

A Review of Genome Sequencing in the Largest Cereal Genome, *Triticum aestivum* L.

Anandi Bierman, Anna-Maria Botha

Department of Genetics, Stellenbosch University, Stellenbosch, South Africa

Email: anandie@sun.ac.za

How to cite this paper: Bierman, A. and Botha, A.-M. (2017) A Review of Genome Sequencing in the Largest Cereal Genome, *Triticum aestivum* L. *Agricultural Sciences*, 8, 194-207.

<https://doi.org/10.4236/as.2017.82014>

Received: December 22, 2016

Accepted: February 24, 2017

Published: February 27, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Sequencing whole plant genomes has advanced rapidly with the development of next generation sequencing (NGS) technologies and bioinformatics, enabling the study of large and complex genomes such as that of the hexaploid cereal, *Triticum aestivum* L. (bread wheat). Despite advances, however, confounding factors such as repetitive elements and low polymorphism still hinder sequencing attempts. Isolation techniques such as sequencing of diploid progenitors and chromosome separation through flow cytometry have showed promise in reducing the size of the genome for sequencing. In this review we discuss the advances and stumbling blocks that have been encountered on the road toward the complete hexaploid wheat genome sequence. We also discuss the latest complimentary techniques and the progression of accumulation of sequence data relevant to wheat genome research.

Keywords

Hexaploid, Next Generation Sequencing, Flow Cytometry, Wheat, Genome

1. Introduction

The wheat genome is one of the largest (17,000 Mbp) and most repeat-rich of all crop plants (76.6% repetitive elements in assembled sequences; [1]). Wheat is an allohexaploid (*Triticum aestivum* L.; $2n = 6x = 42$) of which the genome is divided into three sub-genomes (AA, BB and DD). These sub-genomes act as independent diploid genomes during meiosis due to the action of the *Ph* genes on chromosome 5B [2] [3]. The wheat genome was formed by two polyploidization events. The first polyploidization event formed *Triticum turgidum* ($2n = 4x = 28$; AABB), an allotetraploid which resulted from *Triticum urartu* ($2n = 2x = 14$; AA) and a species which is believed to be *Aegilops speltoides* ($2n = 14$; SS), from the *Sitopsis* section of *Triticum* [4]. The second hybridization event, which resulted in the ancestral *Triticum aestivum* ($2n = 6x = 42$; AABBDD), occurred

between *Triticum turgidum* and the diploid grass, *Aegilops tauschii* (DD) [4] [5]. These progenitor species often provide valuable genetic resources by serving as diploid model organisms for the study of wheat [6]. The sub-genomes in modern hexaploid wheat differ significantly from one another and many genes are not present in triplicate, but are chromosome specific [7]. Genes are also not distributed evenly, but are clustered in gene-rich regions, particularly at the distal regions of chromosomes [8].

The availability of the wheat reference genome sequence will play an absolutely important role in employing molecular tools within breeding programs. By enabling the identification, more efficiently, of genes and markers related to agronomically important traits, breeding schemes can be accelerated to employ and deliver the best stock.

2. First Steps toward Sequencing the Wheat Genome

2.1. Initial Focus on Markers and Genes

Early endeavors toward obtaining wheat sequence information focused on the coding regions of the genome such as cDNA and ESTs. Resources such as 17,000 cDNA sequences, 40,000 UniGenes, 1 million ESTs [9], of which 7000 have been assigned to chromosome-specific bins [10] and great numbers of SNP markers [11] [12] [13] were compiled before a draft genome sequence was available. These initial tools have allowed the development of marker sets and microarrays used in gene expression studies [14] [15] as well as the first views of genome localization and low resolution mapping of phenotypic traits.

Up until 2004, most completed genome sequences had been obtained through a clone-by-clone approach [16]-[22]. Though the data provided was comprehensive, it was an expensive approach, especially if considered for a large genome such as that of wheat. In turn, other methods also had their limitations as whole genome shotgun sequencing was hindered by computational power, especially for genomes with large numbers of repeats and genomic filtration techniques such as methylation filtration [23] and *COt* selection [24] [25] were not well tested enough to guarantee high percentages of gene identification [7]. Ultimately, it was decided to sequence individual chromosomes of wheat, separated by flow cytometry [26] [27] and used to construct Bacterial Artificial Chromosome (BAC) libraries and filtration libraries.

2.2. From BAC Libraries to Isolated Wheat Chromosomes

Construction of BAC libraries from progenitor species provided the first steps toward a reduction of the wheat genome by focusing only on one sub genome within a BAC library. In 1999, Lijavetzky *et al.* [28] constructed a BAC library of the A genome using *T. monococcum* as template. This provided a 5.6 times coverage of the A genome and significantly increased the probability of finding a sequence of interest (>99.6%). However, many of the BAC ends produced included repetitive sequences, and in order to circumnavigate this, a low copy sequence selection step had to be introduced during the end-isolation procedure

which subsequently enabled chromosome walking to obtain fragments of wheat sequence information.

