Received: 9 May 2022 | Accepted: 17 December 2022

DOI: 10.1002/tpg2.20309

The Plant Genome

ORIGINAL ARTICLE

Wheat doubled haploids have a marked prevalence of chromosomal aberrations

Sandesh Shrestha[1] | **Dal-Hoe Koo**[1] | **Byron Evers**[1] | **Shuangye Wu**[1] | **Sean Walkowiak**[2] | **Pierre Hucl**[3] | **Curtis Pozniak**[3] | **Allan Fritz**[4] | **Jesse Poland**[1,5]

[1]Department of Plant Pathology, Kansas State University, Manhattan, KS, USA

[2]Grain Research Laboratory, Canadian Grain Commission, Winnipeg, MB, Canada

[3]Crop Development Centre (CDC), University of Saskatchewan, Saskatoon, SK, Canada

[4]Department of Agronomy, Kansas State University, Manhattan, KS, USA

[5]Center for Desert Agriculture, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

Correspondence
Jesse Poland, Center for Desert Agriculture, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia.
Email: jesse.poland@kaust.edu.sa

Assigned to Associate Editor Mark Sorrells.

Funding information
Kansas Wheat Commission; Canadian Triticum Applied Genomics research project (CTAG2); United States Agency for International Development, Grant/Award Number: AID-OAA-A-13-00051; King Abdullah University of Science and Technology, Grant/Award Number: BAS/1/1098-01-01

Abstract

Double haploid (DH) population development is widely used in many crops, including wheat (*Triticum aestivum* L.), to rapidly produce fixed germplasm for breeding and genetic studies. The genome shock that takes place during DH induction could induce chromosomal aberrations that can impact genome integrity and subsequently plant fitness and agronomic performance. To evaluate the extent of chromosomal aberrations that exist as a result of the DH process, we studied two wheat DH populations: CDC Stanley×CDC Landmark and KS13H9×SYMonument. We utilized high-throughput skim sequencing to construct digital karyotypes of these populations to quantify deletions and aneuploidy with high resolution and accuracy, which was confirmed in selected plants by cytological analysis. The two populations studied showed high proportion of abnormal primary DH lines, 55 and 45%, respectively, based on at least one abnormality per progeny. The chromosomal abnormalities are genetically unstable and were observed segregating in the subsequent generations. These observations have important implications for the use of DH lines in genetics and breeding.

1 | INTRODUCTION

Doubled haploid (DH) technology can generate completely homozygous plants in a single generation, which is advantageous to speed the breeding progress and accelerate the release of new elite cultivars. DH lines are also used extensively to develop mapping populations for genetic studies of various traits (Santra et al., 2017). Therefore, DH lines and their derived populations are an important asset in breeding

Sandesh Shrestha and Dal-Hoe Koo contributed equally to this work.

programs and genetic studies in crop plants and are ubiquitous throughout crop research and improvement.

The development of DH plants requires several steps among which the two most important steps are: (1) inducing haploidy and (2) chromosome doubling to restore the diploid genome state (Niu et al., 2014). Each of these steps has an impact on the successful production of DH lines (Niu et al., 2014) and thus can directly influence the later process of population development for quantitative trait locus and/or gene mapping or development of improved crop varieties. Therefore, there has been much research into the optimization of the DH process for different species for haploid generation and chromosome doubling (Hooghvorst & Nogués, 2020; Ren et al., 2017). A better understanding of the impact on the genome from DH production, therefore, has vast implications for more efficient crop breeding and the development of improved crop varieties.

Bread wheat (*Triticum aestivum* L.) is one of the major sources of carbohydrates and protein consumed around the world (CGIAR, 2020). The production of staple crops, including wheat, needs to be doubled by 2050 to meet the increasing demand. Estimates of yield gains suggest a huge impending deficit compared with the predicted demand (Ray et al., 2013). This concern is heightened by the probability of less favorable climates impacting growing conditions. To meet these challenges of sustainable production, breeding cycles need to be accelerated, leading to the development of wheat varieties having improved yield potential combined with increased biotic and abiotic stress tolerance. Thus, the incorporation of DH technology into wheat breeding programs has been highly prioritized to shorten breeding cycles and accelerate the release of improved varieties. This method has already been used to successfully release cultivars of several agronomically important crops including wheat (Murovec & Bohanec, 2011).

Several approaches of haploid induction have been applied in wheat for DH development. In vitro anther culture (androgenesis or microspore culture) has been successfully used in haploid induction, production of DH plants, and the development of new cultivars (Graf et al., 2003). However, the anther culture technique has several disadvantages, including low efficiency in producing haploid plants from microspore and a high number of albino plants that lack chlorophyll (Jauhar et al., 2009). Wide hybridization is another method of haploid induction. This technique was first used to induce haploidy in barley (*Hordeum vulgare* L.) using interspecific crosses with Bulbous barley (*Hordeum bulbosum* L.) (Kasha & Kao, 1970). Haploid induction in wheat has also been reported but crossability inhibitor genes present in wheat restrict haploid induction using Bulbous barley to specific wheat genotypes (Barclay, 1975; Falk & Kasha, 1981; Snape et al., 1979). In contrast to more closely related species, maize (*Zea mays* L.) mediated haploid induction in wheat is less hindered by crossability inhibiting factors. (Laurie & Bennett, 1986; Laurie & Bennett, 1987). This method is also a gynogenesis process in which the developing haploid embryo consists only of maternal chromosomes (Niu et al., 2014). The wheat–maize wide hybridization system is now a common method to induce haploidy in wheat and has been utilized to develop and release wheat cultivars (Berg et al., 2006; DePauw et al., 2005).

After the successful generation of haploid plants, the chromosomes are doubled to restore haploid plants into fertile diploids. In case of anther culture technique, spontaneous chromosome doubling has been observed with frequency ranging from 15 to 44% (Stober & Hessu, 1997). The antimicrotubule agent, colchicine, has been successfully and widely used to induce chromosome-doubling. Colchicine interferes with microtubule function and results in an abnormal mitotic division known as C-mitosis, where polar segregation of sister chromatids are disrupted, resulting in nonreduced gametes, which subsequently leads to chromosome-doubling (Castillo et al., 2009). The successful production of DH plants presumes a complete doubling of all chromosomes, but the disruption of spindle fibers is an inexact process and disruption could be partial or incomplete (Grant, 1978; Sharma, 1990). Various compounds that induce chromosome doubling, including colchicine, have been reported to cause the loss or gain of chromosomes, also known as aneuploidy, in hexaploid wheat (Sandhu et al., 1991).

Developed DH lines are considered the gold standard for true breeding materials, such that all progeny should be completely homozygous and homogenous. Given the importance of DH lines for genetic studies and the extensive use in wheat breeding programs, any exceptions from this expectation could have considerable implications. The aberrations present in the DH lines could impact the agronomic performance and subsequent breeding effort and genetic studies, or effectively render a DH line as heterogenous. Until now, the extent of chromosomal aberrations present in DH lines has not been explored in depth. Therefore, we set out to

Core Ideas

- Doubled haploids (DH) are central to accelerated breeding efforts being considered as completely uniform.
- We observed high heterogeneity within DH lines from two wheat populations.
- Chromosome aberrations in DH lines included chromosome deletions, duplications, and various aneuploidy.
- Chromosome aberrations and heterogeneity have important implications for the use of DH lines.

systematically characterize the presence of various chromosomal aberrations in newly developed DH populations. We leveraged robust de novo genome assemblies from the Wheat 10+ Genomes Project (Walkowiak et al., 2020) combined with high-throughput skim sequencing to characterize chromosome integrity across the genome at high resolution in large populations. By optimizing and applying this novel skim-sequencing approach, we are able to accurately characterize and classify the dosage of chromosomes and chromosomal segments at high resolution. This approach to quantify DNA copy number at high resolution has been previously termed as "digital karyotyping" (Wang et al., 2002), which we will refer to as the read mapping method we used to characterize chromosomal dosage. Here, we report an extensive analysis of chromosome integrity in wheat DH lines and report many abnormalities in two separate DH populations, representing both spring and winter wheat, developed by wheat breeding programs from Crop Development Centre at University of Saskatchewan and Kansas State University.

2 | RESULTS

2.1 | Skim sequencing of DH populations

To evaluate the chromosome composition across full DH populations, we skim sequenced (Adhikari et al., 2022) one to six individual plants from each of 145 DH lines originating from a CDC Stanley×CDC Landmark DH population (StanleyLandmarkDH), along with the two parents (StanleyDH and LandmarkDH). We also included individual self-pollinated progeny from the individual DH plants for a total of 781 DH plants, which were sequenced to a raw mean coverage of 0.13×. As high-quality checks for variants, we included the previously available sequencing data used for reference genome assembly (Walkowiak et al., 2020) at an average coverage of 8.31×. Similarly, we evaluated 38 DH lines from a KS13H9×SYMonument population (KS13H9×SYMonumentDH) with sampling of three to six individual plants per line for a total of 187 individual plants, generating mean coverage of 0.034×. To generate high-quality alternate variants between two parents, KS13H9 and SYMonument were sequenced at an average coverage of 7.28×.

The two DH populations had similar total and concordant unique alignment when mapped to their respective reference genome (Tables S1 and S2) with average total alignment in the range of 71–74% and concordant unique alignment of 55–60% of the total reads. This indicated strong conservation and utility of the reference genomes. Consistent with divergence between the reference genome and alternate parent being evaluated, we observed consistent regions of lower read mapping from the alternate parent, which segregated at 1:1 ratio in the progeny, indicating highly diverse regions of greater sequence divergence. A prime example is seen with the translocation from *Aegilops ventricosa* on wheat chromosome 2A designated as $2N^{v}S$ (Gao et al., 2020), present in CDC Stanley and absent in CDC Landmark, that appears as a missing segment in CDC Stanley with almost no reads mapping and consistently segregating in the DH progeny (Figure S1). Therefore, we took careful consideration of the density of read mapping in both parents prior to considering a segment as an actual deletion in the DH progeny. Read mapping that matched deviations from the normal 1× coverage in parents were not scored as aberrations in DH progeny. We did a careful manual check of all chromosomes in the parents before scoring aberrations in the DH lines (Figure S2). In addition, we were able to observe normal segregation (e.g., 1:1) for segments where the two parents varied in mapped read depth, which contrasted to segments in a single DH line that were higher or lower than the respective read depth mapping in the parents. In addition, the parental genotypes as described below were confirmed giving an additional check that the observed read depth of that segment matched the corresponding parent from which the segment was inherited.

