Leveraging Breeding Values Obtained from Random Regression Models for Genetic Inference of Longitudinal Traits

Malachy Campbell*, Mehdi Momen, Harkamal Walia, and Gota Morota

M. Campbell, M. Momen, G. Morota, Dep. of Animal and Poultry Sciences, Virginia Polytechnic Inst. and State University, Blacksburg, VA 24061; H. Walia, Dep. of Agronomy and Horticulture, Univ. of Nebraska Lincoln, Lincoln, NE 68583.

ABSTRACT Understanding the genetic basis of dynamic plant phenotypes has largely been limited because of a lack of space and labor resources needed to record dynamic traits, often destructively, for a large number of genotypes. However, the recent advent of image-based phenotyping platforms has provided the plant science community with an effective means to nondestructively evaluate morphological, developmental, and physiological processes at regular, frequent intervals for a large number of plants throughout development. The statistical frameworks typically used for genetic analyses (e.g., genome-wide association mapping, linkage mapping, and genomic prediction) in plant breeding and genetics are not particularly amenable for repeated measurements. Random regression (RR) models are routinely used in animal breeding for the genetic analysis of longitudinal traits and provide a robust framework for modeling trait trajectories and performing genetic analysis simultaneously. We recently used a RR approach for genomic prediction of shoot growth trajectories in rice (*Oryza sativa* L.) from 33,674 single nucleotide polymorphisms. In this study, we have extended this approach for genetic inference by leveraging genomic breeding values derived from RR models for rice shoot growth during early vegetative development. This approach provides improvements over conventional single time point analyses for discovering loci associated with shoot growth trajectories. The RR approach uncovers persistent as well as time-specific transient quantitative trait loci. This methodology can be widely applied to understand the genetic architecture of other complex polygenic traits with repeated measurements.

Abbreviations: BLUP, best-linear unbiased prediction; GEBV, genomic estimated breeding value; GWAS, genome-wide association study; PSA, projected shoot area; QTL, quantitative trait loci; DAT, days after transplant; RR, random regression; SNP, single nucleotide polymorphism; TP, single time point

CORE IDEAS

- Random regression models are an appealing framework for genome-wide association studies (GWAS) of longitudinal traits
- This approach provides improvements over conventional single time point analyses for GWAS.
- We identify quantitative trait loci with transient and persistent effects on shoot growth in rice.
- This is the first application of random regression models for GWAS of longitudinal traits in crops.

A plant's phenotype at any given time is the manifestation of numerous biological processes that have occurred prior to the capture of the phenotype. In most genetic mapping studies, plants are phenotyped at one or few discrete time points. Though this may be sufficient for end point traits, such as yield or grain quality, other agronomically important traits such as plant height or vigor are not static and vary continuously throughout development. Given the dynamic nature of these traits, it is likely that some genes will make a time-dependent contribution to the phenotype. Approaches that consider such infinite-dimensional traits as static fail to fully capture the dynamic processes that have led to the phenotype and may not uncover the contributions of time-specific loci.

Citation: Campbell, M., M. Momen, H. Walia, and G. Morota. 2018. Leveraging breeding values obtained from random regression models for genetic inference of longitudinal traits. Plant Genome 12:180075. doi: 10.3835/plantgenome2018.10.0075

Received 4 Oct. 2018. Accepted 4 Feb. 2019.
*Corresponding author (campbell.malachy@gmail.com).

Recording phenotypic measurements across development in genetic mapping populations is typically limited because of the high space and labor demands needed to record a trait, often destructively, for a large number of genotypes. However, with the advent of image-based phenotyping platforms, researchers can now capture morphological, developmental, and physiological processes nondestructively with higher temporal resolution for a large number of plants (Fraas and Lüthen, 2015; Simko et al., 2016; Shakoor et al., 2017; Tardieu et al., 2017; Araus et al., 2018). Moreover, the growth of the unmanned aerial vehicle industry in recent years has provided many low-cost hardware options that can be outfitted with cameras, facilitating the collection of temporal phenotypes in field settings (Yang et al., 2017). Although the use of these platforms is becoming more routine in plant genetics, the statistical frameworks typically used for genetic analyses (e.g., genome-wide association mapping, linkage mapping, and genomic prediction) in plant breeding and genetics are not amenable for longitudinal traits.

Several studies in recent years have sought to elucidate the genetic basis of longitudinal traits through genome-wide association studies (GWAS) or linkage mapping. For instance Moore et al. (2013) and Würschum et al. (2014) used linkage mapping at discrete time points to identify time-specific quantitative trait loci (QTL) associated with root gravitropism and plant height, respectively. Though these approaches may be effective, because they consider the phenotype at only a single time point, they do not leverage the covariance among time points and may have reduced statistical power compared with approaches that consider the entire trait trajectory in regression modeling. Several studies have leveraged a "two-step" approach for functional association mapping (Bac-Molenaar et al., 2015; Campbell et al., 2017). In the two-step approach, a nonlinear function is fitted to phenotypic records for each genotype that summarizes the trait trajectories by using a few parameters. These parameters are then used as derived phenotypes in subsequent GWAS analyses. However, with these two-step approaches, information is lost between the curve fitting and genetic analysis steps. The residuals from the first curve-fitting step are likely to contain important information regarding persistent environmental effects that are not considered in subsequent genetic analysis. We hypothesize that an approach that unifies the curve fitting and genetic analysis into a single framework is likely to be better than the single time point (TP) or two-step longitudinal approaches.

