

Review

Plants *in silico*: why, why now and what?—an integrative platform for plant systems biology research

Xin-Guang Zhu¹, Jonathan P. Lynch², David S. LeBauer³, Andrew J. Millar⁴, Mark Stitt⁵ & Stephen P. Long⁶

¹CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China, ²Department of Plant Science, Penn State University, University Park, PA 16802, USA, ³Institute for Genomic Biology and National Center for Supercomputer Applications, University of Illinois, 1206 W Gregory Drive, Urbana, IL 61801, USA, ⁴SynthSys and School of Biological Sciences, University of Edinburgh, Midlothian, Scotland, UK, ⁵Max Planck Institute for Molecular Plant Physiology, D-14476 Potsdam Gölms, Germany and ⁶Departments of Crop Sciences and Plant Biology, Institute for Genomic Biology, University of Illinois, Urbana, IL 61801, USA

ABSTRACT

A paradigm shift is needed and timely in moving plant modelling from largely isolated efforts to a connected community endeavour that can take full advantage of advances in computer science and in mechanistic understanding of plant processes. Plants *in silico* (Psi) envisions a digital representation of layered dynamic modules, linking from gene networks and metabolic pathways through to cellular organization, tissue, organ and whole plant development, together with resource capture and use efficiency in dynamic competitive environments, ultimately allowing a mechanistically rich simulation of the plant or of a community of plants *in silico*. The concept is to integrate models or modules from different layers of organization spanning from genome to phenome to ecosystem in a modular framework allowing the use of modules of varying mechanistic detail representing the same biological process. Developments in high-performance computing, functional knowledge of plants, the internet and open-source version controlled software make achieving the concept realistic. Open source will enhance collaboration and move towards testing and consensus on quantitative theoretical frameworks. Importantly, Psi provides a quantitative knowledge framework where the implications of a discovery at one level, for example, single gene function or developmental response, can be examined at the whole plant or even crop and natural ecosystem levels.

Key-words: plant models; crop models; ecosystem models; Earth System models; system analysis; virtual organisms; root architecture; photosynthesis; stomata; plant molecular biology; gene networks; metabolic networks.

WHY 'PLANTS IN SILICO'?

The concept is to integrate functional information from genome to phenome to ecosystem, spanning spatial and temporal scales and capturing the emergent properties of complex, non-linear systems (Fig. 1). This is being realized increasingly

in various microorganisms (e.g. Edwards *et al.* 2000) and now in a mammal, as the virtual physiological rat (Beard *et al.* 2012), providing resources for integration and application of the expertise of the broad community of researchers of these organisms. Such a community activity has not been attempted, to date, by the plant, crop and ecosystem research community. Rather than integration and inter-operability at the model level, our community has attempted to achieve consensus via model inter-comparison, in the hope that by comparing predictions from many locally, sometimes individually, developed models, the truth will emerge from the average (Hanson *et al.* 2004; Li *et al.* 2015). This approach has clear value in identifying strengths and weaknesses of different models and approaches and provides the best consensus currently achievable. However, the consensus occurs at the point of final model output rather than correct representation of the underlying processes. It does not provide a framework for integration, easy linkage to the rapid growth in 'omics or a means to make strong connections with those developing mechanistic understanding of the underlying processes from genome to phenome. We argue for a new approach, a community framework model that takes full advantage of the latest developments in computer software and communications engineering, and accessible to all. This would allow all to share and probe modules developed by domain experts for different levels of organization from gene networks to whole plant and ecosystem responses to the environment. Full representation of even a few crops or representative plant functional types, with robust mechanistic consideration of the key physical, chemical and biological processes, is an ambitious goal. However, Psi would serve now to integrate the emerging process models into one quantitative knowledge framework. Psi will be an unparalleled tool for the analysis and understanding of plants as hierarchies of inter-related biological systems responding to a dynamic environment. It could serve to integrate the community of researchers and develop many synergies (Fig. 2). Its primary objective would be heuristic, to test completeness of knowledge of processes leading to plant growth and development and beyond to ecosystems. The modular framework would allow testing of alternative formulations of underlying processes in the context of the whole. While we envisage Psi as making maximum use of mechanistic knowledge,

Correspondence: S. P. Long; e-mail: slong@illinois.edu

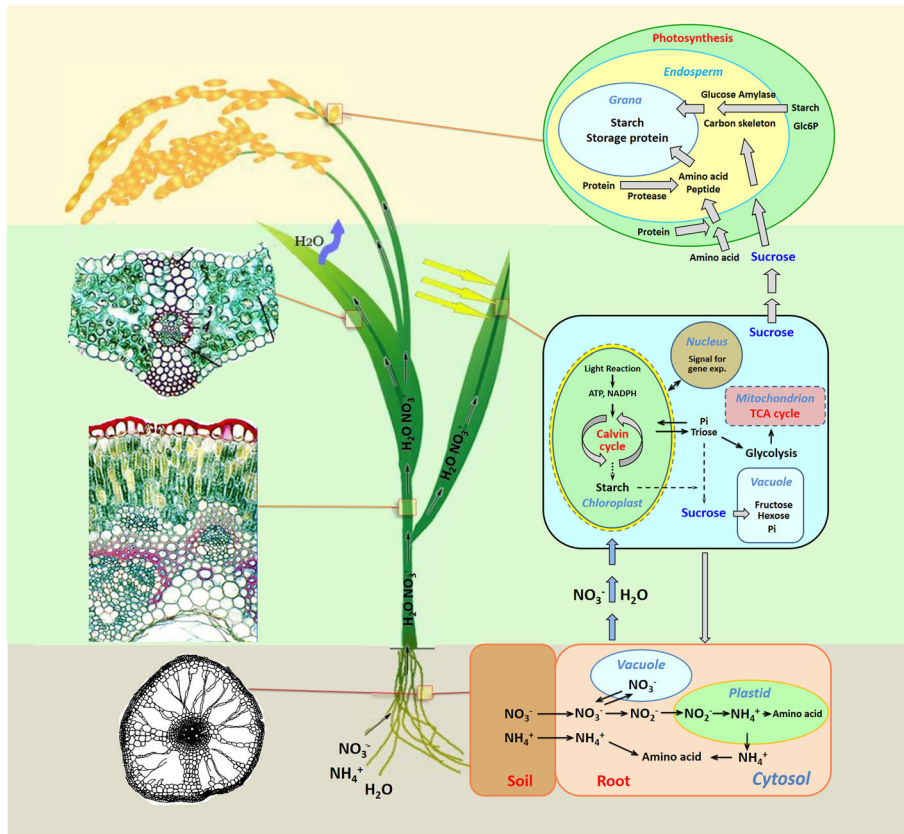


Figure 1. An illustration of the major processes that would be incorporated in *Psi*, based on that previously published (Zhu *et al.* 2011) and presented here with modifications and with the permission of the publisher.

