



Comparison of Single-Trait and Multi-Trait Genome-Wide Association Models and Inclusion of Correlated Traits in the Dissection of the Genetic Architecture of a Complex Trait in a Breeding Program

Lance F. Merrick, Adrienne B. Burke, Zhiwu Zhang and Arron H. Carter*

Department of Crop and Soil Sciences, Washington State University, Pullman, WA, United States

OPEN ACCESS

Edited by:

Lee Hickey,
The University of Queensland,
Australia

Reviewed by:

Philomin Juliana,
International Maize and Wheat
Improvement Center, Mexico
Jianbo He,
Nanjing Agricultural University, China

*Correspondence:

Arron H. Carter
ahcarter@wsu.edu

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 08 September 2021

Accepted: 29 December 2021

Published: 28 January 2022

Citation:

Merrick LF, Burke AB, Zhang Z
and Carter AH (2022) Comparison
of Single-Trait and Multi-Trait
Genome-Wide Association Models
and Inclusion of Correlated Traits
in the Dissection of the Genetic
Architecture of a Complex Trait in a
Breeding Program.
Front. Plant Sci. 12:772907.
doi: 10.3389/fpls.2021.772907

Unknown genetic architecture makes it difficult to characterize the genetic basis of traits and associated molecular markers because of the complexity of small effect quantitative trait loci (QTLs), environmental effects, and difficulty in phenotyping. Seedling emergence of wheat (*Triticum aestivum* L.) from deep planting, has a poorly understood genetic architecture, is a vital factor affecting stand establishment and grain yield, and is historically correlated with coleoptile length. This study aimed to dissect the genetic architecture of seedling emergence while accounting for correlated traits using one multi-trait genome-wide association study (MT-GWAS) model and three single-trait GWAS (ST-GWAS) models. The ST-GWAS models included one single-locus model [mixed-linear model (MLM)] and two multi-locus models [fixed and random model circulating probability unification (FarmCPU) and Bayesian information and linkage-disequilibrium iteratively nested keyway (BLINK)]. We conducted GWAS using two populations. The first population consisted of 473 varieties from a diverse association mapping panel phenotyped from 2015 to 2019. The second population consisted of 279 breeding lines phenotyped in 2015 in Lind, WA, with 40,368 markers. We also compared the inclusion of coleoptile length and markers associated with reduced height as covariates in our ST-GWAS models. ST-GWAS found 107 significant markers across 19 chromosomes, while MT-GWAS found 82 significant markers across 14 chromosomes. The FarmCPU and BLINK models, including covariates, were able to identify many small effect markers while identifying large effect markers on chromosome 5A. By using multi-locus model breeding, programs can uncover the complex nature of traits to help identify candidate genes and the underlying architecture of a trait, such as seedling emergence.

Keywords: covariates, single-locus model, multi-locus model, reduced height alleles, pleiotropic effects, seedling emergence, coleoptile length

INTRODUCTION

Complex traits are controlled by many quantitative trait loci (QTLs) and are influenced by environmental conditions (Bernardo, 2020). Challenges due to complexity, small effect QTLs, and difficulty in phenotyping can make it difficult to characterize the genetic basis of traits and associated molecular markers, especially in biparental populations. Linkage mapping for complex traits can often result in inconsistent estimated QTL effects (Bernardo, 2020; Tibbs Cortes et al., 2021). Unfortunately, many of these complex traits are essential for selection in plant breeding programs, typically being associated with yield potential, end-use quality, and certain biotic and abiotic stress types of tolerance. Therefore, there is a need to increase the knowledge of the inheritance and genetic architecture of these complex traits (Bernardo, 2020).

In recent years, the use of genome-wide association studies (GWASs) has enabled the discovery of QTLs in a collection of diverse populations or diversity panels rather than using a mapping population (Lander and Schork, 1994). GWAS can be performed to dissect the genetic architecture of a trait by exploiting all historical recombination events in the population and allow for the ability to understand the genetic basis by identifying the associations between genetic markers and phenotypes (Lipka et al., 2015). Not only are complex traits influenced by the environment and multiple QTLs, but they also interact with correlated traits that result in a complex genetic architecture. Using heritable covariates and correlated secondary traits can help account for confounding factors that bias marker effects and improve the power of a GWAS model (Aschard et al., 2015). Additionally, using population structure or genetic relatedness controls p -value inflation for each marker and reduces false positives (Tibbs Cortes et al., 2021).

Statistical models have been developed for GWAS to distinguish real associations from false positives caused by population structure and linkage disequilibrium (LD). Some of the first GWAS models, such as general linear models and mixed-linear models (MLMs), were single-locus, single-trait GWAS (ST-GWAS) models created to implement the covariates along with kinship matrices (Yu et al., 2006). However, these simple models resulted in false negatives caused by weakened associations in order to control inflation of p -values due to population structure (Liu et al., 2016). MLMs were improved upon by compressed MLMs (CMLMs), which cluster individuals and use them as random effects rather than individual genotypic effects (Zhang et al., 2010). CMLMs were further improved by using pseudo quantitative trait nucleotides (QTNs) to derive kinship instead of all the genetic markers. The settlement of MLM under progressively exclusive relationship (SUPER) model sorts markers by association and then combines them into bins with the most significant marker designated as pseudo QTNs and then used to derive a reduced kinship matrix (Wang et al., 2014). These methods improved computational efficiency and statistical power over MLMs because of weakening real associations when controlling the p -value inflation while accounting for population structure (Liu et al., 2016). Compared to single-locus models, newer multi-locus models were then developed to test multiple

markers simultaneously (Liu et al., 2016). These multi-locus GWAS models, such as FarmCPU and BLINK, allow for the evaluation of big datasets while also reducing false positives and negatives (Huang et al., 2019). FarmCPU bins markers and fits them as cofactors to control false positives for testing the rest of the markers in a fixed-effect model, and then a random effect model is used to select the associated markers. BLINK eliminates the FarmCPU assumption that causal genes are evenly distributed across the genome, which improves speed, because of the optimization of bin size and number no longer being required (Huang et al., 2019). Additionally, multi-trait GWAS (MT-GWAS) can be used to analyze multiple traits simultaneously. MT-GWAS methods were developed to increase statistical power and identify pleiotropic loci (Porter and O'Reilly, 2017). Correlations and pleiotropy can be used to increase power compared to ST-GWAS (Galeslout et al., 2014). MT-GWAS methods display increase in power even when traits display negative correlation, when only one of the traits is associated with the loci, or when genetic correlations among traits are weak (Galeslout et al., 2014). However, because of the intricate and pleiotropic nature of quantitative traits, there is no best model for all situations, and it is recommended to compare models to dissect the unique genetic architecture of complex traits (Tibbs Cortes et al., 2021).

In Washington state, seedling emergence of deep-sown winter wheat is a complex trait affected by many factors and is dependent on the environment to display variation (Schillinger et al., 1998; Lutcher et al., 2019). Seedling emergence is dependent on deep sowing at depths of 10–15 cm when precipitation is below 150–300 mm annually (Schillinger et al., 1998; Mohan et al., 2013). Under these limited moisture conditions, a winter-wheat summer fallow rotation is employed. Additionally, rod-weeding is used in the fallow year to control weeds as well as limit evapotranspiration, thereby conserving what moisture may be in the soil at depths beneath the rod-weeding implement. In deep-sowing practices, fast-emerging cultivars are required to emerge before precipitation events create soil crusting that can dramatically decrease emergence. Wheat seedlings have decreased emergence when they cannot penetrate the soil surface because of crusting, due to the inability to germinate under dry soil conditions, or short coleoptiles. Seedling emergence and, therefore, stand establishment are vital factors affecting grain yield in these growing regions in Washington state and can reduce grain yields by 35–40% (Schillinger et al., 1998).

Previous studies have shown a significant positive relationship between coleoptile length and seedling emergence (Allan et al., 1962; Sunderman, 1964; Chastain et al., 1995; Schillinger et al., 1998; Botwright et al., 2001; Schillinger, 2011). The reduced height (*Rht*) genes *Rht-B1b* and *Rht-D1b* are mutant alleles that cause the semi-dwarfing stature of wheat (Vogel et al., 1956; Allan et al., 1962). Dwarfing genes are responsible for short stature and have pleiotropic effects that include gibberellin insensitivity, coleoptile length, yield, protein content, and disease resistance (Gale and Yousefian, 1983; Allan, 1989). Semi-dwarf wheat cultivars have improved resistance to lodging and grain yield but reduced coleoptile length by one-half to three-fourths of the standard varieties at the time of their

development (Allan et al., 1961; Allan, 1989; Mohan et al., 2013). The reduced coleoptile length was due to decreased gibberellic acid response, which reduced cell size and elongation (Allan et al., 1959). Historically, when crusting was present, or other unfavorable conditions, the shorter coleoptiles of semi-dwarf cultivars resulted in poor stand establishment and yield potential (Rebetzke et al., 1999).

Recently, after 60 years of breeding, emergence in modern varieties was shown to have a reduced correlation between emergence and coleoptile length (Mohan et al., 2013). Coleoptile length only accounted for 28% of the variability for seedling emergence, and some lines with short coleoptiles had the best emergence rating. The remaining variability is attributed to many factors that affect seedling emergence, leading to a complex system resulting in stand establishment. As stated previously, the two main scenarios that affect seedling emergence include adequate seed-zone water potential and the occurrence of surface soil crust that prevents penetration (Schillinger, 2011). Successful seedling emergence is dependent on the force exerted by the first leaf. The first leaf protrudes through the coleoptile and emerges around 10–12 days after planting. During this time, the first leaf can be prone to buckling before it emerges, which can be affected by coleoptile diameter, speed of emergence, emergence force, and lifting capacity of the first leaf, along with the associated coleoptile length (Arndt, 1965; Schillinger et al., 2017; Lutcher et al., 2019). Adequate seed zone water potential is associated with seed germination and impacts the speed of emergence because of water availability (Evans and Etherington, 1990; Pill, 1995). These studies showed that the genetic basis of seedling emergence is a complex trait not solely controlled by genes for any one factor and results in a poorly understood genetic architecture that is dependent on the environment to display variation (Schillinger et al., 2017; Lutcher et al., 2019). This study presents research to assess the genetic architecture of a complex trait by (1) comparing ST-GWAS and MT-GWAS models for correlated traits and (2) assessing the inclusion of fixed effects to improve the power to explore the genetic architecture of seedling emergence.

MATERIALS AND METHODS

Phenotypic Data

Seedling emergence notes were taken on research plots under low annual precipitation (~150 mm annual precipitation) at the Washington State University Dryland Research Center in Lind, WA (47.001552, -118.565556). The plots were planted using a custom-built deep-furrow planting system to a depth of between 10 and 15 cm, depending on moisture variation among years. The plots were planted 1.5 m wide and 6.1 m long with 31 cm row spacing at a density of 120 seed per square meter. Emergence notes were taken on two populations within the breeding program. The diverse association mapping panel (DP) represents a diverse panel of inbred breeding lines (BLs) from Pacific Northwest breeding programs not selected exclusively in deep-furrow trials. In contrast, the second population is composed of F_{3:5} BLs and represents a population of closely related lines from a single breeding program composed of pedigrees that have been

selected for emergence over previous generations and is a part of the WSU breeding program. The two populations were used to compare GWAS models. The DP was used as the primary population for genetic dissection, and the BL population used as the validating population. The DP was evaluated in 2015, 2017, 2018, and 2019 in Lind, WA (Table 1). The BLs were planted using an unreplicated augmented design that was evaluated in 2015 for emergence (Table 1). In 2016, no data were collected for the DP because of significant soil crusting that was severe enough to impede the seedling emergence of all lines.

Seedling emergence was visually assessed and recorded as a percentage of the total plot that emerged 6 weeks after planting for each trial. Table 1 summarizes location, population, year, and the number of genotyped individuals. The emergence issue for each trial was attributed to moisture stress. Coleoptile length was also measured for the DP in 2014 and in two replicates in 2016 under greenhouse conditions. Coleoptile length was recorded to the nearest millimeter according to Murphy et al. (2008).

Phenotypic Adjustments

Adjusted means from the emergence data collected in the unreplicated trials were adjusted using residuals calculated for the unreplicated lines in individual environments and across environments using the modified augmented complete block design (ACBD) model (Federer, 1956; Goldman, 2019). The adjustments were made following the method implemented in Merrick and Carter (2021), with the full model in a single model as follows:

$$Y_{ij} = \mu + \text{Block}_i + \text{Check}_j + \varepsilon_{ij}, \quad (1)$$

where Y_{ij} is the phenotypic value for the trait of interest of the i -th block and j -th check ($i = 1, \dots, I, j = 1, \dots, J$); μ is the mean effect; Block_i is the fixed effect of the i -th block; Check_j is the fixed effect of the j -th replicated check cultivar; and ε_{ij} are residual errors with a random normal distribution of $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. For adjusted means across environments, the model is as follows:

$$Y_{ijk} = \mu + \text{Block}_i + \text{Check}_j + \text{Env}_k + \text{Block}_i : \text{Env}_k + \text{Check}_j : \text{Env}_k + \varepsilon_{ijk} \quad (2)$$

where Y_{ijk} is the phenotypic value for the trait of interest of the i -th block and j -th check in the k -th environment ($i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, K$); μ is the mean effect; Block_i is the fixed effect of the i -th block; Check_j is the fixed effect of the j -th

TABLE 1 | Populations for seedling emergence screened in unreplicated trials under moisture stress in Lind, WA from 2015 to 2019.