2.2 | SNP distribution

The variant calling of StanleyLandmarkDH population was performed with high coverage sequence data of two parents leading to discovery and genotyping of 15 million SNPs. As we are working with inbred parental lines and DH progeny, each sample was filtered separately to remove missing and heterozygous positions with an average number of SNP per DH of 549,954. For the KS13H9×SYMonumentDH population, 7.8 million SNPs were discovered in the high-coverage parents and genotyped in the DH lines. After removing missing and heterozygous loci, the KS13H9×SYMonumentDH lines had 208,569 usable marker genotypes on average. The SNPs that were homozygous with alternate alleles between two parents were used to define recombination breakpoints and inheritance of chromosomal segments in the DH lines. As noted previously, the genomic segment represented by SNP alleles directly correlated with the read mapping as seen in the two parents. This facilitated the visual scoring of genomic segments in DH lines that deviated from normal 1× coverage due to divergence between the parental genomes (Figure S2).

2.3 | Validation of chromosomal aberrations by karyotype evaluation

To confirm the inferred chromosomal constitution from the read counts, we employed cytological observations with subgenome-specific FISH probes. This can be

considered the gold standard for karyotyping, as it allows identification of each individual chromosome and can confirm the exact karyotype of DH lines. Using the cytogenetic analysis, we were able to confirm the multiple types of aberrations such as deletions and aneuploidy identified by digital karyotyping. For example, one of the $DH_{1:2}$ plants (StanleyLandmarkDH02014-0-2) had nullisomic, monosomic, and trisomic chromosomes identified via digital karyotyping, which was confirmed with cytology (Figures 1a and b). Overall, we completed cytological evaluation of 12 normal plants and 28 plants with abnormal karyotypes with complete agreement between the digital karyotypes and the cytological observations (Figures S3 and S4). An important caveat of the digital karyotyping is that the exact location of chromosomal duplications cannot be unambiguously determined. In these cases, the duplicated segment could be a tandem duplication or a translocation to another chromosome. Given that cytogenetic evaluations are an intense and tedious process, our data support that digital karyotyping from low coverage sequencing can be used as a low-cost, high-throughput alternative to cytological classification of chromosomal aberrations for rapid identification of samples of interest, which can be further investigated with cytology.

2.4 | Chromosomal aberrations in the primary DH generation

Recognizing that these DH plants were heterogenous for deletions and aneuploidy, we employed pedigree tracking within DH lines for individual plant progenies from each primary DH plant. In this "pedigree," we consider the primary DH plant as the DH_0 generation. As these plants were not sampled, we first evaluated multiple individuals from the selfed seed of the primary DH plants, which were designated as $DH_{0:1}$ concordant with a usual filial generational nomenclature (Fehr, 1991). Examining the first-generation DH plants in StanleyLandmarkDH population, digital karyotyping identified 154 (39%) out of 396 $DH_{0:1}$ plants had at least one chromosomal aberration in the form of deletions, aneuploidy, or duplications/translocations (Supplemental Figures S5 and S6 and Table S3). These abnormal plants were found in 79 of the 145 DH lines, meaning that 54.5% of the DH lines had at least one abnormal $DH_{0:1}$ progeny. Given the difference between a total percentage of 39% abnormal plants from 55% of the lines that carried a type of chromosome aberration, it could be inferred that the impact on fitness and survival of abnormal karyotypes is severe enough that a proportion is lost from the first generation.

Many of the $DH_{0:1}$ plants had multiple aberrations resulting in a total of 239 aberrations among the 154 abnormal $DH_{0:1}$ plants (Table 1). Interestingly, aberrations were observed

TABLE 1 The number and type of chromosomal aberrations observed in 154 abnormal out of 396 $DH_{0:1}$ plants in StanleyLandmarkDH population and in 32 abnormal out of 187 $DH_{0:1}$ plants in KS13H9×SYMonumentDH population

Population/ aberration type	1A	1B	1D	2A	2B	2D	3A	3B	3D	4A	4B	4D	5A	5B	5D	6A	6B	6D	7A	7B	7D	Total aberrations	$DH_{0:1}$ plants with at least one aberration	$DH_{0:1}$ with specific aberration (%)
StanleyLandmarkDH																								
Aneuploidy	3	1	3	1	3	4	2	2	2	0	9	6	3	4	5	1	2	7	14	9	0	81	66	17
Deletion	5	1	2	15	8	6	1	7	13	1	8	17	4	3	2	0	3	3	13	2	0	114	95	24
Duplication/translocation	0	2	0	1	3	4	0	1	6	0	6	7	2	0	2	1	2	0	6	1	0	44	37	9
Total	8	4	5	17	14	14	3	10	21	1	23	30	9	7	9	2	7	10	33	12	0	239		
KS13H9×SYMonumentDH																								
Aneuploidy	5	1	1	0	1	1	1	1	4	0	1	4	0	0	1	0	2	4	0	0	1	28	22	12
Deletion	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	4	4	2
Duplication/translocation	0	2	5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	8	8	4
Total	5	3	6	1	1	1	1	2	4	0	2	4	0	0	1	0	4	4	0	0	1	40		

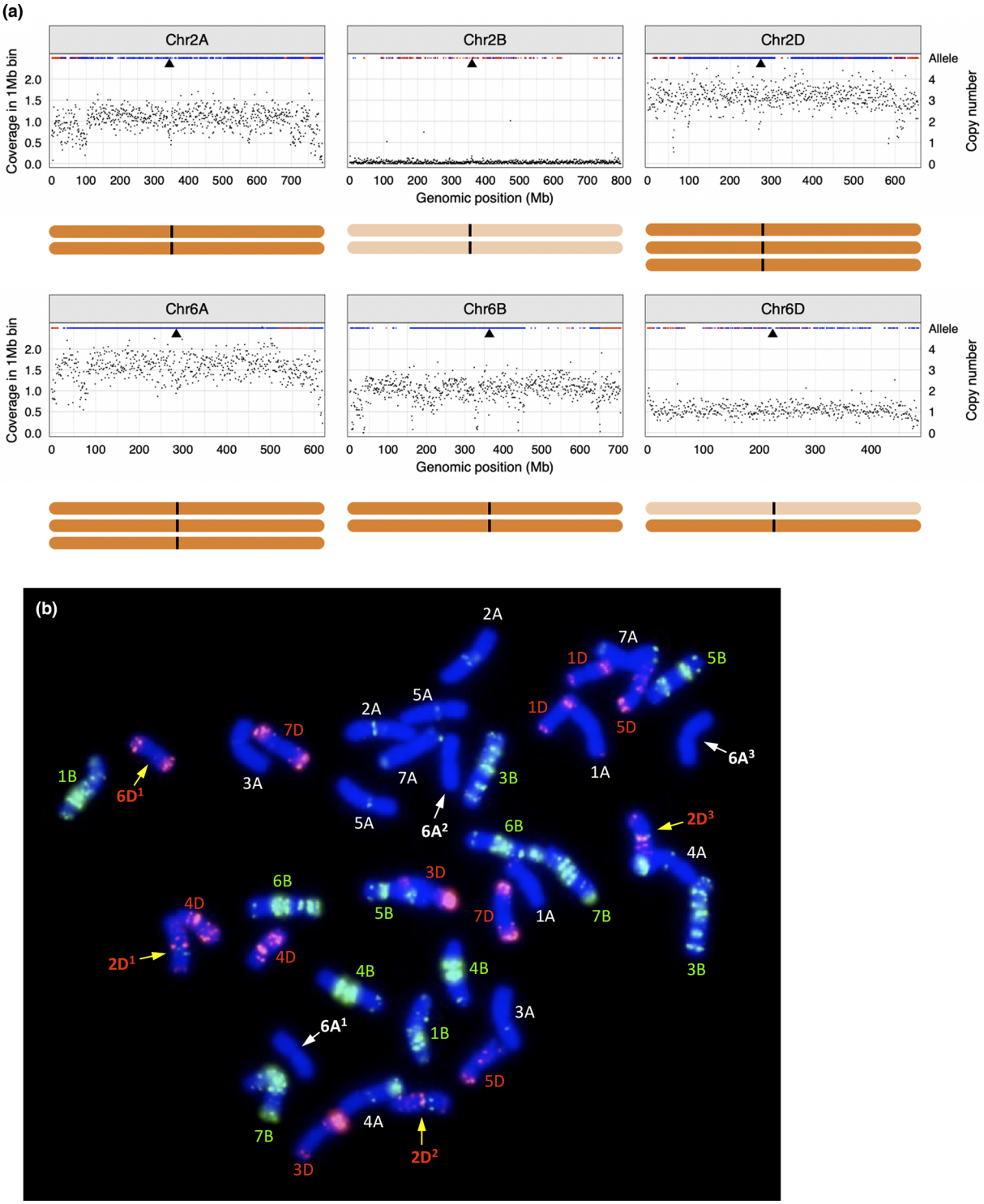

FIGURE 1 The evaluation of a single doubled haploid (DH) plant (StanleyLandmarkDH02014−0−2) derived from CDC Stanley×CDC Landmark that was identified as nullisomic for chromosome 2B, monosomic for chromosome 6D, and trisomic for chromosomes 2D and 6A based on digital karyotype (a) and confirmed with cytology (b); (a) digital karyotype: each panel shows the genomic position of a given chromosome in 1 Mb bins on the physical map of the CDC Landmark reference genome in the *x*-axis with points showing the normalized number of reads mapping to the respective bin. The vertical axis shows both normalized read count (left) and chromosome dosage (right). The parental allele for individual SNP variants is shown on top for CDC Landmark (red) and CDC Stanley (blue) alleles. Centromere positions are showing as black triangle. Pictorial representation of the inferred chromosome constitution is shown below the graphs: dark and light orange shows presence and absence of chromosomal segments, respectively, and (b) cytological analysis: florescence in situ hybridization (FISH) for sites of hybridization with fluorescence-labelled probes, GAAn (green) and pAs1 (red). The observed 41 chromosomes including zero copy of chromosome 2B, one copy of chromosome 6D (yellow arrow), three copies of 2D (yellow arrows) and three copies of 6A (white arrows). Numerical superscript on chromosome designation represents the occurrence of the respective chromosome. The chromosomes were identified according to the standard FISH karyotype (Danilova et al., 2012).

TABLE 2 Distribution of chromosomal aberrations in three subgenomes after length normalization to 5 Gbp subgenome size

Population/subgenome	Number of aberrations	Genome length (bp)	Number of aberrations after genome length normalization	Aberration after normalization (%)
StanleyLandmarkDH		CDC Landmark		
A	73	4966053268	73.5	28.3%
B	77	5204724784	73.9	28.5%
D	89	3982871035	111.7	43.1%
KS13H9×SYMonumentDH		Jagger		
A	7	4983156636	7.0	15.6%
B	12	5219166998	11.5	25.5%
D	21	3970003109	26.4	58.8%

across the genome, involving all chromosomes except chromosome 7D. Most chromosomes were observed with aberrations in multiple $DH_{0:1}$ plants, with up to 33 aberrant DH plants observed for chromosome 7A. After normalizing for variation in subgenome length, the number of aberrations observed in the D genome was highest with 43.1% compared with A and B subgenomes (Table 2).

