

What's in your wheat? Scientists piece together genome of most common bread wheat

November 20 2017



Credit: CC0 Public Domain

Johns Hopkins scientists report they have successfully used two separate gene technologies to assemble the most complete genome sequence to date of *Triticum aestivum*, the most common cultivated species of wheat used to make bread.

A report on the achievement was published in the Oct. 23 issue of *GigaScience* just a few weeks before their related report on the sequencing of the bread wheat's "ancestor," *Aegilops tauschii*, published Nov. 15 in *Nature*.

Together, they say, the wheat genome sequences may help biologists not only better understand the evolutionary history of wheat, but also advance the quest for hardier, more pest- and drought-resistant wheat types to help feed the world's growing population.

"After many years of trying, we've finally been able to produce a high-quality assembly of this very challenging genome," says Steven Salzberg, Ph.D., Bloomberg Distinguished Professor of Biomedical Engineering at the Johns Hopkins University Whiting School of Engineering and the McKusick-Nathans Institute of Genetic Medicine at the Johns Hopkins University School of Medicine.

According to the Johns Hopkins scientists, bread wheat has one of the most complex genomes known to science, containing an estimated 16 billion base pairs of DNA and six copies of seven chromosomes. By comparison, the human genome is about five times smaller, with about three billion base pairs and two copies of 23 chromosomes. Previously published versions of the bread wheat genome have contained large gaps in its highly repetitive DNA sequence.

"The repetitive nature of this genome makes it difficult to fully sequence," says Salzberg. "It's like trying to put together a jigsaw puzzle of a landscape scene with a huge blue sky. There are lots of very similar,

small pieces to assemble."

The newly assembled bread wheat genome, which cost \$300,000 for the sequencing alone, took a year for the Johns Hopkins researchers to assemble 1.5 trillion bases of raw data into a final assembly of 15.34 billion base pairs.

To do it, Salzberg and his team used two types of genome sequencing technology: high throughput short-read sequencing and long-read, single molecule sequencing. As its name implies, high throughput sequencing generates massive amounts of DNA base pairs very quickly and cheaply, although the fragments are very short—just 150 base pairs long for this project. To help assemble the repetitive areas, the Johns Hopkins team used real-time, single molecule sequencing, which reads DNA as it is being synthesized in a tiny, nano-scale well on a chip. The technology enables scientists to read up to 20,000 base pairs at a time by measuring fluorescent signals that are emitted as each DNA base is copied.

Salzberg says that sequencing a genome of this size requires not only genetic expertise, but also very large computing resources available at relatively few research institutions around the world. The team relied heavily on the Maryland Advanced Research Computing Center, a computing center shared by Hopkins and the University of Maryland, which has over 20,000 computer cores (CPUs) and over 20 petabytes of data storage. The team used approximately 100 CPU years to put this genome together.

Salzberg and his team also participated in the collaborative effort reported in the journal *Nature* to sequence an ancestral type of wheat, *Aegilops tauschii*, which is commonly referred to as goatgrass and still found in parts of Asia and Europe. Its genome is approximately one-third the size of the bread wheat genome, but has similar levels of repetition. The work, done as part of a collaborative effort between the

University of California, Davis; Johns Hopkins; and the University of Georgia, took approximately four years to complete. Using ordered-clone genome sequencing, shotgun sequencing and optical genome mapping, the team pieced together the 4.3 billion nucleotides that make up the plant's genetic sequence. With this information, the rest of the team was able to identify sequences that make up the genes responsible for specific characteristics in the plant.

More information: Aleksey V Zimin et al. The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum*, *GigaScience* (2017). [DOI: 10.1093/gigascience/gix097](https://doi.org/10.1093/gigascience/gix097)

Ming-Cheng Luo et al. Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*, *Nature* (2017). [DOI: 10.1038/nature24486](https://doi.org/10.1038/nature24486)

Provided by Johns Hopkins University School of Medicine

Citation: What's in your wheat? Scientists piece together genome of most common bread wheat (2017, November 20) retrieved 29 April 2024 from <https://phys.org/news/2017-11-wheat-scientists-piece-genome-common.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.