The Computable Plant: A Software Architecture for Developmental Modeling in Plants

Eric Mjolsness[1,*], Bruce E Shapiro[2], Victoria Gor[2], Henrik Jönsson[3], G Venugopala Reddy[4], Marcus Heisler[4], and Elliot M Meyerowitz[4]

(1) Institute for Genomics and Bioinformatics, University of California, Irvine CA USA; (2) Jet Propulsion Laboratory, California Insitute of Technology, Pasadena CA, USA; (3) Department of Theoretical Physics, Complex Systems Division, Lund University, SWEDEN; (4) Division of Biology California Insitute of Technology, Pasadena CA, USA.

The Problem: Biological Development

How does an organism's genetic makeup interact with its environment to shape the intricate developmental processes that lead to functional tissues, organs and organisms from undifferentiated cells?

This question has long challenged biologists. Researchers have traditionally used microscopy, mutants and other methods to understand the molecular and cellular bases of development. The amount of data acquired has rapidly increased with recent advances in instrumentation and genomics.

We have therefore embarked upon a multidisciplinary computationally based approach to integrate these data.

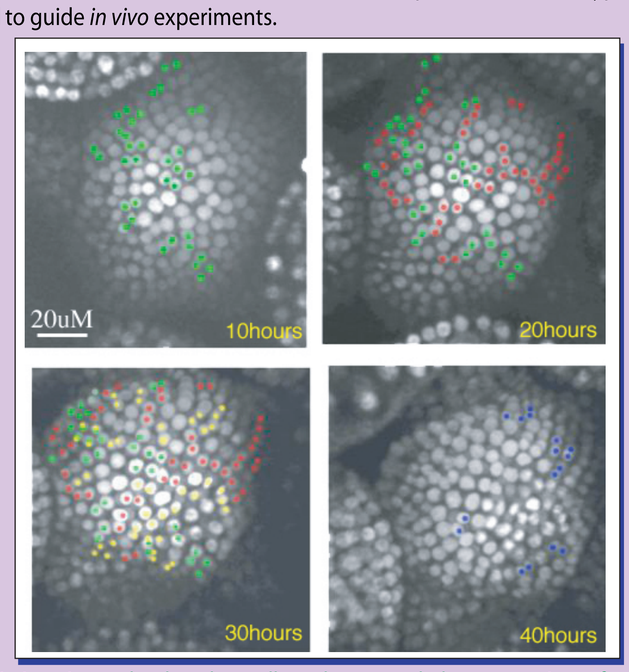
Three views of *Arabidopsis thaliana,* a plant in the mustard family, that is used extensively for biological research, illustrate some of the scales at which developmental processes are seen. The shoot apical meristem is the growing tip of a developing plant stem; the electoromicrgraph at right shows a SAM surrounded by several budding floral meristems.

Enter The Computable Plant Project

The project focuses on a quantitative description of plant development. We have chosen *Arabidopsis thaliana* as a model organism, and shoot apical meristem (SAM) development as a model process. Meristems are the inner plant tissues where regulated cell division, pattern formation and differentiation give rise to plant parts like leaves and flowers.

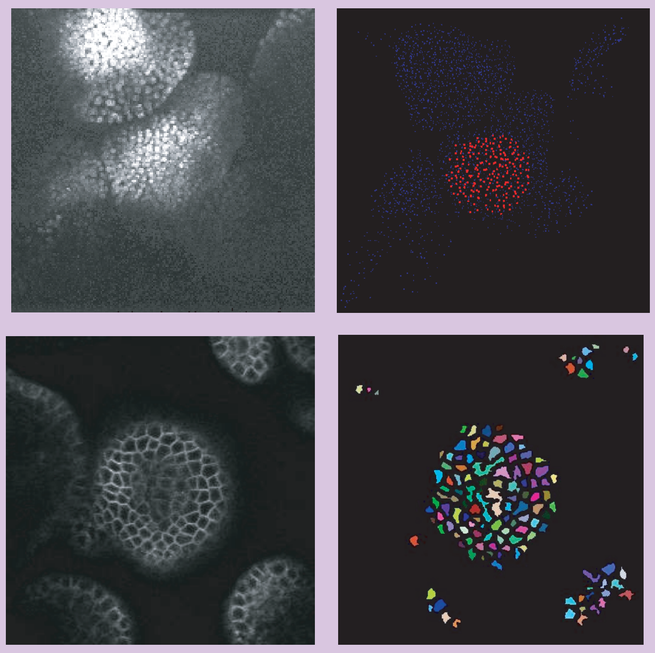
We are using green fluorescent proteins to mark specific cell types in the SAM and image their lineages through development and differentiation leading to specific arrangement of leaves and reproductive growth. Automation of image acquisition and analysis (under development) will help us generate and visualize a vast amount of data. This will be used to model cells and their patterns in the developing meristem and simulate developmental processes under different conditions. These simulations will result in predictions that will be tested experimentally using mutants, altered hormone gradients, and other manipulations.

