Crops in silico: Generating virtual crops using an integrative and multi-scale modeling platform

Amy Marshall-Colon[1*], Stephen P. Long[2, 1, 3*], Douglas K. Allen[4, 5], Gabrielle Allen[6, 7], Daniel A. Beard[8], Bedrich Benes[9], Susanne von Caemmerer[10], AJ Christensen[11], Donna J. Cox[11], John C. Hart[12], Peter M. Hirst[13], Kavya Kannan[1], Daniel S. Katz[11], Jonathan Lynch[14, 15], Andrew J. Millar[16], Balaji Panneerselvam[17], Nathan D. Price[18], Przemyslaw Prusinkiewicz[19], David Raila[11], Rachel G. Shekar[2], Stuti Shrivastava[1], Diwakar Shukla[17], Venkatraman Srinivasan[2], Mark Stitt[20], Matthew J. Turk[21], Eberhard O. Voit[22], Yu Wang[2], Xinyou Yin[23], Xinguang Zhu[24]

[1]Department of Plant Biology, University of Illinois, USA, [2]Carl R. Woese Institute for Genomic Biology, University of Illinois, USA, [3]Department of Crop Sciences, University of Illinois, USA, [4]Agricultural Research Service, United States Department of Agriculture, USA, [5]Donald Danforth Plant Science Center, USA, [6]Department of Astronomy, University of Illinois, USA, [7]College of Education, University of Illinois, USA, [8]Department of Molecular & Integrative Physiology, University of Michigan, USA, [9]Department of Computer Graphics Technology, Purdue University, USA, [10]ARC Centre of Excellence for Translational Photosynthesis, Research School of Biological Sciences, Australian National University, Australia, [11]National Center for Supercomputing Applications, USA, [12]Department of Computer Science, University of Illinois, USA, [13]Department of Horticulture and Landscape Architecture, Purdue University, USA, [14]Department of Plant Science, Pennsylvania State University, USA, [15]Centre for Plant Integrative Biology, University of Nottingham, United Kingdom, [16]SynthSys and School of Biological Sciences, Edinburgh University, United Kingdom, [17]Department of Chemical and Biomolecular Engineering, University of Illinois, USA, [18]Institute for Systems Biology, USA, [19]Department of Computer Science, University of Calgary, Canada, [20]Max Planck Institute of Molecular Plant Physiology, USA, [21]School of Information Science, University of Illinois, USA, [22]The Wallace H. Coulter Department of Biomedical Engineering, Georgia Tech and Emory University, USA, [23]Department of Plant Sciences, Wageningen University, Netherlands, [24]CAS Key Laboratory for Computational Biology and State Key Laboratory for Hybrid Rice, Chinese Academy of Sciences, China

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

All authors have contributed in the write up of the manuscript and edited, reviewed, and approved the final version of the manuscript.

Keywords

crop yield, Multiscale, Computational framework, Model, integration

Abstract

Word count: 219

Multi-scale models can facilitate whole plant simulations by linking gene networks, protein synthesis, metabolic pathways, physiology, and growth. Whole plant models can be further integrated with ecosystem, weather, and climate models to predict how various interactions respond to environmental perturbations. These models have the potential to fill in missing mechanistic details and generate new hypotheses to prioritize directed engineering efforts. Outcomes will potentially accelerate improvement of crop yield, sustainability, and increase future food security. It is time for a paradigm shift in plant modeling, from largely isolated efforts to a connected community that takes advantage of advances in high performance computing and mechanistic understanding of plant processes. Tools for guiding future crop breeding and engineering, understanding the implications of discoveries at the molecular level for whole plant behavior, and improved prediction of plant and ecosystem responses to the environment are urgently needed. The purpose of this perspective is to introduce Crops in silico (cropsinsilico.org), an integrative and multi-scale modeling platform, as one solution that combines isolated modeling efforts toward the generation of virtual crops, which is open and accessible to the entire plant biology community. The major challenges involved both in the development and deployment of a shared, multi-scale modeling platform, which are summarized in this prospectus, were recently identified during the first Crops in silico Symposium and Workshop.

Funding statement

Ethics statements

(Authors are required to state the ethical considerations of their study in the manuscript, including for cases where the study was exempt from ethical approval procedures)

Does the study presented in the manuscript involve human or animal subjects: No

1 **Crops *in silico*: Generating virtual crops using an integrative and multi-scale**

2 **modeling platform**

3

4 Amy Marshall-Colon[1*], Stephen P. Long[1*], Douglas K. Allen[2], Gabrielle Allen[19], Daniel A.

5 Beard[4], Bedrich Benes[5], Susanne von Caemmerer[18], AJ Christensen[3], Donna J. Cox[3], John C.

6 Hart[6], Peter M. Hirst[7], Kavya Kannan[1], Daniel S. Katz[3], Jonathan P. Lynch[8], Andrew J. Millar[9],

7 Balaji Panneerselvam[10], Nathan D. Price[11], Przemyslaw Prusinkiewicz[12], David Raila[3], Rachel G.

8 Shekar[13], Stuti Shrivastava[1], Diwakar Shukla[10], Venkatraman Srinivasan[13], Mark Stitt[14], Matthew

9 J. Turk[20], Eberhard O. Voit[15], Yu Wang[13], Xinyou Yin[16], Xin-Guang Zhu[17]

10

11 [1] Department of Plant Biology, University of Illinois, Urbana-Champaign, IL

12 [2] United States Department of Agriculture, Agricultural Research Service, and Donald Danforth

