

# A Comprehensive Image-based Phenomic Analysis Reveals the Complex Genetic Architecture of Shoot Growth Dynamics in Rice (*Oryza sativa*)

Malachy T. Campbell, Qian Du, Kan Liu, Chris J. Brien, Bettina Berger, Chi Zhang, and Harkamal Walia\*

## Abstract

Early vigor is an important trait for many rice (*Oryza sativa* L.)-growing environments. However, genetic characterization and improvement for early vigor is hindered by the temporal nature of the trait and strong genotype  $\times$  environment effects. We explored the genetic architecture of shoot growth dynamics during the early and active tillering stages by applying a functional modeling and genomewide association (GWAS) mapping approach on a diversity panel of  $\sim$ 360 rice accessions. Multiple loci with small effects on shoot growth trajectory were identified, indicating a complex polygenic architecture. Natural variation for shoot growth dynamics was assessed in a subset of 31 accessions using RNA sequencing and hormone quantification. These analyses yielded a gibberellic acid (GA) catabolic gene, *OsGA2ox7*, which could influence GA levels to regulate vigor in the early tillering stage. Given the complex genetic architecture of shoot growth dynamics, the potential of genomic selection (GS) for improving early vigor was explored using all 36,901 single-nucleotide polymorphisms (SNPs) as well as several subsets of the most significant SNPs from GWAS. Shoot growth trajectories could be predicted with reasonable accuracy using the 50 most significant SNPs from GWAS (0.37–0.53); however, the accuracy of prediction was improved by including more markers, which indicates that GS may be an effective strategy for improving shoot growth dynamics during the vegetative growth stage. This study provides insights into the complex genetic architecture and molecular mechanisms underlying early shoot growth dynamics and provides a foundation for improving this complex trait in rice.

## Core Ideas

- Functional mapping uncovers the genetic architecture of shoot growth dynamics.
- Gibberellic acid is an underlying component for natural variation for shoot growth dynamics in rice.
- Genomic prediction is effective for improving early growth dynamics.

**E**ARLY VIGOR, defined as a plant's ability to accumulate shoot biomass rapidly during early developmental stages, is critical for stand establishment, resource acquisition, and, ultimately, yield. The rapid emergence of leaves leads to early canopy closure, which reduces soil evaporation, thereby improving seasonal water use efficiency and conserving water for later vegetative growth and grain production. In rice, early vigor is a particularly important trait for regions where rice is direct seeded (Mahender et al., 2015). As the cost of labor rises, a shift from the labor-intensive practice of transplanted rice to direct-seeded rice is the expected solution to solve this problem (Mahender et al., 2015).

M.T. Campbell and H. Walia, Dep. of Agronomy and Horticulture, Univ. of Nebraska, Lincoln, NE 68583; Q. Du, K. Liu, and C. Zhang, School of Biological Sciences, Univ. of Nebraska, Lincoln, NE 68583; B. Berger, Australian Plant Phenomics Facility, Univ. of Adelaide, Urrbrae, SA 5064, Australia; C.J. Brien, Phenomics and Bioinformatics Research Centre, Univ. of South Australia, Adelaide, SA 5001, Australia. Received 9 July 2016. Accepted 16 Nov. 2016. \*Corresponding author (hwalia2@unl.edu).

**Abbreviations:** DAT, d after transplanting; GA, gibberellic acid; GEBV, genomic estimated breeding values; GS, genomic selection; GWAS, genomewide association; IH, ImageHarvest; LD, linkage disequilibrium;  $M_0$ , projected shoot area at the start of imaging;  $M_{\text{eff}}$ , effective number of tests; PC, principal component; PSA, projected shoot area; QTL, quantitative trait locus; SNP, single-nucleotide polymorphism.

Published in Plant Genome  
Volume 10. doi: 10.3835/plantgenome2016.07.0064

© Crop Science Society of America  
5585 Guilford Rd., Madison, WI 53711 USA  
This is an open access article distributed under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Several studies have examined seedling vigor in rice and elucidated the underlying genetic basis using conventional phenotyping strategies under field and greenhouse conditions (Redoña and Mackill, 1996; Lu et al., 2007; Cairns et al., 2009; Rebolledo et al., 2012a; 2012b, 2015; Liu et al., 2014). In a recent study by Rebolledo *et al* (2015), multiple vigor-related traits such as plant morphology and nonstructural carbohydrates were quantified in a rice diversity panel of 123 *japonica* varieties (Rebolledo et al., 2015). The authors integrated multiple phenotypic metrics in a functional–structural plant model, called Ecomeristem, and performed GWAS mapping using phenotypic metrics and model parameters as trait values (Luquet et al., 2012; Rebolledo et al., 2015). Such multitrait approaches provide a more comprehensive understanding of the biochemical and genetic basis of early vigor than conventional single trait approaches.

Early vigor is a function of time. The timing of developmental switches that initiate tiller formation and rapid exponential growth are a crucial component of this trait. However, despite this temporal dimension, most studies have assessed the genetic basis of early vigor at one or a few discrete time points (Redoña and Mackill, 1996; Lu et al., 2007; Cairns et al., 2009; Rebolledo et al., 2012a; 2012b, 2015; Liu et al., 2014). Such approaches are overly simplistic and may only provide a snapshot of the genetic determinants that cumulatively influence the final biomass. However, sampling for biomass at high frequencies over a developmental window for mapping populations using conventional destructive phenotyping approaches would require tens to hundreds of thousands of plants and be highly labor-intensive. With the advent of high-throughput image-based phenomic platforms, plants can be phenotyped nondestructively more frequently throughout their growth cycle to examine the temporal dynamics of physiological and morphological traits (Berger et al., 2010; Goltzarian et al., 2011; Busemeyer et al., 2013; Topp et al., 2013; Moore et al., 2013; Würschum et al., 2014; Hairmansis et al., 2014; Slovak et al., 2014; Yang et al., 2014; Honsdorf et al., 2014; Chen et al., 2014; Bac-Molenaar et al., 2015).

Mathematical equations that describe a developmental or physiological process can be applied to this high-resolution temporal data to describe temporal growth trajectories using mathematical parameters. Several models, such as logistic, exponential, and power-law functions, have been used to describe plant growth (Paine et al., 2012). These approaches enhance the temporal resolution of phenotyping and, when combined with association or linkage mapping, improve the power to detect genetic associations for complex traits compared with traditional cross-sectional approaches (Wu and Lin, 2006; Xu et al., 2014; Campbell et al., 2015). However, despite the recent advances in phenotyping technologies, the genetic basis of early growth dynamics in rice or other cereals remains largely unexplored.

Multiple and sometimes uncorrelated phenotypes determine the rate and extent of vegetative growth in

crops. We hypothesize that capturing growth dynamics at a higher temporal resolution can help elucidate the genetic basis of this trait. To this end, we sought to examine the genetic architecture of temporal shoot growth dynamics during the early and active tillering stages (8–27 d after transplanting (DAT) and 19–41 DAT) in rice. A panel of ~360 diverse rice accessions was phenotyped using a nondestructive image-based platform and temporal trends in shoot growth were modeled with a power-law function (Zhao et al., 2011). We provide insights into the genetic basis of shoot growth by using GWAS analysis. The underlying molecular mechanisms were explored using RNA sequencing on a subset of the diversity panel during the early tillering stage. Genomic selection of the model parameters and daily estimates of shoot biomass suggest that GS may be an effective strategy for improving early vigor in rice.

## Materials and Methods

### Plant Materials and Genotyping

A rice diversity panel consisting of 413 accessions was obtained from the USDA-ARS Dale Bumpers Rice Research Center and purified through single seed descent before they were phenotyped (Zhao et al., 2011; Famoso et al., 2011; Eizenga et al., 2014). After removing accessions with low seed or poor seed quality, a subset of 360 and 361 accessions were phenotyped during the early and late tillering stages, respectively. All accessions of the rice diversity panel were genotyped using a 44-k SNP array (Zhao et al., 2011; <https://ricediversity.org/data/sets/44kgwas/>, accessed 18 Apr. 2016). The genotypic dataset consisted of 33,901 markers for 360 accessions. Missing markers were imputed using Beagle with 20 iterations (Browning and Browning, 2016).

### Greenhouse Growth Conditions and Phenotyping

Two experiments were conducted at the Plant Accelerator, Australian Plant Phenomics Facility, at the University of Adelaide, SA, Australia. The first experiment consisted of 361 accessions and was repeated three times from August to November 2013. The experiment consisted of two smarthouses that were used consecutively for three periods, with each period forming a block. In each smarthouse, 216 pots were positioned across 24 lanes. A partially replicated design was used for each period. The plants were phenotyped from 8 to 27 DAT. A complete description of the experimental design is provided in Campbell et al. (2015). The second experiment was replicated three times from September to December 2014. The greenhouse conditions and experimental design were nearly identical to those described above. However, only 360 accessions were used because of seed availability. The plants were phenotyped from 19 to 41 DAT. For each experiment, the plants were imaged daily using a visible and red–green–blue camera (Basler Pilot piA2400–12 gc, Ahrensburg, Germany) from two side-view angles separated by 90° and a single top view.

A total of 142,671 images were generated for the 2013 experiment and 152,997 images were generated from the 2014 experiment.

## Image Processing

All 295,668 images were processed with ImageHarvest (IH) (Knecht et al., 2016). Two processing pipelines were developed to extract plant pixels from the side-view and top-view images. Briefly, the side-view processing pipeline consists of three major steps: (i) preprocessing, (ii) thresholding, and (iii) removal of the pot and carrier from the image. Preprocessing smoothed the image, providing a more uniform background, and was achieved by using the gaussianBlur function in IH. The image was converted to grayscale and the adaptiveThreshold function was used to remove the majority of the nonplant pixels from the image. AdaptiveThreshold separates the image into smaller windows and, in each window, a threshold is applied and pixels are removed based on their intensity. A region of interest was defined around the pot and mean shift segmentation was applied using the meanshift function in IH to extract the plant pixels from this region.

