RESEARCH ARTICLE

WHEAT GENOME

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

International Wheat Genome Sequencing Consortium (IWGSC)*

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 noncoding elements across the A, B, and D subgenomes. 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 coexpression networks by providing 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 subgenome 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.

Wheat (*Triticum aestivum* L.), the most widely cultivated crop on Earth, contributes about a fifth of the total calories consumed by humans and provides more protein than any other food source ([1], [2]). 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 ([3]). 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 ([4]). Wheat has lagged behind other species, primarily owing to the challenges of assembling a large (haploid genome, 1C = 16 Gb) ([5]), hexaploid, and complex genome that contains more than 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 ([6]). 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 ([7–9]) 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 ([10–12]), which allowed the integration of genetic and genomic resources. The completeness and accuracy of IWGSC RefSeq v1.0 provide insights into global genome composition and enable the construction of complex gene coexpression 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 de novo whole-genome assembly (WGA), built from Illumina short-read sequences using NRGene deNovoMagic2 (Fig. 1A, Table 1, and tables S1 and S2), with additional layers of genetic, physical, and sequence data (tables S3 to S8 and figs. S1 and 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) of the sequences assigned and ordered along the 21 chromosomes and almost all of the assigned sequence scaffolds oriented relative to each other (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 were assessed through alignments with radiation hybrid maps for the A, B, and D subgenomes [average Spearman's correlation coefficient (r) of 0.98], the genetic positions of 7832 and 4745 genotyping-by-sequencing derived genetic markers in 88 double haploid and 993 recombinant inbred lines (Spearman's r of 0.986 and 0.987, respectively), and 1.24 million pairs of neighbor insertion site–based polymorphism (ISBP) markers ([13]), of which 97% were collinear and mapped in a similar size range (difference of <2 kb) between the de novo WGA and the available bacterial artificial chromosome (BAC)–based sequence assemblies. Finally, IWGSC RefSeq v1.0 was assessed with independent data derived from coding and noncoding sequences, revealing that 99 and 98% of the previously known coding exons ([6]) and transposable element (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 to 15.8 Gb ([14]) 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 ([6–9]) is the long-range organization, with 90% of the genome represented in superscaffolds larger than 4.1 Mb and with each chromosome represented, on average, by only 76 superscaffolds (Table 1). The largest superscaffold spans 166 Mb, which is half the size of the rice (*Oryza sativa* L.) genome and is larger than the *Arabidopsis thaliana* L. genome ([15], [16]). Moreover, the 21 pseudomolecules position molecular markers for wheat research and breeding [504 single-stranded repeats (SSRs), 3025 diversity array technologies (DArTs), 6689 expressed sequence tags (ESTs), 205,807 single-nucleotide polymorphisms (SNPs), and 4,512,979 ISBPs] (table S9), thus providing a direct link between the genome sequence and genetic loci and 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 subgenomes. Accounting for 85% of the genome, with a relatively equal distribution across the three subgenomes (Table 2), 3,968,974 copies of TEs belonging to 505 families were annotated. Many (112,744) full-length long terminal repeat (LTR)–retrotransposons were identified that have been difficult to define from short-read sequence assemblies (fig. S3). Although the TE content has been extensively rearranged

*All authors with their affiliations are listed at the end of this paper.

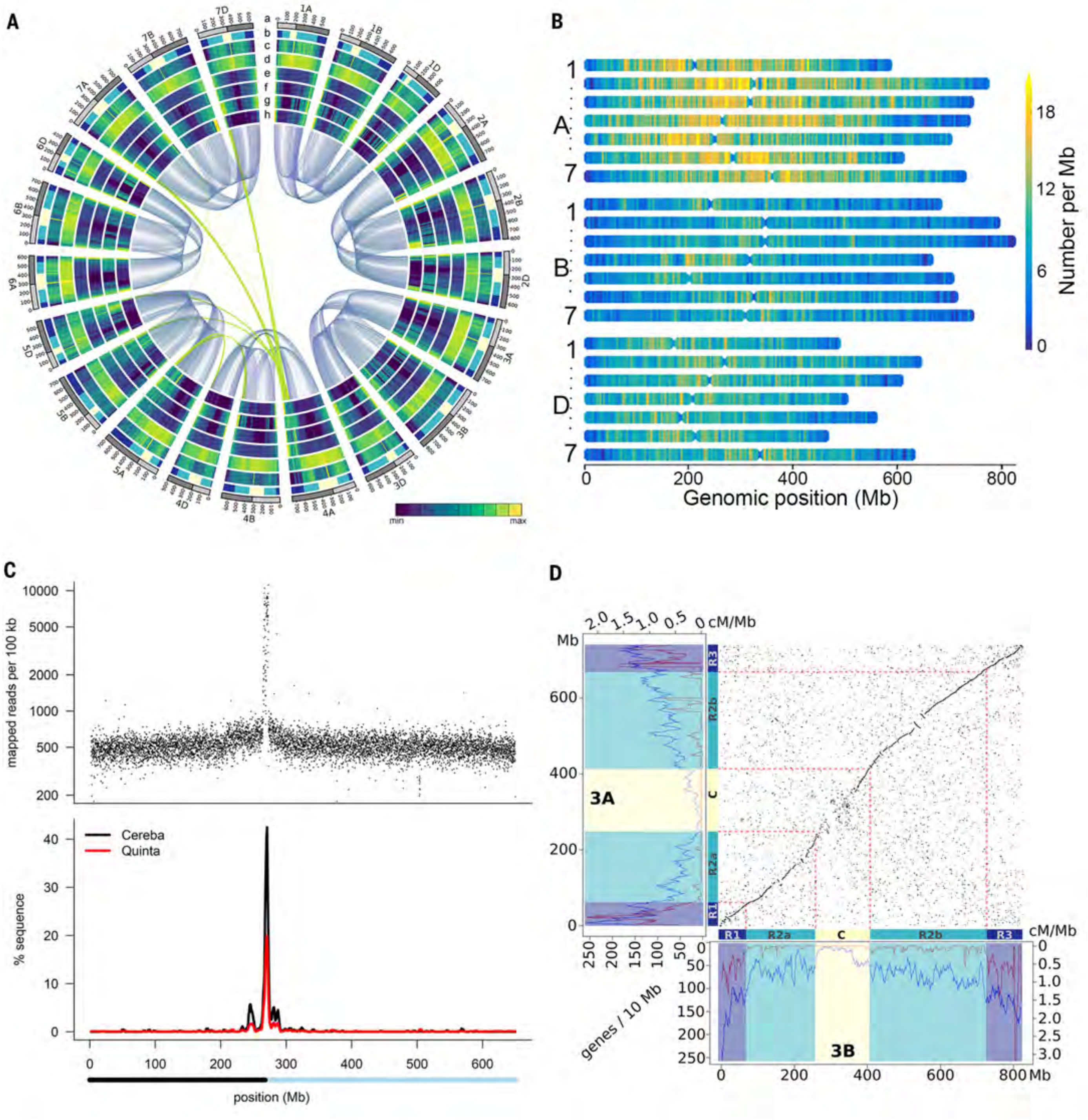

Fig. 1. Structural, functional, and conserved synteny landscape of the 21 wheat chromosomes. (**A**) Circular diagram showing genomic features of wheat. The tracks toward the center of the circle display (a) chromosome name and size (100-Mb tick size; light gray bar indicates the short arm and dark gray indicates the long arm of the chromosome); (b) dimension of chromosomal segments R1, R2a, C, R2b, and R3 [(18) and table S29]; (c) *K*-mer 20-frequencies distribution; (d) LTR-retrotransposons density; (e) pseudogenes density (0 to 130 genes per Mb); (f) density of HC gene models (0 to 32 genes per Mb); (g) density of recombination rate; and (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. Top panel shows density of CENH3 ChIP-seq data along the wheat chromosome. Bottom panel shows distribution and proportion of the total pseudomolecule sequence composed of TEs of the Cereba and Quinta families. The bar below the bottom panel indicates pseudomolecule scaffolds assigned to the short (black) or long (blue) arm on the basis of CSS data (6) mapping. (**D**) Dot-plot visualization of collinearity between homeologous chromosomes 3A and 3B in relation to distribution of gene density and recombination frequency (left and bottom panel boxes: blue and purple lines, respectively). Chromosomal zones R1, R2a, C, R2b, and R3 are colored as in (A). cM, centimorgan.

through rounds of deletions and amplifications since the divergence of the A, B, and D subgenomes 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 1 Mb contributed similar proportions (less than a twofold difference between subgenomes), and only 11 families, accounting for 2% of total TEs, showed a higher than threefold difference between two subgenomes (*17*). TE abundance accounts, in part, for the size differences between subgenomes—for example, 64% of the 1.2-Gb size difference between the B and D subgenomes can be attributed to lower gypsy retrotransposon content. Low-copy DNA content (primarily unclassified sequences) also varied between subgenomes, accounting, for example, for 97 Mb of the 245-Mb size difference between A and B subgenomes (fig. S4). As reported (*18*), 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 the A, B, and D subgenomes at the subfamily (variants) level, as evidenced by subgenome-specific overrepresentation of individual transposon domain signatures (Fig. 1B). See (*17*) 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 noncoding RNAs. We identified eight new microRNA families (fig. S5 and table S10) and the entire complement of tRNAs (which showed an excess of lysine tRNAs, fig. S6). Around 8000 nuclear-inserted plastid DNA segments and 11,000 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 RefSeq v1.0 raw read data (*14*).

Precise positions for the centromeres were defined by integrating Hi-C, CSS (*6*), and published chromatin immunoprecipitation sequencing (ChIP-seq) data for CENH3, a centromere-specific histone H3 variant (*19*). Clear ChIP-seq peaks were evident in all chromosomes and coincided with the centromere-specific repeat families (Fig. 1C, fig. S7, and 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), compared with an average centromere size of ~1.8 Mb in maize (*20*, *21*) and 0.4 to 0.8 Mb in rice (*22*).

