The International Wheat Genome Sequencing Consortium

IWGSC[1], Rudi Appels[2], Ute Baumann[3], Hikmet Budak[4], Isabelle Caugant[1], Jan Dvorak[5], Kellye Eversole[1,6], Mingcheng Luo[5], Etienne Paux[7] and Sébastien Praud[8]

(1) Lee's Summit, MO, USA; (2) Agribio, Australia; (3) University of Adelaide, Australia; (4) Montana BioAg Inc, MT, USA ; (5) University of California, Davis, CA, USA; (6) Eversole Associates, MD, USA; (7) National Institute for Agricultural Research (INRA), Clermont-Ferrand, France; (8) Limagrain, Chappes, France

www.wheatgenome.org

Bread wheat, the staple food for 35% of the world's population, is the last major crop species to benefit from a reference genome sequence. In August 2018, the International Wheat Genome Sequencing Consortium (IWGSC), published the first high quality reference sequence of the bread wheat variety Chinese Spring (IWGSC RefSeq v1.0). With 2,400 members in 68 countries, the IWGSC is an international, collaborative consortium, established in 2005 by a group of wheat growers, plant scientists, and public and private breeders focused on delivering genomic tools and resources for wheat improvement.

The IWGSC RefSeq v1.0 represents 94% of the hexaploid wheat genome organized in 21 chromosome-like pseudomolecules and identifies the structural features of the wheat genome, including 107,891 high confidence gene in their genomic context, along with 4.7 million molecular markers. This sequence allows for analyses to a level that has never been achieved before.

The IWGSC has now entered a new phase of its activities. In particular, an improved version of the reference sequence, IWGSC RefSeq v2.0, which closed a number of gaps and corrected position and orientations of scaffolds, was released to the community in July 2019 under the Toronto pre-publication access agreement. The development of annotation v2.0 integrating functional and manual annotation, as well as alignment with other genomic resources is also in progress. Finally, a wheat diversity project has begun, aiming at completing high quality sequences of landraces representing the worldwide genetic diversity.

Wheat genome deciphered, assembled, and ordered. IWGSC RefSeq v1.0

The International Wheat Genome Sequencing Consortium (IWGSC) et al.
Science 2018;361:eaar7191
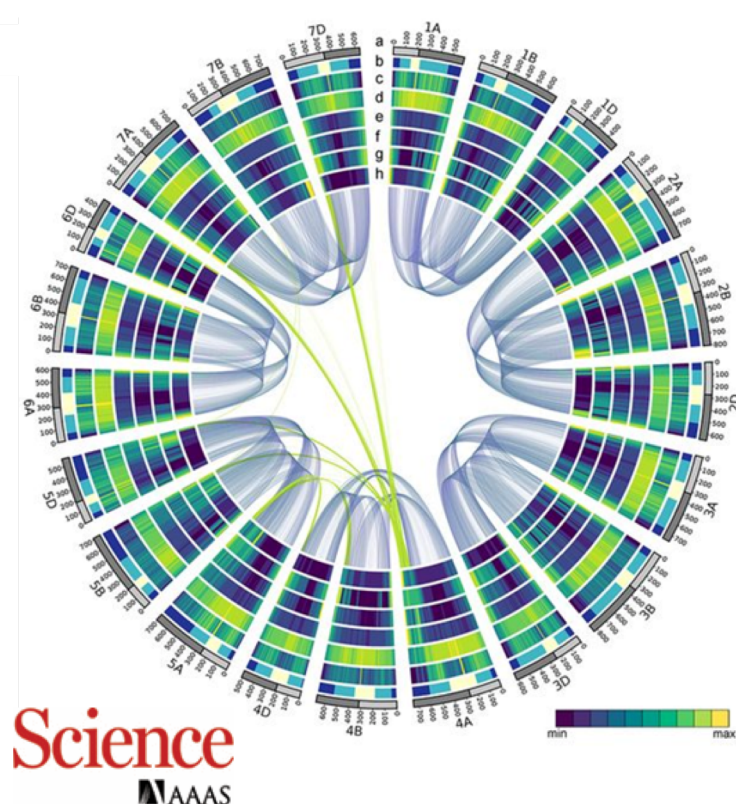
Legend: Circular diagram showing genomic features of wheat. The tracks toward the center of the circle display (a) chromosome name and size (100-Mb tick size; light gray bar indicates the short arm and dark gray indicates the long arm of the chromosome); (b) dimension of chromosomal segments R1, R2a, C, R2b, and R3; (c) K-mer 20-frequencies distribution; (d) LTR-retrotransposons density; (e) pseudogenes density (0 to 130 genes per Mb); (f) density of HC gene models (0 to 32 genes per Mb); (g) density of recombination rate; and (h) SNP density. Connecting lines in the center of the diagram highlight homeologous relationships of chromosomes (blue lines) and translocated regions (green lines).

