**METHODOLOGY AND BIOINFORMATIC PIPELINE ASSOCIATED TO ILLUMINA SNP DATA ANALYSIS USED BY THE INTERNATIONAL DURUM WHEAT GENOME SEQUENCING CONSORTIUM FOR BI-PARENTAL MAPPING AND GDP AND TGC GERMPLASM COLLECTION ANALYSIS IN TETRAPLOID WHEAT**

Adapted from Maccaferri et al. (2019) *Nature Genetics* 51, pages885–895(2019), Supplementary Materials

**BI-PARENTAL MAPPING**

" The genetic maps were produced following a pipeline including: i) scripts **(https://github.com/plantinformatics/Durum_iSelect_90kSNP_GenotypeCalling)** for genotype calling in unrelated samples, sample cluster assignment, confidence score estimates, and final genotype call from Illumina raw data project files; ii) quality check and filtering of genotype calls; iii) marker grouping and ordering in MST-map47. The Script parameters used for genotype calling were as follows:
**-d 3,** sample must be within 3 standard deviations of a known cluster position.
**-r 0.8,** minimum confidence score that sample belongs to the cluster to which it was assigned versus the next closest cluster; a value of 1 indicates highest confidence.
**-g 0.4,** minimum sample genotype call rate before reporting a SNP.
The genotype call outputs for the mapping population data-sets were filtered for SNP call rate (minimum 90%), cross-over rate, presence of identical samples. Marker grouping and linkage mapping were performed in MST-map47 (http://www.mstmap.org/), which can efficiently determine the correct order of markers by computing the Minimum Spanning Tree of an associated graph. For each map, the linkage groups were aligned and oriented based on the tetraploid SNP consensus map48.
The 17 genetic maps, including a revisited Svevo × Zavitan SNP map with 16,372 mapped SNPs, provided genetic and physical map positions for 38,340 Illumina iSelect SNP, 1,341 DArT, 835 SSR, and 109 STS markers as reported in Supplementary materials and in the websites hosting the Svevo genome sequence browser (Interomics website, https://www.interomics.eu/durum-wheat-genome and GrainGenes, https://wheat.pw.usda.gov/GG3/jbrowse_Durum_Svevo)."

**GERMPLASM**

"The raw data (Theta/R) from single Illumina genotyping experiments were jointly analysed for cluster assignment and genotype calling using a custom script for genotype calling in unrelated samples (https://github.com/plantinformatics/Durum_iSelect_90kSNP_GenotypeCalling), as described for the mapping population analysis. In brief, the script assigns samples to clusters previously identified from the genetic mapping analysis. A sample was assigned to a cluster if its probability to belong to that cluster (vs. the next closest cluster) exceeded 0.8. Samples assigned to each cluster were assigned a genotype call if the segregating allele tagged by the cluster could be unambiguously tracked; i.e. the allele it tracks was previously genetically mapped. Based on the complexity of the signal, cluster could be two (best situation corresponding to one single Mendelian locus) or multiples. Thus, the assigned genotype was an arbitrary allelic state, i.e. AA, BB or NC (not called). The cluster file underpinning the script used for genotype calling was based on 38,340 genetically mapped SNP loci mapped across 17 mapping populations.
The script parameters used for genotype calling were: -d 3, called sample were within 3 standard deviations of a known cluster position; -r 0.8, minimum confidence score that sample belongs to the cluster to which it was assigned versus the next closest cluster; a value of 1 indicates highest confidence. The genotype call pipeline allowed us to retrieve 34,543 SNP polymorphic on the complete dataset of 2,558 accessions. This complete dataset was subjected to two consecutive rounds of filtering for redundancy, initially based on passport information (accession name/international code) and then genetic similarity matrix (simple matching genetic similarity) among accessions based on SNP data."