When compared with the StanleyLandmarkDH population, KS13H9×SYMonumentDH had less abnormal $DH_{0:1}$ plants with 32 (17%) out of 187 plants having at least one form of chromosomal aberration (Figure S7 and Table S4). However, when the DH lines are considered as an entity per se, 17 (44.7%) out of 38 DH lines were abnormal based on at least one abnormal $DH_{0:1}$ progeny, which was similar to the StanleyLandmarkDH population. Likewise, the distribution of observed aberrations after subgenome length normalization was higher in the D subgenome (58.82%) compared with the A and B subgenomes (Table 2). Interestingly, in this population, we were able to observe aneuploidy for chromosome 7D, while 7D was the only chromosome without any aberration in the StanleyLandmarkDH population. There were six chromosomes in the KS13H9×SYMonumentDH population without any aberrations (*viz*., 4A, 5A, 5B, 6A, 7A, and 7B.)

The deletions in DH lines were observed in various forms including complete absence of the short or long arm (telosomics), terminal deletions of the short and/or long arm, as well as on the interstitial deletions of various chromosomes (Figure 2). In the $DH_{0:1}$ generation of StanleyLandmarkDH, 95 (24%) out of 396 plants had at least one deletion. We observed a total of 114 different deletions with 15 plants that carried multiple deletions (Tables 1 and S3). Deletions were the most common type of aberration observed in this DH population. In contrast, deletions were the least observed type of aberrations in KS13H9×SYMonumentDH population, which were only identified in four (2.1%) out of 187 $DH_{0:1}$ plants (Tables 1 and S4).

We also observed 66 $DH_{0:1}$ plants (17%) had some form of aneuploidy in the StanleyLandmarkDH population (Figure 3, Tables 1 and S3). The number of aneuploid plants was the second most common type of aberration after deletions, which was observed 81 times in $DH_{0:1}$ of this population, with nine plants having aneuploidy for multiple chromosomes (Tables 1 and S3). In the case of the KS13H9×SYMonumentDH population, the number of aneuploids observed was the highest compared with all other forms of aberration, which were identified in a total of 22 (12%) out of 187 $DH_{0:1}$ plants (Tables 1 and S4). Finally, we observed duplication/translocation events in both populations, with a higher frequency in the KS13H9×SYMonumentDH population (Figure 4 and Table 1). These duplication/translocations were third and second most common when compared with other forms of aberrations in StanleyLandmarkDH and KS13H9×SYMonumentDH, respectively (Table 1).

Beyond clearly defined deletion and duplicated segments, we observed severe forms with complicated aberrations and chromosome dosage that could not easily be interpreted (Figure 5). These complex aberrations were only observed in the StanleyLandmarkDH population and their complexity made it difficult to determine an accurate digital karyotype. The FISH result that we had from StanleyLandmarkDH02074-2 showed all normal chromosomes except homozygous terminal deletion of 7A (Figure S3). Regardless, these particular lines have extremely complex genome rearrangements, involving multiple chromosome breakage and loss and resulting in severely distorted karyotypes. Interestingly, we observed that some of the dosage changes, and hence inferred breakpoints, were apparently at recombination junctions as evidenced by a switch in parental alleles at the same site. However, other breakpoints were found within stretches of chromatin inherited from the same parent.

Strikingly, we observed complete homozygosity in the parental alleles throughout all of the various types of aberrations observed here. Regardless of dosage or complete or partial aneuploid, the genotype calls were uniformly from one of the respective parents and did not show any indications of heterozygosity. This supports that all of these plants did indeed originate from the primary haploid plant(s) and that the chromosomal aberrations during or post haploid induction. The uniform homozygosity also helps to exclude any

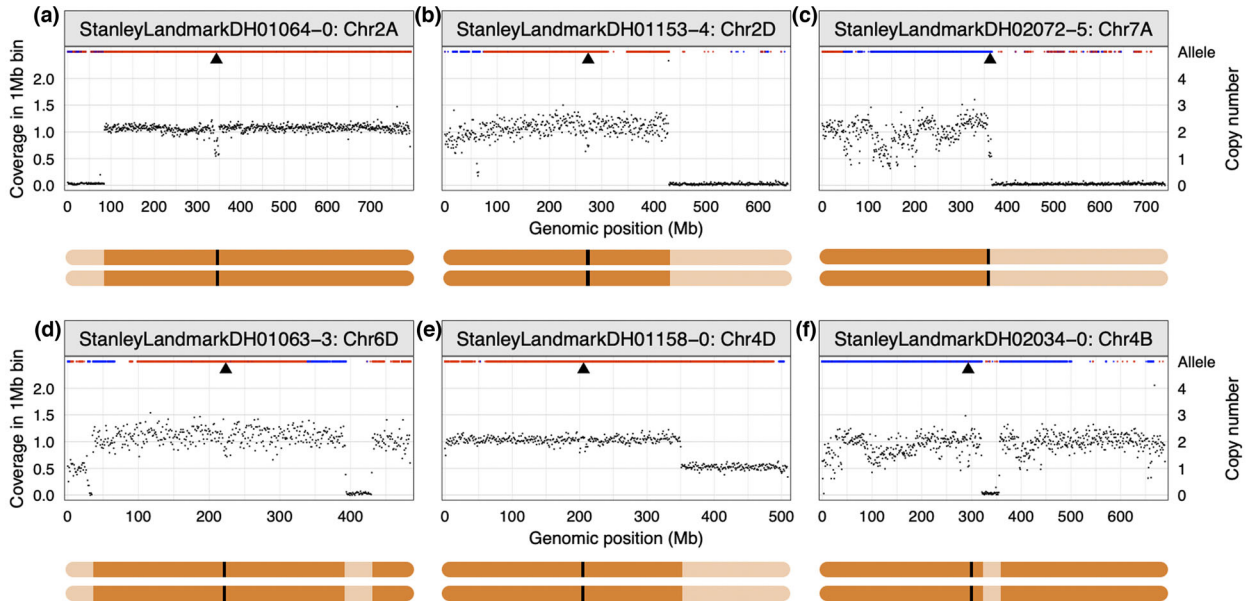

FIGURE 2 Example digital karyotypes carrying deletions in doubled haploid (DH) lines derived from CDC Stanley×CDC Landmark. Each panel shows a given chromosome for an individual DH plant with the genomic position of the chromosome in 1 Mb bins on the physical map of the CDC Landmark reference genome in the *x*-axis with points showing the normalized number of reads mapping to the respective bin. The vertical axis shows both normalized read count (left) and chromosome dosage (right). The parental allele for individual SNP variants is shown on top for CDC Landmark (red) and CDC Stanley (blue) alleles. Centromere positions are showing as black triangle. Pictorial representation of the inferred chromosome constitution is shown below the graphs: dark and light orange shows presence and absence of chromosomal segments, respectively. Inferred digital karyotypes shown are (a) chromosome 2A with a homozygous terminal short arm deletion; (b) chromosome 2D with a homozygous terminal long arm deletion; (c) chromosome 7A with deletions of both long arms (telosomic); (d) chromosome 6D with a hemizygous terminal short arm and a homozygous interstitial long arm deletion; (e) chromosome 4D with a hemizygous terminal long arm deletion and f) chromosome 4B with a homozygous interstitial deletion.

possibility that the observed plants and DH lines were mixed genotypes from outcrossing.

2.5 | Evaluation of chimeric chromosomal aberrations in primary DH lines

Following the observation that many of the deletions and duplications were hemizygous, we hypothesized that the primary DH plants are a chimera of different aberrations. There were 86 primary DHs having at least two $DH_{0:1}$ progeny of which 26 (30.2%) of them produced heterogenous $DH_{0:1}$ plants. The heterogenous $DH_{0:1}$ family included completely normal plants as well as abnormal plants having at least one aberration (Table S5). The remaining primary DH lines generated either all normal ($n = 39$; 45.4%) or all abnormal ($n = 21$; 24.4%) $DH_{0:1}$ plants. This shows that the DH plants have a high level of heterogeneity (heterozygous/hemizygous) within the plant and the resulting DH line. While a combination of normal and abnormal progenies from the same primary DH indicates heterozygousity of the primary DH per se, the hypothesis of a chimeric karyotype within a plant has a prediction of observing more than two haplotypes (e.g., chromosomal karyotypes of the same chromosome) within a diploid individual. Specifically, this would be due to different cells and cell lineages with different aberrations, which leads to a segmental mosaic within the plant. Supporting this hypothesis, we observed two DH lines (lineages), which showed three different chromosome karyotypes for chromosome 6D in the resulting progeny (Figure 6). In the example of progeny from DH plant StanleyLandmarkDH01063 following selfing, this included (i) two normal chromosomes in progeny no. 0 (DH01063-0) (Figure 6a) as well as one normal chromosome in progeny no. 1, 2, and 4 (Figures 6b, 6c, and 6e), (ii) chromosomes with a single deletion in progeny 2, 3, and 4, which showed a deletion on the long arm (Figures 6c, 6d, and 6e), and (iii) chromosomes with two deletions as evidenced in progeny no. 3 that had a chromosome with deletions in both arms (Figure 6d). This DH line also showed a terminal end of the short arm of chromosome 6D, which lacked a centromere, it must have been a duplication on 6D or a segmental translocation to another chromosome. The evidence of three different chromosomal haplotypes originating from a self-pollinated plant with diploid-like inheritance (e.g., allohexaploid), indicated that the parental plant was a mosaic of these different haplotypes.

