



OPEN ACCESS

**Edited by:**

Marion S. Röder,  
Leibniz Institute of Plant Genetics and  
Crop Plant Research (IPK), Germany

**Reviewed by:**

Matthew Rouse,  
United States Department of  
Agriculture, United States  
Sambasivam Periyannan,  
Commonwealth Scientific and  
Industrial Research Organisation  
(CSIRO), Australia  
Colin Hiebert,  
Agriculture and Agri-Food Canada  
(AAFC), Canada

**\*Correspondence:**

Lichao Zhang  
zhanglichao@caas.cn

Xu Liu  
liuxu03@caas.cn  
Xiuying Kong  
kongxiuying@caas.cn

<sup>†</sup>These authors have contributed  
equally to this work

**Specialty section:**

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

**Received:** 15 May 2020

**Accepted:** 29 July 2020

**Published:** 13 August 2020

**Citation:**

Dong C, Zhang L, Chen Z, Xia C, Gu Y,  
Wang J, Li D, Xie Z, Zhang Q, Zhang X,  
Gui L, Liu X and Kong X (2020)  
Combining a New Exome Capture  
Panel With an Effective varBScore  
Algorithm Accelerates BSA-Based  
Gene Cloning in Wheat.  
*Front. Plant Sci.* 11:1249.  
doi: 10.3389/fpls.2020.01249

# Combining a New Exome Capture Panel With an Effective varBScore Algorithm Accelerates BSA-Based Gene Cloning in Wheat

Chunhao Dong<sup>1†</sup>, Lichao Zhang<sup>1\*†</sup>, Zhongxu Chen<sup>2,3†</sup>, Chuan Xia<sup>1</sup>, Yongqiang Gu<sup>5</sup>,  
Jirui Wang<sup>3,4</sup>, Danping Li<sup>1</sup>, Zhencheng Xie<sup>1</sup>, Qiang Zhang<sup>1</sup>, Xueying Zhang<sup>1</sup>,  
Lixuan Gui<sup>2,3</sup>, Xu Liu<sup>1\*</sup> and Xiuying Kong<sup>1\*</sup>

<sup>1</sup> Key Laboratory for Crop Gene Resources and Germplasm Enhancement, MOA, National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China, <sup>2</sup> Department of Life Science, Chengdu Tcuni Technology, Chengdu, China, <sup>3</sup> State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Sichuan Agricultural University, Chengdu, China, <sup>4</sup> Triticeae Research Institute, Sichuan Agricultural University, Chengdu, China, <sup>5</sup> Western Regional Research, United States Department of Agriculture-Agricultural Research Service, Albany, CA, United States

The discovery of functional genes underlying agronomic traits is of great importance for wheat improvement. Here we designed a new wheat exome capture probe panel based on IWGSC RefSeq v1.0 genome sequence information and developed an effective algorithm, varBScore, that can sufficiently reduce the background noise in gene mapping and identification. An effective method, termed bulked segregant exome capture sequencing (BSE-Seq) for identifying causal mutations or candidate genes<sup>5</sup>, was established by combining the use of a newly designed wheat exome capture panel, sequencing of bulked segregant pools from segregating populations, and the robust algorithm varBScore. We evaluated the effectiveness of varBScore on SNP calling using the published dataset for mapping and cloning the yellow rust resistance gene *Yr7* in wheat. Furthermore, using BSE-Seq, we rapidly identified a wheat yellow leaf mutant gene, *yg11*, in an ethyl methanesulfonate (EMS) mutant population and found that a single mutation of G to A at 921 position in the wild type *YGL1* gene encoding magnesium-chelatase subunit chlI caused the leaf yellowing phenotype. We further showed that mutation of *YGL1* through CRISPR/Cas9 gene editing led to a yellow phenotype on the leaves of transgenic wheat, indicating that *yg11* is the correct causal gene responsible for the mutant phenotype. In summary, our approach is highly efficient for discovering causal mutations and gene cloning in wheat.

**Keywords:** wheat exome capture, varBScore, gene cloning, bulked segregant analysis, CRISPR/Cas9, yellow-green leaf mutant

## INTRODUCTION

Wheat (*Triticum aestivum* L.,  $2n = 6x = 42$ , AABBDD) is an important staple crop, providing 20% of all calories consumed by the world population. Identification of genes underlying desirable agronomic traits is of great significance for genetic improvement of wheat. However, wheat functional genome research has lagged behind model plant species due to the lack of complete reference genome sequence in previous years and its complex hexaploid genome. The release of genome sequences of several wheat varieties and progenitors provides an opportunity to explore the rich genetic resources of wheat for crop improvement (Avni et al., 2017; Luo et al., 2017; Zhao et al., 2017; Appels et al., 2018; Ling et al., 2018; Maccaferri et al., 2019).

Dissecting the genetic and molecular mechanisms regulating crop yield and growth is of great importance for crop improvement. Screening of induced mutant populations is an effective approach for discovering new genes underlying phenotypic variations in plants (Schneeberger, 2014; Krasileva et al., 2017). As a traditional forward genetic method, map-based cloning has been widely used to clone genes controlling traits of interest. Using map-based cloning approaches, several very important genes, such as vernalization (Yan et al., 2003; Yan et al., 2004), male sterility (Ni et al., 2017; Xia et al., 2017), and resistance genes (Fu et al., 2009), have been cloned. However, map-based gene cloning usually requires multiple steps, including generating mapping populations, fine mapping to narrow the target gene region to identify genetic markers cosegregating with the target phenotype, candidate gene screening and gene identification by sequencing. This process is often time-consuming and costly, especially in wheat.

With the great advances of next generation sequencing technologies, it is now feasible to dissect all genetic variants within a mutant line and then identify the causal variants associated with the target phenotype through combining bulked segregant analysis (BSA) with the whole genome sequencing strategy (Michelmore et al., 1991; Schneeberger, 2014). Mapping-by-sequencing has been widely implemented in model plants *Arabidopsis* and rice to screen the causal mutations underlying corresponding phenotypes from ethyl methanesulfonate (EMS) mutant populations (Austin et al., 2011; Lim et al., 2014; Schneeberger, 2014). To map the causal mutation, high-throughput sequencing is applied on bulked DNA samples pooled from individuals sharing the mutant and wild phenotype from an  $F_2$  population derived from the cross between the mutant line and a genetically divergent accession or backcross line (Schneeberger et al., 2009; Austin et al., 2011; Schneeberger, 2014). The successful application of methods such as MutMap, MutMap<sup>+</sup>, MutMap-Gap, SHOREmap and MMAPPR proved that combining BSA with high-throughput sequencing (BSA-Seq) is a feasible alternative to map-based cloning for new gene discovery (Schneeberger et al., 2009; Abe et al., 2012; Fekih et al., 2013; Hill et al., 2013; Takagi et al., 2013; Takagi et al., 2015; Yuan et al., 2017).

Although improved high-throughput sequencing technologies have significantly reduced cost and increased data quality, it is still too expensive to generate whole-genome sequence data for wheat

due to its large genome size and highly repetitive content (Choulet et al., 2010). To overcome this limitation, exome capture technologies aiming to cover the majority of the gene coding regions have been developed (Henry et al., 2014; Krasileva et al., 2017; He et al., 2019). Several wheat exome capture probe sets have been developed based on the previously released draft Chinese Spring genome sequence data (Krasileva et al., 2017). Recently, 890 diverse accessions of hexaploid and tetraploid wheat were re-sequenced through exome capture technology to reveal that wild-relative introgression through historic gene flow has made a significant contribution to the adaptive diversity of modern wheat (He et al., 2019). Exome capture has also been used for mapping and gene identification in wheat. A 110-Mb exome capture assay was used to identify candidate natural variants at the *Yr6* locus responsible for yellow rust resistance through mapping-by-sequencing (Gardiner et al., 2016). In addition, an 84-Mb exome capture assay was employed to uncover the coding region mutations in EMS-induced TILLING populations containing 2,735 mutant lines of tetraploid and hexaploid wheat (Krasileva et al., 2017). Using bulked segregant analysis and the exome sequence strategy, Harrington et al. identified a locus controlling a dominant, environmentally dependent chlorosis phenotype within the Durum wheat (*Triticum turgidum*) cv. Kronos (Harrington et al., 2019). Mo et al. (2018) identified a clear peak region on chromosome arm 4BS associated with increased plant height. Martinez et al. (2020) fine-mapped the ABA-hypersensitive mutant ERA8 in a wheat backcross population. Using RenSeq and MutRenSeq strategy, series resistance genes were cloned through developing the exome capture probe panel that specifically captures disease resistance (R) genes (Jupe et al., 2013; Steuernagel et al., 2016). Moreover, AgRenSeq was further developed in exploiting pan-genome variation and rapid R gene cloning by combining association genetics with R gene enrichment sequencing (Arora et al., 2019).

