Shifting the limits in wheat research and breeding using a fully annotated reference genome

International Wheat Genome Sequencing Consortium (IWGSC); Appels, Rudi; Eversole, Kellye; Feuillet, Catherine; Keller, Beat; Rogers, Jane; Stein, Nils; Pozniak, Curtis J; Stein, Nils; Choulet, Frédéric; Distelfeld, Assaf; Eversole, Kellye; Poland, Jesse; Rogers, Jane; Ronen, Gil; Sharpe, Andrew G; Pozniak, Curtis; Ronen, Gil; Stein, Nils; Barad, Omer

DOI:
10.1126/science.aar7191

Title: Shifting the limits in wheat research and breeding using a fully annotated reference genome

Authors: International Wheat Genome Sequencing Consortium (IWGSC)* †.

Affiliations:

*Correspondence to: rudi.appels@unimelb.edu.au (Rudi Appels), eversole@eversoleassociates.com (Kellye Eversole), and stein@ipk-gatersleben.de (Nils Stein).

† All authors with their affiliations appear in the acknowledgements at the end of this paper.

Abstract (100 – 125 words): An annotated reference sequence representing the hexaploid bread wheat genome in 21 pseudomolecules has been analyzed to identify the distribution and genomic context of coding and non-coding elements across the A, B and D sub-genomes. With an estimated coverage of 94% of the genome and containing 107,891 high confidence gene models, this assembly enabled the discovery of tissue and developmental stage related co-expression networks using a transcriptome atlas representing major stages of wheat development. Dynamics of complex gene families involved in environmental adaptation and end-use quality were revealed at sub-genome resolution and contextualized to known agronomic single gene or quantitative trait loci. This community resource establishes the foundation for accelerating wheat research and application through improved understanding of wheat biology and genomics-assisted breeding.

One Sentence Summary (keep under 125 characters): The 21 annotated chromosomes of bread wheat provide a foundation for accelerated innovation in wheat research and breeding.

Main Text: Wheat (*Triticum aestivum* L.), the most widely-cultivated crop on earth, contributes about a fifth of total calories consumed by humans and provides more protein than any other food source (*1*). Breeders strive to develop improved varieties by fine-tuning genetically complex yield and end-use quality parameters while maintaining yield stability and regional adaptation to specific biotic and abiotic stresses (*2*). These efforts are limited, however, by insufficient knowledge and understanding of the molecular basis of key agronomic traits. To meet the demands of human population growth, there is an urgent need for wheat research and breeding to accelerate genetic gain while increasing wheat yield and protecting quality traits. In other plant and animal species, access to a fully annotated and ordered genome sequence, including regulatory sequences and genome diversity information, has promoted the development of systematic and more time-efficient approaches for the selection and understanding of important traits (*3*). Wheat has lagged behind other species primarily due to the challenges of assembling a large (1C=16 Gb) (*4*), hexaploid and complex genome that contains over 85% repetitive DNA.

To provide a foundation for improvement through molecular breeding, the International Wheat Genome Sequencing Consortium (IWGSC) established a road map to deliver a high-quality reference genome sequence of the bread wheat cultivar ‘Chinese Spring’ (CS). A chromosome survey sequence (CSS) intermediate product assigned 124,201 gene loci across the 21 chromosomes and revealed the evolutionary dynamics of the wheat genome through gene loss, gain, and duplication (*5*). The lack of global sequence contiguity and incomplete coverage (only 10 Gb were assembled), however, did not provide the wider regulatory genomic context of genes. Subsequent whole genome assemblies improved contiguity (*6*-*8*) but lacked full annotation, and did not resolve the intergenic space or present the genome in the correct physical order.

Here, we report an ordered and annotated assembly (IWGSC RefSeq v1.0) of the 21 chromosomes of the allohexaploid wheat cultivar CS, an achievement that is built on a rich history of chromosome studies in wheat (*9-11*) that allowed the integration of genetic and genomic resources.. The completeness and accuracy of IWGSC RefSeq v1.0 provides insights into global genome composition and enables the construction of complex gene co-expression networks to identify central regulators in critical pathways, such as flowering time control. The ability to resolve the inherent complexity of gene families related to important agronomic traits demonstrates the impact of IWGSC RefSeq v1.0 on dissecting quantitative traits genetically and implementing modern breeding strategies for future wheat improvement.

Chromosome-scale assembly of the wheat genome

Pseudomolecule sequences representing the 21 chromosomes of the bread wheat genome were assembled by integrating a draft whole genome *de novo* assembly (WGA), built from Illumina short read sequences using NRGene deNovoMagic2 (Fig. 1A, Tables 1, S1, S2) with additional layers of genetic, physical, and sequence data (Tables S3-S8, Figs. S1, S2). In the resulting 14.5 Gb genome assembly, contigs and scaffolds with N50s of 52 kb and 7 Mb, respectively, were linked into superscaffolds (N50 = 22.8 Mb), with 97% (14.1 Gb) assigned and ordered along the 21 chromosomes and almost all of the assigned sequences also oriented (13.8 Gb, 98%). Unanchored scaffolds comprising 481 Mb (2.8% of the assembly length) formed the ‘unassigned chromosome’ (ChrUn) bin. The quality and contiguity of the IWGSC RefSeq v1.0 genome assembly was assessed through alignments with radiation hybrid maps for the A, B, and D sub-genomes (average Spearman’s ρ: 0.98), the genetic positions of 7,832 and 4,745 genotyping-by-sequencing (GBS) derived genetic markers in 88 double haploid and 993 recombinant inbred

lines (Spearman's r: 0.986 and 0.987, respectively), and 1.24 million pairs of neighbor insertion site based polymorphism markers (ISBPs) (*12*) of which 97% were collinear and mapped in a similar size range (difference <2 kb) between the de novo WGA and the available BAC-based sequence assemblies. Finally, IWGSC RefSeq v1.0 was assessed with independent data derived from coding and non-coding sequences revealing that 99% and 98% of the previously known coding exons (*5*) and TE-derived (ISBP) markers (Table S9), respectively, were present in the assembly. The approximate 1 Gb size difference between IWGSC RefSeq v1.0 and the new genome size estimates of 15.4-15.8 Gb (*13*) can be accounted for by collapsed or unassembled sequences of highly-repeated clusters, such as ribosomal RNA coding regions and telomeric sequences.

A key feature distinguishing the IWGSC RefSeq v1.0 from previous draft wheat assemblies (*5-8*) is the long-range organization with 90% of the genome represented in super-scaffolds larger than 4.1 Mb and with each chromosome represented on average by only 76 super-scaffolds (Table 1). The largest super-scaffold spanned 166 Mb, i.e. half the rice (*Oryza sativa* L.) genome, and is larger than the *Arabidopsis thaliana* L. genome (*14, 15*). Moreover, the 21 pseudomolecules position molecular markers for wheat research and breeding (504 SSRs, 3,025 DArTs, 6,689 ESTs, 205,807 SNPs, 4,512,979 ISBPs) (Table S9), thus providing a direct link between the genome sequence and genetic loci / genes underlying traits of agronomic importance.

The composition of the wheat genome

Analyses of the components of the genome sequence revealed the distribution of key elements and enabled detailed comparisons of the homeologous A, B and D sub-genomes. Accounting for 85% of the genome with a relatively equal distribution across the three sub-genomes (Table 2),

3,968,974 copies of transposable elements (TEs) belonging to 505 families were annotated. Many (112,744) full length long terminal-repeat retrotransposons (fl-LTRs) were identified that have been difficult to define from short read sequence assemblies (Fig. S3). Although the TE content has been extensively rearranged through rounds of deletions / amplifications since the divergence of the A, B and D sub-genomes about 5 million years ago, the TE families that shaped the Triticeae genomes have been maintained in similar proportions: 76% of the 165 TE families present in a cumulative length greater than 1Mb contributed similar proportions (<2-fold change between sub-genomes) and only 11 families, accounting for 2% of total TEs, showed a higher than 3-fold change between 2 sub-genomes (*16*). TE abundance accounts, in part, for the size differences between sub-genomes, e.g. 64% of the 1.2 Gb size difference between the B and D sub-genomes can be attributed to lower gypsy retrotransposon content. Low-copy DNA content (primarily unclassified sequences), also varied between sub-genomes, accounting, for example for 97 Mb of the 245 Mb size difference between A and B genomes. (Fig. S4). As reported (*17*), no evidence was found for a major burst of transposition after polyploidization. The independent evolution in the diploid lineages was reflected in differences in the specific composition of A, B and D at the sub-family (variants) level as evidenced by sub-genome specific over-representation of individual transposon domain signatures (Fig. 1B). See (*16*) for a more detailed analysis of the TE content and its impact on the evolution of the wheat genome .

In addition to TEs, annotation of the intergenic space included non-coding RNAs. We identified eight new miRNA families (Fig. S5, Table S10) and the entire complement of tRNAs (showed an excess of lysine tRNAs, Fig. S6). Around 8,000 NUPTs (nuclear inserted plastid DNA segment) and 11,000 NUMTs (nuclear inserted mitochondrial DNA segments) representing respectively 5

and 17 Mb were also revealed by comparing the genome assembly with complete plastid and mitochondrial genomes assembled from the IWGSC RefSeqv1.0 raw read data (*13*).

Precise positions for the centromeres were defined by integrating Hi-C, CSS data (*5*) and published chromatin immuno-precipitation sequencing (ChIP-seq) data for CENH3, a centromere-specific histone H3 variant (*18*). Clear ChIP-Seq peaks were evident in all chromosomes and coincided with the centromere-specific repeat families (Figs. 1C, S7, Table S11). CENH3 targets were also found in unassigned sequence scaffolds (ChrUn) indicating that centromeres of several chromosomes are not yet completely resolved. On the basis of these data, a conservative estimate for the minimal average size of a wheat centromere is 4.9 Mb (6.7 Mb, if including ChrUn, Table S11) contrasting with ~1.8 Mb in maize (*19, 20*) and 0.4-0.8 Mb in rice (*21*).

Gene models were predicted with two independent pipelines previously utilized for wheat genome annotation and then consolidated to produce the RefSeq Annotation v1.0 (Fig. S8). Subsequently, a set of manually-curated gene models was integrated to build RefSeq Annotation v1.1 (Fig. S9, Tables S12-S17). In total, 107,891 high confidence (HC) protein coding loci were identified, with relatively equal distribution across the A, B and D sub-genomes (35,345, 35,643, and 34,212, respectively; Figs. 1D, 2A, S10, Table S18). In addition, 161,537 other protein coding loci were classified as low confidence (LC) genes representing partially supported gene models, gene fragments, and orphans (Table S18). A predicted function was assigned to 82.1% (90,919) of HC genes in RefSeq Annotation v1.0 (Tables S19, S20) and evidence for transcription was found for 85% (94,114), compared to 49% of the LC genes (*22*). Within the

pseudogene category, 25,419 (8%) of 303,818 candidates matched LC gene models. The D sub-genome contained significantly fewer pseudogenes than the A and B sub-genomes (81,905 versus 99,754 and 109,097, respectively; χ^2 P <2.2e-16) (Tables S21, S22, Fig. S10). In ChrUn, 2,691 HC and 675 LC gene models were identified.

The quality of the RefSeq Annotation v1.1 gene set was benchmarked against BUSCO v3 (*23*) representing 1,440 Embryophyta near-universal single-copy orthologs and published annotated wheat gene sets (Figs. 2B, S11). 99% (1,436) of the BUSCO v3 genes were represented in at least one complete copy in RefSeq Annotation v1.1 and 90% (1,292) in three complete copies, an improvement over the 25% (353) and 70% (1,014) identified in the IWGSC (*5*) and TGACv1 (*7*) gene sets, respectively (Fig. 2B). Improved contiguity of sequences in the immediate vicinity of genes was also found: 61% of the HC and LC genes were flanked by at least 10 kb of sequence without Ns, in contrast to 37% and only 5% of TGACv1 and IWGSC CSS gene models, respectively (Fig. S12).

To further characterize the gene-space, a phylogenomic approach was applied to identify gene homeologs and paralogs between and within the wheat sub-genomes and orthologs in other plant genomes (Table S23, Figs. S13-S15). Analysis of a subset of 181,036 genes ("filtered gene set", (*13*), Table 3) comprising 103,757 HC and 77,279 LC genes, identified 39,238 homeologous groups, i.e. clades of A, B and D sub-genome orthologs deduced from gene trees, containing a total of 113,653 genes (63% of the filtered set). Gene losses / retention and gene gains (gene duplications) were determined for all homeologous loci of IWGSC RefSeq v1.0 (Table 3) assuming the presence of a single gene copy at every homeologous locus (referred to as a

"triad"). The percentage of genes in homeologous groups for all configurations (ratios) is highly similar, hence balanced, across the three sub-genomes: 63% (A), 61% (B), and 66% (D). The slightly higher percentage of homeologs on the D sub-genome, together with the lower number of pseudogenes (Table S22) is consistent with its more recent hybridization with the A / B genome progenitor. Although the majority of genes are present in homeologous groups, only 18,595 (47%) of the groups contained triads with one single gene copy per sub-genome (1:1:1 configuration). 5,673 (15%) groups of homeologous genes exhibited at least one sub-genome inparalog, i.e. a gene copy resulting from a tandem or segmental / trans-duplication (1:1:N configuration). The three genomes exhibited similar levels of loss of individual homeologs, affecting 10.7% (0:1:1), 10.3% (1:0:1), and 9.5% (1:1:0) of the homeologous groups in the A, B and D sub-genomes, respectively (Tables 3, S24, S25).

Among the 67,383 (37%) genes of the filtered set not present in homeologous groups, 31,140 genes also had no orthologs in species included in the comparisons outside of bread wheat and comprised, mainly, gene fragments, non-protein-coding loci with open reading frames or other gene calling artifacts. The remaining 36,243 genes had homologs outside of bread wheat and appeared to be sub-genome specific (Table 3). Two of the genes in this category were *granule bound starch synthase*, *GBSS*, on chromosome 4A (1:0:0, a gene that is a key determinant of udon noodle quality) and *ZIP4* within the *Ph1* (*Pairing homeologous 1*) locus on chromosome 5B [0:1:0, a locus critical for the diploid meiotic behavior of the wheat homeologous chromosomes (*24*)]. The phylogenomic analysis indicated the *GBSS* on 4A is a divergent translocated homeolog originally located on chromosome 7B (Fig. S16); whereas, *ZIP4* is a trans-duplication of a chromosome 3B locus (Table S26). Both genes confer important properties on wheat and

illustrate the diversity in origin and function of gene models that are not in a 1:1:1 configuration. No evidence was found for biased partitioning. Rather, our analyses support gradual gene loss and gene movement among the sub-genomes that may have occurred either in the diploid progenitor species, the tetraploid ancestor or following the final hexaploidization event in modern bread wheat (Tables 3, S24, S25). Together with the equal contribution of the three homeologous genomes to the overall gene expression (*22*), this demonstrates the absence of sub-genome dominance ((*25*)).

29,737 bread wheat HC genes (27%) are present as tandem duplicates, which is up to 10% higher than found for other monocotyledonous species (Table S27). Tandemly repeated genes are most prevalent in the B genome (29%), contributing to its higher gene content and larger number of 1:N:1 homeologous groups (Table 3). The postulated hybrid origin of the D sub-genome as a result of inter-specific crossing with AB genome progenitors 1-2 My after they diverged (*26*), is consistent with the synonymous substitution rates of homeologous gene pairs (Fig. S17). Homeologous groups with gene duplicates in at least one sub-genome (1:1:N, 1:N:1, N:1:1) showed elevated evolutionary rates (for the sub-genome carrying the duplicate) compared to strict 1:1:1 or 1:1 groups (Figs. S18-S22). Homeologs with recent duplicates also showed higher levels of expression divergence (Fig. S23), consistent with gene / genome duplications acting as a driver of functional innovation (*27, 28*).

Analysis of synteny between the seven triplets of homeologous chromosomes showed high levels of conservation. There was no evidence for any major rearrangements since the A, B and D sub-genomes diverged ~5 Mya (Fig. 1D), although collinearity between homeologs was disturbed by inversions occurring on average every 74.8 Mb involving blocks of ten genes or more (mean

gene number 48.2 with a mean size of 10.5 Mb) (Fig. 1D, Table S28). Macro-synteny was conserved across centromeric (C) regions, but collinearity (micro-synteny) broke down specifically in these recombination-free, gene-poor regions, for all seven sets of homeologous chromosomes (Figs. 1D, S24-S26, Table S29). Among the 113,653 homeologous genes, 80% (90,232) were found organized in macro-synteny, i.e. still present at their ancestral position (Table S24). At the micro-synteny scale, 72% (82,308) of the homeologs were organized in collinear blocks i.e. intervals with a highly-conserved gene order (Fig. 1D). A higher proportion of syntenic genes was found in the interstitial regions [short arm, R2a (*17*), 46% and long arm, R2b (*17*), 61%] compared to the distal telomeric [short arm, R1 (*17*), 39% and long arm, R3 (*17*), 51%] and centromere regions [C (*17*), 29%], respectively, and the interstitial compartments harbored larger syntenic blocks (Figs. S27, S28). The higher proportions of duplicated genes in distal-terminal regions (34% and 27% versus 13-15% in the other regions; Fig. S29) exerted a strong influence on the decay of syntenic block size and contributed to the higher sequence variability in these regions. Overall, distal chromosomal regions are the preferential targets of meiotic recombination and the fastest evolving compartments. As such, they represent the genomic environment for creating sequence, hence, allelic diversity, providing the basis for adaptability to changing environments.

