PERSPECTIVE

Multiscale computational models can guide experimentation and targeted measurements for crop improvement

Bedrich Benes¹, Kaiyu Guan²,³,⁴, Meagan Lang³, Stephen P. Long⁵,⁶, Jonathan P. Lynch⁷,⁸, Amy Marshall-Colón³,⁴,⁹, *, Bin Peng²,³, James Schnable¹⁰, Lee J. Sweetlove¹¹, and Matthew J. Turk³,¹²

¹Computer Graphics Technology and Computer Science, Purdue University, Knoy Hall of Technology, West Lafayette, IN 47906, USA,
²College of Agricultural, Consumer and Environmental Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, USA,
³National Center of Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, IL, USA,
⁴Center for Advanced Bioenergy and Bioproducts Innovation, University of Illinois at Urbana-Champaign, Urbana, IL, USA,
⁵Carl R. Woese Institute for Genomic Biology, University of Illinois, 1206 West Gregory Drive, Urbana, IL 61801, USA,
⁶Lancaster Environment Centre, University of Lancaster, Lancaster LA1 1YX, UK,
⁷Department of Plant Science, The Pennsylvania State University, University Park, PA 16802, USA,
⁸School of Biosciences, University of Nottingham, Sutton Bonington, Leicestershire LE12 5RD, UK,
⁹Department of Plant Biology, University of Illinois Urbana-Champaign, 265 Morrill Hall, MC-116, 505 South Goodwin Ave., Urbana, IL 61801, USA,
¹⁰Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE 68583, USA,
¹¹Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK, and
¹²School of Information Sciences, University of Illinois, Urbana-Champaign, Urbana, IL, USA

© 2020 Society for Experimental Biology and John Wiley & Sons Ltd

SUMMARY

Computational models of plants have identified gaps in our understanding of biological systems, and have revealed ways to optimize cellular processes or organ-level architecture to increase productivity. Thus, computational models are learning tools that help direct experimentation and measurements. Models are simplifications of complex systems, and often simulate specific processes at single scales (e.g. temporal, spatial, organizational, etc.). Consequently, single-scale models are unable to capture the critical cross-scale interactions that result in emergent properties of the system. In this perspective article, we contend that to accurately predict how a plant will respond in an untested environment, it is necessary to integrate mathematical models across biological scales. Computationally mimicking the flow of biological information from the genome to the phenome is an important step in discovering new experimental strategies to improve crops. A key challenge is to connect models across biological, temporal and computational (e.g. CPU versus GPU) scales, and then to visualize and interpret integrated model outputs. We address this challenge by describing the efforts of the international Crops in silico consortium.

Keywords: photosynthesis, flux modeling, whole-plant architecture, transcriptional regulation, multiscale modeling.

INTRODUCTION: IN SILICO APPROACHES HELP US LEARN WHAT TO MEASURE

Computational models provide simplified representations of real-world systems and structures that help to describe and understand biological complexity. As plant science becomes increasingly computational, models are in effect quantitative hypotheses that may combine multiple lines of understanding to design definitive experiments or direct observations. However, models contain a number of assumptions to substitute for incompletely measured or unmeasured processes, and thus are not complete representations of a system. Likewise, models built on existing knowledge and theory may be difficult to test and validate
if the data needed for comparison do not exist. These aspects of mathematical models can be viewed as limitations or, alternatively, as opportunities for discovery. Gaps in models and validation reflect gaps in data collection, and can serve as inspiration for experimentation or the creation of new technologies to obtain key missing data. For example, mathematical models of the circadian oscillator in Arabidopsis thaliana led to the development of experiments to resolve transcriptional regulation by a negative–negative feedback loop (for review, see Bujdoso and Davis, 2013; Chew et al., 2014). Spatial models of plant cell and organ development have led to the repurposing of existing or design of new imaging technologies (Fernandez et al., 2010; Federici et al., 2012; Jiang et al., 2019). Likewise, model guided bioengineering resulted in field demonstrations of increased photosynthetic efficiency resulting in substantial productivity gains (Kromdijk et al., 2016; South et al., 2019).