Integrating recently developed genome sequence and annotation resources, Gardiner et al. designed two gold standard capture probe sets for hexaploid bread wheat, with the final gene/putative promoter design space of 785,914,746 bp of which 508,889,665 bp was gene and 277,025,081 bp was putative promoter sequences (Gardiner et al., 2019). The application of exome capture assays not only provides us with a fast and convenient method to study domestication and evolution of wheat, but can also generate coding sequence information of a large number of genetic materials that can facilitate reverse genetic screening of a mutant line with known mutation sites. Currently, most wheat exome capture probe sets were designed prior to the recent release of high-quality reference genomes, with a large amount of sequences likely uncovered in the exome capture panels. Therefore, a new exome capture panel with a higher sequence coverage using the latest genome information would be more useful in a wide range of applications.

Leaf color has been extensively studied in rice and other plants because of its large effect on yield by affecting photosynthesis and is an efficient phenotypic marker in hybrid crops (Zhang et al., 2018). However, genetic and molecular characterization of leaf color genes are very limited in wheat,

largely due to its genome complexity and high proportion of repetitive sequences. Most studies have focused on phenotypic characteristics, physiological and biochemical characteristics, and genetic mapping (Wu et al., 2018). In this study, we identified and characterized a yellow-green leaf mutant (*yg11*) from a wheat EMS mutant population. To identify the causal mutation responsible for the phenotype, we designed a new wheat exome capture platform and developed an efficient statistical analysis model, named varBScore, to perform the BSA-Seq analysis and rapidly clone the *YGL1* gene responsible for the yellow-green leaf phenotype in wheat.

## MATERIALS AND METHODS

### Plant Materials

The yellow-green leaf mutant *yg11* originated from mutagenesis of 6,000 seeds of the elite wheat variety YZ4110, which was treated with 1.2% EMS; the germination rates of the EMS-mutagenized seeds were 65% (Zhao et al., 2009), and the yellow-green leaf phenotype was inherited stably after four generations of self-pollination. For the genetic analysis and/or gene mapping, *yg11* was backcrossed with YZ4110 to produce a BC<sub>3</sub>F<sub>2</sub> population consisting of 1,204 individuals. *yg11* was also crossed with a genetically divergent accession Bainong3217, generating an F<sub>2</sub> population that consisted of 431 individuals. The phenotype of all F<sub>2</sub> plants was further validated by using F<sub>3</sub> families. Wheat plants were cultivated in the experimental field of the Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China. In addition, all the transgenic plants were grown in a growth chamber maintained at 24°C, 16/8 h light/dark with 300 μmol m<sup>-2</sup> s<sup>-1</sup> light intensity at 45% humidity.

### Chlorophyll and Carotenoid Content Analysis, and Transmission Electron Microscopy Analysis

The content of chlorophyll and carotenoid was determined with a DU 800 UV/Vis Spectrophotometer according to the method described by Sheng et al. (2017), each sample included ten biological replicates. Chloroplast structures were examined in the four-week-old leaf samples of the wild-type YZ4110 and *yg11* mutant plants using transmission electron microscopy. Leaves were cut to sections approximately 5 mm in length and fixed in a 2% glutaraldehyde solution in phosphate buffer. Subsequently, the sections were fixed in a 1% OsO<sub>4</sub> solution, and the samples were stained with uranyl acetate, followed by dehydration through serial ethanol dilution. The thin sections were embedded in Spurr's medium. Finally, the samples were sliced to 50 nm in thickness and stained again, and then viewed with a JEOL 100 CX electron microscope.

### Exome Capture Design

The capture probe sequences were mainly based on the RefSeq Annotation v1.0, covering 107,891 high-confidence (HC) protein-coding genes and 161,537 low-confidence (LC) genes, 5'UTR and 3'

UTR sequences were included. To capture exons as much as possible in the wheat genome, over 2,600 RNA-seq data downloaded from Sequence Read Archive (SRA) database of the National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/sra/>) were aligned to the IWGSC CS RefSeq v1.0 genome with STAR(v2.6.1b) (Dobin et al., 2013). The novel transcriptome was assembled with Stringtie (v1.3.5) (Pertea et al., 2015) based on IWGSC annotation v1.1 with both HC and LC genes. The novel transcriptome (TMP >2) was used to predict ORFs using TransDecoder (v5.5.0) (Haas et al., 2013) with default parameters. All CDS and untranslated regions (UTR) of novel and HC/LC genes were extracted as redundancy capture targets sets. Basically, a histogram of the frequency of 15mers within the reference genome was constructed, and candidate probes were assigned a score based on probe sequence matches to all 15mers. Probes with scores higher than 100 were rejected. This set contained probes with up to 20 close matches to the genome, as determined by the SSAHA algorithm (Ning et al., 2001), for the purposes of providing better coverage. A probe was aligned to the genome if there were a total of five or fewer variations including single-base insertions, deletions or substitutions between the probe and the genome sequences.

### Exome Capture Sequencing

The genomic DNA of 30 extreme phenotype individuals from different lines of the Bainong3217/*yg11* F<sub>3</sub> population was extracted by means of CTAB method (Chatterjee et al., 2002), and then the genomic DNA was bulked in an equal ratio to generate the wild-type bulked DNA pool and mutant-type bulked DNA pool separately. A total of 4.5 mg genomic DNA (35 ng/ml) was fragmented into 200 bp using the Bioruptor UCD-200 sonicator (Diagenode, Denville, NJ). Pre-libraries were constructed using the Kapa Hyper DNA library prep kit for Illumina HiSeq. Fragmented DNA was end-repaired with an end-repair enzyme, and a deoxyadenosine was added to the 3' ends of the fragments. Kapa barcoded DNA and Kapa indexed adapters were ligated to the sample libraries. The adapter-ligated libraries were selected for an average insert size of 200 bp using next-generation sequencing cleanup and size select (NucleoMag, Macherey-Nagel, Duren, Germany) according to the manufacturer's instructions. The pre-capture amplification was performed using the NimbleGen protocol with eight PCR cycles. Hybridization of sample libraries was performed at 47°C for 60–72 h using the SeqCap EZ library (Roche/NimbleGen, Madison, WI) and custom-designed exome capture probes. The hybrids were enriched using the capture beads from the SeqCap EZ pure capture bead kit, washed, and amplified by ligation-mediated PCR. The quality of the captured libraries was assessed using the Bioanalyzer 4200 (Agilent Technologies, Santa Clara, CA). The libraries were quantified by qRT-PCR and then sequenced using the Illumina HiSeq Nova platform to generate 150-bp paired-end reads.

### The BSE-Seq Pipeline for Rapid Gene Cloning

With the newly designed exome capture probe panel and varBScore, we developed a BSA-based exome capture sequencing

(BSE-Seq) pipeline for rapid gene cloning in wheat. This approach consists of the following steps (**Figure 1**): (1) Develop a population derived from the cross between the mutant line and a genetically divergent accession or a backcross to its parental wild-type plant line. (2) DNAs from the two selected groups showing extreme phenotypes are extracted and bulked in equal amounts, forming two sample pools. (3) The two DNA pools are subjected to exome capture sequencing with deep coverage ( $\sim 70\times$ ); raw sequencing reads are processed using Fastp (v0.12.4) to remove low quality reads and adapter sequences. (4) BWA (v0.7.16) mem is applied to align the high-quality reads to the IWGSC RefSeq v1.0 genome with default parameters. (5) Samtools (v1.9) is used to BAM sort, remove PCR duplicated reads and generate sample alignment statistics. (6) The raw cohort vcf is analyzed with GATK (v4.0.10.1) module. The minimum-mapping-quality parameter is set as 30 for high-quality alignment reads that are used to call variants. (7) bcftools (v1.9) is performed as variants quality filtering with “QUAL>30” and “DP $\geq$ 5”. To annotate variants, customized databases containing IWGSC v1.1 HC/LC genes and annotated genes are created by snpEff (v4.3T). (8) The new statistical model varBScore is then introduced to accurately determine the position of the target candidate gene. (9) The 2,600 RNA-seq data were used as a “background” to identify the causal mutation based on the assumption that the genetic materials carrying the same mutation may have the same phenotype. Using the yellow-green leaf color mutant *ysl1* used in this study as an example, we searched the candidate mutants in the background which is composed of materials with green leaf phenotype. As result, the mutations that appeared in the

