Genome-Wide Association Studies and Genomic Selection for Grain Protein Content Stability in a Nested Association Mapping Population of Spring Wheat

Karansher S. Sandhu ^a, Paul D. Mihalov ^b, Megan J. Lewien ^c, Michael O. Pumphrey ^a, Arron H. Carter ^{a*}

^a Department of Crop and Soil Sciences, Washington State University, Pullman, WA, USA 99164

^b Dewey Scientific, Pullman, WA, USA 99164

^c U.S. Forest Service, Eugene, OR, USA 97401

* Correspondence:

Dr. Arron H. Carter

ahcarter@wsu.edu

Keywords: Finlay-wilkinson regression, genome-wide association study, genomic selection, grain protein content, nested association mapping, ridge regression best linear unbiased prediction, wheat

Abstract

Grain protein content (GPC) is controlled by complex genetic systems and their interactions, and is an important quality determinant for hard spring wheat as it has a positive effect on bread and pasta quality. GPC is variable among genotypes and strongly influenced by environment. Thus, understanding the genetic control of wheat GPC and identifying genotypes with improved stability is an important breeding goal. The objectives of this research were to identify genetic backgrounds with less variation for GPC across environments and identify quantitative trait loci (QTLs) controlling the stability of GPC. A spring wheat nested association mapping (NAM) population of 650 recombinant inbred lines (RIL) derived from 26 diverse founder parents crossed to one common parent, 'Berkut', was phenotyped over three years of field trials (2014-2016). Genomic selection models were developed and compared based on prediction of GPC and GPC stability. After observing variable genetic control of GPC within the NAM population, seven RIL families displaying reduced marker-by-environment interaction were selected based on a stability index derived from Finlay-Wilkinson regression. A genome-wide association study identified seven significant QTLs for GPC stability with a Bonferroni-adjusted P value <0.05 . This study also demonstrated that genome-wide prediction of GPC with ridge regression best linear unbiased estimates reached up to $r = 0.69$. Genomic selection can be used to apply selection pressure for GPC and improve genetic gain for GPC.

Introduction

Grain protein content (GPC) is a high-priority determinant of end-use quality for most cereals (Shewry and Halford 2002), including pasta (*Triticum turgidum* L.) and bread wheat (*Triticum aestivum* L.) where higher GPC is preferred. The large dependence on wheat, rice (*Oryza sativa* L.), and maize (*Zea mays* L.) as primary sources of carbohydrates results in their consumption meeting 80% of dietary protein requirements for humans (Shewry 2007). Compared to other agricultural commodities, cereal grains contain a relatively lower concentration of protein. In a screening of 12,600 lines from the USDA world wheat collection, GPC varied from 7% to 22%, with the genetic component accounting for only a third of the variation (Vogel et al. 1976). Breeding efforts to improve GPC have been difficult due to strong environmental influences and the high variability of GPC across years and locations (Löffler and Busch 1982) combined with variable economic value-based trade-offs between starch and protein yields in grains (DePauw et al. 2007).

Phenotypic plasticity refers to the flexibility for exhibiting changes to the different environments for a particular trait (Kroon et al. 2005). Improved understanding of genotype by environment interactions could provide new strategies for breeding crop varieties that remain stable for performance between changing environments (Kusmec et al. 2017). One of the key challenges to identifying genes controlling environmental stability lies in the quantification of stability as a trait (Kliebenstein et al. 2002). There are various approaches for measuring environmental stability including the use of variance components of individuals across environments (represented by coefficient of variation), the comparison of mean responses of genotypes to the overall mean of individuals in the trial (calculated as coefficient of regression on environment index) (Lin et al. 1986), and expected change in performance of genotype as function of environment effect (Finlay and Wilkinson 1963). Compared to unstructured genotype by environment interaction models, Finlay Wilkinson (FW) regression fits every level of genotype and environment and reveals genotype performance across environments (Lian 2016). GPC stability can potentially help select stable genotypes across increasingly unpredictable environments (Ordas et al. 2008). Various biparental and association mapping studies identified QTLs controlling GPC, but none of them focus on the stability of GPC (Blanco et al. 1996; Huang et al. 2010).

Nested association mapping (NAM) is a multi-parental design with higher allelic variation than bi-parental populations and stronger statistical power than association mapping populations, combining the advantages of each approach (Zhu et al. 2008). A universal parent is crossed to multiple genotypes, followed by inbreeding to make a combination of both full-sib and half-sib recombinant inbred lines (RILs) (Yu and Buckler 2006). NAM population have proven their success in mapping complex traits in barley (*Hordeum vulgare* L.) (Saade et al. 2016), maize (Gage et al. 2020), rice (Fragoso et al. 2017), and arabidopsis (*Arabidopsis thaliana*) (Li et al. 2011). The NAM population design mirrors the structure of several breeding programs where a few superior, diverse, or exotic lines are crossed with elite breeding lines during the pre-breeding pipelines. The cross of exotic germplasm with elite breeding lines for population development aids in normalizing genetic background and selecting segregating alleles in favor of the adapted parent alleles (Blanc et al. 2006). The resolution and power of NAM populations allow for the assessment of complex traits like GPC stability in structured germplasm via genome-wide association studies (GWAS).

Joint linkage association mapping is applied in multi-parental mapping populations where QTL terms are nested within families (McMullen et al. 2009) instead of testing marker effects across the families as in GWAS (Würschum et al. 2012). Larger allelic classes and more balanced allele frequencies increase power to detect QTLs with small effects. NAM combines better

resolution by targeting historical recombination events in parents and also includes more causative events that are likely to segregate in biparental progenies (Tian et al. 2011). Thus, NAM is an effective tool for identifying major and minor QTLs associated with a particular trait (Korte & Ashley 2013).

In plant breeding, accumulation of minor QTLs is a major constraint given the dozens of loci and small population sizes. Hence, genomic selection (GS) is used to sum the effects of genome-wide markers to predict genomic estimated breeding values (GEBVs) (Meuwissen et al. 2001). In GS, a training population is genotyped and phenotyped for the traits of interest to estimate the genetic effect of each marker. The estimated genetic effect of each marker is then used to predict the GEBVs of a related population of breeding lines that is genotyped but has not been phenotyped. The selection of suitable genotypes among the test population is performed based on the predicted favorable allele GEBVs for each trait (Goddard et al. 2010). The success of GS depends upon the prediction accuracy of the GS model, which is measured as the correlation between estimated breeding values and the observed phenotypic values of the selected population (Crossa et al. 2014). GS can translate to higher genetic gains by reducing the number of progenies and cycles needed and also by improving selection intensity and reducing cycle time (Larkin et al. 2019; Robertsen et al. 2019).