We automatically generate specialized, efficient simulation code from models; ultimately we will link this code to suitable bioinformatic datasets through pattern recognition, machine learning, and regulatory circuit inference algorithms. Extensive visualization, image processing, and optimization software will fit predictive models to image data. These simulations will allow us to explore alternative hypotheses in silico and to guide *in vivo* experiments.



Advances in imaging technology have allowed us to track the progression of cell division in the SAM. Colored dots indicate cells that divided in each 10 hour window.

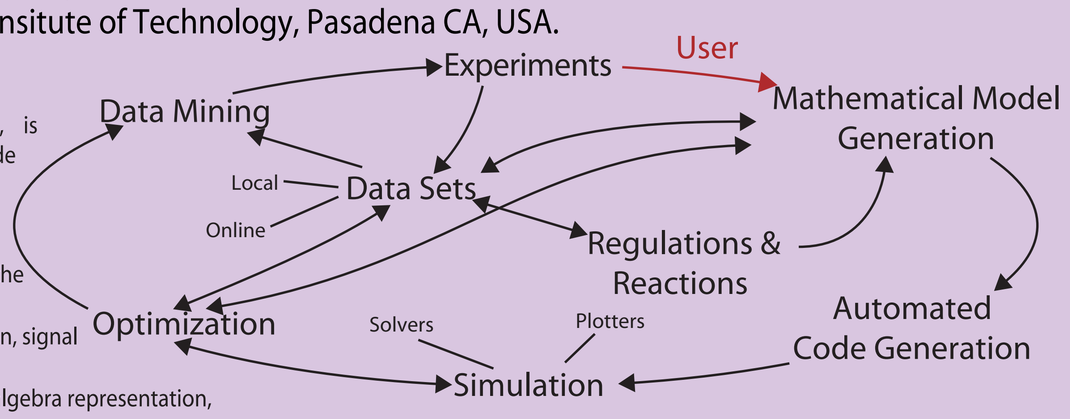
Preliminary image processing algorithms are illustrated below. The images on the left are stained to show cell nuclei (top) and membranes (bottom). Cell centers are found using a 3D lattice of intensity values; a gradient descent method is used to find cells as local intensity maxima (top right). A Laplacian filter applied to membrane image, postprocessed with Iterative Conditional Modes algorithm and Connected Components algorithm (bottom right).



Software Architecture

The architecture, aimed at production-scale model inference , is illustrated in the figure to the right. We generate simulation code from high-level models specified in biological and/or mathematical language. Other computational tools are used to analyze expression imagery and other data sources, and the simulator combined with nonlinear optimization is used to fit the models to the data. Key elements include:

- a *mathematical framework* combining transcriptional regulation, signal transduction, and dynamical mechanical models,
- a *model generation package* (Cellerator) based on a computer algebra representation, including subcellular and tissue-level representations,
- *extensions to SBML* (Systems Biology Modeling Language), an exchangeable model representation format, to include dynamic objects and relationships,
- *a C++ code generator* to translate SBML into highly efficient simulation modules,
- *a simulation engine* including standard numerical solvers and plot capability,
- *a nonlinear optimizer*, and
- ad hoc *image processing* and *data mining* tools.



Images are acquired with a *Zeiss LSM 510 meta upright* laser scanning confocal microscope.

