Genomska selekcija za svojstva kakvoće pšeničnoga zrna

Plavšin, Ivana

Sveučilište u Zagrebu

AGRONOMSKI FAKULTET

Ivana Plavšin

GENOMSKA SELEKCIJA ZA SVOJSTVA KAKVOĆE PŠENIČNOGA ZRNA

DOKTORSKI RAD

Zagreb, 2022.

University of Zagreb

FACULTY OF AGRICULTURE

Ivana Plavšin

GENOMIC SELECTION FOR WHEAT GRAIN QUALITY TRAITS

DOCTORAL THESIS

Zagreb, 2022

Sveučilište u Zagrebu

AGRONOMSKI FAKULTET

IVANA PLAVŠIN

GENOMSKA SELEKCIJA ZA SVOJSTVA KAKVOĆE PŠENIČNOGA ZRNA

DOKTORSKI RAD

Mentori:

prof. dr. sc. Jerko Gunjača

dr. sc. Dario Novoselović,

znanstveni savjetnik u trajnom zvanju

Zagreb, 2022.

University of Zagreb

FACULTY OF AGRICULTURE

Ivana Plavšin

GENOMIC SELECTION FOR WHEAT GRAIN QUALITY TRAITS

DOCTORAL THESIS

Supervisors:

Professor Jerko Gunjača, PhD

Senior Scientific Adviser Dario Novoselović, PhD

Zagreb, 2022

Bibliografski podatci:
Znanstveno područje: Biotehničke znanosti
Znanstveno polje: Poljoprivreda (agronomija)
Znanstvena grana: Genetika i oplemenjivanje bilja, životinja i mikroorganizama
Institucija: Sveučilište u Zagrebu Agronomski fakultet
Voditelji doktorskog rada: prof. dr. sc. Jerko Gunjača, dr. sc. Dario Novoselović
Broj stranica: 157
Broj slika: 10
Broj tablica: 1
Broj priloga: 3
Broj literaturnih referenci: 198
Datum obrane doktorskog rada: 21.07.2022.
Sastav povjerenstva za obranu doktorskog rada:
1. Prof. dr. sc. Zlatko Šatović, Sveučilište u Zagrebu Agronomski fakultet
2. Prof. dr. sc. Hrvoje Šarčević, Sveučilište u Zagrebu Agronomski fakultet
3. Dr. sc. Krešimir Dvojković, Poljoprivredni institut Osijek

Rad je pohranjen u:
Nacionalnoj i sveučilišnoj knjižnici u Zagrebu, Ulica Hrvatske bratske zajednice 4 p.p. 550, 10 000 Zagreb,
Knjižnici Sveučilišta u Zagrebu Agronomskog Fakulteta, Svetošimunska cesta 25, 10 000 Zagreb.

Tema rada prihvaćena je na sjednici Fakultetskog vijeća Agronomskog fakulteta Sveučilišta u Zagrebu, održanoj dana 2. lipnja 2020. godine, te odobrena na sjednici Senata Sveučilišta u Zagrebu, održanoj dana 14. srpnja 2020. godine.

SVEUČILIŠTE U ZAGREBU

AGRONOMSKI FAKULTET

IZJAVA O IZVORNOSTI

Ja, **Ivana Plavšin**, izjavljujem da sam samostalno izradila doktorski rad pod naslovom:

GENOMSKA SELEKCIJA ZA SVOJSTVA KAKVOĆE PŠENIČNOGA ZRNA

Svojim potpisom jamčim:

- da sam jedina autorica ovog doktorskog rada;
- da je doktorski rad izvorni rezultat mojeg rada te da se u izradi istoga nisam koristila drugim izvorima osim onih koji su u njemu navedeni;
- da sam upoznata s odredbama Etičkog kodeksa Sveučilišta u Zagrebu (Čl. 19).

Zagreb, 21.07.2022. godine

_____

Potpis doktorandice

Ocjena doktorskog rada

Doktorski rad je obranjen na Agronomskom fakultetu Sveučilišta u Zagrebu 21.07.2022. godine pred povjerenstvom u sastavu:

1. Prof. dr. sc. Zlatko Šatović _____
   Sveučilište u Zagrebu Agronomski fakultet

2. Prof. dr. sc. Hrvoje Šarčević _____
   Sveučilište u Zagrebu Agronomski fakultet

3. Dr. sc. Krešimir Dvojković _____
   Poljoprivredni institut Osijek

Informacije o mentorima

Prof. dr. sc. Jerko Gunjača rođen je 14. ožujka 1966. u Splitu. Diplomirao je 1990. godine, a magistrirao 1997. godine na Agronomskom fakultetu u Zagrebu. Doktorsku disertaciju pod naslovom "Interakcija genotip × okolina u nebalansiranih serijama pokusa" obranio je 2001. godine, također na Sveučilištu u Zagrebu, Agronomskom fakultetu.

Tijekom poslijediplomskog studija boravio je na dodatnom usavršavanju na School of Biological Sciences, The University of Birmingham, UK. Dobio je tri poslijedoktorske stipendije: TEMPUS IMG za dvomjesečni boravak na Wageningen University and Research Centre, Nizozemska; C.I.H.E.A.M.-a za dvotjedni tečaj održan na I.A.M.Z.-u u Zaragozi, Španjolska; i programa Fulbright za petomjesečni studijski boravak na sveučilištu Cornell u Ithaci, SAD.

Na mjesto mlađeg asistenta na Zavodu za oplemenjivanje bilja, genetiku, biometriku i metodiku istraživanja Agronomskog fakulteta u Zagrebu primljen je 1994. godine. U istom je Zavodu ostao zaposlen sve do danas, te je izabran u znanstveno-nastavno zvanje docenta 2005., izvanrednog profesora 2008., redovitog profesora 2013., te u trajno zvanje redovitog profesora 2018. godine.

Surađivao je na više znanstvenih i stručnih projekata, te je bio voditelj projekta "Povećanje učinkovitosti istraživanja primjenom naprednih biometrijskih modela", financiranog od strane Ministarstva znanosti, obrazovanja i sporta Republike Hrvatske i VIP projekta „Revitalizacija lokalnih sorti heljde za uzgoj obiteljskim poljoprivrednim gospodarstvima sjeverozapadne Hrvatske". U ovom trenutku je suradnik na projektu KK.01.1.1.01.0005 „Bioraznolikost i molekularno oplemenjivanje bilja" Znanstvenog centra izvrsnosti za bioraznolikost i molekularno oplemenjivanje bilja (CroP-BioDiv), financiranog iz Europskog fonda za regionalni razvoj. Rezultate sveukupne znanstvene djelatnosti objavio je u 35 znanstvenih radova iz skupine a1, 16 znanstvenih radova iz skupine a2 i 20 znanstvenih radova iz skupine a3, te prezentirao na 17 međunarodnih i 13 domaćih znanstvenih skupova.

Dr. sc. Dario Novoselović rođen je 29. travnja 1967. godine u Osijeku. Diplomirao je 1992. godine, a magistrirao 1997. godine na Agronomskom fakultetu u Zagrebu. Akademski stupanj doktora biotehničkih znanosti stekao je 2002. godine, također na Agronomskom fakultetu Sveučilišta u Zagrebu, izradom i obranom disertacije pod naslovom "Selekcija i nasljeđivanje kvantitativnih svojstava u ranim generacijama ozime pšenice (*Triticum aestivum* L.)".

Od 1993. godine zaposlen je na Poljoprivrednom institutu Osijek, Odjelu za oplemenjivanje i genetiku strnih žitarica, najprije na radnom mjestu mlađeg asistenta, a od 2009. godine do danas zaposlen je na radnom mjestu znanstvenog savjetnika. Od zaposlenja na Poljoprivrednom institutu Osijek sudjelovao je u izvođenju sljedećih znanstvenih projekata odobrenih i financiranih od Ministarstva znanosti, obrazovanja i sporta Republike Hrvatske: "Genetičko poboljšanje uroda i kakvoće pšenice", "Oplemenjivanje poljoprivrednih kultura u Slavoniji", „Genetika i oplemenjivanje kvantitativnih svojstava pšenice", te u razdoblju od 2007. godine do danas projekta „Genotip i stresni učinci u proizvodnji i kvaliteti sjemena pšenice i ječma". Od 2007. godine do danas nositelj je znanstvenog projekta „Razvoj QTL-a pomoću molekularnih markera za svojstva kvalitete pšenice". Sudjelovao je i u izvedbi dva stručna projekta odobrena i financirana od strane Ministarstva poljoprivrede Republike Hrvatske: „Durum pšenica na poljoprivrednom obiteljskom gospodarstvu" i „Sinergija dušika i suše u okolišu u proizvodnji pšenice i ječma". Od 2016. godine voditelj je stručnog projekta „Razvoj germplazme krušne i durum pšenice za obiteljska poljoprivredna gospodarstva u Republici Hrvatskoj" odobrenog i sufinanciranog od strane Ministarstva poljoprivrede Republike Hrvatske. Bio je voditelj istraživačkog projekta Hrvatske zaklade za znanost „Genetsko poboljšanje i optimizacija potencijala rodnosti pšenice" u razdoblju od 2017. do 2021. godine. Od 2016. godine predsjednik je Znanstvenog odbora Znanstvenog centra izvrsnosti za bioraznolikost i molekularno oplemenjivanje bilja (CroP-BioDiv).

Do danas je kao autor ili koautor objavio više od 100 znanstvenih i stručnih radova, od kojih je 43 objavljeno u časopisima indeksiranim u bazi WOS. Sudjelovao je na više međunarodnih i nacionalnih znanstvenih i stručnih skupova. Od 2009. do 2016. godine bio je član Uređivačkog odbora časopisa „Poljoprivreda" (Osijek, Republika Hrvatska), a od 2011. godine do danas i Izdavačkog saveta časopisa „Ratarstvo i povrtarstvo" (Novi Sad, Republika Srbija). Autor i koautor je 86 sorti ozime pšenice priznatih u Republici Hrvatskoj i 24 sorte priznate u inozemstvu.

U više navrata boravio je u inozemstvu na znanstvenim usavršavanjima, od kojih treba izdvojiti dobivanje Fulbright stipendije i boravak na Sveučilištu Cornell (SAD, 2007./2008. godine) te boravak u Međunarodnom centru za oplemenjivanje kukuruza i pšenice CIMMYT (Meksiko, 2000. godine). Član je Europskog udruženja za istraživanja u oplemenjivanju bilja (EUCARPIA), Hrvatskog genetičkog društva (HGD) i Hrvatskog društva agronoma (HDA).

Doktorski rad je izrađen u sklopu projekta KK.01.1.1.01.0005 „Bioraznolikost i molekularno oplemenjivanje bilja" Znanstvenog centra izvrsnosti za bioraznolikost i molekularno oplemenjivanje bilja voditelja prof. dr. sc. Zlatka Šatovića.

CroP-BioDiv

ZNANSTVENI CENTAR IZVRSNOSTI ZA BIORAZNOLIKOST I MOLEKULARNO OPLEMENJIVANJE BILJA

CENTRE OF EXCELLENCE FOR BIODIVERSITY AND MOLECULAR PLANT BREEDING

EUROPSKI STRUKTURNI I INVESTICIJSKI FONDOVI

Operativni program KONKURENTNOST I KOHEZIJA

Zahvala

Zahvalnost za izradu ovog doktorskog rada dugujem ponajprije svojim mentorima prof. dr. sc. Jerku Gunjači i dr. sc. Dariju Novoseloviću. Hvala na prenesenom iskustvu i znanju, korisnim savjetima, strpljenju i nesebičnoj pomoći prilikom pisanja ovog rada.

Zahvaljujem članovima Povjerenstva za ocjenu i obranu doktorskog rada prof. dr. sc. Zlatku Šatoviću, prof. dr. sc. Hrvoju Šarčeviću i dr. sc. Krešimiru Dvojkoviću na vremenu i trudu uloženom u čitanje i ocjenu ovog rada.

Hvala svim zaposlenicima Poljoprivrednog instituta Osijek koji su, posredno ili neposredno, doprinijeli izradi ovog doktorskog rada. Posebno hvala dragom kolegi Krešimiru na svim lijepim riječima i ohrabrenju tijekom pisanja ovog rada.

Naposlijetku, hvala mojoj obitelji i prijateljima na podršci i razumijevanju koje su mi pružali sve ove godine te neizmjernoj vjeri u moj uspjeh. Veliku zahvalnost dugujem mom prijatelju Vladimiru bez čije bi moralne i tehničke podrške ovaj put bio puno duži i trnovitiji. Vladimire, hvala ti na više od 15 godina bezuvjetnog prijateljstva, na vjeri u mene i onda kada ja to nemam. Lidija, hvala na pregledu ovog rada i hvala što si uvijek spremna slušati. Najveće hvala mojoj sestri Orneli bez koje danas ne bih bila tu gdje jesam.

Sažetak

Pšenica (*Triticum aestivum* L.) je jedna od najvažnijih biljnih vrsta za proizvodnju hrane u svijetu te najvažniji izvor proteina i energije u ljudskoj prehrani. Kakvoća pšenice određuje se na osnovi mnogobrojnih svojstava, a jedan od najvažnijih čimbenika koji utječe na reologiju tijesta je sadržaj glutena i njegova kakvoća. Reološki profil tijesta može se odrediti miksografom koji se zbog male količine brašna potrebne za analizu pokazao prikladnim za korištenje u oplemenjivanju pšenice, osobito u ranim generacijama kada velike količine zrna nisu dostupne. Budući da većina svojstava kakvoće pšenice pokazuje složene obrasce nasljeđivanja, oplemenjivanje na kakvoću, a posebno na pekarsku kakvoću, jedan je od najzahtjevnijih izazova u oplemenjivanju pšenice. Ciljevi ovog istraživanja bili su procijeniti utjecaj interakcije genotip × okoliš i mogućnost korištenja genomske selekcije za svojstva kakvoće zrna pšenice kako bi se postigla učinkovitija selekcija za navedena svojstva te smanjili potencijalni troškovi genotipizacije i fenotipizacije u oplemenjivačkom procesu.

U istraživanju su korištene dvije biparentalne (RIL) populacije pšenice dobivene križanjem roditeljskih sorti Bezostaya-1 × Klara (BK) i Monika × Golubica (MG). Poljski pokusi su provedeni na dvije lokacije u Hrvatskoj tijekom tri godine. U svakom od okoliša određene su vrijednosti sadržaja proteina (GPC), sadržaja vlažnog glutena (WGC) i hektolitarske mase (TW). Reologija tijesta analizirana je pomoću miksografa, a četiri varijable su odabrane za daljnju statističku analizu (MPT, MTW, MTI, MPH). Interakcija genotip × okoliš analizirana je pomoću AMMI modela. RR-BLUP model korišten je kako bi se utvrdila potreba za optimizacijom trenažne populacije na osnovi fenotipske varijance, te ispitao utjecaj veličine trenažne populacije i gustoće biljega na točnost predviđanja genomske selekcije, a analize su provedene za svih sedam svojstava u obje populacije. Za utvrđivanje utjecaja veličine trenažne populacije na točnost predviđanja korištene su tri različite veličine trenažne populacije. Kako bi se utvrdio utjecaj gustoće biljega, genomska selekcija za sva svojstva provedena je korištenjem cijelog skupa biljega te polovice skupa biljega. Učinkovitost RR-BLUP modela uspoređena je s učinkovitošću sedam drugih modela za predviđanje svojstava kakvoće.

Analiza interakcije genotip × okoliš pokazala je određene zajedničke obrasce za dvije promatrane populacije. Za GPC, WGC i TW dominantan izvor fenotipske varijacije bio je okoliš. Na MPT i MTW dominantan utjecaj imala je interakcija genotip × okoliš u BK i genotip u MG populaciji, dok je na MTI i MPH dominantan učinak imao okoliš u BK i interakcija genotip × okoliš u MG populaciji. Općenito, utjecaj interakcije genotip × okoliš imao je važniju ulogu za svojstva miksografa u odnosu na ostala promatrana svojstva. Analizom AMMI2 biplota utvrđeni su neki široko prilagođeni RIL-ovi. Za sva svojstva utvrđene su uglavnom visoke vrijednosti heritabilnosti. Smanjenje veličine trenažne populacije imalo je negativan učinak na dobivenu točnost predviđanja genomske selekcije za sva promatrana svojstva u obje populacije. Dobiveni rezultati nisu podržali optimizaciju trenažne populacije na temelju fenotipske varijance. Također je primijećeno da točnost predviđanja može značajno varirati između okoliša. Kada se usporedi utjecaj različitih gustoća biljega na sposobnost predviđanja svojstava kakvoće unutar MG populacije, vrijednosti točnosti predviđanja dobivene korištenjem veće gustoće biljega bile su više u svim slučajevima. Za većinu kombinacija svojstvo-okoliš model elastične mreže je rezultirao najnižim vrijednostima točnosti predviđanja. Iako se RR-BLUP nije pokazao najuspješnijim modelom u svim slučajevima, nije uočena značajna prednost korištenja bilo kojeg drugog modela. Točnosti predviđanja dobivene u sklopu ovog istraživanja podržavaju primjenu genomske selekcije za oplemenjivanje pšenice na kakvoću, uključujući i oplemenjivanje na neka svojstva dobivena miksograf uređajem.

Ključne riječi: pšenica, svojstva kakvoće, biparentalna populacija, interakcija genotip × okoliš, AMMI model, genomska selekcija, trenažna populacija, heritabilnost, modeli predviđanja

Extended summary

Title of the doctoral thesis in English: Genomic selection for wheat grain quality traits

Wheat (*Triticum aestivum* L.) is one of the most important crops for food production in the world. The importance of wheat emphasizes the fact that wheat products are the most important source of dietary proteins and energy supply for humans. Therefore, achieving suitable wheat quality is of great importance. Wheat quality is determined by a large number of traits and under the strong environmental influence. One of the most important factors affecting dough rheology is gluten content and its strength. Gluten is the most abundant wheat protein, and by its structure, gluten is a complex network of monomeric gliadins and polymeric subunits of glutenin. Among gluten components, high molecular weight glutenin subunits have the greatest impact on dough quality. Different instruments can be used to perform rheological tests, which are necessary to assess wheat baking quality more accurately. Mixograph is a dough mixer that creates a dough rheological profile, providing general information about dough mixing, its behaviour during development, and the strength of the dough. Due to the small amount of flour required, mixograph has shown to be highly suitable for use in wheat breeding, particularly in early generations when availability of grain and flour is still limited.

Since the majority of quality traits have complex inheritance patterns, breeding for improved baking quality is one of the most demanding objectives in wheat breeding. Taking that into account together with often costly and time-consuming phenotyping, predictability of wheat baking quality may be very challenging. The ability to develop a genotype that exhibits both improved performance and high stability of the quality traits is critical to the success of wheat quality improvement. One of the major challenges in plant breeding, in this context, is the occurrence of genotype-by-environment interaction, since its presence makes selecting widely adapted genotypes difficult. The AMMI model is one of the most commonly used methods for the analysis of genotype-by-environment interaction. However, the extensive development of high-throughput genotyping in the last decade has enabled reliable and rapid predictions of breeding values based only on marker information. Genomic selection is one of the recently developed methods that enables the prediction of breeding values of individuals by simultaneously incorporating all available marker information into a model. Genomic selection aims to capture total additive genetic variance based on the sum of the effects of a large number of genetic markers, encompassing all QTLs that contribute to trait variability. In genomic selection, genotypic and phenotypic information of the training population is used to train a model and estimate the marker effects. Obtained data is then applied to the breeding (validation) population of non-phenotyped candidates to estimate their genomic-estimated breeding values (GEBV). The effectiveness of genomic selection is determined by the obtained prediction accuracy, which is affected by a variety of molecular, genetic, and phenotypic factors, as well as the parameters of the selected statistical model. The correct adjustment of factors that can affect prediction accuracy, such as population structure, size of training population, the relatedness of training and validation population, marker density, etc., is the first step toward successful implementation of genomic selection in practical breeding programs. Different prediction models have been developed to solve the problem of high-dimensional datasets occurring in genomic selection. These models differ primarily in their assumptions about the distribution and variance of marker effects, i.e., how marker effects contribute to the trait.

Given the often challenging phenotyping for wheat quality traits, and especially for baking quality traits, the use of classical breeding methods can be costly and time-consuming. Determining the optimal model and parameters of genomic selection would enable the use of molecular markers in the pre-selection process for grain quality traits and the optimization of classical wheat breeding methods. This research aimed to assess the impact of genotype-by-environment interaction and optimize genomic selection for grain quality traits using biparental wheat populations, in order to reduce the potential costs of

genotyping and phenotyping in the breeding process and suggest optimal strategies based on genomic selection for more efficient development of new lines.

Two biparental populations of winter wheat were used in this study. The BK population was derived from the Bezostaya-1 × Klara cross and the MG population from the Monika × Golubica cross. In the BK combination the parental genotypes differed in all high molecular weight glutenin subunits, while the parental genotypes used in the MG combination did not differ in any of the high molecular weight glutenin subunits. The BK and MG populations consisted of 145 and 175 genotypes, respectively, including parental genotypes. Field trials were conducted for three consecutive years (2009 – 2011) at two locations in Croatia (Osijek and Slavonski Brod), i.e., in six different environments. In each environment the field trial was set up according to a row-column design. Genotyping of both populations was done using Diversity Arrays Technology. After marker filtering the final dataset used for genomic selection contained 1087 and 2231 SNP markers for BK and MG population, respectively. Grain protein content, wet gluten content, test weight, and dough rheology were assessed in each environment. Dough rheology was investigated using a mixograph and four variables were selected for further statistical analysis (MPT, MTW, MTI, and MPH).

Genotype-by-environment patterns for the quality and mixograph traits were studied using the AMMI model. The dissection of genotype-by-environment patterns was visualized by a modified version of the AMMI2 biplot, which adds the main effects to the standard AMMI2 biplot using a colour scale. In the first phase of the genomic selection analysis the need for optimization of the training population based on phenotypic variance was assessed using both biparental populations. Additionally, the influence of the training population size and marker density on the prediction accuracy was investigated. For that purpose, three different sizes of training population were used for both BK and MG populations, and two different marker densities for the MG population. The first phase was conducted using only the RR-BLUP model. In the second phase of the genomic selection, the performance of seven different genomic selection models was compared with the performance of the RR-BLUP model. Models included were elastic net, four Bayesian models (BayesA, BayesB, BayesC, and BayesLASSO), random forest, and reproducing kernel Hilbert spaces. This part of the analysis was performed only within the MG population.

Results revealed some positive as well as negative transgressive segregants in both populations for all quality traits although being generally more prevalent in the BK population. This may suggest the dispersion of the alleles with positive (increasor) and negative (decreasor) effects between parental genotypes in both crosses. The environment was the dominant source of variation for grain protein content, wet gluten content, and test weight, accounting for approximately 40% to 85% of the total variation. The pattern was less consistent for mixograph traits for which the dominant source of variation was trait- and population-dependent. Overall, genotype-by-environment interaction was shown to play a more important role for mixograph traits compared to other quality traits. Inspection of the AMMI2 biplot revealed some broadly adapted RILs, among which MG124 is the most interesting, being the prevalent “winner” for grain protein content and wet gluten content, but also the “winner” for non-correlated trait test weight in environment SB10. The broad-sense heritability across environments was high for all traits, except for MPT in the BK population, the heritability of which was 0.45. Although repeatability varied considerably among environments, it was high for most of the trait-environment combinations, with a value above 0.7. These results suggest that heritability itself should not represent a limitation in achieving good prediction accuracy. The results of genomic selection analysis showed that the size of the training population plays an important role in achieving higher prediction accuracies, while marker density does not represent a major limitation. Additionally, the results of the present study did not support the optimization of the training population based on phenotypic variance as a tool to increase prediction accuracy. The performance of eight prediction models was compared and among them, elastic net showed

the lowest prediction accuracy for all traits. Bayesian models provided slightly higher prediction accuracy than the RR-BLUP model. However, this may be considered negligible considering the time required to perform an analysis. Although RR-BLUP was not the best performing model in all cases, no advantage of using any other model studied in this research was observed. Furthermore, strong differences among environments in terms of the prediction accuracy were observed. For example, the prediction accuracy for TW within the MG population was moderate in one environment, while being low in all other environments. Comparing these results to the results of a genotype-by-environment analysis it is noticeable that environments that are characterized by unusually high or low values for prediction accuracy compared to the rest of the environments tend to be those that produce the greatest genotype-by-environment interaction. This suggests that less predictive environments should be excluded from the dataset used to train the prediction model in order to achieve higher prediction accuracies. The prediction accuracies obtained in this study support implementation of genomic selection in wheat breeding for end-use quality, including some mixograph traits.

Key words: wheat, quality traits, biparental population, genotype × environment interaction, AMMI model, genomic selection, training population, heritability, prediction models, GEBV

SADRŽAJ

Popis kratica

AFLP (*amplified fragment length polymorphism)* – polimorfizam dužine amplificiranih ulomaka

AMMI (*additive main effects and multiplicative interaction)* – model aditivnih glavnih učinaka i multiplikativne interakcije

ANOVA (*analysis of variance)* – analiza varijance

BLUP (*best linear unbiased prediction)* – najbolje linearno nepristrano predviđanje

CD (*coefficient of determination*) – koeficijent determinacije

CIMMYT (*International Maize and Wheat Improvement Center)* – Međunarodni centar za oplemenjivanje kukuruza i pšenice

cM – centimorgan

DArT (*Diversity Arrays Technology)*

DH (*doubled haploid*) – udvostručeni haploid

EBV (*estimated breeding value)* – procijenjena oplemenjivačka vrijednost

EM-AMMI (*expectation maximization AMMI*) – AMMI model zasnovan na algoritmu maksimizacije očekivanja

EN (*elastic net*) – model elastične mreže

FAO (*Food and Agriculture Organization)* – Međunarodna organizacija za hranu i poljoprivredu

G-BLUP (*genomic best linear unbiased prediction)* – model genomskog najboljeg linearnog nepristranog predviđanja

GEBV (*genomic estimated breeding value)* – genomska procjena oplemenjivačke vrijednosti

GEI (*genotype-by-environment interaction*) – interakcija genotip × okoliš

GGE (*genotype main effects and genotype × environment interaction)* – glavni učinak genotipa i učinak genotip × okoliš interakcije

GPC (*grain protein content)* – sadržaj proteina u zrnu

HMW (*high molecular weight subunits)* – podjedinice visoke molekularne mase

IPCA (*interaction PCA axis)* – interakcijska PCA os

LASSO (*least absolute shrinkage and selection operator*) – operator najmanjeg apsolutnog skupljanja i odabira

LD (*linkage disequilibrium)* – neravnoteža vezanosti gena

LMW (*low molecular weight subunits)* – podjedinice niske molekularne mase

MAS (*marker assisted selection)* – selekcija potpomognuta biljezima

MPH (*midline peak height*) – visina vrha središnje krivulje miksograma

MPT (*midline peak time)* – vrijeme potrebno da središnja krivulja miksograma dosegne maksimalnu visinu

MSEP (*mean-squared error of prediction*) – prosječna kvadratna pogreška predviđanja

MTI (*midline curve tail)* – površina ispod središnje krivulje miksograma od početne točke zamjesa do kraja procesa miješanja

MTW (*midline curve tail width)* – širina vrha središnje krivulje miksograma na kraju perioda miješanja

NET – nedušične ekstraktivne tvari

NGS (*next-generation sequencing)* – sekvenciranje nove generacije

NIR (*near infrared*)

LOO-CV (*leave-one-out cross-validation*) – pojedinačna unakrsna validacija

PCA (*principal component anaysis)* – analiza glavnih komponenti

PEV (*predictor error variance*) – varijanca pogreške predviđanja

QTL (*quantitative trait loci)* – lokusi kvantitativnih svojstava

RAPD (*random amplified polymorphic DNA)* – sustav nasumično amplificirane polimorfne DNA

RF (*random forest*) – metoda slučajne šume

RFLP (*restriction fragment length polymorphism)* – polimorfizam dužine restrikcijskih ulomaka

RIL (*recombinant inbred line)* – rekombinantna inbred-linija

RKHS (*reproducing kernel Hilbert spaces*) – model Hilbertovih prostora reproducirajućih jezgri

RR (*ridge regression*) – hrbatna regresija

RR-BLUP (*ridge regression best linear unbiased prediction)* – model najboljeg linearnog nepristranog predviđanja po principu hrbatne regresije

SDS (*sodium dodecyl sulfate*) – natrijev dodecil-sulfat

SNP (*single nucleotide polymorphism)* – polimorfizam jednog nukleotida

SPB (*simple parametric bootstrap*) – jednostavna parametrijska *bootstrap* metoda

SSR (*simple sequence repeats)* – jednostavne ponavljajuće sekvence, mikrosateliti

SVD (*singular value decomposition)* – dekompozicija matrice na singularne vrijednosti

SVM (*support vector machine)* – metoda potpornih vektora

TBV (*true breeding value)* – prava oplemenjivačka vrijednost

TKW (*thousand kernel weight)* – masa tisuću zrna

TW (*test weight)* – hektolitarska masa

WGC (*wet gluten content)* – sadržaj vlažnog glutena

Popis tablica

Popis slika

Popis priloga

Popis znanstvenih radova

Objavljeni znanstveni radovi				
Znanstveni rad	Baza	Kategorija	Kvartil	Faktor odjeka (IF)
1. Plavšin, I., Gunjača, J., Šimek, R., Novoselović, D. (2021). Capturing GEI Patterns for Quality Traits in Biparental Wheat Populations. Agronomy 11 (6): 1022.	WoS	A_1	Q_1	3.949
2. Plavšin, I., Gunjača, J., Šatović, Z., Šarčević, H., Ivić, M., Dvojković, K., Novoselović, D. (2021). An Overview of Key Factors Affecting Genomic Selection for Wheat Quality Traits. Plants 10 (4): 745.	WoS	A_1	Q_1	4.658
3. Plavšin, I., Gunjača, J., Galić, V., Novoselović, D. (2022). Evaluation of Genomic Selection Methods for Wheat Quality Traits in Biparental Populations Indicates Inclination towards Parsimonious Solutions. Agronomy 12 (5): 1126.	WoS	A_1	Q_1	3.949

Obrazloženje povezanosti istraživačkih hipoteza i objavljenih znanstvenih radova	
Istraživačka hipoteza	Obrazloženje povezanosti hipoteze sa znanstvenim radom
H1. Korištenjem genomske selekcije moguće je s visokom točnošću predvidjeti svojstva kakvoće zrna pšenice i odabrati RIL-ove s optimalnom GEBV vrijednosti.	Rezultati koji proizlaze iz znanstvenog rada pod rednim brojem 3, u kojem je istražen potencijal genomske selekcije za predviđanje sedam svojstava kakvoće zrna u dvije biparentalne populacije pšenice, djelomično potvrđuju H1. Točnost predviđanja dobivena korištenjem različitih modela varirala je u ovisnosti o okolišu, ali je općenito bila srednja do srednje-visoka za većinu istraživanih svojstava, dok je za jedno od reoloških svojstava pokazala izrazito niske, te u nekim okolišima i negativne vrijednosti.
H2. Na točnost predviđanja modela značajno utječe heritabilnost (nasljednost) svojstva, veličina trenažne populacije i veličina skupa korištenih biljega.	Znanstveni rad pod rednim brojem 2 daje pregled dosadašnjih istraživanja genomske selekcije za svojstva kakvoće pšenice, ističe ključne čimbenike koji utječu na točnost predviđanja te pomaže smjestiti rezultate znanstvenog rada pod rednim brojem 3 u kontekst postojećih istraživanja. U okviru rezultata prikazanih u znanstvenom radu pod rednim brojem 3 djelomično je potvrđena H2. Utvrđeno je da visoka heritabilnost omogućuje predviđanje svojstava kakvoće zrna pšenice sa srednjom do srednje-visokom točnošću, te da veličina trenažne populacije igra važnu ulogu u postizanju viših vrijednosti točnosti predviđanja, dok veličina skupa korištenih biljega ne predstavlja značajno ograničenje.
H3. Na procjenu široko prilagođenih RIL-ova značajno utječe interakcija genotip × okoliš.	Rezultati koji proizlaze iz znanstvenog rada pod rednim brojem 1, u sklopu kojeg je pomoću AMMI modela analizirano postojanje interakcije genotip × okoliš za sedam svojstava kakvoće zrna u dvije biparentalne populacije pšenice, potvrđuju H3. Interakcija genotip × okoliš imala je značajniji utjecaj na reološka svojstva dobivena pomoću miksograf uređaja u odnosu na druga promatrana svojstva kakvoće za koje je okoliš bio dominantan izvor varijacija. Uz pomoć AMMI biplota utvrđeno je postojanje nekoliko široko prilagođenih RIL-ova.

1. UVOD

Pšenica (*Triticum aestivum* L.) je jedan od najvažnijih usjeva u svijetu za proizvodnju hrane. Uzgaja se na širokom području između 30° i 60° sjeverne te između 27° i 40° južne geografske širine (Nuttonson, 1955), a ukupna svjetska proizvodnja 2020. godine iznosila je 760,9 milijuna tona (FAO, 2020). Važnost pšenice za ljudsku prehranu potvrđuje i činjenica da su proizvodi dobiveni od pšenice najvažniji izvor prehrambenih proteina i energije za čovječanstvo (Shewry i Hey, 2015). S obzirom na ubrzani rast svjetske populacije i povećanu potrebu za proizvodnjom hrane, važnost pšenice će sve više rasti, stoga oplemenjivački programi pšenice najveći naglasak stavljaju na povećanje prinosa pšenice. Prinos i sadržaj proteina u zrnu pšenice pokazuju visoku negativnu korelaciju što znači da povećanje prinosa pšenice obično dovodi do smanjenja sadržaja proteina te, posljedično, smanjenja kakvoće pšeničnih proizvoda (Bordes i sur., 2011). Kako bi se ublažila negativna povezanost ova dva važna svojstva pšenice i omogućila istovremena visoka rodnost i zadovoljavajuća kakvoća pšenice oplemenjivački programi bi se trebali snažnije usmjeriti i prema poboljšanju kakvoće zrna. Međutim, poboljšanje kakvoće zrna, a posebno poboljšanje pekarske kakvoće pšenice, jedan je od najzahtjevnijih ciljeva u oplemenjivanju pšenice budući da su svojstava kakvoće većinom kvantitativna svojstva čija je manifestacija pod utjecajem velikog broja gena.

Kakvoća zrna pšenice procjenjuje se na osnovi fizičkih karakteristika zrna, zdravstvenog stanja, te kemijskog sastava zrna. Za ljudsku prehranu među najznačajnijim svojstvima kakvoće su sadržaj proteina u zrnu (eng. *grain protein content*; GPC) te sadržaj različitih frakcija proteina, među kojima je najvažniji gluten. Gluten je najzastupljeniji protein pšenice koji čini otprilike 75 – 80 % ukupnog GPC-a, a njegov sadržaj i kakvoća najvažniji su čimbenici koji određuju tehnološka svojstva brašna i tijesta (Payne i sur., 1987). Kako bi se točnije procijenila pekarska kakvoća pšenice, osim kvalitativne i kvantitativne analize sadržaja proteina, obično su potrebna i reološka ispitivanja koja simuliraju pečenje tijesta pritom ocjenjujući njegova viskoelastična svojstva. Za razliku od ostalih reoloških uređaja miksograf može dati precizne informacije o miješanju i razvoju tijesta, kao i o njegovoj jačini, na temelju vrlo male količine uzorka brašna (2 – 35 g). Na osnovi reološkog profila tijesta dobivenog miksografom i informacija o GPC-u i sadržaju glutena može se dobiti pouzdana procjena pekarske kakvoće (Graybosch i sur., 1999). S obzirom na vrlo malu količinu uzorka koja je potrebna za analizu, miksograf je posebno prikladan za uporabu u oplemenjivanju bilja u ranim generacijama kada nisu dostupne velike količine zrna i brašna (Gras i O'Brien, 1992).

Uspjeh oplemenjivačkog programa s ciljem poboljšanja kakvoće pšenice ovisi o sposobnosti razvoja genotipa s optimalnim vrijednostima i visokom stabilnošću svojstava kakvoće. U tom kontekstu postojanje interakcije genotip × okoliš (eng. *genotype-by-environment interaction*; GEI) jedan je od glavnih izazova s kojima se oplemenjivači susreću jer otežava izbor široko prilagođenog genotipa. Kako bi se postigla učinkovitija selekcija linija u oplemenjivačkim programima i razvile sorte poboljšanih i stabilnih svojstava kakvoće, važno je istražiti utjecaj genotipa, okoliša i GEI na svojstvo od interesa (Bustos-Korts i sur., 2019). Jedan od najčešće korištenih modela za analizu GEI je model aditivnih glavnih učinaka i multiplikativne interakcije (eng. *additive main effects and multiplicative interaction*; AMMI) (Gauch, 1992). Iako je češće korišten u višeokolišnim analizama agronomski važnih svojstva kao što je prinos, AMMI model koristi se također i za analizu GEI i stabilnost svojstava kakvoće pšenice (Mut i sur., 2010).

Većina svojstava kakvoće pšenice su kvantitativna svojstva, a budući da je nasljednost takvih svojstava često niska, tradicionalno korištene metode oplemenjivanja mogu biti dugotrajne, skupe i u konačnici dovesti do nepouzdanih rezultata. Smanjenje troškova i razvoj novih metoda genotipizacije omogućilo je široku dostupnost biljega visoke gustoće te dovelo do njihove učestalije primjene u oplemenjivanju bilja (Crossa i sur., 2013). Jedan od sve češće korištenih pristupa biljezima potpomognutoj selekciji (eng. *marker assisted selection*; MAS) je i genomska selekcija, predstavljena kao inovativan pristup oplemenjivanju još 2001. godine kada su Meuwissen i sur. (2001) zaključili da je pomoću biljega visoke gustoće moguće precizno odrediti oplemenjivačku vrijednost jedinki za koju nisu dostupni fenotipski podatci. Za predviđanje oplemenjivačke vrijednosti nefenotipiziranih linija unutar oplemenjivačke populacije genomska selekcija koristi model koji je utreniran istovremeno koristeći fenotipske i genotipske informacije svih jedinki unutar trenažne populacije. Model procjenjuje učinke biljega na promatrano svojstvo, a koji se dalje koriste kako bi se predvidjele fenotipske vrijednosti linija unutar oplemenjivačke populacije. Na osnovi predviđenih vrijednosti vrši se selekcija najboljih linija (Heffner i sur., 2009). Najvažnija prednost genomske selekcije u odnosu na klasične metode oplemenjivanja je povećanje genetičke dobiti uslijed povećanja preciznosti odabira na genotipskoj razini i skraćivanja selekcijskog ciklusa u oplemenjivačkom procesu budući da omogućava predviđanje samo na osnovi genotipa (Sorrells, 2015). Uspješnost genomske selekcije za predviđanje svojstva od interesa ovisi o dobivenoj točnosti predviđanja, a koja je pod utjecajem različitih čimbenika odabranog svojstva (heritabilnost) i korištene populacije (veličina, srodnost). Na točnost predviđanja u određenoj mjeri utječe i postojanje GEI jer je u tom slučaju otežana selekcija na široko prilagođen genotip budući da može postojati i više od jednog najuspješnijeg genotipa (Heslot i sur., 2014). Stoga se prije uključivanja

genomske selekcije u oplemenjivački program za poboljšanje svojstava kakvoće pšenice treba utvrditi utjecaj GEI kao i optimizirati parametri važni za postizanje optimalne vrijednosti točnosti predviđanja (Zhang i sur., 2015).

U oplemenjivanju bilja, rekombinantne inbred-linije (eng. *recombinant inbred line*; RIL) predstavljaju kolekciju genotipova sa značajnim izvorom genetičke raznolikosti. Međutim, kako bi se odabrale široko prilagođene linije s optimalnim vrijednostima svojstava kakvoće, oplemenjivači moraju istražiti utjecaj GEI za svojstva od interesa. Ranija istraživanja su pokazala da su RIL-ovi stabilniji i bolje podnose promjene u okolišima u usporedbi s tradicijskim kultivarima i sortama (Rodriguez i sur., 2008). Također, za primjenu genomske selekcije na visokosrodnim populacijama kao što su biparentalne populacije, npr. RIL-ovi, potrebna je manja gustoća biljega u odnosu na nesrodne populacije, što značajno smanjuje troškove genotipizacije (Heffner i sur., 2011a). Osim toga, što su populacije srodnije, to je potrebna manja veličina trenažne populacije za postizanje jednake točnosti predviđanja, a što u konačnici smanjuje i troškove fenotipizacije (Jannink i sur., 2010).

Uzimajući u obzir često vremenski i financijski zahtjevnu fenotipizaciju kada je riječ o svojstvima kakvoće pšenice, korištenje klasičnih metoda oplemenjivanja može biti dugotrajno i skupo. Pronalaženje optimalnog modela i parametara genomske selekcije omogućilo bi korištenje molekularnih biljega u postupku predselekcije za svojstva kakvoće zrna te racionalizaciju klasičnih metoda oplemenjivanja pšenice. Cilj istraživanja provedenog u okviru ovog doktorskog rada je procjena utjecaja GEI i optimizacija genomske selekcije za svojstva kakvoće zrna korištenjem biparentalnih populacija pšenice kako bi se smanjili potencijalni troškovi genotipizacije i fenotipizacije u oplemenjivačkom procesu. Dobiveni rezultati pomoći će oplemenjivačima u razvoju strategija koje se zasnivaju na genomskoj selekciji i doprinijeti učinkovitijem razvoju novih linija i sorti.

1.1. Hipoteze i ciljevi istraživanja

Glavne hipoteze ovog istraživanja su:

1. Korištenjem genomske selekcije moguće je s visokom točnošću predvidjeti svojstva kakvoće zrna pšenice i odabrati RIL-ove s optimalnom GEBV vrijednosti.
2. Na točnost predviđanja modela značajno utječe heritabilnost (nasljednost) svojstva, veličina trenažne populacije i veličina skupa korištenih biljega.
3. Na procjenu široko prilagođenih RIL-ova značajno utječe interakcija genotip × okoliš.

Ciljevi ovog istraživanja su:

1. Utvrditi potencijal genomske selekcije za predviđanje svojstava kakvoće zrna pšenice.
2. Optimizirati parametre korištene u genomskoj selekciji s ciljem budućeg smanjenja troškova genotipizacije i fenotipizacije u selekcijskom procesu.
3. Procijeniti utjecaj interakcije genotip × okoliš.
4. Identificirati široko prilagođene RIL-ove s visokim i stabilnim vrijednostima svojstava kakvoće zrna pšenice u ovisnosti o okolišu.

2. PREGLED RELEVANTNE LITERATURE

2.1. Pšenica

2.1.1. Taksonomija i porijeklo pšenice

Pšenica je jedna od najvažnijih biljnih vrsta u svijetu za proizvodnju hrane. Obuhvaća nekoliko vrsta roda *Triticum*, porodice trava (*Poaceae, Gramineae*), a najvažniji predstavnik je vrsta *Triticum aestivum* L., odnosno obična, meka ili krušna pšenica. Iako ostali pripadnici roda *Triticum* postoje kao divlji ili kultivirani oblici, vrsta *T. aestivum* L. postoji isključivo u kultiviranom obliku. Kultivacija pšenice seže daleko u prošlost – prije otprilike 8000 godina (Montenegro i sur., 2017). Porijeklom iz jugozapadne Azije, iz riječne doline Eufrata i Tigrisa, takozvane „kolijevke civilizacije“ (u blizini današnje države Irak), pšenica se do danas proširila po gotovo cijelom svijetu.

2.1.2. Značaj i potreba za pšenicom u svijetu

Kao jedan od najranije kultiviranih usjeva, pšenica je bila osnovni izvor hrane svim velikim civilizacijama Europe, Zapadne Azije i Sjeverne Amerike. Danas se pšenica uzgaja na širokom području između 30° i 60° sjeverne te između 27° i 40° južne geografske širine (Nuttonson, 1955). Prema podatcima Organizacije za hranu i poljoprivredu (eng. *Food and Agriculture Organization*; FAO) za 2020. godinu pšenica se uzgajala na ukupno 219 milijuna hektara u svijetu što je rezultiralo proizvodnjom od 760,9 milijuna tona. U ukupnoj svjetskoj proizvodnji pšenice na Aziju otpada 45,7 %, dok Europa u proizvodnji pšenice sudjeluje s 33,5 %. Najveći proizvođač pšenice u 2020. godini bila je Kina, koju su po proizvodnji pratile Indija i Rusija. U Hrvatskoj se pšenica 2020. godine uzgajala na površini od 147,8 tisuća hektara, a ukupna proizvodnja iznosila je 847,5 tisuća tona (FAO, 2020). Prema procjenama Međunarodnog centra za oplemenjivanje kukuruza i pšenice (eng. *International Maize and Wheat Improvement Center*; CIMMYT) potreba za pšenicom bi uslijed porasta ljudske populacije do 2050. godine mogla porasti za 50 % (CIMMYT, 2022). Važnost pšenice za ljudsku prehranu potvrđuje i činjenica da otprilike 20 % kalorijskog unosa svjetske populacije otpada na pšenicu i proizvode dobivene od pšenice (Shewry i Hey, 2015). Pšenično brašno najčešće se koristi za dobivanje kruha, tjestenine i žitnih pahuljica.

Sustavni rad na oplemenjivanju pšenice započinje početkom 19. stoljeća, a od tada do danas učinjeni su znatni napori kako bi se poboljšala kakvoća pšenice, povećao prinos, otpornost na bolesti i štetnike te različite nepovoljne uvjete abiotičke prirode. S obzirom na ubrzani rast svjetske populacije i povećane potrebe za proizvodnjom hrane, najveći naglasak u oplemenjivanju pšenice stavlja se na povećanje prinosa zrna. Međutim,

povećanje prinosa obično za sobom povlači i smanjenje kakvoće, odnosno sadržaja proteina u zrnu. Stoga bi se u oplemenjivačkim programima jednak naglasak trebao staviti i na poboljšanje svojstava kakvoće zrna (Guzman i sur., 2016).

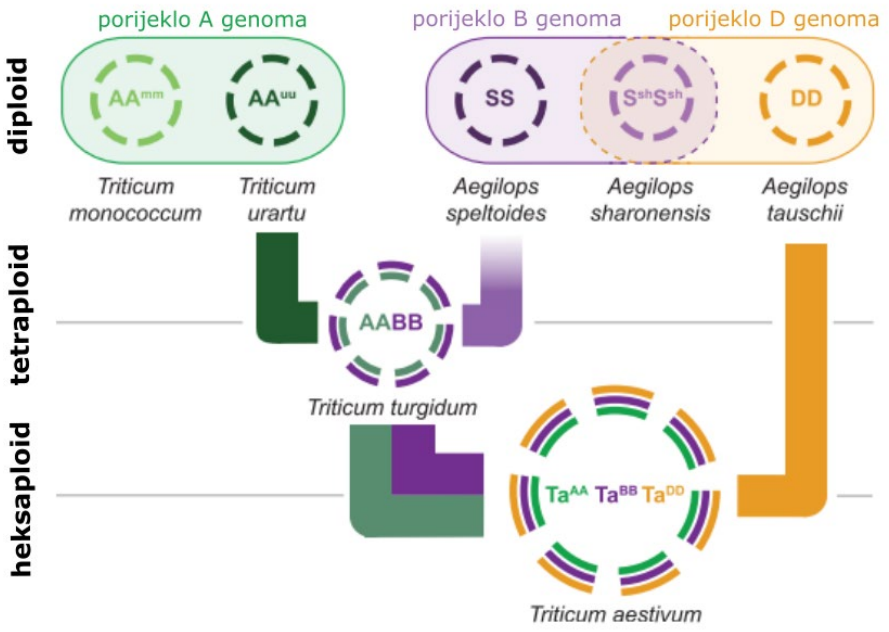
2.1.3. Genom pšenice

Krušna pšenica je heksaploidna vrsta nastala uslijed nekoliko hibridizacijskih događaja između praroditeljskih vrsta (Slika 1) (The International Wheat Genome Sequencing Consortium (IWGSC), 2014). Pretpostavlja se da se prvi takav hibridizacijski događaj koji je vodio razvoju krušne pšenice dogodio prije otprilike 500 000 godina između diploidne divlje vrste *Triticum urartu* (2n = 2x = 14; AA genom) i bliskog srodnika vrste *Aegilops speltoides* (2n = 14; SS genom) čime je nastala tetraploidna vrsta *T. turgidum* (2n = 4x = 28; AABB genom). U drugom hibridizacijskom događaju između vrste *T. turgidum* i diploidne vrste *Aegilops tauschii* (2n = 2x = 14; DD genom) nastao je predak današnje heksaploidne pšenice *T. aestivum* (2n = 6x = 42, AABBDD genom) koji je odonda kultiviran kao krušna pšenica te danas čini više od 95 % pšenice koja se uzgaja diljem svijeta (The International Wheat Genome Sequencing Consortium (IWGSC), 2014). Rezultat takve hibridizacije je da genom vrste *T. aestivum* čini 21 kromosom podijeljen u 7 homeologni grupa (Petersen i sur., 2006). Svaka grupa sadrži po jedan homeologni kromosom iz A, B i D genoma (The International Wheat Genome Sequencing Consortium (IWGSC), 2014). Kopije genoma pšenice pokazuju međusobno više od 95 % sličnosti kodirajućih regija te obično imaju visokokonzerviranu strukturu gena (Adamski i sur., 2020).

Genom krušne pšenice čini oko 16 milijardi parova baza (16 Gbp) s oko 85 % ponavljajućih elemenata, a više od 90 % ukupnog genoma čine nekodirajuće regije (Adamski i sur., 2020). Sadrži više od 133 000 gena od kojih se najveći dio nalazi na B genomu (35 %), zatim na A genomu (33 %), dok se najmanji dio do danas kartiranih gena nalazi na D genomu (32 %) (The International Wheat Genome Sequencing Consortium (IWGSC), 2014). Zahvaljujući napretku postupaka genotipizacije i sekvenciranja tijekom posljednjih desetljeća genom pšenice u potpunosti je sekvenciran i označen 2018. godine (The International Wheat Genome Sequencing Consortium (IWGSC), 2018). Postojanje referentne genomske sekvence za pšenicu omogućava oplemenjivačima lakši pristup informacijama na razini sekvence i proučavanje ekspresije gena u svim razvojnim fazama biljke s ciljem preciznog određivanja ciljanih gena za oplemenjivanje na različita svojstva.

Iako je po strukturi genoma heksaploid, pšenica se genetički gledano ponaša kao diploid budući da geni na *Ph1* lokusu omogućavaju pravilno sparivanje homologni kromosoma i rekombinaciju te sprječavaju sparivanje centromera nehomologni kromosoma (Martinez-Perez i sur., 2001). Zahvaljujući heksaploidnoj strukturi genoma,

krušna pšenica posjeduje sposobnost prilagodbe širokom rasponu klimatskih uvjeta, uključujući i prilagodbu na velike varijacije u vlažnosti zraka tijekom ljetnog perioda, hladnoću tijekom zimskog perioda te kratak fotoperiod (Matsuoka, 2011).



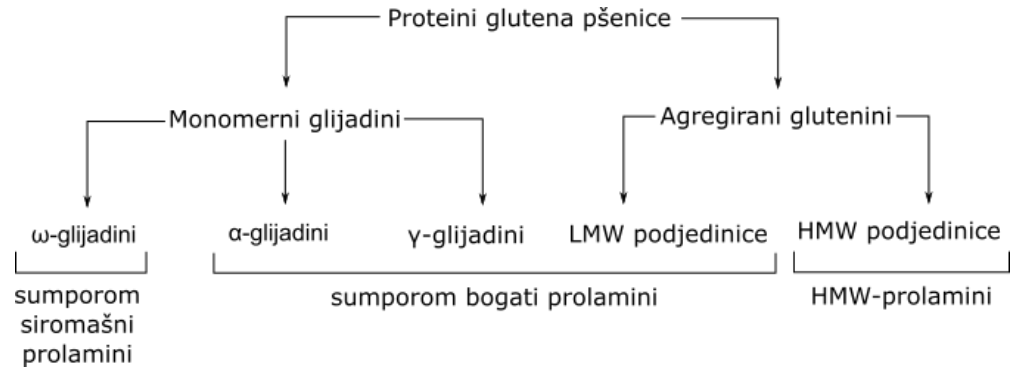
Slika 1. Shematski prikaz genoma pšenice i hibridizacijskih događaja (prema The International Wheat Genome Sequencing Consortium (IWGSC), 2014).

2.2. Kakvoća pšenice

Pod pojmom kakvoće pšenice podrazumijeva se niz fizičkih i zdravstvenih karakteristika zrna, te kemijski sastav zrna. Fizičke karakteristike zrna na osnovi kojih se procjenjuje kakvoća obuhvaćaju svojstva kao što su masa tisuću zrna (eng. *thousand kernel weight*; TKW), hektolitarska masa (eng. *test weight*; TW), apsolutna masa, specifična masa, morfometrijske odlike zrna, boja, čistoća, brašnavost, caklavost i tvrdoća zrna. Zrno visoke kakvoće odlikuje TW veći od 76 kg, TKW u rasponu od 38 do 40 g, te specifična masa u rasponu od 1,32 do 1,42 g/cm^3. Caklavo zrno znak je bogatstva proteinima. Zdravstveno stanje i svježina zrna određuju se organoleptički, odnosno pomoću osjetila vida, mirisa i okusa, pri čemu se ocjenjuju boja, sjaj, miris, okus i klijavost. Komponente kemijskog sastava zrna uključuju vodu, mineralne te organske tvari koje čine najveći dio mase suhog uskladištenog zrna. Najveći maseni udio u organskoj tvari meke pšenice zauzimaju nedušične ekstraktivne tvari (NET) (64 – 76 %), zatim proteini (12 – 16 %), celuloza (2 – 3 %) i masti (2 %). NET većinskim dijelom čini škrob (90 %), dok ostatak čine topivi šećeri (Kovačević i Rastija, 2014).

2.2.1. Proteini pšenice

Za ljudsku prehranu među značajnijim svojstvima kakvoće pšenice su sadržaj i frakcije proteina. Mogućnost prerade pšeničnog brašna u različite proizvode uvelike je određena prisutnošću specifičnih proteina i njihovih frakcija. Frakcije proteina čine dvije skupine: fiziološki aktivni proteini topivi u vodi (albumin i globulin) i rezervni proteini netopivi u vodi (glijadin i glutenin). Rezervni proteini čine 80 – 85 % ukupnih proteina pšeničnog brašna. Prema njihovoj elektroforetskoj pokretljivosti pri niskom pH glijadini se dijele u tri velike skupine: α-, γ- i ω-glijadini (Wieser, 2007). Glutenini su proteinske podjedinice puno veće molekularne mase od glijadina, a dijele se na podjedinice visoke molekularne mase (eng. *high molecular weight*; HMW) i podjedinice niske molekularne mase (eng. *low molecular weight*; LMW) (Slika 2). Krušna pšenica sadrži 3 – 5 HMW podjedinica, te oko 15 LMW podjedinica, a sam udio glutenina neovisan je o razlikama u ukupnom sadržaju proteina (Kolster, 1992). Kompleksna smjesa monomernih glijadina i polimernih podjedinica glutenina nastala miješanjem uz prisutnost vode naziva se gluten ili ljepak, a određuje tehnološka svojstva brašna i tijesta (Elli i sur., 2017; Shewry, 2004). Gluten je najzastupljeniji protein u zrnu pšenice te čini 75 – 80 % ukupnog sadržaja proteina u zrnu (Kristensen i sur., 2018). Sadržaj i kakvoća glutena određeni su omjerom glijadina i glutenina te njihovom kakvoćom, što posljedično utječe i na kakvoću brašna i proizvoda dobivenih od brašna (Macritchie, 1992; Payne i sur., 1987; Rasheed i sur., 2014).



Slika 2. Klasifikacija proteinskih podjedinica glutena pšenice (prema Elli i sur., 2017).

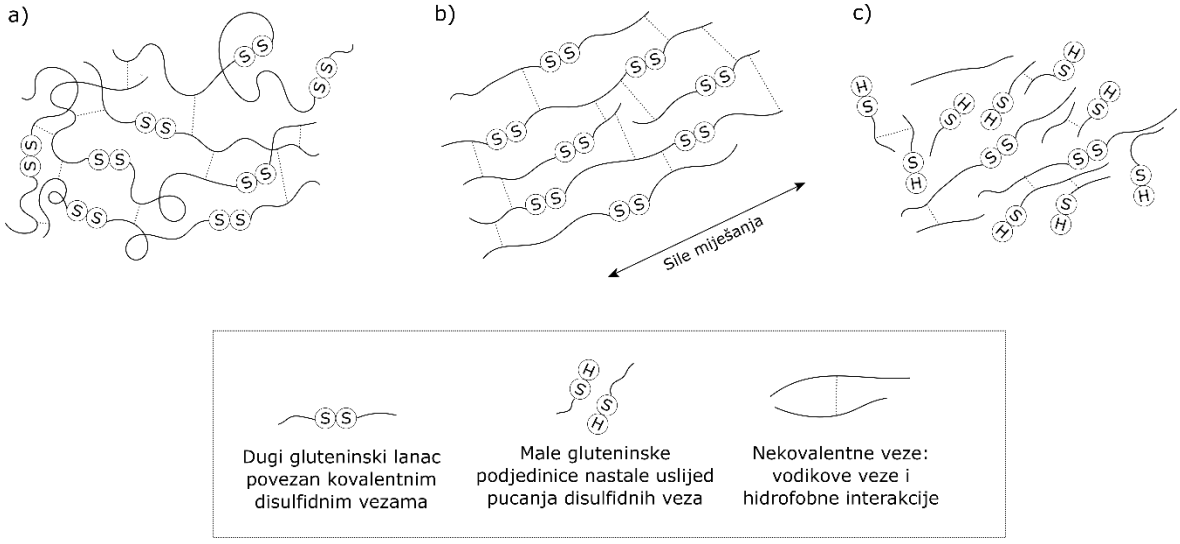
Na količinu proteina i njihovih frakcija u zrnu pšenice značajan utjecaj imaju vremenske prilike (količina padalina, temperatura) pri čemu sušni uvjeti pogoduju povećanju količine i kakvoće proteina (Hernández-Espinosa i sur., 2018). Osim vremenskih prilika, izravan utjecaj na količinu proteina, odnosno dušika, ima mineralna ishrana (Daniel i Triboi, 2000). Povećana ishrana dušikom pokazuje pozitivan utjecaj na povećanje količine proteina što za posljedicu ima i povećanje kakvoće brašna i tijesta (Zörb i sur., 2018). Uslijed povećane ishrane dušikom dolazi do povećanja udjela glijadina te smanjenja udjela albumina i globulina (Kolster, 1992). Prema istraživanju Horvat i sur. (2021) povećanje razine dušika u ishrani dovodi do povećanja sadržaja proteina u zrnu, dok sastav glijadinskih i gluteninskih podjedinica pokazuje manju ovisnost o razini dušika te je izrazito sortno specifičan. Drugo istraživanje Horvat i sur. (2022) pokazalo je da tretman ureom također značajno utječe na povećanje ukupnog GPC-a kao i na poboljšanje svojstava kakvoće povezanih s količinom proteina, dok se sastav skladišnih proteina mijenja u manjoj mjeri. Sadržaj proteina i njegova kakvoća također su sortno specifični te su pod snažnim utjecajem GEI (Groos i sur., 2003).

2.2.2. Čimbenici kakvoće i formiranje tijesta

Kakvoća tijesta složeno je svojstvo pšenice pod utjecajem mnoštva biokemijskih karakteristika kao što su sadržaj i kakvoća proteina, škroba i masti (Payne i sur., 1987), te različitih fizikalno-kemijskih svojstava kao što su sadržaj vlage, kapacitet zadržavanja vode, formiranje šupljina itd. (Huang i sur., 2006). Prilikom miješanja s vodom proteini glutena formiraju glutensku mrežu čija je uloga zadržavanje ugljikovog dioksida koji nastaje fermentacijom uz pomoć kvasca, dok masti u interakciji s proteinima doprinose elastičnosti glutenske mreže. Količina i kakvoća glutena najvažniji su čimbenici kakvoće tijesta pri čemu je također dokazano da su HMW gluteninske podjedinice od najvećeg značaja za kakvoću pšeničnog brašna te tijesta (Kolster, 1992; Payne i sur., 1987; Weegels i sur., 1996). Povećanjem ukupnog sadržaja glutena u brašnu raste i kapacitet zadržavanja vode, vrijeme razvoja tijesta, rastezljivost i energija tijesta. Međutim, osim količine glutena, od iznimne je

važnosti i njegova kakvoća, tj. njegova rastezljivost i elastičnost. Kakvoću glutena određuje omjer glijadina i glutenina u brašnu (Đaković, 1980). Prilikom formiranja tijesta glutenini se kovalentno vežu u veliku elastičnu mrežu dajući tijestu čvrstoću, dok glijadini djeluju kao „plastifikatori“ povećavajući viskoznost i rastezljivost što utječe na reološka svojstva tijesta (Létang i sur., 1999). U pšeničnom brašnu omjer glijadina i glutenina je otprilike 1:1 što rezultira stvaranjem glutena dobre kakvoće koji daje čvrsto i elastično tijesto te porozan kruh dobre kakvoće i velikog volumena.

Reološka svojstva tijesta posljedica su prisutnosti proteinske mreže koju čine dugačke lančaste proteinske molekule međusobno povezane disulfidnim (-S-S-) vezama i vodikovim vezama (Slika 3). Uslijed raskidanja disulfidnih veza nastaju tiolne skupine (-SH). Što je veća koncentracija disulfidnih veza tijesto je čvršće, dok povećanjem koncentracije tiolnih skupina dolazi do omekšavanja tijesta i nepovratnih deformacija viskoznosti tijesta (Létang i sur., 1999). Kakvoća glutena ovisi i o sposobnosti bubrenja te intenzitetu razgradnje proteina. Bubrenje glutena proces je prilikom kojeg dolazi do koloidnog vezanja vode. Bubrenje uzrokuje sve jače razmicanje proteinskih molekula te posljedično njihovu razgradnju. Uslijed presnažnog bubrenja dolazi do razaranja unutrašnje građe glutena pri čemu gluten više ne može zadržati vodu. Gluten gubi elastičnost, postaje rastezljiv, mekan te se lako raspada. Što je gluten nekog brašna sposobniji pri vezivanju vode povećavati volumen, a da se pritom što manje raspada, to je brašno veće kakvoće (Đaković, 1980).



Slika 3. Molekularna interpretacija razvoja glutena (a) na početku zamjesa tijesta, (b) u trenutku optimalnog razvoja, (c) uslijed prekomjernog zamjesa tijesta (prema Létang i sur., 1999).

2.2.3. Metode procjene kakvoće tijesta

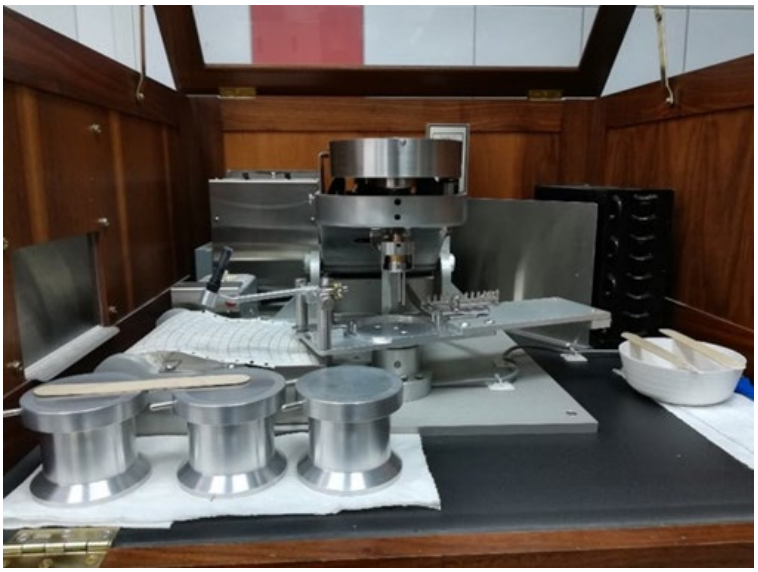
S ciljem boljeg razumijevanja kakvoće brašna i tijesta razvijene su mnogobrojne izravne i neizravne metode procjene. Najčešće korištena izravna metoda je mjerenje volumena kruha. Unatoč tome što zahtijeva veliku količinu brašna uz veliki utrošak vremena, procjena volumena kruha jedno je od najvažnijih i najpouzdanijih mjerila kakvoće tijesta (Weegels i sur., 1996).

Budući da izravne metode uglavnom zahtijevaju veliki utrošak vremena i financijskih sredstava, za procjenu kakvoće tijesta često se koriste tzv. neizravne metode koje uključuju analizu GPC-a, sadržaja vlažnog glutena (eng. *wet gluten content*; WGC), mjerenje TKW i TW, test sedimentacijskog volumena, te procjene reoloških svojstava korištenjem ekstenzografa, alveografa, farinografa i miksografa. GPC se obično koristi kao indikator pekarske kakvoće (Liu i sur., 2016), dok WGC opisuje sposobnost brašna da apsorbira vodu i formira glutensku mrežu te ukazuje na stabilnost tijesta (Lado i sur., 2018). Za dobivanje tijesta i proizvodnju kruha visoke kakvoće GPC bi trebao biti viši od 12,5 % (Turner i sur., 2004).

Dosadašnja istraživanja pokazuju da je GPC često pod snažnim utjecajem okoliša, odnosno da je heritabilnost navedenog svojstva obično vrlo niska (Simmonds, 1995). TKW je jedna od najstabilnijih komponenti prinosa, a zajedno s TW značajno doprinosi kakvoći tijesta utječući na karakteristike mljevenja i izbrašnjavanja (Hook, 1984). Reološki testovi simuliraju pečenje tijesta pri čemu se procjenjuju viskoelastična svojstva te svojstva zamjesa tijesta. Ekstenzograf je reološki uređaj koji mjeri silu potrebnu za razvlačenje tijesta u jednom pravcu pri konstantnoj brzini, dok alveograf mjeri pritisak potreban za napuhati mjehurić zraka u komadu tijesta pritom mjereći sposobnost razvlačenja tijesta u svim smjerovima. Za određivanje jakosti brašna, odnosno otpora tijesta na miješanje i gnječenje, najčešće se koriste farinograf ili miksograf. Iako farinograf ima nekoliko važnih prednosti nad miksografom (kontrola temperature i točnost određivanja apsorpcije vode), njegov glavni nedostatak je velika količina brašna koja je potrebna za provođenje analize (Mani, 2007).

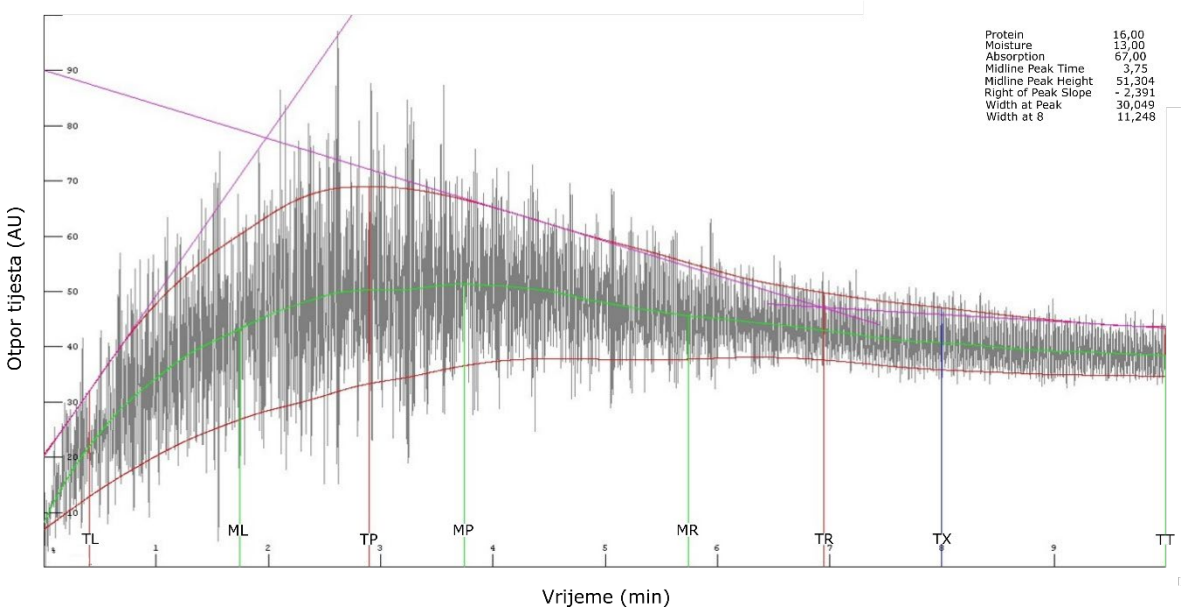
Miksograf je reološki uređaj za mjerenje otpora tijesta na miješanje i ponašanja tijesta prilikom razvoja, prvi put predstavljen još 1933. godine (Swanson i Working, 1933) (Slika 4). Zahvaljujući maloj količini uzorka (2 – 35 g) koja je potrebna za provođenje analize i relativno brzoj interpretaciji dobivenih rezultata, pogodan je za korištenje u oplemenjivanju bilja i to posebno u ranim fazama oplemenjivanja (pojedinačne biljke u cijepajućim generacijama) kada nije dostupna velika količina zrna i brašna (Gras i O'Brien, 1992). Koristi se za dobivanje općenitih informacija o svojstvima miješanja tijesta (vrijeme razvoja tijesta,

otpornost na miješanje) te o snazi tijesta (Johnson i sur., 1943; Swanson, 1993). U kombinaciji s GPC i WGC svojstvima daje pouzdanu procjenu pekarske kakvoće (Graybosch i sur., 1999; Martinant i sur., 1998; Ohm i Chung, 1999).



Slika 4. Miksograf uređaj za određivanje kakvoće brašna.

Rezultati analize miksografom prikazuju se krivuljom – miksogramom (Slika 5). Korištenjem pripadajuće programske podrške za miksograf (MixSmart®) za svaki uzorak kreira se krivulja s 40 zasebnih varijabli koje daju različite informacije o kakvoći tijesta (Chung i sur., 2001; Prashant i sur., 2015).



Slika 5. Primjer miksograma za pšenično brašno.

