Check for updates

The Haplotype-Based Analysis of *Aegilops tauschii* Introgression Into Hard Red Winter Wheat and Its Impact on Productivity Traits

Moses Nyine[1], Elina Adhikari[1], Marshall Clinesmith[2], Robert Aiken[2], Bliss Betzen[1], Wei Wang[1], Dwight Davidson[1], Zitong Yu[1], Yuanwen Guo[1], Fei He[1], Alina Akhunova[3], Katherine W. Jordan[1,4], Allan K. Fritz[2] and Eduard Akhunov[1]*

[1] Department of Plant Pathology, Kansas State University, Manhattan, KS, United States, [2] Department of Agronomy, Kansas State University, Manhattan, KS, United States, [3] Integrated Genomics Facility, Kansas State University, Manhattan, KS, United States, [4] United States Department of Agriculture, Agricultural Research Service Hard Winter Wheat Genetics Research Unit, Manhattan, KS, United States

The introgression from wild relatives have a great potential to broaden the availability of beneficial allelic diversity for crop improvement in breeding programs. Here, we assessed the impact of the introgression from 21 diverse accessions of *Aegilops tauschii*, the diploid ancestor of the wheat D genome, into 6 hard red winter wheat cultivars on yield and yield component traits. We used 5.2 million imputed D genome SNPs identified by the whole-genome sequencing of parental lines and the sequence-based genotyping of introgression population, including 351 $BC_1F_{3:5}$ lines. Phenotyping data collected from the irrigated and non-irrigated field trials revealed that up to 23% of the introgression lines (ILs) produce more grain than the parents and check cultivars. Based on 16 yield stability statistics, the yield of 12 ILs (3.4%) was stable across treatments, years, and locations; 5 of these lines were also high yielding lines, producing 9.8% more grain than the average yield of check cultivars. The most significant SNP- and haplotype-trait associations were identified on chromosome arms 2DS and 6DL for the spikelet number per spike (SNS), on chromosome arms 2DS, 3DS, 5DS, and 7DS for grain length (GL) and on chromosome arms 1DL, 2DS, 6DL, and 7DS for grain width (GW). The introgression of haplotypes from *A. tauschii* parents was associated with an increase in SNS, which was positively correlated with a heading date (HD), whereas the haplotypes from hexaploid wheat parents were associated with an increase in GW. We show that the haplotypes on 2DS associated with an increase in the spikelet number and HD are linked with multiple introgressed alleles of *Ppd-D1* identified by the whole-genome sequencing of *A. tauschii* parents. Meanwhile, some introgressed haplotypes exhibited significant pleiotropic effects with the direction of effects on the yield component traits being largely consistent with the previously reported trade-offs, there were haplotype combinations associated with the positive trends in yield. The characterized repertoire

of the introgressed haplotypes derived from *A. tauschii* accessions with the combined positive effects on yield and yield component traits in elite germplasm provides a valuable source of alleles for improving the productivity of winter wheat by optimizing the contribution of component traits to yield.

Keywords: wheat, *Aegilops tauschii*, wild relative introgression, whole genome sequencing, haplotypes, genotype imputation, grain yield, yield component traits

INTRODUCTION

The gap between population expansion and food production is increasing due to marginal improvements in crop yield, which is attributed to a decline in soil fertility, pests and diseases, and climate change (Bailey-Serres et al., 2019). Wild relatives of wheat are a rich source of novel underutilized allelic diversity with a great potential to improve cultivated wheat through introgression (Placido et al., 2013; Zhang et al., 2017; Hao et al., 2020). The introgression from wild relatives into elite wheat cultivars was reported to increase pest and disease resistance (Periyannan et al., 2013; Saintenac et al., 2013), improve resilience toward environmental stress (Peleg et al., 2005; Placido et al., 2013), and increase yield (Pasquariello et al., 2020). The success of introgression breeding, however, could be affected by the negative epistasis between multiple alleles of wild and cultivated wheat (Nyine et al., 2020), especially in the low recombination regions of chromosomes, where the linkage with the negatively selected alleles could reduce the efficiency of selection for beneficial variants (Hill and Robertson, 1966).

Introgression could exhibit pleotropic effects, affecting multiple, often unrelated traits not directly targeted by selection. For example, the introgression from *Aegilops ventricosa* into the chromosome 7D of wheat was associated with an increase in grain protein content and resistance to eyespot at the expense of reduced yield (Pasquariello et al., 2020). In durum wheat, the introgression of the *GNI-A1* gene from wild emmer increased grain weight by suppressing the fertility of distal florets, resulting in a negative correlation between grain number (GN) and grain weight (Golan et al., 2019). The introgression from *Agropyron elongatum* into the 7DL chromosome arm of wheat that is known to confer leaf rust resistance (*Lr19*) (Wang and Zhang, 1996) also influences root development, resulting in an improved adaption to water stress (Placido et al., 2013) and salinity (Dvořák et al., 1988), and increased biomass (BM) and yield (Reynolds et al., 2001).

Crop yield is a complex trait determined by many component traits, such as 1,000 grain weight (TGW), GN per spike, spikes per unit area, grain width (GW), area, length, etc. (Del Moral et al., 2003; Zhang et al., 2018; Du et al., 2020). Previous studies have shown that the introgression from wild and close relatives improve the yield of hexaploid wheat by changing yield component traits through the pleotropic interaction between introgressed and background alleles of the hexaploid wheat (Jones et al., 2020). Significant trade-offs between yield, yield components, and yield stability have been reported in wheat. A study by Pennacchi et al. (2019) showed that yield and

yield stability have a negative linear relationship in most cases. Other factors such as HD, plant height (PH), and BM influence the source-sink ratio, which in turn affects the harvest index leading to a variation in yield and yield stability (Reynolds et al., 2001, 2020). Balancing the trade-off between yield components is therefore necessary to improve yield, maximize the yield potential, and improve the yield stability in wheat.

Sequence-based genotyping generates high-density SNP marker data that could be used to accurately detect the boundaries of genomic segments introgressed from wild relatives (Kuzay et al., 2019; Nyine et al., 2020), providing a unique opportunity to investigate the effect of introgression on trade-off between the traits affecting total yield. Even though whole-genome sequencing became less expensive, it is still not within the cost range that would allow wheat breeding programs to sequence large populations. Sequencing of founder lines at a high coverage depth and using these genotypes as a reference panel to impute missing and ungenotyped markers in the progeny characterized by low-coverage skim sequencing is a cost-effective alternative. The existing imputation algorithms (Browning and Browning, 2013; Swarts et al., 2014; Davies et al., 2016) provide a highly accurate whole-genome prediction of missing genotypes and were shown to increase the power of genome-wide association (GWAS) scans, thus enabling the identification of SNPs or haplotypes associated with the traits of interest (Li et al., 2010; Nyine et al., 2019). One of the advantages of the increased marker density provided by whole-genome sequencing is the ability to perform association mapping using haplotype information, which improves the detection of quantitative trait loci that would otherwise be missed when using single SNPs (Zhang et al., 2002; Lorenz et al., 2010).

Here, we investigated the impact of the introgression from *A. tauschii* into hard red winter wheat lines on yield and yield component traits and how haplotypes from *A. tauschii* at different genomic loci affect the component traits and total yield. For this purpose, we assessed the phenotypic variability of yield and the component traits, BM, and tenacious glume (Tg) traits in an introgression population derived from *A. tauschii* and hexaploid winter wheat phenotyped under irrigated and rainfed conditions. We applied the SNP diversity data generated by the whole-genome shotgun sequencing at 10× coverage level of the parental lines to impute genotypes in this population (Nyine et al., 2020). This strategy resulted in 5.2 million SNPs that enabled us to identify the introgressed *A. tauschii* haplotypes and assess their effects on the trait variation through the GWAS mapping and haplotype block effect analysis. This introgression population along with high-density SNP genotype data provides

a valuable resource for the effective deployment of *A. tauschii* haplotypes in winter wheat improvement programs.

MATERIALS AND METHODS
Plant Materials

The study population was described in detail by Nyine et al. (2020). In brief, the population was developed by crossing synthetic *A. tauschii*-wheat octoploid lines with recurrent hexaploid winter wheat lines. The octoploid lines were developed by crossing 5 hexaploid winter wheat lines with 21 diverse *A. tauschii* accessions. The resulting F_1 hybrid plants regenerated from the rescued embryos were treated with colchicine to produce synthetic octoploids (Dale et al., 2017). The octoploids were backcrossed once to the respective hexaploid wheat parents or to another wheat line. The BC_1F_1 plants (hexaploid) were self-pollinated and advanced by a single seed descent to the BC_1F_3 generation. Seeds from individual BC_1F_3 plants were bulked and grown in single rows in the field at the Kansas State University, Ashland Research Farm near Manhattan, KS, USA in 2016–2017 growing season. A total of 351 lines were selected from the entire population based on the ability to produce sufficient seeds to allow for yield testing, general fitness, threshability, and phenology corresponding to the adapted wheat cultivars.

Field Phenotyping

The population was phenotyped in 2018 and 2019 at Colby (Manhattan, KS, USA), and in 2020 at (Manhattan, KS, USA). In both locations and years, an augmented design was used to establish the trials. Plots were planted using a New Holland six-row wheat drill. Plot dimensions were 2.5 m long by 0.5 m wide and consisted of three rows with 18 cm row spacing. Starter fertilizer was applied with the seed at planting using granular 18-46-0 diammonium phosphate (DAP) at a rate of 168.1 kg/ha. Additional nitrogen was applied as a topdress in the spring using liquid 28-0-0 urea ammonium nitrate (UAN) at a rate of 67.3 kg/ha. A lateral irrigation system was used at Colby to ensure uniform germination and emergence as well as to provide additional water throughout the growing season in the irrigated treatment. Three hexaploid winter wheat lines well-adapted to Kansas environments (checks) and the hexaploid wheat parents were used as controls with at least three biological replications per block. In 2018 and 2019, two complete blocks were established and one block was irrigated (COI18 and COI19, respectively), whereas the other was rainfed/non-irrigated (CO18 and CO19, respectively), simulating optimal and farmer-field growth conditions. In 2020, only one block was grown under rainfed conditions (AS20).

The population was phenotyped for yield and yield components traits, BM traits, and Tg. Agrobase software (Mulitze, 1990) was used to adjust the grain yield (GY) (bushels per acre (BPA)], for spatial variability. The MARVIN seed imaging system [GTA Sensorik GmbH, Neubrandenburg, Germany) was used to assess the grain morphometric traits such as GN per sample, TGW, grain area (GA), GW, and grain length (GL) from the two technical replicates in 2018, and one measurement in 2019 and 2020. In 2019 and 2020, data were collected for the number of spikes per square foot (SPSF) from two random points within a plot. The 1 × 1 ft square frame was dropped over two rows at least one foot away from the plot edges to avoid the border effect. In 2019, only one row within a frame was cut above the ground level for BM determination, whereas in 2020, both rows were sampled. Biomass samples were collected in paper bags and dried for at least 3 weeks at 32°C (90°F) before processing. We collected data on aboveground dry BM measured as the total weight of the dry sample without the bag, the number of spikes per sample (SPB), the average spikelet number per spike (SNS) from 10 random spikes, and grain weight after threshing [grain sample weight (GSW)]. During threshing, we scored samples for the presence and absence of the Tg trait depending on the threshability. Harvest index (HI) was calculated as the percentage of GSW relative to BM.

In 2020, data for HD were collected from each plot when approximately 50% of the spikes had emerged from the flag leaves. The number of days to heading were calculated as the difference between the heading and planting dates. After all the plots had completed heading, PH, in centimeters was measured on the same day from two random but representative main tillers per plot for the whole field. PH was measured as the distance from the ground surface to the first spikelet of the spike.

