

## ORIGINAL RESEARCH

# Selection signatures in the CIMMYT International Elite Spring and Semi-arid Wheat Yield Trials

Alexandre Mondaini<sup>1,2</sup> | Umesh Rosyara<sup>3,4</sup> | Deepmala Sehgal<sup>3</sup> |  
Susanne Dreisigacker<sup>3</sup> 

<sup>1</sup> Institute of Plant Breeding, Seed Science and Population Genetics, Univ. of Hohenheim, Fruhwirthstrasse 21, Stuttgart 70593, Germany

<sup>2</sup> current address: Univ. of Hong Kong, 21 Sassoon Rd, Pok Fu Lam, Hong Kong

<sup>3</sup> International Maize and Wheat Improvement Center (CIMMYT), Global Wheat Program, Km45 Carretera Mexico-Veracruz, Texcoco, Edo. de México 56237, México

<sup>4</sup> current address: BASF, 26 Davis Dr., Research Triangle, NC 27709, USA

## Correspondence

Susanne Dreisigacker, International Maize and Wheat Improvement Center (CIMMYT), Global Wheat Program, Km. 45 Carretera Mexico-Veracruz, 56237 Texcoco, Edo. de México, México.

Email: [s.dreisigacker@cgiar.org](mailto:s.dreisigacker@cgiar.org)

Assigned to Associate Editor Xianran Li.

## Funding information

Bill and Melinda Gates Foundation, Grant/Award Number: OPP101638; United States Agency for International Development, Grant/Award Number: AIDOAA-A-13-00051; CGIAR Research Program on Wheat

## Abstract

The International Maize and Wheat Improvement Center (CIMMYT) annually distributes advanced wheat (*Triticum aestivum* L.) breeding lines to collaborators worldwide through the International Wheat Improvement Network. Lines are disseminated through international nurseries, including the Elite Spring Wheat Yield Trial (ESWYT) targeted to optimal (irrigated and high production) wheat production areas and the Semi-arid Wheat Yield Trial (SAWYT) targeted to low rainfall environments. A total of 2,184 wheat lines that formed the ESWYT and SAWYT since 1979 and 1992, respectively, were genotyped using genotyping-by-sequencing to explore trends of genetic diversity and selection footprints associated with continuous crop improvement and adaptation. Due to a small population size of each trial, adjacent year trials were pooled into subpopulations. Population structure was evaluated using discriminant analysis of principal components and fixation index. High levels of admixture within and across the ESWYT and SAWYT subpopulations were revealed, indicating that the entire genetic diversity in the overall CIMMYT germplasm pool is harnessed to target core traits to individual mega-environments. Genome wide scans of deviations of minor allele frequencies at each marker identified large linkage blocks in several chromosomes. The scans also revealed that 9.8 and 2.0% of the SNP markers could be associated to selection signatures over time and to environmental adaptation (significant deviations between ESWYT and SAWYT), respectively. Several known genes and previously identified haplotypes associated with grain yield in more recent CIMMYT elite germplasm did fall into genomic regions with directional selection.

**Abbreviations:** BIC, Bayesian information criteria; CIMMYT, International Maize and Wheat Improvement Center; DAPC, discriminant analysis of principal components; DArT, diversity array technology; ESWYT, Elite Spring Wheat Yield Trial;  $F_{ST}$ , fixation index; GBS, genotype-by-sequencing; GWAS, genome wide association studies; MAF, minor allele frequency; ME, mega-environments; PC, principal component; QTL, quantitative trait loci; SAWYT, Semi-arid Wheat Yield Trial; SNP, single nucleotide polymorphism.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *The Plant Genome* published by Wiley Periodicals LLC on behalf of Crop Science Society of America

## 1 | INTRODUCTION

Wheat (*Triticum aestivum* L.) is the most widely grown food crop in the world and the third most important in terms of global production after maize (*Zea mays* L.) and rice (*Oryza sativa* L.). Wheat is of fundamental importance in the human diet by accounting for about 20% energy and protein intake globally (FAO, 2016). Wheat is also a global commodity, being the most exported crop, countries/regions with the largest wheat production do not always correspond with those that have the highest wheat consumption (FAO, 2016). While the rate of world population growth is in general slowing down, it will still reach 8.6 billion people in 2030 and 9.8 billion by 2050 (van Bavel, 2013). Thus, meeting the global demand for wheat will require a substantial increase of grain yield production per unit area, which is currently around 3.3 t ha<sup>-1</sup> (FAO, 2017). Achieving this goal will be challenging as changing climate constraints pose risks not only to the harvested product directly but also hinder wheat growing area expansion.

The International Maize and Wheat Improvement Center (CIMMYT) annually distributes specific nurseries and yield trials as part of a system known as the International Wheat Improvement Network. Nurseries and trials are grown under local conditions across the developing world and the best candidates are selected for direct cultivar release or as parents for new locally made crosses (Baum et al., 2015; Reynolds et al., 2017). The data gathered by collaborators are catalogued, analyzed, and made available to the global wheat research community. Internationally recorded performance data are also crucial to select new parents for subsequent crosses and breeding. Among the annually distributed yield trials, the two most requested are the Elite Spring Wheat Yield Trial (ESWYT) for optimal (irrigated and high production) and the Semi-arid Wheat Yield Trial (SAWYT) for low rainfall environments. Both trials target 45 million ha of wheat production area globally. The ESWYT started in 1979 and the SAWYT in 1992, thus having almost three decades of shared history.

Although studies on the phenotypic performance on the CIMMYT yield trials have been significant (Byth et al., 1976; L. A. Crespo-Herrera et al., 2018; Leonardo A. Crespo-Herrera et al., 2017; Crossa et al., 2007; DeLacy et al., 1993; Mondal et al., 2020), only a few genomic studies have been performed on these trials to date (Crossa et al., 2007; Dawson et al., 2013; Susanne Dreisigacker et al., 2012).

Signatures of selection are reflected by loci or regions in the genome that undergo changes through reduction and an increase or elimination of genetic variation due to natural or artificial selection. Such regions are changed due to their direct functional relevance or indirectly through their association with causative variants. These regions can shed light on evolutionary adaptation or breeding practices, by identi-

### Core Ideas

- Genomic changes may be tracked across internationally distributed CIMMYT wheat lines.
- Structure analysis revealed admixture among elite lines targeted to different environments.
- Genome scans identified large linkage blocks in several chromosomes.
- A higher percentage of selection signatures resulted from selection over time than over environments.

fying changes of allele frequencies over a period that most likely involved relevant genes related to important adaptive and commercial traits (Afzal et al., 2019; Appels et al., 2018; Browning & Browning, 2016; Cavanagh et al., 2013; Laland et al., 2010; Neale & Kremer, 2011). Till date signatures of selection have not been studied in the CIMMYT wheat breeding germplasm. The objectives of this study were therefore to (a) unravel the genetic structures of the lines in the ESWYT and SAWYT yield trials genotyped with genotyping-by-sequencing (GBS) and (b) identify regions under selection on a genome-wide basis.

## 2 | MATERIALS AND METHODS

### 2.1 | Plant materials

We considered spring bread wheat lines included in the two historical international yield trials, ESWYT and SAWYT. In total, 2,184 lines were analyzed; 1,229 lines from ESWYT and 955 lines from SAWYT (Table 1, Supplementary Table S1).

A total of 514 lines were replicated over time, with 28 common lines between the two trials. Each individual trial consisted of about 30–50 lines. The ESWYT was first distributed in 1979 and has been continued ever since, while the SAWYT started in year 1992. Hence, since 1992 both ESWYT and SAWYT are being disseminated in parallel but are targeted to two distinct mega-environments (ME) of which 12 were previously defined by CIMMYT (Braun et al., 1996). The ESWYT lines are targeted to ME1 defined as optimal (irrigated and high production) environments. The ME1 falls into winter temperate climatic areas with 36 million ha of low rainfall in Asia, Africa, and Mexico and has late summer heat stress. These areas are optimally irrigated and may suffer from specific diseases like rust and kernel bunt (*Tilletia indica*). The SAWYT lines are targeted to ME4, low rainfall environments. The ME4 are nonirrigated low rainfall areas that

**TABLE 1** Summary of wheat lines included in two International Maize and Wheat Improvement Center international yield trials, the Elite Spring Wheat Yield Trial (ESWYT) and the Semi-arid Wheat Yield Trial (SAWYT)

International yield trial	Time period in years	Number of trials	Average number of lines per trial	Replicated lines within each trial
ESWYT	1979–2016	38	32	323
SAWYT	1992–2016	25	40	191
Total	1979–2016	63	36	514
<b>Predefined subpopulations</b>				
ESWYT 1–7	1979–1985	7	28	58
ESWYT 8–13	1986–1991	6	27	69
ESWYT 14–25	1992–2003	12	46	129
ESWYT 26–38	2004–2016	13	49	74
SAWYT 1–12	1992–2003	12	44	105
SAWYT 13–25	2004–2016	13	47	79

comprises globally of 12 million ha, which predominantly suffer from water stress. It is subdivided into three major sub-ME: winter rain regions followed by late Mediterranean-type drought (e.g., in Syria, Aleppo), early winter drought followed by late summer rain (e.g., in Marcos Juarez, Argentina) and regions that crop growth is highly dependent on soil-stored moisture after monsoon rains (e.g., Dharwad, India) (Rajaram, 2002).

