Received: 15 November 2022 | Accepted: 28 February 2023

DOI: 10.1002/tpg2.20326

The Plant Genome

ORIGINAL ARTICLE

Special Section: Grain Quality and Nutritional Genomics for Breeding Next Generation Crops

Maintenance of UK bread baking quality: Trends in wheat quality traits over 50 years of breeding and potential for future application of genomic-assisted selection

Nick S. Fradgley[1] | **Alison R. Bentley**[1,2] | **Keith A. Gardner**[1,2] | **Stéphanie M. Swarbreck**[1] | **Matt Kerton**[3]

[1]Genetics and Pre-Breeding Department, National Institute of Agricultural Botany (NIAB), 93 Lawrence Weaver Road, Cambridge, UK

[2]International Maize and Wheat Improvement Center (CIMMYT), Carretera México-Veracruz, México

[3]DSVUK Top Dawkins Barn, Banbury, UK

Correspondence
Nick S. Fradgley, Genetics and Pre-Breeding Department, National Institute of Agricultural Botany (NIAB), 93 Lawrence Weaver Road, Cambridge, CB3 0LE, UK.
Email: nickfradgley@hotmail.com

Assigned to Associate Editor Lee Hickey.

Funding information
Biotechnology and Biological Sciences Research Council, Grant/Award Number: BB/M011194/1; OneCGIAR Accelerated Breeding Initiative

Abstract

Improved selection of wheat varieties with high end-use quality contributes to sustainable food systems by ensuring productive crops are suitable for human consumption end-uses. Here, we investigated the genetic control and genomic prediction of milling and baking quality traits in a panel of 379 historic and elite, high-quality UK bread wheat (*Triticum eastivum* L.) varieties and breeding lines. Analysis of the panel showed that genetic diversity has not declined over recent decades of selective breeding while phenotypic analysis found a clear trend of increased loaf baking quality of modern milling wheats despite declining grain protein content. Genome-wide association analysis identified 24 quantitative trait loci (QTL) across all quality traits, many of which had pleiotropic effects. Changes in the frequency of positive alleles of QTL over recent decades reflected trends in trait variation and reveal where progress has historically been made for improved baking quality traits. It also demonstrates opportunities for marker-assisted selection for traits such as Hagberg falling number and specific weight that do not appear to have been improved by recent decades of phenotypic selection. We demonstrate that applying genomic prediction in a commercial wheat breeding program for expensive late-stage loaf baking quality traits outperforms phenotypic selection based on early-stage predictive quality traits. Finally, trait-assisted genomic prediction combining both phenotypic and genomic selection enabled slightly higher prediction accuracy, but genomic predic-

Abbreviations: 1B/1R, Translocation of rye chromosome 1 to the short arm of wheat chromosome 1B; AACC, American Association of Cereal Chemists; AHDB, Agriculture and Horticulture Development Board; AIC, Akaike Information Criterion; BLAST, basic local alignment search tool; BLUP, Best Linear Unbiased Prediction; CIMMYT, International Maize and Wheat Improvement Center; DSV, Deutsche Saatveredelung; EG-BLUP, Extended GBLUP; G×E, Genotype by environment interaction; GBLUP, genomic best linear unbiased estimate; GBM, Gradient boosted machines; GO, Gene ontology; GRIS, Genetic Resources Information System for Wheat and Triticale; HFN, Hagberg falling number; LASSO, Least Absolute Shrinkage and Selection Operator; Mb, Megabase; NABIM, National Association of British and Irish Millers; NIR, Near-Infrared Reflectance; NMR, Nuclear magnetic resonance; PCoA, Principal coordinate analysis; Pina, Puroindoline a; Pinb, Puroindoline b; QTL, quantitative trait loci; RF, Random forest; RKHS, Reproducing kernel Hilbert space; SDS, Sodium dodecyl sulfate; SNP, single nucleotide polymorphism; TGW, thousand grain weight.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2023 The Authors. *The Plant Genome* published by Wiley Periodicals LLC on behalf of Crop Science Society of America.

tion alone was the most cost-effective selection strategy considering genotyping and phenotyping costs per sample.

1 | INTRODUCTION

Bread wheat (*Triticum eastivum* L.) is a globally important food crop for human consumption, but is also used for livestock feed and biofuel (Shiferaw et al., 2013). Feeding grain crops that contain substantial human digestible protein to livestock is an inefficient use of increasingly scarce resources (Bentley et al., 2022). Enhancing the end-use profile of wheat crops for food uses presents an opportunity to reduce the cost of feeding people per unit of land area and input use (Eisler et al., 2014; Wilkinson & Lee, 2018). High wheat yields are achieved in the United Kingdom (Calderini & Slafer, 1998) albeit with intensive fertilizer use (Lu & Tian, 2017), but only approximately one fifth of wheat varieties bred in the United Kingdom are high-quality milling wheats for direct human consumption (AHDB, 2022), and the proportion of wheat grown in the United Kingdom for food has decreased to less than half in recent decades (Ray et al., 2022). Increasing pressures of climate change and demands of feeding a growing global population will necessitate increased production of high-quality wheat varieties in highly productive northern growing regions such as the United Kingdom.

Since the early days of wheat breeding, a trade-off between grain yield and grain quality has been observed (Percival, 1934), and there is strong evidence supporting a direct genetic trade-off between grain yield and protein content (Michel et al., 2019; Scott et al., 2021; Simmonds, 1995). In wheat breeding programs with an explicit quality focus, simultaneous selection is carried out for a large number of milling and baking quality traits including protein content, high grain specific weight, high Hagberg falling number, good dough rheological properties such as high gluten strength and extensibility, large loaf baking size as well as good crumb structure, texture, and color. Testing the large number of important wheat milling and baking quality traits is expensive and time consuming, so these traits are generally tested in later stages of breeding programs once significant phenotypic and agronomic selection has already been applied. Some high-throughput traits such as grain protein content and specific weight can be measured at earlier stages and are used to make selections and predict more expensive later stage traits. As these quality traits are often only imperfectly correlated to each other, forward selections to later generations are often unreliable and reduced genetic gain is achieved for any single-trait. This challenge, in combination with ideal climatic conditions for obtaining high wheat yields, the wheat protein trade-off, and availability of fertilizer (required to achieve high protein) have resulted in UK wheat breeders primarily focusing on breeding high yielding feed wheat varieties. Over the course of the last century, historical changes in agricultural practice, breeding approaches, and milling and baking processes have all evolved to adapt to this high yield focused wheat cultivation strategy (Belderok et al., 2000). Understanding the trends in past genetic gain in specific component quality traits can shed light on how crop improvements have historically been achieved and can also identify opportunities to inform future selection approaches for enhancing UK wheat quality.

Many breeding programs will run separate selection streams for feed and milling wheat end-use classes. Here, selection of lines with high-quality traits and/or high yield within crosses between milling and feed wheats are particularly important to achieve genetic gain in each stream. Higher throughput and low-cost predictive traits, such as grain protein content and specific weight, are measured for the large number of lines at early-stages in the breeding program and are used both to directly make early-stage selections and to predict more costly later stage traits such as those obtained from loaf baking tests. Enhanced and earlier prediction and selection for these later stage traits would greatly improve capabilities for breeding high-quality milling wheat varieties.

Developments in genotyping technologies have enabled advancements in genomics-assisted breeding in wheat (Adamski et al., 2020). Characterization of specific genetic variants associated with traits through genome wide association studies have identified quantitative trait loci (QTL) that can be used for marker-assisted selection in breeding programs. The role of the three homoeologous high-molecular-weight glutenin genes in determining wheat quality was identified by Payne et al. (1982), and genetic markers for these loci are now routinely used in breeding programs (Liang et al., 2010; Lupton, 2005). Likewise, identification of the genetic control of endosperm texture and grain hardness by the *Puroindoline a (Pina)* and *Puroindoline b (Pinb)* loci (Morris, 2002) and of the negative effects of the 1B/1R translocation from rye on Hagberg falling number and gluten strength (Graybosch, 2001) have proved to be highly applicable in wheat breeding. More recently, genetic control of quality traits for human health, such as increased dietary fiber (Ibba et al., 2021), have been identified and can be employed in breeding. These examples highlight the usefulness of QTL that explain a large proportion of the heritable trait variation and for traits with relatively simple genetic architectures. However, genetic control of the

majority of more complex and highly polygenic quality traits such as grain protein content and flour water absorption have been much more difficult to explain by single QTL genetic effects (Fradgley et al., 2022; Kristensen et al., 2018; Scott et al., 2021). For such traits, genomic selection (Meuwissen et al., 2001) is the preferred approach and has enabled great progress in accelerating genetic gain in modern crop breeding programs (Crossa et al., 2017; Voss-Fels et al., 2019). A wide diversity of genomic prediction models have been developed (Howard et al., 2022) so that testing and selection of the best model for specific trait architectures and breeding scenarios as well as design of the training panel are required for optimum implementation of genomic selection. The costs and benefits of implementing genomic selection must also be pragmatically assessed with regard to the logistics of current breeding pipelines. This is particularly true for making accurate comparisons between phenotypic selection using high throughput phenotyping methods (Robert et al., 2022) and genotyping costs for the number of breeding lines tested at different stages of a commercial breeding program. Genomic selection should be able to increase the accuracy of early-stage selection enough to allow faster recycling of material for crossing to initiate new breeding cycles, and hence accelerate genetic gain.

Here, we assembled a panel of historic high-quality UK wheat varieties combined with breeding lines from multiple recent cycles from the Deutsche Saatveredelung (DSV) UK breeding program to examine the potential application of novel QTL for marker-assisted selection. We also used the panel to test genomic selection to enhance genetic gain for high milling and baking quality varieties within the context of a commercial breeding program. To evaluate how progress in UK quality wheat breeding has historically been made and how selection could be enhanced in the future, we (i) investigated temporal trends in milling and baking quality traits among genotypes in the panel released over the period 1935 to 2020, alongside trends in marker-based kinship among historical cohorts of high-quality wheat lines; (ii) performed a genome wide association study to identify potentially new QTL associated with quality trait variation; (iii) examined trends in frequency of positive alleles of QTL identified over recent decades of breeding and analyzed examples of selection in the historic wheat pedigree to determine how identified QTL have been, or could be, exploited for marker-assisted selection in breeding; and (iv) tested the prediction accuracy of a diverse set of genomic selection models and compared them to the current approaches of phenotypic prediction of final loaf baking and milling quality traits based on early-stage predictive quality traits and trait-assisted genomic prediction that combined both genomic and phenotypic data.

Core Ideas

- Baking quality of bread wheat varieties in the United Kingdom has increased despite declining grain protein content.
- Trends in frequencies of QTL alleles reflect trends in changes in wheat quality traits.
- Genomic selection proved a cost-effective strategy for loaf baking quality traits compared to phenotypic prediction.

2 | MATERIALS AND METHODS

2.1 | Germplasm

A panel of 379 wheat genotypes was assembled to represent genetic diversity in high-quality commercial wheat varieties over recent decades and within a present-day UK and German winter wheat breeding program. These consisted of five groups: (i) 28 previously released commercial wheat varieties that have appeared as either National Association of British and Irish Millers (NABIM) group 1 or 2 milling quality varieties for bread making on the Agriculture and Horticulture Development Board (AHDB) recommended lists between 2004 and 2019 (https://ahdb.org.uk/knowledge-library/recommended-lists-archive); (ii) 45 older and historical released spring and winter wheat varieties that were part of the already available wheat diversity panel as outlined by Fradgley et al. (2019) and noted as high quality by either Belderok et al. (2000) or listed as "Excellent," "High," or "Good" quality on the Genetic Resources Information System for Wheat and Triticale (GRIS; http://wheatpedigree.net/); (iii) 10 additional high- and low-quality advanced breeding lines and released varieties that have recently been used for crossing within the DSV UK breeding program; (iv) 176 advanced breeding lines from six breeding cycles of the DSV UK breeding program; and (v) 131 advanced breeding lines from two cycles of the DSV breeding program in Germany. The names of the recent and historical released varieties used in the panel are given in Table S1.

2.2 | Field trials

Reflecting standard breeding assessment capacity and timeframes, two main field trials were conducted over two growing seasons (2019–2020 and 2020–2021; hereafter denoted year 1 and year 2), at the DSV UK breeding station in Wardington, Oxfordshire, UK (Latitude = 52.103326;

Longitude = −1.2827740). Plots were managed according to field standard practice including fungicides to control foliar diseases, herbicides to control weeds, and a total of 200 kg ha^{-1} of nitrogen fertilizer applied over three applications (50, 100, and 50 kg N ha^{-1}, respectively). A total of 223 genotypes were tested in both years, with an additional 21 genotypes tested in year 1 only (244 genotypes in total tested in year 1) and an additional 28 genotypes tested in year 2 only (251 genotypes in total tested in year 2). Two replicate plots were planted for 56 genotypes (year 1) and 49 genotypes (year 2) while the remaining genotypes were arranged in single unreplicated plots. Both trial years consisted of 300 small plots (1 m^2) in nested sub-blocks design where in year 1, there were 10 sub-blocks of 10 plots nested within three main blocks of 100 plots each. In year 2, there were two sub-blocks of 10 plots nested within each of 15 main blocks of 20 plots each. Trial design was established to closely replicate current testing conditions in a commercial breeding context. Trial plots were harvested with a plot combine, and grain samples were cleaned with an aspirator cleaner and used for further quality analysis.

In addition, data from quality assessments on advanced breeding lines from five trial years from 2017 to 2021 in the DSV UK breeding program were included. These included from 13 to 119 genotypes per year, of which two or three were control varieties common across all trials. These breeding program trials were also conducted at the DSV UK breeding station at Wardington. Across all trial datasets (both main trials and 5 years of breeding program trials), a total of 379 wheat genotypes were a median of seven, and minimum of two genotypes were in common between trials so that analyses across environments could be made.

2.3 | Phenotyping and grain quality assessment

Grain samples from all trial plots were assessed for a range of traits that are predictive of specific aspects of milling and baking quality in line with methods approved by the American Association of Cereal Chemists (AACC, 2000). These included grain protein content measured using near-infrared reflectance (NIR) (AACC method 39-25.01), as well as specific weight (AACC method 55-10.01), both measured on grain using a Foss, Infratec™ NOVA. Alpha amylase activity was assessed using the Hagberg falling number (HFN) test (AACC approved method 56–81.03). Grain morphology traits including mean grain length, width, area, and estimated thousand grain weight (TGW) were assessed on 300 to 400 grains per sample using a MARVIN seed analyzer (GTA Sensorik GmbH).

A subsample of 108 genotypes in year 1 and 96 genotypes in year 2 (15 in common between years), as well as 20 additional advanced breeding lines from other DSV UK breeding program trials, giving a total of 209 genotypes, were assessed for milling and loaf baking quality. As is common within the DSV UK breeding program, samples from each trial were taken from single field plots with no replicated genotypes within each year to maximize the number of tested genotypes for low throughput and expensive phenotyping and within field environmental effects were assumed to be negligible. Grain was milled using a Laboratory Buhler roller mill (AACC approved method 26-21.02), and extraction rate was determined as the percentage of standardized weight of white flour fraction yielded from the total weight of grain milled. Dough rheology traits were then assessed with the milled white flour using a Calibre, DoughLab 2500 following the AACC approved method 54–70.01 where 50 g of flour (adjusted for moisture content) was mixed under high speed for 10 min. The traits assessed by this method include adjusted water absorption required to reach a peak mixing torque of 6.5 N m, dough development time, stability, and softening.