Top-view images were processed using a slightly different pipeline. First, to create a more uniform coloring in the image, the color values were normalized using the normalizeByIntensity function in IH. For each pixel, the normalizeByIntensity function rescales the color value for each channel (red, green, and blue) based on its intensity, which is defined as the sum of all color channels (red plus green plus blue) for that pixel. Next, the image was segmented using the meanshift function. Plant pixels were extracted from the image if the value for the green channel was greater than the value for the red channel. Finally, a morphological closing operation consisting of a dilation and erosion step was used to fill “holes” in the plant that may have been caused by the processing steps. This was achieved using the morphology function in IH.

## Statistical Analysis of Projected Shoot Area

To quantify shoot growth at each time point, the plant pixels extracted from each of the three images (Side View 1, Side View 2, and top view) were summed. This metric, here defined as projected shoot area (PSA) has been shown to be a reliable estimate of shoot biomass (Berger et al., 2010; Campbell et al., 2015). For each period (i.e., August–November 2013 and September–December 2014), PSA was combined across experiments and a linear model was fitted to calculate the adjusted means for each accession using the lsmeans function in the LSMeans package in R (R Core Team, 2014; <https://cran.r-project.org/web/packages/lsmeans/index.html>, accessed 18 Apr. 2017). The experiment and treatment were considered to be fixed effects and accession as a random effect in the linear model. The adjusted means were used for a cross-sectional GWAS of PSA and functional modeling.

## Functional Modeling of Temporal Trends in PSA

For each accession, the adjusted mean for PSA was modeled using the following power function (Paine et al., 2012):

$$M_t = \left[ M_0^{1-\beta} + rt(1-\beta) \right]^{\frac{1}{1-\beta}}, \quad [1]$$

where  $M_t$  is the PSA at time  $t$ ,  $M_0$  is the PSA at the start of imaging, and  $r$  and  $\beta$  are parameters controlling the growth rate. The exponent  $\beta$  allows the relative growth rate to slow as biomass increases (Paine et al., 2012). The choice of model parameters is not trivial in nonlinear regression and is often based on visual inspection or *a priori* knowledge regarding the nature of the data. Therefore, to reduce the labor burden of fitting hundreds of models, a three-step approach was used to obtain optimal estimates of the parameters  $\beta$ ,  $r$ , and  $M_0$  using the nls2, nlrmrt, and nlme packages in R (Grothendieck, 2013; Nash, 2013; Pinheiro et al., 2015). First, rough starting estimates were obtained by using nls2 with the brute-force algorithm with 10,000 iterations. Nls2 performs a grid search to obtain rough estimates of the model parameters using a range of parameter values supplied by the user. The best estimates for each of the model parameters were supplied as starting values to the nlxb function in nlrmrt. Nlrmrt can be more robust in finding solutions than nls, especially when the data to be modeled have small or zero residuals. Finally, the results obtained from nlxb were used to fit the model in nls.

## Genomewide Association Analysis of PSA

To identify the genomic regions associated with PSA at each time point, a mixed model that accounted for kinship and population structure was used for GWAS using the EMMA algorithm (Kang et al., 2008). The mixed linear model can be summarized as  $\mathbf{y} = \mathbf{X}\beta + \mathbf{C}\gamma + \mathbf{Z}\mathbf{u} + \mathbf{e}$ , where  $\mathbf{y}$  is a vector of phenotype,  $\beta$  is a vector of fixed marker effects,  $\gamma$  is a vector of principal component (PC) effects fitted to account for population structure,  $\mathbf{u}$  is a vector of polygenic effects caused by relatedness,  $\mathbf{e}$  is a vector of residuals,  $\mathbf{X}$  is a marker incidence matrix relating  $\beta$  to  $\mathbf{y}$ ,  $\mathbf{C}$  is an incidence matrix relating  $\gamma$  to  $\mathbf{y}$  that consists of the first four PCs resulting from a PC analysis, and  $\mathbf{Z}$  is the corresponding design matrix relating  $\mathbf{y}$  to  $\mathbf{u}$ . It is assumed that  $\mathbf{u} \sim \text{MVN}(0, \mathbf{K}\sigma_u^2)$  and  $\mathbf{e} \sim \text{MVN}(0, I\sigma_e^2)$ , where  $\mathbf{K}$  is a kinship matrix estimated using an allele-sharing matrix calculated from the SNP data. Markers with a minor allele frequency less than 5% were excluded from the analysis. The parameter  $\beta$  during the active tillering stage was logarithmically transformed to provide a normal distribution prior to association mapping.

To test for genetic associations with the model parameters  $M_0$ ,  $\beta$ , and  $r$  in the early and active tillering stages, both univariate and multivariate approaches were implemented using Genomewide Efficient Mixed Model Association (Zhou and Stephens, 2014). To identify SNPs associated with the two power-law parameters describing growth rate,  $\beta$  and  $r$ , a bivariate mixed model was

used. The mixed model that was implemented is similar to that described above. Population and relatedness were accounted for by using the top four PCs and a centered genetic relatedness matrix. For each SNP, a likelihood ratio test was used to test the alternative hypothesis that  $\beta \neq 0$ , against the null hypothesis that  $\beta = 0$ .

To account for multiple testing, the Šidák correction using the effective number of tests ( $M_{\text{eff}}$ ) was applied (Li and Ji, 2005). Briefly, the effective number of independent tests ( $M_{\text{eff}}$ ) was determined via eigenvalue decomposition of the correlation matrix among 34,960 SNPs (MAF < 0.05 for 360 accessions). The test criteria were then adjusted by using the  $M_{\text{eff}}$  with the Šidák (1967) correction below:

$$\alpha_p = 1 - (1 - \alpha_e)^{1/M_{\text{eff}}} \quad [2]$$

where  $\alpha_p$  is the comparison-wise error rate and  $\alpha_e$  is the experiment-wise error rate ( $\alpha = 0.05$ ). Using this approach,  $M_{\text{eff}}$  was determined to be 2203, which is the number of eigenvalues necessary to explain 99% of variation, and  $\alpha_p$  was  $2.33 \times 10^{-5}$ .

Single-nucleotide polymorphisms within 200 kb, which represents the estimated linkage disequilibrium (LD) decay in this population, were considered as a single quantitative trait locus (QTL) (Zhao et al., 2011; Famoso et al., 2011). All genes within 200 kb of significant SNPs were considered as potential candidate genes. ANOVA was used to estimate the proportion of phenotypic variance accounted for by significant SNPs after adjusting for population structure effects. ANOVA was used to compare the linear models  $\mathbf{y} = X\beta + C\gamma + e$  and  $y = C\gamma + e$ , where  $\mathbf{y}$  is a vector of phenotype,  $\beta$  is the SNP effect,  $\gamma$  is a vector of PC effects,  $e$  is a vector of residuals,  $X$  is the SNP genotype, and  $C$  is a matrix of the first four PCs.

## Genomic Selection

Genomic selection was performed using ridge regression best linear unbiased prediction implemented in the rrBLUP package in R (Endelman, 2011). Prediction was performed using all 36,901 SNPs as well as various subsets of the most significant associations from GWAS (GWAS-informed prediction). The inclusion of SNPs was based on GWAS performed in the training population for each fold and replicate. The prediction accuracy was assessed using a fivefold cross-validation. Accessions were randomly assigned to five subsets, with four subsets (288 accessions) used as a training population and the remaining 72 accessions used for validation. To assess prediction accuracy, genomic estimated breeding values (GEBVs) for the 72 accessions were calculated using the marker effects determined from the training population and were correlated to the observed phenotypes for those accessions. Genomewide association was performed by using the GWAS function in the rrBLUP package with the P3D option using the top four PCs being used to account for population structure and the realized additive relationship matrix, calculated with the A.mat function, being

used to account for cryptic relatedness between accessions (Zhang et al., 2010; Endelman, 2011). Single-nucleotide polymorphisms were ranked on the basis of the  $-\log_{10}(p)$  values and the top 50, 100, 200, 500, 1000, 2500, 5000, 7500, 10,000, 15,000, 20,000, 25,000, and 30,000 markers were selected for prediction analysis. The predictive ability was calculated as the correlation between the GEBV of the validation population and the observed values for the validation population. Twenty iterations of the fivefold cross-validation were performed for each trait.

## Hormone Quantification

To quantify GA levels, shoot tissue was collected from 12 accessions at 10 DAT (Supplemental Table S1). This time point was selected to reflect the start of imaging during the early tillering stage experiment. The plants were cultured in a growth chamber with temperatures maintained at 28°C and 25°C during day and night respectively and 60% relative humidity. Lighting was maintained at 800  $\mu\text{mol m}^{-2} \text{s}^{-1}$  using high-pressure sodium lights (Agrolite XT-ED25, Phillips, Somerset, NJ). Seeds were germinated in half-strength Murashige and Skoog media (0.8% agar) (Murashige and Skoog, 1962) and transplanted to pots filled with Turface (Profile Products LLC, Buffalo Grove, IL) 4 d after sowing. At 4 DAT, a half-strength Yoshida solution was provided and pH was maintained at 5.8 (Yoshida et al., 1976). Two plants were grown in each pot. Shoot tissue was from two plants was pooled, flash frozen, and ground in liquid N. Hormones were extracted from 250 mg tissue (fresh weight) following a published extraction protocol and was quantified with high-performance liquid chromatography–mass spectrometry (Pan et al., 2010). Gibberellin A1 (GA1) and gibberellin A4 (GA4) were used as internal standards.

## Growth Conditions for the Transcriptome Experiment

For the gene expression analysis, plants from 31 accessions were grown in a controlled environment growth chamber (Supplemental Table S1). The growth conditions were identical to those described above. At 10 DAT, aerial parts of the seedlings were excised from the roots and frozen immediately in liquid N. The samples were ground with Tissuelyser II (Invitrogen, Waltham, MA) and total RNA was isolated with the RNAeasy isolation kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. On-column DNase treatment was performed to remove genomic DNA contamination (Qiagen). Sequencing was performed on an Illumina HiSeq 2500 (Illumina, San Diego, CA). Sixteen RNA samples were combined in each lane. Two biological replicates were used for each accession (Supplemental Table S1).