Genotyping
Whole-Genome Shotgun Sequencing of Parental Lines

Genomic DNA of 21 *A. tauschii* accessions and 6 hexaploid parents used to create the introgression population was extracted from the leaf tissues of 2-week-old seedlings grown in a greenhouse using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following the protocol of the manufacturer. The quality and concentration of the DNA were assessed by using the PicoGreen dsDNA Assay Kit (Life Technologies, Carlsbad, CA, USA).

Genomic libraries for Illumina sequencing were constructed from ~2 µg of genomic DNA using the PCR-free Illumina protocol at the K-State Integrated Genomics Facility (IGF). The libraries were subjected to size selection using the Pippin Prep system (Sage Scientific, Beverly, MA, USA) to enrich for 400–600 bp fragments. The pooled barcoded libraries were sequenced using the NovaSeq instrument (2 × 150 bp run, S4 flow cell) at Kansas University Medical Center and NextSeq 500 (2 × 150 bp run) at IGF. The PCR-free whole-genome shotgun sequencing libraries generated from 27 parental lines ranged between 200 and 700 bp with an average of 450 bp (**Supplementary Figure 1**). Approximately 14 billion paired-end reads (150 bp long) were generated from the libraries with an average of 0.54 billion reads per sample (data are available at NCBI SRA database BioProject ID: PRJNA745927). The average number of reads per wheat line corresponded to ~10× genome coverage of parental lines. The raw reads with the Phred quality score <15, minimum length <50 bp, and adaptor sequences were filtered out using Trimmomatic v0.38-Java-1.8. The remaining filtered paired-end reads were mapped to the Chinese Spring (CS) RefSeq v1.0

(International Wheat Genome Sequencing Consortium, 2018) using the BWA-mem software v0.7.17. A total of 7.1 billion reads were aligned uniquely to the CS RefSeq v1.0.

The sam files generated by BWA-mem were converted to bam files using the SAMtools v1.10. The Picard Toolkit (http://broadinstitute.github.io/picard) was used to merge bam files from different lanes and sequencers into one bam file per sample. Reads that were aligned to multiple locations within the genome were identified and removed by SAMtools v1.10. The Picard Toolkit was used to prepare the merged unique aligned read bam files for GATK (McKenna et al., 2010) analysis. The preparatory steps included sorting, adding read groups, marking, and removing duplicate reads. The output deduplicated bam files were realigned around INDELs using GATK v3.7 and recalibrated with 90K SNPs (Wang et al., 2014) mapped to CS RefSeq v1.0 as the reference coordinates. The bam files were split into chromosome parts and indexed to reduce the memory and time required to process the files. GATK v3.7 was used to generate the genome variant call format (gvcf) files for each chromosome part. The gvcf files were split into 100 Mb chromosome windows and stored as genomicsDB using the genomicsDBImport tool in GATK v4.0. Joint genotyping of variants from each database was done using GATK v4.0 HaplotypeCaller. The flag <-allow-old-rms-mapping-quality-annotation-data> was set to enable the processing of gvcf files generated by GATK v3.7. All vcf files corresponding to the A, B, and D genomes were concatenated with concat, a Perl-based utility in vcftools v0.1.13. A custom Perl script was used to convert the CS RefSeq v1.0 part coordinates in the concatenated vcf to full coordinates after which vcf-filter tools v0.1.13 were used to remove INDELs, multi-allelic loci, sites with missing data and minor allele frequency (MAF) below 0.05. The filtered SNPs were phased using the Beagle software v4.1 (Browning and Browning, 2013).

The GATK v4.0 HaplotypeCaller identified ~99 million variants including SNPs and INDELs from reads uniquely aligned to the D genome of CS RefSeq v1.0. After excluding INDELs, multi-allelic loci, sites with missing data, and MAF <0.05, 20 million SNPs were retained. These were used to impute missing and ungenotyped SNPs in the D genome of the introgression population.

Genotype Imputation

Sequence-based genotyping of the introgression population was performed previously by Nyine et al. (2020). SNPs with MAF <0.01 were excluded from the raw vcf file using vcf-filter tools v0.1.13. The program conform-gt (https://faculty.washington.edu/ browning/conform-gt.html) was used to check the concordance of D genome SNP positions between the introgression population and the SNPs from the parental lines genotyped by whole-genome shotgun sequencing. Missing and ungenotyped SNPs in the D genome of the introgression population were imputed from the parental lines using Beagle v.5.0. A custom Perl script was used to filter out the imputed SNPs with genotype probability below 0.7 and MAF <0.05, which resulted in 5.2 million SNPs. The filtered SNPs were used in the downstream analyses such as GWAS mapping and identification of the introgressed haplotype blocks.

Introgressed Haplotype Detection

Genetic divergence between the parental lines affects the probability of an accurate detection of the introgressed segments from wild relatives. We used two introgression families, one created by crossing hexaploid wheat with *A. tauschii* ssp. *strangulata* (KanMark x TA1642, aka FAM93) and another one created by crossing hexaploid wheat with *A. tauschii* ssp. *tauschii* (Danby × TA2388, aka FAM97) to identify the introgressed haplotype blocks. FAM93 had 21 introgression lines (ILs), whereas FAM97 had 23 ILs. The R package HaploBlocker (Pook et al., 2019) was used to infer haplotype blocks from 5.2 million SNP sites. Recombinant inbred lines from each family were analyzed together with 21 *A. tauschii* and 6 hexaploid parents. The HaploBlocker parameters used in block calculation were node_min = 2 (default is 5), overlap_remove = TRUE, bp_map, and equal_remove=TRUE. The parameter node_min was used to control the number of haplotypes per node during a cluster-merging step of the block calculation function of HaploBlocker. The reduction in node_min was necessary to account for the low number of haplotype variants within these families. To maintain the SNP position in the haplotype block library, a vector of SNPs was provided *via* the parameter bp_map and prior to block calculation, SNPs in perfect linkage disequilibrium were removed by setting the parameter equal_remove = TRUE. Overlapping haplotypes were removed by setting parameter overlap_remove = TRUE. Custom R and Perl scripts were used to calculate the haplotype block length using the information from haplotype block start and end coordinates. All monomorphic haplotypes between the two parental lines were excluded from a haplotype matrix before calculating the frequency of the introgressed haplotypes per chromosome.

Phenotype Data Analysis
Trait Stability and Heritability

Yield stability is an important trait, which reflects the productivity of the crop under various growth conditions. No single statistic, however, is accurate enough to predict it due to a high variability of genotype and genotype by environment interaction effects. In this study, we used 16 different statistics, including parametric [such as mean variance component (θi), GE variance component [$\theta_{(i)}$], Wricke's ecovalence (W_i^2), regression coefficient ($*b_i$), deviation from regression (S_{di}^2), Shukla's stability variance (σ_i^2), coefficient of variance (CVi), and Kang's rank-sum (*Kang* or *KR*)] and non-parametric [such as Huhn's and Nassar and Huhn's ($S^{(1)}$, $S^{(2)}$, $S^{(3)}$, $S^{(4)}$, $S^{(5)}$, and $S^{(6)}$), and the methods of Thennarasu ($NP^{(1)}$, $NP^{(2)}$, $NP^{(3)}$, and $NP^{(4)}$)] to rank the ILs for yield stability based on their performance across years, locations, and treatments. The description and properties of the statistics are documented at the website: https://manzik.com/stabilitysoft/. The analysis was implemented in R using a script from Pour-Aboughadareh et al. (2019), which is available at: https://github.com/pour-aboughadareh/stabilitysoft. The most stable and/or high yielding lines were identified by sorting them according to their rankings.

Broad sense heritability (H^2) and best linear unbiased predictions (BLUPs) for yield and the component traits were calculated from the 2018 and 2019 data. A mixed linear model with restricted maximum likelihood implemented in R package

lme4 was used to generate the variance components (var) that were used to calculate H^2 as follows:

$$\text{Model} = \text{lmer}(\text{Trait} \sim \text{Trt} + (1|\text{Genotype}) + (1|\text{Genotype} : \text{Trt}), \text{data} = \text{trait\_data})$$

$H^2 = \text{var}(\text{Genotype})/[\text{var}(\text{Genotype}) + \text{var}(\text{Genotype:Trt})/\text{No. of Trt} + \text{var}(\text{Residual})/\text{No. of Trt})]$ where Trt refers to irrigated and non-irrigated treatments. BLUPs were extracted from the linear mixed model as the random effects of genotypes.

Multiple comparisons for the effect of treatment on yield and yield components traits in the introgression population relative to the controls were performed using least squares (LS) means with "Tukey" adjustment method and $\alpha = 0.05$. To further assess the impact of introgressed haplotypes on the traits, lines were sorted in a descending order for each treatment and location/year. The percentage of the ILs that performed better than the best parental lines and checks [percentage of top-performing lines (PTPL)] was calculated for each trait. Similarly, ILs that produced more grains than both parents and checks were identified based on the mean spatial adjusted yield. The percentage increase in yield was calculated as follows:

$$\Delta \bar{y}_{ij} = 100 \times \frac{\bar{y}_i - \bar{y}_j}{\bar{y}_j}$$

where $\Delta \bar{y}_{ij}$ is the percentage change in mean yield, $\bar{y}_i$ is the mean yield for the high yielding ILs, and $\bar{y}_j$ is the mean yield for the controls (parents and checks).

Trade-Off Between Traits

The relationship between yield, yield components, and BM traits were assessed by calculating the Pearson's correlation coefficients. We compared the trend of correlations from different treatments (irrigated/non-irrigated) and years to determine the extent of trade-off between traits within the introgression population and how the environment influenced them.

GWAS Mapping

The genome-wide association analysis was used to identify the genomic regions with SNPs and haplotypes that have a significant association with the traits. We tested the association of 5.2 million SNPs from the D genome with the traits phenotyped in different treatments and years. A total of 15,967 haplotype block windows were identified from 5.2 million SNPs using the R package HaploBlocker v1.5.2 (Pook et al., 2019). Default parameters for HaploBlocker were used except node_min, which were reduced to two (default is five) as most genomic intervals in our data set had <5 haplotype variants. Overlapping haplotypes were removed using the parameter overlap_remove = TRUE, and the SNP coordinates were included in the haplotype library *via* the parameter bp_map. Haplotype blocks were split into windows by setting the parameter return_dataset = TRUE in the block_windowdataset() function. The haplotype variants within a given chromosome interval were recorded as 0, 1, 2, or 3 depending on the total number of haplotype variants present within the interval. In both cases, a mixed linear

model implemented in the R package GAPIT was used to detect the associations. To control for a population structure in SNP-based analyses, 100,000 randomly selected markers were used to calculate the principal components. In haplotype-block-based GWAS, all haplotype blocks were used to calculate the principal components. In both cases, only the first three principal components were used to control the population structure. Two threshold options were used to identify significant associations including a more stringent Bonferroni correction and a less stringent Benjamini and Hochberg (1995) false discovery rate (FDR) at 5% significance level. To control for the effect of treatment and year, GWAS based on BLUPs was also performed, and significant associations were reported only when there were SNP- or haplotype-trait association in at least two independent trials or in the BLUP-based analysis.

SNP- and Haplotype-Trait Effects

Haplotype variation at loci with significant SNPs and their effects on traits in the introgression population were analyzed. The R package HaploBlocker v1.5.2 (Pook et al., 2019) was used to infer haplotypes at the genomic loci with significant SNP-trait associations. Heatmap.2 function provided in the R package gplots was used to visualize the variation in haplotypes from hexaploid wheat and *A. tauschii*. However, at the *Ppd-D1* locus, a visual comparison of the SNPs from the parental lines was done, and the SNPs were annotated using the snpEff v4.3 software to resolve haplotype variants in the *A. tauschii* lines that could not be clearly distinguished by HaploBlocker. SNPs significant at FDR ≤ 0.05 and their estimated allelic effects were selected from the association mapping results and used to verify if the haplotype effect corresponded to the observed phenotype in the introgression population. The mean and the SD of the phenotype were calculated for each group of lines carrying a similar haplotype, and the difference between the means was compared using Tukey's honestly significant difference and the Student's *t*-test.

