An International Model for the Future of Plant Science

Report of the International Plant Science Vision Workshop, March 2009.

Several documents have offered visions for plant science in the past year, which have varied in their geographical and scientific focus. Plant science researchers representing ten nations were invited to the Banbury Center, USA, to provide an international view in the report that follows. Funding for the meeting was provided by three national funding agencies. Membership is detailed in section IX.

Table of Contents

I. Introduction

Today's societal challenges are global. The impacts of, for example, climate change and shrinking arable land do not respect national boundaries. The plants we grow and how we grow them will need to change if we are to continue to meet the pressing need for food, fuel, shelter and novel plant-based products.

Building on prior investments and co-ordination, plant research is now in a unique position to develop an international view to take its activities to the next level. This document sets out a bold, decadal vision for plant science that will maximize the impact of research investments to develop outcomes in the field with benefit to all. This plan has been developed around guiding principles, which are critical to its success:

- There is strength in diversity. The maximum benefit will be gained by employing a diversity of research systems and plant species, and drawing from the broadest range of experiences and perspectives.
- International coordination of investment and research efforts will be essential to leverage national investments and synergize outcomes that lead to step changes.
- The research should lead to the development of a universally-accessible tool kit that will provide modular, off the shelf solutions to enable rapid responses to unforeseen challenges.
- There should be timely and open access to all tools and resources and data generated through internationally coordinated projects.

Following these guiding principles, by the end of the next decade we should have achieved a research-based foundation for a diverse, robust, sustainable and efficient global agricultural system.

II. Executive summary

Plant biology is more important today than ever before, if we are to meet the demands of a growing global population and global climate change. The leap now required in global plant productivity (biomass, food and fibre) can only be secured in the pressing timeframe through quantitative understanding of fundamental mechanisms in plant biology and translating these findings efficiently to crop production. A revolution driven by genome-scale technologies, systems biology and informatics is transforming the practice of plant research. Plant science must therefore balance three goals: understanding fundamental processes of plant biology with sufficient accuracy to build up a mathematical model of the plant from its molecular parts; deciphering the particular biological processes that underlie past and present agriculture, such as hybrid vigour; and applying the new understanding and models to produce crops and protect ecosystems for the decades ahead.

These goals present the research community with conceptual challenges, which demand new research tools and internationally-coordinated infrastructure:

- New experimental methods are required to locate and quantify all molecular components of plant cells, from small chemicals to complex structures, because these link genome sequences to physiological outcomes. Arabidopsis and rice will remain the reference plant species for these studies. A small number of additional species should be adopted for focussed study, with genome sequence data for a wider diversity of plant life and high-throughput measures of plant performance in diverse conditions. A global policy on access to genetic resources will be required to gain the best value from the wealth of national and international germplasm banks.

- The integration of current "'omics" data is not a solved problem, but the goals of plant science require much more than this. Quantitative data on molecular components must now be linked to mathematical models and geographical maps, potentially in "real" time as a plant grows or an epidemic unfolds. The plant science community has provided successful, shared data resources. Open access to pre-competitive data and interpretation require a new commitment to coordination and funding, and may require community-approved enforcement.

- Mathematical models offer the only viable means to understand how plant performance “from seed to seed” derives from interactions among the molecular parts. New methods will be required for modular modelling across multiple scales, in order to link genome sequence and quantitative data to the mechanisms of plant growth, and to the outcomes in crop science and ecology.
- A vigorous pipeline must link testing in the laboratory to validation in field conditions. Establishing this pipeline is a genuinely innovative project covering human resources, interaction across public and private sectors and international regulatory reform. A wholesale reorientation of the educational programmes in biological sciences will be required to multiply the numbers of interdisciplinary scientists and of plant breeders adept in new technologies.

Funding organisations need to provide major and specific support, approximating the scale of biomedical research, to achieve these goals. Our vision is to build from the first, isolated examples of technology-enhanced crop improvement to a systematic process that is implemented on an unprecedented scale, in plant species from traditional row crops to trees and algae. This vision depends upon deeper international cooperation, not only to tackle the “Grand Challenges” in research, but also because projects of high international relevance and urgency cannot wait for funding and organization to emerge piecemeal from local initiatives.

