REVIEW

Reference Genome Sequencing and Advances in Genomic Resources in Common Wheat–Chromosome 6B Project in Japan

Fuminori KOBAYASHI[1]*, Tsuyoshi TANAKA[1], Hiroyuki KANAMORI[1], Jianzhong WU[1] and Hirokazu HANDA[1, 2]

[1] Institute of Crop Science, National Agriculture and Food Research Organization, Tsukuba, Japan
[2] Graduate School of Life and Environmental Sciences, Kyoto Prefectural University, Kyoto, Japan

Abstract

By successfully sequencing the entire genome of common wheat (*Triticum aestivum* L.) in 2018, the International Wheat Genome Sequencing Consortium (IWGSC) achieved its goal of publishing the first reference genome sequence of this important crop. During this period, various technological innovations have dramatically evolved genome sequencing technology; these advances have made it possible to rapidly decode large and complex genomes, including that of polyploid wheat. IWGSC completed reference genome sequence by exploiting these latest technologies via two primary approaches, "whole-genome sequencing" and "sequencing by chromosome." As collaborators in the IWGSC project, a Japanese research team led by the National Agriculture and Food Research Organization was responsible for sequencing chromosome 6B, one of the 21 wheat chromosomes, and contributed to constructing the reference genome sequence. This article presents an outline of reference genome sequence construction and provides an overview of the common wheat genome information obtained. Moreover, it describes the methods used to sequence chromosome 6B, together with the associated analytical data and novel knowledge, including the structural analyses of *Nor-B2* and *Gli-B2* obtained during this work. Furthermore, it describes the recently undertaken sequencing of the genomes of diploid and tetraploid wheat and reviews the overall development of wheat genomic resources.

Introduction

Genome sequence data have become indispensable tools in promoting genetic improvements in crop varieties to achieve better quality, higher yield, adaptation to different environments, and tolerance to biotic stresses. In this regard, information on the rice (*Oryza sativa* L. subsp. *japonica* cv. Nipponbare) genome published in 2004 (IRGSP 2005) paved the way for genome-based breeding. Both basic and applied research on rice have undergone major advances as a consequence of developments in breeding technologies such as quantitative trait locus analysis, gene isolation, the

molecular marker design, and genetic resources development. Moreover, a range of related experimental and bioinformatics data have been released and stored, including primary databases for genome sequences and annotation data (Ouyang et al. 2007, Sakai et al. 2013), followed by secondary databases for omics data, including those of transcriptome studies (Kawahara et al. 2016, Sato et al. 2013). Databases for rice and other crop plants have been developed and can be mined for comparative analyses among plants (Rouard et al. 2011, Tello-Ruiz et al. 2016). This wide range of information plays an important role as a driving force for plant research, not only in rice breeding but also in basic

research at the whole-plant level. Following the ground-breaking sequencing of the rice genome, the motive to decode the genome sequences of other crop species gained momentum worldwide. It was not long before the genome sequences of sorghum (*Sorghum bicolor* L.), maize (*Zea mays* L.), and soybean (*Glycine max* L.) were published (Paterson et al. 2009, Schmutz et al. 2010, Schnable et al. 2009). Furthermore, genome sequencing of *Brachypodium distachyon* (L.), belonging to the same Pooideae subfamily as wheat and barley, has been accomplished (The International Brachypodium Initiative 2010).

Although a diploid plant has a similar genome structure to wheat, sequencing of the genome of barley (*Holdeum vulgare* L.) commenced with establishing the International Barley Genome Sequencing Consortium (IBSC) in 2006. The estimated size of the barley genome is 5.1 Gb, larger than that of most other crops sequenced to date. Although sequencing of this genome presented considerable challenges, a draft genome was published in 2012 (IBSC 2012). The IBSC sequenced the genome of the malting barley cultivar "Morex," the physical map of which comprised 67,000 bacterial artificial chromosome (BAC) clones and 9,265 BAC contigs, covering 4.98 Gb of the barley genome, and 1.9 Gb of sequence data were generated from a deep whole-genome shotgun assembly. This genome sequence was further refined in 2017 (Mascher et al. 2017), based on the use of new technologies such as POPSEQ (Mascher et al. 2013) and Hi-C (Lieberman-Aiden et al. 2009), which facilitated the acquisition of more sophisticated genome information. The revised assembly was released in 2019 using a new *de novo* assembly pipeline, TRITEX (Monat et al. 2019).

Given that common wheat (*Triticum aestivum* L.) is an important crop globally, sequencing of the wheat genome has long been awaited, primarily regarding breeding programs for wheat improvement. However, owing to the complexity of the wheat genomic structure, such as polyploidy (common wheat is allo-hexaploid with the genome formula AABBDD) and large genome size (17 Gb), sequencing transpired to be a relatively protracted endeavor compared with that of other crops. In 2005, the International Wheat Genome Sequencing Consortium (IWGSC) was founded to sequence the common wheat cultivar "Chinese Spring" (CS) (http://www.wheatgenome.org/). As a member of IWGSC, a Japanese research team led by the National Agriculture and Food Research Organization contributed to the project. Following the initial publication of a chromosome-based draft genome sequence in 2014 (IWGSC 2014), the IWGSC released a complete genome assembly of 21 chromosomes in 2018 (IWGSC 2018). This was the first

high-precision reference genome sequence for common wheat and is currently used by many researchers globally for a wide range of research objectives, from basic to applied. From the inception of the IWGSC to the release of a complete genome assembly, researchers developed and utilized a range of innovative technologies for genome sequencing, which have made it possible to achieve highly accurate genome sequencing, even for complex genome structures such as that of common wheat.