Test loaves were then baked with the refined white flour in a highly controlled, industry standard testing lab using a small-scale version of the Chorley Wood Baking Process (Cauvain & Young, 2006). This baking method was used as it is the standard method used to assess wheat baking quality in the United Kingdom. The method involved baking two test loaves for each sample with 600 g of flour. Expected water absorption was determined by near infrared reflectance (NIR) analysis of flour samples. Other ingredients were included as (w/w) percentages of the flour weight: fluid shortening (1%), bakers' yeast (2.7%), sodium chloride salt (1.55%), dough improver containing protease and amylase enzymes to boost yeast activity and ease dough expansion during proofing, oxidizing agents and emulsifiers (2%), soya flour (1.5%), and calcium carbonate (0.22%). Dough samples were mixed in a high-speed rotary mixer with a work input of 15 Wh/kg, while controlling the temperature of the dough at 28 ˚C (± 0.5˚C) with a water cooler. Dough sub-samples of 894 g were then molded, placed in baking tins, and proved for 55 min at 43˚C and 75% humidity. Loaves were then baked at 235˚C for 25 min in a deck oven. Loaf bake height was recorded immediately after baking, and the oven spring was calculated as the difference between bake height and the height of the dough after proving. After leaving baked loaves to cool for 2 h or until the internal temperature reaches 28˚C, loaf volume was measured by three-dimensional laser topography scanning using AACC Method 10–14.01. The color of the sliced crumb was quantified by tristimulus color measurement using a Minolta CR310 including the L* (white vs dark) and b* (yellow vs blue) color space axis as well as calculating L*-b* as a derived value of overall color quality. Visual assessments



FIGURE 1 Examples of loaves baked with the Chorley Wood Baking Process for year 1 and year 2 experiments. Numbers indicate the overall bread score given to each loaf.

of loaf quality characteristics were also made independently by two professional bakers at the test lab of Allied Bakeries, Maidenhead, UK. Crumb structure was scored on a scale of 0 to 5 from "weak," "firm," or "doughy" to "resilient." Crumb texture was scored on a scale of 1 to 4 from "coarse" to "fine." A final overall bread quality score integrating all other loaf quality assessments was also given on a scale of 0 to 9 (Figure 1).

2.4 | Trials analysis

Best linear unbiased prediction (BLUP) predicted means were calculated for each genotype for each baking quality trait across all trials and within each of the two main trial years. BLUPs were calculated by fitting mixed models using the "lme4" package in R. For analyses within each main trial year, effects of blocking structure could be accounted for with Equation (1):

$$Y_{ik} = \mu + G_i + Bb_k + e_i, \quad (1)$$

where Y is a vector of observed values in the response trait, μ is the overall mean, G is the random effect of the ith genotype, Bb is the random effect of the kth sub-block nested with main block, and e is the residual error. For the meta-analysis across all trials, Equation (2) was used for predictive quality traits that had some replication within each environment:

$$Y_{ij} = \mu + G_i + E_j + GE_{ij} + e_{ij}, \quad (2)$$

where, in addition to the components of Equation (1), E is the random effect of the jth trial environment, and GE is the random genotype by environment interaction effect for the ith genotype and the jth trial environment. Blocking structure was not included in unreplicated breeding program trials so blocking effects could not be included in these analyses across all trials. For milling and baking test traits that were not replicated within each environment, a similar model was fitted but without the genotype by environment interaction effect.

Variance components were extracted from fitted mixed models for each trait, and the broad sense heritability (H^2) for each trait across all trials was estimated using Equation (3):

$$H^2 = \frac{\sigma^2{}_G}{\sigma^2{}_G + \frac{\sigma^2{}_{GE}}{nE} + \frac{\sigma^2{}_e}{np}} \quad (3)$$

where $\sigma^2{}_G$ is the genetic variance, $\sigma^2{}_{GE}$ is the genotype by environment variance, nE is the number of trial environments, $\sigma^2{}_e$ is the residual error variance, and np is the mean number of plots per genotype.

2.5 | Genetic markers

The panel was genotyped with a custom 25k single nucleotide polymorphism (SNP) array (provided by SGS—Institut Fresenius TraitGenetics) that was a combined subset of 14,455 markers from the Illumina 90k wheat SNP array (Wang et al., 2014), 2,949 markers from the Affymetrix Axiom 35k wheat breeders' array (Allen et al., 2017), 5,536 markers from another novel and unpublished 135k Affymetrix SNP array, as well as 265 trait specific markers from the literature. From the full set of markers, a subset of 18,493 were selected that had less than 10% missing data and minor allele frequency over 0.05. Remaining missing data were imputed using random forest with the "missForrest" package (Stekhoven & Bühlmann, 2012).

Physical positions on the Chinese Spring RefSeq v1.0 reference genome (IWGSC et al., 2018) for each marker were determined by a basic local alignment search tool (BLAST; Altschul et al., 1990). Correct chromosome assignments were confirmed by comparing linkage disequilibrium values of a marker against all other markers across all genotypes in the panel. For each marker, a set of linked markers ($R^2 > 0.5$) was found and the most common consensus chromosome assigned. If this consensus chromosome was different to that of the focal marker assigned chromosome, then it was corrected to the best BLAST hit on the linked markers consensus chromosome.

2.6 | Wheat panel genetic analysis and diversity trends

A linear kinship matrix for the panel of lines was calculated using the A.mat function in the rrBLUP R package (Endelman, 2011) for a subset of 4,970 SNP markers, that were the result of pruning the full set of markers by removing one of each pair of markers that correlated at $R^2 > 0.8$. Principal coordinate analysis (PCoA) was then conducted on this relationship matrix, and the genotype weightings on the first

two dimensions were plotted to visualize relationships among genotypes and groups of lines.

Trends in genetic diversity among released wheat varieties were evaluated between the release years 1935 and 2020. For this, the mean and median kinship coefficients were calculated for each year for genotypes with a year of release within a sliding window of 20 preceding years. Mean and median kinship coefficients were also calculated within and across each breeding cycle of DSV UK and German breeding lines.

2.7 | Genome-wide association mapping

Quantitative trait locus (QTL) mapping was performed by genome-wide association for each trait BLUP using the full set of 18,493 markers with the GWASpoly package (Rosyara et al., 2016) in R analysis software. Meta-analysis was performed for trait BLUPs across all trials as well as individually for each of the two main trials. The covariance matrix for polygenic effects was estimated using "the leave one chromosome out" method, and population structure was accounted for by including the first three principal components of the marker data as covariates. Thresholds for defining significant marker trait associations at the 90% confidence interval were estimated from 500 permutations of randomly sampling data for each trait. QTL intervals were defined for each QTL identified for each trait as the closest left and right flanking markers that are outside of those within the QTL interval (defined as on the same chromosome, in linkage disequilibrium with $R^2 > 0.5$ to the peak marker and with either a $-\log10(p)$ value within 2 $-\log10(p)$ of the peak marker or above the permutation-based significance threshold for that trait). Similarly to Fradgley et al. (2022), all significant marker trait associations above the threshold were grouped, where pairs of peak markers that had squared correlation coefficients (r^2) above 0.5 were considered as collocating QTL across all traits. Hierarchical clustering using the complete linkage method was then performed using a pairwise distance matrix where pairs of markers that were considered the same QTL had a value of 0 while pairs that were not had a value of 1. All associated markers were then assigned to QTL groups using the cutree function in R using 0.5 as the height threshold. For each trait, all QTL peak markers with the greatest significance per QTL group were fitted in a full multi-QTL model using the fit.QTL function in GWASpoly to test QTL effects by likelihood ratio tests under backward elimination and to determine partial R^2 effects and fitted $-\log10(p)$ values for each QTL peak marker. All fitted QTL peaks with a $-\log10(p)$ value below 3 were removed.

Gene content between left and right QTL flanking markers physical map positions was extracted using the Ensemble Biomart resource (Kinsella et al., 2011) for the IWGS Chinese Spring RefSeq v1.0 genome (IWGSC et al., 2018) for all QTL with intervals less than 10 Mb. Intervals were averaged for pleiotropic QTL found for multiple traits. Locations of likely candidate genes already described in the literature was compared using gene sequences available on the National Center for Biotechnology Information nucleotide database and Basic Local Alignment Search Tool (BLAST; Mount, 2007) search comparisons to the wheat RefSeq v1.0 genome (IWGSC et al., 2018).

2.8 | QTL historic selection trends

Trends over time in marker allele frequencies of QTL peak markers in released milling wheat varieties were examined to determine putative historical selection by breeders. Allele frequencies of the positive allele at each consensus QTL were calculated for released varieties with a 20 preceding year sliding window between 1970 and 2020. The trend was quantified by linear regression of the changes in allele frequency over time.

Specific selection effects by breeders over the historic pedigree were also examined by comparing QTL allele states of the released varieties with that of their parents. Pedigree data were used from the latest version of NIAB wheat pedigree resource (Fradgley 2021) as outlined by (Fradgley et al., 2019). For each QTL, a subset pedigree of simplex families was found where the QTL allele status for released milling quality varieties was different to one of its two parents. Even when marker information was only available for one of the parents, it can be deduced that the released variety was selected by a breeder from a family of segregating lines with a 50% probability of inheriting either QTL allele between polymorphic parents assuming Mendelian inheritance. To increase the number of simplex families that could be considered, parental genotypes were inferred if grandparent genotypes were known to be monomorphic in the pedigree.

2.9 | Genomic prediction models

A diverse set of genomic prediction models that represent contrasting approaches to genomic prediction and modeling various genetic architectures were tested for each quality trait measured across all trials. Cross validation was performed to determine prediction accuracy with five rounds of 10-fold random validation assignment of genotypes where for each fold, models were trained with data from a random sample of 90% of the lines and predictions tested for the remaining 10% of data. In addition, cross validation was performed using breeding lines from each cycle of the DSV UK breeding program as test sets with models trained on all remaining data. Prediction accuracy was estimated as the Pearson's correlation

coefficient (r) between the observed BLUP data and the predicted trait values across all cross-validation folds for each of the five rounds of random cross validation and for each breeding program cycle. The subset of 4970 pruned SNP markers were used for all prediction models. For fair comparison, predictions were made within the same cross validation assignments for all models tested. Six prediction models were tested as follows:

(i) The baseline genomic best linear unbiased estimate (GBLUP) prediction models were implemented using the rrblup package (Endelman, 2011) in R. This includes a linear additive genomic relationship matrix (G) defined using the A.mat function as described by (VanRaden, 2008) and Equation (4):

$$G = \frac{MM'}{p}, \tag{4}$$

where M is the n × m marker matrix and p is the number of markers.

(ii) An extension of the GBLUP (EG-BLUP) was implemented that aims to integrate epistatic marker effects (Jiang & Reif, 2015) with an additional relationship matrix (H) that is the Hadamard product of G with itself: $H = G \odot G$ ($\odot$ indicates the Hadamard product). EG-BLUP was implemented with Bayesian regression models in the BGLR package (Pérez & Campos, 2014) in R. Both G and H matrices were fitted as kernels with 5000 iterations of the Gibbs sampler of which 1000 were burn-in and a default thinning value of five.

(iii) Reproducing kernel Hilbert space (RKHS) regression models were also fitted which model the genomic relationship matrix using non-linear Gaussian kernels. The Gaussian kernel (K) was defined as the exponential relationship between the Euclidian distance among genotypes with Equation (5)

$$K = \exp\left(-hD^2_{ii'}/Q\right), \tag{5}$$

where $D^2_{ii'}$ is the squared Euclidean distance of the n × m marker matrix, Q is the mean distance, and h is the bandwidth parameter as outlined by (Costa-Neto & Fritsche-Neto, 2021; Montesinos-López et al., 2021). The kernel averaging approach was taken by fitting multiple Gaussian kernels with different h values where $h = 1/m \times \{1/5, 1, 5\}$ where m = the median of the squared Euclidean distance (de los Campos et al., 2010; Pérez & Campos, 2014). The same parameters of the Gibbs sampler were used as for EG-BLUP.

(iv) Least absolute shrinkage and selection operator (LASSO) models were used which assign linear additive effects to all SNP markers, but most SNP effects are shrunken to zero. Models were fitted within the glmnet package (Friedman et al., 2010) in R. Optimum values of lambda were determined for each model by eightfold cross validation of each training dataset for each model for 50 lambda values between 1×10^{-24} and 0.4096.

The following non-parametric ensemble and machine learning models were also used:

(v) A fast implementation of random forest (RF) models was fitted in the ranger package (Wright & Ziegler, 2017) in R. For each RF model, 500 regression trees were grown with unlimited interaction depth, but with a minimum leaf node size of 15, and one third of the predictor SNP variables were randomly sampled to be considered for each tree split.

(vi) Stochastic gradient boosted machines (GBM) models (Friedman, 2002) were implemented in the gbm package (Ridgeway, 2007) in R. Performance of these models vary greatly with different hyperparameters, so a forward prediction tuning step was used to determine the optimum hyperparameters to use for fitting each model. This involved setting up a large grid search space including tree interaction depth values of 2, 3, 4, 5, and 6; the bagging fraction of the training set randomly sampled at each step of 0.4, 0.5, 0.6, 0.7, and 0.8; and a sequence of five minimum leaf node size values between 2 and one twentieth of the number of individuals in the training set. Validation error was initially determined for a random sample of 15 components of the grid search space for GBM models with a single random 80:20 split of the training and validation factions. These models for hyperparameter optimization were run with only 50 iterations of regression trees and a large shrinkage value of 0.2 to minimize computation time in this tuning phase. The 15 values of validation error and the sets of three hyperparameters used to fit those models were then used as response and explanatory variables, respectively, in random forest models for forward prediction of the next best set of grid search hyperparameters to try with the lowest predicted validation error. This forward prediction and hyperparameter search iteration was repeated five times to ensure efficient and fast searching and testing of the large hyperparameter space. Once optimum estimated values of interaction depth, bagging fraction, and minimum leaf node size were determined, these parameters were used for larger and slower GBM models with 500 tree iterations and a lower shrinkage value of 0.03. The optimum tree iteration from which to make final predictions was determined by fivefold cross validation to minimize overfitting.

2.10 | Comparing phenotypic and trait-assisted genomic prediction

Current practice within the DSV UK breeding program uses phenotypic data for predictive quality traits measured on large numbers of lines at early generations to predict and make selections for overall baking quality measured at later stages. Therefore, the potential to predict later stage milling and

baking quality traits was assessed using either first-stage quality traits (including grain protein content, specific weight, and HFN) or second-stage quality traits (additionally including dough development time, stability and softening) in comparison to single- or multi-trait genomic prediction models as well as trait-assisted genomic prediction, which combines both genomic data and early-stage phenotypes.

