

REVIEW

Wheat research and breeding in the new era of a high-quality reference genome

Rudi APPELS (✉)^{1,2}

¹ Centre for AgriBioscience/Department of Economic Development, LaTrobe University, Bundoora VIC 3083, Australia
² Department of BioSciences, The University of Melbourne, Victoria 2052, Australia

Abstract The publications of the International Wheat Genome Sequencing Consortium (IWGSC) released in August 2018 are reviewed and placed into the context of developments arising from the availability of the high-quality wheat genome assembly.

Keywords assembly technology, molecular markers, polyploidy, transcript networks, wheat genome

1 Introduction

On 17 August 2018, the International Wheat Genome Sequencing Consortium (IWGSC) coordinated the publication of two papers in *Science*^[1,2], one in *Science Advances*^[3] and four in *Genome Biology*^[4–7] to mark the release of the IWGSC RefSeq v1.0 wheat genome assembly. These papers arose from more than a decade of effort by the IWGSC to develop linearly ordered genome models (pseudomolecules) for each of the 21 chromosomes of the wheat cultivar Chinese Spring (CS). CS was chosen because of the extensive studies by Sears in developing the chromosome-cytogenetic lines (facilitated by crossability to non-wheat lines, such as rye, and its high frequency of misdivision of univalents^[8]) and their subsequent extensive deployment by the wheat research community in early mapping of the genome^[9]. The choice of CS as a reference genome was also notable because it was selected from an awnless Sichuan white landrace Chengduguangtou^[10,11] in the 1950s and thus relatively free of the many alien chromosome introgressions present in modern wheat cultivars.

The assembly of pseudomolecules in IWGSC RefSeq v1.0 provides a high-quality linear assembly of each chromosome comprising 70–80 superscaffolds per chromosome and is freely available^[7]. The assemblies of

human and model organisms have generally included sequencing of bacterial artificial chromosome (BAC)-based physical assemblies, and this was how the IWGSC RefSeq v1.0 assembly was initiated in 2005 using flow sorted chromosome telocentric chromosome arms^[12,13]. The IWGSC RefSeq v1.0 wheat genome assembly was finally achieved by combining a primarily whole genome short-read-based assembly with Hi-C, BAC sequencing and genetic/optical mapping information^[1]. The high-quality IWGSC RefSeq v1.0 assembly covers 94% of the estimated 15.7 Gb of DNA within the genome and thus still requires some local base-level assembly to achieve a finished status, with no gaps, through integrating available data sets^[14,15] as well as new data from optical mapping and more long-read sequencing from CS DNA. The chloroplast and mitochondrial DNA genomes from CS are finished^[16].

2 Why did the sequencing and assembly of the wheat genome take so long?

When the IWGSC project started, the wheat genome project was clearly more complex than any other genome sequenced at the time because of its hexaploid nature and size (about five times the human genome). The flow sorting of telocentric chromosomes and the construction of individual BAC libraries reduced the size and complexity by a factor of 42 and short read sequencing allowed a start to be made to compiling the wheat genome through individual chromosome-based projects by colleagues in different countries and coordinated by the IWGSC. The scale of an individual wheat chromosome project exceeded that of the entire rice genome, and the cost and new bioinformatics/computing requirements limited progress. The high density molecular genetic maps combined with significant breakthroughs from 2015 onwards in applying new optical mapping^[17] and nuclear chromatin conformation capture^[18] technologies to wheat provided a capacity

Received January 18, 2019; accepted April 17, 2019

Correspondence: rudi.appels@unimelb.edu.au

to bridge the large genomic distances between gene islands in the wheat genome (Fig. 1). Linkage of contigs and scaffolds into superscaffolds resulted in a single linear assembly of *c.* 700 Mb of genome sequences for individual chromosomes. As evident in Fig. 1, the repetitive retrotransposable elements between gene islands represented a major obstacle for the assembly algorithms. From 2015, the new “wet chemistry” combined with a major bioinformatics development by NRGene (DeNovoMagic 2, NRGene, Ness-Ziona, Israel) to allow for assembling short Illumina reads (accurate length) from unfractionated wheat nuclear DNA^[1].

Validation of sections of the IWGSC RefSeq v1.0 genome assembly was possible using data from the assemblies from earlier IWGSC chromosome projects^[20] as well as the assemblies from independent efforts^[14,15] (TGAC and PacBio, Fig. 2). The alignments in Fig. 2 indicate that the information exists to produce finished genome sequences^[4].

3 What makes wheat wheat?

IWGSC RefSeq v1.0 now provides the basis for detailed studies of the wheat genetic engine, and an appreciation of how it operates and the depth of its plasticity where up to one third of the high confidence genes annotated on the A, B and D genomes can either be amplified in gene number or deleted (Fig. 3). The phylogenomic analyses^[1] suggest that, in the case of changed numbers in homeologous gene families, many instances have occurred since the formation of the hexaploid wheat some 10000 years ago^[21], correlating with environmental adaptation and end-use quality selection through evolving human agricultural practices.