Atlas of transcription reveals trait associated gene co-regulation networks

The gene annotation coupled with identification of homeologs and paralogs in IWGSC RefSeq v1.0 provide a resource to study gene expression in genome-wide and sub-genome contexts. A total of 850 RNA-Seq samples derived from 32 tissues at different growth stages and/or challenged by different stress treatments were mapped to RefSeq Annotation v1.0 (Database S1,

Fig. 3A, Tables S30, S31, S32). Expression was observed for 94,114 (84.9%) HC genes (Fig. S30) and for 77,920 (49.1%) LC genes, the latter showing lower expression breadth and level [median 6 tissues; average 2.9 transcripts per million (tpm)] than the HC genes (median 20 tissues; average 8.2 tpm) (Fig. S31). This correlated with the higher average methylation status of LC genes (Figs. S32, S33). A principal component analysis (PCA) identified tissue (Fig. 3B), rather than growth stage or stress (Fig. S34), as the main factor driving differential expression between samples, consistent with studies in other organisms (*29-32*). 31.0 % of genes are expressed in over 90% of tissues (average 16.9 tpm, ≥ 30 tissues), and 21.5% of genes are expressed in 10% or fewer tissues (average 0.22 tpm; ≤ 3 tissues; Fig. S31).

8,231 HC genes showed tissue-exclusive expression (Fig. S35). Around half of these were associated with reproductive tissues (microspores, anther and stigma/ovary), consistent with observations in rice (*33*). The tissue-exclusive genes were enriched for response to extra-cellular stimuli and reproductive processes (Database S2). In contrast, 23,146 HC genes expressed across all 32 tissues were enriched for biological processes associated with house-keeping functions such as protein translation and protein metabolic processes. Tissue specific genes were shorter (1,147 ± 8 bp), had fewer exons (2.76 ± 0.3), and were expressed at lower levels (3.4 ± 0.1 tpm) compared to ubiquitous genes (1,429 ± 7 bp; 7.87 ± 0.4 exons, 17.9 ± 0.4 tpm) (Fig. S35).

Genes located in distal regions R1 and R3 (Fig. S25, Table S29) showed lower expression breadth than those in the proximal regions (15.7 and 20.7 tissues, respectively) (Figs. 3C, S36). This correlated with enrichment of Gene Ontology (GO) slim terms such as ‘cell cycle’, ‘translation’, and ‘photosynthesis’ for genes in the proximal regions, whereas, genes enriched for ‘response to stress’ and ‘external stimuli’ were found in the highly recombinant distal R1 and R3

regions (Database S3, Fig. S36, Table S33). The expression breadth pattern was also correlated with the distribution of the repressive H3K27me3 (Pearson correlation coefficient R= -0.76, P < 2.2E-16) and with the active H3K36me3 and H3K9ac (Pearson correlation coefficient R= 0.9 and 0.83, respectively, P < 2.2E-16) histone marks (Fig. S37).

Global patterns of co-expression (*34*) were determined with a weighted gene co-expression network analysis (WGCNA) on 94,114 expressed HC genes. 58% of these genes (54,401) could be assigned to 38 modules (Fig. 3D, Database S4) and, consistent with the PCA, tissues were the major driver of module identity (Fig. 3D, Fig. S38 – S40). The analysis focused initially on the 9,009 triads (syntenic and non-syntenic) with a 1:1:1 A:B:D relationship and for which all homeologs were assigned to a module. 16.4% of the triads had at least one homeolog in a divergent module with the B homeolog most likely to be divergent (37.4% B divergent vs 31.7% A divergent and 30.9% D divergent triads, χ^2 P = 0.007). However, the expression profiles of the majority (83.6%) of triads were relatively consistent with all homeologs in the same (57.6%) or a closely related module (26.0%). The proportion of homeologs found within the same module was higher than expected, pointing to a highly-conserved expression pattern of homeologs across the 850 RNA-Seq samples (Fig. 3E, Table S34). Triads with at least one gene in a non-syntenic position had more divergent expression patterns compared to syntenic triads (21.2% vs 16.2%, χ2 P <0.001) and fewer triads with all homeologs in the same module (48.7%) compared to syntenic triads (58.0%, χ2 P= 0.009). Similar patterns were observed in the 1,933 duplets having a 1:1 relationship between only two homeologs (Table S34). These results are consistent with syntenic homeologs showing similar expression patterns while more dramatic changes in chromosome

context associate with divergent expression and possible sub- or neo-functionalization. These trends were also found across diverse tissue-specific networks (*22*).

To explore the potential of the WGCNA network for identifying novel pathways in wheat, a search was undertaken for modules containing known regulators of wheat flowering time [eg. *PPD1*, (*35*); *FT* (*36*); Fig. 3F]. Genes belonging to this pathway were grouped into specific modules. The upstream genes (*PHYB, PHYC, PPD1, ELF3, VRN2*) were present mainly in modules 1 and 5 and were most highly correlated with expression in leaf/shoot tissues (0.68 and 0.67 respectively, Padj<E-108). In contrast, the integrating gene *FT* and downstream genes *VRN1*, *FUL2* and *FUL3* were found in modules 8 and 11, most highly correlated with expression in spikes (0.69 and 0.65 respectively, Padj<E-101, Table S35). The MADS_II TF family generally associated with the above pathways, was examined more closely with a focus on the gene tree OG0000041 containing 54 of the 118 MADS_II genes in wheat. 24 MADS_II genes from modules 8 and 11 were identified within this gene tree, clustering into two main clades along with Arabidopsis and rice orthologs associated with floral patterning (Fig. S41; Database S5). Within these clades, other MADS_II genes were found that were not in modules 8 or 11 (Fig. 3G), indicating a different pattern of co-expression. None of the 24 MADS_II genes had a simple 1:1 ortholog in Arabidopsis, suggesting that some wheat orthologs function in flowering (those within modules 8 and 11), whereas others could have developed different functions, despite being phylogenetically closely related. Thus, these data provide a framework to identify and prioritize the most likely functional orthologs of known model system genes within polyploid wheat, to characterize them functionally (*37*) and to dissect genetic factors controlling important agronomic traits (*38, 39*). A more detailed analysis of tissue-specific and stress-related networks

(*22*) provides a framework for defining quantitative variation and interactions between homeologs for many agronomic traits (*40*).

Gene family expansion / contraction with relevance to wheat traits

Gene duplication and gene family expansion are important mechanisms of evolution and environmental adaptation, as well as major contributors to phenotypic diversity (*41*, *42*). In a phylogenomic comparative analysis, wheat gene family size and wheat-specific gene family expansion / contraction were benchmarked against nine other grass genomes, including five closely related diploid Triticeae species (Table S23, Figs. S13-15, S42). A total of 30,597 gene families (groups of orthologous genes traced to a last common ancestor in the evolutionary hierarchy of the compared taxa) were defined with 26,080 families containing gene members from at least one of the three wheat sub-genomes (Tables S36-S39). Among the 8,592 expanded wheat gene families (33% of all families), 6,216 were expanded in all three A, B and D sub-genomes (24%; either shared with the wild ancestor or specific to bread wheat, Fig. 4A). Another 1,109 were expanded in only one of the wheat sub-genomes and 2,102 gene families were expanded in either the A or the D genome lineages (Figs. 4A, S43, Table S36). Overall, only 78 gene families were contracted in wheat. Numbers of gene families only expanded in wheat may be overestimated due to limited completeness of the draft progenitor wheat genome assemblies used in this study (*13*) (Table S39). Gene Ontology (GO; ontology of biomedical terms for the areas ‘cellular component’, ‘biological process’, ‘molecular function’), Plant Ontology (PO; ontology terms describing anatomical structures and growth and developmental stages across Viridiplantae) and Plant Trait Ontology (TO; ontology of controlled vocabulary to describe phenotypic traits and QTLs that were physically mapped to a gene in flowering plant species)

analysis identified 1,169 distinct GO/PO/TO terms (15% of all assigned terms) enriched in genes belonging to expanded wheat gene families (Figs. 4B, S44, S45). 'A sub-genome' or 'A-lineage' expanded gene families showed a bias for terms associated with seed formation [overrepresentation of the TO term "plant embryo morphology" (TO:0000064) and several seed, endosperm, and embryo-developmental GO terms] (Fig. S46). Similarly, 'B sub-genome' expanded gene families were enriched for TO terms related to plant vegetative growth and development (Database S6, Fig. S47). Gene families that were expanded in all wheat sub-genomes were enriched for 14 TO terms associated with yield-affecting morphological traits and five terms associated with fertility and abiotic stress tolerance (Fig. 4B), which was also mirrored by enrichment for GO and PO terms associated with adaptation to abiotic stress ('salt stress', 'cold stress') and grain yield and quality ('seed maturation', 'dormancy' and 'germination'). The relationship between the patterns of enriched TO/PO/GO terms for expanded wheat gene families and key characteristics of wheat performance (Fig. S45-S51) provides a resource (Database S6) to explore future QTL mapping and candidate gene identification for breeding.

Many gene families with high relevance to wheat breeding and improvement were among the expanded group and their genomic distribution was analyzed in greater detail (Figs. 4C, S52-S54). Disease resistance related NLR (nucleotide-binding site leucine-rich repeat)-like loci and WAK (wall-associated receptor)-like genes were clustered in high numbers at the distal (R1 and R3) regions of all chromosome arms, with NLRs often co-localizing with known disease resistance loci (Fig. 4C). The Restorer of Fertility-Like (RFL) sub-clade of P class PPR proteins, potentially of interest for hybrid wheat production, comprised 207 genes, nearly three-fold more per haploid sub-genome than in any other plant genome analyzed to date (*43*, *44*). They localized

mainly as clusters of genes in regions on the group 1, 2, and 6 chromosomes, which carry fertility restoration QTLs in wheat (Figs. 4C, S54). Among the dehydrin gene family, implicated with drought tolerance in plants, 25 genes that formed well defined clusters on chromosomes 6A, 6B and 6D (Figs. S53, S55) showed early increased expression under severe drought stress. As the structural variation in the *CBF* genes of wheat is known to be associated with winter survival (*45*), the array of CBF paralogs at the Fr-2 locus (Fig. S56) revealed by IWGSC RefSeqv1.0 provides a basis for targeted allele mining for novel CBF haplotypes from highly frost tolerant wheat genetic resources. Lastly, high levels of expansion and variation in members of grain prolamin gene families (Fig. S52 (*46*)) that can either be related to the response to heat stress or whose protein epitopes are associated with levels of coeliac disease and food allergies (*46*), provide candidates for future selection in breeding programs. From these few examples, it is evident that flexibility in gene copy numbers within the wheat genome has contributed to the adaptability of wheat to produce high quality grain under diverse climates and environments (*47*). Knowledge of the complex picture of the genome-wide distribution of gene families (Fig. 4C), that needs to be considered for selection in breeding programs in the context of distribution of recombination and allelic diversity can now be applied in wheat improvement strategies. This is especially true if ‘must-have traits’ that are allocated in chromosomal compartments with highly contrasting characteristics, are fixed in repulsion, or are found only in incompatible genepools of the respective breeding germplasm.

Rapid trait improvement using physically resolved markers and genome editing

The selection and modification of genetic variation underlying agronomic traits in breeding programs is often complicated if phenotypic selection depends on the expression of multiple loci with quantitative effects that can be strongly influenced by the environment. This dilemma can be overcome if DNA markers in strong linkage disequilibrium with the phenotype are identified through forward genetic approaches, or if the underlying genes can be targeted through genome editing. The potential for IWGSC RefSeq v1.0, together with the detailed genome annotation, to accelerate the identification of potential candidate genes underlying important agronomic traits was exemplified for two targets. A forward genetics approach was used to fully resolve a QTL for stem solidness (*SSt1*) conferring resistance to drought stress and to insect damage (*48*) that was disrupted in previous wheat assemblies by a lack of scaffold ordering and annotation, partial assembly, and/or incomplete gene models (Fig. S57, Tables S40, S41). In IWGSC RefSeq v1.0, *SSt1* contains 160 HC genes (Table S42), of which 26 were differentially expressed (DESeq2, Benjamini-Hochberg-adj p <0.01) between wheat lines with contrasting phenotypes. One of the differentially expressed genes, *TraesCS3B01G608800* was present as a single copy in RefSeq v1.0, but showed copy number variation (CNV) associated with stem-solidness in a diverse panel of hexaploid cultivars (Figs. 5A, S58, Table S43). Using IWGSC RefSeq v1.0, we developed a diagnostic SNP marker physically linked to the CNV that has been deployed to select for stem-solidness in wheat breeding programs (Fig. 5B).

Knowledge from model species can also be used to annotate genes and provide a route to trait enhancement through reverse genetics. The approach here targeted flowering time which is important for crop adaptation to diverse environments and is well-studied in model plants. Six wheat homologues of the *Flowering Locus C* (*FLC*) gene have been identified as having a role in

the vernalization response, a critical process regulating flowering time (*49*). IWGSC RefSeqv1.0 was used to refine the annotation of these six sequences to identify four HC genes and then to design guide RNAs to specifically target by CRISPR/Cas9-based gene editing one of these genes, *TaAGL33,* on all sub-genomes [*TraesCS3A01G435000 (A), TraesCS3B01G470000* (B), and *TraesCS3D01G428000* (D)] (Fig. 5C, (*13*)). Editing was obtained at the targeted gene and led to truncated proteins after the MADS box through small deletions/insertions (Fig. 5D). Expression of all homeologs was high prior to vernalization, dropped during vernalization, and remained low post-vernalization, implying a role for this gene in flowering control. This expression pattern was not strongly affected by the genome edits (Fig. S59). Plants with the editing events in the D-genome flowered 2-3 days earlier than controls (Fig. 5E). Further refinement should help to fully understand the significance of the *TaAGL33* gene for vernalization in monocots. These results exemplify how the IWGSC RefSeqv1.0 could accelerate the development of diagnostic markers and the design of targets for genome editing for traits relevant to breeding.

Conclusions

IWGSC RefSeq v1.0 is a resource that has a potential for disruptive innovation in wheat improvement. By necessity, breeders work with the genome at the whole chromosome level, as each new cross involves the modification of genome-wide gene networks that control the expression of complex traits such as yield. With the annotated and ordered reference genome sequence in place, researchers and breeders can now easily access sequence level information to define changes in the genomes of lines in their programs. While several hundred wheat QTLs have been published, only a small number of genes have been cloned and functionally characterized. IWGSC RefSeq v1.0 underpins immediate application by providing access to

regulatory regions and it will serve as the backbone to anchor all known QTLs to one common annotated reference. Combining this knowledge with the distribution of meiotic recombination frequency, and genomic diversity will enable breeders to tackle more efficiently the challenges imposed by the need to balance the parallel selection processes for adaptation to biotic and abiotic stress, end-use quality, and yield improvement. Strategies can now be defined more precisely to bring desirable alleles into coupling phase, especially in less recombinant regions of the wheat genome. Here the full potential of the newly available genome information may be realized by the implementation of DNA marker platforms and targeted breeding technologies, including genome editing (*50*).

Methods Summary

Whole genome sequencing of cultivar 'Chinese Spring' by short read sequencing-by-synthesis provided the data for de novo genome assembly and scaffolding with the software package DenovoMAGIC2^{TM}. The assembly was super-scaffolded and anchored into 21 pseudomolecules with high density genetic (POPSEQ) and physical (Hi-C and 21 chromosome-specific physical maps) mapping information and by integrating additional genomic resources. Validation of the assembly used independent genetic (de novo GBS maps) and physical mapping evidence (Radiation hybrid maps, BioNano ‘optical maps’ for group 7 homeologous chromosomes). The genome assembly was annotated for genes, repetitive DNA, and other genomic features and in-depth comparative analyses were carried out to analyze the distribution of genes, recombination, position and size of centromeres and the expansion/contraction of wheat gene families. An atlas of wheat gene transcription was built from an extensive panel of 850 independent transcriptome datasets which was then used to study gene co-expression networks. Furthermore, the assembly

was used for the dissection of an important stem solidness QTL and to design targets for genome editing of genes implied in flowering time control in wheat. Detailed methodological procedures are described in the supplementary materials.

References and Notes:

1. Food and Agriculture Organization of the United Nations, FAOSTAT Statistics Database, 78http://www.fao.org/faostat/en/#data/FBS, http://www.fao.org/faostat/en/#data/QC (2017.

2. G. N. Atlin, J. E. Cairns, B. Das, Rapid breeding and varietal replacement are critical to adaptation of cropping systems in the developing world to climate change. Global Food Security 12, 31-37 (2017).

3. J. M. Hickey, T. Chiurugwi, I. Mackay, W. Powell, C. B. P. W. P. Implementing Genomic Selection in, Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. Nat Genet 49, 1297-1303 (2017).

4. K. Arumuganathan, E. D. Earle, Nuclear DNA content of some important plant species. Plant Molecular Biology Reporter 9, 208-218 (1991).

5. The International Wheat Genome Sequencing Consortium, A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. Science 345, (2014).

6. J. A. Chapman et al., A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. Genome Biology 16, 26 (2015).

7. B. J. Clavijo et al., An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. Genome Research 27, 885-896 (2017).

8. A. V. Zimin et al., The first near-complete assembly of the hexaploid bread wheat genome, Triticum aestivum. GigaScience 6, 1-7 (2017).

9. T. R. Endo, B. S. Gill, The Deletion Stocks of Common Wheat. J Hered 87, 295-307 (1996).

10. M. E. Sorrells et al., Comparative DNA sequence analysis of wheat and rice genomes. Genome Res 13, 1818 - 1827 (2003).

11. K. Eversole, J. Rogers, B. Keller, R. Appels, C. Feuillet, in Achieving Sustainable Cultivation of Wheat. (Burleigh-Dodds Science Publishing, 2017), vol. 1, chap. 2.

12. E. Paux et al., Insertion site-based polymorphism markers open new perspectives for genome saturation and marker-assisted selection in wheat. Plant Biotechnology Journal 8, 196-210 (2010).

