

Wheat genome sequencing: a milestone in cereal genomics and its future potential

The cultivated cereals, including rice, maize, wheat, barley, rye, sorghum, oats and millets cater about half of the calories and proteins to the world population and are emerging purveyors of biofuel. In addition to their economic significance, cereals are a prominent choice for comparative genomics as they comprise important crops with diverse native distribution and at least 35-fold variation in genome size (e.g. rice = 420 Mb; wheat = 16,000 Mb). Amongst cereals, wheat (*Triticum* sp.) is the largest cultivated crop (210 million hectares (m ha)¹) produced in a broad-spectrum climatic environments and geographic regions. Bread wheat (*Triticum aestivum*), being a rich source of protein compared to other cereals, constitutes the staple food of the world since bread wheat's domestication ~10,000 years ago² and hence its improvement has vast implications for global food security.

Compared to other cereals, bread wheat genetics is intricate as the genome ($2n = 6x = 42$, AABBDD; ~17 Gb: five times the size of the human genome) is allohexaploid, originating from hybridization between cultivated tetraploid emmer wheat (AABB, *Triticum dicoccoides*) and diploid goat grass (DD, *Aegilops tauschii*) approximately 8000 years ago³. Several molecular breeding and transgene-based approaches have been underway to increase the wheat production through broadening the genetic diversity and examine the key traits. A landmark among these efforts is the release of bread wheat genome sequence⁴. Random shotgun libraries of total DNA and cDNA from *T. aestivum* cv. Chinese Spring (CS42) were sequenced in Roche 454 pyrosequencer using GS FLX titanium and GS FLX + platforms to generate 85 Gb of sequence (220 million reads), equivalent to 5X genome coverage and identified between 94,000 and 96,000 genes⁴. The whole genome sequence can be retrieved from EMBL/GenBank under SRA study at <http://www.ebi.ac.uk/ena/data/view/ERP000319&display=html>.

An orthologous set of representative grass genes, including *Brachypodium distachyon*, sorghum, rice and barley was chosen to map and assemble wheat raw

reads on each orthologous group. The high-resolution synteny maps showed many small interruptions to conserved gene order. Brenchley *et al.*⁴ had also shown that the hexaploid genome is extremely dynamic, with considerable loss of gene family components on polyploidization and domestication, and several gene segments. Generating elite crops with agronomically desirable traits through marker-assisted selection (MAS) is well-accepted from times. The utility of microsatellite markers in MAS is impeccable. Recently, single nucleotide polymorphisms (SNPs) have become the marker of choice in MAS due to their ease of use and scoring and their ability to be automated with relative ease⁵. Brenchley *et al.*⁴ reported about 132,000 SNPs in A, B and D genomes, which facilitates the analysis of QTLs controlling desirable traits such as disease resistance, drought tolerance and yield in wheat.

Though wheat is the largest cultivated cereal globally, maize ranks first in terms of total production. In 2009, wheat was positioned second in global production (685.61 million tonnes; mt) next to maize (818.82 mt), but in 2010, rice (696.32 mt) was ranked second and wheat third (653.65 mt)¹. The latest forecast of the Global Information and Early Warning System (GIEWS) predicts that the world wheat inventories are expected to fall to 163 mt, down 11% from their opening level and 2% (3 mt) less than that reported⁶ in November 2012. This alarming reduction in productivity threatens the food security and provokes the importance of rigorous agricultural research. As an antidote, the wheat whole genome sequencing (WGS) provides direct access to all 96,000 genes and represents an essential step towards a systematic understanding of biology and engineering the cereal crop for valuable traits. Its implications in cereal genetics and breeding include the examination of genome variation, association mapping using natural populations, performing wide crosses and alien introgression, studying the expression and nucleotide polymorphism in transcriptomes, analysing population genetics and evolutionary biology, and studying the epigenetic modifica-

tions. Moreover, the availability of large-scale genetic markers generated through Next Generation Sequencing (NGS) technology will facilitate trait mapping and make marker-assisted breeding more feasible.

Moreover, the WGS data not only facilitate in deciphering the complex phenomena such as heterosis and epigenetics, it may also enable breeders to predict which fragment of a chromosome is derived from which parent in the progeny line, thereby recognizing crossover events occurring in every progeny line and inserting markers on genetic and physical maps without ambiguity. In due course, this will assist in introducing specific chromosomal segments from one cultivar to another. Besides, Brenchley *et al.*³ had identified diverse classes of genes participating in energy production, metabolism and growth that were probably linked with crop yield, which can now be utilized for the development of transgenic wheat. Thus whole genome sequence of wheat and the availability of thousands of SNPs will inevitably permit the breeders to stride towards identifying novel traits, providing biological knowledge and empowering biodiversity-based breeding.

1. Food and Agriculture Organization of the United Nations, FAOSTAT, Report, 2010; <http://faostat3.fao.org/home/index.html#-VISUALIZE>
2. Gupta, P. K., Mir, R. R., Mohan, A. and Kumar, J., *Int. J. Plant Genomics*, 2008, doi: 10.1155/2008/896451.
3. Peng, J. H., Sun, D. and Nevo, E., *Mol. Breed.*, 2011, **28**, 281–301.
4. Brenchley, R. *et al.*, *Nature*, 2012, **491**, 705–710.
5. Liu, Y., He, Z., Appels, R. and Xia, X., *Theor. Appl. Genet.*, 2012, **125**, 1–10.
6. Food and Agriculture Organization of the United Nations, Global information and early warning system, crop prospects and food situation, No. 4, December 2012; <http://www.fao.org/giews/english/cpfs/index.htm>

Mehanathan Muthamilarasan and **Manoj Prasad***, National Institute of Plant Genome Research, Aruna Asaf Ali Marg, New Delhi 110 067, India.

*e-mail: manoj_prasad@nipgr.ac.in