The International Wheat Genome Sequencing Consortium

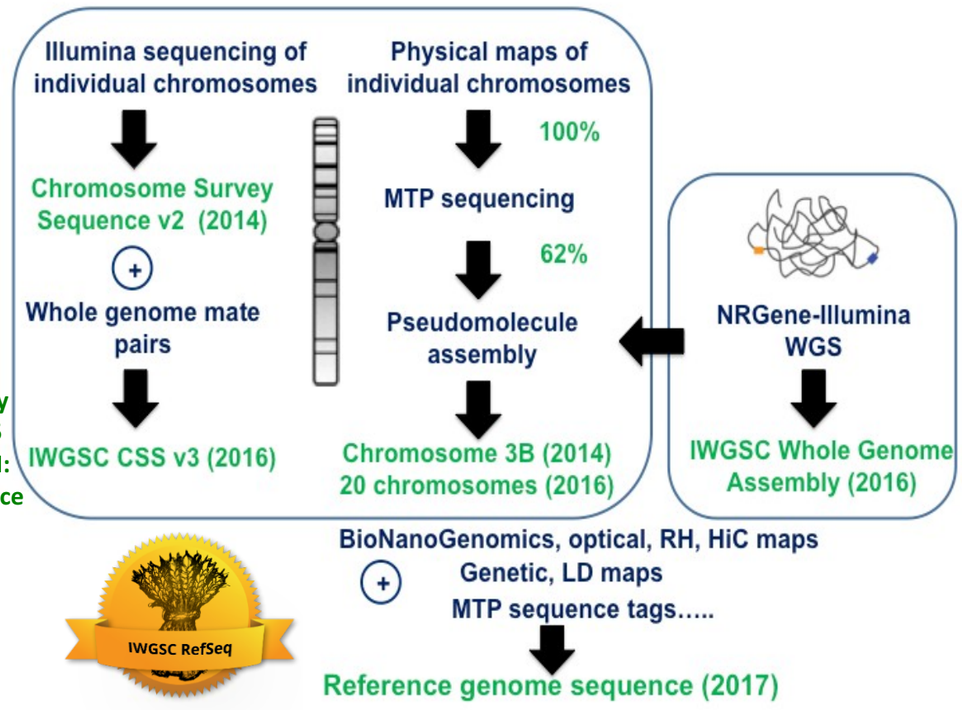
IWGSC[1], Kellye Eversole[2], Jane Rogers[3] and Isabelle Caugant[1]

(1) Lee's Summit, MO, USA; (2) IWGSC, Bethesda, MD, USA; (3) IWGSC, Cambridge, UK

www.wheatgenome.org

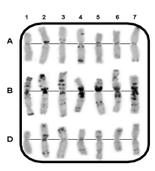
Roadmap to the Wheat Genome Sequence

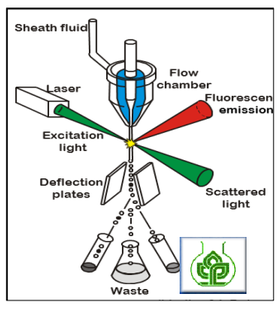
Science 18 July 2014: Vol. 345 no. 6194. DOI: 10.1126/science



To accelerate wheat improvement, the International Wheat Genome Sequencing Consortium (IWGSC) was created in 2005 to deliver to breeders and scientists a gold standard reference genome sequence anchored to genetic maps. The IWGSC strategy focused initially on producing BAC libraries and physical maps for each of the 21 individual bread wheat chromosomes to serve as the framework for the subsequent sequencing of minimal tiling paths of mapped BACs. To complement and complete the BAC-based sequences, the IWGSC produced additional chromosome-specific and whole genome resources. A genome-wide chromosome-based survey sequence provided early access to genic sequences and, more recently, a whole genome sequence assembled with the NRGene DeNovoMAGIC™ 2.0 software was integrated with the BAC-based resources (sequences, physical maps, whole genome profiling (WGP™) tags…) and long range marker maps (genetic maps, Hi-C, radiation hybrid maps) to produce the reference sequence of hexaploid bread wheat cv. *Chinese Spring,* IWGSC RefSeqv1.0. Pre-publication releases of sequence assemblies were made in June 2016 (IWGSC WGA v0.4 – whole genome scaffold assemblies ordered using POPSEQ and Hi-C) and January 2017 (IWGSC RefSeq v1.0). Genome annotation v1.0 was released in June 2017 and is available from http:// wheat-urgi.versailles.inra.fr along with all other IWGSC resources. For additional information on IWGSC RefSeq v1.0, see poster #0995.