Multiscale modeling may refer to the integration of data within and between micro (cellular level) and macro (organ and/or ecosystem levels) scales. For example, multiscale modeling that takes advantage of high-throughput ‘-omics’ data and their integration can reveal interactions within and across molecular scales to uncover emergent properties, such as growth, development, reproduction and senescence, that cannot be attributed to any single level in the system (Fischer, 2008). Likewise, the integration of data from micro-scales can guide observations made at the macro-scale and vice-versa. The idea that an integrated model can be used to help scientists visualize phenomena that they cannot see is particularly helpful for predicting how plant growth and development will respond to different environmental signals (Lynch, 2007).

Compared with mammalian systems, the growth and development of plants is a remarkably plastic process. Even among genetically identical individuals of the same species, biochemical, cellular and architectural characteristics may dramatically change in response to different environmental cues (Sultan, 2000; Des Marais et al., 2013). The abundance of key floral and fruit volatile compounds in genetically identical petunia, tomato, strawberry or blueberry plants varies in response to the relative availability of specific wavelengths of light (Colquhoun et al., 2013). A change in photoperiod can increase the total number of leaves initiated by the same Arabidopsis genotype by six-fold (Morris et al., 2010). Reductions in CO₂ abundance and temperature alter both the vegetative and inflorescence sexuality of teosinte (Piperno et al., 2015). A better understanding of phenotypic plasticity and the genetic and molecular factors that govern responses to the environment may help to identify crop ‘ideotypes’, which are idealized crops able to adapt and thrive in certain environments. The variances in how different plant cultivars will respond to changes in growing conditions, whether positive or negative, can be modeled as genotype by environment interactions (G × E). In some cases in the literature, the general class of ‘environmental’ interactions is subdivided into ‘management’ (environmental factors humans can influence, e.g. planting density, fertilizer application, irrigation practices, etc.) and ‘environment’ (environmental factors not subject to human constraint, e.g. temperature, rainfall, soil type, solar radiation intensity, etc.; Wang et al., 2019). Such G × E × M interactions can be observed and quantified in models that use data from panels of related genotypes, increasing the accuracy with which phenotypic values can be predicted for lines not previously tested in a given environment (Jarquin et al., 2014). However, the absence of mechanistic models for why a given genotype will respond in a certain way in a given environment makes it difficult to predict how a plant will respond in untested environments. Simulations from integrative, multiscale models can quickly investigate new scenarios to test how the system of interest will respond to perturbations, and develop hypotheses to guide experimental design and/or new technologies to obtain measurements that will help to future-proof our food supply. Equally, it may identify root or canopy ideotypes that would better adapt a crop to a given environment, either current or future (Srinivasan et al., 2017).

In this perspective, we provide a number of examples in which mathematical models at individual biological scales have guided experimentation and scientific innovation. We specifically explore the utility of Physiological Crop Growth, Functional-Structural Models of roots, Metabolic Flux Models, and Gene Regulatory Network (GRN) Models (Figure 1). We then contend that by integrating models across scales, we can identify emergent properties of dynamic biological systems. The insights gained from the multiscale models can direct experimental measurements and technological developments to help us better understand how crops will respond to future climates. Model integration also creates new computational challenges for the visualization of multiscale model simulations that will allow researchers to intuitively interpret the results of cross-scale interactions. This challenge is discussed as an important future direction for the field.

