



Unraveling genomic regions and optimizing genomic predictions for grain yield and yield-related traits in oats (*Avena sativa* L.)

Guilherme Oliveira¹ · Sumandeep Bazzar¹ · Jason Fiedler ·
Raja Sekhar Nandety¹ · Jean-Luc Jannik · Melanie Caffe¹

Received: 31 January 2025 / Accepted: 14 April 2025 / Published online: 16 May 2025
© The Author(s) 2025

Abstract Selecting for high grain yield is an intrinsic target of most crop breeding programs. However, the genetic complexity of the trait and its strong interactions with environmental factors pose significant challenges, hindering quick and efficient genetic improvement. This scenario is especially evident in oats (*Avena sativa* L.), as genetic gain for grain yield has historically been low when compared to major crops. The objective of this study was to leverage data collected in an oat breeding program, identifying molecular markers and genomic regions associated with grain yield and related traits and exploring if this genetic information could enhance genomic prediction models for oat grain yield. Association mapping was conducted to investigate genetic associations for grain yield, plant height, heading date, lodging,

snaphack, and crown rust severity in six oat breeding populations evaluated in multiple environments in South Dakota over six different years. Thirty-eight genomic regions associated with the traits were considered most promising because of their significance in multiple populations and/or environments, suggesting their potential as valuable genetic sources. When specific markers from the significant genomic regions were considered as fixed effects in linear and additive genomic prediction models, prediction accuracies for grain yield did not improve. However, using only the markers identified through association mapping, instead of the whole marker set, for model development improved prediction accuracy from approximately 4–6%. These results provide valuable information for increasing genetic gains for grain yield in oats and other crops.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10681-025-03526-3>.

G. Oliveira · S. Bazzar · M. Caffe (✉)
Department of Agronomy, Horticulture, and Plant Science,
South Dakota State University (SDSU), Brookings, SD,
USA
e-mail: melanie.caffe@sdstate.edu

J. Fiedler · R. S. Nandety
Cereal Crops Research Unit, USDA-ARS, Fargo, ND,
USA

J.-L. Jannik
R.W. Holley Center for Agriculture and Health, USDA-
ARS, Ithaca, NY, USA

Keywords *Avena sativa* L. · Grain yield · Genome-wide association study · Genomic prediction

Introduction

Oats (*Avena sativa* L.) the seventh most produced cereal globally (FAOSTAT 2021) is a versatile crop, suitable for both food and feed production, as well as for use as a cover crop. Oat consumption has increased due to its proven benefits in strengthening the immune system and reducing cholesterol and glycemic rates (Tiwari and Cummins 2011).

However, a closer look at the historical data from top oat-producing countries reveals a consistent decline in global oat cultivation area and production since 1960 (FAOSTAT 2021). Consequently, in the current scenario characterized by high demand and limited production areas, oat breeders need to strive for higher genetic gains, mainly in grain yield and grain yield-related traits, to enhance the farmer's production capabilities and economic return, and effectively meet the increasing oat demand. However, grain yield in oats has increased at a low rate when compared to other major cereal crops. According to historical world productions data, oat grain yield has only increased 139%, whereas corn, wheat, and barley yields, increased by 242, 246, and 175%, respectively, within the 1960–2005 timeframe (Strychar 2005). Although breeding alone does not fully explain the historically lower rates of grain yield improvement in oats relative to other crops, this data still supports the necessity for breeders to enhance grain yield in oats. Therefore, given the challenges associated with achieving high genetic gains in oats, a thorough genomic examination of grain yield and yield-related traits would provide valuable information for oat breeders to improve their approach for the development of oat varieties with superior grain yield.

Genome-wide association study (GWAS) has been widely utilized as a technique for identifying genomic regions of interest (Tibbs Cortes et al. 2021). The increasing adoption of GWAS, alongside other genomic mapping techniques such as QTL mapping, can be attributed to its ability to evaluate complex traits with a high mapping resolution, reduced research time for population development, and capacity to evaluate multiple traits within the same population (Zhu et al. 2008). GWAS aims to identify associations between the analyzed trait and a causal genomic variant that influences the phenotype, primarily through linkage disequilibrium (LD), which refers to the nonrandom association of alleles at two or more loci. Commonly, the genomic variants used in GWAS are single nucleotide polymorphism (SNP), due to their cost-effectiveness and the widespread distribution of these markers across the genome (Uffelmann et al. 2021). Diversity panels are often employed in GWAS to explore historical recombination events and to increase genetic diversity under evaluation (Khan et al. 2022). However, following this protocol, an extra step is required to validate the

identified causal genomic variants in breeding populations because, once identified in a diverse panel, they can already be fixed in adapted material or have differences in genetic effects due to differences in genome background. As an alternative, some studies have successfully utilized breeding populations in GWAS, leading to the discovery of more applicable information for direct implementation in plant breeding programs (Begum et al. 2015; Tyrka et al. 2023). While the genetic diversity is expected to be narrower in a breeding population as compared to a diversity panel, the genetic variations behind the existing phenotypic variability can be directly used for selection decisions in breeding programs.

Several GWAS studies focusing on oat yield-related traits have been published in recent years (Montilla-Bascón et al. 2015; Esvelt-Klos et al. 2016/7; Tumino et al. 2017). Not surprisingly, no GWAS has been conducted to evaluate grain yield in oats (grain produced per unit of area). This gap in research can be attributed to the challenges of identifying true associations for traits characterized by multiple genomic regions with small effects on the phenotype, such as grain yield. The GWAS methodology is directly related to the phenotypic variance explained by the marker, and associations are typically identified through significant differences in effects between two allelic variants for a specific marker. Consequently, GWAS on grain yield usually results in no associations or weak associations (Korte and Farlow 2013). However, GWAS research conducted in other cereals such as barley (*Hordeum vulgare* L.) and wheat (*Triticum aestivum* L.), demonstrated that marker associations for grain yield, when identified, are co-localized with strong association from other yield-related traits, usually traits with high heritability (Xu et al. 2018; Tsai et al. 2020; Khan et al. 2022). This scenario creates a new perspective and a new opportunity to conduct GWAS on oat grain yield and yield-related traits simultaneously, aiming to identify some genomic correlations and patterns among yield and yield-related traits.

Despite the identification of important genomic regions through GWAS in oats and the opportunities to consolidate and identify previous and/or new regions affecting oat grain yield, there are still challenges and questions about how to incorporate all the GWAS knowledge and information into applied breeding programs to improve selection. Looking for

potential genomic strategies to be used in association with GWAS, two main genomic strategies use molecular markers to improve selection and predict traits in plant breeding programs: marker-assisted selection (MAS) and genomic prediction (Rice and Lipka 2019). The efficiency of these strategies depends on the genetic architecture of the traits. MAS, which uses one or a few markers, is better suited for predicting qualitative traits or quantitative traits that contain one or a few major QTLs. In contrast, genomic prediction models utilize all the available molecular markers to predict the best individuals and have proved to be a superior strategy for quantitative traits such as grain yield (Meuwissen et al. 2001; Bernardo 2014). However, it may be valuable to combine information from both approaches, GWAS and genomic prediction. Studies in wheat, and in a non-domesticated crop, intermediate wheatgrass (*Thinopyrum intermedium*), have shown that genomic regions with a large effect on grain yield, as identified by GWAS when incorporated as fixed effects in genomic prediction models, increase accuracies in predicting superior lines for this trait (Bajgain et al. 2019; Sehgal et al. 2020). Another possibility is to use a subset of markers identified through GWAS, rather than all available markers, to attempt to achieve higher prediction accuracies (Alemu et al. 2024). The rationale is to retain only the relevant SNPs while discarding phenotypically neutral ones, thereby enhancing prediction accuracy by increasing the correlation between the markers and the phenotype (Alemu et al. 2024). However, the genetic architecture of the traits likely plays a significant role, as the selected SNPs would need to account

for nearly all the genetic variation of the trait to perform as well as or better than using all markers (Jeong et al. 2020). It is currently unexplored whether markers identified for oat grain yield and yield-related traits identified by GWAS can improve genomic prediction models for predicting oat yield, and it is currently unknown if some regions are significantly associated with oat grain yield and oat grain yield related traits across environments.

Therefore, the objectives of this study are to utilize GWAS to identify and dissect genomic regions associated with oat grain yield and other related traits in oat breeding populations across environments, and to explore whether these results can be incorporated in genomic prediction models to improve prediction accuracy for oat grain yield. For this, we leveraged existing breeding data collected for multiple sets of experimental lines in multiple environments.

Material and methods

Breeding datasets

The oat breeding program at South Dakota State University (SDSU) annually evaluates preliminary yield trials (PYTs) across multiple locations. For this study, we used breeding populations derived from PYTs evaluated in 2015, 2016, 2017, 2020, 2021, and 2022. Each PYT was conducted across a different range of testing sites (Table 1). The locations evaluated were Aurora (AUR: 44° 16' N, −96° 41' W), Beresford (BRF: 43° 04' N, −96° 46' W), Brookings (BRK: 44°

Table 1 Experiment locations, number of lines and SNPs markers summary

Year	Locations	Traits ^a	Number	
			Lines	SNP markers
2015	BRF, SSH, VLG, WIN	CRS, GYL, HDT, PHT, SPK	221	8299
2016	SSH, VLG, WIN	CRS, GYL, HDT, PHT, SPK	148	11,699
2017	BRF, SSH, VLG, WIN	CRS, GYL, HDT, PHT, SPK	225	11,356
2020	AUR, BRF, SSH, VLG, WIN	CRS, GYL, HDT, LDG, PHT, SPK	224	9301
2021	BRK, BRF, PIE, SSH, VLG	GYL, HDT, LDG, PHT, SPK	230	8625
2022	BRK, BRF, PIE, SSH, VLG	CRS, GYL, HDT, LDG, PHT, SPK	233	10,902

^aSome traits were not evaluated in some specific environments. For a detailed breakdown of traits evaluation in any environment, refer to the Supplementary Table 1 provided. AUR: Aurora, BRF: Beresford, BRK: Brookings, PIE: Pierre, SSH: South Shore, VLG: Volga, WIN: Winner. CRS: Crown rust severity, GYL: Grain yield, HDT: Heading date, LDG: Lodging, PHT: Plant height, SPK: Snapback

18° N, −96° 47" W), Pierre (PIE: 44° 22' N, −100° 21' W), South Shore (SSH: 45° 6' N, −96° 55' W), Volga (VLG −44° 19' N, −96° 55' W), and Winner (WIN: 43° 22' N, −99° 51' W) (Fig. 1). Plot sizes were either 1.5 by 1.8 m or 1.5 by 3.7 m depending on the location and year, and the seeding rate was 9 g per square meter. Temperature and rainfall were collected through weather stations present in proximity to the experimental sites (Supplementary Fig. 1). These experiments consist of 270 oat plots following an augmented design with four check cultivars, each replicated nine times. The checks used were the same across locations each year but changed over years and were as follow: Hayden (Caffe-Treml et al. 2017) was used as a check in 2015, 2016, 2017, 2020, 2021, and 2022; Horsepower in 2015, 2016, 2017, 2020, and 2021; Natty (Caffe-Treml et al. 2020) in 2015, 2016, 2017, and 2020; Shelby 427 in 2015, 2016, and 2017; Warrior in 2020, 2021, and 2022; MN-Pearl in 2021 and 2022; and Rushmore (Caffe et al. 2023) in 2022. Except for MN-Pearl, developed by the University of Minnesota, all the checks were developed by SDSU oat breeding program. In addition, due to external factors such as missing data and genotypic filtering applied in subsequent steps, the final number of breeding lines used varied across years (Table 1). Overall, the breeding datasets encompass 6 years, 26 environments (year-location combinations), and 1281 breeding lines (Table 1, Supplementary Table 1).

