Realizing the Potential of Plant Genomics: From Model Systems to the Understanding of Diversity

Report of the NSF-Sponsored Workshop:

"New Directions in Plant Biological Research"

April 1999

TABLE OF CONTENTS

Foreword to the Report

Since the discovery of the structure of DNA, the field of biology has experienced a series of revolutionary changes in the ways in which problems are approached. The recent development of high capacity methods for analyzing the structure and function of genes, which may be collectively termed "genomics", represents a new paradigm with broad implications for biologists. Instead of characterizing genes one or a few at a time, it is now possible to determine the complete nucleotide sequence of all of the genes in an organism and to measure the amount of mRNA corresponding to all of the genes. Although comprehensive information of this kind is currently available for only a few organisms, it seems likely that comparable levels of information will rapidly become available for most widely studied organisms, including several higher plants. Access to this information, and new tools that exploit it, will profoundly alter the ways in which biologists select and approach questions. This, in turn, will directly impact the application of directed genetic methods to the improvement of economically important plants.

Although future developments in a rapidly emerging field can be difficult to predict, some projection of trajectory can be helpful in planning for the development of community resources and other initiatives that may facilitate progress. Indeed, the success of the *Arabidopsis* genome

project stands as an example of productive foresight. Now that the sequence of *Arabidopsis* is nearing completion, it was deemed useful to reinitiate the process that led to the formulation of the original vision of the project (NSF document 90-80). Toward this end a group of plant biologists (Appendix I) met to discuss the subject at an NSF-sponsored workshop "New Directions in Plant Biological Research" held on November 23 & 24 1998 at the Carnegie Institution of Plant Biology, Stanford University. This document summarizes the perspectives that emerged from these discussions and presents the views/vision of a diverse group of plant scientists. The report also including input from the North American *Arabidopsis* Steering Committee (NAASC) as well as from the plant biology community at large.

Mission Statement:

To exploit the revolution in plant genomics by understanding the function of all plant genes within their cellular, organismal and evolutionary context and to create an information structure to coordinate, integrate, analyze and make accessible knowledge about plant biology.

Specific Recommendations:

- Development of and ready access to comprehensive data, materials sets, tools and enabling technologies including a complete inventory of full-length cDNAs, gene knockouts, and genome chips/arrays

- Promoting "systems" approaches to biological processes using integrative and functional genomic information

- Development of a computerized knowledge base "the electronic plant" to capture all plant functional genomic and systems data

- Training of graduate students and postdoctoral fellows in "systems approaches" to understand plant biological processes

- Establishment of training courses in computational plant genomics/systems biology to prepare students to integrate information about a variety of plant systems to formulate testable hypotheses

I. INTRODUCTION

Plant Genomics: Early Lessons from the *Arabidopsis* Genome Sequence

Plant biologists have been at the forefront of the shift to genomics. As a result, one of the first eukaryotic organisms to be completely sequenced will be the small mustard species *Arabidopsis thaliana* (Meinke et al., Science 1998, 282:662). Because of certain technical advantages,

including a highly compact genome of about 130 Mb with little interspersed repetitive DNA, *Arabidopsis* is one of the most widely used model organisms for studying the biology of higher plants. It is also close relative of many food plants such as canola, cabbage, cauliflower, broccoli, turnip, rutabaga, kale, brussels sprouts, kohlrabi and radish. More importantly, *Arabidopsis* genome sequence information will be applicable to major crops including corn and soybeans, because of the close evolutionary similarities among the genomes of all flowering plants. An international consortium called The *Arabidopsis* Genome Initiative (AGI), which includes six research groups in Japan, Europe and the USA are collaborating to sequence of the genome. About 50% of the genome sequence is currently available in public databases and a large proportion of the genes are also represented by partial cDNA sequences. It is currently anticipated that the complete genome sequence will be available by the end of the year 2000.

The results to date of the *Arabidopsis* Genome Initiative have revealed a new world of scientific processes and opportunities in plants. With greater than one-third of the genome sequence finished, completely unanticipated functions have been found. These recent findings reveal our present ignorance of the mechanisms by which plants grow and function, but also provide the exact experimental material necessary to gain a new understanding: the genes mediating the newly discovered processes. New technical capabilities such as gene chips and gene replacement technologies now, for the first time, allow us to proceed rapidly and surely from gene sequence to function. Thus, there is an unprecedented opportunity to follow up on these recent genomic discoveries by learning the functions of all plant genes, to reveal the previously unknown processes that are fundamental to the mechanisms of plant life.

