As the databases fill up with a wealth of new plant gene sequences, researchers are turning to several innovative techniques to decipher the genes' functions

Reaping the Plant Gene Harvest

Plant scientists are on the verge of reaping a bounteous harvest—not of golden grains of wheat or corn, but of raw data. Just last month, for example, researchers reported the first complete sequences of two chromosomes from *Arabidopsis thaliana*, a tiny mustard widely used as a model plant, and they expect to decipher the rest of its genome by the end of this year. Projects aimed at sequencing the genomes of major crop plants, including rice, are also beginning to bear fruit. But as these efforts pack DNA databases with a bumper crop of sequences, scientists are facing a mind-boggling challenge: how to figure out what all the new plant genes they are discovering actually do.

In some ways, the challenge looms even greater than that facing animal researchers, whose own coffers are fast filling with genes of unknown function. Plant genomes tend to be even bigger and more complicated than those of mammals, but the scientists studying them lack the funding and attention doled out to their colleagues through high-profile ventures such as the Human or Mouse Genome Projects. Nevertheless, plant scientists are employing new techniques and innovative strategies to process their cornucopia of sequence information.

Some of these techniques have been borrowed from mammalian genomics, such as the "microarray" technology now coming into vogue for analyzing how gene expression patterns change as conditions vary (*Science*, 15 October 1999, p. 444). Other methods apply only to plants: For example, researchers are using snippets of DNA known as transposable elements to generate wholesale lots of mutant plants that can then be screened for interesting trait changes. Both of these technologies can help identify genes that plants turn on or off in response to stresses such as drought or salty soils—information the biotech industry welcomes eagerly.

But plant researchers are already venturing beyond that into the vast frontier of "metanomics." The field tracks the effects of particular mutations or environmental changes on a plant's entire metabolic repertoire. To begin, scientists generate countless profiles or fingerprints of a plant's metabolites, such as the cadre of sugar derivatives produced during starch biosynthesis in potatoes, under various conditions. Later, researchers can get a quick idea of what biochemical pathway an unknown gene might perturb, or the side effects of a particular gene alteration, by comparing the metabolic profile of a genetically altered plant with those already logged into the databases.

"These advances in tools and technologies are going to have a tremendously positive impact on the basic understanding of plant biology," predicts Susan Martino-Catt, a plant molecular biologist at Pioneer Hybrid in Johnston, Iowa. Already, she says, "the applications are allowing us to make progress that we've not been able to make until now."

DIGGING UP PLANT GENES

Name of Method	What It Is	How It Is Used
BLAST searching	Computer scanning of DNA databases	Assigns gene function based on similarity to known genes
Synteny	Compares chromosomal gene maps from different species	Assigns gene function based on similarities in sequence and location
Chimeraplasty	Inserts DNA/RNA hybrids into cells	Generates specific gene mutations
TUSC (Trait Utility System for Corn)	Inserts DNA that jumps into genes	Generates wholesale mutations in maize
Activation tagging	Inserts DNA enhancers via a plant cell–infecting bacterium	Generates wholesale mutations in plants
RNA silencing/Gene overexpression	Infects tobacco plants with genetically altered TMV	Turns tobacco genes either on or off
Microarray analysis	DNA snippets on chips	Tracks gene activity
Proteomics	Two-dimensional gels of protein expression	Tracks protein expression by cells
Metanomics	Gas chromatography and mass spectrometry profiles	Tracks metabolite expression by cells