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

Assembly characteristics	Values
Assembly size	14.5 Gb
Number of scaffolds	138,665
Size of assembly in scaffolds ≥ 100 kb	14.2 Gb
Number of scaffolds ≥ 100 kb	4,443
N50 contig length	51.8 kb
Contig L50 number	81,427
N90 contig length	11.7 kb
Contig L90 number	294,934
Largest contig	580.5 kb
Ns in contigs	0
N50 scaffold length	7.0 Mb
Scaffold L50 number	571
N90 scaffold length	1.2 Mb
Scaffold L90 number	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 superscaffold length	22.8 Mb
Superscaffold L50 number	166
N90 superscaffold length	4.1 Mb
Superscaffold L90 number	718
Largest superscaffold	165.9 Mb
Sequence assigned to chromosomes	14.1 Gb (96.8%)
Sequence ≥ 100 kb assigned to chromosomes	14.1 Gb (99.1%)
Number of superscaffolds on chromosomes	1,601
Number of oriented superscaffolds	1,243
Length of oriented sequence	13.8 Gb (95%)
Length of oriented sequence ≥ 100 kb	13.8 Gb (97.3%)
Smallest number of superscaffolds per subgenome chromosome	35 (7A), 68 (2B), 36 (1D)
Largest number of superscaffolds per subgenome chromosome	111 (4A), 176 (3B), 90 (3D)
Average number of superscaffolds per chromosome	76

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 and tables S12 to S17). In total, 107,891 high-confidence (HC) protein-coding loci were identified, with relatively equal distribution across the A, B, and D subgenomes (35,345, 35,643, and 34,212, respectively; Figs. 1D and 2A, fig. S10, and 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 and S20), and evidence for transcription was found for 85% (94,114) of the HC genes versus 49% of the LC genes (*23*). Within the pseudogene category, 25,419 (8%) of 303,818 candidates matched LC gene models. The D subgenome contained significantly fewer pseudogenes than the A and B subgenomes (81,905 versus 99,754 and 109,097, respectively; χ^2 test, $P < 2.2 \times 10^{-16}$) (tables S21 and S22 and fig. S10). In ChrUn, 2691 HC and 675 LC gene models were identified.

The quality of the RefSeq Annotation v1.1 gene set was benchmarked against BUSCO v3 (*24*), representing 1440 Embryophyta near-universal single-copy orthologs and published annotated wheat gene sets (Fig. 2B and fig. S11). Of the BUSCO v3 genes, 99% (1436) were represented in at least one complete copy in RefSeq Annotation v1.1 and 90% (1292) in three complete copies, an improvement over the 25% (353) and 70% (1014) of BUSCO v3 genes that were identified in the IWGSC (*6*) and TGACv1 (*8*) 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 ambigous bases (Ns), in contrast to 37% and only 5% of the HC and LC genes in the 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 subgenomes and orthologs in other plant genomes (table S23 and figs. S13 to S15). Analysis of a subset of 181,036 genes ["filtered gene set," (*14*) and Table 3] comprising 103,757 HC and 77,279 LC genes identified 39,238 homeologous groups—that is, clades of A, B, and D subgenome orthologs deduced from gene trees—containing a total of 113,653 genes (63% of the filtered set). Gene losses or 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 subgenomes: 63% (A), 61% (B), and 66% (D). The slightly higher percentage of homeologs in the D subgenome, together with the lower number of pseudogenes (table S22), is consistent with its more recent hybridization with the AABB tetraploid genome progenitor. Although most of the genes are present in homeologous groups, only 18,595 (47%) of the groups contained triads with a single gene copy per subgenome (an A:B:D configuration of 1:1:1). Of the groups of homeologous genes, 5673 (15%) exhibited at least one subgenome inparalog, that is, a gene copy resulting from a tandem or a segmental trans duplication (1:1:N A:B:D configuration; N indicates a minimum of one additional paralog per respective subgenome). 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 subgenomes, respectively (Table 3 and tables S24 and S25).

Of 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 mainly comprised 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 subgenome 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 *pairing homeologous 1* (*Ph1*) locus on chromosome 5B [0:1:0, a locus critical for the diploid meiotic behavior of the wheat homeologous chromosomes (25)]. The phylogenomic analysis indicated that the *GBSS* on 4A is a divergent translocated homeolog originally located on chromosome 7B (fig. S16), whereas *ZIP4* is a transduplication 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 subgenomes that may have occurred in either the diploid progenitor species or the tetraploid ancestor or following the final hexaploidization event in modern bread wheat (Table 3 and figs. S24 and S25). Together with the equal contribution of the three homeologous genomes to the overall gene expression (23), this demonstrates the absence of subgenome dominance (26).

Of the bread wheat HC genes, 29,737 (27%) are present as tandem duplicates, which is up to 10% higher than that found for other monocotyledonous species (table S27). Tandemly repeated genes are most prevalent in the B subgenome (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 subgenome, as a result of interspecific crossing with AABB tetraploid genome progenitors 1 to 2 million years after they diverged (27), is consistent with the synonymous substitution rates of homeologous gene pairs (fig. S17). Homeologous groups with gene duplicates in at least one subgenome (1:1:N, 1:N:1, or N:1:1) showed elevated evolutionary rates (for the subgenome carrying the duplicate) as compared with strict 1:1:1 or 1:1 groups (figs. S18 to S22). Homeologs with recent duplicates also showed higher levels of expression divergence (fig. S23), consistent with gene and genome duplications acting as a driver of functional innovation (28, 29).

Analysis of synteny between the seven triplets of homeologous chromosomes showed high levels

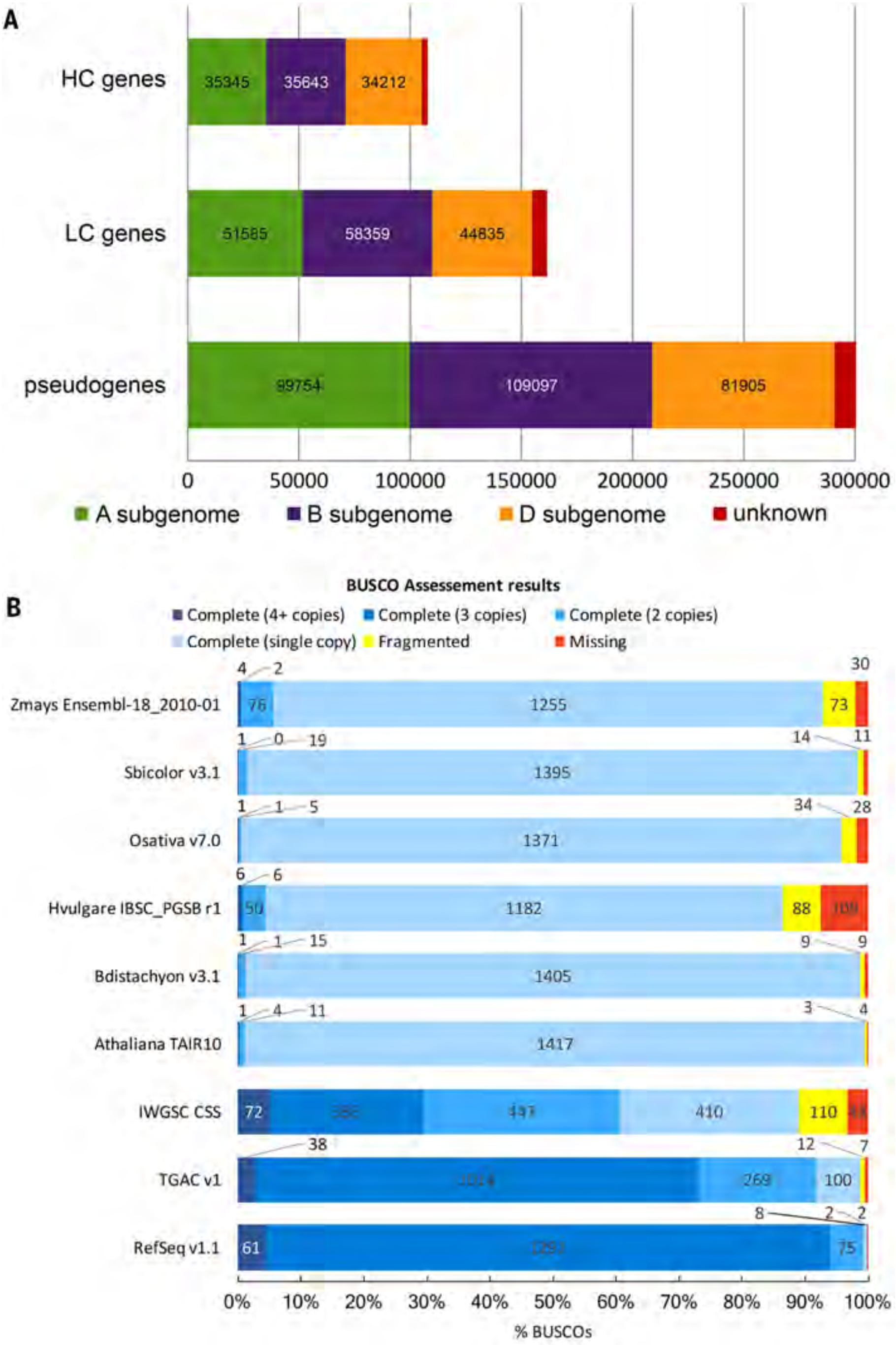

Fig. 2. Evaluation of automated gene annotation. (**A**) Selected gene prediction statistics of IWGSC RefSeq Annotation v1.1, including number and subgenome distribution of HC and LC genes as well as pseudogenes. (**B**) BUSCO v3 gene model evaluation comparing IWGSC RefSeq Annotation v1.1 to earlier 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.