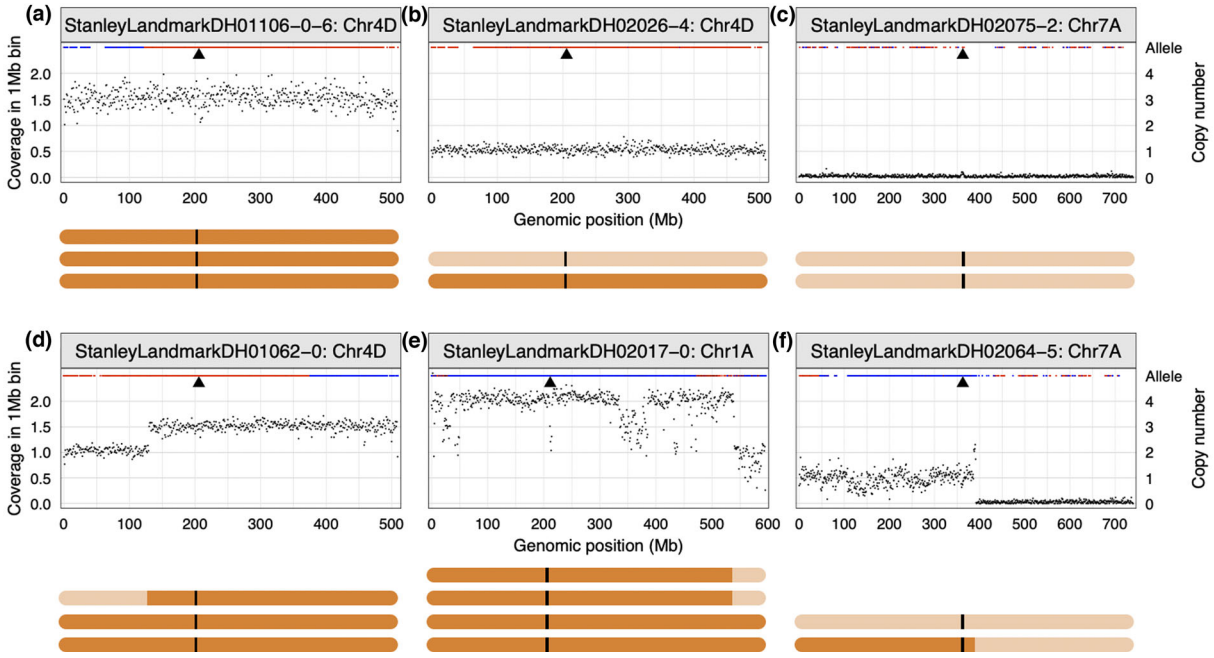

FIGURE 3 Example digital karyotypes carrying aneuploidy in doubled haploid (DH) lines derived from CDC Stanley×CDC Landmark. Each panel shows a specific chromosome from an individual DH plant. The genomic position of the respective chromosome in 1 Mb bins on the physical map of the CDC Landmark reference genome is on the *x*-axis with points showing the normalized number of reads mapping to the respective bin. The vertical axis shows both normalized read count (left) and chromosome dosage (right). The parental allele for individual SNP variants is shown on top for CDC Landmark (red) and CDC Stanley (blue) allele. Centromere positions are showing as black triangles. Pictorial representation of the inferred chromosome constitution is shown below the graphs: dark and light orange shows presence and absence of chromosomal segments, respectively. Inferred digital karyotypes shown are (a) trisomic chromosome 4D, (b) monosomic chromosome 4D, and (c) nullisomic chromosome 7A. Examples of partial aneuploid in DH lines; (d) three copies of chromosome 4D with a terminal short arm deletion in a copy; (e) four copies of chromosome 1A with a terminal long arm deletion in two copies and (f) one copy of chromosome 7A carrying a long arm deletion.

2.6 | Segregation of chromosomal aberrations in the $DH_{1:2}$ generation

To evaluate and confirm the heterogeneity of DH lines and the inheritance of the observed deletions, we evaluated sets of progenies from respective $DH_{0:1}$ plants to observe the expected segregation in the following $DH_{1:2}$ generation (Figure S8 and Table S6). We expected that the abnormal plants with hemizygous aberration will have segregation of the aberration in $DH_{1:2}$ plants. All of the 19 $DH_{0:1}$ that had observed genomic abnormalities and had progeny in the $DH_{1:2}$ generation, demonstrated segregation of the aberrations seen in $DH_{0:1,}$ such as a monosomic chromosome 5B segregating in $DH_{1:2}$ generation as monosomic and disomic (StanleyLandmarkDH01038-0; Figure S9). This further confirms the observations that these chromosome aberrations were being accurately scored in the $DH_{0:1}$ plants and that resulting progeny are essentially segregating for null alleles.

There were some notable exceptions where we observed progeny carrying a chromosomal segment that was completely absent in the parental line. This phenomenon was observed in $DH_{1:2}$ progenies of three $DH_{0:1}$ plants (Figure S10). For example, a homozygous deletion in StanleyLandmarkDH02009-0 was observed as hemizygous and complete chromosome 4D in the progeny of this plant (Figure S10a). We made a careful assessment for any error and cross-checked the assignment of plant ID and sequencing/genotyping data to ensure data integrity. Given that the progenies are hemizygous for the deletion at the exact same breakpoint as the homozygous deletion in the $DH_{0:1}$ generation, however, supports that these are the correct lineage with this pedigree. Further to this, individual plants within the lines all carry the same homozygous variants from SNP genotypes across the genome, confirming that they are indeed the same DH line and not a seed mixture, contamination, or cross-pollination from another line. While perplexing, it remains to be studied if the first-generation DH plants are mosaic for the given deletions, a possible explanation of genotyping from a leaf sample that could differ in genetic composition from other sectors of the meristem, which include the eventual spike. Supporting that the primary and other DH are mosaic, we observed both complete chromosomes and chromosomes with the same deletion (same

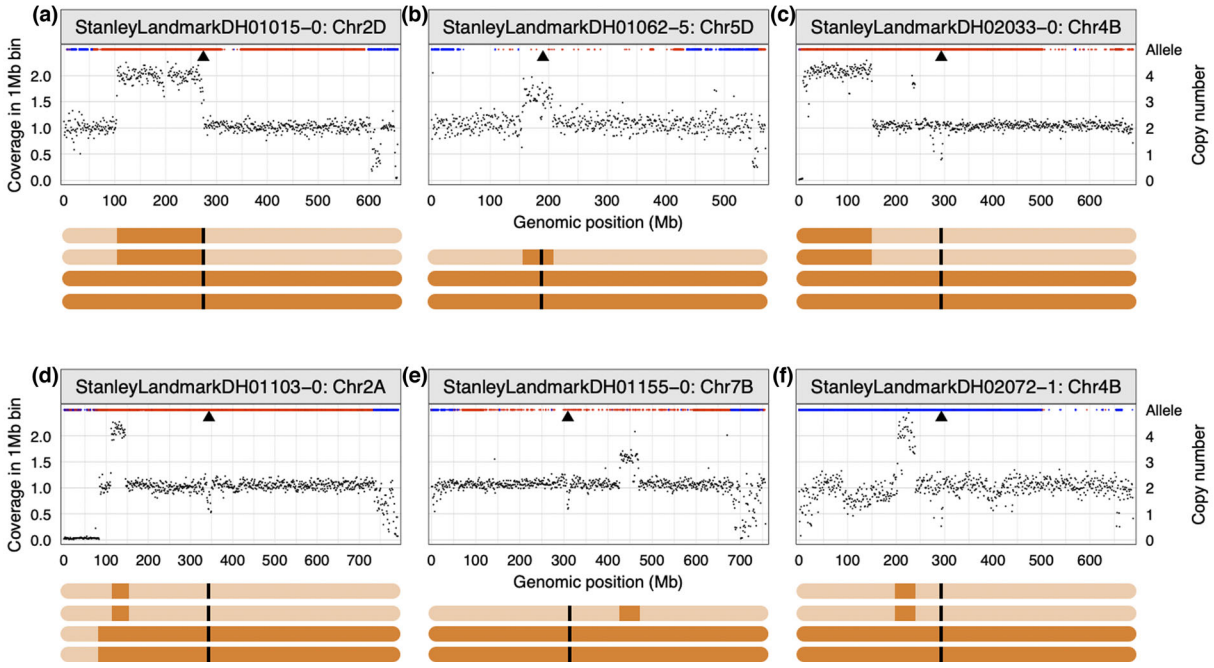

FIGURE 4 Example digital karyotypes carrying duplications and/or duplicated translocations in doubled haploid (DH) lines derived from CDC Stanley×CDC Landmark. Each panel shows a specific chromosome from an individual DH plant with the genomic position of the chromosome in 1 Mb bins on the physical map of the CDC Landmark reference genome in the *x*-axis with points showing the normalized number of reads mapping to the respective bin. The vertical axis shows both normalized read count (left) and chromosome dosage (right). The parental allele for individual SNP variants is shown on top for CDC Landmark (red) and CDC Stanley (blue) alleles. Centromere positions are showing as black triangles. Pictorial representation of the inferred chromosome constitution is shown below the graphs: dark and light orange shows presence and absence of chromosomal segments, respectively. Inferred digital karyotypes shown are (a) two copies of an interstitial duplication/translocation in chromosome 2D. (b) One copy of an interstitial duplication/translocation in chromosome 5D. (c) Two copies of a duplication/translocation in the terminal end of the short arm of chromosome 4B. (d) A homozygous terminal short arm deletion and two copies of an interstitial duplication/translocation of chromosome 2A. (e) One copy of an interstitial duplication/translocation of chromosome 7B and (f) two copies of an interstitial duplication/translocation in chromosome 4B.

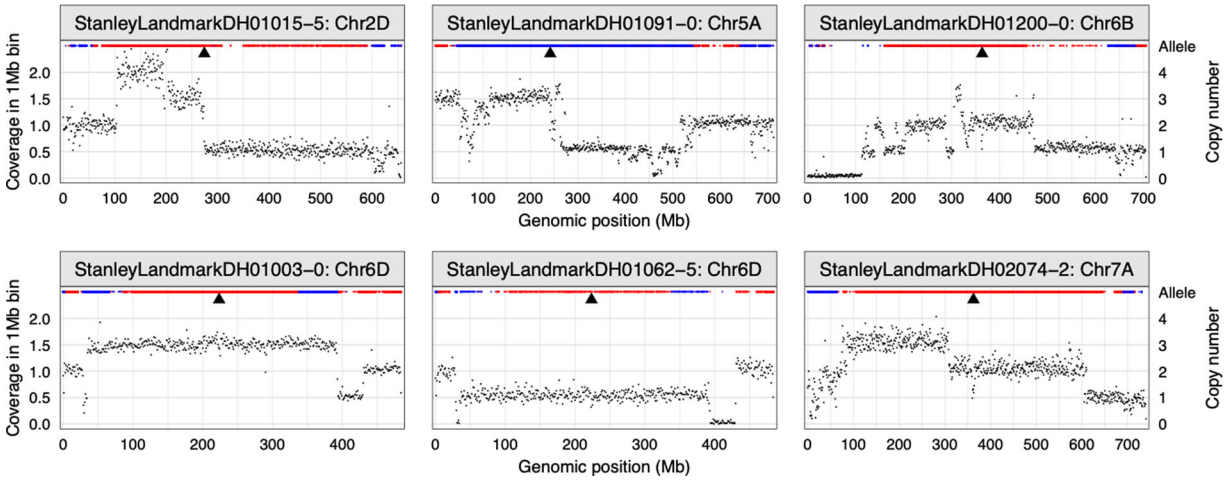

FIGURE 5 Example digital karyotypes carrying complicated aberrations in doubled haploid (DH) lines derived from CDC Stanley×CDC Landmark. Each panel shows a single chromosome from an individual DH plant with the genomic position of the chromosome in 1 Mb bins on the physical map of the CDC Landmark reference genome on the *x*-axis with points showing the normalized number of reads mapping to the respective bin. The vertical axis shows both normalized read count (left) and chromosome dosage (right). The parental allele for individual SNP variants is shown on top for CDC Landmark (red) and CDC Stanley (blue) alleles. Centromere positions are showing as black triangles.

