S2B, (Wang et al., 2008; Zhang et al., 2013)). Based on all of the above results, *Yr26* was located in the interval 300-330 Mb.

Molecular mapping of *Yr26* using $F_{2:3}$ lines

To validate putative SNPs and generate a genetic map of the *Yr26* locus, we designed KASP assays. Initially, 125 SNPs from the SNP arrays and BSR-Seq were selected for conversion to KASP markers and KASP assays were performed on the parents and pools to confirm their specificity before being genotyped on the 240 $F_{2:3}$ lines from the cross NIL-S/NIL-R. Four SSR and STS markers (*Xgwm11*, *WE173*, *STS-BQ33* and *Xbarc181*) linked with *Yr26* were also used to build a skeleton genetic map. A linkage group within chr 1B (12.3 cM in length) was constructed using data for the 4 SSR and 32 KASP markers from the $F_{2:3}$ lines. The *Yr26* gene was preliminarily located between KASP markers *WRS270* and *WRS290* in an interval of 2.7 cM (Fig. 3D, S4A, B), corresponding to the physical interval 300.0-343.6 Mb.

Confirming the *Yr26* interval in different genetic populations

Representing different genetic backgrounds 1,034 $F_{2:6}$ RILs (NIL-S/NIL-R), 273 $F_{2:8}$ RILs (YM5/92R137) and 156 $F_{2:7}$ RILs (AvS/92R137) were used to validate the *Yr26* interval. Additionally, 45 of 110 KASP markers within the 300.0-343.6 Mb interval showing polymorphisms between parents and pools were used to construct the genetic maps. As the marker orders for these maps were similar all maps were integrated into a single consensus genetic map and *Yr26* was located between KASP markers *WRS435* and *WRS312* in an interval of 0.04 cM (Fig. 4B, S4C, D), corresponding to the physical interval 325.3-329.5 Mb

(Table S2A). Sequences of the KASP markers are listed in Table S7.

Physical mapping and comparative genomic analysis

After determining the physical positions of polymorphic or putative SNPs, the order of KASP markers in the genetic map was compared with that in IWGSC RefSeq v1.0; the results showed good conformity. To assess the relationship between wheat and rice, barley, *Brachypodium*, maize and sorghum, the relevant wheat gene sequences were selected to identify orthologous genes in comparative genomic regions in the five graminaceous species. As shown in Table S8 399 SNPs between *WRS270* and *WRS290* corresponded to the 111 CDSs of wheat, 92 CDSs of rice and 87 CDSs of barley, *Brachypodium*, maize and sorghum. For simplification only 49 polymorphic KASP-SNP markers located between *WRS270* and *WRS290* are listed in Fig. 5. In general, the 49 KASP markers representing the region in wheat showed good collinearity with rice chr 10 (Os10), barley chr 1 (Hv1), *Brachypodium* chr 3 (Bd3), maize chr 1 (Zea1), and sorghum chr 1 (Sb1), but some differences including reverse orders and small syntenous blocks were observed. For example, an inverted segment was present in rice (from *OS10G0495600* to *OS10G0488800*), *Brachypodium* (from *BRADI3G29917* to *BRADI3G29700*), and sorghum (from *SORBI_001G212300* to *SORBI_001G208700*) compared to wheat, barley and maize. The syntenous blocks were more fragmented and scattered indicating the evolutionary complexity of graminaceous plant genomes. This information will be valuable for further high-resolution mapping and map-based cloning of *Yr26*.

Recombinants, high-density genetic map, and candidate genes

We used flanking markers *WRS435* and *WRS312* to genotype 13,128 F_2 plants from the NIL-S/NIL-R cross and 2,341 F_2 plants from the cross AvS/92R137. Thirty-six plants carrying recombination events were identified and used to saturate the *Yr26* vicinity. Once the locus was mapped more precisely, we developed closer flanking KASP markers based on their physical positions in the target region and reduced the number of plants carrying recombination events in the critical region to five, each of which was self-pollinated and the corresponding F_3 families were tested with *Pst* race CYR32 to verify their phenotypes. *Yr26* was finally located between the KASP markers *CM1461* and *WRS467* in an interval of 0.003 cM (Fig. 4C), corresponding to a physical interval of 488 kb. This region in the hexaploid wheat reference genome contained no typical resistance gene, but there were three genes that have been implicated in disease response, coding for an E3 ubiquitin ligase, an ABC

transporter, and a WAT-related protein (Table S9A, Fig. 4D). Similarly, there was no typical resistance gene in the wild emmer reference genome (Table S9B, Fig. 4E).

Haplotype variation of the *Yr26* region

The wheat 660K SNP array provides an extremely rich avenue of inquiry of natural variation in germplasm and permits identification of further SNPs tightly linked to target genes. The genomic segment of chr 1B containing 48 SNPs was extracted from 384 Chinese cultivars with 660K SNP genotyping data to observe haplotype and phylogenetic clustering (Table S10). The phylogenetic population structure based on the *Yr26* region using the SNP data revealed four distinct clusters. As shown in Fig. 6, the branch for the *Yr26* region of line 92R137 (*Yr26*), AvSYr24 (*Yr24*), Chuanmai 42 (*YrCH42*), Guinong 22 (*YrGn22*) and their derivatives is considerably differentiated from the other branches for the corresponding regions in ancestral donors of *Yr9, Yr10, Yr15, Yr29/Lr46, Yr64, Yr65* and lines without *Yr26/Yr24/YrCH42/YrGn22*.

KASP marker validation among the CARS wheat germplasm collection

Although *Yr26* was fine mapped to a considerably small genomic region, the linked SNPs distinguished *Yr26* from other *Yr* loci in chromosome 1B. In order to evaluate the robustness of markers linked to *Yr26*, 10 KASP-SNP markers covering a 1.34 Mb genomic region encompassing the *Yr26* locus were extracted to generate SNP genotypic data from the group comprising 1,322 accessions (Table S11). Markers *CM1461, CM501* and *WRS467* clearly differentiated cultivars harbouring *Yr26/Yr24/YrCH42/YrGn22* from others with >97%, >99% and >98% success rates, respectively (Fig. S4E, F, G), indicating that they could be used to detect *Yr26*. Moreover, the combination of *CM1461, CM501* and *WRS467* appeared to be most predictive of *Yr26*, based on panels of varieties analyzed in previous studies (Han et al. 2010, 2015; Zeng et al. 2014; Wu et al. 2016, 2018b; Li et al. 2017).

Discussion

Super pooling analysis with high-throughput sequencing and genotyping provided a reliable result for gene localization

Conventional gene/QTL mapping requires phenotyping and genotyping of every individual in a mapping population and is a time-consuming, laborious, and costly process. The availability of next generation sequencing (NGS) technologies and optimized reference genome sequences now available for wheat have improved the efficiency of genotyping and

generating new ways to accelerate genetic analysis and shorten the breeding cycle time. However, accurate, single sequencing of a small population can generate a huge amount of redundant data caused by 'noisy' allele frequency estimates that are not beneficial to further analysis especially for hexaploid wheat. Combining high-throughput NGS technology with a pooling strategy can eliminate biological variability and generate more reproducible results especially when the pool size is increased (Schlötterer et al., 2014; Xu et al., 2017). In this study, the stripe rust phenotypes were evaluated on RIL populations to ensure accuracy. Then we demonstrated that BSR-Seq coupled with three super-pool comparisons successfully identified SNPs differentiating resistance and susceptibility to stripe rust and allowed rapid discovery of the target region. Moreover, paired sets of three super DNA pools from different genetic populations coupled with SNP arrays assisted the filtering of superfluous data and allowed greater focus on the target region. They accurately identified SNPs linked to *Yr26*, which in turn enabled development of robust markers.

Comparing BSR-Seq with SNP arrays

Recent development of methods for SNP discovery and detection in polyploid species provides an ability to accelerate fine-mapping and cloning of genes in common wheat. Trick et al. (2012) made a first attempt at mapping the cloned high grain protein content gene *GPC-B1* in tetraploid wheat by BSR-Seq. Subsequently Ramirez-Gonzalez et al. (2015) mapped stripe rust gene *Yr15* in hexaploid wheat by the same method. Wang et al. (2017) also positioned a ~4 Mb candidate region associated with resistance to stripe rust in a similar way. BSR-Seq generally provides a rough location of a particular gene/QTL, and additional genetic analysis is needed to refine the location and narrow the chromosomal interval. In this study *Yr26* was initially mapped to a ~230 Mb candidate region (i.e., 120-195 Mb, 277-349 Mb and 408-488 Mb, respectively) on chr 1B that was consistent with the result of BFR for three main reasons. First, recombination frequencies vary along the length of the chromosome and recombination frequency is significantly restricted in pericentromeric regions compared to distal regions. *Yr26* is located near the centromere where recombination is reduced or absent (Philippe et al., 2013). This leads to a large physical interval that makes it difficult to refine the target region. Second, a part of the differential expression levels of alleles at a single locus may be influenced both spatially and temporally in different tissues, especially for induced gene expression and these kinds of SNPs cannot be identified in comparisons between non-inoculated resistant lines and susceptible lines (Schlötterer et al., 2014). Last, positioning results were completely dependent on the integrity of the reference genome sequence. In

addition, genetic backgrounds between the cross parents and Chinese Spring also affect mapping outcomes. In contrast, SNP arrays are based on DNA sequencing by hybridization. SNP loci from SNP arrays are genome-specific and can discriminate subgroups among the A, B, and D genomes in hexaploid accessions. Moreover, they possess a high resolution in wheat chromosomes and cover whole-genomes including exons and introns. They are not affected by gene expression. However, it is difficult for SNP arrays to distinguish heterozygous loci and multiple copy genes. Hence, it is a priority that BSR-Seq and SNP arrays are complementary and allow the advantages of each to assist rapid and efficient gene isolation.

