The iPlant Collaborative – Grand Challenge Workshops 2008

-distributed by Ruth Bastow (ruth@arabidopsis.info)

Background

The iPlant Collaborative (http://iplantcollaborative.org/) has been established by a grant ($50 million for the first 5-year phase) from the US National Science Foundation to the University of Arizona. It is intended to address the need to capitalize on the large volumes of data emerging from plant science research. The aim is to develop a "cyberinfrastructure" of computer hardware, software and skilled staff, to enable "Grand Challenges" in plant sciences to be addressed.

The iPlant Collaborative program will run for an initial 5-year period (2008-2013), with funding for a further 5 years a possibility.

iPlant uses the concept of "Grand Challenge" topics. The plant science community was invited to propose "Grand Challenges", a subset of which (those considered of sufficient importance, and amenable to being addressed) would be supported by the iPlant team.

Although workshops were not an absolute requirement for proposal of a "Grand Challenge", the community was encouraged to organize discussion workshops, which would also be attended by iPlant personnel.

In 2008, nine workshop proposals were received, and five of these proposals were selected as workshop topics. A summary and feedback from UK PI that attended these five workshops, and the subsequent proposals that resulted are contained in this document.

Timescale

The overall timescale for identifying and beginning to address "Grand Challenges" is outlined below. Project proposals from the first set of workshop are to be submitted by February 2009, with decisions on priorities being made in March and the first projects being initiated in April or soon after. Another set of workshops are expected to be held in 2009.

Phase I: Community Selects Board of Directors (Spring 2008)
Phase II: Kick-off Conference (2008)
Phase III: Grand Challenge Workshops (GCW) (September-December 2008)
Phase IV: Grand Challenge teams form
                    GC Projects proposed (by ~ 1 February 2009)
                    Projects prioritized by Board of Directors (by end March 2009)
Phase V: iPC & GC teams develop discovery environments in collaboration
                    (starting ~ April 2009)
Phase VI: Additional GCW and proposals considered (Summer + Autumn 2009)

Aims and scope of iPlant

The role of iPlant is to provide the cyberinfrastructure to facilitate addressing the Grand Challenges identified as priorities. It will *not* fund any "wet biology"; if unavailability of data is a barrier to meeting a Grand Challenge, a separate funding bid (to NSF and/or other bodies, including non-US organizations) must be made to enable data acquisition. Each Grand Challenge project to be funded must meet the following criteria set out by NSF: it must

> address a problem or problems that are not easily solved without the Plant Science Cyberinfrastructure Collaborative (iPlant for short)
> need a sophisticated cyberinfrastructure which is not currently available
> use data which are (largely) already available
> be on an exceptional scale and have exceptional scope and impact

Once a Grant Challenge topic has been approved, the team proposing it, and others from the relevant community, will work with core iPlant staff to develop a Discovery Environment for that Grand Challenge. Discovery Environments are defined as shared, open source, web-based software platforms that address the grand challenge questions in plant sciences by allowing community members to integrate, visualize, analyze and annotate large-scale datasets and associated computational models.

iPlant organisation and key players

The iPlant project personnel are listed on http://iplantcollaborative.org/about-ipc/principal-personnel. Key players include

- Richard Jorgensen, Collaborative Director, Plant Sciences/BIO5, University of Arizona
- Steve Goff, Director of Community Interactions, University of Arizona
- Rick Blevins, Director of Cyberinfrastructure Development, University of Arizona
- Martha Narro, Director of Education, Outreach and Training, University of Arizona

Grand Challenge Workshops (GCW) 2008

1. Impacts of Climate Change and Mechanistic Basis of Plant Adaptation
September 30th – October 3rd 2008

Impact of Climate Change Workshop Summary
Workshop Lead Organizer -: Ruth Grene (Virginia Tech)
Anticipated global climate change will require directed adaptations of crop species on an unprecedented magnitude in order to sustain agricultural production. This GCW seeks to design a proposal that would dramatically improve quantitative prediction of phenotypes by facilitating the characterization of "molecular pathways" of ecologically and agriculturally important traits affected by climate change. If successful the project would also create computational tools that enable the integration of phenotypes across diverse species. A key focus of this project will be the development of infrastructure tools for the global integration of all available high throughput data for effects of abiotic stress factors, associated with climate change, that shape crop and non-crop plant performance in natural environments.