13. Supplementary Materials.

14. The Arabidopsis Genome Initiative, Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408, 796-815 (2000).

15. International Rice Genome Sequencing Project, The map-based sequence of the rice genome. Nature 436, 793-800 (2005).

16. T. Wicker et al., Impact of transposable elements on genome structure and evolution in wheat. BioRxiv DOI: https://doi.org/10.1101/363192 (2018).

17. F. Choulet et al., Structural and functional partitioning of bread wheat chromosome 3B. Science 345, (2014).

18. X. Guo et al., De Novo Centromere Formation and Centromeric Sequence Expansion in Wheat and its Wide Hybrids. PLOS Genetics 12, e1005997 (2016).

19. K. Wang, Y. Wu, W. Zhang, R. K. Dawe, J. Jiang, Maize centromeres expand and adopt a uniform size in the genetic background of oat. Genome Research 24, 107-116 (2014).

20. Y. Jiao et al., Improved maize reference genome with single-molecule technologies. Nature 546, 524 (2017).

21. H. Yan et al., Intergenic Locations of Rice Centromeric Chromatin. PLoS Biology 6, e286 (2008).

22. R. Ramirez-Gonzalez et al., The transcriptional landscape of hexaploid wheat across tissues and cultivars. Science DOI: 10.1126/science.aar6089 (2018).

23. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210-3212 (2015).

24. M.-D. Rey et al., Exploiting the ZIP4 homologue within the wheat Ph1 locus has identified two lines exhibiting homoeologous crossover in wheat-wild relative hybrids. Molecular Breeding 37, 95 (2017).

25. F. Cheng et al., Gene retention, fractionation and subgenome differences in polyploid plants. Nature Plants 4, 258-268 (2018).

26. T. Marcussen et al., Ancient hybridizations among the ancestral genomes of bread wheat. Science 345, (2014).

27. Y. Van de Peer, S. Maere, A. Meyer, The evolutionary significance of ancient genome duplications. Nature Reviews Genetics 10, 725 (2009).

28. P. S. Soltis, D. E. Soltis, Ancient WGD events as drivers of key innovations in angiosperms. Current Opinion in Plant Biology 30, 159-165 (2016).

29. M. Melé et al., The human transcriptome across tissues and individuals. Science 348, 660-665 (2015).

30. S. C. Stelpflug et al., An expanded maize gene expression atlas based on RNA sequencing and its use to explore root development. The Plant Genome 9, (2016).

31. F. He et al., Large-scale atlas of microarray data reveals the distinct expression landscape of different tissues in Arabidopsis. The Plant Journal 86, 472-480 (2016).

32. X. Wang et al., Comparative genomic analysis of C4 photosynthetic pathway evolution in grasses. Genome Biology 10, R68 (2009).

33. L. Xia et al., Rice Expression Database (RED): An integrated RNA-Seq-derived gene expression database for rice. Journal of Genetics and Genomics 44, 235-241 (2017).

34. R. J. Schaefer, J.-M. Michno, C. L. Myers, Unraveling gene function in agricultural species using gene co-expression networks. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1860, 53-63 (2017).

35. J. Beales, A. Turner, S. Griffiths, J. Snape, D. Laurie, A pseudo-response regulator is misexpressed in the photoperiod insensitive Ppd-D1a mutant of wheat (Triticum aestivum L.). Theoretical and Applied Genetics 115, 721-733 (2007).

36. L. Yan et al., The wheat and barley vernalization gene VRN3 is an orthologue of FT. Proceedings of the National Academy of Sciences 103, 19581-19586 (2006).

37. K. V. Krasileva et al., Uncovering hidden variation in polyploid wheat. Proceedings of the National Academy of Sciences 114, E913-E921 (2017).

38. Y. Wang et al., Transcriptome Association Identifies Regulators of Wheat Spike Architecture. Plant Physiology 175, 746-757 (2017).

39. M. Pfeifer et al., Genome interplay in the grain transcriptome of hexaploid bread wheat. Science 345, (2014).

40. P. Borrill, N. Adamski, C. Uauy, Genomics as the key to unlocking the polyploid potential of wheat. New Phytologist 208, 1008-1022 (2015).

41. F. A. Kondrashov, Gene duplication as a mechanism of genomic adaptation to a changing environment. Proceedings of the Royal Society B: Biological Sciences 279, 5048-5057 (2012).

42. P. Schiffer, J. Gravemeyer, M. Rauscher, T. Wiehe, Ultra large gene families: a matter of adaptation or genomic parasites? Life 6, 32 (2016).

43. T. Sykes et al., In Silico Identification of Candidate Genes for Fertility Restoration in Cytoplasmic Male Sterile Perennial Ryegrass (Lolium perenne L.). Genome Biology and Evolution 9, 351-362 (2017).

44. J. Melonek, J. D. Stone, I. Small, Evolutionary plasticity of restorer-of-fertility-like proteins in rice. Scientific Reports 6, 35152 (2016).

45. T. Würschum, C. F. H. Longin, V. Hahn, M. R. Tucker, W. L. Leiser, Copy number variations of CBF genes at the Fr-A2 locus are essential components of winter hardiness in wheat. The Plant Journal 89, 764-773 (2017).

46. A. Juhász et al., Wheat proteins as a source of food intolerance: Genome mapping and influence of environment. Science Advances DOI: 10.1126/sciadv.aar8602 (2018).

47. M. Feldman, A. A. Levy, in Alien Introgression in Wheat: Cytogenetics, Molecular Biology, and Genomics, M. Molnár-Láng, C. Ceoloni, J. Doležel, Eds. (Springer International Publishing, Cham, 2015), pp. 21-76.

48. K. T. Nilsen et al., High density mapping and haplotype analysis of the major stem-solidness locus SSt1 in durum and common wheat. PLOS ONE 12, e0175285 (2017).

49. N. Sharma et al., A flowering locus C homolog is a vernalization-regulated repressor in Brachypodium and is cold regulated in wheat. Plant Physiology 173, 1301-1315 (2017).

50. H. Puchta, Applying CRISPR/Cas for genome engineering in plants: the best is yet to come. Current Opinion in Plant Biology 36, 1-8 (2017).

51. H.-Q. Ling et al., Draft genome of the wheat A-genome progenitor Triticum urartu. Nature 496, 87 (2013).

52. J. Jia et al., Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation. Nature 496, 91-95 (2013).

53. Y. Ishida, M. Tsunashima, Y. Hiei, T. Komari, in Agrobacterium Protocols: Volume 1, K. Wang, Ed. (Springer New York, New York, NY, 2015), pp. 189-198.

54. M. Alaux et al., Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data. BioRxiv DOI: https://doi.org/10.1101/363259 (2018).

55. R. Avni et al., Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science 357, 93-97 (2017).

56. F. Choulet et al., Megabase Level Sequencing Reveals Contrasted Organization and Evolution Patterns of the Wheat Gene and Transposable Element Spaces. The Plant Cell 22, 1686-1701 (2010).

57. G. Keeble-Gagnère et al., Optical and physical mapping with local finishing enables megabase-scale resolution of agronomically important regions on wheat chromosome 7A. BioRxiv DOI: BIORXIV/2018/363465 (2018).

58. R. Kalhor, H. Tjong, N. Jayathilaka, F. Alber, L. Chen, Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. Nat Biotech 30, 90-98 (2012).

59. E. Lieberman-Aiden et al., Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science 326, 289-293 (2009).

60. S. Beier et al., Construction of a map-based reference genome sequence for barley, Hordeum vulgare L. Scientific Data 4, 170044 (2017).

61. J. Šafář et al., Dissecting large and complex genomes: flow sorting and BAC cloning of individual chromosomes from bread wheat. The Plant Journal 39, 960-968 (2004).

62. J. Šafář et al., Development of chromosome-specific BAC resources for genomics of bread wheat. Cytogenetic and Genome Research 129, 211-223 (2010).

63. M.-C. Luo et al., High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. Genomics 82, 378 (2003).

64. J. van Oeveren et al., Sequence-based physical mapping of complex genomes by whole genome profiling. Genome Research 21, 618-625 (2011).

65. C. Soderlund, S. Humphray, I. Dunham, L. French, Contigs built with fingerprints, markers, and FPC V4.7. Genome Res 11, 934 - 941 (2000).

66. Z. Frenkel, E. Paux, D. Mester, C. Feuillet, A. Korol, LTC: a novel algorithm to improve the efficiency of contig assembly for physical mapping in complex genomes. BMC Bioinformatics 11, 584 (2010).

67. H. Staňková et al., BioNano genome mapping of individual chromosomes supports physical mapping and sequence assembly in complex plant genomes. Plant Biotechnology Journal 14, 1523-1531 (2016).

68. N. Poursarebani et al., Whole-genome profiling and shotgun sequencing delivers an anchored, gene-decorated, physical map assembly of bread wheat chromosome 6A. The Plant Journal 79, 334-347 (2014).

69. F. Kobayashi et al., A high-resolution physical map integrating an anchored chromosome with the BAC physical maps of wheat chromosome 6B. BMC Genomics 16, 595 (2015).

70. M. Kubaláková, J. Vrána, J. Číhalíková, H. Šimková, J. Doležel, Flow karyotyping and chromosome sorting in bread wheat (Triticum aestivum L.). Theoretical and Applied Genetics 104, 1362-1372 (2002).

71. V. K. Tiwari et al., A whole-genome, radiation hybrid mapping resource of hexaploid wheat. The Plant Journal 86, 195-207 (2016).

72. H. Rimbert et al., High throughput SNP discovery and genotyping in hexaploid wheat. PLOS ONE 13, e0186329 (2018).

73. S. de Givry, M. Bouchez, P. Chabrier, D. Milan, T. Schiex, Carhta Gene: multipopulation integrated genetic and radiation hybrid mapping. Bioinformatics 21, 1703-1704 (2005).

74. M. E. Sorrells et al., Reconstruction of the Synthetic W7984 × Opata M85 wheat reference population. Genome 54, 875-882 (2011).

75. J. A. Poland, P. J. Brown, M. E. Sorrells, J.-L. Jannink, Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE 7, e32253 (2012).

76. Y. Wu, P. R. Bhat, T. J. Close, S. Lonardi, Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. PLoS Genet 4, e1000212 (2008).

77. J. Doležel, M. Kubaláková, E. Paux, J. Bartoš, C. Feuillet, Chromosome-based genomics in the cereals. Chromosome Research 15, 51-66 (2007).

78. A. A. Myburg et al., The genome of Eucalyptus grandis. Nature 510, 356 (2014).

79. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754-1760 (2010).

80. H. Li et al., The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079 (2009).

81. R. Whitford et al., Hybrid breeding in wheat: technologies to improve hybrid wheat seed production. Journal of Experimental Botany 64, 5411-5428 (2013).

82. J. N. Burton et al., Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nature Biotechnology 31, 1119 (2013).

83. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841-842 (2010).

84. J. Daron et al., Organization and evolution of transposable elements along the bread wheat chromosome 3B. Genome Biology 15, 546 (2014).

85. S. Kurtz et al., Versatile and open software for comparing large genomes. Genome Biology 5, R12 (2004).

86. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10-12 (2011).

87. P. Leroy et al., TriAnnot: A Versatile and High Performance Pipeline for the Automated Annotation of Plant Genomes. Frontiers in Plant Science 3, 5 (2012).

88. A. F. Smit, Identification of a new, abundant superfamily of mammalian LTR-transposons. Nucleic Acids Research 21, 1863-1872 (1993).

89. S. F. Altschul et al., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl. Acids Res. 25, 3389-3402 (1997).

90. L. Pingault et al., Deep transcriptome sequencing provides new insights into the structural and functional organization of the wheat genome. Genome Biology 16, 29 (2015).

91. L. Dong et al., Single-molecule real-time transcript sequencing facilitates common wheat genome annotation and grain transcriptome research. BMC Genomics 16, 1039 (2015).

92. The International Barley Genome Sequencing Consortium (IBSC) et al., A physical, genetical and functional sequence assembly of the barley genome. Nature 491, 711-716 (2012).

93. G. S. C. Slater, E. Birney, Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6, 31 (2005).

94. M. Stanke, S. Waack, Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 2, 215-225 (2003).

95. N. Amano, T. Tanaka, H. Numa, H. Sakai, T. Itoh, Efficient plant gene identification based on interspecies mapping of full-length cDNAs. DNA Research 17, 271-279 (2010).

96. T. D. Wu, C. K. Watanabe, GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21, 1859-1875 (2005).

97. C. Trapnell et al., Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology 28, 511 (2010).

98. D. Kim, B. Langmead, S. L. Salzberg, HISAT: a fast spliced aligner with low memory requirements. Nature Methods 12, 357 (2015).

99. D. W. Barnett, E. K. Garrison, A. R. Quinlan, M. P. Strömberg, G. T. Marth, BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics 27, 1691-1692 (2011).

100. M. Pertea et al., StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature Biotechnology 33, 290 (2015).

101. The UniProt Consortium, UniProt: the universal protein knowledgebase. Nucleic Acids Research 45, D158-D169 (2017).

102. G. Gremme, V. Brendel, M. E. Sparks, S. Kurtz, Engineering a software tool for gene structure prediction in higher organisms. Information and Software Technology 47, 965-978 (2005).

103. J. Keilwagen et al., Using intron position conservation for homology-based gene prediction. Nucleic Acids Research 44, e89-e89 (2016).

104. K. Mochida, T. Yoshida, T. Sakurai, Y. Ogihara, K. Shinozaki, TriFLDB: a database of clustered full-length coding sequences from Triticeae with applications to comparative grass genomics. Plant Physiology 150, 1135-1146 (2009).

105. S. Ghosh, C.-K. K. Chan, in Plant Bioinformatics: Methods and Protocols, D. Edwards, Ed. (Springer New York, New York, NY, 2016), pp. 339-361.

106. S. R. Eddy, Accelerated Profile HMM Searches. PLOS Computational Biology 7, e1002195 (2011).

107. L. Venturini, S. Caim, G. Kaithakottil, D. L. Mapleson, D. Swarbreck, Leveraging multiple transcriptome assembly methods for improved gene structure annotation. bioRxiv doi.org/10.1101/216994, (2017).

108. D. Mapleson, L. Venturini, G. Kaithakottil, D. Swarbreck, Efficient and accurate detection of splice junctions from RNAseq with Portcullis. bioRxiv doi.org/10.1101/217620, (2017).

109. A. Kozomara, S. Griffiths-Jones, miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Research 39, D152-D157 (2011).

110. S. J. Lucas, H. Budak, Sorting the wheat from the chaff: identifying miRNAs in genomic survey sequences of Triticum aestivum chromosome 1AL. PLOS ONE 7, e40859 (2012).

111. N. R. Markham, M. Zuker, in Bioinformatics: Structure, Function and Applications, J. M. Keith, Ed. (Humana Press, Totowa, NJ, 2008), pp. 3-31.

112. H. B. Cagirici, S. Biyiklioglu, H. Budak, Assembly and annotation of transcriptome provided evidence of miRNA mobility between wheat and wheat stem sawfly. Frontiers in Plant Science 8, 1653 (2017).

113. B. A. Akpinar, M. Kantar, H. Budak, Root precursors of microRNAs in wild emmer and modern wheats show major differences in response to drought stress. Functional & Integrative Genomics 15, 587-598 (2015).

114. T. M. Lowe, S. R. Eddy, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research 25, 955-964 (1997).

115. The Gene Ontology Consortium, Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Research 45, D331-D338 (2017).

116. L. Cooper et al., The Plant Ontology as a tool for comparative plant anatomy and genomic analyses. Plant and Cell Physiology 54, e1-e1 (2013).

117. E. Arnaud et al., Towards a reference Plant Trait Ontology for modeling knowledge of plant traits and phenotypes. DOI:10.13140/2.1.2550.3525, (2012).

118. P. Borrill, R. Ramirez-Gonzalez, C. Uauy, expVIP: a customizable RNA-seq data analysis and visualization platform. Plant Physiology 170, 2172-2186 (2016).

119. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. Nat Biotech 34, 525-527 (2016).

120. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559 (2008).

121. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15, 550 (2014).

122. Y. Benjamini, D. Yekutieli, The Control of the false discovery rate in multiple testing under dependency. The Annals of Statistics 29, 1165-1188 (2001).

123. F. Krueger, S. R. Andrews, Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27, 1571-1572 (2011).

124. A. Veluchamy et al., LHP1 regulates H3K27me3 spreading and shapes the three-dimensional conformation of the Arabidopsis genome. PLOS ONE 11, e0158936 (2016).

125. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114-2120 (2014).

126. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357 (2012).

127. A. D. Zimmer et al., Reannotation and extended community resources for the genome of the non-seed plant Physcomitrella patens provide insights into the evolution of plant gene structures and functions. BMC Genomics 14, 498 (2013).

128. J. S. Bernardes, F. R. J. Vieira, G. Zaverucha, A. Carbone, A multi-objective optimization approach accurately resolves protein domain architectures. Bioinformatics 32, 345-353 (2016).

129. D. M. Emms, S. Kelly, OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biology 16, 157 (2015).

130. J. Huerta-Cepas, H. Dopazo, J. Dopazo, T. Gabaldón, The human phylome. Genome Biology 8, R109 (2007).

131. J. Huerta-Cepas, F. Serra, P. Bork, ETE 3: reconstruction, analysis, and visualization of phylogenomic data. Molecular Biology and Evolution 33, 1635-1638 (2016).

132. S. Mirarab, T. Warnow, ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31, i44-i52 (2015).

133. D. Lang et al., Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity. Genome Biology and Evolution 2, 488-503 (2010).

134. J. T. Garland, A. W. Dickerman, C. M. Janis, J. A. Jones, Phylogenetic analysis of covariance by computer simulation. Systematic Biology 42, 265-292 (1993).