Genetic and physical mapping of *Yr26*

In the case of *Yr26*, chromosomal localization was not amenable for mapping due to its proximal location on the long arm of chr 1B. Indeed, infrequent recombinants among 2,341 F_2 plants observed by Zhang et al. (2013) were not sufficient to delimit *Yr26* to a 0.1 cM region. Mutant R gene enrichment sequencing (MutRenSeq) was considered to enable rapid identification of genes responsible for resistance without positional fine mapping (Steuernagel et al. 2016). However, we failed isolate *Yr26* in seven susceptible EMS-induced mutants by this method (data not shown; Evans Lagudah, personal communication). This negative result is consistent with the putative gene contents (ABC transporter, WAT protein, and E3 ligase). Consequently, we pursued map-based cloning by expanding the number F_2 plants to approximately 15,000 (>30,000 gametes) and identified key recombinants that narrowed the target region to 488 kb.

Candidate gene analysis

To date, 10 race-specific rust-resistance genes have been cloned in wheat, including *Sr22*, *Sr33*, *Sr35*, *Sr45*, *Sr50* (resistance to stem rust), *Lr1*, *Lr10*, *Lr21*, *Lr22* (resistance to leaf rust), and *Yr10* (resistance to stripe rust), and all involve nucleotide-binding and leucine-rich repeat (NLR) receptor proteins (Ellis et al. 2014; Krattinger and Keller 2016; Periyannan et al. 2017; Thind et al. 2017). However, barley stem rust resistance gene *Rpg1* encodes a receptor-like serine/threonine kinase with tandem kinase domains (Brueggeman et al. 2002). *Yr26* displays all-stage resistance (ASR) and likely contains an NLR structure. However, a mass of repetitive sequences but no typical LRR gene is present in the target region based on Chinese Spring IWGSC RefSeqv1.0 and Zavitan WEWSeq v1.0 information. Nevertheless we cannot rule out the possibility that an NBS-LRR gene may be missing in the *Yr26* candidate regions

in Chinese Spring and Zavitan. Efforts are now underway to delimit the physical region harboring *Yr26* by screening BACs from line 92R137. This will further increase marker discovery and enable identification of further candidate genes in the region.

Haplotype analysis and marker-assisted selection

Wheat chr 1B is rich in resistance genes. To date, twelve *Yr* genes (or putative genes) have been reported on chr 1B; namely, *Yr9*, *Yr10*, *Yr15*, *Yr24/26/CH42/YrGn22*, *Yr29/Lr46*, *Yr64*, *Yr65*, *YrH52*, *YrH62*, *YrL693*, *YrSM139*, *YrAlp* and *YrExp1* (Cheng et al. 2014; Huang et al. 2014; Lin and Chen 2007, 2008; McIntosh et al. 2018; Nagy et al. 2003; Peng et al. 1999; Sun et al. 1997; Wang et al. 2002; William et al. 2003; Zhang et al. 2016) (Cheng et al., 2014; Huang et al., 2014; Lin and Chen, 2007, 2008; Nagy et al., 2003; Peng et al., 1999; Sun et al., 1997; Wang et al., 2002; William et al., 2003; Wu et al., 2018b; Zhang et al., 2016). All of the above genes except *Yr29/Lr46* and *YrH62* confer all-stage resistance (ASR) or race-specific resistance and most of them are positioned in the pericentromeric region. Haplotype variation of the *Yr24/26/CH42/YrGn22* region between 92R137, AvSYr24, Chuanmai 42 and Guinong 22 and materials carrying other *Yr* genes indicated no commonality between the *Yr24*, *26*, *CH42*, *YrGn22* group locus/loci and other known *Yr* genes mapped in overlapping or adjacent regions.

Marker-assisted selection provides a targeted approach to detect and track *Yr* genes in breeding programs and will increase as markers become gene-based and as high-throughput genotyping platforms continue to become more affordable. However, DNA-based markers generated by these technologies must be well validated before they can be used in marker-assisted breeding. Marker validation of a gene of interest can be verified by introgression to diverse genetic backgrounds (or advanced breeding lines) that can then be phenotypically evaluated in multiple environments (Brown 2015; Chen 2013; St. Clair 2010). Accordingly, the efficacy of markers is directly relevant to how closely linked the marker is to the targeted gene/QTL and the stability of expression of the responsible gene in different genetic backgrounds. In this study, many SNPs linked to *Yr26* were obtained by BSR-Seq and SNP arrays. Different genetic backgrounds of RILs were involved in mapping the *Yr26* region and provided confirmation of the mapping results. However, we failed to convert most of the SNPs to KASP markers. The most serious causes for the low the KASP conversion rates in wheat is the high level of repetitive sequences and sequence similarities among the respective AA, BB and DD genomes that give rise to short fragment assembly errors (Uauy et al. 2017b).

Thus, there are many false-positive SNPs between extreme pools and hence a large proportion of KASP markers fail to distinguish between the genomes. Another reason is genetic background differences between 92R137 (donor parent of RNA-Seq) and Chinese Spring (the hexaploid wheat reference genome), and especially between 92R137 and Zavitan (the wild emmer wheat reference genome). In our experiments, the SNPs from ADR based on IWGSC RefSeq v1.0 were more easily converted to KASP than the SNPs from BFR based on WEWSeq v.1.0. Identification of major haplotypes associated with *Yr26* by SNP-based haplotype analysis provided a reliable approach for developing robust, breeder-friendly KASP markers. From haplotype analysis we identified two flanking markers and one associated marker specific for the stripe rust resistance locus *Yr26*. These tightly linked molecular markers validated in a set of 1,322 wheat accessions not only can be used in marker-assisted selection of *Yr26* in breeding programs but will be crucial for screening BACs and cloning the causal gene.

Conclusion

This work demonstrated that RNA-Seq and SNP arrays supported by super pooling analysis can rapidly and reliably identify trait-associated candidate genomic regions in hexaploid wheat. Through a series of comprehensive filtering criteria, masses of redundant data were eliminated and putative SNPs harboring the target region were preferably selected. Using these SNPs, *Yr26* was fine mapped to a smaller interval providing a solid foundation for map-based cloning. Moreover, we successfully exploited a method for validating robust, breeder-friendly and cost-effective KASP markers by SNP-based haplotype analysis. These findings provide more reliable, if not 'perfect', markers for MAS of *Yr26*. This integration strategy can be applied to generate many markers closely linked to target genes/QTL for a trait of interest in wheat and other polyploid species.

Acknowledgments

The authors are grateful to Prof. R. A. McIntosh, Plant Breeding Institute, University of Sydney, for critical review of this manuscript; Prof. Peidu Chen and Prof. Aizhong Cao, Cytogenetics Institute, Nanjing Agricultural University, for providing *Yr26* germplasms and genetic populations. This study was financially supported by International S&T Cooperation Program of China (2015DFG32340), National Natural Science Foundation of China (31371924), the National Key Research and Development Program of China (Grant No. 2016YFE0108600), the earmarked funds for Modern Agro-industry Technology Research

System (No. CARS-3-1-11) and National Natural Science Foundation for Young Scientists of China (Grant 31701421).

References

Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, et al. (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. Nat Biotechnol 30:174-178

Allen AM, Barker GLA, Berry ST, Coghill JA, Gwilliam R, et al. (2011) Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (*Triticum aestivum* L.). Plant Biotechnol J 9:1086-1099

Allen AM, Barker GLA, Wilkinson P, Burridge A, Winfield M, et al. (2013) Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum* L.). Plant Biotechnol J 11:279-295

Avni R, Nave M, Barad O, Baruch K, Twardziok SO, et al. (2017) Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science 357:93-97

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B 57:289-300

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120

Brown JK (2015) Durable resistance of crops to disease: a Darwinian perspective. Annu Rev Phytopathol 53:513-539

Brueggeman R, Rostoks N, Kudrna D, Kilian A, Han F, et al. (2002) The barley stem rust-resistance gene *Rpg1* is a novel disease-resistance gene with homology to receptor kinases. Proc Natl Acad Sci USA 99:9328-9333

Cao A, Xing L, Wang X, Yang X, Wang W, et al. (2011) Serine/threonine kinase gene *Stpk-V*, a key member of powdery mildew resistance gene *Pm21*, confers powdery mildew resistance in wheat. Proc Natl Acad Sci USA 108:7727-7732

Chapman JA, Mascher M, Buluc AN, Barry K, Georganas E, et al. (2015) A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. Genome Biol 16:26

Chen WQ, Wu LR, Liu TG, Xu SC, Jin SL, et al. (2009) Race dynamics, diversity, and virulence evolution in *Puccinia striiformis* f. sp. *tritici*, the causal agent of wheat stripe rust in China from 2003 to 2007. Plant Dis 93:1093-1101

Chen X (2013) High-temperature adult-plant resistance, key for sustainable control of stripe rust. Amer J Plant Sci 04:608-627

Chen XM (2005) Epidemiology and control of stripe rust [*Puccinia striiformis* f. sp. *tritici*] on wheat. Can J Plant Pathol 27:314-337

Chen XM (2014) Integration of cultivar resistance and fungicide application for control of wheat stripe rust. Can J Plant Pathol 36:311-326

Cheng P, Xu LS, Wang MN, See DR, Chen XM (2014) Molecular mapping of genes *Yr64*

and *Yr65* for stripe rust resistance in hexaploid derivatives of durum wheat accessions PI 331260 and PI 480016. Theor Appl Genet 127:2267-2277

Clavijo BJ, Venturini L, Schudoma C, Accinelli GG, Kaithakottil G, et al. (2017) An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. Genome Res 27:885-896

Earl DA, VonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359-361

Ellis JG, Lagudah ES, Spielmeyer W, Dodds PN (2014) The past, present and future of breeding rust resistant wheat. Front Plant Sci 5

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611-2620

Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8:186-194

Fu D, Uauy C, Distelfeld A, Blechl A, Epstein L, et al. (2009) A kinase-START gene confers temperature-dependent resistance to wheat stripe rust. Science 323:1357-1360

Garcia V, Bres C, Just D, Fernandez L, Tai FW, et al. (2016) Rapid identification of causal mutations in tomato EMS populations via mapping-by-sequencing. Nat Protoc 11:2401-2418