its modular nature recognizes that simpler empirical or phenomenological sub-models of different biological processes will often be more effective, depending on context. Equally, it provides a means to test when and where simpler sub-models requiring less parameterization can be as or more effective than more complete mechanistic models in simulating the growth and development of a whole plant. Beyond this, it may lead to better predictive plant productivity models, including those of

crops, which may draw on the incremental improvement in representation of processes provided by *Psi*. Here, under the ‘Why Plants *In Silico*?’ section, we explain why this new approach is needed now, new applications *Psi* would allow in the ‘How *Psi* Might Be Applied to Enhance Capabilities in Plant Sciences’ section, why this is timely is discussed under the ‘Why Now?’ section and what *Psi* might look like in the ‘Potential Framework’ section.

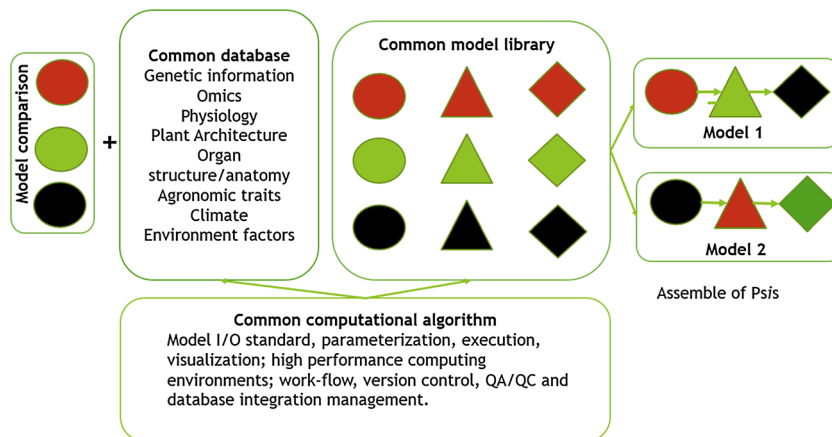


Figure 2. The overall structure of plants *in silico* (*Psi*) that would include a common module library, representing individual molecular, physiological and developmental processes that can be assembled to form models for different plants and applications. The *Psi* would include a common data source, which can be used to conduct model comparison.

We see five major drivers for why this is timely and needed, although these do not represent an exclusive list.

1 *The weak link in land–atmosphere interaction models.*

Local and regional atmospheric transport and soil reactive transport models are based largely on exact fundamental laws of physics and chemistry. In modelling the soil–plant–atmosphere continuum at a local, regional or global scale, these advanced physico-chemical models of the atmosphere and soil are being linked with vegetation models that, by comparison, are largely empirical and phenomenological, with little mechanistic depth. Nevertheless, it is plants that convert sunlight energy into the chemical energy that drives ecosystem carbon uptake and many elemental cycles while strongly influencing the terrestrial water cycle. Plants are therefore by far the ‘weakest link’ in both Earth Systems and local ecosystem models, including crop systems. This adds huge uncertainties to models in the context of the global, regional and local carbon, hydrological and nutrient cycles.

2 *A new approach to crop models.* Crop monocultures represent the simplest assemblages of plants and have the largest databases, at least in terms of predicting biomass production and yield over a range of environments. With projections of forward shortfalls in supply over demand, their efficacy has gained even greater importance (Long *et al.* 2015). Most models of the major crops have evolved at single or limited locations and developed at a time when both knowledge of underlying mechanism was limited and computational power even more limited in its ability to represent such mechanism. They have therefore evolved in a vertical manner at different locations attempting to deal with all layers of expertise from molecular to global production. While today, high-speed high-volume communication would allow horizontal integration allowing experts at different levels of organization to participate from dispersed locations. The problems of the current paradigm are perhaps illustrated where five of the most advanced wheat models were used to predict the direct effect of just one global change variable, rising CO₂, on yield. The average of this inter-model comparison overestimated the measured effect by 100%, and the individual models differed in their predictions by 200%, yet this was despite many shared assumptions within the models (Ainsworth *et al.* 2008b). If our models for the best studied plants in the field miss the mark so badly, can we really expect any real precision in predictions for unmanaged multi-species systems, for which there is far less parameterization data and knowledge of the plants concerned? Equally, it suggests a very poor foundation from which to predict future global food production.

3 *Mechanistically rich models for predicting beyond experience.* As discussed in the succeeding texts, the current major families of models of crops and vegetation are largely empirical containing very limited mechanistic rigor and with gross simplifications of key processes. In the past, this may have been necessitated by limited computational power and inefficient languages. While such empirical approaches can work well within the bounds of existing data and experienced environments, they are fundamentally unsuited for

extrapolation to new scenarios. The only models suited to extrapolate beyond empirical parameterization are models that are based on the underlying mechanisms of response or those that capture emergent phenomena correctly. The mechanistic nature of *Psi*, as envisaged, would in the longer term provide a more feasible approach for quantitative prediction of the responses of plants to multi-factor environmental changes. This is of particular importance, given that we lack sufficient data under realistic field conditions to allow reliable generalization into empirical models – for example, the interactive effects of simultaneously rising [CO₂], [O₃] and temperature have not been investigated under open-air field conditions for most crops and where they have it has been at just one location.

4 *The need to integrate expertise across levels of organization.* Much contemporary plant scientific research has focused on the function of individual genes, proteins and processes in a relatively isolated and descriptive manner. Because of this isolation, when the conclusions are tested in a different environment, they can fail, not just quantitatively but qualitatively. For example, a long-held view developed from understanding of photosynthesis of individual leaves in full sunlight was that increasing Rubisco specificity would benefit crop productivity. However, when the heterogeneity of the light environment inside a crop canopy is considered, it is found that Rubisco specificity should be decreased, not increased, to improve crop productivity (Zhu *et al.* 2004). Another example is that increased temperature has been assumed to decrease plant carbon gain; however, when the interaction of rising atmospheric [CO₂] and temperature are considered, the opposite can be the case (Long 1991). By incorporating all key processes and their inherent interactions in a unified framework, *Psi* could be used as an effective tool to evaluate the consequences of interactions between different processes and environmental factors. It would highlight when unanticipated emergent responses may occur and identify such interactions as an important area for further investigation. By being modular, it would also allow the cost-benefit assessment of simple versus complex representations of processes, that is, where the simple and more tractable are sufficient for a given purpose is.

