MASC Systems Biology- 2007 Subcommittee Report

Prepared by Philip Benfey (Chair, philip.benfey@duke.edu)

A primary goal of the subcommittee is to further the use of Systems Biology among *Arabidopsis* researchers to elucidate the structure, dynamics, and organizational principles of the regulatory and metabolic networks that support living cells. Although there is no widely agreed upon definition of Systems Biology, most work in this field can be characterized as an effort to identify the molecular interactions that underlie cellular function and to integrate them in a coherent model. What distinguishes Systems Biology from fields like Molecular Genetics, which have the same ultimate goal, is the approaches that are used and the perspective taken. Systems Biology uses high-dimensional data that are usually acquired through high-throughput approaches such as microarrays, together with more focused high-resolution and high-content data, particularly time series. The elements within the large datasets are seen as a parts list and an initial goal is to find the connections among the parts. This leads to the identification of modules that frequently have properties that would be difficult to predict from the collection of parts. These are referred to as emergent properties. When connections are found between modules it can lead to the identification of networks. Once networks are identified, Systems Biologists seek to determine how the networks function. This is done by perturbing them with external stimuli or by removing or increasing the activity of one of the parts. Perturbation of the network can lead to an understanding of the dynamics "over" the network - for example how information flows through a signaling network.

Systems Biology also seeks to understand the dynamics of networks in the sense of how networks become modified as cells alter their states. A salient feature of Systems Biology is the tight integration of quantitative reasoning with biological experimentation. Modeling approaches are drawn from statistics, dynamic systems, linear algebra and topology, to name a few. The reliance on modeling approaches is driven by the high dimensionality of the data and the complexity of the networks that govern cellular function. Most Systems Biology today is carried out by interdisciplinary teams of scientists with backgrounds in Computer Science, Mathematics, Physics, Engineering, Statistics, and Biology.

- Given the shortage of theoreticians currently working on questions in plant science, a short-term goal of the subcommittee is to bring together additional researchers from primarily quantitative backgrounds with experimentalists working on *Arabidopsis* to address Systems Biology questions.
- A second goal is to promote public data repositories and benchmark data sets, which allow theoretical work with wider scope and on larger scales.
- A medium-term goal is to facilitate the development of international consortia in *Arabidopsis* systems biology by promoting communication among emerging national Centers and research programs.
- A long-term goal will be to promote the application of Systems Biology approaches to a wide range of plant science problems.

Subcommittee plans
- The subcommittee will plan a workshop to coincide with the next International *Arabidopsis* Conference (Montreal, 2008).
- It will also work with the other MASC subcommittees that interface closely with work in Systems Biology, and build links to the relevant community organizations in Systems Biology, theoretical and computational biology. One aim here is to recruit theoretical researchers by publicizing some of the contemporary questions in *Arabidopsis* research, and making the relevant *Arabidopsis* data very easily accessible.

- Another aim is to benefit from existing models and modeling expertise in other organisms and fields. The subcommittee also will seek input from the rest of the plant community at the next Plant Biology Conference to identify what their needs are, and how *Arabidopsis*, as a model system, can best assist systems biology efforts in other species and benefit from experience of modeling in other areas of plant science. This will facilitate the investigation of agronomically-important and medicinal species. It is anticipated this would the first of many such discussions.

Recommendations of the Subcommittee

The amount of *Arabidopsis* experimental data being generated by high throughput methods is growing at a tremendous pace. Indeed, data integration is one of the biggest challenges that we face today in systems biology research. Achieving effective data integration for data analysis and interpretation is crucial to advance plant systems biology.

- This subcommittee will work closely with the MASC Bioinformatics subcommittee to promote the creation of software platforms and data sharing standards to allow systems biology research to be carried out effectively by the *Arabidopsis* community. A key set of challenges center around the creation of experiments and experimental data in a form that is maximally accessible to bioinformatics analysis, biologist-mining and mathematical modeling, and in parallel, the promotion of model-building as a central component of *Arabidopsis* research.
- Comparisons of analysis of the same benchmark data sets by different groups could facilitate our understanding of the strengths and weaknesses of different computational methods in different contexts. For example, one idea, is a systems biology competition, where an *Arabidopsis* dataset is given out with "known information" intentionally missing, and a prize is given to the student who can best predict the missing data.
- Integration of systems biology approaches into teaching, by updating the mathematics content of undergraduate courses in biology (to include modeling, not just statistics) and by updating the plant science examples used in mathematics courses (to include functional genomics and development, not just agronomics).