

CHAPTER 5

SYSTEMS BIOLOGY IN PLANT RESEARCH

Fumiaki Katagiri

INTRODUCTION

Generally speaking, systems biology research in plants has not reached an advanced stage, a situation largely due to the low level of funding for basic plant research. There are two major reasons for this. First, crop species to which plant research outcomes could be applied are numerous and diverse, and many problems that need to be solved are species-specific. Therefore, resources are spread thinly among many different plant systems. There is also strong political pressure to shift plant research funding toward crop species at the expense of model systems, even though the crop systems are generally less tractable. Since quality, quantity, and correlativity of experimental data and a repeating cycle of experimental and theoretical work are critical for success in systems biology, it would be best to focus on model plant species at this stage. Second, the agribusiness industry is showing declining interest in applications of biotechnology. Due to the anti-GMO (genetically modified organism) movement in developed countries, the industry expects difficulty in marketing biotech-based high-value products, such as functional foods, which could involve complex metabolic engineering. The declining research activities in the agribusiness industry also reduce the employment prospects for scientists trained in plant biology. Low funding levels are directly related to slower progress in new basic research areas, such as systems biology, especially when the research requires relatively large early investments and long periods of time for outcomes to be apparent. Furthermore, interdisciplinary research areas such as systems biology need to attract researchers from different disciplines to particular biological systems, which is difficult when funding levels are low and future job prospects are poor. In this context, it was disappointing to learn that a major plant systems biology collaborative program in Japan, the rice genome simulator project, was abruptly cancelled last year without clear explanation.

Another challenge, which is not specific to the plant research field, is that the number of senior principal investigators (PIs) who can properly evaluate systems biology projects is very small. Senior PIs are the ones who would organize large projects, make major changes in curricula, review manuscripts, evaluate grant proposals, and promote young researchers. Due to the enormous success of molecular genetics for many years, a large percentage of senior PIs cannot appreciate projects that are not reductionism-oriented and do not yield clear yes/no answers based on highly simplified hypotheses. The effect of this on systems biology funding is severe. This bias constitutes a drag on the adoption of systems biology by the research community. Training the next generation of researchers in systems biology is another important issue, and it is discussed in detail in Chapter 6. However, this senior PI issue is, in a sense, more important than training young researchers. Young researchers would easily be discouraged if they were not properly evaluated and encouraged by people with the power to influence their careers.

To accurately model a biological network, researchers need to know the identity and function of a sufficient number of molecular components that correspond to nodes in the network, although what fraction is sufficient is frequently debated. Molecular networks in many plants are generally less well-studied than in animals and microbes, and it is believed by many that not enough is yet known to move to modeling. Therefore, in many networks, identification of components is a major task at present. Functional genomics, i.e. discovery of gene function on a large scale, is a popular approach for this purpose. Public funding, such as *Arabidopsis* 2010 by the National Science Foundation (NSF) and programs funded by the EU, is supporting many ongoing functional genomics efforts in the U.S., Europe, and Japan. The panel visited some

of the major sites for these efforts, such as Dr. Beynon's group for the Complete *Arabidopsis* Transcriptome MicroArray (CATMA) program (<http://www.catma.org>) in Europe (Hilson et al., 2004) and Dr. Shinozaki's group at the RIKEN Genome Sciences Center in Japan (Sakurai et al., 2005). However, plant research programs that have advanced beyond the stage of component identification do exist. In the following sections, some examples of advanced plant systems biology research are reviewed. The sections are divided according to the areas of research rather than by geographic areas. This is because clear regional differences were not apparent and because it is crucial for the plant research community to coordinate and cooperate at the global level since the total funds available for research in these areas are limited.

METABOLIC NETWORKS

The structures of major metabolic networks are well established. Some groups are already studying dynamics of the networks. While metabolic networks can be illustrated as conventional metabolic maps, it is not clear which routes are really important to explain the fluxes in these networks. Dr. Fell (Oxford Brookes University) applied the elementary mode analysis (Schuster et al., 2002) to phase 3 (daylight metabolism of stored malate with no net CO_2 uptake) of the Crassulacean acid metabolism (CAM). CAM plants are typically plants of arid climates. They open stomata during the night to assimilate CO_2 into malate, and during the day they use the carbon stored in malate for Rubisco (ribulose 1,5-bisphosphate carboxylase/oxygenase)-driven carbon fixation without opening stomata. In this way they minimize loss of water. The analysis predicted six distinct pathways that could be used to accomplish CAM whereas only five pathways were known. Later, *Mesembryanthemum crystallinum* (common ice plant) was found to use the sixth pathway for CAM, validating the prediction.