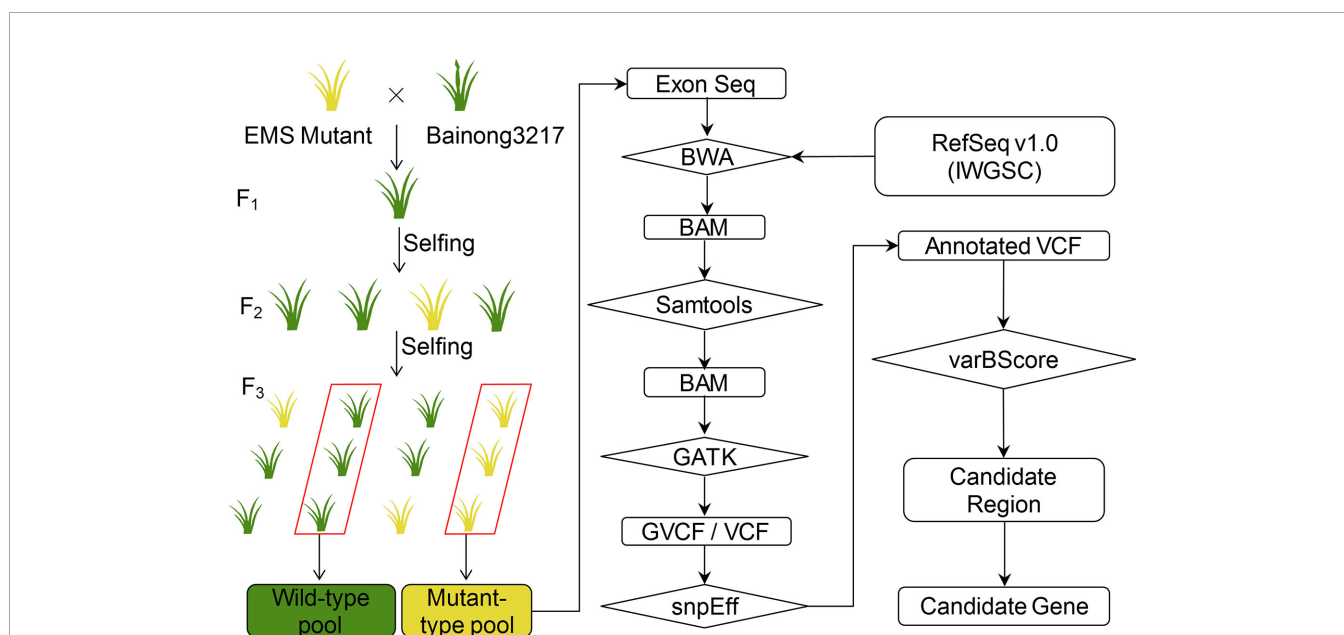
background would be filtered out, which is helpful in identifying the causative mutation.

## Data Processing

Raw sequence reads were processed using Fastp (v0.12.4) (Chen et al., 2018) to remove low quality reads and adapters. BWA (v0.7.16) (Li and Durbin, 2009) mem was applied to align the high-quality reads to IWGSC RefSeq v1.0 genome with default parameters. Samtools (v1.9) (Li et al., 2009) was used to bam sort, remove read PCR duplications and for sample alignment statistics. After that, raw cohort vcf was worked out with GATK (v4.0.10.1) (McCormick et al., 2015) module BaseRecalibrator, ApplyBQSR, HaplotypeCaller, CombineGVCFs and GenotypeGVCFs were used. The minimum-mapping-quality parameter was set as 30 for only high-quality alignment reads used to call variants. bcftools (v1.9) (Narasimhan et al., 2016) was performed for variants quality filtering with “QUAL > 30” and “DP  $\geq$  5”. To annotate variants, customized databases containing IWGSC v1.1 HC/LC genes were generated by snpEff (v4.3T) (Cingolani et al., 2012).

## Phylogenetic Analysis

All the protein sequences were obtained from Ensembl Plants ([http://plants.ensembl.org/Triticum\\_aestivum/Tools/Blast](http://plants.ensembl.org/Triticum_aestivum/Tools/Blast)) using the YGL1 protein sequence as a query, and the homoeologs that shared more than 60% identity with YGL1 were selected for analysis. After alignment of amino acid sequences by ClustalW with default parameters, the phylogenetic tree was constructed using MEGA 7.0 (Kumar et al., 2016) based on the neighbor-joining method, and the bootstrap values were estimated with 1,000 replicates.



**FIGURE 1** | Flow chart of the BSE-Seq analysis method. The seeds of a wheat cultivar were mutagenized by ethyl methanesulfonate (EMS). Then, the mutant was crossed to the genetically divergent accession Bainong3217. The resulting  $F_1$  plants were self-pollinated to obtain  $F_2$  segregating progeny. DNAs of  $F_3$  displaying the wild-type and mutant phenotypes were bulked and subjected to exome capture sequencing followed by alignment to the reference sequence. Then, the SNPs were called and analyzed to identify the causal SNP for the mutant phenotype.



## Transgene Construction

The target sites of guide RNA were designed by a web-based toolkit E-CRISP (Heigwer et al., 2014). A binary vector *TaU3p-413* that can carry two guide RNAs and the intermediate vector *pCBC-MT1T2* were used to create the CRISPR construct (Xing et al., 2014) by the Golden gate method (Engler et al., 2008). Briefly, the target site (5'-CAACAGGGGGATACTGTATG-3') was incorporated into both PCR forward and reverse primers (Supplementary Table S3). Then the PCR fragment with the target site that was amplified from *pCBC-MT1T2* was assembled into the binary vector *TaU3p-413* using *BsaI* and T4 Ligase (New England Biolabs). Then the plasmid was transformed into the *Agrobacterium tumefaciens* strain EHA105 and introduced into immature embryos of the wheat cultivar 'Fielder' via an *Agrobacterium tumefaciens*-mediated transformation method using licensed protocols of 'PureWheat' (Ishida et al., 2015).

## RESULTS

### Development of a New Wheat Exome Capture Panel

The exome capture panel was developed based on the IWGSC RefSeq v1.0 sequence information (Appels et al., 2018) and the Roche NimbleGen probe synthesis technology (OID46587, IRNSOW1123). A total of 299,387,477 bp regions were selected in the probe design, mainly including two aspects: (1) the 107,891 High Confidence (HC) and 161,537 Low Confidence (LC) genes containing the sequences from the 5'UTR (untranslated region) to the 3'UTR, which come from RefSeq Annotation v1.1, and (2) novel transcripts annotated from 2,600 RNA-seq data downloaded from the Sequence Read Archive (SRA) database of the National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/sra/>); these transcripts were highly expressed but beyond the scope of RefSeq Annotation v1.1. Approximately 54,241,943 bp regions were absent due to the unused sequence including gaps and sequence repeats, and a final probe length of 245,145,534 bp was accumulated. In addition, a single 120 bp probe can enrich up to 500 bp routinely with adequate sequencing coverage, with

**TABLE 1** | Summary of the new designed 260 Mb exome capture panel.

Genome	IWGSC_ReferSeq_v1.0	
Number of regions	680,432	
Length of regions (bp)	299,387,477	
Statistics	Probe Coverage	Estimated Coverage
Target Bases Covered	245,145,534	277,596,390
% Target Bases Covered	81.88	92.72
Targets with no coverage	37,897	36,717
Target Bases Not Covered	54,241,943	21,791,087
Due to sequence gaps	144,188	64,900
Due to repeats	21,620,096	142,93,102
% Target Bases Not Covered	18.12	7.28
Due to sequence gaps	0.05	0.02
Due to repeats	7.22	4.77
Total capture targets	1,806,109	
Total capture space (bp)	268,933,489	

the flanking sequence besides the exons could be captured simultaneously (Gardiner et al., 2015), the coverage length is approximately 277,596,390 bp estimated by a 200 bp fragment library (Table 1). We aligned the 245,145,534 bp probe sequences to the reference genome sequences from IWGSC annotation v1.1 and identified 107,400 high confidence genes and 132,688 HC transcripts.

### Rationale Behind varBScore

One of the challenges of sequencing-based methods to identify causal mutation sites is the high background noise, attributable to the large number of resulting potential SNPs and sequencing/mapping errors, particularly in large and complex genomes such as wheat. We developed a new sequence data analysis model, varBScore, which calculates the variance between the observed allele frequency and the expected allele frequency in a sliding window, which could reduce the noise in the process of mapping the causative mutation to a large extent.

Theoretically, according to the Mendelian inheritance, for a population generated from the selfing of a heterozygous parent which carries plenty of SNPs, every SNP locus has a distribution of AA : Aa : aa = 1:2:1. For instance, a stable EMS-induced mutant which is controlled by a recessive gene crossed with its wild-type and generating a heterozygote, then producing an F<sub>2</sub> population by selfing, can be separated into two pools by phenotype: the wild-type pool, which contains homozygotes (AA) and heterozygotes (Aa), and the mutant-type pool, composed of homozygous mutants (aa). The expected allele frequency of the causative mutation in wild-type pool (AA : Aa = 1:2) would be 1/3; similarly, its expected allele frequency in mutant-type pool (aa) would be 1. The expected allele frequency of other SNPs would be related to their linkage to the causative mutation, and those SNPs that have a close linkage with the causative mutation may share the similar expected allele frequency with the causative mutation. The varBScore algorithm screens the segments that meet the expected allele expectations by calculating the Standard Deviation (SD) between the expected allele frequency and the observed allele frequency of every SNP.

The varBScore can be calculated as follows:

$$\text{var } BScore_k^j = -\log_{10} \left( \sqrt{\frac{+(M_k^j - \bar{m})^2}{n-1}} + c \right) \\ * -\log_{10} \left( \sqrt{\frac{+(W_k^j - \bar{w})^2}{n-1}} + c \right)$$

Frequency of alternate alleles (AF) different > *df* between wild type and mutant type,  $df = |M_k^j - W_k^j|$ .