Within the framework of the IWGSC project, a Japanese group, including our group, was assigned the task of sequencing chromosome 6B, which is one of the largest chromosome among the 21 chromosomes of wheat. Chromosome 6B has several interesting features from basic and applied aspects, such as secondary structures that serve as a site for nucleolus organizer region (NOR), designated *Nor-B2,* and the location of the *Gli-B2* locus for α/β-gliadins, major components of the seed storage proteins and important factors for the quality of wheat flour.

In this review, we outline the approaches and outcomes of the IWGSC genome sequencing project, focusing on the sequencing of chromosome 6B and related research, which our research group worked on as part of the IWGSC initiative.

Genome assemblies of common wheat produced by the IWGSC

The most distinctive feature of the common wheat genome is its large size, estimated to be 17 Gb (Šafář et al. 2010). This is ~130 times larger than that of *Arabidopsis thaliana* (135 Mb), the whole genome of which was the first to be sequenced among higher plants, and 44 times larger than that of rice (389 Mb). The size of the entire *Arabidopsis* genome is but a fraction of that of a wheat chromosome, whereas the entire rice genome would fit within a single arm of a wheat chromosome. Note that if the 17 Gb of the common wheat genome was sequenced at the speed at which the whole rice genome was sequenced, it would have taken 350 y. Accordingly, the initial research strategy adopted by the IWGSC was the use of next-generation sequencing (NGS) technology. An important challenge in this project was that the chromosomal arms were separated by flow cytometry using double-ditelosomic lines of CS for each chromosome as the source material (Doležel et al. 2007), and the sequencing was performed separately for each chromosomal arm. However, although this approach benefits from reducing the complexity of the genome structure and enhancing the accuracy of analysis, even if

the wheat genome is divided into chromosomal arms, these still have large sizes ranging from 230 to 580 Mb, far exceeding the size of the *Arabidopsis* genome. Accordingly, using this approach, it would have been extremely difficult to sequence the common wheat genome without the availability of NGS technology, which can facilitate the production of large-scale data. Before the construction of the reference genome sequence, the IWGSC conducted a short-read NGS (Illumina) analysis for each chromosome arm and published a survey sequence comprising all chromosomes (IWGSC 2014). This survey sequence covered 10.2 Gb, equivalent to 61% of the 17 Gb of the common wheat genome, wherein 124,201 putative loci were identified with high confidence, which is similar to the 32,000-38,000 loci per genome that have been identified in other diploid plants. The survey sequence is considered to cover the genic regions of common wheat. It was assembled based on the separate sequencing of the individual chromosome. The sequences of the three homoeologous genomes were distinguishable. However, the survey sequence had poor coverage of 61%. The sequences of the individually assembled contigs were only a few kilobases in size, and the positional relationships of individual contigs on a given chromosome were unclear. Hence, the sequence did not reach the level of completeness obtained for other sequenced reference genomes, such as that of rice.

To obtain a higher quality wheat genome assembly, the IWGSC introduced new technology to the project, referred to as DeNovoMAGIC (NRGene, Ness-Ziona, Israel), which can be used to substantially accelerate the whole-genome sequencing of challenging genomes, such as that of polyploid wheat, as well as enhancing the quality of the assembled genome. With a DeNovoMAGIC 2 pipeline, the IWGSC constructed a draft *de novo* whole-genome assembly (WGA) from Illumina short reads using the whole genome of CS as a material, which was integrated into pseudomolecule sequences representing the 21 chromosomes of common wheat, with additional data including the survey sequence of each chromosome arm. All unanchored scaffolds were clustered as an additional pseudomolecule (ChrUn). The resulting construct was released as IWGSC RefSeq v1.0, a comprehensive chromosome-level genome assembly of 14.5 Gb covering 94% of the common wheat genome (based on the new genome size estimates of 15.4 to 15.8 Gb according to IWGSC [2018]) and contains 107,891 high-confidence gene models (IWGSC 2018). Unlike the survey sequence that was finely divided, this reference sequence consists of the pseudomolecules for 21 wheat chromosomes. Therefore, the positional relationship of

genes within the chromosome can be discerned with greater clarity and is extremely accurate. This information is available at "Wheat URGI" (http://wheat-urgi. versailles.inrae.fr/) or "Ensemble Plants" (https://plants. ensembl.org/Triticum_aestivum/Info/Index). As with barley, POPSEQ and Hi-C data were integrated for the construction of RefSeq v1.0, and the quality of genome assembly was assessed through alignment with a radiation hybrid map and high-resolution genetic map, which consists of genotyping-by-sequencing (GBS)-derived genetic markers and insertion site-based polymorphism (ISBP) markers (IWGSC 2018).