5 *A functional framework for utilizing the explosion of ‘omics data.* *Psi* would be a platform to link genome to phenome, as seen in emerging integrative models (Hill *et al.* 2013; Chew *et al.* 2014). Modern high-throughput technologies enable rapid accumulation of large amounts of genomic, transcriptomic, epigenomic, proteomic, metabolomic and phenomic data. Various bioinformatic approaches have been developed to extract information based on such rich data, such as identifying differentially expressed genes, identifying motifs enriched in a particular pathway or identifying clusters of genes that might be co-regulated (Usadel *et al.* 2009; Rhee & Mutwil 2014), as well as identifying sets of transcripts or metabolites that are correlated with more complex emergent phenotypes (Sulpice *et al.* 2009; Riedelsheimer *et al.* 2012; Sulpice *et al.* 2013). However, these analyses are inherently

correlation based and hence do not necessarily provide new mechanistic understanding for genes identified through such approaches (Stitt 2013). As a result, these analyses provide a far more comprehensive catalogue of changes at multiple levels, compared with previous traditional single-gene knock-out approaches. However, such information cannot be directly used to predict changes at the physiological or whole plant levels. Mechanism-based modelling from the *Psi* framework would provide a direct linkage between these molecular level variations and physiological responses, building on the scaling from gene expression to plant growth illustrated by Chew *et al.* (2014). *Psi* would also provide a unique platform to link genomic-scale constraint-based models to physiological outcome. Such constraint-based models, now mature within the microbial research community (Benedict *et al.* 2014), have also shown potential in *Arabidopsis* (Poolman *et al.* 2009; Chew *et al.* 2014). Such models can allow prediction of protein synthesis and by implication, growth rates and energy consumption from quantitative molecular information about ribosome and transcript abundance (Piques *et al.* 2009; Pal *et al.* 2013). One caveat of such a modelling approach is that while it reflects the stoichiometry of substrates and products within a network, they ignore the non-linearity of many relationships of reaction velocities to substrates, products, effectors and inhibitors, even though these are often mathematically well defined (Price *et al.* 2004). However, the predictions from constraint-based models do provide boundary conditions for the metabolic pool dynamics and physiological parameters of a *Psi*. Hence, fusing constraint-based models with kinetic models within *Psi* would provide a unique opportunity to incorporate genomic-scale information with more detailed kinetic models for the better known processes, such as clock-based processes, photosynthesis and respiration (Zhu *et al.* 2013; Chew *et al.* 2014).

Psi would offer a number of immediate practical applications in crop breeding and engineering. The plants *in silico* concept would be particularly valuable in developing *ideotypes*, which are ideal phenotypes for a given or future environment, as an aid to germplasm selection and breeding. The ideotype or optimization concept can be applied at the level of a single process, for example, photosynthetic carbon metabolism (Zhu *et al.* 2007), an individual plant (Sarlikioti *et al.* 2011) or a field crop canopy (Drewry *et al.* 2014). Mechanistic modelling is uniquely valuable in ideotype development because it permits the evaluation of many potential phenotypes for different environments and selection by application of evolutionary algorithms, Pareto efficiency or other optimization algorithms – geared to a specific outcome. Because the utility of specific traits may interact with the expression of other traits, potential phenotypes of interest may include the large number of combinations among traits of interest. For example, if 20 individual traits are of interest, each of which exists in only two states (e.g. present versus absent, many versus few, large versus small, etc.), there exist 2^{20} (over one million) distinct phenotypes combining these traits. The utility of these phenotypes as ideotypes would require their evaluation in various possible environments representing the edaphic, climatic and management scenarios of interest. These permutations result in a very large number

of scenarios of interest, which vastly exceed the capabilities of empirical research. An effective *Psi* would allow the researcher to narrow down the combinations to those most likely to fit their test environment. Mechanistic modelling is therefore the ideal avenue for the development of robust ideotypes that consider traits, trait interactions and a wide range of spatial and temporal environments (Poorter *et al.* 2013).

HOW *PSI* MIGHT BE APPLIED TO ENHANCE CAPABILITIES IN PLANT SCIENCES

Here, we give two examples, but many more may be envisaged.

1 Designing crops for current and future environments, with particular reference to bioenergy. Development of crop ideotypes is especially important for accelerating the development of sustainable second-generation bioenergy crops, which have not yet been subject to significant selection for either productivity or other desired ecosystem services. It is entirely possible or even probable that the ideal phenotype (ideotype) of a biofuel crop is significantly different from the range of phenotypes present in existing germplasm. While our existing major food crops represent centuries of selection and a huge intensity of research investment over the past 50 years, our second-generation bioenergy crops, such as willows and Miscanthus, are little different from the wild plants from which they were selected (Somerville *et al.* 2010). Such a means to accelerate breeding also has value for the staple food crops in the developing world, which have not received the intensity of breeding effort of the major crops. As an example, see the poor rate of improvement in global yields of cassava, the staple of much of sub-Saharan Africa versus wheat (Long *et al.* 2015). Even for the major food crops, *Psi* provides new opportunities for designing crops with higher yield potential or/and resource use efficiencies.

Why is there an opportunity to increase productive potential and resource use efficiency in crops when millennia of evolution that has surely selected for efficient plants? Firstly, during evolution, fecundity not productivity is the driving force. As a result, many of the features that confer a competitive advantage at the level of the individual plant, which is the basic unit on which evolution operates, are not necessarily beneficial in a monotypic crop stand. One example is that our modern crops may have more chlorophyll in their leaves than is optimal for yield (Ort *et al.* 2011). Dark green leaves help to shade out competitors but in a crop situation result in the upper leaves absorbing more light than they can use in photosynthesis, while the layers of lower leaves are starved for light. As a large part of the protein in leaves is invested in the photosynthetic apparatus, the distribution of chlorophyll is likely to reflect the distribution of protein. This lowers nitrogen use efficiency because of inappropriate allocation of nitrogen in the stand, with respect to maximizing productivity. Secondly, the evolved features that are highly adapted for previous environments are not necessarily optimal for the current and certainly not optimal for future atmospheric conditions. A major issue with respect to global change is rising atmospheric $[CO_2]$. The ancestors of modern