Multiple linear regression was used for phenotypic prediction models where predictive quality traits from either first and/or second stages were fit as explanatory variables with each later stage milling and baking quality trait as response variables. Stepwise model simplification was implemented using the step function in R to remove non-significant ($p > 0.05$) explanatory variables and to select a fitted model with the lowest Akaike information criterion (AIC). These phenotypic predictions were compared with single- and multi-trait Bayesian linear regression genomic prediction models that were fitted with the BGLR package in R (Pérez & Campos, 2014) using a linear genomic relationship matrix as outlined above. Separate models for single-trait models while multi-trait models were fitted using the Multitrait function, as described by Pérez-Rodríguez and Campos (2022), where all trait data for test genotypes were masked as missing. In addition, we tested trait-assisted genomic prediction models; all trait data except the predictive quality traits for either first or second stage groups were masked for test genotypes. Models used 10,000 iterations of the Gibbs sampler, of which 1000 were burn-in.

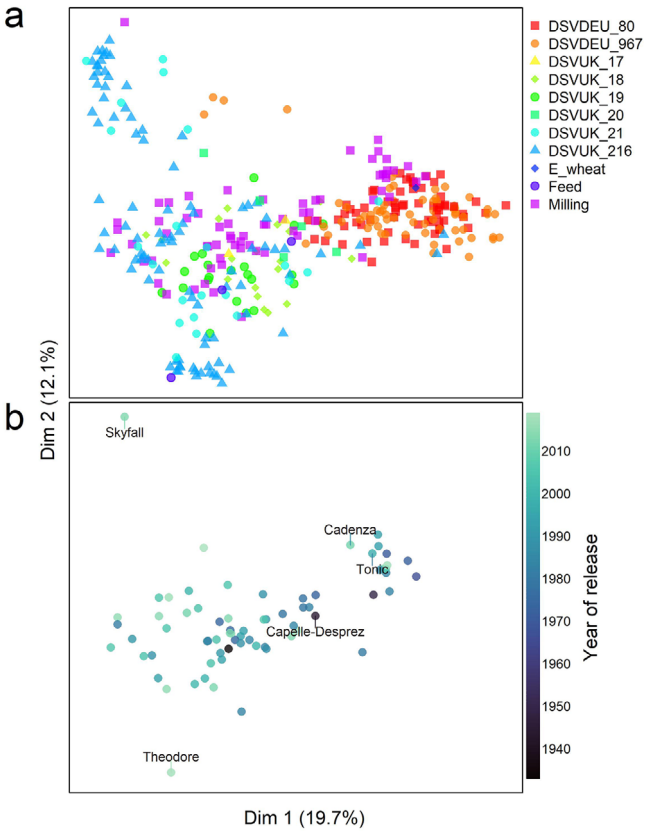


FIGURE 2 Principal coordinate analysis identifies population structure in the panel of wheat genotypes. The upper plot (a) shows differentiation of genotypes from the German and UK cycles of the Deutsche Saatveredelung (DSV) breeding program as well as released varieties. Details of genotypes in each group are shown in Table S1. DSVUK and DSVDEU groups indicate breeding lines form different cycles of the UK and German DSV breeding programs. E wheat varieties are a class of high-quality and high protein content varieties in Germany. The lower plot (b) shows the distribution of released varieties according to year of release.

3 | RESULTS

3.1 | Genetic structure reflects spatial and temporal trends in breeding for end-use quality

A panel of wheat varieties was assembled that represent major high-quality milling wheat varieties released in the United Kingdom over recent decades and material within the DSV UK and Germany wheat breeding program. Genotypic relationships were characterized using data from a SNP genotyping array, and PCoA revealed clear structure in the panel (Figure 2). The first dimension, which explained 19.7% of the variation, largely differentiated between the German and UK DSV breeding lines (Figure 2a). The released milling wheat varieties were distributed evenly across the first PCoA dimension, but the older of these varieties, such as Cappelle Desprez, or more recent spring type varieties, such as Cadenza and Tonic, were more central between the UK and German groups, reflecting their ancestral significance across European breeding programs. The second PCoA dimension explained 12.1% of the variation and contrasted between clusters of milling and feed wheat quality DSV breeding lines that are more closely related to the modern but genetically distinct milling variety Skyfall and the modern feed wheat variety Theodore, respectively (Figure 2b). Differentiation between classes of German breeding lines was unclear within the first two PCoA dimensions, but a few lines were clustered toward the group of recently developed UK breeding lines, suggesting more recent exchange of material between countries.

A comparison of the relative kinship relationships among varieties released over time, and among breeding lines within each cycle of the DSV breeding program, revealed trends in genetic diversity (Figure 3a). It also allowed visualization of the level of diversity that breeders utilize within a breeding program (Figure 3b). The decrease in average kinship among released milling wheat varieties, particularly between 1950 and 1970, represents an incease in genetic diversity in that period, which has been stably maintained (Figure 3a). In comparison, the mean kinship among breeding lines within each of the UK and German cycles of the DSV breeding

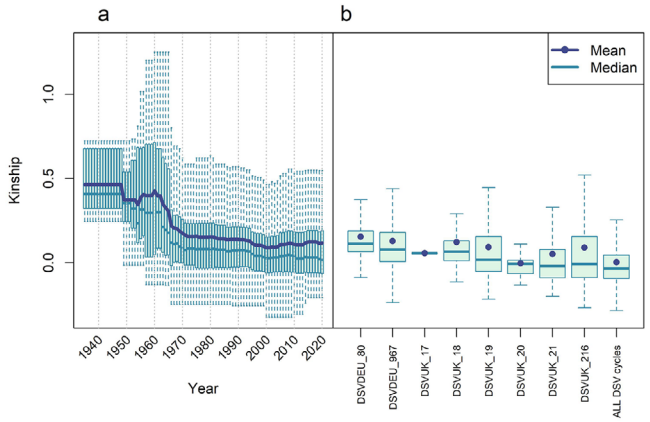


FIGURE 3 Trends in kinship indicate high and stable diversity among released milling wheat varieties over time (a), as well as within and among cycles of the UK and German DSV breeding program (b). Boxes indicate the upper and lower interquartile range and dashes indicate 1.5 × the inter quartile range.

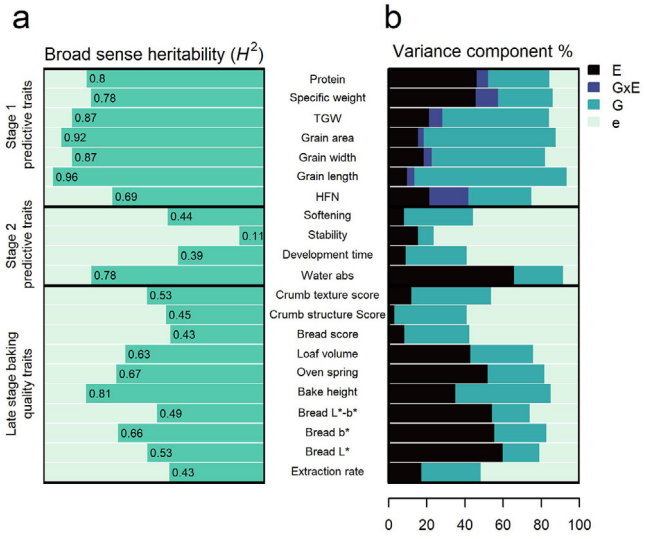


FIGURE 4 Comparison of trait heritability (a) and variance components (b) for grain and quality traits. E, environmental effects; G, genetic effects; G×E, genotype by environment interactions; e, residual error. G×E variance components could only be estimated for traits measured with replication within trials. HFN, Hagberg falling number; TGW, thousand grain weight; bread b*, bread yellowness reflectance; bread L*: bread whiteness reflectance.

program are low (Figure 3b). Furthermore, mean kinship among lines across all UK and German breeding program cycles was even lower, suggesting greater genetic diversity is available to breeders where material is exchanged and crossed among countries and breeding cycles.

3.2 | Phenotypic quality traits vary across years and breeding cycles

The panel was characterized for a comprehensive set of predictive quality traits that are measured at early-stages of the breeding program as well as more costly milling and loaf baking quality traits that are typically measured at later stages. These trials were also integrated with advanced breeding line trials as part of the DSV UK breeding program so that a meta-analysis of predicted trait values across several years of unbalanced trials was possible and performed within a commercial breeding context. Estimation of trait heritability across these environments revealed that grain dimension traits (length, width, area, and thousand grain weight) had the highest heritability ($H^2 > 0.87$) with only small environmental effects (Figure 4). Early-stage predictive quality traits, including grain protein content, specific weight, and HFN, were generally also highly heritable (Figure 4a), but with larger environmental effects, and HFN had a relatively large genotype by environment interaction (G×E) variance component across the seven trial year environments (Figure 4b). Dough rheology stability had very low heritability ($H^2 = 0.11$) due to a large residual error component, while the similar trait softening was greater ($H^2 = 0.45$). Among later stage milling and baking quality traits, loaf baking height and oven spring had the highest heritability. In contrast, the subjective loaf scores for overall baking quality showed only small environmental effects but had large error variance and therefore lower heritability. Bread color traits had much larger environmental effects but with only small error, and yellowness (bread b*) was more heritable than whiteness (bread L*) (Figure 4a).

3.3 | Temporal trend analysis indicates quality trait changes over time

Estimates of wheat quality traits were made for UK commercial milling quality varieties that were released over the last 85 years. Trends in these quality traits were quantified over this period of breeding by simple linear regression of the variety trait values against their year of release (Figure 5). The most prominent trend was a decrease in grain protein content despite its historical value as a trait for wheat bread quality. The data indicate that it has decreased by an average of 0.023% points per year so that a decrease of approximately 15% (from over 14.5% to just over 12.5% grain protein content) was observed over the period represented in this study (Figure 5). Despite this decrease in grain protein content, overall bread quality score has significantly increased (Figure 5b). This seems to be related to associated improvements and increases in loaf quality traits such as bread color (decreased bread b* and increased bread L*), larger loaf size (increased loaf volume, bake height, and oven spring), as well as better crumb texture and structure scores (Figure 5a). Other dough rheology traits including dough softening and stability also

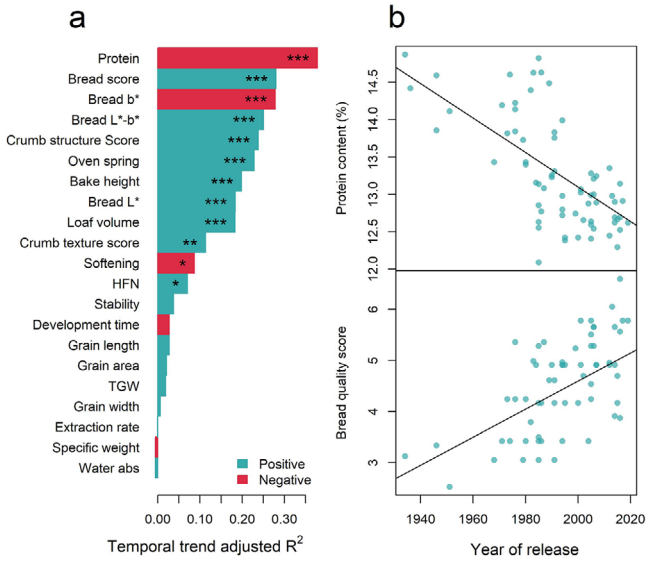


FIGURE 5 Temporal trends in wheat quality traits among released high-quality UK wheat varieties. HFN, Hagberg falling number; TGW, thousand grain weight. For each trait, R2 is shown in (a), while (b) shows the linear regression for grain protein content and bread quality score against variety year of release. *** indicates $p < 0.001$, ** indicates $p < 0.01$ and * indicates $p < 0.05$.

improved, but with a less clear trend. This generally suggests that the decrease in protein quantity has been compensated by increased gluten quality and has resulted in increased overall baking quality for the Chorley Wood baking process method tested here. HFN also showed a slight improvement, but other important milling quality traits such as flour water absorption, grain size, specific weight, and flour extraction rate show very little temporal trend.

3.4 | Meta-QTL analysis identifies known and novel genomic regions contributing to wheat quality traits

QTL mapping was performed by genome wide association for all traits in each of the two trial years and for trait values estimated across all environments by meta-analysis. Across all traits and analyses, 337 marker trait associations were found, which could be grouped into 24 independent QTL based on linkage disequilibrium. Multi-QTL models were fitted for each trait. Seven of the QTL were found to collocate for different traits (Table 1). QTL physical intervals and flanking markers were defined and gene content on the Chinese Spring RefSeq v1.0 genome was extracted. QTL intervals contained an average of 15 genes with a wide range of functions so identifying candidate causal genes was difficult (Table S2). The *QTgw.niab-5A* QTL spanned a particularly large interval over 98 Mb across the centromeric region of chromosome 5A. Gene lists from identified QTL may be further reviewed for targeting potential as candidate genes.

Some of the QTL could be related to previously characterized QTL in the literature based on physical map position. The largest effect QTL found for several baking loaf size and bread quality traits measured in year 1 and in the meta-analysis was mapped to a similar position as found by Fradgley et al. (2022) for dough rheological traits, and BLAST search for the gene sequence provided by Anderson et al. (1989) indicated that TraesCS1D02G317200 is the likely candidate gene in the interval. This QTL was, therefore, considered to distinguish between the positive effect 5+10 allele and other lower score alleles at the *Glu-1D* locus (Table 1). Similarly, a QTL on chromosome 1A was found to have an effect on bread color reflectance traits, was close to the homoeologous region to *Glu-1D* as also identified by (Fradgley et al., 2022), and was within 4 Mb of results from the BLAST search for gene sequence provided by De Bustos et al. (2000) and so was considered to probably be the *Glu-1A* locus (Table 1). Although *Glu-1A* would be expected to have an effect on gluten strength and loaf volume traits, the novel effect on bread color may be due to an associated measured color difference in enhanced bread crumb texture. The allele that significantly increased bread color yellowness (b*) in year 2, also decreased the crumb structure score from a mean of 3.8 to 2.6 and loaf volume by 8.2% in year 2, but these effects were not genome-wide significant.