Recent unexpected discoveries derived from genomic sequence information including the identification of:

1) **Glutamate receptor genes.** Glutamate receptors are components of neurons, where they serve as ion channels in rapid synaptic transmission. Recently two glutamate receptor genes have been found in *Arabidopsis*, each encoding proteins with all of the signature domains of the animal proteins. What role could such previously brain-specific proteins play in plants? Use of channel inhibitors shows that blocking the plant channels seems to affect the ability of the plants to perceive light. What role do the channels play? Several glutamate channel blockers used in brain research derive from plants; they were previously considered to be used in defense against herbivores. Now it appears that these compounds may function in plants to regulate plant channels. How many other bioactive plant compounds, such as caffeine or nicotine, have direct roles in plants? What are the roles?

2) **Receptor kinases.** The *Arabidopsis* genome sequences to date reveal nearly 100 different genes that appear to code for transmembrane receptor serine/threonine kinases, implying that the complete genome will have on the order of 300 such genes. The mutant phenotype of only three is known; two are involved in regulation of cell division and differentiation in meristems, and one appears to be a hormone receptor. A similar gene in rice mediates response to a bacterial pathogen. What are the roles of the hundreds of these proteins? Their existence implies a massive network of cell-cell and environment-plant communication, via a series of ligands yet to be

discovered. Understanding this network will give us an entirely new view of plant development, environmental response, and organismal integration.

3) Nuclear architecture. Use of modern imaging technologies with nuclear proteins (such as photoreceptors and transcription factors) discovered in *Arabidopsis* reveals a series of striking and novel subnuclear localization patterns. What are the nuclear microcompartments that have been revealed, and what is their role in gene expression and genomic function?

4) Multiple drug resistance – ABC transporter genes. There are on the order of 50 genes coding for ABC transporters already found in the *Arabidopsis* genome project and thus, 150 are expected in the complete genome. What do they transport? Some appear to be mitochondrial. What is the nature of this mitochondrial traffic, and the role it plays in nuclear-cytoplasmic interaction and energy metabolism?

5) G-protein coupled receptors. It appears that there are already many candidate genes that could code for 7-transmembrane G-protein coupled receptors in *Arabidopsis*. There is, however, only a single known heterotrimeric G-protein alpha subunit. Do all of the putative receptors act through a single G-protein, or do these plant receptors act in an entirely new signal transduction pathway, different from the pathway in which the cognate animal proteins acts?

6) Blue light photoreceptors. One indication of the amazing degree of conservation in biological systems is the discovery of molecules in plants that are important in human disease. For example, the conserved role of the blue light photoreceptors cryptochrome in regulating both circadian plant growth and rhythmic sleep behavior in animals.

These unanticipated discoveries have opened a new world of experimentation and understanding that we must explore now.

II. Functional and Integrative Genomics: The Next Decade of Plant Research

The goal of the field is to understand plant biology completely. In particular, this means being able to explain in detail how all aspects of plant growth and development occur, how plants respond to change in their environment, the molecular basis of variation between species and how plants respond in complex communities (ie., ecosystems). As integral part of this goal is to make the knowledge broadly accessible. In this respect our traditions for organizing academic knowledge are no longer appropriate to the task and will be even less so in the future.

In order to progress toward that goal in the most efficient manner we need to have a sense of what kinds of discoveries are likely to be most useful so that we can allocate resources and make the many community decisions that shape modern science on a daily basis. It is also essential to identify the factors that will facilitate broad progress in understanding plant biology rather than only focusing on one or two model species. The biomedical community is largely focused on one organism and is not a good model for how plant biology should develop. The completion of the

goals for plant biology described in our mission statement will enable dramatic increases in our understanding of mechanistic aspects of plant gene function.

Assigning Functions to All Plant Genes. An important and revolutionary aspect of the proposal – *to understand the function of all plant genes by the year 2010* - is that it implicitly endorses the allocation of resources to attempts to assign function to genes that have no known function. This represents a significant departure from the common practice of defining and justifying a scientific goal based on the biological phenomena. The rationale for endorsing this radical change is that for the first time it is feasible to envision a whole system approach to function. This whole system approach promises to be orders of magnitude more efficient than the conventional approach. We envision that once the efficiencies of genomics have been realized, presumably within the next decade, there will be a renewed emphasis on problem-oriented approaches and an expanded emphasis on understanding diversity.

In order to attain this level of understanding, a set of scientific problem-oriented goals were defined. The attainment of these goals relies heavily upon the development of a set of process-oriented, enabling technologies (tools and materials) and on further human resource development in the plant sciences.

A. Scientific Problem-Oriented Goals. The natural world provides a virtually endless source of biological phenomena that are amenable to description and explanation. However, experience suggests that certain problems yield insights of broad relevance whereas others do not. What are the problems that are likely to yield the greatest advances?

- ***Identification of the language of transcriptional control elements*** will allow us to be predictive about gene expression (and to control gene expression).