Layered across the changes in gene number is the possibility of differential levels of expression of homeologous genes in different tissues^[2] and in response to a specific feature of the environment, such as temperature^[3]. The study of Juhász et al.^[3] focused on the prolamin family

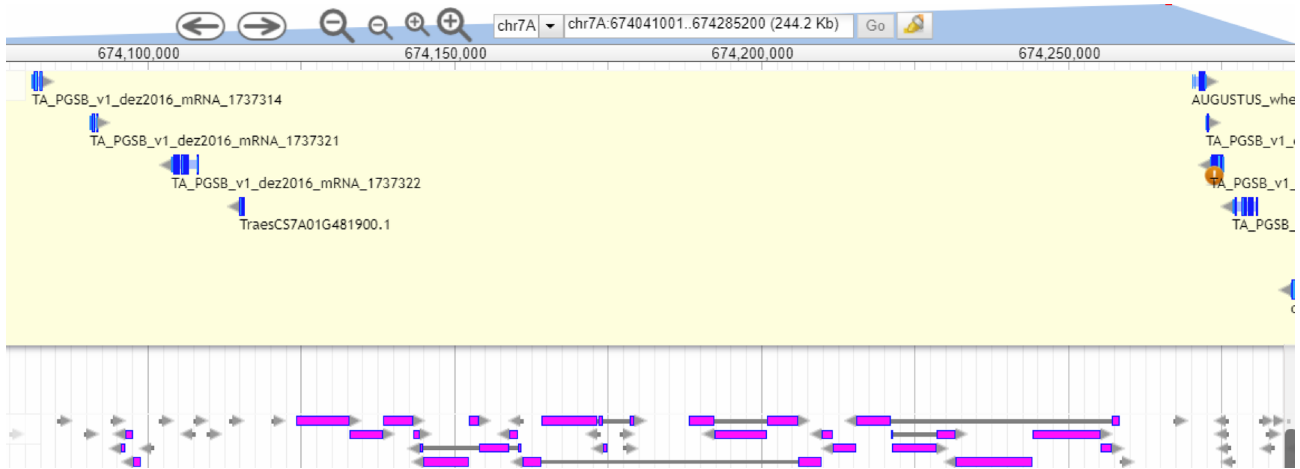


Fig. 1 A representative view of *c.* 200 kb of the wheat genome showing the gene islands (blue) and retrotransposable elements (pink). The scale near the top of the image is in the genome position scale base pairs. The image is a screen capture of IWGSC RefSeq v1.0 viewed in Apollo^[19] as established in the Earlham Institute, Norwich Research Park, UK.

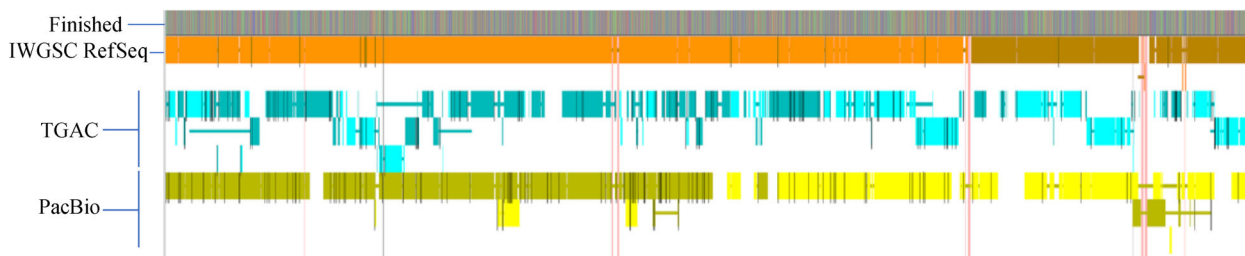


Fig. 2 Detail of a 2.4-Mb region from chromosome 7AS (modified from Keeble-Gagnère et al.^[4]). The matching regions from IWGSC RefSeq v1.0 (orange), TGAC (cyan) and PacBio (yellow) assemblies show the alignments of the genome sequences and the black bars indicate differences between genome sequences. The vertical pink bars indicate regions of the finished sequence not present in any other assembly.

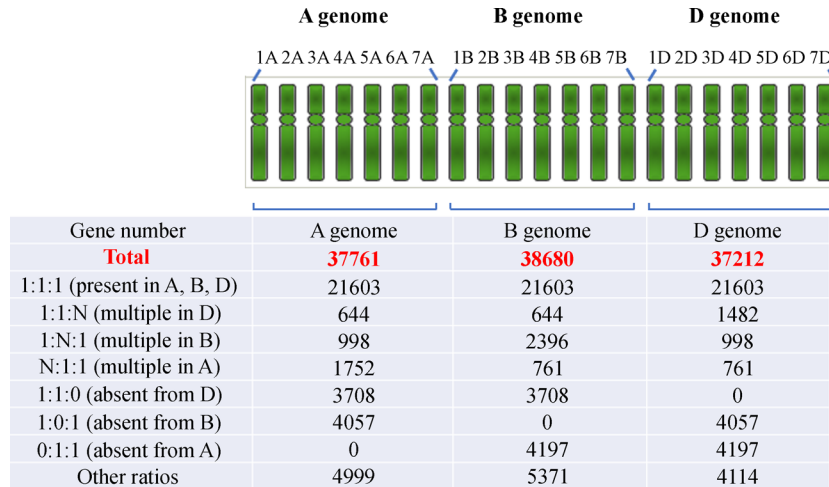


Fig. 3 Summary of overall diversity in gene numbers within the wheat genome: modified from IWGSC^[1]. The ratio, 1:1:1, indicates homeologous genes present on the A, B and D genomes. The ratio, 1:1:N, indicates the respective homeologous gene on the D genome has multiple copies, similarly for 1:N:1 and N:1:1 designators as indicated in the table. The ratio, 1:1:0, indicates the respective homeologous gene on the D genome is absent, similarly for 1:0:1 and 0:1:1 designators as indicated in the table.