Phenotyping and data analysis

In total, six traits were evaluated: (i) oat grain yield (GYL) was measured in kilogram per hectare at

14% moisture; (ii) plant height (PHT) was determined by the distance in centimeters from the soil surface to the tip of the tallest panicle in the plants; (iii) heading date (HDT) was determined by the difference in days from the planting date to the day when 50% of the panicles emerged from the main stem in each plot; (iv) lodging (LDG) was measured in a scale 1 to 5 (1: no lodging, 5: all plants completely lodged in the plot), (v) snapback (SPK) was collected by drawing an oat row back and rating the straw stiffness following a scale of 1 to 5 (1: high strength and springiness of the stem, 5: low strength and springiness of the stem); and (vi) crown rust severity (CRS) was quantified visually based on the percentage of leaves covered with pustules for the plants inside the plots.

Boxplots were developed for the traits that were measured quantitatively, while bar graphs were created for the categorical traits, both using the R package ggplot2 (Wickham 2016). Given the unreplicated nature of the experiment, adjusted means, obtained from the best linear unbiased estimates (BLUEs), were calculated for all traits across locations within year (Eq. 1) and within each location (Eq. 2) using the lme4 R package (Bates et al. 2015):

$$y_{ij} = \mu + Gen_i + Loc_j + (Gen \times Loc)_{ij} + e_{ij} \quad (1)$$

$$y_{ijk} = \mu + Gen_i + Row_j + Column_k + e_{ijk} \quad (2)$$

where y is the phenotype, μ is the overall mean (intercept of the model), Gen_i is the effect of the genotypes, Row_j is the effect of the row, Loc_j is the effect of the location, $Column_k$ is the effect of the columns, $(Gen \times Loc)_{ij}$ is the interaction effect of genotype and location, and e is the error effect (residuals). All effects were treated as random, except for the genotype effect (fixed) in the BLUE value calculations. The across locations within year estimates were used to run a principal component analysis (PCA) biplots employed in the R packages FactoMineR and factoextra (Lê et al. 2008), discerning how the GYL is related to other related traits and how these traits can influence GYL under genotype-by-environment interactions, while the within each location estimates were used in the association mapping and genomic prediction analysis.



Fig. 1 Localization of South Dakota State University oat breeding experimental sites in South Dakota, USA. AUR: Aurora, BRF: Beresford, BRK: Brookings, PIE: Pierre, SSH: South Shore, VLG: Volga, WIN: Winner

Genotyping-by-sequencing

Genotyping-by-sequencing (GBS) was performed at the USDA-ARS Genotyping Lab (Fargo, North Dakota) following a protocol adapted from Poland et al. (2012), previously described in the literature (Huang et al. 2014), which uses the double restrictions enzymes, PstI and MspI. GBS libraries were sequenced with Illumina NextSeq 2000 system with 150 bp paired ends (Illumina, San Diego, USA). Sequence data were quality filtered and trimmed for adapters with *bbduk* (BBTools, <https://jgi.doe.gov/data-and-tools/bbtools>). The trimmed paired-end sequences were aligned to the oat reference genome, *Avena sativa* OT3098v2 (<https://wheat.pw.usda.gov/jb/?data=ggds/oat-ot3098v2-pepsico>) using Bowtie2 (Langmead and Salzberg 2012). The aligned reads were then sorted and indexed using SAMtools (Danecek et al. 2011, 2021), and vcf files were generated with mapping quality score of 30, using BCFtools (Danecek et al. 2021). The variant file was further filtered using vcftools, with a max-missing of 50%, a minor allele frequency (MAF) of 0.01, and a minimum read depth coverage of 4.

A total of 38,070 SNP markers comprised the initial genotypic data. The missing markers were imputed using the LD-kNNi LD KNNi method from TASSEL v 5.0 (Bradbury et al. 2007). For each year, the data was isolated and curated independently, focusing on the lines present in that year. The genotypes and SNPs within each genomic dataset were filtered using the following criteria: (1) genotypes that had more than 60% of SNPs data missing, (2) genotypes with heterozygosity proportion higher than 15%, (3) SNPs with heterozygosity higher than 5%, and (4) SNPs with MAF lower than 5%. After the filtering process, each genomic dataset was suitable for employment in GWAS, and the final number of SNPs is presented in Table 1.

Association mapping

The association between the SNPs markers and the traits analyzed was performed in the GAPIT R package (Wang and Zhang 2021). The association model selected was 'BLINK' because of its improved ability to avoid false positives and to identify true positive associations as compared to other models (Huang et al. 2019). More specifically, BLINK' select the

highest associated markers (pseudo quantitative trait nucleotides (QTNs)) based on fixed effect model (FEM) and LD information and conducts a second FEM to optimize the selected markers to get the best association predictions (Huang et al. 2019). This process is repeated several times until stabilization of the model (Huang et al. 2019). As a result, the 'BLINK' model increases the statistical power and decreases false associations, being a better option than commonly used mixed linear models.

Population structure was accounted for to understand the structure of SDSU breeding populations and determine the best parameters for the association analyses. Characterized by PCA analyses using the 'prcomp' function in R, all breeding populations showed a decrease in the genetic variance after the third or fourth principal component (Supplementary Fig. 2). Therefore, three principal components were used for the PYTs from 2015, 2016, 2020 and 2022, while four principal components were used for the PYTs from 2017 and 2021. Models using additional principal components (four or five, depending on the population) resulted in similar findings. To identify and classify SNP markers as significantly associated with specific traits, the threshold selected was the Bonferroni correction method. Additionally, we conducted a LD analysis, estimating the squared allele correlation frequency (r^2) for each chromosome separately and for the whole genome, using the LDcorSV R package (Mangin et al. 2012; Supplementary Fig. 3). Although some chromosomes from genome 'C' exhibited high LD decays, the average decay across chromosomes was between 25–33 Mbp (Supplementary Fig. 3), with the LD for the whole genome decaying at 28 Mbp. Thus, SNPs separated by less than 28 Mbp were considered to be located in the same genomic region. Genomic regions with multiple significance were contemplated if they harbored a minimum of two distinct significant markers, independently of whether these markers were identified for the same trait and/or environment.

Genomic prediction

Based on the phenotypic data analysis and GWAS association results, SNP markers that showed a shared association between GYL and other GYL-related traits, were selected to be used as fixed effects in genomic prediction (GP) scenarios. This

was done in addition to a ‘control’ genomic prediction scenario (GPM1) where no markers were fitted as fixed effects. Also, markers subsets (MS) to be used in the genomic prediction models as random effects were chosen based on the GWAS results. The first subset uses all the markers available (MS1), the second subset used only the markers identified by GWAS and present in the potential genomic regions (MS2), and the third subset used (MS3) all the markers inside the potential genomic regions. The genomic prediction model selected to run the scenarios was grounded in a linear mixed model that facilitates the partitioning of random and fixed effects (Eq. 3), where γ is the vector of phenotypic observations, X is a full-ranking design matrix for the fixed effects β , Z is the design matrix for the random effects u with $u \sim N(G\sigma^2, \mu)$, where G is a genomic relationship matrix calculated using the whole set of molecular markers available (Endelman 2011). In essence, what sets GPM1 apart from the other models is that while the latter incorporate the allele dosage of their respective fixed markers in the X matrix, removing these markers from the G matrix calculation, GPM1 includes all markers in the G matrix calculation. In the same sense, what sets MS1 apart for the other models is that the latter incorporated a subset of markers in the G matrix calculation, and MS1 includes all markers in the G matrix calculation.

$$\gamma = X\beta + Zu + e \quad (3)$$

To assess the effectiveness of each scenario in predicting grain yield, a five-fold cross-validation strategy on the PYT 2022 oat breeding populations datasets was employed. The rationale behind using breeding lines from the most recent PYT (2022) is the possibility to assess whether GWAS findings from the previous PYTs (previous breeding cycles) could be applied to the newest lines (current breeding cycle). The PYT 2022 dataset was partitioned randomly into five folds (80% training–20% testing), maintaining approximately the same number of oat genotypes in each fold. The same fold composition was used for all the different scenarios. Accuracies were measured using the Pearson correlation coefficient r between the adjusted phenotypes and the predicted phenotypes based on the estimated genetic values. Subsequently, mean accuracy

was calculated by averaging the results across all the five-fold. This strategy was replicated 100 times, ensuring different individuals’ combinations in the folds during the several replications. Therefore, the final accuracy of each genomic prediction model was the mean of 100 replications. It is important to mention that due to the incomplete similarity between the genomic datasets, specific significant markers, which theoretically should be tested as fixed effects or included in the marker subsets as random effects in a genomic prediction scenario, may not be tested because of their absence in the 2022 genomic dataset. Additionally, conducting a GWAS on a population and then using these markers as fixed effects in genomic prediction on the same population could lead to an overfitted model (Wray et al. 2013). Therefore, markers that showed a shared association between GYL and other GYL-related traits identified in 2022 were not considered as fixed effects for the described cross-validation scheme using the 2022 dataset, only markers identified in previous years were considered.

Results

Phenotypic variation in the oat breeding datasets

The SDSU breeding datasets showed a considerable phenotypic variation for all traits analyzed, despite being advanced breeding lines (Fig. 2). The lowest grain yields were obtained in WIN in 2017, averaging 2084 kg ha⁻¹, while the highest yields were observed in SSH in 2020, averaging 10,973 kg ha⁻¹ (Fig. 2). When comparing years, grain yields were lowest in 2021, with most of the location’s averages inferior to 4500 kg ha⁻¹ (Fig. 2). The contrast between 2021 and other years was also observed for PHT and HDT, with the shortest and earliest flowering plants observed in 2021 (Fig. 1). These results align with the weather data for 2021, particularly a drought during the month of June, which affected the growth and development of the plants in 2021 (Supplementary Fig. 1). Most of the environments in which crown rust was observed had an average CRS below 50%, except for VLG in 2016 which had a high crown rust pressure with an average severity of 61% (Fig. 1). Generally, the SPK scores were well distributed in all the environments analyzed, with scores 2 and 3 being the most common