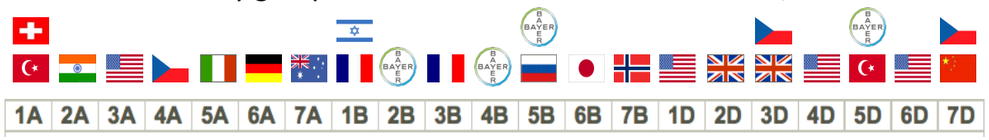
Chromosome-based Resources for Wheat





Chromosome / chromosome arm flow-sorting by J. Dolozel et al., IEB, Czech Republic.

- BAC libraries for 3 chromosomes + 34 chromosome arms: 2,713,728 BAC clones constructed at the Institute of Experimental Botany: http://olomouc.ueb.cas.cz/dna-libraries/cereals.
- All BAC libraries are archived and available at INRA CNRGV, Toulouse, France. For access, see http://cnrgv.toulouse.inra.fr.
- Physical maps constructed from chromosome-specific BACs for 21 wheat chromosomes can be viewed and are available for download from URGI. The maps were contributed by groups from 16 countries across the world, as indicated below.



- Minimal tile path BAC tag sequences (WGP™) generated through the IWGSC-Bayer CropScience project for all chromosomes, except 3B: 4,305,249 tags for 693,697 BACs.
- BAC sequence assemblies for 8 chromosomes (1A, 1B, 3B, 3D, 6B, 7A, 7B, 7D) + 2 arms (4AL, 5BS).
- Illumina chromosome arm shotgun libraries for survey sequence (CSS) (**Science 18 July 2014: Vol. 345 no. 6194. DOI: 10.1126/science)** provided chromosome-specific sequence contig assemblies (N_{50} contig length = 5.9 kb) used to assign ~100,00 genes, POPSEQ data and whole genome sequence assemblies to chromosomes. They were also used to design Illumina 90k and Affymetrix 420K SNP Chips.
- Bionano Genomics optical maps of 7A,7B,7D (1,335 Bionano contigs).

Physical maps and sequence data are available via the IWGSC sequence Repository at URGI INRA Versailles, France
http://wheat-urgi.versailles.inra.fr

Bread Wheat Reference Genome

The IWGSC Chinese Spring Bread Wheat Reference Sequence (IWGSC RefSeq v1.0) integrates a *de novo* assembly of Illumina short sequence reads built with NRGene DeNovoMAGIC™ 2.0 software (IWGSC WGA) with whole genome (Hi-C, POPSEQ and RH maps) and chromosome-based resources. For more details, see poster #0995.

Main features include:

- 14.5 Gb assembled sequence: 14.1 Gb assigned to chromosomes; 262 Mb gaps (Ns)
- N_{50} super-scaffold 22.8 Mb; N_{50} scaffold = 7.1 Mb (max 46 Mb)
- 107,891 high confidence gene models annotated
- 74% genes are homoeologs
- 27% HC genes duplicated (cf. 15-20% in other grasses)
- >3.9m TE elements (84% of genome)
- Transcriptome atlas comprising 850 RNASeq samples from 32 tissues (http://wheat-expression.com)

Next Step: IWGSC Phase II

With the completion of IWGSC RefSeq v1.0, the Consortium is moving into phase II activities:

- Pan-genome based on *de novo* sequencing and assembly of multiple wheat genomes (landraces and elite varieties) that cover the full breadth of wheat diversity
- Pan-genome database, visualization, and tools that will benefit public breeders and the full range of industry partners, from SMEs to large companies
- IWGSC Exome Array based on the reference sequence
- Pipeline for including community-based manual and functional annotation in annual annotation updates
- Continued development and maintenance of an integrated database for the wheat genome sequence.

Thanks to IWGSC sponsors

Arcadia Biosciences · biogemma · Bayer CropScience · CIMMYT · Céréales Vallée · DuPont The miracles of science · Florimond Desprez · INRA Science & Impact · Institute of Experimental Botany of the AS CR, v. v. i. · KeyGene · Limagrain from earth to life · Kansas Wheat · RAGT · syngenta · The University of Adelaide