$M_k^j$ : *j* locus AF of mutant bulk;  $\bar{m}$ : expectation AF of mutant bulk;  $W_k^j$ : *j* locus AF of wild bulk;  $\bar{w}$ : expectation AF of wild bulk; *n*: number of SNP in windows *k*; *c* = 0.01, *c* = 0.001, *c* = 0.0001...;

### Validation of varBScore on a Known Mutant in Wheat

*Yr7* is a major Yellow stripe rust (YR) resistance gene that is identified in wheat. Gardiner et al. (2020) firstly used a

traditional QTL mapping method and mapped *Yr7* to a low-recombination region near the centromere of Chromosome 2B by a double haploid (DH) population, then the association mapping panel and mapping-by-sequencing strategy were then conducted to fine map the *Yr7* gene. Finally, a 60 Mb region that contains the *Yr7* locus was identified on chromosome 2B (633,430,001–663,920,001 bp, 681,790,001–693,030,001 bp and 702,660,001–721,190,001 bp). The candidate region contains 589 high confidence genes; 10 disease resistance genes that are mainly clustered together in the region 683,043,955–686,815,417 bp (~4 Mbp) are considered as candidate genes for *Yr7* resistance.

To test the varBScore algorithm, we download the BSA data set of *Yr7* (<https://www.ebi.ac.uk/ena>, study PRJEB36010) and then analyzed the data by using the varBScore algorithm, which showed a very distinct peak on chromosome 2B along each chromosomal pseudomolecule (Figure 2A). Further analysis showed that the highest scoring points were located on the regions (586,775,469–88,621,811 bp, 655,257,794–658,193,278 bp, and 680,759,276–682,847,904 bp) of chromosome 2B (Figure 2B), and the third peak region was close to the candidate region. In our result, the candidate gene could be mapped to such a region without any other approaches, indicating that varBScore could be effective on gene mapping.

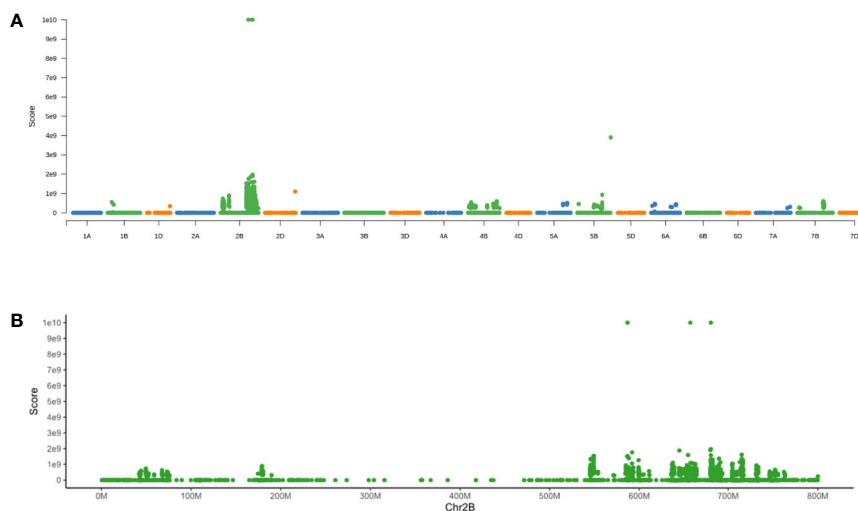
## Rapid Cloning of a Wheat Yellow-Green Leaf Gene *YGL1*

To further verify the effectiveness of BSE-Seq, a chlorotic leaf phenotype mutant *ygl1*, obtained by ethyl methanesulfonate (EMS) mutation of wheat variety Yanzhan4110 (YZ4110), was used to conduct the analysis. Compared with the wild type, *ygl1* exhibited obvious yellow leaf color from seedling to mature stage (Figures 3A, B). The contents of chlorophyll *a* (Chl *a*), chlorophyll *b* (Chl *b*), and carotene were significantly reduced

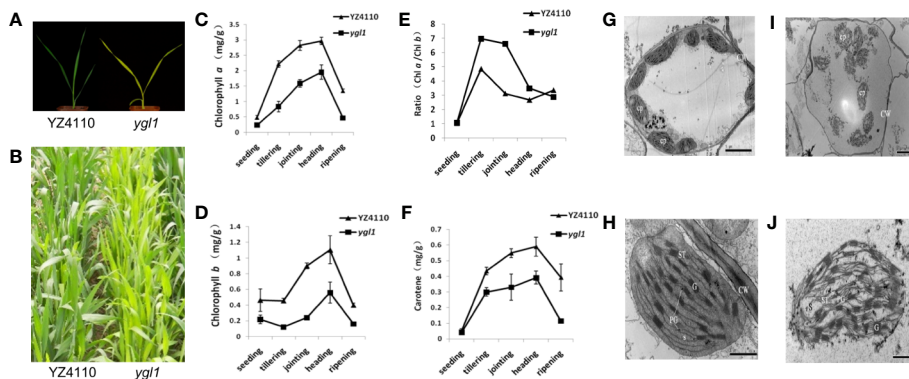
(Figures 3C–F). Transmission electron microscopy (TEM) analysis revealed that the shape and size of the chloroplast of the wild type and the mutant were not significantly different. However, in the mutant, the outer membrane structures of many chloroplasts are broken, the inner chloroplast matrix and thylakoid are exposed, and the number of normal chloroplasts is less than that of wild-type plants (Figures 3G–J). Genetic analysis of 1,204 BC<sub>3</sub>F<sub>2</sub> population plants derived from a backcross between the mutant line *ygl1* and wild-type YZ4110 showed that the yellow leaf phenotype of *ygl1* was controlled by a recessive nuclear gene (Supplementary Table S1).

We first developed a F<sub>2:3</sub> segregating population (431 individuals) derived from the cross between *ygl1* and a genetically divergent accession Bainong3217. The yellow and green pools bulked from the homozygous genotype of 30 yellow and 30 green F<sub>3</sub> individual plants were constructed, respectively. Exome capture and high-throughput sequencing were conducted and generated approximately 20 Gb of sequence data for each pool. Compared with the Chinese Spring reference genome, a total of 777,780 and 792,839 sequence variations were detected in yellow and green pools, respectively. By projecting these variations to chromosomes, we found that the most significant varBScore was at the 671 Mb position on chromosome 7A (Figure 4A).

Given the characteristics of EMS mutagenesis, certain variations should be filtered out based on the following criteria (Figure 4B): (1) Allelic differences between the yellow and green pools. A total of 3,604 SNPs were discovered, and the SNPs that had allelic differences (the green pool had an identical genotype with the reference genome, and the yellow pool had a mutant genotype, or reverse) between pools were 1,731. (2) Because EMS-type mutations (base changes from C to T or G to A) constitute over 99% across the population (Krasileva et al., 2017), approximately 542 EMS-type mutations were retained. (3) Omitting mutations that could lead to missense, loss of start



**FIGURE 2 |** Results for mapping of *Yr7* by varBScore. **(A)** Manhattan plot of the varBScore along each chromosome. **(B)** Manhattan plot of the varBScore along chromosome 2B.

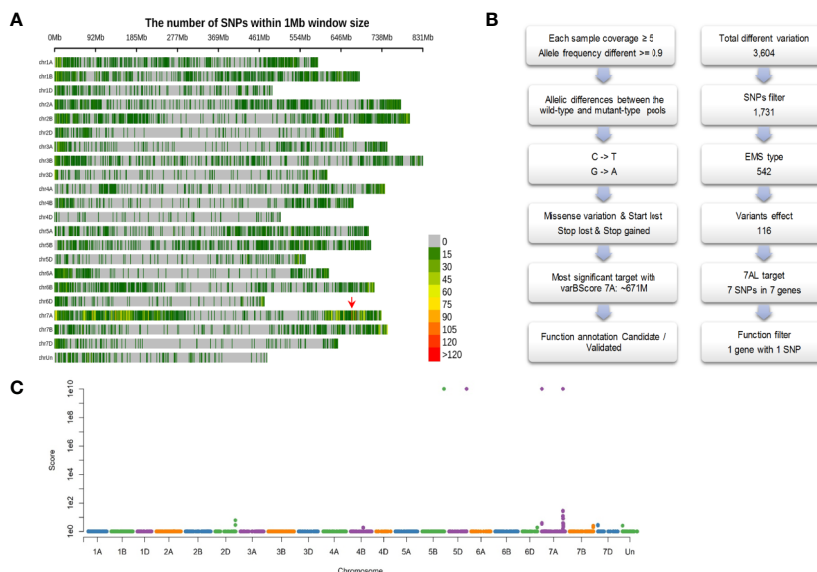


**FIGURE 3 |** Characterization of wheat yellow-green leaf mutant *ygl1*. **(A, B)** Phenotype of YZ4110 and the *ygl1* mutant at seedling and mature stage, respectively; **(C, F)** Chlorophyll a **(C)**, Chlorophyll b **(D)**, the ratio of Chlorophyll a and Chlorophyll b **(E)** and carotene contents **(F)** in leaves of the YZ4110 and the *ygl1* mutant; Data indicate average  $\pm$  standard deviation (SD) (n = 10). **(G, J)** Transmission electron microscopy analysis of chloroplasts in the YZ4110 **(G, H)** and the *ygl1* mutant **(I, J)**. CP, chloroplast; PG, plastoglobule; G, grana; S, stroma; CW, cell wall; ST, stroma thylakoid. Scale bar = 5  $\mu$ m in G and H, scale bar = 1  $\mu$ m in I, J.

codon, loss of stop codon and gain of coding amino acids, and 116 SNPs were left. (4) Further screening was conducted by filtering out the SNPs that could be detected in the “background”, which is comprised of the materials with green leaf; this step could filter out the SNPs generated from the sequencing errors, mapping errors, homoologous genes and the same variations existing in the genetic materials. Through the above filtering steps, we finally obtained the seven most significant SNPs on chromosome 7A at the 671 Mb position (**Figure 4C, Table 2**). Functional annotation of the genes carrying these seven SNPs (**Table 2**) revealed that *TraesCS7A02G480700* (**Figure 5A**)

encoding the chloroplast magnesium-chelatase subunit *chlI*, a key enzyme in chlorophyll synthesis and chloroplast development, may be involved in the regulation of leaf color. Sequence analysis indicated that *TraesCS7A02G480700* encodes an ortholog of *OsChlI*, whose single-recessive mutants exhibiting a yellowish-green leaf color (Zhang et al., 2006) (**Supplementary Figure S1**), which has a similar phenotype with the *ygl1* mutant.

