

Scientists decode newly sequenced strawberry genome

December 26 2010

An international research consortium has sequenced the genome of the woodland strawberry, according to a study published in the Dec. 26 advance online edition of the journal *Nature Genetics*. The development is expected to unlock possibilities for breeding tastier, hardier varieties of the berry and other crops in its family.

"We've created the strawberry parts list," said the consortium's leader Kevin Folta, an associate professor with the University of Florida's Institute of Food and Agricultural Sciences. "For every organism on the planet, if you're going to try to do any advanced science or use molecular-assisted breeding, a parts list is really helpful. In the old days, we had to go out and figure out what the parts were. Now we know the components that make up the strawberry plant."

From a genetic standpoint, the woodland strawberry, formally known as *Fragaria vesca*, is similar to the cultivated strawberry but less complex, making it easier for scientists to study. The 14-chromosome woodland strawberry has one of the smallest genomes of economically significant plants, but still contains approximately 240 million base pairs.

The consortium of 75 researchers from 38 institutions that sequenced the [genome](#) included two Georgia Tech researchers. They are Mark Borodovsky, a Regents professor with a joint appointment in the Wallace H. Coulter Department of Biomedical Engineering at Georgia Tech and Emory University and the Georgia Tech School of Computational Science and Engineering, and Paul Burns, who worked

on the project as a bioinformatics Ph.D. student.

Once the consortium uncovered the [genomic sequence](#) of the woodland strawberry, Borodovsky and Burns led the efforts in identifying protein-coding genes in the sequence. Using a newly developed pattern recognition program called GeneMark.hmm-ES+, Borodovsky and Burns identified 34,809 genes, of which 55 percent were assigned to gene families.

The GeneMark.hmm-ES+ program iteratively identified the correct algorithm parameters from the DNA sequence and transcriptome data. The program used a probabilistic model called the Hidden Markov Model to pinpoint the boundaries between coding sequences -- called exons -- and non-coding sequences, which could be either introns or intergenic regions.

In identifying the genes, prediction and training steps were repeated, each time detecting a larger set of true coding and non-coding sequences used to further improve the model employed in statistical pattern recognition. When the new sequence breakdown coincided with the previous one, the researchers recorded their final set of predicted genes.

"GeneMark.hmm-ES+ is a hybrid program that uses both DNA and RNA sequences to predict protein-coding genes," said Borodovsky, who is also director of Georgia Tech's Center for [Bioinformatics](#) and Computational Genomics.

Borodovsky developed the first version of GeneMark in 1993. In 1995, this program was used to find genes in the first completely sequenced genomes of bacteria and archaea. The research team then developed self-training versions of the gene finding program for prokaryotic (organisms that lack a cell nucleus) and eukaryotic (organisms that contain a cell nucleus) genomes in 2001 and 2005, respectively. Development of these

programs has been supported by the National Institutes of Health since 1993.

Most recently, Borodovsky's team predicted genes in the genomes of the green alga *Chlorella variabilis* NC64A and the mushroom *Coprinopsis cinerea*, with reports published in 2010 in the journals *The Plant Cell* and *Proceedings of the National Academy of Sciences*, respectively.

"Our approach to gene prediction in the strawberry genome proved highly effective, with 90 percent of the genes predicted by the hybrid gene model supported by transcript-based evidence," added Borodovsky.

Further analysis of the woodland strawberry genome revealed genes involved in key biological processes, such as flavor production, flowering and response to disease. Additional examination also revealed a core set of signal transduction elements shared between the strawberry and other plants.

The woodland strawberry is a member of the Rosaceae family, which consists of more than 100 genera and 3,000 species. This large family includes many economically important and popular fruit, nut, ornamental and woody crops, including the cultivated strawberry, almond, apple, peach, cherry, raspberry and rose.

In the long term, breeders will be able to use the information to create plants that can be grown with less environmental impact, better nutritional profiles and larger yields.

"The wealth of genetic information collected by this strawberry genome sequencing project will help spur the next wave of research into the improvement of [strawberry](#) and other fruit crops," added Borodovsky.

Provided by Georgia Institute of Technology

Citation: Scientists decode newly sequenced strawberry genome (2010, December 26) retrieved 25 April 2024 from

<https://phys.org/news/2010-12-scientists-decode-newly-sequenced-strawberry.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.