

Arabidopsis Genome. A Milestone in Plant Biology

The completion of the *Arabidopsis* genome sequence is the culmination of a remarkable decade of world-wide growth and collaboration in developing *Arabidopsis* as a model genetic and genomic system. Since the 1930s, plant geneticists have been drawn to the study of *Arabidopsis* because of its small size, its predilection to self-pollinate, its quick generation time, and its copious production of tiny seeds. Parallels with *Drosophila melanogaster* were commonly discussed, but with the exception of a few pioneers, *Arabidopsis* was not widely embraced by the broader plant biology community until the last 10 years when it became apparent that genomic tools had the potential to revolutionize plant biology.

I made the decision to switch the focus of my laboratory from work on symbiotic nitrogen fixation to *Arabidopsis*-pathogen interactions in the summer of 1984, but the appeal of *Arabidopsis* had already been percolating in my mind for a number of years. Influenced by the work of George Rédei and John Langridge on the isolation of *Arabidopsis* biosynthetic mutants (Langridge, 1965; Rédei and Li, 1969), I devoted about a year of my postdoctoral training in 1973/1974 to the development of an *Arabidopsis* somatic cell system.

As an assistant professor in 1975, the major goal of my first National Science Foundation (NSF) grant was the "transfer of functioning bacterial nitrogen fixation genes to plants." Although the plant in this proposal was *Arabidopsis*, the seeds that I had worked on as a postdoc remained in an unopened vial in my desk drawer for many years. In fact, I did not fully appreciate the power of *Arabidopsis* as a genetic model until the early 1980s when I heard Chris Somerville describe the genetic analysis of photorespiratory mutants at a Plant Molecular Biology Gordon Conference, became aware that Elliot Meyerowitz's laboratory had shown not only that *Arabidopsis* has a small genome but also had begun construction of a correlated physical genetic map, and took a close look at Maarten Koornneef's and David Meinke's work on the isolation of *Arabidopsis* developmental mutants (Meinke and Sussex, 1979; Koornneef et al., 1980; Somerville and Ogren, 1982; Leutwiler et al., 1984). I then concluded (with some feeling of regret that my *Arabidopsis* seeds had lain dormant for so long) that it was just a matter of time before *Arabidopsis* would be adopted as a universal model system for plant biologists and that it would be a good idea to get in on the ground floor. The amount of research on *Arabidopsis* has grown from a trickle in the 1970s and 1980s to a steadily increasing flood in the 1990s (Fig. 1). The completion of the *Arabidopsis* genome sequence no doubt will stimulate even more *Arabidopsis*-related work.

It is noteworthy that the highly productive *Arabidopsis* genomics enterprise that we have today is not simply a consequence of plant biologists systematically recognizing the advantages of a model plant system. On the contrary, in the 1970s and 1980s many influential plant scientists resisted attempts to adopt a model system, arguing that the diversity of crop plants and the realities of funding plant-related research demanded that the majority of plant research should be carried out on agronomically important species. Interestingly, I think that the development of recombinant DNA technologies initially reinforced the view that a model plant system was not necessary because rDNA technology allowed genes to be cloned from any plant. This led to many unrealistic predictions about the coming of a second green revolution, but in reality, because of a lack of funding and a lack of focus, plant research languished in comparison to the stunning advances that were being made by our colleagues working with animal systems.

Thus, I believe that it was important, at least in the United States, that the NSF began to orchestrate a worldwide *Arabidopsis* genome project. In so doing, NSF had the willing assistance of an activist group of *Arabidopsis* converts who organized themselves into the Multinational Coordinated *Arabidopsis* Genome Research Project and the North American *Arabidopsis* Steering Committee. Gradually, an *Arabidopsis* infrastructure was established that included stock centers in the United States and Europe, a database, an annual meeting, recombinant inbred mapping lines, identification of PCR-based mapping markers, an expressed sequence tag sequencing project, and a coordinated genome sequencing project. More recent additions to the infrastructure include knock-out libraries with more than 200,000 lines in the public domain, expression tagged lines, transcriptional profiling tools, and a large database of single nucleotide polymorphisms and small insertions and deletions between the Columbia and Landsberg *erecta* accessions. The latter database, which should greatly facilitate map-based cloning (Drenkard et al., 2000), is notable because it was posted as a public service by Cereon Genomics, setting an important precedent for what will hopefully be the beginning of a new era of cooperation between industry and academia.