We further examined the genotypes of these seven SNPs in both parents and progenies of the backcross population. The results showed that the genotype of *TraesCS7A02G480700* in wild-type YZ4110 and 43 homozygous F<sub>3</sub> individual plants with green leaves



**FIGURE 4 |** Identification of the candidate causal mutation. **(A)** The distribution of SNPs and significant candidate selected region (indicated with red arrow) across the 21 wheat chromosomes; **(B)** The principle of filtering variation based on EMS mutagenesis characteristics; **(C)** Manhattan plot of varBScore along each chromosomal pseudomolecule.

was G at the position 921, while it was A in the mutant *ysl1* and 34 homozygous F<sub>3</sub> individual plants with yellow leaf phenotype. The rest of the SNPs have some recombination events among the progenies, which means they are not casual mutations.

### Verification of the Candidate Gene *YGL1* Through Genome Editing With CRISPR/Cas9

A genome-editing experiment using the CRISPR/Cas9 system was further implemented; a single guide RNA was designed to target the site within a conserved region of three *YGL1* homoeologous genes (*TraesCS7A02G480700*, *TraesCS7B02G382800*, *TraesCS7D02G467500*) on chromosomes 7A, 7B, and 7D genomes (Figure 5A); the primers used for vector construction and editing identification are listed in Supplementary Table S3. In the transgenic lines, 22 out of 42 T<sub>0</sub> plants were found to carry the mutant allele in *YGL1* homoeologous genes. Among these 22 gene-edited transgenic plants, TW181-36, TW181-3, and TW181-1 plants were homozygous or bi-allelic mutants at the edited sites of the 7A, 7B, and 7D genomes, respectively (Supplementary Figure S2, Supplementary Table S2), and these plants showed chlorotic leaf phenotype like the mutant *ysl1* (Figure 5B), indicating the three *YGL1* homoeologs have conserved functions. In addition, the seedlings of the plants (TW181-6, 34, 35, 43) with homozygous or bi-allelic mutants in all three genomes also showed chlorotic leaf phenotype and were shorter and weaker than other plants, suggesting that the function of the three genes had a dose effect (Figure 5B, Supplementary Table S2). The remaining plants that carried heterozygous mutations or the non-gene-edited plants showed green leaf phenotypes similar to that of the wild type. These results indicate that the targeted mutation co-segregated with the phenotypic variation of chlorosis leaf and demonstrate that *TraesCS7A02G480700* is the functional gene of *YGL1*.

### varBScore Can Identify Causal Mutations With Less Background Noise Than SNP-Index

We compared varBScore with the popular SNP-Index method (Abe et al., 2012) in identifying potential causal SNPs for the *ysl1*

mutant. As shown in Figure 6A, the  $\Delta$ SNP-index has very dramatic changes in the multiple regions of 0–10 M, 90–200 M, and 600–700 M on chr7A. Due to the much higher noise level (several sites were close to 1 or –1) with SNP-Index, the SNP-Index method could not determine the candidate region. In contrast, the results with varBScore show only two distinct candidate segments (Figure 6B), and one of the two segments was the candidate gene target.

## DISCUSSION

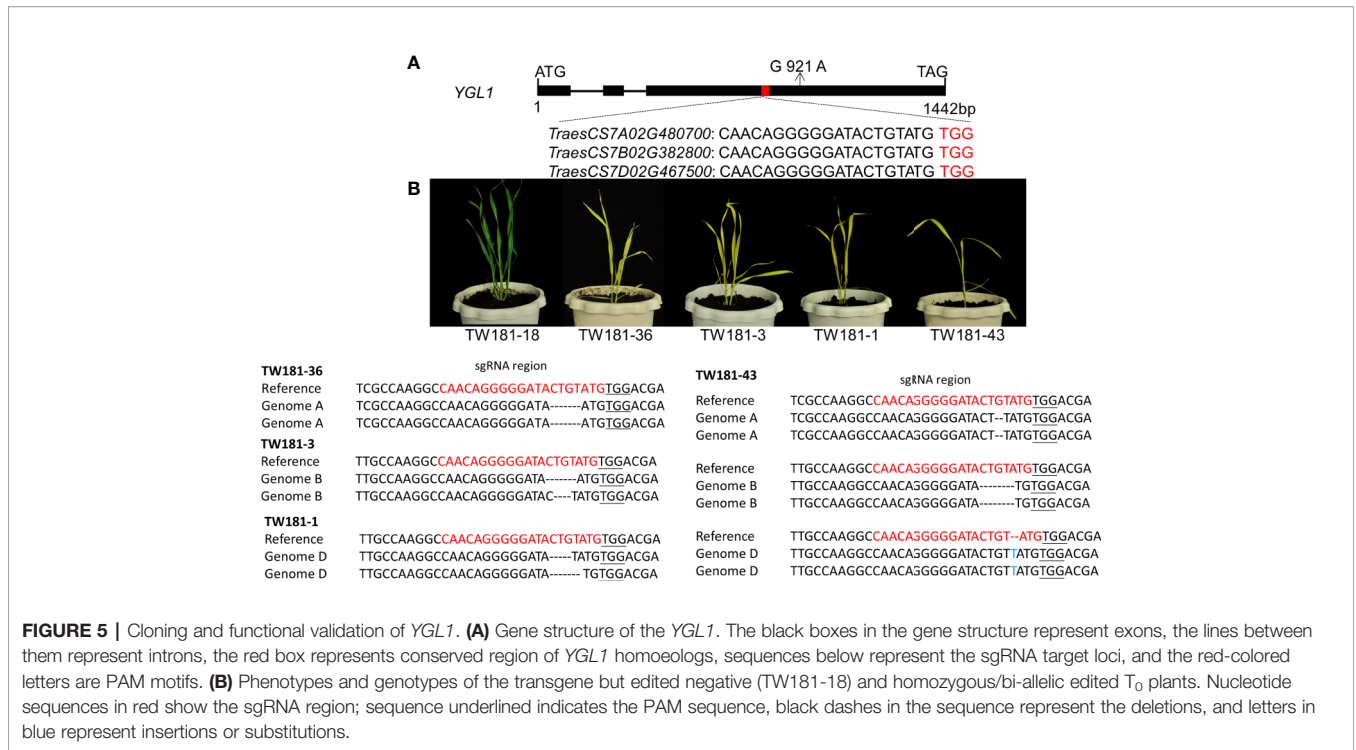
In the present study, we devised a BSA-Seq method based on a newly designed wheat exome capture panel and the varBScore algorithm and successfully applied it to clone the causal mutation in yellow-green leaf mutant *ysl1* in a much faster and efficient manner compared to the conventional map-based cloning strategy. Like other BSA-based cloning methods, our approach started with construction of a population derived from a cross between the isolated mutant line and a genetically divergent accession, followed by the use of an F<sub>3</sub> population to bulk the green and yellow samples for exome capture sequencing with approximately 70× average depth. We then implemented a novel algorithm model, varBScore, and a series of bioinformatic analyses to fine-map *YGL1* within seven genes. Through functional annotation of these seven genes and validation of the gene function by the CRISPR/Cas9 method, we confirmed that the magnesium-chelatase subunit chlI encoded by *TraesCS7A02G480700* was the candidate gene carrying the causal mutation for yellow leaf phenotype of *ysl1*. Thus, BSE-Seq is a time- and cost-saving strategy for discovering the causal mutation responsible for the mutant phenotype.