Vrijeme (izraženo u minutama ili sekundama) potrebno da središnja krivulja dosegne svoju maksimalnu visinu (eng. *midline peak time;* MPT) odgovara optimalnom vremenu zamjesa tijesta tj. optimalnom vremenu potrebnom za razvoj tijesta i postizanje maksimalnog otpora tijesta. Otpornost tijesta na miješanje (eng. *overmixing*) opisana je pomoću nekoliko različitih varijabli kao što su visina krivulje u specifičnom vremenu nakon postignutog vrhunca krivulje te kut između rastućeg i padajućeg dijela krivulje. Uzlazni nagib krivulje govori u kojoj se mjeri razvija tijesto dok silazni pokazuje slabljenje tijesta i njegovu stabilnost. Što je kut između rastućeg i padajućeg nagiba tuplji to je veća tolerancija tijesta prema miješanju. Širina vrha središnje krivulje na kraju perioda miješanja (eng. *midline peak width*; MTW) izražena u postotku ukazuje na konzistentnost i stabilnost tijesta na kraju procesa miješanja. Površina ispod središnje krivulje od početne točke zamjesa do kraja procesa miješanja (eng. *midline curve integral*; MTI) opisuje energiju uloženu tijekom procesa zamjesa tijesta, a ovisi o snazi tijesta i njegovoj otpornosti (Johnson i sur., 1943). Visina vrha središnje krivulje (eng. midline *peak height;* MPH) izražena u postotku daje informaciju o snazi brašna (Mani, 2007). Brašna dobre kakvoće na miksogramu će pokazati visoku apsorpciju vode, umjereno vrijeme miješanja (3 – 6 minuta), visoku snagu glutena te dobru otpornost na miješanje tijesta.

2.3. Oplemenjivanje pšenice na kakvoću zrna

Oplemenjivanje je postupak kojim se poboljšavaju nasljedne osobine ciljnih svojstava bilja, a koji rezultira stvaranjem novih, genetički poboljšanih sorti (Borojević, 1981). Tradicionalni oplemenjivački programi oslanjaju se uglavnom na ocjenjivanje ciljnih svojstava (fenotipova) pojedinačnih biljaka te njihovih srodnika, od stanične razine do razine cijelog organizma. Na osnovi zabilježenog, odabiru se jedinke koje postižu najbolje vrijednosti poželjnih svojstava, a koje ulaze u daljnji proces selekcije i križanja. Fenotipizacija bilja najčešće se provodi na više različitih svojstava u velikim populacijama uzgajanim na nekoliko lokacija kroz dvije ili više godina (Singh i sur., 2016).

S obzirom na ubrzani rast svjetske populacije i povećanu potrebu za hranom najveći naglasak u oplemenjivanju pšenice stavlja se na povećanje prinosa zrna. Međutim, povećanje prinosa obično za sobom povlači smanjenje kakvoće zrna (Simmonds, 1995). Kako bi se izbjeglo njezino značajno smanjenje oplemenjivački programi trebaju jednak naglasak stavljati i na poboljšanje kakvoće pšenice (Guzman i sur., 2016). Poboljšanje kakvoće jedan je od najzahtjevnijih ciljeva u oplemenjivanju pšenice budući da većina svojstava kakvoće pokazuje složene obrasce nasljeđivanja. Klasični oplemenjivački programi usmjereni su uglavnom na poboljšanje sadržaja i sastava proteina što je najvažniji kriterij kakvoće pšenice, posebno u pogledu pekarske kakvoće. Najčešće korištena metoda u oplemenjivanju je križanje roditeljskih genotipova s ciljem razvoja populacije sa širokom genetičkom varijabilnošću. Nakon križanja, različita svojstva roditeljskih genotipova rekombiniraju se u potomstvu iz kojeg se odabiru stabilne, homozigotne linije. Kako bi se osiguralo da se svaka moguća rekombinacija u segregirajućoj populaciji dogodi najmanje jednom, potrebno je razviti dovoljno veliku populaciju koja omogućuje odabir genotipova s poželjnim svojstvima, a čija je frekvencija pojavljivanja niska. Međutim, oplemenjivanje na kakvoću pšenice često je izazovno jer fenotipizacija zahtijeva veće količine zrna i brašna, koji obično nisu dostupni u ranoj fazi oplemenjivačkog procesa (Bedo i sur., 2017).

Klasične metode oplemenjivanja, uključujući fenotipsku selekciju poželjnih svojstava, predstavljaju usko grlo oplemenjivanja iz više razloga. Naime, velik broj promatranih svojstava kao i velik broj parcela ili biljaka na kojima se svojstva procjenjuju, zahtijevaju velik utrošak vremena te sofisticiranu opremu i mjerne uređaje što mjerenja ujedno čini i financijski zahtjevnima. Osim toga, fenotipske vrijednosti kvantitativnih svojstava, kao što su svojstva kakvoće pšenice, snažno ovise o utjecaju okoliša što posljedično zahtijeva procjenu svojstva u nekoliko ponavljanja i u različitim okolišima. Zbog utjecaja okoliša na fenotipske vrijednosti svojstava i pogrešaka prilikom mjerenja, fenotip najčešće nije savršeni pokazatelj genotipskog potencijala biljke (Newell i Jannink, 2014).

2.4. Interakcija genotip × okoliš

Glavni cilj oplemenjivača bilja može se opisati kao stvaranje i selekcija genotipova koji će biti prilagođeni na različite uvjete uzgoja (meteorološke prilike, svojstva tla i čimbenike upravljanja prisutne na određenoj lokaciji). Genotipovi često pokazuju različitu osjetljivost na promjene okolišnih uvjeta što se objašnjava pojavom GEI. Fenotipske vrijednosti kvantitativnih svojstava pod snažnim su utjecajem okoliša te često i interakcije između genotipa i okoliša. Stoga se fenotip, odnosno očekivana vrijednost fenotipa, može iskazati sljedećom jednadžbom:

$$P = \mu + G + E + (G \times E)$$

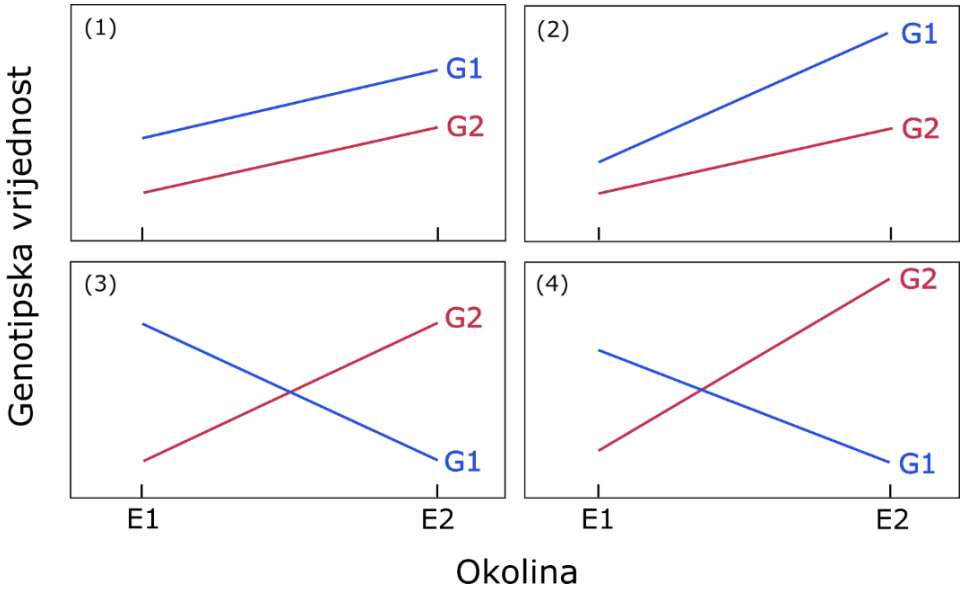
gdje je P vrijednost fenotipa kvantitativnog svojstva, μ ukupna srednja vrijednost, G učinak genotipa, E učinak okoliša, a $G \times E$ učinak GEI (Singh i Singh, 2015).

GEI se može definirati kao razlika između opažene fenotipske vrijednosti i očekivane fenotipske vrijednosti uzimajući u obzir odgovarajuće genotipske i okolišne vrijednosti (Baker, 1988), odnosno kao varijacija uzrokovana udruženim učincima genotipa i okoliša (Dickerson, 1962). Okoliš predstavlja sve negenetičke čimbenike koji mogu utjecati na fenotipske vrijednosti određenog genotipa. Okolišne varijable predstavljaju sve fizikalne i kemijske odlike tla, klimatske čimbenike kao što su količina padalina i temperatura, količina, raspodjela i kvaliteta sunčeve svjetlosti, te žive organizme (npr. patogene) kojima su biljke izložene (Bernardo, 2010).

Jačina GEI odražava se na adaptabilnost i stabilnost genotipa. Adaptabilnost genotipa definira se kao genetička sposobnost genotipa da se prilagodi različitim okolišnim uvjetima (Finlay i Wilkinson, 1963). Ovisno o širini raspona okolišnih uvjeta kojima je genotip sposoban prilagoditi se, može se govoriti o generalnoj ili širokoj adaptabilnosti (genotip je prilagođen na cijeli ili veći dio raspona ciljanih okoliša) te uskoj ili specifičnoj adaptabilnosti (genotip se pokazuje superiornim samo na manjem dijelu raspona ciljanih okoliša). Stabilnost genotipa ovisi o njegovoj sposobnosti reagiranja na okolišne uvjete, a ta pojava se naziva još i fenotipska plastičnost (Bradshaw, 1965). Fenotipska plastičnost može se opisati i kao raspon fenotipova koje jedan genotip može proizvesti u različitim okolišima (Grogan i sur., 2016). Ovisna je o genotipskoj kompoziciji samog genotipa ili sorte, a odražava reakciju pojedinačnog genotipa i cijele populacije na promijenjene okolišne uvjete (Borojević, 1981).

Budući da je cilj svakog oplemenjivačkog procesa stvaranje nove, stabilne sorte poboljšanih svojstava, postojanje GEI jedan je od najvažnijih izazova za oplemenjivače bilja. Da bi se odluka o priznavanju nove sorte mogla donijeti s većom objektivnošću i

sigurnošću, fenotipska procjena svojstava od interesa mora se provoditi u različitim agroekološkim uvjetima odnosno različitim okolišima (na većem broju lokacija kroz nekoliko godina) (Borojević, 1981). Što je učinak interakcije izraženiji, to je teže izabrati superiorne genotipove. Prema Haldane (1947) postojanje GEI važno je samo ako ona dovodi do promjene poretka genotipova između različitih okoliša. Međusobni odnos genotipova u različitim okolišima može pratiti jedan od četiri uzorka (Slika 6) (Bernardo, 2010; De Leon i sur., 2016):



Slika 6. Međusobni odnos genotipova u različitim okolišima (prema Bernardo, 2010).

(1) Genotip 1 (G1) je superiorniji u odnosu na genotip 2 (G2) u oba okoliša (E1 i E2), a razlika između njih je konstantna. U ovom slučaju ne postoji GEI što oplemenjivaču značajno olakšava odabir superiornog genotipa. Genotipove nije potrebno ocjenjivati u višeokolišnim pokusima budući da ovakav odnos označava da će genotip koji je superiorniji u E1 također biti superiorniji i u E2. Genetička varijanca je u ovom slučaju homogena, a između okoliša ne postoji korelacija.

(2) G1 je superiorniji u odnosu na G2 u oba okoliša, ali razlika između njih nije konstantna. U ovom slučaju postoji GEI, ali ona ne uključuje promjenu poretka genotipova. Ovakav odnos naziva se kvantitativna ili *non-crossover* interakcija, a može se pokazati nepouzdanim ukoliko se ciljani raspon okoliša proširi. Genetička varijanca je u ovom slučaju, za razliku od prethodnog, heterogena, a između okoliša ne postoji korelacija.

(3) G2 je superiorniji u odnosu na G1 u okolišu E1, ali za E2 vrijedi suprotno. U ovom odnosu dolazi do promjene poretka genotipova, odnosno isti genotipovi se ne pokazuju jednako uspješnim u različitim okolišima. Ovakav odnos naziva se

kvalitativna ili *crossover* interakcija. U ovom slučaju, iako dolazi do promjene poretka genotipova, njihova apsolutna razlika u uspješnosti ostaje konstantna, odnosno korelacija između uspjeha genotipa u dva različita okoliša je -1, a genetička varijanca je homogena.

(4) Ovaj uzorak također predstavlja kvalitativnu ili *crossover* interakciju, ali za razliku od prethodnog, apsolutna razlika u uspješnosti dva genotipa nije konstantna i mijenja se između različitih okoliša (korelacija je veća od -1). U ovom slučaju genetička varijanca je heterogena.

Sa stajališta oplemenjivanja bilja važnija je kvalitativna ili *crossover* interakcija budući da je cilj oplemenjivanja razviti sorte koje će biti stabilne u različitim okolišima. U slučaju kvalitativne interakcije redoslijed genotipova u različitim okolišima se mijenja, te najuspješniji genotipovi u jednom okolišu mogu u drugom okolišu pokazati izrazito nisku uspješnost (Gauch i Zobel, 1997). Prisutnost GEI negativno utječe i na heritabilnost svojstva. Što je interakcija veća, to je heritabilnost manja što može ograničavati i napredak u selekciji.

Istraživanja na prinosu pšenice u Ujedinjenom Kraljevstvu su pokazala da je u periodu od 1946. do 1977. doprinos okoliša iznosio 40 – 60 %, genotipa 25 – 40 %, dok je doprinos GEI iznosio 15 – 25 % (Simmonds, 1981). U pokusu provedenom na 299 genotipova ozime pšenice u 11 različitih okoliša, Grogan i sur. (2016) potvrdili su značajan utjecaj GEI na svojstva prinos i datum klasanja. Analizirajući varijabilnost navedenih svojstava u odnosu na njihovu srednju vrijednost (plastičnost svojstva), zaključili su i da je dugogodišnja selekcija dovela do pojave da datum klasanja pokazuje smanjeni odaziv na okolišne uvjete, dok je odaziv u slučaju prinosa povećan. Također, značajan doprinos GEI za svojstva kao što su TW i TKW utvrdili su i Brancourt-Hulmel i sur. (2000).

Istraživanja koja su uključivala svojstva kakvoće pšenice pokazala su da je najveći dio fenotipske varijance za svojstvo GPC posljedica negenetičkih čimbenika među kojima je i snažan utjecaj okoliša (Groos i sur., 2003). Iako sadržaj glutena pokazuje pozitivnu korelaciju s GPC, istraživanja pokazuju da je kakvoća glutena pod snažnim utjecajem genotipa (Šimić i sur., 2006). U usporedbi s GPC, reološka svojstva tijesta pokazala su manju ovisnost o utjecaju GEI u istraživanjima Williams i sur. (2008), Hernández-Espinosa i sur. (2018) i Drezner i sur. (2010), dok su druga istraživanja pokazala da je varijanca reoloških svojstava tijesta uzrokovana prisutnošću GEI jednaka ili veća od varijance uzrokovane samo genotipom ili samo okolišem (Grausgruber i sur., 2000; Peterson i sur., 1998). Analizirajući agronomska svojstva i svojstva kakvoće 10 sorti ozime pšenice u periodu od 1997. do 2002. godine, Drezner i sur. (2010) utvrdili su da je utjecaj GEI bio

značajno manji u odnosu na utjecaj genotipa, osim u slučaju prinosa, TKW i Hagbergovog broja padanja. Međutim, unatoč tome što utjecaj GEI za svojstva kakvoće (GPC, WGC, gluten indeks) nije bio velik, autori smatraju da procjena učinka GEI može pružiti bolji uvid u promjene kakvoće različitih sorata u odnosu na okolišne uvjete. Za svojstva GPC, WGC te reološka svojstva mjerena farinografom i ekstenzografom, Grausgruber i sur. (2000) utvrdili su da je GEI bila jednaka ili veća od genotipske komponente ili komponente okoliša. Istraživanje provedeno na 30 genotipova pšenice u 17 okoliša pokazalo je da, iako u manjoj mjeri doprinosi varijabilnosti svojstava u odnosu na genotip ili okoliš, GEI ima značajan utjecaj na svojstva kao što su sadržaj proteina u brašnu, kompozicija proteina, sedimentacijski volumen, te svojstva dobivena analizom na miksograf uređaju (Peterson i sur., 1998). Promatrajući reološka svojstva pšenice dobivena alveografom važna za dobru kakvoću kruha, Robert i Denis (1996) zaključili su da GEI doprinosi varijaciji svojstava s 8 do 21,4 % te u nekim slučajevima značajno utječe na stabilnost svojstava kakvoće.

Prisutnost GEI sugerira postojanje značajnih razlika u okolišnim uvjetima, koje mogu dovesti do promjene u poretku genotipova u različitim okolišima što oplemenjivačima znatno otežava selekciju na poželjna svojstva. Ovisno o jačini GEI, oplemenjivači moraju odlučiti hoće li težiti stvaranju široko ili usko prilagođene sorte te na osnovu toga definirati ciljani raspon okoliša za provođenje pokusa (Bustos-Korts i sur., 2019). Povećanje selekcijske dobiti ovisi o sposobnosti predviđanja uspjeha specifične kombinacije genotip-okoliš, stoga predviđanje budućih kombinacija zahtijeva dublje razumijevanje prethodnih što je uvjetovano detaljnim prikupljanjem podataka i pravilnom analizom istih (van Eeuwijk i sur., 2016). U oplemenjivanju bilja za opisivanje i analizu GEI koriste se različiti statistički modeli koji pružaju informacije o utjecaju okoliša na genotip (osjetljivost genotipa, adaptabilnost, identifikacija najuspješnijeg genotipa u podskupu okoliša). Neki od najčešće korištenih statističkih modela su:

(1) ANOVA – najjednostavniji, potpuni aditivni model;

(2) Regresija na srednju vrijednost okoliša ili Finlay-Wilkinson regresija – najčešće korišten model, ali neinformativan za GEI u slučaju malog broja genotipova ili kombinacija genotip-okoliš;

(3) Bilinearni modeli (AMMI, GGE) – pružaju bolji analitički pristup razumijevanju GEI;

(4) Modeli faktorijalne regresije;

(5) Linearni mješoviti modeli – prikladni za procjenu strukture varijance-kovarijance okoliša te za predviđanje uspjeha genotipova u određenim okolišima (Bustos-Korts i sur., 2019).

2.4.1. AMMI model

AMMI model predstavlja metodu analize GEI koja spaja aditivne i multiplikativne komponente u jedinstvenu analizu (Gauch, 1992). Generalno, AMMI model može se zapisati kao:

$$Y_{ge} = \mu + \alpha_g + \beta_e + \sum_n \lambda_n \gamma_{gn} \delta_{en} + \rho_{ge}$$

gdje je Y_{ge} opažena fenotipska vrijednost genotipa g u okolišu e, μ ukupna srednja vrijednost, α_g odstupanje genotipa g od ukupne srednje vrijednosti, β_e odstupanje okoliša e od ukupne srednje vrijednosti, λ_n svojstvena vrijednost IPCA osi n, γ_{gn} vrijednost svojstvenog vektora genotipa g za IPCA os n, δ_{en} vrijednost svojstvenog vektora okoliša e za IPCA os n, a ρ_{ge} ostatak (Gauch, 2013, 1988).

Prvi dio modela je ANOVA koja uključuje samo glavne učinke (genotip i okoliš), a ostatak se koristi za konstrukciju matrice GEI. Drugi dio modela odnosi se na dekompoziciju matrice GEI na svojstvene vrijednosti i ortogonalne vektore (eng. *singular value decomposition*; SVD) koji tvore od 1 do K broja osi. Navedene osi nazivaju se interakcijske PCA (eng. *principal component analysis*) osi ili IPCA osi. Primjenom AMMI modela GEI se rastavlja na niz K multiplikativnih komponenti (vektora) za svaki genotip i za svaki okoliš čiji odnosi opisuju različitu osjetljivost genotipova na različite (uglavnom nepoznate) prevladavajuće stresove u ispitivanom okolišu (Bustos-Korts i sur., 2019), odnosno. Krajnji rezultat analize AMMI modelom su genotipske i okolišne vrijednosti za K zadržanih osi te razmjeri sume kvadrata interakcije objašnjene multiplikativnim komponentama. Ako je K = 0 (AMMI0 model), nema IPCA osi. U ovom slučaju riječ je o ANOVA modelu, a za opis matrice podataka koristi se aditivni genotipski i okolišni prosjek, na osnovu kojeg se genotipovi rangiraju podjednako u svim okolišima pri čemu se zanemaruje GEI. AMMI0 model u praksi je najčešće nedovoljan, a podcjenjuje učinak interakcija u modelu. Za K = 1 (AMMI1 model) model je predstavljen glavnim učincima (genotip i okoliš) te osi prve glavne interakcijske komponente (IPCA1) za interpretaciju matrice ostatka. Ako je K = 2 (AMMI2), zadržane su prve dvije IPCA osi (IPCA1 i IPCA2). Model je u tom slučaju predstavljen glavnim učincima, a neaditivna varijabilnost tj. interakcija (matrica ostatka) objašnjena je pomoću dvije glavne komponente. Uključivanje dodatnih glavnih komponenata može ići sve do potpunog AMMI modela (AMMIF). Međutim, AMMIF može precijeniti učinak interakcije te se stoga u praksi uglavnom zanemaruju multiplikativne komponente višeg reda (komponente interakcije koje nisu statistički značajne) te se najčešće koriste AMMI1 ili AMMI2 modeli. Također, u svrhu detaljnije interpretacije AMMI rezultata i definiranja točnih okolišnih ili genetičkih čimbenika koji utječu na GEI mogu se izračunati korelacijski

koeficijenti za pojedine IPCA osi i različite genetičke, fiziološke ili okolišne podatke (Gauch i Zobel, 1997; Gauch, 2013).

Za vizualizaciju rezultata AMMI modela najčešće se koristi biplot, vrlo informativan vizualizacijski alat koji prikazuje glavne i interakcijske učinke genotipova i okoliša (Gauch i Zobel, 1997). Smještaj genotipa g na biplotu određen je genotipskim vrijednostima $a1_g$ i $a2_g$, dok je smještaj okoliša e određen vektorima okolišnih vrijednosti $b1_e$ i $b2_e$ (Bustos-Korts i sur., 2019). Netipične vrijednosti genotipova ili okoliša, ili njihovo grupiranje, vrlo se lako mogu uočiti na biplotu (van Eeuwijk i sur., 2016). Udaljenost genotipa ili okoliša od ishodišta proporcionalna je GEI koju taj genotip ili okoliš proizvodi. Genotipovi koji su na biplotu smješteni blizu jedan drugome pokazuju sličnu adaptabilnost. Okoliši smješteni bliže jedan drugome proizvode sličnu GEI, a kut između bilo koja dva vektora okoliša ukazuje na sličnost između okoliša, tj. na genetičku korelaciju. Udaljenost projekcije genotipa g na vektor okoliša e od ishodišta daje informaciju o apsolutnoj dimenziji interakcije genotipa g u okolišu e. Predznak GEI može se procijeniti i iz kuta kojeg zatvaraju vektor genotipa i vektor okoliša. Ako je kut među njima pravi (90°), GEI nije prisutna. Također, okoliši čiji vektori zatvaraju kut od 90° proizvode sličnu GEI za sve genotipove. Ako vektori dva okoliša zatvaraju kut od 180°, genotip koji je dobro prilagođen jednom okolišu istovremeno će biti vrlo slabo prilagođen drugom okolišu. U slučaju da vektori dva okoliša zatvaraju kut od 90°, ponašanje genotipa u jednom okolišu bit će sasvim nezavisno od njegovog ponašanja u drugom okolišu (genetička korelacija je nula) (Bustos-Korts i sur., 2019). Fenotipska stabilnost genotipa može također biti procijenjena AMMI biplotom i to kao udaljenost vrijednosti genotipa od ishodišta značajne genotip × okoliš SVD komponente (Elias i sur., 2016). Osim identifikacije široko prilagođenih genotipova koji postižu visoke vrijednosti promatranih svojstava, AMMI također daje uvid u grupiranje okoliša u homogene megaokoliše (podskup okoliša u kojima se isti, ili približno isti, genotipovi pokazuju superiornima) koji se u daljnjem istraživanju mogu promatrati kao zasebni ciljani raspon okoliša (Gauch, 2013; Gauch i Zobel, 1997).

Dvije glavne prednosti koje čine AMMI model široko korištenim u analizi višeokolišnih pokusa su mogućnost boljeg razumijevanja kompleksne GEI, uključujući i identifikaciju megaokoliša i odabir široko prilagođenih genotipova, i povećanje točnosti kako bi se poboljšale preporuke, ponovljivost i odabir te povećala genetička dobit (Gauch, 2013). U analizi GEI za različita svojstva pšenice, AMMI se koristi još od ranih 90-ih godina 20. stoljeća. Analizirajući podatke CIMMYT-a za prinos 18 genotipova pšenice uzgajanih na 25 lokacija diljem svijeta Crossa i sur. (1991) utvrdili su da je AMMI model iznimno pogodan za raščlanjivanje GEI, utvrđivanje megaokoliša, kao i za identifikaciju genotipova visokog prinosa u određenim okolišima i identifikaciju široko prilagođenih genotipova. U navedenom

istraživanju AMMI1 model se pokazao najpogodnijim za utvrđivanje megaokoliša, dok je AMMI3 model pokazao optimalnu točnost predviđanja. Annicchiarico i Perenzin (1994) su u svom istraživanju utvrdili povezanost AMMI rezultata za prinos pšenice s okolišnim uvjetima, odnosno da IPCA1 os AMMI modela odražava otpornost na mraz, dok IPCA2 os predstavlja otpornost na polijeganje i terminalnu sušu. Iako je najčešće korišten u višeokolišnim analizama komercijalno važnih svojstava poput prinosa pšenice (Groos i sur., 2003; Gauch, 2006; Yang i sur., 2009), AMMI model također se koristi i za procjenu stabilnosti svojstava kakvoće ozime pšenice (reološki parametri dobiveni farinografom, ekstenzografom i miksografom, GPC, WGC, gluten indeks i dr.) (Grausgruber i sur., 2000; Mani, 2007; Drezner i sur., 2010; Mut i sur., 2010). Istraživanje predžetvenog proklijavanja pšenice na 197 genotipova biparentalne populacije pšenice uzgajane u 14 okoliša, pokazalo je da primjena AMMI modela može poboljšati pronalazak i razumijevanje lokusa kvantitativnih svojstava (eng. *quantitative trait loci*; QTL) (Gauch i sur., 2011). Naime, utvrđeno je da AMMI povećava točnost predviđanja svojstva što povećava vjerojatnost pronalaska QTL-a, a grupiranje okoliša otkriva sustavne trendove čiji su uzroci često ekološke ili biološke prirode. AMMI model uspješno je korišten i u analizi GEI za svojstva kakvoće pšenice u RIL populacijama (Groos i sur., 2004). Za svojstva kakvoće ispitana korištenjem RIL populacija pokazalo se da okoliš, te združeni učinak okoliša i GEI, imaju najveći utjecaj na GPC (Krishnappa i sur., 2019). Elangovan i sur. (2011) potvrdili su prevladavajući utjecaj okoliša za svojstva GPC i TW. Ispitujući izvore varijacija za svojstva miksografa Prashant i sur. (2015) utvrdili su sličan obrazac, odnosno značajan doprinos okoliša i GEI fenotipskoj varijanci za svojstva miksografa. S druge strane, neka istraživanja su pokazala da AMMI model nije primjenjiv s istom učinkovitošću na sva svojstva kakvoće pšenice. Bez obzira na visok i značajan učinak GEI, Elangovan i sur. (2008) su pokazali da AMMI model nije uspješan u identificiranju stabilnih genotipova za volumen kruha.

2.5. Molekularni pristupi oplemenjivanju pšenice

Budući da je fenotipizacija često dugotrajan i opsežan posao oplemenjivači se sve više okreću alternativnim metodama oplemenjivanja kako bi se smanjila potreba za fenotipizacijom te ubrzao proces selekcije. Visoki troškovi genotipizacije i sekvenciranja genoma još su u donedavnoj prošlosti ograničavali primjenu molekularnih biljega u oplemenjivačkom procesu. Relativno mali broj biljega ciljane regije genoma korišten je za odabiranje linija na osnovu prisutnosti ili odsutnosti agronomski važnih alela. Smanjenje troškova i razvoj novih metoda genotipizacije (eng. *next-generation sequencing*; NGS) omogućilo je genotipizaciju visoke gustoće te dovelo do sve veće primjene molekularnih biljega u oplemenjivanju (Crossa i sur., 2013). Stoga se danas u oplemenjivanju bilja, pa i pšenice, sve češće koriste metode temeljene na molekularnim biljezima koje daju preciznije procjene te povećavaju učinkovitost oplemenjivanja. Noviji pristupi oplemenjivanju uvelike skraćuju vrijeme potrebno za jedan oplemenjivački ciklus, poboljšavaju preciznost selekcije te omogućuju učinkovitije korištenje genetičke varijance u svrhu poboljšanja genetičke dobiti u oplemenjivačkim programima (Lorenz i sur., 2011). Molekularni biljezi mogu biti korišteni i za analizu poligenskih svojstava, odnosno za QTL analizu.

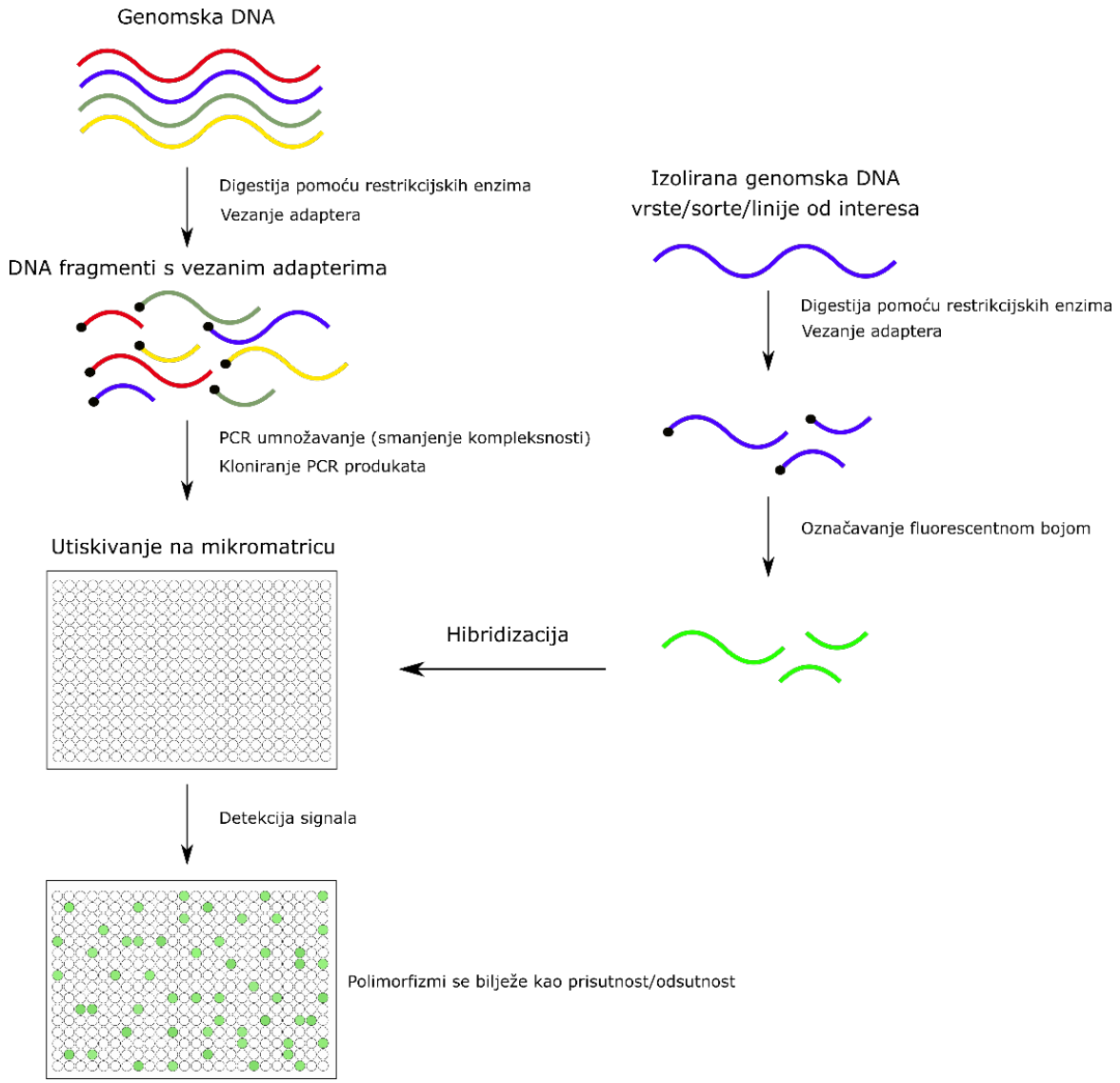
2.5.1. Molekularni biljezi

U biologiji biljaka razlikuju se dva osnovna tipa biljega: morfološki i molekularni biljezi. Morfološki biljezi predstavljaju bilo koje lako uočljivo morfološko svojstvo od kojih su mnogi vidljivi samo na određenim dijelovima biljke i u određenim stadijima razvoja. Broj morfoloških biljega je ograničen, a njihova ekspresija često je nestalna. U molekularne biljege ubrajamo izoenzimske biljege i biljege na razini DNA. Prednost molekularnih nad morfološkim biljezima je ta da nisu pod utjecajem okolišnih uvjeta te da se mogu detektirati u svim fazama razvoja biljke (Mohan i sur., 1997). Izoenzimski biljezi su molekularni oblici istog enzima, ali čija je pojavnost pod kontrolom različitih gena. Oni ne utječu na morfološke ili fiziološke karakteristike jedinke.

Biljezi na razini DNA ono su što najčešće podrazumijevamo pod pojmom molekularnih biljega. To su dijelovi (fragmenti ili sekvence) DNA, duljine jednog ili više parova baza, a koji otkrivaju mjesta varijacije između organizama ili vrsta. Najčešće se nalaze u nekodirajućim regijama DNA tzv. intronima, a pretpostavlja se da nemaju biološki značajnu funkciju. Nazivaju se biljezima jer su često povezani sa specifičnim genima te služe kao pokazatelj njihove prisutnosti u genomu (Collard i sur., 2005). Polimorfizmi takvih biljega unutar genoma su mnogobrojni što omogućuje relativno jednostavno povezivanje njihove prisutnosti s pojavom određenog fenotipa. Tijekom posljednja četiri desetljeća razvijeni su brojni biljezi na razini DNA koji se međusobno razlikuju u složenosti i cijeni

analize, količini informacija koju otkrivaju te broju otkrivenih polimorfizama po reakciji, a uključuju RFLP (eng. *restriction fragment length polymorphism*), RAPD (eng. *random amplified polymorphic DNA*), AFLP (eng. *amplified fragment length polymorphism*), mikrosatelitne odnosno SSR biljege (eng. *simple sequence repeats*) i dr. Zbog sve učestalije potrebe profiliranja cijelog genoma, navedene biljege u upotrebi sve više zamjenjuju biljezi kao što su polimorfizmi jednog nukleotida (eng. *single nucleotide polymorphism,* SNP). SNP biljezi predstavljaju promjene na razini jednog nukleotida na odgovarajućim dijelovima genoma različitih jedinki. Polimorfizam nukleotida na određenom mjestu u genomu smatra se SNP-om samo u slučaju da je frekvencija najmanje frekventnog alela ≥ 1 %. Svaki SNP lokus može imati četiri alela pri čemu svaki alel predstavlja prisutnost jednog od četiri moguća DNA nukleotida. SNP-ovi su mnogobrojni (jedan SNP se pojavljuje na otprilike svakih 100 – 300 parova baza), imaju nisku stopu mutacije te se relativno jednostavno mogu detektirati što ih čini idealnim genetičkim biljezima. Češće se nalaze u nekodirajućim regijama genoma (Singh i Singh, 2015).

DArT (eng. *Diversity Arrays Technology*) tehnologija omogućava istovremeni pronalazak svih promjena prisutnih u DNA lancu (insercije, delecije, SNP) te otkrivanje od nekoliko stotina do nekoliko tisuća polimorfnih biljega (Kilian i sur., 2005). Sama tehnologija zasniva se na hibridizaciji pomoću DArT mikromatrica (Slika 7). DArT analiza sastoji se od dva glavna koraka: (1) konstrukcija mikromatrice (sadrži klonirane prikaze pripremljene pomoću digestije DNA restrikcijskim endonukleazama koristeći genske zalihe genotipova koji predstavljaju genetičku raznolikost vrste), i (2) genotipizacija jedinki od interesa zasnovana na hibridizaciji njihovih genomskih fragmenata i mikromatrice (Singh i Singh, 2015). Mikromatrice se zatim ispiru, skeniraju te se analiziraju prisutni polimorfizmi. Zabilježeni polimorfizmi označavaju prisutnost, odnosno odsutnost hibridizacije na svakom pojedinom elementu mikromatrice. Oni odražavaju varijacije u DNA sekvenci koje otkrivaju koje genomske sekvence su prisutne u genomskom prikazu (Jaccoud i sur., 2001). DArT biljezi su bialelni biljezi koji mogu biti dominantni ili kodominantni. Budući da se DArT tehnologija zasniva na redukciji genoma restrikcijskim endonukleazama, odabrana metoda redukcije odredit će koju vrstu polimorfizama DArT biljezi pretražuju u genomu. Npr. korištenjem restrikcijskih enzima osjetljivih na metilaciju DNA u genomu će se identificirati biljezi koji odražavaju promjene u sekvenci (SNP biljezi) te polimorfizmi metilacije DNA (Kilian i sur., 2005). Za potrebe sekvenciranja genoma pšenice, najčešće se koristi kombinacija dvije restrikcijske endonukleaze – *PstI* (prepoznaje i cijepa sekvencu 5' CTGCA/G 3' – 3' G/ACGTC 5') i *TaqI* (prepoznaje i cijepa sekvencu 5' T/CGA 3' – 3' AGC/T 5') budući da je dokazano da najbolje otkriva polimorfizme unutar genoma pšenice (Singh i Singh, 2015).



Slika 7. Pojednostavljeni prikaz DArT tehnologije.

Za razliku od prvobitne DArT metode zasnovane na fluorescentnom bojanju, DArTseq metoda koristi platforme za sekvenciranje nove generacije pri čemu istovremeno analizira nekoliko desetaka tisuća biljega što u konačnici omogućuje visoku propusnost genotipizacije i rezultira otkrivanjem velikog broja SNP-ova na razini cijelog genoma u vrlo kratkom vremenu (Kilian i sur., 2005).

2.5.2. Biljezima potpomognuta selekcija

MAS koristi veliki broj molekularnih biljega kako bi se utvrdila njihova povezanost s QTL-ima za pojedina svojstva od interesa. Raniji pristupi MAS-u zahtijevali su dugotrajan proces razvoja kartirajućih populacija za utvrđivanje povezanosti biljega s QTL-ima za jednostavna agronomska svojstva. Za korištenje biljega u QTL kartiranju biljeg i QTL moraju međusobno biti u neravnoteži vezanosti gena (eng. *linkage disequilibrium*; LD). Ako biljeg i QTL međusobno nisu u LD-u, što je često slučaj u genetski udaljenijim (*outbreeding*)

populacijama, svi haplotipovi su prisutni u nasumičnim frekvencijama pa biljeg ne može dati pouzdanu informaciju o QTL-u (Dekkers i Hospital, 2002).

Kada se vezanost biljega i QTL-a za promatrano svojstvo potvrdi, MAS omogućuje da se samo na osnovu genotipa predvidi fenotipska vrijednost, odnosno omogućuje selekciju svojstva od interesa na osnovu biljega koji je u LD-u s njim (Mani, 2007). MAS se do danas pokazao kao iznimno uspješna metoda u oplemenjivanju bilja, uspješnija i od fenotipske selekcije. Istraživanja su pokazala da će za svojstva pod kontrolom većeg broja QTL-a MAS biti uspješniji od fenotipske selekcije u svim slučajevima, osim u slučaju kad vrijednost heritabilnosti promatranog svojstva iznosi 1,0 (Dudley, 1993). Uz razvoj jeftinijih metoda genotipizacije (NGS platforme) uključivanje genomskih alata u tradicionalno korištenu fenotipsku selekciju pomaže skratiti trajanje selekcijskog ciklusa, povećati preciznost selekcije te povećati stopu genetičke dobiti u oplemenjivačkim programima (Varshney i sur., 2017). Danas MAS sve više zamjenjuje genomska selekcija koja istovremeno koristi biljege cijelog genoma.

2.6. Genomska selekcija

Genomska selekcija jedan je od novije razvijenih pristupa MAS-u koji omogućuje predviđanje uzgojnih vrijednosti jedinki istovremenim korištenjem molekularnih biljega cijelog genoma. Takav pristup oplemenjivanju prvi put je opisan još 2001. godine (Meuwissen i sur., 2001) te je ponajprije bio korišten u oplemenjivačkim programima životinja, a otada je stekao široku primjenu i među oplemenjivačima bilja (Krishnappa i sur., 2021). Za razliku od ostalih pristupa MAS-u, genomska selekcija ne zahtijeva utvrđivanje biljega povezanih s QTL-ima svojstva od interesa. Tablica 1 navodi osnovne sličnosti i razlike u pristupu oplemenjivanju zasnovanom na MAS-u i na genomskoj selekciji. Genomska selekcija pokušava obuhvatiti ukupnu aditivnu genetičku varijancu na temelju zbroja učinaka velikog broja molekularnih biljega, pritom obuhvaćajući sve QTL-e koji doprinose varijabilnosti svojstva od interesa (Bernardo i Yu, 2007). Budući da istovremeno koristi sve dostupne molekularne biljege osnovna genetička kontrola svojstva od interesa u genomskoj selekciji nije nužno poznata.

U osnovi se genomska selekcija zasniva na korištenju velikog broja molekularnih biljega koji gusto i ravnomjerno prekrivaju genom, tzv. biljega visoke gustoće. Najčešće korišten tip molekularnih biljega u genomskoj selekciji su SNP biljezi čija je dostupnost, brzina analize i financijska isplativost danas daleko veća u odnosu na druge tipove molekularnih biljega. Visoka gustoća biljega korištena u genomskoj selekciji nastoji osigurati da su svi QTL-i svojstva od interesa u LD-u s barem jednim biljegom. Pomoću biljega visoke gustoće moguće je precizno predvidjeti oplemenjivačku vrijednost jedinki za koje ne postoje fenotipski, već samo genotipski podatci (Meuwissen i sur., 2001). Kako bi se izračunale genomske procjene oplemenjivačke vrijednosti (eng. *genomic estimated breeding value*; GEBV) jedinki za koje su poznati samo genotipski podatci, genomska selekcija koristi dva skupa podataka:

(1) trenažnu populaciju (eng. *training population*) tj. kalibracijsku populaciju – referentna populacija za koju su dostupni genotipski i fenotipski podatci, a na osnovu koje se procjenjuju učinci biljega za svojstvo od interesa,

(2) oplemenjivačku populaciju (eng. *breeding population*) tj. validacijsku populaciju (eng. *validation population*) – obuhvaća selekcijske kandidate za koje su poznati genotipski, ali ne i fenotipski podatci, i za koje se na osnovu procijenjenih učinaka biljega korištenjem trenažne populacije izračunavaju GEBV vrijednosti (Sorrells, 2015).

Selekcija se dalje provodi na oplemenjivačkoj populaciji, koja nije fenotipizirana, na osnovu izračunatih GEBV vrijednosti (Slika 8). Sama GEBV vrijednost ne daje informaciju o funkciji QTL-a ili gena za promatrano svojstvo, ali može poslužiti kao odličan kriterij u selekciji (Lorenz i sur., 2011; Voss-Fels i sur., 2019).

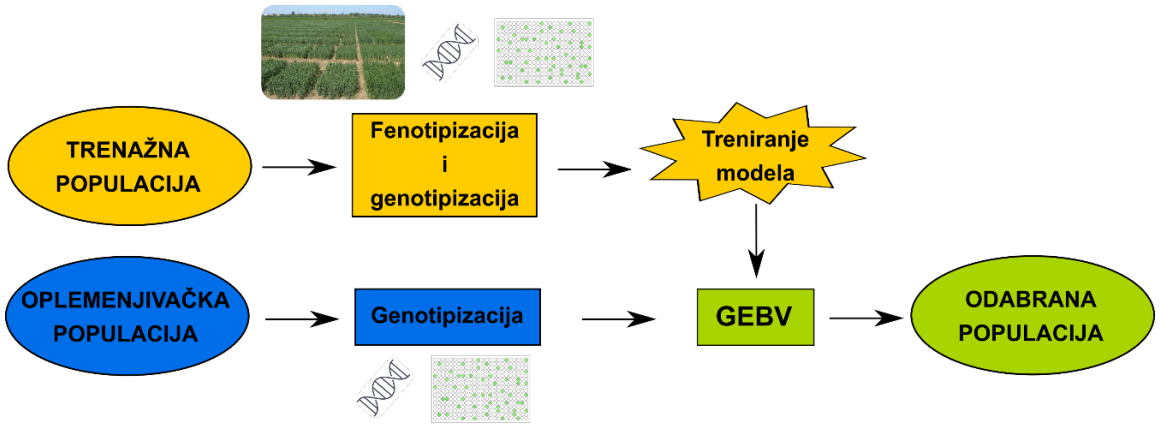
Tablica 1. Osnovne sličnosti i razlike između pristupa oplemenjivanju zasnovanog na biljezima potpomognutoj selekciji i genomskoj selekciji (prema Singh i Singh, 2015).

Biljezima potpomognuta selekcija		**Genomska selekcija**
QTL-i sa značajnim i velikim učincima	**Ciljani QTL-i**	Svi koji utječu na svojstvo
Biljezi	**Osnova selekcije**	GEBV vrijednost procijenjena pomoću biljega
Nekoliko biljega povezanih s QTL-ima svojstva od interesa	**Broj korištenih biljega**	Veliki broj biljega duž cijelog genoma
Potrebno	**Pronalazak, potvrda i provjera QTL-a**	Nije potrebno, procjenjuju se učinci biljega povezani sa svojstvom
Ne mora biti u srodstvu s populacijom na kojoj se provodi MAS	**Populacija korištena za treniranje modela / pronalazak QTL-a**	U srodstvu s populacijom na kojoj se provodi genomska selekcija
Populacija se uglavnom ne održava	**Dugovječnost populacije korištene za treniranje modela / pronalazak QTL-a**	Populacija se održava i redovno ažurira dodavanjem novih linija
Tijekom pronalaska, potvrde i provjere QTL-a	**Fenotipska procjena**	Ograničena na trenažnu populaciju
Introgresija/Akumulacija ciljanih QTL-a	**Cilj oplemenjivačkog programa**	Poboljšanje ciljanih kvantitativnih svojstava
Koristi se zaseban skup podataka o biljezima za svaki QTL	**Selekcija na više svojstava**	Koristi se isti skup podataka o biljezima za sva svojstva

Kako bi se procijenili učinci biljega genomska selekcija koristi različite statističke modele predviđanja. Točnost predviđanja genomske selekcije mjeri se Pearsonovom korelacijom između GEBV vrijednosti i prave (genetičke) oplemenjivačke vrijednosti (eng. *true breeding value*; TBV) jedinki unutar oplemenjivačke populacije. Budući da u praktičnoj primjeni nije moguće unaprijed znati TBV vrijednost selekcijskih kandidata, za izračun korelacije s GEBV koristi se procijenjena oplemenjivačka vrijednost (eng. *estimated breeding value*; EBV) jedinki kako bi se utvrdila uspješnost modela predviđanja i samim time i uspješnost genomske selekcije (Heffner i sur., 2009; Ward i sur., 2019).

Prije uključivanja genomske selekcije u oplemenjivački program potrebno je procijeniti njezinu potencijalnu uspješnost za svojstvo od interesa unutar ciljane populacije korištenjem unakrsne validacije (eng. *cross-validation*). U postupku unakrsne validacije

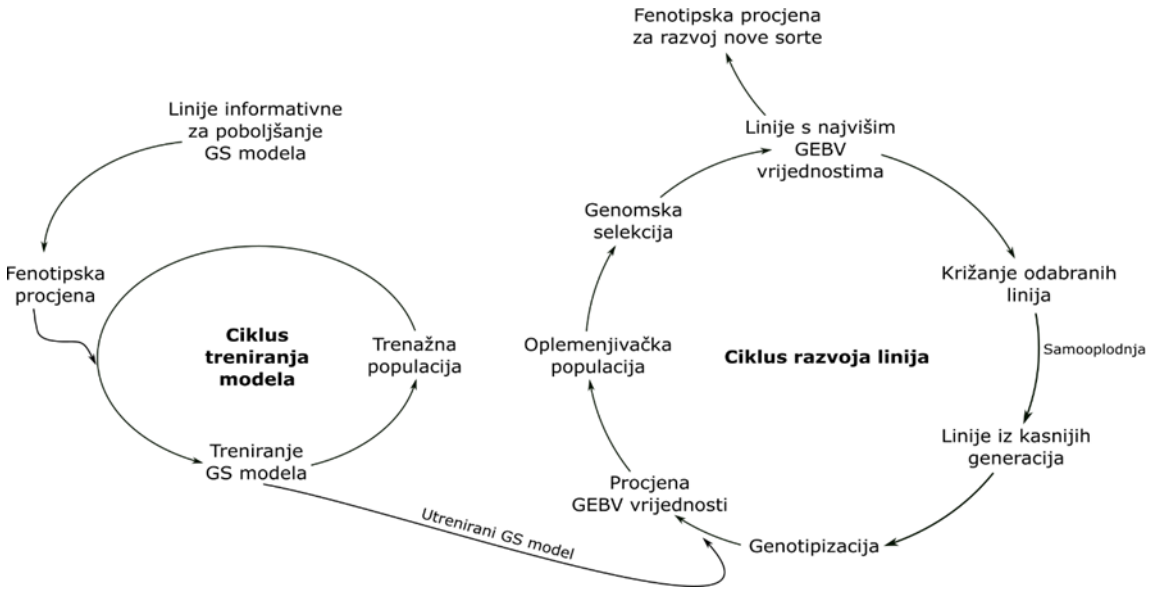
populacija za koju su dostupni fenotipski i genotipski podatci podijeli se na dva skupa podataka: trenažni skup podataka i validacijski skup podataka. Razdioba na ova dva skupa podataka može biti nasumična ili slijediti neki od algoritama za optimizirano uzorkovanje. Za odabir trenažne populacije najčešće se koristi nasumično uzorkovanje, a može se koristiti i stratificirano uzorkovanje, uzorkovanje na temelju srednje vrijednosti koeficijenta determinacije (eng. *coefficient of determination*; CD), srednje vrijednosti varijance pogreške predviđanja (eng. *predictor error variance*; PEV), i dr. (Isidro i sur., 2015; Marulanda i sur., 2015; Rincent i sur., 2012). Odabir same strategije optimizacije trenažne populacije ovisi ponajprije o strukturi populacije. Prema rezultatima istraživanja Isidro i sur. (2015) za strukturirane populacije poželjno je za trenažnu populaciju odabrati jedinke koje daju najveću fenotipsku varijancu kako bi se postigla visoka točnost predviđanja.



Slika 8. Dijagram procesa genomske selekcije (prema Plavšin i sur., 2021).

Najvažnija prednost genomske selekcije u odnosu na tradicionalno korištene metode oplemenjivanja je povećanje genetičke dobiti uslijed skraćivanja selekcijskog ciklusa u oplemenjivačkom procesu (Heffner i sur., 2010; Sorrells, 2015; Voss-Fels i sur., 2019). Genomska selekcija poboljšava točnost selekcije i omogućuje odabir uspješnih linija ranije u oplemenjivačkom ciklusu, čime se smanjuje potencijalni trošak fenotipizacije u kasnijim generacijama (Belamkar i sur., 2018; Lorenz i sur., 2011). Prema istraživanju Heffner i sur. (2010) za svaki provedeni ciklus MAS-a moguće je provesti 2.33 oplemenjivačka ciklusa zasnovana na genomskoj selekciji kada je riječ o oplemenjivanju na agronomski važna svojstva ozime pšenice. Također, ako se u oplemenjivačkom ciklusu zasnovanom na genomskoj selekciji postigne točnost predviđanja od 0,5, moguće je postići genetičku dobit koja na godišnjoj razini za dva puta premašuje dobit postignutu korištenjem MAS-a.

Prema Heffner i sur. (2009) strategija koja bi omogućila učinkovito iskorištenje i implementaciju genomske selekcije u oplemenjivačke programe bilja sastoji se od dva ciklusa: ciklusa treniranja modela i ciklusa razvoja linija, a koji se međusobno nadopunjuju (Slika 9). Ciklus treniranja modela ima za cilj kontinuirano poboljšanje točnosti predviđanja, a što se postiže uključivanjem informativnih linija u trenažnu populaciju. Fenotipski i genotipski podatci informativnih linija koriste se za poboljšanje predviđanja modela. Linije informativne za poboljšanje modela dobivene su kao rezultat ciklusa razvoja linija u kojem se utrenirani model koristi za procjenu GEBV vrijednosti na osnovu kojih se odabiru najuspješnije linije. Najuspješnije linije mogu se koristiti kao osnova za razvoj nove sorte ili se iz njih mogu dalje razvijati cijepajuće generacije koje će se genotipizirati i nastavljati ciklus razvoja linija. Uspješnost ovakvih strategija ovisi ponajprije o optimizaciji parametara za provođenje genomske selekcije na određenoj kulturi i za određeno svojstvo.



Slika 9. Shematski prikaz strategije uključivanja genomske selekcije u oplemenjivački program bilja (prema Heffner i sur., 2009).

Budući da genomska selekcija uzima u obzir sve dostupne biljege bez njihove prethodne selekcije pokazalo se da je posebno korisna za predviđanje poligenskih svojstava čija je ekspresija pod kontrolom velikog broja QTL-a s malim učinkom, kao što su svojstva kakvoće pšenice (Kristensen i sur., 2018). Unatoč važnosti svojstava kakvoće pšenice u kontekstu prehrane, istraživanja genomske selekcije za svojstva kakvoće pšenice još uvijek nisu tako brojna kao što je to slučaj za agronomski važna svojstva, primjerice prinos. Pregledni znanstveni rad pod naslovom „An Overview of Key Factors Affecting Genomic Selection for Wheat Quality Traits“ (Plavšin i sur., 2021), koji je uključen u ovaj

doktorski rad, daje detaljan pregled dosadašnjih istraživanja genomske selekcije za svojstva kakvoće pšenice (Prilog 3).

2.6.1. Čimbenici koji utječu na točnost predviđanja genomske selekcije

Prvi korak ka uspješnoj implementaciji genomske selekcije u praktični oplemenjivački program je ispravna prilagodba svih čimbenika koji mogu utjecati na točnost predviđanja. Uspješnost oplemenjivačkog programa zasnovanog na genomskoj selekciji ovisit će o dobivenoj točnosti predviđanja, a koja je pod utjecajem različitih genetičkih čimbenika (raspodjela i jačina LD-a između biljega i QTL-a, kolinearnost između biljega, razlike u frekvenciji alela između trenažne i oplemenjivačke populacije, struktura i srodnost trenažne i oplemenjivačke populacije) (Edwards i sur., 2019; Habier i sur., 2007; Isidro i sur., 2015) i fenotipskih čimbenika vezanih uz promatrano svojstvo (heritabilnost, fenotipska varijanca trenažne populacije) (Heffner i sur., 2009; Ornella i sur., 2012). Osim navedenog, značajan utjecaj na točnost predviđanja imaju i veličina trenažne populacije korištena za procjenu učinaka biljega, gustoća i vrsta biljega, ali i karakteristike odabranog statističkog modela za predviđanje (Heslot i sur., 2012; Poland i sur., 2012).

Prema rezultatima dosadašnjih istraživanja najvažniju ulogu u postizanju visoke točnosti predviđanja ima međusobni odnos triju čimbenika: strukture promatranih populacija, veličine trenažne populacije i gustoće biljega (Riedelsheimer i sur., 2013; Riedelsheimer i Melchinger, 2013; Robertsen i sur., 2019). Prilikom dizajniranja trenažne populacije u obzir se treba uzeti i oplemenjivačka populacija, odnosno trenažna populacija mora biti dizajnirana tako da se postigne željeni ishod (visoka točnost predviđanja) u oplemenjivačkoj populaciji (Crossa i sur., 2016; Jannink i sur., 2010). Da bi se postigla zadovoljavajuća točnost predviđanja trenažna populacija bi trebala biti visokosrodna s oplemenjivačkom populacijom ili sadržavati genotipove srodne genotipovima prisutnima u oplemenjivačkoj populaciji (Asoro i sur., 2011; Hickey i sur., 2014). Brojna su istraživanja, jednako u biljnim i životinjskim oplemenjivačkim programima, pokazala da je točnost predviđanja genomske selekcije značajno smanjena u slučaju da trenažna i oplemenjivačka populacija nisu srodne (Windhausen i sur., 2012; Ly i sur., 2013; Albrecht i sur., 2014; Lorenz i Smith, 2015; Rutkoski i sur., 2015). Točnost predviđanja može rezultirati i nultim vrijednostima ili vrijednostima vrlo blizu nuli, ako se prilikom odabiranja trenažne i oplemenjivačke populacije koriste isključivo međusobno nesrodne jedinke (Riedelsheimer i sur., 2013). Točnost predviđanja raste s porastom veličine trenažne populacije, kao i s porastom gustoće biljega sve dok ne dosegne tzv. plato predviđanja nakon kojeg daljnje povećanje nema značajan utjecaj na točnost predviđanja (Arruda i sur., 2015; Heffner i sur., 2011a; Maulana i sur., 2019). Što su trenažna i oplemenjivačka populacija genetski srodnije,

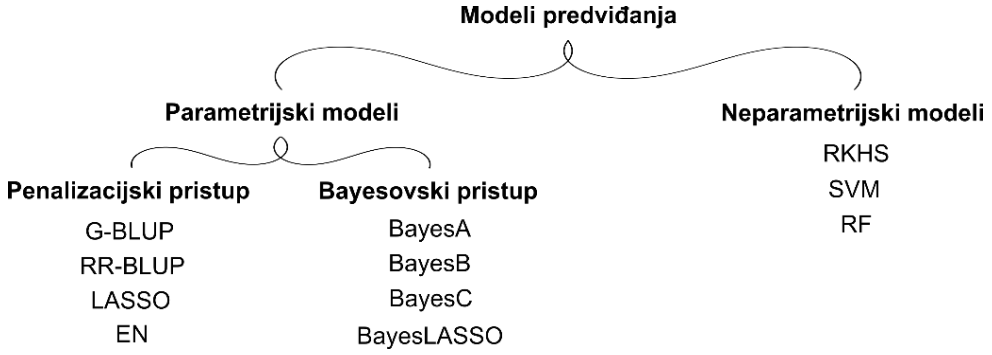
to su veličina trenažne populacije i gustoća biljega potrebni za postizanje platoa točnosti predviđanja manji (Herter i sur., 2019; Rutkoski i sur., 2015a). Također, veličina trenažne populacije dostatna za postizanje zadovoljavajuće vrijednosti točnosti predviđanja manja je u slučaju samooplodnih u odnosu na stranooplodne biljne vrste (Singh i Singh, 2015). Osim toga, stopa LD-a utječe na gustoću biljega potrebnu za postizanje određene vrijednosti točnosti predviđanja (Sorrells, 2015). Visoka stopa LD-a između QTL-a i biljega prisutna u visokosrodnim populacijama, kao što su biparentalne populacije, osigurava da je svaki QTL obuhvaćen s više biljega. Mali broj genskih rekombinacija prisutan u biparentalnim populacijama biljaka koje nisu nasumično sparivane uzrokuje postojanje velikih tzv. vezanih blokova (eng. *linkage blocks*) što ograničava lokalizaciju QTL-a na relativno velike intervale od 10 do 20 cM te uzrokuje visoku stopu LD-a između biljega i QTL-a. Stoga je niža gustoća biljega potrebna kako bi se dosegao plato točnosti predviđanja kada se genomska selekcija primjenjuje u biparentalnim populacijama u odnosu na nesrodne populacije (Lorenzana i Bernardo, 2009).

Osim o navedenim čimbenicima, točnost predviđanja također značajno ovisi o heritabilnosti svojstva. Što je heritabilnost promatranog svojstva manja, to se manja točnost predviđanja može očekivati za to svojstvo (Bernardo i Yu, 2007; Combs i Bernardo, 2013; Zhang i sur., 2015). U slučaju niske heritabilnosti svojstva, za postizanje relativno visoke točnosti predviđanja potrebno je značajno povećati broj fenotipskih opažanja (Meuwissen i sur., 2001; Sorrells, 2015). Istraživanja su također pokazala da na točnost predviđanja u određenoj mjeri utječe i postojanje GEI budući da može dovesti do promjene poretka uspješnosti genotipova (Heslot i sur., 2014). Stoga neka istraživanja predlažu uključivanje GEI u model za predviđanje (Jarquín i sur., 2017; Lado i sur., 2016; Lopez-Cruz i sur., 2015) ili uključivanje informacija iz koreliranih okoliša kako bi se poboljšala točnost predviđanja (Burgueño i sur., 2012; Ornella i sur., 2012). Relativno mali broj okoliša potreban je za postizanje visoke točnosti predviđanja (Riedelsheimer i Melchinger, 2013), a postojanje replikacija unutar okoliša dodatno pospješuje točnost predviđanja (Endelman i sur., 2014).

Brojni znanstveni radovi daju opsežne preglede čimbenika koji utječu na točnost predviđanja genomske selekcije za različite biljne vrste i različita svojstva (Combs i Bernardo, 2013; Krishnappa i sur., 2021; Wang i sur., 2018), a detaljan pregled utjecaja različitih čimbenika na točnost predviđanja za svojstva kakvoće pšenice, kao i vrijednosti heritabilnosti zabilježene u dosadašnjim istraživanjima, prikazani su u znanstvenom radu pod naslovom „An Overview of Key Factors Affecting Genomic Selection for Wheat Quality Traits“ (Plavšin i sur., 2021), koji je uključen u ovaj doktorski rad (Prilog 3).

2.6.2. Modeli predviđanja korišteni u genomskoj selekciji

Do danas su razvijeni različiti modeli predviđanja potrebni za rješavanje problema visokodimenzioniranih skupova podataka u genomskoj selekciji koji proizlaze iz velike količine podataka prikupljenih genotipizacijom visoke propusnosti. Kako bi se smanjio problem koji nastaje zbog prevelikog broja varijabli predviđanja (broj biljega) u odnosu na broj opažanja (fenotipski podatci) u genomskoj selekciji se koriste modeli zasnovani na različitim metodama sažimanja (de los Campos i sur., 2013; Heffner i sur., 2009; Merrick i Carter, 2021). Klasifikacija najčešće korištenih modela prikazana je na Slika 10. Osnovna razlika između parametrijskih i neparametrijskih modela je u pretpostavci veze između fenotipskih i genotipskih vrijednosti pri čemu parametrijski modeli pretpostavljaju linearnost te veze i u obzir uzimaju samo aditivne učinke gena, dok neparametrijski modeli pretpostavljaju da je veza nelinearna te tako omogućuju da se u obzir uzmu i neaditivni učinci gena (Desta i Ortiz, 2014). Modeli predviđanja korišteni u genomskoj selekciji uglavnom se razlikuju u pretpostavkama distribucije i varijance učinaka biljega (Liu i sur., 2018; Meuwissen i sur., 2001). S obzirom da se modeli razlikuju u pretpostavkama kako učinci biljega doprinose ukupnoj varijanci promatranog svojstva, uspješnost predviđanja nekog modela ovisit će ponajprije o genetičkoj arhitekturi tog svojstva (Larkin i sur., 2019).



Slika 10. Klasifikacija najčešće korištenih modela predviđanja u genomskoj selekciji.

BLUP modeli parametrijski su modeli koji se zasnivaju na predviđanju po principu najboljih linearnih nepristranih predviđanja (eng. *best linear unbiased prediction*; BLUP), a u koje se ubrajaju G-BLUP (eng. *genomic BLUP*) i RR-BLUP (eng. *ridge regression BLUP*). Zbog svoje robusnosti i pouzdanosti rezultata koje pruža RR-BLUP je najčešće korišteni model u genomskoj selekciji. U matričnom zapisu RR-BLUP model može se prikazati sljedećom jednadžbom:

$$y = WGu + e$$

gdje je y vektor fenotipskih vrijednosti, W je matrica dizajna, G je matrica genotipova, u je vektor učinaka biljega koji prate normalnu distribuciju i imaju zajedničku varijancu

($u \sim N(0, I\sigma_u^2)$, a e je pogreška ostatka. BLUP rješenje za izračun učinaka biljega može se zapisati kao:

$$u = (Z^T Z + \lambda I)^{-1} Z^T y$$

gdje je λ parametar hrbatne regresije (eng. *ridge regression*; RR), I je matrica identiteta, a $Z = WG$. Vrijednost parametra λ izračunava se kao omjer varijance pogreške i varijance biljega (σ_e^2/σ_u^2), a ista vrijednost penalizacije primjenjuje se na sve učinke biljega što uzrokuje njihovo jednako sažimanje prema nuli, bez obzira na veličinu učinka pojedinog biljega (Endelman, 2011). Glavna pretpostavka ovog modela je da svi biljezi doprinose svojstvu što često nije realistična pretpostavka. U genomskoj selekciji pšenice RR-BLUP je korišten za predviđanje važnih agronomskih svojstava kao što su prinos zrna (He i sur., 2016; Heslot i sur., 2012; Isidro i sur., 2015), otpornost na fuzarijsku palež klasa (Arruda i sur., 2015; Rutkoski i sur., 2012), otpornost na crnu žitnu hrđu (Ornella i sur., 2012; Rutkoski i sur., 2014), svojstva kakvoće (Heffner i sur., 2011a; Ward i sur., 2015) i dr.

Za razliku od RR-BLUP-a, G-BLUP model uz pomoć biljega procjenjuje matricu odnosa između jedinki u trenažnoj i oplemenjivačkoj populaciji, a koja se kasnije koristi za procjenu matrice varijanci-kovarijanci genetičkih vrijednosti (Tan i sur., 2017; Zhong i sur., 2009). Međutim, istraživanja su pokazala da su navedena dva modela istovjetna pod uvjetima koji se obično susreću u praktičnoj primjeni (Habier i sur., 2007; Hayes i sur., 2009).

Osim BLUP modela, u parametrijske modele s penalizacijskim pristupom ubrajaju se i operator najmanjeg apsolutnog skupljanja i odabira (eng. *least absolute shrinkage and selection operator*; LASSO) i elastična mreža (eng. *elastic net*; EN). Osnova LASSO modela je istovremena primjena metode odabira varijabli (iz para usko povezanih biljega izabire jedan, a ostale zanemaruje) i metode sažimanja regresijskih koeficijenata biljega kako bi se procijenili njihovi učinci (Wang i sur., 2018). EN model primjenjuje ponderiranu kombinaciju penalizacije metodama RR i LASSO pri čemu parametar P_α određuje kojoj od dvije metode će analiza biti sličnija. Što je vrijednost α niža, to je EN model sličniji RR-u (za koji je α = 0), dok je EN model s vrijednostima α bližima 1 sličniji LASSO analizi (za koju je α = 1). Prema tome, EN istovremeno odabire skupine koreliranih biljega i kontinuirano sažima učinke biljega (Friedman i sur., 2010; Zou i Hastie, 2005).

Bayesovski modeli svoju osnovu nalaze u Bayesovom teoremu, a za razliku od RR-BLUP modela koji sažima sve učinke biljega jednako (što može dovesti do prevelikog sažimanja vrlo velikih učinaka), osmišljeni su kako bi se preciznije modelirali učinci biljega različitih veličina (Heffner, 2011). Svi bayesovski modeli mogu se opisati pomoću sljedeće jednadžbe:

$$y = \mu + \sum_{k=1}^{m} x_k \beta_k + e$$

gdje je y vektor fenotipskih vrijednosti, μ je srednja vrijednosti, x_k je vektor genotipova za biljeg k, β_k je učinak biljega k, m je broj biljega, a e je vektor pogrešaka koji slijedi distribuciju $e \sim N(0, I\sigma_e^2)$. Bayesovski modeli međusobno se razlikuju u prethodnim pretpostavkama distribucije učinaka biljega (β_k), odnosno svaki model pretpostavlja drugačiju prethodnu distribuciju za procjenu učinaka biljega. BayesA i BayesB modeli pretpostavljaju da β_k slijedi obrnutu hi-kvadrat distribuciju, a vrijednost parametra π određuje vjerojatnost da je učinak biljega k jednak nuli. Za BayesA model vrijedi da je π = 0, odnosno model pretpostavlja da svi biljezi imaju određeni učinak na promatrano svojstvo (Meuwissen i sur., 2001). BayesB model primjenjuje distribuciju s glavninom mase na nuli što omogućava da mnogi biljezi imaju učinak jednak nuli (Habier i sur., 2011; Meuwissen i sur., 2001). Sa stajališta oplemenjivača BayesB predstavlja realističniju pretpostavku s obzirom da postoje regije genoma koje nisu u vezi s QTL-om te stoga nemaju svi biljezi učinak na svojstvo od interesa. Pretpostavka BayesC modela je da jedan dio biljega (1 - π razmjer) ima učinak jednak nuli, dok učinak drugog dijela biljega (π razmjer) prati normalnu distribuciju (Hu i sur., 2019). BayesLASSO model predstavlja oblik LASSO regresije (Park i Casella, 2008). Za procjenu učinaka biljega ovaj model primjenjuje dvostruku eksponencijalnu distribuciju pri čemu dodjeljuje jedinstvenu varijancu svim biljezima. Na taj način uzrokuje jače sažimanje regresijskih koeficijenata bliže nuli (biljezi s malim učinkom) i slabije sažimanje regresijskih koeficijenata čija je apsolutna vrijednost visoka (biljezi s velikim učinkom) (Heslot i sur., 2012). Navedeni bayesovski modeli pokazali su se boljima u odnosu na druge pristupe u slučajevima kada postoji velika gustoća biljega, a relativno mali broj fenotipskih podataka (Singh i Singh, 2015). U istraživanjima genomske selekcije na pšenici navedeni su modeli, s većom ili manjom točnošću predviđanja, korišteni za predviđanje važnih agronomskih svojstava kao što su prinos zrna (He i sur., 2016; Heffner i sur., 2011a, 2011b; Storlie i Charmet, 2013), TKW (Heslot i sur., 2012), otpornost na hrđu (Ornella i sur., 2012; Rutkoski i sur., 2014), otpornost na fuzarijsku palež klasa (Rutkoski i sur., 2012), visina (Zhao i sur., 2014) i dr. U simulacijskim istraživanjima BayesB model pokazao se uspješnijim u odnosu na RR-BLUP model pod pretpostavkom aditivnog učinka gena (Habier i sur., 2007), dok su istraživanja na ječmu pokazala upravo suprotno (Zhong i sur., 2009). BayesB može predstavljati bolji izbor od RR-BLUP modela jedino u slučaju kada su biljezi snažno povezani s QTL-ima, a što je očekivano kada QTL-i imaju veliki učinak na promatrano svojstvo (Jannink i sur., 2010).