Dr. Shachar-Hill's group (Michigan State University) combined the Elementary Mode Analysis with experimental measurements of mass balance, enzyme activity and stable isotope labeling in a study of carbohydrate conversion to oil through glycolysis in immature green seeds of *Brassica napus* (oilseed rape) (Schwender et al., 2004). The efficiency of conversion was higher than expected. They discovered that refixation by Rubisco of released CO_2 explains the high efficiency. This was the first description of the role of Rubisco in this context.

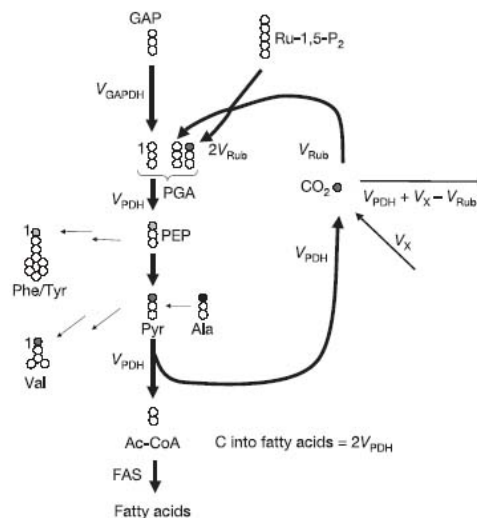


Figure 5.1. A newly discovered pathway for refixation of CO_2 released during conversion of carbohydrate to storage fat (Schwender et al., 2004).

Major efforts to collect correlated messenger ribonucleic acid (mRNA) profiles, metabolite profiles, and other phenotypes from many genetically perturbed *Arabidopsis* plants are under way at the Max Planck Institute for Molecular Plant Physiology (<http://www.mpimp-golm.mpg.de/>) in Golm, Germany and by members of a collaboration led by the Kazusa DNA Research Institute (KDRI) (<http://www.kazusa.or.jp/eng/index.html>) in Chiba, Japan. The research in Germany is led by Drs. Willmitzer and Stitt and focuses on primary metabolism. In addition to use of the Affymetrix GeneChip array, they collect mRNA profiles of most transcription factor genes using a high-throughput real-time reverse transcription-polymerase chain

reaction (RT-PCR) (Czechowski et al., 2004). They also measure protein levels of major enzymes and use other specialized profiling methods. The research in Japan is led by Drs. Shibata (KDRI) and Saito (Chiba University) and focuses on secondary metabolism. They use sophisticated equipment (gas chromatography time-of-flight mass spectrometry—GC-TOF-MS, Liquid chromatography photo-diode-array-detection mass spectrometry—LC-PDA-MS, Capillary electrophoresis mass spectrometry—CE-MS, liquid chromatography fourier transformation mass spectrometry—LC-FT-MS, and liquid chromatography time-of-flight mass spectrometry—LC-TOF-MS) for metabolomic measurements in addition to mRNA profiling with the Agilent microarray (Hirai et al., 2004). Accurate mass information obtained by Fourier transformation mass spectrometry (FT-MS) is a great help in identification of metabolites. The Japanese group uses a well-established suspension culture cell line to ease the issue of establishing consistent growth conditions. Both German and Japanese groups have invested in bioinformatic tools, including viewers that integrate expression and metabolite information along metabolic maps (Thimm et al., 2004).

To understand the dynamics of metabolic networks, metabolic flux measurements are important. Dr. Shanks (Iowa State University) has developed a computer-assisted method to estimate metabolic fluxes of several pathways using biosynthetically directed fractional ^{13}C -labeling and two-dimensional [^{13}C , ^1H] nuclear magnetic resonance (NMR) (Sriram et al., 2004). This method could be applicable for many known metabolic networks.

REGULATORY NETWORKS IN DEVELOPMENTAL PROCESSES

Thanks to extensive genetic analysis, many important components of regulatory networks that control several developmental processes are known. If key information can be collected about the important components, sufficient information may be obtained to model such regulatory networks at a practical level. Some developmental processes are explained by transcription-regulatory cascades. In such cases, the activity of each gene is well correlated with the mRNA level of the gene. For plants to be successful in evolutionary terms, it is crucial to have flowers open and seeds set at the right time in the growing season. Many factors, such as photoperiod (day length), the plant hormones called gibberellins, and experience of cold weather, can affect flowering time. Dr. Welch (Kansas State University) modeled the *Arabidopsis* flowering time control system using mRNA levels of the important molecular components of the process (Welch et al., 2003). Circadian rhythms are important in developmental controls as well as in physiological controls. For example, plants measure photoperiod by comparing it with their own circadian rhythms. Dr. Millar and his collaborators (University of Warwick, U.K.) modeled the *Arabidopsis* circadian clock (Locke et al., 2005). Although circadian clocks exist in diverse organisms, such as cyanobacteria, fungi, animals, and plants, the molecular machinery in different organisms seems to be quite diverse. In the case of cyanobacteria, an *in vitro* system composed of three proteins that can generate a 24-hour period has recently been reconstituted (Nakajima et al., 2005). It is of interest to see whether the network structures and the system control are also diverse, even though the resulting clocks are all robust in the maintaining period.