Mechanistic Basis of Plant Adaptation Workshop Summary
Workshop Lead Organizer :- David Salt (Purdue University)
One of the most pressing grand challenges in biology is to understand how plants adapt to complex and unpredictable biotic and abiotic environments.
This GCW brought together varied participants to discuss and investigate if an integrated cyberinfrastructure would improve the ability of researchers to efficiently determine the fundamental genetic, biochemical and physiological mechanisms underlying plant adaption by improving cooperation and integration across laboratories, field sites and species

Feedback from UK Participant – Dr. Sascha Ott (University of Warwick)

The first two iPlant workshops on "Plant Responses to Climate Change" and "Mechanistic Adaptation of Plants" were held concurrently. These started out separately, but were joined together later as many of the challenges involved in each were similar. Participants from different backgrounds (experimental and computational at different levels - molecular to ecological) faced the challenges of talking a common language and of identifying a "grand challenge" that is grand enough to meet iPlant's aspirations, but can be solved by a software development project that is relatively short-term. Time for free discussions was limited as a large number of short talks filled the schedule. Lessons learned from these workshops were useful for the organization of subsequent workshops.

Conclusions
Overall the workshop did not fully discover what biologists needed in this area. Lack of discussion time and large number of participants meant that there were too many different ideas discussed during the workshop to draw clear conclusions.

2. Developing Common Models for Molecular Mechanisms
7-10th November 2008

Workshop Summary
Workshop Lead Organizer – Katherine Denby (University of Warwick)

An increasing world population and predicted changes in the global climate demand improvements in plant stress tolerance and crop productivity. These traits are determined by a complex web of interactions between plant genotype and biotic and abiotic factors in the local environment. The ability to understand, predict, and then manipulate plant responses to the environment will be critical for increased food security in the 21st century.
This GCW proposed to identify the barriers to the generation of user–friendly models and to develop a plan to overcome these challenges. Using the integration of circadian, diurnal, and environmental response models as a biological framework, the group planed to generate models that are broadly useful, easy to manipulate, and yield predictions that can be tested in the lab and the field.

Feedback from UK Participants: Jim Beynon, (University of Warwick), Vicky Buchanan-Wollaston (University of Warwick) and Andrew Millar (University of Edinburgh).

The Mechanistic Modeling workshop was organized by Stacey Harmer, with about 45 participants from a broad and international range of backgrounds. A few participants from the earlier Salt and Grene workshops also attended. The iPlant initiative was outlined but questions remained regarding delivery mechanisms, which apparently had not been resolved. Lively discussions covered the desirable bioinformatic and modeling infrastructures to integrate emerging data, and to support multi-scale modeling from intracellular pathways (metabolic or gene regulatory) to tissues, whole plants, fields and ecosystems. The level of discussion was quite general, reflecting a dynamic that was led by biologists with a leavening of modeling, systems biology and computer science researchers, little input from resource providers and (by design) almost no specific input from iPlant or NSF staff. Three Grand Challenge proposals were mooted at the end of workshop, focusing on Flowering Time (led by Steve Welch, Kansas State), Building C4 Rice (led by Tom Brutnell, Cornell and Xinguang Zhu, University of Illinois), and Plant Adaptation to Environmental Stress (led by Ruth Grene, Virginia Tech). Each has been followed up with a proposal.

Many of the proposed requirements for cyberinfrastructure appeared to be common between the three proposals. A general list of cyberinfrastructure requests from this workshop included:

1. Model sharing, model repository, graphical interface, modeling tools.
2. Data compilation, data retrieval, nomenclature/ontologies, need easy access to all available data.
3. Software that is easy to use – Graphical User Interface at front end - iPlant to host and manage.
4. Visualization tools at different organizational and temporal levels.
5. Collection of new data should follow a process with iPlant cyberinfrastructure in mind.