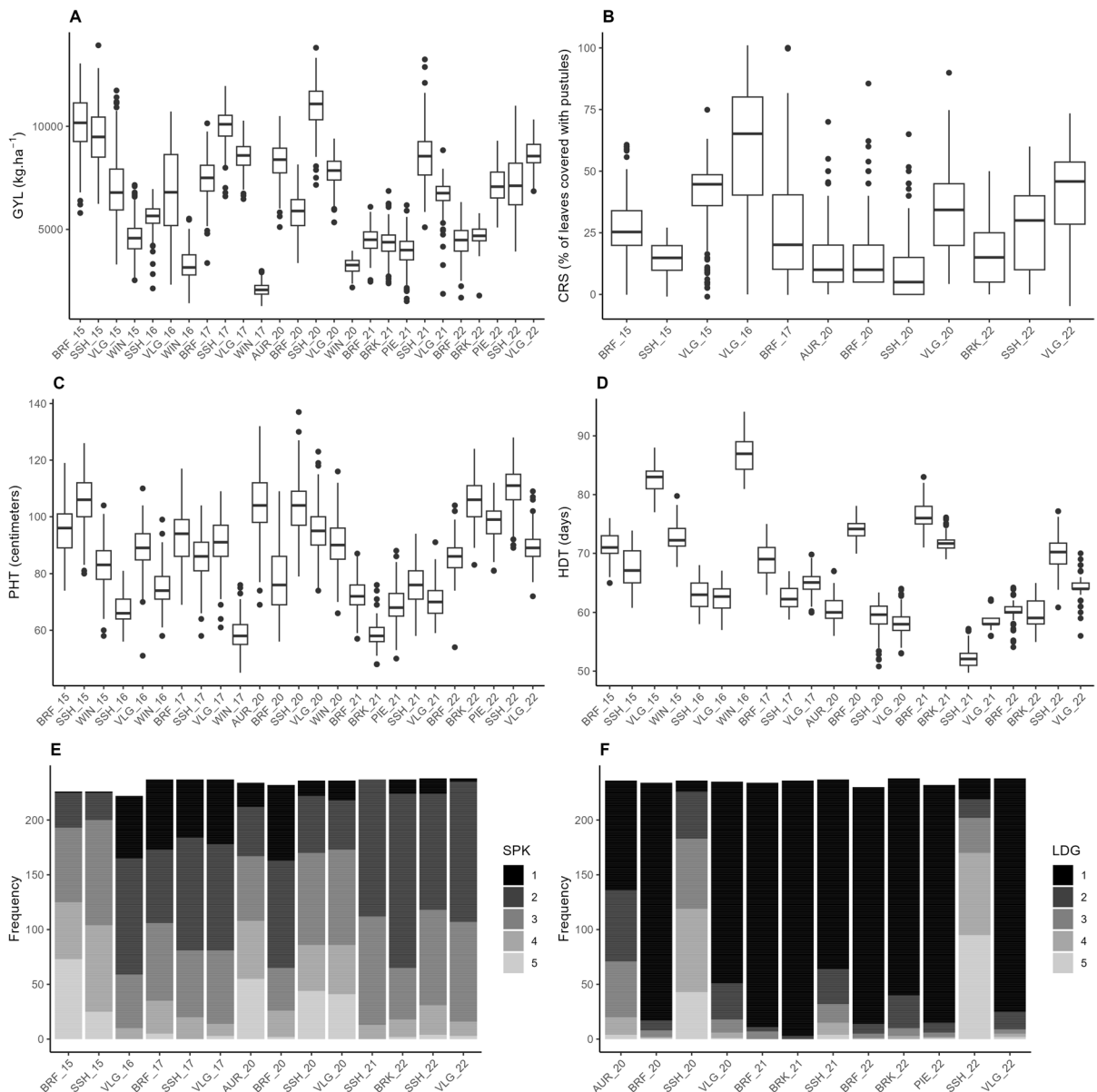


Fig. 2 Phenotypic distribution of grain yield (GYL) and GYL-related traits for South Dakota State University oat breeding populations. A=GYL, B=CRS (crown rust severity), C=PHT (plant height), D=HDT (heading date), E=SPK (snapback), F=LDG (Lodging) The evaluated locations were:

AUR: Aurora, BRF: Beresford, BRK: Brookings, PIE: Pierre, SSH: South Shore, VLG: Volga, WIN: Winner. The evaluated years were: 15=2015, 16=2016, 17=2017, 20=2020, 21=2021, 22=2022

among the different oat breeding populations (Fig. 2). For LDG, the scores varied considerably among environments, indicating significant environmental influence (Fig. 2). For example, two environments conducted in SSH, during the years 2020 and 2022, presented high degrees of 4 and 5 scores, meaning a

high percentage of plants in the plots were lodged in these specific environments (Fig. 2).

In years where environmental conditions were favorable to the development of CRS, the PCA biplot showed a highly negative significant correlation between CRS and GYL, with vectors

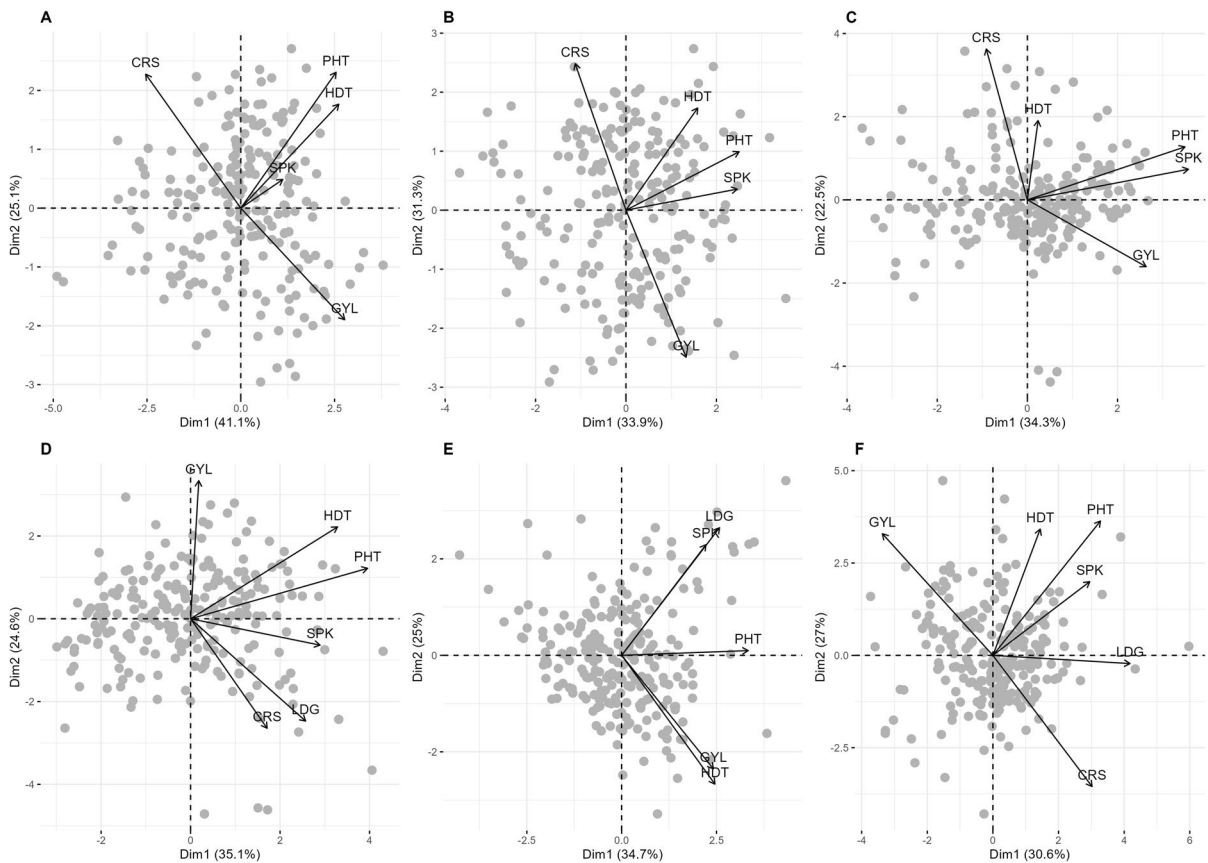


Fig. 3 PCA biplot of grain yield (GYL) and GYL-related traits for South Dakota State University oat breeding populations in six different years. A=2015, B=2016, C=2017,

D=2020, E=2021, F=2022. For a detailed breakdown of locations in each year, refer to the Supplementary Table 1 provided

for these traits in completely opposite directions based on their quadrants (Fig. 3). The relationship between GYL and the other traits was variable and depended on the year (Fig. 3). In 2015 and 2016, the GYL vector and the other trait vectors were close to 90° , indicating no correlation. However, in 2017, the angles between the GYL and SPK/PHT vectors narrowed, suggesting a higher correlation. The angle between the GYL and HDT vectors differed depending on the year and revealed a significant positive correlation between the two traits in 2021. The biplots also displayed a negative correlation between GYL and LDG in environments where LDG could be recorded. Finally, the correlations observed between PHT, SPK, and LDG varied to a degree but remained positive over the years.

Association mapping

A total of 193 SNPs were significantly associated with GYL and GYL-related traits over the different sets of breeding populations and environments (Supplementary Table 2). These markers were located on all of the 21 chromosomes, representing a wide distribution across the oat genome (Supplementary Table 2). The markers significantly associated with CRS ($n=35$), GYL ($n=32$), HDT ($n=38$), LDG ($n=28$), PHT ($n=43$), and SPK ($n=26$), were distributed across 14, 14, 14, 11, 16, and 14 chromosomes, and the markers with the highest variance had a PVE% value of 34.87, 34.81, 33.98, 59.18, 53.52, and 49.91%, respectively to each trait (Supplementary Table 2).

The significant molecular markers (193) were grouped into 76 genomic regions (Supplementary Table 2) based on their proximity on the oat genome and the extent of LD in those populations. However, some of these genomic regions were specific to a single marker being significant in a single environment (Supplementary Table 2) and consequently, were considered ‘weak’ regions. Those ‘weak’

genomic regions were filtered out and only genomic regions that had at least two distinct markers with high significance were considered, totaling 153 significant SNPs distributed on 38 regions (Figs. 4, 5, and Table 3). Despite the filter, the genomic regions are present on 18 oat chromosomes out of a total of 21, still representing a wide abundance of significant markers on the oat genome controlling

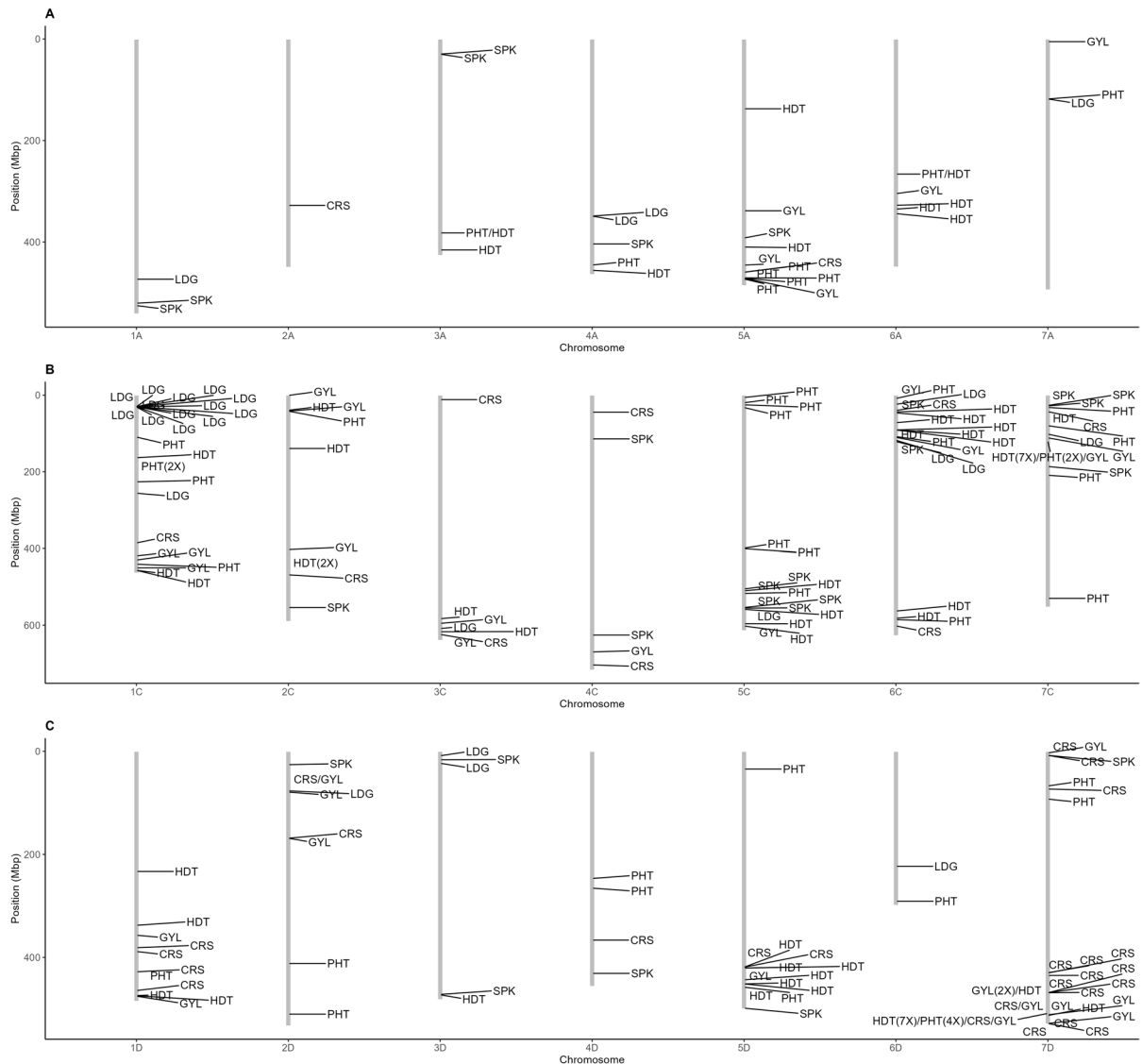


Fig. 4 Genomic mapping for the significant markers associated with grain yield (GYL), CRS (crown rust severity), PHT (plant height), HDT (heading date), SPK (snapback), and LDG (Lodging) in the South Dakota State University oat breeding populations. A = Genome A, B = Genome C, C = Genome D.

