

Plant Functional Genomics

Chris Somerville* and Shauna Somerville

Nucleotide sequencing of the *Arabidopsis* genome is nearing completion, sequencing of the rice genome has begun, and large amounts of expressed sequence tag information are being obtained for many other plants. There are many opportunities to use this wealth of sequence information to accelerate progress toward a comprehensive understanding of the genetic mechanisms that control plant growth and development and responses to the environment.

The recent completion of the genome sequences of a number of bacterial species and several eukaryotes (1) has demonstrated the feasibility and utility of sequencing large genomes. Most biologists now envision the day when the complete genome sequence of their favorite organisms, or a proxy thereof, will be available in powerful electronic databases. Access to this information, and new tools that exploit it, will profoundly alter the ways we select and approach questions in biology. This, in turn, will directly affect the application of genetic methods for improving economically important species. Although future developments in a rapidly emerging field are difficult to predict, we believe many of the major developments in genomics that will influence basic research in plant biology and plant improvement during the next decade can be anticipated. Some of these possibilities are summarized here as well as in recent articles (2, 3).

A DNA Sequence Transect

One of the first eukaryotic organisms that will be completely sequenced is the small mustard species *Arabidopsis thaliana* (4) (Fig. 1). During the past decade, *Arabidopsis* has emerged as one of the most widely used model organisms for studying the biology of higher plants. As a member of the mustard family, it is closely related to many food plants such as canola, cabbage, cauliflower, broccoli, turnip, rutabaga, kale, brussels sprouts, kohlrabi, and radish. It was chosen for sequencing because it has a highly compact genome of about 130 Mb with little interspersed repetitive DNA. Six research groups in Japan, Europe, and the United States are collaborating on the sequencing. About 59% of the genome sequence is currently available in public databases and a large proportion of the genes are also represented by partial cDNA sequences (4, 5). It is currently anticipated that the complete ge-

nome sequence of *Arabidopsis* will be available by the end of the year 2000.

Because *Arabidopsis* is only distantly related to the cereal crops that provide the bulk of the world's food supply, the genome of rice will also be sequenced during the next decade (6). Rice was chosen because, in addition to its importance as a food source for about one-quarter of the human population, it has one of the most compact genomes among the cereals. It contains about 3.5 times as much DNA as *Arabidopsis* but only about 20% as much DNA as maize and about 3% as much DNA as wheat (7). However, the genome organization of the cereals appears to be very highly conserved; rice, wheat, maize, sorghum, millet, and other cereals exhibit a high degree of synteny (8). The differences in genome size are primarily due to amplification of interspersed repetitive sequences (9); there is no evidence that angiosperms with large amounts of DNA per cell have substantially greater numbers of functional genes than

angiosperms with relatively small amounts of DNA. Because of extensive synteny among the cereal genomes, knowledge of gene order and organization in rice may be used to isolate and characterize the corresponding genes in other cereals (8, 10). Thus, for instance, if a genetic locus where a useful trait is encoded is mapped between a pair of closely linked molecular markers in wheat, it may be possible to identify candidate genes for the rice ortholog by analyzing the rice genome sequence located between the rice orthologs of the molecular markers.

The sequences of *Arabidopsis* and rice will provide two foci from which the genome contents of other higher plants can be extrapolated. It appears likely that, as the costs of DNA sequencing continue to decrease, additional plant genomes may eventually be sequenced. However, during the next decade additional complete plant genome sequences probably will not be publicly available because of the high cost of sequencing the whole genome of any of the major crops. For instance, the cost of sequencing the maize genome is expected to be about the same as the cost of sequencing the human genome. However, extensive partial cDNA sequence information will be publicly available for most of the genes from many important plant species (11). There are currently more than

