bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Multi-trait ensemble genomic prediction and simulations of recurrent selection highlight importance of complex trait genetic architecture in long-term genetic gains in wheat

Nick Fradgley[1*], Keith A. Gardner[1†], Alison R. Bentley[1†], Phil Howell[1], Ian J. Mackay[1‡], Michael F. Scott[2$], Richard Mott[2], James Cockram[1]

[1]NIAB, 93 Lawrence Weaver Road, Cambridge, CB3 0LE, United Kingdom

[2]University College London (UCL) Genetics Institute, Gower St, London, WC1E 6BT, United Kingdom

[†]Current address: International Maize and Wheat Improvement Center (CIMMYT), Carretera México-Veracruz, México

[‡]Current address: Scotland's Rural College (SRUC), Kings Buildings, West Mains Road, Edinburgh, EH9 3JG, United Kingdom

[$]Current address: School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, United Kingdom

[*]Corresponding author: nick.fradgley@niab.com

Key words

Triticum aestivum L., recurrent selection, simulation, genomic prediction, multiparent advanced generation inter-cross (MAGIC) population.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

Cereal crop breeders have achieved considerable genetic gain in genetically complex traits, such as grain yield, while maintaining genetic diversity. However, focus on selection for yield has negatively impacted other important traits. To better understand selection within a breeding context, and how it might be optimised, we analysed genotypic and phenotypic data from a diverse, 16-founder wheat multi-parent advanced generation inter-cross (MAGIC) population.

Compared to single-trait models, multi-trait ensemble genomic prediction models increased prediction accuracy for almost 90% of traits, improving grain yield prediction accuracy by 3-52%. For complex traits, non-parametric models (Random Forest) also outperformed simplified, additive models (LASSO), increasing grain yield prediction accuracy by 10-36%. Simulations of recurrent genomic selection then showed that sustained greater forward prediction accuracy optimised long-term genetic gains.

Simulations of selection on grain yield found indirect responses in related traits, which involved optimisation of antagonistic trait relationships. We found multi-trait selection indices could be used to optimise undesirable relationships, such as the trade-off between grain yield and protein content, or combine traits of interest, such as yield and weed competitive ability.

Simulations of phenotypic selection found that including Random Forest rather than LASSO genetic models, and multi-trait rather than single-trait models as the true genetic model, accelerated and extended long-term genetic gain whilst maintaining genetic diversity. These results suggest important roles of pleiotropy and epistasis in the wider context of wheat breeding programmes and provide insights into mechanisms for continued genetic gain in a limited genepool and optimisation of multiple traits for crop improvement.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1. Introduction

Classical plant breeding aims to achieve continuous genetic gain by recurrent selection of important traits over many generations. However, the biological and genetic processes that allow continued genetic gain within a finite genepool are still unclear. For example, the Illinois maize long-term selection experiment achieved continuous increases in seed oil and protein concentration in a closed population for more than 100 generations without apparent loss of genetic variation (Dudley, 2007). Long-term trends in wheat (*Triticum aestivum* L.) breeding also reflect this, where significant genetic gain in traits such as yield has been achieved in the last century (Mackay et al., 2011; McCraig et al., 1995; Tadesse et al., 2019), whilst molecular studies have not found the expected reductions in genetic diversity over the same period of modern plant breeding (Fu, 2015; van de Wouw et al., 2010; White et al., 2008).

Selection on one trait can have positive or negative pleiotropic effects on other traits. For example, the Illinois long-term selection experiment found correlated responses to selection for oil and protein content and indirect effects on other traits, such as starch content. Wheat breeding requires selection for multiple traits of economic importance, including grain yield and quality traits, as well as other agronomically important or physiologically adaptive traits, such as developmental stage, plant architecture and disease resistance. In many cases, positive correlated responses in combinations of desirable traits can be achieved, but there are often complex trade-offs between antagonistically related traits. Considerable work has succeeded in identifying underlying quantitative trait loci (QTL) controlling individual yield components, such as grain size (e.g. Brinton et al., 2017) and spikelet number (e.g. Kuzay et al., 2019; Muqaddasi et al., 2019). However, the effects of such yield component loci rarely have consistent positive effects on yield in broader genetic backgrounds due to compensatory effects which trade-off against other yield components. For example, increased grain number per inflorescence in wheat is commonly associated with reductions in other yield components such as thousand grain weight, or tiller number (Corsi et al., 2021; Quintero et al., 2018; Xie and Sparkes, 2021).

In general, long-term increases in wheat yields have been achieved phenotypically by optimisation of harvest index (the ratio of grain to total shoot dry matter) to reduce intra-crop competition (Fischer and Kertesz, 1976), as well as through increased grain filling with starch carbohydrates (Lovegrove et al., 2020; Shewry et al., 2020). However, these have led to negative trade-offs in other valuable traits. Decreased competitive ability of modern wheat varieties with weeds (Murphy et al., 2008; Vandeleur and Gill, 2004) necessitates increased reliance on herbicides as well as potentially poorer uptake of soil nutrients (Ruisi et al., 2015). Yield loss from weed competition has become even more problematic in intensified cropping systems (Storkey et al., 2021). Additionally, increased yield and starch grain filling has been subject to the long-standing trade-off between yield and grain protein content (Simmonds, 1995; White et al., 2021), and has led to dilution of wheat grain protein content (Austin et al., 1980; Fufa et al., 2005) and mineral nutrient density (Davis, 2009; Shrewry et al., 2016). This has also led to higher optimum nitrogen fertiliser application rates to meet milling wheat grain protein requirements with diminishing increases in yield, and thus poorer nitrogen use efficiency (Hawkesford, 2014). Trade-offs between grain yield and both protein content and weed competitive ability seem not to have been generally addressed by commercial breeding due to yield being considered the highest economically important trait. Recent analysis by Raherison et al. (2020) suggested that negative pleiotropic genetic effects in wheat have rarely been compensated for and optimised by breeding, and Yang et al. (2022) showed that breeders' selections

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

have almost always been in favour of yield at the expense of protein. Changing economic, legislational, environmental and societal factors mean that breeding focus will increasingly need to consider how to deliver sustainable intensification of food supply, ensuring yield stability of our future crops in the face of such pressures. Plant breeding will play a role in delivering these goals, and will likely require the application of new breeding approaches and methodologies.

Genomic selection models aim to predict as large a proportion of heritable phenotypic variation as possible using genome-wide marker data to allocate estimated breeding values to untested individuals (Jannink et al., 2010; Meuwissen et al., 2001), and are likely to be a major source of improvement in plant breeding in the coming decades (Mackay et al., 2021). Genomic prediction models include genetic effects that don't necessarily reach genome-wide significance in QTL mapping, which only detects large additive genetic effects and often fails to account for a large proportion of heritable trait variation in traits with complex genetic architectures, despite extensive genomic and phenotypic characterisation (Goddard et al., 2016). However, the role of non-additive epistatic effects in complex trait genetic architectures (i.e the interactions between genes) remains understudied, and is often overlooked (Carlborg and Haley, 2004) – likely due to the high computational requirements to model high order interactions (Jiang and Reif, 2015). Genomic prediction models that take epistatic effects into account have recently been developed, including the extension of the genomic best linear unbiased prediction (GBLUP) (Jiang and Reif, 2015) and machine/ensemble learning methods such as Random Forest (Schmalohr et al., 2018; Wright et al., 2016), which are often able to improve prediction accuracies in real datasets (Charmet et al., 2020).

The NIAB Diverse MAGIC (Multi-parent Advanced Generation Inter-Cross) wheat population (NDM) was recently developed to investigate the genetic architecture of a range of traits in wheat (Scott et al., 2021). It consists of 16 founders genotyped via exon and promotor capture sequencing and 504 recombinant inbred lines genotyped via whole-genome sequencing and imputation, resulting in ~1.1M single nucleotide polymorphisms (SNPs) between genotypes, or 55k SNPs after filtering for linkage disequilibrium (LD) (Scott et al., 2021). The 16 founders are wheat varieties that span 70 years of breeding and capture a large proportion of the northwest European genetic diversity. The genetic diversity present in the NDM is efficiently recombined through multiple generations of inter-crossing, eroding LD accumulated in the founders over long-term selective breeding. For this reason, traditional genomic prediction models, such as GBLUP, that make use of kinship relationships (Clark et al., 2011) may perform poorly in MAGIC where causal variants can be considered more independently (Scott et al., 2021). The NDM is ideal for investigating trait relationships, due to intensive phenotyping and the lack of the confounding effects of age and origin that are present in panels of selectively bred varieties (Scott et al., 2020). Furthermore, this population provides a good test for multi-trait selection indices, such as grain yield protein deviation (GYPD; Michel et al., 2019), that have been proposed to help minimise trade-offs between traits.

Here we use NDM resources as a microcosm of long-term selection in wider wheat breeding programmes to test differing approaches to selection and genetic models. Within the overall context of understanding the phenotypic and genetic mechanisms that may enable enhanced genetic gain in the future, we (i) investigate complex trait relationships relating to yield in the observed population of lines. We then (ii) develop multi-trait genomic prediction models that increase prediction accuracy by exploiting pleiotropic effects among traits, and (iii) investigate how increased prediction accuracy translates to greater genetic gain in simulation of long term-recurrent genomic selection within the NDM. We also (iv) simulated both phenotypic and genotypic effects of recurrent phenotypic selection within the population comparing different true genetic models based on genomic

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

prediction models trained on the observed population. Our results reveal correlated responses in a wide range of traits when selection is purely on yield, as well as the potential to achieve genetic gain in several traits of interest that trade-off by using use multi-trait selection indices. Comparison of response to selection under differing genomic prediction models (simplified models with a minimal number of additive effects versus more complex polygenic models that take higher order epistatic interaction effects into account) also highlights the important role of both pleiotropy and epistasis as potential mechanisms for continued genetic gain in crop breeding.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2. Methods

2.1 Germplasm, phenotypic and genotypic datasets

Genotypic and phenotypic data for the NDM wheat population was sourced from Scott et al. (2021). Briefly, the population of 504 NDM recombinant inbred lines derived from the 16 founders was phenotyped for a wide range of traits over two successive seasons (2016-2017 and 2017-2018) in the United Kingdom (UK). All but one of the 73 traits described by Scott et al. (2021) (**Table 1**) were used, the exception being seed germination rate (GR) due to a large proportion of missing data. Traits measured in each year were considered separately. Missing data for all remaining traits (at <1.2%) were imputed with the median trait value. Line genotypes were previously characterised by skim sequencing and imputed using the founder haplotypes (Scott et al., 2021). Of 1.1M SNPs identified from founder exome and promoter sequencing, we use the subset of ~55k SNPs after pruning for LD for our analyses. Missing marker data (~1%) were imputed using the 'missForest' package (Buhlman, 2011) in R (R Core Team, 2020), which uses non-parametric Random Forest prediction models to iteratively predict and impute missing data on a marker-by-marker basis.

Table 1. Abbreviations of traits phenotyped in the NDM, as described by Scott et al. (2021). All data are from field trials, except where noted. Nursery = data collected from 1×1m unreplicated plots. Field = data collected from 2×6m replicated plots. Trait groups indicate groups of strongly positively or negatively correlated traits that grouped together through hierarchical clustering as shown in **Figure 1**. Some traits were phenotyped at multiple time points (GLA) and in both trail years so that a total of 72 traits were included.

Abbreviation	Trait	Trait group	Abbreviation	Trait	Trait group
PHS	Pre-harvest sprouting	1	GW	Grain width	6
PIG	General pigmentation	1	HET	Height to ear tip	6
SW	Specific weight	1	HFLB	Height to flag leaf base	6
FLS	Flag leaf senescence	2	LOD	Lodging	6
GS39	Flag leaf emergence date	2	TGW	Thousand grain weight	6
GS55	Ear emergence date	2	TIS	Tip infertile spikelets	6
GS65	Anthesis date	2	AWN	Presence of awns	7
JGH	Juvenile growth habit (Nursery)	2	EL	Ear length	7
SH	Spring habit	2	ETA	Ear taper	7
FLA	Flag leaf angle	3	GLAU	Glaucosity	7
FLF	Flag leaf floppiness	3	SPIG	Stem pigmentation	7
FLL	Flag leaf length	3	BIS	Basal infertile spikelets	8
GLA#	Green leaf area (10 time points)	4	EW	Ear weight	8
GPC	Grain protein content	5	FLW	Flag leaf width	8
GY	Yield	5	GPE	Grains per ear	8
FLED	Flag leaf to ear distance	6	GPS	Grains per spikelet	8
GA	Grain area	6	TS	Total spikelets	8
GL	Grain length	6	YR	Yellow rust infection (Field and Nursery)	8

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2.2 Statistical analysis

All analyses were conducted using R statistical analysis software. Pearson's correlation coefficients among all investigated traits were calculated. Hierarchical clustering of traits was performed using the 'hclust' R function and 'complete' method, where the distance matrix (d) was derived from the equation:

$$d = 2(1 - \sqrt{c^2})$$

where c represents the trait correlation matrix. Traits were then assigned to eight groups using the 'cutree' R function.

2.2.1 Genomic prediction models

Two contrasting genomic prediction models were compared for both single-trait (ST) and multi-trait (MT) models. These included generalised linear models including the Lasso penalty (LASSO) implemented in the 'glmnet' R package (Friedman et al., 2010) where the majority of SNP effects are shrunk to zero. For comparison, a non-linear, statistical learning approach was also used which generally included much larger numbers of SNPs with non-additive effects in each model: Random Forest (RF), implemented in the 'randomForest' R package (Breiman, 2001). For LASSO models, the value of lambda (a shrinkage penalty) used for each prediction was optimised using 8-fold cross validation. For RF models, 300 trees were run per model and default parameters of one third of variables randomly sampled at each split, and a minimum of five observations in terminal leaf nodes were used. Previous work by Scott et al. (2021) found that ridge regression models that include all marker effects with a small, but non-zero effect, did not have as high prediction accuracy as LASSO in the MAGIC population, and so were not further tested here.

Two types of MT models were implemented. Firstly, by performing single value decomposition (SVD) of the matrix of all phenotypes, as proposed by Montesinos-López et al. (2019a), whereby each of the decomposed and uncorrelated vectors from all the traits were predicted as traits themselves using the same genomic models as for ST predictions. The predictions of vectors were then back-transformed to the original trait scales to derive the MT predictions per-trait. Secondly, a multi-trait stacked ensemble method was also used which employs an approach often used in machine learning (Spyromitros-Xioufis et al., 2016) and has previously been applied for Bayesian multi-output regression of multi-trait predictions (Montesinos-López et al., 2019b; Sapkota et al., 2020). For this, a two-step model was used where each trait was first predicted from genomic data with the same genomic models as for ST predictions, and then all trait predictions were used as explanatory variables (features) in a second multi-trait ensemble model to again predict each response trait. Both first and second stage predictions were fitted only on data from the training fraction and predictions were independently made for test lines with only genetic marker data. Either LASSO or RF models were used to fit first stage ST models, but only RF models were used to flexibly include non-linear multi-trait relationships for the second stage ensemble models. Information from related traits is therefore used in a model, trained only on the training fraction to adjust single-trait predictions made directly from genomic data. Unlike trait-assisted genomic prediction, such as used by Fernandes et al. (2018), no observed trait data are used in the tested cross-validation fraction. As both MT prediction approaches can be applied with any genomic prediction model for each ST prediction or for SVD vectors, we were able to compare LASSO and RF genomic models for both ST and MT approaches.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Prediction accuracies were determined by performing three rounds of 10-fold random cross-validation among all lines in the dataset and averaging the three Pearson's correlation coefficients between observed and predicted trait values across all cross-validation folds. Valid comparisons of prediction accuracy were ensured by testing all prediction models using the same cross-validation fold assignments. After model cross-validation, full prediction models were fitted using the entire dataset for combinations of both ST and MT models with both LASSO and RF genomic models. Variable importance scores for each SNP marker in RF genomic prediction models and for each trait covariate in MT ensemble models were derived from the full models as the mean decrease in mean square error using the 'importance' function in the 'randomForest' R package. Effect sizes for each SNP marker were also derived from full LASSO models where the majority of SNP effects were shrunk to zero.

2.2.2 Simulations of recurrent genomic selection

We first simulated a recurrent genomic selection programme within the NDM to assess the performance of different prediction models to achieve long-term genetic gain in grain yield. This was done within a framework of assuming a true inherited genetic model based on the MT ensemble RF genetic model as outlined above and trained on the observed genetic and phenotype data. True phenotypes were derived from predictions from this model for genotypes at each cycle of simulations and the different genomic prediction models outlined above were trained on the individuals in the first cycle. New cycles of genotypes derived from crossing selected fractions of lines were simulated using a genetic map of ~55,000 SNPs.

The genetic map (Supplementary Table S1 and Supplementary Figure S1) was constructed using the 'qtl2' R package (Broman et al., 2019) with the marker data ordered by physical map position (RefSeq v1,0, IWGSC, 2018). The genetic map distance was then re-estimated using the 'est_map' function with 1,000 maximum iterations and an assumed genotyping error probability of 0.001. The cross object was considered as a 16-way multi-parent recombinant inbred line population, so the differing local recombination effects for each founder haplotypes were preserved for subsequent simulations. 23 markers were removed from the full set which caused genetic map distortion.

Selection of lines at each generation were made based on predicted phenotypes from the genomic prediction model. To reduce excessive inbreeding and loss of genetic variance, 15 lines from different 16-way or bi-parental families with the highest selection index values were selected. 30 offspring inbred line genotypes were then simulated for each of 105 possible pairwise cross combinations among the selected lines so that the following generation comprised of 3,150 lines from 105 biparental families. The phenotypes of these were again predicted from the genomic prediction models trained on the true phenotypes of the first generation and the process repeated for 20 cycles of recurrent selection. 20 iteration repeats of the simulations were run for each genomic prediction model. Genomic prediction models were fitted as detailed above for ST and MT, LASSO and RF models, but additionally RF models were run that were restricted to a tree depth of one (RF1) to completely limit the degree of marker interaction effects. 2,000 trees were used for RF1 models.