Lines with more than one trait indicate multiple associations by a single unique marker, and values in parentheses indicate how many times the marker was associated with the trait in different environments

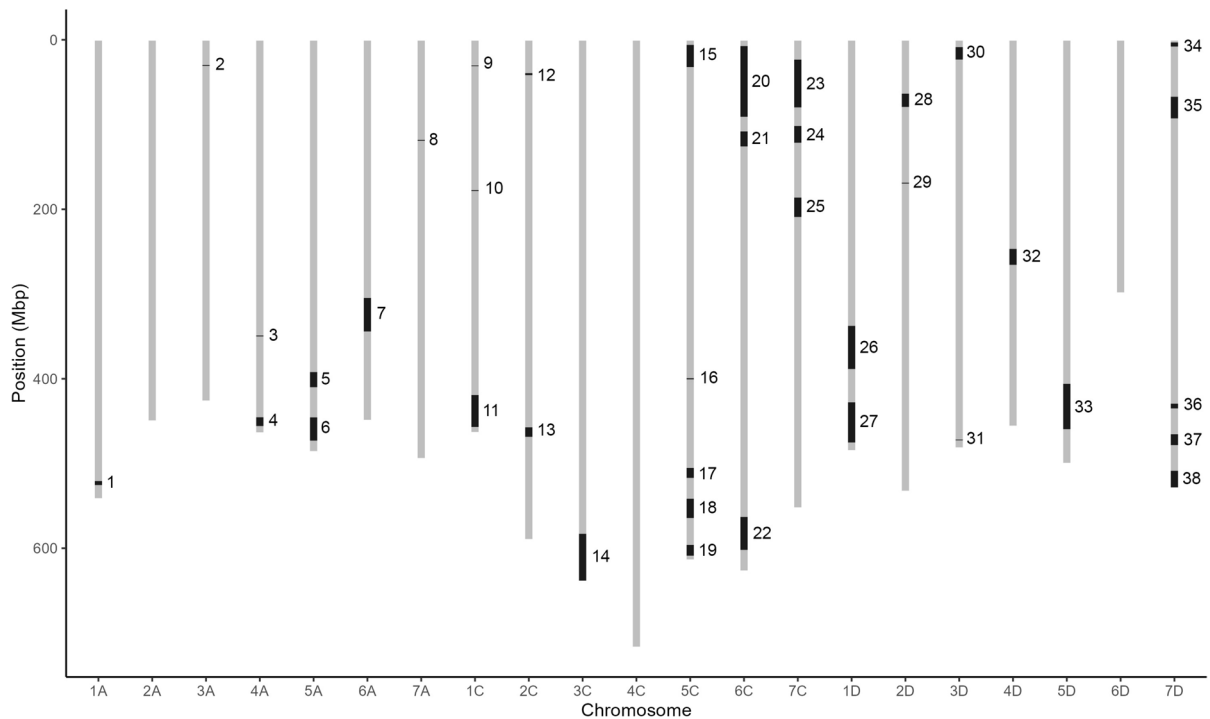


Fig. 5 Genomic regions controlling grain yield (GYL) and GYL-related traits in South Dakota State University oat breeding populations

GYL and GYL-related traits (Fig. 5 and Table 3). The number of markers within each region varies from 2 to 12 (Table 3). In addition, only considering the most significant marker within each region, the $-\log_{10}(p)$ values ranged from 5.43 to 22.28, while the PVE% was within the range of 4.35 to 59.18% (Table 3). A description of the markers identification for each trait follows:

Crown rust severity

After filtering, a total of 29 markers associated with CRS were identified (Fig. 4, Supplementary Table 1). Those markers were distributed across 16 genomic regions, presented on the chromosomes 5A, 2C, 3C, 6C, 7C, 1D, 2D, 5D, and 7D. Most of the markers were located on chromosome 7D, with 15 markers distributed in five genomic regions (Fig. 4, Supplementary Table 2). Among all the markers, the marker Schr7D_468526702 (chromosome 7D) was the only one identified in two environments: AUR and SSH in 2020.

Grain yield

27 markers were associated with GYL, encompassing 17 genomic regions over eleven oat chromosomes (5A, 6A, 1C, 2C, 3C, 5C, 6C, 7C, 1D, 2D, 5D, 7D). It is worth mentioning that only one marker was detected twice in the association mapping analysis: the marker Schr7D_466989127 (chromosome 7D). It was identified for BRF in 2017 and for AUR in 2020 (Supplementary Table 2).

Heading date

HDT was the trait with the highest number of genomic regions, with 33 markers detected across twenty genomic regions (Fig. 4, Supplementary Table 2). Additionally, it had the highest number of markers being detected in multiple environments, totaling six markers (Schr2C_457309640, Schr5D_420549221, Schr5D_451380445, Schr6C_90813775, Schr7C_121437154, Schr7D_508618782) (Supplementary Table 2). Interestingly, these six markers exhibit a year-association pattern. For instance,

the marker Schr6C_90813775 was exclusively identified in 2015 (SSH, VLG, and WIN), while Schr5D_451380445 was only detected in 2017 (SSH and VLG). Similarly, Schr2C_457309640 (SSH and VLG) and Schr5D_420549221 (BRK and BRF) were only identified in 2022. There are exceptions, such as the markers Schr7C_121437154, identified in multiple locations in 2020 and 2021, and Schr7D_508618782, detected for multiple locations in 2015, 2016, and 2017.

Lodging

LDG was the trait with the lowest number of genomic regions, with 24 markers detected across ten genomic regions over nine chromosomes (4A, 7A, 1C, 3C, 5C, 6C, 7C, 2D, 3D) (Fig. 4, Supplementary Table 2). None of the 24 markers were found more than once in the analyses, however, 12 of them were found on the same chromosome (1C) and for the same environment (BRF 2022).

Plant height

34 markers were significantly associated with PHT. These markers were distributed on chromosomes 4A, 5A, 7A, 1C, 2C, 5C, 6C, 7C, 1D, 4D, 7D (Fig. 4, Supplementary Table 2) across 21 genomic regions and included: Schr1C_177417158, identified for VLG and WIN in 2020; Schr7C_121437154, identified for BRF in 2020 and for SSH in 2022; and Schr7D_508618782, identified for BRF and SSH in 2015 and 2017. The last two markers were also associated with GYL/HDT and CRS/GYL/HDT,

respectively, in mostly the same environments (Supplementary Table 2).

Snapback

Our study identified 12 genomic regions containing markers associated with SPK (Fig. 4, Supplementary Table 1). These markers (19) were identified on the chromosomes 1A, 3A, 5A, 5C, 6C, 7C, 3D, 7D (Fig. 4, Supplementary Table 2). Each marker was uniquely detected within a singular environment.

Shared markers between grain yield and grain yield-related traits

Six markers (Schr7C_121437154, Schr2D_63564425, Schr7D_3202345, Schr7D_466989127, Schr7D_478260270, and Schr7D_508618782) were associated with GYL and one or more GYL-related traits in multiple environments, not necessarily in the same environment for all traits, representing potential candidates to be explored to improve oat grain yield (Table 2, Supplementary Table 2). Notably, Schr7C_121437154 exhibited a robust association, co-localizing with GYL, PHT and HDT across seven environments. Markers Schr2D_63564425, Schr7D_3202345, and Schr7D_478260270 displayed environment-specific associations with GYL and CRS in VLG (2016), VLG (2016), and SSH (2022), respectively, while Schr7D_466989127 was detected for GYL and PHT in three distinct environments (BRF in 2017, AUR in 2020, and PIE in 2022). Schr7D_508618782 demonstrated the broadest and strongest association, co-localizing with CRS,

Table 2 Description of the molecular markers with shared association between GYL and GYL-related traits

Marker	Chromosome	Position (bp)	Environment(s) association ^a	Other traits identified than GYL
Schr7C_121437154	7C	121,437,154	AUR_20, SSH_20, BRF_20, VLG_20, SSH_21, BRF_21, VLG_21	PHT and HDT
Schr2D_63564425	2D	63,564,425	VLG_16	CRS
Schr7D_3202345	7D	3,202,345	VLG_16	CRS
Schr7D_466989127	7D	466,989,127	BRF_17, AUR_20, PIE_22	HDT
Schr7D_478260270	7D	478,260,270	SSH_22	CRS
Schr7D_508618782	7D	508,618,782	BRF_15, SSH_15, VLG_15, WIN_15, SSH_16, WIN_16, BRF_17, SSH_17, VLG_17	CRS, HDT and PHT

^aThe three letters represent the location, while the last two represent the year. AUR: Aurora, BRF: Beresford, PIE: Pierre, SSH: South Shore, VLG: Volga, WIN: Winner. 15: 2015, 16: 2016, 17: 2017, 20: 2020, 21: 2021, 22 = 2022

HDT, and PHT across nine environments. Given Schr7D_508618782 significance, its allele distributions were analyzed to explore the relationship between this marker and these traits (Supplementary Fig. 4). The presence of the variant ‘CC’, compared to the ‘GG’, results in higher PHT, later HDT, higher GYL and lower CRS medians (Supplementary Fig. 4).

Genomic prediction with association mapping information

Genomic prediction strategies were evaluated to determine if the GWAS results identified in this study, conducted over multiple environments, and with different breeding populations, could be used to improve prediction accuracy for GYL (Fig. 6). To complete this task, markers Schr2D_63564425 (GPM2), Schr7D_3202345 (GPM3), and

Schr7D_508618782 (GPM4) were fitted as fixed effects in separate models using the 2022 population. Markers Schr7C_121437154 (absent in the 2022 genomic dataset), Schr7D_466989127 (associated with PHT in the 2022 breeding population), and Schr7D_478260270 (associated with GYL in the 2022 breeding population) were not included in this analysis. Additionally, three marker subsets—MS1 (all markers, $n=10,902$), MS2 (only markers significantly associated with GYL and/or GYL related traits, $n=132$), and MS3 (markers in genomic regions associated with GYL and/or GYL related traits, $n=1636$)—were tested alongside the fixed-effect markers, resulting in fifteen genomic prediction strategies. Comparing the models that incorporated markers as fixed effects (GPM2, GPM3, GPM4) with the control model that did not incorporate markers as fixed effects (GPM1), no increase in prediction accuracy was observed (Fig. 6), regardless of the marker