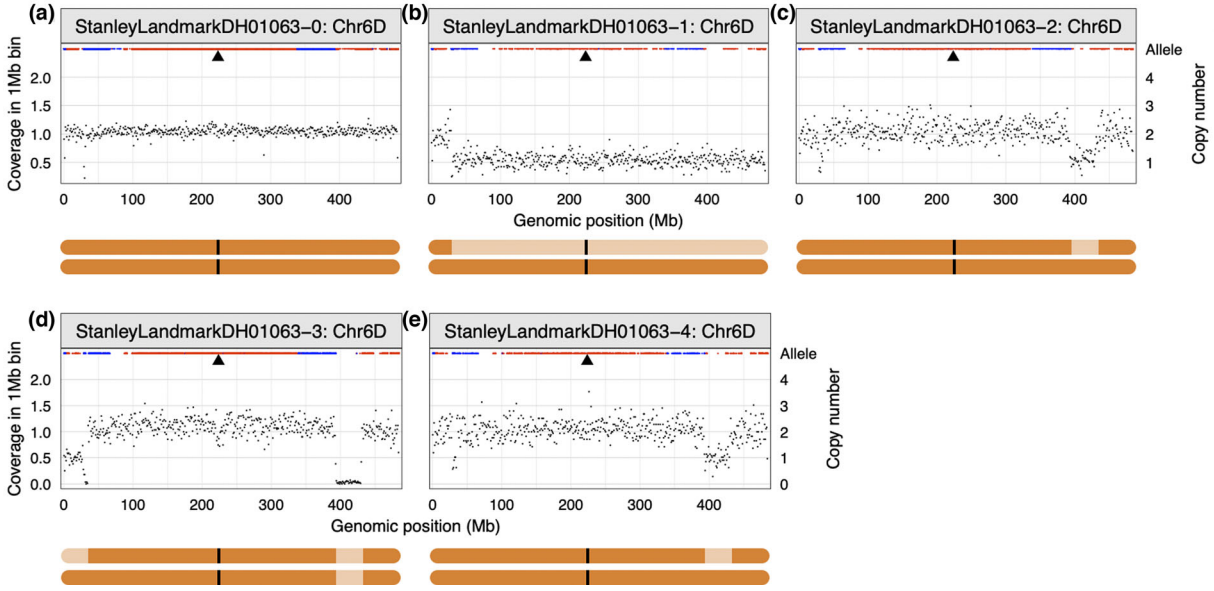

FIGURE 6 Digital karyotype of chromosome 6D of five progenies (no. 0, 1, 2, 3, and 4) of a chimeric primary double haploid plant which formed the doubled haploid (DH) line StanleyLandmarkDH01063 derived from CDC Stanley×CDC Landmark. Each panel is showing chromosome 6D from a single progeny for (a) StanleyLandmarkDH01063-0; (b) StanleyLandmarkDH01063-1; (c) StanleyLandmarkDH01063-2; (d) StanleyLandmarkDH01063-3; and (e) StanleyLandmarkDH01063-4. The *x*-axis shows the genomic position of the chromosome in 1 Mb bins on the physical map of the CDC Landmark reference genome with points showing the normalized number of reads mapping to the respective bin. The vertical axis shows both normalized read count (left) and chromosome dosage (right). The parental allele for individual SNP variants is shown on top for CDC Landmark (red) and CDC Stanley (blue) alleles. Centromere positions are showing as black triangles. Pictorial representation of the inferred chromosome constitution is shown below the graphs: dark and light orange shows presence and absence of chromosomal segments, respectively.

chromosome and breakpoint) in other $DH_{0:1}$ plants from these lines (e.g., Figures S10b and c).

3 | DISCUSSION

The DH technology speeds up the breeding process by producing homozygous pure lines for evaluation in a short period of time (Wędzony et al., 2009) but the stress exerted on the genome during this process can induce significant structural changes that can undermine the integrity of the genome in the resulting DH lines. The induction of chromosomal aberrations such as aneuploidy has been reported due to colchicine in plants as well as animals (Schmid et al., 1999; Sharma, 1990). Here, we observed that the extreme pressure exerted on the genome in wheat during the process of haploid induction and chromosome doubling has resulted in significant chromosomal aberrations in the DH progeny, including duplication, deletion, translocation, and aneuploidy. The evaluation of DH lines from two independent wheat breeding programs showed a high level of chromosomal aberrations exists in DH lines, confirming that this is a prevalent phenomenon in the wheat breeding DH process. Many of these deletions and duplications effectively rendered the DH line as heterozygous. From these populations, we observed multiple different aberrations for every chromosome, supporting that the induced chromosome changes are not isolated, but genomewide. The difference in the overall level of aberrations in the two populations could be due to the wheat genotypes or could have been affected by the developmental process, which was carried out by separate breeding programs and different DH laboratories, or even the genetic background of the germplasm and the growth habit (spring vs. winter). Even though there are varying levels of aberrations in DH lines, ranging from single to multiple aberrations in chromosomes, a single aberration could make the DH line unsuitable for further development unless corrected by single plant selection and purification.

The use of DH lines in breeding programs and for genetic studies presupposes that the lines are completely homozygous and stable euploids. Some of the aberrations observed in the two DH populations were hemizygous and segregating within the respective DH line and in the subsequent generations. For breeding populations, where selection is heavily applied on plant type and agronomics, it could be expected that any genomic deficiencies may be quickly eliminated. Further experiments are needed to determine the effectiveness of typical breeder selections in removing the aberrant DH lines. However, this would also implicate, and perhaps necessitate, single plant purification of advanced lines and breeders' seed out of DH-derived materials. Another real possibility with the heterogeneity of these DH lines is that superior lines are discarded because aneuploids within the line are negatively affecting the aggregate performance of the line for visual and phenotypic selection even though the euploid

genotype is superior but confounded with heterogenous aneuploidy.

Another implication for breeding is that the DH process is essentially creating a substantial amount of null genetic mutations which could be considered an increased genetic load if not continually removed through selection. Any terminal deletions have also rendered a given chromosome unable to pair with the normal intact homologous chromosome, limiting any further use in breeding (Gill & Friebe, 1998). This becomes a critical limitation for breeders because even if a given DH line is vastly superior but contains even a small terminal deletion, it loses value as a good parent due to a lack of pairing and recombining. Likewise, any null alleles present in the lines and the breeding population are likely to increase the incidences of sterility which may be manifested in various progeny or different conditions.

While the selection in breeding populations is expected to quickly remove chromosome abnormities, DH populations intended for genetic studies are often advanced without selection and evaluated as complete populations (Costa et al., 2001). An important consideration in mapping is that the DH lines will be genotyped as homozygous for a given allele though it actually could be hemizygous. As observed with the genotype calls of parental alleles in this study, the DH plants and lines could be considered completely homozygous. Depending on the genetic architecture of the trait under study, this could have implications on accurately mapping the trait if the association test does not account for potentially hemizygous and/or null alleles. The presence of aneuploidy and chromosomal deletions in these materials would persist and could have important implications on phenotyping and genetic associations.

Our study demonstrated that the newly developed digital karyotyping method using low-coverage skim sequencing could be used in deciphering the segmental or chromosomal dosage along the chromosomes. While many genotyping technologies including SNP arrays will only identify the allelic variants, skim sequencing of DH lines enables both genotyping and dosage assessment (Adhikari et al., 2022; Allen et al., 2017; Mason et al., 2017). Our finding shows that even very low-coverage sequencing can easily detect aberrations of various forms. Cytogenetics techniques such as FISH or flow cytometry have also been used in the identification of chromosomal aberrations (Danilova et al., 2012; Kim et al., 2003; Pfosser et al., 1995) but the applications of these methods on thousands of lines are not practical. Since multiplexed skim sequencing with low coverage per sample is cheaper and parallel bioinformatics processing is faster, the digital karyotype approach could be an effective and quicker way of scanning thousands of DH lines to detect abnormalities. Only one limitation of the digital karyotyping is the inability to distinguish between segmental duplications and translocation because the reads originating from these aberrations will map on the same segment on the chromosome. This limitation will not hinder the digital karyotyping in differentiating normal DH lines from abnormal ones unless the goal is to accurately assess if a segment is found in a duplication versus a translocation.

The development of DH lines for crop improvement is an expensive enterprise. The DH plant that is nullisomic for a pair of any chromosome or has a homozygous deletion in any chromosome cannot be recovered and becomes unsuitable for further breeding. As observed here, the genomic abnormalities generated during the DH process necessitate careful consideration for using DH populations in both breeding and genetics. While novel technologies such as skim sequencing can quickly and effectively screen DH lines, there are still implications for how to handle abnormal DH lines in the breeding cycle. Selection and removal of abnormal DH plants and lines will further increase the associated cost of DH production and breeding with DH. At the same time, however, there is continual investment and research into optimizing the DH process in different species. Commensurate with this work, the presented methodology and observations in this study can add an important dimension to the optimization of DH methods to focus not only on the overall induction rate but also on the integrity of the resulting DH genome. Alternatively, both the overall cost of DH and the genomic implications presented here raise additional cost/benefit factors into optimal breeding strategies where a method such as speed breeding to rapidly advance generations and generate homozygous recombinant inbred lines (RILs) for testing and crossing could become more advantageous (Watson et al., 2018). Similar follow-up studies are needed to examine the chromosome integrity in "speed-bred RILs." If other factors including time and costs are equivalent, the possibility of increased genome integrity from inbreeding compared with doubling haploids could prove advantageous.