## RNA-seq Mapping and Analysis

After being examined with the package FastQC, short reads obtained via Illumina 101-bp single-end RNA sequencing were screened and trimmed using Trimmomatic (Bolger et al., 2014) to ensure each read has

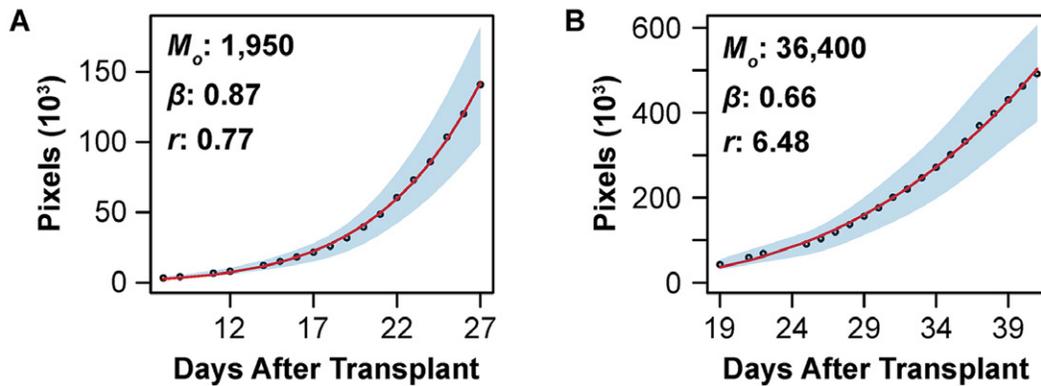


Fig. 1. Shoot growth trajectories of the rice diversity panel during the early (A) and active tillering (B) stages. The mean growth across all rice diversity panel accessions was fitted using a power-law function and is indicated by the red line. The SD is indicated by the light blue shadow. The points indicate the mean growth across all accessions at each individual time point.

average quality score larger than 30 and longer than 15 bp (Andrews, 2010; Bolger et al., 2014). The trimmed short reads were mapped against the rice genome (*Oryza sativa* ‘Nipponbare’ MSU, Release 6.0) using TopHat (version 2.0.10), allowing up to two base mismatches per read. Reads that mapped to multiple locations were discarded (Trapnell et al., 2009). The number of reads for each gene in the MSU version 6 annotation was counted using HTSeq and the “union” resolution mode was used (Anders et al., 2014).

The Bioconductor packages limma and EdgeR were used to identify genes that were differentially expressed between allelic groups at id3013397 and id2010960 (Gentleman et al., 2004; Robinson and Oshlack, 2010; Ritchie et al., 2015). Briefly, genes with a total read count below 50 across all 64 samples were excluded from the differential expression analysis. The libraries were normalized using the trimmed mean of M-values in EdgeR (Robinson and Oshlack, 2010). The voom approach, implemented in limma, was used to identify differentially expressed genes between the two allelic groups for the SNPs id3013397 and id2010960 (Law et al., 2014). Benjamini and Hochberg’s method was used to control the false discovery rate, and genes with an adjusted  $p$ -value  $\leq 0.1$  were considered to be differentially expressed (Benjamini and Hochberg, 1995).

## Results

### Considerable Natural Variation for Shoot Growth Dynamics is Present in the Rice Diversity Panel

We performed image-based high-throughput phenotyping of a subset of the rice diversity panel (Famoso et al., 2011; Zhao et al., 2011) consisting of ~360 accessions belonging to each of the five major rice subpopulations (*indica*, *aus*, *aromatic*, *temperate japonica*, and *tropical japonica*) with the goal of discovering novel genetic variants influencing temporal growth dynamics during vegetative growth. Two separate triplicated experiments were

conducted in which the diversity panel was phenotyped over periods of 19 and 22 d at the early vegetative stage (8–27 DAT; 361 accessions) and the active tillering stage (19–41 d DAT; 360 accessions), respectively. To quantify growth trajectories as a function of time, the shoot growth dynamics were modeled for each accession using a power-law function (Paine et al., 2012) (Eq. [1]). Although rice exhibits a determinate growth pattern, the overlapping developmental windows selected for these studies captured the accelerating phase of shoot growth (Fig. 1A,B).

Considerable natural variation was observed in the diversity panel for all model parameters ( $M_0$ ,  $\beta$ , and  $r$ ) as well as discrete measurements for PSA (Fig. 2; Supplemental File S1). Heritability estimates for model parameters ranged from 0.32 to 0.69 during the early tillering stage, whereas estimates during the late tillering stage ranged from 0.46 to 0.72. In both developmental stages,  $M_0$ , the initial biomass on the first day of imaging, showed the highest heritability (0.69 and 0.72 at the early and active tillering stages, respectively). Model parameters associated with growth rate ( $\beta$  and  $r$ ) displayed moderate heritability estimates ( $\beta$ : 0.46 and 0.47;  $r$ : 0.32 and 0.46 during the early and active tillering stages, respectively). The moderate to high heritability for all model parameters suggests that a portion of the phenotypic variation for growth dynamics is under genetic control. The slightly lower genotypic contribution observed for model parameters describing growth rate compared with  $M_0$  indicates a smaller genotypic effect for these rate parameters and suggests that they may be influenced more by environmental conditions.

To examine the relationship between shoot growth dynamics at the early and active tillering stages, Pearson correlation analysis was conducted for each of the model parameters obtained from the early tillering and active tillering stages, as well as PSA at each time point in the experiment (Supplemental File S2). Significant positive relationships were observed for each model parameter between the two developmental stages, suggesting that the growth characteristics observed at early tillering stage partly persists during the later tillering stages. The

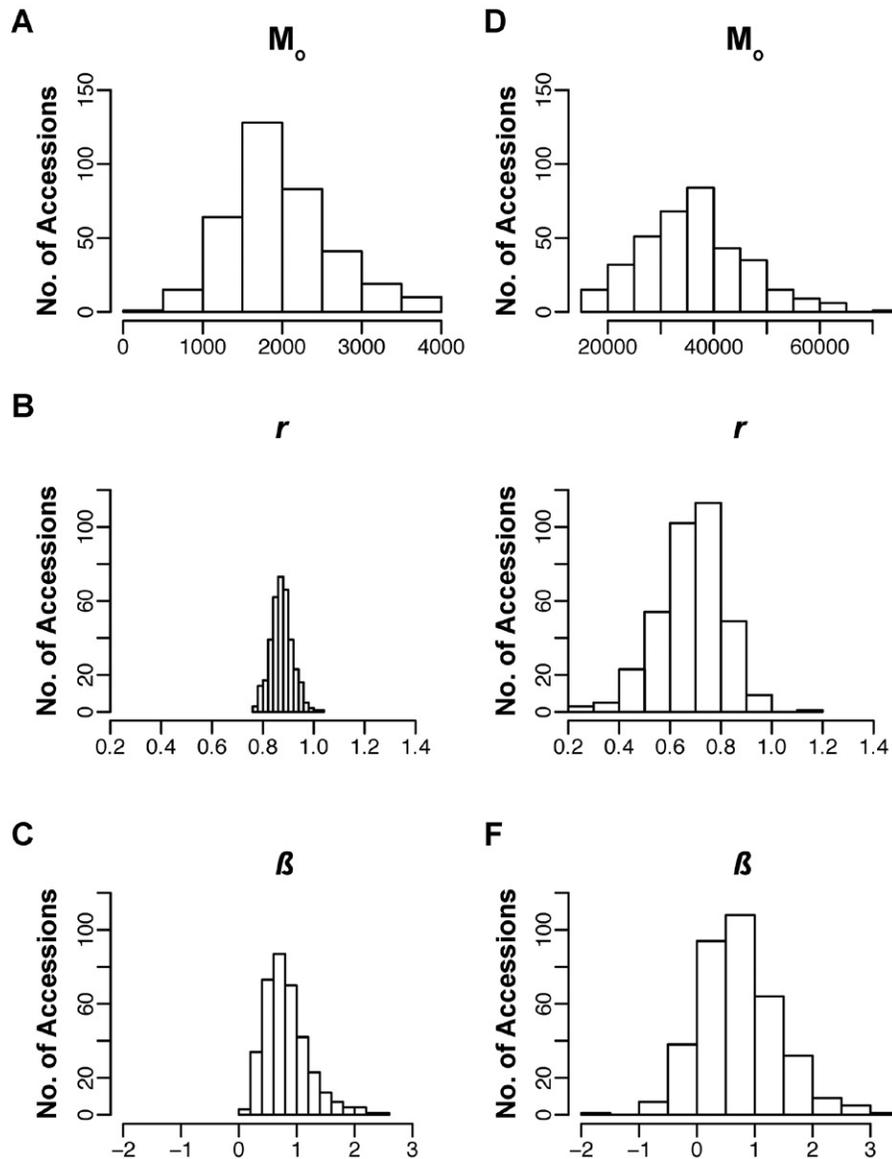


Fig. 2. Distribution for each model parameter during the early (A–C) and active (D–E) tillering stages in rice. The parameters  $\beta$  (C, F) and  $r$  (B, E) control the rate of growth, whereas  $M_0$  (A, D) represents the estimate of the biomass at the start of the experiment. Values for  $\beta$  during the active tillering stage (F) are log-transformed.

model parameter  $r$  (from both developmental stages) displayed a significant negative correlation with PSA during both the early and later tillering stages. Conversely,  $\beta$  displayed a significant positive correlation with PSA, which is not surprising, considering that as values of  $\beta$  approach 1, the power-law function begins to behave more like an exponential function, with constant relative growth rates being achieved when  $\beta = 1$ . Thus plants exhibiting near-exponential growth tend to be very large.

