ORIGINAL ARTICLE

Leveraging genomic prediction to surpass current yield gains in spring barley

Johanna Åstrand[1,2] · Firuz Odilbekov[2] · Ramesh Vetukuri[1] · Alf Ceplitis[2] · Aakash Chawade[1]

Abstract

Key message **Genetic gain in Nordic spring barley varieties was estimated to 1.07% per year. Additionally, genomic predictive ability for yield was 0.61 in a population of breeding lines.**

Abstract Barley is one of the most important crops in Europe and meeting the growing demand for food and feed requires continuous increase in yield. Genomic prediction (GP) has the potential to be a cost-efficient tool in breeding for complex traits; however, the rate of yield improvement in current barley varieties is unknown. This study therefore investigated historical and current genetic gains in spring barley and how accounting for row-type population stratification in a breeding population influences GP results. The genetic gain in yield was estimated using historical data from field trials from 2014 to 2022, with 22–60 market varieties grown yearly. The genetic gain was estimated to 1.07% per year for all varieties, serving as a reference point for future breeding progress. To analyse the potential of using GP in spring barley a population of 375 breeding lines of two-row and six-row barley were tested in multi-environment trials in 2019–2022. The genetic diversity of the row-types was examined and used as a factor in the predictions, and the potential to predict untested locations using yield data from other locations was explored. This resulted in an overall predictive ability of 0.61 for yield (kg/ha), with 0.57 and 0.19 for the separate two-row and the six-row breeding lines, respectively. Together this displays the potential of implementing GP in breeding programs and the genetic gain in spring barley market varieties developed through GP will help in quantifying the benefit of GP over conventional breeding in the future.

Introduction

Barley (*Hordeum vulgare*) is the third most important crop in Europe with 91 million tons produced in 2022 (FAOSTAT 2024). While mainly used for feed and malting, the nutritious cereal is gaining popularity for human consumption (Baik and Ullrich 2008; Meints et al. 2016). Breeding for higher sustainable yields in a changing climate and introducing new varieties into new locations is vital to meet the growing demand for food stability. The success of breeding programs has been measured by the level of adoption of the newly released varieties by the growers (Brennan and Byerlee 1991). Another measurement is the rate of genetic

gain in varieties on the market (Rutkoski 2019). Genetic gain, or genetic response, is a measure of the improvement in average genetic or phenotypic value in a population due to repeated artificial selection (Hazel and Lush 1942). The genetic gain per unit time is estimated through the breeder's equation $R = \frac{i\sigma_A r}{t}$ and is dependent on the selection intensity (i), the square root of the additive genetic variance (σ_a), the selection accuracy (r) which is the narrow sense heritability (h^2) in phenotypic selection, and the cycle time (t). Several studies have investigated the genetic gain of barley in breeding programs around the world from the late 1800s and onwards, summarized in Table 1 and reviewed in Cossani et al. (2022). These have generally found advancement in genetic gain of 14.9–74 kg/ha per year or 0.4–1.1% per year; however, the studies often involve very few samples, sometimes with single varieties representing each decade of breeding (Table 1). There also seems to be a difference in the achieved genetic gain between the row-types in barley where the two-rowed lines have improved more in yield compared to the six-rowed varieties tested (Martiniello et al. 1987; Munoz et al. 1998). In addition, the genetic improvement

✉ Aakash Chawade
  aakash.chawade@slu.se

1   Department of Plant Breeding, Swedish University of Agricultural Sciences, Alnarp, Sweden

2   Lantmännen Lantbruk, Svalöv, Sweden

Table 1 Genetic gain estimates of spring and winter barley in breeding programs in different regions 1880–2013

Country/Region	Type and number of lines	Years	Genetic gain in yield per year	References
Argentina	Spring: 9 two-row	1944–1998	41 kg/ha or 0.36%	Abeledo et al. (2003)
Australia	Spring: 13 two-row	1942–2013	16 kg/ha or 0.43%	Cossani et al. (2022)
Brazil	Spring: 5 two-row 1 per decade	1968–2008	No gain until 1980, then 59.9 kg/ha	Rodrigues et al. (2020)
Canada	Spring: 20 six-row	1910–1988	26 kg/ha	Bulman et al. (1993)
Canada	Spring: 20 genotypes	1910–1987	41.1 kg/ha or 0.17% of harvest index	Jedel and Helm (1994)
England and Wales	Spring: 37 genotypes	1880–1980	15 kg/ha (0.39% per year) in 1880–1953, and 41 kg/ha (0.84% per year) in 1953–1980	Riggs et al. (1981)
Finland	Spring: 7 six-row 1 per decade	1927–1980	No gain	Peltonen-Sainio and Karjalainen (1991)
Italy	Winter: 5 two-row, 12 six-row	1960–1980	74 kg/ha or 1.1% (two-row) 52 kg/ha or 0.75% (six-row)	Martiniello et al. (1987)
Nordic Region	Spring: 90 two-row, 29 six-row	1930–1991	22 kg/ha or 0.28%	Ortiz et al. (2002)
Spain	Winter and spring: 10 two-row, 10 six-row	1930–1990	40.7 kg/ha for two-row 32.9 kg/ha for six-row	Munoz et al. (1998)
USA	Spring: 6 genotypes *1 per decade*	1920–1982	45.7 kg/ha or 0.9%	Wych and Rasmusson (1983)
USA	Spring: 10 genotypes	1920–1984	15.7 kg/ha per year	Boukerrou and Rasmusson (1990)
USA	Spring: 98 genotypes	1958–1998	14.9 kg/ha or 0.4%	Condón et al. (2009)

has not been steady over time, instead the increase in genetic gain for yield was higher in the second half of the 1900s (Riggs et al. 1981; Rodrigues et al. 2020).

In the Nordic region, the genetic gain has been analyzed in small populations from 1927 to 1980, where no gain was found (Peltonen-Sainio and Karjalainen 1991), and in 119 varieties from 1930 to 1991 with a gain of 22 kg/ha or 0.28% (Ortiz et al. 2002), but to our knowledge there have been no recent studies evaluating current progress in plant breeding. By periodically tracking the genetic gain in the current market varieties, it is possible to provide a reference point that can be used to evaluate new breeding strategies.

The continued improvement of spring barley is needed, which can be aided by incorporating genomic prediction (GP) in breeding programs. GP aims to predict breeding and genetic values based on marker effects (Bernardo 1994; Meuwissen et al. 2001). In GP, the genotypic and phenotypic data of a training population are used to find the genomic estimated breeding values (GEBVs) of individuals in a testing population where there is genetic but not phenotypic data available (Meuwissen et al. 2001). The GEBV is the sum of all marker effects estimated in the model from the accessions with both genotype and phenotype records and can be used to make predictions of the breeding value of non-phenotyped individuals based on their genotypic scores. The accuracy of GEBVs is defined as the Pearson correlation between GEBV and the true breeding value. Breeding programs with incorporated GP methods have the potential to achieve a higher genetic gain in a variety of species compared to breeding based on only phenotypic selection, through the optimization of the parameters in the breeder's equation, i.e., increased selection accuracy, selection intensity, and cycling time (Crossa et al. 2017). GP as a breeding strategy also led to a higher stability in the genetic gain over time when genotype by year interaction is present (Gaynor et al. 2017).