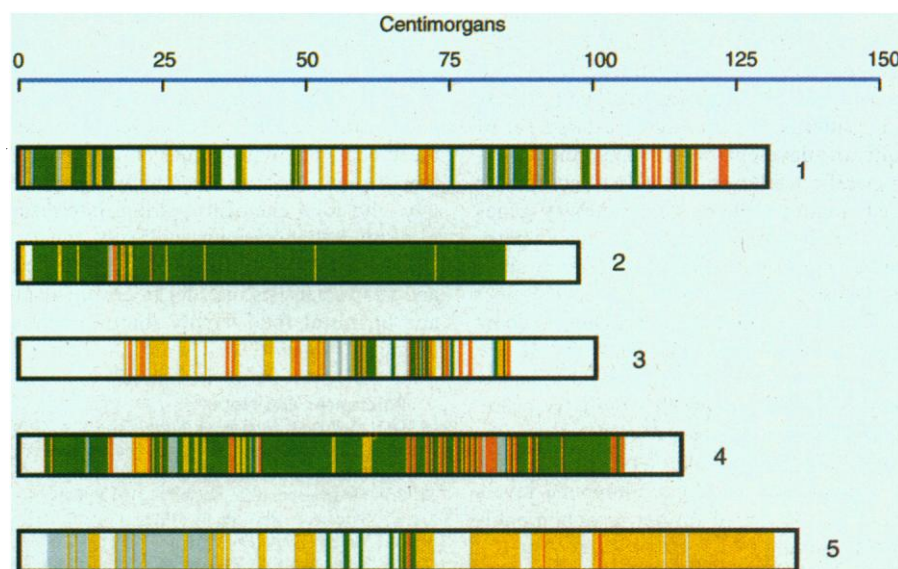


Fig. 1. Status of the *Arabidopsis* genome sequencing project. The five chromosomes are represented by rectangles; length is approximately to scale. Green regions represent annotated sequences available in GenBank; yellow represents regions completed and largely available in various databases; orange indicates regions that are currently being sequenced; gray indicates regions in preparation for sequencing. From the *Arabidopsis* database AtDB (<http://genome-www3.stanford.edu/cgi-bin/AtDB/Schrom>) with permission.

Carnegie Institution of Washington, Department of Plant Biology, 260 Panama Street, Stanford CA 94305, USA.

*To whom correspondence should be addressed. E-mail: crs@andrew2.stanford.edu

127,000 expressed sequence tag (EST) sequences from 19 plant species in public databases and the number is expected to grow rapidly during the next several years. These sequences will provide isomorphisms between the model genomes and other species, forming a kind of transect through genome diversity in higher plants that is anchored in comprehensive knowledge of the two representative species. Thus, as genes associated with functions or traits in one plant are cloned, it usually will be possible to identify the orthologs responsible for the trait in other plant species by a database search or by using the sequence information to clone the corresponding gene from the species of interest.

Flowering Plants Contain the Same Genes

Although flowering plants have evolved during the past 150 million years or so and therefore might be expected to be very similar at the genetic level, substantial developmental and metabolic diversity exists. Understanding the basis for this diversity is a key to understanding how to effect rational improvements in the productivity and utility of crop species. Knowledge of the genetic basis for intraspecies variation in specific traits should be useful for selecting or creating useful variation within a species.

The availability of extensive EST information for many species, in conjunction with the complete sequences of rice and *Arabidopsis*, will allow unambiguous insight into the question of how similar the genomes of higher plants are. When the *Arabidopsis* and rice sequences are complete, it will be possible to directly compare all the EST and other available sequences from various plants with the genomic sequences from the model genomes. Our preliminary analysis of available sequences suggests that most gene products from higher plants exhibit adequate sequence similarity to deduced amino acid sequences of other plant genes to permit assignment

of probable gene function, if it is known, in any higher plant. This is illustrated by a comparison of sequence identity of a random sample of putatively orthologous rice and *Arabidopsis* proteins (Fig. 2). Because of the relatively recent radiation of the angiosperms, we consider it likely that there will be very few protein-encoding angiosperm genes that do not have orthologs or paralogs in *Arabidopsis* or rice. Therefore, understanding the genetic basis for diversity may devolve to identifying the relevant differences in the control of expression or the function of essentially the same set of genes. Indeed, it has been hypothesized that the developmental diversity of higher plants may be largely due to changes in the cis-regulatory sequences of transcriptional regulators (12).