Chromosomal aberrations observed in $DH_{0:1}$ segregated as expected in $DH_{1:2}$ progenies with a few exceptions where we observed the gain of chromosomal segments that were absent in the respective parents. As the genome-wide genotypes and chromosomal breakpoints in progenies which had chromosomal segment gain are identical to parents, this helps to rule out possible seed mixture or DNA contamination. It is possible that this observation of non-Mendelian inheritance could be from individual primary DH plants that were chimeric for the deleted chromosome segment. Therefore, individual leaf samples for DNA (and digital karyotype) could be a complete deletion, while other parts of the plant were still hemizygous, with cell lines producing hemizygous germ cells. It is known that genomic instability can lead to sectored mosaic plants, particularly in the case of active transposons or other genomic instabilities (Frank & Chitwood, 2016). The mitotic cell division in meristematic tissue with a dicentric chromosome could result in homozygous deletion while other parts remain hemizygous. The phenomenon of anaphase bridge formation in dicentric chromosome leading to chromosomal breakage has been described in maize with breakage–fusion–bridge cycles

in subsequent nuclear divisions resulting in various chromosomal aberrations including homozygous duplication and deficiency (McClintock, 1942).

In StanleyLandmarkDH population, severe forms of aberrations were observed with uninterpretable chromosomal karyotypes having multiple duplications and deletions. Regardless, the observed aberrations clearly originated from multiple chromosome breakage events within the same chromosome. This could be a result of initial break–fusion–bridge cycles within the primary DH from unstable chromosomes without telomeres, or multiple translocation events within the primary DH. At the same time, however, the homozygosity of allelic variants within the DH plants, supports that these all originated from haploid plants and the induced genome shock from haploid induction and then chromosome doubling.

While we have identified many cases of chromosomal abnormalities, the mechanism whereby these abnormalities are generated could be further investigated. It is known that spindle fiber arrestment is often incomplete, and therefore attachment of fibers to a monosomic chromosome during the haploid doubling will induce chromosome breakage or movement to a pole and chromosome loss (Grant, 1978; Sharma, 1990). Furthermore, the presence of these aberrations needs to be evaluated in other species where DH are regularly used. Owing to the polyploid nature of wheat and strong genome buffering capacity, the prevalence of the large chromosomal aberrations might be much higher in wheat or tolerated better than in diploid species where abnormal karyotypes would be more lethal. Regardless of the mechanism, however, the prevalence of significant chromosomal aberrations in DH plants has important implications for genetics studies and plant breeding and warrants careful consideration both in the design and implementation of DH breeding strategies.

4 | MATERIALS AND METHODS

4.1 | Development of DH lines

The two DH populations used in this study were developed using the common method of wheat–maize wide hybridization as described previously (Knox et al., 2000; Santra et al., 2017) and summarized as shown in Figure 7. The first population was derived from the crossing of two Canadian hard red spring wheat cultivars CDC Stanley and CDC Landmark developed by Crop Development Centre, University of Saskatchewan, and was designated as StanleyLandmarkDH population. For the purpose of this study, it was important to differentiate the primary DH, as well as first- and second-generation following DH production, and individual plants within each of these generations. The primary DH plant was not sampled, but designated by line number (e.g., StanleyLandmarkDH01002). The primary DH plants were self-pollinated to produce seeds of the $DH_{0:1}$ generation (designated as, e.g., StanleyLandmarkDH01002-0), which were tissue sampled as individual plants and genotyped with skim sequencing. To further validate the presence of aneuploidy and deletions and confirm the expected segregation within DH lines, the $DH_{0:1}$ plants were selfed to produce $DH_{1:2}$ generation designated with two generations within the line pedigree numbers (e.g., StanleyLandmarkDH01002-0-01).

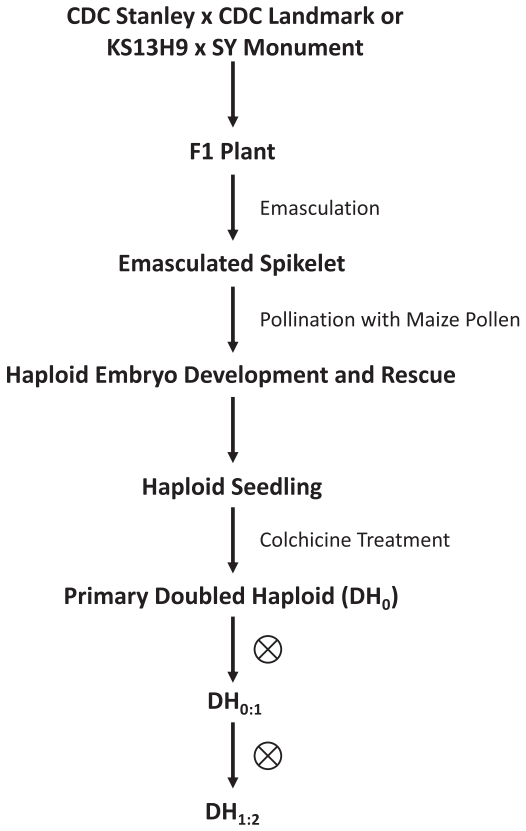

FIGURE 7 Schematic representation of DH development process. CDC Stanley×CDC Landmark and KS13H9×SYMonument populations were designated as StanleyLandmarkDH and KS13H9×SYMonumentDH, respectively.

In total, StanleyLandmarkDH population included 145 primary DH lines composed of 396 individual $DH_{0:1}$ plants and 383 $DH_{1:2}$ plants. The primary DH plants had limited seed. The selfed seed from primary DH plants was recovered and germinated to produce $DH_{0:1}$ plants ranging from one to six plants per line (Table S7) with 86 lines having two or more $DH_{0:1}$. There were 47 lines that were advanced via single seed descent to the $DH_{1:2}$ generation with at least one to a maximum of 13 plants (Table S8).

Similarly, two winter wheat cultivars were crossed, KS13H9×SYMonument, and DH lines were developed by the KSU wheat breeding program as described earlier, which is hereafter designated as the KS13H9×SYMonumentDH population. In total, the KS13H9×SYMonumentDH population

had 38 primary DH plants and the numbers were designated as, for example, KS13H9SYMonumentDH0001. As before, the primary DH plants were not sampled and the self-pollinated seed was generated to form the $DH_{0:1}$ generation from which a total of 187 plants were sampled (designated as, e.g., KS13H9SYMonumentDH0001-1) ranging from three to five plants per line (Table S9).

4.2 | DNA extraction and sequencing

For DNA extraction, about 2 cm of tissue was collected at young two-leaf stage, lyophilized for 2 days with Labconco FreeZone Plus 4.5 (Labconco Corporation, Kansas City, MO, USA), and then ground into a fine powder using Mixer Mill MM 400 (Retsch GmbH, Haan, Germany). The genomic DNA in a 96-well plate was extracted using BioSprint 96 DNA Plant Kit (QIAGEN, Hilden, Germany) following the manufacturer's recommendations.

To skim sequence the population, we optimized a low-volume high-throughput variation of the Illumina Nextera library preparation kit (Adhikari et al., 2022). This library preparation method can multiplex samples from multiple 96-well plates into a single library, which can be sequenced in a single lane and later demultiplex into individual samples based on dual-index barcodes. The quantified final libraries were used for paired-end sequencing in NovoSeq 6000 or HiSeqX10 (Illumina, Inc., San Diego, CA, USA).

From a previous study to develop the methods for skim sequencing, we sequenced 48 DH lines from the CDC Stanley×CDC Landmark DH population (Adhikari et al., 2022). For StanleyLandmarkDH population, remnant seed stock from the two original parents used for crossing was also skim sequenced with the DH population and designated as StanleyDH and LandmarkDH. In addition to skim sequencing of parents, the previously published paired-end 2×265 bp raw reads generated from 470 bp libraries used for de novo assembly of CDC Stanley and CDC Landmark were also included in this study for variant calling (Walkowiak et al., 2020). These were designated as Stanley470 and Landmark470 and accessed from NCBI SRA (archive no. PRJNA544491). The two parents from KS13H9×SYMonumentDH population were also sequenced at higher coverage. The WGS of KS13H9 and SYMonument was done using a PCR-free library with TruSeq DNA PCR-Free LT Sample Prep Kit (Illumina, Inc.) followed by paired-end sequencing in HiSeqX10 (Illumina, Inc.).

4.3 | Digital karyotyping

The sequence based digital karyotyping was performed by read mapping technique to decipher chromosomal dosage along the chromosome. The multiplexed skim sequencing of DH plants was received as a single fastq file, which was demultiplexed into individual samples based on the combination of i7 and i5 indexes used during library preparation. A custom Perl script was used in searching the paired barcode and assigning paired-end reads to the respective sample (Adhikari et al., 2022). The paired-end reads of each DH line from two populations were aligned to separate reference genomes using Hisat2 (Kim et al., 2019). The default alignment parameters were selected with the additional flags of *–no-spliced-alignment* and *–no-unal*. The output alignment files were sorted with SAMtools sort option (Li et al., 2009). To recover reads with high-quality mapping, any paired-ends reads with unexpected insert size deviated from sequence library fragment size and mapped to multiple locations on the genome were discarded by filtering for concordant (*YT:Z:CP*) and unique single site mapping (*NH:i:1*) based on SAM tags present in the alignment file, respectively.

The StanleyLandmarkDH population with a total of 783 genotyped DH plants including four parental sequences were aligned to both of the parental reference genomes. The complete assembled reference genomes of both parents are available but CDC Landmark reference genome (170831_Landmark_pseudomolecules_v1) was selected as the reference to align the reads from DH plants (Walkowiak et al., 2020). In case of KS13H9×SYMonumentDH population, none of the parents have the reference genome so the related US wheat cultivar Jagger (180529_Jagger_pseudomolecule_v1.1) assembled by 10+ Genome Project was used as a reference genome to align reads from 189 DH plants including two parents (Walkowiak et al., 2020).

The high-quality reads mapped across the genomes recovered after filtering were used to identify the distribution of mapped reads along the chromosomes. A custom script was used to count the number of raw reads per 1 Mb bin (14,454 and 14,564 bins in the CDC Landmark and Jagger reference genomes, respectively) and normalized the read count per bin along 21 chromosomes (https://github.com/sandeshsth/05_chromosomal_aberrations.git).

To compare relative read depth among DH lines, each 1 Mb bin within a sample was normalized to 1x coverage based on the total coverage for that respective sample.

First, the coverage of the sample was calculated as:

$$\text{sample coverage} = \frac{2 \times \text{no. of concordant unique paired} - \text{end reads mapped} \ \text{read length}}{\text{reference genome assembly}} \quad (1)$$

where the number of reads was the total number of read pairs, which passed the above filters and the read length was 150 bp except for the two parents which were 265 bp.