In barley, GP has been carried out with different combinations of models, marker densities, and population

compositions with varying predictive abilities (PAs). Rembe et al. (2022) achieved a PA of 0.73 for grain yield; however, with a leave-one-out strategy of different locations and years the PA of the model decreased to a range of 0.14–0.46. In another study with a low-density SNP dataset of 337 markers and only six-rowed breeding lines, the PA of yield was 0.54 (Tiede and Smith 2018). Thorwarth et al. (2017) found only minor differences between GP models when comparing eleven different GP models for German six-rowed winter barley and suggested instead that the population structure significantly affected the results.

A complicating factor in barley breeding programs is the genetic difference between the two-row and the six-rowed barley lines (Wonneberger et al. 2023); however, studies on GP using combined populations of two-rowed and six-rowed barley are not common. One study on naked barley germplasm used a diversity panel of two-row and six-row barley for the prediction of threshability where the subpopulations were found to differ significantly (Massman et al. 2023). This resulted in an overall PA of 0.86; however, within five out of six separate subpopulation the PA was lower than within the combined population (Massman et al. 2023). Where significant subpopulation structures with differing between-group PA, e.g., where both barley row-types exist within a breeding programe, it is important to consider population effects when designing a training population to avoid an inflated estimate of the PA for the entire population.

In this study, the genetic gain in spring barley varieties was investigated. A set of lines available on the market and prospective new lines were compared against cultivars with current or previously large market shares in Sweden. Historical data from official trials were used to estimate the progress of commercial lines released and registered to the market between 2001 and 2023. Field trial data from a breeding programe of two-row and six-row spring barley were also used to develop GP models. Yield data collected from multi-environment trials were used to develop and test models for predicting yield in specific targeted environments. The influence of the row-type and the breeding set on the predictive ability was investigated.

Materials and methods

Genetic gain analysis

Field data were collected from national multi-environment trials of market varieties of two-row spring barley typically grown in Sweden through Sverigeförsöken, conducted by Hushållningssällskapet and the Swedish University of Agricultural Sciences. For 2014–2022, data were collected from four locations in southern Sweden each year, with twelve different trial sites used. In each location the lines were tested

in an alpha lattice field design in two–three replicates. Each replicate consisted of five to eight sub-blocks containing four to eight lines depending on the trial site and the number of lines tested. Statistical analysis of the field trials was carried out per trial basis by Sverigeförsöken (https://sverigeforsoken.se/). Trials with three replicates were analyzed by using an incomplete block design with one treatment factor and the replicates as a random effect. Trials with two replicates were analyzed using a split-plot incomplete block analysis with replications with either random or fixed effects. The fields were treated according to local best practices for each location, adding fertilizer, fungicide, and herbicide. The number of lines tested each year varied from 22 to 60, with a total of 174 unique lines that represented common market varieties grown in Sweden, registered from 2001 to 2023, Supplemental Table 1 (Online resource 1). Yield was given as tons/hectare harvested grain at 15% moisture content.

Each year three checks were included: KWS Irina, RGT Planet, and Dragoon. The adjusted yield from the field design analysis of these three checks was averaged within each trial site to form a synthetic tester. To obtain a relative yield of each line, the yield of the tested line was divided by the yield of the synthetic tester within each environment. Genetic gain (GG) over time was estimated in two ways: as a population of varieties and as individual varieties available on the market. For the population of market varieties, the GG was estimated by plotting the relative yield of each line in the year it was tested, and performing regression analysis where the slope of the line over the years was used as the GG estimate. For the estimation of the improvement of released individual market varieties, the relative yield of an accession was averaged in all its tested years to obtain a measurement of its performance across environments. This relative yield was then plotted against the year of registration which was obtained from VarietyFinder (CPVO, www.cpvo.europa.eu) to determine the genetic gain over time using individual lines.

Plant material and field data for genomic prediction

Spring barley breeding lines from the Lantmännen plant breeding company were tested in three locations in Sweden for three years, 2019–2021. The population comprised 375 spring barley breeding lines with 222 two-row and 153 six-row lines. The lines were developed in two breeding sets with 178 lines in SWA18 (74 two-row and 104 six-row) and 195 (147 two-row and 48 six-row) in SWA19 and included two checks, Anneli (two-row) and SWJudit (six-row). The lines were tested at locations of different latitudes of 63.2 (Location 1), 59.6 (Location 2) and 58.3 (Location 3) to study the performance at different daylight and temperature conditions. All 375 lines were tested in all locations and

years and were planted in 15 m^2 plots in a non-replicated trial consisting of 42 blocks. Each block contained 39 entries in each block of which nine were checks; either Anneli or SWJudit was used in each block. The fields were treated according to local best practices with fertilizer and herbicide in all locations and fungicide in location 2 and 3. Yield was given per plot as kg/hectare at a target water moisture content of 15%.

Phenotype analysis

To control for differences in field conditions, each yield plot was normalized to a nearby check to estimate the adjusted mean yield of each genotype. This was carried out by using the linear methods of moving means analysis (Townley-Smith and Turd, 1973) in Genovix (Agronomix Software 2022) to obtain yield values adjusted to local checks within each environment. The adjusted means of each line were used to calculate the best linear unbiased predictor (BLUP) values using the lme4 package version 1.1–35.1 implemented in R 4.2.2 (Bates et al. 2015) with genotype, location, and year set as random effects. The following formula module was used to estimate BLUPs for each line in the population:

$$\text{BLUP} = yield \sim (1|gen) + (1|loc) + (1|year) + (1|gen:loc)$$
$$+ (1|gen:year) + (1|loc:year) + (1|loc:year:block) \quad (1)$$

where yield is the adjusted mean yield in kg/ha, gen is the genotype, loc is the trial location (Location 1–3), year is the trial year (2019–2021), gen:loc is the genotype and location interaction, gen:year is the genotype and year interation, loc:year is the interaction between the location and the year, and loc:year:block is the interaction between the location, year and the block effect of the non-replicated trial design. For downstream prediction evaluations within subgroups the BLUP values were calculated with lines unique to the subgroup. The lines belonging to different row-types and breeding sets were subset prior to the separate BLUP value estimation. For single environment predictions all locations and years for all lines were used for the BLUP estimation except for the environment to be predicted, i.e., two years with three locations and one year with two locations. BLUP values for individual locations and years were calculated using the entire population but with the following adjustments of the Eq. 1 for year and location. Individual years were calculated as:

$$\text{BLUP} = yield \sim (1|gen) + (1|loc) + (1|gen:loc) + (1|loc:block) \quad (2)$$

$$\text{BLUP} = yield \sim (1|gen) + (1|year)$$
$$+ (1|gen:year) + (1|year:block) \quad (3)$$

Broad sense heritability was calculated according to:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{gy}^2}{n_y} + \frac{\sigma_{gl}^2}{n_l} + \frac{\sigma_r^2}{n_y n_l}} \quad (4)$$

where σ_g^2, σ_{gy}^2, σ_{gl}^2, and σ_r^2 are the variance components of genotype, genotype×year, genotype×location, and residual variance, respectively, and the n_y, and n_l are the number of years, and locations per environment (Schmidt et al. 2019). The yield in each environment was plotted against the BLUP values estimated from all environments using ggplot2 version 3.4.4 (Wickham 2016) to visualize the relationship between yield in single environments and the adjusted values.

