

PLEASE NOTE:

MD5 checksums are provided to ensure complete data transfer of files, please make sure that the MD5 checksum of your copy corresponds to the one provided with the files.

QUESTIONS

In case of questions about the IWGSC Data Repository, please contact Michael Alaux from URGI in the first instance at: urg-support@inrae.fr

In case of questions about the data content, you can contact the data providers.

ARTICLE

Zhu et al., Optical maps refine the bread wheat *Triticum aestivum* cv Chinese Spring genome assembly, *Plant J.* 2021 Apr 24, <https://doi.org/10.1111/tbj.15289>

Assembly

CREDIT: IWGSC RefSeq v2.1 assembly is the result of team efforts under the leadership of Mingcheng Luo and Jan Dvorak (UC Davis, CA, USA) and with funding from the US National Science Foundation grant IOS-1929053 and the USDA Agricultural Research Service CRIS project 2030-21430-014-00-D, an improved version of the reference wheat genome has been completed and is available for use without restriction.

The genome assembly of *Triticum aestivum* cv. Chinese Spring (IWGSC RefSeq v1.0; [IWGSC, 2018](#)) was improved using whole genome optical maps and contigs assembled from whole-genome-shotgun (WGS) PacBio SMRT reads ([Zimin et al. 2017](#)).

Optical maps were used to detect and resolve chimeric scaffolds, anchor unassigned scaffolds, correct ambiguities in positions and orientations of scaffolds, create super-scaffolds, and estimate gap sizes more accurately. PacBio contigs were used for gap closing.

Pseudomolecules of the Chinese Spring 21 chromosomes were re-constructed to develop a new reference sequence, IWGSC RefSeq v2.0. The RefSeq v2.0 was further refined by using the CDS of genes from the IWGSC annotation v1.1 (IWGSC. 2018). Among the CDS from 269,428 genes, mis-matches, “SNP” or one-base indels, were found in 9416 genes, likely base errors in the PacBio assembly remained after polishing. The sequence harboring the “problematic” CDS were therefore replaced with corresponding contigs in the RefSeq v1.0, and the final genome assembly RefSeq v2.1 was generated.

All revisions involved approximately 10% sequence length of the IWGSC RefSeq v1.0.