ARTICLES
https://doi.org/10.1038/s41588-019-0496-6
nature genetics

Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics

Philomin Juliana[1], Jesse Poland[2], Julio Huerta-Espino[3], Sandesh Shrestha[2], José Crossa[1], Leonardo Crespo-Herrera[1], Fernando Henrique Toledo[1], Velu Govindan[1], Suchismita Mondal[1], Uttam Kumar[4], Sridhar Bhavani[1], Pawan K. Singh[1], Mandeep S. Randhawa[5], Xinyao He[1], Carlos Guzman[1,6], Susanne Dreisigacker[1], Matthew N. Rouse[7], Yue Jin[7], Paulino Pérez-Rodríguez[8], Osval A. Montesinos-López[9], Daljit Singh[2], Mohammad Mokhlesur Rahman[2,10], Felix Marza[11] and Ravi Prakash Singh[1*]

Bread wheat improvement using genomic tools is essential for accelerating trait genetic gains. Here we report the genomic predictabilities of 35 key traits and demonstrate the potential of genomic selection for wheat end-use quality. We also performed a large genome-wide association study that identified several significant marker–trait associations for 50 traits evaluated in South Asia, Africa and the Americas. Furthermore, we built a reference wheat genotype–phenotype map, explored allele frequency dynamics over time and fingerprinted 44,624 wheat lines for trait-associated markers, generating over 7.6 million data points, which together will provide a valuable resource to the wheat community for enhancing productivity and stress resilience.

Wheat is an important cereal that is the staple food for more than 2.5 billion people globally[1]. Beyond its nutritional and health benefits[2], wheat contributes substantially to food security by providing 20% of dietary calories and protein worldwide[3]. Wheat is also more widely cultivated than any other crop[4] on an area of 220 million hectares with an annual production of 722.4 million metric tons[5]. However, the low annual rate (0.9%) of yield increase[6], stagnating yields[7] and the impacts of diseases[8], climate change[9], drought and heat stresses[10] leading to a decreased yield remain key challenges to wheat production. Hence, to accelerate wheat breeding for higher yield potential and stress resilience, integration of genomic tools that can facilitate accurate selection and provide insights into the molecular basis of key traits is essential.

A genomics-based breeding strategy that has transformed the dairy industry is genomic selection, in which the genomic-estimated breeding values obtained from genome-wide molecular markers are used for the selection of superior animals[11,12]. The potential of genomic selection for wheat improvement is substantial, as it can facilitate more accurate selection, reduce the cycle time and phenotyping costs and subsequently accelerate genetic gains from selection, especially for complex traits of low heritability[13,14]. However, the practical implementation of genomic selection in applied wheat breeding programs is limited, with few comprehensive assessments of the genomic predictabilities of various traits that breeders collectively select for. In addition, the cost of genotyping is a key constraint to large-scale implementation of genomic selection, and it is important to evaluate a cost-effective genotyping technology that small breeding programs and developing countries can also use to screen their lines.

An extensive understanding of the genetic architecture of key traits is critical for making accurate selection decisions and for combining desired allelic combinations. In this regard, genome-wide association studies (GWAS) are a powerful strategy for dissecting the genetic basis of complex traits and identifying marker–trait associations based on the linkage disequilibrium (LD) between a marker and the causal polymorphism[15,16]. Although a substantial number of wheat GWAS have been reported, several studies involving complex traits are underpowered owing to the small size of the association panels used, and there are few studies that focus on multiple traits. These factors highlight the necessity for a holistic multiple-traits GWAS in a large panel, that would help to gain insight into the co-localization of loci for different traits. In addition, for complex traits like grain yield (GY) that are highly influenced by the environment, multi-environment GWAS[17,18] are pivotal to understand the genetic basis of GY stability across environments.

The recent availability of the International Wheat Genome Sequencing Consortium's reference sequence (RefSeq v.1.0)[19] of bread wheat has created new opportunities for genomics-based breeding in wheat. The wealth of genomic information from the RefSeq v.1.0, coupled with extensive phenotyping information from the global wheat trials of the International Maize and Wheat Improvement Center (CIMMYT) could serve as a roadmap for future wheat improvement. CIMMYT plays a key role in developing

[1]International Maize And Wheat Improvement Center (CIMMYT), Texcoco, Mexico. [2]Wheat Genetics Resource Center, Department of Plant Pathology, Kansas State University, Manhattan, KS, USA. [3]Campo Experimental Valle de Mexico, Instituto Nacional de Investigaciones Forestales, Agricolas y Pecuarias (INIFAP), Chapingo, Mexico. [4]Borlaug Institute for South Asia (BISA), New Delhi, India. [5]International Maize and Wheat Improvement Center (CIMMYT), Nairobi, Kenya. [6]Departamento de Genética, Escuela Técnica Superior de Ingeniería Agronómica y de Montes, Campus de Rabanales, Universidad de Córdoba, Cordoba, Spain. [7]United States Department of Agriculture, Agricultural Research Service (USDA-ARS), Cereal Disease Laboratory and Department of Plant Pathology, University of Minnesota, St Paul, MN, USA. [8]Colegio de Post graduados, Montecillos, Mexico. [9]Facultad de Telematica, Universidad de Colima, Colima, Mexico. [10]Regional Agricultural Research Station, Bangladesh Agricultural Research Institute (BARI), Jamalpur, Bangladesh. [11]Instituto Nacional de Innovación Agropecuaria y Forestal (INIAF), La Paz, Bolivia. *e-mail: r.singh@cgiar.org

global wheat germplasm with improved yield potential, disease resistance, abiotic stress tolerance, end-use quality, broad adaptation to diverse environments and has a major share in wheat varieties released worldwide[20]. Hence, we aligned 78,606 genotyping-by-sequencing (GBS) markers[21] for 44,624 lines that included five panels of wheat breeding lines from CIMMYT's first-year yield trials (YTs) evaluated between 2014 and 2018 to the RefSeq v.1.0. The anchored markers were used in genomic prediction models for a subset of 3,485 lines phenotyped for 35 key traits, to understand the impact of genomic coverage and training populations on the genomic predictabilities of traits. We also performed GWAS to dissect the genetic architecture of 50 traits, evaluated in South Asia, Africa and the Americas, and anchored the significant markers to a reference wheat genotype–phenotype map, aligned to the RefSeq v.1.0. Furthermore, we generated the genomic fingerprints of 44,624 wheat lines for trait-associated markers and examined the marker allele frequency dynamics to characterize the role of selection in shaping patterns of allelic variation over time.

Results

Phenotyping data overview. The key trait analyzed in this study was GY, which was evaluated in the irrigated environments of Obregon (Mexico), Afghanistan, Bangladesh, Canada, Egypt, Bahawalpur (Pakistan), Islamabad (Pakistan), Ludhiana (India), Pusa (India), Morocco and Sudan, and the drought-stressed, early-sown heat-stressed and late-sown heat-stressed environments of Obregon (Mexico). We also analyzed agronomic traits including days to heading (DTHD), days to maturity (DTMT), lodging and plant height. Disease resistance was analyzed as field resistance to stem rust race TTKSK/Ug99 and its derivatives, yellow rust evaluated in Mexico and Ludhiana, *Septoria tritici* blotch (STB), spot blotch and wheat blast, and seedling resistance to stem rust races QTHJC, RCRSC, RKRQC, TKTTF, TPMKC, TRTTF and TTTTF and to leaf rust race MBJ/SP. The end-use quality traits analyzed included alveograph P/L (the ratio of the peak height to the length of the alveogram), alveograph W (work value), flour protein content, flour sedimentation, flour yield, grain color, grain hardness, grain protein content (GPC), loaf volume, mixing time, test weight and thousand-kernel weight (TKW) (Supplementary Data 1, Supplementary Table 1 and Supplementary Fig. 1a–d). The phenotypic correlations (Supplementary Fig. 2a) indicated high positive correlations between traits like DTHD and DTMT, alveograph W and mixing time, and flour protein content and GPC.

Genomic predictabilities of traits. We used four panels of wheat breeding lines comprising CIMMYT's second-year elite yield trials (EYTs) evaluated during the crop seasons of 2014–2017 for genomic predictions. These lines were genotyped using GBS[22,23] and from the set of 78,606 unfiltered markers (Supplementary Fig. 2b), subsets of markers with less than 70%, 50% and 10% missing data were used for genomic predictions. The genomic coverage associated with the subsets clearly showed a decreasing trend towards the proximal centromeric regions with stringent filtering for missing data (Fig. 1), and thereby served as ideal sets for evaluating the effect of genomic coverage on the predictabilities of traits (16,072 markers showed high coverage, 9,285 markers had moderate coverage and 2,253 markers showed low coverage). Genomic prediction accuracies were obtained with the genomic best linear unbiased prediction (GBLUP) approach using fivefold cross-validations within each EYT panel, in which folds comprising 153–196 lines were predicted from four other folds comprising 613–784 lines, and predictions across EYT panels, for which 766–980 lines in a panel were predicted from three other panels comprising 2,505–2,719 lines. We observed similar accuracies at the different levels of genomic coverage, with the high-coverage marker subset providing an average increase of only 0.02 ± 0.02 (mean ± s.d.) in accuracy compared to the low-coverage subset, in both cross-validations and prediction across panels, across all traits (Fig. 2). To further understand the impact of marker densities, we filtered the marker subset with 10% missing data for pairwise correlations greater than 0.8, 0.5 and 0.3, and observed that the subsets after filtering for correlations greater than 0.5 (374–504 markers), resulted in an average decrease of only 0.05 ± 0.03 in accuracy both within and across panels. However, when markers filtered for pairwise correlations less than 0.3 were used (77–160 markers), it led to an average decrease of 0.23 ± 0.05 in prediction accuracy within panels and an average decrease of 0.13 ± 0.06 in prediction accuracy across panels (Fig. 2).

The traits grain color, seedling and field resistance to stem rust, mixing time, alveograph W, flour sedimentation, loaf volume, flour protein content, GPC and TKW had the highest genomic predictabilities (0.60–0.85), whereas all of the other traits were moderately predictable in the cross-validations. In predictions across panels, the traits with high predictabilities were seedling resistance to stem rust races QTHJC, TTTTF, TKTTF and TPMKC, grain color, mixing time, alveograph W, field resistance to stem rust and flour sedimentation with an average decrease of only 0.07 ± 0.05 in accuracy from the corresponding cross-validation accuracies. However, we observed low across-panel predictabilities of traits like GY and DTMT in all environments, DTHD in the early-sown heat-stressed environment of Mexico, DTHD in the irrigated environment of Mexico, STB, spot blotch, stem rust resistance to race RCRSC and race RKRQC and field resistance to yellow rust, with an average decrease of 0.20 ± 0.06 from the corresponding cross-validation accuracies. We also evaluated the Bayes B approach for predictions and observed that it resulted in similar accuracies as the GBLUP approach and the high genomic predictabilities of some traits were due to the large effects of the key trait-linked markers for those traits (Supplementary Data 2).

GWAS. A large EYT panel of 3,485 lines (for 24 traits) and several other panels (157–7,887 lines) were used to dissect the genetic architecture of 50 key trait–environment combinations using GWAS. Moderate population structure was present in the EYT panel as indicated by the first two principal components and the three ancestral sub-populations (Supplementary Fig. 3a,b). We identified hundreds of significant marker–trait associations (Figs. 3 and 4) after Bonferroni correction for multiple testing and report the marker *P* values, additive effects and percentage variation explained by each marker (Supplementary Table 2). We also used the LD between markers to delineate about 138 quantitative trait loci (QTLs) and sub-QTLs associated with various traits (Supplementary Data 3 and Supplementary Tables 3, 4 and 5). In addition, we performed a systematic review of literature to determine the relative positions of the QTLs identified in this study and observed that 44.4% of the QTLs coincided with previously reported QTLs or genes, and the remaining 55.6% QTLs were novel. The most significant trait-associated markers in each chromosome and previously reported genes or QTLs in proximity to the significant markers were then included in a reference genotype–phenotype map aligned to the RefSeq v.1.0 (Fig. 5).

Genomic regions associated with GY. We identified 31 QTLs that were significantly associated with GY in 14 locations, of which *Qcim.2A.1*, *Qcim.3B.2*, *Qcim.6A.7* and *Qcim.4D.1* were significant in five or more environments and could potentially confer GY stability across environments. Among these, *Qcim.2A.1* was located in the 2NS translocation[24], *Qcim.3A.3* flanked the grain size and TKW-linked *Tags5-3A* gene[25] and a GY QTL[26], *Qcim.5B.1* coincided with two loci that are linked to a heat-susceptibility index[27] and to the ferritin gene, *Tafer-5B*, which enhances tolerance to heat stress[28], and *Qcim.7B.2* was located near the sucrose synthase gene *Tasus1-7B*, which is associated with TKW[29].