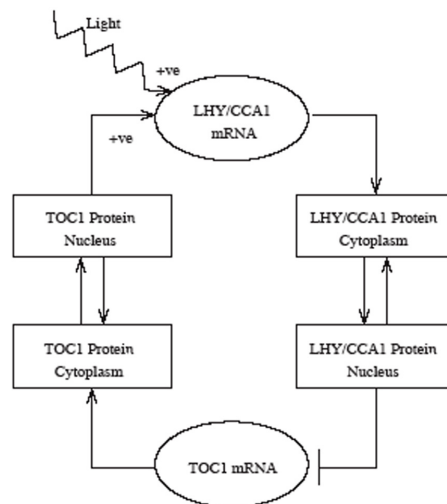


Figure 5.2. Model for the central feedback in the *Arabidopsis* clock (Locke et al., 2005).

An ambitious project called The Computable Plant has been initiated by a consortium led by Dr. Meyerowitz (Cal Tech) (<http://www.computableplant.org/>). This project aims to model development of the shoot apical meristem (SAM) in *Arabidopsis*. Meristems are the inner plant tissues, where regulated cell division, pattern formation, and differentiation give rise to plant parts like leaves and flowers. The project includes modeling efforts as well as experimental efforts to monitor the cell lineage of specific cell types in real time (Reddy et al., 2004). This process will ultimately be automated.

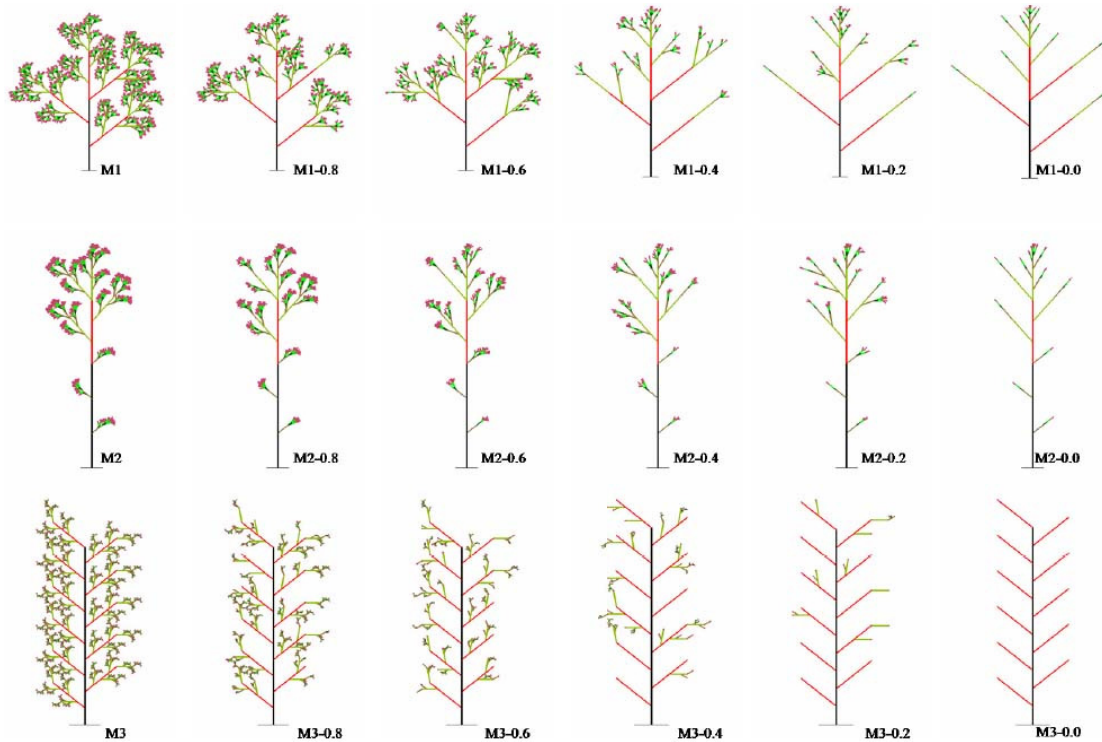


Figure 5.3. Variation of theoretically generated plant architectures (Ferraro et al., 2005).