III. Societal and Scientific Justification

The food price surges and food riots around the world in 2008 were a stark warning that feeding an additional 2.2 billion people by 2050 will not be an easy task, but it is one where plant science will make a decisive contribution. The doubling of world food production over the past half century has come at a cost. Almost all accessible agricultural land has been used and much of our existing farmland has been degraded by salination and overuse of fertilizers and pesticides. In the next 40 years food production must double again, while restoring and maintaining the quality of the planet’s natural resources for future generations. Failing to achieve these goals could have devastating consequences.

Plant genetics has contributed half of the historical increase in crop yields. Future success will similarly depend on novel, higher-yielding crops. The other challenges that face us have never been seen before – severely reduced water availability, global warming and volatile weather patterns, competition from non-food crops and urban encroachment as the growing population moves from the farms to the cities. The new crops must therefore combine higher yield with improved quality, with tolerance to drought and other stresses, and with resistance to pests and disease. Better agricultural practices, national infrastructures and imaginative new international policies will also be needed to ensure that a global population of nine billion are fed equitably and sustainably, but increased and stable crop yields are the crucial foundation.

Plant research is both a foundation for agriculture and a foundation for all biology. Major scientific advances and new concepts from past and current plant science include the cell theory, the laws of genetics, genetic regulation, small regulatory RNA’s, and epigenetics. In addition, the functions of conserved genes in protein degradation, photoperception and many other areas have been determined first in plant species. Undoubtedly plant science will yield unexpected findings in the future, which will bring fundamental, new dimensions to our understanding of all biology.

Plant science is now at the brink of a revolution driven by genome-scale technologies, systems biology and informatics, which will underpin the next generations of crops. Research covering many areas will contribute, including the mechanisms of development and growth, regulatory and metabolic networks and the molecular basis of evolutionary and adaptive change. A major challenge will be to achieve these fundamental advances while ensuring that the results flow rapidly to practical crop plant breeding. Combining model species research with GM approaches offers a demonstrated acceleration in plant improvement. New paradigms will be required to develop non-food crops, for bioproducts - biofuels and pharmaceuticals. Entirely new crop species must be developed and evaluated.

The next decade will be critical for plant science. Increased food production will be required to sustain the dramatic population jump expected within the decade. The policies, the investment and the science conducted in these years will determine how successfully we meet the greatest and most immediate challenge ever faced by the human race.

IV. Decadal goals in plant science

Research in plants will target three transformative goals:

(i) to understand the central processes of plant biology with sufficient accuracy that a mathematical model of the plant can be built up from its molecular parts;

(ii) to decipher particular plant processes that underpin agriculture; and,

(iii) combining new understanding with the modelling approach, to design the food, fuel and fibre crops for the decades ahead.

Among the central processes of plant biology, plant science should aim:

- To gain a molecular-level understanding of how structural, regulatory and metabolic networks, including epigenetic processes, are integrated in living plant cells.
- To understand how cells constrained within cell walls can communicate to form an entire organism.
- To deconstruct the mechanisms that plants use to integrate the internal, developmental and environmental signals that control adaptive responses in the ecosystem.
- To understand the molecular basis of evolutionary change in key species over longer timescales and how this links to biodiversity.

Uncovering the biological foundations of agriculture, plant science should aim:

- To decipher the biological processes underlying key plant traits including heterosis, apomixis, perenniality, self-incompatibility and the control of recombination.
- To understand the evolution and genome dynamics involved in the domestication of crops from wild plants.

Applying new understanding and dynamic models, plant science should aim:

- To link the knowledge gained from understanding of these regulatory networks and evolutionary driving forces to create crop ideotypes - idealized plant types with specific combinations of favorable characteristics.
- To move plant breeding from empirical-based approaches to a predictive science using models of plant growth and behaviour based on molecular data - the computable plant.
- To preserve ecosystems and their interrelationships with agriculture, by applying knowledge of the molecular basis of biodiversity.

Realization of these goals will allow the decoding of plant life “from seed to seed”.

V. Conceptual challenges

Plant cells process many inputs to achieve adaptive responses in metabolism and development. Each plant’s particular combination of adaptive programmes must be decoded in terms of specific, underlying biological processes. In parallel, our understanding of each molecular process must be integrated in order to reassemble models of a whole plant. During the last decade, an impressive amount of data on these processes has already been obtained, confronting us with a new set of conceptual challenges.