13 Plant Science Center, Louis, MO

14 [3] National Center for Supercomputing Applications, University of Illinois, Urbana-Champaign, IL

15 [4] Department of Molecular & Integrative Physiology, University of Michigan, Ann Arbor, MI

16 [5] Department of Computer Graphics Technology, Purdue University, West Lafayette, IN

17 [6] Department of Computer Science, University of Illinois, Urbana-Champaign, IL

18 [7] Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, IN

19 [8] Department of Plant Science, Penn State University, University Park, PA and Centre for Plant

20 Integrative Biology, University of Nottingham, UK

21 [9] SynthSys and School of Biological Sciences, Edinburgh University, UK

22 [10] Department of Chemical and Biomolecular Engineering, University of Illinois, Urbana-

23 Champaign, IL

24 [11] Institute for Systems Biology, Seattle, WA

25 [12] Department of Computer Science, University of Calgary, AB, Canada

26 [13] Carl R. Woese Institute for Genomic Biology, University of Illinois, Urbana-Champaign, IL

27 [14] Max Planck Institute of Molecular Plant Physiology, Golm, Germany

28 [15] The Wallace H. Coulter Department of Biomedical Engineering, Georgia Tech and Emory

29 University, Atlanta, GA

30 [16] Centre for Crop Systems Analysis, Department of Plant Sciences, Wageningen University &

31 Research, Wageningen, The Netherlands

1   [17]CAS Key Laboratory for Computational Biology and State Key Laboratory for Hybrid Rice,

2   Partner Institute for Computational Biology, Chinese Academy of Sciences, Shanghai, China

3   [18]Australian National University, ARC Centre of Excellence for Translational Photosynthesis,

4   Research School of Biological Sciences

5   [19]Department of Astronomy and College of Education, University of Illinois, Urbana-Champaign,

6   IL

7   [20]School of Information Science, University of Illinois, Urbana-Champaign, IL

8

9   **Corresponding Author:** Amy Marshall-Colon, 193 Edward R. Madigan Laboratory, University

10 of Illinois, 1201 W. Gregory Dr., Urbana, IL   61801-3838, USA; amymc@illinois.edu

11

12 total number of words: 3015 (not including section headings)

13 figures: 2

Summary (words: 219)

Multi-scale models can facilitate whole plant simulations by linking gene networks, protein synthesis, metabolic pathways, physiology, and growth. Whole plant models can be further integrated with ecosystem, weather, and climate models to predict how various interactions respond to environmental perturbations. These models have the potential to fill in missing mechanistic details and generate new hypotheses to prioritize directed engineering efforts. Outcomes will potentially accelerate improvement of crop yield, sustainability, and increase future food security. It is time for a paradigm shift in plant modeling, from largely isolated efforts to a connected community that takes advantage of advances in high performance computing and mechanistic understanding of plant processes. Tools for guiding future crop breeding and engineering, understanding the implications of discoveries at the molecular level for whole plant behavior, and improved prediction of plant and ecosystem responses to the environment are urgently needed. The purpose of this perspective is to introduce Crops *in silico* (cropsinsilico.org), an integrative and multi-scale modeling platform, as one solution that combines isolated modeling efforts toward the generation of virtual crops, which is open and accessible to the entire plant biology community. The major challenges involved both in the development and deployment of a shared, multi-scale modeling platform, which are summarized in this prospectus, were recently identified during the first Crops *in silico* Symposium and Workshop.

Introduction

Designing crops with higher yield potential and enhanced resource use efficiency is a desirable goal for future food security and sustainability. However, this is a difficult task for crop breeding and engineering programs due to unforeseen, complex traits that arise from interactions among genotype, environment, and management (GxExM). Tools that predict emergent phenotypes in response to genetic or environmental perturbations by identifying metabolic pathways or canopy forms for modification are needed to evaluate and ameliorate risks to crop yield (Yin & Struik, 2008, Srinivasan et al., 2017). The information obtained from these tools can be used to direct breeding efforts to design new germplasm (ideotypes) that can thrive in a variety of environmental scenarios.

1

The accurate reconstruction of organisms *in silico* is a timely solution toward increasing our predictive capabilities. This prospect has been at the forefront of vertebrate and microbial modeling efforts for the last two decades. Many successes have been realized from community projects based on integrative, multi-scale modeling built around a central framework and supported by their respective research communities, such as the Virtual Physiological Human (VPH) (Hunter *et al.*, 2013), the Virtual Physiological Rat (VPR) (Beard *et al.*, 2012), and in a whole cell model of *Mycoplasma genitalium* (Karr *et al.*, 2012). The VPH and VPR projects have made significant strides toward the realization of predictive medicine via working examples of integrative and multi-scale modeling (Tewari *et al.*, 2016).

