

iPlant Grand Challenge Workshop in Computational Approaches to Plant Development: Computational Morphodynamics of Plants

Organizers

Primary Contact: Eric Mjolsness, University of California, Irvine, Institute for Genomics and Bioinformatics (IGB), and Departments of Computer Science and Mathematics

ph.: (949) 824-3533, fax: (949) 824-4056
emj@uci.edu ; <http://www.ics.uci.edu/~emj>

Eric Mjolsness is an Associate Professor in the Department of Computer Science, with a joint appointment in Mathematics, and is program leader for systems biology in the Institute for Genomics and Bioinformatics at the University of California, Irvine. By means of the Computable Plant collaboration co-organized with Meyerowitz, he has initiated a wave of successful plant development modeling at the level of cells and tissues. He has developed numerous mathematical modeling methods for systems biology.

Philip N. Benfey, Duke University, Department of Biology

ph.: (919) 613-8182, fax: (919) 660-7338
philip.benfey@duke.edu ; <http://fds.duke.edu/db/aas/Biology/faculty/philip.benfey/>

Christophe Godin, INRIA Montpellier, Division of Plant Modelling and Computer Science

ph.: 33 (0)4 67 61 65 77 , fax: 33 (0)4 67 61 56 68
christophe.godin@inria.fr; <http://www-sop.inria.fr/virtualplants/personnel/Christophe.Godin/godin.php>

B.S. Manjunath, UC Santa Barbara, Electrical and Computer Engineering Department

ph.: (805) 893-7112, fax: (805) 893-3262
manj@ece.ucsb.edu; <http://vision.ece.ucsb.edu/manjunath/>

Manjunath's research interests include image/video analysis (including texture and shape analysis, segmentation, registration), multimedia databases and data mining (feature extraction, content based access, high dimensional indexing and similarity search), steganography (data hiding in images and video, and their detection), and signal/image processing for bio-informatics. He directs the NSF/ITR funded Bio-Image Informatics Center and NSF sponsored IGERT program on Interactive Digital Multimedia. He is also associated with the interdisciplinary Multimedia Arts and Technology program and the Computer Science Department.

Elliot Meyerowitz, California Institute of Technology, Division of Biology

ph.: (626) 395-6889, fax: (626) 449-0756
meyerow@caltech.edu ; <http://www.its.caltech.edu/~plantlab/>

Meyerowitz is an Arabidopsis developmental geneticist, with experience in plant genetics, genomics, and molecular biology whose lab is currently pioneering methods for live imaging and live image analysis of growing plants at the cellular and subcellular level. He has started, with others, three outreach programs - one an NIH-funded program connecting Caltech graduate students, postdocs, and faculty with national teacher-training efforts organized at the Exploratorium in San Francisco, and the other an NSF-FIBR and Arthur Vining Davis Foundation-funded summer teacher training program, and an American Society of Plant Biologists-funded school-term teacher training program.

Edgar Spalding, University of Wisconsin, Department of Botany

ph.: (608) 265-5294
spalding@wisc.edu; <http://www.botany.wisc.edu/spalding/bio.htm>

Summary of the Proposed Workshop:

A grand challenge facing the plant biology community is to understand how plants develop and grow - that is, how genomic and environmental information combine to produce a three-dimensional and highly dynamic plant consisting of intercommunicating cells and tissues. The proposed workshop will develop plans to create a novel computational approach that will bridge existing gaps between biochemical, cellular, and organismal levels of understanding. Central to the approach is the notion that much of the necessary bridging information can be computationally extracted from images and used to inform computational models that reveal the causal relationships between biochemical and genetic activities of cells and cellular activities in development on one hand, and between cellular activities in development and the structure and function of tissues at the organismal level, on the other.

