Standard breeding schemes are not always optimal – particularly when several genes spread over the entire genome are involved. Bayer scientists have now represented genotypes and crossbreeding steps in mathematical formulae: computational breeding can make breeding for new plant varieties faster, cheaper and better.

**Mathematics accelerates breeding**

Crop breeders have to experiment for years with thousands of plants to develop a variety with enhanced traits. Their breeding methods are successful but have a hard time keeping up with the increasing complexity of customer needs. In response to this challenge, Bayer mathematicians have developed a software program that can greatly facilitate breeding. It delivers a recommended formula for arriving at the target plant with significantly fewer cross-breeding generations.

**Standard breeding scheme**

- 5,000 plants
- 6 generations

**Supported by mathematical methods**

- 1,700 plants
- 5 generations

- **66% fewer plants**
- **17% faster**

Standard breeding schemes are not always optimal – particularly when several genes spread over the entire genome are involved. Bayer scientists have now represented genotypes and crossbreeding steps in mathematical formulae: computational breeding can make breeding for new plant varieties faster, cheaper and better.
It is not enough any more for tomatoes, cucumbers and other vegetables to look good on grocery store shelves. Customers, for example, prefer tomatoes that are aromatic and firm, but juicy. They’re not interested in tomatoes that get mushy or moldy. Farmers meanwhile want tomato plants that can resist pathogens and deliver high yields. Plant breeders over the years have managed to instill many of these characteristics in tomatoes by means of time-consuming and targeted breeding. Similar success has been achieved with other plants, such as cotton, where breeders have developed varieties resistant to certain pests.

But breeders are increasingly reaching the limits of what is possible with their methods when it comes to further enhancing today’s quality plants and combining the best traits of two varieties into a new one, for example breeding tomatoes that are both very large and resistant to several diseases. “Some traits simply cannot be bred into a new plant after just a few generations,” says Geert De Meyer, head of the Computational Life Science team dealing with Biometrics and Breeding Research at Breeding and Trait Development in Ghent, Belgium.

The process was considerably easier back in the day of Mendel’s breeding experiments. In the 19th century, Austrian monk Gregor Mendel, now known as the father of genetics, cross-bred two plants, one of which bore red flowers, the other white. According to the laws of Nature, some of the various offspring produced red flowers, others white or pink. His experiment was relatively simplistic, because the trait “flower color” is determined by a single gene found at a specific locus in the DNA with different versions passed on by the “mother” and “father.” However, things get much more complex when it comes to other traits.

Researchers are now using computer models to optimize tomatoes and cotton

But breeders are increasingly reaching the limits of what is possible with their methods when it comes to further enhancing today’s quality plants and combining the best traits of two varieties into a new one, for example breeding tomatoes that are both very large and resistant to several diseases. “Some traits simply cannot be bred into a new plant after just a few generations,” says Geert De Meyer, head of the Computational Life Science team dealing with Biometrics and Breeding Research at Breeding and Trait Development in Ghent, Belgium.

The process was considerably easier back in the day of Mendel’s breeding experiments. In the 19th century, Austrian monk Gregor Mendel, now known as the father of genetics, cross-bred two plants, one of which bore red flowers, the other white. According to the laws of Nature, some of the various offspring produced red flowers, others white or pink. His experiment was relatively simplistic, because the trait “flower color” is determined by a single gene found at a specific locus in the DNA with different versions passed on by the “mother” and “father.” However, things get much more complex when it comes to other traits.

A number of different genes are responsible for a tomato’s taste

“The flavor of a tomato or the high yield of a cotton plant is encoded by many genes at various regions in the DNA – the technical term is in the genetic background. We therefore refer to such traits as complex traits,” explains De Meyer. “It is impossible to merge simple traits of a father plant and complex traits from a mother plant in one single breeding step, because the DNA

The fine difference

Genetic engineering and plant breeding have more in common than is commonly believed: both methods involve transferring genes. One of the major differences is in how this transfer takes place. Using genetic engineering methods, researchers can introduce a gene for a specific trait into a plant. It is also possible to insert genes from other organisms. Plant breeders, meanwhile, cross-breed plants and thus combine their different characteristics – including sometimes unwanted properties. The new plant may then, for example, bear sweeter fruit but might also have more fibrous fruit flesh.
and chromosomes from the father and mother naturally mix 50:50, meaning that we always miss some pieces.” As a result, dozens of cross-breeding steps and several thousand plants are required to combine a complex trait and simple traits in a new plant. In some cases, the process goes beyond the limits of a breeding experiment. “We would need immense greenhouses for the many generations it would take to finally arrive at the right plant,” says De Meyer. “At present it’s difficult for us to combine different traits into one genetic background,” confirms Frank Millenaar, a tomato prebreeder at Vegetable Seeds in the Netherlands. Prebreeders deliver new plant traits from wild varieties or distantly related plant material to breeders. “Optimal crossbreeding schemes will help us greatly to reach our goals more quickly and efficiently.”

De Meyer therefore decided to ask Bayer’s Applied Mathematics group for help. This team develops mathematical models and algorithms to solve complex problems in a variety of business areas and currently also in plant breeding. In this case, they extended the computer application Gene Stacker developed in an earlier collaboration with the University of Ghent in Belgium. It calculates the specific crossbreeding steps and the number of plants needed to combine a set of desired traits. Working together with colleagues, co-developer Dr. Kathrin Hatz has enabled Gene Stacker to construct the right breeding schedule for complex background traits. It provides breeders with a recommended formula for crossbreeding.

