AGRICULTURE

Unlocking big data doubled the accuracy in predicting the grain yield in hybrid wheat

Yusheng Zhao[1], Patrick Thorwarth[2], Yong Jiang[1], Norman Philipp[3], Albert W. Schulthess[1], Mario Gils[4], Philipp H. G. Boeven[5], C. Friedrich H. Longin[2], Johannes Schacht[5], Erhard Ebmeyer[6], Viktor Korzun[7,8], Vilson Mirdita[9], Jost Dörnte[10], Ulrike Avenhaus[11], Ralf Horbach[12], Hilmar Cöster[13], Josef Holzapfel[14], Ludwig Ramgraber[15], Simon Kühnle[16], Pierrick Varenne[17], Anne Starke[5], Friederike Schürmann[14], Sebastian Beier[1], Uwe Scholz[1], Fang Liu[1], Renate H. Schmidt[1], Jochen C. Reif[1]*

The potential of big data to support businesses has been demonstrated in financial services, manufacturing, and telecommunications. Here, we report on efforts to enter a new data era in plant breeding by collecting genomic and phenotypic information from 12,858 wheat genotypes representing 6575 single-cross hybrids and 6283 inbred lines that were evaluated in six experimental series for yield in field trials encompassing ~125,000 plots. Integrating data resulted in twofold higher prediction ability compared with cases in which hybrid performance was predicted across individual experimental series. Our results suggest that combining data across breeding programs is a particularly appropriate strategy to exploit the potential of big data for predictive plant breeding. This paradigm shift can contribute to increasing yield and resilience, which is needed to feed the growing world population.

INTRODUCTION

Wheat is one of the most important crops providing one-fifth of the world's food calories and proteins (*1*). To supply an estimated world population of 9 billion in 2050, global wheat production must be increased by 60% compared with 2005–2007 levels (*2*). Hybrid breeding has the potential to become a critical factor in increasing grain yield and resilience of wheat (*3*). A key challenge in hybrid breeding is to predict the most promising combination of parents leading to a high-yielding hybrid (*4*, *5*).

The genetic basis of hybrid grain yield in wheat is complex and determined by many quantitative trait loci and their interactions, each of which has little impact on the phenotypic variation (*3*, *6*). Genome-wide prediction, which was initially implemented in animal breeding (*7*) but is now also routinely used in plant breeding (*1*) and human genetics (*8*), is the method of choice for hybrid prediction of complex traits such as grain yield (*9*). The success of genome-wide prediction depends on accurate prediction equations, especially for new lines that have not yet been evaluated in the field, i.e., out-of-sample scenarios. For example, threefold higher prediction accuracies were observed for wheat hybrids, which had both parents in common with the training set that had been assessed regarding phenotype and genotype than for hybrids that that did not share a single parent with the lines that were evaluated in the training set (*6*). The low prediction accuracies that were observed in previous studies for out-of-sample scenarios are of concern as these impede due to decreasing relatedness with ongoing selection the long-term selection gains in recurrent genomic selection programs (*10*), which are crucial for the success of hybrid wheat breeding (*11*). Consequently, there is an urgent need to develop reliable prediction models for out-of-sample scenarios that are commonplace in continuously operating breeding programs. Theoretical and empirical results have shown that prediction abilities rise with a combination of increasing heritability of traits (*12*, *13*), increasing ratio of sample size to effective population size (*14*, *15*), marker density (*16*), and the genetic and also environmental similarity between training and test datasets (*6*). The development of accurate predictive models using comprehensive data was convincingly demonstrated in the UK Biobank cohort study consisting of half a million participants (*13*). It appears, therefore, meaningful to assemble also for crop plants large and diverse training populations for which high-quality phenotypic and genomic data are available, but this is hardly feasible or cost-effective for small- to medium-scale hybrid breeding programs. So far, the grain yield performance of hybrids has been predicted in populations based on a maximum of 2000 single crosses and up to 700 parents (*6*, *17*–*19*). Wheat breeding organizations, often in joint efforts in public-private partnerships, have generated extensive phenotypic and genomic data in individual studies with many of these studies (*6*, *18*–*20*) focusing on hybrid wheat breeding in Europe. It remains to be explored whether the integration of these different datasets may permit the development of reliable prediction models as the aggregation of several medium-sized datasets for different populations of wheat lines and/or hybrids into a large one is challenging for several reasons. First, different types of genomic data need to be integrated. Owing to the availability of a high-quality reference genome for wheat, this can be efficiently realized (*21*). Second, marker density needs to be adjusted to the population size (*16*). Considering the genetic relatedness between lines, the genome size of wheat, and

[1]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), 06466 Stadt Seeland, Germany. [2]State Plant Breeding Institute, University of Hohenheim, Fruwirthstr. 21, 70593 Stuttgart, Germany. [3]Syngenta Seeds GmbH, Kroppenstedterstr. 4, 39398 Hadmersleben, Germany. [4]Nordsaat Saatzucht GmbH, , Böhnshauserstr. 1, 38895 Langenstein, Germany. [5]Limagrain GmbH, Salderstr. 4, 31226 Peine-Rosenthal, Germany. [6]KWS LOCHOW GmbH, Ferdinand-von-Lochow-Str. 5, 29303 Bergen, Germany. [7]KWS SAAT SE & Co. KGaA, Grimsehlstr. 31, 37574 Einbeck, Germany. [8]Federal State Budgetary Institution of Science Federal Research Center, "Kazan Scientific Center of Russian Academy of Sciences," ul. Lobachevskogo, 2/31, Kazan, 420111 Tatarstan, Russian Federation. [9]BASF Agricultural Solutions Seed GmbH, OT Gatersleben, Am Schwabeplan 8, 06466 Seeland, Germany. [10]Deutsche Saatveredelung AG, Leutewitz 26, 01665 Käbschütztal, Germany. [11]W. von Borries-Eckendorf GmbH & Co. KG, Hovedisserstr. 92, 33818 Leopoldshöhe, Germany. [12]Saatzucht Bauer GmbH & Co. KG, Hofmarkstr.1, 93083 Niederträubling, Germany. [13]RAGT2n, Steinesche 5a, 38855 Silstedt, Germany. [14]Secobra Saatzucht GmbH, Feldkirchen 3, 85368 Moosburg, Germany. [15]Saatzucht Josef Breun GmbH & Co. KG, Amselweg 1, 91074 Herzogenaurach, Germany. [16]Pflanzenzucht Oberlimpurg, Oberlimpurg 2, 74523 Schwäbisch Hall, Germany. [17]Limagrain Europe, Ferme de l'Etang BP3, 77390 Verneuil l'Etang, France.
*Corresponding author. Email: reif@ipk-gatersleben.de

the sizes of the individual populations, marker density does not represent a major limiting factor for genome-wide prediction studies of individual datasets (*22*), as sufficient marker coverage of more than 10,000 single-nucleotide polymorphisms (SNPs) can be achieved by using high-density SNP chips, genotyping-by-sequencing, or exome capture sequencing (*23*, *24*). Third, the integrated analyses of phenotypic data pose problems because of their unbalanced and often noisy structure: The individual experimental series usually lack proper links across experimental series and/or environments.

The aggregation of several medium-sized experimental series into a large dataset has features in common with the dimensions volume, velocity, variety, and veracity that have been used to characterize "big data" and which depend on advanced analysis approaches for the generation of new insights or the optimization of processes (*25*). Here, we report on the collection and integration of large-scale genomic and/or phenotypic data of 12,858 wheat entries, consisting of 6575 single-cross hybrids and 6283 inbred lines, which were investigated in six experimental series in multienvironmental yield trials. The six experimental series were based on different crossing designs comprising factorial mating and topcross designs and include not only a very broad genetic diversity of the European wheat breeding pool but also plant genetic resources. We have used this comprehensive dataset to explore the challenges and potentials of entering the era of big data in crop breeding. The low overlap between common genotypes in the six experimental series resulting from the intrinsic structure when aggregating several medium-sized datasets clearly caused challenges because of interaction effects between genotypes and experimental series. Accordingly, hybrid prediction across two different experimental series resulted in low prediction abilities, but twofold higher values were achieved by integrating data from several experimental series into the training populations. This finding calls for the compilation of comprehensive datasets to train models for hybrid prediction.

RESULTS

Genetic structure of the meta-population

We fingerprinted a subset of 5042 unique inbreds (Fig. 1A and figs. S1 and S2) out of 6283 winter wheat lines and parents of single-cross hybrids that were used for six large-scale field experimental series (table S1) with marker arrays derived from a public 90,000 SNP chip (*26*), resulting in 10,522 high-quality markers. The lines of experimental series I, II, III, IV, and VI represented a comprehensive selection of the current elite bread wheat breeding pool developed for Central Europe, i.e., Germany, Poland, Denmark, Switzerland, Austria, Northern France, Netherlands, Czech Republic, and Slovakia, and were provided by 14 wheat-breeding companies (*18*). Experimental series VI encompassed lines only (table S1), whereas experimental series I to V included lines and hybrids. Three different crossing schemes shown in fig. S3 were used to generate in total 5643 single-cross hybrids from current elite lines (experimental series I to IV in table S1). This panel was supplemented by 267 former elite varieties from the last five decades and 357 genetic resources preserved at the IPK gene bank in Gatersleben (experimental series V). The former elite cultivars and genetic resources were crossed with elite lines and produced 932 hybrids (experimental series V in table S1 and fig. S4).

The effective population size was estimated using the marker data and amounted to 95 for the analysis across the six experimental series with a range within experimental series from 23 (experimental series IV) to 70 (experimental series VI; table S2). The larger effective population size of experimental VI in comparison to the other series comprising elite lines and derived hybrids only is also reflected in the results of the principal coordinate analysis (Fig. 1A and fig. S2). Lines of experimental series V were separated from current elite lines as emphasized by an average fixation index of F_{ST} of 0.06 (Fig. 1B), which was further supported by the results of the principal coordinate analysis (Fig. 1A), distribution of genetic distances within experimental series V and across the different experimental series (Fig. 1C), and overall lower linkage disequilibrium (LD) phases observed for experimental series V as compared with the other five experimental series (Fig. 1D). In summary, we assembled a diverse sample of plant material for this study.