Random regression models provide a robust framework for modeling trait trajectories and performing genetic analysis simultaneously (Schaeffer and Dekker, 1994; Huisman et al., 2002; Schaeffer, 2004; Sun et al., 2017). Covariance functions, such as spline or polynomial functions, are used to model trait trajectories for each line and are sufficient to capture the covariance across time points while estimating fewer parameters (Kirkpatrick et al., 1990, 1994; Meyer, 1998; White et al., 1999; Strabel and Misztal, 1999; Pool and Meuwissen, 2000; Huisman et al., 2002; Schaeffer, 2004; Misztal, 2006; Sun et al., 2017). In a recent study, Sun et al. (2017) used a RR approach with cubic splines in wheat (*Triticum aestivum* L.) to obtain best linear unbiased predictions of secondary traits derived from high-throughput hyperspectral and thermal imaging. Regression coefficients are treated as random effects, and therefore allow the values to vary between individuals. Genomic estimated breeding values (GEBVs) for regression coefficients are obtained through a mixed model, and using simple algebra, GEBVs can be obtained for any time throughout the continuous trait trajectory (Mrode, 2014).

Genomic estimated breeding values represent the summation of all additive genetic effects across the genome for a given individual. Goddard (2009) showed that GEBVs predicted from genomic relationships [e.g., genomic best linear unbiased prediction (BLUP)] are equivalent to those predicted from regression on markers. Given this equivalence, marker effects can be easily calculated from GEBVs and thus genetic inference (e.g., GWAS) can be performed. Although this approach is different from conventional single-marker regression GWAS approaches, it offers several advantages. First, hundreds of thousands of statistical tests are typically run for single-marker regression GWAS; as a result, a stringent *p*-value threshold must be used to limit false discoveries (Hayes, 2013). Thus, loci recovered via single-marker regression GWAS approaches typically account for only a fraction of the total genetic variance for a trait (Yang et al., 2010). Whole-genome BLUP approaches [e.g., single nucleotide polymorphism (SNP)-BLUP or genomic BLUP] assume an infinitesimal model in which all loci make some, albeit small, contribution to the phenotype (Hayes, 2013). Thus, when we consider all markers simultaneously, small-effect QTL are recovered and more genetic variation can be captured than in single-marker regression GWAS (Yang et al., 2010). Best linear unbiased prediction approaches shrink the marker effects toward zero and thus may not be appropriate for simple traits that are regulated by few loci with large effects. However, for complex polygenic traits, these assumptions are reasonable and should yield biologically meaningful results. In the case of RR, GEBVs can be calculated at each time point and can be leveraged to examine the contribution of loci across a trait trajectory or the time axis.

In a recent study, we used a RR approach for genomic prediction of shoot growth trajectories in rice (Campbell et al., 2018). The use of longitudinal phenotypes with RR captured greater genetic variation than the TP approach and significantly improved prediction accuracy. In the current study, we have leveraged GEBVs derived from RR models to examine the genetic architecture of shoot growth through a 20-d period during early vegetative development. We show that this approach can be used for genetic inference of shoot growth trajectories and uncovers persistent as well as time-specific QTL. Furthermore, we show that the RR approach uncovers considerably more associations than a conventional TP analysis.

MATERIALS AND METHODS

High-Throughput Phenotyping

Phenotypic data was collected for 357 diverse rice accessions from the Rice Diversity Panel 1 (Zhao et al., 2011). The plant materials, experimental design, and image processing are described in detail in Campbell et al., (2018). Briefly, 378 lines were phenotyped at the Plant Accelerator, Australian Plant Phenomics Facility, at the University of Adelaide, SA, Australia, from February to April 2016. In this period, three experiments were conducted, where experiment consisted of a partly replicated design with 54 randomly selected lines that had two replicates in each experiment. The plants were grown on greenhouse benches for 10 d after transplanting (DAT) and were loaded on the imaging system and watered to 90% field capacity at 11 DAT.

The plants were imaged daily from 13 to 33 DAT with a visible (red–green–blue) camera (Pilot piA2400–12 gc, Basler, Ahrensburg, Germany) from two side-view angles separated by 90° and a single top view. LemnaGrid software was used to extract plant pixels from the red–green–blue images via a color classification strategy. Noise (i.e., small areas of nonplant pixels) in the image was removed with a series of erosion and dilation steps. Projected shoot area (PSA) was calculated as the sum of the plant area projected in two-dimensional space from each of the three RGB images and was used as a measure of shoot biomass. Previous studies have shown a high correlation between PSA and conventional destructive measures of shoot biomass (Golzarian et al., 2011; Campbell et al., 2015; Neilson et al., 2015; Knecht et al., 2016). A depiction of PSA collected from RGB images is provided as Supplemental Figure S1. Outlier plants at each time point were detected via the 1.5 interquartile range rule. Briefly, the distribution of PSA each day was split into quartiles and the interquartile range was calculated as the difference between the third and first quartiles. Points that were either less than Quartile 1 – 1.5 interquartile range or greater than Quartile 3 + 1.5 interquartile range were considered as outliers. Outliers were plotted and those that exhibited abnormal growth patterns were removed. A total of 2604 plants remained for downstream analyses.

Predicting Genomic Breeding Values

Random Regression

Trajectories for PSA across the 20 time points were modeled with a RR model with Legendre polynomials. The model is the same as was used for genomic prediction in Campbell et al., (2018). The model is described below in the notation of Mrode (2014):

$$\mathrm{PSA}_{tij} = \mu + \sum_{k=0}^{2} \phi(t)_{jtk}\beta_k + \sum_{k=0}^{2} \phi(t)_{jtk} u_{jk} + \sum_{k=0}^{1} \phi(t)_{jtk} s_{ik} + e_{tij}, \quad [1]$$

where PSA_{tij} is PSA on Day t for Line j within Experiment i, β_k is the fixed second-order Legendre polynomial used to model the mean PSA trajectory for all lines, u_{jk} and s_{ik} are the k^{th} random regression coefficients for additive genetic effect and random experiment effects, and e_{tij} is the random residual. The order of b was selected on the basis of visual inspections of the PSA over the 20 d. The random additive genetic effects (u) are modeled by a second-order Legendre polynomial and the experiment effects (s) are modeled by a first-order Legendre polynomial.