of conservation. There was no evidence that any major rearrangements occurred since the A, B, and D subgenomes diverged ~5 million years ago (Fig. 1D), although collinearity between homeologs was disturbed by inversions occurring, on average, every 74.8 Mb, involving blocks of 10 genes or more (mean gene number of 48.2 with a mean size of 10.5 Mb) (Fig. 1D and table S28). Macrosynteny was conserved across centromere (C) regions, but collinearity (microsynteny) broke down specifically in these recombination-free, gene-poor regions for all seven sets of homeologous chromosomes (Fig. 1D, figs. S24 to S26, and table S29). Of the 113,653 homeologous genes, 80% (90,232) were found organized in macrosynteny, that is, still present at their ancestral position (table S24). At the microsynteny scale, 72% (82,308) of the homeologs were organized in collinear blocks, that is, intervals with a highly conserved gene order (Fig. 1D). A higher proportion of syntenic genes was found in the interstitial regions [short arm, R2a (18), 46% and long arm, R2b (18), 61%] than in the distal telomeric [short arm, R1 (18), 39% and long arm, R3 (18), 51%] and centromere regions [C (18), 29%], and the interstitial compartments harbored larger syntenic blocks (figs. S27 and S28). The higher proportions of duplicated genes in distal-terminal regions (34 and 27% versus 13 to 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, provides a resource to study gene expression in genome-wide and subgenome 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 (Fig. 3A, database S1, and tables S30 to 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 six 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 and S33). A principal component analysis 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 (30–33). Of the total number of genes, 31.0% are expressed in more than 90% of tissues (average 16.9 tpm, ≥30 tissues), and 21.5% are expressed in 10% or fewer tissues (average 0.22 tpm, ≤3 tissues; fig. S31).

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. Abbreviations in parentheses under the headings "Class 1" and "Class 2" indicate transposon types.

Major elements	Wheat subgenome			
	AA	BB	DD	Total
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-retrotransposons (RLX)	2.6	3.5	3.7	3.2
Non-LTR-retrotransposons				
Long interspersed nuclear elements (RIX)	0.81	0.96	0.93	0.90
Short interspersed nuclear elements (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 terminal inverted repeats	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
Unannotated DNA	13.2	14.4	15.7	14.4
(Pre)-microRNAs	0.039	0.057	0.046	0.047
tRNAs	0.0056	0.0050	0.0068	0.0057

Of the HC genes, 8231 showed tissue-exclusive expression (fig. S35). About half of these were associated with reproductive tissues (microspores, anther, and stigma or ovary), consistent with observations in rice (34). The tissue-exclusive genes were enriched for response to extracellular stimuli and reproductive processes (database S2). By contrast, 23,146 HC genes expressed across all 32 tissues were enriched for biological processes associated with housekeeping functions such as protein translation and protein metabolic processes. Tissue-specific genes were shorter [1147 ± 8 base pairs (bp)], had fewer exons (2.76 ± 0.3), and were expressed at lower levels (3.4 ± 0.1 tpm) compared with ubiquitous genes (1429 ± 7 bp, 7.87 ± 0.4 exons, and 17.9 ± 0.4 tpm) (fig. S35).

Genes located in distal regions R1 and R3 (fig. S25 and table S29) showed lower expression breadth than those in the proximal regions (15.7 and 20.7 tissues, respectively) (Fig. 3C and fig. 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, and table S33). The expression breadth pattern was also correlated with the distribution of the repressive H3K27me3 (trimethylated histone H3 lysine 27) (Pearson $r = -0.76$, $P < 2.2 \times 10^{-16}$) and with the active H3K36me3 and H3K9ac (acetylated H3K9) (Pearson $r = 0.9$ and 0.83, respectively; $P < 2.2 \times 10^{-16}$) histone marks (fig. S37).

Global patterns of coexpression (35) were determined with a weighted gene coexpression network analysis (WGCNA) on 94,114 expressed HC genes. Of these genes, 58% (54,401) could be assigned to 38 modules (Fig. 3D and database S4), and, consistent with the principal component analysis, tissues were the major driver of module identity (Fig. 3D and figs. S38 to S40). The analysis focused initially on the 9009 triads (syntenic and nonsyntenic) with a 1:1:1 A:B:D relationship and for which all homeologs were assigned to a module. Of the triads, 16.4% had at least one homeolog in a divergent module, with the B homeolog most likely to be divergent (37.4% B-divergent versus 31.7% A-divergent and 30.9% D-divergent triads, χ^2 test $P = 0.007$). However, the expression profiles of most (83.6%) of the triads were relatively consistent with all homeologs in the same (57.6%) or a closely related (26.0%) module. 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 and table S34). Triads with at least one gene in a nonsyntenic position had a higher amount of

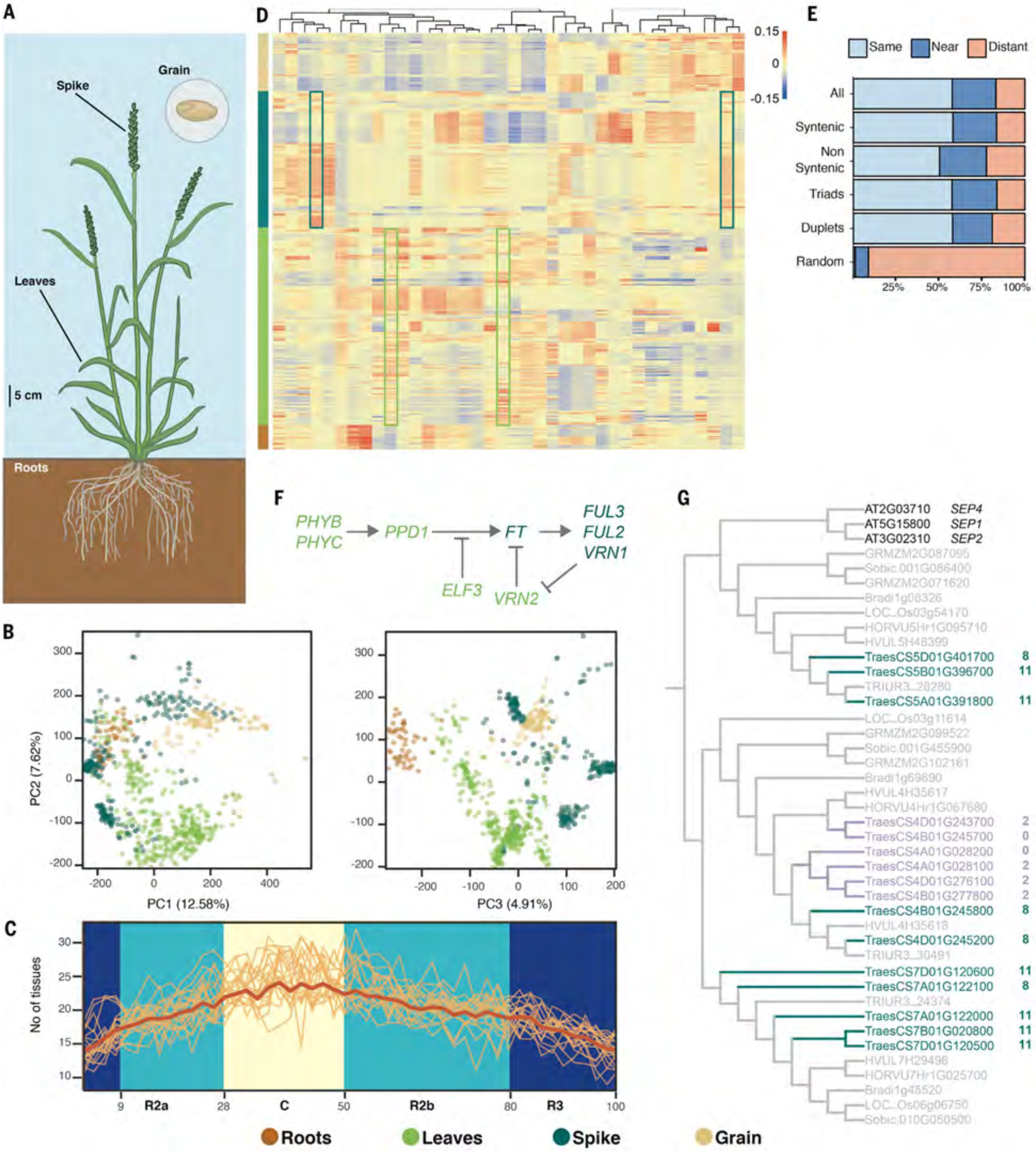

Fig. 3. Wheat atlas of transcription. (**A**) Schematic illustration of a mature wheat plant and high-level tissue definitions for "roots," "leaves," "spike," and "grain" used in the further analysis. (**B**) Principal component (PC) analysis plots for similarity of overall transcription, with samples colored 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 (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 on the basis of a scaled position of each gene within the corresponding genomic compartment (blue, aqua, and light yellow background) across the 21 chromosomes (orange lines). (**D**) Heatmap illustrating the expression of a representative gene (eigengene) for the 38 coexpression 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 (C). DESeq2-normalized expression levels are shown. Modules 1 and 5 (light 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 the WGCNA network. (**F**) Simplified flowering pathway in polyploid wheat. Genes are colored according to their assignment to leaf (light 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). Gray branches indicate non-wheat genes.

