

The Reference Sequence for the Bread Wheat Genome

The International Wheat Genome Sequencing Consortium¹, and Rudi Appels²

(1) Lee's Summit, MO, USA; (2) Murdoch University, Perth, Australia

www.wheatgenome.org

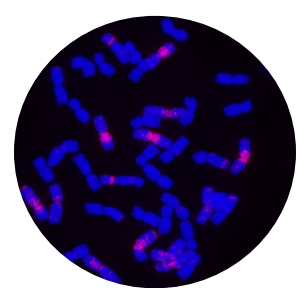
The goal of the IWGSC since its inception in 2005 has been the generation of a high quality reference genome sequence for bread wheat that integrates genetic and genomic resources and supports the acceleration of wheat improvement and production to keep pace with projected rises in human population.

The IWGSC RefSeq v1.0 has been assembled from a whole genome shotgun assembly of allohexaploid bread wheat *cv. Chinese Spring* and integrated with a wealth of IWGSC resources including: chromosome survey sequences (IWGSC CSS); chromosome-specific BAC-based physical maps, whole genome profiling (WGPTM) tag sequences, and optical maps; POPSEQ genetic maps, Hi-C, and radiation hybrid maps.

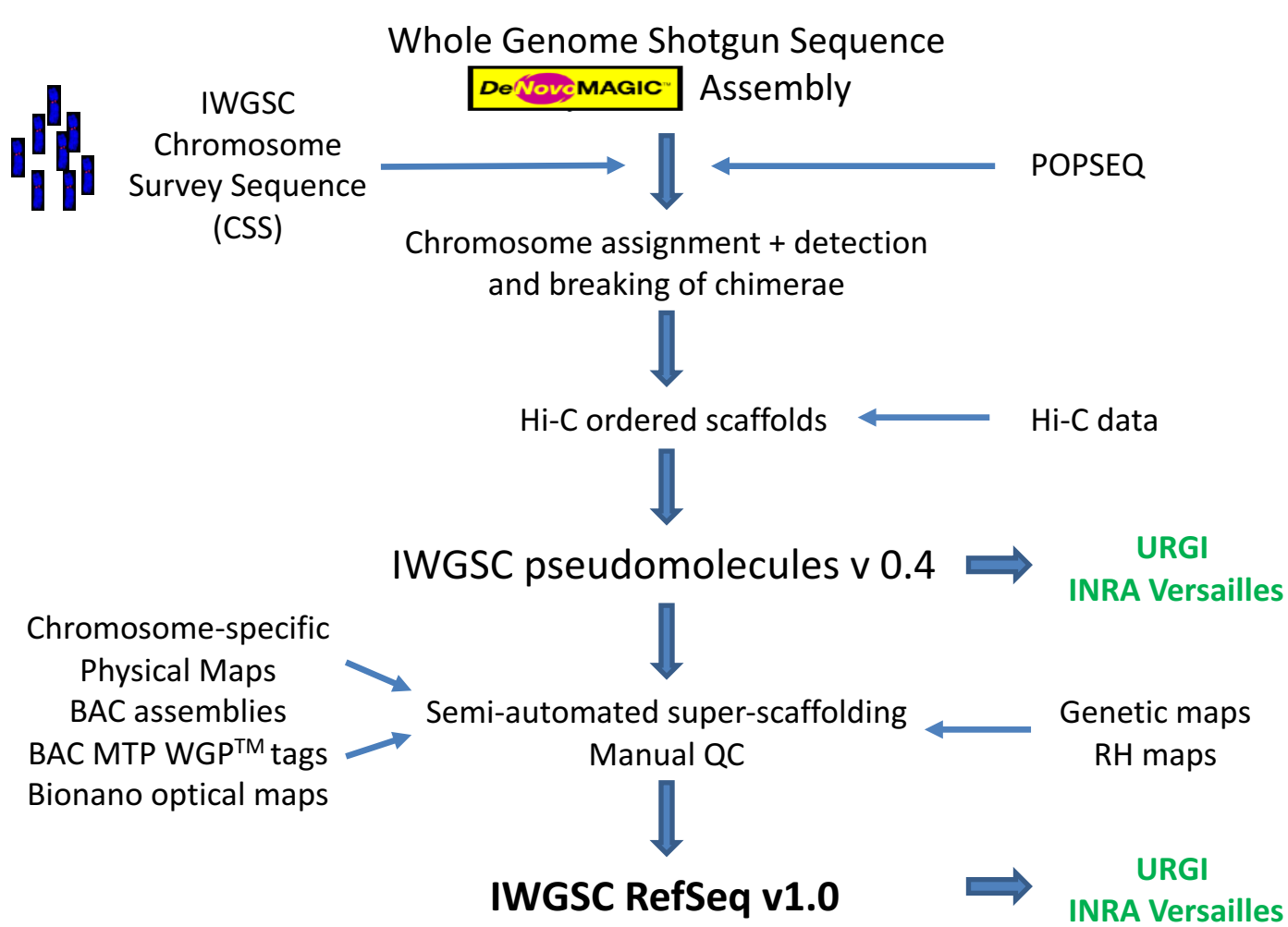
The assembly represents ~94% of the predicted wheat genome size in large super-scaffolds (N50 22.8Mb) that are assigned and ordered along the 21 wheat chromosomes. 107,891 high confidence gene models have been annotated and further sets of incompletely supported gene models and pseudogenes identified. The annotated genes have been used to analyze the distribution of homeologous genes across A,B,D genomes, together with gene duplications and losses that play important roles in wheat evolution. Insights into gene expression and its regulation have been revealed using a transcriptome atlas developed from 850 RNASeq datasets representing all stages of wheat phenological development.