Of the seven QTL found for HFN, two could be related to genes or QTL previously found in the literature. First, the QTL on the long arm of chromosome 4A with a peak marker at 605,664,305 bp was found to match the preharvest sprouting (*Phs-A1*) locus because it had close synteny with the *Triticum dicocoides* 4A region between 595.3 and 596.7 Mb as described by Shorinola et al. (2017). BLAST search for the sequence of the *TaMKK3-A* candidate gene provided by Torada et al. (2016) mapped to an unknown chromosome, but of the gene models in the defined interval, TraesCS4A02G315500 included "embryo development ending in seed dormancy" as a Gene ontology (GO) term. Second, the QTL on chromosome 7B (*QHfn.niab-7B.2*) at 750,605,912 bp was found to approximately collocate with a QTL at 750,082,927 bp that explained a relatively large proportion of the phenotypic variance in HFN by (White et al., 2022) and is in a similar region to QTL identified by Mohler et al. (2014) and a near significant QTL identified in Danish breeding lines by Kristensen et al. (2018). However, neither of these QTL was found to explain a large proportion of the trait variation in HFN across multiple environments ($R^2 < 0.04$). Furthermore, the data showed that *QHfn.niab-7B.2* also had a pleiotropic positive effect on dough softening. A single QTL was found for flour water absorption at the distal region of the short arm of chromosome 5D and is a likely colocation with the known *Ha* hardness locus where puroindoline

TABLE 1 Analysis identified QTL for all traits across all environments (Meta) and within each main trial year (Yr1 and Yr2)

QTL	Trait	Chrom.	−log10(*p*)	Effect	R^2
QHfn.niab-1A	HFN Meta	1A	4.00	19.258	0.04
Glu-1A	Bread b* Yr2	1A	4.24	1.508	0.16
Glu-1A	Bread L*-b* Yr2	1A	4.91	−4.063	0.18
QOsp.niab-1A	Oven spring Yr2	1A	5.48	−13.516	0.21
QGl.niab-1B	Grain area Meta	1B	4.65	1.251	0.07
QGl.niab-1B	Grain length Meta	1B	5.33	0.309	0.08
QGl.niab-1B	Grain length Yr1	1B	4.91	0.310	0.08
QGl.niab-1B	Grain length Yr2	1B	4.69	0.322	0.07
QHfn.niab-1D	HFN Meta	1D	3.58	−23.379	0.04
QSpw.niab-1D	Specific weight Yr1	1D	5.01	−1.032	0.08
Glu-1D	Bake height Meta	1D	8.30	−7.825	0.15
Glu-1D	Oven spring Meta	1D	7.34	−5.399	0.14
Glu-1D	Bread score Meta	1D	6.88	−0.535	0.13
Glu-1D	Crumb structure Score Meta	1D	6.31	−0.365	0.12
Glu-1D	Bake height Yr1	1D	10.38	−15.870	0.33
Glu-1D	Oven spring Yr1	1D	11.03	−15.131	0.35
Glu-1D	Bread score Yr1	1D	9.44	−2.398	0.31
Glu-1D	Crumb structure Score Yr1	1D	7.32	−1.632	0.24
QGpc.niab-2A	Protein Yr2	2A	6.31	0.435	0.10
QBcl.niab-2B.1	Bread b* Yr1	2B	5.50	0.674	0.18
QBcl.niab-2B.2	Bread L* Yr1	2B	5.06	−3.202	0.17
QBcl.niab-2B.2	Bread L*-b* Yr1	2B	5.02	−3.960	0.17
QBcl.niab-2B.2	Bread L Meta	2B	4.23	−0.789	0.08
QBctx.niab-2B	Crumb texture score Yr2	2B	5.21	−2.481	0.20
QBsc.niab-2D	Bread score Meta	2D	6.11	−0.535	0.11
QBsc.niab-2D	Crumb texture score Meta	2D	7.87	−0.368	0.15
QSpw.niab-3A	Specific weight Yr2	3A	5.22	1.245	0.08
QBcl.niab-3D	Bread b* Yr1	3D	5.79	−1.166	0.19
Phs-A1	HFN Meta	4A	3.52	15.447	0.04
Phs-A1	HFN Yr1	4A	4.80	−49.083	0.07
QTgw.niab-5A	TGW Meta	5A	6.12	2.344	0.09
QTgw.niab-5A	Grain area Meta	5A	5.89	0.764	0.08
QTgw.niab-5A	Grain length Meta	5A	5.97	0.178	0.09
QTgw.niab-5A	TGW Yr1	5A	5.53	2.502	0.09
QTgw.niab-5A	Grain area Yr1	5A	4.03	1.465	0.06
QTgw.niab-5A	Grain area Yr2	5A	6.68	0.758	0.11
QTgw.niab-5A	Length Yr2	5A	6.54	0.193	0.10
Pinb-D1	Water abs Meta	5D	6.54	−1.668	0.14
Pinb-D1	Water abs Yr2	5D	5.86	−2.822	0.22
QSof.niab-6B	Softening Meta	6B	4.92	−7.299	0.10
QHfn.niab-6D	HFN Yr2	6D	5.66	29.151	0.09
QGrw.niab-7A	Grain width Meta	7A	5.24	−0.064	0.07
QHfn.niab-7B.1	HFN Meta	7B	3.41	−23.252	0.03
QHfn.niab-7B.2	Softening Yr1	7B	7.31	56.067	0.31
QHfn.niab-7B.2	HFN Meta	7B	3.18	−28.770	0.03
QExr.niab-7D	Extraction rate Yr1	7D	4.95	1.349	0.16
QHfn.niab-Un	HFN Meta	Unmapped	3.33	−17.254	0.03

Note: R^2 and −log10(*p*) values were derived from multi-QTL models for each trait.

Abbreviations: HFN, Hagberg falling number; QTL, quantitative trait loci; TGW, thousand grain weight.

protein genes are located (Morris, 2002) and matched a BLAST search from sequence from Ayala et al. (2016).

Some potentially novel QTL that could not be related to others in the literature were found for predictive quality traits. Grain size traits (grain length, width, area, and TGW) had high heritability ($H^2 > 0.87$), and two QTL on the long arm of chromosome 1B (*QGl.niab-1B*) and over a wide interval of the centromeric region of 5A (*QTgw.niab-5A*) were consistently found in both main trial analyses and meta-analysis. Despite relatively high heritability of grain protein content and specific weight ($H^2 = 0.80$ and 0.78, respectively), only single QTL were identified for each of these traits on chromosomes 2A (*QGpc.niab-2A*) and 3A (*QSpw.niab-3A*), respectively, and these explained a small proportion of the phenotypic variation ($R^2 < 0.1$) in only one of the main trial environments. HFN was found to be of lower heritability ($H^2 = 0.69$), but several QTL were found that explained a small proportion of the trait variation ($0.03 < R^2 < 0.09$). A particularly large G×E variance component was found for HFN, and none of the HFN QTL were identified in both main trial years.

Several further potentially novel QTL were also identified for loaf baking quality traits. In particular, two QTL were identified on chromosome 2B (*QBcl.niab-2B.1* and *QBcl.niab-2B.2*), and one on 3D (*QBcl.niab-3D*) for loaf color traits measured by color reflectance, all explaining large proportions of the trait variation in year 1 main trials ($R^2 > 0.17$). However, none of these were identified in year 2 or the meta-analysis. A potentially novel QTL on chromosome 2D (*QBsc.niab-2D*) for overall bread score and crumb texture score was identified for the all-trials meta-analysis. Despite relatively low heritability of flour extraction rate ($H^2 = 0.43$), a QTL on chromosome 7D (*QExr.niab-7D*) was found for this trait in year 1 trials only, explaining a reasonably high proportion of the phenotypic variation ($R^2 = 0.16$).

3.5 | Changing allele frequencies provide evidence for historic selection of specific QTL

The temporal trends in allele frequency of beneficial peak marker alleles identified by the QTL analysis among the panel of released varieties were examined using a 20-preceding years sliding window moving average. This considered examples of varieties in the historic pedigree where breeders had selected the variety from a cross between polymorphic parents. Considering groups of QTL based on their effect on different wheat quality traits, the observed trends and shifts in allele frequency largely reflected and explained many of the changes found for different quality traits. The positive allele for *QGpc.niab-2A* was found to be common among varieties before 1970, such as Bersee and Highbury, but rapidly declined in frequency among varieties released since then (Figure S1a). All four examples of milling wheat varieties that were found to be selected from parents that were polymorphic for *QGpc.niab-2A* (Malacca, Maris-Widgeon, Caxton and Chilton) inherited the low grain protein content allele suggesting that breeders have been actively selecting against this QTL.

The compensatory increase in loaf baking quality traits despite decreased protein content was explained by the steady increase in frequency of positive alleles for four QTL found here affecting loaf baking quality and crumb color traits (Figure S1b,c). Only one of these QTL (*QOsp.niab-1A*) had a sufficient number of examples of varieties with known polymorphic parents to test pedigree selection effects. This analysis found that six of the seven varieties selected from polymorphic parents (Spark, Solitaire, Wembley, Ketchum, Gallant, and Kingdom) inherited the positive *QOsp.niab-1A* allele, while only Xi-19 did not. This suggests that breeders were actively selecting positive genetic effects within families of segregating breeding lines. While the QTL for loaf quality traits are almost all fixed for the positive allele among the most recently released milling wheat varieties, they generally remain polymorphic among breeding lines within the DSV UK breeding program, which include both high milling quality and low-quality feed wheat varieties (Figure S1b). QTL for loaf color traits (such as *QBcl.niab-3D*), show an increasing trend but are still at lower frequency in released varieties and among DSV breeding lines.

In contrast, the overall frequencies of positive alleles for identified QTL for other traits, such as HFN, grain size traits, and specific weight, generally remained neutral (Figure S1a). Although the positive allele for *QSpw.niab-1D* seemed to increase in frequency, *QSpw.niab-3A* showed the opposite trend and this also reflects the lack of temporal trends in these traits (Figure 5a). Taken together, these results show evidence of selection in line with changes in wheat quality traits over time and support potential use of QTL effects identified from GWAS.

3.6 | Genomic prediction accuracy increases with model optimization based on trait genetic architecture

To determine the suitability of using genomic selection for wheat milling and baking quality traits, a diverse range of genomic prediction models whose assumptions encompass several different trait genetic architectures were tested. Under a random 10-fold cross validation scheme, generally high prediction accuracies were achieved considering the theoretical maximum accuracy based on the estimated heritability of each trait (Figure 6). More than half of the heritable genetic variation could be predicted for traits such as grain protein content, specific weight, or HFN, as well as some of the loaf quality traits including overall bread score and baking height. For

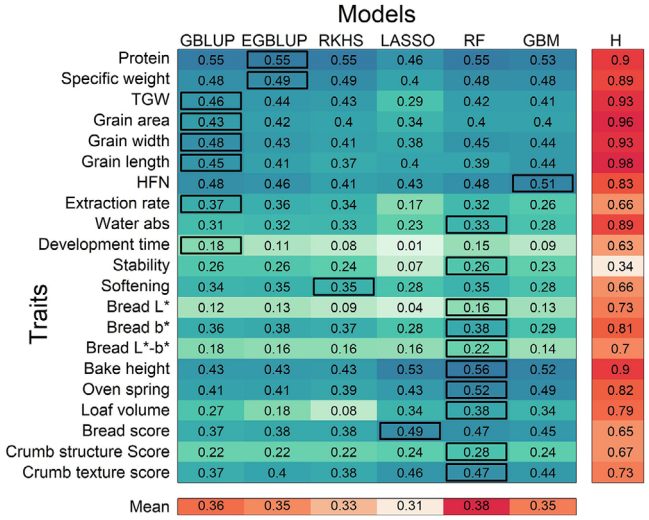


FIGURE 6 Model testing identified variation in mean prediction accuracy for all traits from random cross validation. Prediction accuracy is calculated as the Pearson correlation between the observed and predicted trait values. Black boxes indicate the most accurate model for each trait. The theoretical maximum prediction accuracy derived from the square root of the trait broad sense heritability (H), indicating the theoretical maximum prediction accuracy, is shown to the right. TGW, thousand grain weight; HFN, Hagberg falling number.

traits including flour water absorption and bread L*, a much smaller proportion of the heritable variation could be predicted. Overall, much greater potential to predict heritable trait variation could be demonstrated from genomic prediction than the combination of all identified QTL. This was particularly so for traits such as protein content or specific weight for which QTL were only found in single environments and explained less than 10% of the phenotypic variation, whereas cross validation of genomic prediction models showed that over half of the heritable phenotypic variation could be successfully predicted. This suggests that genomic prediction models can successfully integrate a much larger number of smaller effect QTL to predict highly polygenic traits.

There was no single prediction model that performed best for all traits (Figure 6). However, as expected, the differences in model prediction accuracy appear to reflect the underlying genetic architecture of each trait. For example, LASSO models performed poorly in comparison to models based on estimates of kinship (GBLUP, EG-BLUP and RKHS) for traits such as grain protein content, specific weight, and TGW suggesting that these are highly polygenic traits with many genetic effects distributed across the genome. Despite relatively high heritability, only a few QTL of small effect size, with low R^2 values, were detected for these traits (Table 1). LASSO was the most accurate model for overall bread score and performed relatively well for other related traits such as loaf volume and crumb texture score. Together with the large R^2 values for QTL identified for these traits, this suggests that a larger fraction of the variation is explained by a few large-effect QTLs.

There was rarely an advantage of the more complex models that aim to account for epistatic genetic effects (EG-BLUP and RKHS) over the baseline GBLUP models that include linear genomic relationship matrices, suggesting a minimal role of epistasis for most traits in this dataset. However, the machine learning based models (RF and GBM) were able to flexibly integrate and predict non-linear and non-additive effects, and these performed competitively for the range of traits. RF was often the most accurate model (for 10 out of 21 traits) or comparable with the most accurate model, especially for the loaf quality traits such as overall bread score or bake height. RF and GBM models share similar model structure, but RF performed better than GBM for all traits except HFN when GBM was the overall most accurate model.

Practically, within a breeding program, genomic selection is used for forward prediction from past to future breeding cycles and genotypes become more distantly related to the training panel relative to random k-fold cross-validation across cohorts. Therefore, the accuracy from cross validation of genomic prediction models were also tested where lines from each cycle of the DSV UK and German breeding program were used as the test fraction and models were trained using all other data. Prediction accuracies were much more variable among breeding cycle cross validations than for multiple comparisons of random cross validation (Figures 6 and 7 and Figure S2). This was likely due to the smaller number of lines within some breeding cycles making accurate estimation of prediction accuracy difficult. While there was large variation in prediction accuracies between cycles, they were generally lower on average than for random cross validation for several traits. Traits such as grain protein content and HFN were predicted only slightly less accurately compared to random cross validation, whereas specific weight and grain size traits prediction accuracy was much lower for cross validation among breeding cycles.