Genetic analysis of the population

The population was genotyped with the Illumina Infinium XT 15k SNP array by TraitGenetics (SGS Institut Fresenius, Germany), with 13,847 SNP markers. The markers were filtered by removing markers with more than 10% missing values in the entire population and markers without an annotated position in the barley genome, leaving 12,559 markers. A total of 9589 markers remained when filtering the entire population for a minor allele frequency (MAF) of > 0.05, whereas 7351 and 5624 markers remained when filtering two-row (221 lines) and six-row (152 lines) separately. When filtering for MAF above 0.05 in the entire population 4.6% of the polymorphic markers within each row-type were removed, 1.5% and 3.1% from the two-rowed and six-rowed, due to a too low MAF in the combined population. Missing data were imputed by the mean of each marker using the RR-BLUP package in R 4.2.2 (Endelman 2011). The imputation was carried out with the combined population for predictions within the combined population, and within each row-type and breeding set separately, for predictions with separate populations. The VanRaden kinship matrix was conducted and visualized using the GAPIT 3.0 package in R 4.2.2 (Wang and Zhang 2021) and principal component analysis (PCA) was conducted using the GAPIT 3.0 package in R 4.2.2 (Wang and Zhang 2021) and visualized using the plotly package in R 4.2.2 (Sievert 2020).

Genomic prediction

Genomic prediction was carried out using the ridge regression best linear unbiased prediction (RR-BLUP) model (Endelman 2011) in R 4.2.2. The RR-BLUP model used was:

$$y = WGu + \varepsilon \tag{5}$$

where W is the design matrix relating lines to observations, G is the genotype matrix, u is a vector of marker effects and ε is the residuals. The predictive ability was estimated with fivefold cross-validation using 80% of the training population to predict the remaining 20% set as the validation population. The model was iterated with a randomized training population 500 times. The average accuracy of the predictions was determined by correlating the predicted values against the observed values of the validation population and average this correlation over the iterations. For predictions for single environments the BLUPs were used to predict the adjusted mean yield values which were used as the validation, i.e., a single year and location prediction.

Table 2 Number of lines and yield of spring barley varieties tested in official trials in 2014–2022

Year tested	Number of varieties	Relative yield	Absolute yield (tons/ha)
2014	60	0.96	6.03
2015	54	0.93	7.20
2016	55	0.98	6.49
2017	48	0.96	6.83
2018	50	1.01	4.49
2019	49	1.01	6.87
2020	22	1.00	5.98
2021	23	1.05	4.57
2022	37	1.02	7.65

Relative yield is the average yield of all varieties tested compared to the synthetic tester within each year. The absolute yield is the average yield for all varieties tested in each year

Results

Realized genetic gain in market varieties

The genetic gain in spring barley market varieties (Table 2) was investigated as i) improvement of the population of market varieties available to growers, and ii) as improvement of registered market varieties. In the first scenario, the average increase in yield of the population of market varieties tested in individual years was 1.07% per year (Fig. 1a) suggesting that the population of varieties available to growers each year had improved yield compared to the synthetic tester. The increase in yield per year for the top five highest yielding varieties was 0.94%, whereas the lowest yielding varieties improved by 1.57% per year for the population of market varieties. In the second scenario, the average increase in yield of individual varieties compared to the synthetic tester was 0.63% per year for varieties registered 2001–2023 (Fig. 1b), suggesting that released varieties showed a genetic gain over this time.

Population structure and genetic variance in a barley breeding programe

A principal component analysis (PCA) of the population detected the presence of two distinct subpopulations in the breeding population (Fig. 2a). The first principal component (PC) accounted for 25.8% of the genetic variation and separated the population depending on row-type, whereas the second PC accounted for only 8% variance. Within the two-row lines, there is some separation depending on the breeding set where the SWA19 appears more genetically diverse in the two-row lines. However, there is no such clear division within the six-row lines. This was also seen in the heat map of the kinship matrix where the lines were clearly
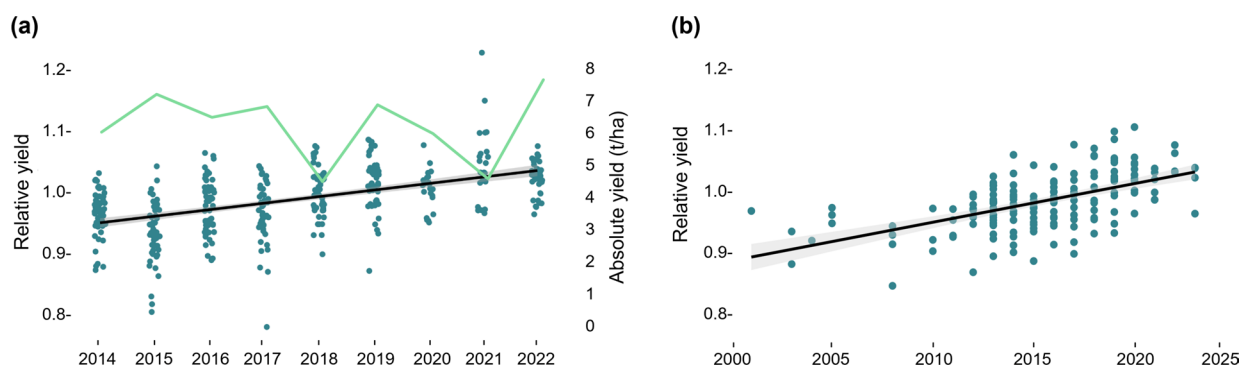