of genes amplified in wheat and complemented the extensive analyses of the respective proteins in wheat^[22], including the recent studies by Altenbach et al.^[23] and Kawaura et al.^[24]. The prolamin family of genes encode the gliadin and glutenin proteins that accumulate in endosperm-protein bodies and provide a major source of nitrogen for the germinating embryo as well as nutrition for consumers of the grain. The gene families contributing to starch synthesis to form the endosperm-starch granules are not as complex as the prolamins but do show 1:0:1 (Fig. 3) type variation that can have major implications for quality attributes associated with flour from the milled grain. The granule bound starch synthase (*GBSS*) on chromosome 4A (1:0:0) is a key determinant of udon noodle quality^[25] and phylogenomic analyses^[1] confirmed that this gene was a divergent, translocated, homeolog of a gene originally located on chromosome 7B, syntenic to the *GBSS* genes on 7A and 7D. The detailed genomic analysis of a large collection of wheat lines^[1] indicated that the natural deletion of *GBSS* on chromosome 4A^[25] was specifically limited to the gene itself, thus minimizing any other adverse effects on the plant.

Human health aspects of consuming wheat relating to celiac disease and allergic reactions^[3,22–24] could be addressed in detail at the genomics level with IWGSC RefSeq v1.0^[3]. The study of Juhász et al.^[3] established a new reference map for immunostimulatory wheat proteins as a basis for selecting wheat lines and developing diagnostics for products with more favorable consumer attributes. The changing profiles of target grain proteins resulting from variation in temperature at which the wheat spike matured are of particular interest for the production of food that can be used as a safe and healthy long-term alternative to the currently highly restrictive and short-term

approaches that rely on absolute wheat and gluten avoidance.

On a broader scale, the transcriptome study by Ramírez-González et al.^[2] using IWGSC RefSeq v1.0 generated tissue co-expression networks that accounted for differential expression throughout development. These networks, alongside detailed gene expression atlases, indicated the importance of polyploidy in shaping the response of wheat to stress and generates tissue specific expression. Homeologous genes with high variation in expression between tissues were, as expected, more varied for cis-regulatory elements in the promoter regions and featured more frequent transposable element insertions. The transcriptome study also indicated a significantly greater divergence in spatial expression within non-syntenic gene models (where a gene falls outside the expected order between blocks of A, B and D subgenome sequences) in all tissue networks compared to syntenic gene models (conservation of gene order in blocks of A, B and D subgenome sequences) and this was consistent with the possibility that non-syntenic gene models were under a more relaxed selection pressure. In both syntenic and non-syntenic homeologous genes there was a small but significant bias toward D genome genes showing a more stable expression pattern in all 15 tissues studied by Ramírez-González et al.^[2]. This bias toward the D genome was also observed for *H3K27me3* (histone H3 lysine 27 trimethylation) across gene models which showed a lower distribution of this methylation feature associated with repression of transcription.

The study of the retrotransposable element (TEs) space in the A, B and D subgenomes^[26] covered 85% of the total genome and indicated that while syntenic TEs have been replaced by the insertion of novel elements, the distances

between genes have remained relatively constant. Overall 3968974 copies of TEs belonging to 505 families were annotated including 112744 full-length long-terminal repeat (LTR) retrotransposons that are normally difficult to identify. Importantly the composition of TEs within 2 kb of gene models differed from those in the core intergenic regions (large TEs, such as LTR-RTs and CACTAs) and were enriched in small TEs (< 1000 bp) in the miniature inverted-repeat transposable elements, SINEs and mutator classes^[26]. This bias was shared by A, B and D genomes, and may relate to patterns of homeologous gene expression; this needs further study^[2].

At the whole subgenome level, the TE families of the A, B and D genomes are in similar proportions to those found in the diploid progenitors^[26] although within the sub-genomes certain individual transposons domains show differences in content; 11 out of 505 TE families showed a greater than threefold difference between sub-genomes^[1,26]. The TEs are evidently a key feature of the plasticity that exists within the wheat genome at a quantitative rather than qualitative level. However, the possibility that TEs provide contact points for unequal crossovers to generate indels has been precluded by a detailed comparison between the genome sequences for chromosome 2D in CS and the cultivar CH Campala Lr22a^[27].