Across all breeding cycles, trends among model accuracies generally reflected those from random cross validation. For example, LASSO models performed poorly for traits such as grain protein content, specific weight, and water absorption, but were more accurate for loaf baking quality traits such as overall bread score and crumb texture score. Similar to random cross validation, specific weight was better predicted using kinship-based prediction models (GBLUP, EG-BLUP, and RKHS), and HFN was best predicted using GBM. Trends in accuracy between different breeding cycles varied among traits. For example, lines from the DSVUK.18 breeding cycle were accurately predicted for most loaf baking quality traits, such as baking height and loaf volume, but poorly for grain size traits, such as TGW and grain length and width. Also, genotypes from UK and German breeding cycles were generally more genetically distinct (Figure 2), and

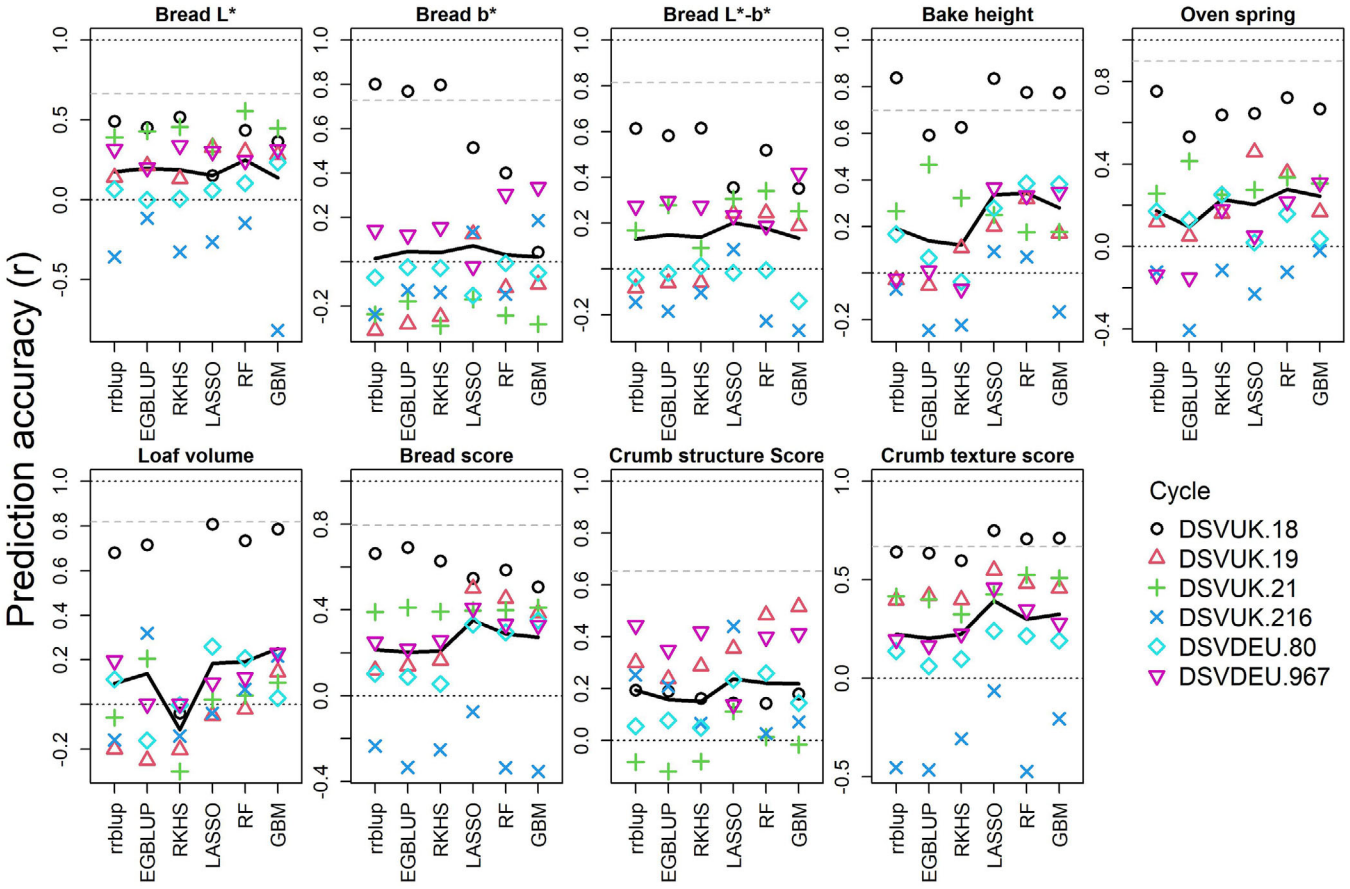


FIGURE 7 Prediction accuracy of different models for a subset of predictive and baking quality traits from cross validation for groups of lines from each DSV breeding program cycle. Pots for all traits are shown in Figure S2. Prediction accuracy was tested on lines from each cycle of the Deutsche Saatveredelung (DSV) UK and German breeding program (i.e., DSVUK.18) from models trained on all other data. Prediction accuracy is calculated as the Pearson correlation between the observed and predicted trait values. Solid black lines indicate the mean accuracy across all breeding cycles. Horizontal grey dashed lines indicate the theoretical maximum prediction accuracy derived from the square root of the trait heritability.

prediction accuracy of both German cycles was lower for grain protein content when the models were trained on the remaining, mostly UK origin, genotypes. However, this trend was less evident for other traits.

3.7 | Genomic versus phenotypic selection for high value, late-stage quality traits

We considered the practical implementation of genomic prediction in a commercial context where early-stage predictive quality traits are used to predict and make selections for later stage loaf baking quality traits. To do this, the prediction accuracy of genomic prediction models and phenotypic prediction models based on early-stage predictive quality traits was compared. To be effectively used at earlier stages of the breeding cycle, genomic selection must, therefore, add significant predictive ability for the more costly later stage milling and loaf baking quality traits in addition to the predictive phenotypic data that are already available at this stage.

Considering correlations between individual predictive traits at the first stage (protein content, specific weight, HFN, and grain size traits) and second stage (additionally water absorption, development time, stability, and softening) of the breeding program, these traits were partially predictive of later stage loaf baking quality (Figure S3). Although correlations were generally low ($r < 0.43$), directions were as expected. Specific weight had the strongest correlation with flour extraction rate ($r = 0.27$), compared to grain size traits that were much lower ($r < 0.09$). Both grain protein content and grain width, rather than length, correlated positively with flour water absorption ($r > 0.25$). Specific weight also had clear correlations with other baking quality traits, particularly for bread whiteness L* ($r = 0.26$), while correlations with grain protein content were surprisingly low considering its importance as a basic measure of quality. Protein content was even negatively associated with some bread color traits where higher protein lines often had lower bread L* values ($r = -0.13$). The correlation between grain protein content and overall bread score was near zero, but this relationship

was found to be non-linear where the optimum protein content was estimated to be just over 13%, while either excessively high or low protein content lines generally had poorer bread score (Figure S4). Although there were many lines with protein content around 13% and very low overall bread score, the top 10% of lines with the highest bread score all had protein contents within a narrow range of 12.36% and 13.52%. Therefore, grain protein content is a poor predictor of baking quality alone, but must be around 13% for lines to be of high potential baking quality. HFN was also positively but weakly associated with many later stage baking quality traits (Figure S4).

Addition of second stage predictive traits for gluten quality showed that only dough stability and softening had strong associations with loaf baking quality traits (Figure S4). Softening had the strongest association with all loaf baking quality traits except bread b* and crumb color score suggesting good predictive potential for use in selection. Flour water absorption and development time had much weaker associations with baking quality traits. Overall, these results suggest that different predictive quality traits provide information relating to independent aspects of loaf baking quality so forward selection based on multiple traits would be required in a breeding program.

Considering the relationships among predictive quality traits and later stage milling and loaf baking quality traits, the predictive accuracy of multi-trait phenotypic, single- and multi-trait genomic, and trait-assisted genomic prediction approaches were compared. Phenotypic prediction models were fitted using simple multiple regression with either sets of first or second stage predictive quality traits as explanatory variables. These were compared to single-trait genomic selection using Bayesian mixed models as well as multi-trait models for completely unphenotyped test lines. Furthermore, a trait-assisted selection scenario was investigated where multi-trait models were fitted and data for predictive quality traits were not masked from the test lines so that both genomic and trait data were used to predict test line phenotypes. Only the Bayesian mixed models with linear genomic relationship matrices were used here for comparisons between single and multi-trait genomic prediction models.

Simple, multiple linear regression models were found to be the best approach for combining a small number of predictive traits but phenotypic selection with only the first stage predictive quality traits was generally inaccurate (Figure 8). Inclusion of the second stage dough rheology predictive traits increased accuracy of phenotypic selection considerably for all traits except flour extraction rate which was largely predicted by the strong association with grain specific weight. Genomic prediction using the single-trait GBLUP model outperformed phenotypic selection with first stage predictive traits for all late stage milling and baking quality traits except for bread L* and bread L*−b* (Figure 8). Second stage phenotypic selection was more accurate than the single-trait GBLUP for most loaf baking traits (Figure 8), but the best single-trait genomic prediction models compared in Figure 6 were still generally more accurate than phenotypic selection for these traits, with the exception of bread L* and L*−b*.

These results suggest that similar accuracy of selection can be achieved from genomic selection as for phenotyping advanced second stage predictive quality traits. Multi-trait genomic prediction models that consider covariance among traits did not significantly improve accuracy over single-trait genomic prediction models for any trait when predicting completely unphenotyped lines. However, when using a trait-assisted genomic prediction scenario where a breeder has genetic marker data for breeding lines that have already been phenotyped for either first or second stage predictive quality traits, multi-trait models generally did have increased accuracy over single-trait models. Similar to trends in phenotypic selection, the value of including first rather than second stage predictive traits in trait-assisted genomic prediction models was greater for flour extraction rate, but for most loaf baking quality traits, there was little advantage in accuracy until second stage dough rheology traits are also measured on the test lines.

Although combining all stages of phenotyping with genomic prediction did increase prediction accuracy, this advantage may not be large enough to justify the costs of both phenotyping and genotyping. Considering the estimated genotyping and phenotyping costs involved in implementing phenotypic, genomic, and trait-assisted genomic selection (Figure 9a), genomic selection offers the greatest value per cost prediction accuracy of overall bread score to enable efficient selection at only a slightly greater cost than for first stage predictive trait phenotyping (Figure 9b). Implementing trait-assisted genomic selection where combined costs of both genotyping and phenotyping are much higher only offers a relatively small increase in accuracy. These results indicate strong potential for genomic prediction to be applied in a breeding program instead of costly second stage predictive quality trait phenotyping where the best suited single-trait genomic prediction model for each trait can be applied with great effect.

4 | DISCUSSION

Novel QTL identified here may be useful for early-stage marker assisted selection within segregating families in breeding programs. While evaluation of wheat varieties' suitability for baking quality has traditionally been based on several predictive quality tests, data show that direct genomic prediction of advanced baking quality traits holds great potential for accurate genomic selection compared to selection based on additive QTL detected through GWAS or phenotypic selection based on early-stage predictive quality traits.

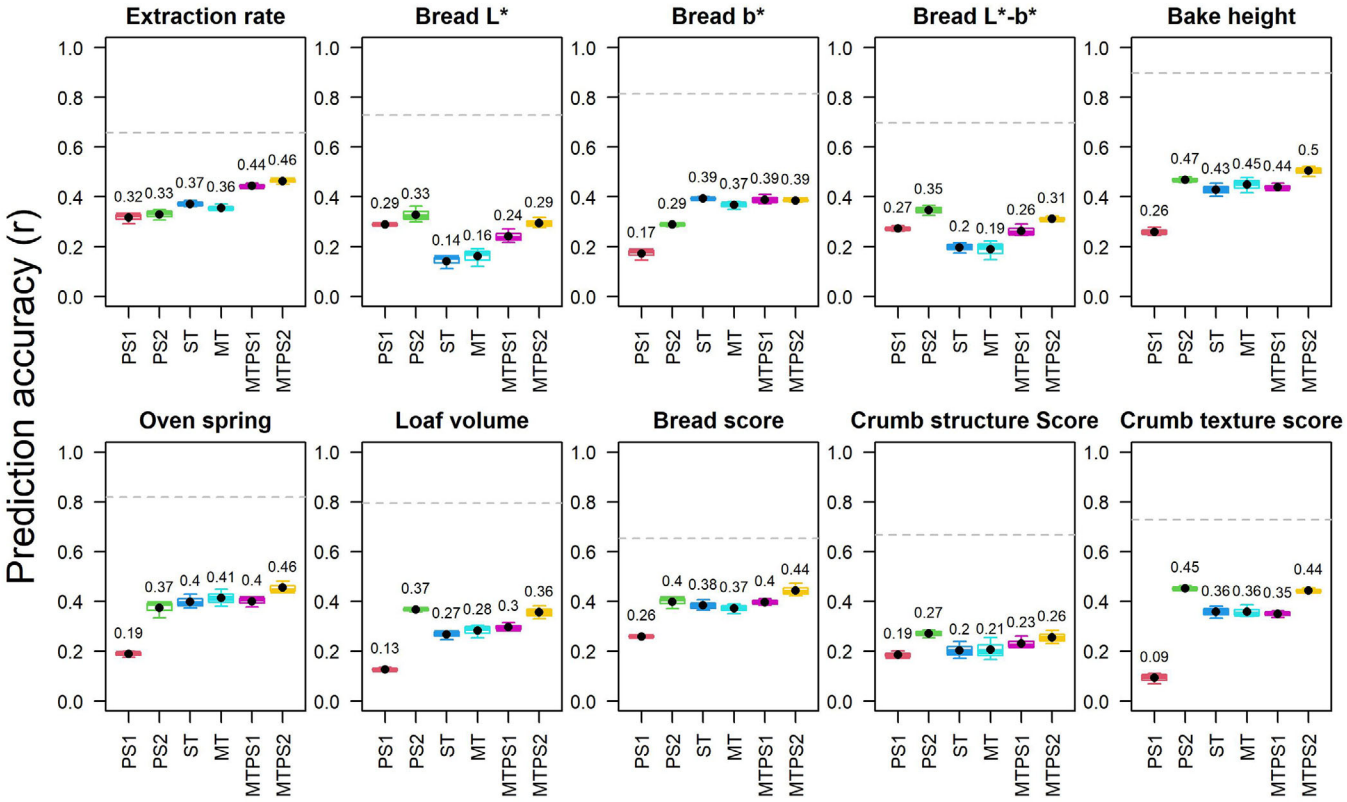


FIGURE 8 Prediction accuracy varies when comparing phenotypic as well as single- and multi-trait genomic prediction models. PS1, phenotypic selection using stage 1 predictive traits (grain protein content, specific weight, and Hagberg falling number); PS2, phenotypic selection using stage 2 predictive traits (PS1 traits as well as DoughLab development time, stability, and softening); ST, single-trait genomic selection; MT, multi-trait genomic selection for completely unphenotyped test lines; MTPS1, multi-trait-assisted genomic selection with PS1 traits included for test lines; MTPS2, genomic selection with PS2 traits included as covariates. Boxplots show the distribution of prediction accuracies from each round of 10-fold cross validation and black points indicate the mean. Horizontal dashed lines indicate the theoretical maximum prediction accuracy derived from the square root of the trait heritability.

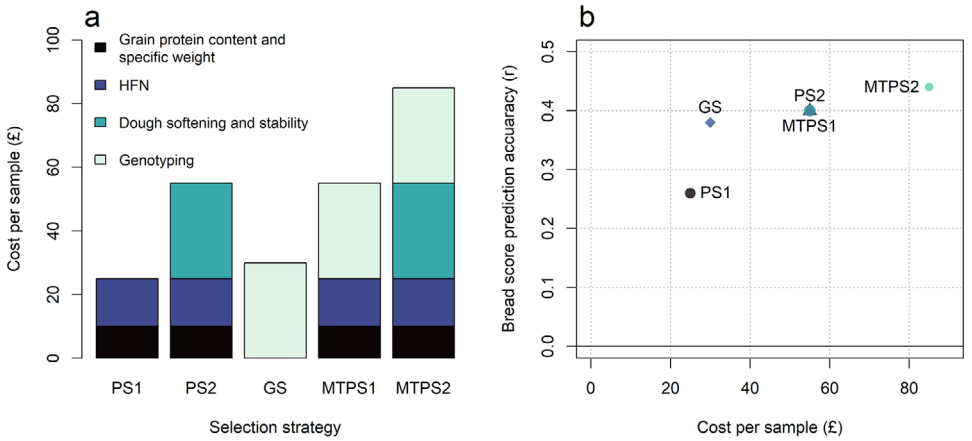


FIGURE 9 Cost per unit of accuracy varies between selection strategies: (a) breakdown of estimated costs among different phenotypic and genomic selection strategies and (b) comparison of selection strategy costs with the prediction accuracy of overall bread score. PS1, phenotypic selection using stage 1 predictive traits (grain protein content, specific weight, and Hagberg falling number); PS2, phenotypic selection using stage 2 predictive traits (PS1 traits as well as DoughLab development time, stability, and softening); GS, genomic selection; MTPS1, multi-trait-assisted genomic selection with PS1 traits included for test lines; MTPS2, genomic selection with PS2 traits included as covariates.