U genomskoj selekciji sve se češće koriste i različite neparametrijske metode zasnovane na strojnom učenju kao što su metoda potpornih vektora (eng. *support vector machine*; SVM) i metoda slučajne šume (eng. *random forest*; RF), a koje su se pokazale učinkovitima u otkrivanju interakcija između biljega (Desta i Ortiz, 2014; Jannink i sur., 2010). Model RF zasniva se na konstrukciji velikog broja stabala odluke koji su jednako distribuirani. Za svako pojedinačno stablo odluke izračunava se zasebno predviđanje korištenjem regresijskih modela, a srednja vrijednost izračuna za sva stabla odluke predstavlja konačnu vrijednosti predviđanja. *Bootstrap* metoda koristi se za pronalaženje optimalnog podskupa podataka koji će služiti za konstruiranje stabala odluke, a dijeljenje skupa podataka na svakom čvoru stabla odluke odvija se tako da se sa svakim sljedećim uzorkom smanji funkcija gubitaka (Breiman, 2001). RF model se može prikazati sljedećom jednadžbom:

$$\hat{y}_i = \frac{1}{B}\sum_{b=1}^{B} T_b(x_i)$$

gdje je $\hat{y}_i$ fenotipska procjena genotipa x_i, T je broj stabala odluke, a B je broj *bootstrap* uzoraka. Međutim, kao što je slučaj i s drugim korištenim modelima, istraživanja su pokazala da uspješnost modela strojnog učenja ovisi o promatranom svojstvu. Dok su neka istraživanja pokazala da pristupi temeljeni na metodama strojnog učenja nadmašuju konvencionalne modele (npr. RR-BLUP) za predviđanje svojstava kakvoće pšenice (Sandhu i sur., 2021a, 2021b), druga istraživanja nisu utvrdila značajno poboljšanje točnosti predviđanja korištenjem metoda strojnog učenja (Battenfield i sur., 2016). Osim navedenih, genomska selekcija koristi i neke druge modele previđanja kao što su jezgrene metode, npr. Hilbertovi prostori reproducirajućih jezgri (eng. *reproducing kernel Hilbert spaces*; RKHS). RKHS model provodi poluparametrijsku regresiju na genotipske podatke, a za kontrolu distribucije učinaka biljega koristi genetsku udaljenost i jezgrenu funkciju zasnovanu na Euklidskoj mjeri sličnosti biljega (De Los Campos i sur., 2010; Gianola i Van Kaam, 2008).

3. REZULTATI I RASPRAVA

3.1. Pregled objavljenih kvalifikacijskih znanstvenih radova

Prošireni sažetak: Znanstveni rad 1.

Naslov znanstvenog rada na hrvatskom jeziku: Interakcija genotip × okoliš za svojstva kakvoće u biparentalnim populacijama pšenice

Postojanje GEI često predstavlja izazov za oplemenjivače bilja jer otežava odabir stabilnih ili superiornih genotipova. Kako bi se smanjile prepreke uzrokovane postojanjem GEI i postigla učinkovitija selekcija za svojstva kakvoće pšenice, važno je pravilno procijeniti učinke genotipa, okoliša i GEI na svojstvo od interesa. U ovom radu, GEI za svojstva kakvoće proučavana je pomoću AMMI modela. Ciljevi istraživanja u sklopu ovog znanstvenog rada bili su istražiti GEI i pojavu transgresije za svojstva GPC, WGC, TW i četiri reološka svojstva dobivena miksograf uređajem u dvije biparentalne populacije pšenice.

U istraživanju su korištene dvije biparentalne (RIL) populacije pšenice dobivene iz križanja roditeljskih sorti Bezostaya-1 × Klara (BK) i Monika × Golubica (MG). Roditeljske sorte korištene u BK populaciji razlikovale su se u svim HMW gluteninskim podjedinicama, dok se roditeljske sorte u MG populaciji nisu razlikovale niti u jednoj od HMW gluteninskih podjedinica (Horvat i sur., 2006). Nakon križanja i samooplodnje biljke su nasumično odabirane sve do F7 generacije koja je služila za provođenje poljskih pokusa, a koja se sastojala od 145 (BK) i 175 (MG) genotipova, uključujući i roditeljske genotipove. Poljski pokusi su provedeni na dvije lokacije u Hrvatskoj (Osijek i Slavonski Brod) tijekom tri godine (2009. – 2011.), odnosno u 6 različitih okoliša koji su označeni kombinacijom lokacije i godine (OS09 – OS11, SB09 – SB11). U svakom okolišu pokus je postavljen prema redno-stupčanom dizajnu. GPC i TW su određeni na bazi cijelog zrna korištenjem NIR spektroskopije, a WGC je određen prema standardiziranoj ICC metodi br. 155. Reološka svojstva tijesta određena su miksograf uređajem. Za daljnju analizu odabrane su četiri varijable za koje je utvrđena najbolja ponovljivost, a koje daju sveobuhvatan uvid u razvoj tijesta: MPT, MTW, MTI i MPH (Prashant i sur., 2015). U prvom dijelu statističke analize skupni podatci iz pojedinačnih pokusa analizirani su prema mješovitom modelu koji je uključivao fiksne učinke genotipa, okoliša i repeticija unutar okoliša, te nasumične učinke redova i stupaca unutar repeticije unutar okoliša. Tako procijenjene vrijednosti za sve kombinacije genotip-okoliš korištene su za izračun korelacija te za analizu GEI korištenjem AMMI modela. Imputacija nedostajućih podataka provedena je korištenjem EM-AMMI (eng. *expectation maximization AMMI*) algoritma (Paderewski, 2013). Za procjenu značajnosti i

odabir broja IPCA osi koje će biti zadržane u konačnom modelu korištena su tri različita pristupa: jednostavna parametrijska *bootstrap* metoda (eng. *simple parametric bootstrap;* SPB), F_R-test i pojedinačna unakrsna validacija (eng. *leave-one-out cross-validation;* LOO-CV) (Forkman i Piepho, 2014; Paderewski, 2013; Piepho, 1995). Rezultati AMMI analize su prikazani korištenjem modificiranog AMMI2 biplota koji je osim IPCA1 i IPCA2 osi uključivao i srednje vrijednosti svojstava prikazane skalom u boji.

Rezultati analize transgresivne segregacije pokazali su da u obje kombinacije križanja postoje pozitivne i negativne segregirajuće linije. Raspon vrijednosti promatranih svojstava BK populacije bio je puno širi od raspona vrijednosti zabilježenih za roditeljske sorte, a za sva ispitivana svojstva pronađene su i pozitivne (38,46 – 77,62 %) i negativne (5,59 – 46,85 %) transgresivne segregirajuće linije. Srednje vrijednosti za BK populaciju bile su slične srednjim vrijednostima roditelja za sva svojstva, osim za svojstva miksografa MPT i MTW za koja su srednje vrijednosti bile veće od oba roditeljska genotipa. Srednje vrijednosti MG populacije bile su unutar raspona vrijednosti roditeljskih genotipova za sva svojstva, a što je rezultat šireg raspona vrijednosti genotipova Monika i Golubica u odnosu na genotipove Bezostaya-1 i Klara. Posljedično, niži omjer pozitivnih i negativnih segregirajućih linija opažen je u MG populaciji u odnosu na BK populaciju. Vrlo visoke pozitivne korelacije između svojstava GPC i WGC ($r \geq 0,89$) i svojstava MTI i MPH ($r \geq 0,91$), te slabe negativne korelacije između svojstava GPC/WGC i TW utvrđene su u obje promatrane populacije. TW je pokazao negativne, ali niske korelacije sa sva četiri svojstva miksografa u BK populaciji, dok su u MG populaciji navedene korelacije bile uglavnom pozitivne. Pozitivna korelacija između svojstava GPC i WGC i miksograf svojstava MTI i MPH utvrđena je za obje populacije, iako su korelacije bile puno više u BK u odnosu na MG populaciju. Nasuprot tome, druga dva svojstva miksografa (MPT i MTW) pokazala su različite trendove u pogledu korelacije s GPC/WGC u dvjema populacijama kao i u različitim okolišima. Za neke parove svojstava uočene su značajne varijacije u pogledu jačine i smjera korelacije, a što može ukazivati na snažan utjecaj okoliša. Usporedba različitih pristupa za procjenu značajnosti i odabir broja IPCA osi koje će biti zadržane u konačnom modelu pokazala je znatno neslaganje u pogledu dobivenih rezultata. Za polovicu slučajeva mogućih kombinacija populacija-svojstvo SPB i F_R-test su odabrali isti AMMI model, dok su za drugu polovicu slučajeva ova dva pristupa odabrala modele koji se znatno razlikuju u pogledu broja osi koje je potrebno zadržati. Općenito, F_R-test se pokazao kao najliberalniji pristup, dok je LOO-CV bio najkonzervativniji pristup koji je kao optimalan model predložio aditivni model (AMMI0) za sva svojstva osim za MTI u BK populaciji i MTW u MG populaciji. GEI analiza korištenjem AMMI modela pokazala je neke zajedničke obrasce za dvije promatrane populacije. Za GPC, WGC i TW svojstva dominantan izvor fenotipske varijacije

bio je okoliš, koji je činio oko 75 % ukupne varijacije u BK, te 40 – 85 % varijacije u MG populaciji. Također, za navedena svojstva učinci genotipa i GEI bili su podjednaki, s time da je u BK populaciji učinak genotipa bio dvostruko veći u odnosu na učinak GEI. U MG populaciji je učinak genotipa bio otprilike 1,5 puta veći od učinka GEI za GPC i WGC, dok je za TW učinak GEI bio veći od učinka genotipa. Na MPT i MTW dominantan utjecaj imala je GEI u BK i genotip u MG populaciji, dok je najslabiji učinak na ova svojstva imao okoliš, osim u slučaju MTW svojstva u MG populaciji gdje je učinak okoliša bio nešto veći od učinka GEI. Na MTI i MPH dominantan učinak imao je okoliš u BK i GEI u MG populaciji, dok je učinak genotipa uvijek bio na drugom mjestu. Općenito, utjecaj GEI imao je važniju ulogu za svojstva miksografa u odnosu na ostala promatrana svojstva. Uzimajući u obzir rezultate dobivene AMMI2 modelom, IPCA1 je doprinijela GEI u udjelu od 30 do 70 %, dok je zajednički doprinos IPCA1 i IPCA2 osi iznosio između 54 i 86 %. Najveći doprinos IPCA1 ili IPCA1+IPCA2 osi zabilježen je za svojstvo TW. Analizom AMMI2 biplota utvrđeni su neki široko prilagođeni RIL-ovi, među kojima je najzanimljiviji RIL MG124 koji se pokazao najuspješnijim genotipom za svojstva GPC i WGC u većini promatranih okoliša, ali i najuspješnijim genotipom za nekorelirano svojstvo TW u okolišu SB10.

Dvije promatrane biparentalne populacije pokazale su različitosti u pogledu prilagodljivosti, a što se može pripisati korištenju široko prilagođene sorte Bezostaya-1 kao roditeljskog genotipa u jednom križanju (Bezostaya-1 × Klara), dok su u drugom križanju korištene dvije sorte (Monika × Golubica) porijeklom iz istog uzgojnog programa, a koje nikada nisu uzgajane u regiji porijekla u velikim razmjerima. Nadalje, transgresivne segregirajuće linije bile su mnogo zastupljenije u BK populaciji, a što osigurava širu osnovu za odabir RIL-ova sa širokom ili specifičnom prilagodbom u ranijim generacijama. AMMI analiza se pokazala kao dobar alat za identifikaciju potencijalno zanimljivih RIL-ova jednostavnom analizom biplota, a odabrani RIL-ovi pokazuju potencijal za korištenje u oplemenjivačkim programima usmjerenima na poboljšanje kakvoće pšenice.

Ključne riječi: kakvoća pšenice, miksograf, biparentalna populacija, GEI, AMMI, EM-AMMI

Prošireni sažetak: Znanstveni rad 2.

Naslov znanstvenog rada na hrvatskom jeziku: Pregled ključnih čimbenika koji utječu na genomsku selekciju za svojstva kakvoće pšenice

Selekcija za svojstva kakvoće pšenice korištenjem klasičnih metoda često može biti dugotrajna i financijski zahtjevna jer uvjetuje iscrpnu fenotipizaciju u završnim fazama razvoja novih linija i sorti. Razvojem visokopropusne genotipizacije tijekom posljednjeg desetljeća došlo je do široke primjene metoda za pouzdano i brzo predviđanje uzgojnih vrijednosti samo na temelju genotipskih podataka. Jedna od takvih metoda je i genomska selekcija koja omogućuje predviđanje uzgojnih vrijednosti jedinki istovremenim korištenjem svih dostupnih biljega za treniranje modela predviđanja. Uspješnost genomske selekcije ovisi o dobivenoj točnosti predviđanja, a na koju utječu različiti molekularni, genetički i fenotipski čimbenici, kao i čimbenici odabranog statističkog modela. Većina istraživanja genomske selekcije na pšenici odnosi se na istraživanja vezana uz prinos i otpornost na bolesti koji se smatraju ključnim svojstvima za uspješnu proizvodnju pšenice. Takav snažan naglasak na prinos pšenice razumljiv je sa stajališta rastuće potrebe za pšenicom kao izvorom hrane. Međutim, s obzirom na ulogu pšenice kao glavne sirovine za proizvodnju mnogih proizvoda neophodnih za prehranu čovječanstva, jednak naglasak treba biti stavljen i na poboljšanje svojstava kakvoće, a posebno na poboljšanje svojstava koja utječu na krajnju kakvoću proizvoda od pšenice. Ciljevi ovog znanstvenog rada bili su dati pregled dosadašnjih istraživanja o genomskoj selekciji za svojstva kakvoće pšenice i istaknuti ključne čimbenike koji utječu na točnost predviđanja, kako bi se predložio najprimjenjiviji pristup genomskoj selekciji za svojstva kakvoće pšenice.

Prvo istraživanje genomske selekcije za svojstva kakvoće pšenice objavljeno je 2011. godine (Heffner i sur., 2011b). Koristeći više nesrodnih populacija pšenice autori su ispitali mogućnost predviđanja 13 različitih svojstava pšenice, uključujući i svojstva kakvoće kao što su GPC, TW, izbrašnjavanje, retencijska sposobnost brašna i dr. Rezultati su pokazali da točnost predviđanja korištenjem genomske selekcije premašuje točnost predviđanja korištenjem MAS-a za 30 %. Uspoređujući genomsku selekciju s fenotipskom selekcijom utvrdili su 95 %-tnu sličnost predviđanja te predložili uključivanje genomske selekcije i u oplemenjivačke programe usmjerene na kakvoću pšenice budući da je genomska selekcija pokazala potencijal povećanja genetičke dobiti po jedinici vremena i troškova. Od tada do danas objavljeno je više istraživanja genomske selekcije za različita svojstva kakvoće pšenice, a ovaj znanstveni rad donosi pregled 21 relevantnog istraživanja te daje sažetak o najvažnijim čimbenicima koji su ispitani u navedenim istraživanjima, a koji utječu na točnost predviđanja genomske selekcije, kao što su istraživana svojstva, vrsta i veličina korištene populacije, vrsta i broj korištenih biljega, korišteni model predviđanja i dr.

Istraživanja obuhvaćena ovim znanstvenim radom sugeriraju da su veličina trenažne populacije, struktura populacije i gustoća biljega tri glavna čimbenika koja utječu na točnost predviđanja genomske selekcije, a učinci kojih su međusobno isprepleteni (Bassi i sur., 2016; Robertsen i sur., 2019). Već prva istraživanja genomske selekcije za svojstva kakvoće pšenice pokazala su da veličina trenažne populacije ima značajan utjecaj na točnost predviđanja te da veličina trenažne populacije potrebna za postizanje optimalne točnosti predviđanja također ovisi i o srodnosti trenažne i oplemenjivačke (validacijske) populacije. Iz ovog znanstvenog rada vidljivo je da su u dosadašnjim istraživanjima genomske selekcije za svojstva kakvoće korištene različite populacije, u smislu srodnosti i veličine, te da su vrijednosti točnosti predviđanja dobivene korištenjem manjih, ali visokosrodnih populacija usporedive s točnostima predviđanja dobivenim korištenjem većih, nesrodnih populacija. Veličina korištenih biparentalnih populacija kretala se od 128 (Michel i sur., 2018) do 282 DH (eng. *doubled haploid*) linija (Hu i sur., 2019), dok je kod nesrodnih populacija veličina dosezala i 5520 linija (Battenfield i sur., 2016). Gustoća biljega korištena u navedenim istraživanjima kretala se od 973 biljega dobivena s različitim genotipizacijskim platformama (Heffner i sur., 2011a) do 78606 SNP-ova (Juliana i sur., 2019), s tom razlikom da je manja gustoća biljega korištena za biparentalne populacije, dok su značajno veće gustoće biljega bile potrebne kada su korištene nesrodne populacije. Istraživanja provedena za svojstva kakvoće pšenice također su pokazala da je heritabilnost jedan od glavnih čimbenika koji utječu na točnost predviđanja te da svojstva s visokom heritabilnošću obično pokazuju i visoke vrijednosti točnosti predviđanja (Yao i sur., 2018). Svojstva s niskom heritabilnošću, npr. Zeleny indeks sedimentacije, pokazala su izrazito nisku točnost predviđanja (0,1) (Tsai i sur., 2020). Nasuprot tome, za svojstva s visokim vrijednostima heritabilnosti točnost predviđanja dosezala je i 0,96 (npr. SDS sedimentacija) (Liu i sur., 2016). Dobivene vrijednosti heritabilnosti za svojstva kakvoće pšenice kretale su se u rasponu od 0,35 (Tsai i sur., 2020) do 0,96 (Liu i sur., 2016), iako su za većinu istraživanih svojstava vrijednosti heritabilnosti bile umjerene do visoke. Prema rezultatima dosadašnjih istraživanja i utvrđenim vrijednostima, malo je vjerojatno da će heritabilnost predstavljati ograničavajući čimbenik u genomskoj selekciji za svojstva kakvoće pšenice. Usporedba različitih modela korištenih za predviđanje pokazala je da bayesovski modeli ne pokazuju jasnu superiornost u odnosu na druge modele kada je riječ o predviđanju svojstava kakvoće pšenice, odnosno da točnost predviđanja genomske selekcije za svojstva kakvoće pšenice obično nije pod velikim utjecajem odabranog modela predviđanja (Yao i sur., 2018).

Za većinu svojstava kakvoće pšenice genomska selekcija se pokazala kao uspješna metoda predviđanja budući da svojstva kakvoće pokazuju složene obrasce nasljeđivanja, a genomska selekcija istovremeno uzima u obzir učinke svih QTL-a za određeno svojstvo,

bez obzira na veličinu učinka. Istraživanja obuhvaćena ovim znanstvenim radom pokazala su da genomska selekcija omogućuje dobivanje zadovoljavajuće točnosti predviđanja, a što omogućuje njeno uključivanje u oplemenjivačke programe usmjerene na kakvoću pšenice. Također, genomska selekcija može biti od velikog značaja za predviđanje uspješnosti linija u ranim generacijama, jačanje stabilnosti svojstava kakvoće te za učinkovitiji odabir linija pšenice visoke kakvoće.

Ključne riječi: kakvoća pšenice, genomska selekcija, GEBV, točnost predviđanja, trenažna populacija, validacijska populacija, heritabilnost

Prošireni sažetak: Znanstveni rad 3.

Naslov znanstvenog rada na hrvatskom jeziku: Procjena metoda genomske selekcije za svojstva kakvoće pšenice u biparentalnim populacijama ukazuje na potrebu za jednostavnijim rješenjima

Oplemenjivanje za svojstva kakvoće pšenice, a posebno na svojstva kakvoće krajnje upotrebe, često je izazovno jer fenotipska analiza zahtijeva veće količine brašna i zrna koji obično nisu dostupni u ranim fazama oplemenjivačkog procesa. Korištenje miksografa kao brze i učinkovite metode procjene kakvoće tijesta zajedno s genomskom selekcijom može pomoći u procesu selekcije linija visoke kakvoće ranije u oplemenjivačkom procesu čime se postiže veća dobit po jedinici vremena i troška. U ovom znanstvenom radu istražen je potencijal genomske selekcije za predviđanje sedam svojstava kakvoće, uključujući i neka svojstva dobivena miksograf uređajem, u dvije biparentalne populacije pšenice. Ciljevi ovog znanstvenog rada bili su utvrditi potrebu za optimizacijom trenažne populacije na temelju fenotipske varijance, identificirati učinak veličine trenažne populacije i gustoće biljega na točnost predviđanja pomoću RR-BLUP modela te utvrditi uspješnost RR-BLUP modela i sedam drugih modela za predviđanje svojstava kakvoće pšenice.

U ovom istraživanju korištene su dvije biparentalne populacije ozime pšenice: Bezostaya-1 × Klara (BK) i Monika × Golubica (MG). BK populacija sastojala se od 139, a MG od 153 RIL-a. Svi genotipovi su genotipizirani korištenjem DArT tehnologije, a što je nakon filtriranja podataka rezultiralo s 1087 SNP biljega za BK i 2231 SNP biljega za MG populaciju. Poljski pokusi s obje populacije provedeni su na dvije lokacije u Hrvatskoj (Osijek i Slavonski Brod) tijekom tri godine (2009. – 2011.), odnosno u 6 različitih okoliša koje su označene kombinacijom lokacije i godine (OS09 – OS11, SB09 – SB11), a u svakom okolišu pokus je postavljen prema redno-stupčanom dizajnu. Analizirano je sedam svojstava kakvoće: GPC, WGC, TW te miksograf svojstva MPT, MTW, MTI i MPH. U prvom dijelu statističke analize skupni podatci iz pojedinačnih pokusa analizirani su prema mješovitom modelu koji je uključivao fiksne učinke genotipa, okoliša i repeticija unutar okoliša, te nasumične učinke redova i stupaca unutar repeticije unutar okoliša. Tako procijenjene vrijednosti za sve kombinacije genotip-okoliš korištene su za izračun heritabilnosti u širem smislu za sve okoliše zajedno i ponovljivosti za svaki okoliš zasebno, te kao ulazni podatci za genomsku selekciju koja je provedena za svaki okoliš zasebno. RR-BLUP model, koji je robustan i manje kompjuterski zahtjevan u odnosu na druge modele, korišten je u prvom dijelu analize kako bi se utvrdila potreba za optimizacijom trenažne populacije na osnovu fenotipske varijance, te ispitao utjecaj veličine trenažne populacije i gustoće biljega na točnost predviđanja. Za svaku nasumično odabranu trenažnu populaciju izračunata je fenotipska varijanca kako bi se utvrdilo postoji li korelacija između fenotipske varijance

trenažne populacije i dobivene točnosti predviđanja, a korištene veličine trenažne populacije bile su 25, 50 i 75 linija za obje promatrane populacije. Za utvrđivanje utjecaja veličine trenažne populacije na točnost predviđanja korištene su tri različite veličine trenažne populacije, a koje su se sastojale od 50, 65 i 80 % nasumično odabranih linija od ukupnog broja linija unutar svake od promatranih biparentalnih populacija, dok je ostatak od 50, 35 i 20 % linija služio kao validacijska populacija. Kako bi se utvrdio utjecaj gustoće biljega, genomska selekcija za sva svojstva unutar MG populacije provedena je korištenjem cijelog skupa biljega (2231 SNP) te polovice skupa biljega (1123 SNP-a). U drugom dijelu analize uspoređena je učinkovitost RR-BLUP modela i sedam drugih modela predviđanja, a koji su uključivali parametrijske modele (EN, BayesA, BayesB, BayesC i BayesLASSO) i neparametrijske modele (RF i RKHS). Ovaj dio analize proveden je samo na MG populaciji, korištenjem cijelog skupa biljega te veličinom trenažne populacije od 122 RIL-a (80 % od ukupnog broja linija).

Rezultati ovog istraživanja pokazali su da je u obje promatrane populacije heritabilnost za GPC, WGC i TW bila visoka (0,78 – 0,92), dok je heritabilnost svojstava miksografa bila također visoka, iako nešto niža u odnosu na druga svojstva (≥ 0,71). Jedino je MPT pokazao umjerene vrijednosti heritabilnosti u BK populaciji (0,45). Ponovljivosti unutar okoliša bile su jednake ili nešto veće unutar BK u odnosu na MG populaciju za većinu kombinacija svojstvo-okoliš, a najviše vrijednosti su zabilježene u okolišu OS09, dok su najniže vrijednosti za MG populaciju zabilježene u okolišu OS11. Rezultati su pokazali i da je za neke kombinacije svojstvo-okoliš korelacija između fenotipske varijance odabrane trenažne populacije i točnosti predviđanja bila gotovo nula. U nekim kombinacijama populacija-svojstvo-okoliš uočen je blagi pad te pomak s pozitivnih na negativne vrijednosti koeficijenata korelacije, dok je u drugim slučajevima uočeno vrlo blago povećanje korelacije uslijed povećanja veličine trenažne populacije. Općenito, jačina i smjer korelacija pokazali su značajna variranja između različitih okoliša, zabilježene vrijednosti su bile niske ($r \leq 0,35$) te nije uočen nikakav dosljedan obrazac u jačini i smjeru korelacija za bilo koju od ispitanih kombinacija populacija-svojstvo-okoliš. Budući da nije utvrđena veza između fenotipske varijance i točnosti predviđanja, za sve daljnje izračune korištene su samo nasumično odabrane trenažne populacije. Smanjenje veličine trenažne populacije s 85 na 50 % imalo je negativan učinak na dobivenu točnost predviđanja u svim promatranim kombinacijama populacija-svojstvo-okoliš. Također je primijećeno da točnost predviđanja ovisi o okolišu, bez obzira na veličinu trenažne populacije, te može značajno varirati, npr. najveća razlika u točnosti predviđanja uočena je pri veličini trenažne populacije od 80 % za TW u MG populaciji, a iznosila je 0,06 u okolišu OS11 te 0,49 u okolišu OS10. Kada se usporedi utjecaj različitih gustoća biljega na sposobnost predviđanja svojstava kakvoće unutar MG

populacije, vrijednosti točnosti predviđanja dobivene korištenjem veće gustoće biljega bile su više za sve kombinacije svojstvo-okoliš i sve korištene veličine trenažne populacije. Ipak, utvrđene razlike u dobivenoj točnosti predviđanja nisu bile velike. Uzimajući u obzir veličinu trenažne populacije od 80 % najveća razlika u točnosti predviđanja utvrđena je za WGC u okolišu SB11, s vrijednostima od 0,20 i 0,32 kada je korišteno 2231 u odnosu na 1123 SNP biljega. Uspoređujući dvije populacije i slučajeve u kojima je korišten približno isti broj biljega, uočeno je da je točnost predviđanja za GPC, WGC, i TW veća u BK populaciji, dok su svojstva miksografa pokazala bolju predvidljivost u MG populaciji. Općenito je utvrđeno da točnost predviđanja znatno varira između promatranih okoliša. Kada su uspoređene točnosti predviđanja dobivene korištenjem različitih modela u MG populaciji, uočeno je da odabir najuspješnijeg modela ovisi više o okolišu nego o promatranom svojstvu. Za većinu kombinacija svojstvo-okoliš EN model je rezultirao najnižim vrijednostima točnosti predviđanja, a koje su u velikom broju slučajeva bile i negativne. Za 35 od 42 moguće kombinacije svojstvo-okoliš bayesovski modeli su se pokazali najuspješnijima, dok je RF model bio najuspješniji u pet, a RKHS u sedam slučajeva. Iako je RR-BLUP model dao najviše vrijednosti točnosti predviđanja u samo jednom slučaju, točnosti predviđanja dobivene drugim modelima nisu bile znatno više (u prosjeku za 0,05). Jedna od najvećih razlika među korištenim modelima je vrijeme potrebno za jednu analizu, a u tom pogledu se najmanje zahtjevnim pokazao EN model (19 min za jednu analizu), dok su najzahtjevniji bili bayesovski modeli (do 171 min potrebne za jednu analizu).

U ovom istraživanju utvrđeno je da veličina trenažne populacije igra važnu ulogu u postizanju viših vrijednosti točnosti predviđanja, dok gustoća biljega ne predstavlja značajno ograničenje. Dobiveni rezultati nisu podržali optimizaciju trenažne populacije na temelju fenotipske varijance. Iako se RR-BLUP nije pokazao kao najuspješniji model u svim slučajevima, nije uočena značajna prednost korištenja bilo kojeg drugog modela. Točnosti predviđanja dobivene u sklopu ovog istraživanja podržavaju primjenu genomske selekcije za oplemenjivanje pšenice na kakvoću, uključujući i oplemenjivanje na neka svojstva dobivena miksograf uređajem.

Ključne riječi: pšenica, svojstva kakvoće, genomska selekcija, biparentalna populacija, RIL, modeli predviđanja, trenažna populacija, fenotipska varijanca

3.2. Objedinjena rasprava

3.2.1. Transgresivna segregacija, fenotipska korelacija i heritabilnost svojstava kakvoće pšenice u biparentalnim populacijama

U ovom istraživanju korištene su dvije biparentalne populacije pšenice dobivene iz križanja roditeljskih sorti Bezostaya-1 × Klara (BK populacija) i Monika × Golubica (MG populacija). Roditeljske sorte korištene u BK populaciji međusobno su se razlikovale u svim HMW gluteninskim podjedinicama, dok se roditeljske sorte u MG populaciji nisu razlikovale niti u jednoj od HMW gluteninskih podjedinica (Horvat i sur., 2006). BK populacija predstavlja primjer križanja koje se često primjenjuje u praktičnim oplemenjivačkim programima gdje se križaju fenotipski značajno različiti roditelji, dok je MG kombinacija odabrana kako bi se fenotipske varijacije svojstava kakvoće suzile na genetičke čimbenike isključujući razlike uzrokovane posjedovanjem različitih HMW gluteninskih podjedinica.

Za proizvodnju tijesta i kruha visoke kakvoće pšenica mora posjedovati odgovarajuću kakvoću, a koja se često procjenjuje na osnovu GPC-a i WGC-a. Iako se poželjni raspon vrijednosti svojstava kakvoće može značajno razlikovati ovisno o namjeni pšeničnog brašna, za proizvodnju kruha visoke kakvoće GPC ne bi trebao biti niži od 12,5 % (Turner i sur., 2004), dok je minimalna poželjna vrijednost WGC-a za pšenično brašno 24 % (Singh i Singh, 2015). Nasuprot tome, istraživanja su pokazala da kakvoća i sastav gluteninskih podjedinica, a posebno HMW gluteninskih podjedinica, imaju važniju ulogu u određivanju kakvoće tijesta od samog sadržaja glutena (Horvat i sur., 2006; Payne, 1987; Shewry i sur., 1995). Ranija istraživanja kakvoće koja su uključivala sorte pšenice korištene u ovom istraživanju pokazala su da je Bezostaya-1 sorta više kakvoće u odnosu na preostale tri sorte, a što se pripisuje prisutnosti 5+10 HMW gluteninskih podjedinica na *Glu-D1* lokusu. Iako Klara, Monika i Golubica posjeduju 2+12 HMW gluteninske podjedinice, čija se prisutnost obično dovodi u vezu s lošijom kakvoćom, ove sorte su pokazale dobru kakvoću, a što se može objasniti većim udjelom ukupnih HMW gluteninskih podjedinica (Horvat i sur., 2006). Rezultati ovog istraživanja (znanstveni rad 1) također potvrđuju da sva četiri roditeljska genotipa posjeduju zadovoljavajuću kakvoću u pogledu zabilježenih vrijednosti GPC-a i WGC-a u svim promatranim okolišima. Izuzetak je bila jedino sorta Monika u okolišu SB11 čija je GPC vrijednost bila 12,0 %. Iako nije uvijek jednostavno definirati poželjne vrijednosti svojstava dobivenih miksografom jer one ovise o promatranoj vrsti populacije i namjeni brašna, općenito vrijedi da tijesta dobre kakvoće karakterizira jak gluten na što ukazuje dulje vrijeme razvoja tijesta (MPT), veća konzistencija i stabilnost nakon miješanja (MTW), veća energija uložena u proces miješanja (MTI) i veća jakost tijesta (MPH) (Campbell i sur., 2001). Za svih sedam svojstava kakvoće utvrđeno je postojanje

pozitivnih i negativnih transgresivnih segregirajućih linija, iako je njihova pojavnost bila općenito veća u BK populaciji, a što osigurava širu osnovu za odabir široko prilagođenih ili specifično prilagođenih RIL-ova. Takvi rezultati su očekivani kad se u obzir uzme činjenica da roditeljski genotipovi Bezostaya-1 i Klara imaju različitu genetičku osnovu kakvoće (različite HMW gluteninske podjedinice na svim *Glu-1* lokusima) (Horvat i sur., 2006). Srednje vrijednosti RIL-ova bile su slične srednjim vrijednostima roditeljskih genotipova za sva svojstva, osim za MPT i MTW čije su srednje vrijednosti bile više od srednjih vrijednosti oba roditelja što može biti pokazatelj jačih epistatskih učinaka za navedena dva svojstva. S druge strane, Monika i Golubica dijele istu genetičku osnovu kakvoće (iste HMW gluteninske podjedinice na svim *Glu-1* lokusima), ali pokazuju veći fenotipski raspon u usporedbi s genotipovima Bezostaya-1 i Klara što posljedično rezultira manjom zastupljenošću transgresivnih segregirajućih linija. Pojava transgresivne segregacije za sva svojstva kakvoće korištena u ovom istraživanju ukazuje na prisutnost alela koji bi se mogli koristiti za poboljšanje svojstava kakvoće u sva četiri roditeljska genotipa.

Visoka pozitivna korelacija između GPC-a i WGC-a zabilježena u obje populacije (znanstveni rad 1) bila je očekivana, uzimajući u obzir da je gluten najzastupljeniji protein pšenice i da je međusobna povezanost ova dva svojstva već dobro dokumentirana (Kristensen i sur., 2018; Laidig i sur., 2017). Osim toga, u obje populacije uočena je visoka pozitivna korelacija između MTI i MPH što može upućivati na to da je u procesu miješanja tijesta veće jakosti potrebno uložiti veću energiju i da taj odnos nije ovisan o genetičkoj osnovi kakvoće pšenice. Korelacije između WGC i MTI/MPH bile su jače u BK u usporedbi s MG populacijom, što upućuje na to da WGC može utjecati na jakost tijesta u slučaju kada barem jedan roditeljski genotip posjeduje HMW gluteninske podjedinice povezane s dobrom kakvoćom. S druge strane, vrlo niske pozitivne, pa čak i negativne korelacije uočene između WGC i MPT, odnosno MTW, upućuju na to da optimalno vrijeme razvoja i konzistencija tijesta ne ovise o količini glutena, već o njegovoj kakvoći koja je uglavnom uvjetovana sastavom HMW gluteninskih podjedinica (Payne i sur., 1987). Važnost *Glu-D1* lokusa u manifestaciji reoloških svojstava potvrđena je u ranijim istraživanjima (Campbell i sur., 2001; Zhang i sur., 2009; Zheng i sur., 2013). Dobiveni rezultati ukazuju na to da genetička osnova kakvoće roditeljskih genotipova nema utjecaja na svojstva kakvoće kao što su GPC, WGC i TW ukoliko su njihove vrijednosti već unutar poželjnog raspona, ali može značajno utjecati na reološke osobine tijesta.

Heritabilnost u širem smislu (znanstveni rad 3) procijenjena za sve okoliše zajedno za svojstva kakvoće bila je visoka, uz izuzetak svojstva MPT u BK populaciji čija je heritabilnost bila umjerena. Iako je ponovljivost unutar okoliša značajno varirala, u većini slučajeva je bila visoka. Zabilježena visoka heritabilnost ukazuje na to da genetički

čimbenici imaju jači učinak na vrijednosti svojstava kakvoće, a što omogućuje njihovo predviđanje korištenjem genomske selekcije koja nastoji obuhvatiti ukupnu genetičku varijancu na osnovu velikog broja biljega (Meuwissen i sur., 2001). Općenito, najniže vrijednosti ponovljivosti uočene su u okolišu SB09 za BK populaciju i u okolišu OS11 za MG populaciju što ukazuje na prisutnost jačeg negenetičkog učinka unutar tih okoliša. Vrijednosti heritabilnosti i ponovljivosti za sedam svojstava kakvoće promatranih u sklopu ovog istraživanja jednake su ili više u odnosu na vrijednosti dobivene u ranijim istraživanjima svojstava kakvoće (znanstveni rad 2), a što upućuje na to da heritabilnost ne bi trebala predstavljati značajnu prepreku za postizanje dobre točnosti predviđanja korištenjem genomske selekcije (Kristensen i sur., 2019; Michel i sur., 2017; Yao i sur., 2018). Unatoč tome, istraživanja su pokazala da točnost predviđanja u nekim okolišima može biti niska iako je heritabilnost visoka, a što se objašnjava različitim okolišnim čimbenicima (Dawson i sur., 2013; Heslot i sur., 2013).

3.2.2. Interakcija genotip × okoliš za svojstva kakvoće pšenice

Tijekom posljednja četiri desetljeća u literaturi o GEI raspravlja se problematika pronalaženja odgovarajućih metoda za testiranje značajnosti IPCA osi prilikom provođenja analize AMMI modelom. Do danas su predloženi različiti testovi, a uspješnost nekih od njih uspoređena je u nekoliko ranije objavljenih istraživanja (Forkman i Piepho, 2014; Piepho, 1995). Međutim, zbog specifične prirode istraživanja temeljenih na RIL-ovima koja obično uključuju veliki broj genotipova testiranih u samo nekoliko okoliša, u ovom istraživanju (znanstveni rad 1) primijenjena su tri različita testa i uspoređeni su dobiveni rezultati s ciljem utvrđivanja optimalnog testa za primjenu na RIL populacijama. Za odabir broja osi koje je potrebno zadržati u konačnom AMMI modelu korištena su tri pristupa: (1) SPB (Forkman i Piepho, 2014), (2) F_R-test (Piepho, 1993) i (3) LOO-CV (Paderewski, 2013). Rezultati dobiveni primjenom navedene tri metode za testiranje značajnosti IPCA osi pokazali su nedosljednost u pogledu odabranog optimalnog AMMI modela. Od tri korištena pristupa, LOO-CV se pokazala kao najkonzervativnija metoda, odabirući potpuni aditivni model (AMMI0) za gotovo sve promatrane slučajeve. Odabir AMMI0 modela kao najprikladnijeg ukazuje na potpunu odsutnost GEI, a što za svojstva promatrana u ovom istraživanju nije prihvatljivo rješenje, osobito u slučajevima kada se veći dio ukupne sume kvadrata pripisuje GEI. Od dva preostala pristupa, F_R-test se pokazao znatno liberalnijim predlažući zadržavanje većeg broja IPCA osi u konačnom modelu. Međutim, odabir složenijih modela ujedno dovodi i do prekomjernog zadržavanja šuma u modelu (eng. *overfitting*). Slijedom navedenih razloga SPB metoda odabrana je kao najprikladnija za odabir broja IPCA osi u konačnom AMMI modelu za promatrana svojstva kakvoće u RIL populacijama. Takav odabir u skladu je s preporukama istraživanja koje su proveli Forkman i Piepho (2014), a

koji su zaključili da SPB metoda nadmašuje F_R-test i LOO-CV u pogledu niske vjerojatnosti dobivanja lažno pozitivnih rezultata. S obzirom da je odabran kao optimalan u većini slučajeva te da odabir složenijih modela (AMMI3 u ovom istraživanju) može dovesti do prekomjernog zadržavanja šuma u modelu (Gauch i Zobel, 1990), daljnja GEI analiza provedena je koristeći AMMI2 model za svih sedam svojstava kakvoće. Problematika prekomjernog zadržavanja šuma u modelu može biti još naglašenija u slučaju velike količine nedostajućih podataka (Paderewski i Rodrigues, 2014) pa su za takve skupove podataka složenije metode za imputaciju podataka prikladnije od EM-AMMI metode korištene u ovom istraživanju (Arciniegas-Alarcón i sur., 2020; Paderewski, 2013).

Primjenom AMMI modela za analizu GEI (znanstveni rad 2) utvrđeno je da je za tri svojstva kakvoće (GPC, WGC i TW) u obje populacije dominantan izvor varijabilnosti bio okoliš. Relativni doprinosi genotipa i GEI bili su podjednaki za sva tri svojstva u BK populaciji. U MG populaciji je uočena manja razlika između doprinosa genotipa i GEI za navedena tri svojstva, odnosno doprinos genotipa bio je otprilike 1,5 puta veći od doprinosa GEI (za GPC i WGC) ili čak i manji (za TW). Četiri svojstva miksografa mogu se prema sličnosti dobivenih rezultata podijeliti u dvije skupine: (1) MPT i MTW za koje je dominantan izvor varijabilnosti u BK populaciji bila GEI, dok je u MG populaciji dominantan izvor bio genotip, i (2) MTI i MPH za koje je dominantan učinak u BK populaciji imao okoliš, dok je u MG populaciji dominantan učinak na ova dva svojstva imala GEI. Prema Wiliams i sur. (2008) dominantan učinak okoliša za svojstva GPC, WGC i TW opažen je u brojnim prijašnjim istraživanjima iako postoje i iznimke od ovog pravila. Međutim, s obzirom na specifičnosti istraživanja koje koriste RIL-ove, kao što su velik broj genotipova i manji broj okoliša, rezultate dobivene ovim istraživanjem trebalo bi usporediti sa sličnim istraživanjima temeljenim na RIL populacijama. Ako se u obzir uzmu rezultati novijih istraživanja na RIL-ovima, dominantan učinak okoliša gotovo da i nije zabilježen. Prashant i sur. (2015) utvrdili su dominantan učinak okoliša samo za GPC, dok su Echeverry-Solarte i sur. (2015) i Krishnappa i sur. (2019) pokazali da genotip i okoliš imaju jednak učinak na GPC. Nasuprot tome, prema Goel i sur. (2019) učinak okoliša bio je manji od svih ostalih učinaka za svojstva GPC, WGC i TW. Najčešće istraživano svojstvo miksografa u RIL populacijama je MPT za koje su Prashant i sur. (2015) utvrdili da je pod dominantnim učinkom okoliša, kao što je to slučaj i u ovom istraživanju za BK populaciju. U MG populaciji za isto svojstvo dominantan učinak je imala GEI, a što odgovara rezultatima istraživanja Goel i sur. (2019). Neka istraživanja su pokazala da na MPT dominantan učinak ima genotip (Echeverry-Solarte i sur., 2015; Jin i sur., 2016). Dominantan učinak GEI na MTW i okoliša na MTI utvrdili su i Prashant i sur. (2015), a što odgovara rezultatima ovog istraživanja za BK populaciju. Navedena istraživanja na RIL populacijama razlikuju se po mnogim genotipskim i okolišnim

čimbenicima, što može biti jedan od glavnih razloga utvrđenih razlika u rezultatima GEI analize.

Analizom modificiranih AMMI2 biplota utvrđeno je postojanje nekih najuspješnijih genotipova u obje populacije (znanstveni rad 1). Za visokokorelirana svojstva, npr. GPC i WGC, u većini promatranih okoliša isti su se RIL-ovi pokazali najuspješnijima (npr. BK032, BK007, MG124). Niti u jednoj od promatranih populacija nisu uočeni zajednički najuspješniji RIL-ovi za svojstva koja nisu bila međusobno u visokoj korelaciji. Općenito, nisu utvrđeni jasni obrasci između okoliša ili godina pa je pretpostavka da uspješnost genotipova uglavnom ovisi o prisutnosti pojedinačnih povoljnih ili nepovoljnih kombinacija lokacije i godine. Jedan takav primjer je okoliš OS10 koji se pokazao kao povoljan okoliš za većinu svojstava BK populacije, odnosno kao nepovoljan za TW. Za MG populaciju isti okoliš se pokazao povoljnim samo za svojstva GPC i WGC. U oplemenjivanju bilja naglasak se obično stavlja na pronalaženje široko ili usko prilagođenih RIL-ova. Iz AMMI2 biplota vidljivo je postojanje nekih zajedničkih najuspješnijih RIL-ova koji u selekciji predstavljaju široko prilagođene genotipove, npr. BK012 za TW, BK042 za MPT i MG124 za GPC i WGC. Sa stajališta oplemenjivanja zanimljiv primjer je genotip MG124 koji se pokazao kao dominantni najuspješniji genotip za GPC te kao najuspješniji genotip u nekoliko okoliša za visokokorelirano svojstvo WGC, ali također i kao najuspješniji genotip za nekorelirano svojstvo TW u okolišu SB10. Poseban slučaj predstavljaju i oni genotipovi čija se boja na modificiranom AMMI2 biplotu ne poklapa s bojom okoliša za koju su prilagođeni, a što ukazuje na to da je njihova srednja vrijednost puno viša ili niža od srednje vrijednosti svojstva za određeni okoliš. Npr. za genotip BK007 zabilježena je visoka vrijednost GPC-a u okolišu SB09 za koji je srednja vrijednost GPC-a niska te ga se zbog toga treba smatrati široko, a ne usko prilagođenim genotipom jer posjeduje sposobnost zadržavanja visokog GPC-a čak i u nepovoljnim okolišima. S druge strane, utvrđeno je postojanje i nekih najuspješnijih genotipova čija se visoka ili niska srednja vrijednost podudara sa srednjom vrijednosti okoliša kojem su prilagođeni, npr. MG016 i okoliš SB09 za MTW te BK059 i okoliš SB09 za MTI. Međutim, za navedena svojstva miksografa nije nužno poželjno da pokazuju visoku vrijednost te podudarnost visoke vrijednosti najuspješnijeg genotipa i okoliša može biti i negativna ako učinak GEI pomiče vrijednosti ovih svojstava iznad njihove poželjne vrijednosti. MTW ukazuje na stabilnost tijesta na kraju procesa miješanja, a MTI na energiju uloženu tijekom procesa zamjesa te visoke vrijednosti ovih svojstava mogu ukazivati na tijesto loše kakvoće koje je kruto i neelastično.

Rezultati koji proizlaze iz znanstvenog rada 1 potvrđuju hipotezu 3 ovog doktorskog rada te pokazuju da je GEI imala znatno važniji učinak na svojstva miksografa u odnosu na preostala tri promatrana svojstva kakvoće. Uz pomoć AMMI2 biplota utvrđeno je postojanje

nekoliko široko i specifično prilagođenih RIL-ova, a koji pokazuju potencijal za korištenje u oplemenjivačkim programima usmjerenima na poboljšanje kakvoće pšenice.

3.2.3. Genomska selekcija za svojstva kakvoće pšenice

Prije donošenja odluke o uključivanju genomske selekcije u oplemenjivački program potrebno je prilagoditi sve čimbenike koji mogu utjecati na točnost predviđanja. Iako su se troškovi genotipizacije značajno smanjili posljednjih godina, troškovi fenotipizacije za svojstva kakvoće, a posebno za svojstva pekarske kakvoće, ostali su još uvijek relativno visoki (Kristensen i sur., 2018). Stoga je prvi korak ka primjeni genomske selekcije za predviđanje svojstava kakvoće optimizacija trenažne populacije kako bi se smanjili potencijalni troškovi fenotipizacije uz održavanje zadovoljavajuće razine točnosti predviđanja. Većina istraživanja koja se bavila problematikom optimizacije trenažne populacije je pokazala da veličina trenažne populacije ima značajan utjecaj na točnost predviđanja (Battenfield i sur., 2016; Kristensen i sur., 2019; Liu i sur., 2016; Lorenz, 2013; Michel i sur., 2016). Osim toga, ranija istraživanja su pokazala da srodnost trenažne i validacijske populacije igra važnu ulogu u određivanju optimalne veličine trenažne populacije (Edwards i sur., 2019). U ovom istraživanju (znanstveni rad 3) korištene su tri različite veličine trenažne populacije (50, 65 i 80 % od ukupnog broja RIL-ova u svakoj od populacija) kako bi se utvrdio njihov utjecaj na točnost predviđanja sedam svojstava kakvoće pšenice unutar biparentalnih populacija. Najveća točnost predviđanja za sva svojstva postignuta je korištenjem 80 % RIL-ova kao trenažne populacije, odnosno kada je trenažna populacija uključivala 111 RIL-ova za BK i 122 RIL-a za MG populaciju. Smanjenje veličine trenažne populacije s 85 na 50 % od ukupnog broja linija u populaciji imalo je negativan utjecaj na postignutu točnost predviđanja u svim promatranim kombinacijama populacija-svojstvo-okoliš. Međutim, iako negativan, utjecaj smanjenja trenažne populacije nije bio velik što ukazuje na činjenicu da se čak i uz korištenje manje trenažne populacije može postići umjerena do visoka točnost predviđanja. Prijašnja istraživanja su također pokazala da je relativno mala veličina trenažne populacije dovoljna za postizanje visoke točnosti predviđanja (Kristensen i sur., 2019; Lado i sur., 2018; Lozada i sur., 2019; Verges i van Sanford, 2020), osobito u visokosrodnim populacijama kao što su biparentalne populacije (Heffner i sur., 2011a; Lorenzana i Bernardo, 2009). Veličina trenažne populacije koja će biti korištena za provođenje genomske selekcije važna je u pogledu pravilnije raspodjele resursa, tj. za donošenje odluke o broju genotipova koji će biti uključeni u pokus i fenotipizirani, a što je od posebnog značaja kada je postupak fenotipizacije dugotrajan ili financijski zahtjevan. Veličine korištenih populacija u ovom istraživanju u skladu su s veličinama biparentalnih populacija na kojima su ranije provedena istraživanja genomske selekcije za svojstva kakvoće pšenice (znanstveni rad 2).

Osim optimizacije u smislu veličine trenažne populacije neka istraživanja predlažu i odabir jedinki koje će činiti trenažnu populaciju na osnovu zadanog kriterija kao što je PEV ili CD vrijednost, kako bi se povećala točnost predviđanja. Koristeći dvije populacije hibrida kukuruza Rincent i sur. (2012) su pokazali da optimizacija trenažne populacije na temelju CD-a povećava pouzdanost genomske selekcije budući da dovodi do smanjenja varijance uslijed veće srodnosti jedinki u odabranoj trenažnoj populaciji. U istraživanju Marulanda i sur. (2015) proučavan je utjecaj različitih parametara trenažne populacije na točnost predviđanja korištenjem simulirane biparentalne populacije kukuruza. Od svih parametara uključenih u istraživanje samo je za fenotipsku varijancu trenažne populacije utvrđena pozitivna korelacija s dobivenim točnostima predviđanja. Na osnovu dobivenih rezultata autori su predložili optimizaciju trenažne populacije tako da se poveća njena fenotipska varijanca kako bi se postigle veće vrijednosti točnosti predviđanja. U ovom istraživanju (znanstveni rad 3) opaženi koeficijenti korelacije između fenotipske varijance nasumično odabrane trenažne populacije i dobivene točnosti predviđanja bili su niski ($r \leq 0,35$) i nije uočen dosljedan obrazac u jačini ili smjeru korelacije za bilo koju od korištenih kombinacija populacija i svojstva. Budući da nije pronađena jaka korelacija između fenotipske varijance i točnosti predviđanja, nije utvrđena opravdana potreba da se trenažna populacija optimizira na osnovu fenotipske varijance kako bi se postigla veća točnost predviđanja za svojstva kakvoće pšenice te je za sve naknadne analize korištena isključivo nasumično odabrana trenažna populacija.

Iako su troškovi genotipizacije značajno smanjeni u posljednjih nekoliko godina, genotipizacija i dalje za oplemenjivače ostaje jedna od financijski zahtjevnijih strana genomske selekcije. Stoga je važno optimizirati gustoću biljega koja će biti korištena, a da pritom točnost predviđanja ostane zadovoljavajuća. Prema rezultatima ranijih istraživanja, povećanje gustoće biljega ima pozitivan učinak na točnost predviđanja. Međutim, nakon postizanja određene gustoće biljega doseže se plato nakon kojeg daljnje povećanje gustoće više nema značajan utjecaj na točnost predviđanja (Haile i sur., 2018; Heffner i sur., 2011a). Gustoća biljega potrebna za dostizanje platoa predviđanja ovisi o srodnosti korištenih populacija. Liu i sur. (2016) su utvrdili da do postizanja platoa dolazi prilikom korištenja 3000 biljega, odnosno da se plato može postići već i pri gustoći od 500 biljega kada su trenažna i validacijska populacija genetski srodnije. Zbog niske stope rekombinacije genetski srodnije populacije obično imaju visoku stopu LD-a između biljega i QTL-a. Posljedično, potrebna je niža gustoća biljega kako bi se dosegao plato točnosti predviđanja u biparentalnim populacijama (Lorenzana i Bernardo, 2009). Prema Haile i sur. (2018) plato predviđanja za GPC unutar DH populacije postignut je pri gustoći od 2000 biljega. Istraživanje Juliana i sur. (2019) je pokazalo da, nakon što se postigne određena rezolucija, povećanje gustoće

biljega ima neznatan učinak na točnost predviđanja svojstava kakvoće u biparentalnim populacijama pšenice. Za svojstva kakvoće ova rezolucija može biti postignuta čak i pri niskoj gustoći biljega, tj. korištenjem 256 biljega u biparentalnim populacijama (Heffner i sur., 2011a) i 768 biljega u manje srodnim populacijama (Heffner i sur., 2011b). U ovom istraživanju (znanstveni rad 3) uspoređene su točnosti predviđanja dobivene korištenjem cijelog dostupnog skupa podataka o biljezima (2231 SNP-ova) i polovice skupa podataka o biljezima (1123 SNP-ova) u MG populaciji. Prema rezultatima iz postojeće literature (znanstveni rad 2) gustoća od 1123 biljega bi trebala biti dovoljna za postizanje prihvatljivih vrijednosti točnosti predviđanja u biparentalnoj populaciji. Kada se usporede točnosti predviđanja svojstava kakvoće unutar MG populacije dobivene korištenjem različitih gustoća biljega vidljivo je da su veće vrijednosti točnosti predviđanja dobivene kada je korištena veća gustoća biljega, iako razlike u dobivenoj točnosti predviđanja nisu bile velike. Uzimajući u obzir veličinu trenažne populacije od 80 % najveća razlika u točnosti predviđanja utvrđena je za svojstvo WGC u okolišu SB11 gdje je zabilježena vrijednost iznosila 0,20 kada je korištena manja gustoća biljega, odnosno 0,32 kada je korištena veća gustoća biljega. Općenito, dobiveni rezultati pokazuju da je za svojstva GPC, WGC i MPT porast točnosti predviđanja bio nešto veći u odnosu na preostala četiri svojstva (TW, MTW, MTI i MPH). Ovakvi rezultati upućuju na to da je manji broj biljega već dovoljan za postizanje dobre točnosti predviđanja za neka svojstva te da je s većim brojem biljega možda dosegnut plato točnosti predviđanja za ovu populaciju i promatrana svojstva, dok za druga svojstva daljnje povećanje gustoće biljega može poboljšati točnost predviđanja. Prema Gorjanc i sur. (2017) korištenje biljega niske gustoće u kombinaciji s većom trenažnom populacijom omogućuje da se udvostruči dobivena vrijednost točnosti predviđanja u biparentalnim populacijama, a što upućuje na to da veličina trenažne populacije ima važniju ulogu u postizanju visoke točnosti predviđanja od gustoće biljega (Lorenzana i Bernardo, 2009).

Ranija istraživanja koja su uključivala svojstva kakvoće pšenice već su zabilježila dovoljno visoke točnosti predviđanja koje omogućuju uključivanje genomske selekcije u oplemenjivačke programe pšenice i selekciju u ranim generacijama oplemenjivačkog ciklusa. Pregled rezultata provedenih istraživanja i dobivenih vrijednosti točnosti predviđanja prikazan je u znanstvenom radu 2. Uspoređujući točnosti predviđanja dobivene korištenjem RR-BLUP modela za dvije populacije korištene u ovom istraživanju (znanstveni rad 3) može se uočiti da je točnost predviđanja za neka svojstva veća u jednoj, odnosno u drugoj populaciji. Općenito, za GPC, WGC i dva svojstva miksografa (MPT i MTW) utvrđena je umjerena predvidljivost s vrijednostima točnosti predviđanja do 0,57. Predvidljivost svojstva TW i druga dva svojstva miksografa (MTI i MPH) bila je niska i znatno je varirala između okoliša što je u nekim slučajevima rezultiralo i negativnim vrijednostima točnosti

predviđanja. Uz iznimku nekih okoliša, veća je točnost predviđanja za GPC, WGC i TW utvrđena u BK populaciji, dok su svojstva miksografa MPT, MTW i MPH pokazala bolju predvidljivost u MG populaciji. Svojstvo MTI u obje populacije pokazalo je nisku točnost predviđanja (uglavnom < 0,2 te u nekim slučajevima i negativne vrijednosti) uz visoke vrijednosti prosječne kvadratne pogreške predviđanja (eng. *mean-squared error of prediction*; MSEP) što može upućivati na to da reološka svojstva koja opisuju snagu ili otpornost tijesta nisu dobra ciljana svojstva za provođenje genomske selekcije. Točnost predviđanja svojstava kakvoće u ovom istraživanju značajno je varirala između različitih promatranih okoliša, ali su dobivene vrijednosti općenito usporedive s vrijednostima iz postojeće literature (Heffner i sur., 2011a; Michel i sur., 2018). U istraživanju Kristensen i sur. (2018) točnosti predviđanja svojstava kakvoće, uključujući i GPC, bile su 0,5 ili više kada je korištena nasumično odabrana trenažna populacija, dok su točnosti predviđanja dobivene u istraživanju Charmet i sur. (2014) za TW iznosile do 0,7 što je više nego u ovom istraživanju gdje su najviše vrijednosti točnosti predviđanja za TW dosezale 0,6. Umjerena točnost predviđanja (do 0,62) za svojstva miksografa i niska točnost predviđanja za TW utvrđeni su u istraživanju Battenfield i sur. (2016). Nasuprot tome, točnosti predviđanja za osam svojstava kakvoće, uključujući WGC i svojstva miksografa, u istraživanju Lado i sur. (2018) bile su niske do umjerene (0,24 – 0,43). Međutim, i niže vrijednosti točnosti predviđanja mogu biti korištene za predviđanje uspješnosti linija u ranim generacijama, ako se odabir linija provodi istovremeno na temelju GEBV i BLUP vrijednosti (Belamkar i sur., 2018).

Rezultati koji proizlaze iz znanstvenog rada 3, a odnose se na utjecaj različitih čimbenika na točnost predviđanja svojstava kakvoće, djelomično potvrđuju hipotezu 2 ovog doktorskog rada. Utvrđeno je da visoka heritabilnost omogućuje predviđanje svojstava kakvoće zrna pšenice sa srednjom do srednje-visokom točnošću te da veličina trenažne populacije igra važnu ulogu u postizanju viših vrijednosti točnosti predviđanja, dok veličina skupa korištenih biljega ne predstavlja značajno ograničenje.

Osim fenotipskih i genotipskih čimbenika promatrane populacije i svojstva te veličine trenažne populacije i gustoće biljega, odabrani model predviđanja također može utjecati na točnost predviđanja genomske selekcije. U ovom istraživanju (znanstveni rad 3) uspoređena je uspješnost RR-BLUP modela s uspješnošću pet drugih parametrijskih modela (EN, BayesA, BayesB, BayesC i BayesLASSO) i dva neparametrijska modela (RF i RKHS). Kada se usporedi uspješnost predviđanja različitih modela uočljivo je da odabir najuspješnijeg modela uvelike ovisi o promatranom okolišu, a ne toliko o svojstvu. Najniže vrijednosti točnosti predviđanja za većinu kombinacija svojstvo-okoliš dobivene su korištenjem EN modela koji je ujedno bio i model s najvećim brojem slučajeva u kojima je

dobivena negativna vrijednost točnosti predviđanja. U samo jednom slučaju je EN model postigao najveću točnost predviđanja (svojstvo MPT u okolišu SB10). Međutim, u nekim je slučajevima EN bio jednako uspješan kao RR-BLUP, a s obzirom da je ujedno bio i kompjuterski najmanje zahtjevan model može se preporučiti za slučajeve u kojima oplemenjivački program uključuje veliki broj linija te je važnije selekciju provesti u što kraćem roku nego s vrlo visokom preciznošću. U najvećem broju kombinacija svojstvo-okoliš bayesovski modeli (BayesA, BayesB i BayesC) pokazali su se najuspješnijima. Međutim, bayesovski modeli su u ovom istraživanju bili kompjuterski najintenzivniji modeli koji su zahtijevali više od 2 sata za jednu analizu. Stoga bi se bayesovski modeli mogli preporučiti za oplemenjivačke programe s manjim brojem linija gdje je potrebno selekciju provesti s većom preciznošću. Ipak, vrijednosti točnosti predviđanja dobivene bayesovskim modelima nisu bile znatno više od vrijednosti dobivenih RR-BLUP modelom, a što se može objasniti visokom srodnošću trenažne i validacijske populacije korištene u ovom istraživanju (Gao i sur., 2013; Kristensen i sur., 2018). BayesLASSO model je bio najuspješniji u samo dva slučaja, a slijedili su ga RF model s pet i RKHS model sa sedam slučajeva. S obzirom da su modeli RF i RKHS nadmašili RR-BLUP u samo nekoliko slučajeva, ne mogu se preporučiti kao modeli od izbora za svojstva kakvoće pšenice, kao što je to bio slučaj u nekim ranijim istraživanjima (Sandhu i sur., 2021a). Iako je model RR-BLUP bio najuspješniji u samo jednom slučaju (svojstvo TW u okolišu OS10) uspješnost svih ostalih korištenih modela nije bila znatno veća u usporedbi s RR-BLUP modelom. Točnost predviđanja najuspješnijeg modela bila je u prosjeku za samo 0,05 viša i kretala se od 0 (TW u okolišu OS10 gdje su BayesA i BayesC modeli imali istu točnost predviđanja kao RR-BLUP) do 0,14 (MPT u okolišu OS09 gdje je najuspješniji model bio BayesB s točnošću predviđanja od 0,18 u usporedbi s 0,04 za RR-BLUP model). Iako su do danas razvijeni brojni modeli predviđanja za genomsku selekciju, niti jedan nije pokazao jasnu prednost u odnosu na druge modele u smislu postizanja veće točnosti predviđanja bez obzira na promatrano svojstvo (Heslot i sur., 2012; Yao i sur., 2018). Neka istraživanja pokazala su da nema značajne razlike u uspješnosti predviđanja između BLUP i bayesovskih modela za većinu svojstava kakvoće pšenice (Heffner i sur., 2011a; Kristensen i sur., 2018; Tsai i sur., 2020), dok su druga pokazala da su bayesovski modeli učinkovitiji u hvatanju LD-a između biljega i QTL-a, što ih čini uspješnijima za predviđanja u slučajevima kada trenažna i validacijska populacija nisu visokosrodne (Gao i sur., 2013; Zhao i sur., 2014). Prema istraživanju Sandhu i sur. (2021b) modeli dubokog učenja postigli su 10 % veću točnost predviđanja za svojstva GPC i TW, dok je drugo istraživanje Sandhu i sur. (2021a) potvrdilo da su modeli dubokog učenja i modeli strojnog učenja, među kojima i RF, općenito superiorniji u odnosu na druge modele za predviđanje svojstava kakvoće pšenice. Prema rezultatima dobivenim u ovom istraživanju RF model bio je jedan od najmanje uspješnih

modela za predviđanje svojstava kakvoće pšenice. Iako u ovom istraživanju RR-BLUP nije bio najuspješniji model u svim promatranim slučajevima, nije uočena značajna prednost korištenja bilo kojeg drugog modela. Budući da nije dokazana jasna superiornost jednog modela u odnosu na drugi u smislu postignute točnosti predviđanja, najbolji izbor predstavljaju kompjuterski manje zahtjevni modeli koji također postižu prihvatljive vrijednosti točnosti predviđanja, kao što je RR-BLUP.

Još jedan od izazova s kojima se susreću oplemenjivači prilikom uključivanja genomske selekcije u oplemenjivačke programe je prisutnost GEI (Heslot i sur., 2014). Ranija istraživanja već su pokazala da točnost predviđanja za određeno svojstvo može značajno varirati između različitih okoliša (Crossa i sur., 2016a, 2010), a rezultati dobiveni ovim istraživanjem (znanstveni rad 3) to su i potvrdili. Uzimajući u obzir točnosti predviđanja unutar MG populacije uočljivo je da je točnost predviđanja za neka svojstva (npr. TW) u jednom okolišu (OS10) umjerena, dok je u svim ostalim okolišima niska. Sličan obrazac može se uočiti i za MTW čija je točnost predviđanja u okolišu OS10 niska, dok je u svim ostalim okolišima umjerena do visoka, te za WGC i MPT čija je točnost predviđanja znatno niža u okolišima SB09 i OS09. Uspoređujući ove rezultate s rezultatima GEI analize za ista svojstva (znanstveni rad 1), može se uočiti da su okoliši u kojima se postižu neuobičajeno visoke ili niske vrijednosti točnosti predviđanja u usporedbi s ostalim okolišima obično one koje proizvode najveću GEI. Najjasniji primjer za to je TW čija je točnost predviđanja najviša u okolišu OS10, a koji je na AMMI2 biplotu bio jedini okoliš koji nije bila grupiran zajedno s ostalim okolišima. Kako bi se riješila problematika prisutnosti GEI i postigle veće točnosti predviđanja neka istraživanja predlažu uključivanje GEI u model za genomsku selekciju (Jarquín i sur., 2017; Lado i sur., 2016; Lopez-Cruz i sur., 2015) ili uključivanje informacija iz visokokoreliranih okoliša u model (Burgueño i sur., 2012; Ornella i sur., 2012). Prema Ornella i sur. (2012) ako su dva okoliša visokokorelirana moguće je predvidjeti uspješnost genotipova unutar jednog okoliša na temelju modela utreniranog korištenjem podataka iz drugog okoliša. Osim toga, identificiranje okoliša koji pokazuju manju točnost predviđanja, i njihovo uklanjanje iz skupa podataka korištenog za treniranje modela, pokazalo se kao dobar pristup za postizanje viših vrijednosti točnosti predviđanja (Heslot i sur., 2013; Michel i sur., 2016). Ovakav pristup mogao bi biti uspješan za svojstva WGC, MPT i MTW iz ovog istraživanja, a za koja je utvrđena niska točnost predviđanja u samo jednom od okoliša dok je u svim drugim okolišima točnost predviđanja bila umjerena.

Rezultati koji proizlaze iz znanstvenog rada 3, a odnose se na predvidljivost različitih svojstava kakvoće i utjecaj odabranog modela na točnost predviđanja, djelomično potvrđuju hipotezu 1 ovog doktorskog rada. Točnost predviđanja dobivena korištenjem različitih modela varirala je u ovisnosti o okolišu, ali je općenito bila srednja do srednje-visoka za

većinu istraživanih svojstava, dok je za jedno od reoloških svojstava (MTI) pokazala izrazito niske, te u nekim okolišima i negativne vrijednosti.

4. ZAKLJUČCI

- Primjenom AMMI modela za analizu GEI utvrđeno je da je za GPC, WGC i TW u obje populacije dominantan izvor varijabilnosti bio okoliš. GEI je imala znatno važniji učinak na svojstva miksografa u odnosu na preostala tri promatrana svojstva kakvoće. Dominantan učinak GEI zabilježen je za MPT i MTW u BK populaciji te za MTI i MPH u MG populaciji.
- Utvrđeno je postojanje nekoliko široko i specifično prilagođenih RIL-ova, a koji pokazuju potencijal za korištenje u oplemenjivačkim programima usmjerenima na poboljšanje kakvoće. Genotip MG124 primjer je široko prilagođenog RIL-a.
- Heritabilnost u širem smislu za sva promatrana svojstva kakvoće bila je visoka, uz iznimku svojstva MPT u BK populaciji čija je heritabilnost bila umjerena. Postignute vrijednosti heritabilnosti ne predstavljaju ograničenje za postizanje dobre točnosti predviđanja korištenjem genomske selekcije.
- Smanjenje veličine trenažne populacije imalo je negativan utjecaj na točnost predviđanja u svim promatranim slučajevima. Nije utvrđena potreba za optimizacijom trenažne populacije na osnovu fenotipske varijance kako bi se postigla veća točnost predviđanja za svojstva kakvoće pšenice.
- Povećanje gustoće biljega imalo je pozitivan utjecaj na točnost predviđanja u svim promatranim slučajevima, iako razlike u dobivenoj točnosti predviđanja korištenjem manje i veće gustoće biljega nisu bile velike.
- Najniže vrijednosti točnosti predviđanja dobivene su korištenjem EN modela, dok su bayesovski modeli bili najuspješniji u predviđanju svojstava kakvoće. Razlike u točnosti predviđanja između najuspješnijeg modela za pojedino svojstvo i RR-BLUP modela nisu bile velike, stoga RR-BLUP model predstavlja najbolji izbor budući da je računalno znatno manje zahtjevan od bayesovskih modela.
- Točnost predviđanja dobivena korištenjem različitih modela varirala je u ovisnosti o okolišu, ali je općenito bila srednja do srednje-visoka za većinu istraživanih svojstava, dok je za jedno od reoloških svojstava (MTI) pokazala izrazito niske te u nekim okolišima i negativne vrijednosti.
- Dobivene točnosti predviđanja podržavaju primjenu genomske selekcije za svojstva kakvoće pšenice, uključujući i neka reološka svojstva dobivena miksograf uređajem, uz iznimku MTI svojstva koje se nije pokazalo kao dobro ciljano svojstvo za genomsku selekciju.

5. POPIS LITERATURE

1. Adamski N.M., Borrill P., Brinton J., Harrington S.A., Marchal C., Bentley A.R., Bovill W.D., Cattivelli L., Cockram J., Contreras-Moreira B., Ford B., Ghosh S., Harwood W., Hassani-Pak K., Hayta S., Hickey L.T., Kanyuka K., King J., Maccaferrri M., Naamati G., Pozniak C.J., Ramirez-Gonzalez R.H., Sansaloni C., Trevaskis B., Wingen L.U., Wulff B.B.H., Uauy C. (2020). A roadmap for gene functional characterisation in crops with large genomes: Lessons from polyploid wheat. Elife 9: e55646. doi:10.7554/eLife.55646

2. Albrecht T., Auinger H.J., Wimmer V., Ogutu J.O., Knaak C., Ouzunova M., Piepho H.P., Schön C.C. (2014). Genome-based prediction of maize hybrid performance across genetic groups, testers, locations, and years. Theor Appl Genet 127 (6): 1375–1386. doi:10.1007/s00122-014-2305-z

3. Annicchiarico P., Perenzin M. (1994). Adaptation patterns and definition of macro-environments for selection and recommendation of common-wheat genotypes in Italy. Plant Breed 113 (3): 197–205. doi:10.1111/j.1439-0523.1994.tb00723.x

4. Arciniegas-Alarcón S., García-Peña M., Canas Rodrigues P. (2020). New multiple imputation methods for genotype-by-environment data that combine singular value decomposition and Jackknife resampling or weighting schemes. Comput Electron Agric 176: 105617. doi:10.1016/j.compag.2020.105617

5. Arruda M.P., Brown P.J., Lipka A.E., Krill A.M., Thurber C., Kolb F.L. (2015). Genomic selection for predicting *Fusarium* head blight resistance in a wheat breeding program. Plant Genome 8 (3): eplantgenome2015.01.0003. doi:10.3835/plantgenome2015.01.0003

6. Asoro F.G., Newell M.A., Beavis W.D., Scott M.P., Jannink J.-L. (2011). Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. Plant Genome J 4 (2): 132–144. doi:10.3835/plantgenome2011.02.0007

7. Baker R.J. (1988). Tests for crossover genotype-environmental interactions. Can J Plant Sci 68: 405–410.

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

9. Battenfield S.D., Guzmán C., Gaynor R.C., Singh R.P., Peña R.J., Dreisigacker S., Fritz A.K., Poland J.A. (2016). Genomic selection for processing and end-use quality traits in the CIMMYT spring bread wheat breeding program. Plant Genome 9 (2): plantgenome2016.01.0005. doi:10.3835/plantgenome2016.01.0005

10. Bedo Z., Láng L., Rakszegi M. (2017). Breeding for grain-quality traits. U: Cereal Grains: Assessing and Managing Quality (ur. Wrigley C., Batey I., Miskelly D.), Woodhead Publishing, str. 425–452. doi:10.1016/B978-0-08-100719-8/00016-4

11. Belamkar V., Guttieri M.J., Hussain W., Jarquín D., El-basyoni I., Poland J., Lorenz A.J., Baenziger P.S. (2018). Genomic selection in preliminary yield trials in a winter wheat breeding program. G3 Genes, Genomes, Genet 8 (8): 2735–2747. doi:10.1534/g3.118.200415

12. Bernardo R. (2010). Genotype x environment interaction. U: Breeding for Quantitative Traits in Plants, Stemma Press, Woodbury, Minnesota, USA, str. 422.