Genetic gain for each trait over the selection simulations were determined by comparing the mean true trait value of all lines at each generation to the mean true trait value in the first generation. The divergence from this mean among different traits was standardised to the standard deviation of trait values in the first generation.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The accuracy of genomic prediction models was also determined as the Pearson's correlation coefficient between the true and predicted trait values among genotypes at each simulation cycle.

2.2.3 Simulations of recurrent phenotypic selection

In addition to simulations of different genomic selection procedures within a simulated true genetic model, we also compared simulations with different true genetic models to assess both phenotypic and genomic response to selection with different genetic model assumptions of trait genetic architecture. Simulations were run as above but selections of individuals were based on the true phenotypes derived from different genetic models so that it was assumed that the simulated breeder could make perfect estimates of trait values from phenotypic selection. Different simulations were run for ST and MT as well as RF and LASSO models as outlined above and for different selection indices as outlined below. Genetic response to selection was also characterised as the change in allele frequency for all ~55,000 SNPs at each generation, and the genetic diversity was calculated as the number of polymorphic SNPs at each generation. For each set of simulations, traits or SNP markers were considered under selection rather than drift if their response to selection was significantly different to 0 considering all 20 simulation repeats using a t-test.

2.2.4 Selection indices

Indices for simulated selection were defined as follows:

1. Grain yield measured in each trial year.
2. Multi-trait index including grain yield and traits known to be associated with weed competitiveness. The weed competitive ability selection index (Weeds_ESIM) was based on the restricted eigenvector selection index method (RESIM) (Cerón-Rojas et al., 2008). For this, principal component analysis was performed on a selection of desirable traits based on the literature, which included yield measured in both years as well as traits previously identified as valuable for weed competitive ability. These included high early vigour, measured as green leaf area (GLA) over the development phase before flowering time, and planophile (horizontal) flag leaf angle (FLA) (Andrew et al., 2015; Kissing Kucek et al., 2021; Korres and Froud-Williams, 2002; Mwendwa et al., 2020). To mitigate risk of lodging (i.e. the permanent displacement of a stem from vertical), mean crop height (HET) between both years was also then restricted to values between 60-65 cm. The vector weightings on the first principal component with mean HET values between this range therefore represented the Weeds_ESIM. Most traits involved in this selection index were positively correlated so the first principal component could be assumed to provide a desirable combined index for selection.

3. Grain yield protein deviation (GYPD). The GYPD selection index was calculated as the sum of the scaled and centred mean yield and protein across both years using the equation:

$$GYPD = \frac{Y - \bar{Y}}{\sqrt{\frac{\sum(Y - \bar{Y})^2}{n - 1}}} + \frac{P - \bar{P}}{\sqrt{\frac{\sum(P - \bar{P})^2}{n - 1}}}$$

where Y and P is the mean line yield and protein across years, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

3. Results

3.1 Grain yield is correlated with multiple traits in the observed population

We analysed data from a genetically diverse and highly recombined 16-founder wheat MAGIC population for a wide range of agronomically important traits over multiple trial years (72 trait – year combinations) to investigate complex trait-trait relationships and their implications for breeding. Correlation analysis across all traits and years revealed complex trait relationships and substantial differences in patterns between the two trial years (**Figure 1** and Supplementary Table S2). Considering grain yield (GY) as the primary trait of interest, many other secondary traits were found to be correlated (**Figure 1**). The strong negative trade-off between yield and grain protein content (GPC) was mediated by yield component traits. For example, grain size traits (such as thousand grain weight; TGW), grains per spikelet (GPS) and total spikelets per ear (TS), were all positively correlated with GY, but negatively correlated with protein content and with each other. Therefore, potential benefits of selecting for any one of the yield component traits in isolation are buffered by problematic trade-offs with other yield component traits and likely have negative effects on protein content.

Differential relationships between yield and other developmental stage and plant architecture traits between the two trial years were also found. In year 1, taller and later flowering genotypes were generally higher yielding (yield – height to ear tip correlation = 0.20; yield – heading date correlation = 0.32), whereas in year 2, the correlation between height and yield was negative (correlation = -0.11) and between yield and heading date was non-significant. Therefore, strong genotype-by-environment interaction (G×E) effects on yield means that selection for yield, or related adaptive traits, in any single environment may have limited potential to increase yield in another environment. However, contrasting patterns of rainfall and temperature between the two yield trial years (**Figure 2**Figure 5), in which year 1 was characterised by high temperatures and drought before anthesis (March and April) whilst year 2 was characterised by extreme terminal heat and drought after anthesis (June and July), may explain the differences in relationships between adaptive traits and yield in this study. Both of the two trial years experienced different extremes of monthly climate variables compared to the distributions of the last 56 years, so were considered separately in these analyses.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

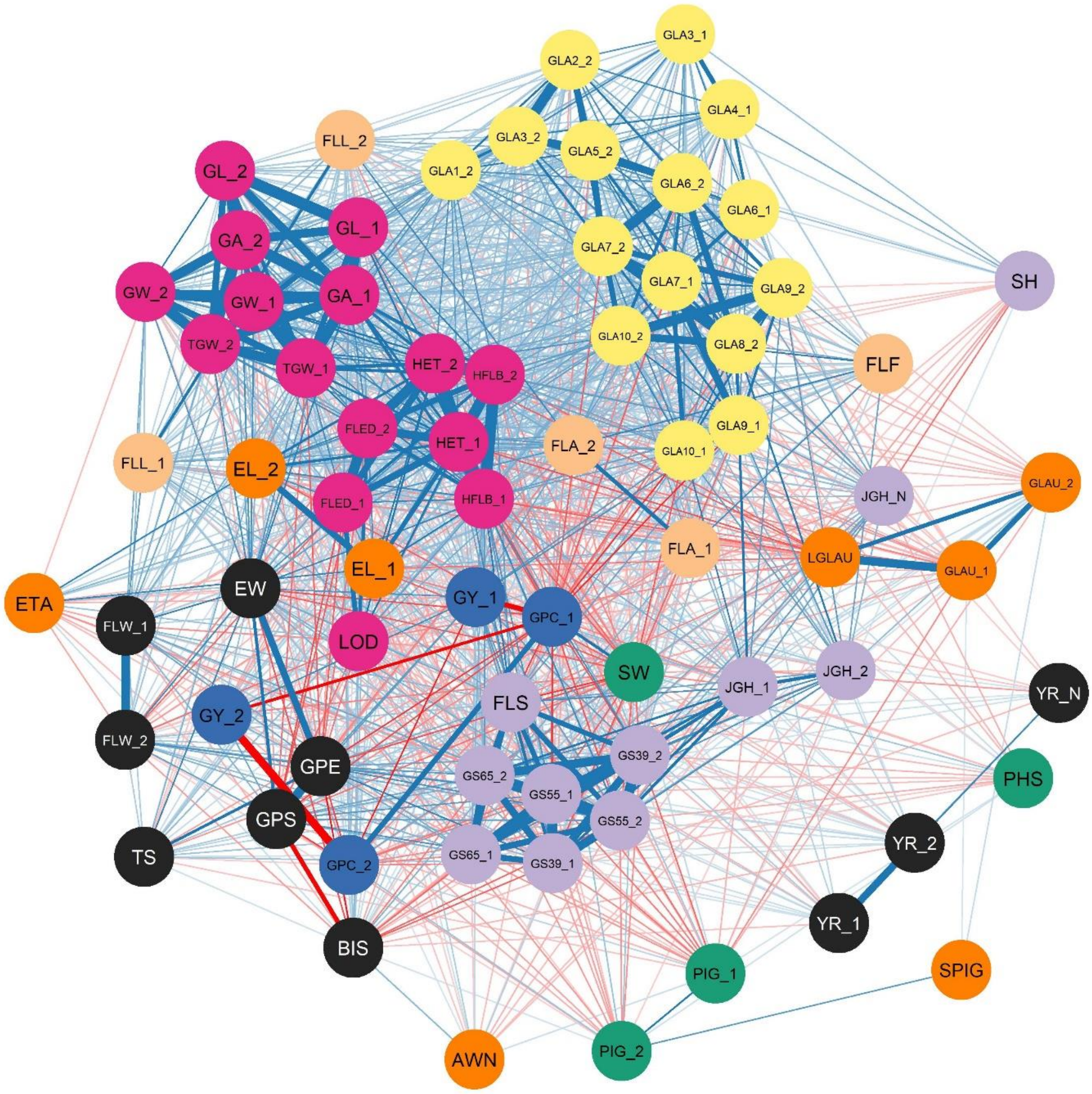


Figure 1. Correlation network for 72 traits measured in two trial years among 504 NIAB Diverse MAGIC lines. Grain yield = GY. Abbreviations for all additional trait names are given in Table 1. Trait node colours indicate the eight groups of related traits as identified using hierarchical clustering. The _1 and _2 designations used after trait abbreviations refer to trial year 1 and trial year 2, respectively. Blue and red connecting lines indicate positive and negative correlations, respectively, while line thickness is relative to correlation *p*-value significance.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

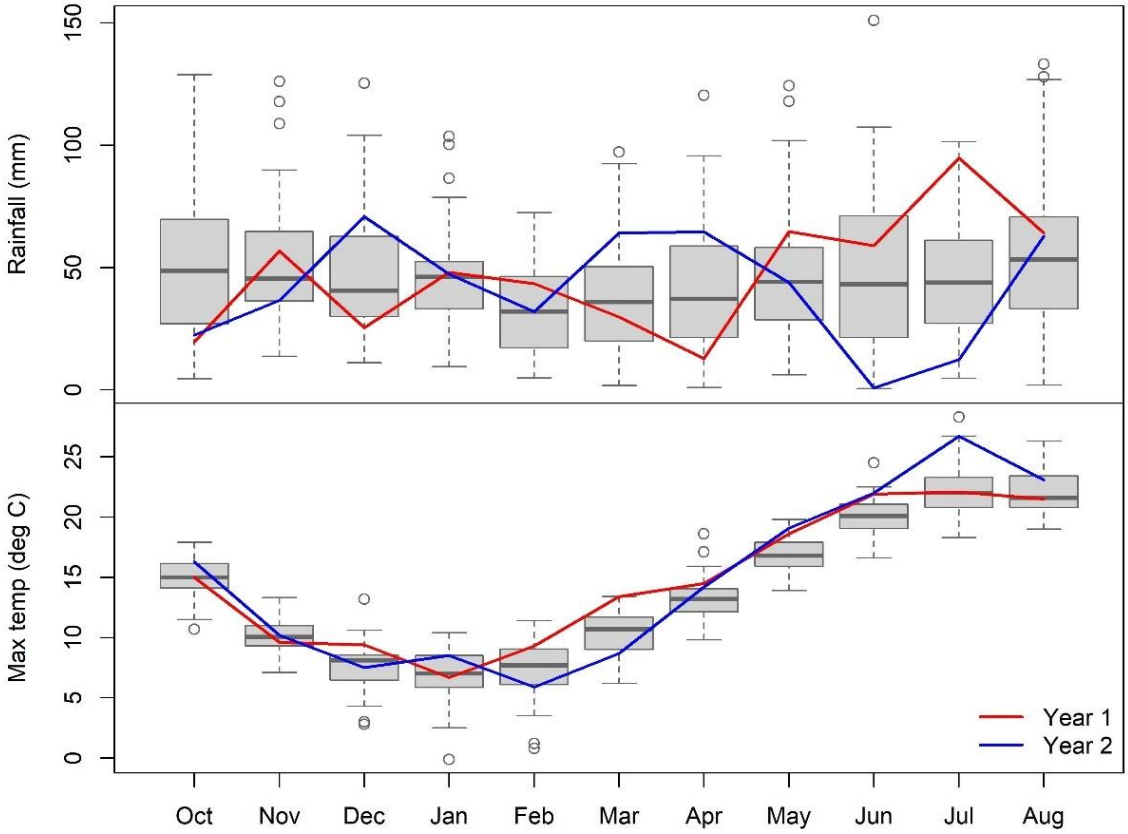


Figure 2. Contrasting patterns of monthly rainfall and maximum temperature over the growing season for the two trial years. Boxplots indicate historic variation in data for each month from 1960 to 2016 (horizontal line = median, boxes = interquartile range, whiskers = 1.5 times the interquartile range and points = values outside 1.5 times the interquartile range).

Other plant architecture traits, such as green leaf area (GLA) in the development phase, juvenile growth habit and flag leaf morphology had weak but positive correlations with yield, suggesting potential relevance of these traits as mechanisms to increase yield, or as valuable traits themselves to select for in combination with yield to increase crop competitive ability with weeds. However, the strong positive correlations between GLA traits and plant height traits mean that increasing these traits without increasing lodging risk may be problematic. Optimising combinations of important traits therefore requires consideration of correlated responses due to pleiotropy and linkage.

3.2 Genomic prediction of complex traits

We tested the accuracy of several genomic prediction approaches to determine the genetic architecture of the multiple related traits, using both single-trait (ST) and multi-trait (MT) models that take into account relationships among correlated traits. LASSO represents models including trait genetic architecture controlled by a minimal number of additive genetic effects across the genome, while Random Forest (RF) represents models including a much greater number of interacting genetic effects. RF outperformed LASSO for most traits in ST models and was particularly advantageous for traits with generally low genomic prediction accuracy, such as GLA and grain yield in both years (**Figure 3a**). Prediction accuracy was increased from 0.34 and 0.20 in ST LASSO models to

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

0.38 and 0.27 in ST RF models for grain yield measured in each year respectively. In contrast, traits with greater prediction accuracy, such as plant height or grain dimension traits, were better predicted by LASSO. This suggests that RF can successfully predict genetic effects in traits with more complex genetic architecture, potentially using non-additive and epistatic effects.

Multi-trait ensemble (MT ens) models consistently produced more accurate predictions for almost all traits for both RF (86% of traits) and particularly for LASSO (90% of traits) models (**Figure 3b-c**). On the other hand, single vector decomposition (SVD) MT models were less reliable and generally resulted in lower prediction accuracies (**Figure 3b-c),** although RF notably outperformed LASSO in SVD MT models. Therefore, including information from predictions of other related traits in a multi-trait ensemble model was advantageous over attempting to predict pleiotropic effects directly in a decomposed trait matrix. Traits that were poorly predicted by LASSO compared to RF for ST models (**Figure 3a**, left panel) had particularly increased prediction accuracies via the MT ensemble predictions for LASSO (**Figure 3b**, left panel). This suggests traits that were predicted better by either of the multi-trait models have few direct or large genetic effects and are rather the culmination of many other component traits. These indirect genetic effects may be picked up in complex RF prediction models, but are best captured by the use of MT models that can directly model pleiotropic trait trade-offs

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

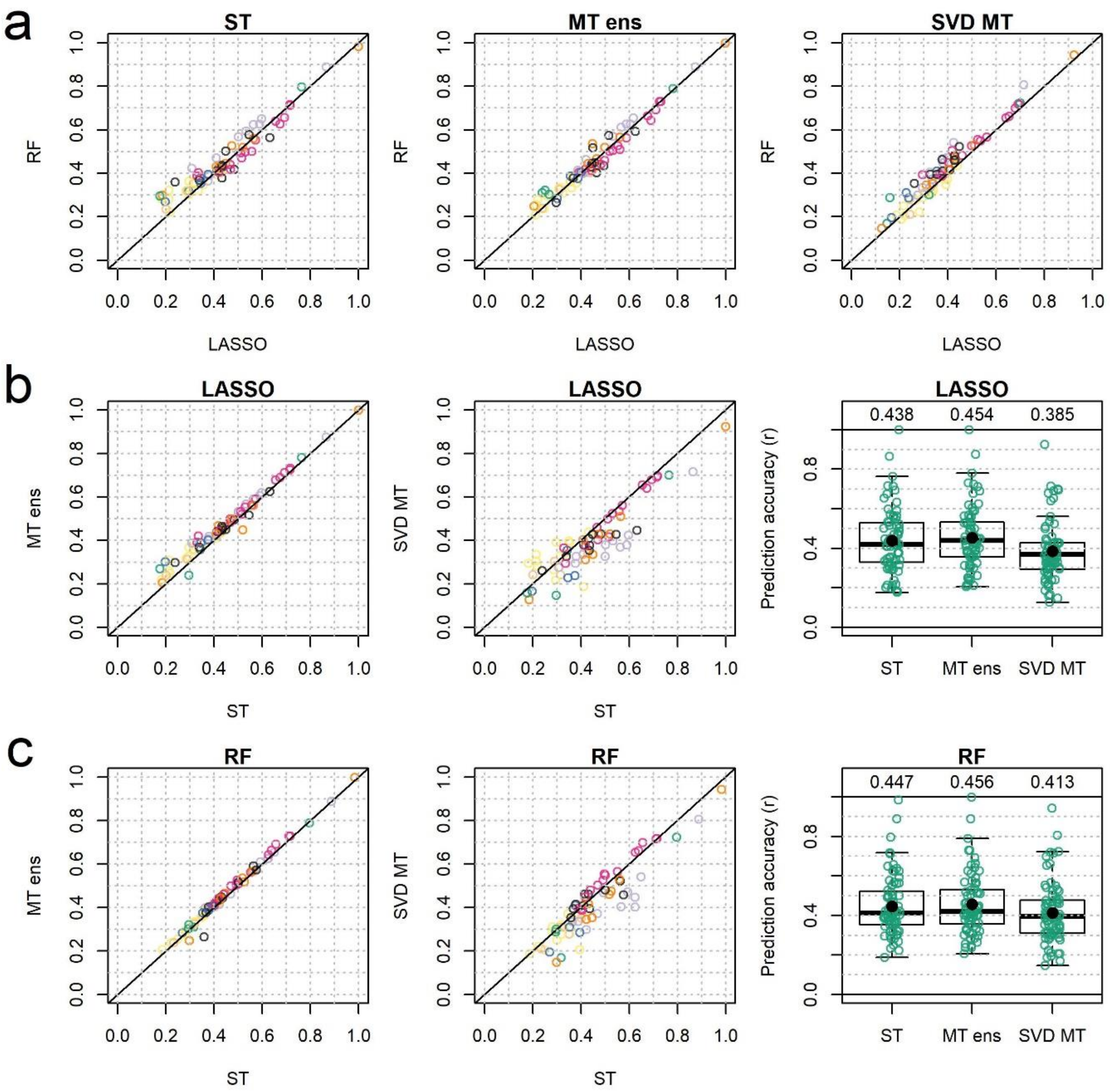


Figure 3. Comparison of genomic prediction accuracies for all traits. Each circle represents a trait-year combination, with circles colour coded according to trait group, as in **Figure 1**) between LASSO and Random Forest (RF) models using single-trait (ST), multi-trait ensemble (MT ens) and Single Vector Decomposition (SVD) approaches. Row (a) compares LASSO and RF prediction models, row (b) compares ST with both MT approaches for LASSO prediction models, and row (c) compares ST with both MT approaches for RF prediction models. Horizontal lines in boxplots represent the median and black dots represent the mean prediction accuracy across all traits, which is also shown above each boxplot.