The isolation of individual wheat chromosomes further increased the discriminatory power of BAC libraries in wheat. Kubaláková *et al.* (2002) [29] first demonstrated that wheat chromosomes could be flow sorted intact, though due to the lack of size differences between chromosomes, initially only 3B could be isolated with high purity. Estimated at 995 Mbp [1] [30], chromosome 3B is twice as large as the entire rice genome (370 Mbp; [31]), which made it an attractive candidate for initial attempts at physical mapping and sequencing [26]. After flow sorting ± 1.8 million 3B chromosomes, a BAC library could be constructed representing 6.2 equivalents of the 3B chromosome [27]. As flow cytometric techniques improved, smaller chromosomes could be separated using telosomic wheat lines and lines carrying isochromosomes [29]. Aneuploid or ditelosomic wheat lines or wheat-rye addition lines [32] also provided select chromosomes with sizes differentiable from the remaining chromosomal complement.

2.3. Sequencing the Wheat Progenitors

In 2013, Jia *et al.* [33] published a draft genome sequence of the diploid D genome progenitor, *Ae. tauschii*, and obtained an assembly representative of 97% of the 4.36 Gbp genome using the Roche-454 sequencing platform. The estimated 65.9% repetitive elements in this genome are lower than that of hexaploid bread wheat, while it is estimated to contain 34,498 Protein Coding Genes (PCGs) which does correspond, roughly, to a third of the total number of genes in hexaploid bread wheat [1]. A draft genome assembly of *Triticum urartu* [34] was also produced through whole genome shotgun approaches.

During discussions at the International Wheat Genome Sequencing Consortium (IWGSC) meeting in 2003 though, many positive outcomes were listed for starting to sequence the hexaploid wheat genome in its entirety. Modern wheat has diverged substantially from its progenitors and its polyploid nature could reveal much about its evolution as well as polyploid speciation [7], which is seen in many agriculturally significant plant species. The final argument was that agriculturally important genes might be chromosome specific and possibly overlooked if only one of the progenitors were selected for sequencing [7].

As the sequencing of isolated wheat chromosomes and progenitor genomes became a reality, the first IWGSC was assembled and a workshop was held in Crystal City, Washington, DC in November 2003. This workshop would set into motion the unraveling of the genome of hexaploid wheat. The aim of the workshop was to discuss the status and future of wheat genomics with particular focus on the possibility of sequencing the complete wheat genome. At the meeting, the cultivar, “Chinese Spring” was selected as universal choice for sequencing of the wheat genome due to the availability of genetic stocks of this cultivar [7].

3. 2012 Draft Genome Sequence

Nine years after the initial conceptualization of a genome sequence for hexaploid

wheat, Brenchley *et al.* (2012) [35] published the first draft sequence obtained through whole genome shotgun sequencing. The authors obtained lower coverage (5 fold) with longer read lengths on the Roche 454 sequencing platform. Using diploid wheat relatives and progenitors the authors could classify homeologous relationships and build orthologous gene family frameworks. Ultimately, a catalog of 132,000 SNPs was created in the various sub-genomes of wheat. Repeat elements were identified to make up 79% of the sequenced genome and the number of genes was estimated at 94,000 to 96,000. A representation of nearly all wheat genes could be identified as constructed orthologous groups matched 90% of metabolic genes in *Arabidopsis* and 92% of publicly available wheat full-length cDNA. Overall, a trend toward gene family size reduction was observed in hexaploid wheat despite its recent evolution as a hexaploid. Expanded gene families' overrepresented categories included proteins involved in energy metabolism, defense, growth and nutrition. The authors also found gene families to be classified as A, B or D-genome derived, which indicates that transcriptional regulatory networks are maintained in a genome-specific manner in wheat.

4. 2014 IWGSC Genome

In 2014 the IWGSC published their paper on the wheat genome sequence [1]—an initiative that was started more than 10 years prior. The hexaploid nature of the wheat genome allows for a buffering effect in aneuploid wheat lines and viable mono-, tri and tetrasomic cytogenetic stocks as well as nullisomics have been developed in addition to aneuploids of every chromosome arm [36] [37]. Mayer *et al.* (2014) [1] used double ditelosomic wheat lines [37] of the cultivar, “Chinese Spring” to isolate individual chromosomes for sequencing on the Illumina platform. By sequencing the chromosomes individually, the sequencing consortium reduced much of the complexity of the genome and was able to differentiate genes which showed multiple copies as well as conserved homologs. Sequencing depth obtained ranged between 30x and 241x [1].