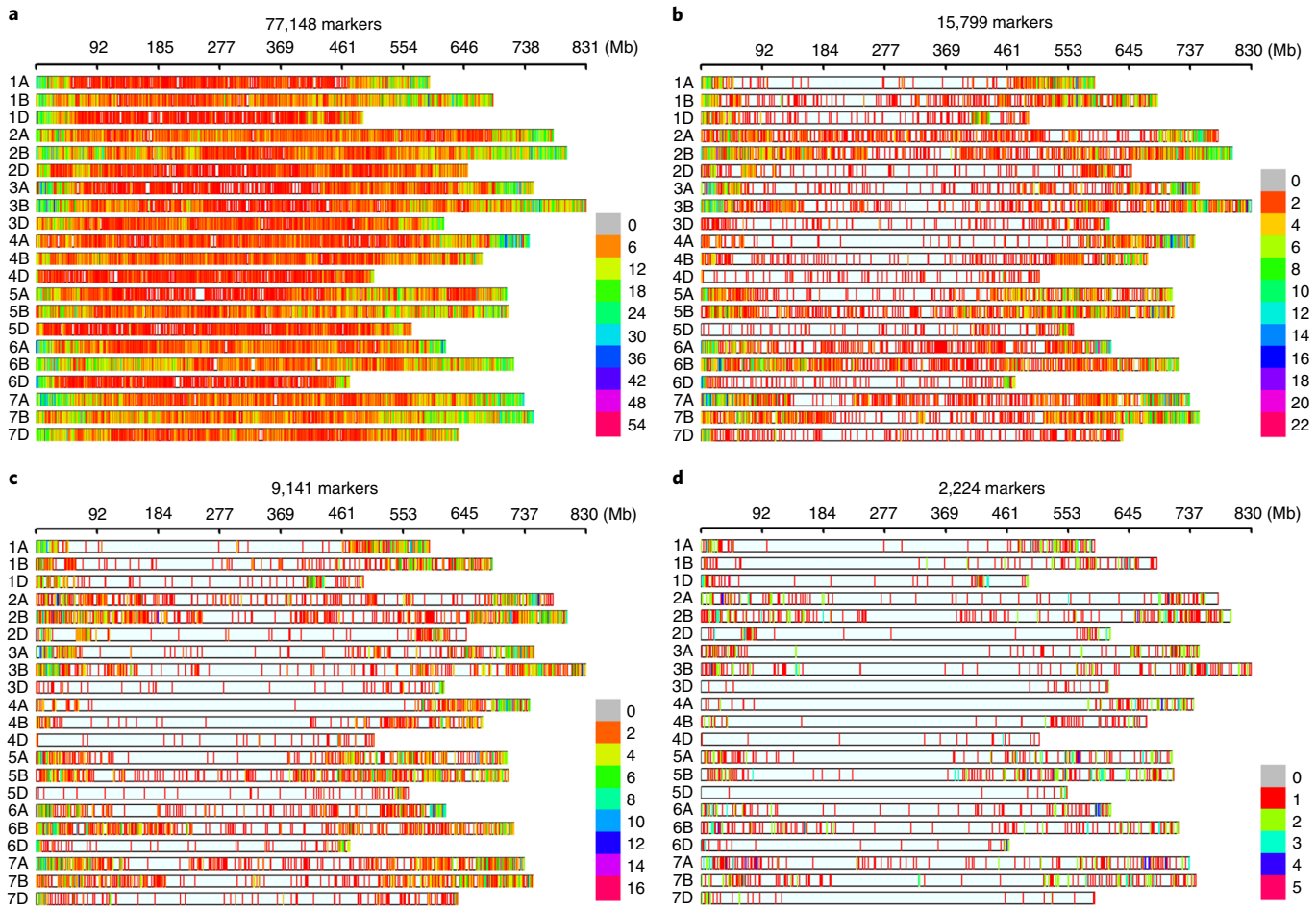

Fig. 1 | Distribution of GBS markers in the bread wheat genome. a–d, The 77,148 markers denote the unfiltered set (**a**), and subsets of 15,799 (**b**), 9,141 (**c**) and 2,224 (**d**) markers denote filtered sets with 70%, 50% and 10% of missing data, respectively (the unaligned markers are not included). The color key with marker densities indicates the number of markers within a window size of 1 Mb. The genomic coverage associated with the marker subsets shows a decreasing trend toward the proximal centromeric regions with stringent filtering for missing data.

Genomic regions associated with agronomic traits. We identified 15 QTLs that were significantly associated with the traits DTHD and DTMT, including seven common QTLs. Among these QTLs, *Qcim.2B.2* was in the location of the *Photoperiod-B1* (*Ppd-B1*) gene[30] and *Qcim.3A.2* was 6.8 Mb away from *Xbarc45*, which is associated with ear emergence[31]. On the homeologous groups of chromosome 5, *Qcim.5A.4*, *Qcim.5B.6* and *Qcim.5D.2* flanked the cloned *Vrn-A1*, *Vrn-B1* and *Vrn-D1* genes that control vernalization in wheat[32]. Similarly, on the homeologous groups of chromosome 7, *Qcim.7A.4* was in the same location as the QTL linked to ear emergence[31], *Qcim.7B.1* was 1.25 Mb away from the vernalization gene *Vrn-B3*[33] and *Qcim.7D.2* was 9.8 Mb proximal to the *Vrn-D3* gene[33]. Lodging and plant height were associated with ten and seven QTLs, respectively, with two overlapping QTLs, *Qcim.2A.1* and *Qcim.3B.6*, that were in the location of the 2NS translocation and the *Tamyb10-B1* gene[34], respectively.

Genomic regions associated with disease resistance. We identified 27 QTLs that were associated with disease resistance, of which 10 were novel QTLs. Seedling resistance to leaf rust race MBJ/SP was conferred by *Qcim.1D.2*, which was 1.3 Mb from the *Lr42* gene-linked marker[35] and *Qcim.2B.1*, which was 0.2 Mb proximal to the *Lr16* gene-linked marker[36]. Seedling resistance to stem rust races QTHJC, RCRSC, RKRQC and TPMKC was conferred by 3–4 QTLs, with *Qcim.2A.1* and *Qcim.3A.1* associated with all of these stem rust races. We also observed the co-occurrence of other identified QTLs with previously reported genes: *Qcim.2D.1* was 0.7 Mb away from the *Sr6* gene[37], which is effective against stem rust races TPMKC and QTHJC, *Qcim.6B.10* flanked the *Sr11* gene, which is effective against race TKTTF[38], *Qcim.4A.5* was in the same position as the *Sr7a* gene, which is effective against race TTTTF[39,40], and a QTL effective against race RCRSC[37] and *Qcim.6A.2* was in the location of the *Sr8a* gene[41], which is effective against stem rust race TRTTF.

Field resistance to stem rust was associated with seven QTLs, of which *Qcim.3B.1* and *Qcim.3B.2* flanked the *Sr2* gene, which is known to confer durable rust resistance[42] and *Qcim.7D.1* was 0.4 Mb distal to the cloned broad-spectrum resistance gene, *Lr34-Yr18-Sr57*[43]. The QTL *Qcim.3B.4* is likely to be in the location of the *Sr12* gene[44] or another gene important for field resistance to stem rust that co-localizes with *Sr12*[45], both of which are in the interval of the significant markers. The QTLs on chromosome 6AS with the highest effect on field resistance to stem rust were novel sources of resistance, and were therefore explored further using a fine-mapping approach. The biparental population for fine-mapping was developed using Kenya Fahari/2*Kachu as the resistant parent (which was the most likely donor of the 6AS QTL) and the most significant marker in the fine-mapping analysis corresponded to the location of *Qcim.6A.3*. We also observed that the 6AS telomeric region exhibited high recombination, and the peri-centromeric region in several progenies was inherited from the susceptible

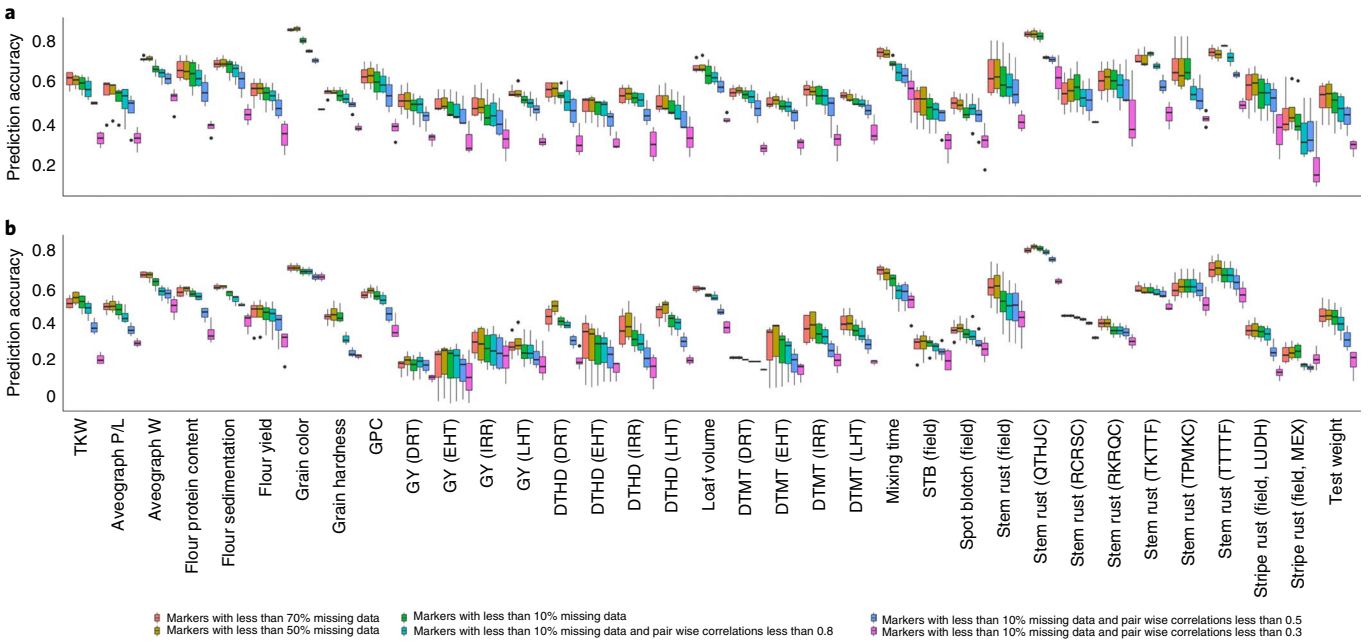

Fig. 2 | Genomic prediction accuracies for DTHD, DTMT and GY in different environments, disease resistance and end-use quality traits, within and across panels at different marker densities using the genomic-best linear unbiased prediction approach. a, The within-panel accuracies are the average accuracies obtained from fivefold cross-validations, where folds comprising 153–196 lines were predicted from four other folds comprising 613–784 lines, within each EYT evaluated during the 2013–2014, 2014–2015, 2015–2016 and 2016–2017 seasons using markers with less than 70% missing data (12,798–14,260 markers in the four individual seasons), markers with less than 50% missing data (7,737–8,586 markers in the four individual seasons), markers with less than 10% missing data (1,290–1,889 markers in the four individual seasons), markers with less than 10% missing data and pairwise correlations less than 0.8 (781–958 markers in the four individual seasons), markers with less than 10% missing data and pairwise correlations less than 0.5 (374–447 markers in the four individual seasons) and markers with less than 10% missing data and pairwise correlations less than 0.3 (77–97 markers in the four individual seasons). **b**, The across-panel accuracies are the average prediction accuracies obtained from predicting each EYT panel of 766–980 lines from three other panels of 2,505–2,719 lines, except for seedling disease resistance, using markers with less than 70% missing data (16,072 markers), markers with less than 50% missing data (9,285 markers), markers with less than 10% missing data (2,253 markers), markers with less than10% missing data and pairwise correlations less than 0.8 (1,091 markers), markers with less than 10% missing data and pairwise correlations less than 0.5 (504 markers) and markers with less than 10% missing data and pairwise correlations less than 0.3 (160 markers). The upper and lower hinges of the boxes in the box plot indicate the first (25th percentile) and third (75th percentile) quartiles, the center line indicates the median (50th percentile) corresponding to the prediction accuracies across evaluation panels and the upper and lower whiskers extend from the hinge to the largest or smallest values that are less than or equal to the hinge and 1.5× the interquartile range, respectively. TKW, thousand kernel weight; GPC, grain protein content; GY, grain yield; DRT, drought stressed; EHT, early-sown heat stressed; IRR, irrigated; LHT, late-sown heat stressed; DTHD, days to heading; DTMT, days to maturity; STB, *Septoria tritici* blotch; LUDH, Ludhiana; MEX, Mexico. Except for the smallest marker subset with 10% missing data and pairwise correlations less than 0.3, which led to an average decrease of 0.23 ± 0.05 in prediction accuracy within panels and an average decrease of 0.13 ± 0.06 in prediction accuracy across panels, all other marker subsets resulted in similar accuracies or had only a marginal difference in accuracies in both cross-validations and prediction across panels, across all traits.

parent, Apav (Supplementary Fig. 4a–c). Field resistance to yellow rust in Mexico was associated with the *Qcim.2A.1* or the 2NS translocation and *Qcim.3A.1*, which were also associated with wheat blast. However, for yellow rust in Ludhiana (India), resistance was associated with two QTLs on chromosome 2B and the presence of the *Yr31* flanking marker[46] near the most significant marker confirms that one or both of the QTLs indicate *Yr31* resistance, which is present in several CIMMYT wheat cultivars[47,48]. Resistance to STB and spot blotch was associated with the centromeric region of chromosome 2A, and spot blotch was also associated with *Qcim.3B.3*, which is probably the *Sb3* gene that was 1 Mb distal to this QTL[49].