MT ensemble models increased genomic prediction accuracy of grain yield from an average of 0.27 to 0.33 for LASSO models and from 0.32 to 0.34 for RF models on average across both years and cross-validations. Variable importance of trait covariates in the Random Forests used in the MT ensemble models indicates the influence of these traits in the model. Across all traits, most highly influential traits had strong positive or negative correlations in each year among the observed lines (**Figure 4a**). Considering grain yield in each year as the primary trait of interest, highly correlated traits such as GPC and grain yield measured in the other year, were highly important in MT ensemble models for grain yield, suggesting that pleiotropic effects mediating the grain yield

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and protein content trade-off are useful for predicting grain yield itself (**Figure 4b**). Base model predictions of yield in the other year as the focal yield trait were also included in models with high importance suggesting that the ensemble model effectively takes yield G×E effects into account. Developmental stage traits including dates of growth stages GS39 (flag leaf blade all visible), GS55 (ear half emerged) and GS65 (flowering half-way complete) and flag leaf senescence (FLS) were particularly important covariate traits and were positively correlated with yield in year 1 to a greater extent than in year 2 (**Figure 1**). These traits also featured with greater importance in MT ensemble models when grain yield in year 1, compared to year 2, was predicted (**Figure 4b**). This indicates that later-developing lines were predicted to be higher yielding in the year without terminal drought stress. Other yield component traits including grains per spikelet (GPS) and grains per ear (GPE), but not grain size traits, were found to be of high importance (**Figure 4b**). Many GLA traits, particularly when measured in the spring, were also included with fairly high importance in MT ensemble models (**Figure 4b**), suggesting a role of the crop development phase in resource acquisition for final grain yield. These results not only identify important traits for inclusion in multi-trait prediction models, but also physiological mechanisms for grain yield improvement.

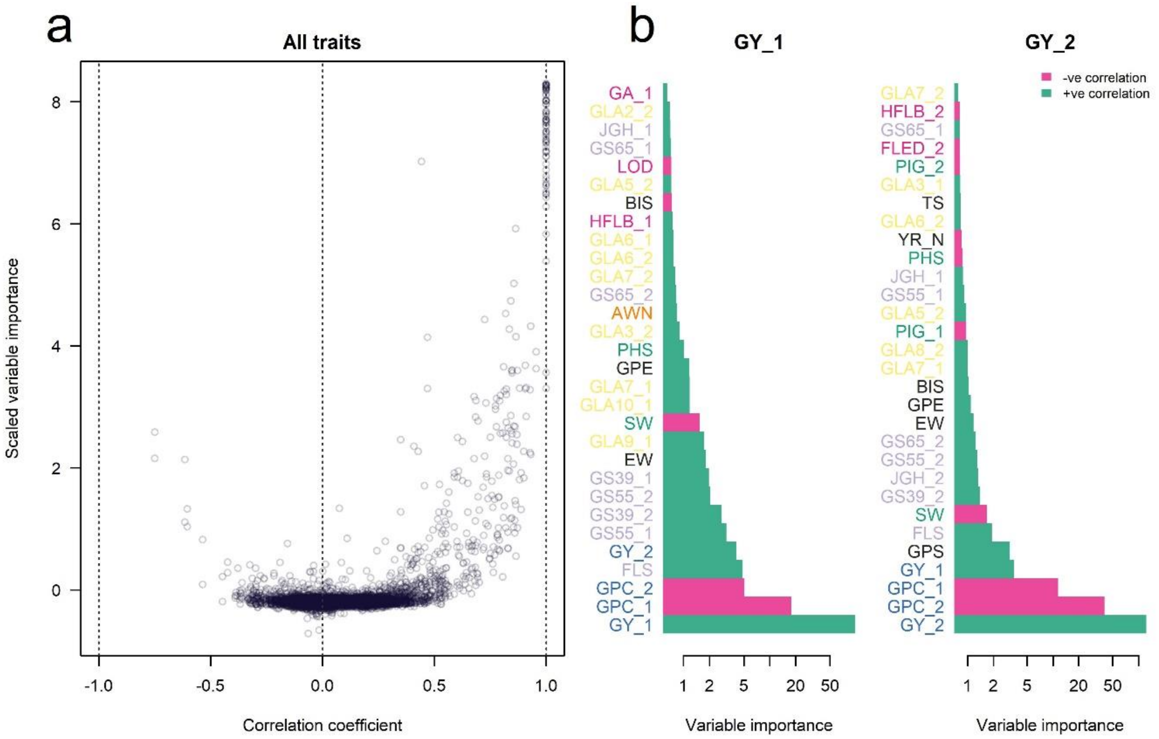


Figure 4. The Influence of related traits in multi-trait ensemble prediction models. (**a**) The relationship between pairwise correlation coefficients among all traits and years and the variable importance score in Random Forest multi-trait ensemble models for all target traits. (**b**) The 30 most important trait variables used in Random Forest multi-trait ensemble models for prediction of grain yield in each year (GY_1 and GY_2). All trait abbreviations are as listed in **Table 1** and colour coded according to trait group, as in **Figure 1**. The _1 and _2 designations used after trait abbreviations refer to trial year 1 and trial year 2, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

3.3 Accurate genomic prediction models increase long-term genetic gain in simulation of recurrent genomic selection

We then investigated the potential for different genomic prediction models to achieve genetic gain in yield through simulation of a recurrent genomic selection programme within the NDM. MT ensemble RF models were found to be the most accurate model from cross validation within the observed population (**Figure 3**) and so were used as the true genetic model to define the true phenotypes of simulated lines. The genomic prediction models were then trained on the simulated true phenotype data of lines in the first generation and genomic predictions of phenotypes were used to make selections for subsequent cycles of lines.

The accuracy of genomic prediction models over the course of the selection simulations generally reflected those in cross validated of the observed data. Prediction accuracy of all models decreased in later cycles of the simulations, but models that were more accurate in the observed data and maintained accuracy for longer, such as MT RF and ST RF (**Figure 5a**), achieved greater long-term genetic gain in yield (**Figure 5b**). RF models that included restricted trees with an interaction depth of one so that genetic marker interaction effects could not be included were much less accurate and led to less genetic gain, particularly for grain yield in year 1 (**Figure 5**). This suggests an important role for prediction of non-additive genetic effects in RF models for continued accuracy of genomic prediction models, particularly as breeding cycles become more distantly related to the training set.

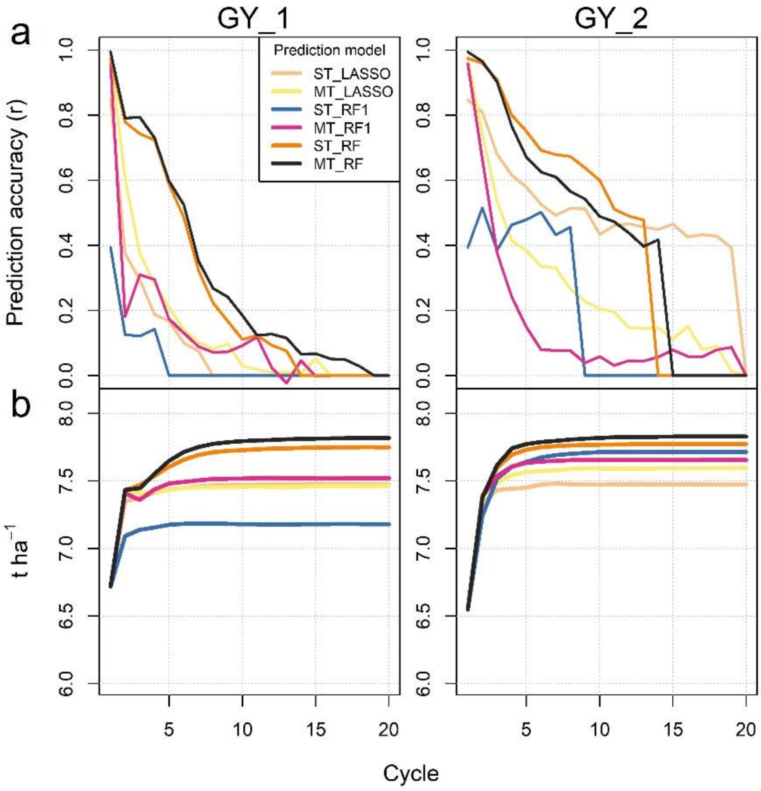


Figure 5. Trends in accuracy of (a) genomic prediction models and (b) resulting genetic gain over a simulated recurrent genomic selection programme for grain yield (GY) measured in two years trial years. Genomic prediction models include single- (ST) and multi-trait (MT) ensembles models for LASSO and Random Forest (RF). RF1 indicates Random Forest models with a restricted interaction depth of one. Lines represent the averages across 20 simulation repeats.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

3.4 Simulation of recurrent phenotypic selection for grain yield reveals indirect effects on multiple traits

Phenotypic correlations identified traits that may be under similar genetic control through either pleiotropy or linkage, so we investigated the potential for recurrent selection to achieve genetic gains in traits directly under selection as well as indirect effects on other traits. Simulations of a phenotypic recurrent selection programme were run that compared different selection indices and true genetic models. For these, selections were made based on the true phenotypic values rather than the predicted phenotypes as for simulation of recurrent genomic selection.

Firstly, we simulated selection based purely on grain yield measured in each trial year. Considering the MT ensemble RF as the true genetic model, which achieved the greatest prediction accuracy across traits (**Figure 3**)**,** selection on grain yield *per se* resulted in rapid genetic gain in the yield trait under direct selection as well as indirect effects on other related traits (**Figure 6**). These included selection for combinations of related traits that were complementary to, as well as those that were antagonistic to, their correlation in the unselected population. As an example of complementary trait selection, grain protein content (GPC) was shown to be strongly negatively correlated with grain yield in the original population (**Figure 1**) and therefore rapidly decreased as grain yield was selected for. Although both grain yield and GPC are both positively valued traits, here we define these as under complementary selection were their trait correlations and selection covariance are in the same direction. In contrast, antagonistic trait selection could be demonstrated by plant height traits and grain dimension traits that were all positively correlated with each other in the original population (**Figure 1**), but covaried negatively over time in the simulated population under selection for yield measured in each trial year; large grain size traits increased over time (GA, GL, GW, TGW), whereas plant height traits (HET, HFLB) decreased (**Figure 6**). Similarly, green leaf area (GLA) traits over the foundation development stage were generally positively correlated with plant height traits but increased over time as grain yield was selected for, while plant height traits decreased (**Figure 6**). Although the majority of trait relationships had complementary rather than antagonistic trait correlations in the original population and covariances in the simulated population (55.4 and 57.8% of pairwise relationship when grain yield selected in each year respectively had both positive or both negative correlations and covariances under selection; **Figure 6**), the significant remaining proportion did not. This indicates that antagonistic trait trade-offs were required to be optimised to achieve the genetic gains in yield simulated in the population and highlights the benefit of the multi-trait prediction approach.

As suggested by the low correlation between yields in each year, selection for yield in either year had only limited effects on yield in the other year where approximately half the genetic gain in yield in the alternate year was achieved in simulated selection for yield in either year. This G×E effect for yield was reflected by how the yield component traits were co-selected with yield between the two years. Grains per ear (GPE) and grains per spikelet (GPS) were increased when selection was for grain yield in year 2 but remained mostly neutral for grain yield in year 1 (**Figure 6**). Further to the differential importance of traits in the multi-trait ensemble models outlined above, differential selection responses of yield component traits according to yield in differing environments highlights the capacity for G×E interactions to buffer response to selection for grain yield. However, when G×E is predictable, in certain target environments, contrasting yield component strategies could be used to adapt the crop to the environment. For example, it may be supposed that a genotype that can be high yielding by

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

producing many grain sites per ear throughout an extended development phase due to being later flowering would be better adapted to environments without terminal drought stress.

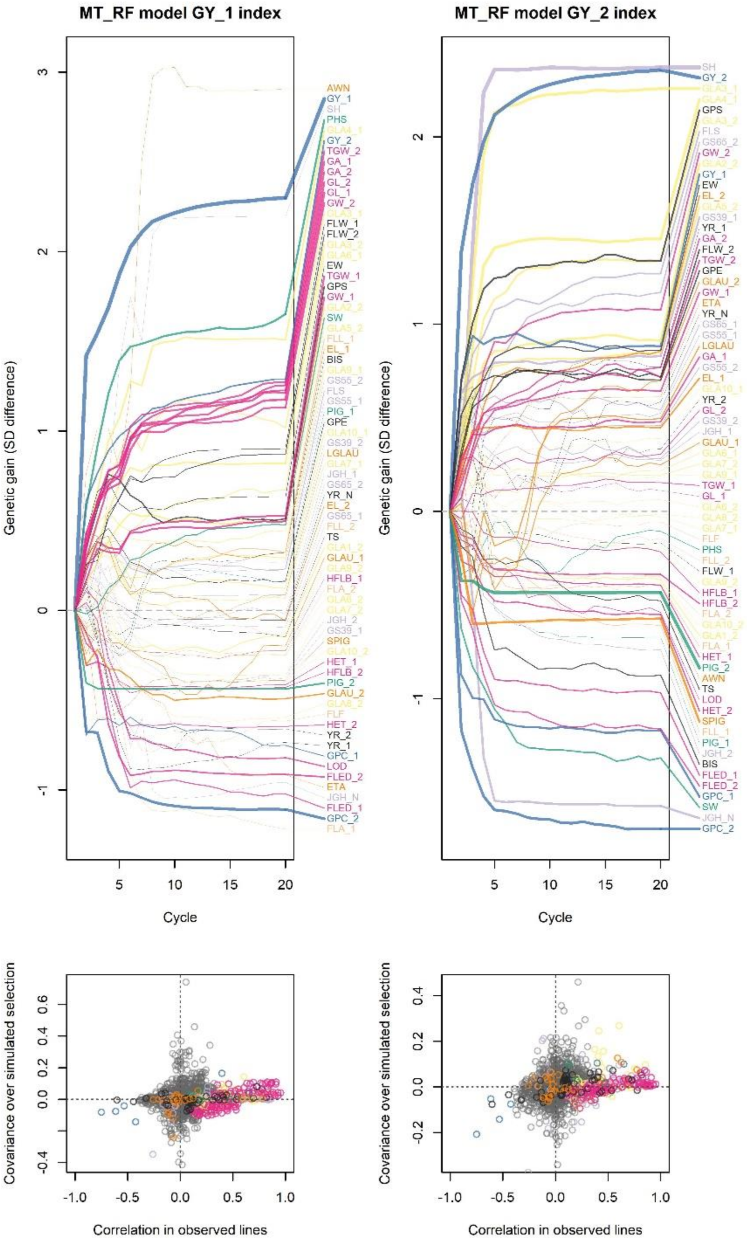


Figure 6. **Simulated phenotypic response to selection on grain yield measured in two years.** Upper plots indicate response of all 72 traits under simulated recurrent selection for grain yield in each of the two trial years based on multi-trait Random Forest (MT RF) genomic prediction models. Genetic gain was calculated as difference in population mean trait values to generation one and scaled to the standard deviation of the trait values in generation one. Line colours relate to trait groups identified by hierarchical clustering shown in **Figure 1**. The _1 and _2 designations used after trait abbreviations refer to trial year 1 and trial year 2, respectively. Line widths are relative to the t-test significance of each trait genetic gain from cycle 0 to 20 across all 20 simulation repeats. Lower plots compare the correlations between all pairs of traits in the original population and the covariance between trait pairs over time in the simulated population under selection. Points in upper right or lower left quadrants indicate both positive or negative correlation and covariance which demonstrates complementary trait selection. Points in the lower right or upper left quadrants represent differing positive or negative correlation and covariance, suggesting antagonistic trait selection. Point colours indicate pairs of traits that are both in the same trait group following the colour scheme in **Figure 1**, while grey points indicate trait pairs from different groups.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

3.5 Selection indices enable optimisation of trait trade-offs

We next tested whether multi-trait selection indices could be employed to simultaneously optimise selection for yield and other traits of interest, such as grain protein content (GPC) or crop architecture traits that aid competition with weeds. As outlined by the observed trait correlations, early season green leaf area (GLA) traits and grain yield were found to be slightly positively correlated, so could be co-selected, but the additional association between GLA and plant height would need to be restricted to limit risk of lodging. We therefore simulated effects of a phenotypic selection strategy based on selection index to increase important traits for crop competitive ability with weeds in combination with grain yield, whilst restricting changes in plant height as well as an index for high grain yield protein deviation to combine both negatively correlated traits.