There are three specific areas of need: acquisition and analysis of images of cells, tissues, and organ systems actively undergoing morphogenesis; computational modeling, to represent and test specific hypotheses for the mechanisms by which cellular activities and cell-cell communication and coordinated growth result in defined and dynamic cellular patterns; and mechanical modeling, to provide computer representations of plant tissues in which each cell, its mechanical influence on its tissue, and the results of mechanical force on its enlargement and division combine to provide a realistic and testable substrate for developmental models. The grand challenge will be greatly served by a cyberinfrastructure that facilitates the flow of information about growth and development from image acquisition devices to interoperable databases to modeling functions.

The 2-day workshop is expected to involve approximately 30 participants. (A followup day of strictly optional technical discussions may also be scheduled.) The opportunities and present barriers to progress in the three different areas of need will be assessed. Participants will create a strategic plan for developing cyberinfrastructure required for image acquisition and analysis at the subcellular, cellular and organ levels; computational modeling of pattern formation in developing plant tissues; and mechanical models of plant tissues using finite element models at the cellular and tissue level. An expected outcome is that close and well-defined links between the iPlant central organization and the individual laboratories that have already developed tools and/or data sets will emerge as a key element of a feasible Grand Challenge project. Workshop time will also be dedicated to exploring educational opportunities, with emphasis on plant and computation lessons for middle and high school biology classes, and on the possibilities for the products of the project to influence undergraduate research and pedagogy.

Grand Challenge Workshop Description

We will convene a workshop directed to the grand challenge of *discovering how plant morphology arises from and feeds back on cellular, biochemical and informational processes*. This fundamental question of developmental biology when applied to plants necessarily includes the novel aspect of information flowing not only from cellular to morphological levels via developmental and growth processes, but also from morphology to cellular and biochemical process through cytoskeletal and hormonal responses to mechanical force, which conditions differential cell growth, cellular anisotropy via cellulose deposition, and cell division planes. We believe that cyberinfrastructure deficiencies presently impede the image analysis and modeling activities required for understanding in a unified way the two-way flow of developmentally relevant information between the biochemical, cellular, morphological, and organismal levels of organization. The workshop will result in a plan to address the key deficiencies.

Proposed Workshop

The workshop will address three interrelated cyberinfrastructure areas: image analysis, modeling of gene and cell-cell communication networks, and modeling of the mechanical properties of plant cells and tissues; and also how progress in these areas can contribute to outreach and educational opportunity. New cyberinfrastructure is required for image analysis, from the point of image acquisition through the extraction of machine-usable information. The principle has been proved – useful information can be automatically extracted from images of roots, shoots, leaves, flowers, and embryos, sometimes even as

they grow and with high spatiotemporal resolution, resulting in high resolution on growth patterns and growth rates. It has also been demonstrated that biochemical/developmental modeling employing systems of differential or stochastic equations can describe cellular and tissue activities such as enzymatic pathways, gene regulatory networks, and networks of cell-cell communication that define growth and differentiation. Realities of plant development further require modeling that accounts for cellular level mechanics reflective of plant-unique turgor pressures and cell wall properties, and large-scale tissue movements (such as buckling) that feed back into cellular activities. What is needed is a robust cyberinfrastructure that supports and links each of these areas of activity.

The workshop will bring together biologists and computational scientists with experience, interests, and/or data related to the three main scientific areas, each of which will be the subject of a specific session. Sessions will consist of plenary level lectures followed by parallel breakout groups. Discussions will be focused on identifying the most critical cyberinfrastructure deficiencies impeding specific avenues of research. Each session will have the goal of outlining the type of data to be used and the eventual computational product to be produced. Each group will also be asked to supply names of others who should be informed and invited to participate.

Each session will feature teams of participants with different expertise and experience. The image analysis discussions will address whole plant imaging (led by Edgar Spalding), analysis of confocal microscope images, and image databases (to be covered by Robert Murphy and B.S. Manjunath). Computational modeling of development at the cellular level will be led by Eric Mjolsness, and will include discussions of models of pattern formation in the shoot apical meristem, root apical meristem, leaf, and flower. Mechanical models will be led by Marcus Heisler, and will include discussions of finite element models (FEM) of plant tissues at the cellular level, and mechanical models of plant tissues at the organismal level. A session on outreach and educational opportunities and strategies will be led by Martha Kirouac.