4 Future developments

The publication of the IWGSC RefSeq v1.0 genome sequence now provides a good set of genomes from plants that are important in agriculture to conduct fundamental studies to define attributes in, for example, rice and then examine/test the functional ideas in a polyploid situation in the case of wheat. To meet the demands of changes in the environment, the major advances in all the levels of genomics through to bioinformatics need to be captured for choosing parents to be used in crossing programs so that the genetic makeup of wheat can be refined and continue to provide up to a fifth of the total calories consumed by humans globally. Furthermore, the analysis of progeny from crosses for advancing yield and end-use quality, while adapting the crop to specific regional diseases and abiotic stresses, will require accurate and early assessment of traits before larger scale phenotyping is undertaken. Although these are not new issues, the technological advances provide for some novel approaches.

4.1 New markers for choosing parents and tracking progeny

The established application of molecular markers is to deploy diagnostic DNA sequences that are tightly linked to a phenotype of interest. Early studies deployed repetitive sequences to track so-called alien chromosome segments

from grasses related to wheat that were linked to the phenotypic traits they introduced^[28]. The availability the 21 pseudomolecules in IWGSC RefSeq v1.0 has increased the diagnostic DNA marker capacity for wheat research and breeding. The genome sequence provides for the positioning of 504 simple sequence repeats, 3025 diversity array technologies, 6689 expressed sequence tags, 205807 single nucleotide polymorphisms (SNPs) and 4512979 ISBPs, and establishes a closer link between the genome sequence and genetic loci/genes associated with traits of agronomic importance^[1]. The broader coverage of molecular markers facilitates genome wide association studies across different environments to associate these markers with traits for surveying wheat genetic resources in gene banks and identify new parents for use in cross breeding^[29]. The IWGSC RefSeq v1.0 reference genome can also assign these new markers to haplotype blocks in the genome and increase the resolution of genomic selection so that genomic estimated breeding values for candidate regions can be selected before phenotypic assessments are made^[30]. The genome-based resources and technology thus combine to establish diagnostics for tracking the complex network of genes that impact on yield potential and stability across diverse environments^[29].

The diagnostic DNA sequence approach was being used for stem solidness (*SSt1*^[1]), a trait for conferring resistance to drought stress and insect damage, when it ran into difficulties due to the location of the trait being in a region where wheat assemblies (prior to IWGSC RefSeq v1.0) lacked scaffold ordering and annotation, partial assembly and/or incomplete gene models. In IWGSC RefSeq v1.0, the *SSt1* region was found to contain 160 high confidence genes^[1] including 26 genes showing differential expression between wheat lines with contrasting phenotypes (solid vs. hollow stems). One gene of significance in this study, *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, which allowed diagnostic SNP markers to be physically linked to the *Sst1* phenotype for more efficient screening of the required genotypes.

The use of IWGSC RefSeq v1.0 to solve molecular marker issues in real time will most likely become the standard to further improve the availability of molecular markers in wheat breeding programs.

4.2 Variation in the number of genes in a family

The rice blast locus is a well-studied example of a disease resistance locus where a repetitive gene family confers resistance. The locus in rice on chromosome 8, does not comprise the typical *NLR* (nucleotide-binding leucine-rich repeat) genes but instead consists of germin-like genes (GLPs) that code for proteins with oxalate oxidase domains, in a cluster of 12 GLP coding genes that each

contribute to the final resistance. Loss of function of individual GLP in the cluster quantitatively reduces the resistance phenotype^[31].

The suggestion that the GLP cluster comprises a broad-spectrum, basal mechanism conserved among the Gramineae^[31] has been supported by the indication that at the *Sr2* locus in wheat there exists an array of 10 GLPs^[32] on chromosome 3B in the cultivar Hope and is now considered to represent this locus. The availability of IWGSC RefSeq v1.0 allows other tandem arrays of GLPs to be readily identified, as shown in Fig. 4, for chromosome 5A for associating with disease resistance loci on this arm.

The GLPs do not conform to the usual gene models associated with many disease resistance genes. The gene models that encode proteins with the *NLR* domains are more common and they were a primary point of interest in a detailed comparison between chromosome 2D from CS and the cultivar CH Campala Lr22a^[27]. These authors observed extensive variation in the *NLR* copy number between CS and CH Campala Lr22a including a 786 kb region which had 16 *NLR* genes compared to the equivalent region in CS where 21 kb carried only two *NLR* genes. In a different region CS had 10 *NLR* genes in a 716 kb region compared to only two in an equivalent 39 kb region in CH Campala Lr22a. The break points defining the apparent indels between the two genomes could be precisely assigned to *NLR* genes suggesting recombination events were involved.

The observations of Thind et al.^[27] focused on the type of differences, such as small deletions or duplications as well as allele differences, that may be driving phenotypic

variation between parents and their progeny and emphasized the importance for accurate assemblies to be made available to the wheat research community.

4.3 Fine-scale modification of gene structures

A particularly valuable outcome of the release of IWGSC RefSeq v1.0 is that the wheat genome reference sequence can take its place alongside other cereal genome sequences so that, as a group, the data sets can be cross-checked and cross-referenced to annotate genes and provide a route to trait enhancement using reverse genetics. An example in IWGSC^[1] related to flowering time, which is so significant for the adaptation of wheat^[34,35] to diverse environments and is also well studied in model plants. IWGSC RefSeq v1.0 has been used^[1] to refine the annotation of six homeologs identified by Sharma et al.^[34] for the *FLOWERING LOCUS C* gene and four high confidence genes could be identified in the reference genome. These were expected to have a role in the vernalization response and hence flowering time. To truncate one of these genes (*TaAGL33*), CRISPR-Cas9-based gene editing, was performed for the respective genes on subgenomes A, B and D. Although expression patterns of the genes were not strongly affected by the genome edits, the edits on *TaAGL33* in the D genome resulted in plants flowering 2–3 days earlier. The outputs of this research were argued to “exemplify how IWGSC RefSeq v1.0 could accelerate the development of diagnostic markers and the design of targets for genome editing for traits relevant to breeding”^[1].