Location	Trial	Year	Individuals
Lind	DP*	2015	473
Lind	DP	2017	473
Lind	DP	2018	473
Lind	DP	2019	473
Lind	F _{3:5}	2015	276

*DP: quality association mapping diversity panel.

replicated check cultivar; Env_k is the fixed effect of the k -th environment; and ϵ_{ijk} are residual errors with a random normal distribution of $\epsilon \sim N(0, \sigma_\epsilon^2)$.

Best linear unbiased predictors (BLUPs) for heritability were calculated for each trial and across trials using a mixed linear model for the full augmented randomized complete block design in a single environment and is as follows:

$$Y_{ijk} = \mu + \mathbf{Block}_i + \mathbf{Check}_j + \mathbf{Gen}_{l(j)} + \epsilon_{ijk} \quad (3)$$

where Y_{ij} is the phenotypic value for the trait of interest of the l -th genotype nested in the j -th check in the i -th block ($i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, K$); μ is the mean effect; \mathbf{Block}_i is the random effect of the i -th block with the distribution $\mathbf{Block} \sim N(0, \sigma_{Block}^2)$; \mathbf{Check}_j is the fixed effect of the j -th replicated check cultivar; $\mathbf{Gen}_{k(j)}$ is the genotype k in the j -th check with the distribution $\mathbf{Gen} \sim N(0, \sigma_{Gen}^2)$; and ϵ_{ij} are residual errors with a random normal distribution of $\epsilon \sim N(0, \sigma_\epsilon^2)$. For the full model (3) across environments is as follows:

$$Y_{ijkl} = \mu + \mathbf{Block}_i + \mathbf{Check}_j + \mathbf{Gen}_{l(j)} + \mathbf{Env}_k + \mathbf{Block}_i : \mathbf{Env}_k + \mathbf{Check}_j : \mathbf{Env}_k + \mathbf{Gen}_{l(j)} : \mathbf{Env}_k + \epsilon_{ijkl} \quad (4)$$

where Y_{ijkl} is the phenotypic value for the trait of interest of the l -th genotype nested in the j -th Check of the i -th block and k -th environment ($i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, K, l = 1, \dots, L$); μ is the mean effect; \mathbf{Block}_i is the random effect of the i -th block with the distribution $\mathbf{Block} \sim N(0, \sigma_{Block}^2)$; \mathbf{Check}_j is the fixed effect of the j -th replicated check cultivar; $\mathbf{Gen}_{l(j)}$ is the random effect of the genotype l in the j -th check with the distribution $\mathbf{Gen} \sim N(0, \sigma_{Gen}^2)$; \mathbf{Env}_k is the random effect of the k -th environment with the distribution $\mathbf{Env} \sim N(0, \sigma_{Env}^2)$; and ϵ_{ijkl} are residual errors with a random normal distribution of $\epsilon \sim N(0, \sigma_\epsilon^2)$. Heritability on a genotype-difference basis for broad-sense heritability was calculated using the variance components from models 3 and 4 implemented in Merrick and Carter (2021) and using BLUPs for both individual environments and across environments using the formula from Cullis et al. (2006):

$$H_{Cullis}^2 = 1 - \frac{\overline{V}_{\Delta..}^{BLUP}}{2\sigma_g^2} \quad (5)$$

where σ_g^2 and $\overline{V}_{\Delta..}^{BLUP}$ are the genotype variance and mean-variance, respectively, of a difference between two BLUPs for the genotypic effect BLUPs (Schmidt et al., 2019). Trial evaluation and significant differences were evaluated using the coefficient of variation, and by analysis of variance (ANOVA) in individual and across trials using BLUP models 3 and 4. BLUPs for coleoptile length were calculated across trials using a mixed linear model as follows:

$$Y_{klm} = \mathbf{Gen}_l + \mathbf{Env}_k + \mathbf{Rep}_{m(k)} + \mathbf{Gen}_l : \mathbf{Env}_k + \epsilon_{klm} \quad (6)$$

where Y_{klm} is the trait of interest of the l -th genotype in the k -th environment of the m -th replication ($k = 1, \dots, K, l = 1, \dots, L, m = 1, \dots, M$); $\mathbf{Gen}_{l(j)}$ is the random effect of the l -th genotype with $\mathbf{Gen} \sim N(0, \sigma_{Gen}^2)$; \mathbf{Env}_k is the random effect of the k -th environment with $\mathbf{Env} \sim N(0, \sigma_{Env}^2)$; $\mathbf{Rep}_{m(k)}$ is the random

effect of the replication m in the k -th environment; and ϵ_{ijk} are residual errors with a random normal of distribution of $\epsilon \sim N(0, \sigma_\epsilon^2)$.

Adjusted means for coleoptile length were calculated across trials using a linear model as follows:

$$Y_{lm} = \mathbf{Env}_k + \mathbf{Rep}_{m(k)} + \epsilon_{lm} \quad (7)$$

where Y_{lm} is the phenotypic value for the trait of interest of the k -th environment of the m -th replication ($l = 1, \dots, L, m = 1, \dots, M$); \mathbf{Env}_k is the fixed effect of the k -th environment; $\mathbf{Rep}_{m(k)}$ is the fixed effect of the replication m in the k -th environment; and ϵ_{ijk} are residual errors with a random normal distribution of $\epsilon \sim N(0, \sigma_\epsilon^2)$. Means were adjusted for coleoptile length following the method for models 1 and 2.

Phenotypic correlations were conducted between seedling emergence in the DP across years along with coleoptile length. Due to the unreplicated nature of the seedling emergence phenotypes, genetic correlations between seedling emergence in the DP and coleoptile length was calculated using the R package "sommer" with the multivariate Newton-Raphson algorithm used for multiple random effects and covariance structures using the multivariate model in Covarrubias-Pazarán (2018):

$$\begin{bmatrix} Y_A \\ Y_B \end{bmatrix} = \begin{bmatrix} X_A & \mathbf{0} \\ \mathbf{0} & X_B \end{bmatrix} \begin{bmatrix} \beta_A \\ \beta_B \end{bmatrix} + \begin{bmatrix} Z_A & \mathbf{0} \\ \mathbf{0} & Z_B \end{bmatrix} \begin{bmatrix} u_A \\ u_B \end{bmatrix} + \begin{bmatrix} \epsilon_A \\ \epsilon_B \end{bmatrix} \quad (8)$$

where Y_A and Y_B are vectors ($n \times 1$; n = number of lines) of trait-adjusted means for emergence using models 1 and 2, and adjusted means for coleoptile length using model 7, respectively. β_i is a vector [$p \times 1$; p = three principal components (PCs)] of fixed effects of PCs, u_i is a vector ($1 \times m$; m = number of markers) of random effects for individuals with $u \sim N(0, \sigma_u^2)$, and ϵ_i is a vector ($n \times 1$; n = number of lines) of residuals with $\epsilon \sim N(0, \sigma_\epsilon^2)$ for each trait ($I = A..B$). X and Z are incidence matrices for fixed effects ($n \times p$; n = number of lines, p = three PCs) and random genetic effects ($n \times m$; n = number of lines, m = number of markers), respectively, for each trait. The distribution of the multivariate response and phenotypic variance-covariance V following the models in Covarrubias-Pazarán (2018) are:

$$Y = X\beta + Zu + \epsilon$$

$$Y \sim MVN(X\beta, V)$$

$$Y = \begin{bmatrix} Y_A \\ Y_B \end{bmatrix}$$

$$X = \begin{bmatrix} X_A & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & X_B \end{bmatrix}$$

$$V = \begin{bmatrix} Z_A K \sigma_{u_A}^2 Z_A' + I \sigma_{\epsilon_A}^2 & \cdots & Z_A K \sigma_{u_{A,B}} Z_B' + I \sigma_{\epsilon_{A,B}}^2 \\ \vdots & \ddots & \vdots \\ Z_A K \sigma_{u_{A,B}} Z_B' + I \sigma_{\epsilon_{A,B}}^2 & \cdots & Z_B K \sigma_{u_B}^2 Z_B' + I \sigma_{\epsilon_B}^2 \end{bmatrix} \quad (9)$$

where \mathbf{K} is the additive genetic relationship matrix ($n \times n$; the number of lines) calculated by $K = \frac{WW'}{2 \sum_k p_k q_k}$ using the $n \times m$ -centered genotype matrix W , with p and q being allele one and two for the k -th genotype and was implemented using the “a.mat” function in R for the k -th random effect ($u = 1, \dots, k$); and \mathbf{I} is an identity matrix ($n \times n$; n = number of lines) for the residual term. The terms $\sigma_{uk_i}^2$ and $\sigma_{\epsilon_i}^2$ denote the genetic and residual variance of the trait i , respectively, and $\sigma_{uk_{AB}}$ and $\sigma_{\epsilon_{AB}}$ are the genetic and residual covariance between traits A and B. Genetic correlations (r_G) between seedling emergence and coleoptile length were then calculated using the function “cov2cor” in the R package sommer (Covarrubias-Pazarán, 2018; R Core Team, 2018).

$$r_G = \frac{\sigma_{uk_{AB}}}{\sqrt{\sigma_{uk_A}^2 * \sigma_{uk_B}^2}} \quad (10)$$

Genotypic Data

Lines were genotyped by using genotyping-by-sequencing (GBS; Elshire et al., 2011) through the North Carolina State Genomics Sciences Laboratory in Raleigh, NC, United States, using the restriction enzymes *MspI* and *PstI* (Poland et al., 2012). Genomic DNA was isolated from seedlings in the one- to three-leaf stage using Qiagen BioSprint 96 Plant kits and the Qiagen BioSprint 96 workstation (Qiagen, Germantown, MD, United States). DNA libraries were prepared following the protocol of DNA digestion with *PstI* and *MspI* restriction enzymes (Poland et al., 2012). Genotyping by sequencing (GBS; Elshire et al., 2011) was conducted at North Carolina State University Genomic Sciences Laboratory with either an Illumina HiSeq 2500 or a NovaSeq 6000. DNA library barcode adapters, DNA library analysis, and sequence single-nucleotide polymorphism (SNP) calling were provided by the USDA Eastern Regional Small Grains Genotyping Laboratory (Raleigh, NC, United States). Sequences were aligned to the Chinese Spring International Wheat Genome Sequencing Consortium (IWGSC) RefSeq v1.0 (Appels et al., 2018) using the Burrows-Wheeler Aligner (BWA) 0.7.17 (Li and Durbin, 2009). GBS SNP markers were called using the TASSEL-GBS v2 SNP calling pipeline in Tassel v5 (Bradbury et al., 2007; Glaubitz et al., 2014). Since all the lines in both populations were considered inbred lines, genetic markers with heterozygote calls over 10% were filtered out. Genetic markers common across both populations were combined for quality control, and markers with more than 20% missing data, minor allele frequency of less than 5%, and those that were monomorphic were removed. The markers were then imputed using Beagle version 5.0 and filtered once more for markers under a 5% minor allele frequency (Browning et al., 2018). A total of 40,368 SNP markers remained.

All winter wheat lines in the DP were genotyped with Kompetitive Allele Specific PCR (KASP®) assays in the WSU Winter Wheat Breeding Laboratory using allele-specific SNP markers for semi-dwarf causing mutant alleles *Rht-B1b* and *Rht-D1b* previously reported in Grogan et al. (2016) and Rasheed et al. (2016). The KASP assays were performed using PACE™ Genotyping Master Mix (3CR Bioscience, Harlow, United Kingdom) following the instructions of the manufacturer, and endpoint genotyping was conducted from fluorescence

using a Lightcycler 480 Instrument II (Roche, Indianapolis, IN, United States). The *Rht* markers were coded as lines with *Rht-B1b* (1), *Rht-D1b* (2), *Rht-B1b* heterozygous (3), *Rht-D1b* heterozygous (4), *Rht-D1b* with a heterozygous *Rht-B1b* (5), *Rht-B1b* with a heterozygous *Rht-D1b* (6), and both *Rht-B1b* and *Rht-D1b* heterozygous (7).

Linkage disequilibrium between marker pairs was evaluated using JMP Genomics v.9.0 (SAS Institute, Inc, 2011). Significant marker pairs in the same chromosome were considered in LD at a p -value < 0.05. Population structure within both populations was analyzed using PC analysis biplots and k -means clustering using the markers in the DP and BL populations individually using the function “prcomp” and “cluster” in R, respectively (R Core Team, 2018).