A major challenge to understanding the genetic basis of interspecies diversity is that, in at least some cases, minor changes in the structure or expression of a gene may lead to major changes in phenotype. This was recently illustrated for the genes that control modifications of fatty acids in plants (13). Higher plants collectively produce more than 200 fatty acids, which accumulate as storage oils in seeds. These fatty acids differ primarily because of the presence of double bonds, hydroxyls, epoxy groups, triple bonds, or secondary modifications of these functional groups at various carbons along the fatty acyl chains. It has recently become apparent that these functional groups are produced by a family of closely related fatty acyl desaturase-like enzymes (14). The observation that as few as four amino acid substitutions can convert a desaturase to a hydroxylase illustrates how new chemical constituents can accumulate without the evolution of an obviously distinct enzyme (13). Large gene families have also been observed for cytochrome P450s, enzymes involved in polysaccharide biosynthesis, disease-related genes, transcription factors, protein kinases, and phos-

phatases to name a few. Thus, a major challenge associated with exploiting the explosion of genome information will be in deducing rules that can predict the precise function of members of gene families.

One promising avenue, termed phylogenomics, exploits the use of evolutionary information to facilitate assignment of gene function (15). The approach is based on the idea that functional predictions can be greatly improved by focusing on how genes became similar in sequence during evolution instead of focusing on the sequence similarity itself. Because the power of the analysis increases in proportion to the number of sequences that are available, this method should become more useful as the database of plant sequences expands.

Assigning Function to Genes

One of the major efficiencies that has emerged from plant genome research to date is that about 54% of higher plant genes can be assigned some degree of function by comparing them with the sequences of genes of known function (16) (Fig. 3). In effect, a universal biology has coalesced from the common language of gene and protein sequences. Unfortunately, knowing the general function frequently does not provide an insight into the specific role in the organism. For instance, on the basis of sequence analysis, about 13% of *Arabidopsis* genes are inferred to be involved in transcription or signal transduction (16). However, knowing that a gene encodes a kinase or transcription factor does not provide any useful information about what processes are controlled by these genes. Thus, completion of the genome sequences of *Arabidopsis* and rice will be followed by a second phase of large-scale functional genomics in which all of the 20,000 to 25,000 genes that make up the basic angiosperm genome will be assigned function on the basis of experimental evidence. Considering that the combined efforts of the plant biology com-

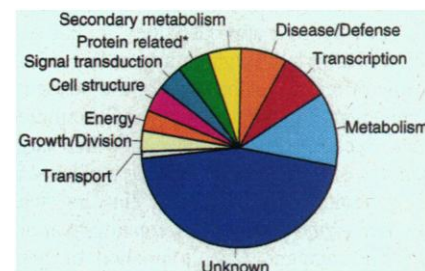
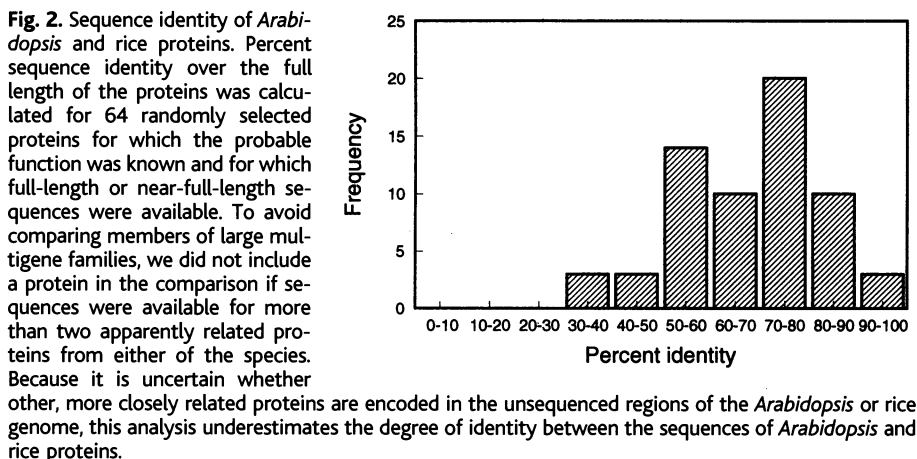