Several studies have reported the application of GS for wheat yield and disease resistance (Sun et al. 2017; Sandhu et al. 2021a), but not thoroughly for quality traits, especially GPC and GPC stability. GS has been tested in soft winter wheat for end-use quality, where Heffner et al. (2011) found that end-use quality and processing traits are more predictive than grain yield. The prediction accuracy for flour yield and flour protein was 0.56 and 0.39, respectively, using a ridge regression model. GS accuracy for different agronomic traits and their stability was predicted in a winter wheat population of 273 lines and it was observed that GS accuracy varied from 0.33 to 0.67 for yield with yield stability having higher accuracy (Huang et al. 2016). Although several studies have focused on genomic selection for yield stability, none of them have investigated GPC stability. The main objectives of this study were to (1) detect marker-traits associations for GPC stability and GPC; and (2) identify ability of GS models for predicting GPC and GPC stability under cross and independent validations.

Materials and Methods

Plant material and trait measurement

Thirty-two spring wheat accessions from the USDA-ARS National Small Grains Collection were chosen as parental lines for the creation of the NAM population. These parents were crossed with the common cultivar 'Berkut' ('Irena'/'Babax'/'Pastor'; released in 2002) to create 32 half-sib families (Blake et al. 2019). Berkut was included in this study because it is a broadly adapted photoperiod insensitive and semi-dwarf cultivar developed by the International Maize and Wheat Improvement Center (CIMMYT), Mexico.

Twenty-six families with genotype data provided by Kansas State University (Jordan et al. 2018) were then selected, and 25 random RILs from each family (650 total RIL; named NAM₆₅₀) were planted between 2014 and 2016 at Spillman Agronomy Farm in Pullman, WA, under rainfed conditions (Sandhu et al. 2021b). A modified augmented design was used in each trial with three check cultivars (Berkut, 'Thatcher', and 'McNeal' (Lanning et al. 1994)). Planting was completed on 5th May 2014, 8th April 2015, and 10th April 2016 based on field conditions. The agronomic practices, including nitrogen fertilizer goals, were uniform in each environment. Percentage of

protein content in the grain was measured using a Perten DA 7000 NIR analyzer (Perkin Elmer, Sweden).

Statistical analysis

Best linear unbiased estimates (BLUEs) were calculated with the augmented complete block design (ACBD) in the R program for individual environments (R Development Core Team. 2013; Rodríguez et al. 2018) using the model

$$Y_{ij} = \mu + \text{Gen}_j + \text{Check}_j + \text{Block}_i + e_{ij}$$

where Y_{ij} is the grain protein content of an individual line; μ is the mean effect; Block_i is the fixed effect of the i^{th} block, Gen_j represents the fixed effect of unreplicated genotypes; Check_j is the effect of the replicated checks within each block; and e_{ij} is the standard normal error. All the effects were considered fixed in BLUE calculations.

The significance of differences in GPC were analyzed across years and between families using analysis of variance (ANOVA). Standard deviation and coefficient of variation were calculated within the different families to identify those with lower variation for GPC across years. The GPC variation within families in each environment and across years was used to select families that had less variation for GPC. Broad sense heritability for GPC was obtained as

$$H^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_e^2)$$

where σ_g^2 and σ_e^2 are the genotypic and error variance components, respectively.

Stability analysis

The stability of each RIL was assessed using a Finlay Wilkinson (FW) program implemented in the FW package in R (Lian et al. 2016). The FW package jointly estimates the parameters of the FWR equation

$$Y_{ij} = \mu + g_i + (1+b_i)t_j + e_{ij}$$

where Y_{ij} is the GPC of the i^{th} RIL from the j^{th} environment, μ is the mean effect, t_j is the main effect of the j^{th} environment, g_i is the main effect of the i^{th} RIL, $(1+b_i)$ is the change in expected performance of the i^{th} RIL per unit change in the environmental effect (Stability index), and e_{ij} is an error term assumed to be independent and identical distributed with mean zero and variance σ_e^2 . All parameters are treated as random effects with distributions: $\mathbf{g} \sim N(0, \mathbf{A}\sigma_g^2)$, $\mathbf{b} \sim N(0, \mathbf{A}\sigma_b^2)$, and $\mathbf{t} \sim N(0, \mathbf{T}\sigma_t^2)$, where $\mathbf{A}$ is an $N \times N$ kinship matrix for all the RILs calculated from the complete set of genetic markers, and $\mathbf{T}$ is an $M \times M$ Pearson variance-covariance matrix, describing the relationship of phenotypic values among M environments.

The stability of each RIL was calculated using GPC values for a selected 175 RILs (NAM₁₇₅) in each environment as dependent variables. The FW package returned the environmental stable genotypic effect (g_i), environmental effect (t_j), and the stability index of each RIL (b_i). The stability index for each RIL provided an idea about plasticity across the environments (Kusmec et al. 2017). A stability index of 1 and -1 means that genotype is highly plastic and responds according to environmental changes whereas a stability index of 0 suggests the genotype performs stable under different environments. Loci associated with GPC stability were identified by GWAS of the stability index absolute values. The parents for the selected 175 RILs were ‘Dharwar Dry’, ‘PI210945’, ‘Citr15144’, ‘PI92569’, ‘PI92569’, ‘Citr4174’, and ‘PI43355’ (Supplementary Table 1 & 2).

Genotyping

The NAM population genotyping, curation methods, and population maps were previously reported (Sandhu et al. 2021a, c). The initial genotypic data used in this study composed of 73,345 high-quality markers which were anchored to the Chinese Spring RefSeqv1 reference map (Marcussen et al. 2014) and NAM₁₇₅ selected for stable GPC were used. Individual RIL with missing GPC data were removed before filtering. Markers with more than 20% of data missing were removed for further analysis. Individual RIL missing more than 10% allele data were removed before culling based on markers with less than 10% minor allele frequency (MAF) to lower the probability of Type I error during GWAS. After filtering, 175 RILs from NAM₁₇₅ and 38,588 markers were used for GWAS analysis.

Population structure and genome-wide association studies

Kinship matrix and structure parameters were calculated by GAPIT (Lipka et al. 2012). The VanRaden algorithm was used to derive the kinship parameter from marker genotypes (VanRaden 2008). Population structure was analyzed using principal component analysis (PCA) implemented with the R function 'prcomp' in the software GAPIT (Price et al. 2006). Genetic relatedness among NAM₁₇₅ and NAM₆₅₀ was calculated using population differentiation coefficient (F_{st}) (SAS Institute Inc. 2011). The GWAS was conducted in the GAPIT R package using Bayesian information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) model (Huang et al. 2018). Association studies were performed on the NAM₁₇₅ using GPC stability index and environmentally-stable phenotypic GPC derived from the FW regression. Selection of PCA groups for GWAS was performed by evaluating Q-Q plot results for each model. PCA groups were included as covariates in the BLINK model to account for population structure. The Bonferroni correction with a stringent $\alpha=0.05$ was used to identify highly significant associations.

