Genomic and Molecular Characterization of Wheat Streak Mosaic Virus Resistance Locus *2* (*Wsm2*) in Common Wheat (*Triticum aestivum* L.)

Yucong Xie[1†], Punya Nachappa[2], Vamsi J. Nalam[2] and Stephen Pearce[1*†]

[1] Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO, United States, [2] Department of Agricultural Biology, Colorado State University, Fort Collins, CO, United States

Wheat streak mosaic virus (WSMV) is an economically important viral pathogen that threatens global wheat production, particularly in the Great Plains of the United States. The *Wsm2* locus confers resistance to WSMV and has been widely deployed in common wheat varieties adapted to this region. Characterizing the underlying causative genetic variant would contribute to our understanding of viral resistance mechanisms in wheat and aid the development of perfect markers for breeding. In this study, linkage mapping in a doubled-haploid (DH) mapping population confirmed *Wsm2* as a major locus conferring WSMV resistance in wheat. The *Wsm2* flanking markers were mapped to a 4.0 Mbp region at the distal end of chromosome 3BS containing 142 candidate genes. Eight haplotypes were identified from seventeen wheat genotypes collected from different agroecological zones, indicating that *Wsm2* lies in a dynamic region of the genome with extensive structural variation and that it is likely a rare allele in most available genome assemblies of common wheat varieties. Exome sequencing of the variety "Snowmass", which carries *Wsm2*, revealed several loss-of-function mutations and copy number variants in the 142 candidate genes within the *Wsm2* interval. Six of these genes are differentially expressed in "Snowmass" compared to "Antero," a variety lacking *Wsm2*, including a gene that encodes a nucleotide-binding site leucine-rich repeat (NBS-LRR) type protein with homology to RPM1. A *de novo* assembly of unmapped RNA-seq reads identified nine transcripts expressed only in "Snowmass," three of which are also induced in response to WSMV inoculation. This study sheds light on the variation underlying *Wsm2* and provides a list of candidate genes for subsequent validation.

Keywords: wheat, Wheat streak mosaic virus (WSMV), Wsm2, RPM1, structural variation, antiviral mechanism, viral resistance

INTRODUCTION

Common wheat (*Triticum aestivum* L.) provides approximately 20% of the calories and proteins consumed by the human population (FAOSTAT, 2021). First observed in 1922, Wheat streak mosaic virus (WSMV) is the species type of the genus *Tritimovirus* within the family *Potyviridae* (Stenger et al., 1998; Singh et al., 2018) and is an economically important viral

pathogen that threatens wheat production around the globe (Navia et al., 2013). In the United States, WSMV mainly affects wheat grown in the Great Plains, causing average annual yield losses of approximately 5%, although severe localized infections can result in complete crop failure (Singh and Kundu, 2018; McKelvy et al., 2021). Once infected with WSMV, wheat leaves exhibit a characteristic yellow and green streaked mosaic pattern (Hadi et al., 2011). For winter wheat varieties, symptoms are most severe when infection occurs during tillering, which results in stunting, poor fertility, and reduced grain set (Hunger et al., 1992).

The transmission vector for WSMV is the eriophyid wheat curl mite (WCM) *Aceria tosichella* Keifer, which has a body length of ∼200 μm and is spread between plants by the wind (Slykhuis, 1955). The WSMV can be acquired by WCM from infected host plants during a 10- to 30-min feeding time and remains active in WCM for 7–9 days (Orlob, 1966). Upon landing on wheat plants, the WCM remains hidden in rolled and curled leaves and leaf sheaths, where it can feed and survive for several months (Navia et al., 2013). As a result, miticides are ineffective in controlling WCM populations (Navia et al., 2013). Moreover, volunteer wheat, other monocots, and wild weeds, including oats (*Avena sativa*), barley (*Hordeum vulgare*), rye (*Secale cereale*), corn (*Zea mays*), and foxtail millet (*Setaria italica*), can serve as a "green bridge" for WCM to complete their life cycle between wheat cropping seasons (Singh and Kundu, 2018). This broad host range makes it ineffective and impractical for many growers to use cultural practices to eradicate WCM from infected fields (Singh et al., 2018). Therefore, the most effective long-term strategy to prevent damage caused by WCM and WSMV is to develop wheat cultivars with genetic resistance to the WSMV-WCM disease complex (Harvey et al., 1999; Nachappa et al., 2021).

To date, four quantitative trait loci (QTL) associated with WCM resistance (known as *Curl mite colonization* or *Cmc* genes) have been identified from grass species and ancestors of common wheat (Thomas and Conner, 1986; Whelan and Hart, 1988; Malik et al., 2003). Although these resistance alleles inhibit WCM reproductive potential and reduce its transmission rate in the field, their effectiveness varies between WCM populations and environmental conditions (Murugan et al., 2011; Dhakal et al., 2017). Moreover, all four *Cmc* genes are derived from alien introgressions and are associated with reduced yields due to linkage drag (Harvey et al., 1999).

In addition to introgressing genetic resistance to WCM, four genes (*Wsm1, Wsm2, Wsm3,* and *c2652*) for WSMV resistance have been identified (Haley et al., 2002; Sharp et al., 2002; Divis et al., 2006; Haber et al., 2006). Both *Wsm1* and *Wsm3* originated in intermediate wheatgrass (*Thinopyrum intermedium*) and were transferred into common wheat varieties through alien translocation (Wells et al., 1982; Friebe et al., 2009; Liu et al., 2011; Danilova et al., 2017). When deployed in elite varieties, these alleles confer a yield penalty due to linkage drag, limiting their value in wheat breeding programs. For example, in the absence of WSMV infection, *Wsm1* confers a yield penalty of up

to 30% (Seifers et al., 1995; Sharp et al., 2002). The *Wsm3* gene has been introduced to common wheat only recently, and its effect on performance remains to be fully characterized (Nachappa et al., 2021). Although *c2652* was identified from a hard red spring wheat population (Haber et al., 2006), it has not been utilized in wheat germplasm development to date. The most widely deployed QTL is *Wsm2*, which was first identified in the wheat breeding line CO960293-2 and most likely originated in a common wheat background (Haley et al., 2002).

Over the last 15 years, *Wsm2* has provided strong resistance to WSMV (Seifers et al., 2006; Lu et al., 2012), leading to low WSMV incidence in field conditions (McKelvy et al., 2021). However, WSMV resistance-breaking strains have been reported in infected wheat carrying *Wsm2* (Fellers et al., 2019; Redila et al., 2021; Albrecht et al., 2022) and from *Setaria viridis* (Kumssa et al., 2019). Although *Wsm2* is temperature sensitive and less effective in the field at temperatures above 18°C (Seifers et al., 2006), there is no evidence that it has deleterious impacts on yield or other agronomic traits (Lu et al., 2012). Due to these advantages, *Wsm2* has been introduced into several common wheat varieties by recombination, including "RonL" (Martin et al., 2007), "Snowmass" (Haley et al., 2011), "Clara CL" (Martin et al., 2014), "Oakley CL" (Zhang et al., 2015), and "Joe" (Zhang et al., 2016).

Linkage mapping in two $F_{2:3}$ populations showed that the WSMV resistance conferred by *Wsm2* is controlled by a single dominant allele located on chromosome arm 3BS (Lu et al., 2011). Subsequent studies using a recombinant inbred line (RIL) population mapped the *Wsm2* locus to a 6.5 cM region (Assanga et al., 2017; Tan et al., 2017). Three SNP markers, each within 1 cM of *Wsm2*, were transformed into KASP assays and validated in a RIL population ("CO960293" × "TAM111") and in two doubled haploid (DH) populations ("RonL" × "Ripper" and "Snowmass" × "Antero"), from which haplotypes associated with WSMV resistance and susceptibility were identified (Tan et al., 2017). A genome-wide association study (GWAS) on 597 wheat breeding lines identified 10 other significant SNP markers associated with WSMV resistance (Dhakal et al., 2018). These 10 SNPs mapped to a 17.1–18.9 Mbp telomeric region on chromosome 3B in the IWGSC RefSeq v1.0 wheat reference genome assembly, coinciding with the *Wsm2* locus. This region contains a cluster of fourteen genes encoding Bowman-Birk inhibitors (BBIs) and is highly diverse between wheat varieties, with evidence of structural variation (Xie et al., 2021).

Despite the importance of *Wsm2*, the underlying causative gene has not been cloned, which will be required to develop perfect markers for breeding programs, and to further our understanding of viral resistance mechanisms in crops. In the current study, genome assemblies for the landrace "Chinese Spring" (IWGSC et al., 2018; Zhu et al., 2021) and sixteen other wheat varieties were used to characterize haplotypic variation at the *Wsm2* locus and to study their association with WSMV resistance. Exome and transcriptome sequencing datasets in the variety "Snowmass," which carries *Wsm2*, were analyzed to identify several possible causative genes underlying this locus.

MATERIALS AND METHODS

Plant Materials

Seeds of the wheat varieties "Jagger," "SY Mattis," "Robigus," "Mace," "Paragon," "Landmark," "Stanley," "Claire," "Weebill," "Cadenza," "Kronos," and "Chinese Spring" were obtained from Seedstor (https://www.seedstor.ac.uk) and used to perform WSMV phenotyping assays. A *T. aestivum* L. DH population (*n* = 116) produced by Heartland Plant Innovations Inc. was developed by wheat-maize wide hybridization (Santra et al., 2017) from the parents "Snowmass" (WSMV resistant) and "Antero" (WSMV susceptible) and used for linkage mapping. "Snowmass" and "Antero" leaf tissues were used to quantify WSMV coat protein transcript levels by qRT-PCR in a time course from 0, 5, 10, and 15-days post-inoculation (dpi). For the RNA-seq study, eight individuals homozygous for the *Wsm2* locus were selected from the DH population (**Supplementary Table S1**).

WSMV Inoculation and Phenotype Evaluation

An isolate of WSMV originally collected from Akron, Colorado, in 2017 was propagated in the greenhouse by mechanically inoculating the susceptible winter wheat genotype "Longhorn" every 6 months. Leaf tissues with a yellow streaking or mosaic pattern typical of WSMV were collected, frozen at −80 °C, and used to prepare fresh inoculum. The inoculum was prepared with 1:10 (w/v) dilution of the WSMV-infected wheat leaf tissue and 0.01 M of potassium phosphate buffer (pH 7.4) and inoculated on two-week-old seedlings. WSMV phenotyping was performed on two-week-old wheat plants grown in a PGR15 growth chamber (Conviron, Manitoba, Canada) in a 12 h photoperiod with temperatures set to 18°C day/15°C night. Mechanical inoculation was performed using a soft sponge soaked with the inoculum gently rubbed on the surface of wheat seedling leaves that were previously dusted with carborundum powder. Mock inoculation with phosphate buffer was used as a control. Plants were phenotyped 2 weeks after inoculation with WSMV by examining visual symptoms based on a modified 0–5 scale; (0 = no symptoms; 1 = one to a few chlorotic streaks on only one leaf of a plant; 2 = one to a few chlorotic streaks on < 10% of leaf tissues; 3 = moderate mosaic on >50% leaf tissues; 4 = severe mosaic on >80% leaf tissues; 5 = complete mosaic, necrosis, and yellowing) (Tan et al., 2017). The DH population was phenotyped using four biological replicates of each line in three independent phenotyping trials using a completely randomized design. Mean phenotype scores were used to perform linkage mapping. Individual wheat varieties were phenotyped using five biological replicates in two independent experiments. Varieties with median phenotypic scores ≤ 2 were considered resistant, and varieties with scores >2 were considered susceptible.

Haplotype Analysis for Within-Species Variation at the *Wsm2* Locus

The *Wsm2* flanking markers wsnp_Ex_c3005_5548573 (IWA3260) and BS00026471_51 (IWB7629), one left boundary flanking marker Ku_c663_1869), and three other markers tightly linked to *Wsm2* (IAAV6442, BS00018764_51, and IWA7647,

Supplementary Table S2) were used for haplotype analysis. The basic local alignment search tool (BLAST) alignment of the sequence 100 bp upstream and downstream from each SNP position was used to identify the physical position of *Wsm2* on the wheat reference genomes IWGSC RefSeq v1.0 (IWGSC 2018) and RefSeq v2.1 (Zhu et al., 2021). These SNPs were also mapped to the genome assemblies of sixteen other wheat varieties (Mace, Lancer, CDC Stanley, CDC Landmark, Julius, Norin61, ArinaLrFor, Jagger, Cadenza, Paragon, Kronos, Robigus, Claire, Spelt, Weebill, and SY Mattis) to analyze genomic variation using the Galaxy platform (Afgan et al., 2018).

For validation of the genomic variation between varieties, the genomic sequence of *Wsm2* in these wheat cultivars was extracted using the "bedtools" getfasta command. The FASTA files of *Wsm2* in different wheat varieties were subjected to pairwise alignment, and dot plots were generated with "D-genies" using the minimap function (Cabanettes and Klopp, 2018).

Linkage Mapping Analysis for the DH Population

The DH population was subjected to genotyping-by-sequencing (GBS, Elshire et al., 2011), and data were processed as described by Liu et al. (2016). The GBS markers were mapped to the wheat reference genome IWGSC RefSeq v1.0 (IWGSC 2018) and annotated based on their position. QTL analysis was performed with the R version 4.0.3 packages "R/qtl" (Arends et al., 2010) and "ASMap" (Taylor and Butler, 2017) using the mean phenotyping scores from two weeks post-inoculation of the four biological replicates of each DH line.