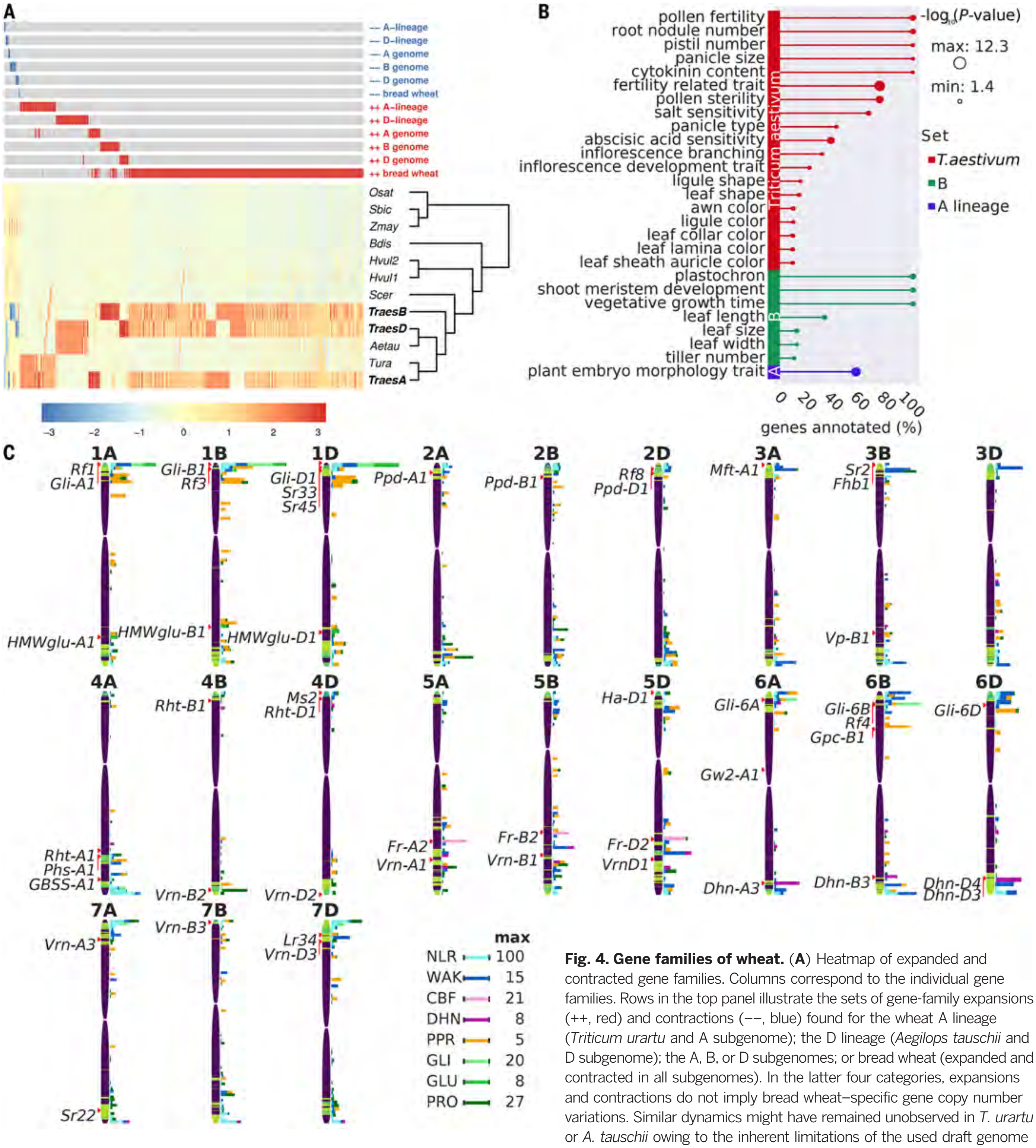

Fig. 4. Gene families of wheat. (**A**) Heatmap of expanded and contracted gene families. Columns correspond to the individual gene families. Rows in the top panel illustrate the sets of gene-family expansions (++, red) and contractions (−−, blue) found for the wheat A lineage (*Triticum urartu* and A subgenome); the D lineage (*Aegilops tauschii* and D subgenome); the A, B, or D subgenomes; or bread wheat (expanded and contracted in all subgenomes). In the latter four categories, expansions and contractions do not imply bread wheat–specific gene copy number variations. Similar dynamics might have remained unobserved in *T. urartu* or *A. tauschii* owing to the inherent limitations of the used draft genome assemblies (53, 54). Rows in the bottom panel heatmap (color scheme on *z*-score scale) indicate the fold expansion and contraction of gene families for the taxa and species included in the analysis [*Oryza sativa* (Osat), *Sorghum bicolor* (Sbic), *Zea mays* (Zmay), *Brachypodium distachyon* (Bdis), *Hordeum vulgare* (Hvul1/2), *Secale cereale* (Scer), *A. tauschii* (Aetau), *T. urartu* (Tura), and wheat A (TraesA), B (TraesB), and D (TraesD) subgenomes]. (**B**) All enriched TO terms for the gene families depicted in (A). Overrepresented TO terms were found for expanded families in bread wheat (all subgenomes, red), the B subgenome (green), and the A lineage (*T. urartu* and A subgenome, 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* (−$\log_{10}$) significance of expansion. (**C**) Genomic distribution of gene families associated with adaptation to biotic (light and dark blue) or abiotic stress (light and dark pink), RNA metabolism in organelles and male fertility (orange), or end-use quality (light, medium, and dark green). Known positions of agronomically important genes and loci are indicated by red arrows and arrowheads to the left of the chromosome bars. Recombination rates are displayed as heatmaps in the chromosome bars [7.2 cM/Mb (light green) to 0 cM/Mb (black)].

divergent expression patterns compared to syntenic triads (21.2 versus 16.2%, χ^2 test $P < 0.001$) and fewer such triads shared all homeologs in the same module (48.7%) compared to syntenic triads (58.0%, chi-square test $P = 0.009$). Similar patterns were observed in the 1933 duplets that have a 1:1 relationship between only two homeologs (table S34). These results are consistent with syntenic homeologs showing similar expression patterns, whereas more dramatic changes in chromosome context associate with divergent expression and possible sub- or neofunctionalization. These trends were also found across diverse tissue-specific networks (23).

To explore the potential of the WGCNA network for identifying previously uncharacterized pathways in wheat, a search was undertaken for modules containing known regulators of wheat flowering time [e.g., *PPD1* (36) and *FT* (37); Fig. 3F]. Genes belonging to this pathway were grouped into specific modules. The upstream genes (*PHYB*, *PHYC*, *PPD1*, *ELF3*, and *VRN2*) were present mainly in modules 1 and 5 and were most highly correlated with expression in leaf and shoot tissues (0.68 and 0.67, respectively; adjusted $P < 1 \times 10^{-108}$). By 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; adjusted $P < 1 \times 10^{-101}$; table S35). The MADS_II transcription factor family that is generally associated with the above pathways was examined more closely, with a focus on the gene tree OG0000041, which contains 54 of the 118 MADS_II genes in wheat. Twenty-four 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 and 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 coexpression. 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 (38), and to dissect genetic factors controlling important agronomic traits (39, 40). A more detailed analysis of tissue-specific and stress-related networks (23) provides a framework for defining quantitative variation and interactions between homeologs for many agronomic traits (41).

Gene-family expansion and 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 (42, 43). In a phylogenomic comparative analysis, wheat gene-family size and wheat-specific gene-family expansion and contraction were benchmarked against nine other grass genomes, including five closely related diploid Triticeae species (table S23 and figs. S13 to S15 and 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

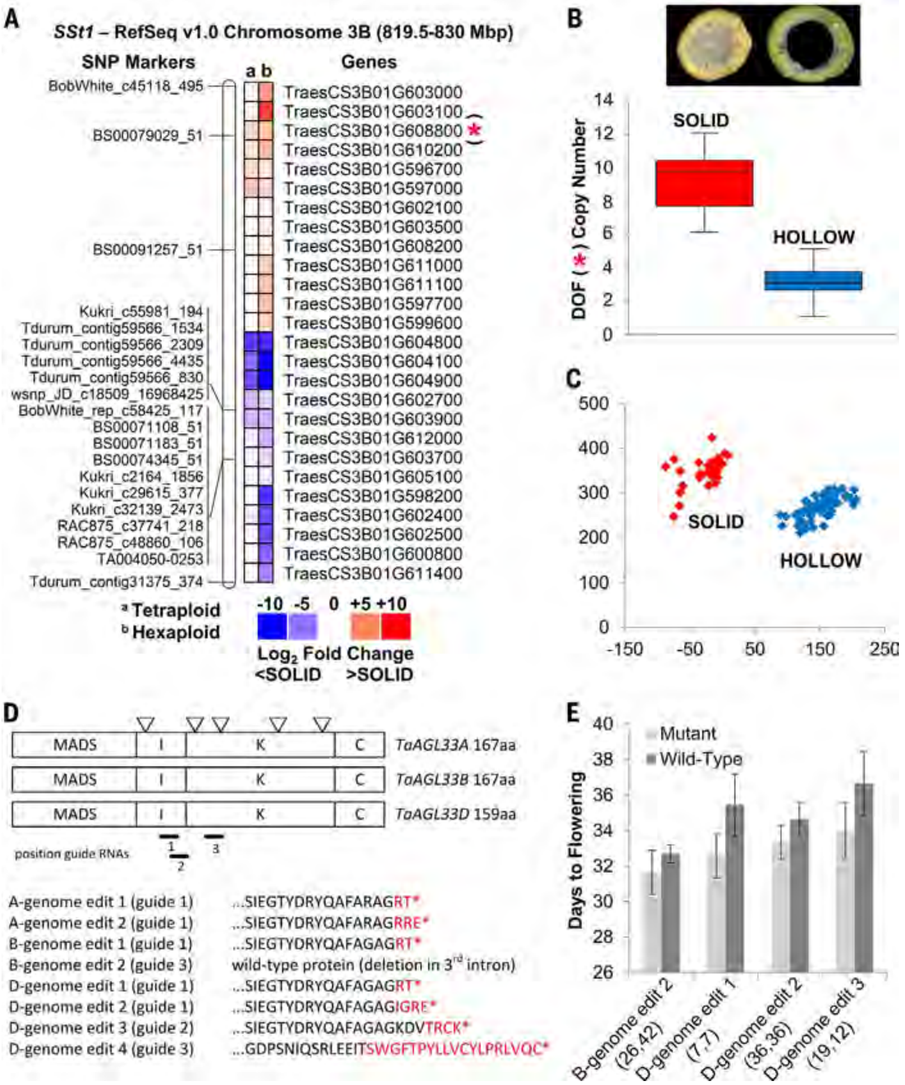

Fig. 5. IWGSC RefSeq v1.0–guided dissection of *SSt1* and *TaAGL33*. (**A**) The Lillian-Vesper population genetic map was anchored to IWGSC RefSeq v1.0 (left), and differentially expressed genes were identified between solid- and hollow-stemmed lines of hexaploid (bread) and tetraploid (durum) wheat (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 (DNA-binding one-zinc finger) 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. Relative intensity of the fluorophores (FAM and HEX) used in KASPar analysis are shown. Vertical axis shows FAM signal; horizontal axis shows HEX signal. (**D**) Schematic of the three TaAGL33 proteins, showing the typical MADS, I, K, and C domains. Triangles indicate the position of the five introns that occur in all three homeologs. Bars indicate the position of single-guide RNAs designed for exons 2 and 3. Three T-DNA vectors—each containing the *bar* selectable marker gene, CRISPR nuclease, and one of three single-guide RNA sequences—were used for *Agrobacterium*-mediated wheat transformation, essentially as described earlier (55). Transgenic plants were obtained with edits at the targeted positions in all TaAGL33 homeologs. The putatively resulting protein sequence is displayed starting close to the edits, with wild-type amino acids (aa) 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 gray bars) and the respective homozygous wild-type segregants (dark gray bars). Numbers in parentheses refer to the number of edited and wild-type plants examined, respectively. Error bars display SEM. Growth conditions were as described in (50).