So now, only 10 years after the establishment of the Multinational Coordinated *Arabidopsis* Genome Project, there is a complete *Arabidopsis* sequence. Remarkably, the sequence was completed only 1 year after the *Caenorhabditis elegans* and a few months after the *D. melanogaster* genomes were finished, a major achievement for plant biology. It is important to explicitly and forcefully make the point that the com-

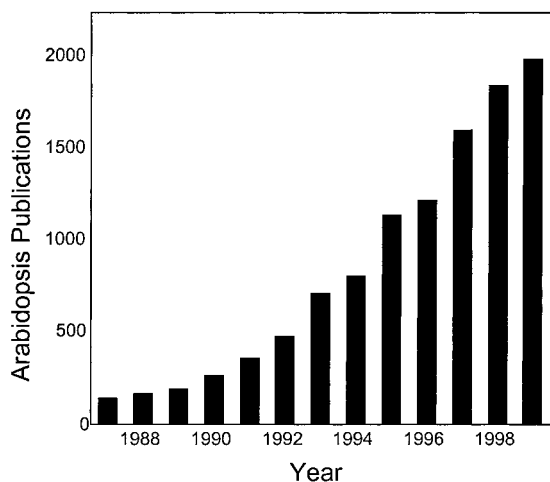


Figure 1. Number of papers that contain the word Arabidopsis in the title or abstract as determined by a search of the Institute for Scientific Information Web of Science and the BIOSIS databases.

pletion of the Arabidopsis sequence is important and relevant to all plant biologists, not just those who work on Arabidopsis. Recent discoveries, which show an unexpected degree of synteny in the genomes of various plant species, support the conclusion that not only the scientific information about specific genes, pathways, and processes will be useful in the genome analysis of other plants, but that many of the genomic resources and tools assembled for Arabidopsis can be modified for use with other plants as well (Schmidt, 2000). In fact, because of the sheer volume and extent of Arabidopsis-related research, integration of genetic, genomic, biochemical, physiological, and morphological data about Arabidopsis should reveal new insights into many universal aspects of plant biology.

Thus, in the same way that a concentration of resources on the genetic and genomic analysis of *Saccharomyces cerevisiae*, *C. elegans*, and *D. melanogaster* have led to the identification of what Hartwell et al. (1999) have called "functional modules" that are conserved across phyla, the Arabidopsis genome project should lead to the identification of novel plant-specific functional modules. Functional modules are a collection of proteins (that may include small molecules) that carry out relatively complex biochemical processes such as signal transduction or the synthesis of complex secondary products. Although many functional modules were first identified using traditional genetic and biochemical approaches, genomics should greatly accelerate this process. For example, because the genome sequence of *D. melanogaster* is available, tandem mass spectrometry was used to microsequence 30 peptides derived from a purified chromatin remodeling complex, immediately identifying the components of the complex and their corresponding genes (B. Kingston, personal communication). Another example is the use of a so-called

"compendium" of transcriptional profiles obtained under different environmental conditions and from a variety of mutants that has been used in budding yeast to identify putative functional modules by pattern matching (Hughes et al., 2000). A final example is a project headed by Jen Sheen (Department of Molecular Biology, Massachusetts General Hospital, Boston), in which I am participating, that is designed to identify all of the Arabidopsis mitogen-activated protein kinase (MAPK)-, MAPKK-, and MAPKKK-encoding genes and then to sort them into signal transduction cascades using high-throughput functional analysis of epitope-tagged or constitutively activated kinases in a protoplast transient expression system (<http://xanadu.mgh.harvard.edu/sheenweb/>).