Mathematical Framework

Let v_i be a vector of concentrations in a particular cell, and let v_i^n be the corresponding vector in a neighboring cell . Then the following genetic regulatory network (GRN) is fit to the data:

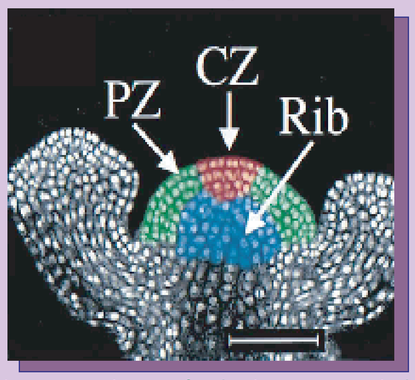
$$\frac{dv_i}{dt}=\frac{1}{\tau}[g(u_i+h_i)-\lambda_i v_i]+\frac{dv_{i,Cellerator}}{dt}$$

where

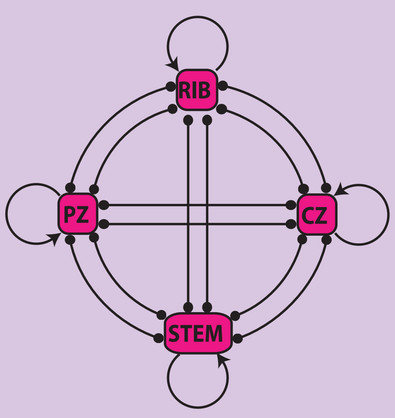
$$u_i=\sum_j T_{ij}v_j+\sum_{n,j} S_{ij}v_j^n+\sum_{n,j,k} P_{ij}Q_{jk}v_j v_k^n$$

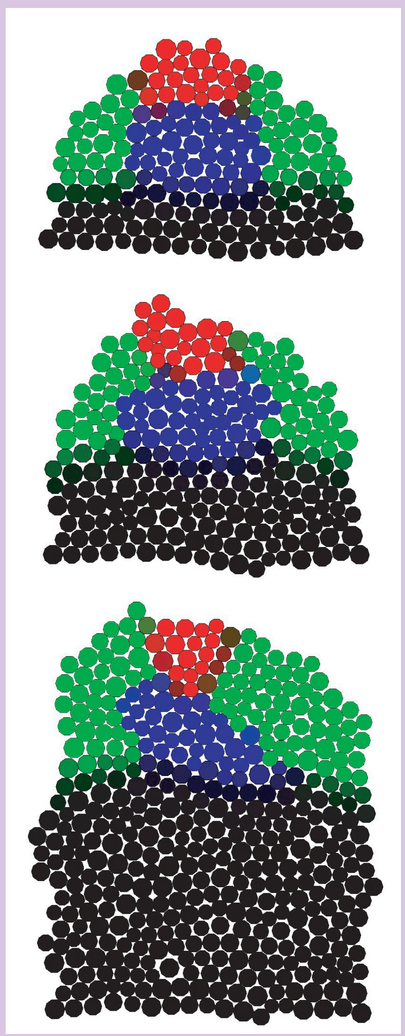
Here T, S, P and Q are connection matrices, describing, respectively the effect of (T) species j on species i; (S) species i in neigboring cell n on species j; (P) the effect of receptor j activation on species i; and (Q) the effect of ligand k, produced in neighboring cell n, on receptor j activation; $g(x)=[1+x/(1+x^2)^{1/2}]/2$; τ, λ and h are constants; and $dv_{i,Cellerator}/dt$ are differential equaitons produced by cellerator from known or postulated specific biochemical reaction networks.Spatial dynamics are modeled by a breakable nonlinear "spring" force between neighboring cells that is set to zero between non-interacting cells.

As an example, consder the naive four-protein GRN is illustrated below to the left, in which each region and the stem are identified by differential expression of region-specific proteins. Despite its simplicity, this model produces remarkably realistic-looking simulations. The resulting differentiation pattern at three different time points is shown below on the right.