Despite the considerable sequencing cost reduction due to the increasing data volume per sequencing run in different high-throughput sequencing platforms, it is still quite expensive to perform the genetic variation analyses in wheat by whole genome resequencing. In wheat, the over 90% sequence identity in coding regions from the A, B, and D homoeologous genomes as well as the high content of repetitive DNA and large retroelements has

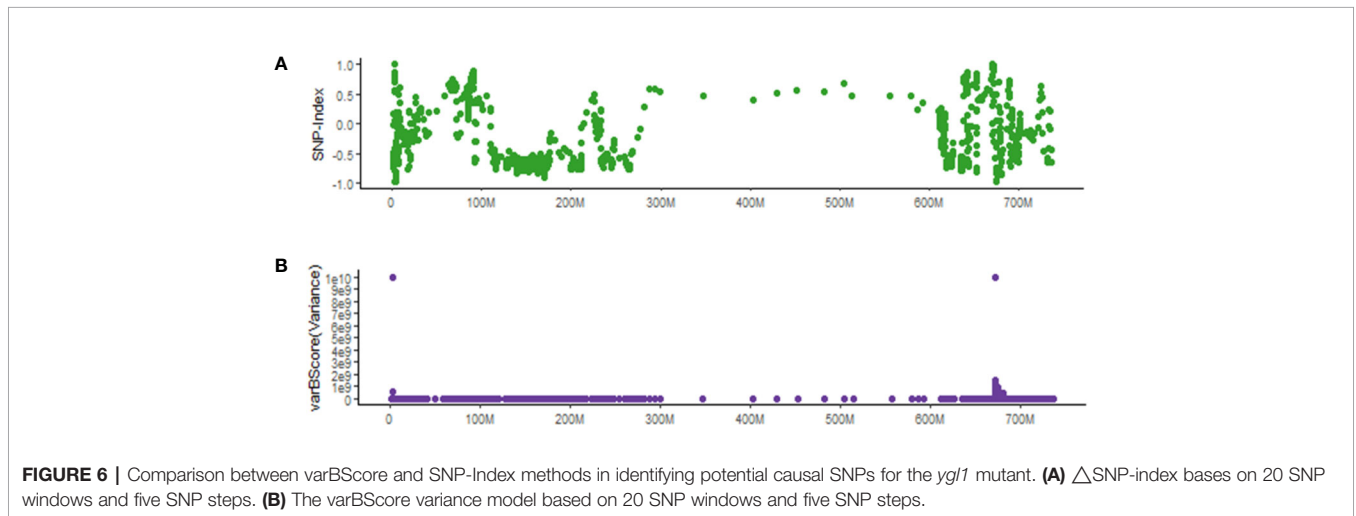
**TABLE 2 |** Candidate genes and nucleotide variations of targets with the most significant varBScore (Chr7A: –671 M).

Chr	Pos	Reference	mutant base	Feature	Gene	Wildtype Pool	Mutant Pool	Functional Annotation
chr7A	630,870,068	G	A	Missense variant	<i>TraesCS7A02G436400</i>	5 0	0 5	Ethylene-responsive transcription factor
chr7A	655,451,742	C	T	Missense variant	<i>TraesCS7A02G460100</i>	31 0	2 32	Protease inhibitor/seed storage/lipid transfer protein family protein
chr7A	66,2134,333	C	T	Missense variant	<i>TraesCS7A02G465600</i>	11 0	0 6	Ubiquitin family
chr7A	662,143,366	C	T	Missense variant	<i>TraesCS7A02G465700</i>	38 0	1 19	Transcription factor FapR
chr7A	672,873,436	C	T	Missense variant	<i>TraesCS7A02G480700</i>	34 0	0 53	Mg-protoporphyrin IX chelatase
chr7A	680,628,523	C	T	Missense variant	<i>TraesCS7A02G492000</i>	18 0	1 10	ATP-dependent zinc metalloprotease
chr7A	682,175,198	C	T	Missense variant	<i>TraesCS7A02G493600</i>	4 0	0 4	Heat shock family protein





**FIGURE 5 |** Cloning and functional validation of *YGL1*. **(A)** Gene structure of the *YGL1*. The black boxes in the gene structure represent exons, the lines between them represent introns, the red box represents conserved region of *YGL1* homoeologs, sequences below represent the sgRNA target loci, and the red-colored letters are PAM motifs. **(B)** Phenotypes and genotypes of the transgene but edited negative (TW181-18) and homozygous/bi-allelic edited  $T_0$  plants. Nucleotide sequences in red show the sgRNA region; sequence underlined indicates the PAM sequence, black dashes in the sequence represent the deletions, and letters in blue represent insertions or substitutions.



**FIGURE 6 |** Comparison between varBScore and SNP-Index methods in identifying potential causal SNPs for the *ygl1* mutant. **(A)**  $\Delta$ SNP-index bases on 20 SNP windows and five SNP steps. **(B)** The varBScore variance model based on 20 SNP windows and five SNP steps.

brought extra challenges for analyzing the large body of data (Uauy, 2017). Exons constitute approximately 0.95% of the wheat genome but code for the primary protein components that play essential roles in plant growth and development. Therefore, exome capture is an alternative method to effectively sequence coding regions with low cost in wheat. Our new wheat exome capture probe set with an accumulated probe length of 245 Mb can cover 277 Mb gene regions considering the capture of the flanking sequences including promoter regions, untranslated regions neighboring to the exons, and introns.

SNP-index and  $\Delta$ SNP-index have been widely used for identifying potential causal SNPs. Compared to  $\Delta$ SNP-index,

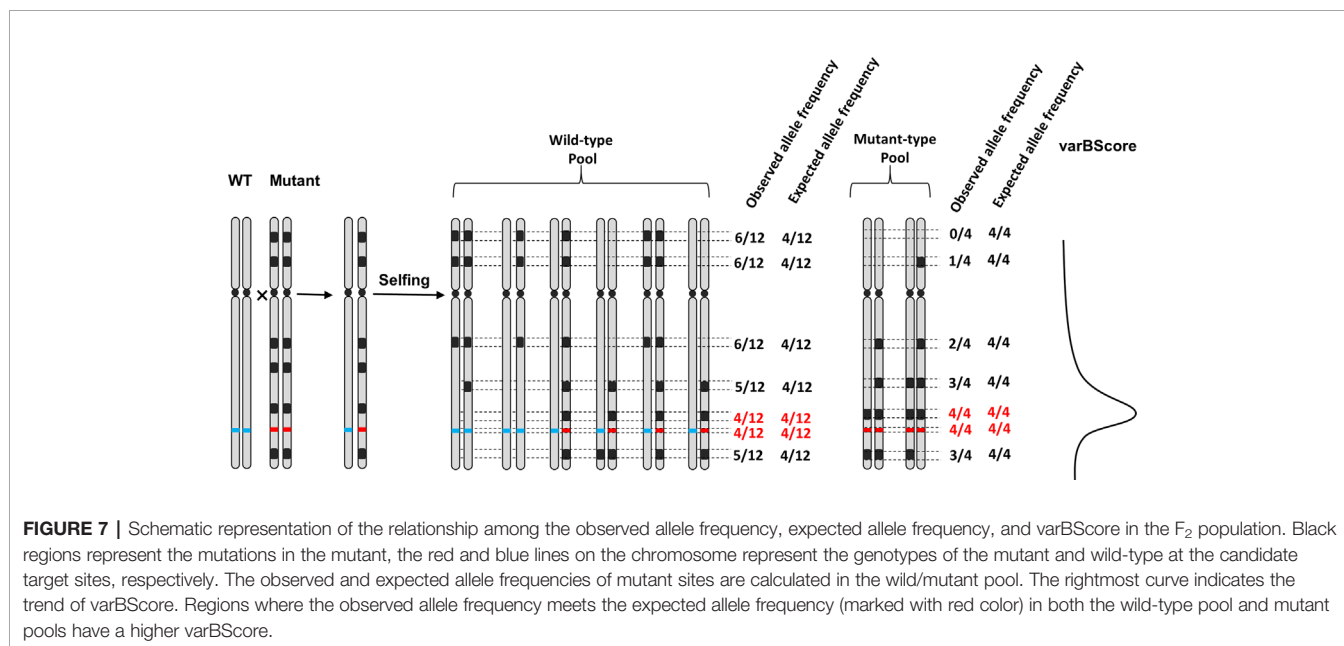
which is directly obtained from the subtraction of observed allele frequency between bulked pools, the expected allele frequency is introduced as another parameter in varBScore algorithm, which calculates the standard deviation (SD) of the observed allele frequency and expected allele frequency, leading to a higher accuracy in evaluating the degree of variation. According to the linkage inheritance rule, molecular markers close to the gene of interest on a chromosome tend to be inherited together during meiosis, thus they may share a similar expected allele frequency with the gene of interest. Therefore, the minimum variation occurs only when the observed allele frequencies of the mutant and wild-type pool both meet their

expected allele frequency simultaneously. By evaluating the logarithm of the variation, then the varBScore would be higher than those regions that do not satisfy the condition (Figure 7). As a result, the variances in the bulked pool that meet the expected allele frequency would show a peak in the varBScore plot. The advantage of the varBScore is that standard deviation was used to evaluate the variation, which would be more precise than the subtraction. This algorithm can reduce the background noise to some extent and provides an alternative method to identify causal mutations.

