Journal of Experimental Botany, Vol. 70, No. 9 pp. 2561–2574, 2019
doi:10.1093/jxb/erz090 Advance Access Publication 2 March, 2019
This paper is available online free of all access charges (see https://academic.oup.com/jxb/pages/openaccess for further details)

RESEARCH PAPER

A framework for genomics-informed ecophysiological modeling in plants

Diane R. Wang[1,*,†], Carmela R. Guadagno [2,†], Xiaowei Mao[3,†], D. Scott Mackay[1], Jonathan R. Pleban[1], Robert L. Baker[4], Cynthia Weinig[2], Jean-Luc Jannink[3,5] and Brent E. Ewers[2]

[1] Geography Department, University at Buffalo, Buffalo, NY 14226, USA
[2] Botany Department, University of Wyoming, Laramie, WY 82071, USA
[3] Plant Breeding and Genetics Section, Cornell University, Ithaca, NY 14853, USA
[4] Biology Department, Miami University, Oxford, OH 45056, USA
[5] USDA-ARS, Ithaca, NY, 14853, USA

[†]These authors contributed equally to this work.

*Correspondence: drwang@buffalo.edu

Received 14 December 2018; Editorial decision 18 February 2019; Accepted 18 February 2019

Editor: Bertrand Muller, INRA, France

Abstract

Dynamic process-based plant models capture complex physiological response across time, carrying the potential to extend simulations out to novel environments and lend mechanistic insight to observed phenotypes. Despite the translational opportunities for varietal crop improvement that could be unlocked by linking natural genetic variation to first principles-based modeling, these models are challenging to apply to large populations of related individuals. Here we use a combination of model development, experimental evaluation, and genomic prediction in *Brassica rapa* L. to set the stage for future large-scale process-based modeling of intraspecific variation. We develop a new canopy growth submodel for *B. rapa* within the process-based model Terrestrial Regional Ecosystem Exchange Simulator (TREES), test input parameters for feasibility of direct estimation with observed phenotypes across cultivated morphotypes and indirect estimation using genomic prediction on a recombinant inbred line population, and explore model performance on an *in silico* population under non-stressed and mild water-stressed conditions. We find evidence that the updated whole-plant model has the capacity to distill genotype by environment interaction (G×E) into tractable components. The framework presented offers a means to link genetic variation with environment-modulated plant response and serves as a stepping stone towards large-scale prediction of unphenotyped, genetically related individuals under untested environmental scenarios.

Keywords: Abiotic stress, development, G×E, growth, genomic prediction, process-based models.

Introduction

Beginning in the 1960s, dynamic process-based models for plants have been implemented successfully to address research questions and inform decision-making across a variety of fundamental and applied settings (Alberda and Sibma, 1968; de Wit *et al.*, 1970; Hoogenboom, 2000; Jones *et al.*, 2003; Keating *et al.*, 2003; Sinclair *et al.*, 2010; Messina *et al.*, 2015; Thorp *et al.*, 2017; Johnson *et al.*, 2018). These models enable the study of time-series responses of a plant system to various inputs by formalizing

physiological processes into assemblies of mathematical functions. While a diversity of tools exist, with names such as 'biophysical models', 'ecophysiological models', and 'crop growth models', one unifying feature is that they can readily synthesize frontier knowledge from divergent disciplines (Bouman *et al.*, 1996), thereby adapting continuously to a changing research landscape. Since the rise of the genomics era, the capacity of process-based modeling to handle temporal climatic variation and extend predictions to novel scenarios such as stress conditions has made it attractive to scientific communities interested in dissecting the genetics of environmentally plastic traits or in predicting performance for breeding (White and Hoogenboom, 1996, 2003; Hammer *et al.*, 2005, 2006; Messina *et al.*, 2006; Struik, 2012; Yin *et al.*, 2018). Despite the translational opportunities proffered by successful linking of genetic variation and physiological dynamics via process-based modeling, major challenges for extending these models to handle the sizable populations relevant to plant geneticists and breeders remain. How can we systematically generate large sets of genotype-specific input parameters needed by biophysical models? Literature values available at the species or genus level often used for modeling may be suitable for conserved traits with little intraspecific variation, but are not appropriate for traits expressing high variability at the genotype level. Conversely, how does process model structure need to be modified to utilize genotype-specific inputs? First principles-based models are conventionally developed without comprehensive evaluation of the genetic relevance of parameters.

Innovative approaches to address the challenge of combining process modeling and natural genetic variation have been investigated with varying degrees of success. Early efforts involved modeling the effects of quantitative trait loci (QTLs) discovered in mapping populations to inform process model parameters or using modeling outcomes for QTL discovery itself (Yin *et al.*, 1999, 2005; Yin, 2000). Another early example is the incorporation of seven genes that affected phenology, growth habit, and seed size into BEANGRO, a process model for the common bean, *Phaseolus vulgaris* L (White and Hoogenboom, 1996). One downside to a gene- or QTL-based approach for process model parameterization is that typically only loci with large or moderate effect sizes can be used as predictors because loci of small effects are likely to be missed in the discovery process. An alternative is genomic prediction (GP), which entails estimating breeding values or other traits of unphenotyped individuals using whole-genome models trained on related phenotyped individuals (Meuwissen *et al.*, 2001). Compared with a QTL-based approach, genetic effects are estimated for each marker regardless of effect sizes, and then are summed up to predict the overall breeding value. GP was first implemented in cattle breeding to increase the rate of genetic gain (Schaeffer, 2006), and has since been adopted across plant breeding communities for black-box (empirical) prediction of agronomically important traits (Heffner *et al.*, 2011; Rutkoski *et al.*, 2015; Spindel *et al.*, 2015; Thavamanikumar *et al.*, 2015; Minamikawa *et al.*, 2017; Okeke *et al.*, 2017; Wolfe *et al.*, 2017). Recently, Technow *et al.* developed a method to capture biological information in predictions by integrating crop modeling directly into a genomic prediction framework, which was later applied to a maize data set (Technow *et al.*, 2015; Cooper *et al.*, 2016). The reverse application, namely utilization of genomic prediction to inform model parameters, has been proposed (Tardieu *et al.*, 2018) but, to our knowledge, it has not yet been implemented.

Here we explore a modeling framework in *Brassica rapa* L. that leverages genomic prediction to drive genotype-dependent variability of a plant process model, Terrestrial Regional Ecosystem Exchange Simulator (TREES) (Mackay *et al.*, 2015). This model was selected for its capacity to model rhizosphere and xylem hydraulics, critical processes in drought-stressed conditions. Recognizing that growth and development are likely to be highly genotype specific, we build and iteratively update a new canopy growth submodel within TREES as the principal link between genetic complexity and dynamic physiological response, testing it under non-stressed and water-limited scenarios. We ask several fundamental questions. Can input parameters be directly estimated from real-world observation? How does uncertainty at the parameter level propagate to outcome uncertainty? Are parameters genetically relevant and can they then be predicted using genomic information?

Materials and methods

Model development

The process model TREES couples soil–plant hydraulics with plant physiological routines such as photosynthesis and carbon allocation. The model has been well tested under various natural and experimental conditions primarily for long-lived woody species (Mackay *et al.*, 2015; McDowell *et al.*, 2016; Tai *et al.*, 2017; Johnson *et al.*, 2018), where development was assumed to occur across very long time frames (e.g. years or decades). In order to extend the application of TREES for use in genotypes of short-lived herbaceous plants that express dramatic ontogenetic changes over relatively short time frames (e.g. days, weeks, and months), we developed a new submodel to capture temperature-dependent canopy growth based on the emergence and expansion events of individual leaves. The other subprocesses within TREES were retained, with the rationale that the mechanisms underlying fundamental biophysical processes such as C_3 photosynthesis and hydraulics are conserved and therefore can be dealt with via proper parameterization versus a design overhaul. The new canopy submodel uses 24 additional input parameters (Table 1), selected based on feasibility of being either empirically measured or predicted with genomic models, and is designed to begin model simulations at the initiation of the third epicotylar leaf. The primary governing equation used for computation of theoretical relative rate of leaf growth is:

$$r_{ij} = \frac{A'(\tau_{ij})}{A(\tau_{ij})} \quad (1)$$

where τ_{ij} is thermal time of leaf i spent in process j (cell division or leaf expansion). $A(\tau_{ij})$ is the thermal time-dependent leaf area following a logistic growth function

$$A(\tau_{ij}) = \frac{K}{1 + \frac{K - N_0}{N_0} \cdot \exp(-r \cdot \tau_{ij})} \quad (2)$$

parameterized by three inputs K, N_0, and r, and $A'(\tau_{ij})$ is its first-order derivative. Details on model development including other governing equations are presented in Supplementary Appendix S1 at *JXB* online, and a pipeline describing how to derive leaf area parameter information for the new submodel in TREES is presented in Supplementary Fig. S1. Scripts associated with the pipeline may be accessed at https://github.com/DRWang3/leaf_model_TREES_paper (last accessed 6 March 2019) along with the version of TREES used in this study. For sensitivity analysis simulations of stochastic leaf growth parameters, we sampled from

Table 1. *Input parameters for the new canopy growth submodel*

Type	Parameter	Description	Unit
Module on/off	*useLeafModule*	1 to grow individual leaves; 0 otherwise	Unitless
Leaf growth and development	*leafAreaMax**	Parameter *K* from logistic growth function	cm^2
	*initialLeafSize**	Parameter N_0 from logistic growth function	cm^2
	*leafArea_Rate**	Parameter *r* from logistic growth function	Unitless
	useLeafGamma	1 for stochastic variation in leaf growth; 0 for deterministic simulation	Unitless
	*Kalpha**	Shape parameter of gamma distribution for *leafAreaMax*	Unitless
	*Kbeta**	Rate parameter of gamma distribution for *leafAreaMax*	Unitless
	*Nalpha**	Shape parameter of gamma distribution for *initialLeafSize*	Unitless
	*Nbeta**	Rate parameter of gamma distribution for *initialLeafSize*	Unitless
	*ralpha**	Shape parameter of gamma distribution for *leafArea_Rate*	Unitless
	*rbeta**	Rate parameter of gamma distribution for *leafArea_Rate*	Unitless
	dur_LeafExpansion	Thermal time to 95% of *leafAreaMax*	°Ch
	SLA_low	Genotype-specific lower limit for specific leaf area	kg C m^{-2}
	leaf_insertAngle	Angle between leaf petiole and main stem	°
	leaf_len_to_width	Ratio between leaf length and width	Unitless
Temperature-dependent development	*phyllochron*	Thermal time between leaf appearance events	°Ch
	floweringTime	Thermal time to flowering	°Ch
	Tbase	Developmental base temperature	°C
Model initiation	*therm_plant_init*	Thermal time to third leaf initiation (approximated by leaf emergence)	°Ch
	projectedArea_init	Projected shoot area at third leaf initiation (approximated by leaf emergence)	cm^2
	SLA_init	Specific leaf area used to compute state variables at initialization	kg C m^{-2}
Ground area	*pot_size*	Area of pot	cm^2
Growth scaling	*root_to_shoot*	Root to shoot carbon allocation	mg mg^{-1}
	leaf_to_stem	Leaf to stem carbon allocation	mg mg^{-1}

*Asterisks indicate parameters that were used for genomic prediction in this study.

distributions for all three leaf parameters, *K*, N_0, and *r*, at either low or high uncertainty levels (40th–60th quantile and 10th–90th quantile, respectively). Original TREES parameters were held invariant across the genotypes.

