# How supercomputers and plant genetics provide farmers better seeds

As genomic sequencing has decreased in price, the amount of data available to researchers has grown exponentially. But without advanced computing, the secrets of plant genetics would go largely untapped. As a result, more and more scientists are turning to computational tools to help organize and analyze genetic data. Managing the data overload could be key to feeding a growing world.

One plant biologist who's turned to advanced computing is Thomas Juenger, a faculty member in the Department of Integrative Biology of the University of Texas at Austin. He and his team wished to use existing data on exposure of mustard seed green (*Arabidopsis thaliana*) to cold and drought to test for a dozen environmental variables. Their research could help provide farmers with hardier plants that can adapt to climate change.

But their dataset included thousands of strains of the plant, which each had hundreds of thousands of genomic markers.

Another co-author of the study, Jesse Lasky, an Earth Institute fellow at Columbia University, explained: "to run these models across the genome, you quickly run out of time. It's really just a problem where you do lots of little things many, many times. It's much easier to accomplish that when you can run that problem on many cores across a cluster. That was the challenge."

Scientists in Australia faced similar data overload. Computational biologist Jill Gready, a professor at Australian National University's John Curtin School of Medical Research, wanted to improve food security by enriching the understanding of photosynthesis.

Gready is using the capacity of Australia's most powerful computer, Raijin, to hunt through existing seed bank data for plants that have efficiently working Rubisco, an enzyme that pulls carbon dioxide from the air facilitating the conversion of a plant's carbon into plant biomass.

"Computer simulation offers us the possibility to obtain detailed information on the conformation, reactions, interactions and other properties of proteins which is unobtainable by experimental means," she said.

For both projects, the work would be nearly impossible without supercomputing abilities.

Access to computer infrastructure powerful enough to handle massive datasets has been a problem for institutions that lack budgetary resources comparable to large multinational corporations.

In some cases, a lack of computational skills left researchers unable to select the right tool to solve problems facing farmers.

Now, a number of institutions are trying to fill that gap, especially for public or non-profit scientists.

Such is the case with the National Science Foundation, which formed the iPlant Collaborative in 2008 to

provide open-access tools, infrastructure and training to researchers in the U.S.

Juenger and Lasky received assistance from iPlant Collaborative to study *Arabidopsis*. Juenger's lab has studied *Arabidopsis* for over a decade. Considered the "lab rat" of plants, the mustard green was the first plant genome to be sequenced. Gene expression can differ widely in the plant that flourishes in a range of environments including Scandinavia, North Africa and Central Asia, making it ideal to study.

Because plants are rooted, they cope with environmental fluctuations like changes in moisture, temperature or insects by changing their gene expression. Juenger explained:

> As a plant starts to sense dropping temperatures, a cascade of gene expression can allow the plant to acclimatize to cold temperatures, and in effect prepare itself for the coming freezing conditions.

The study's results contribute to understanding about how plants evolve and could help plant breeders ange, which is crucial in securing the world's food supply.

The iPlant collaborative has also aided research in genomic variation of how tomato leaves respond in different light environments and how plants communicate on a molecular level using thousands of genes.

Beyond genetics, iPlant has helped forecast climate changes' impact on the growing season in Malawi and other African nations and link DNA crosses to cancer, to name a few.

Gready is pairing her computer-assisted search for genetic data with lab experiments to develop Rubisco that absorbs carbon dioxide more efficiently. She has already seen progress that could result in drought-tolerant crops. She explained:

> Our studies aim to find out why Rubisco is so inefficient, and to use this information to re-engineer it for improved efficiency. Even modest improvements offer major scope to enhance light, water and nutrient utilization by plants and, hence, to create higher-yielding food crops, to green deserts and to restore degraded landscapes. Another application could be improved tree

Rubiscos that can lock up more CO2 and fight climate change, or adaptation of suitable plants for sustainable high-yield biofuel production.

Gready's team isn't alone in counting on supercomputers to make photosynthesis more efficient. In a [paper](#) published in March in *Cell*, lead author Stephen Long, a plant biologist at the University of Illinois at Urbana-Champaign, recognized better supercomputing as the future of studying photosynthesis:

> We have unprecedented computational resources that allow us to model every stage of photosynthesis and determine where the bottlenecks are, and advances in genetic engineering will help us augment or circumvent those steps that impede efficiency.

Unprecedented computational resources will, ultimately, lead to a greater number of projects not yet imagined in the world of plant genetics. These projects could take on sizes previously thought of as unattainable for their gargantuan size, which makes advancements in supercomputing abilities and plant genetics an area of study to watch in the future.

**Rebecca Randall is a journalist focusing on global food and agriculture issues. Follow her [@beccawrites](#).**

**Additional resources:**

- [Photosynthesis upgrade proposed to raise crop yields](#), Scidev.net
- [Robots find genetic mutations in crops, accelerating breeding solutions](#), IEEE Spectrum