Getting started

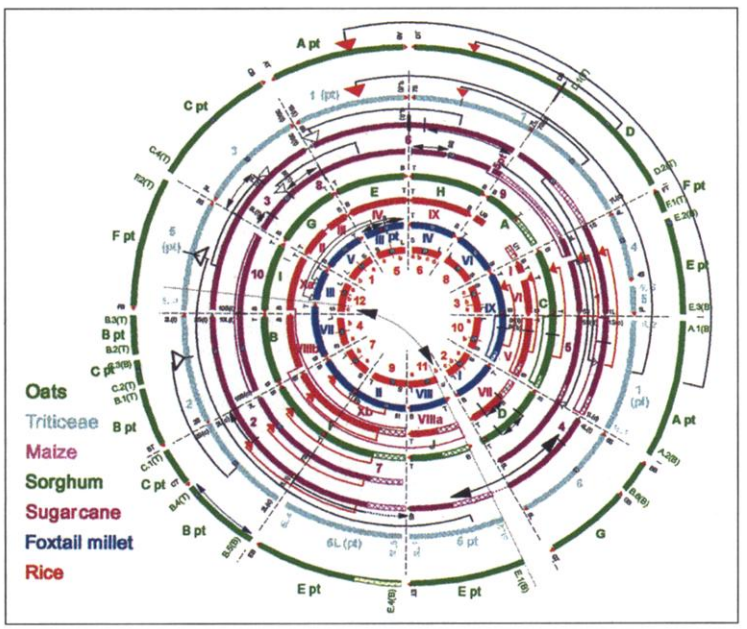
The first step in trying to figure out what a newfound gene does is to feed its sequence into a database. If the new gene matches the sequence of a gene whose function is known, it probably performs a similar role. This so-called BLAST searching can be a powerful technique, but the chances of finding a useful match will depend, in part, on the number of sequences already deposited in a database and what investigators know about them—which is one reason plant researchers are excited about the current progress with *Arabidopsis*.

Two of the plant's chromosomes, about 30% of the genome, have already debuted—the results appear in the 16 December issue of *Nature*—while the other three are expected to come out by the end of the year. But although the *Arabidopsis* genome may serve as a Rosetta stone for deciphering other plant genomes, the lowly mustard will inevitably come up short in some respects. For one, 40% of the genes on the two completed chromosomes have no assigned functions. And for another, *Arabidopsis* sports a very tiny genome—roughly 110 megabases—which is easily dwarfed by the genomes of plants such as maize, which carries 4500 megabases of DNA.

Fortunately, plant researchers have many other options besides BLAST searches. A quick and dirty way to guess the function of an unknown gene relies on a phenomenon called synteny in which genes devoted to particular functions, such as leaf development, tend to cluster in the genome and maintain their chromosomal organization from one organism to another. In fact, synteny is turning out to be so pervasive in grasses—including the important crop plants, wheat, rice, and maize—that researchers can draw a map in

which they position the chromosomes of some seven different grass species in concentric circles and then draw straight spokes to connect similar genes in similar places on those chromosomes (see diagram). "We are not talking about related genes, but the same genes," says Jeffrey Bennetzen, a plant geneticist at Purdue University in West Lafayette, Indiana.

Still, wheat is not maize, and rice is not millet. "Clearly there are differences as well," Bennetzen adds. And there is another limitation to comparative genetics: Researchers can often tell what class of protein a gene encodes, but they still may have no clue about the specific role it plays. For example, cells contain numerous kinase enzymes. They all add phosphate groups to other proteins, but act in a myriad of disparate biochemical pathways. "Five years ago, it would have been, 'Wow, we found a new kinase,' " says Charles Arntzen at the Boyce Thompson Institute for Plant Research at Cornell University. "Today it is, "Oh no, we've found another new kinase.' "



Circles of genes. The diagram shows how the genomes of seven grasses can be arranged so that regions carrying similar genes are aligned.

To get a handle on more specific information, genetic engineers are approaching the problem from a different direction: mutating plant genes and seeing what effect that has on the plant. Investigators already have developed a panoply of techniques, some of which have pinpoint specificity, while others create mutations in large numbers of genes. On the high-specificity end sits the fledgling technology dubbed chimeraplasty, which uses hybrids of synthetic DNA and RNA—chimeras—to create specific single-letter mutations in the genes of plants and other organisms (*Science*, 16 July 1999, p. 316). Researchers, including Arntzen and Gregory May of the Samuel Roberts Noble Foundation in Ardmore, Oklahoma, have successfully used chimeraplasty to mutate plant genes whose functions are already known, and they are now applying it to more obscure genes. "We are pursuing this with vigor," May says.

Martino-Catt's team at Pioneer Hybrid, which also reported success with chimeraplasty, has taken an additional, more broad-brush approach: They are creating random mutations in maize with a so-called transposable element—a bit of DNA that can jump into or out of genes in maize, interrupting their sequences, and effectively inactivating them. The Pioneer group has already let the transposable element, called mutator, loose in thousands of fertile maize plants, and have stored the resulting seeds. To find out whether mutator has interrupted any interesting genes, the researchers simply grow the seeds and then screen the resulting plants for changes, such as drought tolerance or sweeter kernels.