Genomic regions associated with end-use quality. We identified 60 QTLs associated with end-use quality, of which 36 were not previously reported. The trait grain color had the simplest genetic architecture and was associated with *Qcim.3A.4*, which was 2.9 Mb away from the red grain color MYB-type transcription factor *Tamyb10-A1*[34], and *Qcim.3B.6*, which flanked the *Tamyb10-B1* gene[34]. Grain hardness was associated with *Qcim.5D.1*, which was 8.6 Mb distal to the markers linked to the two grain-hardness-determining puroindoline genes, located in the *Hardness* locus[50]. The traits alveograph W and mixing time were associated with several common loci, some of which were also associated with loaf volume. Among them, *Qcim.1A.1* was in the location of the gamma-gliadin, omega-gliadin and low-molecular-weight gluten genes encoded by the tightly linked *Gli-A1* and *Glu-A3* loci, which are associated with dough strength and mixing time[51–53]; *Qcim.1B.1*–*Qcim.1B.3* were in the location of the gamma-gliadin (*Gli-B1*) and low-molecular-weight gluten genes (*Glu-B3*) that affect dough strength and bread-making quality[54]; *Qcim.1B.6* was in the location of the high-molecular-weight *Glu-B1* gene associated with dough mixograph parameters[55]; *Qcim.1D.1* was 0.005 Mb away from the low-molecular-weight gluten gene (*Glu-A3*) linked to gluten strength-related parameters[56]; and *Qcim.1D.3* coincided with the high-molecular-weight *Glu-D1* gene that is known to affect dough strength and bread-making quality[57].

The traits flour protein content and GPC were associated with eight common loci, of which *Qcim.6A.5* and *Qcim.6B.5* were associated

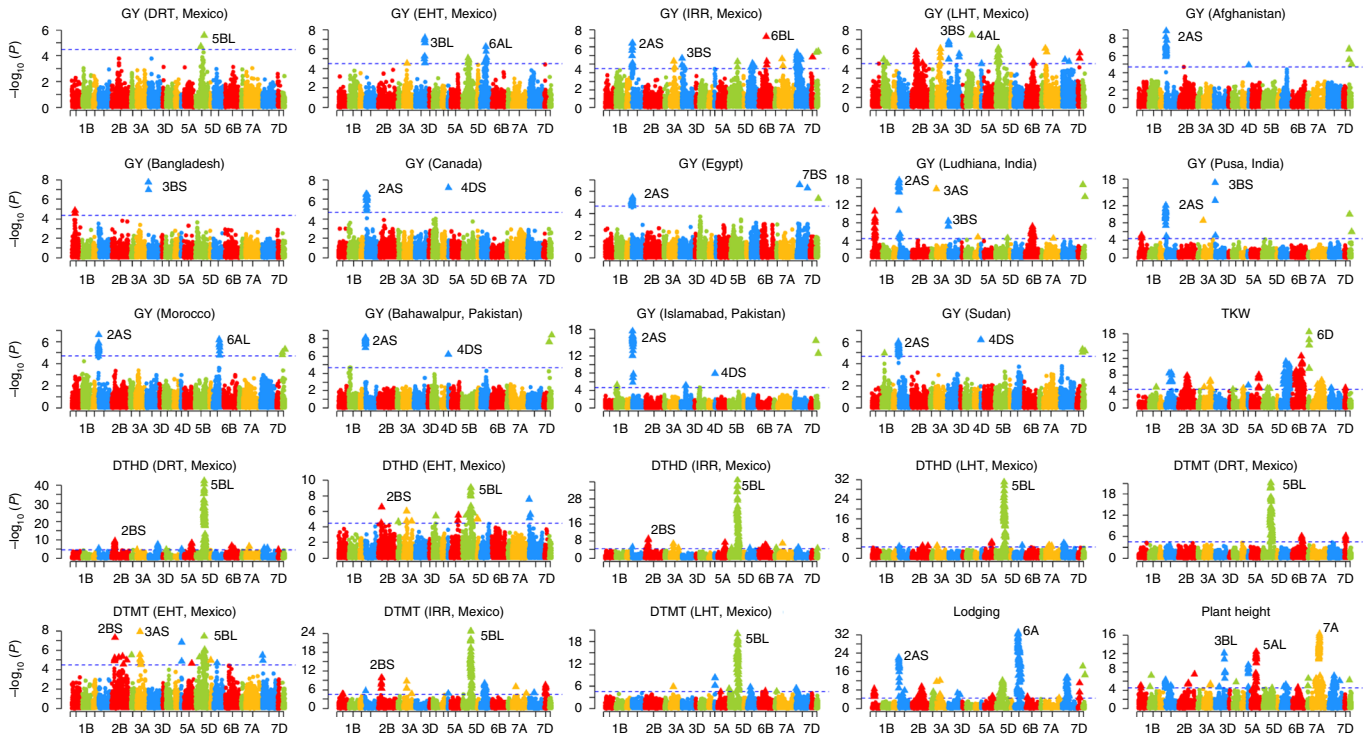

Fig. 3 | Marker–trait associations for GY and agronomic traits. Manhattan plots showing marker–trait associations for GY and agronomic traits from a genome-wide association mapping study. A Bonferroni α level of 0.20 was used to correct for multiple testing and identify significant markers. The 2AS chromosomal region was significantly associated with grain yield in several environments, whereas the 5BL and 2BS chromosomal regions were associated with days to heading and maturity in several environments.

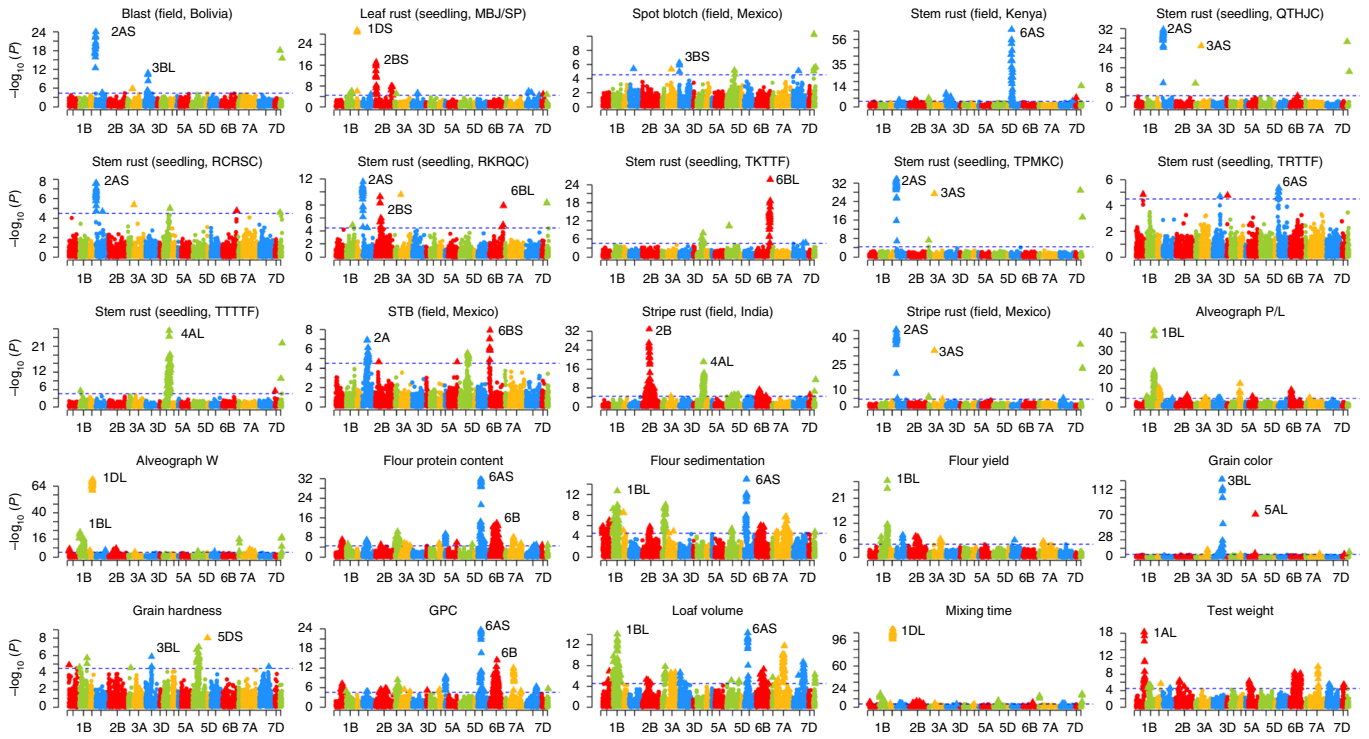

Fig. 4 | Marker–trait associations for disease resistance and quality-related traits. Manhattan plots showing marker–trait associations for disease resistance and quality-related traits from a genome-wide association mapping study. A Bonferroni α level of 0.20 was used to correct for multiple testing and identify significant markers. The 2AS chromosomal region was associated with field resistance to blast and stripe rust (India) and seedling resistance to race QTHJC, race RCRSC, race RKRQC and race TPMKC. A region on chromosome 1DL was associated with Alveograph W and mixing time, and regions on chromosomes 6AS and 6B were associated with both grain and flour protein content.

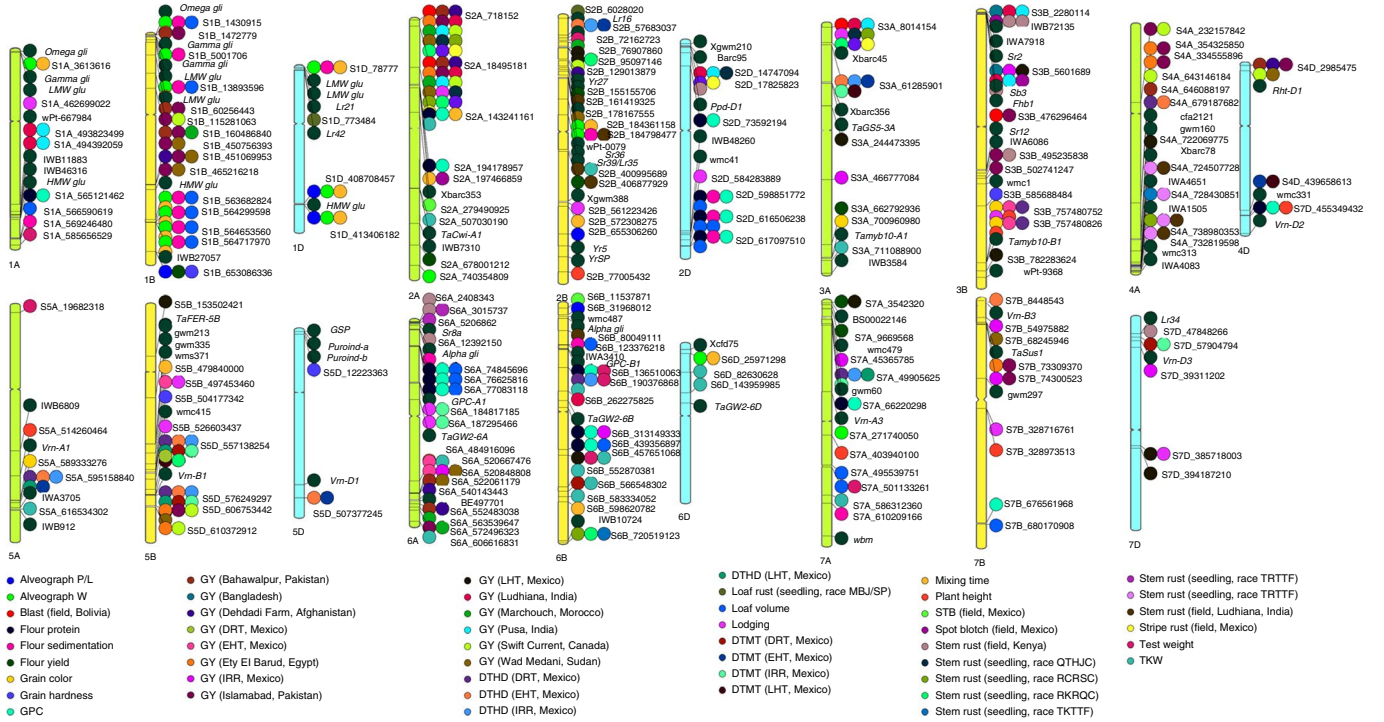

Fig. 5 | The reference genotype–phenotype map. A reference genotype–phenotype map with the most significant trait-associated markers in each chromosome aligned to the reference sequence of bread wheat (RefSeq v.1.0), along with previously reported genes, QTLs or linked markers. HMW, high molecular weight; LMW, low molecular weight.