## 2.2 | Genotyping

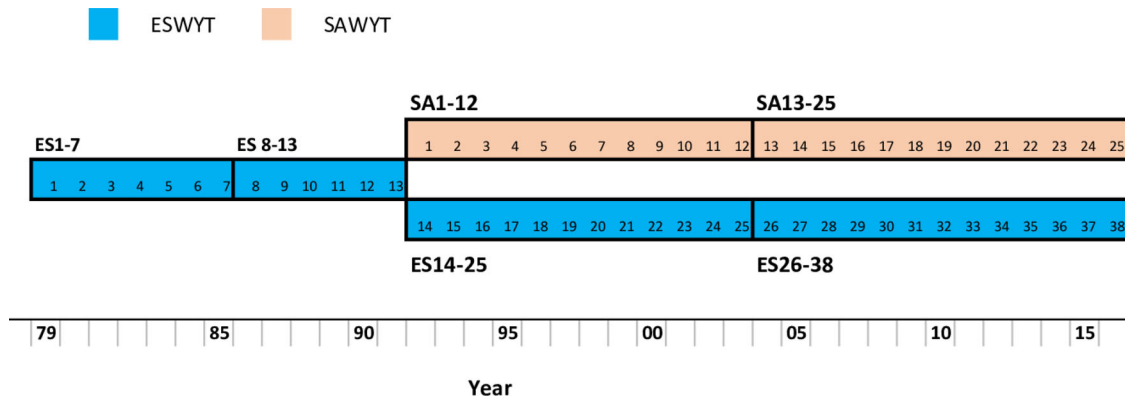
All 2,184 lines were genotyped using GBS (Elshire et al., 2011; Poland et al., 2012) to capture genome wide markers. Genotyping was performed on an Illumina HiSeq 2500 (Illumina Inc.) with each lane pooling 190 samples. Read length consisted initially of 100 bp, but after trimming for barcode removal, 64 bp sequence tags were derived. Single nucleotide polymorphisms (SNPs) were called using TASSEL 5 GBSv2 pipeline (Glaubitz et al., 2014) and anchored to the International Wheat Genome Sequencing Consortium reference sequence (RefSeq version 1.0). A total of 106,511 SNPs were reported, and three different filtering methods implemented by successive steps were applied at each SNP locus: (a) inbreeding coefficient  $\geq 80\%$ , (b) Fisher's exact Test ( $p < .001$ ) and (c) chi-square (expect inbreeding of 96%). Single nucleotide polymorphisms that passed at least one of the filters were recovered with a total of 82,697 SNPs retained. SNPs passing one (a), two (a and b), or three (a, b, and c) filters were 17,603, 34,998, and 30,096, respectively. Missing marker data was imputed with Beagle version 4.1 (Browning & Browning, 2016) by performing linear interpolation of ungenotyped variants. Markers with greater than 80% missing values and less than 5% minor allele frequency (MAF) were also removed.

## 2.3 | Subpopulations

As each individual yield trial was small with only 30–50 individuals, we pooled adjacent year trials into overall six subpopulations (Figure 1). From 1979 to 1991 only ESWYT was disseminated. We divided this time-period into two subpopulations from year 1979 to 1985 (ES1–ES7) and from year 1986 to 1991 (ES8–ES13). Starting in 1992 both yield trials, ESWYT and SAWYT, were distributed in parallel. Similarly, we divided each yield trial distributed across this time-period (1992–2016) into two subpopulations, from 1992 to 2003 (ES14–25, SA1–12) and from 2003 to 2016 (ES26–38, SA13–25). The two subpopulations of the initial germplasm pool of the ESWYT from 1979 to 1991 comprised a shorter year span than the later subpopulations (6 and 12 yr), to explore their relationship to the SAWYT, which expected to derive from the last years of the ESWYT (ES8–ES13).

## 2.4 | Population structure

Population structure analysis was performed with R package 'adegenet' v.2.1.1 (Jombart & Ahmed, 2011), using discriminant analysis of principal components (DAPC). Discriminant analysis of principal components requires prior groups/populations to be defined, thus we implemented two approaches to look for these groups/populations: (a) clustering by unsupervised learning k-means algorithm (Hartigan & Wong, 1979) and (b) supervised learning with predefined six subpopulations described in the previous section. When estimating clusters by k-means, the maximum number of discoverable clusters was set to  $k = 40$ , with the number of randomly chosen initial centroids in each run set to  $n = 25$  to better provide algorithmic convergence. The result of all sequential k clusters was then compared with the Bayesian information criteria (BIC) (Supplementary Figure S1), which should be



**FIGURE 1** Timeline of predefined subpopulations. ESWYT, Elite Spring Wheat Yield Trial; SAWYT, Semi-arid Wheat Yield Trial

lowest for the number of  $k$  or having the biggest decrease in BIC, thus the final number of clusters was set to  $k = 6$ . When visually comparing plots generated by both approaches, labeling populations a priori yielded better results of population differentiation, hence our preferred method. Discriminant analysis of principal components, as with principal component analysis, is a method of dimensionality reduction but in contrast focuses on optimizing the between group variance, whereas principal component analysis looks at the global variance often overlooking differences between groups. It accomplishes that by seeking discriminant functions (synthetic variables) that are linear combinations of the original variables (alleles). Discriminant analysis of principal components first converts the raw genotypic data into a standard principal component (PC) matrix, requests a prior number of principal components to be retained, and then creates the discriminant functions on the retained PCs. However, a trade-off must be taken when choosing the number of PCs to be retained. While too few PCs would cause a reduction on the statistical power of discrimination between groups, too many can cause over-fitting. Thus, we computed an alpha-score optimization statistic that represents the proportion of successful reassignments of observed discriminant functions to values obtained by randomizing groups, in this manner we could determine the best number of PCs to be retained which was set 216.

## 2.5 | Fixation index

To further explore population differentiation, we defined the fixation index ( $F_{ST}$ ) for the two yield trials.  $F_{ST}$  is one of the most widely used descriptive statistics in population and evolutionary genetics (Holsinger & Weir, 2009). It provides a measure of genetic differentiation and is directly related to the variance of allele frequency among populations and, conversely, to the amount of similarity among individuals within populations.  $F_{ST}$  values were calculated accord-

ing to (Nei, 1988) using the R package ‘hierfstat’ v.0.04-22 (Goudet, 2005).

For the  $F_{ST}$  comparisons, we further divided the ESWYT and SAWYT subpopulations into 3-to-4-yr groups from its primary until its most recent trials (e.g., ES1–4, ES5–7, ES8–10, ES11–13, ES14–16, SA1–3, etc.). This was done to observe environmental and temporal trends within the subpopulations under study. We analyzed  $F_{ST}$  values in three ways:

1. by calculating  $F_{ST}$  relative to the primary four-year group (ES1-4) to obtain genetic differentiation over time (e.g., between ES1–4 and ES5–7, ES1–4 and ES8–10, ES1–4 and ES11–13, etc.);
2. by calculating  $F_{ST}$  between temporally consecutive groups to derive potential genetic shifts over a shorter time period (e.g., between ES1–4 and ES5–7, ES5–7 and ES8–10, SA1–3 and SA4–6, etc.); and
3. by calculating  $F_{ST}$  between parallel groups of ESWYT and SAWYT trials to explore the trends of genetic differentiation between the two yield trials (e.g., SA1–3 and ES14–16).

## 2.6 | Signatures of Selection

Minor allele frequency was computed for each SNP marker in the six subpopulations using the R package ‘hierfstat’ v.0.04-22 (Goudet, 2005). The MAF values were plotted against physical marker position for each chromosome using R package lattice v.0.20-38 (Sarkar, 2008).

