New tool helps annotate plant genes more quickly

June 11 2014, by Mickie Anderson

Since the first plant genome sequence was obtained for the plant Arabidopsis in 2000, scientists have gene-sequenced everything from cannabis to castor bean.

University of Florida scientists were part of a research team that this week unveiled a new tool that will help all plant scientists label ("annotate" in researcher parlance) genes far more quickly and accurately and is expected to give a big boost to traditional and nontraditional plant breeders.

Christopher Henry, a computational biologist at the University of Chicago who had a leading role in creating the database, called PlantSEED, said it is an important step toward the engineering of improved crops, such as creating rice that grows more efficiently or is more drought resistant.

Or creating perennial corn.

"Imagine if you didn't have to plant seeds for crops – if crops were just like your flowers and your maize just came up year after year," he said.

Andrew Hanson, a UF eminent scholar in horticultural sciences, said he believes PlantSEED – the capstone of the team's three-year effort – will prove even more of a boon to traditional breeders and should help them create better cultivars, faster.

"It's really the future. It'll be a new tool in the hands of the next generation of plant breeders, just as similar tools for bacteria are now widely used in microbial metabolic engineering," he said.

While scientists have been documenting and annotating genome sequences for plants at an amazing clip since breaking through with Arabidopsis, the work has not been without challenges.

In documenting genome sequences, scientists must sort through millions of bits of genetic code to identify what function each gene is responsible for (such as telling a plant how tall to grow or how to transport an amino acid throughout the plant). They base those identifications on evidence from previous studies.

That can be an imperfect process, Hanson said, because with 20,000 to 30,000 genes in a typical plant, scientists can't possibly conduct experiments to find out what each and every gene is responsible for. And they don't.

That is where the team's PlantSEED system comes in.

The open-access system, described in a paper published online Monday by *Proceedings of the National Academy of Sciences*, integrates data from plant scientists around the world into a common platform, which should result in better, more quickly-updated plant models for everyone using them.

PlantSEED will help plant scientists begin to make better use of genome information by helping them create consistently accurate models for all plant genomes contained in the database.

Hanson, a faculty member with UF's Institute of Food and Agricultural Sciences, likens the new tool to models aeronautical engineers use when

testing new equipment. They don't build a brand new jet every time they want to test a new material, but instead, test it by plugging information into computer models.

Because of tools like PlantSEED, plant scientists will eventually be able to do the same, he said.

"You can't really make as much use of the genome information as we should be able to until you can do that kind of modeling for plants, as well," he said. "And that's pretty much what this project is about."

Provided by University of Florida