with the alleles at the *Gpc-A1* and *Gpc-B1* loci[58,59], located in those intervals. Both alveograph P/L and flour yield were associated with *Qcim.1B.7*, while flour sedimentation and loaf volume were associated with *Qcim.1B.1*, *Qcim.1B.3*, *Qcim.1B.6* and *Qcim.2D.5*. Test weight was associated with five QTLs, one of which coincided with a previously reported QTL[60]. Finally, TKW was associated with ten QTLs including *Qcim.2A.2*, which was 1 Mb away from the cell wall invertase gene, *Tacwi-A1*, which is known to increase kernel weight[61]; *Qcim.3A.5* was localized to same location as the TKW gene *TaTGW6-A1*[62]; and *Qcim.6A.7.1*, *Qcim.6B.4* and *Qcim.6D.2* were close to the grain weight-associated *TaGW2* genes[63].

Genomic regions associated with different classes of traits. Several QTLs that were associated with more than one class of trait were identified. For example, the *Qcim.2A.1* or the 2NS translocation was significantly associated with GY in ten environments, seedling resistance to four stem rust races, yellow rust in Mexico, wheat blast, lodging, DTHD, DTMT and plant height. On the telomeric end of chromosome 3BS, two linked QTLs were associated with GY in several environments, stem rust and spot blotch. Another QTL, *Qcim.3B.6*, was associated with grain color (*Tamyb10-B1*), GY evaluated in the early-sown heat-stressed environment, DTHD in the drought-stressed environment, plant height and lodging. We also observed that *Qcim.6A.5* and *Qcim.6B.5* were associated with protein content, TKW and GY, and *Qcim.6B.7* (TKW) had sub-QTLs associated with DTMT.

Genomic fingerprinting and allele frequency dynamics. A large resource (Supplementary Table 6a–d) containing genomic fingerprints of 44,624 wheat lines for 195 trait-associated markers comprising 7.6 million data points was generated (Fig. 6, Supplementary Fig. 5a–c). The alleles with increasing effects on GY, agronomic and end-use quality traits, and decreasing effects on lodging and diseases (referred to as favorable alleles), along with their changes in frequencies due to selection from 2014 to 2018 were obtained (Supplementary Fig. 6a–d). For GY, 45 of the 65 fingerprinted markers had favorable alleles in more than 65% of the lines, with the highest increase over the years observed for marker S2D_14747094 and *Qcim.2A.1* (28–38% increase). A benchmark high-yielding CIMMYT-derived Mexican variety, BORLAUG100 F2014 had favorable alleles for GY in 35 of the 39 non-missing fingerprinted markers, which substantiates its high yield potential. We also analyzed the frequency dynamics of the favorable alleles for GY due to selection for 15 years (2003–2017) using 47 GY-associated markers in the globally distributed elite spring wheat yield trials (ESWYTs, Fig. 7). Considering the last three ESWYTs, we identified several markers at *Qcim.1B.4*, *Qcim.3B.8*, *Qcim.4A.1* and *Qcim.4D.1* that were fixed or near fixation for the favorable alleles for GY, with frequencies greater than 96%. The highest change in the frequency of favorable alleles from 2003 to 2015–2017 was observed for markers in *Qcim.2A.1* (24- to 28-fold increase), *Qcim.7B.2.1* and *Qcim.1B.5*. To test whether the change in the variance of the favorable alleles is what would be expected solely due to random genetic drift, we compared the expected variance of the favorable alleles over time with the observed allelic variance for all the markers. A higher variance of favorable alleles for the markers at *Qcim.5B.7*, *Qcim.2A.1*, *Qcim.4A.1*, *Qcim.4A.2*, *Qcim.7B.2.1* and *Qcim.1B.4* compared to the expected variance (Supplementary Table 7) indicated that artificial selection might have been a stronger force than genetic drift in driving the frequencies of favorable alleles higher.

The favorable alleles for lodging resistance at the most significant marker (S6A_187295466) had a very low frequency (0.13), whereas the allele that increased plant height at the most significant marker was present at a high frequency (0.74). Fingerprinting for DTHD and DTMT revealed that the frequencies of favorable alleles were high (greater than 76%) for markers S4D_439658613, S6A_76625816, S5A_595158840 and S2B_57683037. Disease resistance fingerprinting revealed that favorable alleles at markers S2D_14747094,

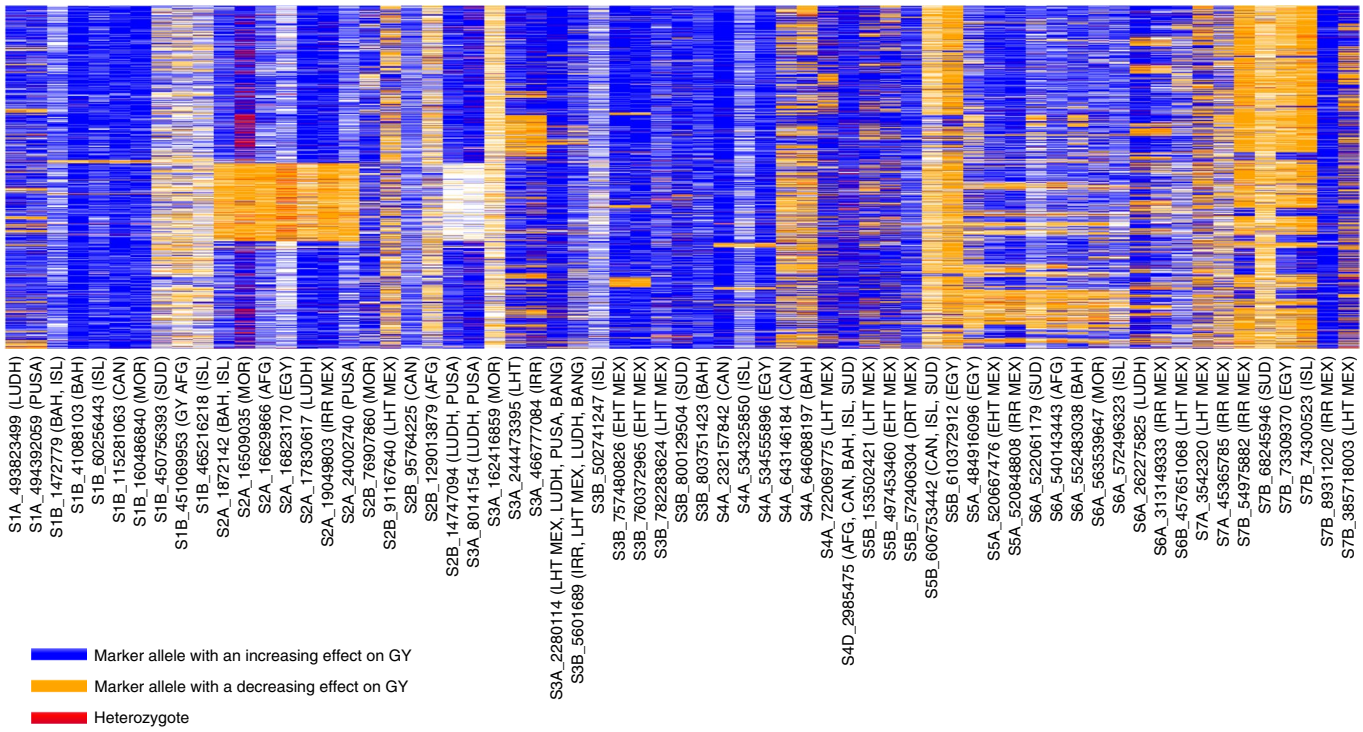

Fig. 6 | Genomic fingerprinting analysis for markers that are significantly associated with GY across different environments. Genomic fingerprinting analysis of 44,624 wheat lines from the International Wheat and Maize Improvement Center's bread wheat germplasm for markers that were significantly associated with GY in different environments, including the irrigated environments of Bahawalpur, Pakistan (BAH), Dehdadi Farms, Afghanistan (AFG), Jamalpur, Bangladesh (BANG), Swift Current, Canada (CAN), Ety El Barud, Egypt (EGY), Obregon, Mexico (IRR MEX), Islamabad, Pakistan (ISL), Ludhiana, India (LUDH), Marchouch, Morocco (MOR), Pusa, India (PUSA), Wad Medani, Sudan (SUD), and in the drought-stressed (DRT MEX), early-sown heat-stressed (EHT MEX) and late-sown heat-stressed (LHT MEX) environments of Obregon, Mexico. A high frequency of GY-favorable alleles was observed, and more than 65% of the lines had the favorable alleles in 45 of the 65 fingerprinted markers.

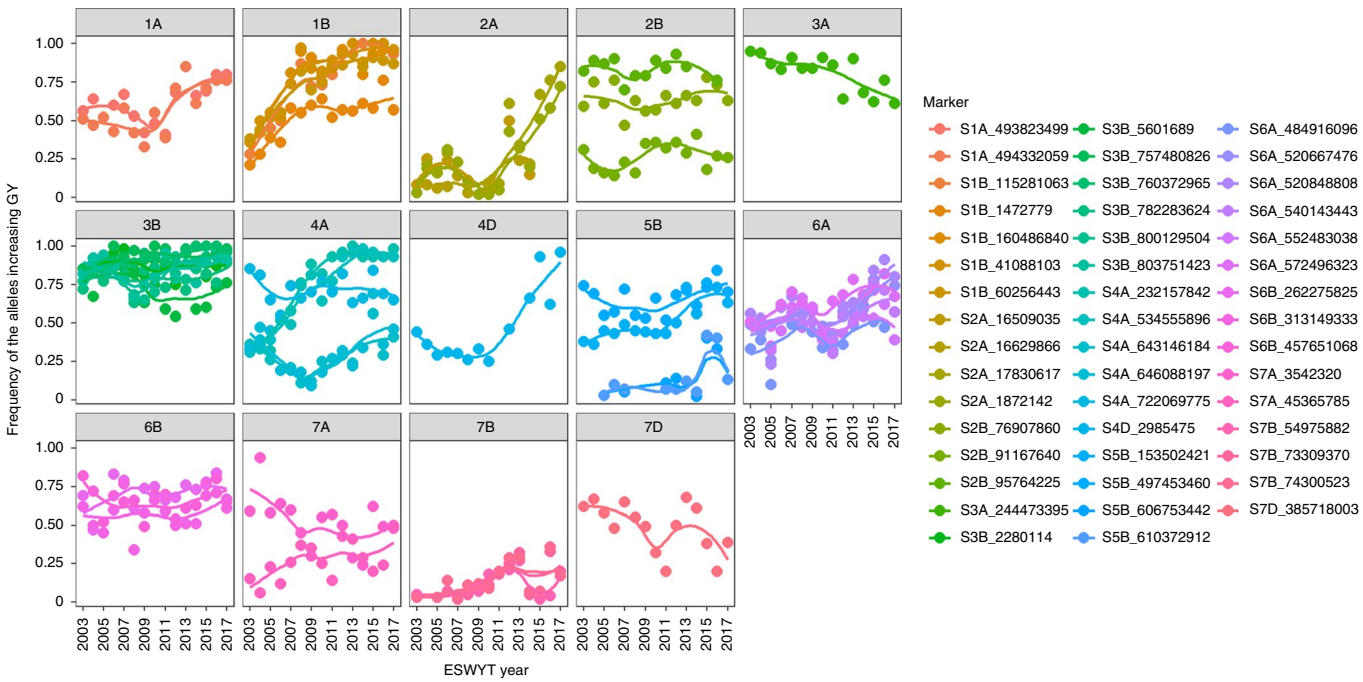

Fig. 7 | Trends in the favorable allele frequencies of 47 GY-associated markers in the ESWYTs over 15 years (2013–2017). The alleles of several markers in chromosomes 1A, 1B, 2A, 2D, 3B, 4A, 4D, 6A and 6B show patterns of fixation or near fixation.