In their chromosome-based sequence data of the wheat genome, Mayer *et al.*, (2014) [1] found 81% of raw sequencing reads and 76.6% of assembled sequences to contain repeats. Retro-elements were most abundant in the A sub-genome chromosomes (A > B > D in order of abundance) while transposons were most abundant in the D sub-genome (D > B > A in order of abundance). Of 270 miRNA molecules identified, 49 had not been reported previously and correspondence to 98,068 predicted miRNA-coding loci was identified. At least one target gene (putative) could be identified for 87% of these miRNA-coding loci. Gene annotation of the chromosome-based wheat genome sequence relied on comparisons to genes in related grass species such as *Brachypodium distachyon* [38], *Oryza sativa* [39], *Sorghum bicolor* [40], and *Hordeum vulgare* [41] in addition to full length wheat cDNAs [9] and RNA-seq data from “Chinese Spring” tissue at different stages of development. Mayer *et al.* (2014) [1] estimated a count of 106,000 protein coding genes in wheat, which is a slightly greater figure than that estimated by Brenchley *et al.* (2012) [35] though consistent

with estimates for the diploid sub-genomes, ranging between 32,000 and 38,000 as well as estimates for the diploid progenitors [33] [34] [35]. The distribution of genes across the sub-genomes was highest in the B sub-genome (35%) followed by A (33%) and D (32%) though this distribution pattern was not reflected at the level of the homeologous groups of chromosomes. Variation of up to 2.4 fold was observed among the different chromosome arms. This observation is also seen in rye [42].

Lineage specific intra-chromosomal duplication was determined for genes on each wheat chromosome. On average for all chromosomes, 23.6% of genes are duplicated, though Mayer *et al.* (2014) [1] state that this is likely an underestimation. Comparisons with intra-chromosomal duplicates in other species such as rice, sorghum, barley, maize and millet, show the gene duplications in wheat to be significantly higher (17% - 20%); [1] [38] [39] [40] [43] [44]. In comparing gene family sizes of hexaploid wheat with *Ae. tauschii* and *T. urartu* [33] [34], the authors found that genes belonging to expanded families were mainly affected by gene loss as was also reported by Brenchley *et al.* (2012) [35]. On the other hand, genes without paralogous copies, *i.e.* singletons, were not subject to gene loss and the retention rate of genes was very similar across the three sub-genomes of hexaploid wheat. The authors found no evidence for a gradual gene loss after polyploidization although the D genome, the most recent addition to hexaploid wheat, did show the lowest levels of gene loss of the three sub-genomes. In addition to their similarity in gene distribution, the gene content of the A, B and D sub-genomes was very similar with only a small number of truly unique genes (1.3% to 1.7%).

Sequence conservation in chromosomes of hexaploid wheat showed highly conserved gene sequences between the sub-genomes and their diploid relatives. In addition, the wheat sub-genomes have high regulatory autonomy with little regulation between the sub-genomes [45]. The patterns of gene expression in hexaploid wheat contrasts to that observed in other species such as allopolyploid cotton [46] [47], mesopolyploid *Brassica rapa* [48], synthetic allotetraploid *Arabidopsis* [49], and paleopolyploid maize [50], which all have one sub-genome more transcriptionally active than the others. Knowledge still outstanding, according to Mayer *et al.* (2014) [1], concerns the distribution of genes and their positions along the chromosomes of hexaploid wheat as well as gene evolution during wheat's development.

5. New Developments

BioNano mapping or optical mapping (genome mapping in nanochannel arrays) [51] is fast becoming a useful tool as it generates short sequence maps along DNA that stretches thousands of base pairs in length, thus allowing for novel techniques for mapping and assembling gigabase sized genomes. As the IWGSC wheat reference genome is produced by the sequencing of physical maps from individually isolated chromosomes, applying the BioNano mapping strategy here seems practical. Šimková *et al.* (2016) [52] discussed their strategy of generating

a BioNano high resolution map for chromosome 7DS. The authors generated 371 contigs with N50 of 1.3 Mbp. Chromosome 7DS sequence assemblies obtained through clone-by-clone sequencing were anchored to the 7DS BioNano map. This proved valuable in improving BAC-contig physical maps and validating sequence assembly [52].

The BioNano technology has been implemented on a genome-wide scale by Zhu *et al.* (2016) [53] in order to study genome structure. The authors discuss the construction of whole genome BioNano maps for both wheat and *Ae. tauschii* and the subsequent comparison in order to detect structural differences between the D sub-genome of wheat and the *Ae. tauschii* genome. Their comparison yielded a large amount of indels and through using the BioNano maps, novel information on these differences could be obtained. Indels that occurred during the evolution of wheat could be discerned, located exactly and their role in recombination defined [53].