Genomic selection

Genome-wide marker effects for GPC and GPC stability were estimated using ridge regression best linear unbiased prediction (rrBLUP; Endelman 2011), according to the model:

$$y = \mu + Zu + e$$

where y is an $N \times 1$ vector of BLUEs for GPC or GPC stability for each RIL, μ is the overall mean, Z is an $N \times M$ matrix of markers, u is a vector of marker effects, and e is a vector of residuals. GS was performed with five-fold cross-validation by including 80% of the samples in the training population and predicting the GEBVs of the remaining 20% of the samples under each environmental condition. For accuracy assessment, two 50 replication sets were performed, where each replication consisted of five model iterations.

Genomic selection models were developed for the NAM population of 650 RILs (NAM₆₅₀) and separately for the 175 RILs (NAM₁₇₅) from seven families selected based on lower variation for GPC across environments. Furthermore, independent validations were performed using both sets of RILs for predicting GEBVs for GPC. During independent validations, GS models were trained on the previous year data set, and predictions were made for the upcoming year. Models trained on 2014 GPC data were used for predictions in 2015 and 2016. Similarly, the 2015 GPC training model was used for 2016 predictions (Sandhu et al. 2021c).

Results

Variation of grain protein content across environments

The GPC values of the NAM₆₅₀ population ranged from 11.2-18.0% in 2014, 8.7-16.8% in 2015 and 9.7-17.0% in 2016 (**Figure 1**), with 2014 having the highest mean GPC. Distributions of GPC values were normal in the three environments (**Supplementary Figure 1**) with significant positive correlations ($R^2=0.49$ for 2014 and 2015, $R^2=0.42$ for 2014 and 2016, and $R^2=0.57$ for 2015 and 2016). Broad-sense heritability for GPC was moderate to high, ranging between $H^2=0.62$ in 2014, $H^2=0.36$ in 2015 and $H^2=0.68$ in 2016. The average GPC for the 26 families evaluated in this study is presented in **Supplementary Table 1**. Using this data, seven families totaling 175 RILs (NAM₁₇₅) were selected for low variation in GPC (**Supplementary Table 2**).

Stability analysis

Stability index (b), environmental effect (t), and genotypic effect (g) values were obtained for the NAM₁₇₅ population. Absolute values of stability index for NAM₁₇₅ RILs ranged from 0.00 to 2.504 and were normally distributed. Thirty RILs were identified which had no significant difference of stability index from 0 using individual t test ($p < 0.05$). Environment 2014 was observed to be the most favorable for high GPC. NAM₁₇₅ population was divided into five categories based upon the GPC and stability index, the trends of which are depicted in **Figure 2**. Categories 2, 3, and 5 includes the stable GPC lines with GPC in the range of 15-17%, 14-15%, and 9-14%, respectively (**Figure 2**). Categories 1 and 4 represent the plastic lines with reversal of GPC when moving to other environments. Categories 1, 2, 3, 4, and 5 includes 81, 15, 9, 64, and 6 RILs, respectively.

Population structure analysis

Population structure analyzed with PCA separated the NAM₁₇₅ population into seven separate groups based on their different parents, in addition to the common parent 'Berkut' (**Figure 3**). The PC1 accounted for 7.0% of the variation, whereas the PC2 explained 5.0% of the genetic variation (**Supplementary Figure 2**). Kinship plots obtained from VanRaden algorithms in GAPIT also separated the population into seven different groups (data not shown). Inclusion of different number of PCs as covariate in the GWAS model demonstrated that the first three PCs best control the false positives and false negatives, as evident from the Q-Q plots (**Supplementary Figure 4**). F_{st} coefficient was 0.11 for the NAM₁₇₅ and 0.18 for that NAM₆₅₀.

Marker-trait associations for the stability of grain protein content

Seven QTLs controlling GPC stability were identified on six chromosomes using a stringent Bonferroni correction of $\alpha=0.05$ (**Table 1; Supplementary Figure 3**). The variation explained by each locus ranged from -4.19 to 3.68%. Two significant QTLs were on chromosome 3B and chromosomes 1A, 1B, 2A, 2B, and 4A each had one significant association. Cumulatively, these loci explained 18.40% of the phenotypic variation. The QTLs on 1A and 2A have a positive effect on increasing GPC stability. The QTL on 1A has the favorable allele from Berkut and Dharwar Dry, while the QTL on 2A has the favorable allele from the landraces, namely CIt15144, PI210945, PI92001, and Dharwar Dry (**Table 1**). Five other QTLs had a negative effect on the GPC stability, and the parents of origin for associated alleles are provided in **Table 1**. Removal of

those alleles from a breeding program by selecting for alternative alleles as these loci will favor the development of lines having increased GPC stability.

Four significant QTLs were identified using the GPC values obtained across the environments as the response trait (**Table 2; Supplementary Figure 5**). GPC was associated with four different loci located on chromosomes 1A, 3B, 4A, and 7A that individually explained -7.30 to 6.78% of the variation for this trait and collectively 16.74% of the total phenotypic variation. The QTL on 1A positively affects GPC, and the associated allele originated from Berkut and Dharwar Dry. The remaining three QTLs negatively affected GPC and removing those alleles will increase GPC. Parents of origin for the associated alleles are provided in **Table 2**. As there are multiple alleles with favorable effects identified from various different parental lines, pre-breeding efforts will be required to introgress all QTL identified. As some QTLs were identified from landraces, selection during this pre-breeding process may help reduce linkage drag.

Prediction for grain protein content and stability

The prediction accuracy for GPC in the NAM₆₅₀ population was $r=0.50$ in 2014, $r=0.55$ in 2015, and $r=0.53$ in 2016. Overall, GPC stability was less predictable with prediction accuracy values between $r=0.34$ and 0.44 and a mean of 0.40 . The prediction accuracy of GPC in the selected NAM₁₇₅ subset ranged between $r=0.56$ to 0.69 . The maximum prediction accuracy of $r=0.69$ was achieved again for the 2015 environment, while the lowest prediction accuracy of $r=0.56$ was obtained for the 2014 environment. Comparison of prediction accuracies for each environment using the two different sets of population is presented in **Figure 4**. Prediction accuracy is generally high when using the NAM₁₇₅, increasing by 15% as compared to the NAM₆₅₀.