Exome Capture Analysis for Genetic Variation Underlying *Wsm2* Locus

Exome reads of "Snowmass," "Antero," "Brawl," "Byrd," "Hatcher," "CO940610," and "Platte" were captured using the NimbleGen SeqCap EZ wheat whole-genome assay and sequenced as described by Jordan et al. (2015) and He et al. (2019). The 150 bp paired-end Illumina reads were filtered for quality using "fastp" (Chen et al., 2018). Reads were then aligned to the wheat reference IWGSC RefSeq v1.0 assembly using "bowtie2" v. 2.3.5 (Langmead and Salzberg, 2012) with the following parameters: -k 2 -N 1 -L 22 -D 20 -R 3. The alignments were subjected to "samtools" v1.11 to generate sorted BAM files and then "bcftools" v1.11 (Danecek et al., 2021) was used to call variants within the *Wsm2* locus with the "mpileup" command. The "SnpEff" (Cingolani et al., 2012) tool was used to predict the effects of genetic variants, including SNPs, indels, and multiple-nucleotide polymorphisms. The sorted BAM files from "Antero," "Brawl," "Byrd," "Hatcher," "CO940610," and "Platte" were used as a reference set to assay copy number variation (CNV) for the test sample "Snowmass" (*Wsm2*+) using the "ExomeDepth R" package v.1.1.12 (Plagnol et al., 2012). The default parameters were used, except for a transition probability of 0.001 ("CallCNVs") and a minimum overlap of 0.01 ("AnnotateExtra").

To calculate the total variant number in the defined regions, the sorted BAM files were subjected to "bcftools" to generate VCF files and specify the defined region using (-r) in the "bcftools mpileup" command. The output VCF files were then

subjected to "SnpEff" to predict the total number of variants within each region. The total exome or protein-coding gene length was calculated using the GFF3 file from IWGSC RefSeq v1.1 (http://ftp.ensemblgenomes.org/pub/plants/release-52/gff3/triticum_aestivum/) to extract the start and end position for each "gene" feature and sum the length of each "gene" feature within the selected interval. The variant rate was calculated by dividing the total exome length by the total variant number. The total variant rate for "Snowmass" was calculated for each individual chromosome and for the 4.0 Mbp interval spanning *Wsm2*. In addition, six other selected 4.0 Mbp intervals on telomeric or centromeric regions of the chromosomes (3B:40-44Mbp, 3B:200-204Mbp, 3B:345-349Mbp, 1A:15-19Mbp, 2B:15-19Mbp, and 6D:15-19Mbp), together with the *Wsm2* interval in six other wheat varieties were used as comparisons for the overall variant rate.

WSMV Quantification

Two weeks after germination, "Antero" (*Wsm2-*, WSMV susceptible) and "Snowmass" (*Wsm2+*, WSMV resistant) plants were subjected to WSMV inoculation as described above. Leaf samples were taken by cutting the whole leaf tissues from five biological replicates over the time course of 0-, 5-, 10-, and 15-dpi. Leaf tissues were ground and homogenized in liquid nitrogen for subsequent total RNA isolation with the Spectrum™ Plant Total RNA Isolation Kit (Sigma, USA), followed by on-column DNase I digestion treatment (Sigma-Aldrich, USA) to remove genomic DNA according to the manufacturer's instructions. The one-step qRT-PCR reaction was carried out using the TaqMan® RNA-to-Ct™ 1-step kit (Applied Biosystems™, USA) on a QuantStudio™3 Real-Time PCR system (Applied Biosystems, USA). Reaction conditions and parameters adopted were as described previously by Price et al. (2010). WSMV was detected *via* qRT-PCR using coat protein-specific primers and the probe listed in **Supplementary Table S3**. To quantify the absolute amount of WSMV coat protein transcript levels, a regression line was plotted using plant RNA that contains 10 ng/µL of WSMV and subjected to 10-fold serial dilutions to 1×10^{-5} ng/µL. The C_q values for each dilution were plotted against total RNA transcript levels, and the regression line was considered at $R^2 > 0.996$.

RNA-Seq Library Preparation

Sixteen samples were collected for the RNA-seq experiment, including four biological replicates of two genotypes (*Wsm2+* and *Wsm2-*) from the control condition (C - mock inoculation with phosphate buffer) and treatment condition (T - WSMV inoculation). Whole leaves were harvested at 10 dpi, stored at $-80°C$, and ground to a homogenized fine powder in liquid nitrogen. Total RNA was isolated with Spectrum™ Plant Total RNA Isolation Kit (Sigma, USA) and quantified using Qubit™ Flex Flouremeter (Invitrogen™, USA). The Agilent 2100 bioanalyzer (RNA Nano Chip, Agilent, CA) was used to check RNA integrity. The library construction and sequencing *via* Illumina HiSeq 2000 were performed by Novogene Co., Ltd (Sacramento, CA, USA) and approximately 150 bp paired-end raw reads were generated. Raw sequencing data are available

from the NCBI Gene Expression Omnibus under accession number GSE190382.

Transcript Abundance, Differential Expression (DE) and GO Enrichment Analysis

To quantify WSMV reads in the RNA-seq samples, the coding sequence from WSMV isolate KSHm2014 (9,384 bp) was retrieved from the NCBI database (MK318278.1, https://www.ncbi.nlm.nih.gov/). This sequence was concatenated to the IWGSC RefSeq v1.0 wheat genome (IWGSC 2018) as an additional FASTA entry and used as the combined wheat-WSMV reference genome to build index files for alignments. Raw reads of each paired-end library were examined for sequence quality and adaptor sequences were removed using "fastp" with default settings (Chen et al., 2018). Trimmed paired-end RNA-seq reads were aligned to the reference genome using "STAR" 2.7.3 (Dobin and Gingeras, 2015) with parameters "-outFilterMismatchNmax 6 -alignIntronMax 10000". Non-normalized reads were counted with "featureCounts" (Liao et al., 2014) with parameters "-t gene -p" and used as input for the R package "DEseq2" v3.14 (Love et al., 2014). Read counts of WSMV in each sample were normalized to "fastp" trimmed reads of the corresponding samples for count per million (cpm) of WSMV, and log-transformed into LogCPM. Differentially expressed genes (DEGs) were identified from pairwise comparisons for treatment effect (WSMV-treated *vs.* mock-treated) and for genotypic effect under each condition: Resistant vs. Susceptible under WSMV-treated condition (T_{Wsm2-}^{Wsm2+}) and mock-treated condition (C_{Wsm2-}^{Wsm2+}). The *P*-value threshold was determined using Benjamini and Hochberg's approach (Benjamini and Hochberg, 1995) for controlling the false discovery rate (FDR < 0.01) without controlling the $\log_2$ fold change (FC). Venn diagrams were drawn using VENNY software (Oliveros, 2007). Gene ontology (GO) enrichment analyses were performed with the "TopGO" R package v3.14 and Fisher tests were conducted to identify significant GO terms ($P < 0.01$).

De novo Transcriptome Assembly of Unmapped Reads in *Wsm2+* and Presence/Absence Analysis

During the STAR alignment, reads that did not map to IWGSC RefSeq v1.0 were collected with parameter "-outReadsUnmapped" and assembled with "Trinity" tool v2.14.0 (Grabherr et al., 2011) using the following parameters: "-seqType fq –samples_file<input_file> -max_memory 10G -CPU 20." The proportion of reads mapped to the assembly was assessed with "Bowtie2" v2.3.5 (Langmead and Salzberg, 2012). Then "CD-HIT" (cd-hit-est – c 0.95) was used to remove redundant transcripts. To annotate the gene functions, the assembled transcripts were used as BLASTx queries against the NCBI NR database of non-redundant proteins (cutoff: 1e-5). The DEG analysis was performed on unmapped reads against assembled transcriptomes to identify differentially expressed transcripts (DETs) in four pairwise comparisons: T_{Wsm2-}^{Wsm2+}, C_{Wsm2-}^{Wsm2+}, $Wsm2+_C^T$, and $Wsm2-_C^T$.

Presence and absence variation (PAV) analysis was performed by BLASTn (default: 1e-3) on the assembled transcript sequences against the coding sequences (CDS) of other wheat varieties using the Galaxy platform (Afgan et al., 2018). The transcript was determined to be present (+) when the top BLAST hit exhibited a CDS similarity percentage >96%; otherwise the transcript was determined to be absent from that genome assembly. For unique transcripts in *Wsm2+* that are also present in other wheat varieties, the corresponding gene ID and physical position were extracted using the "Galaxy" platform (Afgan et al., 2018).

Gene Expression Validation With qRT-PCR

Transcript levels of selected candidate DEGs identified from RNA-seq experiments were validated with qRT-PCR in the same samples used for WSMV quantification. First-strand complementary DNA (cDNA) was synthesized from 2 μg of total RNA using SuperScript IV Reverse Transcriptase Kit (Thermo Fisher Scientific, USA) according to the manufacturer's instructions. The qRT-PCR reactions were performed using PowerUp SYBR Green Master Mix (Thermo Fisher Scientific, USA) in a 20 μL reaction with 100 ng cDNA and 1 μL of a 10 μM solution for each primer. Relative gene expression analysis was calculated using *ACTIN* as the internal control gene and 2^{-CT} method was used for relative quantification. Primer efficiency and specificity were determined by analyzing amplification in a four-fold dilution series and checking the dissociation curve for a single amplified product and calculated as: Efficiency (%) = $(4^{\frac{1}{slope}} - 1) \times 100$. All primers in this study had an efficiency >90% and are listed in **Supplementary Table S3**.

RESULTS

Genomic Characterization of *Wsm2* in Common Wheat

Analysis of Within-Species Genomic Variation for *Wsm2* in Seventeen Wheat Varieties

The *Wsm2* flanking markers wsnp_ex_c3005_5548573 (referred to as SNP1 hereafter) and BS00026471_51 (SNP6) were used to define the physical position of the *Wsm2* locus in genome assemblies of the landrace "Chinese Spring" (**Supplementary Table S2**). SNP1 mapped to two locations on chromosome arm 3BS, 14,985,292 bp, and 26,650,801 bp. At both positions, 100 bp flanking sequences were identical except that the former carries the "C" allele type, whereas the latter carries the "T" allele type for this SNP. As a result, the SNP1 is an ambiguous marker to define the left boundary of *Wsm2*. Therefore, an alternative marker Ku_c663_1896 (SNP2) which is 0.3 cM from SNP1 in the RIL population and 680 bp downstream of SNP1 mapped uniquely at position 14,985,972 bp (**Supplementary Table S2**) and was used to define the left boundary of *Wsm2* in subsequent analyses.

The markers, SNP2 and SNP6 span a 4.0 Mbp interval on chromosome arm 3BS in both the IWGSC RefSeq v1.0 (15.0 - 19.0 Mbp) and RefSeq v 2.1 (20.5 Mbp - 24.5 Mbp) genome assemblies (**Supplementary Table S2**). In both assemblies, this region included the same 142 annotated gene models (70
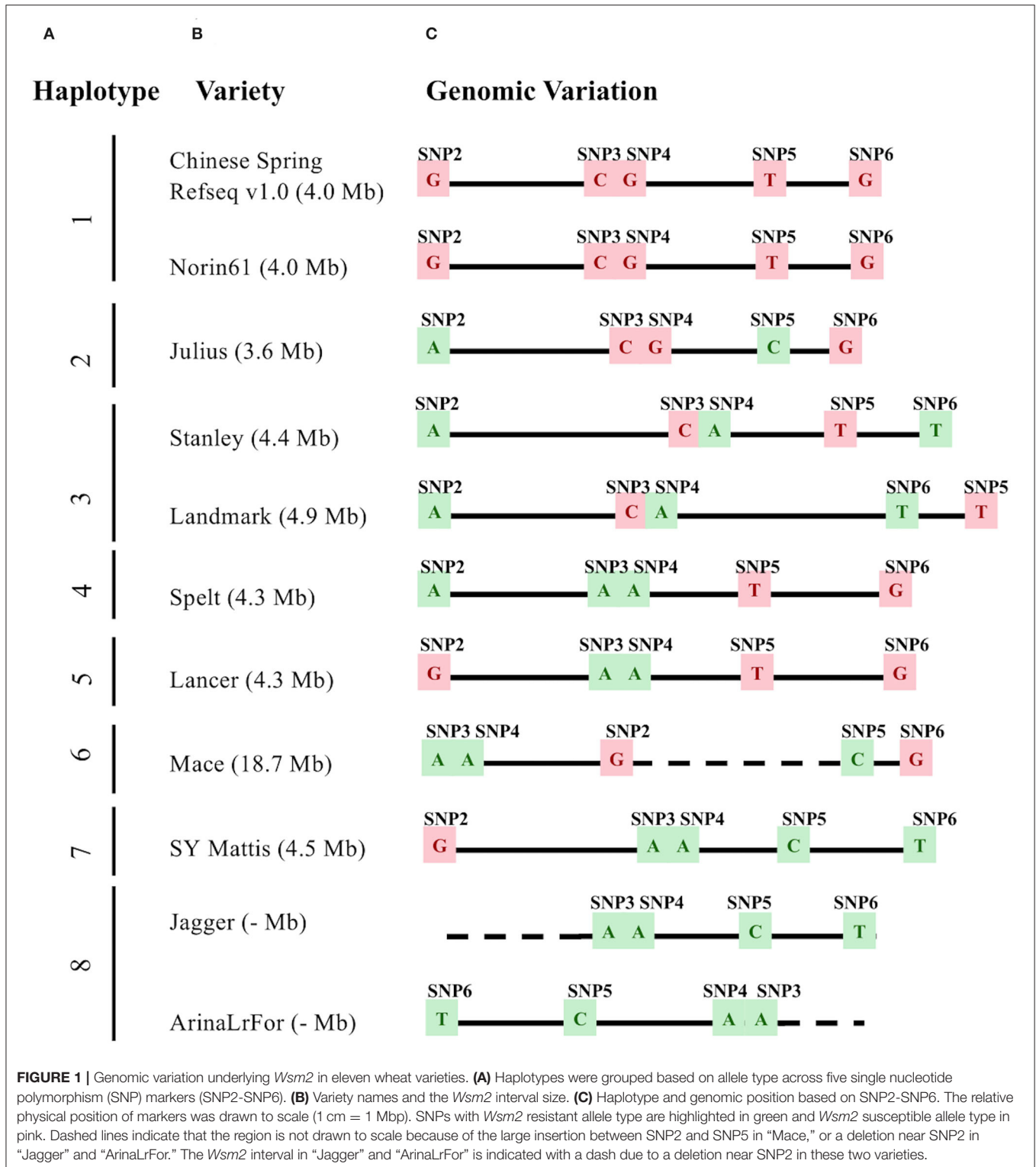
high confidence, 72 low confidence; **Supplementary Table S4**). Three other SNP markers, each within 1 cM of the *Wsm2* (Tan et al., 2017), are referred to as SNP3, SNP4, and SNP5, together with the flanking marker SNP6, were used to define the haplotypes across the *Wsm2* region. "Chinese Spring" carries the "CGTG" haplotype (**Figure 1**) and is consistent with its median phenotypic score of 4 (based on a 0–5 scale of visual symptoms where ≤ 2.0 indicates resistance and > 2.0 indicates susceptibility) (**Table 1**). The susceptible phenotype of "Chinese Spring" indicates that the wheat reference genome likely does not contain the *Wsm2* genetic variant.