A. Develop multidisciplinary approaches, in particular modelling

Comprehensive models of plant growth offer the only viable means to integrate this range of information and apply it to address the societal and environmental challenges outlined above. There is widespread agreement that new types of mathematical modelling are required to understand plant biology in an integrative manner. Detailed models of particular subsystems must be combined within a larger theoretical framework to understand how plant performance “from seed to seed” derives from interactions among the parts. This requires new theoretical concepts for modular modelling across multiple scales in time and space, with multiple physics (diffusion, mechanics). Such methods will be equally applicable to modelling other multicellular organisms.

B. Extend studies to a larger number of reference systems.

Arabidopsis and rice will remain the reference plant species, partly for practical reasons of efficiency in the introduction and validation of new approaches, and partly for conceptual reasons arising from the depth and concentration of biological understanding. However, plant science will move beyond the focus on these pioneering model species, in order to translate current understanding from model to crop species, to enable comparative studies that enhance understanding of all plants, and to study traits that are otherwise inaccessible. The adoption of a small number of additional reference species from unicellular algae to trees will expose more complex genomes and biology to detailed study. Sequencing technology will rapidly bring genome data for a still wider diversity of plant life, including “orphan crops” and species with unusual adaptations.

C. Coordinate data management at the level of the whole scientific community

Simultaneous transformations in the rate and diversity of data acquisition, together with the wider range of species under study, risk overwhelming the research community’s ability to assimilate the experimental results. The conceptual challenges in data integration and data management will limit our ability to extract scientific value, especially at the network-wide scale. The difficulties can be acutely felt by researchers engaged in modelling, who may be in different locations from those generating data. The success of predictive biology depends upon connecting these groups, so the data integration issue must be addressed in a coordinated manner across the international plant research community.

D. Translate fundamental plant science to applications

The models derived from fundamental research must extend beyond the laboratory to the complexity of natural and field environments. In order to create plants capable of producing dependable, robust yields in the face of changing environmental conditions, a vigorous pipeline must link genome assembly and testing in the laboratory to phenotyping and validation in field conditions. Systems biology models can thus be linked to the ongoing modelling traditions of crop science and ecology, and metagenomic sequencing can inform research on biodiversity and climate change. The establishment of this pipeline is a genuinely innovative project that will require a multi-faceted approach, covering human resources, interaction across public and private sectors and international regulatory reform.

VI. Scientific needs: tools and their applications

Plants are complex systems, i.e. they are composed of elements with particular properties that interact to form structures at a higher scale with emergent properties depending on these interactions. Thus molecules assemble into complexes, which assemble into cells, tissues, organs and finally entire plants. Plants, in turn will interact with their environment as part of ecosystems. If we want to understand how these assemblies function we have to study the components at each level of organisation. More particularly, we need to define what the parts are, how they interact and the dynamics of their interactions. In this context quantitative approaches will be central at all levels of organisation.

A. *Intra-cellular level*

A full knowledge of all the molecules, from enzymes to metabolites, that make up plant cells is now required, including their spatio-temporal distribution within the cell or plant. This information must be obtained using quantitative biochemical assays, ideally reaching single-cell and single-molecule resolution, including proteomic and metabolomic approaches.

In vivo imaging and spectroscopic techniques will be essential to locate and quantify the molecular components. Large collections of protein-tagged transgenic plants can be developed rapidly as a first step, to localise individual proteins. However, detection methods must be extended to locate and quantify small molecules (such as hormones).

The next challenge will be to identify how these parts interact to form a robust molecular network. Mapping and modelling the genetic regulatory network has been started using a range of 'omics' approaches focused on transcriptional regulation. Essential quantitative and spatial information is still missing regarding transcription levels, epigenetic regulation, protein stability and dynamic distribution. This must now be provided using new experimental techniques. However, the regulatory network cannot be described by the activity of genes and transcription alone, but also depends on other cellular components, for example when cytoskeletal elements transport regulatory molecules. Efforts should therefore be aimed at elucidating how the chemical constituents interact to form physical structures: for example, how the cell wall assembles.

The structural, metabolic and regulatory molecular networks can only be understood using integrative approaches that combine diverse data types into models that represent spatial structure in detail. Ultimately, this should enable us to understand how a physical, dynamical structure such as the plant cell is assembled, responds and functions.