The GS accuracy was lower for independent validation of each population set and under different environments compared to cross-validation GS accuracies (**Figure 5**). Independent prediction accuracies for the NAM₆₅₀ population ranged between $r=0.30$ to 0.39 and ranged between $r=0.35$ to 0.43 for the NAM₁₇₅ for GPC. The independent prediction accuracies were higher for the NAM₁₇₅, and the same results were observed during the cross-validation prediction scenario. The highest independent prediction accuracy was obtained by a training model on 2015 GPC to 2016 GPC (**Figure 5**).

Discussion:

Stability of genotypes across environments

A primary goal of plant breeding programs is to select germplasm with superior adaptation to the targeted environments. The performance of genotypes may range from those that are very well adapted to a more narrow set of environments, and perform below average in others, to genotypes that perform consistently relative to others across a wider range of environments and are considered to have greater stability. Herein, we used a static stability concept, targeting a predetermined value of GPC across all the environments. There are numerous statistical tools for analyzing static stability but here we applied FW regression analysis as it is capable of summarizing the interactions in comprehensible ways (Eberhart and Russell 1966).

Crossover and non-crossover interactions were observed in our FW regression analysis which included reversals in rank and scale effects for GPC due to environmental effects and G*E interactions. Identification of only thirty lines with stable GPC across the environments out of 650 lines highlights the challenge of maintaining adequate population size when selecting for a complex quantitative trait in genetically diverse germplasm. Evaluating this population in a greater number of environments would provide more insight into the genetic control of this trait. Additionally, the founder parents of NAM population are primarily landraces that have not undergone routine selection for GPC (Blake et al. 2019) as has been done in contemporary germplasm for at least the past 60 years. The stable lines varied in their average GPC, but lines having high and stable GPC (Category 2) demonstrates the ability to select genotypes for these traits. The lines identified in this study having high GPC and stability are particularly useful for introgression into modern breeding germplasm to expand available allelic variation for this important trait.

Genomic regions controlling stability of GPC

The QTLs associated with GPC stability and *per se* GPC performance were not detected in the same genomic regions, suggesting that GPC stability and GPC are under the control of different genes. Accounting for the different genetic architecture of these two traits could aid in indexed selection for the desired GPC along with stability. Different QTL controlling yield stability and *per se* yield in wheat has similarly been documented (Berke et al. 1992). In maize, yield and yield stability were observed to be independent, demonstrating the potential for simultaneous selection for both traits (Tollenaar and Lee 2002). Loci controlling yield stability were located in the same regions as QTLs for yield and yield-component in a barley mapping effort (Kraakman et al. 2004). A critical factor that is not captured or quantifiable in each of these yield-related studies are the precise abiotic or biotic factors limiting yield potential in different environments. Thus, stability is an important aspect to consider and our study suggests that breeding programs may select for both GPC and GPC stability by treating them as separate traits for developing varieties for unpredictable environmental variables.

Genomic regions controlling GPC stability have not been investigated in wheat based on available literature. In the present study, seven QTLs controlling GPC stability distributed over six chromosomes were identified that individually explained -4.19 to 3.68% of variation for GPC stability. The loci mapped on 1A and 2A were aiding in increasing the stability of GPC while other five regions were decreasing GPC stability. Similar results were obtained by Sehgal et al. (2017) while mapping genomic regions for yield stability, where 11 QTLs associated with yield stability were distributed on seven chromosomes. The amount of phenotypic variation explained by their yield stability QTLs varied from 3.2 to 8.1%. Thus, although stability is an important consideration when selecting for complex quantitative traits like yield and GPC, appropriate index weighting in genomic selection approaches is most likely the better alternative in selection schemes.

Marker trait associations have reported QTLs for GPC on all the 21 chromosomes of wheat (Blanco et al. 2006; Groos et al. 2003; Joppa et al. 1997; Perretant et al. 2000; Prasad et al. 2003). Loci controlling GPC were mapped to chromosomes 1A, 3B, 4A, and 7A in this study. The locus linked to *SpringWheatNAM_tag_190170* on 1A was previously identified in a DH population using composite interval mapping (Mahjourimajd et al. 2016) and Groos et al. (2003) using an F₇ RIL population grown in five environments. The favorable allele for this locus was identified from

Berkut and Dharwar Dry, both of which are present-day cultivars, suggesting that this favorable locus is fixed in the wheat germplasm and not in the landraces. The 7A GPC locus was discovered in the same region as another investigation (Mahjourimajd et al. 2016; Rapp et al. 2018). Those studies also reported that this locus had a negative effect on the total GPC. Loci on chromosomes 3B and 4A have not been previously reported for GPC in wheat (Blanco et al. 1996; Joppa et al. 1997; Heo and Sherman 2013), which is due to no prior utilization of these landraces in mapping studies (Blake et al. 2019). These two loci had a negative effect on GPC with associated alleles identified in the landraces. The absence of these loci in the cultivars Berkut and Dharwar Dry strengthens the fact that these negative loci have been removed from the germplasm with continuous selection by the breeders. The utilization of diverse landraces in this study provides information about different genomic regions which are not absent in the present-day cultivars.

Similar to GPC stability loci, QTLs controlling GPC *per se* explained a small portion of the variance (16.74%). Given the number of small-effect loci controlling each trait, the GS approach proposed in this study should improve the ability to select for these traits simultaneously during a breeding process. The utilization of indexed genomic selection would assist in selecting simultaneously for GPC stability and GPC (Schulthess et al. 2016). Rapp et al. (2018) demonstrated the use of phenotypic and genomic selection indices to select durum wheat lines for high GPC and grain yield. Similarly, lines having GPC and grain yield were selected using index selection in multi-variate GS model for wheat (Michel et al. 2019). These studies suggested that utilization of index selection in multi-variate GS models would potentially aid in selecting lines having stable GPC in addition to high GPC and grain yield.

Accuracy for predicting GPC and GPC stability

Prediction accuracy for GPC ranged between 0.50 and 0.69, which is moderately higher than prediction accuracies for grain yield (Battenfield et al. 2016). Heritability of GPC, the effect of the NAM population, and population sizes used for training the GS models would each affect prediction accuracy (Heffner et al. 2011; Poland et al. 2012). Our results are consistent with previous studies where high heritability has resulted in better prediction accuracies in cereals (Windhausen et al. 2012). The heritability of GPC is usually higher than yield, which ultimately resulted in better predictions for GPC (Battenfield et al. 2016). In a genomic prediction experiment for grain yield in oats, lower prediction accuracy was obtained because the experiment was planted under diverse environmental conditions resulting in reduced genetic variance as compared to G*E interactions, and thereby reduced heritability (Asoro et al. 2011). The NAM population in the current study was investigated in more homogenous target environments for wheat production in the PNW. This should result in a relatively lower G*E variance relative to genetic variance, leading to a higher heritability estimate and an increase in genomic selection accuracy.