Considering the ST RF as the true genetic model, the combined grain yield + weed competition selection index succeeded in increasing desirable competitive traits including GLA, flag leaf area (FLA) as well as grain yield in both years, whilst maintaining plant height at an acceptable level (**Figure 7**). Indirect effects on other traits included rapid early selection for spring-type growth habit (SH) up to the fifth breeding cycle, but which then remained at around 90% frequency in the population without fixation in any of the of the simulation repetitions. The GYPD selection index also achieved genetic gain in desirable traits (grain yield and protein content in both years) and had some indirect effects on related traits (**Figure 7**). As an example of one trait that was co-selected with GYPD, flag leaf width (FLW) increased in all simulation repeats, increasing by 3% and 4.2% for the trait when measured in year 1 and year 2, respectively. Whilst most of the trait relationships selected for in the weed competition index were positively correlated and complementary traits, such as all of the GLA traits and grain yield, the GYPD selection index included more antagonistic trait relationships (positive correlation and negative covariance or negative correlation and positive covariance) that were required to be optimised in addition to yield and protein trade off (**Figure 7**). For example, under GYPD selection, grains per spikelet (GPS) correlated negatively with GPC in each year (correlation = -0.33 and -0.36 in each year respectively), but covaried positively over simulated selection (covariance = 0.52 and 0.35 in each year respectively), where GPS increased by an average of 0.23 over the course of simulated selection while GPC measured in each year also increased by 2.48% and 1.57% respectively. Furthermore, the flag leaf to ear distance (FLED) and GPC measured in year 2 correlated positively (correlation = 0.21), but covaried negatively over the simulated selection (covariance = -0.25), where FLED decreased by 2.73cm while GPC increased by 1.57% over the course of simulated selection. These provide examples of trait mechanisms by which yield and GPC could be simultaneously selected.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

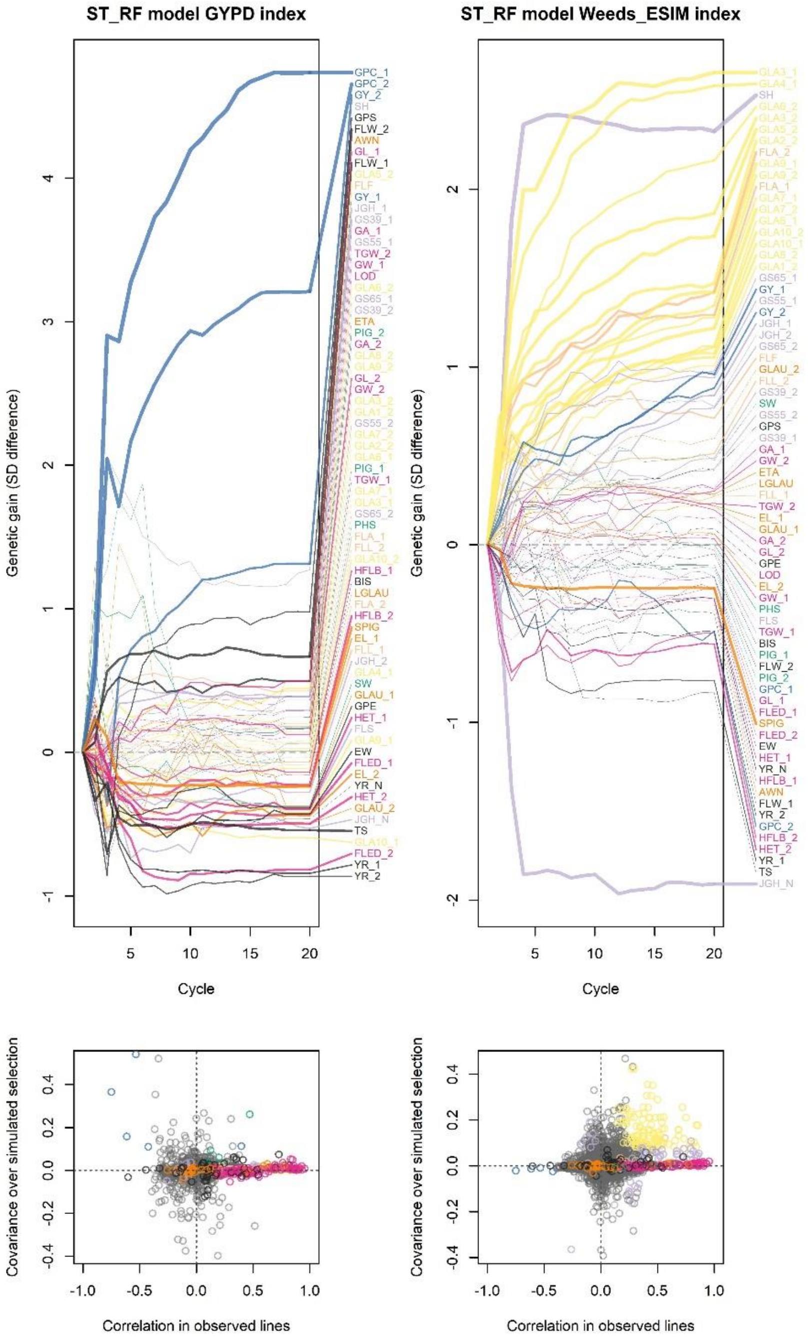


Figure 7. Simulated phenotypic response to selection on two selection indices. Upper plots indicate response of all traits under simulated recurrent selection for two multi-trait selection indices based on single-trait Random Forest (ST RF) genomic prediction models. GYPD = selection to increase both grain yield and protein (grain yield protein deviation); Weeds_ESIM = selection to increase yield as well as weed competitive traits whilst limiting change in plant height. Line colours relate to trait groups identified by hierarchical clustering and correlations shown in Figure 1. Line widths are relative to the t-test significance of each trait genetic gain from cycle generation 0 to 20 across all 20 simulation repeats. Lower plots compare the correlations between all pairs of traits in the original population and the covariance between trait pairs over time in the simulated population under selection. Points in upper right or lower left quadrants indicate both positive or negative correlation and covariance which demonstrates, suggesting complementary trait selection. Points in the lower right or upper left quadrants represent differing positive or negative correlation and covariance which demonstrates, suggesting antagonistic trait selection. Point colours indicate pairs of traits that are both in the same trait group following the colour scheme in Figure 1, while grey points indicate trait pairs from different groups.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Selection on multiple traits, which requires optimisation of multiple trait trade-offs, slowed the rate of genetic gain in grain yield for each of the selection indices: mean grain yield across both trial years increased when using both selection indices, but at a slower rate, particularly for grain yield in year 1 under GYPD selection, compared to when grain yield was selected for *per se* (**Figure 8**). However, in comparison to gain in grain yield in either one of the two years when selection was for grain yield in the other year, both GYPD and the yield + weed competition selection indices achieved generally comparable gains for yield whilst also increasing other favourable traits (**Figure 8**). These results show that antagonistic trait relationships are generally possible to optimise through appropriate selection. However, while this may slow genetic gain to some extent in the primary traits of interest, such as grain yield, this more realistically represents the balance of selection for multiple traits that occurs in breeding programmes.

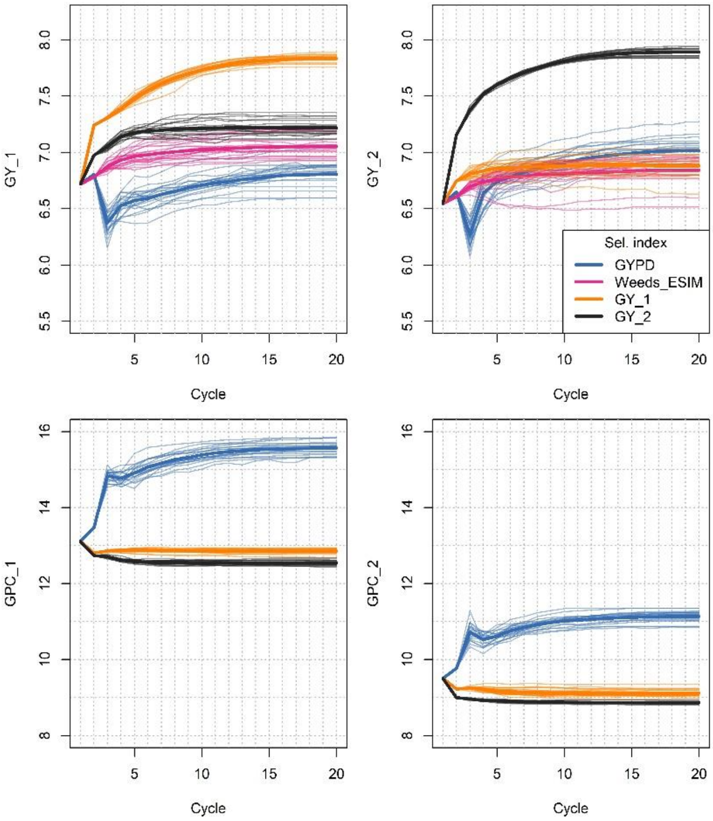


Figure 8. **Simulated response to selection in grain yield (GY) measured in two trial years (_1 = year1, _2 = year2) under different selection indices based on single-trait Random Forest (ST RF) genomic prediction models**. Narrow lines represent each of 20 simulation repeats while thicker lines represent the mean across all simulation repeats. GYPD = selection to increase both grain yield and protein (grain yield protein deviation); Weeds_ESIM = selection to increase yield as well as weed competitive traits whilst limiting change in plant height.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

3.6 Different true genetic models affect long-term response to simulated selection

After comparing simulated response to different selection indices, we then tested how using different true genetic models that are based on the different genomic prediction models trained on the observed NDM population (LASSO versus RF; using either ST or MT approaches) affect phenotypic and genomic response to selection. Both RF and MT ensemble models were shown above to generally increase the prediction accuracy across traits (**Figure 3**), and here we show these predictions had a lower degree of shrinkage towards the mean of predicted trait values in comparison to the grain yield trait values of observed lines (**Figure 9a**). The MT RF models in fact had comparable variances in prediction values to the observed grain yield data, indicating their realistic prediction of phenotypic variation. While this would be expected from an overfitting model, cross validation with the observed data showed an increased accuracy of these models.

Simulations with MT RF true genetic models tended to have the largest and longest increase of genetic gain over the course of simulated selection of grain yield (**Figure 9b**). Genetic gain in grain yield plateaued, at a relatively low level, after only around six cycles of selection using ST LASSO genetic model simulations but both cycle time to plateaux and plateaux height (maximum genetic gain) were both extended by either using a RF or MT model. This pattern of faster and higher genetic gain in RF or MT models was accompanied by the retention of higher phenotypic (**Figure 9c**) and genetic (**Figure 9d**) variance, particularly over long-term selection in the MT models. Almost all non-zero LASSO SNP effects were fixed after 8 cycles of selection in any simulation repeat for selection for grain yield in either year, limiting further genetic gain (**Figure 9**). Continued loss of genetic diversity once all genetic effects that affect phenotypic variance were fixed was down to genetic drift. Many of the SNP with highest variable importance in RF models were in common with the largest LASSO SNP effect coefficients, and the largest of these were fixed in the first few cycles of selection at a similar rate for both ST RF and ST LASSO (**Figure 9e**), where almost all of the ten SNPs with the largest LASSO effect or RF variable importance were fixed after five cycles of selection for both models. However, RF models included many more SNPs with non-zero importance (~20,000) than non-zero LASSO effects (61 and 87 for grain yield in years 1 and 2, respectively) and many more of these small or non-additive genetic effects in RF models remained polymorphic for longer (**Figure 9e**). For example, a significant proportion of these (14.3 and 13.8% for grain yield in years 1 and 2, respectively) remained polymorphic after 10 cycles of selection while genetic gain in yield still continued to increase (**Figure 9b**). This suggests that accumulation of the SNP effects, that were too small to be included in LASSO, or complex non-additive SNP by SNP epistatic genetic effects, made a large contribution to continued long-term genetic gain in RF models even after large effect QTL are fixed.

Furthermore, simply adding a MT second step to LASSO models to include indirect pleiotropic effects also increased and extended long-term genetic gain to a similar or greater extent to ST or MT RF models (**Figure 9b**). Using MT models, LASSO SNP effects were fixed at a much slower rate (**Figure 9e**) and phenotypic and genetic variance was maintained for much longer (**Figure 9c-d**), where on average 12% of the ten largest LASSO SNP effects for each single trait were polymorphic after 10 cycles of selection for across all simulation repeats with selection for grain yield in both years. This also suggests that the greater degree of pleiotropy present in MT models, which increased prediction accuracy for low accuracy LASSO models in particular (**Figure 3**), meant that the number of small effect loci involved in each trait was greatly increased. However, the number of indirect

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

pleiotropic LASSO SNP effects across non-additive ensemble models could not be quantified. Selection could therefore act on more complex trait relationships driven by pleiotropy and/or linkage.

Linkage among antagonistic genetic effects could be shown to partly limit genetic gain. On average, only 0.8 and 5% of non-zero ST LASSO model SNP coefficients were negatively fixed resulting in an average of 0.28 and 2.15% loss of the maximum yield after 20 cycles of selection for yield in each year respectively. However, this was exacerbated in MT LASSO genetic models where 16.4 and 28.7% of ST LASSO SNP effects with negative effects on the trait under selection, were incorrectly fixed resulting in 14.1 and 23.5% loss of genetic gain. This further indicates insufficient recombination to completely decouple antagonistic linked QTL that were not directly involved in ST LASSO models for grain yield directly but pleiotropically linked through MT models.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

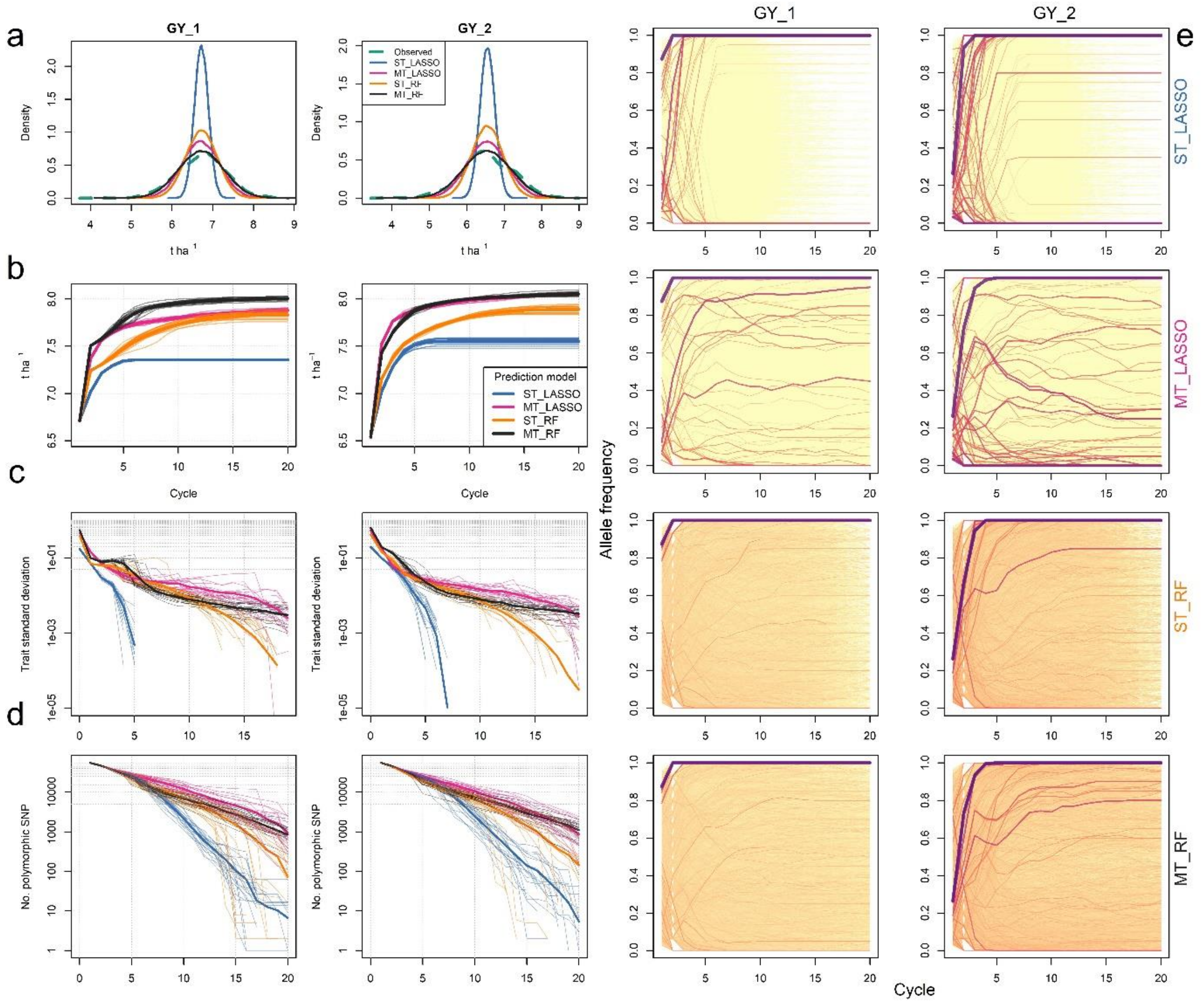


Figure 9. Simulated phenotypic and genetic response to selection with different genetic models. (**a**) Comparisons among distributions of observed and predicted trait values for two grain yield (GY) scenarios for different prediction models. LASSO and Random Forest (RF) genomic prediction models in conjunction with single- (ST) or multi-trait (MT) ensemble models are compared. (**b**) Rates of genetic gain in GY in each GY scenario when GY is directly selected under 20 cycles of simulated recurrent selection comparing different genomic prediction models. Comparisons in the rate of reductions in (**c**) phenotypic and (**d**) genetic variation in the NIAB Diverse MAGIC (NDM) population under recurrent selection comparing the same prediction models as above, and colour coded in the same way. Narrow lines represent each of 20 simulation repeats while thicker lines represent the mean across all simulation repeats. (**e**) Changes in mean allele frequency for all ~55,000 SNP markers across 20 simulation repeats for the NDM population under simulated selection for grain yield in two yield scenarios using the four models (ST LASSO, ST RF, MT LASSO, MT RF). Line widths and colour are proportional to SNP effect size in LASSO models and variable importance score for RF models.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

4. Discussion

A complex structure of trait relationships that interact with environmental conditions were found to be involved in prediction of grain yield. Through simulation of recurrent selection within a genetically diverse highly recombined multi-founder wheat population, and based on observed genomic and phenotypic data, we tested several contrasting genetic models and quantitative genetic approaches to recurrent selection. We found that, in comparison to a simplifying LASSO genetic model where each trait was predicted directly from a minimal subset of markers with additive effects, prediction accuracies were increased both by using a multi-trait ensemble approach and Random Forest prediction models, which potentially incorporate pleiotropy and epistatic effects respectively. This was particularly so for complex traits with low prediction accuracy, and in simulations of recurrent selection these models also increased the rate and extent of long-term genetic gain, whilst maintaining phenotypic and genetic variance. Thus, genomic prediction models that include more complex genetic effects such as epistasis, and pleiotropy may better reflect how continued genetic gain is achieved through breeding.