Datasets:

The datasets that currently exist, or that are expected in the next couple of years include large sets of live imaging data of a variety of reporter genes and proteins in shoot apical meristems (see Heisler et al., 2005; Gordon et al., 2007 for examples); live imaging data at the cellular level of root growth in a variety of conditions (e.g. Grieneisen et al., 2007) and <http://www.plantsci.cam.ac.uk/Haseloff/>; thousands of time-lapse image series of seedling roots and hypocotyls responding to gravity and light (see Miller et al., 2007 and <http://phytomorph.wisc.edu/phytomorph.htm>); large number of electron tomographs with ultrahigh subcellular resolution obtained by Marisa Otegui. In addition to these Arabidopsis data sets, we expect to use live images of growth of the alga *Coleochaete* (<http://www.plantsci.cam.ac.uk/Haseloff/>). Confocal microscope imagery of the shoot apex will be made available to the iPlant project from at least the laboratories of Jan Traas (Lyon), Chris Kuhlemeier (Bern), and Elliot Meyerowitz. Sepal imagery will be provided by Adrienne Roeder, currently in Meyerowitz's laboratory. Furthermore, a major consortium of UK laboratories is developing live imaging data for root development, and they have agreed to the eventual release of all imagery. This project is the CPIB (Centre for Plant Integrative Biology) at the University of Nottingham - <http://www.cpiib.info/>. Each of these is sufficient for the development of the image analysis infrastructure. For computational modeling, the existing datasets are relevant to specific problems. Phyllotaxis is represented by the literature on auxin transport and response of shoot apical meristem cells to auxin, as summarized in Heisler et al., 2005. Root cell type specification has its own literature and data sets, on gene expression in specific cell types it is (Birnbaum et al., 2003), for auxin gradient control (Grieneisen et al. 2007) and the references therein, and for responses to gravity (<http://phytomorph.wisc.edu/phytomorph.htm>). Input information for development of modeling infrastructure also includes early attempts at modeling environments, for example Cellerator (Shapiro et al., 2003).

Current State of Computational Thinking in the Field

Computational thinking is beginning to seep into the field of plant development, although the tools

are far too difficult to use to engage most of the plant biologists whose research could benefit. Substantial hardware resources are virtually unused, though there is clearly a need for them in both large-scale tissue simulation and in automated parameter optimization/model selection computations. The Workshop will discuss the means by which these existing tools can be suited to use by plant scientists.

Leading computational technologies in the field of plant developmental modeling include those generated by the Computable Plant project (Mjolsness, Meyerowitz, and collaborators) outlined below, as well as L-Systems simulation and virtual laboratory software as long developed by P. Prusinkiewicz and collaborators, C. Godin's OpenAlea software platform for virtual plants, and Cellular Potts Models (Grieneisen et al. 2007), among others. The Computable Plant project approaches include the simple Cellzilla extension of Cellerator to generate fixed-template tissue models from reaction network arrows, weak spring models, the Organism/FEM C++ program for simulating weak spring and FEM mechanics combined with ODE regulatory models, and the fundamental Dynamical Grammars (DG) framework for variable-structure systems such as arise in development, along with its "Plenum" prototype implementation. DG's are fundamental because they combine time-evolution operators in the mathematically simplest way: addition. A potentially related project is the Sigmoid web-accessible GUI and model database wrapped around Cellerator, which currently however is limited to fixed-structure models.

Other developmental modeling approaches used in biological development generally include noncellular reaction-diffusion partial differential equations, membrane energy functions accounting for stretch and adhesion, P-systems and "membrane computing", λ -calculus (Danos and Laveve 2004), level set methods, the MGS simulator with dynamical topology by Giavitto, dynamic topological cell complexes (Brisson 1993), polygonal vertex models (Hufnagel et al. 2007), and spatial stochastic modeling using the R-leap algorithm (Auger et al. 2006). Mathematical objects of interest that may be captured in such models include emergent symmetries, instabilities, and patterns of deformation.