“The basic algorithm is built on Mendel’s laws of heredity. Then we feed in the genetic information for the specific plants that we have available at the start,” explains Hatz, “using genetic markers to probe relevant DNA segments.”

For discrete simple traits, the genetic marker information is provided by the breeders who have identified the genome positions of interest in prior experiments. The genetic background is typically tracked by a set of markers positioned at regular intervals across the genome. “It would be ideal if we could merge the whole background information of a high-quality mother plant with a father plant carrying simple traits like the fruit size,” says De Meyer, “but in fact it is more likely that we would spoil the mother’s complex, elite traits with the father’s background. In order to transfer all the background information for the mother’s elite traits, we need many more steps.”

The computer then tests different breeding schedules until it hits the target. What sounds simple is very complex mathematically, because the background genes are passed on bit-by-bit, step-by-step from one generation to the next. The process can add up to several million possible combinations.

The computer starts with one mother and one father plant, which produce a first filial generation. Crossbreeding continues, for instance between two plants from one generation. In the third step, it may be necessary to backcross a progeny with the mother plant.

How computer models forecast customer needs

The work of the Applied Mathematics group at Bayer also helps to solve entirely different problems. For Bayer’s Consumer Health Division, experts investigated how the positioning of non-prescription drugs can be altered so that customers would be more likely to buy them. They combined classical findings in behavioral research about the limbic system with mathematical analysis of current internet searches. Experts refer to this combination as “predictive limbic modeling.” An analysis of search terms showed trends in society during the past few years by means of attributes such as success, speed and determination.

When combining plant traits, there are often several million breeding possibilities

For discrete simple traits, the genetic marker information is provided by the breeders who have identified the genome positions of interest in prior experiments. The genetic background is typically tracked by a set of markers positioned at regular intervals across the genome. “It would be ideal if we could merge the whole background information of a high-quality mother plant with a father plant carrying simple traits like the fruit size,” says De Meyer, “but in fact it is more likely that we would spoil the mother’s complex, elite traits with the father’s background. In order to transfer all the background information for the mother’s elite traits, we need many more steps.”

The computer then tests different breeding schedules until it hits the target. What sounds simple is very complex mathematically, because the background genes are passed on bit-by-bit, step-by-step from one generation to the next. The process can add up to several million possible combinations.

The computer starts with one mother and one father plant, which produce a first filial generation. Crossbreeding continues, for instance between two plants from one generation. In the third step, it may be necessary to backcross a progeny with the mother plant.
One element of the program is a specifically developed “branch-and-bound” algorithm, a method of mathematical optimization. “The algorithm first calculates possible combinations for mixing genes from one generation to the next. The number of branches grows, like on a tree,” explains Hatz. “Branch for branch, the method, tailored to solve the problem in question, examines efficiently in promising branches which combination will lead to the target. Branches that indicate early they won’t result in the target genotype are cut off and discarded until only one is left; that is what mathematicians refer to as ‘bounded.’

**Gene Stacker software predicts which breeding steps will lead to the targeted plant**

The calculation ends with a precise breeding schedule. The method recommends to breeders which plants should be crossbred in the next generation. Because there is a certain degree of probability involved in inheriting or not inheriting genes, the application further recommends the minimum number of breeding plants required to ensure that the target genes end up in at least one of the offspring. The new method is used by breeders in Bayer's Crop Science Division, who develop seeds for customers. "We have dozens of breeding centers worldwide, where we are currently working to introduce the new software application as a tool for optimizing plant breeding," says De Meyer. He, Kathrin Hatz, and the rest of the Gene Stacker team were able to demonstrate how well the program works in preliminary tests with cotton in simulation mode. Breeders currently need six years and 5,000 plants to introduce a background trait into a cotton plant. As the simulation showed, the Gene Stacker application reduces the number of plants to 1,700 and the breeding time to five years.

So thanks to mathematics, breeders can get a new breed to market faster. Furthermore, it can simplify breeding and therefore could reduce costs significantly, by as much as 66 percent in the cotton experiment. For Dr. Linus Görilitz, head of Applied Mathematics, this was more than just an isolated case. "In the era of big data and accelerating digitalization we can deliver very different and new solutions from which many areas of Bayer can benefit."

In the case of plant breeding, Görilitz and his team have already demonstrated the potential benefit of mathematics. "Getting products to customers a whole year earlier, with less resource input, is not only a competitive advantage, but also shows how innovation and sustainability accompany each other. Gene Stacker is just one example of how we can create value via the application of mathematical methods and the use of computer models."

---

Mathematical modeling is important in many industries today. You use it for advertising. How?

We have known for some time that humans process stimuli in advertising in the limbic system, the part of our brain that is responsible for emotions. These stimuli fall into three categories: balance, dominance and stimulation. Dominance is expressed by status symbols, success or honor. Balance refers to attributes such as security and stability. Stimulation is linked to adventure, fascination or the pursuit of new things. We try to understand what the customer feels: which of the three categories should we specifically appeal to in an advertisement to reach customers in the best emotional way? Thanks to computational modeling, we can now support the answer with statistical and empirical evidence.

To what extent can emotions be expressed mathematically?

We use limbic modeling. This method analyzes tremendous volumes of anonymous data, for instance on customer buying behavior or interests. It recognizes trends as well as the prevailing limbic predispositions among customers. Companies can then make decisions concerning new advertising strategies based on this mathematical analysis. In the past, they would have had to rely heavily on intuition.

---

research spoke with Marco Casanova, managing partner of the Branding Institute in Switzerland, about the importance of mathematical modeling.