135. M.-C. Luo et al., Genome sequence of the progenitor of the wheat D genome Aegilops tauschii. Nature, (2017).

136. S. Grossmann, S. Bauer, P. N. Robinson, M. Vingron, Improved detection of overrepresentation of Gene-Ontology annotations with parent–child analysis. Bioinformatics 23, 3024-3031 (2007).

137. S. Aibar, C. Fontanillo, C. Droste, J. De Las Rivas, Functional Gene Networks: R/Bioc package to generate and analyse gene networks derived from functional enrichment and clustering. Bioinformatics 31, 1686-1688 (2015).

138. G. Su, A. Kuchinsky, J. H. Morris, D. J. States, F. Meng, GLay: community structure analysis of biological networks. Bioinformatics 26, 3135-3137 (2010).

139. N.-p. D. Nguyen, S. Mirarab, K. Kumar, T. Warnow, Ultra-large alignments using phylogeny-aware profiles. Genome Biology 16, 124 (2015).

140. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2 – approximately maximum-likelihood trees for large alignments. PLOS ONE 5, e9490 (2010).

141. I. Letunic, P. Bork, Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Research 44, W242-W245 (2016).

142. N. M. Glover, H. Redestig, C. Dessimoz, Homoeologs: what are they and how do we infer them? Trends in Plant Science 21, 609-621 (2016).

143. Y. Wang et al., MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Research 40, e49-e49 (2012).

144. M. Mascher et al., A chromosome conformation capture ordered sequence of the barley genome. Nature 544, 427-433 (2017).

145. L. Al Ait, Z. Yamak, B. Morgenstern, DIALIGN at GOBICS—multiple sequence alignment using various sources of external information. Nucleic Acids Research 41, W3-W7 (2013).

146. D. Wang, Y. Zhang, Z. Zhang, J. Zhu, J. Yu, KaKs_Calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies. Genomics, Proteomics & Bioinformatics 8, 77-80 (2010).

147. J. C. Zadoks, T. T. Chang, C. F. Konzak, A decimal code for the growth stages of cereals. Weed Research (Oxford) 14, 415-421 (1974).

148. A. Dobin et al., STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15-21 (2013).

149. M. Pertea, D. Kim, G. M. Pertea, J. T. Leek, S. L. Salzberg, Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat. Protocols 11, 1650-1667 (2016).

150. K. J. Livak, T. D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. Methods 25, 402-408 (2001).

151. A. H. Paterson et al., The Sorghum bicolor genome and the diversification of grasses. Nature 457, 551 (2009).

152. P. S. Schnable et al., The B73 Maize genome: complexity, diversity, and dynamics. Science 326, 1112-1115 (2009).

153. E. Bauer et al., Towards a whole-genome sequence for rye (Secale cereale L.). The Plant Journal 89, 853-869 (2017).

154. M. J. Sanderson, Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Molecular Biology and Evolution 19, 101-109 (2002).

155. P.-A. Christin et al., Molecular dating, evolutionary rates, and the age of the grasses. Systematic Biology 63, 153-165 (2014).

156. X. C. Zhao, P. J. Sharp, Production of all eight genotypes of null alleles at 'waxy' loci in bread wheat, Triticum aestivum L. Plant Breeding 117, 488-490 (1998).

157. K. Jordan et al., A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. Genome Biology 16, 48 (2015).

158. M. Krzywinski et al., Circos: an information aesthetic for comparative genomics. Genome Res 19, 1639 - 1645 (2009).

159. C. E. Niederhuth et al., Widespread natural variation of DNA methylation within angiosperms. Genome Biology 17, 194 (2016).

160. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) 57, 289-300 (1995).

161. G. Yu et al., GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinformatics 26, 976-978 (2010).

162. S. Fischer et al., in Current Protocols in Bioinformatics. (John Wiley & Sons, Inc., 2002).

163. S. Cheng et al., Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. The Plant Journal 85, 532-547 (2016).

164. Z. Li et al., SSR analysis and identification of fertility restorer genes Rf1 and Rf4 of Triticum timopheevii cytoplasmic male sterility (T-CMS) in wheat (Triticum aestivum L.). Journal of Agricultural Biotechnology 22, 1114-1122 (2014).

165. M. Geyer, T. Albrecht, L. Hartl, V. Mohler, Exploring the genetics of fertility restoration controlled by Rf1 in common wheat (Triticum aestivum L.) using high-density linkage maps. Molecular Genetics and Genomics 293, 451-462 (2017).

166. P. Sinha, S. M. S. Tomar, Vinod, V. K. Singh, H. S. Balyan, Genetic analysis and molecular mapping of a new fertility restorer gene Rf8 for Triticum timopheevi cytoplasm in wheat (Triticum aestivum L.) using SSR markers. Genetica 141, 431-441 (2013).

167. J. Breen et al., A physical map of the short arm of wheat chromosome 1A. PLOS ONE 8, e80272 (2013).

168. S. J. Lucas et al., Physical mapping integrated with syntenic analysis to characterize the gene space of the long arm of wheat chromosome 1A. PLOS ONE 8, e59542 (2013).

169. D. Raats et al., The physical map of wheat chromosome 1BS provides insights into its gene space organization and evolution. Genome Biology 14, R138 (2013).

170. R. Philippe et al., A high density physical map of chromosome 1BL supports evolutionary studies, map-based cloning and sequencing in wheat. Genome Biology 14, R64 (2013).

171. E. Paux et al., A physical map of the 1-Gigabase bread wheat chromosome 3B. Science 322, 101-104 (2008).

172. K. Holušová et al., Physical map of the short arm of bread wheat chromosome 3D. The Plant Genome 10, (2017).

173. O. Shorinola et al., Association mapping and haplotype analysis of the pre-harvest sprouting resistance locus Phs-A1 reveals a causal role of TaMKK3-A in global germplasm. bioRxiv, 10.1101/131201 (2017).

174. D. Barabaschi et al., Physical mapping of bread wheat chromosome 5A: an integrated approach. The Plant Genome 8, (2015).

175. E. A. Salina et al., Features of the organization of bread wheat chromosome 5BS based on physical mapping. BMC Genomics 19, 80 (2018).

176. B. A. Akpinar et al., The physical map of wheat chromosome 5DS revealed gene duplications and small rearrangements. BMC Genomics 16, 453 (2015).

177. T. Belova et al., Utilization of deletion bins to anchor and order sequences along the wheat 7B chromosome. Theoretical and Applied Genetics 127, 2029-2040 (2014).

178. Z. Tulpová et al., Integrated physical map of bread wheat chromosome arm 7DS to facilitate gene cloning and comparative studies. New Biotechnology, 10.1016/j.nbt.2018.03.003, (2018).

179. F.-H. Lu et al., Independent assessment and improvement of wheat genome assemblies using Fosill jumping libraries. bioRxiv, 10.1101/219352 (2017).

180. M. A. Nesterov et al., Identification of microsatellite loci based on BAC sequencing data and their physical mapping into the soft wheat 5B chromosome. Russian Journal of Genetics: Applied Research 6, 825-837 (2016).

181. E. M. Sergeeva et al., Fine organization of genomic regions tagged to the 5S rDNA locus of the bread wheat 5B chromosome. BMC Plant Biology 17, 183 (2017).

182. M.-C. Luo et al., A 4-gigabase physical map unlocks the structure and evolution of the complex genome of Aegilops tauschii, the wheat D-genome progenitor. Proceedings of the National Academy of Sciences 110, 7940-7945 (2013).

183. X. Zeng et al., The draft genome of Tibetan hulless barley reveals adaptive patterns to the high stressful Tibetan Plateau. Proceedings of the National Academy of Sciences 112, 1095-1100 (2015).

184. Y. Kawahara et al., Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. Rice 6, 4 (2013).

Acknowledgments: The IWGSC would like to thank the following individuals: M. Burrell and C. Bridson (Norwich Biosciences Institute) for computational support of RNA-Seq data; I. Christie (Graminor AS) and H. Rudi (Norwegian University of Life Sciences) for assistance with chromosome 7B; R. P. Davey (Earlham Institute) for assistance with RNA-Seq data; J. Deek (Tel Aviv University) for growing the source plants and DNA extraction used for whole genome sequencing; Z. Dubská, E. Jahnová, M. Seifertová, R. Šperková, R. Tušková, and J. Weiserová (Institute of Experimental Botany, Olomouc) for assistance with flow cytometric chromosome sorting, BAC library construction, and estimation of genome size; S. Durand, V. Jamilloux, M. Lainé, and C. Michotey (URGI, INRA) for assistance with and access to the IWGSC sequence repository; A. Fiebig of the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) for submitting the Hi-C data; T. Florio for the design of the wheat schematic for the expression atlas and Sst Figure (www.flozbox.com/Science_Illustrated); C. Karunakaran and T. Bond of the Canadian Light Source for performing CT imaging; J. Kawai, N. Kondo H. Sano, N. Suzuki, M. Tagami, H. Tarui of RIKEN and H. Fujisawa, Y. Katayose, K. Kurita, S. Mori, Y. Mukai, and H. Sasaki of the Institute of Crop Science, NARO, and T. Matsumoto of Tokyo University of Agriculture for assistance with deep sequencing of chromosome 6B; P. Lenoble and C. Orvain of Genoscope for assistance in the sequencing of chromosome 1B; A. J. Lukaszewski (University of California, Riverside), B. Friebe and J. Raupp (Kansas State University) for providing seeds of wheat telosomic lines for chromosome sorting; C. Maulis (https://polytypo.design, https://propepper.net) for design and graphics of the prolamin super-family chromosome map; M. Seifertová and H. Tvardíková of the Institute of Experimental Botany for assistance with BAC DNA extraction and sequencing for chromosomes 3DS, 4A, and 7DS; and I. Willick and K. Tanino of the University of Saskatchewan for their assistance in sample preparation and the use of lab facilities.

Funding: The authors would like to thank the following for their financial support of research that enabled the completion of the IWGSC RefSeq v1.0 Project: Agence Nationale pour la Recherche (ANR), ANR-11-BSV5-0015- Ploid-Ploid Wheat- Unravelling bases of polyploidy success in wheat and ANR-16-TERC-0026-01- 3DWHEAT; Agriculture and Agri-Food Canada National Wheat Improvement Program and the AgriFlex Program; Alberta Wheat Development Commission through the Canadian Applied Tricticum Genomics (CTAG2); Australian Government, Department of Industry, Innovation, Climate Change, Science, Research and Tertiary Education: Australia China Science and Research Fund Group Mission (Funding Agreement ACSRF00542); Australian Research Council Centre of Excellence in Plant Energy Biology (CE140100008); Australian Research Council Laureate Fellowship (FL140100179); Bayer CropScience; Biotechnology and Biological Sciences Research Council (BBSRC) 20:20 Wheat® (project number BB/J00426X/1), Institute Strategic Programme grant [BB/J004669/1], Designing Future Wheat (DFW) Institute Strategic Programme (BB/P016855/1), the Wheat Genomics for Sustainable Agriculture (BB/J003557/1), and the Anniversary Future Leader Fellowship (BB/M014045/1); Canada First Research Excellence Fund through the Designing Crops for Global Food Security initiative at the University of Saskatchewan; Council for Agricultural Research and Economics, Italy, through CREA-Interomics; Department of Biotechnology, Ministry of Science and Technology, Govt. of India File No. F. No.BT/IWGSC/03/TF/2008; DFG (SFB924) for support of KFXM; European Commission through the *TriticeaeGenome* (FP7-212019); France Génomique (ANR-10-INBS-09,) Genome Canada through the CTAG2 project; Genome Prairie through the CTAG2 project; German Academic Exchange Service (DAAD) PPP Australien 1j16; German Federal Ministry of Food and Agriculture grant 2819103915 WHEATSEQ"; German Ministry of Education and Research grant 031A536 “de.NBI”; Global Institute for Food Security Genomics and Bioinformatics fund; Gordon and Betty Moore Foundation Grant GBMF4725 to Two Blades Foundation; Grain Research Development Corporation (GRDC) Australia; Graminor AS NFR project 199387 - Expanding the technology base for Norwegian wheat breeding; Sequencing wheat chromosome 7B; illumina; INRA - French National Institute for Agricultural Research; International Wheat Genome Sequencing Consortium and its sponsors; Israel Science Foundation grants 999/12, 1137/17, and 1824/12; Junta de Andalucía, Spain, project P12-AGR-0482; MINECO (Spanish Ministry of Economy, Industry and Competitiveness) project BIO2011-15237-E; Ministry of Agriculture, Forestry and Fisheries of Japan through Genomics for Agricultural Innovation, KGS-1003 and through Genomics-based Technology for Agricultural Improvement, NGB-1003; Ministry of Education and Science of Russian Federation project RFMEFI60414X0106 and project RFMEFI60414 X0107; Ministry of Education, Youth and Sport of the Czech Republic Award no. LO1204 (National Program of Sustainability I); Nisshin Flour Milling Inc.; National Research Council of Canada Wheat Flagship program; Norwegian University of Life Sciences (NMBU) NFR project 199387 - Expanding the technology base for Norwegian wheat breeding - Sequencing wheat chromosome 7B; National Science Foundation, United States, Award (FAIN) 1339389, GPF-PG: Genome Structure and Diversity of Wheat and Its Wild Relatives, Award DBI-0701916, and Award IIP-1338897; Russian Science Foundation project 14-14-00161; Saskatchewan Ministry of Agriculture through the CTAG2 project; Saskatchewan Wheat Development Commission through the CTAG2 project; The Czech Science Foundation Award no. 521/06/1723 (Construction of BAC library and physical mapping of the wheat chromosome

3D), Award no. 521-08-1629 (Construction of BAC DNA libraries specific for chromosome 4AL, and positional cloning of gene for adult plant resistance to powdery mildew in wheat), Award no. P501/10/1740 (Physical map of the wheat chromosome 4AL and positional cloning of a gene for yield), Award no. P501/12/2554 (Physical map of wheat chromosome arm 7DS and its use to clone a Russian wheat aphid resistance gene), Award no. P501/12/G090 (Evolution and Function of Complex Plant Genomes), Award no. 14-07164S (Cloning and molecular characterization of wheat QPm-tut-4A gene conferring seedling and adult plant race nonspecific powdery mildew resistance), and Award no. 13-08786S (Chromosome arm 3DS of bread wheat: its sequence and function in allopolyploid genome); The Research Council of Norway (NFR) project 199387 - Expanding the technology base for Norwegian wheat breeding; Sequencing wheat chromosome 7B; United States Department of Agriculture NIFA 2008-35300-04588, the University of Zurich; Western Grains Research Foundation through the CTAG2 project; Western Grains Research Foundation National Wheat Improvement Program; and the Winifred-Asbjornson Plant Science Endowment Fund. The research leading to these results also has received funding from the French Government managed by the Research National Agency (ANR) under the Investment for the Future programme (BreedWheat project ANR-10-BTBR-03), from FranceAgriMer (2011-0971 and 2013-0544), French Funds to support Plant Breeding (FSOV) and from INRA. Axiom genotyping was conducted on the genotyping platform GENTYANE at INRA Clermont-Ferrand (gentyane.clermont.inra.fr). This research was supported in part by the NBI Computing infrastructure for Science (CiS) group through the HPC cluster.

Author contributions: The International Wheat Genome Sequencing Consortium (IWGSC). Authorship of this paper should be cited as "International Wheat Genome Sequencing Consortium" (IWGSC, 2018). Participants are arranged by working group and contributions with leaders/co-leaders and major contributors listed alphabetically first and then other contributors follow alphabetically. Corresponding authors (*), major contributors (†) and working group leader(s) or co-leaders (‡) are indicated.

IWGSC RefSeq Principal Investigators: Rudi Appels[1,36*‡] (rudi.appels@unimelb.edu.au), Kellye Eversole[2,3*‡] (eversole@eversoleassociates.com), Catherine Feuillet[17] (feuillet@bayer.com), Beat Keller[41] (bkeller@botinst.uzh.ch), Jane Rogers[6‡] (janerogersh@gmail.com), and Nils Stein[4,5*‡].

IWGSC Whole Genome Assembly Principal Investigators: Curtis J. Pozniak[11‡] (curtis.pozniak@usask.ca), Nils Stein[4,5*‡] (stein@ipk-gatersleben.de), Frédéric Choulet[7] (frederic.choulet@inra.fr), Assaf Distelfeld[25] (adistel@tauex.tau.ac.il), Kellye Eversole[2,3*] (eversole@eversoleassociates.com), Jesse Poland[28] (jpoland@ksu.edu), Jane Rogers[6] (janerogersh@gmail.com), Gil Ronen[12] (gil@nrgene.com), and Andrew G. Sharpe[43] **(**andrew.sharpe@gifs.ca**).**

Whole Genome Sequencing and Assembly: Curtis Pozniak[11‡] (curtis.pozniak@usask.ca), Gil Ronen[12‡] (gil@nrgene.com), Nils Stein[4,5*‡] (stein@ipk-gatersleben.de), Omer Barad[12†] (omerb@nrgene.com), Kobi Baruch[12†] (kobi@nrgene.com), Frédéric Choulet[7†] (frederic.choulet@inra.fr), Gabriel Keeble-Gagnère[1†] (gabriel.keeble-

gagnere@ecodev.vic.gov.au), Martin Mascher[4,67†] (mascher@ipk-gatersleben.de), Andrew G. Sharpe[43†] (andrew.sharpe@gifs.ca), Gil Ben-Zvi[12†] (bzgil@nrgene.com), and Ambre-Aurore Josselin[7] (ambre-aurore.josselin@inra.fr),

Hi-C Data Based Scaffolding: Nils Stein[4,5*‡] (stein@ipk-gatersleben.de), Martin Mascher[4,67‡] (mascher@ipk-gatersleben.de), and Axel Himmelbach[4] (himmelba@ipk-gatersleben.de).