families containing gene members from at least one of the three wheat subgenomes (tables S36 to S39). Among the 8592 expanded wheat gene families (33% of all families), 6216 were expanded in all three A, B, and D subgenomes (24%; either shared with the wild ancestor or specific to bread wheat, Fig. 4A). Another 1109 were expanded in only one of the wheat subgenomes, and 2102 gene families were expanded in either the A or the D genome lineages (Fig. 4A, fig. S43, and table S36). Overall, only 78 gene families were contracted in wheat. The number of gene families that are only expanded in wheat may be overestimated owing to limited completeness of the draft progenitor wheat genome assemblies used in this study (*14*) (table S39). Gene Ontology (GO; ontology of biomedical terms for the areas "cellular component," "biological process," and "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 quantitative trait loci (QTLs) that were physically mapped to a gene in flowering plant species] analyses identified 1169 distinct GO, PO, and TO terms (15% of all assigned terms) enriched in genes belonging to expanded wheat gene families (Fig. 4B and figs. S44 and S45). "A-subgenome" 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-subgenome" expanded gene families were enriched for TO terms related to plant vegetative growth and development (database S6 and fig. S47). Gene families that were expanded in all wheat subgenomes 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" and "cold stress") and grain yield and quality ("seed maturation," "dormancy," and "germination"). The relationship between the patterns of enriched TO, PO, and GO terms for expanded wheat gene families and key characteristics of wheat performance (figs. S45 to 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 (Fig. 4C and figs. S52 to 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) subclade of P class pentatricopeptide repeat (PPR) proteins, potentially of interest for hybrid wheat production, comprised 207 genes, nearly threefold more per haploid subgenome than have been identified in any other plant genome analyzed to date (*44*, *45*). They localized mainly as clusters of genes in regions on the group 1, 2, and 6 chromosomes, which carry fertility-restoration QTLs in wheat (Fig. 4C and fig. S54). Within 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 and 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 (*46*), the array of *CBF* paralogs at the *Fr-2* locus (fig. S56) revealed by IWGSC RefSeq v1.0 provides a basis for targeted allele mining for previously uncharacterized *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 and (*47*)] either related to the response to heat stress or whose protein epitopes are associated with levels of celiac disease and food allergies (*47*) 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 in diverse climates and environments (*48*). Knowledge of the complex picture of the genome-wide distribution of gene families (Fig. 4C), which needs to be considered for selection in breeding programs in the context of

Table 3. Groups of homeologous genes in wheat. Homeologous genes are "subgenome 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 subgenome-specific (orphan) genes are found only in one subgenome 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 subgenomes. Nonconserved orphans are either singletons or duplicated in the respective subgenome, but neither have obvious homologs in the other subgenomes 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)	Number in wheat genome	Composition of groups (%)	Number of genes in A	Number of genes in B	Number of genes in D	Total number of 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 subgenome orphans			12,412	12,987	10,844	36,243
Nonconserved subgenome singletons			10,084	12,185	8,679	30,948
Nonconserved subgenome duplicated orphans			71	83	38	192
Total (filtered)			**60,328**	**63,935**	**56,773**	**181,036**

distribution of recombination and allelic diversity, can now be applied to 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 gene pools 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 insect damage ([49]) that was disrupted in previous wheat assemblies by a lack of scaffold ordering and annotation, partial assembly, and/or incomplete gene models (fig. S57 and tables S40 and S41). In IWGSC RefSeq v1.0, *SSt1* contains 160 HC genes (table S42), of which 26 were differentially expressed (DESeq2, Benjamini-Hochberg adjusted $P < 0.01$) between wheat lines with contrasting phenotypes. One of the differentially expressed genes, *TraesCS3B01G608800*, was present as a single copy in IWGSC RefSeq v1.0 but showed copy number variation associated with stem solidness in a diverse panel of hexaploid cultivars (Fig. 5A, fig. S58, and table S43). Using IWGSC RefSeq v1.0, we developed a diagnostic SNP marker physically linked to the copy number variation 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 homologs of the *FLOWERING LOCUS C* (*FLC*) gene have been identified as having a role in the vernalization response, a critical process regulating flowering time ([50]). IWGSC RefSeq v1.0 was used to refine the annotation of these six sequences to identify four HC genes and then to design guide RNAs to specifically target, with CRISPR-Cas9–based gene editing, one of these genes, *TaAGL33*, on all subgenomes [*TraesCS3A01G435000* (A), *TraesCS3B01G470000* (B), and *TraesCS3D01G428000* (D)] [Fig. 5C and ([14])]. Editing was obtained at the targeted gene and led to truncated proteins after the MADS box through small deletions and insertions (Fig. 5D). Expression of all homeologs was high before 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 subgenome flowered 2 to 3 days earlier than controls (Fig. 5E). Further refinement should help to fully elucidate the importance of the *TaAGL33* gene for vernalization in monocots. These results exemplify how the IWGSC RefSeq v1.0 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 the 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. Although 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 more efficiently tackle 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 through the implementation of DNA-marker platforms and targeted breeding technologies, including genome editing ([51]).

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. The assembly was superscaffolded 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 genotyping-by-sequencing 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 and contraction of wheat gene families. An atlas of wheat gene transcription was built from an extensive panel of 850 independent transcriptome datasets and was then used to study gene coexpression networks. Furthermore, the assembly was used for the dissection of an important stem-solidness QTL and to design targets for genome editing of genes implicated 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, Food balance sheets (2017); www.fao.org/faostat/en/#data/FBS.
2. Food and Agriculture Organization of the United Nations, FAOSTAT statistics database, Crops (2017); www.fao.org/faostat/en/#data/QC.
3. 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. *Glob. Food Sec.* **12**, 31–37 (2017). doi: 10.1016/j.gfs.2017.01.008; pmid: 28580238
4. J. M. Hickey, T. Chiurugwi, I. Mackay, W. Powell; Implementing Genomic Selection in CGIAR Breeding Programs Workshop Participants, Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat. Genet.* **49**, 1297–1303 (2017). doi: 10.1038/ng.3920; pmid: 28854179
5. K. Arumuganathan, E. D. Earle, Nuclear DNA content of some important plant species. *Plant Mol. Biol. Report.* **9**, 208–218 (1991). doi: 10.1007/BF02672069
6. International Wheat Genome Sequencing Consortium (IWGSC), A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* **345**, 1251788 (2014). doi: 10.1126/science.1251788; pmid: 25035500
7. J. A. Chapman *et al*., A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biol.* **16**, 26 (2015). doi: 10.1186/s13059-015-0582-8; pmid: 25637298
8. 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 Res.* **27**, 885–896 (2017). doi: 10.1101/gr.217117.116; pmid: 28420692
9. A. V. Zimin *et al*., The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum*. *Gigascience* **6**, 1–7 (2017). doi: 10.1093/gigascience/gix097; pmid: 29069494
10. T. R. Endo, B. S. Gill, The deletion stocks of common wheat. *J. Hered.* **87**, 295–307 (1996). doi: 10.1093/oxfordjournals.jhered.a023003
11. M. E. Sorrells *et al*., Comparative DNA sequence analysis of wheat and rice genomes. *Genome Res.* **13**, 1818–1827 (2003). pmid: 12902377
12. K. Eversole, J. Rogers, B. Keller, R. Appels, C. Feuillet, in *Breeding, Quality Traits, Pests and Diseases*, vol. 1 of *Achieving Sustainable Cultivation of Wheat*, P. Langridge, Ed. (Burleigh-Dodds Science Publishing, 2017), chap. 2.
13. E. Paux *et al*., Insertion site-based polymorphism markers open new perspectives for genome saturation and marker-assisted selection in wheat. *Plant Biotechnol. J.* **8**, 196–210 (2010). doi: 10.1111/j.1467-7652.2009.00477.x; pmid: 20078842
14. See supplementary materials.
15. Arabidopsis Genome Initiative, Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000). doi: 10.1038/35048692; pmid: 11130711
16. International Rice Genome Sequencing Project, The map-based sequence of the rice genome. *Nature* **436**, 793–800 (2005). doi: 10.1038/nature03895; pmid: 16100779
17. T. Wicker *et al*., International Wheat Genome Sequencing Consortium, Impact of transposable elements on genome

structure and evolution in wheat. *Genome Biol.* doi: 10.1186/s13059-018-1479-0 (2018).