Prior studies are not available to compare accuracy of GS for GPC stability with stability index values obtained from FW regression in wheat. GS accuracies for GPC stability were significantly ($p < 0.05$) lower than GPC, suggesting a more complex architecture of GPC stability. Huang et al. (2016) conducted GS for grain yield, test weight, and flour protein content stability using an additive main effect and multiplicative interaction (AMMI) model in wheat. They observed an accuracy from 0.14 to 0.31 for the stability of flour protein using four different GS models namely Bayesian ridge regression (0.14), elastic net (0.27), rrBLUP (0.31), and RKHS (0.31). Their study also demonstrated that the stability index for quality traits has less prediction

accuracy than the trait itself. Our results, coupled with those of Huang et al. (2016), suggest that GS can be used for predicting GPC stability. Furthermore, the prediction accuracy for GPC stability could be improved by obtaining a stability index from a larger number of field trials, as a large number of environments are useful for more reliable estimations of stability (Piepho 1998). The GS models can be retrained by incorporating data from additional environments in subsequent years. High prediction accuracies for stability could better predict the performance of genotypes at multiple locations during breeding cycles.

There was an improvement of prediction accuracy when model was trained on NAM₁₇₅ compared to the NAM₆₅₀ population. This is in contrast to other studies where it is observed that prediction accuracies improve when the number of individuals in the training population increases (Isidro et al. 2015; Lorenz et al. 2011; Lorenzana and Bernardo 2009). This argument was strengthened by the population differentiation coefficient (F_{st}), suggesting lines in the NAM₁₇₅ were more genetically related compared to the NAM₆₅₀. It has been observed that prediction accuracies increase when training and testing are more genetically related to each other as was the case in NAM₁₇₅ population set (Heffner et al. 2010; Lorenz & Smith 2015). These results suggest that selection for lines having GPC stability also aids in the improvement of GS accuracies for GPC. This study opens up the avenue for utilization of GS in spring wheat breeding program for selecting lines having GPC stability in addition to high GPC.

Conclusions

Selection for GPC is often secondary to grain yield in terms of breeding objectives for spring wheat, although it is a primary trait producers consider when selecting varieties. We report the first large-scale study of nested association mapping and evaluation of GS for GPC stability in wheat. This study identified wheat lines having less variation for GPC and mapped QTLs controlling GPC stability, an important and often overlooked trait. The identification of stable genotypes with high GPC could help in developing cultivars that can perform similarly in multiple environments. The QTLs identified in this study explained a small amount of phenotypic variation, demonstrating the complexity of the trait which suggests a GS approach could best address breeding for GPC and GPC stability. Prediction accuracy is sufficiently high for the implementation of GS for GPC and GPC stability in this study. With implementation of GS for these traits, predictions can be available while evaluating grain yield, enabling selection for GPC along with agronomic and yield traits.

Conflict of Interest

The authors declare no conflict of interest.

Authors contributions

KS: analyzed data, conceptualized the idea, and drafted the manuscript; PM: assisted in data analysis and edited the manuscript; ML: collected the phenotypic data; MP: conducted field trials, and obtained the funding for the project; AC: edited the manuscript, conducted field trials, supervised the study and obtained the funding for the project. All authors read and approved the final manuscript.

Funding

This project was supported by the Agriculture and Food Research Initiative Competitive Grant 2017-67007-25939 (WheatCAP) and 2016-68004-24770 from the USDA National Institute of Food and Agriculture and Hatch project 1014919.

References

- Asoro FG, Newell MA, Beavis WD, et al (2011) Accuracy and Training Population Design for Genomic Selection on Quantitative Traits in Elite North American Oats. *Plant Genome* 4:132–144. <https://doi.org/10.3835/plantgenome2011.02.0007>
- Battenfield SD, Guzmán C, Chris Gaynor R, et al (2016) Genomic selection for processing and end-use quality traits in the CIMMYT spring bread wheat breeding program. *Plant Genome* 9:. <https://doi.org/10.3835/plantgenome2016.01.0005>
- Berke TG, Baenziger PS, Morris R (1992) Chromosomal location of wheat quantitative trait loci affecting agronomic performance of seven traits, using reciprocal chromosome substitutions. *Crop Sci* 32:621–627. <https://doi.org/https://doi.org/10.2135/cropsci1992.0011183X003200030010x>
- Blake NK, Pumphrey M, Glover K, et al (2019) Registration of the triticeae-cap spring wheat nested association mapping population. *J Plant Regist* 13:294–297. <https://doi.org/10.3198/jpr2018.07.0052crmp>
- Blanc G, Charcosset A, Mangin B, et al (2006) Connected populations for detecting quantitative trait loci and testing for epistasis: An application in maize. *Theor Appl Genet* 113:206–224. <https://doi.org/10.1007/s00122-006-0287-1>
- Blanco A, De Giovanni C, Laddomada B, et al (1996) Quantitative trait loci influencing grain protein content in tetraploid wheats. *Plant Breed* 115:310–316. <https://doi.org/https://doi.org/10.1111/j.1439-0523.1996.tb00925.x>
- Blanco A, Simeone R, Gadaleta A (2006) Detection of QTLs for grain protein content in durum wheat. *Theor Appl Genet* 112:1195–1204. <https://doi.org/10.1007/s00122-006-0221-6>
- Crossa. Jose, Pérez P, Hickey JB, et al (2014) Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity (Edinb)* 112:48–60. <https://doi.org/10.1038/hdy.2013.16>
- De Kroon Hans, Heidrun Huber, Josef F. Stuefer and JMVG (2005) A modular concept of phenotypic plasticity in plants. *New Phytol* 166:73–82. <https://doi.org/DOI: 10.1111/j.1469-8137.2004.01310.x>
- DePauw RM, Knox RE, Clarke FR, et al (2007) Shifting undesirable correlations. In: *Euphytica*. Springer, pp 409–415
- Eberhart SA, Russell WA (1966) Stability Parameters for Comparing Varieties1. *Crop Sci* 6:36–40. <https://doi.org/10.2135/cropsci1966.0011183x000600010011x>
- Endelman JB (2011) Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *Plant Genome* 4:250–255. <https://doi.org/10.3835/plantgenome2011.08.0024>