Chromosome arm-based genomic resources

Although originally generated to construct pseudomolecules for individual chromosome arms, a BAC-based physical map and sequence assemblies were also used to construct RefSeq v1.0 (IWGSC 2018). The IWGSC originally adopted a BAC-by-BAC sequencing approach using NGS technology for each chromosome arm. The sorting of single chromosomes or chromosome arms from CS and its aneuploid lines enabled the construction of chromosome (arm)-specific BAC libraries (Šafář et al. 2010). Finally, 43 such BAC libraries were constructed, consisting of 38 libraries for each chromosome arm (including duplication of some chromosome arms), two libraries for the entirety of chromosome 3B, and three libraries for a pool of chromosomes 1D/4D/6D. These BAC libraries are available from CNRGV, INRAe (https://cnrgv.toulouse. inrae.fr/en/Library/Wheat).

BAC libraries have served as essential resources for the development of physical maps and map-based genome sequencing. Construction of the physical maps of common wheat chromosomes and chromosome arms involved groups in 16 countries and one company participating in the physical mapping projects of IWGSC (https://www.wheatgenome.org/Projects/IWGSC-Bread-Wheat-Projects/Physical-mapping). Physical maps for chromosome 3B and 20 chromosome arms were successfully constructed and published (Table 1). The genome sequences of the minimal-tiling path (MTP) clones that comprise the physical map of chromosome 3B, the largest chromosome in wheat, were deciphered, and the pseudomolecule on a wheat chromosome was constructed for the first time by a French group (Choulet et al. 2014). Furthermore, BAC MTP sequences of chromosome arms 3DL, 4AL, and 5BS and chromosome 7A were assembled and released (Table 1).

In addition to BAC-by-BAC sequencing, shotgun sequence analyses for the flow-sorted chromosomal

DNAs of chromosomes 1A, 1B, 1D, 4A, 4D, 5A, 5B, 5D, and 6B, using a Roche 454 FLX Titanium sequencer (Roche, CT, USA), and chromosome arms 7BS and 7DS, using an Illumina sequencer, have been performed. Table 1 presents the details.

Genome sequencing project for chromosome 6B

Collaborating with IWGSC, our group tackled a genome sequencing of chromosome 6B, with financial support from the Japanese Ministry of Agriculture, Forestry and Fisheries ("Genomics for Agricultural Innovation" project and "Genomics-based Technology for Agricultural Improvement" project) and Nisshin Flour Milling Inc. This project had two primary approaches: (1) whole-chromosome shotgun sequencing and (2) BAC-based physical mapping and genomic sequencing. The size of chromosome 6B is estimated to be 914 Mb, of which the short arm (6BS) and long arm (6BL) account for 415 Mb and 498 Mb, respectively (Šafář et al. 2010). We used chromosome arm-specific DNA isolated from the flow-sorted 6B mitotic chromosome arms of the double-ditelosomic lines to reduce the complexity of the genome for the construction of physical maps and genomic sequencing (Tanaka et al. 2014).

1. Whole-chromosome shotgun sequencing

Samples of DNA derived from 6BS and 6BL were independently sequenced using a 454 GS-FLX Plus

Table 1. Publication of chromosome and chromosome arm-based genomic resources of common wheat

Chromosome, chromosome arm	Shotgun sequencing	BAC physical mapping	MTP sequencing
1AS		Breen et al. (2013) (Switzerland)	-
1AL	Wicker et al. (2011) (Switzerland, Germany, Czech Republic)	Lucas et al. (2013) (Turkey)	-
1BS		Raats et al. (2013) (Israel)	-
1BL		Philippe et al. (2013) (France)	-
1D		-	-
3AS	-	Sehgal et al. (2012) (USA)	-
3B	-	Paux et al. (2008) (France)	Choulet et al. (2014) (France)
3DS	-	Holušová et al. (2017) (Czech Republic)	-
3DL	-	Lu et al. (2018) (UK)	Lu et al. (2018) (UK)
4AS	Hernandez et al. (2012) (Czech Republic)	-	-
4AL		-	Shorinola et al. (2017) (Czech Republic)
4D	Helguera et al. (2015) (Argentina)	-	-
5A	Vitulo et al. (2011) (Italy)	Barabaschi et al. (2015) (Italy)	-
5BS	Sergeeva et al. (2014) (Russia)	Salina et al. (2018) (Russia)	Nesterov et al. (2016) Sergeeva et al. (2017) (Russia)
5BL		-	-
5D	Lucas et al. (2014) (Turkey)	Akpinar et al. (2015) (Turkey)	-
6A	-	Poursarebani et al. (2014) (Germany)	-
6B	Tanaka et al. (2014) (Japan)	Kobayashi et al. (2015) (Japan)	Kobayashi et al. (this review) (Japan)
7A	-	Keeble-Gagnère et al. (2018) (Australia)	Keeble-Gagnère et al. (2018) (Australia)
7BS	Berkman et al. (2012) (Australia)	Belova et al. (2014) (Norway)	-
7BL	-		-
7DS	Berkman et al. (2011) (Australia)	Staňková et al. (2016) Tulpová et al. (2019) (Czech Republic)	-

sequencer. From the 4.94 and 5.51 Gb shotgun sequence data for 6BS and 6BL, 235 and 273 Mb sequences were assembled to cover 56.6% and 54.9% of the estimated length of each arm, respectively (Tanaka et al. 2014). Repetitive sequences were found to account for 76.6% and 85.5% of the assembled sequences of 6BS and 6BL, respectively, and ~13% of repetitive regions comprised novel repeat elements. We succeeded in detecting the transcribed genes, tRNA genes, miRNAs, and rDNA sequences, thereby providing a broad overview of the sequence features of chromosome 6B. Although these genomic sequences were fragmented and did not completely cover the entire chromosome 6B, we identified 16,728 simple sequence repeats (SSRs) in the non-repetitive region and were accordingly able to efficiently generate new SSR markers, which will be useful for various aspects of genetic and genomic research, including BAC-based physical map construction, as described in the next section.