**MATHEMATICAL MODELING GUIDES EXPERIMENTATION AND DISCOVERY**

**Physiological crop growth models**

Physiological crop growth models (CGMs), i.e. mathematical models dictating processes related to crop growth at above-organ (organ, plant, field, ecosystem) scales, are designed to simulate crop responses to environmental variations and their integrated impacts on productivity (Jones et al., 2003; Holzworth et al., 2014). CGMs often include a set of parameters that describe a specific crop or
specific genotype of a crop, and a set of time series environmental data to drive the simulation. These CGMs are built upon the collective experimental evidences and theoretical understanding of how a plant will develop throughout the growing season under different environmental conditions. In the last two decades, CGMs have been expanded to include genotypic data to predict the emergent phenotypic outcome from complex interactions among genotype, environment and management (Wang et al., 2019). Such models have been shown to increase the accuracy with which plant phenotypic outcomes can be predicted, particularly in unobserved environments and genotypes (Cooper et al., 2016; Messina et al., 2018). These CGMs provide the potential to predict performance not only of observed plants, but also of hypothetical plants with modestly different parameter/trait values (Messina et al., 2018). These CGMs provide the potential to predict performance not only of observed plants, but also of hypothetical plants with modestly different parameter/trait values (Messina et al., 2018). These CGMs provide the potential to predict performance not only of observed plants, but also of hypothetical plants with modestly different parameter/trait values (Messina et al., 2018).

**Crop growth models identify bottleneck traits for further study.** Crop growth models have been used to identify bottleneck traits for high productivity under different environmental and management conditions (Hammer et al., 2006, 2010; Messina et al., 2019). In one example, crop model simulations were used to explore the potential drivers related to canopy and root architectures for the observed yield trend in the US Corn Belt (Hammer et al., 2009). Recent developments in cross-scale crop models have enabled *in silico* testing of the impact that different, and even hypothetical, photosynthetic manipulations will have on yield under real environmental conditions (Yin and Struik, 2017; Wu et al., 2019). Wu et al. (2019) tested the coupling and decoupling of photosynthesis and stomatal conductance (A-gs) in their model, and found that coupled A-gs results in rapid depletion of soil moisture early in the growing season, thus decoupled A-gs is the preferred scenario in water-limiting conditions to improve crop growth. However, whether a decoupled A-gs scenario is possible for any photosynthetic manipulation remains unclear. The results of this study suggest that better representation of the photosynthesis–stomatal conductance link in the crop model is required to recapitulate experimental observations taken over different environmental conditions, especially those under water-stressed conditions.

**Computational crop growth models can assess management adaptations and environmental impacts.** Besides environmental variations, genetic gain in crop production at large scales is also affected by management practices. CGMs can be used to design better management practices, which can complement with genetic improvement for higher crop productivity in the field. Many studies have explored different management changes to sustain crop production under climate change. Among these efforts, Hammer et al. (2014) used CGMs to optimize the crop management practices (planting window, density, and row configuration) for either a fixed, broadly adapted, or region-specific adapted sorghum genotypes in Australia, demonstrating that environment-specific adaptation can confer both yield and production risk advantages at industry scale. Modeling was used to optimize genotype selection and agronomy in response to altered management with the introduction of the more efficient two-row sugarcane harvesters in Brazil (Wang et al., 2017). With more attention to sustainable agriculture (Gliessman, 1990; Horrigan et al., 2002; Conway and Barbier, 2013), the environmental impacts of agriculture cultivation should also be considered in optimizing the genetic and management
improvement strategies (Peng et al., 2020), which will require interfacing field-scale CGMs with large-scale hydrological, ecosystem, climate and economic models (van Ittersum et al., 2008; Sacks and Kucharik, 2011; Levis et al., 2012; Jägermeyr et al., 2016; Peng et al., 2018).