- 467 Finlay KW, Wilkinson GN (1963) The analysis of adaptation in a plant-breeding programme. Aust J
468 Agric Res 14:742–754. <https://doi.org/10.1071/AR9630742>
- 469 Fragoso CA, Moreno M, Wang Z, et al (2017) Genetic architecture of a rice nested association mapping
470 population. G3 Genes, Genomes, Genet 7:1913–1926. <https://doi.org/10.1534/g3.117.041608>
- 471 Francisco Rodríguez; Gregorio Alvarado; Ángela Pacheco; Juan Burgueño (2018) ACBD-R. Augmented
472 Complete Block Design with R for Windows. Version 4.0
- 473 Gage JL, Monier B, Giri A, Buckler ES (2020) Ten years of the maize nested association mapping
474 population: Impact, limitations, and future directions. Plant Cell 32:2083–2093
- 475 Goddard ME, Hayes BJ, Meuwissen THE (2010) Genomic selection in livestock populations. Genet Res
476 (Camb) 92:413–421. <https://doi.org/10.1017/S0016672310000613>
- 477 Groos C, Robert N, Bervas E, Charmet G (2003) Genetic analysis of grain protein-content, grain yield
478 and thousand-kernel weight in bread wheat. Theor Appl Genet 106:1032–1040.
479 <https://doi.org/10.1007/s00122-002-1111-1>
- 480 Heffner EL, Jannink JL, Iwata H, et al (2011) Genomic selection accuracy for grain quality traits in
481 biparental wheat populations. Crop Sci 51:2597–2606. <https://doi.org/10.2135/cropsci2011.05.0253>
- 482 Heffner EL, Lorenz AJ, Jannink JL, Sorrells ME (2010) Plant breeding with Genomic selection: Gain per
483 unit time and cost. Crop Sci 50:1681–1690. <https://doi.org/10.2135/cropsci2009.11.0662>
- 484 Heo H, Sherman J (2013) Identification of QTL for Grain Protein Content and Grain Hardness from
485 Winter Wheat for Genetic Improvement of Spring Wheat. Plant Breed Biotechnol 1:347–353.
486 <https://doi.org/10.9787/pbb.2013.1.4.347>
- 487 Huang M, Cabrera A, Hoffstetter A, et al (2016) Genomic selection for wheat traits and trait stability.
488 Theor Appl Genet 129:1697–1710. <https://doi.org/10.1007/s00122-016-2733-z>
- 489 Huang M, Liu X, Zhou Y, et al (2018) BLINK: A package for the next level of genome-wide association
490 studies with both individuals and markers in the millions. Gigascience 8:1–12.
491 <https://doi.org/10.1093/gigascience/giy154>
- 492 Huang X, Wei X, Sang T, et al (2010) Genome-wide association studies of 14 agronomic traits in rice
493 landraces. Nat Genet 42:961–967. <https://doi.org/10.1038/ng.695>
- 494 Isidro J, Jannink JL, Akdemir D, et al (2015) Training set optimization under population structure in
495 genomic selection. Theor Appl Genet 128:145–158. <https://doi.org/10.1007/s00122-014-2418-4>
- 496 Joppa LR, Hareland GA, Du C, Hart GE (1997) Mapping gene(s) for grain protein in tetraploid wheat
497 (*Triticum turgidum* L.) using a population of recombinant inbred chromosome lines. Crop Sci
498 37:1586–1589. <https://doi.org/doi:10.2135/cropsci1997.0011183X003700050030x>
- 499 Jordan, Katherine W., Wang S, He F, Chao S, et al (2018) The genetic architecture of genome-wide
500 recombination rate variation in allopolyploid wheat revealed by nested association mapping. Plant J
501 95:1039–1054. <https://doi.org/10.1111/tpj.14009>
- 502 Kliebenstein DJ, Figuth A, Mitchell-olds T (2002) Genetic Architecture of Plastic Methyl Jasmonate
503 Responses in *Arabidopsis thaliana*. 1696:1685–1696
- 504 Korte A, Ashley F (2013) The advantages and limitations of trait analysis with GWAS : a review. Plant

505 Methods 9:29. <https://doi.org/https://doi.org/10.1186/1746-4811-9-29>

506 Kraakman ATW, Niks RE, Van Den Berg PMMM, et al (2004) Linkage disequilibrium mapping of yield
507 and yield stability in modern spring barley cultivars. *Genetics* 168:435–446.
508 <https://doi.org/10.1534/genetics.104.026831>

509 Kusmec A, Srinivasan S, Nettleton D, Schnable PS (2017) Distinct genetic architectures for phenotype
510 means and plasticities in *Zea mays*. *Nat Plants* 3:715–723. [https://doi.org/10.1038/s41477-017-](https://doi.org/10.1038/s41477-017-0007-7)
511 0007-7

512 Lanning SP, Talbert LE, McGuire CF, et al (1994) Registration of ‘McNeal’ Wheat. *Crop Sci* 34:1126.
513 <https://doi.org/10.2135/cropsci1994.0011183x003400040060x>

514 Larkin DL, Lozada DN, Mason RE (2019) Genomic Selection—Considerations for Successful
515 Implementation in Wheat Breeding Programs. *Agronomy* 9:479.
516 <https://doi.org/10.3390/agronomy9090479>

517 Le Gouis J, Bordes J, Ravel C, et al (2012) Genome-wide association analysis to identify chromosomal
518 regions determining components of earliness in wheat. *Theor Appl Genet* 124:597–611.
519 <https://doi.org/10.1007/s00122-011-1732-3>

520 Li H, Bradbury P, Ersoz E, et al (2011) Joint QTL linkage mapping for multiple-cross mating design
521 sharing one common parent. *PLoS One* 6:. <https://doi.org/10.1371/journal.pone.0017573>

522 Lian. Lian and G de los C (2016) GENOMIC SELECTION FW : An R Package for Finlay – Wilkinson
523 Regression that Incorporates Genomic / Pedigree Information and Covariance Structures Between
524 Environments. *G3 Genes, Genomes, Genet* 6:589–597. <https://doi.org/10.1534/g3.115.026328>

525 Lin Chuang-Sheng, Michael R. Binns and LPL (1986) Stability Analysis: Where Do We Stand?l. *Crop*
526 *Sci* 26:894–900. <https://doi.org/https://doi.org/10.2135/cropsci1986.0011183X002600050012x>

527 Lipka AE, Tian F, Wang Q, et al (2012) GAPIT: Genome association and prediction integrated tool.
528 *Bioinformatics* 28:2397–2399. <https://doi.org/10.1093/bioinformatics/bts444>

529 Löffler CM, Busch RH (1982) Selection for Grain Protein, Grain Yield, and Nitrogen Partitioning
530 Efficiency in Hard Red Spring Wheat1. *Crop Sci* 22:591–595.
531 <https://doi.org/10.2135/cropsci1982.0011183x002200030038x>

532 Lorenz AJ, Chao S, Asoro FG, et al (2011) *Genomic Selection in Plant Breeding : Knowledge and*
533 *Prospects*, 1st edn.