Giovannoni JJ, Wing RA, Ganal MW, Tanksley SD (1991) Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. Nucleic Acids Res 19:6553-6568

Han D, Wang Q, Zhang L, Wei G, Zeng Q, et al. (2010) Evaluation of resistance of current wheat cultivars to stripe rust in Northwest China, North China and the Middle and Lower Reaches of Changjiang River epidemic area. Sci Agric Sin 43:2889-2896

Han DJ, Wang QL, Chen XM, Zeng QD, Wu JH, et al. (2015) Emerging *Yr26*-virulent races of *Puccinia striiformis* f. sp. *tritici* are threatening wheat production in the Sichuan Basin, China. Plant Dis 99:754-760

Hovmøller MS, Walter S, Justesen AF (2010) Escalating threat of wheat rusts. Science 329:369

Huang Q, Li X, Chen WQ, Xiang ZP, Zhong SF, et al. (2014) Genetic mapping of a putative *Thinopyrum intermedium*-derived stripe rust resistance gene on wheat chromosome 1B. Theor Appl Genet 127:843-853

Jia J, Zhao S, Kong X, Li Y, Zhao G, et al. (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. Nature 496:91-95

Kosambi DD (1943) The estimation of map distances from recombination values. Ann Eugen 12:172-175

Krattinger SG, Keller B (2016) Molecular genetics and evolution of disease resistance in cereals. New Phytol 212:320-332

Krattinger SG, Lagudah ES, Spielmeyer W, Singh RP, Huerta-Espino J, et al. (2009) A

putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. Science 323:1360-1363

Li B, Xu Q, Yang Y, Wang Q, Zeng Q, et al. (2017) Stripe rust resistance and genes in Chongqing wheat cultivars and lines. Sci Agric Sin 50:413-425

Li GQ, Li ZF, Yang WY, Zhang Y, He ZH, et al. (2006) Molecular mapping of stripe rust resistance gene *YrCH42* in Chinese wheat cultivar Chuanmai 42 and its allelism with *Yr24* and *Yr26*. Theor Appl Genet 112:1434-1440

Li Q, Ma D, Li Q, Fan Y, Shen X, et al. (2016) Genetic analysis and molecular mapping of a stripe rust resistance gene in Chinese wheat differential Guinong 22. J Phytopathology 164:476-484

Li S, Chou HH (2004) LUCY2: an interactive DNA sequence quality trimming and vector removal tool. Bioinformatics 20:2865-2866

Li ZQ, Zeng SM (eds) (2002) Wheat Rust in China. China Agriculture Press, Beijing

Lin F, Chen XM (2007) Genetics and molecular mapping of genes for race-specific all-stage resistance and non-race-specific high-temperature adult-plant resistance to stripe rust in spring wheat cultivar Alpowa. Theor Appl Genet 114:1277-1287

Lin F, Chen XM (2008) Molecular mapping of genes for race-specific overall resistance to stripe rust in wheat cultivar Express. Theor Appl Genet 116:797-806

Line RF, Qayoum A (1992) Virulence, aggressiveness, evolution, and distribution of races of *Puccinia striiformis* (the cause of stripe rust of wheat) in North America 1968-1987. US Department of Agriculture Technical Bulletin No. 1788, pp 74

Ling HQ, Zhao S, Liu D, Wang J, Sun H, et al. (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. Nature 496:87-90

Liu B, Liu T, Zhang Z, Jia Q, Wang B, et al. (2017) Discovery and pathogenicity of CYR34, a new race of *Puccinia striiformis* f. sp. *tritici* in China. Acta Phytopathol Sin doi:10.13926/j.cnki.apps.000071

Liu S, Yeh C, Tang HM, Nettleton D, Schnable PS (2012) Gene mapping via bulked segregant RNA-Seq (BSR-Seq). PLoS ONE 7:e36406

Liu W, Frick M, Huel R, Nykiforuk CL, Wang X, et al. (2014) The stripe rust resistance gene *Yr10* encodes an evolutionary-conserved and unique CC-NBS-LRR sequence in wheat. Mol Plant 7:1740-1755

Ma J, Zhou R, Dong Y, Wang L, Wang X, et al. (2001) Molecular mapping and detection of the yellow rust resistance gene *Yr26* in wheat transferred from *Triticum turgidum* L. using microsatellite markers. Euphytica 120:219-226

McIntosh RA, Dubcovsky J, Rogers J, Morris C, Appels R, et al. (2016) Catalogue of gene symbols for wheat: 2016 Supplement. https://shigen.nig.ac.jp/wheat/komugi/genes/macgene/supplement2015.pdf

McIntosh RA, Dubcovsky J, Rogers J, Morris C, Appels R, et al. (2017) Catalogue of gene symbols for wheat: 2017 Supplement. https://shigen.nig.ac.jp/wheat/komugi/genes/macgene/supplement2017.pdf

McIntosh RA, Lagudah ES (2000) Cytogenetical studies in wheat. XVIII. Gene *Yr24* for

resistance to stripe rust. Plant Breeding 119:81-83

McIntosh RA, Mu J, Han D, Kang Z (2018) Wheat stripe rust resistance gene *Yr24/Yr26*: A retrospective review. The Crop J doi:10.1016/j.cj.2018.02.001

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 88:9828-9832

Moore JW, Herrera-Foessel S, Lan C, Schnippenkoetter W, Ayliffe M, et al. (2015) A recently evolved hexose transporter variant confers resistance to multiple pathogens in wheat. Nat Genet 47:1494-1498

Nagy ED, Eder C, Molnár-Láng M, Lelley T (2003) Genetic mapping of sequence-specific PCR-based markers on the short arm of the 1BL.1RS wheat-rye translocation. Euphytica 132:243-249

Peng JH, Fahima T, R O Der MS, Li YC, Dahan A, et al. (1999) Microsatellite tagging of the stripe-rust resistance gene *YrH52* derived from wild emmer wheat, *Triticum dicoccoides*, and suggestive negative crossover interference on chromosome 1B. Theor Appl Genet 98:862-872

Periyannan S, Milne RJ, Figueroa M, Lagudah ES, Dodds PN (2017) An overview of genetic rust resistance: From broad to specific mechanisms. PLoS Pathog 13:e1006380

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945-959

Ramirez-Gonzalez RH, Segovia V, Bird N, Fenwick P, Holdgate S, et al. (2015) RNA-Seq bulked segregant analysis enables the identification of high-resolution genetic markers for breeding in hexaploid wheat. Plant Biotechnol J 13:613-624

Ramirez-Gonzalez RH, Uauy C, Caccamo M (2015) PolyMarker: A fast polyploid primer design pipeline. Bioinformatics 31:2038-2039

Rasheed A, Hao Y, Xia X, Khan A, Xu Y, et al. (2017) Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. Mol plant 10:1047-1064

Schlötterer C, Tobler R, Kofler R, Nolte V (2014) Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. Nat Rev Genet 15:749-763

Singh VK, Khan AW, Saxena RK, Kumar V, Kale SM, et al. (2016) Next-generation sequencing for identification of candidate genes for Fusarium wilt and sterility mosaic disease in pigeonpea (*Cajanus cajan*). Plant Biotechnol J 14:1183-1194

Song WN, Ko L, Henry RJ (1994) Polymorphisms in the α-*amy1* gene of wild and cultivated barley revealed by the polymerase chain reaction. Theor Appl Genet 89:509-513

St. Clair DA (2010) Quantitative disease resistance and quantitative resistance loci in breeding. Annu Rev Phytopathol 48:247-268

Steuernagel B, Periyannan SK, Hernández-Pinzón I, Witek K, Rouse MN, et al. (2016) Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture.

Nat Biotechnol 34:652-655

Stubbs RW (1985) Stripe rust. In: Roelfs, A. P., Bushnell W. R., (eds) The cereal rusts. Vol. II. New York. Academic Press pp 61-101

Sun GL, Fahima T, Korol AB, Turpeinen T, Grama A, et al. (1997) Identification of molecular markers linked to the *Yr15* stripe rust resistance gene of wheat originated in wild emmer wheat, *Triticum dicoccoides*. Theor Appl Genet 95:622-628

Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, et al. (2013) QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J 74:174-183

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular rvolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731-2739

Thind AK, Wicker T, Simkova H, Fossati D, Moullet O, et al. (2017) Rapid cloning of genes in hexaploid wheat using cultivar-specific long-range chromosome assembly. Nat Biotechnol 35:793-796

Trick M, Adamski N, Mugford SG, Jiang C, Febrer M, et al. (2012) Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. BMC Plant Biology 12:14

Uauy C (2017a) Wheat genomics comes of age. Curr Opin Plant Biol 36:142-148

Uauy C, Wulff BBH, Dubcovsky J (2017b) Combining traditional mutagenesis with new high-throughput sequencing and genome editing to reveal hidden variation in polyploid wheat. Annu Rev Genet 51:435-454

United Nations Department Of Economic And Social Affairs PD (2015) World population prospects: the 2015 revision. Working Paper no. ESA/P/WP.241. [WWW document] URL https://esa.un.org/unpd/wpp/ [accessed 27 February 2017].