Genome-Wide Association Models

To dissect the genetic architecture of a complex trait (seedling emergence), the ST-GWAS models were implemented using the Genome Association and Prediction Integrated Tool (GAPIT; Liu et al., 2016; Tang et al., 2016; Huang et al., 2019). Both the ST-GWAS and MT-GWAS models were implemented with three PCs fitted as fixed effects. Three PCs were used based on BIC values using model selection in GAPIT (Tang et al., 2016). The GWAS models were conducted on seedling emergence using the adjusted means mentioned previously and on the BLUPs for coleoptile length. The DP was used as the primary population for genetic dissection, and the BL population was used as the validating population. Three ST-GWAS models were used for comparison. The single-locus ST-GWAS model used was an MLM, and the multi-locus models were BLINK and FarmCPU. Within the DP, we compared each model within each year combination without covariates, then with the *Rht* markers as covariates, coleoptile length BLUPs, and both *Rht* markers and coleoptile length as covariates. This resulted in 28 datasets for ST-GWAS for seedling emergence in the DP. The GWAS models were then conducted within the BL without covariates to validate the significant markers. In addition, the ST-GWAS models were used to dissect coleoptile length within the DP for further validation to determine whether the significant markers affected coleoptile length.

Additionally, MTMM was implemented using the “sommer” package for MT-GWAS to identify pleiotropic interactions between seedling emergence and coleoptile length within the DP using the multivariate models 8 and 9. The MT-GWAS model was then implemented in the package “sommer” to obtain marker effects using the inverse of the phenotypic variance matrix (V) and is a generalized linear model of the form:

$$\mathbf{b} = (\mathbf{X}'\mathbf{V} - \mathbf{X})\mathbf{X}'\mathbf{V} - \mathbf{y} \quad (11)$$

$$\text{with } \mathbf{X} = \mathbf{Z}\mathbf{M}_i \quad (12)$$

where \mathbf{b} is the marker effect (1×2), \mathbf{y} is the multivariate response variable (1×2), \mathbf{V} is the inverse of the phenotypic variance matrix (2×2), \mathbf{Z} is the incidence matrix for the random effect to perform the GWAS, and \mathbf{M}_i is the i -th column of the marker

matrix. Furthermore, we implemented three additional *F*-tests for joint analysis using the scripts proposed in Korte et al. (2012). The full (FULL) model, which includes the effect of the marker genotype and its interaction, was tested against a null model and identified both loci with common and interaction effects. The interaction model (IE) for identifying the interaction effects between the traits tested the full model against a genetic model. Finally, we identified common (COM) genetic effects and tested the genetic model against a null model.

Significant associations were based on a Benjamini–Hochberg false discovery rate (FDR) threshold (Benjamini and Hochberg, 1995). The phenotypic variation and effect explained by each significant marker were calculated by conducting stepwise linear regression between phenotypic and genotypic data and calculating the difference between the effects and variation when a single significant marker was added to the null model in R (R Core Team, 2018; Lozada et al., 2019). Significant markers were deemed consistent when they were significant in at least 2 years or in both populations and were used to display reliability of the effect and significance of the markers. We then tested the additive effect of pyramiding the consistent markers identified across populations and the consistent markers identified in multiple years from the DP in each population individually. Manhattan plots were created using the “ggplot2” package, and QQ plots were plotted using the package “CMplots” in R (R Core Team, 2018; Yin et al., 2021).

RESULTS

Phenotypic Data

Heritability for seedling emergence was moderately high in the BLs and a few years in the DP, whereas the heritability decreased in combined years. The highest heritability in a single trial was 0.88 in the DP in 2018, and the BLs had a heritability of 0.77 (Table 2). The trials with larger negative adjusted mean values have a larger SD, which indicates a wider range of seedling emergence and increased environmental pressure and phenotypic variation for selection purposes (Supplementary Figure 1). Heritability for coleoptile length was very high in the DP (0.89) (Supplementary Table 1).

Since the variation for seedling emergence depends on environmental effects, it is important to examine the variance components of trials. ANOVAs were conducted using models 3 and 4. Genetic variances were significant for the DP in 2015, 2017, and for all the combined trials (Supplementary Table 2). The environmental effect was not significant in any of the combined trials but had a very large variance for the combined trials. However, for the nested genotype to environment and nested block to environment had large significant variances displaying significant genotype-by-environment interaction (GEI) over the combined trials of 2015–2018 and 2015–2019 (Supplementary Table 2). Furthermore, phenotypic correlations allow for us to compare the results in our GWAS models. The DP trials are significantly positively correlated to each other except for three scenarios: DP 2015 to DP 2018, DP 2017 to DP 2018, and DP 2017 to DP 2019 (Supplementary Table 3). The BL $F_{3;5}$ trial

in 2015 was significantly correlated to DP 2017. The genetic correlations between the DP seedling emergence to coleoptile length showed moderate to large correlations in and across years (Table 3). The highest genetic correlation was found in DP 2019, with a correlation of 0.66. However, this was not the case for the phenotypic correlations. The phenotypic correlations in all years were near zero between seedling emergence and coleoptile length (Table 3 and Figure 1).

Genotypic Data

The PC biplot using the SNP markers for the DP displayed four clusters according to the elbow method with much overlap between clusters 1 and 2, and a slight overlap between clusters 3 and 4 (Figure 2A). PC1 explained 12.9% of the variation, and PC2 explained 6.9% of the variation. However, most of the lines clustered within a single cluster in the BL population even though the elbow method revealed four clusters using *k*-means, and the biplot explained less variation with 9 and 5.2% for PC1 and PC2, respectively (Figure 2B). We can visually see the larger genetic variation in the DP than in the BLs; therefore, it is important to be accounted for in our GWAS models.

The frequency of the *Rht* alleles in the DP population can be seen in Supplementary Table 3. The majority of the lines in the DP had either *Rht-B1b* (0.564) or *Rht-D1b* (0.347), with *Rht-B1b* conveying higher mean seedling emergence than *Rht-D1b* (Figure 3). Furthermore, Figure 3 displays outliers for seedling emergence for lines with certain *Rht* alleles. The outliers were not removed because the cause of poor seedling emergence was indeterminate because of the complexity between phenotypic variation and genotypic effect.

Single-Trait Genome-Wide Association Studies

To investigate the pleiotropic effects of significant markers, the ST-GWAS was conducted on coleoptile length within the DP. Genome-wide association for coleoptile length displayed three unique markers using MLM, BLINK, and FarmCPU (Table 4). BLINK and MLM both identified a marker, *S1A_14084576*, on chromosome 1A. Marker *S1A_14084576* conveyed the largest R^2 value of any significant marker with a value of 7% and an effect of 14.5 mm. The marker with the next largest R^2 of 0.05 was *S6A_543395015*. It was identified by both BLINK and FarmCPU on chromosome 6A, and it also conveyed the largest effect with 34.4 mm. A third marker was identified only by FarmCPU, *S2A_765677087*, and was located on chromosome 2A. However, *S2A_765677087* had a rather small effect and R^2 of 3.2 mm and 0.04, respectively. The three markers significantly associated with coleoptile length were not identified in any population, year, or model for seedling emergence (Supplementary File 1).

There were 107 unique markers over all the combinations of ST-GWAS for seedling emergence (Supplementary File 1). Of the 107 significant markers, 96 markers were significant in the DP and 15 were significant in the BLs, with four markers significant in both populations. Seventy-five of the markers were significant in at least two combinations over both populations. Seventy-one and five markers were significant in at least two combinations in

TABLE 2 | Cullis heritability and trial statistics for adjusted means for deep-sowing seedling emergence in individual and combined trials for the diversity panel (DP) population and breeding line (BL) population phenotyped from 2015 to 2019 and 2015, respectively.

Population	Trial	Year	Heritability	CV ^a	Max ^b (%)	Mean (%)	Min ^c (%)	SD ^d
DP	DP	2015	0.75	78.91	139	48	-43	38
DP	DP	2017	0.70	24.18	118	87	-26	21
DP	DP	2018	0.88	45.46	118	62	-43	28
DP	DP	2019	0.68	106	135	43	-45	45
DP	DP	2015–2017	0.63	34.53	126	68	-17	23
DP	DP	2015–2018	0.61	28	110	66	2	18
DP	DP	2015–2019	0.64	31.58	101	60	1	19
BL	F _{3,5}	2015	0.77	53.55	121	61	-59	33

^aCV: coefficient of variation.

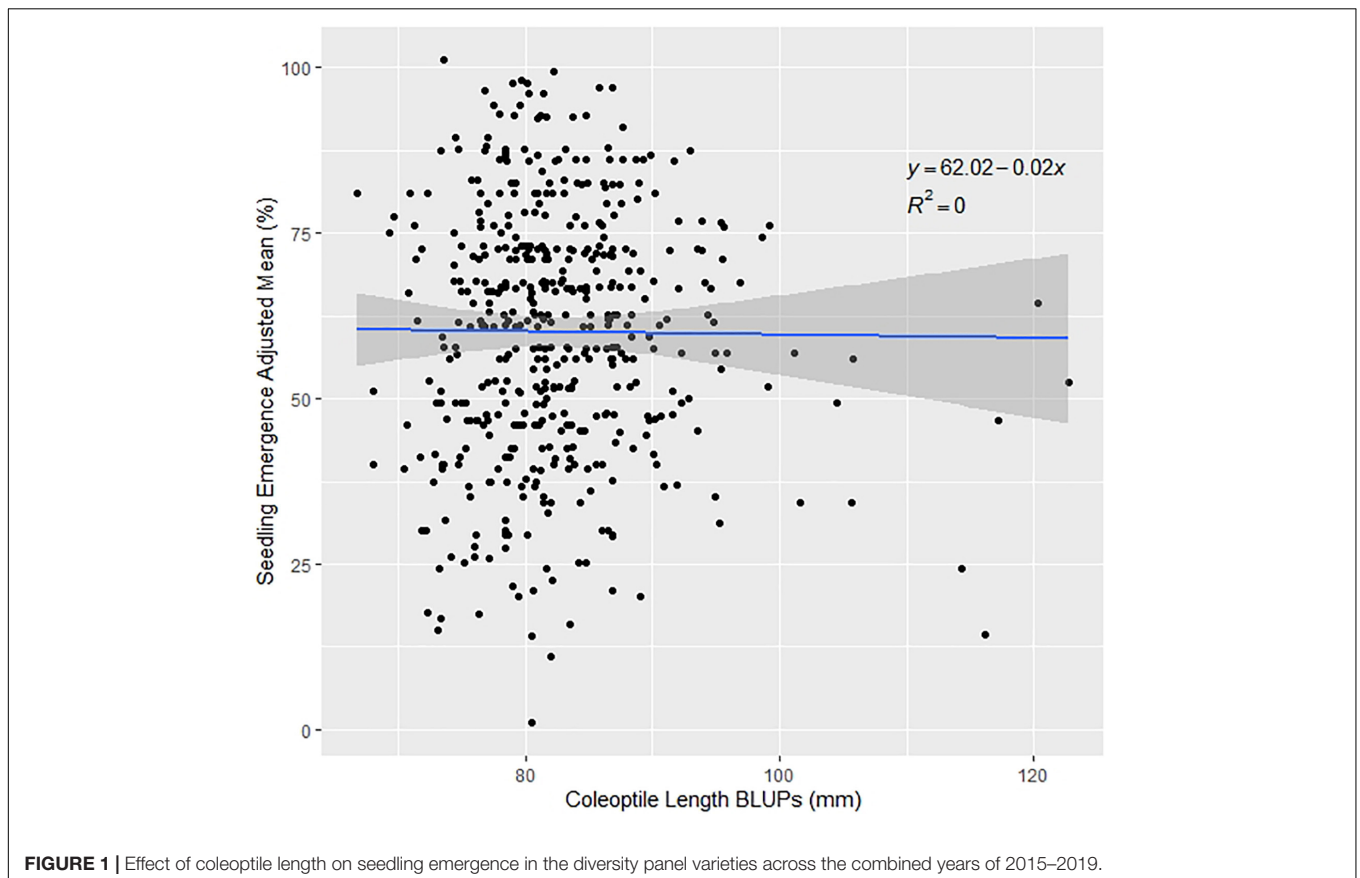
^bMax: maximum.

^cMin: minimum.

^dSD: standard deviation.

TABLE 3 | Genetic and phenotypic correlations between adjusted means for coleoptile length and deep-sowing seedling emergence in individual and combined trials for the DP population phenotyped from 2015 to 2019.