13. Bernardo R., Yu J. (2007). Prospects for genomewide selection for quantitative traits in maize. Crop Sci 47 (3): 1082–1090. doi:10.2135/cropsci2006.11.0690

14. Bordes J., Ravel C., Le Gouis J., Lapierre A., Charmet G., Balfourier F. (2011). Use of a global wheat core collection for association analysis of flour and dough quality traits. J Cereal Sci 54 (1): 137–147. doi:10.1016/j.jcs.2011.03.004

15. Borojević S. (1981). Principi i metodi oplemenjivanja bilja. Izdavački centar Radničkog univerziteta Radivoj Ćirpanov, Novi Sad, Srbija, 386 str.

16. Bradshaw A.D. (1965). Evolutionary significance of phenotypic plasticity in plants. Adv Genet 13: 115–155. doi:10.1016/S0065-2660(08)60048-6

17. Brancourt-Hulmel M., Denis J.B., Lecomte C. (2000). Determining environmental covariates which explain genotype environment interaction in winter wheat through probe genotypes and biadditive factorial regression. Theor Appl Genet 100 (2): 285–298. doi:10.1007/s001220050038

18. Breiman L. (2001). Random forests. Mach Learn 45: 5–32. doi:10.1007/978-3-030-62008-0_35

19. Burgueño J., de los Campos G., Weigel K., Crossa J. (2012). Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. Crop Sci 52 (2): 707–719. doi:10.2135/cropsci2011.06.0299

20. Bustos-Korts D., Romagosa I., Borràs-Gelonch G., Slafer Gustavo, Eeuwijk F. (2019). Genotype by environment interaction and adaptation. U: Crop Science. Encyclopedia of Sustainability Science and Technology Series (ur. Savin R., Slafer G.), Springer, New York, USA, str. 29–71. doi:10.1007/978-1-4614-5797-8_199

21. Campbell K.G., Finney P.L., Bergman C.J., Gualberto D.G., Anderson J.A., Giroux M.J., Siritunga D., Zhug J., Gendre F., Roué C., Vérel A., Sorrells M.E. (2001). Quantitative trait loci associated with milling and baking quality in a soft x hard wheat cross. Crop Sci 41 (4): 1275–1285. doi:10.2135/cropsci2001.4141275x

22. Charmet G., Storlie E., Oury F.X., Laurent V., Beghin D., Chevarin L., Lapierre A., Perretant M.R., Rolland B., Heumez E., Duchalais L., Goudemand E., Bordes J., Robert O. (2014). Genome-wide prediction of three important traits in bread wheat. Mol Breed 34 (4): 1843–1852. doi:10.1007/s11032-014-0143-y

23. Chung O.K., Ohm J.B., Caley M.S., Seabourn B.W. (2001). Prediction of baking characteristics of hard winter wheat flours using computer-analyzed mixograph parameters. Cereal Chem 78 (4): 493–497. doi:10.1094/CCHEM.2001.78.4.493

24. CIMMYT. (2022). International Maize and Wheat Improvement Center. Dostupno na: https://www.cimmyt.org/work/wheat-research/; pristup: 21.03.2022.

25. Collard B.C.Y., Jahufer M.Z.Z., Brouwer J.B., Pang E.C.K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142: 169–196. doi:10.1007/s10681-005-1681-5

26. Combs E., Bernardo R. (2013). Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers. Plant Genome 6 (1): plantgenome2012.11.0030. doi:10.3835/plantgenome2012.11.0030

27. Crossa J., Beyene Y., Semagn K., Pérez P., Hickey J.M., Chen C., de los Campos G., Burgueño J., Windhausen V.S., Buckler E., Jannink J.L., Cruz M.A.L., Babu R. (2013). Genomic prediction in maize breeding populations with genotyping-by-sequencing. G3 Genes, Genomes, Genet 3 (11): 1903–1926. doi:10.1534/g3.113.008227

28. Crossa J., De Los Campos G., Maccaferri M., Tuberosa R., Burgueño J., Pérez-Rodríguez P. (2016a). Extending the marker × environment interaction model for genomic-enabled prediction and genome-wide association analysis in durum wheat. Crop Sci 56 (5): 2193–2209. doi:10.2135/cropsci2015.04.0260

29. Crossa J., de los Campos G., Pérez P., Gianola D., Burgueño J., Araus J.L., Makumbi D., Singh R.P., Dreisigacker S., Yan J., Arief V., Banziger M., Braun H.J. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186 (2): 713–724. doi:10.1534/genetics.110.118521

30. Crossa J., Fox P.N., Pfeiffer W.H., Rajaram S., Gauch H.G. (1991). AMMI adjustment for statistical analysis of an international wheat yield trial. Theor Appl Genet 81 (1): 27–37. doi:10.1007/BF00226108

31. Crossa J., Jarquín D., Franco J., Pérez-Rodríguez P., Burgueño J., Saint-Pierre C., Vikram P., Sansaloni C., Petroli C., Akdemir D., Sneller C., Reynolds M., Tattaris M., Payne T., Guzman C., Peña R.J., Wenzl P., Singh S. (2016b). Genomic prediction of gene bank wheat landraces. G3 Genes, Genomes, Genet 6 (7): 1819–1834. doi:10.1534/g3.116.029637

32. Daniel C., Triboi E. (2000). Effects of temperature and nitrogen nutrition on the grain composition of winter wheat: Effects on gliadin content and composition. J Cereal Sci 32 (1): 45–56. doi:10.1006/jcrs.2000.0313

33. Dawson J.C., Endelman J.B., Heslot N., Crossa J., Poland J., Dreisigacker S., Manès Y., Sorrells M.E., Jannink J.L. (2013). The use of unbalanced historical data for genomic selection in an international wheat breeding program. F Crop Res 154: 12–22. doi:10.1016/j.fcr.2013.07.020

34. De Leon N., Jannink J.L., Edwards J.W., Kaeppler S.M. (2016). Introduction to a special issue on genotype by environment interaction. Crop Sci 56 (5): 2081–2089. doi:10.2135/cropsci2016.07.0002in

35. De Los Campos G., Gianola D., Rosa G.J.M., Weigel K.A., Crossa J. (2010). Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. Genet Res (Camb) 92 (4): 295–308. doi:10.1017/S0016672310000285

36. Dekkers J.C.M., Hospital F. (2002). The use of molecular genetics in the improvement of agricultural populations. Nat Rev Genet 3 (1): 22–32. doi:10.1038/nrg701

37. Desta Z.A., Ortiz R. (2014). Genomic selection: Genome-wide prediction in plant improvement. Trends Plant Sci 19 (9): 592–601. doi:10.1016/j.tplants.2014.05.006

38. Dickerson G.E. (1962). Implications of genetic-environmental interaction in animal breeding. Anim Prod 4 (1): 47–63. doi:10.1017/S0003356100034395

39. Drezner G., Gunjača J., Novoselović D., Horvat D. (2010). Interpretation of GEI effect analysis for some agronomic and quality traits in ten winter wheat (*Triticum aestivum* L.) cultivars. Cereal Res Commun 38 (2): 259–265. doi:10.1556/crc.38.2010.2.12

40. Dudley J.W. (1993). Molecular markers in plant improvement: Manipulation of genes affecting quantitative traits. Crop Sci 34 (1): 660–668. doi:10.2135/cropsci1994.0011183x003400010094x

41. Echeverry-Solarte M., Kumar A., Kianian S., Simsek S., Alamri M.S., Mantovani E.E., McClean P.E., Deckard E.L., Elias E., Schatz B., Xu S.S., Mergoum M. (2015). New QTL alleles for quality-related traits in spring wheat revealed by RIL population derived from supernumerary × non-supernumerary spikelet genotypes. Theor Appl Genet 128 (5): 893–912. doi:10.1007/s00122-015-2478-0

42. Edwards S.M.K., Buntjer J.B., Jackson R., Bentley A.R., Lage J., Byrne E., Burt C., Jack P., Berry S., Flatman E., Poupard B., Smith S., Hayes C., Gaynor R.C., Gorjanc G., Howell P., Ober E., Mackay I.J., Hickey J.M. (2019). The effects of training population design on genomic prediction accuracy in wheat. Theor Appl Genet 132 (7): 1943–1952. doi:10.1007/s00122-019-03327-y

43. Elangovan M., Dholakia B.B., Rai R., Lagu M.D., Tiwari R., Gupta R.K., Gupta V. (2011). Mapping QTL associated with agronomic traits in bread wheat (*Triticum aestivum* L.). J Wheat Res 3 (1): 14–23.

44. Elangovan M., Rai R., Dholakia B.B., Lagu M.D., Tiwari R., Gupta R.K., Rao V.S., Röder M.S., Gupta V.S. (2008). Molecular genetic mapping of quantitative trait loci associated with loaf volume in hexaploid wheat (*Triticum aestivum*). J Cereal Sci 47 (3): 587–598. doi:10.1016/j.jcs.2007.07.003

45. Elias A.A., Robbins K.R., Doerge R.W., Tuinstra M.R. (2016). Half a century of studying genotype × environment interactions in plant breeding experiments. Crop Sci 56 (5): 2090–2105. doi:10.2135/cropsci2015.01.0061

46. Elli L., Villalta D., Roncoroni L., Barisani D., Ferrero S., Pellegrini N., Bardella M.T., Valiante F., Tomba C., Carroccio A., Bellini M., Soncini M., Cannizzaro R., Leandro G. (2017). Nomenclature and diagnosis of gluten-related disorders: A position statement by the Italian Association of Hospital Gastroenterologists and Endoscopists (AIGO). Dig Liver Dis 49 (2): 138–146. doi:10.1016/j.dld.2016.10.016

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

48. Endelman J.B., Atlin G.N., Beyene Y., Semagn K., Zhang X., Sorrells M.E., Jannink J.-L. (2014). Optimal design of preliminary yield trials with genome-wide markers. Crop Sci 54 (1): 48–59. doi:10.2135/cropsci2013.03.0154

49. FAO. (2020). Food and Agriculture Organization of the United Nations. FAOSTAT Statistical Database. Dostupno na: https://www.fao.org/faostat/en/#data; pristup: 21.03.2022.

50. Finlay K.W., Wilkinson G.N. (1963). The analysis of adaptation in a plant-breeding programme. Aust J Agric Res 14: 742–754.

51. Forkman J., Piepho H.P. (2014). Parametric bootstrap methods for testing multiplicative terms in GGE and AMMI models. Biometrics 70 (3): 639–647. doi:10.1111/biom.12162

52. Friedman J., Hastie T., Tibshirani R. (2010). Regularization paths for generalized linear models via coordinate descent. J Stat Softw 33 (1): 1–22. doi:10.1163/ej.9789004178922.i-328.7

53. Gao H., Su G., Janss L., Zhang Y., Lund M.S. (2013). Model comparison on genomic predictions using high-density markers for different groups of bulls in the Nordic Holstein population. J Dairy Sci 96 (7): 4678–4687. doi:0.3168/jds.2012-6406

54. Gauch H.G. (2013). A simple protocol for AMMI analysis of yield trials. Crop Sci 53 (5): 1860–1869. doi:10.2135/cropsci2013.04.0241

55. Gauch H.G. (2006). Statistical analysis of yield trials by AMMI and GGE. Crop Sci 46 (4): 1488–1500. doi:10.2135/cropsci2005.07-0193

56. Gauch H.G. (1992). Statistical analysis of regional yield trials: AMMI analysis of factorial designs. Elsevier, New York, USA, 278 str.

57. Gauch H.G. (1988). Model selection and validation for yield trials with interaction. Biometrics 44 (3): 705–715.

58. Gauch H.G., Rodrigues P.C., Munkvold J.D., Heffner E.L., Sorrells M. (2011). Two new strategies for detecting and understanding QTL × environment interactions. Crop Sci 51 (1): 96–113. doi:10.2135/cropsci2010.04.0206

59. Gauch H.G., Zobel R.W. (1997). Identifying mega-environments and targeting genotypes. Crop Sci 37 (2): 311–326. doi:10.2135/cropsci1997.0011183X003700020002x

60. Gauch H.G., Zobel R.W. (1990). Imputing missing yield trial data. Theor Appl Genet 79 (6): 753–761. doi:10.1007/BF00224240

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

62. Goel S., Singh K., Singh B., Grewal S., Dwivedi N., Alqarawi A.A., Abd Allah E.F., Ahmad P., Singh N.K. (2019). Analysis of genetic control and QTL mapping of essential wheat grain quality traits in a recombinant inbred population. PLoS One 14 (3): e0200669. doi:10.1371/journal.pone.0200669

63. Gorjanc G., Dumasy J.F., Gonen S., Gaynor R.C., Antolin R., Hickey J.M. (2017). Potential of low-coverage genotyping-by-sequencing and imputation for cost-effective genomic selection in biparental segregating populations. Crop Sci 57 (3): 1404–1420. doi:10.2135/cropsci2016.08.0675

64. Gras P.W., O'Brien L. (1992). Application of a 2-gram mixograph to early generation selection for dough strength. Cereal Chem 69 (3): 254–257.

65. Grausgruber H., Oberforster M., Werteker M., Ruckenbauer P., Vollmann J. (2000). Stability of quality traits in Austrian-grown winter wheats. F Crop Res 66 (3): 257–267. doi:10.1016/S0378-4290(00)00079-4

66. Graybosch R.A., Peterson C.J., Hareland G.A., Shelton D.R., Olewnik M.C., He H., Stearns M.M. (1999). Relationships between small-scale wheat quality assays and commercial test bakes. Cereal Chem 76 (3): 428–433. doi:10.1094/CCHEM.1999.76.3.428

67. Grogan S.M., Anderson J., Stephen Baenziger P., Frels K., Guttieri M.J., Haley S.D., Kim K.S., Liu S., McMaster G.S., Newell M., Prasad P.V.V., Reid S.D., Shroyer K.J., Zhang G., Akhunov E., Byrne P.F. (2016). Phenotypic plasticity of winter wheat heading date and grain yield across the US great plains. Crop Sci 56 (5): 2223–2236. doi:10.2135/cropsci2015.06.0357

68. Groos C., Bervas E., Charmet G. (2004). Genetic analysis of grain protein content, grain hardness and dough rheology in a hard x hard bread wheat progeny. J Cereal Sci 40 (2): 93–100. doi:10.1016/j.jcs.2004.08.006

69. Groos C., Robert N., Bervas E., Charmet G. (2003). Genetic analysis of grain protein-content, grain yield and thousand-kernel weight in bread wheat. Theor Appl Genet 106 (6): 1032–1040. doi:10.1007/s00122-002-1111-1

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

71. Habier D., Fernando R.L., Dekkers J.C.M. (2007). The impact of genetic relationship information on genome-assisted breeding values. Genetics 177 (4): 2389–2397. doi:10.1534/genetics.107.081190

72. Habier D., Fernando R.L., Kizilkaya K., Garrick D.J. (2011). Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics 12: 186. doi:10.1186/1471-2105-12-186

73. Haile J.K., N'Diaye A., Clarke F., Clarke J., Knox R., Rutkoski J., Bassi F.M., Pozniak C.J. (2018). Genomic selection for grain yield and quality traits in durum wheat. Mol Breed 38: 75. doi:10.1007/s11032-018-0818-x

74. Haldane J.B.S. (1947). The interaction of nature and nurture. Ann Eugen 13: 197–205.

75. Hayes B.J., Visscher P.M., Goddard M.E. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. Genet Res (Camb) 91 (1): 47–60. doi:10.1017/S0016672308009981

76. He S., Schulthess A.W., Mirdita V., Zhao Y., Korzun V., Bothe R., Ebmeyer E., Reif J.C., Jiang Y. (2016). Genomic selection in a commercial winter wheat population. Theor Appl Genet 129 (3): 641–651. doi:10.1007/s00122-015-2655-1

77. Heffner E.L. (2011). Predicting genetic value of breeding lines using genomic slection in a winter wheat breeding program. Doktroska disertacija. Cornell University, USA, 234 str.

78. Heffner E.L., Jannink J.-L., Iwata H., Souza E., Sorrells M.E. (2011a). Genomic selection accuracy for grain quality traits in biparental wheat populations. Crop Sci 51: 2597–2606. doi:10.2135/cropsci2011

79. Heffner E.L., Jannink J.-L., Sorrells M.E. (2011b). Genomic selection accuracy using multifamily prediction models in a wheat breeding program. Plant Genome 4 (1): 65–75. doi:10.3835/plantgenome2010.12.0029

80. Heffner E.L., Lorenz A.J., Jannink J.L., Sorrells M.E. (2010). Plant breeding with genomic selection: Gain per unit time and cost. Crop Sci 50 (5): 1681–1690. doi:10.2135/cropsci2009.11.0662

81. Heffner E.L., Sorrells M.E., Jannink J.L. (2009). Genomic selection for crop improvement. Crop Sci 49: 1–12. doi:10.2135/cropsci2008.08.0512

82. Hernández-Espinosa N., Mondal S., Autrique E., Gonzalez-Santoyo H., Crossa J., Huerta-Espino J., Singh R.P., Guzmán C. (2018). Milling, processing and end-use quality traits of CIMMYT spring bread wheat germplasm under drought and heat stress. F Crop Res 215: 104–112. doi:10.1016/j.fcr.2017.10.003

83. Herter C.P., Ebmeyer E., Kollers S., Korzun V., Würschum T., Miedaner T. (2019). Accuracy of within- and among-family genomic prediction for *Fusarium* head blight and *Septoria tritici* blotch in winter wheat. Theor Appl Genet 132 (4): 1121–1135. doi:10.1007/s00122-018-3264-6

84. Heslot N., Akdemir D., Sorrells M.E., Jannink J.L. (2014). Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. Theor Appl Genet 127 (2): 463–480. doi:10.1007/s00122-013-2231-5

85. Heslot N., Jannink J.L., Sorrells M.E. (2013). Using genomic prediction to characterize environments and optimize prediction accuracy in applied breeding data. Crop Sci 53 (3): 921–933. doi:10.2135/cropsci2012.07.0420

86. Heslot N., Yang H.P., Sorrells M.E., Jannink J.L. (2012). Genomic selection in plant breeding: A comparison of models. Crop Sci 52 (1): 146–160. doi:10.2135/cropsci2011.06.0297

87. Hickey J.M., Dreisigacker S., Crossa J., Hearne S., Babu R., Prasanna B.M., Grondona M., Zambelli A., Windhausen V.S., Mathews K., Gorjanc G. (2014). Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. Crop Sci 54 (4): 1476–1488. doi:10.2135/cropsci2013.03.0195

88. Hook S.C.W. (1984). Specific weight and wheat quality. J Sci Food Agric 35 (10): 1136–1141. doi:10.1002/jsfa.2740351013

89. Horvat D., Drezner G., Jurković Z., Šimić G., Magdić D., Dvojković K. (2006). The importance of high-molecular-weight glutenin subunits for wheat quality evaluation. Poljoprivreda 12: 53–57

90. Horvat D., Dvojković K., Novoselović D., Tucak M., Andrić L., Magdić D., Drezner G. (2022). Response of wheat yield and protein-related quality on late-season urea application. Agronomy 12: 886. doi:10.3390/agronomy12040886

91. Horvat D., Šimić G., Dvojković K., Ivić M., Plavšin I., Novoselović D. (2021). Gluten protein compositional changes in response to nitrogen application rate. Agronomy 11: 325. doi:10.3390/agronomy11020325

92. Hu X., Carver B.F., Powers C., Yan L., Zhu L., Chen C. (2019). Effectiveness of genomic selection by response to selection for winter wheat variety improvement. Plant Genome 12 (3): 180090. doi:10.3835/plantgenome2018.11.0090

93. Huang X.Q., Cloutier S., Lycar L., Radovanovic N., Humphreys D.G., Noll J.S., Somers D.J., Brown P.D. (2006). Molecular detection of QTLs for agronomic and quality traits in a doubled haploid population derived from two Canadian wheats (*Triticum aestivum* L.). Theor Appl Genet 113 (4): 753–766. doi:10.1007/s00122-006-0346-7

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

95. Jaccoud D., Peng K., Feinstein D., Kilian A. (2001). Diversity Arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res 29 (4): e25. doi:10.1093/nar/29.4.e25

96. Jannink J.L., Lorenz A.J., Iwata H. (2010). Genomic selection in plant breeding: From theory to practice. Briefings Funct Genomics Proteomics 9 (2): 166–177. doi:10.1093/bfgp/elq001

97. Jarquín D., Lemes da Silva C., Gaynor R.C., Poland J., Fritz A., Howard R., Battenfield S., Crossa J. (2017). Increasing genomic-enabled prediction accuracy by modeling genotype × environment interactions in kansas wheat. Plant Genome 10 (2): /plantgenome2016.12.0130. doi:10.3835/plantgenome2016.12.0130

98. Jin H., Wen W., Liu J., Zhai S., Zhang Y., Yan J., Liu Z., Xia X., He Z. (2016). Genome-wide QTL mapping for wheat processing quality parameters in a Gaocheng

8901/Zhoumai 16 recombinant inbred line population. Front Plant Sci 7: 1032. doi:10.3389/fpls.2016.01032

99. Johnson J.A., Swanson C.O., Bayfield E.G. (1943). The correlation of mixogram with baking results. Cereal Chem 20: 625–644.

100. Juliana P., Poland J., Huerta-Espino J., Shrestha S., Crossa J., Crespo-Herrera L., Toledo F.H., Govindan V., Mondal S., Kumar U., Bhavani S., Singh P.K., Randhawa M.S., He X., Guzman C., Dreisigacker S., Rouse M.N., Jin Y., Pérez-Rodríguez P., Montesinos-López O.A., Singh D., Mokhlesur Rahman M., Marza F., Singh R.P. (2019). Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics. Nat Genet 51 (10): 1530–1539. doi:10.1038/s41588-019-0496-6

101. Kilian A., Huttner E., Wenzl P., Jaccoud D., Carling J., Caig V., Evers M., Heller-Uszynska K., Uszynski G., Cayla C., Patararapuwadol S., Xia L., Yang S., Thomson B. (2005). The fast and the cheap : SNP and DArT-based whole genome profiling for crop improvement. U: In the Wake of the Double Helix: From the Green Revolution to the Gene Revolution, Bologna, Italy, str. 443–461.

102. Kolster P. (1992). High molecular weight glutenin subunits of wheat: Qualitative and quantitative variation in relation to breadmaking quality. University of Wageningen, Netherlands, 135 str.

103. Kovačević V., Rastija M. (2014). Žitarice. Sveučilište J. J. Strossmayera u Osijeku, Poljoprivredni fakultet, Osijek, Hrvatska, 235 str.

104. Krishnappa G., Ahlawat A.K., Shukla R.B., Singh S.K., Singh S.K., Singh A.M., Singh G.P. (2019). Multi-environment analysis of grain quality traits in recombinant inbred lines of a biparental cross in bread wheat (*Triticum aestivum* L.). Cereal Res Commun 47 (2): 334–344. doi:10.1556/0806.47.2019.02

105. Krishnappa G., Savadi S., Tyagi B.S., Singh S.K., Mamrutha H.M., Kumar S., Mishra C.N., Khan H., Gangadhara K., Uday G., Singh G., Singh G.P. (2021). Integrated genomic selection for rapid improvement of crops. Genomics 113 (3): 1070–1086. doi:10.1016/j.ygeno.2021.02.007

106. Kristensen P.S., Jahoor A., Andersen J.R., Cericola F., Orabi J., Janss L.L., Jensen J. (2018). Genome-wide association studies and comparison of models and cross-validation strategies for genomic prediction of quality traits in advanced winter wheat breeding lines. Front Plant Sci 9: 69. doi:10.3389/fpls.2018.00069

107. Kristensen P.S., Jensen J., Andersen J.R., Guzmán C., Orabi J., Jahoor A. (2019). Genomic prediction and genome-wide association studies of flour yield and alveograph quality traits using advanced winter wheat breeding material. Genes (Basel) 10 (9): 669. doi:10.3390/genes10090669

108. Lado B., Barrios P.G., Quincke M., Silva P., Gutiérrez L. (2016). Modeling genotype × environment interaction for genomic selection with unbalanced data from a wheat breeding program. Crop Sci 56 (5): 2165–2179. doi:10.2135/cropsci2015.04.0207

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

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

111. Larkin D.L., Lozada D.N., Mason R.E. (2019). Genomic selection - Considerations for successful implementation in wheat breeding programs. Agronomy 9: 479. doi:10.3390/agronomy9090479

112. Létang C., Piau M., Verdier C. (1999). Characterization of wheat flour-water doughs. Part I: Rheometry and microstructure. J Food Eng 41 (2): 121–132. doi:10.1016/S0260-8774(99)00082-5

113. Liu G., Zhao Y., Gowda M., Longin C.F.H., Reif J.C., Mette M.F. (2016). Predicting hybrid performances for quality traits through genomic-assisted approaches in Central European wheat. PLoS One 11 (7): e0158635. doi:10.1371/journal.pone.0158635

114. Liu X., Wang Hongwu, Wang Hui, Guo Z., Xu X., Liu J., Wang S., Li W.X., Zou C., Prasanna B.M., Olsen M.S., Huang C., Xu Y. (2018). Factors affecting genomic selection revealed by empirical evidence in maize. Crop J 6 (4): 341–352. doi:10.1016/j.cj.2018.03.005

115. Lopez-Cruz M., Crossa J., Bonnett D., Dreisigacker S., Poland J., Jannink J.L., Singh R.P., Autrique E., de los Campos G. (2015). Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. G3 Genes, Genomes, Genet 5 (4): 569–582. doi:10.1534/g3.114.016097

116. Lorenz A.J. (2013). Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: A simulation experiment. G3 Genes, Genomes, Genet 3 (3): 481–491. doi:10.1534/g3.112.004911

117. Lorenz A.J., Chao S., Asoro F.G., Heffner E.L., Hayashi T., Iwata H., Smith K.P., Sorrells M.E., Jannink J.L. (2011). Genomic selection in plant breeding: Knowledge and prospects. Adv Agron 110: 77–123. doi:10.1016/B978-0-12-385531-2.00002-5

118. Lorenz A.J., Smith K.P. (2015). Adding genetically distant individuals to training populations reduces genomic prediction accuracy in barley. Crop Sci 55 (6): 2657–2667. doi:10.2135/cropsci2014.12.0827

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

120. Lozada D.N., Mason R.E., Sarinelli J.M., Brown-Guedira G. (2019). Accuracy of genomic selection for grain yield and agronomic traits in soft red winter wheat. BMC Genet 20: 82. doi:10.1186/s12863-019-0785-1

121. Ly D., Hamblin M., Rabbi I., Melaku G., Bakare M., Gauch H.G., Okechukwu R., Dixon A.G.O., Kulakow P., Jannink J.L. (2013). Relatedness and genotype × environment interaction affect prediction accuracies in genomic selection: A study in cassava. Crop Sci 53 (4): 1312–1325. doi:10.2135/cropsci2012.11.0653

122. Macritchie F. (1992). Physicochemical properties of wheat proteins in relation to functionality. Adv Food Nutr Res 36: 1–87. doi:10.1016/S1043-4526(08)60104-7

123. Mani E. (2007). Molecular dissection of breadmaking quality in wheat (*Triticum aestivum* L.). Doktroska disertacija. University of Pune, India, 113 str.

124. Martinant J.P., Nicolas Y., Bouguennec A., Popineau Y., Saulnier L., Branlard G. (1998). Relationships between mixograph parameters and indices of wheat grain quality. J Cereal Sci 27 (2): 179–189. doi:10.1006/jcrs.1997.0156

125. Martinez-Perez E., Shaw P., Moore G. (2001). The *Ph1* locus is needed to ensure specifc somatic and meiotic centromere association. Nature 411: 204–207. doi:10.1088/1742-6596/678/1/012022

126. Marulanda J.J., Melchinger A.E., Würschum T. (2015). Genomic selection in biparental populations: Assessment of parameters for optimum estimation set design. Plant Breed 134 (6): 623–630. doi:10.1111/pbr.12317

127. Matsuoka Y. (2011). Evolution of polyploid *Triticum* wheats under cultivation: The role of domestication, natural hybridization and allopolyploid speciation in their diversification. Plant Cell Physiol 52 (5): 750–764. doi:10.1093/pcp/pcr018

128. Maulana F., Kim K.-S., Anderson J.D., Sorrells M.E., Butler T.J., Liu S., Baenziger P.S., Byrne P.F., Ma X.-F. (2019). Genomic selection of forage quality traits in winter wheat. Crop Sci 59: 1–11. doi:10.2135/cropsci2018.10.0655

129. Meuwissen T.H.E., Hayes B.J., Goddard M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819–1829. doi:10.1016/j.aquaculture.2016.04.008

130. Michel S., Ametz C., Gungor H., Epure D., Grausgruber H., Löschenberger F., Buerstmayr H. (2016). Genomic selection across multiple breeding cycles in applied bread wheat breeding. Theor Appl Genet 129 (6): 1179–1189. doi:10.1007/s00122-016-2694-2

131. Michel S., Gallee M., Löschenberger F., Buerstmayr H., Kummer C. (2017). Improving the baking quality of bread wheat using rapid tests and genomics: The prediction of dough rheological parameters by gluten peak indices and genomic selection models. J Cereal Sci 77: 24–34. doi:10.1016/j.jcs.2017.07.012

132. Michel S., Kummer C., Gallee M., Hellinger J., Ametz C., Akgöl B., Epure D., Löschenberger F., Buerstmayr H. (2018). Improving the baking quality of bread wheat by genomic selection in early generations. Theor Appl Genet 131 (2): 477–493. doi:10.1007/s00122-017-2998-x

133. Mohan M., Nair S., Bhagwat A., Krishna T.G., Yano M., Bhatia C.R., Sasaki T. (1997). Genome mapping, molecular markers and marker-assisted selection in crop plants. Mol Breed 3: 87–103. doi:10.1007/978-981-10-2961-5_12

134. Montenegro J.D., Golicz A.A., Bayer P.E., Hurgobin B., Lee H.T., Chan C.K.K., Visendi P., Lai K., Doležel J., Batley J., Edwards D. (2017). The pangenome of hexaploid bread wheat. Plant J 90 (5): 1007–1013. doi:10.1111/tpj.13515

135. Mut Z., Aydin N., Orhan Bayramoglu H., Ozean H. (2010). Stability of some quality traits in bread wheat (*Triticum aestivum*) genotypes. J Environ Biol 31 (4): 489–495.

136. Nuttonson M.Y. (1955). Wheat-climate relationships and the use of phenology in ascertaining the thermal and photo-thermal requirements of wheat based on data of North America and of some thermally analogous areas of North America in the Soviet

Union and in Finland. American Institute of Crop Ecology, Washington, D.C., USA, 388 str.

137. Ohm J.B., Chung O.K. (1999). Gluten, pasting, and mixograph parameters of hard winter wheat flours in relation to breadmaking. Cereal Chem 76 (5): 606–613. doi:10.1094/CCHEM.1999.76.5.606

138. Ornella L., Sukhwinder-Singh, Perez P., Burgueño J., Singh R., Tapia E., Bhavani S., Dreisigacker S., Braun H.J., Mathews K., Crossa J. (2012). Genomic prediction of genetic values for resistance to wheat rusts. Plant Genome 5 (3): 136–148. doi:10.3835/plantgenome2012.07.0017

139. Paderewski J. (2013). An R function for imputation of missing cells in two-way data sets by EM-AMMI algorithm. Commun Biometry Crop Sci 8 (2): 60–69.

140. Paderewski J., Rodrigues P.C. (2014). The usefulness of EM-AMMI to study the influence of missing data pattern and application to Polish post-registration winter wheat data. Aust J Crop Sci 8 (4): 640–645.

141. Park T., Casella G. (2008). The Bayesian Lasso. J Am Stat Assoc 103 (482): 681–686. doi:10.1198/016214508000000337

142. Payne P.I. (1987). Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. Annu Rev Plant Physiol 38: 141–153.

143. Payne P.I., Nightingale M.A., Krattiger A.F., Holt L.M. (1987). The relationship between HMW glutenin subunit composition and the bread-making quality of british-grown wheat varieties. J Sci Food Agric 40: 51–65.

144. Petersen G., Seberg O., Yde M., Berthelsen K. (2006). Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). Mol Phylogenet Evol 39 (1): 70–82. doi:10.1016/j.ympev.2006.01.023

145. Peterson C.J., Graybosch R.A., Shelton D.R., Baenziger P.S. (1998). Baking quality of hard winter wheat: Response of cultivars to environment in the Great Plains. Euphytica 100: 157–162. doi:10.1007/978-94-011-4896-2_30

146. Piepho H.P. (1993). Robustness of statistical tests for multiplicative terms in the additive main effects and multiplicative interaction model for cultivar trials. Crop Sci 33: 1186–1193. doi:10.1007/BF00221987

147. Plavšin I., Gunjača J., Šatović Z., Šarčević H., Ivić M., Dvojković K., Novoselović D. (2021). An overview of key factors affecting genomic selection for wheat quality traits. Plants 10 (4): 745. doi:10.3390/plants10040745

148. Poland J.A., Brown P.J., Sorrells M.E., 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 (2): e32253. doi:10.1371/journal.pone.0032253

149. Prashant R., Mani E., Rai R., Gupta R.K., Tiwari R., Dholakia B., Oak M., Röder M., Kadoo N., Gupta V. (2015). Genotype × environment interactions and QTL clusters underlying dough rheology traits in *Triticum aestivum* L. J Cereal Sci 64: 82–91. doi:10.1016/j.jcs.2015.05.002

150. Rasheed A., Xia X., Yan Y., Appels R., Mahmood T., He Z. (2014). Wheat seed storage proteins: Advances in molecular genetics, diversity and breeding applications. J Cereal Sci 60 (1): 11–24. doi:10.1016/J.JCS.2014.01.020

151. Riedelsheimer C., Endelman J.B., Stange M., Sorrells M.E., Jannink J.L., Melchinger A.E. (2013). Genomic predictability of interconnected biparental maize populations. Genetics 194 (2): 493–503. doi:10.1534/genetics.113.150227

152. Riedelsheimer C., Melchinger A.E. (2013). Optimizing the allocation of resources for genomic selection in one breeding cycle. Theor Appl Genet 126 (11): 2835–2848. doi:10.1007/s00122-013-2175-9

153. Rincent R., Laloë D., Nicolas S., Altmann T., Brunel D., Revilla P., Rodríguez V.M., Moreno-Gonzalez J., Melchinger A., Bauer E., Schoen C.C., Meyer N., Giauffret C., Bauland C., Jamin P., Laborde J., Monod H., Flament P., Charcosset A., Moreau L. (2012). Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: Comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.). Genetics 192 (2): 715–728. doi:10.1534/genetics.112.141473

154. Robert N., Denis J.B. (1996). Stability of baking quality in bread wheat using several statistical parameters. Theor Appl Genet 93: 172–178. doi:10.1007/BF00225742

155. Robertsen C., Hjortshøj R., Janss L. (2019). Genomic selection in cereal breeding. Agronomy 9 (2): 95. doi:10.3390/agronomy9020095

156. Rodriguez M., Rau D., Papa R., Attene G. (2008). Genotype by environment interactions in barley (*Hordeum vulgare* L.): Different responses of landraces, recombinant inbred lines and varieties to Mediterranean environment. Euphytica 163 (2): 231–247. doi:10.1007/s10681-007-9635-8

157. Rutkoski J., Benson J., Jia Y., Brown-Guedira G., Jannink J.L., Sorrells M. (2012). Evaluation of genomic prediction methods for *Fusarium* head blight resistance in wheat. Plant Genome 5 (2): 51–61. doi:10.3835/plantgenome2012.02.0001

158. Rutkoski J.E., Poland J.A., Singh R.P., Huerta-Espino J., Bhavani S., Barbier H., Rouse M.N., Jannink J.L., Sorrells M.E. (2014). Genomic selection for quantitative adult plant stem rust resistance in wheat. Plant Genome 7 (3): plantgenome2014.02.0006. doi:10.3835/plantgenome2014.02.0006

159. Rutkoski J.E., Singh R.P., Huerta-Espino J., Bhavani S., Poland J., Jannink J.-L., Sorrells M.E. (2015a). Efficient use of historical data for genomic selection: A case study of stem rust resistance in wheat. Plant Genome 8 (1): plantgenome2014.09.0046. doi:10.3835/plantgenome2014.09.0046

160. Rutkoski J.E., Singh R.P., Huerta-Espino J., Bhavani S., Poland J.A., Jannink J.-L., Sorrells M.E. (2015b). Genetic gain from phenotypic and genomic selection for quantitative resistance to stem rust of wheat. Plant Genome 8 (2): plantgenome2014.10.0074. doi:10.3835/plantgenome2014.10.0074

161. Sandhu K.S., Aoun M., Morris C.F., Carter A.H. (2021a). Genomic selection for end-use quality and processing traits in soft white winter wheat breeding program with machine and deep learning models. Biology (Basel) 10 (7): 689. doi:10.3390/biology10070689

162. Sandhu K.S., Lozada D.N., Zhang Z., Pumphrey M.O., Carter A.H. (2021b). Deep learning for predicting complex traits in spring wheat breeding program. Front Plant Sci 11: 613325. doi:10.3389/fpls.2020.613325

163. Shewry P.R. (2004). Improving the protein content and quality of temperate cereals: wheat, barley and rye. U: Impacts of Agriculture on Human Health and Nutrition (ur. Cakmak I., Welch R.), USDA, ARS, U.S. Plant, Soil and Nutrition Laboratory, Cornell University, USA, str. 118–137.

164. Shewry P.R., Hey S.J. (2015). The contribution of wheat to human diet and health. Food Energy Secur 4 (3): 178–202. doi:10.1002/FES3.64

165. Shewry P.R., Tatham A.S., Barro F., Barcelo P., Lazzeri P. (1995). Biotechnology of breadmaking: Unraveling and manipulating the multi-protein gluten complex. Bio/Technology 13 (11): 1185–1190. doi:10.1038/nbt1195-1185

166. Šimić G., Horvat D., Jurković Z., Drezner G., Novoselović D., Šimić G., Horvat D., Jurković Z., Drezner G., Novoselović D., Dvojković K. (2006). The genotype effect on

the ratio of wet gluten content to total wheat grain protein. J Cent Eur Agric 7 (1): 13–18. doi:10.5513/jcea.v7i1.350

167. Simmonds N.W. (1995). The relation between yield and protein in cereal grain. J Sci Food Agric 67 (3): 309–315. doi:10.1002/jsfa.2740670306

168. Simmonds N.W. (1981). Genotype (G), environment (E) and GE components of crop yields. Exp Agric 17 (4): 355–362. doi:10.1017/S0014479700011807

169. Singh A., Ganapathysubramanian B., Singh A.K., Sarkar S. (2016). Machine learning for high-throughput stress phenotyping in plants. Trends Plant Sci 21 (2): 110–124. doi:10.1016/j.tplants.2015.10.015

170. Singh B.D., Singh A.K. (2015). Marker-assisted plant breeding: Principles and practices. Springer, New Delhi, India, 514 str. doi:10.1007/978-81-322-2316-0

171. Sorrells M.E. (2015). Genomic selection in plants: Empirical results and implications for wheat breeding. U: Advances in Wheat Genetics: From Genome to Field (ur. Ogihara Y., Takumi S., Handa H.), Springer Japan KK, Yokohama, Japan, str. 401–409. doi:10.1007/978-4-431-55675-6_36

172. Storlie E., Charmet G. (2013). Genomic selection accuracy using historical data generated in a wheat breeding program. Plant Genome 6 (1): plantgenome2013.01.0001. doi:10.3835/plantgenome2013.01.0001

173. Swanson C.O., Working E.B. (1933). Testing of the quality of flour by the recording dough mixer. Cereal Chem 10: 1–29.

174. Tan B., Grattapaglia D., Martins G.S., Ferreira K.Z., Sundberg B., Ingvarsson P.K. (2017). Evaluating the accuracy of genomic prediction of growth and wood traits in two *Eucalyptus* species and their F1 hybrids. BMC Plant Biol 17: 110. doi:10.1186/s12870-017-1059-6

175. The International Wheat Genome Sequencing Consortium (IWGSC). (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361 (6403): eaar7191. doi:10.1126/science.aar7191

176. The International Wheat Genome Sequencing Consortium (IWGSC). (2014). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345 (6194): 1251788. doi:10.1126/science.1251788

177. Tsai H.Y., Janss L.L., Andersen J.R., Orabi J., Jensen J.D., 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 (1): 3347. doi:10.1038/s41598-020-60203-2

178. Turner A.S., Bradburne R.P., Fish L., Snape J.W. (2004). New quantitative trait loci influencing grain texture and protein content in bread wheat. J Cereal Sci 40 (1): 51–60. doi:10.1016/j.jcs.2004.03.001

179. van Eeuwijk F.A., Bustos-Korts D. V., Malosetti M. (2016). What should students in plant breeding know about the statistical aspects of genotype × environment interactions? Crop Sci 56 (5): 2119–2140. doi:10.2135/cropsci2015.06.0375

180. Varshney R.K., Roorkiwal M., Sorrells M.E. (2017). Genomic selection for crop improvement: An introduction. U: Genomic Selection for Crop Improvement: New Molecular Breeding Strategies for Crop Improvement (ur. Varshney R.K., Roorkiwal M., Sorrells M.E.), Springer, str. 1–258. doi:10.1007/978-3-319-63170-7_1

181. Verges V.L., van Sanford D.A. (2020). Genomic selection at preliminary yield trial stage: Training population design to predict untested lines. Agronomy 10 (1): 60. doi:10.3390/agronomy10010060

182. Voss-Fels K.P., Cooper M., Hayes B.J. (2019). Accelerating crop genetic gains with genomic selection. Theor Appl Genet 132 (3): 669–686. doi:10.1007/s00122-018-3270-8

183. Wang X., Xu Y., Hu Z., Xu C. (2018). Genomic selection methods for crop improvement: Current status and prospects. Crop J 6 (4): 330–340. doi:10.1016/j.cj.2018.03.001

184. Ward B.P., Brown-Guedira G., Tyagi P., Kolb F.L., van Sanford D.A., Sneller C.H., Griffey C.A. (2019). Multienvironment and multitrait genomic selection models in unbalanced early-generation wheat yield trials. Crop Sci 59 (2): 491–507. doi:10.2135/cropsci2018.03.0189

185. Ward J., Rakszegi M., Bedo Z., Shewry P.R., Mackay I. (2015). Differentially penalized regression to predict agronomic traits from metabolites and markers in wheat. BMC Genet 16: 19. doi:10.1186/s12863-015-0169-0

186. Weegels P.L., Hamer R.J., Schofield J.D. (1996). Functional properties of wheat glutenin. J Cereal Sci 23 (1): 1–18. doi:10.1006/jcrs.1996.0001

187. Wieser H. (2007). Chemistry of gluten proteins. Food Microbiol 24 (2): 115–119. doi:10.1016/j.fm.2006.07.004

188. Williams R.M., O'Brien L., Eagles H.A., Solah V.A., Jayasena V. (2008). The influences of genotype, environment, and genotype x environment interaction on wheat quality. Aust J Agric Res 59 (2): 95–111. doi:10.1071/AR07185

189. Windhausen V.S., Atlin G.N., Hickey J.M., Crossa J., Jannink J.-L.L., Sorrells M.E., Raman B., Cairns J.E., Tarekegne A., Semagn K., Beyene Y., Grudloyma P., Technow F., Riedelsheimer C., Melchinger A.E. (2012). Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. G3 Genes, Genomes, Genet 2 (11): 1427–1436. doi:10.1534/g3.112.003699

190. Yang R.C., Crossa J., Cornelius P.L., Burgueño J. (2009). Biplot analysis of genotype × environment interaction: Proceed with caution. Crop Sci 49 (5): 1564–1576. doi:10.2135/cropsci2008.11.0665

191. Yao J., Zhao D., Chen X., Zhang Y., Wang J. (2018). Use of genomic selection and breeding simulation in cross prediction for improvement of yield and quality in wheat (*Triticum aestivum* L.). Crop J 6 (4): 353–365. doi:10.1016/j.cj.2018.05.003

192. Zhang X., Pérez-Rodríguez P., Semagn K., Beyene Y., Babu R., López-Cruz M.A., San Vicente F., Olsen M., Buckler E., Jannink J.-L.L., Prasanna B.M., Crossa J. (2015). Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs. Heredity (Edinb) 114: 291–299. doi:10.1038/hdy.2014.99

193. Zhang Yelun, Wu Y., Xiao Y., Yan J., Zhang Yong, Zhang Yan, Ma C., Xia X., He Z. (2009). QTL mapping for milling, gluten quality, and flour pasting properties in a recombinant inbred line population derived from a Chinese soft × hard wheat cross. Crop Pasture Sci 60 (6): 587–597. doi:10.1071/CP08392

194. Zhao Y., Mette M.F., Gowda M., Longin C.F.H., Reif J.C. (2014). Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. Heredity (Edinb) 112 (6): 638–645. doi:10.1038/hdy.2014.1

195. Zheng F., Deng Z. ying, Shi C. lan, Zhang X. ye, Tian J. chun. (2013). QTL mapping for dough mixing characteristics in a recombinant inbred population derived from a waxy × strong gluten wheat (*Triticum aestivum* L.). J Integr Agric 12 (6): 951–961. doi:10.1016/S2095-3119(13)60472-4

196. Zhong S., Dekkers J.C.M., Fernando R.L., Jannink J.L. (2009). Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: A barley case study. Genetics 182: 355–364. doi:10.1534/g3.118.200740

197. Zörb C., Ludewig U., Hawkesford M.J. (2018). Perspective on wheat yield and quality with reduced nitrogen supply. Trends Plant Sci 23 (11): 1029–1037. doi:10.1016/J.TPLANTS.2018.08.012

198. Zou H., Hastie T. (2005). Regularization and variable selection via the elastic net. J R Stat Soc Ser B Stat Methodol 67 (2): 301–320. doi:10.1111/j.1467-9868.2005.00503.x

ŽIVOTOPIS AUTORA

Ivana Plavšin rođena je u Osijeku 16. rujna 1990. godine. Godine 2009. završila je III. gimnaziju u Osijeku prirodoslovno-matematičkog obrazovnog programa, te upisala preddiplomski sveučilišni studij Biologija na Odjelu za biologiju Sveučilišta J. J. Strossmayera u Osijeku. Akademski naziv sveučilišne prvostupnice biologije stekla je 2012. godine i upisala diplomski sveučilišni studij Biologija, smjer znanstveni, također na Odjelu za biologiju u Osijeku. Godine 2014. upisala je paralelni diplomski studij Ekološka poljoprivreda na Poljoprivrednom fakultetu u Osijeku. Akademski naziv magistra biologije stekla je 2015., a godinu kasnije i akademski naziv magistra inženjerka ekološke poljoprivrede. Dobitnica je dvije Rektorove nagrade (2013. i 2016. godine), dvije Pročelničke nagrade za najboljeg studenta (2011. i 2013. godine) i jedne Dekanove nagrade (2015. godine). Tijekom diplomskog studija uključila se u znanstveno-istraživački rad na Odjelu za biologiju u Osijeku pod mentorstvom prof. dr. sc. Branimira K. Hackenbergera što je rezultiralo sudjelovanjem na nekoliko međunarodnih konferencija, objavom tri znanstvena rada u časopisima indeksiranim u Current Contents bazi i jednog poglavlja u knjizi. Nakon završetka studija odradila je stručno osposobljavanje u Hrvatskoj agenciji za hranu u Osijeku te bila zaposlena kao analitičar u laboratoriju za glikobiologiju u tvrtci Genos d.o.o.

Na Odjelu za oplemenjivanje i genetiku strnih žitarica Poljoprivrednog instituta Osijek zaposlila se u ožujku 2018. godine kao doktorandica na projektu „Bioraznolikost i molekularno oplemenjivanje bilja“ Znanstvenog centra izvrsnosti za bioraznolikost i molekularno oplemenjivanje bilja čiji je voditelj prof. dr. sc. Zlatko Šatović. Iste godine upisala je i poslijediplomski doktorski studij Poljoprivredne znanosti na Agronomskom fakultetu u Zagrebu. Tijekom 2020. položila je i posljednje ispite na doktorskom studiju s ukupnim prosjekom ocjena 5,00.

U periodu od 2018. do 2021. godine kao suradnik je bila uključena u rad istraživačkog projekta „Genetsko poboljšanje i optimizacija potencijala rodnosti pšenice“ Hrvatske zaklade za znanost voditelja dr. sc. Darija Novoselovića. Za vrijeme doktorskog studija započela je rad na oplemenjivačkom programu pšenice pod vodstvom mentora dr. sc. Darija Novoselovića te sudjelovala u razvoju sorte ozime pšenice „Barba“ priznate 2020. godine od Ministarstva poljoprivrede Republike Hrvatske. Tijekom 2018. provela je tri mjeseca na usavršavanju iz područja spektroskopskih metoda korištenih u fenotipizaciji biljaka na University of Applied Sciences and Arts Northwestern Switzerland u mjestu Windisch, Švicarska. Godine 2022. boravila je tri mjeseca na usavršavanju iz područja bolesti strnih žitarica i visokopropusne fenotipizacije na Julius Kühn institutu u mjestu Quedlinburg, Njemačka. Od početka doktorskog studija sudjelovala je na nekoliko radionica i većem broju međunarodnih kongresa te objavila šest radova u časopisima indeksiranim u Current Contents bazi i jednu monografiju kao koautorica.

Popis publikacija

Znanstveni radovi objavljeni u časopisima citiranim u Current Contents bazi:

Plavšin, I., Gunjača, J., Galić, V., Novoselović, D. (2022). Evaluation of Genomic Selection Methods for Wheat Quality Traits in Biparental Populations Indicates Inclination towards Parsimonious Solutions. Agronomy 12 (5): 1126. doi:10.3390/agronomy12051126

Plavšin, I., Gunjača, J., Šimek, R., Novoselović, D. (2021). Capturing GEI Patterns for Quality Traits in Biparental Wheat Populations. Agronomy 11 (6): 1022. doi:10.3390/agronomy11061022

Plavšin, I., Gunjača, J., Šatović, Z., Šarčević, H., Ivić, M., Dvojković, K., Novoselović, D. (2021). An Overview of Key Factors Affecting Genomic Selection for Wheat Quality Traits. Plants 10 (4): 745. doi:10.3390/plants10040745

Ivić, M., Grljušić, S., **Plavšin, I.**, Dvojković, K., Lovrić, A., Rajković, B., Černe, M., Popović, B., Lončarić, Z., Bentley, A. R., Swarbreck, S. M., Šarčević, H., Novoselović, D. (2021). Variation for Nitrogen Use Efficiency Traits in Wheat Under Contrasting Nitrogen Treatments in South-Eastern Europe. Front Plant Sci 12: 2545. doi:10.3389/fpls.2021.682333

Ivić, M., Grljušić, S., Popović, B., Andrić, L., **Plavšin, I.**, Dvojković, K., Novoselović, D. (2021). Screening of Wheat Genotypes for Nitrogen Deficiency Tolerance Using Stress Screening Indices. Agronomy 11 (8): 1544. doi:10.3390/agronomy11081544

Horvat, D., Šimić, G., Dvojković, K., Ivić, M., **Plavšin, I.**, Novoselović, D. (2021). Gluten Protein Compositional Changes in Response to Nitrogen Application Rate. Agronomy 11 (2): 325. doi:10.3390/agronomy11020325

Plavšin, I., Velki, M., Ećimović, S., Vrandečić, K., Ćosić, J. (2017). Inhibitory effect of earthworm coelomic fluid on growth of the plant parasitic fungus *Fusarium oxysporum*. Eur J Soil Biol 78: 1–6. doi:10.1016/j.ejsobi.2016.11.004

Plavšin, I., Staškova, T., Šery, M., Smykal, V., Hackenberger, B. K.; Kodrik, D. (2015). Hormonal enhancement of insecticide efficacy in *Tribolium castaneum*: Oxidative stress and metabolic aspects. Comp Biochem Physiol Part - C: Toxicol Pharmacol 170: 19–27. doi:10.1016/j.cbpc.2015.01.005

Velki, M., **Plavšin, I.**, Dragojević, J., Hackenberger, B. K. (2014). Toxicity and repellency of dimethoate, pirimiphos-methyl and deltamethrin against *Tribolium castaneum* (Herbst) using different exposure methods. J Stored Prod Res 59: 36–41. doi:10.1016/j.jspr.2014.04.005

Poglavlja u knjigama:

Kodrík, D., **Plavšin, I.**, Velki, M., Stašková, T. (2015). Enhancement of insecticide efficacy by adipokinetic hormones. U: Insecticides: Occurrence, Global Threats and Ecological Impact (ur. Montgomery, J.), Nova Science Publishers, New York, USA, str. 77–92. ISBN: 978-1-63483-475-9

Monografije:

Novoselović, D., Dvojković, K., Vrandečić, K., Brkić-Bubola, K., Grljušić, S., Horvat, D., Drezner, G., Oplanić, M., **Plavšin, I.**, Ivić, M., Lutrov, K. (2018). Razvoj germplazme krušne i durum pšenice za obiteljska poljoprivredna gospodarstva u Republici Hrvatskoj. Poljoprivredni institut Osijek, Osijek. ISBN: 978-953-7843-08-3

Kongresna priopćenja:

Plavšin, I., Gunjača, J., Novoselović, D. (2021). The prediction accuracy of genetic values is affected by imputation method within a wheat biparental population. Book of Abstracts of 6th Conference on Cereal Biotechnology and Breeding, 3.-5.11.2021., Budimpešta, Mađarska, str. 101–102.

Plavšin, I., Gunjača, J., Šatović, Z., Novoselović, D. (2021). Comparison of genomic selection models to predict wheat quality traits in a biparental population. Book of Abstract of 56th Croatian and 16th International Symposium on Agriculture, Faculty of Agrobiotechnical Sciences Osijek, University Josip Juraj Strossmayer in Osijek, 5.-10.09.2021., Vodice, Hrvatska, str. 130.

Horvat, D., Šimić, G., Dvojković, K., Ivić, M., **Plavšin, I.**, Novoselović, D. (2019). Wheat proteins: Quantitative distribution under nitrogen fertilisation. Book of Abstracts of 10th International and 12th Croatian Congress of Cereal Technologists „Flour-Bread '19“, Faculty of Food Technology Osijek, Josip Juraj Strossmayer University of Osijek, 11.-14.6.2019., Osijek, Hrvatska, str. 68.

Novoselović, D., Ivić, M., **Plavšin, I.**, Lovrić, A., Černe, M., Maričević, M., Rajković, B., Šarčević, H. (2019). NUE variation within a panel of selected winter wheat cultivars under South-Eastern European conditions. Book of Abstracts of 5th Conference on Cereal Biotechnology and Breeding, 4.-7.11.2019., Budimpešta, Mađarska, str. 83–84.

Dvojković, K., Novoselović, D., Horvat, D., Ivić, M., **Plavšin, I.**, Drezner, G. (2019). Wheat Breeding Challenges Related to Climate Changes. Book of Abstracts of 10th International and 12th Croatian Congress of Cereal Technologists „Flour-Bread '19“, Faculty of Food Technology Osijek, Josip Juraj Strossmayer University of Osijek, 11.-14.6.2019., Osijek, Hrvatska, str. 4.

Plavšin, I., Ivić, M., Novoselović, D., Resan, B. (2019). Elemental LIBS analysis of wheat samples using nJ femtosecond laser. Proceedings of SPIE 11370, Ultrafast Optics 2019, 6.-11.10.2019., Bol, Hrvatska, str. 333–337.

Lovrić, A., Šarčević, H., Maričević, M., Ikić, I., Černe, M., Ivić, M., **Plavšin, I.**, Dvojković, K., Rajković, B., Novoselović, D. (2019). Relativna učinkovitost indirektne i direktne selekcije za prinos zrna pri niskoj i visokoj razini dušika kod ozime pšenice. Zbornik sažetaka 54. hrvatskog i 14. međunarodnog simpozija

agronoma, Sveučilište u Zagrebu, Agronomski fakultet, 17.-22.2.2019., Vodice, Hrvatska, str. 85–86.

Ivić, M., **Plavšin, I.**, Černe, M., Popović, B., Maričević, M., Lovrić, A., Šarčević, H., Novoselović, D. (2018). Utjecaj dušičnog stresa na neka svojstva pšenice u ovisnosti o sorti i okolini. Knjiga sažetaka "Potencijal tla i zemljišnih resursa: ključne uloge znanosti i učinkovitih politika", 13. kongres Hrvatskog tloznanstvenog društva s međunarodnim sudjelovanjem, 10-14.09.2018., Vukovar, Hrvatska, str. 47.

Grljušić, S., Dvojković, K., Černe, M., Ivić, M., **Plavšin, I.**, Drezner, G., Novoselović, D. (2018). Interakcija genotip x okolina za urod durum pšenice. Zbornik sažetaka 11. međunarodnog kongresa "Oplemenjivanje bilja, sjemenarstvo i rasadničarstvo", Hrvatsko agronomsko društvo, Zagreb, 7-9.11.2018., 37–38.

Plavšin, I., Stašková, T., Kodrík, D. (2015). Interactions of insecticides with adipokinetic hormones in *Tribolium castaneum.* Book of Abstracts of 4th Young Environmental Scientists (YES) meeting, 14.-19.3.2015., Petnica, Valjevo, Srbija, str. 37.

Kodrík, D., **Plavšin, I.**, Velki, M., Stašková, T. (2015). Prospective utilization of insect stress hormones in pest control. Book of Abstracts of Society for Experimental Biology Meeting Prague 2015, 30.6.-3.7.2015., Prag, Češka, str. 28.

Plavšin, I., Hackenberger Kutuzović, B. (2013). Residual film method and modified residual film methods used to evaluate toxicity of three pesticides on *Tribolium castaneum.* Book of Abstracts of 3rd Young Environmental Scientists (YES) meeting "Interdisciplinary discourse on current environmental challenges", 11.-13.2.2013., Krakow, Poljska, str. 62.

PRILOZI

Prilog 1. Znanstveni rad: Plavšin, I., Gunjača, J., Šimek, R., Novoselović, D. (2021). Capturing GEI Patterns for Quality Traits in Biparental Wheat Populations. Agronomy 11 (6): 1022. doi:10.3390/agronomy11061022

agronomy

MDPI

Article

Capturing GEI Patterns for Quality Traits in Biparental Wheat Populations

Ivana Plavšin [1,2], Jerko Gunjača [2,3,*], Ruđer Šimek [1,†] and Dario Novoselović [1,2]

1 Department for Cereal Breeding and Genetics, Agricultural Institute Osijek, Južno Predgrađe 17, 31 000 Osijek, Croatia; ivana.plavsin@poljinos.hr (I.P.); rudjer.simek@gmail.com (R.Š.); dario.novoselovic@poljinos.hr (D.N.)
2 Centre of Excellence for Biodiversity and Molecular Plant Breeding (CoE CroP-BioDiv), Svetošimunska Cesta 25, 10 000 Zagreb, Croatia
3 Department of Plant Breeding, Genetics and Biometrics, University of Zagreb Faculty of Agriculture, Svetošimunska Cesta 25, 10 000 Zagreb, Croatia
* Correspondence: jgunjaca@agr.hr; Tel.: +385-1-239-3938
† Current address: Arguo d.o.o., Zadarska 14, 31 000 Osijek, Croatia.

Abstract: Genotype-by-environment interaction (GEI) is often a great challenge for breeders since it makes the selection of stable or superior genotypes more difficult. In order to reduce drawbacks caused by GEI and make the selection for wheat quality more effective, it is important to properly assess the effects of genotype, environment, and GEI on the trait of interest. In the present study, GEI patterns for the selected quality and mixograph traits were studied using the Additive Main Effects and Multiplicative Interaction (AMMI) model. Two biparental wheat populations consisting of 145 and 175 RILs were evaluated in six environments. The environment was the dominant source of variation for grain protein content (GPC), wet gluten content (WGC), and test weight (TW), accounting for approximately 40% to 85% of the total variation. The pattern was less consistent for mixograph traits for which the dominant source of variation has been shown to be trait and population-dependent. Overall, GEI has been shown to play a more important role for mixograph traits compared to other quality traits. Inspection of the AMMI2 biplot revealed some broadly adapted RILs, among which, MG124 is the most interesting, being the prevalent "winner" for GPC and WGC, but also the "winner" for non-correlated trait TW in environment SB10.

Keywords: wheat quality; mixograph; biparental population; GEI; AMMI; EM-AMMI

Citation: Plavšin, I.; Gunjača, J.; Šimek, R.; Novoselović, D. Capturing GEI Patterns for Quality Traits in Biparental Wheat Populations. *Agronomy* **2021**, *11*, 1022. https://doi.org/10.3390/agronomy11061022

Academic Editors: Folkard Asch and Sabine Stürz

Received: 24 April 2021
Accepted: 19 May 2021
Published: 21 May 2021

1. Introduction

Before release and widespread use for human consumption, wheat cultivars must possess suitable end-use quality. However, baking quality improvement is one of the most demanding objectives in wheat breeding, since the majority of quality traits exhibit complex inheritance patterns. Wheat baking quality could be evaluated by a large number of traits, which are generally controlled by many minor-effect and a few large-effect genes [1]. One of the most important factors affecting dough characteristics is gluten content and its strength [2,3]. Gluten is the most abundant wheat protein, which contributes to approximately 75–80% of total grain protein content (GPC) [4]. Such a large abundance consequently causes a high positive correlation of GPC and gluten content with other quality traits [5]. Structurally, gluten is a complex network of monomeric gliadins and polymeric subunits of glutenin [6,7], the quality and strength of which is determined by the proportion of its components and their quality [8,9], rather than by the total GPC or gluten content. Among all gluten components, high molecular weight glutenin subunits (HMW-GS) have the greatest impact on bread quality [8]. GPC is usually used as an indicator of baking quality [10], while wet gluten content (WGC) illustrates dough water absorption and its ability to form a gluten network, hence indicating the stability of the dough [11]. Among

other grain quality traits, test weight (TW) is often used as a predictor of flour extraction and thus wheat quality [12]. Taking into account often costly and time-consuming phenotyping for wheat baking quality, as well as the existence of complex interaction between wheat proteins and other components, such as pentosans and puroindolins, predictability of dough baking quality may be very difficult [13,14]. Therefore, in addition to the qualitative and quantitative composition analyses, rheological tests are required in order to assess baking quality more accurately. Rheological tests simulate the baking performance of the dough, evaluating its mixing and viscoelastic properties. Different devices can be used to perform rheological tests, i.e., farinograph, extensograph, and mixograph. Mixograph is a dough mixer that creates dough rheological profile based on numerous variables, which together with grain protein and gluten content, can give a reliable estimation of baking quality [15]. It is used to obtain general information about dough mixing and behavior during development, as well as the strength of the dough [16,17]. Due to the small amount of flour required (2, 10, or 35 g) and relatively fast interpretation of the obtained results, mixograph is highly suitable for use in plant breeding especially in early generations [18].

The success of wheat quality improvement depends on the ability to develop a genotype with both superior performance and high stability of the quality traits. In this context, one of the major challenges in plant breeding is the existence of genotype-by-environment interaction (GEI). In order to achieve more effective selection in wheat breeding programs, it is of the utmost importance to understand how genotype, environment, and GEI each contribute to the trait of interest [19]. Differential response of genotypes depending on the environmental conditions has already been determined for the majority of the wheat grain quality traits [20–23]. The greatest part of the phenotypic variation for GPC is due to non-genetic factors among which a strong influence of environment has been well documented [20,24,25]. Although gluten content shows a positive correlation with GPC, it is considered that for its qualitative characteristics genotype plays the most important role [26]. Compared to GPC, dough rheological traits are showed to be less influenced by GEI in some cases [27–29], while other research showed that variation due to GEI is equal or higher than variation caused by genotype or environment alone [20,23].

One of the most widely used methods for GEI analysis is the Additive Main Effects and Multiplicative Interaction (AMMI) model. It represents the combination of analysis of variance (ANOVA) procedure for main effects and the singular value decomposition (SVD) of GEI term [30]. Although most commonly used in multi-environment analyses of the most important traits such as wheat yield, AMMI model was also used to analyze GEI and stability of wheat quality traits [20,31,32]. In order to investigate GEI, the phenotypic performance of a collection of genotypes must be assessed in multiple environments. In plant breeding, recombinant inbred lines (RILs) represent a collection of genotypes with a noteworthy source of genetic diversity. However, in order to select RILs with wide trait adaptation, breeders should be able to reveal GEI effects underlying the trait of interest and to identify stable genotypes. The ability of AMMI model to explore GEI and to identify stable high-yielding genotypes in wheat populations consisting of RILs has already been proven in several studies [33–35]. Rodriguez et al. [36] concluded that RILs were more stable and tolerable to variable environmental conditions when compared to landraces and varieties. Although still not numerous, studies have shown that AMMI model can be used to successfully analyze GEI and its structure for wheat quality traits in RIL populations [37]. Among the traits tested in RIL populations, GPC is shown to be under the largest influence of environment, as well as of joint effects of environment and GEI (>98% and >99% of contribution to the total sum of squares, respectively) [38]. The predominant influence of environment, with more than 90% of phenotypic variation due to the joint influence of environment and GEI was reported by Elangovan et al. [39] for both GPC and TW. Furthermore, Prashant et al. [40] examined sources of variation affecting mixograph traits and determined a similar pattern, namely, the substantial contribution of environment and GEI to phenotypic variation of mixograph traits as well. On the other hand, there is some evidence that the AMMI model is not applicable with the same efficiency to all quality

traits. Specifically, regardless of the high and significant effect of GEI, AMMI model was not particularly successful in identifying stable genotypes for loaf volume [41]. Nevertheless, AMMI model has been able to successfully identify specifically adapted as well as stable wheat genotypes for the majority of quality traits examined in RIL populations [38].

The objectives of this study were to investigate transgressive and GEI patterns for selected quality and mixograph traits in two biparental wheat populations derived from the crosses of pairs of parental cultivars that either did not differ at all or differed in all HMW-GS, which play a key role in breeding for improved bread-making quality.

2. Materials and Methods

2.1. Plant Material

Two biparental (RIL) winter wheat populations were used in this study. The BK and MG population were derived from the Bezostaya-1 × Klara and Monika × Golubica crosses, respectively. In BK combination parental cultivars differed in all HMW-GS, while in MG combination parental cultivars did not differ in any HMW-GS [42]. The BK combination represented an example of crossing that is often applied in practical breeding programs where parents of significant phenotypic divergence are crossed, while the MG combination was chosen to narrow phenotypic variation to genetic factors excluding the differences in HMW-GS. After crossing and selfing, plants were randomly selected up to the F7 generation. The BK and MG populations consisted of 145 and 175 genotypes, respectively, including parental cultivars. Klara, Monika, and Golubica are high-yielding cultivars of good bread-making quality developed at the Agricultural Institute Osijek, while Bezostaya-1 is a Russian cultivar with good technological grain quality.

2.2. Description of Field Trials

The field trials were conducted during three consecutive years (2009–2011) at two locations in Croatia—Osijek (OS) and Slavonski Brod (SB), and both populations were evaluated in these six environments (location–year combinations). Soil type represented at locations Osijek and Slavonski Brod is eutric cambisol and eugley, respectively. Osijek and Slavonski Brod locations are classified as having a moderately warm and rainy oceanic climate (Cfb) by the Köppen–Geiger climate classification. A summary of meteorological data for both locations during three growing seasons is presented in Tables S1–S3 (Supplementary Materials). On average, no substantial difference was observed in daily temperatures and soil temperatures at 5 cm soil depth between Osijek and Slavonski Brod. The biggest difference in total rainfall between locations was observed during the 2008/2009 season, while no substantial difference was observed during two remaining seasons. In overall, the highest amount of precipitation was recorded during the 2009/2010 season.

In each environment, the field trial was set as a row–column design with two replicates, in 16 rows by 19 or 22 columns (for BK or MG population, respectively). The initial plot size was 4.86 m^2, and prior to harvesting, the front and back end of plots were trimmed to final net plot size of 2.7 m^2 for all genotypes and in all trials. Basic fertilization prior to planting was applied by adding 100 kg ha^{-1} of urea (46% N) and 300 kg ha^{-1} NPK (7:20:30) at both locations and all seasons. The N applied at top-dressings was 27 kg N ha^{-1} at tillering and stem extension growth stages, respectively. Total amount of applied macronutrients was 121 kg N ha^{-1}, 60 kg P_2O_5 ha^{-1} and 90 kg K_2O ha^{-1}. All other cultural practices including application of herbicides, insecticides, and fungicides to control major weeds, insects and foliar diseases were typical for commercial wheat production in Croatia. In the trials with BK population, parental cultivars were sown twice per replicate and another control (L84-2004) was added to both trial populations, to fill the grid. Due to the low flour sample quality and possible unreliability of obtained mixograph results, data collected for the BK population at the location Slavonski Brod in the year 2011 were not used in the analysis. In the text, tables, and graphical representations that follow, environments will be denoted using the location-year combination abbreviations: OS09 (Osijek—2009),

OS10 (Osijek—2010), OS11 (Osijek—2011), SB09 (Slavonski Brod—2009), SB10 (Slavonski Brod—2010), and SB11 (Slavonski Brod—2011).