Integrated analysis highlights the quality of the phenotypic data

Extensive grain yield data were compiled by evaluating not only hybrids but also inbred lines in field trials in 125,422 plots in Central Europe. The phenotypic data were collected in six experimental series. We evaluated the quality of the raw data of the individual experimental series (fig. S1) and removed 519 plots as outliers (*27*), resulting in estimates of broad-sense heritability for grain yield in the range 0.64 to 0.92 for the individual experimental series (table S1 and Fig. 2A). The different experimental series were linked by up to 37 overlapping genotypes detected by genomic data (table S3). Pairwise Rogers' distances between all genotypes were calculated, and genotypes with Rogers' distances less than 0.03 were considered to be overlapping genotypes. The set of overlapping genotypes allowed an integrated analysis.

We used the grain yield data from overlapping genotypes to assess the presence of potential biases and reduced correlations due to interaction effects between genotypes and environments in estimating the performance across the six experimental series. To detect a potential bias between experimental series, a sufficient number of common genotypes need to be present in the different pairs of environments to obtain reliable data, but it is equally important that experimental series do not have too many environments in common because this may lead to a systematic underestimation of a potential bias. Contrasting experimental series I and VI with the combined set of experimental series II, III, IV, and V fulfilled these requirements. Using the combined phenotypic data, we selected a genotype that was common in at least one of the pairs of experimental sets described above (I versus II–V and VI versus II–V) and coded it differently in the two sets. Grain yield was then estimated and recorded for the selected common genotypes for experimental series I and VI and the combined set of experimental series II, III, IV, and V. Repeating this procedure for all overlapping genotypes resulted in two sets of yield estimates. The correlation between the two estimates for all overlapping genotypes was high and amounted to 0.68 ($P < 0.001$) with a regression coefficient of 1.001, which was not significantly ($P = 0.81$) different from 1 (Fig. 2B). This clearly suggests absence of a systematic bias; nonetheless, the reduced correlation (Fig. 2B) provides evidence for interaction effects between genotypes and experimental series.

The integrated phenotypic analysis resulted in 4491 lines and 6246 high-quality hybrids (figs. S1, S3, and S4) for which marker profiles were available (table S2) or, in the case of hybrids, could be derived from their parents' information. The hybrids and lines were

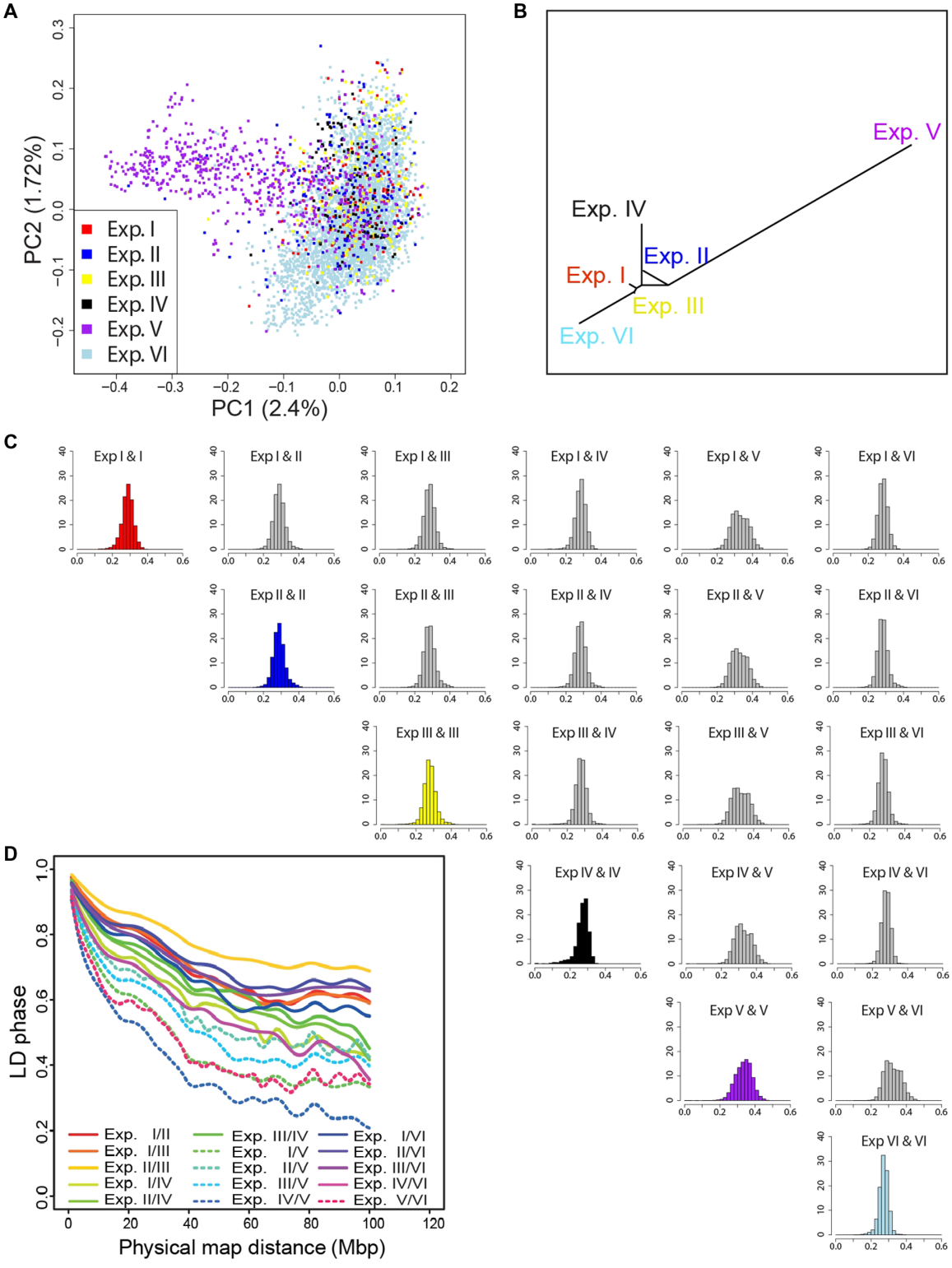

Fig. 1. Population genomic analyses of parental lines grouped into six experimental series. (**A**) Principal coordinate analysis of the inbred lines based on Rogers' distances matrix. Percentages in parentheses refer to the proportion of genotypic variance explained by the first and second principal coordinates (PCs). (**B**) Neighbor-joining tree based on the results of F_{ST} statistics for the six experimental series (Exp.). (**C**) Distribution of Rogers' distances for inbred lines within and across experimental series. In each histogram plot, the range of Rogers' distances is displayed on the *x* axis; on the *y* axis the percentage of line pairs is provided. (**D**) Persistence of the LD phase between the six experimental series.

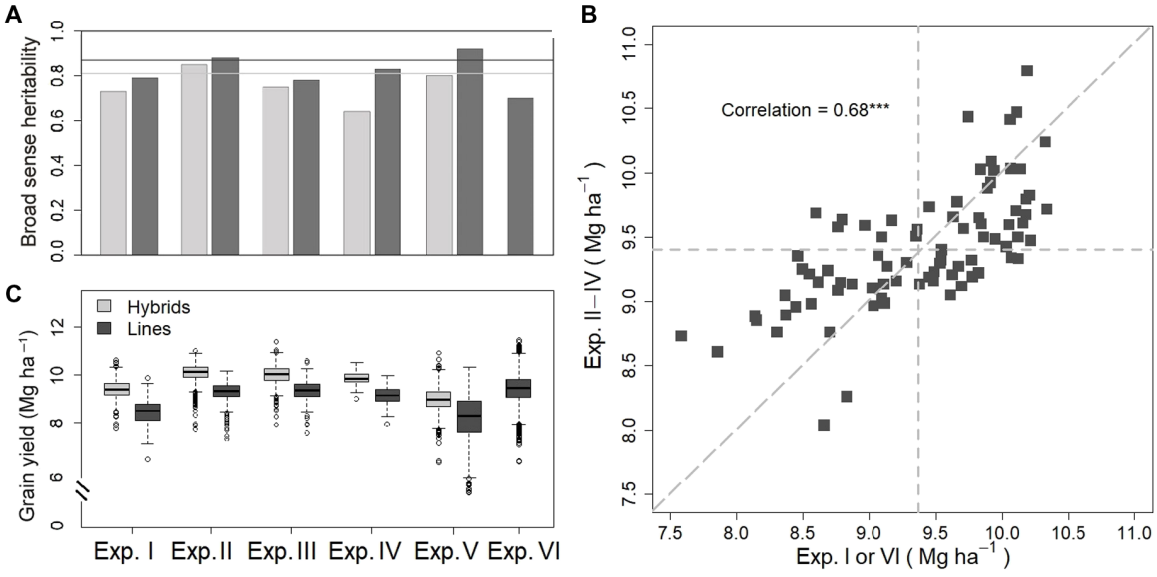

Fig. 2. Grain yield performance assessed in multienvironmental field trials. (**A**) Broad-sense heritability values for hybrids and lines within experimental series are shown as bars and across experimental series as vertical lines. Light and dark gray refer to hybrids and lines, respectively. (**B**) Assessing a potential bias in grain yield estimates triggered by merging nonorthogonal phenotypic data across experimental series. Grain yield was estimated on the basis of the combined phenotypic data of all but one overlapping genotypes. For this genotype, grain yield was then estimated separately for experimental series (Exp.) I or VI and a combined set of experimental series II, III, IV, and V. Repeating this procedure for all overlapping genotypes resulted in two sets of estimates. The correlations between these estimates are plotted. ***$P < 0.001$. (**C**) Distribution of best linear unbiased estimations for grain yield (Mg ha^{-1}) of the genotypes included in the six experimental series.

evaluated on average in 9.3 and 5.4 environments resulting in broad-sense heritability estimates of 0.81 and 0.87 (Fig. 2A and table S4), respectively. Within experimental series, hybrids outperformed their parents and checks by on average 9.4% (Fig. 2C). This was also observed when inspecting the results across experimental series: hybrids (9.7 Mg ha^{-1}) outyielded on average all lines and checks by 5.5% (9.2 Mg ha^{-1}), indicating the potential to increase wheat yield by implementing hybrid breeding programs.

Prediction ability of hybrid grain yield is determined mainly by relatedness

Because each of the parents in experimental series I, II, and III was tested in several hybrid combinations, we investigated the ability to predict hybrid grain yield performance using a genomic-based unbiased prediction model incorporating both additive and dominance genomic relationships and a chessboard-like cross-validation with three different level of relatedness: T_2, T_1, and T_0 (fig. S5). The validation in test set T_2, which included only hybrids originating from the same group of parents as the hybrids in the training set, showed the highest prediction ability of 0.73 averaged over experimental series I, II, and III (Fig. 3A). This value decreased to 0.25 for the test set T_0, which contained only hybrids that did not share parents with the training set. Thus, the decreasing trend in prediction ability reflected the diminishing relatedness from the T_2 to the T_0 scenario. The declines in prediction ability observed in our wheat experiments were more pronounced than in maize (*28*). This can be explained by a lower effective population size for a single maize breeding program compared with the diversity panels sampled across multiple wheat breeding programs in our study.