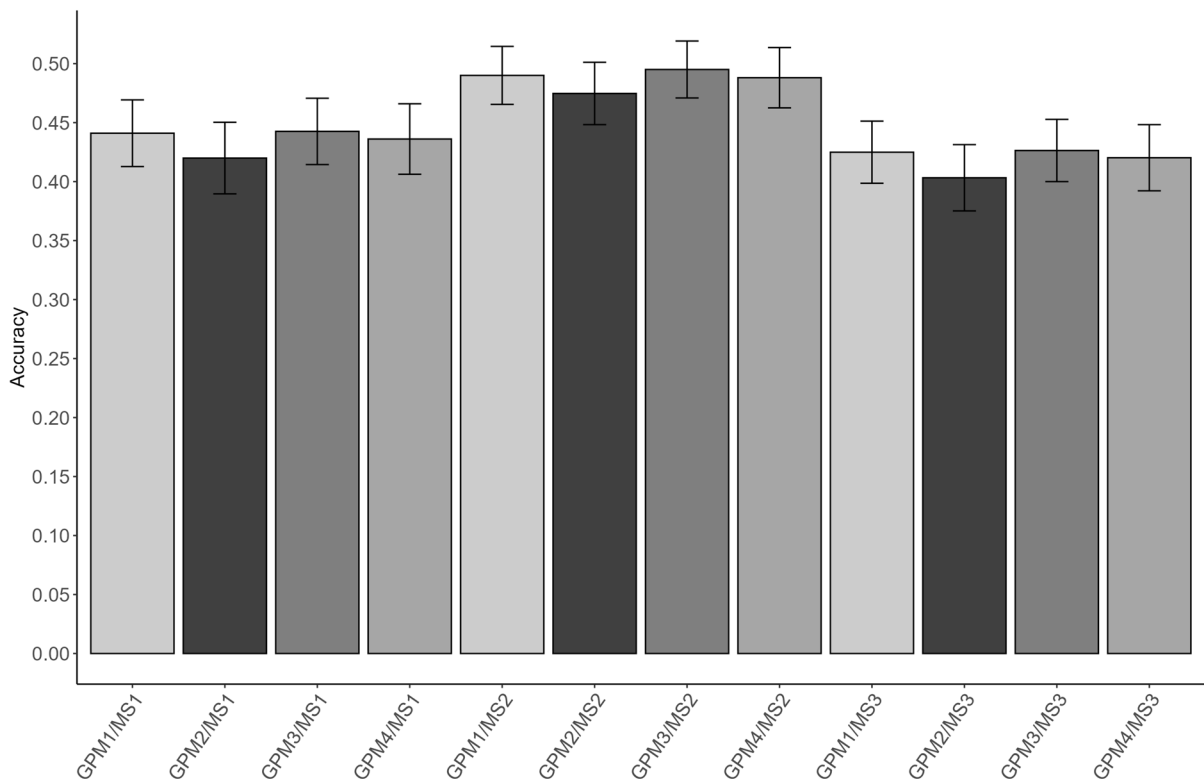


Fig. 6 Grain yield (GYL) predictive accuracies of genomic prediction models using different sets of significant markers as random and fixed effects. GPM1=genomic prediction model 1 (no markers as fixed effect), GPM2=genomic prediction model 2 (Schr2D_63564425 as fixed effect), GPM3=genomic

prediction model 3 (Schr7D_3202345 as fixed effect), GPM4=genomic prediction model 4 (Schr7D_508618782 as fixed effect), MS1=marker subset 1 ($n=10,902$), MS2=marker subset 2 ($n=132$), MS3=marker subset 3 ($n=1636$)

Table 3 Major regions containing significant SNP for grain yield and grain yield-related traits in oats

Chr	Trait/Environment(s) ^a	ID	Gen. region	N	Sig. SNPs	Gen. region (Mbp)		-log ₁₀ (p)	PVE (%)	Previous identification for the same trait(s)
						Start	End			
1A	SPK_SSH15, SPK_VLG15	1	520.7	2	2	525.4	5.47–7.93	2.30–19.47		
3A	SPK_BRF20	2	29.7	2	2	29.7	19.20–22.28	10.34–49.94		
4A	LDG_BRK21	3	348.9	2	2	348.9	5.39–5.43	0.0–38.56		Wooten et al. (2009)
4A	HDT_BRF21, PHT_BRF21	4	445.4	2	2	445.4	5.30–7.15	5.69–19.71		
5A	HDT_SSH15, SPK_VLG16	5	392.1	2	2	409.9	6.61–6.93	5.85–27.96		
5A	GYL_BRF20, CRS_BRF20, PHT_VLG20, PHT_BRF15, PHT_BRK22, PHT_SSH21, GYL_SSH21, PHT_SSH22	6	445.6	7	7	472.9	5.43–5.77	3.87–15.52		Siripoonwiwat et al. (1996), Holland et al. (1997), Jackson et al. (2008)
6A	GYL_AUR20, HDT_VLG15, HDT_SSH15, HDT_PIE22	7	304.6	4	4	344.1	5.37–7.13	1.38–4.92		Siripoonwiwat et al. (1996), Portyanko et al. (2005), Nava et al. (2012), Herrmann et al. (2014)
7A	LDG_VLG22, PHT_VLG17	8	118.1	2	2	118.5	5.94–5.98	3.90–20.30		Herrmann et al. (2014)
1C	LDG_BRF22	9	30.2	12	12	31	5.56–12.45	0.0–59.18		Tumino et al. (2017)
1C	HDT_VLG20, PHT_VLG20, PHT_WIN20	10	163.2	2	2	177.4	5.38–10.60	22.43–52.37		Sunstrum et al. (2019)
1C	GYL_BRF17, GYL_BRK21, PHT_BRF20, GYL_BRF17, HDT_SSH22, HDT_VLG22	11	419.2	6	6	456.8	5.45–10.59	3.40–24.14		Sunstrum et al. (2019)
2C	HDT_VLG15, PHT_AUR20, GYL_AUR20	12	39.4	3	3	41.7	5.53–12.28	2.45–7.19		Esvelt-Klos et al. (2017)
2C	HDT_SSH22, HDT_VLG22, CRS_VLG22	13	457.3	2	2	468.5	6.14–8.00	1.48–22.83		Siripoonwiwat et al. (1996), Holland et al. (1997), Tumino et al. (2016), Canales et al. (2021)
3C	HDT_SSH16, GYL_SSH22, LDG_VLG21, HDT_VLG15, CRS_BRF17, GYL_WIN20	14	583.0	6	6	638.1	5.36–7.31	0.0–34.81		Siripoonwiwat et al. (1996), Holland et al. (1997), Jackson et al. (2010), McNish et al. (2020)
5C	PHT_SSH15, PHT_VLG17, PHT_SSH22, PHT_PIE21	15	6.0	4	4	32.2	6.31–7.04	3.55–15.64		Siripoonwiwat et al. (1996)
5C	PHT_BRK21, PHT_SSH21	16	399.3	2	2	400.7	6.14–7.85	15.30–29.58		Siripoonwiwat et al. (1996)
5C	SPK_SSH15, SPK_VLG20, HDT_SSH21, PHT_SSH15	17	505.2	4	4	516.9	5.54–8.36	0.0–20.28		Siripoonwiwat et al. (1996)
5C	SPK_BRF20, HDT_AUR20, LDG_VLG20	18	541.6	5	5	564.3	6.02–30.92	0.0–20.02		Siripoonwiwat et al. (1996), Holland et al. (1997)
5C	HDT_VLG15, HDT_BRF22, GYL_PIE22	19	596.1	3	3	608.8	5.88–6.20	2.38–9.07		Siripoonwiwat et al. (1996), Holland et al. (1997), Bekele et al. (2018)
6C	GYL_VLG15, PHT_VLG22, LDG_SSH21, SPK_VLG20, CRS_VLG15, HDT_SSH15, HDT_WIN15, HDT_SSH17, HDT_VLG15, HDT_WIN15	20	7.4	9	9	90.8	5.34–14.72	2.63–28.23		Hizbai et al. (2012)

Table 3 (continued)

Chr	Trait/Environment(s) ^a	ID	N	Gen. region (Mbp)		-log10(p)	PVE (%)	Previous identification for the same trait(s)	
				Start	End				
6C	PHT_BRF21, GYL_AUR20, LDG_VLG21, LDG_PIE22, SPK_BRF20	21	5	108.1	125.8	5.44–23.75	0.0–13.61		
6C	HDT_BRF15, PHT_SSH21, CRS_BRF15	22	4	563.1	601.9	5.55–8.76	1.65–18.61		
7C	SPK_BRF15, SPK_VLG22, SPK_BRF17, PHT_SSH20, CRS_SSH20, HDT_SSH20, PHT_BRF22	23	7	23.4	79.7	5.50–7.90	3.55–37.11	Holland et al. (1997), Holland et al. (2002), De Koeyer et al. (2004), Esvelt-Klos et al. (2017)	
7C	LDG_VLG22, GYL_VLG21, HDT_AUR20, HDT_SSH20, HDT_BRF20, PHT_BRF20, HDT_VLG20, HDT_SSH21, PHT_SSH21, GYL_SSH21, HDT_BRF21, HDT_VLG21	24	3	101.7	121.4	5.80–13.81	6.48–33.46		
7C	SPK_BRF20, PHT_SSH22	25	2	186.2	209.1	5.74–14.44	0.00–4.37		
1D	HDT_BRF22, GYL_SSH15, CRS_VLG22, CRS_VLG15	26	4	337.7	388.5	5.36–7.28	1.89–13.05	Siripoonwiwat et al. (1996), Holland et al. (1997), Winkler et al. (2016),	
1D	CRS_BRF15, PHT_VLG17, CRS_SSH22, HDT_PIE22, HDT_SSH22, GYL_BRK21	27	6	427.8	475.1	5.41–6.17	2.35–9.54	Siripoonwiwat et al. (1996), Holland et al. (1997), De Koeyer et al. (2001), Esvelt-Klos et al. (2017)	
2D	CRS_VLG16, GYL_VLG16, LDG_SSH21, GYL_SSH15	28	3	63.6	79.2	5.58–7.38	5.48–12.48	Winkler et al. (2016), Esvelt-Klos et al. (2017)	
2D	GYL_BRF21, CRS_VLG15	29	2	168.7	168.7	5.87–7.26	0.0–4.35	Babiker et al. (2015)	
3D	LDG_VLG22, SPK_SSH15, LDG_PIE22	30	3	8.6	23.3	6.06–7.33	6.16–14.53	Tumino et al. (2017)	
3D	SPK_VLG22, HDT_AUR20	31	2	47.2	47.3	5.58–7.21	3.83–4.40	Holland et al. (2002)	
4D	PHT_BRK22, PHT_BRF15	32	2	24.7	26.6	8.86–9.80	6.30–11.68		
5D	CRS_VLG22, HDT_VLG15, CRS_VLG15, HDT_BRK22, HDT_BRF22, GYL_PIE22, HDT_VLG21, HDT_SSH17, HDT_VLG17, PHT_SSH15, HDT_WIN16	33	11	406.0	459.5	5.44–11.11	3.79–20.46	Siripoonwiwat et al. (1996), Holland et al. (1997), Holland et al. (2002), Tanhuanpää et al. (2012), Esvelt-Klos et al. (2016)	
7D	CRS_VLG16, GYL_VLG16, SPK_BRF20, CRS_BRF15	34	4	3.2	7.9	7.78–8.83	0.0–22.54	Harder et al. (1980), Siripoonwiwat et al. (1996), De Koeyer et al. (2001), Wight et al. (2004), Jackson et al. (2010), Sunstrum et al. (2019), Admassu-Yimer et al. (2022)	
7D	PHT_WIN17, CRS_BRF15, PHT_VLG16	35	3	67.2	92.7	5.69–17.41	0.0–39.98	Harder et al. (1980), De Koeyer et al. (2001), Wight et al. (2004), Jackson et al. (2010), Sunstrum et al. (2019), Admassu-Yimer et al. (2022)	

Table 3 (continued)

