The Promise of AI in Plant Breeding

How integrating the digital and biological worlds will help ensure food security, protect the planet and support farmers' livelihoods

Artificial intelligence is changing the world. That's not hyperbole; it's already happening. From improving manufacturers' production efficiency to quickly identifying financial fraud to significantly accelerating the complex act of drug discovery, AI is quickly revolutionizing a wide array of businesses.

Its greatest potential, however, may lie in its ability to advance one of society's oldest industries: agriculture. Seeing patterns where humans cannot and making accurate predictions out of vast data sets, AI is rapidly chipping away at what has been considered a holy grail of biology — understanding the causal relationships between genes and what they impact. Combined with the groundbreaking ability to perform precise gene edits, **it is helping turn the previously time-consuming craft of plant breeding into an efficient, data-driven process that will help ensure food security for a growing global population, while better caring for our planet and improving farmer well-being through economically sustainable solutions.**

MILLENNIA OF PROGRESS

Plant breeding has accomplished incredible feats since humans started to choose, then cross plants with features they liked more than 10,000 years ago. These historical actions initiated a process of domestication that would ultimately produce an array of crops that better met society's needs, many of which we still benefit from today. Not only has plant breeding helped ensure we can feed today's global population of 8.1 billion and counting, but together with advances in medicine, it has helped increase humans' life expectancy as overall poverty rates decline.

Speed, however, has never been plant breeding's strong suit. Breeders have to work through multiple iterations over multiple years to reach their goals, and even when successful, they would rarely know at a genetic level what produced the desired crop improvements. (There's a reason plant breeding is often referred to as more art than science.) Historically, this painstaking pace has not been a problem. **But with a global population expected to hit 11 billion by 2050 and climate change already disrupting long-standing temperature patterns, drying up centuries-old aquifers and causing more frequent extreme weather events, it's clear that plants need to evolve — and fast.**

That's where AI will play a critical role, helping to power a transition from breeding through selection to breeding through design. Instead of having to spend generations crossing, growing, checking, choosing and crossing again, breeders — with assistance from AI — can design a plant and develop it in just a few short years.

While this might have sounded impossible even just a decade or two ago, the convergence of three key breakthroughs makes it possible today:

- The discovery of CRISPR as a tool to precisely edit DNA
- A revolution in genomic data discovery that is unlocking volumes of new knowledge and insights
- The technology that can expertly parse this information and guide breeders to know exactly where and what to edit: AI

What AI can do for plant breeding is not all that different from what it has done for drug discovery in human health. By rapidly analyzing vast datasets to identify potential drug candidates, predicting their efficacy and safety, and optimizing chemical structures, **AI's deep learning capabilities have already significantly shortened a process that previously took more than a decade**. For example, the Massachusetts Institute of Technology in 2020 discovered a new antibiotic after a computer model screened more than 100 million chemical compounds — in a matter of days.[1]

There is one challenge to this comparison, however: AI success hinges on data, and compared to humans, plant genomic data is lagging far behind. Fortunately, AI's first big impact in plant breeding is in fixing this problem.

[1] news.mit.edu/2020/artificial-intelligence-identifies-new-antibiotic-0220

LEARN MORE

Scan the QR code to read our Multiplex Gene Editing white paper

FILLING THE KNOWLEDGE GAP

There are several reasons we know comparatively little about plants' inner workings.

First, as humans, we've (understandably) prioritized genomic research into our own species. The second is that plant genomes are simply much more complicated. **The wheat genome, for example, contains about 120,000 genes — six times the roughly 20,000 genes in the human genome.**

Third, and perhaps most significant, is that the characteristics that make plants more resilient are incredibly complex, influenced not by individual genes but by the interactions of vast gene networks as well as the surrounding environment. Some human diseases are being cured with single-gene edits; understanding and improving what determines a plant's water use efficiency will never be that straightforward.

As mentioned earlier, it should also be noted that **connecting genes to a feature is neither linear nor obvious, and as a result, any associated research has historically been incredibly time-consuming.** (The most common approach is colloquially referred to as the "one gene, one protein, one graduate student" model, and it involves scientists typically spending years characterizing one gene at a time. Try applying that to all 120,000 wheat genes.)

Although the data gap presents a challenge for plant breeding, AI is itself already helping to fix the problem. Through its ability to synthesize vast amounts of research, recognize patterns and rapidly process large datasets, the technology has begun to generate all-new knowledge at an exponential rate.