Year	DP 2015	DP 2017	DP 2018	DP 2019	DP 2015–2017	DP 2015–2018	DP 2015–2019
Genetic correlations (R^2)	0.42	-0.40	0.38	0.66	0.19	0.32	0.52
Phenotypic correlations (R^2)	0.00	-0.01	-0.06	0.03	0.00	-0.04	-0.01



the DP and BLs, respectively. Significant markers were found on 19 of the 21 chromosomes in the DP, with the majority (34) of the significant markers located on chromosome 5A. In the BLs, 15

unique markers spanned nine chromosomes. All the three models had more significant markers in the DP population than in the BL population. The MLM identified 36 unique markers over all



GWAS, three in the BLs, and 23 in the DP. FarmCPU identified 65 unique markers overall, nine in the BLs, and 57 in the DP. BLINK identified 31 unique significant markers overall, eight in the BL, and 23 in the DP. The MLM identified significant markers mainly on chromosome 5A with many neighboring significant markers. In contrast, FarmCPU and BLINK identified the same markers on chromosome 5A and identified more markers on various other chromosomes (**Supplementary File 1**).

Additionally, there was a differential effect of BLINK and FarmCPU to detect significant markers by including covariates. First, we compared the number of significant markers identified within the DP (**Figure 4**). For FarmCPU, including both coleoptile length and Rht alleles individually and in combination decreased the number of significant markers compared to FarmCPU without covariates. However, for BLINK, Rht alleles, and the combination of Rht alleles and coleoptile length increased the number of significant markers. Coleoptile length alone decreased the number of significant markers compared to BLINK without covariates. Conversely, the effect of covariates was limited in the MLM. For example, including coleoptile length and Rht alleles as covariates alone resulted in a similar number of significant markers (64 and 63) compared to the MLM without covariates. Including both Rht alleles and coleoptile length did reduce the number of significant markers to 55. Furthermore, a large number of markers in the LD increased the number of significant markers identified by the MLM compared to the other ST-GWAS models.

Second, we compared the effect of including covariates by examining QQ plots. The markers should follow the quantile line in the QQ plot with the exception of a few markers toward the end of the line. A large deviation for a few markers, with the remaining markers on the quantile line, can indicate the power to identify significant markers and fewer false positives. However, a large number of markers deviating from the line can indicate an increase in false positives. For example, in the combined analysis of 2015–2018 and in the BLs in 2015, the MLM was consistent regardless of covariates, which show little effect on the ability to identify significant markers (**Supplementary Figure 2**). However, this was not the case with BLINK or FarmCPU. Within the DP, the inclusion of Rht or Rht and coleoptile length increased the deviation from the quantile line for the markers at the end of the quantile line more than the models without covariates (**Supplementary Figures 3, 4**). This displays the advantage of using covariates, such as Rht alleles, within the DP for an increase in detection of significant markers when the trait in question is genetically correlated with another trait. In addition, the QQ plots reveal the difference between the models in the different populations. For both BLINK and FarmCPU, the BL population had the largest deviations compared to the DP GWAS, whereas the opposite was seen in the MLM.

In addition, the environmental impact on seedling emergence had a large effect on the power to identify significant markers. The varying number of significant markers showed the GEI for seedling emergence in different years for the DP. The GWAS

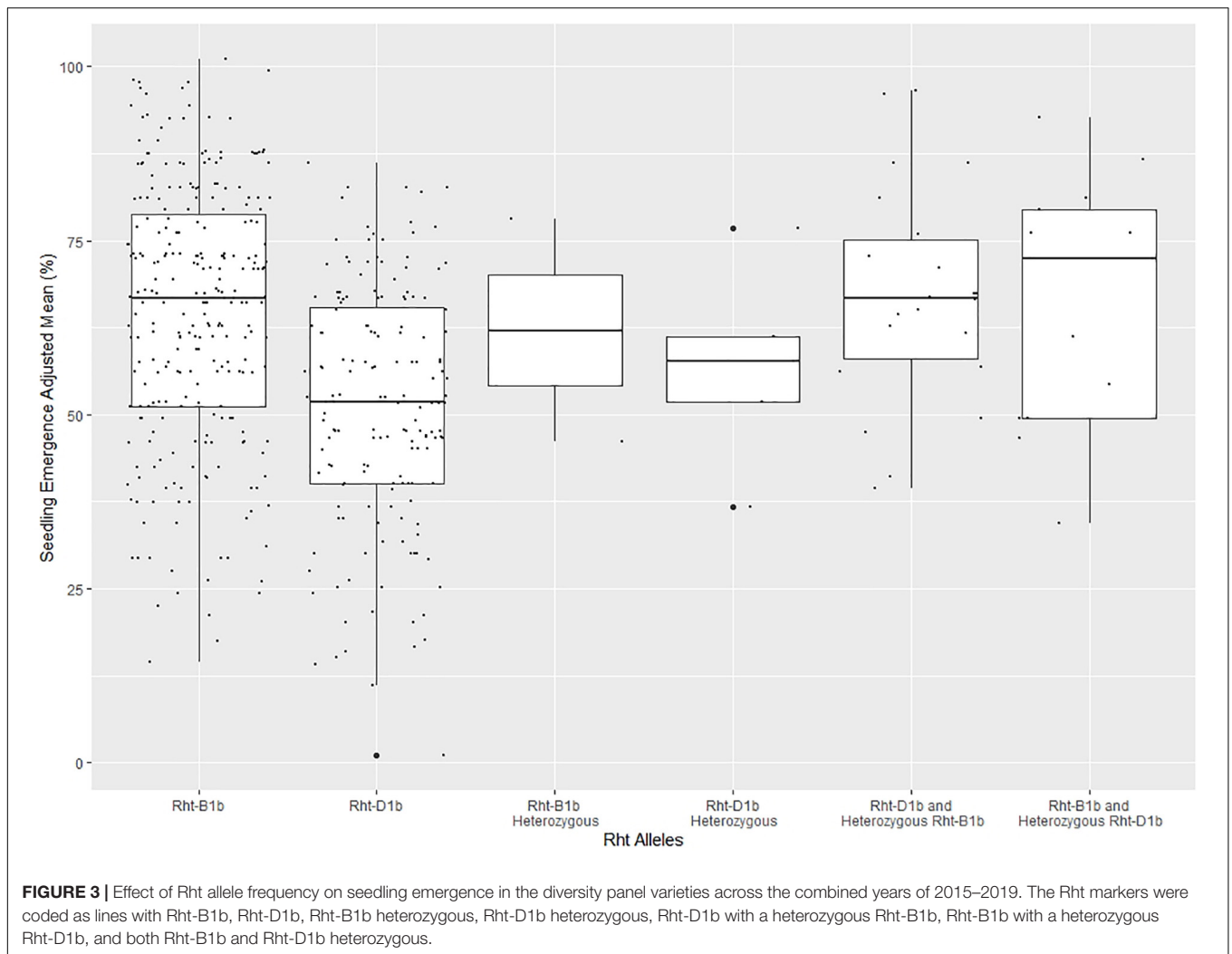


TABLE 4 | Significant markers for coleoptile length in a Pacific Northwest winter wheat diversity panel using Bayesian information and linkage-disequilibrium iteratively nested keyway (BLINK), fixed and random model circulating probability unification (FarmCPU), and mixed linear (ML) genome-wide association study (GWAS) models.

Marker	Position ^a	Alleles ^b	Chr ^a	Model	p-Value	MAF ^c	Effect (mm) ^d	R ^{2e}
S1A_14084576	14084576	<u>T</u> /A	1A	BLINK	1.82E−11	0.01	14.52	0.07
				MLM	4.78E−08	0.01	14.52	0.07
S2A_765677087	7.66E+08	G/ <u>T</u>	2A	FarmCPU	4.22E−09	0.14	3.23	0.04
S6A_543395015	5.43E+08	<u>T</u> /G	6A	BLINK	4.70E−08	0.00	34.44	0.05
				FarmCPU	6.32E−14	0.00	34.44	0.05

^aChromosomes and positions of markers were determined according to IWGSC RefSeq v.1.1.

^bAllele: favorable allele is underlined.

^cMAF: minor allele frequency.

^dEffect: increase of seedling emergence (%) with the inclusion of favorable allele of associated marker.

^eR²: phenotypic variance explained by associated marker.

models were able to identify the most significant markers in the 2015 trial, with 132 markers (**Supplementary Figure 5**). However, only a few significant markers were identified in the other individual years, with two in 2017 and five in 2019. Furthermore, combining years and accounting for GEI in our phenotypic adjustments allowed for the GWAS models to increase the ability and power to dissect seedling emergence

compared to individual years. All 3-year combinations increased the number of significant markers consistently compared to individual years, with 109 in year combinations 2015–2017 and 105 in 2015–2018. Additionally, there was a decrease in significant markers identified with an increase in year combinations with the combination of all years (2015–2019), identifying only 47 markers (**Supplementary Figure 5**).

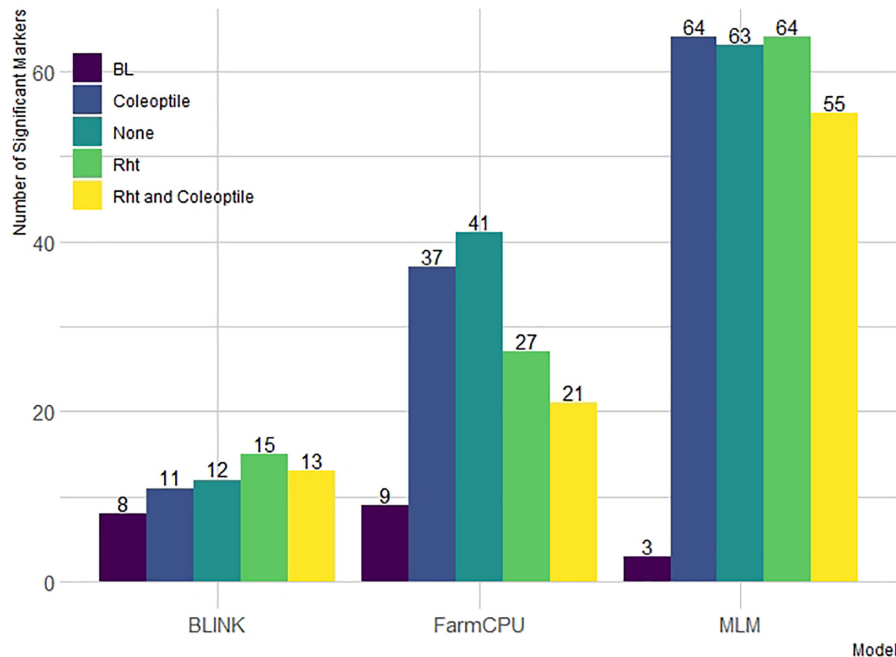


FIGURE 4 | Models compared are Bayesian information and linkage-disequilibrium iteratively nested keyway (BLINK), fixed and random model circulating probability unification (FarmCPU), and mixed-linear model (MLM) with and without Rht and coleoptile length as covariates in the DP and without covariates in the breeding line (BL).

Multi-Trait Genome-Wide Association Studies

The MT-GWAS models were used to identify pleiotropic loci between seedling emergence and coleoptile length. The MT-GWAS models displayed very similar results for the individual traits compared to the ST-GWAS models, especially in comparison to the single trait results from the MT-GWAS model. However, we used the Bonferroni cut-off with an $\alpha = 0.05$ instead of FDR because of the large deviations and inflation of p -values seen on the QQ plots for the joint models (Supplementary Figure 6). Using FDR resulted in 924 unique markers across the majority of chromosomes. In comparison, the Bonferroni cut-off resulted in 82 unique significant markers across 14 chromosomes with the majority of the large effect alleles on chromosome 5A (Supplementary File 2). In comparing the results for MT-GWAS FULL, IE, and COM F -tests to the single trait results from the MT-GWAS, we see no significant markers for coleoptile length for any other model (Figure 5). The significant markers for coleoptile length were the same two significant markers on chromosomes 1A and 6A identified in the ST-GWAS models. Additionally, the significant markers for seedling emergence were located on chromosomes 2B (1) and 5A (15).