To compare within-species genomic diversity at the *Wsm2* locus, the corresponding genomic region was identified in ten additional wheat varieties with pseudomolecule genome assemblies by mapping the physical position of each single nucleotide polymorphism (SNP) (**Figure 1**, **Supplementary Table S5**). This region contained structural variation between varieties. For example, the physical distance between these five markers ranges from 3.6 Mbp in "Julius" to 18.7 Mbp in "Mace," due to a 17.9 Mbp insertion and inversion between markers SNP2 and SNP5 in the latter variety (**Figure 1**). An inversion between markers, SNP3 and SNP6 was also detected in "ArinaLrFor," while SNP2 is deleted from the "Jagger" and "ArinaLrFor" assemblies (**Figure 1**). An alignment of genomic sequence flanking this region suggests that these two varieties carry a deletion near the left boundary of *Wsm2* (**Supplementary Figure S1**).

To further analyze *Wsm2* haplotypes, six other wheat varieties with scaffold-level genome assemblies ("Kronos," "Cadenza," "Weebill," "Claire," "Paragon," and "Robigus") and two winter wheat parental lines ("Antero" and "Snowmass") were included in the analysis. Eight haplotypes in this region were identified considering all nineteen wheat varieties (**Table 1**). The "CGTG" haplotype associated with WSMV susceptibility in "Chinese Spring" was also identified in "Norin61" and "Antero" (**Table 1**). Although "Norin61" was not phenotyped in the current study, both "Chinese Spring" and "Antero" have a median score greater than 2, consistent with the association between this haplotype and WSMV susceptibility (**Table 1**). However, the five varieties carrying the "AACT" haplotype exhibited variable phenotypic scores, including a median score of 1 in "Snowmass," 2 in "Jagger," and 3 in "SY Mattis" and "Robigus" (**Table 1**), indicating that this haplotype is not consistently associated with WSMV resistance. Six other haplotypes were identified in the remaining eleven wheat varieties, all of which were susceptible to WSMV infection (median score > 2, **Table 1**). Taken together, these results show that *Wsm2* is in a dynamic region of the wheat genome and is likely absent from all wheat varieties with assembled genomes.

Linkage Mapping Confirmed That *Wsm2* Confers Resistance to WSMV in "Snowmass"

To validate the association between *Wsm2* and WSMV resistance, linkage mapping was performed in a DH mapping population (*n* = 116) derived from "Snowmass" (mean phenotypic score 0.8, "AACT haplotype") and "Antero" (mean phenotypic score 3.3, "CGTG" haplotype). Four significant QTL for WSMV

FIGURE 1 | Genomic variation underlying *Wsm2* in eleven wheat varieties. **(A)** Haplotypes were grouped based on allele type across five single nucleotide polymorphism (SNP) markers (SNP2–SNP6). **(B)** Variety names and the *Wsm2* interval size. **(C)** Haplotype and genomic position based on SNP2–SNP6. The relative physical position of markers was drawn to scale (1 cm = 1 Mbp). SNPs with *Wsm2* resistant allele type are highlighted in green and *Wsm2* susceptible allele type in pink. Dashed lines indicate that the region is not drawn to scale because of the large insertion between SNP2 and SNP5 in "Mace," or a deletion near SNP2 in "Jagger" and "ArinaLrFor." The *Wsm2* interval in "Jagger" and "ArinaLrFor" is indicated with a dash due to a deletion near SNP2 in these two varieties.

resistance were identified [logarithm of odds (LOD) > 3, $P <$ 0.001] on chromosomes 3B, 3D, 5B, and 7B (**Figure 2**). The strongest association was identified on chromosome 3B, where 60 significant markers (LOD > 3) were mapped, including

45 within the region between 11.9 Mbp and 28.5 Mbp, co-located with the previously defined *Wsm2* region (15.0 - 19.0 Mbp) (**Supplementary Table S6**). In addition, two other significant markers were mapped to chromosome 3D at physical

TABLE 1 | Haplotype analysis of *Wsm2* in 19 wheat varieties.

Marker Name	SNP3	SNP4	SNP5	SNP6	Phenotype Score (Median)	Haplotype Group
Marker Position (bp)	16,443,466	16,455,416	17,770,942	18,873,524		
Antero	C	G	T	G	3	1
Chinese Spring	C	G	T	G	4	
Norin61	C	G	T	G	a	
Julius	C	G	C	G	a	2
Kronos	C	A	C	G	4	3
Cadenza	C	A	T	G	3	4
Weebill	C	A	T	T	3	5
Claire	C	A	T	T	3	
CDC Stanley	C	A	T	T	4	
CDC Landmark	C	A	T	T	3.5	
Spelt	A	A	T	G	a	6
Lancer	A	A	T	G	a	
Paragon	A	A	C	G	2.5	7
Mace	A	A	C	G	4	
Robigus	A	A	C	T	3	8
SY Mattis	A	A	C	T	3	
Jagger	A	A	C	T	2	
ArinaLrFor	A	A	C	T	a	
Snowmass	A	A	C	T	1	

[a]*Gray highlight means plant materials were unavailable to phenotype.*
The haplotype alleles shared with "Snowmass" (WSMV resistant) are highlighted in blue, whereas the haplotype shared with "Antero" (WSMV susceptible) is highlighted in pink. Phenotyping was performed 2 weeks after virus inoculation based on a 0–5 scale of visual symptoms. A median score ≤2 was considered resistant, whereas a median score >2 was considered susceptible.

positions 4,397,505 bp and 5,446,355 bp. Sequence alignment showed that this region is not syntenic to the *Wsm2* locus on chromosome 3B (**Supplementary Figure S2**). Five other significant markers were mapped to chromosome 7B, and one significant marker was mapped to chromosome 5B. Full details of each marker and their associated LOD score are provided in **Supplementary Table S6**. The results confirmed that "Snowmass" contains the *Wsm2* variant, and that this locus confers WSMV resistance. Therefore, "Snowmass" can be used to identify and characterize genetic variation underlying the *Wsm2* locus.
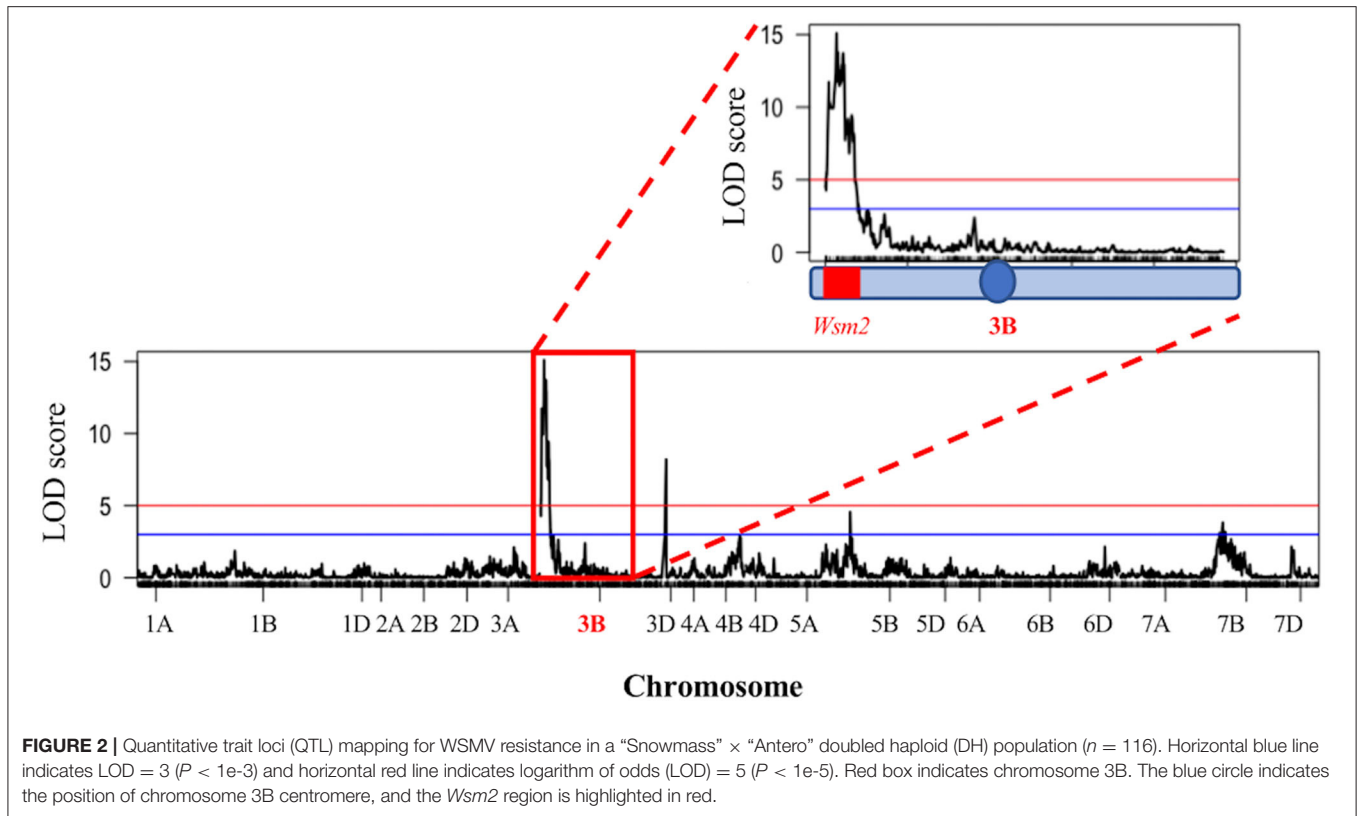
Exome Sequencing Revealed Genetic and Copy Number Variation at the *Wsm2* Locus

Exome sequencing reads from "Snowmass" were mapped to the IWGSC RefSeq v1.0 genome to characterize natural genetic variation in protein-coding genes between "Chinese Spring" (*Wsm2*-) and "Snowmass" (*Wsm2*+). Within the *Wsm2* interval, 1,191 SNPs (96%) and 50 small Indels (4%) were identified (**Supplementary Table S7**). This translates to a rate of one variant per 126 bp of the coding sequence, higher than the mean rate of one variant per 503 bp across the whole exome, and six other 4 Mbp regions sampled from different regions of the genome (**Supplementary Table S8**). However, a high rate of variation in this region was also observed in six other wheat genotypes ("Antero," "Brawl," "Byrd," "Hatcher," "CO940610," and "Platte"), ranging from one variant per 125

bp to 189 bp exome length (**Supplementary Table S8**). None of these varieties exhibit resistance to WSMV, suggesting that the high rate of variation is unrelated to the presence of *Wsm2*, consistent with its origins in common wheat genetic material.

The variants within *Wsm2* are predicted to induce 4,382 genetic effects within spliced transcript sequences (**Supplementary Table S7**). Six high-impact variants were predicted, leading to either premature introduction of a stop codon within the coding sequence or a shift in the open reading frame (**Supplementary Table S7**). Of the six high impact variants in "Snowmass," four were also present in the WSMV susceptible parent "Antero" (**Supplementary Table S7**). One of the genetic variants unique to "Snowmass" is a 2 bp insertion in *TraesCS3B02G042400LC*, a non-translating gene, and the other is a 10 bp deletion in *TraesCS3B02G038300* (Bowman-Birk trypsin inhibitor) that introduces a premature stop codon at amino acid 156 (T156*) and is likely to encode a non-functional protein (**Supplementary Table S7**).

To search for structural variation within the *Wsm2* interval in "Snowmass," exome sequencing read depth from "Snowmass" was compared to the mean depth from the exomes of six other wheat varieties. Ten of the 142 candidate genes within *Wsm2* exhibited copy number variation (**Figure 3**, **Supplementary Figure S3**). Six genes underwent an expansion in "Snowmass" and are predicted to have two to three copies compared to the mean coverage in the reference exome set

FIGURE 2 | Quantitative trait loci (QTL) mapping for WSMV resistance in a "Snowmass" × "Antero" doubled haploid (DH) population (*n* = 116). Horizontal blue line indicates LOD = 3 (*P* < 1e-3) and horizontal red line indicates logarithm of odds (LOD) = 5 (*P* < 1e-5). Red box indicates chromosome 3B. The blue circle indicates the position of chromosome 3B centromere, and the *Wsm2* region is highlighted in red.