Chr	Trait/Environment(s) ^a	ID	N	Gen. region (Mbp)		-log10(p)	PVE (%)	Previous identification for the same trait(s)
				Start	End			
7D	CRS_VLG15, CRS_BRF15, CRS_VLG15, CRS_VLG22	36	4	429.4	434.9	12.77–15.10	4.78–19.61	Harder et al. (1980), Siripoonwiwat et al. (1996), Jackson et al. (2010), Wight et al. (2004), Sunstrum et al. (2019), Admassu-Yimer et al. (2022)
7D	CRS_BRF17, GYL_BRF17, GYL_BRF17, GYL_AUR20, HDT_PIE22, CRS_AUR20, CRS_SSH20, CRS_BRF20, CRS_SSH22, GYL_SSH22	37	6	465.5	478.2	5.75–17.41	3.70–24.65	Harder et al. (1980), Siripoonwiwat et al. (1996), Holland et al. (1997), De Koeyer et al. (2001), Holland et al. (2002), Wight et al. (2004), Locatelli et al. (2006), Jackson et al. (2010), Esvelt-Klos et al. (2017), Bekele et al. (2018), Sunstrum et al. (2019), Admassu-Yimer et al. (2022)
7D	HDT_SSH15, PHT_SSH15, PHT_SSH15, CRS_BRF15, GYL_BRF15, HDT_BRF15, PHT_BRF15, HDT_VLG15, HDT_WIN15, HDT_SSH16, HDT_WIN16, HDT_SSH17, PHT_SSH17, PHT_BRF17, HDT_VLG17, GYL_BRF21, GYL_PIE22, HDT_SSH17, CRS_BRK22, GYL_SSH22, CRS_VLG22, CRS_SSH22	38	8	508.6	528.1	5.26–18.39	0.0–55.17	Harder et al. (1980), Siripoonwiwat et al. (1996), Holland et al. (1997), De Koeyer et al. (2001), Holland et al. (2002), Wight et al. (2004), Locatelli et al. (2006), Jackson et al. (2008), Jackson et al. (2010), Esvelt-Klos et al. (2017), Bekele et al. (2018), Tumino et al. (2017), Zimmer et al. (2018), Sunstrum et al. (2019), McNish et al. (2020)

^aThe first three letters represent the trait; the subsequent three letters represent the location, while the last two numbers represent the year. The evaluated locations were: AUR: Aurora, BRF: Beresford, BRK: Brookings, PIE: Pierre, SSH: South Shore, VLG: Volga, WIN: Winner. The evaluated years were: 15 = 2015, 16 = 2016, 17 = 2017, 20 = 2020, 21 = 2021, 22 = 2022

subset explored (MS1, MS2, or MS3). Additionally, a decrease in prediction accuracy was noted in GPM2 (Schr2D_63564425 as fixed effect) compared to GPM1. However, differences in prediction among the marker subset were observed. Notably, MS2, which used the lowest number of markers in the *G* matrix, increased prediction accuracy of approximately 4–6% compared to the models in MS1 and MS3.

Discussion

Data collected as part of breeding program activities can be leveraged to increase our understanding of the genetic architecture of important traits, to identify genetic variants that are directly relevant to genetic improvement efforts, and to evaluate the applicability of new breeding methods (Bazzer et al. 2025). The current research assessed GYL and GYL-related traits in six SDSU PYT oat populations, which amounted to 1281 oat lines and consisted of phenotypic data collected in 26 environments. The phenotypic variability for GYL and GYL-related traits within the breeding populations studied was broad and the phenotypic variability was also wide among the set of environments considered (Fig. 2). Large phenotypic variation within these populations is not only essential for successful association mapping analysis, but it is fundamental to the oat breeding program, as genetic variability within breeding populations is necessary for achieving genetic gain and oat improvement (Swarup et al. 2021).

We evaluated the phenotypic and genotypic relationships between GYL and GYL-related traits within the SDSU oat breeding target mega-environments to gain insights to develop approaches to increase genetic gain for GYL in oats. GYL variability and genotype by environment interactions can be attributed to the presence or absence of favorable alleles for GYL-related traits which affect directly or indirectly the genotypes responses to abiotic and biotic stresses. Uncovering/exploring these genomic regions can be a key to keep developing high yield oat varieties. The significant relationship between GYL and CRS in some environments is an example of the effect of biotic stress on GYL. In environments where crown rust incidence was sufficiently high to be recorded in the field, there was a high negative association between GYL and CRS,

meaning that most of the genotypes with high GYL exhibited low CRS (Fig. 3). This pattern was also observed when performing GWAS for those environments, with four markers (Schr2D_63564425, Schr7D_478260270, Schr7D_3202345 and Schr7D_508618782) associated with both GYL and CRS (Table 2). Additional markers significantly associated with GYL were in the same genomic regions as markers significantly associated with CRS (Table 3). These observations agree with the pleiotropic effect of crown rust resistance genes on grain yield.

The positive correlation between GYL and HDT in 2021 (Fig. 3) is an example of the effect of abiotic stress on GYL within the SDSU oat breeding target mega-environments. There was a significant positive correlation between GYL and HDT in 2021 (Fig. 3). Higher GYL was observed for breeding lines with later heading dates as compared to the early flowering lines. Based on observations made in the field and weather data, we suspect that the positive correlation between GYL and HDT that year was related to early lines suffering a severe heat and drought stress, affecting grain yield, whereas the late lines were able to escape as they had not yet reached a critical phenological stage during the peak of heat and drought conditions and were able to benefit from rain and lower heat prior to reaching this critical phenological stage (Fig. 2, Supplementary Fig. 1). Similarly, the relationship between HDT and GYL was observed in the GWAS results for 2021; we identified a shared marker between GYL and HDT for 2021, Schr7C_121437154.

The phenotypic correlation between GYL and PHT, SPK, and LDG was low and variable depending on the environment. None of the significant markers for LDG and SPK co-localized with markers associated with GYL, nor did they co-localize with each other. However, markers on 3C, 6C, 7C, 2D, and 7D were significantly detected in the same genomic regions for GYL and/or LDG/SPK. The severity of the lodging may not have been sufficient to considerably affect GYL in some environments while the lodging may have been too severe in other environments to differentiate among genotypes (Figs. 2 and 3), decreasing the capability to identify shared co-localization between LDG and GYL. However, a strong genomic region for lodging was identified on chromosome 1C for the BRF location in 2022 (Table 3,

Fig. 4). This region overlaps with one of the associations found by Tumino et al. (2017).

Some genomic regions controlling GYL and GYL-related traits in this study were not reported previously, mainly the regions on genome A (Table 3). Interestingly, most of these genomic regions contain only markers associated with SPK, a trait not widely studied in oats (Table 3). This limited research is likely one of the reasons for the lack of broad knowledge about genomic regions affecting it (Table 3). However, overall, most of the genomic regions were not novel and have been found previously by other researchers (Table 3). The genomic regions on the chromosome 7D stand out due to the high number of significant markers for CRS, GYL, PHT and HDT and their detection across environments in our study (Figs. 4, 5 and Table 3). Previous reports suggest the potential localization of crown rust resistance genes (*Pc38*, *Pc54*, *Pc58*, *Pc62*, *Pc63*, *Pc96*) on chromosome 7D (Wight et al. 2004, Esvelt-Klos et al. 2017, Admassu-Yimer et al. 2022, Abdullah et al. 2023). Four out of six significant markers affecting GYL and at least one other trait in our study (Schr7D_3202345, Schr7D_466989127, Schr7D_478260270, and Schr7D_508618782) were located on this chromosome (Table 2). Additionally, numerous loci were previously reported on chromosome 7D for other traits than those analyzed in our study (Wight et al. 2024), including Barley yellow dwarf virus (BYDV) resistance (Siripoonwiwat et al. 1996; Barbosa-Neto et al. 2000). These findings indicate the importance of chromosome 7D on GYL in oats and suggest that the genetic variability in the 7D region identified in different genetic background is also present in the SDSU advanced breeding material (Table 3) and can be exploited within the oat breeding program. Interestingly, recent findings showed that the chromosome 7D suffers from a lack of recombination, probably caused by historical inversions and other genetic rearrangements on this chromosome (Tinker et al. 2022). However, this scenario may create a challenging barrier to finding segregating lines with high grain yield, crown rust resistance, short stature, and early maturity, as the allele for marker Schr7D_508618782 that was associated with increased CR resistance (lowest CRS) was also associated with increased plant height and later maturity (Supplementary Fig. 4), which are both undesirable characteristics in the target mega-environments of the breeding program.

Other genomic regions caught our attention, specifically some regions on chromosomes 7C and 2D. These regions included markers Schr7C_121437154 and Schr2D_63564425, which were associated with GYL and at least one other trait in our study. The chromosome 2D potentially harbors different crown rust resistance genes, such as *Pc45*, *Pc53*, and *PcKM* (Park et al. 2022; Wight et al. 2024). Based on current knowledge, it remains difficult to confirm whether these genes are in the same Schr2D_63564425 genomic region. For the genomic region on the chromosome 7C, most of the markers, including Schr7C_121437154, were identified for HDT in 2020 or 2021 (Supplementary Table 2), in drought environment. Interestingly, this is the first report of markers associated with GYL or GYL-related traits within this genomic region. This leads us to suspect that drought conditions in these specific environments played a significant role in the observed associations. For this specific case, we scanned the GrainGenes T3/Oat database, which contains an inventory of oat genes according to the *Avena sativa* OT3098 oat reference genome (v2) (<https://wheat.pw.usda.gov/jb/?data=/ggds/oat-ot3098v2-pepsico>). Within this region, we identified a gene AVESA.00001b.r3.7Cg0001447 (5863 bp, located at chr7C:121419863..121425725) related with the enzyme production 'Dehydrogenase/reductase SDR family member 7'. Notably this gene is located only 11,429 bp away from the Schr7C_121437154 marker. In *Arabidopsis thaliana* L., SDR-related activities were associated with drought stress due to their role in abscisic acid (ABA) biosynthesis (Cheng et al. 2002).

Significant genomic regions were also identified on other chromosomes. The wide range of significant SNPs found (153), as well as the large number of significant genomic regions (38) indicates the complex genetic architecture of these GYL and GYL related traits and confirms the large effects of genotype $\times$ environment interactions, specifically for GYL (Korte and Farlow 2013). The potential genotype $\times$ environment effect on the associations, aligned to the use of different populations, can also be observed due to the high number of markers or genomic regions that are not consistently detected across the multi-environment trials (Supplementary Table 2). This pattern was observed not only for GYL but also for the GYL-related traits, which typically exhibit higher heritability than GYL (Supplementary

Table 2). These findings reinforce the importance of conducting GWAS in multiple environments to identify potential genetic variants underlying the traits analyzed. Additionally, the complexity of these traits highlights the challenges in implementing strategies, such as MAS, that rely on major genes or QTLs. Instead, results confirm that prioritization of genomic prediction methods would be more effective. However, the GWAS detected strong and shared associations between traits in SDSU oat breeding populations (Table 3), and this information could be considered in genomic prediction model development. Markers that showed shared associations with GYL and other GYL-related traits were prioritized for exploration as fixed effects, while markers subsets to be used as random effects were selected based on our GWAS results.

In our attempt to account for GWAS results in a linear and additive genomic prediction model, models that incorporated specific markers as fixed effects did not increase grain yield predictions when compared to our control models across different marker subsets scenarios (Fig. 6). Rice and Lipka (2019) analyzing simulated traits for maize and sorghum, demonstrated that no increase or even decreases in prediction accuracies occur when significant markers were fitted as fixed effects, particularly when the genetic architecture of the traits became more complex. Our results are consistent with this scenario; although we identified strong genomic regions associated with GYL-related traits, their effects as fixed factors were not large enough to significantly impact the modelling and, consequently, the predictions accuracies for a complex trait such as GYL. On the other hand, prediction accuracies were considerably affected by marker density in the model (Fig. 6). Optimizing marker density and distribution has been extensively explored as means to enhance genomic prediction accuracies in several crops (Norman et al. 2018; DoVale et al. 2022; Aalborg et al. 2024). However, using GWAS results as criteria to reduce the number of markers for genomic prediction models has not been widely exploited (Jeong et al. 2020). In our study, the genomic prediction models performed slightly better when only the markers identified through GWAS ($n=132$) were used, compared to the full set of available markers ($n=10,902$) (Fig. 6). These results primarily support the importance of these markers identified via GWAS in influencing GYL in oats.