18. F. Choulet *et al*., Structural and functional partitioning of bread wheat chromosome 3B. *Science* **345**, 1249721 (2014). doi: 10.1126/science.1249721; pmid: 25035497
19. X. Guo *et al*., De novo centromere formation and centromeric sequence expansion in wheat and its wide hybrids. *PLOS Genet.* **12**, e1005997 (2016). doi: 10.1371/journal.pgen.1005997; pmid: 27110907
20. 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 Res.* **24**, 107–116 (2014). doi: 10.1101/gr.160887.113; pmid: 24100079
21. Y. Jiao *et al*., Improved maize reference genome with single-molecule technologies. *Nature* **546**, 524–527 (2017). pmid: 28605751
22. H. Yan *et al*., Intergenic locations of rice centromeric chromatin. *PLOS Biol.* **6**, e286 (2008). doi: 10.1371/journal.pbio.0060286; pmid: 19067486
23. R. H. Ramírez-González *et al*., The transcriptional landscape of polyploid wheat. *Science* **361**, eaar6089 (2018). doi: 10.1126/science.aar6089
24. 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). doi: 10.1093/bioinformatics/btv351; pmid: 26059717
25. 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. *Mol. Breed.* **37**, 95 (2017). doi: 10.1007/s11032-017-0700-2; pmid: 28781573
26. F. Cheng *et al*., Gene retention, fractionation and subgenome differences in polyploid plants. *Nat. Plants* **4**, 258–268 (2018). doi: 10.1038/s41477-018-0136-7; pmid: 29725103
27. T. Marcussen *et al*., Ancient hybridizations among the ancestral genomes of bread wheat. *Science* **345**, 1250092 (2014). pmid: 25035499
28. Y. Van de Peer, S. Maere, A. Meyer, The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* **10**, 725–732 (2009). doi: 10.1038/nrg2600; pmid: 19652647
29. P. S. Soltis, D. E. Soltis, Ancient WGD events as drivers of key innovations in angiosperms. *Curr. Opin. Plant Biol.* **30**, 159–165 (2016). doi: 10.1016/j.pbi.2016.03.015; pmid: 27064530
30. M. Melé *et al*., The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015). doi: 10.1126/science.aaa0355; pmid: 25954002
31. S. C. Stelpflug *et al*., An expanded maize gene expression atlas based on RNA sequencing and its use to explore root development. *Plant Genome* **9**, 0025 (2016). pmid: 27898762
32. F. He *et al*., Large-scale atlas of microarray data reveals the distinct expression landscape of different tissues in Arabidopsis. *Plant J.* **86**, 472–480 (2016). doi: 10.1111/tpj.13175; pmid: 27015116
33. X. Wang *et al*., Comparative genomic analysis of C4 photosynthetic pathway evolution in grasses. *Genome Biol.* **10**, R68 (2009). doi: 10.1186/gb-2009-10-6-r68; pmid: 19549309
34. L. Xia *et al*., Rice Expression Database (RED): An integrated RNA-Seq-derived gene expression database for rice. *J. Genet. Genomics* **44**, 235–241 (2017). doi: 10.1016/j.jgg.2017.05.003; pmid: 28529082
35. R. J. Schaefer, J.-M. Michno, C. L. Myers, Unraveling gene function in agricultural species using gene co-expression networks. *Biochim. Biophys. Acta* **1860**, 53–63 (2017). pmid: 27485388
36. J. Beales, A. Turner, S. Griffiths, J. W. Snape, D. A. Laurie, A *Pseudo-Response Regulator* is misexpressed in the photoperiod insensitive *Ppd-D1a* mutant of wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* **115**, 721–733 (2007). doi: 10.1007/s00122-007-0603-4; pmid: 17634915
37. L. Yan *et al*., The wheat and barley vernalization gene *VRN3* is an orthologue of *FT*. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19581–19586 (2006). doi: 10.1073/pnas.0607142103; pmid: 17158798
38. K. V. Krasileva *et al*., Uncovering hidden variation in polyploid wheat. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E913–E921 (2017). doi: 10.1073/pnas.1619268114; pmid: 28096351
39. S. Wang *et al*., Cytological and transcriptomic analyses reveal important roles of *CLE19* in pollen exine formation. *Plant Physiol.* **175**, 1186–1202 (2017). doi: 10.1104/pp.17.00439; pmid: 28916592
40. M. Pfeifer *et al*., Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science* **345**, 1250091 (2014). doi: 10.1126/science.1250091; pmid: 25035498
41. P. Borrill, N. Adamski, C. Uauy, Genomics as the key to unlocking the polyploid potential of wheat. *New Phytol.* **208**, 1008–1022 (2015). doi: 10.1111/nph.13533; pmid: 26108556
42. F. A. Kondrashov, Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc. Biol. Sci.* **279**, 5048–5057 (2012). doi: 10.1098/rspb.2012.1108; pmid: 22977152
43. P. H. Schiffer, J. Gravemeyer, M. Rauscher, T. Wiehe, Ultra large gene families: A matter of adaptation or genomic parasites? *Life* **6**, 32 (2016). doi: 10.3390/life6030032; pmid: 27509525
44. T. Sykes *et al*., In silico identification of candidate genes for fertility restoration in cytoplasmic male sterile perennial ryegrass (*Lolium perenne* L.). *Genome Biol. Evol.* **9**, 351–362 (2017). pmid: 26951780
45. J. Melonek, J. D. Stone, I. Small, Evolutionary plasticity of restorer-of-fertility-like proteins in rice. *Sci. Rep.* **6**, 35152 (2016). doi: 10.1038/srep35152; pmid: 27775031
46. 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. *Plant J.* **89**, 764–773 (2017). doi: 10.1111/tpj.13424; pmid: 27859852
47. A. Juhász *et al*., Genome mapping of seed-borne allergens and immune-responsive proteins in wheat. *Sci. Adv.* **4**, eaar8602 (2018). doi: 10.1126/sciadv.aar8602
48. M. Feldman, A. A. Levy, in *Alien Introgression in Wheat: Cytogenetics, Molecular Biology, and Genomics*, M. Molnár-Láng, C. Ceoloni, J. Doležel, Eds. (Springer, 2015), pp. 21–76.
49. 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). doi: 10.1371/journal.pone.0175285; pmid: 28399136
50. N. Sharma *et al*., A flowering locus C homolog is a vernalization-regulated repressor in *Brachypodium* and is cold regulated in wheat. *Plant Physiol.* **173**, 1301–1315 (2017). doi: 10.1104/pp.16.01161; pmid: 28034954
51. H. Puchta, Applying CRISPR/Cas for genome engineering in plants: The best is yet to come. *Curr. Opin. Plant Biol.* **36**, 1–8 (2017). doi: 10.1016/j.pbi.2016.11.011; pmid: 27914284
52. International Wheat Genome Sequencing Consortium, Gene family expansion and contraction in the genome of bread wheat cv. Chinese Spring. e!DAL (2018). doi: 10.5447/IPK/2018/5
53. H.-Q. Ling *et al*., Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature* **496**, 87–90 (2013). doi: 10.1038/nature11997; pmid: 23535596
54. J. Jia *et al*., *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* **496**, 91–95 (2013). doi: 10.1038/nature12028; pmid: 23535592
55. Y. Ishida, M. Tsunashima, Y. Hiei, T. Komari, in *Agrobacterium Protocols: Volume 1*, K. Wang, Ed. (Springer, 2015), pp. 189–198.

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 *SSt1* 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, and H. Tarui of RIKEN for assistance with deep sequencing of chromosome 6B; H. Fujisawa, Y. Katayose, K. Kurita, S. Mori, Y. Mukai, and H. Sasaki of the Institute of Crop Science, NARO for assistance with deep sequencing of chromosome 6B; 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 of the University of California, Riverside, and B. Friebe and J. Raupp of 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 superfamily 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, Government of India File No. F grant 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; U.S. 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 has also received funding from the French Government managed by the ANR under the Investment for the Future program (BreedWheat project ANR-10-BTBR-03), from FranceAgriMer (2011-0971 and 2013-0544), French Funds to support Plant Breeding (FSOV), and 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: See below, where authors are arranged by working group and contributions; leaders, co-leaders, and major contributors are listed alphabetically first and then other contributors follow alphabetically. **Competing interests:** The authors declare no competing interests. Bayer CropScience holds a patent application (WO2015000914A1) that covers the modulation of flowering time in monocots using the *FLC* gene. **Data and materials availability:** The IWGSC RefSeq v1.0 assembly and annotation data, physical maps for all chromosomes and chromosome arms, and all data related to this study are available in the IWGSC Data Repository hosted at URGI: https://wheat-urgi.versailles.inra.fr/Seq-Repository, Assembly and annotation data are also available at ENSEMBL-Plants: https://plants.ensembl.org/Triticum_aestivum/Info/Index. The BAC libraries for all chromosomes and 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 cultivar Chinese Spring are provided in database S6 at http://dx.doi.org/10.5447/IPK/2018/5 (*52*). The raw sequencing data used for de novo whole-genome assembly are available from the Sequence Read Archive under accession number SRP114784. RNA-seq data are available at SRA under accession numbers PRJEB25639, PRJEB23056, PRJNA436817, PRJEB25640, SRP133837, and 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 are available under project number SRP133674 (SRR6792673 to SRR6792689). Organellar DNA sequences were deposited at NCBI GenBank (MH051715 and MH051716). Further details on data accessibility are outlined in the supplementary materials and methods.