S2B_734266688, S3A_8014154, S3B_2280114 (spot blotch), S2A_197466859 and QTLs *Qcim.2A.1*, *Qcim.3B.4*, *Qcim.6A.2* and *Qcim.6B.3* had a very high frequency (more than 85%) in several panels, whereas markers S1D_7734848, S3B_2280114 (stem rust) and S6A_7224185 had a very low frequency of favorable alleles (less than 10%). Fingerprinting for end-use quality traits showed that

the overall frequencies of favorable alleles were greater than 85% for markers S1D_413406182, S5B_479840000, S1D_408708457, S2B_155155706 and S6A_77083118. The markers with very low frequencies of favorable alleles, including S3A_711088900, S3B_757480826, S5D_12223363 and S5A_589333276, were associated with TKW, grain color and grain hardness. Over the years, an increasing trend (20–24%) in the favorable allele frequencies was observed for markers S2D_616506238, S2D_617097510 and S1B_653086336.

Discussion

We designed a comprehensive genomic prediction study that incorporates several economically important traits, thus providing a concrete foundation for future bread wheat improvement using genomic selection. Our results provide evidence that the across-population accuracies using different training populations will be lower than the within-population accuracies for low-heritable traits, like GY, that are subject to high environmental interactions[64–66] and for some foliar diseases and phenological traits. We also provide evidence that genomic selection will be a potential selection tool for several end-use quality-related traits, field resistance to stem rust and seedling disease resistance with a simple genetic architecture. As some of these traits involve laborious, time-consuming and expensive phenotyping, it is evident from the predictabilities reported in this study that genomic selection could be valuable for these traits, with respect to scaling-up phenotyping to large unphenotyped populations, making precise selections in early generations and reducing the generation interval through rapid cycling, in which cases genomic selection has added value to livestock breeding[67–70]. In addition, genomic selection for these traits has the attractive potential to render value over the phenotyping investment by minimizing replication costs for breeding programs and selecting superior varieties for target developing countries where expensive phenotyping infrastructure and resources may not be available.

An interesting observation in this study was that different levels of genomic coverage had minimal impact on the genomic predictabilities of traits, implying that once the genomic resolution has been reached in a high LD crop like wheat, marker number is no longer a critical limiting factor for prediction accuracies, as also observed in low-density marker panels in other species[71–73]. Although high genotyping costs impede the application of genomic selection on a global scale, we have successfully demonstrated the robustness of a relatively inexpensive genotyping technology (US$10 per sample) like GBS in achieving sufficient predictive accuracies for most traits that can be effectively used by breeding programs in developing countries.

This study reports the results of GWAS for an array of globally important traits that provide valuable insights into the genetic architecture and co-localization of loci for various traits, some of which are highlighted. An intriguing co-localization identified in this study was the association of the 2NS translocation from *Aegilops ventricosa* with GY in 10 different environments, disease resistance, lodging, phenology and plant height. Although this region has previously been reported to carry resistance to all three rusts[74] and wheat blast[75], we furthermore report its potential association with GY and stability, demonstrating its potential value for wheat breeding. An interesting region on the telomeric end of chromosome 3BS, which was favorably associated with GY and spot blotch, but unfavorably associated with stem rust, indicates that the two associated genes could be linked in repulsion or the same gene might have an opposite pleiotropic effect on these traits. A repulsion phase linkage between the *Sr2* gene and *Fhb1* gene in this region has been reported previously and the *Fhb1* gene-linked marker[76] was distal to this important region. However, it is unclear whether the stem rust-associated region is linked to the *Sr2* gene, because this gene is known to be widely deployed in the CIMMYT germplasm, whereas the allele frequencies at the two markers indicated that only 8–14% of the yield trial lines had the favorable allele for stem rust. A QTL on chromosome 3BL in the location of the red grain color gene *Tamyb10-B1* had alleles that led to red grain color, and increased plant height also associated with decreased DTHD in the drought-stressed environment and GY in the early-sown heat-stressed environment. This could be due to another locus in this interval or the pleiotropic effect of the MYB-transcription factor that is well-known to be associated with tolerance to abiotic stresses[77] and to increase plant height[78]. Two other QTLs associated with the alleles at the *Gpc-A1* and *Gpc-B1* loci had opposite effects on TKW and GY, respectively, consistent with negative associations between these traits[79]. The TKW-increasing allele located in the *TaGW2-6B* gene was also associated with early heading and maturity, substantiating previous reports[80].

The genotype–phenotype map anchoring significant markers to the RefSeq highlights the application of the RefSeq as a platform for comparing and validating GWAS results. It will also serve as a community resource providing opportunities for accelerating genomics-assisted wheat breeding through the targeted selection of desired regions. The genomic fingerprints of a large panel of lines comprising several key varieties cultivated worldwide provide an important leap in the understanding of the genetic basis of traits in superior varieties. For example, we provide evidence for the genetic basis of high yield in the variety BORLAUG100 F2014 and for the high resistance to stem rust in an old Kenyan variety (Kenya Fahari). The progressive trend and near fixation of favorable alleles for GY over 15 years not only show the effective impact of selection, but also emphasize the need for a continued effort of breeders to introduce novel sources of favorable alleles and the importance of integrating genomic tools in achieving accelerated enrichment of favorable alleles. Overall, the extensive datasets and results presented in this study provide a framework for breeders to design strategies to efficiently tackle alarming stresses like stem rust[81], wheat blast[82], STB[83] and climate change[84,85], while ensuring food-sustainability and security.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41588-019-0496-6.

Received: 11 March 2019; Accepted: 13 August 2019;
Published online: 23 September 2019

References

1. CGIAR Research Program on Wheat. *Wheat in the World* https://wheat.org/wheat-in-the-world/ (CRP, 2018).
2. Shewry, P. R. & Hey, S. J. The contribution of wheat to human diet and health. *Food Energy Secur.* **4**, 178–202 (2015).
3. Shiferaw, B. et al. Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. *Food Secur.* **5**, 291–317 (2013).
4. Curtis, T. & Halford, N. G. Food security: the challenge of increasing wheat yield and the importance of not compromising food safety. *Ann. Appl. Biol.* **164**, 354–372 (2014).
5. *FAOSTAT* http://www.fao.org/faostat/ (FAO, 2018).
6. Ray, D. K., Mueller, N. D., West, P. C. & Foley, J. A. Yield trends are insufficient to double global crop production by 2050. *PLoS ONE* **8**, e66428 (2013).
7. Ray, D. K., Ramankutty, N., Mueller, N. D., West, P. C. & Foley, J. A. Recent patterns of crop yield growth and stagnation. *Nat. Commun.* **3**, 1293 (2012).
8. Singh, R. P. et al. Disease impact on wheat yield potential and prospects of genetic control. *Annu. Rev. Phytopathol.* **54**, 303–322 (2016).
9. Wheeler, T. & von Braun, J. Climate change impacts on global food security. *Science* **341**, 508–513 (2013).
10. Zampieri, M., Ceglar, A., Dentener, F. & Toreti, A. Wheat yield loss attributable to heat waves, drought and water excess at the global, national and subnational scales. *Environ. Res. Lett.* **12**, 064008 (2017).