Whole Genome Assembly QC & Analyses: Frédéric Choulet[7‡] (frederic.choulet@inra.fr), Gabriel Keeble-Gagnère[‡] (gabriel.keeble-gagnere@ecodev.vic.gov.au), Martin Mascher[4,67‡](mascher@ipk-gatersleben.de), Jane Rogers[6‡] (janerogersh@gmail.com), François Balfourier[7] (francois.balfourier@inra.fr), Juan Gutierrez-Gonzalez[30] (jgutierr@umn.edu), Matthew Hayden[1] (matthew.hayden@ecodev.vic.gov.au), Ambre-Aurore Josselin[7] (ambre-aurore.josselin@inra.fr), ChuShin Koh[43] (kevin.koh@gifs.ca), Gary Muehlbauer[30] (muehl003@umn.edu), Raj K Pasam[1] (raj.pasam@ecodev.vic.gov.au), Etienne Paux[7] (etienne.paux@inra.fr), Curtis J. Pozniak[11] (curtis.pozniak@usask.ca), Philippe Rigault[39] (prigault@gydle.com), Andrew G. Sharpe[43] **(**andrew.sharpe@gifs.ca**),** Josquin Tibbits[1] (josquin.tibbits@ecodev.vic.gov.au), and Vijay Tiwari[54] (vktiwari@umd.edu).

Pseudomolecule Assembly: Frédéric Choulet[7‡] (frederic.choulet@inra.fr), Gabriel Keeble-Gagnère[1‡] (gabriel.keeble-gagnere@ecodev.vic.gov.au), Martin Mascher[4,67‡] (mascher@ipk-gatersleben.de), Ambre-Aurore Josselin[7] (ambre-aurore.josselin@inra.fr), and Jane Rogers[6] (janerogersh@gmail.com).

RefSeq Genome Structure and Gene Analyses: Manuel Spannagl[9‡] (manuel.spannagl@helmholtz-muenchen.de), Frédéric Choulet[7‡] (frederic.choulet@inra.fr), Daniel Lang[9†] (daniel.lang@helmholtz-muenchen.de), Heidrun Gundlach[9] (h.gundlach@helmholtz-muenchen.de), Georg Haberer[9] (g.haberer@helmholtz-muenchen.de), Gabriel Keeble-Gagnère[1] (gabriel.keeble-gagnere@ecodev.vic.gov.au), Klaus F.X. Mayer[9,44] (k.mayer@helmholtz-muenchen.de), Danara Ormanbekova[9,48] (danara.ormanbekova2@unibo.it), Etienne Paux[7] (etienne.paux@inra.fr), Verena Prade[9] (verena.prade@helmholtz-muenchen.de), Hana Šimková[8] (simkovah@ueb.cas.cz), and Thomas Wicker[41] (wicker@botinst.uzh.ch).

Automated Annotation: Frédéric Choulet[7‡] (frederic.choulet@inra.fr), Manuel Spannagl[9‡] (manuel.spannagl@helmholtz-muenchen.de), David Swarbreck[50‡] (david.swarbreck@earlham.ac.uk), Hélène Rimbert[7†] (helene.rimbert@inra.fr), Marius Felder[9] (marius.felder@helmholtz-muenchen.de), Nicolas Guilhot[7] (nicolas.guilhot@inra.fr), Heidrun Gundlach[9] (h.gundlach@helmholtz-muenchen.de), Georg Haberer[9] (g.haberer@helmholtz-muenchen.de), Gemy Kaithakottil[50] (Gemy.Kaithakottil@earlham.ac.uk), Jens Keilwagen[40] (jens.keilwagen@julius-kuehn.de), Daniel Lang[9] (daniel.lang@helmholtz-muenchen.de), Philippe Leroy[7] (philippe.leroy.2@inra.fr), Thomas Lux[9] (thomas.lux@helmholtz-muenchen.de), Klaus F.X. Mayer[9,44] (k.mayer@helmholtz-muenchen.de), Sven Twardziok[9] (sven.twardziok@posteo.de), and Luca Venturini[50] (Luca.Venturini@earlham.ac.uk).

Manual Gene Curation: Rudi Appels[1,36‡*] (rudi.appels@unimelb.edu.au), Hélène Rimbert[7†] (helene.rimbert@inra.fr), Frédéric Choulet[7] (frederic.choulet@inra.fr), Angéla Juhász[36,37]

(A.Juhasz@murdoch.edu.au), and Gabriel Keeble-Gagnère[1] (gabriel.keeble-gagnere@ecodev.vic.gov.au).

Sub-Genome Comparative Analyses: Frédéric Choulet[7‡] (frederic.choulet@inra.fr), Manuel Spannagl[9‡] (manuel.spannagl@helmholtz-muenchen.de), Daniel Lang[9†] (daniel.lang@helmholtz-muenchen.de), Michael Abrouk[8,19] (abrouk@ueb.cas.cz), Georg Haberer[9] (g.haberer@helmholtz-muenchen.de), Gabriel Keeble-Gagnère[1] (gabriel.keeble-gagnere@ecodev.vic.gov.au), Klaus F.X. Mayer[9,44] (k.mayer@helmholtz-muenchen.de), and Thomas Wicker[41] (wicker@botinst.uzh.ch).

Transposable Elements: Frédéric Choulet[7‡] (frederic.choulet@inra.fr), Thomas Wicker[41†](wicker@botinst.uzh.ch), Heidrun Gundlach[9†] (h.gundlach@helmholtz-muenchen.de), Daniel Lang[9] (daniel.lang@helmholtz-muenchen.de), and Manuel Spannagl[9] (manuel.spannagl@helmholtz-muenchen.de).

Phylogenomic Analyses: Daniel Lang[9‡] (daniel.lang@helmholtz-muenchen.de) Manuel Spannagl[9‡] (manuel.spannagl@helmholtz-muenchen.de), Rudi Appels[1,36]* (rudi.appels@unimelb.edu.au), and Iris Fischer[9] (iris.fischer@helmholtz-muenchen.de).

Transcriptome Analyses & RNASeq Data: Cristobal Uauy[10‡] (cristobal.uauy@jic.ac.uk), Philippa Borrill[10†] (Philippa.Borrill@jic.ac.uk), Ricardo H. Ramirez-Gonzalez[10†] (Ricardo.Ramirez-Gonzalez@jic.ac.uk), Rudi Appels[1,36]* (rudi.appels@unimelb.edu.au), Dominique Arnaud[63] (dominiquearnaud.fr@gmail.com), Smahane Chalabi[63] (smahane.chalabi@gmail.com), Boulos Chalhoub[62,63] (boulos.chalhoub@yahoo.com), Frédéric Choulet[7] (frederic.choulet@inra.fr), Aron Cory[11] (aron.cory@usask.ca), Raju Datla[22] (raju.datla@nrc-cnrc.gc.ca), Mark W. Davey[18] (mark.davey@bayer.com), Matthew Hayden[1] (matthew.hayden@ecodev.vic.gov.au), John Jacobs[18] (j.jacobs@bayer.com), Daniel Lang[9] (daniel.lang@helmholtz-muenchen.de), Stephen J. Robinson[52] (steve.robinson@agr.gc.ca), Manuel Spannagl[9] (manuel.spannagl@helmholtz-muenchen.de), Burkhard Steuernagel[10] (burkhard.steuernagel@jic.ac.uk), Josquin Tibbits[1] (josquin.tibbits@ecodev.vic.gov.au), Vijay Tiwari[54] (vktiwari@umd.edu), Fred van Ex[18] (frederic.vanex@bayer.com), and Brande B. H. Wulff[10] (brande.wulff@jic.ac.uk).

Whole Genome Methylome: Curtis J. Pozniak[11‡] (curtis.pozniak@usask.ca), Stephen J. Robinson[52‡] (steve.robinson@agr.gc.ca), Andrew G. Sharpe[43‡] (andrew.sharpe@gifs.ca), and Aron Cory[11] (aron.cory@usask.ca).

Histone Mark Analyses: Moussa Benhamed[15‡] (moussa.benhamed@u-psud.fr), Etienne Paux[7‡] (etienne.paux@inra.fr), Abdelhafid Bendahmane[15] (abdel.bendahmane@u-psud.fr), Lorenzo Concia[15] (lorenzo.concia@u-psud.fr), and David Latrasse[15] (david.latrasse@u-psud.fr).

BAC Chromosome MTP IWGSC-Bayer Whole Genome Profiling(WGP™) Tags: Jane Rogers[6‡] (janerogersh@gmail.com), John Jacobs[18‡] (j.jacobs@bayer.com), Michael Alaux[13] (michael.alaux@inra.fr), Rudi Appels[1,36]* (rudi.appels@unimelb.edu.au), Jan Bartoš[8] (bartos@ueb.cas.cz), Arnaud Bellec[20] (arnaud.bellec@inra.fr), Hélène Berges[20]

(helene.berges@inra.fr), Jaroslav Doležel[8] (dolezel@ueb.cas.cz), Catherine Feuillet[17] (feuillet@bayer.com), Zeev Frenkel[26] (zvfrenkel@gmail.com), Bikram Gill[28] (bsgill@ksu.edu), Abraham Korol[26] (korol@research.haifa.ac.il), Thomas Letellier[13] (thomas.letellier@inra.fr), Odd-Arne Olsen[56] (odd-arne.olsen@nmbu.no), Hana Šimková[8] (simkovah@ueb.cas.cz), Kuldeep Singh[65] (kuldeep35@pau.edu), Miroslav Valárik[8](valarik@ueb.cas.cz), Edwin van der Vossen[64] (edwin.van-der-vossen@keygene.com), Sonia Vautrin[20] (sonia.vautrin@inra.fr), and Song Weining[66] (sweining2002@yahoo.com).

Chromosome LTC Mapping & Physical Mapping Quality Control: Abraham Korol[26‡] (korol@research.haifa.ac.il), Zeev Frenkel[26†] (zvfrenkel@gmail.com), Tzion Fahima[26†] (fahima@research.haifa.ac.il), Vladimir Glikson[29] (lvglikson@gmail.com), Dina Raats[50] (dina.raats@earlham.ac.uk), and Jane Rogers[6] (janerogersh@gmail.com).

RH Mapping: Vijay Tiwari[54‡] (vktiwari@umd.edu), Bikram Gill[28] (bsgill@ksu.edu), Etienne Paux[7] (etienne.paux@inra.fr), and Jesse Poland[28] (jpoland@ksu.edu).

Optical Mapping: Jaroslav Doležel[8‡] (dolezel@ueb.cas.cz), Jarmila Číhalíková[8] (cihalikovaj@seznam.cz), Hana Šimková[8] (simkovah@ueb.cas.cz), Helena Toegelová[8] (toegelova@ueb.cas.cz), and Jan Vrána[8] (vrana@ueb.cas.cz).

Recombination Analyses: Pierre Sourdille[‡7] (pierre.sourdille@inra.fr) and Benoit Darrier[7] (benoit.darrier@inra.fr).

Gene Family Analyses: Rudi Appels[1,36*‡] (rudi.appels@unimelb.edu.au), Manuel Spannagl[9‡] (manuel.spannagl@helmholtz-muenchen.de), Daniel Lang[9†] (daniel.lang@helmholtz-muenchen.de), Iris Fischer[9] (iris.fischer@helmholtz-muenchen.de), Danara Ormanbekova[9,48] (danara.ormanbekova2@unibo.it), and Verena Prade[9] (verena.prade@helmholtz-muenchen.de).

CBF gene family: Delfina Barabaschi[16‡] (delfina.barabaschi@crea.gov.it) and Luigi Cattivelli[16] (luigi.cattivelli@crea.gov.it).

Dehydrin gene family: Pilar Hernandez[33‡] (phernandez@ias.csic.es), Sergio Galvez[27‡] (galvez@uma.es), and Hikmet Budak[14] (hikmet.budak@montana.edu).

NLR gene family: Burkhard Steuernagel[10‡] (burkhard.steuernagel@jic.ac.uk), Jonathan D. G. Jones[35] (jonathan.jones@sainsbury-laboratory.ac.uk), Kamil Witek[35] (kamil.witek@sainsbury-laboratory.ac.uk), Brande B. H. Wulff[10] (brande.wulff@jic.ac.uk), and Guotai Yu[10], (guotai.yu@jic.ac.uk).

PPR gene family: Ian Small[45‡] (ian.small@uwa.edu.au), Joanna Melonek[45†] (joanna.melonek@uwa.edu.au), and Ruonan Zhou[4], (zhou@ipk-gatersleben.de).

Prolamin gene family: Angéla Juhász[36,37‡] (A.Juhasz@murdoch.edu.au), Tatiana Belova[56†] (tatiana.belova@nmbu.no), Rudi Appels[1,36*] (rudi.appels@unimelb.edu.au), and Odd-Arne Olsen[56] (odd-arne.olsen@nmbu.no).

WAK gene family: Kostya Kanyuka[38‡] (kostya.kanyuka@rothamsted.ac.uk), Robert King[42†] (robert.king@rothamsted.ac.uk)

Stem Solidness (Sst1) QTL Team: Kirby Nilsen[11‡] (kirby.nilsen@usask.ca), Sean Walkowiak[11‡] (sean.walkowiak@usask.ca), Curtis J. Pozniak[11‡] (curtis.pozniak@usask.ca), Richard Cuthbert[21] (richard.cuthbert@agr.gc.ca), Raju Datla[22] (raju.datla@nrc-cnrc.gc.ca), Ron Knox[21] (ron.knox@agr.gc.ca), Krysta Wiebe[11] (k.wiebe@usask.ca), and Daoquan Xiang[22] (daoquan.xiang@nrc-cnrc.gc.ca).

Flowering Locus C (FLC) Gene Team: Antje Rohde[72‡] (antje.rohde@bayer.com) and Timothy Golds[18‡] (timothy.golds@bayer.com)

Genome Size Analysis: Jaroslav Doležel[8‡] (dolezel@ueb.cas.cz), Jana Čížková[8] (cizkova@ueb.cas.cz), and Josquin Tibbits[1] (josquin.tibbits@ecodev.vic.gov.au).

MicroRNA and tRNA annotation: Hikmet Budak[14‡] (hikmet.budak@montana.edu), Bala Ani Akpinar[14] (aniakpinar@gmail.com), and Sezgi Biyiklioglu[14] (sezgi.biyiklioglu@montana.edu).

Genetic Maps and Mapping: Gary Muehlbauer[30‡] (muehl003@umn.edu), Jesse Poland[28‡] (jpoland@ksu.edu), Liangliang Gao[28] (lianggao@ksu.edu), Juan Gutierrez-Gonzalez[30] (jgutierr@umn.edu), and Amidou N'Daiye[11](amidou.ndaiye@usask.ca).

BAC libraries and Chromosome Sorting: Jaroslav Doležel[8‡] (dolezel@ueb.cas.cz), Hana Šimková[8†] (simkovah@ueb.cas.cz), Jarmila Číhalíková[8] (cihalikovaj@seznam.cz), Marie Kubaláková[8] (kubalakovam@seznam.cz), Jan Šafář[8] (safar@ueb.cas.cz), and Jan Vrána[8] (vrana@ueb.cas.cz).

BAC Pooling, BAC library Repository, and Access: Hélène Berges[20‡] (helene.berges@inra.fr), Arnaud Bellec[20] (arnaud.bellec@inra.fr), and Sonia Vautrin[20] (sonia.vautrin@inra.fr).

IWGSC Sequence & Data Repository and Access: Michael Alaux[13‡] (michael.alaux@inra.fr), Françoise Alfama[13] (francoise.alfama-depauw@inra.fr), Anne-Françoise Adam-Blondon[13] (anne-francoise.adam-blondon@inra.fr), Raphael Flores[13] (raphael.flores@inra.fr), Claire Guerche[13] (claire.guerche@inra.fr), Thomas Letellier[13] (thomas.letellier@inra.fr), Mikaël Loaec[13] (mikael.loaec@inra.fr), and Hadi Quesneville[13] (hadi.quesneville@inra.fr).

Physical Maps and BAC-based Sequences:

1A BAC Sequencing & Assembly: Curtis J. Pozniak[11‡] (curtis.pozniak@usask.ca), Andrew G. Sharpe[22,43‡] **(**andrew.sharpe@gifs.ca**),** Sean Walkowiak[‡11] (sean.walkowiak@usask.ca), Hikmet Budak[14] (hikmet.budak@montana.edu), Janet Condie[22] (Janet.Condie@nrc-cnrc.gc.ca), Jennifer Ens[11] (jennifer.ens@usask.ca), ChuShin Koh[43] (kevin.koh@gifs.ca), Ron Maclachlan[11] (ron.maclachlan@usask.ca), Yifang Tan[22] (yifang.tan@nrc-cnrc.gc.ca), and Thomas Wicker[41] (wicker@botinst.uzh.ch).

1B BAC Sequencing & Assembly: Frédéric Choulet[7‡] (frederic.choulet@inra.fr), Etienne Paux[7‡] (etienne.paux@inra.fr), Adriana Alberti[61] (aalberti@genoscope.cns.fr), Jean-Marc Aury[61] (jmaury@genoscope.cns.fr), François Balfourier[7] (francois.balfourier@inra.fr), Valérie Barbe[61] (vbarbe@genoscope.cns.fr), Arnaud Couloux[61] (acouloux@genoscope.cns.fr), Corinne Cruaud[61] (cruaud@genoscope.cns.fr), Karine Labadie[61] (klabadie@genoscope.cns.fr), Sophie Mangenot[61] (mangenot@genoscope.cns.fr), and Patrick Wincker[61,68,69] (pwincker@genoscope.cns.fr).

1D, 4D, 6D Physical Mapping: Bikram Gill[28‡] (bsgill@ksu.edu), Gaganpreet Kaur[28] (gaganchahal@gmail.com), Mingcheng Luo[34] (mcluo@ucdavis.edu), and Sunish Sehgal[53] (sunish.sehgal@sdstate.edu).