crops evolved through 25 million years when $[\text{CO}_2]$ averaged 220 ppm. Within just 200 years, the human population has raised $[\text{CO}_2]$ to 400 ppm, with most of that rise in just the past 50 years, giving few generations of crop breeding for any adaptation (Zhu *et al.* 2004). Because of the costs of Free-air CO_2 enrichment systems (FACE), it will be very expensive to empirically breed for increased performance under simulated future atmospheric CO_2 conditions, and any such activities will be best integrated within a strong predictive context (Ainsworth *et al.* 2008a; Bishop *et al.* 2015; Calfapietra *et al.* 2010). The mechanistically rich proposed formulation of *Psi* could be particularly valuable in predicting the optimal crop features and agronomic practices for both food and energy crops under current and forecast future climates. Recent mechanistic modelling of this issue has shown that by modifying canopy architecture under future $[\text{CO}_2]$, simultaneous gains in productivity and water use efficiency could be obtained (Drewry *et al.* 2014). However, these predictions were based on simulations of static crop canopies. More robust predictions would be possible within the framework of the proposed *Psi*, where root growth and distribution, water relations and growth stage could all affect outcome and yet be included in assessing optimization. Despite thousands of years of selection, the genetic diversity of our major crops is low relative to the diversity of their wild relatives and landraces (Xu *et al.* 2009). *Psi* would provide a framework to identify non-obvious morphological and physiological traits of value in this wider genetic base.

2 A heuristic tool. Models are particularly valuable when they permit analysis of processes that are difficult to measure empirically. In this context, *Psi* will be especially useful in the analysis of plant–soil interactions. Root systems are exceedingly dynamic complex entities that intimately interact with the soil, an exceedingly dynamic, complex and opaque environment (Fig. 3). Our understanding of plant interactions with the soil, including its living components, is far from complete. In this context, as in all, iterative integration of modelling and empirical research will be synergistic, in that *Psi* can identify key scenarios and processes deserving empirical investigation, while empirical research permits increasing refinement of *Psi*.

Psi can be used to predict the emergent biology of plants as complex systems. *Psi* will also have value as a heuristic tool, that is, as a mechanism to explore and improve the adequacy of our conceptual understanding of plants. Indeed, *Psi* will be a platform to integrate our conceptual understanding of plants in a formal mathematical environment.

WHY NOW?

There are many individual models of plant processes and growth of specific crops already available. So, why another model now? Firstly, *Psi* does not aim to replace these but integrate and evaluate these as part of a community whole. It would take advantage of contemporary computational, cloud storage and communication tools allowing different process formulations to be incorporated and tested in a plug-and-play manner. Secondly, a number of our existing plant, crop and

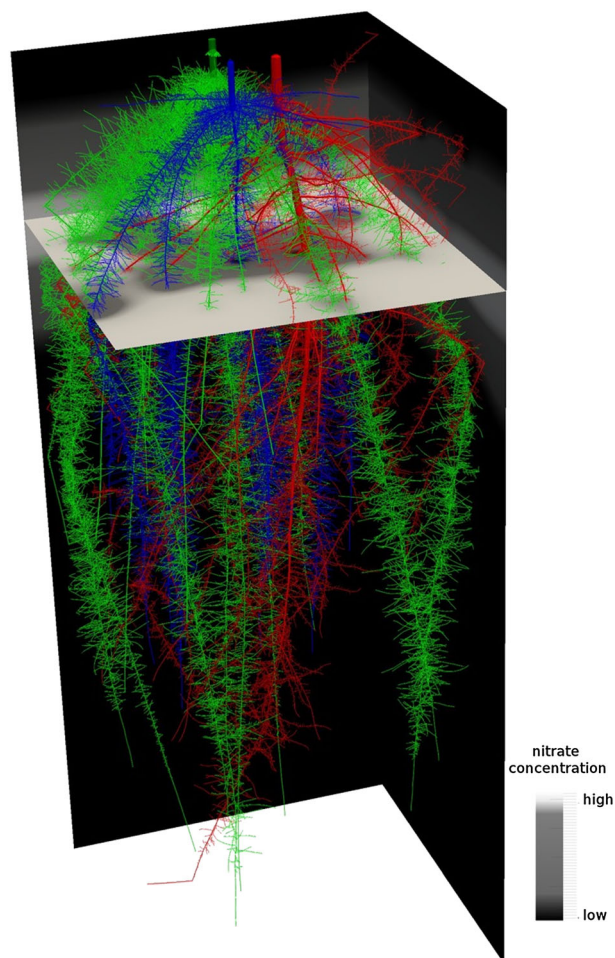


Figure 3. Computer simulation of the root systems, as an example of visualization. Here, the ancient ‘three sisters’ of traditional Central American polyculture, that is, maize (green), bean (blue) and squash (red), are examined. Following developmental rules, it shows how the individual systems develop to have root architectures that permit efficient nitrate capture by the system as a whole (Postma & Lynch 2012). Image courtesy of Johannes Postma.

vegetation models are the product of single labs rather than the community and often rely on heritage and sometimes undocumented code. This can make integration with modular community frameworks and parallelization cumbersome or impractical. Evolution in different labs has resulted in very different structures and degrees of complexity for models of different plant groups and crop species. As a result, it is very difficult to attribute differences in responses to a given environmental variable versus differences in the biology of the crop versus differences in the model structure. The *Psi* proposed here aims to provide a framework that can enable assembly of the diverse models developed by domain experts, enable comparison of the representation of processes and hence improve the accuracy of prediction from such models. Through this practice, the most appropriate models for individual processes in given contexts can be identified. The integrated whole, based on such model components, will consequently be more predictive.

A mechanistic, multi-scale plants *in silico* platform, which faithfully simulates the dynamic responses of plants to environmental change at levels spanning from molecular to multi-organismic scales, would be a revolutionary community discovery platform. A near-term tangible goal of Psi should be to establish a framework that will be able to mechanistically predict the responses of growth and developmental processes in a few model species, such as *Arabidopsis*, and model crop species such as rice and maize, to variation in light quantity, day length, [CO₂], water and nutrient availability. In this way, the model can predict responses beyond experimental observations with some confidence.

A number of recent developments in both plant and computer sciences now make developing such a Psi possible (Chew *et al.* 2014). Figure 4 provides what is perhaps the first example of scaling through layers of organization, from gene expression in response to environmental fluctuations to the dynamics of growth and development of an individual, in this case *Arabidopsis*.