Morphogenesis has been a central interest in developmental biology. Some modeling work has been done to explain plant morphogenesis at various scales. Dr. Coen's group (John Innes Center, Norwich, U.K.) measured cell growth in the snapdragon petal using clonal analysis (Rolland-Lagan et al., 2003). Then they modeled the complex shape of the petal based on the cell growth pattern. Simulation of the model found that the direction of growth that was maintained parallel to the proximodistal axis of the flower was crucial for formation of the asymmetric petal shape, but other factors, such as changes in cell shape, were not. Dr. Prusinkiewicz's work (University of Calgary, Canada) involves mathematical modeling of the plant architecture using L-systems with a small number of parameters (Prusinkiewicz, 2004). For example, diverse inflorescence shape patterns seen among various dicot plant species could be explained by two parameters that represent characteristics of a putative factor. This approach can be considered as reverse engineering of the plant architecture. It will be interesting to see if an actual molecular counterpart of the putative factor exists.

CONCLUDING REMARKS

Progress of systems biology research in the plant field has been slow. However, several advanced studies shed light on unique aspects of plants. There are no clear national differences in plant research in systems biology. Several actions are needed to promote systems biology of plants.

To make the most out of limited funding:

- Focus on model plant species. It is clear that the majority of advanced studies have been performed with model plant species, such as *Arabidopsis*.
- Cooperate rather than compete at the global level.

To compensate for the PI population bias against promotion of systems biology in the research community:

- Implement a sustaining, targeted funding program in plant systems biology.

To raise the next generation of researchers:

- Train biology-major students in quantitative science.
- Recruit students oriented to mathematics, engineering, physics, and chemistry into plant biology.

REFERENCES

- Czechowski, T., R. P. Bari, M. Stitt, W. R. Scheible, and M. K. Udvardi. 2004. Real-time RT-PCR profiling of over 1400 *Arabidopsis* transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes. *Plant J* 38: 366–379.
- Ferraro, P. C. Godin, and P. Prusinkiewicz. 2005. Toward a quantification of self-similarity in plants. *Fractals* 13(2): 91–109.
- Hilson, P., J. Allemeersch, T. Altmann, S. Aubourg, A. Avon, J. Beynon, R. P. Bhalerao, F. Bitton, M. Caboche, B. Cannoot, *et al.* 2004. Versatile gene-specific sequence tags for *Arabidopsis* functional genomics: transcript profiling and reverse genetics applications. *Genome Res* 14: 2176–2189.
- Hirai, M. Y., M. Yano, D. B. Goodenowe, S. Kanaya, T. Kimura, M. Awazuhara, M. Arita, T. Fujiwara, and K. Saito. 2004. Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 101: 10205–10210.
- Locke, J. C., A. J. Millar, and M. S. Turner. 2005. Modeling genetic networks with noisy and varied experimental data: the circadian clock in *Arabidopsis thaliana*. *J Theor Biol* 234: 383–393.
- Nakajima, M., K. Imai, H. Ito, T. Nishiwaki, Y. Murayama, H. Iwasaki, T. Oyama, and T. Kondo. 2005. Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation *in vitro*. *Science* 308: 414–415.
- Prusinkiewicz, P. 2004. Modeling plant growth and development. *Curr Opin Plant Biol* 7: 79–83.
- Reddy, G. V., M. G. Heisler, D. W. Ehrhardt and E. M. Meyerowitz. 2004. Real-time lineage analysis reveals oriented cell divisions associated with morphogenesis at the shoot apex of *Arabidopsis thaliana*. *Development* 131: 4225–4237.
- Rolland-Lagan, A. G., J. A. Bangham, and E. Coen. 2003. Growth dynamics underlying petal shape and asymmetry. *Nature* 422: 161–163.
- Sakurai, T., Satou, M., Akiyama, K., Iida, K., Seki, M., Kuromori, T., Ito, T., Konagaya, A., Toyoda, T., and Shinozaki, K. 2005. RARGE: a large-scale database of RIKEN *Arabidopsis* resources ranging from transcriptome to phenome. *Nucleic Acids Res* 33: D647–650.
- Schuster, S., C. Hilgetag, J. H. Woods, and D. A. Fell. 2002. Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism. *J Math Biol* 45: 153–181.
- Schwender, J., F. Goffman, J. B. Ohlrogge, and Y. Shachar-Hill. 2004. Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. *Nature* 432: 779–782.
- Sriram, G., D. B. Fulton, V. V. Iyer, J. M. Peterson, R. Zhou, M. E. Westgate, M. H. Spalding, J. V. and Shanks. 2004. Quantification of compartmented metabolic fluxes in developing soybean embryos by employing biosynthetically directed fractional (¹³C) labeling, two-dimensional [(¹³C), (¹H)] nuclear magnetic resonance, and comprehensive isotopomer balancing. *Plant Physiol* 136: 3043–3057.
- Thimm, O., O. Blasing, Y. Gibon, A. Nagel, S. Meyer, P. Kruger, J. Selbig, L. A. Muller, S. Y. Rhee, and M. Stitt. 2004. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J* 37: 914–939.
- Welch, S. M., J. L. Roe, and Z. Dong. 2003. A genetic neural network model of flowering time control in *Arabidopsis thaliana*. *Agron J* 95: 71–81.