Gene-editing has emerged as a powerful tool for functional gene validation and mutant creation in both animals and plants (Chen et al., 2019). However, a problem that cannot be ignored is that there is a chance of off-target editing, which may introduce some unexpected phenotypes (Jin et al., 2019). Selecting specific target sites for editing is a better strategy to avoid or reduce the problem of off-target editing. To design specific target sites for gene editing, we aligned the sequence of the target locus with the wheat reference genome and found that the sequence did not significantly align with any other genomic regions in wheat, indicative of a relatively low possibility of off-target editing. As we expected, our observation of  $T_0$  plants carrying the homozygous mutations showed that the targeted mutation co-segregated with phenotypic variation of chlorosis leaf (Figure 5B, Supplementary S2). Therefore, the functional validation supports the hypothesis that *TraesCS7A02G480700* is the functional gene *YGL1*. As a hexaploid crop, many genes have three homoeologs present in each of the three wheat genomes of bread wheat. The three homoeologs could be functionally redundant or have additive effects. We showed that a mutation in any of them produced a yellow leaf phenotype like that of the *ysl1* mutant and simultaneous knock-out of all three homoeologs of *YGL1* not only conferred the

yellow phenotype of the edited plant but also showed slower and weaker growth than the plant with only one of copy edited (Figure 5B). These results suggested that the three homoeologs share the same function but may also have additive effects.

Combining BSA and the exome sequence strategy could accelerate gene mapping, especially in polyploid species with large and complex genome such as wheat. First, although generating the segregant populations is still a rate-limiting step compared with technologies like MutRenSeq and AgRenSeq (Steuernagel et al., 2016; Arora et al., 2019), BSE-Seq could be easily used to identify the linked interval regardless of the multiple gene copies and high similarity among the homoeologs. An additional benefit is that whole-exome capture sequencing could obtain most of the variations existing in the coding regions of genes, which is helpful for the construction of linkage map across the whole genome. Specifically, capturing R genes leads to the reduction of data analysis and the cost of sequencing, while it carries the risk of missing of those non-typical NLRs resistance genes. In this sense, BSE-Seq has a wider range of adaptability. Secondly, compared with the 84-Mb and 110-Mb exome capture probe panels developed previously (Gardiner et al., 2016; Krasileva et al., 2017), the newly developed panel in our research or larger capture probe sets designed recently (Gardiner et al., 2019) could capture more sequence information, leading to a higher resolution in gene mapping and identification; even if the mutant gene was not captured, SNPs along the two flanks of the candidate gene would be sufficient to identify the linked interval. Thirdly, the varBScore algorithm raised here provides an alternative method to identify causal mutations and fits for various segregant populations including  $F_2$ ,  $F_{2,3}$ , RILs (recombinant inbred lines), and DHs (doubled haploids), which may have a wider application in gene mapping and identification.



## DATA AVAILABILITY STATEMENT

Customized exome capture design files and inhouse scripts will be made available by the authors upon request. Sequence data have been submitted to GenBank under Bioproject no. PRJNA579285.

## AUTHOR CONTRIBUTIONS

XK, LZ, ZC, and XL conceived the study and designed the experiments. CX, CD, DL, ZX, QZ, and XZ, performed the experiments. ZC and LG performed the data analysis. JW provided the computing resources. YG, LZ, ZC, and XK wrote the manuscript. All authors contributed to the article and approved the submitted version.

## REFERENCES

- Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., et al. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotech.* 30 (2), 174–178. doi: 10.1038/nbt.2095
- Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361, eaar7191. doi: 10.1126/science.aar7191
- Arora, S., Steuernagel, B., Gaurav, K., Chandramohan, S., Long, Y. M., Matny, O., et al. (2019). Resistance gene cloning from a wild crop relative by sequence capture and association genetics. *Nat. Biotechnol.* 37, 139–143. doi: 10.1038/s41587-018-0007-9
- Austin, R. S., Vidaurre, D., Stamatiou, G., Breit, R., Provart, N. J., Bonetta, D., et al. (2011). Next-generation mapping of Arabidopsis genes. *Plant J.* 67 (4), 715–725. doi: 10.1111/j.1365-313X.2011.04619.x
- Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S. O., Gundlach, H., et al. (2017). Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science* 357 (6346), 93–97. doi: 10.1126/science.aan0032
- Chatterjee, A., Moulik, S. P., Majhi, P. R., and Sanyal, S. K. (2002). Studies on surfactant-biopolymer interaction. I. Microcalorimetric investigation on the interaction of cetyltrimethylammonium bromide (CTAB) and sodium dodecylsulfate (SDS) with gelatin (Gn), lysozyme (Lz) and deoxyribonucleic acid (DNA). *Biophys. Chem.* 98 (3), 313–327. doi: 10.1016/s0301-4622(02)00107-2
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34 (17), i884–i890. doi: 10.1093/bioinformatics/bty560
- Chen, K., Wang, Y., Zhang, R., Zhang, H., and Gao, C. (2019). CRISPR/Cas Genome Editing and Precision Plant Breeding in Agriculture. *Annu. Rev. Plant Biol.* 70, 667–697. doi: 10.1146/annurev-arplant-050718-100049
- Choulet, F., Wicker, T., Rustenholz, C., Paux, E., Salse, J., Leroy, P., et al. (2010). Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. *Plant Cell* 22 (6), 1686–1701. doi: 10.1105/tpc.110.074187
- Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain *w<sup>1118</sup>*; *iso-2*; *iso-3*. *SnpEff. Fly* 6 (2), 80–92. doi: 10.4161/fly.19695
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29 (1), 15–21. doi: 10.1093/bioinformatics/bts635
- Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. *PLoS One* 3, e3647. doi: 10.1371/journal.pone.0003647

## FUNDING

We thank Prof. Qixin Sun's group for providing the CRISPR/Cas9 vector system. This work was supported by the National Transgenic Research Project (2016ZX08009001-001-004), the National Key Research and the Development Program of China (2016YFD0101001), the Agricultural Science and Technology Innovation Program of Chinese Academy of Agricultural Sciences (CAAS) and the Tianshan Innovation Team Plan (2020D14002).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2020.01249/full#supplementary-material>

- Fekih, R., Takagi, H., Tamiru, M., Abe, A., Natsume, S., Yaegashi, H., et al. (2013). MutMap+: genetic mapping and mutant identification without crossing in rice. *PLoS One* 8 (7), e68529. doi: 10.1371/journal.pone.0068529
- Fu, D., Uauy, C., Distelfeld, A., Blechl, A., Epstein, L., Chen, X., et al. (2009). A kinase-START gene confers temperature-dependent resistance to wheat stripe rust. *Science* 323 (5919), 1357–1360. doi: 10.1126/science.1166289
- Gardiner, L. J., Quinton-Tulloch, M., Olohan, L., Price, J., Hall, N., and Hall, A. (2015). A genome-wide survey of DNA methylation in hexaploid wheat. *Genome Biol.* 16, 273. doi: 10.1186/s13059-015-0838-3
- Gardiner, L. J., Bansept-Basler, P., Olohan, L., Joynson, R., Brenchley, R., Hall, N., et al. (2016). Mapping-by-sequencing in complex polyploid genomes using genic sequence capture: a case study to map yellow rust resistance in hexaploid wheat. *Plant J.* 87 (4), 403–419. doi: 10.1111/tpj.13204
- Gardiner, L. J., Brabbs, T., Akhunov, A., Jordan, K., Budak, H., Richmond, T., et al. (2019). Integrating genomic resources to present full gene and putative promoter capture probe sets for bread wheat. *GigaScience* 8 (4), 1–13. doi: 10.1093/gigascience/giz018
- Gardiner, L. J., Bansept-Basler, P., El-Soda, M., Hall, A., and O'Sullivan, D. M. (2020). A framework for gene mapping in wheat demonstrated using the *Yr7* yellow rust resistance gene. *PLoS One* 15 (4), e0231157. doi: 10.1371/journal.pone.0231157
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8 (8), 1494–1512. doi: 10.1038/nprot.2013.084
- Harrington, S. A., Cobo, N., Karafiatova, M., Dolezel, J., Borrill, P., and Uauy, C. (2019). Identification of a Dominant Chlorosis Phenotype Through a Forward Screen of the *Triticum turgidum* cv. Kronos TILLING Population. *Front. Plant Sci.* 10:963:963. doi: 10.3389/fpls.2019.00963
- He, F., Pasam, R., Shi, F., Kant, S., Keeble-Gagnere, G., Kay, P., et al. (2019). Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. *Nat. Genet.* 51 (5), 896–904. doi: 10.1038/s41588-019-0382-2
- Heigwer, F., Kerr, G., and Boutros, M. (2014). E-CRISP: fast CRISPR target site identification. *Nat. Methods* 11 (2), 122–123. doi: 10.1038/nmeth.2812
- Henry, I. M., Nagalakshmi, U., Lieberman, M. C., Ngo, K. J., Krasileva, K. V., Vasquez-Gross, H., et al. (2014). Efficient Genome-Wide Detection and Cataloging of EMS-Induced Mutations Using Exome Capture and Next-Generation Sequencing. *Plant Cell* 26 (4), 1382–1397. doi: 10.1105/tpc.113.121590
- Hill, J. T., Demarest, B. L., Bisgrove, B. W., Gorski, B., Su, Y. C., and Yost, H. J. (2013). MMAPP: mutation mapping analysis pipeline for pooled RNA-seq. *Genome Res.* 23 (4), 687–697. doi: 10.1101/gr.146936.112
- Ishida, Y., Tsunashima, M., Hiei, Y., and Komari, T. (2015). Wheat (*Triticum aestivum* L.) transformation using immature embryos. *Methods Mol. Biol.* 1223, 189–198. doi: 10.1007/978-1-4939-1695-5\_15