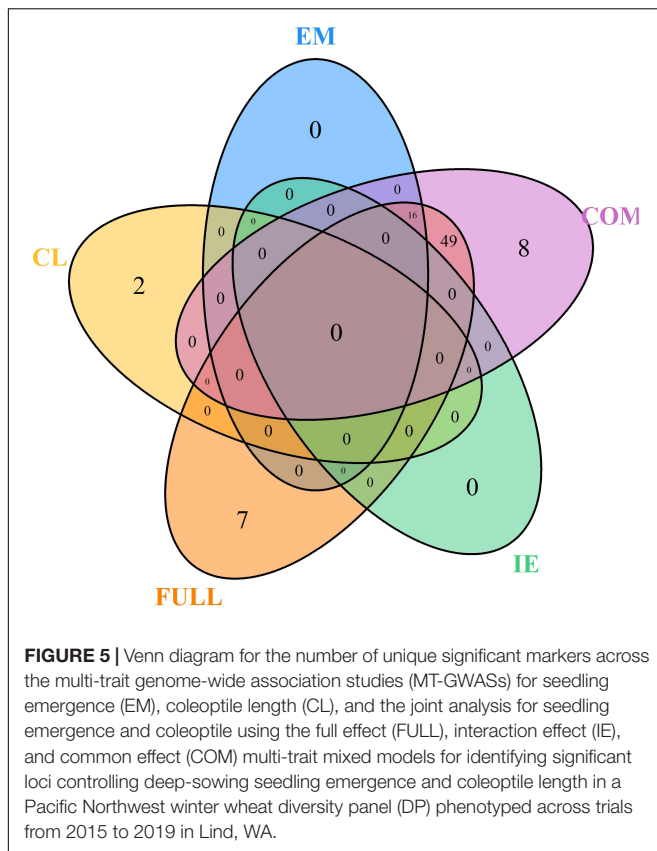
Furthermore, there were no significant markers in the IE, indicating no contradictory effect markers between seedling emergence and coleoptile length. The lack of significant IE markers is why the FULL and COM models are very similar. The COM model also identified significant markers in every year in the DP compared to the single trait results for seedling

emergence, which did not identify significant markers in the 2017 through 2019 individual years (Supplementary File 2). Additionally, there were 64 significant markers in the COM model, with 16 significant markers also found significant for seedling emergence (Figure 5). These markers indicate the potential for confounding effects and the ability to identify them using a joint analysis model compared to the single trait results implemented in the MT-GWAS models.

Consistent Significant Markers

Because of the lack of identified loci for seedling emergence, the identified markers were compared for significance across years, populations, and correlated traits to insure consistency. No markers were significantly associated for both seedling emergence and coleoptile length in either ST-GWAS or MT-GWAS. Using ST-GWAS models within the DP, 23 out of 107 unique markers were found significant in more than 1 year. BLINK only identified one marker, *S5A_522153944*, across years. FarmCPU identified six markers across years on chromosomes 2A (2), 2B (1), and 5A (3). The MLM identified 16 markers across years, with all but one on chromosome 5A, and with the other marker located on chromosome 5B. For these markers in the DP, only *S5A_522153944* was identified by all three models.

For the number of significant markers across chromosomes, QQ plots, covariate effect, and populations, FarmCPU displayed the ability to identify both the consistent large effect and small effect markers with fewer false positives because of LD on chromosome 5A (Figure 6). In comparison, the MLM consistently identified 13 markers on 5A with R^2 LD values



above 0.8. The two largest effect consistent markers in the DP, *S5A_522153944*, and *S5A_522153953*, were found on chromosome 5A and improved seedling emergence by 30%. These two markers had a maximum R^2 of 10% in the DP. Other consistent markers identified across years in the DP were located on chromosomes 2A (2), 2B (3), 5A (15), 5B (1), and 7A (1). Overall, the consistent markers across years accounted for 30% of the combined 2015–2019 DP variation. The GWAS on the BL population displayed more consistent results than in the DP, and only identified consistent markers on chromosome 5A. There were four significant consistent markers identified in the BLs (Table 5). The four consistent markers were all found in the DP, and had effect ranges of 17–30%, with R^2 values of 12–13%. The consistent markers in the BL population accounted for 18% of the total variation for seedling emergence, displaying more consistent dissection of the complex trait compared to the DP.

Out of the 107 unique significant markers, only four were identified in both populations. Furthermore, only *S5A_522153944*, *S5A_522153953*, and *S5A_523025549* were identified across both populations with the same model (Table 5). The MLM identified all three of these markers in both populations. Whereas BLINK did not identify any markers in both populations, FarmCPU only identified *S5A_522153944*. The effect of covariates on identifying the consistent markers was inconsistent in the DP. The MLM displayed no response to the inclusion of covariates (Supplementary Figure 7). However, FarmCPU identified the consistent marker *S5A_522153944* on

chromosome 5A using the Rht alleles, coleoptile length, and Rht alleles with coleoptile length in combination as covariates (Supplementary Figure 8). Furthermore, BLINK was able to detect *S5A_522153944* with and without covariates but was not able to identify it in the BL population (Supplementary Figure 9).

The MT-GWAS models displayed very similar results for identifying consistent markers across years for both the single trait and joint models as the ST-GWAS models. Using single trait MT-GWAS models, consistent markers for seedling emergence were located on 2B and 5A, whereas for coleoptile length, they were identified on chromosomes 1A and 6A. These results are similar to those found using ST-GWAS models. The COM MT-GWAS model identified 19 consistent markers on chromosome 5A, including the large effect marker *S5A_522153944* (Figure 7). In addition, the COM model identified consistent markers on chromosomes 2B (1), 5B (1), and 7B (3). The 2B and 5B markers were *S2B_301162652* and *S5B_491273019*. The 7B markers are all completely linked with R^2 values of 1, with the marker *S7B_663828309* having the largest effect. The 7B markers were the only consistent markers in MT-GWAS not found consistent in the ST-GWAS models.

Effect of Combining Favorable Alleles

The frequencies of the favorable allele for consistent markers across populations and the consistent markers across years for seedling emergence are important for understanding the selection and makeup of the populations. The consistent markers with LD above 0.8 R^2 were binned together with the marker, and the largest effect marker was identified to represent the bin. This resulted in three bins for the markers identified across populations remaining and 12 bins for the markers identified across years remaining (Supplementary Table 5). In addition, the significant marker on 7B identified in the COM MT-GWAS model was included with the markers consistent across years. These bins were used to display the additive effect for seedling emergence. Most of the markers had high favorable allele frequencies (~90%) in both the DP and BLs (Supplementary Table 5). This shows that these markers have been indirectly selected in both populations.

Even though the consistent markers across populations and years had high favorable allele frequency in the populations, the differences were shown when we combined them and compared the cumulative effect of favorable alleles on seedling emergence (Figures 8A–D). For the consistent markers across populations, the majority of lines in both populations had favorable alleles for all the three bins, which were all located on chromosome 5A. Both populations showed an additive effect, but in the DP, the lines with all three bins had a lower mean compared to the lines with just two bins (Figure 8B). However, in the BLs, there was an increasing trend in emergence with the accumulation of favorable alleles (Figure 8A). The consistent markers across the populations only accounted for 8% of the total variation in the DP and 19% in the BLs. For consistent markers across years, the DP showed an additive effect for all 12 bins, whereas the BLs showed a diminishing return, signifying some of the bins may not have a large effect within the BL population (Figures 8C,D). For the DP and BLs, the majority of lines had either eight or nine favorable

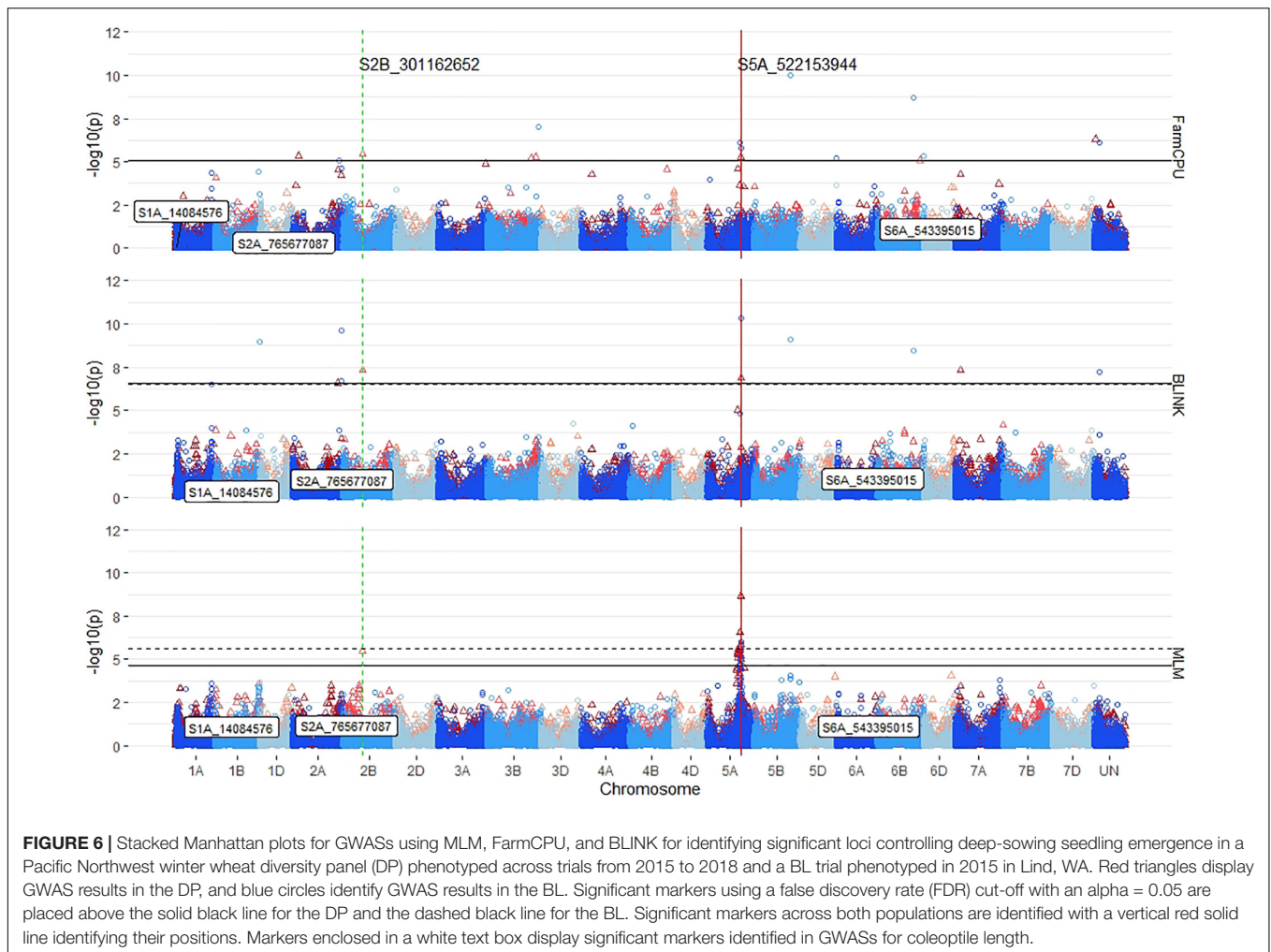


FIGURE 6 | Stacked Manhattan plots for GWASs using MLM, FarmCPU, and BLINK for identifying significant loci controlling deep-sowing seedling emergence in a Pacific Northwest winter wheat diversity panel (DP) phenotyped across trials from 2015 to 2018 and a BL trial phenotyped in 2015 in Lind, WA. Red triangles display GWAS results in the DP, and blue circles identify GWAS results in the BL. Significant markers using a false discovery rate (FDR) cut-off with an $\alpha = 0.05$ are placed above the solid black line for the DP and the dashed black line for the BL. Significant markers across both populations are identified with a vertical red solid line identifying their positions. Markers enclosed in a white text box display significant markers identified in GWASs for coleoptile length.

alleles. The consistent markers across years accounted for only 23% of the variation in the BLs and 31% in the DP.

DISCUSSION

Complex Traits and Seedling Emergence

Complex traits are quantitative in nature and are affected by many small-effect QTLs (Holland, 2007). The challenges that impede the understanding of complex traits are the inability to statistically detect and map minor effect QTLs, accurately understand GEI, and account for pleiotropic effects (Luo et al., 2017). This study attempted to characterize one such trait, seedling emergence for deep-sown winter wheat, to be used as a model for other complex traits.

Our study found no significant associations between seedling emergence and coleoptile length. In previous literature, major QTLs for coleoptile length was reported on chromosomes 4BS, 4BL, and 5AL in wheat (Rebetzke et al., 2001). The QTLs on 4B were reported on either side of the *Rht-B1* gene. This study was further enhanced by a subsequent QTL analysis that resolved the two 4B QTLs directly to the *Rht-B1* locus (Rebetzke et al., 2007).

Our study found no significant markers on chromosome 4B for coleoptile length or seedling emergence. Two small effect markers were found on chromosome 4D using the joint COM and FULL MT-GWAS models but were not consistent across years. The major consistent markers validated across populations for seedling emergence in our study were identified on chromosome 5A. In the follow-up study by Rebetzke et al. (2007), QTLs for coleoptile length on chromosome 5A were identified but were not repeatable across populations. In the same study, Rebetzke et al. (2007) identified small-effect QTLs for coleoptile length on wheat chromosomes 1A, 2B, 2D, 3A, 3B, 5A, and 6A. In our study, significant markers for coleoptile length were found on chromosomes 1A, 2A, and 6B and only in the DP, but none were also significant for seedling emergence. Therefore, we conclude that the GWAS models tested are not selecting major markers for coleoptile length, and that the seedling emergence we are dissecting is indeed due to other factors affecting seedling emergence, similar to what was indicated in Mohan et al. (2013).