At the Plant and Animal Genome Conference in San Diego, 2016, Zimin *et al.* [54] made the remark that Illumina sequencing technologies, though the most popular sequencing technology, produces short reads (up to 300 bp) which often further complicates the analysis of already complicated genomes due to the fragmentary nature of the data. Longer reads would be advantageous and PacBio has recently stepped up to that challenge [55]. Though expensive, PacBio produces reads in the order of tens of thousands of base pairs. The error rate of PacBio is still quite significant and therefore, these authors propose a hybrid strategy leveraging the accuracy of the shorter Illumina reads with the length of the PacBio reads to construct an assembly of the 4.5 Gbp *Ae. tauschii* genome [54]. This hybrid approach making use of Illumina short reads in combination with the longer reads offered by PacBio was also used to construct the completed genome sequence of *T. urartu*, for which the draft genome was published in 2013 [34] and the genome sequence completed and presented 3 years later at the Plant and Animal Genome conference in 2016 [56].

Clavijo *et al.* (2016) [57] stated that despite all the improvements to date, an accurate and nearly complete assembly of the wheat genome is still out of reach. They state that the first draft sequence available [35], based on orthologous group assemblies of related grass protein sequences was highly fragmented and though the subsequent, chromosome-based assembly (IWGSC chromosome survey sequence (CSS); [1]) was able to identify homeologous relationships between the sub-genomes, it too remained fragmented. These authors, therefore, provide a sequence assembly and annotation of “Chinese Spring” wheat with more in-depth analysis of its sequence and structure. With 33x sequence coverage of the genome, the scaffolds produced through their study, contain 13.4 Gbp of the total genome with an N50 value of 88.8 kbp. Their annotation contains 104,091 protein coding genes of high confidence as it is supported by transcriptome sequence data and full length cDNA sequence data. This new assembly, termed TGAC1, was classified into chromosome arms by alignment to raw reads from the IWGSC-CSS data [1] and represents nearly 80% of the 17 Gbp wheat

genome which provides a 60% improvement in genome coverage. In confirmation with previous studies, more than 80% of the TGAC1 assembly comprised of transposable elements of which 70% were retro-elements and 13% DNA transposons [57].

Miss-assembly and gaps in the sequence data are recurring problems in the assembly of any genome whether by whole genome shotgun approaches [58] [59] [60] or clone-by-clone approaches [61] [62]. Applications providing longer reads have fast become very popular for solving some of the issues associated with genome assemblies and platforms such as PacBio [55] and BioNano [51] show promise to improve the quality of shotgun assemblies. In an attempt at the wheat reference genome, the IWGSC adopted a clone-by-clone sequencing approach based on physical maps constructed from BAC libraries specific to individual chromosome arms. In order to reconstruct the genome from BAC sequences, contigs from the physical map are anchored—a process which relies heavily on many genetic markers present on the contigs, and with known positions on the chromosomes. Though high-density linkage maps are available for wheat, their resolution is often limited due to small sizes of mapping populations and regions of low recombination along wheat chromosomes [63] [64]. To circumvent this particular problem, approaches independent of recombination, such as the BioNano genome mapping system, have proven highly valuable [65].

6. What Is Currently Available

The IWGSC (<http://www.wheatgenome.org/About>) have released several whole genome assemblies of both wheat and its progenitors. As of January 2014 the IWGSC released genome survey sequence assemblies [1], population sequencing (POPSEQ) data and Genome Zipper data as well as annotated chromosome 3B pseudomolecule data for download by the public, followed shortly by the release of diploid and other progenitor wheat species' whole genome sequences for download by the public. The initial wheat genome assembly (IWGSC1 + POPSEQ; http://archive.plants.ensembl.org/Triticum_aestivum/Info/Index) was a chromosome survey sequence of the “Chinese Spring” cultivar of hexaploid wheat, refined into chromosomal pseudomolecules using POPSEQ data [66]. By August 2014, a new version of the wheat survey sequence gene model was made publicly available (v2.2). Up until the release of the latest TGAC version of the wheat reference genome assembly, three versions of the IWGSC-CSS data had been released to the public with improvements such as the cleaning of data by removal of duplicates (version 2) and the incorporation of 185 Gbp of mate pair sequence data (version 3).

In April 2016 the TGACv1 wheat genome assembly was made publicly available on Ensembl Plants. The new assembly (TGAC, [57]) has a greater representation of the genome, at 78% with N50 for the scaffolds of 88.8 kbp. This new assembly combined Illumina and PacBio sequencing platforms to identify 104,091 protein coding genes and 10,156 non-coding RNA genes. In June of that same year the IWGSC WGA v0.4 was made publicly available for BLAST analy-

sis or download (IWGSC WGA v0.4; http://plants.ensembl.org/Triticum_aestivum/Info/Index; [57]). This new assembly consists of 13.4 Gbp in contigs greater than 500 bp with 99% of the total number of genes from the previous assembly, located in the new TGAC assembly. In addition, alignments of RNA-seq data have also been added to this assembly. The whole genome assembly for wheat, available through the IWGSC (<https://wheat-urgi.versailles.inra.fr/Seq-Repository/>; [1]) provides researchers with a chromosome-based draft sequence of “Chinese Spring” wheat annotated with 124,201 gene loci and representing 61% of the genome sequence. The Illumina short sequence reads produced scaffolds adding to a total of 14.5 Gbp (L50 7.1 Mbp).