Allelic frequency changes in temporal dimension were calculated as deviations of MAF from the primary subpopulation (ES1–7) to all others. A SNP marker was declared undergoing significant directional selection by applying a threshold of 1% quantile from both tails of the distribution of MAF deviation values. Markers under selection derived by environmental adaptation (SNP markers selected in different directions in

the ESWYT and SAWYT) were explored by the deviation of MAF of each SNP from the two most recent subpopulations of each different yield trial (ES26–38; SA13–25) and applying the same percent quantile threshold. To further explore our results, we plotted several genes known to have historical relevance in CIMMYT germplasm into the genome-wide scans and exploited the shifts in allele frequency of recently identified quantitative trait loci (QTL) consistently associated with grain yield (Sehgal et al., 2020). The QTL were detected based on a large haplotype-based genome wide association study (GWAS) that included more than 6,000 CIMMYT advanced lines, evaluated in elite yield trials across five environments, in the last 7 years. The gene positions of the known genes were based on highest probability sequence alignment (BLASTn) of functional or closely linked gene-based markers to the International Wheat Genome Sequencing Consortium reference sequence (RefSeq vs 1.0) performed in Ensembl Plants ([https://plants.ensembl.org/Triticum\\_aestivum/Info/Index](https://plants.ensembl.org/Triticum_aestivum/Info/Index)). Estimation of the frequency changes of the recently discovered QTL was possible because the same genotyping platform was used in both studies. The favorable allele of the first SNP of each haplotype was used to examine the allele frequency change over time in each yield trial, separately.

## 2.7 | Statistical analysis

All data processing and statistical analyses, otherwise specified in above, were implemented in R v.3.5.1 (R Development Core Team, 2018).

## 3 | RESULTS

### 3.1 | Genotypic data

A final number of 40,530 SNP markers remained after data filtering and imputation, which were spread across all 21 chromosomes (Supplementary Table S3). The largest number of SNPs were physically positioned in genome B and the lowest in genome D. Only 1.3% of the SNP markers could not be assigned to any chromosome. Chromosomes 2B, 6B, and 7A were the three most densely covered chromosomes with on average one marker per 0.23 Mbp, while chromosomes 4D, 5D, and 2D were the least densely covered chromosomes averaging one marker per 1.00 Mbp.

### 3.2 | Population structure

We superimposed the k-means grouping structure by color-coding every entry by its respective predefined subpopulation. The superimposition was done due to the unclear

number of clusters to be chosen when analyzing the optimal number of  $k$  against the BIC (Supplementary Figure S1). The alpha-score optimization statistic had its optimum at 216 PC (Supplementary Figure S2). All subpopulations showed admixture in the unsupervised DAPC (Figure 2a). However, the lines assigned to the first two subpopulations ES1–7 and ES8–13 tended to concentrate in the second quadrant of the DAPC plot and subpopulation ES26–38 tended to concentrate toward the third quadrant. The lines in subpopulation SA13–25 tended to group opposite to the previous subpopulations between the first and the fourth quadrant. Lines in the two subpopulations, SA1–12 and ES14–25, showed a random scatter in the plot without any tendency of grouping. Supervised DAPC showed that summarizing the data with predefined subpopulations provided an overall U-shaped structure (Figure 2b). Like the unsupervised DAPC, the first two ESWYT subpopulations, ES1–7 and ES8–13, grouped closely together, next to subpopulation ES14–25. The ESWYT subpopulations formed the first half of the U-shaped structure. The two SAWYT subpopulations formed the second half of the U-shaped structure, while the subpopulation SA14–25 was related more closely to all ESWYT populations than the population SA1–12.

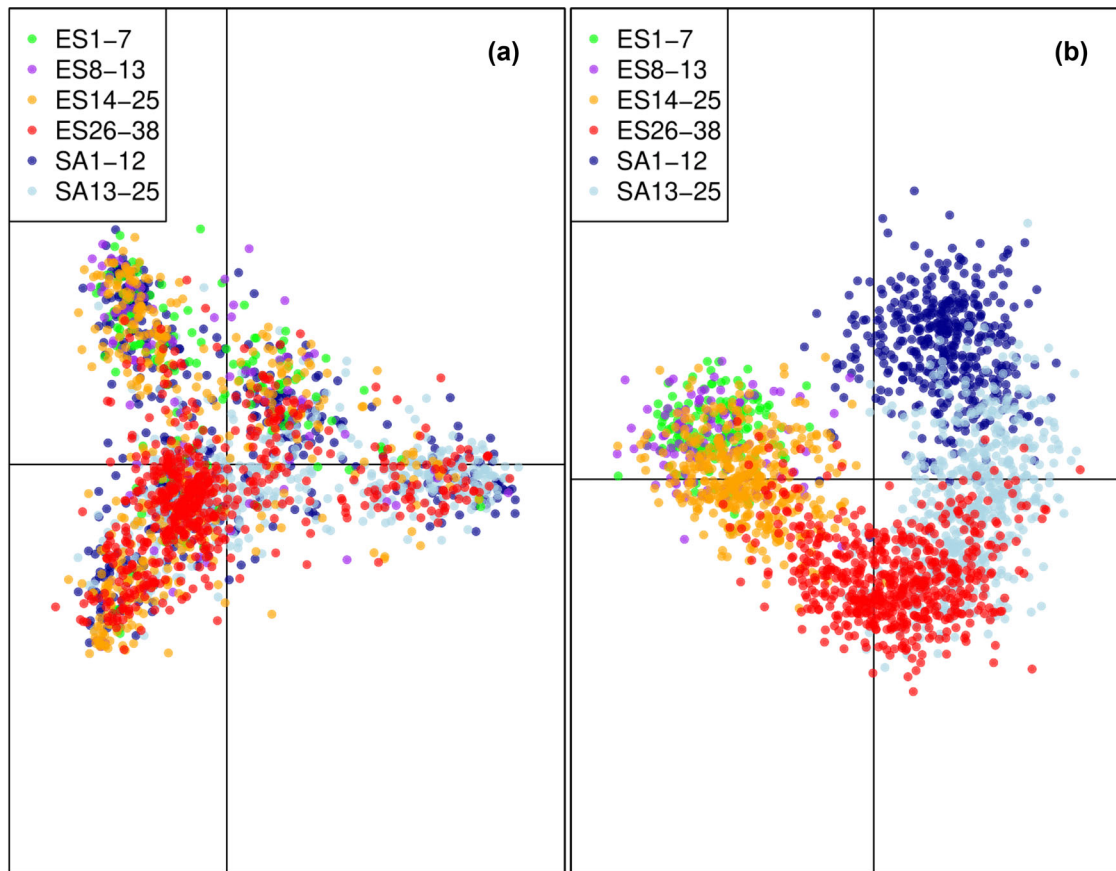
### 3.3 | $F_{ST}$

The  $F_{ST}$  values between 3- and 4-yr groups were overall low. The  $F_{ST}$  values comparing each group to the initial group of four yield trials (ES1–4) increased over time for each yield trial (Figure 3a). While up to 2000 (SA7–9 and ES20–22), the  $F_{ST}$  values relative to the initial group were always higher for the ESWYT than for the SAWYT, this relationship reversed after 2000.

The  $F_{ST}$  values of consecutive groups were very similar in the ESWYT and SAWYT and ranged from 0.004 to 0.033. Comparisons between parallel groups of the two different yield trials from 1992 until 2016 showed increasing  $F_{ST}$  values up to year 2006 (SA13–15 and ES26–28), which then declined (Figure 3b).

### 3.4 | Changes in allele frequency over time and selection signatures

We plotted all SNP markers based on MAF deviations to visualize patterns along the chromosomes (Supplementary Figure S3 and S4). Overall, we observed regions that showed large linkage blocks, specially flanking the centromere (on chromosomes 2A, 2B, 3D, 5A, 5B, 6A, 6B, and 7A). However, some large linkage blocks were also outside the centromeric region (e.g., on chromosome 1D). Such regions showed constant MAF values for each of the six subpopulations under study.