1 Many basic modules are available but rarely effectively integrated. Several of the basic modules required to form Psi have emerged. These include models of the complete C₃ and C₄ photosynthetic process from biochemistry (Zhu *et al.* 2013; Wang *et al.* 2014) to 3D reaction diffusion models of the mesophyll cell (Tholen & Zhu 2011; McGrath & Long 2014), 3D models of actual plant canopies (Zhu *et al.* 2012; Song *et al.* 2013), models of stomatal action (Buckley & Mott 2013), respiration (Sweetlove *et al.* 2013), phloem and xylem flow (Hall & Minchin 2013), growth and development (Steinacher *et al.* 2012; Bennett *et al.* 2014), shoot patterning (Domagalska & Leyser 2011), flowering (Song *et al.* 2012), root structural and functional dynamics (Lynch 2013; Dyson *et al.* 2014) and linkage to gene regulatory networks (Hill *et al.* 2013; Chew *et al.* 2014). Now, there is the opportunity, for the first time, to integrate all within a single framework to grow a whole plant, complete with shoot, root and reproductive organs *in silico* and reconstructed in 3D (Figs. 3 & 4).

2 Rapid growth of data for model parameterization and validation. Modern high-throughput phenotyping and genotyping technologies offer an unprecedented capacity for parameterization and validation of a Psi that would be difficult to implement in the legacy plant models. Advances in large-scale measurement of transcriptomes, proteomes and metabolomes (Gibon *et al.* 2004) make it possible for the first time to parameterize mechanistic models down to the actual molecular components involved. Increasingly high-throughput methods allow a thorough characterization of the dynamics of metabolite pools, as well as time-resolved estimates of growth rates (Chew *et al.* 2014, Sulpice *et al.* 2014). Progress in the detailed measurement of flux distribution at the leaf or even the whole plant level also provides an unprecedented opportunity to validate such mechanistic models (Heise *et al.* 2014). Added to these high-throughput molecular level data, rapid advances in high-throughput phenotyping also make it realistic to rapidly reconstruct 3D canopy architecture and quantify dynamics of organ growth or even dynamics of metabolic networks. Modern imaging technologies now also enable 3D reconstruction

of the detailed structure and anatomy of different tissues or even cells at scales fine enough to enable development of morphogenetic models. These do not just provide large data sets but the opportunity to digitally represent actual 3D details of real shoot and root structures, down to cellular and sub-cellular detail within a Psi. At the same time, software developments allow linkage of database and modelling frameworks across many remote locations, which can automatically update model parameterization as new data is evaluated and incorporated, using, for example, the approach of PeCAN (Dietz *et al.* 2013).

3 Need to represent signalling at all levels of plant organization. Increasing knowledge about plant signalling pathways and plant microbiome signalling will provide a framework to model complex interactive responses. For example, current models for carbon allocation and starch turnover are rigid because they are empirically parameterized to a small set of conditions and break down in a wider range of conditions (Chew *et al.* 2014). It should soon be possible to replace them with models that use an understanding of how starch turnover is regulated by the clock and sugar sensing and are able to predict how the allocation and use of photosynthate responds in a far wider range of conditions. Other examples include use of emerging insights into the links between sugars, nutrient resources and hormone synthesis and action and between abiotic and biotic defence responses and growth to understand and model, and in turn probe *in silico*, the interactions and trade-offs between metabolism, growth and defence in a complex multi-factorial environment (Howe & Jander 2008; Liu *et al.* 2009; Bolouri-Moghaddam *et al.* 2010; Rouached *et al.* 2010; Domagalska & Leyser 2011; Sulpice *et al.* 2013).

POTENTIAL FRAMEWORK

Internet connectivity together with open-source software such as R-project and high-performance computing (HPC) allow plant system models to evolve far beyond the current paradigm where separate groups model the same plant or the same process in parallel, with limited or no collaboration. This current paradigm is failing to fully realize the synergies that can now be achieved by high-speed and high-volume connectivity, sharing of tools and use of modern open-source version control software to collaboratively develop a model over continents. What is envisaged to solve this? Psi would be hierarchical, with a modular framework to accept exchangeable and inter-operable components available from a library and from individuals. This library would contain representations, and alternative representations, of independent mechanisms such as photosynthesis, respiration, other metabolic pathways, underlying gene networks, organ development, root interactions with the soil, stomatal action, hydraulic flows, nutrient and carbon allocation, phenology, canopy microclimate and many others. Psi would therefore be flexible. Forms could be constructed from these components based on scientific objectives, mechanistic understanding, genomic information and/or predictive capacities. This inter-operability will require common and clearly defined names and application programming