7. Future Prospects and Conclusions

From the data generated on the hexaploid wheat genome thus far and the number of mapping studies on this cereal, it is clear that novel and combined approaches are required in order to characterize genes of agronomic importance. Novelty can be found in the advances of sequencing platforms such as PacBio (<http://www.pacificbiosciences.com>). The platform provides a method of sequencing individual DNA molecules in real time with read lengths greater than 1000 bp [67]. This already provides a ten-fold improvement in sequence read length and in combination with optical mapping (BioNano; [68]), may provide exciting advances to the analysis of complex genomes such as hexaploid wheat. This technology may provide the highly needed bridge between DNA sequence and chromosome, as the inability to *de novo* assemble these complex genomes often fail to bring the two together [69] [70]. Using optical mapping technology, Hastie *et al.* (2013) [70] assembled a 2.1 Mbp repetitive region of the *Aegilops tauschii* genome. The technique makes direct visualization of sequence motifs on long single DNA molecules possible, and previously unplaced sequence contigs can therefore be anchored. Following this approach, Hastie *et al.* (2013) [70] improved the genome assembly of *Ae. tauschii* significantly by increasing its completeness from 75% complete to 95% complete.

Genome zippers (linear gene order models) are fast gaining popularity in mapping studies [71]. The Genome Zipper approach [72] makes use of comparisons of shotgun sequences with reference grass genomes such as *Brachypodium*, rice and sorghum. Genes identified within the syntenic regions generated by these comparisons are used to construct a genomic build along a scaffold of markers, taking into account the order of the sequence tagged genes within their reference genomes and the order deduced from the scaffold of markers [73]. Vitulo *et al.* (2011) [74] generated a genome zipper for the short and long arms of chromosome 5A of wheat by identifying conserved homologies in *Brachypodium*, rice and sorghum. The authors stated that the basis of the genome zipper concept is that the virtual gene order in one species can be constructed based on synteny with closely related species. Also in 2011, Mayer *et al.* [72] were able to assemble 21,766 genes in a putative linear order in barley through the use of

genome zippers.

Understanding wheat biology not only serves academic purposes but paves the way toward increasing of yields and the identification of phenotypes that protect against biotic and abiotic stresses. Genome sequencing of wheat is one way to increase our understanding of this hexaploid cereal [7]. Sequencing the wheat genome facilitates our understanding of related grass species through annotation based on conservation of orthologous groups between the grasses [7]. The availability of a polyploid, such as wheat's, genome sequence also provides an attractive model toward the study of genome changes behind the evolution of polyploidy in plants [65].

The wheat genome sequence has posed such a challenge to the scientific community, that the importance of compiling its genome has allowed for the implementation and advancement of new sequencing technologies such as Pac-Bio and BioNano [52] [54]. These rapid advancements in science and technology, so akin to the green revolution of the 1960's, have been pivotal in the completion of the wheat genome sequence [1].

References

- [1] Mayer, K.F.X., Rogers, J., Doležel, J., Pozniak, C., Eversole, K., Feuillet, C., *et al.* (2014) A Chromosome-Based Draft Sequence of the Hexaploid Bread Wheat (*Triticum aestivum*) Genome. *Science*, **345**, Article ID: 1251788. <https://doi.org/10.1126/science.1251788>
- [2] Gautier, M.F., Cosson, P., Guirao, A., Alary, R. and Joudrier, P. (2000) Puroindoline Genes Are Highly Conserved in Diploid Ancestor Wheats and Related Species but Absent in Tetraploid *Triticum* Species. *Plant Science*, **153**, 81-91. [https://doi.org/10.1016/S0168-9452\(99\)00258-7](https://doi.org/10.1016/S0168-9452(99)00258-7)
- [3] Martinez-Perez, E., Shaw, P. and Moore, G. (2001) The *Ph1* Locus Is Needed to Ensure Specific Somatic and Meiotic Centromere Association. *Nature*, **411**, 204-207. <https://doi.org/10.1038/35075597>
- [4] Petersen, G., Seberg, O., Yde, M. and Berthelsen, K. (2006) Phylogenetic Relationships of *Triticum* and *Aegilops* and Evidence for the Origin of the A, B, and D Genomes of Common Wheat (*Triticum aestivum*). *Molecular Phylogenetics and Evolution*, **39**, 70-82. <https://doi.org/10.1016/j.ympev.2006.01.023>
- [5] Nesbitt, M. and Samuel, D. (1996) From Staple Crop to Extinction? The Archaeology and History of the Hulled Wheats. *Hulled Wheats*, **4**, 41-100.
- [6] Kumar, A., Seetan, R., Mergoum, M., Tiwari, V.K., Iqbal, M.J., Wang, Y., *et al.* (2015) Radiation Hybrid Maps of the D-Genome of *Aegilops tauschii* and Their Application in Sequence Assembly of Large and Complex Plant Genomes. *BMC Genomics*, **16**, 800. <https://doi.org/10.1186/s12864-0152030-2>
- [7] Gill, B.S., Appels, R., Botha-Oberholster, A.-M., Buell, C.R., Bennetzen, J.L., Chalhouh, B., *et al.* (2004) A Workshop Report on Wheat Genome Sequencing: International Genome Research on Wheat Consortium. *Genetics*, **168**, 1087-1096. <https://doi.org/10.1534/genetics.104.034769>
- [8] Gupta, P.K., Mir, R.R., Mohan, A. and Kumar, J. (2008) Wheat Genomics: Present Status and Future Prospects. *International Journal of Plant Genomics*, **2008**, Article ID: 896451. <https://doi.org/10.1155/2008/896451>
- [9] Mochida, K., Yoshida, T., Sakurai, T., Ogihara, Y. and Shinozaki, K. (2009)