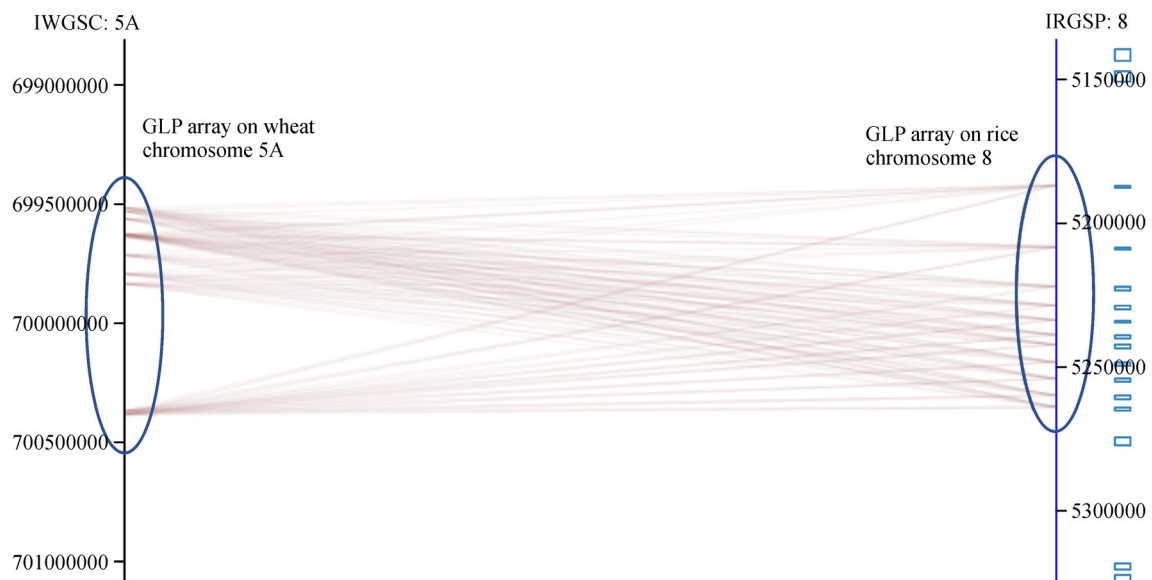


Fig. 4 Synteny alignment of a well-known rice blast locus on rice chromosome 8 to a new location on the long arm of wheat chromosome 5A. The synteny based alignment was performed with Pretzel software^[33].

4.4 Predicting new targets for complex traits

Extensive transcriptome studies^[1,2] have established a framework of gene transcripts (covering 88% of the gene models) comprising 78 modules and representing co-expression networks from non-stressed grain, root, leaf and spike tissues. The module-partitioned transcriptome was then used to identify new genes belonging to the flowering time pathway through co-location with associated regulators that grouped into specific modules. Genes, such as *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. The integrating genes from the flowering time pathway with *VRN1*, *FUL2* and *FUL3* were found in modules 8 and 11 and were most highly correlated with expression in spikes. Partitioning of key elements in the flowering time pathway was then used to examine a large group of transcription factors (MADS-II transcription factor family) and it was evident that 24 of 118 MADS-II gene models were in modules 8 and 11. Notably, none of the 24 MADS-II genes had a straight forward relationship with an equivalent gene in *Arabidopsis*, suggesting that new functions within the flower time pathway for these genes had evolved.

The capacity to use the module-based transcriptome framework in association-type analysis for studying the complex traits provides the basis for understanding “what makes wheat wheat” as well as defining novel pathways that enable researchers and breeders to refine the knowledge of wheat lines to be used as parents in crossing programs or CRISPR-Cas9-based gene editing.

5 Concluding comments

As with all other genome assembly projects, the reference wheat genome (IWGSC RefSeq v1.0) is an ongoing project to close gaps and confirm gene annotations as well as to assign functions to gene models. The key driver for wheat is to ensure that this crop continues to provide an efficient and sustainable use of agricultural resources and reduce risk factors in production systems and processing industries. Developing wheat cultivars to better suit their production environments by targeting specific environmental stresses, such as early morning frost linked to rapid warming of the crop during the day, is critical because such stresses can cause severe yield loss and threaten food security. The wheat research community can now use outputs for barley, maize and rice to speed up defining genome targets for choosing better parents in crossing programs. In the food processing industry, the genome provides the basis for new diagnostics that can track contaminants, point of origin and composition of products.

Acknowledgements The author is grateful to colleagues in the wheat community for their stimulating insights into the wheat genome.

Compliance with ethics guidelines Rudi Appels declares that there is no conflict of interest or financial conflict to disclose.