Fig. 3. Functional classification of predicted genes in a 1.9-Mb region of the *Arabidopsis* genome. Protein related refers to gene products involved in synthesis, degradation, modification, storage, and targeting of proteins and in intracellular trafficking. Analysis was based on 389 predicted or known genes. From (16).

munity have resulted in direct functional analysis of only about 1000 genes to date (5), this may seem like a tall order. However, it appears likely that the efficiency gained by "reverse genetics" will fundamentally change this equation. Large collections of insertion mutants are available for *Arabidopsis*, maize, petunia, and snapdragon, and collections of insertion mutants will probably be created in several other species including rice. These collections can be screened for an insertional inactivation of any gene by using the polymerase chain reaction (PCR) primed with oligonucleotides based on the sequences of the target gene and the insertional mutagen (3, 17). The presence of an insertion in the target gene is indicated by the presence of a PCR product. By multiplexing DNA samples, hundreds of thousands of lines can be screened and the corresponding mutant plants can be identified with relatively small effort. In addition, several groups are embarking on sequencing the genomic DNA flanking a large number of insertions so that an insertion in virtually any gene can be identified by a computer search (2, 18). Analysis of the phenotype and other properties of the corresponding mutant will frequently provide an insight into the function of the gene.

A major limitation to the analysis of gene function by mutation is that a high degree of gene duplication is apparent in *Arabidopsis* (16) and is, therefore, probably a common feature of plant genomes. Because many of the gene duplications in *Arabidopsis* are very tightly linked, it usually will not be feasible to produce double mutants by genetic recombination. A possible solution may be to use homologous recombination to eliminate tandem genes simultaneously by gene replacement (19). Alternatively, a method for producing point mutations by using RNA-DNA hybrids may be useful (20). By targeting mutagenic sequences that introduce stop codons to regions that are conserved among all duplicates it may be possible to generate all combinations of null mutations in all members of a multigene family from one experiment. Although these methods are amenable to a gene-by-gene approach, they are not well suited to a high throughput approach because of low efficiency or because of the necessity of regenerating plants from single cultured cells. Because of the ease with which large numbers of transgenic *Arabidopsis* plants can be generated by infecting flowers with *Agrobacterium tumefaciens* containing an insertional mutagen (21), a method of gene silencing based on producing double-stranded RNA from bidirectional transcription of genes in transgenic plants may be broadly useful for high-throughput gene inactivation (22). This method could, in principle, be designed to use promoters that are expressed in only a few cell types or at a particular devel-

opmental stage or in response to an external stimulus. This could significantly obviate problems associated with the lethality of some mutations. Although the mechanism is not yet understood, it bears some resemblance to double-stranded RNA-mediated gene silencing in nematodes (23).

For many applications, particularly in species other than *Arabidopsis* where production of tens of thousands of transformants is slow and time-consuming, virus-induced gene silencing may be the most facile method for suppressing gene function (24). This method exploits the fact that some or all plants have a surveillance system that can specifically recognize viral nucleic acids and mount a sequence-specific suppression of viral RNA accumulation. By inoculating plants with a recombinant virus containing part of a plant gene, it is possible to rapidly silence the endogenous plant gene (25).

It is expected that application of these and related methods will lead to the assignment of some degree of gene function to all genes in the basic angiosperm genome within the next decade. In addition, parallel studies of the function of genes in other, nonplant, organisms will contribute a great deal to understanding gene function. This comprehensive approach to understanding gene function will greatly facilitate the creation of plant improvements that are based on knowledge of the entire system instead of on a gene-by-gene basis.

Impact of Gene Chips and Microarrays

One of the most important experimental approaches for discovering the function of genes promises to be gene chips and microarrays. In principle, DNA sequences representing all the genes in an organism can be placed on miniature solid supports and used as hybridization substrates to quantitate the expression of all the genes represented in a complex mRNA sample (26). Thus, we may expect to have extensive databases of quantitative information about the degree to which each gene responds to pathogens, pests, drought, cold, salt, photoperiod, and other environmental variation. Similarly, we will have extensive information about which genes respond to changes in developmental processes such as germination and flowering. In addition, we will soon know which genes respond to the phytohormones, growth regulators, safeners, herbicides, and related agrichemicals. Perhaps less obviously, we may expect to have similar information for many mutants or natural accessions that differ in some way that cannot be readily assigned to genetic variation by other criteria. Knowledge of which genes exhibit changes in expression of a mutant of interest will be useful for formulating hypotheses about the

roles of the gene affected by the mutation (27).

