

# Big data and the science of the Christmas tree

December 18 2014

---



Jill Wegrzyn, assistant research professor of ecology and evolutionary biology. Wegrzyn is helping develop bioinformatics tools that will enable more scientists to find meaning in massive amounts of data, such as those on crop production. Credit: Sean Flynn/UConn Photo

Often called the "Cadillac of Christmas trees," the Fraser Fir has everything a good Christmas tree should have: an even triangular shape, a sweet piney fragrance, and soft needles that (mostly) stay attached and won't leave tiny stabs in your fingers.

But even Frasers eventually turn, and by the New Year, what was once a beautiful sapling has started to smell like decomposing wood and litter its needles across your living room floor.

So scientists who refine the breeding of these and other practically perfect crops are always looking for new ways to understand how trees grow best.

Now, a federally funded initiative at UConn and partner universities will make it easy for plant scientists and other researchers to do just this using "big data."

By developing software that will connect genetic, physical, and environmental data housed in more than 15 major plant databases, assistant research professor Jill Wegrzyn of the Department of Ecology and Evolutionary Biology in the College of Liberal Arts and Sciences and her colleagues will create tools not only to benefit crop science, but to help address important ecological issues like reforestation and climate change.

"It's one of the ongoing hurdles of data analysis. As a scientist, I might be aware of these important data sources, but they are in different formats and locations, and are often much too large for a single desktop machine to analyze," says Wegrzyn. "So the question becomes: How do we enable access so that more scientists can start finding meaning in these massive amounts of data? My job as a bioinformaticist is to help biologists achieve this in the era of next-generation sequencing and high-throughput phenotyping."



Researchers now use drone technology to survey forests and orchards. Credit: Sean Flynn/UConn Photo

Wegrzyn and her colleagues, including project principal investigator Stephen Ficklin at Washington State University, recently received \$1.5 million from the National Science Foundation to develop a cyber infrastructure, called Tripal Gateway, that will allow scientists to access, visualize, and analyze data anywhere in the world. The infrastructure builds on Tripal, an existing open-source toolkit designed to assist with the construction of online genomic and genetic databases.

The infrastructure will serve thousands of scientists from industries, universities, and nonprofits worldwide, and is part of a \$31 million program of the NSF Data Infrastructure Building Blocks program.

Many people think of [big data](#) in the life sciences as solely genetic information, but Wegrzyn points out that many scientific databases also contain large amounts of phenotypic data, or information about the physical attributes of organisms, as well as environmental data.

For example, researchers now use drone technology to survey and monitor forests and orchards. Combining this information with [environmental data](#) such as soil and climatic conditions, as well as genetic information, says Wegrzyn, can help scientists understand big questions, like how a forest's biodiversity is changing under climate change, or what individuals within a given species will survive in a reforested landscape.

The databases, which have names like CottonGen, The Citrus Genome Database, PeanutBase, and TreeGenes, are housed at universities and government agencies such as the U.S. Department of Agriculture. Wegrzyn serves as curator of TreeGenes, which is hosted jointly at UConn and the University of California, Davis, serves more than 2,000 researchers worldwide and has information on 1,200 tree species, including Christmas tree species like the famed Fraser Fir.

In another current project under the USDA Specialty Crop Initiative, she and lead investigator John Frampton of North Carolina State University will mine TreeGenes for associations of genes with traits like needle retention and disease resistance.

"Trees live a long time and provide a great genetic record," she says. "These tools can help us understand what genes contribute to specific traits."

Wegrzyn is also the lead scientist at the UConn Bioinformatics Facility, which is housed in the Institute for Systems Genomics. There she serves as a resource for UConn students and researchers in Storrs and in

Farmington at UConn Health. Currently she has ongoing collaborations with College of Liberal Arts and Sciences faculty in the departments of ecology and evolutionary biology and molecular and cell biology, and in the College of Agriculture, Health, and Natural Resources' plant science and animal science departments, among others.

"The most exciting aspect is to enable research through these tools," Wegrzyn says. "This project will develop an infrastructure that will contribute to discoveries in ecology, conservation, and agriculture."

Provided by University of Connecticut

Citation: Big data and the science of the Christmas tree (2014, December 18) retrieved 7 August 2024 from <https://phys.org/news/2014-12-big-science-christmas-tree.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.