From Corn-Bred Statistics to High-Tech Breeding

SCOTT BAUER (K9255-2)

Geneticist Ed Buckler measures a maize ear for statistical analysis.

In 1999, U.S. farmers produced more than 9.4 billion bushels of corn grain worth nearly $18 billion. Some quick math tells you they reaped an average of only $1.90 per bushel. With such a low price margin, farmers need corn varieties that thrive under the lowest chemical inputs, stand up straight under wind and rain, and produce grain with the highest nutrient value.

Those needs prompted the National Science Foundation to fund the Maize Evolutionary Genomics Project. It involves six researchers at three universities and the Agricultural Research Service's Plant Science Research Unit, located on the North Carolina State University campus in Raleigh. ARS plant geneticist Edward S. Buckler and N.C. State professor of crop sciences Major Goodman are two of the scientists involved.

With Goodman's collaboration, John Doebley at the University of Minnesota and Brandon Gaut at the University of California, Irvine, are searching the genomes of 1,300 inbred lines and wild races of maize and their close relatives, the teosintes, to sketch a genetic family tree. The researchers have selected a core set of 100 lines of maize that represent a broad range of genetic diversity. That diversity is due to mutations, or changes in the arrangement of the thousands of bases that spell out the code for a given gene. (See illustration on next page.)

Ultimately, breeders or genetic engineers will be able to use some of the exceptional variations to produce hardier plants or grain with better protein. But first, they must know which mutations on which genes control a desired trait. That requires statistical associations.

That's where ARS' Buckler, who has extensive expertise in statistical genetics, comes in. Goodman is also collaborating on the statistical aspect of the project.

Statistically First

"Our group is among the first to associate natural allelic [base] diversity with trait variation in plants by using statistical methods," says Buckler. He's adapting statistics developed for human genetic research to the world of corn and creating some new equations in the process.

Once developed, "the statistical methodology will be useful to study any type of plant or fungi, and of course humans," says Buckler.

Goodman explains that the human gene pool is far more homogeneous than the corn gene pool. "There's a lot more variability between inbred corn lines. But very few of the mutations—less than 1 percent—have any effect," he says. "We need statistical tests to kick out the false associations."

Buckler says his predecessors have earmarked so many genes which could possibly control traits of interest that, without statistics, "we may select the right gene only 10 percent of the time. These statistical approaches will allow us to find the right one 95 percent of the time."

Busy as Bees

Now in his second year on this project, Buckler is surveying 19 genes from 100 inbred lines of maize to test for statistical associations. The genes are known to contribute to traits like stalk length and plant height; flowering time; disease resistance; and the protein, carbohydrate, and oil content of the kernel. He says his group will probably end up working with 30 genes.

Buckler's laboratory looks a bit like a beehive, with seven technicians and students and a postdoctoral molecular biologist busily working at automated gene sequencers. Before he can apply any statistics, each of those 30 genes has to be sequenced—or spelled out base by base—for each of the 100 core lines. When the work is finished, the group will have sequenced some 10 million bases, he says.

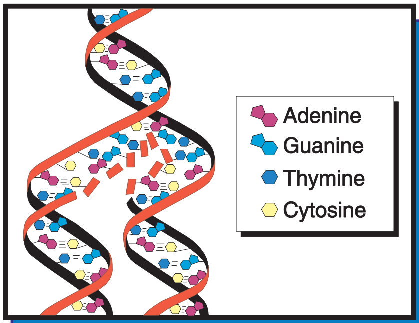
Then, after he makes the statistical associations and zeros in on the mutations most likely to affect the traits he is interested in, he has to prove the associations. So Buckler, Goodman, and co-workers are making hundreds of crosses between corn lines with the promising mutations and growing them to see if they really do make the plant flower earlier or shorten its stalk or change the nutritional profile of the kernels. The results will help Buckler fine-tune the statistics.

A-MAIZE-ingly Complex

Medical researchers have associated single-gene mutations with close to 100 human diseases. And they are starting to identify some of the mutations that play into the more complex diseases, such as heart disease or cancer. Likewise, the complex traits Buckler's group works with are controlled by several genes on different chromosomes. Each mutation adds a little to, or subtracts a little from, the normal trait. Thus they are called quantitative traits.

Trouble is, there are usually dozens—maybe hundreds—of base changes on each of the genes connected with a trait. But only one or two specific changes may actually affect the trait in question. So finding the exact bases that contribute to the corn plant's height or flowering time or amino acid profile is like trying to find specific addresses in an unfamiliar city without a map.

Complicating things further is the fact that "the corn genome is much more diverse than the human genome," according to Buckler, who notes that an average of 2.5 percent of the bases differ between any two corn lines. "That's 25 times more diversity than in the human genome." In fact, "corn and its wild relatives make up one of the most diverse plant genera in the world," he adds.



DNA is a dynamic molecule, constantly duplicating to create new cells. So it is prone to develop errors in the sequence of the four bases—Adenine, Guanine, Cytosine, and Thymine—that spell out the genetic code. Over time, genes accidentally change, or mutate. They may lose or gain a few extra bases, invert a section of bases so it reads backwards, or swap chunks of bases with other genes.

Thanks to groundwork laid by corn geneticists, particularly ARS' Charles W. Stuber, who led corn genetics research at Raleigh until his retirement in 1998, "we have a good map of markers," says Buckler. Stuber and others used markers to find the general location of genes that contribute to economically important traits. That's the street map.

But the map is far from complete. "We know maybe only 10 percent of the genes involved in plant height. We don't know which mutations in these genes affect the trait. And we don't know how the subtle changes affect traits," says Buckler. He and Goodman are looking for those bases—the specific addresses—on the streets mapped by Stuber and others.

A Look at Lineage

How do scientists sort out the one or two operative bases from the impostors? By using statistics that incorporate knowledge of ancestry. When a specific mutation shows up with much higher frequency in, say, tall corn than in short corn, there's a good chance that the mutation affects plant height. But there's also a chance that it just got passed down the ancestral line along with an operative mutation and doesn't really govern corn height.

For example, explains Buckler, take the human population of any large city. If you associate people's traits with their genotypes, you may get false positives if you assume that the population mated randomly. There are ethnic groups in any city, and people in those groups are more likely to marry and bear children within their group than outside it. So, too, corn has lots of little "ethnic" groups. "If you ignore ancestry, you may get a lot of associations that aren't real," says Buckler, noting that this happened in research on the human genome.

Fortunately, the science of statistics is always evolving. Oxford University statistician Jonathan Pritchard recently published new equations that improve accuracy in identifying the mutations behind human diseases across different populations.

"Disease is a yes-or-no situation—you either have it or you don't," says Buckler. He is extending Pritchard's equations to apply to quantitative traits such as plant height, which are less straightforward.—By **Judy McBride,** ARS.

This research is part of Plant, Microbial, and Insect Genetic Resources, Genomics and Genetic Improvement (#301) and Plant Biological and Molecular Processes (#302), two ARS National Programs described on the World Wide Web at http://www.nps.ars.usda.gov.

Edward S. Buckler is in the USDA-ARS Plant Science Research Unit, Gardner Hall, Room 2523, P.O. Box 7614, North Carolina State University, Raleigh, NC 27695-7614; phone (919) 513-1475, fax (919) 856-4816, e-mail buckler@statgen.ncsu.edu.