These databases of gene expression information will provide insights into the "pathways" of genes that control complex responses and will be a first step toward an ecology of the genome in which the genome is viewed as a whole and the relationships of gene products to each other will be considered from at least one perspective (relative level of expression). Perhaps the types of models that ecologists currently use for understanding the interactions in ecosystems will prove useful (28). Indeed, because microarrays can be made for any organism for which complementary DNAs can be isolated it seems likely that ecological applications will be found. It is not necessary to know the sequence of the genes on a DNA microarray beforehand—this can be determined after the arrays have been used to identify genes that may be of interest by some criterion.

The accumulation of DNA microarray or gene chip data from many different experiments will create a potentially very powerful opportunity to assign functional information to genes of otherwise unknown function. The conceptual basis of the approach is that genes that contribute to the same biological process will exhibit similar patterns of expression. Thus, by clustering genes based on the similarity of their relative levels of expression in response to diverse stimuli or developmental or environmental conditions, it should be possible to assign hypothetical functions to many genes based on the known function of other genes in the cluster (29). Work with plant microarrays is just beginning but there is every reason to believe that this approach will soon be a standard component of the repertoire of plant biologists (30). The principal challenge, at present, is to develop methods for databasing and interrogating the massive amounts of data that result from this type of experiment.

One of the most important advances in plant improvement was the discovery of hybrid vigor and the exploitation of this phenomenon in modern breeding programs. In spite of extensive speculation about the mechanistic basis for hybrid vigor, it is poorly understood (31). It will be very interesting to compare whole genome microarrays of inbred parental lines with the heterotic hybrids. We speculate that the hybrids will exhibit significant differences in the expression of clusters of functionally related genes and that different hybrids will have different patterns of expression. If this proves to be the case, it may be possible to progress toward more predictive development of heterotic hybrids by breeding for certain patterns of gene expression. It may also provide a much needed linkage between the breeding of different plants—that is, if it is found that variation in the expression of certain pathways or processes is associated with enhanced yield or quality in one species, this may provide testable targets for

rational improvement in other species.

In contrast to microarrays, which are produced by directly spotting DNA on a matrix, gene chips are produced by synthesizing oligonucleotides on a solid support by photolithography or other methods (32). This method has the potential to produce arrays that contain several hundred thousand oligonucleotides. Thus, assuming further improvements of the technology, it is possible to envision gene chips with sufficient complexity to represent an entire plant genome. Gene chips have been used to measure the expression of all genes in the yeast genome with minimal concern about cross hybridization of structurally related genes (33). By hybridizing yeast genomic DNA to such chips, 3714 single-nucleotide polymorphisms between two genotypes could be identified in a single hybridization (34). Although the chips are currently too costly for routine use in many breeding programs, it seems likely that technical innovations and the efficiencies associated with expanded use will drive the costs down. The application of such chips or other oligonucleotide array technologies to genotyping individuals in segregating populations will revolutionize genetic mapping and marker-assisted breeding.

Artificial Plant Chromosomes

As the genomics of *Arabidopsis* and rice progress, one of the principal challenges will be to develop the methods by which advanced knowledge about these organisms is translated into useful insights about the hundred or more plant species of economic importance. At the single gene level, excellent tools are being developed for comparing the functions of plant genes. It is easy to produce large numbers of stable transgenic plants of *Arabidopsis*. Thus, to test the function of a cloned gene from a higher plant, a facile method is to determine whether it complements a mutation in the corresponding *Arabidopsis* gene or in another host. Although the results may sometimes be difficult to interpret when the trait controlled by the gene is highly divergent between the host and the gene donor, this should be a broadly useful method. It seems likely that, when the analysis of the rice genome is more fully developed, a comprehensive collection of rice mutations may provide a similarly useful alternative host for analyzing gene function.