In matrix notation, the model is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{Qs} + e, \quad [2]$$

where $\mathbf{y}$ is a vector of PSA over the 20 d and is of order n, where n is the number of observations. $\mathbf{X}$ is an $n \times k_f$ covariable matrix where the number of columns is equal to the order of Legendre polynomials used to model fixed effects (k_f). The matrices $\mathbf{Z}$ and $\mathbf{Q}$ are covariable matrices for the random additive genetic and random experimental effects, respectively. The number of rows for $\mathbf{Z}$ is n and the number of columns corresponds to the number of lines used multiplied by the order of the Legendre polynomial used to fit the additive genetic effect ($q \times k_g = 357 \times 3 = 1071$). The dimension of $\mathbf{Q}$ is $n \times e \times k_s$, where k_s is the order of the Legendre polynomial used to fit the permanent environmental effects and e is the number of experiments. We assume that $\mathbf{u} \sim N(0, \mathbf{G} \otimes \Omega)$, $\mathbf{s} \sim N(0, \mathbf{I} \otimes \mathbf{P})$, and $\mathbf{e} \sim N(0, \mathbf{I} \otimes \mathbf{D})$. Here, $\mathbf{\Omega}$ and $\mathbf{P}$ are the covariance matrices for the RR coefficients for the additive genetic and permanent environmental effects, and $\mathbf{D}$ is a diagonal matrix that allows for heterogeneous variances over the 20 time points.

A genomic relationship matrix ($\mathbf{G}$) was calculated using VanRaden (2008):

$$\mathbf{G} = \frac{\mathbf{W_{sc}} \mathbf{W'_{sc}}}{m}, \quad [3]$$

where $\mathbf{W_{sc}}$ is a centered and scaled $q \times m$ matrix, where m represents 33,674 SNPs. The variance components and genomic BLUPs were obtained via ASReml version 4.0 (Gilmour et al., 2015).

Solving the mixed model equation will give three RR coefficients for each line. With these RR coefficients, GEBVs at each time point can be obtained. For line j, the predicted GEBVs at each time point are given by $GEBV_j = \mathbf{\Phi}_g \hat{u}_j$ (Mrode, 2014). $\mathbf{\Phi}_g$ is the matrix of Legendre polynomials of time covariates. A detailed explanation of the RR model is provided in the Supplemental Methods.

Single Time Point Analysis

The following mixed model approach was used to fit genomic BLUPs at each time point:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{Qs} + \mathbf{e}, \quad [4]$$

The matrices X, Z, and Q correspond to incidence matrices for the fixed, random additive genetic, and random experimental effects, respectively. Moreover, the dimensions for X, Z, and Q are $n \times 1$, $n \times q$, and $n \times e$. We assume the random terms are distributed as follows $\mathbf{u} \sim N(0, \mathbf{G}\sigma_g^2)$, $\mathbf{s} \sim N(0, \mathbf{I}\sigma_s^2)$, and $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$. A genomic

relationship matrix (**G**) was calculated as above and used for predicting breeding values at each time point.

Genome-wide Association Analyses

Estimating Marker Effects from GEBVs

The GEBVs ($\hat{\mathbf{u}}$) can be parameterized as $\hat{\mathbf{u}} = \hat{\beta}\mathbf{W}_{sc}$, where $\mathbf{W}_{sc}$ is a matrix of marker genotypes, as defined above, and $\hat{\beta}$ is a vector of allele substitution effects. $\hat{\beta}$ can be obtained from BLUPs as follows:

$$BLUP(\hat{\beta}) = \mathbf{W}'_{sc}\mathbf{G}^{-1}\left[\mathbf{I}+\mathbf{G}^{-1}\frac{\sigma_e^2}{\sigma_g^2}\right]^{-1}\mathbf{y}, \quad [5]$$

where σ_g^2 and σ_e^2 are the genetic and residual variances, respectively (Morota and Gianola, 2014).

The BLUP of GEBVs is:

$$BLUP(\hat{\mathbf{u}}) = \left[\mathbf{I}+\mathbf{G}^{-1}\frac{\sigma_e^2}{\sigma_g^2}\right]^{-1}\mathbf{y}, \quad [6]$$

The BLUP of marker effects can be obtained via the following linear transformation:

$$BLUP(\hat{\beta}) = \mathbf{W}_{sc}'\mathbf{G}^{-1}BLUP(\hat{\mathbf{u}}). \quad [7]$$

This relationship was leveraged to solve for marker effects from breeding values at each time point for both RR and TP analyses.

Variance of SNP Effects

The variance of marker effects was calculated following the methods outlined by Duarte et al. (2014). Briefly, the variance of the marker effects can be obtained via linear transformation of the variance of GEBVs ($\hat{\mathbf{u}}$) as follows:

$$\begin{aligned}\mathrm{Var}(\hat{\beta}) &= \mathrm{Var}(\mathbf{W}_{sc}'\mathbf{G}^{-1}\hat{\mathbf{u}}) \\ &= \mathbf{W}_{sc}'\mathbf{G}^{-1}\mathrm{Var}(\hat{\mathbf{u}})\mathbf{G}^{-1}\mathbf{W}_{sc}\end{aligned}. \quad [8]$$

The prediction error variance (PEV) of $\hat{\mathbf{u}}$ is:

$$\begin{aligned}\mathrm{PEV}(\hat{\mathbf{u}}) = \mathbf{C}^{22} &= \mathrm{Var}(\mathbf{u}) - \mathrm{Var}(\hat{\mathbf{u}}) \\ &= \mathbf{G}\sigma_g^2 - \mathrm{Var}(\hat{\mathbf{u}})\end{aligned}, \quad [9]$$

where $\mathbf{C}^{22}$ is obtained by inverting the coefficient matrix of the mixed model equation (provided in the Supplemental Methods) and extracting the elements corresponding to additive genetic effects (Henderson, 1984). Thus by rearranging Eq. [9] for $\mathrm{PEV}(\hat{\mathbf{u}})$, the variance of the predicted breeding values is

$$\mathrm{Var}(\hat{\mathbf{u}}) = \mathbf{G}\sigma_g^2 - \mathbf{C}^{22}\sigma_e^2. \quad [10]$$