The topcross mating design of experimental series IV with only four tester lines (fig. S3) and of experimental series V with an average of two tester lines (fig. S4) prevented chessboard-like cross-validation, and we, therefore, applied random fivefold cross-validation (fig. S5). The prediction ability was 0.61 for experimental series IV and 0.71 for experimental series V (Fig. 3A). This, as expected, corresponds to the values observed for the T_2 scenario in experimental series I, II, and III and highlights again that within–experimental series relatedness is the driving force of the prediction ability for hybrid grain yield in our study.

Analysis of a comprehensive inbred line population

Experimental series VI was the most comprehensive series in our study, with 3448 lines being genotyped and phenotyped (table S2). It represented grain yield data from a commercial line breeding program. The fivefold cross-validation showed a high prediction ability of 0.69 and thus approached the mean value of 0.73, which had been established for the T_2 scenario of experimental series I, II, and III (Fig. 3A). At first glance, this may appear unexpected because the genetic distance between training and test populations was higher in experimental series VI (average 5% quantile of genetic distances equaled 0.22) than the average value observed for the other five experimental series (average 5% quantile of genetic distances equaled 0.16; table S5 and fig. S6). Nevertheless, the observed range of phenotypic values in experimental series VI was much larger than that in experimental series I, II, and III (Fig. 2C). Thus, the prediction ability in experimental series I, II, and III might have been suppressed due to range restriction, a phenomenon that the correlation is reduced when the sample has a restricted range of scores (*29*). In addition, the degrees of freedom to estimate the additive effects of SNPs in the inbred population depend on the number of lines, but in the case of hybrids, on the number of parents. Thus, the increased degrees of freedom allowed a more precise estimation of the additive effects to predict grain yields in experimental series VI compared with the other series. This, together, explains why the prediction

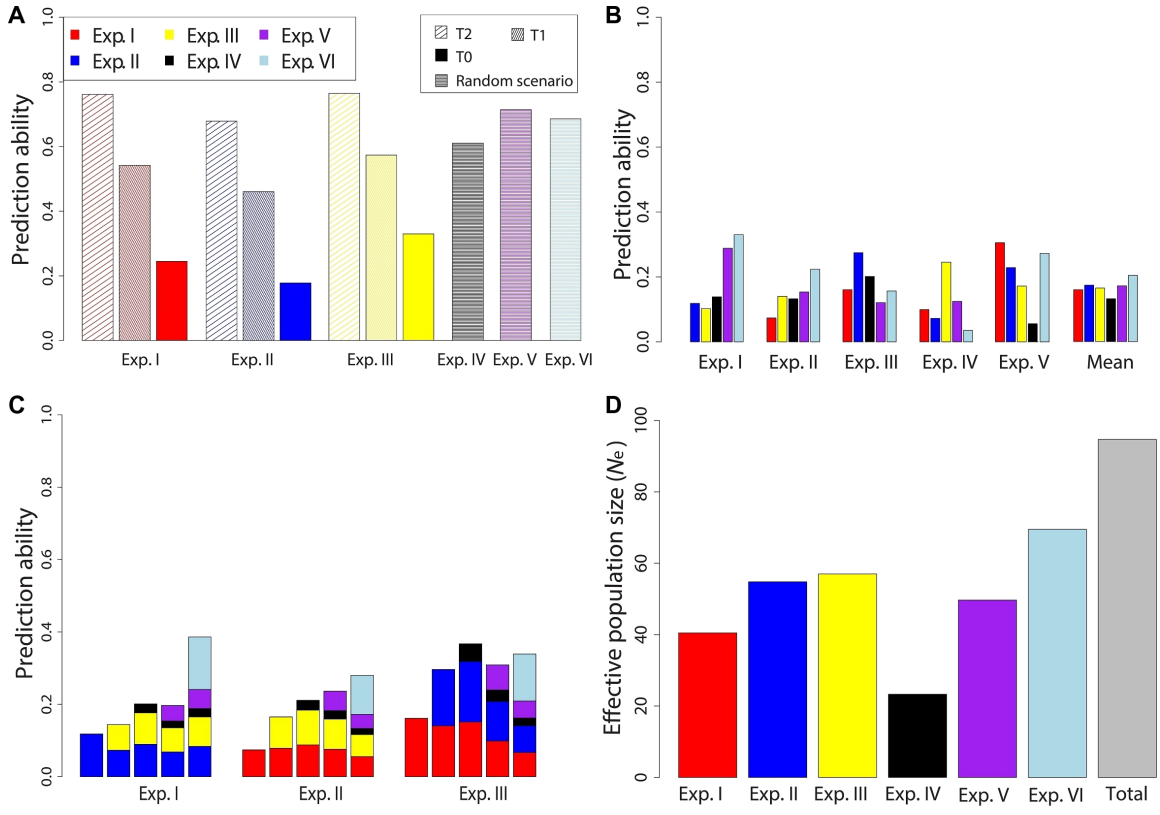

Fig. 3. Prediction abilities and the effective population size within and across the six experimental series. (**A**) Prediction ability within experimental series (Exp.) was estimated for related or unrelated training populations by using a chessboard-like cross-validation in experimental series I, II, and III or by fivefold cross-validation based on random sampling of genotypes (random scenario) in experimental series IV, V, and VI. (**B**) Prediction abilities across different experimental series. For each of the training populations shown on the *x* axis, the prediction abilities for the different test populations are displayed as colored bars. (**C**) Increase in prediction ability combining incremental data across experimental series. The lengths of the colored boxes in each bar represent the proportions of the genotypes of the different experimental series used as training sets. (**D**) Effective population size within and across the experimental series. The different experimental series are color coded according to the key in (A).

ability of 0.69 in experimental series VI is almost as high as the average value of 0.73, which was observed for the T_2 scenario in experimental series I, II, and III (Fig. 3A).

Interactions between genotypes and experimental series affect across series prediction ability

The ability to predict the hybrid performance from one experiment to another across experimental series I, II, III, or IV was lower (0.16; table S6 and Fig. 3B) compared with the prediction ability observed for the T_0 scenario within experimental series (Fig. 3A). This decrease cannot be explained by an increased genetic distance between the parental lines of experimental series I, II, III, and IV compared with the distance of the T_0 scenarios within experimental series (table S5 and fig. S6).

The prediction abilities from one experimental series to another varied almost 10-fold (0.035 to 0.330; table S6 and Fig. 3B). In several instances, values for the prediction abilities across experimental series outperformed the mean value of 0.25, which had been observed for the T_0 scenario within experimental series I, II, and III (Fig. 3A). As shown in table S7, the different experimental series shared certain fractions of parental lines leading to T_1 and in rare cases to T_2 hybrids; on average, 16% of the predicted hybrids across experimental series are corresponding to a T_1 scenario. Thus, it is tempting to speculate that the variation in relatedness provides an explanation for the variation in prediction ability across experimental series, but the proportion of T_1 hybrids between pairs of experimental series was not significantly correlated ($r = 0.12$; $P > 0.1$) with the prediction abilities from one experimental series to another. Instead, the lower prediction abilities across two experimental series (table S6 and Fig. 3B) compared with the prediction ability observed for the T_0 scenario within individual experimental series (Fig. 3A) may reflect genotype-by-experiment interaction effects (Fig. 2B). To assess this in more detail, only experimental series II was used as training set to predict the performance of 148 previously untested hybrids that had both parental lines in common with experimental series II (T_2 hybrids). These 148 hybrids were phenotyped in a separate validation experiment for grain yield in eight environments, which had not been used for experimental series II. The prediction ability for these previously untested T_2 hybrids tested in a different set of environments reached only 0.54 (fig. S7), whereas within experimental series II, a prediction ability of 0.68 had been observed for the T_2 scenario (Fig. 3A). This demonstrates the pronounced impact of interaction between genotypes and experimental series on the prediction ability.

By using experimental series VI as the training population and the remaining experimental series as test populations, the prediction ability for the parental lines (0.36; table S8) was on average 77% higher than for the hybrids (0.20; table S6 and Fig. 3B). The use of a comprehensive population of inbred lines as training set, as experimental

series VI, cannot account for the general heterosis effect observed in hybrids (Fig. 2C). Nonetheless, the highest average prediction ability of hybrid performance from one experiment to another was observed using the comprehensive line training data (0.20; table S6), i.e., experimental series VI. This finding illustrates the potential to increase the predictive power for hybrids by exploiting the precision of estimating additive effects in the large population of inbred lines.

The potential of big data for hybrid prediction

One of the important tasks in hybrid wheat breeding is to predict for new environments the single-cross performance of parental lines that have not yet been evaluated in other hybrids. Despite the large number of hybrids evaluated in each experimental series in our study (table S2), the average prediction ability was low (0.17) when predicting the hybrid performance from one experimental series to another (table S6), but the prediction ability was doubled when integrating data across experimental series in the training population in a leave-one-experimental-series-out scenario (table S9 and Fig. 3C). In accordance to quantitative genetic theory (*16*), this increase reflects the ratio of the number of parental components/inbred lines versus the effective population size (N:N_e), which ranged from 3.5 to 9.7 for experimental series I to V but amounted to 47.4 in the total population (table S2 and Fig. 3D). We assessed the relevance of N:N_e in more detail by randomly sampling subpopulations out of experimental series VI representing a range of N:N_e from 2 to 60 (Fig. 4, A and B) and observed a nonlinear increase in the prediction

