Crop researchers harness artificial intelligence to breed crops for the changing climate

Carolyn Beans, *Science Writer*

Until recently, the field of plant breeding looked a lot like it did in centuries past. A breeder might examine, for example, which tomato plants were most resistant to drought and then cross the most promising plants to produce the most drought-resistant offspring. This process would be repeated, plant generation after generation, until, over the course of roughly seven years, the breeder arrived at what seemed the optimal variety.

Now, with the global population expected to swell to nearly 10 billion by 2050 (1) and climate change shifting growing conditions (2), crop breeder and geneticist Steven Tanksley doesn't think plant breeders have that kind of time. "We have to double the productivity per acre of our major crops if we're going to stay on par with the world's needs," says Tanksley, a professor emeritus at Cornell University in Ithaca, NY.

To speed up the process, Tanksley and others are turning to artificial intelligence (AI). Using computer science techniques, breeders can rapidly assess which plants grow the fastest in a particular climate, which genes help plants thrive there, and which plants, when crossed, produce an optimum combination of genes for a given location, opting for traits that boost yield and stave off the effects of a changing climate. Large seed companies in particular have been using components of AI for more than a decade. With computing power rapidly advancing, the techniques are now poised to accelerate breeding on a broader scale.

AI is not, however, a panacea. Crop breeders still grapple with tradeoffs such as higher yield versus

Researchers at ETH Zürich use standard color images and thermal images collected by drone to determine how plots of wheat with different genotypes vary in grain ripeness. Image credit: Norbert Kirchgessner (ETH Zürich, Zürich, Switzerland).

Nature Source Improved Plants (NSIP) speeds up its papaya breeding program in southeastern Mexico by using decision-making approaches more common in engineering. Image credit: Nature Source Improved Plants/Jesús Morales.

In 1997, Tanksley began chairing a committee at Cornell that aimed to incorporate data-driven research into the life sciences. There, he encountered an engineering approach called operations research that translates data into decisions. In 2006, Tanksley cofounded the Ithaca, NY-based company Nature Source Improved Plants on the principle that this engineering tool could make breeding decisions more efficient. "What we've been doing almost 15 years now," says Tanksley, "is redoing how breeding is approached."

A Manufacturing Process

Such approaches try to tackle complex scenarios. Suppose, for example, a wheat breeder has 200 genetically distinct lines. The breeder must decide which lines to breed together to optimize yield, disease resistance, protein content, and other traits. The breeder may know which genes confer which traits, but it's difficult to decipher which lines to cross in what order to achieve the optimum gene combination. The number of possible combinations, says Tanksley, "is more than the stars in the universe."

An operations research approach enables a researcher to solve this puzzle by defining the primary objective and then using optimization algorithms to predict the quickest path to that objective given the relevant constraints. Auto manufacturers, for example, optimize production given the expense of employees, the cost of auto parts, and fluctuating global currencies. Tanksley's team optimizes yield while selecting for traits such as resistance to a changing climate. "We've seen more erratic climate from year to year, which means you have to have crops that are more robust to different kinds of changes," he says.

For each plant line included in a pool of possible crosses, Tanksley inputs DNA sequence data, phenotypic data on traits like drought tolerance, disease resistance, and yield, as well as environmental data for the region where the plant line was originally developed. The algorithm projects which genes are associated with which traits under which environmental conditions and then determines the optimal combination of genes for a specific breeding goal, such as drought tolerance in a particular growing region, while accounting for genes that help boost yield. The algorithm also determines which plant lines to cross together in which order to achieve the optimal combination of genes in the fewest generations.

Nature Source Improved Plants conducts, for example, a papaya program in southeastern Mexico where the once predictable monsoon season has become erratic. "We are selecting for varieties that can produce under those unknown circumstances," says Tanksley. But the new papaya must also stand up to ringspot, a virus that nearly wiped papaya from Hawaii altogether before another Cornell breeder developed a resistant transgenic variety (6). Tanksley's papaya isn't as disease resistant. But by plugging "rapid growth rate" into their operations research approach, the team bred papaya trees that produce copious fruit

marketable appearance. And even the most sophisticated AI cannot guarantee the success of a new variety. But as AI becomes integrated into agriculture, some crop researchers envisage an agricultural revolution with computer science at the helm.