A temporal trend in correlation was observed between the model parameters  $r$  and  $\beta$  and PSA during both developmental stages. For instance, during the early tillering stage  $\beta$  displayed a weak positive relationship with PSA at 8 to 16 DAT ( $r = 0.25$ – $0.48$ ). However, the strength of the correlation became progressively larger as time progressed with the highest being observed at 23 DAT. The parameter  $r$ , on the other hand, displayed

a similar trend, with a weak negative relationship being observed during the early time points and a progressively stronger negative correlation being observed at the later time points. Similar temporal trends were observed for  $\beta$  and  $r$  during the active tillering stage. These results indicate that the parameters obtained from functional modeling ( $\beta$  and  $r$ ) can be used to describe growth characteristics other than plant size that otherwise would not be identified via conventional cross-sectional approaches.

### Genetic Basis of Shoot Biomass

To identify QTLs that exhibit a time-specific effect on shoot biomass we conducted GWAS at each of the 38 time points during the early tillering and active tillering stages. A total of seven QTLs (26 SNPs) were significantly associated with PSA at one or more time points during the early and active tillering stages ( $p < 2.33 \times 10^{-5}$ ;

**Table 1. Quantitative trait loci (QTLs) associated with projected shoot area (PSA) during the early or late vegetative stage. Single-nucleotide polymorphisms within 200 kb were merged and considered to be a single QTL.**

Chromosome	QTL position (bp)	Traits
1	2,136,478–2,334,214	PSA (active tillering)
2	34,616,145–34,630,711	PSA (early tillering); PSA (active tillering)
4	5,923,594–6,033,595	PSA (early tillering)
5	4,871,496–5,071,497	PSA (early tillering)
6	26,235,553–26,631,930	PSA (early tillering); PSA (active tillering)
8	2,901,247–3,101,248	PSA (active tillering)
10	15,397,674–15,659,792	PSA (early tillering); PSA (active tillering)

Table 1; Supplemental File S3). Five QTLs (nine SNPs) were identified during the early tillering stage and five (21 SNPs) were identified during the active tillering stage (Table 1; Supplemental File S3).

Interestingly, nearly a half of the QTLs (three) persisted across developmental stages, suggesting that they may impact growth throughout the vegetative phases captured by the two experiments. For instance, the most significant persistent association was identified in a region on chromosome 6 (~26.2–26.6 Mb) (Fig. 3). Significant signals were identified for 16 of the 38 time points, with the earliest association observed at 22 DAT and the latest at 39 DAT. This QTL explained only a small portion of the phenotypic variation for PSA at 35 DAT (~5%). Several QTL were identified that displayed a time-specific effect. A single QTL at ~5.9 Mb on chromosome 4 was associated with PSA during the early tillering stage at 14 DAT and from 17 to 26 DAT. Two QTLs had effects on PSA during the active tillering stage only. Of these two, a QTL located at ~3 Mb on chromosome 8 had the largest effect on PSA and explained approximately 10% of the variation for PSA at 19 DAT (this QTL was only identified during the active tillering stage). This QTL was associated with PSA at 19 to 22 DAT. These results indicate that shoot biomass may be regulated by multiple loci with small effects that act in both a transient and persistent manner throughout early vegetative growth.

### Functional GWAS Analysis

Although GWAS of PSA at discrete intervals provides information regarding the genetic basis of plant size over time, modeling plant growth allows for the data to be reduced to a mathematical equation that is defined by a small number of parameters, which can then be used as traits for genetic analysis. This approach allows for the detection of the genomic regions associated with growth trajectories, rather than discrete estimates of plant size. To identify genetic regions associated with shoot growth dynamics, GWAS was performed by using the model parameters that were obtained by modeling shoot biomass accumulation at the early tillering and active tillering stages separately. A total of 19 SNPs ( $p < 2.33 \times 10^{-5}$ ),

corresponding to seven unique QTLs were identified for model parameters at both developmental stages (Table 2; Fig. 4). The early tillering stage displayed a higher number of associations, with five QTLs detected, though two QTLs were detected for the model parameter  $M_0$  during the active tillering stage. For all significant markers, the percentage of variation explained by individual QTLs was low ( $\leq 0.10$ ). The most significant association was detected at ~25.2 Mb on chromosome 2 for  $r$  during the early tillering stage and explained approximately 6% of total variation for this trait. No QTL were identified that were common across all developmental stages, suggesting that shoot growth dynamics during the early and active tillering stages may be regulated by distinct genetic mechanisms.

In the power-law function, the parameters  $r$  and  $\beta$  were used to control the rate of growth over time. Both model parameters displayed strong phenotypic correlations during both the early and active tillering stages (–0.93 and –0.54, the for early and active tillering stages, respectively). Joint analysis of correlated phenotypes may improve power to detect genetic associations in GWAS compared to univariate approaches (Ferreira and Purcell, 2009; Kim and Xing, 2009; Korte et al., 2012; O’Reilly et al., 2012; Zhou and Stephens, 2014). With this in mind, model parameters  $r$  and  $\beta$ , which control growth rate, were used in a bivariate mixed model to detect associations in the early and active tillering stages (Zhou and Stephens, 2014). Overall, the bivariate approach identified nine QTLs (19 SNP) in total compared with the univariate approach; however, more than twice as many QTLs were identified during the active tillering stage if the bivariate approach was used (Fig. 4D,4H). During the early tillering stage, four QTLs were identified, two of which were identified by using the bivariate approach only. At the active tillering stage, five QTLs were identified, all of which were detected via the bivariate approach. The most significant association in the early tillering stage was detected at ~10.1 Mb on chromosome 8 ( $p = 6.11 \times 10^{-8}$  for the SNP wd8001592). These results suggest that the bivariate approach is more effective for detecting loci with small effects that regulate shoot growth dynamics.

### Shoot Growth Dynamics during Early Tillering may be Regulated by GA

To identify candidate genes that may regulate vegetative growth dynamics in rice, we treated significant SNPs within a 200-kb window of one another as a single QTL and genes within 200 kb of each QTL were used for further analysis. The 200-kb window was selected on the basis of the estimated LD decay in this diversity panel (Zhao et al., 2011; Famoso et al., 2011). A gene encoding a GA2-oxidase protein (*OsGA2ox7*; LOC\_Os02g41954), which is involved in GA catabolism, was identified within the QTL at ~25.2 Mb on chromosome 2 that was associated with  $r$  during the early tillering stage. *OsGA2ox7* is located approximately 54 kb away from the most significant SNP in this region, which is within

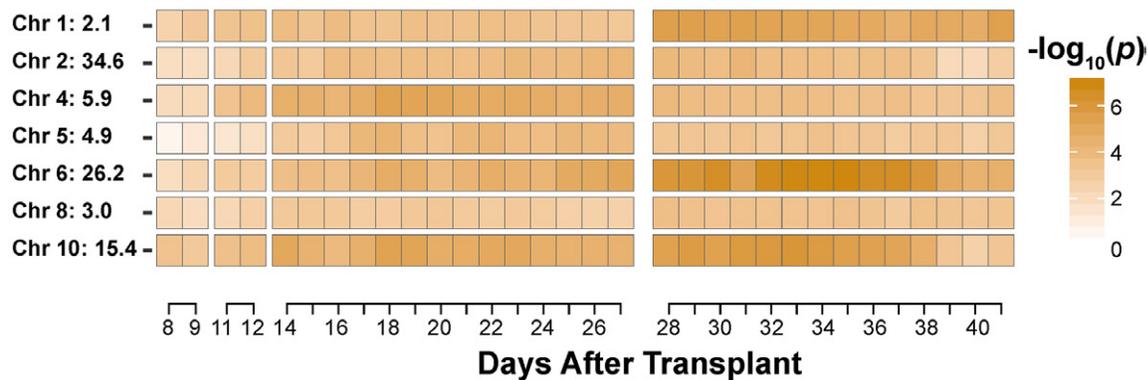


Fig. 3. Genomewide association (GWAS) analysis of projected shoot area (PSA) at each time point during the early and active tillering stages in rice. Genomewide association was performed at each time point and significant single-nucleotide polymorphisms (SNPs) ( $p < 2.33 \times 10^{-5}$ ) within a 200-kb window were considered to be a single quantitative trait locus. Quantitative trait locus positions are indicated on the left of the heatmap. Quantitative trait loci detected during the overlapping time points are provided in Supplemental Table S2.

**Table 2. Subset of quantitative trait loci (QTLs) associated with the model parameters ( $M_0$ ,  $r$ ,  $\beta$ ) during the early and late vegetative stages. Single-nucleotide polymorphisms within 200 kb were merged and considered to be a single QTL.**

Chromosome	QTL position (bp)	Trait†	Candidate genes
1	2,136,478–2,334,214	$M_0$ (AT)	
1	38,578,456–38,578,457	MV (ET)	<i>OsSD1</i> (Sasaki et al., 2002)
2	22,247,678–22,447,679	MV (ET)	
2	25,230,196–25,486,029	$r$ (ET); MV (ET)	<i>OsGA2ox7</i> (Lo et al., 2008)
3	28,521,947–28,621,948	MV (AT)	
4	26,365,306–26,565,307	MV (AT)	
5	4,884,744–5,084,745	$r$ (ET)	
6	26,277,256–26,366,993	MV (AT)	
8	2,901,212–3,101,248	$M_0$ (AT)	
8	8,348,656–8,448,657	$r$ (ET)	
8	10,138,560–10,145,645	MV (AT)	
9	17,857,737–18,361,253	$\beta$ (ET); MV (ET)	<i>OsSG1</i> (Nakagawa et al., 2012), <i>OsHOX4</i> (Dai et al., 2008)
12	3,943,968–4,146,493	$r$ (ET)	
12	22,231,779–22,331,780	MV (AT)	<i>OsT1D1</i> (Sunohara et al., 2009)

† ET, early tillering; AT, active tillering; MV, multivariate

the estimated LD within this region (60 kb). To examine the differences between allelic groups at this QTL, we performed RNA sequencing and GA quantification on a subset of accessions (Supplemental File S5; Supplemental Table S1). A linear model was fitted in the Bioconductor package limma to examine the differences in expression between the two allelic groups at the most significant SNPs for this QTL (Gentleman et al., 2004; Ritchie et al., 2015). Interestingly, the accessions in the major allelic group (A) displayed faster growth rate and significantly lower expression of this transcript ( $p = 0.005$ , SNP id2010960) than minor allelic groups (G) (Fig. 5A–C; Supplemental File S6).