Uricaru R, Rizk G, Lacroix V, Quillery E, Plantard O, et al. (2015) Reference-free detection of isolated SNPs. Nucleic Acids Res 43:e11

Van Ooijen JW (2006) JoinMap4, software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, The Netherlands

van Poecke RMP, Maccaferri M, Tang J, Truong HT, Janssen A, et al. (2013) Sequence-based SNP genotyping in durum wheat. Plant Biotechnol J 11:809-817

Varshney RK, Terauchi R, McCouch SR (2014) Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. PLoS biology 12:e1001883

Voorrips RE (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. J Hered 93:77-78

Wang C, Zhang Y, Han D, Kang Z, Li G, et al. (2008) SSR and STS markers for wheat stripe rust resistance gene *Yr26*. Euphytica 159:359-366

Wang L, Ma J, Zhou R, Wang X, Jia J (2002) Molecular tagging of the yellow rust resistance gene *Yr10* in common wheat, PI178383 (*Triticum aestivum* L.). Euphytica 124:71-73

Wang S, Wong D, Forrest K, Allen A, Chao S, et al. (2014) Characterization of polyploid

wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnol J 12:787-796

Wang Y, Xie J, Zhang H, Guo B, Ning S, et al. (2017) Mapping stripe rust resistance gene *YrZH22* in Chinese wheat cultivar Zhoumai 22 by bulked segregant RNA-Seq (BSR-Seq) and comparative genomics analyses. Theor Appl Genet 130:2191-2201

Wellings CR (2011) Global status of stripe rust: a review of historical and current threats. Euphtica 179:129-141

William M, Singh RP, Huerta-Espino J, Islas SO, Hoisington D (2003) Molecular marker mapping of leaf rust resistance gene *Lr46* and its association with stripe rust resistance gene *Yr29* in wheat. Phytopathology 93:153-159

Wu J, Liu S, Wang Q, Zeng Q, Mu J, et al. (2018a) Rapid identification of an adult plant stripe rust resistance gene in hexaploid wheat by high-throughput SNP array genotyping of pooled extremes. Theor Appl Genet 131:43-58

Wu J, Wang Q, Xu L, Chen X, Li B, et al. (2018b) Combining SNP genotyping array with bulked segregant analysis to map a gene controlling adult-plant resistance to stripe rust in wheat line 03031-1-5 H62. Phytopathology 108:103-113

Wu JH, Wang QL, Chen XM, Wang MJ, Mu JM, et al. (2016) Stripe rust resistance in wheat breeding lines developed for central Shaanxi, an overwintering region for *Puccinia striiformis* f. sp. *tritici* in China. Can J Plant Pathol 38:317-324

Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26:873-881

Xu Y, Li P, Zou C, Lu Y, Xie C, et al. (2017) Enhancing genetic gain in the era of molecular breeding. J Exp Bot 68:2641-2666

Zeng Q, Han D, Wang Q, Yuan F, Wu J, et al. (2014) Stripe rust resistance and genes in Chinese wheat cultivars and breeding lines. Euphytica 196:271-284

Zhang H, Zhang L, Wang C, Wang Y, Zhou X, et al. (2016) Molecular mapping and marker development for the *Triticum dicoccoides*-derived stripe rust resistance gene *YrSM139-1B* in bread wheat cv. Shaanmai 139. Theor Appl Genet 129:369-376

Zhang X, Han D, Zeng Q, Duan Y, Yuan F, et al. (2013) Fine mapping of wheat stripe rust resistance gene *Yr26* based on collinearity of wheat with *Brachypodium distachyon* and rice. PLoS ONE 8:e57885

Zou C, Wang P, Xu Y (2016) Bulked sample analysis in genetics, genomics and crop improvement. Plant Biotechnol J 14:1941-1955

Legends for all figures, supporting figures and supporting tables

Fig. 1. Schematic representation of experimental methodology. (**A**) Development of recombinant inbred lines (RILs) by single seed descent from crosses YM5/92R137,

Avs/92R137 and NIL-S/NIL-R. (**B**) Stripe rust phenotyping of YM5, AvS, Y158, NIL-S, 92R137, NIL-R and their corresponding progenies at both the seedling and adult plant stages after inoculation with *Pst* race CYR32. Progenies were pooled according to resistant or susceptible phenotype. (**C**) The RNA of super- pools were sequenced in an Illumina HiSeq2500. An analytical pipeline was performed to align the transcript reads to the common wheat genome IWGSC RefSeq v1.0 and wild emmer genome WEWSeq v1.0, respectively, followed by SNP calling and filtering by Allele Difference Ratio (ADR) and Bulk Frequency Ratios (BFR), respectively. (**D**) DNAs from super-pools were genotyped by the 90K and 660K SNP arrays. Homozygous SNP genotypes were identified and aligned to IWGSC RefSeq v1.0, and polymorphic SNPs were counted and clustered according to their physical locations. (**E**) Putative SNPs were validated by KASP assays and were then used to construct a high-density genetic map and subjected to haplotype analysis.

Fig. 2. Overview of analyses by the 90K and 660K SNP arrays. **A, D**: Venn diagrams of polymorphic SNPs; **B, E**: Distribution of polymorphic SNPs in each chromosome was based on 90K and 660K genetic maps (Wang et al., 2014; Jizeng Jia, personal communication); **C, F**: Physical positions of polymorphic SNPs in the chr 1B reference genome were determined by best alignment to IWGSC RefSeq v1.0. SNPs were clustered by location (windows size: 1 Mb) and selected SNPs (in red ellipses) were analyzed in KASP assays.

Fig. 3. Overview of analyses of BSR-Seq by Allele Difference Ratio (ADR). (**A**) Trimming raw data, alignment of sequencing reads to the wheat genome assembly and discovering SNPs. (**B, C**) Filtering SNPs by multiple criteria and identifying 530 trait-associated SNPs on chromosome 1B. (**C**) Physical positions of the 530 SNPs and several genetic markers linked to *Yr26* in IWGSC RefSeq v1.0. Genetic markers proximal to *Yr26* are designated in red rhombi, whereas markers distal to *Yr26* gene are designated in green triangles. Based on the physical positions of genetic markers in previous studies, SNPs in the red frame were selected as candidate SNPs and used for KASP assays. (**D**) Genetic map constructed with KASP markers. Blue markers indicate SNPs identified by ADR.

Fig. 4. Genetic and physical mapping of *Yr26*. (**A**) Deletion bin location of *Yr26* in 1BL. (**B**) Graphical genotyping of the *Yr26* region using data for F_8 and F_6 RILs derived from YM5/92R137 and NIL-S/NIL-R, respectively. (**C**) Physical positions of KASP-SNP markers (according to IWGSC RefSeq v1.0) used for fine mapping of the *Yr26* locus. Recombinants involving adjacent markers and numbers of lines in each group are indicated in parentheses.

Green, orange and slash bars represent homozygous resistant (92R137 and NIL-R), homozygous susceptible (YM5 and NIL-S) and heterozygous resistant genotypes, respectively. Phenotypic differences of each recombinant group are listed on the right. R, resistant; S, susceptible; H, heterozygous resistant. The phenotypes of F_2 plants were confirmed by progeny testing the $F_{2:3}$ lines. (**D**) Annotated genes were predicted to occur between the *CM1461* and *WRS467* markers in the IWGSC RefSeq v1.0 genome. Colored arrows indicate disease resistance-related genes and their transcriptional orientations. (**E**) Annotated genes predicted to occur between the *CM1461* and *WRS467* markers in the WEWSeq v1.0 genome. Colored arrows indicate genes in the target region and their transcriptional orientations.

Fig. 5. Collinearity analysis of the *Yr26* genetic map and rice, barley, *Brachypodium*, maize and sorghum physical maps. Consensus genetic map of *Yr26* on wheat chr 1BL produced from results from F_6 and F_8 RILs (Ta-1BL). Each color corresponds to a chromosome on the circle. Orthologous genomic regions of *Yr26* on rice chr 10 (Os10), barley chr 1 (Hv1), *Brachypodium* chr 3 (Bd3) and maize chr 1 (Zea1), and sorghum chr 1 (Sb1). The lines represent the relationships between mapped genes on wheat chr 1BL and orthologous genes in rice, barley, *B. distachyon*, sorghum, and maize. The black wedges indicate the *Yr26* region in wheat and its homologous regions in other graminaceous species. The figure was trimmed to eliminate redundant information, namely, more than one SNP corresponding to wheat genes matching the same rice/barley/*B. distachyon*/maize/sorghum CDS region were merged. More detail is provided in Supplementary Table S8.

Fig. 6. Haplotype variation of the *Yr26* region among wheat genotypes revealed by SNPs distributed in the target region. Phylogenetic tree constructed with SNPs in the *Yr26* region. The color of each accession in the tree was according to the topological structure groups in (A). The details of results based on Wheat660K SNP markers in supplementary Table S10.

Fig. S1. Main *Yr* genes used across wheat growing areas in China and the proportion of *Yr26* in each *Pst* epidemic region. A group of 1,322 wheat entries from the CARS wheat germplasm collection was used for analysis. They included 667 cultivars, 636 breeding lines and 19 ancestral donors of *Yr* genes on 1B and susceptible checks. The data were from our previous studies.

Fig. S2. Statistics of RNA-seq reads alignment to the bread wheat reference genome.

Fig. S3. Physical location of SNPs with bulk frequency ratios (BFR) >6 mapped to wild emmer chr 1B

Fig. S4. Genotypic plots from selected KASP assays. The X- and Y-axes indicate FAM- and HEX-fuorescence units, respectively. The central cluster (green) is comprised of heterozygous individuals, whereas clusters near the axes are homozygous for either the AvS, YM5 and NIL-S (HEX; red), or 92R137 and NIL-R (FAM; blue) alleles. Black dots in the lower left indicate a water control (non-template control, NTC) and pink dots represent missing or failed data. **(A, B)** Marker *WRS270* and *WRS290* results for the RIL population each showing two clusters. **(C, D)** *WRS435* and *WRS312* results for the F_2 population showing three major clusters. **(E, F, G, H)** Results of genotyping with *WRS467*, *CM501*, *CM1461* and *CM1641* in 1,322 diverse cultivars and breeding lines, each showing two main clusters

Table S1. Genetic analysis of seedling resistance to to *Pst* race CYR32 in different populations

Table A in S2. The physical positions of all markers in IWGSC RefSeq V.1.0 using BLAST

Table B in S2. The physical positions of EST sequences distributed in different 1B bins in IWGSC RefSeq V.1.0 using BLAST

Table S3. Summary of RNA-Seq data and alignments

Table S4. Summary of SNP discovery. A total of 338,731 (73.17% of 462,943) SNPs were located on 21 chromosomes