- TriFLDB: A Database of Clustered Full-Length Coding Sequences from Triticeae with Applications to Comparative Grass Genomics. *Plant Physiology*, **150**, 1135-1146. <https://doi.org/10.1104/pp.109.138214>
- [10] Qi, L.L., Echaliier, B., Chao, S., Lazo, G.R., Butler, G.E., Anderson, O.D., *et al.* (2004) A Chromosome Bin Map of 16,000 Expressed Sequence Tag Loci and Distribution of Genes among the Three Genomes of Polyploid Wheat. *Genetics*, **168**, 701-712. <https://doi.org/10.1534/genetics.104.034868>
- [11] Allen, A.M., Barker, G.L.A., Wilkinson, P., Burrige, A., Winfield, M., Coghill, J., *et al.* (2013) Discovery and Development of Exome-Based, Co-Dominant Single Nucleotide Polymorphism Markers in Hexaploid Wheat (*Triticum aestivum* L.). *Plant Biotechnology Journal*, **11**, 279-295. <https://doi.org/10.1111/pbi.12009>
- [12] Krasileva, K.V., Buffalo, V., Bailey, P., Pearce, S., Ayling, S., Tabbita, F., *et al.* (2013) Separating Homeologs by Phasing in the Tetraploid Wheat Transcriptome. *Genome Biology*, **14**, R66. <https://doi.org/10.1186/gb-2013-14-6-r66>
- [13] Saintenac, C., Jiang, D., Wang, S. and Akhunov, E. (2013) Sequence-Based Mapping of the Polyploid Wheat Genome. *G3: Genes Genomes Genetics*, **3**, 1105-1114. <https://doi.org/10.1534/g3.113.005819>
- [14] Coram, T.E., Settles, M.L., Wang, M. and Chen, X. (2008) Surveying Expression Level Polymorphism and Single-Feature Polymorphism in Near-Isogenic Wheat Lines Differing for the *Yr5* Stripe Rust Resistance Locus. *Theoretical and Applied Genetics*, **117**, 401-411. <https://doi.org/10.1007/s00122-008-0784-5>
- [15] Bernardo, A.N., Bradbury, P.J., Ma, H., Hu, S., Bowden, R.L., Buckler, E.S., *et al.* (2009) Discovery and Mapping of Single Feature Polymorphisms in Wheat Using Affymetrix Arrays. *BMC Genomics*, **10**, 251. <https://doi.org/10.1186/1471-2164-10-251>
- [16] Waterston, R. (1998) Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology. *Science*, **282**, 2012-2018. <https://doi.org/10.1126/science.282.5396.2012>
- [17] The, A.G.I. (2000) Analysis of the Genome Sequence of the Flowering Plant *Arabidopsis thaliana*. *Nature*, **408**, 796-815. <https://doi.org/10.1038/35048692>
- [18] McPherson, J.D., Marra, M., Hillier, L., Waterston, R.H., Chinwalla, A., Wallis, J., *et al.* (2001) A Physical Map of the Human Genome. *Nature*, **409**, 934-941. <https://doi.org/10.1038/35057157>
- [19] Feng, Q., Zhang, Y., Hao, P., Wang, S., Fu, G., Huang, Y., *et al.* (2002) Sequence and Analysis of Rice Chromosome 4. *Nature*, **420**, 316-320. <https://doi.org/10.1038/nature01183>
- [20] Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T. and Katayose, Y. (2002) The Genome Sequence and Structure of Rice Chromosome 1. *Nature*, **420**, 312-316. <https://doi.org/10.1038/nature01184>
- [21] Wood, V., Gwilliam, R., Rajandream, M.A., Lyne, M., Lyne, R., Stewart, A., *et al.* (2002) The Genome Sequence of *Schizosaccharomyces pombe*. *Nature*, **415**, 871-880. <https://doi.org/10.1038/nature724>
- [22] Rice Chromosome Sequencing Consortium (2003) In-Depth View of Structure, Activity and Evolution of Rice Chromosome. *Science*, **300**, 1566-1569. <https://doi.org/10.1126/science.1083523>
- [23] Rabinowicz, P.D., Schutz, K., Dedhia, N., Yordan, C., Parnell, L.D., Stein, L., *et al.* (1999) Differential Methylation of Genes and Retrotransposons Facilitates Shotgun Sequencing of Maize Genome. *Nature Genetics*, **23**, 305-308. <https://doi.org/10.1038/15479>