2AL Physical Mapping: Kuldeep Singh[65‡] (kuldeep35@pau.edu), Parveen Chhuneja[65] (pchhuneja@pau.edu), Om Prakash Gupta[65] (opgupta@pau.edu), Suruchi Jindal[65] (suruchi-coasab@pau.edu), Parampreet Kaur[65] (parampreet.pau@gmail.com), Palvi Malik[65] (palvimalik@pau.edu), Priti Sharma[65] (pritisharma@pau.edu), and Bharat Yadav[65] (bharat_yadav@pau.edu).

2AS Physical Mapping: Nagendra K. Singh[70‡] (nksingh4@gmail.com), Jitendra P. Khurana[71‡] (khuranaj@genomeindia.org), Chanderkant Chaudhary[71] (ckryptone@gmail.com), Paramjit Khurana[71] (param@genomeindia.org), Vinod Kumar[70] (kumar.vinod81@gmail.com), Ajay Mahato[70] (ajaybioinfo@gmail.com), Saloni Mathur[71] (saloni@genomeindia.org), Amitha Sevanthi[70] (amithamithra.nrcpb@gmail.com), Naveen Sharma[71] (naveenlalosharma@gmail.com), and Ram Sewak Tomar[70] (rsstomar@rediffmail.com).

2B, 2D, 4B, 5BL, & 5DL IWGSC-Bayer Whole Genome Profiling(WGP™) Physical Maps: Jane Rogers[6‡] (janerogersh@gmail.com), John Jacobs[18‡] (j.jacobs@bayer.com), Michael Alaux[13] (michael.alaux@inra.fr), Arnaud Bellec[20] (arnaud.bellec@inra.fr), Hélène Berges[20] (helene.berges@inra.fr), Jaroslav Doležel[8] (dolezel@ueb.cas.cz), Catherine Feuillet[17] (feuillet@bayer.com), Zeev Frenkel[26] (zvfrenkel@gmail.com), Bikram Gill[28] (bsgill@ksu.edu), Abraham Korol[26] (korol@research.haifa.ac.il), Edwin van der Vossen[64] (edwin.van-der-vossen@keygene.com), and Sonia Vautrin[20] (sonia.vautrin@inra.fr).

3AL Physical Mapping: Bikram Gill[28‡] (bsgill@ksu.edu), Gaganpreet Kaur[28] (gaganchahal@gmail.com), Mingcheng Luo[34] (mcluo@ucdavis.edu), and Sunish Sehgal[53] (sunish.sehgal@sdstate.edu).

3DS Physical Mapping & BAC Sequencing & Assembly: Jan Bartoš[8‡] (bartos@ueb.cas.cz), Kateřina Holušová[8] (holusovak@ueb.cas.cz), and Ondřej Plíhal[49] (ondrej.plihal@upol.cz).

3DL BAC Sequencing & Assembly: Matthew D. Clark[50,73] (matt.clark@nhm.ac.uk), Darren Heavens[50] (Darren.Heavens@earlham.ac.uk), George Kettleborough[50] (kettleg@gmail.com), and Jon Wright[50] (Jon.Wright@earlham.ac.uk).

4A Physical Mapping, BAC Sequencing, Assembly, & Annotation: Miroslav Valárik[8‡] (valarik@ueb.cas.cz), Michael Abrouk[8,19] (abrouk@ueb.cas.cz), Barbora Balcárková[8]

(balcarkova.bara@seznam.cz), Kateřina Holušová[8] (holusovak@ueb.cas.cz), Yuqin Hu (yqhu@ucdavis.edu), and Mingcheng Luo[34] (mcluo@ucdavis.edu).

5BS BAC Sequencing, & Assembly: Elena Salina[47‡] **(**salina@bionet.nsc.ru**),** Nikolai Ravin[23,51‡] (nravin@biengi.ac.ru), Konstantin Skryabin[23,51‡] (skryabin@biengi.ac.ru), Alexey Beletsky[23] (mortu@yandex.ru), Vitaly Kadnikov[23] (vkadnikov@bk.ru), Andrey Mardanov[23] (mardanov@biengi.ac.ru), Michail Nesterov[47] (mikkanestor@bionet.nsc.ru), Andrey Rakitin[23] (rakitin@biengi.ac.ru), and Ekaterina Sergeeva[47] (sergeeva@bionet.nsc.ru).

6B BAC Sequencing & Assembly: Hirokazu Handa[31‡] (hirokazu@affrc.go.jp), Hiroyuki Kanamori[31] (kanamo@affrc.go.jp), Satoshi Katagiri[31] (skatagiri@affrc.go.jp), Fuminori Kobayashi[31] (kobafumi@affrc.go.jp), Shuhei Nasuda[46] (nasushu@kais.kyoto-u.ac.jp), Tsuyoshi Tanaka[31] (tstanaka@affrc.go.jp), and Jianzhong Wu[31] (jzwu@affrc.go.jp).

7A Physical Mapping & BAC Sequencing: Rudi Appels[1,36*‡] (rudi.appels@unimelb.edu.au), Matthew Hayden[1] (matthew.hayden@ecodev.vic.gov.au), Gabriel Keeble-Gagnère[1] (gabriel.keeble-gagnere@ecodev.vic.gov.au), Philippe Rigault[39] (prigault@gydle.com), and Josquin Tibbits[1] (josquin.tibbits@ecodev.vic.gov.au).

7B Physical Mapping, BAC Sequencing, & Assembly: Odd-Arne Olsen[56‡] (odd-arne.olsen@nmbu.no), Tatiana Belova[56†] (tatiana.belova@nmbu.no), Federica Cattonaro[58] (cattonaro@igatechnology.com), Min Jiumeng[60] (minjm@bgi.com), Karl Kugler[9] (Kg.kugler@gmail.com), Klaus F.X. Mayer[9,44] (k.mayer@helmholtz-muenchen.de), Matthias Pfeifer[9] (matthiaspfeifer@gmx.net), Simen Sandve[57] (simen.sandve@nmbu.no), Xu Xun[59] (xuxun@genomics.cn), and Bujie Zhan[56†] (bujie.zhan@gmail.com).

7DS BAC Sequencing & Assembly: Hana Šimková[8‡] (simkovah@ueb.cas.cz), Michael Abrouk[8,19] (abrouk@ueb.cas.cz), Jacqueline Batley[24] (jacqueline.batley@uwa.edu.au), Philipp E. Bayer[24] (philipp.bayer@uwa.edu.au), David Edwards[24] (Dave.Edwards@uwa.edu.au), Satomi Hayashi[32] (satomi.hayashi@qut.edu.au), Helena Toegelová[8] (toegelova@ueb.cas.cz), Zuzana Tulpová[8] (tulpova@ueb.cas.cz), and Paul Visendi[55] (P.Muhindira@greenwich.ac.uk),

7DL Physical Mapping & BAC Sequencing: Song Weining[66‡] (sweining2002@yahoo.com), Licao Cui[66] (juelianjunjie@foxmail.com), Xianghong Du[66] (xianghongdu@nwsuaf.edu.cn), Kewei Feng[66] (fkwyc@hotmail.com), Xiaojun Nie[66] ours2011@163.com), Wei Tong[66] (tongw@nwsuaf.edu.cn), and Le Wang[66] (lerwang@ucdavis.edu).

Figures: Philippa Borrill[10] (Philippa.Borrill@jic.ac.uk), Heidrun Gundlach[9] (h.gundlach@helmholtz-muenchen.de), Sergio Galvez[27] (galvez@uma.es), Gemy Kaithakottil[50] (Gemy.Kaithakottil@earlham.ac.uk), Daniel Lang[9] (daniel.lang@helmholtz-muenchen.de), Thomas Lux[9] (thomas.lux@helmholtz-muenchen.de), Martin Mascher[4,67] (mascher@ipk-gatersleben.de), Danara Ormanbekova[9,48] (danara.ormanbekova2@unibo.it), Verena Prade[9] (verena.prade@helmholtz-muenchen.de), Ricardo H. Ramirez-Gonzalez[10] (Ricardo.Ramirez-Gonzalez@jic.ac.uk), Manuel Spannagl[9] (manuel.spannagl@helmholtz-muenchen.de), Nils

Stein[4,5]* (stein@ipk-gatersleben.de) Cristobal Uauy[10] (cristobal.uauy@jic.ac.uk), and Luca Venturini[50] (Luca.Venturini@earlham.ac.uk).

Manuscript Writing Team: Nils Stein[4,5]*‡ (stein@ipk-gatersleben.de), Rudi Appels[1,36]*‡ (rudi.appels@unimelb.edu.au), Kellye Eversole[2,3]* (eversole@eversoleassociates.com), Jane Rogers[6†] (janerogersh@gmail.com), Philippa Borrill[10] (Philippa.Borrill@jic.ac.uk), Luigi Cattivelli[16] (luigi.cattivelli@crea.gov.it), Frédéric Choulet[7] (frederic.choulet@inra.fr), Pilar Hernandez[33] (phernandez@ias.csic.es), Kostya Kanyuka[38] (kostya.kanyuka@rothamsted.ac.uk), Daniel Lang[9] (daniel.lang@helmholtz-muenchen.de), Martin Mascher[4,67] (mascher@ipk-gatersleben.de), Kirby Nilsen[11] (kirby.nilsen@usask.ca), Etienne Paux[7] (etienne.paux@inra.fr), Curtis J. Pozniak[11] (curtis.pozniak@usask.ca), Ricardo H. Ramirez-Gonzalez[10] (Ricardo.Ramirez-Gonzalez@jic.ac.uk), Hana Šimková[8] (simkovah@ueb.cas.cz), Ian Small[45](ian.small@uwa.edu.au), Manuel Spannagl[9] (manuel.spannagl@helmholtz-muenchen.de), David Swarbreck[50], (david.swarbreck@earlham.ac.uk), and Cristobal Uauy[10] (cristobal.uauy@jic.ac.uk).

[1]AgriBio, Centre for AgriBioscience, Department of Economic Development, Jobs, Transport and Resources, 5 Ring Rd, La Trobe University, Bundoora, Victoria 3083 Australia.

[2]International Wheat Genome Sequencing Consortium (IWGSC), 5207 Wyoming Road, Bethesda, Maryland, 20816, United States.

[3]Eversole Associates, 5207 Wyoming Road, Bethesda, Maryland, 20816, United States.

[4]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Genebank, Correnssstr. 3, 06466 Stadt Seeland, Germany.

[5]The University of Western Australia (UWA), School of Agriculture and Environment, 35 Stirling Highway, Crawley WA 6009, Australia.

[6]International Wheat Genome Sequencing Consortium (IWGSC), 18 High Street, Little Eversden, Cambridge CB23 1HE, United Kingdom.

[7]GDEC (Genetics, Diversity and Ecophysiology of Cereals), INRA, Université Clermont Auvergne (UCA), 5 chemin de Beaulieu, 63039 Clermont-Ferrand, France.

[8]Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Šlechtitelů 31, CZ-78371, Olomouc, Czech Republic.

[9]Helmholtz Center Munich, Plant Genome and Systems Biology (PGSB), Ingolstaedter Landstr. 1 85764 Neuherberg, Germany.

[10]John Innes Centre, Crop Genetics, Norwich Research Park, Norwich NR4 7UH, United Kingdom.

[11]University of Saskatchewan, Crop Development Centre, Agriculture Building, 51 Campus Drive, Saskatoon SK, S7N 5A8, Canada.

[12]NRGene Ltd., 5 Golda Meir St., Ness Ziona 7403648, Israel.

[13]URGI, INRA, Université Paris-Saclay, 78026 Versailles, France.

[14]Montana State University, Plant Sciences and Plant Pathology, Cereal Genomics Lab, 412 Leon Johnson Hall, Bozeman, MT 59717, USA.

[15]Institute of Plant Sciences - Paris-Saclay, Biology Department, Bâtiment 630, rue de Noetzlin, Plateau du Moulon, CS80004, 91192 - Gif-sur-Yvette Cedex, France.

[16]Council for Agricultural Research and Economics (CREA), Research Centre for Genomics & Bioinformatics, via S. Protaso, 302, I -29017 Fiorenzuola d'Arda, Italy.

[17]Bayer CropScience, Crop Science Division, Research & Development, Innovation Centre, 3500 Paramount Parkway, Morrisville, NC 27560, United States.

[18]Bayer CropScience, Trait Research, Innovation Center, Technologiepark 38, 9052, Gent, Belgium.

[19]King Abdullah University of Science and Technology, Biological and Environmental Science & Engineering Division, Thuwal, 23955-6900, Kingdom of Saudi Arabia.

[20]INRA, CNRGV, Chemin de Borde Rouge CS 52627 31326 Castanet Tolosan cedex, France.

[21]Agriculture and Agri-Food Canada, Swift Current Research and Development Centre, Box 1030, Swift Current, SK S9H 3X2, Canada.

[22]National Research Council Canada, Aquatic and Crop Resource Development, 110 Gymnasium Place, Saskatoon SK S7N 0W9, Canada.

[23]Research Center of Biotechnology of the Russian Academy of Sciences, Institute of Bioengineering, Leninsky Ave. 33, bld 2, Moscow 119071, Russia.

[24]University of Western Australia, School of Biological Sciences and Institute of Agriculture, University of Western Australia, Perth, 6009 Australia.

[25]School of Plant Sciences and Food Security, Tel Aviv University, Ramat Aviv 69978, Israel.

[26]University of Haifa, Institute of Evolution and the Department of Evolutionary and Environmental Biology, 199 Abba-Hushi Avenue,Mount Carmel, Haifa 3498838, Israel.

[27]Universidad de Málaga, Lenguajes y Ciencias de la Computación, Campus de Teatinos, 29071 Málaga, Spain.

[28]Kansas State University, Plant Pathology, Throckmorton Hall, Kansas State University, Manhattan KS, 66506, United States.

[29]Multi-QTL Ltd., University of Haifa, Haifa, Israel.

[30]University of Minnesota, Department of Agronomy and Plant Genetics, 411 Borlaug Hall, St. Paul, MN 55108.

[31]Institute of Crop Science, NARO (former NIAS), 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8518, Japan.

[32]Queensland University of Technology, Earth, Environmental and Biological Sciences, Brisbane, Queensland, Australia.

[33]Instituto de Agricultura Sostenible (IAS-CSIC), Consejo Superior de Investigaciones Científicas, Alameda del Obispo s/n, 14004 Córdoba, Spain.

[34]University of California, Davis, Department of Plant Sciences, One Shield Avenue, Davis, CA 95617, United States.

[35]The Sainsbury Laboratory, Norwich Research Park, NR4 7UH, Norwich, United Kingdom.

[36]Murdoch University, Australia China Centre for Wheat Improvement, School of Veterinary and Life Sciences, 90 South Street, Murdoch WA 6150, Australia.

[37]Agricultural Institute, MTA Centre for Agricultural Research, Applied Genomics Department, 2 Brunszvik Street, Martonvásár H 2462, Hungary.

[38]Rothamsted Research, Biointeractions and Crop Protection, West Common, Harpenden, AL5 2JQ, United Kingdom.

[39]GYDLE, Suite 220, 1135 Grande Allée, Ouest, Suite 220, Québec, QC G1S 1E7, Canada.

[40]Julius Kühn-Institut, Institute for Biosafety in Plant Biotechnology, Erwin-Baur-Str. 27 06484 Quedlinburg, Germany.

[41]University of Zurich, Department of Plant and Microbial Biology, Zollikerstrasse 107, 8008 Zurich, Switzerland.

[42]Rothamsted Research, Computational and Analytical Sciences, West Common, Harpenden, AL5 2JQ, United Kingdom.

[43]University of Saskatchewan, Global Institute for Food Security, 110 Gymnasium Place Saskatoon SK S7N 4J8, Canada.

[44]Technical University of Munich, School of Life Sciences, Weihenstephan, Germany.

[45]The University of Western Australia, School of Molecular Sciences, ARC Centre of Excellence in Plant Energy Biology, 35 Stirling Highway, Crawley WA 6009, Australia.

[46]Kyoto University, Graduate School of Agriculture, Kitashirakawaoiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan.

[47]The Federal Research Center Institute of Cytology and Genetics, SB RAS, pr. Lavrentyeva 10, Novosibirsk 630090, Russia.

[48]University of Bologna, Department of Agricultural Sciences,Viale Fanin, 44 40127 Bologna, Italy.

[49]Palacký University, Centre of the Region Haná for Biotechnological and Agricultural Research, Department of Molecular Biology, Šlechtitelů 27, CZ-78371 Olomouc, Czech Republic.

[50]Earlham Institute, Core Bioinformatics, Norwich, NR4 7UZ, United Kingdom.

[51]Moscow State University, Faculty of Biology, Leninskie Gory, 1, Moscow, 119991, Russia.

[52]Agriculture and Agri-Food Canada, Saskatoon Research and Development Centre, 107 Science Place, Saskatoon, SK, S7N 0X2, Canada.

[53]South Dakota State University, Agronomy Horticulture and Plant Science, 2108 Jackrabbit Dr, Brookings, SD 57006, United States.

[54]University of Maryland, Plant Science and Landscape Architecture, 4291 Fieldhouse Road, 2102 Plant Sciences Building College Park, MD 20742, United States.

[55]University of Greenwich, Natural Resources Institute, Central Avenue, Chatham, Kent ME4 4TB, United Kingdom.

[56]Norwegian University of Life Sciences, Faculty of Bioscience, Department of Plant Science, Arboretveien 6, 1433 Ås, Norway.

[57]Norwegian University of Life Sciences, Faculty of Bioscience, Department of Animal and Aquacultural Sciences, Arboretveien 6, 1433 Ås, Norway.