11. Meuwissen, T. H. E., Hayes, B. J. & Goddard, M. E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829 (2001).
12. Meuwissen, T., Hayes, B. & Goddard, M. Genomic selection: a paradigm shift in animal breeding. *Anim. Front.* **6**, 6–14 (2016).
13. Heffner, E. L., Sorrells, M. E. & Jannink, J.-L. Genomic selection for crop improvement. *Crop Sci.* **49**, 1–12 (2009).
14. Crossa, J. et al. Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.* **22**, 961–975 (2017).
15. Yu, J. & Buckler, E. S. Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.* **17**, 155–160 (2006).
16. Thornsberry, J. M. et al. *Dwarf8* polymorphisms associate with variation in flowering time. *Nat. Genet.* **28**, 286–289 (2001).
17. Quarrie, S. A. A. et al. A high-density genetic map of hexaploid wheat (*Triticum aestivum* L.) from the cross Chinese Spring × SQ1 and its use to compare QTLs for grain yield across a range of environments. *Theor. Appl. Genet.* **110**, 865–880 (2005).
18. Snape, J. W. et al. Dissecting gene × environmental effects on wheat yields via QTL and physiological analysis. *Euphytica* **154**, 401–408 (2007).
19. International Wheat Genome Sequencing Consortium (IWGSC) Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361**, eaar7191 (2018).
20. Lantican, M. A. et al. *Impacts of International Wheat Improvement Research, 1994–2014* (CIMMYT, 2016).
21. Poland, J. et al. Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome* **5**, 103–113 (2012).
22. Poland, J. A. & Rife, T. W. Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome* **5**, 92–102 (2012).
23. Elshire, R. J. et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**, e19379 (2011).
24. Helguera, M., Khan, I. A., Kolmer, J., Lijavetzky, D. & Dubcovsky, J. PCR assays for the *Lr37-Yr17-Sr38* cluster of rust resistance genes and their use. *Crop Sci.* **43**, 1839–1847 (2003).
25. Ma, L. et al. *TaGS5-3A*, a grain size gene selected during wheat improvement for larger kernel and yield. *Plant Biotechnol. J.* **14**, 1269–1280 (2016).
26. Rustgi, S. et al. Genetic dissection of yield and its component traits using high-density composite map of wheat chromosome 3A: bridging gaps between QTLs and underlying genes. *PLoS ONE* **8**, e70526 (2013).
27. Mason, R. E., Mondal, S., Beecher, F. W. & Hays, D. B. Genetic loci linking improved heat tolerance in wheat (*Triticum aestivum* L.) to lower leaf and spike temperatures under controlled conditions. *Euphytica* **180**, 181–194 (2011).
28. Zang, X. et al. Overexpression of wheat ferritin gene *TaFER-5B* enhances tolerance to heat stress and other abiotic stresses associated with the ROS scavenging. *BMC Plant Biol.* **17**, 14 (2017).
29. Hou, J. et al. Global selection on sucrose synthase haplotypes during a century of wheat breeding. *Plant Physiol.* **164**, 1918–1929 (2014).
30. Díaz, A., Zikhali, M., Turner, A. S., Isaac, P. & Laurie, D. A. Copy number variation affecting the *Photoperiod-B1* and *Vernalization-A1* genes is associated with altered flowering time in wheat (*Triticum aestivum*). *PLoS ONE* **7**, e33234 (2012).
31. Griffiths, S. et al. Meta-QTL analysis of the genetic control of ear emergence in elite European winter wheat germplasm. *Theor. Appl. Genet.* **119**, 383–395 (2009).
32. Yan, L. et al. Positional cloning of the wheat vernalization gene *VRN1*. *Proc. Natl Acad. Sci. USA* **100**, 6263–6268 (2003).
33. Yan, L. et al. The wheat and barley vernalization gene *VRN3* is an orthologue of *FT*. *Proc. Natl Acad. Sci. USA* **103**, 19581–19586 (2006).
34. Himi, E. & Noda, K. Red grain colour gene (R) of wheat is a Myb-type transcription factor. *Euphytica* **143**, 239–242 (2005).
35. Sun, X., Bai, G., Carver, B. F. & Bowden, R. Molecular mapping of wheat leaf rust resistance gene *Lr42*. *Crop Sci.* **50**, 59 (2010).
36. Kassa, M. T. et al. Highly predictive SNP markers for efficient selection of the wheat leaf rust resistance gene *Lr16*. *BMC Plant Biol.* **17**, 45 (2017).
37. Edae, E. A., Pumphrey, M. O. & Rouse, M. N. A genome-wide association study of field and seedling response to individual stem rust pathogen races reveals combinations of race-specific genes in North American spring wheat. *Front. Plant Sci.* **9**, 52 (2018).
38. Nirmala, J. et al. Markers linked to wheat stem rust resistance gene *Sr11* effective to *Puccinia graminis* f. sp. *tritici* race TKTTF. *Phytopathology* **106**, 1352–1358 (2016).
39. Turner, M. K., Jin, Y., Rouse, M. N. & Anderson, J. A. Stem rust resistance in 'Jagger' winter wheat. *Crop Sci.* **56**, 1719–1725 (2016).
40. Gao, L. et al. Genetic characterization of stem rust resistance in a global spring wheat germplasm collection. *Crop Sci.* **57**, 2575–2589 (2017).
41. Hiebert, C. W., Rouse, M. N., Nirmala, J. & Fetch, T. Genetic mapping of stem rust resistance to *Puccinia graminis* f. sp. *tritici* race TRTTF in the Canadian wheat cultivar harvest. *Phytopathology* **107**, 192–197 (2017).
42. Mago, R. et al. An accurate DNA marker assay for stem rust resistance gene *Sr2* in wheat. *Theor. Appl. Genet.* **122**, 735–744 (2011).
43. Krattinger, S. G. et al. A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. *Science* **323**, 1360–1363 (2009).
44. Rouse, M. N., Talbert, L. E., Singh, D. & Sherman, J. D. Complementary epistasis involving *Sr12* explains adult plant resistance to stem rust in Thatcher wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* **127**, 1549–1559 (2014).
45. Hiebert, C. W. et al. Major gene for field stem rust resistance co-locates with resistance gene *Sr12* in 'Thatcher' wheat. *PLoS ONE* **11**, e0157029 (2016).
46. Yang, E. N. et al. QTL analysis of the spring wheat 'Chapio' identifies stable stripe rust resistance despite inter-continental genotype × environment interactions. *Theor. Appl. Genet.* **126**, 1721–1732 (2013).
47. McDonald, D. B., McIntosh, R. A., Wellings, C. R., Singh, R. P. & Nelson, J. C. Cytogenetical studies in wheat XIX. Location and linkage studies on gene *Yr27* for resistance to stripe (yellow) rust. *Euphytica* **136**, 239–248 (2004).
48. Singh, R. P., William, H. M., Huerta-Espino, J. & Crosby, M. Identification and mapping of gene *Yr31* for resistance to stripe rust in *Triticum aestivum* cultivar Pastor. In *Proc. 10th International Wheat Genetics Symposium*. (eds Pogna N. E. et al.) 411–413 (Instituto Sperimentale per la Cerealicoltura, 2003).
49. Lu, P. et al. Fine genetic mapping of spot blotch resistance gene *Sb3* in wheat (*Triticum aestivum*). *Theor. Appl. Genet.* **129**, 577–589 (2016).
50. Morris, C. F. Puroindolines: the molecular genetic basis of wheat grain hardness. *Plant Mol. Biol.* **48**, 633–647 (2002).
51. Færgestad, E. M. et al. Relationships between storage protein composition, protein content, growing season and flour quality of bread wheat. *J. Sci. Food Agric.* **84**, 877–886 (2004).
52. Zhen, S. et al. Deletion of the low-molecular-weight glutenin subunit allele *Glu-A3a* of wheat (*Triticum aestivum* L.) significantly reduces dough strength and breadmaking quality. *BMC Plant Biol.* **14**, 367– (2014).
53. Bonafede, M. D., Tranquilli, G., Pflüger, L. A., Peña, R. J. & Dubcovsky, J. Effect of allelic variation at the *Glu-3/Gli-1* loci on breadmaking quality parameters in hexaploid wheat (*Triticum aestivum* L.). *J. Cereal Sci.* **62**, 143–150 (2015).
54. Wang, Y. et al. Low molecular weight glutenin subunit gene *Glu-B3h* confers superior dough strength and breadmaking quality in wheat (*Triticum aestivum* L.). *Sci. Rep.* **6**, 27182 (2016).
55. Cooper, J. K., Stromberger, J. A., Morris, C. F., Bai, G. & Haley, S. D. End-use quality and agronomic characteristics associated with the *Glu-B1al* high-molecular-weight glutenin allele in U.S. hard winter wheat. *Crop Sci.* **56**, 2348–2353 (2016).
56. Maucher, T., Figueroa, J. D. C., Reule, W. & Peņa, R. J. Influence of low molecular weight glutenins on viscoelastic properties of intact wheat kernels and their relation to functional properties of wheat dough. *Cereal Chem.* **86**, 372–375 (2009).
57. Guzmán, C. et al. Sources of the highly expressed wheat bread making (*wbm*) gene in CIMMYT spring wheat germplasm and its effect on processing and bread-making quality. *Euphytica* **209**, 689–692 (2016).
58. Uauy, C., Distelfeld, A., Fahima, T., Blechl, A. & Dubcovsky, J. A. A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. *Science* **314**, 1298–1301 (2006).
59. Avni, R. et al. Functional characterization of *GPC-1* genes in hexaploid wheat. *Planta* **239**, 313–324 (2014).
60. Assanga, S. O. et al. Mapping of quantitative trait loci for grain yield and its components in a US popular winter wheat TAM 111 using 90K SNPs. *PLoS ONE* **12**, e0189669 (2017).
61. Ma, D., Yan, J., He, Z., Wu, L. & Xia, X. Characterization of a cell wall invertase gene *TaCwi-A1* on common wheat chromosome 2A and development of functional markers. *Mol. Breed.* **29**, 43–52 (2012).
62. Hanif, M. et al. *TaTGW6-A1*, an ortholog of rice *TGW6*, is associated with grain weight and yield in bread wheat. *Mol. Breed.* **36**, 1 (2016).
63. Qin, L. et al. *TaGW2*, a good reflection of wheat polyploidization and evolution. *Front. Plant Sci.* **8**, 318 (2017).
64. Juliana, P. et al. Prospects and challenges of applied genomic selection—a new paradigm in breeding for grain yield in bread wheat. *Plant Genome* **11**, 180017 (2018).
65. Crossa, J. et al. Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity* **112**, 48–60 (2014).
66. Reif, J. C., Zhao, Y., Würschum, T., Gowda, M. & Hahn, V. Genomic prediction of sunflower hybrid performance. *Plant Breed.* **132**, 107–114 (2013).
67. Voss-Fels, K. P., Cooper, M. & Hayes, B. J. Accelerating crop genetic gains with genomic selection. *Theor. Appl. Genet.* **132**, 669–686 (2019).
68. Pryce, J. E. & Daetwyler, H. D. Designing dairy cattle breeding schemes under genomic selection: a review of international research. *Anim. Prod. Sci.* **52**, 107–114 (2012).

69. Hayes, B. J., Bowman, P. J., Chamberlain, A. J. & Goddard, M. E. Invited review: Genomic selection in dairy cattle: progress and challenges. *J. Dairy Sci.* **92**, 433–443 (2009).
70. García-Ruiz, A. et al. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proc. Natl Acad. Sci. USA* **113**, E3995–E4004 (2016).
71. Luan, T. et al. The accuracy of genomic selection in Norwegian red cattle assessed by cross-validation. *Genetics* **183**, 1119–1126 (2009).
72. Lenz, P. R. N. et al. Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of black spruce (*Picea mariana*). *BMC Genom.* **18**, 335 (2017).
73. Moser, G., Khatkar, M. S., Hayes, B. J. & Raadsma, H. W. Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers. *Genet. Sel. Evol.* **42**, 37 (2010).
74. Bariana, H. S. & Mcintosh, R. A. Cytogenetic studies in wheat. XV. Location of rust resistance genes in VPM1 and their genetic linkage with other disease resistance genes in chromosome 2A. *Genome* **36**, 476–482 (1993).
75. Cruz, C. D. et al. The 2NS translocation from *Aegilops ventricosa* confers resistance to the *Triticum* pathotype of *Magnaporthe oryzae*. *Crop Sci.* **56**, 990–1000 (2016).
76. Zhang, X., Rouse, M. N., Nava, I. C., Jin, Y. & Anderson, J. A. Development and verification of wheat germplasm containing both *Sr2* and *Fhb1*. *Mol. Breed.* **36**, 85 (2016).
77. Zhao, Y. et al. Characterization of wheat *MYB* genes responsive to high temperatures. *BMC Plant Biol.* **17**, 208 (2017).
78. Zhang, Y. et al. OsMPH1 regulates plant height and improves grain yield in rice. *PLoS ONE* **12**, e0180825 (2017).
79. Brevis, J. C. & Dubcovsky, J. Effects of the chromosome region including the Gpc-B1 locus on wheat grain and protein yield. *Crop Sci.* **50**, 93–104 (2010).
80. Su, Z., Hao, C., Wang, L., Dong, Y. & Zhang, X. Identification and development of a functional marker of *TaGW2* associated with grain weight in bread wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* **122**, 211–223 (2011).
81. Singh, R. P. et al. Emergence and Spread of new races of wheat stem rust fungus: continued threat to food security and prospects of genetic control. *Phytopathology* **105**, 872–884 (2015).
82. Cruz, C. D. & Valent, B. Wheat blast disease: danger on the move. *Trop. Plant Pathol.* **42**, 210–222 (2017).
83. Torriani, S. F. F. et al. *Zymoseptoria tritici*: a major threat to wheat production, integrated approaches to control. *Fungal Genet. Biol.* **79**, 8–12 (2015).
84. Tack, J., Barkley, A. & Nalley, L. L. Effect of warming temperatures on US wheat yields. *Proc. Natl Acad. Sci. USA* **112**, 6931–6936 (2015).
85. Trnka, M. et al. Adverse weather conditions for European wheat production will become more frequent with climate change. *Nat. Clim. Change* **4**, 637–643 (2014).

Acknowledgements

This research was supported by the Feed the Future project through the US Agency for International Development (USAID), under the terms of contract no. AID-OAA-A-13-00051 (J.P. and R.P.S.). The opinions expressed herein are those of the authors and do not necessarily reflect the views of the USAID. We thank the innovation laboratory at Kansas State University, the CGIAR Research Program on Wheat, the Indian Council of Agricultural Research (ICAR), the Australian Centre for International Agricultural Research (ACIAR), several national partners (Afghanistan, Bangladesh, Canada, Egypt, India, Morocco, Pakistan and Sudan) and field technicians for their support in generating the genotyping and phenotyping data.

Author contributions

P.J. drafted the manuscript and performed the analyses. R.P.S., J.P., J.H.-E. and J.C. planned the study and supervised the analysis. F.H.T., P.P.-R. and O.A.M.-L. performed some of the analyses. L.C.-H., V.G., S.M., U.K., S.B., P.K.S., M.S.R., X.H., C.G., M.N.R., Y.J., D.S., M.M.R. and F.M. generated the phenotyping data. S.D. performed the DNA extraction and S.S. called the marker polymorphisms.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0496-6.

Correspondence and requests for materials should be addressed to R.P.S.

Methods

Populations. The populations used include six panels of lines from CIMMYT's global bread wheat breeding program. Panel 1 comprises 3,485 lines from the EYTs or the second-year yield trials. The 3,485 lines include 766 lines from EYT 2013–2014, 775 lines from EYT 2014–2015, 964 lines from EYT 2015–2016 and 980 lines from EYT 2016–2017. Panel 2 comprises 2,142 lines from the South Asia bread wheat yield trials (SABWYT, a subset of the EYTs) or the third-year evaluations of GY in India and Bangladesh. This panel includes 536 lines from SABWYT 2014–2015, 531 lines from SABWYT 2015–2016, 535 lines from SABWYT 2015–2016 and 540 lines from SABWYT 2016–2017. Panel 3 comprises 1,956 lines from the international bread wheat screening nursery (IBWSN). Panel 4 comprises 7,887 lines from the first-year yield trials evaluated in 2014–2015. Panel 5 comprises 863 lines that were candidates for the stem rust resistance screening nurseries (SRRSNs) and subsets of the EYTs. This panel includes 219 lines evaluated in 2014, 179 lines evaluated in 2015, 234 lines evaluated in 2016 and 231 lines evaluated in 2017. Panel 6 comprises 676 lines from the elite spring wheat yield trials (ESWYTs) evaluated from 2003 to 2017.

The cohorts of breeding populations were developed using the selected-bulk breeding scheme, for which the selected plants in early generations were bulked until the head- or plant-row stage[64]. The selections from this pre-yield testing or head-row stage comprise the lines in the first-year yield trials. Genomic fingerprinting was carried out for the 44,624 lines that included 919 lines from EYT 2013–2014, 6,927 lines from the 2013–2014 yield trial, 8,762 lines from 2014–2015 yield trial, 9,041 lines from 2015–2016 yield trial, 9,034 lines from 2016–2017 yield trial, 8,088 lines from 2017–2018 yield trial, 1,384 lines from the IBWSN and 469 parents from CIMMYT's crossing blocks 2009–2018. A biparental mapping population (Apav/#1//Kenya Fahari/2*Kachu) of 200 individuals with the resistant parent Kenya Fahari/2*Kachu and the susceptible parent Apav/#1 was developed for the dissection of the genetic basis of stem rust resistance in Kenya Fahari/2*Kachu.