- ***Elucidation of the molecular mechanisms that regulate pattern formation and morphogenesis*** will allow us to understand and control the morphological diversity of higher plants.

- ***Structural characterization of key enzymes*** will allow us to efficiently identify genes and proteins responsible for chemical diversity in plants.

- ***Elucidation of processes that are not found in Arabidopsis or rice*** will provide a useful basis for future work in understanding plant diversity.

- ***Elucidation of basic cellular processes*** that are unique to plants or which differ between plants and animals will be essential to understanding the mechanisms underlying growth and development.

The explicit scientific problems can best be stated in the form of the central unanswered questions in the areas of cell biology, biochemistry/genetics, development and evolution of plants. These questions can and must now be answered.

1. Cell Biology. The function and structure of cells determines the architecture of the plant. We will endeavor to understand the number and function of plant cell types and the physiological contributions and interactions amongst them, including a detailed molecular understanding of each cell type. Major unanswered questions include:

How is cell shape, size, and density regulated—how is plant morphology determined?

What is the basis for broad structural diversity amongst plant species?

How does organ morphology derive from cell physiology and structure and result in adaptive characteristics?

How are asymmetric growth signals coordinated?

To what extent do pathway components physically group within cells?

What are the subcellular signaling domains that control various responses such as asymmetric cell elongation?

How does cell-cell communication happen in plants—what are the roles of plasmodesmata, receptor kinases, adhesins, structural cell wall components, other small molecules?

How does differentiation of plant cell types take place in the absence of cell migration?

What is the role of the cytoskeleton in plant cell differentiation?

What are the unique mechanical properties of plants? How does mechanosensing control differentiation?

Cell wall deposition—how does it happen—how does it relate to expansion and division and differentiation?

Intracellular dynamics of plant cells—how does it differ from other systems?

What is the function of cytoplasmic streaming and how is it regulated?

What are the processes that control/coordinate organelle biogenesis?

Short- and long-range signaling—how does it happen for different stimuli?

How do tissues coordinate their responses to a diversity of signals?--light (quality, quantity, photoperiod), water, ions, gravity, microbes (pathogens and symbionts and epiphytes—both rhizoshpere and phylloplane-associated), and insects?

What are the modes of signaling (chemical/ electrophysiological/ mechanical/ osmotic/bulk flow)?

Nuclear topology/3-D chromatin structure—how does it relate to coordinate expression of sets of genes?

How does cell migration occur in cases such as pollen tube growth?

Regulation of apotosis (programmed cell death), suspensor, xylem element differentiation, role of cell death in leaf morphology—what are the signals—how does pathogenesis relate to developmentally programmed cell death?

What is the basis of non-host resistance?

What is the nature of interorganellar communication—control of nuclear gene expression by chloroplast (and mitochondria)—is nuclear gene expression controlled by other organelles?

Protein and oil bodies—how are they structured? How does the packing work?

2) Biochemical and Genetic Functions. While the past ten years has seen major advances in our understanding the genetic and biochemical basis for gene function in a variety of plants, most notably *Arabidopsis*, major unanswered questions remain. Examples of scientific questions to be addressed include:

-Sensing and responding to environmental cues including light, cold, drought, gravity and many other abiotic signals by which plants interact with their surroundings

-Interacting with an enormous (largely unknown/unexplored) variety of other organisms including bacteria, fungi, insects and other plants via chemical and other signals

-The defining of pathways for biosynthesis of numerous primary and secondary metabolites and for the biogenesis of cellular membranes, walls, and organelles including lipid, protein and polysaccharide biosynthesis.

-The nature of biotic signals and signal transduction mechanisms including hormones, light and a variety of stress producing stimuli.

-The mechanisms of gene expression, including transcription, translation and regulatory networks controlled by kinase cascades and other protein-protein interactions including the assembly and destruction of macromolecular complexes via chaperones, proteasomes etc.

-Epigenetic effects on gene expression, cellular memory, imprinting, gene silencing and paramutation.

-Genetic mechanisms of repair, replication, recombination, transposition, meiotic and mitotic inheritance of the genetic material

Chromosome dynamics, kinetochore function

Functional discovery by sequence comparison with prokaryotes and eukaryotic genomes, such as the discovery of novel universal pathways of protein translocation.

3) Developmental Biology. The final form of a higher plant is produced through a complex interaction of gene expression, preexisting cellular components and the environment. For the first time we will have a catalog of all of the genes of a higher plant as a starting point for developmental analysis. This together with new enabling technologies will now allow an understanding of the fundamental events underlying developmental processes. Identifying the function of all genes will allow an understanding of a diversity of plant process including:

How are whole sets of genes coordinately regulated to produce complex plant structures?

How can genetic differences account for the tremendous diversity of angiosperms

How do plant cells communicate with each other, and how does this communication regulate cellular behavior and differentiation?

How does communication between cells leads to regulation of gene expression?

What is the nature of regulatory networks responsible for regulation of development?

How are environmental signals perceived and how are the perceived signals transduced to modulate gene expression and development?

What are the nature of sequences responsible for the regulation of gene expression?

The nature of regulatory transcription factors and how they interact with specific DNA sequences to regulate gene expression

The interplay between the genome, the cytoskeleton and the cell wall and how this process mediates plant morphogenesis

How plants measure time and respond to timing events (annual an biennial regulatory processes, circadian rhythms)

4) Genome Evolution. Although flowering plants have evolved during the past 150 million years or so, and might therefore be expected to be very similar at the genetic level, substantial developmental and chemical diversity is apparent. Understanding the basis for this diversity is a key to understanding how to effect rational improvements in the productivity and utility of crop species. Knowledge of the genetic basis for intraspecies variation in specific traits should be useful in selecting or creating useful variation within a species**.** Examples of scientific questions to be addressed include:

> How are genomes similar within and between families which chromosomal segments have been conserved (macro and micro)
>
> How does higher order structure compare?
>
> What are the functional consequences of conservation?
>
> What are the differences – interspecific how do these arise (duplication, ploidy, repetitive DNA) contribution to speciation
>
> What are the differences – intraspecies?
>
> How is allelic diversity defined and quantified?
>
> What is the complete gene repertoire of the plant kingdom?
>
> How representative is the genic content of rice and *Arabidopsis*?
>
> How does phylogenetic/ecological diversity predict genic novelty?

B. Process-Oriented Goals: For biologists to fully realize the benefit from the availability of fully sequenced plant genomes, we must facilitate a dramatic change in the current laboratory landscape. Current studies of single genes will transition to large-scale analysis of entire gene regulatory networks. Moreover, with the expected further reduction in the cost of DNA sequencing technologies, comparisons of a few genomes will be replaced by a broad sampling of plant genetic diversity. Ultimately, the future plant biologist will be integrating an array of tools that will allow an appreciation of the complexity which we can currently only imagine.

The proposal made about ten years ago to sequence the *Arabidopsis* genome was a process goal. It was based on the concept that it would be tremendously enabling for the entire community to have the complete sequence of the *Arabidopsis* genome.

> *- To assign a function to all higher plant genes on the basis of experimental evidence by 2010.* An implicit assumption is that higher plants share essentially the same basic set of genes and that by knowing the function of a small number of representatives, it will be possible to infer the probable function in other species. We recognize that in many cases, the exact function of many genes will remain elusive beyond the year 2010. However, we believe that it will be possible to

experimentally assign a general function to each gene in the genome of the model organisms or to establish that inactivation or overexpresion of the gene causes no observable effect on the organism. In order to establish gene function, we believe that, at the very least, each gene must be characterized with respect to the phenotypes cased by inactivation or overexpression, the mode and specificity of regulation, the subcellular localization of the gene product, the identity of interacting proteins and the identity of coregulated genes. By extending the goal to include all higher plants, we intend that research on diversity of gene function must play an increasingly important role.

- ***To determine the complete sequence of a second plant, preferably rice.*** The sequencng of rice must be accompanied by a *similar* kind of infrastructure development that was used to advance *Arabidopsis* (eg., maps, mutants and infrastructure).

- ***To acquire deep EST coverage for all plant species of utility or academic interest.*** This will be very useful in many respects but, in particular, will help to identify genes that are highly divergent between *Arabidopsis*, rice and other plant species.

-***To develop high resolution genetic and physical maps of select plant species*** in particular Grasses (maize, sorghum, wheat, barley) Legumes (soybean) Solanaceae (potato, tomato) and other important crops.

-***To produce low resolution genetic/physical maps and sample sequencing of a variety of flowering plants*** including gymnosperms, mosses, and liverworts (using ESTs, BAC ends, low coverage genomic sequencing)

- ***To characterize the expression of all Arabidopsis and cereal genes*** *under all developmental and environmental circumstances.*

- ***To develop new information management methods that will provide universal access to all information about basic plant biology.*** The large amounts of data that are being generated by genome-related approaches cannot be distributed by print media. There is a pressing need for the development of new information retrieval and analysis approaches. In the short term, these will need to be publicly supported in much the same way the internet was subsidized. Eventually, electronic databases will merge with print media and appropriate cost/benefit models will be implemented.

There is a shortage of people trained in the development and use of bioinformatics. Support for graduate and postgraduate training is needed. Support for the development of academic programs is also needed.