(**Figure 3**). These include a cluster of four adjacent genes encoding UDP-glycosyltransferase proteins and a gene encoding a nucleotide-binding site leucine-rich repeat (NBS-LRR) type disease resistance protein (**Figure 3**).
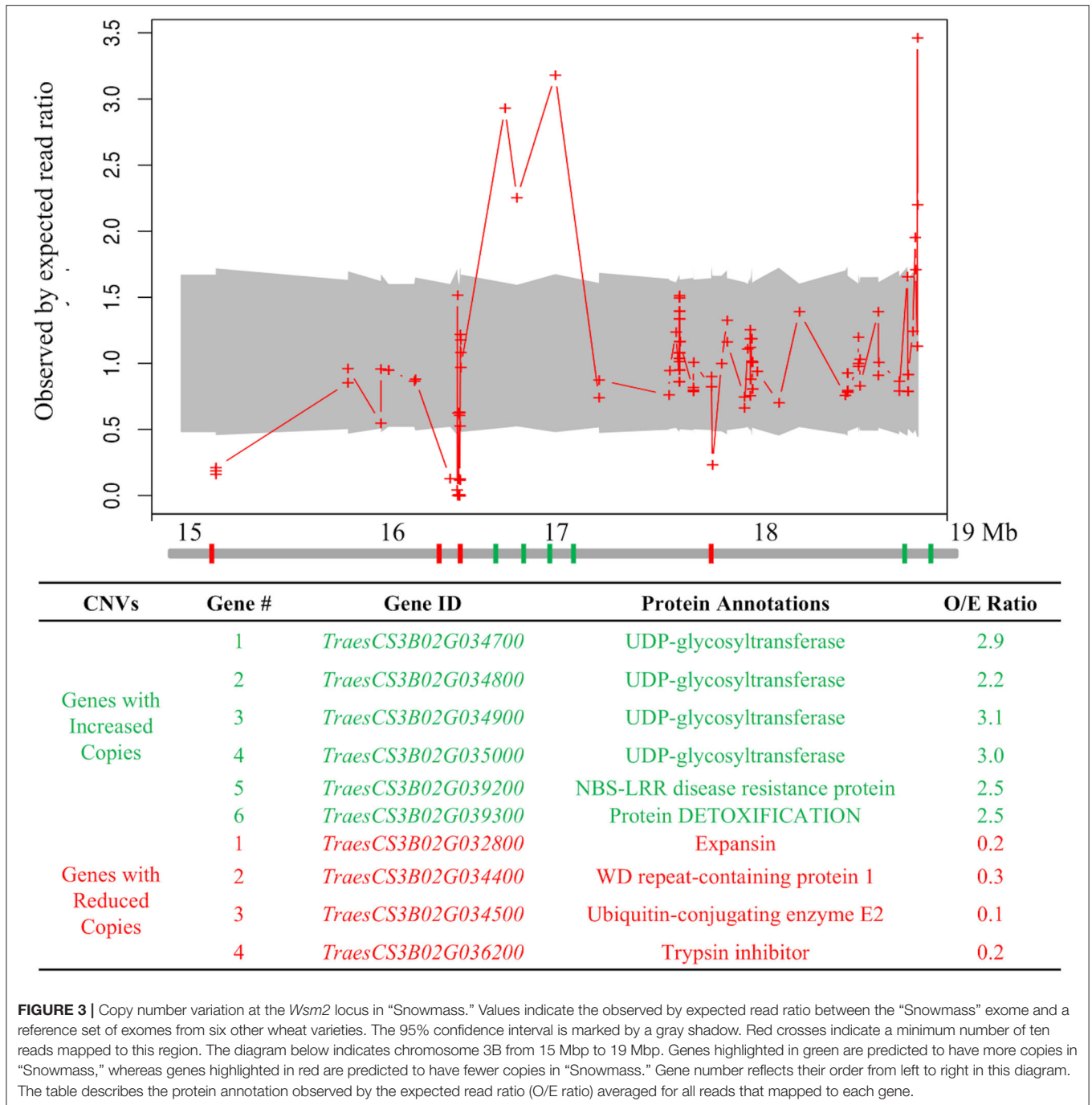
Transcriptomics to Characterize Host Response to WSMV Infection
WSMV Accumulation in *Wsm2+* and *Wsm2-* Genotypes

To compare the accumulation of WSMV in resistant and susceptible wheat, RNA was extracted from whole leaf tissues from "Snowmass" (*Wsm2+*, WSMV resistant) and "Antero" (*Wsm2-*, WSMV susceptible) at four time points after WSMV inoculation (0, 5, 10, and 15 dpi, **Figure 4A**). WSMV was accumulated in both genotypes throughout the time course, but at a much lower rate in *Wsm2+* compared to *Wsm2-* (**Figure 4A**). There were no significant differences in WSMV coat protein transcript levels between genotypes at either 0 or 5 dpi (*P* > 0.05), but at both 10 dpi (4.4-fold, *P* < 0.001) and 15 dpi (4.7-fold, *P* < 0.05) *Wsm2+* contained significantly lower levels of WSMV transcripts than *Wsm2-* (**Figure 4A**). This result was consistent with visual symptoms; whereas *Wsm2-* individual plants showed characteristic streaked and mosaic patterns on their leaves beginning at 10 dpi, the leaves of *Wsm2+* individuals remained asymptomatic throughout the time course (**Figure 4B**).
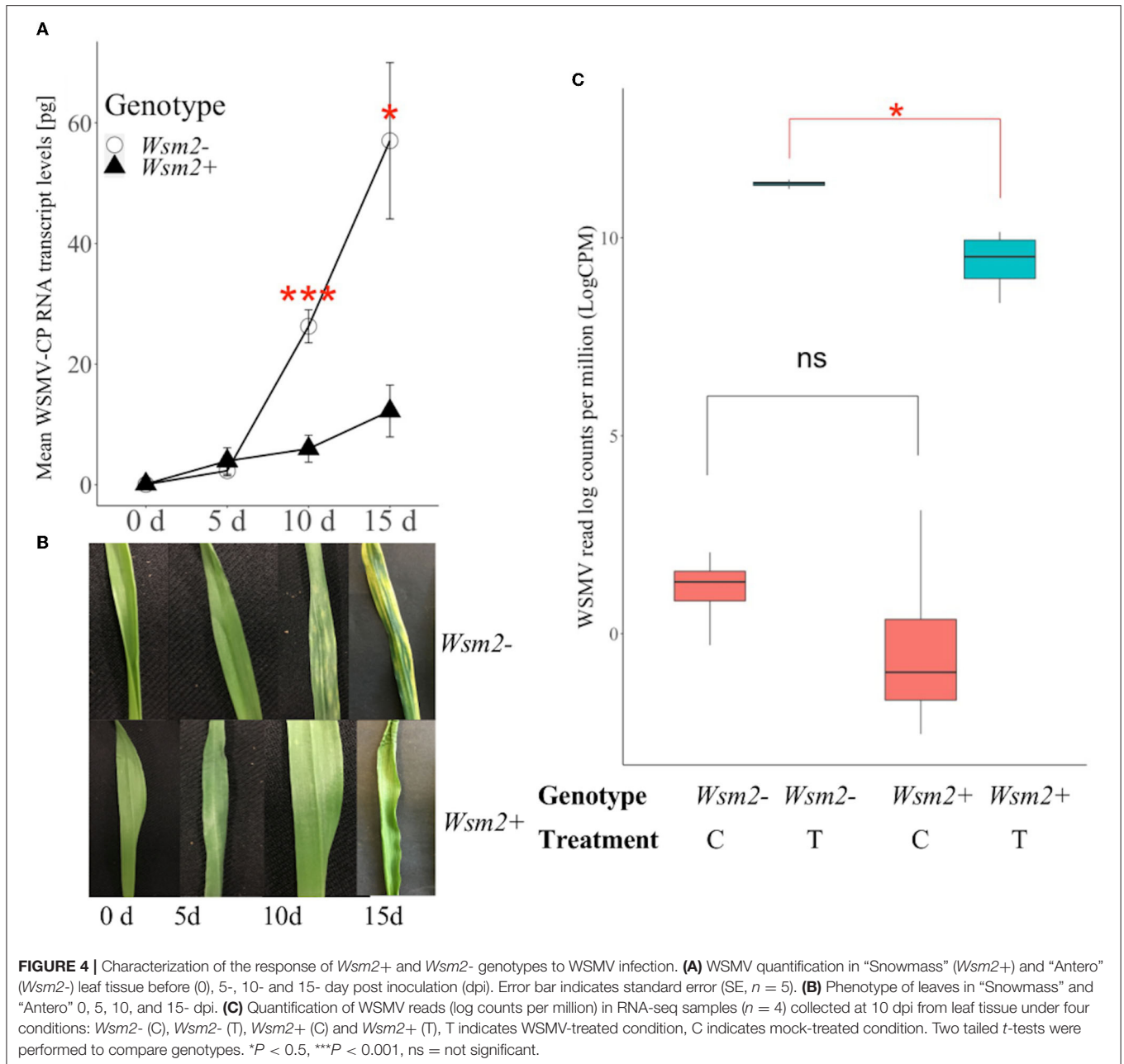
Summary Statistics of the RNA-Seq Experiment

Based on the time course results, 10 dpi was selected to characterize the early transcriptomic response of wheat plants to WSMV infection. In this RNA-seq study, 16 samples were collected from the leaf tissue of two wheat genotypes under two treatments (WSMV-treated, mock-treated). Using GBS markers mapped to the IWGSC RefSeq v1.0 genome, four DH individuals were shown to have an identical haplotype to "Snowmass," and another four had an identical haplotype to "Antero," of which these eight individuals were selected as plant materials representing two genotypes (*Wsm2+* and *Wsm2-*, **Supplementary Table S1**). Phenotype scores also confirmed that *Wsm2+* DH lines are resistant to WSMV, whereas *Wsm2-* DH lines are susceptible (**Supplementary Table S1**). After adaptor trimming and removal of low-quality reads, an average of 26.7 million reads were retained (**Supplementary Table S9**) and mapped to a combined wheat-WSMV reference genome described in detail in Section Transcript Abundance, Differential Expression (DE) and GO Enrichment Analysis. Although WSMV reads were detected in all samples, the levels were more than 10,000-fold higher in WSMV-treated (T) samples [mean 50,418 WSMV read counts per million (cpm)] compared to mock-treated (C) samples (mean 5 cpm, **Supplementary Table S10**). WSMV levels were also significantly higher in *Wsm2-* (T) samples (mean 11.4 LogCPM) than in *Wsm2+* (T) samples (mean 9.4 LogCPM, *P* < 0.05, **Figure 4C**), consistent with qRT-PCR results at 10 dpi (**Figure 4A**).

FIGURE 3 | Copy number variation at the *Wsm2* locus in "Snowmass." Values indicate the observed by expected read ratio between the "Snowmass" exome and a reference set of exomes from six other wheat varieties. The 95% confidence interval is marked by a gray shadow. Red crosses indicate a minimum number of ten reads mapped to this region. The diagram below indicates chromosome 3B from 15 Mbp to 19 Mbp. Genes highlighted in green are predicted to have more copies in "Snowmass," whereas genes highlighted in red are predicted to have fewer copies in "Snowmass." Gene number reflects their order from left to right in this diagram. The table describes the protein annotation observed by the expected read ratio (O/E ratio) averaged for all reads that mapped to each gene.

The overall mapping rate was 97.2% across 16 samples. The average unique mapping rate for *Wsm2-* wheat samples was 84.5 ± 3.3%, whereas for *Wsm2+* samples, the rate was 81.0 ± 1.3% (**Supplementary Table S9**). In the principal component analysis (PCA), PC1 explained 45% of the variance in the overall transcriptome between samples, and most samples were separated according to treatment (**Supplementary Figure S4**). There were three ambiguous samples, of which two belong to *Wsm2+* (T) (R2 and R4), and one belongs to *Wsm2-* (C) (R4)

(**Supplementary Figure S4**). The counts of WSMV reads in the two ambiguous *Wsm2+* (T) samples were 8.4 and 9.2 LogCPM, respectively, compared to an average of 10.0 LogCPM in other *Wsm2+* (T) samples (**Supplementary Table S10**), suggesting that transcriptome variation in these samples may be due to variation in WSMV inoculation and infection. Samples were not grouped according to their genotype, indicating that transcriptomic variation between genotypes was comparatively smaller than between treatments.

FIGURE 4 | Characterization of the response of *Wsm2+* and *Wsm2-* genotypes to WSMV infection. **(A)** WSMV quantification in "Snowmass" (*Wsm2+*) and "Antero" (*Wsm2-*) leaf tissue before (0), 5-, 10- and 15- day post inoculation (dpi). Error bar indicates standard error (SE, *n* = 5). **(B)** Phenotype of leaves in "Snowmass" and "Antero" 0, 5, 10, and 15- dpi. **(C)** Quantification of WSMV reads (log counts per million) in RNA-seq samples (*n* = 4) collected at 10 dpi from leaf tissue under four conditions: *Wsm2-* (C), *Wsm2-* (T), *Wsm2+* (C) and *Wsm2+* (T), T indicates WSMV-treated condition, C indicates mock-treated condition. Two tailed *t*-tests were performed to compare genotypes. *$P < 0.5$, ***$P < 0.001$, ns = not significant.