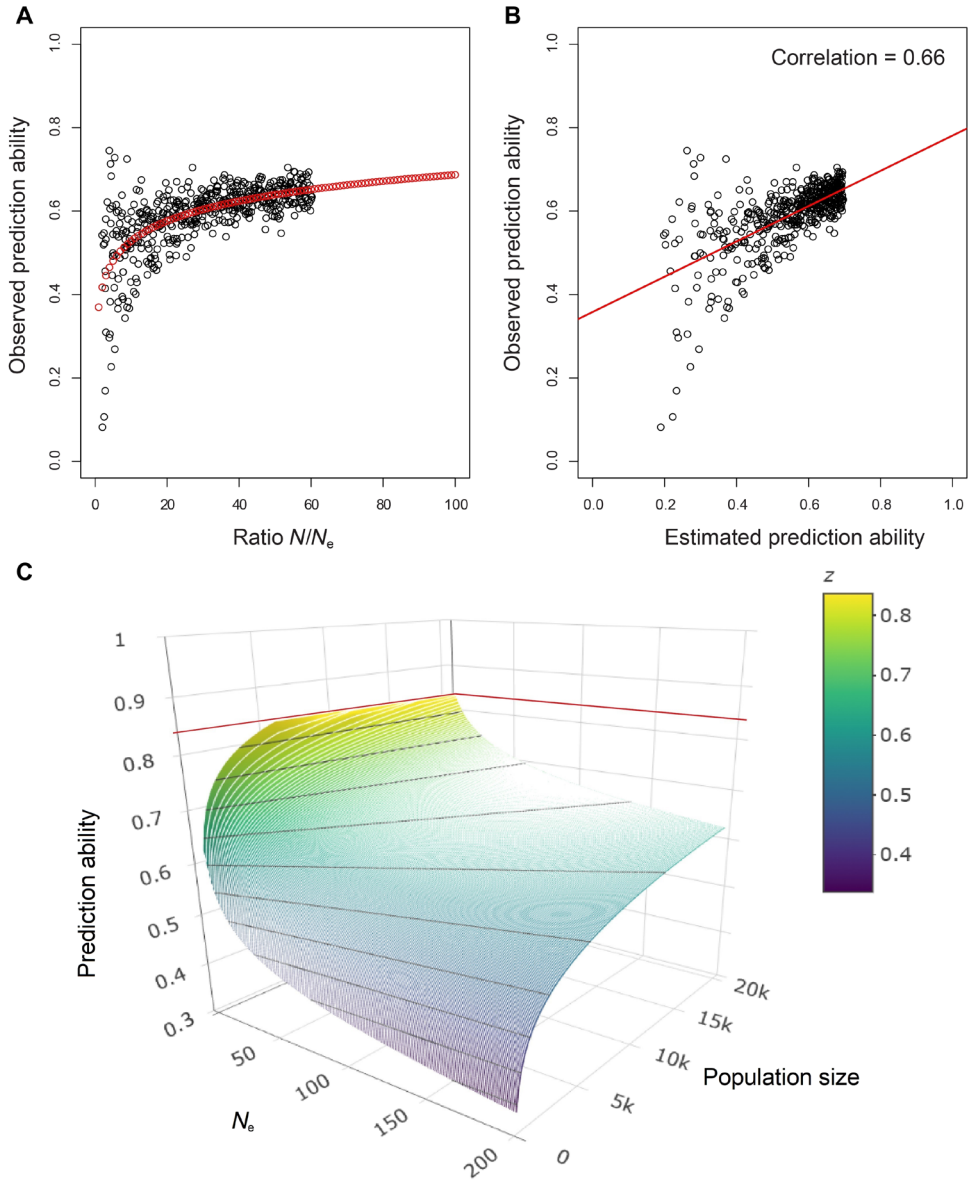

Fig. 4. Relationship between prediction ability and effective population size (N_e) in experimental series VI. (**A**) Biplot of observed prediction ability and the ratio of sample population size (N) versus the effective population size (N_e) in 500 subsamples ranging from $N = 100$ to $N = 3100$ drawn randomly out of experimental series VI. (**B**) Association between observed and estimated prediction accuracies in 500 subsamples drawn randomly out of experimental series VI. (**C**) Projection of prediction ability for population size N from 200 to 20,000 and N_e from 2 to 200; the red line corresponds to the square root of heritability, which represents the upper limit of the prediction ability.

ability with increasing N:N_e, with the largest increase occurring in the range of N:N_e between 2 and 40. We used a nonlinear regression based on data from experimental series VI, estimated relevant parameters, and extrapolated the prediction abilities for N from 200 to 20,000 and N_e from 2 to 200 (Fig. 4C). The trend not only highlights the potential to increase substantially the prediction ability by accumulating more data but also shows the limit determined by the heritability of the phenotypic data of the training population. The latter represents a major obstacle, as plant breeding programs use multistage selection: In the first stages, a large number of individuals are phenotyped in few environments to maximize the selection intensity, and only in the last stage are a small number of individuals evaluated in a large number of environments to maximize heritability. Thus, to assemble a large training population that has been intensively phenotyped, data must be collected across years or even across breeding programs.

We further studied whether marker density is a major limiting factor for the prediction ability by resampling subsets of markers in the entire dataset comprising experimental series I to VI. The fivefold cross-validation showed a substantial increase in prediction abilities when the number of markers was increased from 106 (1%) to 3508 (33.3%), but the increase flattened out when the number of markers exceeded 5261 (50%; fig. S8). The marker density, therefore, is not a major limiting factor in our study. In summary, the compilation and integration of comprehensive datasets for the training of models for hybrid prediction not only have the potential to improve out-of-sample hybrid prediction ability but also challenge current breeding practices.

DISCUSSION

The characteristics of big data differ between crop and animal breeding in comparison to human genetics. In crop and animal breeding, heritability can be increased by massive phenotyping of progenies (*30*, *31*), the effective population sizes are much smaller compared with human genetics (*20*, *32*), and genetic variance among offspring is the key source exploited in selection programs (*8*). In human genetics, on the other hand, the heritability of a particular trait often depends on nonrepeated observations and can hardly be influenced by the experimenter, populations with large effective size and unrelated individuals are often used, while genetic variance within families is considered as noise (*8*, *12*). The latter assumption is now also relevant for crop breeding when recurrent speed-breeding programs with several selection cycles per year are implemented, resulting in a decrease in relatedness between training and test populations (*33*). Recent results in not only human genetics but also animal breeding (*8*) have shown that big data lead to a substantial reduction in the gap between trait heritability and the genotypic variance, which can be explained with a genomic predictor. Accordingly, our study, as an attempt to aggregate medium-sized datasets of around 2000 genotypes into big data of around 13,000 genotypes in wheat breeding, showed that it was possible to increase the prediction ability of the hybrid performance by 34% when comparing the average value for the T_0 scenario within individual experimental series (Fig. 3, A and C) with those across several experimental series (table S9). The prediction abilities that were achieved in our study exceeded those expected based on studies simulating individual breeding programs (*34*). This can be explained by the higher heritability and the larger training population in our study compared with the simulation study. However, the assumed heritability in the simulation study is based on the optimal allocation of resources in multistage selection programs, and thus, much higher heritabilities in individual breeding programs are unlikely. This suggests that it is a particular adequate strategy to combine data across individual breeding programs to fully exploit the potential of predictive plant breeding. Likewise, increasing the ratio between the number of parents tested on grain yield in hybrid backgrounds and the effective population size is a promising approach to improve the prediction ability of the hybrid performance. In this context, incomplete factorial mating designs with balanced missing hybrid patterns, as used in experimental series II and III (fig. S3), or topcross designs, as used in experimental series IV and V, are efficient and can be optimized by maximizing the connectivity and diversity between training and test populations (*35*). Moreover, comprehensive line datasets such as experimental series VI can also provide valuable data for increasing the prediction abilities of hybrids at least in the transitional phase from line to hybrid breeding.

Reliable genome-wide prediction models based on extensive training populations allow the exploration of a large potential genetic space by predicting the performance of millions of single-cross hybrids (*4*–*6*, *36*). For example, this is fundamental for genome-based establishment of heterotic groups (*6*). Moreover, genome-wide predictions can boost the selection intensity in hybrid breeding and, thus, the selection gain (*5*). We have illustrated the latter point by predicting the grain yield of all 10,082,295 potential single hybrids of the 4491 lines of experimental series I to V (fig. S9). In total, 3591 untested single-cross hybrids had a predicted yield higher than that of the best predicted parental line. The best predicted hybrid not tested to date is expected to exceed with 11.7 Mg ha^{-1} the parental line with the highest predicted yield of 11.2 Mg ha^{-1}.

The era of genome-wide selection using big data could further benefit from a revision of the genetic (*6*, *35*) as well as experimental designs of grain yield trials presently in use. The current medium-sized datasets in plant breeding often reflect sequential experimental phenotyping series that lead to a block(experiment)–wise missing value structure in the integrated phenotypic dataset: A subset of genotypes is evaluated in a subset of environments with a small number of overlapping entries (table S3). The latter allow an estimation of the main effects of the environments but do not allow separation of the genotype main effects from the interaction effects between genotype and environmental series for those genotypes, which have not been evaluated across series. A key challenge in further improving prediction ability is, therefore, to reduce the influence of interaction effects between genotypes and experimental series. The pronounced interaction effects between genotypes and experimental series are most likely the result of a reduced representation of the environmental diversity in the different experimental series during phenotyping. On the basis of data of experimental series II, a simple approach to increase the environmental diversity while keeping the number of plots in a very similar range is shown (Fig. 5, A to C). As a baseline, we randomly sampled grain yield data of all lines and hybrids for 3 of the 12 environments, corresponding to 6072 plots, and estimated the correlation between grain yield estimates for the data of the subsets and the total 12 environments (Fig. 5, A and D). The correlation varied widely with a 25% quantile of $r = 0.66$, revealing the pronounced variation due to a low environmental diversity in the subset of three environments. This variation in correlation is very much reduced, albeit at a lower mean