Empirical evaluations

Baseline experiment

Experiments to characterize variation with respect to leaf growth and developmental rates across *B. rapa* morphotypes were performed at the University of Wyoming (Laramie, WY, USA). Plants were grown in the greenhouse (1800 µmol photons m^{-2} s^{-1} maximum photosynthetic photon flux density (PPFD), 20.75–25.7 °C/20.9–21.3 °C day/night, 26.5–73.8% relative humidity with an average of 47.1% in the autumn of 2017. Three morphotypes of *B. rapa* were evaluated: five replicates of Chinese cabbage (CC), five replicates of vegetable turnip (VT), and 16 replicates of oilseed (R500). Seeds of CC and VT were obtained from the Dutch Crop Genetic Resources Center (CGN) in Wageningen—VT (CGN10995) and CC (CGN06867)—while seeds of R500, *B. rapa* subsp. *trilocularis* (Yellow Sarson), were part of a collection bulked at the University of Wyoming in 2011. All seeds were planted in pots (500 ml) filled with a soil mix [Miracle-Gro Moisture control Potting Mix (20% v/v), Marysville, OH, and Profile Porous Ceramic (PPC) Greens Grade (80% v/v), Buffalo Grove, IL] amended with 2 ml of Osmocote 18-6-12 fertilizer (Scotts, Marysville, OH) per pot. All plants were watered daily to maintain soil volumetric water content (VWC) at 38±5% (ECH_2O; METER Group, Pullman, WA). Individual leaf area was determined throughout the experimental period at least three times per week using Easy Leaf Area software (Easlon and Bloom, 2014) and its standard 4 cm^2 red marker. Photographs were batch processed using the following parameter values: leaf minimum green RBG value (40–50 range), leaf green ratio G/R (1.06), leaf green ratio G/B (1.32), scale minimum red RGB value (0), scale red ratio R/G and R/B (1.48), scale area (4.0 cm^2), processing speed (4), and minimum leaf size (0). Processed photos were inspected visually to ensure adequate green pixel coverage of each leaf. Leaf area data for these morphotypes may be found at https://github.com/DRWang3/leaf_model_TREES_paper (last accessed 6 March 2019).

Mild drought experiment

In the spring of 2018, the mild drought experiment was conducted under greenhouse conditions at the University of Wyoming. Seeds of R500 and CC were planted using the same pot size and soil mix utilized in the baseline experiment of 2017. Plants were randomized and divided into three cohorts: a well-watered control group (WW) and two drought treatments [drought scenario 1 (D1) and drought scenario 2 (D2)] where drought was applied at two different developmental stages for each genotype. In total, the experimental set-up included 288 individual plants (54 D1 + 54 D2 + 36 WW per genotype) to account for necessary destructive sampling of whole plants for leaf water potential throughout the experimental period. In D1 and D2, water was completely withheld at the emergence of leaf four and emergence of leaf five, respectively, and plants were recovered just after emergence of leaf six and seven, respectively. The experimental overview is presented in Supplementary Fig. S2a. Leaf area was measured approximately daily after the emergence of the third leaf using the easy leaf area protocol described for the baseline experiment (Supplementary Dataset S1). Projected canopy photos were taken at the emergence of the third leaf. Nine representative plants were tracked for leaf area in each drought treatment (total 9 replicates×2 genotypes×2 drought groups=36 drought plants) while four plants were tracked in well-watered groups (total 4 replicates×2 genotypes×1 group=8 well-watered plants). The leaf relative rate of expansion (RER) was computed following Granier and Tardieu (1999) as the local slope of the relationship between the natural logarithm of leaf area and thermal time.

Gas exchange and leaf water potential measurements, for computation of hydraulic conductance, were taken twice a day (pre-dawn and midday) at regular time points throughout the experiment such that each genotype by treatment combination was evaluated twice during drought and once during the day after recovery (Supplementary Fig. S2a; Dataset S1). The leaf water potential assay required whole-plant destructive sampling, so at each time point and time of day, six replicates per genotype by treatment

(D1, D2) combination and two well-watered controls per genotype were measured. All measured plants were randomly selected, and sample size was chosen according to successful experiments previously performed on *B. rapa* (Greenham *et al.*, 2017; Guadagno *et al.*, 2017). Pre-dawn and mid-day gas exchange measurements (Long and Bernacchi, 2003) were taken on the youngest expanded leaf. Photosynthetic rate and stomatal conductance were measured, using a portable gas exchange system (LI-COR-6400XT; LI-COR Biosciences, Lincoln, NE). Midday measurements (~10.00 h to 13.00 h) were taken in the conservatory where plants were growing, and environmental conditions in the cuvette matched ambient conditions in the greenhouse. Specifically, the cuvette settings were: flow rate, 300 μmol s^{-1}; CO_2, 400 μmol^{-1} mol air; vapor pressure deficit (VPD), 1.1±0.7 kPa, photosynthetically active radiation (PAR) 600 μmol photon m^{-2} s^{-1}, and block temperature set at 21 °C, with cuvette fan on fast mode. Pre-dawn measurements (~03.00 h to 05.00 h) were taken with the same cuvette settings except in the dark and temperature set at 20 °C; measurements at pre-dawn were taken using a dim green light for visibility. For pre-dawn and midday leaf water potential measurements (Koide *et al.*, 1989), plants were cut with a razor blade at the base of the stem and the main stem was immediately inserted into a Scholander pressure chamber (model 600 Pressure Chamber Instrument; PMS Instrument, Albany, OR). Since the destructive sampling over the course of the experiment decreased the number of plants, individuals were randomized within their genotype by treatment trays after each time point.

Soil VWC (%) (ECH_2O; METER Group) was monitored for all cohorts of plants to maintain the treatments for the duration of the experiment as shown in Supplementary Fig. S2b. This was measured daily in the evening for randomly selected pots within a genotype by treatment tray as well as at each time point for all pots of the plants being measured. A total of 589 VWC observations were taken in this manner over the course of 15 d. Soil VWC after recovery for all treatments was maintained at 38±5% by keeping ~4 cm of water at the bottom of each tray. For both empirical data collections (autumn 2017 and spring 2018), all plants were harvested for above- and below-ground biomass evaluation. In the mild drought experiment, this was assessed at each time point in drought and at three additional days during recovery to assay biomass allocation over time. Above-ground tissue was cut at the soil level with a razor blade, weighed (fresh biomass), oven-dried for 10 d at 65 °C, and weighed again for dry biomass. The same procedure was applied to roots after harvest and manual gentle rinsing with deionized water to eliminate soil and plant debris. Data collected from the mild drought experiment (leaf water potential, biomass, transpiration, leaf area data, and soil moisture) are found in Supplementary Dataset S1 as separate worksheets.

RIL leaf area analysis

Time-series leaf growth metrics (length and width of the second leaf) on the BraIRRi recombinant inbed line (RIL) population (Iniguez-Luy *et al.*, 2009) were obtained from original observations under non-stressed growth conditions from a previous study also conducted at the University of Wyoming (Baker *et al.*, 2015). To derive an estimated leaf area metric, leaves were assumed to have elliptical morphology, so area could be computed as

$$A = 1/4 \times l \times w \times \mathrm{p} \tag{3}$$

where A is leaf area, l is leaf length, and w is leaf width. Thermal time was calculated based on the original study's base temperature of 0.96 °C and meteorological data provided by Baker *et al.* (2015). Hourly temperatures were 15.48–30.61 °C with an average of 21.4 °C, which encompassed the range experienced in the baseline experiment. Relative humidity ranged from 35% to 48% with an average of 42%, which was a smaller range than in the baseline experiment.

Genotype parameterization

Genotype-specific parameters of logistic growth curves that model leaf area response to thermal time for the morphotypes (CC, VT, and R500) and the RIL population were estimated with a hierarchical Bayesian model with genotype and leaf levels. The genotype level was included because we were specifically interested in genotype-specific parameters and the leaf level was included because longitudinal measurements were taken on individual leaves, which may have similar or different behavior. Estimation was carried out using Gibbs sampling implemented in the R package *rjags* (Plummer, 2013). Prior distributions for K and r were weakly informed from empirical means of the morphotypes and RIL population: dnorm(10, 25), dnorm(0.0007, 1e08), respectively, where the first value indicates the mean and the second indicates the precision. For N_0, a prior of dnorm(0.01, 25) was selected based on theoretical knowledge that the initial value of a leaf growth curve should be very small (e.g. 0.01 cm^2). After a burn-in period of 5000 steps, four independent chains were sampled every 50th sample, and 1000 steps were saved per genotype per parameter. Univariate and multivariate scale reduction factors ($\sqrt{\widehat{R}}$ and $\sqrt{\widehat{R_M}}$) were evaluated to check for model convergence, ensuring that values were ~1.0 and never >1.2 (Brooks and Gelman, 1998). Resultant posteriors of K, N_0, and r per genotype were fit to gamma distributions using maximum likelihood estimation implemented in the R package *fitdistrplus* (Delignette-Muller and Dutang, 2015) in order to generate gamma shape and rate parameters used by TREES as parameters for stochastic simulations.

Genomic prediction

Genomic prediction was performed on the BraIRRi RIL population (Iniguez-Luy *et al.*, 2009) for the median, shape, and rate estimates of each leaf growth parameter, K, N_0, and r. A published genome-wide 1482 single nucleotide polymorphism (SNP) data set (Markelz *et al.*, 2017) was utilized for 123 individuals that also had phenotypic data. After quality control, 1481 SNPs with minor allele frequency >0.01 were retained for genomic prediction. A total of eight approaches were analyzed, including a genomic best linear unbiased prediction (GBLUP) model, Bayesian marker regression models, and machine learning models [Reproducing Kernel Hilbert Spaces (RKHS) and Random Forest (RF)]. The models differed in their assumptions of SNP effects and the counts of non-additive genetic effects. Five-fold cross-validation was performed using 50 replicates for each analysis, and accuracy of genomic prediction was computed as the Pearson correlation between genomic breeding values and observed values.