2.3. Phenotyping

The TW and GPC traits were determined by near-infrared spectroscopy on whole grains using the Infratec 1221 Grain Analyzer and reported at 14% moisture basis, and expressed in kg hL^{-1} and %, respectively. Prior to WGC and mixograph analyses, grain samples were tempered and milled using a laboratory mill. The WGC was determined according to ICC standard method No 155 using the Glutomatic 2200 Gluten System and Glutomatic Centrifuge 2015, Perten, using a 10 g flour sample. The WGC is expressed as a percentage of mass relative to the initial sample mass, and calculated according to Equation (1):

$$\text{WGC } (\%) = \frac{\text{total wet gluten } (\text{g})}{10 \text{ g}} \times 100 \quad (1)$$

Dough rheology was assessed using the Swanson and Working Mixograph (National MFG Co., National Manufacturing Company, Lincoln, NE, USA). The amount of flour required for analysis is determined according to the sample protein amount and sample moisture. Flour moisture was measured using a Mettler Toledo HR83 moisture analyzer. Prior to mixing, the required amount of water was added to the sample, calculated according to the Equation (2):

$$\% \text{ abs} = (1.5 \times \% \text{ protein} + 43.6) \times 10 \text{ [mL]} \quad (2)$$

The results of the analysis were processed using MixSmart software (v 3.40). The best repeatability was determined for the variables of the central curve, which provide a comprehensive view of the optimal dough development. Therefore, the following variables were used as input for statistical analysis: MPT (midline peak time [min])—time required to achieve maximum dough resistance, i.e., time required for optimal dough development; MTW (midline curve tail width [%])—width of the peak at the end of the mixing period that indicates the consistency and stability of the dough at the end of mixing process; MTI (midline curve tail integral)—area below the midline curve from the starting point to the end of the mixing process that describes energy used during the mixing process; and MPH (midline peak height [%])—indicates the dough strength [40].

2.4. Statistical Analysis

At the first stage of the analysis, pooled data from individual trials were analyzed using the mixed model:

$$Y = G + E + G{\cdot}E + REP{\cdot}E \ : \ ROW{\cdot}REP{\cdot}E + COL{\cdot}REP{\cdot}E \quad (3)$$

that included the fixed effects of genotype (G), environment (G), GEI (G·E), and replicates within environments (REP·E), as well as the random effects of rows and columns within replicates within environments (ROW·REP·E and COL·REP·E, respectively). By allowing for all nested effects to vary across environments and removing the zero-variance effects, the optimal model was built using the Wald test and AIC (Akaike's information criterion) as selection criteria. Predicted values for all genotype-environment combinations were then taken as the input for the second stage of the analysis. The first stage of the analysis was performed within R environment [43], using the commercial package "asreml" [44] and freely available companion package "asremlPlus" [45].

The second stage of analysis began with an assessment of Pearson's correlations between the traits within and across available environments, using estimated values for all genotype-environment combinations from the first stage of the analysis. Mean values for parental cultivars together with mean values and ranges for the RILs were calculated

for all traits within as well as across environments. Matrix of genotype by environment estimates was decomposed using the AMMI model:

$$Y = G + E + (G \cdot E)_{pattern} + (G \cdot E)_{noise} \quad (4)$$

where Y is the predicted value of the dependent variable, G is the genotypic effect, E is the environmental effect, and interaction effect G·E is divided into selected AMMI model estimate—$(G \cdot E)_{pattern}$ and discarded residual—$(G \cdot E)_{noise}$. $(G \cdot E)_{pattern}$ effect is the sum of the appropriate (for a certain genotype-environment combination) matrix elements for k selected AMMI axes, where each matrix element is the product of the singular value for axis k (λ_k) and the appropriate elements of genotypic and environmental vectors for the same axis (γ_k and δ_k, respectively) [46,47].

In order to handle missing data that occurred for mixograph traits in both populations, the analysis for those traits was performed by using the Expectation–Maximization AMMI (EM–AMMI) algorithm according to Paderewski [48]. Three different approaches were used for the selection of terms that should be retained in the final model: (1) Simple parametric bootstrap (SPB) [49], (2) F_R-test [50], and (3) Leave-one-out cross-validation (LOO CV) procedure [48]. The contributions of genotype, environment, and GEI terms to the total variance were calculated as the ratio of the sum of squares (SS) of the corresponding term and the total SS, and expressed as a percentage [51]. The contributions of interaction principal component axis (IPCA) scores to GEI were calculated on the same principle, using the corresponding IPCA SS and GEI SS.

Instead of standard AMMI biplots, AMMI1 (main effects vs. IPCA1), and AMMI2 (IPCA1 vs. IPCA2), AMMI dissection of GEI patterns was visualized rather by using a modified version of AMMI2 biplot. It represents an attempt to combine properties of both standard types by adding the main effects to the AMMI2 biplot using a color scale. In order to avoid point cluttering, genotypic and environmental IPCA scores were calculated by applying different scaling (by multiplying their appropriate eigenvectors with two times square root and half square root of eigenvalues, respectively).

Second stage of the analysis was likewise performed within R environment [43], using the following packages: "corrplot" [52], "Hmisc" [53], "reshape2" [54], "dplyr" [55], "ggplot2" [56], and "ggrepel" [57]. An R function "EM.AMMI" [48] has been applied for the data imputation prior to AMMI analysis using two IPCAs, 1000 iterations, and precision level of 0.001 to run the EM–AMMI procedure. For the SPB test the adjusted R script from Forkman and Piepho [49] was used and the probability value of the test was obtained using 100,000 bootstrap samples. The F_R-test was performed as suggested by Piepho [50] using the degrees of freedom estimated for IPCA terms according to Gollob [58]. An R function "CV.LOO" [48] was used to perform LOO CV procedure using four (BK dataset) and five (MG dataset) IPCAs, and the permissible minimum number of observed values (MNO) in each row and each column of the data matrix set to three.

3. Results

3.1. The First Stage of the Analysis

Starting with the full model including separate nested effects for each environment, for each trait and population, models were gradually reduced until the optimal model was reached. The nested effects structure in selected optimal models is shown in the Supplementary Table S4. For random effects full model was reduced by removing all zero-variance effects (if any); there was one case where one of the effects was reduced to a single effect for all environments and the other completely removed from the model, thus providing the substantial reduction of AIC (MTW in MG). The fixed effect of replicates was truncated by keeping it in the model only for environments in which the Wald test was significant if that resulted in lower AIC compared to the model with full effect. Overall, no obvious pattern could be observed, as the optimal models tend to be specific for each trait/population combination.

3.2. Transgressive Segregation in Quality Traits

Overall descriptive statistics for seven traits of both populations are shown in Table 1, while Tables S5 and S6 (Supplementary Materials) are showing descriptive statistics per environment for BK and MG population, respectively. Means for parental cultivars are followed by means and ranges of RILs. Transgressive segregants are denoted as "positive" if they exhibited values higher than the parental cultivar with higher trait value, and as "negative" if their values were lower compared to the parental cultivar with lower trait value. Across environments, Bezostaya-1 had a higher mean compared to Klara for all traits except TW, MTI, and MPH, even though the difference in means for these two parental cultivars was not highly pronounced. The range of BK RILs was much wider than the range of parental cultivars, and both positive and negative transgressive segregants are found for all traits examined (in the range of 38.46–77.62% and 5.59–46.85% for positive and negative segregants, respectively). This may suggest the presence of increaser as well as decreaser alleles for quality traits in both parents. Mean values for BK RILs were similar to parental means for all traits, except for the mixograph traits MPT and MTW, the mean values of which were higher than the higher-performing parental cultivar (Table 1). These were also the traits for which the highest proportion of positive and the lowest proportion of negative segregants was observed, while for all the other traits across environments positive and negative segregants were equally represented. In the MG population, Golubica can be identified as the parent with a higher value for most of the traits, both within and across environments (Table 1). Mean values for MG RILs were within ranges of parental values for all traits as they were much wider than ranges between Bezostaya-1 and Klara. Consequently, a lower ratio of both positive and negative transgressive segregants was noticed in the MG population compared to the BK population in general. Trait-wise, the lowest ratios for both segregant groups were observed for mixograph traits MTI and MPH, while the largest disproportion between them was recorded in WGC where there were 5.4 times more positive than negative transgressive segregants. Considering individual environments, it is noticeable that means for MG RILs vary substantially between environments (Table S3 in Supplementary Materials). Interestingly, the strongest predominance of negative segregants over positive ones is recorded for mixograph traits MTI and MPH in the OS11 environment, where more than 90% negative and approximately 1% positive segregants were present. On the contrary, the same two traits showed the highest predominance of positive over negative segregants in environment SB11, with approximately 50% of positive and almost no negative segregants present.

Table 1. Summary of parental means, RIL means, and ranges, and rates of transgressive segregants across environments for seven quality traits assessed in BK and MG RIL wheat populations.

Trait	Parental Cultivars Mean		RILs			Positive Transgressive Segregants [3]		Negative Transgressive Segregants [4]	
	P1 [1]	P2 [2]	Min	Mean	Max	N	%	N	%
BK population									
GPC [5]	14.3	13.9	10.6	14.0	17.5	55	38.5	60	42.0
WGC	34.7	34.1	20.5	34.1	43.9	61	42.7	62	43.4
TW	79.1	79.8	64.8	79.5	86.9	71	49.7	46	32.2
MPT	5.2	4.6	1.3	5.4	10.0	86	60.1	19	13.3
MTW	20.1	18.8	10.8	21.3	35.1	111	77.6	8	5.6
MTI	359.6	367.2	237.3	361.3	504.9	65	45.5	66	46.2
MPH	41.6	42.4	25.5	41.8	58.9	66	46.2	67	46.9
MG population									
GPC	13.0	14.1	11.0	13.8	17.3	56	32.4	21	12.1
WGC	29.9	35.3	21.8	33.5	43.9	27	15.6	5	2.9
TW	79.3	80.4	65.7	79.6	86.2	46	26.6	70	40.5
MPT	4.6	4.9	1.6	4.8	9.1	95	54.9	63	36.4

Table 1. *Cont.*

Trait	Parental Cultivars Mean		RILs			Positive Transgressive Segregants [3]		Negative Transgressive Segregants [4]	
	P1 [1]	P2 [2]	Min	Mean	Max	N	%	N	%
MTW	9.3	17.9	4.4	15.2	45.9	46	26.6	19	11.0
MTI	347.5	438.8	285.5	382.8	527.8	3	1.7	10	5.8
MPH	41.5	52.2	30.8	45.4	73.2	5	2.9	18	10.4

[1] Bezostaya-1 and Monika for BK and MG population, respectively. [2] Klara and Golubica for BK and MG population, respectively. [3] RILs that exhibited values higher than the parental cultivar with higher trait value. [4] RILs that exhibited values lower than the parental cultivar with lower trait value. [5] Abbreviations: grain protein content (GPC), wet gluten content (WGC), test weight (TW), midline peak time (MPT), midline curve tail width (MTW), midline curve tail integral (MTI), midline peak height (MPH).

3.3. Phenotypic Correlations between Quality Traits

A summary of Pearson's correlation coefficients for both populations is presented in a form of correlograms across environments (Figure 1) and within environments (Figures S1 and S2 in Supplementary Materials). Very high positive correlations were observed across environments between GPC and WGC ($r \geq 0.89$), and between MTI and MPH ($r \geq 0.91$), and they were constant in both populations as well as within environments. Although not being very strong, negative correlations were observed in both populations between GPC/WGC and TW across environments. On the other hand, within environments, these correlations were mostly positive for the BK population with the lowest values recorded in the SB10 environment, while for the MG population almost no correlation was observed between these traits, except in the SB11 environment where correlations were positive but very weak ($r \sim 0.2$). Similarly, TW exhibited negative but low correlations with all four mixograph traits in the BK population, while in the MG population these correlations were mostly positive across environments. Substantial variations in terms of correlation strength and direction were observed for these trait pairs in both populations within environments. GPC and WGC showed positive correlations with mixograph traits MTI and MPH across environments in both populations, with the difference that correlations were much higher in BK compared to the MG population. These correlation patterns were also consistent within environments but showed higher variation in the case of the MG population (ranging between 0.18 and 0.66). In contrast, the correlations of the other two mixograph traits (MPT and MTW) with GPC/WGC exhibited quite different trends. MPT was weakly correlated with both GPC and WGC with opposite signs in two populations; MTW was moderately negatively correlated with them in the MG population, but not correlated in the BK population. This pattern was highly inconsistent within environments. Variations between environments observed for some trait pairs (mostly between TW and other traits), especially regarding the direction of correlations, should indicate a strong environmental impact.

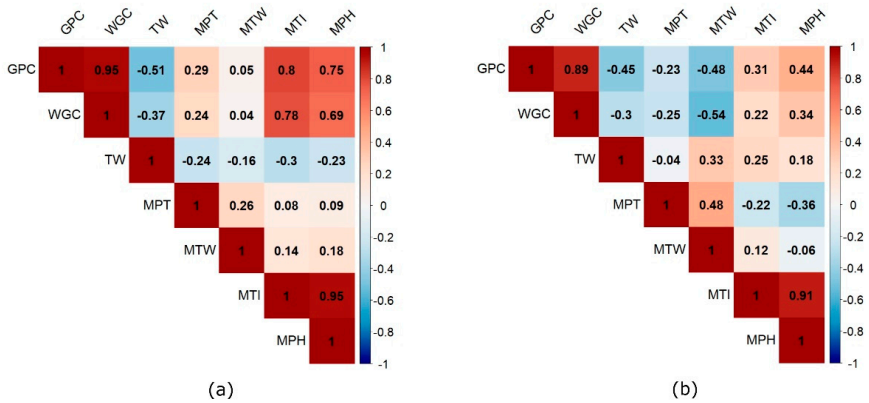


Figure 1. Pearson's correlation coefficients across environments for (**a**) BK population and (**b**) MG population.

3.4. Selection of the Appropriate AMMI Model

In the model selection process, i.e., deciding upon how many significant IPCA terms should be retained in the model, three different methods were compared: the SPB method, F_R-test, and LOO CV method. For all four mixograph traits in both populations, a small proportion of missing data occurred (<1%), so the presence of missing data implied the use of the EM–AMMI algorithm. Unlike LOO CV, which can be embedded into EM–AMMI analysis, for two other model selection methods, this creates the drawback caused by preselecting the model, i.e., beforehand decision upon the number of axes to be retained in the model. To resolve this issue, tests were first applied on the EM–AMMI imputed data based on all possible models, from AMMI0 to AMMIF (although for AMMIF algorithm never converged). Since no difference was observed regards the model selected using the simple parametric bootstrap method and F_R-test (Tables S7 and S8 in Supplementary Materials), all AMMI1 imputed data were used in all further calculations.

Tables 2 and 3 summarize test outcomes for all IPCA terms used, for the BK (4) and MG (5) population, stating their SS and contributions to GEI SS, as well as the test statistics and corresponding probability values for all methods applied. The optimal number of IPCA terms for SPB and F_R-test was determined by *p*-value taking into account all terms until a non-significant value is obtained (at the level of significance $p < 0.05$). The selection by the LOO CV method is based on Root Mean Square Prediction Difference (RMSPD), and the optimal model was the one with the lowest RMSPD. Generally, there is a large disagreement between the test outcomes. In exactly half of the trait-population combinations, SPB and F_R-test selected the same AMMI model, while in the other half they selected the models that could substantially differ in the number of axes retained. F_R-test was overall more liberal, and more prone to the tendency to declare all tests significant. On the other side, there is the LOO CV, as the most conservative criterion. It selected the additive model (AMMI0) as the most appropriate one for all except one trait in each population (MTI in BK, and MTW in MG population).

Table 2. Results of tests for IPCA terms for seven quality traits assessed in BK population.

IPCA	Sum of Squares			Simple Bootstrap		F_R-Test		Cross-Validation
	IPCA SS	%	Total %	*T*	*p* Value	*F*	*p* Value	*RMSPD*
				GPC				
0								14.7
1	58.0	42.6	42.6	0.43	0.000	2.16	0.000	21.1
2	40.0	29.4	71.9	0.51	0.000	2.05	0.000	49.9
3	21.6	15.8	87.8	0.57	0.315	1.27	0.076	52.0
4	16.7	12.3	100.0					
				WGC				
0								44.4
1	524.7	42.2	42.2	0.42	0.000	2.13	0.000	54.1
2	343.1	27.6	69.8	0.48	0.001	1.79	0.000	105.9
3	192.8	15.5	85.3	0.51	0.955	1.04	0.413	197.9
4	183.2	14.7	100.0					
				TW				
0								39.6
1	562.4	56.9	56.9	0.57	0.000	3.85	0.000	94.1
2	284.1	28.7	85.6	0.67	0.000	3.91	0.000	97.7
3	89.9	9.1	94.7	0.63	0.007	1.69	0.001	118.0
4	52.5	5.3	100.0					
				MPT				
0								39.3
1	400.2	41.2	41.2	0.41	0.000	2.05	0.000	46.9
2	268.8	27.7	68.8	0.47	0.001	1.73	0.000	102.3
3	209.9	21.6	90.4	0.69	0.000	2.23	0.000	108.8
4	93.3	9.6	100.0					

Table 2. *Cont.*

IPCA	Sum of Squares			Simple Bootstrap		F_R-Test		Cross-Validation
	IPCA SS	%	Total %	*T*	*p* Value	*F*	*p* Value	*RMSPD*
				MTW				
0								74.1
1	1774.3	48.7	48.7	0.48	0.000	2.68	0.000	102.0
2	800.3	22.0	70.7	0.43	0.061	1.46	0.003	119.9
3	571.6	15.7	86.4	0.54	0.684	1.14	0.215	235.5
4	496.8	13.6	100.0					
				MTI				
0								796.8
1	126404.6	31.3	31.3	0.32	0.255	1.37	0.009	1043.1
2	116513.4	28.8	60.1	0.41	0.210	1.35	0.016	1983.6
3	94460.8	23.4	83.5	0.59	0.120	1.40	0.024	133.1
4	66703.0	16.5	100.0					
				MPH				
0								102.6
1	1999.9	30.2	30.2	0.30	0.606	1.27	0.036	128.3
2	1866.4	28.1	58.3	0.40	0.296	1.32	0.026	157.6
3	1466.6	22.1	80.4	0.53	0.792	1.11	0.275	252.0
4	1300.9	19.6	100.0					

Table 3. Results of tests for IPCA terms for seven quality traits assessed in MG population.

IPCA	Sum of Squares			Simple Bootstrap		F_R-Test		Cross-Validation
	IPCA SS	%	Total %	*T*	*p* Value	*F*	*p* Value	*RMSPD*
				GPC				
0								20.7
1	108.2	36.6	36.6	0.37	0.000	2.25	0.000	42.2
2	71.9	24.4	61.0	0.38	0.000	1.83	0.000	69.2
3	45.1	15.3	76.3	0.39	0.376	1.26	0.035	84.0
4	40.5	13.7	90.0	0.58	0.124	1.35	0.024	75.1
5	29.6	10.0	100.0					
				WGC				
0								62.8
1	855.4	31.6	31.5	0.32	0.000	1.79	0.000	69.3
2	748.9	27.6	59.2	0.40	0.000	1.99	0.000	178.6
3	432.3	15.9	75.1	0.39	0.385	1.26	0.036	295.1
4	405.4	15.0	90.1	0.60	0.028	1.49	0.005	262.5
5	268.4	9.9	100.0					
				TW				
0								44.7
1	957.0	69.7	69.7	0.70	0.000	8.92	0.000	48.6
2	151.6	11.0	80.7	0.36	0.001	1.68	0.000	69.2
3	116.9	8.5	89.2	0.44	0.009	1.55	0.000	90.3
4	77.3	5.6	94.8	0.52	0.864	1.07	0.319	115.7
5	71.1	5.2	100.0					
				MPT				
0								35.6
1	352.0	39.8	39.8	0.40	0.000	1.93	0.000	50.4
2	154.0	17.4	57.2	0.29	0.817	0.80	0.953	162.4
3	149.7	16.9	74.1	0.40	0.302	0.65	0.998	185.4
4	128.2	14.5	88.6	0.56	0.283	2.54	0.000	175.8
5	100.6	11.4	100.0					

Table 3. *Cont.*

IPCA	Sum of Squares			Simple Bootstrap		F_R-Test		Cross-Validation
	IPCA SS	%	Total %	*T*	*p* Value	*F*	*p* Value	*RMSPD*
				MTW				
0								114.4
1	3996.4	43.2	43.2	0.43	0.000	2.24	0.000	114.3
2	1726.7	18.7	61.9	0.31	0.402	0.87	0.858	385.1
3	1391.4	15.0	76.9	0.39	0.332	0.64	0.998	336.6
4	1165.1	12.6	89.5	0.54	0.563	2.34	0.000	454.4
5	973.6	10.5	100.0					
				MTI				
0								1078.0
1	255111.6	32.0	32.0	0.32	0.000	1.38	0.003	1182.6
2	195556.4	24.5	56.5	0.36	0.001	1.11	0.211	1473.6
3	137022.0	17.2	73.7	0.40	0.311	0.65	0.998	2656.3
4	114022.2	14.3	88.0	0.54	0.509	2.38	0.000	2750.6
5	95698.0	12.0	100.0					
				MPH				
0								150.9
1	4709.2	30.1	30.1	0.30	0.001	1.26	0.025	172.8
2	3671.7	23.5	53.6	0.34	0.027	1.00	0.506	217.1
3	2779.2	17.8	71.4	0.38	0.535	0.61	0.999	494.6
4	2562.1	16.4	87.8	0.57	0.155	2.68	0.000	684.3
5	1903.4	12.2	100.0					

3.5. GEI Patterns in Quality Traits

A preliminary insight to the biplot disclosure of GEI patterns is given in Table 4, showing relative contributions of the different AMMI2 model terms to the total variability. Besides population or trait-specific, some common general patterns can be observed as well. For three non-mixograph traits (GPC, WGC, and TW) the environment (E) was the dominant source of variation, consistently accounting for approximately three-quarters of the total variation in the BK population and varying between 40% and 85% in the MG population. The relative contributions of the two remaining effects are equally uniform in the BK population, where genotypic (G) variation is approximately twice as large as interaction (GEI) variation. This difference is reduced in the MG population, where G is approximately 1.5 times larger (GPC and WGC), or even smaller than GEI (TW). Finally, E is always much larger than the other two terms, except for GPC in the MG population where it is just slightly larger than G. Four mixograph traits could be roughly divided into two pairs that exhibit similar patterns. Dominant effect on MPT and MTW has GEI in BK and G in MG population. For this first pair of traits the E always has the weakest effect, except for MTW in the MG population, where the effect of E is slightly bigger than the effect of GEI. The other pair consists of MTI and MPH, characterized by the dominant effect of E in the BK and GEI in the MG population; the size of the contribution of G is always between the other two effects. The amount of GEI pattern captured by the first IPCA varies from 30% to 70%, which increases to a cumulative 54% to 86% for the first two IPCAs. The trait with the largest amount of pattern captured by either first IPCA or first and second together is TW in both populations.

Table 4. AMMI2 GEI patterns in both populations.

Trait	BK Population					MG Population				
	Contribution to Total SS (%)			Contribution to GEI (%)		Contribution to Total SS (%)			Contribution to GEI (%)	
	G	E	GEI	IPCA1	IPCA1 + IPCA2	G	E	GEI	IPCA1	IPCA1 + IPCA2
GPC	17.7	74.9	7.4	42.6	71.9	36.1	42.3	21.6	36.6	61.0
WGC	17.6	75.9	6.5	42.2	69.8	24.7	59.7	15.7	31.6	59.2
TW	18.3	71.6	10.2	56.9	85.6	6.9	84.9	8.1	69.7	80.7
MPT	26.8	11.0	62.2	41.2	68.8	47.1	12.6	40.4	39.8	57.2
MTW	39.8	9.2	50.9	48.7	70.7	45.5	33.7	20.8	43.2	61.9
MTI	29.9	45.9	24.2	31.3	60.1	31.2	23.7	45.2	32.0	56.5
MPH	30.7	40.3	29.0	30.2	58.3	36.7	15.7	47.5	30.1	53.6

Before inspecting the biplots, it is necessary to explain all the modifications made to the standard AMMI2 biplot. The color scale used on biplots indicates the mean value of the corresponding trait RILs, parental cultivars, and environments. RILs are designated by points; environments and parental cultivars by abbreviated labels (with or without a frame, respectively). "Winner" RILs are labeled and linked with their winning environments by a matching super/subscripted letter. The list of "winner" RILs within each environment is given in Supplementary Table S9. Interestingly, common "winner" RILs were not observed among traits that were not highly correlated in none of the two populations. In order to reduce the number of presented biplots, traits were paired based on correlations, the similarity of GEI patterns, and common "winners". One biplot from each pair is included in Figure 2 (BK population) or Figure 3 (MG population), and the other one moved to Figure S3 or Figure S4 in the Supplementary Materials. Generally, there are no clear patterns of distinction between neither locations nor years, so environmental variation is mostly due to the presence of some favorable or unfavorable combinations of locations and years. One such example is OS10, which is a favorable environment for the BK population for most of the traits (and unfavorable for TW); while the same statement is only partially applicable to the MG population. For breeding purposes, the focus is usually directed at finding generally or specifically adapted RILs. Proper candidates for selection as broadly adapted RILs are common "winners", such as BK012 for TW (Figure 2b), BK042 for MPT (Figure 2c), and MG124 for GPC and WGC (Figure S4a and Figure 3a, respectively). Especially interesting could be the last one, MG124, being the almost universal "winner" for GPC and dominant "winner" for its correlated trait WGC, but also "winner" for non-correlated trait TW in SB10. In some cases, the majority of "winning" RILs had low means of the corresponding trait, e.g., MTI in the BK population (Figure 2d) and MTW in the MG population (Figure 3c). Specifically adapted RILs are all exclusive "winners", and generally, all RILs that are located close to a certain environment. Out of all possible examples that could be found across all biplots, of particular interest could be those with closely matched color or completely mismatched color with the environment to which they are adapted. A mismatching example is high GPC line BK007 in the low GPC environment SB09, which due to high average protein content should be considered "broad" rather than "specifically" adapted RIL, capable to retain high protein content even in unfavorable environments. Before commenting on two matching cases for mixograph traits, it should be stressed that, for them, either low or high values are not immediately considered as desirable. Therefore, matching high values, i.e., MG016 with SB09 for MTW (Figure 3c) could be a negative feature if added interaction effect pushes already high values over the upper desirable limit. On the other side, matching low values, i.e., BK059 with SB09 for MTI, should generally be considered as positive features.

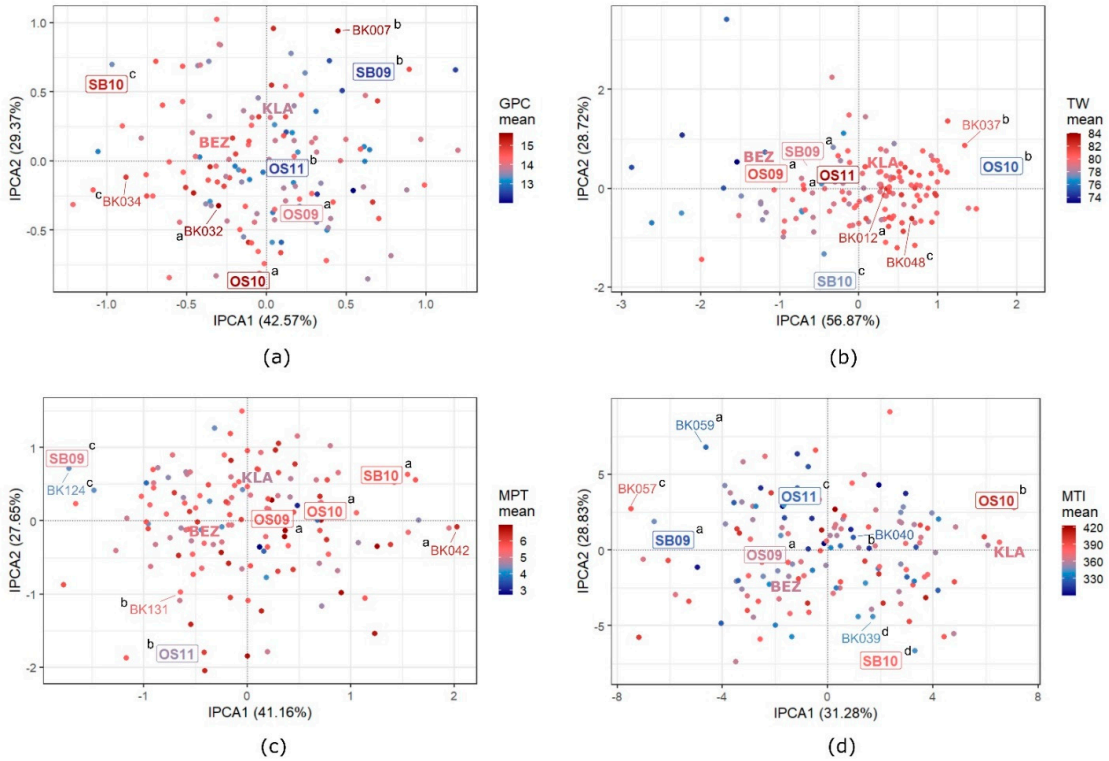


Figure 2. Modified AMMI2 biplots for (**a**) GPC, (**b**) TW, (**c**) MPT, and (**d**) MTI traits of BK RIL population. RILs are marked with dots, while environments and parental cultivars (Bezostaya-1 and Klara) are marked by abbreviated labels (with or without a frame, respectively). Within each environment one "winning" RIL is labeled and linked with its winning environment by a matching super/subscripted letter. The color indicates the mean value of the trait.

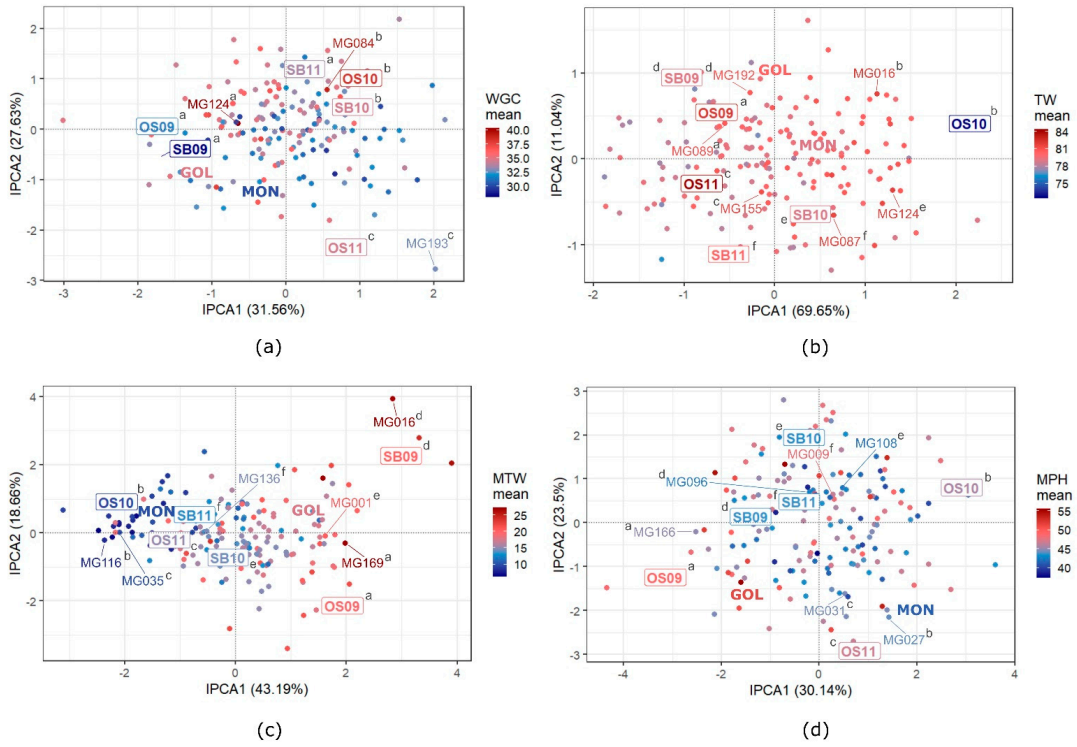


Figure 3. Modified AMMI2 biplots for (**a**) WGC, (**b**) TW, (**c**) MTW, and (**d**) MPH traits of MG RIL population. RILs are marked with dots, while environments and parental cultivars (Bezostaya-1 and Klara) are marked by abbreviated labels (with or without a frame, respectively). Within each environment, one "winning" RIL is labeled and linked with its winning environment by a matching super/subscripted letter. The color indicates the mean value of the trait.

4. Discussion

4.1. Transgressive Segregation and Phenotypic Correlations among Quality Traits

To produce high-quality dough and bread, wheat has to possess a suitable quality which is often assessed by protein and gluten content. Although the preferred value range of quality traits can vary considerably depending on the purpose of the wheat flour, for bread making GPC should be at least 12.5% [59]. WGC is an indicator of protein quality and according to Singh and Singh [60], its minimal value for wheat flour should be 24%. On the other hand, relevant research showed that gluten composition and its quality play a more important role compared to the quantity alone [7,8]. Horvat et al. [42] reported that, together with the composition of HMW-GS, the proportion of subunits must also be taken into account when assessing wheat quality. Research on the quality that included wheat cultivars explored in the present study showed that Bezostaya-1 was the cultivar of generally higher quality when compared to Monika and Golubica, which was attributed to the presence of 5 + 10 HMW-GS on D1 locus. However, regardless of possessing 2 + 12 HMW-GS type on D1 locus, which is generally associated with lower quality, Klara, Monika, and Golubica manifested very good quality. This elevated quality of cultivars with unfavorable HMW-GS was explained by higher proportion of total HMW-GS [42].

The extensibility and elasticity of gluten define its quality and consequently the quality of the dough, which can be estimated using the mixograph. Although it is difficult to exactly define desirable values for the mixograph traits as they vary depending on the type of population being studied and the purpose of the flour, generally it can be concluded that the good quality of dough is characterized by strong gluten which is indicated by longer optimal dough development time (MPT), greater consistency, and stability after mixing (MTW), higher energy used during mixing (MTI), and higher dough strength (MPH) [61]. The greatest advantage of using mixograph is that it requires a small amount of flour for analysis thus enabling the prediction of dough properties in early generation progenies.

In the present study, positive as well as negative transgressive segregants were detected in both populations although being generally more prevalent in the BK population. This result is expected considering that the parental cultivars Bezostaya-1 and Klara have different genetic bases of quality (different HMW-GS on all Glu-1 loci) [42]. For all examined traits, RIL means were positioned close to the parental means, except for the mixograph traits MPT and MTW within the BK population, the mean values of which were higher than the higher-performing parental cultivar Bezostaya-1. For these traits, this could be an indication of the presence of stronger epistatic effects. On the other hand, Monika and Golubica share the same genetic basis of quality, meaning that they possess the same type of HMW-GS on all Glu-1 loci; but show a higher phenotypic difference compared to Bezostaya-1 and Klara, resulting in a generally lower representation of segregants in MG RIL population. However, the occurrence of transgressive segregation for all quality traits examined in this study in both RIL populations suggests the presence of increaser and decreaser alleles in all four parental cultivars.

A high positive correlation between GPC and WGC recorded in both populations is expected, taking into account that gluten is the most abundant wheat protein and that the interrelatedness of these two traits is already well documented [4,5]. Additionally, in both populations, a high positive correlation was observed between mixograph traits MTI and MPH suggesting that higher energy must be used in the mixing process of the dough with higher strength and that this relationship is not dependent on the genetic background of wheat quality. When comparing correlations between WGC and mixograph traits for both populations in the present study, somewhat opposite patterns can be observed. The correlations between WGC and MTI/MPH were stronger in BK compared to the MG population suggesting that when at least one parental cultivar possesses HMW-GS associated with good bread-making quality, WGC can affect the dough strength. On the other hand, negligible positive and even negative correlations observed between WGC and MPT/MTW imply that optimal development time and the consistency of the dough do not

depend on the gluten quantity, but rather on its quality which is mostly influenced by the composition of HMW-GS [8]. These findings suggest that the parental genetic background of quality may not have an impact on traits such as GPC, WGC, and TW, if their phenotypic performance is already satisfactory, but may considerably affect mixograph traits, i.e., dough rheology. The importance of Glu-D1 in controlling mixograph traits has been confirmed in previous studies [61–63].

4.2. Selection of the Appropriate AMMI Model

The problem of finding the appropriate tests for IPCA terms is present in the literature on GEI over the last forty years. Numerous different tests were proposed, and some of them have already been compared in several published studies [49,50]. However, due to the specific nature of studies based on RILs (a large number of genotypes tested in just a few environments), it seemed interesting to employ three different tests in this study, and compare their outcomes. The LOO CV method turned out to be the most conservative, almost exclusively selecting the additive model (AMMI0), thus suggesting the complete absence of GEI. This does not seem to be a realistic conclusion, especially in cases where half or more of the total SS can be attributed to GEI. Out of the two remaining tests, F_R-test was slightly more liberal, tending to select more complex models, thus overfitting noise. This promotes the SPB test to the most appropriate one, which is in agreement with Forkman and Piepho's [49] conclusion. They have highly recommended the SPB test because it outperformed both F_R-test and cross-validation method in terms of performance power and probability of getting false-positive results, and accordingly, it was used in the present study as the criterion for the decision on the number of terms to be retained in the model. However, some caution should be taken when selecting more complex models (AMMI3 in this study), as they tend to overfit the noise [64], especially in the presence of missing data [65]. For the data sets with the larger amount of missing data, some more complex methods of data imputation [66] are probably more appropriate than the simple method used here [48].

4.3. GEI Patterns

According to the review by Williams et al. [27], the dominant effect of E for traits like GPC, WGC, and TW has been observed in numerous studies in the past, although they have cited more than a few exceptions to this general rule. However, rather than making comparisons over all sorts of GEI studies in wheat, it would make much more sense to search specifically for studies based on RILs, as they typically include a large number of genotypes tested over just a few environments. If some recent studies based on RILs are considered, there is almost no evidence for the dominant effect of E. It was reported for GPC only by Prashant et al. [40], while Echeverry-Solarte et al. [67] and Krishnappa et al. [38] detected equal effects of G and E for the same trait. On the contrary, in Goel et al. [68], the effect of E was smallest for GPC, WGC, and TW. The most common mixograph trait used in studies based on RILs is MPT, for which Prashant et al. [40] detected the dominant effect of E, the same as is in the BK population from the present study. In the other population, MG, the dominant effect for MPT is GEI, corresponding to the findings of Goel et al. [68]. The remaining effect, G, was found to be dominant for MPT in studies by both Echeverry-Solarte et al. [67] and Jin et al. [69]. Prashant et al. [40] have also detected dominant effects of GEI for MTW, and E for MTI, both corresponding to the same findings for the BK population in the present study. Generally, all the cited studies differ in many genotypic (type of parents, parental differences, etc.) as well as environmental (width of the range of environmental differences, climate conditions, etc.) factors, which all could be used as explanation why no common pattern could be established for any of the considered traits. However, GEI seems to be much more important for mixograph traits than for other quality traits (with one exception).

The use of standard AMMI biplots in studies involving a large number of genotypes can create an unreadable clutter of points or labels [34,70,71] unless a reduced set of lines

was selected for biplot representation [38]. While point cluttering can be avoided by different scaling of genotypic and environmental scores, label cluttering was prevented by labeling only parents and "winners". Once this problem was sorted out, the informativeness of standard biplots can be increased by carefully adding some extra features. Probably the most important modification is the addition of main effects that enable the integration of AMMI1 and AMMI2 biplots. Although there are some other solutions [19], the use of color scale proved to work well if a large number of genotypes need to be plotted.

4.4. Consequences for a Breeding Program

The most plausible explanation for the differences in adaptability between two RIL populations could be attributed to the use of widely adapted cultivar Bezostaya-1 in one cross (Bezostaya-1 × Klara) as opposed to the use of two cultivars from the same breeding program in another cross (Monika × Golubica) which were never grown on a large scale in the region of origin. It fits well the expected outcome and confirms the soundness of the standard crossing approach/strategy where one of the parents used for crossing should be widely adapted cultivar for the trait of interest. Furthermore, BK cross produced much more transgressive segregants, thus providing a wider base for selection of RILs with broad or specific adaptation in earlier generations. AMMI analysis provides the means for an easy identification of the potentially interesting RILs by simple visual inspection of biplots. The selected RILs have promising potential for use in the breeding programs aimed at quality improvement.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/agronomy11061022/s1, Figure S1: Pearson's correlation coefficients across environments for BK population, Figure S2: Pearson's correlation coefficients across environments for MG population, Figure S3: Modified AMMI2 biplots for (a) WGC, (b) MTW, and (c) MPH traits of BK RIL population, Figure S4: Modified AMMI2 biplots for (a) GPC, (b) MPT, and (c) MTI traits of MG RIL population, Table S1: Average daily temperatures (°C) per location during three growing seasons, Table S2: Rainfalls (mm) per location during three growing seasons, Table S3: Soil temperatures (°C) at 5 cm soil depth per location during three growing seasons, Table S4: Structure of nested effects in optimal models for each trait in both RIL populations examined, Table S5: Summary of parental means, RIL means and ranges, and rates of transgressive segregants within environments for seven quality traits assessed in BK RIL wheat population, Table S6: Summary of parental means, RIL means and ranges, and rates of transgressive segregants within environments for seven quality traits assessed in MG RIL wheat population, Table S7: Results of tests for IPCA terms for four mixograph traits containing missing values in the BK population, Table S8: Results of tests for IPCA terms for four mixograph traits containing missing values in the MG population, Table S9: "Winner" RILs within each environment for both populations examined.

Author Contributions: Conceptualization, I.P., J.G., and D.N.; methodology, R.Š.; formal analysis, I.P. and J.G.; investigation, R.Š.; writing—original draft preparation, I.P. and J.G.; writing—review and editing, J.G. and D.N.; visualization, I.P.; supervision, J.G. and D.N. All authors have read and agreed to the published version of the manuscript.

Funding: This study has been fully supported by the project KK.01.1.1.01.0005 Biodiversity and Molecular Plant Breeding, Centre of Excellence for Biodiversity and Molecular Plant Breeding (CoE CroP-BioDiv), Zagreb, Croatia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data were obtained from the Agricultural Institute Osijek and are available on request from the corresponding author with the permission of the Agricultural Institute Osijek.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bordes, J.; Ravel, C.; Le Gouis, J.; Lapierre, A.; Charmet, G.; Balfourier, F. Use of a global wheat core collection for association analysis of flour and dough quality traits. *J. Cereal Sci.* **2011**, *54*, 137–147. [CrossRef]
2. Payne, P.I. Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. *Annu. Rev. Plant Physiol.* **1987**, *38*, 141–153. [CrossRef]
3. Shewry, P.R. Improving the Protein Content and Quality of Temperate Cereals: Wheat, Barley and Rye. In *Impacts of Agriculture on Human Health and Nutrition*; Cakmak, I., Welch, R., Eds.; USDA, ARS, U.S. Plant, Soil and Nutrition Laboratory, Cornell University: Ithaca, NY, USA, 2004; Volume 2.
4. Kristensen, P.S.; Jahoor, A.; Andersen, J.R.; Cericola, F.; Orabi, J.; Janss, L.L.; Jensen, J. Genome-wide association studies and comparison of models and cross-validation strategies for genomic prediction of quality traits in advanced winter wheat breeding lines. *Front. Plant Sci.* **2018**, *9*, 69. [CrossRef]
5. Laidig, F.; Piepho, H.P.; Rentel, D.; Drobek, T.; Meyer, U.; Huesken, A. Breeding progress, environmental variation and correlation of winter wheat yield and quality traits in German official variety trials and on-farm during 1983–2014. *Theor. Appl. Genet.* **2017**, *130*, 223–245. [CrossRef]
6. Elli, L.; Villalta, D.; Roncoroni, L.; Barisani, D.; Ferrero, S.; Pellegrini, N.; Bardella, M.T.; Valiante, F.; Tomba, C.; Carroccio, A.; et al. Nomenclature and diagnosis of gluten-related disorders: A position statement by the Italian Association of Hospital Gastroenterologists and Endoscopists (AIGO). *Dig. Liver Dis.* **2017**, *49*, 138–146. [CrossRef]
7. Shewry, P.R.; Tatham, A.S.; Barro, F.; Barcelo, P.; Lazzeri, P. Biotechnology of breadmaking: Unraveling and manipulating the multi-protein gluten complex. *Bio/Technology* **1995**, *13*, 1185–1190. [CrossRef]
8. Payne, P.I.; Nightingale, M.A.; Krattiger, A.F.; Holt, L.M. The Relationship between HMW Glutenin Subunit Composition and the Bread-making Quality of British-grown Wheat Varieties. *J. Sci. Food Agric.* **1987**, *40*, 51–65. [CrossRef]
9. MacRitchie, F. Physicochemical properties of wheat proteins in relation to functionality. *Adv. Food Nutr. Res.* **1992**, *36*, 1–87. [CrossRef]
10. Liu, G.; Zhao, Y.; Gowda, M.; Longin, C.F.H.; Reif, J.C.; Mette, M.F. Predicting hybrid performances for quality traits through genomic-assisted approaches in Central European wheat. *PLoS ONE* **2016**, *11*, e0158635. [CrossRef] [PubMed]
11. Lado, B.; Vázquez, D.; Quincke, M.; Silva, P.; Aguilar, I.; Gutiérrez, L. Resource allocation optimization with multi-trait genomic prediction for bread wheat (*Triticum aestivum* L.) baking quality. *Theor. Appl. Genet.* **2018**, *131*, 2719–2731. [CrossRef]
12. Hook, S.C.W. Specific weight and wheat quality. *J. Sci. Food Agric.* **1984**, *35*, 1136–1141. [CrossRef]
13. Békés, F. New aspects in quality related wheat research: 1. Challenges and achievements. *Cereal Res. Commun.* **2012**, *40*, 159–184. [CrossRef]
14. Hayes, B.J.; Panozzo, J.; Walker, C.K.; Choy, A.L.; Kant, S.; Wong, D.; Tibbits, J.; Daetwyler, H.D.; Rochfort, S.; Hayden, M.J.; et al. Accelerating wheat breeding for end-use quality with multi-trait genomic predictions incorporating near infrared and nuclear magnetic resonance-derived phenotypes. *Theor. Appl. Genet.* **2017**, *130*, 2505–2519. [CrossRef] [PubMed]
15. Graybosch, R.A.; Peterson, C.J.; Hareland, G.A.; Shelton, D.R.; Olewnik, M.C.; He, H.; Stearns, M.M. Relationships between small-scale wheat quality assays and commercial test bakes. *Cereal Chem.* **1999**, *76*, 428–433. [CrossRef]
16. Swanson, C. Testing of the quality of flour by the recording dough mixer. *Cereal Chem.* **1993**, *10*, 1–29.
17. Johnson, J.A.; Swanson, C.O.; Bayfield, E.G. The correlation of mixogram with baking results. *Cereal Chem.* **1943**, *20*, 625–644.
18. Gras, P.W.; O'Brien, L. Application of a 2-gram mixograph to early generation selection for dough strength. *Cereal Chem.* **1992**, *69*, 254–257.
19. Bustos-Korts, D.; Romagosa, I.; Borràs-Gelonch, G.; Slafer, G.; Eeuwijk, F. Genotype by Environment Interaction and Adaptation. In *Crop Science. Encyclopedia of Sustainability Science and Technology Series*; Savin, R., Slafer, G., Eds.; Springer: New York, NY, USA, 2019; pp. 29–71.
20. Grausgruber, H.; Oberforster, M.; Werteker, M.; Ruckenbauer, P.; Vollmann, J. Stability of quality traits in Austrian-grown winter wheats. *Field Crop. Res.* **2000**, *66*, 257–267. [CrossRef]
21. Robert, N.; Denis, J.B. Stability of baking quality in bread wheat using several statistical parameters. *Theor. Appl. Genet.* **1996**, *93*, 172–178. [CrossRef]
22. Battenfield, S.D.; Guzmán, C.; Chris Gaynor, R.; Singh, R.P.; Peña, R.J.; Dreisigacker, S.; Fritz, A.K.; Poland, J.A. Genomic selection for processing and end-use quality traits in the CIMMYT spring bread wheat breeding program. *Plant Genome* **2016**, *9*, 1–12. [CrossRef]
23. Peterson, C.J.; Graybosch, R.A.; Shelton, D.R.; Baenziger, P.S. Baking quality of hard winter wheat: Response of cultivars to environment in the Great Plains. *Euphytica* **1998**, *100*, 157–162. [CrossRef]
24. Simmonds, N.W. The relation between yield and protein in cereal grain. *J. Sci. Food Agric.* **1995**, *67*, 309–315. [CrossRef]
25. Groos, C.; Robert, N.; Bervas, E.; Charmet, G. Genetic analysis of grain protein-content, grain yield and thousand-kernel weight in bread wheat. *Theor. Appl. Genet.* **2003**, *106*, 1032–1040. [CrossRef]
26. Šimić, G.; Horvat, D.; Jurković, Z.; Drezner, G.; Novoselović, D.; Dvojković, K. The genotype effect on the ratio of wet gluten content to total wheat grain protein. *J. Cent. Eur. Agric.* **2006**, *7*, 13–18. [CrossRef]
27. Williams, R.M.; O'Brien, L.; Eagles, H.A.; Solah, V.A.; Jayasena, V. The influences of genotype, environment, and genotype × environment interaction on wheat quality. *Aust. J. Agric. Res.* **2008**, *59*, 95–111. [CrossRef]

28. Hernández-Espinosa, N.; Mondal, S.; Autrique, E.; Gonzalez-Santoyo, H.; Crossa, J.; Huerta-Espino, J.; Singh, R.P.; Guzmán, C. Milling, processing and end-use quality traits of CIMMYT spring bread wheat germplasm under drought and heat stress. *Field Crop. Res.* **2018**, *215*, 104–112. [CrossRef]
29. Drezner, G.; Gunjača, J.; Novoselović, D.; Horvat, D. Interpretation of GEI effect analysis for some agronomic and quality traits in ten winter wheat (*Triticum aestivum* L.) cultivars. *Cereal Res. Commun.* **2010**, *38*, 259–265. [CrossRef]
30. Gauch, H.G. *Statistical Analysis of Regional Yield Trials: AMMI Analysis of Factorial Designs*; Elsevier: Amsterdam, The Netherlands, 1992.
31. Mani, E. Molecular Dissection of Breadmaking Quality in Wheat (*Triticum aestivum* L.). Ph.D. Thesis, University of Pune, Pune, India, October 2007.
32. Mut, Z.; Aydin, N.; Orhan Bayramoglu, H.; Ozean, H. Stability of some quality traits in bread wheat (*Triticum aestivum*) genotypes. *J. Environ. Biol.* **2010**, *31*, 489–495.
33. Khan, M.A.U.; Mohammad, F.; Khan, F.U.; Ahmad, S.; Ullah, I. Additive main effect and multiplicative interaction analysis for grain yield in bread wheat. *J. Anim. Plant Sci.* **2020**, *30*, 677–684. [CrossRef]
34. Sardouei-Nasab, S.; Mohammadi-Nejad, G.; Nakhoda, B. Yield stability in bread wheat germplasm across drought stress and non-stress conditions. *Agron. J.* **2019**, *111*, 175–181. [CrossRef]
35. Khan, M.A.U.; Mohammad, F.; Khan, F.U.; Ahmad, S.; Raza, M.A.; Kamal, T. Comparison among different stability models for yield in bread wheat. *Sarhad J. Agric.* **2020**, *36*, 282–290. [CrossRef]
36. Rodriguez, M.; Rau, D.; Papa, R.; Attene, G. Genotype by environment interactions in barley (*Hordeum vulgare* L.): Different responses of landraces, recombinant inbred lines and varieties to Mediterranean environment. *Euphytica* **2008**, *163*, 231–247. [CrossRef]
37. Groos, C.; Bervas, E.; Charmet, G. Genetic analysis of grain protein content, grain hardness and dough rheology in a hard X hard bread wheat progeny. *J. Cereal Sci.* **2004**, *40*, 93–100. [CrossRef]
38. Krishnappa, G.; Ahlawat, A.K.; Shukla, R.B.; Singh, S.K.; Singh, S.K.; Singh, A.M.; Singh, G.P. Multi-environment analysis of grain quality traits in recombinant inbred lines of a biparental cross in bread wheat (*Triticum aestivum* L.). *Cereal Res. Commun.* **2019**, *47*, 334–344. [CrossRef]
39. Elangovan, M.; Dholakia, B.B.; Rai, R.; Lagu, M.D.; Tiwari, R.; Gupta, R.K.; Gupta, V.S. Mapping QTL associated with agronomic traits in bread wheat (*Triticum aestivum* L.). *J Wheat Res.* **2011**, *3*, 14–23.
40. Prashant, R.; Mani, E.; Rai, R.; Gupta, R.K.; Tiwari, R.; Dholakia, B.; Oak, M.; Röder, M.; Kadoo, N.; Gupta, V. Genotype × environment interactions and QTL clusters underlying dough rheology traits in *Triticum aestivum* L. *J. Cereal Sci.* **2015**, *64*, 82–91. [CrossRef]
41. Elangovan, M.; Rai, R.; Dholakia, B.B.; Lagu, M.D.; Tiwari, R.; Gupta, R.K.; Rao, V.S.; Röder, M.S.; Gupta, V.S. Molecular genetic mapping of quantitative trait loci associated with loaf volume in hexaploid wheat (*Triticum aestivum*). *J. Cereal Sci.* **2008**, *47*, 587–598. [CrossRef]
42. Horvat, D.; Drezner, G.; Jurković, Z.; Šimić, G.; Magdić, D.; Dvojković, K. The importance of high-molecular-weight glutenin subunits for wheat quality evaluation. *Poljoprivreda* **2006**, *12*, 53–57.
43. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020. Available online: http://www.R-project.org (accessed on 20 May 2021).
44. Butler, D.G.; Cullis, B.R.; Gilmour, A.R.; Gogel, B.J.; Thompson, R. *ASReml-R Reference Manual Version 4*; VSN International Ltd.: Hemel Hempstead, UK, 2017.
45. Brien, M.C. asremlPlus: Augments 'ASReml-R' in Fitting Mixed Models and Packages Generally in Exploring Prediction Differences. 2021. R Package Version 4.2-32. Available online: https://cran.r-project.org/web/packages/asremlPlus/index.html (accessed on 20 May 2021).
46. Gauch, H.G. Model Selection and Validation for Yield Trials with Interaction. *Biometrics* **1988**, *44*, 705–715. [CrossRef]
47. Gauch, H.G. A simple protocol for AMMI analysis of yield trials. *Crop Sci.* **2013**, *53*, 1860–1869. [CrossRef]
48. Paderewski, J. An R function for imputation of missing cells in two-way data sets by EM-AMMI algorithm. *Commun. Biometry Crop Sci.* **2013**, *8*, 60–69.
49. Forkman, J.; Piepho, H.P. Parametric bootstrap methods for testing multiplicative terms in GGE and AMMI models. *Biometrics* **2014**, *70*, 639–647. [CrossRef]
50. Piepho, H.P. Robustness of statistical tests for multiplicative terms in the additive main effects and multiplicative interaction model for cultivar trials. *Theor. Appl. Genet.* **1995**, *90*, 438–443. [CrossRef]
51. Tarakanovas, P.; Ruzgas, V. Additive main effect and multiplicative interaction analysis of grain yield of wheat varieties in Lithuania. *Agron. Res.* **2006**, *4*, 91–98.
52. Wei, T.; Simko, V. corrplot: Visualization of a Correlation Matrix 2017. R Package Version 0.84. Available online: https://cran.r-project.org/web/packages/corrplot/index.html (accessed on 20 May 2021).
53. Harrell, F.E. Hmisc: Harrell Miscellaneous. 2021. R Package Version 4.4-2. Available online: https://CRAN.R-project.org/package=Hmisc (accessed on 20 May 2021).
54. Wickham, H. Reshaping Data with the reshape Package. *J. Stat. Softw.* **2007**, *21*, 1–20. [CrossRef]
55. Wickham, H.; François, R.; Henry, L.; Müller, K. dplyr: A Grammar of Data Manipulation 2020. R Package Version 1.0.2. Available online: https://CRAN.R-project.org/package=dplyr (accessed on 20 May 2021).
56. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016.

57. Slowikowski, K. ggrepel: Automatically Position Non-Overlapping Text Labels with "ggplot2" 2020. R Package Version 0.8.2. Available online: https://CRAN.R-project.org/package=ggrepel (accessed on 20 May 2021).
58. Gollob, H.F. A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika* **1968**, *33*, 73–115. [CrossRef] [PubMed]
59. Turner, A.S.; Bradburne, R.P.; Fish, L.; Snape, J.W. New quantitative trait loci influencing grain texture and protein content in bread wheat. *J. Cereal Sci.* **2004**, *40*, 51–60. [CrossRef]
60. Singh, B.; Singh, N. Physico-chemical, water and oil absorption and thermal properties of gluten isolated from different indian wheat cultivars. *J. Food Sci. Technol.* **2006**, *43*, 251–255.
61. Campbell, K.G.; Finney, P.L.; Bergman, C.J.; Gualberto, D.G.; Anderson, J.A.; Giroux, M.J.; Siritunga, D.; Zhug, J.; Gendre, F.; Roué, C.; et al. Quantitative trait loci associated with milling and baking quality in a soft × hard wheat cross. *Crop Sci.* **2001**, *41*, 1275–1285. [CrossRef]
62. Zhang, Y.; Wu, Y.; Xiao, Y.; Yan, J.; Zhang, Y.; Zhang, Y.; Ma, C.; Xia, X.; He, Z. QTL mapping for milling, gluten quality, and flour pasting properties in a recombinant inbred line population derived from a Chinese soft × hard wheat cross. *Crop Pasture Sci.* **2009**, *60*, 587–597. [CrossRef]
63. Zheng, F.; Deng, Z.; Shi, C.; Zhang, X.; Tian, J. QTL Mapping for Dough Mixing Characteristics in a Recombinant Inbred Population Derived from a Waxy × Strong Gluten Wheat (*Triticum aestivum* L.). *J. Integr. Agric.* **2013**, *12*, 951–961. [CrossRef]
64. Gauch, H.G.; Zobel, R.W. Imputing missing yield trial data. *Theor. Appl. Genet.* **1990**, *79*, 753–761. [CrossRef] [PubMed]
65. Paderewski, J.; Rodrigues, P.C. The Usefulness of EM-AMMI to Study the Influence of Missing Data Pattern and Application to Polish Post-Registration Winter Wheat Data. *Aust. J. Crop Sci.* **2014**, *8*, 640–645.
66. Arciniegas-Alarcón, S.; García-Peña, M.; Rodrigues, P.C. New Multiple Imputation Methods for Genotype-by-Environment Data That Combine Singular Value Decomposition and Jackknife Resampling or Weighting Schemes. *Comput. Electron. Agric.* **2020**, *176*, 105617. [CrossRef]
67. Echeverry-Solarte, M.; Kumar, A.; Kianian, S.; Simsek, S.; Alamri, M.S.; Mantovani, E.E.; McClean, P.E.; Deckard, E.L.; Elias, E.; Schatz, B.; et al. New QTL alleles for quality-related traits in spring wheat revealed by RIL population derived from supernumerary × non-supernumerary spikelet genotypes. *Theor. Appl. Genet.* **2015**, *128*, 893–912. [CrossRef]
68. Goel, S.; Singh, K.; Singh, B.; Grewal, S.; Dwivedi, N.; Alqarawi, A.A.; Abd Allah, E.F.; Ahmad, P.; Singh, N.K. Analysis of genetic control and QTL mapping of essential wheat grain quality traits in a recombinant inbred population. *PLoS ONE* **2019**, *14*, e0200669. [CrossRef]
69. Jin, H.; Wen, W.; Liu, J.; Zhai, S.; Zhang, Y.; Yan, J.; Liu, Z.; Xia, X.; He, Z. Genome-wide QTL mapping for wheat processing quality parameters in a Gaocheng 8901/Zhoumai 16 recombinant inbred line population. *Front. Plant Sci.* **2016**, *7*, 1032. [CrossRef] [PubMed]
70. Merrick, L.F.; Lyon, S.R.; Balow, K.A.; Murphy, K.M.; Jones, S.S.; Carter, A.H. Utilization of evolutionary plant breeding increases stability and adaptation of winter wheat across diverse precipitation zones. *Sustainability* **2020**, *12*, 9728. [CrossRef]
71. Sjoberg, S.M.; Carter, A.H.; Steber, C.M.; Garland-Campbell, K.A. Unraveling complex traits in wheat: Approaches for analyzing genotype × environment interactions in a multienvironment study of falling numbers. *Crop Sci.* **2020**, *60*, 3013–3026. [CrossRef]

Supplementary Figures

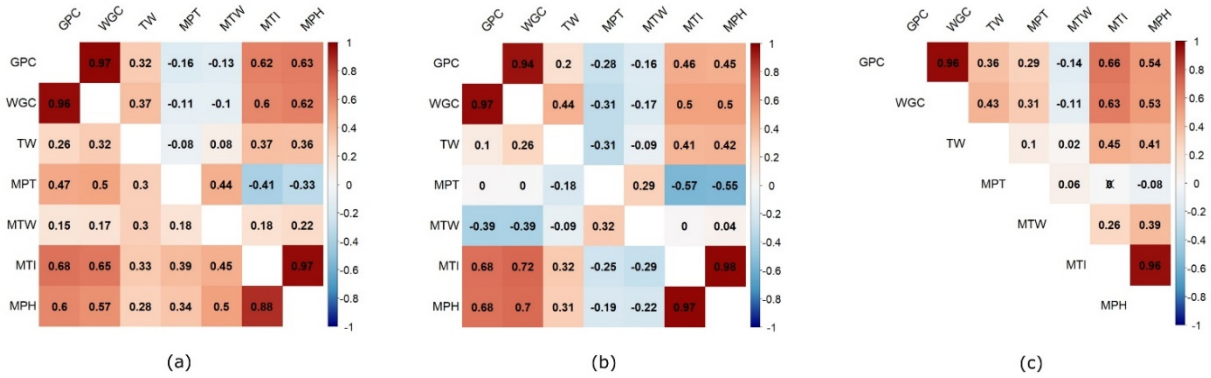


Figure S1: Pearson's correlation coefficients across environments for BK population. Years of experiment are marked as following: (a) 2009, (b) 2010, and (c) 2011. The data in the upper-right triangle represent the location Osijek, while those in the lower-left triangle represent location Slavonski Brod.

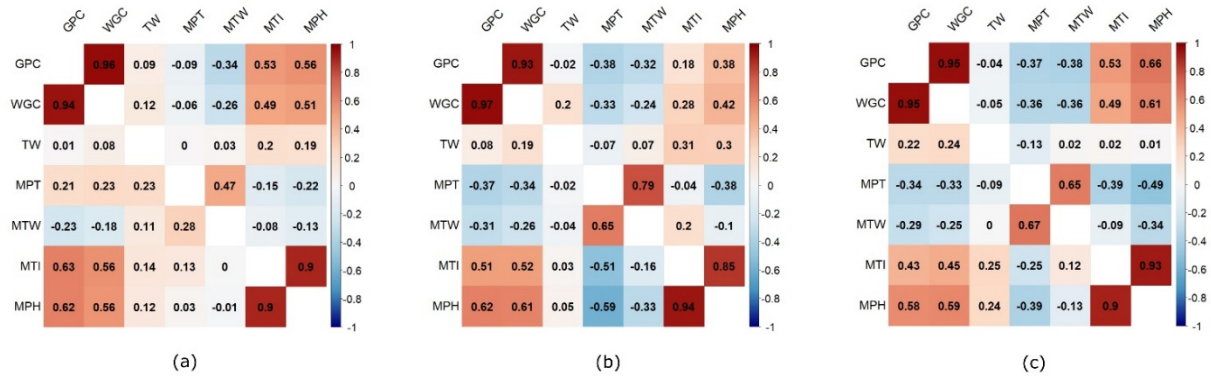


Figure S2: Pearson's correlation coefficients across environments for MG population. Years of experiment are marked as following: (a) 2009, (b) 2010, and (c) 2011. The data in the upper-right triangle represent the location Osijek, while those in the lower-left triangle represent location Slavonski Brod.

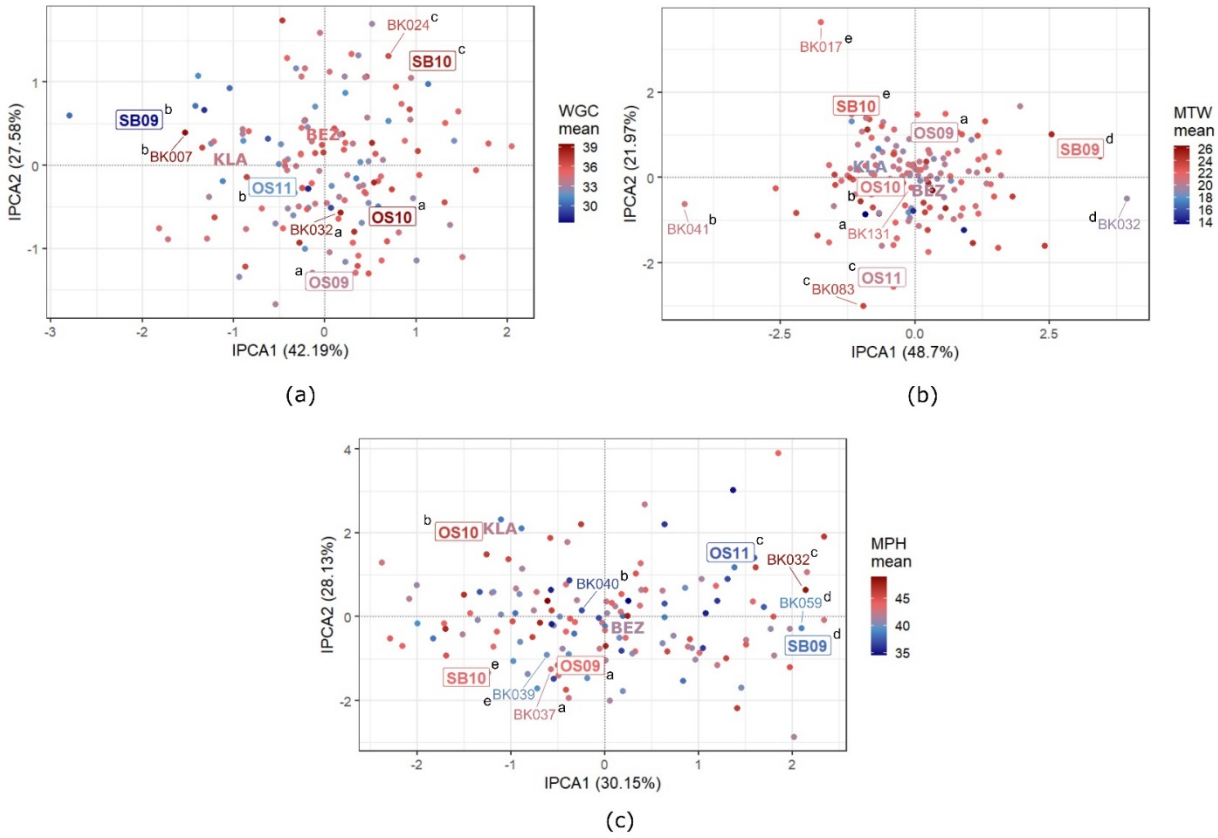


Figure S3: Modified AMMI2 biplots for (a) WGC, (b) MTW, and (c) MPH traits of BK RIL population. RILs are marked with dots, while environments and parental cultivars (Bezostaya-1 and Klara) are marked by abbreviated labels (with or without a frame, respectively). Within each environment one "winning" RIL is labeled and linked with its winning environment by a matching super/subscripted letter. The color indicates the mean value of the trait.

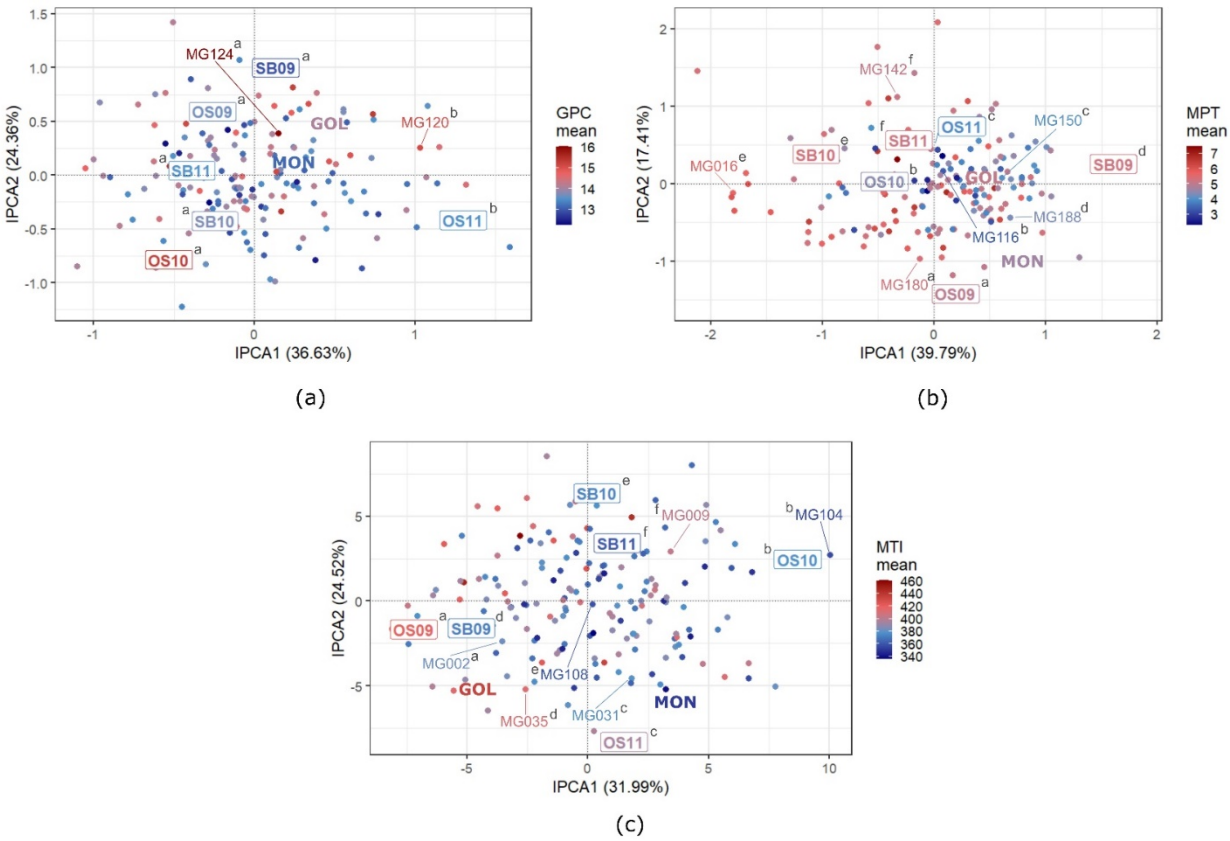


Figure S4: Modified AMMI2 biplots for (a) GPC, (b) MPT, and (c) MTI traits of MG RIL population. RILs are marked with dots, while environments and parental cultivars (Monika and Golubica) are marked by abbreviated labels (with or without a frame, respectively). Within each environment one "winning" RIL is labeled and linked with its winning environment by a matching super/subscripted letter. The color indicates the mean value of the trait.