RESULTS

Trait Variation in the Introgression Population

Broad sense H^2 of GY was 0.7 in 2018 and 0.64 in 2019, whereas for the yield component traits such as TGW, GA, GW, and GL, H^2 values were 0.85, 0.89, 0.83, and 0.95, respectively. The agronomic performance of the ILs relative to the controls (parent/checks) was assessed by comparing their yield and yield component traits under different treatments. The effect of treatment on yield was significant in 2018 ($p < 2.2e-16$), but not in 2019 ($p = 0.24$) at 95% confidence level. The latter is partially associated with more abundant rainfall in 2019 that reduced the difference in the water availability stress levels between the irrigated and non-irrigated field trials in Colby, KS, USA. Based on the LS means, significant differences in yield between controls and ILs was observed in 2019 and 2020, but not in 2018 (**Table 1**). The yield data collected from irrigated and rainfed (non-irrigated) field trials conducted between 2018 and 2020 revealed that up to 23% of the lines with introgressions produce

TABLE 1 | Comparison of grain yield (GY) between introgression lines (ILs; progeny) and controls (checks/hexaploid parents) per treatment within a year using least squares (LS) means.

Year	Treatment	Group	LSmean	SE	df	Lower.CL	Upper.CL	Group
2018	Irrigated	Progeny	68.6	0.77	410	66.8	70.5	a
2018	Irrigated	Check	72.1	7.64	410	53.8	90.4	a
2018	Irrigated	Parent	75.7	3.6	410	67.1	84.3	a
2018	Non-Irrigated	Progeny	50.1	6.43	395	34.6	65.5	a
2018	Non-Irrigated	Check	51.6	0.66	395	50	53.2	a
2018	Non-Irrigated	Parent	57.1	2.95	395	50	64.2	a
2019	Irrigated	Progeny	103	0.82	379	100.6	105	a
2019	Irrigated	Parent	119	3.8	379	109.5	128	b
2019	Irrigated	Check	121	9.05	379	99.3	143	ab
2019	Non-Irrigated	Progeny	101	0.86	382	99.2	103	a
2019	Non-Irrigated	Parent	116	3.85	382	106.3	125	b
2019	Non-Irrigated	Check	116	9.44	382	93.5	139	ab
2020	Non-Irrigated	Progeny	47.1	0.43	385	46	48.1	a
2020	Non-Irrigated	Parent	55.3	2.01	385	50.5	60.1	b
2020	Non-Irrigated	Check	59.3	4.79	385	47.8	70.8	b

The 2018 and 2019 trials were conducted at Colby, whereas the 2020 trial was conducted at Ashland, Manhattan, KS, USA. Groups with similar letters are not significantly different at 95% confidence level.

more grain than the controls (**Supplementary Figure 2**). In 2018, however, 3.2 and 48% of the ILs produced more grain than the parental lines and checks, respectively, in the non-irrigated trial (CO18), suggesting that the check cultivars were more sensitive to drought stress than the parental lines.

The proportion of ILs outperforming the checks and parental lines for the measured traits varied between treatments and years with a minimum of 0.8% for BM in the non-irrigated trial in 2019 (CO19) and a maximum of 73% for TGW in the 2018 irrigated trial (COI18). The percentage increase in yield for the ILs that outperformed both checks and parent varied from 11 to 57%, whereas the number of lines that produced more grains varied from 6 to 94 (**Table 2**). Under drought stress conditions in 2018 (CO18), the mean yield of top-performing ILs was 1.6- and 1.4-folds higher than the checks and parents, respectively (**Figure 1**). These results suggest that some ILs carry the alleles that confer drought tolerance, thus ensuring high productivity under stressful conditions. The highest yield potential of both ILs and controls was observed in 2019. The mean of the top yielding ILs reached 134 BPA, whereas that of the parental lines and checks reached 119 and 121 BPA, respectively (**Table 2**).

The yield stability analyses were performed to identify ILs that are both high yielding and stable under various environmental conditions. Ranking by mean yield showed that 6% of the lines carrying introgression produced more grains than most parental lines, except Larry (**Supplementary File 1**). The yield of these lines ranged from 84.4 to 92.7 BPA. The average rank (AR) of 16 stability statistics revealed 12 lines with introgressions showing high yield stability. Five of these lines were both stable and high yielding according to KR when compared to the controls. The yield of these five ILs (KS15SGDCB110-11, KS15SGDCB098-1, KS15SGDCB103-6, KS15SGDCB098-14, and KS15SGDCB111-1) varied between 82 and 93 BPA. The yield of the most stable and high yielding IL (92.7 BPA) was 9.8% higher than the average

TABLE 2 | Percentage of mean yield difference between top-performing ILs and controls (parents and checks) per treatment in each year and location.

Group ID	No. of IL	Mean yield IL	Mean yield controls	% yield diff.
AS20_ILP	6	68.6	55.2	24.3
AS20_ILC	6	68.6	59.3	15.8
CO18_ILP	21	78.7	57.1	37.8
CO18_ILC	21	78.7	50.1	57.1
COI18_ILP	94	86.9	75.7	14.8
COI18_ILC	94	86.9	72.1	20.6
CO19_ILP	11	130.5	115.5	13.0
CO19_ILC	11	130.5	116.1	12.4
COI19_ILP	6	134.2	119.2	12.6
COI19_ILC	6	134.2	121.0	10.9

AS, Ashland non-irrigated trial; CO, Colby non-irrigated trial; COI, Colby irrigated trial; ILP, introgression lines by parents' GY comparison; ILC, introgression lines by checks GY comparison.

yield of the controls (84.4 BPA). These results indicate that novel alleles from *A. tauschii* have the potential to increase the adaptive potential of hard red winter wheat to different environmental conditions. In addition, the stability statistics could help to prioritize ILs for deployment in different agroecological zones depending on their ranking in stability and yield. Lines that are moderately high yielding but show good yield stability could be deployed in marginal environments, whereas less stable but high yielding lines could be deployed in less stressful environments to achieve high productivity.

Harvest index, a measure of source-sink capacity, was also assessed for stability in irrigated and rainfed trials. About 92 ILs showed a higher average HI (47.4–52.8) than the best parental line KS061406LN-26 (47.3). The AR based on the 16 stability
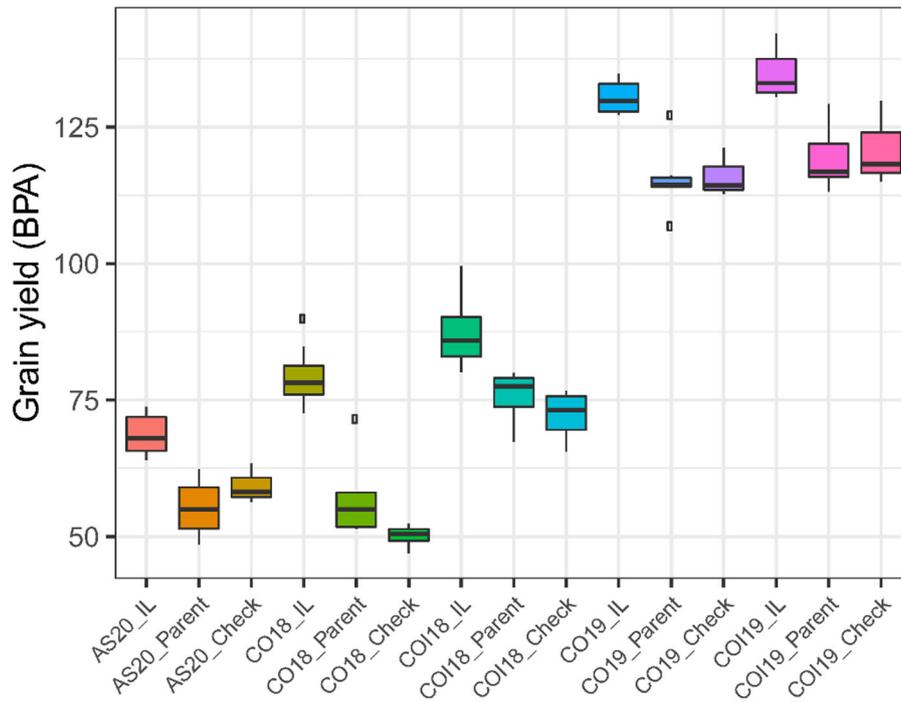
FIGURE 1 | Boxplots comparing the mean grain yield (GY) between the top-performing introgression lines (ILs) and the controls (parents and checks) across treatments, years, and locations. AS20 refers to Ashland rainfed trial in 2020, CO18 is Colby rainfed trial in 2018, COI18 is Colby irrigated trial in 2018, CO19 is Colby rainfed trial in 2019, and COI19 is Colby irrigated trial in 2019.

statistics placed 11 out of 92 lines in the top 20 most stable lines (**Supplementary File 2**). Line KS15SGDCB111-1, which is high yielding and stable, also ranked in the top five lines with a stable and high average HI.
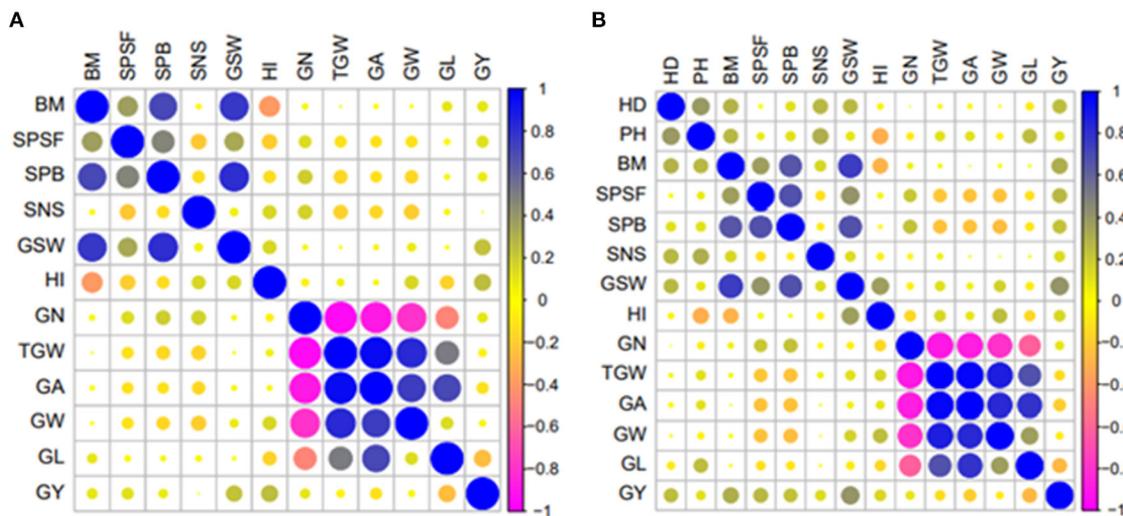
Trade-Off Between Yield and Yield Component Traits

Pearson's correlation coefficients between the average yield and yield stability based on AR of the 16 stability statistics was −0.44 ($p < 0.001$; **Supplementary File 1**). However, the correlation between yield and KR was −0.71 ($p < 0.001$), indicating that the most stable ILs were not necessarily the highest grain yielders although there were some exceptions. Similarly, the correlation between average HI and AR was −0.42 ($p < 2.2e$-16), whereas the correlation between HI and KR was −0.73 ($p < 2.2e$-16; **Supplementary File 2**).