Second, the coverage in each 1 Mb bin was calculated as:

$$1\,\text{Mbbincoverage} = \frac{\text{no. of readsin 1 Mbbin} \times \text{readlength}}{1e6} \quad (2)$$

The expectation is that the coverage in each 1 Mb bin will be equal to the coverage of the corresponding sample, that is,

$$\frac{1\,\text{Mb bin coverage}}{\text{sample coverage}} = 1 \quad (3)$$

Any deviation from the expected 1× coverage was used in the identification of loss or gain of chromosomal segments by digital karyotyping.

4.4 | Variant calling

In both populations, the filtered alignment files with concordant unique reads were used to discover SNPs with BCFtools version 1.10.2 with the following parameters: *bcftools mpileup –annotate AD,DP,INFO/AD –skip-indels -f reference.fa -b bam_list.txt -B | bcftools call -m –variants-only –skip-variants indels –output-type v -o out.vcf –group-samples*—(Li, 2011). To expedite the variant calling pipeline in both populations, individual chromosomes were processed to call SNPs, filtered, and later merged together.

For StanleyLandmarkDH population, each SNP location was filtered for the following criteria: raw total read depth of at least 30 and alternate allele depth of at least 3, SNPs having homozygous differences between Landmark470 and Stanley470, minor allele frequency of at least 0.05, and remove SNPs where the Landmark470 allele does not match the reference allele. In the DH population lines, as they were sequenced at very low coverage, we expect only a few percentage of genotype calls present. Thus, we removed SNP positions with greater than 98% missing data. Similarly, after discovering SNPs in KS13H9×SYMonumentDH population, each SNP position was filtered and recovered when the following criteria were met: total read depth of at least 6 and at most of 40, alternate allele depth of at least 2, and alternate allele between the two parents, KS13H9 and SYMonument.

The SNPs identified in the two populations were used in typing parental alleles along the chromosomes in DH lines leading to combined analysis of parental chromosomes inherited (SNP allele) along with chromosome dosage (read count). The genomic segments along the chromosomes of DH lines exhibit either of two alleles inherited from two parents, which could be visualized as a virtual chromosome with recombination. For plotting the allele along the chromosomes, all missing and heterozygous positions in each sample were filtered individually. The read coverage in 1 Mb bins along with the allele type designation from two parents along the chromosome were plotted with ggplot2 in R (Wickham, 2016).

4.5 | Scoring chromosomal aberrations

The aberrations along the chromosomes were determined based on visual scoring of digital karyotypes. A normal karyotype will have two copies of each chromosome resulting in 1× normalized read coverage in 1 Mb bins along the chromosome. The deviation from normal 1× coverage will show chromosomal aberrations in the form of deletion, aneuploidy, or duplication/translocation. The deviation of 1 Mb bins to 0× or 0.5× in a specific region was used to indicate a homozygous or heterozygous (hemizygous) deletion, respectively. If a certain region had a deviation of 1.5× or 2× coverage, this indicated a copy number increase from a duplication/translocation with one or two additional copies, respectively. As the read depth is only an indication of copy number, higher read depth could indicate a tandem duplication or also a translocation to another chromosome as the translocated sequencing reads can map to the reference region from where it originated. Therefore, both terms were used together as the genomic data do not differentiate a duplication from a translocation. If the dosage deviation was observed for the complete chromosome then this indicated aneuploidy.

In the StanleyLandmarkDH population, the normalized graphs of parents (Landmark470, LandmarkDH, Stanley470, and StanleyDH) were considered during the scoring of aberrations in DH lines. Since the reads were mapped to the CDC Landmark reference genome, the variant regions in CDC Stanley compared with CDC Landmark may not have normalized to 1× coverage due to structural divergence between these genomes. Such regions with variations,

deletion, or duplications were considered and accounted scoring DH lines based on allele type (CDC Landmark or CDC Stanley) present along the chromosome. In the case of the KS13H9×SYMonumentDH population, the DH lines and two parents were mapped to the Jagger reference genome, so any regions that deviated from normal 1× coverage in KS13H9 and SYMonument were accounted while scoring aberration in the DH lines.

4.6 | Fluorescence in situ hybridization

The cytogenetic examination of DH plants was performed to validate the chromosomal aberrations identified by digital karyotyping. We used the FISH technique to examine the karyotype of DH lines, especially to identify the chromosomal abnormalities at subgenome level. Mitotic chromosomes were prepared as described by Koo & Jiang (2009) with minor modifications. Root tips were collected from plants and placed in a nitrous oxide gas chamber for 2 h. The root tips were fixed overnight in ethanol:glacial acetic acid (3:1) and then squashed in a drop of 45% acetic acid. All preparations were stored at −70°C until use. DNA probes of the pAs1 and GAA repeats (Danilova et al., 2017; Koo et al., 2015) were labeled with either digoxigenin-11-dUTP or biotin-16-dUTP (Roche, Indianapolis, IN) using a nick translation reaction. Individual chromosomes were identified using the pAs1 and GAA repeats and karyotypes as originally described by Mukai et al. (1993) and Cuadrado et al. (2000, 2008). Biotin- and digoxigenin-labeled probes were detected with Alexa Fluor 488 streptavidin antibody (Invitrogen, Grand Island, NY) and rhodamine-conjugated antidigoxigenin antibody (Roche), respectively. Chromosomes were counterstained with 4′,6-diamidino-2-phenylindole in Vectashield antifade solution (Vector Laboratories, Burlingame, CA). The images were captured with a Zeiss Axioplan 2 microscope (Carl Zeiss Microscopy LLC, Thornwood, NY) using a cooled CCD camera CoolSNAP HQ2 (Photometrics, Tucson, AZ) and AxioVision 4.8 software. The final contrast of the images was processed using Adobe Photoshop CS5 software (Adobe Inc., San Jose, CA). Using this methodology, representative normal and abnormal DH plants identified by digital karyotyping from both populations were selected for karyotype validation with FISH.

DATA AVAILABILITY STATEMENT

The sequence data are available in National Center for Biotechnology Information (NCBI) database as a BioProject accession PRJNA729723. The DH populations developed from CDC Stanley×CDC Landmark and KS13H9×SYMonument are deposited in sequence read archive (SRA) database with accession SRS8963504 and SRS9552497, respectively.

ACKNOWLEDGMENTS

This project was completed with support from the Kansas Wheat Commission and the Kansas Wheat Alliance, King Abdullah Univesity of Science and Technology, Feed the Future Innovation Lab for Applied Wheat Genomics through the U.S. Agency for International Development under the terms of Contract No AID-OAA-A-13-00051, and the Canadian Triticum Applied Genomics research project (CTAG2) funded by Genome Canada, Western Grain Research Foundation, Government of Saskatchewan, Saskatchewan Wheat Development Commission, Alberta Wheat Commission, Manitoba Crop Alliance, DuPont Pioneer, Viterra, and Secan.

AUTHOR CONTRIBUTION

Sandesh Shrestha: Methodology, software, validation, formal analysis investigation, data curation, visualization, writing—original draft, writing—review and editing; **Dal-Hoe Koo**: Conceptualization, methodology, validation, investigation, visualization. writing—review and editing; **Byron Evers**: Validation, investigation, resources, data curation, writing—review and editing; **Shuangye Wu**: Validation, investigation, writing—review and editing; **Sean Walkowiak**: Methodology, validation, resources, writing—review and editing; **Pierre Hucl**: Methodology, Resources, writing—review and editing, funding acquisition; **Curtis Pozniak**: Conceptualization, validation, resources, writing—review and editing, funding acquisition; **Allan Fritz**: Conceptualization, methodology, validation, investigation, resources, writing—review and editing, funding acquisition; **Jesse Poland**: Conceptualization, methodology, software, resources, data curation, visualization, writing—original draft, writing—review and editing, supervision, project administration, funding acquisition.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

Sandesh Shrestha https://orcid.org/0000-0001-5527-8999
Dal-Hoe Koo https://orcid.org/0000-0003-4578-8519
Byron Evers https://orcid.org/0000-0003-1840-5842
Sean Walkowiak https://orcid.org/0000-0002-3866-5038
Curtis Pozniak https://orcid.org/0000-0002-7536-3856
Allan Fritz https://orcid.org/0000-0003-2574-8675
Jesse Poland https://orcid.org/0000-0002-7856-1399

REFERENCES

Adhikari, L., Shrestha, S., Wu, S., Crain, J., Gao, L., Evers, B., Wilson, D., Ju, Y., Koo, D. -H., Hucl, P., Pozniak, C., Walkowiak, S., Wang, X., Wu, J., Glaubitz, J. C., DeHaan, L., Friebe, B., & Poland, J. (2022). A high-throughput skim-sequencing approach for genotyping, dosage

estimation and identifying translocations. *Scientific Reports*, *12*(1), 17583. https://doi.org/10.1038/s41598-022-19858-2

Allen, A. M., Winfield, M. O., Burridge, A. J., Downie, R. C., Benbow, H. R., Barker, G. L., Wilkinson, P. A., Coghill, J., Waterfall, C., & Davassi, A. (2017). Characterization of a wheat breeders' array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (*Triticum aestivum*). *Plant Biotechnology Journal*, *15*(3), 390–401. https://doi.org/10.1111/pbi.12635

Barclay, I. (1975). High frequencies of haploid production in wheat (*Triticum aestivum*) by chromosome elimination. *Nature*, *256*(5516), 410–411. https://doi.org/10.1038/256410a0

Berg, J., Stougaard, R., Cook, C., Kephart, K., Carlson, G., Wichman, D., Eckhoff, J., Riveland, N., Kushnak, G., & Nash, D. (2006). Registration of 'MT1159CL' wheat. *Crop Science*, *46*(3), 1395–1397. https://doi.org/10.2135/cropsci2005.11-0405

Castillo, A., Cistué, L., Vallés, M., & Soriano, M. (2009). Chromosome doubling in monocots. *Advances in haploid production in higher plants*. (pp. 329–338) Springer:.

