

Wheat genome sequence completed

— Towards accelerating the development of new improved varieties —

An international consortium, to which NARO and Kyoto University participate, has completed the sequencing of the wheat genome. The high-quality genome sequence corresponding to the 21 chromosomes of wheat revealed over 100,000 genes that determine various traits. The genome sequence will serve as the foundation for isolation of useful genes and identification of DNA markers towards the development of new improved varieties of wheat.

Overview

The International Wheat Genome Sequencing Consortium (IWGSC), to which the National Agriculture and Food Research Organization (NARO) and Kyoto University participate, has succeeded in deciphering the sequence of the wheat genome.

The IWGSC was established in 2005 with the aim of sequencing the genome of wheat, one of the world's three leading food crops that include rice and corn. In 2014, a chromosome-based draft sequence was published. However, since the draft sequence consisted of fine-scale sequenced fragments with no information on the location of genes in specific chromosomes, it is highly desirable to elucidate an accurate reference genome sequence assembled into individual chromosomes.

This time, the IWGSC has successfully achieved the decoding of the high-quality reference sequence covering 94% of the wheat genome, identified a total of 107,891 genes, and provided information on genes involved in various traits such as disease resistance and quality of wheat flour.

Wheat accounts for about 20% of the food calories consumed by the human population all over the world. The completed genome sequence will provide the foundation for advancement of research in crop improvement and accelerate efforts in addressing global food problems due to increasing population and climate change. In addition, it will also facilitate the development of new wheat varieties with strong resistance against major diseases, enhanced bread-making qualities etc.

Research Background

Wheat is one of the world's three leading cereal crops that include rice and corn, providing about 20% of the calories we consume every day, and is also a leading source of protein. However, the global wheat supply and demand has become critical in recent years due to global environmental changes and population growth, and an increase in wheat production by more than 60% over the next 40 years has been deemed indispensable. In addition, in response to the emergence of new threats such as the outbreak of wheat blast disease, it is urgent to develop wheat varieties that are resistant to environmental changes for sustainable production. In Japan, it is also necessary to increase domestic wheat production and develop varieties with high yield, excellent milling quality, as well as varieties with sufficient resistance to pre-harvest sprouting, *Fusarium* head blight etc. often caused by Japan's wet climate.

Research Timeline

In order to facilitate the breeding of improved wheat varieties, it is necessary to identify agronomically important genes based on the genome information, and to develop DNA markers linked to such traits. However, the genetic make-up of the bread wheat that is currently grown throughout the world is

difficult to decipher because it is a complex hexaploid (AABBDD) species consisting of three genomes that originated from three different ancestral species. In addition, homoeologous genes with similar nucleotide sequences diverged from the same ancestor to each of the three genomes so that even if the sequence of a specific gene located on the A genome chromosome has been identified, it is not known whether the same gene exists in B genome chromosome or D genome chromosome and are therefore difficult to use as DNA markers. Moreover, the genome size of wheat which is almost 40 times the size of the rice genome, combined with the complexity of a polyploid genome, makes it difficult to use genome information.

To overcome these situations, the International Wheat Genome Sequencing Consortium (IWGSC) was organized in 2005 in order to accurately decipher the genome sequence of each wheat chromosome. Researchers from participating countries selected the bread wheat variety "Chinese Spring" (Fig. 1) and began to decipher the genome by sharing the 21 chromosomes (3 genomes × 7 chromosomes) which comprise the entire wheat genome. Japan led by the National Institute of Agrobiological Sciences (presently incorporated into the National Agriculture and Food Research Organization) joined the consortium and became in charge of sequencing chromosome 6B. In 2014, the IWGSC released a chromosome-based draft sequence corresponding to 61% of the entire genome with gene sequences assigned to individual chromosomes (July 17, 2014 press release of the National Institute of Agrobiological Sciences entitled "Genetic blueprint of the wheat genome revealed").

However, although the draft sequence provided the chromosomal position of the genes, the mutual positional relationships of the genes remain unclear. Therefore, the IWGSC continued the collaborative research towards determining the genome sequence of each individual chromosome to produce a reference genome sequence with high accuracy (Fig. 2).

Contents and Merits

The IWGSC has succeeded in decoding the 14.5 Gb (14.5 billion base pairs) reference sequence covering 94% of the bread wheat genome. Unlike the chromosome-based draft sequence, this reference genome sequence consists of finely divided 1.7 to 8.9 kb (1700 to 8900 base pairs) segments, corresponding to a single sequence assembly for each chromosome (average of 670 million base pairs). It is highly accurate and the positional relationships of genes in the chromosome have been clarified. This makes it possible to clarify the number of genes as well as the kind and position of the genes present in specific regions of the chromosomes. This would also greatly facilitate efficient gene isolation by narrowing down the position of genes with markers and eventually accelerate the development of new wheat varieties.

A comprehensive analysis of this reference genome sequence revealed a total of 107,891 genes. Among them are more than 2,000 genes which belong to the NB-LRR (nucleotide-binding, leucine rich repeat proteins) gene family and are thought to be involved in disease resistance, and more than 1,000 genes belonging to the prolamin gene family, which consists of seed storage proteins involved in the quality of wheat flour. It is expected that these genes could be used in developing new varieties for disease resistance and quality improvement.

In addition, from the information derived from 850 gene expression data derived from various plant organs and plant growth stages, the homoeologous genes as well as genes which are commonly expressed were found, leading to the elucidation of gene functions and gene networks in polyploid species.