Supplementary Tables

Table S1. Average daily temperatures (°C) per location during three growing seasons (from planting to harvesting).

Month	2008/09		2009/10		2010/11	
	Osijek	Slavonski Brod	Osijek	Slavonski Brod	Osijek	Slavonski Brod
X	13	12.2	11.5	11	9.1	8.9
XI	7.5	7.3	8.2	7.8	8.9	8.1
XII	3.8	3.8	3.1	3.4	0.2	0.3
I	-1.1	-1.6	-0.8	0	1.1	0.8
II	2.3	2.9	1.4	1.8	0.7	1
III	6.8	7.1	6.8	6.8	6.4	6.4
IV	14.6	14.2	12.4	12.3	13.2	13.1
V	18.3	18.1	16.5	16.2	16.7	16.3
VI	19.2	19.3	20.4	20.2	20.8	20.6
VII	23.2	22.6	23.2	22.7	22.2	22.6
Average	10.76	10.59	10.27	10.22	9.93	9.81

Table S2. Rainfalls (mm) per location during three growing seasons (from planting to harvesting).

Month	2008/09		2009/10		2010/11	
	Osijek	Slavonski Brod	Osijek	Slavonski Brod	Osijek	Slavonski Brod
X	34.4	44.2	55.3	45	67.1	58.2
XI	44.9	64.9	67.8	68.1	56.3	71.5
XII	40.5	46.8	100.8	106.1	73.5	68.5
I	60.3	62.9	83.9	79.5	23.6	27.2
II	28.6	26.2	58.6	81.8	18.4	16.6
III	26.5	41.4	22.2	49.5	37.1	35.9
IV	18.7	13	71.1	52.9	19.4	17.7
V	39.4	43.6	120.8	161.4	81.2	43.8
VI	62.8	103.8	234	176.9	49.9	47.4
VII	13.8	61.1	31.5	42.3	73.9	108.7
Total	369.9	507.9	846	863.5	500.4	495.5

Table S3. Soil temperatures (°C) at 5 cm soil depth per location during three growing seasons (from planting to harvesting).

Month	2008/09		2009/10		2010/11	
	Osijek	Slavonski Brod	Osijek	Slavonski Brod	Osijek	Slavonski Brod
X	13.9	13.9	13.4	13.3	10.6	11.5
XI	8.5	8.4	8.3	8.3	8.6	8.7
XII	4.5	4.2	4.8	4.8	2.3	2.2
I	0.4	0.8	1.6	1.7	1.6	1.5
II	2.8	3.2	1.9	2.0	1.5	2.0
III	6.9	7.3	6.7	7.2	6.2	6.9
IV	15.0	15.4	13.9	13.6	13.6	14.8
V	21.5	21.0	18.3	18.4	18.4	19.4
VI	22.8	21.7	22.8	21.9	23.8	24.1
VII	27.1	26.1	25.8	26.0	25.0	25.7
Average	12.3	12.2	11.8	11.7	11.2	11.7

Table S4. Structure of nested effects in optimal models for each trait in both RIL populations examined.

Trait	Population	Environments[1]		
		Rep	Row	Col
GPC	BK	3	All[2]	2, 4, 5
	MG	1, 3	All	1, 3, 4, 5
WGC	BK	Single[3]	All	4, 5
	MG	1, 3, 6	1, 3, 4, 5, 6	All
TW	BK	Single	1, 2, 4, 5	2, 4, 5
	MG	1, 2, 5, 6	2, 4, 5, 6	2, 4, 6
MPT	BK	2, 5	2, 3, 4, 5	4, 5
	MG	1, 3, 5	4, 5, 6	1, 3, 4, 6
MTW	BK	2	4	2, 4
	MG	Single	Single	None[4]
MTI	BK	Single	2, 3, 4, 5	2, 4, 5
	MG	1, 2, 3, 6	1, 3, 4, 5, 6	3, 4
MPH	BK	3, 5	2, 3, 4, 5	4, 5
	MG	Single	All	All

[1] Numbers indicate environments for which this effect was included (1 – OS09, 2 – OS10, 3 – OS11,
4 – SB09, 5 – SB10, 6 – SB11).
[2] „All" indicates different effect in all environments.
[3] „Single" indicates unique effects for all environments.
[4] „None" indicates that model did not include this effect.

Table S5. Summary of parental means, RIL means and ranges, and rates of transgressive segregants within environments for seven quality traits assessed in BK RIL wheat population.

Trait	Parental cultivars Mean		RILs			Positive transgressive segregants[1]		Negative transgressive segregants[2]	
	Bezostaya-1	Klara	Min	Mean	Max	N	%	N	%
OS09 environment									
GPC	14.1	13.9	11.9	14.1	15.9	74	51.8	57	39.9
WGC	33.4	33.7	25.6	33.6	39.5	73	51.1	63	44.1
TW	81.4	82.0	77.0	81.6	84.5	64	44.8	57	39.9
MPT	5.8	5.2	3.4	5.8	7.7	74	51.8	43	30.1
MTW	19.3	16.9	12.4	20.4	27.8	104	72.7	8	5.6
MTI	373.8	320.5	249.8	363.5	463.4	58	40.6	13	9.1
MPH	45.6	39.4	29.8	43.9	55.1	49	34.3	21	14.7
OS10 environment									
GPC	16.0	14.9	13.9	15.7	17.5	47	32.9	29	20.3
WGC	39.9	37.7	32.9	39.4	43.9	64	44.8	33	23.1
TW	73.9	76.8	64.8	75.3	79.8	53	37.1	38	26.6
MPT	6.9	4.6	2.5	5.8	9.2	27	18.9	21	14.7
MTW	20.8	20.8	13.9	21.2	29.6	71	49.7	71	49.7
MTI	378.0	456.1	311.3	409.5	504.9	14	9.8	27	18.9
MPH	43.0	52.5	36.7	46.2	58.9	11	7.7	37	25.9
OS11 environment									
GPC	12.7	12.7	11.0	12.5	14.6	54	37.8	88	61.5
WGC	32.5	32.5	26.4	31.6	37.3	48	33.6	94	65.7
TW	85.1	84.2	78.9	84.1	86.9	30	20.9	69	48.3
MPT	4.2	2.9	1.3	4.6	9.9	82	57.3	18	12.6
MTW	18.3	17.7	14.7	20.4	30.3	108	75.5	22	15.4
MTI	342.7	359.8	259.1	328.6	418.4	24	16.8	97	67.8
MPH	37.7	38.9	28.4	37.1	51.2	43	30.1	86	60.1
SB09 environment									
GPC	12.5	12.6	10.6	12.4	14.9	49	34.3	86	60.1
WGC	27.7	28.7	20.5	27.4	35.6	41	28.7	81	56.6
TW	79.6	79.3	73.7	79.3	82.5	66	46.2	67	46.9
MPT	5.0	4.9	1.3	5.1	8.2	79	55.2	60	41.9
MTW	21.1	16.3	10.8	21.7	35.1	67	46.9	3	2.1
MTI	304.1	290.9	237.3	323.5	411.3	96	67.1	26	18.2
MPH	36.7	34.9	25.5	38.0	49.8	88	61.5	39	27.3
SB10 environment									
GPC	16.0	15.1	12.6	15.3	17.2	38	26.6	58	40.6
WGC	40.1	37.6	30.7	38.6	43.9	44	30.8	47	32.9
TW	75.5	76.9	65.6	77.4	81.5	94	65.7	17	11.9
MPT	4.3	5.5	1.8	5.8	10.0	86	60.1	24	16.8
MTW	20.9	22.0	12.1	22.9	33.3	87	60.8	30	21.0
MTI	399.2	408.4	284.6	381.5	464.6	41	28.7	93	65.0
MPH	44.8	46.3	32.1	43.6	53.2	42	29.4	82	57.3

[1] RILs that exhibited values higher than the parental cultivar with higher trait value.

[2] RILs that exhibited values lower than the parental cultivar with lower trait value.

Table S6. Summary of parental means, RIL means and ranges, and rates of transgressive segregants within environments for seven quality traits assessed in MG RIL wheat population.

Trait	Parental cultivars Mean		RILs			Positive transgressive segregants[1]		Negative transgressive segregants[2]	
	Monika	Golubica	Min	Mean	Max	N	%	N	%
OS09 environment									
GPC	13.1	14.0	11.6	13.7	16.1	55	31.8	50	28.9
WGC	28.2	34.0	25.4	31.6	39.0	39	22.5	23	13.3
TW	81.5	84.3	77.9	82.1	84.3	0	0.00	44	25.4
MPT	6.4	5.1	2.0	4.9	7.6	31	17.9	92	53.2
MTW	10.0	22.2	6.5	19.6	35.4	53	30.6	19	11.0
MTI	359.4	510.4	329.9	420.8	527.8	4	2.3	10	5.8
MPH	42.1	60.8	36.9	49.3	73.2	8	4.6	20	11.6
OS10 environment									
GPC	14.8	15.1	13.5	15.3	17.3	106	61.3	51	29.5
WGC	32.3	39.5	32.0	38.4	43.9	59	34.1	1	0.6
TW	73.5	73.8	65.7	72.4	77.6	54	31.2	110	63.6
MPT	3.2	4.7	1.9	4.4	8.5	64	37.0	33	19.1
MTW	5.8	9.2	4.4	9.4	23.4	55	31.8	13	7.5
MTI	334.0	394.6	301.6	377.2	470.8	47	27.2	19	11.0
MPH	44.2	47.1	34.8	45.8	59.3	52	30.0	73	42.2
OS11 environment									
GPC	13.0	14.3	11.1	13.5	16.2	28	16.2	59	34.1
WGC	33.5	36.7	29.6	34.6	42.2	29	16.8	54	31.1
TW	84.6	84.5	81.9	84.4	86.2	74	42.8	89	51.5
MPT	2.9	3.4	1.8	3.9	7.4	126	72.8	22	12.7
MTW	11.6	13.3	6.9	15.2	23.7	127	73.4	23	13.3
MTI	442.8	485.4	327.1	394.0	489.3	1	0.6	162	93.6
MPH	53.8	58.0	36.1	46.8	62.9	2	1.2	157	90.8
SB09 environment									
GPC	12.6	14.0	10.9	12.9	15.6	16	9.3	61	35.3
WGC	26.1	32.7	21.8	28.0	36.4	8	4.6	40	23.1
TW	78.6	80.7	75.4	79.1	81.6	13	7.5	52	30.1
MPT	6.1	5.7	2.0	5.3	8.5	52	30.1	100	57.8
MTW	13.2	26.6	4.8	20.2	45.9	24	13.9	26	15.0
MTI	314.6	442.3	298.7	370.9	472.7	4	2.3	4	2.3
MPH	36.3	53.7	33.8	43.6	58.4	5	2.9	7	4.1
SB10 environment									
GPC	12.8	13.6	12.1	13.7	16.3	97	56.1	19	11.0
WGC	30.0	34.9	29.6	34.6	42.1	73	42.2	3	1.7
TW	78.4	79.1	75.3	79.0	82.4	85	49.1	51	29.5
MPT	2.8	4.3	1.6	5.3	9.1	122	70.5	7	4.1
MTW	8.1	21.2	5.1	14.3	30.1	14	8.1	30	17.3
MTI	347.5	436.9	285.5	372.4	502.9	9	5.2	44	25.4
MPH	39.2	50.5	30.8	43.1	60.8	16	9.3	47	27.2
SB11 environment									
GPC	12.0	13.3	11.7	13.5	16.4	103	59.5	5	2.9
WGC	29.4	33.8	28.0	33.7	41.4	84	48.6	9	5.2
TW	79.6	79.9	76.1	80.3	83.0	112	64.7	47	27.2
MPT	6.1	6.1	2.1	5.2	8.0	51	29.5	118	68.2
MTW	7.1	15.3	4.9	12.9	27.2	57	32.9	24	13.9
MTI	286.7	363.2	286.7	360.6	451.4	76	43.9	0	0.00
MPH	33.2	42.9	33.0	43.7	54.5	90	52.0	1	0.6

[1] RILs that exhibited values higher than the parental cultivar with higher trait value.

[2] RILs that exhibited values lower than the parental cultivar with lower trait value.

Table S7. Results of tests for IPCA terms for four mixograph traits containing missing values in the BK population. Tests were conducted on the imputed data sets based on different AMMI models.

		MPT						MTW						MTI						MPH					
		Sum of Squares		Simple bootstrap		FR-test		Sum of Squares		Simple bootstrap		FR-test		Sum of Squares		Simple bootstrap		FR-test		Sum of Squares		Simple bootstrap		FR-test	
	IPCA	IPCA SS	%	*T*	*p value*	*F*	*p value*	IPCA SS	%	*T*	*p value*	*F*	*p value*	IPCA SS	%	*T*	*p value*	*F*	*p value*	IPCA SS	%	*T*	*p value*	*F*	*p value*
AMMI0	1	400.2	45.5	0.41	0.000	2.04	0.000	1581.1	53.3	0.46	0.000	2.45	0.000	127383.3	38.2	0.32	0.269	1.36	0.009	1993.2	37.4	0.30	0.650	1.25	0.042
	2	268.8	30.6	0.47	0.001	1.74	0.000	797.4	26.9	0.42	0.084	1.44	0.005	111755.5	33.5	0.41	0.202	1.36	0.015	1869.9	35.1	0.40	0.281	1.32	0.024
	3	209.9	23.9	0.69	0.000	2.22	0.000	587.2	19.8	0.54	0.627	1.16	0.188	94516.5	28.3	0.59	0.120	1.40	0.024	1470.9	27.6	0.53	0.761	1.12	0.254
AMMI1	1	400.7	45.6	0.41	0.000	2.05	0.000	1712.4	55.5	0.48	0.000	2.68	0.000	127752.5	38.3	0.32	0.255	1.37	0.009	2008.5	37.7	0.30	0.606	1.27	0.036
	2	268.4	30.5	0.47	0.001	1.73	0.000	798.9	25.9	0.43	0.061	1.46	0.003	111537.2	33.4	0.41	0.210	1.35	0.016	1863.0	34.9	0.40	0.296	1.32	0.026
	3	210.2	23.9	0.69	0.000	2.23	0.000	573.7	18.6	0.54	0.684	1.14	0.215	94565.2	28.3	0.59	0.120	1.40	0.024	1463.8	27.4	0.53	0.792	1.11	0.275
AMMI2	1	400.2	45.5	0.41	0.000	2.04	0.000	1774.3	56.4	0.49	0.000	2.77	0.000	126404.6	37.5	0.31	0.370	1.33	0.015	1999.9	37.5	0.30	0.630	1.26	0.039
	2	268.8	30.6	0.47	0.001	1.74	0.000	800.3	25.4	0.43	0.059	1.47	0.003	116513.4	34.5	0.42	0.109	1.42	0.007	1866.4	35.0	0.40	0.286	1.32	0.025
	3	209.9	23.9	0.69	0.000	2.22	0.000	571.6	18.2	0.54	0.707	1.13	0.226	94460.8	28.0	0.59	0.120	1.40	0.024	1466.6	27.5	0.53	0.777	1.11	0.265
AMMI3	1	411.3	45.7	0.41	0.000	2.06	0.000	1654.6	54.6	0.47	0.000	2.58	0.000	183845.8	45.5	0.39	0.000	1.87	0.000	1970.9	36.6	0.30	0.772	1.22	0.063
	2	266.9	29.6	0.46	0.004	1.66	0.000	797.7	26.3	0.43	0.064	1.46	0.004	125226.6	31.0	0.44	0.030	1.52	0.001	1899.7	35.3	0.40	0.272	1.33	0.023
	3	222.8	24.7	0.71	0.000	2.37	0.000	576.7	19.0	0.54	0.662	1.15	0.204	94788.9	23.5	0.59	0.111	1.40	0.022	1514.9	28.1	0.54	0.639	1.16	0.193

Table S8. Results of tests for IPCA terms for four mixograph traits containing missing values in the MG population. Tests were conducted on the imputed data sets based on different AMMI models.

		MPT						MTW						MTI						MPH					
		Sum of Squares		Simple bootstrap		FR-test		Sum of Squares		Simple bootstrap		FR-test		Sum of Squares		Simple bootstrap		FR-test		Sum of Squares		Simple bootstrap		FR-test	
	IPCA	IPCA SS	%	*T*	*p value*	*F*	*p value*	IPCA SS	%	*T*	*p value*	*F*	*p value*	IPCA SS	%	*T*	*p value*	*F*	*p value*	IPCA SS	%	*T*	*p value*	*F*	*p value*
AMMI0	1	337.1	43.7	0.39	0.000	1.85	0.000	3879.2	48.5	0.43	0.000	2.23	0.000	254432.1	36.3	0.32	0.000	1.37	0.004	4704.7	34.3	0.30	0.001	1.26	0.025
	2	154.3	20.0	0.29	0.822	0.80	0.953	1560.6	19.5	0.31	0.390	0.87	0.855	195682.8	27.9	0.36	0.001	1.11	0.210	3670.4	26.8	0.34	0.028	1.00	0.508
	3	151.4	19.6	0.40	0.254	0.66	0.997	1392.2	17.4	0.39	0.330	0.64	0.998	137255.2	19.6	0.40	0.301	0.65	0.998	2770.1	20.2	0.38	0.561	0.61	0.999
	4	128.3	16.6	0.56	0.266	2.55	0.000	1160.2	14.5	0.54	0.556	2.35	0.000	113363.1	16.2	0.54	0.571	2.34	0.000	2567.0	18.7	0.57	0.156	2.67	0.000
AMMI1	1	349.8	44.7	0.40	0.000	1.93	0.000	3882.6	48.6	0.43	0.000	2.24	0.000	255288.9	36.4	0.32	0.000	1.38	0.003	4708.1	34.3	0.30	0.001	1.26	0.025
	2	154.0	19.7	0.29	0.817	0.80	0.953	1558.5	19.5	0.31	0.402	0.87	0.858	195528.2	27.9	0.36	0.001	1.11	0.211	3671.4	26.8	0.34	0.027	1.00	0.506
	3	149.7	19.2	0.40	0.302	0.65	0.998	1392.1	17.4	0.39	0.332	0.64	0.998	136972.1	19.5	0.40	0.311	0.65	0.998	2776.9	20.2	0.38	0.535	0.61	0.999
	4	128.3	16.4	0.56	0.283	2.54	0.000	1160.0	14.5	0.54	0.563	2.34	0.000	114204.8	16.3	0.54	0.509	2.38	0.000	2563.2	18.7	0.57	0.155	2.68	0.000
AMMI2	1	352.0	44.9	0.40	0.000	1.94	0.000	3996.4	48.3	0.43	0.000	2.23	0.000	255111.6	36.4	0.32	0.000	1.38	0.003	4709.2	34.3	0.30	0.001	1.27	0.024
	2	154.0	19.6	0.29	0.815	0.80	0.952	1726.7	20.9	0.31	0.397	0.96	0.612	195556.4	27.9	0.36	0.001	1.11	0.210	3671.7	26.8	0.34	0.027	1.00	0.505
	3	149.7	19.1	0.40	0.301	0.65	0.998	1391.3	16.8	0.39	0.333	0.64	0.998	137022.0	19.5	0.40	0.313	0.65	0.998	2779.2	20.3	0.38	0.538	0.62	0.999
	4	128.2	16.4	0.56	0.284	2.53	0.000	1165.1	14.1	0.54	0.561	2.38	0.000	114022.2	16.3	0.54	0.511	2.37	0.000	2562.1	18.7	0.57	0.155	2.68	0.000
AMMI3	1	538.3	54.5	0.49	0.000	2.87	0.000	3959.1	48.3	0.43	0.000	2.23	0.000	254779.7	36.3	0.32	0.000	1.38	0.004	4724.7	34.3	0.30	0.001	1.27	0.024
	2	171.1	17.3	0.31	0.291	0.89	0.816	1687.1	20.6	0.32	0.104	0.94	0.677	195614.1	27.9	0.36	0.001	1.11	0.210	3675.9	26.7	0.34	0.029	0.99	0.511
	3	152.9	15.5	0.40	0.196	0.67	0.996	1391.5	17.0	0.39	0.326	0.64	0.998	137126.5	19.6	0.40	0.305	0.65	0.998	2813.2	20.4	0.39	0.456	0.62	0.999
	4	126.2	12.8	0.56	0.327	2.50	0.000	1164.4	14.2	0.54	0.507	2.38	0.000	113691.5	16.2	0.54	0.544	2.36	0.000	2551.2	18.5	0.57	0.159	2.67	0.000
AMMI4	1	351.5	44.4	0.39	0.000	1.91	0.000	4205.2	48.6	0.44	0.000	2.27	0.000	256897.9	36.5	0.32	0.000	1.39	0.003	4718.0	34.3	0.30	0.001	1.27	0.024
	2	158.9	20.1	0.29	0.680	0.82	0.928	1897.0	21.9	0.35	0.005	1.06	0.333	195329.5	27.7	0.36	0.001	1.10	0.220	3674.1	26.7	0.34	0.028	1.00	0.507
	3	152.8	19.3	0.40	0.210	0.66	0.996	1390.8	16.1	0.39	0.330	0.64	0.998	136641.2	19.4	0.39	0.346	0.64	0.998	2798.2	20.4	0.39	0.483	0.62	0.999
	4	127.6	16.1	0.56	0.272	2.55	0.000	1167.3	13.5	0.55	0.494	2.39	0.000	116014.4	16.5	0.55	0.434	2.42	0.000	2555.0	18.6	0.57	0.154	2.68	0.000

Table S9. „Winner" RILs within each environment for both populations examined.

Environment	Trait						
	GPC	**WGC**	**TW**	**MPT**	**MTW**	**MTI**	**MPH**
	BK population						
OS09	BK032	BK032	BK012	BK042	BK131	BK059	BK037
OS10	BK032	BK032	BK037	BK042	BK041	BK040	BK040
OS11	BK007	BK007	BK012	BK131	BK083	BK057	BK032
SB09	BK007	BK007	BK012	BK124	BK032	BK059	BK059
SB10	BK034	BK024	BK048	BK042	BK017	BK039	BK039
	MG population						
OS09	MG124	MG124	MG089	MG180	MG169	MG002	MG166
OS10	MG124	MG084	MG016	MG116	MG116	MG104	MG027
OS11	MG120	MG193	MG155	MG150	MG035	MG031	MG031
SB09	MG124	MG124	MG192	MG188	MG016	MG035	MG096
SB10	MG124	MG084	MG124	MG016	MG001	MG108	MG108
SB11	MG124	MG124	MG087	MG142	MG136	MG009	MG009

Prilog 2. Znanstveni rad: Plavšin, I., Gunjača, J., Šatović, Z., Šarčević, H., Ivić, M., Dvojković, K., Novoselović, D. (2021). An Overview of Key Factors Affecting Genomic Selection for Wheat Quality Traits. Plants 10 (4): 745. doi:10.3390/plants1004074

plants

MDPI

Review

An Overview of Key Factors Affecting Genomic Selection for Wheat Quality Traits

Ivana Plavšin [1,2], Jerko Gunjača [2,3], Zlatko Šatović [2,4], Hrvoje Šarčević [2,3], Marko Ivić [1], Krešimir Dvojković [1,*] and Dario Novoselović [1,2]

1 Department for Cereal Breeding and Genetics, Agricultural Institute Osijek, Južno predgrađe 17, 31000 Osijek, Croatia; ivana.plavsin@poljinos.hr (I.P.); marko.ivic@poljinos.hr (M.I.); dario.novoselovic@poljinos.hr (D.N.)
2 Centre of Excellence for Biodiversity and Molecular Plant Breeding (CoE CroP-BioDiv), Svetošimunska cesta 25, 10000 Zagreb, Croatia; jgunjaca@agr.hr (J.G.); zsatovic@agr.hr (Z.Š.); hsarcevic@agr.hr (H.Š.)
3 Department of Plant Breeding, Genetics and Biometrics, Faculty of Agriculture, University of Zagreb, Svetošimunska cesta 25, 10000 Zagreb, Croatia
4 Department of Seed Science and Technology, Faculty of Agriculture, University of Zagreb, Svetošimunska cesta 25, 10000 Zagreb, Croatia
* Correspondence: kresimir.dvojkovic@poljinos.hr; Tel.: +385-31-515-512

Abstract: Selection for wheat (*Triticum aestivum* L.) grain quality is often costly and time-consuming since it requires extensive phenotyping in the last phases of development of new lines and cultivars. The development of high-throughput genotyping in the last decade enabled reliable and rapid predictions of breeding values based only on marker information. Genomic selection (GS) is a method that enables the prediction of breeding values of individuals by simultaneously incorporating all available marker information into a model. The success of GS depends on the obtained prediction accuracy, which is influenced by various molecular, genetic, and phenotypic factors, as well as the factors of the selected statistical model. The objectives of this article are to review research on GS for wheat quality done so far and to highlight the key factors affecting prediction accuracy, in order to suggest the most applicable approach in GS for wheat quality traits.

Keywords: wheat quality; genomic selection; GEBV; prediction accuracy; training population; validation population; heritability

Citation: Plavšin, I.; Gunjača, J.; Šatović, Z.; Šarčević, H.; Ivić, M.; Dvojković, K.; Novoselović, D. An Overview of Key Factors Affecting Genomic Selection for Wheat Quality Traits. *Plants* **2021**, *10*, 745. https://doi.org/10.3390/plants10040745

Academic Editor: Gianni Barcaccia

Received: 18 March 2021
Accepted: 8 April 2021
Published: 11 April 2021

1. Introduction

According to estimates of the International Maize and Wheat Improvement Center (CIMMYT), the need for wheat (*Triticum aestivum* L.) and its products could increase by at least 50% until 2050 [1] as a result of extensive human population growth and dietary changes. Given the accelerated growth of the world's population and the increased need for food production, the greatest emphasis in wheat breeding is placed on increasing grain yield. However, an increase in yield usually entails a decrease in protein content or grain quality [2,3]. Therefore, in breeding programs strong emphasis should be placed on improving grain quality [4]. In the context of wheat quality, the most important traits are grain protein content (GPC) and gluten content (GC), as they directly affect the technological properties of flour and dough [5–7]. The majority of these traits are characterized mostly by low heritability due to strong environmental impact [8–10].

The extensive development of high-throughput genotyping in the last couple of decades has led to the increasing use of molecular markers in plant breeding, which enabled the development of prediction methods based only on marker information such as genomic selection (GS) [11,12]. The first GS studies in wheat were published more than a decade ago [13,14]. The results of these studies showed that models based on genomic markers outperform models based only on pedigree relationships and that GS could successfully enhance rates of genetic gain, which provided a strong foundation for

further research on GS in wheat. Later studies also showed that, if the traits of interest are complex and influenced by many quantitative trait loci (QTLs) each controlling a small proportion of phenotypic variation, GS will be more relevant than marker-assisted selection (MAS) [15,16].

Currently, the majority of researchers of GS in wheat consider grain yield and disease resistance as key traits for successful wheat production [17–21]. Such a strong focus on grain yield is understandable from a point of view where grain yield is not improving fast enough to fill the gap between production and projected demands in the near future [22]. Considering wheat's role as the main ingredient in many different products fundamental to the nutrition of humankind, equal emphasis should be given to the quality traits, especially those related to the end-use quality of wheat products. An overview of GS research studies for traits such as grain yield, *Fusarium* head blight, stripe and brown rust resistance, plant height, days to heading, and preharvest sprouting (PHS) tolerance is given by Rutkoski et al. [23]. Despite their importance in the context of nutrition, research on GS for wheat quality traits is still scarce. In this context, the objectives of the present study are to review research on GS for wheat quality conducted so far and to highlight the key factors affecting GS accuracy in order to suggest the most applicable approach in GS for wheat quality traits.

2. Genomic Selection and Prediction Models Used

Genomic selection is one of the newly developed methods that enables the prediction of breeding values of individuals by simultaneously incorporating all available marker information into a model [12]. Unlike other molecular breeding methods, GS does not require the identification of markers associated with QTLs of traits of interest. GS attempts to capture total additive genetic variance based on the sum of the effects of a large number of genetic markers, encompassing all QTLs that contribute to trait variability [24]. Therefore, the underlying genetic control in GS is not necessarily known. In GS, training population (TP) is genotyped using one of the methods of high-throughput genotyping and phenotyped for desired traits in a target set of environments. Obtained data are used to train a model that will be applied to the breeding population (BP) of unphenotyped individuals (selection candidates) to calculate their genomic-estimated breeding values (GEBVs) using only the marker scores [12] (Figure 1). The most important advantage of GS over phenotypic selection (PS) is the increase of genetic gain due to the shortening of the selection cycle in breeding process by reducing the need for phenotyping [25,26].

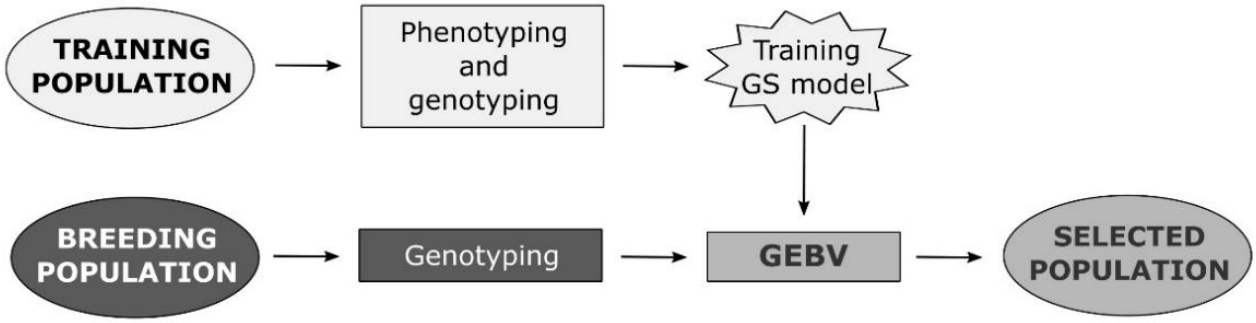


Figure 1. Flow diagram of a plant breeding program based on genomic selection.

High-throughput genotyping generates a large amount of marker data, which are then used in GS. When the number of predictor variables (markers) is much greater than the number of observations (phenotypic values), the result is an infinite number of marker effect estimates. In order to reduce the problem of highly dimensioned data, different parametric and nonparametric models have been developed and used in GS (Figure 2).

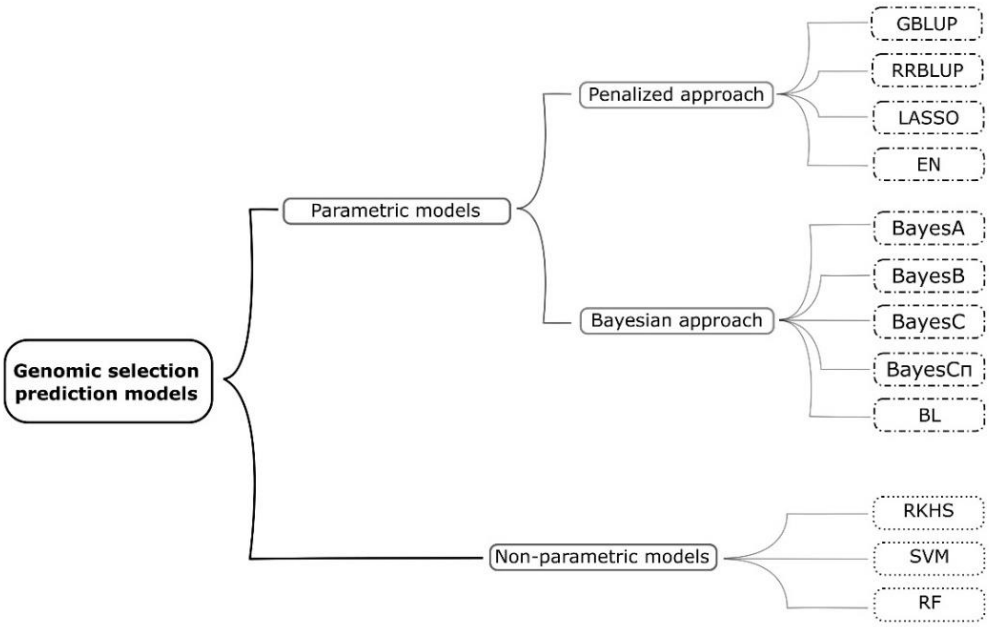


Figure 2. Classification of the most frequently used genomic selection prediction models.

The prediction models differ mainly in the assumptions of the distribution of the marker effects, i.e., the assumptions of how the marker effects contribute to the overall variance [12]. Detailed features of GS models and their (dis)advantages are already given elsewhere [27–29], so they will be just briefly discussed here. Genomic best linear unbiased predictor (GBLUP) uses a genomic-estimated relationship matrix, assigns equal variance to all markers, and assumes that they are equally contributing to the trait of interest [29]. Ridge-regression best linear unbiased predictor (RRBLUP) assumes that all markers have common variance but allows that markers have unequal effects. RRBLUP shrinks all marker effects equally towards zero, regardless of the size of their effect, which can possibly lead to overshrinking of large-effect loci [12,15,30]. On the other hand, Bayesian alphabet models (BayesA, BayesB, BayesC, BayesCπ) assign different types of prior distribution to marker effects, thus having a more realistic assumption of marker effects [12,13,21]. LASSO (Least absolute shrinkage and selector operator) and Bayesian LASSO (BL) models use both variable selection and shrinkage methods, with the difference that BL additionally applies prior exponential distribution on marker variances [28]. Like RRBLUP, Bayesian ridge regression (BRR) shrinks marker effects equally towards zero but additionally uses prior Gaussian distribution for marker effects [31]. Elastic net (EN) uses two penalty methods—the LASSO and ridge regression, which results in averaging markers that are highly correlated and then using the averaged gene for the model [32]. Random forest (RF) and support vector machines (SVM) are nonparametric models based on supervised learning methods, which have been proved to be effective in detecting interactions between markers [27,33]. Reproducing kernels Hilbert spaces regression (RKHS) is another nonparametric model that is able to capture nonadditive effects [34].

3. Factors Affecting Prediction Accuracies of Genomic Selection in Wheat

The prediction accuracy of GS is commonly estimated using cross-validation, in which a set of individuals that are both genotyped and phenotyped is divided into a training set (training population) and validation set (validation population, VP), with marker effects estimated in the training set used to predict GEBVs for the validation set [35]. The accuracy is then measured as the correlation between GEBVs and true breeding values (observed phenotypes) of individuals from the validation set. Prediction accuracy of GS is influenced by various molecular, genetic, and phenotypic factors, as well as the features of the selected statistical model. Genetic factors include the distribution and strength of

linkage disequilibrium (LD) between markers and QTL, marker collinearity, population size and structure, differences in allele frequency between TP and VP, etc. [17,30,36–38]. Phenotypic factors include factors related to traits themselves, such as heritability and phenotypic variance of the TP [24,39,40]. Other factors that affect the accuracy of the prediction are the number and type of molecular markers, the similarity of the TP and the VP, TP size, and the features of the selected statistical model [28,39,41–45].

Three major factors that affect the GS accuracy are population structure, TP size, and marker density, the effects of which are highly interrelated [46,47]. Population structure can give rise to a false association between a marker and QTL, thus causing structure-generated LD, which can lead to overestimated genomic heritability and biased GS prediction accuracies [37,48]. Meuwissen et al. [49] estimated the minimum number of markers required to reach high prediction accuracy (approximately 0.9) when using unrelated individuals to be equal to 10 times the product of the effective population size and the genome size in Morgans ($10 \times N_e \times L$), while the minimum size of the population was estimated to be $2 \times N_e \times L$. In the case of wheat (the L of which is approximately 30 Morgans) and assumption of $N_e = 50$, that would amount to 15,000 markers and 3000 individuals in an unrelated population. However, those estimations were obtained using simulations, while empirical studies on wheat have reported acceptable accuracies for a much lower number of markers and smaller populations, depending on the population structure [17,21,30]. As in other species, studies on wheat also showed that larger TP reduces bias and decreases the marker effect variance, thus resulting in higher prediction accuracy [31]. The interrelatedness of marker density and population structure seems to play an important role in optimizing GS in wheat. Namely, it has been shown that the higher the relatedness between TP and VP, the smaller the response to increased marker density [48].

4. Overview of Genomic Selection Research for Wheat Quality Improvement

The first GS study for wheat quality traits was published in 2011 [21]. The study was conducted in a population composed of multiple wheat families and showed that GS accuracy surpasses MAS accuracy for wheat quality traits by roughly 30% and that GS was about 95% as accurate as PS. Authors also concluded that, regardless of the inferiority when compared to PS, GS has the potential to increase genetic gain per unit of time and costs when applied in a breeding program.

A study by Heffner et al. [30] based on two biparental wheat populations, examined the potential of GS to predict nine wheat quality traits. The authors of the study have found that the mean accuracy obtained by GS was 1.4 times greater than the one obtained by MAS, but that both GS and MAS were inferior to PS. However, those findings were expected due to the polygenic nature and medium to high heritability of all examined traits. Liu et al. [50] reported that, when predicting wheat hybrid performance for seven quality traits, GS extensively outperformed MAS, while giving similar results as PS in the case of higher relatedness of TP and VP. It was only in the case of lower relatedness of TP and VP that GS was preferred over PS, thus emphasizing the importance of additive effects in wheat quality traits. According to Battenfield et al. [51], genetic gain was 1.4 to 2.7 times higher when comparing GS to PS for processing and end-use quality traits since GS requires only marker data and a much larger population can be genotyped than phenotyped for wheat quality traits. Michel et al. [52] investigated the use of GS for predicting dough rheological traits in early generations and proved its substantial benefit over MAS. These findings imply that GS can capture more of the genetic variance of wheat quality traits when compared to MAS since it considers both small effect loci in addition to major QTLs. Nevertheless, all of the above-mentioned studies showed that the accuracy of GS for wheat quality traits is under the influence of many factors, with underlined heritability of the trait, genetic relationship between TP and VP, and size of the TP being the most important driving forces of GS accuracy. A summary of the most relevant GS studies for wheat quality traits, with an overview of factors affecting prediction accuracy covered, is given in Table 1.

Table 1. Overview of references for genomic selection studies covering wheat quality traits.

Reference	Quality Traits Examined [1]	Population Type and Size [2]	Platform and Number of Markers [2]	GS Prediction Model [3]	Factors Affecting Prediction Accuracy Examined [2]				Comparison to Other Types of Selection [2]	Single-Trait (ST) or Multitrait (MT) Analysis
					Selected Model	TP Size	TP/VP Relatedness	Marker Density		
[30]	TW, PHS, FY, KH, LA-SRC, NaCO-SRC, Suc-SRC, H_2O-SRC	2 biparental populations (209/174 DHs)	399 multiple platforms/574 DArTs	RRBLUP, BayesCπ	Yes	Yes	No	Yes	Yes (MAS and PS)	ST
[21]	TW, PHS, FY, FP, LA-SRC, NaCO-SRC, Suc-SRC, H_2O-SRC, KH	374 lines	1158 DArTs	RRBLUP, BayesA, BayesB, BayesCπ	Yes	Yes	No	Yes	Yes (MAS and PS)	ST
[51]	TKW, TW, GPC, FY, FP, LV, KH, SDS sedimentation, mixograph and alveograph traits	5520 lines	3075 SNPs	RRBLUP, GAUSS, PLSR, EN, RF	Yes	Yes	No	No	Yes	ST
[53]	TKW, TW, GPC, KH, SDS sedimentation	8416/2403 landrace accessions	~23,000/~33,000 DArT SNPs	GBLUP	No	Yes	Yes	No	No	ST
[54]	TW, FY, FP, LA-SRC, NaCO-SRC, Suc-SRC, H_2O-SRC, KH	273 elite lines and cultivars	3919/13,198 SNPs	RRBLUP, BRR, RKHS, EN	Yes	Yes	Yes	Yes	No	ST
[50]	TKW, TW, GPC, GC, SC, KH, Zeleny sedimentation	135 inbred lines,1604 hybrids	17,372 SNPs	RRBLUP, W-BLUP, BayesCπ	Yes	Yes	Yes	Yes	Yes (PS)	ST
[55]	GPC, PY	659 lines	9500 DArT SNPs	RRBLUP	No	Yes	Yes	No	No	ST
[56]	TKW, TW, GPC, FY, FP, SC, amylose content, FN, LV, LT, MIXT, KH, starch damage, viscosity, farinograph and extensograph traits	2076 varieties and synthetic derivative lines	51,208 SNPs	Multivariate model	No	No	No	No	No	ST + MT
[57]	GPC, farinograph, extensograph, and alveograph traits	128 DHs	6600 DArT SNPs	RRBLUP	No	No	No	No	No	MT
[58]	GPC, gluten index, alveograph traits	170 varieties and advanced lines, 154 DHs	9752/5153 SNPs	RRBLUP, GBLUP, BayesA, BayesB, BL, RKHS, MT-BayesA, MT-Matrix, MT-SI	Yes	No	Yes	Yes	No	ST + MT
[59]	TKW, TW, GPC, FN, Zeleny sedimentation	635 lines (159 full-sib families)	10,802 SNPs	GBLUP, BL	Yes	Yes	Yes	Yes	No	ST
[60]	TW, GPC, WGC, SV, alveograph and mixograph traits	495 lines	6655 SNPs	BRR, Bayes multivariate Gaussian model	No	Yes	No	No	No	ST + MT

Table 1. *Cont.*

Reference	Quality Traits Examined [1]	Population Type and Size [2]	Platform and Number of Markers [2]	GS Prediction Model [3]	Factors Affecting Prediction Accuracy Examined [2]				Comparison to Other Types of Selection [2]	Single-Trait (ST) or Multitrait (MT) Analysis
					Selected Model	TP Size	TP/VP Relatedness	Marker Density		
[52]	GPC, farinograph and extensograph traits	840 lines	4598 DArT SNPs	RRBLUP, W-BLUP	Yes	No	No	No	Yes (MAS)	ST + MT
[61]	TKW, GPC, mixograph, farinograph, and extensograph traits	57 cultivars and lines	7588 SNPs	RRBLUP, BayesA BayesB, BL, BRR	Yes	No	No	No	No	ST + MT
[62]	TKW, GPC, SDS sedimentation	282 DHs	7426 SNPs	RRBLUP, BL, RF, RKHS	Yes	Yes	No	No	Yes (PS)	ST
[63]	TKW, TW, GPC, FY, FP, FS, LV, MIXT, KH, grain color, alveograph traits	3485 lines	78,606 SNPs	GBLUP, BayesB	Yes	No	Yes	Yes	No	ST
[64]	TKW, GPC, FN, Zeleny sedimentation	1152 lines	11,058 SNPs	GBLUP, Bayesian SNP-BLUP	Yes	Yes	Yes	No	Yes (MAS)	ST + MT
[65]	FY, alveograph traits	635 lines (159 full-sib families)	10,802 SNPs	GBLUP, BL	Yes	Yes	Yes	No	No	ST
[66]	GPC, PY, extensograph and farinograph traits	480 lines	7300 DArT SNPs	GBLUP, W-BLUP	Yes	No	No	No	No	ST + MT
[67]	TKW, TW, GPC, FP, LV, KH, SDS sedimentation, mixograph and alveograph traits	~1400 lines	78,606 SNPs before filtering *	BMTME, MTR	Yes	No	No	No	No	MT
[68]	GPC, Zeleny sedimentation	1325 lines	9290 SNPs	RRBLUP, BL	Yes	No	No	No	No	ST

* Final number of markers used for analysis is not mentioned. [1] TKW—thousand-kernel weight, TW—test weight, GPC—grain protein content, FY—flour yield, FP—flour protein, FS—flour sedimentation, WGC—wet gluten content, PY—protein yield, GC—gluten content, KH—kernel hardness, SC—starch content, FN—falling number, LV—loaf volume, LT—loaf texture, MIXT—mixing time, SV—sedimentation volume, PHS—preharvest sprouting, LA-SRC—lactic acid solvent retention capacity, NaCO-SRC—sodium carbonate solvent retention capacity, H_2O-SRC—water solvent retention capacity, Suc-SRC—sucrose solvent retention capacity, SDS—sodium dodecyl sulfate. [2] DH—double haploid, SNP—single nucleotide polymorphism, TP—training population, VP—validation population, PS—phenotypic selection, MAS—marker-assisted selection, ST—single-trait, MT—multitrait. [3] GS—genomic selection, RRBLUP—ridge regression best linear unbiased prediction, GBLUP—genomic best linear unbiased prediction, BL—Bayesian least absolute shrinkage and selector operator (LASSO), BRR—Bayesian ridge regression, GAUSS—Gaussian kernel, PLSR—partial least squares regression, RKHS—reproducing kernel Hilbert space, EN—elastic net, W-BLUP—weighted best linear unbiased prediction, BMTME—Bayesian multitrait multienvironment, MTR—multitrait ridge regression, MT-SI—multitrait selection index, RF—random forest.

4.1. Effect of Training Population Size

As early as with the first studies of GS for wheat quality traits, it was demonstrated that TP size (N_{TP}) significantly impacts the GS accuracy. The average accuracy for nine wheat quality traits was reported to be roughly 1.6-fold higher for N_{TP} = 96 compared to N_{TP} = 24 when applied to a biparental population [30]. A similar pattern was observed in a study by Heffner et al. [21] in which a population consisting of multiple wheat families was used to predict some quality and agronomic traits. Increasing N_{TP} from 96 to 288 resulted in an overall increase in accuracy by approximately 30%. It is interesting to note that in order to achieve approximately the same GS prediction accuracy, a TP that is three times greater should be used in multifamily populations compared to biparental populations (mean accuracies of 0.58 and 0.52 correspond to N_{TP} = 96 and N_{TP} = 288 in biparental and multifamily population, respectively). The positive influence of an increased number of lines in TP was observed for GPC and protein yield (PY) traits [55], where maximum accuracy was reached at maximum TP size (N_{TP} = 240) and amounted to 0.51 and 0.16 for GPC and PY, respectively. When investigating the influence of using different proportions of the entire population as TP (20–80%), Hu et al. [62] concluded that average prediction accuracy benefited from larger TP size when predicting wheat quality traits such as SDS (sodium dodecyl sulfate) sedimentation volume and thousand-kernel weight (TKW). In agreement with previous studies, Kristensen et al. [59] reported that the highest accuracies were recorded for all examined traits in the case of LOO (leave-one-out) type of cross-validation (the largest possible TP scenario), while the *k*-fold cross-validation proved that the use of smaller TP resulted in slightly lower GS prediction abilities. Similar results for flour yield (FY) and alveograph traits were reported by Kristensen et al. [65].

Overall, the size of the TP depends on the genetic relatedness between TP and VP. The more related the two populations are, the smaller the size of the TP will be needed to obtain satisfying GS prediction accuracies for wheat quality traits [54]. Battenfield et al. [51] also reported enhanced accuracy as a result of increasing TP size and random assignment of full-sibs to TP and VP, therefore, creating a greater genetic relationship. Considering that the phenotyping of wheat quality traits can be both costly and time-consuming, designing a TP that at the same time maximizes genetic diversity and enhances GS accuracy, while being small enough to achieve rapid phenotyping, is key for the successful implementation of GS in a breeding program [53].

4.2. Relatedness of Training and Validation Population

As for the other wheat traits [69], it has also been observed in other studies of GS for wheat quality traits that, in order to achieve high GS accuracy, TP and VP have to be closely related. In research by Liu et al. [50] three scenarios with low, intermediate, and high relatedness of TP and VP were created in order to predict seven quality traits of wheat hybrids. As expected, results showed that GS accuracy enhances with an increase of population relatedness, regardless of the prediction model used. However, for the scenario of high relatedness, GS and PS resulted in similar prediction accuracies, suggesting that for highly related populations, PS could be hardly outperformed by GS, whereas in lowly related populations, GS will be a method of choice. Poor prediction accuracies were observed for quality traits in durum wheat when the performance of one population type (doubled-haploid) was predicted based on another population type (breeding panel consisted of varieties and advanced lines) [58]. Kristensen et al. [59] used different types of cross-validations to study the impact of genetic distance of TP and VP on GS prediction accuracy. In LOO cross-validation, the GEBV of each individual is predicted based on the rest of the population, thus representing a scenario where the size of TP and genetic relatedness between TP and VP is as large as possible. Leave-family-out (LFO) cross-validation represents a scenario where different levels of genetic relatedness of TP and VP are present since the GEBV of individuals in each family is predicted based on the remaining families in a given population. Comparing LOO and LFO (lower relatedness), Kristensen et al. [59] concluded that genetic relatedness had a bigger impact on GS accuracy

than the size of TP. The predictive abilities decreased the most in the case of GPC (0.5 and 0.2 for LOO and LFO, respectively) when increasing the genetic distance between populations, while the smallest impact of increased genetic distance was recorded in the case of Zeleny sedimentation (0.79 and 0.68 for LOO and LFO, respectively). Similar results were reported for FY and alveograph traits, where the decrease of GS accuracy in a range of 24% to 35% was observed when comparing LOO and LFO cross-validation methods [65], and for Zeleny sedimentation, GPC, TKW, and test weight (TW) [64], suggesting that genetic composition of TP is crucial for achieving accurate genomic predictions.

Prediction accuracies for GPC and PY traits showed a strong bias when predicting within the breeding cycle (lower relatedness) compared to predicting between-cycle (higher relatedness). According to Michel et al. [55], the highest bias for GPC was 86%, whereas PY was overestimated in a range from 17% all the way up to 712%. A study by Juliana et al. [63] has provided evidence that for traits with lower heritability the influence of using lowly related populations will be even more pronounced. Therefore, in order to achieve reliable predictions, the use of a diverse TP is recommended.

4.3. Effect of Marker Density

Studies of GS for wheat quality traits investigating the effect of marker density (i.e., number of markers, N_M) all led to the same conclusion that the accuracy of the prediction enhances with increasing marker density until it reaches a plateau, after which a further increase in marker density has no effect on accuracy [30]. Since required marker density is primarily determined by the extent of LD in the examined population, it is assumed that lower marker density will be sufficient for closely related populations (e.g., biparental population) than for distant populations to achieve satisfying GS prediction accuracy. In a study conducted using two biparental wheat populations [30], average GS prediction reached a plateau at N_M = 256, after which a slight drop in accuracy was observed (N_M = 384), while in a multifamily approach [21] increasing N_M from 192 to 1158 increased GS accuracy by approximately 10%, after which response reached a plateau. Huang et al. [54] reported no significant differences in GS accuracy when using the complete set of markers (N_M = 13,198) and a subset of 3919 markers, implying that lower N_M is already sufficient for predicting quality traits in wheat elite lines and varieties. Juliana et al. [63] confirmed those findings using subsets of a marker data set that contained less than 70%, 50%, and 10% missing data, which corresponded to a scenario of high coverage (N_M = 16,072), moderate coverage (N_M = 9285), and low marker coverage (N_M = 2253). They concluded that marker density had a minimal impact on GS accuracy, suggesting that when a genomic resolution is reached in a high LD species (i.e., wheat), marker density no longer represents a limiting factor.

The interdependence of marker density and relatedness of TP and VP in the context of GS was illustrated in a study by Liu et al. [50] in which three scenarios representing low, intermediate, and high relatedness were used. In the case of lowly related TP and VP, the plateau was reached after ~3000 markers, whereas in the case of intermediate and highly related TP and VP, the plateau was reached at ~2000 and ~500 markers, respectively.

4.4. Effect of Heritability of the Trait

Numerous studies up to date showed that GS accuracy is strongly influenced by heritability, i.e., the fraction of the phenotypic variance of the trait due to genetic variance. Although there is no unambiguous categorization, the majority of studies on wheat categorize heritability values as low (<0.4), moderate (0.4–0.7), and high (>0.7) [21,30]. Generally, traits with high heritability show high GS accuracy and vice versa. The predictive ability of GS for wheat quality traits parallels their heritability which is often showed to be moderate to high. An overview of heritability and GS prediction accuracy ranges reported for some wheat quality traits is given in Table 2.

Table 2. Overview of heritability and GS prediction accuracy reported in studies covering wheat quality traits.

Reference	Quality Traits Examined [1]	Heritability Type	Heritability Strength	Heritability Range	GS Prediction Accuracy Range [3]
[21]	PHS, GPC, TW, Suc-SRC, LA-SRC, KH, FY	broad-sense	high	0.71–0.93	0.45–0.76
[30]	TW, PHS, FY, KH, LA-SRC, NaCO-SRC, Suc-SRC, H_2O-SRC	broad-sense	moderate—high	0.67–0.95	0.27–0.74
[51]	TKW, TW, GPC, FY, FP, SDS sedimentation, KH, LV, mixograph and alveograph traits	narrow-sense	moderate	0.41–0.68	0.42–0.71
[54]	TW, FY, FP, KH, LA-SRC, NaCO-SRC, Suc-SRC, H_2O-SRC	alternative calculation for unbalanced data [2]	high	0.75–0.95	0.31–0.67
[50]	TKW, TW, GPC, GC, SC, KH, Zeleny sedimentation	broad-sense	moderate—high	0.63–0.96	0.35–0.96 [4]
[57]	GPC, farinograph, extensograph, and alveograph traits	alternative calculation for unbalanced data [2]	moderate—high	0.69–0.83	0.16–0.61 [4]
[59]	TKW, TW, GPC, FN, Zeleny sedimentation	narrow-sense	moderate—high	0.56–0.81	0.2–0.79
[60]	TW, GPC, WGC, SV, alveograph and mixograph traits	broad-sense	moderate	0.36–0.64	0.24–0.43 [4]
[52]	GPC, farinograph and extensograph traits	narrow-sense	moderate	0.4–0.66	0.3–0.53
[61]	TKW, GPC, mixograph, farinograph, and extensograph traits	broad-sense	high	0.78–0.93	0.25–0.42
[65]	FY, alveograph traits	narrow-sense	moderate—high	0.38–0.72	0.3–0.79
[68]	GPC, SC, Zeleny sedimentation	narrow-sense	low—moderate	0.35–0.62	0.1–0.3

[1] TKW—thousand-kernel weight, TW—test weight, GPC—grain protein content, FY—flour yield, FP—flour protein, WGC—wet gluten content, GC—gluten content, KH—kernel hardness, SC—starch content, FN—falling number, LV—loaf volume, SV—sedimentation volume, PHS—preharvest sprouting, LA-SRC—lactic acid solvent retention capacity, NaCO-SRC—sodium carbonate solvent retention capacity, H_2O-SRC—water solvent retention capacity, Suc-SRC—sucrose solvent retention capacity, SDS—sodium dodecyl sulfate. [2] According to Piepho and Möhring [70]. [3] Accuracy across all used models or scenarios. [4] Accuracy of single-trait genomic selection model.

Studies on wheat quality traits showed that heritability was the main factor that affected the accuracy of GS [61]. Interestingly, not all highly heritable traits showed high GS accuracy. While for most of the highly heritable traits (TW, sucrose solvent retention capacity (Suc-SRC), water solvent retention capacity (H_2O-SRC), and lactic-acid solvent retention capacity (LA-SRC)), mean GS accuracy across the four models used was 0.6 and higher, for FY and KH, accuracies were 0.45 and 0.38, respectively, despite their heritability values being > 0.9 [54]. Low heritability traits would require larger TP in order to attain the same prediction accuracy as in the case of traits with moderate to high heritability [56]. According to the reported heritabilities (Table 2), it is highly unlikely that the heritability will present a limiting factor in GS for wheat quality traits.

4.5. Effect of Model Used

A broad range of models can be used to predict the phenotypic performance of wheat, but the performance of each model is interrelated with the genetic architecture of the examined trait and relatedness of TP and VP. As it is presented in Table 1, the majority of GS studies for wheat quality traits used GBLUP and RRBLUP models, the performance of which was usually compared to one of the Bayesian models.

Little or no difference in prediction accuracy was detected between RRBLUP and Bayesian models in a study by Heffner et al. [21], which suggested that all examined quality traits were controlled by many QTLs of small effect. RRBLUP was comparable by Bayesian models for highly polygenic quality traits in biparental populations while being surpassed in the case of populations with a high genetic variance of examined traits [30]. RRBLUP and BayesCπ showed no significant differences when predicting hybrid performance [50]. BL gave similar or slightly higher prediction accuracies than GBLUP for GPC, TW, TKW, falling number (FN), FY, and alveograph traits, while the biggest difference was recorded in the case of Zeleny sedimentation [59,65]. Those findings may be due to the better performance of Bayesian models in case of lower relatedness of TP and VP, and in case of traits controlled by few major QTLs, since they shrink small effects stronger while shrinking large effects much weaker. Zeleny sedimentation has been proved to be controlled by few QTLs of large effect, hence obtaining higher GS accuracies when Bayesian models were used [59]. Similar results were observed when comparing RRBLUP and BL models for GPC and Zeleny sedimentation [68]. Hu et al. [62] compared two nonparametric (RKHS and RF) and two parametric models (RRBLUP and BL) when predicting SDS sedimentation volume, GPC, and TKW, and concluded that their performance was strongly influenced by prediction scenario (predicting within the same year and across years where years represented different drought conditions). Namely, nonparametric models outperformed parametric in the cross-year prediction which represented a more realistic setting, while in the same-year prediction average performances of RF, RKHS, and RRBLUP were similar, with RF showing significant variations among growing seasons. Only a study by Battenfield et al. [51] showed that when GS accuracy was obtained by cross-validation, Gaussian kernel (GAUSS) was the best model for predicting all quality traits within a population consisting of multiple families, thus outperforming EN, partial least square regression (PLSR), and RRBLUP.

Bayesian models usually require longer computation time compared to GBLUP or RRBLUP [12,58,62] but show no clear superiority over the other models across wheat quality traits [61,68], i.e., the accuracy of GS for wheat quality traits was generally not under the large influence of prediction model applied. Therefore, RRBLUP showed to be a model of choice in the majority of GS studies for wheat quality traits [54,58] due to its robustness and shorter computational time [55].

5. Multitrait Genomic Selection

Wheat quality traits can often be hard to improve, since they usually require a large amount of flour and/or labor to be invested, thus limiting the size of the TP that can be phenotyped which leads to insufficient GS accuracy. Incorporating additional phenotypic information in the multitrait approach for GS could help to overcome the problem of potentially low GS accuracy obtained for wheat quality traits. Multitrait GS data obtained utilizing rapid quality tests are used for predicting parameters of more laborious wheat quality tests. Rapid tests such as near-infrared (NIR) and nuclear magnetic resonance (NMR) methods are less labor-intensive and require a small amount of flour. It has been proved that incorporating NIR and NMR data into the multitrait approach increases the accuracy of GS for some wheat quality traits (accuracy ranged between 0 and 0.47, and between 0 and 0.69 in a single-trait approach and multitrait approach, respectively) [56]. Incorporating easily obtained gluten peak indices into multitrait GS analysis improved average prediction accuracy by roughly 20% in comparison to single-trait GS for dough rheology traits [57]. Including metabolomics data in GS resulted in increased accuracies for some wheat quality traits (GPC, GC, FN, FY, Zeleny sedimentation, KH) compared to GS based on DArT markers only [71]. According to Haile et al. [58], the multitrait approach resulted in higher prediction accuracy only in the case of yield, whereas for quality traits, all single-trait models applied gave better prediction accuracy compared to the multitrait approach. Lado et al. [60] showed that no multitrait model used performed better than a single-trait model, but that using highly correlated traits in multitrait GS

for wheat quality allows reduction of TP up to 30% without significantly affecting the predictive ability of the model. Further research studies showed that using different GS indices in simultaneous selection for yield and wheat quality traits still does not outperform single-trait prediction for GPC, PY, and the dough rheological traits, but suggested that simultaneous improvement of yield and wheat quality should target protein quality, rather than GPC [66]. A significant gain of multitrait approach is expected only for low heritable traits that are incorporated with high heritable traits, between which high genetic correlation exists [64]. Data for traits incorporating together in a multitrait analysis must be already available or easy to obtain on a large number of samples in a short period of time [67].

6. Conclusions

Due to the complex nature of inheritance for the majority of wheat quality traits, GS seems to be the method of choice because it simultaneously accounts for small and medium effect loci as well as for major QTLs. Numerous studies in the last decade proved that GS has sufficient accuracy for implementation in the breeding programs targeting wheat quality. Genomic selection can be helpful in predicting the performance of lines in early generations and preselecting high-performing lines, boosting trait stability, and efficiently selecting superior genotypes for wheat quality traits. There is some evidence that GS could also be used to address one of the biggest problems in wheat breeding—how to simultaneously select for grain yield and quality traits since the existence of a strong negative correlation between those traits is well known and documented. Nevertheless, before implementing GS in the breeding for wheat quality traits, some limitations considering trait heritability, genetic relationship between TP and VP, and size of the TP must be taken into account.

Author Contributions: I.P., J.G., and D.N. wrote the draft for this review. The draft was carefully reviewed by Z.Š., H.Š., M.I., and K.D. All authors have read and agreed to the published version of the manuscript.