In addition to the gene-by-gene approach, it would be useful to transfer large numbers of genes among plant species. For instance, because of the large numbers of genes that typically encode seed storage proteins in plants, it may be necessary to manipulate dozens of modified seed storage protein genes in order to be able to tailor the amino acid content of seeds. It may also be useful to be able to simultaneously introduce large numbers of genes in order to

explore the genetic basis of complex traits. In principle, it may be possible to identify the genes for useful pathways or traits by fragmenting a donor genome into large pieces—say 50 gene segments—and then introducing them into a recipient plant such as *Arabidopsis* and testing for components of the phenotype of interest. This will be useful only if the presence of the introduced gene confers a dominant or semi-dominant phenotype such as the presence of a new enzyme activity, an altered disease reaction, or modification of a developmental process. By introducing 50 genes at a time into *Arabidopsis*, only about 500 transgenic plants would need to be assayed in order to explore the entire genome of a typical diploid angiosperm at 1X coverage.

It seems likely that this type of analysis will be accomplished by making plant artificial chromosome (PLAC) libraries of a number of plant species with divergent properties and small genomes. The centromeres in *Arabidopsis* have been mapped (35) and current genome sequencing efforts will soon extend through these regions, facilitating identification and manipulation of centromere-containing regions of chromosomes. Although there is substantial uncertainty about what factors other than DNA sequence may be required to reconstitute a functional plant centromere, it may be possible to develop new vectors that contain both yeast and *Arabidopsis* centromeres. Because *Arabidopsis* telomeres are very similar to those in yeast (36) it may be possible to use a hybrid sequence of alternating plant and yeast sequences that function in both types of organisms. Thus, it may be possible to develop yeast artificial chromosome-PLAC libraries of many plants in yeast and then introduce them into a suitable plant host to evaluate the phenotypic consequences. By providing a defined chromosomal environment for cloned genes, the use of PLACs may also enhance our ability to reproducibly produce transgenic plants with defined levels of gene expression.

Rational Plant Improvement

The implications of genomics with respect to food, feed, and fiber production can be envisioned on many fronts. At the most fundamental level, the advances in genomics will greatly accelerate the acquisition of knowledge and that, in turn, will directly affect many aspects of the processes associated with plant improvement. Knowledge of the function of all plant genes, in conjunction with further development of tools for modifying and interrogating genomes, will lead to development of a robust genetic engineering discipline in which rational changes can be designed and modeled from first principles.