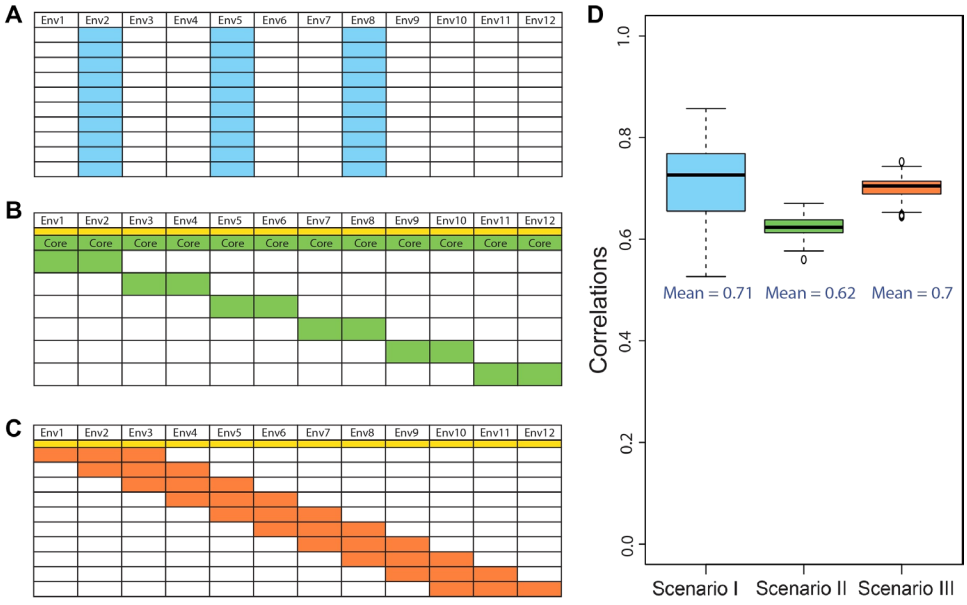

Fig. 5. Optimized field designs to reduce genotype-by-environment interaction effects exemplified on the basis of yield trials of experimental series II in 12 environments. (**A**) In scenario I, all lines and hybrids are tested in a subset of three environments (Env). (**B**) In scenario II, a core of 10% of the lines and hybrids is sampled and tested in all 12 environments together with 11 check varieties (yellow color). The remaining 90% of lines and hybrids are divided into six groups of equal size and tested in two environments. (**C**) In scenario III, the lines and hybrids are divided into 10 subgroups, each of which is tested in only three environments, with the restriction that two environments overlap with those of the next group. All 12 environments are linked with 11 check varieties (yellow color). (**D**) Correlation between grain yield estimates for the data of the subsets of scenario I, II, and III and those for all 12 environments.

value, when testing a core set of 10% of the lines and hybrids and 11 check varieties in all 12 environments and the remaining 90% of lines and hybrids in six groups of equal size in 2 environments, requiring 6465 plots (Fig. 5, B and D). Reducing the variation of correlations at a comparable mean value to the base line model is, however, possible by dividing the lines and hybrids into 10 subgroups, each of which is tested in only three environments, with the restriction that two environments overlap with those of the next group. All environments are linked with 11 check varieties, giving a total of 6455 plots (Fig. 5, C and D). Further improvement may be possible by selecting the subgroups of hybrids in such a way that a close relationship (T_2 scenario) between the different subgroups is ensured, and this relationship is taken into account in the phenotypic data analyses. Alternatively, innovations in phenotyping can be used to identify environmental drivers for interaction effects between genotypes and environments (*37*). The information on the environmental drivers can then be integrated as covariables into the statistical analyses to obtain more accurate estimates of the genotype main effects, thus reducing the estimation bias caused by interaction effects between genotypes and experimental series (*38*, *39*). In summary, an optimized design of multienvironment yield trials in the era of genomic selection coupled with innovations in an integrated analysis of field trials promises to increase the accuracy of predictive plant breeding based on big data.

MATERIALS AND METHODS

Plant materials and field trials

The study includes plant material and phenotypic data from six experimental series. Experimental series I was based on 135 elite winter bread wheat lines and their 1604 single-cross hybrid progenies. Details of the plant material and phenotypic data have been published in a previous study (*6*). Parental lines have been chosen to reflect a wide range of the diversity that exists in Central Europe. The lines were divided into a female pool of 120 lines and a male pool of 15 lines, depending on pollination capacity, plant height, and flowering time. A factorial mating design was used to produce 1604 single-cross hybrids (fig. S3). The 135 parental lines, 1604 hybrids, and 10 other check varieties were tested for grain yield (Mg ha^{-1}) in 11 environments (5 sites in 2012 and 6 sites in 2013) in Germany (table S10). In each environment, the experimental design consisted of three trials. In each trial, a partially replicated alpha lattice design was used. Different genotypes were evaluated in different trials linked by 10 common checks. Plot sizes ranged from 5 to 7.4 m^2.

Experimental series II was based on 226 elite winter bread wheat lines and their 1815 single-cross hybrid progenies. Details of the plant material and phenotypic data have been published in a previous study (*18*). Briefly, parental lines have been chosen to reflect a wide range of the diversity that exists in Central Europe. The lines were divided into a female pool of 185 lines and a male pool of 41 lines, depending on pollination capacity, plant height, and flowering time. A factorial mating design was used to produce 1815 single-cross hybrids (fig. S3). The 226 parental lines, 1815 hybrids, and 11 common checks were tested for grain yield (Mg ha^{-1}) in 12 environments (6 sites in 2016 and 6 sites in 2017) in Germany (table S10). In each environment, the experimental design consisted of three trials. In each trial, an unreplicated alpha lattice design was used. Different genotypes were evaluated in different trials linked by the 11 common checks. Plot sizes ranged from 5.70 to 10.00 m^2.

Experimental series III was based on 236 elite winter bread wheat lines and their 1744 single-cross hybrid progenies. Parental lines have been chosen to reflect a wide range of the diversity that exists

in Central Europe. The lines were divided into a female pool of 196 lines and a male pool of 40 lines, depending on pollination capability, plant height, and flowering time. A factorial mating design was used to produce 1744 single-cross hybrids (fig. S3). The 236 parental lines, 1744 hybrids, and 11 additional check varieties were evaluated for grain yield (Mg ha^{-1}) in six sites in 2018 in Germany (table S10). In each environment, the experimental design consisted of three trials. In each trial, an unreplicated alpha lattice design was used. Different genotypes were evaluated in different trials linked by the 11 common checks. Plot sizes ranged from 5.70 to 9.00 m^2.

Experiment IV was based on 128 elite winter bread wheat lines and their 480 single-cross hybrid progenies. Parental lines have been chosen to reflect a wide range of the diversity that exists in Central Europe. The lines were divided into a female pool of 8 lines and a male pool of 120 lines, depending on pollination capability, plant height, and flowering time. A factorial mating design was used to produce 480 single-cross hybrids (fig. S3). The 128 parental lines and 480 hybrids were split into two series linked by 16 common checks. Series 1 was evaluated for grain yield (Mg ha^{-1}) in 11 environments (6 sites in 2016 and 5 sites in 2017) in Germany (table S10). Series 2 was also evaluated for grain yield (Mg ha^{-1}) in 12 environments (6 sites in 2017 and 6 sites in 2018) in Germany (table S10). An unreplicated alpha lattice design was used. Plot sizes ranged from 5.7 to 10.50 m^2.

Experimental series V included 932 hybrids between elite lines and historic varieties or accessions obtained from the gene bank of the IPK Gatersleben. Six hundred sixty-seven hybrids were produced by crossing 45 elite winter bread wheat lines adapted to the growing conditions of Central Europe with 361 diverse accessions. Here, the elite lines were used as females in hybrid seed production. The accessions were used as male parents and were selected by screening a sample of 4575 gene bank accessions from the gene bank of IPK Gatersleben for pronounced anther extrusion. According to already published passport data (*40*) and information directly obtained from the Genebank Information System of IPK Gatersleben (GBIS: https://gbis.ipk-gatersleben.de/gbis2i/; *41*), the acquisition date of ~47% of these accessions predates the year 1970, and more than 60 worldwide origins were represented in this gene bank sample. In addition, 265 hybrids were produced by crossing 258 historic varieties with plants originating from four different seed mixtures, each including either two or three elite male lines (fig. S4). Elite lines with good anther extrusion but which showed different flowering times were combined in the four mixtures and used as male crossing partners to optimize the hybrid seed production by an almost perfect match of flowering time between male and female lines and to guarantee the unambiguous identification of hybrids. The historic varieties originated from all over Europe from the past four decades and were characterized by a short plant height. The parental lines, 932 hybrids, and 28 to 32 additional common checks were evaluated for grain yield (Mg ha^{-1}) in up to five sites in three trials in Germany (table S10). Trials 1, 2, and 3 included 621, 618, and 500 entries evaluated in the years 2016, 2017, and 2018, respectively. An unreplicated alpha lattice design was used. Plot sizes ranged from 6 to 9 m^2.

Experimental series VI was based on 4972 Central European elite winter wheat lines of the breeding program of KWS LOCHOW GmbH (Bergen, Germany). Part of the phenotypic data of the lines evaluated in it have been published in a previous study (*20*). Briefly, the lines were evaluated in the years 2012, 2013, 2014, and 2015 for grain yield in up to 10 sites in Germany. The lines were divided into 13 to 18 individual trials connected through five to six common checks. The experimental design for each trial followed an alpha design with one to three replications per site, with the number of entries per trial ranging from 30 to 306. Plot size ranged from 6.05 to 17.25 m^2. In all six experimental series, harvesting was performed mechanically, and the harvest was adjusted to a moisture content of 140 g H_2O kg^{-1}.

Curation of phenotypic data

A linear mixed model was used including the effects of genotypes, trials, replications nested within trials, and blocks nested within trials and replications. All data were screened for outliers (fig. S1) using the method 4 "Bonferroni-Holm with rescaled median absolute deviation standardized residuals" as suggested previously (*27*). Outliers were removed, and best linear unbiased estimations (BLUEs) of the genotypes in each environment were obtained as outlined in detail elsewhere (*18*) and served as the input for the subsequent analyses. All linear mixed models were implemented using the software ASReml-R 3.0 (*42*).

Genomic data analyses

The genomic profiles of 5042 lines were determined using 15,000 or 90,000 SNP arrays based on an Illumina Infinium assay (*26*). The number of markers in each experimental series ranged from 11,736 to 81,489. To reduce the risk of a high proportion of missing values in the integrated data, we used only common SNP markers across all six experiments. For the remaining 10,564 common SNP markers, we observed that the missing values were less than 10%, and these missing data were imputed with software IMPUTE2 (*43*). After imputation, we removed the monomorphic markers, and the remaining 10,522 SNP markers were used for subsequent analyses. Marker profiles of the hybrids were deduced from the corresponding parental lines. On the basis of the SNP profiles, a principal coordinate analysis was performed for 5042 lines, for which both phenotypic and genotypic data were available and/or if they represented parental lines of hybrids. The F_{ST} statistic for each pair of experiments was estimated using the method of Weir and Cockerham (*44*) as implemented in the R package "hierfstat" (*45*) and visualized by a neighbor-joining tree using the R package "ape" (*46*). LD between all pairs of SNP markers within each chromosome was calculated as the squared Pearson correlation coefficient (r^2) between vectors of SNP alleles using the 5042 lines. The persistence of linkage phase between the experiments was inferred by analyzing how similar or dissimilar the correlations between pairs of markers were following the approach suggested previously (*47*). Briefly, as the r^2 values do not allow to differentiate between a positive and a negative correlation, we calculated LD among all pairs of markers with a physical distance smaller than 100 Mbp as the correlation coefficient r, where r can take values between −1 and 1. The squared correlation between r values of two different experiments was defined as LD phase and plotted against the physical map distance to fit natural smoothing splines.