Host Transcriptomic Response to WSMV Infection

To characterize host transcriptomic responses to WSMV infection, DEGs between mock- and WSMV-treated susceptible materials were analyzed (**Supplementary Table S11**). In total, 8,975 DEGs were detected between *Wsm2-* (T) and *Wsm2-* (C) conditions ($P < 0.01$), of which 5,031 were upregulated and 3,944 were downregulated after WSMV infection (**Figure 5A**).

Downregulated genes were most significantly enriched ($P < 0.01$) for the biological process GO terms relating to "photosynthesis" (GO:0015979) and metabolic processes, such as "purine ribonucleoside metabolic process" and "lipid metabolic process" (**Figure 5A**, **Supplementary Table S12**), indicating

that host plants suppress growth-related metabolic activity in response to WSMV infection. In contrast, upregulated genes were significantly enriched for the biological process GO terms related to "transport" (GO:0006810) and "localization" (GO:0051179) (**Figure 5A**, **Supplementary Table S12**). Among these upregulated genes three were *Pathogenesis-related 1* (*PR1*) genes, six *PR2* (β-1,3-glucanase) and *PR3* (chitinase) genes, four *PR5* (thaumatin-like protein) genes, and fifteen *PR9* (peroxidase) genes, together with two homoeologous genes encoding RNA-binding proteins ($P < 0.01$, **Supplementary Table S13**), suggesting their potential role in the host response to viral infection.
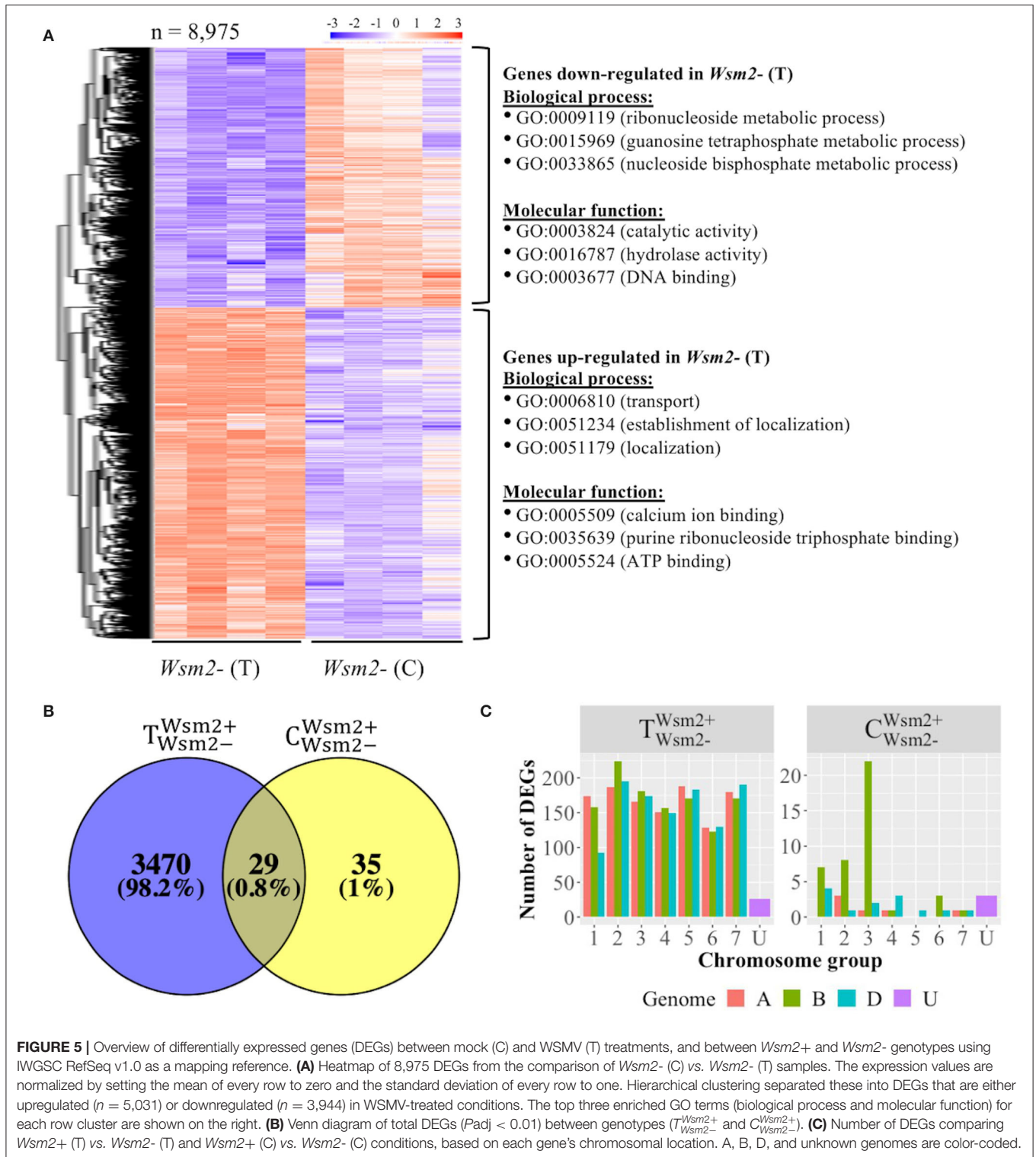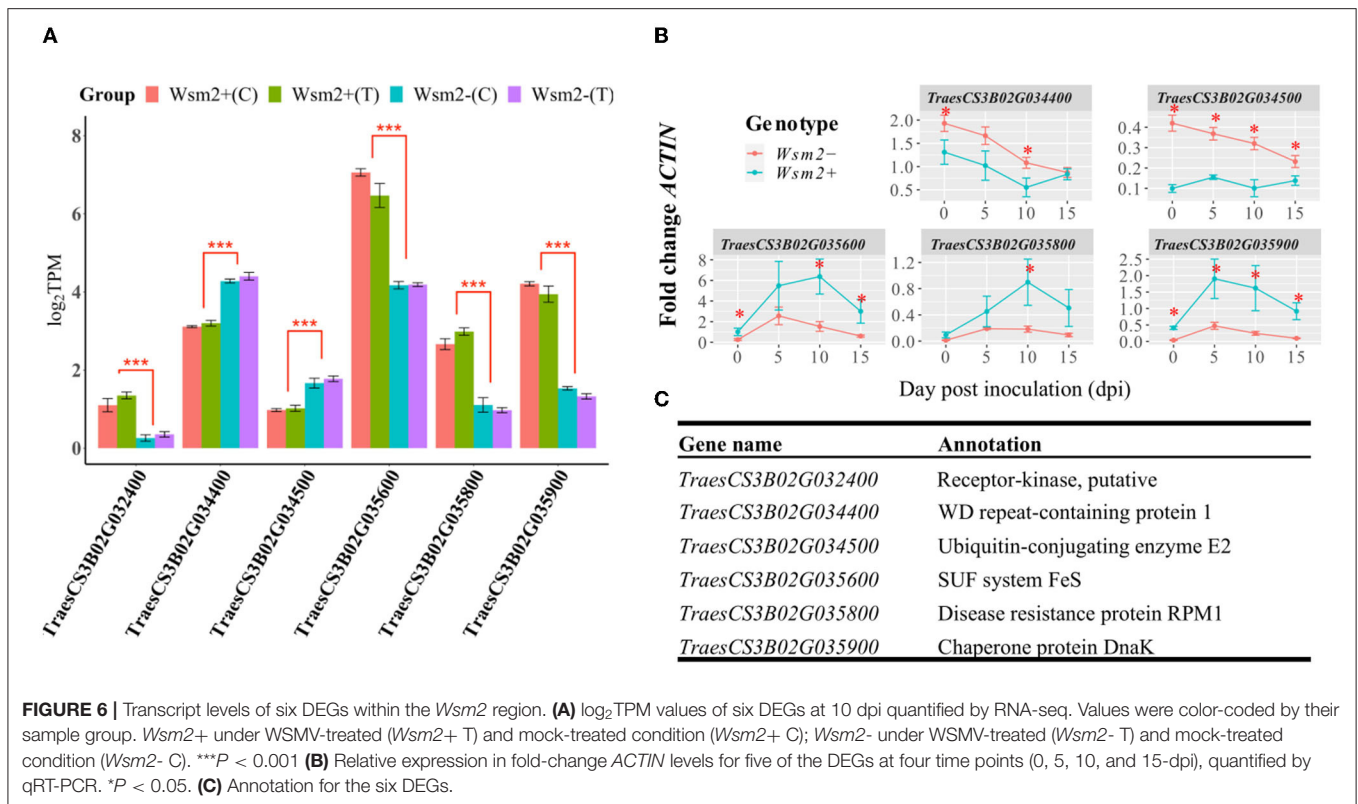
FIGURE 5 | Overview of differentially expressed genes (DEGs) between mock (C) and WSMV (T) treatments, and between *Wsm2+* and *Wsm2-* genotypes using IWGSC RefSeq v1.0 as a mapping reference. **(A)** Heatmap of 8,975 DEGs from the comparison of *Wsm2-* (C) *vs. Wsm2-* (T) samples. The expression values are normalized by setting the mean of every row to zero and the standard deviation of every row to one. Hierarchical clustering separated these into DEGs that are either upregulated ($n = 5,031$) or downregulated ($n = 3,944$) in WSMV-treated conditions. The top three enriched GO terms (biological process and molecular function) for each row cluster are shown on the right. **(B)** Venn diagram of total DEGs (*P*adj < 0.01) between genotypes (T_{Wsm2-}^{Wsm2+} and C_{Wsm2-}^{Wsm2+}). **(C)** Number of DEGs comparing *Wsm2+* (T) *vs. Wsm2-* (T) and *Wsm2+* (C) *vs. Wsm2-* (C) conditions, based on each gene's chromosomal location. A, B, D, and unknown genomes are color-coded.

Difference in Transcriptomic Response Between Genotypes

To characterize transcriptional changes between genotypes, DEGs were analyzed under mock-treated (C_{Wsm2-}^{Wsm2+}) and WSMV-treated (T_{Wsm2-}^{Wsm2+}) conditions. Only sixty-four genes were differentially expressed between genotypes in mock-treated conditions (**Figure 5B**), of which 28 were more highly expressed in *Wsm2+* genotypes, and 36 were more highly

FIGURE 6 | Transcript levels of six DEGs within the *Wsm2* region. **(A)** log$_2$TPM values of six DEGs at 10 dpi quantified by RNA-seq. Values were color-coded by their sample group. *Wsm2+* under WSMV-treated (*Wsm2+* T) and mock-treated condition (*Wsm2+* C); *Wsm2-* under WSMV-treated (*Wsm2-* T) and mock-treated condition (*Wsm2-* C). ***$P < 0.001$ **(B)** Relative expression in fold-change *ACTIN* levels for five of the DEGs at four time points (0, 5, 10, and 15-dpi), quantified by qRT-PCR. *$P < 0.05$. **(C)** Annotation for the six DEGs.

expressed in *Wsm2-* genotypes (**Supplementary Table S11**). Twenty-two (34.4%) of these genes are located on chromosome 3B (**Figure 5C**, **Supplementary Table S14**), and six are among the 142 candidate genes underlying *Wsm2* (see Section Examination of Candidate Genes Within *Wsm2* Found Six DEGs Between Genotypes).

In comparison, 3,499 genes were differentially expressed between genotypes under WSMV-treated conditions, of which 1,920 were more highly expressed in *Wsm2+* genotypes, and 1,579 were more highly expressed in *Wsm2-* genotypes (**Supplementary Table S11**). Twenty-nine genes were differentially expressed between the genotypes in both mock- and WSMV-treated conditions, while 3,470 genes were differentially expressed only under WSMV-treated conditions (**Figure 5B**). These results indicate that the response of the host plant to WSMV infection varies depending on the presence or absence of *Wsm2*. These 3,470 DEGs are significantly enriched for GO terms relating to different metabolic processes (GO:0046128, GO:0072521, GO:0033865) and catalytic activity (GO:0003824, GO:0016757) (**Supplementary Table S12**), indicating that the differences between genotypes in the days following WSMV infection include modified cellular metabolism and catalytic activity. However, examination of enriched GO terms for these DEGs did not find any "defense response," "hormone regulation," "signaling transduction," "cell wall biogenesis," or "photosynthesis"-related terms (**Supplementary Table S12**).

Analysis of Transcriptomes to Identify Candidate Genes Underlying *Wsm2*
Examination of Candidate Genes Within *Wsm2* Found Six DEGs Between Genotypes

Of the 142 annotated candidate genes within the *Wsm2* interval in the IWGSC RefSeq v1.0 genome assembly, six were differentially expressed between *Wsm2+* and *Wsm2-* genotypes in both WSMV-treated (T) and mock-treated (C) conditions (**Figure 6A**). The differential transcript levels at 10 dpi between genotypes of five of these genes were validated using qRT-PCR (**Figure 6B**), demonstrating the reliability of RNA-seq in quantifying transcript levels. Some were also significantly differentially expressed at earlier or later time points following inoculation, demonstrating they exhibit sustained differences in expression between genotypes (**Figure 6B**).

Two genes, encoding a WD repeat-containing protein 1 (*TraesCS3B02G034400*) and Ubiquitin-conjugating enzyme E2 (*TraesCS3B02G034500*), were more highly expressed in *Wsm2-* (**Figure 6C**, **Supplementary Table S15**). Both genes exhibit reduced copy number in "Snowmass" (**Figure 3**), potentially explaining their lower transcript levels in *Wsm2+* genotypes. The other four candidate genes were more highly expressed in *Wsm2+* genotypes, and encode a putative receptor kinase (*TraesCS3B02G032400*), an SUF system FeS protein (*TraesCS3B02G035600*), NBS-LRR type protein with homology to RPM1 (*TraesCS3B02G035800*), and a Chaperone protein DnaK (*TraesCS3B02G035900*) (**Figure 6C**).