For the TP approach, $\mathbf{C}^{22}$ is a $q \times q$ matrix, and diagonal elements correspond to the prediction error variance of breeding values. Since the mixed model equation is solved for each time point independently, this procedure can be used to obtain the variance of SNP effects on each day. However for the RR approach, $\mathbf{C}^{22}$ is $(q \times k_g) \times (q \times k_g)$ and represents the prediction error variance for the additive genetic RR coefficients. Therefore, to obtain $\mathrm{Var}(\hat{\mathbf{u}})$ at each time point, we define a new matrix $\mathbf{C}^{22^*}$, namely $(q \times d) \times (q \times d)$, where d is the number of time points (e.g., 20). This is given by:

$$\mathbf{C}^{22^*} = \mathbf{\Phi}_{\mathbf{g}}^{*}\mathbf{C}^{22}\mathbf{\Phi}_{\mathbf{g}}^{*\prime}, \quad [11]$$

where $\mathbf{\Phi}_{\mathbf{g}}^{*}$ is a $(q \times d) \times (k_g \times d)$ block matrix, in which the diagonal submatrices consist of Legendre polynomials at each standardized time interval. This approach is analogous to that described by Mrode (2014) and is described in greater detail in the Supplemental Methods.

Obtaining P-Values for Marker Effects

The SNP effects for SNP_j at time t were divided by their corresponding Var ($\hat{\beta}$) values according to:

$$\mathrm{SNP}_{jt} = \frac{\hat{\beta}}{\sqrt{\mathrm{Var}(\hat{\beta})}}. \quad [12]$$

The p-values for marker effects were calculated as 1 minus the cumulative probability density of the absolute value of SNP_{jt}. This number was subsequently multiplied by 2. This is summarized as follows.

$$p-\mathrm{value}_{SNP_{jt}} = 2(1-\phi(|\mathrm{SNP}_{jt}|)). \quad [13]$$

In line with Zhao et al. (2011), a threshold of 1×10^{-4} was used to define significant loci.

Data Accessibility

The full datasets and all code used in this study are available via GitHub (https://github.com/malachycampbell/Leveraging-RR-GEBVs-for-genomic-inference-of-longitudinal-traits, accessed 15 Apr. 2019) and the Wheat and Rice Center for Heat Resilience website (WRCHR.org).

RESULTS AND DISCUSSION

To identify loci associated with shoot growth trajectories in rice, we used a novel RR approach that allows trait trajectories to be modeled across time points. Shoot growth trajectories were recorded for 357 diverse rice accessions over a period of 20 d during early vegetative growth (13–33 DAT). A RR model was fitted to the shoot growth trajectories, which included a fixed second-order Legendre polynomial, a random second-order Legendre polynomial for the additive genetic effect, a first-order Legendre polynomial for the environmental effect, and heterogeneous residual variances. Genomic estimated breeding values were predicted for each accession at each of the 20 time points as described in Campbell et al. (2018) and were used to estimate marker effects at each time point (Supplemental File S1). The results from the RR were compared with a conventional TP approach in which GEBVs were predicted at each time point via a conventional mixed model and were used to estimate the marker effects.

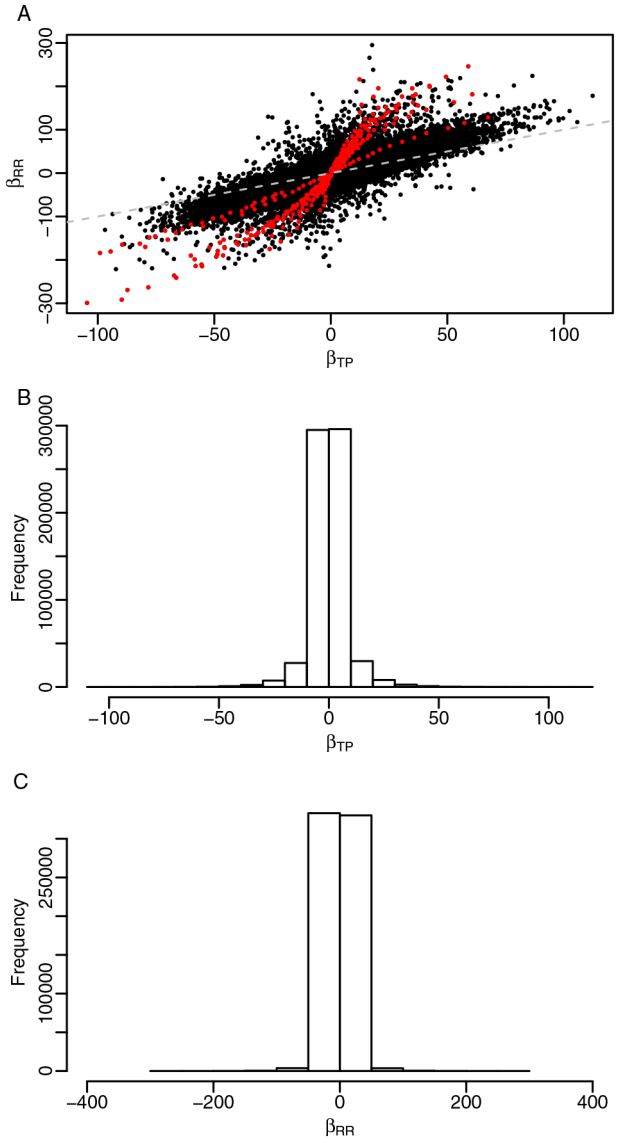

Fig. 1. Correlation and distribution of SNP effects from random regression (RR) and single time point (TP) analysis. (A) Correlation between SNP effects for the RR (β_{RR}) and TP analyses (β_{TP}). Single nucleotide polymorphisms (SNPs) highlighted in red are those that were statistically significant in the RR approach ($p < 1 \times 10^{-4}$). The gray broken lines depict a one-to-one relationship between β_{RR} and β_{TP}. The distribution of SNP effects across all 20 time points from (B) the TP analyses and(C) RR analysis.