**Functional–structural models**

*Three-dimensional simulations of root systems and canopies guide the breeding of more resilient crops.* Root and canopy phenotypes with superior ability to acquire water and nutrients are promising selection targets to develop the resource-efficient, climate-resilient crops urgently needed in global agriculture (Lynch, 2007; Srinivasan et al., 2017). A number of technologies, such as high-throughput phenotyping, have been developed to rapidly measure canopy architecture. However, a primary obstacle for exploring root phenes is the complexity and opacity of the fitness landscape of root phenotypes. Root systems are highly complex, dynamic entities growing in a highly complex, dynamic, diverse and opaque medium. Empirical approaches to analyze the structure–function relationships of root phenotypes are invasive, difficult, and subject to serious artifacts. This problem is therefore well suited to *in silico* approaches (Vos et al., 2010; Dunbabin et al., 2013; Lobet et al., 2017; Passot et al., 2018; Postma and Black, 2020). In this section, we provide a case study of the utility of modeling approaches to understand root/soil interactions, using as an example the heuristic, functional–structural plant model *SimRoot* (Lynch et al., 1997; now *OpenSimRoot*, Postma et al., 2017), which has been critically important in several aspects of this effort (Figure 2), including: (i) evaluating trait utility; (ii) discovering new traits; (iii) estimating processes that we cannot control or measure; (iv) exploring decision spaces that are too vast to explore empirically; (v) evaluating phenotypes and environments that do not exist in nature; and (vi) augmenting empirical phenotyping platforms.

**Evaluating trait utility.** Root cortical aerenchyma (RCA) is formed by conversion of living parenchyma cells to air space via programmed cell death, thereby reducing the metabolic cost (i.e. C and nutrient requirements) of root tissue. To test the hypothesis that reduced root metabolic costs would allow greater soil exploration and nutrient capture, *SimRoot* was used to estimate the effects of varying RCA formation in maize and common bean for P capture, and in maize for the capture of N, P and K, as a function of nutrient availability (Postma and Lynch, 2010, 2011). Results showed a substantial benefit of RCA for nutrient capture in maize, which as a monocot has a more persistent cortex than dicot species like bean (Strock et al., 2018). Simulation results were later validated by empirical research, which showed that in maize, natural genotypic variation for RCA formation is associated with greater capture of N (Saengwilai et al., 2014a), P (Galindo-Castaneda et al., 2018) and water (Zhu et al., 2010; Chimungu et al., 2015) in the field and in controlled environments in which those resources limited growth.

**Discovering new traits.** *SimRoot* studies examining how root phenes (i.e. elements of a phenotype; York et al., 2013) interact to affect soil exploration highlighted the potential importance of the number of crown roots (CN) in maize root phenotypes (York et al., 2015). Initially CN was varied as ‘background’ phenotypic variation in studies focused on the growth angle of crown roots, but the strong effect of CN on nutrient capture *in silico* prompted a closer examination of natural genotypic variation for CN in maize. In the field and controlled environments, reduced CN was associated with greater root depth, and greater capture of N under N stress (Saengwilai et al., 2014b), and water under water deficit stress (Gao and Lynch, 2016). In contrast, greater CN was associated with reduced root depth and therefore greater capture of P from low-P soil (Sun et al., 2018).

**Estimating processes that are difficult to control or measure.** Competition for soil resources within and among neighboring root systems is an important element of root foraging efficiency that is exceedingly difficult to measure empirically. *SimRoot* was used to test the hypothesis that ancient maize/bean and maize/bean/squash polycultures are more productive on low-fertility soils than their respective monocultures, because of niche complementarity.
Simulations predicted better soil exploration and N capture in the polyculture than in monoculture (Postma and Lynch, 2012), which was later validated in field studies (Zhang et al., 2014a).

Exploring decision spaces that are too vast to explore empirically. The interactions of dozens of root phenes with each other and with a wide range of soil and atmospheric conditions create a vast decision space that would be infeasible to explore through empirical experimentation. Furthermore, SimRoot is able to vary individual phenes and conditions in isolation, which is very challenging to achieve empirically. As an example, a SimRoot study of how axial root phenotypes in common bean interact to influence the capture of N and P found that the utility of a root architectural phenotype is determined by whether the constituent phenes are synergistic or antagonistic, and that no single phenotype is optimal across contrasting environments (Rangarajan et al., 2018). This approach is now being extended to optimize root phenotypes across a wide range of soils and geographic locations.