Funding: This study has been fully supported by the project KK.01.1.1.01.0005 Biodiversity and Molecular Plant Breeding, Centre of Excellence for Biodiversity and Molecular Plant Breeding (CoE CroP-BioDiv), Zagreb, Croatia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. CIMMYT. International Maize and Wheat Improvement Center. Available online: https://www.cimmyt.org/work/wheat-research/ (accessed on 16 March 2021).
2. Shewry, P.R. Improving the protein content and quality of temperate cereals: Wheat, barley and rye. In *Impacts of Agriculture on Human Health and Nutrition*; Cakmak, I., Welch, R., Eds.; USDA, ARS, U.S. Plant, Soil and Nutrition Laboratory, Cornell University: Ithaca, NY, USA, 2004; Volume 2.
3. Simmonds, N.W. The Relation between Yield and Protein in Cereal Grain. *J. Sci. Food Agric.* **1995**, *67*, 309–315. [CrossRef]
4. Guzman, C.; Peña, R.J.; Singh, R.; Autrique, E.; Dreisigacker, S.; Crossa, J.; Rutkoski, J.; Poland, J.; Battenfield, S. Wheat Quality Improvement at CIMMYT and the Use of Genomic Selection on It. *Appl. Transl. Genom.* **2016**, *11*, 3–8. [CrossRef]
5. Groos, C.; Robert, N.; Bervas, E.; Charmet, G. Genetic Analysis of Grain Protein-Content, Grain Yield and Thousand-Kernel Weight in Bread Wheat. *Theor. Appl. Genet.* **2003**, *106*, 1032–1040. [CrossRef]
6. Payne, P.I.; Nightingale, M.A.; Krattiger, A.F.; Holt, L.M. The Relationship between HMW Glutenin Subunit Composition and the Bread-Making Quality Of British-grown Wheat Varieties. *J. Sci. Food Agric.* **1987**, *40*, 51–65. [CrossRef]
7. Wieser, H. Chemistry of Gluten Proteins. *Food Microbiol.* **2007**, *24*, 115–119. [CrossRef] [PubMed]
8. Grausgruber, H.; Oberforster, M.; Werteker, M.; Ruckenbauer, P.; Vollmann, J. Stability of Quality Traits in Austrian-Grown Winter Wheats. *Field Crop. Res.* **2000**, *66*, 257–267. [CrossRef]
9. Robert, N.; Denis, J.B. Stability of Baking Quality in Bread Wheat Using Several Statistical Parameters. *Theor. Appl. Genet.* **1996**, *93*, 172–178. [CrossRef] [PubMed]

10. Simmonds, N.W. Genotype (G), Environment (E) and GE Components of Crop Yields. *Exp. Agric.* **1981**, *17*, 355–362. [CrossRef]
11. Crossa, J.; Beyene, Y.; Semagn, K.; Pérez, P.; Hickey, J.M.; Chen, C.; de los Campos, G.; Burgueño, J.; Windhausen, V.S.; Buckler, E.; et al. Genomic Prediction in Maize Breeding Populations with Genotyping-by-Sequencing. *G3 Genes Genom. Genet.* **2013**, *3*, 1903–1926. [CrossRef] [PubMed]
12. Meuwissen, T.H.E.; Hayes, B.J.; Goddard, M.E. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **2001**, *157*, 1819–1829. [CrossRef] [PubMed]
13. de los Campos, G.; Naya, H.; Gianola, D.; Crossa, J.; Legarra, A.; Manfredi, E.; Weigel, K.; Cotes, J.M. Predicting Quantitative Traits with Regression Models for Dense Molecular Markers and Pedigree. *Genetics* **2009**, *182*, 375–385. [CrossRef] [PubMed]
14. Crossa, J.; de los Campos, G.; Pérez, P.; Gianola, D.; Burgueño, J.; Araus, J.L.; Makumbi, D.; Singh, R.P.; Dreisigacker, S.; Yan, J.; et al. Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers. *Genetics* **2010**, *186*, 713–724. [CrossRef] [PubMed]
15. Heffner, E.L.; Sorrells, M.E.; Jannink, J.L. Genomic Selection for Crop Improvement. *Crop Sci.* **2009**, *49*, 1–12. [CrossRef]
16. Heslot, N.; Jannink, J.L.; Sorrells, M.E. Perspectives for Genomic Selection Applications and Research in Plants. *Crop Sci.* **2015**, *55*, 1–12. [CrossRef]
17. Arruda, M.P.; Brown, P.J.; Lipka, A.E.; Krill, A.M.; Thurber, C.; Kolb, F.L. Genomic Selection for Predicting Fusarium Head Blight Resistance in a Wheat Breeding Program. *Plant Genom.* **2015**, *8*, 1–12. [CrossRef]
18. Crossa, J.; Pérez, P.; Hickey, J.M.; Burgueño, J.; Ornella, L.; Cerón-Rojas, J.; Zhang, X.; Dreisigacker, S.; Babu, R.; Li, Y.; et al. Genomic Prediction in CIMMYT Maize and Wheat Breeding Programs. *Heredity* **2014**, *112*, 48–60. [CrossRef]
19. Poland, J.; Endelman, J.B.; Dawson, J.; Rutkoski, J.; Wu, S.; Manes, Y.; Dreisigacker, S.; Crossa, J.; Sánchez-Villeda, H.; Sorrells, M.; et al. Genomic Selection in Wheat Breeding Using Genotyping-by-Sequencing. *Plant Genom.* **2012**, *5*, 103–113. [CrossRef]
20. Rutkoski, J.; Benson, J.; Jia, Y.; Brown-Guedira, G.; Jannink, J.L.; Sorrells, M. Evaluation of Genomic Prediction Methods for Fusarium Head Blight Resistance in Wheat. *Plant Genom.* **2012**, *5*, 51–61. [CrossRef]
21. Heffner, E.L.; Jannink, J.-L.; Sorrells, M.E. Genomic Selection Accuracy Using Multifamily Prediction Models in a Wheat Breeding Program. *Plant Genom.* **2011**, *4*, 65–75. [CrossRef]
22. Ray, D.K.; Mueller, N.D.; West, P.C.; Foley, J.A. Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLoS ONE* **2013**, *8*, e66428. [CrossRef]
23. Rutkoski, J.E.; Crain, J.; Poland, J.; Sorrells, M.E. Genomic Selection for Small Grain Improvement. In *Genomic Selection for Crop Improvement: New Molecular Breeding Strategies for Crop Improvement*; Varshney, R.K., Roorkiwal, M., Sorrells, M.E., Eds.; Springer International Publishing AG: Cham, Switzerland, 2017; pp. 99–130.
24. Bernardo, R.; Yu, J. Prospects for Genomewide Selection for Quantitative Traits in Maize. *Crop Sci.* **2007**, *47*, 1082–1090. [CrossRef]
25. Sorrells, M.E. Genomic selection in plants: Empirical results and implications for wheat breeding. In *Advances in Wheat Genetics: From Genome to Field*; Ogihara, Y., Takumi, S., Handa, H., Eds.; Springer Japan KK: Yokohama, Japan, 2015; pp. 401–409.
26. Voss-Fels, K.P.; Cooper, M.; Hayes, B.J. Accelerating Crop Genetic Gains with Genomic Selection. *Theor. Appl. Genet.* **2019**, *132*, 669–686. [CrossRef] [PubMed]
27. Desta, Z.A.; Ortiz, R. Genomic Selection: Genome-Wide Prediction in Plant Improvement. *Trends Plant Sci.* **2014**, *19*, 592–601. [CrossRef] [PubMed]
28. Wang, X.; Xu, Y.; Hu, Z.; Xu, C. Genomic Selection Methods for Crop Improvement: Current Status and Prospects. *Crop J.* **2018**, *6*, 330–340. [CrossRef]
29. Zhang, H.; Yin, L.; Wang, M.; Yuan, X.; Liu, X. Factors Affecting the Accuracy of Genomic Selection for Agricultural Economic Traits in Maize, Cattle, and Pig Populations. *Front. Genet.* **2019**, *10*, 189. [CrossRef]
30. Heffner, E.L.; Jannink, J.-L.; Iwata, H.; Souza, E.; Sorrells, M.E. Genomic Selection Accuracy for Grain Quality Traits in Biparental Wheat Populations. *Crop Sci.* **2011**, *51*, 2597–2606. [CrossRef]
31. de los Campos, G.; Hickey, J.M.; Pong-Wong, R.; Daetwyler, H.D.; Calus, M.P.L. Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. *Genetics* **2013**, *193*, 327–345. [CrossRef] [PubMed]
32. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1–22. [CrossRef]
33. Jannink, J.L.; Lorenz, A.J.; Iwata, H. Genomic Selection in Plant Breeding: From Theory to Practice. *Brief. Funct. Genom. Proteom.* **2010**, *9*, 166–177. [CrossRef]
34. Gianola, D.; Van Kaam, J.B.C.H.M. Reproducing Kernel Hilbert Spaces Regression Methods for Genomic Assisted Prediction of Quantitative Traits. *Genetics* **2008**, *178*, 2289–2303. [CrossRef]
35. van den Berg, I.; Meuwissen, T.H.E.; MacLeod, I.M.; Goddard, M.E. Predicting the Effect of Reference Population on the Accuracy of within, across, and Multibreed Genomic Prediction. *J. Dairy Sci.* **2019**, *102*, 3155–3174. [CrossRef] [PubMed]
36. Habier, D.; Fernando, R.L.; Dekkers, J.C.M. The Impact of Genetic Relationship Information on Genome-Assisted Breeding Values. *Genetics* **2007**, *177*, 2389–2397. [CrossRef]
37. Isidro, J.; Jannink, J.L.; Akdemir, D.; Poland, J.; Heslot, N.; Sorrells, M.E. Training Set Optimization under Population Structure in Genomic Selection. *Theor. Appl. Genet.* **2015**, *128*, 145–158. [CrossRef] [PubMed]
38. Lorenzana, R.E.; Bernardo, R. Accuracy of Genotypic Value Predictions for Marker-Based Selection in Biparental Plant Populations. *Theor. Appl. Genet.* **2009**, *120*, 151–161. [CrossRef] [PubMed]

39. Combs, E.; Bernardo, R. Accuracy of Genomewide Selection for Different Traits with Constant Population Size, Heritability, and Number of Markers. *Plant Genome* **2013**, *6*, 1–7. [CrossRef]
40. Zhang, X.; Pérez-Rodríguez, P.; Semagn, K.; Beyene, Y.; Babu, R.; López-Cruz, M.A.; San Vicente, F.; Olsen, M.; Buckler, E.; Jannink, J.-L.L.; et al. Genomic Prediction in Biparental Tropical Maize Populations in Water-Stressed and Well-Watered Environments Using Low-Density and GBS SNPs. *Heredity* **2015**, *114*, 291–299. [CrossRef]
41. Brauner, P.C.; Müller, D.; Molenaar, W.S.; Melchinger, A.E. Genomic Prediction with Multiple Biparental Families. *Theor. Appl. Genet.* **2020**, *133*, 133–147. [CrossRef]
42. Edwards, S.M.K.; Buntjer, J.B.; Jackson, R.; Bentley, A.R.; Lage, J.; Byrne, E.; Burt, C.; Jack, P.; Berry, S.; Flatman, E.; et al. The Effects of Training Population Design on Genomic Prediction Accuracy in Wheat. *Theor. Appl. Genet.* **2019**, *132*, 1943–1952. [CrossRef]
43. Heslot, N.; Yang, H.P.; Sorrells, M.E.; Jannink, J.L. Genomic Selection in Plant Breeding: A Comparison of Models. *Crop Sci.* **2012**, *52*, 146–160. [CrossRef]
44. Lorenz, A.J.; Smith, K.P. Adding Genetically Distant Individuals to Training Populations Reduces Genomic Prediction Accuracy in Barley. *Crop Sci.* **2015**, *55*, 2657–2667. [CrossRef]
45. Rutkoski, J.E.; Singh, R.P.; Huerta-Espino, J.; Bhavani, S.; Poland, J.; Jannink, J.-L.; Sorrells, M.E. Efficient Use of Historical Data for Genomic Selection: A Case Study of Stem Rust Resistance in Wheat. *Plant Genome* **2015**, *8*, 1–10. [CrossRef] [PubMed]
46. Bassi, F.M.; Bentley, A.R.; Charmet, G.; Ortiz, R.; Crossa, J. Breeding Schemes for the Implementation of Genomic Selection in Wheat (*Triticum* Spp.). *Plant Sci.* **2016**, *242*, 23–36. [CrossRef]
47. Robertsen, C.; Hjortshøj, R.; Janss, L. Genomic Selection in Cereal Breeding. *Agronomy* **2019**, *9*, 95. [CrossRef]
48. Norman, A.; Taylor, J.; Edwards, J.; Kuchel, H. Optimising Genomic Selection in Wheat: Effect of Marker Density, Population Size and Population Structure on Prediction Accuracy. *G3 Genes Genom. Genet.* **2018**, *8*, 2889–2899. [CrossRef] [PubMed]
49. Meuwissen, T.H.E. Accuracy of Breeding Values of "unrelated" Individuals Predicted by Dense SNP Genotyping. *Genet. Sel. Evol.* **2009**, *41*, 35. [CrossRef] [PubMed]
50. Liu, G.; Zhao, Y.; Gowda, M.; Longin, C.F.H.; Reif, J.C.; Mette, M.F. Predicting Hybrid Performances for Quality Traits through Genomic-Assisted Approaches in Central European Wheat. *PLoS ONE* **2016**, *11*, e0158635. [CrossRef]
51. Battenfield, S.D.; Guzmán, C.; Gaynor, R.C.; Singh, R.P.; Peña, R.J.; Dreisigacker, S.; Fritz, A.K.; Poland, J.A. Genomic Selection for Processing and End-Use Quality Traits in the CIMMYT Spring Bread Wheat Breeding Program. *Plant Genom.* **2016**, *9*, 1–12. [CrossRef]
52. Michel, S.; Kummer, C.; Gallee, M.; Hellinger, J.; Ametz, C.; Akgöl, B.; Epure, D.; Löschenberger, F.; Buerstmayr, H. Improving the Baking Quality of Bread Wheat by Genomic Selection in Early Generations. *Theor. Appl. Genet.* **2018**, *131*, 477–493. [CrossRef]
53. Crossa, J.; Jarquín, D.; Franco, J.; Pérez-Rodríguez, P.; Burgueño, J.; Saint-Pierre, C.; Vikram, P.; Sansaloni, C.; Petroli, C.; Akdemir, D.; et al. Genomic Prediction of Gene Bank Wheat Landraces. *G3 Genes Genom. Genet.* **2016**, *6*, 1819–1834. [CrossRef] [PubMed]
54. Huang, M.; Cabrera, A.; Hoffstetter, A.; Griffey, C.; Van Sanford, D.; Costa, J.; McKendry, A.; Chao, S.; Sneller, C. Genomic Selection for Wheat Traits and Trait Stability. *Theor. Appl. Genet.* **2016**, *129*, 1697–1710. [CrossRef]
55. Michel, S.; Ametz, C.; Gungor, H.; Epure, D.; Grausgruber, H.; Löschenberger, F.; Buerstmayr, H. Genomic Selection across Multiple Breeding Cycles in Applied Bread Wheat Breeding. *Theor. Appl. Genet.* **2016**, *129*, 1179–1189. [CrossRef]
56. Hayes, B.J.; Panozzo, J.; Walker, C.K.; Choy, A.L.; Kant, S.; Wong, D.; Tibbits, J.; Daetwyler, H.D.; Rochfort, S.; Hayden, M.J.; et al. Accelerating Wheat Breeding for End-Use Quality with Multi-Trait Genomic Predictions Incorporating near Infrared and Nuclear Magnetic Resonance-Derived Phenotypes. *Theor. Appl. Genet.* **2017**, *130*, 2505–2519. [CrossRef] [PubMed]
57. Michel, S.; Gallee, M.; Löschenberger, F.; Buerstmayr, H.; Kummer, C. Improving the Baking Quality of Bread Wheat Using Rapid Tests and Genomics: The Prediction of Dough Rheological Parameters by Gluten Peak Indices and Genomic Selection Models. *J. Cereal Sci.* **2017**, *77*, 24–34. [CrossRef]
58. Haile, J.K.; N'Diaye, A.; Clarke, F.; Clarke, J.; Knox, R.; Rutkoski, J.; Bassi, F.M.; Pozniak, C.J. Genomic Selection for Grain Yield and Quality Traits in Durum Wheat. *Mol. Breed.* **2018**, *38*, 75. [CrossRef]
59. Kristensen, P.S.; Jahoor, A.; Andersen, J.R.; Cericola, F.; Orabi, J.; Janss, L.L.; Jensen, J. Genome-Wide Association Studies and Comparison of Models and Cross-Validation Strategies for Genomic Prediction of Quality Traits in Advanced Winter Wheat Breeding Lines. *Front. Plant Sci.* **2018**, *9*, 69. [CrossRef] [PubMed]
60. Lado, B.; Vázquez, D.; Quincke, M.; Silva, P.; Aguilar, I.; Gutiérrez, L. Resource Allocation Optimization with Multi-Trait Genomic Prediction for Bread Wheat (Triticum Aestivum L.) Baking Quality. *Theor. Appl. Genet.* **2018**, *131*, 2719–2731. [CrossRef] [PubMed]
61. Yao, J.; Zhao, D.; Chen, X.; Zhang, Y.; Wang, J. Use of Genomic Selection and Breeding Simulation in Cross Prediction for Improvement of Yield and Quality in Wheat (Triticum aestivum L.). *Crop J.* **2018**, *6*, 353–365. [CrossRef]
62. Hu, X.; Carver, B.F.; Powers, C.; Yan, L.; Zhu, L.; Chen, C. Effectiveness of Genomic Selection by Response to Selection for Winter Wheat Variety Improvement. *Plant Genom.* **2019**, *12*, 180090. [CrossRef]
63. Juliana, P.; Poland, J.; Huerta-Espino, J.; Shrestha, S.; Crossa, J.; Crespo-Herrera, L.; Toledo, F.H.; Govindan, V.; Mondal, S.; Kumar, U.; et al. Improving Grain Yield, Stress Resilience and Quality of Bread Wheat Using Large-Scale Genomics. *Nat. Genet.* **2019**, *51*, 1530–1539. [CrossRef]
64. Kristensen, P.S.; Jahoor, A.; Andersen, J.R.; Orabi, J.; Janss, L.; Jensen, J. Multi-Trait and Trait-Assisted Genomic Prediction of Winter Wheat Quality Traits Using Advanced Lines from Four Breeding Cycles. *Crop. Breed. Genet. Genom.* **2019**, *1*, e1900010. [CrossRef]

65. Kristensen, P.S.; Jensen, J.; Andersen, J.R.; Guzmán, C.; Orabi, J.; Jahoor, A. Genomic Prediction and Genome-Wide Association Studies of Flour Yield and Alveograph Quality Traits Using Advanced Winter Wheat Breeding Material. *Genes* **2019**, *10*, 669. [CrossRef] [PubMed]
66. Michel, S.; Löschenberger, F.; Ametz, C.; Pachler, B.; Sparry, E.; Bürstmayr, H. Combining Grain Yield, Protein Content and Protein Quality by Multi-Trait Genomic Selection in Bread Wheat. *Theor. Appl. Genet.* **2019**, *132*, 2767–2780. [CrossRef]
67. Ibba, M.I.; Crossa, J.; Montesinos-López, O.A.; Montesinos-López, A.; Juliana, P.; Guzman, C.; Delorean, E.; Dreisigacker, S.; Poland, J. Genome-Based Prediction of Multiple Wheat Quality Traits in Multiple Years. *Plant Genom.* **2020**, *13*, e20034. [CrossRef]
68. Tsai, H.Y.; Janss, L.L.; Andersen, J.R.; Orabi, J.; Jensen, J.D.; Jahoor, A.; Jensen, J. Genomic Prediction and GWAS of Yield, Quality and Disease-Related Traits in Spring Barley and Winter Wheat. *Sci. Rep.* **2020**, *10*, 3347. [CrossRef] [PubMed]
69. Charmet, G.; Storlie, E.; Oury, F.X.; Laurent, V.; Beghin, D.; Chevarin, L.; Lapierre, A.; Perretant, M.R.; Rolland, B.; Heumez, E.; et al. Genome-Wide Prediction of Three Important Traits in Bread Wheat. *Mol. Breed.* **2014**, *34*, 1843–1852. [CrossRef] [PubMed]
70. Piepho, H.P.; Möhring, J. Computing Heritability and Selection Response from Unbalanced Plant Breeding Trials. *Genetics* **2007**, *177*, 1881–1888. [CrossRef] [PubMed]
71. Ward, J.; Rakszegi, M.; Bedo, Z.; Shewry, P.R.; Mackay, I. Differentially Penalized Regression to Predict Agronomic Traits from Metabolites and Markers in Wheat. *BMC Genet.* **2015**, *16*, 19. [CrossRef]

Prilog 3. Znanstveni rad: Plavšin, I., Gunjača, J., Galić, V., Novoselović, D. (2022). Evaluation of Genomic Selection Methods for Wheat Quality Traits in Biparental Populations Indicates Inclination towards Parsimonious Solutions. Agronomy 12 (5): 1126. doi:10.3390/agronomy12051126

agronomy

MDPI

Article

Evaluation of Genomic Selection Methods for Wheat Quality Traits in Biparental Populations Indicates Inclination towards Parsimonious Solutions

Ivana Plavšin [1,2], Jerko Gunjača [2,3,*], Vlatko Galić [4] and Dario Novoselović [1,2]

[1] Department for Cereal Breeding and Genetics, Agricultural Institute Osijek, Južno predgrađe 17, 31000 Osijek, Croatia; ivana.plavsin@poljinos.hr (I.P.); dario.novoselovic@poljinos.hr (D.N.)
[2] Centre of Excellence for Biodiversity and Molecular Plant Breeding (CoE CroP-BioDiv), Svetošimunska cesta 25, 10000 Zagreb, Croatia
[3] Department of Plant Breeding, Genetics and Biometrics, Faculty of Agriculture, University of Zagreb, Svetošimunska cesta 25, 10000 Zagreb, Croatia
[4] Department for Maize Breeding and Genetics, Agricultural Institute Osijek, Južno predgrađe 17, 31000 Osijek, Croatia; vlatko.galic@poljinos.hr
* Correspondence: jgunjaca@agr.hr; Tel.: +385-1-239-3938

Abstract: Breeding for end-use quality traits is often challenging since their assessment requires larger quantities of grain and flour samples, which are usually not available early in the breeding process. Using the mixograph as a fast and effective method of evaluating dough quality together with genomic selection (GS) can help in pre-selecting high-performing progenies earlier in the breeding process and achieve a higher gain per unit of time and cost. In the present study, the potential of GS to predict seven end-use quality traits, including mixograph traits, in two biparental wheat populations was investigated. Field trials with both populations were conducted at two locations in Croatia (Osijek and Slavonski Brod) over three years. Results showed that the size of the training population (TP) plays an important role in achieving higher prediction accuracies, while marker density is not a major limitation. Additionally, results of the present study did not support the optimization of TP based on phenotypic variance as a tool to increase prediction accuracy. The performance of eight prediction models was compared and among them elastic net showed the lowest prediction accuracy for all traits. Bayesian models provided slightly higher prediction accuracy than the ridge regression best linear unbiased prediction (RR-BLUP) model, which is negligible considering the time required to perform an analysis. Although RR-BLUP was not the best performing model in all cases, no advantage of using any other model studied here was observed. Furthermore, strong differences between environments in terms of the prediction accuracy achieved were observed, suggesting that environments that are less predictive should be removed from the dataset used to train the prediction model. The prediction accuracies obtained in this study support implementation of GS in wheat breeding for end-use quality, including some mixograph traits.

Keywords: wheat; quality traits; genomic selection; biparental population; RIL; prediction models; training population; phenotypic variance

Citation: Plavšin, I.; Gunjača, J.; Galić, V.; Novoselović, D. Evaluation of Genomic Selection Methods for Wheat Quality Traits in Biparental Populations Indicates Inclination towards Parsimonious Solutions. *Agronomy* **2022**, *12*, 1126. https://doi.org/10.3390/agronomy12051126

Received: 25 February 2022
Accepted: 5 May 2022
Published: 7 May 2022

1. Introduction

The importance of wheat (*Triticum aestivum* L.) is underlined by the fact that wheat products are the most important source of dietary proteins and energy supply for humankind [1,2]. Therefore, suitable wheat quality is of great importance. Many traits have been identified to determine wheat quality, namely grain protein content (GPC), wet gluten content (WGC), gluten quality, grain hardness, test weight (TW), etc. [3]. High wheat quality is determined by high GPC content, while TW is often used as an indicator of flour yield. On the other hand, the baking quality of wheat is mainly influenced by gluten content and, more importantly, its composition and quality [4]. However, breeding for end-use quality

traits, and especially baking quality traits, is often challenging because larger quantities of grain and flour samples are needed, which are usually not available early in the breeding process. The mixogram allows for the precise evaluation of flour quality using a relatively small sample (2–35 g), making it ideal for plant breeding, especially in the early breeding generations. It is a dough mixer that rapidly develops a dough sample and establishes its rheological profile, which provides the information on gluten quality and dough strength, optimum development time, etc. [5].

Reduced costs and the introduction of novel genotyping technologies have enabled high-density genotyping and increased the use of molecular markers in plant breeding. Because phenotyping is often very time consuming, breeders are increasingly turning to alternative breeding approaches to reduce the need for phenotyping and speed up the selection process. Phenotyping for end-use and baking quality of wheat is not only time consuming but also costly. For an equal number of lines, phenotyping for end-use quality and processing traits can be up to fifty times more expensive than high-density genotyping-by-sequencing (GBS), which is widely available today [6]. Therefore, approaches based on molecular markers are increasingly used in plant breeding, including breeding for wheat quality. One of the widely used methods of marker-based selection is genomic selection (GS), which was first proposed as a promising breeding strategy in 2001 when Meuwissen et al. [7] revealed that high-density markers can be used to estimate the breeding values of non-phenotyped genotypes. In GS, the training population (TP) is used to evaluate marker effects, followed by model validation in a related (preselected) validation population (VP), while selection is performed in a test or a target breeding population that contains candidate genotypes for which phenotypic data are not available. Marker effects estimated by predictive statistical models using genotypic and phenotypic data of the TP are used to calculate the genomic estimated breeding value (GEBV) of the selection candidates. GS helps to reduce the duration of a breeding cycle, improves selection accuracy, and allows more effective use of genetic diversity to increase genetic gain in breeding programs [8,9]. It allows for the selection of lines earlier in the breeding cycle, thereby reducing the potential cost of phenotyping [10]. Since GS takes into account all available markers without pre-selecting them, it has been reported to be particularly suitable for predicting polygenic traits, the expression of which is influenced by a large number of low-effect loci, such as the end-use quality traits of wheat [11].

The first step towards the successful implementation of GS in practical breeding programs is the correct adjustment of the parameters that can affect prediction accuracy. An overview of these parameters has been given elsewhere [12–15], and, as reported, the interrelatedness of the population structure, the TP size, and the marker density plays the most important role [16,17]. When designing the TP, the VP must be taken into account, i.e., the TP should be designed in accordance with the desired outcomes in the VP [18,19]. To achieve acceptable prediction accuracies, the TP should be highly related to the VP or contain genotypes that are related to the genotypes present in the VP [20,21]. Prediction accuracy increases with the size of the TP as well as with the marker density until it reaches a plateau [22–24]. The more closely related the TP and VP are, the smaller the TP and marker density required to reach the plateau of prediction accuracy [25,26]. In addition, the extent of linkage disequilibrium (LD) affects the number of markers required to reach a given level of GS prediction accuracy [9]. The extensive LD between quantitative trait loci (QTL) and markers in highly related populations, such as biparental populations, ensures that more than one marker accompanies each QTL. Consequently, a lower marker density is required to reach a plateau of prediction accuracy when GS is applied in biparental populations [27]. Optimization of the TP for GS has been shown to be important in achieving higher prediction accuracy. Many sampling algorithms have been proposed for TP optimization, namely random sampling, stratified sampling, sampling based on coefficient of determination (CD) mean or predictor error variance (PEV) mean, etc. [28,29]. Isidro et al. [28] showed that the strategy to optimize the TP depends on the population structure. According to their results, for structured populations, it is preferable to build

a TP with the largest phenotypic variance to achieve high prediction accuracy. Using a biparental population, Marulanda et al. [30] demonstrated that optimization strategies based on genetic properties of the population do not lead to an increase in prediction accuracy. It was shown that only phenotypic variance in the TP is related to prediction accuracy. Since phenotyping is currently the most expensive phase of GS, determining the optimal size of the TP and the potential need for its optimization is critical to reduce the need for phenotyping. One of the key elements for the successful and cost-effective implementation of GS in plant breeding programs is achieving the desired prediction accuracy in conjunction with effective resource allocation.

Different prediction models are developed to address the problem of high-dimensional data sets in GS, which arises from the large amount of data collected in high-throughput genotyping. Most of the differences between prediction models relate to different assumptions about the distribution and variance of marker effects, i.e., how marker effects contribute to the overall variance of the observed trait [7]. The specific features of the different prediction models have been presented in detail in previous publications [31–34]. Due to its robustness and reliability of results, ridge regression best linear unbiased prediction (RR-BLUP) is the most widely used prediction model in GS [35]. In addition to the traditionally used models, such as the genomic best linear unbiased prediction (G-BLUP), RR-BLUP, and Bayesian alphabet models and machine learning and deep learning approaches, have also been applied to GS in recent years. As in other models, the performance of machine and deep learning models has been shown to be trait-dependent. While some studies suggest that approaches based on deep learning methods outperform conventional models in predicting wheat quality traits [36–38], other studies have not shown significant improvement in performance [39].

This study investigated the potential of GS to predict seven different end-use quality traits, including mixograph traits, in two recombinant inbred line (RIL) winter wheat populations. Specific objectives included: (1) assessment of the need for TP optimization based on phenotypic variance, (2) identification of the effect of TP size and marker density on prediction accuracy using the RR-BLUP model, and (3) evaluation of the performance of the RR-BLUP model and seven other prediction models, including one machine learning model. The results obtained should provide insights and recommendations for the implementation of GS in wheat breeding for end-use and baking quality.

2. Materials and Methods

2.1. Plant Material, Field Trials, and Phenotyping

Two biparental (RIL) winter wheat populations were used in the present study: Bezostaya-1 × Klara (BK) and Monika × Golubica (MG). Pedigree of used genotypes is given in Table S4 in the Supplementary Material. After crossing of parental cultivars and selfing, plants were randomly selected up to the F7 generation, which was used for field trials in the growing season of 2008/2009. Originally, the BK population consisted of 145 genotypes and the MG population 175 genotypes, including parental cultivars. Due to the insufficient quality of samples, some genotypes could not be successfully genotyped, so a total of 139 and 153 RILs were used for this study for the BK and MG populations, respectively. Field trials with both populations were conducted at two locations in Croatia (Osijek and Slavonski Brod) over three years (2009–2011, denoting the year of harvest). Each individual year–location combination represented an environment designated by the following abbreviations: OS09 (Osijek–2009), OS10 (Osijek–2010), OS11 (Osijek–2011), SB09 (Slavonski Brod–2009), SB10 (Slavonski Brod–2010), and SB11 (Slavonski Brod–2011). In each of the six environments, a field trial was set up with two replicates according to a row–column design. Data collected for the BK population in the SB11 environment were discarded due to the low quality of the flour samples. The analysis included seven quality traits, which are listed and described in Table 1. Further details on the selection of parental cultivars, the experimental design, the soil type and weather conditions at

the experimental sites, the fertilization rate applied, and the phenotyping procedure are described in a previously published article [40].

Table 1. Quality trait abbreviations, descriptions, and measurement method used in present study.

Trait Abbreviation	Description	Unit	Measuring Instrument
GPC	Grain protein content measured on whole grain samples	percent	Infratec 121 Grain Analyzer
WGC	Wet gluten content measured using flour samples	percent	Glutomatic 2200 Gluten System/Glutomatic Centrifuge 2015 (Perten)
TW	Test weight measured on whole grain samples	kg hL^{-1}	Infratec 121 Grain Analyzer
MPT	Midline peak time measured using flour samples (denotes time required for optimal dough development) [41]	min	Mixograph (National MFG Co., National Manufacturing Company, Lincoln, NE, USA); MixSmart software (v 3.40)
MTW	Midline curve tail width measured using flour samples (designates the consistency and stability of the dough) [41]	percent	
MTI	Midline curve tail integral measured using flour samples (describes energy used during the mixing process) [41]	unitless	
MPH	Midline peak height measured using flour samples (denotes dough strength) [41]	percent	

2.2. Statistical Analysis and Heritability Estimation

The combined data from individual trials were subjected to the prediction of genotypic best linear unbiased estimates (BLUE) using the mixed model:

$$Y = G + E + G{\cdot}E + REP{\cdot}E + ROW{\cdot}REP{\cdot}E + COL{\cdot}REP{\cdot}E, \quad (1)$$

which included the fixed effects of genotype (G), environment (E), genotype-by-environment interaction (G·E), and replicates within environments (REP·E), as well as the random effects of rows and columns within replicates within environments (ROW·REP·E and COL·REP·E, respectively). The resulting predicted values for all genotype–environment combinations were used as input for all subsequent GS analyses. A more detailed description of the calculation procedure can be found in previously published paper [40].

Broad-sense heritability (H^2) for all traits was calculated as the ratio of total genetic variance to total phenotypic variance. Across-environmental heritability was assessed using the following equation:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ge}^2}{e} + \frac{\sigma_e^2}{er}}, \quad (2)$$

where σ_g^2 is genotypic variance, σ_{ge}^2 is genotype-by-environment interaction (GEI) variance, and σ_e^2 is the error variance component, while e and r represent the number of environments and the number of replicates per environment, respectively. For the assessment of heritability across environments, the variance components were calculated using model (1) treating all effects as random.

For the assessment of within environment repeatability, each environment was analyzed separately using the model:

$$Y = G + REP + ROW{\cdot}REP + COL{\cdot}REP \quad (3)$$

which includes the random effects of genotype (G), replicate (REP), and the effects of rows and columns within replicates (ROW·REP and COL·REP, respectively). The variance components thus obtained were used to calculate within-environment repeatability according to the following equation:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_e^2}{r}}, \quad (4)$$

which is a simplified version of Equation (2) that only takes σ_g^2 as genotypic and σ_e^2 as the error variance component and r as the number of replicates per environment. Statistical analysis and heritability/repeatability estimations were performed within the R environment [42] using the commercial package "asreml" [43] and the free add-on package "asremlPlus" [44].

2.3. Genotypic Characterization

For genotyping, all genotypes (F7 filial generation) were grown in a climate chamber. Leaf tissue samples were collected at the 4–5 leaf stage and used for the isolation of genomic DNA. The collected samples were frozen and subjected to a lyophilization procedure at a temperature of −50 °C and a pressure of 0.1 millibar. The samples prepared in this way were ground in a Retsch Schwing mill using Eppendorf tubes and metal beads before DNA isolation. The further procedure for DNA isolation followed the protocol described in Karp et al. [45] for the isolation of DNA from plant material. The concentration and purity of the isolated DNA was determined spectrophotometrically using an Eppendorf BioPhotometer instrument. After DNA quantification, the isolated DNA was diluted to the optimal concentration for subsequent analysis (50–100 ng/µL). DNA samples were sent to Diversity Arrays Technology (DArT) at the University of Canberra, Bruce, Australia for sequencing. For the purposes of this study, DArT SNP markers were used. The initial SNP call provided 4874 and 7192 markers for the BK and MG populations, respectively. After excluding markers with incomplete chromosomal position data, heterozygosity greater than 30% and minor allele frequency (MAF) <0.05, missing data were imputed using Beagle software, version 5.1 [46]. The final dataset used for GS contained 1087 (BK population) and 2231 (MG population) filtered and imputed SNPs. Figure S1 in the Supplementary Materials shows the distribution of SNPs on each chromosome and genetic map relative to the population.

2.4. Cross-Validation Strategies for Assessment of Different Genomic Selection Problems

The first phase of the GS analysis aimed to determine whether the TP needed to be optimized based on phenotypic variance, what size of TP was required, and to investigate the influence of marker density, all with the goal of improving prediction accuracy. Because it is robust and less computationally intensive compared to other models, only the RR-BLUP model was used to estimate the marker effects at this stage of the analysis.

The RR-BLUP is a parametric prediction model which can be represented by the following equation in matrix notation:

$$y = WGu + e \quad (5)$$

where y is a vector of phenotypic values, W and G are the design and the genotype matrix, respectively, u is a vector of marker effects which follow normal distribution and have a common variance $u \sim N(0, I\sigma_u^2)$, and the residual error is represented by $e \sim N(0, I\sigma_e^2)$. The BLUP solution for the calculation of marker effects can be written as:

$$u = (Z^T Z + \lambda I)^{-1} Z^T y \quad (6)$$

where λ is a ridge regression parameter calculated as the ratio of residual and marker variances (σ_e^2 / σ_u^2), I is an identity matrix, and $Z = WG$. The same penalty parameter is applied to all marker effects causing an equal shrinkage towards zero regardless of the

size of the marker effect. This model applies Restricted Estimated Maximum Likelihood (REML) function for marker effect estimation [35].

GS analysis using RR-BLUP was evaluated for both populations and all quality traits separately in each available environment. Prediction accuracy was estimated using 100 independent (nTimes) 10-fold cross-validations. Pearson correlation between GEBVs and actual phenotypic values was calculated for each replicate (nTimes) and final prediction accuracy was expressed as the mean value over nTimes. The mean-squared error of prediction (MSEP) value was reported as the mean over nTimes as a criterion for the quality of the model. The MSEP value was reported as the mean over nTimes. The RR-BLUP model was implemented within the R environment [42], using the "BWGS" pipeline [47] and the "glmnet" package [48].

2.4.1. Effect of Training Population Phenotypic Variance on Prediction Accuracy

For each randomly selected TP, phenotypic variance was calculated for 100 independent 10-fold cross-validations to determine if there was a correlation between the phenotypic variance of the TP and the resulting prediction accuracy. The TP sizes were set to 25, 50, and 75 lines for both BK and MG populations (representing percentages of approximately 15, 35, and 50 of the total number of lines in the population). The aim of this step of the analysis was to evaluate whether phenotypic variance should be taken into account when optimizing the TP to achieve higher prediction accuracy. Because no strong correlation was found between phenotypic variance and prediction accuracy, all subsequent analyses were conducted using only randomly selected TP.

2.4.2. Effect of Marker Density on Prediction Accuracy

The final dataset used for GS for the MG population contained twice as many SNPs (2231) compared to the BK population (1087). To determine whether the differences in prediction accuracy between populations reflected differences in marker density, an additional subset of marker data was generated for the MG population. To avoid potential bias from a completely random subset of the data, marker pruning was performed based on LD. For each pair of markers with coefficient of LD greater than 0.9, only one SNP was left in the final data set. This resulted in a subset of markers that contained 1123 SNPs, which was used to approximate the number of SNPs in the full marker dataset for the BK population.

2.4.3. Effect of Training Population Size on Prediction Accuracy

Three different TP sizes were used to investigate the effect of TP size on prediction accuracy. The population size corresponding to 50, 65, and 80% of the total number of lines was used as the TP, while the remaining lines (50, 35, and 20%) served as the VP. The percentages given correspond to the 70, 90, and 111 lines for the BK population and 77, 99 and 122 lines for the MG population. In each scenario, the TP was randomly selected for all traits. This procedure was performed for the BK population and the MG population using both the full and reduced marker datasets.

2.5. Comparison of Genomic Selection Models

The second phase of the analysis compares the performance of different GS models, for predicting all seven quality traits examined in this study with the performance of the RR-BLUP model. This part of the analysis was performed only with the MG population, as it had a larger number of markers and lines compared to the BK population. The entire marker data set available for the MG population without reduction was used to estimate marker effects. The TP included 80% of the lines (122 RILs), while the VP included the remaining 20% (31 RILs). In addition to the RR-BLUP model, which is explained in more detail in the Section 2.4, five other parametric models and two semi-parametric models were used. The parametric models included elastic net (EN) and four Bayesian models—BayesA (BA), BayesB (BB), BayesC (BC), and BayesLASSO (BL). The semi-parametric models used in this study were random forest (RF) and reproducing kernel Hilbert spaces (RKHS).

All Bayesian regression models can be described using the following equation:

$$y = \mu + \sum_{k=1}^{m} x_k \beta_k + e$$

where y is a vector of phenotypic values, μ is the overall mean, x_k is the vector of genotypes for the kth marker, β_k is the effect of the kth marker, m is the number of markers, and e is a vector of residuals with the assumptions of $e \sim N(0, I\sigma_e^2)$. Bayesian models differ in the prior assumptions of the effects of markers (β_k), i.e., they assign distinct prior distribution for the estimation of marker effects. In the BA and BB models, β_k follows the inverted chi-square distribution and the π value determines the probability that the marker has zero effect. For the BA model $\pi = 0$, which assumes that all markers have non-zero effect [7]. The BB model applies a distribution with point mass at zero, thus, allowing for many markers to have a zero effect [7,49]. From a breeders' point of view, it is a more realistic assumption given that certain regions of the genome are not associated with QTL; hence, the effects of some markers would be absent. The BC model assumes that some of the markers (1—π) have zero effect, while the rest of them (π) follow a Gaussian distribution [36]. The BL model represents the L1 regularization norm in a Bayesian framework, to obtain a form of least absolute shrinkage and selection operator (LASSO) regression described by Park and Casella [50]. This model applies double exponential distribution for the estimation of marker effects and assigns unique variance to all markers, thus, causing stronger shrinkage of regression coefficients closer to zero (markers with small effect) and weaker shrinkage of coefficient with high absolute value (markers with greater effect) [51].

The EN model applies the weighted combination of penalization represented in the RR (L2 regularization of marker effects) and LASSO (L1 regularization of marker effects) methods. It introduces the elastic-net penalty P_α, which determines how much weight is given to each of the two methods. The lower the α value, the more similar EN performs to RR ($\alpha = 0$), while EN with α value closer to 1 is more equivalent to LASSO ($\alpha = 1$). Therefore, EN can make the selection of groups of correlated markers while performing automatic variable selection and continuous shrinking simultaneously [48,52].

The RF is a machine learning model which can be represented by the following equation:

$$\hat{y}_i = \frac{1}{B} \sum_{b=1}^{B} T_b(x_i)$$

where $\hat{y}_i$ is the phenotypic prediction of the genotype x_i, T is the number of trees, and B represents the number of bootstrap samples. Briefly, RF method is based on the construction of numerous identically distributed trees. For each tree, the individual prediction using the regression model is made and the final prediction value represents an average of outputs from all trees. The bootstrap method is used to find the optimal subset of training data for the construction of each tree. The splitting at the tree node is carried out in such a way that the loss function is reduced with each bootstrapped sample [53].

The RKHS model carries out the semi-parametric regression on marker genotypes. To control the distribution of marker effects, this model uses genetic distance and a kernel function which is based on the Euclidean measure of marker similarity [54]. Briefly, the covariance structure is constructed using the markers $Cov(g_i, g_i') \propto K(x_i, x_i')$ where x_i and x_i' are vectors of marker genotypes and $K(.,.)$ is a positive definite function, e.g., reproducing kernel [55].

For each of the models used, 100 independent 10-fold cross-validations were performed. As with the RR-BLUP model, prediction accuracy and MSEP were reported as the average of the cross-validation replicates for each model. The parameters for iterative models (Bayesian models and RKHS) were set to 5000 iterations with a burn-in (number of discarded samples) of 1000 and a thinning of three. This part of the analysis was performed within the R environment [42] using the "BWGS" pipeline [47] and packages "glmnet" [48] (EN model), "BGLR" [56] (Bayesian models and RKHS), and "randomForest" [53] (RF).

3. Results

3.1. Heritability and Repeatability of the Traits

Estimates of heritability for all traits and both populations across environments and repeatability within environments are shown in Table 2. In both observed populations, heritabilities for the GPC, WGC, and TW traits, were high ranging, from 0.78 to 0.92. In general, mixograph traits had high heritabilities ($\geq$0.71), but these were somewhat lower compared with the other three quality traits. Only the MPT trait in the BK population had moderate heritability, with a value of 0.45. Comparing the within-environmental repeatabilities between the two populations, it is noticeable that they were equal or slightly higher in the BK population for most trait–environment combinations. In the BK population, repeatabilities were moderate to high, ranging from 0.55 to 0.95 for all trait-environment combinations, except in the SB09 environment where values were mostly moderate, ranging from 0.13 (MPT) to 0.78 (TW). Overall, in the BK population, the highest repeatabilities for all traits were observed in the OS09 environment. In the MG population, repeatabilities for the majority of trait–environment combinations were moderate to high, ranging from 0.52 to 0.96, with the exception of the OS11 environment, where repeatabilities were mostly low, ranging from 0.18 (GPC) to 0.43 (MPT).

Table 2. Broad-sense heritability (across environments, H^2)/repeatability (within environment) estimates for both biparental populations used.

	GPC [2]	WGC	TW	MPT	MTW	MTI	MPH
Within environment [1]	Bezostaya-1/Klara (BK) population						
OS09	0.95	0.94	0.91	0.75	0.79	0.87	0.88
OS10	0.86	0.86	0.92	0.61	0.77	0.84	0.85
OS11	0.93	0.92	0.92	0.57	0.73	0.87	0.84
SB09	0.49	0.54	0.78	0.13	0.52	0.49	0.51
SB10	0.65	0.73	0.77	0.55	0.74	0.83	0.77
Across environments	0.91	0.92	0.88	0.45	0.71	0.81	0.77
Within environment	Monika/Golubica (MG) population						
OS09	0.86	0.83	0.86	0.68	0.9	0.8	0.84
OS10	0.85	0.86	0.87	0.94	0.96	0.86	0.88
OS11	0.18	0.25	0.19	0.43	0.31	0.24	0.35
SB09	0.76	0.72	0.65	0.52	0.89	0.73	0.72
SB10	0.82	0.82	0.83	0.81	0.93	0.83	0.86
SB11	0.66	0.71	0.79	0.7	0.89	0.73	0.77
Across environments	0.90	0.89	0.78	0.84	0.91	0.72	0.76

[1] Environment abbreviations represent a combination of year and location of experiment and are as follows: OS09 (Osijek–2009), OS10 (Osijek–2010), OS11 (Osijek–2011), SB09 (Slavonski Brod–2009), SB10 (Slavonski Brod–2010), and SB11 (Slavonski Brod–2011). [2] Trait abbreviations: grain protein content (GPC), wet gluten content (WGC), test weight (TW), midline peak time (MPT), midline curve tail width (MTW), midline curve tail integral (MTI), midline peak height (MPH).

3.2. The Influence of Training Population Phenotypic Variance on Prediction Accuracy

Scatter plots showing the relationship between the phenotypic variance of randomly selected TP in three different sizes and the prediction accuracy obtained with the RR-BLUP model for two traits of each population are included in Figure 1 (GPC and TW for the BK population, MTI and MPH for the MG population), and all other plots are included in Figure S2 (BK population) or Figure S3 (MG population) in the Supplementary Materials. The number in the angle of each scatter plot represents the observed correlation coefficient. For some trait–environment combinations, the correlation was nearly zero, regardless of the TP size used (Figure 1d, Figures S2c,d and S3c–e in the Supplementary Materials). Looking at each population–trait–environment combination separately, in some cases a slight decrease in the correlation coefficient was observed along with a shift from positive to negative values as the TP size increased (Figure 1a, Figures S2a and S3a in the Supplementary Materials).

In some cases, the correlation coefficient even increased with increasing TP size (Figure 1b: OS11 environment, Figure 1d: OS09 and SB10 environment). In addition, the strength and direction of the correlation varied considerably among the different environments of the same combination of population and trait. In general, the observed correlation coefficients were low ($r \leq 0.35$) and no consistent pattern in the strength or direction of correlation was observed for any of the population–trait combinations examined.

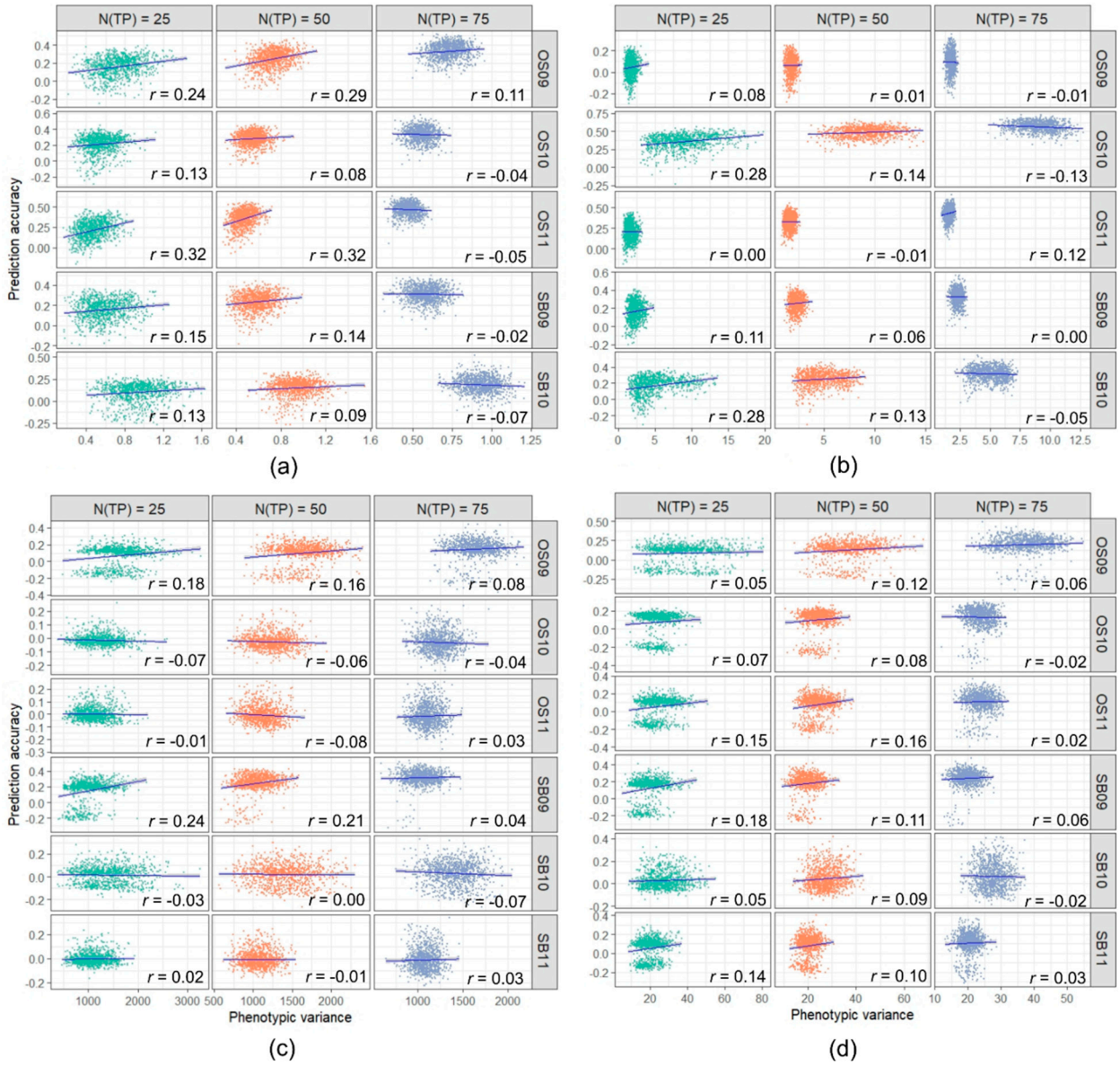


Figure 1. Phenotypic variance of randomly selected TP plotted against prediction accuracy values obtained using RR-BLUP model. The top two scatter plots refer to observations for (**a**) GPC and (**b**) TW traits of BK population, and the bottom two to (**c**) MTI and (**d**) MPH traits of MG population. For each population–trait–environment combination, three different sizes of TP were used (25, 50, and 75 lines) with remaining lines serving as VP.

3.3. The Influence of Training Population Size and Marker Density on Prediction Accuracy

The boxplots in Figure 2 and Figures S4 and S5 (in the Supplementary Materials) show the influence of the size of TP and marker density (N_M) on the prediction accuracy of the RR-BLUP model in each environment tested. MSEP values for all combinations are shown in Table S1 in the Supplementary Materials. Figure 2 includes three traits per population for

which prediction accuracy was highest in all or most environments, i.e., GPC, WGC, and TW for the BK population (Figure 2a–c) and MPT, MTW, and MPH for the MG population (Figure 2d–f in the case of N_M = 1123, and Figure 2g–i in the case of N_M = 2231). Boxplots for all other traits are included in Figure S4 (BK population) and Figure S5 (MG population) in the Supplementary Materials. When comparing two populations and cases in which approximately the same number of markers were used (N_M = 1087 and N_M = 1123 for the BK and MG populations, respectively), it is noticeable that the prediction accuracy for the traits GPC, WGC, and TW was higher in the BK population, whereas the mixograph traits showed better predictability in the MG population. An exception is the MTI trait, the predictability of which was higher in the BK population, although it was still low in most environments (<0.3). Reducing the TP size from 85% to 50% of the total number of lines in a population had a negative effect on the achieved prediction accuracy in all observed population–trait–environment combinations. Although the effect was negative, it was not as severe, implying that even using 50% of the population as TP, the prediction accuracy can still be moderate to high (Figure 2a–c). It is also noticeable that the prediction accuracy strongly depends on the environment, regardless of the TP size, and it can vary substantially, e.g., the largest difference in prediction accuracy was observed at TP size 80% for the trait TW in the MG population, ranging from 0.06 in the OS11 environment to 0.49 in the OS10 environment.

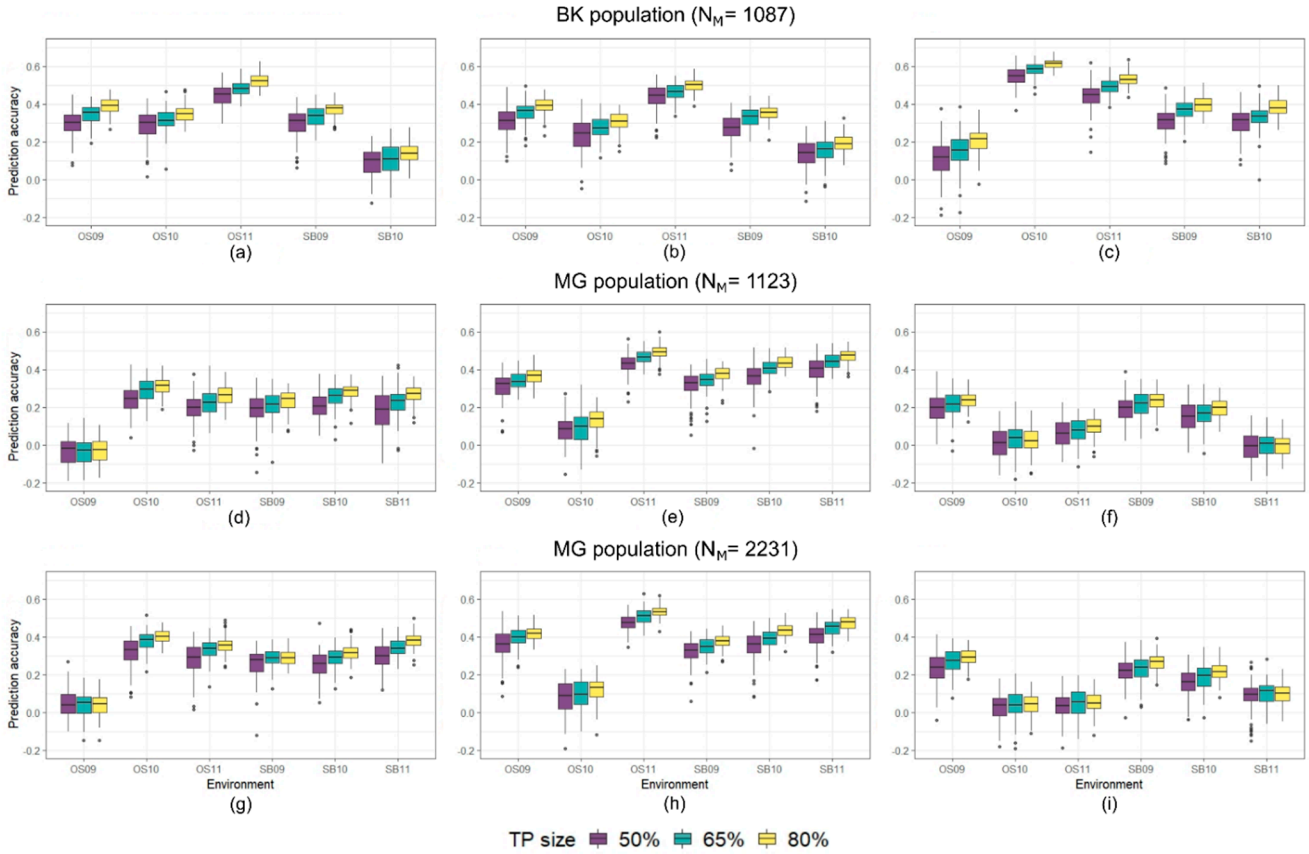


Figure 2. Prediction accuracy values obtained using RR-BLUP model and three different sizes of TP (50%, 65%, and 80% of the total number of lines in a population). The top three boxplots represent results obtained for BK population and traits (**a**) GPC, (**b**) WGC, and (**c**) TW. The middle three boxplots refer to observations for MG population and traits (**d**) MPT, (**e**) MTW, and (**f**) MPH in the case when half of the marker dataset was used (N_M = 1123 SNPs), while the bottom three (**g**–**i**) refer to the same traits, respectively, but using the whole available marker dataset (N_M = 2231 SNPs).

When comparing the influence of different marker densities on the predictability of quality traits within the MG population (Figure 2d–i and Figure S5 in the Supplementary

Materials), it is noticeable that higher values of prediction accuracy were obtained when a higher marker density was used (N_M = 2231 compared to N_M = 1123) for all trait–environment combinations and all TP sizes used. Nevertheless, the differences in found in prediction accuracy were not large. Considering the TP size of 80%, the largest difference in prediction accuracy was found for trait WGC in environment SB11, with values of 0.20 and 0.32 when N_M = 1123 and N_M = 2231 were used, respectively (Figure S5b,f in the Supplementary Materials). These results suggest that a lower number of markers is already sufficient to achieve good predictability, and that with a higher number of markers a plateau of prediction accuracy may have been reached for this population and the observed traits.

3.4. Performance of Different Prediction Models

Figure 3 shows the mean prediction accuracy for the MG population and the traits (a) GPC, (b) TW, (c) MTW, and (d) MPH obtained with eight different prediction models. Results for the remaining three traits (WGC, MPT, and MTI) are shown in Figure S6 in the Supplementary Materials. In both Figure 3 and Figure S6, the error bars indicate the standard deviation. The mean prediction accuracies over 100 independent 10-fold cross-validations are shown along with the standard deviations in Table S2 in the Supplementary Materials. The highest prediction accuracy achieved for each trait–environment combination is shown in bold. The mean MSEP values for all trait–environment–model combinations are listed in Table S3 in the Supplementary Materials. In general, trait predictability was found to be good in some environments while being low in other environments. For example, the prediction accuracy of TW (Figure 3b) was low (−0.08–0.31) in all environments except environment OS10, where the prediction accuracy was moderate (0.36–0.49) for all models examined, whereas the prediction accuracy of WGC (Figure S6a in the Supplementary Materials) was moderate for most environment–model combinations except environment SB09, where the prediction accuracy was low and even negative for the model EN. However, overall, GPC, WGC, and two mixograph traits (MPT and MTW) showed moderate predictability with prediction accuracies up to 0.57 (Figure 3a,c and Figure S6a,b and Table S2 in the Supplementary Materials). The predictability of TW and the other two mixograph traits (MTI and MPH) was rather low and varied substantially between environments, resulting in negative values of prediction accuracy in some cases (Figure 3b,d and Figure S6c and Table S2 in the Supplementary Materials). When comparing the performance of the different models, it is noticeable that the model with best performance depends strongly on the observed environment and not so much on the trait. According to Table S2, the model EN had the lowest values of prediction accuracy for the most combinations of traits and environments and was also the model with the highest number of cases with negative prediction accuracy values. In only one case did the EN model achieve the highest prediction accuracy (MPT trait in environment SB10). In 35 of 42 possible trait–environment combinations, the Bayesian alphabet models (BA, BB, and BB) proved to be superior, whereas the BL model performed best in only two cases, followed by the RF model with five and the RKHS model with seven cases. Although the RR-BLUP model performance was superior in only one case (TW trait in OS10 environment), the performance of all other models was not substantially better compared to it. Indeed, the prediction accuracy of the best performing model was on average only 0.05 points higher and ranged from 0 (in the case of the TW trait in the OS10 environment where the BA and BC models had the same prediction accuracy as RR-BLUP) to 0.14 (in the case of the MPT trait in the OS09 environment, where the superior model was BB with a prediction accuracy of 0.18 compared to 0.04 of the RR-BLUP model). Observed MSEP values (Table S3 in the Supplementary Materials) were relatively low (0.44 or lower) for the majority of the trait–environment combinations with little or no difference among prediction models used. The highest MSEP values were recorded for MTI, ranging from 0.49 to 3.42. One of the biggest differences among the implemented models is their computational efficiency, i.e., the time required for one analysis. In the present study, conducted on a 64-bit Windows 10 workstation with a 2.90 GHz Intel (R) Xeon (R)

processor and 32 GB RAM, the least demanding model was EN, which took approximately 19 min to compute a prediction accuracy. It was followed by RR-BLUP, RF, and RKHS with computation times of 38, 71, and 92 min, respectively. The most demanding were the Bayesian models, which required almost 3 h to compute an analysis (166, 162, 161, and 171 for BA, BB, BC, and BL, respectively).

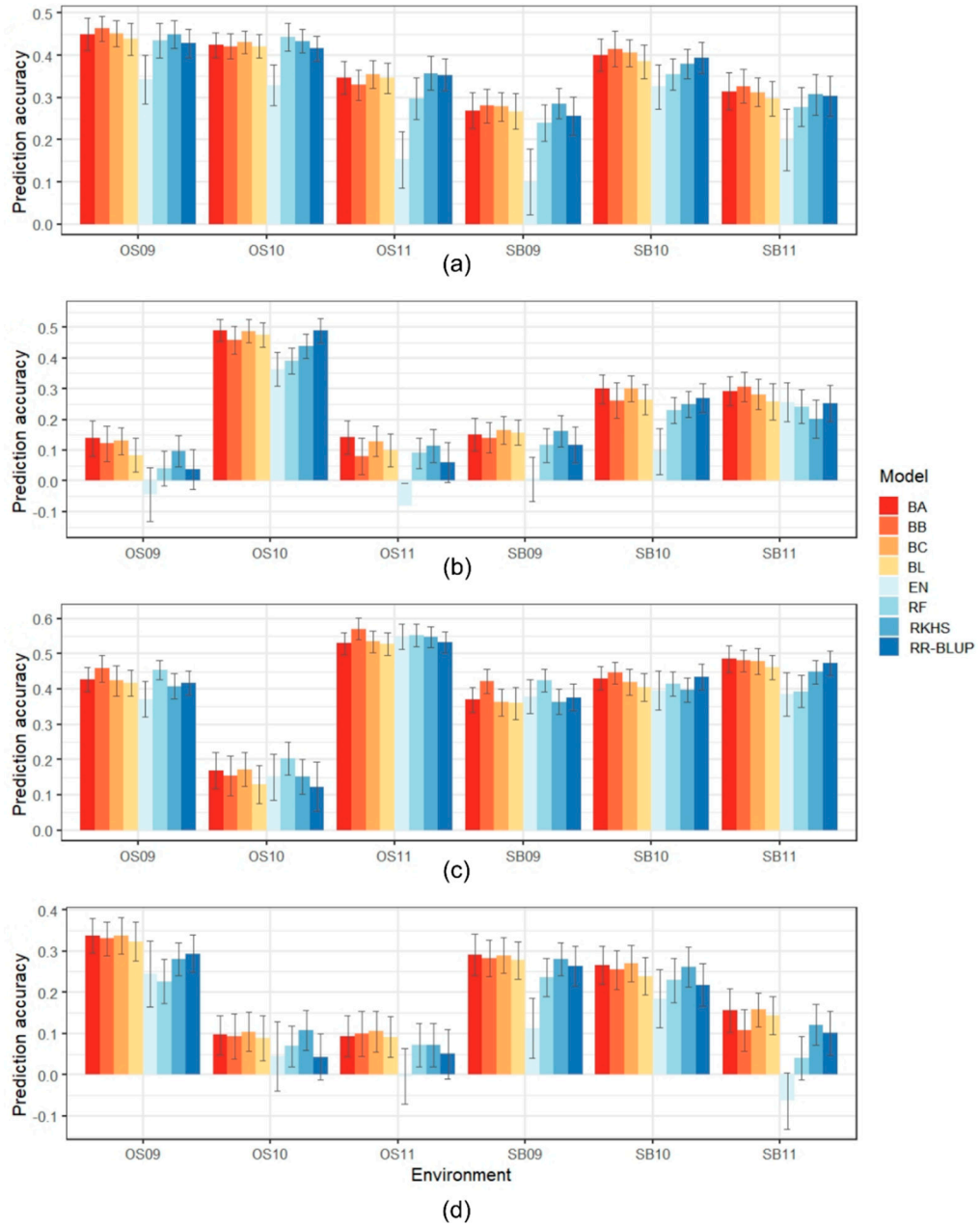


Figure 3. Prediction accuracies for MG population and traits (**a**) GPC, (**b**) TW, (**c**) MTW, and (**d**) MPH evaluated with eight different prediction models. Error bars denote standard deviation.

4. Discussion

The present study investigated the potential of the GS approach for predicting seven end-use quality traits, among which are some rheological properties of the dough obtained by the mixograph. While assessment of dough rheological traits is typically labor-intensive and time-consuming, the mixograph can provide good insight into baking quality by using only a small flour sample. When combined with GS, it has the potential to support end-use quality improvement well, especially in filial or early wheat breeding generations.

4.1. Heritability

In the present study, the broad-sense heritability estimated across all environments was high (>0.7) for all traits, with the exception of MPT in the BK population, the heritability of which was 0.45 (Table 2). Although repeatability varied considerably within environments, it was high for most of the trait–environment combinations, with a value above 0.7, suggesting that heritability itself should not be a barrier to achieving good prediction accuracy. High heritability generally indicates that genetic factors account for the majority of trait variance, making GS the ideal approach for predicting these traits, as it takes into account all available markers while attempting to capture the total genetic variance [57]. In general, the lowest repeatabilities for the BK and MG population were obtained in the SB09 and OS11 environments, respectively, indicating the presence of a stronger non-genetic variance effect within each environment. Two recent studies by Sandhu et al. [37,38] reported similar heritabilities for TW, but the reported values for GPC were lower (0.35–0.63) than those obtained in the present study. Lado et al. [58] reported low to moderate heritabilities for WGC, TW, and some mixograph traits, while Hayes et al. [59] showed that the vast majority of dough rheology and baking traits had higher heritabilities compared to grain traits such as GPC and TW. Nevertheless, the heritabilities reported to date for wheat end-use quality traits appear to be sufficient to achieve acceptable prediction accuracy and have not been reported to be a significant limiting factor for GS [60–62]. Regardless, several previous studies have shown that predictability can be low in some environments despite high heritability [63], which can be explained by various environmental factors [64]. Additionally, Sandhu et al. [38] have shown that although heritability for a single trait can vary substantially between tested environments, this variation does not significantly affect prediction accuracy.

4.2. Optimization of Training Population and Marker Density

When deciding to include GS in a breeding program, it is essential to take the right steps to optimize the factors that might lead to unnecessarily high costs. Although the cost of genotyping has decreased significantly in recent years, the cost of phenotyping wheat traits for end-use quality have remained relatively high [11]. Therefore, it is important to optimize TP to reduce the potential cost of phenotyping while maintaining the same level of prediction accuracy. Prediction accuracy was significantly affected by the size of the TP in almost all studies that investigated this issue [39,62,65–67]. In addition, the relatedness of TP and VP was found to play an important role in choosing the optimal size of TP [68]. The results of this study are consistent with those of previous studies [24,27]. In the present study, the highest prediction accuracy for all traits was achieved when 80% of the dataset was used as TP, i.e., when the TP included 111 and 122 RILs for the BK and MG populations, respectively (Figure 2, Figures S4 and S5 in the Supplementary Materials). On the other hand, the results obtained with half of the dataset as TP (70 and 77 RILs for BK and MG, respectively) show that acceptable levels of prediction accuracy values may be achieved even with a smaller TP. Previous research has also shown that a reasonably small TP size is sufficient to achieve high prediction accuracies [58,62,69,70], especially in highly related populations such as biparental populations [24]. This is particularly important for resource allocation, i.e., deciding on the number of genotypes to include in the experiment, especially when phenotyping is time-consuming or expensive. Additionally, some other criteria for selecting TP individuals, such as PEV mean or CD mean, were recommended in previous research in order to maximize prediction accuracy. Using two diverse groups of maize inbreeds, Rincent et al. [29] showed that TP optimization based on CD mean values maximizes the reliability of GS. The authors justified the reported results by stating that the CD mean reduces the variance due to the higher relatedness of individuals in the selected TP. Marulanda et al. [30] studied the influence of different parameters on the variability of prediction accuracy using a simulated biparental maize population. Of the parameters studied, only TP phenotypic variance was found to be positively correlated with the prediction accuracy and was suggested as a tool for the optimization of TP. This

correlation was stronger when smaller TP was used, while it was weaker and even negative in the case of larger TP. In the present study, no consistent correlation was found between phenotypic variance and prediction accuracy, regardless of the trait or size of TP used (Figure 1, Figures S2 and S3 in the Supplementary Materials). Therefore, results of the present study do not support the optimization of TP based on phenotypic variance as a tool to increase prediction accuracy for wheat end-use quality traits.

Although the cost of genotyping has decreased substantially over the years, it remains a significant source of expense for breeders. Therefore, it is important to optimize the marker density used for GS. According to previous studies, increasing marker density has a positive effect on prediction accuracy but reaches a plateau after which further increase have no significant effect on prediction accuracy [24,71]. Whether the plateau is reached with a lower or higher marker density depends on the relatedness of the population. Liu et al. [66] have shown that the plateau is reached at approximately 3000 markers and can be as low as 500 markers when TP and VP are more closely related. Due to the low rate of recombination, closely related plant populations usually have large linkage blocks, resulting in a high LD between markers and QTL. Consequently, a lower marker density is required to reach the plateau of prediction accuracy in biparental populations, which have a high LD compared to populations with low relatedness [27]. Using a double-haploid (DH) population and a breeding panel, Haile et al. [71] showed that the plateau for GPC is reached at 2000 markers. Juliana et al. [72] reported that once genomic resolution is achieved, increasing marker density has little effect on the predictability of quality traits in biparental wheat populations. It has been reported that, for wheat quality traits, this genomic resolution can be achieved even at low marker density, i.e., 256 markers in biparental populations and 768 markers in multi-family populations [24,73]. In the present study using the MG population, we compared the prediction accuracies when the entire available marker dataset (N_M = 2231) and half of it (N_M = 1123) were used (Figure 2 and Figure S5 in the Supplementary Materials). According to the existing literature, 1123 markers should be sufficient to achieve acceptable prediction accuracy in the biparental population. In general, no large increase in prediction accuracy was achieved when 2231 markers were used compared with 1123 markers (Figure 2, Figures S4 and S5 in the Supplementary Materials). However, the results presented in this study show that for some traits, such as GPC (Figure S5a,e in the Supplementary Materials), WGC (Figure S5b,f in the Supplementary Materials), and MPT (Figure 2d,g), the increase was slightly larger compared with the other four studied traits. This suggests that for some traits, such as TW, MTW, MTI, and MPH, the plateau was already reached at 1123 markers, whereas for other traits, a further increase in marker density may still improve prediction accuracy. Gorjanc et al. [74] reported that low coverage GBS combined with increased TP size doubles the value of prediction accuracy and can be successfully used for GS in biparental populations. These results suggest that the size of TP plays a more important role in achieving high prediction accuracy than marker density [27].

4.3. Prediction Accuracies of Different Models

Sufficiently high prediction accuracies to allow for the inclusion of GS in the breeding program and selection early in the breeding cycle have already been reported for end-use quality traits [59,66]. Comparing the prediction accuracy of RR-BLUP model for two populations examined in this study, it can be seen that some traits are more predictable in one population than another (Figure 2, Figures S4 and S5 in the Supplementary Materials). With the exception of some environments, higher prediction accuracy was obtained for GPC, WGC, and TW in the BK population, while the mixograph traits MPT, MTW, and MPH showed better predictability in the MG population. MTI showed low predictability in both populations together with high MSEP values, from which it can be concluded that this trait is not a good target trait for GS. Trait predictabilities observed in this study varied by environment but were generally comparable to results from the existing literature [24,75]. When lines were randomly assigned to TP or VP, Kristensen et al. [11] achieved a prediction

accuracy of 0.5 or higher for wheat quality traits, including GPC. Using different prediction models for biparental wheat populations, Charmet et al. [76] reported accuracies up to 0.7 for TW, which is higher than that presented by the results of the present study, where the highest prediction accuracy of TW was approximately 0.6. Lado et al. [58] showed prediction accuracies ranging from 0.24 to 0.43 for eight bread baking quality traits, including WGC and mixograph traits. On the other hand, Battenfield et al. [39] obtained moderate prediction accuracies (up to 0.62) for several mixograph traits while showing low predictability of TW. Nevertheless, some authors reported that lower prediction accuracies can be successfully used to exploit GS in early generations when the selection of lines is performed simultaneously based on GEBV and BLUP values [10].

Although numerous prediction models have been developed to date for GS, none has shown a clear advantage over other models by achieving higher prediction accuracy regardless of the trait being evaluated [51,60]. Previous research has shown that there is no significant difference in performance between BLUP and Bayesian models for most wheat end-use quality traits [11,57]. When comparing the performance of RR-BLUP and BC models for wheat quality traits in two biparental populations, Heffner et al. [24] found little or no difference in average performance of the two models. However, when looking at each population separately, they concluded that RR-BLUP performed better than BC in one population, while it was less accurate in the other population, which the authors explained by the different marker effects in each population. Some studies have shown that Bayesian models are better at capturing LD between markers and QTL and are, therefore, better for predicting genotype performance when TP and VP have low relatedness [11,77,78]. Of the seven prediction models used, Battenfield et al. [39] reported the lowest prediction accuracies for RF in general. Sandhu et al. [38] found that deep learning models outperformed the RR-BLUP model in a biparental population. The RR-BLUP model provided a prediction accuracy of 0.48 and 0.45 for the GPC and TW, respectively, while one of the deep learning models used provided approximately 10% higher prediction accuracy. In another study by Sandhu et al. [37], the authors confirmed that deep learning models are generally superior to other models used for predicting end-use quality and processing traits in wheat breeding populations. Average prediction accuracy across all traits was highest for deep learning models (0.63–0.64), followed by machine learning models (0.63 for both RF and support vector machine, SVM) and RR-BLUP (0.61). The lowest average prediction accuracies were obtained for Bayesian models. In addition to the deep learning models, SVM and RF were the best performing models for the traits GPC and TW, followed by RR-BLUP and Bayesian models, although the differences in prediction accuracy were minor. In the present study, we compared the performance of eight models, including one machine learning model (RF) (Figure 3, Figure S6 and Table S2 in the Supplementary Materials). In general, the model with the lowest prediction accuracy was EN. However, in some cases EN was as successful as RR-BLUP and it was also the least computationally intensive model, so it is recommended for cases where a breeding program includes a large number of lines and selection needs to be carried out quickly and not with high precision. The RF and RKHS models outperformed RR-BLUP only for some trait–environment combinations and, therefore, cannot be recommended as models of choice for end-use traits in general, as in some previous studies [37]. For the majority of trait–environment combinations, Bayesian models (BA, BB, and BC) had the highest prediction accuracy. Nevertheless, the obtained values were not substantially higher than those of RR-BLUP, which could be explained by the high relatedness of TP and VP in the present study [11,77]. Bayesian models were also the most computationally intensive and time-consuming models in the present study, requiring more than 2 h for one analysis. Therefore, Bayesian models could be recommended for breeding programs with fewer lines where selection must be performed with a higher degree of precision. Since no clear superiority of one model over another in terms of achieved prediction accuracy could be shown in the present study, less computationally intensive models that also achieve a reasonable level of prediction accuracy, such as RR-BLUP, represent the best choice.

4.4. Genotype-by-Environment Interaction

Another major challenge in implementing GS in breeding programs represents GEI [79]. According to Bernardo [80], there are three possibilities of how to deal with the GEI when breeding for quantitative traits in plants. The first approach is to ignore it, the second is to reduce it, and the third is to exploit it. Due to the high heritability of the traits investigated in the present study, the first approach was applied, and the analysis was performed for each of the environments separately. Indeed, it has already been reported that the prediction accuracy varies considerably between the environments tested [81,82], and the results presented in this study are no exception. Looking at the prediction accuracy within the MG population (Figure 3 and Figure S6 in the Supplementary Materials), it is clear that the prediction accuracy for some traits, such as TW (Figure 3b), is moderate in one environment (OS10), while it is low in all other environments. A similar pattern can be observed for MTW (Figure 3c), which has low predictability in the OS10 environment, while the prediction accuracy is moderate to high in the other environments, and for WGC (Figure S6a in the Supplementary Materials) and MPT (Figure S6b in the Supplementary Materials), for which substantially lower prediction accuracy was observed in the SB09 and OS09 environments, respectively. Comparing these results with those from our previous publication in which we examined GEI in the same dataset [40], certain assumptions can be made. Environments that are characterized by unusually high or low values for prediction accuracy compared to the rest of the environments tend to be those that produce the greatest GEI and are more pronounced. The clearest example of this is the TW, the predictability of which was highest in the OS10 environment. This was the only outstanding environment, while all others were grouped together on the AMMI2 biplot (see Plavšin et al. [40] Figure 3b). Nevertheless, the stability of the prediction models and the accuracies achieved in different environments are still largely unknown. Some research suggests that modeling GEI in GS [83–85] or incorporating information from correlated environments [86,87] leads to higher prediction accuracy. Ornella et al. [87] showed that high correlation between environments allows for the prediction of one environment based on a model trained with data from another environment. Furthermore, identifying and removing environments from the dataset used to train the prediction model proved to be a successful strategy to improve prediction accuracy [64,65]. This would be a good strategy for the WGC, MPT, and MTW traits from this study, as only one environment was found to be less predictive, while moderate predictive abilities were seen in all the others.

5. Conclusions

In the present study, the potential of GS to predict seven end-use quality traits in two biparental wheat populations was investigated. As in previous studies, it was found that the size of TP plays an important role in achieving high prediction accuracies, while marker density is not a major limitation nowadays due to the use of high-throughput genotyping. Moreover, no advantage of TP optimization based on phenotypic variance was found in this study. Although RR-BLUP was not the best performing model in all cases presented, no significant advantage of using any other model studied here was observed. Some Bayesian models provided slightly higher prediction accuracy than RR-BLUP, which can be considered negligible considering the time required to perform an analysis. Furthermore, we observed strong differences between environments in terms of the prediction accuracy achieved, suggesting that environments that are less predictive should be removed from the dataset used to train the prediction model. Nonetheless, we provided evidence that GS is a good potential selection tool for end-use quality traits, including some mixograph traits. End-use quality traits, and especially dough rheology traits, are typically difficult to breed for because their evaluation is time-consuming and requires a larger quantity of seed, which is usually not available in early generations. Therefore, using the mixograph as a fast and effective method of evaluating dough quality together with GS can help in pre-selecting high-performing lines earlier in the breeding process and achieve a higher gain per unit of time and cost.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agronomy12051126/s1, Figure S1: The distribution of SNPs on each chromosome for (a) BK population and (c) MG population together with genetic map of all available SNPs for (c) BK population and (d) MG population, Figure S2: Phenotypic variance of randomly selected TP plotted against prediction accuracy values obtained using RR-BLUP model for following traits of BK population: (a) WGC, (b) MPT, (c) MTW, (d) MTI, and (e) MPH, Figure S3: Phenotypic variance of randomly selected TP plotted against prediction accuracy values obtained using RR-BLUP model for following traits of MG population: (a) GPC, (b) WGC, (c) TW, (d) MPT, and (e) MTW, Figure S4: Prediction accuracy values obtained using RR-BLUP model and three different sizes of TP (50%, 65% and 80% of the total number of lines in a population) for following traits of BK population: (a) MPT, (b) MTW, (c) MTI, and (d) MPH, Figure S5: Prediction accuracy values obtained using RR-BLUP model and three different sizes of TP (50%, 65% and 80% of the total number of lines in a population) for following traits of MG population: (a) GPC, (b) WGC, (c) TW, and (d) MTI, Figure S6: Prediction accuracies for MG population and traits (a) WGC, (b) MPT, and (c) MTI evaluated with eight different prediction models. Error bars denote standard deviation; Table S1: Mean MSEP values estimated for both populations using RR-BLUP model. Standard deviation values are indicated in parenthesis, Table S2: Mean prediction accuracy values estimated for MG population using eight different prediction models. Standard deviation values are indicated in parenthesis, Table S3: Mean MSEP values estimated for MG population using eight different prediction models. Standard deviation values are indicated in parenthesis, Table S4: Pedigree of winter wheat genotypes used in the study.