2. BAC-based physical mapping and genomic sequencing

More than five million chromosome arms were corrected by flow sorting for the construction of chromosome arm-specific BAC libraries. Using the chromosomal DNA extracted from each arm, two BAC libraries were constructed for 6BS and 6BL, which comprised 57,600 and 76,032 BACs with an average insert length of ~130 kb, representing 15.3 and 18 times equivalents of their estimated sizes, respectively (Kobayashi et al. 2015). BAC clones were fingerprinted using the whole-genome profiling (WGP™) method to construct the physical map. WGP™ is an alternative method for establishing chromosome physical maps developed with reference NGS-based technology (van Oeveren et al. 2011). WGP™ tags were used not only for BAC fingerprinting but also for super-scaffolding to construct RefSeq v1.0 (IWGSC 2018). Using FingerPrinted Contigs (FPC) software (Soderlund et al. 1997), the BAC contigs for the 6BS and 6BL were successfully assembled, representing 91% of the entire physical length of chromosome 6B (Kobayashi et al. 2015).

The constructed BAC contigs were integrated into a radiation hybrid (RH) map of chromosome 6B, which was a vital step in determining the physical location of the BAC contigs on this chromosome. RH mapping is a powerful tool for high-resolution mapping in wheat (Tiwari et al. 2012). For chromosome 6B assembly, we produced an RH panel of 355 lines based on a cross between the Nullisomic-6B Tetrasomic-6A (N6BT6A) line and freshly γ-irradiated pollen of CS (Watanabe et al. 2014). DNA markers are also essential for anchoring BAC contigs to their specific genomic positions, and thus far, a total of 8,313 markers including 208 publicly genetic markers, 751 PLUG markers, and 7,354 newly developed markers (Table 2), updated from the report of Kobayashi et al. (2015), have been developed and subsequently validated for quality and effectiveness. Hence, we were able to localize 3,166 markers on individual BAC clones (Table 2). Thus, we established an RH map of a single linkage group of total length 2,537.1 cR, consisting of 970 markers (681 loci), wherein 2,069 obligate breaks were identified.

Analysis of the overlap among neighboring clones within each BAC contig using the FPC MTP module enabled us to select 8,130 MTP clones, providing a basis for map-based genomic sequencing of the chromosome 6B. Regarding these MTP clones, an individually tagged pooled library and mate-pair library (3 kb and 8 kb) were prepared and sequenced using the 454 GS-FLX Plus and Illumina platforms, respectively, to construct an assembly, and the sequence data thus obtained were used to align the assembly (scaffolding). In summary, the initial version of the BAC-based 6B genome assembly (version 1.0) comprised 871 BAC contigs consisting of 7,086 clones with a total length of 688 Mb. It comprised 7,635 scaffolds with maximum and average lengths of ~3.14 Mb and 90 kb, respectively (Table 3).
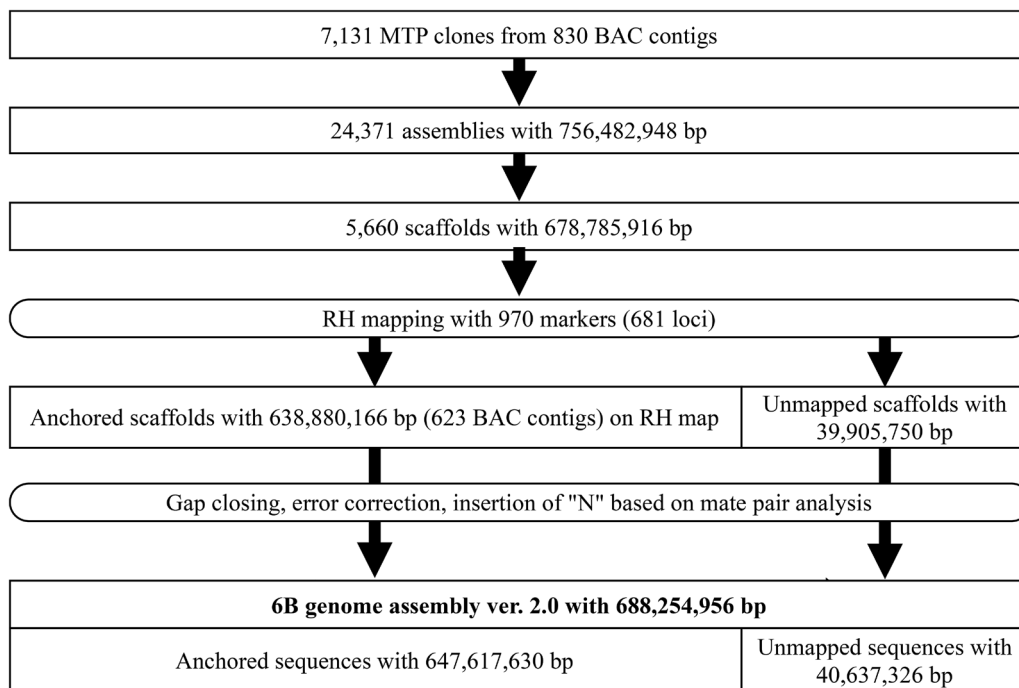
Moreover, the sequence data were used to investigate