An Art and a Science

During the "green revolution" of the 1960s, researchers developed new chemical pesticides and fertilizers along with high-yielding crop varieties that dramatically increased agricultural output (3). But the reliance on chemicals came with the heavy cost of environmental degradation (4). "If we're going to do this sustainably," says Tanksley, "genetics is going to carry the bulk of the load."

Plant breeders lean not only on genetics but also on mathematics. As the genomics revolution unfolded in the early 2000s, plant breeders found themselves inundated with genomic data that traditional statistical techniques couldn't wrangle (5). Plant breeding "wasn't geared toward dealing with large amounts of data and making precise decisions," says Tanksley.

within a year, before the virus accumulates in the plant.

"Plant breeders need operations research to help them make better decisions," says William Beavis, a plant geneticist and computational biologist at Iowa State in Ames, who also develops operations research strategies for plant breeding. To feed the world in rapidly changing environments, researchers need to shorten the process of developing a new cultivar to three years, Beavis adds.

The big seed companies have investigated use of operations research since around 2010, with Syngenta, headquartered in Basel, Switzerland, leading the pack, says Beavis, who spent over a decade as a statistical geneticist at Pioneer Hi-Bred in Johnston, IA, a large seed company now owned by Corteva, which is headquartered in Wilmington, DE. "All of the soybean varieties that have come on the market within the last couple of years from Syngenta came out of a system that had been redesigned using operations research approaches," he says. But large seed companies primarily focus on grains key to animal feed such as corn, wheat, and soy. To meet growing food demands, Beavis believes that the smaller seed companies that develop vegetable crops that people actually eat must also embrace operations research. "That's where operations research is going to have the biggest impact," he says, "local breeding companies that are producing for regional environments, not for broad adaptation."

In collaboration with Iowa State colleague and engineer Lizhi Wang and others, Beavis is developing operations research-based algorithms to, for example, help seed companies choose whether to breed one variety that can survive in a range of different future growing conditions or a number of varieties, each tailored to specific environments. Two large seed companies, Corteva and Syngenta, and Kromite, a Lambertville, NJ-based consulting company, are partners on the project. The results will be made publicly available so that all seed companies can learn from their approach.

Drones and Adaptations

Useful farming AI requires good data, and plenty of it. To collect sufficient inputs, some researchers take to the skies. Crop researcher Achim Walter of the Institute of Agricultural Sciences at ETH Zürich in Switzerland and his team are developing techniques to capture aerial crop images. Every other day for several years, they have deployed image-capturing sensors over a wheat field containing hundreds of genetic lines. They fly their sensors on drones or on cables suspended above the crops or incorporate them into handheld devices that a researcher can use from an elevated platform (7).

Meanwhile, they're developing imaging software that quantifies growth rate captured by these images (8). Using these data, they build models that predict how quickly different genetic lines grow under different weather conditions. If they find, for example, that a subset of wheat lines grew well despite a dry spell,

then they can zero in on the genes those lines have in common and incorporate them into new drought-resistant varieties.

Research geneticist Edward Buckler at the US Department of Agriculture and his team are using machine learning to identify climate adaptations in 1,000 species in a large grouping of grasses spread across the globe. The grasses include food and bioenergy crops such as maize, sorghum, and sugar cane. Buckler says that when people rank what are the most photosynthetically efficient and water-efficient species, this is the group that comes out at the top. Still, he and collaborators, including plant scientist Elizabeth Kellogg of the Donald Danforth Plant Science Center in St. Louis, MO, and computational biologist Adam Siepel of Cold Spring Harbor Laboratory in NY, want to uncover genes that could make crops in this group even more efficient for food production in current and future environments. The team is first studying a select number of model species to determine which genes are expressed under a range of different environmental conditions. They're still probing just how far this predictive power can go.

Such approaches could be scaled up—massively. To probe the genetic underpinnings of climate adaptation for crop species worldwide, Daniel Jacobson, the chief researcher for computational systems biology at Oak Ridge National Laboratory in TN, has amassed "climatype" data for every square kilometer of land on Earth. Using the Summit supercomputer, they then compared each square kilometer to every other square kilometer to identify similar environments (9). The result can be viewed as a network of GPS points connected by lines that show the degree of environmental similarity between points.

> ### "For me, breeding is much more like art. I need to see the variation and I don't prejudge it. I know what I'm after, but nature throws me curveballs all the time, and I probably can't count the varieties that came from curveballs."
>
> —Molly Jahn

In collaboration with the US Department of Energy's Center for Bioenergy Innovation, the team combines this climatype data with GPS coordinates associated with individual crop genotypes to project which genes and genetic interactions are associated with specific climate conditions. Right now, they're focused on bioenergy and feedstocks, but they're poised to explore a wide range of food crops as well. The results will be published so that other researchers can conduct similar analyses.