The trade-off between yield and yield components was influenced by the treatment, year, and location as evidenced by a variation in the levels of correlations (**Supplementary File 3**). Higher positive correlations were observed among grain morphometric traits such as TGW, GA, GW, and GL, ranging from 0.13 (between GW and GL) to 0.96 (between TGW and GA) (**Figure 2**). HI and GSW positively correlated with GY, whereas the correlation between GY and GN was positive but nonsignificant in all trials, except for the Colby irrigated trial in 2019 (COI19) (**Supplementary File 3**). BM correlated negatively with HI but showed a positive correlation with GSW

(**Supplementary File 3**). In some cases, an increase in the SNS resulted in a reduction in the TGW, GA, GW, or GL, which was consistent with the previously observed trade-off between these traits (Kuzay et al., 2019). In contrast, HD positively correlated with the SNS and PH, which is in agreement with the previous findings (Shaw et al., 2013; Muqaddasi et al., 2019).

To further understand the contribution of different yield components to the final yield, we compared the phenotypes of the top yielding ILs to those of the controls across all treatments (**Supplementary File 4**). In the CO18 trial under non-irrigated conditions, the ILs that outperformed the controls in yield had the highest TGW, GA, and GL, whereas under irrigated conditions (COI18), all yield component traits showed the highest levels of expression in the top yielding ILs. The top yielding ILs in the CO19 trial had the highest HI, GW, SNS, and BM while TGW and GA were comparable to those of the parental lines. In the COI19 trial, the TGW, GA, and GW traits contributed more toward the final yield compared to the GL, HI, and BM traits. In the AS20 trial, high levels of heterogeneity were observed among the top yielding lines for the TGW, GA, GW, and GL traits. However, these lines showed a higher BM than the controls, resulting in a reduced HI.

Previously, it was suggested that the introgression from wild relatives might have a negative impact on agronomic traits due to a negative epistasis between the alleles of wild and cultivated wheat (Nyine et al., 2020). We investigated the relationship between the total size of the introgressed

FIGURE 2 | Pearson's correlation coefficients between yield and yield component traits at Colby in 2018 rainfed trial **(A)** and Ashland in 2020 rainfed trial **(B)**. Where HD, heading date; PH, plant height; BM, aboveground dry biomass; SPSF, spikes per square foot; SPB, spikes per bag; SNS, spikelet number per spike; GSW, grain sample weight; HI, harvest index; GN, grain number; TGW, 1,000 grain weight; GA, grain area; GW, grain width; GL, grain length; GY, grain yield.

genomic segments and phenotype. We found a positive linear relationship between GA, GL, SNS, and the total size of the introgressed segments (**Figure 3**, **Supplementary File 5**). For the TGW, however, a positive linear relationship was only observed under drought stress conditions indicating that some wheat lines with large introgressions are efficient in utilizing the limited soil moisture and nutrients during grain filling. There was a negative relationship between GY, HI, GW, TGW under irrigated conditions, and the size of introgression.

Genotyping and Imputation

To identify *A. tauschii* haplotypes in the D genome of introgression population, we generated high-density SNP data. By the whole-genome sequencing of 6 hexaploid parental lines and 21 *A. tauschii* accessions used for generating octoploid parents, we identified about 20 million high-quality SNP variants (MAF $\geq$ 0.05) and used them for genotype imputation in the introgression population genotyped by complexity-reduced sequencing. The total number of D genome SNPs retained after filtering out SNPs with genotype probability below 0.7 and MAF < 0.05 was 5.2 million.
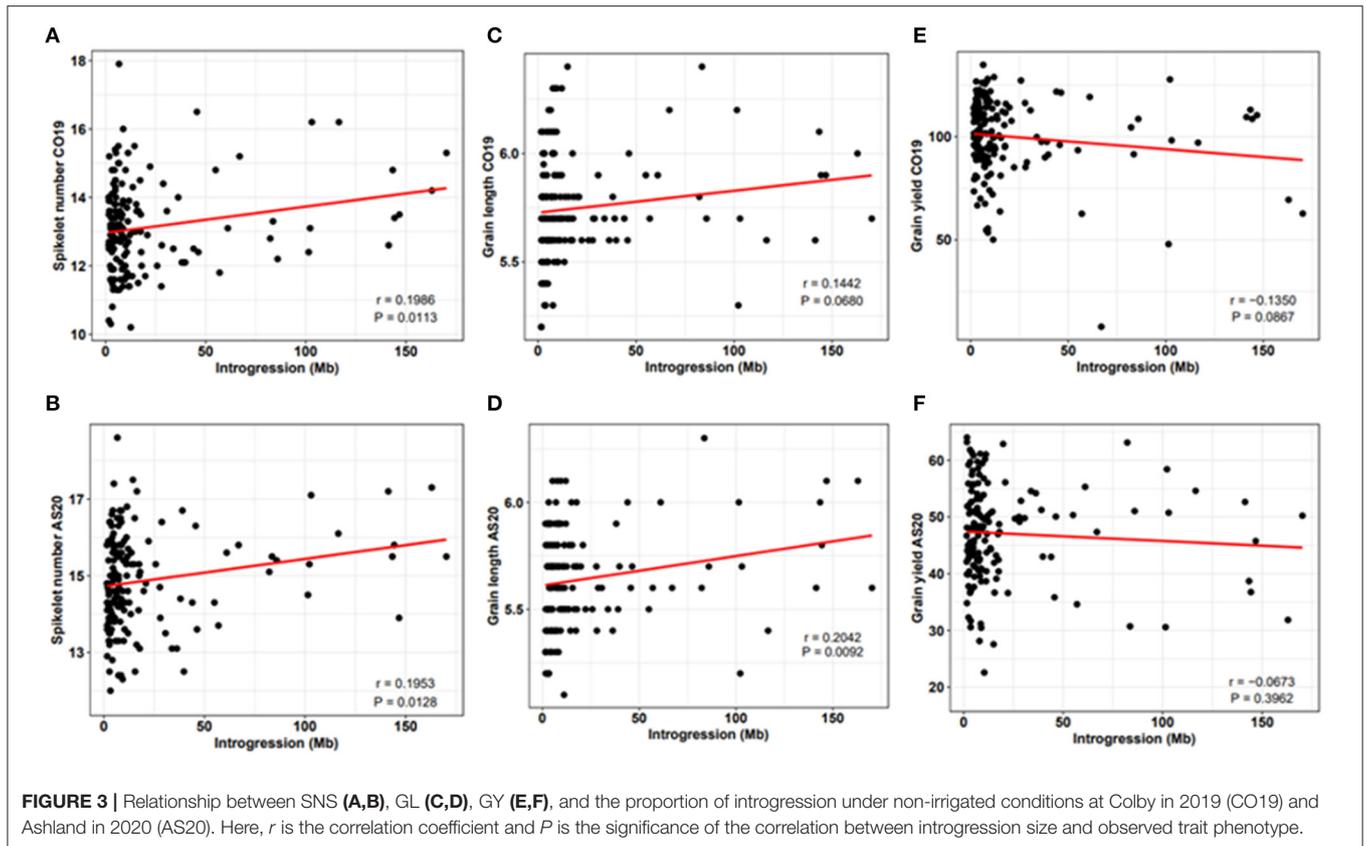
Haplotypic Variation Between ssp. *strangulata* and ssp. *tauschii* Families

Using HaploBlocker v1.5.2, we identified 4,764 and 6,429 non-overlapping haplotype blocks in the *A. tauschii* ssp. *strangulata* (FAM93) and *A. tauschii* ssp. *tauschii* (FAM97) families, respectively. After filtering out the monomorphic haplotypes between the parental lines, 869 (18%) and 3,020 (47%) segregating haplotypes were retained in FAM93 and FAM97, respectively (**Table 3**, **Supplementary File 6**). The low proportion of segregating haplotypes between hexaploid wheat and ssp. *strangulata* D genomes is in agreement with the finding that *A. tauschii* ssp. *strangulata* was the donor of the D genome

of hexaploid wheat (Wang et al., 2013). These results also suggest that the high similarity between the genome of ssp. *strangulata* and the D genome of hexaploid wheat could result in the underestimation of the proportion of the introgressed haplotypes. The average genome-wide haplotype block length in FAM93 was higher (2 Mb) than that in the FAM97 (1 Mb) (**Supplementary File 6**). There was a significant difference in the introgressed haplotype length between the lines in FAM93 and FAM97 based on the *t*-test (p = 3.1e-16). The longest haplotype introgressed in all lines from FAM93 was 44 Mb on chromosome arm 3DL, whereas in FAM97 only four lines had a haplotype >32 Mb on chromosome arm 7DL. The number of segregating haplotypes in FAM93 varies from 32 (3D) to 336 (2D), whereas in FAM97 the number of segregating haplotypes varies from 173 (3D) to 617 (5D) (**Table 3**). In FAM93 and FAM97, the average frequency of each haplotype from *A. tauschii* parents in the ILs was 11 and 4, respectively (**Supplementary File 6**).

SNP- and Haplotype-Based GWAS Mapping

Genome-wide association study was performed in the *A. tauschii* introgression population to assess the effects of introgression into the D genome on the variance of traits related to BM, yield and yield components, and Tg. The marker-trait association analyses were based on both individual SNPs and haplotype blocks identified by HaploBlocker from the 5.2 million imputed variants. We report only those associations that are replicated in at least two independent field trials and show a significant association with both SNPs and haplotypes at FDR 0.05 (**Supplementary Table 1**). Several genomic loci with significant associations distributed on the D genome chromosomes were detected for GL, GW, and SNS. For other traits such as GY, TGW, GN, GA, HI, BM, GSW, and SPSF, no consistent associations replicated in independent trials were detected.

FIGURE 3 | Relationship between SNS **(A,B)**, GL **(C,D)**, GY **(E,F)**, and the proportion of introgression under non-irrigated conditions at Colby in 2019 (CO19) and Ashland in 2020 (AS20). Here, *r* is the correlation coefficient and *P* is the significance of the correlation between introgression size and observed trait phenotype.