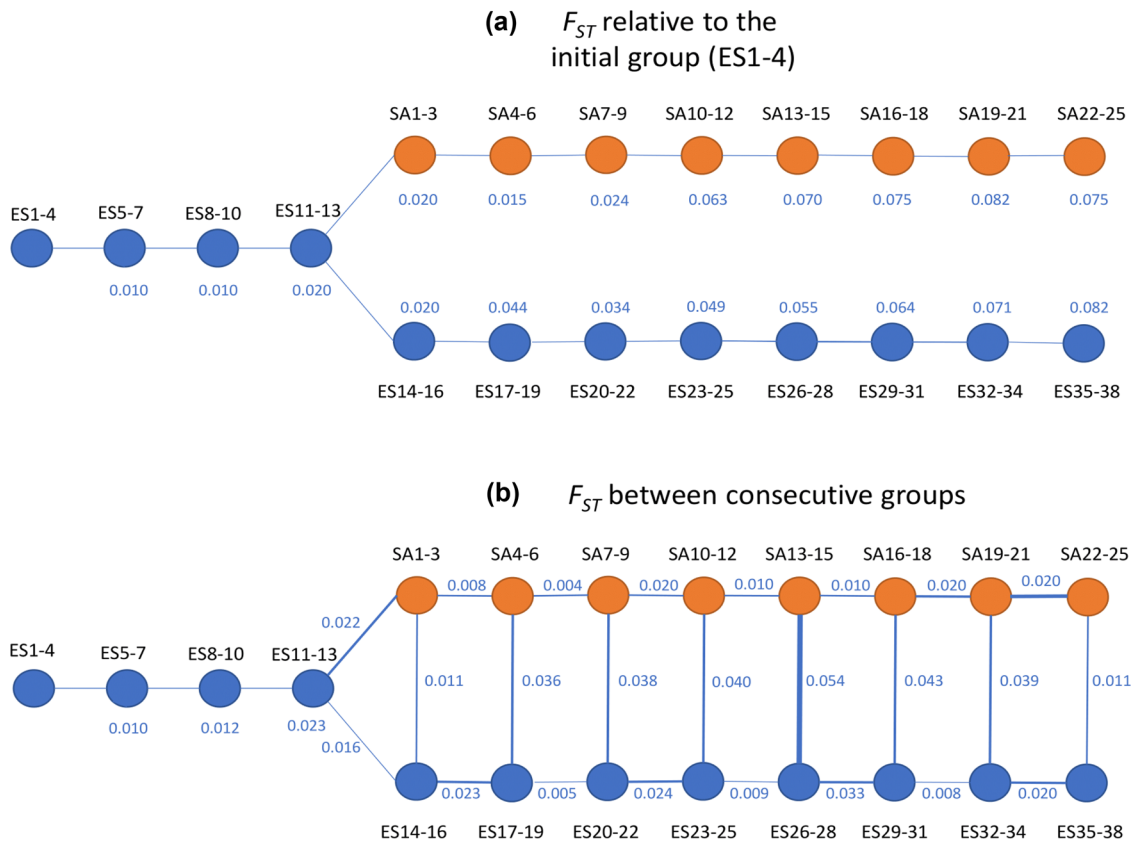
**FIGURE 2** Discriminant analysis of principal components (DAPC) using two methods: (a) DAPC grouped by k-means algorithm with color code superimposition for the predefined subpopulations; (b) DAPC grouped by predefined subpopulations

Conversely, regions around the telomere showed a noisy pattern, most likely due to high rates of recombination. Density distribution plots of MAF deviations with their respective quantile boundary thresholds were devised for the different signatures of selection under study (Supplementary Figure S3). MAF deviations in temporal dimension had a rather smooth distribution that concentrated toward zero with the upper- and lower- boundaries of one percent quantiles beginning around 0.25 and  $-0.25$  (Supplementary Figure S3a). The MAF deviations in environmental dimension (among the two yield trials) showed a peak shifted toward negative values, with the upper and lower boundaries beginning around 0.2 and  $-0.2$  (Supplementary Figure S3a). From the total number of markers, 9.8% could be determined to be under selection over time following our threshold (Table 2). When considering the largest time span of MAF deviations (i.e., ES1–7 to ES26–38; SA14–25), they accounted for almost all the temporal differences (8.2%) leaving only 1.6% to other subpopulations. Environmental differences accounted for 2.0% using the deviation from the two most recent subpopulations of each yield trial (ES26–38 and SA13–25). Most selection signatures derived from the MAF deviations of the first (ES1–7) to the two most recent populations (SA14–25; ES26–38), with 1,871 markers

identified between ES1–7 and SA14–25 and 1,473 between ES1–7 to ES26–38, accounting together for 84% of the overall markers under selection. Chromosomes 6A, 6B, 4A, 4B, and 4D showed clear shifts of MAF deviation into one direction, while other chromosomes (e.g., 3A, 1B, 5A) showed allele frequency shifts into both directions. Allele frequencies shifts spanned substantial physical distances representing large linkage blocks, often (but not always) around the centromere (Figure 4). When considering MAF deviations of environmental adaptation, we detected 797 markers under selection. Chromosomes 3B, 1B, 5B, 2A, and 5A in descending order had the higher concentration of markers with a range of 48 to 159 (Figure 5). Chromosomes 2A, 2B, and 6B showed a larger number of markers concentrated in big linkage blocks.

### 3.5 | Exploring known genes and QTL within the genome-wide scans

We integrated several known genes and recently identified grain yield QTL in our genome-wide scans to test if any of the genes fall into signatures of selection. On chromosome 1, the *Glu-A1*, *Glu-B1*, and *Glu-D1* genes were flanked by only few



**FIGURE 3** Fixation index ( $F_{ST}$ ) using two different approaches: (a)  $F_{ST}$  values comparing each bin to the initial group; (b)  $F_{ST}$  comparisons between consecutive groups

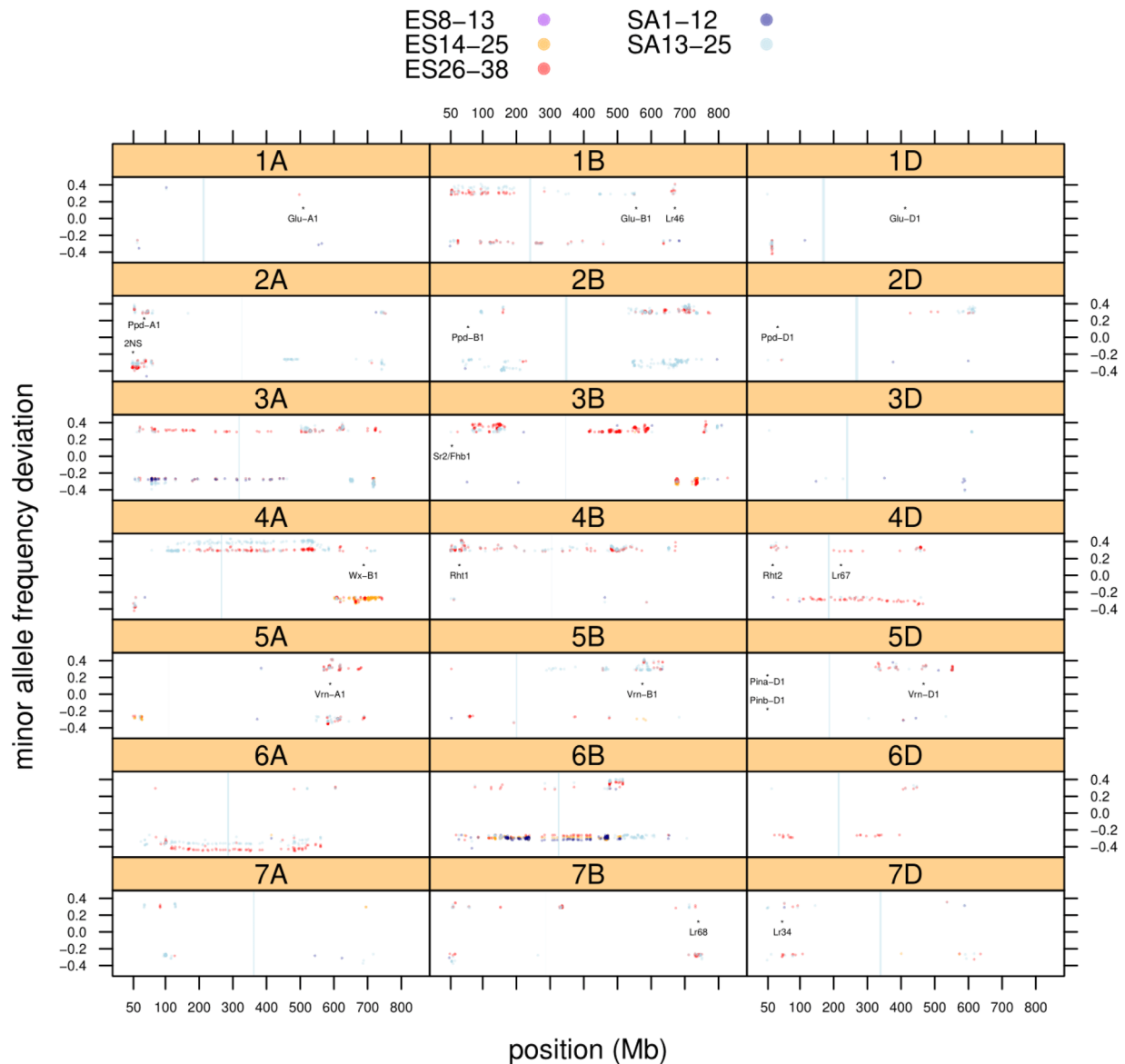
**TABLE 2** Summary of observed signatures of selection across International Maize and Wheat Improvement Center's the Elite Spring Wheat Yield Trial and the Semi-arid Wheat Yield Trial

Selection signature type	Numbers of markers (%)	Populations
Over time	3,975 (9.8%)	ES1–7 – all
Over largest time span	3,344 (8.2%)	ES1–7 – ES26–38; SA13–25
Over environments/trials	7,97 (2.0%)	ES26–38 – SA13–25

markers. In contrast, gene *Lr46* located on chromosome 1B conferring slow-rusting resistance was located within a small linkage block on temporal dimension with maximized allele frequency shifts in ES26–38 and SA13–25. On chromosome 2, the photoperiod response gene *Ppd-A1* and the 2NS translocation introgressed from *Triticum ventricosum* (Tausch) Cess. carrying several disease resistance genes were flanked by linkage blocks in the temporal selection. Both loci showed a unidirectional selection pattern toward populations ES26–38 and SA13–25. On chromosome 3, the two closely linked disease resistance genes *Sr2* and *Fhb1* were only sparsely covered with markers. On chromosome 4, gene *Wx-B1* was positioned within clear linkage blocks when in temporal dimension. The plant height genes *Rht1* and *Rht2* were located within linkage blocks with decreasing minor allele frequency over time. The rust resistance gene *Lr67* on chromosome 4D fell into a

larger linkage block with selection in the opposite direction of *Rht2*. On chromosome 5, the vernalization response gene *Vrn-A1* showed a bidirectional shift of the minor allele, while the minor alleles of markers under selection close to *Vrn-B1* and *D1* decreased. The gene *Vrn-B1* showed a clear sweep among environments. The grain hardness genes *Pina-D1*, *Pinb-D1* had low coverage. On chromosome 7, markers flanking the two rust resistance genes, *Lr68* and *Lr34*, showed an increase of the minor allele over time.