Many robust models have been developed to simulate biological processes and phenotypic responses of crops to environmental parameters, including models of: the C3 and C4 photosynthetic process (Zhu et al., 2013; Wang et al., 2014); 3D plant canopies (Song et al., 2013); stomatal action (Buckley and Mott, 2013); respiration (Sweetlove et al., 2013); phloem and xylem flow (Hall and Minchin, 2013); growth and development (Prusinkiewicz and Runions, 2012); flowering (Song et al., 2012); root structural and functional dynamics (Lynch, 2013); and gene regulatory networks (Chew et al., 2014; 2017), among others. However, many isolated crop models focus on a narrow range of spatial and temporal scales, limiting their ability for extrapolation beyond measured data and resulting in inadequate prediction of crop response to new scenarios produced by perturbations (Zhu et al., 2016). There is a need to rebuild crop growth models to include the underlying mechanisms of response, reaching from gene networks and metabolic pathways through to cellular organization, tissue and organ development, and resource capture in dynamic competitive environments and ecosystems.

Despite the rich history of robust plant systems modeling (Tardieu, 2010), no coordinated effort toward the creation of a virtual physiological plant, based on integrative and multi-scale modeling, has been initiated or sustained in the plant community. While the mammalian, microbial, earth systems, hydrological, and astrophysical communities, among others, have developed methods and tools to overcome many obstacles in integrative and multi-scale modeling, and which can be adapted toward modeling plant growth, several challenges are unique to the plant community.

Some specific challenges to integrative and multi-scale modeling in plants were identified by the international community of scientists at the first symposium and workshop on Plants *in silico* (now renamed as Crops *in silico*). First, while there is only one species of human, there are dozens of food and bioenergy crops in production with important differences in primary and secondary metabolism, plant architecture, and phenology, which require different models for accurate simulations. Second, isolated modeling efforts have resulted in redundancy and a collection of models written in different markup and scripting languages due to a lack of community standards. Code and documentation for many legacy models are either difficult to find or are publicly unavailable, while many new models lack a user interface or meaningful data visualization. Another technical barrier is integrating models at different spatial and temporal scales, while social barriers include issues of intellectual property and ownership of code and model inputs/outputs. High level plant models are often inaccessible to the research community that does not have computational expertise. Many of these models could be improved with better estimates of model parameters supplied by domain experts.

To overcome these limitations, this first workshop discussed the following aims and goals. The long-term aim of Crops *in silico* is to reconstruct a functioning crop plant and community of plants from the genes upward. A secondary, but equally important, aim is to transition crop and plant modeling from many siloed efforts to a whole community effort that can benefit from the synergies that are largely absent today. Toward this aim the following goals must be fulfilled. 1) Provide a framework that enables integration of models at different levels from gene and metabolic networks to organ development and whole crop productivity. 2) Develop the framework to be crop independent, to avoid recreating common processes for each crop, such as photosynthesis (i.e. crop specification in parameter files rather than hard-wired into the code). 3) Plug-and-play capability for fine- or coarse-graining biological processes. 4) Provide a user-friendly graphical interface to facilitate use by domain experts. 5) Deliver outputs as 3D visualizations and animations. To achieve these goals, the plant sciences community must develop a close partnership with computer science to achieve a joint mission.

The Crops in silico initiative

We propose to meet the above stated challenges and goals through the creation of a discovery platform called Crops *in silico* (Cis) (http://cropsinsilico.org). Developments in high-performance computing (HPC), open-source software, and functional knowledge of plants render the Cis concept realistic and timely. Crops *in silico* is envisioned as a central framework of tools and modules that can be interconnected to solve user-defined biological questions. With a large enough collection of tools and modules we can achieve an accurate representation of a reference plant (*Arabidopsis thaliana*) as well as individuals and fields of crop plants such as rice (*Oryza sativa*), maize (*Zea mays*), soybean (*Glycine max*), and cassava (*Manihot esculenta*), spanning from the fine-grained atomistic scale, up to the coarse-grained whole plant or ecosystem scale (Figure 1). A suggested roadmap for future directions was developed by the participants of the symposium, which is outlined here, and includes: i) building a community of researchers from different domains of expertise; ii) construction of a central framework to build the virtual crops; iii) incorporation of existing, or development of new resources; and iv) continual improvement of model integration and mathematical descriptions of underlying natural processes.

Roadmap

I. *The Research Community:* To achieve transformational advances in the development of virtual crops, it is essential to build an international Crops *in silico* community comprised of experts in experimentation, agronomy, physiology, phenotyping, breeding, modeling, computer science, software development, and visualization. It is crucial for a community of domain experts to come together in collaborations and conversations to understand diverse needs, best estimates of parameters, and ongoing biological and technical questions to drive the development of the Cis framework. A Cis community has the capacity to share models and unique data sets, providing the information needed to form more complete crop models with more accurate data.