Table 2. Molecular markers developed for the physical mapping of chromosome 6B

Marker category	Source	Developed	Anchored to BAC contigs	Anchoring on RH map
Publicly genetic markers	GrainGenes, NBRP-Wheat	208	126	17
PLUG markers	Ishikawa et al. (2007, 2009)	751	366	9
Genic markers	EST	1,643	793	25
ISBP markers	Shotgun and MTP sequences	3,965	1,666	742
SSR markers	Shotgun and MTP sequences	1,476	78	78
Others	Shotgun and MTP sequences	270	137	99
	Total	8,313	3,166	970

Table 3. Scaffolding of BAC-based genome assemblies of chromosome 6B

	Version 1.0	Version 2.0
Number of MTP clones	7,086	7,131
Number of scaffolds	7,635	5,660
Total length (bp)	687,555,957	678,785,916
Average length (bp)	90,053	119,927
Maximum length (bp)	3,140,367	3,998,456
N50 (bp)	459,245	544,679
Number of super-scaffolds	871	830
Average length (bp)	789,353	817,814

7,131 MTP clones from 830 BAC contigs

↓

24,371 assemblies with 756,482,948 bp

↓

5,660 scaffolds with 678,785,916 bp

↓

RH mapping with 970 markers (681 loci)

↓

Anchored scaffolds with 638,880,166 bp (623 BAC contigs) on RH map	Unmapped scaffolds with 39,905,750 bp

Gap closing, error correction, insertion of "N" based on mate pair analysis

↓

6B genome assembly ver. 2.0 with 688,254,956 bp

Anchored sequences with 647,617,630 bp	Unmapped sequences with 40,637,326 bp

Fig. 1. The entire process for construction of the 6B genome assembly from a BAC-based physical map

the overlapping portions between MTP clones. Although we confirmed most overlapping regions (97.2% of the entire overlapped area), there appeared to be certain inconsistencies in the remaining 2.8%. We searched for bridge clones to correct the abnormal contigs and eventually obtained the final BAC contig alignment in chromosome 6B, which consisted of 830 contigs. Thereafter, additional sequencing of the bridge clones and repositioning of scaffolds was performed, and the 6B genome sequence assembly was further refined (version 2.0), yielding a construct of 679 Mb in length consisting of 7,131 clones. The assembly comprised 5,660 scaffolds with maximum and average 4.0 Mb and 120 kb lengths, respectively (Table 3). Although the total length of the sequence had been shortened compared with ver.1.0 due to the deletion of sequences with low reliability, the average length, maximum length, and N50

had all increased, and the quality of scaffolding had been improved (Table 3). Gaps were subsequently filled, and sequencing errors were corrected using the IWGSC survey sequence data (IWGSC 2014). Furthermore, on the basis of the mate-pair library sequencing data, we estimated the gap lengths between individual assemblies and inserted "N" corresponding to these values. Finally, a chromosome 6B genomic sequence assembly with a length of 688 Mb was constructed and designated as "BAC-based 6B genome assembly ver.2.0" (Fig. 1). We subsequently succeeded in adding positional information to 95% (648 Mb) of this new chromosome 6B assembly, based on the results of the RH map (Fig. 1).

Regarding the construction of RefReq v1.0, we provided the IWGSC with fingerprint data of BAC clones (WGP™ tags) and the assembled sequence data of MTP clones. These data were used to correct WGA scaffolds'

positions and orientation and fill gaps between assemblies (IWGSC 2018). The genomic resources of chromosome 6B have been deposited in the database of the Komugi Genome Sequencing Program (https://komugigsp.dna. affrc.go.jp/research/download/index.html).

Application of genomic resources for analyses of the structural features of chromosome 6B

Integration of the physical map and shotgun sequences of chromosome 6B enabled us to characterize several features that are specific to the genomic composition and structure of chromosome 6B (Kobayashi et al. 2015). For example, a comparison of the chromosomal alignment of genes identified from the integrated genomic data with those of other grass genomes revealed local rearrangements on chromosome 6B, indicating that at least five large inversions have occurred during wheat evolution (Kobayashi et al. 2015).