A botanically universal modeling environment must simulate heterogeneous dynamics, which is possible using shallow, black-box style integration of disparate solvers (for example via second-order operator splitting methods) or much more efficiently by less general but deeper integration of numerical methods. Examples of the latter include the combination of stochastic discrete events and continuously operating ODE's in Plenum and (less accurately for the ODE's) L-system software, and many other solution algorithms (some yet to be invented) that require full access to the solution algorithm source code. Model exchange formats such as SBML with suitable future extensions, or CellML+FieldML, may be able to usefully represent developmental models.

Thus, there is a wealth of useful starting points for developmental modeling cyberinfrastructure.

But taken individually, none of these methods is fully scalable to the Cyberinfrastructure challenge.

Problems of inadequate use of standard mathematics, such as for stochastic processes and their integration with accurate ODE solvers in continuous time, are widespread. It is by a synthesis of the best that the developmental modeling community has to offer, that a useful cyberinfrastructure will be delivered.

Barriers that currently prevent effective utilization of available data:

Image analysis

Barrier: Absence of image analysis tools sufficient for segmentation of plant images at the cellular and subcellular level, and absence of ability to correspond features in time-series images in cellular and whole-plant images. The variety of image formats, not necessarily compatible with all tools, is a problem being addressed by the Open Microscopy Environment (<http://www.openmicroscopy.org/>). The need for distributed or high throughput computing and workflow tools to manage the analysis of many images generated by automated acquisition platforms is unmet.

Finite element mechanical modeling

Barrier: The lack of any FEM elastic dynamics code that can incorporate continual cell growth and frequent cell division. Cell division requires frequent remeshing of the dividing cells and often their neighbors. The usual assumption is for infrequent global remeshing, which is inadequate. Very strong 3D

data structures and algorithms will be required to support these operations especially on large, organism-sized simulations that may have to be parallelized even within a single run. Multiple commercial engineering codes have been tried (at Caltech CACR and at CPIB) and found lacking in these areas.

Barrier: The lack of available-source and open-source FEM code that does the foregoing. Source availability is essential for deep integration, beyond the low-order accuracy available through operator splitting (via interleaved execution), of simulation engines for mechanical processes, (possibly stochastic) biochemical processes, and growth processes. This combination is virtually the definition of Computational Morphodynamics, and the key to plant developmental modeling. Open-source code is a key facilitating factor for continued progress in these areas.

Developmental modeling

Barrier: The lack of a mathematically and physically well-founded, computationally well-engineered plant development modeling system.

Barrier: The lack of parallel computing in either single-run simulations of large 3D tissue models, or even in parameter searches.

Barrier: The lack of a standardized systems biology model exchange format (such as SBML) that can be used to model large, developing tissues. SBML has a disastrous built-in deficiency from this point of view: the lack of array and dynamical array constructs. This implies that as a modeler increases the number of cells in a tissue model, the size of the SBML file must increase proportionately, and that an arbitrary-sized growing tissue cannot be represented at all. This is a barrier to developmental modeling to at least the extent that a standardized model exchange language is useful for coordinating or combining many modeling efforts. We have made a formal SBML proposal for dynamical arrays (Shapiro et al. 2004), but no-one admits to having the resources to pursue it.

Barriers that can be addressed with additional datasets:

These can only be identified after the development of the cyberinfrastructure using the existing datasets described above - once knowledge has been generated from the existing data, it should become clear if there is a need for any different amount of type of data to gain full understanding of plant growth and development.