- [24] Peterson, D.G., Schulze, S.R., Sciara, E.B., Lee, S.A., Bowers, J.E., Nagel, A., *et al.* (2002) Integration of Cot Analysis, DNA Cloning and High-Throughput Sequencing Facilitates Genome Characterization and Gene Discovery. *Genome Research*, **12**, 795-807. <https://doi.org/10.1101/gr.226102>
- [25] Yuan, Y., SanMiguel, P. and Bennetzen, J. (2003) High-Cot Sequence Analysis of the Maize Genome. *The Plant Journal*, **34**, 249-255. <https://doi.org/10.1046/j.1365-313X.2003.01716.x>
- [26] Vrána, J., Kubaláková, M., Šimková, H., Čihalíková, J., Lysák, M.A., Doležel, J. (2000) Flow Sorting of Mitotic Chromosomes in Common Wheat (*Triticum aestivum* L.). *Genetics*, **156**, 2033-2041
- [27] Šáfář, J., Bartos, J., Janda, J., Bellec, A., Kubaláková, M., Valárik, M., *et al.* (2004) Dissecting Large and Complex Genomes: Flow Sorting and BAC Cloning of Individual Chromosomes from Bread Wheat. *Plant Journal*, **39**, 960-968. <https://doi.org/10.1111/j.1365-313X.2004.02179.x>
- [28] Lijavetzky, D., Muzzi, G., Wicker, T., Keller, B., Wing, R. and Dubcovsky, J. (1999) Construction and Characterization of a Bacterial Artificial Chromosome (BAC) Library for the A Genome of Wheat. *Genome*, **42**, 1176-1182. <https://doi.org/10.1139/g99-076>
- [29] Kubaláková, M., Vrána, J., Čihalíková, J., Šimková, H. and Doležel, J. (2002) Flow Karyotyping and Chromosome Sorting in Bread Wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics*, **104**, 1362-1372. <https://doi.org/10.1007/s00122-002-0888-2>
- [30] Gill, B.S., Friebe, B. and Endo, T.R. (1991) Standard Karyotype and Nomenclature System for Description of Chromosome Bands and Structural Aberrations in Wheat (*Triticum aestivum*). *Genome*, **34**, 830-839. <https://doi.org/10.1139/g91-128>
- [31] Itoh, T., Tanaka, T., Barrero, R.A., Yamasaki, C., Fujii, Y., Hilton, P.B., *et al.* (2007) Curated Genome Annotation of *Oryza sativa ssp. japonica* and Comparative Genome Analysis with *Arabidopsis thaliana*. *Genome Research*, **17**, 175-183. <https://doi.org/10.1101/gr.5509507>
- [32] Pfosser, M., Amon, A., Lelley, T. and Heberle-Bors, E. (1995) Evaluation of Sensitivity of Flow Cytometry in Detecting Aneuploidy in Wheat Using Disomic and Ditelosomic Wheat-Rye Addition Lines. *Cytometry*, **21**, 387-393. <https://doi.org/10.1002/cyto.990210412>
- [33] Jia, J., Zhao, S., Kong, X., Li, Y., Zhao, G., He, W., *et al.* (2013) *Aegilops tauschii* Draft Genome Sequence Reveals a Gene Repertoire for Wheat Adaptation. *Nature*, **496**, 91-95. <https://doi.org/10.1038/nature12028>
- [34] Ling, H.-Q., Zhao, S., Liu, D., Wang, J., Sun, H., Zhang, C., *et al.* (2013) Draft Genome of the Wheat A-Genome Progenitor *Triticum urartu*. *Nature*, **496**, 87-90. <https://doi.org/10.1038/nature11997>
- [35] Brechley, R., Spannagl, M., Pfeifer, M., Barker, G.L.A., D'Amore, R., Allen, A.M., *et al.* (2012) Analysis of the Bread Wheat Genome Using Whole-Genome Shotgun Sequencing. *Nature*, **491**, 705-710. <https://doi.org/10.1038/nature11650>
- [36] Sears, E.R. (1954) The Aneuploids of Common Wheat. Research Bulletin No. 572, Agricultural Experiment Station, University of Missouri, Columbia, 1-58.
- [37] Sears, E.R. and Sears, L.M.S. (1978) The Telocentric Chromosomes of Common Wheat. *Proceedings of the 5th International Wheat Genetics Symposium*, New Delhi, 23-28 February 1978, 389-407.
- [38] Vogel, J.P., Garvin, D.F., Mockler, T.C., Schmutz, J., Rokhsar, D. and Bevan, M.W. (2010) Genome Sequencing and Analysis of the Model Grass *Brachypodium distachyon*. *Nature*, **463**, 763-768. <https://doi.org/10.1038/nature08747>