Phenotyping. *GY and agronomic traits.* GY for all of the 3,485 lines in panel 1 was evaluated in the Norman E. Borlaug Research station, Ciudad Obregon, Sonora, Mexico (27°29′N, 109°56′W), during four seasons (2013–2014 to 2016–2017). The lines in each season were sown in 39 trials, with each trial comprising 28 lines and two high-yielding check varieties (Kachu and BORLAUG100 F2014) in six blocks. In addition, all the entries were replicated thrice in four environments, which included the following conditions: (1) Optimally irrigated; the lines were sown in raised beds, during the optimum planting time (mid-November) in an optimally irrigated environment that received 500 mm of water in five irrigations. (2) Drought stressed; the lines were sown in the flat planting system, during the optimum planting time and grown with about 180 mm of water supplemented through drip irrigation. (3) Early-sown heat stressed; the lines were sown in raised beds, about 30 days before the optimum planting time and received optimal irrigation. (4) Late-sown heat stressed; the lines were sown on raised beds about 90 days after the optimal planting time, to naturally expose them to high-temperature stress during that time, and they received optimal irrigation. In addition, phenological traits like DTHD (number of days from germination to 50% of spike emergence) and DTMT (number of days from germination to 50% physiological maturity) in all the environments were recorded for the lines in panel 1. As GY in the different environments had moderate to high correlations with DTHD, the lines at the tails of the DTHD distributions were removed. Plant height (the height of the plant in cm) and lodging (ordinal scale from 0 to 5) were measured for the 7,887 lines in 2014–2015 yield trial with two replicates. The 2,142 lines from the SABWYTs were evaluated in the irrigated environments of the Borlaug Institute for South Asia in Ludhiana (30°54′ N, 75°51′ E) and Pusa (25° 59′ N, 85° 41′ E), India, in addition to Jamalpur (24° 55′ N, 89° 57′ E), Bangladesh with two replicates and the same field design as in Mexico. Some or all the 676 ESWYT lines were also evaluated for GY in the irrigated environments of Bahawalpur, Pakistan (28° 49′ N and 71° 39′ E), Dehdadi Farms, Afghanistan (36° 39′ N, 66° 59′ E), Swift Current, Canada (50° 17′ N, 107° 48′ W), Ety El Barud, Egypt (31° 7′ N, 30° 48′ E), Islamabad, Pakistan (33° 45′ N, 73° 6′ E), Marchouch, Morocco (33.55° 3′ N, 6.69° 38′ W) and Wad Medani, Sudan (14° 24′ N, 33° 29′E), with both CIMMYT and local check varieties during 2003–2017.

Disease response evaluations. Seedling response to leaf rust. Seedling response to leaf rust caused by *Puccinia triticina* Eriks. race MBJ/SP[86] was evaluated in CIMMYT's greenhouses at El Batan, Mexico for the 1,956 lines in panel 3. Inoculum preparation and inoculation were carried out as described previously[87] and the seedling responses were evaluated 10 days after inoculation, using a 0–4 scale[88].

Seedling response to stem rust. Seedling response to stem rust caused by *Puccinia graminis* Pers. f. sp. *tritici* was evaluated at the USDA-ARS Cereal Disease Laboratory. The seven *P. graminis* f. sp. *tritici* races that were evaluated in the SRRSN include QTHJC (isolate 75ND717C), RCRSC (isolate 77ND82A-1), RKRQC (isolate 99KS76A-1), TKTTF (isolate 13ETH18-1), TPMKC (isolate 74MN1409), TRTTF (isolate 06YEM34-1) and TTTTF (isolate 01MN84A-1)[89]. Among these, race TPMKC was used for evaluation of all four SRRSN candidates, races QTHJC and RCRSC in 2014 and 2015, race RKQQC in 2014, 2015 and 2016, race TRTTF in 2014, race TKTTF in 2015, 2016 and 2017, and race TTTTF in 2016 and 2017. Inoculum preparation and inoculation were carried out as described previously[90,91] and the infection types were scored on a 0–4 scale[92]. The leaf and stem rust seedling scores were linearized to a 0–9 scale for the analyses[93].

Field response to stem rust and yellow rust. The EYT lines were evaluated for field response to stem rust at the Kenya Agricultural and Livestock Research Organization, Njoro, Kenya (0° 19′ N, 35° 56′ E) during the 2013–2017 main and off-seasons with race TTKST in 2013–2016 and with a mixture of TTKST and TTKTT in 2017, both belonging to the Ug99 lineage[94]. Similarly, the biparental population was evaluated for field response to stem rust during the 2016 off-season for the races mentioned above. The EYTs were also evaluated for yellow rust (caused by *Puccinia striiformis* West.) response to the Mexican isolates Mex96.11, Mex08.13 and Mex14.191 at CIMMYT's research station, Toluca, Mexico (19° 17′ N, 99° 11′ W) as described previously[93] and to the predominant races collected from popular cultivars PBW343 (during 2013–2016) and HD2967 (during 2017) in Ludhiana, India. Rust response was scored twice or thrice between the early- and late-dough stages, at weekly to 10-day intervals after the severity of the susceptible checks reached 80%. The percentage of infected tissue (0–100%) was assessed using the modified Cobb scale[95].

Field response to STB. Field response of the EYT lines to STB, which is caused by *Zymoseptoria tritici* Desm., was evaluated at Toluca, Mexico during the 2013–2017 crop seasons as described previously[96], using the double-digit scale (00–99) for rating foliar diseases[97,98]. We performed three to four evaluations, and calculated the disease severity percentages using the double-digit scores, from which the area under the disease progression curve (AUDPC)[99] and the relative AUDPC were calculated.

Field response to spot blotch. Field response of the EYT lines to spot blotch caused by *Bipolaris sorokiniana* Sacc. was evaluated at CIMMYT's research station, Agua Fria, Mexico (19° 59′ N, 97° 50′ W), during the 2013–2017 crop seasons. The planting design was similar to that for STB, with the lines sown during November and harvested in March. A mixture of virulent races collected from leaf samples that were naturally infected in Agua Fria was used for inoculation[100] and the relative AUDPC was calculated, similar to STB.

Field response to wheat blast. The field response of 271 lines from the IBWSN to wheat blast caused by *Magnaporthe oryzae* Catt. was evaluated in the Department of Santa Cruz, Quirusillas, Bolivia in two replications. The lines were sown in the fourth week of December 2017 in 1-m double rows with 20-cm spacing between the rows. Local varieties Urubo and Atlax were used as resistant and susceptible checks, respectively. A locally collected *M. oryzae* isolate was used for field inoculation at anthesis, as well as 2 days after anthesis at a concentration of 50,000 spores ml^{-1} using a hand-held sprayer. Disease evaluation was performed at 21 days after the first inoculation on 10 spikes that were tagged at anthesis. The total and the infected number of spikelets were recorded for each of the 10 spikes, and then the wheat blast index was calculated using the formula: wheat blast index = incidence × severity, where incidence indicates the proportion of spikes with wheat blast infection and severity indicates the average percentage of infected spikelets.

End-use quality. The end-use quality traits were evaluated for all of the lines in the EYTs, with slightly modified methods from the American Association of Cereal Chemists (AACC)[101] standards. A mixograph (National Manufacturing Company) according to AACC method 54–40A[101] and the Chopin alveograph (Tripette & Renaud), AACC method 54–30A, were used to analyze dough rheological properties. The mixograph was then used to obtain the optimal mixing time (min), whereas the alveograph was used to measure the dough strength, or the work value under the curve (alveograph W) and the tenacity versus extensibility, which is the ratio of the height to the length of the curve (alveograph P/L, $mm\,mm^{-1}$). The flour SDS sedimentation volume (ml) was determined using 1 g of flour[102], and bread loaf volume (cm^3) was assessed by the rape seed displacement method according to the AACC method 10–05.01[101], from pup loaves that were baked as pan bread using the slightly modified AACC method 10–09[101]. The optimal water absorption for the mixograph, alveograph and bread-making tests were calculated based on solvent retention capacity[103].

We also measured GPC (on a 12.5% moisture basis), grain hardness (particle size index) and moisture content using near-infrared spectroscopy (NIR system 6500, Foss) in accordance to the methods AACC 39–10, 39–70A and 39–00, respectively[101]. The grain samples were then milled using the Brabender Quadrumat Jr. (C. W. Brabender OHG), after being optimally tempered (13 to 16.5%), according to the hardness. Both flour protein and moisture content were then determined with the Antaris II FT-NIR analyzer (Thermo). Calibration for particle size index (AACC method 55–30), moisture (AACC method 44–15A) and protein (AACC Method 46–11A) were performed in both near-infrared spectroscopy instruments. Test weight ($kg\,hl^{-1}$) was obtained by weighing a 37.81-ml sample, and TKW (g) was obtained by weighing the kernels that were counted

using the digital image system SeedCount SC5000 (Next Instruments). Grain color was scored as a binary trait, where 0 represented white and 1 represented red and flour yield was obtained as the percentage recovered from milling.

Statistical analysis of the phenotypic data. *Removal of outliers.* The Huber's robust fit outliers method[104] was used to remove outliers in the phenotypic data using the JMP statistical software (https://www.jmp.com).

Best linear unbiased estimates. The best linear unbiased estimates (BLUEs) for GY within each environment were calculated using the ASREML statistical package[105] with the following mixed linear model:

$$y_{ijkl} = \mu + g_i + t_j + r_{k(j)} + b_{l(jk)} + \varepsilon_{ijkl} \quad (1)$$

where y_{ijkl} is the GY of the ith genotype in the jth trial, kth replicate and lth block, μ is the mean, g_i is the fixed effect of the genotype i, t_j is the random effect of the trial $t_j \sim \text{NIID}(0, \sigma_t^2)$ where NIID stands for normal, independent and identically distributed, $r_{k(j)}$ is the random effect of the replicate within the trial, $r_{k(j)} \sim \text{NIID}(0, \sigma_r^2)$, $b_{l(jk)}$ is the random effect of the incomplete block within the trial and the replicate $b_{l(jk)} \sim \text{NIID}(0, \sigma_b^2)$ and ε_{ijkl} is the residual $\varepsilon_{ijkl} \sim \text{NIID}(0, \sigma_\varepsilon^2)$. The random effect of the year was included in equation (1) for GY BLUEs across years in the EYT and ESWYT panels and the BLUEs for all other traits.

Phenotypic correlations. The phenotypic correlations among all the trait–environment combinations evaluated in the 3,485 EYT lines were obtained using the Pearson's product–moment estimator.

Genotyping. The GBS approach[22,106] was used to obtain genome-wide markers for all of the lines. The Illumina platform was used for genotyping at Kansas State University, and marker polymorphisms were called using TASSEL v.5 (trait analysis by association evolution and linkage) GBS v.2 pipeline[106]. The minor allele frequency for single-nucleotide polymorphism (SNP) discovery was set to 0.01 and about 6,075,743 unique tags were anchored to the International Wheat Genome Sequencing Consortium's first version of the reference sequence (RefSeq v.1.0) assembly of the bread wheat variety Chinese Spring[19], with Bowtie2[107]. This resulted in an overall alignment rate of 63.98%, with 28.92% unique alignments and 35.06% multiple alignments. The SNPs were initially filtered for an inbred coefficient of >80%, $P < 0.001$ using Fisher's exact test and a χ^2 value less than the critical value of 9.2 with 2 d.f. and $\alpha = 0.01$. The 78,606 SNPs that passed at least one of these filters were further filtered for a minor allele frequency greater than 5% and heterozygosity less than 5%. For the four evaluation nurseries, marker subsets with less than 70% missing data (12,798–14,260 markers), 50% missing data (7,737–8,586 markers), 10% missing data (1,290–1,889 markers), 10% missing data and pairwise correlations less than 0.8 (781–958 markers), 10% missing data and pairwise correlations less than 0.5 (374–447 markers) and 10% missing data and pairwise correlations less than 0.3 (77–97 markers) were created and used for within-panel genomic predictions. For across-panel predictions, the different marker subsets used include those with less than 70% missing data (16,072 markers), 50% missing data (9,285 markers), 10% missing data (2,253 markers), 10% missing data and pairwise correlations less than 0.8 (1,091 markers), 10% missing data and pairwise correlations less than 0.5 (504 markers) and 10% missing data and pairwise correlations less than 0.3 (160 markers). For GWAS, a subset of markers with missing data less than 40% was used, resulting in 6,355 markers for the EYT panel and 9,171–9,704 markers for the seven ESWYT panels. Similarly, for the biparental population, markers with missing data greater than 30%, minor allele frequency lesser than 5% and heterozygosity greater than 10% were removed resulting in 1,501 markers. Missing data were imputed with the k-nearest neighbor genotype imputation method based on LD using LinkImpute[108] in TASSEL[109] v.5.