B. *Cellular and Inter-cellular level*

Cells interact biochemically and physically to form tissues, in which differentiated cell types are characterised by distinct functions (photosynthesis, uptake of nutrients, development of reproductive structures) and, importantly, distinct growth dynamics.

In the last decade considerable advances have been made regarding major signalling pathways that coordinate cell behaviour. Examples include the light regulation of gene expression, or the hormone signalling pathways. By contrast, we do not know the functions of hundreds of receptor-like molecules and putative peptide hormones that have been identified using genomics. The potential cell-to-cell signalling functions of metabolites and RNAs largely remain to be elucidated. Cells also interact physically. The physical interactions are translated into cellular responses via the structural components such as the cell wall and the cytoskeleton. New methods will be required to interrogate these little-known processes.

The molecular regulatory networks interact with structural elements of the cells to determine their size and shape, through modification of the cell wall. This is a major issue in developmental biology, which is not understood in detail for any cell. Quantitative understanding will require comprehensive analysis of growth patterns with cellular resolution, combined with dynamic perturbations of the regulatory network.

C. Organismal level

The molecular networks making up cells and tissues ultimately lead, in interaction with the environment, to the formation of a plant with a specific size and architecture. The behaviour of the whole plant cannot be predicted from the properties of molecules and cells, so the plant must continue to be studied at the scale of tissues and organs. Quantitative measurement of growth patterns in a range of environments has led to the development of descriptive models. Integrating molecular and cellular information will greatly enhance their predictive power. Such multi-scale models will link genotype to phenotype, for example connecting genomic information arising from next-generation sequencing with crop traits at the whole-plant level. To connect environmental inputs to whole-plant responses will require precision phenotyping methods that can be applied in the field and in natural environments.

The analysis above is applicable to all plant species and refers to common mechanisms. Comparative analyses will highlight the differences among species that lead to their distinct functional characteristics, and the relevant mechanisms of genome evolution. Plants offer particular opportunities to unravel the evolutionary consequences and opportunities arising from hybridisation, polyploidy, and genome duplications.

D. Supra-organismal level

The plant's place in the ecosystem exposes it to abiotic factors and to other organisms, both beneficial and harmful. Metagenomic analysis of the phytosphere, investigating phytopathology and symbiosis, will resolve genetic diversity in unprecedented detail, potentially tracking evolution in near-real-time. Combined understanding of genetic diversity and environmental responses will reveal the spectrum of selective pressures and the molecular origins of adaptive change. In the context of crops in the field, this integrated understanding of genotype, phenotype and agronomic environment is the foundation for the specification of crop ideotypes that match local conditions. Actually realising these ideotypes will depend on a greater understanding of natural DNA transactions (such as recombination), and will demand enhanced transformation technologies for rapid and specific manipulation of crop genomes.

E. Common strategies across levels of organisation.

The data regarding the interactions of hundreds of components (molecules, cells, or plants) in time and space will be impossible to grasp without the appropriate informatics and modelling tools. Integrated databases are needed to connect different data types and accelerate the development of quantitative models. Models in the form of virtual cells and tissues are being developed, but complex spatial representation remains a challenge. Multi-scale modelling will require new theoretical frameworks that populate the spatial structures with biochemistry, physics and cell biology. This area is already testing computing power to the limit, and will require innovative software and hardware. Both theoretical and technical developments will then be required to apply the systems approach on a large scale: first to build the models, and then to automate data acquisition, analysis and model parameterisation, in order to test the effects of many environmental signals, genetic manipulations or chemical perturbagens.

VII. A Global Infrastructure for Plant Science

The next five years will see a revolution in the amount of genome-scale data available for plants. Not only will the genome sequence of all important crop and model species be deciphered, but also the sequences of multiple varieties of these species, unlocking the wealth of collections of germplasm amassed over the past 150 years. This will be supplemented by extensive sampling of genetic variation in wider populations; equivalent data from plant pathogens and symbionts; and parallel studies of gene expression and regulation. As new technologies revolutionise both the volume of available data, and our potential to address ever-larger challenges, the importance of developing appropriate, internationally-coordinated infrastructure and human resources is growing. The plant science community has demonstrated profound commitment to and successful provision of shared resources for data acquisition, dissemination and analysis. Three essential pillars are envisioned to build upon this foundation to accelerate international research in plant science over the next decade:

A. Reference resources for biodiversity

National germplasm banks together with those of international centres hold an underutilised wealth of plant diversity. Technological advances offer the possibility of identifying valuable genes and gene combinations at a genome-wide scale. A rational, global genetic resources policy will be essential to ensure that this work is coordinated and that the results are available to all, thereby, gaining best value for research and crop improvement.