Our study demonstrated that chromosome 5A is associated with seedling emergence, with four large effect markers identified consistently across both years in the DP and validated in the BLs. Therefore, a major effect of QTL for seedling emergence, with

TABLE 5 | Consistent significant markers across both the diversity panel and breeding lines for controlling deep-sowing seedling emergence in a Pacific Northwest winter wheat breeding trial using BLINK, FarmCPU, and ML genome-wide association studies (GWAS) models using covariates for coleoptile length and Rht alleles.

Marker	Position ^a	Alleles ^b	Chr ^a	Pop ^c	Year	Model	Covariate ^d	p-Value	MAF ^e	R ^{2f}	Effect (%) ^g
S5A_514260464	5.14E+08	<u>C</u> /T	5A	BL	15	FarmCPU	N	7.86E-07	0.23	0.13	17
				DP	15	MLM	N, R, C, B	1.81E-09	0.11	0.08	19
				DP	15–17	MLM	N, R, C, B	6.69E-07	0.11	0.05	9
				DP	15–18	MLM	N, R, C, B	6.48E-06	0.11	0.02	5
S5A_522153944	5.22E+08	<u>C</u> /T	5A	DP	15	BLINK	N, C	1.58E-11	0.04	0.09	29
				DP	15–17	BLINK	N, C, B	3.27E-12	0.04	0.08	17
				DP	15–18	BLINK	N, R, C	1.83E-06	0.04	0.10	15
				DP	15–19	BLINK	N, R, C, B	1.82E-09	0.04	0.06	12
				BL	15	FarmCPU	N	1.69E-06	0.04	0.12	30
				DP	15	FarmCPU	N, C	2.44E-06	0.04	0.09	29
				DP	15–18	FarmCPU	N, R	5.13E-06	0.04	0.10	15
				DP	15–19	FarmCPU	R, C, B	6.15E-06	0.04	0.06	12
				BL	15	MLM	N	9.29E-07	0.04	0.12	30
				DP	15	MLM	N, R, C, B	3.62E-09	0.04	0.09	29
				DP	15–17	MLM	N, R, C, B	7.77E-09	0.04	0.08	17
				S5A_522153953	5.22E+08	<u>A</u> /G	5A	DP	15–17	BLINK	R
DP	15	FarmCPU	B					1.03E-07	0.04	0.09	29
DP	15–18	FarmCPU	B					4.54E-07	0.04	0.10	15
BL	15	MLM	N					9.29E-07	0.04	0.12	30
DP	15	MLM	N, R, C, B					3.62E-09	0.04	0.09	29
DP	15–17	MLM	N, R, C, B					7.77E-09	0.04	0.08	17
DP	15–18	MLM	N, R, C, B					2.67E-09	0.04	0.10	15
S5A_523025549	5.23E+08	<u>A</u> /G	5A	BL	15	BLINK	N	4.56E-11	0.04	0.12	31
				BL	15	MLM	N	2.57E-06	0.04	0.12	31
				DP	15	MLM	N, R, C, B	1.03E-06	0.05	0.06	23
				DP	15–17	MLM	N, R, C, B	1.15E-05	0.05	0.05	13
				DP	15–18	MLM	N, R, C, B	1.33E-06	0.05	0.07	12

^aChromosomes and positions of markers were determined according to IWGSC RefSeq v.1.1.

^bAllele: favorable allele is underlined.

^cPop: BL, breeding lines; DP, diversity panel.

^dCovariate: N, none; R, reduced height alleles; C, BLUPs for coleoptile length; B, both reduced height and coleoptile length.

^eMAF: minor allele frequency.

^fR²: phenotypic variance explained by associated marker.

^gEffect: increase of seedling emergence (%) with the inclusion of favorable allele of associated marker.

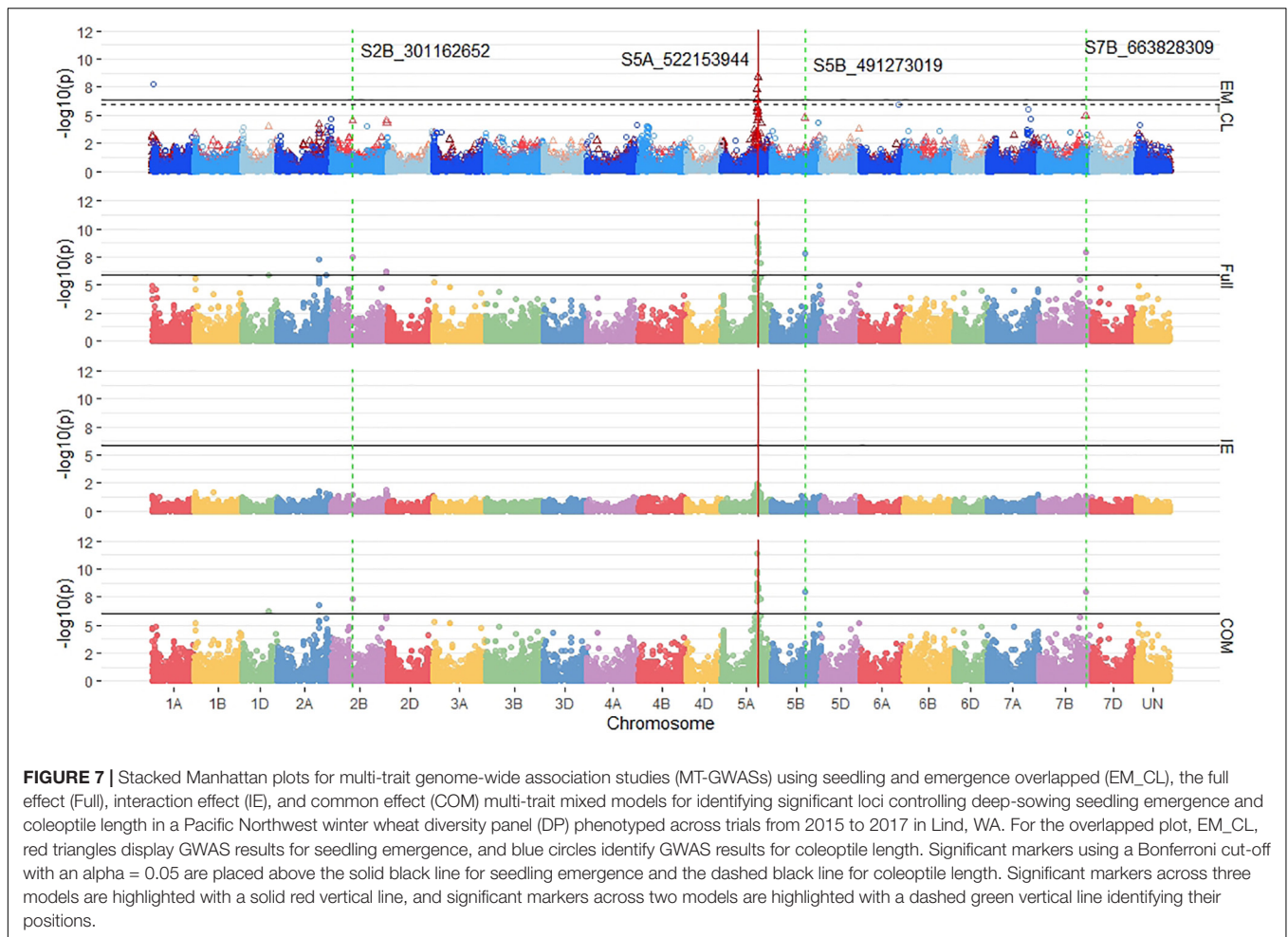
an increase of up to 30%, may be present in that chromosomal location. According to Bernardo (2014), a major marker accounts for >10% of the phenotypic variation. Since this 5A locus is associated with seedling emergence and not with coleoptile length, other associated traits are improving seedling emergence. These associated traits may be fast emergence, the ability to germinate and/or grow under moisture stress, or other unknown mechanisms. Since the variation for emergence in the DP and BLs was due to moisture stress, the 5A locus may affect both fast emergence and the ability to germinate and grow under low moisture. The identification of a few large effects and many small effects significant markers confirms that seedling emergence is a complex trait controlled by multiple pleiotropic factors other than coleoptile length.

Genome-Wide Association Models

Genetic mapping through association studies has been performed to dissect the genetic architecture of various traits in wheat

(Adhikari et al., 2012; Lozada et al., 2017, 2019; Bajgain et al., 2019; Lozada and Carter, 2020). However, GWAS has not been implemented for deep-sowing seedling emergence in winter wheat but has been conducted in rice (Zhao et al., 2018). Even further, few studies have compared ST-GWAS models, and fewer have compared FarmCPU and BLINK (Liu et al., 2016; Huang et al., 2019). By comparing multiple covariates and GWAS models, we attempted to compare the ability of models and covariates to identify significant markers to dissect a trait with unknown genetic architecture. This allowed us to identify large effect markers with up to 30% effect and 10% R² values. Therefore, we can conclude that these markers are important for seedling emergence.

The single-locus model used in our study was MLM. The MLM identified significant markers mainly on chromosome 5A and was consistent in identifying large effect markers without the inclusion of covariates. However, the MLM displays low ability in identifying both major markers and other lower effect significant

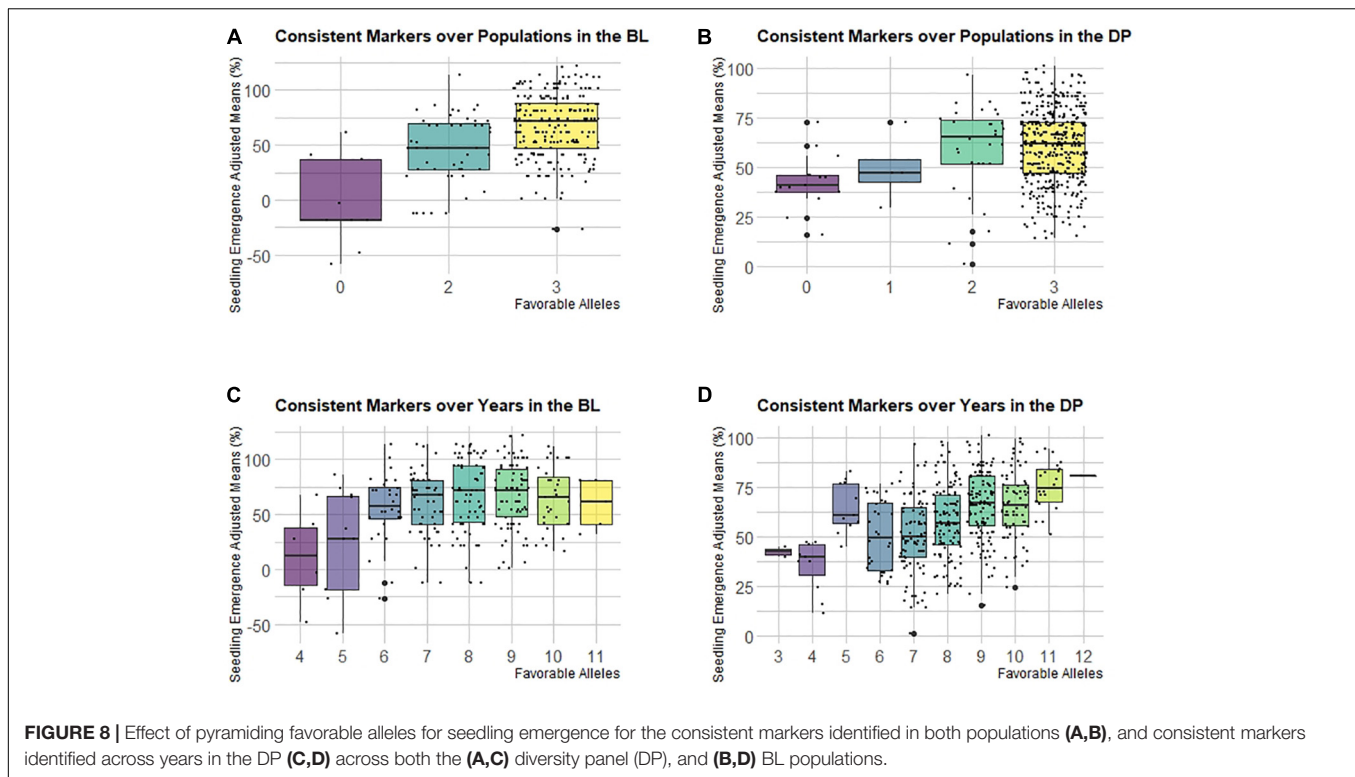


markers in the presence of high levels of LD. Furthermore, while MLMs have been shown to be comparable to FarmCPU (a multi-locus model) for simple traits, the MLMs have been shown to have difficulties with low power and false positives for complex traits (Habier et al., 2011; Liu et al., 2016; Ward et al., 2019). MLMs evaluate the relationship and overall variation of each genetic marker independently; therefore, as the trait increases in complexity and number of pleiotropic effects, the proportion of variation due to a locus decreases (Miao et al., 2019).