Of the 28 grain yield QTL considered, 14 QTL showed an overall increase of the favorable allele over time for both yield trials (see Figure 6). In the ESWYT, the favorable allele of 19 (67%) QTL were increasing, while for nine (33%) QTL the favorable allele frequency decreased. In the SAWYT, the favorable allele of 16 (57%) QTL were increasing, while for 12 (43%) QTL the favorable allele frequency decreased over



**FIGURE 4** Temporal minor allele frequency (MAF) deviations, shaded areas are centromeric physical regions

time. Some QTL were clearly selected; for example, QTL ‘*S5D\_550192169*’ progressed from a frequency of 0.6 for the favorable allele to 0.9 in both most recent subpopulations. Other QTL were clearly selected in the opposite direction; for example, ‘*S5B\_47584429*’, which departed from 0.45 frequencies in the most ancient population to 0.25 in the most recent.

## 4 | DISCUSSION

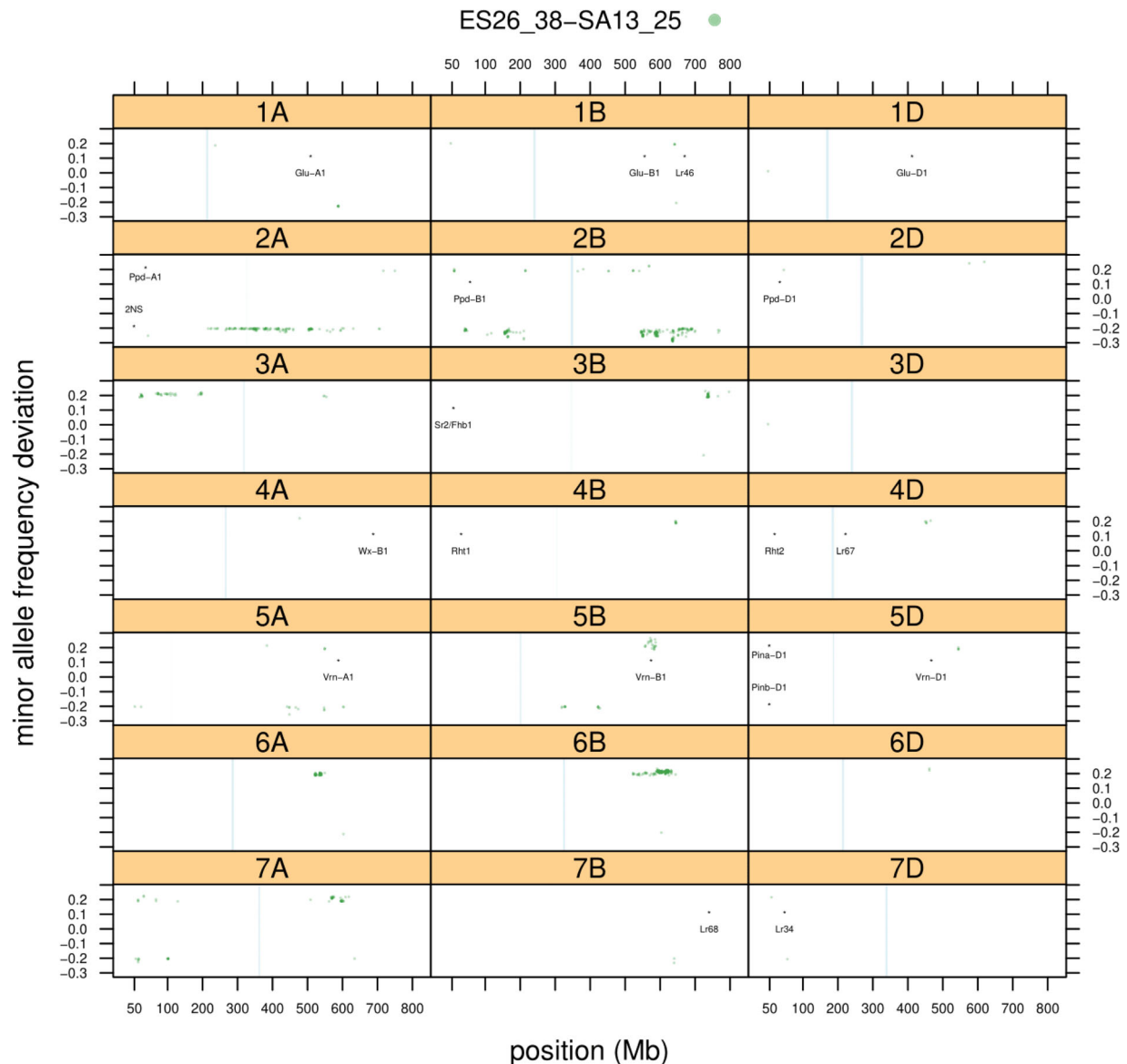
The CIMMYT Global Wheat Program aims to develop better adapted wheat cultivars for farmers in the developing world. The International Wheat Improvement Network is thereby crucial in disseminating CIMMYT wheat lines carrying core traits of high and stable grain yield, disease resistance, and

quality to national partners, thus representing a unique delivery pathway for new potential cultivars (Reynolds et al., 2017). The yield trials utilized in this study are disseminated to two out of 12 defined MEs, ME1 and ME4, which cover a large wheat production area globally (Rajaram, 2002). We genotyped the lines included in ESWYT and SAWYT since their initial distribution and investigated their genetic structure as well as signatures of selection within and among the trials.

### 4.1 | Population stratification

Genotyping-by-sequencing is a simple highly multiplexed system for constructing reduced library representations (Elshire et al., 2011). It provides large numbers of SNPs