Table S5. Information of *Yr26*-linked SNPs derived from BSR-Seq

Table S6. SNPs in chr 1B vs total number of SNPs with BFR >6

Table S7. Kompetitive Allele Specific PCR (KASP) primers used to genotype individuals in the F_2 and RIL populations for mapping *Yr26* and haplotype analysis

Table S8. Best blast hit of wheat gene sequences corresponding to SNP-flanked sequence against protein or CDS sequences of rice, barley, *Brachypodium*, maize and sorghum

Table A in S9. Genes in the fine-mapped chr 1B interval. Gene models between flanking markers *WRS435* and *WRS312* (positions 325,289,375 and 329,486,125, respectively, on chr 1B according to Chinese Spring IWGSC RefSeq v1.0)

Table B in S9. Genes in the fine-mapped chr 1B interval. Gene models between flanking

markers *WRS435* and *WRS312* (positions 325,764,533 and 333,201,982, respectively, according to Zavitan WEWSeq v1.0)

Table S10. Haplotype variation in the *Yr26* region among the set of 384 wheat genotypes revealed by SNPs in the target region

Table S11. Alleles of KASP markers flanking *Yr26* in 1,322 accessions including ancestral donors of *Yr* genes on 1B, susceptible checks, wheat cultivars and advanced breeding lines

Figure

Fig.1

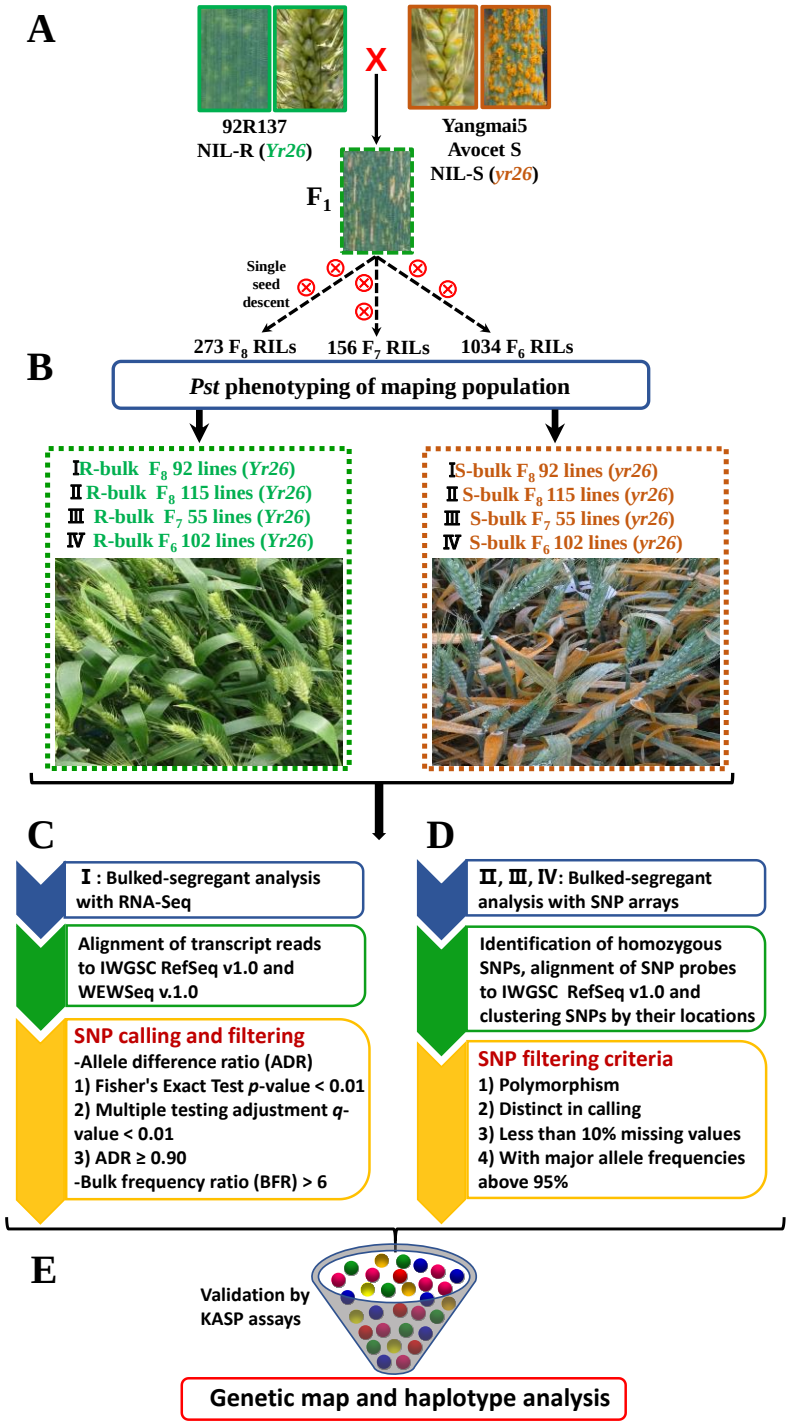


Fig.2

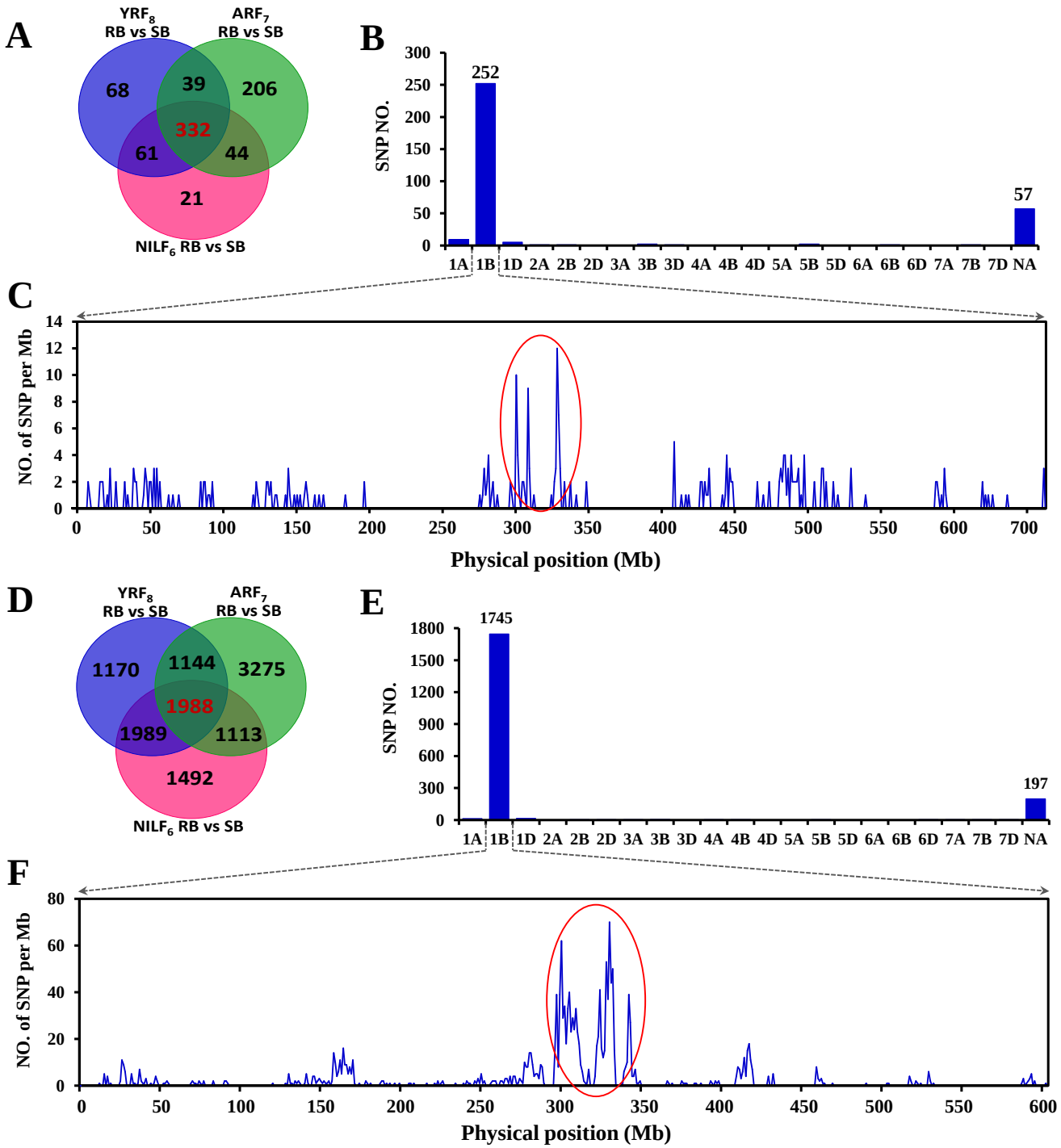

A
YRF8 RB vs SB
ARF7 RB vs SB
NILF6 RB vs SB
68
39
206
332
61
44
21
B
SNP NO.
252
57
1A 1B 1D 2A 2B 2D 3A 3B 3D 4A 4B 4D 5A 5B 5D 6A 6B 6D 7A 7B 7D NA
C
NO. of SNP per Mb
Physical position (Mb)
D
YRF8 RB vs SB
ARF7 RB vs SB
NILF6 RB vs SB
1170
1144
3275
1988
1989
1113
1492
E
SNP NO.
1745
197
1A 1B 1D 2A 2B 2D 3A 3B 3D 4A 4B 4D 5A 5B 5D 6A 6B 6D 7A 7B 7D NA
F
NO. of SNP per Mb
Physical position (Mb)