4.1 | Trends in UK quality wheat breeding

Genetic analysis showed a decrease in mean kinship among released varieties, most notably until 1970, indicating that the diversity of wheat varieties released by breeders for quality has increased. There has been no subsequent net loss of genetic diversity due to breeders' selection, despite significant genetic gain for both yield and quality traits being achieved during this time. This concurs with other studies that found little decline in diversity over recent decades of UK (Donini et al., 2000; Srinivasan et al., 2003; White et al., 2008) and German (Reif et al., 2005) wheat breeding. Concerted pre-breeding efforts to introgress genetic diversity from the tertiary genepool (Leigh et al., 2022), as well as historic exchange of breeding material among countries and regions (Fradgley et al., 2019) are both likely to have countered loss of genetic diversity due to continued breeders' selections and fixation of positive genetic effects controlling specific traits.

Results from analysis of kinship among wheat genotypes suggest that significant trends in genetic gain in quality traits have been achieved without net loss of genetic diversity. More modern milling wheat varieties in the United Kingdom generally have lower grain protein content compared to older varieties despite high grain protein being a desirable quality trait for selection. This aligns with other studies which found decreasing trends in protein content in line with the negative genetic trade-off between yield and protein content (Scott et al., 2021; Shewry et al., 2016) and underlines the focus that breeding has had on increasing yield. However, the clear increase in loaf baking quality scores indicate that the reductions in protein content have been more than offset by increased gluten protein quality. Whereas other studies have found that grain protein content is well correlated with loaf volume (Groos et al., 2007; Ibba et al., 2020), the lack of a strong association presented here is likely due to the Chorley Wood baking process method used here. At the time of its development, this process reduced the protein quantity requirements of UK grown wheat (Cauvain & Young, 2006). We propose that selection for increased gluten quality in combination with changes in the industrial bread baking process have enabled the large increases in yield of wheat achieved through breeding in recent decades (Calderini & Slafer, 1998; Mackay et al., 2011). Similar trends in decreased protein content have been observed from studies in Germany (Laidig et al., 2017), Spain (Sanchez-Garcia et al., 2015), and the United States (Fufa et al., 2005). However, analysis of the International Maize and Wheat Improvement Center (CIMMYT) wheat breeding program in Mexico, that focuses on a broader range of wheat end-uses, suggest that increases in yield and functional quality have been possible without decreased protein content (Guzmán et al., 2017). CIMMYT has also taken the similar approach of specifically selecting for positive high- and low-molecular-weight glutenin gene alleles within crosses at early-stages (Guzman et al., 2016). However, several other side-effects of grain protein levels, both positive and negative, are important to consider in wheat breeding. Maintaining high grain protein content is important to prevent the dilution of wheat proteins such as albumins and globulins which are not functional for gluten strength but have valuable amino acid profiles (Lasztity, 2017), as well as associated mineral micronutrient density (Tabbit et al., 2017). Development of wheat varieties that have good loaf baking quality at low protein levels may also help reduce nitrogen fertilizer application requirements for milling wheat crops (Shewry et al., 2020), which is an increasingly important requirement for mitigating environmental costs and political uncertainties of food production (Bentley, 2022). Selecting for wheat varieties that optimize trade-offs between yield, protein quantity, quality, and nutritional value will require a clearer understanding of the genetic control and prediction of specific wheat milling and baking quality traits.

4.2 | Selection history of wheat quality QTL

Trends in QTL frequencies identified here for protein content and loaf quality traits reflect trends in these quality traits over time and suggest how historic progress has been made, as well as identifying loci which still have potential for future selection of positive alleles. The positive allele for the protein content QTL on chromosome 2A had been negatively selected and remains at low frequency in current breeding material so is likely involved in the trade-off between yield and protein. It would therefore be a poor candidate for marker assisted selection. In contrast to the protein content QTL, the four QTL for loaf baking quality have been under strong positive selection and are all almost fixed in recent high quality milling wheat varieties but are more frequently polymorphic among the recent cycles of DSV UK breeding lines. While some of these QTL, such as *Glu-1D*, are already well characterized and widely used in breeding programs, the novel QTL identified here, such as for oven spring and crumb texture, will be more easily applied in breeding, particularly for crosses between milling quality and higher yielding feed wheats, as practiced in the DSV breeding program.

Other important quality traits such as HFN, specific weight, and extraction rate were not found to have increased in more recent varieties as much as loaf baking quality traits, but some novel QTL for these traits were identified here. In particular, two QTL for HFN were likely co-located with the *Phs-A1* gene on chromosome 4A that has recently been fine mapped by Shorinola et al. (2017) as well as a QTL on 7A, which is likely concurrent with QTL identified by White et al. (2022) and Mohler et al. (2014). Although both of these QTL were

only found in single trial years, there was a relatively large estimated G×E variance component of HFN. Therefore, use of multiple markers for environment specific QTL will be particularly important to select for increased stability of HFN in future years with unpredictable G×E effects rather than finding a QTL that has a consistent effect in all environments.

4.3 | Genomic prediction offers improved selection over marker-assisted and phenotypic selection based on predictive traits

QTL identified here will be useful for marker-assisted selection using single markers to screen segregating families of breeding lines as early as the F2 generations to fix positive alleles. This works well for large effect QTL that explain a significant proportion of the trait variation. For complex traits such as grain protein content and specific weight, using multiple detected QTL can only account for a relatively small proportion of the trait variation and so making predictions of trait values is difficult. Genomic prediction models tested here made more accurate predictions explaining more than half of the heritable trait variation for most traits when using random cross validation. Cross validated prediction accuracy among cycles of breeding lines was lower but still demonstrated similar trends among prediction model performance and highlights the suitability of genomic selection in an applied breeding context.

Testing diverse genomic prediction models showed that no single model was the most accurate for all traits, thus highlighting the importance of testing multiple approaches for realistic breeding prediction scenarios and traits with unknown genetic architecture. In general, prediction accuracies were as expected for traits such as grain protein content and thousand grain weight, but lower for some traits such as specific weight, in comparison to other studies (Battenfield et al., 2016; Sandhu et al., 2021, 2022), but more similarly to Lado et al. (2018) which may be due to the lower number of genotypes in the panel analyzed here. Heritability of dough rheology traits including softening and stability was lower than expected in comparison to a previous study using the same method (Fradgley et al., 2022), and prediction accuracy was lower than for similar alveograph traits used in other studies (Kristensen et al., 2019; Lado et al., 2018). Similarly, prediction accuracy of loaf baking quality traits were high in comparison to some earlier stage predictive traits but were still lower than for studies with large breeding program datasets (Battenfield et al., 2016; Ibba et al., 2020), but were similar to Hayes et al. (2017) and higher than results from Longin et al. (2020) who used a much smaller panel.

In terms of differences among models, results presented here were in contrast to other studies where semi-parametric models such as EG-BLUP that include epistatic effects (Jiang & Reif, 2015; Martini et al., 2017) or modeling kinship using a Gaussian kernel with RKHS (de los Campos et al., 2010; Gianola & van Kaam, 2008) generally increase prediction accuracy over standard GBLUP. The poor performance of EG-BLUP and RKHS models presented here may be down to how optimum values of the bandwidth parameter must be determined or the strong kinship structure in the panel. LASSO models performed most unreliably of all models, reflecting the polygenic genetic architecture of most traits investigated here. LASSO models work particularly well for relatively simple traits with only a few and additive genetic effects (Scott et al., 2021), so the good performance of these models for most loaf baking quality traits supports the evidence from the multiple major QTL effects such as the *Glu-1D* locus that explained a high proportion of variation in these traits in GWAS. The non-parametric machine learning models including random forest and gradient boosting machines seemed to perform competitively for most traits demonstrating their flexibility to model a broad range of genetic architectures. This is likely due to their ability to take complex non-additive genetic effects into account (Niel et al., 2015). Charmet et al. (2020) also found that random forest had the highest accuracy for predicting wheat yield compared to a wide range of other models. Despite the complex and comprehensive hyperparameter tuning method used here, which requires extended computational time, gradient boosting machine models performed slightly worse than random forest for all traits except HFN. However, other studies have found that GBM models are able to offer better accuracy over random forest (Ogutu et al., 2011), and Montesinos-López et al. (2022) found that GBM outperformed a Bayesian threshold GBLUP model in multiple datasets. These results demonstrate good prospects for more generally applying machine learning approaches to genomic prediction, and other deep learning approaches are likely to become increasingly useful for larger datasets (Montesinos-López et al., 2021).

Multi-trait prediction models which model the covariance structure among response variable traits were also tested. For cross validation scenarios where no phenotypic data are available for test lines, no increase in accuracy was found compared to single-trait models. This is in contrast to findings for predicting some similar quality traits by Sandhu et al. (2022). However, cross validation schemes which included some predictive trait data for test lines did often show improved prediction accuracy over both phenotypic selection and single-trait genomic prediction models. This demonstrates potential to maximize prediction accuracy by combining both trait-assisted genomic selection programs. The benefit of multi-trait models only when correlated traits are measured on test genotypes aligns with several others taking this approach, such as Zhang-Biehn et al. (2021) who found similar results for predicting wheat baking quality traits and Lado et al. (2018) who used predictive quality traits to aid

prediction of dough alveograph traits. In theory, a multi-trait approach should take advantage of correlations among related traits with phenotypes only in the training set for simulated datasets (Calus & Veerkamp, 2011; Jia & Jannink, 2012). However, even other studies making multi-trait predictions of yield do not find this in real data (Sun et al., 2017). The advantage of these models are likely to be greatest when secondary traits are highly correlated and have greater heritability than the target trait (Montesinos-López et al., 2021, 2022), but this is rarely the case in breeding datasets. Here, we found that the dough rheology traits including stability and softening had surprisingly low heritability compared to other studies (Battenfield et al., 2016; Lado et al., 2018). This seemed to be due to a large error component, particularly in the second year of main trials, and may explain the lack of advantage of multi-trait genomic prediction models.

The combined costs of both genotyping for genomic selection and phenotyping may not outweigh the benefit of increased accuracy from trait-assisted genomic selection. Therefore, multi-trait prediction of completely unphenotyped breeding lines that would be produced from speed breeding and genomic selection programs remains a challenge but of high relevance when using genomic predictions in multi-trait selection indices (Moeinizade et al., 2020). Trait-assisted genomic selection also likely achieves an increase in accuracy by the secondary traits accounting for G×E in the test material and models that integrate covariances between traits, genotypes, and environments have been developed (Montesinos-López et al., 2016). Therefore, the enhanced prediction accuracy of trait-assisted genomic prediction models found here within the same set of environments is likely overestimated compared to making predictions into separate unknown environments. Furthermore, the value of genomic prediction models that are trained on several years of previous trials will be greater for achieving long-term genetic gain in future years with unpredictable G×E effects compared to phenotypic selection that can only be made in a limited number of years.

4.4 | Best use of genomic selection in an applied wheat breeding program

In the context of an applied breeding program, a few generations of inbreeding to generate recombinant inbred lines (typically F3 of F4) are required before genotyping for genomic selection. Genomic prediction must therefore be competitive in terms of cost and accuracy compared to phenotypic selection based on predictive traits that can be measured on large numbers of breeding lines at the same stage. In the DSV UK breeding program compared here, the first stage predictive traits: protein content, specific weight, and HFN are tested on F4 breeding lines while second stage predictive traits dough stability, softening, and development time are tested using a DoughLab at F5 generations and are used to select a much smaller subset of lines to go forward for expensive loaf baking tests for late-stage selection. Using first-stage predictive traits in multiple regression models made generally inaccurate predictions of loaf baking quality traits but relatively good predictions of flour extraction rate and bread crumb whiteness (L*) due to the high correlation between these traits and grain specific weight. Groos et al. (2007) also showed the difficulty in predicting baking performance from predictive quality traits in a French study but found a stronger positive association between protein content and bread score which may be due to the differences in bread making methods used. Here, phenotypic selection was only competitive with the best genomic selection models for loaf baking quality traits once second stage dough rheology traits were measured, and these are much lower throughput and costly to run. However, other higher throughput measures of gluten quality, such as Zeleny and SDS sedimentation values (Preston et al., 1982; Zeleny et al., 1960) or solvent retention capacity (Kweon et al., 2011), may offer an effective intermediate phenotyping stage prior to the stage 2 dough rheology testing presented here. Hayes et al. (2017) also presented a novel method for integrating near infrared (NIR) and nuclear magnetic resonance (NMR) measurements on only small samples of flour in genomic prediction models. Nevertheless, comparison of results between phenotypic and genomic prediction demonstrated that effective selections for bread baking quality can be made using genomic predictions at earlier stages of the breeding program and at lower costs than early-stage phenotyping, while predicting late-stage baking traits with similar accuracy to the predictions from stage 2 phenotypic data.

As well as improving and accelerating forward selection of breeding lines with good expected loaf baking quality, genomic selection without the need for slow early-stage phenotyping from field grown material enables much faster early generation times to develop sufficiently inbred line material by making use of recently advanced methods in speed breeding technologies (Cha et al., 2022; Ghosh et al., 2018). In this way, breeding lines with high genomic estimated breeding value can be recycled for crossing at earlier stages to minimize the breeding cycle time and maximize genetic gain in applied breeding programs (Alahmad et al., 2022; Voss-Fels et al., 2019). Once a sufficiently large genomic prediction training dataset is established, genomic selections can be made on large numbers of completely unphenotyped breeding lines, and the training dataset can be kept up to date and accurate by incorporating phenotypic data from previous cycles of advanced line evaluation trials. In practice, genomic selection is of greater value as other important traits such as yield and disease resistance can also be included and predicted at the same cost of genotyping and selection indices can be

applied to optimize genetic gain in relation to economic value of multiple traits (Céron-Rojas & Crossa, 2018).

5 | CONCLUSIONS

We report trends in wheat quality traits over the last half a century of intensive wheat breeding in the United Kingdom. Continuous selection for decreased protein content has been compensated for by increased gluten quality to achieve improved bread baking quality using current industrial baking methods. Identified QTL for quality traits changed in frequency over time, revealing the genetic basis for trait variation over time as well as future opportunities for marker-assisted selection and improvement of quality traits that have been historically overlooked. We also demonstrate that genomic selection and trait-assisted genomic selection hold great potential for cost-effective selection and improvement of otherwise expensive milling and loaf baking quality traits in a commercial wheat breeding program context.

AUTHOR CONTRIBUTIONS

Nick Fradgley: conceptualization; data curation; formal analysis; investigation; methodology; validation; visualization; writing—original draft; writing—review and editing. **Alison Bentley**: funding acquisition; investigation; methodology; project administration; supervision; validation; writing—review and editing. **Keith Gardner**: conceptualization; formal analysis; methodology; project administration; supervision; validation; writing—review and editing. **Stéphanie Swarbreck**: methodology; project administration; supervision; validation; writing—review and editing. **Matt Kerton**: data curation; funding acquisition; investigation; methodology; project administration; resources; supervision; validation; writing—review and editing.