GBLUP

The GBLUP model is:

$$\boldsymbol{y} = \mathbf{1}\,\mu + \boldsymbol{Za} + \boldsymbol{e} \tag{4}$$

where $\boldsymbol{y}$ is the vector of the phenotype; $\mathbf{1}$ is the vector of ones; μ is the overall mean; $\boldsymbol{a}$ is the vector of additive genetic effects accounted for by SNPs; $\boldsymbol{Z}$ is the incidence matrix relating $\boldsymbol{a}$ to $\boldsymbol{y}$; and $\boldsymbol{e}$ is the vector of random residuals. It is assumed that $\boldsymbol{a} \sim \boldsymbol{N}(0, \boldsymbol{G}\sigma^2_a)$ and $\boldsymbol{e} \sim \boldsymbol{N}(0, \boldsymbol{I}\sigma^2_e)$. The $\boldsymbol{G}$ is the genomic relationship matrix constructed from SNPs following method 2 of VanRaden (2008) and σ^2_a is the additive genetic variance explained by SNPs. The analyses using the GBLUP model was performed using the R package *sommer* (Covarrubias-Pazaran, 2016).

Bayesian marker regression models

The Bayesian marker regression models we applied were BayesA (Meuwissen *et al.*, 2001), BayesB (Meuwissen *et al.*, 2001), BayesC (Habier *et al.*, 2011), Bayes Lasso (BL) (Park and Casella, 2008), and Bayes Ridge Regression (BRR) (Hoerl and Kennard, 1970). Models could be generalized as the following:

$$\boldsymbol{y} = \mathbf{1}\,\mu + \boldsymbol{Xg} + \boldsymbol{e} \tag{5}$$

where $\boldsymbol{y}$, $\mathbf{1}$, μ, and $\boldsymbol{e}$ are the same as in the GBLUP model; $\boldsymbol{g}$ is the vector of effects for SNPs; and $\boldsymbol{X}$ is the genotype matrix, which has been centered and scaled for all genotyped individuals. Among four Bayesian marker regression models: BayesA and BL assume that SNP effects $\boldsymbol{g}$ have non-zero effects and the effect variances are drawn from scaled-t and double-exponential distributions, respectively; BayesB assumes that SNP effects $\boldsymbol{g}$ are drawn from a two-component mixture of zero with probability π and a scaled-t distribution; BayesC assumes that SNP effects $\boldsymbol{g}$ are drawn from

a two-component mixture of zero with probability π and a normal distribution; BRR assumes that SNP effects $\boldsymbol{g}$ are all drawn from a normal distribution with the same variance. Each Bayesian model was run as a single chain with a total length of 10 000 Markov chain samples, where the first 1000 iterations were discarded as burn-in. Every fifth sample of the remaining 9000 iterations was saved for the posterior analysis. The analyses using the Bayesian marker regression models were performed using the R package *BGLR* (Pérez and De Los Campos, 2014). Since the genetic background of the traits analyzed was unknown, default settings for hyperparameters were used; for example, in the models BayesB and BayesC, the default rule is set to π_0=0.5 and p_0=10.

Reproducing Kernel Hilbert Spaces

Compared with the GBLUP model, the RKHS model used a Gaussian kernel function to measure the relationships among individuals (Gianola and Van Kaam, 2008). Due to the non-linearity of this Gaussian kernel function, RKHS could capture both additive and non-additive effects (Gianola and Van Kaam, 2008). The analyses using the RKHS model were performed using the R package *BGLR* (Pérez and De Los Campos, 2014). The bandwidth parameter of our analysis was set to 0.7, and the setting for Markov chain was the same as that for Bayesian marker regression models.

Random Forest

Implementations of RF have been shown to capture epistatic effects accurately (Motsinger-Reif *et al.*, 2008; Michaelson *et al.*, 2010; Picotti *et al.*, 2013). Our RF analyses were carried out using the *randomForest* package in R (Liaw and Wiener, 2002). The validation sets were predicted using a collection of regression trees with a subset of the total data set that is bootstrapped over the total data sets. Then the predictions were averaged over all trees. The number of SNPs sampled at each split (mtry) was fine-tuned prior to the analysis using the package function tuneRF (ntreeTry=500, stepFactor=1.5, improve=0.01). RF was then applied with 10 000 trees (ntree).

Model exploration on an in silico *population*

The updated TREES model was examined on a simulated population of 123 individuals, which varied in the leaf growth module parameters *leafAreaMax*, *initialLeafSize*, *leafAreaRate*, *floweringTime*, and *SLA_low*, and informed by the RIL population, BraIRRi (i.e. directly estimated from data). All other parameters were fixed across the population in order to monitor the sensitivity of the model to growth and development constraints under varying environmental conditions. The *in silico* population was simulated under well-watered and water-limited scenarios. Forty non-flowering individuals were examined for ratios (water-limited to well-watered) for the following modeled traits: volumetric soil water content (%), leaf area index (LAI; cm^2 cm^{-2}), canopy-level modeled transpiration (mmol H_2O m^{-2} s^{-1}), maximum potential transpiration (mmol H_2O m^{-2} s^{-1}), photosynthesis (μmol CO_2 m^{-2} s^{-1}), and stomatal conductance (mol H_2O m^{-2} s^{-1}). Three individuals, selected based on *leafAreaMax* (at the 25th, 50th, and 75th quantile of population distribution of *leafAreaMax*), were further simulated using TREES's stochastic mode to compare model results when informed by directly estimated parameters from data versus predicted by GP. Twenty-five stochastic simulations were run per individual per treatment per parameterization method [25 simulations×3 individuals×2 treatments×2 method (predicted versus direct)=300 simulations]. Parameter and drivers can be found at https://github.com/DRWang3/leaf_model_TREES_paper (last accessed 6 March 2019).

Results and discussion

Model development and parameterization

To adapt the process-based model TREES for studying short-lived crop species such as *B. rapa*, we replaced its original plant growth module with a new growth and development routine based on the sequential emergence and expansion of individual leaves. The general organization for this new module hinges on well-studied relationships between developmental rates and thermal time (Wilhelm and McMaster, 1995; Granier and Tardieu, 1998; Pantin *et al.*, 2011) (Fig. 1, step 1), which are governed by equations presented in Supplementary Appendix S1. The 24 parameters of the submodel describe leaf growth functions, temperature-dependent development, and resource allocation, each of which may be gathered from empirical evaluation (Table 1). To streamline a protocol for collecting input parameter data, we conducted a pilot experiment under non-stressed greenhouse growth conditions and characterized time-series growth traits of three *B. rapa* morphotypes of cultivated *B. rapa*, which differ dramatically in life history traits and biomass allocation [oilseed (R500), vegetable turnip (VT), and Chinese cabbage (CC)] (Supplementary Fig. S3). Of these cultivars, R500 was the only one to flower prior to the end of the experiment at 23 850 °Ch (993.75 °Cd) (Supplementary Table S2). CC exhibited a high leaf emergence rate (i.e. a short phyllochron at 1027 °Ch) and large final leaf size at 38.6 cm^2, while VT and R500 displayed similar phyllochron (1980 °Ch and 1971 °Ch, respectively), but R500 grew smaller leaves that took a longer time to expand. The observed relative ranks were consistent with selection pressures that would have been necessary to drive the evolutionary divergence of these morphotypes; for example, quickly emerging and expanding leaves and delayed transition to reproduction would have been favorable features for genotypes such as CC that are cultivated for consumption of vegetative organs. Of note, in all three genotypes, the first two epicotylar leaves were consistently smaller at their maximum size than leaves that developed later (Supplementary Table S3a).

The central function that is computed at each 30 min time step in the updated TREES submodel yields the relative rate of leaf growth under non-stressed conditions, equivalent to the logarithmic derivative of a leaf logistic growth function parameterized by K, N_0, and r (see the Materials and methods; Supplementary Appendix S1). To derive genotype-specific inputs to the model, we estimated logistic leaf growth curve parameters for R500, VT, and CC off-line with the time-series leaf area data collected in the non-stressed baseline experiment (Materials and methods; Supplementary Fig. S4). Estimated posterior distribution of the parameter K, the upper asymptote of the function, was the most differentiated across the three genotypes as compared with N_0 and r, reflections of initial leaf area and growth rate, respectively. Using medians of K, N_0, and r for R500, CC, and VT to constrain the TREES model, we ran simulations for 1125 half-hour time steps (~47 d) and evaluated longitudinal leaf growth. Results were consistent with expected leaf development in the three genotypes, with CC attaining high leaf area quickly and R500 growing small leaves slowly (Supplementary Fig. S3, bottom row).

Effects of trait uncertainty

Most process-based plant models are deterministic; that is, one set of input parameters and environmental data yields a single time-series outcome. Such tools are useful to study system