As a community we can address many of the issues that commonly plague integrative and multi-scale modeling efforts in plants. Some examples include: the adoption of a common lexicon, sustainable data and model standards to facilitate the design, implementation of software tools that enable model interoperability, and agreed rules on sharing and archiving of model components.

Others in the biological modeling community have addressed these issues by establishing organizations such as the Wheat Data Interoperability Working Group (https://rd-alliance.org/groups/wheat-data-interoperability-wg.html), or the Computational Modeling in Biology Network (COMBINE) (http://co.mbine.org/). Similarly, the Cis community must adopt and support standards and network with the larger computational biology neighborhood to take full advantage of open resources for model interoperability.

To enable this communication, annual meetings will facilitate collaboration among plant scientists and the larger modeling community in two forums. The annual science symposium and workshops will share research from isolated efforts and organize collaborative thinking toward generating virtual crops. The symposium will comprise presentations on recent advancements in modeling plants at different levels of organization and computational tools that could accelerate achieving *in silico* plants. During the workshop, participants will i) discuss the latest developments within and the linkages across science domains, ii) share new methods and resources during hands-on tutorials for users, and iii) develop strategic plans to advance the Cis project; specifically realizing the integrated framework. A separate hackathon event will develop a community of software tool developers and infrastructure resource providers, to collaborate on standards development and software support, including software integration.

II. *The Central Framework:* Critical to the success of existing multi-scale modeling efforts in other fields has been the establishment of a centralized framework capable of connecting, integrating and running models. A modeling infrastructure facilitates sharing of data and tools and has sufficient interface development to enable users at all levels of expertise to take advantage of plug-and-play modeling. Complex module integration will require collaboration among engineers, mathematicians, computer scientists, and biologists to develop a suite of pre-processing and processing resources to: i) expedite semantic reconciliation among model languages; ii) give new emphasis to model annotation using standardized ontologies and vocabularies; and iii) perform biologically driven module reduction.

Many successful frameworks already exist that the Cis community can build upon and learn from. Within the plant community there is the Virtual Laboratory (VLab) / L-studio environment

(Prusinkiewicz, 2004), which is widely used to model plant development at scales ranging from cells and tissues to individual plants to plant ecosystems using the integrative power of the L-system formalism (http://algorithmicbotany.org/virtual_laboratory/), and the OpenAlea (Pradal et al., 2008) framework, which takes a top-down approach and focuses on integrating ecophysiological and agronomic processes with plant architecture models. Cactus (Goodale *et al.*, 2003) is one such open-source component framework from the astrophysics community, for HPC in which modularity allows components to be run at different scales for different applications. Similarly, the OpenMI interface (Moore & Tindall, 2005), enables multiscale hydrology modeling and provides adopted community standards and a framework to exchange data between environmental and water management models at runtime.

The Cancer, Heart and Soft Tissue Environment (CHASTE) supports biological multi-scale models, combining models of different types in a modular fashion for several common scenarios (49 publications since 2008; http://www.cs.ox.ac.uk/chaste/). As another example, the VPR project created a modeling framework called SemGen (Neal et al., 2015) that contains software allowing users to perform fine- and coarse-graining of aggregate models at the organ and cellular levels, where specific pathways can be extracted as modules. Whereas Cactus and OpenMI rely on developers to write models in a supported language, SemGen takes advantage of adopted community standards such as CellML (https://www.cellml.org/) and SBML (http://sbml.org/Main_Page), and provides tools to annotate existing models with rich semantics when standards fail with certain model types (*e.g.*, partial differential equations). For loosely coupled simulations, where many components interact strictly through file exchanges, the Swift framework (Wilde *et al.*, 2011), which takes advantage of a variety of computational resources, might be appropriate. The proto-Crops *in silico* community includes researchers from the Cactus, VPR, and Swift projects who will advise on the construction of an appropriate framework for the plant science community.

III. *Integration of Models, Data, Tools, and Visualization*: Long-term model integration will benefit greatly from coupling well-documented transmission standards with packaged communication libraries, which will alleviate common issues related to ensuring bit-level interoperability, and allow researchers to focus on semantic interoperability. Some of these

existing resources are described below. The inclusion of information from many research groups, species, and environments will improve the quality of outputs and expand the utility of Cis to answer a diverse set of research questions. Existing software, tools, and visualization resources can be leveraged to create the modular framework capable of performing the necessary simulations. One initial method for model communication put forward in the workshop is the Advanced Message Queuing Protocol (AMQP; amqp.org). This protocol is suitable for coarse-grained communication among models written in different languages (Python, Matlab, C/C++, etc.) and operating at different time steps. MPI (open-mpi.org) is an alternative messaging protocol for models requiring fine-grained communication, while the Swift (swift-lang.org) approach was suggested to enable parallelism at different levels of granularity. We envision that the Cis framework will include a set of libraries that provide interoperability and communication between models, including the AMQP or MPI. The Cis platform will be the total environment, encompassing the user interface for viewing and launching models, libraries, data repositories, and so forth.