Random Regression GWAS Recovers More Significant Associations and Increases Predicted Marker Effect Sizes

With RR models, incorporating the covariance structure of multiple measurements should lead to a more accurate partitioning of phenotypic variation into the genetic and environmental components, and improve genetic inference. To demonstrate the advantages of a longitudinal genetic inference approach over a conventional TP approach, significant marker effects were compared between the RR and TP approaches (Fig. 1). A 131% increase in the number of significant associations ($p < 10^{-4}$) was observed with the RR approach compared with the conventional TP model (Fig. 2). A total of 442 SNPs were found to be significantly associated with shoot growth trajectories at one or more time points in the RR approach, whereas 191 were found in the TP approach. Correlations in SNP effects and $-\log_{10}(p)$-values estimated via the two approaches showed very high agreement ($r = 0.85$ and 0.79, respectively); however, the predicted marker effects ($\hat{\beta}$) obtained with the RR were considerably larger than the TP analysis (Fig. 1). For instance, $\hat{\beta}$ for the RR approach ranged from –299.1 to 295.0 across all days, whereas for the TP approach, $\hat{\beta}$ ranged from –104.6 to 112.3. These differences are evident in the distribution of marker effects pictured in Fig. 1. Manhattan plots for each of the 20 time points is provided as Supplemental Fig. S2, Supplemental Fig.S3, Supplemental Fig. S6, and Supplemental Fig. S7. The corresponding Q–Q plots are provided as Supplemental Fig. S4, Supplemental Fig. S5, Supplemental Fig. S8, and Supplemental Fig. S9. These results indicate that the use of information across all time points with the RR approach improves the ability to detect significant associations and also increases the predicted marker effect sizes compared with a model that uses information at only a single time point.

These results suggest that the inclusion of the time axis for genetic inference will improve the ability to recover significant associations. Several other studies have shown similar improvements in the estimation of variance components and genetic inference via different approaches for longitudinal traits. For instance, De Andrade et al. (2002) showed a longitudinal approach that leveraged pedigree data and systolic blood pressure measurements collected at three time points improved heritability estimates compared with a TP approach. In the context of GWAS, Das et al. (2011) used a novel functional GWAS approach and identified several new variants associated with body mass index collected at four time points in humans. Moreover, by using simulated data, the authors showed that the statistical power exceeded 0.8, with a false positive rate of less than 0.1 for sample sizes greater than 1000. Similar gains for GWAS have been demonstrated in plants, animals, and humans (Xu et al., 2014; Campbell et al., 2015; Yi et al., 2015; Lund et al., 2008).

Random Regression GWAS Reveals the Dynamic Genetic Architecture of Shoot Growth Responses in Rice

For many traits, such as growth, genetic effects are expected to vary across time. These temporal genetic effects can be effectively captured by a RR approach. To examine the dynamic genetic architecture of shoot growth trajectories, significant SNPs from the RR approach were selected and those within a 200-kb window were merged to a single QTL. The 200-kb window that we used corresponded to the average linkage disequilibrium in rice (Zhao et al., 2011; Huang et al., 2010). For the RR approach, 26 significant QTL were detected at one or more time points; for the TP approach, only 15 significant QTL were detected (Supplemental File S1).

To dissect the dynamic genetic architecture of shoot growth in rice, significant QTL were classified into four