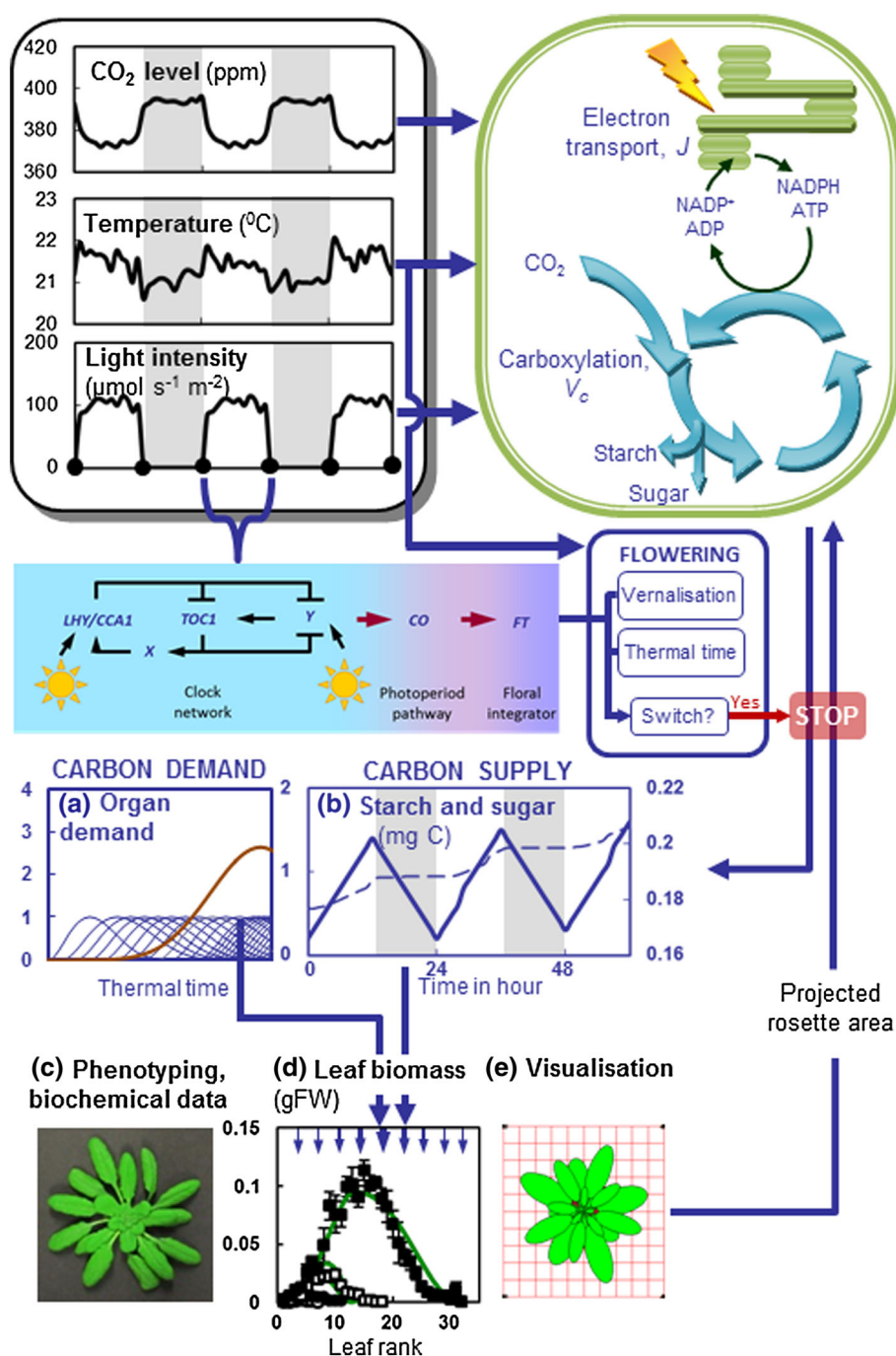


Figure 4. The *Arabidopsis* framework model's workflow predicts whole plant and individual organ growth from underlying molecular network responses. Carbon supply (b) is used as sugar (dashed line) or stored as starch (solid line). Carbon is allocated at each hourly time step according to the demand of individual leaves (a, thin blue lines) and roots (brown line). The model was validated using phenotyping and biochemical data (c). Simulation of individual leaves for the Col wild type (green lines) closely match experimental data (d) at 18 (open circles), 25 (filled circles), 27 (open squares) and 38 (filled squares) days after sowing. Leaves are ranked according to the order of appearance. The integrated model uses simulated sizes of individual leaves to calculate the projected rosette area (e) for photosynthesis, considering the spiral leaf arrangement (phyllotaxy) and upward (zenithal) angle, which determines carbon supply through photosynthesis net of respiration. Simulation of vegetative growth stops at flowering, as determined by a phenology model linked to the clock-regulated photoperiod pathway. Experimental conditions: $\sim 21.3^{\circ}\text{C}$; 12:12 light/dark cycle; light flux, $110\ \mu\text{mol m}^{-2}\text{ s}^{-1}$; mean daytime CO_2 level, 375 ppm. The error bars show the standard errors for five plants. Image adapted from Chew *et al.* (2014).

interfaces for the data and model components. The implementation of common semantics and interfaces will allow researchers to collaborate across scales and disciplines, sharing code and

data while working together on common components that are applied within different frameworks and scales (Zeigler *et al.* 2000; Grimm *et al.* 2006). Psi's modular approach will enable

researchers to evaluate different theories by exchanging and testing alternative components within a common framework, such as alternative models of respiratory carbon loss (Fig. 2).

Psi would be developed as a community resource, under an open-source license and on a publicly accessible collaborative platform such as GitHub. A suite of metrics should be available to guide ranking of modules and identify the relevant processes and level of complexity necessary to simulate a given plant or process in developing a model for a specific context. This context would be defined by the question being addressed and the relative importance of available information, including data, modules and understanding. This will allow evaluation of modules describing each process in a relevant context based on explanatory and predictive merit.

Hence, *Psi* will not be one massive program that encompasses everything a plant does. Indeed, this would seem far from possible given that it will require a complete understanding of all mechanisms underlying growth, development and interactions with environments of every plant. Furthermore, that would also violate a cardinal tenet of modelling, which is parsimony: A model should be as simple as possible to capture the relevant processes, but not simpler. The *Psi* framework would allow assessment of the cost/benefit of increasing complexity in model construction, and it would also allow groups to call upon more complex sub-models as needed. Most importantly, *Psi* should become an organized community of scholars skilled in the relevant domain expertise whose research is enabled by careful coordination and curation of modern software ecosystems.

CONCLUSION

Developments in HPC and functional knowledge of plants make achieving the plants *in silico* concept a realistic possibility, one that can be extended from model plants such as *Arabidopsis* to the major crops and to functional types in natural ecosystems. Importantly, it provides a quantitative knowledge framework where the implications of a discovery at one level, for example, single gene function or developmental response, can be examined at the whole plant or even crop and natural ecosystem levels. It requires a paradigm shift from the isolated development of disparate models that communicate via inter-model comparisons to a community effort where collaboration occurs throughout development. Here, experts in each domain would co-develop modules of a larger *Psi* framework and provide access to all, both as a teaching and research tool, via user-friendly interfaces. *Psi* will therefore generate new insights and access into plant biology as well as new modalities for collaboration and integration among plant scientists.

ACKNOWLEDGMENTS

We acknowledge the US Department of Energy workshop on 'Computational Challenges for Mechanistic Modeling of Terrestrial Environments' for initiating the discussion between JPL, SPL and XGZ that led to this Forward Look. SPL acknowledges support by the Chinese Academy of Sciences

for a Senior Scientist Visiting Professorship and from the Bill & Melinda Gates Foundation. XGZ and MS acknowledge EU Project 3to4 (number 289582). XGZ acknowledges support from Designer Breeding by Molecular Modules (XDA08020301). DSLB was supported in part by an award from the National Center for Supercomputer Applications. MS was supported by the Max Planck Society. AJM and MS acknowledge support by the EU (collaborative project TiMet under contract no. 245143).