Evaluating phenotypes and environments that do not exist in nature. Whether it is evaluation of unique root phenotypes or simulating future climates, *in silico* approaches have unique value in the analysis of scenarios that do not exist in nature. For example, SimRoot was useful in simulating N capture by maize roots with altered nitrate uptake kinetics, to show that existing phenotypes may be improved through breeding (York et al., 2016). In another example, SimRoot was useful in showing how specific root phene states contribute to the greater N capture of modern maize cultivars in response to density and fertilizer regimes over 100 years of maize breeding (York et al., 2015).

Augmenting empirical phenotyping platforms. High-throughput phenotyping of crop germplasm is a challenge for root phenotypes. *In silico* tools like OpenSimRoot will be useful in extending to entire root systems what can only be readily observed in root crowns, for example (Trachsel et al., 2011; Colombi et al., 2015; Burridge et al., 2016), or in extending measurements of young plants, which are easier to measure, to mature plants, which are more challenging.

SimRoot and its successor OpenSimRoot have been critically useful in understanding the fitness landscape of root phenotypes for water and nutrient capture, which has led to the breeding of more stress-tolerant, resource-efficient crops (Burridge et al., 2019). The integration of root models at the tissue, organ, organism and stand scales, as exemplified by SimRoot and OpenSimRoot, with models at cellular and subcellular scales, together with robust models of shoots, soil, microbes and diverse agroecologies, will comprise increasingly powerful tools in guiding the development of the crops and cropping systems needed to sustain a growing human population amidst a degrading environment. This has become possible with the advent of model interconnectors that can effectively exchange inputs and outputs from one model (e.g. OpenSimRoot) to another model, such as a shoot model, even though the two may be coded in different languages (Lang, 2019).

Parallel 3D representation of canopies has similarly allowed both understanding of the impacts of dynamic changes in light at the leaf level, leading to interventions that have improved productivity, and have identified ideotypes that would allow increased productivity and water use efficiency (for review, see Wang et al., 2020).

Metabolic flux and Gene Regulatory Network models

The genetic manipulation of crop growth and architecture requires knowledge about the underlying metabolic fluxes that underpin nutrient assimilation and growth. Functional—structural models typically represent metabolic processes in a relatively simplistic form, and there is substantial scope for more detailed models of metabolism to be embedded within whole-plant models (Baldazzi et al., 2012). If we consider the crop in terms of the efficiency with which it assimilates and utilizes absorbed inorganic nutrients, then models of the integrated metabolic systems for nutrient assimilation and conversion into biomass components are of key importance. In this section, we will discuss how the predictions of metabolic system models can be exploited. We refer the readers to recent reviews that cover the methodological approaches for modeling metabolism and the scope/limitations of metabolic models (Nägele and Weckwerth, 2012; Rohwer, 2012; Kruger and Ratcliffe, 2015; Dersch et al., 2016; Basler et al., 2018). For this purpose, we will assume a metabolic model is in place that simulates the hundreds of reactions of central metabolism with good accuracy. We will also assume that the goal is to manipulate the existing metabolic system rather than to introduce entirely new, synthetic pathways. Although synthetic pathway design holds great promise for improvement of plant metabolism (Schwander et al., 2016; Erb et al., 2017; Trudeau et al., 2018), the use of models to guide synthetic pathway design in plants is in its infancy but has resulted in a recent success in terms of increased productivity (Küken and Nikoloski, 2019; South et al., 2019).