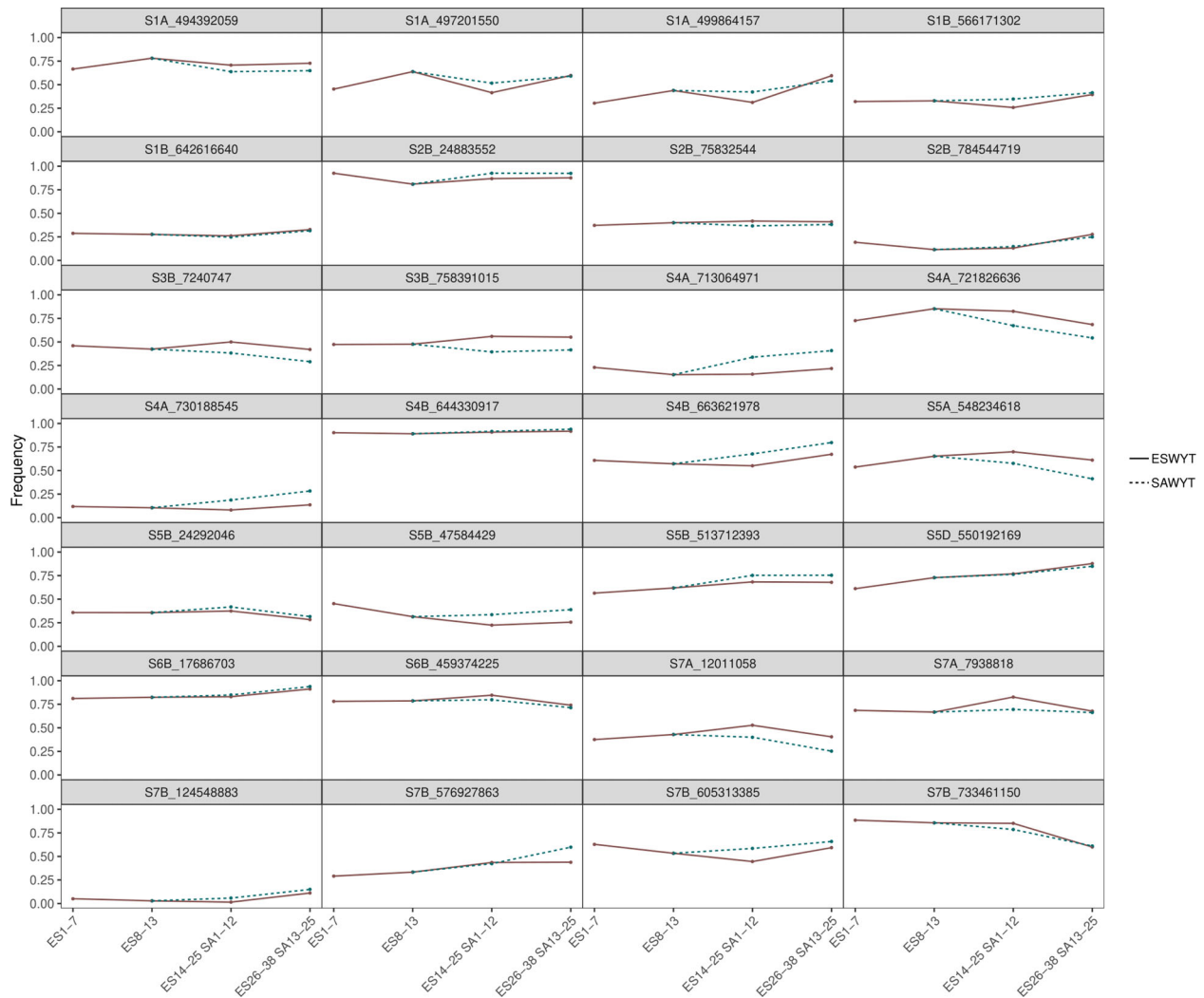


**FIGURE 5** Environmental minor allele frequency (MAF) deviations, shaded areas are centromeric physical regions

for use in genotyping (Beissinger et al., 2013) while keeping the costs low, reducing sample handling, and providing efficient barcoding without reference sequence limits. With GBS, marker discovery and genotyping occur simultaneously, resulting in minimum ascertainment bias. Therefore, we had an extensive coverage of the wheat genome compared with previous studies that used different marker types and significantly less markers (Balfourier et al., 2007; S. Dreisigacker et al., 2004; White et al., 2008).

Population stratification is often driven by pedigree and individuals are rarely drawn from panmictic populations. We observed considerable admixture across the lines included in the ESWYT and SAWYT. This result supports previous findings on population stratification analysis in CIMMYT wheat germplasm. For instance, Dreisigacker et al. (2012) looked at the ESWYT and found instability using diversity array tech-

nology (DArT) markers and the software STRUCTURE on determining that the optimal number of K subpopulations also suggested strong relatedness of ESWYT germplasm. Another study using the same DArT marker technology and software determined the population stratification within individual ESWYT trials and found an overall pattern of unstructured populations (Crossa et al., 2007). More recently, Juliana et al. (2018) performed a principal component analysis using GBS and found no strong population structure among lines in CIMMYT elite yield trials. In contrast, clear population structure is observed in other breeding populations; for example, Canadian durum wheat (*Triticum durum* Desf.) (N'Diaye et al., 2018) with clear differences in pedigree and breeding program sources in Canada involving relatively little introduction of new material. Similarly, clear clusters were observed across 185 commercially released cultivars across the Great Plains



**FIGURE 6** Favorable alleles over time of 28 grain yield quantitative trait loci (QTL). ESWYT, Elite Spring Wheat Yield Trial; SAWYT, Semi-arid Wheat Yield Trial

in the United States (Ayalew et al., 2020). The CIMMYT is known for the consistent use of diverse parents and substantial genetic resources (e.g., landraces and synthetic hexaploid wheat) that have been vital to meet the requirements of global adaptation and enable the adoption of newly developed lines by cooperating wheat research programs globally (Dreisgacker et al., 2004; Rosyara et al., 2019). Hence, our results of substantial admixture in subpopulations were expected. However, some trends were also visible in our DAPC plot. Wheat lines from the later ESWYT and SAWYT tended to group apart from the first subpopulation including the initial 13 years of ESWYT only (from 1979 to 1992). In addition, lines specific to the later ESWYT and SAWYT concentrated in larger parts in different clusters.

Admixture of subpopulations in the ESWYT and SAWYT trials were also reflected by the overall low  $F_{ST}$  values ( $0.01 < F_{ST} < 0.05$ ). The  $F_{ST}$  can be thought of as a measure of the correlation of genes between populations.

Low  $F_{ST}$  values indicate gene flow between populations, which is expected in breeding materials as parent and subsequently their offspring are re-cycled with every breeding cycle. Higher  $F_{ST}$  values may arise by genetic drift or artificial selection (Suzuki, 2010). The increasing  $F_{ST}$  values relative to the initial 4-yr group suggests continued line improvement over time through artificial selection (Figure 5). Allele frequency shifts for both ESWYT and SAWYT on trends of genetic gains have been investigated in earlier studies by phenotypic data analysis. Crespo-Herrera et al. (2018) estimated grain yield gains over eight years of ESWYT, 2006–2007 (ES27) to 2014–2015 (ES34) relative to the widely grown cultivar ‘Attila’ (GID14337) and found 1.67% improvement over time. Crespo-Herrera et al. (2018) looked at grain yield improvement in SAWYT germplasm grown during 2002–2003 (SA11) to 2012–2013 (SA23) relative to four drought-tolerant wheat lines used as constant checks and found an overall rate of increase of 1.8% in low yield environments

and 1.4% in medium yield environments. The lower frequency shifts in the SAWYT from its initiating year in 1992 to 2000 together with the rather random scatter of the same lines in the DAPC plot, probably reflects the time that was required to establish the core traits and distribute the germplasm to ME4 (Trethowan et al., 2002). Among trials,  $F_{ST}$  values increased until 2006 (ES26–28; SA13–15) and decreased from 2007 until 2016. During this period, established, large-scale, and simultaneous testing pipelines for grain yield potential under several environmental conditions, including both, drought stress, and fully irrigated, were routinely performed. Substantial spillover was identified, for example, advanced lines targeted to ME1 were recognized as suitable parents to be used in ME4. Consequently, advanced lines are currently extracted from one larger diverse gene pool with specific emphasis on good performance across trial and management conditions, to maintain yield stability, a key feature of CIMMYT wheat germplasm and specifically important to cope with interannual variability in the yield driven by climate change (Morton, 2007).

## 4.2 | Selection signatures

Loci under directional selection are expected to have higher interpopulation variability than intrapopulation variability when compared with neutral loci. Applied to our study, loci showing larger amounts of differentiation between subpopulations may reveal genomic regions that have been subject to diversifying selection. This differentiation can be assessed by patterns of variance in allele frequency or by examining  $F_{ST}$  values between populations (Kirk & Freeland, 2011; Konijnendijk et al., 2015). A significant amount (9.8%) of loci indicated strong evidence of selection over 38 yr of breeding in CIMMYT's bread wheat elite lines. N'Diaye et al. (2018) explored loci under selection in breeding for durum wheat through three different methods: using the total variance  $F_{ST}$ -based outlier detection method (Lositan), the hierarchical island model (Arlequin), and the Bayesian genome scan method (BayeScan). Although breeding programs and species cannot be compared directly, the authors found a similar percentage of SNPs under selection. The percentage of SNPs under selection was higher over time than over trials/environments. This may be due to several reasons: (a) a longer time period when estimating selection signatures in temporal (38 years) versus environmental dimension (25 yr); (b) the above-mentioned spillover effect, so genes conferring fitness to, for example, high yields or disease resistance, are similar in both sets, ESWYT and SAWYT, thus genes conferring fitness to one ME may confer fitness to several MEs; (c) both yield trials are targeted to different ME's, however these are assembled to maintain a broad genetic base in CIMMYT germplasm; (d) adaptation to a specific ME is expected

to be based on a rather limited number of specific genes, thus a lower percentage of SNP under selection is expected. To the best of our knowledge, this study is the first that define gene regions driving adaptation to two MEs in wheat. A more detailed scrutiny of these regions will shed light on the genetic architecture of key adaptive traits in CIMMYT spring bread wheat elite lines.