Model Repositories: Using the packaged Cis framework interoperability and communication libraries, models will be linked during execution (Table 1). BioModels (Chelliah *et al.*, 2015) is a large model repository which provides access to published pathway models, and has automatic conversion tools that provide downloads in multiple formats. PlaSMo (Tindal *et al.*, 2010) is a smaller and specific portal for plant growth models, which may be published or private. These range from general crop level models, such as LINTUL (http://models.pps.wur.nl/node/943), which simulates plant biomass accumulation based on light interception and efficiency in crops, to more specific plant processes, including models that cannot be represented in SBML. PlaSMo supports SBML models, including updated versions of models that are published in BioModels.

Data Repositories: The Cis framework will support access to existing experimental data repositories, such as GEO (Edgar *et al.*, 2002), the AgMIP Data Interchange (https://data.agmip.org/), CyVerse (Merchant *et al.*, 2016), KiMOSys (Costa *et al.*, 2014), and BetyDB (https://www.betydb.org/). Importantly, Cis must be able to assess the origin and reliability of data used in the various modules to prevent the propagation of errors. This will require the community to gain a sustainable consensus on strategies and metrics for evaluating the

1 credibility of both the input data and the integrated model outputs, including uncertainty

2 quantification, sensitivity analysis, error documentation, version tracking, and validation with

3 experimental data. Recognizing that the most influential models in plant sciences have resulted

4 from close interaction between model and experiment, Cis should be designed to enable all

5 interested labs to exploit the modeling framework in advancing their experimental and

6 observational studies, in turn providing validation and improvements.

7

8 *Software and Tools:* The technical integration of models requires software to overcome the dual

9 challenge of reconciling the semantics of models at different scales and successfully leveraging

10 existing tools developed for the different model types (see Table 1). Cis hopes to uncover general

11 strategies for tool and model integration and grasp where new tool development is necessary.

12

13 *Data Visualization*: In the era of "Big Data," a key challenge faced by the plant sciences

14 community is effective visualization of large experimental or simulation datasets to reveal hidden

15 insights. The Cis interface will provide access to model integration tools and enable visualization

16 of model outputs as graphs, tables, and animations. It is anticipated that interactive visualizations

17 of integrated model outputs will intuitively convey simulation dynamics and reveal emergent

18 behaviors that will help researchers to identify new questions. With the proposed modular

19 framework, it will be possible to integrate current, open-source visualization tools into the Cis

20 platform, such as Houdini (https://www.sidefx.com/) and L-Studio (Prusinkiewicz, 2004). High-

21 quality visualizations of the results from integrated and multi-scale modeling will be valuable not

22 only to domain experts, but also to inform producers, farmers, breeders, and the broad public. This

23 transition from investigator-based interactive visualization to end-user and public-based

24 presentation visualization can increase the transparency of scientific research and make it

25 understandable to a broad audience.

26

27 IV. *Sustainability*. Robert Burns wrote that the best laid schemes of mice and men often go awry

28 (English translation). This phrase can be applied to many well-meaning attempts to create

29 biological tools designed to make the lives of scientists easier or our data more meaningful.

30 However, without community buy-in and nurturing, the best laid schemes often fall by the

31 wayside. At the outset of Crops *in silico* it is critical to devote thought to sustaining this platform

1 into the future, through both financial support and cyberinfrastructure. Aside from traditional

2 sources for funding (federal and private), one option for financial support is to form a not-for-

3 profit association dedicated to sustaining the Cis platform. This route has been successfully

4 followed by the OpenMI platform, which was established in 2007 and sustained through today.

5 Key to the long-term success of Cis is frequent and open conversations among stakeholders to

6 encourage usage. This will be facilitated through the annual meetings described earlier, but also

7 by hosting workshops and other convenings to introduce, teach, and improve the Cis framework

8 and software. The Cis framework must also maintain and expand its user base by including

9 thorough documentation and tutorials that are easy to follow. To be truly user-friendly, it must

10 have an intuitive and easy-to-navigate interface; effective error handling; and, at the very least,

11 work (*i.e.*, have a robust underlying structure). This idealized tool can come to fruition by including

12 computer scientists, information technologists, and graphic artists in the Cis community.

13

14 ***Conclusion***

15

16 Famous biological models such as the Lotka-Volterra predator-prey models (Lotka, 1920;

17 Volterra, 1931), the Farquhar, von Caemmerer, and Berry model of photosynthesis (Farquhar *et*

18 *al.*, 1980), and the Hodgkin-Huxley membrane potential models (Hodgkin & Huxley, 1952) have

19 provided the scientific community with unprecedented understanding of biological processes

20 through simulations of unknown states. These and other models demonstrate the kinds of insights

21 that are only achievable through modeling. We are at a point in history where we have both the

22 need and the capability to use information and models at multiple levels to model whole systems

23 and to achieve greater insights into how whole plants and ecosystems will respond to genetic

24 changes, as well as environmental challenges never before encountered. The Crops *in silico*

25 initiative has the potential to be a powerful discovery tool in which dozens of simulations across

26 multiple scenarios can be accomplished in a few hours. It will be the first framework to enable

27 customized integration of coherent subsets of existing plant models to address specific biological

28 questions. The incorporation of an intuitive user interface with advanced visualization of integrated

29 model outputs also makes the Cis framework unique. It is evident that many technical and social

30 challenges in the development of Cis remain and will arise. However, with open communication

1 and support of the scientific community across domains of expertise, this improbable vision can

2 become a reality.