Fig. 1 Relative and absolute yield of Swedish barley market varieties. **a** The genetic gain in a population of common market varieties tested in individual years increased by 1.07% from 2014 to 2022. The green line shows the absolute yield in tons/ha for the lines tested in each year as an average of the locations. **b** The genetic gain in varieties depending on year of release was 0.63% per year for market varieties registered 2001–2023
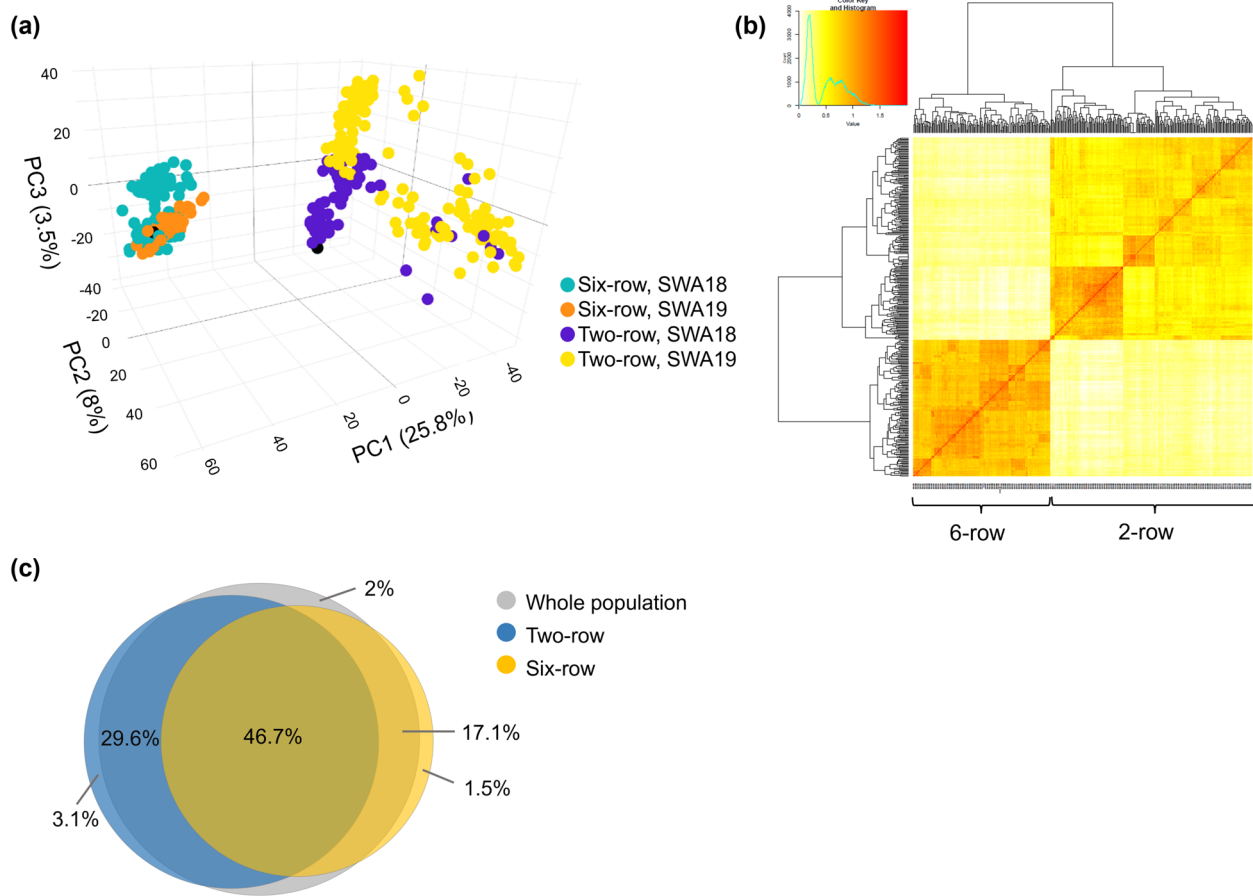
Fig. 2 PCA plot, kinship heat map and markers with a MAF above 0.05 in the barley breeding lines. **a** Clustering of the genotypes depending on the row-type can be clearly identified from the PCA. The six-rowed genotypes clustered very tightly; however, within the two-rows, there was some diversity attributed to the breeding set in which the lines were developed. **b** Heat map of the VanRaden kinship matrix showing the relatedness of the lines depending on the row-type of the lines. The two-rowed lines display more clustering within the subpopulation compared to the six-rowed genotypes. **c** The proportion of markers with a MAF > 0.05 differed between the entire population and the groups of row-types, where 32.7% and 18.6% of the markers were polymorphic only in the two-row or the six-row lines, respectively

separated depending on the row-type and some clustering was observed within the two-row genotypes (Fig. 2b). Although the difference between the breeding sets were only minor they were of interest for the possible effect on accuracy when conducting predictions between-breeding sets.

To analyze the difference in marker polymorphism in the breeding population the minor allele frequency (MAF) within the two-rowed and the six-rowed groups was calculated separately. The number of polymorphic markers, i.e., with a MAF above 0.05, differed depending on the row-type where 32.7% and 18.6% of the markers were uniquely polymorphic for either the two-row or the six-row lines, respectively, suggesting a significant difference in allele diversity between the row-types (Fig. 2c). Of the markers in the whole population, 2% were polymorphic only in the combined population, due to the fixation of different alleles in the separate row-type groups, i.e., non-informative markers

for either row-type. Overall, only 46.7% of the markers were polymorphic in both row-type groups, suggesting that more than half of the markers were non-informative for the combined population.

Grain yield is influenced by the interplay of row-type and environment

The broad sense heritability for yield (kg/ha) in the whole population was high (0.73) with a genetic variance of 0.089 (Table 3). The overall yield was higher in the two-rowed varieties than in the six-rowed, (p-value 9.4×10^{-6}) with a higher genetic variance in the two-row lines. The two-rowed lines also showed a wider distribution of yield compared to the six-rowed, suggesting more diversity in the population (Table 3). In the individual environments 2021 was the lowest yielding year in all locations, whereas 2019 and 2020

Table 3 Broad sense heritability, genetic variance components, and population averages for yield (kg/ha) using BLUP values in spring barley

	All	Two-row	Six-row
Heritability (H^2)	0.727	0.791	0.597
Genetic variance	0.089	0.104	0.056
Gen:year variance	0.011	0.003	0.026
Gen:loc variance	0.010	0.013	0.008
Year	0.705	0.720	0.648
Location	1.311	1.031	1.621
Year:location	0.107	0.235	0.025
Residual variance	0.239	0.197	0.239
Average (kg/ha)	5.74	5.95	5.51
Min (kg/ha)	4.47	4.97	4.47
Max (kg/ha)	7.13	7.13	6.55

were more comparable (Fig. 3), similar to the results from the genetic gain analysis (Fig. 1a). Location 2 was the highest yielding location, whereas the southernmost Location 3 was the lowest yielding location overall. The two-row varieties were generally higher yielding than the six-rowed varieties in the northernmost Location 1 and the reverse was true in the southernmost Location 3 (Fig. 3), suggesting that the row-type in this population might be of importance for the yield at different latitudes and climates.

Genomic selection of targeted groups

The predictive ability (PA) for yield (kg/ha) in the entire population was 0.609 (Table 4, Scenario A). Due to the genetic difference between row-types, the breeding lines were divided depending on row-type to improve the PA. Interestingly, the PA decreased slightly to 0.572 within the two-row population and to 0.193 for the six-row lines (Table 4, Scenario B). Using markers which were only polymorphic within one row-type did not affect the PA, resulting in 0.579 and 0.189 for two-row and six-row lines, respectively. Using one row-type to predict the other (Table 4, Scenario C) resulted in low PA for both populations, suggesting that between row-type predictions are suboptimal.

The predictive ability within-breeding sets was slightly higher in the SWA19 set (0.527) than for the SWA18 set (0.437) (Table 4, Scenario D). The SWA19 generation had a higher proportion of two-rowed varieties which could explain why the PA was higher within this generation compared to the SWA18 which had a higher proportion of six-rowed lines. Using one breeding set to predict the GEBVs of another generation lead to a decrease in PA to 0.37 for SWA18 to SWA19 and 0.326 for SWA19 to SWA18 (Table 4, Scenario E), suggesting that between-breeding set predictions result in lower PA than within-breeding sets but

is still a viable option. In addition, the division of the population into breeding sets and row-types resulted in smaller population sizes used to evaluate the prediction models, which could also have contributed to the lowering of the PA.