Optimizing flux distributions. There are three principal ways in which the information from metabolic models can be used. The first is to exploit the fact that many metabolic models predict optimal flux distributions with respect to a specific objective, often growth. By comparison of the optimal flux distribution against reality, inefficiencies in the crop can be identified. For example, when a multi-tissue
source-sink model of metabolism was optimized for whole-plant growth, it was found that it was more efficient to transfer all nitrate taken up by the roots to the shoot for assimilation into amino acids (Shaw and Cheung, 2018). The nitrogen requirement of the root was serviced by transporting amino acids back to the root from the shoot. In reality, a variable but significant amount of nitrogen assimilation typically occurs directly in the root (Andrews, 1986). The comparison of model versus reality therefore leads to the hypothesis that crop growth would be increased if more nitrate assimilation occurred in the shoot (Amthor et al., 2015). High-efficiency flux modes in the leaf metabolic network have also been identified, requiring altered flows of ATP and reducing equivalents between subcellular compartments and increased use of mitochondrial oxidative phosphorylation during the light (Shameer et al., 2019).

Capacity limitations of metabolic systems. The second way in which metabolic models can be used, is to identify capacity limitations within the metabolic network. The $V_{\text{max}}$ of enzymes in central metabolism are generally in excess of the flux (Junker et al., 2007; Colombié et al., 2015; Schwender et al., 2015), meaning that capacity limitations are rare. However, the rapid change in environmental conditions in recent decades, such as the increase in atmospheric CO$_2$, leads to a situation in which the capacity of the metabolic system has not caught up with the higher carbon assimilation rates that result. Metabolic modeling has been used to identify enzymes in photosynthetic carbon metabolism that have become limiting under modern CO$_2$ conditions (Zhu et al., 2007). This was done using an evolutionary algorithm to optimize the amount of enzymes (assuming a constant amount of total enzyme) in a kinetic model of the chloroplast electron transport chain, the Calvin–Benson–Bassham (CBB) cycle, photorespiration, starch biosynthesis and sucrose biosynthesis. When the actual amounts of these enzymes were compared with the model-optimized amounts (with the model run under the then current atmospheric CO$_2$ concentration of 380 mmol mol$^{-1}$), it was predicted that greater photosynthetic assimilation would be achieved with substantial increases in the amounts of the CBB enzymes sedoheptulose bisphosphatase and fructose bisphosphate aldolase and the starch biosynthetic enzyme ADP-glucose pyrophosphorylase. Further, if increased triose phosphate/3PGA export from the chloroplast was allowed, then the model also predicted an increase in cytosolic fructose bisphosphatase to increase cytosolic sucrose biosynthesis.

The predicted effects of increasing the two CBB cycle enzymes have been confirmed in a range of transgenic plants (Miyagawa et al., 2001; Lefebvre et al., 2005; Uematsu et al., 2012; Köhler et al., 2017; Simkin et al., 2017). Additionally, transgenic plants with increased leaf ADP-glucose pyrophosphorylase (Gibson et al., 2011) and increased leaf sucrose synthesis (Jonik et al., 2012) also showed increased photosynthetic assimilation rate and enhanced growth. There is clearly potential to extend this type of analysis to the wider plant metabolic network and to include nitrogen assimilation as well as carbon assimilation.