With a sequence assembly that now supports the resolution of complex gene families associated with important traits such as yield, grain quality or disease resistance, wheat has a key resource in place to anchor all QTL knowledge to the reference sequence and stimulate new molecular approaches for the future.

IWGSC RefSeq v1 Assembly and Annotation

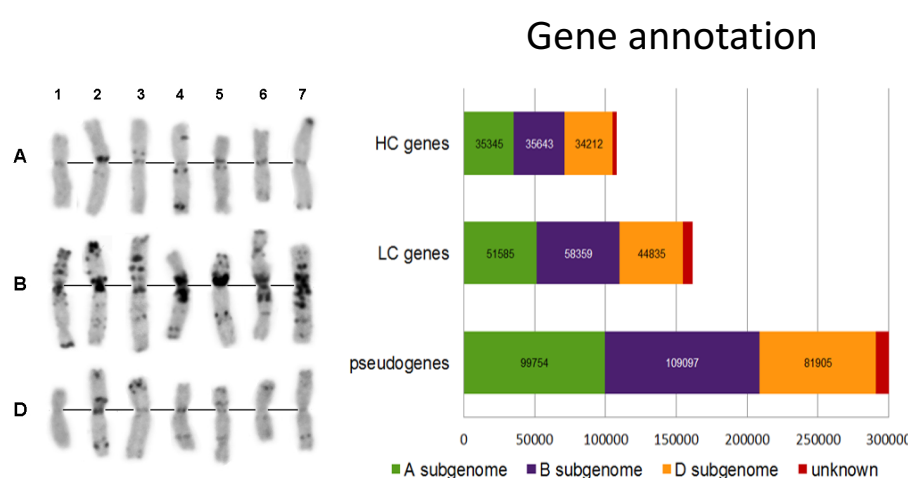


Triticum aestivum cv. Chinese Spring
Hexaploid 15.4 Gb genome



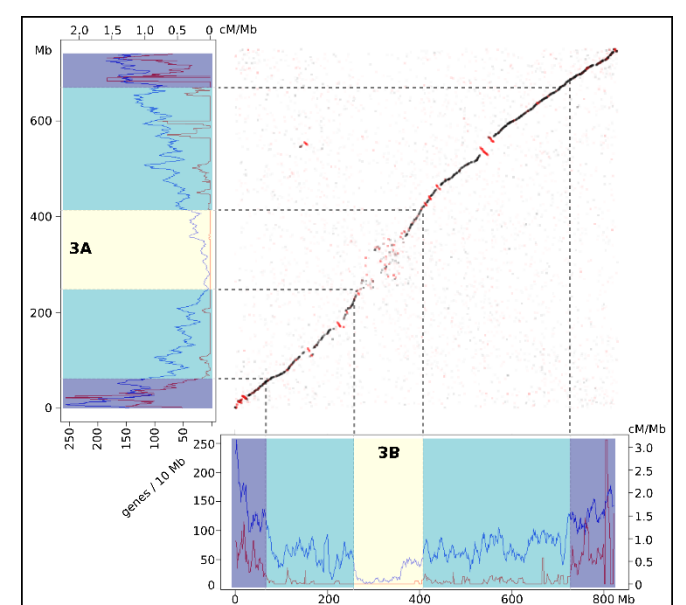
IWGSC RefSeq v1.0 contains: 14.1 Gb (>90%) assigned and ordered on chromosomes + 0.4Gb unassigned (Chr Un)

Super-scaffold N50 = 22.8Mb
Largest super-scaffold 166Mb
Average no. super-scaffolds per chromosome = 76



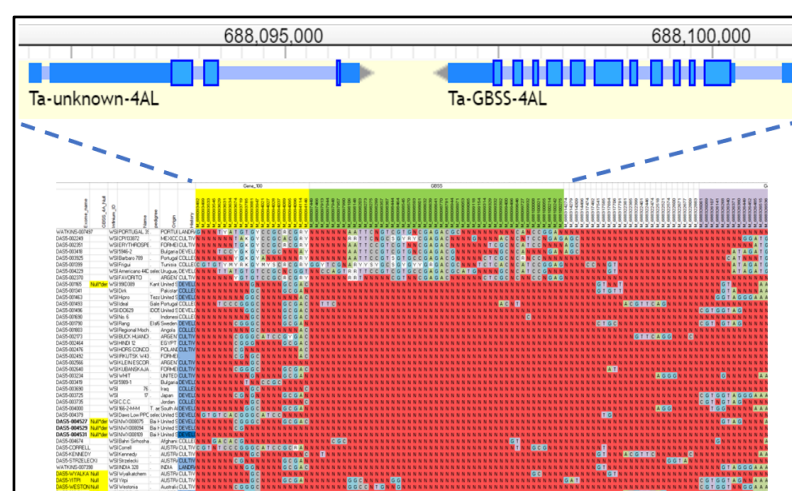
Wheat Genome Structure and Function

A,B,D genome comparisons reveal high levels of sequence collinearity between homeologous chromosomes. Gene order, gene density distribution (blue line), and recombination frequency (red line) (illustrated here for 3A and 3B (A)) are also highly conserved.



Five chromosomal regions have been defined for all chromosomes. R1 and R3 (purple) corresponding to chromosome extremities are recombination-rich, with high gene- and low TE-densities and contain the highest proportions of duplicated genes and genes expressed in fewest conditions. R2a and R2b (teal blue) regions are intermediate in character between R1/R3 and the gene-poor, recombination-free C regions, where collinearity between homeologs breaks down.

With high recombination, distal R1 and R3 regions provide the genomic environment for creating sequence diversity which gives the basis for the wheat plant's ability to adapt to changing environments.



Phylogenomic analysis identified 74% of genes (114,495) present in homeologous groups (i.e. clades of A,B,D sub-genome orthologs). In 54.6% of the groups genes are present in 1:1:1 ratio. The remaining homeologs are found in different copy number configurations. 30% of homeologs had lost a gene from one sub-genome, but no evidence was seen for any biased loss (i.e. sub-genome dominance).

Non-syntenic homeologs showed evidence of altered gene expression and, in some cases gene function. One example is illustrated (left) by Granule-Bound Starch Synthase (GBSS). This analysis shows GBSS to be a highly divergent homeolog of a group 7 gene, presumably resulting from the classic 7B→4A translocation. The allele on chromosome 4A is associated with udon noodle quality and was found to be deleted in 3.9% (25) of a set of 644 hexaploid wheats and landraces undertaken by AgriBio, Melbourne, Australia.

Data Resources

IWGSC data are available at URGI, INRA Versailles, for BLAST searches, download and browser format. <https://wheat-urgi.versailles.inra.fr>.

Thanks to IWGSC sponsors