Fig.3

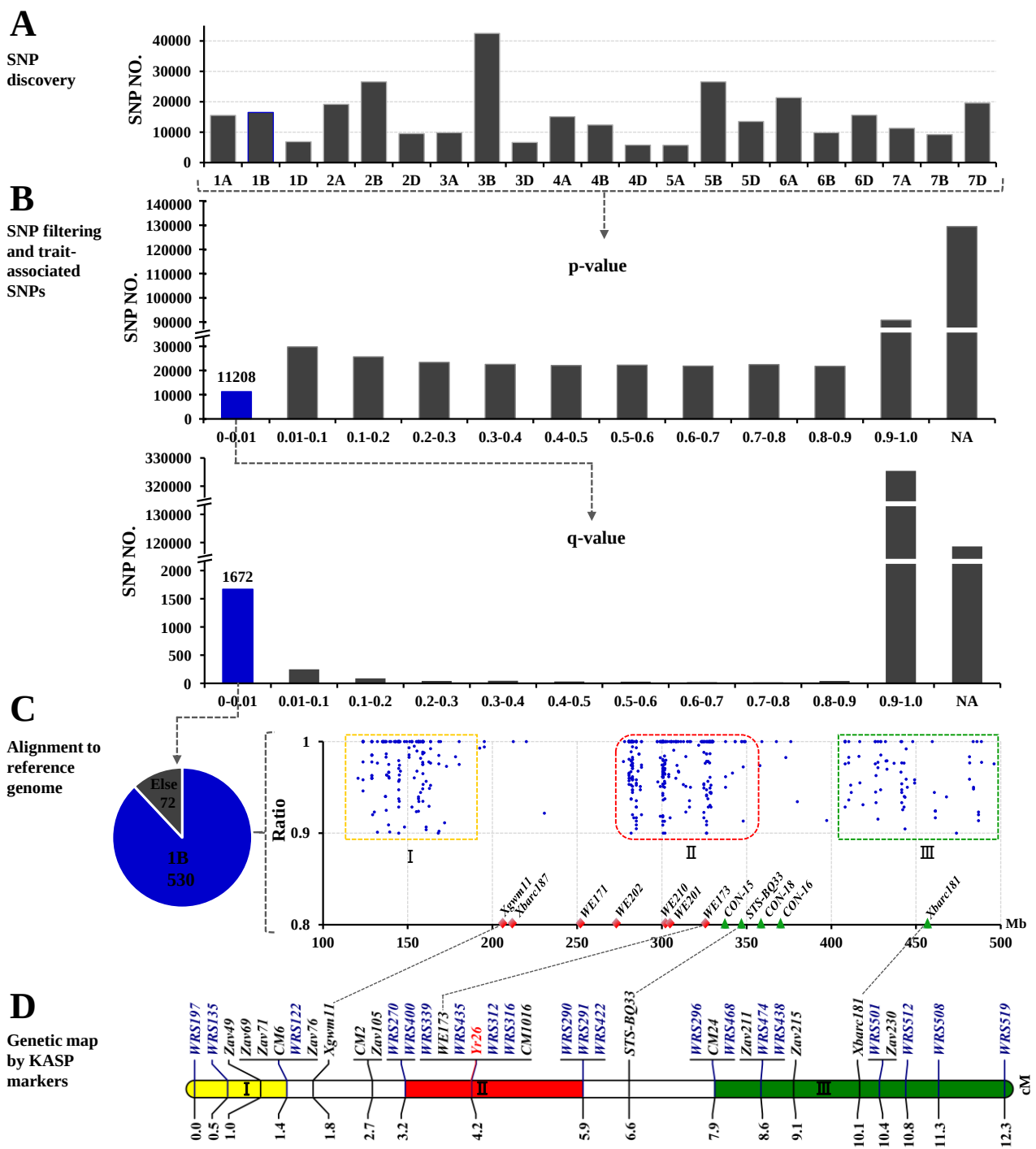

A
SNP discovery
SNP NO.
40000
30000
20000
10000
0
1A 1B 1D 2A 2B 2D 3A 3B 3D 4A 4B 4D 5A 5B 5D 6A 6B 6D 7A 7B 7D
B
SNP filtering and trait-associated SNPs
p-value
SNP NO.
140000
130000
120000
110000
100000
90000
30000
20000
10000
0
11208
0-0.01 0.01-0.1 0.1-0.2 0.2-0.3 0.3-0.4 0.4-0.5 0.5-0.6 0.6-0.7 0.7-0.8 0.8-0.9 0.9-1.0 NA
q-value
SNP NO.
330000
320000
130000
120000
2000
1500
1000
500
0
1672
0-0.01 0.01-0.1 0.1-0.2 0.2-0.3 0.3-0.4 0.4-0.5 0.5-0.6 0.6-0.7 0.7-0.8 0.8-0.9 0.9-1.0 NA
C
Alignment to reference genome
Else
72
1B
530
Ratio
1
0.9
0.8
I
II
III
Xgwm11 Xbarc187 WE171 WE202 WE210 WE201 WE173 CON-15 STS-BQ33 CON-18 CON-16 Xbarc181
100 150 200 250 300 350 400 450 500
Mb
D
Genetic map by KASP markers
WRS197 WRS135 Zav49 Zav69 Zav71 CM6 WRS122 Zav76 Xgwm11 CM2 Zav105 WRS270 WRS400 WRS339 WE173 WRS435 Yr26 WRS312 WRS316 CM1016 WRS290 WRS291 WRS422 STS-BQ33 WRS296 CM24 WRS468 Zav211 WRS474 WRS438 Zav215 Xbarc181 WRS501 Zav230 WRS512 WRS508 WRS519
I
II
III
cM
0.0 0.5 1.0 1.4 1.8 2.7 3.2 4.2 5.9 6.6 7.9 8.6 9.1 10.1 10.4 10.8 11.3 12.3

Fig.4

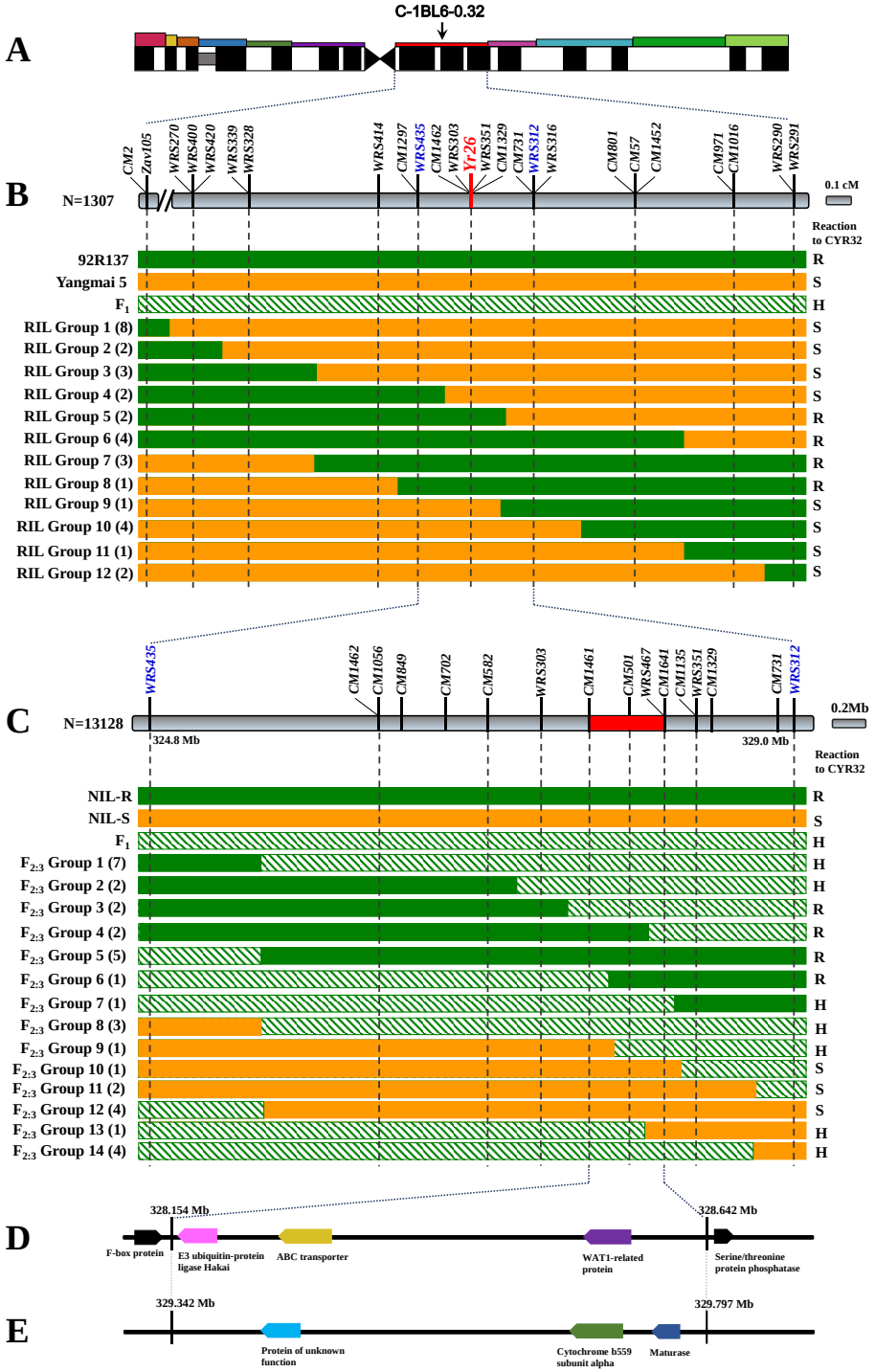

A
C-1BL6-0.32
B
N=1307
0.1 cM
CM2
Zav105
WRS270
WRS400
WRS420
WRS339
WRS328
WRS414
CM1297
WRS435
CM1462
WRS303
Yr26
WRS351
CM1329
CM731
WRS312
WRS316
CM801
CM57
CM1452
CM971
CM1016
WRS290
WRS291
Reaction to CYR32
92R137 R
Yangmai 5 S
F1 H
RIL Group 1 (8) S
RIL Group 2 (2) S
RIL Group 3 (3) S
RIL Group 4 (2) S
RIL Group 5 (2) R
RIL Group 6 (4) R
RIL Group 7 (3) R
RIL Group 8 (1) R
RIL Group 9 (1) S
RIL Group 10 (4) S
RIL Group 11 (1) S
RIL Group 12 (2) S
C
N=13128
0.2Mb
324.8 Mb
329.0 Mb
WRS435
CM1462
CM1056
CM849
CM702
CM582
WRS303
CM1461
CM501
WRS467
CM1641
CM1135
WRS351
CM1329
CM731
WRS312
Reaction to CYR32
NIL-R R
NIL-S S
F1 H
F2:3 Group 1 (7) H
F2:3 Group 2 (2) H
F2:3 Group 3 (2) R
F2:3 Group 4 (2) R
F2:3 Group 5 (5) R
F2:3 Group 6 (1) R
F2:3 Group 7 (1) H
F2:3 Group 8 (3) H
F2:3 Group 9 (1) H
F2:3 Group 10 (1) S
F2:3 Group 11 (2) S
F2:3 Group 12 (4) S
F2:3 Group 13 (1) H
F2:3 Group 14 (4) H
D
328.154 Mb
328.642 Mb
F-box protein
E3 ubiquitin-protein ligase Hakai
ABC transporter
WAT1-related protein
Serine/threonine protein phosphatase
E
329.342 Mb
329.797 Mb
Protein of unknown function
Cytochrome b559 subunit alpha
Maturase