Broad-sense heritability for grain yield

A two-step procedure was applied to analyze the grain yield data across environments (*48*). In the first step, the data for each environment were analyzed separately. A linear mixed model was used including the effects of genotypes, trials, replications nested within trials, and blocks nested within trials and replications. BLUEs of the genotypes in each environment were obtained and served as the

input of the second step, where a linear mixed model was applied including the effects of environments and genotypes. Fixed genotypic effects were assumed to obtain the BLUEs of the genotypic values of the hybrids and their parents. Within each experimental series, broad-sense heritability was calculated with a one-step model. For each experimental series, a submodel of the following general model was applied, and only those factors relevant to the experimental design and the population used in a certain experimental series were retained

$$\begin{aligned}\text{Yield}\sim&\text{Group} + \text{Env} + \text{Env:Trial} + \text{Env:Trial:Block} + \text{Line} + \\ &\text{GCA}_\text{F} + \text{GCA}_\text{M} + \text{SCA} + \text{Env:Line} + \text{Env:GCA}_\text{F} + \text{Env:} \\ &\text{GCA}_\text{M} + \text{Env:SCA} + \text{residual(Env)}\end{aligned} \quad (1)$$

While the "Group" effect included specific means of hybrids and lines, "Env," "Trial," and "Block" were the effects of environments, trials, and blocks, respectively. The main effect of lines was denoted as "Line". The main effect of hybrids was decomposed into "GCA_F", "GCA_M", and "SCA" effects, which refer to the general combining ability (GCA) effects of females and males, and specific combining ability (SCA) effect of hybrids, respectively. The following terms were genotype-by-environment interaction effects. For integrated analysis across experimental series, the broad-sense heritability was calculated on the basis of the BLUEs within each environment with model

$$\begin{aligned}\text{Yield}\sim&\text{Group} + \text{Exp} + \text{Env} + \text{Line} + \text{GCA}_\text{F} + \text{GCA}_\text{M} + \text{SCA} + \text{Env:} \\ &\text{Line} + \text{Env:GCA}_\text{F} + \text{Env:GCA}_\text{M} + \text{Env:SCA} + \text{residual(Env)}\end{aligned} \quad (2)$$

While "Exp" refers to the effects of the experimental series, the other parameters are the same as model 1. All effects except "Group" and "Exp" are set as random effects, and we use a heterogeneous variance model for the residuals in each environment. All linear mixed models were implemented using the software ASReml-R 3.0 (*42*).

Combining phenotypic and genomic data

Groups of genotypes with pairwise Rogers' distances below 0.03 were defined to be duplicates and merged for the integrated phenotypic data analyses. Together, 484 duplicate groups were identified representing 1168 lines or hybrids. These included 78 groups of hybrids and 406 groups of lines. The final genomic dataset comprised 10,737 unique lines (4491) and hybrids (6246); the latter were derived by crossing 456 male and 720 female lines. This population of 10,737 genotypes for which 10,522 high-quality SNP markers had been assessed was used for the genome-wide prediction analyses.

Genomic prediction and validation scenarios of the prediction ability

We used a genomic best linear unbiased prediction model (G-BLUP) including additive and dominance effects. The G-BLUP model has the following form

$$Y = T + G_\text{A} + G_\text{D} + e \quad (3)$$

where T is a fixed effect of the overall means within the experimental series or effect of the type of genotype (either lines or hybrids) across the experimental series, and G_A corresponds to the additive genetic values and G_D refers to the dominance genetic values. The additive genetic value was modeled as $G_\text{A} \sim N(0, A\sigma_a^2)$, with the additive relationship matrix being $A = \frac{WW^\text{T}}{2\sum_{k=1}^{p} p_k(1-p_k)}$, $W = C_n Z_\text{A}$, and C_n is the centering matrix, n is the number of genotypes, Z_A is design matrix for additive markers, p_k is the allele frequency of k-th marker, and p is the number of markers. The dominance genetic values were modeled as $G_\text{D} \sim N(0, D\sigma_d^2)$, with the dominance relationship matrix being $D = \frac{VV^\text{T}}{4\sum_{k=1}^{p} p_k^2(1-p_k)^2}$, and V is the general orthogonal design matrix for dominance marker effects (*49*). Details of the method can be found in (*6*). The above models were implemented using the R package BGLR (*50*) with 30,000 iterations, with the first 3000 iteration used as burn-in.

We used chessboard-like (experimental series I, II, and III) and random fivefold cross-validations (experimental series IV, V, and VI) to evaluate the prediction ability of genomic prediction within experimental series (fig. S5). Basically, data were divided into two sets, a training set and a test set. The G-BLUP model was trained using genomic and phenotypic data of the training set. The genomic data of the test set were used to predict the genetic values of hybrids and lines. The prediction ability for each test set was estimated as the Pearson correlation coefficient between the predictions and the observed phenotypic values.

In addition, we tested the prediction ability across experimental series using different combinations of training sets. In the first scenario, we used one out of the six experimental series as training set. Each of the other experimental series was used as test set. The prediction ability for each test set was estimated as the Pearson correlation coefficient between the predicted and the observed hybrid performances. In the second scenario, we used either experimental series I, II, or III as test set. For the training sets, we incrementally added the experimental series except the one used as test set. The prediction ability for each test set was estimated again as the Pearson correlation coefficient between the predicted and the observed hybrid performances.

Effective population size and estimation of prediction ability

We estimated the effective population size N_e as

$$N_\text{e} = \frac{k}{3*\left(\widehat{r^2} - \frac{1}{n}\right)} \quad (4)$$

where n is the harmonic mean of the sample size, $\widehat{r^2}$ is the expected LD between unlinked loci, and $k = 1$ for monoecious and $k = 2$ for dioecious plants (*51*, *52*).

When the heritability (h^2) is known, the estimated prediction ability ($\widehat{p_\text{est}}$) is the square root of heritability multiplied by the estimated prediction accuracy ($\widehat{r_\text{est}}$). The estimated prediction accuracy depends on the heritability of the trait and the ratio between the effective number of segments in the genome and the number of individuals in the training population (*16*)

$$\begin{cases}\widehat{r_\text{est}} = \sqrt{\dfrac{h^2}{h^2 + \frac{4N_\text{e}Lv}{N}}} \\ \widehat{p_\text{est}} = h*\widehat{r_\text{est}}\end{cases} \quad (5)$$

where L denotes the genome size in Morgan, $4N_eLv$ is the effective number of segments in the genome assuming an infinitesimal model,

and N is the size of the training population. From Eq. 5, it can be concluded that if heritability and marker density are fixed, the ability of genome-wide prediction is positively correlated with $\frac{N}{N_e}$, where $\frac{N}{N_e}$ is the ratio between the size of the training population and the effective population size. We randomly sampled 500 subpopulations from experiments VI with N ranging from 100 to 3100 and a range in the ratio of N/N_e from 2 to 60 between the different subsets. For each subpopulation, a fivefold cross-validation was applied to obtain the observed prediction ability $\widehat{p_{obs}}$. We also calculated the broad-sense heritability of each subpopulation and used Eq. 5 to obtain $\widehat{p_{est}}$. The correlation between $\widehat{p_{obs}}$ and $\frac{N}{N_e}$ or between $\widehat{p_{obs}}$ and $\widehat{p_{est}}$ was analyzed. Moreover, we used a nonlinear regression model $\widehat{p_{obs}} = a + b*\ln\left(\frac{N}{N_e}\right) + \varepsilon$ (*53*) to estimate the relationship between $\frac{N}{N_e}$ and $\widehat{p_{obs}}$. All the above analyses use software R version 3.6.0 (*54*).

Optimized field designs to reduce genotype-by-environment interaction effects

On the basis of the data of experimental series II, we tested in silico three scenarios of field designs, each requiring a similar number of plots. The full data of experimental series II included 11 checks, 226 elite lines, and 1815 single-cross hybrid progenies. In scenario I, a balanced missing design was considered in which all lines and hybrids were tested in three randomly selected environments, corresponding to 6072 plots. We analyzed all 220 combinations of three environments and estimated the across environment BLUEs in each subset. In scenario II, a core of 10% of the elite lines and 10% of hybrids were sampled and tested in all 12 environments with 11 checks. The remaining 90% of elite lines and hybrids were divided into six groups of equal size and tested in two environments. The random sampling was run for 220 times as in scenario I, and the average number of plots used in this scenario was 6465. In scenario III, all lines and hybrids were divided into 10 subgroups, each of which was tested in only three environments, with the restriction that two environments overlapped with those of the next group. The 11 checks were tested in all environments to estimate the environmental effects, and the average number of plots used in this scenario corresponded to 6455. The model used to estimate BLUEs for all three scenarios was $Y = \mu + G + G * E + e$, where μ is overall mean, G is the genotypic effect of lines and hybrids, $G * E$ is the genotype-by-environment interaction effect, and e corresponds to the residuals. The Pearson correlation between BLUEs from subsets of scenarios I, II, and III, and the total data including 12 environments were used to estimate the precision of the estimates of the genotype effects.

Validation experiment to estimate the role of interactions between genotypes and experimental series

To assess the role of interactions between genotypes and experimental series, we generated 148 previously untested T_2 hybrids that had both parental lines in common with experimental series II but were not tested in experimental series II. We selected the 148 hybrids out of the 23,610 potential hybrids using the predicted grain yield performance and further information on producibility of single-cross hybrids, i.e., anther extrusion, plant height, and flowering time. Ninety-six of the 148 hybrids belong to the highest yielding hybrids (>90% quantile), and 30 belong to the lowest yielding hybrids (< 10% quantile). The yield of the other 22 hybrids is somewhere in between these two groups. The 148 hybrids were phenotyped in a separate validation experiment for grain yield in eight environments in Germany in the year 2019. We estimated the BLUEs as outlined above and studied their correlation with the predicted hybrid performance using data from experimental series II.