## 4.3 | Known genes and recently observed QTL

Several genes and QTL were observed as having physical proximity to the loci we have identified being under selection over time and through environmental adaptation. Physical proximity relevance on chromosome 1 can be attributed to gene *Lr46* which is known to confer slow-rusting resistance and was first described by (Singh et al., 1998) in cultivar 'Pavon 76'. On chromosome 2, the photoperiod sensitivity gene *Ppd-A1* had closer proximity to a cluster of markers that were both present in our temporal selection signatures. Photoperiod insensitivity is being targeted by CIMMYT since N. E. Borlaug implemented the shuttle breeding concept (Rajaram, 2002). Almost all CIMMYT elite lines are photoperiod insensitive (showing high frequency of *Ppd-D1a* allele), due to selection in two photoperiod contrasting environments during the shuttle. The 2NS translocation introgressed from *Ae. ventricum* and carrying at least three disease resistance genes (*Lr37*, *Yr17*, *Sr28*) were flanking several markers also on chromosome 2. Cultivars with the CIMMYT line 'Milan' in their pedigree are known to contain the 2NS translocation, which recently also showed high levels of resistance to wheat blast under field conditions (Cruz et al., 2016; Juliana et al., 2018). Interestingly the same genomic region was also identified as selection sweeps in the studies by (Ayalew et al., 2020; Dadshani et al., 2021). On chromosome 4, the waxy gene *Wx-B1* was physically located close to markers under temporal selection. Also, dwarfing genes *Rht1* and *Rht2* were under temporal selection. All CIMMYT lines today are semi-dwarf and carry, in comparison with the early green revolution lines mainly the dwarfing allele at *Rht1*, some the dwarfing allele at *Rht2*. On chromosome 5B, all three homologous vernalization genes were under temporal selection and *Vrn-B1* was also under environmental selection. These results were in line with CIMMYT internal trait-based marker data related to various phenology genes, which showed an overtime increasing frequency of the spring alleles at *Vrn-B1* and *Vrn-D1* and of the winter allele at *Vrn-A1*, specifically in the SAWYT. On chromosome 7, the slow-rusting resistance gene *Lr34* and the adult-plant resistance gene *Lr68* were in proximity of markers selected in temporal dimension, emphasizing CIMMYT efforts on providing durable rust resistance.

A logical next step would be to relate the selection signatures to several key phenotypic traits. However, phenotypic data (especially from the initial years) have shown to be highly unbalanced. Therefore, we plotted allele frequency changes of recent GWAS results for grain yield derived from an extensive data set of CIMMYT advanced lines (Sehgal et al., 2020). In both yield trials, for more than half of the QTL (57% and 67% for the SAWYT and ESWYT, respectively) the favorable allele increased in frequency over time. This tendency corroborates the trait relevance of the identified QTL. Most importantly, an important genomic region on chromosome 5D was identified, which was under strong selection. This genomic region that showed an increase in frequency over time in trials was also identified by Sehgal et al. (2020) using haplotype-based GWAS. Monitoring and additional pyramiding of these QTL offer breeders new opportunities for further improvement facilitating marker-assisted or genomic selection.

## ACKNOWLEDGMENTS

GBS genotyping was supported by the ‘Next frontier for rapid gains in maize and wheat improvement’ project through the Bill & Melinda Gates Foundation Grant No.: OPP101638; the Feed the Future Innovation Lab for Applied Wheat Genomics project through the US Agency for International Development (USAID), under the terms of contract no. AIDOAA-A-13-00051.


## AUTHOR CONTRIBUTIONS

Alexandre Mondaini: Data curation; Formal analysis; Writing-original draft; Writing-review & editing. Umesh Rosyara: Conceptualization; Data curation; Formal analysis; Writing-review & editing. Deepmala Sehgal: Data curation; Writing-review & editing. Susanne Dreisigacker: Conceptualization; Data curation; Writing-original draft; Writing-review & editing.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ORCID

Susanne Dreisigacker  <https://orcid.org/0000-0002-3546-5989>

## REFERENCES

- Afzal, F., Li, H., Gul, A., Subhani, A., Ali, A., Mujeeb-Kazi, A., Ogbonnaya, F., Trethowan, R., Xia, X., He, Z., & Rasheed, A. (2019). Genome-wide analyses reveal footprints of divergent selection and drought adaptive traits in synthetic-derived wheats. *G3: Genes, Genomes, Genetics*, 9(6), 1957–1973. <https://doi.org/10.1534/g3.119.400010>
- Appels, R., Eversole, K., Stein, N., Feuillet, C., Keller, B., Rogers, J., Pozniak, C. J., Choulet, F., Distelfeld, A., Poland, J., Ronen, G., Sharpe, A. G., Barad, O., Baruch, K., Keeble-Gagnère, G., Mascher, M., Ben-Zvi, G., Josselin, A., Himmelbach, A., ... Wang, L. e (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, 361(6403). <https://doi.org/10.1126/science.aar7191>
- Ayalew, H., Sorrells, M. E., Carver, B. F., Baenziger, P. S., & Ma, X.-F. (2020). Selection signatures across seven decades of hard winter wheat breeding in the Great Plains of the United States. *Plant Genome*, 13(3), e20032. <https://doi.org/10.1002/tpg2.20032>
- Balfourier, F., Roussel, V., Strelchenko, P., Exbrayat-Vinson, F., Sourdille, P., Boutet, G., Koenig, J., Ravel, C., Mitrofanova, O., Beckert, M., & Charmet, G. (2007). A worldwide bread wheat core collection arrayed in a 384-well plate. *Theoretical and Applied Genetics*, 114(7), 1265–1275. <https://doi.org/10.1007/s00122-007-0517-1>
- Baum, M., Tadesse, W., Nachit, M., Abdalla, O., Rajaram, S., Singh, R., Payne, T., Ammar, K., Morgounov, A., & Braun, H. (2015). Global crop improvement networks to bridge technology gaps. In Y. Ogihara, S. Takumi, & H. Handa (Eds.), *Advances in wheat genetics: From genome to field* (pp. 387–399). Springer. [https://doi.org/10.1007/978-4-431-55675-6\\_44](https://doi.org/10.1007/978-4-431-55675-6_44)
- Beissinger, T. M., Hirsch, C. N., Sekhon, R. S., Foerster, J. M., Johnson, J. M., Muttoni, G., Vaillancourt, B., Buell, C. R., Kaeppler, S. M., & De Leon, N. (2013). Marker density and read depth for genotyping populations using genotyping-by-sequencing. *Genetics*, 193(4), 1073–1081. <https://doi.org/10.1534/genetics.112.147710>
- Braun, H.-J., Rajaram, S., & Van Ginkel, M. (1996). CIMMYT’s approach to breeding for wide adaptation. *Euphytica*, 92(1–2), 175–183. [https://doi.org/10.1007/978-94-015-8806-5\\_25](https://doi.org/10.1007/978-94-015-8806-5_25) <https://doi.org/10.1007/BF00022843>
- Browning, B. L., & Browning, S. R. (2016). Genotype imputation with millions of reference samples. *American Journal of Human Genetics*, 98(1), 116–126. <https://doi.org/10.1016/j.ajhg.2015.11.020>
- Byth, D. E., Eisemann, R. L., & De Lacy, I. H. (1976). Two-way pattern analysis of a large data set to evaluate genotypic adaptation. *Heredity*, 37(2), 215–230. <https://doi.org/10.1038/hdy.1976.84>
- Cavanagh, C. R., Chao, S., Wang, S., Huang, B. E., Stephen, S., Kiani, S., Forrest, K., Saintenac, C., Brown-Guedira, G. L., Akhunova, A., See, D., Bai, G., Pumphrey, M., Tomar, L., Wong, D., Kong, S., Reynolds, M., Da Silva, M. L., Bockelman, H., ... Akhunov, E. (2013). Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proceedings of the National Academy of Sciences of the United States of America*, 110(20), 8057–8062. <https://doi.org/10.1073/pnas.1217133110>
- Crespo-Herrera, L. A., Crossa, J., Huerta-Espino, J., Vargas, M., Mondal, S., Velu, G., Payne, T. S., Braun, H., & Singh, R. P. (2018). Genetic gains for grain yield in cimmyt’s semi-arid wheat yield trials grown in suboptimal environments. *Crop Science*, 58(5), 1890–1898. <https://doi.org/10.2135/cropsci2018.01.0017>
- Crespo-Herrera, L. A., Crossa, J., Huerta-Espino, J., Autrique, E., Mondal, S., Velu, G., Vargas, M., Braun, H. J., & Singh, R. P. (2017). Genetic yield gains in CIMMYT’S international elite spring wheat yield trials by modeling the genotype × environment interaction. *Crop Science*, 57(2), 789–801. <https://doi.org/10.2135/cropsci2016.06.0553>
- Crossa, J., José, Burgueño, J., Dreisigacker, S., Vargas, M., Herrera-Foessel, S. A., Lillemo, M., Singh, R. P., Trethowan, R., Warburton, M., Franco, J., Reynolds, M., Crouch, J. H., & Ortiz, R. (2007). Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. *Genetics*, 177(3), 1889–1913. <https://doi.org/10.1534/genetics.107.078659>