The accomplishments of the past 10 years speak for themselves. The question of the moment, however, is where do we go from here? A publicized goal is the elucidation of the function of all of the estimated 25,000 Arabidopsis genes by the year 2010 (Chory et al., 2000). This is a demanding goal, but one that I think is likely to be achieved if the focus is not on individual genes but rather on sorting the genome into sets of functional modules and then determining the function(s) of each of the modules. This greatly simplifies the task, perhaps by more than an order of magnitude. Evolutionary analysis may also result in simplifications, as it is now becoming apparent that the Arabidopsis genome is likely to have undergone at least one round of duplication in its evolutionary history. Other insights into the function of many Arabidopsis genes may come from teasing out which Arabidopsis genes originated from the ancient integration of a cyanobacterial genome into the plant lineage to form the chloroplast. This may in part account for the vast metabolic capability of plants. A final simplification may come from the comparison of Arabidopsis accession Columbia, the sequenced line, with the many other Arabidopsis accessions. Subtle differences in the patterns of gene expression, perhaps related in turn to the expression or function of transcription factors or other regulatory proteins, may lead to significant insights into the evolution of so-called "quantitative" traits and the underlying genes and modules that are most likely the major driving force in the evolution of multicellular organisms.

Whether or not the plant genome can be deconvoluted into a relatively small number of functional modules, a relatively complete understanding of the Arabidopsis genome will once again require a coordinated international effort. One aspect of international cooperation that I think is extremely important is the development of plant-related databases. For example, in the United States the publicly funded Arabidopsis databases, an *Arabidopsis thaliana* Data Base followed by the Arabidopsis Information Service, have played a key role in compiling Arabidopsis-

related data in recent years. However, as we look to the future, more sophisticated databases that contain a variety of "functional" information will be required to store, display, distribute, and manipulate the disparate types of data on plant biology that are generated by the Arabidopsis and other plant genome projects. An important role for these databases will be to integrate data from a variety of disciplines, including physiology, biochemistry, genetics, and genomics, and from a variety of plants as well as microbes, parasites, and herbivores that interact with them. Integration of these data will be a formidable task, but will be greatly simplified if each of the plant biology sub-communities can agree on sets of standardized experimental procedures so that experiments carried out in different laboratories can be compared directly. This will require a level of cooperation that far surpasses the coordinated efforts that were required to sequence the Arabidopsis genome, but it is now being explored by those interested in transcriptional profiling analysis (<http://genome-www.stanford.edu/email/plantarrays.html>).

Evidence that cooperation in Arabidopsis functional genomics is currently in a healthy state is presented in this issue of *Plant Physiology*, in which six solicited articles are devoted to "Resources and Opportunities" for Arabidopsis genome research. The purpose of these articles is not only to make widely known what public resources are available, but also to recognize the significant contributions of a number of dedicated individuals. Specifically, take a look at the articles concerning seed and molecular resources for Arabidopsis by Scholl et al. (2000), the Arabidopsis genome projects by Federspiel (2000), the Arabidopsis Gene-Chip project by Zhu and Wang (2000), the Arabidopsis microarray project by Wisman and Ohlrogge (2000), the Arabidopsis knockout facility by Sussman et al. (2000), and the Arabidopsis Information Service by Rhee (2000). This issue also contains 26 research articles, four *Scientific Correspondence* articles, and four *Updates*, all concerning Arabidopsis-related work. In the future, *Plant Physiology* will publish additional issues that will highlight Arabidopsis research. The next such issue will be in June 2001 with John Browse and Ben Scheres serving as editors-in-charge.

I want to close with a warning. Although the advancement of plant genomic-related science will undoubtedly result in unprecedented opportunities for engineering agronomically important plant varieties, the growing public concern about the use of genetically modified organisms (GMOs) in agriculture poses a threat to the continued financial support for plant science research. This has already happened in Europe, and aggressive protests against GMOs by a consortium of environmental groups appear to be having an adverse affect on public opinion concerning GMOs in the United States. What is the responsibility of plant scientists in this debate? Although it

may be expected that scientists will adopt a self-serving stance in favor of GMOs and sustained funding for basic research, it is nevertheless critically important for plant scientists to become involved in educating both the public as well as relevant legislators about the underlying scientific issues. It is also critically important for scientists when discussing GMOs to clearly differentiate between scientific concerns and how these concerns might be addressed and/or alleviated by sensible and appropriate regulation, and the role that GMOs have played or might play in the business plans of multinational corporations. Several of our colleagues have vigorously entered the fray and their efforts on our behalf should be applauded. Now is the time for more of us to join their ranks. Some information on how to get started can be found at <http://plantbio.berkeley.edu/~outreach/>.