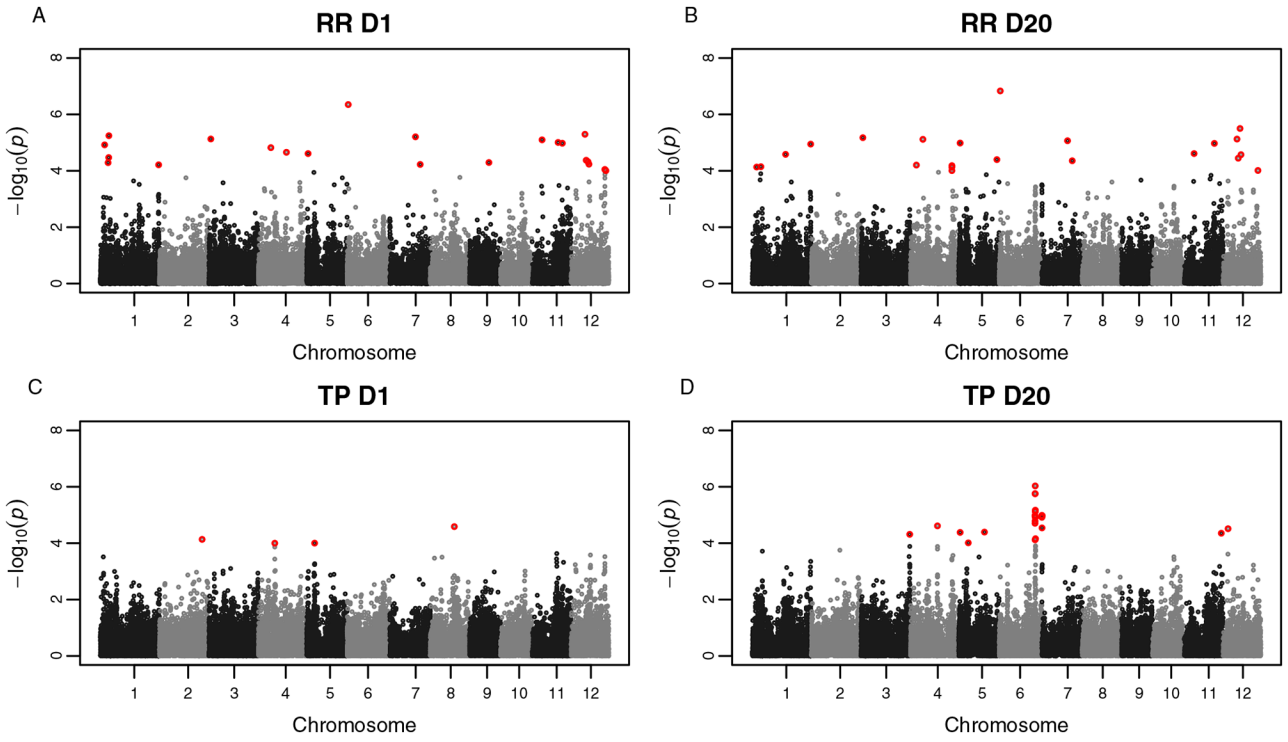

Fig. 2. Manhattan plots for random regression (RR) and single time point (TP) approaches on Days 1 and 20. (A,B) Manhattan plots for the RR approach on Days 1 and 20, respectively. (C,D) Manhattan plots for the TP approach on Days 1 and 20, respectively. $-\log_{10}(p)$ is shown on the *y*-axis. Statistically significant single nucleotide polymorphisms are highlighted in red ($p < 1 \times 10^{-4}$).

categories: persistent QTL (QTL detected at all 20 time points), long-duration QTL (those with significant associations at more than 12 but fewer than 20 time points), mid-duration QTL (QTL with associations at 6 to 12 time points), and short-duration QTL (those with associations at fewer than six time points). Of these categories, far more persistent QTL were detected, with a total of 13 observed at all 20 time points (Fig. 3). Short-duration QTL also showed the smallest number of significant QTL (two); five and six QTL were detected for long and mid-duration QTL, respectively. The frequencies of significant QTLs for each category were calculated at each time point and plotted as a function of time (Supplemental Fig. S10). The majority of long-duration QTL were detected toward the end of the experiment (Day 8 onward), whereas short-duration QTL were detected only from Days 1 to 4. Mid-duration QTL were detected at all time points. The *p*-values across all 20 time points for all significant QTL are provided in Supplemental Fig. S2 and Supplemental Fig. S3. Collectively, these results indicate that the shoot growth is regulated by numerous loci that have both transient and persistent effects throughout early vegetative growth.

The importance of time-specific QTL has been demonstrated in both plants and animals (Moore et al., 2013; Bac-Molenaar et al., 2015; Campbell et al., 2015, 2017). For instance, by using a TP linkage mapping approach, Moore et al. (2013) showed several time-specific QTL associated with root gravitropic responses in *Arabidopsis thaliana* (L.) Heynh. Moreover, many of these QTL harbored candidate genes known to influence root growth, root gravitropism, or hormone transport and signaling. Bac-Molenaar et al. (2015) collected rosette growth trajectories over a period of 20 d for a diverse panel of 324 *A. thaliana* accessions. A growth function was fitted for each accession and the model parameters were used for GWAS. The authors showed that many associations detected for model parameters were also detected at a few time points via a TP GWAS approach. Although few longitudinal studies have been performed in rice, our previous studies have identified time-specific QTL for shoot growth and salt stress responses (Campbell et al., 2015, 2017).

CONCLUSION

New phenotyping platforms have provided the plant science community with a suite of tools to collect high-dimensional temporal phenotypic data. With these temporal datasets, quantitative genetic approaches that can leverage the covariance across time points must be fully utilized to realize the potential of these data for genomic prediction and genetic inference. In this study, we show that the RR framework that has been extensively developed in animal breeding can be extended to genetic inference in plants. This approach can be used effectively to identify QTL with time-specific effects. To date, this is the first application of RR models for genetic inference of a longitudinal trait in a major crop.

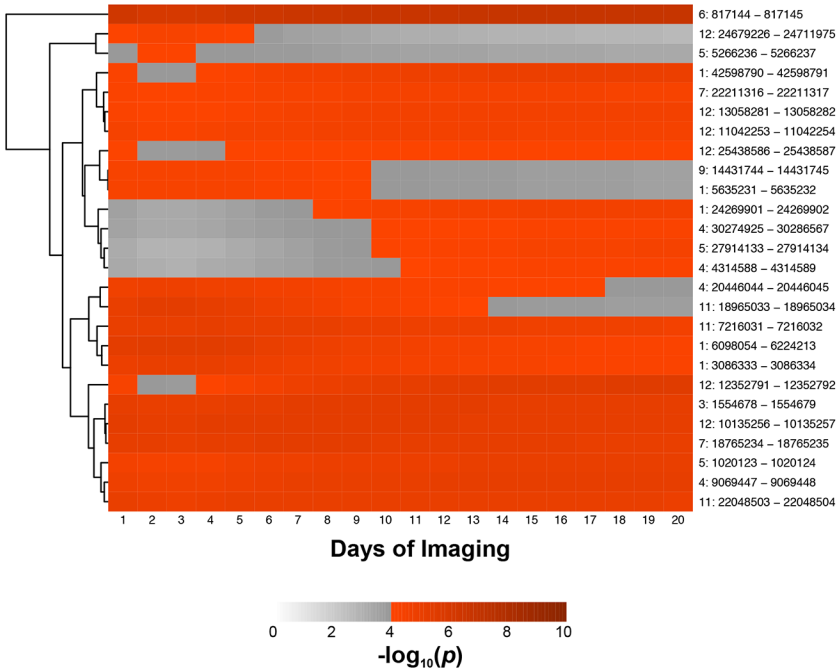

Fig. 3. Heatmap showing time-specific quantitative trait loci (QTL). A subset of significant QTL identified via the random regression (RR) approach are pictured. The *x*-axis indicates the days of imaging and the *y*-axis shows the chromosome and intervals for the QTL. For each of the QTL, the most significant single nucleotide polymorphism within the interval at each time point were selected. The gray color scale indicates a nonsignificant association; the red color scale indicates a statistically significant association ($p < 1 \times 10^{-4}$).