To determine whether there were differences in GA content between allelic groups at the QTLs harboring *OsGA2ox7*, we quantified GA1, GA4, and GA20 levels in shoot tissue of 12 accessions within the major and minor allelic groups at this QTL. Samples were collected during the early tillering stage (10 DAT) to replicate the developmental timing of this QTL. Significant differences were observed between allelic groups for GA4, with the faster growing A allele displaying higher levels of GA4 ( $p < 0.039$ ; Fig. 6A). Although slightly higher levels of GA1 were also observed in the A allelic groups, the differences were not significant at the chosen  $\alpha$  level ( $\alpha = 0.05$ ; Fig. 6B). These results indicate that natural variation for shoot growth dynamics during the early tillering stage may be partly explained by differences in their bioactive GA levels. The differences in GA levels may be caused of higher expression of the GA catalytic enzyme, *OsGA2ox7*.

### Prediction of Shoot Growth Dynamics and PSA

Since the QTLs identified using GWAS explained only a small portion of the total variance for PSA and model parameters, an approach that captures the small effects of many markers, such as GS, may be more advantageous for trait improvement than single marker strategies. With this in mind, GS analysis was performed by using ridge regression best linear unbiased prediction to examine the potential for improving shoot growth dynamics and PSA in rice (Endelman, 2011). The accuracy of GS was assessed using a fivefold cross-validation using 36,901 SNPs, as well as 14 sets of varying size of the top SNPs identified from GWAS with 20 iterations for each SNP set. For each accession in the training population, the GEBVs were calculated from marker effects estimated in the training population and were correlated with phenotypes for the 72 accessions in the validation population. The relationship between GEBVs calculated with all 36,901 markers and observed phenotypes is presented in Fig. 7 and Supplemental Fig. S1 to Supplemental Fig. S28.

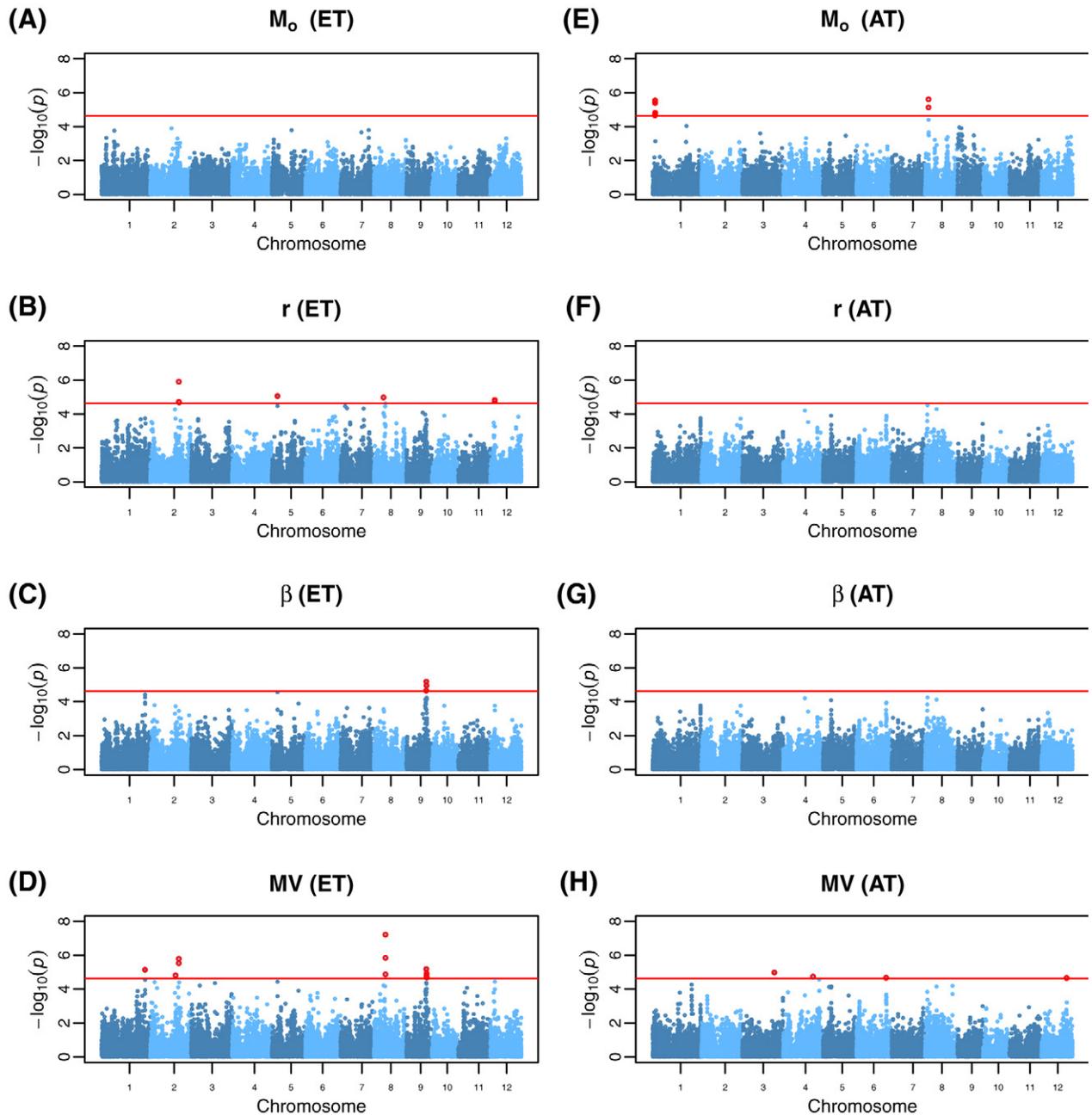


Fig. 4. Genomewide association analysis of the model parameters during the early (A–D) and active tillering (E–H) stages in rice [ $M_0$  (A, E),  $r$  (B, F),  $\beta$  (C, G), multivariate (D, H)]. The red horizontal line indicates the genomewide significance level ( $p < 2.33 \times 10^{-5}$ ) as determined via the effective number of tests ( $M_{\text{eff}}$ ) method with the Šidák correction (Li and Ji, 2005). Significant single-nucleotide polymorphisms are highlighted in red. MV, multivariate; ET, early tillering; AT, active tillering;  $M_0$ , biomass at the start of the experiment.

Prediction accuracies ranged from 0.39 to 0.73 (averaged across all SNP sets) with the highest accuracy observed for PSA at 30 DAT (0.73, averaged across all SNP sets; Supplemental File S7). The mean and SD across the 20 iterations is provided as Supplemental File S7. Of the three model parameters,  $M_0$  exhibited the highest accuracy (0.55 and 0.61 during the early and active tillering stages, respectively) (Fig. 8; Supplemental File S7). Although traits could be predicted with reasonable accuracy using the 50 most significant SNPs from

GWAS (0.37–0.68), the accuracy of prediction improved by including more markers. These trends in accuracy are consistent with a polygenic architecture where hundreds to thousands of loci contribute small effects to the phenotype (Kooke et al., 2016). These results indicate that GS may be an effective strategy for improving shoot biomass and enhancing shoot growth dynamics during the vegetative growth stage in rice.

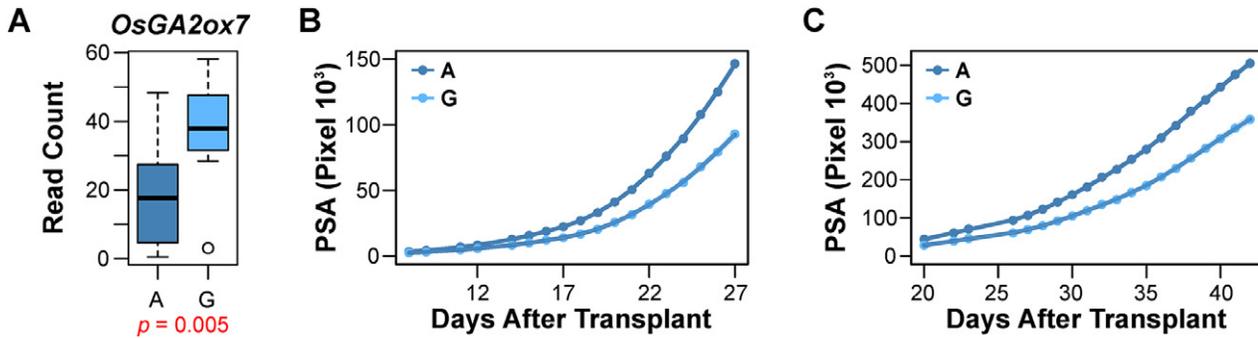


Fig. 5. Role of gibberellic acid (GA) in the contrasting growth responses of rice between allelic groups at ~25.2 Mb on chromosome 2: (A) expression of *OsGA2ox7* between allelic groups at id2010960; (B,C) mean growth trajectories for allelic groups at the most significant single-nucleotide polymorphism (SNP) associated with the model parameter  $r$  (id2010960) during the (B) early and (C) active tillering stages. Statistical significance was determined using the Bioconductor packages *limma* and *edgeR*. The resulting  $p$ -value is indicated in red. Both genes displayed significance difference in expression between allelic groups at the corresponding SNP [false discovery rate (FDR) < 0.1].