FIGURE 7 | Expression analysis of *de novo* assembled transcripts absent from "Chinese Spring." Venn diagrams show the total number of differentially expressed transcripts (DETs) (*P*adj < 0.01) between **(A)** $Wsm2+_C^T$ and $Wsm2-_C^T$, **(B)** T_{Wsm2-}^{Wsm2+} and C_{Wsm2-}^{Wsm2+}, and **(C)** T_{Wsm2-}^{Wsm2+} and $Wsm2+_C^T$. **(D)** Expression in Log$_2$TPM of three DETs significant in both T_{Wsm2-}^{Wsm2+} and $Wsm2+_C^T$ contrasts. ***P < 0.001. **(E)** Annotations of three DETs.

None of the four UDP-glycosyltransferase genes predicted to exhibit increased copy number in "Snowmass" (**Figure 3**) were differentially expressed between genotypes (**Supplementary Table S15**), suggesting these genes are unlikely to contribute to WSMV resistance. Despite their potential roles in biotic stress resistance, all twelve BBI genes within the *Wsm2* interval exhibit low expression levels (TPM < 0.4) across all samples and were not differentially expressed between genotypes (**Supplementary Table S15**).

De novo Assembly of Unmapped Reads Revealed Transcripts Absent From the Wheat Reference Genome

To identify potential causative genes underlying *Wsm2* that are absent from the IWGSC RefSeq v1.0 genome, a *de novo* assembly of RNA-seq reads that did not map to this reference was performed. A total of 23,066,200 unmapped reads (5.4% of all reads, **Supplementary Table S9**) were combined from all 16 samples and assembled into 161,210 non-redundant transcripts.

The unmapped RNA-seq reads from each sample were mapped back to the *de novo* assembled transcriptome, revealing 245 transcripts that were differentially expressed in at least one of the four pairwise comparisons ($Wsm2+_C^T$, $Wsm2-_C^T$, T_{Wsm2-}^{Wsm2+}, and C_{Wsm2-}^{Wsm2+}, **Supplementary Table S16**). Of these, 56 were annotated as sequences from non-plant species, including 42 matching WSMV, and were excluded from the analysis. Of

the remaining 189 transcripts, nine were absent or expressed at very low levels in *Wsm2-* genotypes (defined as transcript levels >0.2 TPM in both WSMV- and mock-treated conditions, **Supplementary Table S17**). Using BLAST, these nine transcripts were confirmed to be absent from "Chinese Spring" and, with two exceptions, absent from the genomes of ten other wheat varieties (**Supplementary Tables S16, S17**). Among these *Wsm2+*-specific transcripts, one encodes an LRR receptor, one encodes a BBI trypsin inhibitor, and three transcripts encode Leaf rust 10 resistance proteins (**Supplementary Table S17**). Sixteen other transcripts were absent from *Wsm2+* genotypes, including one predicted to encode a negative regulator of resistance protein (**Supplementary Table S17**).

There were 116 transcripts differentially expressed between WSMV-treated and mock-treated conditions, including 43 only in the *Wsm2+* genotype, 63 only in the *Wsm2-* genotype, and ten shared between both genotypes (**Figure 7A**). Additionally, a total of 90 transcripts were differentially expressed between genotypes (18 only in mock-treated conditions, 56 only in WSMV-treated conditions, and 16 in both conditions, **Figure 7B**). Among these transcripts, three were induced by *WSMV* treatment in the *Wsm2+* genotype (**Figure 7C**). One transcript encoding a leaf rust 10 disease resistance locus receptor-like protein kinase (XM_037561459.1) was exclusively expressed in *Wsm2+* (T) conditions (TPM = 5, *P* < 0.001) (**Figure 7D, Supplementary Table S16**). This transcript was present in six

wheat varieties and in each case, was located on chromosome 3B, approximately 20 Mbp downstream of the left boundary marker for the *Wsm2* locus (**Supplementary Table S18**).

The two other transcripts exhibited significantly higher expression in *Wsm2*- genotypes compared to *Wsm2*+ genotypes in WSMV-treated conditions ($P < 0.001$), indicating they may be negatively associated with *Wsm2*-mediated resistance (**Figure 7D**, **Supplementary Table S16**). One transcript encodes a cytochrome P450 (XM_037560405.1) and was present at the distal end of chromosome arm 3BL in ten wheat varieties with genome assemblies (**Supplementary Table S18**). This is more than 800 Mbp from the *Wsm2* locus, suggesting it is unlikely to be located in the *Wsm2* region. The other transcript encodes a lectin-like receptor kinase gene (MT027257.1) and was absent from all wheat varieties, indicating that this likely represents a rare transcript in *Wsm2*+ genotypes (**Supplementary Table S18**).

DISCUSSION

Although the *Wsm2* locus has been deployed in multiple wheat varieties to confer WSMV resistance, the underlying gene has yet to be identified, limiting our understanding of viral resistance mechanisms in crops. In the current study, genomic analyses in seventeen wheat varieties revealed that *Wsm2* lies in a dynamic and variable region of the genome and is likely rare among modern wheat varieties. Transcriptomic analysis was used to characterize the molecular responses of host plants to WSMV infection and to identify a set of candidate genes that exhibit variation between *Wsm2*+ and *Wsm2*- genotypes.

Wsm2 Lies in a Highly Dynamic Region of the Genome and Is Likely Absent From Many Modern Wheat Varieties

The *Wsm2* locus was previously mapped to a 6.5 cM telomeric region of chromosome arm 3BS (Lu et al., 2012; Assanga et al., 2017; Tan et al., 2017), which corresponds to a 4.0 Mbp interval in the IWGSC RefSeq v1.0 wheat reference genome assembly (**Supplementary Table S4**). The region of this genome is dynamic and variable, with at least eight haplotypic groups and several instances of large insertions and deletions among common wheat varieties (**Figure 1**). Telomeric regions of the chromosomes are associated with high recombination rates, resulting in frequent duplication and divergence events that potentially contribute to this variation (See et al., 2006; Saintenac et al., 2009). In addition to structural variation, exome regions of *Wsm2* in "Snowmass" exhibited a higher variant rate than the average variant rate across all chromosomes (**Supplementary Table S8**). Functionally constrained regions of the genome containing essential genes exhibit reduced mutation rates and are subject to stronger purifying selection (Monroe et al., 2022). In contrast, resistance (*R*) genes tend to have a higher rate of variants than other genes, and many are located in clusters (Dolatabadian et al., 2017, 2020). It is more likely that the causative genes underlying *Wsm2* belong to an *R* gene type

rather than having an essential developmental role. The reduced purifying selection could contribute to the higher variant rate within this region.

The high rate of variants in the *Wsm2* region is shared between all studied varieties, including those that carry a susceptible *Wsm2* allele type (**Supplementary Table S8**). It is consistent with *Wsm2* originating from a common wheat genetic background. However, based on their broad WSMV-susceptibility, it is likely that the *Wsm2* resistance variant is absent from all seventeen wheat genotypes with sequenced genomes, including IWGSC RefSeq v1.0, which is derived from "Chinese Spring" (**Table 1**). This agrees with previous studies showing that "Chinese Spring" is susceptible to WSMV (Tan et al., 2017) and that *Wsm2* is absent from wild *Brachypodium* accessions, a monocot ancestor of wheat (Zhang and Hua, 2018). Therefore, although the wheat pangenome is a powerful resource to exploit natural variation and characterize genetic variants associated with agronomic traits and stress resistance (Walkowiak et al., 2020), the absence of rare genetic variants from sequenced wheat germplasm might limit their application for some gene discovery projects.

Variation within the *Wsm2* region may explain previous findings of inconsistent marker order across this locus in different mapping populations (Tan et al., 2017). It may complicate the application of marker-assisted selection. Despite the diverse haplotypes in these 17 wheat varieties, including four exhibiting the "AACT" haplotype, all genotypes exhibited a WSMV susceptible phenotype, although the result in "Jagger" is ambiguous (**Table 1**). The lack of association between these SNPs and WSMV resistance highlights the complexity of this locus, and that marker-assisted selection should be approached with care. Cloning *Wsm2* would allow for the development of perfect markers to confirm its presence in different wheat germplasm for introgressing this resistance allele in breeding programs. In addition to *Wsm2*, several other regions of the genome with weaker associations with WSMV resistance were detected (**Figure 2**). The peak marker information provided in **Supplementary Table S6** can be used to develop assays to validate and select these alleles during germplasm development for WSMV-resistant wheat varieties.

Candidate Genes Underlying *Wsm2*

That *Wsm2* lying in a dynamic region of the genome is consistent with a previous study showing multiple tandem duplication events involving clusters of BBI genes at this locus (Xie et al., 2021). Further genetic variation in the BBI family was detected in "Snowmass." Compared to the wheat reference genome, "Snowmass" is predicted to carry a non-functional allele of one BBI gene (*TraesCS3B02G038300*, **Supplementary Table S7**) and deletion of another (*TraesCS3B02G036200*, **Figure 3**). Although one *Wsm2*+ specific transcript annotated as a BBI was detected in "Snowmass" (**Supplementary Table S17**), overall, BBIs showed low transcript levels that were not significantly different between genotypes under WSMV-treated conditions, suggesting they are unlikely to confer WSMV resistance.

Wsm2 is a dominant allele (Lu et al., 2012), so it is likely that the variant is of gain-of-function. Therefore, while the "Snowmass" genome contains reduced copies of four genes within the *Wsm2* region (**Figure 3**) and lacks sixteen transcripts present in *Wsm2*- genotypes (**Supplementary Table S17**), these are unlikely to be causative variants for *Wsm2*.

Six candidate genes within the *Wsm2* interval in "Chinese Spring" were differentially expressed between genotypes (**Figure 6**). Two DEGs were upregulated in *Wsm2*+ throughout the WSMV infection time course (**Figure 6B**) and encode SUF system FeS (*TraesCS3B02G035600*) and Chaperone protein DnaK (*TraesCS3B02G035900*). The chaperone protein DnaK (HSP70) responds to both biotic and abiotic stress by helping to prevent the accumulation of excessive newly synthesized proteins and ensure proper protein folding during their transition process (Park and Seo, 2015). Another upregulated candidate in *Wsm2*+ (*TraesCS3B02G035800*) encodes a CC-NB-LRR domain protein with homology to *RPM1*. In *Arabidopsis*, this protein recognizes the avirulence factor, AvrRpm1 from the bacterial pathogen, *Pseudomonas syringae* pv. *maculicola* and triggers plant ETI defense responses (Grant et al., 1995).

It is also possible that *Wsm2* is a novel gene absent from the "Chinese Spring" reference genome. Additional copies of tandemly duplicated UDP-glycosyltransferase genes were identified in "Snowmass" (**Figure 3**). Although these additional copies were not associated with increased transcript levels at 10 dpi (**Supplementary Table S15**), they might be induced at other time points and play a role in plant defense against viral pathogens. UDP-glycosyltransferase genes have diverse roles in plant immunity against various types of pathogens. For example, UDP-glycosyltransferase proteins have been shown to function as negative regulators of the necrotrophic fungus, *Botrytis cinerea* in *Arabidopsis* (Castillo et al., 2019) and to promote resistance to the hemi-biotrophic bacterial pathogen, *Pseudomonas syringae* pv *tomato* carrying the *AvrRpm1* gene (Langlois-Meurinne et al., 2005). Moreover, the tomato gene *Twi1*, which encodes a glycosyltransferase, was shown to play a role in plant defense against tomato spotted wilt virus *via* secondary metabolites (Campos et al., 2019).

Through *de novo* assembly of unmapped RNA-seq reads, nine transcripts present in "Snowmass" but absent from "Antero," "Chinese Spring," and most other wheat genomes were identified (**Supplementary Table S17**). Five were annotated as disease-related genes (Leaf rust 10 disease resistance receptor protein kinase, Bowman-Birk inhibitor, and LRR receptor-like kinase). To determine whether these genes are responsible for WSMV resistance, it will be necessary to map their position in the "Snowmass" genome and perform functional characterization either by developing gene knockouts in "Snowmass" or transforming each gene into WSMV-susceptible varieties. Three novel transcripts exhibited differential expression between genotypes. They were also induced in response to WSMV infection and encode a leaf rust 10 disease resistance locus protein kinase, a lectin-like receptor kinase, and a cytochrome P450 (**Figure 7**, **Supplementary Table S18**). Cytochrome P450 proteins function in phytoalexin biosynthesis,

hormone metabolism regulation, and the biosynthesis of secondary metabolites and other defensive signaling molecules that regulate plant immunity against various pathogen types (Xu et al., 2015). The lectin-like receptor kinase gene (*LecRLK*) is a class of RLK that contains a lectin/lectin-like ectodomain that can bind to carbohydrates (Sun et al., 2020). *LecRLKs* are involved in plant basal defense against biotrophic and necrotrophic pathogens through carbohydrate signal perception, which triggers the PTI response (Sun et al., 2020). However, whether *LecRLKs* are also involved in ETI or play a role in plant response to viral infection remains unknown. The gene annotated as leaf rust 10 disease-resistance locus receptor-like protein kinase-like (*LRK10*) was first identified in wheat, providing resistance to the fungal pathogen, *Puccinia triticina*, which causes wheat brown rust (Feuillet et al., 1997, 1998). The *LRK10* gene was later characterized as an NLR-class of *R* gene in wheat with a strong diversifying selected N-terminal CC domain, suggesting a complex molecular mechanism of pathogen detection and signal transduction (Loutre et al., 2009). Although no evidence suggests that this *LRK10* is involved in plant defense response to viral pathogens, it is possible that this candidate could directly or indirectly interact with viral molecules and could be involved in downstream signal transduction pathways important in immunity.

