

DATA ACCESS DISCLAIMER:

These data are in open access. While scientists may freely publish using the IWGSC data, IWGSC does request that the source of the data be properly acknowledged.

For questions regarding data access and publication policy please contact IWGSC (communications@wheatgenome.org).

IWGSC RefSeq v1.1 annotation:

This annotation refers to the IWGSC RefSeq v1.0 assembly.

So it is a new version of the annotation of the same assembly.

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 please contact Michael Alaux from URGI in the first instance at:

urgi-support@inra.fr

CREDITS:

This data set has been generated as a collaborative effort between INRA-GDEC Clermont Ferrand, France (INRA) (Frederic Choulet, Philippe Leroy, Helene Rimbart), Plant Genome and Systems Biology, Helmholtz Zentrum, Munich, Germany (PGSB) (Klaus Mayer, Manuel Spannagl, Sven Twardziok, Heidrun Gundlach) and Earlham Institute, Norwich, UK (EI) (David Swarbreck, Luca Venturini, Gemy Kaithakottil) under the coordination of IWGSC (International Wheat Genome Sequencing Consortium) (Kellye Eversole, Jane Rogers).

Gene Annotation

NEWS:

In comparison of the v1.0 annotation, 3 modifications were done:

- add wrongly removed genes during the integration
- remove LC which have an overlap with manually curated genes (IWGSC_v1.1_LC_removed.ids)
- update ids of TE-LC genes coming from the HC set in order to fit with the LC naming and numbering (IWGSC_v1.1_LC.correspondanceTEHC.txt).

IDs of gene models follow this nomenclature:

v1.0: TraesCS1A01GXXXXXX.1

v1.1: TraesCS1A02GXXXXXX.1

Traes: Triticum aestivum (species name)

CS: Chinese Spring (variety name)

1A: chromosome name

O[12]: version of annotation 01=v1.0 / 02=v1.1
G: genomic sequence
XXXXXX: 6 digit number
.1: splicing variant #1

While updating the RefSeq Annotation from v1.0 to v1.1, the "6 digit number" was NOT modified. For instance, TraesCS1A01G000100.1 and TraesCS1A01G000100.1 define the two IDs of the same gene model in versions v1.0 and v1.1, respectively.

Manually curated gene families

CREDITS:

Manually curation of gene families have been performed by different groups (cf. <http://science.sciencemag.org/content/361/6403/ear7191>).

CONTENT:

Theses gene are already available in the gene v1.1 annotation.
These files have been generated to provide an easy access to gff3 dedicated to each family.

cbf.gff3 CRT-Binding Factors (CBF) gene family
nlr.gff3 Nucleotide-binding and Leucine-rich Repeat (NLR) gene family
ppr.gff3 Pentatricopeptide repeats (PPR) gene family
prol_glob.gff3 Prolamin and Globulin genes
rk_aatset9.gff3 Amino acid transporters
wak.gff3 Cell Wall-Associated Kinase (WAK)

RNA-seq mapping

CREDITS:

Ricardo Ramirez-Gonzalez, Philippa Borrill, Cristobal Uauy

This folder contains the TPMs and Counts for the public samples

Mapping protocol:

To calculate the read counts and TPMs we used [Kallisto version 0.42.3](<https://pachterlab.github.io/kallisto/about>). The index was built with the default values and K=31. Kallisto was run with the default parameters. All the fastq files from a single sample are run together.

The reference used was RefSeqv1.1, including the High and Low quality genes.