Conclusions
Workshop participants had some difficulty in evaluating which areas of cyberinfrastructure the iPlant P.I. and Co-I.'s will best be able to deliver. Specialized expertise from non-iPlant groups will almost certainly be required for some projects, but it was unclear how such groups will be engaged. iPlant-funded personnel have been working in the Meyerowitz group, for example; iPlant and NSF staff indicated that further such arrangements would be considered as one of several options, but it was unclear on what scale.

3. Tree of Life
19th -23rd November 2008

<u>Workshop Summary</u>
Workshop Lead Organizer - Michael Donoghue (Yale University)

Ever since Darwin heralded "the great Tree of Life," biologists have attempted to infer the precise order and timing of the branching events that link all species that have ever existed. Owing to the shear magnitude of this problem, and to its fundamental importance, reconstructing the tree of life is one of the most profound scientific endeavors ever undertaken. Tackling this grand challenge will require the integration of data from multiple sources, including the morphology of living and extinct organisms, and, increasingly, vast quantities of genomic data. It will also require a quantum leap in the capabilities of algorithms to infer phylogenies from these data at the scale imagined.

<u>Feedback from UK Participant</u> - Helen Ougham (Aberystwyth University)
The 45 participants that attended this workshop reached the consensus that the community needs to assemble an accurate and comprehensive tree of (plant) life in all branches. This outcome was very much driven by the priorities of the biologists. Such a project would :-
- Help further our current understanding and conserving biodiversity.
- Aid the identification of newly-discovered species.
- Assist in predicting the consequences of climate change for plant survival.
- Facilitate the analysis of the evolution of diversity at the morphological and functional level.
- Help to determine the basis of speciation.
- Further our understanding the role of polyploidisation events in plant evolution
- Promote education and outreach in this area, including countering the creationist standpoint.
The workshop identified the following main priorities for a cyberinfrastructure project in this area
- Tools for assembling "the tree" (or "summary tree") of all plants
- Development of new or improved algorithms for supertrees, parallelization, reticulation
- Access to high-performance computing for handling very large datasets

- Better tools for whole-genome-based phylogenies
- Methods to produce accurate trees from markers that have been considered to be too difficult to align
- Ultra-fast Maximum-Likelihood analyses (with bootstrapping)
- Alignment methods that properly represent all homologies
- Improved models of evolution (and simulators for these models) to test methods
- Tools for reconciling gene family phylogenies with the "summary tree"
- Tools for determining rates of evolution and timing of polyploidisation events; dating the nodes of the tree
- Ancestral state reconstruction
- Visualisation tools for very large trees
- A means of mapping traits (in the broad sense including e.g., physiology, ecology, biogeography) to internodes of the "summary tree"
- An ontology for naming each node of the iPToL "summary tree" formally and precisely
- Analytical tools to take data from samples (e.g., vouchers, DNA samples) to phylogenetic analysis (phylogenetic data) or specimen databases (vouchers)
- Phylogenetic analysis 'on demand'
- Tool for estimating environmental traits in geographical space
- Tool for georeferencing historical specimen data that are not in latitude/longitude coordinates now
- Tool for mapping specimen data along with environmental data
- Interfaces appropriate for different types of user (researchers, schoolteachers, policy-makers, general public)

Conclusions

A grand challenge proposal to assemble the tree of life will be submitted to the iPlant board of directors. International participation was welcomed but there were very few invited international participants: 2 UK, 2 German, 2 Canada attended the workshop and there was relatively little discussion of related international projects.

4. Computational Morphodynamics
15-19th December 2008

<u>Workshop Summary</u>
Workshop Lead Organizer – Eric Mjolsness (UC Irvine)

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. This GCW workshop aimed to 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.