3

4 *Acknowledgements*

1   *Tables:*

2

3   **Table 1. Existing tools and resources for integrative and multi-scale modeling.**

	Resource	Description	Citation
Frameworks	Cactus	Problem solving framework that enables parallel computation across scales.	(Goodale *et al.*, 2003)
	SemGen	Tool to automate modular composition and decomposition of biosimulation models	(Gennari *et al.*, 2011; Neal *et al.*, 2015)
	FLAME	Agent-based modeling system that scales from laptops to HPC and parallel super computing	(Holcombe *et al.*, 2006; Kiran *et al.*, 2010)
	OpenMI	Software for independent model exchange at run time.	(Moore and Tindall, 2005)
	Swift	Parallel scripting system for many task workflows.	(Wilde, *et al.*, 2011)
	VLab/ L-studio	Modeling and simulation of plant development from genes to ecosystems	(Prusinkiewicz 2004)
	OpenAlea	Visualization and modeling of plant architecture.	(Pradal *et al*., 2008)
Model/data repositories	PlaSMo	Database for plant growth models and interface,	(Tindal *et al.*, 2010)
	BioModels	Database with biochemical and non-biochemical models, MIRIAM compliant	(Chelliah *et al.*, 2015)
	GEO	Data repository for high throughput genomic datasets, utilizing MIAME standards	(Edgar *et al.*, 2002)
	CyVerse	Repository for tools for developing data storage pipeline	(Hanlon *et al.*, 2015; Merchant *et al.*, 2016)
Semantic reconciliation	SBOL	Standard synthetic biology open language	(Roehner *et al.*, 2016)
	JSim	Utilizes mathematical modeling language for writing models and annotation	(Butterworth *et al.*, 2013)
	COMBINE	Initiative to develop a set of interoperable and non-overlapping standards for modeling	(Hucka *et al.*, 2015)

4

5

13

1   *Figures Legends:*

2

3   Figure 1. Layers of organization of biological models across temporal and spatial scales. The y-

4   axis represents real-time in which changes occur at each biological level; the x-axis represents the

5   relative size or space which the biological level encompasses. The arrows indicate possible direct

6   interactions among scales. Organ level image is from Kim *et al.*, 2001.

1 *References*

2

3 Beard DA, Neal ML, Tabesh-Saleki N, Thompson CT, Bassingtwaighte JB, Shimoyama M,

4 Carlson BE. 2012. Multiscale modeling and data integration in the Virtual Physiological Rat

5 Project. *Annals of Biomedical Engineering*, **40**: 2365–2378. doi: 10.1007/s10439-012-0611-7.

6

7 Buckley, T. N. and Mott, K. 2013. Modelling Stomatal Conductance in Response to

8 Environmental Factors. *Plant, Cell, and Environment* 36: 1691–1699. doi:10.1111/pce.12140

9

10 Butterworth E, Jardine BE, Raymond GM, Neal ML, Bassingthwaighte JB. 2013. JSim, an open-

11 source modeling system for data analysis. *F1000Research* **2**: 288. doi:10.12688/f1000research.2-

12 288.v1.

13

14 Chelliah V, Juty N, Ajmera I, Ali R, Dumousseau M, Glont M, Hucka M, Jalowicki G, Keating

15 S, Knight-Schrijver V, *et al.* 2015. BioModels: Ten-year anniversary. *Nucleic Acids Research*

16 **43**: D542–D548. doi: 10.1093/nar/gku1181.

17

18 Chew, Y.H.,. Seaton, D.D., Mengin, V., Flis, A., George, G.M., , Mugford, S.T., Moulin, M.,

19 Zardilis, A., FitzPatrick, T., Smith, A.M., Zeeman, S.C., Stitt, M., and J. Millar, A.J. 2017.

20 Linking circadian time to growth rate quantitatively via carbon metabolism. *bior$_x$iv* doi:

21 https://doi.org/10.1101/105437

22

23 Chew, Y. H., Wenden, B., Flis, A., Mengin, V., Taylor, J., Davey, C. L., Tindal, C., Thomas, H.,

24 Ougham, H. J., Reffye, P. De, Stitt, M., Williams, M., Muetzelfeldt, R., Halliday, K. J., Millar,

25 A. J. 2014. Multiscale Digital Arabidopsis Predicts Individual Organ and Whole-Organism

26 Growth. *PNAS,* USA. e4127-e4136. doi:10.1073/pnas.1410238111

27

28 Costa RS, Veríssimo A, Vinga S. 2014. KiMoSys : a web-based repository of experimental data

29 for KInetic MOdels of biological SYStems. *BMC Systems Biology* **8**: 85.