A more comprehensive analysis of genetic variation would require genomic sequencing of "Snowmass." Although whole-genome sequencing would be feasible, targeted sequencing of a chromosome arm isolated by flow-sorting may be more cost-effective. This approach has been successfully applied to clone the broad-spectrum of *Lr22a* leaf-rust resistance gene in wheat (Thind et al., 2017) and would be a valuable approach to studying the *Wsm2* locus and other rare alleles in wheat germplasm.

Insight Into Wheat Host Transcriptomic Response to WSMV Infection

Despite the threat that viral pathogens pose to crop production, we have only a limited understanding of host antiviral immune mechanisms in monocot crops (Mandadi and Scholthof, 2013; Huang, 2021). The current study revealed that metabolic processes and photosynthesis were suppressed in susceptible hosts following WSMV infection (**Supplementary Table S12**). Our findings are consistent with previous studies, which showed that pathogen infection leads to the suppression of gene expression and protein production in photosynthetic processes (Bilgin et al., 2010; Göhre et al., 2012) due to the growth-to-defense tradeoff to optimize plant fitness and the efficient use of resources (Huot et al., 2017).

PR genes are induced upon pathogen infection and encode proteins associated with host defense responses (Ren et al., 2020). *PR1* genes are considered markers for plant resistance to biotrophic pathogens (Van Loon and Van Strien, 1999) and three wheat *PR1* genes were induced by WSMV infection (**Supplementary Table S13**), suggesting they may play a role in host response. The induction of *PR1* is usually associated

with the accumulation of salicylic acid (SA), a phytohormone involved in plant defense that may stimulate host antiviral responses through the inhibition of viral replication, cell-to-cell movement, and long-distance movement (Singh et al., 2004). However, there was no enrichment for phytohormone biosynthesis or signaling functional terms among genes responding to WSMV infection (**Supplementary Table S12**), and no SA, jasmonic acid (JA), or ethylene signaling genes were found among these DEGs (**Supplementary Table S11**). This suggests that these phytohormones may not be involved in the response of wheat to WSMV at this time point, although dedicated studies to characterize the role of SA or JA more broadly during WSMV infection will be required. Members of the RNA-binding proteins (RBPs) family can also regulate SA-mediated plant immune responses in *Arabidopsis* (Qi et al., 2010). One wheat RBP protein (TaUBA2C) was recently shown to interact with cysteine-rich protein from the Chinese wheat mosaic virus, which activates downstream defense responses to inhibit viral infection (Li et al., 2022). In the current study, *TaUBA2C* was induced two-fold following WSMV treatment (**Supplementary Table S13**), indicating that RBPs may play a role in the host response to WSMV infection.

In conclusion, genomic analyses indicate that *Wsm2* lies in a dynamic region of the wheat genome and is likely a rare allele in modern common varieties. This high rate of variation observed in *Wsm2* may complicate marker-assisted selection for virus resistance in breeding programs. Several candidate genes were identified, including genes found only in *Wsm2+* genotypes. Sequencing the "Snowmass" genome will facilitate the identification of *Wsm2*, which will expand our knowledge of viral resistance mechanisms in crops.

DATA AVAILABILITY STATEMENT

The RNA-seq dataset presented in this study can be found in the online Gene Expression Omnibus repository under accession number GSE190382. Full details of all data analysis steps and outputs are provided in supplementary files.

AUTHOR CONTRIBUTIONS

YX performed the experiment and bioinformatic data analysis and wrote the first draft of the manuscript. YX and SP designed the study and prepared the manuscript. PN and VN provided comments in the paper revision. All authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

We are grateful to Dr. Eduard Akhunov (Kansas State University) for providing exome sequencing data, to Dr. Shuyu Liu (Texas A & M University) for helpful advice on WSMV phenotyping, and to Dr. Scott Haley, Emily Hudson-Arns and Tori Anderson for providing GBS markers and DH plant materials used in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.928949/full#supplementary-material

REFERENCES

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., and Cech, M., et al. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucl Acids Res.* 46, 537–544. doi: 10.1093/nar/gky379

Albrecht, T., White, S., Layton, M., Stenglein, M., Haley, S., and Nachappa, P. (2022). Occurrence of wheat curl mite and mite-vectored viruses of wheat in Colorado and insights into the wheat virome. *Plant Dis.* doi: 10.1094/PDIS-02-21-0352-RE

Arends, D., Prins, P., Jansen, R. C., and Broman, K. W. (2010). R/qtl: high-throughput multiple QTL mapping. *Bioinform. Oxf. Engl.* 26, 2990–2992. doi: 10.1093/bioinformatics/btq565

Assanga, S., Zhang, G., Tan, C. T., Rudd, J. C., Ibrahim, A., Xue, Q., et al. (2017). Saturated genetic mapping of Wheat streak mosaic virus resistance gene *Wsm2* in wheat. *Crop Sci.* 57, 332–339. doi: 10.2135/cropsci2016.04.0233

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bilgin, D. D., Zavala, J. A., Zhu, J., Clough, S. J., Ort, D. R., and Delucia, E. H. (2010). Biotic stress globally downregulates photosynthesis genes. *Plant Cell Environ.* 33, 1597–1613. doi: 10.1111/j.1365-3040.2010.02167.x

Cabanettes, F., and Klopp, C. (2018). D-Genies: dot plot large genomes in an interactive, efficient and simple way. *PeerJ.* 6, e4958. doi: 10.7717/peerj.4958

Campos, L., López-Gresa, M. P., Fuertes, D., Bellés, J. M., Rodrigo, I., and Lisón, P. (2019). Tomato glycosyltransferase *Twi1* plays a role in flavonoid glycosylation and defence against virus. *BMC Plant Biol.* 19, 450. doi: 10.1186/s12870-019-2063-9

Castillo, N., Pastor, V., Chávez, Á., Arr,ó, M., Boronat, A., Flors, V., et al. (2019). Inactivation of UDP-glucose sterol glucosyltransferases enhances *Arabidopsis* resistance to *Botrytis cinerea*. *Front. Plant Sci.* 10, 1162–1162. doi: 10.3389/fpls.2019.01162

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)* 6, 80–92. doi: 10.4161/fly.19695

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10, giab008. doi: 10.1093/gigascience/giab008

Danilova, T. V., Zhang, G., Liu, W., Friebe, B., and Gill, B. S. (2017). Homoeologous recombination-based transfer and molecular cytogenetic mapping of a Wheat streak mosaic virus and *Triticum* mosaic virus resistance gene *Wsm3* from *Thinopyrum intermedium* to wheat. *Theor. Appl. Genet.* 130, 549–556. doi: 10.1007/s00122-016-2834-8

Dhakal, S., Tan, C. T., Anderson, V., Yu, H., Fuentealba, M. P., Rudd, J. C., et al. (2018). Mapping and KASP marker development for wheat curl mite resistance in "TAM 112" wheat using linkage and association analysis. *Mol. Breed.* 38, 119–119. doi: 10.1007/s11032-018-0879-x

Dhakal, S., Tan, C. T., Paezold, L., Fuentealba, M. P., Rudd, J. C., Blaser, B. C., et al. (2017). Wheat curl mite resistance in hard winter wheat in the US great plains. *Crop Sci.* 57, 53–61. doi: 10.2135/cropsci2016.02.0121

Divis, L. A., Graybosch, R. A., Peterson, C. J., Baenziger, P. S., Hein, G. L., Beecher, B. B., et al. (2006). Agronomic and quality effects in winter wheat of a gene conditioning resistance to Wheat streak mosaic virus. *Euphytica* 152, 41–49. doi: 10.1007/s10681-006-9174-8

Dobin, A., and Gingeras, T. R. (2015). Mapping RNA-seq reads with STAR. *Curr. Protoc. Bioinforma.* 51, 11.14.1–11.14.19. doi: 10.1002/0471250953.bi1114s51

Dolatabadian, A., Bayer, P. E., Tirnaz, S., Hurgobin, B., Edwards, D., and Batley, J. (2020). Characterization of disease resistance genes in the *Brassica napus* pangenome reveals significant structural variation. *Plant Biotechnol. J.* 18, 969–982. doi: 10.1111/pbi.13262

Dolatabadian, A., Patel, D. A., Edwards, D., and Batley, J. (2017). Copy number variation and disease resistance in plants. *Theor. Appl. Genet.* 130, 2479–2490. doi: 10.1007/s00122-017-2993-2

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6, e19379–e19379. doi: 10.1371/journal.pone.0019379

FAOSTAT (2021). *Food and Agriculture Organization of the United Nations (FAO).* FAOSTAT Database. Available online at: http://faostat.fao.org/site/291/default. aspx

Fellers, J. P., Webb, C., Fellers, M. C., Shoup Rupp, J., and De Wolf, E. (2019). Wheat virus identification within infected tissue using nanopore sequencing technology. *Plant Dis.* 103, 2199–2203. doi: 10.1094/PDIS-09-18-1700-RE

Feuillet, C., Reuzeau, C., Kjellbom, P., and Keller, B. (1998). Molecular characterization of a new type of receptor-like kinase (*wlrk*) gene family in wheat. *Plant Mol. Biol.* 37, 943–953. doi: 10.1023/A:1006062016593

Feuillet, C., Schachermayr, G., and Keller, B. (1997). Molecular cloning of a new receptor-like kinase gene encoded at the *Lr10* disease resistance locus of wheat. *Plant J.* 11, 45–52. doi: 10.1046/j.1365-313X.1997.11010045.x

Friebe, B., Qi, L. L., Wilson, D. L., Chang, Z. J., Seifers, D. L., Martin, T. J., et al. (2009). Wheat-*Thinopyrum intermedium* recombinants resistant to Wheat streak mosaic virus and *Triticum* mosaic virus. *Crop Sci.* 49, 1221–1226. doi: 10.2135/cropsci2008.09.0513

Göhre, V., Jones, A. M. E., Sklenár, J., Robatzek, S., and Weber, A. P. M. (2012). Molecular crosstalk between PAMP-triggered immunity and photosynthesis. *Mol. Plant-Microbe Interactions®* 25 1083–1092. doi: 10.1094/MPMI-11-11-0301

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Grant, M. R., Godiard, L., Straube, E., Ashfield, T., Lewald, J., Sattler, A., et al. (1995). Structure of the *Arabidopsis RPM1* gene enabling dual specificity disease resistance. *Science* 269, 843–846. doi: 10.1126/science.7638602

Haber, S., Seifers, D. L., and Thomas, J. (2006). A new source of resistance to Wheat streak mosaic virus (WSMV) in spring wheat. *Can. J. Plant Pathol.* 28, 324.

Hadi, B. A. R., Langham, M. A. C., Osborne, L., and Tilmon, K. J. (2011). Wheat streak mosaic virus on wheat: Biology and management. *J. Integr. Pest Manag.* 2, J1–J5. doi: 10.1603/IPM10017

Haley, S. D., Johnson, J. J., Peairs, F. B., Stromberger, J. A., Heaton, E. E., Seifert, S. A., et al. (2011). Registration of 'Snowmass' wheat. *J. Plant Regist.* 5, 87–90. doi: 10.3198/jpr2010.03.0175crc

Haley, S. D., Martin, T. J., Quick, J. S., Seifers, D. L., Stromberger, J. A., Clayshulte, S. R., et al. (2002). Registration of "CO960293-2" wheat germplasm resistant to Wheat streak mosaic virus and russian wheat aphid. *Crop Sci.* 42, 1381–1382. doi: 10.2135/cropsci2002.1381

Harvey, T. L., Seifers, D. L., Martin, T. J., Brown-Guedira, G., and Gill, B. S. (1999). Survival of wheat curl mites on different sources of resistance in wheat. *Crop Sci.* 39, 1887–1889. doi: 10.2135/cropsci1999.3961887x

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, 896–904. doi: 10.1038/s41588-019-0382-2

Huang, C. (2021). From player to pawn: viral avirulence factors involved in plant immunity. *Viruses* 13, 688. doi: 10.3390/v13040688

Hunger, R., Sherwood, J., Evans, C. K., and Montana, J. (1992). Effects of planting date and inoculation date on severity of Wheat streak mosaic in hard red winter wheat cultivars. *Plant Dis.* 76, 1056–1060. doi: 10.1094/PD-76-1056

Huot, B., Castroverde, C. D. M., Velásquez, A. C., Hubbard, E., Pulman, J. A., Yao, J., et al. (2017). Dual impact of elevated temperature on plant defence and bacterial virulence in *Arabidopsis*. *Nat. Commun.* 8, 1–11. doi: 10.1038/s41467-017-01674-2

IWGSC, Appels, R., Eversole, K., Stein, N., Feuillet, C., Keller, B., 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

Jordan, K. W., Wang, S., Lun, Y., Gardiner, L. J., MacLachlan, R., Hucl, P., et al. (2015). A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. *Genome Biol.* 16, 48–48. doi: 10.1186/s13059-015-0606-4

Kumssa, T. T., Rupp, J. S., Fellers, M. C., Fellers, J. P., and Zhang, G. (2019). An isolate of Wheat streak mosaic virus from foxtail overcomes *Wsm2* resistance in wheat. *Plant Pathol.* 68, 783–789. doi: 10.1111/ppa.12989

Langlois-Meurinne, M., Gachon, C. M. M., and Saindrenan, P. (2005). Pathogen-responsive expression of glycosyltransferase genes *UGT73B3* and *UGT73B5* is necessary for resistance to *Pseudomonas syringae* pv *tomato* in *Arabidopsis*. *Plant Physiol.* 139, 1890–1901. doi: 10.1104/pp.105.067223

