REVIEW OF BREAK OUT SESSION ON GRAND CHALLENGES IN PLANT BIOLOGY

Three separate groups were formed to discuss potential grand challenge questions that would relate to broad swaths of plant biology/plant science research. Interestingly, in all three independent sessions, a good portion of the 2 allotted hours was spent discussing just what constituted a grand challenge in plant biology. In addition, significant time was spent in each session discussing what were viewed by most participants to be very important infrastructure challenges/opportunities. Finally, each group came up with a list of potential grand challenge questions. Some of these were almost identical across groups, some were unique to each group. These discussions and ideas were then presented to the combined group. After these presentations, much of the discussion time in the combined session centered around questions related to the way that budget funds from the iPC would be allocated and regarding grand challenge workshop application instructions, which were answered briefly in this session and in much more detail the following morning. In addition, the iPC team gave a clarification of the role of the iPC relative to individual biologist or CS grants. The general message from the iPC team was that “iPC will free up resources for you to use for science, not compete with you”. By the end of the Wednesday morning session, most meeting participants understood this concept and began to make plans to develop grand challenge teams.

Summary of Definition of a Grand Challenge:

What is the nature of a grand challenge that is both compelling and tractable?

One breakout group suggested at first that to be truly “grand” the question should attack problems at multiple scales and be potentially intractable, at least in the immediate future. It was generally agreed that this was more appropriately called a “vision”, as Richard Jefferson pointed out in his plenary talk following the combined session discussion.

So, what constitutes a grand challenge question?

Other communities have wrestled with this question and provided some direction, which one audience member provided to one of the breakout groups:

A grand challenge question should:

- Have a clear compelling solution (even if this is not obvious immediately)
- Have a well defined measure of success in solving it
- Give some predictive power in the end
- Be decomposable & diagnostic
- Be visionary but not unrealistic
- Be compelling to general public
- Provide motivation to many researchers

What types of impact should a question have to be considered "grand"?

> **Societal**: Agriculture, Bioenergy, Crops, Fuel, Climate Change resistance, Nutritional Quality
>
> **Scientific Impact**: Tractable system for modeling and systems biology and synthetic biology. Tools and concepts will be useful for all aspects of biology. Impact on other related areas of plant biology: growth, development and ecology (e.g from seeds to senescence).
>
> **Tractable**: Metabolic pathways provide a framework for building on existing knowledge and relate to physiology and plant performance. New genome scale data (e.g.. metabolomics, proteomics, interactomics) and can be interpreted and integrated into this framework to build and validate testable networks and models.
>
> **Infrastructure**: Requires tool development whose impact is cross cutting and universal.

An important concern was voiced by many that: work on "model systems" should not completely "shade out" today's non-model systems, especially given the fact that a deluge of genome sequences will be made available during iPlant's tenure, completely changing the definition of "model system".

***POTENTIAL GRAND CHALLENGES**:*

<u>1. INTEGRATED MODELING</u>

iPlant: " Empowering new plant biology". An important new aspect of biology is modeling. Models function to integrate data, generate and test hypothesis, and enhance communication within the scientific community. This is the new aspect that iPlant is empowering. Models of plant biology at all levels are being developed. This point came up in all sessions and was the foundation for many of the grand challenge (GC) questions that were proposed.

Applications: Optimize biomass production, N-use efficiency, water use, mineral use, CO2 fixation, solar energy, stress response, and more.

GC: Build a unicellular model for plant metabolism. (Distinguished for cell types).
GC: Build a Virtual Plant Cell.
Transcription, Protein interactions, Metabolites
Cellular and subcellular localization
GC: Model the biogenesis of an orgenelle.
GC: Build a Virtual plant. Use to develop predictive and testable models.
Need: infrastructure to work with different inputs, take two species and identify differences between 2 species based on a property. Predict from an input genome how the differential of those genomes leads to different in output.
Can one build a computer model that incorporates all –omics and other databases and, when run, shows how a single cell develops into a whole plant?
GC: Integrate metabolism with development and natural variation.
GC: Systems Biology & Synthetic biology to address efficiency in plant growth:
Increase Use efficiency of water, C02, light, nitrogen & minerals
GC: Perform in silico reconstruction of biochemical signaling and genetic networks of key plant species. Understanding these networks, network topology and robustness. Similar to the work that Bernard Paulsson is doing in yeast and E coli. Much harder in plants because of multicellularity. We should embrace this challenge.
GC: Make sense of intraspecific variation and microevolution. Can we really get the two to completely explain each other? If not, what else is missing?

***POTENTIAL GRAND CHALLENGES**:*

2. FUNDAMENTAL BIOLOGICAL QUESTIONS:

GC: How does plant gene regulation work? Identify all TFs, ALL binding sites, good set of expression data across all cell types. Chromatin and microRNA modifications.

GC: What are the underlying networks that support plant growth and development? (metabolism, gene regulation, etc.)

GC: How do plant hormones regulate growth and development? How did these mechanisms evolve? Identify in complete detail how all plant hormones are produced and how they act, including identification of all factors that stimulate or modulate their production.