30

1 Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and

2 hybridization array data repository. *Nucleic acids research* **30**: 207–10. doi: 10.1186/s12918-

3 014-0085-3.

4

5 Farquhar, GD, von Caemmerer S, Berry JA. 1980. A biochemical model of photosynthetic CO2

6 assimilation in leaves of C 3 species. *Planta* **149(1)**:78-90. doi: 10.1007/BF00386231.

7

8 Gennari JH, Neal ML, Galdzicki M, Cook DL. 2011. Multiple Ontologies in Action: Composite

9 Annotations for Biosimulation Models. *Journal of Biomedical Informatics* **44(1)**: 146–154.

10 doi:10.1016/j.jbi.2010.06.007 .

11

12 Goodale T, Allen G, Lanfermann G, Massó J, Radke T, Seidel E, Shalf J. 2003. The Cactus

13 framework and toolkit: Design and applications. In Vector and Parallel Processing –

14 VECPAR'2002, 5th International Conference, Lecture Notes in Computer Science, Berlin,

15 Springer. doi: 10.1007/3-540-36569-9_13.

16

17 Hall, A. J. and Minchin, P. E. H. 2013. A Closed-Form Solution for Steady-State Coupled

18 Phloem/xylem Flow Using the Lambert-W Function. *Plant, Cell, and Environment* 36: 2150–

19 2162. doi:10.1111/pce.12125

20

21 Hanlon MR, Vaughn M, Mock S, Dooley R, Moreira W, Stubbs J, Town C, Miller J,

22 Krishnakumar V, Ferlanti E, *et al.* 2015. Araport: An application platform for data discovery.

23 *Concurrency Computation* **27**: 4412–4422. doi:10.1002/cpe.3542.

24

25 Hodgkin AL, Huxley AF. 1952. A quantitative description of membrane current and its

26 application to conduction and excitation in nerve. *The Journal of Physiology* **117(4)**: 500–544.

27 doi:10.1113/jphysiol.1952.sp004764.

28

29 Holcombe M, Coakley S, Smallwood R. 2006. A General Framework for Agent-based

30 Modelling of Complex Systems. *Proceedings of the 2006 European Conference on Complex*

Systems. URL
http://www.cabdyn.ox.ac.uk/complexity_PDFs/ECCS06/Conference_Proceedings/PDF/p25.pdf

Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, and the rest of the SBML
Forum:, Arkin AP, Bornstein BJ, Bray D, *et al.* 2003. The systems biology markup language
(SBML): a medium for representation and exchange of biochemical network models.
Bioinformatics **19**: 524–531. doi: 10.1093/bioinformatics/btg015.

Hunter P, Chapman T, Coveney PV, de Bono B, Diaz V, Fenner J, Frangi AF, Harris P, Hose R,
Kohl P, *et al.* 2013. A vision and strategy for the virtual physiological human: 2012 update.
Interface Focus **3**: 20130004. doi: 10.1098/rsfs.2013.0004.

Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival B Jr, Assad-Garcia N,
Glass JI, Covert MW. 2012. A whole-cell computational model predicts phenotype from
genotype. *Cell* **150(2)**:389-401. doi: 10.1016/j.cell.2012.05.044.

Kim MS, McMurtrey JE, Mulchi CL, Daughtry CST, Chappelle EW, Chen YR. 2001. Steady-
state multispectral fluorescence imaging system for plant leaves. *Applied Optics* **40**:157-166. doi:
10.1364/AO.40.000157.

Kiran M, Richmond P, Holcombe M, Chin LS, Worth D, Greenough C (2010) FLAME:
simulating large populations of agents on parallel hardware architectures. *Proceedings of the 9th
International Conference on Autonomous Agents and Multiagent Systems*. URL
http://dl.acm.org/citation.cfm?id=1838517.

Lotka AJ. 1920. Analytical Note on Certain Rhythmic Relations in Organic Systems. *Proc. Natl.
Acad. Sci. U.S.A.* **6:** 410–415.

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. doi:10.1093/aob/mcs293

1   McGrath, J. M. and Long, S. P. 2014. Can the Cyanobacterial Carbon-Concentrating Mechanism

2   Increase Photosynthesis in Crop Species? A Theoretical Analysis. *Plant Physiology* **164**: 2247–

3   61. doi:10.1104/pp.113.232611

4

5   Merchant N, Lyons E, Goff S, Vaughn M, Ware D, Micklos D, Antin P. 2016. The iPlant

6   Collaborative: Cyberinfrastructure for Enabling Data to Discovery for the Life Sciences. *PLOS*

7   *Biology* **14**: e1002342. doi: 10.1371/journal.pbio.1002342.

8

9   Moore, R.V., and Tindall, C.I. (2005). An overview of the open modelling interface and

10   environment (the OpenMI). *Environmental Science & Policy* **8**, 279-286. doi:

11   10.1016/j.envsci.2005.03.009.

12

13   Neal ML, Carlson BE, Thompson CT, James RC, Kim KG, Tran K, Crampin EJ, Cook D,

14   Gennari JH. 2015. Semantics-based composition of integrated cardiomyocyte models motivated

15   by real-world use cases. *PLoS One* **10(12)**:e0145621. doi: 10.1371/journal.pone.0145621.