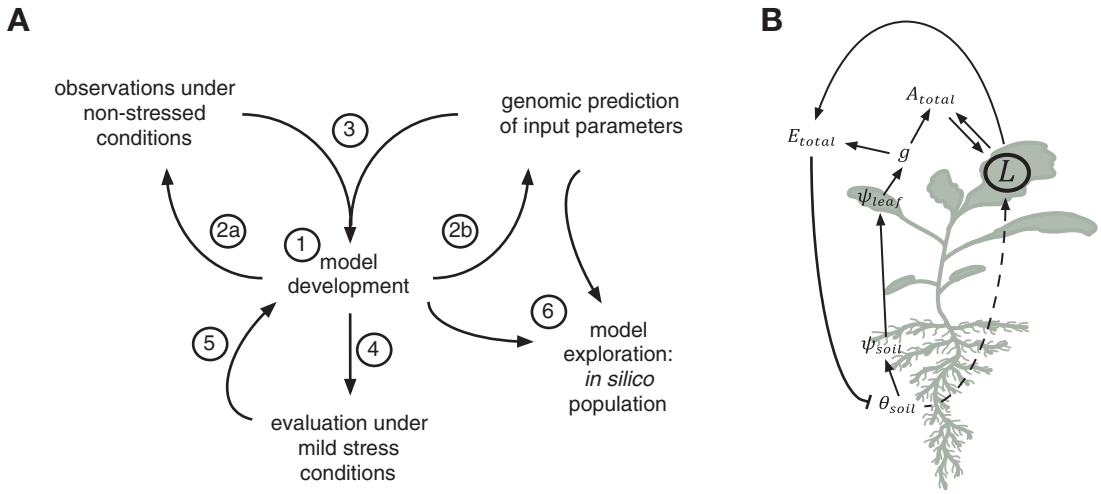

Fig. 1. Study overview. (A) Summary of modeling, experimental, and analysis efforts carried out in this study. (1) A generalized sequential leaf growth submodel is framed based on theoretical knowledge of dicot leaf development and integrated into the whole-plant model, TREES. (2a, b) New model input parameters are tested for ease of direct measurement and for feasibility of indirect estimation using genomic prediction models. (3) Outcomes learned from 2a and b inform updates to model architecture. (4) Leaf growth and physiological traits are evaluated in response to mild drought. (5) A function for modulating leaf growth under water stress is incorporated into the refined model. (6) The updated model is used to explore effects of growth and development variation on physiological response to mild water stress on a simulated population. Details, including plant material used at each stage, are given in Supplementary Table S1. (B) A simplified systems diagram of TREES showing the high-level relationships between state variables of the original TREES model and the new leaf growth module (symbols: L, leaf growth module; A_{total}, photosynthesis; E_{total}, transpiration; g, stomatal conductance; ψ_{leaf}, leaf water potential; ψ_{soil}, soil water potential; θ_{soil}, soil volumetric water content). The dashed arrow represents the empirical function from Stage 5 of (A).

changes in response to time-varying environmental drivers; however, they are incapable of capturing within-genotype stochastic variation that would be useful for comparisons across genotypes over multiple simulation scenarios. In order to model a genotype's phenotypic spectrum that is apparent even under experimental, non-stressed conditions, we asked whether we could inform process model inputs with parameter distributions in lieu of point estimates for leaf development. We fit gamma distributions to the posterior distributions derived from R500, CC, and VT for K, N_0, and r parameters, which resulted in a shape and a rate estimate for each leaf growth parameter per genotype (Supplementary Fig. S5; Table S2). We then modified TREES accordingly to provide an option for internally sampling K, N_0, and r out of distributions parameterized by rate and shape values (Fig. 1, step 3; Supplementary Appendix S1). To explore the effects of general growth parameters and uncertainty in leaf-specific parameters on other traits modeled by TREES, we ran a series of stochastic simulations constrained for R500, VT, and CC growth and development. Simulated genotypes differed in LAI ($cm^2\ cm^{-2}$) (Supplementary Fig. S6a, b), which was expected due to its direct relationship with leaf area, but they also differed in whole-plant hydraulic conductance (mmol $m^{-2}\ s^{-1}\ MPa^{-1}$) (Supplementary Fig. S6c, d) and transpiration traits (mmol $m^{-2}\ s^{-1}$) (Supplementary Fig. S7). Leaf water potential (MPa) was less differentiated overall but diverged in the latter part of the simulations, especially between CC and the other two genotypes (Supplementary Fig. S6e, f). Modeling outcomes suggested that leafy morphotypes such as CC may be more vulnerable to water stress than oeliferous types such as R500 (Supplementary Fig. S7), unless other intrinsic traits (e.g. those related to gas exchange) were concurrently divergent in these genotypes. Increased uncertainty of leaf parameters was observed to affect some traits more than others (a, c, e compared with b, d, f in Supplementary Figs S6 and S7); for example, while actual transpiration did not change appreciably, the time-series distributions of modeled maximum potential transpiration became considerably wider, giving rise to large ranges of the hydraulic safety margin (i.e. the difference between transpiration and maximum potential transpiration) (Sperry *et al.*, 1998). Taken together, our results indicated that variation in growth parameters has the potential to give rise to variation in dynamic physiological responses. This was manifest despite only manipulating variation in growth parameters, highlighting how process-based models can capture internal interactions found inherently within plant systems.

Genomic prediction of model parameters

Dicot leaf expansion, characterized by growth in two dimensions, is challenging to measure longitudinally in an automated fashion, with the exception of plants that form rosettes, such as *Arabidopsis thaliana* (Pantin *et al.*, 2011). We therefore next asked whether genomic prediction can serve as an alternative to inform genotype-specific parameters in our process model, as part of assessing its potential scalability to large germplasm panels (Fig. 1, step 2b). We utilized a previously collected genomic data set (Markelz *et al.*, 2017) (Supplementary Fig. S8a) and time-series morphometrics data (Baker *et al.*, 2015) of the second leaf on a RIL population, BraIRRi, derived from a cross between R500 and Imb211, a fast-cycling genotype (Iniguez-Luy *et al.*, 2009). From the panel, we grew out two contrasting RILs to first evaluate development of the second leaf versus older and younger leaves. The size of the second leaf was observed to be more closely associated with sizes

of the other leaves in these inbred lines than in the cultivars (Supplementary Table S3b), probably due to the early-flowering nature of these RILs. This observation supported the use of the previously collected data on second leaves to inform growth model parameters in TREES for this population. Genomic prediction of K, N_0, and r posterior medians using a genomic best linear unbiased prediction model (GBLUP) were 0.64, 0.42, and 0.31 in accuracy, respectively. Notably, predictions of rate and shape gamma parameters were more accurate than predictions of the medians for N_0 (shape, 0.51; rate, 0.50) and r (shape, 0.59; rate, 0.60) (Fig. 2A). In contrast, for K, whose median already had high prediction accuracy, accuracy decreased slightly (shape, 0.58; rate, 0.60). Simulations of parental leaf growth curves using 100 sets of K, N_0, and r that were sampled from truncated gamma distributions (10th–90th quantiles) parameterized by predicted rate and shape values yielded outcomes that clearly discriminated the two genotypes; Imb211 curves overlapped with observed growth while R500 deviated from observations in the latter part of growth due to underprediction of the K parameter distribution (Fig. 2B–E).

We next compared genomic prediction performance across a suite of eight diverse models including GBLUP, five Bayesian marker regression models (BayesA, BayesB, BayesC, BL, and BRR), and two machine learning models (RF and RKHS). GBLUP and Bayesian regression models were selected because they are widely used in genomic prediction for plant and animal breeding, while the machine learning models were included for their potential to capture non-additive genetic effects. Genomic prediction accuracy is influenced by trait genetic architecture, which can be broadly defined in two dimensions: heritability (proportion of genetic variance to phenotypic variance) and complexity (number of genes controlling the phenotype). Overall, leaf model parameters showed moderate heritability (0.34–0.60 mean estimates across methods), and estimates were higher for the rate and shape parameters of N_0 and r as compared with their medians (Supplementary Table S4). Prediction accuracies averaged across models ranged from 0.32 for r (median) to 0.63 for K (median) and largely reflected their heritability ranks (Table 2; Supplementary Fig. S8b–l). The five Bayesian marker regression models yielded similar prediction accuracies to GBLUP; these models perform similarly for traits of moderate complexity (Wang *et al.*, 2018), suggesting moderate complexity of leaf growth parameters. To confirm the moderate complexity of leaf growth parameters, genome-wide association studies (McCarthy *et al.*, 2008) could be carried out in future studies. RF, which has been previously shown to uncover epistatic effects, was found here to outperform Bayesian marker regression models and GBLUP for N_0 (median) and r (median) but fared worse for other parameters. Overall, RKHS was favored, consistently achieving the most accurate prediction across all tested parameters, suggesting that non-additive genetic effects may play important roles for determining leaf growth traits. This was especially notable in predictions for rate and shape of N_0 where RKHS yielded accuracies 6% greater than other models. Interestingly, there was an appreciable increase in accuracy for rate and shape parameters of N_0 and r distributions versus medians observed across all evaluated methods; this result provided additional support to modify TREES architecture to incorporate within-genotype uncertainty using rate and shape parameters (Fig. 1, step 3).

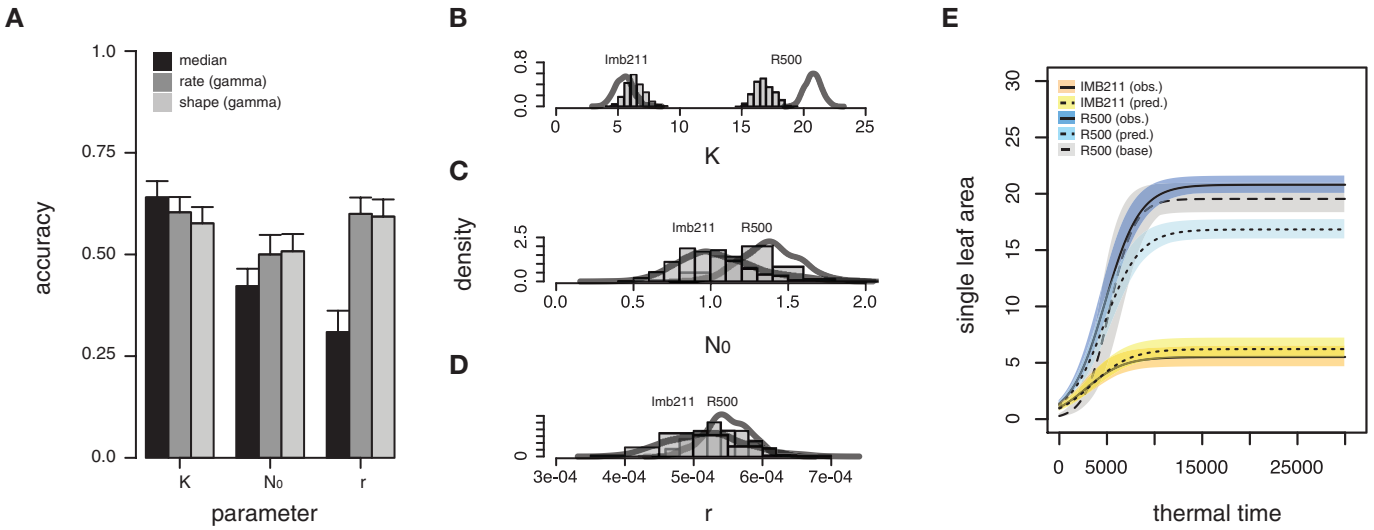

Fig. 2. Genomic prediction of leaf growth parameters using a genomic best linear unbiased prediction (GBLUP) model. (A) Means from 5-fold cross-validation for predicting either single descriptors (medians) or distribution parameters of K, N_0, and r posteriors in the Imb211×R500 recombinant inbred line (RIL) population. The bar shows the 95% confidence interval of accuracy from 50 replicates of 5-fold cross-validation. Boxplots that show result distributions are given in Supplementary Fig. S8. Accuracy is measured as the Pearson correlation between genomic breeding values and observed values. (B–D) Gamma distributions of leaf growth parameters (K, N_0, and r) of parental lines R500 and Imb211 parameterized by rate and shape values either directly estimated (density curves) or predicted with ridge regression (histograms). Each density curve and histogram represents 500 samples. (E) Simulated leaf growth curves of R500 and Imb211 resulting from sampling from K, N_0, and r truncate gamma distributions (10th–90th quantiles) parameterized by rate and shape values either directly estimated ('obs.') or predicted using genomic prediction ('pred.'). Also shown is simulated R500 leaf growth curves following the same sampling scheme using parameters directly estimated from the baseline experiment ('base'). Lines indicate the mean of the samples, and shaded regions indicate the range of 100 curves per class.