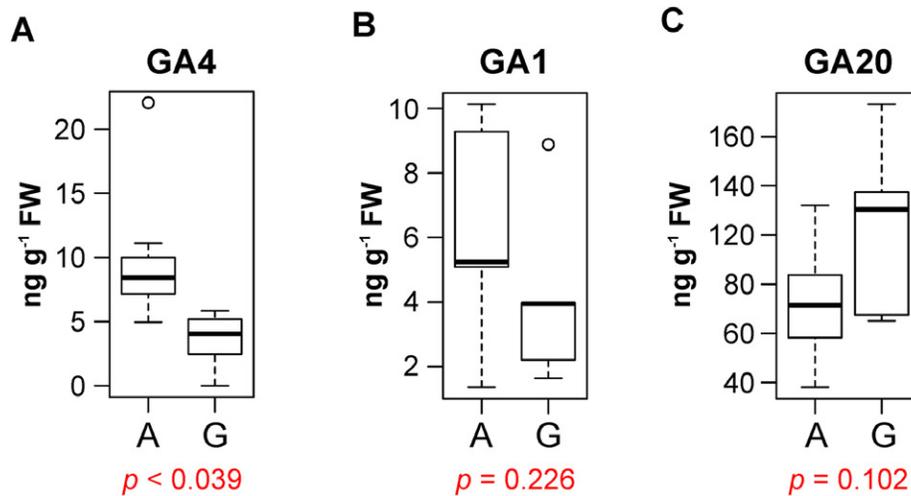


Fig. 6. Gibberellic acid (GA) content in rice within allelic groups at id2010960. Gibberellic acid levels were quantified in shoot tissue at ~10 d after transplantation (DAT). Differences between allelic groups were determined via Student's  $t$ -test and the resulting  $p$ -value is indicated in red ( $n = 12$ ).

## Discussion

Growth is a complex phenotype that is greatly influenced by environmental and developmental cues. The observable phenotype is the cumulative result of many biological processes that occur over time. In grasses, the timing of developmental switches that initiate tiller development or a transition to the reproductive phase has a large influence on vigor and plant size. The temporal nature of such genetic effects are evidenced by the transient QTLs associated with PSA, as well as the small overlap in QTLs associated with model parameters across developmental stages. With the advent of high-throughput phenomics platforms, high-resolution temporal data can be collected nondestructively for large mapping populations. Mathematical equations that describe a developmental or physiological process can be applied to these data to reduce the temporal growth trajectories to a few mathematical parameters and may capture components of the phenotype that may not be detected through cross-sectional phenotyping approaches. The additional

phenotypic information provided by this approach is supported by the temporal correlation trend observed between model parameters and PSA, as well as the identification of 11 unique QTLs that were identified only with the functional GWAS approach. Only four QTLs were identified with both the cross-sectional and functional GWAS approaches. The identification of 11 QTLs that are unique to model parameters may be a result of the greater power to detect alleles with small effects by using phenotypic data across time points; alternatively, the model parameters may be elucidating component(s) of growth phenotype that are not intuitively derived from discrete measurements of plant size.

The influence of plant hormones on agronomic growth-related traits is well documented in rice and other cereals (Peng et al., 1999; Yamamuro et al., 2000; Ikeda et al., 2001; Chandler et al., 2002; Chono et al., 2003; Li et al., 2003; Nakamura et al., 2006; Sakamoto, 2006). Gibberellic acid is a key regulator of plant growth and development and has contributed substantially to

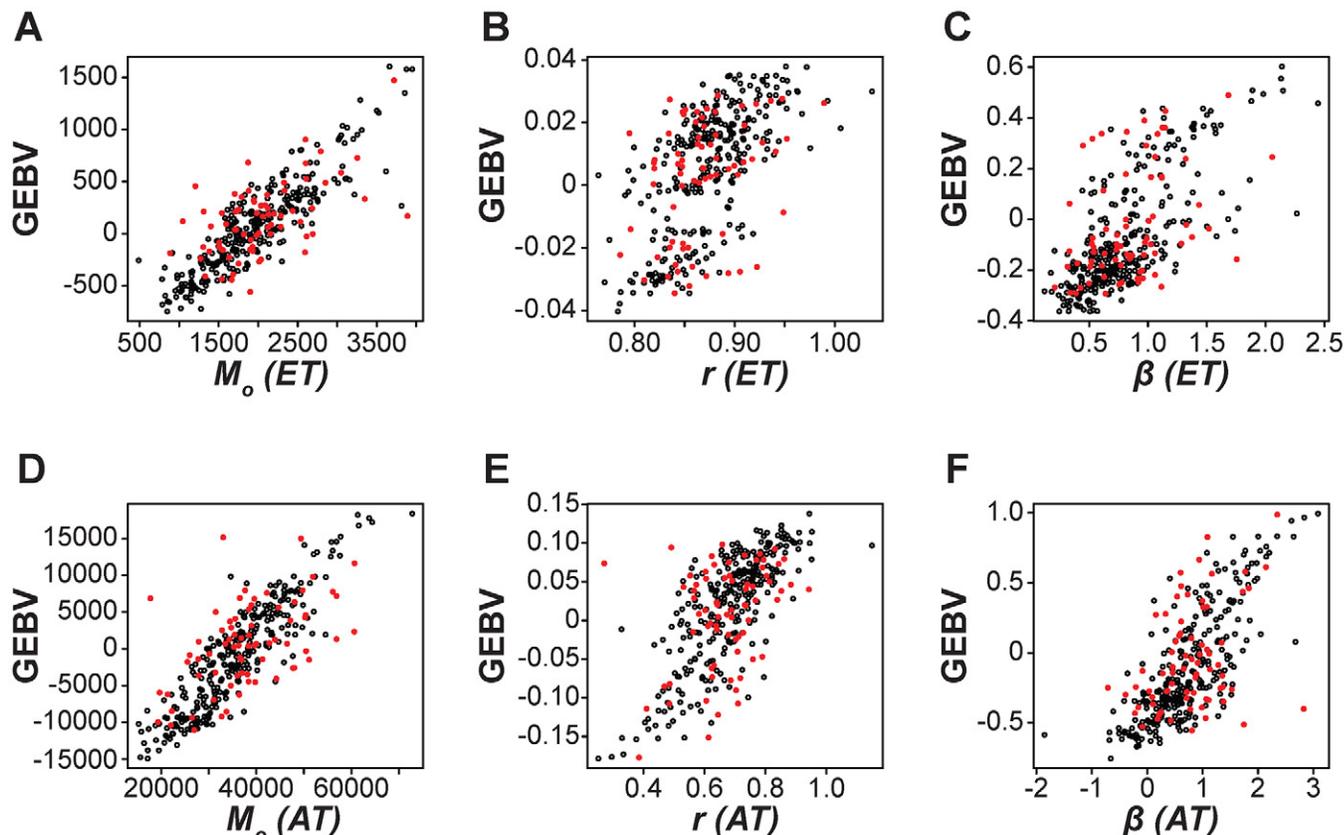


Fig. 7. Comparison of genomic estimated breeding values (GEBVs) and observed phenotypes for model parameters at the early (A–C) and active tillering stages (D–F) of rice [ $M_o$  (A, D),  $r$  (B, E),  $\beta$  (C, F)]. Genomic estimated breeding values were estimated using all 36,901 markers using a training population of 288 accessions. The black open points in each figure represent the accessions in the training population; filled red points indicate accessions in the validation population (72 accessions). ET, early tillering; AT, active tillering.

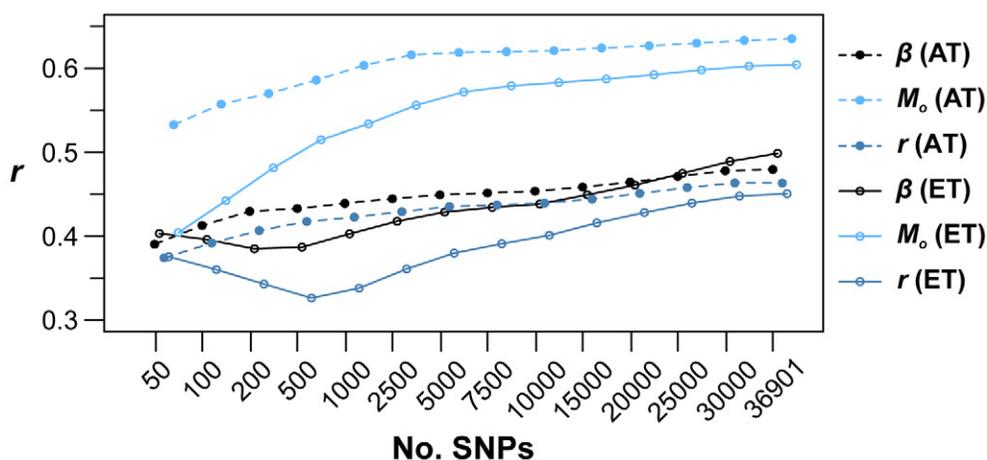


Fig. 8. Prediction accuracies for model parameters during the early and active tillering stages of rice. Prediction accuracy was determined by using a fivefold cross-validation for each of the 14 sets of single-nucleotide polymorphisms (SNPs). The values represent the mean correlation between the genomic estimated breeding values and the observed values for the validation population for 20 iterations of the fivefold cross-validation.

grain production in the 20th century through the development of high-yielding semidwarf cereals (Peng et al., 1999; Sasaki et al., 2002). In rice, the most widely used gene for modifying plant architecture in modern varieties is *sd1*, which encodes a 20-oxidase GA biosynthetic protein (Sasaki et al., 2002). A single nonsynonymous substitution within this gene has a drastic effect on

plant architecture, resulting in a deficiency of bioactive GA and dwarf stature (Sasaki et al., 2002). In our study, GWAS provided evidence for GA in the regulation of early vigor; however, the effects of the QTLs were minor compared with that of *sd1*. A QTL at  $\sim 25.2$  Mb on chromosome 2 was associated with  $r$  during the early tillering stage and harbored a GA catabolic gene, *OsGA2ox7*

(Busov et al., 2003; Lo et al., 2008; Rieu et al., 2008; Wud-dineh et al., 2015). Lo et al. (2008) showed that higher expression of *OsGA2ox7*, relative to wild-type plants, resulted in reduced stature and growth rate. In our study, accessions belonging to the major allele group displayed significantly higher expression of this gene, slower growth rate during the early tillering stage, and lower GA4 levels, which is consistent with the previous report for *OsGA2ox7* (Lo et al., 2008).

