Technology issues in plant development

Barbara Mazur

For much of the 1980s and 1990s, the challenge to researchers in plant biotechnology was the isolation of potentially useful genes and the development of methods of transforming and then expressing those genes in appropriate crop strains.

Within the next decade, the task will be very different. Genes, in essence, will be in surfeit. The new challenge will be to make sense of the genetic information and to build outward from it to understand how we can control the way genes act in concert to produce crop traits of interest to farmers, processors, and consumers.

Genomics

The starting point for much of the gene discovery work in crop biotechnology is the expressed sequence tag (EST) approach pioneered by Craig Venter at The Institute for Genomic Research (Bethesda, MD) and Celera (Bethesda, MD), by Human Genomic Sciences (Rockville, MD), and by Incyte Pharmaceuticals (Palo Alto, CA). Comparisons of partial DNA sequences from cDNA libraries can be used to show which genes are transcribed at high levels in which species under a variety of conditions.

Once a gene is fully sequenced and characterized in one species, comparisons of ESTs provide a short-cut to identification of the homologous gene in other crops. This approach changes the gene discovery paradigm (in that the best way of finding a gene relevant to a particular crop may be to look in another species altogether), it improves product commercialization timelines, and increases the efficiency of cross-species gene cloning.

The next phase of genomic work in plants goes beyond the discovery of one or two genes. Most agronomically important properties of crops are genetically complex. Characteristics such as responses to physicochemical or environmental stresses, the production of high yields of a given energy storage compound, or intrinsic disease resistance will be associated with many genes or gene complexes and their regulatory elements. A much more complex level of genetic analysis will be necessary.

To undertake positional cloning of these quantitative trait loci, research will need to produce an infrastructure both of experimental procedures and of bioinformatics for managing the information generated.

It will be necessary to develop physical maps of plant genomes, a set of overlapping DNA fragments from genomic libraries held in yeast or bacterial chromosomes. The physical maps give a baseline on which to order both existing EST clones and other genetic markers. That will form a platform for using positional cloning techniques to draw out the association between observed traits and particular genetic markers, thereby narrowing the search for the genes responsible. One of the challenges in this field will be to characterize the contributions of multiple genes and environmental factors to the trait of interest in the crop.

Many of what will become standard tools in functional genomic analysis will find gainful employment in plant biotechnology.

Functional genomics

Gene mapping and positional cloning can tell us which genes are important and their importance, but tells us nothing about how they behave and what they actually do.

Many of what will become standard tools in functional genomic analysis will find gainful employment in plant biotechnology. Thus, the kinds of microarray technologies (often misnamed DNA chips) being developed at companies such as Incyte, Affymetrix (Palo Alto, CA), or Sequenom (Hamburg, Germany) to follow the messenger RNA levels of known genes will be applied extensively to panels of plant genes. Similarly, techniques such as SAGE[1] (serial analysis of gene expression), which do not require prior gene identification, will be used for comparative plant gene expression profiling.

Protein profiling using techniques such as two-dimensional polyacrylamide gel electrophoresis combined with mass spectrometry (see "Proteomics and molecular medicine," p. 19) is now also powerful enough to follow not only changes in the levels of the principle protein components of a cell but also changes resulting from posttranslational modifications. Other generic techniques for gaining insight into protein function include monoclonal antibody-based assays and the yeast two-hybrid system.

In order to identify genes expressed in a particular location of the plant or during a particular development stage, gene trapping methods using transposable elements have been developed. Sundaresan et al.[2], for instance, used transposable elements of the Ac/Ds system carrying GUS reporter genes to identify genes in *Arabidopsis thaliana* with expression patterns specific to particular organs, tissues, or cell types and those functioning during the development of organs, pattern formation, or cell differentiation.

Bioinformatics

The coordination and correlation of the huge quantities of information researchers are generating will require a substantial effort in bioinformatics. A bioinformatics system within the corporate crop breeding environment has to link all operations in crop development on the molecular constitution of plants and on the interactions between molecules.

Laboratory information management systems will be used extensively, ensuring the quality of experimental protocols, whether conducted by researchers or by laboratory robots undertaking automated processes. Such systems would be essential in handling and tracing the origins of large amounts of experimental data.

One of the key resources that automated procedures will produce are databases of mapping and sequencing information covering a range of species. These will be linked to others that cover, for instance, protein structural and functional data, the activities of agrochemical candidates, and plant behavior data under greenhouse or field conditions. The bioinformatics system also needs to integrate different parts of the research operation so that, for instance, laboratory researchers

Barbara Mazur is director of biotechnology research at DuPont Agricultural Products Experimental Station, P.O. Box 80402, Wilmington, DE 19880-0402, USA (Barbara.J.Mazur-1@usa.dupont.com).

have the opportunity to get early feedback on the field performance of crops containing transgene combinations of interest.

Separate parts of a bioinformatics system will be needed for modeling the metabolic perturbances that follow the addition of transgenes, what some would call the plant "physiome."

Although improvements in the throughput and efficiency of plant cell transformation (especially in elite lines) will still be needed in order to achieve commercial goals, a bigger challenges lies in assessing rapidly the consequences of transformation. Phenotypes such as herbicide tolerance show up in transformed single cells: traits such as high seed yields do not.

What R&D teams need are high-throughput, rapid, small-scale assays for traits of interest. It is insufficient (although it is necessary) to have tests that look for the gene or phenotype: the insertion or expression of a transgene may be responsible for pleiotropic effects such as the production of catabolites or changes in metabolic flux in pathways elsewhere in the cell.