Predictions between years and locations were also carried out to investigate if predicting performance in an untested environment could be successful. The highest PA between different years was achieved by using data from 2020 and 2021, for lines in 2019, with a PA of 0.389 (Table 4, Scenario F). Of the years tested, the yield was the most stable across locations in 2019 which might have contributed to the higher PA. Interestingly, when using data from two locations to predict the third location, the PA was the lowest for the location with the overall highest yields, i.e., Location 2 (0.279), and highest for the location with the lowest overall yields, Location 3 (0.365) (Table 4, Scenario G).

Considering the different performance of the row-types depending on the location (Fig. 3), single location predictions were carried out to determine if the accuracy could be increased within a location, resulting in a PA of 0.357, 0.307, and 0.487 within locations 1, 2 and 3, respectively (Table 4, Scenario H). Again, the highest PA was achieved for the lowest yielding Location 3 and the lowest PA for the highest yielding location 2.

A more realistic scenario for a breeding programme that targets a specific region, for example Location 1, would be to test some material in the target region and investigate whether adding data from other regions would improve the PA of untested lines. In this approach, the PA varied significantly depending on the year within Location 1 (Table 4, Scenario I). Using 2019 and 2021 to predict the yield in 2020, which was the highest yielding year, resulted in the highest PA (0.441), whereas using data from 2019 and 2020 for the prediction of 2021, the lowest yielding year, resulted in the lowest PA (0.069). There was more variability in the PA when predicting only one environment compared to predicting one location with three years of data. This follows naturally from the increase in variability in the line performance in each environment.

Discussion

Genetic gain

The continued yield gain increasingly depends on progress from plant breeding (Lillemo et al. 2010; Mackay et al. 2011). Understanding the historical and current gains achieved through breeding is essential to set future targets and evaluate the performance of new breeding strategies. Here, field trial data were collected to compare varieties commonly grown in Sweden, i.e., KWS Irina, Dragoon and RGT Planet, to a population of current and candidate
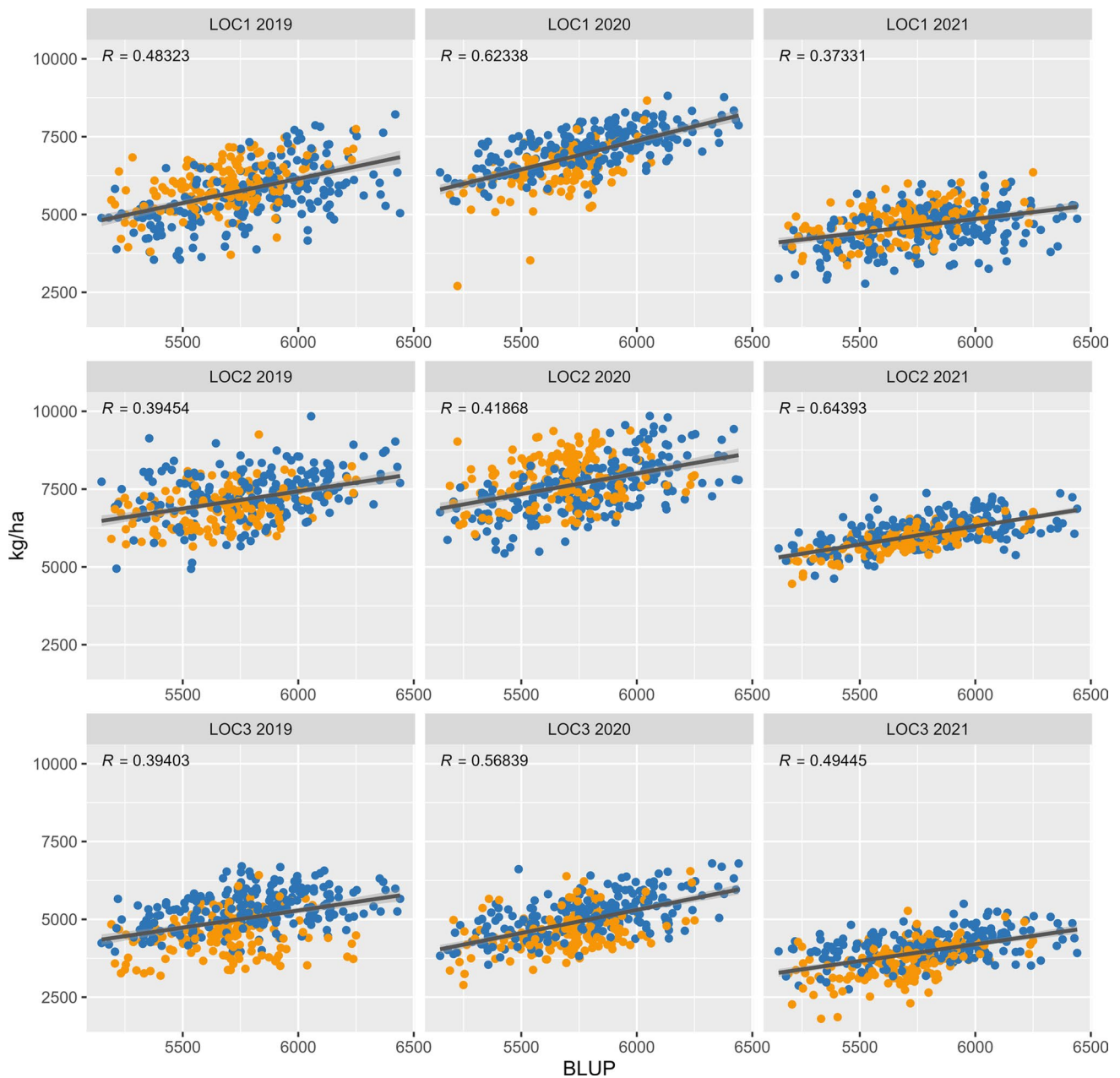
Fig. 3 Correlation of the yield (kg/ha) in individual environments compared to BLUP values. All individual environments showed a medium to high correlation to the BLUP values as displayed with the Pearsons correlation coefficient (R) for each environment. The two-row lines (blue) were generally higher yielding compared to the six-row lines (orange) (colour figure online)

market varieties. This approach made it possible to investigate the genetic gain of material currently available to farmers. Within this population of varieties marketed for this region, there was a trend of genetic gain increase for the yield of 1.07% per year, compared to the checks, suggesting that there is progress in the breeding for better-performing lines. The yield of the top performers in each year showed a smaller improvement of 0.95% per year, whereas the lowest yielding lines improved by 1.57% per year. It is unclear whether this improvement of the worst-performing lines was

due to improvement or to the removal of lines included in the field trials. However, new lines were continuously added to the population over the tested years, and they performed better than the lines removed from the trials with an average of 0.63% genetic gain in line performance per year.