TABLE 3 | Variation of introgressed haplotypes between *Aegilops tauschii* ssp. *strangulata* (FAM93) and *A. tauschii* ssp. *tauschii* (FAM97) families.
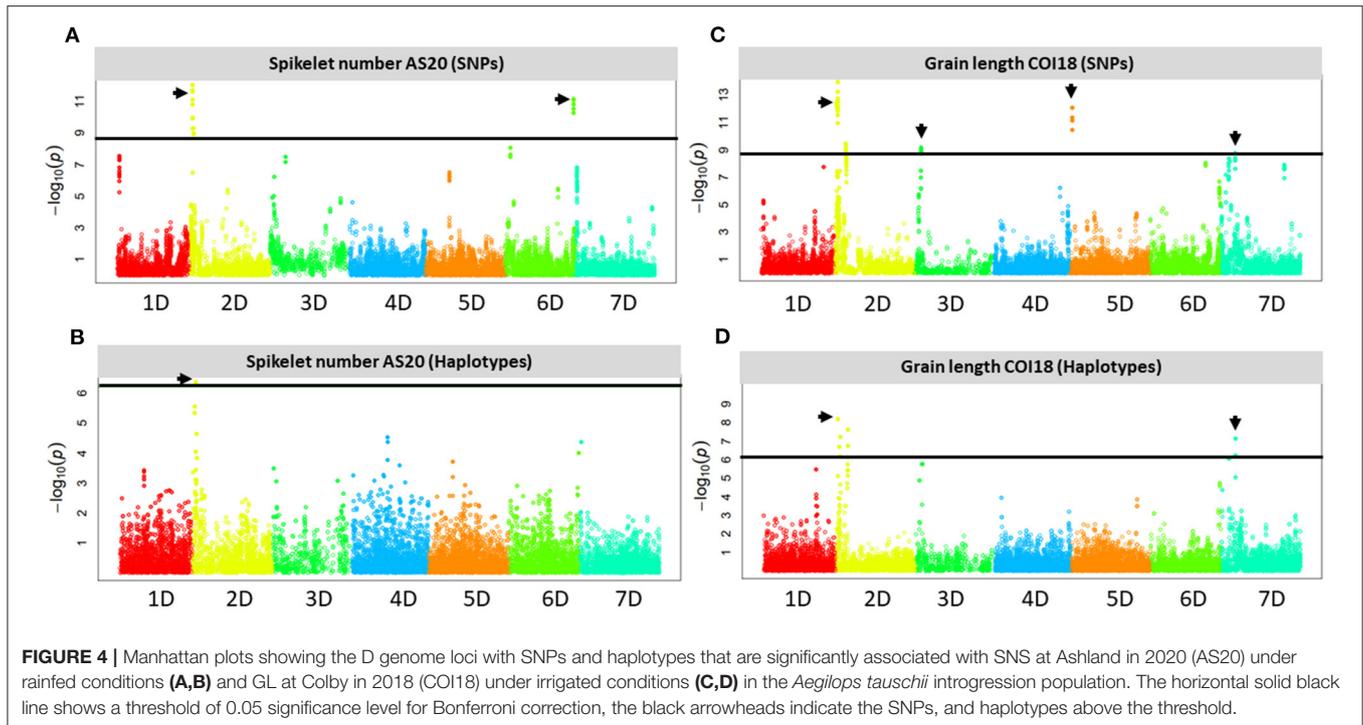
Family	Pedigree	Subspecies	Chrom	NH	MHL (bp)	MHF	Range (bp)
FAM93	KanMark × TA1642	*strangulata*	chr1D	105	1,896,815	4.6	1,788–8,639,976
FAM93	KanMark × TA1642	*strangulata*	chr2D	336	1,326,620	20.0	813–30,656,640
FAM93	KanMark × TA1642	*strangulata*	chr3D	32	6,653,571	4.9	2,525–44,264,371
FAM93	KanMark × TA1642	*strangulata*	chr4D	143	1,865,609	4.3	2,942–12,070,089
FAM93	KanMark × TA1642	*strangulata*	chr5D	138	2,302,767	7.1	556–15,103,840
FAM93	KanMark × TA1642	*strangulata*	chr6D	65	2,519,945	3.0	3,455–13,483,300
FAM93	KanMark × TA1642	*strangulata*	chr7D	50	2,667,416	2.04	1,627–11,349,447
FAM97	Danby × TA2388	*tauschii*	chr1D	495	829,609	3.1	762–9,169,049
FAM97	Danby × TA2388	*tauschii*	chr2D	352	1,305,439	1.8	813–15,991,790
FAM97	Danby × TA2388	*tauschii*	chr3D	173	2,968,356	3.0	884–27,225,468
FAM97	Danby × TA2388	*tauschii*	chr4D	540	750,352	0.8	1,982–20,010,593
FAM97	Danby × TA2388	*tauschii*	chr5D	617	770,483	10.1	1,243–15,396,133
FAM97	Danby × TA2388	*tauschii*	chr6D	373	830,168	3.8	764–23,208,199
FAM97	Danby × TA2388	*tauschii*	chr7D	465	1,000,899	3.7	1,397–32,547,936

NH, Number of haplotype blocks; MHL, Mean haplotype block length; MHF, Mean haplotype frequency.

We identified multiple significant SNP- and haplotype-trait associations from all trials on chromosome arms 2DS and 7DS for GL (**Figure 4**). The most significant SNPs were located in haplotype block windows 22,262,355–22,289,017, 30,582,113–30,595,115, and 80,864,297–81,398,316 bp on chromosome arm 2DS, and 11,024,311–11,374,767 bp on chromosome arm 7DS (**Supplementary Table 1**). Association analysis based on BLUPs confirmed the results obtained

from individual trials for GL on these two chromosomes. Other significant associations detected in at least two trials were identified on chromosome arms 3DS and 5DS (**Supplementary File 7**).

We identified haplotypes with significant SNPs associated with GW on 1DL, 2DS, 6DL, and 7DS from at least two independent trials that were confirmed by BLUP-based analysis (**Supplementary File 7**). Haplotype block windows

FIGURE 4 | Manhattan plots showing the D genome loci with SNPs and haplotypes that are significantly associated with SNS at Ashland in 2020 (AS20) under rainfed conditions **(A,B)** and GL at Colby in 2018 (COI18) under irrigated conditions **(C,D)** in the *Aegilops tauschii* introgression population. The horizontal solid black line shows a threshold of 0.05 significance level for Bonferroni correction, the black arrowheads indicate the SNPs, and haplotypes above the threshold.

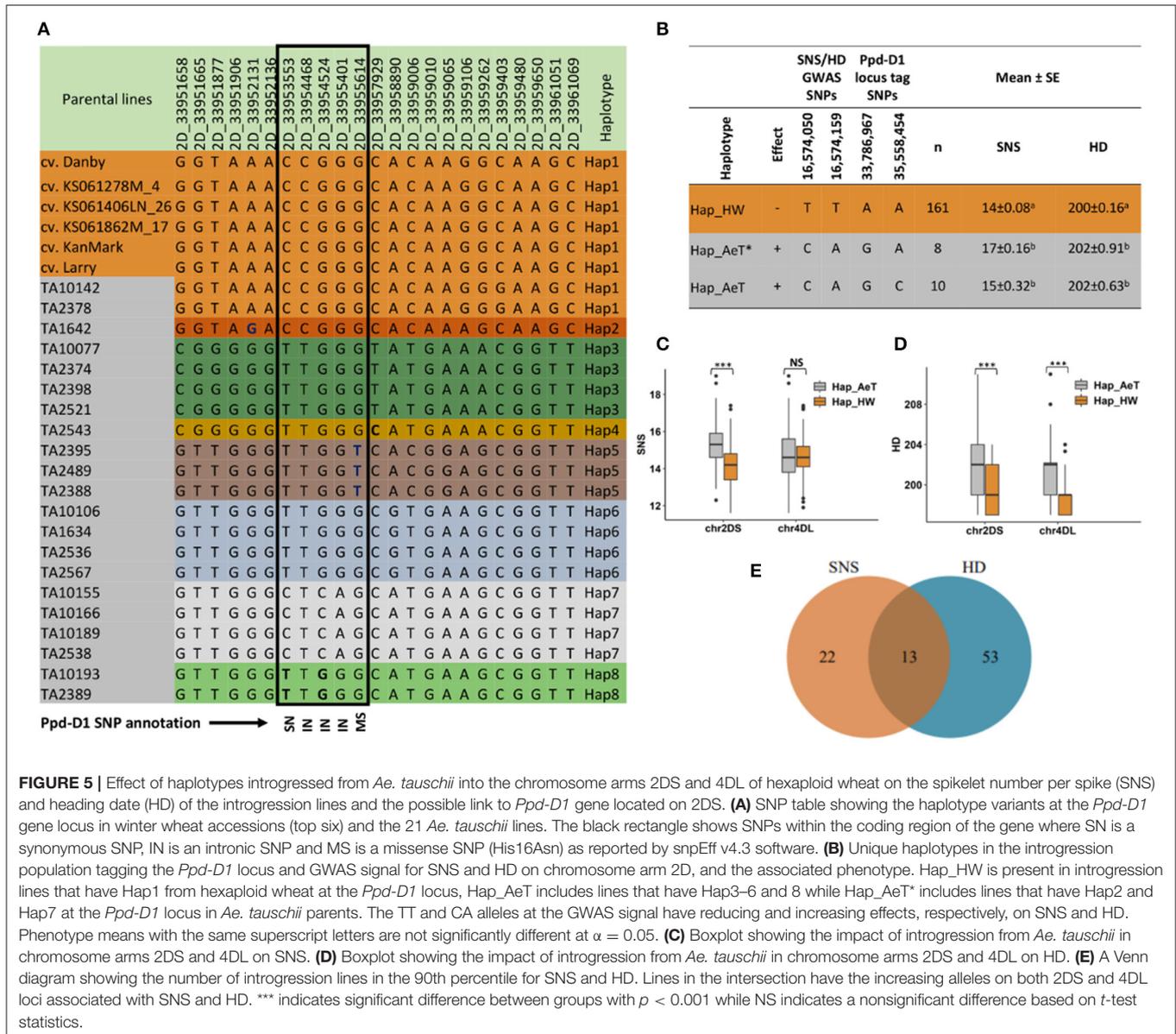65,964,778–66,124,103 and 66,265,325–66,266,089 bp showed the most significant association on 2DS.

At 95% confidence level, the most significant SNP-trait associations were identified on chromosome arms 2DS and 6DL for SNS from the three independent trials (COI19, CO19, and AS20), (**Figure 4**, **Supplementary File 7**). The most significant associations are located at 16.5 and 463.8 Mb on 2DS and 6DL, respectively. Haplotype-trait analysis confirmed the association on 2DS for SNS at the 16.5 Mb locus located within the haplotype block window 16,497,666–16,548,006 bp. At FDR < 0.05, there was no haplotype block window on 6DL locus that overlapped with a significant SNP-trait association.

Previous studies have shown that SNS is linked to HD (Shaw et al., 2013; Muqaddasi et al., 2019). In the current study, we detected significant associations with SNS on chromosome arms 2DS and 6DL. We had 1 year data for HD and PH collected from Ashland in 2020, which provided us a good opportunity to validate this link in the *A. tauschii* introgression population. GWAS mapping detected significant associations with HD on chromosome arms 2DS and 4DL, whereas all D genome chromosomes showed a significant association with PH but the strongest signals were observed on 1DS, 3DS, and 6DL. The haplotype block window 16,548,753–16,639,561 bp on 2DS with the most significant SNPs for HD overlapped the locus showing a significant association with SNS, which is in proximity to another haplotype block overlapping with the most significant SNPs for SNS (16,497,666–16,548,006 bp). These results suggest that the expression of these two traits could be co-regulated.

For HD, the haplotype block windows on chromosome arm 4DL 442,735,095–442,751,954 bp and 459,271,685–459,290,731

bp had the most significant SNP-trait associations. The three traits (SNS, HD, and PH) are known to be affected by the *Rht8* and *Ppd-D1* genes on 2DS, in addition to *Rht1* on 4D, which control PH and flowering time (Borojevic and Borojevic, 2005; Chen et al., 2018). Due to the lack of SNPs located near the *Ppd-D1* gene locus at ~34 Mb (33,961,438–33,951,651 bp interval in CS RefSeq v1.0), we could not directly validate its association with these traits. However, significant associations for SNS were detected at ~3 Mb next to the *Ppd-D1* locus in the CO19 and AS20 trials on haplotype blocks 2D: 30,192,335–30,264,745 bp and 2D: 28,829,778–28,937,705 bp, respectively. In the parental lines with high-density SNPs (~20 million), the *Ppd-D1* locus had SNPs, which allowed us to precisely map the haplotypes from *A. tauschii* and hexaploid wheat lines. The results obtained from HaploBlocker showed that all hexaploid parents carry an identical haplotype, which is distinct from that of *A. tauschii* accessions.