Genomic prediction. We used the GBLUP, which is a robust and widely used model for genomic predictions[96,110,111]. The GBLUP model can be represented as:

$$y_i = \mu + g_i + \varepsilon_i \quad (2)$$

where y_i is the response phenotype or the adjusted best linear unbiased predictions for individual i, μ is the general mean, the vector $\mathbf{g} = (g_1, \ldots, g_i)'$ contains the genomic values of the lines that follows a multivariate normal density such that $\mathbf{g} = \{g_i\} \sim N\left(\mathbf{0}, \mathbf{G}\sigma_g^2\right)$ where $\mathbf{G} = \mathbf{XX}'/p$ is the genomic relationship matrix[112], $\mathbf{X}$ is the centered and standardized genomic relationship matrix, σ_g^2 is the genomic variance and p the number of markers and ε_i is the error term (assuming that the joint distribution of $\boldsymbol{\varepsilon}$ is $N(0, \mathbf{I}\sigma_e^2)$ where σ_e^2 is the residual variance). We also used the Bayes B approach[113] for comparison with the GBLUP prediction accuracies and estimation of marker effects. The BGLR package in R[114] was used to fit the GBLUP and Bayes B models with 100,000 iterations and 10,000 burn-ins. BGLR treats the parameter π that is the proportion of non-null effects as unknown and assigns a beta prior parameterized such that the expected value is $E(\pi) = \pi_0$, where π_0 is the number of prior counts. Prediction accuracies were calculated as the Pearson's correlation between the phenotypic values or BLUEs and the predicted genomic-estimated breeding values.

The impact of missing marker data and genomic coverage on predictions was evaluated with marker subsets of 2,253, 9,285 and 16,072 markers. Genomic predictions within the panel were done using fivefold cross-validations, for which folds comprising 153–196 lines were predicted from four other folds comprising 613–784 lines, and the mean of 100 iterations was taken. We also performed across-panel genomic predictions, for which panels comprising 766–980 lines were predicted from three other panels of 2,505–2,719 lines, except for seedling resistance traits.

Genome-wide association mapping. Genome-wide association mapping was implemented in TASSEL v.5, using a mixed linear model[115] that accounts for both population structure and kinship. Population structure was accounted for with the first two principal components[116], whereas kinship was accounted for using the pedigree–relationship matrix. We also used the R package LEA (Landscape and Ecological Association studies) to estimate the individual ancestry coefficients and the number of ancestral sub-populations[117] The optimal level of compression and the 'population parameters previously determined' method[118] was used to run the mixed linear model and a Bonferroni threshold level of 0.20 was used to correct for multiple testing and identify the significant markers in the panel of EYTs. The significant markers were then delineated into QTLs based on the LD between markers, where markers with $P < 0.001$ for the existence of LD were included in the same QTL. The genetic positions of the markers were obtained through publicly available mapped markers in the Triticeae Toolbox database (https://triticeaetoolbox.org). The trait-associated markers and previously reported genes or QTLs near the significant markers were anchored onto a genotype–phenotype map aligned to the RefSeq v.1.0 and visualized using Phenogram (http://visualization.ritchielab.org/phenograms/plot).

Biparental mapping. In the biparental mapping population, QTL mapping and genetic map construction were done using the Minimum Spanning Tree algorithm[119] in the ASMap R package[120]. In addition, the recombinants were profiled for the proportion of resistant and susceptible parental genomes.

Genomic fingerprinting and allele frequency dynamics. The estimated allelic effects of 195 trait-associated markers (the most significant marker for each trait in each QTL) from GWAS were used to generate the genomic fingerprints of 44,624 wheat lines. The favorable alleles were defined as those that had increasing effects on GY, agronomic traits and end-use quality related traits and decreasing effects on diseases. The progressive trend of the favorable alleles and the alleles with increasing effects were analyzed in the five yield trial panels from 2014 to 2018. In addition, the frequency dynamics of the favorable alleles for GY in the ESWYTs due to selection for 15 years (2013–2017) was assessed using 47 markers associated with GY in the EYTs and ESWYTs. Furthermore, to determine whether the change in allele frequencies is solely due to genetic drift, we calculated the expected variance due to random genetic drift using the equation:

$$V_t \sim p(1-p)\left(1 - \exp\left(-\frac{t}{2N_e}\right)\right) \quad (3)$$

where p is the initial allele frequency, t is the number of generations and N_e is the effective population size[121] and compared it with the observed variance of the favorable allele frequencies for GY.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The phenotyping data for the lines used in this study are available in Supplementary Data 1. The marker P values, additive effects and percentage variation explained by each marker are available in Supplementary Table 2. The genomic fingerprints of 44,624 wheat lines for 195 trait-associated markers are available in Supplementary Table 4a–d. The raw genotyping data for the lines are available in FigShare (https://doi.org/10.6084/m9.figshare.8940257.v1).

References

86. Herrera-Foessel, S. A. et al. Lr68: A new gene conferring slow rusting resistance to leaf rust in wheat. *Theor. Appl. Genet.* **124**, 1475–1486 (2012).
87. Juliana, P. et al. Genomic and pedigree-based prediction for leaf, stem, and stripe rust resistance in wheat. *Theor. Appl. Genet.* **130**, 1415–1430 (2017).
88. Roelfs, A. P., Singh, R. P. & Saari, E. E. *Rust Diseases of Wheat: Concepts and Methods of Disease Management* (CIMMYT, 1992).
89. Chen, S. et al. Fine mapping and characterization of *Sr21*, a temperature-sensitive diploid wheat resistance gene effective against the *Puccinia graminis* f. sp. *tritici* Ug99 race group. *Theor. Appl. Genet.* **128**, 645–656 (2015).
90. Jin, Y. et al. Characterization of seedling infection types and adult plant infection responses of monogenic *Sr* gene lines to race TTKS of *Puccinia graminis* f. sp. *tritici*. *Plant Dis.* **91**, 1096–1099 (2007).

91. Rouse, M. N. & Jin, Y. Stem rust resistance in A-genome diploid relatives of wheat. *Plant Dis*. **95**, 941–944 (2011).
92. Stakman, E. C., Stewart, D. M. & Loegering, W. Q. *Identification of Physiologic Races of* Puccinia graminis *var.* tritici USDA-ARS E-617 (USDA, 1962).
93. Juliana, P. et al. Genome-wide association mapping for resistance to leaf rust, stripe rust and tan spot in wheat reveals potential candidate genes. *Theor. Appl. Genet*. **131**, 1405–1422 (2018).
94. Randhawa, M. S. et al. Identification and validation of a common stem rust resistance locus in two bi-parental populations. *Front. Plant Sci.* **9**, 1788 (2018).
95. Peterson, R. F., Campbell, A. B. & Hannah, A. E. A diagrammatic scale for estimating rust intensity on leaves and stems of cereals. *Can. J. Res.* **26c**, 496–500 (1948).
96. Juliana, P. et al. Comparison of models and whole-genome profiling approaches for genomic-enabled prediction of Septoria tritici blotch, Stagonospora nodorum blotch, and tan spot resistance in wheat. *Plant Genome* https://doi.org/10.3835/plantgenome2016.08.0082 (2017).
97. Saari, E. E. & Prescott, J. A scale for appraising the foliar intensity of wheat diseases. *Plant Dis. Rep.* **59**, 376–381 (1975).
98. Eyal, Z., Scharen, A. L., Prescott, J. M. & van Ginkel, M. *The Septoria Diseases of Wheat: Concepts and Methods of Disease Management* (CIMMYT, 1987).
99. Simko, I. & Piepho, H.-P. The Area under the disease progress stairs: calculation, advantage, and application. *Phytopathology* **102**, 381–389 (2012).
100. Singh, P. et al. Resistance to spot blotch in two mapping populations of common wheat is controlled by multiple QTL of minor effects. *Int. J. Mol. Sci.* **19**, 4054 (2018).
101. AACC. *Approved Methods of Analysis* 11th edn (American Association of Cereal Chemists, 2000); https://doi.org/10.1094/AACCIntMethod-10-05.01
102. Pena, R. J., Amaya, A., Rajaram, S. & Mujeeb-Kazi, A. Variation in quality characteristics associated with some spring 1B/1R translocation wheats. *J. Cereal Sci*. **12**, 105–112 (1990).
103. Guzmán, C., Posadas-Romano, G., Hernández-Espinosa, N., Morales-Dorantes, A. & Peña, R. J. A new standard water absorption criteria based on solvent retention capacity (SRC) to determine dough mixing properties, viscoelasticity, and bread-making quality. *J. Cereal Sci*. **66**, 59–65 (2015).
104. Huber P. J. & Ronchetti, E. M. *Robust Statistics* 2nd edn (John Wiley & Sons, 2009).
105. Gilmour, A. R. ASREML for testing fixed effects and estimating multiple trait variance components. *Proc. Assoc. Adv. Anim. Breed. Genet.* **12**, 386–390 (1997).
106. Glaubitz, J. C. et al. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS ONE* **9**, e90346 (2014).
107. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
108. Money, D. et al. LinkImpute: fast and accurate genotype imputation for nonmodel organisms. *G3 (Bethesda)* **5**, 2383–2390 (2015).
109. Bradbury, P. J. et al. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).
110. Heslot, N., Yang, H., Sorrells, M. E. & Jannink, J.-L. Genomic selection in plant breeding: a comparison of models. *Crop Sci.* **52**, 146–160 (2012).
111. Rutkoski, J. et al. Evaluation of genomic prediction methods for fusarium head blight resistance in wheat. *Plant Genome* **5**, 51–61 (2012).
112. VanRaden, P. M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **91**, 4414–4423 (2008).
113. Habier, D., Fernando, R. L., Kizilkaya, K. & Garrick, D. J. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics* **12**, 186 (2011).
114. Pérez, P. & de los Campos, G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* **198**, 483–495 (2014).
115. Yu, J. et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **38**, 203–208 (2006).
116. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
117. Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G. & François, O. Fast and efficient estimation of individual ancestry coefficients. *Genetics* **196**, 973–983 (2014).
118. Zhang, Z. et al. Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **42**, 355–360 (2010).
119. Wu, Y., Bhat, P. R., Close, T. J. & Lonardi, S. Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLoS Genet*. **4**, e1000212 (2008).
120. Taylor, J. & Butler, D. ASMap: Linkage Map Construction using the MSTmap Algorithm. R version 0.4-4 (2015).
121. Barton, N. H., Briggs, D. E. G., Eisen, J. A., Goldstein, D. B. & Patel, N. H.*Evolution* (Cold Spring Harbor Laboratory Press, 2007).

nature research

Corresponding author(s): Dr. Ravi Prakash Singh

Last updated by author(s): Jul 16, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection
Data analysis	The Huber's robust fit outliers method was used for removing outliers in the phenotypic data using the 'JMP' statistical software (www.jmp.com). The BGLR package in 'R' was used to fit the Genomic-Best linear unbiased prediction model for genomic prediction. TASSEL version 5 was used for genome-wide asscoiation mapping. QTL mapping and genetic map construction were done using the Minimum Spanning Tree algorithm in the ASMap 'R' package.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The phenotyping data for the lines used in this study is available in Supplemental File S1. The marker p-values, additive effects and percent variation explained by each marker are available in Supplementary Table 2. The genomic-fingerprints of 44,624 wheat lines for 195 traits-associated markers is available in Supplementary Tables 4a-d. The raw genotyping data for all the 44,624 lines is available at https://doi.org/10.6084/m9.figshare.8940257.v1

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The sample size of 44,624 lines represents CIMMYT's wheat breeding lines genotyped for the past 5 years and a subset of 3,485 lines that had extensive phenotyping information was used for genome-wide association mapping and genomic prediction.
Data exclusions	No data was excluded from the analysis.
Replication	Most of the phenotyping data used in this study was collected on a set of replicates and for disease evaluations, atleast 3-4 evaluations across time during the progression of the disease were performed. For grain yield measurements, all the entries were replicated thrice and evaluated in trials with two checks per trial.
Randomization	Randomization is applicable only to the assignment of individual lines to the different folds for cross-validations in genomic predictions and we used 100 iterations to re-allocate the lines to different folds every time and only the mean of 100 iterations is reported. The only covariate to be considered was the days to heading that was moderately associated with grain yield and it was controlled by removing the lines in the tails of the days to heading distribution and only including lines that headed in a week's interval.
Blinding	Blinding was not relevant to this study and we have used all the available phenotypic and genotypic data.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging