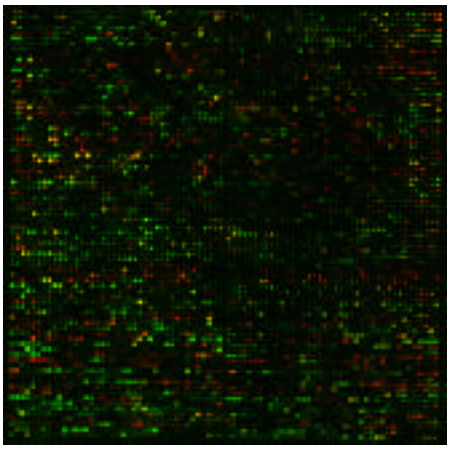
NASA Institute for Advanced Concepts

**Phase I - Final Report**
**NASA Institute for Advanced Concepts**
**NASA Prime Contract Number NAS5-98051**
**Subcontract Agreement Number 07600-073**

**January 7, 2002**

# A Flexible Architecture for Plant Functional Genomics in Space Environments



**Dr. Terri L. Lomax**
**Department of Botany and Plant Pathology**
**Oregon State University**

## ABSTRACT

Plants will play an essential role in providing life support for long-term space exploration and habitation. Our NIAC Phase I project explored a flexible architecture for measuring the response of plants to any unique space condition and for determining the function of genes to optimize plant performance under those conditions. The proposed architecture combines the recent and forthcoming rapid advances in the field of plant genomics with the power of homologous recombination to determine function via gene alterations or knock-outs. Discoveries made with this functional genomics approach will allow the engineering or selection of plants optimized to thrive in specific space environments. The proposed functional genomics architecture will advance NASA's mission of human exploration, use, and development of space. Concepts developed here will be useful both in the near- to mid-term on the International Space Station, in the far-term for longer missions (i.e. to Mars), and for eventual space habitation.

The novel combination of three technologies allowed us to validate the viability of the proposed concept and define the major feasibility issues during Phase I. Testing the feasibility of the proposed plant functional genomics architecture for space recently became possible with the availability of the first complete genome sequence for a plant (***Arabidopsis***), the development of microarray technology for measuring gene expression, and the demonstration of efficient homologous recombination in a model green land plant, a moss (***Physcomitrella***).

We used commercially available microarray chips and the gravitropic response of ***Arabidopsis*** to:

1. verify that microarray technology has enormous power to simultaneously measure plant genetic responses to a variety of environmental conditions, and

2. define major feasibility issues including reproducibility, sensitivity, developmental and environmental control, biocomputing requirements, and automation.

We also explored several mechanisms for determining gene function:

1. Homologous recombination in ***Physcomitrella*** provides a currently available mechanism for establishing or validating gene function; however, it remains unwieldy and time-consuming;

2. Novel antisense oligomers have the potential to rapidly establish the function for some genes, although feasibility issues regarding delivery and applicability remain;

3. Project 2010, a large international effort, aims to determine the function of every gene in ***Arabidopsis*** by 2010. In Phase II, we will assess the capacity for applying the results from the ***Arabidopsis*** 2010 project to those plant species most likely to be used in space environments.

As more plant genomes are sequenced and technologies for determining gene function are expanded to other species, the plant functional genomics architecture will optimize the performance of any plant in any space environment. Future additions to the architecture will include the technical advances necessary for remote collection and evaluation of data. The result will be a flexible module for measuring and optimizing plant responses that will play an integral, essential role in all manned NASA enterprises.

**TABLE OF CONTENTS**


| | Page |
|---|---|
| 1. Introduction …….. | 4 |
| 2. Advanced Concept Description …….. | 4 |
| 3. Monitoring Plant Genetic Responses ……... | 5 |
| 3.1. Current MicroarrayTechnology ……... | 6 |
| 3.1.1. Microarray Feasibility Issues …….. | 7 |
| 3.1.2. Potential of Microarray Technology …… | 10 |
| 3.2. Emerging Microarray Technologies …… | 10 |
| 4. Functional Genomics …….. | 11 |
| 4.1. Homologous Recombination …… | 11 |
| 4.1.1. Homologous Recombination Feasibility Issues …… | 13 |
| 4.1.2. Potential of Homologous Recombination for Gene Function Studies.. | 13 |
| 4.2. Antisense Oligomers …….. | 13 |
| 4.2.1. Antisense Oligomer Feasibility Issues …… | 15 |
| 4.2.2. Potential of the PMO Technology for Plant Gene Function Studies.. | 15 |
| 4.3. Project 2010 ……... | 15 |
| 5. Future Advanced Concept Directions – Phase II ……... | 16 |
| 5.1. Develop New Genetic Technologies ……... | 16 |
| 5.2. Define Important Genomes ……... | 17 |
| 5.3. Plant Genomic Environmental Monitoring and Control System…… | 17 |
| 5.4. Computation, Bioinformatics, and Automation …… | 17 |
| 6. Literature Cited …… | 19 |

## 1. INTRODUCTION

"To advance human exploration, use, and development of space"

Inherent within the short-term and long-term goals of NASA's strategic plan is the ever-increasing presence of humans as part of extended missions and as occupants of both the International Space Station (ISS) and planetary settlements. Included in the physical requirements of a prolonged human presence in space are the components necessary for manageable and affordable subsistence. As on Earth, the most efficient and self-sustaining system to provide food, oxygen, and raw materials is a plant-centered biosystem. Plants produce oxygen and food while eliminating carbon dioxide from the environment. Plant functions will be essential to generating a balanced, sustainable habitat in the closed environment of space travel and colonization.

Because plants will eventually be major contributors to life support systems in long term space flight and habitation, the effects of space environments on plants is an important area of research. One major question is: "How will novel space environments affect the life cycle and productivity of plants individually and within the larger context of a balanced ecosystem?" There are many aspects of enclosed environments that have already been considered. Some of these conditions, such as increased $CO_2$ levels and low temperature, can be simulated and tested on Earth. However, there are other factors and novel combinations of stimuli that cannot be recreated and yet may significantly affect individual plant performance and overall balance within the enclosed biological community. The goal of this Phase I NIAC project was to develop the concept for an architecture that will ultimately be capable of investigating plant responses to the space environment and optimizing plants to serve necessary functions within this setting.

## 2. ADVANCED CONCEPT DESCRIPTON

As NASA plans for missions of ever-increasing duration and seeks to establish safe and routine access to space in support of permanent human operations, plants will play an increasingly important role in providing life support or even artificial environments. In the space environment, plants will be faced with extreme conditions (e.g. high $CO_2$, low temperature) or even new conditions not found on Earth (e.g. microgravity, space radiation). In order to prepare for any space environment, it is important to understand how plants will respond to such challenges. Functional genomics is the generation and analysis of systematically produced information about what genes do (Brent, 2000). Such knowledge will allow us to optimize plant growth and performance under a variety of conditions.

The advanced concept underlying this Phase I project was the development of a flexible architecture for determining the identity and function of the genes underlying plant responses to any given stimulus or environmental condition (Fig. 1). The proposed plant functional genomics architecture for space is based on the availability of the first complete genome sequence for a plant (***Arabidopsis thaliana***), the development of microarray technology for simultaneously measuring the expression of thousands of genes, and the demonstration of efficient homologous recombination in a model green land plant, the moss ***Physcomitrella***, which provides the opportunity to knock-out or alter genes in order to investigate their function. DNA microarrays are designed to systematically analyze the entire complement of genes expressed for a specific organism. RNA isolated from test plants can be applied to corresponding microarrays to allow the identification of genes regulated by a specific environmental condition. Identification of the gene sequence from the databases used to generate the microarrays then allows the isolation of

a gene in order to more specifically determine its function via gene disruption. After functional analysis of plants containing the disrupted gene, targeted genetic engineering can be used to optimize performance of that species in response to that particular space environmental condition. Testing the flexibility of the proposed architecture comes both from the ability to analyze the response to any stimulus in any plant, and from the adaptability of using new information and technologies as they are developed. We are in an unprecedented growth phase in the accumulation of plant genomic knowledge and the proposed architecture will be designed such that it will change and adapt to best utilize that knowledge and technology growth.

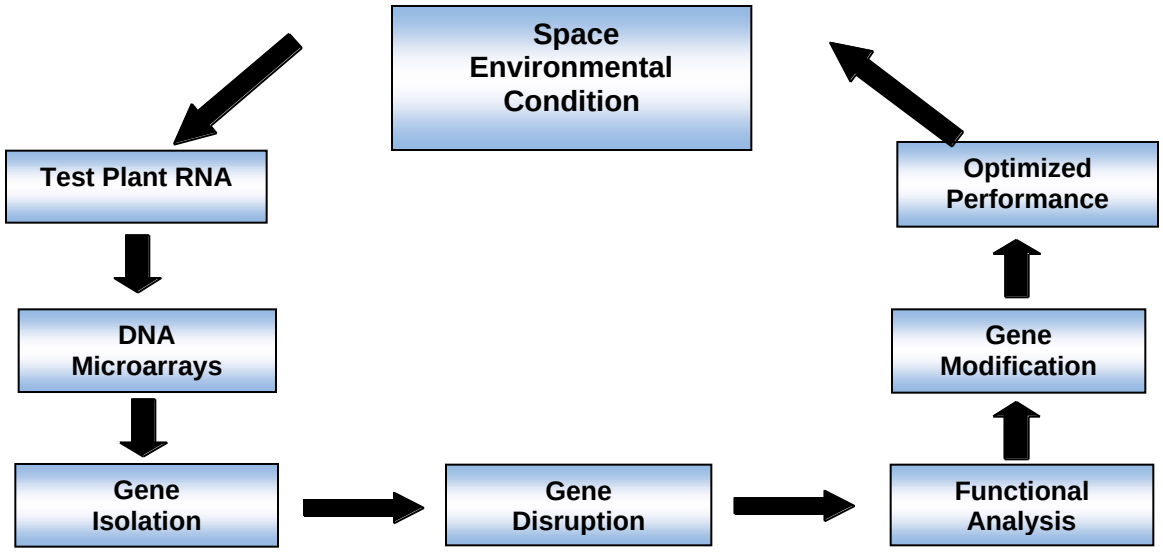



**Figure 1. The basic architecture for genomic analysis and manipulation.**

## 3. MONITORING PLANT GENETIC RESPONSES

Recently, scientists funded by the government, as well as several foundations and private companies, have completed the first complete genome sequence for a plant, ***Arabidopsis*** (see http://www.nature.com/nature/fow/001214.html). Additionally, governments and private industry are funding researchers to complete the sequencing of the entire genomes of several species of crop plants, such as rice, corn, tomato, and soybean. Expectations are that the entire DNA sequences of these organisms will be determined in the next few years, with many more plant genome sequences available in the more distant future.

The gene sequence information discovered to date provides a vast and diverse starting point for gene function research. However, in the absence of other information, a gene's sequence does not provide any clues to its function. Without understanding gene function, researchers have limited ability to translate genetic sequence information into a means of manipulating plant productivity and adaptability. An initial indication of the function of a gene is provided by determining when and where that gene is expressed, both during plant development and in response to varying environmental conditions. A new technology called microarrays now allows rapid analysis of the expression patterns of every gene in a genome.

In Phase I, we evaluated the feasiblility of using microarrays to monitor plant genetic responses to environmental conditions and to identify genes and pathways used in those

responses. Here we describe the microarray technology and provide compelling evidence of the necessity for using microarrays for sustaining human life in space. In addition, we outline efforts needed to further develop the microarray technology to optimize plant biosystems.

### 3.1. Current Microarray Technology

The relatively new microarray technology is made possible by genome sequencing projects. Robots use the resulting gene sequence data to mass-produce gene segments on membranes, silicon chips, or glass slides. The basic principle behind a microarray is that a piece of DNA will bind only to the exact complementary strand, i.e. DNA comes in matched sets so you can use one strand to find the other. In microarrays, from 1,000 to over 20,000 genes are arrayed in a defined pattern and used to identify the complementary strands of DNA previously synthesized from cell and tissue extracts of mRNA (the "expressed" portion of the gene that will result in the synthesis of a protein) that are "washed" over the chip. To detect the pairing, samples are tagged with fluorescent dyes that light up under a multi-laser scanner (Figure 2). Basic microarray data comes in the form of probe cell light intensities, measured when the microarray is scanned. Those intensities can then be compared between samples. Microarray technology thus allows investigators to determine the activity levels of genes. In addition, microarrays show which genes are "off" or to what extent they are "on" by indicating how much transcript is being made in a given tissue at a given time. The power of the microarray technology is that a single analysis can identify entire pathways that a plant uses to respond to an environmental stimulus.

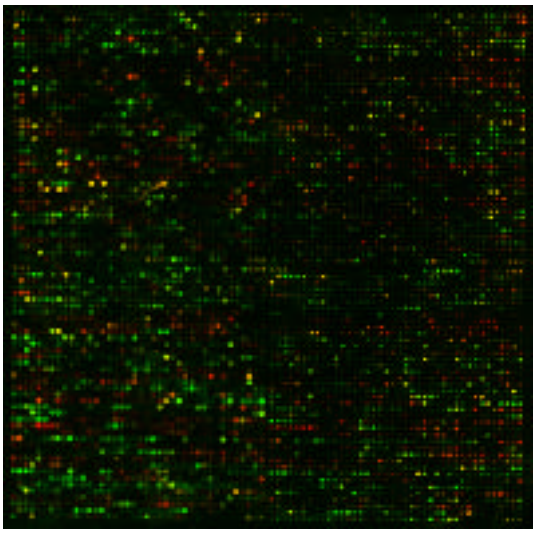


**Figure 2. Microarray analysis provides information on gene involvement in a process or pathway.** In this false color analysis of the comparison between different environmental conditions, each cell represents a different gene, with increases in gene activity represented by green, decreases in red and expression that is not changed in yellow.

Since the entire ***Arabidopsis*** genome sequence is now known, microarrays representing a large proportion of the genes in the genome are available. ***Arabidopsis*** is a useful model organism because it is genetically related to soybeans, cotton, vegetables, and oil seed crops. ***Arabidopsis*** is an efficient model organism because it has a short life cycle of seven weeks, a small genome, and yet performs all normal plant functions. ***Arabidopsis*** is well characterized for

many different physiological, biochemical, and molecular biological characteristics and can be effectively utilized to monitor biological responses and identify potential gene modifications that will enhance plant productivity.

In our Phase I study, DNA microarrays were employed to examine gene expression in ***Arabidopsis*** plants responding to gravistimulation. We were able to establish that over 100 genes in ***Arabidopsis*** significantly change their expression pattern in response to gravistimulation. Additional data was evaluated from recently published research reports. Although the microarray technology is only in its infancy, a picture of the type, sensitivity, and complexity of the data generated from DNA microarrays is beginning to emerge. That picture suggests that DNA microarrays and a functional genomics approach to physiological assessment is both feasible and necessary to meet the demands of a plant-based life support system.

3.1.1***.*** Microarray Feasibility Issues***.*** The current feasibility of examining a physiological system by interpreting DNA microarray data was examined in our Phase I study using gravistimulated ***Arabidopsis*** seedlings and commercially available ***Arabidopsis*** genomic DNA microarrays from Affymetrix (http://www.affymetrix.com). Affymetrix microarrays employ silicon chip technology where each gene is represented by a complex of oligonucleotides. Each current-generation Affymetrix "chip" contains over 8,000 ***Arabidopsis*** genes. Chips with the entire Arabidopsis genome (ca. 25,000 genes) are expected to be available within the next six months. We chose the Affymetrix silicon chip technology for evaluation because many more genes can be arrayed in a much smaller space using their patented photolithography methodology compared to the other major arraying approach of spotting DNA onto glass slides. In addition, silicon chips are smaller and lighter than glass slides which are important considerations for NASA missions.

***Reproducibility.*** In Phase I, we completed our initial microarray expression analysis utilizing data generated in three experiments. In each experiment, gene expression in gravistimulated ***Arabidopsis*** seedlings was compared to that observed in control plants. The microarray allowed the simultaneous screening of the plants' total RNA (the expressed portion of the genome) for the potential expression of 8,297 different ***Arabidopsis*** genes. This current-generation Affymetrix chip represents approximately one-third of the complete genome complement. The goals of these initial analyses were:

1. to establish a set of indicators that will evaluate overall microarray performance and determine feasibility issues;

2. to demonstrate that microarray analysis can reveal changes in plant genomic function in response to an environmental change relevant to space, in this case, gravity; and

3. to compare the genes that have been identified as changing expression in ***Arabidopsis*** with the genes in the ***Physcomitrella patens*** genome to further test the functional genomics architecture (see below).

To accomplish these purposes, both absolute and comparative analyses of the ***Arabidopsis*** microarray data were performed. RNA transcripts that hybridize to the microarray probes are labeled with fluorescence and read by a scanner. Absolute analysis utilizes the light intensities measured for all relevant probe cells to determine if a particular transcript is present within a sample. Comparative analysis is used to determine the relative change in abundance for each transcript between a baseline (e.g. a control) and an experimental sample (in this case, gravistimulated ***Arabidopsis*** seedlings). Both types of data analysis were performed using the ***Data Mining Tool***, an Affymetrix software program.

***Indicators of Microarray performance.*** Six indicators of microarray performance were defined and evaluated by applying them to the data generated. Specific indicators include:

1. Overall level of transcription activity

2. Level of background signal intensity and noise observed

3. Ratios of sample binding to different ends of DNA sequences representing internal controls and constituitively expressed genes

4. Consistency of absolute calls between pairs of similar samples (e.g. controls)

5. Consistency of difference calls between matched control-experimental pairs

6. Incidence of false changes encountered when comparative analysis is applied to pairs of similar samples (e.g. two controls)

In each experiment, gravistimulation initiated a significantly altered expression pattern for a large number of genes. However, the profile of gene expression that changed was different for each experiment. The microarray performance indicators described above were utilized to identify the source of the unexpected variation. In each case, the microarray was found to have functioned properly. In one experiment, the variation with regard to expected results was related to differences in RNA extraction techniques or with the subsequent ***in vitro*** transcription reactions. In the remaining two experiments, the cause of the variation was biological in origin and possibly due to the effects of circadian rhythms derived from different sampling times of the tissue (see below). These results suggest that replications and repetitions of microarray experiments may be necessary to ensure that measured changes in gene expression are the consequence of a particular change in external stimuli. Statistical studies indicate that designing microarray experiments with at least three replicates will greatly reduce misclassification rates (Lee et al., 2000).

***Sensitivity***. The ability of microarrays to reveal gene transcripts expressed at very low levels is essential. Greatest confidence in microarray results can be achieved when a large fold change in highly expressed genes is observed. Unfortunately, essential regulatory proteins often occur in low quantities, requiring only low levels of gene expression. For example, the regulatory proteins involved in a plant's gravitropic response are not known and could not be specifically identified in this Phase I research. Confident detection of changes in these low level transcripts requires additional replication of the experiment and the use of multiple microarrays. With multiple microarray data sets, real changes in gene expression can be separated from apparent changes in expression that reflect random biological variation. Nevertheless, the ability of DNA microarrays to detect changes in low level expression signals within the context of background signal and noise remains an important feasibility issue.

***Environmental and Developmental Control***. There is growing use of DNA microarrays to assess changes in global gene expression patterns in plants responding to a wide variety of stimuli. Utilizing the data from the microarrays that met the desired performance standards, we identified genes that appeared to have either increased or decreased expression in response to gravistimulation. Of the 8,297 transcripts probed on the microarray, 57 were increased and 42 were decreased in the gravistimulated seedlings compared to controls. Data filters were developed to screen and rank this set of transcripts according to level of expression and

magnitude of expression change. Twenty-six transcripts appeared to have significantly increased expression in the gravistimulated seedlings compared to controls. Nine of these showed greater than a two-fold increase in expression. Fifteen transcripts appeared to have significantly decreased expression in the gravistimulated seedlings. While this analysis is based on only a few experiments, the transcripts identified code for an intriguing set of proteins that includes hormone receptors and key signal transduction components.

As part of our feasibility assessment, we also performed an extensive literature search and evaluated other studies that have used DNA microarray analysis to determine entire genome expression patterns (transcriptome) for plant responses to other environmental conditions: salt stress (Bohnert et al., 2001; Kawaski et al., 2001), oxidative stress (Desikan et al., 2001), growth in darkness (Desprez et al., 1998), iron deficiency (Thimm et al., 2001), and exposure to inducers of systemic acquired resistance (Maleck et al., 2000; Peterson et al., 2000). Generally, the number of genes observed to have an altered expression level ranged from 100 to 200; however, some stimuli induced much greater increases. For example, following three days of iron deficiency stress, ***Arabidopsis*** shoots demonstrated a significant increase in the expression of 2,240 genes (Thimm et al., 2001). The large number of genes that changed expression in response to a single environmental condition reflects the inherent complexity of the underlying biological system.

The complexity of plant genetic responses is further revealed in the variations of gene expression that occur with plant tissue type, with developmental stage, and with diurnal and circadian rhythms. DNA microarrays are beginning to provide a quantitative measure of these normal variations. For example, of the 6,052 genes found to be expressed in five-week-old ***Arabidopsis*** roots, 426 were found to be expressed exclusively in the root while the remainder were simultaneously expressed in other tissues: 4,995 in the leaves, 5,036 in the stems, and 5,243 in the flowers (Zhu et al., 2001). In another ***Arabidopsis*** microarray study, the expression of 433 genes (of 1,443 genes identified) were found to be at least two-fold higher in leaves compared to roots (Ruan et al., 1998). A study of 7,800 ***Arabidopsis*** genes via DNA microarrays revealed that 816 genes have at least a two-fold change in expression level associated with a diurnal rhythm (Schaffer et al., 2001). These normal variations in gene expression have important implications for tissue sampling protocols and for the ultimate interpretation of data as a means to assess physiological function.

Through our Phase I analysis, we have determined that tissue sampling must be performed under strictly defined conditions. As highlighted by microarray studies from us and others, gene expression may be influenced by a wide variety of environmental stimuli and also by biological factors related to tissue type, development, and circadian rhythms. When biological factors are not taken into account, apparent inconsistencies in microarray analyses will result.

***Requirements for Biocomputing Capacity***. Ultimately, the major challenge may be using gene expression data to determine the status of physiological processes. Consider that ***Arabidopsis*** has a background of 15,000 expressed genes of which perhaps 300 will have altered expression when responding to a specific change in the environment. Perhaps 2,000 Ara***bidopsis*** genes will change with a diurnal rhythm. Plants growing in space will be exposed to many environmental stimuli that will inevitably lead to complex patterns of gene expression. The products of gene expression in space may interact in many ways because physiological systems are interlinked and complex. DNA microarrays and functional genomics produce large quantities of data that will provide invaluable insights into physiological systems. A functional genomics approach to physiological monitoring in plants will use supercomputers to manage the microarray data and model the complexity of physiological systems. Efforts to develop needed

computational capacities are underway (see Future Advanced Concept Directions). The trend of funding and research in both the public and private sector is to develop biocomputing so that microarrays can be used on Earth for sustainable agricultural and forest systems. The space program should capitalize on the research investment underway. Ultimately, the issues of life support and sustainability in space are not different from those on Earth.

***Automation.*** Microarray analysis for monitoring plant environmental responses in space can be accomplished through automation. Robotics will be used to standardize sampling protocols, processing, and data analysis. Information from microarray analysis can be communicated via telemetry from manned and unmanned settings in space. Information from monitoring plant genome functions with microarray analysis can be used to change environmental conditions needed to optimize plant growth and other life support functions.

3.1.2. Potential of Microarray Technology. Microarrays have the potential to become a vital tool to survey the activity of a plant's entire genome in response to a given stimulus. Stimuli encountered in a space environment can result in the altered expression of hundreds of genes within a plant. Changes in plant gene expression profiles must be understood within the context of normal variations of gene expression that occur with tissue development and circadian rhythms.

As evidenced by the large number of genes involved, the underlying physiological processes that govern a plant's response to external stimuli are extraordinarily complex. The sensitivity of DNA microarray analysis will increase over time such that it can detect changes in the expression of even low-abundance regulatory protein genes involved in these complex physiological processes. DNA microarrays make it feasible to simultaneously assess the expression levels of hundreds to thousands of genes that may be involved in a particular physiological response. Microarrays are the only technology currently on the horizon that will provide this kind of detailed information.

### 3.2. Emerging Microarray Technologies

Presently, about 75% of microarray research involves broad gene expression profiling using high-density arrays to screen cellular messenger RNA samples across as many polynucleotide probes as possible. Affymetrix (http://www.affymetrix.com) currently holds about 80% of that market due to its photolithographic technology, which is similar to that used by the semiconductor industry. Photolithography enables Affymetrix to make the highest density DNA microarrays available (Bouchie, 2002). Affymetrix can currently build over 400,000 features (multiple copies of a probe) on a 1.64 $cm^2$ chip, but a chip bearing over 40 million features is theoretically possible. A competing firm, Agilent (http://www.chem.agilent.com), plans to use ink jet technology to generate microarrays of about 100,000 features per 1.64 $cm^2$. Affymetrix is the only firm that can place entire plant genomes on a single chip for the foreseeable future (Bouchie, 2002).

Much of the future growth in the microarray sector is expected to be in low-density microarrays. Once a relatively small number of genes has been associated with a disease or physiological process (the goal of current research using high-density arrays), low-density arrays can be used to screen many patients or plants, for example, for diagnostic and/or research purpose. Associating genes with disease or environmental responses is currently at a bottleneck, and many more population genetic studies and improved informatics algorithms are needed before the low-density technology is useful (Smith, 2001). However, in the 10-40 year NIAC timeframe, low-density chips optimized for the specific species and environmental challenges

associated with plant-based life-support systems may provide optimal and cost effective day-to-day biomonitoring.

One promising technology for low-density chips targeted to specific assays is the electrochemical technology currently being used by CombiMatrix (http://www.combimatrix.com), and recently licensed by NASA Ames Research Center to conduct research in both terrestrial laboratories and in space. With this technology, scientists can carry out experiments on Earth, then upload the data to the Space Station where the experiment could be duplicated using CombiMatrix's technology, thus surmounting the space and weight restrictions imposed by the ISS. While the primary use at this time is planned to be for the NASA biotechnology space initiative with a goal of understanding how human DNA behaves in space, we plan to work with Dr. Viktor Stolc at NASA Ames to determine how the technology can be adapted to monitor plant genetic responses. In the near term, NASA scientists hope to use the CombiMatrix technology to monitor astronaut health on the International Space Station. The expected long-term outcome is that NASA scientists will be able to use CombiMatrix's system in space to design and produce customized biochips, conduct experiments, analyze the results, and produce additional biochips incorporating modified test parameters, without having to return to Earth (http://www.combimatrix.com/press.shtml#).

## 4. FUNCTIONAL GENOMICS

While DNA microarray analysis can provide important information on the potential involvement of a gene in a given process or pathway, other methods are needed to confirm the function of the gene (Brent, 2000; Hohn and Puchta, 1999). Knowing a gene's function and how that function is interconnected with the function of other genes is necessary before the determination of gene expression can provide a basis for physiological assessment. The progression of understanding required can be thought of as a four step process: 1) identify all the genes in the genome and their pattern of expression; 2) determine the function of the protein encoded by each gene; 3) determine how multiple genes and their products are involved in a particular physiological process; and 4) examine how multiple genes and their associated physiological systems interact and modulate each other to dictate the organism's overall physiological response. Although we are only beginning to develop the genetic concepts needed for using plants in space, the rapid pace of current research and discovery ensures that a functional genomics approach to plant physiological assessment will be feasible within the next 10-40 years.

### 4.1. Homologous Recombination

The type of physiological assessment needed for a plant-based life support biosystem is not currently feasible using DNA microarray data alone because the function of over one-half of the genes in plant genomes is unknown. Such knowledge will likely be available in the next 10-40 years (see below). In the meantime, it appears necessary to devise other means of determining gene function. Therefore, in Phase I we proposed using homologous recombination to test the feasibility of applying genomic information to determine and ultimately alter gene function.

The elucidation of gene function has most powerfully been carried out by analysis of gene malfunction. Although it is not currently feasible to selectively knock-out a specific gene in most plants, selective gene knock-out by the process of homologous recombination is possible in the moss, ***Physcomitrella patens*** (Puchta and Hohn, 1996). If a given ***Physcomitrella*** gene is very

similar to one in higher plants, it is possible to gain clues about the function of the higher plant gene by knocking-out the similar gene in moss and observing the resultant moss phenotype. In contrast to the inefficiency of homologous recombination in all other plants, targeting efficiencies over 90% have been reported for ***Physcomitrella*** (Schaefer and Zÿrd, 1997, Figure 3A). Recent experiments have used homologous recombination in ***Physcomitrella*** to elucidate the biological function of two previously unknown ***Physcomitrella*** genes, an acyl lipid desaturase (Girke et al., 1998) and a protein involved in organelle division, FtsZ (Strepp et al., 1998). Figure 3B illustrates the vector design for homologous gene recombination and gene knock-out.

Recent sequence data analyses reveal a high degree of homology between ***Physcomitrella*** and seed plants. The moss is more closely related to dicots such as ***Arabidopsis*** than the dicots are related to monocots like rice (Reski, 1998a). The genome size of ***Physcomitrella*** is about three times that of ***Arabidopsis***. A variety of genes have now been cloned from ***Physcomitrella*** and they are remarkably homologous to their cognate higher plant genes (Reski, 1998a,b). In addition, many fundamental physiological and genetic processes can be studied in ***Physcomitrella*** (Cove, 1992; Reski, 1998a). There are no significant differences in promoter or codon usage between ***Physcomitrella*** and ***Arabidopsis***. The simplicity of the gene knock-out technology using homologous recombination allows developmental analysis to be defined at the cellular level.

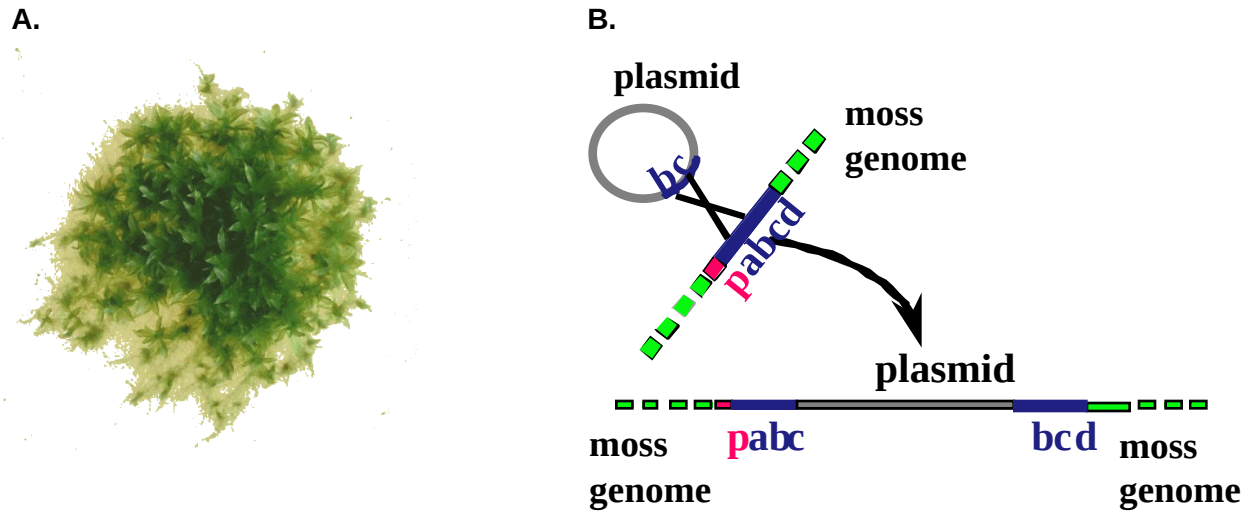



**Figure 3. The moss species used in gene knock-out technology, *Physcomitrella* (A) and the gene knock-out scheme used in homologous recombination (B).** Insertion of the plasmid into the target gene using surrounding homologous sequences destroys gene function.

Importantly, homologous recombination in ***Physcomitrella*** was recently used to determine the cellular function of a gene isolated from ***Arabidopsis***. Girod et al. (1999) found that generating disruptions in the gene for a subunit of the 26S proteosome in ***Physcomitrella*** resulted in developmental arrest and formation of abnormal vegetative tissue that was unable to form buds or gametophores. Complementation of the phenotype with the plant hormones auxin and cytokinin restored bud formation and partial development of gametophores. Thus, the homologous recombination studies in ***Physcomitrella*** demonstrated previously unknown functions for this protein, and the role of the ubiquitin/26S proteosome pathway in plant developmental processes triggered by hormones.

4.1.1. Homologous Recombination Feasibility Issues. During the Phase I study we examined the feasibility of the gene knock-out technique using ***Physcomitrella***. The normal response of ***Physcomitrella*** to gravity was observed. Genes that change expression in gravistimulated ***Arabidopsis*** plants were used to screen a ***Physcomitrella*** database to identify similar genes. ***Physcomitrella*** protoplasts were isolated and successfully transformed with plasmid vectors. The next step in the analysis would be to use the plasmid vectors to insert a DNA sequence into the middle of a moss gene homologous to one shown to be involved in ***Arabidopsis*** gravitropism by microarrays, disrupt its function, and observe the resultant ***Physcomitrella*** for alterations in its normal gravitropic response. However, a complete experiment to knock-out a ***Physcomitrella*** gene was not possible during the 6-month Phase I study, both due to the time-consuming nature of the technique and to the incomplete status of the ***Physcomitrella*** sequencing project, which made it difficult to find target genes.

A potential drawback to the use of ***Physcomitrella*** for gene function discovery is that some genes identified in higher plants will not have homologous counterparts in the moss genome even once it is completed. Also, the function served by a gene in moss may not be identical to the function served by the homologous gene in a vascular plant. For example, the gravitropic response of the moss takes place in a filament of single cells, whereas gravitropism in higher plants requires the coordinated response of an entire multicellular tissue. Nevertheless, the approach of using homologous recombination does appear to be feasible, and to have potential for providing valuable clues about gene function for some genes.

4.1.2. Potential of Homologous Recombination for Gene Function Studies. Homologous recombination, while efficient in ***Physcomitrella*** and important to the ability to currently assess gene function, is relatively unwieldy, time-consuming, and can only be applied to a select, single gene. The technology appears better suited to the analysis of single genes rather than genomic scale studies. Therefore, we turned our attention to more rapid technologies and those that appeared to have the potential to contribute to genome-wide understanding in the 10-40 year NIAC timeline.

### 4.2. Antisense Oligomers

We further defined gene function within model plant systems by testing a new class of antisense oligonucleotides, phosphorodiamidate morpholino oligomers (PMO, Figure 4). Antisense oligomers have been demonstrated to be a valuable research tool because they cause dysfunction of specific, targeted genes (Ghosh and Iversen, 2000). Antisense PMOs are uncharged molecules containing a six-membered morpholino ring that replaces the sugar moiety of the nucleic acid, and a nonionic phosphorodiamidate linkage that replaces the phosphodiester linkage of the helix backbone. The unique structure of PMOs prevents enzymatic degradation of the molecules while still permitting their solubility in water. Antisense oligomers prevent gene function either by inhibiting protein translation via steric blockages of the ribosome protein complex onto the mRNA strand (Ghosh et al., 2000) or by interference with pre-mRNA splicing (Giles et al., 1999). As a rapid method to alter gene expression patterns and elucidate gene function, antisense technology relies upon: (1) hybridization of the nucleic acid oligomer to target mRNA sequences, (2) sufficient biological stability of the antisense oligomer, and (3) the capacity of the oligomer:mRNA complex to inhibit gene expression. Phosphorodiamidate morpholino oligomers (PMO) contain the same basic structure as naturally occurring nucleic acids and will bind to mRNA by Watson-Crick base-pairing (Ghosh et al., 1999). Due to altered linkages between the components of the oligomer backbone, these short fragments (ca. 20 bases long)

are resistant to nuclease digestion within the cell and do not serve as a substrate for RNase H, an enzyme responsible for destroying mRNA when duplexed to DNA strands (Stein et al., 1997).

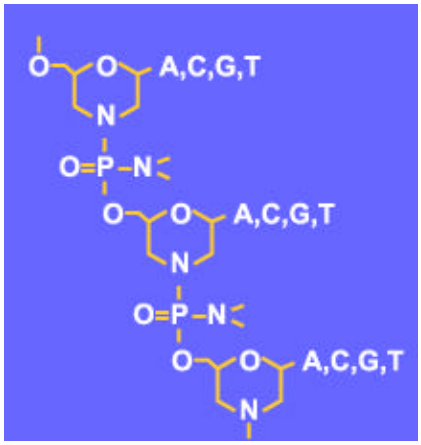



**Figure 4. The structure of the phosphorodiamidate morpholino oligomer (PMO) tested in Phase I studies.**

During Phase I, we were able to arrange to test this new type of antisense technology on a plant system for the first time. In a cooperative effort between AVI BioPharma, Inc. (Corvallis, OR) and our laboratory, a novel antisense oligomer was tested with moss protoplasts to determine uptake and detection feasibility. Since PMOs had not previously been used with plant cells, the first question was whether the oligomers could be efficiently delivered to plant cells. We began with protoplast because their cell wall had already been removed and we had already found them to be amenable to the uptake of nucleic acids in transformation assays. A non-targeted PMO covalently linked to fluorescein (a compound used for detection) was transformed into moss protoplasts and evaluated against untransformed, control protoplasts (Figure 5). A distinct punctate green signal was detected only in those protoplasts transformed with the PMO. Untransformed protoplasts exhibited only the background fluorescence of chloroplasts. The results from our preliminary experiments indicated that the oligomer could be introduced into plant cells and then detected by illumination under appropriate wavelengths. A clear benefit of this approach is how quickly results can be generated from each experiment. Protoplasts were incubated for only one day before cells were evaluated for the presence of the tagged oligomer. Homologous recombination experiments with moss protoplasts require regeneration for several weeks and additional manipulation of plant material and DNA before confirmation of gene expression can be attempted.

**A.** **B.**

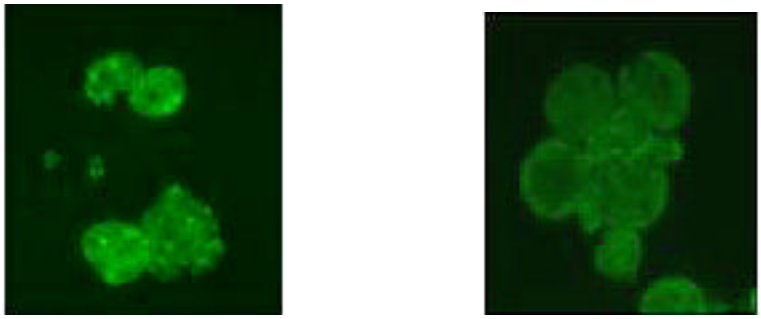


**Figure 5. Moss protoplasts transformed with antisense PMO (A) and untransformed controls (B).**

4.2.1. Antisense Oligomer Feasibility Issues – To extend the use of PMOs to plant systems besides protoplasts, feasibility studies must include the delivery of the oligomer into intact plant cells, which are surrounded by a tough cell wall not present around animal cells. Therefore, the delivery of the oligomer into plant cells by more vigorous means will include the use of a biolistic delivery system. Additionally, AVI BioPharma is working in conjunction with our laboratory to generate plant-specific targeted PMOs containing several different indicators that will permit our investigation of the efficiency and effectiveness of the delivery system for specific gene suppression, for the identification of specific protein expression levels, and for the continued presence of the targeted PMO within plant tissue.

4.2.2. Potential of the PMO Antisense Technology for Plant Gene Function Studies - As mentioned earlier, the experiments performed by our laboratory in this Phase I study are the first use of this type of antisense molecule with a plant system. However, the application and efficiency of the PMO technology within human and animal systems has been established as a valuable research tool that indicated gene function (Arora et al., 2000) and altered viral pathology (Stein et al., 2001). Other applications that have been demonstrated for this novel antisense approach include as a pharmacological delivery system in cancer immunotherapy, as a therapeutic treatment to reduce coronary restenosis, as a therapeutic agent to increase the effectiveness of other drugs by preventing their degradation by a key liver enzyme, and as a means to prevent or control viral infection and disease. Clearly, its effectiveness in suppressing gene expression in human and animal systems is now established and we are optimistic that further studies during our Phase II study will demonstrate its utility for plants as well.

### 4.3. The *Arabidopsis* 2010 Project

A promising resource that is currently being developed for the study of functional genomics is the Multinational Coordinated ***Arabidopsis*** 2010 Project. To fully take advantage of the wealth of information provided by the ***Arabidopsis*** genome sequencing effort, plant biologists have proposed an important and revolutionary new initiative: to determine the function of all genes within this reference species and to place genes within their cellular, organismal, and evolutionary context by the year 2010 (http://www.arabidopsis.org/workshop.html). The ***Arabidopsis*** 2010 project involves the collaborative efforts of an international community of research scientists. Each group will select a set of genes as the subject of their research. All available means of creative and innovative research are proposed to determine the function of this network of genes within ***Arabidopsis***. In the decades that follow, the analysis will be expanded to other species. At the same time, research will examine how genes function together to define a plant's physiological processes. The prospect of having such complete functional genomics information provided concurrent with the continued design of a flexible, modular architecture illustrates the adaptability of our advanced concept to utilize information and new technology as it becomes available. As technology and information becomes more accessible, such as increased bioinformatics and knowledge about the interaction of molecular signal cascades, these components can be added to the cyclic architecture to better design the next adaptative strategies. In Phase II, we plan to monitor the progress of the 2010 project and forecast the state of knowledge and availability of technologies that will result from it, and investigate their benefit to NASA strategic enterprises over the next 10-40 years.

## 5. FUTURE ADVANCED CONCEPT DIRECTIONS – PHASE II

The field of molecular biology is experiencing an explosive degree of research and discovery. The continued development of techniques, combined with our growing understanding of physiology, will make a genomics-based physiological monitoring and control system feasible within a ten to forty year time frame. Much of the key research required is already underway. In the next few decades, several developments are anticipated.

### 5.1. Development of New Genetic Technologies

New methods and technologies are under development to manipulate the expression of genes and facilitate their transfer between organisms. Our Phase I study explored the use of microarrays and functional genomics for space applications, and identified several areas that will benefit from further feasibility studies. There is a need to determine the precision, sensitivity, and consistency of gene expression profiles generated by DNA microarrays. More replications of microarray analysis are needed to define the sources of variation in plant genomes. Many reports in the literature are based on three or fewer microarray experiments. Our preliminary experience with microarrays indicates that three replications are the minimum required to attribute apparent changes in genome function detected by microarrays to environmental changes.

There is a need to explore the capability of microarrays to detect changes in low-level transcripts. Such transcripts may encode important regulatory proteins. Affymetrix, in November 2001, provided new microarray analysis software, which is more statistically based, and may facilitate the distinction of low-level expression signal from background noise. If microarray data is going to be used to gain a picture of a physiological system, it is important to know what limits exist within this technology.

There is a need to establish gene expression profiles for plant responses to specific environmental conditions. How many genes must be monitored to define a particular response? How much overlap and redundancy are there in the expression profiles observed following different environmental stimuli?

There is a need to continue to explore methods to determine gene function in plants. The use of gene knock-out by homologous recombination in the moss, ***Physcomitrella patens,*** will not be further developed, but the potential of novel antisense PMO technology to rapidly test gene function will be explored. At the same time, we will monitor the progress of the ***Arabidopsis*** 2010 project and assess how the information emerging from that effort will apply to plants used in space life support systems. Ultimately, a variety of techniques will be needed in order to piece together how the functions of different genes interact to create physiological systems.

Finally, there is a need to assess the latest developments in automation and robotics, to determine both the feasibility of performing the required tasks automatically, and the physical characteristics such an automated instrument would have. Can a self-contained functional-genomics based monitor be made of sufficiently small weight and size to be practical for a space application?

### 5.2. Definition of Important Plant Genomes

Large-scale efforts are currently underway to generate sequence information for the complete genomes of crop plant species that will be useful for life support in space. The initial focus was to sequence the entire genome of the model plant ***Arabidopsis thaliana*** and the

Directorate for Biological Sciences (BIO) of the National Science Foundation recently announced its intention to support research to determine the function of all genes in ***Arabidopsis*** by the year 2010 (the 2010 project). Information gained from the 2010 project will provide invaluable insight into functional genomics that extends beyond ***Arabidopsis***, to all plant species. Sequencing efforts are also underway for other species, including tomato, soybeans, and maize. Significant work has already been done by NASA and its collaborators to identify the plant species that will be most feasible and beneficial to include in life support systems for long-range missions. As part of Phase II, we plan to determine the extent to which those species will be included in future genome sequencing and functional genomic projects and to advise NASA on important research directions to that end.

### 5.3. Plant Genomic Environmental Monitoring and Control System

Phase I of our project has established the capacity for using plant genomes for monitoring environmental factors. The value of such monitoring will be to use genetic information from plants to not only provide a record of conditions during their growth, but also to indicate changes in environment necessary to optimize plant growth and life support functions. For example, if plant gene expression changes indicate that temperatures are too low for optimal growth, the information determined by automated microarray analysis will be sent by telemetry to an environmental control system that will increase temperature. The environmental adjustments made following the analysis of plant genetic responses can include temperature, humidity, nutrient supply, photoperiod, light level, and concentrations of atmospheric gases.

The results of our Phase I project also reveal a new potential for using plant genetic responses to interact with environmental factors. Analysis of plant gene expression will be used to do much more than simply adjust environmental conditions to optimize plant growth. Life support for humans in space will require plant growth not only optimized for food production, but for an array of other functions. Life support functions provided by plants include controlling humidity via transpiration, balancing carbon dioxide and oxygen concentrations with the processes of photosynthesis and respiration, providing shade with leaf structure, and recycling water and solid wastes. Genomic analysis with microarray analysis can reveal the nature of the ***in situ*** plant/environment relationship, and provide a basis for making decisions about changing environmental conditions for life support needs other than food production. For example, the environmental conditions optimal for food production may not be the same as conditions needed to increase oxygen concentrations in the atmosphere. Plant genomic analysis with microarrays could indicate just how environmental factors should be adjusted to increase emissions of oxygen by plants.

Of course, the life support functions provided by plants will never be a single function, but a balance of many roles. Our Phase I studies show that plant gene response analysis with microarrays provide a key mechanism for adjusting environments to solve the multiple life support needs of all life forms in space.

### 5.4. Computation, Bioinformatics, and Automation

Application of the computational power of supercomputers to the modeling of complex physiological processes is expanding rapidly. On October 16, 2001, NASA announced a collaborative partnership between NASA's Center for Computational Astrobiology and Fundamental Biology (NCCAFB), and Stanford's Center for Biomedical Computation (CBMC) to conduct multi-disciplinary research and development in the emerging field of computational

biology. As stated in the NASA news release, “Computational biology has become indispensable in modern biology because it lets scientists gather, store, and analyze vast amounts of data obtained from gene sequencing, the use of microarrays, and the study of proteins and cell physiology.” The collaboration will employ NASA’s state-of-the-art supercomputers and will initially focus on the integration of diverse databases and the simulation of physical models. While their initial intention is not to work with plants, they will be developing the tools that could one day be used to implement the genomics-based physiological monitor for a plant centered life-support system. As part of Phase II, we plan to work with the partnership to explore ways that their technology can be applied to plant genetic response systems.

In Phase II, we also plan to collaborate with a recently formed consortium between the Oregon University System, Oregon Health Sciences University, and the Department of Energy’s Pacific Northwest National Laboratory. Through the collaboration, advanced computer data mining programs will be developed that allow scientists to examine the information produced by gene expression monitoring and protein analysis and to integrate biological response mechanisms.

Much of the development in biocomputing power will take place in the private sector (e.g. Intel is developing a major effort in biocomputing). We will assess these capabilities by attending industry conferences, such as Biochip 2002, and visiting especially promising firms to initiate collaborations on visioning the progression and uses of biocomputing over the next 10-40 years that will prove useful to our proposed concept

Automation of many of the steps involved with sampling and processing tissue for RNA extraction and for performing analyses with DNA microarrays will be close at hand. Automation is necessary in space as those who depend upon plants for life support may not be biologists. Robotic collection of plant material for analysis, mechanical processing of samples, and computerized data analysis will constitute an essential part of the plant biosystem genetic monitoring and control system.

## 6. LITERATURE CITED


Arora, V., Knapp, D.C., Smith, B.L., Statdfield, M.L., Stein, D.A., Reddy, M.T., Weller, D.D., and Iverson, P.L. 2000. c-Myc antisense limits rat liver regeneration and indicates role for c-Myc in regulating cytochrome P-450 3A activity. J. Pharm. Exp. Ther. 292:921-928.

Bohnert, H., et al. 2001. A genomics approach towards salt stress tolerance. Plant Physiol. Bioch. 39:295-311.

Bouchie, A. 2002. Shift anticipated in DNA microarray market. Nature Biotechnology 20:8.

Brent, R. 2000. Genomic biology. Cell 100:169-183.

Cove, D.J. 1992. Regulation of development in the moss, ***Physcomitrella patens***. In: Brody, S., Cove, D.J., Ottolenghi, S., and Russo, V.E.A. (eds) Developmental Biology: A Molecular Approach, pp. 179-193. Springer Verlag, Heidelberg.

Desikan, R., Mackerness, S., Hancock, J., and Neill, S.J. 2001. Regulation of the ***Arabidopsis*** transcriptome by oxidative stress. Plant Physiol. 127:159-172.

Desprez, T., Amselem, J., Caboche, M., and Hofte, H. 1998. Differential gene expression in ***Arabidopsis*** monitored using cDNA arrays. Plant J. 14:643-652.

Ghosh, C. and Iversen, P.L. 2000. Intracellular delivery strategies for antisense phosphorodiamidate morpholino oligomers. Antisense Nuc. Acid Drug Devel. 10:263-274.

Ghosh, C., Stein, D., Weller, D., and Iversen, P.L. 2000. Evaluation of antisense mechanisms of actions. Methods Enzymol. 313:135-143.

Giles, R.V., Spiller, D.G., Clarke, R.E., and Tidd, D.M. 1999. Antisense morpholino oligonucleotide analog induces missplicing of ***c-myc*** mRNA. Antisense Nuc. Acid Drug Devel. 9:213-220.

Girke, T., Schmidt, H., Zähringer, U., Reski, R., and Heinz, E. 1998. Identification of a novel 6-acyl-group desaturase by targeted gene disruption in ***Physcomitrella patens***. Plant J. 15:39-48.

Girod, P-A., Fu, H., Zÿrd, J-P., and Vierstra, R.D. 1999. Multiubiquitin chain binding subunit MCB1 (RPN10) of the 26S proteosome is essential for developmental progression in ***Physcomitrella patens***. Plant Cell 11:1457-1471.

Hohn, B. and Puchta, H. 1999. Gene therapy in plants. Proc. Natl. Acad. Sci. USA 96:8321-8323.

Kawaski, S., Borchert, C., and Deyholos, M. 2001. Gene expression profiles during the initial phase of salt stress in rice. Plant Cell 13:889-905.

Lee, M-L.T., Kuo, F.C., Whitmore, G.A., and Sklar, J. 2000. Importance of replication in microarray gene expression studies: Statistical methods and evidence from repetitive cDNA hybridizations. Proc. Natl. Acad. Sci. USA 97:9834-9839.

Maleck, K., Levine, A., Eulgem, T., Moragn, A., Schmid, J., Lawton, K.A., Dangl, J.L., and Dietricj, R.A. 2000. The transcriptome of ***Arabidopsis thaliana*** during systemic acquired resistance. Nat. Genet. 264:403-410.

Peterson, M., et al. 2000. ***Arabidopsis*** MAP kinase 4 negatively regulates systemic acquired resistance. Cell 103:1111-1120.

Puchta, H. and Hohn, B. 1996. From centimorgans to base pairs - homologous recombination in plants. Trends Plant Sci. 1:340-348.

Reski, R. 1998a. ***Physcomitrella*** and ***Arabidopsis***: The David and Goliath of reverse genetics. Trends Plant Sci. 3:209-210.

Reski, R. 1998b. Moss (***Physcomitrella patens***) expressed sequence tags include several sequences which are novel for plants. Bot. Acta 111:145-151.

Ruan, Y., Gilmore, J., and Conner, T. 1998. Towards ***Arabidopsis*** genome analysis: Monitoring expression profiles of 1400 genes using cDNA microarrays. Plant J. 15:821-833.

Schaefer, D.G. and Zÿrd, J-P. 1997. Efficient gene targeting in the moss ***Physcomitrella patens***. Plant J. 11:1195-1206.

Schaffer, R., Landgraf, J., Accerbi, M., Simon, V., Larson, M., and Wisman, E. 2001. Microarray analysis of diurnal and circadian-regulated genes in ***Arabidopsis***. Plant Cell 13:113-123.

Smith, K. 2001. DNA microarrays: A strategic market analysis. http://www.frontlinesmc.com/.

Stein, D., Foster, E., Huang, S., Weller, D., and Summerton, J. 1997. A specificity comparison of four antisense types: Morpholino, 2'O-methyl RNA, DNA, and phosphorothioate DNA. Antisense Nuc. Acid Drug Devel. 7:151-157.

Stein, D.A., Skilling, D.E., Iverson, P.L., and Smith, A.W. 2001. Inhibition of vesivirus infections in mammalian tissue culture with antisense morpholino oligomers. Antisense Nuc. Acid Drug Devel. 11:317-325.

Strepp, R., Scholz, S., Kruse, S., Speth, V., and Reski, R. 1998. Plant nuclear gene knockout reveals a role in plastid division for the homolog of the bacterial cell division protein FtsZ, an ancestral tubulin. Proc. Natl. Acad. Sci. USA 95:4368-4373.

Thimm, O., Essigmann, B., Kloska, S., Altmann, T., and Buckhout, T.J. 2001. Response of ***Arabidopsis*** to iron deficiency stress as revealed by microarray analysis. Plant Physiol. 1273:1030-1043.

Zhu, T., Budworth, P., Han, B., Brown, D., Chang, H-S., Zou, G., and Wang, X. 2001. Toward elucidating the global gene expression patterns of developing ***Arabidopsis***: Parallel analysis of 8,300 genes by a high density oligonucleotide probe array. Plant Physiol. Biochem. 39:221-242.