Feedback from UK Participants: Andrew Bangham (University of East Anglia) and Ruth Bastow (GARNet)

This workshop was organized and led by Eric Mjolsness and Elliot Meyerowitz and was made up of a broad range of participants including a number of key international representatives. After numerous breakout sessions and vibrant discussion groups, delegates concluded that they required a low cost set of tools and user friendly GUIs that would assist in image analysis and finite element modeling (FEM) for cells/organs would be central to any project in this field. A FEM based project did not appear to be well supported by the iPlant team despite this being an area of key importance to the biologists. Pilot projects had been initiated between the Meyerowitz, Mjolsness groups and the iPlant team but these appeared to suffer from a gap between the computational scientists view of the area and the reality of the biological questions being addressed. This clearly reflected that delivering only 10% of the desired outcomes of this (or any other) workshop will be a Grand Challenge.

Conclusions
A proposal on computational morphodynamics will be submitted to the iPlant board of directors. However, workshop attendees expressed concern at the lack of explanation of the extent and areas of cyberinfrastructure that the iPlant team would/could deliver in the given time frame.

General Conclusions from all workshops
Biology instead of the cyberinfrastructure was the driving force in each workshop. Although the biological question asked in each case was of interest it was unclear if the current program of workshops and resulting proposals will be the most successful way to fully utilize the funding that is available and generate the computational infrastructure that is currently needed in this area.
In addition the role of the IPlant team and Cyberinfrastructure was never clearly articulated to any workshop making it difficult to comprehend how the iPlant team will interact with the biologists and tackle the projects submitted.

Current iPlant Grand Challenge Proposals
http://iplantcollaborative.org/grand-challenge-proc/proposed-challenges

Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity
Project Leaders: -Brian J Enquist (University of Arizona), Richard Condit (Center for Tropical Forest Science), Robert Peet (University of North Carolina at Chapel Hill), Mark Schildhauer (National Centre for Ecological Analysis and Synthesis (NCEAS)) and Barbara M Thiers (William and Lynda Steere Herbarium)

Summary
This project proposes to network core databases and data networks to create a novel resource for quantitative plant biodiversity science. The grand challenge will be to assemble and share the world's rapidly accumulating botanical information from plots and collections to create a distributed, web-accessible, readily analyzable data resource. With such a resource, the group aim to answer major questions of direct relevance to plant ecology, plant evolution, plant geography, conservation, global change biology, and protection of biodiversity and ecosystem services. In particular, *how does climate influence the distribution and abundance of plant species, how does the phylogenetic diversity of plants vary across broad environmental and climatic gradients, and how are plants assembled into ecological communities?* The team propose to create a data resource of unprecedented size and scope together with the tools for

its use, thereby empowering botanists and the general public to better address fundamental issues in plant ecology and global change biology.

Computational Morphodynamics - Grand Challenge Project for iPlant Collaborative

Project Leaders: - BS Manjunath (University of California, Santa Barbara), Elliot Meyerowitz (California Institute of Technology), Eric Mjolsness (University of California, Irvine), Edgar Spalding (University of Wisconsin) and Anne Sylvester (University of Wyoming).

Summary

This proposal aims to further our knowledge of how systems of intercommunicating cells and tissues produce a highly dynamic plant. Achieving this level of understanding of plant growth and development is a multiscale spatio-temporal problem. Its solution requires integrating what can be known about molecular processes and networks, the emergent properties that result from their cellular contexts, and the physical forces that are important drivers and as well as consequences of plant form. Much of the necessary bridging information can be computationally extracted from electronic images at various spatial scales and used to inform computational models.

This proposal aims to facilitate the flow of information from images of various types into appropriate modeling activities via major development and integration in three specific areas (i) computer-assisted acquisition and analysis of images of cells, tissues, organ systems and whole plants actively undergoing morphogenesis; (ii) biochemical/developmental modeling, needed to represent and test specific hypotheses for the mechanisms by which cellular activities, cell–cell communication and coordinated growth produce dynamic cellular patterns; and (iii) biomechanical modeling, needed to provide computer representations of plant tissues in which each cell, its mechanical influence on its tissue, and the feedback of mechanical force on the enlargement and division of cells combine to provide a realistic and testable substrate for development.