534 Lorenz AJ, Smith KP (2015) Adding genetically distant individuals to training populations reduces
535 genomic prediction accuracy in Barley. *Crop Sci* 55:2657–2667.
536 <https://doi.org/10.2135/cropsci2014.12.0827>

537 Lorenzana RE, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in
538 biparental plant populations. *Theor Appl Genet* 120:151–161. [https://doi.org/10.1007/s00122-009-](https://doi.org/10.1007/s00122-009-1166-3)
539 1166-3

540 Mahjourimajd S, Taylor J, Rengel Z, et al (2016) The genetic control of grain protein content under
541 variable nitrogen supply in an Australian wheat mapping population. *PLoS One* 11:1–13.
542 <https://doi.org/10.1371/journal.pone.0159371>

543 Marcussen T, Sandve SR, Heier L, et al (2014) A chromosome-based draft sequence of the hexaploid

544 bread wheat (*Triticum aestivum*) genome Ancient hybridizations among the ancestral genomes of
545 bread wheat Genome interplay in the grain transcriptome of hexaploid bread wheat Structural and
546 functional pa. *Science* (80-) 345:1251788. <https://doi.org/10.1126/science.1251788>

547 McMullen MD, Kresovich S, Villeda HS, et al (2009) Genetic Properties of the Maize Nested Association
548 Mapping Population. *Science* (80-) 325:737–740. <https://doi.org/10.1126/science.1174320>

549 McMullen MD, Kresovich S, Villeda HS, et al (2009) Supporting online material for: Genetic Properties
550 of the Maize Nested Association Mapping Population. *Science* (80-) 325:737–741.
551 <https://doi.org/10.1126/science.1174320>

552 Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide
553 Dense Marker Maps. *Genetics* 157:1819–29

554 Michel S, Löschenberger F, Ametz C, et al (2019) Simultaneous selection for grain yield and protein
555 content in genomics-assisted wheat breeding. *Theor Appl Genet* 132:1745–1760.
556 <https://doi.org/10.1007/s00122-019-03312-5>

557 Ordas B, Malvar RA, Hill WG (2008) Genetic variation and quantitative trait loci associated with
558 developmental stability and the environmental correlation between traits in maize. *Genet Res*
559 (Camb) 90:385–395. <https://doi.org/10.1017/S0016672308009762>

560 Perretant MR, Cadalen T, Charmet G, et al (2000) QTL analysis of bread-making quality in wheat using a
561 doubled haploid population. *Theor Appl Genet* 100:1167–1175.
562 <https://doi.org/https://doi.org/10.1007/s001220051420>

563 Piepho H-P (1998) Methods for Comparing the Yield Stability of Cropping Systems. *J Agron Crop Sci*
564 180:193–213. <https://doi.org/10.1111/j.1439-037X.1998.tb00526.x>

565 Poland J, Endelman J, Dawson J, et al (2012) Genomic Selection in Wheat Breeding using Genotyping-
566 by-Sequencing. *Plant Genome J* 5:103–113. <https://doi.org/10.3835/plantgenome2012.06.0006>

567 Prasad M, Kumar N, Kulwal PL, et al (2003) QTL analysis for grain protein content using SSR markers
568 and validation studies using NILs in bread wheat. *Theor Appl Genet* 106:659–667.
569 <https://doi.org/https://doi.org/10.1007/s00122-002-1114-y>

570 Prasad M, Varshney RK, Kumar A, et al (1999) A microsatellite marker associated with a QTL for grain
571 protein content on chromosome arm 2DL of bread wheat. *Theor Appl Genet* 99:341–345.
572 <https://doi.org/10.1007/s001220051242>

573 Price AL, Patterson NJ, Plenge RM, et al (2006) Principal components analysis corrects for stratification
574 in genome-wide association studies. *Nat Genet.* <https://doi.org/10.1038/ng1847>

575 R Development Core Team. (2013) R: A language and environment for statistical computing. R
576 Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900:201

577 Rapp M, Lein V, Lacoudre F, et al (2018) Simultaneous improvement of grain yield and protein content
578 in durum wheat by different phenotypic indices and genomic selection. *Theor Appl Genet*
579 131:1315–1329. <https://doi.org/10.1007/s00122-018-3080-z>

580 Robertsen C, Hjortshøj R, Janss L (2019) Genomic Selection in Cereal Breeding. *Agronomy* 9:95.
581 <https://doi.org/10.3390/agronomy9020095>

582 Saade S, Maurer A, Shahid M, et al (2016) Yield-related salinity tolerance traits identified in a nested

583 association mapping (NAM) population of wild barley. Sci Rep 6:32586.
584 <https://doi.org/10.1038/srep32586>

585 Sandhu KS, Lozada DN, Zhang Z, et al (2021a) Deep Learning for Predicting Complex Traits in Spring
586 Wheat Breeding Program. Front Plant Sci 11:613325. <https://doi.org/10.3389/fpls.2020.613325>

587 Sandhu KS, Mihalyov PD, Lewien MJ, et al (2021b) Combining Genomic and Phenomic Information for
588 Predicting Grain Protein Content and Grain Yield in Spring Wheat. Front Plant Sci 12:170.
589 <https://doi.org/10.3389/fpls.2021.613300>

590 Sandhu KS, Patil SS, Pumphrey MO, Carter AH (2021c) Multi-Trait Machine and Deep Learning Models
591 for Genomic Selection using Spectral 1 Information in a Wheat Breeding Program. bioRxiv
592 2021.04.12.439532. <https://doi.org/10.1101/2021.04.12.439532>

593 Schulthess AW, Wang Y, Miedaner T, et al (2016) Multiple-trait- and selection indices-genomic
594 predictions for grain yield and protein content in rye for feeding purposes. Theor Appl Genet
595 129:273–287. <https://doi.org/10.1007/s00122-015-2626-6>

596 Sehgal D, Autrique E, Singh R, et al (2017) Identification of genomic regions for grain yield and yield
597 stability and their epistatic interactions. Sci Rep 7:1–12. <https://doi.org/10.1038/srep41578>

598 Shewry PR (2007) Improving the protein content and composition of cereal grain. J Cereal Sci 46:239–
599 250. <https://doi.org/10.1016/j.jcs.2007.06.006>

600 Shewry PR, Halford NG (2002) Cereal seed storage proteins: Structures, properties and role in grain
601 utilization. J Exp Bot 53:947–958. <https://doi.org/10.1093/jexbot/53.370.947>