4.1 The value of multi-trait models

We showed that modelling relationships among traits is valuable for increasing genomic prediction accuracy. Traditional multi-trait genomic prediction models consider the covariance structure of related traits across multiple environments and replicates and increase genomic prediction accuracy for cross-validation schemes when test fractions include partially phenotyped individuals in the test environment (Jia and Jannink, 2012). However, other studies often do not find an advantage to multi-trait models for untested genotypes in real datasets (Bhatta et al., 2020; Ward et al., 2019). We present results from multi-trait ensembles that integrate predictions of multiple traits into the same model (Van der Laan et al., 2007; He et al., 2016). These ensemble models consistently outperformed single trait models, while a contrasting approach using single vector decomposition of the multi-trait matrix performed poorly and more variably across traits. Although the increase in prediction accuracy was small for most traits, the advantage of multi-trait ensemble models was particularly great for traits that were poorly predicted by LASSO models, suggesting that ensemble methods efficiently incorporate additional information from large numbers of small pleiotropic genetic effects among related traits, which ST LASSO models would otherwise overlook when each trait is considered independently. Traits such as grain yield are polygenic and few genetic markers with large and consistent effects have been identified and applied in breeding (Bernardo, 2016). However, predictions of component traits of yield, many of which have simpler genetic architectures (Scott et al., 2021), can improve the ensemble prediction model for yield. We found that many highly correlated traits were used as covariates with high importance in multi-trait models. Furthermore, the covariate importance scores of traits in the ensemble models highlight physiological mechanisms for trait improvement and enable optimisation of antagonistic trait relationships (**Figure 4**). Where yield components correlate negatively with each other, the multi-trait ensemble model is able to optimise the interplay among these traits to increase the prediction accuracy of yield as the primary trait of interest. Similar to the approach taken by Powell et al. (2022) who modelled multiple systems biology development processes to bridge the gap between genotype to complex phenotype, we used multiple physiological traits in more agnostic models without defined crop growth parameters to aid prediction of the complex processes behind grain yield.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Inter-year environmental variation can modulate relationships between traits. We noted strongly contrasting weather conditions between the two trial years in which phenotypic data was collected (**Figure 2**). The covariate importances of traits for predicting yield changed with the year scenario being predicted, revealing some mechanisms controlling G×E for yield (**Figure 4**). For example, growth stage phenotypes were more important covariates in year 1. Similarly, trade-offs in plant size and earliness likely maintain polygenic trait variation due to varying environmental pressures in the wild plant *Mimulus guttatus* (Troth et al., 2018). In breeding, any single strategy to achieve high yield may be hampered by unpredictable year-to-year environmental variation, and thus limit response to selection and reduction in genetic variance. While our simulations of future genetic gain cannot account for unmeasured environments in the future, commercial wheat breeders often take this into account and make selections of promising lines with a diversity of phenological or plant height traits to ensure adaptive potential.

4.2 The potential to optimise trait trade-offs that conventional breeding has neglected

Using multi-trait data from a MAGIC population that controls for confounding effects of population structure (Scott et al., 2020), we found that pleiotropy and/or tight genetic linkage are significant causes of correlated trait responses to selection. These data also shed light on the combination of traits that would be required to be co-selected or optimised to achieve continuous gains in grain yield as a primary trait under selection. Furthermore, we find antagonistic trade-offs among traits that have been problematic for wheat crop improvement. We suggest that historic enhancement of grain yield by breeders at the cost of key traits such as weed competitive ability, or grain protein content, has been due to the over-riding value placed on grain yield as a primary selection criterion during variety testing, as well as market pressures, leaving little scope for compromise with other traits. Integration of novel trait variation to optimise these trade-offs within an elite wheat genepool, which has been under such strong directional selection, would therefore be difficult. However, simulations presented here show that with appropriate selection indices, genetic gain in both yield and other valuable but negatively correlated traits was possible to some extent.

4.3 The value of complex genomic prediction models for continued genetic gain

Simulating selection using prediction models with contrasting genetic architectures (i.e., in terms of additive or epistatic and direct or pleiotropic genetic effects) had major impacts on the outcomes of recurrent selection. The greater complexity of these models both increased cross-validated prediction accuracy of complex traits in the observed population and extended the accuracy of genomic predictions in simulations of recurrent genomic selection. Furthermore, simulations of phenotypic selection assuming a complex genetic model demonstrated accelerated and extended potential for genetic gain while maintaining genetic and phenotypic variance. The role of non-additive genetic effects has been demonstrated elsewhere to preserve genetic variance over long-term selection in simulated populations (Wientjes et al., 2021). Wang et al. (2004) used simulations of selection within

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the CIMMYT wheat breeding programme to compare genetic models, finding that inclusion of epistasis in genetic models greatly reduced the rate that additive genetic variance is lost due to selection. Although the role of epistasis is thought to contribute little to overall genetic variance, at least in outbred populations (Hill and Mäki-Tanila, 2015), evolutionary theory supports these observations in crop breeding; selection can enable conversion of epistatic to additive genetic effects, allowing hidden or cryptic genetic variation to then be unlocked (Carlborg et al., 2006; Hill, 2017). The limits to trait variation in our study are likely underestimates because as allele frequencies shift and trait genetic architectures evolve under selection new additive genetic effects would be unlocked for selection which cannot be modelled or predicted in the observed population. Supporting this, we found that prediction models soon become out of date and suffer loss of prediction accuracy, particularly for simple additive prediction models (LASSO), when target genotypes become more distantly related to the training set (Edwards et al., 2019). However, in realistic scenarios of a wheat breeding programme practicing genomic selection, the training model is continually updated with data from advanced breeding line testing which would enable more linear continued genetic gain. Continuous novel mutations may also play an important role in regenerating genetic variation and extending limits to long-term selection in large populations (Hill, 1982) but were not considered in simulations reported here. Pre-breeding programmes can also introduce novel genetic diversity from the primary, secondary and tertiary wheat genepool (Balfourier et al., 2019). Nevertheless, the population we study is representative of diverse north-west European wheats across 70 years. We found that additive variation included in minimal LASSO prediction models was quickly depleted during simulated selection. We propose that pleiotropic and epistatic genetic effects and G×E interactions have played a major role in maintaining wheat genetic diversity despite strong selection and will be particularly important for applied genomic selection of elite varieties in already highly selected breeding populations.

4.4 Potential for applied crop breeding

MAGIC populations have proven valuable resources for direct generation of commercial varieties of some less intensively bred crops than wheat (Scott et al., 2020). In simulated breeding programmes, Bernardo (2021) suggested that multi-parent crossing schemes may be valuable for maintaining genetic diversity. However, the diverse MAGIC wheat population described here is unlikely to generate commercially competitive varieties due to the broad genetic basis and historic founders. Instead, this MAGIC population samples and recombines genetic diversity across 70 years and can therefore be considered a microcosm of past and future selective breeding. In this context, our simulations rerun alternate histories to test different selection models and approaches and reveal physiological and genetic mechanisms for future breeding. We suggest that this approach, including multi-trait ensembles, could be further integrated with environmental information to inform crop models (Cooper et al., 2021; Stöckle and Kemanian, 2020). Considering that traditional wheat breeding programme cycles generally extend over at least five years, our simulations of twenty cycles of recurrent selection represent an equivalent of over one hundred years of traditional wheat breeding (albeit it with no further input from genotypes outside of the 16 founders from which the MAGIC population was constructed). Current wheat breeding programmes are also likely to be at a point towards the later stages of selection simulations presented here where the majority of large effect QTL are either fixed or well accounted for. Further genetic gain in current breeding programmes will therefore likely be achieved through optimisation of small and complex genetic effects (Gorjanc et al., 2018).

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Through selection, breeders appear to often maintain linkage disequilibrium across unexpectedly large genomic regions because they contain several beneficial alleles (Brinton et al., 2020; Fradgley et al., 2019), which could also interact epistatically. Therefore, prediction models that capture relevant haplotype blocks and the most recently unlocked epistatic effects will be increasingly important for forward prediction of high performing breeding lines. Comparisons between patterns of linkage disequilibrium in commercial selected varieties and in simulations presented here would validate this process and uncover valuable sites for further marker assisted selection. Our simulations suggest that beneficial variation is often lost during breeding. This diverse MAGIC population is a reservoir for such this genetic diversity and the phenotypic and genomic data allow beneficial alleles to be identified. Through targeted rapid recurrent selection and breeding technologies that reduce generation time (Watson et al., 2018; Cha et al., 2021), this population could therefore be used to provide useful pre-breeding material for commercial breeding programmes to deliver accelerated and continued genetic gain. Optimum contribution selection can also be applied to maximise long-term genetic gain by maintaining genetic variance for selection at later generations.

4.5 Conclusions

In summary, we demonstrated the value of multi-trait ensemble models for genomic prediction of complex traits and simulated recurrent selection using these genetic models based empirically on an extensively genotyped and phenotyped NDM population. We consider this a microcosm of wider wheat breeding programmes working with the wider pool of wheat germplasm so that our results provide insights into the trends and mechanisms by which the considerable progress and genetic gain has been made in modern wheat breeding without apparent genetic diversity loss. These findings highlight the importance of models and approaches that take into account these mechanisms to maximize further genetic gain in the future.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

5. Acknowledgements

This work was funded by Biotechnology and Biological Sciences Research Council (BBSRC) grants BB/M011666/1 awarded to JC, BB/M011585/1 to RM and BB/M011194/1 as part of the PhD research of NF. We also thank Dr Stephanie Swarbreck for useful comments on the manuscript.

6. Author contributions

NF conceived of and undertook the analyses. JC and RM were awarded project funding. NF wrote the first draft of the manuscript and all authors reviewed, edited, and approved the final manuscript.

7. Data availability

Phenotypic and genotypic data used in this study were as presented by Scott et al. (2021) and can be freely accessed at http://mtweb.cs.ucl.ac.uk/mus/www/MAGICdiverse/.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

8. References

Andrew, I.K.S., Storkey, J. and Sparkes, D.L., 2015. A review of the potential for competitive cereal cultivars as a tool in integrated weed management. *Weed Research*, 55(3), pp.239-248.

Austin, R.B., Bingham, J., Blackwell, R.D., Evans, L.T., Ford, M.A., Morgan, C.L. and Taylor, M., 1980. Genetic improvements in winter wheat yields since 1900 and associated physiological changes. *The Journal of Agricultural Science*, 94(3), pp.675-689.

Balfourier, F., Bouchet, S., Robert, S., De Oliveira, R., Rimbert, H., Kitt, J., Choulet, F., International Wheat Genome Sequencing Consortium, BreedWheat Consortium and Paux, E., 2019. Worldwide phylogeography and history of wheat genetic diversity. *Science advances*, 5(5), p.eaav0536.

Bernardo, R., 2016. Bandwagons I, too, have known. Theoretical and Applied Genetics, 129(12), pp.2323-2332.

Bernardo, R., 2021. Multiparental populations in line development: Genetic gain, diversity, and practical limitations. *Crop Science*, 61(6), pp.4139-4150.

Bhatta, M., Gutierrez, L., Cammarota, L., Cardozo, F., Germán, S., Gómez-Guerrero, B., Pardo, M.F., Lanaro, V., Sayas, M. and Castro, A.J., 2020. Multi-trait genomic prediction model increased the predictive ability for agronomic and malting quality traits in barley (*Hordeum vulgare* L.). *G3: Genes, Genomes, Genetics*, 10(3), pp.1113-1124.

Bogard, M., Allard, V., Brancourt-Hulmel, M., Heumez, E., Machet, J.M., Jeuffroy, M.H., Gate, P., Martre, P. and Le Gouis, J., 2010. Deviation from the grain protein concentration–grain yield negative relationship is highly correlated to post-anthesis N uptake in winter wheat. *Journal of Experimental Botany*, 61(15), pp.4303-4312.

Breiman, L., 2001. Random Forests. *Machine Learning*, 45(1), pp.5-32.

Brinton, J., Simmonds, J., Minter, F., Leverington-Waite, M., Snape, J. and Uauy, C., 2017. Increased pericarp cell length underlies a major quantitative trait locus for grain weight in hexaploid wheat. *New Phytologist*, 215(3), pp.1026-1038.

Brinton, J., Ramirez-Gonzalez, R.H., Simmonds, J., Wingen, L., Orford, S., Griffiths, S., Haberer, G., Spannagl, M., Walkowiak, S., Pozniak, C. and Uauy, C., 2020. A haplotype-led approach to increase the precision of wheat breeding. *Communications Biology*, 3(1), pp.1-11.

Broman, K.W., Gatti, D.M., Simecek, P., Furlotte, N.A., Prins, P., Sen, Ś., Yandell, B.S. and Churchill, G.A., 2019. R/qtl2: software for mapping quantitative trait loci with high-dimensional data and multiparent populations. *Genetics*, 211(2), pp.495-502.

Buhlmann, D.J.S.P., 2011. MissForest-nonparametric missing value imputation for mixed-type data. *Bioinformatics*, 28, pp.113-118.

Carlborg, Ö. and Haley, C.S., 2004. Epistasis: too often neglected in complex trait studies? *Nature Reviews Genetics*, 5(8), pp.618-625.

Carlborg, Ö., Jacobsson, L., Åhgren, P., Siegel, P. and Andersson, L., 2006. Epistasis and the release of genetic variation during long-term selection. *Nature Genetics*, 38(4), pp.418-420.

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. and Kim, K.M., 2022. Speed vernalization to accelerate generation advance in winter cereal crops. *Molecular Plant*, 15(8), pp 1300-1309.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Cerón-Rojas, J.J., Sahagún-Castellanos, J., Castillo-González, F., Santacruz-Varela, A. and Crossa, J., 2008. A restricted selection index method based on eigenanalysis. *Journal of Agricultural, Biological, and Environmental Statistics*, 13(4), pp.440-457.

Charmet, G., Tran, L.G., Auzanneau, J., Rincent, R. and Bouchet, S., 2020. BWGS: AR package for genomic selection and its application to a wheat breeding programme. *PloS One*, 15(4), p.e0222733.

Clark, S.A., Hickey, J.M. and Van der Werf, J.H., 2011. Different models of genetic variation and their effect on genomic evaluation. *Genetics Selection Evolution*, 43(1), pp.1-9.

Cooper, M., Powell, O., Voss-Fels, K.P., Messina, C.D., Gho, C., Podlich, D.W., Technow, F., Chapman, S.C., Beveridge, C.A., Ortiz-Barrientos, D. and Hammer, G.L., 2021. Modelling selection response in plant-breeding programs using crop models as mechanistic gene-to-phenotype (CGM-G2P) multi-trait link functions. *in silico Plants*, 3(1), p.diaa016.

Corsi, B., Obinu, L., Zanella, C.M., Cutrupi, S., Day, R., et al., 2021. Analysis of a German multi-parental population identifies eight genetic loci controlling two or more yield components in wheat, including the genetic loci *Rht24*, *WAPO-A1* and *WAPO-B1* and multi-trait QTL on chromosomes 5A and 6A. *Theoretical and Applied Genetics*, 134, pp.1435-1454.

Davis, D.R., 2009. Declining fruit and vegetable nutrient composition: What is the evidence? *Horticultural Science*, 44(1):15-19. https://doi.org/10.21273/HORTSCI.44.1.15

Dudley, J.W., 2007. From means to QTL: The Illinois long-term selection experiment as a case study in quantitative genetics. *Crop Science*, 47, pp.S20-S31.

Edwards, S.M., Buntjer, J.B., Jackson, R., Bentley, A.R., Lage, J., Byrne, E., Burt, C., Jack, P., Berry, S., Flatman, E. and Poupard, B., 2019. The effects of training population design on genomic prediction accuracy in wheat. *Theoretical and Applied Genetics*, 132(7), pp.1943-1952.

Fernandes, S.B., Dias, K.O., Ferreira, D.F. and Brown, P.J., 2018. Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. *Theoretical and Applied Genetics*, 131(3), pp.747-755.

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

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

Fischer, R.A. and Kertesz, Z., 1976. Harvest index in spaced populations and grain weight in microplots as indicators of yielding ability in spring wheat 1. *Crop Science*, 16(1), pp.55-59.

Fu, Y.B., 2015. Understanding crop genetic diversity under modern plant breeding. *Theoretical and Applied Genetics*, 128(11), pp.2131-2142.

Fufa, H., Baenziger, P.S., Beecher, B.S., Graybosch, R.A., Eskridge, K.M. and 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), pp.187-198.

Goddard, M.E., Kemper, K.E., MacLeod, I.M., Chamberlain, A.J. and Hayes, B.J., 2016. Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture. *Proceedings of the Royal Society B: Biological Sciences*, 283(1835), p.20160569.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Gorjanc, G., Gaynor, R.C. and Hickey, J.M., 2018. Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. *Theoretical and applied genetics*, 131(9), pp.1953-1966.

Guttieri, M.J., Baenziger, P.S., Frels, K., Carver, B., Arnall, B. and Waters, B.M., 2015. Variation for grain mineral concentration in a diversity panel of current and historical Great Plains hard winter wheat germplasm. *Crop Science*, 55(3), pp.1035-1052.

Hawkesford, M.J., 2014. Reducing the reliance on nitrogen fertilizer for wheat production. *Journal of Cereal Science*, 59(3), pp.276-283.

He, D., Kuhn, D. and Parida, L., 2016. Novel applications of multitask learning and multiple output regression to multiple genetic trait prediction. *Bioinformatics*, 32(12), pp.i37-i43.

Hill, W.G., 1982. Predictions of response to artificial selection from new mutations. *Genetics Research*, 40(3), pp.255-278.

Hill, W.G., 2017. “Conversion” of epistatic into additive genetic variance in finite populations and possible impact on long-term selection response. *Journal of Animal Breeding and Genetics*, 134(3), pp.196-201.