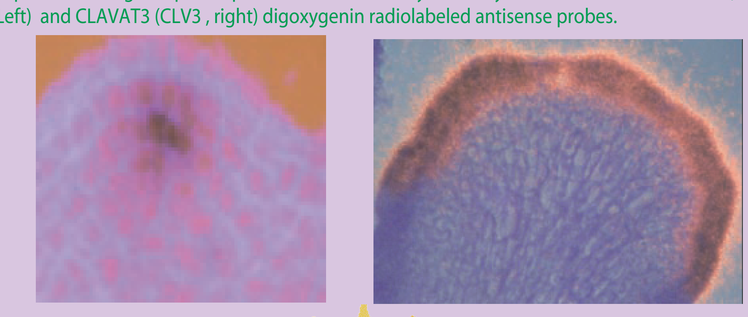


Laser scanning confocal microscope optical section of the arabidopsis SAM and adjacent floral meristems, stained with propidium iodide to show nuclei, colored to show typical SAM zonation: CZ, central zone; PZ, peripheral zone; Rib, rib meristem. Scale bar: 50 microns.

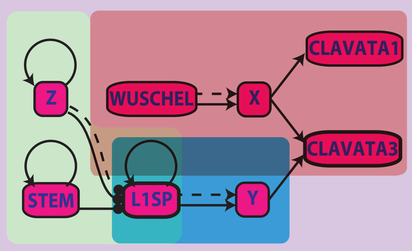


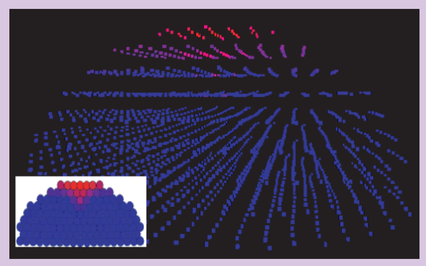


Expression of region-specific proteins is illustrated by In situ hybridization with CLAVATA1 (CLV1, Left) and CLAVATA3 (CLV3 , right) digoxygenin radiolabeled antisense probes.

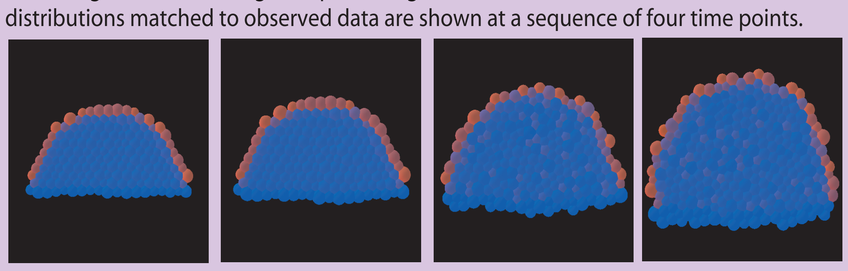


A more sophisticated model (below) is based on new gene expression imagery of key regulatory proteins in the SAM's development, such as CLAVATA3, CLAVATA1 (a receptor kinase), WUSCHEL (a homeodomain transcription factor), and a layer-1 specific protein (L1SP) that is only expressed in the surface layer. The pink network in the upper right of the model has been inferred from experiments; a hypothetical molecule (or sub-network) X is proposed to diffuse between cells. In the blue box a second diffusive signal originates from L1SP expressing cells and diffuses into the rest of the meristem; CLAVATA3 is turned on only if the sum X+Y exceeds threshold. A third hypothesis is illustrated in the green area, in which Z, which diffuses throughout the meristem, creates an L1 specific expression pattern for L1SP; STEM is a gene expressed only in the lowest layer of the meristem. Dashed lines in the figure indicate intercellular interactions, solid lines intracellular.





The initial distribution of CLAVATA3, for example, is illustrated in the figure on the right. A simulation-optimization process is used to tune parameters. Initial values are chosen randomly; this is followed by greedy local search to find energy minima. Simulated annealing and Levenberg-Marquardt algorithms are also available. Concentration distributions matched to observed data are shown at a sequence of four time points.