Genomes

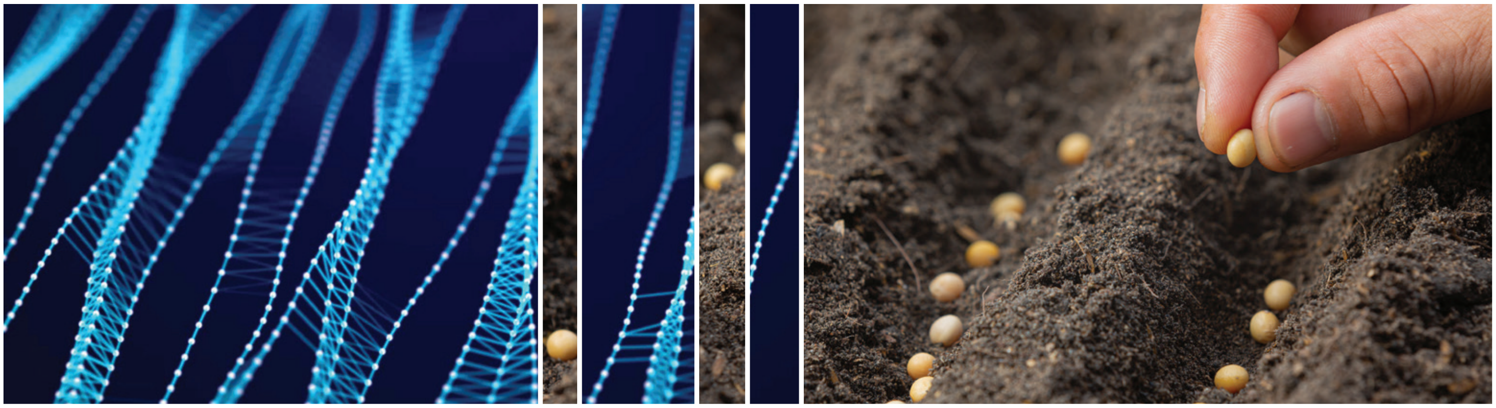
BY THE NUMBERS

~20K
genes in humans

~30K
genes in corn

~50K
genes in soybeans

~120K
genes in wheat



ACCELERATING DISCOVERY

Take the SEEDesign™ company Inari as an example. With an objective to increase the yields of crops such as soybeans by as much as 20%, its scientists are working on identifying the genetic changes most likely to produce this outcome. Relying solely on scientific literature, they were limited to just 50 soybean genes known to be somehow associated with yield-related characteristics. (Of the 50,000-gene soybean genome, literature has up to now only characterized the functions of about 200.)

So, taking a page out of the pharmaceutical industry's predictive genomics playbook, **they developed an AI-powered Predictive Design platform to analyze troves of genetics and so-called OMICS data from literature and proprietary research using state-of-the-art computational methods.** Rapidly evaluating insights from multiple genomic levels, Inari researchers were first able to identify more than 50 new yield-related gene candidates within a single year. Just 6 months later, with additional advanced AI-based methods, the platform pointed to 600 more.

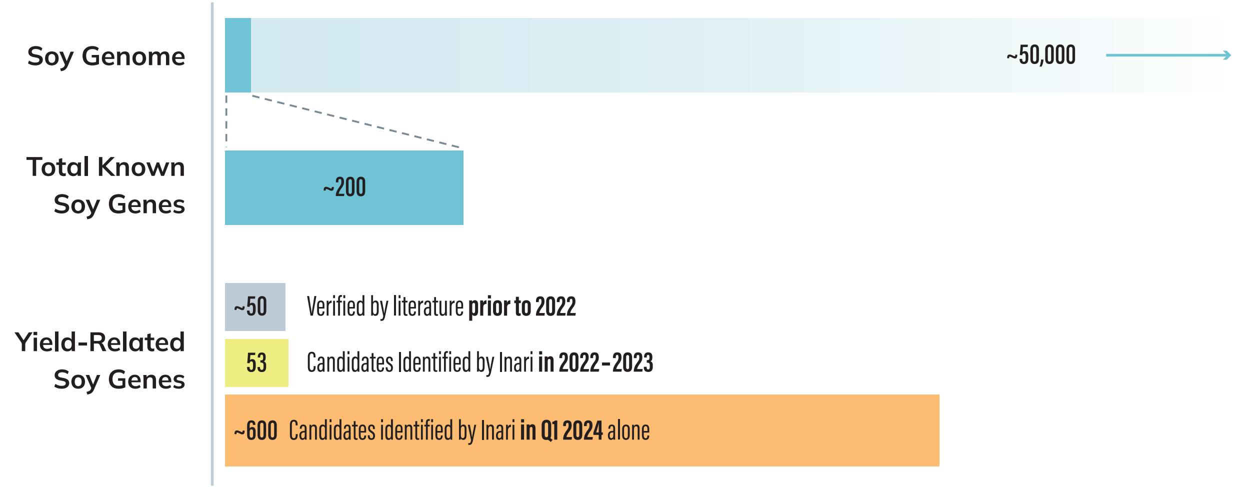
Even if only half of these genes are ultimately verified to impact yield, this would mean **AI helped to successfully increase knowledge of soybeans' genetic yield drivers more than six-fold in just 18 months**—a landmark in accelerated learning.

