Important milestone in wheat research announced

November 12 2015

A more complete and accurate wheat genome assembly is being made available to researchers, by The Genome Analysis Centre (TGAC). This landmark resource builds on international efforts in wheat genomics and will help wheat breeders accelerate their crop improvement programs and researchers to discover genes for key traits such as yield, nutrient use and bread making quality.

A more complete and accurate wheat genome assembly is being made available to researchers, by The Genome Analysis Centre on Nov. 12, 2015. This landmark resource builds on international efforts in this area and will help wheat breeders accelerate their crop improvement programs and researchers to discover genes for key traits such as yield, nutrient use and bread making quality. As wheat is one of the world's most vital crops, the new genomics resources will help secure future food supplies.

The wheat genome is now assembled into fewer and much larger chunks of DNA and covers regions that previous assemblies did not reach, such as complicated highly repetitive regions that form about 80 per cent of the DNA sequences. "Furthermore", said Matt Clark, Group Leader at TGAC (Co-Principal Investigator on the grant), who led the sequencing work, "wheat has a very large and complex genome made by the hybridisation of three closely related grasses, each of which has a large genome itself. It's has been a complex problem that has confounded scientists for several years."

Reaching this milestone has been a major UK-based effort to identify and understand wheat genes and develop insights into the links between them to aid breeding programmes. In this latest development, billions of bases needed to be sequenced and the assembly (a gigantic jigsaw puzzle using billions of pieces that are very similar to each other) took three weeks to complete on one of the UK's largest supercomputers, which was specially configured for work on wheat.

A more complete and accurate wheat genome assembly is being made available to researchers, by The Genome Analysis Centre (TGAC). This landmark resource builds on international efforts in wheat genomics and will help wheat breeders accelerate their crop improvement programs and researchers to discover genes for key traits such as yield, nutrient use and bread making quality.

To assemble the wheat genome, Bernardo Clavijo, Algorithms Research and Development Team Leader at TGAC, made major modifications to a software, called DISCOVAR, developed by the Broad Institute, Cambridge US, (previously used for specialist applications in human genome assembly) in a collaboration established by Federica Di Palma, Director of Science of TGAC and Visiting Scientist at the Broad

Institute. In order to ensure all the complexity of the DNA sequence was preserved during assembly, he made a series of major overhauls to the software: "We centred our approach on achieving maximal coverage of the genome, by distinguishing repeats. We were very careful to use newly generated high-quality input data."

These advances now mean the software can assemble several wheat genomes with high speed and great precision. This sets the stage for rapidly generating useful assemblies of many varieties of wheat, which is an essential step for breeding and research. Mike Bevan from the John Innes Centre (JIC) (Co-Principal Investigator), said: "The capacity to sequence and assemble many wheat genomes efficiently breaks down major barriers to wheat crop improvement. We will now be able to exploit genetic variation from ancestral wheat varieties for crop improvement in new ways."

Ksenia Krasileva, Group Leader at TGAC and TSL, who has conducted an initial assessment of the assemblies, said: "One of the most complex and large groups of genes in wheat are those that contribute to the nutritional and bread-making quality of the grain. These are all present in complete copies in the genome, suggesting other hard-to-assemble genes are also accurately represented."

Steve Visscher, Deputy Chief Executive of the Biotechnology and Biosciences Research Council (BBSRC), who funded the project, said: "BBSRC is delighted to have supported this work, which has made an important contribution to the G20-sponsored international Wheat Initiative. Many research groups are contributing to the global research effort to develop a fully assembled and aligned wheat genome sequence to access, understand and apply the richness of wheat genetic diversity to increase wheat yield, improve wheat's tolerance to stresses, pathogens and pests, and improve the sustainability wheat production. It is fitting that this important step in unravelling the complex wheat genome, which

is five times the size of the human genome, has adapted specialist software developed for the human genome assembly."

The early release of the data as a new resource for the world wheat researchers and breeders reflects the Wheat Initiative's founding principles of sharing data and seeking synergy through collaboration to help tackle the global grand challenge of feeding a population of nearly 10 billion by 2050. The data will be available for sequence searches (BLAST) at TGAC's Grassroot Genomics platform from November 12 2015. The full data set, with genes identified, will be publicly available from the European Bioinformatics Institute's (EBI) Ensembl database at the end of 2015. This is a key milestone in the BBSRC funded research project "Triticeae Genomics for Sustainable Agriculture" in collaboration with TGAC, JIC, the European Bioinformatics Institute and Rothamsted Research.

More information: www.tgac.ac.uk/grassroots-genomics/

Provided by The Genome Analysis Centre