By using SNPs identified by the whole-genome sequencing of parental lines, we characterized a haplotypic diversity at the *Ppd-D1* locus (**Figure 5A**). All hexaploid wheat lines carried the same *Ppd-D1* haplotype (Hap1) while seven haplotypes of the *Ppd-D1* gene (Hap2–Hap8) were identified in *A. tauschii*. The whole-genome sequencing of 21 *A. tauschii* revealed a broader range of *Ppd-D1* diversity compared to a previous study (Guo et al., 2009), which identified only three *Ppd-D1* haplotypes. The *A. tauschii* ssp. *strangulata* accessions carried the haplotypes that were identical to hexaploid wheat, except for Hap2 in TA1642, which had one SNP at position 33,952,131 bp (**Figure 5A**). The *Ppd-D1* genic region in *A. tauschii* ssp. *tauschii* accessions has one synonymous (SN), three intronic (IN), and one missense (MS) SNPs. The MS variant at position 33,955,614 bp results in His16Asn change, which is predicted to have a moderate

FIGURE 5 | Effect of haplotypes introgressed from *Ae. tauschii* into the chromosome arms 2DS and 4DL of hexaploid wheat on the spikelet number per spike (SNS) and heading date (HD) of the introgression lines and the possible link to *Ppd-D1* gene located on 2DS. **(A)** SNP table showing the haplotype variants at the *Ppd-D1* gene locus in winter wheat accessions (top six) and the 21 *Ae. tauschii* lines. The black rectangle shows SNPs within the coding region of the gene where SN is a synonymous SNP, IN is an intronic SNP and MS is a missense SNP (His16Asn) as reported by snpEff v4.3 software. **(B)** Unique haplotypes in the introgression population tagging the *Ppd-D1* locus and GWAS signal for SNS and HD on chromosome arm 2D, and the associated phenotype. Hap_HW is present in introgression lines that have Hap1 from hexaploid wheat at the *Ppd-D1* locus, Hap_AeT includes lines that have Hap3–6 and 8 while Hap_AeT* includes lines that have Hap2 and Hap7 at the *Ppd-D1* locus in *Ae. tauschii* parents. The TT and CA alleles at the GWAS signal have reducing and increasing effects, respectively, on SNS and HD. Phenotype means with the same superscript letters are not significantly different at $\alpha = 0.05$. **(C)** Boxplot showing the impact of introgression from *Ae. tauschii* in chromosome arms 2DS and 4DL on SNS. **(D)** Boxplot showing the impact of introgression from *Ae. tauschii* in chromosome arms 2DS and 4DL on HD. **(E)** A Venn diagram showing the number of introgression lines in the 90th percentile for SNS and HD. Lines in the intersection have the increasing alleles on both 2DS and 4DL loci associated with SNS and HD. *** indicates significant difference between groups with $p < 0.001$ while NS indicates a nonsignificant difference based on *t*-test statistics.

functional impact, and only present in the lines with haplotype Hap5 (**Figure 5A**). Next, we inferred the parental haplotypes of the *Ppd-D1* locus in the introgression population by using SNPs within the ~1–2 Mb region surrounding the *Ppd-D1* locus. About 82% of the ILs carried haplotype Hap1.

Further, we evaluated the linkage of *Ppd-D1* haplotypes with SNP alleles showing a significant association with a variation in SNS and HD. For this purpose, we used two SNP sites, 2D_33786967 and 2D_35558454, which flank the *Ppd-D1* locus on both sides and have genotyping information in the introgression population. We compared them to SNP alleles that were significantly associated with SNS and HD in a haplotype block window 2D: 16,548,753–16,639,561 bp (2D_16574050 and 2D_16574159), spanning ~17 Mb region (**Figure 5B**). We found that the GWAS alleles associated with an increase in SNS and HD in the introgression population are also linked with two

A. tauschii haplotypes (Hap_AeT* and Hap_AeT), whereas the GWAS alleles associated with decreasing effects were in LD with Hap_HW contributed by the hexaploid wheat parents. The Hap_AeT* group of haplotypes was contributed by the *A. tauschii* parents having Hap2 and Hap7 at the *Ppd-D1* locus.

The Phenotypic Effects of Haplotype Block Variants
Average SNS and HD

Significant haplotype-trait associations were identified on chromosome arms 2DS and 4DL that influence SNS and HD. Chromosome 2DS had multiple introgressed haplotypes that are significantly associated with a variation in SNS and HD, with the most significant haplotypes located at 16,497,666–16,548,006 and 16,548,753–16,639,561 bp for SNS and HD, respectively. The haplotype variants with the increasing effect at these loci

TABLE 4 | ANOVA for the effect of *Ppd-D1* haplotype variants from hexaploid wheat and *A. tauschii* on the spikelet number per spike (SNS) in three experimental trials of the introgression population.

Haplotype	Estimate	Std. error	*t*-value	Pr(>|t|)
Hexaploid wheat (HW)	13.38	0.13	−3.41	0.000677***
A. tauschii (AeT)	13.83	0.12	115.84	<2e-16***
HW_Hap1	13.38	0.84	1.79	0.07317
AeT_Hap2	13.79	0.94	2.05	0.04092*
AeT_Hap3	13.74	0.91	2.06	0.03958*
AeT_Hap4	13.50	1.19	1.37	0.17002
AeT_Hap5	13.53	0.92	1.80	0.07190
AeT_Hap6	13.69	0.93	1.96	0.05069
AeT_Hap7	**14.62**	**0.92**	**2.99**	**0.00288****
AeT_Hap8	13.77	1.19	1.60	0.11051

*** = 0.001, ** = 0.01, and * = 0.05.
Bold values indicate the haplotype that shows the highest significant increase in the spikelet number.

were from *A. tauschii* parents, whereas those with a reducing effect were from the hexaploid wheat lines (**Figures 5C,D**). The verification of GWAS results for an allelic effect at 2DS locus associated with SNS and HD supports the abovementioned observation (**Supplementary File 8**). We observed a positive Pearson's correlation coefficient between SNS and HD; and lines having haplotypes from either parent showed significant differences in the phenotype based on a *t*-test ($r = 0.23$, $p = 3.31e-07$) at 95% confidence level. Haplotypes on 4DL had a smaller effect on SNS compared to HD. Among 35 and 66 ILs having SNS and HD trait values above the 90th percentile of trait distribution, respectively, 13 lines had the increasing alleles from *A. tauschii* at both 2DS and 4DL loci associated with SNS and HD traits (**Figure 5E**).

Initially, we did not detect a significant GWAS signal directly associated with SNPs within the *Ppd-D1* gene due to the lack of high-quality imputed SNPs in this region. Further analysis of parental haplotypes identified SNP variants linked with both the *Ppd-D1* haplotypes and haplotypes at 28 and 30 Mb region, showing a significant haplotype-trait association in two trials. These haplotypes were within ∼3 Mb from the *Ppd-D1* locus and likely overlap with *Ppd-D1*. We performed the ANOVA to determine the effect of different haplotype variants identified in the parental accessions on the SNS in the introgression population using data from the three experimental trials (COI19, CO19, and AS20) (**Table 4**). The results show that both hexaploid and *A. tauschii* haplotypes have a significant effect on SNS ($p < 0.001$; **Table 4**, **Figure 5B**). Among the *A. tauschii* haplotypes, Hap7 had the highest impact on SNS ($p = 0.003$) followed by Hap2 ($p = 0041$) and Hap3 ($p = 0.040$). In contrast, Hap5 with the His16Asn MS mutation had a negative effect on SNS and was not significantly different from Hap1 present in hexaploid wheat lines ($p = 0.072$). Consistent with previous studies, these results showed that the *Ppd-D1* gene located at ∼34 Mb (33,961,438–33,951,651 bp interval in CS RefSeq v.1.0), which plays a role in flowering time regulation in wheat, also has a strong effect on the variation in the SNS (Beales et al., 2007; Guo et al., 2009).

TABLE 5 | Chromosome 2D haplotypes variants associated with the spikelet number and HD and how they influence other traits in the introgression population.

Trait	Haplotype	Loc1 Freq.	chr2D:16548753–16639561	SD
HD	Hap_HW	53	199[b]	2.28*
HD	**Hap_AeT**	**156**	**202[a]**	**2.86[†]**
HD	Hap_HW/AeT	142	200[c]	2.29[‡]
PH	Hap_HW	53	68.27[ab]	6.79
PH	Hap_AeT	156	70.82[a]	6.97
PH	Hap_HW/AeT	142	65.9[b]	8.68
BM	Hap_HW	53	155.18[a]	28.21
BM	Hap_AeT	156	156.26[a]	31.01
BM	Hap_HW/AeT	142	157.52[a]	35.82
SNS	Hap_HW	53	14.72[b]	1.2
SNS	**Hap_AeT**	**156**	**15.27[a]**	**1.12**
SNS	Hap_HW/AeT	142	14.05[c]	1.03
GSW	Hap_HW	53	66.84[a]	14.71
GSW	Hap_AeT	156	65.82[a]	13.67
GSW	Hap_HW/AeT	142	66.47[a]	13.57
HI	Hap_HW	53	42.85[a]	4.35
HI	Hap_AeT	156	42.23[a]	4.24
HI	Hap_HW/AeT	142	42.89[a]	6.28

*Haplotype with a reducing effect, [†]Haplotype with an increasing effect, [‡]Haplotype with an intermediate effect, Loc1 = chr2D: 16,575,276–16,635,669 bp; Loc1 Freq. = Number of introgression lines having the haplotype; HD is a heading date, PH is an average plant height, BM is the aboveground total dry biomass, SNS is a spikelet number per spike, GSW is the grain sample weight from a biomass sample, and HI is the harvest index. Means with the same superscript letters are not statistically significant for each trait based on Tukey's honestly significant difference at 95% confidence level.
Bold values indicate the haplotype that shows the highest increase in both the SNS and HD.

Pleiotropic Effects of Haplotypes on Yield Component Traits

We also evaluated the effects of distinct haplotypes associated with HD on other traits. Haplotype Hap_AeT from chromosome 2D located at 16,548,753–16,639,561 bp is associated with a significant increase in the days to heading and SNS without any significant impact on BM, HI, and GSW (**Table 5**).

We compared the effects of two haplotypes associated with GW on 2DS (2D: 65,964,778–66,124,103 bp and 2D: 66,265,325–66,266,089 bp), where Hap_HW from hexaploid parents and Hap_AeT_st from *A. tauschii* ssp. *strangulata* increase GW, and Hap_AeT_ta from the *A. tauschii* ssp. *tauschii* parents reduces GW (**Table 6**). In Colby 2018 non-irrigated trial data, the haplotype at the 2D: 65,964,778–66,124,103 bp locus that was associated with an increase in GW and GL was also linked with an increase in GA and a decrease in GN. Meanwhile, both Hap_AeT_st and Hap_AeT_ta haplotypes at 2D: 65,964,778–66,124,103 bp were associated with an increase in GL, only Hap_AeT_ta was linked with a significant increase in GN (**Table 6**). The Hap_AeT_ta haplotype at the 2D: 66,265,325–66,266,089 bp haplotype block had similar effects on GN although the observed difference was not significant.

Similarly, haplotype block 6D: 463,775,852–463,809,722 bp associated with PH and SNS has three variants (**Supplementary File 7**). Haplotype variant Hap1_HW&AeT is associated with an increase in SNS, PH, BM, and GY. It showed

TABLE 6 | Chromosome 2D haplotypes variants associated with the grain width (GW) in Colby 2018 rainfed trial and their effects on other traits in the introgression population.