Wheat Exome Panel

IWGSC Collaboration with Arbor Biosciences
- Based on Chinese Spring (Hexaploid)
- Compatible with other cultivars
- Focused on high confidence exons
- 200 Mb of 17 Gbp genome
- 30x coverage with 18 Gb sequencing

myBaits Wheat Exome v1
- Various kit sizes
  - 8, 48, 96 reaction kit sizes
  - Reaction sufficient for 8 pooled samples
- Services offered through myReads™
- New modules coming soon
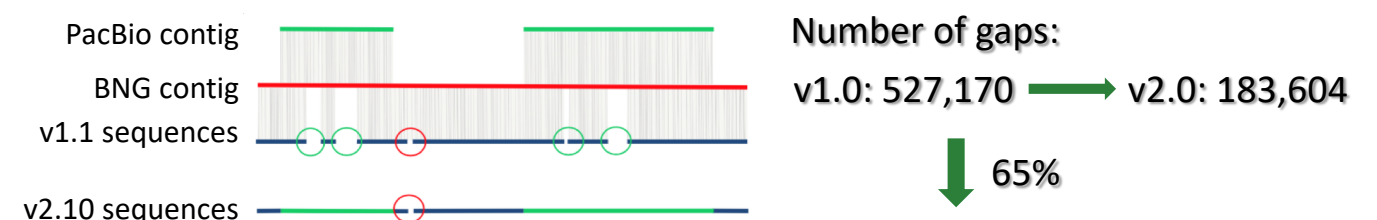 - Low-confidence annotations
  - Promoter regions

https://arborbiosci.com/products/mybaits-expert-wheat-exome-panel
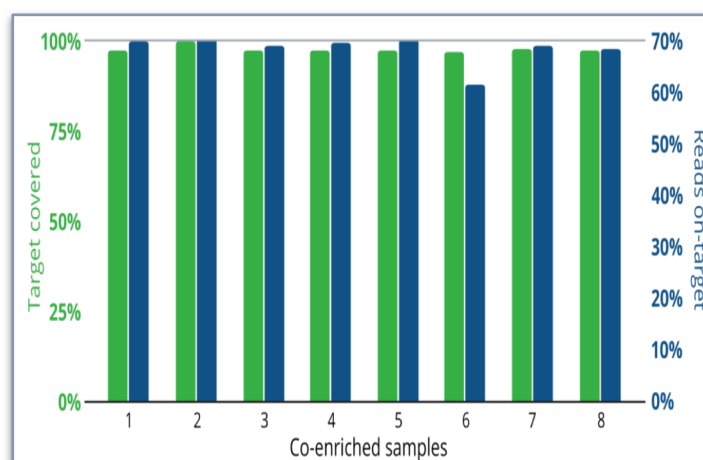
IWGSC RefSeq v2.0

Under the leadership of Mingcheng Luo and Jan Dvorak (UC Davis, CA, USA), with funding from the US National Science Foundation

Gap closing with PacBio contigs

Number of gaps:
v1.0: 527,170 → v2.0: 183,604
↓ 65%

Improvements v2.0 vs v1.0:
- Resolved 75 chimeric scaffolds
- Anchored 283 additional scaffolds from ChrUn of v1.0
- Corrected orientation errors for 357 scaffolds
- Revised locations (change of order) for 244 scaffolds
- Change of the total length of pseudomolecules: + 158,997,657 bp
- Change of the total effective length of pseudomolecules: + 167,843,761 bp
- Percentage of sequences anchored on chromosomes: 97.65% in v2.0 (96.69% in v1.0)
- All revisions: 1,380 Mb (≈ 10%) of the v1.0 assembly

Data available at the IWGSC data repository hosted by URGI-INRA (registration required) in advance of publication under the terms of the Toronto agreement. https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies

IWGSC Phase II

Phase II activities will expand the foundation for genomics-based crop improvement in wheat in response to challenges imposed by population expansion and climate change.

Phase II priorities:
- IWGSC Wheat Diversity project based on de novo sequencing and assembly of multiple wheat genomes (landraces and elite varieties) that cover the breadth of wheat diversity
- IWGSC Gold Standard: gap filling, maintenance, updates, and integration of manual and functional annotation of IWGSC RefSeq v1.0 & v2.0
- Develop a wheat proteome atlas
- Develop additional exome resources through the Arbor Bioscience collaboration
- Improved user-friendly, integrated databases and tools

All IWGSC data are available at URGI, INRA Versailles, for BLAST searches & download https://wheat-urgi.versailles.inra.fr/Seq-Repository