Improving early vigor has remained a major challenge in rice breeding programs because of the complex genetic basis and large genotype-by-environment effect. GWAS has proven to be an indispensable approach to identifying causal genes underlying traits in a variety of different organisms. Genomewide association is particularly effective for traits that are regulated by a few loci with large effects (Korte and Farlow, 2013). Several loci were identified for model parameters and PSA that harbored genes known to regulate growth in rice and other species but despite the moderate to high heritability observed for the traits, each individual locus only explained a small portion of the phenotypic variation for the trait. These patterns are typical of complex polygenic traits. With this type of genetic architecture, single marker strategies (e.g., marker-assisted selection) may not be the most effective approach for genetic improvement. Genomic selection, on the other hand, does not focus on individual markers and thus is a more effective approach to breeding for traits controlled by many loci with small effects compared with marker-assisted selection (Jannink et al., 2010). Prediction accuracies for PSA and model parameters were relatively high, ranging from 0.39 to 0.73, which indicates that GS may be an effective strategy for improving shoot growth dynamics in rice. Moreover, high prediction accuracies were obtained by using a subset of informative markers from association mapping, which may further reduce genotyping costs.

This study provides insights into the complex genetic architecture and molecular mechanisms underlying early shoot growth dynamics in rice. Although early vigor is of interest in many breeding programs, the complex genetic basis, the temporal component of the phenotype, and large genotype  $\times$  environment effects have hindered genetic improvement. The evaluation of large mapping populations in a controlled environment using nondestructive phenomics provides high-resolution phenotypic data and reduces the influence of the environment on this complex trait. Moreover, the use of a large diversity panel allows for a greater allelic diversity to be queried, thus providing broader insight into the genetic basis of this trait than that from studies using biparental populations. Our study presents a foundational approach for the elucidation of vegetative growth dynamics and early vigor in rice. The approach of combining high-resolution image-based phenotyping coupled with functional mapping and genome prediction could be widely applicable for complex traits across numerous crop species.

## Supplemental Information

File S1: Phenotypic data for PSA and model parameters for all accessions of Rice Diversity Panel 1 (RDP1).

File S2: Pearson correlation coefficients for model parameters and PSA.

File S3: Genomewide association results for daily measurements of PSA. The values in each column represent the  $-\log_{10}p$ -value for each SNP.

File S4: Genes within 200 kb of significant SNPs for the GWAS of model parameters during the early and active tillering stages.

File S5: Read counts for the 2417 genes found within 200 kb of a significant SNP from the GWAS with model parameters.

File S6: Genes displaying significant differences in expression between allelic groups at id2010960 (FDR < 0.1).

File S7: Prediction accuracies for daily measurements of PSA during the early and active tillering stages.

Figures S1–S28. Comparison of genomic estimated breeding values (GEBV) and observed phenotypes for projected shoot area (PSA). GEBV were estimated using all 36,901 markers using a training population of 288 accessions. The black open points in each figure represent the accessions in the training population; the filled red points indicate accessions in the validation population (72 accessions). ET, early tillering; AT, active tillering.

Table S1: Accessions used for RNA sequencing, GA quantification or both.

Table S2: Quantitative trait loci associated with PSA detected on overlapping days (19–27 d after transplant) in the early and active tillering experiments. DAT, d after transplantation.

## Conflict of Interest Disclosure

The authors declare no conflicts of interest.

## Acknowledgments

This work is funded by a National Science Foundation Plant Genome Research grant (Award no. 1238125) to HW. The Plant Accelerator, Australian Plant Phenomics Facility, is supported under the National Cooperative Research Infrastructure Strategy of the Australian Government.

## References

- Anders, S., P.T. Pyl, and W. Huber. 2014. HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2):166–169. doi:10.1093/bioinformatics/btu638
- Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data. Babraham Inst., Cambridge, UK. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed 18 Apr. 2017).
- 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
- Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57:289–300.
- Berger, B., B. Parent, and M. Tester. 2010. High-throughput shoot imaging to study drought responses. *J. Exp. Bot.* 61:3519–3528. doi:10.1093/jxb/erq201

- Bolger, A.M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120. doi:10.1093/bioinformatics/btu170.
- Browning, B.L., and S.R. Browning. 2016. Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* 98(1):116–126. doi:10.1016/j.ajhg.2015.11.020
- Busemeyer, L., A. Ruckelshausen, K. Möller, A.E. Melchinger, K. V Alheit, H.P. Maurer, et al. 2013. Precision phenotyping of biomass accumulation in triticale reveals temporal genetic patterns of regulation. *Sci. Rep.* 3:2442. doi:10.1038/srep02442
- Busov, V.B., R. Meilan, D.W. Pearce, C. Ma, S.B. Rood, and S.H. Strauss. 2003. Activation tagging of a dominant gibberellin catabolism gene (GA 2-oxidase) from poplar that regulates tree stature. *Plant Physiol.* 132:1283–1291. doi:10.1104/pp.103.020354
- Cairns, J.E., O.S. Namuco, R. Torres, F.A. Simborio, B. Courtois, G.A. Aquino, et al. 2009. Investigating early vigour in upland rice (*Oryza sativa* L.): II. Identification of QTLs controlling early vigour under greenhouse and field conditions. *Field Crops Res.* 113:207–217. doi:10.1016/j.fcr.2009.05.007
- 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.* 168:1476–1489. doi:10.1104/pp.15.00450
- Chandler, P.M., A. Marion-Poll, M. Ellis, and F. Gubler. 2002. Mutants at the *Slender1* locus of barley cv Himalaya. Molecular and physiological characterization. *Plant Physiol.* 129:181–190. doi:10.1104/pp.010917
- Chen, D., K. Neumann, and S. Friedel. 2014. Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis. *Plant Cell* 26:4636–4655. doi:10.1105/tpc.114.129601
- Chono, M., I. Honda, H. Zeniya, K. Yoneyama, D. Saisho, K. Takeda, et al. 2003. A semidwarf phenotype of barley uzu results from a nucleotide substitution in the gene encoding a putative brassinosteroid receptor. *Plant Physiol.* 133:1209–1219. doi:10.1104/pp.103.026195
- Dai, M., Y. Hu, Q. Ma, Y. Zhao, and D.-X. Zhou. 2008. Functional analysis of rice *HOMEBOX4* (*Oshox4*) gene reveals a negative function in gibberellin responses. *Plant Mol. Biol.* 66:289–301. doi:10.1007/s11103-007-9270-8
- Eizenga, G.C., M. Ali, R.J. Bryant, K.M. Yeater, A.M. McClung, S.R. McCouch, et al. 2014. Registration of the Rice Diversity Panel 1 for genomewide association studies. *J. Plant Reg.* 8:109–116. doi:10.3198/jpr2013.03.0013crmp
- Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* 4:250–255. doi:10.3835/plantgenome2011.08.0024
- Famoso, A.N., K. Zhao, R.T. Clark, C.-W. Tung, M.H. Wright, C. Bustamante, et al. 2011. Genetic architecture of aluminum tolerance in rice (*Oryza sativa*) determined through genome-wide association analysis and QTL mapping. *PLoS Genet.* 7(8):e1002221. doi:10.1371/journal.pgen.1002221
- Ferreira, M.A.R., and S.M. Purcell. 2009. A multivariate test of association. *Bioinformatics* 25:132–133. doi:10.1093/bioinformatics/btn563
- Gentleman, R.C., V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, S. Dudoit, et al. 2004. Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol.* 5:R80. doi:10.1186/gb-2004-5-10-r80
- 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* 7:2. doi:10.1186/1746-4811-7-2
- Grothendieck, G. 2013. nls2: Non-linear regression with brute force. R Package version. R Project for Statistical Computing. <https://CRAN.R-project.org/package=nls2> (accessed 5 May 2017).
- Hairmansis, A., B. Berger, M. Tester, and S.J. Roy. 2014. Image-based phenotyping for non-destructive screening of different salinity tolerance traits in rice. *Rice* 7(1):16. doi:10.1186/s12284-014-0016-3
- Honsdorf, N., T. March, B. Berger, M. Tester, and K. Pillen. 2014. High-throughput phenotyping to detect drought tolerance QTL in barley introgression lines. *PLoS ONE* 9(5):e97047. doi:10.1371/journal.pone.0097047
- Ikedo, A., M. Ueguchi-Tanaka, Y. Sonoda, H. Kitano, M. Koshioka, Y. Futsuhara, et al. 2001. *slender* rice, a constitutive gibberellin response mutant, is caused by a null mutation of the *SLR1* gene, an ortholog of the height-regulating gene *GAI/RGA/RHT/D8*. *Plant Cell* 13:999–1010. doi:10.1105/tpc.13.5.999
- Jannink, J.-L., A.J. Lorenz, and H. Iwata. 2010. Genomic selection in plant breeding: From theory to practice. *Brief. Funct. Genomics* 9:166–177. doi:10.1093/bfpg/elq001
- Kang, H.M., N.A. Zaitlen, C.M. Wade, A. Kirby, D. Heckerman, M.J. Daly, et al. 2008. Efficient control of population structure in model organisms association mapping. *Genetics* 178:1709–1723. doi:10.1534/genetics.107.080101
- Kim, S., and E.P. Xing. 2009. Statistical estimation of correlated genome associations to a quantitative trait network. *PLoS Genet.* 5(8):e1000587. doi:10.1371/journal.pgen.1000587
- 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.* 67:3587–3599. doi:10.1093/jxb/erw176
- Kooke, R., W. Kruijjer, R. Bours, F.F.M. Becker, A. Kuhn, J. Buntjer, et al. 2016. Genome-wide association mapping and genomic prediction elucidate the genetic architecture of morphological traits in *Arabidopsis thaliana*. *Plant Physiol.* 170(4): 2187–2203. doi:10.1104/pp.15.00997
- Korte, A., and A. Farlow. 2013. The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods* 9:29. doi:10.1186/1746-4811-9-29
- Korte, A., B.J. Vilhjálmsson, V. Segura, A. Platt, Q. Long, and M. Nordborg. 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
- Law, C.W., Y. Chen, W. Shi, and G.K. Smyth. 2014. Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 15:R29. doi:10.1186/gb-2014-15-2-r29
- Li, J., and L. Ji. 2005. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* 95:221–227. doi:10.1038/sj.hdy.6800717
- Li, X., Q. Qian, Z. Fu, Y. Wang, G. Xiong, D. Zeng, et al. 2003. Control of tillering in rice. *Nature* 422:618–621. doi:10.1038/nature01518
- Liu, L., Y. Lai, J. Cheng, L. Wang, W. Du, Z. Wang, et al. 2014. Dynamic Quantitative Trait Locus Analysis of Seed Vigor at Three Maturity Stages in Rice. *PLoS ONE* 9(12):e115732. doi:10.1371/journal.pone.0115732
- Lo, S.-F., S.-Y. Yang, K.-T. Chen, Y.-I. Hsing, J.A.D. Zeevaert, L.-J. Chen, and S.-M. Yu. 2008. A novel class of gibberellin 2-oxidases control semidwarfism, tillering, and root development in rice. *Plant Cell* 20:2603–2618. doi:10.1105/tpc.108.060913
- Lu, X.-L., A.-L. Niu, H.-Y. Cai, Y. Zhao, J.-W. Liu, Y.-G. Zhu, et al. 2007. Genetic dissection of seedling and early vigor in a recombinant inbred line population of rice. *Plant Sci.* 172(2):212–220. doi:10.1016/j.plantsci.2006.08.012
- Luquet, D., J.C. Soulié, M.C. Rebolledo, L. Rouan, A. Clément-Vidal, and M. Dingkuhn. 2012. Developmental dynamics and early growth vigour in rice 2: Modelling genetic diversity using ecomeristem. *J. Agron. Crop Sci.* 198:385–398. doi:10.1111/j.1439-037X.2012.00527.x
- Mahender, A., A. Anandan, and S.K. Pradhan. 2015. Early seedling vigour, an imperative trait for direct-seeded rice: An overview on physio-morphological parameters and molecular markers. *Planta* 241(5):1027–1050. doi:10.1007/s00425-015-2273-9
- Moore, C.R., L.S. Johnson, I.-Y. 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
- Murashige, T., and F. Skoog. 1962. A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant.* 15(3):473–497.
- Nakagawa, H., A. Tanaka, T. Tanabata, M. Ohtake, S. Fujioka, H. Nakamura, et al. 2012. *SHORT GRAIN1* decreases organ elongation and brassinosteroid response in rice. *Plant Physiol.* 158(3):1208–1219. doi:10.1104/pp.111.187567