Hill, W.G. and Mäki-Tanila, A., 2015. Expected influence of linkage disequilibrium on genetic variance caused by dominance and epistasis on quantitative traits. *Journal of animal breeding and genetics*, 132(2), pp.176-186.

International Wheat Genome Sequencing Consortium (IWGSC), Appels, R., Eversole, K., Stein, N., Feuillet, C., Keller, B., Rogers, J., Pozniak, C.J., Choulet, F., Distelfeld, A. and Poland, J., 2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, 361(6403), p.eaar7191.

Jannink, J.L., Lorenz, A. and Iwata, H., 2010. Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics*, 9, pp.166–177.

Jia, Y. and Jannink, J.L., 2012. Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics*, 192(4), pp.1513-1522.

Jiang, Y. and Reif, J.C., 2015. Modeling epistasis in genomic selection. Genetics, 201(2), pp.759-768.

Kissing Kucek, L., Mallory, E.B., Darby, H.M., Dawson, J.C. and Sorrells, M.E., 2021. Breeding wheat for weed-competitive ability: I. Correlated traits. *Euphytica*, 217(11), pp.1-15.

Korres, N. E., and Froud-Williams, R. J. 2002. Effects of winter wheat cultivars and seed rate on the biological characteristics of naturally occurring weed flora. *Weed Research*, 42(6), pp.417-428.

Kuzay, S., Xu, Y., Zhang, J., Katz, A., Pearce, S., Su, Z., Fraser, M., Anderson, J.A., Brown-Guedira, G., DeWitt, N. and Haugrud, A.P., 2019. Identification of a candidate gene for a QTL for spikelet number per spike on wheat chromosome arm 7AL by high-resolution genetic mapping. *Theoretical and Applied Genetics*, 132(9), pp.2689-2705.

Lovegrove, A., Pellny, T.K., Hassall, K.L., Plummer, A., Wood, A., Bellisai, A., Przewieslik-Allen, A., Burridge, A.J., Ward, J.L. and Shewry, P.R., 2020. Historical changes in the contents and compositions of fibre components and polar metabolites in white wheat flour. *Scientific Reports*, 10(1), pp.1-9.

Mackay, I.J., Cockram, J., Howell, P., Powell, W. 2021. Understanding the classics: the unifying concepts of transgressive segregation, inbreeding depression and heterosis and their central relevance for crop breeding. *Plant Biotechnology Journal*, 19, pp.26-34.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Mackay, I., Horwell, A., Garner, J., White, J., McKee, J., Philpott, H., 2011. Reanalysis 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), pp.225-238.

McCaig, T.N. and DePauw, R.M., 1995. Breeding hard red spring wheat in western Canada: Historical trends in yield and related variables. *Canadian Journal of Plant Science*, 75(2), pp.387-393.

Meuwissen, T., Hayes, B.J. and Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157, pp.1819–1829.

Michel, S., Löschenberger, F., Ametz, C., Pachler, B., Sparry, E. and Bürstmayr, H., 2019. Simultaneous selection for grain yield and protein content in genomics-assisted wheat breeding. *Theoretical and Applied Genetics*, 132(6), pp.1745-1760.

Montesinos-López, O.A., Montesinos-López, A., Crossa, J., Ramírez-Alcaraz, J.M., Singh, R., Mondal, S. and Juliana, P., 2019a. A singular value decomposition Bayesian multiple-trait and multiple-environment genomic model. *Heredity*, 122(4), pp.381-401.

Montesinos-López, O.A., Montesinos-López, A., Luna-Vázquez, F.J., Toledo, F.H., Pérez-Rodríguez, P., Lillemo, M. and Crossa, J., 2019b. An R package for Bayesian analysis of multi-environment and multi-trait multi-environment data for genome-based prediction. *G3: Genes, Genomes, Genetics*, 9(5), pp.1355-1369.

Muqaddasi, Q.H., Brassac, J., Koppolu, R., Plieske, J., Ganal, M.W. and Röder, M.S., 2019. *TaAPO-A1*, an ortholog of rice *ABERRANT PANICLE ORGANIZATION 1*, is associated with total spikelet number per spike in elite European hexaploid winter wheat (*Triticum aestivum* L.) varieties. *Scientific Reports*, 9(1), pp.1-12.

Murphy, K.M., Dawson, J.C. and Jones, S.S., 2008. Relationship among phenotypic growth traits, yield and weed suppression in spring wheat landraces and modern cultivars. Field Crops Research, 105(1-2), pp.107-115.

Mwendwa, J.M., Brown, W.B., Weidenhamer, J.D., Weston, P.A., Quinn, J.C., Wu, H. and Weston, L.A., 2020. Evaluation of commercial wheat cultivars for canopy architecture, early vigour, weed suppression, and yield. *Agronomy*, 10(7), p.983.

Powell, O.M., Barbier, F., Voss-Fels, K.P., Beveridge, C. and Cooper, M., 2022. Investigations into the emergent properties of gene-to-phenotype networks across cycles of selection: A case study of shoot branching in plants. *in silico Plants*, *4*(1), p.diac006.

Quintero, A., Molero, G., Reynolds, M.P. and Calderini, D.F., 2018. Trade-off between grain weight and grain number in wheat depends on G×E interaction: a case study of an elite CIMMYT panel (CIMCOG). *European Journal of Agronomy*, 92, pp.17-29.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Raherison, E., Majidi, M.M., Goessen, R., Hughes, N., Cuthbert, R., Knox, R. and Lukens, L., 2020. Evidence for the accumulation of nonsynonymous mutations and favorable pleiotropic alleles during wheat breeding. *G3: Genes, Genomes, Genetics*, 10(11), pp.4001-4011.

Ruisi, P., Frangipane, B., Amato, G., Frenda, A.S., Plaia, A., Giambalvo, D. and Saia, S., 2015. Nitrogen uptake and nitrogen fertilizer recovery in old and modern wheat genotypes grown in the presence or absence of interspecific competition. *Frontiers in Plant Science*, 6, p.185.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Sapkota, S., Boatwright, J.L., Jordan, K., Boyles, R. and Kresovich, S., 2020. Multi-trait regressor stacking increased genomic prediction accuracy of sorghum grain composition. *Agronomy*, 10(9), p.1221.

Scott, M.F., Fradgley, N.S., Bentley, A., Brabbs, T., Corke, F., Gardner, K., Horsnell, R., Howell, P., Ladejobi, O., Mackay, I.S., Mott, R. and Cockram, J. 2021. Limited haplotype diversity underlies polygenic trait architecture across 70 years of wheat breeding. *Genome Biology*, 22, p137.

Scott, M.F., Ladejobi, O., Amer, S., Bentley, A.R., Biernaskie, J., Boden, S.A., Clark, M., Dell'Acqua, M., Dixon, L.E., Filippi, C.V., Fradgley, N., Gardner, K.A., Mackay, I.J., O'Sullivan, D., Percival-Alywn, L., Roorkiwal, M., Singh, R.K., Thudi, M., Varshney, R.K., Venturini, L., Whan, A., Cockram, J. and Mott, R. 2020. Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. *Heredity*, 125(6), pp.396-416.

Schmalohr, C.L., Grossbach, J., Clément-Ziza, M. and Beyer, A., 2018. Detection of epistatic interactions with Random Forest. *bioRxiv*, p.353193.

Shewry, P.R., Pellny, T.K,, and Lovegrove, A., 2016. Is modern wheat bad for health? *Nature Plants*, 2, p.16097. https://doi.org/10.1038/nplants.2016.97

Shewry, P.R., Hassall, K.L., Grausgruber, H., Andersson, A.M., Lampi, A.M., Piironen, V., Rakszegi, M., Ward, J.L. and Lovegrove, A., 2020. Do modern types of wheat have lower quality for human health? *Nutrition Bulletin*, 45(4), pp.362-373.

Simmonds, N.W., 1995. The relation between yield and protein in cereal grain. *Journal of the Science of Food and Agriculture*, 67(3), pp.309-315.

Spyromitros-Xioufis, E., Tsoumakas, G., Groves, W. and Vlahavas, I., 2016. Multi-target regression via input space expansion: treating targets as inputs. *Machine Learning*, 104(1), pp.55-98.

Stöckle, C.O. and Kemanian, A.R., 2020. Can crop models identify critical gaps in genetics, environment, and management interactions? *Frontiers in Plant Science*, 11, p.737.

Storkey, J., Mead, A., Addy, J. and MacDonald, A.J., 2021. Agricultural intensification and climate change have increased the threat from weeds. *Global Change Biology*, 27(11), pp.2416-2425.

Tadesse, W., Sanchez-Garcia, M., Assefa S.G., Amri, A., Bishaw, Z., Ogbonnaya, F.C., Baum, M., 2019. Genetic gains in wheat breeding and its role in feeding the world. *Crop Breeding, Genetics and Genomics*, 1, e190005.

Troth, A., Puzey, J.R., Kim, R.S., Willis, J.H., Kelly, J.K. 2018. Selective trade-offs maintain alleles underpinning complex trait variation in plants. *Science,* 361(6401), pp.475-478.

Vandeleur, R.K. and Gill, G.S., 2004. The impact of plant breeding on the grain yield and competitive ability of wheat in Australia. *Australian Journal of Agricultural Research*, 55(8), pp.855-861.

van de Wouw, M., van Hintum, T., Kik, C., van Treuren, R. and Visser, B., 2010. Genetic diversity trends in twentieth century crop cultivars: a meta analysis. *Theoretical and Applied Genetics*, 120(6), pp.1241-1252.

Van der Laan, M.J., Polley, E.C. and Hubbard, A.E., 2007. Super learner. *Statistical Applications in Genetics and Molecular Biology*, 6(1), 25.

Wang, J., van Ginkel, M., Trethowan, R., Ye, G., DeLacy, I., Podlich, D. and Cooper, M., 2004. Simulating the effects of dominance and epistasis on selection response in the CIMMYT Wheat Breeding Program using QuCim. *Crop Science*, 44(6), pp.2006-2018.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.08.515457; this version posted November 9, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Ward, B.P., Brown-Guedira, G., Tyagi, P., Kolb, F.L., Van Sanford, D.A., Sneller, C.H. and Griffey, C.A., 2019. Multienvironment and multitrait genomic selection models in unbalanced early-generation wheat yield trials. *Crop Science*, 59(2), pp.491-507.

Watson, A., Ghosh, S., Williams, M.J., Cuddy, W.S., Simmonds, J., Rey, M.D., Hatta, M.A.M., Hinchliffe, A., Steed, A., Reynolds, D., Adamski, N.M., et al., 2018. Speed breeding is a powerful tool to accelerate crop research and breeding. *Nature Plants*, 4(1), pp.23-29.

Wientjes, Y.C., Bijma, P., Calus, M.P., Zwaan, B.J., Vitezica, Z.G. and van den Heuvel, J., 2022. The long-term effects of genomic selection: 1. Response to selection, additive genetic variance, and genetic architecture. *Genetics Selection Evolution*, 54(1), pp.1-21.

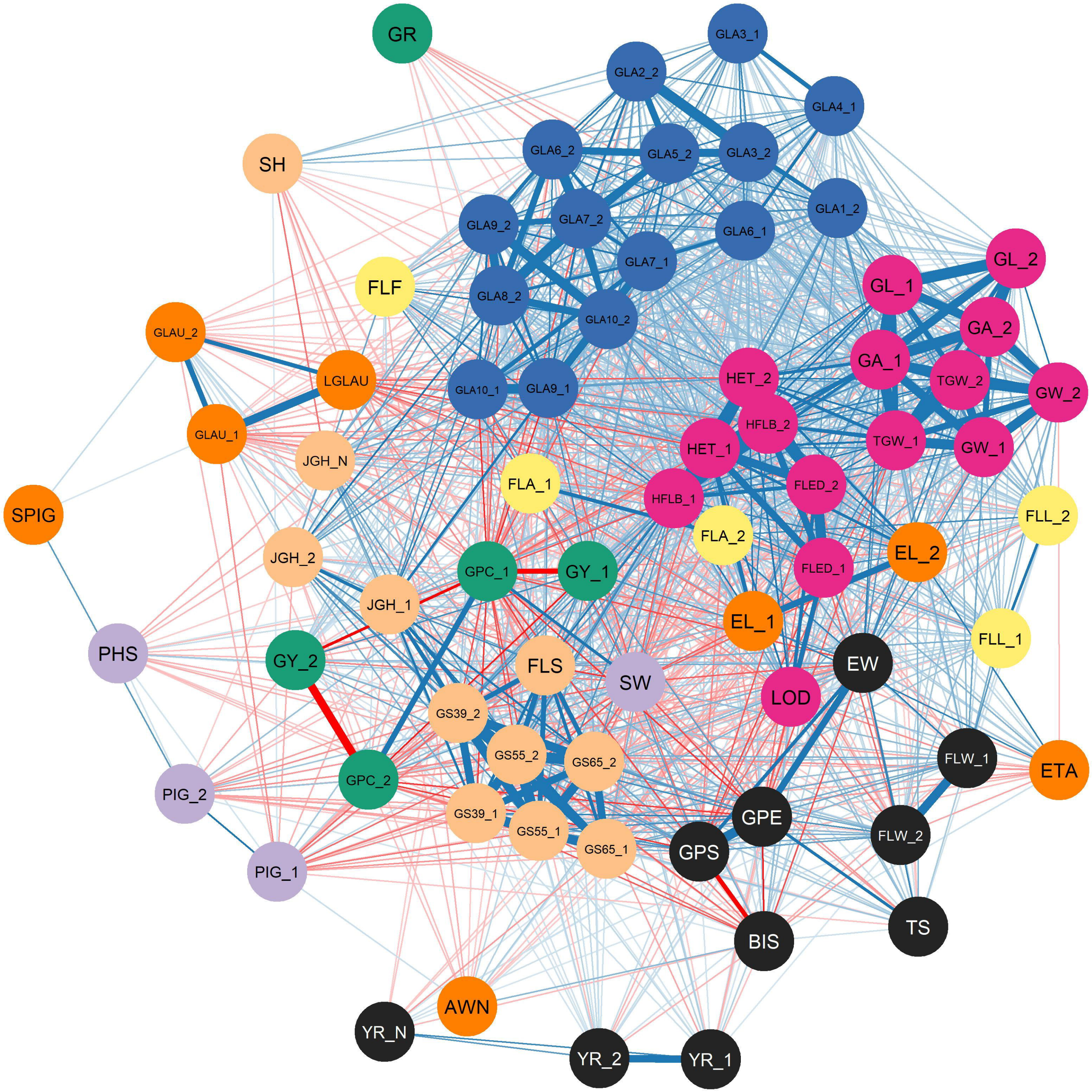
White, J., Law, J.R., MacKay, I., Chalmers, K.J., Smith, J.S.C., Kilian, A. and Powell, W., 2008. The genetic diversity of UK, US and Australian cultivars of *Triticum aestivum* measured by DArT markers and considered by genome. *Theoretical and Applied Genetics*, 116(3), pp.439-453.

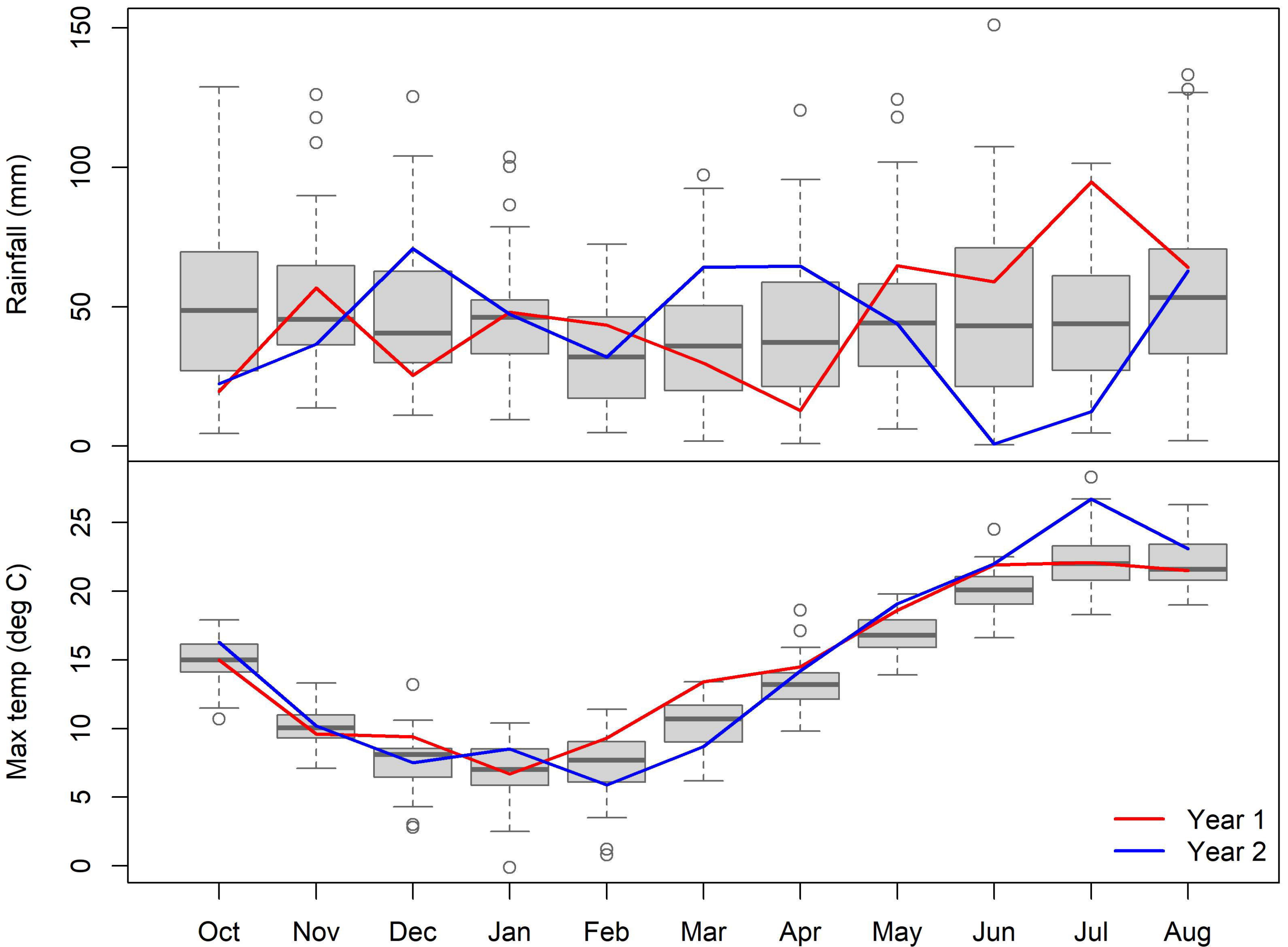
White, J., Sharma, R., Balding, D., Cockram, J. and Mackay, I.J., 2021. Genome-wide association mapping of Hagberg falling number, protein content, test weight, and grain yield in U.K. wheat. *Crop Science*, 62, pp.965-981,