This small subset effectively captured the genetic variation for this trait without compromising accuracy. Although a much smaller subset ($n=132$), the markers selected were obtained by performing GWAS for GYL and GYL-related traits in 26 different environments, therefore, likely capturing important loci contributing to GYL in a range of environmental conditions. However, several implications and limitations must be considered when using a small subset of markers to predict GYL over the long term. First, the introduction of new genetic variability in breeding programs could potentially reduce the efficiency of the selected markers over time, as newly important regions would not be accounted for. Second, selection pressure in breeding program could fix favorable alleles, thereby reducing the effects of some markers in the model. Third, in our case, markers from specific chromosomes were not included, for example the chromosomes 2A, 3A, 4C and 6D, which could be a barrier to obtaining higher prediction accuracies. Further research is needed to optimize and fully understand the short- and long-term effects of using markers selected through GWAS in genomic prediction approaches.

Conclusion

In summary, the association mapping performed across different environments in an array of oat breeding populations revealed the genetic complexity involved in controlling GYL in oats. The GWAS identified 153 markers across 38 genomic regions significantly associated with GYL and related traits. Many of these genomic regions have been previously recognized by other researchers. Our results contribute to a more comprehensive understanding of the genetic architecture and genomic regions affecting GYL-related traits. Integrating GWAS information to enhance GYL prediction accuracy was not successful when specific markers from certain genomic regions were incorporated into the genomic prediction models as fixed effect, but an increase in accuracy was observed when only the significant markers were considered in the model as compared to all markers. These markers were associated with GYL-related traits. Results also highlight the importance of performing GWAS in multiple environments to thoroughly detect regions associated with the trait over a

range of environmental conditions. Although the proposed methodology was successful in increasing prediction accuracy, results indicated that it is crucial to select specific markers carefully and understand the effects of those genomic regions not only on GYL but also on a range of breeding-targets to avoid genetic drift and a reduction in phenotypic variability.

Acknowledgements We want to acknowledge Nicholas Hall and part-time student employees for conducting the field trials and assisting with the collection of phenotypic data. Thank you to Terrance Peterson and Mary Osenga at the North Central Small Grains Genotyping Laboratory for acquiring genotyping data.

Author contributions G.O and M.C contributed to the study of conception and design. Material preparation, data collection and analysis were performed by G.O., M.C., S.B., J. F., R. S. N. and J.L.J. The first draft of the manuscript was written by G.O. and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This study was supported by the South Dakota Agricultural Experiment Station, South Dakota State University Agronomy Horticulture and Plant Science Department, the National Institute of Food and Agriculture under hatch proposal SD00H529-14, the South Dakota Crop Improvement Association, Grain Millers, General Mills Foundation, and USDA-ARS Project #3060-21000-046-000-D.

Data availability The genotyping dataset is available on T3/Oat (triticeaetoolbox.org).

Declarations Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

Conflict of interest The authors have no relevant financial or non-financial interests to disclose

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

- Aalborg T, Sverrisdóttir E, Kristensen HT, Nielsen KL (2024) The effect of marker types and density on genomic prediction and GWAS of key performance traits in tetraploid potato. *Front Plant Sci.* <https://doi.org/10.3389/fpls.2024.1340189>
- Abdullah S, Gordon T, Admassu-Yimer B, Paczos-Grzęda E, Harrison SA, Menzies JG, Esvelt-Klos K (2023) Mapping and identification of molecular markers for the *Pc96* gene conferring resistance to crown rust. *PLoS ONE* 18:e0283769. <https://doi.org/10.1371/journal.pone.0283769>
- Admassu-Yimer B, Klos KE, Griffiths I, Cowan A, Howarth C (2022) Mapping of crown rust (*Puccinia coronata* f. sp. avenae) resistance gene *Pc54* and a novel quantitative trait locus effective against powdery mildew (*Blumeria graminis* f. sp. avenae) in the Oat (*Avena sativa*) line *Pc54*. *Phytopathology* 112:1316–1322. <https://doi.org/10.1094/PHYTO-10-21-0445-R>
- Alemu A, Åstrand J, Montesinos-López OA, Sánchez JJ, Fernández-González J, Tadasse W, Vetukuri RR, Carlsson AS, Ceplitis A, Crossa J, Ortiz R, Chawade A (2024) Genomic selection in plant breeding: key factors shaping two decades of progress. *Mol Plant* 17:552–578. <https://doi.org/10.1016/j.molp.2024.03.007>
- Bajgain P, Zhang X, Anderson JA (2019) Genome-wide association study of yield component traits in intermediate wheatgrass and implications in genomic selection and breeding. *G3 (Bethesda)* 9:2429–2439. <https://doi.org/10.1534/g3.119.400073>
- Babiker EM, Gordon TC, Jackson EW, Chao S, Harrison SA, Carson ML, Obert DE, Bonman JM (2015) Quantitative trait loci from two genotypes of oat (*Avena sativa*) conditioning resistance to *Puccinia coronata*. *Phytopathology* 105:239–245. <https://doi.org/10.1094/PHYTO-04-14-0114-R>
- Barbosa-Neto JF, Siripoonwiwat W, O'Donoghue LS, Gray SM, Smith DM, Kolb FL, Gourmet C, Brown CM, Sorrells ME (2000) Chromosomal regions associated with barley yellow dwarf virus resistance in oat. *Euphytica* 114:67–76. <https://doi.org/10.1023/A:1003936220509>
- Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear mixed-effects models using lme4. *J Stat Softw* 67:1–48
- Bazzzer SK, Oliveira G, Fiedler JD, Nandety RS, Jannink J-L, Caffè M (2025) Genomic strategies to facilitate breeding for increased β -Glucan content in oat (*Avena sativa* L). *BMC Genom* 26:35. <https://doi.org/10.1186/s12864-024-11174-5>
- Begum H, Spindel JE, Lalusin A, Borromeo T, Gregorio G, Hernandez J, Virk P, Collard B, McCouch SR (2015) Genome-wide association mapping for yield and other agronomic traits in an elite breeding population of tropical rice (*Oryza sativa*). *PLoS ONE* 10:e0119873. <https://doi.org/10.1371/journal.pone.0119873>
- Bekele WA, Wight CP, Chao S, Howarth CJ, Tinker NA (2018) Haplotype-based genotyping-by-sequencing in oat genome research. *Plant Biotechnol J* 16:1452–1463. <https://doi.org/10.1111/pbi.12888>

- Bernardo R (2014) Genomewide selection when major genes are known. *Crop Sci* 54:68–75. <https://doi.org/10.2135/cropsci2013.05.0315>
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635. <https://doi.org/10.1093/bioinformatics/btm308>
- Caffe-Treml M, Hall L, Bauer R, Kleinjan J, Hall N, Ingemansen JA (2017) Registration of oat cultivar ‘Hayden.’ *J Plant Registrations* 11:95–99. <https://doi.org/10.3198/jpr2016.10.0056rcr>
- Caffe-Treml M, Hall L, Bauer R, Hall N, Kleinjan J, Graham C, Ingemansen JA, Turnipseed B, Krishnan P (2020) Registration of oat cultivar ‘Natty.’ *J Plant Registrations* 14:281–287. <https://doi.org/10.1002/plr2.20074>
- Caffe M, Hall L, Hall N, Bauer R, Kleinjan J, Graham C, Ingemansen JA, Turnipseed B, Krishnan P (2023) Registration of oat cultivar ‘Rushmore.’ *J Plant Registrations* 17:247–254. <https://doi.org/10.1002/plr2.20282>
- Canales FJ, Montilla-Bascón G, Bekele WA, Howarth CJ, Langdon T, Rispail N, Tinker NA, Prats E (2021) Population genomics of Mediterranean oat (*A. sativa*) reveals high genetic diversity and three loci for heading date. *Theor Appl Genet* 134:2063–2077. <https://doi.org/10.1007/s00122-021-03805-2>
- Cheng W-H, Endo A, Zhou L, Penney J, Chen H-C, Arroyo A, Leon P, Nambara E, Asami T, Seo M, Koshihara T, Sheen J (2002) A unique short-chain dehydrogenase/reductase in arabidopsis glucose signaling and abscisic acid biosynthesis and functions. *Plant Cell* 14:2723–2743. <https://doi.org/10.1105/tpc.006494>
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H (2021) Twelve years of SAMtools and BCFtools. *GigaScience*. <https://doi.org/10.1093/gigascience/giab008>
- De Koeber DL, Phillips RL, Stuthman DD (2001) Allelic shifts and quantitative trait loci in a recurrent selection population of oat. *Crop Sci* 41:1228–1234. <https://doi.org/10.2135/cropsci2001.4141228x>
- De Koeber DL, Tinker NA, Wight CP, Deyl J, Burrows VD, O'Donoghue LS, Lybaert A, Molnar SJ, Armstrong KC, Fedak G, Wesenberg DM, Rosnagel BG, McElroy AR (2004) A molecular linkage map with associated QTLs from a hullless × covered spring oat population. *Theor Appl Genet* 108:1285–1298. <https://doi.org/10.1007/s00122-003-1556-x>
- DoVale JC, Carvalho HF, Sabadin F, Fritsche-Neto R (2022) Genotyping marker density and prediction models effects in long-term breeding schemes of cross-pollinated crops. *Theor Appl Genet* 135:4523–4539. <https://doi.org/10.1007/s00122-022-04236-3>
- Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome*. <https://doi.org/10.3835/plantgenome2011.08.0024>
- Esvelt Klos K, Huang Y-F, Bekele WA, Obert DE, Babiker E, Beattie AD, Bjørnstad Å, Bonman JM, Carson ML, Chao S, Gnanesh BN, Griffiths I, Harrison SA, Howarth CJ, Hu G, Ibrahim A, Islamovic E, Jackson EW, Jannink J-L, Kolb FL, McMullen MS, Mitchell Fetch J, Murphy JP, Ohm HW, Rines HW, Rosnagel BG, Schlueter JA, Sorrells ME, Wight CP, Yan W, Tinker NA (2016) Population genomics related to adaptation in elite oat germplasm. *Plant Genome* 9. <https://doi.org/10.3835/plantgenome2015.2010.0103>
- Esvelt-Klos K, Yimer BA, Babiker EM, Beattie AD, Bonman JM, Carson ML, Chong J, Harrison SA, Ibrahim AMH, Kolb FL, McCartney CA, McMullen M, Fetch JM, Mohammadi M, Murphy JP, Tinker NA (2017) Genome-wide association mapping of crown rust resistance in oat elite germplasm. *Plant Genome*. <https://doi.org/10.3835/plantgenome2016.10.0107>
- FAOSTAT (2021) Crops and livestock products. Food and Agriculture Organization of the United States <https://www.fao.org/faostat/en/#data/QCL>. Accessed 5 May 2024
- Harder DE, McKenzie RIH, Martens JW (1980) Inheritance of crown rust resistance in three accessions of *Avena sterilis*. *Can J Genet Cytol* 22:27–33. <https://doi.org/10.1139/g80-005>
- Herrmann MH, Yu J, Beuch S, Weber WE (2014) Quantitative trait loci for quality and agronomic traits in two advanced backcross populations in oat (*Avena sativa* L.). *Plant Breed* 133:588–601. <https://doi.org/10.1111/pbr.12188>
- Hizbati BT, Gardner KM, Wight CP, Dhanda RK, Molnar SJ, Johnson D, Frégeau-Reid J, Yan W, Rosnagel BG, Holland JB, Tinker NA (2012) Quantitative trait loci affecting oil content, oil composition, and other agronomically important traits in oat. *Plant Genome*. <https://doi.org/10.3835/plantgenome2012.07.0015>
- Holland J, Portyanko V, Hoffman D, Lee M (2002) Genomic regions controlling vernalization and photoperiod responses in oat. *Theor Appl Genet* 105:113–126. <https://doi.org/10.1007/s00122-001-0845-5>
- Holland JB, Moser HS, O'Donoghue LS, Lee M (1997) QTLs and epistasis associated with vernalization responses in oat. *Crop Sci* 37:1306–1316. <https://doi.org/10.2135/cropsci1997.0011183X003700040047x>
- Huang M, Liu X, Zhou Y, Summers RM, Zhang Z (2019) BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience*. <https://doi.org/10.1093/gigascience/giy154>
- Huang YF, Poland JA, Wight CP, Jackson EW, Tinker NA (2014) Using genotyping-by-sequencing (GBS) for genomic discovery in cultivated oat. *PLoS ONE* 9:e102448. <https://doi.org/10.1371/journal.pone.0102448>
- Jackson EW, Obert DE, Avant JB, Harrison SA, Chong J, Carson ML, Bonman JM (2010) Quantitative trait loci in the Ogle/TAM O-301 oat mapping population controlling resistance to *Puccinia coronata* in the field. *Phytopathology* 100:484–492
- Jackson EW, Obert DE, Menz M, Hu G, Bonman JM (2008) Qualitative and quantitative trait loci conditioning resistance to *Puccinia coronata* pathotypes NQMG and LGCG in the oat (*Avena sativa* L.) cultivars Ogle and TAM O-301. *Theor Appl Genet* 116:517–527. <https://doi.org/10.1007/s00122-007-0687-x>