ACKNOWLEDGMENTS

This work was funded by Biotechnology and Biological Sciences Research Council (BBSRC) grant number: BB/M011194/1 as part of the PhD research project of NF. DSV UK provided data and in-kind contribution to the project. ARB is supported by the OneCGIAR Accelerated Breeding Initiative. We would also like to thank the two anonymous reviewers whose comments have greatly improved the manuscript.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The genotypic and phenotypic datasets generated and analyzed in the current study are available at https://figshare.com/projects/Wheat_milling_and_baking_quality_dataset/153093. However, genetic marker names and genotype id names have both been anonymized due to commercial sensitivity of data relevant to the DSV UK breeding program.

ORCID

Nick S. Fradgley https://orcid.org/0000-0002-7868-8795
Alison R. Bentley https://orcid.org/0000-0001-5519-4357
Keith A. Gardner https://orcid.org/0000-0002-4890-301X
Stéphanie M. Swarbreck https://orcid.org/0000-0001-8355-7354
Matt Kerton https://orcid.org/0000-0002-9587-8523

REFERENCES

AACC. (2000). *AACC approved methods of analysis* (11th ed.). Cereals and Grains Association. https://www.cerealsgrains.org/resources/methods/Pages/default.aspx

Adamski, N. M., Borrill, P., Brinton, J., Harrington, S. A., Marchal, C., Bentley, A. R., Bovill, W. D., Cattivelli, L., Cockram, J., Contreras-Moreira, B., Ford, B., Ghosh, S., Harwood, W., Hassani-Pak, K., Hayta, S., Hickey, L. T., Kanyuka, K., King, J., Maccaferrri, M., … Uauy, C. (2020). A roadmap for gene functional characterisation in crops with large genomes: Lessons from polyploid wheat. *ELife*, *9*, e55646. https://doi.org/10.7554/eLife.55646

AHDB. (2022). *AHDB recommended lists for cereals and oilseeds 2022–23* (1st ed.). Agriculture and Horticulture Development Board.

Alahmad, S., Rambla, C., Voss-Fels, K. P., & Hickey, L. T. (2022). Accelerating breeding cycles. In M. P. Reynolds, & H. -J. Braun (Eds.), *Wheat improvement: Food security in a changing climate* (pp. 557–571). Springer International Publishing. https://doi.org/10.1007/978-3-030-90673-3_30

Allen, A. M., Winfield, M. O., Burridge, A. J., Downie, R. C., Benbow, H. R., Barker, G. L. A., Wilkinson, P. A., Coghill, J., Waterfall, C., Davassi, A., Scopes, G., Pirani, A., Webster, T., Brew, F., Bloor, C., Griffiths, S., Bentley, A. R., Alda, M., Jack, P., … Edwards, K. J. (2017). Characterization of a wheat breeders' array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (*Triticum aestivum*). *Plant Biotechnology Journal*, *15*(3), 390–401. https://doi.org/10.1111/pbi.12635

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Anderson, O. D., Greene, F. C., Yip, R. E., Halford, N. G., Shewry, P. R., & Malpica-Romero, J. -M. (1989). Nucleotide sequences of the two high-molecular-weight glutenin genes from the D-genome of a hexaploid bread wheat, *Triticum aestivum* L. cv Cheyenne. *Nucleic Acids Research*, *17*(1), 461–462. https://doi.org/10.1093/nar/17.1.461

Ayala, M., Guzmán, C., Peña, R. J., & Alvarez, J. B. (2016). Genetic diversity and molecular characterization of puroindoline genes (Pina-D1 and Pinb-D1) in bread wheat landraces from Andalusia (Southern Spain). *Journal of Cereal Science*, *71*, 61–65. https://doi.org/10.1016/j.jcs.2016.07.017

Battenfield, S. D., Guzmán, C., Gaynor, R. C., Singh, R. P., Peña, R. J., Dreisigacker, S., Fritz, A. K., & Poland, J. A. (2016). Genomic selection for processing and end-use quality traits in the CIMMYT Spring Bread Wheat Breeding Program. *The Plant*

Genome, *9*(2), plantgenome2016.01.0005. https://doi.org/10.3835/plantgenome2016.01.0005

Belderok, B., Mesdag, J., Mesdag, H., & Donner, D. A. (2000). *Bread-making quality of wheat: A century of breeding in Europe*. Springer Science & Business Media.

Bentley, A. (2022). Broken bread—Avert global wheat crisis caused by invasion of Ukraine. *Nature*, *603*(7902), 551–551. https://doi.org/10.1038/d41586-022-00789-x

Bentley, A. R., Donovan, J., Sonder, K., Baudron, F., Lewis, J. M., Voss, R., Rutsaert, P., Poole, N., Kamoun, S., Saunders, D. G. O., Hodson, D., Hughes, D. P., Negra, C., Ibba, M. I., Snapp, S., Sida, T. S., Jaleta, M., Tesfaye, K., Becker-Reshef, I., & Govaerts, B. (2022). Near- to long-term measures to stabilize global wheat supplies and food security. *Nature Food*, 1–4. https://doi.org/10.1038/s43016-022-00559-y

Calderini, D. F., & Slafer, G. A. (1998). Changes in yield and yield stability in wheat during the 20th century. *Field Crops Research*, *57*(3), 335–347. https://doi.org/10.1016/S0378-4290(98)00080-X

Calus, M. P., & Veerkamp, R. F. (2011). Accuracy of multi-trait genomic selection using different methods. *Genetics Selection Evolution*, *43*(1), 26. https://doi.org/10.1186/1297-9686-43-26

Cauvain, S. P., & Young, L. S. (2006). *The Chorleywood bread process*. Woodhead Publishing.

Céron-Rojas, J. J., & Crossa, J. (2018). *Linear selection indices in modern plant breeding*. Springer Nature. https://doi.org/10.1007/978-3-319-91223-3

Cha, J. -K., O'Connor, K., Alahmad, S., Lee, J. -H., Dinglasan, E., Park, H., Lee, S.-M., Hirsz, D., Kwon, S. -W., Kwon, Y., Kim, K. -M., Ko, J. -M., Hickey, L. T., Shin, D., & Dixon, L. E. (2022). Speed vernalization to accelerate generation advance in winter cereal crops. *Molecular Plant*, *15*(8), 1300–1309. https://doi.org/10.1016/j.molp.2022.06.012

Charmet, G., Tran, L.-G., Auzanneau, J., Rincent, R., & Bouchet, S. (2020). BWGS: A R package for genomic selection and its application to a wheat breeding programme. *PLoS ONE*, *15*(4), e0222733. https://doi.org/10.1371/journal.pone.0222733

Costa-Neto, G., & Fritsche-Neto, R. (2021). Enviromics: Bridging different sources of data, building one framework. *Crop Breeding and Applied Biotechnology*, *21*(spe), 393521–393533. https://doi.org/10.1590/1984-70332021V21SA25

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., Burgueño, J., González-Camacho, J. M., Pérez-Elizalde, S., Beyene, Y., Dreisigacker, S., Singh, R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J., & Varshney, R. K. (2017). Genomic selection in plant breeding: Methods, models, and perspectives. *Trends in Plant Science*, *22*(11), 961–975. https://doi.org/10.1016/j.tplants.2017.08.011

De Bustos, A., Rubio, P., & Jouve, N. (2000). Molecular characterisation of the inactive allele of the gene Glu-A1 and the development of a set of AS-PCR markers for HMW glutenins of wheat. *Theoretical and Applied Genetics*, *100*(7), 1085–1094. https://doi.org/10.1007/s001220051390

de los Campos, G., Gianola, D., Rosa, G. J. M., Weigel, K. A., & Crossa, J. (2010). Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genetics Research*, *92*(4), 295–308. https://doi.org/10.1017/S0016672310000285

Donini, P., Law, J. R., Koebner, R. M. D., Reeves, J. C., & Cooke, R. J. (2000). Temporal trends in the diversity of UK wheat. *Theoretical and Applied Genetics*, *100*(6), 912–917. https://doi.org/10.1007/s001220051370

Eisler, M. C., Lee, M. R. F., Tarlton, J. F., Martin, G. B., Beddington, J., Dungait, J. A. J., Greathead, H., Liu, J., Mathew, S., Miller, H., Misselbrook, T., Murray, P., Vinod, V. K., Van Saun, R., & Winter, M. (2014). Agriculture: Steps to sustainable livestock. *Nature*, *507*(7490), Article 7490. https://doi.org/10.1038/507032a

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R Package rrBLUP. *The Plant Genome*, *4*(3), 250–255. https://doi.org/10.3835/PLANTGENOME2011.08.0024

Fradgley, N., Gardner, K., Kerton, M., Swarbreck, S. M., & Bentley, A. R. (2022). Trade-offs in the genetic control of functional and nutritional quality traits in UK winter wheat. *Heredity*, *128*(6), 420–433. https://doi.org/10.1038/s41437-022-00503-7

Fradgley, N., Gardner, K. A., Cockram, J., Elderfield, J., Hickey, J. M., Howell, P., Jackson, R., & Mackay, I. J. (2019). A large-scale pedigree resource of wheat reveals evidence for adaptation and selection by breeders. *PLoS Biology*, *17*(2), e3000071.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367–378.

Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22.

Fufa, H., Baenziger, P. S., Beecher, B. S., Graybosch, R. A., Eskridge, K. M., & Nelson, L. A. (2005). Genetic improvement trends in agronomic performances and end-use quality characteristics among hard red winter wheat cultivars in Nebraska. *Euphytica*, *144*(1), 187–198. https://doi.org/10.1007/s10681-005-5811-x

Ghosh, S., Watson, A., Gonzalez-Navarro, O. E., Ramirez-Gonzalez, R. H., Yanes, L., Mendoza-Suárez, M., Simmonds, J., Wells, R., Rayner, T., Green, P., Hafeez, A., Hayta, S., Melton, R. E., Steed, A., Sarkar, A., Carter, J., Perkins, L., Lord, J., Tester, M., … Hickey, L. T. (2018). Speed breeding in growth chambers and glasshouses for crop breeding and model plant research. *Nature Protocols*, *13*(12), 12. https://doi.org/10.1038/s41596-018-0072-z

Gianola, D., & van Kaam, J. B. C. H. M. (2008). Reproducing Kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics*, *178*(4), 2289–2303. https://doi.org/10.1534/genetics.107.084285

Graybosch, R. A. (2001). Mini review: Uneasy unions: Quality effects of rye chromatin transfers to wheat. *Journal of Cereal Science*, *33*(1), 3–16. https://doi.org/10.1006/jcrs.2000.0336

Groos, C., Bervas, E., Chanliaud, E., & Charmet, G. (2007). Genetic analysis of bread-making quality scores in bread wheat using a recombinant inbred line population. *Theoretical and Applied Genetics*, *115*(3), 313–323. https://doi.org/10.1007/s00122-007-0563-8

Guzmán, C., Autrique, E., Mondal, S., Huerta-Espino, J., Singh, R. P., Vargas, M., Crossa, J., Amaya, A., & Peña, R. J. (2017). Genetic improvement of grain quality traits for CIMMYT semi-dwarf spring bread wheat varieties developed during 1965–2015: 50 years of breeding. *Field Crops Research*, *210*, 192–196. https://doi.org/10.1016/j.fcr.2017.06.002

Guzman, C., Peña, R. J., Singh, R., Autrique, E., Dreisigacker, S., Crossa, J., Rutkoski, J., Poland, J., & Battenfield, S. (2016). Wheat quality improvement at CIMMYT and the use of genomic selection on it. *Applied & Translational Genomics*, *11*, 3–8. https://doi.org/10.1016/j.atg.2016.10.004

Hayes, B. J., Panozzo, J., Walker, C. K., Choy, A. L., Kant, S., Wong, D., Tibbits, J., Daetwyler, H. D., Rochfort, S., Hayden, M. J., &

Spangenberg, G. C. (2017). Accelerating wheat breeding for end-use quality with multi-trait genomic predictions incorporating near infrared and nuclear magnetic resonance-derived phenotypes. *Theoretical and Applied Genetics*, *130*(12), 2505–2519. https://doi.org/10.1007/s00122-017-2972-7

Howard, R., Jarquin, D., & Crossa, J. (2022). Overview of genomic predictions (GP) methods and the associated assumptions on the variance of marker effect, and on the architecture of the target trait. In N. Ahmadi, & J. Bartholomé (Eds.), *Complex trait prediction: Methods and protocols*(pp. 139–156). Springer. https://doi.org/10.1007/978-1-0716-2205-6_5

Ibba, M. I., Crossa, J., Montesinos-López, O. A., Montesinos-López, A., Juliana, P., Guzman, C., Delorean, E., Dreisigacker, S., & Poland, J. (2020). Genome-based prediction of multiple wheat quality traits in multiple years. *The Plant Genome*, *13*(3), e20034. https://doi.org/10.1002/tpg2.20034

Ibba, M. I., Juliana, P., Hernández-Espinosa, N., Posadas-Romano, G., Dreisigacker, S., Sehgal, D., Crespo-Herrera, L., Singh, R., & Guzmán, C. (2021). Genome-wide association analysis for arabinoxylan content in common wheat (*T. Aestivum* L.) flour. *Journal of Cereal Science*, *98*, 103166. https://doi.org/10.1016/j.jcs.2021.103166

IWGSC, A., R, E., K, S., N, F., C, K., B, R., J, P., C, J., Choulet, F., Distelfeld, A., Poland, J., Ronen, G., Sharpe, A. G., Barad, O., Baruch, K., Keeble-Gagnère, G., Mascher, M., Ben-Zvi, G., Josselin, A.-A., … Wang, L. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, *361*(6403), eaar7191. https://doi.org/10.1126/science.aar7191

Jia, Y., & Jannink, J. -L. (2012). Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics*, *192*(4), 1513–1522. https://doi.org/10.1534/genetics.112.144246

Jiang, Y., & Reif, J. C. (2015). Modeling epistasis in genomic selection. *Genetics*, *201*(2), 759–768. https://doi.org/10.1534/genetics.115.177907

Kinsella, R. J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., Kersey, P., & Flicek, P. (2011). Ensembl BioMarts: A hub for data retrieval across taxonomic space. *Database*, *2011*, bar030. https://doi.org/10.1093/database/bar030

Kristensen, P. S., Jahoor, A., Andersen, J. R., Cericola, F., Orabi, J., Janss, L. L., & Jensen, J. (2018). Genome-wide association studies and comparison of models and cross-validation strategies for genomic prediction of quality traits in advanced winter wheat breeding lines. *Frontiers in Plant Science*, *9*. https://www.frontiersin.org/article/10.3389/fpls.2018.00069

Kristensen, P. S., Jensen, J., Andersen, J. R., Guzmán, C., Orabi, J., & Jahoor, A. (2019). Genomic prediction and genome-wide association studies of flour yield and alveograph quality traits using advanced winter wheat breeding material. *Genes*, *10*(9), 9. https://doi.org/10.3390/genes10090669

Kweon, M., Slade, L., & Levine, H. (2011). Solvent retention capacity (SRC) testing of wheat flour: Principles and value in predicting flour functionality in different wheat-based food processes and in wheat breeding—A review. *Cereal Chemistry*, *88*(6), 537–552. https://doi.org/10.1094/CCHEM-07-11-0092

Lado, B., Vázquez, D., Quincke, M., Silva, P., Aguilar, I., & Gutiérrez, L. (2018). Resource allocation optimization with multi-trait genomic prediction for bread wheat (*Triticum aestivum* L.) baking quality. *Theoretical and Applied Genetics*, *131*(12), 2719–2731. https://doi.org/10.1007/s00122-018-3186-3

Laidig, F., Piepho, H.-P., Rentel, D., Drobek, T., Meyer, U., & Huesken, A. (2017). Breeding progress, environmental variation and correlation of winter wheat yield and quality traits in German official variety trials and on-farm during 1983–2014. *Theoretical and Applied Genetics*, *130*(1), 223–245. https://doi.org/10.1007/s00122-016-2810-3

Lasztity, R. (2017). *The chemistry of cereal proteins*. Routledge.