Transcriptional regulation of metabolic networks. So far, we have considered the identification of direct targets within the metabolic system itself – i.e. genes encoding enzymes or transporters of metabolites. However, it is also possible to link the changes in the metabolic system to those of the GRN that regulates enzyme levels (Lakshmanan et al., 2016; Mohanty et al., 2016), which is the third way in which metabolic models can be used. Regulatory information can be layered on top of metabolic networks using information from the transcriptome. GRN analysis can be done to identify putative transcriptional regulators of transporters and metabolic genes. One successful way to construct a GRN is to use available genome-level information for a species to identify and analyze putative promoter regions for regulatory elements such as transcription factor binding motifs (Gutierrez et al., 2008; Para et al., 2014; Varala et al., 2018). When promoter analysis is combined with co-expression analysis of transcripts, robust predictions can be made about positive and negative regulatory relationships (Vandepoele et al., 2009; Kulkarni et al., 2018). Network analysis can then be used to identify important regulatory nodes in the GRN. Nodes that exert a large amount of control over the entire network are referred to as hub genes that can sometimes be master regulators of all or portions of biological pathways and processes. Thus, network analysis can help to identify genes to target for experimental, functional analysis (for review, see Gehan et al., 2015; Marshall-Colón and Kliewenstein, 2019). Likewise, GRN analysis can identify co-regulated genes involved in common pathways and processes, which can also direct experimental efforts by revealing what biological level may uncover a phenotype in response to a genetic or environmental perturbation that alters the transcriptome. This was recently demonstrated in a multiscale model of soybean photosynthesis, in which GRN, metabolic flux and micro-meteorological models were integrated to simulate soybean response to elevated CO$_2$. This study examined key regulatory points in metabolic flux and then explored how the genes encoding the most influential enzymes in the light reactions of photosynthesis are regulated at the transcriptional level (Kannan et al., 2019). Other examples in maize include the use of GRNs to predict tissue-specific regulation by transcription factors (Huang et al., 2018) and the molecular mechanisms contributing to the process of natural leaf senescence (Zhang et al., 2014b). Likewise, a multiscale digital model
of Arabidopsis linked multiple interaction processes from regulation of gene expression to plant structure and biomass, and accurately predicted the difference in aboveground biomass between two Arabidopsis accessions (Chew et al., 2014).

CURRENT EFFORTS TOWARD MULTISCALE MODELING AND FUTURE DIRECTIONS

The above examples show that mathematical models at single scales (e.g. micro or macro) can form explanations of biological phenomena and guide the development of measurements to test model-generated hypotheses. Additional gaps in our knowledge and missing measurements may not be revealed until models are integrated across scales in an attempt to reveal emergent properties of the system, particularly in tight feedback loops where non-linear behaviors may manifest through the combination of systems governed by different methods (e.g. between ‘rule-based’ and stochastic systems, as described in Qu et al., 2011). By combining modeling efforts into an integrated, multiscale model where boundaries are well-identified but also high-dimensional, in silico experimentation can test ‘what if’ scenarios, and potentially identify meaningful new measurements and manipulations to consider. Scenario testing with integrated models is one goal of the Crops in silico consortium (cropsinsilico.org; Marshall-Colon, et al., 2017). Optimization of model representations at the level of photosynthesis, the canopy, or the root system, has already led to improvements at the plant and field level. However, the greatest prizes will clearly come from the ability to mathematically optimize more complete integrated systems.

As experimentation can only cover limited G × E × M space in reality, crop modeling provides a feasible way to assess the impacts of genetic improvements at large scales (Sinclair et al., 2010, 2019; Messina et al., 2015, 2018). However, this requires an effective bridge to fine-scale (i.e. gene network, molecular, metabolic, tissue and organ) models and enable the testing of gene-level improvements on broad geographic regions using crop models. The challenge lies in the inherent scale differences between physiological crop models and fine-scale cellular models that are on different biological, temporal and computational scales (Baladazzi et al., 2012; Band et al., 2012; Hill et al., 2013).

Developing flexible, robust mechanisms for connecting independently developed models that operate across a wide range of spatial and temporal scales has substantial impact both within the domain of crop sciences and external to it. The development of an intuitive computational framework to facilitate model integration is another key goal of the Crops in silico consortium. The yggdrasil framework (Lang, 2019) provides a mechanism for building flexible, and yet compactly-defined, connections between models that are robust and extremely low-latency. This polyglot framework, enhanced with flow control, logic and in situ visualization, provides a platform onto which analysis, instrumentation and provenance tracking can be built. While this is similar to large-scale workflow projects such as Kepler (Ludascher et al., 2006) and DataWolf (Navarro et al., 2019), the structure of yggdrasil is inverted; rather than focusing on the overall workflow and building connections as a result of this, the connections are the fundamental operations, and the data passed between them are augmented with semantic information regarding its content, such as the provenance of the data, the units associated with it, and any information about the logical conditions that led to its sharing. The yggdrasil framework facilitates asynchronous communication among models written in different languages and that operate at different scales, resolving the historical problems associated with integrative and multiscale modeling (Lang, 2019).