SAVING TIME AND RESOURCES

Once candidate genes have been identified, Inari's scientists run extensive tests to determine which genes should be edited in which manner to generate optimal results.

AI plays a critical role here as well, as the first round of experiments in the product design journey is run "in silico"—i.e. through computer modeling. Using different machine learning, deep learning, and Generative AI methods, these so-called "dry lab"

Example: AI-Powered Yield Gene Discovery

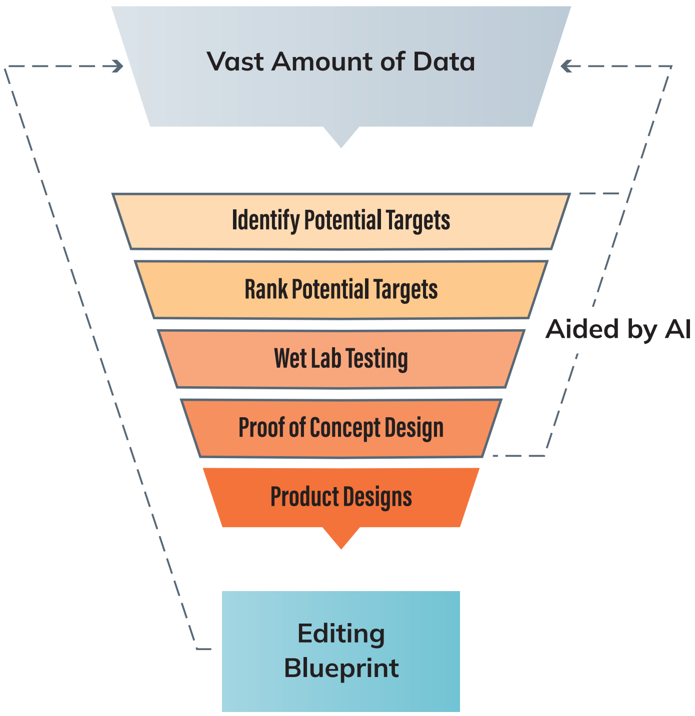


experiments offer an initial ranking of optimal gene editing targets and reduce the number of candidate genes to be tested in "wet lab" conditions. **By the time hypotheses move to biological testing, this digital filtering has helped ensure a high probability of success, meaning scientists are spending less time and resources on possibilities that likely would have failed later in the pipeline — another R&D gamechanger.**

The wet lab results, followed by proof-of-concept tests, determine which combinations of gene edits are provided to Inari's product team as blueprints for implementation into the production pipeline.

Once proven in greenhouse and field trials, the product can move toward commercialization, a mere 2-5 years after the initial "in silico" processes begin. **This is as little as one-third the time it typically takes for a traditionally bred product to come to market.**

The Journey from Tens of Thousands of Genes to Product Designs



ONLY THE BEGINNING

With AI's applications in biology still in relative infancy, new functions will only further accelerate the plant breeding process. Inari, for example, is using the technology to not only inform where and how to edit, but also improve the company's editing toolbox itself. Through AI-driven protein engineering, scientists are discovering and developing Cas enzymes that deliver new levels of accuracy and efficiency.

Other applications like AI chatbots are also being introduced to help scientists finetune their hypotheses. Just like ChatGPT is designed to generate human-like text based on a wide array of conversational contexts, Inari uses genomic language models that specialize in interpreting and predicting sequences of nucleotides within genetic data, enabling insights into biological functions and genetic variations.

Meanwhile, recognizing the value of high-quality data, Inari operates extensive, interconnected lab-to-field feedback loops that generate new insights at every step in both the discovery and design processes. **In collaboration with AWS, they are even developing techniques to overcome remaining data limitations using generative AI.**

This is only the first wave in the rapid evolution of AI-powered plant breeding. As both algorithms and data mature, systems will become more intelligent, efficient and effective — helping to ensure plant breeding can continue to support and secure a thriving future for humanity.

ABOUT INARI™

inari.com

Inari is the SEEDesign™ company, using new breeding technology to push the boundaries of what is possible by designing nature-positive seeds for a more sustainable food system. A combination of AI-powered predictive design and a cutting-edge multiplex gene editing toolbox enables Inari to unlock the full potential of seed and advance critical solutions with broad applications for growing more food with fewer resources. This includes products that will exponentially increase yield while reducing the environmental impact on land, water and nitrogen use.