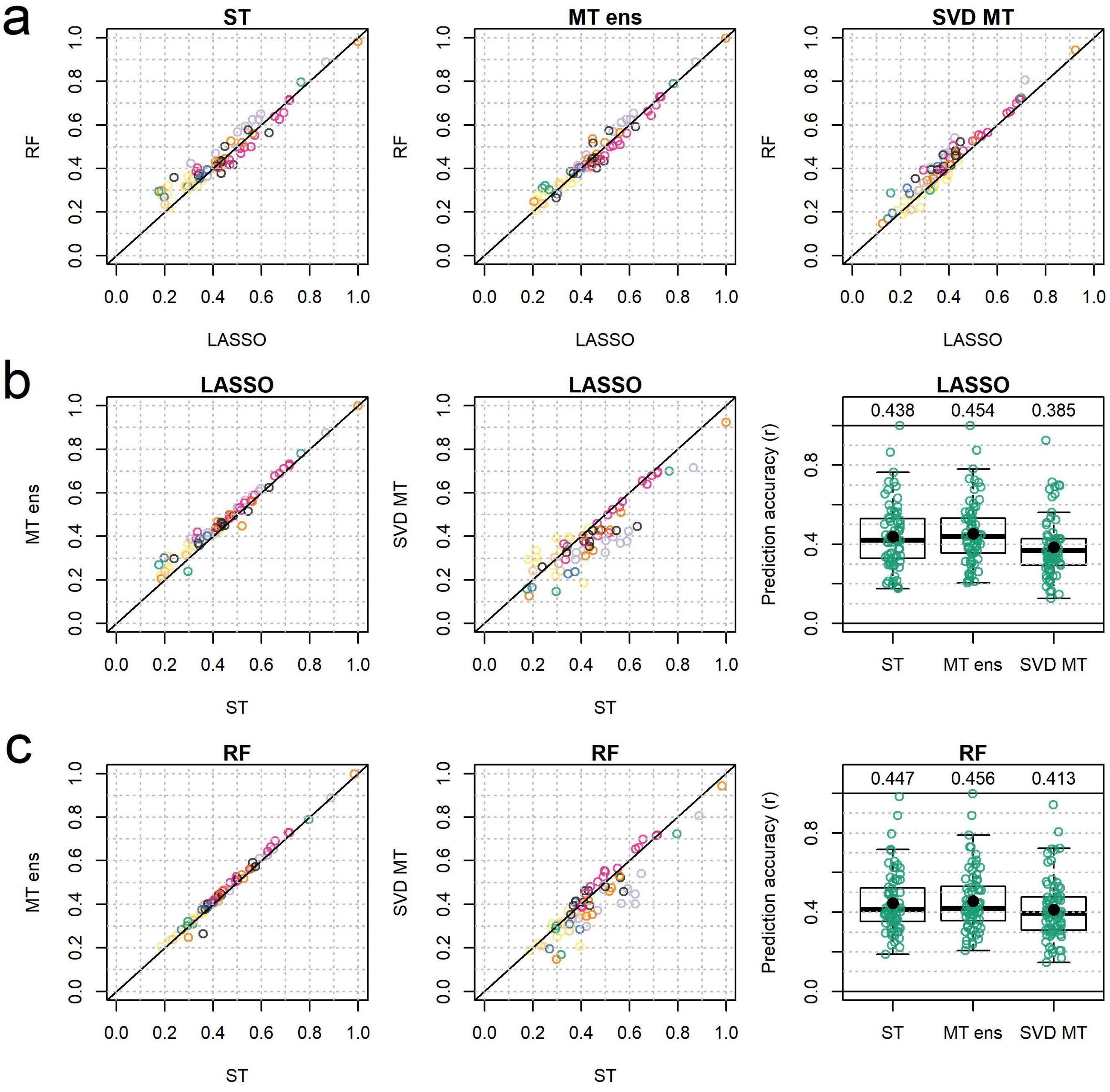
Wright, M.N., Ziegler, A. and König, I.R., 2016. Do little interactions get lost in dark random forests? BMC *Bioinformatics*, 17(1), pp.1-10.

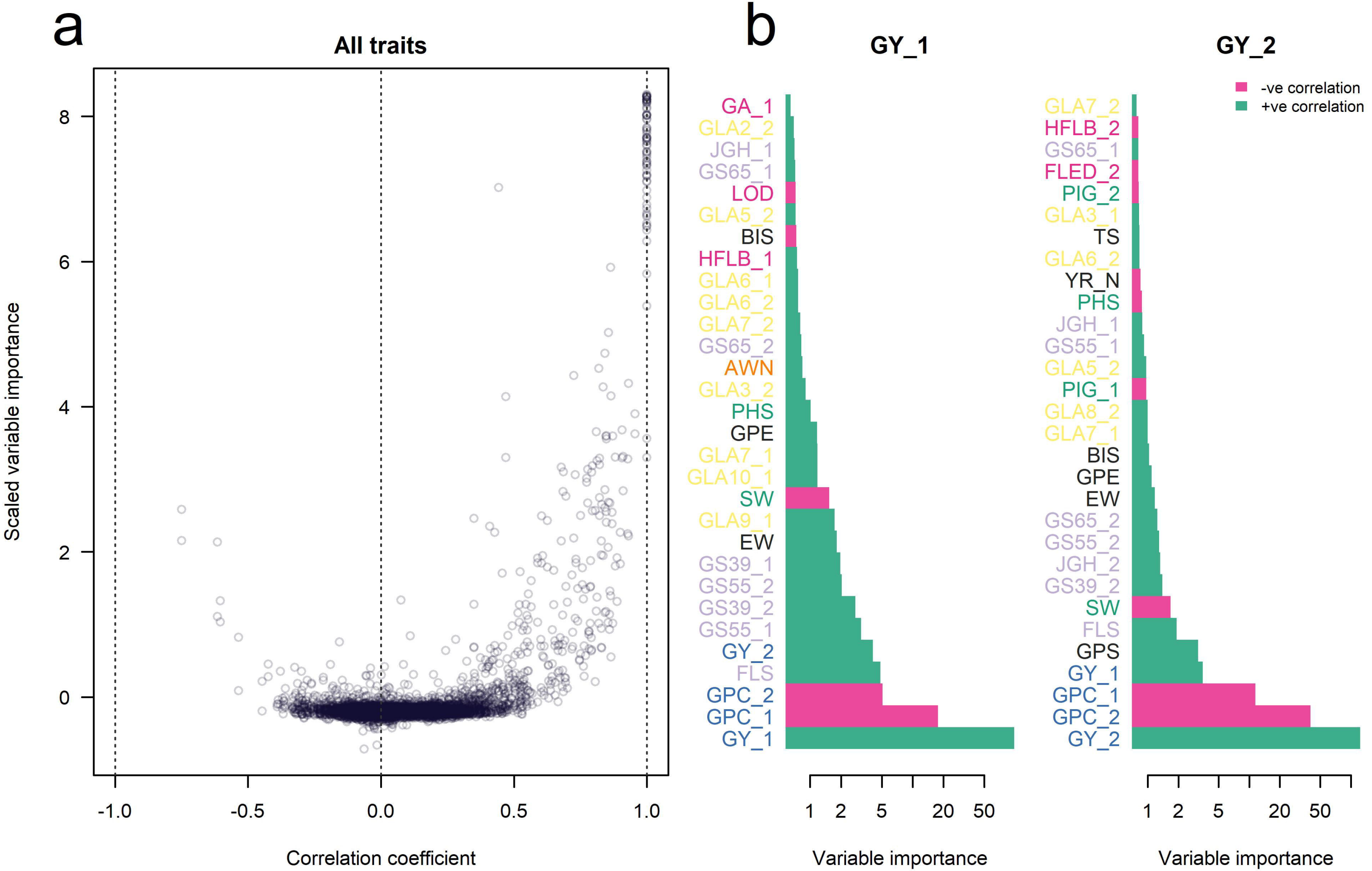
Xie, Q. and Sparkes, D.L., 2021. Dissecting the trade-off of grain number and size in wheat. *Planta*, 254(1), pp.1-16.

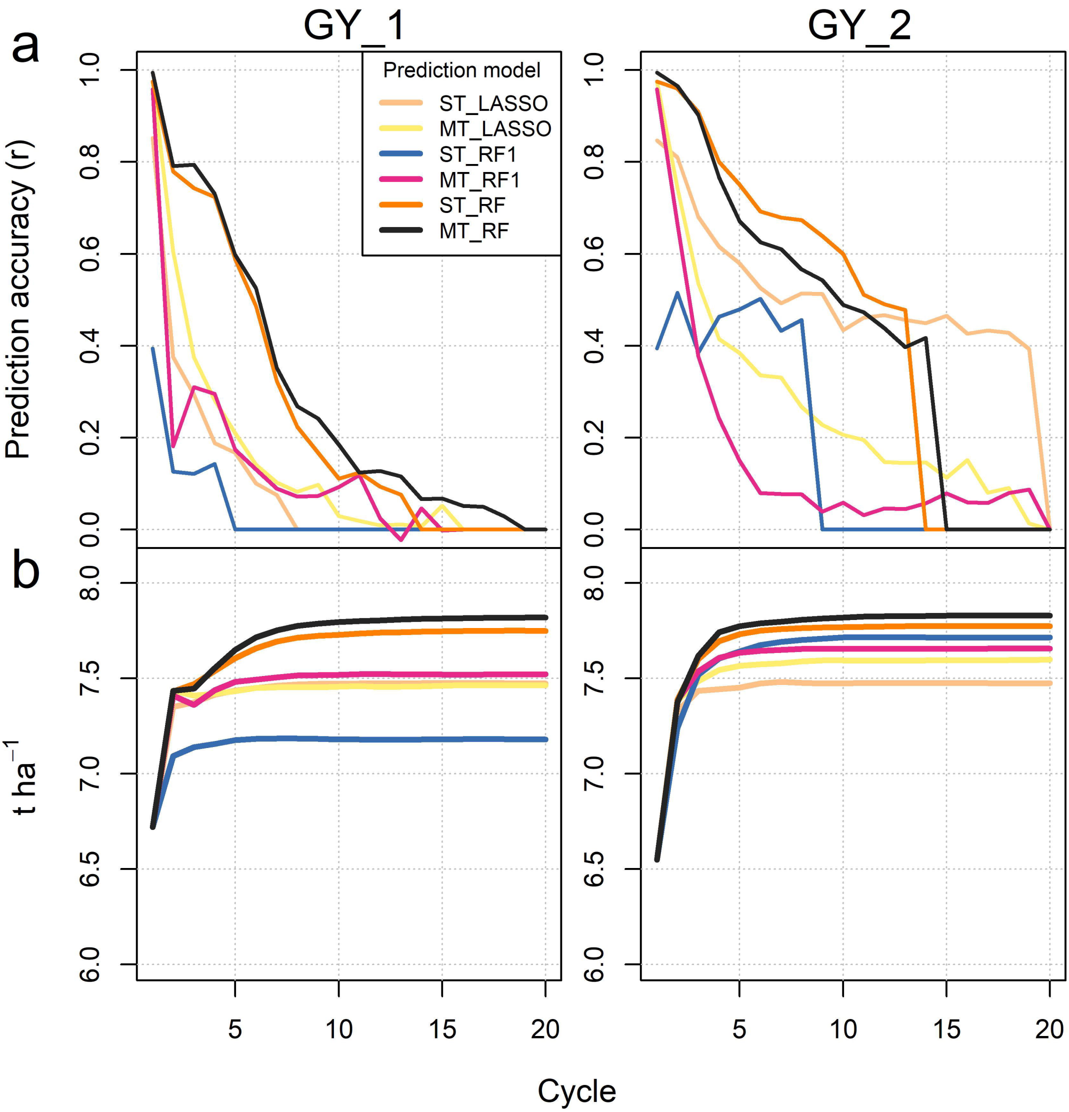
Yang, C.J., Ladejobi, O., Mott, R., Powell, W. and Mackay, I., 2022. Analysis of historical selection in winter wheat. *Theoretical and Applied Genetics*, 135, pp.3005-3023.

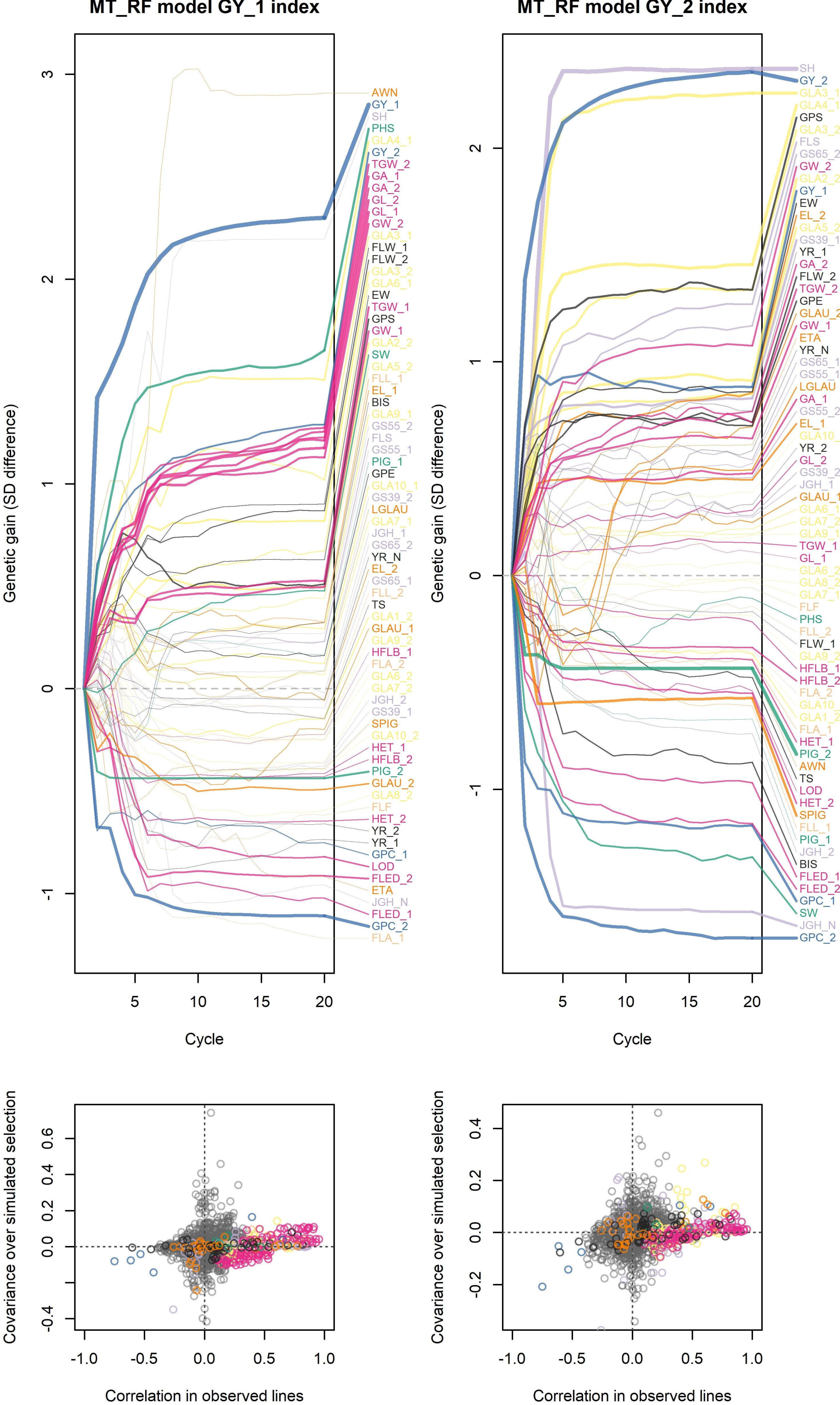

GR
GLA3_1
GLA2_2
GLA4_1
SH
GLA6_2
GLA5_2
GLA3_2
GLA1_2
GLA9_2
GLA7_2
GLA6_1
GLA7_1
GL_2
FLF
GLA8_2
GLA10_2
GL_1
GLAU_2
GA_2
GA_1
LGLAU
HET_2
TGW_2
GW_2
GLA10_1
GLA9_1
GLAU_1
HFLB_2
HET_1
TGW_1
GW_1
JGH_N
FLA_1
FLED_2
HFLB_1
SPIG
FLL_2
FLA_2
EL_2
JGH_2
GPC_1
GY_1
FLED_1
JGH_1
EL_1
FLL_1
PHS
GY_2
FLS
SW
EW
LOD
GS39_2
FLW_1
GS55_2
GS65_2
ETA
GPC_2
PIG_2
GS39_1
GPE
GS55_1
FLW_2
GPS
GS65_1
PIG_1
TS
BIS
AWN
YR_N
YR_2
YR_1