This article is a review and does not contain any studies with human or animal subjects performed by the author.

References

1. The International Wheat Genome Sequencing Consortium (IWGSC), Appels R, Eversole K, Stein N, Feuillet C, Keller B, Rogers J, Pozniak C J, Choulet F, Distelfeld A, Eversole K, Poland J, Ronen G, Sharpe A G, Barad O, Baruch K, Keeble-Gagnère G, Mascher M, Ben-Zvi G, Josselin A A, Himmelbach A, Balfourier F, Gutierrez-Gonzalez J, Hayden M, Koh C, Muehlbauer G, Pasam R K, Paux E, Rigault P, Tibbits J, Tiwari V, Spannagl M, Lang D, Gundlach H, Haberer G, Mayer K F X, Ormanbekova D, Prade V, Šimková H, Wicker T, Swarbreck D, Rimbart H, Felder M, Guilhot N, Kaithakottil G, Keilwagen J, Leroy P, Lux T, Twardziok S, Venturini L, Juhász A, Abrouk M, Fischer I, Uauy C, Borrill P, Ramirez-Gonzalez R H, Arnaud D, Chalabi S, Chalhoub B, Cory A, Datla R, Davey M W, Jacobs J, Robinson S J, Steuernagel B, van Ex F, Wulff B B H, Benhamed M, Bendahmane A, Concia L, Latrasse D, Bartoš J, Bellec A, Berges H, Doležel J, Frenkel Z, Gill B, Korol A, Letellier T, Olsen O A, Singh K, Valárik M, van der Vossen E, Vautrin S, Weining S, Fahima T, Glikson V, Raats D, Čihalíková J, Toegelová H, Vrána J, Sourdille P, Darrier B, Barabaschi D, Cattivelli L, Hernandez P, Galvez S, Budak H, Jones J D G, Witek K, Yu G, Small I, Melonek J, Zhou R, Belova T, Kanyuka K, King R, Nilsen K, Walkowiak S, Cuthbert R, Knox R, Wiebe K, Xiang D, Rohde A, Golds T, Čížková J, Akpinar B A, Biyiklioglu S, Gao L, N'Daiye A, Kubaláková M, Šafář J, Alfama F, Adam-Blondon A F, Flores R, Guerche C, Loaec M, Quesneville H, Condie J, Ens J, Maclachlan R, Tan Y, Alberti A, Aury J M, Barbe V, Couloux A, Cruaud C, Labadie K, Mangenot S, Wincker P, Kaur G, Luo M, Sehgal S, Chhuneja P, Gupta O P, Jindal S, Kaur P, Malik P, Sharma P, Yadav B, Singh N K, Khurana J P, Chaudhary C, Khurana P, Kumar V, Mahato A, Mathur S, Sevanthi A, Sharma N, Tomar R S, Holušová K, Plíhal O, Clark M D, Heavens D, Kettleborough G, Wright J, Balcárková B, Hu Y, Salina E, Ravin N, Skryabin K, Beletsky A, Kadnikov V, Mardanov A, Nesterov M, Rakitin A, Sergeeva E, Handa H, Kanamori H, Katagiri S, Kobayashi F, Nasuda S, Tanaka T, Wu J, Cattonaro F, Jiumeng M, Kugler K, Pfeifer M, Sandve S, Xun X, Zhan B, Batley J, Bayer P E, Edwards D, Hayashi S, Tulpová Z, Visendi P, Cui L, Du X, Feng K, Nie X, Tong W, Wang L. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, 2018, **361** (6403): 361–374
2. Ramírez-González R H, Borrill P, Lang D, Harrington S A, Brinton J, Venturini L, Davey M, Jacobs J, van Ex F, Pasha A, Khedikar Y, Robinson S J, Cory A T, Florio T, Concia L, Juery C, Schoonbeek H, Steuernagel B, Xiang D, Ridout C J, Chalhoub B, Mayer K F X, Benhamed M, Latrasse D, Bendahmane A, International Wheat Genome Sequencing Consortium, Wulff B B H, Appels R, Tiwari V, Datla R, Choulet F, Pozniak C J, Provart N J, Sharpe A G, Paux E, Spannagl M, Bräutigam A, Uauy C. The transcriptional landscape of polyploid wheat. *Science*, 2018, **361**(6403): eaar6089