Ortiz et al. (2002) studied genetic gain in Nordic spring barley and found a genetic gain of 13% over 60 years of breeding. Finnish spring barley varieties showed no genetic gain in varieties released from 1927 to 1980 (Peltonen-Sainio and Karjalainen 1991). The lines used in the current

Table 4 Predictive ability of genomic prediction of different training and validation populations for yield (kg/ha) in spring barley

Scenario	TP	VP	Pop size	TP size	VP size	PA yield
A. Whole population	All (80%)	All (20%)	375	300	75	0.609
B. Within row-type	Two-row	Two-row	222	178	44	0.572
	Six-row	Six-row	153	122	31	0.193
C. Between row-type	Two-row	Six-row	222	178	44	0.102
	Six-row	Two-row	153	122	31	0.210
D. Within-breeding set	SWA18	SWA18	178	142	36	0.437
	SWA19	SWA19	195	156	39	0.526
E. Between-breeding set	SWA18	SWA19	178	178	195	0.370
	SWA19	SWA18	195	195	178	0.326
F. Between years	2020 + 2021	2019	375	300	75	0.389
	2019 + 2021	2020	375	300	75	0.334
	2019 + 2020	2021	375	300	75	0.260
G. Between locations	Loc 1 + Loc 2	Loc 3	375	300	75	0.365
	Loc 1 + Loc 3	Loc 2	375	300	75	0.279
	Loc 2 + Loc 3	Loc 1	375	300	75	0.309
H. Within locations	Loc 1	Loc 1	375	300	75	0.357
	Loc 2	Loc 2	375	300	75	0.307
	Loc 3	Loc 3	375	300	75	0.487
I. Single location	All-Loc 1, 2019	Loc 1 in 2019	375	300	75	0.326
	All-Loc 1, 2020	Loc 1 in 2020	375	300	75	0.441
	All-Loc 1, 2021	Loc 1 in 2021	375	300	75	0.069

Yield data for 375 spring barley breeding lines tested in three locations and three years was used to test PA in different scenarios of training (TP) and validation (VP) population composition using fivefold cross-validation in RR-BLUP. PA, predictive ability

study were all registered on the Swedish market from 2001 to 2023 and constitute an estimate of the genetic gain in the varieties available to farmers. The overall gain of the varieties in this study of 1.07% per year is one of the highest reported for spring barley, suggesting a steady increase in yield potential over time; however, there is a need to improve the genetic gain to meet future demands.

Increasing genetic gain in future market varieties remains a challenge but also provides opportunities for new technology. Through simulated trials with varying population sizes and trait heritability, the genomic selection strategy constantly outperformed phenotypic or marker-assisted selection (MAS) in predicted genetic gain over time (Bernardo and Yu 2007; Iwata and Jannink 2011; Jannink 2010). In barley breeding, GP using RR-BLUP resulted in a higher genetic gain increase in comparison with traditional phenotypic selection which was especially true for traits with low heritability and for populations with large sample sizes (Iwata and Jannink 2011). By using GP, the expected genetic gain per year can be two to three times higher for wheat and maize in comparison with the use of MAS or phenotypic selection (Heffner et al. 2010; Tessema et al. 2020). Integrating GP in breeding programs can potentially increase genetic gain while reducing cost per unit time (Heffner et al. 2010; Lorenz 2013).

Genomic selection for targeted regions

For growers, the selection of varieties for cultivation could be assisted by variety testing in locations resembling their own conditions (Cooper et al. 1996; Smith et al. 2015). However, for breeding, the prediction of the performance of lines in larger target regions still lacks in the predictive ability of new environments (Smith et al. 2015). In this study, using observed environments to predict untested environments (the combination of a year and a location) led to moderate predictive ability (Table 4). The variation between trial sites within a region can be as large as the variation between regions suggesting that non-static genotype by environment interactions due to seasonal factors can be higher than regional differences (Cullis et al. 2006).

A higher PA depends on designing the right training set from a carefully designed breeding scheme (Bassi et al. 2016). Training population size is one of the most important considerations for the PA in GS (Asoro et al. 2011; Isidro et al. 2015; Lorenzana and Bernardo 2009), more so than the marker density and number, as it affects the diversity and the genetic variance that can be identified in a crop. The accuracy of predictions is usually decreased if there is little relatedness between the training and the breeding population (Gianola et al. 2009; Habier et al. 2007, 2010; Zhong et al.

2009). In wheat, the PA for a validation population of little relatedness to the training population could be improved by increasing the diversity of the training set (Norman et al. 2018). In contrast, using populations with low relatedness decreased the PA from an overall of 0.8–0.4 (Norman et al. 2018).

The population structure caused by the row-types in spring barley is of importance for the implementation of GP methods in breeding programs. Although the genetics resulting in the fertility of the lateral spikes, and thereby the two- or six-rowed phenotype, is relatively simplistic and well-known, the extent of the difference in genetics resulting from the separation of the populations is less known. In a study with four wild barley (three *Hordeum ssp. spontaneum* and one *Hordeum spp. agriocrithon*) and a diversity panel of cultivated inbred barley lines, the most genetic variation was attributed to the row-type, rather than the cultivation status, and accounted for 15.1% of the genetic variation (Casale et al. 2022). Another study found differences in allele frequency between two- and six-rowed barley, potentially due to the incorporation and fixation of important loci related to yield, disease resistance, abiotic stress resistance, and flowering time (Wonneberger et al. 2023). Breeding of barley in modern times has focused on adapting barley for a variety of different purposes and environments. Higher protein content for varieties aimed for animal feed, and low protein and large grain size for malting barley varieties could have contributed to the differences in the germplasms. In addition, alleles for powdery mildew resistance (Friedt et al. 2011; Jørgensen 1992), selection for late flowering, and increased yield in European two-rowed spring barley (Tondelli et al. 2013) led to further separation between row-types.

The spring barley population used in the current study consisted of lines of different row-types and from different breeding sets. Here, there was significant genetic difference depending on row-type, accounting for 25.8% of the genetic variation, with less diversity assigned to breeding set. When all lines were used as one population the PA was high, 0.609, whereas predictions within the individual row-type groups achieved accuracies of 0.572 and 0.193 for two-row and six-row accessions, respectively. The genetic separation of the row-types in barley also resulted in very low PA when using one row-type to predict the other whereas the inter-breeding set prediction was still moderate (Table 4, Scenario C and E). This could indicate that the high PA for the whole population was inflated, due to the clustering of the genetic subpopulations depending on row-types. Predictions using whole populations can lead to higher PAs compared to within-subpopulation predictions (Guo et al. 2014; Werner et al. 2020) suggesting that the accounting for population structure is important for accurate prediction estimates.

Another factor influencing GP accuracy is the size of the training population. When separating the population based on row-types or breeding sets there was a significant reduction in effective population size for each training population which could have contributed to the lowering of the PA compared to the use of the entire population. In a study on barley, Nielsen et al. (2016) found that reducing the training set to less than 200 lines decreased the PA. In addition, with a lower training population the importance of the population structure and the relatedness of the lines is increased (Nielsen et al. 2016). To account for population structure such as the row-type in barley through the division of the training population, the overall size of the training population decreases, resulting in a reduced PA. Including more observations could be necessary to improve the predictions to accurately account for the differences between the of the spring barley row-types.

Conclusion

Genetic gain per year in current market varieties was estimated as 1.07% for the varieties tested. This estimate can serve as a checkpoint for the evaluation of new varieties currently developed. Introducing genomic prediction in plant breeding programs has the potential to streamline selection in breeding programs. In barley separating the population based on row-type to increase the predictive ability in genomic selection could be beneficial if the number of lines in the breeding population is large enough. Large populations might be needed to account for population differences depending on row-type to improve the predictive ability of a divided population, whereas programs with a low number of one or both row-types might benefit from keeping the population intact. Targeted breeding for untested regions could significantly reduce phenotyping costs; however, the effects of the variations in the environments, i.e., years and locations, somewhat reduced the predictive ability.

