

Wheat Genome Sequencing: Lessons Emerging

Vijeta Gupta¹ and Ekta Kamboj²

¹Dept. of Genetics & Plant Breeding, College of Agriculture, Chaudhary Charan Singh Haryana Agricultural University, Hisar, Haryana

²Dept. of Agronomy, College of Agriculture, Chaudhary Charan Singh Haryana Agricultural University, Hisar, Haryana

SUMMARY

Wheat is the most important crop in the world, yet it has one of the most challenging genomes. Bread wheat is a hexaploid, with three complete genomes termed A, B and D in the nucleus of each cell. Wheat has lagged behind other species, primarily owing to the challenges of assembling a large (haploid genome, 1C = 17 Gb) and complex genome that contains more than 85% repetitive DNA. A chromosome survey sequence (CSS) intermediate product assigned 124,201 gene loci across the 21 chromosomes and revealed the evolutionary dynamics of the wheat genome through gene loss, gain, and duplication.

INTRODUCTION

Sequencing the wheat genome has long been considered an insurmountable challenge, due to the high complexity of the wheat genome. Less genomic conservation between rice and wheat has also restricted comparative genomic studies for genetic enhancement of wheat. But improving average wheat yields has become a major objective with genome sequencing as its prerequisite. Significant advances in the understanding of the wheat plant and grain biology must be achieved to increase absolute yields and protect the crop from an estimated average annual loss of 25% caused by biotic (pests) and abiotic stresses (heat, frost, drought, and salinity). Genome sequencing is a widely accepted mechanism for accelerating achievement of these objectives and enables more rapid genetic improvement. In addition to food security, wheat genome sequencing will lead to improved human health and nutrition. The availability of high quality reference genome would accelerate the breeding of climate-resilient wheat varieties to feed the ever-increasing world population.

The International Wheat Genome Sequencing Consortium (IWGSC) was created in 2005 by a group of

- Wheat growers
- Plant scientists
- Public and private breeders to change this paradigm.

Today, the international public-private collaborative consortium has more than 1500 members in 60 countries. The goal of the IWGSC is to make a high quality genome sequence of the bread wheat cv. *Chinese Spring* publicly available, in order to serve as a foundation for the accelerated development of improved varieties and to empower all aspects of basic and applied wheat science.

An international wheat genome project could be established through the following steps:

- Constructing an accurate, sequence-ready, global physical (BAC-contig) map anchored to the high-resolution genetic and deletion maps of the 21 chromosomes of the hexaploid wheat genotype Chinese Spring.
- Exploring the use of flow-sorted chromosome- and arm-specific libraries in the assembly of the global physical map and in preparation for the sequencing of the gene-containing regions of homeologous chromosome groups.
- Identifying genomic sequence tags using gene-enrichment procedures such as ESTs and full-length cDNAs of 2x, 4x, and 6x wheat for an accurate estimation of the wheat unigene set.
- Leveraging rice sequence and wheat-rice gene synteny, comparative genetics, and wheat unigenes toward the development of high-resolution genetic and deletion maps of the 21 chromosomes of Chinese Spring wheat.
- Sample sequencing of BACs from different ploidy wheats and genotypes should also be undertaken.
- Integrating bioinformatics at every step for project management, data analysis, improved methods of sequence annotation, and dissemination of data.
- Engaging all wheat stakeholders and educational institutions globally, especially in developing countries, and locally in all aspects of the research, technology transfer, workforce training, and promotion of science.
- Maintaining all data, materials, and resources in the public domain and free of intellectual property rights.

- Organizing an international steering committee to coordinate and execute all aspects of the wheat genome sequencing project.

Milestones in Wheat Genome Sequencing

- The first milestone was reached on July 2014 with the publication of the chromosome-based draft genome sequence in the journal *Science*.
- The physical maps for all chromosomes (milestone 2) were completed by the end of 2015.
- In June 2016, a whole genome shotgun assembly (IWGSC WGA v0.4) was made available pre-publication. The whole genome assembly was subsequently integrated with physical maps and other chromosome-based sequence resources to generate the first version of the chromosome-based reference sequence (RefSeq v1.0), which was made available pre-publication in January 2017 (milestone 3 & 4).
- Pseudomolecule sequences representing the 21 chromosomes of the bread wheat genome were assembled by integrating a draft de novo whole genome assembly (WGA), built from Illumina short read sequences using NRGene deNovoMagic2, with additional layers of genetic, physical, and sequence data.

In the resulting 14.5-Gb genome assembly, contigs and scaffolds with N50s of 52 kb and 7 Mb, respectively, were linked into super scaffolds (N50 = 22.8 Mb), with 97% (14.1 Gb) of the sequences assigned and ordered along the 21 chromosomes and almost all of the assigned sequence scaffolds oriented relative to each other (13.8 Gb, 98%). Unanchored scaffolds comprising 481 Mb (2.8% of the assembly length) formed the “unassigned chromosome” (ChrUn) bin

SUMMARY

The world's first genome sequence on bread wheat has been completed with the help of Chinese scientists. The International Wheat Genome Sequencing Consortium successfully mapped the genome of "hexaploid bread wheat with an annotated reference sequence," according to a study released in the US-based academic journal *Science* on Friday. It marked the first time a hexaploid wheat genome sequence was completed, *People's Daily* reported Monday. The Chinese team physically mapped the chromosome 7DL and analyzed its sequence. They also participated in analysis of the entire genome. Wheat is the last important crop to see its genome fully sequenced. The research will help wheat adapt to dry land and saline-alkali soil as well as increase overall production in the future and the sequencing can also contribute to wheat's resistance to diseases and pests. The IWGSC will now focus on producing a genome-sequence based toolbox for breeders and scientists to use for wheat improvement. It will involve several projects, such as maintaining and improving the current reference genome to ultimately produce a “Gold Standard” reference genome sequence that is manually and functionally annotated; sequencing other varieties of wheat in order to represent the worldwide diversity of wheat; and continuing to develop a database for the wheat community to access all these genomic resources.

REFERENCES

- Guan, J., Garcia, D.F., Zhou, Y., Appels, R., Li, A. and Mao, L., 2020. The Battle to Sequence the Bread Wheat Genome: A Tale of the Three Kingdoms. *Genomics, proteomics & bioinformatics*.
- Moolhuijzen, P., Dunn, D.S., Bellgard, M., Carter, M., Jia, J., Kong, X., Gill, B.S., Feuillet, C., Breen, J. and Appels, R., 2007. Wheat genome structure and function: genome sequence data and the International Wheat Genome Sequencing Consortium. *Australian Journal of Agricultural Research*, 58(6), pp.470-475.