REFERENCES

- Ainsworth E.A., Beier C., Calfapietra C., Ceulemans R., Durand-Tardif M., Farquhar G.D. & Godbold D.L. (2008a) Next generation of elevated CO₂ experiments with crops: a critical investment for feeding the future world. *Plant, Cell and Environment* **31**, 1317–1324.
- Ainsworth E.A., Leakey A.D.B., Ort D.R. & Long S.P. (2008b) FACE-ing the facts: inconsistencies and interdependence among field, chamber and modeling studies of elevated [CO₂] impacts on crop yield and food supply. *New Phytologist* **179**, 5–9.
- Beard D.A., Neal M.L., Tabesh-Saleki N., Thompson C.T., Bassingthwaite J.B., Shimoyama M. & Carlson B.E. (2012) Multiscale Modeling and Data Integration in the Virtual Physiological Rat Project. *Annals of Biomedical Engineering* **40**, 2365–2378.
- Benedict M.N., Henriksen J.R., Metcalf W.W., Whitaker R.J. & Price N.D. (2014) ITEP: an integrated toolkit for exploration of microbial pan-genomes. *BMC Genomics* **15**, 8.
- Bennett T., Hines G. & Leyser O. (2014) Canalization: what the flux? *Trends in Genetics* **30**, 41–48.
- Bishop K.A., Betzelberger A.M., Long S.P. & Ainsworth E.A. (2015) Is there potential to adapt soybean (*Glycine max* Merr.) to future CO₂? An analysis of the yield response of 18 genotypes in free-air CO₂ enrichment. *Plant Cell and Environment* **38**, 1765–1774.
- Bolouri-Moghaddam M.R., Le Roy K., Xiang L., Rolland F. & Van den Ende W. (2010) Sugar signalling and antioxidant network connections in plant cells. *Febs Journal* **277**, 2022–2037.
- Buckley T.N. & Mott K.A. (2013) Modelling stomatal conductance in response to environmental factors. *Plant, Cell and Environment* **36**, 1691–1699.
- Calfapietra C., Ainsworth E.A., Beier C., De Angelis P., Ellsworth D.S., Godbold D.L., ... Ceulemans R. (2010) Challenges in elevated CO₂ experiments on forests. *Trends in Plant Science* **15**, 5–10.
- Dietze M.C., Lebauer D.S. & Kooper R. (2013) On improving the communication between models and data. *Plant Cell and Environment* **36**, 1575–1585.
- Domagalska M.A. & Leyser O. (2011) Signal integration in the control of shoot branching. *Nature Reviews. Molecular Cell Biology* **12**, 211–221.
- Drewry D.T., Kumar P. & Long S.P. (2014) Simultaneous improvement in productivity, water use, and albedo through crop structural modification. *Global Change Biology* **20**, 1955–1967.
- Dyson R.J., Vizcay-Barrena G., Band L.R., Fernandes A.N., French A.P., Fozard J.A., ... Hodgman T.C. (2014) Mechanical modelling quantifies the functional importance of outer tissue layers during root elongation and bending. *New Phytologist* **202**, 1212–1222.
- Edwards J.S. & Palsson B.O. (2000) The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 5528–5533.
- Gibon Y., Blaessing O.E., Hannemann J., Carillo P., Hohne M., Hendriks J.H.M. & Palacios. (2004) A robot-based platform to measure multiple enzyme activities in *Arabidopsis* using a set of cycling assays: comparison of changes of enzyme activities and transcript levels during diurnal cycles and in prolonged darkness. *Plant Cell* **16**, 3304–3325.
- Grimm V., Berger U., Bastiansen F., Eliassen S., Ginot V., Giske J. & Goss-Custard J. (2006) A standard protocol for describing individual-based and agent-based models. *Ecological Modelling* **198**, 115–126.
- Hall A.J. & Minchin P.E.H. (2013) A closed-form solution for steady-state coupled phloem/xylem flow using the Lambert-W function. *Plant, Cell and Environment* **36**, 2150–2162.
- Hanson P.J., Amthor J.S., Wullschlegel S.D., Wilson K.B., Grant R.F., Hartley A., ... Cushman R.M. (2004) Oak forest carbon and water simulations: Model intercomparisons and evaluations against independent data. *Ecological Monographs* **74**, 443–489.