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Li, J., Feng, H., Liu, S., Liu, P., Chen, X., Yang, J., et al. (2022). Phosphorylated viral protein evades plant immunity through interfering the function of RNA-binding protein. *PLoS Pathog.* 18, e1010412. doi: 10.1371/journal.ppat.1010412

Liao, Y., Smyth, G. K., and Shi, W. (2014). FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Liu, S., Assanga, S. O., Dhakal, S., Gu, X., Tan, C.-T., Yang, Y., et al. (2016). Validation of chromosomal locations of 90K array single nucleotide polymorphisms in US wheat. *Crop Sci.* 56, 364–373. doi: 10.2135/cropsci2015.03.0194

Liu, W., Seifers, D. L., Qi, L. L., Friebe, B., and Gill, B. S. (2011). A compensating wheat–*Thinopyrum intermedium* robertsonian translocation conferring resistance to Wheat streak mosaic virus and *Triticum* mosaic virus. *Crop Sci.* 51, 2382–2390. doi: 10.2135/cropsci2011.03.0118

Loutre, C., Wicker, T., Travella, S., Galli, P., Scofield, S., Fahima, T., et al. (2009). Two different CC-NBS-LRR genes are required for *Lr10*-mediated leaf rust resistance in tetraploid and hexaploid wheat. *Plant J.* 60, 1043–1054. doi: 10.1111/j.1365-313X.2009.04024.x

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550–550. doi: 10.1186/s13059-014-0550-8

Lu, H., Kottke, R., Devkota, R., Amand, P. S., Bernardo, A., Bai, G., et al. (2012). Consensus mapping and identification of markers for marker-assisted selection of *Wsm2* in wheat. *Crop Sci.* 52, 720–728. doi: 10.2135/cropsci2011.07.0363

Lu, H., Price, J., Devkota, R., Rush, C., and Rudd, J. (2011). A dominant gene for resistance to Wheat streak mosaic virus in winter wheat line "CO960293-2." *Crop Sci.* 51, 5–12. doi: 10.2135/cropsci2010.01.0038

Malik, R., Brown-Guedira, G. L., Smith, C. M., Harvey, T. L., and Gill, B. S. (2003). Genetic mapping of wheat curl mite resistance genes *Cmc3* and *Cmc4* in common wheat. *Crop Sci.* 43, 644–650. doi: 10.2135/cropsci2003.0644

Mandadi, K. K., and Scholthof, K. B. G. (2013). Plant immune responses against viruses: How does a virus cause disease? *Plant Cell* 25, 1489–1505. doi: 10.1105/tpc.113.111658

Martin, J. T., Fritz, A. K., Seifers, D. L., and Shroyer, J. P. (2007). "RonL" hard white wheat. *Kans. State Univ. Agric. Exp. Stn. Coop. Ext. Serv.* 926, 3.

Martin, T. J., Zhang, G., Fritz, A. K., Miller, R., and Chen, M. S. (2014). Registration of 'Clara CL' wheat. *J. Plant Regist.* 8, 38–42. doi: 10.3198/jpr2013.07.0040crc

McKelvy, U., Brelsford, M., Sherman, J., and Burrows, M. (2021). Reactions of winter wheat, spring wheat, and barley cultivars to mechanical inoculation with Wheat streak mosaic virus. *Plant Health Prog.* 23, 72–81. doi: 10.1094/PHP-10-20-0083-RS

Monroe, J. G., Srikant, T., Carbonell-Bejerano, P., Becker, C., Lensink, M., Exposito-Alonso, M., et al. (2022). Mutation bias reflects natural selection in Arabidopsis thaliana. *Nature.* 602, 101–105. doi: 10.1038/s41586-021-04269-6

Murugan, M., Cardona, P. S., Duraimurugan, P., Whitfield, A. E., Schneweis, D., Starkey, S., et al. (2011). Wheat curl mite resistance: Interactions of mite feeding with Wheat streak mosaic virus infection. *J. Econ. Entomol.* 104, 1406–1414. doi: 10.1603/EC11112

Nachappa, P., Haley, S., and Pearce, S. (2021). Resistance to the wheat curl mite and mite-transmitted viruses: challenges and future directions. *Curr. Opin. Insect Sci.* 45, 21–27. doi: 10.1016/j.cois.2020.11.003

Navia, D., de Mendonça, R. S., Skoracka, A., Szydło, W., Knihinicki, D., Hein, G. L., et al. (2013). Wheat curl mite, *Aceria tosichella*, and transmitted viruses: an expanding pest complex affecting cereal crops. *Exp. Appl. Acarol.* 59, 95–143. doi: 10.1007/s10493-012-9633-y

Oliveros, J. C. (2007). Venny. An interactive tool for comparing lists with venn diagrams. Available online at: https://bioinfogp.cnb.csic.es/tools/venny/ (accessed August 06, 2022).

Orlob, G. B. (1966). Feeding and transmission characteristics of *aceria tulipae* keifer as vector of Wheat streak mosaic virus. *J. Phytopathol.* 55, 218–238. doi: 10.1111/j.1439-0434.1966.tb02228.x

Park, C. J., and Seo, Y. S. (2015). Heat shock proteins: a review of the molecular chaperones for plant immunity. *Plant Pathol. J.* 31, 323–333. doi: 10.5423/PPJ.RW.08.2015.0150

Plagnol, V., Curtis, J., Epstein, M., Mok, K. Y., Stebbings, E., Grigoriadou, S., et al. (2012). A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics* 28, 2747–2754. doi: 10.1093/bioinformatics/bts526

Price, J. A., Smith, J., Simmons, A., Fellers, J., and Rush, C. M. (2010). Multiplex real-time RT-PCR for detection of Wheat streak mosaic virus and *Tritcum* mosaic virus. *J. Virol. Methods* 165, 198–201. doi: 10.1016/j.jviromet.2010.01.019

Qi, Y., Tsuda, K., Joe, A., Sato, M., Nguyen, L. V., Glazebrook, J., et al. (2010). A putative RNA-binding protein positively regulates salicylic acid–mediated immunity in *Arabidopsis*. *Mol. Plant Microbe Interact.* 23 1573–1583. doi: 10.1094/MPMI-05-10-0106

Redila, C. D., Phipps, S., and Nouri, S. (2021). Full genome evolutionary studies of Wheat streak mosaic-associated viruses using high-throughput sequencing. *Front. Microbiol.* 12, 699078. doi: 10.3389/fmicb.2021.699078

Ren, Z., Liu, J., Din, G. M. U., Zhang, H., Du, Z., Chen, W., et al. (2020). Transcriptome analysis of wheat spikes in response to *Tilletia controversa* Kühn which cause wheat dwarf bunt. *Sci. Rep.* 10, 21567. doi: 10.1038/s41598-020-78628-0

Saintenac, C., Falque, M., Martin, O. C., Paux, E., Feuillet, C., and Sourdille, P. (2009). Detailed recombination studies along chromosome 3B provide new insights on crossover distribution in wheat (*Triticum aestivum* L.). *Genetics* 181, 393–403. doi: 10.1534/genetics.108.097469

Santra, M., Wang, H., Seifert, S., and Haley, S. (2017). "Doubled haploid laboratory protocol for wheat using wheat–maize wide hybridization," in *Methods in Molecular Biology*, eds. Bhalla, P. L., and Singh, M. B., (New York, NY: Springer New York), 235–249.

See, D. R., Brooks, S., Nelson, J. C., Brown-Guedira, G., Friebe, B., and Gill, B. S. (2006). Gene evolution at the ends of wheat chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* 103, 4162–4167. doi: 10.1073/pnas.0508942102

Seifers, D. L., Martin, T. J., Harvey, T. L., and Gill, B. S. (1995). Temperature sensitivity and efficacy of Wheat streak mosaic virus resistance derived from *Agropyron intermedium*. *Plant Dis.* 79, 1104–1106. doi: 10.1094/PD-79-1104

Seifers, D. L., Martin, T. J., Harvey, T. L., Haber, S., and Haley, S. D. (2006). Temperature sensitivity and efficacy of Wheat streak mosaic virus resistance derived from "CO960293" wheat. *Plant Dis.* 90, 623–628. doi: 10.1094/PD-90-0623

Sharp, G. L., Martin, J. M., Lanning, S. P., Blake, N. K., Brey, C. W., Sivamani, E., et al. (2002). Field evaluation of transgenic and classical sources of Wheat streak mosaic virus resistance. *Crop Sci.* 42, 105–110. doi: 10.2135/cropsci2002.1050

Singh, D. P., Moore, C. A., Gilliland, A., and Carr, J. P. (2004). Activation of multiple antiviral defence mechanisms by salicylic acid. *Mol. Plant Pathol.* 5, 57–63. doi: 10.1111/j.1364-3703.2004.00203.x

Singh, K., and Kundu, J. K. (2018). Wheat streak mosaic virus. *Plant Viruses* 18:131−148. doi: 10.1201/b22221-8

Singh, K., Wegulo, S. N., Skoracka, A., and Kundu, J. K. (2018). Wheat streak mosaic virus: a century old virus with rising importance worldwide. *Mol. Plant Pathol.* 19, 2193–2206. doi: 10.1111/mpp.12683

Slykhuis, J. T. (1955). *Aceria tulipae* keifer (acarina: eriophyidae) in relation to the spread of Wheat streak mosaic. *Phytopathology* 45, 116–128. doi: 10.1111/j.1399-3054.1955.tb07793.x

Stenger, D. C., Hall, J. S., Choi, I. R., and French, R. (1998). Phylogenetic relationships within the family *Potyviridae*: wheat streak mosaic virus and brome streak mosaic virus are not members of the genus *Rymovirus*. *Phytopathology* 88, 782–787. doi: 10.1094/PHYTO.1998.88.8.782

Sun, Y., Qiao, Z., Muchero, W., and Chen, J. G. (2020). Lectin receptor-like kinases: the sensor and mediator at the plant cell surface. *Front. Plant Sci.* 11, 1–11. doi: 10.3389/fpls.2020.596301

Tan, C. T., Assanga, S., Zhang, G., Rudd, J. C., Haley, S. D., Xue, Q., et al. (2017). Development and validation of KASP markers for wheat streak mosaic virus resistance gene *Wsm2*. *Crop Sci.* 57, 340–349. doi: 10.2135/cropsci2016.04.0234

Taylor, J., and Butler, D. (2017). R package ASMap: Efficient genetic linkage map construction and diagnosis. *J. Stat. Softw.* 79, 1–29. doi: 10.18637/jss.v079.i06

Thind, A. K., Wicker, T., Šimková, H., Fossati, D., and Moullet, O., Brabant, C., et al. (2017). Rapid cloning of genes in hexaploid wheat using cultivar-specific long-range chromosome assembly. *Nat. Biotechnol.* 35, 793–796. doi: 10.1038/nbt.3877

Thomas, J. B., and Conner, R. L. (1986). Resistance to colonization by the wheat curl mite in *Aegilops squarrosa* and its inheritance after transfer to common wheat. *Crop Sci.* 26, 527–530. doi: 10.2135/cropsci1986.0011183X002600030019x

Van Loon, L. C., and Van Strien, E. A. (1999). The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. *Physiol. Mol. Plant Pathol.* 55, 85–97. doi: 10.1006/pmpp.1999.0213

Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M. T., Brinton, J., et al. (2020). Multiple wheat genomes reveal global variation in modern breeding. *Nature* 588, 277–283. doi: 10.1038/s41586-020-2961-x

Wells, D. G., Kota, R. S., Sandhu, H. S., Gardner, W. S., and Finney, K. F. (1982). Registration of one disomic substitution line and five translocation lines of winter wheat germplasm resistant to Wheat streak mosaic virus. *Crop Sci.* 22, 114–123. doi: 10.2135/cropsci1982.0011183X002200060083x

Whelan, E. D. P., and Hart, G. E. (1988). A spontaneous translocation that transfers wheat curl mite resistance from decaploid *Agropyron elongatum* to common wheat. *Genome* 30, 289–292. doi: 10.1139/g88-050

Xie, Y., Ravet, K., and Pearce, S. (2021). Extensive structural variation in the Bowman-Birk inhibitor family in common wheat (*Triticum aestivum* L.). *BMC Genomics* 22, 218–218. doi: 10.1186/s12864-021-07475-8

Xu, J., Wang, X. Y., and Guo, W. Z. (2015). The cytochrome P450 superfamily: key players in plant development and defense. *J. Integr. Agric.* 14, 1673–1686. doi: 10.1016/S2095-3119(14)60980-1

Zhang, G., and Hua, Z. (2018). Genome comparison implies the role of *Wsm2* in membrane trafficking and protein degradation. *PeerJ.* 2018, 1–18. doi: 10.7717/peerj.4678

Zhang, G., Martin, T. J., Fritz, A. K., Miller, R., Chen, M.-S., Bowden, R. L., et al. (2015). Registration of 'Oakley CL' wheat. *J. Plant Regist.* 9, 190–195. doi: 10.3198/jpr2014.04.0023crc

Zhang, G., Martin, T. J., Fritz, A. K., Miller, R., Chen, M.-S., Bowden, R. L., et al. (2016). Registration of 'Joe' hard white winter wheat. *J. Plant Regist.* 10, 283–286. doi: 10.3198/jpr2016.02.0007crc

Zhu, T., Wang, L., Rimbert, H., Rodriguez, J. C., Deal, K. R., De Oliveira, R., et al. (2021). Optical maps refine the bread wheat *Triticum aestivum* cv. Chinese Spring genome assembly. *Plant J.* 107, 303–314. doi: 10.1111/tpj.15289