Author Contributions: Conceptualization, I.P., J.G. and D.N.; methodology, I.P., V.G. and J.G.; formal analysis, I.P. and J.G.; writing—original draft preparation, I.P.; writing—review and editing, J.G., V.G. and D.N.; visualization, I.P.; supervision, J.G. and D.N. All authors have read and agreed to the published version of the manuscript.

Funding: This study has been fully supported by the project KK.01.1.1.01.0005 Biodiversity and Molecular Plant Breeding, Centre of Excellence for Biodiversity and Molecular Plant Breeding (CoE CroP-BioDiv), Zagreb, Croatia.

Data Availability Statement: The data were obtained from the Agricultural Institute Osijek and are available upon request from the corresponding author with the permission of the Agricultural Institute Osijek.

Acknowledgments: The authors gratefully acknowledge Ruđer Šimek for his valuable technical assistance.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shewry, P.R. Improving the protein content and composition of cereal grain. *J. Cereal Sci.* **2007**, *46*, 239–250. [CrossRef]
2. Shewry, P.R.; Hey, S.J. The contribution of wheat to human diet and health. *Food Energy Secur.* **2015**, *4*, 178–202. [CrossRef] [PubMed]
3. Bordes, J.; Ravel, C.; Le Gouis, J.; Lapierre, A.; Charmet, G.; Balfourier, F. Use of a global wheat core collection for association analysis of flour and dough quality traits. *J. Cereal Sci.* **2011**, *54*, 137–147. [CrossRef]
4. Shewry, P.R.; Tatham, A.S.; Barro, F.; Barcelo, P.; Lazzeri, P. Biotechnology of breadmaking: Unraveling and manipulating the multi-protein gluten complex. *Bio/Technology* **1995**, *13*, 1185–1190. [CrossRef] [PubMed]
5. Swanson, C.O.; Working, E.B. Testing of the quality of flour by the recording dough mixer. *Cereal Chem.* **1933**, *10*, 1–29.
6. Guzman, C.; Peña, R.J.; Singh, R.; Autrique, E.; Dreisigacker, S.; Crossa, J.; Rutkoski, J.; Poland, J.; Battenfield, S. Wheat quality improvement at CIMMYT and the use of genomic selection on it. *Appl. Transl. Genom.* **2016**, *11*, 3–8. [CrossRef]
7. Meuwissen, T.H.E.; Hayes, B.J.; Goddard, M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **2001**, *157*, 1819–1829. [CrossRef]
8. Lorenz, A.J.; Chao, S.; Asoro, F.G.; Heffner, E.L.; Hayashi, T.; Iwata, H.; Smith, K.P.; Sorrells, M.E.; Jannink, J.L. Genomic selection in plant breeding: Knowledge and prospects. *Adv. Agron.* **2011**, *110*, 77–123. [CrossRef]
9. Sorrells, M.E. Genomic selection in plants: Empirical results and implications for wheat breeding. In *Advances in Wheat Genetics: From Genome to Field*; Ogihara, Y., Takumi, S., Handa, H., Eds.; Springer Japan KK: Yokohama, Japan, 2015; pp. 401–409.
10. Belamkar, V.; Guttieri, M.J.; Hussain, W.; Jarquín, D.; El-basyoni, I.; Poland, J.; Lorenz, A.J.; Baenziger, P.S. Genomic selection in preliminary yield trials in a winter wheat breeding program. *G3 Genes Genomes Genet.* **2018**, *8*, 2735–2747. [CrossRef]

11. Kristensen, P.S.; Jahoor, A.; Andersen, J.R.; Cericola, F.; Orabi, J.; Janss, L.L.; Jensen, J. Genome-wide association studies and comparison of models and cross-validation strategies for genomic prediction of quality traits in advanced winter wheat breeding lines. *Front. Plant Sci.* **2018**, *9*, 69. [CrossRef]
12. Plavšin, I.; Gunjača, J.; Šatović, Z.; Šarčević, H.; Ivić, M.; Dvojković, K.; Novoselović, D. An overview of key factors affecting genomic selection for wheat quality traits. *Plants* **2021**, *10*, 745. [CrossRef]
13. Wang, X.; Xu, Y.; Hu, Z.; Xu, C. Genomic selection methods for crop improvement: Current status and prospects. *Crop J.* **2018**, *6*, 330–340. [CrossRef]
14. Combs, E.; Bernardo, R. Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers. *Plant Genome* **2013**, *6*, plantgenome2012-11. [CrossRef]
15. Krishnappa, G.; Savadi, S.; Tyagi, B.S.; Singh, S.K.; Mamrutha, H.M.; Kumar, S.; Mishra, C.N.; Khan, H.; Gangadhara, K.; Uday, G.; et al. Integrated genomic selection for rapid improvement of crops. *Genomics* **2021**, *113*, 1070–1086. [CrossRef] [PubMed]
16. Robertsen, C.; Hjortshøj, R.; Janss, L. Genomic Selection in Cereal Breeding. *Agronomy* **2019**, *9*, 95. [CrossRef]
17. Riedelsheimer, C.; Endelman, J.B.; Stange, M.; Sorrells, M.E.; Jannink, J.L.; Melchinger, A.E. Genomic predictability of interconnected biparental maize populations. *Genetics* **2013**, *194*, 493–503. [CrossRef]
18. Jannink, J.L.; Lorenz, A.J.; Iwata, H. Genomic selection in plant breeding: From theory to practice. *Brief. Funct. Genom. Proteom.* **2010**, *9*, 166–177. [CrossRef]
19. Crossa, J.; Jarquín, D.; Franco, J.; Pérez-Rodríguez, P.; Burgueño, J.; Saint-Pierre, C.; Vikram, P.; Sansaloni, C.; Petroli, C.; Akdemir, D.; et al. Genomic prediction of gene bank wheat landraces. *G3 Genes Genomes Genet.* **2016**, *6*, 1819–1834. [CrossRef]
20. Asoro, F.G.; Newell, M.A.; Beavis, W.D.; Scott, M.P.; Jannink, J.-L. Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. *Plant Genome J.* **2011**, *4*, 132–144. [CrossRef]
21. Hickey, J.M.; Dreisigacker, S.; Crossa, J.; Hearne, S.; Babu, R.; Prasanna, B.M.; Grondona, M.; Zambelli, A.; Windhausen, V.S.; Mathews, K.; et al. Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop Sci.* **2014**, *54*, 1476–1488. [CrossRef]
22. Maulana, F.; Kim, K.S.; Anderson, J.D.; Sorrells, M.E.; Butler, T.J.; Liu, S.; Baenziger, P.S.; Byrne, P.F.; Ma, X.-F. Genomic selection of forage quality traits in winter wheat. *Crop Sci.* **2019**, *59*, 2473–2483. [CrossRef]
23. Arruda, M.P.; Brown, P.J.; Lipka, A.E.; Krill, A.M.; Thurber, C.; Kolb, F.L. Genomic selection for predicting *Fusarium* head blight resistance in a wheat breeding program. *Plant Genome* **2015**, *8*, plantgenome2015-01. [CrossRef] [PubMed]
24. Heffner, E.L.; Jannink, J.-L.; Iwata, H.; Souza, E.; Sorrells, M.E. Genomic selection accuracy for grain quality traits in biparental wheat populations. *Crop Sci.* **2011**, *51*, 2597–2606. [CrossRef]
25. Rutkoski, J.; Singh, R.P.; Huerta-Espino, J.; Bhavani, S.; Poland, J.; Jannink, J.-L.; Sorrells, M.E. Efficient use of historical data for genomic selection: A case study of stem rust resistance in wheat. *Plant Genome* **2015**, *8*, eplantgenome2014-09. [CrossRef] [PubMed]
26. Herter, C.P.; Ebmeyer, E.; Kollers, S.; Korzun, V.; Würschum, T.; Miedaner, T. Accuracy of within- and among-family genomic prediction for *Fusarium* head blight and *Septoria tritici* blotch in winter wheat. *Theor. Appl. Genet.* **2019**, *132*, 1121–1135. [CrossRef]
27. Lorenzana, R.E.; Bernardo, R. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theor. Appl. Genet.* **2009**, *120*, 151–161. [CrossRef]
28. Isidro, J.; Jannink, J.L.; Akdemir, D.; Poland, J.; Heslot, N.; Sorrells, M.E. Training set optimization under population structure in genomic selection. *Theor. Appl. Genet.* **2015**, *128*, 145–158. [CrossRef]
29. Rincent, R.; Laloë, D.; Nicolas, S.; Altmann, T.; Brunel, D.; Revilla, P.; Rodríguez, V.M.; Moreno-Gonzalez, J.; Melchinger, A.; Bauer, E.; et al. Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: Comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.). *Genetics* **2012**, *192*, 715–728. [CrossRef]
30. Marulanda, J.J.; Melchinger, A.E.; Würschum, T. Genomic selection in biparental populations: Assessment of parameters for optimum estimation set design. *Plant Breed.* **2015**, *134*, 623–630. [CrossRef]
31. de los Campos, G.; Hickey, J.M.; Pong-Wong, R.; Daetwyler, H.D.; Calus, M.P.L. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* **2013**, *193*, 327–345. [CrossRef]
32. Hayes, B.J.; Visscher, P.M.; Goddard, M.E. Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res.* **2009**, *91*, 47–60. [CrossRef] [PubMed]
33. Heffner, E.L.; Sorrells, M.E.; Jannink, J.L. Genomic selection for crop improvement. *Crop Sci.* **2009**, *49*, 1–12. [CrossRef]
34. Merrick, L.F.; Carter, A.H. Comparison of genomic selection models for exploring predictive ability of complex traits in breeding programs. *Plant Genome* **2021**, *14*, e20158. [CrossRef] [PubMed]
35. Endelman, J.B. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome J.* **2011**, *4*, 250–255. [CrossRef]
36. Hu, X.; Carver, B.F.; Powers, C.; Yan, L.; Zhu, L.; Chen, C. Effectiveness of Genomic Selection by Response to Selection for Winter Wheat Variety Improvement. *Plant Genome* **2019**, *12*, 180090. [CrossRef] [PubMed]
37. Sandhu, K.S.; Aoun, M.; Morris, C.F.; Carter, A.H. Genomic selection for end-use quality and processing traits in soft white winter wheat breeding program with machine and deep learning models. *Biology* **2021**, *10*, 689. [CrossRef] [PubMed]
38. Sandhu, K.S.; Lozada, D.N.; Zhang, Z.; Pumphrey, M.O.; Carter, A.H. Deep Learning for Predicting Complex Traits in Spring Wheat Breeding Program. *Front. Plant Sci.* **2021**, *11*, 613325. [CrossRef] [PubMed]

39. Battenfield, S.D.; Guzmán, C.; Gaynor, R.C.; Singh, R.P.; Peña, R.J.; Dreisigacker, S.; Fritz, A.K.; Poland, J.A. Genomic Selection for Processing and End-Use Quality Traits in the CIMMYT Spring Bread Wheat Breeding Program. *Plant Genome* **2016**, *9*, plantgenome2016-01. [CrossRef]
40. Plavšin, I.; Gunjača, J.; Šimek, R.; Novoselović, D. Capturing GEI patterns for quality traits in biparental wheat populations. *Agronomy* **2021**, *11*, 1022. [CrossRef]
41. Prashant, R.; Mani, E.; Rai, R.; Gupta, R.K.; Tiwari, R.; Dholakia, B.; Oak, M.; Röder, M.; Kadoo, N.; Gupta, V. Genotype × environment interactions and QTL clusters underlying dough rheology traits in *Triticum aestivum* L. *J. Cereal Sci.* **2015**, *64*, 82–91. [CrossRef]
42. R Core Team R. *A Language and Environment for Statistical Computing*; Foundation for Statistical Computing: Vienna, Austria, 2020.
43. Butler, D.G.; Cullis, B.R.; Gilmour, A.R.; Gogel, B.J.; Thompson, R. *ASReml-R Reference Manual Version 4*; VSN International Ltd.: Hemel Hempstead, UK, 2017.
44. Brien, C. Asremlplus: Augments "ASReml-R" in Fitting Mixed Models and Packages Generally in Exploring Prediction Differences 2021. Package Version 4.2-32. Available online: https://cran.r-project.org/web/packages/asremlPlus/index.html (accessed on 20 October 2021).
45. Karp, A.; Isaac, P.G.; Ingram, D.S. *Molecular Tools for Screening Biodiversity*; Karp, A., Isaac, P.G., Ingram, D.S., Eds.; Chapman & Hall: London, UK, 1998; ISBN 9789401064965.
46. Browning, B.L.; Zhou, Y.; Browning, S.R. A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am. J. Hum. Genet.* **2018**, *103*, 338–348. [CrossRef] [PubMed]
47. Charmet, G.; Tran, L.G.; Auzanneau, J.; Rincent, R.; Bouchet, S. BWGS: A R package for genomic selection and its application to a wheat breeding programme. *PLoS ONE* **2020**, *15*, e0222733. [CrossRef] [PubMed]
48. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1–22. [CrossRef] [PubMed]
49. Habier, D.; Fernando, R.L.; Kizilkaya, K.; Garrick, D.J. Extension of the bayesian alphabet for genomic selection. *BMC Bioinform.* **2011**, *12*, 186. [CrossRef] [PubMed]
50. Park, T.; Casella, G. The Bayesian Lasso. *J. Am. Stat. Assoc.* **2008**, *103*, 681–686. [CrossRef]
51. Heslot, N.; Yang, H.P.; Sorrells, M.E.; Jannink, J.L. Genomic selection in plant breeding: A comparison of models. *Crop Sci.* **2012**, *52*, 146–160. [CrossRef]
52. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2005**, *67*, 301–320. [CrossRef]
53. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
54. Gianola, D.; Van Kaam, J.B.C.H.M. Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics* **2008**, *178*, 2289–2303. [CrossRef]
55. De Los Campos, G.; Gianola, D.; Rosa, G.J.M.; Weigel, K.A.; Crossa, J. Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genet. Res.* **2010**, *92*, 295–308. [CrossRef]
56. Pérez, P.; de los Campos, G. BGLR: A Statistical Package for Whole Genome Regression and Prediction. *Genetics* **2014**, *198*, 483–495. [CrossRef] [PubMed]
57. Tsai, H.Y.; Janss, L.L.; Andersen, J.R.; Orabi, J.; Jensen, J.D.; Jahoor, A.; Jensen, J. Genomic prediction and GWAS of yield, quality and disease-related traits in spring barley and winter wheat. *Sci. Rep.* **2020**, *10*, 3347. [CrossRef]
58. Lado, B.; Vázquez, D.; Quincke, M.; Silva, P.; Aguilar, I.; Gutiérrez, L. Resource allocation optimization with multi-trait genomic prediction for bread wheat (*Triticum aestivum* L.) baking quality. *Theor. Appl. Genet.* **2018**, *131*, 2719–2731. [CrossRef] [PubMed]
59. Hayes, B.J.; Panozzo, J.; Walker, C.K.; Choy, A.L.; Kant, S.; Wong, D.; Tibbits, J.; Daetwyler, H.D.; Rochfort, S.; Hayden, M.J.; et al. Accelerating wheat breeding for end-use quality with multi-trait genomic predictions incorporating near infrared and nuclear magnetic resonance-derived phenotypes. *Theor. Appl. Genet.* **2017**, *130*, 2505–2519. [CrossRef] [PubMed]
60. Yao, J.; Zhao, D.; Chen, X.; Zhang, Y.; Wang, J. Use of genomic selection and breeding simulation in cross prediction for improvement of yield and quality in wheat (*Triticum aestivum* L.). *Crop J.* **2018**, *6*, 353–365. [CrossRef]
61. Michel, S.; Gallee, M.; Löschenberger, F.; Buerstmayr, H.; Kummer, C. Improving the baking quality of bread wheat using rapid tests and genomics: The prediction of dough rheological parameters by gluten peak indices and genomic selection models. *J. Cereal. Sci.* **2017**, *77*, 24–34. [CrossRef]
62. Kristensen, P.S.; Jensen, J.; Andersen, J.R.; Guzmán, C.; Orabi, J.; Jahoor, A. Genomic Prediction and Genome-Wide Association Studies of Flour Yield and Alveograph Quality Traits Using Advanced Winter Wheat Breeding Material. *Genes* **2019**, *10*, 669. [CrossRef]
63. Dawson, J.C.; Endelman, J.B.; Heslot, N.; Crossa, J.; Poland, J.; Dreisigacker, S.; Manès, Y.; Sorrells, M.E.; Jannink, J.L. The use of unbalanced historical data for genomic selection in an international wheat breeding program. *Filed Crops Res.* **2013**, *154*, 12–22. [CrossRef]
64. Heslot, N.; Jannink, J.L.; Sorrells, M.E. Using genomic prediction to characterize environments and optimize prediction accuracy in applied breeding data. *Crop Sci.* **2013**, *53*, 921–933. [CrossRef]
65. Michel, S.; Ametz, C.; Gungor, H.; Epure, D.; Grausgruber, H.; Löschenberger, F.; Buerstmayr, H. Genomic selection across multiple breeding cycles in applied bread wheat breeding. *Theor. Appl. Genet.* **2016**, *129*, 1179–1189. [CrossRef]

66. Liu, G.; Zhao, Y.; Gowda, M.; Longin, C.F.H.; Reif, J.C.; Mette, M.F. Predicting hybrid performances for quality traits through genomic-assisted approaches in Central European wheat. *PLoS ONE* **2016**, *11*, e0158635. [CrossRef] [PubMed]
67. Lorenz, A.J. Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: A simulation experiment. *G3 Genes Genomes Genet.* **2013**, *3*, 481–491. [CrossRef] [PubMed]
68. Edwards, S.M.K.; Buntjer, J.B.; Jackson, R.; Bentley, A.R.; Lage, J.; Byrne, E.; Burt, C.; Jack, P.; Berry, S.; Flatman, E.; et al. The effects of training population design on genomic prediction accuracy in wheat. *Theor. Appl. Genet.* **2019**, *132*, 1943–1952. [CrossRef] [PubMed]
69. Verges, V.L.; van Sanford, D.A. Genomic selection at preliminary yield trial stage: Training population design to predict untested lines. *Agronomy* **2020**, *10*, 60. [CrossRef]
70. Lozada, D.N.; Mason, R.E.; Sarinelli, J.M.; Brown-Guedira, G. Accuracy of genomic selection for grain yield and agronomic traits in soft red winter wheat. *BMC Genet.* **2019**, *20*, 82. [CrossRef]
71. Haile, J.K.; N'Diaye, A.; Clarke, F.; Clarke, J.; Knox, R.; Rutkoski, J.; Bassi, F.M.; Pozniak, C.J. Genomic selection for grain yield and quality traits in durum wheat. *Mol. Breed.* **2018**, *38*, 75. [CrossRef]
72. Juliana, P.; Poland, J.; Huerta-Espino, J.; Shrestha, S.; Crossa, J.; Crespo-Herrera, L.; Toledo, F.H.; Govindan, V.; Mondal, S.; Kumar, U.; et al. Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics. *Nat. Genet.* **2019**, *51*, 1530–1539. [CrossRef]
73. Heffner, E.L.; Jannink, J.-L.; Sorrells, M.E. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genome* **2011**, *4*, 65–75. [CrossRef]
74. Gorjanc, G.; Dumasy, J.F.; Gonen, S.; Gaynor, R.C.; Antolin, R.; Hickey, J.M. Potential of low-coverage genotyping-by-sequencing and imputation for cost-effective genomic selection in biparental segregating populations. *Crop Sci.* **2017**, *57*, 1404–1420. [CrossRef]
75. Michel, S.; Kummer, C.; Gallee, M.; Hellinger, J.; Ametz, C.; Akgöl, B.; Epure, D.; Löschenberger, F.; Buerstmayr, H. Improving the baking quality of bread wheat by genomic selection in early generations. *Theor. Appl. Genet.* **2018**, *131*, 477–493. [CrossRef]
76. Charmet, G.; Storlie, E.; Oury, F.X.; Laurent, V.; Beghin, D.; Chevarin, L.; Lapierre, A.; Perretant, M.R.; Rolland, B.; Heumez, E.; et al. Genome-wide prediction of three important traits in bread wheat. *Mol. Breed.* **2014**, *34*, 1843–1852. [CrossRef] [PubMed]
77. Gao, H.; Su, G.; Janss, L.; Zhang, Y.; Lund, M.S. Model comparison on genomic predictions using high-density markers for different groups of bulls in the Nordic Holstein population. *J. Dairy Sci.* **2013**, *96*, 4678–4687. [CrossRef] [PubMed]
78. Zhao, Y.; Mette, M.F.; Gowda, M.; Longin, C.F.H.; Reif, J.C. Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. *Heredity* **2014**, *112*, 638–645. [CrossRef] [PubMed]
79. Heslot, N.; Akdemir, D.; Sorrells, M.E.; Jannink, J.L. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theor. Appl. Genet.* **2014**, *127*, 463–480. [CrossRef] [PubMed]
80. Bernardo, R. Genotype x Environment Interaction. In *Breeding for Quantitative Traits in Plants*; Stemma Press: Woodbury, MN, USA, 2010; p. 422.
81. Crossa, J.; De Los Campos, G.; Maccaferri, M.; Tuberosa, R.; Burgueño, J.; Pérez-Rodríguez, P. Extending the marker × environment interaction model for genomic-enabled prediction and genome-wide association analysis in durum wheat. *Crop Sci.* **2016**, *56*, 2193–2209. [CrossRef]
82. Crossa, J.; de los Campos, G.; Pérez, P.; Gianola, D.; Burgueño, J.; Araus, J.L.; Makumbi, D.; Singh, R.P.; Dreisigacker, S.; Yan, J.; et al. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* **2010**, *186*, 713–724. [CrossRef]
83. Lado, B.; Barrios, P.G.; Quincke, M.; Silva, P.; Gutiérrez, L. Modeling genotype × environment interaction for genomic selection with unbalanced data from a wheat breeding program. *Crop Sci.* **2016**, *56*, 2165–2179. [CrossRef]
84. Jarquín, D.; Lemes da Silva, C.; Gaynor, R.C.; Poland, J.; Fritz, A.; Howard, R.; Battenfield, S.; Crossa, J. Increasing genomic-enabled prediction accuracy by modeling genotype × environment interactions in kansas wheat. *Plant Genome* **2017**, *10*, plantgenome2016-12. [CrossRef]
85. Lopez-Cruz, M.; Crossa, J.; Bonnett, D.; Dreisigacker, S.; Poland, J.; Jannink, J.L.; Singh, R.P.; Autrique, E.; de los Campos, G. Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. *G3 Genes Genomes Genet.* **2015**, *5*, 569–582. [CrossRef]
86. Burgueño, J.; de los Campos, G.; Weigel, K.; Crossa, J. Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. *Crop Sci.* **2012**, *52*, 707–719. [CrossRef]
87. Ornella, L.; Sukhwinder-Singh; Perez, P.; Burgueño, J.; Singh, R.; Tapia, E.; Bhavani, S.; Dreisigacker, S.; Braun, H.J.; Mathews, K.; et al. Genomic prediction of genetic values for resistance to wheat rusts. *Plant Genome* **2012**, *5*, 136–148. [CrossRef]

Supplementary Figures

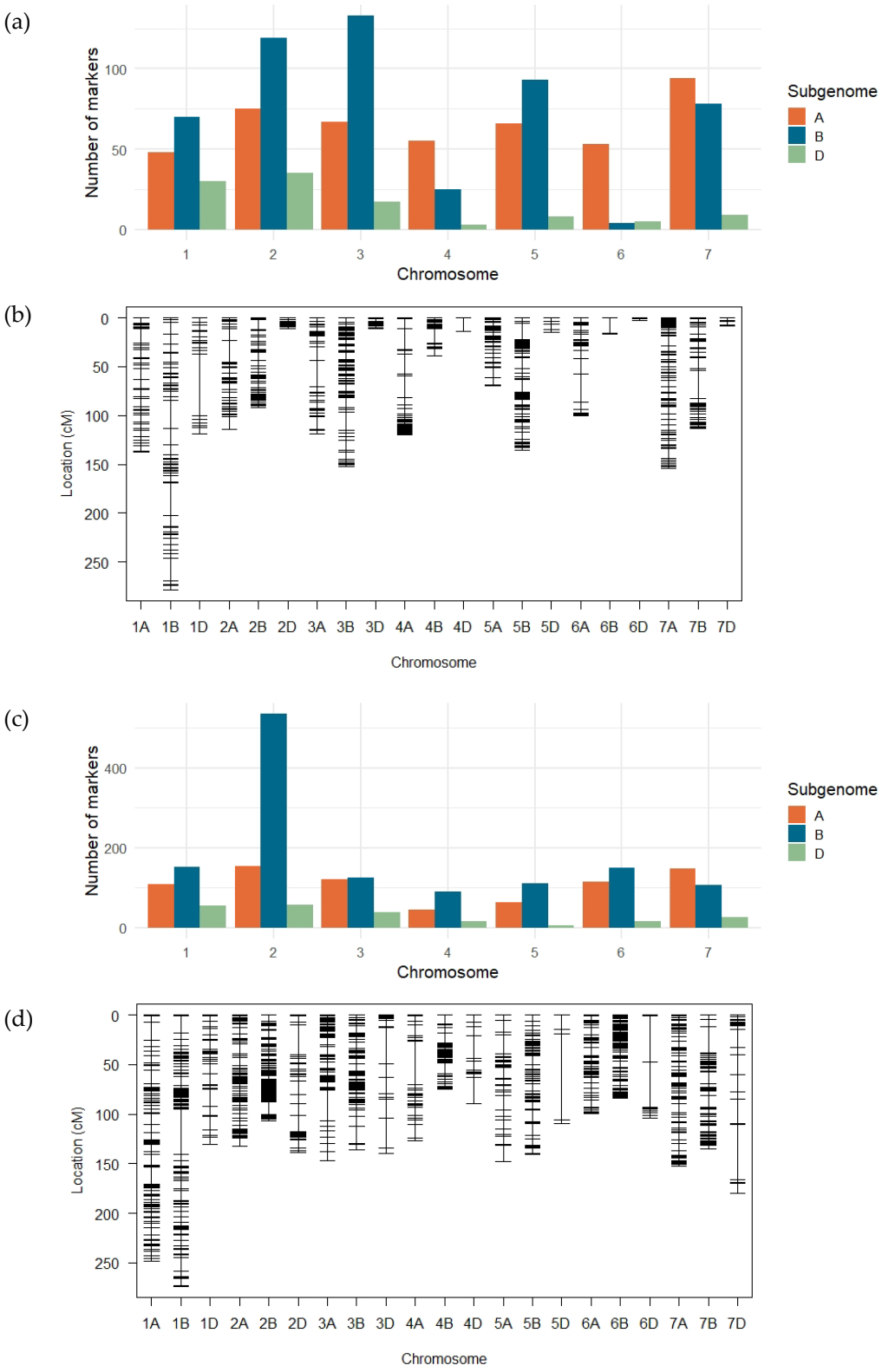


Figure S1. The distribution of SNPs on each chromosome for (a) BK population and (c) MG population together with genetic map of all available SNPs for (c) BK population and (d) MG population.

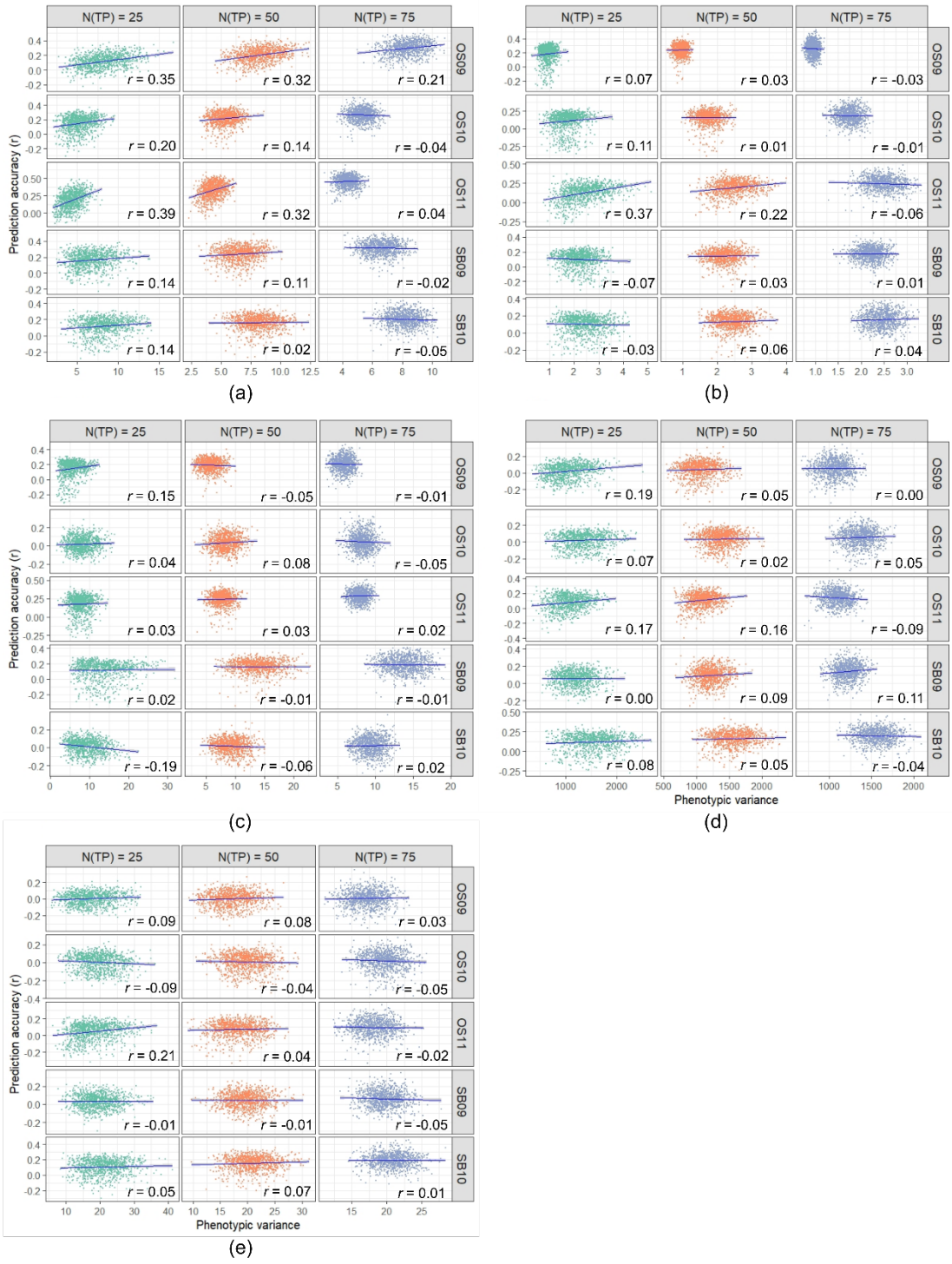


Figure S2. Phenotypic variance of randomly selected TP plotted against prediction accuracy values obtained using RR-BLUP model for following traits of BK population: (a) WGC, (b) MPT, (c) MTW, (d) MTI, and (e) MPH. For each population-trait-environment combination, three different sizes of TP were used (25, 50 and 75 lines) with remaining lines serving as VP. The number in the angle of each scatter plot represents the observed correlation coefficient.

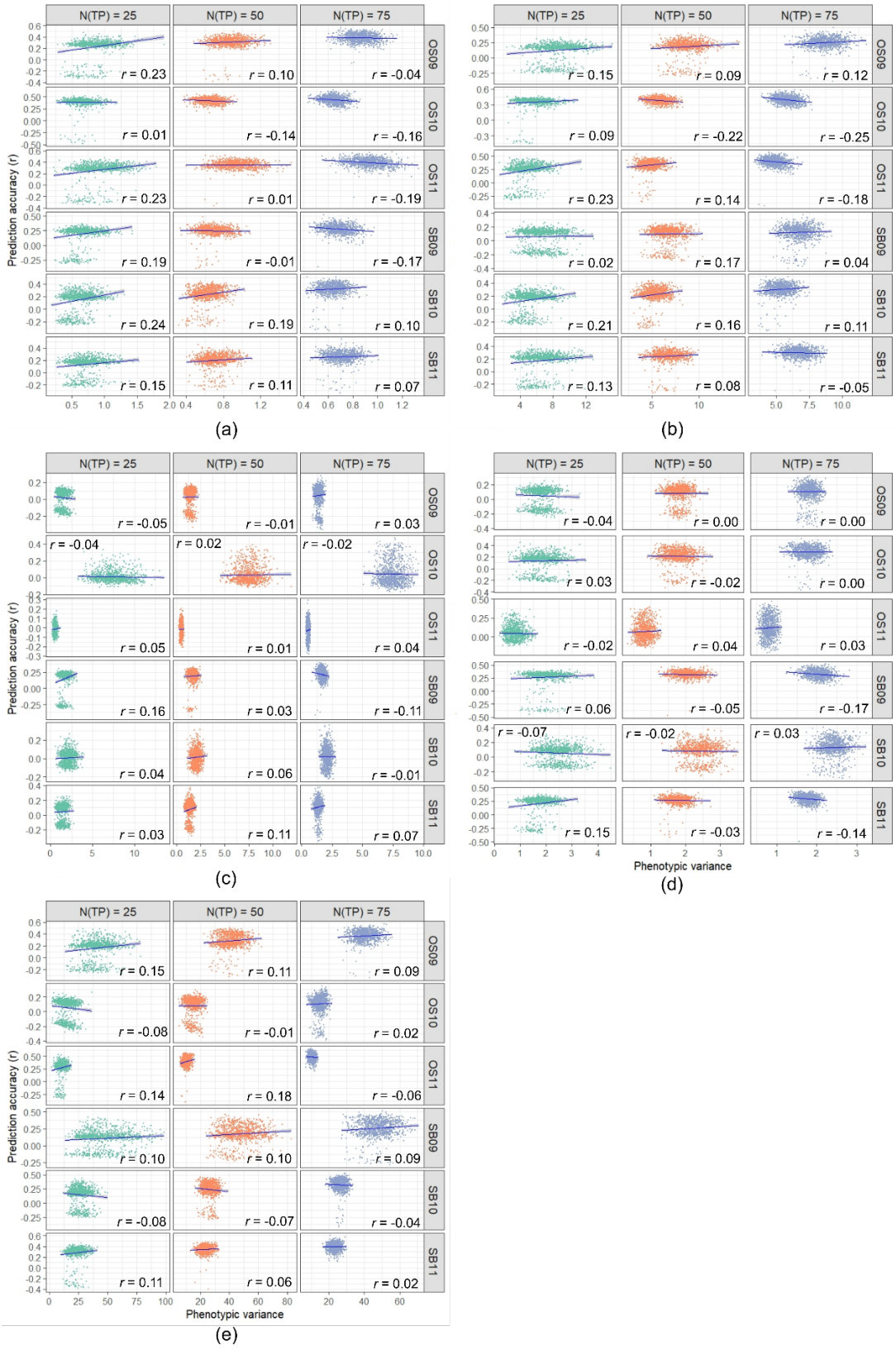


Figure S3. Phenotypic variance of randomly selected TP plotted against prediction accuracy values obtained using RR-BLUP model for following traits of MG population: (a) GPC, (b) WGC, (c) TW, (d) MPT, and (e) MTW. For each population-trait-environment combination, three different sizes of TP were used (25, 50 and 75 lines) with remaining lines serving as VP. The number in the angle of each scatter plot represents the observed correlation coefficient.

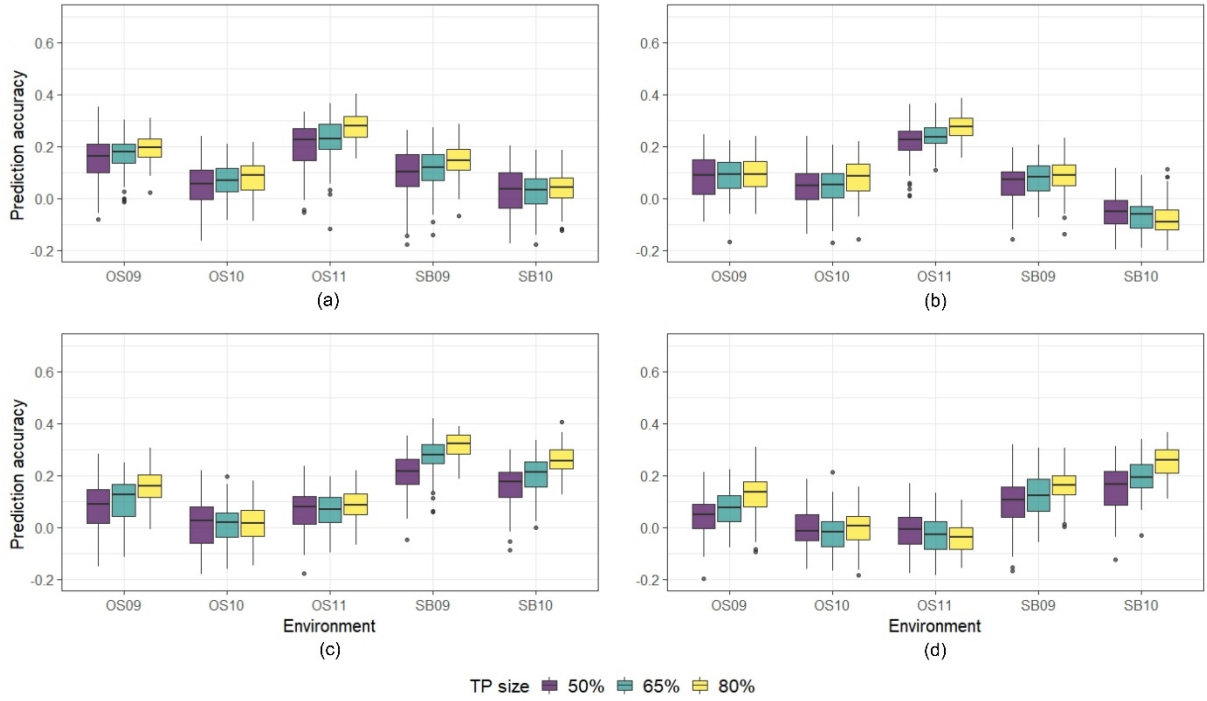


Figure S4. Prediction accuracy values obtained using RR-BLUP model and three different sizes of TP (50%, 65% and 80% of the total number of lines in a population) for following traits of BK population: (a) MPT, (b) MTW, (c) MTI, and (d) MPH.

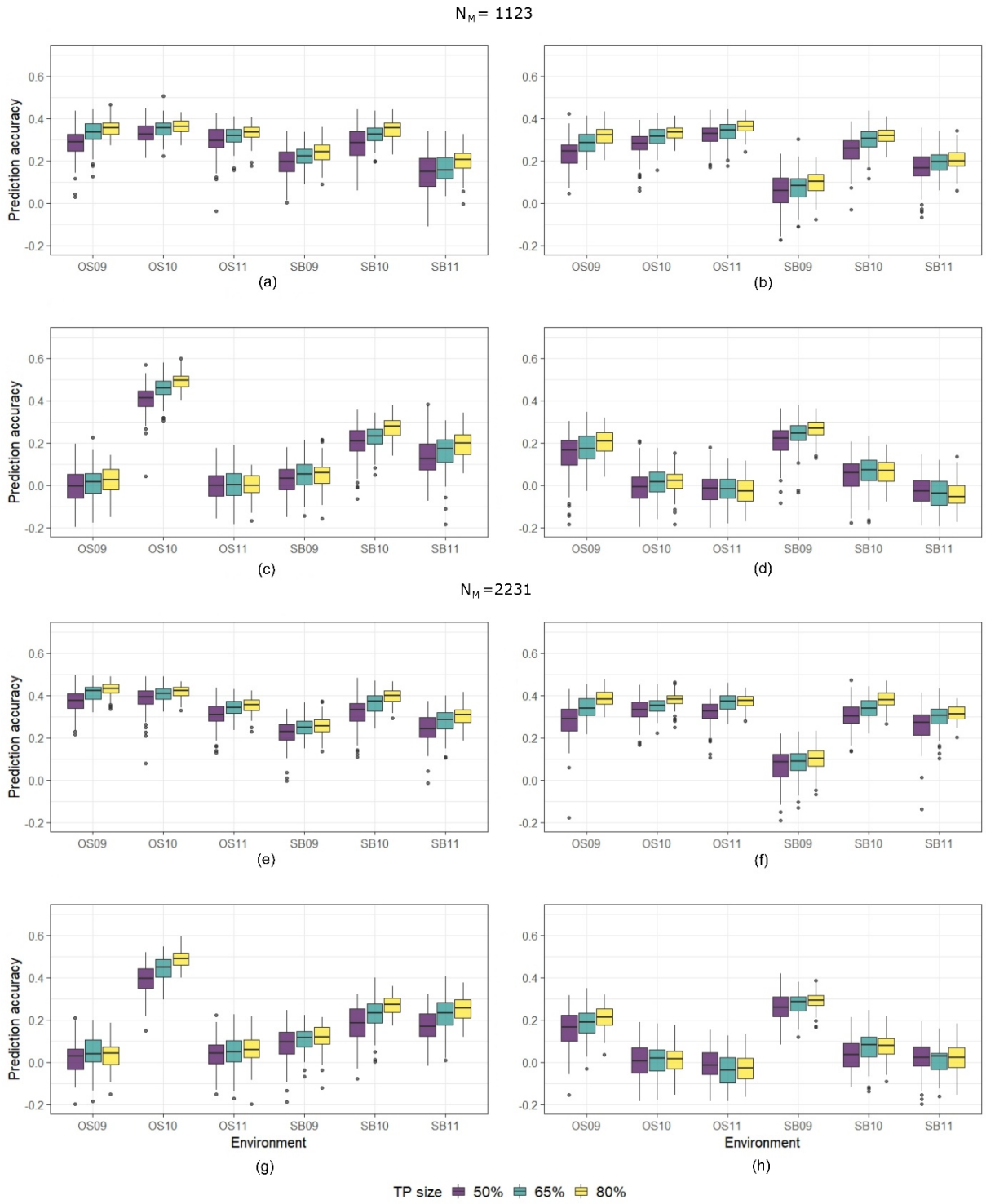


Figure S5. Prediction accuracy values obtained using RR-BLUP model and three different sizes of TP (50%, 65% and 80% of the total number of lines in a population) for following traits of MG population: (a) GPC, (b) WGC, (c) TW, and (d) MTI.

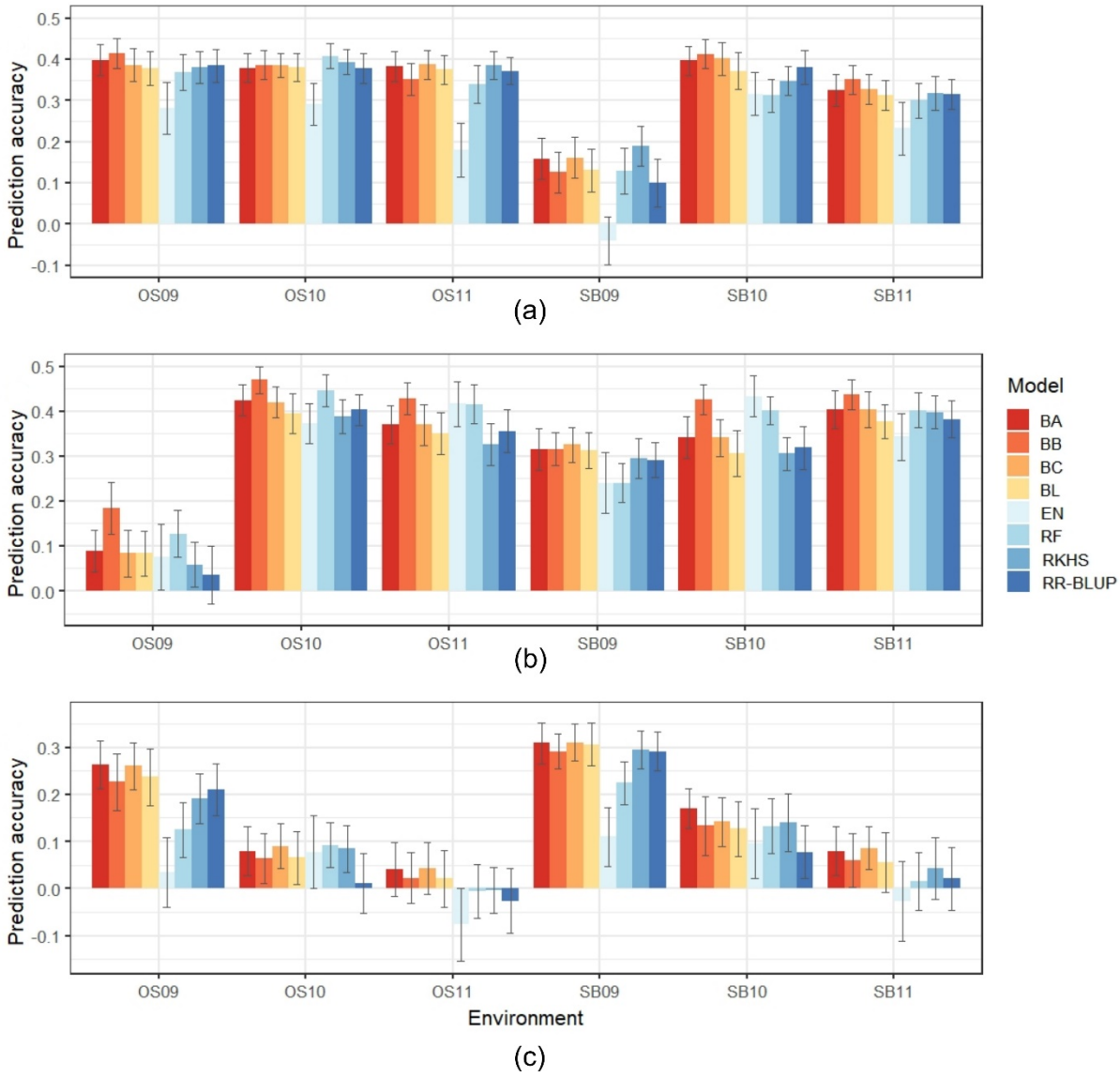


Figure S6. Prediction accuracies for MG population and traits (a) WGC, (b) MPT, and (c) MTI evaluated with eight different prediction models. Error bars denote standard deviation.

Supplementary Tables

Table S1. Mean MSEP values estimated for both populations using RR-BLUP model. Standard deviation values are indicated in parenthesis.

Env.	TP size	GPC	WGC	TW	MPT	MTW	MTI	MPH
		BK population						
OS09	50%	0.04 (0.02)	0.08 (0.05)	0.17 (0.02)	0.07 (0.02)	0.21 (0.06)	0.82 (0.63)	0.09 (0.06)
	65%	0.02 (0.01)	0.06 (0.04)	0.17 (0.02)	0.07 (0.02)	0.19 (0.05)	0.94 (0.81)	0.1 (0.07)
	80%	0.02 (0.01)	0.05 (0.03)	0.17 (0.02)	0.06 (0.01)	0.2 (0.03)	1.19 (0.69)	0.12 (0.08)
OS10	50%	0.05 (0.02)	0.07 (0.05)	0.16 (0.06)	0.16 (0.03)	0.22 (0.05)	1.67 (0.70)	0.16 (0.07)
	65%	0.04 (0.01)	0.06 (0.04)	0.17 (0.05)	0.15 (0.03)	0.23 (0.06)	1.52 (0.74)	0.15 (0.07)
	80%	0.03 (0.01)	0.05 (0.03)	0.18 (0.04)	0.16 (0.02)	0.25 (0.04)	1.29 (0.65)	0.14 (0.06)
OS11	50%	0.03 (0.01)	0.1 (0.05)	0.06 (0.03)	0.17 (0.05)	0.1 (0.05)	0.82 (0.36)	0.13 (0.04)
	65%	0.03 (0.01)	0.1 (0.04)	0.06 (0.02)	0.17 (0.03)	0.09 (0.04)	0.71 (0.38)	0.12 (0.03)
	80%	0.02 (0.01)	0.08 (0.03)	0.06 (0.02)	0.16 (0.02)	0.08 (0.04)	0.59 (0.28)	0.12 (0.02)
SB09	50%	0.10 (0.02)	0.34 (0.05)	0.12 (0.03)	0.07 (0.02)	0.04 (0.02)	4.82 (0.82)	0.32 (0.08)
	65%	0.09 (0.01)	0.32 (0.04)	0.11 (0.02)	0.07 (0.02)	0.05 (0.03)	4.69 (0.75)	0.29 (0.09)
	80%	0.09 (0.01)	0.32 (0.03)	0.12 (0.01)	0.06 (0.02)	0.05 (0.03)	4.14 (0.66)	0.26 (0.07)
SB10	50%	0.02 (0.01)	0.1 (0.07)	0.18 (0.05)	0.14 (0.03)	0.11 (0.04)	1 (0.71)	0.15 (0.11)
	65%	0.02 (0.01)	0.12 (0.06)	0.19 (0.04)	0.13 (0.03)	0.11 (0.03)	1.09 (0.85)	0.22 (0.09)
	80%	0.01 (0.004)	0.13 (0.05)	0.2 (0.03)	0.13 (0.02)	0.11 (0.02)	1.31 (0.78)	0.28 (0.07)

Env.	TP size	MG population						
OS09	50%	0.06 (0.02)	0.21 (0.07)	0.03 (0.01)	0.1 (0.005)	0.13 (0.08)	1.57 (0.74)	0.19 (0.11)
	65%	0.05 (0.02)	0.2 (0.06)	0.02 (0.01)	0.1 (0.005)	0.1 (0.07)	1.57 (0.55)	0.17 (0.10)
	80%	0.04 (0.02)	0.19 (0.05)	0.03 (0.01)	0.1 (0.005)	0.07 (0.05)	1.43 (0.52)	0.13 (0.09)
OS10	50%	0.01 (0.003)	0.07 (0.03)	0.25 (0.07)	0.21 (0.03)	0.15 (0.07)	2.54 (0.46)	0.28 (0.07)
	65%	0.01 (0.004)	0.06 (0.03)	0.21 (0.08)	0.2 (0.03)	0.13 (0.07)	2.45 (0.41)	0.28 (0.05)
	80%	0.01 (0.004)	0.06 (0.02)	0.19 (0.05)	0.21 (0.02)	0.14 (0.06)	2.46 (0.26)	0.29 (0.04)
OS11	50%	0.01 (0.004)	0.03 (0.02)	0.02 (0.01)	0.01 (0.004)	0.31 (0.07)	0.93 (0.33)	0.05 (0.04)
	65%	0.01 (0.005)	0.02 (0.01)	0.02 (0.01)	0.01 (0.004)	0.36 (0.07)	0.91 (0.26)	0.04 (0.03)
	80%	0.01 (0.005)	0.01 (0.005)	0.02 (0.005)	0.01 (0.004)	0.36 (0.06)	0.9 (0.19)	0.03 (0.02)
SB09	50%	0.01 (0.005)	0.04 (0.03)	0.01 (0.005)	0.02 (0.01)	0.26 (0.18)	1.12 (0.63)	0.12 (0.09)
	65%	0.01 (0.004)	0.03 (0.02)	0.02 (0.01)	0.02 (0.01)	0.24 (0.12)	0.93 (0.56)	0.11 (0.09)
	80%	0.01 (0.004)	0.02 (0.01)	0.01 (0.005)	0.01 (0.005)	0.23 (0.11)	0.84 (0.46)	0.13 (0.08)
SB10	50%	0.02 (0.01)	0.05 (0.04)	0.09 (0.03)	0.09 (0.03)	0.3 (0.11)	1.27 (0.69)	0.2 (0.14)
	65%	0.02 (0.01)	0.05 (0.04)	0.09 (0.02)	0.1 (0.03)	0.27 (0.11)	1.4 (0.61)	0.25 (0.14)
	80%	0.02 (0.01)	0.05 (0.04)	0.09 (0.02)	0.1 (0.02)	0.24 (0.09)	1.52 (0.68)	0.26 (0.12)
SB11	50%	0.06 (0.02)	0.15 (0.06)	0.02 (0.01)	0.15 (0.02)	0.1 (0.08)	2.27 (0.75)	0.09 (0.11)
	65%	0.06 (0.02)	0.14 (0.06)	0.02 (0.01)	0.15 (0.02)	0.1 (0.08)	2.12 (0.54)	0.08 (0.07)
	80%	0.05 (0.02)	0.11 (0.05)	0.02 (0.01)	0.16 (0.02)	0.09 (0.06)	2.18 (0.44)	0.08 (0.07)

Table S2. Mean prediction accuracy values estimated for MG population using eight different prediction models. Standard deviation values are indicated in parenthesis.

Trait	Env.	Model							
		RR-BLUP	**EN**	**BA**	**BB**	**BC**	**BL**	**RF**	**RKHS**
GPC	OS09	0.43 (0.03)	0.34 (0.06)	0.45 (0.04)	**0.46** (0.03)	0.45 (0.03)	0.44 (0.04)	0.43 (0.04)	0.45 (0.03)
	OS10	0.42 (0.03)	0.33 (0.05)	0.42 (0.03)	0.42 (0.03)	0.43 (0.03)	0.42 (0.03)	**0.44** (0.03)	0.43 (0.03)
	OS11	0.35 (0.04)	0.15 (0.07)	0.35 (0.04)	0.33 (0.04)	0.35 (0.03)	0.35 (0.04)	0.3 (0.05)	**0.36** (0.04)
	SB09	0.26 (0.04)	0.1 (0.08)	0.27 (0.04)	0.28 (0.04)	0.28 (0.03)	0.27 (0.04)	0.24 (0.04)	**0.29** (0.04)
	SB10	0.39 (0.04)	0.33 (0.05)	0.4 (0.04)	**0.41** (0.04)	**0.41** (0.03)	0.38 (0.04)	0.35 (0.04)	0.38 (0.03)
	SB11	0.3 (0.05)	0.2 (0.07)	0.31 (0.04)	**0.33** (0.04)	0.31 (0.03)	0.3 (0.04)	0.28 (0.05)	0.31 (0.05)
WGC	OS09	0.38 (0.04)	0.28 (0.06)	0.4 (0.04)	**0.41** (0.04)	0.39 (0.04)	0.38 (0.04)	0.37 (0.04)	0.38 (0.04)
	OS10	0.38 (0.04)	0.29 (0.05)	0.38 (0.04)	0.39 (0.04)	0.39 (0.03)	0.38 (0.03)	**0.41** (0.03)	0.39 (0.03)
	OS11	0.37 (0.03)	0.18 (0.07)	0.38 (0.04)	0.35 (0.04)	**0.39** (0.03)	0.37 (0.04)	0.34 (0.05)	**0.39** (0.03)
	SB09	0.1 (0.06)	-0.04 (0.06)	0.16 (0.05)	0.13 (0.05)	0.16 (0.05)	0.13 (0.05)	0.13 (0.05)	**0.19** (0.05)
	SB10	0.38 (0.04)	0.32 (0.05)	0.4 (0.04)	**0.41** (0.04)	0.4 (0.04)	0.37 (0.04)	0.31 (0.04)	0.35 (0.04)
	SB11	0.32 (0.04)	0.23 (0.06)	0.32 (0.04)	**0.35** (0.03)	0.33 (0.04)	0.31 (0.04)	0.3 (0.04)	0.32 (0.04)
TW	OS09	0.04 (0.06)	-0.04 (0.09)	**0.14** (0.06)	0.12 (0.06)	0.13 (0.04)	0.08 (0.05)	0.04 (0.06)	0.1 (0.05)
	OS10	**0.49** (0.04)	0.36 (0.06)	**0.49** (0.04)	0.46 (0.04)	**0.49** (0.04)	0.48 (0.04)	0.39 (0.04)	0.44 (0.04)
	OS11	0.06 (0.07)	-0.08 (0.07)	**0.14** (0.05)	0.08 (0.06)	0.13 (0.05)	0.1 (0.05)	0.09 (0.05)	0.11 (0.05)
	SB09	0.12 (0.06)	0.01 (0.07)	0.15 (0.05)	0.14 (0.05)	**0.16** (0.04)	**0.16** (0.04)	0.12 (0.06)	**0.16** (0.05)
	SB10	0.27 (0.05)	0.1 (0.07)	**0.3** (0.05)	0.26 (0.06)	**0.3** (0.04)	0.26 (0.05)	0.23 (0.04)	0.25 (0.04)
	SB11	0.25 (0.06)	0.26 (0.06)	0.29 (0.05)	**0.31** (0.05)	0.28 (0.05)	0.26 (0.06)	0.24 (0.06)	0.2 (0.06)
MPT	OS09	0.04 (0.06)	0.08 (0.07)	0.09 (0.05)	**0.18** (0.06)	0.08 (0.05)	0.08 (0.05)	0.13 (0.05)	0.06 (0.05)
	OS10	0.4 (0.03)	0.37 (0.04)	0.42 (0.03)	**0.47** (0.03)	0.42 (0.03)	0.4 (0.04)	0.45 (0.04)	0.39 (0.04)
	OS11	0.36 (0.05)	0.42 (0.05)	0.37 (0.04)	**0.43** (0.04)	0.37 (0.05)	0.35 (0.05)	0.42 (0.04)	0.33 (0.05)
	SB09	0.29 (0.04)	0.24 (0.07)	0.31 (0.05)	0.32 (0.04)	**0.33** (0.04)	0.31 (0.04)	0.24 (0.04)	0.29 (0.04)
	SB10	0.32 (0.05)	**0.43** (0.05)	0.34 (0.05)	**0.43** (0.03)	0.34 (0.04)	0.31 (0.05)	0.4 (0.03)	0.31 (0.04)
	SB11	0.38 (0.04)	0.34 (0.05)	0.4 (0.04)	**0.44** (0.03)	0.4 (0.04)	0.38 (0.04)	0.4 (0.04)	0.4 (0.04)
MTW	OS09	0.42 (0.03)	0.37 (0.05)	0.43 (0.03)	**0.46** (0.04)	0.42 (0.04)	0.42 (0.04)	0.45 (0.03)	0.41 (0.04)
	OS10	0.12 (0.07)	0.15 (0.07)	0.17 (0.05)	0.15 (0.06)	0.17 (0.05)	0.13 (0.05)	**0.2** (0.05)	0.15 (0.05)
	OS11	0.53 (0.03)	0.55 (0.04)	0.53 (0.03)	**0.57** (0.03)	0.53 (0.03)	0.53 (0.03)	0.55 (0.03)	0.55 (0.03)
	SB09	0.38 (0.04)	0.38 (0.05)	0.37 (0.03)	**0.42** (0.03)	0.36 (0.04)	0.36 (0.05)	**0.42** (0.03)	0.36 (0.04)
	SB10	0.43 (0.04)	0.4 (0.06)	0.43 (0.03)	**0.45** (0.03)	0.42 (0.04)	0.41 (0.04)	0.41 (0.03)	0.4 (0.03)
	SB11	0.47 (0.03)	0.39 (0.06)	**0.48** (0.04)	**0.48** (0.03)	**0.48** (0.04)	0.46 (0.04)	0.39 (0.05)	0.45 (0.03)
MTI	OS09	0.21 (0.06)	0.03 (0.07)	**0.26** (0.05)	0.23 (0.06)	**0.26** (0.05)	0.24 (0.06)	0.12 (0.06)	0.19 (0.05)
	OS10	0.01 (0.06)	0.08 (0.08)	0.08 (0.05)	0.06 (0.05)	**0.09** (0.05)	0.07 (0.06)	**0.09** (0.05)	**0.09** (0.05)
	OS11	-0.03 (0.07)	-0.08 (0.08)	**0.04** (0.06)	0.02 (0.05)	**0.04** (0.06)	0.02 (0.06)	-0.01 (0.06)	0 (0.05)
	SB09	0.29 (0.04)	0.11 (0.06)	**0.31** (0.04)	0.29 (0.04)	**0.31** (0.04)	**0.31** (0.05)	0.22 (0.04)	0.3 (0.04)
	SB10	0.08 (0.06)	0.1 (0.07)	**0.17** (0.04)	0.13 (0.06)	0.14 (0.05)	0.13 (0.06)	0.13 (0.06)	0.14 (0.06)
	SB11	0.02 (0.07)	-0.03 (0.08)	0.08 (0.05)	0.06 (0.06)	**0.09** (0.05)	0.06 (0.06)	0.02 (0.06)	0.04 (0.06)
MPH	OS09	0.29 (0.05)	0.24 (0.08)	**0.34** (0.04)	0.33 (0.04)	**0.34** (0.04)	0.32 (0.05)	0.23 (0.05)	0.28 (0.04)
	OS10	0.04 (0.06)	0.04 (0.08)	0.1 (0.05)	0.09 (0.05)	0.1 (0.05)	0.09 (0.05)	0.07 (0.05)	**0.11** (0.05)
	OS11	0.05 (0.06)	0 (0.07)	0.09 (0.05)	0.1 (0.06)	**0.11** (0.05)	0.09 (0.05)	0.07 (0.05)	0.07 (0.05)
	SB09	0.26 (0.05)	0.11 (0.07)	**0.29** (0.05)	0.28 (0.04)	**0.29** (0.04)	0.28 (0.05)	0.24 (0.05)	0.28 (0.04)
	SB10	0.22 (0.05)	0.18 (0.07)	**0.27** (0.05)	0.25 (0.05)	**0.27** (0.04)	0.24 (0.05)	0.23 (0.05)	0.26 (0.05)
	SB11	0.1 (0.05)	-0.06 (0.07)	**0.16** (0.05)	0.11 (0.05)	**0.16** (0.04)	0.14 (0.05)	0.04 (0.05)	0.12 (0.05)

Table S3. Mean MSEP values estimated for MG population using eight different prediction models. Standard deviation values are indicated in parenthesis.

Trait	Env.	Model							
		RR-BLUP	**EN**	**BA**	**BB**	**BC**	**BL**	**RF**	**RKHS**
GPC	OS09	0.04 (0.02)	0.04 (0.02)	0.04 (0.01)	0.04 (0.01)	0.04 (0.01)	0.05 (0.01)	0.06 (0.01)	0.06 (0.005)
	OS10	0.01 (0.004)	0.01 (0.005)	0.01 (0.005)	0.01 (0.005)	0.01 (0.005)	0.01 (0.004)	0.01 (0.005)	0.01 (0.004)
	OS11	0.01 (0.005)	0.01 (0.005)	0.02 (0.01)	0.01 (0.005)	0.02 (0.01)	0.01 (0.005)	0.01 (0.005)	0.01 (0.004)
	SB09	0.01 (0.004)	0.02 (0.01)	0.01 (0.005)	0.01 (0.005)	0.01 (0.01)	0.01 (0.005)	0.01 (0.005)	0.01 (0.004)
	SB10	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)	0.01 (0.005)	0.02 (0.01)	0.01 (0.003)
	SB11	0.05 (0.02)	0.06 (0.02)	0.04 (0.02)	0.05 (0.01)	0.05 (0.01)	0.06 (0.01)	0.05 (0.02)	0.06 (0.01)
WGC	OS09	0.19 (0.05)	0.18 (0.06)	0.18 (0.04)	0.18 (0.04)	0.19 (0.04)	0.20 (0.04)	0.20 (0.04)	0.23 (0.03)
	OS10	0.06 (0.02)	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)	0.06 (0.02)	0.05 (0.03)	0.06 (0.02)
	OS11	0.01 (0.005)	0.04 (0.03)	0.03 (0.02)	0.02 (0.02)	0.03 (0.02)	0.02 (0.01)	0.04 (0.02)	0.01 (0.005)
	SB09	0.02 (0.01)	0.03 (0.02)	0.03 (0.02)	0.03 (0.02)	0.03 (0.02)	0.02 (0.01)	0.05 (0.03)	0.02 (0.01)
	SB10	0.05 (0.04)	0.07 (0.05)	0.05 (0.03)	0.06 (0.03)	0.05 (0.03)	0.03 (0.02)	0.06 (0.03)	0.02 (0.01)
	SB11	0.11 (0.05)	0.14 (0.05)	0.08 (0.05)	0.10 (0.04)	0.10 (0.04)	0.13 (0.04)	0.11 (0.04)	0.14 (0.03)
TW	OS09	0.03 (0.01)	0.02 (0.01)	0.03 (0.01)	0.02 (0.01)	0.03 (0.01)	0.03 (0.01)	0.02 (0.01)	0.03 (0.01)
	OS10	0.19 (0.05)	0.19 (0.07)	0.18 (0.05)	0.17 (0.04)	0.20 (0.05)	0.21 (0.05)	0.28 (0.03)	0.27 (0.03)
	OS11	0.02 (0.005)	0.02 (0.01)	0.03 (0.01)	0.02 (0.01)	0.03 (0.01)	0.02 (0.01)	0.04 (0.01)	0.03 (0.01)
	SB09	0.01 (0.005)	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)	0.01 (0.005)	0.02 (0.01)	0.02 (0.01)
	SB10	0.09 (0.02)	0.09 (0.02)	0.08 (0.02)	0.08 (0.02)	0.08 (0.02)	0.09 (0.01)	0.09 (0.01)	0.09 (0.01)
	SB11	0.02 (0.01)	0.03 (0.02)	0.02 (0.01)	0.03 (0.02)	0.02 (0.01)	0.01 (0.005)	0.01 (0.005)	0.01 (0.005)
MPT	OS09	0.10 (0.005)	0.11 (0.01)	0.10 (0.02)	0.09 (0.02)	0.10 (0.02)	0.10 (0.01)	0.09 (0.02)	0.10 (0.02)
	OS10	0.21 (0.02)	0.20 (0.03)	0.19 (0.02)	0.20 (0.02)	0.20 (0.02)	0.20 (0.02)	0.17 (0.03)	0.18 (0.02)
	OS11	0.01 (0.004)	0.02 (0.01)	0.01 (0.005)	0.01 (0.01)	0.01 (0.01)	0.01 (0.005)	0.02 (0.01)	0.01 (0.004)
	SB09	0.01 (0.005)	0.03 (0.02)	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)	0.01 (0.005)	0.02 (0.01)	0.01 (0.005)
	SB10	0.10 (0.02)	0.10 (0.03)	0.09 (0.02)	0.10 (0.03)	0.09 (0.02)	0.08 (0.02)	0.05 (0.02)	0.08 (0.01)
	SB11	0.16 (0.02)	0.16 (0.03)	0.15 (0.02)	0.15 (0.02)	0.15 (0.01)	0.15 (0.01)	0.15 (0.02)	0.14 (0.01)
MTW	OS09	0.07 (0.05)	0.12 (0.07)	0.07 (0.05)	0.09 (0.06)	0.08 (0.07)	0.08 (0.05)	0.13 (0.08)	0.14 (0.07)
	OS10	0.14 (0.06)	0.13 (0.08)	0.12 (0.06)	0.14 (0.07)	0.13 (0.06)	0.13 (0.05)	0.08 (0.06)	0.10 (0.06)
	OS11	0.36 (0.06)	0.40 (0.09)	0.36 (0.05)	0.39 (0.07)	0.36 (0.05)	0.34 (0.05)	0.42 (0.05)	0.27 (0.04)
	SB09	0.23 (0.11)	0.43 (0.11)	0.23 (0.12)	0.38 (0.11)	0.24 (0.11)	0.20 (0.09)	0.44 (0.10)	0.14 (0.09)
	SB10	0.24 (0.09)	0.20 (0.10)	0.29 (0.09)	0.22 (0.08)	0.30 (0.07)	0.31 (0.08)	0.18 (0.07)	0.35 (0.06)
	SB11	0.09 (0.06)	0.11 (0.09)	0.10 (0.06)	0.07 (0.05)	0.09 (0.05)	0.07 (0.05)	0.09 (0.06)	0.09 (0.06)
MTI	OS09	1.43 (0.52)	1.61 (0.60)	1.43 (0.49)	1.46 (0.55)	1.35 (0.52)	1.45 (0.45)	1.56 (0.63)	1.57 (0.36)
	OS10	2.46 (0.26)	2.36 (0.50)	2.68 (0.36)	2.48 (0.35)	2.61 (0.31)	2.57 (0.28)	2.31 (0.47)	2.35 (0.26)
	OS11	0.9 (0.19)	0.94 (0.22)	0.75 (0.33)	0.76 (0.26)	0.66 (0.32)	0.74 (0.25)	0.77 (0.32)	0.85 (0.25)
	SB09	0.84 (0.46)	1.25 (0.58)	0.49 (0.35)	0.57 (0.42)	0.52 (0.39)	0.70 (0.44)	0.60 (0.39)	0.84 (0.36)
	SB10	1.52 (0.68)	1.29 (0.73)	2.1 (0.68)	1.71 (0.51)	2.03 (0.46)	1.68 (0.54)	1.63 (0.39)	1.45 (0.30)
	SB11	2.18 (0.44)	2.01 (0.77)	2.85 (0.52)	2.57 (0.45)	2.85 (0.47)	2.50 (0.56)	3.42 (0.71)	2.38 (0.38)
MPH	OS09	0.13 (0.09)	0.26 (0.12)	0.08 (0.06)	0.12 (0.09)	0.09 (0.06)	0.12 (0.08)	0.14 (0.11)	0.15 (0.08)
	OS10	0.29 (0.04)	0.32 (0.04)	0.33 (0.07)	0.31 (0.04)	0.33 (0.05)	0.29 (0.04)	0.33 (0.05)	0.27 (0.03)
	OS11	0.03 (0.02)	0.05 (0.04)	0.06 (0.04)	0.05 (0.03)	0.06 (0.04)	0.05 (0.04)	0.03 (0.02)	0.03 (0.02)
	SB09	0.13 (0.08)	0.13 (0.11)	0.18 (0.01)	0.16 (0.08)	0.15 (0.08)	0.09 (0.07)	0.17 (0.06)	0.07 (0.04)
	SB10	0.26 (0.12)	0.17 (0.12)	0.32 (0.01)	0.26 (0.09)	0.28 (0.09)	0.22 (0.10)	0.22 (0.06)	0.20 (0.05)
	SB11	0.08 (0.07)	0.06 (0.05)	0.16 (0.07)	0.09 (0.06)	0.14 (0.06)	0.08 (0.05)	0.16 (0.08)	0.07 (0.05)

Table S4. Pedigree of winter wheat genotypes used in the study.

Name	Type	Pedigree
Monika	Commercial variety	F21078-82 / Srpanjka
Golubica	Commercial variety	Slavonija / Gemini
Bezostaya-1	Commercial variety	Skorospelka 2 / Lutescens 17
Klara	Commercial variety	Slavonija / Zg 5328-75
MG	RIL	Monika / Golubica
BK	RIL	Bezostaya-1 / Klara