- Cruz, C. D., Peterson, G. L., Bockus, W. W., Kankanala, P., Dubcovsky, J., Jordan, K. W., Akhunov, E., Chumley, F., Baldelomar, F. D., & Valent, B. (2016). The 2NS translocation from *Aegilops ventricosa* confers resistance to the Triticum pathotype of *Magnaporthe oryzae*. *Crop Science*, *56*(3), 990–1000. <https://doi.org/10.2135/cropsci2015.07.0410>
- Dadshani, S., Mathew, B., Ballvora, A., Mason, A. S., & Léon, J. (2021). Detection of breeding signatures in wheat using a linkage disequilibrium-corrected mapping approach. *Scientific Reports*, *11*(1), 5527. <https://doi.org/10.1038/s41598-021-85226-1>
- Dawson, J. C., Endelman, J. B., Heslot, N., Crossa, J., Poland, J., Dreisigacker, S., Manès, Y., Sorrells, M. E., & Jannink, J. - L. (2013). The use of unbalanced historical data for genomic selection in an international wheat breeding program. *Field Crops Research*, *154*, 12–22. <https://doi.org/10.1016/j.fcr.2013.07.020>
- Delacy, I. H., Fox, P. N., Corbett, J. D., Crossa, J., Rajaram, S., Fischer, R. A., & Van Ginkel, M. (1993). Long-term association of locations for testing spring bread wheat. *Euphytica*, *72*(1–2), 95–106. <https://doi.org/10.1007/BF00023777>
- Dreisigacker, S., Zhang, P., Warburton, M. L., Van Ginkel, M., Hoisington, D., Bohn, M., & Melchinger, A. E. (2004). SSR and pedigree analyses of genetic diversity among CIMMYT wheat lines targeted to different megaenvironments. *Crop Science*, *44*(2), 381–388. <https://doi.org/10.2135/cropsci2004.3810>
- Dreisigacker, S., Shewayrga, H., Crossa, J., Arief, V. N., Delacy, I. H., Singh, R. P., Dieters, M. J., & Braun, H. - J. (2012). Genetic structures of the CIMMYT international yield trial targeted to irrigated environments. *Molecular Breeding*, *29*(2), 529–541. <https://doi.org/10.1007/s11032-011-9569-7>
- Elshire, R. J., Glaubitz, J. C., Sun, Q.i, Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *Plos One*, *6*(5), 19379. <https://doi.org/10.1371/journal.pone.0019379>
- FAO. (2016). FAOSTAT Wheat Production Statistics. [www.fao.org/faostat/en/#data/QCL](http://www.fao.org/faostat/en/#data/QCL)
- FAO. (2017). FAOSTAT Wheat Production Statistics. [www.fao.org/faostat/en/#data/QCL](http://www.fao.org/faostat/en/#data/QCL)
- Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q.i, & Buckler, E. S. (2014). TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *Plos One*, *9*(2), e90346. <https://doi.org/10.1371/journal.pone.0090346>
- Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*(1), 184–186. <https://doi.org/10.1111/j.1471-8286.2004.00828.x>
- Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Applied Statistics*, *28*(1), 100. <https://doi.org/10.2307/2346830>
- Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: Defining, estimating and interpreting FST. *Nature Reviews Genetics*, *10*(9), 639–650. <https://doi.org/10.1038/nrg2611>
- Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. <https://doi.org/10.1093/bioinformatics/btr521>
- Juliana, P., Singh, R. P., Poland, J., Mondal, S., Crossa, J., Montesinos-López, O. A., Dreisigacker, S., Pérez-Rodríguez, P., Huerta-Espino, J., Crespo-Herrera, L., & Govindan, V. (2018). Prospects and challenges of applied genomic selection—a new paradigm in breeding for grain yield in bread wheat. *The Plant Genome*, *11*(3), 180017. <https://doi.org/10.3835/plantgenome2018.03.0017>
- Kirk, H., & Freeland, J. R. (2011). Applications and implications of neutral versus non-neutral markers in molecular ecology. *International Journal of Molecular Sciences*, *12*(6), 3966–3988. <https://doi.org/10.3390/ijms12063966>
- Konijnendijk, N., Shikano, T., Daneels, D., Volckaert, F. A. M., & Raeymaekers, J. A. M. (2015). Signatures of selection in the three-spined stickleback along a small-scale brackish water–freshwater transition zone. *Ecology and Evolution*, *5*(18), 4174–4186. <https://doi.org/10.1002/ece3.1671>
- Laland, K. N., Odling-Smee, J., & Myles, S. (2010). How culture shaped the human genome: Bringing genetics and the human sciences together. *Nature Reviews Genetics*, *11*(2), 137–148. <https://doi.org/10.1038/nrg2734>
- Mondal, S., Dutta, S., Crespo-Herrera, L., Huerta-Espino, J., Braun, H. J., & Singh, R. P. (2020). Fifty years of semi-dwarf spring wheat breeding at CIMMYT: Grain yield progress in optimum, drought and heat stress environments. *Field Crops Research*, *250*, 107757. <https://doi.org/10.1016/j.fcr.2020.107757>
- Morton, J. F. (2007). The impact of climate change on smallholder and subsistence agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(50), 19680–19685. <https://doi.org/10.1073/pnas.0701855104>
- N'diaye, A., Haile, J. K., Nilsen, K. T., Walkowiak, S., Ruan, Y., Singh, A. K., Clarke, F. R., Clarke, J. M., & Pozniak, C. J. (2018). Haplotype loci under selection in canadian durum wheat germplasm over 60 years of breeding: Association with grain yield, quality traits, protein loss, and plant height. *Frontiers in Plant Science*, *871*, 1589. <https://doi.org/10.3389/fpls.2018.01589>
- Neale, D. B., & Kremer, A. (2011). Forest tree genomics: Growing resources and applications. *Nature Reviews Genetics*, *12*(2), 111–122. <https://doi.org/10.1038/nrg2931>
- Marks, J. (1988). Molecular evolutionary genetics. *American Journal of Physical Anthropology*, *75*(3), 428–429. <https://doi.org/10.1002/ajpa.1330750317>
- Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J. - L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *Plos One*, *7*(2), 32253. <https://doi.org/10.1371/journal.pone.0032253>
- Rajaram, S. (2002). *CIMMYT international wheat breeding*. <http://www.fao.org/3/y4011e/y4011e09.htm>
- Reynolds, M. P., Braun, H. J., Cavalieri, A. J., Chapotin, S., Davies, W. J., Ellul, P., Feuillet, C., Govaerts, B., Kropff, M. J., Lucas, H., Nelson, J., Powell, W., Quilligan, E., Rosegrant, M. W., Singh, R. P., Sonder, K., Tang, H., Visscher, S., & Wang, R. (2017). Improving global integration of crop research. *Science*, *357*, (6349), 359–360. <https://doi.org/10.1126/science.aam8559>
- Rosyara, U., Kishii, M., Payne, T., Sansaloni, C. P., Singh, R. P., Braun, H. - J., & Dreisigacker, S. (2019). Genetic contribution of synthetic hexaploid wheat to CIMMYT's spring bread wheat breeding germplasm. *Scientific Reports*, *9*(1). <https://doi.org/10.1038/s41598-019-47936-5>
- Sarkar, D. (2008). *Lattice: Multivariate data visualization with R*. Springer Science & Business Media.
- Sehgal, D., Mondal, S., Crespo-Herrera, L., Velu, G., Juliana, P., Huerta-Espino, J., Shrestha, S., Poland, J., Singh, R., & Dreisigacker, S. (2020). Haplotype-based, genome-wide association study reveals stable genomic regions for grain yield in CIMMYT spring bread wheat. *Frontiers in Genetics*, *11*, 1427. <https://doi.org/10.3389/fgene.2020.589490>

- Singh, R. P., Mujeeb-Kazi, A., & Huerta-Espino, J. (1998). Lr46: A gene conferring slow-rusting resistance to leaf rust in wheat. *Phytopathology*, 88(9), 890–894. <https://doi.org/10.1094/PHTO.1998.88.9.890>
- Suzuki, Y. (2010). Statistical methods for detecting natural selection from genomic data. *Genes & Genetic Systems*, 85, 359–376. <https://doi.org/10.1266/ggs.85.359>
- Trethowan, R. M., Ginkel, M., & Rajaram, S. (2002). Progress in breeding wheat for yield and adaptation in global drought affected environments. *Crop Science*, 42(5), 1441–1446. <https://doi.org/10.2135/cropsci2002.1441>
- Van Bavel, J. (2013). The world population explosion: Causes, backgrounds and -projections for the future. *Facts, Views & Vision in ObGyn*, 5(4), 281–291. <http://www.ncbi.nlm.nih.gov/pubmed/24753956>
- White, J., Law, J. R., Mackay, I., Chalmers, K. J., Smith, J. S. C., Kilian, A., & Powell, W. (2008). The genetic diversity of UK, US and Australian cultivars of *Triticum aestivum* measured by DARt markers

and considered by genome. *Theoretical and Applied Genetics*, 116(3), 439–453. <https://doi.org/10.1007/s00122-007-0681-3>

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Mondaini, A., Rosyara, U., Sehgal, D., & Dreisigacker, S. (2022). Selection signatures in the CIMMYT International Elite Spring and Semi-arid Wheat Yield Trials. *Plant Genome*, 15:e20165. <https://doi.org/10.1002/tpg2.20165>