The presence of a nucleolus organizer region (NOR), *Nor-B2*, is one of the characteristic features of chromosome 6B, which is a locus for rRNA genes (rDNA locus) located on the short arm of 6B (Morrison 1953, Pikaard 2000). We identified the rRNA gene sequence in three BAC contigs located on 6BS, and from their positions on the RH map, we were able to estimate the position of the *Nor-B2* locus on the physical map (Kobayashi et al. 2015). Enrichment of the genomic sequence data via the construction of BAC-based genome assemblies and RefSeq v1.0 has enabled us to further our analysis of chromosome 6B. We estimated that the total copy number of the rDNA units in the wheat genome is 11,160, 60.9% of which is located in *Nor-B2* (Handa et al. 2018). The characterization of the region's structural features surrounding *Nor-B2* was performed by comparison of gene positions among the homoeologous chromosomes 6A, 6B, and 6D. We accordingly detected a non-syntenic segment in the region adjacent to *Nor-B2*, which was also identified in *Nor-B1* on chromosome 1B (Handa et al. 2018). Further characterization of the regions surrounding the NOR, such as repeat contents and epigenetic status as well as the expression analysis of rRNA genes, revealed that these regions probably play important roles in autoregulation of the associated major NORs as well as in silencing minor NORs (Handa et al. 2018).

Gliadins are major components of the storage proteins in wheat endosperm and are important for the quality of wheat flour, and their quantity and quality directly affect the properties of wheat gluten. The genes encoding α/β-gliadins are located at the *Gli-2* loci on the short arms of homoeologous group 6 chromosomes (McIntosh et al. 2013). These *Gli-2* loci consist of multiple gene copies with highly similar sequences, among which are numerous pseudogenes (Kawaura et al. 2012). Hence, based on NGS data, these multiple sequences are not reflected in the complete assembly and impede the complete characterization of the *Gli-2* loci. Indeed, details of the *Gli-2* loci remain incomplete, even in RefSeq v1.0 (IWGSC 2018). Our studies have identified the *Gli-B2* locus in the BAC-based genomic assembly of chromosome 6B (Kobayashi et al. 2015). These data and our BAC-based resources have enabled us to identify BAC clones in the *Gli-B2* region for resequencing using long-read NGS. The resulting data were accordingly integrated into the previous assembly to obtain a more accurate sequence for *Gli-B2* (Handa et al. 2019). Characterization of the resequenced *Gli-B2* region is ongoing, and we anticipate that more accurate sequence and gene content information will be available in the near future.

Available wheat genomic resources

The initial version of the wheat WGA (RefSeq v1.0) has subsequently been updated. IWGSC RefSeq v2.1 is now freely available at the IWGSC data repository hosted by URGI-INRAE (https://wheat-urgi.versailles.inrae.fr/ Seq-Repository/Assemblies). For this updated assembly, whole-genome optical mapping and long-read PacBio contigs were used to refine further the positions and orientations of scaffolds, including the anchoring of unassigned scaffolds and gap closure (Zhu et al. 2021). The IWGSC Annotation v2.1 is completed by conducting reannotation of RefSeq v2.1, and also available (https:// wheat-urgi.versailles.inrae.fr/Seq-Repository/ Annotations).

Subsequent to the sequencing of CS, 10 high-precision genome assemblies and five scaffold-level assemblies of common wheat accession have been produced through the 10+ Wheat Genome Project (Walkowiak et al. 2020), an international project inaugurated to provide a foundation for analysis of the genomic variation in wheat (http://www.10wheatgenomes. com). The Japanese research group involved in this project focuses on the evolutionary genome and cytogenetic analyses as well as sequence the common wheat cultivar "Norin 61," a representative Japanese accession. All assemblies and annotations relating to this project are available for download under the Toronto Agreement (https://wheat.ipk-gatersleben.de). The BLAST server is also available at IPK Wheat BLAST Server (https://galaxy-web.ipk-gatersleben.de/), Earlham Institute (https://grassroots.tools/public/service/blast_blastn),

and the National BioResource Project, Japan-Wheat (https://shigen.nig.ac.jp/wheat/komugi/about/norin61GenomeSequence.jsp).

Other than hexaploid wheat, the genome assemblies of tetraploid wheat (genome formula AABB), such as those of wild emmer wheat (*T. dicoccoides*) "Zavitan" (Avni et al. 2017, Zhu et al. 2019) and two durum wheat (*T. durum*) accessions, "Svevo" (Maccaferri et al. 2019) and "Kronos" (Walkowiak et al. 2020), are also available. For diploid wheat, the genome assemblies of two wheat ancestors, wild einkorn wheat (*T. urartu*, A genome progenitor) and *Aegilops tauschii* (D genome progenitor), have similarly been released (Jia et al. 2013, Ling et al. 2013, 2018, Luo et al. 2017, Zhao et al. 2017). Published and publicly available wheat resources and tools as well as their application in research are summarized in detail by Adamski et al. (2020).

Concluding remarks

Despite the large genome size, complexities of polyploidy, and high frequency of repetitive elements, the reference genome sequence of common wheat was completed in 2018. New tools for genome-wide analyses continually enhance the quality of genome sequencing, and annotation data and additional information have been used in several applied studies on wheat, including breeding programs. By exploiting the various technological innovations that have been developed over the course of sequencing wheat and other organisms, it is now possible to construct a whole-genome sequence of desirable varieties, as has been achieved in the 10+ Wheat Genome Project. Nonetheless, NGS technology has certain limitations regarding the combination of whole-genome shotgun sequencing with short-read and high-precision genome assembling technology. These are currently the mainstream approaches. Nevertheless, BAC-based analyses are still effective for sequencing regions characterized by complex structures, such as the *Gli-2* region. Long-read type NGS technologies, including PacBio (https://www.pacb.com) and Oxford Nanopore Technologies (https://nanoporetech.com), are proving to be effective in resolving these problems. Accordingly, research focusing on the complex genomic structure of wheat will be further facilitated by using a combination of established genomic resources and newly developed technologies.