The Next Agricultural Revolution

Despite these advances, the transition to AI can be unnerving. Operations research can project an ideal combination of genes, but those genes may interact in

Beans

unpredictable ways. Tanksley's company hedges its bets by engineering 10 varieties for a given project in hopes that at least one will succeed.

On the other hand, such a directed approach could miss happy accidents, says Molly Jahn, a geneticist and plant breeder at the University of Wisconsin–Madison. "For me, breeding is much more like art. I need to see the variation and I don't prejudge it," she says. "I know what I'm after, but nature throws me curveballs all the time, and I probably can't count the varieties that came from curveballs."

There are also inherent tradeoffs that no algorithm can overcome. Consumers may prefer tomatoes with a leafy crown that stays green longer. But the price a breeder pays for that green calyx is one percent of the yield, says Tanksley.

Image recognition technology comes with its own host of challenges, says Walter. "To optimize algorithms to an extent that makes it possible to detect a certain trait, you have to train the algorithm thousands of times." In practice, that means snapping thousands of crop images in a range of light conditions. Then there's the ground-truthing. To know whether the models work, Walter and others must measure the trait they're after by hand. Keen to know whether the model accurately captures the number of kernels on an ear of corn? You'd have to count the kernels yourself.

Despite these hurdles, Walter believes that computer science has brought us to the brink of a new agricultural revolution. In a 2017 PNAS Opinion piece, Walter and colleagues described emerging "smart farming" technologies—from autonomous weeding vehicles to moisture sensors in the soil (10). The authors worried, though, that only big industrial farms can afford these solutions. To make agriculture more sustainable, smaller farms in developing countries must have access as well.

Fortunately, "smart breeding" advances may have wider reach. Once image recognition technology becomes more developed for crops, which Walter expects will happen within the next 10 years, deploying it may be relatively inexpensive. Breeders could operate their own drones and obtain more precise ratings of traits like time to flowering or number of fruits in shorter time, says Walter. "The computing power that you need once you have established the algorithms is not very high."

The genomic data so vital to AI-led breeding programs is also becoming more accessible. "We're really at this point where genomics is cheap enough that you can apply these technologies to hundreds of species, maybe thousands," says Buckler.

Plant breeding has "entered the engineered phase," adds Tanksley. And with little time to spare. "The environment is changing," he says. "You have to have a faster breeding process to respond to that."

1 United Nations, Department of Economic and Social Affairs, Population Division, *World Population Prospects 2019: Highlights*, (United Nations, New York, 2019).

2 N. Jones, "Redrawing the map: How the world's climate zones are shifting" Yale Environment 360 (2018). https://e360.yale.edu/features/redrawing-the-map-how-the-worlds-climate-zones-are-shifting. Accessed 14 May 2020.

3 P. L. Pingali, Green revolution: Impacts, limits, and the path ahead. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 12302–12308 (2012).

4 D. Tilman, The greening of the green revolution. *Nature* **396**, 211–212 (1998).

5 G. P. Ramstein, S. E. Jensen, E. S. Buckler, Breaking the curse of dimensionality to identify causal variants in Breeding 4. *Theor. Appl. Genet.* **132**, 559–567 (2019).

6 D. Gonsalves, Control of papaya ringspot virus in papaya: A case study. *Annu. Rev. Phytopathol.* **36**, 415–437 (1998).

7 N. Kirchgessner *et al.*, The ETH field phenotyping platform FIP: A cable-suspended multi-sensor system. *Funct. Plant Biol.* **44**, 154–168 (2016).

8 K. Yu, N. Kirchgessner, C. Grieder, A. Walter, A. Hund, An image analysis pipeline for automated classification of imaging light conditions and for quantification of wheat canopy cover time series in field phenotyping. *Plant Methods* **13**, 15 (2017).

9 J. Streich *et al.*, Can exascale computing and explainable artificial intelligence applied to plant biology deliver on the United Nations sustainable development goals? *Curr. Opin. Biotechnol.* **61**, 217–225 (2020).

10 A. Walter, R. Finger, R. Huber, N. Buchmann, Opinion: Smart farming is key to developing sustainable agriculture. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 6148–6150 (2017).