The semantically defined networks utilized by models connected through yggdrasil provide the opportunity for in situ visualization of connections between models. Such visualization can provide researchers with deeper insights about individual models by revealing mismatches between expectations and model-assumptions upon integration with another model. Likewise, representing the output of the integrated, multiscale model simulations as 3D geometry is a natural and easily understandable representation for human comprehension. Further development of the yggdrasil framework will provide researchers with a comprehensive workbench for developing, studying and understanding multiscale modeling tasks. The visualization of multiscale model simulations is key to understanding the emergent properties that arise from the interactions within a complex biological system. Thus, effective visualization of data and simulations is a grand challenge for multiscale modeling efforts, and the Crops in silico consortium hope to overcome a number of the technological and computational challenges that exist in visualization.

Biological structures have huge variation that depends on the environment and plant genetics. Recent advances in geometric modeling of vegetation in computer graphics (Pirk et al., 2016) allow for the generation of very high-quality 3D geometries (Pirk et al., 2012), while models written directly in higher programming languages such as C++ or Python (Fišer et al., 2017) allow for high user control, and interactive experimentation. Likewise, modern Graphical Processing Units (hardware devices for processing geometry) allow for the generation of highly detailed geometries consisting of millions of polygons in real-time. However, non-destructive capturing and reconstructing of fully detailed 3D plant geometry and its development belong among the biggest challenges of computer vision. In order to validate the biological models, robust and compact geometric representations are necessary, as well as...
the ability to reconstruct real plants into the geometric model.

One promising approach is inverse procedural modeling that allows for functional–structural representation of vegetation (Stava et al., 2014) by estimating the parameters of a growth model. However, the input of this algorithm is the reconstructed plant 3D geometry that is difficult to obtain. Moreover, optimization approaches require metrics that compare the generated geometry with the captured geometric model. Deep learning has the potential to aid the problems in 3D geometry reconstruction. Deep learning is an umbrella term for a set of algorithms that use deep neural networks (LeCun et al., 2015), and it has been successfully applied in biology-related fields, for example in leaf detection and counting (Ubbens et al., 2018; Kuzničov et al., 2019).

Attempts to reconstruct 3D plant geometry have revealed a need for more appropriate input data, such as 360-degree depth-images or point clouds of plants rather than 2D photographs. Appropriate data for these models are rapidly becoming available through advances in high-throughput phenotyping and sensor development, in particular in-depth cameras and LiDAR technologies. These can be mounted on gimbals attached to autonomous robots that can roam through the crop rows, and to unmanned aerial vehicles (drones) generating massive high-quality and stable data streams. These in turn will be able to interact with models, providing, for example, re-parameterization, so that combined with agricultural weather forecasting, futures may be predicted and updated as a season progresses.

CONCLUSIONS

Advances in technology for model integration and intuitive visualization of integrated model simulations are opening avenues for in silico experimentation and scenario testing. An outcome of this endeavor, which is beginning to emerge, is the acceleration of directed engineering and breeding for crop ideotypes that will result in enhanced crop production and sustainability under challenging future environmental conditions. Likewise, integrative and multiscale modeling has the potential to reveal data gaps, and guide new experimentation, measurement and the development of enabling technologies.

ACKNOWLEDGEMENTS

Research reported in the publication was supported by the Foundation for Food and Agriculture Research under award number – grant ID: 602757. The content of this publication is solely the responsibility of the authors, and does not necessarily represent the official views of the foundation for Food and Agriculture Research.

CONFLICT OF INTEREST

The authors of this manuscript claim no conflict of interest.

AUTHOR CONTRIBUTIONS

All authors contributed equally to the conception of the presented idea and the writing of the manuscript.

DATA AVAILABILITY STATEMENT

There are no data associated with this manuscript.

REFERENCES


Model directed measurements for crop improvement