Fixed and random model circulating probability unification and multi-locus models provide the best trade-off between power and false positives (Liu et al., 2016; Miao et al., 2019). Multi-locus models increase the proportion of genetic variance using major effect markers as fixed effects and identifying more significant markers (Miao et al., 2019). This is why our multi-locus models were able to dissect the complex trait of seedling emergence accurately and identify both major and minor effect markers. The multi-locus models used in our study were FarmCPU and BLINK and had very similar results in the BL population, but there were fewer significant markers discovered by BLINK in the DP. The difference between BLINK and FarmCPU is that Bayesian Information Criteria in a fixed-effects model replaces REML in the random-effects model, and LD information is used to replace

the bin method implemented in FarmCPU. Therefore, since the LD is high among significant markers on the same chromosome in both populations, the amount of markers grouped together may be different, which reduces the ability to identify significant markers on 5A across populations. Apart from the markers on 5A that the MLM also found, most markers found significant by FarmCPU and BLINK had relatively small effects and R^2 values, indicating the ability of multi-locus models to identify small-effect markers for complex traits.

There was little to no increase in identifying pleiotropic loci using MT-GWAS compared to ST-GWAS models. This may be due to the lack of similar significant markers or pleiotropic loci between seedling emergence and coleoptile length, along with the inflated p -values we observed in the joint analyses. We observed such large inflation of p -values that we had to use the more stringent Bonferroni cut-off as the threshold. The MT-GWAS model we used implemented a single-locus model and, therefore, suffered limitations in identifying small effect loci similar to the MLM. However, the COM model did identify consistent markers across years on chromosome 7A, whereas the ST-GWAS model only identified a single-year combination. The identification of a significant marker in the COM model not seen in the ST-GWAS, combined with the lack of significant interaction



effect loci, confirms the positive genetic correlation seen in the majority of years between seedling emergence and coleoptile length. However, due to the lack of evidence of pleiotropic loci in our other methods, the genetic correlation may be due to the correlation of shared loci such as the *Rht* alleles.

The lack of increase in ability to identify pleiotropic loci using the MT-GWAS models may also be due to the lack of phenotypic correlation between seedling emergence and coleoptile length (Korte et al., 2012). Even though we see a positive genetic correlation, the phenotypic correlation is near zero, indicating a small genetic effect in comparison to the environmental effect. In this scenario, it has been shown that the MT-GWAS models do not outperform ST-GWAS models and confirm why our MT-GWAS models did not outperform the ST-GWAS models (Korte et al., 2012). However, MT-GWAS can still be used as a complement to ST-GWAS models to identify interactions and pleiotropic loci such as the loci on chromosome 7A.

Covariates

Covariates are commonly used in GWAS studies to allow for models to differentiate genetic and environmental effects (Tibbs Cortes et al., 2021). For a complex trait, more confounding associations such as correlated traits or pleiotropic effects may account for a portion of seedling emergence variation (Saltz et al., 2017). To account for these correlated traits, we compared the use of covariates for *Rht* alleles and coleoptile length. These correlated traits were used because of previous studies reporting large associations between the traits (Allan et al., 1962; Sunderman, 1964; Chastain et al., 1995; Schillinger et al., 1998; Botwright et al., 2001; Schillinger, 2011).

The *Rht* and coleoptile length covariates had an effect on identifying significant markers for the multi-locus models but not for the MLM. The multi-locus models required the inclusion of covariates in order to identify the major effect of significant markers on chromosome 5A, but not for identifying the small effect markers. Including covariates or secondary traits can account for potential confounding factors that bias marker effect estimates (Aschard et al., 2015). There were no common significant markers between seedling emergence and coleoptile length, including covariates, that affected the ability of the multi-locus models to identify the large effect markers. This indicates that there may be confounding effects between coleoptile length and seedling emergence and that the relationship between coleoptile length and *Rht* alleles in our populations is no longer linear. Therefore, even though modern varieties are no longer dependent on coleoptile length for improved emergence, it may still confer some confounding effects. This is further confirmed by the results of our MT-GWAS models and moderate genetic correlations. Therefore, we recommend using both multi-locus models and covariates for correlated traits to identify both small and large effect loci.

Association Mapping Populations and Environments

Closer relationships between genotypes indicate that fewer recombination events have occurred, which preserves marker LD and requires a less genetic variation to be accounted for by the models (Habier et al., 2007; Bernardo, 2020). The DP had a more distinct population structure compared to the BLs,

indicating that the BL population is more genetically related. The BL population was purposely selected over generations for deep-sowing seedling emergence in the Washington State University breeding program. Even though these lines have diverse pedigrees, they are based on specific founder lines with good emergence, which may be the reason for the increased relatedness. In contrast, the DP is composed of varieties from various breeding programs in the Pacific Northwest, with the majority of lines not bred specifically for deep-sowing seedling emergence. Even with the differences in population structure, we were able to identify the consistent markers on chromosome 5A in both populations.

In addition to the differences between the composition and genetic relatedness of the populations, the number of environments examined between the two populations differed. The BLs only had emergence data for 2015, whereas the DP, an unselected population for seedling emergence, had trait data for multiple years. The combining of years and accounting for GEI in our phenotypic adjustments of the DP allowed for the GWAS models to increase the power to dissect seedling emergence compared to individual years. Furthermore, the lack of consistency between GWAS models and populations confirms the complexity of seedling emergence and the need for evaluation across multiple environments. Since the environment can create phenotypic variation in seedling emergence, it displays GEI. If a trait displays GEI, it follows that so would the QTLs responsible for the phenotypic expression (Bernardo, 2020). A change in the ranking of QTLs across environments indicates QTL-by-Environment Interaction (QEI), with the detection of QTLs in some environments and not others (Bernardo, 2020). Therefore, QEI can be seen with the differing number of significant markers in individual and combined years.

Furthermore, the difficulty in dissecting seedling emergence within and across years can be seen in the difference between the varying genetic and phenotypic correlations between years and traits. The differences in genetic correlations for seedling emergence and coleoptile length from year to year and the near-zero phenotypic correlations can be explained by the large effect of the environment and the multitude of factors that affect seedling emergence. The phenotypic effect can be partitioned into both genetic and non-genetic effects (Searle, 1961). Since genetic correlation only takes into account the genetic effect, we still see moderate values. The negative genetic correlation in 2017 can be explained by the different factors that affect seedling emergence, such as coleoptile diameter, force, and speed of emergence, and not coleoptile length. The fact that other factors are affecting seedling emergence is further confirmed because of the lack of significant markers for the interaction effect in the MT-GWAS models (Korte et al., 2012). If the negative genetic correlation was due to the significant loci for coleoptile length, we would see significant markers for the interaction effect (Korte et al., 2012). Additionally, the near-zero phenotypic correlations display that coleoptile length, while having moderate genetic correlations, has a much smaller genetic effect than the environmental effect, and therefore has little overall phenotypic effect for seedling emergence. In other terms, when the environment has large control over the expression of the trait but a low correlation, the

phenotypic correlation would be expected to be lower than the genetic correlation (Searle, 1961).

Combining Favorable Alleles

We see an increase in seedling emergence and additive effect of accumulating favorable alleles using the consistent markers identified across years in the DP. In contrast, in the BLs, there was an additive effect of accumulating favorable alleles for the consistent markers identified over populations. The difference between the two sets of markers was that the consistent markers over populations were exclusively located on chromosome 5A, whereas the consistent markers over years were located on various chromosomes. All the consistent markers displayed high frequencies in both populations, which demonstrated the success of accumulating favorable alleles through traditional phenotypic selection.

There was a general additive effect with the accumulation of favorable alleles for both marker sets. The consistent markers identified across years were based on the results of the DP and showed a continual increase in seedling emergence as favorable alleles were accumulated. In the BLs, seedling emergence improved through additivity but only to a point; once the number of favorable alleles was high, no improvement was found with continual accumulation. This may be because the DP is an unselected population, whereas the BLs had previously been selected for improved emergence, and therefore may already have fixed a high number of alleles for emergence. Since there was not a large decrease in seedling emergence with the accumulation of favorable markers, we can presume that seedling emergence is generally additive with the possibility of some interaction or non-additive gene action, as seen in other complex traits (Bonnafous et al., 2018). Pyramiding favorable alleles has shown to be successful for disease resistance traits (Naruoka et al., 2015; Lewien et al., 2018; Lozada et al., 2019) and can be equally successful for abiotic traits such as seedling emergence.

CONCLUSION

This study displayed the ability of GWAS to dissect the genetic architecture of a complex trait such as deep-sown seedling emergence. Both the ST-GWAS and MT-GWAS models identified a few large effects and many small effect markers for seedling emergence. Additionally, neither the ST-GWAS nor the MT-GWAS models identified large pleiotropic effect markers between seedling emergence and coleoptile length. The ST-GWAS and MT-GWAS models did not identify the same significant markers for seedling emergence or coleoptile length, and the MT-GWAS models did not identify any interaction effect markers. However, by using multi-locus models in conjunction with covariates for correlated traits, we were able to identify more small effect loci over single-locus models to dissect a complex trait in both the DP and BL populations. Additionally, the DP displayed the necessity for combining years for consistent identification of significant markers for a trait dependent on the environment for phenotypic variation. Furthermore, the MT-GWAS models displayed a lower ability to identify small effect

loci over ST-GWAS models for single-trait analysis and inflated p -values for joint analysis but still identified the large effect markers on 5A. Therefore, using multi-locus models combined with covariates (such as major genes), breeding programs can uncover the complex nature of traits to help identify candidate genes and the underlying architecture of a trait to make more efficient breeding decisions and selection methods.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://github.com/lfmerrick21/GWAS-Complex-Traits>.

AUTHOR CONTRIBUTIONS

LM conceptualized the idea, analyzed the data, and drafted the manuscript. AB genotyped the KASP markers, and reviewed and edited the manuscript. ZZ reviewed and edited the manuscript. AC supervised the study, conducted field trials, edited the manuscript, and obtained the funding for the project. All authors contributed to the article and approved the submitted version.

REFERENCES

- Adhikari, T. B., Gurung, S., Hansen, J. M., Jackson, E. W., and Bonman, J. M. (2012). Association Mapping of quantitative trait loci in spring wheat landraces conferring resistance to bacterial leaf streak and spot blotch. *Plant Genome J.* 5:1. doi: 10.3835/plantgenome2011.12.0032
- Allan, R. E. (1989). Agronomic comparisons between Rht1 and Rht2 semi-dwarf genes in winter wheat. *Crop Sci.* 29, 1103–1108. doi: 10.2135/cropsci1989.0011183X002900050001x
- Allan, R. E., Vogel, O. A., Burleigh, J. R., and Peterson, C. J. (1961). Inheritance of coleoptile length and its association with culm length in four winter wheat crosses 1. *Crop Sci.* 1, 328–332. doi: 10.2135/cropsci1961.0011183X000100050009x
- Allan, R. E., Vogel, O. A., and Craddock, J. C. Jr. (1959). Comparative response to gibberellic acid of dwarf, semi-dwarf, and standard short and tall winter wheat varieties 1. *Agron. J.* 51, 737–740. doi: 10.2134/agronj1959.00021962005100120013x
- Allan, R. E., Vogel, O. A., and Peterson, C. J. (1962). Seedling emergence rate of fall-sown wheat and its association with plant height and coleoptile length 1. *Agron. J.* 54, 347–350. doi: 10.2134/agronj1962.00021962005400040022x
- Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. doi: 10.1126/science.aar7191
- Arndt, W. (1965). The impedance of soil seals and the forces of emerging seedlings. *Soil Res.* 3, 55–68.
- Aschard, H., Vilhjálmsson, B. J., Joshi, A. D., Price, A. L., and Kraft, P. (2015). Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *Am. J. Hum. Genet.* 96, 329–339. doi: 10.1016/j.ajhg.2014.12.021
- Bajgain, P., Zhang, X., and Anderson, J. A. (2019). Genome-wide association study of yield component traits in intermediate wheatgrass and implications in genomic selection and breeding. *G3amp58 Genes Genomes Genet.* 9, 2429–2439. doi: 10.1534/g3.119.400073

FUNDING

This research was partially funded by the National Institute of Food and Agriculture (NIFA) of the US Department of Agriculture (Award number 2016-68004-24770), Hatch project 1014919, and the OA Vogel Research Foundation at Washington State University.