Table 2. *Comparison of accuracy across genomic prediction methods to predict leaf growth model parameters*

Trait	BayesA	BayesB	BayesC	BL	BRR	GBLUP	Random Forest	RKHS	Mean
K	0.64	0.63	0.64	0.64	0.64	0.64	0.57	**0.65**	0.63
K_rate	0.6	0.6	0.6	0.6	0.6	0.6	0.55	**0.64**	0.6
K_shape	0.57	0.57	0.58	0.57	0.58	0.58	0.52	**0.61**	0.57
N_0	0.42	0.42	0.42	0.42	0.42	0.42	**0.45**	**0.46**	0.43
N_0*_rate*	0.5	0.49	0.5	0.5	0.5	0.5	0.48	**0.56**	0.5
N_0*_shape*	0.5	0.5	0.51	0.5	0.5	0.51	0.5	**0.57**	0.51
r	0.31	0.31	0.31	0.31	0.31	0.31	**0.34**	**0.35**	0.32
r_rate	0.6	0.59	0.59	0.59	0.6	0.6	0.57	**0.63**	0.6
r_shape	0.59	0.59	0.59	0.59	0.59	0.59	0.56	**0.63**	0.59
Mean	0.52	0.52	0.53	0.52	0.53	0.53	0.51	0.57	0.53

Bold numbers indicate the best-performing method(s) for each parameter. Accuracy is measured as the Pearson correlation between genomic breeding values and observed values, and represent means from 50 replications of 5-fold cross-validation.

These empirical results supported the use of genomic prediction to inform genotype-specific leaf growth parameters as inputs into the updated TREES model.

Leaf area modulation by mild drought

While empirical evaluation and genomic prediction provide critical information on growth progression under non-stressed conditions, the value of first principles-based process modeling lies in extending its predictive capabilities to novel environmental conditions, which may include changing environmental stressors such as transient water limitation. TREES is one of the few process-based models that couples rhizosphere and xylem hydraulics with carbon processes and has been shown to predict water stress response accurately across species such as aspen (*Populus tremuloides*), honey mesquite (*Prosopis glandulosa*), juniper (*Juniperus ashei*), and Texas persimmon (*Diospyros texana*) (Mackay *et al.*, 2015; Tai *et al.*, 2017; Johnson *et al.*, 2018). Because there is evidence for the direct control of hydraulics on leaf expansion (Pantin *et al.*, 2011; Caldeira *et al.*, 2014), we set out to refine our growth model to capture leaf area modulation in response to water limitation in *B. rapa* by conducting a mild drought and recovery experiment using the genotypes CC and R500 (Fig. 1, step 4). We imposed two staggered treatments of water withholding (D1 and D2), with their timings tailored according to leaf emergence events across the two genotypes (see the Materials and methods and Supplementary Fig. S2a for an experimental overview). Although the water stress imposed was mild and only lasted between 5 d and 6 d, droughted plants displayed visible loss of turgor prior to recovery (Supplementary Fig. S9) and mean biomass allocation had shifted to below ground under both drought treatments in each genotype (Supplementary Table S5). The relative rate of leaf expansion (RER) showed a depression during drought followed by a rapid increase after recovery (Fig. 3; Supplementary Fig. S10).

Using the data collected from the mild drought experiment, we sought to integrate functions into the TREES leaf growth module that could allow it to model leaf expansion under water-constrained conditions (Fig. 1, step 5). Due to high variability and limited observations of hydraulic conductance (see Supplementary Dataset S1 for leaf water potential and transpiration measurements), we opted instead to link the modulation of leaf area expansion to a measure of relative water status in the soil. This variable expressed a consistent signature across treatments (Supplementary Fig. S2b), and the empirical function related a soil water content ratio to a coefficient that modulates the relative leaf area expansion rate (Supplementary Appendix S1). To explore the updated whole-plant model, we simulated the behavior of genotypes CC and R500 under their respective growth conditions beginning at the initiation of leaf 3 and compared modeled results with observed daily 'slow' traits (leaf area and volumetric soil water content) along with pre-dawn and midday 'fast' traits (leaf water potential and leaf-level measures of photosynthesis, stomatal conductance, and transpiration) (Fig. 4; Supplementary Figs. S11–S13). Overall, the model performed better for genotype CC than for R500 in nearly all traits, especially matching CC soil water content and leaf area progression well (root mean square errors of 3.02% and 4.61 cm^2). This is likely to indicate the need for better genotype-specific model parameterization for parameters that we had assumed to be species-level values due to lack of prior cultivar-level information; the results suggest that those values may have been better suited for CC than R500. For leaf water potential, a challenging trait to monitor on tender herbaceous crops, noise found in the observed data in both genotypes led to a lack of strong differentiation between pre-dawn and midday values (Supplementary Figs S11c, S12e) while in the model there was always a clear drop at midday due to the dependence of leaf water potential on hydraulics, leading to poorly matched modeled results for this trait. During periods of water deficit, the relative expansion rate of simulated leaf growth (Supplementary Fig. S13) followed dynamics similar to what was observed (Fig. 3; Supplementary Fig. S10). However, the empirical function we incorporated did not account for observed water-limited RER surpassing well-watered RER upon recovery, suggesting a knowledge gap in the mechanism of leaf expansion response to

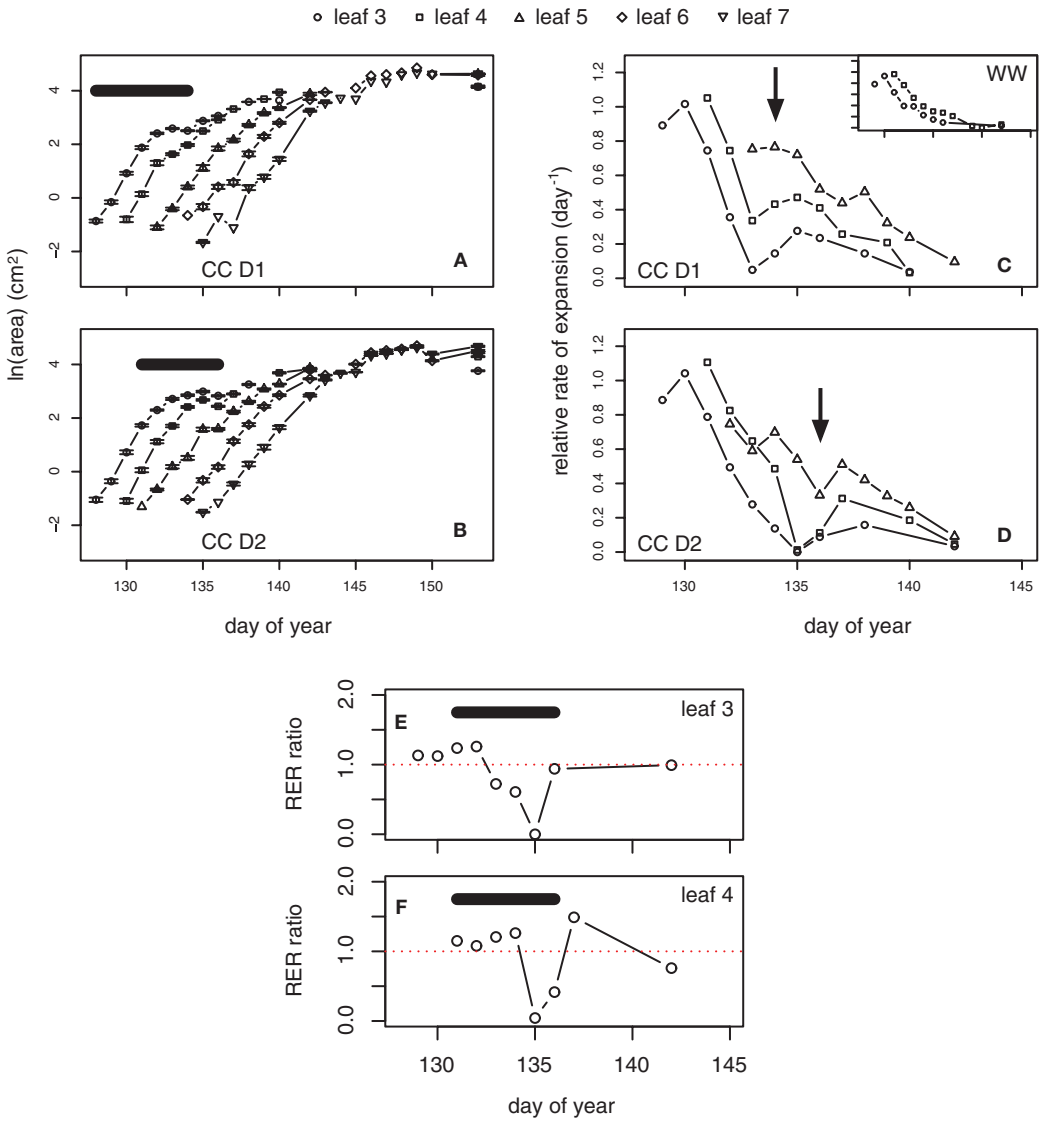

Fig. 3. Leaf growth under mild drought. (A and B) Leaf area and (C and D) mean relative expansion rate of CC plants in response to mild water deficit and recovery. Error bars in (A) and (B) indicate the standard error of the mean. Relative expansion rates (RER) in (C) and (D) were computed using means from (A) and (B). (E and F) Ratio of the relative expansion rate (RER) in leaf 3 and leaf 4 of CC drought scenario 2 (D2) relative to CC well-watered treatment (WW); dotted line at 1.0 indicates equal RER in WW and D2. Arrows in (C and D) indicate onset of re-watering. Bars in (A), (B), (E), and (F) indicate periods of water limitation.

water recovery that may allow mildly droughted plants to 'catch up' with well-watered plants. Overall, the modeling link between soil water deficit and leaf expansion modulation serves as a functional placeholder until more data can be gathered to uncover how leaf RER responds to plant hydraulics.