- ***To develop infrastructure and community processes that will integrate basic and applied research in plant biology.***

C. Enabling technologies:To realize many of the process-oriented goals, the following enabling technologies, materials, comprehensive data sets and tools need to be developed and made available to the community:

> **•the complete sequence and inventory of full-length cDNAs** for predicted open reading frames for model species. These will be essential for correct identification of all coding regions in these genomes. These data sets can then be used as a standard for comparison in exploration other species for novel genes not found in the model species

> **•a complete set of insertion mutants in all plant genes** with sequence of flanking regions for rapid mutant identification.

> **•an expression map of every gene by *in situ* and/or enhancer/gene traps.** Alternatively, GFP or similar fusion should be produced for every gene predicted by open reading frame.

> **• tools for parallel analysis of the expression of all plant genes,** using whole genome sequence information (e.g., chip-based technology).

> **•tools for genome-wide, rapid mapping for mutants,** whole genome chips with appropriate sampling redundancy for identification of all differences between two genomes should be assembled and made available. This would enable direct identification of mutations in a single experiment.

> **•tools for genome-wide, rapid mapping of QTLs,** to identify naturally occurring alleles affecting quantitative traits (e. g. chip-based technology for segregating populations).

> **•tools for proteomics,** such as those currently available for yeast that allow easy determination of the identity of purified proteins.

> **• artificial plant chromosome vectors**, to enable the facile movement of large gene sets from plant to plant.

> **• methods for efficient gene replacement by homologous recombination,** to enable precise and reproducible gene function studies.

It is recognized that the community will never assemble these tools unless concerted efforts are made. Large consortia of investigators working on each of these topics, independently of their application, will need to be established. While this is perhaps not the most immediate creative science, the benefits to overall progress clearly justifies the investment of resources. Directed assembly of these tools will prevent the counterproductive "reinvention of the wheel" by all

independent labs wishing to utilize these new technologies. Independent investigators will then be able to focus their efforts on the biological questions at hand.

A key concept for all these methods/tools is access. Access to all these tools must include means of dissemination of appropriate arrays, resources for their use, and means of storing and accessing data generated by their use. The investment in their development will only be worthwhile if it includes establishment of effective mechanisms for ensuring the methods and data will be available to the entire biology research community. The mechanisms for rapid dissemination of information developed by the U. S. component of the *Arabidopsis* Genome Initiative, whereby even preliminary genome sequence data is rapidly disseminated, should serve as a model for this. Under no circumstances should consortia be established under conditions where non-members would be excluded from immediate access/utilization of the products of the consortia. Efforts to assemble these tools need to be established as soon as possible.

It is expected that the application of these and related methods will lead to the assignment of some degree of gene function to most higher plant genes. In addition, parallel studies of the function of all genes in the other, non plant, model organisms will contribute a great deal to understanding gene function. Presumably, this comprehensive approach to understanding gene function will usher in an era of true genetic engineering in which rational changes can be devised from some level of knowledge of the entire system, rather than the one-gene tinkering that has marked the beginning of the era.

III. Human Resource Development.

The availability of massive quantities of plant genome sequence data plus an increasing understanding of plant gene function will bring about rapid changes in all areas of plant biology, from biochemistry and molecular biology to ecology and evolution. Advances in these areas of basic plant biology will be rapidly transformed into practical benefits that will change the face of agriculture and will trigger rapid growth in the agriculturally related biotechnology industry. These anticipated advances will require a new generation of basic and applied plant biologists who possess new combinations of skills, both among college graduates and among those who pursue more advanced training. Providing these highly skilled plant scientists will require the coordinated efforts of colleges, universities, and federal funding agencies, with appropriate support from the private sector and professional societies.

The most important characteristic of the new generation of plant scientists is that they will need to be broadly trained in plant biology. Regardless of whether they are biochemists or ecologists, they will need to have basic training in all areas of plant biology and must be able to work as effective members of multi disciplinary teams. Although a solid foundation in biology will be the most important requirement, the next generation of plant scientists will also need to be highly skilled in a wide array of genomic techniques including the ability to use computers to access and analyze large quantities of biological data.

Although the training of future plant biologists will occur primarily in colleges and universities, federal funding agencies will be important partners in this effort. New training programs are needed at all levels; i.e. baccalaureate, MS, Ph.D., and post-doctoral. New interdisciplinary

training programs at the undergraduate level will require changes in the curriculum requirements in many traditional majors. At the graduate and postdoctoral levels, new and expanded financial support mechanisms are needed. Funding for graduate students and postdoctoral associates has traditionally been tied to individual research grants, which often limits the training possibilities available to the trainees. New funding mechanisms, including both expanded training grants and fellowships to individual students, are needed. These mechanisms should provide trainees with a maximum of flexibility to pursue their training in top quality labs and to gain the various experiences and expertise needed.

Training Goals:

Establish new individual predoctoral and postdoctoral fellowship programs in plant biology.

Expand existing training grant programs, with a focus on high quality programs that provide multi disciplinary training experiences for both predoctoral and postdoctoral trainees in plant biology.

IV. Benefits of the Research.

The availability of the sequence of *Arabidopsis*, together with the comprehensive data sets and tools outlined above will not only dramatically accelerate research on this species, but also will facilitate studies on other plants and higher eukaryotes. As noted above, the availability of a comprehensive set of gene knockouts can be used to make rapid progress on analysis of gene function for currently "unknown" genes. The majority of such genes are common to all plants, and a large fraction are common to all higher eukaryotes. Thus, much of the work on *Arabidopsis* will be directly applicable to these other systems, and *Arabidopsis* will be the system of choice for answering many questions on gene function. In addition, syntenic relationships are already being established between the chromosomes of *Arabidopsis* and a variety of crop species. Exploitation of these relationships will dramatically facilitate molecular identification of genes associated with specific desirable traits in crops. The perceived viability of this approach is clearly evidenced by the significant investment of the private sector in *Arabidopsis*/crop comparative genomics.

On the biochemical and cell biological side, proteomics tools will enable rapid identification of proteins from very small samples. This will engender a resurgence in biochemical approaches to biological questions. Small-scale protein purification and determination of the nature of the protein will be a routine method of gene identification. Obtaining and examining the knockouts of the identified genes will immediately lead to further characterization of biological function. Similarly, partners of proteins, or other components of protein complexes will be more easily identified and analyzed.

On the genetic side, isolation of genes and understanding their molecular/biochemical function will no longer be a bottleneck. For example, allowing the immediate ability to compare sequence similarities between members of gene families. This will lead to a renewed focus on genetic screens "rational genetics"- aimed at understanding gene function at a tissue or organismal level. Researchers will be able to afford to invest more time on relatively complex screens, including

those that might yield a lot of "false hits". They will be able to afford to discard many of the mutants coming from these screens because of the ease of isolation and analysis. Genetic screens are still one of the most effective tools to identify gene function, even in the age of cDNA arrays, and are unlikely to be superseded soon by *in silico* analyses.

In addition to analysis of small sets of genes or proteins, genome-wide analysis will be enabled. For example, such approaches have already led to identification of new, previously undetected regulatory sequences in yeast. The ability to simultaneously study regulation of whole sets of genes, or even of all genes, will enable unprecedented insight into complex biological processes. If properly implemented, it should be possible to assemble a database of expression profiles of each condition, as it is determined. Other researchers could draw on and add to this database as they perform their own analysis. This should form the basis of a comprehensive map of regulation of genes during development and in response to environmental conditions.

In the long term, such a database could be integrated with the growing body of biological information on gene function, plant form and plant biochemistry. The eventual goal would be assembly of a virtual plant—click on a leaf at three days of age and get a list of all genes up-regulated in a variety of physiological conditions; click on a root and watch it grow with a dynamic list of changes at the molecular level accessible. Thus, the data would be accessible to all researchers in an easily utilized form.

The development and application of genome sequences and enabling technologies will lead to the removal of a series roadblocks to progress, putting the answers to all of these questions in reach. For example, genomic technology will enable the discovery of:

> **-novel plant functions previously obscure to conventional genetic and biochemical analysis**
>
> **-the discovery of novel signaling molecules, including peptide and other small molecules responsible for intracellular, intercellular and interorganismal interactions**
>
> **-the interactions among and between metabolic pathways**
>
> **-integrative approaches to complex cellular functions such as cell growth, division and morphogenesis by parallel analysis of regulatory networks in many different areas (e.g. metabolism and hormone signaling)**

The whole range of plant developmental processes will be understood at a molecular enabling manipulation of specific development processes including the regulation of plant size, the time to flowering**,** the mechanisms controlling seed dormancy and embryogenesi**s** Understanding of the mechanisms regulating plant development using genomic technology will enable rational alteration of this process in agronomically useful directions by allowing:

> **-the identification of candidate genes underlying traits**

-gene prospecting (looking for novel variants)

-engineering novel alleles/traits

-rapid characterization and managing germplasm resources

The ability to study gene expression and genetic function in parallel for multiple genes, networks, and their products will revolutionize mechanistic approaches to plant physiology. Understanding the plant genome structure will enable mechanistic dissection of chromosomal biology, including the assembly of complete, interacting pathways for biosynthesis of novel compounds, and will allow whole genome analysis of epigenetic mechanisms and their relationship to genome organization and combinatorial approaches to genetic analysis via functional libraries

Rapid methods for comprehensive analysis of coordinated regulation of gene expression will enable new means of discovery of gene function through rapid analysis of mutants, as moving to a gene will no longer be a bottleneck.. Moreover, these tools will allow testing of models of functions of single genes, and even construction and reintroduction of novel assemblies of multiple genes.

V. Application of the Intellectual Advances: Potential Benefit to

Applied Science

A. Molecular Analysis of Traits. Molecular genetics has to this point been dominated by the identification of single genes that control phenotypes of interest. However, this approach is of limited value in understanding the molecular basis for complex phenotypes that are complicated integrals of the action of many genes. For instance, yield, although exceedingly important, is relatively undefined at the biochemical and physiological level. This lack of definition stems not only from the complexity of the phenotype, but also from the limited number of specific biochemical or physiological measurements that can be made.

The identification of every gene in a particular plant species gives the possibility to analyze expression of each gene in particular varieties or genotypes, in different environments, and in response to various stresses. In addition, modern chromatographic methods, coupled with on-line detection by mass spectrometry, are allowing the resolution of increasingly complex mixtures of small molecules and proteins. As a result of these parallel developments, the assignment of surrogate end-points for detecting complex phenotypes will become increasingly commonplace and facile.

Understanding patterns of gene regulation will directly lead to the ability to express genes at appropriate places and times in development or in response to specific extra-organismal cues. This knowledge will be applicable to genes engineered into plants from any source, including those outside the plant kingdom.

The result of understanding how complex properties are regulated will be an ability to breed crops that are improvements of current varieties, as well as new crops that make altered amounts or types of certain components of that crop's output (*e.g.* starch, oil ,protein). Moreover, the ability to precisely control gene expression and redirect metabolic output will make possible the production of novel materials in plants at economically attractive levels.

B. Molecular Breeding. As the resolution of genetic maps in the major crops increases, and as the molecular basis for specific traits or physiological responses becomes better elucidated, it will be increasingly possible to associate candidate genes, discovered in model species, with corresponding loci in crop plants. Appropriate relational databases will make it possible to freely associate across genomes with respect to gene sequence, putative function, or genetic map position. Once such tools have been implemented, the distinction between quantitative genetics (breeding) and molecular genetics will become blurred. Breeders will routinely use computer models to formulate predictive hypotheses to create phenotypes of interest from complex allele combinations, and then construct those combinations by scoring large populations for very large numbers of genetic markers.

The vast resource comprising breeding knowledge gathered over the last several decades will become directly linked to basic plant biology, and enhance the ability to elucidate gene function in model organisms. For instance, traits that are poorly defined at the biochemical level but well established as a visible phenotype can be associated by high resolution mapping with candidate genes. Orthologous genes in a model species, such as *Arabidopsis* or rice, may not have a known association with a quantitative trait like that seen in the crop, but might have been implicated in a particular pathway or signaling chain by genetic or biochemical experiments. This kind of cross-genome referencing will lead to a convergence of economically relevant breeding information with basic molecular genetic information. The specific phenotypes of commercial interest that we expect to be dramatically improved by these advances include both the improvement of factors that traditional limit agronomic performance (input traits) and the alteration of the amount and kinds of materials that crops produce (output traits). Examples include:

> **-abiotic stress tolerance (cold, drought, salt)**
>
> **-biotic stress tolerance (fungal, bacterial, viral, chewing and sucking insect feeding)**
>
> **-nutrient use efficiency (N,P,K)**
>
> **-manipulation of plant architecture and development (size, organ shape, number, and position, timing of development, senescence)**
>
> **-metabolite partitioning (redirecting of carbon flow among existing pathways, or shunting into new pathways)**

C. Rational Plant Improvement. The implications of genomics with respect to food, feed and fibre production can be envisioned on many fronts. At the most fundamental level, the advances in genomics will greatly accelerate the acquisition of knowledge and that, in turn, will directly

impact many aspects of the processes associated with plant improvement. Knowledge of the function of all plant genes, in conjunction with the further development of tools for modifying and interrogating genomes, will lead to the development of a genuine genetic engineering paradigm in which rational changes can be designed and modeled from first principles.

