Brief Communication

Bulked segregant CGT-Seq-facilitated map-based cloning of a powdery mildew resistance gene originating from wild emmer wheat (Triticum dicoccoides)

Qiuhong Wu1,†, Fei Zhao2,3,†, Yongxing Chen1,3,†, Panpan Zhang1,3,†, Huaiyue Zhang1,3,†, Huaizhi Zhang1,3,†, Guanghao Guo1,3,†, Jingzhong Xie1,†, Lingli Dong1,†, Ping Lu1,†, Miaomiao Li1,†, Shengwei Ma1,†, Tzion Fahima4,†, Eviatar Nevo4,†, Hongjie Li5,†, Yijing Zhang2,3,*,† and Zhiyong Liu1,†,*,†

1State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, The Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing, China
2National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai, China
3University of Chinese Academy of Sciences, Beijing, China
4Institute of Evolution and the Department of Evolutionary and Environmental Biology, University of Haifa, Haifa, Israel
5The National Engineering Laboratory of Crop Molecular Breeding, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China

Keywords: BSA, ChIP-Seq, NLR, PAV.

Powdery mildew, caused by Blumeria graminis f. sp. tritici (Bgt), is a widely occurring foliar disease of wheat worldwide. Wild emmer wheat (WEW, Triticum dicoccoides) (AABB, 2n = 4x = 28), the progenitor of the cultivated tetraploid and hexaploid wheat, is highly resistant to powdery mildew, and many resistance alleles were identified in this wild species. However, only Pm41 that encodes a nucleotide-binding leucine-rich repeat (NLR) protein has been cloned using a map-based cloning approach (Li et al., 2020). NLR proteins account for the major gene families of disease resistance genes (Kourelis and van Beyma, 2013) and are subjected to epigenetic regulation including histone acetylation and DNA methylation (Li et al., 2019). Only 31%–34% of wheat NLRs were shared across all genomes, and the numbers of unique NLR varied from 22 to 192 per wheat cultivar (Walkowiak et al., 2020). These presence–absence variations (PAVs) of NLRs are obstacles for isolating disease resistance genes based on reference genome sequences. Bulked segregant analysis (BSA), an approach for identifying genetic variations associated with disease resistance genes, has been proven effective in mapping and cloning wheat PmSe gene in combination with next-generation sequencing (NGS) technology (Xie et al., 2020). The combination of BSA and core genome targeted sequencing (CGT-Seq), an approach to capture the major types of histone modification in the genome (Qi et al., 2018), might be used as an effective approach to capture sequence variations and PAVs in wheat NLRs between disease resistant and susceptible genotypes.

Powdery mildew resistance gene MIWE18 was identified in common wheat line 3D249 (Figure 1a), an introgression derivative of WEW accession WE18 (G-360-M), and initially mapped within a 23.5 cM genetic interval on chromosome 7AL (Han et al., 2009). In order to fine mapping and cloning of MIWE18, a CGT-Seq based BSA pipeline was developed (Figure 1b). We selected 30 homozygous resistant and 30 susceptible F2 families from the Xuezao × 3D249 mapping population to construct a pair of phenotypically contrasting DNA bulks. Chromatin immunoprecipitation sequencing (ChIP-Seq) was performed for three histone marks H3K27me3, H3K4me3, and H3K36me3 that are closely associated with gene activity. After quality control, 423,924,608 and 358,879,346 reads for the resistant and susceptible DNA bulks, respectively, were obtained for subsequent analysis.

Firstly, candidate chromosome region potentially associated with MIWE18 was located. Following stringent mapping and filtering steps, 8,461 high confidence single nucleotide polymorphisms (SNPs) were detected (Figure 1c). The sSNP-index method (Takagi et al., 2013) was applied to calculate the quantitative SNP differences between the two ChiPed DNA bulks for each of consecutive chromosome window. Further screening with regard to the read numbers and genotypes resulted in 5.32 Mb interval significantly associated with disease resistance on chromosome 7AL (724.21–729.53 Mb) (Figure 1d). To detect sequence variations associated with the resistance allele, the ChiPed sequences from the susceptible and resistant pools were de novo assembled, resulting in 655,192 and 685,858 contigs, respectively. The sequences specifically present in the resistant pool represented the candidate sequences within or near by the MIWE18 locus. The resistant bulk-specific contigs within the physical interval detected above were selected and mapped to the Chinese Spring (CS) reference genome. Since CS was susceptible to powdery mildew,