The International Wheat Genome Sequencing Consortium (IWGSC)

IWGSC RefSeq principal investigators: Rudi Appels[1,36]*‡, Kellye Eversole[2,3]*‡, Catherine Feuillet[17], Beat Keller[41], Jane Rogers[6]‡, Nils Stein[4,5]*‡.
IWGSC whole-genome assembly principal investigators: Curtis J. Pozniak[11]‡, Nils Stein[4,5]*‡, Frédéric Choulet[7], Assaf Distelfeld[25], Kellye Eversole[2,3]*, Jesse Poland[28], Jane Rogers[6], Gil Ronen[12], Andrew G. Sharpe[43].
Whole-genome sequencing and assembly: Curtis Pozniak[11]‡, Gil Ronen[12]‡, Nils Stein[4,5]*‡, Omer Barad[12]†, Kobi Baruch[12]†, Frédéric Choulet[7]†, Gabriel Keeble-Gagnère[1]†, Martin Mascher[4,67]†, Andrew G. Sharpe[43]†, Gil Ben-Zvi[12]†, Ambre-Aurore Josselin[7].
Hi-C data-based scaffolding: Nils Stein[4,5]*‡, Martin Mascher[4,67]‡, Axel Himmelbach[4].
Whole-genome assembly quality control and analyses: Frédéric Choulet[7]‡, Gabriel Keeble-Gagnère[1]‡, Martin Mascher[4,67]‡, Jane Rogers[6]‡, François Balfourier[7], Juan Gutierrez-Gonzalez[30], Matthew Hayden[1], Ambre-Aurore Josselin[7], ChuShin Koh[43], Gary Muehlbauer[30], Raj K. Pasam[1], Etienne Paux[7], Curtis J. Pozniak[11], Philippe Rigault[39], Andrew G. Sharpe[43], Josquin Tibbits[1], Vijay Tiwari[54].
Pseudomolecule assembly: Frédéric Choulet[7]‡, Gabriel Keeble-Gagnère[1]‡, Martin Mascher[4,67]‡, Ambre-Aurore Josselin[7], Jane Rogers[6].
RefSeq genome structure and gene analyses: Manuel Spannagl[9]‡, Frédéric Choulet[7]‡, Daniel Lang[9]†, Heidrun Gundlach[9], Georg Haberer[9], Gabriel Keeble-Gagnère[1], Klaus F. X. Mayer[9,44], Danara Ormanbekova[9,48], Etienne Paux[7], Verena Prade[9], Hana Šimková[8], Thomas Wicker[41].
Automated annotation: Frédéric Choulet[7]‡, Manuel Spannagl[9]‡, David Swarbreck[50]‡, Hélène Rimbert[7]†, Marius Felder[9], Nicolas Guilhot[7], Heidrun Gundlach[9], Georg Haberer[9], Gemy Kaithakottil[50], Jens Keilwagen[40], Daniel Lang[9], Philippe Leroy[7], Thomas Lux[9], Klaus F. X. Mayer[9,44], Sven Twardziok[9], Luca Venturini[50].
Manual gene curation: Rudi Appels[1,36]*‡, Hélène Rimbert[7]†, Frédéric Choulet[7], Angéla Juhász[36,37], Gabriel Keeble-Gagnère[1].
Subgenome comparative analyses: Frédéric Choulet[7]‡, Manuel Spannagl[9]‡, Daniel Lang[9]†, Michael Abrouk[8,19], Georg Haberer[9], Gabriel Keeble-Gagnère[1], Klaus F. X. Mayer[9,44], Thomas Wicker[41].
Transposable elements: Frédéric Choulet[7]‡, Thomas Wicker[41]†, Heidrun Gundlach[9]†, Daniel Lang[9], Manuel Spannagl[9].
Phylogenomic analyses: Daniel Lang[9]‡, Manuel Spannagl[9]‡, Rudi Appels[1,36]*, Iris Fischer[9].
Transcriptome analyses and RNA-seq data: Cristobal Uauy[10]‡, Philippa Borrill[10]†, Ricardo H. Ramirez-Gonzalez[10]†, Rudi Appels[1,36]*, Dominique Arnaud[63], Smahane Chalabi[63], Boulos Chalhoub[62,63], Frédéric Choulet[7], Aron Cory[11], Raju Datla[22], Mark W. Davey[18], Matthew Hayden[1], John Jacobs[18], Daniel Lang[9], Stephen J. Robinson[52], Manuel Spannagl[9], Burkhard Steuernagel[10], Josquin Tibbits[1], Vijay Tiwari[54], Fred van Ex[18], Brande B. H. Wulff[10].
Whole-genome methylome: Curtis J. Pozniak[11]‡, Stephen J. Robinson[52]‡, Andrew G. Sharpe[43]‡, Aron Cory[11].
Histone mark analyses: Moussa Benhamed[15]‡, Etienne Paux[7]‡, Abdelhafid Bendahmane[15], Lorenzo Concia[15], David Latrasse[15].
BAC chromosome MTP IWGSC–Bayer Whole-Genome Profiling (WGP) tags: Jane Rogers[6]‡, John Jacobs[18]‡, Michael Alaux[13], Rudi Appels[1,36]*, Jan Bartoš[8], Arnaud Bellec[20], Hélène Berges[20], Jaroslav Doležel[8], Catherine Feuillet[17], Zeev Frenkel[26], Bikram Gill[28], Abraham Korol[26], Thomas Letellier[13], Odd-Arne Olsen[56], Hana Šimková[8], Kuldeep Singh[65], Miroslav Valárik[8], Edwin van der Vossen[64], Sonia Vautrin[20], Song Weining[66].
Chromosome LTC mapping and physical mapping quality control: Abraham Korol[26]‡, Zeev Frenkel[26]†, Tzion Fahima[26]†, Vladimir Glikson[29], Dina Raats[50], Jane Rogers[6].
RH mapping: Vijay Tiwari[54]‡, Bikram Gill[28], Etienne Paux[7], Jesse Poland[28].
Optical mapping: Jaroslav Doležel[8]‡, Jarmila Číhalíková[8], Hana Šimková[8], Helena Toegelová[8], Jan Vrána[8].
Recombination analyses: Pierre Sourdille‡[7], Benoit Darrier[7].
Gene family analyses: Rudi Appels[1,36]*‡, Manuel Spannagl[9]‡, Daniel Lang[9]†, Iris Fischer[9], Danara Ormanbekova[9,48], Verena Prade[9].
CBF gene family: Delfina Barabaschi[16]‡, Luigi Cattivelli[16].
Dehydrin gene family: Pilar Hernandez[33]‡, Sergio Galvez[27]‡, Hikmet Budak[14].
NLR gene family: Burkhard Steuernagel[10]‡, Jonathan D. G. Jones[35], Kamil Witek[35], Brande B. H. Wulff[10], Guotai Yu[10].
PPR gene family: Ian Small[45]‡, Joanna Melonek[45]†, Ruonan Zhou[4].
Prolamin gene family: Angéla Juhász[36,37]‡, Tatiana Belova[56]†, Rudi Appels[1,36]*, Odd-Arne Olsen[56].
WAK gene family: Kostya Kanyuka[38]‡, Robert King[42]†.
Stem solidness (*SSt1*) QTL team: Kirby Nilsen[11]‡, Sean Walkowiak[11]‡, Curtis J. Pozniak[11]‡, Richard Cuthbert[21], Raju Datla[22], Ron Knox[21], Krysta Wiebe[11], Daoquan Xiang[22].
Flowering locus C (*FLC*) gene team: Antje Rohde[72]‡, Timothy Golds[18]‡.
Genome size analysis: Jaroslav Doležel[8]‡, Jana Čížková[8], Josquin Tibbits[1].
MicroRNA and tRNA annotation: Hikmet Budak[14]‡, Bala Ani Akpinar[14], Sezgi Biyiklioglu[14].
Genetic maps and mapping: Gary Muehlbauer[30]‡, Jesse Poland[28]‡, Liangliang Gao[28], Juan Gutierrez-Gonzalez[30], Amidou N'Daiye[11].
BAC libraries and chromosome sorting: Jaroslav Doležel[8]‡, Hana Šimková[8]†, Jarmila Číhalíková[8], Marie Kubaláková[8], Jan Šafář[8], Jan Vrána[8].
BAC pooling, BAC library repository, and access: Hélène Berges[20]‡, Arnaud Bellec[20], Sonia Vautrin[20].
IWGSC sequence and data repository and access: Michael Alaux[13]‡, Françoise Alfama[13], Anne-Françoise Adam-Blondon[13], Raphael Flores[13], Claire Guerche[13], Thomas Letellier[13], Mikaël Loaec[13], Hadi Quesneville[13].
Physical maps and BAC-based sequences:
1A BAC sequencing and assembly: Curtis J. Pozniak[11]‡, Andrew G. Sharpe[22,43]‡, Sean Walkowiak‡[11], Hikmet Budak[14], Janet Condie[22], Jennifer Ens[11], ChuShin Koh[43], Ron Maclachlan[11], Yifang Tan[22], Thomas Wicker[41].
1B BAC sequencing and assembly: Frédéric Choulet[7]‡, Etienne Paux[7]‡, Adriana Alberti[61], Jean-Marc Aury[61], François Balfourier[7], Valérie Barbe[61], Arnaud Couloux[61], Corinne Cruaud[61], Karine Labadie[61], Sophie Mangenot[61], Patrick Wincker[61,68,69].
1D, 4D, and 6D physical mapping: Bikram Gill[28]‡, Gaganpreet Kaur[28], Mingcheng Luo[34], Sunish Sehgal[53].
2AL physical mapping: Kuldeep Singh[65]‡, Parveen Chhuneja[65], Om Prakash Gupta[65], Suruchi Jindal[65], Parampreet Kaur[65], Palvi Malik[65], Priti Sharma[65], Bharat Yadav[65].
2AS physical mapping: Nagendra K. Singh[70]‡, Jitendra P. Khurana[71]‡, Chanderkant Chaudhary[71], Paramjit Khurana[71], Vinod Kumar[70], Ajay Mahato[70], Saloni Mathur[71], Amitha Sevanthi[70], Naveen Sharma[71], Ram Sewak Tomar[70].
2B, 2D, 4B, 5BL, and 5DL IWGSC–Bayer Whole-Genome Profiling (WGP) physical maps: Jane Rogers[6]‡, John Jacobs[18]‡, Michael Alaux[13], Arnaud Bellec[20], Hélène Berges[20], Jaroslav Doležel[8], Catherine Feuillet[17], Zeev Frenkel[26], Bikram Gill[28], Abraham Korol[26], Edwin van der Vossen[64], Sonia Vautrin[20].
3AL physical mapping: Bikram Gill[28]‡, Gaganpreet Kaur[28], Mingcheng Luo[34], Sunish Sehgal[53].
3DS physical mapping and BAC sequencing and assembly: Jan Bartoš[8]‡, Kateřina Holušová[8], Ondřej Plíhal[49].
3DL BAC sequencing and assembly: Matthew D. Clark[50,73], Darren Heavens[50], George Kettleborough[50], Jon Wright[50].
4A physical mapping, BAC sequencing, assembly, and annotation: Miroslav Valárik[8]‡, Michael Abrouk[8,19], Barbora Balcárková[8], Kateřina Holušová[8], Yuqin Hu[34], Mingcheng Luo[34].
5BS BAC sequencing and assembly: Elena Salina[47]‡, Nikolai Ravin[23,51]‡, Konstantin Skryabin[23,51]‡, Alexey Beletsky[23], Vitaly Kadnikov[23], Andrey Mardanov[23], Michail Nesterov[47], Andrey Rakitin[23], Ekaterina Sergeeva[47].
6B BAC sequencing and assembly: Hirokazu Handa[31]‡, Hiroyuki Kanamori[31], Satoshi Katagiri[31], Fuminori Kobayashi[31], Shuhei Nasuda[46], Tsuyoshi Tanaka[31], Jianzhong Wu[31].
7A physical mapping and BAC sequencing: Rudi Appels[1,36]*‡, Matthew Hayden[1], Gabriel Keeble-Gagnère[1], Philippe Rigault[39], Josquin Tibbits[1].
7B physical mapping, BAC sequencing, and assembly: Odd-Arne Olsen[56]‡, Tatiana Belova[56]†, Federica Cattonaro[58], Min Jiumeng[60], Karl Kugler[9], Klaus F. X. Mayer[9,44], Matthias Pfeifer[9], Simen Sandve[57], Xu Xun[59], Bujie Zhan[56]†.
7DS BAC sequencing and assembly: Hana Šimková[8]‡, Michael Abrouk[8,19], Jacqueline Batley[24], Philipp E. Bayer[24], David Edwards[24], Satomi Hayashi[32], Helena Toegelová[8], Zuzana Tulpová[8], Paul Visendi[55].
7DL physical mapping and BAC sequencing: Song Weining[66]‡, Licao Cui[66], Xianghong Du[66], Kewei Feng[66], Xiaojun Nie[66], Wei Tong[66], Le Wang[66].
Figures: Philippa Borrill[10], Heidrun Gundlach[9], Sergio Galvez[27], Gemy Kaithakottil[50], Daniel Lang[9], Thomas Lux[9], Martin Mascher[4,67], Danara Ormanbekova[9,48], Verena Prade[9], Ricardo H. Ramirez-Gonzalez[10], Manuel Spannagl[9], Nils Stein[4,5]*, Cristobal Uauy[10], Luca Venturini[50].
Manuscript writing team: Nils Stein[4,5]*‡, Rudi Appels[1,36]*‡, Kellye Eversole[2,3]*, Jane Rogers[6]†, Philippa Borrill[10], Luigi Cattivelli[16], Frédéric Choulet[7], Pilar Hernandez[33], Kostya Kanyuka[38], Daniel Lang[9], Martin Mascher[4,67], Kirby Nilsen[11], Etienne Paux[7], Curtis J. Pozniak[11], Ricardo H. Ramirez-Gonzalez[10], Hana Šimková[8], Ian Small[45], Manuel Spannagl[9], David Swarbreck[50], Cristobal Uauy[10].