Acknowledgements

References

Adamski, N. M. et al. (2020) A roadmap for gene functional characterisation in crops with large genomes: Lessons from polyploid wheat. *eLife*, **9**, e55646.

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

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

Barabaschi, D. et al. (2015) Physical mapping of bread wheat chromosome 5A: An integrated approach. *Plant Genome*, **8**, 1-24.

Belova, T. et al. (2014) Utilization of deletion bins to anchor and order sequences along the wheat 7B chromosome. *Theor. Appl. Genet.*, **127**, 2029-2040.

Berkman, P. J. et al. (2011) Sequencing and assembly of low copy and genic regions of isolated *Triticum aestivum* chromosome arm 7DS. *Plant Biotechnol. J.*, **9**, 768-775.

Berkman, P. J. et al. (2012) Sequencing wheat chromosome arm 7BS delimits the 7BS/4AL translocation and reveals homoeologous gene conservation. *Theor. Appl. Genet.*, **124**, 423-432.

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

Choulet, F. et al. (2014) Structural and functional partitioning of bread wheat chromosome 3B. *Science*, **345**, 1249721.

Doležel, J. et al. (2007) Chromosome-based genomics in the cereals. *Chromosome Res.*, **15**, 51-66.

Handa, H. et al. (2018) Structural features of two major nucleolar organizer regions (NORs), *Nor-B1* and *Nor-B2*, and chromosome-specific rRNA gene expression in wheat. *Plant J.*, **96**, 1148-1159.

Handa, H. et al. (2019) The accurate genome sequences combined with the expressed proteins data for the complete grasp of the gliadin locus, *Gli-B2*, on wheat chromosome 6B. *1st International Wheat Congress Abstract Proceedings Poster Presentations*.

Helguera, M. et al. (2015) New insights into the wheat chromosome 4D structure and virtual gene order, revealed by survey pyrosequencing. *Plant Sci.*, **233**, 200-212.

Hernandez, P. et al. (2012) Next-generation sequencing and syntenic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content. *Plant J.*, **69**, 377-386.

Holušová, K. et al. (2017) Physical map of the short arm of bread wheat chromosome 3D. *Plant Genome*, **10**, 1-11.

International Rice Genome Sequencing Project (IRGSP) (2005)

The map-based sequence of the rice genome. *Nature*, **436**, 793-800.

Ishikawa, G. et al. (2007) PCR-based landmark unique gene (PLUG) markers effectively assign homoeologous wheat genes to A, B and D genomes. *BMC Genomics*, **8**, 135.

Ishikawa, G. et al. (2009) Localization of anchor loci representing five hundred annotated rice genes to wheat chromosomes using PLUG markers. *Theor. Appl. Genet.*, **118**, 499-514.

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

Kawahara, Y. et al. (2016) TENOR: Database for comprehensive mRNA-Seq experiments in rice. *Plant Cell Physiol.*, **57**, e7.

Kawaura, K. et al. (2012) Genome change in wheat observed through the structure and expression of α/β-gliadin genes. *Funct. Integr. Genomics*, **12**, 341-355.

Keeble-Gagnère, G. et al. (2018) Optical and physical mapping with local finishing enables megabase-scale resolution of agronomically important regions in the wheat genome. *Genome Biol.*, **19**, 112.

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

Lieberman-Aiden, E. et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289-293.

Ling, H. Q. et al. (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu. Nature*, **496**, 87-90.

Ling, H. Q. et al. (2018) Genome sequence of the progenitor of wheat A subgenome *Triticum urartu. Nature*, **557**, 424-428.

Lu, F. H. et al. (2018) Independent assessment and improvement of wheat genome sequence assemblies using Fosill jumping libraries. *Gigascience*, **7**, giy053.

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

Lucas, S. J. et al. (2014) Next-generation sequencing of flow-sorted wheat chromosome 5D reveals lineage-specific translocations and widespread gene duplications. *BMC Genomics*, **15**, 1080.

Luo, M. C. et al. (2017) Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii. Nature*, **551**, 498-502.

Maccaferri, M. et al. (2019) Durum wheat genome highlights past domestication signatures and future improvement targets. *Nat. Genet.*, **51**, 885-895.

Mascher, M. et al. (2013) Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *Plant J.*, **76**, 718-727.

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

McIntosh, R. A. et al. (2013) Catalogue of gene symbols for wheat. http://www.shigen.nig.ac.jp/wheat/komugi/genes/download.jsp.

Monat, C. et al. (2019) TRITEX: Chromosome-scale sequence assembly of *Triticeae* genomes with open-source tools. *Genome Biol.*, **20**, 284.