A process-based modeling framework for decomposing G×E

While the new leaf growth module will benefit from future improved mechanistic understanding of drought response, its capacity to account for genotypic variation and within-genotype stochasticity afforded the opportunity to examine how process-based models might be useful for studying whole-plant genotype by environment interaction (G×E) (Fig. 1, step 6). As a modeling exercise, we explored the behavior of a simulated population of *B. rapa* under 17 d well-watered and water-limited scenarios (Supplementary Fig. S14). Individuals (genotypes) of this *in silico* population varied for five leaf module parameters (*leafAreaMax*, *initialLeafSize*, *leafAreaRate*, *floweringTime*, and *SLA_low*) associated with phenology and growth that were informed (i.e. directly estimated from data) by a real *B. rapa* population, BraIRRi, and fixed for all other parameters. The water-limited scenario differed from the well-watered environment by having a 6 d window of water deficit about halfway through the simulation (Supplementary Fig. S15).

We examined resulting trait ratios (water-limited to well-watered) of individuals in the population. Final ratios for LAI ranged from 0.787 to 0.931. This variation was largely associated with phenology differences in the set of genotypes that had transitioned to reproductive phase prior to the end of the simulation, with a correlation of –0.70 (P-value $<1.34e10^{-13}$) (Supplementary Fig. S16). Due

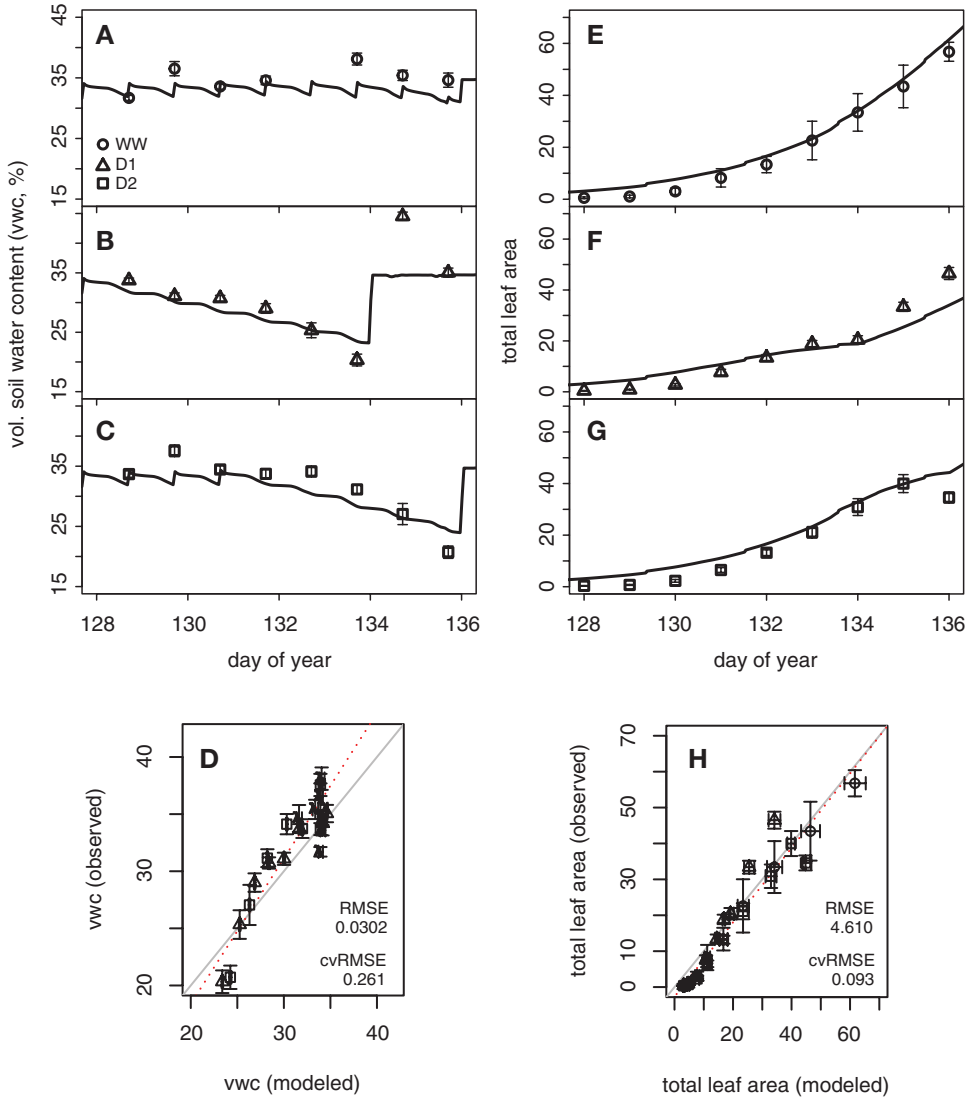

Fig. 4. Confrontation of the model with observed soil water content and leaf area. TREES simulations parameterized with genotype CC leaf growth and forced with environmental data collected in the mild drought experiment. Time-series of soil volumetric water content (%) (A–C) and total leaf area (cm^2) (E–G). In one-to-one plots (D, H), solid lines represent the one-to-one line while dotted lines indicate best fit. Observed points are means, and error bars show their SEs. Each modeled value shown in (D) and (H) is averaged from a 12 h time interval corresponding to when the daily observations were made, and bars indicate value range in the 12 h intervals. Circle, WW (well-watered); triangle, D1 (drought scenario 1); square, D2 (drought scenario 2). cvRMSE refers to the mean normalized root mean square error.

to this association, we restricted the following examination to the 40 individuals that did not flower prior to the end of simulation. For volumetric soil water content, LAI, and transpiration traits, behavior was consistent across this population subset (Fig. 5A, B; Supplementary Figs S17–S21). In contrast, modeled canopy-level photosynthesis and canopy-level stomatal conductance exhibited behavior that diverged across this population subset (Fig. 5C, D; Supplementary Figs S22, S23). This divergent behavior was most apparent for stomatal conductance. After onset of recovery (days 7–13 in Fig. 5), ratio ranges clearly encompassed the value 1.0, meaning that while some genotypes of the water-limited group had greater values after drought recovery than their well-watered counterparts, others had lesser values. The genotype rankings themselves were also dynamic, with some genotypes having >1.0 values at some time points in the recovery but <1.0 at others (Fig. 5D). Rhizosphere–xylem hydraulic interactions modeled in TREES can result in lower photosynthesis rates in high leaf area plants at high soil moisture due to lower root areas limiting water supply. This mechanism in the model would explain why some genotypes have lower canopy-level photosynthesis in well-watered versus water-limited treatment upon recovery despite having greater LAI.

These results were emergent properties of the system; in other words, outcomes of interacting model functions dependent on different combinations of variables integrated over time. While emergent properties are general phenomena of modeling complex systems, the specific example that we report here as emergent is a case of G×E, namely the differential response across genotypes to the environmental scenario (well-watered versus water-limited in this case). By exploring the updated model using an *in silico* population as opposed to a real population, we are able to safely conclude that the G×E we observed resulted

from emergent properties of the model rather than other biological complexities characteristic of plant populations that are often challenging to account for (e.g. transgressive genetic variation). Additional stochastic simulations using three individuals contrasting in maximum leaf size (see the Materials and methods) supported genomic prediction as a means to parameterize growth in whole-plant models for modeling across environmental scenarios (Supplementary Fig. S24). Taken together, the results supported the updated model's capacity to leverage genotypic-level differences in growth to drive intraspecific variability in response to environmental factors.

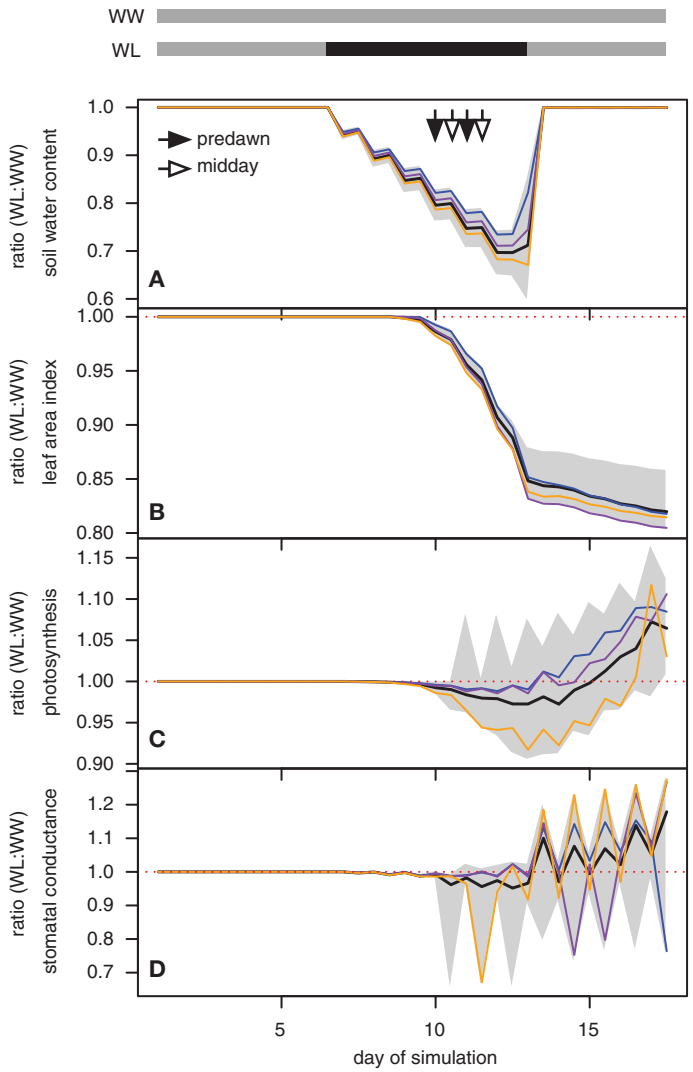

Fig. 5. Behavior of a simulated *B. rapa* population under well-watered (WW) and water-limited (WL) scenarios. The 40 individuals of the *in silico* population that remained in the vegetative phase for the entirety of the simulation period are compared for their behavior under WW and WL conditions (Supplementary Fig. S15). Watering regimes for each scenario are shown as gray and black bars at the top of the figure (gray indicates daily water input; black indicates no water input). Pre-dawn (05.00 h time step) and midday (13.00 h time step) ratios of WW to WL for (A) soil water content (B) leaf area index, (B) canopy-level photosynthesis, and (C) canopy-level stomatal conductance are shown here. The shaded gray regions show the range of values across the 40 individuals; the solid black curves depict the mean. The dotted line indicates 1.0, where WW and WL have the same value. Blue, purple, and orange lines depict three different individuals out of the 40.