- Jeong S, Kim JY, Kim N (2020) GMStool: GWAS-based marker selection tool for genomic prediction from genomic data. *Sci Rep* 10:19653. <https://doi.org/10.1038/s41598-020-76759-y>
- Khan H, Krishnappa G, Kumar S, Mishra CN, Krishna H, Devate NB, Rathana ND, Parkash O, Yadav SS, Srivastava P, Biradar S, Kumar M, Singh GP (2022) Genome-wide association study for grain yield and component traits in bread wheat (*Triticum aestivum* L.). *Front Genet.* <https://doi.org/10.3389/fgene.2022.982589>
- Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9:29. <https://doi.org/10.1186/1746-4811-9-29>
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>
- Lê S, Josse J, Husson F (2008) FactoMineR: an R package for multivariate analysis. *J Stat Softw* 25:1–18
- Locatelli AB, Federizzi LC, Milach SCK, Wight CP, Molnar SJ, Chapados JT, Tinker NA (2006) Loci affecting flowering time in oat under short-day conditions. *Genome* 49:1528–1538. <https://doi.org/10.1139/g06-108>
- Mangin B, Siberchicot A, Nicolas S, Doligez A, This P, Cierco-Ayrolles C (2012) Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity* 108:285–291. <https://doi.org/10.1038/hdy.2011.73>
- McNish IG, Zimmer CM, Susko AQ, Heuschele DJ, Tiede T, Case AJ, Smith KP (2020) Mapping crown rust resistance at multiple time points in elite oat germplasm. *Plant Genome* 13:e20007. <https://doi.org/10.1002/tpg2.20007>
- Meuwissen TH, 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>
- Montilla-Bascón G, Rispail N, Sánchez-Martín J, Rubiales D, Mur LAJ, Langdon T, Howarth CJ, Prats E (2015) Genome-wide association study for crown rust (*Puccinia coronata* f. sp. *avenae*) and powdery mildew (*Blumeria graminis* f. sp. *avenae*) resistance in an oat (*Avena sativa*) collection of commercial varieties and landraces. *Front Plant Sci* 6. <https://doi.org/10.3389/fpls.2015.00103>
- Nava IC, Wight CP, Pacheco MT, Federizzi LC, Tinker NA (2012) Tagging and mapping candidate loci for vernalization and flower initiation in hexaploid oat. *Mol Breed* 30:1295–1312. <https://doi.org/10.1007/s11032-012-9715-x>
- 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/Genetics* 8(9):2889–2899. <https://doi.org/10.1534/g3.118.200311>
- Park RF, Boshoff WHP, Cabral AL, Chong J, Martinelli JA, McMullen MS, Fetch JWM, Paczos-Grzęda E, Prats E, Roake J, Sowa S, Ziems L, Singh D (2022) Breeding oat for resistance to the crown rust pathogen *Puccinia coronata* f. sp. *avenae*: achievements and prospects. *Theor Appl Genetics* 135:3709–3734. <https://doi.org/10.1007/s00122-022-04121-z>
- Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* 7:e32253. <https://doi.org/10.1371/journal.pone.0032253>
- Portyanko VA, Chen G, Rines HW, Phillips RL, Leonard KJ, Ochocki GE, Stuthman DD (2005) Quantitative trait loci for partial resistance to crown rust *Puccinia coronata* in cultivated oat *Avena sativa* L. *Theor Appl Genetics* 111(2):313–324. <https://doi.org/10.1007/s00122-005-2024-6>
- Rice B, Lipka AE (2019) Evaluation of RR-BLUP genomic selection models that incorporate peak genome-wide association study signals in maize and sorghum. *Plant Genome* 12:180052. <https://doi.org/10.3835/plantgenom.e2018.07.0052>
- Sehgal D, Rosyara U, Mondal S, Singh R, Poland J, Dreisigacker S (2020) Incorporating genome-wide association mapping results into genomic prediction models for grain yield and yield stability in CIMMYT spring bread wheat. *Front Plant Sci* 11:197. <https://doi.org/10.3389/fpls.2020.00197>
- Siripoonwiwat W, O'Donoghue LS, Wesenberg D, Hoffman DL, Barbosa-Neto JF, Sorrells ME (1996) Chromosomal regions associated with quantitative traits in oat. *J Agric Genom* 2:1–13
- Strychar R (2005) World oat production, trade and usage. In: Webster FH, Woods PJ (eds) *Oats: chemistry and technology*, 2nd edn. American Association of Cereal Chemist, St. Paul, pp 1–10
- Sunstrum FG, Bekele WA, Wight CP, Yan W, Chen Y, Tinker NA (2019) A genetic linkage map in southern-by-spring oat identifies multiple quantitative trait loci for adaptation and rust resistance. *Plant Breed* 138:82–94. <https://doi.org/10.1111/pbr.12666>
- Swarup S, Cargill EJ, Crosby K, Flagel L, Kniskern J, Glenn KC (2021) Genetic diversity is indispensable for plant breeding to improve crops. *Crop Sci* 61:839–852. <https://doi.org/10.1002/csc2.20377>
- Tanhuanpää P, Manninen O, Beattie A, Eckstein P, Scoles G, Rosnagel B, Kiviharju E (2012) An updated doubled haploid oat linkage map and QTL mapping of agronomic and grain quality traits from Canadian field trials. *Genome* 55:289–301. <https://doi.org/10.1139/g2012-017>
- Tibbs Cortes L, Zhang Z, Yu J (2021) Status and prospects of genome-wide association studies in plants. *Plant Genome* 14:e20077
- Tinker NA, Wight CP, Bekele WA, Yan W, Jellen EN, Renhuldt NT, Sirijovski N, Lux T, Spannagl M, Mascher M (2022) Genome analysis in *Avena sativa* reveals hidden breeding barriers and opportunities for oat improvement. *Commun Biol* 5:474. <https://doi.org/10.1038/s42003-022-03256-5>
- Tiwari U, Cummins E (2011) Meta-analysis of the effect of β -glucan intake on blood cholesterol and glucose levels. *Nutrition* 27:1008–1016. <https://doi.org/10.1016/j.nut.2010.11.006>
- Tsai H-Y, Janss LL, Andersen JR, Orabi J, Jensen JD, Jahoor A, Jensen J (2020) Genomic prediction and GWAS of yield, quality and disease-related traits in spring barley and winter wheat. *Sci Rep* 10:3347. <https://doi.org/10.1038/s41598-020-60203-2>
- Tumino G, Voorrips RE, Rizza F, Badeck FW, Morcia C, Ghizzoni R, Germeier CU, Paulo M-J, Terzi V, Smulders MJM

- (2016) Population structure and genome-wide association analysis for frost tolerance in oat using continuous SNP array signal intensity ratios. *Theor Appl Genet* 129:1711–1724. <https://doi.org/10.1007/s00122-016-2734-y>
- Tumino G, Voorrips RE, Morcia C, Ghizzoni R, Germeier CU, Paulo M-J, Terzi V, Smulders MJM (2017) Genome-wide association analysis for lodging tolerance and plant height in a diverse European hexaploid oat collection. *Euphytica* 213:163. <https://doi.org/10.1007/s10681-017-1939-8>
- Tyrka M, Krajewski P, Bednarek PT, Rączka K, Drzazga T, Matysik P, Martofel R, Woźna-Pawlak U, Jasińska D, Niewińska M, Ługowska B, Ratajczak D, Sikora T, Witkowski E, Dorczyk A, Tyrka D (2023) Genome-wide association mapping in elite winter wheat breeding for yield improvement. *J Appl Genet* 64:377–391. <https://doi.org/10.1007/s13353-023-00758-8>
- Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, Martin HC, Lappalainen T, Posthuma D (2021) Genome-wide association studies. *Nat Rev Methods Primers* 1:59. <https://doi.org/10.1038/s43586-021-00056-9>
- Wickham H (2016) *ggplot2: Elegant graphics for data analysis*. Springer, New York. https://doi.org/10.1007/978-3-319-24277-4_9
- Wang J, Zhang Z (2021) GAPIT Version 3: boosting power and accuracy for genomic association and prediction. *Genom Proteom Bioinform* 19:629–640. <https://doi.org/10.1016/j.gpb.2021.08.005>
- Winkler LR, Michael Bonman J, Chao S, Admassu Yimer B, Bockelman H, Esvelt Klos K (2016) Population structure and genotype-phenotype associations in a collection of oat landraces and historic cultivars. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2016.01077>
- Wight CP, O'Donoghue LS, Chong J, Tinker NA, Molnar SJ (2004) Discovery, localization, and sequence characterization of molecular markers for the crown rust resistance genes *Pc38*, *Pc39*, and *Pc48* in cultivated oat (*Avena sativa* L.). *Mol Breed* 14:349–361. <https://doi.org/10.1007/s11032-004-0148-z>
- Wight CP, Blake VC, Jellen EN, Yao E, Sen TZ, Tinker NA (2024) One hundred years of comparative genetic and physical mapping in cultivated oat (*Avena sativa*). *Crop Pasture Sci*. <https://doi.org/10.1071/CP23246>
- Wooten DR, Livingston DP III, Lyerly HJ, Holland JB, Jellen EN, Marshall DS, Murphy JP (2009) Quantitative trait loci and epistasis for oat winter-hardiness component traits. *Crop Sci* 49:1989–1998. <https://doi.org/10.2135/cropsci2008.10.0612>
- Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM (2013) Pitfalls of predicting complex traits from SNPs. *Nat Rev Genet* 14:507–515. <https://doi.org/10.1038/nrg3457>
- Xu X, Sharma R, Tondelli A, Russell J, Comadran J, Schnaithmann F, Pillen K, Kilian B, Cattivelli L, Thomas WTB, Flavell AJ (2018) Genome-wide association analysis of grain yield-associated traits in a Pan-European barley cultivar collection. *Plant Genome* 11:170073. <https://doi.org/10.3835/plantgenome2017.08.0073>
- Zimmer CM, Ubert IP, Pacheco MT, Federizzi LC (2018) Molecular and comparative mapping for heading date and plant height in oat. *Euphytica* 214:101. <https://doi.org/10.1007/s10681-018-2182-7>
- Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. *Plant Genome* 1:5–20. <https://doi.org/10.3835/plantgenome2008.02.0089>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.