Fig.5

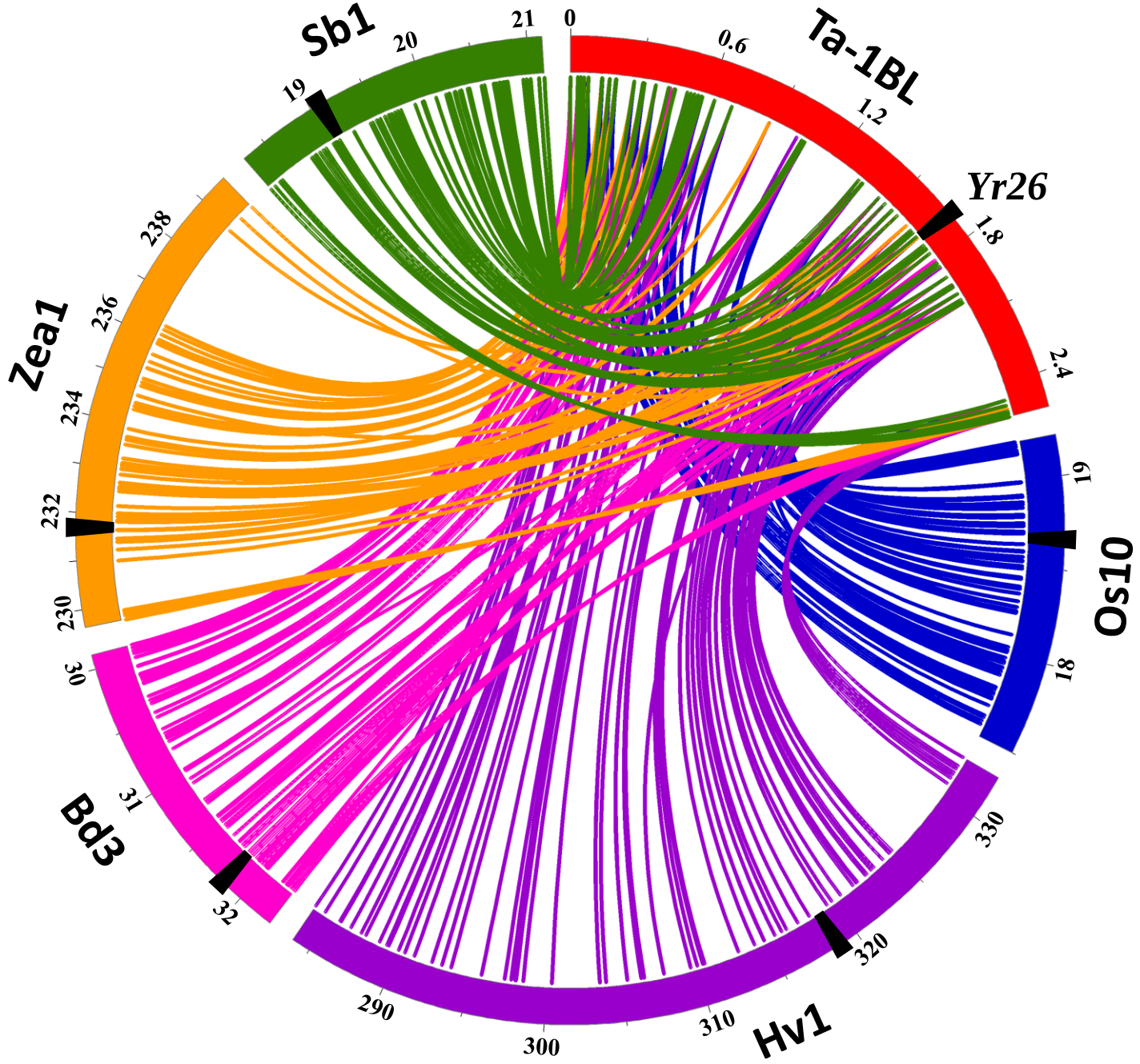
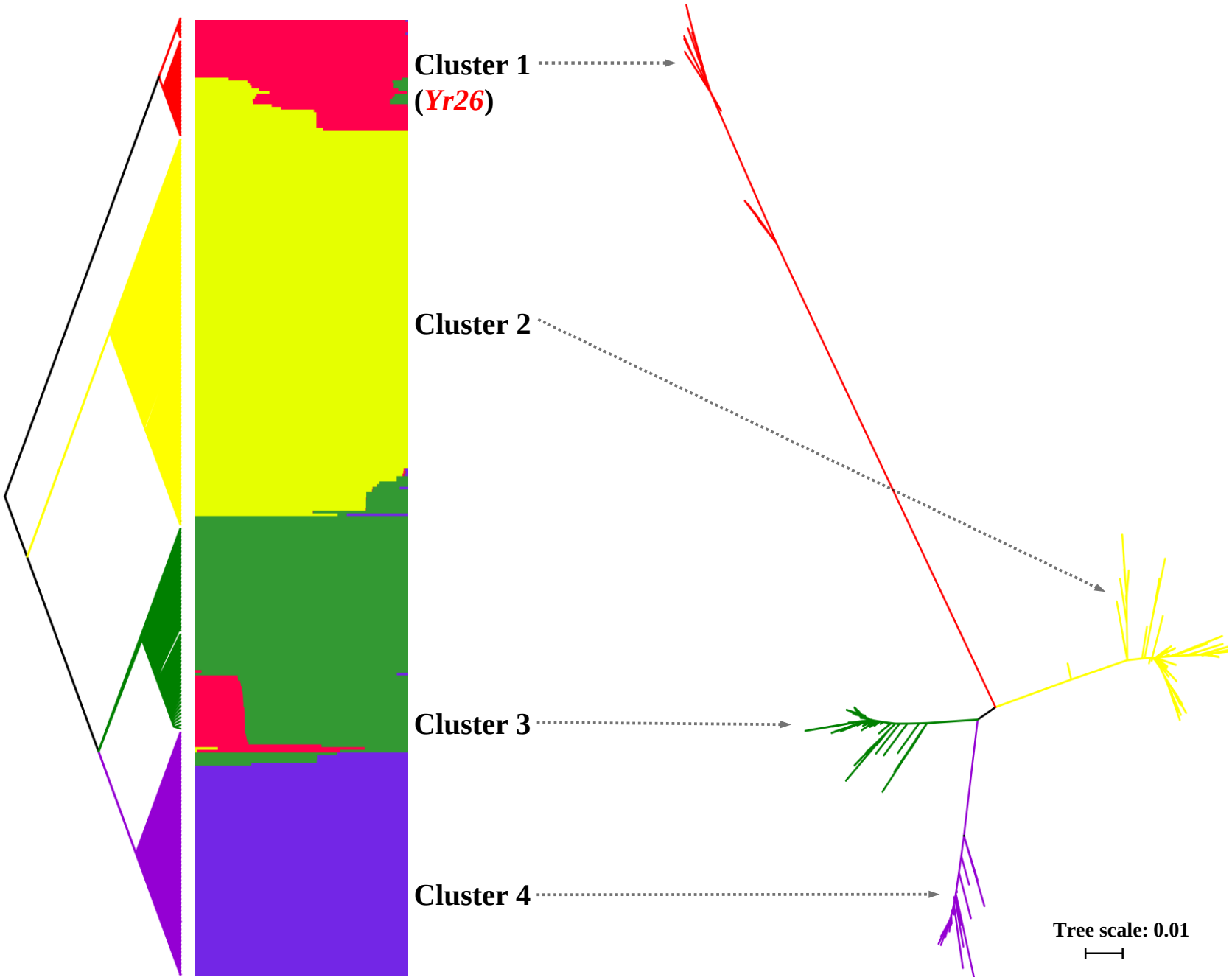

Sb1
Ta-1BL
Yr26
Os10
Hv1
Bd3
Zea1
0
0.6
1.2
1.8
2.4
19
18
320
330
290
300
310
31
32
30
230
232
234
236
238
19
20
21

Fig.6


Cluster 1
(Yr26)
Cluster 2
Cluster 3
Cluster 4
Tree scale: 0.01
A. NJ tree
B. Structure
C. Splits tree

Click here to access/download

Supplementary Material (download option from within PDF)

Supplementary Table S1.xlsx

Click here to access/download

Supplementary Material (download option from within PDF)

Supplementary Table S2.xlsx

Click here to access/download

Supplementary Material (download option from within PDF)

Supplementary Table S3.xlsx

Click here to access/download

Supplementary Material (download option from within PDF)

Supplementary Table S4.xlsx

Click here to access/download

Supplementary Material (download option from within PDF)

Supplementary Table S5.xlsx

Click here to access/download

Supplementary Material (download option from within PDF)

Supplementary Table S6.xlsx

Click here to access/download

Supplementary Material (download option from within PDF)

Supplementary Table S7.xlsx

Click here to access/download

Supplementary Material (download option from within PDF)

Supplementary Table S8.xlsx

Click here to access/download

Supplementary Material (download option from within PDF)

Supplementary Table S9.xlsx

Click here to access/download

Supplementary Material (download option from within PDF)

Supplementary Table S10.xlsx

Click here to access/download

Supplementary Material (download option from within PDF)

Supplementary Table S11.xlsx

Change of authorship request form (pre-acceptance)

Please read the important information on page 4 before you begin

This form should be used by authors to request any change in authorship including changes in corresponding authors. Please fully complete all sections. Use black ink and block capitals and provide each author's full name with the given name first followed by the family name.