Author contributions Johanna Åstrand, Aakash Chawade and Alf Ceplitis conceived of the study. Ramesh Vetukuri, Aakash Chawade and Alf Ceplitis received the funding. Data analysis and graph preparation were carried out by Johanna Åstrand. Firuz Odilbekov and Alf Ceplitis provided the genotype and phenotype data for the genomic prediction study. The first draft of the manuscript was written by Johanna Åstrand. All authors reviewed and edited previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability The datasets analyzed for genetic gain estimation during the current study are available at the Sverigeförsöken repository (https://sverigeforsoken.se/).

Declarations

Conflict of interest This work was supported by SLU Grogrund (SLU. ltv.2022.4.3-403). ACe and JÅ were employed by Lantmännen Lantbruk. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abeledo LG, Calderini DF, Slafer GA (2003) Genetic improvement of barley yield potential and its physiological determinants in Argentina (1944–1998). Euphytica 130:325–334. https://doi.org/10.1023/A:1023089421426

Agronomix Software I (2022) Plant breeding software from agronomix. https://www.agronomix.com/. Accessed Jan 2024

Asoro FG, Newell MA, Beavis WD, Scott MP, Jannink J (2011) Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. Plant Genome. https://doi.org/10.3835/plantgenome2011.02.0007

Baik BK, Ullrich SE (2008) Barley for food: characteristics, improvement, and renewed interest. J Cereal Sci 48:233–242. https://doi.org/10.1016/j.jcs.2008.02.002

Bassi FM, Bentley AR, Charmet G, Ortiz R, Crossa J (2016) Breeding schemes for the implementation of genomic selection in wheat (Triticum spp.). Plant Sci 242:23–36. https://doi.org/10.1016/j.plantsci.2015.08.021

Bates D, Mächler M, Bolker BM, Walker SC (2015) Fitting linear mixed-effects models using lme4. J Stat Softw. https://doi.org/10.18637/jss.v067.i01

Bernardo R (1994) Prediction of maize single-cross performance using RFLPs and information from related hybrids. Crop Sci 34(1):20–25. https://doi.org/10.2135/cropsci1994.0011183X003400010003x

Bernardo R, Yu J (2007) Prospects for genomewide selection for quantitative traits in maize. Crop Sci 47:1082–1090. https://doi.org/10.2135/cropsci2006.11.0690

Boukerrou L, Rasmusson DD (1990) Breeding for high biomass yield in spring barley. Crop Sci 30:31–35. https://doi.org/10.2135/cropsci1990.0011183X003000010007x

Brennan JP, Byerlee D (1991) The rate of crop varietal replacement on farms: measures and empirical results for wheat. Plant Var Seeds 4:99–106

Bulman P, Mather DE, Smith DL (1993) Genetic improvement of spring barley cultivars grown in eastern Canada from 1910 to 1988. Euphytica 71:35–48. https://doi.org/10.1007/BF00023465

Casale F, Van Inghelandt D, Weisweiler M, Li J, Stich B (2022) Genomic prediction of the recombination rate variation in barle—A route to highly recombinogenic genotypes. Plant Biotechnol J 20(4):676–690. https://doi.org/10.1111/pbi.13746

Condón F, Rasmusson DC, Schiefelbein E, Velasquez G, Smith KP (2009) Effect of advanced cycle breeding on genetic gain and phenotypic diversity in barley breeding germplasm. Crop Sci 49(5):1751–1761. https://doi.org/10.2135/cropsci2008.10.0585

Cooper M, Brennan PS, Sheppard JA (1996) A strategy for yield improvement of wheat which accommodates large genotype by environment interactions. In: Cooper M, Hammer GL (eds) Plant adaptation and crop improvement. CAB International, pp 487–511

Cossani CM, Palta J, Sadras VO (2022) Genetic yield gain between 1942 and 2013 and associated changes in phenology, yield components and root traits of Australian barley. Plant Soil 480:151–163. https://doi.org/10.1007/s11104-022-05570-7

Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de Los Campos G, Burgueño J, González-Camacho JM et al (2017) Genomic selection in plant breeding: methods, models, and perspectives. Trends Plant Sci 22(11):961–975. https://doi.org/10.1016/j.tplants.2017.08.011

Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. J Agric Biol Environ Stat 11(4):381–393. https://doi.org/10.1198/108571106X154443

Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4(3):250–255. https://doi.org/10.3835/plantgenome2011.08.0024

FAO-STAT. (2024) Crops and livestock products. FAO; Library of Congress. https://www.fao.org/faostat/en/#data/QCL Accessed Jan 2024

Friedt W, Horsley RD, Harvey BL, Poulsen DME, Lance RCM, Ceccarelli S, Grando S, Capettini F (2011) Barley breeding history, progress, objectives, and technology. In: Ullrich SE (ed) Barley: production, improvment and uses. Blackwell Publishing Ltd., Oxford, pp 160–220

Gaynor RC, Gorjanc G, Bentley AR, Ober ES, Howell P, Jackson R, Mackay IJ, Hickey JM (2017) A two-part strategy for using genomic selection to develop inbred lines. Crop Sci 57(5):2372–2386. https://doi.org/10.2135/cropsci2016.09.0742

Gianola D, De Los Campos G, Hill WG, Manfredi E, Fernando R (2009) Additive genetic variability and the bayesian alphabet. Genetics 183(1):347–363. https://doi.org/10.1534/genetics.109.103952

Guo Z, Tucker DM, Basten CJ, Gandhi H, Ersoz E, Guo B, Xu Z, Wang D, Gay G (2014) The impact of population structure on genomic prediction in stratified populations. Theor Appl Genet 127(3):749–762. https://doi.org/10.1007/s00122-013-2255-x

Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genome-assisted breeding values. Genetics 177(4):2389–2397. https://doi.org/10.1534/genetics.107.081190

Habier D, Tetens J, Seefried FR, Lichtner P, Thaller G (2010) The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genet Sel Evol. https://doi.org/10.1186/1297-9686-42-5

Hazel LN, Lush JL (1942) The efficiency of three methods of selection. J Hered. https://doi.org/10.1093/oxfordjournals.jhered.a105102

Heffner EL, Lorenz AJ, Jannink JL, Sorrells ME (2010) Plant breeding with genomic selection: gain per unit time and cost. Crop Sci 50(5):1681–1690. https://doi.org/10.2135/cropsci2009.11.0662

Isidro J, Jannink JL, Akdemir D, Poland J, Heslot N, Sorrells ME (2015) Training set optimization under population structure in genomic selection. TAG. Theor Appl Genet 128(1):145–158. https://doi.org/10.1007/s00122-014-2418-4

Iwata H, Jannink JL (2011) Accuracy of genomic selection prediction in barley breeding programs: a simulation study based on

the real single nucleotide polymorphism data of barley breeding lines. Crop Sci 51(5):1915–1927. https://doi.org/10.2135/cropsci2010.12.0732

