Small group scoops international effort to sequence huge wheat genome

Just six scientists conquer one of the most complicated genomes ever read.

Ewen Callaway

31 October 2017



Nico van Kappel/Minden Pictures/Getty

The genome of wheat (Triticum aestivum) is huge, and full of repetitive sequences.

The wheat genome is finally complete. A giant international consortium of academics and companies has been trying to finish the challenging DNA sequence for more than a decade, but in the end, it was a small US-led team that scooped the prize. Researchers hope that the genome of bread wheat (*Triticum aestivum*) — described in the journal *GigaScience* this month[1] — will aid efforts to study and improve a staple crop on which around 2 billion people rely.

The wheat genome is crop geneticists' Mount Everest. It is huge — more than five times the size of a single copy of the human genome — and harbours six copies of each chromosome, adding up to between 16 billion and 17 billion letters of DNA. And more than 80% of it is made of repetitive sequences. These stretches are especially vexing for scientists trying to assemble the short DNA segments generated by sequencing machines into much longer chromosome sequences.

It's like putting together a jigsaw puzzle filled with pieces of blue sky, says Steven Salzberg, a genomicist at Johns Hopkins University in Baltimore, Maryland, who led the latest sequencing effort. "The wheat genome is full of blue sky. All these pieces look like a lot of other pieces, but they're not exactly alike."

As a result, previous wheat-genome sequences contained gaps that made it hard for scientists to locate and examine any particular gene, says Klaus Mayer, a plant genomicist at the Helmholtz Center in Munich, Germany, and one of 1,800 members of the International Wheat Genome Sequencing Consortium (IWGSC) that have been tackling the genome since 2005.

Related stories

- Devastating wheat fungus appears in Asia for first time
- Ancient DNA reveals how wheat came to prehistoric Britain
- Fiendish wheat genome reveals grain's history

More related stories

Devastating wheat

Related stories

- fungus appears in Asia for first time
- Ancient DNA reveals how wheat came to prehistoric Britain
- Fiendish wheat genome reveals grain's history

More related stories

A sequence released by the consortium in 2014 covered about two-thirds of the genome, but it was highly fragmented and lacked details about the sequences between genes². Improved versions were released in 2016 and 2017, but the use of these data is restricted until the IWGSC publishes its analysis (Mayer says the team is preparing to submit its report to a journal). The sequence was also produced using proprietary software from a company called NRGene, preventing other scientists from reproducing the effort.

Puzzle pieces

Salzberg, who specializes in assembling genome sequences, and his five colleagues decided to tackle the problem themselves. To overcome the challenge of ordering repetitive DNA — the puzzle pieces of blue sky — the researchers used a sequencing technology that generates very long DNA stretches

(often in excess of 10,000 DNA letters). They also created much shorter, but highly accurate sequences, using another technology.

Stitching these 'reads' together — which amounted to 1.5 trillion DNA letters and consumed 880,000 hours of processor time on a cluster of parallel computers — resulted in nearly continuous chromosome sequences that encompassed 15.3 billion letters of the wheat genome.

Mayer calls the new sequence "a major leap forward". Postdocs can spend whole fellowships locating a single wheat gene of interest, he says. "Those genes which took 10 man- or woman-years to clone, this will melt down to a couple of months, hopefully." The results of such research should help breeders to develop strains of wheat that are better able to tolerate climate change, disease and other stresses.

Some scientists are already using the new wheat genome — including, Salzberg says, members of the IWGSC working on one particular chromosome. But if it is to be of widespread use, all of the genes and sequences will need to be identified and labelled, a laborious process known as annotation. Salzberg says that a collaborator of his is planning to do this, "unless someone does it sooner".

Neil Hall, a genomicist and director of the Earlham Institute, a genomics research centre in Norwich, UK, sees Salzberg's approach as a sign of the times. If the wheat genome — considered one of the most complicated to be tackled by scientists — can be sequenced by a small team using the latest technology, almost any genome could.

"I think we've moved beyond the era where genome projects have to be these monolithic international cooperations," Hall says. "Genomics is more like the gig economy now."

Nature doi:10.1038/nature.2017.22924

References

- 1. Zimin, A. V. et al. GigaScience gix097 (2017).
- The International Wheat Genome Sequencing Consortium (IWGSC) Science 345, 1251788 (2014).