Plan for diverse participation:

Our plan is to have broad geographical representation, including international participation; and attendance of scientists from different types of institution of higher education, from large state universities (e.g. Mjolsness, University of California; Spalding, University of Wisconsin), large private universities (e.g. Szymanski, Purdue and Murphy, Carnegie Mellon), small private research universities (e.g. Meyerowitz, Caltech), small liberal arts colleges (e.g. Kramer, Simon's Rock College) and in addition research institutes (e.g. Jackson, Cold Spring Harbor labs); and industry, both large (e.g. Liu, Monsanto) and small (e.g. Michaels, Phenotype Screening Inc.). We will also have representatives of private educational institutions that are not universities, e.g. Kirouac, Huntington Botanical Gardens). As the list of potential participants makes clear, gender diversity is also accounted for, including representation from women's colleges (e.g. Shepard, Barnard).

Assessment of Education, Outreach and Training opportunities that derive from potential GC questions in this field.

There is an opportunity to engage students in the life sciences through dissemination of web-accessible and visually appealing interactive simulations of plant development. Once available, software to measure plant growth or root curvature could form the basis for new student laboratories at the high school or college level. The methods developed could also be part of teacher education programs (such as the

Grounding in Botany program at the Huntington Botanical Garden) in which the teachers develop new lesson plans and high school laboratories that would use computational approaches equally with standard laboratory procedures. Thus, new and existing outreach efforts could be greatly bolstered by appropriate use of such cyberinfrastructure.

Collateral activities

We will submit a proposal to the Kavli Institute of Theoretical Physics, UC Santa Barbara, as we have been invited to convene the best physics-based plant developmental modeling minds and directions as a followup to the Spring 2008 Morphogenesis workshop at this prominent theoretical physics institute.

References:

- Auger, A, Chatelain, P, Koumoutsakos, P (2006) R-leaping: Accelerating the stochastic simulation algorithm by reaction leaps. *Journal Of Chemical Physics* **125**, 084103.
- Birnbaum, K, Shasha, DE, Wang, JY, Jung, JW, Lambert, GM, Galbraith, DW, Benfey, PN (2003) A gene expression map of the Arabidopsis root. *Science* **302**, 1956-1960.
- Brisson, Representing structures in d dimensions: Topology and order. *Discrete Comput. Geom.* **9**, 387-426), 1993.
- Danos and Laveve, Formal molecular biology. *Theoretical Computer Science* 325(1) 2004, 69-110.
- Hufnagel L, Teleman AA, Rouault H, Cohen SM, Shraiman BI, On the mechanism of wing size determination in fly development. *Proc Natl Acad Sci U S A.* 2007 Mar 6;104(10):3835-40.
- Grieneisen, VA, Xu, J, Maree, AFM, Hogeweg, P, Scheres, B (2007) Auxin transport is sufficient to generate a maximum and gradient guiding root growth. *Nature* 449, 1008-1013.
- Marée A.F.M. & Hogeweg P. (2001), How amoeboids self-organize into a fruiting body: Multicellular coordination in Dictyostelium discoideum. *Proc. Natl. Acad. Sci. USA*, 98, 3879-3883.
- Miller ND, Parks BM, Spalding EP (2007) Computer-vision analysis of seedling responses to light and gravity. *The Plant Journal* 52: 374-381.
- Shapiro, B.E., Levchenko, A., Meyerowitz, E.M. Wold, B.J., and Mjolsness, E.D. (2003) Cellerator: Extending a computer algebra system to include biochemical arrows for signal transduction. *Bioinformatics* **19**, 677-678.
- Shapiro, B.E., Gor, V. and Mjolsness, E. "Systems Biology Markup Language (SBML) Level 3 Proposal: Dynamic Arrays", computableplant.ics.uci.edu/DynamicArrays.pdf ; <http://www.sbml.org/Forums/index.php?t=tree&goto=1399&rid=0> . December 2004

URLs:

L Systems: <http://www.algorithmicbotany.org>

MGS: <http://mgs.ibisc.univ-evry.fr/>

P Systems: <http://ppage.psystems.eu/>

OpenAlea: <http://openalea.gforge.inria.fr/dokuwiki/doku.php>

CellML: <http://www.cellml.org>