16

17   Pradal C, Dufour-Kowalski S, Boudon F, Fournier C, Godin C. 2008. OpenAlea: A Visual

18   Programming and Component-Based Software Platform for Plant Modeling. *Functional Plant*

19   *Biology* **35**: 751–760. doi:10.1071/FP08084.

20

21   Prusinkiewicz P. 2004. Art and science for life: Designing and growing virtual plants with L-

22   systems. *Acta Horticulturae* **630**: 15-28. doi: 10.17660/ActaHortic.2004.630.1

23   Prusinkiewicz, P. and Runions, A. 2012. Computational models of plant development and form.

24   *New Phytologist* **193**: 549-569. **doi:** 10.1111/j.1469-8137.2011.04009.x

25

26   Roehner N, Beal J, Clancy K, Bartley B, Misirli G, Grünberg R, Oberortner E, Pocock M, Bissell

27   M, Madsen C, *et al.* 2016. Sharing Structure and Function in Biological Design with SBOL 2.0.

28   *ACS Synthetic Biology*, acssynbio.5b00215. doi:10.1021/acssynbio.5b00215.

29

1 Srinivasan V, Kumar P, Long SP. 2017. Decreasing, not increasing, leaf area wil raise crop

2 yields under global atmospheric change. *Global Change Biology* **23**(4): 1626-1635.

3

4 Song, Y. H., Smith, R. W., To, B. J., Millar, A. J., Imaizumi, T. 2012. FKF1 Conveys Timing

5 Information for CONSTANS Stabilization in Photoperiodic Flowering. *Science* **336**: 1045–1049.

6 doi:10.1126/science.1219644

7

8 Song, Q., Zhang, G., and Zhu, X. G. 2013. Optimal Crop Canopy Architecture to Maximise

9 Canopy Photosynthetic CO 2 Uptake under Elevated CO2-A Theoretical Study Using a

10 Mechanistic Model of Canopy Photosynthesis. *Functional Plant Biology* **40**: 109–124.

11 doi:10.1071/FP12056

12

13 Sweetlove, L. J., Williams, T. C. R., Cheung, C. Y. M., and Ratcliffe, R. G. 2013. Modelling

14 Metabolic CO2 Evolution - A Fresh Perspective on Respiration. *Plant, Cell and Environment* **36**:

15 1631–1640. doi:10.1111/pce.12105

16

17 Tardieu, F. 2010. Why work and discuss the basic principles of plant modelling 50 years after

18 the first plant models? *Journal of Experimental Botany*, **61**: 2039-2041.

19

20 Tewari, S.G., Bugenhagen, S.M., Palmer, B.M. and Beard, D.A., 2016. Dynamics of cross-

21 bridge cycling, ATP hydrolysis, force generation, and deformation in cardiac muscle. *Journal of*

22 *molecular and cellular cardiology*, **96:** 11-25. doi: 10.1016/j.yjmcc.2015.02.006..

23

24 Tindal C, Muetzelfeldt R, Ougham HJ, Millar A (ed). 2010. Plant Systems Modeling portal,

25 PlaSMo. [WWW publication] URL http://www.research.ed.ac.uk/portal/en/publications/plant-

26 systems-modeling-portal-plasmo(71191cd6-422b-4e64-8031-b0c08d5efbed).html. [accessed 6

27 August 2016].

28

29 Volterra V. 1931. Variations and fluctuations of the number of individuals in animal species

30 living together in Animal Ecology, Chapman, R.N. (ed), McGraw–Hill.

1

Wang, Y., Long, S. P., and Zhu, X.-G. 2014. Elements Required for an Efficient NADP-Malic

Enzyme Type C4 Photosynthesis. *Plant Physiology* **164**: 2231–46. doi:10.1104/pp.113.230284

4

Wilde M, Hategan N, Wozniak JM, Clifford B, Katz DS, Foster I. 2011. Swift: A language for

distributed parallel scripting. *Parallel Computing* **37(9)**: 633-652. DOI:

10.1016/j.parco.2011.05.005.

8

Yin X, Struik PC. 2008. Applying modelling experiences from the past to shape crop systems

biology: The need to converge crop physiology and functional genomics. *New Phytologist* **179**:

629-642. 10.1111/j.1469-8137.2008.02424.x.

12

Zhu, X.-G., Y. Wang, D. Ort, and S. Long. 2013. e-Photosynthesis: a comprehensive dynamic

mechanistic model of C3 photosynthesis: from light capture to sucrose synthesis. *Plant Cell &*

Environment **36**:1711-1727.

16

Zhu XG, Lynch JP, LeBauer DS, Millar AJ, Stitt M, Long SP. 2016. Plants *in silico*: why, why

now and what?—an integrative platform for plant systems biology research. *Plant, Cell, and*

Environment **39(5)**: 1049-1057. DOI: 10.1111/pce.12673.

20

Figure 1.JPEG