Previously, TREES utilized functions that depended on environmental variables (e) and/or population-level parameters (p): generically $f(e)$, $f(p)$, $f(e,p)$. In the current study, additional layering of inter- and intragenotypic variation allows for four supplemental relationship types constrained by genotype-level parameters (g): $f(g)$, $f(p,g)$, $f(g,e)$, and $f(p,e,g)$. By formalizing the components that can give rise to emergent G×E, this update represents a theoretical advancement in the TREES model as it allows the dimension of time (and correlated environment) to interface with genetic variation. In the model, the relationship between genotype-specific leaf growth and temperature is an example of $f(g,e)$ where g are the leaf growth parameters and e is thermal time; the new empirical modulation of relative leaf area by relative soil water content represents an $f(e)$ function; computation of instantaneous LAI depends on functions that use g (leaf growth parameters) and p (parameter value for leaf angle fixed in the population for this case) (Supplementary Appendix S1). Exploration of the *in silico* population provides evidence that the interaction of these and other new functions with original TREES routines integrated over time can give rise to whole-plant G×E (Fig. 6). Extending the modeling to real plant populations with proper genotype-level parameterization on all parts of the model (growth and development, nutrient cycling, carbon processes, and hydraulics) is needed to explore the predictive capacities of the model at-scale under stressed conditions. On the whole, this work suggests that the systematic identification, development, and testing of modeling components that can be refined by accounting for g may be key to utilizing process-based models for making predictions at-scale and identifying contributing elements underlying whole-plant G×E.

Conclusion

In this study, we introduced and tested the various components of a framework aimed at finding the modeling connections between genomics-informed design and environment-driven biophysics. First, we overhauled the development module within TREES to adapt it for short-lived crop plants and modified model architecture to handle across- and within-genotype variation. Then we showed that genomic prediction offers a means to indirectly inform model input parameters for deterministic and stochastic simulations. Finally, we demonstrated the utility of a new empirical relationship in TREES to modulate the expansion of leaf growth under water-limited scenarios. The assemblage of experiments, analyses, and model updates laid out here represents a 'staging area' that supports a long-term goal of equipping plant process-based models with the ability to simulate large intraspecific populations. As technological advancements in plant phenotyping improve our ability to gather temporally resolved data on large germplasm panels, models that can integrate natural genetic variation with physiological response will be key to harnessing this flood of information for application across both basic and translational domains.

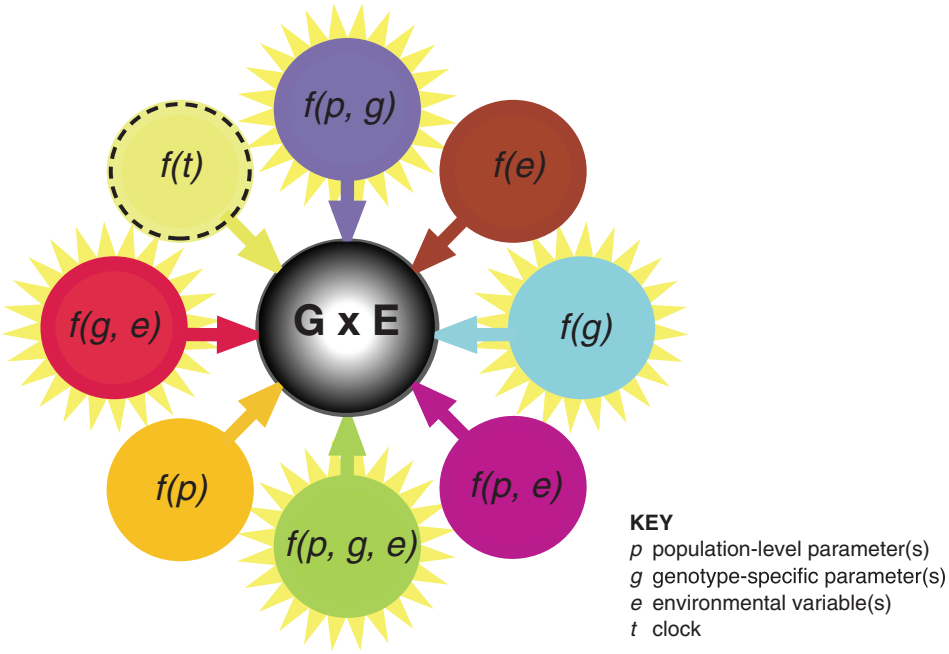

Fig. 6. Decomposing genotype by environment interaction (G×E) with a biophysical framework. Genomics-informed process-based models are capable of formalizing the subprocesses that can give rise to the observed G×E. G×E at the whole-plant level, an outcome of integration across multiple dimensions, is represented here by a '3D' sphere. '2D' circles are mathematical functions within the process model that are computed at each time step of simulations. Not depicted are the routines that tie these various functions together; these may be thought of as single- and double-headed arrows that link the 2D circles with each other. In this conceptual figure, the dimension of time extends straight out towards the reader, with multiple planes of circles stacked in layers on top of each other through time. Model state variables, which are carried over throughout a simulation from one time step to the next, relate one layer to the next one above it. Starbursts mark model functions that utilize genotype-specific parameters, representing the parts of the model that can refined to account for genotypic variation. The *f*(*t*) circle with a dashed outline is a function of the clock (i.e. circadian rhythm), another area for future model improvement.

Supplementary data

Supplementary data are available at *JXB* online.

Fig. S1. A pipeline to inform leaf growth input parameters of the TREES canopy growth submodel.

Fig. S2. Evaluation of *B. rapa* genotypes under mild drought.

Fig. S3. Vegetative leaf growth variation across *Brassica rapa* morphotypes.

Fig. S4. Posterior parameter distributions of leaf growth curves across *B. rapa* morphotypes.

Fig. S5. Stochastic leaf growth curves for CC.

Fig. S6. Using genotype-specific stochastic parameters to inform whole-plant simulations.

Fig. S7. Transpiration outcomes in simulations with stochastic leaf growth.

Fig. S8. Genomic prediction of leaf growth model parameters.

Fig. S9. Recovery of droughted plants.

Fig. S10. R500 leaf growth under mild drought.

Fig. S11. Simulated versus observed traits in genotype CC.

Fig. S12. Simulated versus observed traits in genotype R500.

Fig. S13. Modulation of relative rate of leaf expansion in response to water limitation.

Fig. S14. Model exploration using a simulated population.

Fig. S15. Environmental drivers used for modeling population responses to water deficit.

Fig. S16. Relationship of final leaf area index ratio (water-limited to well-watered) and flowering time in simulated population.

Fig. S17. Transpiration of a simulated *B. rapa* population under well-watered and water-limited scenarios.

Fig. S18. Heat map of volumetric soil water content ratios (WL:WW) across simulation time.

Fig. S19. Heat map of leaf area index ratios (WL:WW) across simulation time.

Fig. S20. Heat map of midday transpiration ratios (WL:WW) across simulation time.

Fig. S21. Heat map of midday maximum transpiration ratios (WL:WW) across simulation time.

Fig. S22. Heat map of midday photosynthesis ratios (WL:WW) across simulation time.

Fig. S23. Heat map of midday stomatal conductance ratios (WL:WW) across simulation time.

Fig. S24. Comparison of simulation results informed by directly estimated parameters versus informed by predicted parameters.

Table S1. Summary table expansion of Fig. 1a.

Table S2. Growth parameter values of *B. rapa* morphotypes used for TREES modeling.

Table S3. Maximum leaf sizes across leaf number.

Table S4. Genomic heritability of leaf growth parameters.

Table S5. Observed ratio of shoot to root under mild drought.

Dataset S1. Data collected from the mild drought experiment.

Appendix S1. TREES leaf growth submodel.

Acknowledgements

We thank Tim L. Setter for background resources that enabled initial conceptualization for the sequential leaf growth routine, and James Berry for growth chamber access. We are grateful to Marta Szumski, Chris Nieters, Sarah Lemli Weber, Hunter Peterson, Rachel Shrode, Keegan Ferris, and Daniel Beverly for technical assistance. This study was funded by the National Science Foundation Division of Integrative Organismal Systems (NSF-IOS) #1547796.

References

Alberda T, Sibma L. 1968. Dry matter production and light interception of crop surfaces: 3. Actual herbage production in different years as compared with potential values. Grass and Forage Science **23,** 206–215.

Baker RL, Leong WF, Brock MT, Markelz RJ, Covington MF, Devisetty UK, Edwards CE, Maloof J, Welch S, Weinig C. 2015. Modeling development and quantitative trait mapping reveal independent genetic modules for leaf size and shape. New Phytologist **208,** 257–268.

Bouman BAM, Van Keulen H, Van Laar HH, Rabbinge R. 1996. The 'School of de Wit' crop growth simulation models: a pedigree and historical overview. Agricultural Systems **52,** 171–198.

Brooks SP, Gelman A. 1998. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics **7,** 434–455.

Caldeira CF, Jeanguenin L, Chaumont F, Tardieu F. 2014. Circadian rhythms of hydraulic conductance and growth are enhanced by drought and improve plant performance. Nature Communications **5,** 5365.

Cooper M, Technow F, Messina C, Gho C, Totir LR. 2016. Use of crop growth models with whole-genome prediction: application to a maize multi-environment trial. Crop Science **56,** 2141–2156.

Covarrubias-Pazaran G. 2016. Genome-assisted prediction of quantitative traits using the R package sommer. PLoS One **11,** e0156744.

Delignette-Muller ML, Dutang C. 2015. fitdistrplus: an R package for fitting distributions. Journal of Statistical Software **64,** 1–34.

de Wit CT, Brouwer R, De Vries FWTP. 1970. The simulation of photosynthetic systems. Proceedings of the IBP/PP technical meeting, Trebon, 14–21 September 1969. 47–70.

Easlon HM, Bloom AJ. 2014. Easy Leaf Area: automated digital image analysis for rapid and accurate measurement of leaf area. Applications in Plant Science **2**, pii: apps.1400033.

Gianola D, van Kaam JB. 2008. Reproducing kernel hilbert spaces regression methods for genomic assisted prediction of quantitative traits. Genetics **178,** 2289–2303.