Please cite this article as: Wu, Q., Zhao, F., Chen, Y., Zhang, P., Zhang, H., Guo, G., Xie, J., Dong, L., Lu, P., Li, M., Ma, S., Fahima, T., Nevo, E., Li, H., Zhang, Y. and Liu, Z. (2021) Bulked segregant CGT-Seq-facilitated map-based cloning of a powdery mildew resistance gene originating from wild emmer wheat (Triticum dicoccoides). Plant Biotechnol J., https://doi.org/10.1111/pbi.13609
we focussed on sequences that were partially mapped to CS (containing unmapped fragments, or with insertions or deletions). A total of 479 CGT-Seq-derived contigs were identified in the resistant pool and sequence annotation revealed that five contigs contained NB-ARC domain, with lengths ranging from 595 to 2128 bp and sequence identities ranging from 15% to 56% between each other. Significantly, one 2128 bp contig (resk36_3047604) was highly homologous (99% identity) to the

---

**Figure 1** Map-based cloning of powdery mildew resistance gene MlWE18 introgressed from wild emmer into common wheat using a bulked segregant Core Genome Targeted sequencing (CGT-Seq) approach. (a) Reactions and Coomassie Brilliant Blue staining of 3D249 and Xuezao to Bgt isolate E09. (b) Workflow of CGT-seq based bulked segregant analysis. (c) The number of high-quality SNPs that differ between ChiPed resistant and susceptible DNA pools. (d) ΔSNP-index plot for the MlWE18 genomic region. (e) The CGT-Seq detected contig sequence (green) showed highly sequence similarity with Pm60. (f) High-density genetic map of MlWE18 and predicated genes in the 334 kb physical interval of Chinese Spring 7AL. (g) The RT-PCR expression patterns of the predicted genes and the 4 CGT-Contigs in the mapping interval in 3D249. (h) The relative expression levels of NLRWE18 examined by qRT-PCR at different h post inoculation (hpi) in 3D249. (i) Structure of the NLRWE18 and sequence difference with Pm60. The positions of SNPs and amino acids variations in the CDS (c.) and NLR protein were illuminated, respectively. (j) Plasmid construct of ProNLRWE18::NLRWE18 used for transformation. (k) Powdery mildew responses of NLRWE18 transgenic T 1 plants. +/-, presence/absence of the transgene.
Cloning wheat disease resistance gene by BSA-CGT-Seq

diploid *T. urartu* wheat *Pm60* sequence (Zou et al., 2018), with only 4 single nucleotide variations (SNVs) (Figure 1e).

In order to fine map the *MlWE18* locus, the five contigs identified by CGT-Seq together with sequences from the 5.32 Mb 7AL physical intervals of *CS* were used to develop polymorphic markers (*CGT1-CGT5; M306-M405*) linked to *MlWE18*. Since *MlIW172* was mapped to the same chromosomal region as *MlWE18* (Ouyang et al., 2014), polymorphisms of markers *WGGC4660, WGGC4662*, and *WGGC4657* linked to *MlIW172* were also tested. After screening a mapping population *MlWE18* region as markers *NLRWE18 resk36_3047604 (CGT3, (Figure 1f). RT-PCR was conducted to test the expression of the 7 with leucine-rich repeats proteins named NLR1, NLR2, and NLR3 domain-containing proteins, and three nucleotide-binding sites genomic region identified a MYB-related protein, a major region in the CS 7AL reference genome. Annotation of the CGT2 and *WGGC4657* and within a 0.09 cM genetic interval between markers *M306, M303, CGT1, CGT2*, and *CGT3* were found to co-segregated with *MlWE18*, whereas *CGT5* was closely linked (~0.05 cM) (Figure 1f). No polymorphism was detected between Xuezao and 3D249 for CGT4, CGT1, CGT2, and CGT3 were dominant markers detected only in the resistant 3D249, but not in the susceptible Xuezao, CS, *T. urartu G1812*, and WEW Zavitian, indicating PAVs of these NLR-like sequences between genotypes. Finally, *MlWE18* was mapped within a 0.09 cM genetic interval between markers *WGGC4660* and *WGGC4657*, and co-segregated with *M306, M303, CGT1, CGT2*, and *CGT3* (Figure 1f), corresponding to a 334 kb genomic region in the CS 7AL reference genome. Annotatio...