In later stages of development, the use of DNA markers can accelerate the breeding process. Restriction fragment length polymorphism (RFLP) maps for most of the major commercial crops have already been developed. RFLP markers (and other markers such as single nucleotide polymorphisms) are used to track the inheritance of genotypes in the breeding process, but it will be even more essential to do it efficiently when dealing with complex polygenic traits.

Gene expression

In the first generation of transgenic crops, control of gene expression was reasonably straightforward. Traits such as herbicide tolerance, insect resistance, or disease resistance just required overexpression of one gene—coding for one enzyme, one toxic protein, or one virus coat protein. In order to improve both agronomic and grain quality traits in transgenic crops, however, it will be important to precisely control gene expression in both a tissue-specific and temporal manner. The corollary of the use of genomic studies to uncover the multigenic nature of certain traits is that plant genetic engineers will need to learn how to coordinate the expression of several genes over the growing period of the crop. For crops used to produce industrial feedstocks such as oils, starches, or other polymers, high expression levels (perhaps of the genes for all the enzymes in a particular pathway) may be necessary to make production economically competitive.

Activities in genomics will identify many new promoter sequences[2] across a range of species. However, other avenues are also receiving increasing attention.

A number of researchers are on the trail of transcription factors, for instance. In tobacco, abscisic acid has been shown[3] to increase the production of mRNA for two seed storage proteins while repressing the expression of isocitrate lyase mRNA. The response to abscisic acid appear to be modulated both developmentally and by exogenous sucrose and calcium. In other work, on the anthocyanin pigment synthesis pathway in maize, Vicki Chandler and colleagues at the University of Oregon have shown that regulation of pigment synthesis is dependent on the allelic composition at a particular locus[4].

In order to improve both agronomic and grain quality traits in transgenic crops, however, it will be important to precisely control gene expression in both a tissue-specific and temporal manner.

Another phenomenon that continues to be the focus of several research programs, especially for high-level expression of transgenes, is the influence of scaffold- or matrix-attachment regions (SAR, MAR)[5,6]. These regions of DNA are thought to be of fundamental importance for the organization of nuclear material and for the regulation of gene expression. When flanked with a SAR from tobacco, for instance, the expression per copy of a beta-glucuronidase reporter gene was almost 140-fold higher than a control with no SAR[6].

SARs are one possible response to reduce homology-dependent gene silencing mechanisms, a phenomenon apparent when multiple copies of a transgene are present. However, they are less effective with large numbers of a given transgene. In this regard, improved knowledge of antisense[7] and of cosuppression mechanisms that plants have evolved to combat transposable elements or virus infections[8] would aid in improving the efficiency of expression.

Site-specific recombination systems such as Cre-lox[9] and chemically regulated promoters[10] are being used to study the control and timing of specific gene expression in detail.

Genetic variation

The palette of the molecular plant breeders is, of course, now much broader than it was traditionally. Bacterial genes for herbicide tolerance and *Bacillus thuringiensis* insecticidal proteins have already been used in genetically engineered crops widely grown in North America and elsewhere. Bacterial enzymes have also been used to increase the lysine content of experimental crops: dihydropicolinic acid synthase encoded by a gene from a *Corynebacterium* species and an aspartokinase gene from *E. coli* expressed in the seeds of canola and soybean increased the seed lysine content around two- to five-fold[11]. The comparable plant enzymes are subject to lysine feedback inhibition whereas the bacterial enzymes are not.

Non-crop plants and insects will also provide abundant sources of genes for the development of novel crops. So, too, will animals and humans—particularly in the production of "pharming" crops. Furthermore, in vitro mechanisms of molecular evolution promise to generate genes encoding entirely novel proteins, albeit based solidly on nature's original designs. Thus, DNA shuffling has been used to develop a version of the green fluorescent protein that has over 40-fold higher fluorescence than the wild-type protein[12].

Another source of variability will be shuffling of the modular enzymatic units within multifunctional enzyme complexes. A group at KOSAN Biosciences (Burlingame, CA) and Stanford University (Stanford, CA), for instance, has shown that the substrate specificity of polyketide synthases is influenced by a small "hypervariable region" in one enzyme module. They suggest this region could be manipulated by combinatorial mutagenesis to produce novel enzyme specificities, thereby providing a means for synthesizing new types of polyketides.

In taking crop improvement forward over the next decade or so, researchers will have to capture and correlate the vast amounts of information stemming from studies of plant genomes, expression patterns, and the physiology of whole plants. To harness this new knowledge in novel crops, they will need efficient methods, not only for transforming crops, but also for analyzing the outcomes of transformation at an early stage.

1. Velculescu, V.E. et al. *Science* **270**, 484–487 (1995).
2. Sundaresan, V. et al. *Genes and Development* **15**, 1797–1810 (1995).
3. Bustos, M.M., Iyer, M. & Gagliardi, S.J. *Plant Mol. Biol.* **37**, 265–274 (1998).
4. Hollick, J.B. & Chandler, V.L. *Genetics* **150**, 891–897 (1998).
5. Meier, I. et al. *Plant Cell* **8**, 2105–2115 (1996).
6. Allen, G.C. et al. *Plant Cell* **8**, 899–913 (1996).
7. Baulcombe, D.C. *Plant Mol. Biol.* **32**, 79–88 (1996).
8. Matzke, M.A. & Matzke, A.J. *Cell Mol. Life Sci.* **54**, 94–103 (1998).
9. Odell, J.T., Hoopes, J.L. & Vermerris, W. *Plant Physiol.* **106**, 447–458 (1994).
10. McNellis, T.W. et al. *Plant J.* **14**, 247–257 (1998).
11. Falco, S.C. et al. 1995. *Bio/Technology* **13:**577-587.
12. Crameri, A. et al. *Nat. Biotechnol.* **14**, 315–319 (1996).