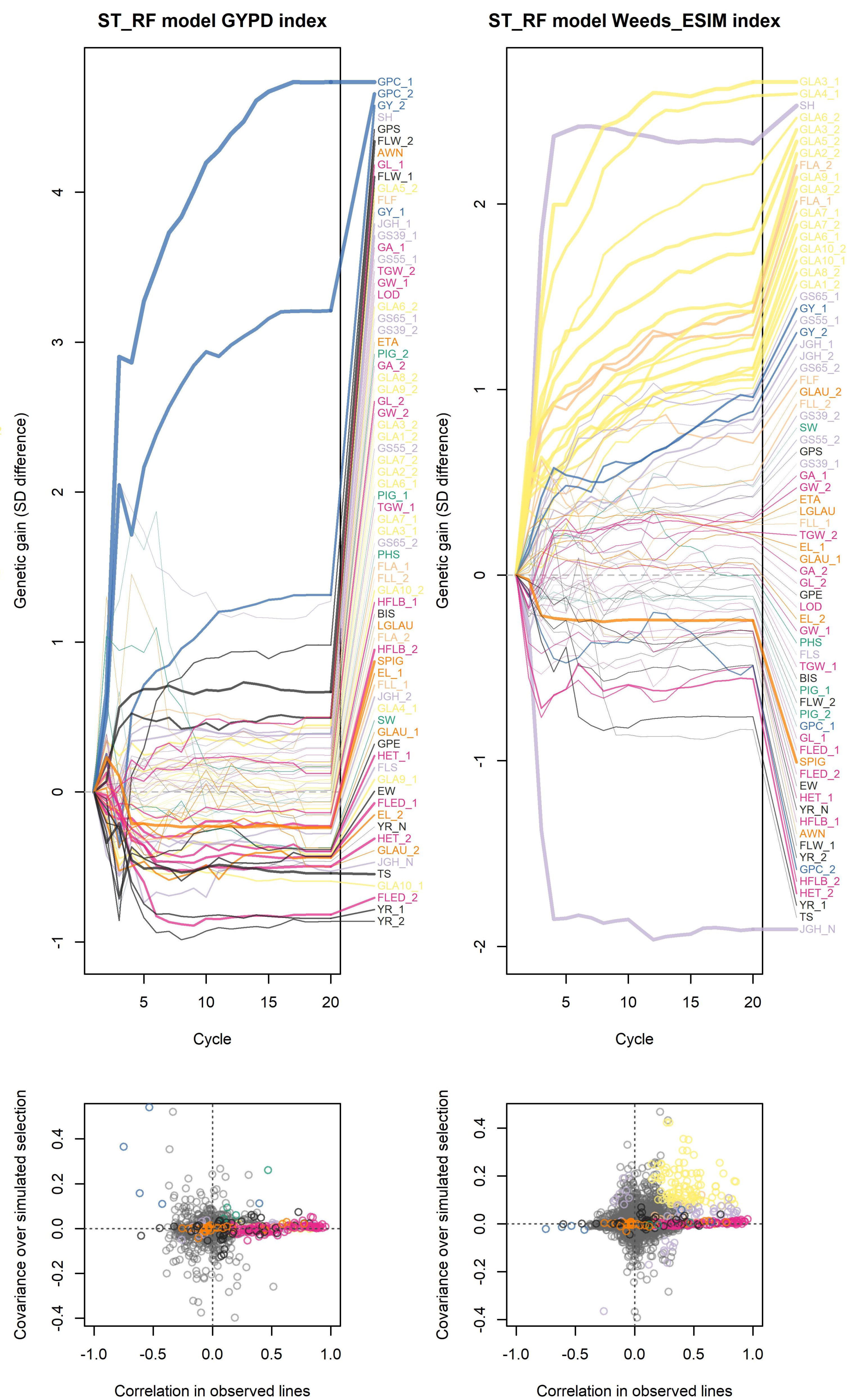

Rainfall (mm)
0
50
100
150
Max temp (deg C)
0
5
10
15
20
25
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Year 1
Year 2

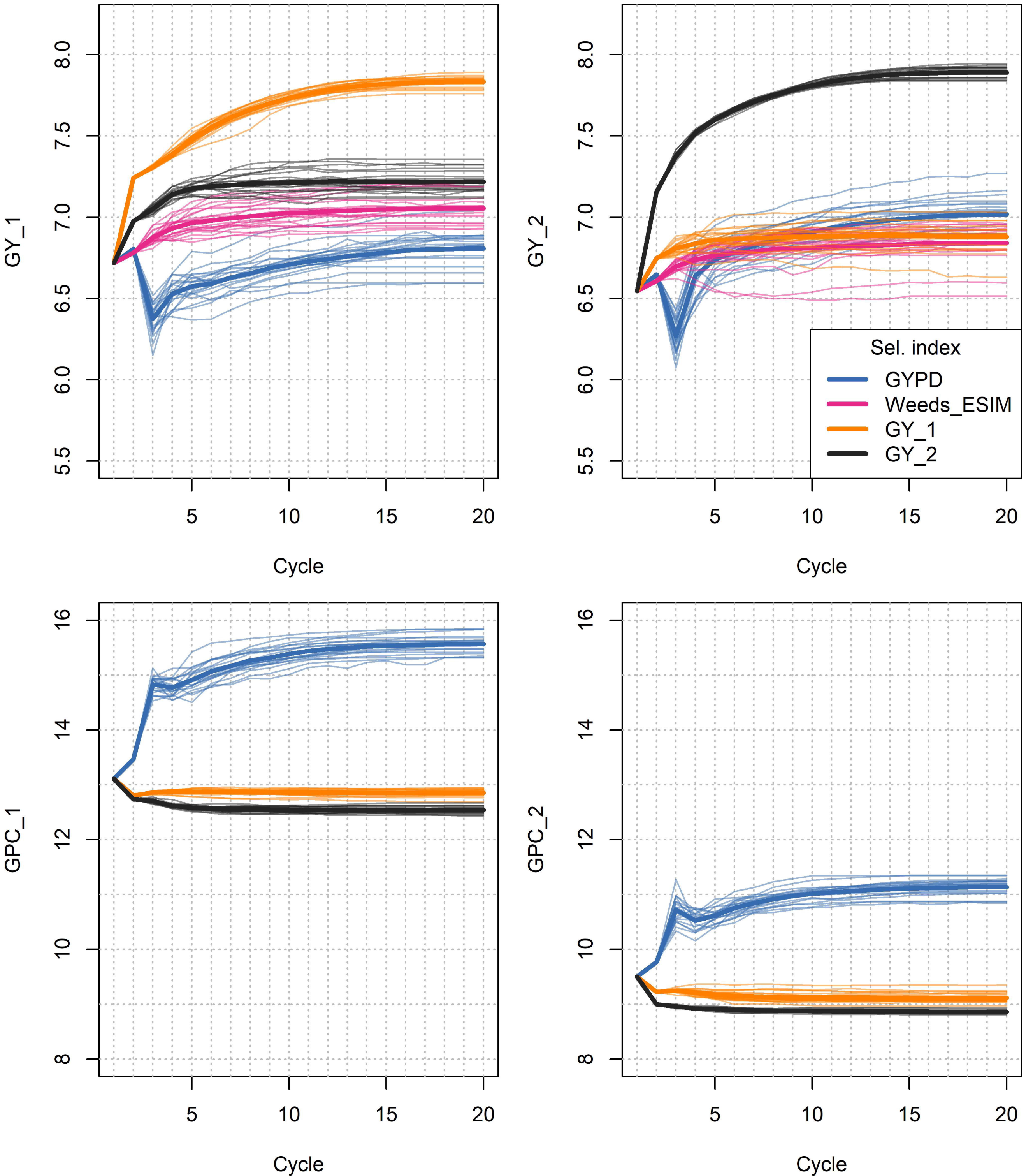

a
ST
MT ens
SVD MT
RF
LASSO
0.0
0.2
0.4
0.6
0.8
1.0
b
LASSO
MT ens
SVD MT
ST
Prediction accuracy (r)
0.438
0.454
0.385
c
RF
0.447
0.456
0.413

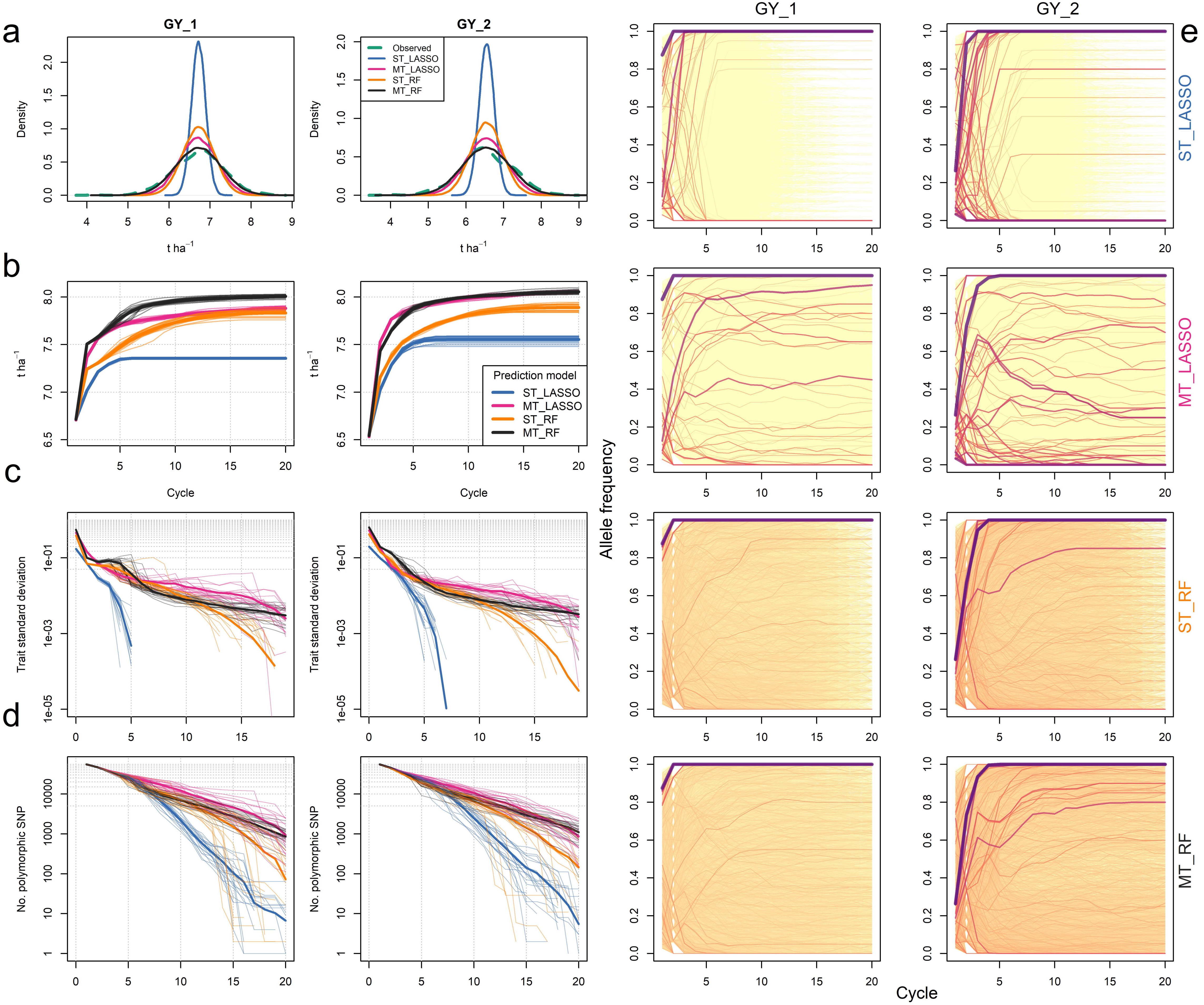

a
All traits
Scaled variable importance
8
6
4
2
0
-1.0
-0.5
0.0
0.5
1.0
Correlation coefficient
b
GY_1
GA_1
GLA2_2
JGH_1
GS65_1
LOD
GLA5_2
BIS
HFLB_1
GLA6_1
GLA6_2
GLA7_2
GS65_2
AWN
GLA3_2
PHS
GPE
GLA7_1
GLA10_1
SW
GLA9_1
EW
GS39_1
GS55_2
GS39_2
GS55_1
GY_2
FLS
GPC_2
GPC_1
GY_1
1
2
5
20
50
Variable importance
GY_2
🟥 -ve correlation
🟩 +ve correlation
GLA7_2
HFLB_2
GS65_1
FLED_2
PIG_2
GLA3_1
TS
GLA6_2
YR_N
PHS
JGH_1
GS55_1
GLA5_2
PIG_1
GLA8_2
GLA7_1
BIS
GPE
EW
GS65_2
GS55_2
JGH_2
GS39_2
SW
FLS
GPS
GY_1
GPC_1
GPC_2
GY_2
1
2
5
20
50
Variable importance


GY_1
GY_2
a
b
Prediction accuracy (r)
t ha⁻¹
Prediction model
ST_LASSO
MT_LASSO
ST_RF1
MT_RF1
ST_RF
MT_RF
1.0
0.8
0.6
0.4
0.2
0.0
8.0
7.5
7.0
6.5
6.0
5
10
15
20
Cycle


MT_RF model GY_1 index
MT_RF model GY_2 index
Genetic gain (SD difference)
Cycle
5
10
15
20
3
2
1
0
-1
AWN
GY_1
SH
PHS
GLA4_1
GY_2
TGW_2
GA_1
GA_2
GL_2
GL_1
GW_2
GLA3_1
FLW_1
FLW_2
GLA3_2
GLA6_1
EW
TGW_1
GPS
GW_1
GLA2_2
SW
GLA5_2
FLL_1
EL_1
BIS
GLA9_1
GS55_2
FLS
GS55_1
PIG_1
GPE
GLA10_1
GS39_2
LGLAU
GLA7_1
JGH_1
GS65_2
YR_N
EL_2
GS65_1
FLL_2
TS
GLA1_2
GLAU_1
GLA9_2
HFLB_1
FLA_2
GLA6_2
GLA7_2
JGH_2
GS39_1
SPIG
GLA10_2
HET_1
HFLB_2
PIG_2
GLAU_2
GLA8_2
FLF
HET_2
YR_2
YR_1
GPC_1
LOD
FLED_2
ETA
JGH_N
FLED_1
GPC_2
FLA_1
2
1
0
-1
SH
GY_2
GLA3_1
GLA4_1
GPS
GLA3_2
FLS
GS65_2
GW_2
GLA2_2
GY_1
EW
EL_2
GLA5_2
GS39_1
YR_1
GA_2
FLW_2
TGW_2
GPE
GLAU_2
GW_1
ETA
YR_N
GS65_1
GS55_1
LGLAU
GA_1
GS55_2
EL_1
GLA10_
YR_2
GL_2
GS39_2
JGH_1
GLAU_1
GLA6_1
GLA7_2
GLA9_1
TGW_1
GL_1
GLA6_2
GLA8_2
GLA7_1
FLF
PHS
FLL_2
FLW_1
GLA9_2
HFLB_1
HFLB_2
FLA_2
GLA10_
GLA1_2
FLA_1
HET_1
PIG_2
AWN
TS
LOD
HET_2
SPIG
FLL_1
PIG_1
JGH_2
BIS
FLED_1
FLED_2
GPC_1
SW
JGH_N
GPC_2
Covariance over simulated selection
Correlation in observed lines
0.6
0.4
0.2
0.0
-0.2
-0.4
-1.0
-0.5
0.0
0.5
1.0
0.4
0.2
0.0
-0.2
-1.0
-0.5
0.0
0.5
1.0


ST_RF model GYPD index
ST_RF model Weeds_ESIM index
Genetic gain (SD difference)
Cycle
4
3
2
1
0
-1
-2
5
10
15
20
GPC_1
GPC_2
GY_2
SH
GPS
FLW_2
AWN
GL_1
FLW_1
GLA5_2
FLF
GY_1
JGH_1
GS39_1
GA_1
GS55_1
TGW_2
GW_1
LOD
GLA6_2
GS65_1
GS39_2
ETA
PIG_2
GA_2
GLA8_2
GLA9_2
GL_2
GW_2
GLA3_2
GLA1_2
GS55_2
GLA7_2
GLA2_2
GLA6_1
PIG_1
TGW_1
GLA7_1
GLA3_1
GS65_2
PHS
FLA_1
FLL_2
GLA10_2
HFLB_1
BIS
LGLAU
FLA_2
HFLB_2
SPIG
EL_1
FLL_1
JGH_2
GLA4_1
SW
GLAU_1
GPE
HET_1
FLS
GLA9_1
EW
FLED_1
EL_2
YR_N
HET_2
GLAU_2
JGH_N
TS
GLA10_1
FLED_2
YR_1
YR_2
GLA3_1
GLA4_1
SH
GLA6_2
GLA3_2
GLA5_2
GLA2_2
FLA_2
GLA9_1
GLA9_2
FLA_1
GLA7_1
GLA7_2
GLA6_1
GLA10_2
GLA10_1
GLA8_2
GLA1_2
GS65_1
GY_1
GS55_1
GY_2
JGH_1
JGH_2
GS65_2
FLF
GLAU_2
FLL_2
GS39_2
SW
GS55_2
GPS
GS39_1
GA_1
GW_2
ETA
LGLAU
FLL_1
TGW_2
EL_1
GLAU_1
GA_2
GL_2
GPE
LOD
EL_2
GW_1
PHS
FLS
TGW_1
BIS
PIG_1
FLW_2
PIG_2
GPC_1
GL_1
FLED_1
SPIG
FLED_2
EW
HET_1
YR_N
HFLB_1
AWN
FLW_1
YR_2
GPC_2
HFLB_2
HET_2
YR_1
TS
JGH_N
Covariance over simulated selection
Correlation in observed lines
0.4
0.2
0.0
-0.2
-0.4
-1.0
-0.5
0.0
0.5
1.0


GY_1
8.0
7.5
7.0
6.5
6.0
5.5
5
10
15
20
Cycle
GY_2
8.0
7.5
7.0
6.5
6.0
5.5
Sel. index
GYPD
Weeds_ESIM
GY_1
GY_2
5
10
15
20
Cycle
GPC_1
16
14
12
10
8
5
10
15
20
Cycle
GPC_2
16
14
12
10
8
5
10
15
20
Cycle


a
b
c
d
e
GY_1
GY_2
Density
t ha^{-1}
Observed
ST_LASSO
MT_LASSO
ST_RF
MT_RF
Cycle
Prediction model
Trait standard deviation
No. polymorphic SNP
Allele frequency
ST_LASSO
MT_LASSO
ST_RF
MT_RF
Cycle