Supplemental Materials

Supplemental File S1. Results of the GWAS. *P*-values are provided for each all SNPs at each time point for the RR and TP approaches.

Supplemental Fig. S1. Visual depiction of the process of assessing shoot biomass from PSA.

Supplemental Fig. S2. Manhattan plots for the RR approach at Days 1 to 10.

Supplemental Fig. S3. Manhattan plots for the RR approach at Days 10 to 20.

Supplemental Fig. S4. Q–Q plots for the RR approach at Days 1 to 10.

Supplemental Fig. S5. Q–Q plots for the RR approach at Days 11 to 20.

Supplemental Fig. S6. Manhattan plots for the TP approach at Days 1 to 10.

Supplemental Fig. S7. Manhattan plots for the TP approach at Days 11 to 20.

Supplemental Fig. S8. Q–Q plots for the TP approach at Days 1 to 10.

Supplemental Fig. S9. Q–Q plots for the TP approach at Days 11 to 20.

Supplemental Fig. S10. Frequency of time-specific QTL, where frequency was determined by dividing the number of QTL detected at time *t* by the total number of QTL for a given class.

Supplemental Methods. Detailed explanation of the RR approach and how exact *P*-values were obtained for marker effects from RR-derived GEBVs.

Conflict of Interest Disclosure

The authors declare that there is no conflict of interest.

Author Contributions

The study was conceived by HW, GM, and MC; phenotyping was performed by MC and HW; MC, MM, and GM performed all analyses; MC wrote the manuscript; and editorial comments were provided by MM, HW, and GM.

ACKNOWLEDGMENTS

Funding for this research was provided by the National Science Foundation (United States) through Award No. 1238125 to HW, and Award No. 1736192 to HW and GM.

REFERENCES

Araus, J.L., S.C. Kefauver, M. Zaman-Allah, M.S. Olsen, and J.E. Cairns. 2018. Translating high-throughput phenotyping into genetic gain. Trends Plant Sci. 23(5):451–466. doi:10.1016/j.tplants.2018.02.001

Bac-Molenaar, J.A, D. Vreugdenhil, C. Granier, and J.J.B. Keurentjes. 2015. Genome-wide association mapping of growth dynamics detects time specific and general quantitative trait loci. J. Exp. Bot. 66(18):5567–5580. doi:10.1093/jxb/erv176

Campbell, M.T, Q. Du, K. Liu, C.J. Brien, B. Berger, C. Zhang, et al. 2017. A comprehensive image-based phenomic analysis reveals the complex genetic architecture of shoot growth dynamics in rice. Plant Genome 10(2). doi:10.3835/plantgenome2016.07.0064

Campbell, M.T., A.C. Knecht, B. Berger, C.J. Brien, D. Wang, and H. Walia. 2015. Integrating image-based phenomics and association analysis to dissect the genetic architecture of temporal salinity responses in rice. Plant Physiol. 168(4):1476–1789. doi:10.1104/pp.15.00450

Campbell, M.T., H. Walia, and G. Morota. 2018. Utilizing random regression models for genomic prediction of a longitudinal trait derived from high-throughput phenotyping. Plant Direct 2(9):e00080.

Das, K., J. Li, Z. Wang, C. Tong, G. Fu, Y. Li, et al. 2011. A dynamic model for genome-wide association studies. Hum Genet 129(6):629–639. doi:10.1007/s00439-011-0960-6

De Andrade, M., R. Guéguen, S. Visvikis, C. Sass, G. Siest, and C.I. Amos. 2002. Extension of variance components approach to incorporate temporal trends and longitudinal pedigree data analysis. Genet. Epidemi. 22(3):221–232. doi:10.1002/gepi.01118

Duarte, J.L.G, R.J.C. Cantet, R.O. Bates, C.W. Ernst, N.E. Raney, and J.P. Steibel. 2014. Rapid screening for phenotype–genotype associations by linear transformations of genomic evaluations. BMC Bioinformatics 15(1): 246.

Fraas, S., and H. Lüthen. 2015. Novel imaging-based phenotyping strategies for dissecting crosstalk in plant development. J. Exp. Bot. 66(16):4947–4955. doi:10.1093/jxb/erv265

Gilmour, A.R., B.J. Gogel, B.R. Cullis, S.J. Welham, and R. Thompson. 2015. ASReml user guide release 4.1 stsuctural Specification. VSN International Ltd.: Hemel Hempstead.

Goddard, M. 2009. Genomic selection: Prediction of accuracy and maximisation of long term response. Genetica 136(2):245–257. doi:10.1007/s10709-008-9308-0

Golzarian, M.R., R.A. Frick, K. Rajendran, B. Berger, S. Roy, M. Tester, et al. 2011. Accurate inference of shoot biomass from high-throughput images of cereal plants. Plant Methods 7(1):2. doi:10.1186/1746-4811-7-2

Hayes, B. 2013. Overview of statistical methods for genome-wide association studies (GWAS) In C. Gondro, J. van der Werf, B. Hayes, editors, Genome-wide association studies and genomic prediction.. Springer, Basel, Switzerland. p. 149–169.

Henderson, C.R. 1984. Applications of linear models in animal breeding. Univ. of Guelph: Guelph.

Huang, X., T. Sang, Q. Zhao, Q. Feng, Y. Zhao, C. Li, et al. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. Genet. 42(11):961–967. doi:10.1038/ng.695

Huisman, A.E., R.F. Veerkamp, and J.A.M. Van Arendonk. 2002. Genetic parameters for various random regression models to describe the weight data of pigs. J. Anim. Sci. 80(3):575–582. doi:10.2527/2002.803575x

Kirkpatrick, M., W.G. Hill, and R. Thompson. 1994. Estimating the covariance structure of traits during growth and ageing, illustrated with lactation in dairy cattle. Genet. Res. (Camb.) 64(1): 57–69. doi:10.1017/S0016672300032559

Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. Genetics 124(4):979–993.