B. An informatics infrastructure to support predictive biology

Ultra-high throughput sequencing technologies will provide extensive evidence for gene expression and regulation; the availability of more reference genomes will increase our ability to infer function in species less directly studied. Yet even the relatively simple integration of “omics” data is not a solved problem, with both informatics and organisational factors promoting data fragmentation and redundancy. If the potential offered by high-throughput data generation is not to be wasted, and if predictive biology is to be realised, we will require much more than this: the ability to link such data to quantitative chemistry, to developmental models and geographical information systems, and to provide systems offering access to such data in a “real” time frame (e.g. that defined by the growth of an organ, or the spread of an epidemic).

We need an infrastructure that will facilitate both the development of leading-edge modelling techniques and their subsequent application to real world problems. This should be constructed using a layered approach, building plant-centric services on top of universal resources shared with other domains of biology.

Essential elements within this system include:

1. Compulsory use of public archives and established data standards for primary data and metadata (for DNA sequences, those maintained by NCBI, EBI, DDBJ) enforced by funders, publishers and the wider community. The development of an open informatics platform and public database for plant images and phenotyping is a priority.
2. Use of standard informatics platforms for integration of 'omics data, to save costs on redundant development and to provide a unified interface for downstream researchers.
3. Community-led resources for individual species/groups of related species to define reference sequence, genes etc.; and to take charge of data integration and curation, according to the priorities of each community.
4. High-level resources to provide an integrative, comparative view across the plant/ plant pathogen/ plant symbiont domains.
5. Links between stock collections and 'omics-centric databases should be seamless and efficient.
6. The informatics framework needed to support predictive systems biology is not yet fully specified, and its definition will require contributions from scientists trained in many disciplines outside plant science (mathematics, physics, chemistry, computer science). Special initiatives will be needed to deliver the services that will nurture this inter-disciplinary research.

C. Coordinated international activity for strategic priorities and grand challenges

Projects of high international relevance and urgency cannot wait for funding and organization to emerge piecemeal from local initiatives. Such projects range from the implementational, for example, the determination of reference genome sequences for the most important plant species or the establishment of a coordinated informatics infrastructure for ‘omics-scale data, to conceptual “grand challenges” aimed at overcoming the key bottlenecks to the next generation of plant science, for example the development of multiscale modeling frameworks for systems biology. This will require the development of new mechanisms to support the harmonization of research goals and access to funds.

VIII. Foreseeable outcomes

The leap now required in global plant productivity (biomass, food and fibre) can only be secured in the pressing timeframe through quantitative understanding of fundamental mechanisms in plant biology and translating these findings efficiently to crop production sciences. Concrete examples from contemporary crop improvement programmes have already demonstrated the relevance of this “model species+GM” approach and the rapid progress that has been achieved, relative to conventional breeding approaches. Our vision is to build from the first, isolated examples to a systematic process that is implemented on an unprecedented scale, in plant species from traditional row crops to trees and algae.

A new framework will allow the multi-scale modeling of regulatory and metabolic networks, development and environmental signals in multi-cellular organisms. These models will predict the impact of plant genotypes on responses to environmental challenges, precisely and systematically. At a later stage, these will specify optimal ideotypes for sustainable agriculture.

Sequence-based measures of genetic diversity combined with phenotypic analyses will enhance the effective management of global genetic resources enabling preservation and capture of biodiversity. This diversity will be a source of valuable genes which may be used in developing crop varieties and plant-derived bioproducts.