With probes specific for mutator, the group can pull out and clone the DNA sequences containing the transposable element and, thus, figure out which genes it disrupted to produce a new trait. "It's a very robust system," says Martino-Catt, who estimates that Pioneer can easily grow plants and fully screen about 100 genes a year using the approach. But although it works well in maize, mutator targets few other plants. And even in corn, the transposable element can't easily expose genes with less obvious functions, such as the production of a novel metabolite that might turn out to have medicinal properties.

A more general approach to mutating plant genes involves a technique called "activation tagging." The tags are stretches of DNA containing transcriptional enhancers —naturally occurring sequences that permanently turn on genes when inserted on the front or back end of the gene or even just nearby. "The beauty of enhancers is that they don't have to be in any specific place within a gene in order to turn on expression," says May, whose group is among those using the technique.

But enhancers don't jump into genes the way mutator does in maize—and that initially thwarted hopes of using them. However, researchers have been able to engineer the plant-infecting *Agrobacterium tumefaciens* to shuttle viral enhancers into plant cells. For example, May has used an engineered *Agrobacterium* to create over 300,000 lines of the alfalfa relative *Medicago truncatula* with inserts in approximately 95% of the plant's genes. His team is now screening for resulting trait changes. The enhancer elements, like the mutator sequence, then also serve as flags to identify the genes responsible for the changes.

In a similar vein, a team led by molecular biologist Guy della-Cioppa of Biosource Technologies Inc. in Vacaville, California, is using tobacco mosaic virus (TMV) to shuttle genes into plant cells to trace their function. In one type of application, the researchers have created "libraries" by separately cloning thousands of genes from a plant, such as *Arabidopsis*, into TMV. They then infect tobacco plants in the greenhouse with the altered TMVs and screen the resulting plants for changes, such as disease or drought resistance, conferred by the transplanted gene. Della-Cioppa says Biosource researchers have gleaned clues to various gene functions, but he declines to discuss the details for proprietary reasons.

The technique can also be used to inactivate tobacco plant genes that are counterparts of a gene from another plant. In this case, Della-Cioppa's team clones the transferred genes backward into the TMV, so that the messenger RNAs (mRNAs) they produce in the resulting plants will bind to the mRNAs produced by tobacco genes that have a similar sequence, effectively silencing the genes. Screening for trait changes should then shed light on the silenced gene and, ultimately, the inserted one. "It's a neat new aspect of virology to modify DNA sequences that we don't yet understand," Cornell's Arntzen says.

Questions remain about how useful the technique will be, however. For example, because TMV is used to infect only full-grown plants, the modified viruses may reveal little or nothing about the genes involved in early plant development. And there are concerns about the virus contaminating nearby plants in field tests, although Della-Cioppa says that the group uses a disabled virus that is less virulent than normal strains.

Looking at the big picture

All these techniques involve direct manipulation of the plant genes themselves. But researchers are also stepping back to get a broader picture of how a plant alters its pat-

CREDIT: MICHAEL GALE AND KATRIEN M. DEVOS, *PROC. NATL. ACAD. SCI. USA* **95**, 1971 (1998)

terns of gene expression or biochemistry over time or in response to changing environmental pressures. Some are borrowing techniques from the Human Genome Project, such as microarrays—chips containing hundreds or thousands of gene snippets laid out in precise arrays that provide quick snapshots of the expression of whole suites of genes simultaneously.

To find out which genes are expressed in a plant cell or tissue, a researcher isolates the mRNAs produced by the cell's genes, copies them into DNAs (called complementary DNAs or cDNAs), adds a fluorescent tag for tracking purposes, and then washes a solution of the labeled cDNAs over a chip. Each DNA snippet on the chip will only pick up the cDNA from the corresponding gene. Hence, by measuring the fluorescence of the chip DNAs, a researcher can learn which genes crank up or slow down their expression in response to a particular change, say the introduction of a gene of unknown function into a strawberry plant.

If a series of genes wink on and off together, they are likely to be operating in the same pathway. And if an unknown gene either over- or underexpressed in a plant affects the members of such a pathway, the unknown gene can be assigned to that pathway as well.