Predicting the yield performance of all potential single-cross hybrids from the 4491 lines

We used the complete dataset and a ridge regression BLUP model with additive and dominance marker effects to predict the hybrid performance of all potential single-cross hybrids between the 4491 inbred lines. The model is as follows

$$Y = T + Z_A a + Z_D d + e \quad (6)$$

where Z_A and Z_D are the design matrices for additive and dominance markers, the elements of Z_A are −1, 0, 1, while the homozygotes are coded as −1 and 1, and the heterozygotes are coded as 0. The elements of Z_D are 0 and 1, while the two homozygote classes are coded as 0, and the heterozygotes are coded as 1. We use the ridge regression BLUP model in this section because it is not efficient to predict all 10,082,295 potential hybrids using GBLUP.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/24/eabf9106/DC1

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. P. Juliana, J. Poland, J. Huerta-Espino, S. Shrestha, J. Crossa, L. Crespo-Herrera, F. H. Toledo, V. Govindan, S. Mondal, U. Kumar, S. Bhavani, P. K. Singh, M. S. Randhawa, X. He, C. Guzman, S. Dreisigacker, M. N. Rouse, Y. Jin, P. Perez-Rodriguez, O. A. Montesinos-Lopez, D. Singh, M. M. Rahman, F. Marza, R. P. Singh, Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics. *Nat. Genet.* **51**, 1530–1539 (2019).
2. D. K. Ray, N. D. Mueller, P. C. West, J. A. Foley, Yield trends are insufficient to double global crop production by 2050. *PLOS ONE* **8**, e66428 (2013).
3. Y. Jiang, R. H. Schmidt, Y. Zhao, J. C. Reif, A quantitative genetic framework highlights the role of epistatic effects for grain-yield heterosis in bread wheat. *Nat. Genet.* **49**, 1741–1746 (2017).
4. A. R. Rogers, J. C. Dunne, C. Romay, M. Bohn, E. S. Buckler, I. A. Ciampitti, J. Edwards, D. Ertl, S. Flint-Garcia, M. A. Gore, C. Graham, C. N. Hirsch, E. Hood, D. C. Hooker, J. Knoll, E. C. Lee, A. Lorenz, J. P. Lynch, J. McKay, S. P. Moose, S. C. Murray, R. Nelson, T. Rocheford, J. C. Schnable, P. S. Schnable, R. Sekhon, M. Singh, M. Smith, N. Springer, K. Thelen, P. Thomison, A. Thompson, M. Tuinstra, J. Wallace, R. J. Wisser, W. Xu, A. R. Gilmour, S. M. Kaeppler, N. D. Leon, J. B. Holland, The importance of dominance and genotype-by-environment interactions on grain yield variation in a large-scale public cooperative maize experiment. *G3 (Bethesda)* **11**, jkaa050 (2021).
5. S. Xu, D. Zhu, Q. Zhang, Predicting hybrid performance in rice using genomic best linear unbiased prediction. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 12456–12461 (2014).
6. Y. Zhao, Z. Li, G. Liu, Y. Jiang, H. P. Maurer, T. Würschum, H. P. Mock, A. Matros, E. Ebmeyer, R. Schachschneider, E. Kazman, J. Schacht, M. Gowda, C. F. H. Longin, J. C. Reif, Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15624–15629 (2015).
7. T. H. Meuwissen, B. J. Hayes, M. E. Goddard, Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829 (2001).
8. N. R. Wray, K. E. Kemper, B. J. Hayes, M. E. Goddard, P. M. Visscher, Complex trait prediction from genome data: Contrasting EBV in livestock to PRS in humans: Genomic prediction. *Genetics* **211**, 1131–1141 (2019).
9. X. Huang, S. Yang, J. Gong, Q. Zhao, Q. Feng, Q. Zhan, Y. Zhao, W. Li, B. Cheng, J. Xia, N. Chen, T. Huang, L. Zhang, D. Fan, J. Chen, C. Zhou, Y. Lu, Q. Weng, B. Han, Genomic architecture of heterosis for yield traits in rice. *Nature* **537**, 629–633 (2016).
10. J. M. Hickey, T. Chiurugwi, I. Mackay, W. Powell, Implementing Genomic Selection in CGIAR Breeding Programs Workshop Participants, Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat. Genet.* **49**, 1297 (2017).
11. M. Rembe, Y. Zhao, Y. Jiang, J. C. Reif, Reciprocal recurrent genomic selection: An attractive tool to leverage hybrid wheat breeding. *Theor. Appl. Genet.* **132**, 687–698 (2019).