Future Prospects

The reference genome sequence information of the wheat decoded this time will be indispensable in the breeding of wheat varieties, and the development of new varieties of which are resistant to environmental fluctuations for sustainable production, and wheat varieties with high crop yield in Japan and overseas.

Using the reference genome sequence information, NARO will promote the identification and elucidation of the genes that improve the quality genes such as resistance genes against *Fusarium*

head blight, yellow mosaic virus, preharvest sprouting etc. as well as genes involved in bread quality. The genome sequence information will also be used to promote the development of DNA markers that can be used for efficient breeding of high-quality wheat varieties capable of stable production and contribute further to the promotion of the domestic wheat industry.

The application of genomic sequence and DNA marker information will not be limited to the identification of agriculturally useful genes and wheat improvement. A comparative analysis of the wheat genome with the genomes which have been already completely decoded so far such as rice, corn, barley, sorghum etc. is expected to make a great contribution to fundamental research such as obtaining clues that can be used to understand evolution cereal crops which are important food resources for mankind.

Publication

International Wheat Genome Sequencing Consortium (2018) **Shifting the limits in wheat research and breeding through a fully annotated and anchored reference genome sequence.** *Science* doi: 10.1126/science.aar7191.

Reference Information



Fig. 1. Cultivated wheat (*Triticum aestivum* L.)

The leading Japanese wheat variety "Kitahonami" (background) and the wheat variety "Chinese Spring" (inset) which has been completely sequenced.

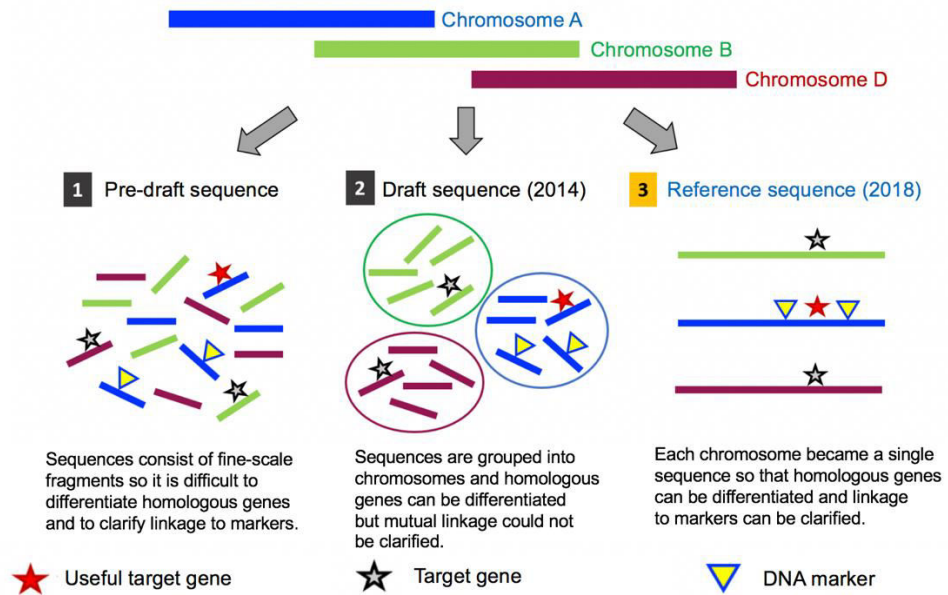


Fig. 2. Wheat genome information generated by distinct levels of sequencing

The DNA sequences of three genomes that have differentiated from a common ancestor were separately characterized and the sequences for each genome were further connected to facilitate the elucidation of a more accurate and easy-to-use genome information.

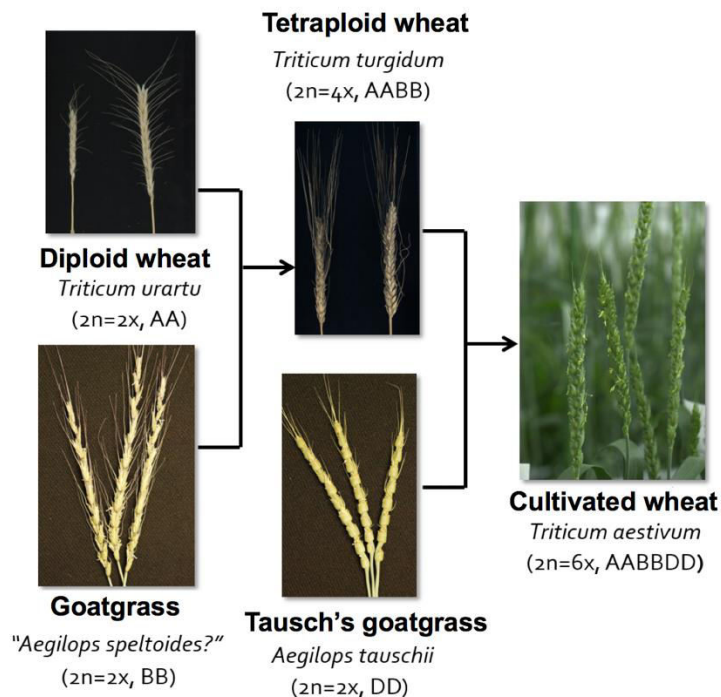


Fig. 3. Origin of cultivated wheat from related species

The cultivated wheat currently used for making bread and udon (wheat flour noodle) is a result of mutual crossing of three diploid species (*Triticum urartu*, *Aegilops speltoides*, *Aegilops tauschii*) that differentiated from a common ancestor.