Morrison, J. W. (1953) Chromosome behavior in wheat monosomics. *Heredity*, **7**, 203-217.

Nesterov, M. A. et al. (2016) Identification of microsatellite loci based on BAC sequencing data and their physical mapping into the soft wheat 5B chromosome. *Russ. J. Genet. Appl. Res.*, **6**, 825-837.

Ouyang, S. et al. (2007) The TIGR Rice Genome Annotation Resource: Improvements and new features. *Nucleic Acids Res.*, **35**, D883-D887.

Paterson, A. H. et al. (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, **457**, 551-556.

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

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

Pikaard, C. S. (2000) The epigenetics of nucleolar dominance. *Trends Genet.*, **16**, 495-500.

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

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

Rouard, M. et al. (2011) GreenPhylDB v2.0: Comparative and functional genomics in plants. *Nucleic Acids Res.*, **39**, D1095-D1102.

Šafář, J. et al. (2010) Development of chromosome-specific BAC resources for genomics of bread wheat. *Cytogenet. Genome Res.*, **129**, 211-223.

Sakai, H. et al. (2013) Rice Annotation Project Database (RAP-DB): An integrative and interactive database for rice genomics. *Plant Cell Physiol.*, **54,** e6.

Salina, E. A. et al. (2018) Features of the organization of bread wheat chromosome 5BS based on physical mapping. *BMC Genomics*, **19 (Suppl 3)**, 80.

Sato, Y. et al. (2013) RiceXPro version 3.0: expanding the informatics resource for rice transcriptome. *Nucleic Acids Res.*, **41**, D1206-D1213.

Schmutz, J. et al. (2010) Genome sequence of the palaeopolyploid soybean. *Nature*, **463**, 178-183.

Schnable, P. S. et al. (2009) The B73 maize genome: Complexity, diversity, and dynamics. *Science*, **326**, 1112-1115.

Sehgal, S. K. et al. (2012) Chromosome arm-specific BAC end sequences permit comparative analysis of homoeologous chromosomes and genomes of polyploid wheat. *BMC Plant Biol.*, **12**, 64.

Sergeeva, E. M. et al. (2014) Common wheat chromosome 5B composition analysis using low-coverage 454 sequencing. *Plant Genome*, **7**, 1-16.

Sergeeva, E. M. et al. (2017) Fine organization of genomic regions tagged to the 5S rDNA locus of the bread wheat 5B chromosome. *BMC Plant Biol.*, **17 (Suppl 1)**, 183.

Shorinola, O. et al. (2017) Association mapping and haplotype analysis of the pre-harvest sprouting resistance locus *Phs-A1* reveals a causal role of TaMKK3-A in global germplasm. *Front. Plant Sci.*, **8**, 1555.

Soderlund, C. et al. (1997) FPC: A system for building contigs from restriction fingerprinted clones. *Comput. Appl. Biosci.*, **13**, 523-535.

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

Tanaka, T. et al. (2014) Next-generation survey sequencing and the molecular organization of wheat chromosome 6B. *DNA Res.*, **21**, 103-114.

Tello-Ruiz, M. K. et al. (2016) Gramene 2016: Comparative plant genomics and pathway resources. *Nucleic Acids Res.*, **44**, D1133-D1140.

The International Barley Genome Sequencing Consortium (2012) A physical, genetic and functional sequence assembly of the barley genome. *Nature*, **491**, 711-716.

The International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*, **463**, 763-768.

The International Wheat Genome Sequencing Consortium (IWGSC) (2014) A chromosome-based draft sequence of the hexaploid bred wheat genome. *Science*, **345**, 1251788.

The International Wheat Genome Sequencing Consortium (IWGSC) (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, **361**, eaar7191.

Tiwari, V. K. et al. (2012) Endosperm tolerance of paternal aneuploidy allows radiation hybrid mapping of the wheat D-genome and a measure of γ ray-induced chromosome breaks. *PLoS ONE*, **7**, e48815.

Tulpová, Z. et al. (2019) Integrated physical map of bread wheat chromosome arm 7DS to facilitate gene cloning and comparative studies. *New Biotechnol.*, **48**, 12-19.

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

Vitulo, N. et al. (2011) First survey of the wheat chromosome 5A composition through a next generation sequencing approach. *PLoS ONE*, **6**, e26421.

Walkowiak, S. et al. (2020) Multiple wheat genomes reveal global variation in modern breeding. *Nature*, **588**, 277-283.

Watanabe, S. et al. (2014) A tip to generate a radiation hybrid map of wheat chromosome: How could we connect linkage groups separated at the pericentromeric region on chromosome 6B? *Wheat Inf. Ser.*, **117**, 5-7.

Wicker, T. et al. (2011) Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives. *Plant Cell*, **23**, 1706-1718.

Zhao, G. et al. (2017) The *Aegilops tauschii* genome reveals multiple impacts of transposons. *Nat. Plants*, **3**, 946-955.

Zhu, T. et al. (2019) Improved genome sequence of wild emmer wheat Zavitan with the aid of optical maps. *G3 (Bethesda)*, **9**, 619-624.

Zhu, T. et al. (2021) Optical maps refine the bread wheat *Triticum aestivum* cv. Chinese Spring genome assembly. *Plant J.*