Granier C, Tardieu F. 1998. Is thermal time adequate for expressing the effects of temperature on sunflower leaf development? Plant, Cell & Environment **21,** 695–703.

Greenham K, Guadagno CR, Gehan MA, Mockler TC, Weinig C, Ewers BE, McClung CR. 2017. Temporal network analysis identifies early physiological and transcriptomic indicators of mild drought in *Brassica rapa*. eLife **6,** e29655.

Guadagno CR, Ewers BE, Speckman HN, Aston TL, Huhn BJ, DeVore SB, Ladwig JT, Strawn RN, Weinig C. 2017. Dead or alive? Using membrane failure and chlorophyll a fluorescence to predict plant mortality from drought. Plant Physiology **175,** 223–234.

Habier D, Fernando RL, Kizilkaya K, Garrick DJ. 2011. Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics **12,** 186.

Hammer GL, Chapman S, Van Oosterom E, Podlich DW. 2005. Trait physiology and crop modelling as a framework to link phenotypic complexity to underlying genetic systems. Australian Journal of Agricultural Research **56,** 947–960.

Hammer G, Cooper M, Tardieu F, Welch S, Walsh B, van Eeuwijk F, Chapman S, Podlich D. 2006. Models for navigating biological complexity in breeding improved crop plants. Trends in Plant Science **11,** 587–593.

Heffner EL, Jannink J-L, Iwata H, Souza E, Sorrells ME. 2011. Genomic selection accuracy for grain quality traits in biparental wheat populations. Crop Science **51,** 2597–2606.

Hoerl AE, Kennard RW. 1970. Ridge regression: biased estimation for nonorthogonal problems. Technometrics **12,** 55–67.

Hoogenboom G. 2000. Contribution of agrometeorology to the simulation of crop production and its applications. Agricultural and Forest Meteorology **103,** 137–157.

Iniguez-Luy FL, Lukens L, Farnham MW, Amasino RM, Osborn TC. 2009. Development of public immortal mapping populations, molecular markers and linkage maps for rapid cycling *Brassica rapa* and *B. oleracea*. Theoretical and Applied Genetics **120,** 31–43.

Johnson DM, Domec J, Berry ZC, Schwantes AM, Woodruff DR, McCulloh KA, Polley HW, Wortemann R, Swenson JJ, Mackay DS. 2018. Co-occurring woody species have diverse hydraulic strategies and mortality rates during an extreme drought. Plant, Cell & Environment **41,** 576–588.

Jones JW, Hoogenboom G, Porter CH, Boote KJ, Batchelor WD, Hunt LA, Wilkens PW, Singh U, Gijsman AJ, Ritchie JT. 2003. The DSSAT cropping system model. European Journal of Agronomy **18,** 235–265.

Keating BA, Carberry PS, Hammer GL, Probert ME, Robertson MJ, Holzworth D, Huth NI, Hargreaves JNG, Meinke H, Hochman Z. 2003. An overview of APSIM, a model designed for farming systems simulation. European Journal of Agronomy **18,** 267–288.

Koide RT, Robichaux RH, Morse SR, Smith CM. 1989. Plant water status, hydraulic resistance and capacitance. In: Pearcy RW, Ehleringer JR, Mooney HA, Rundel PW, eds. Plant physiological ecology. Dordrecht: Springer, 161–183.

Liaw A, Wiener M. 2002. Classification and regression by random forest. R News **2,** 18–22.

Long SP, Bernacchi CJ. 2003. Gas exchange measurements, what can they tell us about the underlying limitations to photosynthesis? Procedures and sources of error. Journal of Experimental Botany **54,** 2393–2401.

Mackay DS, Roberts DE, Ewers BE, Sperry JS, McDowell NG, Pockman WT. 2015. Interdependence of chronic hydraulic dysfunction and canopy processes can improve integrated models of tree response to drought. Water Resources Research **51,** 6156–6176.

Markelz RJC, Covington MF, Brock MT, Devisetty UK, Kliebenstein DJ, Weinig C, Maloof JN. 2017. Using RNA-seq for genomic scaffold placement, correcting assemblies, and genetic map creation in a common *Brassica rapa* mapping population. G3 (Bethesda, Md.) **7,** 2259–2270.

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nature Reviews. Genetics **9,** 356–369.

McDowell NG, Williams AP, Xu C, Pockman WT, Dickman LT, Sevanto S, Pangle R, Limousin J, Plaut J, Mackay DS. 2016. Multi-scale predictions of massive conifer mortality due to chronic temperature rise. Nature Climate Change **6,** 295.

Messina CD, Jones JW, Boote KJ, Vallejos CE. 2006. A gene-based model to simulate soybean development and yield responses to environment. Crop Science **46,** 456–466.

Messina CD, Sinclair TR, Hammer GL, Curan D, Thompson J, Oler Z, Gho C, Cooper M. 2015. Limited-transpiration trait may increase maize drought tolerance in the US Corn Belt. Agronomy Journal **107,** 1978–1986.

Meuwissen TH, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics **157,** 1819–1829.

Michaelson JJ, Alberts R, Schughart K, Beyer A. 2010. Data-driven assessment of eQTL mapping methods. BMC Genomics **11,** 502.

Minamikawa MF, Nonaka K, Kaminuma E, *et al*. 2017. Genome-wide association study and genomic prediction in citrus: potential of genomics-assisted breeding for fruit quality traits. Scientific Reports **7,** 4721.

Motsinger-Reif AA, Reif DM, Fanelli TJ, Ritchie MD. 2008. A comparison of analytical methods for genetic association studies. Genetic Epidemiology **32,** 767–778.

Okeke UG, Akdemir D, Rabbi I, Kulakow P, Jannink JL. 2017. Accuracies of univariate and multivariate genomic prediction models in African cassava. Genetics, Selection, Evolution **49,** 88.

Pantin F, Simonneau T, Rolland G, Dauzat M, Muller B. 2011. Control of leaf expansion: a developmental switch from metabolics to hydraulics. Plant Physiology **156,** 803–815.

Park T, Casella G. 2008. The Bayesian Lasso. Journal of the American Statistical Association **103,** 681–686.

Pérez P, de los Campos G. 2014. Genome-wide regression and prediction with the BGLR statistical package. Genetics **198,** 483–495.

Picotti P, Clément-Ziza M, Lam H, *et al*. 2013. A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. Nature **494,** 266–270.

Plummer M. 2013. rjags: Bayesian graphical models using MCMC. R Package version 3.

Rutkoski J, Singh RP, Huerta-Espino J, Bhavani S, Poland J, Jannink JL, Sorrells ME. 2015. Genetic gain from phenotypic and genomic selection for quantitative resistance to stem rust of wheat. The Plant Genome **8,** doi: 10.3835/plantgenome2014.10.0074

Schaeffer LR. 2006. Strategy for applying genome-wide selection in dairy cattle. Journal of Animal Breeding and Genetics **123,** 218–223.

Sinclair TR, Messina CD, Beatty A, Samples M. 2010. Assessment across the United States of the benefits of altered soybean drought traits. Agronomy Journal **102,** 475–482.

Sperry JS, Adler FR, Campbell GS, Comstock JP. 1998. Limitation of plant water use by rhizosphere and xylem conductance: results from a model. Plant, Cell & Environment **21,** 347–359.

Spindel J, Begum H, Akdemir D, Virk P, Collard B, Redoña E, Atlin G, Jannink JL, McCouch SR. 2015. Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. PLoS Genetics **11,** e1004982.

Struik PC. 2012. Modelling gene–trait–crop relationships: past experiences and future prospects. Acta Horticulturae **957,** 181–189.

Tai X, Mackay DS, Anderegg WR, Sperry JS, Brooks PD. 2017. Plant hydraulics improves and topography mediates prediction of aspen mortality in southwestern USA. New Phytologist **213,** 113–127.

Tardieu F, Simonneau T, Muller B. 2018. The physiological basis of drought tolerance in crop plants: a scenario-dependent probabilistic approach. Annual Review of Plant Biology **69,** 733–759.

Technow F, Messina CD, Totir LR, Cooper M. 2015. Integrating crop growth models with whole genome prediction through approximate Bayesian computation. PLoS One **10,** e0130855.

Thavamanikumar S, Dolferus R, Thumma BR. 2015. Comparison of genomic selection models to predict flowering time and spike grain number in two hexaploid wheat doubled haploid populations. G3 (Bethesda, Md.) **5,** 1991–1998.

Thorp KR, Hunsaker DJ, Bronson KF, Andrade-Sanchez P, Barnes EM. 2017. Cotton irrigation scheduling using a crop growth model and FAO-56 methods: Field and simulation studies. Transactions of the ASABE **60,** 2023–2039.

VanRaden PM. 2008. Efficient methods to compute genomic predictions. Journal of Dairy Science **91,** 4414–4423.

Wang J, Zhou Z, Zhang Z, Li H, Liu D, Zhang Q, Bradbury PJ, Buckler ES, Zhang Z. 2018. Expanding the BLUP alphabet for genomic prediction adaptable to the genetic architectures of complex traits. Heredity **121,** 648–662.

White JW, Hoogenboom G. 1996. Simulating effects of genes for physiological traits in a process-oriented crop model. Agronomy Journal **88,** 416–422.

White JW, Hoogenboom G. 2003. Gene-based approaches to crop simulation. Agronomy Journal **95,** 52–64.

Wilhelm WW, McMaster GS. 1995. Importance of the phyllochron in studying development and growth in grasses. Crop Science **35,** 1–3.

Wolfe MD, Del Carpio DP, Alabi O, Ezenwaka LC, Ikeogu UN, Kayondo IS, Lozano R, Okeke UG, Ozimati AA, Williams E. 2017. Prospects for genomic selection in cassava breeding. Plant Genome **10,** doi: 10.3835/plantgenome2017.03.0015.

Yin X. 2000. A generic equation for nitrogen-limited leaf area index and its application in crop growth models for predicting leaf senescence. Annals of Botany **85,** 579–585.

Yin X, Kropff MJ, Stam P. 1999. The role of ecophysiological models in QTL analysis: the example of specific leaf area in barley. Heredity **82 Pt 4,** 415–421.

Yin X, Struik PC, van Eeuwijk FA, Stam P, Tang J. 2005. QTL analysis and QTL-based prediction of flowering phenology in recombinant inbred lines of barley. Journal of Experimental Botany **56,** 967–976.

Yin X, van der Linden CG, Struik PC. 2018. Bringing genetics and biochemistry to crop modelling, and vice versa. European Journal of Agronomy **100,** 132–140.