References and Notes

1. R. D. Fleischmann *et al.*, *Science* **269**, 496 (1995); C. J. Bult *et al.*, *ibid.* **273**, 1058 (1996); F. R. Blattner *et al.*, *ibid.* **277**, 1453 (1997); S. T. Cole *et al.*, *Nature* **393**, 537 (1998); H. W. Mewes *et al.*, *ibid.* **387** (suppl.), 7 (1997); A. Goffeau *et al.*, *Science* **274**, 546 (1996); The *C. elegans* Sequencing Consortium, *ibid.* **282**, 2012 (1998).
2. D. Bouchez and H. Höfte, *Plant Physiol.* **118**, 725 (1998).
3. R. A. Martienssen, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2021 (1998).
4. D. W. Meinke, J. M. Cherry, C. Dean, S. D. Rounsley, M. Koornneef, *Science* **282**, 662 (1998); S. Rounsley, X. Y. Lin, K. A. Ketchum, *Curr. Opin. Plant Biol.* **1**, 136 (1998).
5. S. D. Rounsley *et al.*, *Plant Physiol.* **112**, 1177 (1996).
6. S. A. Goff, *Curr. Opin. Plant Biol.* **2**, 86 (1999); T. Sasaki, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2027 (1998).
7. M. D. Bennett and J. B. Smith, *Philos. Trans. R. Soc. London Ser. B* **334**, 309 (1991).
8. M. D. Gale and K. M. Devos, *Science* **282**, 656 (1998).
9. J. L. Bennetzen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1975 (1998).
10. S. McCouch, *ibid.*, p. 1983.
11. E. Pennisi, *Science* **282**, 652 (1998).
12. J. Doebley and L. Lukens, *Plant Cell* **10**, 1075 (1998).
13. P. Broun, J. Shanklin, E. Whittle, C. Somerville, *Science* **282**, 1315 (1998).
14. M. Lee *et al.*, *ibid.* **280**, 915 (1998).
15. J. A. Eisen, *Genome Res.* **8**, 163 (1998).
16. The EU *Arabidopsis* Genome Project, *Nature* **391**, 485 (1998).
17. E. C. McKinney *et al.*, *Plant J.* **8**, 613 (1995); P. J. Krysan, J. C. Young, F. Tax, M. R. Sussman, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8145 (1996); R. G. Winkler, M. R. Frank, D. W. Galbraith, R. Feyereisen, K. Feldmann, *Plant Physiol.* **118**, 743 (1998); M. Montana *et al.*, *Science* **274**, 1537 (1996).
18. A. F. Tissier *et al.*, *Plant Cell*, in press.
19. S. A. Kempin *et al.*, *Nature* **389**, 802 (1997).
20. A. Cole-Strauss *et al.*, *Science* **273**, 1386 (1996); B. T. Kren, P. Bandyopadhyay, C. J. Steer, *Nature Med.* **4**, 285 (1998); V. Alexeev and K. Yoon, *Nature Biotechnol.* **16**, 1343 (1998).
21. S. J. Clough and A. F. Bent, *Plant J.* **17**, 735 (1999).
22. P. M. Waterhouse, M. W. Graham, M. B. Wang, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13959 (1998).
23. A. Fire *et al.*, *Nature* **391**, 806 (1998); M. K. Montgomery, S. Q. Xu, A. Fire, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15502 (1998).
24. D. C. Baulcombe, *Curr. Opin. Plant Biol.* **2**, 109 (1999).
25. S. Kjemtrup *et al.*, *Plant J.* **14**, 91 (1998); M. T. Ruiz, O. Voinnet, D. C. Baulcombe, *Plant Cell* **10**, 937 (1998).
26. M. Schena, D. Shalon, R. W. Davis, P. O. Brown, *Science* **270**, 467 (1995); J. L. DeRisi, V. R. Iyer, P. O. Brown, *ibid.* **278**, 680 (1997).
27. F. C. P. Holstete *et al.*, *Cell* **95**, 717 (1998).
28. R. F. Service, *Science* **284**, 83 (1999).
29. S. Chu *et al.*, *ibid.* **282**, 699 (1998); M. B. Eisen, P. T. Spellman, P. O. Brown, D. Botstein, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14863 (1998); V. R. Iyer *et al.*, *Science* **283**, 83 (1999).
30. D. Baldwin, V. Crane, D. Rice, *Curr. Opin. Plant Biol.* **2**, 96 (1999); D. M. Kehoe, P. Villand, S. Somerville, *Trends Plant Sci.* **4**, 38 (1999); Y. Ruan, J. Gilmore, T. Conner, *Plant J.* **15**, 821 (1998).
31. B. V. Milborrow, *J. Exp. Bot.* **49**, 1063 (1998).
32. S. P. A. Fodor *et al.*, *Science* **251**, 767 (1991); R. J. Lipshutz, S. P. A. Fodor, T. R. Gingeras, D. J. Lockhart, *Nature Genet.* **21**, 20 (1999).
33. L. Wodicka, H. Dong, M. Mittmann, M. H. Ho, D. J. Lockhart, *Nature Biotechnol.* **15**, 1359 (1997).
34. E. A. Winzler *et al.*, *Science* **281**, 1194 (1998).
35. G. P. Copenhaver, W. E. Browne, D. Preuss, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 247 (1998); G. P. Copenhaver and D. Preuss, *Curr. Opin. Plant Biol.* **2**, 104 (1999).
36. E. J. Richards and F. M. Ausubel, *Cell* **53**, 127 (1988).
37. We thank R. Beachy, J. Ecker, C. Gasser, P. Gilna, R. Hangarter, S. Kay, K. Keegstra, R. Martienssen, S. McCouch, D. Meinke, E. Meyerowitz, J. Verbeke, E. Ward, and S. Wessler for stimulating discussion.