3. Juhász A, Belova T, Florides C G, Maulis C, Fischer I, Gell G, Birinyi Z, Ong J, Keeble-Gagnère G, Maharajan A, Ma W, Gibson P, Jia J, Lang D, Mayer K F X, Spannagl M, Tye-Din J A, Appels R, Olsen O A. Genome mapping of seed-borne allergens and immunoresponsive proteins in wheat. *Science Advances*, 2018, **4** (8): eaar8602
4. Keeble-Gagnère G, Rigault P, Tibbits J, Pasam R, Hayden M, Forrest K, Frenkel Z, Korol A, Huang B E, Cavanagh C, Taylor J, Abrouk M, Sharpe A, Konkin D, Sourdille P, Darrier B, Choulet F, Bernard A, Rochfort S, Dimech A, Watson-Haigh N, Baumann U, Eckermann P, Fleury D, Juhász A, Boisvert S, Nolin M A, Doležel J, Šimková H, Toegelová H, Šafář J, Luo M C, Câmara F, Pfeifer M, Isdale D, Nyström-Persson J, Iwagc, Koo D H, Tinning M, Cui D, Ru Z, Appels R. Optical and physical mapping with local finishing enables megabase-scale resolution of agronomically important regions in the wheat genome. *Genome Biology*, 2018, **19**(1): 112
5. Thind A K, Wicker T, Müller T, Ackermann P M, Steuernagel B, Wulff B B H, Spannagl M, Twardziok S O, Felder M, Lux T, Mayer K F X, Keller B, Krattinger S G. Chromosome-scale comparative sequence analysis unravels molecular mechanisms of genome dynamics between two wheat cultivars. *Genome Biology*, 2018, **19**(1): 104
6. Wicker T, Gundlach H, Spannagl M, Uauy C, Borrill P, Ramirez-González R H, de Oliveira R, Mayer K F X, Paux E, Choulet F. Impact of transposable elements on genome structure and evolution in bread wheat. *Genome Biology*, 2018, **19**(1): 103
7. Alaux M, Rogers J, Letellier T, Flores R, Alfama F, Pommier C, Mohellibi N, Durand S, Kimmel E, Michotey C, Guerche C, Loaec M, Lainé M, Steinbach D, Choulet F, Rimbart H, Leroy P, Guilhot N, Salse J, Feuillet C, Paux E, Eversole K, Adam-Blondon A F, Quesneville H. Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data. *Genome Biology*, 2018, **19**(1): 111
8. Sears E R. Misdivision of univalents in common wheat. *Chromosoma*, 1952, **4**(6): 535–550
9. Sears E R. Chromosome mapping with the aid of telocentrics. *Hereditas*, 1966, **2**: 370–381
10. Sears E R, Miller T. The history of Chinese Spring wheat. *Cereal Research Communications*, 1985, **13**: 261–263
11. Liu D, Zhang L, Hao M, Ning S, Yuan Z, Shoufen Dai S, Huang L, Wu B, Yan Z, Lan X, Zheng Y. Wheat breeding in the hometown of Chinese Spring. *Crop Journal*, 2018, **6**(1): 82–90
12. Gill B S, Appels R, Botha-Oberholster A M, Buell C R, Bennetzen J L, Chalhou B, Chumley F, Dvorák J, Iwanaga M, Keller B, Li W, McCombie W R, Ogihara Y, Quetier F, Sasaki T. A workshop report on wheat genome sequencing: International Genome Research on Wheat Consortium. *Genetics*, 2004, **168**(2): 1087–1096
13. Doležel J, Doleželová M, Suchánková, P, Šafář J, Kovářová P, Bartoš J, Čihalíková J, Šimková H. Flow cytogenetic analysis of the wheat genome. *Frontiers of Wheat Bioscience*, 2005, **Memorial Issue** (Wheat Information Service No.100): 3–15
14. Clavijo B J, Venturini L, Schudoma C, Accinelli G G, Kaithakottil G, Wright J, Borrill P, Kettleborough G, Heavens D, Chapman H, Lipscombe J, Barker T, Lu F H, McKenzie N, Raats D, Ramirez-Gonzalez R H, Coince A, Peel N, Percival-Alwyn L, Duncan O, Trösch J, Yu G, Bolser D M, Namaati G, Kerhornou A, Spannagl M, Gundlach H, Haberer G, Davey R P, Fosker C, Palma F D, Phillips A L, Millar A H, Kersey P J, Uauy C, Krasileva K V, Swarbreck D, Bevan M W, Clark M D. 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*, 2017, **27**(5): 885–896
15. Zimin A V, Puiu D, Hall R, Kingan S, Clavijo B J, Salzberg S L. The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum*. *GigaScience*, 2017, **6**(11): 1–7
16. Ogihara Y. Genome science of polyploid wheat. *Wheat Information Service*, 2005, **100**: 169–184
17. Appels R, Nyström-Persson J, Keeble-Gagnère G. Advances in genome studies in plants and animals. *Functional & Integrative Genomics*, 2014, **14**(1): 1–9
18. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature Genetics*, 2006, **38**(11): 1348–1354
19. Lee E, Helt G A, Reese J T, Munoz-Torres M C, Childers C P, Buels R M, Stein L, Holmes I H, Elsik C G, Lewis S E. Web Apollo: a web-based genomic annotation editing platform. *Genome Biology*, 2013, **14**(8): R93
20. The International Wheat Genome Sequencing Consortium (IWGSC), Mayer K F X, Rogers J, Doležel J, Pozniak C, Eversole K, Feuillet C, Gill B, Friebe B, Lukaszewski A J, Sourdille P, Endo T R, Kubaláková M, Čihalíková J, Dubsá Z, Vrána J, Sperková R, Šimková H, Febrer M, Clissold L, McLay K, Singh K, Chhuneja P, Singh N K, Khurana J, Akhunov E, Choulet F, Alberti A, Barbe V, Wincker P, Kanamori H, Kobayashi F, Itoh T, Matsumoto T, Sakai H, Tanaka T, Wu J, Ogihara Y, Handa H, Maclachlan P R, Sharpe A, Klassen D, Edwards D, Batley J, Olsen O A, Sandve S R, Lien S, Steuernagel B, Wulff B, Caccamo M, Ayling S, Ramirez-Gonzalez R H, Clavijo B J, Wright J, Pfeifer M, Spannagl M, Martis M M, Mascher M, Chapman J, Poland J A, Scholz U, Barry K, Waugh R, Rokhsar D S, Muehlbauer G J, Stein N, Gundlach H, Zytnicki M, Jamilloux V, Quesneville H, Wicker T, Faccioli P, Colaiacovo M, Stanca A M, Budak H, Cattivelli L, Glover N, Pingault L, Paux E, Sharma S, Appels R, Bellgard M, Chapman B, Nussbaumer T, Bader K C, Rimbart H, Wang S, Knox R, Kilian A, Alaux M, Alfama F, Couderc L, Guilhot N, Viseux C, Loaec M, Keller B, Praud S. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*, 2014, **345** (6194): 1251788
21. Ogbonnaya F C, Halloran G M, Lagudah E S. D genome of wheat: 60 years on from Kihara, Sears and McFadden. *Wheat Information Service*, 2005, **100**: 205–220
22. Bromilow S, Gethings L A, Buckley M, Bromley M, Shewry P R, Langridge J I, Clare Mills E N. A curated gluten protein sequence database to support development of proteomics methods for determination of gluten in gluten-free foods. *Journal of Proteomics*, 2017, **163**: 67–75
23. Altenbach S B, Chang H C, Simon-Buss A, Jang Y R, Denery-Papini S, Pineau F, Gu Y Q, Huo N, Lim S H, Kang C S, Lee J Y. Towards reducing the immunogenic potential of wheat flour: omega