- Nakamura, A., S. Fujioka, H. Sunohara, N. Kamiya, Z. Hong, Y. Inukai, et al. 2006. The role of *OsBRII* and its homologous genes, *OsBRL1* and *OsBRL3*, in rice. *Plant Physiol.* 140(2):580–590. doi:10.1104/pp.105.072330
- Nash, J. 2013. nlmrt: Functions for nonlinear least squares solutions, 2013. R Package version 9(0): 50. R Project for Statistical Computing. <https://CRAN.R-project.org/package=nlmrt> (accessed 5 May 2017).
- O'Reilly, P.F., C.J. Hoggart, Y. Pomyen, F.C.F. Calboli, P. Elliott, M.-R. Jarvelin, et al. 2012. MultiPhen: Joint model of multiple phenotypes can increase discovery in GWAS. *PLoS ONE* 7(5):e34861. doi:10.1371/journal.pone.0034861
- Paine, C.E.T., T.R. Marthens, D.R. Vogt, D. Purves, M. Rees, A. Hector, et al. 2012. How to fit nonlinear plant growth models and calculate growth rates: An update for ecologists. *Methods Ecol. Evol.* 3:245–256. doi:10.1111/j.2041-210X.2011.00155.x
- Pan, X., R. Welti, and X. Wang. 2010. Quantitative analysis of major plant hormones in crude plant extracts by high-performance liquid chromatography–mass spectrometry. *Nat. Protoc.* 5:986–992. doi:10.1038/nprot.2010.37
- Peng, J., D.E. Richards, N.M. Hartley, G.P. Murphy, K.M. Devos, J.E. Flintham, et al. 1999. “Green revolution” genes encode mutant gibberellin response modulators. *Nature* 400:256–261. doi:10.1038/22307
- Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, and R Core Team. 2015. {nlme}: Linear and nonlinear mixed effects models. R package version 3.1-131. R Project for Statistical Computing. <https://CRAN.R-project.org/package=nlme> (accessed May 5 2017).
- R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.r-project.org/> (accessed 18 Apr. 2017).
- Rebolledo, M.-C., M. Dingkuhn, A. Clement-Vidal, L. Rouan, and D. Luquet. 2012a. Phenomics of rice early vigour and drought response: Are sugar related and morphogenetic traits relevant? *Rice (N. Y.)* 5(1):22. doi:10.1186/1939-8433-5-22
- Rebolledo, M.C., M. Dingkuhn, B. Courtois, Y. Gibon, A. Clément-Vidal, D.F. Cruz, et al. 2015. Phenotypic and genetic dissection of component traits for early vigour in rice using plant growth modelling, sugar content analyses and association mapping. *J. Exp. Bot.* 66(18):5555–5566. doi:10.1093/jxb/erv258
- Rebolledo, M.C., M. Dingkuhn, P. Péré, K.L. McNally, and D. Luquet. 2012b. Developmental dynamics and early growth vigour in rice. I. Relationship between development rate (1/Phyllochron) and growth. *J. Agron. Crop Sci.* 198(5):374–384. doi:10.1111/j.1439-037X.2012.00528.x
- Redoña, E.D., and D.J. Mackill. 1996. Genetic variation for seedling vigor traits in rice. *Crop Sci.* 36(2):285–290. doi:10.2135/cropsci1996.0011183X003600020012x
- Rieu, I., S. Eriksson, S.J. Powers, F. Gong, J. Griffiths, L. Woolley, et al. 2008. Genetic analysis reveals that C19-GA 2-oxidation is a major gibberellin inactivation pathway in *Arabidopsis*. *Plant Cell* 20(9):2420–2436. doi:10.1105/tpc.108.058818
- Ritchie, M.E., B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, et al. 2015. *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43(7):e47. doi:10.1093/nar/gkv007
- Robinson, M.D., and A. Oshlack. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11:R25. doi:10.1186/gb-2010-11-3-r25
- Sakamoto, T. 2006. Phytohormones and rice crop yield: Strategies and opportunities for genetic improvement. *Transgenic Res.* 15(4):399–404. doi:10.1007/s11248-006-0024-1
- Sasaki, A., M. Ashikari, M. Ueguchi-Tanaka, H. Itoh, A. Nishimura, D. Swapan, et al. 2002. Green revolution: A mutant gibberellin-synthesis gene in rice. *Nature* 416(6882):701–702. doi:10.1038/416701a
- Šidák, Z. 1967. Rectangular confidence regions for the means of multivariate normal distributions. *J. Am. Stat. Assoc.* 62(318): 626–633.
- Slovak, R., C. Göschl, X. Su, K. Shimotani, T. Shiina, and W. Busch. 2014. A scalable open-source pipeline for large-scale root phenotyping of *Arabidopsis*. *Plant Cell* 26(6):2390–2403. doi:10.1105/tpc.114.124032
- Sunohara, H., T. Kawai, S. Shimizu-Sato, Y. Sato, K. Sato, and H. Kitano. 2009. A dominant mutation of *TWISTED DWARF 1* encoding an  $\alpha$ -tubulin protein causes severe dwarfism and right helical growth in rice. *Genes Genet. Syst.* 84(3):209–218. doi:10.1266/ggs.84.209
- Topp, C.N., A.S. Iyer-Pascuzzi, J.T. Anderson, C.-R. Lee, P.R. Zurek, O. Symonova, et al. 2013. 3D phenotyping and quantitative trait locus mapping identify core regions of the rice genome controlling root architecture. *Proc. Natl. Acad. Sci. USA* 110(18):E1695–E1704. doi:10.1073/pnas.1304354110
- Trapnell, C., L. Pachter, and S.L. Salzberg. 2009. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* 25:1105–1111. doi:10.1093/bioinformatics/btp120
- Wu, R., and M. Lin. 2006. Functional mapping—how to map and study the genetic architecture of dynamic complex traits. *Nat. Rev. Genet.* 7:229–237. doi:10.1038/nrg1804
- Wuddineh, W.A., M. Mazarei, J. Zhang, C.R. Poovaiah, D.G.J. Mann, A. Ziebell, et al. 2015. Identification and overexpression of *gibberellin 2-oxidase (GA2ox)* in switchgrass (*Panicum virgatum* L.) for improved plant architecture and reduced biomass recalcitrance. *Plant Biotechnol. J.* 13:636–647. doi:10.1111/pbi.12287
- 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 Genet.* 15:59. doi:10.1186/1471-2156-15-59
- Xu, Z., X. Shen, and W. Pan. 2014. Longitudinal analysis is more powerful than cross-sectional analysis in detecting genetic association with neuroimaging phenotypes. *PLoS ONE* 9(8):e102312. doi:10.1371/journal.pone.0102312
- Yamamuro, C., Y. Ihara, X. Wu, T. Noguchi, S. Fujioka, S. Takatsuto, et al. 2000. Loss of function of a rice *brassinosteroid insensitive1* homolog prevents internode elongation and bending of the lamina joint. *Plant Cell* 12:1591–1605. doi:10.1105/tpc.12.9.1591
- Yang, W., Z. Guo, C. Huang, L. Duan, G. Chen, N. Jiang, et al. 2014. Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* 5:5087. doi:10.1038/ncomms6087
- Yoshida, S., D.A. Forno, and J.H. Cock. 1976. Laboratory manual for physiological studies of rice. IRRI, Manila, Philippines.
- Zhang, Z., E. Ersoz, C.-Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, et al. 2010. Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42:355–360. doi:10.1038/ng.546
- Zhao, K., C.-W. 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. doi:10.1038/ncomms1467
- Zhou, X., and M. Stephens. 2014. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* 11(4): 407–409. doi:10.1038/nmeth.2848