602 Sun J, Rutkoski JE, Poland JA, et al (2017) Multitrait , Random Regression , or Simple Repeatability
603 Model in High-Throughput Phenotyping Data Improve Genomic Prediction for Wheat Grain Yield.
604 <https://doi.org/10.3835/plantgenome2016.11.0111>

605 Tian F, Bradbury PJ, Brown PJ, et al (2011) Genome-wide association study of leaf architecture in the
606 maize nested association mapping population. Nat Genet 43:159–162.
607 <https://doi.org/10.1038/ng.746>

608 Tollenaar M, Lee EA (2002) Yield potential, yield stability and stress tolerance in maize. F Crop Res
609 75:161–169. [https://doi.org/10.1016/S0378-4290\(02\)00024-2](https://doi.org/10.1016/S0378-4290(02)00024-2)

610 Vanraden PM (2008) Efficient Methods to Compute Genomic Predictions. J Dairy Sci 91:4414–4423.
611 <https://doi.org/10.3168/jds.2007-0980>

612 Vogel KP, Johnson VA, Mattern PJ (1976) Protein and Lysine Content of Grain, Endosperm, and Bran
613 of Wheats from the USDA World Wheat Collection 1 . Crop Sci 16:655–660.
614 <https://doi.org/10.2135/cropsci1976.0011183x0016000500014x>

615 Windhausen VS, Atlin GN, Hickey JM, et al (2012) Effectiveness of Genomic Prediction of Maize
616 Hybrid Performance in Different Breeding Populations and Environments. G3 Genes, Genomes,
617 Genet 2:1427–1436. <https://doi.org/10.1534/g3.112.003699>

618 Würschum T, Liu W, Gowda M, et al (2012) Comparison of biometrical models for joint linkage
619 association mapping. Heredity (Edinb) 108:332–340. <https://doi.org/10.1038/hdy.2011.78>

620 Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. Curr Opin
621 Biotechnol 17:155–160. <https://doi.org/10.1016/j.copbio.2006.02.003>

622 Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic Design and Statistical Power of Nested
623 Association Mapping in Maize. *Genetics* 178:539–551. <https://doi.org/10.1534/genetics.107.074245>

624 Zhu C, Gore M, Buckler ES, Yu J (2008) Status and Prospects of Association Mapping in Plants. *Plant*
625 *Genome J* 1:5–20. <https://doi.org/10.3835/plantgenome2008.02.0089>

626 (2011) SAS Institute Inc. "SAS® 9.3 system options: Reference.

627

628

629

630

631

632

633

634

635

636

Table 1: Significant markers representing quantitative trait loci for grain protein content stability in a nested association mapping of hard spring wheat.

Marker description ^Δ						Allelic effect [‡]		Significance values
SNP name	Chromosome	Position on chromosome (cM)	Alleles [⊃]	Parental line with bolded allele	Minor allele frequency	Cumulative R ²	Effect	P-value (Bonf.)
SpringWheatNAM_tag_81337:59	1A	50917564	C/T	Berkut, Dharwar Dry	0.40	1.15	+	7.91E-08
SpringWheatNAM_tag_302718	1B	381359110	A/G	CItr15144	0.12	4.19	-	1.32E-07
SpringWheatNAM_tag_94853	2A	569963539	T/G	CItr15144, PI210945, PI92001, Dharwar Dry	0.36	3.68	+	1.06E-06
SpringWheatNAM_tag_269074	2B	155136626	A/C	PI210945, PI43355	0.11	3.62	-	5.67E-12
SpringWheatNAM_tag_84633	3B	23359725	A/T	PI92569	0.14	3.98	-	1.05E-08
SpringWheatNAM_tag_7037	3B	508522245	G/T	Berkut, Dharwar Dry, PI92569	0.42	1.27	-	2.84E-09
SpringWheatNAM_tag_281164	4A	309647389	T/C	Berkut, Dharwar Dry, CItr15144, PI210945	0.30	0.51	-	2.84E-09

^Δ Description of markers representing the tag of SNP, allele form of the tag SNP, minor allele frequency, chromosome, and position.

[‡] Phenotypic variation explained by the marker polymorphism: + increasing effect of minor allele; - decreasing effect of minor allele.

[⊃] Minor alleles are bolded

643 **Table 2:** Significant markers representing quantitative trait loci for grain protein content in a nested association mapping of hard spring wheat.

Marker description ^Δ						Allelic effect [‡]		Significance values
SNP name	Chromosome	Position on chromosome (cM)	Alleles [⊖]	Parental line with bolded allele	Minor allele frequency	Cumulative R ²	Effect	P-value (Bonf.)
SpringWheatNAM_tag_190170	1A	1.24E+08	T/C	Berkut, Dharwar Dry	0.49	6.78	+	1.60E-08
SpringWheatNAM_tag_32264	3B	125543693	A/T	Berkut, PI43355	0.34	7.30	-	1.26E-06
SpringWheatNAM_tag_82154	4A	6.4E+08	C/A	CItr15144, PI210945, CItr4175.	0.26	2.34	-	1.23E-07
SpringWheatNAM_tag_136322	7A	6.67E+08	T/G	Berkut, PI43355	0.41	0.32	-	1.55E-08

644 ^Δ Description of markers representing the tag of SNP, allele form of the tag SNP, minor allele frequency, chromosome, and position.

645 [‡] Phenotypic variation explained by the marker polymorphism: + increasing effect of minor allele; - decreasing effect of minor allele.

646 [⊖] Minor alleles are bolded

647

648

649

650

651

652

653

654

Figure 1: Variation of grain protein content for all the NAM₆₅₀ across the three environments. The X-axis depicts the three environments, namely 2014, 2015, and 2016; the Y-axis shows the GPC in percentage.

Figure 2: Trends for stability index and GPC observed in the NAM₁₇₅ population. The slopes represent the stability index, X-axis depicts the environmental effect, and Y-axis shows the grain protein content.

Figure 3: Population structure inferred from principal component analysis and illustrated with the first two principal component's help. The whole population is subgrouped into seven groups, each group having a single NAM family.

Figure 4: Comparison of GS prediction accuracies for the NAM₆₅₀ and NAM₁₇₅ populations for GPC. The X-axis represents the combination of population and environment, and the Y-axis represents the prediction accuracies.

Figure 5: Comparison of independent prediction accuracies for the NAM₆₅₀ and NAM₁₇₅ populations for GPC. The X-axis represents the combination of population and environment where first number represents the testing environment while second number represents the environment on which the GS model was trained, and the Y-axis represents the prediction accuracies.

674

675

676

677

678

679