Please note: In author collaborations where there is formal agreement for representing the collaboration, it is sufficient for the representative or legal guarantor (usually the corresponding author) to complete and sign the Authorship Change Form on behalf of all authors.

Manuscript ID no. in case of unpublished manuscript:	TAAG-D-17-00473R1
DOI in case of published manuscript:	
ISBN (for books):	

Section 1: Please provide the current title of manuscript
(For journals: Please provide the manuscript ID, title and/or DOI if available.)
(For books: Please provide the title, ISBN and/or DOI if available.)

Title: SNP-based pool genotyping and haplotype analysis accelerates fine mapping of the wheat genomic region containing stripe rust resistance gene *Yr26*

Section 2: Please provide the previous authorship, in the order shown on the manuscript before the changes were introduced. Please indicate the corresponding author by adding (CA) behind the name.

Notice: This is the authors list submitted last time, but it is not the list in the previous manuscript. We have explained it, please see it below.

	First name(s)	Family name	ORCID or SCOPUS id, if available
1st author			
2nd author	Zhensheng	Kang (CA)	
3rd author	Jianhui	Wu	
4th author	Qingdong	Zeng	
5th author	Qilin	Wang	
6th author	Shengjie	Liu	
7th author	Shizhou	Yu	
	Jingmei	Mu	

Please use an additional sheet if there are more than 7 authors.

8th author	Shuo	Huang
9th author	Hanan	Sela
10th author	Assaf	Distelfeld
11th author	Lili	Huang
12th author	Dejun	Han (CA)

Change of authorship request form (pre-acceptance)

Section 3: Please provide a justification for change. Please use this section to explain your reasons for changing the authorship of your manuscript, e.g. what necessitated the change in authorship? Please refer to the (journal) policy pages for more information about authorship. Please explain why omitted authors were not originally included and/or why authors were removed on the submitted manuscript.

Here, we have to apologize that we made a mistake. When we started to submit the manuscript, we didn't know how to rearrange the author's order as corresponding author was always in the first place. Then we have to acquiesce it. However, in the first version the authors list submitted was inconsistent with the list in manuscript. Because the corresponding author Zhensheng Kang was in the first place. This time we have known that how to operate to adjust the authors' order and put the corresponding author Zhensheng Kang in the end. Now the list is the same as that in manuscript. In fact, the previous version of manuscript is the same as this version. You can check it. We feel sorry about this.

Section 4: Proposed new authorship. Please provide your new authorship list in the order you would like it to appear on the manuscript. Please indicate the corresponding author by adding (CA) behind the name. If the corresponding author has changed, please indicate the reason under section 3.

	First name(s)	Family name (this name will appear in full on the final publication and will be searchable in various abstract and indexing databases)
1st author		
2nd author	Jianhui	Wu
3rd author	Qingdong	Zeng
4th author	Qilin	Wang
5th author	Shengjie	Liu
6th author	Shizhou	Yu
7th author	Jingmei	Mu

Please use an additional sheet if there are more than 7 authors.

8th author	Hanan	Sela
9th author	Assaf	Distelfeld
10th author	Lili	Huang
11th author	Dejun	Han (CA)
12th author	Zhensheng	Kang (CA)

Change of authorship request form (pre-acceptance)

Section 5: Author contribution, Acknowledgement and Disclosures. Please use this section to provide a new disclosure statement and, if appropriate, acknowledge any contributors who have been removed as authors and ensure you state what contribution any new authors made (if applicable per the journal or book (series) policy). Please ensure these are updated in your manuscript - after approval of the change(s) - as our production department will not transfer the information in this form to your manuscript.

New acknowledgements:

The authors are grateful to Prof. R. A. McIntosh, Plant Breeding Institute, University of Sydney, for critical review of this manuscript; Prof. Peidu Chen and Prof. Aizhong Cao, Cytogenetics Institute, Nanjing Agricultural University, for providing Yr26 germplasms and genetic populations. This study was financially supported by International S&T Cooperation Program of China (2015DFG32340), National Natural Science Foundation of China (31371924), the National Key Research and Development Program of China (Grant No. 2016YFE0108600), the earmarked funds for Modern Agro-industry Technology Research System (No. CARS-3-1-11) and National Natural Science Foundation for Young Scientists of China (Grant 31701421).

New Disclosures (financial and non-financial interests, funding):

Not applicable

New Author Contributions statement (if applicable per the journal policy):

Not applicable

State 'Not applicable' if there are no new authors.

Change of authorship request form (pre-acceptance)

Section 6: Declaration of agreement. *All* authors, unchanged, new and removed *must* sign this declaration.
(NB: Please print the form, sign and return a scanned copy. Please note that signatures that have been inserted as an image file are acceptable as long as it is handwritten. Typed names in the signature box are unacceptable.) *** Please delete as appropriate. Delete all of the bold if you were on the original authorship list and are remaining as an author.**

	First name	Family name		Signature	Affiliated institute	Date
1st author	Jianhui	Wu	I agree to the proposed new authorship shown in section 4 **/and the addition/removal*of my name to the authorship list.**			
2nd author	Qingdong	Zeng	I agree to the proposed new authorship shown in section 4 **/and the addition/removal*of my name to the authorship list.**			
3rd author	Qilin	Wang	I agree to the proposed new authorship shown in section 4 **/and the addition/removal*of my name to the authorship list.**			
4th authors	Shengjie	Liu	I agree to the proposed new authorship shown in section 4 **/and the addition/removal*of my name to the authorship list.**			
5th author	Shizhou	Yu	I agree to the proposed new authorship shown in section 4 **/and the addition/removal*of my name to the authorship list.**			
6th author	Jingmei	Mu	I agree to the proposed new authorship shown in section 4 **/and the addition/removal*of my name to the authorship list.**			
7th author	Shuo	Huang	I agree to the proposed new authorship shown in section 4 **/and the addition/removal*of my name to the authorship list.**			

Please use an additional sheet if there are more than 7 authors.

Change of authorship request form (pre-acceptance)

Important information. Please read.

- Please return this form, fully completed, to Springer Nature. We will consider the information you have provided to decide whether to approve the proposed change in authorship. We may choose to contact your institution for more information or undertake a further investigation, if appropriate, before making a final decision.

- By signing this declaration, all authors guarantee that the order of the authors are in accordance with their scientific contribution, if applicable as different conventions apply per discipline, and that only authors have been added who made a meaningful contribution to the work.

- Please note, we cannot investigate or mediate any authorship disputes. If you are unable to obtained agreement from all authors (including those who you wish to be removed) you must refer the matter to your institution(s) for investigation. Please inform us if you need to do this.

- If you are not able to return a fully completed form within **30 days** of the date that it was sent to the author requesting the change, we may have to withdraw your manuscript. We cannot publish manuscripts where authorship has not been agreed by all authors (including those who have been removed).

- Incomplete forms will be rejected.

Authorship change form

	First name	Family name		Signature	Affiliated institute	Date
1st author	Jianhui	Wu	I agree to the proposed new authorship shown in section 4	Jianhui Wu	State Key Laboratory of Crop Stress Biology for Arid Areas, Northwest A&F University	2018.3.30
2nd author	Qingdong	Zeng	I agree to the proposed new authorship shown in section 4	Qingdong Zeng	State Key Laboratory of Crop Stress Biology for Arid Areas, Northwest A&F University	2018.3.30
3rd author	Qilin	Wang	I agree to the proposed new authorship shown in section 4	Qilin Wang	State Key Laboratory of Crop Stress Biology for Arid Areas, Northwest A&F University	2018.3.30
4th author	Shengjie	Liu	I agree to the proposed new authorship shown in section 4	Shengjie Liu	State Key Laboratory of Crop Stress Biology for Arid Areas, Northwest A&F University	2018.3.30
5th author	Shizhou	Yu	I agree to the proposed new authorship shown in section 4	Shizhou Yu	State Key Laboratory of Crop Stress Biology for Arid Areas, Northwest A&F University	2018.3.30
6th author	Jingmei	Mu	I agree to the proposed new authorship shown in section 4	Jingmei Mu	State Key Laboratory of Crop Stress Biology for Arid Areas, Northwest A&F University	2018.3.30
7th author	Shuo	Huang	I agree to the proposed new authorship shown in section 4	Shuo Huang	State Key Laboratory of Crop Stress Biology for Arid Areas, Northwest A&F University	2018.3.30
8th author	Hanan	Sela	I agree to the proposed new authorship shown in section 4		The Institute for Cereal Crops Improvement Tel-Aviv University	2018.3.30
9th author	Assaf	Distelfeld	I agree to the proposed new authorship shown in section 4	Assaf Distelfeld	School of Plant Sciences and Food Security, Tel Aviv University	2018.3.30
10th author	Lili	Huang	I agree to the proposed new authorship shown in section 4	Lili Huang	State Key Laboratory of Crop Stress Biology for Arid Areas, Northwest A&F University	2018.3.30
11th author	Dejun	Han	I agree to the proposed new authorship shown in section 4	Dejun Han	State Key Laboratory of Crop Stress Biology for Arid Areas, Northwest A&F University	2018.3.30
12th author	Zhensheng	Kang	I agree to the proposed new authorship shown in section 4	Zhensheng Kang	State Key Laboratory of Crop Stress Biology for Arid Areas, Northwest A&F University	2018.3.30

Click here to download Authorship change form Sign names.JPG