- Heise R., Arrivault S., Szcwoka M., Tohge T., Nunes-Nesi A., Stitt M., ... Nikoloski Z. (2014) Flux profiling of photosynthetic carbon metabolism in intact plants. *Nat. Protocols* **9**, 1803–1824.
- Hill K., Porco S., Lobet G., Zappala S., Mooney S., Draye X., ... Bennett M.J. (2013) Root systems biology: integrative modeling across scales, from gene regulatory networks to the rhizosphere. *Plant Physiology* **163**, 1487–1503.
- Howe G.A. & Jander G. (2008) Plant immunity to insect herbivores. *Annual Review of Plant Biology* **59**, 41–66.
- Li T., Hasegawa T., Yin X.Y., Zhu Y., Boote K., Adam M., ... Bouman B. (2015) Uncertainties in predicting rice yield by current crop models under a wide range of climatic conditions. *Global Change Biology* **21**, 1328–1341.
- Liu T.Y., Chang C.Y. & Chiou T.J. (2009) The long-distance signaling of mineral macronutrients. *Current Opinion in Plant Biology* **12**, 312–319.
- Long S.P. (1991) Modification of the response of photosynthetic productivity to rising temperature by atmospheric CO₂ concentrations: has its importance been underestimated? *Plant, Cell and Environment* **14**, 729–739.
- Long S.P., Marshall-Colon A. & Zhu X.G. (2015) Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. *Cell* **161**, 56–66.
- Lynch J.P. (2013) Steep, cheap and deep: an ideotype to optimize water and N acquisition by maize root systems. *Annals of Botany* **112**, 347–357.
- McGrath J.M. & Long S.P. (2014) Can the cyanobacterial carbon-concentrating mechanism increase photosynthesis in crop species? A theoretical analysis. *Plant Physiology* **164**, 2247–2261.
- Ort D.R., Zhu X.G. & Melis A. (2011) Optimizing antenna size to maximize photosynthetic efficiency. *Plant Physiology* **155**, 79–85.
- Pal S.K., Liput M., Piques M., Ishihara H., Obata T., Martins M.C.M. & Sulpice R. (2013) Diurnal changes of polysome loading track sucrose content in the rosette of wild-type *Arabidopsis* and the starchless *pgm* mutant. *Plant Physiology* **162**, 1246–1265.
- Piques M., Schulze W.X., Hohne M., Usadel B., Gibon Y., Rohwer J. & Stitt M. (2009) Ribosome and transcript copy numbers, polysome occupancy and enzyme dynamics in *Arabidopsis*. *Molecular Systems Biology* **5**.
- Poolman M.G., Miguet L., Sweetlove L.J. & Fell D.A. (2009) A genome-scale metabolic model of *Arabidopsis* and some of its properties. *Plant Physiology* **151**, 1570–1581.
- Poorter H., Anten N.P.R. & Marcelis L.F.M. (2013) Physiological mechanisms in plant growth models: do we need a supra-cellular systems biology approach? *Plant, Cell and Environment* **36**, 1673–1690.
- Postma J.A. & Lynch J.P. (2012) Complementarity in root architecture for nutrient uptake in ancient maize/bean and maize/bean/squash polycultures. *Annals of Botany* **110**, 521–534.
- Price N.D., Reed J.L. & Palsson B.O. (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews Microbiology* **2**, 886–897.
- Rhee S.Y. & Mutwil M. (2014) Towards revealing the functions of all genes in plants. *Trends in Plant Science* **19**, 212–221.
- Riedelsheimer C., Lisek J., Czedik-Eysenberg A., Sulpice R., Flis A., Grieder C., Altmann T. (2012) Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 8872–8877.
- Rouached H., Arpat A.B. & Poirier Y. (2010) Regulation of phosphate starvation responses in plants: signaling players and cross-talks. *Molecular Plant* **3**, 288–299.
- Sarikioti V., de Visser P.H.B., Buck-Sorlin G.H. & Marcelis L.F.M. (2011) How plant architecture affects light absorption and photosynthesis in tomato: towards an ideotype for plant architecture using a functional-structural plant model. *Annals of Botany* **108**, 1065–1073.
- Somerville C., Youngs H., Taylor C., Davis S.C. & Long S.P. (2010) Feedstocks for lignocellulosic biofuels. *Science* **329**, 790–792.
- Song Q., Zhang G. & Zhu X.-G. (2013) Optimal crop canopy architecture to maximize canopy photosynthetic CO₂ uptake under elevated CO₂—a theoretical study using a mechanistic model of canopy photosynthesis. *Functional Plant Biology* **40**, 109–124.
- Song Y.H., Smith R.W., To B.J., Millar A.J. & Imaizumi T. (2012) FKF1 conveys timing information for CONSTANS stabilization in photoperiodic flowering. *Science* **336**, 1045–1049.
- Steinacher A., Leyser O. & Clayton R.H. (2012) A computational model of auxin and pH dynamics in a single plant cell. *Journal of Theoretical Biology* **296**, 84–94.
- Stitt M. (2013) Systems-integration of plant metabolism: means, motive and opportunity. *Current Opinion in Plant Biology* **16**, 381–388.
- Sulpice R., Flis A., Ivakov A.A., Apelt F., Krohn N., Encke B. & Abel C. (2014) *Arabidopsis* coordinates the diurnal regulation of carbon allocation and growth across a wide range of photoperiods. *Molecular Plant* **7**, 137–155.
- Sulpice R., Nikoloski Z., Tschöep H., Antonio C., Kleessen S., Larhlmi A. & Selbig J. (2013) Impact of the carbon and nitrogen supply on relationships and connectivity between metabolism and biomass in a broad panel of *Arabidopsis* accessions. *Plant Physiology* **162**, 347–363.
- Sulpice R., Pyl E.T., Ishihara H., Trenkamp S., Steinfath M., Witucka-Wall H., Gibon Y. (2009) Starch as a major integrator in the regulation of plant growth. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 10348–10353.
- Sweetlove L.J., Williams T.C.R., Cheung C.Y.M. & Ratcliffe R.G. (2013) Modelling metabolic CO₂ evolution—a fresh perspective on respiration. *Plant, Cell and Environment* **36**, 1631–1640.
- Tholen D. & Zhu X.-G. (2011) The mechanistic basis of internal conductance: a theoretical analysis of mesophyll cell photosynthesis and CO₂ diffusion. *Plant Physiology* **156**, 90–105.
- Usadel B., Obayashi T., Mutwil M., Giorgi F.M., Bassel G.W., Tanimoto M. & Chow A. (2009) Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. *Plant, Cell and Environment* **32**, 1633–1651.
- Wang Y., Long S.P. & Zhu X.G. (2014) Elements required for an efficient NADP-Malic enzyme type C4 photosynthesis. *Plant Physiology* **164**, 2231–2246.
- Xu Y.B., This D., Pausch R.C., Vonhof W.M., Coburn J.R., Comstock J. & McCouch S.R. (2009) Leaf-level water use efficiency determined by carbon isotope discrimination in rice seedlings: genetic variation associated with population structure and QTL mapping. *Theoretical and Applied Genetics* **118**, 1065–1081.
- Zeigler B.P., Praehofer H. & Kim T.G. (2000) *Theory of modeling and simulation: integrating discrete event and continuous complex dynamic systems*. Academic Press, San Diego, 510.
- Zhu X.-G., Portis A.R. Jr. & Long S.P. (2004) Would transformation of C3 crop plants with foreign Rubisco increase productivity? A computational analysis extrapolating from kinetic properties to canopy photosynthesis. *Plant, Cell and Environment* **27**, 155–165.
- Zhu X.-G., Song Q.-F. & Ort D.R. (2012) Elements of a dynamic systems model of canopy photosynthesis. *Current Opinion in Plant Biology* **15**, 237–244.
- Zhu X.-G., de Sturler E. & Long S.P. (2007) Optimizing the distribution of resources between enzymes of carbon metabolism can dramatically increase photosynthetic rate: a numerical simulation using an evolutionary algorithm. *Plant Physiology* **145**, 513–526.
- Zhu X.G., Wang Y., Ort D.R. & Long S.P. (2013) e-photosynthesis: a comprehensive dynamic mechanistic model of C3 photosynthesis: from light capture to sucrose synthesis. *Plant, Cell and Environment* **36**, 1711–1727.
- Zhu X., Zhang G., Tholen D., Wang Y., Xin C. & Song Q. (2011) The next generation models for crops and agro-ecosystems. *Science China-Information Sciences* **54**, 589–597.

Received 17 October 2015; accepted for publication 17 October 2015