12. J. Yang, J. Zeng, M. E. Goddard, N. R. Wray, P. M. Visscher, Concepts, estimation and interpretation of SNP-based heritability. *Nat. Genet.* **49**, 1304–1310 (2017).
13. H. Kim, A. Grueneberg, A. I. Vazquez, S. Hsu, G. de los Campos, Will big data close the missing heritability gap? *Genetics* **207**, 1135–1145 (2017).
14. H. D. Daetwyler, B. Villanueva, J. A. Woolliams, Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLOS ONE* **3**, e3395 (2008).
15. P. M. Visscher, J. Yang, M. E. Goddard, A commentary on 'common SNPs explain a large proportion of the heritability for human height' by Yang *et al*. (2010). *Twin Res. Hum. Genet.* **13**, 517–524 (2010).
16. T. H. Meuwissen, Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. *Genet. Sel. Evol.* **41**, 35 (2009).
17. T. A. Schrag, W. Schipprack, A. E. Melchinger, Across-years prediction of hybrid performance in maize using genomics. *Theor. Appl. Genet.* **132**, 933–946 (2019).
18. P. H. G. Boeven, Y. Zhao, P. Thorwarth, F. Liu, H. P. Maurer, M. Gils, R. Schachschneider, J. Schacht, E. Ebmeyer, E. Kazman, V. Mirdita, J. Dörnte, S. Kontowski, R. Horbach, H. Cöster, J. Holzapfel, A. Jacobi, L. Ramgraber, C. Reinbrecht, N. Starck, P. Varenne, A. Starke, F. Schürmann, M. Ganal, A. Polley, J. Hartung, S. Beier, U. Scholz, C. F. H. Longin, J. C. Reif, Y. Jiang, T. Würschum, Negative dominance and dominance-by-dominance epistatic effects reduce grain-yield heterosis in wide crosses in wheat. *Sci. Adv.* **6**, eaay4897 (2020).
19. B. R. Basnet, J. Crossa, S. Dreisigacker, P. Perez-Rodriguez, Y. Manes, R. P. Singh, U. R. Rosyara, F. Camarillo-Castillo, M. Murua, Hybrid wheat prediction using genomic, pedigree, and environmental covariables interaction models. *Plant Genome* **12**, 180051 (2019).
20. S. He, J. C. Reif, V. Korzun, R. Bothe, E. Ebmeyer, Y. Jiang, Genome-wide mapping and prediction suggests presence of local epistasis in a vast elite winter wheat populations adapted to Central Europe. *Theor. Appl. Genet.* **130**, 635–647 (2017).
21. International Wheat Genome Sequencing Consortium (IWGSC), R. Appels, K. Eversole, N. Stein, C. Feuillet, B. Keller, J. Rogers, C. J. Pozniak, F. Choulet, A. Distelfeld, J. Poland, G. Ronen, A. G. Sharpe, O. Barad, K. Baruch, G. Keeble-Gagnère, M. Mascher, G. Ben-Zvi, A.-A. Josselin, A. Himmelbach, F. Balfourier, J. Gutierrez-Gonzalez, M. Hayden, C. S. Koh, G. Muehlbauer, R. K. Pasam, E. Paux, P. Rigault, J. Tibbits, V. Tiwari, M. Spannagl, D. Lang, H. Gundlach, G. Haberer, K. F. X. Mayer, D. Ormanbekova, V. Prade, H. Šimková, T. Wicker, D. Swarbreck, H. Rimbert, M. Felder, N. Guilhot, G. Kaithakottil, J. Keilwagen, P. Leroy, T. Lux, S. Twardziok, L. Venturini, A. Juhász, M. Abrouk, I. Fischer, C. Uauy, P. Borrill, R. H. Ramirez-Gonzalez, D. Arnaud, S. Chalabi, B. Chalhoub, A. Cory, R. Datla, M. W. Davey, J. Jacobs, S. J. Robinson, B. Steuernagel, F. van Ex, B. B. H. Wulff, M. Benhamed, A. Bendahmane, L. Concia, D. Latrasse, J. Bartoš, A. Bellec, H. Berges, J. Doležel, Z. Frenkel, B. Gill, A. Korol, T. Letellier, O.-A. Olsen, K. Singh, M. Valárik, E. van der Vossen, S. Vautrin, S. Weining, T. Fahima, V. Glikson, D. Raats, J. Číhalíková, H. Toegelová, J. Vrána, P. Sourdille, B. Darrier, D. Barabaschi, L. Cattivelli, P. Hernandez, S. Galvez, H. Budak, J. D. G. Jones, K. Witek, G. Yu, I. Small, J. Melonek, R. Zhou, T. Belova, K. Kanyuka, R. King, K. Nilsen, S. Walkowiak, R. Cuthbert, R. Knox, K. Wiebe, D. Xiang, A. Rohde, T. Golds, J. Čížková, B. A. Akpinar, S. Biyiklioglu, L. Gao, A. N'Daiye, M. Kubaláková, J. Šafář, F. Alfama, A.-F. Adam-Blondon, R. Flores, C. Guerche, M. Loaec, H. Quesneville, J. Condie, J. Ens, R. Maclachlan, Y. Tan, A. Alberti, J.-M. Aury, V. Barbe, A. Couloux, C. Cruaud, K. Labadie, S. Mangenot, P. Wincker, G. Kaur, M. Luo, S. Sehgal, P. Chhuneja, O. P. Gupta, S. Jindal, P. Kaur, P. Malik, P. Sharma, B. Yadav, N. K. Singh, J. P. Khurana, C. Chaudhary, P. Khurana, V. Kumar, A. Mahato, S. Mathur, A. Sevanthi, N. Sharma, R. S. Tomar, K. Holušová, O. Plíhal, M. D. Clark, D. Heavens, G. Kettleborough, J. Wright, B. Balcárková, Y. Hu, E. Salina, N. Ravin, K. Skryabin, A. Beletsky, V. Kadnikov, A. Mardanov, M. Nesterov, A. Rakitin, E. Sergeeva, H. Handa, H. Kanamori, S. Katagiri, F. Kobayashi, S. Nasuda, T. Tanaka, J. Wu, F. Cattonaro, M. Jiumeng, K. Kugler, M. Pfeifer, S. Sandve, X. Xun, B. Zhan, J. Batley, P. E. Bayer, D. Edwards, S. Hayashi, Z. Tulpová, P. Visendi, L. Cui, X. Du, K. Feng, X. Nie, W. Tong, L. Wang, Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361**, eaar7191 (2018).
22. A. Norman, J. Taylor, J. Edwards, H. Kuchel, Optimising genomic selection in wheat: Effect of marker density, population size and population structure on prediction accuracy. *G3* **8**, 2889–2899 (2018).
23. V. Belamkar, M. J. Guttieri, W. Hussain, D. Jarquín, I. El-basyoni, J. Poland, A. J. Lorenz, P. S. Baenziger, Genomic selection in preliminary yield trials in a winter wheat breeding program. *G3* **8**, 2735–2747 (2018).
24. F. Liu, Y. Zhao, S. Beier, Y. Jiang, P. Thorwarth, C. F. H. Longin, M. Ganal, A. Himmelbach, J. C. Reif, A. W. Schulthess, Exome association analysis sheds light onto leaf rust (*Puccinia triticina*) resistance genes currently used in wheat breeding (*Triticum aestivum* L.). *Plant Biotechnol. J.* **18**, 1396–1408 (2020).
25. J. S. Ward, A. Barker, Undefined by data: A survey of big data definitions (2013); arXiv:1309.5821.
26. S. Wang, D. Wong, K. Forrest, A. Allen, S. Chao, B. E. Huang, M. MacCaferri, S. Salvi, S. G. Milner, L. Cattivelli, A. M. Mastrangelo, A. Whan, S. Stephen, G. Barker, R. Wieseke, J. Plieske; International Wheat Genome Sequencing Consortium, M. Lillemo, D. Mather, R. Appels, R. Dolferus, G. Brown-Guedira, A. Korol, A. R. Akhunova, C. Feuillet, J. Salse, M. Morgante, C. Pozniak, M.-C. Luo, J. Dvorak, M. Morell, J. Dubcovsky, M. Ganal, R. Tuberosa, C. Lawley, I. Mikoulitch, C. Cavanagh, K. J. Edwards, M. Hayden, E. Akhunov, Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnol. J.* **12**, 787–796 (2014).
27. A. M. Bernal-Vasquez, H. F. Utz, H. P. Piepho, Outlier detection methods for generalized lattices: A case study on the transition from ANOVA to REML. *Theor. Appl. Genet.* **129**, 787–804 (2016).
28. F. Technow, T. A. Schrag, W. Schipprack, E. Bauer, H. Simianer, A. E. Melchinger, Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. *Genetics* **197**, 1343–1355 (2014).
29. L. D. Goodwin, N. L. Leech, Understanding correlation: Factors that affect the size of r. *J. Exp. Educ.* **74**, 249–266 (2006).
30. M. S. Lund, A. P. de Roos, A. G. de Vries, T. Druet, V. Ducrocq, S. Fritz, F. Guillaume, B. Guldbrandtsen, Z. Liu, R. Reents, C. Schrooten, F. Seefried, G. Su, A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genet. Sel. Evol.* **43**, 43 (2011).
31. K. P. Voss-Fels, M. Cooper, B. J. Hayes, Accelerating crop genetic gains with genomic selection. *Theor. Appl. Genet.* **132**, 669–686 (2019).
32. A. Tenesa, P. Navarro, B. J. Hayes, D. L. Duffy, G. M. Clarke, M. E. Goddard, P. M. Visscher, Recent human effective population size estimated from linkage disequilibrium. *Genome Res.* **17**, 520–526 (2007).
33. L. T. Hickey, A. N. Hafeez, H. Robinson, S. A. Jackson, S. C. Leal-Bertioli, M. Tester, C. Gao, I. D. Godwin, B. J. Hayes, B. B. Wulff, Breeding crops to feed 10 billion. *Nat. Biotechnol.* **37**, 744–754 (2019).
34. D. Müller, P. Schopp, A. E. Melchinger, Persistency of prediction accuracy and genetic gain in synthetic populations under recurrent genomic selection. *G3* **7**, 801–811 (2017).
35. T. Guo, X. Yu, X. Li, H. Zhang, C. Zhu, S. Flint-Garcia, M. D. McMullen, J. B. Holland, S. J. Szalma, R. J. Wisser, J. Yu, Optimal designs for genomic selection in hybrid crops. *Mol. Plant* **12**, 390–401 (2019).
36. Y. Cui, R. Li, G. Li, F. Zhang, T. Zhu, Q. Zhang, J. Ali, Z. Li, S. Xu, Hybrid breeding of rice via genomic selection. *Plant Biotechnol. J.* **18**, 57–67 (2020).
37. E. J. Millet, W. Kruijer, A. Coupel-Ledru, S. A. Prado, L. Cabrera-Bosquet, S. Lacube, A. Charcosset, C. Welcker, F. van Eeuwijk, F. Tardieu, Genomic prediction of maize yield across European environmental conditions. *Nat. Genet.* **51**, 952–956 (2019).
38. J. Burgueño, G. de los Campos, K. Weigel, J. Crossa, Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. *Crop. Sci.* **52**, 707–719 (2012).
39. X. Li, T. Guo, Q. Mu, X. Li, J. Yu, Genomic and environmental determinants and their interplay underlying phenotypic plasticity. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 6679–6684 (2018).
40. N. Philipp, S. Weise, M. Oppermann, A. Börner, J. Keilwagen, B. Kilian, D. Arend, Y. Zhao, A. Graner, J. C. Reif, A. W. Schulthess, Historical phenotypic data from seven decades of seed regeneration in a wheat ex situ collection. *Sci. Data* **6**, 137 (2019).
41. M. Oppermann, S. Weise, C. Dittmann, H. Knüpffer, GBIS: The information system of the German Genebank. *Database* **2015**, bav021 (2015).
42. D. G. Butler, B. R. Cullis, A. R. Gilmour, B. J. Gogel, R. Thompson, *ASReml-R reference manual (version 3)* (The State of Queensland, Department of Primary Industries and Fisheries: Brisbane, Qld, 2009).
43. B. N. Howie, P. Donnelly, J. Marchini, A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLOS Genet.* **5**, e1000529 (2009).
44. B. S. Weir, C. C. Cockerham, Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
45. J. Goudet, Hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* **5**, 184–186 (2005).
46. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
47. Y. M. Badke, R. O. Bates, C. W. Ernst, C. Schwab, J. P. Steibel, Estimation of linkage disequilibrium in four US pig breeds. *BMC Genomics* **13**, 24 (2012).
48. J. Möhring, H. P. Piepho, Comparison of weighting in two-stage analysis of plant breeding trials. *Crop. Sci.* **49**, 1977–1988 (2009).
49. J. M. Álvarez-Castro, Ö. Carlborg, A unified model for functional and statistical epistasis and its application in quantitative trait loci analysis. *Genetics* **176**, 1151–1167 (2007).
50. P. Pérez, G. de Los Campos, Genome-wide regression and prediction with the BGLR statistical package. *Genetics* **198**, 483–495 (2014).
51. W. G. Hill, Estimation of effective population size from data on linkage disequilibrium. *Genet. Res.* **38**, 209–216 (1981).
52. R. S. Waples, A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conserv. Genet.* **7**, 167 (2006).

53. D. M. Bates, D. G. Watts, *Nonlinear Regression Analysis and Its Applications*, vol. 2 (Wiley, 1988).
54. R Core Team, *A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2015).

Acknowledgments: We thank GFPi and proWeizen for project coordination. **Funding:** This work was funded by BMEL through BLE within the "ZUCHTWERT" project (grant ID: 2814604113). The Federal Ministry of Education and Research of Germany is acknowledged for funding A.W.S. (grant FKZ031B0184A). **Author contributions:** J.C.R. and Y.Z. conceived and designed the study. P.T., N.P., M.G., P.H.G.B., C.F.H.L., J.S., E.E., V.K., V.M., J.D., U.A., R.H., H.C., J.H., L.R., S.K., P.V., A.S., and F.S. acquired and contributed data. Y.Z., P.T., and P.H.G.B. processed the data, performed the analyses, and analyzed the results. S.B. and U.S. did data management. Y.J., F.L., and A.W.S. supported in data analyses. Y.Z., P.T., Y.J., R.H.S., and J.C.R. interpreted the results and wrote the manuscript, and all coauthors provided input. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Data of experimental series I were published earlier (https://doi.org/10.1073/pnas.1514547112). The genomic and phenotypic data of experimental series II, III, IV, and V that supported the findings of this study are available at the electronic data archive library edal (http://dx.doi.org/10.5447/ipk/2020/17). The data of the experimental series VI can be provided by KWS LOCHOW GmbH pending scientific review and a completed material transfer agreement. Requests for the data of the experimental series VI should be submitted to Patrick Thorwarth (patrick.thorwarth@kws.com). Code availability: R codes for implementing genomic prediction is available at the following link: https://github.com/ysz98/Genomic-prediction-of-hybrid-wheat

Submitted 27 November 2020
Accepted 28 April 2021
Published 11 June 2021
10.1126/sciadv.abf9106

Citation: Y. Zhao, P. Thorwarth, Y. Jiang, N. Philipp, A. W. Schulthess, M. Gils, P. H. G. Boeven, C. F. H. Longin, J. Schacht, E. Ebmeyer, V. Korzun, V. Mirdita, J. Dörnte, U. Avenhaus, R. Horbach, H. Cöster, J. Holzapfel, L. Ramgraber, S. Kühnle, P. Varenne, A. Starke, F. Schürmann, S. Beier, U. Scholz, F. Liu, R. H. Schmidt, J. C. Reif, Unlocking big data doubled the accuracy in predicting the grain yield in hybrid wheat. *Sci. Adv.* **7**, eabf9106 (2021).

Unlocking big data doubled the accuracy in predicting the grain yield in hybrid wheat

Yusheng Zhao, Patrick Thorwarth, Yong Jiang, Norman Philipp, Albert W. Schulthess, Mario Gils, Philipp H. G. Boeven, C. Friedrich H. Longin, Johannes Schacht, Erhard Ebmeyer, Viktor Korzun, Vilson Mirdita, Jost Dörnte, Ulrike Avenhaus, Ralf Horbach, Hilmar Cöster, Josef Holzapfel, Ludwig Ramgraber, Simon Kühnle, Pierrick Varenne, Anne Starke, Friederike Schürmann, Sebastian Beier, Uwe Scholz, Fang Liu, Renate H. Schmidt and Jochen C. Reif

Sci Adv **7** (24), eabf9106.
DOI: 10.1126/sciadv.abf9106