[58]Instituto di Genomica Applicata, Via J. Linussio 51, Udine, 33100, Italy.

[59]BGI-Shenzhen, BGI Genomics, Yantian, Shenzhen, Guangdong, China.

[60]BGI-Shenzhen, BGI Genomics, Building No.7, BGI Park, No.21 Hongan 3rd Street, Yantian District, Shenzhen, China.

[61]CEA - Institut de Biologie François-Jacob, Genoscope, 2 Rue Gaston Cremieux 91057 Evry Cedex, France.

[62]Monsanto SAS, 28000 Boissay, France.

[63] Institut National de la Recherche Agronomique (INRA), 2 rue Gaston Crémieux, 9057 Evry, France.

[64]Keygene, N.V., Agro business Park 90, 6708 PW Wageningen, The Netherlands.

[65]Punjab Agricultural University, Ludhiana, School of Agricultural Biotechnology, ICAR-National Bureau of Plant Genetic Resources, Dev Prakash Shastri Marg, New Delhi 110012, India.

[66]Northwest A&F University, State Key Laboratory of Crop Stress Biology in Arid Areas, College of Agronomy, Northwest A&F University, Yangling 712101, Shaanxi, China.

[67]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany.

[68]CNRS, UMR 8030, CP5706, Evry, France.

[69]Université d'Evry, UMR 8030, CP5706, Evry, France.

[70]ICAR-National Research Centre on Plant Biotechnology, LBS Building, Pusa Campus, New Delhi 110012, India.

[71]University of Delhi South Campus, Interdisciplinary Center for Plant Genomics & Department of Plant Molecular Biology, Benito Juarez Road, New Delhi-110021, India.

[72]Bayer CropScience, Breeding & Trait Development, Technologiepark 38, 9052, Gent, Belgium.

[73]Department of Lifesciences, Natural History Museum, Cromwell Road, London SW7 5BD, U.K.

Competing interests: Authors declare no competing interests.

Data and materials availability: The IWGSC RefSeq v1.0 assembly and annotation data, physical maps for all chromosomes/chromosome arms, as well as all data related to this study are available in the IWGSC Data Repository hosted at URGI: https://wheat-urgi.versailles.inra.fr/Seq-Repository. The BAC libraries for all chromosomes/chromosomes arms are available at the CNRGV-INRA: https://cnrgv.toulouse.inra.fr/en/Library/Wheat. Details

on gene family expansion and contraction in the genome of bread wheat cv. Chinese Spring are provided in databaseS6 at http://dx.doi.org/10.5447/IPK/2018/5. The raw sequencing data used for *de novo* whole genome assembly is available from the Sequence Read Archive under accession number SRP114784. RNAseq data is available at SRA under accession IDs PRJEB25639, PRJEB23056, PRJNA436817, PRJEB25640, SRP133837, PRJEB25593. Hi-C sequence data are available under accession number PRJEB25248. ChIP seq data are available under SRA study PRJNA420988 (SRP1262229). CS bisulfite sequencing data is available under project ID SRP133674 (SRR6792673-SRR6792689. Organellar DNA sequences were deposited at NCBI Genbank (MH051715, MH051716). Further details on data accessibility are outlined in the supplementary Materials and Methods.

Supplementary Materials:

Materials and Methods

Figures S1-S59

Tables S1-S43

External Databases S1-S6

References (54-184)

Figure captions

Fig. 1. Structural, functional, and conserved synteny landscape of the 21 wheat chromosomes. (A) Circular diagram visualizing genomic features of wheat. The tracks towards the center of the circle display: a - chromosome name and size (100 Mb tick size, light grey bar = short arm, dark grey= long arm of the chromosome); b - dimension of chromosomal segments R1, R2a, C, R2b, R3 ((*17*)Table S29); c - Kmer 20 frequencies distribution; d - LTR-retrotransposons density; e - pseudogenes density (0 to 130 genes per Mb); f - density of high confidence gene models (HC; 0 to 32 genes per Mb); g - density of recombination rate; h- SNP density. Connecting lines in the center of the diagram highlight homeologous relationships of chromosomes (blue lines) and translocated regions (green lines). (B) Distribution of PFAM domain PF08284 ‘retroviral aspartyl protease’ signatures across the different wheat chromosomes. (C) Positioning of the centromere in the 2D pseudomolecule. Upper panel: Density of CENH3 ChIP-seq data along wheat

chromosome. Lower panel: Distribution and proportion of the total pseudomolecule sequence composed of TE of the Cereba/Quinta families. The bar below the lower panel indicates pseudomolecule scaffolds assigned to the short (black) or long (blue) arm based on CSS data (*5*) mapping. (D) Dot pot visualization of collinearity between homeologous chromosomes 3A and 3B in relation to distribution of gene density and recombination frequency (left and lower panel boxes: blue and purple lines, respectively). Chromosomal zones R1, R2a, C, R2b, R3 colored as per in Fig. 1A.

Fig. 2. Evaluation of automated gene annotation. (A) Selected gene prediction statistics of IWGSC RefSeq annotation version 1.1 including number and sub-genome distribution of high confidence (HC) and low confidence (LC) genes as well as pseudogenes. (B) BUSCO v3 gene model evaluation comparing IWGSC RefSeq annotation v1.1 to earlier published bread wheat whole genome annotations as well as to annotations of related grass reference genome sequences. BUSCO provides a measure for the recall of highly conserved gene models.

Fig. 3. Wheat atlas of transcription. (A) Schematic illustration of a mature wheat plant and high-level tissue definitions ‘roots’, ‘leaves’, ‘spike’ and ‘grain’ used in the further analysis. (B) Principal component analysis plots for similarity of overall transcription with samples coloured according to their high-level tissue of origin (as introduced in A). The color key for tissue is shown at the bottom of the figure under panel C. (C) Chromosomal distribution of the average expression breadth [number of tissues in which genes are expressed (total number of tissues, n=32)]. The average (dark orange line) is calculated based on a scaled position of each gene within the corresponding genomic compartment (blue, aqua and white background) across the 21 chromosomes (orange lines). (D) Heatmap illustrating the expression of a representative gene

(eigengene) for the 38 co-expression modules defined by WGCNA. Modules are represented as columns, with the dendrogram illustrating eigengene relatedness. Each row represents one sample; colored bars to the left indicate the high-level tissue of origin; the color key is shown at the bottom of the figure under panel C. DESeq2 normalised expression levels are shown. Modules 1 and 5 (pale green boxes) were most correlated with high-level 'leaf tissue' whereas modules 8 and 11 (dark green boxes) were most correlated with 'spike'. (E) Bar plot of module assignment (same, near or distant) of homeologous triads and duplets in WGCNA network. (F) Simplified flowering pathway in polyploid wheat. Genes are coloured according to their assignment to 'leaf' (pale green) or 'spike' (dark green) correlated modules. (G) Excerpt from phylogenetic tree for MADS transcription factors including known Arabidopsis flowering regulators *SEP1*, *SEP2* and *SEP4* (black) (for the full phylogenetic tree see Fig. S38). Green branches represent wheat orthologs of modules 8 and 11, whereas purple branches are wheat orthologs assigned to other modules (0 and 2). Grey branches indicate non-wheat genes.

Fig. 4. Gene families of wheat. (A) Heatmap of expanded and contracted gene families. Columns correspond to the individual gene families. Rows in the upper panel illustrate the sets of gene family expansions (++; red) and contractions (–; blue) found for the wheat A lineage (*T. urartu* and A sub-genome), the D lineage (*A. tauschii* and D sub-genome), the A, B or D sub-genomes or bread wheat (expanded/contracted in all sub-genomes). In the latter four categories, expansions/contractions do not imply bread-wheat specific gene copy number variations. Similar dynamics might have remained unobserved in *T. urartu* or *A. tauschii* due to the inherent limitations of the used draft genome assemblies (*51*, *52*). Rows in the lower panel heatmap (color scheme on z-score scale) indicate the fold expansion and contraction of gene families for the taxa

/ species included in the analysis [*Oryza sativa* (Osat), *Sorghum bicolor* (Sbic), *Zea mays* (Zmay), *Brachypodium distachyon* (Bdis), *Hordeum vulgare* (Hvul1/2), *Secale cereale* (Scer), *Aegilops tauschii* (Aetau), *Triticum urartu* (Tura), wheat A (TraesA), B (TraesB) and D (TraesD) sub-genomes]. (B) All enriched Plant Trait Ontology (TO) terms for the gene families depicted in (A). Over-represented TO terms were found for expanded families in bread wheat (all sub-genomes; red), the B sub-genome (green) and the A lineage (*T. urartu* and A sub-genome; blue) only, respectively. The x-axis represents the percentage of genes annotated with the respective TO term that were contained in the gene set in question. The size of the bubbles corresponds to the p-value (-log10) significance of expansion. (C) Genomic distribution of gene families associated with adaptation to biotic (light/dark blue) or abiotic stress (light/dark pink), RNA metabolism in organelles and male fertility (orange) or end-use quality (light/medium/dark green). Known positions of agronomically important genes / loci are indicated by red arrows / arrowheads to the left of the chromosome bars. Recombination rates are displayed as heat maps in the chromosome bars (light green = 7.2 cM/Mb to black = 0 cM/Mb).

Fig. 5. IWGSC RefSeq v1.0 guided dissection of *SSt1* and *TaAGL33*. (A) The Lillian/Vesper population genetic map was anchored to IWGSC RefSeq v1.0 (left) and differentially expressed genes were identified between solid and hollow-stemmed lines of hexaploid- (bread) and tetraploid (durum) wheat (right). (B) Cross-sectioned stems of ‘Lillian’ (solid) and ‘Vesper’ (hollow) are shown as a phenotypic reference (top). Increased copy number of *TraesCS3B01G608800* (annotated as a DOF transcription factor) is associated with stem phenotypic variation (bottom). (C) A high-throughput SNP marker tightly linked to *TraesCS3B01G608800* reliably discriminates solid from hollow-stemmed wheat lines. (D)

Schematic of the three TaAGL33 proteins, showing the typical MADS, I, K and C domains. Triangles indicate the position of the 5 introns that occur in all three homeologs. Bars indicate the position of sgRNAs designed for exons 2 and 3. Three T-DNA vectors each containing the *bar* selectable marker gene, CRISPR nuclease and one of three sgRNA sequences were used for *Agrobacterium*-mediated wheat transformation, essentially as described earlier (*53*). Transgenic plants were obtained with edits at the targeted positions in all TaGL33 homeologs. The putatively resulting protein sequence is displayed starting close to the edits with wild-type amino acids in black font and amino acids resulting from the induced frame shifts in red font. * indicates premature termination codons. (E) Mean days to flowering (after 8 weeks of vernalization) for progeny of four homozygous edited plants (light grey bars) and the respective homozygous wild-type segregants (dark grey bars). Numbers in brackets refer to the number of edited and wild-type plants examined, respectively. Error bars display SEM. Growth conditions were as described in (*49*).

Table 1. Assembly statistics of IWGSC Refseq v1.0.

Assembly size	14.5 Gb
Number of scaffolds	138,665
Size of assembly in scaffolds >= 100Kb	14.2 Gb
Number of scaffolds >= 100Kb	4,443
N50 contig length	51.8 Kb
Contig L50	81,427
N90 contig length	11.7 Kb
Contig L90	294,934
Largest contig	580.5 Kb
Ns in contigs	0
N50 scaffold length	7.0 Mb
Scaffold L50	571
N90 scaffold length	1.2 Mb
Scaffold L90	2,390
Largest scaffold	45.8 Mb
Ns in scaffolds	261.9 Mb
Gaps filled with BAC sequences	183 (1.7 Mb)
Average size of inserted BAC sequence	9.5 Kb
N50 super-scaffold length	22.8 Mb
Super-scaffold L50	166
N90 super-scaffold length	4.1 Mb
Super-scaffold L90	718
Largest super-scaffold	165.9 Mb
Sequence assigned to chromosomes	14.1 Gb (96.8%)
Sequence >=100Kb assigned to chromosomes	14.1 Gb (99.1%)
Number of super-scaffolds on chromosomes	1,601
Number of oriented super-scaffolds	1,243
Length of oriented sequence	13.8 Gb (95%)
Length of oriented sequence >= 100Kb	13.8 Gb (97.3%)
Smallest number of super-scaffolds per sub-genome chromosome	35 (7A) / 68 (2B) / 36 (1D)
Highest number of super-scaffolds per sub-genome chromosome	111 (4A) / 176 (3B) / 90 (3D)
Average number of super-scaffolds per chromosome	76

Table 2. Relative proportions of the major elements of the wheat genome. Proportions of TEs are given as the percentage of sequences assigned to each superfamily relative to genome size.

			AA	BB	DD	AABBDD
Assembled sequence assigned to chromosomes (Gb)			4.935	5.180	3.951	14.066
Size of TE-related sequences (Gb)			4.240	4.388	3.285	11.913
%TEs			85.9%	84.7%	83.1%	84.7%
Class 1	LTR-retrotransposons					
		Gypsy (RLG)	50.8%	46.8%	41.4%	46.7%
		Copia (RLC)	17.4%	16.2%	16.3%	16.7%
		Unclassified LTR-RT (RLX)	2.6%	3.5%	3.7%	3.2%
	Non-LTR-retrotransposons					
		LINE (RIX)	0.81%	0.96%	0.93%	0.90%
		SINE (SIX)	0.01%	0.01%	0.01%	0.01%
Class 2	DNA transposons					
		CACTA (DTC)	12.8%	15.5%	19.0%	15.5%
		Mutator (DTM)	0.30%	0.38%	0.48%	0.38%
		Unclassified with TIRs	0.21%	0.20%	0.22%	0.21%
		Harbinger (DTH)	0.15%	0.16%	0.18%	0.16%
		Mariner (DTT)	0.14%	0.16%	0.17%	0.16%
		Unclassified class#2	0.05%	0.08%	0.05%	0.06%
		hAT (DTA)	0.01%	0.01%	0.01%	0.01%
	Helitrons	(DHH)	0.0046%	0.0044%	0.0036%	0.0042%
Unclassified repeats			0.55%	0.85%	0.63%	0.68%
Coding DNA			0.89%	0.89%	1.11%	0.95%
Un-annotated DNA			13.2%	14.4%	15.7%	14.4%
(pre)-miRNAs			0.039%	0.057%	0.046%	0.047%
tRNAs			0.0056%	0.0050%	0.0068%	0.0057%

Table 3. Groups of homeologous genes in wheat. Homeologous genes are "sub-genome orthologs" and were inferred by species tree reconciliation in the respective gene family. Numbers include both HC and LC genes filtered for TEs ("filtered gene set"). Conserved sub-genome-specific (orphan) genes are found only in one sub-genome but have homologs in other plant genomes used in this study. This includes orphan outparalogs resulting from ancestral duplication events and conserved only in one of the sub-genomes. Non-conserved orphans are either singletons or duplicated in the respective sub-genome, but do neither have obvious homologs in the other sub-genomes or the other plant genomes studied. Microsynteny is defined as the conservation and collinearity of local gene ordering between orthologous chromosomal regions. Macrosynteny is defined as the conservation of chromosomal location and identity of genetic markers like homeologs, but may include the occurrence of local inversions, insertions or deletions. Additional data are presented in Table S24.

homeologous group (A:B:D)	# in wheat genome	% of groups	# genes in A	# genes in B	# genes in D	# total genes
1:1:1	21,603	55.1%	21,603	21,603	21,603	64,809
1:1:N	644	1.6%	644	644	1,482	2,770
1:N:1	998	2.5%	998	2,396	998	4,392
N:1:1	761	1.9%	1,752	761	761	3,274
1:1:0	3,708	9.5%	3,708	3,708	0	7,416
1:0:1	4,057	10.3%	4,057	0	4,057	8,114
0:1:1	4,197	10.7%	0	4,197	4,197	8,394
other ratios	3,270	8.3%	4,999	5,371	4,114	14,484
1:1:1 in microsynteny	18,595	47.4%	18,595	18,595	18,595	55,785
total in microsynteny	30,339	77.3%	27,240	27,063	28,005	82,308
1:1:1 in macrosynteny	19,701	50.2%	19,701	19,701	19,701	59,103
total in macrosynteny	32,591	83.1%	29,064	30,615	30,553	90,232
total in homeologous groups	**39,238**	**100.0%**	**37,761**	**38,680**	**37,212**	**113,653**

conserved sub-genome orphans	12,412	12,987	10,844	36,243
non-conserved sub-genome singletons	10,084	12,185	8,679	30,948
non-conserved sub-genome duplicated orphans	71	83	38	192
total (filtered)	60,328	63,935	56,773	181,036