Assembling the Tree of Life for the Plant Sciences (iPTOL)

Project Leaders: Michael Sanderson (University of Arizona), Michael Donoghue (Yale University), Pamela Soltis (Florida Museum of Natural History), Douglas Soltis (University of Florida), Val Tannen (University of Pennsylvania), Alexandros Stamatkis (University of Muchen) and Todd Vision (University of North Carolina)

Summary

This project proposes to bring plant and computer scientists together to focus on the grand challenge of assembling all knowledge of the phylogeny of green plants to build a comprehensive phylogeny fully 100 times larger than the largest existing trees, to build a cyberinfrastructure for the dissemination of data associated with trees, and to implement scalable "post-tree" analysis tools to enable integration of the plant tree of life with the rest of the botanical sciences.

Cyberinfrastructural Support for the Genetic and Ecophysiological Decipherment of Plant Phenological Control in Complex and Changing Environments

Project Leaders: Justin Borevitz (University of Chicago), Edward Buckler (Cornell University), Melanie Correll (University of Florida), Chad Myers (University of Minnesota), Johanna Schmitt (Brown University), Stephen Welch (Kansas State University) and Jeffery White (USDA/ARS).

Summary

This proposal aims to further our understanding of the complex regulation of plant phenology so that researchers can predict this process across a broad range of species and environments to increase the fitness of crops in the field.

To achieve this aim the project proposes to
1. Develop adapt, and/or adopt an ontology to unify plant data bases
2. Generate tools for integrating metabolomic, transcriptomic, and current/future high-throughput data; this will include (*i*) construction of software links to diverse data types and (*ii*) development of an integrative network explorer to overlay and peruse the data
3. Advance current text mining capabilities
4. Incorporate tools for generating gene- and network-level hypotheses from diverse data;
5. Integrate data on *G x E* interactions for species, landraces, and genotypes and facilitate new technologies for collecting such information
6. Establish a resource for species-specific knowledge of metabolism
7. Create quantitative models from the gene- and network-levels to ecophysiology

Grand Challenge Project Proposal: Fundamental Understanding of Photosynthesis through Integration of Complex Molecular Data and Modeling Tools
Project Leaders: Thomas Brutnell (Cornell University), Richard Bruskiewich (IRRI), Xinguang Zhu (University of Illinois), Christopher Myers (Cornell University), Martin Gent (Connecticut Research Station), Pankaj Jaiswal (Oregon State University) and Shaoying Kathy Lu (University of Illinois).

Summary
This goal of this project is to provide innovative cyberinfrastructure for scientific research into the process of photosynthesis to advance our knowledge and help to solve two major global challenges in plant biology:
a) Enhancing photosynthetic efficiency in crop plants of agronomic importance; and
b) Identifying a strategy to convert agronomically important plants, such as rice and wheat, to use the more efficient variant of photosynthesis which is found in species like maize and sorghum.
To achieve these scientific objectives, the project aims to provide the following :
1. A community portal for photosynthesis research
2. A public data repository for photosynthesis research.
3. A knowledge discovery environment for photosynthesis systems biology.
4. A simulation environment for systems modeling of photosynthesis.
5. An education, outreach and training portal for photosynthesis science.

Cyberinfrastructural Support for Model-Building In Plant Stress Biology: Effects of Climate Change Factors on Crops Across and Within Species
Project Leaders: Ruth Grene (Virginia Tech), Asa Ben Hur (Colorado State University), Doina Caragea (Kansas State Manhattan), Melanie Correll (University of Florida), Pankaj Jaiswal (Oregon State University), and Pamela Ronald (UC Davis).

Summary
This project proposes to dramatically improve quantitative prediction (modeling) of phenotypes to facilitate the characterization of "molecular pathways" affecting agriculturally important stress related traits affected by climate change. The team aim to identify stress responsive mechanisms that may be specific to given genera, species, or genotypes, as well as those that are common across species. To achieve these goals will require the integration of current and future data over multiple scales of organization and across species, with a primary focus on putative abiotic stress resistance mechanisms, from the genomic through the whole plant level, including yield in the cases of crop plants. The project will focus on building a common ontology, development of tools for text-mining, data integration, modeling, and inference of possible regulatory networks from compilation and modeling of existing data.