CGIAR (2020). Wheat in the World. https://wheat.org/wheat-in-the-world/

Costa, J., Corey, A., Hayes, P., Jobet, C., Kleinhofs, A., Kopisch-Obusch, A., Kramer, S., Kudrna, D., Li, M., & Riera-Lizarazu, O. (2001). Molecular mapping of the Oregon Wolfe barleys: A phenotypically polymorphic doubled-haploid population. *Theoretical and Applied Genetics*, *103*(2-3), 415–424. https://doi.org/10.1007/s001220100622

Cuadrado, A., Cardoso, M., & Jouve, N. (2008). Increasing the physical markers of wheat chromosomes using SSRs as FISH probes. *Genome*, *51*(10), 809–815. https://doi.org/10.1139/G08-065

Cuadrado, A., Schwarzacher, T., & Jouve, N. (2000). Identification of different chromatin classes in wheat using *in situ* hybridization with simple sequence repeat oligonucleotides. *Theoretical and Applied Genetics*, *101*(5), 711–717. https://doi.org/10.1007/s001220051535

Danilova, T. V., Friebe, B., & Gill, B. S. (2012). Single-copy gene fluorescence *in situ* hybridization and genome analysis: Acc-2 loci mark evolutionary chromosomal rearrangements in wheat. *Chromosoma*, *121*(6), 597–611. https://doi.org/10.1007/s00412-012-0384-7

Danilova, T. V., Zhang, G., Liu, W., Friebe, B., & 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. *Theoretical and Applied Genetics*, *130*(3), 549–556. https://doi.org/10.1007/s00122-016-2834-8

DePauw, R., Townley-Smith, T., Humphreys, G., Knox, R., Clarke, F., & Clarke, J. (2005). Lillian hard red spring wheat. *Canadian Journal of Plant Science*, *85*(2), 397–401. https://doi.org/10.4141/P04-137

Falk, D., & Kasha, K. (1981). Comparison of the crossability of rye (*Secale cereale*) and *hordeum bulbosum* onto wheat (*Triticum aestivum*). *Canadian Journal of Genetics and Cytology*, *23*(1), 81–88. https://doi.org/10.1139/g81-010

Fehr, W., "Principles of Cultivar Development: Theory and Technique" (1991). Agronomy Books. 1. https://lib.dr.iastate.edu/agron_books/1

Frank, M. H., & Chitwood, D. H. (2016). Plant chimeras: The good, the bad, and the 'Bizzaria'. *Developmental Biology*, *419*(1), 41–53. https://doi.org/10.1016/j.ydbio.2016.07.003

Gao, L., Koo, D. -H., Juliana, P., Rife, T., Singh, D., da Silva, C. L., Lux, T., Dorn, K. M., Clinesmith, M., & Silva, P. (2020). The *Aegilops ventricosa* 2N^{v}S segment in bread wheat: Cytology, genomics and breeding. *Theoretical and Applied Genetics*, 1–14.

Gill, B. S., & Friebe, B. (1998). Plant cytogenetics at the dawn of the 21st century. *Current Opinion in Plant Biology*, *1*(2), 109–115. https://doi.org/10.1016/S1369-5266(98)80011-3

Graf, R., Hucl, P., Orshinsky, B., & Kartha, K. (2003). McKenzie hard red spring wheat. *Canadian Journal of Plant Science*, *83*(3), 565–569. https://doi.org/10.4141/P02-115

Grant, W. F. (1978). Chromosome aberrations in plants as a monitoring system. *Environmental Health Perspectives*, *27*, 37–43. https://doi.org/10.1289/ehp.782737

Hooghvorst, I., & Nogués, S. (2020). Chromosome doubling methods in doubled haploid and haploid inducer-mediated genome-editing systems in major crops. *Plant Cell Reports*, 1–16.

Jauhar, P. P., Xu, S. S., & Baenziger, P. S. (2009). Haploidy in cultivated wheats: Induction and utility in basic and applied research. *Crop Science*, *49*(3), 737–755. https://doi.org/10.2135/cropsci2008.08.0462

Kasha, K., & Kao, K. (1970). High frequency haploid production in barley (*Hordeum vulgare* l.). *Nature*, *225*(5235), 874–876. https://doi.org/10.1038/225874a0

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kim, K. -M., Baenziger, P. S., Rybczynski, J. J., & Arumuganathan, K. (2003). Characterization of ploidy levels of wheat microspore-derived plants using laser flow cytometry. *In Vitro Cellular & Developmental Biology-Plant*, *39*(6), 663–668.

Knox, R., Clarke, J., & DePauw, R. (2000). Dicamba and growth condition effects on doubled haploid production in durum wheat crossed with maize. *Plant Breeding*, *119*(4), 289–298. https://doi.org/10.1046/j.1439-0523.2000.00498.x

Koo, D. H., & Jiang, J. (2009). Super-stretched pachytene chromosomes for fluorescence *in situ* hybridization mapping and immunodetection of DNA methylation. *The Plant Journal*, *59*(3), 509–516. https://doi.org/10.1111/j.1365-313X.2009.03881.x

Koo, D. -H., Sehgal, S. K., Friebe, B., & Gill, B. S. (2015). Structure and stability of telocentric chromosomes in wheat. *PLoS ONE*, *10*(9), e0137747. https://doi.org/10.1371/journal.pone.0137747

Laurie, D., & Bennett, M. (1986). Wheat × maize hybridization. *Canadian Journal of Genetics and Cytology*, *28*(2), 313–316. https://doi.org/10.1139/g86-046

Laurie, D., & Bennett, M. (1987). The effect of the crossability loci *Kr1* and *Kr2* on fertilization frequency in hexaploid wheat x maize crosses. *Theoretical and Applied Genetics*, *73*(3), 403–409. https://doi.org/10.1007/BF00262508

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Mason, A. S., Higgins, E. E., Snowdon, R. J., Batley, J., Stein, A., Werner, C., & Parkin, I. A. (2017). A user guide to the *brassica* 60K illumina infinium™ SNP genotyping array. *Theoretical and Applied Genetics*, *130*(4), 621–633. https://doi.org/10.1007/s00122-016-2849-1

McClintock, B. (1942). The fusion of broken ends of chromosomes following nuclear fusion. *Proceedings of the National Academy of Sciences*, *28*(11), 458. https://doi.org/10.1073/pnas.28.11.458

Mukai, Y., Nakahara, Y., & Yamamoto, M. (1993). Simultaneous discrimination of the three genomes in hexaploid wheat by multicolor fluorescence in situ hybridization using total genomic and highly repeated DNA probes. *Genome*, *36*(3), 489–494. https://doi.org/10.1139/g93-067

Murovec, J., & Bohanec, B. (2011). Haploids and doubled haploids in plant breeding. Plant Breeding, Dr. I., Abdurakhmonov, (Ed.).–2012.–P:87–106.

Niu, Z., Jiang, A., Abu Hammad, W., Oladzadabbasabadi, A., Xu, S. S., Mergoum, M., & Elias, E. M. (2014). Review of doubled haploid production in durum and common wheat through wheat × maize hybridization. *Plant Breeding*, *133*(3), 313–320. https://doi.org/10.1111/pbr.12162

Pfosser, M., Amon, A., Lafferty, J., Heberle-Bors, E., & Lelley, T. (1995). Gain or loss of single chromosomes in wheat-rye addition lines and in 6x triticale detected by flow cytometry. *Plant Breeding*, *114*(6), 555–557. https://doi.org/10.1111/j.1439-0523.1995.tb00858.x

Ray, D. K., Mueller, N. D., West, P. C., & Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PLoS ONE*, *8*(6), e66428. https://doi.org/10.1371/journal.pone.0066428

Ren, J., Wu, P., Trampe, B., Tian, X., Lübberstedt, T., & Chen, S. (2017). Novel technologies in doubled haploid line development. *Plant Biotechnology Journal*, *15*(11), 1361–1370. https://doi.org/10.1111/pbi.12805

Sandhu, S., Dhesi, J., Gill, B., & Svendsgaard, D. (1991). Evaluation of 10 chemicals for aneuploidy induction in the hexaploid wheat assay. *Mutagenesis*, *6*(5), 369–373. https://doi.org/10.1093/mutage/6.5.369

Santra, M., Wang, H., Seifert, S., & Haley, S. (2017). Doubled haploid laboratory protocol for wheat using wheat–maize wide hybridization. *Wheat biotechnology*. (pp. 235–249) Springer.

Schmid, T. E., Xu, W., & Adler, I. -D. (1999). Detection of aneuploidy by multicolor FISH in mouse sperm after *in vivo* treatment with acrylamide, colchicine, diazepam or thiabendazole. *Mutagenesis*, *14*(2), 173–179. https://doi.org/10.1093/mutage/14.2.173

Sharma, C. (1990). Chemically induced aneuploidy in higher plants. *Mutagenesis*, *5*(2), 105–126. https://doi.org/10.1093/mutage/5.2.105

Snape, J., Chapman, V., Moss, J., Blanchard, C., & Miller, T. (1979). The crossabilities of wheat varieties with *Hordeum bulbosum*. *Heredity*, *42*(3), 291–298. https://doi.org/10.1038/hdy.1979.32

Stober, A., & Hessu, D. (1997). Spike pretreatments, anther culture conditions, and anther culture response of 17 German varieties of spring wheat (*Triticum aestivum* l.). *Plant Breeding*, *116*(5), 443–447. https://doi.org/10.1111/j.1439-0523.1997.tb01028.x

Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M. T., Brinton, J., Ramirez-Gonzalez, R. H., Kolodziej, M. C., Delorean, E., Thambugala, D., Klymiuk, V., Byrns, B., Gundlach, H., Bandi, V., Siri, J. N., Nilsen, K., Aquino, C., Himmelbach, A., Copetti, D., … Pozniak, C. J. (2020). Multiple wheat genomes reveal global variation in modern breeding. *Nature*, *588*(7837), 277–283. https://doi.org/10.1038/s41586-020-2961-x

Wang, T. -L., Maierhofer, C., Speicher, M. R., Lengauer, C., Vogelstein, B., Kinzler, K. W., & Velculescu, V. E. (2002). Digital karyotyping. *Proceedings of the National Academy of Sciences*, *99*(25), 16156–16161. https://doi.org/10.1073/pnas.202610899

Watson, A., Ghosh, S., Williams, M. J., Cuddy, W. S., Simmonds, J., Rey, M. -D., Asyraf Md Hatta, M., Hinchliffe, A., Steed, A., Reynolds, D., Adamski, N. M., Breakspear, A., Korolev, A., Rayner, T., Dixon, L. E., Riaz, A., Martin, W., Ryan, M., Edwards, D., … & Hickey, L. T. (2018). Speed breeding is a powerful tool to accelerate crop research and breeding. *Nature Plants*, *4*(1), 23–29. https://doi.org/10.1038/s41477-017-0083-8

Wędzony, M., Forster, B., Żur, I., Golemiec, E., Szechyńska-Hebda, M., Dubas, E., & Gotębiowska, G. (2009). Progress in doubled haploid technology in higher plants. *Advances in haploid production in higher plants*. (pp. 1–33) Springer.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag. https://ggplot2.tidyverse.org

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Shrestha, S., Koo, D.-H., Evers, B., Wu, S., Walkowiak, S., Hucl, P., Pozniak, C., Fritz, A., & Poland, J. (2023). Wheat doubled haploids have a marked prevalence of chromosomal aberrations. *The Plant Genome*, *16,* e20309. https://doi.org/10.1002/tpg2.20309