Knecht, A.C., M.T. Campbell, A. Caprez, D.R. Swanson, and H. Walia. 2016. Image Harvest: An open-source platform for high-throughput plant image processing and analysis. J. Exp. Bot. 67(11): 3587–3599.

Lund, M.S., P. Sorensen, P. Madsen, and F. Jaffrézic. 2008. Detection and modelling of time-dependent QTL in animal populations. Genet. Sel. Evol. 40(2):177. doi:10.1186/1297-9686-40-2-177

Meyer, K. 1998. Estimating covariance functions for longitudinal data using a random regression model. Genet Sel Evol 30(3): 221. doi:10.1186/1297-9686-30-3-221

Misztal, I. 2006. Properties of random regression models using linear splines. J. Anim. Breed. Genet. 123(2):74–80. doi:10.1111/j.1439-0388.2006.00582.x

Moore, C.R., L.S. Johnson, I. Kwak, M. Livny, K.W. Broman, and E.P Spalding. 2013. High-throughput computer vision introduces the time axis to a quantitative trait map of a plant growth response. Genetics 195(3):1077–1086. doi:10.1534/genetics.113.153346

Morota, G., and D. Gianola. 2014. Kernel-based whole-genome prediction of complex traits: A review. Front. Genet. 5:363. doi:10.3389/fgene.2014.00363

Mrode, R.A. 2014. Linear models for the prediction of animal breeding values. CABI, Wallingford, UK.

Neilson, E.H., A.M. Edwards, C.K. Blomstedt, B. Berger, B. Lindberg Møller, and R.M. Gleadow. 2015. Utilization of a high-throughput shoot imaging system to examine the dynamic phenotypic responses of a C4 cereal crop plant to nitrogen and water deficiency over time. J Exp Bot 66(7):1817–1832. doi:10.1093/jxb/eru526

Pool, M.H., and T.H.E. Meuwissen. 2000. Reduction of the number of parameters needed for a polynomial random regression test day model. Livest. Prod. Sci. 64(2–3): 133–145. doi:10.1016/S0301-6226(99)00166-9

Schaeffer, L.R. and J.C.M. Dekkers. 1994. Random regressions in animal models for test-day production in dairy cattle. In C. Smith, editor, Proceedings of the 5th World Congress of Genetics Applied Livestock Production, Guelph, ON, Canada. 7–12 Aug. 1994. Dept. of Animal & Poultry Science, Univ. of Guelph, Guelph, ON, Canada. p. 443–446.

Schaeffer LR,. Random regressions in animal models for test-day production in dairy cattle. Proceedings of the 5th World Congress on Genetics Applied to Livestock Production;. 1994. pp. 443–446

Schaeffer, L.R. 2004. Application of random regression models in animal breeding. Livest. Prod. Sci. 86(1–3): 35–45. doi:10.1016/S0301-6226(03)00151-9

Shakoor, N., S. Lee, and T.C. Mockler. 2017. High throughput phenotyping to accelerate crop breeding and monitoring of diseases in the field. Curr. Opin. Plant Biol. 38: 184–192. doi:10.1016/j.pbi.2017.05.006

Simko, I., J.A. Jimenez-Berni, and X.R.R Sirault. 2016. Phenomic approaches and tools for phytopathologists. Phytopathology 107(1): 6–17.

Strabel, T., and I. Misztal. 1999. Genetic parameters for first and second lactation milk yields of Polish black and white cattle with random regression test-day models. J. Dairy Sci. 82(12): 2805–2810. doi:10.3168/jds.S0022-0302(99)75538-4

Sun, J., J.E. Rutkoski, J.A. Poland, J. Crossa, J. Jannink, and M.E. Sorrells. 2017. Multitrait, random regression, or simple repeatability model in high-throughput phenotyping data improve genomic prediction for wheat grain yield. Plant Genome 10(2). doi:10.3835/plantgenome2016.11.0111

Tardieu, F., L. Cabrera-Bosquet, T. Pridmore, and M. Bennett. 2017. Plant phenomics, from sensors to knowledge. Curr. Biol. 27(15): R770–R783. doi:10.1016/j.cub.2017.05.055

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci 91(11): 4414–4423. doi:10.3168/jds.2007-0980

White, I.M.S., R. Thompson, and S. Brotherstone. 1999. Genetic and environmental smoothing of lactation curves with cubic splines. J. Dairy Sci. (3) 632–638. doi:10.3168/jds.S0022-0302(99)75277-X

Würschum, T., W. Liu, L. Busemeyer, M.R. Tucker, J.C. Reif, E.A. Weissmann, et al. 2014. Mapping dynamic QTL for plant height in triticale. BMC Genetics 15(1): 59. doi:10.1186/1471-2156-15-59

Xu, Z., Xiaotong S., Wei P., and Alzheimer's Disease Neuroimaging Initiative. 2014. Longitudinal analysis is more powerful than cross-sectional analysis in detecting genetic association with neuroimaging phenotypes. PloS One 9(8): E102312.

Yang, G., J. Liu, C. Zhao, Z. Li, Y. Huang, H. Yu, et al. 2017. Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. Front Plant Sci 8:1111. doi:10.3389/fpls.2017.01111

Yang, J., B. Benyamin, B.P. McEvoy, S. Gordon, A.K. Henders, D.R. Nyholt, et al. 2010. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42(7):565. doi:10.1038/ng.608

Yi, G., M. Shen, J. Yuan, C. Sun, Z. Duan, L. Qu, et al. 2015. Genome-wide association study dissects genetic architecture underlying longitudinal egg weights in chickens. BMC Genomics 16(1): 746. doi:10.1186/s12864-015-1945-y

Zhao, K., C. Tung, G.C. Eizenga, M.H. Wright, M.L. Ali, A.H. Price, et al. 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nat. Commun. 2: 467.