GC: Subset of previous question: **How has the auxin signaling system (transporters, receptors, AUX/IAA genes, etc) evolved?** This nicely connects molecular to phyllogenetic interests, in a way that is perhaps doable on a 5-10 yr time scale.

GC: How do plant cells divide? This needs to be addressed at ALL levels of biology, from the gene expression level to the physical constraint level, etc.

GC: Just what is the epigenome? Is there such a thing as a single epigenome? Can it really be measured?

GC: Can we identify all metabolic steps that we are missing from present knowledge of existing pathways?

GC: How diverse is natural product chemistry across the plant kingdom? This should be understood from one or more model systems (such as Arabidopsis thaliana and close relatives, identifying the diversity across accessions) to across the diversity of the plant kingdom.

GC: C4 photosynthesis evolved "52 times": What is the problem with RuBisCo?

GC: How is plant diversity impacted by global climate change? This is somewhat related to the previous question about photosynthesis and RuBisCo.

GC: How can we effectively link metabolic pathways with structural biology and evolutionary information to generate a broader understanding of how metabolism really functions in plants on multiple time scales?

GC: How can plants reach their productivity potential sustainably? (agriculture-oriented, social benefit)

GC: How do biotic and abiotic factors determine the mix of organisms and total biomass of an ecosystem? (ecology, pathogenesis, metals tolerance, etc. etc. etc. This overlaps with the section above regarding modeling.)

GC: Can one systematically connect results from model organisms to non-model organisms? (This overlaps with the section above regarding modeling.

INFRASTRUCTURE & DATASETS NEEDED IN GENERAL

Gold standard data: Curate a set of time-series data for RNA, protein level at organ, cell-specific and sub cellular level. Used for validation, develop quantiative networks.

Published data extraction: Develop tools to extract all data (not mere text mining) on a GC topic.
Develop tools for ortholog identification: Enable cross species network analysis.

Data Integration & Visualization : Need user friendly methods to overlay different types of data onto metabolic pathways and other network data.

Formalize metadata and data acquisition tools. Establish standards, software tools to facilitate use of standards to allow data integration and interoperability. Built on formalized ontologies and controlled vocabularies.

Education & Outreach

Train new generation of scientists: Biology + Computer Science.
Senators understand food, fuel and fiber needs.

Important points to consider when developing infrastructure and tools:

- Standards for data representation should be developed
- Integration of various "omics" data visually via a graphical interface that is accessible to all would be very useful
- Good quality descriptions of metadata that allows links to other relevant data must be required
- Applications should be able to be used by web-savvy students and researchers, not just the experts in the narrow field
- The infrastructure should capture knowledge (analysis and interpretation) along with raw data.
- Planning around use by amateurs should be a driving force
    - All researchers will be amateurs at multiple points on the system. Nevertheless, all will want to be able to integrate their data with data from fields in which they are not experts.
- Educators, students, hobbyists and others are likely to be users and contributors of data.
- Tools should be accessible, and work as advertised.

INFRASTRUCTURE TOOLS THAT iPC COULD WORK TO DEVELOP IMMEDIATELY

The discussions invariably led to many audience participants suggesting (sometimes repeated in a slightly different way from what someone else had mentioned earlier) a number of what were widely viewed as important deficiencies in currently available cyberinfrastructure tools. It was also widely viewed that it would be very appropriate and definitely desirable for the iPlant team to begin addressing some of these immediately if possible.

Suggested infrastructure tools or efforts to produce tools:

1. Produce user-friendly tools to allow for production and analysis of computer models of plant development, such that users could change parameters and perhaps gene activity to see the effects on the model organism. These tools would be flexible enough to be useful to a broad range of plant biologists.

2. Many participants pointed out that the available plant databases are scattered across the web and difficult to access collectively. In addition, there are MANY new databases under construction currently. Produce a portal (accessible through the iPlant web site) that guides users through these data bases and also updates the community on upcoming data resources. This could be a library or catalog of existing and upcoming data bases, or it could be a superdatabase that allows through one database interface access to all other databases.

3. Produce and/or make available free software tools for image analysis of images and movies that would allow for the identification of nuclei or cell walls or other organelles automatically. Perhaps these could be built as plug-ins to the widely-used ImageJ software, originally developed by the NIH.

4. Produce a simple to use interface (that doesn't require the individual to program it each time themselves for their data) that allows for the viewing and interpretation of networks of all sorts, such as gene interaction, protein:protein interaction, metabolic, transcription factor, etc., and of data types of all sorts, such as all "omics" datatypes

5. Build a phylogenetic coordinate system, broadly construed, onto which other data types could mapped or mashed. The general view was that such a coordinate system needs to be able to accommodate change, so that when the phylogenies change, all of the data mapped onto them can change commensurately ("otherwise you have a train wreck").

This summary was compiled by David Gang from notes contributed by the Breakout Session summary teams. Session 1: David Gang, Gloria Coruzzi, David Salt, Klaas Van Wijk, & Pascal Braun; Session 2: Edgar Spalding; Session 3: Steven Slater. As well as from notes on the summary session contributed by Ann Stapleton.