Transformation technology combined with marker-assisted selection and molecular understanding of gene networks involved in conferring specific traits will allow **the rapid development of new crop varieties with high yield and yield stability.** Varieties engineered to contain genes responsible for desirable traits such as efficient nutrient and water use, tolerance to environmental stress and disease, will support sustainable agricultural systems. These newly-bred or -engineered crop varieties will be able to be grown in marginal lands such as water-limited, water-logged, or high salinity areas, increasing total crop production without the need to create new farmland, thereby allowing the preservation of forests in particular. In addition, the new varieties for a wide range of crops, including those important in developing countries, will maintain production under changing climates.

Synthetic metabolic engineering with genes from plants or other organisms will increase the range and quantity of **bio-products** such as pharmaceuticals by engineering novel genes in biochemical pathways, and for increased production of valuable existing compounds.

From the knowledge of the gene networks involved in the domestication of existing crops, **the conversion of wild species to high value crops** will be accelerated, for example offering new alternatives to fossil fuels that need not compete with food production for land.

Synergies from **international coordination** of research and exchange of practices, genetic resources, databases, tools, and personnel will be captured for the international research community. Worldwide data integration will accelerate the attainment of future research goals, and reduce duplicated activities. This organisation of knowledge constitutes a new type of global intellectual capital, which will lay the foundation for progress in the life sciences in future decades. For example, we anticipate complete plant genome synthesis based on ideotypes designed at nucleotide resolution, that reliably deliver desired plant characteristics.

IX. Recommendations

A. Funding the scientific and infrastructure programme

Funding organisations need to provide major and specific support, approximating the scale of biomedical research, to capture and integrate molecular, cellular, organismal and ecological data, to understand how living plants develop, function and adapt to their environment, and to conduct practical crop improvement in the light of this knowledge. The experimental focus should be on dynamic and quantitative data at multiple scales. In parallel, new levels of effort and investment will be needed to secure interdisciplinary research that delivers modelling and data integration systems.

B. Data capture and open access

All generated data needs to have a funded path to integration and complete downstream dissemination over the period of its usefulness. Open access to pre-competitive data and interpretation provides a strategic advantage to the community, which is not being realised and may require community-approved enforcement mechanisms.

C. Training and education

A wholesale reorientation of the educational programmes in biological sciences will be required in order to multiply the numbers of interdisciplinary scientists and of plant breeders adept in new technologies. Plant science needs to exploit its competitive advantages (culture of data accessibility, international outlook, success in practical application and moral imperatives in ecological and societal goals), in order to attract expertise in theoretical disciplines. Programmes should be introduced to promote direct interaction of laboratory researchers and plant breeders, in order to gain value from new ideas and technologies combined with the existing skills of the breeders.

D. International coordination in research and access standards

1. Research coordination

Plant science has benefited tremendously from collaborative international funding mechanisms. The vision for a coordinated, international effort towards whole-plant models crucially depends upon deeper cooperation, with international agreement of scientific priorities and strategic commitment to international implementation and delivery mechanisms. Joint funding programmes should be broadened to encompass further nations and disciplinary areas, with a substantial increase in funding, in order to gain best value from unique infrastructures and scarce human resources. Support for exchanges of researchers at all career stages should be increased, to broaden their direct experience, extend best practice and maximise the use of specialised research facilities.

2. Streamlining the regulatory framework

Standardising and simplifying regulatory processes at all levels will be required to facilitate open access to data, resources and materials. Plant science should adopt international standards in this area, where these are agreed by the relevant community. The success of translational research from the laboratory to the field has been hampered by intractable, non-science-based regulatory processes. International dialogue should be accelerated to simplify and rationalize this process.

X. Contributors, Acknowledgements, References and Reports.

Participants
Liz Dennis, CSIRO, Australia
Joe Ecker, Salk Institute, USA (Co-Chair)
Mike Gale, John Innes Centre, UK
Klaus Harter, University of Tubingen, Germany
Paul Kersey, EBI, UK
Suk-ha Lee, Seoul National University, South Korea
Minami Matsui, RIKEN Plant Science Centre, Japan
Andrew Millar, University of Edinburgh, UK (Co-Chair)
Isobel Parkin, Agriculture Canada, Canada

Jan Traas, INRA/CNRS/ENS, Lyon, France

Funders
Ruth Bastow and Rowan McKibbin, Biotechnology and Biological Science Research Council (BBSRC), UK
Catherine Kistner, Deutsche Forschungsgemeinschaft (DFG), Germany
Jane Silverthorne, National Science Foundation (NSF), USA