Trait	Haplotype	Haploblock chr2D:65964778–66124103			Haploblock chr2D:66265325–66266089		
		Freq.	Trait mean	Trait SD	Freq.	Trait mean	Trait SD
GN	Hap_AeT_st	5	475.50[b]	27.86	20	504.15[a]	55.28
GN	Hap_HW	140	499.04[b]	44.3	304	502.97[a]	41.6
GN	Hap_AeT_ta	202	510.94[a]	43.38	27	540.37[a]	47.94
TGW	Hap_AeT_st	5	32.74[a]	2.15	20	31.54[a]	2.74
TGW	Hap_HW	140	31.40[a]	2.2	304	31.30[a]	2
TGW	Hap_AeT_ta	202	31.01[a]	1.98	27	29.54[a]	1.76
GA	Hap_AeT_st	5	14.38[a]	1.18	20	13.89[a]	1.03
GA	Hap_HW	140	13.50[b]	0.69	304	13.53[a]	0.66
GA	Hap_AeT_ta	202	13.53[b]	0.66	27	13.28[a]	0.61
GW	Hap_AeT_st	5	3.14[a]	0.05	20	3.09[ab]	0.08Δ
GW	Hap_HW	140	3.13[a]	0.09	304	3.09[a]	0.09Δ
GW	Hap_AeT_ta	202	3.06[b]	0.08	27	3.00[b]	0.07#
GL	Hap_AeT_st	5	6.25[a]	0.51	20	6.11[a]	0.4
GL	Hap_HW	140	5.85[b]	0.22	304	5.93[b]	0.24
GL	Hap_AeT_ta	202	6.00[a]	0.24	27	6.01[a]	0.22
GY	Hap_AeT_st	5	59.19[a]	14.46	20	58.76[a]	14.77
GY	Hap_HW	140	60.51[a]	12.04	304	61.26[a]	12.03
GY	Hap_AeT_ta	202	61.28[a]	12.54	27	59.55[a]	13.41

Δhaplotype that increases GW, # haplotype that reduces GW. Means with the same superscript letters are not statistically significant for each trait based on Tukey's honestly significant difference at 95% confidence level.

no association with HD and grain traits, except GN where an intermediate effect was observed. This haplotype variant is present in two *A. tauschii* lines (TA2389 and TA2398) and the hexaploid parents excluding KanMark and KS061862M-17. Haplotype variant Hap0_HW&AeT, which is found in KanMark and KS061862M-17 and two ssp. *strangulata* accessions (TA1642 and TA2378), has an intermediate effect on GY, reducing it by two bushels compared to Hap1_HW&AeT. The third haplotype variant (Hap_AeT) is present only in the *A. tauschii* lines.

The haplotype Hap_AeT contributed by *A. tauschii* at haplotype block 7D: 14,185.651–14,596,748 bp, was associated with a significant increase in SNS compared to haplotypes present in winter wheat (**Supplementary Table 2**). This increase was associated with a significant decrease in GL and had no significant effect on GY. At 7D: 14,722,457–14,817,138 bp, the Hap_AeT haplotype contributed by *A. tauschii* was also linked with a significant increase in SNS compared to Hap1_HW&AeT detected in both hexaploid wheat and *A. tauschii* parents. However, an increase in SNS for this haplotype was connected with a decrease in both GL and GY. At this haplotype block (7D: 14,722,457–14,817,138 bp), the most significant increase in GY was observed for lines carrying haplotype Hap0_HW&AeT, which was associated with a moderate increase in both SNS and GL (**Supplementary Table 2**).

DISCUSSION

Here, we performed the sequence-based analysis of haplotypes in the wild-relative introgression population developed by crossing a diverse panel of *A. tauschii* accessions with winter wheat cultivars. Our results demonstrate that the whole-genome sequencing of wild and cultivated wheat founder

lines in combination with the complexity-reduced sequencing of a derived introgression population provides an effective framework for SNP imputation. Because most breeding populations are based on a limited number of founders, often including 10–30 lines, their whole-genome sequencing is feasible in crops even with large genomes such as wheat, and provides a comprehensive description of allelic diversity present in a breeding population. The latter makes sequenced founders an ideal reference panel for imputing genotypes in a breeding population genotyped using low-coverage or complexity-reduced sequencing. This was recently demonstrated by imputing genotypes in the wheat MAGIC population genotyped by low-coverage sequencing (Scott et al., 2021). The composition of our introgression population, including multiple biparental cross families (Nyine et al., 2020), also shifts the population allele frequency toward more common variants, which could be imputed with a higher accuracy than rare variants (Huang et al., 2015). In addition, the high levels of LD in the introgression population should increase the length of haplotype blocks and facilitate the detection of matching haplotypes in the reference panel of founders using even sparse genotyping data generated by low-coverage or complexity-reducing sequencing. Consistent with these assumptions, an imputation algorithm implemented in Beagle (Browning and Browning, 2013) allowed us to impute nearly 5.2 million SNPs in the introgression population with high genotype call probabilities above 0.7 using SNPs generated by complexity-reduced sequencing of this population and nearly 20 million variants identified in 27 founders. This high-density SNP marker data permitted a detailed characterization of the introgressed haplotypes (Pook et al., 2019) and assessing their effects on productivity traits.

Our results demonstrate that wild-relative introgressions into the D genome of wheat, the least diverse amongst the three subgenomes (Chao et al., 2010; Jordan et al., 2015; Singh et al., 2019), is associated with the increased levels of variation in yield and yield component traits. The analysis of data from several years and locations under irrigated and non-irrigated conditions revealed many superior ILs that produce more grains or show higher yield stability than the control cultivars. The yield increase in top-performing ILs was driven by a combination of yield component traits, and in many cases, it was associated with increased grain size, grain weight, and BM or improved harvest index. These results suggest that wild-relative introgression has the potential to positively affect source-sink balance, which was suggested to be one of the important factors contributing to yield potential (Reynolds et al., 2017). Many of these high yielding lines (∼23%) were also among the top lines showing the highest levels of yield stability, indicating that the introgression from *A. tauschii* likely improves the adaptive potential of hard red winter wheat in different environmental conditions. Consistent with this conclusion, the highest impact of introgression on yield was found in a non-irrigated trial, indicating that alleles from *A. tauschii* likely improve the adaption of hexaploid wheat to water-limiting conditions. The *A. tasuchii* accessions used to generate the introgression population represent both L1 and L2 lineages (Wang et al., 2013) and originate from a broad range of geographical locations with variable climatic conditions, likely capturing adaptive haplotypes from the regions prone to drought stress.

Heading date is one of the key agronomic traits linked with wheat adaptation to different geographical locations and improvement in yield (Jung and Müller, 2009). In our population, a positive correlation was observed between HD and the SNS, with some lines showing up to 2-week delay in HD. Several haplotype blocks on chromosome arms 2DS and 4DL were significantly associated with a variation in spikelet number and HD. The haplotypes with increasing effects at both loci were derived from *A. tauschii,* indicating their potential for modulating both traits in bread wheat. Chromosome 2DS is known to harbor the *Ppd-D1* and *Rht8* genes that control flowering time and PH, respectively, and also could affect the spikelet number (Shaw et al., 2013; Muqaddasi et al., 2019). The overlapping haplotype blocks associated with the spikelet number and HD were identified on 2DS, confirming that the two traits co-segregating in the population have a common genetic basis. We demonstrated that these 2DS haplotypes are associated with the different allelic variants of the *Ppd-D1* gene from *A. tauschii*. Consistent with the earlier studies, these results demonstrated that the different alleles of the *Ppd-D1* gene have distinct effects on HD and SNS (Beales et al., 2007, Guo et al., 2009). These effects were correlated with the relative expression levels of each *Ppd-D1* allele (Guo et al., 2009), suggesting that functional mutations within the *Ppd-D1* coding region and the modifier mutations in the regulatory region of the gene likely to account for a variation in these traits in the *A. tauschii* winter wheat introgression population. The developmental plasticity modulated by *Ppd-D1* is mediated by the changes in the expression of flowering time genes (Gol et al., 2021). It was

shown that the *Ppd-H1* from wild barley is capable of integrating environmental signals to control HD and minimize the negative impact of transient drought stress on spikelet number (Gol et al., 2021). Consistent with this observation, in the current study, ILs that have a high proportion of *A. tauschii* segments produced more grain under drought stress in the Colby 2018 trial, raising the possibility that the *A. tauschii* alleles of *Ppd-D1* also have the potential to protect wheat from the physiological effects of stress that lead to low yield.

Our study reveals that some haplotypes associated with the productivity trait variation in the introgression population also exhibit significant pleiotropic effects. Meanwhile, the direction of effects on various traits was largely consistent with the previously reported trade-offs among component traits (Griffiths et al., 2015; Reynolds et al., 2017; Quintero et al., 2018), the combined effects of some introgressed haplotypes were associated with the positive trends in yield. For example, a haplotype contributed by *A. tauschii* ssp. *tauschii* at the chromosome 2D haplotype block at 65,964,778–66,124,103 bp was associated with an increase in GL, size, and number with a moderate positive effect on GY. At the haplotype block on chromosome 7D located between 14,722,457–14,817,138 bp, the Hap0_HW&AeT haplotype shared between hexaploid wheat and *A. tauschii* parents and associated with a moderate increase in both SNS and GL was also associated with the most significant increase in GY. Analyses of the pleotropic effects of the introgressed haplotypes suggest that these haplotypes on chromosomes 2D and 7D could be utilized in breeding programs to improve yield component traits without negative effects on other productivity traits.

CONCLUSIONS

The imputation of markers from whole-genome-sequenced reference panels into skim-sequenced inference populations is increasingly becoming a common practice in plant breeding program due to its cost-effectiveness (Happ et al., 2019; Jensen et al., 2020). Our study demonstrates the utility of this strategy for detecting introgression in the wheat genome and contributes to developing genomic resources for deploying wild-relative diversity in wheat breeding programs. We show that the haplotype-based analysis of trait variation in this population has the potential to improve our knowledge on the genetic effects of the introgressed diversity on productivity traits and identify novel haplotypes for improving yield potential in wheat.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

AF and MC generated the *A. tauschii* introgression population. MN, EAd, MC, RA, BB, WW, DD, ZY, YG, FH, KJ, and AF

phenotyped the introgression population. MN, BB, and AA generated the genotyping data. AA performed next-generation sequencing of parental lines and introgression population. MN and FH performed the bioinformatical analyses of data. MN performed the statistical analysis. EAk conceived the idea and interpreted results. MN and EAk wrote the manuscript. All authors contributed to the manuscript revision, read, and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.716955/full#supplementary-material

REFERENCES

Bailey-Serres, J., Parker, J. E., Ainsworth, E. A., Oldroyd, G. E. D., and Schroeder, J. I. (2019). Genetic strategies for improving crop yields. *Nature*. 575, 109–118. doi: 10.1038/s41586-019-1679-0

Beales, J., Turner, A., Griffiths, S., Snape, J. W., and Laurie, D. A. (2007). A pseudo-response regulator is misexpressed in the photoperiod insensitive *Ppd-D1a* mutant of wheat (*Triticum aestivum* L.). *Theor. Appl. Genet*. 115, 721–733. doi: 10.1007/s00122-007-0603-4

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* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Borojevic, K., and Borojevic, K. (2005). The transfer and history of "reduced height genes" (Rht) in Wheat from Japan to Europe. *J. Hered*. 96, 455–459. doi: 10.1093/jhered/esi060

Browning, B. L., and Browning, S. R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* 194, 459–471. doi: 10.1534/genetics.113.150029

Chao, S., Dubcovsky, J., Dvorak, J., Luo, M.-C., Baenziger, S. P., Matnyazov, R., et al. (2010). Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). *BMC Genom*. 11:727. doi: 10.1186/1471-2164-11-727

Chen, L., Du, Y., Lu, Q., Chen, H., Meng, R., Cui, C., et al. (2018). The photoperiod-insensitive allele Ppd-D1a promotes earlier flowering in Rht12 dwarf plants of bread wheat. *Front. Plant Sci*. 9:1312. doi: 10.3389/fpls.2018.01312

Dale, Z., Jie, H., Luyu, H., Cancan, Z., Yun, Z., Yarui, S., et al. (2017). An advanced backcross population through synthetic octaploid wheat as a "bridge": development and QTL detection for seed dormancy. *Front. Plant Sci*. 8:2123. doi: 10.3389/fpls.2017.02123

Davies, R. W., Flint, J., Myers, S., and Mott, R. (2016). Rapid genotype imputation from sequence without reference panels. *Nat. Genet*. 48, 965–969. doi: 10.1038/ng.3594

Del Moral, L. G., Rharrabti, Y., Villegas, D., and Royo, C. (2003). Evaluation of grain yield and its components in durum wheat under Mediterranean conditions: an ontogenic approach. *Agron. J*. 95, 266–274. doi: 10.2134/agronj2003.0266