APPENDIX I: ROSTER OF WORKSHOP PARTICPANTS

Roger N. Beachy, Ph.D.
Donald Danforth Plant Science Center
St. Louis, Missouri
rnbeachy@danforthcenter.org

Joseph R. Ecker, Ph.D.
University of Pennsylvania
Department of Biology
415 S. University Ave
Philadelphia, PA 19104-6018
Email:jecker@atgenome.bio.upenn.edu
Tel: (215) 898-9384
FAX: (215) 898-8780

Charles S. Gasser, Ph.D.
University of California,Davis
Molecular and Cellular Biology
1 Shields Ave.
Davis, CA 95616
Email: csgasser@ucdavis.edu
Tel:(530) 752-1013
FAX:(530)752-3085

Steve A. Kay, Ph.D.
The Scripps Research Institute
Department of Cell Biology
10550 North Torrey Pines Rd.
La Jolla, CA 92037
Email:stevek@scripps.edu
Tel: (619) 784-2360
FAX: (619)784-2368

Kenneth Keegstra, Ph.D.
MSU-DOE Plant Research Laboratory
Michigan State University
East Lansing, MI 48824
Email:keegstra@pilot.msu.edu
Tel:(517) 353-2270
FAX:(517) 353-9168

Rob Martienssen, Ph.D.
Cold Spring Harbor Laboratory
1 Bungtown Road
Cold Spring Harbor
NY 11724
martiens@cshl.org
Tel (516) 367 8322
Fax (516) 367 8369

Susan R. McCouch, Ph.D.
Cornell University
Department of Plant Breeding
418 Bradfield Hall
Ithaca, NY 14853
Email: srm4@cornell.edu
Tel: (607) 255-0420
FAX: (607)-255-6683

Elliot Meyerowitz, Ph.D.
California Institute of Technology
Division of Biology 156-29
1200 East California Boulevard
Pasadena, California 91125, USA
email: meyerow@cco.caltech.edu
telephone: (626) 395-6889
fax: (626) 449-0756

Chris Somerville, Ph.D.
Carnegie Institution of Washington
Stanford University
260 Panama Street
Stanford, CA 94305

Tel:(650)325-1521, ext 203
FAX:(650)-325-6857

Eric Ward, Ph.D.
Novartis Agribusiness Biotech Research,Inc.
3054 Cornwallis Road
Research Triangle Park, NC 27709-2257
Email:eric.ward@nabri.Novartis.com

Susan Wessler, Ph.D.
University of Georgia
Dept. of Genetic
Life Sciences Bldg.
Athens, GA 30602
Email: sue@dogwood.botany.uga.edu
Tel:706-542-1870
FAX: 706-542-3910

Observers:
David Meinke, Ph.D.
OAD/BIO
National Science Foundation
NSF Plant Genome Research Program
4201 Wilson Boulevard
Arlington, VA 22230
Email:dmeinke@nsf.gov
Tel: (703) 306-1470
FAX:(703) 306-0356

Paul Gilna, Ph.D.
BIO/DBI
National Science Foundation
4201 Wilson Boulevard
Arlington, Virginia 22230
Email:PGILNA@nsf.gov
Tel: (703) 306-1469
FAX: (703) 306-0356

Roger Hangarter, Ph.D.
National Science Foundation

BIO/IBN
4201 Wilson Boulevard
Arlington, Virginia 22230
Email: rhangart@nsf.gov
Tel:(703) 306-1422
FAX: (703) 306-0349

Judith A. Verbeke. Ph.D.
National Science Foundation
BIO/IBN
4201 Wilson Boulevard
Arlington, Virginia 22230
Email:jverbeke@nsf.gov
Tel:(703) 306-1422
FAX: (703) 306-0349

Workshop Agenda

"New Directions in Plant Biological Research"

Carnegie Institution of Plant Biology

Stanford, California

November 23-24, 1998

<u>Host</u>

Chris Somerville, Carnegie-Stanford

<u>Chair</u>

Joe Ecker, University of Pennsylvania

Monday, November 23

9:00 a.m. Welcome-Meeting Logistics-Chris Somerville

9:15 a.m. Introduction- Charge of the Workshop- Joe Ecker

9:30 a.m. Begin Participant Perspectives: "The Future of Plant Research"

9:30 a.m. Elliot Meyerowitz

10:00 a.m. Chris Somerville

10:30 a.m. Break

10:45 a.m. Steve Kay

11:15 a.m. Chuck Gasser

11:45 a.m. Sue Wessler

12:15 p.m. Lunch

1:30 p.m. Rob Martienssen

2:00 p.m. Susan McCouch

2:30 p.m. Roger Beechey

3:00 p.m. Ken Keegstra

3:30 p.m. Eric Ward

4:00 p.m. Break

4:15 p.m. David Meinke

4:45 p.m. Paul Gilna

5:15 p.m. General discussion of issues/questions raised:

All participants

7:15 p.m. Dinner

November 24

9:00 a.m. General Discussion of the Direction of Future of Plant Research: Entire group

12:00 p.m. Lunch

1:30 p.m. Formation of break out groups (3 persons/group) responsible for further development of a specific recommendations (1 page each)

3:00 - 3:15 Break

3:15 p.m. Assembly of entire group for discussion of break-out group reports

5:15 p.m. Wrap-up session- writing assignments

5:30 p.m. Workshop adjourns

6:00 p.m. Reception/Dinner