LITERATURE CITED

- Chory J, Ecker JR, Briggs S, Caboche M, Coruzzi GM, Cook D, Dangl J, Grant S, Gueriot ML, Henikoff S, Martienssen R, Okada K, Raikhel NV, Somerville CR, Weigel D (2000) National Science Foundation-sponsored workshop report, "The 2010 Project". Functional genomics and the virtual plant: a blueprint for understanding how plants are built and how to improve them. *Plant Physiol* **123**: 423–426
- Drenkard E, Richter BG, Rozen S, Stutius LM, Angell NA, Mindrinos M, Cho RJ, Oefner PJ, Davis RW, Ausubel FM (2000) A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in Arabidopsis. *Plant Physiol* **124**: 1483–1492
- Federspiel N (2000) Deciphering a weed: genomic sequencing of Arabidopsis. *Plant Physiol* **124**: 1456–1459
- Hartwell LH, Hopfield JJ, Leibler S, Murray AW (1999) From molecular to modular cell biology. *Nature* **402**: C47–C52
- Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD, Kidd MJ, King AM, Meyer MR, Slade D, Lum PY, Stepaniants SB, Shoemaker DD, Gachotte D, Chakraburty K, Simon J, Bard M, Friend SH (2000) Functional discovery via a compendium of expression profiles. *Cell* **102**: 109–126
- Koornneef M, Rolff E, Spruit CJP (1980) Genetic control of light-induced hypocotyl elongation in *Arabidopsis thaliana* (L.) Heynh. *Z Pflanzenphysiol* **106**: 147–160
- Langridge J (1965) Temperature-sensitive, vitamin-requiring mutants of *Arabidopsis thaliana*. *Aust J Biol Sci* **18**: 311–321
- Leutwiler LS, Hough-Evans BR, Meyerowitz EM (1984) The DNA of *Arabidopsis thaliana*. *Mol Gen Genet* **194**: 15–23
- Meinke DW, Sussex IM (1979) Embryo-lethal mutants of *Arabidopsis thaliana*: a model system for genetic analysis of plant embryo development. *Dev Biol* **72**: 50–61

- Rédei GP, Li SL** (1969) Effects of x-rays and ethyl methanesulfonate on the chlorophyll B locus in the soma and on the thiamine loci in the germline of *Arabidopsis*. *Genetics* **61**: 453–459
- Rhee SY** (2000) Bioinformatic resources, challenges, and opportunities using *Arabidopsis* as a model organism in a post-genomic era. *Plant Physiol* **124**: 1460–1464
- Schmidt R** (2000) Synteny: recent advances and future prospects. *Curr Opin Plant Biol* **2**: 97–102
- Scholl R, May S, Ware D** (2000) Seed and molecular resources for *Arabidopsis*. *Plant Physiol* **124**: 1477–1480
- Somerville CR, Ogren WL** (1982) Mutants of the cruciferous plant *Arabidopsis thaliana* lacking glycine decarboxylase activity. *Biochem J* **202**: 373–380
- Sussman MR, Amasino R, Young JC, Krysan PJ, Austin-Phillips S** (2000) The *Arabidopsis* knockout facility at the University of Wisconsin–Madison. *Plant Physiol* **124**: 1465–1467
- Wisman E, Ohlrogge J** (2000) *Arabidopsis* microarray service facilities. *Plant Physiol* **124**: 1468–1471
- Zhu T, Wang X** (2000) Large-scale profiling of the *Arabidopsis* transcriptome. *Plant Physiol* **124**: 1472–1476

Frederick M. Ausubel
Department of Genetics, Harvard Medical School
and Department of Molecular Biology
Massachusetts General Hospital
Boston, MA 02114