Du, Y., Cui, B., Wang, Z., Sun, J., and Niu, W. (2020). Effects of manure fertilizer on crop yield and soil properties in China: a meta-analysis. *Catena* 193:104617. doi: 10.1016/j.catena.2020.104617

Dvorák, J., Edge, M., and Ross, K. (1988). On the evolution of the adaptation of *Lophopyrum elongatum* to growth in saline environments. *PNAS* 85, 3805–3809. doi: 10.1073/pnas.85.11.3805

Gol, L., Haraldsson, E. B., and von Korff, M. (2021). Ppd-H1 integrates drought stress signals to control spike development and flowering time in barley. *J. Exp. Bot*. 72, 122–136. doi: 10.1093/jxb/eraa261

Golan, G., Ayalon, I., Perry, A., Zimran, G., Ade-Ajayi, T., Mosquna, A., et al. (2019). GNI - A1 mediates trade-off between grain number and grain weight in tetraploid wheat. *Theor. Appl. Genet*. 132, 2353–2365. doi: 10.1007/s00122-019-03358-5

Griffiths, S., Wingen, L., Pietragalla, J., Garcia, G., Hasan, A., Miralles, D., et al. (2015). Genetic dissection of grain size and grain number trade-offs in CIMMYT wheat germplasm. *PLoS ONE* 10:e0118847. doi: 10.1371/journal.pone.0118847

Guo, Z., Song, Y., Zhou, R., Ren, Z., and Jia, J. (2009). Discovery, evaluation and distribution of haplotypes of the wheat Ppd-D1 gene. *New Phytol*. 185, 841–851. doi: 10.1111/j.1469-8137.2009.03099.x

Hao, M., Zhang, L., Ning, S., Huang, L., Yuan, Z., Wu, B., et al. (2020). The resurgence of introgression breeding, as exemplified in wheat improvement. *Front. Plant Sci*. 11:252. doi: 10.3389/fpls.2020.00252

Happ, M. M., Wang, H., Graef, G. L., and Hyten, D. L. (2019). Generating High density, low cost genotype data in soybean [Glycine max (L.) Merr.]. *G3 Genes Genomes Genet*. 9, 2153–2160. doi: 10.1534/g3.119.400093

Hill, W. G., and Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genet. Res*. 8, 269–294. doi: 10.1017/S0016672300010156

Huang, J., Howie, B., McCarthy, S., Memari, Y., Walter, K., Min, J. L., et al. (2015). Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat. Commun*. 6:8111. doi: 10.1038/ncomms9111

International Wheat Genome Sequencing Consortium (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361, 1–16. doi: 10.1126/science.aar7191

Jensen, S. E., Charles, J. R., Muleta, K., Bradbury, P. J., Casstevens, T., Deshpande, S. P., et al. (2020). A sorghum practical haplotype graph facilitates genome-wide imputation and cost-effective genomic prediction. *Plant Genome* 13:e20009. doi: 10.1002/tpg2.20009

Jones, B. H., Blake, N. K., Heo, H.-Y., Kalous, J. R., Martin, J. M., Torrion, J. A., et al. (2020). Improving hexaploid spring wheat by introgression of alleles for yield component traits from durum wheat. *Crop Sci*. 60, 759–771. doi: 10.1002/csc2.20011

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. doi: 10.1186/s13059-015-0606-4

Jung, C., and Müller, A. E. (2009). Flowering time control and applications in plant breeding. *Trends Plant Sci*. 14, 563–573. doi: 10.1016/j.tplants.2009.07.005

Kuzay, S., Xu, Y., Zhang, J., Katz, A., Pearce, S., Su, Z., et al. (2019). Identification of a candidate gene for a QTL for spikelet number per spike on wheat chromosome arm 7AL by high-resolution genetic mapping. *Theor. Appl. Genet*. 132, 2689–2705. doi: 10.1007/s00122-019-03382-5

Li, Y., Willer, C. J., Ding, J., Scheet, P., and Abecasis, G. R. (2010). MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol*. 34, 816–834. doi: 10.1002/gepi.20533

Lorenz, A. J., Hamblin, M. T., and Jannink, J.-L. (2010). Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in Barley. *PLoS ONE* 5:e14079. doi: 10.1371/journal.pone.0014079

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

Mulitze, D. K. (1990). AGROBASE/4: a microcomputer database management and analysis system for plant breeding and agronomy. *Agron. J.* 82, 1016–1021. doi: 10.2134/agronj1990.00021962008200050035x

Muqaddasi, Q. H., Brassac, J., Koppolu, R., Plieske, J., Ganal, M. W., and Röder, M. S. (2019). TaAPO-A1, an ortholog of rice aberrant panicle organization 1, is associated with total spikelet number per spike in elite European hexaploid winter wheat (*Triticum aestivum* L.) varieties. *Sci. Rep.* 9:13853. doi: 10.1038/s41598-019-50331-9

Nyine, M., Adhikari, E., Clinesmith, M., Jordan, K. W., Fritz, A. K., and Akhunov, E. (2020). Genomic patterns of introgression in interspecific populations created by crossing wheat with its wild relative. *G3 Genes Genomes Genet.* 10, 3651–3661. doi: 10.1534/g3.120.401479

Nyine, M., Wang, S., Kiani, K., Jordan, K., Liu, S., Byrne, P., et al. (2019). Genotype imputation in winter wheat using first-generation haplotype map SNPs improves genome-wide association mapping and genomic prediction of traits. *G3 Genes Genomes Genet.* 9, 125–133. doi: 10.1534/g3.118.200664

Pasquariello, M., Berry, S., Burt, C., Uauy, C., and Nicholson, P. (2020). Yield reduction historically associated with the Aegilops ventricosa 7D V introgression is genetically and physically distinct from the eyespot resistance gene Pch1. *Theor. Appl. Genet.* 133, 707–717. doi: 10.1007/s00122-019-03502-1

Peleg, Z., Fahima, T., Abbo, S., Krugman, T., Nevo, E., Yakir, D., et al. (2005). Genetic diversity for drought resistance in wild emmer wheat and its ecogeographical associations. *Plant Cell Environ.* 28, 176–191. doi: 10.1111/j.1365-3040.2005.01259.x

Pennacchi, J. P., Carmo-Silva, E., Andralojc, P. J., Lawson, T., Allen, A. M., Raines, C. A., et al. (2019). Stability of wheat grain yields over three field seasons in the UK. *Food Energy Secur.* 8:e00147. doi: 10.1002/fes3.147

Periyannan, S., Moore, J., Ayliffe, M., Bansal, U., Wang, X., Huang, L., et al. (2013). The gene Sr33, an ortholog of barley mla genes, encodes resistance to wheat stem rust race Ug99. *Science* 341, 786–788. doi: 10.1126/science.1239028

Placido, D. F., Campbell, M. T., Folsom, J. J., Cui, X., Kruger, G. R., Baenziger, P. S., et al. (2013). Introgression of novel traits from a wild wheat relative improves drought adaptation in wheat. *Plant Physiol.* 161, 1806–1819. doi: 10.1104/pp.113.214262

Pook, T., Schlather, M., Campos, G., de los, Mayer, M., Schoen, C. C., and Simianer, H. (2019). HaploBlocker: creation of subgroup-specific haplotype blocks and libraries. *Genetics* 212, 1045–1061. doi: 10.1534/genetics.119.302283

Pour-Aboughadareh, A., Yousefian, M., Moradkhani, H., Poczai, P., and Siddique, K. H. M. (2019). STABILITYSOFT: a new online program to calculate parametric and non-parametric stability statistics for crop traits. *Appl. Plant Sci.* 7:e01211. doi: 10.1002/aps3.1211

Quintero, A., Molero, G., Reynolds, M. P., and Calderini, D. F. (2018). Trade-off between grain weight and grain number in wheat depends on GxE interaction: a case study of an elite CIMMYT panel (CIMCOG). *Eur. J. Agron.* 92, 17–29. doi: 10.1016/j.eja.2017.09.007

Reynolds, M., Chapman, S., Crespo-Herrera, L., Molero, G., Mondal, S., Pequeno, D. N. L., et al. (2020). Breeder friendly phenotyping. *Plant Sci.* 295:110396. doi: 10.1016/j.plantsci.2019.110396

Reynolds, M. P., Calderini, D. F., Condon, A. G., and Rajaram, S. (2001). Physiological basis of yield gains in wheat associated with the LR19 translocation from *Agropyron elongatum*. *Euphytica* 119, 139–144. doi: 10.1023/A:1017521800795

Reynolds, M. P., Pask, A. J. D., Hoppitt, W. J. E., Sonder, K., Sukumaran, S., Molero, G., et al. (2017). Strategic crossing of biomass and harvest index—source and sink—achieves genetic gains in wheat. *Euphytica* 213, 1–23. doi: 10.1007/s10681-017-2040-z

Saintenac, C., Zhang, W., Salcedo, A., Rouse, M. N., Trick, H. N., Akhunov, E., et al. (2013). Identification of wheat gene Sr35 that confers resistance to Ug99 stem rust race group. *Science* 341, 783–786. doi: 10.1126/science.1239022

Scott, M. F., Fradgley, N., Bentley, A. R., Brabbs, T., Corke, F., Gardner, K. A., et al. (2021). Limited haplotype diversity underlies polygenic trait architecture across 70 years of wheat breeding. *Genome Biol.* 22:137. doi: 10.1186/s13059-021-02354-7

Shaw, L. M., Turner, A. S., Herry, L., Griffiths, S., and Laurie, D. A. (2013). Mutant alleles of photoperiod-1 in wheat (*Triticum aestivum* L.) that confer a late flowering phenotype in long days. *PLoS ONE* 8:e79459. doi: 10.1371/journal.pone.0079459

Singh, N., Wu, S., Tiwari, V., Sehgal, S., Raupp, J., Wilson, D., et al. (2019). Genomic analysis confirms population structure and identifies inter-lineage hybrids in *Aegilops tauschii*. *Front. Plant Sci.* 10:9. doi: 10.3389/fpls.2019.00009

Swarts, K., Li, H., Romero Navarro, J. A., An, D., Romay, M. C., Hearne, S., et al. (2014). Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. *Plant Genome* 7, 1–12. doi: 10.3835/plantgenome2014.05.0023

Wang, J., Luo, M.-C., Chen, Z., You, F. M., Wei, Y., Zheng, Y., et al. (2013). *Aegilops tauschii* single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic diversity and pinpoint the geographic origin of hexaploid wheat. *New Phytol.* 198, 925–937 doi: 10.1111/nph.12164

Wang, R. R.-C., and Zhang, X.-Y. (1996). Characterization of the translocated chromosome using fluorescence in situ hybridization and random amplified polymorphic DNA on two *Triticum aestivum*–thinopyrum intermedium translocation lines resistant to wheat streak mosaic or barley yellow dwarf virus. *Chromosome Res.* 4, 583–587. doi: 10.1007/BF02261721

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E., et al. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90000 single nucleotide polymorphism array. *Plant Biotechnol. J.* 12, 787–796. doi: 10.1111/pbi.12183

Zhang, H., Mittal, N., Leamy, L. J., Barazani, O., and Song, B.-H. (2017). Back into the wild–Apply untapped genetic diversity of wild relatives for crop improvement. *Evol. Appl.* 10, 5–24. doi: 10.1111/eva.12434

Zhang, J., Gizaw, S. A., Bossolini, E., Hegarty, J., Howell, T., Carter, A. H., et al. (2018). Identification and validation of QTL for grain yield and plant water status under contrasting water treatments in fall-sown spring wheats. *Theor. Appl. Genet.* 131, 1741–1759. doi: 10.1007/s00122-018-3111-9

Zhang, K., Calabrese, P., Nordborg, M., and Sun, F. (2002). Haplotype block structure and its applications to association studies: power and study designs. *Am. J. Hum. Genet.* 71, 1386–1394. doi: 10.1086/344780