Jannink JL (2010) Dynamics of long-term genomic selection. Genet Sel Evol. https://doi.org/10.1186/1297-9686-42-35

Jedel PE, Helm JH (1994) Assessment of western Canadian barleys of historical interest: I yield and agronomic traits. Crop Sci 34(4):922–927. https://doi.org/10.2135/cropsci1994.0011183X003400040015x

Jørgensen H (1992) Discovery, characterization and exploitation of Mlo powdery mildew resistance in barley. Euphytica 63:141–152. https://doi.org/10.1007/BF00023919

Lillemo M, Reitan L, Bjørnstad Å (2010) Increasing impact of plant breeding on barley yields in central Norway from 1946 to 2008. Plant Breed 129(5):484–490. https://doi.org/10.1111/j.1439-0523.2009.01710.x

Lorenz AJ (2013) Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: a simulation experiment. G3 Gen Genomes Genet 3(3):481–491. https://doi.org/10.1534/g3.112.004911

Lorenzana RE, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor Appl Genet 120(1):151–161. https://doi.org/10.1007/s00122-009-1166-3

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

Martiniello P, Delogu G, Odoardi M, Boggini G, Stanca AM (1987) Breeding progress in grain yield and selected agronomic characters of winter barley (*Hordeum vulgare* L.) over the last quarter of a century. Plant Breed 99:289–294. https://doi.org/10.1111/j.1439-0523.1987.tb01184.x

Massman C, Meints B, Hernandez J, Kunze K, Smith KP, Sorrells ME, Hayes PM, Gutierrez L (2023) Genomic prediction of threshability in naked barley. Crop Sci 63(2):674–689. https://doi.org/10.1002/csc2.20907

Meints B, Cuesta-Marcos A, Fisk S, Ross A, Hayes P (2016) Food Barley Quality Improvement and Germplasm Utilization. In *Exploration, Identification and Utilization of Barley Germplasm*, Elsevier, pp. 41–73

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Munoz P, Voltas J, Araus JL, Igartua E, Romagosa I (1998) Changes over time in the adaptation of barley releases in north-eastern Spain. Plant Breed 117:531–835. https://doi.org/10.1111/j.1439-0523.1998.tb02202.x

Nielsen NH, Jahoor A, Jensen JD, Orabi J, Cericola F, Edriss V, Jensen J (2016) Genomic prediction of seed quality traits using advanced barley breeding lines. PLoS ONE. https://doi.org/10.1371/journal.pone.0164494

Norman A, Taylor J, Edwards J, Kuchel H (2018) Optimising genomic selection in wheat: effect of marker density, population size and population structure on prediction accuracy. G3 Genes Genomes Genet 8(9):2889–2899. https://doi.org/10.1534/g3.118.200311

Ortiz R, Nurminiemi M, Madsen S, Arne Rognli O, Bjørnstad Å (2002) Genetic gains in Nordic spring barley breeding over sixty years. Euphytica 126:283–289. https://doi.org/10.1023/A:1016302626527

Peltonen-Sainio P, Karjalainen R (1991) Genetic yield improvement of cereal varieties in northern agriculture since 1920. Acta Agric Scand 41(3):267–273. https://doi.org/10.1080/00015129109439909

Rembe M, Zhao Y, Wendler N, Oldach K, Korzun V, Reif JC (2022) The potential of genome-wide prediction to support parental selection evaluated with data from a commercial barley breeding program. Plants. https://doi.org/10.3390/plants11192564

Riggs TJ, Hanson PR, Start ND, Miles DM, Morgan CL, Ford MA (1981) Comparison of spring barley varieties grown in England and Wales between 1880 and 1980. J Agric Sci 97:599–610. https://doi.org/10.1017/S0021859600036935

Rodrigues O, Minella E, Costenaro ER (2020) Genetic improvement of barley (*Hordeum vulgare* L.) in Brazil: yield increase and associated traits. Agric Sci 11(4):417–430. https://doi.org/10.4236/as.2020.114025

Rutkoski JE (2019) A practical guide to genetic gain. In *Advances in Agronomy* . Academic Press Inc. pp 217–249

Schmidt P, Hartung J, Rath J, Piepho H-P (2019) Estimating broad-sense heritability with unbalanced data from agricultural cultivar trials. Crop Sci. https://doi.org/10.2135/cropsci2018.06.0376

Sievert C (2020) Interactive web-based data visualization with R, plotly, and shiny. Chapman and Hall/CRC, Boca Raton

Smith AB, Ganesalingam A, Kuchel H, Cullis BR (2015) Factor analytic mixed models for the provision of grower information from national crop variety testing programs. Theor Appl Genet. 128(1):55–72. https://doi.org/10.1007/s00122-014-2412-x

Tessema BB, Liu H, Sørensen AC, Andersen JR, Jensen J (2020) Strategies using genomic selection to increase genetic gain in breeding programs for wheat. Front Genet. https://doi.org/10.3389/fgene.2020.578123

Thorwarth P, Ahlemeyer J, Bochard AM, Krumnacker K, Blümel H, Laubach E, Knöchel N, Cselényi L, Ordon F, Schmid KJ (2017) Genomic prediction ability for yield-related traits in German winter barley elite material. Theor Appl Genet 130(8):1669–1683. https://doi.org/10.1007/s00122-017-2917-1

Tiede T, Smith KP (2018) Evaluation and retrospective optimization of genomic selection for yield and disease resistance in spring barley. Mol Breed. https://doi.org/10.1007/s11032-018-0820-3

Tondelli A, Xu X, Moragues M, Sharma R, Schnaithmann F, Ingvardsen C, Manninen O, Comadran J et al (2013) Structural and temporal variation in genetic diversity of european spring two-row barley cultivars and association mapping of quantitative traits. Plant Genome. https://doi.org/10.3835/plantgenome2013.03.0007

Townley-Smith TF, Hurd EA (1973) Use of moving means in wheat yield trials. Can J Plant Sci 53(3):447–450

Wang J, Zhang Z (2021) GAPIT version 3: boosting power and accuracy for genomic association and prediction. Genomics Proteom Bioinform 19(4):629–640. https://doi.org/10.1016/j.gpb.2021.08.005

Werner CR, Gaynor RC, Gorjanc G, Hickey JM, Kox T, Abbadi A, Leckband G, Snowdon RJ, Stahl A (2020) How population structure impacts genomic selection accuracy in cross-validation: implications for practical breeding. Front Plant Sci. https://doi.org/10.3389/fpls.2020.592977

Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer-Verlag, New York

Wonneberger R, Schreiber M, Haaning A, Muehlbauer GJ, Waugh R, Stein N (2023) Major chromosome 5H haplotype switch structures the European two-rowed spring barley germplasm of the past 190 years. Theor Appl Genet. https://doi.org/10.1007/s00122-023-04418-7

Wych RD, Rasmusson DC (1983) Genetic improvement in malting barley cultivars since 1920. Crop Sci 23:1037–1040. https://doi.org/10.2135/cropsci1983.0011183X002300060004x

Zhong S, Dekkers JCM, Fernando RL, Jannink JL (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. Genetics 182(1):355–364. https://doi.org/10.1534/genetics.108.098277