Although microarray techniques can place new genes into known pathways, they may not reveal exactly what they do in those pathways. What's more, finding that a gene is making its mRNA doesn't necessarily mean that the mRNA is actually making the ultimate protein product. Thus, some scientists have turned to "proteomics," which compares patterns of proteins, not mRNAs, under different conditions. "You have to catch the protein to know what metabolic pathway a gene is involved in," says geneticist Hervé Thiellement, whose team at the University of Geneva in Switzerland is among those using such techniques.

Proteomics uses two-dimensional (2D) gel electrophoresis to separate the proteins in a cell or tissue by size and pH characteristics, followed by mass spectrometry to help identify each component of the resulting gel pattern. Using this technology, researchers are building protein expression pattern databases for *Arabidopsis*, rice, maize, and pine trees. So far, these include comparisons of protein patterns from different plant organs and tissues, and they also show how the patterns change as a result of seasonal variations or water restriction. Visitors to the databases can call up pictures of 2D gels with links to previous publications about a protein of interest, the plant tissue from which it was derived, and genetic data such as sequence information and the localization of the corresponding genes on a linkage map.

The 2D gel technique can be time-consuming and technically challenging, however. And, in the case of proteins with general functions such as kinases, changes in protein patterns can't always reveal exactly what a specific protein does in a particular pathway of the plant. Some answers may come, however, from the newest technique turning heads in the plant genomic community: metabolic profiling. The idea has its roots in toxicology and blood screening, where chemists have long used state-of-the-art gas chromatography to separate components of a liquid, followed by mass spectrometry to identify and quantify them.

Plant researchers vary the application slightly: They want to create metabolic "maps" or generalized profiles of the metabolites produced by various plants and plant tissues. Profiles are likely to change when the plant is subjected to various environmental conditions or when an unknown gene is mutated, for example. Thus, after logging the results in a database, researchers can then compare how the maps change when they mutate a plant's own genes or introduce new ones.

Mutating with TMV. The plant above has been infected with TMV carrying a gene that turns up orange pigment synthesis, while the one at the upper right has been infected with a construct that blocks pigment production.

It was a problem with such a gene transfer experiment 3 years ago that prompted biochemist Richard Trethewey, then at the Max Planck Institute for Molecular Plant Physiology in Golm, Germany, to develop "metanomics," as he calls it. His team was trying to reduce the sugar levels and boost the starch content of potatoes, an alteration considered desirable by makers of potato chips and by those who want to improve crop yields. To do that, Trethewey and his colleagues introduced the yeast gene for the enzyme invertase and a bacterial gene for another enzyme, glucokinase, into potatoes. Invertase converts the disaccharide sugar, sucrose, into its two constituents, fructose and glucose, which can in turn be converted to starch with the aid of the glucokinase. But much to the researchers' surprise, the transgenic tubers lost *both* sugar and starch.

Not until the German team put the potatoes through an exhaustive chemical and physical analysis did they solve the puzzle: The excess invertase and the sugar molecules it produced had perturbed the potato plant's metabolism. The plants ended up rerouting the sugar molecules to create a host of alternate metabolites—none of which could ultimately fuel starch formation. Thus, the group was stuck with potatoes drained of both sugar and starch but abounding in unusual metabolites. Although the original experiment was seemingly a failure, the results drove home the realization that tracing such perturbations could provide another way to get insights into gene function.

In collaboration with the German pharmaceutical company BASF, Trethewey and colleagues then founded a company, also called Metanomics, located in Berlin. There his team is fast and furiously cataloging metabolic profiles from *Arabidopsis* plants exposed to various conditions. The team is also working with mutated *Arabidopsis* lines to glean the function of unknown genes.

Others at larger agro-pharmaceutical companies are also quietly and quickly creating their own version of metanomics. "Metabolic profiling will definitely play a considerable role in the future," says Egon Moesinger at Novartis in Basel, Switzerland. Martino-Catt at Pioneer Hybrid agrees: "I see this as becoming more and more important as we move forward."

Moving forward in genomics may actually carry plant researchers back to their more biological roots where they can learn more about the metabolic pathways of plants. "When you have the chance to stand back and look at *all* the genes at one time," Martino-Catt says, "you learn so much more by letting the plant teach you what it is doing."

–TRISHA GURA

CREDIT: MONTO KUMAGAI/BIOSOURCE TECHNOLOGIES INC.