- gliadins encoded by the D genome of hexaploid wheat may also harbor epitopes for the serious food allergy WDEIA. *BMC Plant Biology*, 2018, **18**(1): 291
24. Kawaura K, Miura M, Kamei Y, Ikeda T M, Ogihara Y. Molecular characterization of gliadins of Chinese Spring wheat in relation to celiac disease elicitors. *Genes & Genetic Systems*, 2018, **93**(1): 9–20
 25. Zhao X C, Batey I L, Sharp P J, Crosbie G, Barclay I, Wilson R, Morell M K, Appels R. A single genetic locus associated with starch granule and noodle quality in wheat. *Journal of Cereal Science*, 1998, **27**(1): 7–13
 26. Wicker T, Gundlach H, Spannagl M, Uauy C, Borrill P, Ramírez-González R H, de Oliveira R, Mayer K F X, Paux E, Choulet F. Impact of transposable elements on genome structure and evolution in bread wheat. *Genome Biology*, 2018, **19**(1): 103
 27. Thind A K, Wicker T, Müller T, Ackermann P M, Steuernagel B, Wulff B B H, Spannagl M, Twardziok S O, Felder M, Lux T, Mayer K F X, Keller B, Krattinger S G. Chromosome-scale comparative sequence analysis unravels molecular mechanisms of genome dynamics between two wheat cultivars. *Genome Biology*, 2018, **19**(1): 104
 28. Mukai Y. Perspectives in molecular cytogenetics of wheat. *Wheat Information Service*, 2005, **100**: 17–32
 29. Rasheed A, Hao Y, Xia X, Khan A, Xu Y, Varshney R K, He Z. Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. *Molecular Plant*, 2017, **10**(8): 1047–1064
 30. Poland J, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, Dreisigacker S, Crossa J, Sánchez-Villeda H, Sorrells M, Jannink J L. Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome*, 2012, **5**(3): 103–113
 31. Manosalva P M, Davidson R M, Liu B, Zhu X, Hulbert S H, Leung H, Leach J E. A germin-like protein gene family functions as a complex quantitative trait locus conferring broad-spectrum disease resistance in rice. *Plant Physiology*, 2009, **149**(1): 286–296
 32. Mago R, Tabe L, Vautrin S, Šimková H, Kubaláková M, Upadhyaya N, Berges H, Kong X, Breen J, Doležel J, Appels R, Ellis J G, Spielmeyer W, Spielmeyer W. Major haplotype divergence including multiple germin-like protein genes, at the wheat Sr2 adult plant stem rust resistance locus. *BMC Plant Biology*, 2014, **14**(1): 379
 33. Keeble-Gagnere G, Isdale D, Suchecki R, Kruger A, Lomas K, Carroll D, Li S, Whan A, Hayden M, Tibbits J. Integrating past, present and future wheat research with Pretzel. *bioRxiv*, 2019 (preprint). doi:10.1101/517953
 34. Sharma N, Ruelens P, D'hauw M, Maggen T, Dochy N, Torfs S, Kaufmann K, Rohde A, Geuten K. A flowering locus C homolog is a vernalization-regulated repressor in *Brachypodium* and is cold regulated in wheat. *Plant Physiology*, 2017, **173**(2): 1301–1315
 35. Shaw L M, Lyu B, Turner R, Li C, Chen F, Han X, Fu D, Dubcovsky J. FLOWERING LOCUS T2 regulates spike development and fertility in temperate cereals. *Journal of Experimental Botany*, 2019, **70**(1): 193–204