ACKNOWLEDGMENTS

We would like to acknowledge Washington State University Winter Wheat Breeding Program personnel Gary Shelton and Kyall Hagemeyer for plot maintenance and data collection under field conditions and Kerry Barlow for coleoptile length data collection. We would also like to thank Gina Brown-Guedira, Jared Smith, Brian Ward, and staff at the Eastern Regional Small Grains Genotyping Laboratory for their assistance with DNA library prep and GBS sequencing and analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.772907/full#supplementary-material>

- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Bernardo, R. (2014). Genomewide selection when major genes are known. *Crop Sci.* 54:68. doi: 10.2135/cropsci2013.05.0315
- Bernardo, R. (2020). *Breeding for Quantitative Traits in Plants*, 3rd Edn. Woodbury, Minnesota: Stemma Press.
- Bonnafous, F., Fievet, G., Blanchet, N., Boniface, M.-C., CarrÁire, S., Gouzy, J., et al. (2018). Comparison of GWAS models to identify non-additive genetic control of flowering time in sunflower hybrids. *Theor. Appl. Genet.* 131, 319–332. doi: 10.1007/s00122-017-3003-4
- Botwright, T., Rebetzke, G., Condon, T., and Richards, R. (2001). The effect of rht genotype and temperature on coleoptile growth and dry matter partitioning in young wheat seedlings. *Funct. Plant Biol.* 28, 417–423. doi: 10.1071/pp01010
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308
- Browning, B. L., Zhou, Y., and Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* 103, 338–348. doi: 10.1016/j.ajhg.2018.07.015
- Chastain, T. G., Ward, K. J., and Wysocki, D. J. (1995). Stand establishment response of soft white winter wheat to seedbed residue and seed size. *Crop Sci.* 35:cropsci1995.0011183X003500010040x. doi: 10.2135/cropsci1995.0011183X003500010040x
- Covarrubias-Pazarán, G. (2018). *Software Update: Moving the R Package Sommer to Multivariate Mixed Models for Genome-Assisted Prediction*. Available online at: <https://www.biorxiv.org/content/10.1101/354639v1> (accessed September 8, 2021)
- Cullis, B. R., Smith, A. B., and Coombes, N. E. (2006). On the design of early generation variety trials with correlated data. *J. Agric. Biol. Environ. Stat.* 11:381. doi: 10.1198/108571106X154443
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach

- for high diversity species. *PLoS One* 6:e19379. doi: 10.1371/journal.pone.0019379
- Evans, C. E., and Etherington, J. R. (1990). The effect of soil water potential on seed germination of some British plants. *New Phytol.* 115, 539–548. doi: 10.1111/j.1469-8137.1990.tb00482.x
- Federer, W. F. (1956). *Experimental Design, Theory and Application*. New York, NY: Macmillan.
- Gale, M. D., and Youssefian, S. (1983). “Pleiotropic effects of the Norin 10 dwarfing genes, Rht1 and Rht2, and interactions in response to chlormequat,” in *Proceedings of the Sixth International Wheat Genetics Symposium*, ed. S. Sakamoto (Kyoto: Plant Germ-Plasm Institute, Faculty of Agriculture, Kyoto University).
- Galesloot, T. E., Van Steen, K., Kiemeneij, L. A., Jans, L. L., and Vermeulen, S. H. (2014). A comparison of multivariate genome-wide association methods. *PLoS One* 9:e95923. doi: 10.1371/journal.pone.0095923
- Glaubitz, J. C., Castevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., et al. (2014). TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS One* 9:e90346. doi: 10.1371/journal.pone.0090346
- Goldman, I. (2019). *Plant Breeding Reviews*. Hoboken, NJ: John Wiley & Sons.
- Grogan, S. M., Brown-Guedira, G., Haley, S. D., McMaster, G. S., Reid, S. D., Smith, J., et al. (2016). Allelic variation in developmental genes and effects on winter wheat heading date in the U.S. Great Plains. *PLoS One* 11:e0152852. doi: 10.1371/journal.pone.0152852
- Habier, D., Fernando, R. L., and Dekkers, J. C. M. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177, 2389–2397. doi: 10.1534/genetics.107.081190
- Habier, D., Fernando, R. L., Kizilkaya, K., and Garrick, D. J. (2011). Extension of the bayesian alphabet for genomic selection. *BMC Bioinform.* 12:1–12. doi: 10.1186/1471-2105-12-186
- Holland, J. B. (2007). Genetic architecture of complex traits in plants. *Curr. Opin. Plant Biol.* 10, 156–161. doi: 10.1016/j.pbi.2007.01.003
- Huang, M., Liu, X., Zhou, Y., Summers, R. M., and Zhang, Z. (2019). BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience* 8:154. doi: 10.1093/gigascience/giy154
- Korte, A., Vilhjálmsson, B. J., Segura, V., Platt, A., Long, Q., and Nordborg, M. (2012). A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat. Genet.* 44, 1066–1071. doi: 10.1038/ng.2376
- Lander, E. S., and Schork, N. J. (1994). Genetic dissection of complex traits. *Science* 265, 2037–2048.
- Lewin, M. J., Murray, T. D., Jernigan, K. L., Garland-Campbell, K. A., and Carter, A. H. (2018). Genome-wide association mapping for eyespot disease in US Pacific Northwest winter wheat. *PLoS One* 13:e0194698. doi: 10.1371/journal.pone.0194698
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Lipka, A. E., Kandianis, C. B., Hudson, M. E., Yu, J., Drnevich, J., Bradbury, P. J., et al. (2015). From association to prediction: statistical methods for the dissection and selection of complex traits in plants. *Curr. Opin. Plant Biol.* 24, 110–118. doi: 10.1016/j.pbi.2015.02.010
- Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* 12:1005767. doi: 10.1371/journal.pgen.1005767
- Lozada, D., Godoy, J. V., Murray, T. D., Ward, B. P., and Carter, A. H. (2019). Genetic dissection of snow mold tolerance in US pacific northwest winter wheat through genome-wide association study and genomic selection. *Front. Plant Sci.* 10:1337. doi: 10.3389/fpls.2019.01337
- Lozada, D. N., and Carter, A. H. (2020). Insights into the genetic architecture of phenotypic stability traits in winter wheat. *Agronomy* 10:368. doi: 10.3390/agronomy10030368
- Lozada, D. N., Mason, R. E., Babar, M. A., Carver, B. F., Guedira, G.-B., Merrill, K., et al. (2017). Association mapping reveals loci associated with multiple traits that affect grain yield and adaptation in soft winter wheat. *Euphytica* 213:222. doi: 10.1007/s10681-017-2005-2
- Luo, Z., Wang, M., Long, Y., Huang, Y., Shi, L., Zhang, C., et al. (2017). Incorporating pleiotropic quantitative trait loci in dissection of complex traits: seed yield in rapeseed as an example. *Theor. Appl. Genet.* 130, 1569–1585. doi: 10.1007/s00122-017-2911-7
- Lutcher, L. K., Wuest, S. B., and Johlke, T. R. (2019). First leaf emergence force of three deep-planted winter wheat cultivars. *Crop Sci.* 59, 772–777. doi: 10.2135/cropsci2018.08.0495
- Merrick, L. F., and Carter, A. H. (2021). Comparison of genomic selection models for exploring predictive ability of complex traits in breeding programs. *bioRxiv* [preprint]. doi: 10.1101/2021.04.15.440015
- Miao, C., Yang, J., and Schnable, J. C. (2019). Optimising the identification of causal variants across varying genetic architectures in crops. *Plant Biotechnol. J.* 17, 893–905. doi: 10.1111/pbi.13023
- Mohan, A., Schillinger, W. F., and Gill, K. S. (2013). Wheat seedling emergence from deep planting depths and its relationship with coleoptile length. *PLoS One* 8:e73314. doi: 10.1371/journal.pone.0073314
- Murphy, K., Balow, K., Lyon, S. R., and Jones, S. S. (2008). Response to selection, combining ability and heritability of coleoptile length in winter wheat. *Euphytica* 164, 709–718. doi: 10.1007/s10681-008-9692-7
- Naruoka, Y., Garland-Campbell, K. A., and Carter, A. H. (2015). Genome-wide association mapping for stripe rust (*Puccinia striiformis* F. sp. *tritici*) in US Pacific Northwest winter wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 128, 1083–1101. doi: 10.1007/s00122-015-2492-2
- Pill, W. G. (1995). “Low water potential and presowing germination treatments to improve seed quality,” in *Seed Quality: Basic Mechanisms and Agricultural Implications*, ed. A. S. Basra (Binghamton, NY: The Haworth Press), 319–359.
- Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., et al. (2012). Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome J.* 5:103. doi: 10.3835/plantgenome2012.06.0006
- Porter, H. F., and O’Reilly, P. F. (2017). Multivariate simulation framework reveals performance of multi-trait GWAS methods. *Sci. Rep.* 7, 1–12. doi: 10.1038/srep38837
- Rasheed, A., Wen, W., Gao, F., Zhai, S., Jin, H., Liu, J., et al. (2016). Development and validation of KASP assays for genes underpinning key economic traits in bread wheat. *Theor. Appl. Genet.* 129, 1843–1860. doi: 10.1007/s00122-016-2743-x
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rebetzke, G. J., Appels, R., Morrison, A. D., Richards, R. A., McDonald, G., Ellis, M. H., et al. (2001). Quantitative trait loci on chromosome 4B for coleoptile length and early vigour in wheat (*Triticum aestivum* L.). *Aust. J. Agric. Res.* 52, 1221–1234. doi: 10.1071/AR01042
- Rebetzke, G. J., Ellis, M. H., Bonnett, D. G., and Richards, R. A. (2007). Molecular mapping of genes for coleoptile growth in bread wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 114, 1173–1183. doi: 10.1007/s00122-007-0509-1
- Rebetzke, G. J., Richards, R. A., Fischer, V. M., and Mickelson, B. J. (1999). Breeding long coleoptile, reduced height wheats. *Euphytica* 106, 159–168. doi: 10.1023/A:1003518920119
- Saltz, J. B., Hessel, F. C., and Kelly, M. W. (2017). Trait correlations in the genomics era. *Trends Ecol. Evol.* 32, 279–290. doi: 10.1016/j.tree.2016.12.008
- SAS Institute, Inc (2011). *SAS® 9.3 System Options: Reference*. Cary, NC: SAS Institute Inc.
- Schillinger, W. F. (2011). Rainfall impacts winter wheat seedling emergence from deep planting depths. *Agron. J.* 103, 730–734. doi: 10.2134/agronj2010.0442
- Schillinger, W. F., Donaldson, E., Allan, R. E., and Jones, S. S. (1998). Winter wheat seedling emergence from deep sowing depths. *Agron. J.* 90, 582–586. doi: 10.2134/agronj1998.00021962009000050002x
- Schillinger, W. F., Schofstoll, S. E., Smith, T. A., and Jacobsen, J. A. (2017). Laboratory method to evaluate wheat seedling emergence from deep planting depths. *Agron. J.* 109, 2004–2010. doi: 10.2134/agronj2016.12.0715
- Schmidt, P., Hartung, J., Bennewitz, J., and Piepho, H.-P. (2019). Heritability in plant breeding on a genotype-difference basis. *Genetics* 212, 991–1008. doi: 10.1534/genetics.119.302134
- Searle, S. R. (1961). Phenotypic, genetic and environmental correlations. *Biometrics* 17, 474–480. doi: 10.2307/2527838
- Sunderman, D. W. (1964). Seedling emergence of winter wheats and its association with depth of sowing, coleoptile length under various conditions, and plant height. *Agron. J.* 56, 23–25. doi: 10.2134/agronj1964.00021962005600010008x

- Tang, Y., Liu, X., Wang, J., Li, M., Wang, Q., Tian, F., et al. (2016). GAPIT version 2: an enhanced integrated tool for genomic association and prediction. *Plant Genome* 9:120. doi: 10.3835/plantgenome2015.11.0120
- Tibbs Cortes, L., Zhang, Z., and Yu, J. (2021). Status and prospects of genome-wide association studies in plants. *Plant Genome* 14:e20077. doi: 10.1002/tpg2.20077
- Vogel, O. A., Craddock, J. C. Jr., Muir, C. E., Everson, E. H., and Rohde, C. R. (1956). Semidwarf growth habit in winter wheat improvement for the pacific northwest I. *Agron. J.* 48, 76–78. doi: 10.2134/agronj1956.00021962004800020009x
- Wang, Q., Tian, F., Pan, Y., Buckler, E. S., and Zhang, Z. (2014). A super powerful method for genome wide association study. *PLoS One* 9:e107684. doi: 10.1371/journal.pone.0107684
- Ward, B. P., Brown-Guedira, G., Kolb, F. L., Van Sanford, D. A., Tyagi, P., Sneller, C. H., et al. (2019). Genome-wide association studies for yield-related traits in soft red winter wheat grown in virginia. *PLoS One* 14:e0208217. doi: 10.1371/journal.pone.0208217
- Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., et al. (2021). rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics Proteomics Bioinformatics*. doi: 10.1016/j.gpb.2020.10.007
- Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208. doi: 10.1038/ng1702
- Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42, 355–360. doi: 10.1038/ng.546
- Zhao, Y., Zhao, W., Jiang, C., Wang, X., Xiong, H., Todorovska, E. G., et al. (2018). Genetic architecture and candidate genes for deep-sowing tolerance in rice revealed by non-syn GWAS. *Front. Plant Sci.* 9:332. doi: 10.3389/fpls.2018.00332

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Merrick, Burke, Zhang and Carter. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.