Code Autogeneration

Cellerator is used to design signal transduction networks (STN) based on traditional kinetic interactions (e.g., Mass Action, Michaelis-Menten, etc) as well as the GRN framework. Input to *Cellerator* is via a set of arrow-based reactions representing the network; cellerator than uses computer algebra to translate the STN into a system of differential equations. *Cellerator* then uses *MathSBML* to write the model as extended-SBML. A C++ program parses the SBML (via *libSBML)* and generates derivative functions for each chemical species in the model. The deriv functions are then linked with the simulation engine (numerical solver) and run to generate time course predictions.

```
void SBMLrule6::derivs(Compartment
&compartment,int species,
      std::vector< std::vector<double> > &y,
      std::vector< std::vector<double> > &dydt
) {
double ngbr_dot_sum1;
  int n;
  ngbr_dot_sum1 = 0.0;
  for( n=0; n<compartment.numNeighbor();n++)
    ngbr_dot_sum1+=
    dot_sum(parameter_[variableIndex_[1][0]],
    y[compartment.neighbor(n)]
    [compartment.speciesStart()],
    numVariableIndex(1));
  dydt[compartment.index()][species] +=
    (sigmoid(parameter(2)+
    dot_sum(parameter_[variableIndex_[0][0]],
    y[compartment.index()]
    [compartment.speciesStart()],
    numVariableIndex(0))+ ngbr_dot_sum1) -
    parameter(1)*
    y[compartment.index()][species]) /
    parameter(0);
}
```

Educational Outreach

We are currently developing a new set of techniques for high school, pre-service science teachers, and undergraduate students through our partnership with the Huntington Botanical Gardens in San Marino, CA. Outreach activities will culminate in a summer institute in which 30 high school students will develop a public kiosk to display a computable plant model for exhibit at the Huntington, which hosts 500,000 visitors per year. This program holds remarkable promise for linking cutting-edge knowledge and techniques with K-12 teachers' and students' understanding of plant development and integrative biology.

References

Brand U, Fletcher JC, Hobe M, Meyerowitz EM, Simon R, "Dependence of Stem Cell Fatein Arabidopsis on a Feedback Loop Regulated by CLV3 Activity", *Science* 289:617-619 (2000).

Finney A, Gor V, Bornstein B, Mjolsness ED, "Systems Biology Markup Language (SBML) Level 3 Proposal: Array Features," at http://www.sbml.org (2003).

Fletcher JC, Brand U, Running MP, Simon R, Meyerowitz, "Signaling of Cell Fate Decisionsby CLAVATA3 in Arabidopsis Shoot Meristems," *Science* 283:1911-1914 (1999).

Hucka M et al, "The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models", *Bioinformatics* 19: 524-531 (2003).

Jönsson H, Shapiro BE, Meyerowitz EM, Mjolsness ED, "Signaling in Multicellular Models of Plant Development, " In S. Kumar and P.J. Bentley (Es.) "*On Growth, Form, and Computers*", Academic Press (2003).

Meyerowitz, EM, "Genetic control of cell division patterns in developing plants," *Cell*, 88:299-308 (1997).

Mjolsness ED, Jönnson H, Shapiro BE, Meyerowitz EM, "Modeling plant development with gene regulation networks including signaling and cell division," in *Bioinformatics of Genome Regulation and Structure*, eds R. Hofestaedt and N. Kolchanov, Kluwer Publications (in press 2003).

Mjolsness ED, Sharp DH, Reinitz J, "A connectionist model of development", *J. Theor. Biol.*, 152:429-453 (1991).

Shapiro BE, Hucka M, Finney A, Doyle J, "MathSBML: A package for manipulating SBML based biological models," submitted for publication (2003).

Shapiro BE, Levchenko A, Meyerowitz EM, Wold BJ, Mjolsness ED, "Cellerator: extending a computer algebra system to include biochemical arrows for signal transduction simulations," *Bioinformatics* 18:677-678 (2003).

Shapiro BE, Mjolsness ED, "Developmental Simulations with Cellerator", paper presented at the Second International Conference on Systems Biology, Pasadena, CA, Novemeber 2001.

http://www.computableplant.org

*address all inquiries to emj@uci.edu

Sponsored by the National Science Foundation Frontiers in Integrative Biological Research Program