- Jin, S., Zong, Y., Gao, Q., Zhu, Z., Wang, Y., Qin, P., et al. (2019). Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* 364 (6437), 292–295. doi: 10.1126/science.aaw7166
- Jupe, F., Witek, K., Verweij, W., Sliwka, J., Pritchard, L., Etherington, G. J., et al. (2013). Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J.* 76, 530–544. doi: 10.1111/tj.12307
- Krasileva, K. V., Vasquez-Gross, H. A., Howell, T., Bailey, P., Paraiso, F., Clissold, L., et al. (2017). Uncovering hidden variation in polyploid wheat. *Proc. Natl. Acad. Sci. U.S.A.* 114 (6), E913–E921. doi: 10.1073/pnas.1619268114
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25 (14), 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi: 10.1093/bioinformatics/btp352
- Lim, J. H., Yang, H. J., Jung, K. H., Yoo, S. C., and Paek, N. C. (2014). Quantitative trait locus mapping and candidate gene analysis for plant architecture traits using whole genome re-sequencing in rice. *Mol. Cells* 37 (2), 149–160. doi: 10.14348/molcells.2014.2336
- Ling, H. Q., Ma, B., Shi, X., Liu, H., Dong, L., Sun, H., et al. (2018). Genome sequence of the progenitor of wheat A subgenome *Triticum urartu*. *Nature* 557 (7705), 424–428. doi: 10.1038/s41586-018-0108-0
- Luo, M. C., Gu, Y. Q., Puiui, D., Wang, H., Twardziok, S. O., Deal, K. R., et al. (2017). Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature* 551 (7681), 498–502. doi: 10.1038/nature24486
- Maccaferri, M., Harris, N. S., Twardziok, S. O., Pasam, R. K., Gundlach, H., Spannagl, M., et al. (2019). Durum wheat genome highlights past domestication signatures and future improvement targets. *Nat. Genet.* 51 (5), 885–898. doi: 10.1038/s41588-019-0381-3
- Martinez, S. A., Shorinola, O., Conselman, S., See, D., Skinner, D. Z., Uauy, C., et al. (2020). Exome sequencing of bulked segregants identified a novel *TaMKK3-A* allele linked to the wheat *ERA8* ABA-hypersensitive germination phenotype. *Theor. Appl. Genet.* 133 (3), 719–736. doi: 10.1007/s00122-019-03503-0
- McCormick, R. F., Truong, S. K., and Mullet, J. E. (2015). RIG: Recalibration and interrelation of genomic sequence data with the GATK. *G3 (Bethesda)* 5 (4), 655–665. doi: 10.1534/g3.115.017012
- Michelmore, R. W., Paran, I., and Kesseli, R. V. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Natl. Acad. Sci. U.S.A.* 88, 9828–9832. doi: 10.1073/pnas.88.21.9828
- Mo, Y., Howell, T., Vasquez-Gross, H., Haro, L. A., Dubcovsky, J., and Pearce, S. (2018). Mapping causal mutations by exome sequencing in a wheat TILLING population: a tall mutant case study. *Mol. Genet. Genomics* 293 (2), 463–477. doi: 10.1007/s00438-017-1401-6
- Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., and Durbin, R. (2016). BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* 32 (11), 1749–1751. doi: 10.1093/bioinformatics/btw044
- Ni, F., Qi, J., Hao, Q., Lyu, B., Luo, M. C., Wang, Y., et al. (2017). Wheat *Ms2* encodes for an orphan protein that confers male sterility in grass species. *Nat. Commun.* 8:15121. doi: 10.1038/ncomms15121
- Ning, Z., Cox, A. J., and Mullikin, J. C. (2001). SSAHA: a fast search method for large DNA databases. *Genome Res.* 11 (10), 1725–1729. doi: 10.1101/gr.194201
- Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33 (3), 290–295. doi: 10.1038/nbt.3122
- Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A. H., Nielsen, K. L., et al. (2009). SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nat. Methods* 6 (8), 550–551. doi: 10.1038/nmeth0809-550
- Schneeberger, K. (2014). Using next-generation sequencing to isolate mutant genes from forward genetic screens. *Nat. Rev. Genet.* 15 (10), 662–676. doi: 10.1038/nrg3745
- Sheng, Z. H., Lv, Y. S., Li, W., Luo, R. J., Wei, X. J., Xie, L. H., et al. (2017). *Yellow-Leaf 1* encodes a magnesium-protoporphyrin IX monomethyl ester cyclase, involved in chlorophyll biosynthesis in rice (*Oryza sativa* L.). *PLoS One* 12 (5), e0177989. doi: 10.1371/journal.pone.0177989
- Steuernagel, B., Periannan, S. K., Hernandez-Pinzon, I., Witek, K., Rouse, M. N., Yu, G., et al. (2016). Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. *Nat. Biotechnol.* 34, 652–655. doi: 10.1038/nbt.3543
- Takagi, H., Uemura, A., Yaegashi, H., Tamiru, M., Abe, A., Mitsuoka, C., et al. (2013). MutMap-Gap: whole-genome resequencing of mutant F2 progeny bulk combined with *de novo* assembly of gap regions identifies the rice blast resistance gene *Pii*. *New Phytol.* 200 (1), 276–283. doi: 10.1111/nph.12369
- Takagi, H., Tamiru, M., Abe, A., Yoshida, K., Uemura, A., Yaegashi, H., et al. (2015). MutMap accelerates breeding of a salt-tolerant rice cultivar. *Nat. Biotechnol.* 33 (5), 445–449. doi: 10.1038/nbt.3188
- Uauy, C. (2017). Wheat genomics comes of age. *Curr. Opin. Plant Biol.* 36, 142–148. doi: 10.1016/j.pbi.2017.01.007
- Wu, H., Shi, N., An, X., Liu, C., Fu, H., Cao, L., et al. (2018). Candidate Genes for Yellow Leaf Color in Common Wheat (*Triticum aestivum* L.) and Major Related Metabolic Pathways according to Transcriptome Profiling. *Int. J. Mol. Sci.* 19 (6), 1594. doi: 10.3390/ijms19061594
- Xia, C., Zhang, L., Zou, C., Gu, Y., Duan, J., Zhao, G., et al. (2017). A TRIM insertion in the promoter of *Ms2* causes male sterility in wheat. *Nat. Commun.* 8:15407. doi: 10.1038/ncomms15407
- Xing, H. L., Dong, L., Wang, Z. P., Zhang, H. Y., Han, C. Y., Liu, B., et al. (2014). A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biol.* 14:327. doi: 10.1186/s12870-014-0327-y
- Yan, L., Loukoianov, A., Tranquilli, G., Helguera, M., Fahima, T., and Dubcovsky, J. (2003). Positional cloning of the wheat vernalization gene *VRN1*. *Proc. Natl. Acad. Sci. U.S.A.* 100 (10), 6263–6268. doi: 10.1073/pnas.0937399100
- Yan, L., Loukoianov, A., Blechl, A., Tranquilli, G., Ramakrishna, W., SanMiguel, P., et al. (2004). The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. *Science* 303 (5664), 1640–1644. doi: 10.1126/science.1094305
- Yuan, J. H., Li, J. H., Yuan, J. J., Jia, K. L., Li, S. F., Deng, C. L., et al. (2017). The application of MutMap in forward genetic studies based on whole-genome sequencing. *Yi Chuan* 39 (12), 1168–1177. doi: 10.16288/jyczz.17-095
- Zhang, H., Li, J., Yoo, J., Yoo, S., Cho, S., et al. (2006). Rice Chlorina-1 and Chlorina-9 encode ChlD and ChlI subunits of Mg-chelatase, a key enzyme for chlorophyll synthesis and chloroplast development. *Plant Mol. Biol.* 62, 325–337. doi: 10.1007/s11103-006-9024-z
- Zhang, T., Feng, P., Li, Y., Yu, P., Yu, G., Sang, X., et al. (2018). *VIRESCENT-ALBINO LEAF 1* regulates leaf colour development and cell division in rice. *J. Exp. Bot.* 69 (20), 4791–4804. doi: 10.1093/jxb/ery250
- Zhao, T. X., Kong, X. Y., Zhou, R. H., Gao, S. C., and Jia, J. Z. (2009). Morphological identification and analysis of EMS mutants from hexaploid wheat cultivar Yanzhan 4110. *Sci. Agric. Sin.* 42, 755–764. doi: 10.3864/j.issn.0578-1752.2009.03.001
- Zhao, G., Zou, C., Li, K., Wang, K., Li, T., Gao, L., et al. (2017). The *Aegilops tauschii* genome reveals multiple impacts of transposons. *Nat. Plants* 3 (12), 946–955. doi: 10.1038/s41477-017-0067-8

**Conflict of Interest:** Authors ZC and LG were employed by the company Chengdu Tcuni Technology.

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

Copyright © 2020 Dong, Zhang, Chen, Xia, Gu, Wang, Li, Xie, Zhang, Zhang, Gui, Liu and Kong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.