[1]AgriBio, Centre for AgriBioscience, Department of Economic Development, Jobs, Transport, and Resources, 5 Ring Road, La Trobe University, Bundoora, VIC 3083, Australia. [2]International Wheat Genome Sequencing Consortium (IWGSC), 5207 Wyoming Road, Bethesda, MD 20816, USA. [3]Eversole Associates, 5207 Wyoming Road, Bethesda, MD 20816, USA. [4]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Genebank, Corrensstr. 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, UK. [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, UK. [11]University of Saskatchewan, Crop Development Centre, Agriculture Building, 51 Campus Drive, Saskatoon, SK S7N 5A8, Canada. [12]NRGene Ltd., 5 Golda Meir Street, Ness Ziona 7403648, Israel. [13]URGI, INRA, Université Paris-Saclay, 78026 Versailles, France. [14]Plant Sciences and Plant Pathology, Cereal Genomics Lab, Montana State University, 412 Leon Johnson Hall, Bozeman, MT 59717, USA. [15]Biology Department, Institute of Plant Sciences–Paris-Saclay, 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 and Bioinformatics, via S. Protaso, 302, I -29017

Fiorenzuola d'Arda, Italy. [17]Bayer CropScience, Crop Science Division, Research and Development, Innovation Centre, 3500 Paramount Parkway, Morrisville, NC 27560, USA. [18]Bayer CropScience, Trait Research, Innovation Center, Technologiepark 38, 9052 Gent, Belgium. [19]Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, 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 Avenue 33, Building 2, Moscow 119071, Russia. [24]School of Biological Sciences and Institute of Agriculture, University of Western Australia, Perth, WA 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]Plant Pathology, Throckmorton Hall, Kansas State University, Manhattan, KS 66506, USA. [29]MultiQTL Ltd., University of Haifa, Haifa 3498838, Israel. [30]Department of Agronomy and Plant Genetics, University of Minnesota, 411 Borlaug Hall, St. Paul, MN 55108, USA. [31]Institute of Crop Science, NARO, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8518, Japan. [32]Queensland University of Technology, Earth, Environmental and Biological Sciences, Brisbane, QLD 4001, Australia. [33]Instituto de Agricultura Sostenible (IAS-CSIC), Consejo Superior de Investigaciones Científicas, Alameda del Obispo s/n, 14004 Córdoba, Spain. [34]Department of Plant Sciences, University of California, Davis, One Shield Avenue, Davis, CA 95617, USA. [35]The Sainsbury Laboratory, Norwich Research Park, Norwich NR4 7UH, UK. [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, UK. [39]GYDLE, Suite 220, 1135 Grande Allée, Ouest, Québec, QC G1S 1E7, Canada. [40]Julius Kühn-Institut, Institute for Biosafety in Plant Biotechnology, Erwin-Baur-Str. 27, 06484 Quedlinburg, Germany. [41]Department of Plant and Microbial Biology, University of Zurich, Zollikerstrasse 107, 8008 Zurich, Switzerland. [42]Rothamsted Research, Computational and Analytical Sciences, West Common, Harpenden AL5 2JQ, UK. [43]University of Saskatchewan, Global Institute for Food Security, 110 Gymnasium Place, Saskatoon, SK S7N 4J8, Canada. [44]School of Life Sciences Weihenstephan, Technical University of Munich, 85354 Freising, Germany. [45]School of Molecular Sciences, ARC Centre of Excellence in Plant Energy Biology, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia. [46]Graduate School of Agriculture, Kyoto University, 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]Department of Agricultural Sciences, University of Bologna, Viale Fanin, 44 40127 Bologna, Italy. [49]Department of Molecular Biology, Centre of the Region Haná for Biotechnological and Agricultural Research, Palacký University, Šlechtitelů 27, CZ-78371 Olomouc, Czech Republic. [50]Earlham Institute, Core Bioinformatics, Norwich NR4 7UZ, UK. [51]Faculty of Biology, Moscow State University, 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]Agronomy Horticulture and Plant Science, South Dakota State University, 2108 Jackrabbit Drive, Brookings, SD 57006, USA. [54]Plant Science and Landscape Architecture, University of Maryland, 4291 Fieldhouse Road, 2102 Plant Sciences Building, College Park, MD 20742, USA. [55]University of Greenwich, Natural Resources Institute, Central Avenue, Chatham, Kent ME4 4TB, UK. [56]Faculty of Bioscience, Department of Plant Science, Norwegian University of Life Sciences, Arboretveien 6, 1433 Ås, Norway. [57]Faculty of Bioscience, Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Arboretveien 6, 1433 Ås, Norway. [58]Instituto di Genomica Applicata, Via J. Linussio 51, Udine 33100, Italy. [59]BGI-Shenzhen, BGI Genomics, Yantian District, Shenzhen 518083, Guangdong, China. [60]BGI-Shenzhen, BGI Genomics, Building No. 7, BGI Park, No. 21 Hongan 3rd Street, Yantian District, Shenzhen 518083, China. [61]CEA–Institut de Biologie François-Jacob, Genoscope, 2 rue Gaston Crémieux, 91057 Evry Cedex, France. [62]Monsanto SAS, 28000 Boissay, France. [63]Institut National de la Recherche Agronomique (INRA), 2 rue Gaston Crémieux, 91057 Evry Cedex, France. [64]Keygene, N.V., Agro Business Park 90, 6708 PW Wageningen, 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]State Key Laboratory of Crop Stress Biology in Arid Areas, College of Agronomy, Northwest A&F University, Yangling, Shaanxi 712101, China. [67]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany. [68]CNRS, UMR 8030, CP5706, 91057 Evry, France. [69]Université d'Evry, UMR 8030, CP5706, 91057 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 and Department of Plant Molecular Biology, Benito Juarez Road, New Delhi 110021, India. [72]Bayer CropScience, Breeding and Trait Development, Technologiepark 38, 9052 Gent, Belgium. [73]Department of Lifesciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK.

Authorship of this paper should be cited as "International Wheat Genome Sequencing Consortium" (IWGSC, 2018).
*Corresponding author. Email: rudi.appels@unimelb.edu.au (R.A.); eversole@eversoleassociates.com (K.E.); stein@ipk-gatersleben.de (N.S.)
†Major contributors
‡Working group leader(s) or co-leaders

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6403/eaar7191/suppl/DC1
Materials and Methods
Figs. S1 to S59
Tables S1 to S43
References (*56*–*186*)
Databases S1 to S5

13 December 2017; accepted 11 July 2018
10.1126/science.aar7191