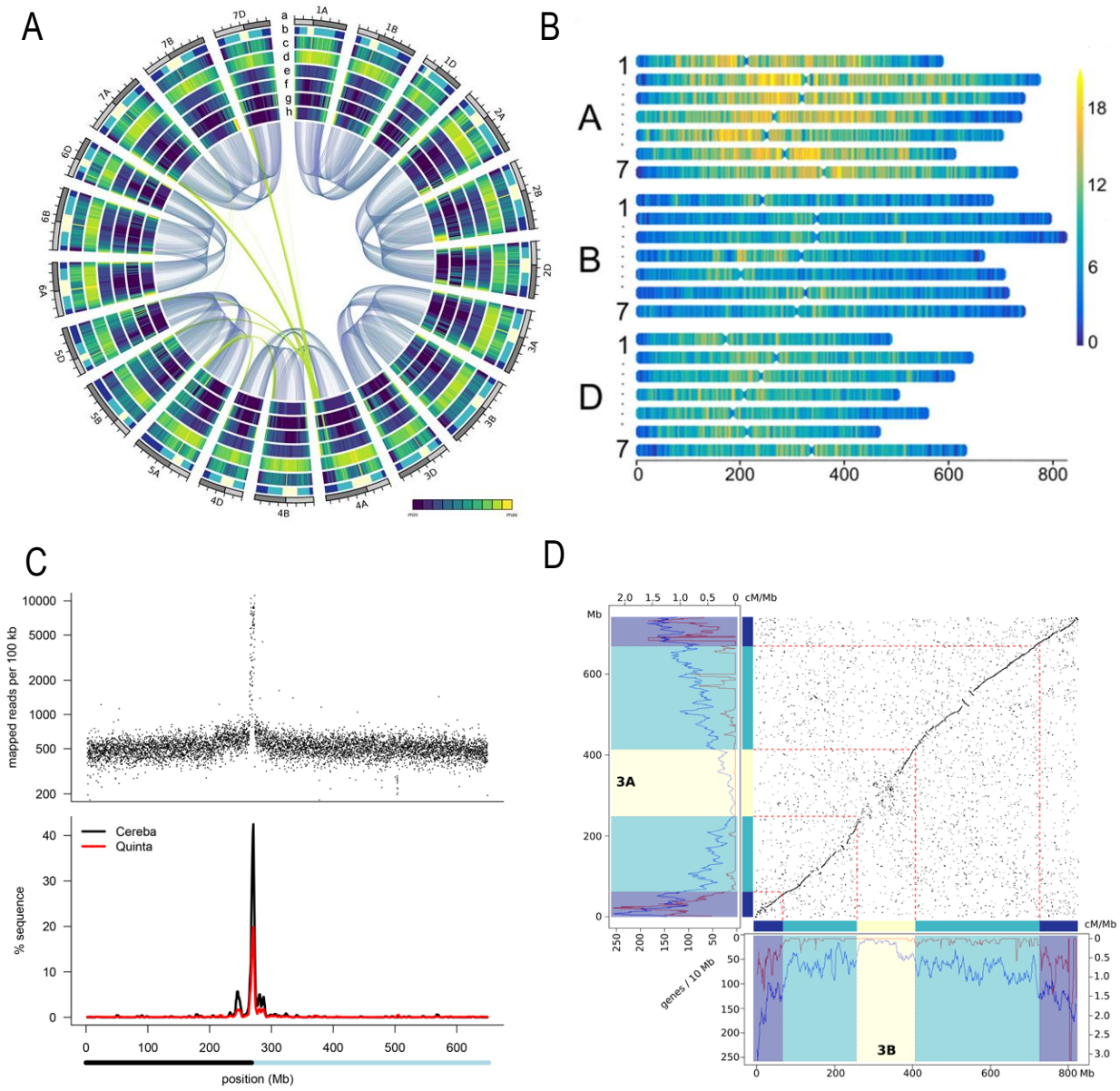

Fig. 1. Structural, functional, and conserved synteny landscape of the 21 wheat chromosomes.

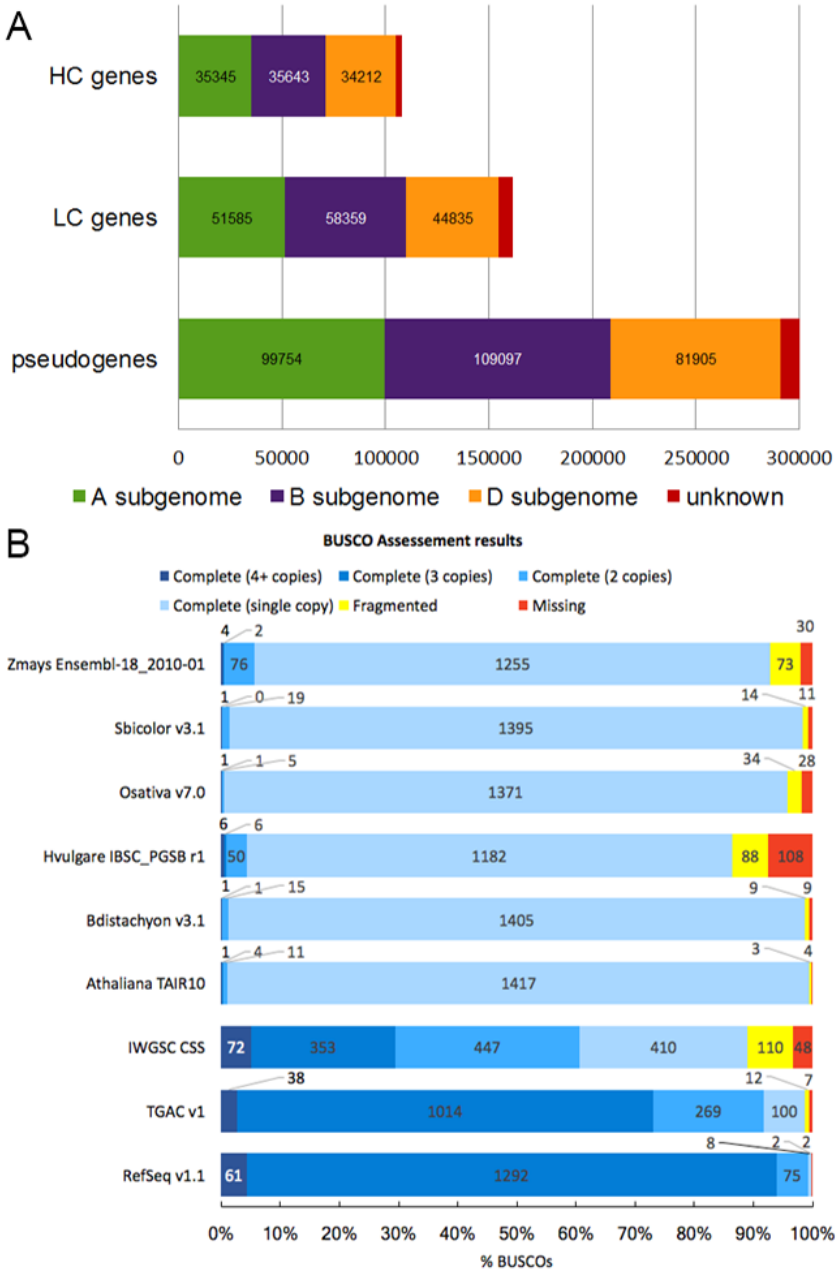

Fig. 2. Evaluation of automated gene annotation.

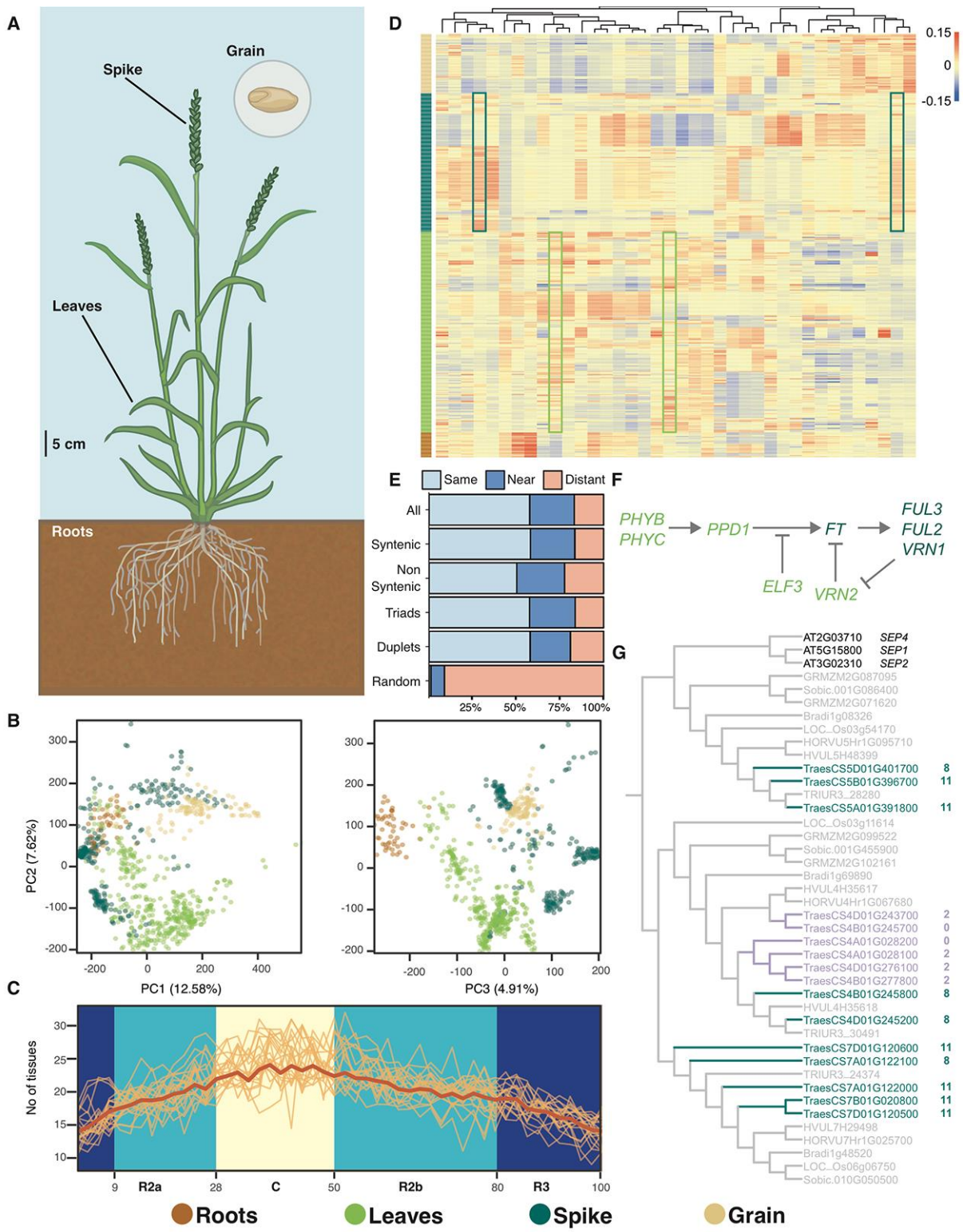

Fig. 3. Wheat atlas of transcription.

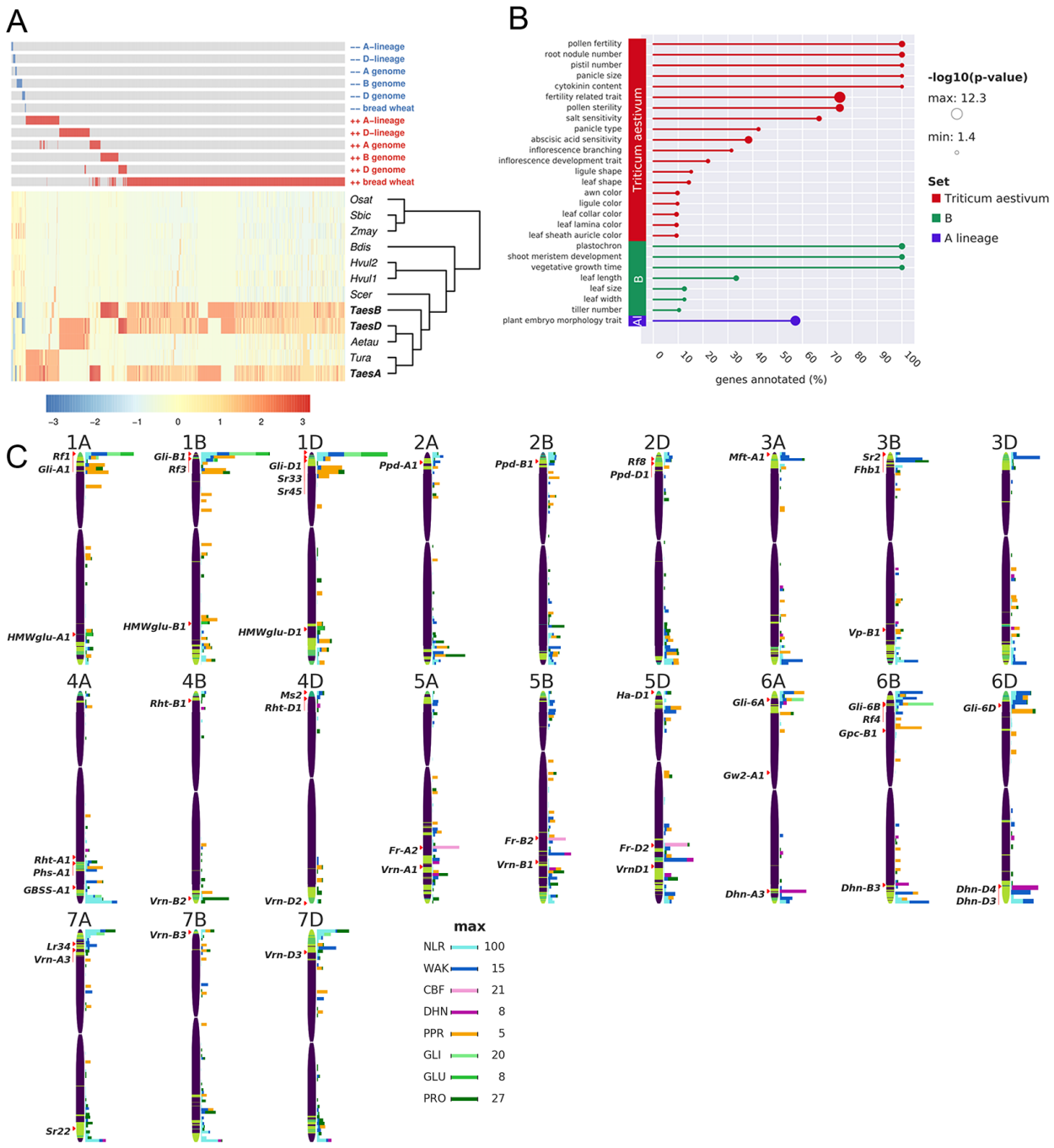

Fig. 4. Analysis of gene families of wheat.

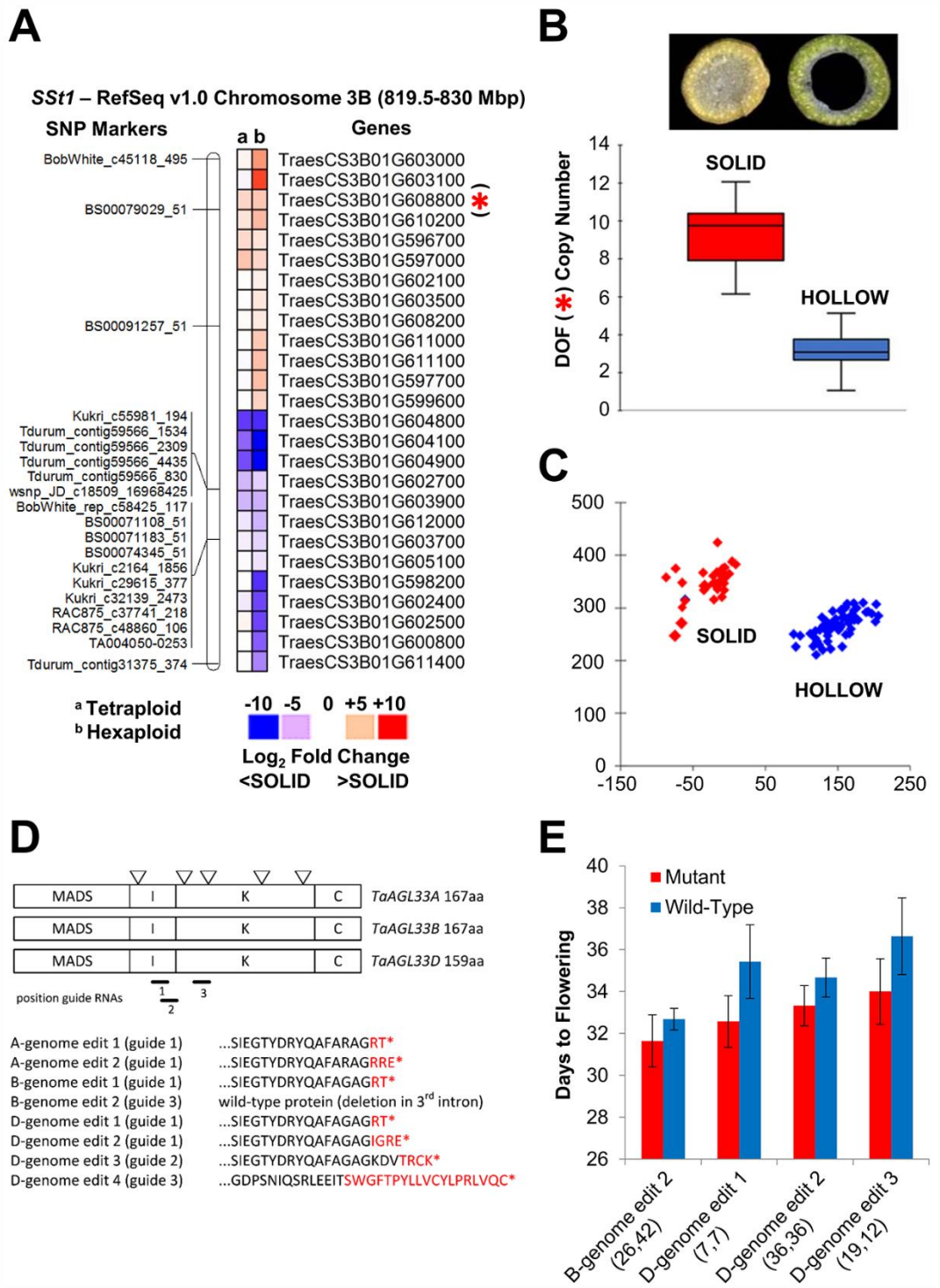

Fig. 5. IWGSC RefSeq v1.0 guided dissection of *SSt1* and *TaAGL33*.