Leigh, F. J., Wright, T. I. C., Horsnell, R. A., Dyer, S., & Bentley, A. R. (2022). Progenitor species hold untapped diversity for potential climate-responsive traits for use in wheat breeding and crop improvement. *Heredity*, *128*(5), 5. https://doi.org/10.1038/s41437-022-00527-z

Liang, D., Tang, J., Peña, R. J., Singh, R., He, X., Shen, X., Yao, D., Xia, X., & He, Z. (2010). Characterization of CIMMYT bread wheats for high- and low-molecular weight glutenin subunits and other quality-related genes with SDS-PAGE, RP-HPLC and molecular markers. *Euphytica*, *172*(2), 235–250. https://doi.org/10.1007/s10681-009-0054-x

Longin, F., Beck, H., Gütler, H., Heilig, W., Kleinert, M., Rapp, M., Philipp, N., Erban, A., Brilhaus, D., Mettler-Altmann, T., & Stich, B. (2020). Aroma and quality of breads baked from old and modern wheat varieties and their prediction from genomic and flour-based metabolite profiles. *Food Research International*, *129*, 108748. https://doi.org/10.1016/j.foodres.2019.108748

Lu, C., & Tian, H. (2017). Global nitrogen and phosphorus fertilizer use for agriculture production in the past half century: Shifted hot spots and nutrient imbalance. *Earth System Science Data*, *9*(1), 181–192. https://doi.org/10.5194/essd-9-181-2017

Lupton, F. (2005). Advances in work on breeding wheat with improved grain quality in the twentieth century. *The Journal of Agricultural Science*, *143*(2–3), 113–116. https://doi.org/10.1017/S0021859604004617

Mackay, I. J., Horwell, A., Garner, J., White, J., McKee, J., & Philpott, H. (2011). Reanalyses of the historical series of UK variety trials to quantify the contributions of genetic and environmental factors to trends and variability in yield over time. *Theoretical and Applied Genetics*, *122*(1), 225–238. https://doi.org/10.1007/s00122-010-1438-y

Martini, J. W. R., Gao, N., Cardoso, D. F., Wimmer, V., Erbe, M., Cantet, R. J. C., & Simianer, H. (2017). Genomic prediction with epistasis models: On the marker-coding-dependent performance of the extended GBLUP and properties of the categorical epistasis model (CE). *BMC Bioinformatics [Electronic Resource]*, *18*(1), 3. https://doi.org/10.1186/s12859-016-1439-1

Meuwissen, T. H., Hayes, B. J., & Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*(4), 1819–1829.

Michel, S., Löschenberger, F., Ametz, C., Pachler, B., Sparry, E., & Bürstmayr, H. (2019). Simultaneous selection for grain yield and protein content in genomics-assisted wheat breeding. *Theoretical and Applied Genetics*, *132*(6), 1745–1760. https://doi.org/10.1007/s00122-019-03312-5

Moeinizade, S., Kusmec, A., Hu, G., Wang, L., & Schnable, P. S. (2020). Multi-trait genomic selection methods for crop improvement. *Genetics*, *215*(4), 931–945. https://doi.org/10.1534/genetics.120.303305

Mohler, V., Albrecht, T., Mrva, K., Schweizer, G., & Hartl, L. (2014). Genetic analysis of falling number in three bi-parental common winter wheat populations. *Plant Breeding*, *133*(4), 448–453. https://doi.org/10.1111/pbr.12177

Montesinos-López, A., Montesinos-López, O. A., Montesinos-López, J. C., Flores-Cortes, C. A., de la Rosa, R., & Crossa, J. (2021). A guide for kernel generalized regression methods for genomic-enabled prediction. *Heredity*, *126*(4), 4. https://doi.org/10.1038/s41437-021-00412-1

Montesinos-López, O. A., Gonzalez, H. N., Montesinos-López, A., Daza-Torres, M., Lillemo, M., Montesinos-López, J. C., & Crossa, J. (2022). Comparing gradient boosting machine and Bayesian threshold BLUP for genome-based prediction of categorical traits in wheat breeding. *The Plant Genome*, *15*(3), e20214. https://doi.org/10.1002/tpg2.20214

Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Toledo, F. H., Pérez-Hernández, O., Eskridge, K. M., & Rutkoski, J. (2016). A genomic Bayesian multi-trait and multi-environment model. *G3 Genes|Genomes|Genetics*, *6*(9), 2725–2744. https://doi.org/10.1534/g3.116.032359

Montesinos-López, O. A., Montesinos-López, A., Mosqueda-Gonzalez, B. A., Montesinos-López, J. C., & Crossa, J. (2022). Accounting for correlation between traits in genomic predictions. In N. Ahmadi, & J. Bartholomé (Eds.), *Complex trait prediction: Methods and protocols* (pp. 285–327). Springer US. https://doi.org/10.1007/978-1-0716-2205-6_10

Montesinos-López, O. A., Montesinos-López, A., Pérez-Rodríguez, P., Barrón-López, J. A., Martini, J. W. R., Fajardo-Flores, S. B., Gaytan-Lugo, L. S., Santana-Mancilla, P. C., & Crossa, J. (2021). A review of deep learning applications for genomic selection. *BMC Genomics [Electronic Resource]*, *22*(1), 19. https://doi.org/10.1186/s12864-020-07319-x

Morris, C. F. (2002). Puroindolines: The molecular genetic basis of wheat grain hardness. *Plant Molecular Biology*, *48*(5), 633–647. https://doi.org/10.1023/A:1014837431178

Mount, D. W. (2007). Using the basic local alignment search tool (BLAST). *Cold Spring Harbor Protocols*, *2007*(7), pdb.top17. https://doi.org/10.1101/pdb.top17

Niel, C., Sinoquet, C., Dina, C., & Rocheleau, G. (2015). A survey about methods dedicated to epistasis detection. *Frontiers in Genetics*, *6*. https://www.frontiersin.org/article/10.3389/fgene.2015.00285

Ogutu, J. O., Piepho, H. - P., & Schulz-Streeck, T. (2011). A comparison of random forests, boosting and support vector machines for genomic selection. *BMC Proceedings*, *5*(3), S11. https://doi.org/10.1186/1753-6561-5-S3-S11

Payne, P. I., Holt, L. M., Lawrence, G. J., & Law, C. N. (1982). The genetics of gliadin and glutenin, the major storage proteins of the wheat endosperm. *Plant Foods for Human Nutrition*, *31*(3), 229–241. https://doi.org/10.1007/BF01108632

Percival, J. (1934). *Wheat in Great Britain*. Duckworth London.

Pérez, P., & de los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics*, *198*(2), 483–495. https://doi.org/10.1534/genetics.114.164442

Pérez-Rodríguez, P., & de los Campos, G. (2022). Multitrait Bayesian shrinkage and variable selection models with the BGLR-R package. *Genetics*, *222*(1), iyac112. https://doi.org/10.1093/genetics/iyac112

Preston, K. R., March, P. R., & Tipples, K. H. (1982). An assessment of the SDS-sedimentation test for the prediction of Canadian bread wheat quality. *Canadian Journal of Plant Science*, *62*(3), 545–553.

Ray, D. K., Sloat, L. L., Garcia, A. S., Davis, K. F., Ali, T., & Xie, W. (2022). Crop harvests for direct food use insufficient to meet the UN's food security goal. *Nature Food*, *3*, 367–374. https://doi.org/10.1038/s43016-022-00504-z

Ridgeway, G. (2007). Generalized Boosted Models: A guide to the gbm package. *Update*, *1*(1), 2007.

Robert, P., Brault, C., Rincent, R., & Segura, V. (2022). Phenomic selection: A new and efficient alternative to genomic selection(GS). In N. Ahmadi, & J. Bartholomé (Eds.), *Complex trait prediction: Methods and protocols* (pp. 397–420). Springer US. https://doi.org/10.1007/978-1-0716-2205-6_14

Sanchez-Garcia, M., Álvaro, F., Peremarti, A., Martín-Sánchez, J. A., & Royo, C. (2015). Changes in bread-making quality attributes of bread wheat varieties cultivated in Spain during the 20th century. *European Journal of Agronomy*, *63*, 79–88. https://doi.org/10.1016/j.eja.2014.11.006

Sandhu, K. S., Aoun, M., Morris, C. F., & Carter, A. H. (2021). Genomic selection for end-use quality and processing traits in soft white winter wheat breeding Program with machine and deep learning models. *Biology*, *10*(7), 7. https://doi.org/10.3390/biology10070689

Sandhu, K. S., Patil, S. S., Aoun, M., & Carter, A. H. (2022). Multi-trait multi-environment genomic prediction for end-use quality traits in winter wheat. *Frontiers in Genetics*, *13*. https://www.frontiersin.org/articles/10.3389/fgene.2022.831020

Scott, M. F., Fradgley, N., Bentley, A. R., Brabbs, T., Corke, F., Gardner, K. A., Horsnell, R., Howell, P., Ladejobi, O., Mackay, I. J., Mott, R., & Cockram, J. (2021). Limited haplotype diversity underlies polygenic trait architecture across 70 years of wheat breeding. *Genome Biology*, *22*(1), 137. https://doi.org/10.1186/s13059-021-02354-7

Shewry, P. R., Pellny, T. K., & Lovegrove, A. (2016). Is modern wheat bad for health? *Nature Plants*, *2*(7), 1–3.

Shewry, P. R., Wood, A. J., Hassall, K. L., Pellny, T. K., Riche, A. B., Hussain, A., Hawkesford, M. J., Griffiths, S., Penson, S., & Tucker, G. (2020). *Low protein wheat for bread making (No. 621)*. AHDB Cereals and Oilseeds.

Shiferaw, B., Smale, M., Braun, H. - J., Duveiller, E., Reynolds, M., & Muricho, G. (2013). Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. *Food Security*, *5*(3), 291–317. https://doi.org/10.1007/s12571-013-0263-y

Shorinola, O., Balcárková, B., Hyles, J., Tibbits, J. F. G., Hayden, M. J., Holušova, K., Valárik, M., Distelfeld, A., Torada, A., Barrero, J. M., & Uauy, C. (2017). Haplotype analysis of the pre-harvest sprouting resistance locus Phs-A1 reveals a causal role of TaMKK3-A in global germplasm. *Frontiers in Plant Science*, *8*. https://www.frontiersin.org/article/10.3389/fpls.2017.01555

Simmonds, N. W. (1995). The relation between yield and protein in cereal grain. *Journal of the Science of Food and Agriculture*, *67*(3), 309–315. https://doi.org/10.1002/jsfa.2740670306

Srinivasan, C. S., Thirtle, C., & Palladino, P. (2003). Winter wheat in England and Wales, 1923–1995: What do indices of genetic diversity reveal? *Plant Genetic Resources*, *1*(1), 43–57. https://doi.org/10.1079/PGR20031

Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—Non-parametric missing value imputation for mixed-type data. *Bioinformatics*, *28*(1), 112–118.

Sun, J., Rutkoski, J. E., Poland, J. A., Crossa, J., Jannink, J. - L., & Sorrells, M. E. (2017). Multitrait, random regression, or simple repeatability model in high-throughput phenotyping data improve genomic prediction for wheat grain yield. *The Plant Genome*, *10*(2), plantgenome2016.11.0111. https://doi.org/10.3835/plantgenome2016.11.0111

Tabbita, F., Pearce, S., & Barneix, A. J. (2017). Breeding for increased grain protein and micronutrient content in wheat: Ten years of the GPC-B1 gene. *Journal of Cereal Science*, *73*, 183–191. https://doi.org/10.1016/j.jcs.2017.01.003

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, *91*(11), 4414–4423. https://doi.org/10.3168/jds.2007-0980

Voss-Fels, K. P., Cooper, M., & Hayes, B. J. (2019). Accelerating crop genetic gains with genomic selection. *Theoretical and Applied Genetics*, *132*(3), 669–686. https://doi.org/10.1007/s00122-018-3270-8

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E., Maccaferri, M., Salvi, S., Milner, S. G., Cattivelli, L., Mastrangelo, A. M., Whan, A., Stephen, S., Barker, G., Wieseke, R., Plieske, J., Consortium, I. W. G. S., Lillemo, M., Mather, D., … Akhunov, E. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnology Journal*, *12*(6), 787–796. https://doi.org/10.1111/pbi.12183

White, J., Law, J. R., MacKay, I., Chalmers, K. J., Smith, J. S. C., Kilian, A., & Powell, W. (2008). The genetic diversity of UK, US and Australian cultivars of Triticum aestivum measured by DArT markers and considered by genome. *Theoretical and Applied Genetics*, *116*(3), 439–453. https://doi.org/10.1007/s00122-007-0681-3

White, J., Sharma, R., Balding, D., Cockram, J., & Mackay, I. J. (2022). Genome-wide association mapping of Hagberg falling number, protein content, test weight, and grain yield in U.K. wheat. *Crop Science*, *62*(3), 965–981. https://doi.org/10.1002/csc2.20692

Wilkinson, J. M., & Lee, M. R. F. (2018). Review: Use of human-edible animal feeds by ruminant livestock. *Animal*, *12*(8), 1735–1743. https://doi.org/10.1017/S175173111700218X

Wright, M. N., & Ziegler, A. (2017). ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, *77*(1). https://doi.org/10.18637/jss.v077.i01

Zeleny, L., Greenaway, W. T., Gurney, G. M., Flfield, C. C., & Lebsock, K. (1960). Sedimentation value as an index of dough-mixing characteristics in early-generation wheat selections. *Cereal Chemistry*, *37*, 673–678.

Zhang-Biehn, S., Fritz, A. K., Zhang, G., Evers, B., Regan, R., & Poland, J. (2021). Accelerating wheat breeding for end-use quality through association mapping and multivariate genomic prediction. *The Plant Genome*, *14*(3), e20164. https://doi.org/10.1002/tpg2.20164

SUPPORTING INFORMATION

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

How to cite this article: Fradgley, N. S., Bentley, A. R., Gardner, K. A., Swarbreck, S. M., & Kerton, M. (2023). Maintenance of UK bread baking quality: Trends in wheat quality traits over 50 years of breeding and potential for future application of genomic-assisted selection. *The Plant Genome*, *16,* e20326. https://doi.org/10.1002/tpg2.20326