- [39] Burr, B. (2005) The Map-Based Sequence of the Rice Genome. *Nature*, **436**, 793-800. <https://doi.org/10.1038/nature03895>
- [40] Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., *et al.* (2009) The *Sorghum bicolor* Genome and the Diversification of Grasses. *Nature*, **457**, 551-556. <https://doi.org/10.1038/nature07723>
- [41] Mayer, K.F., Waugh, R., Brown, J.W., Schulman, A., Langridge, P., Platzer, M., *et al.* (2012) A Physical, Genetic and Functional Sequence Assembly of the Barley Genome. *Nature*, **491**, 711-716.
- [42] Martis, M., Zhou, R., Haseneyer, G., Schmutzer, T., Vrána, J., Kubaláková, M., *et al.* (2013) Reticulate Evolution of the Rye Genome. *The Plant Cell*, **25**, 3685-3698. <https://doi.org/10.1105/tpc.113.114553>
- [43] Bennetzen, J., Schmutz, J., Wang, H., Percifield, R., Hawkins, J., Pontaroli, A.C., *et al.* (2012) Reference Genome Sequence of the Model Plant *Setaria*. *Nature Biotechnology*, **30**, 555-561. <https://doi.org/10.1038/nbt.2196>
- [44] Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Pasternak, S., Liang, C., *et al.* (2009) The B73 Maize Genome: Complexity, Diversity and Dynamics. *Science*, **326**, 1112-1115. <https://doi.org/10.1126/science.1178534>
- [45] Pfeifer, M., Kugler, K.G., Sandve, S.R., Zhan, B., Rudi, H., Hvidsten, T.R., *et al.* (2014) Genome Interplay in the Grain Transcriptome of Hexaploid Bread Wheat. *Science*, **345**, Article ID: 1250091. <https://doi.org/10.1126/science.1250091>
- [46] Adams, K., Cronn, R., Percifield, R. and Wendel, J.F. (2003) Genes Duplicated by Polyploidy Show Unequal Contributions to the Transcriptome and Organ-Specific Reciprocal Silencing. *Proceedings of the National Academy of Sciences*, **100**, 4649-1654. <https://doi.org/10.1073/pnas.0630618100>
- [47] Yoo, M.J., Szadkowski, E. and Wendel, J.F. (2013) Homoeolog Expression Bias and Expression Level Dominance in Allopolyploid Cotton. *Heredity*, **110**, 171-180. <https://doi.org/10.1038/hdy.2012.94>
- [48] Cheng, F., Wu, J., Fang, L., Sun, S., Liu, B., Lin, K., *et al.* (2012) Biased Gene Fractionation and Dominant Gene Expression among the Subgenomes of *Brassica rapa*. *PLoS ONE*, **7**, e36442. <https://doi.org/10.1371/journal.pone.0036442>
- [49] Chen, Z. (2007) Genetic and Epigenetic Mechanisms for Gene Expression and Phenotypic Variation in Plant Polyploids. *Annual Review of Plant Biology*, **58**, 377. <https://doi.org/10.1146/annurev.arplant.58.032806.103835>
- [50] Schnable, J., Springer, N. and Freeling, M. (2011) Differentiation of the Maize Subgenomes by Genome Dominance and Both Ancient and Ongoing Gene Loss. *Proceedings of the National Academy of Sciences*, **108**, 4069-4074. <https://doi.org/10.1073/pnas.1101368108>
- [51] Mikheyev, A. and Tin, M. (2014) A First Look at the Oxford Nanopore MinION Sequencer. *Molecular Ecology Resources*, **14**, 1097-1102. <https://doi.org/10.1111/1755-0998.12324>
- [52] Šimková, H., Staňková, H., Hastie, A., Chan, S., Vrána, J., Tulpova, Z., *et al.* (2016) BioNano Genome Map of Bread Wheat 7DS Arm Supports Sequence Assembly and Analysis. *Proceedings of 24th Plant and Animal Genome Conference*, San Diego, 9-13 January 2016.
- [53] Zhu, T., Rodriguez, J.C., Deal, K.R., Van, S., Dvorak, J. and Luo, M.-C. (2016) Evolution of Genome Structure in Polyploid Wheat Revealed by Comparison of Wheat and *Aegilops tauschii* Whole-Genome BioNano Maps. *Proceedings of 24th Plant and Animal Genome Conference*, San Diego, 9-13 January 2016, 1-2.
- [54] Zimin, A. (2016) Assembly of the 4.5 Gb Ancestral Wheat D-Genome from Hybrid

- PacBio and Illumina Data. *Proceedings of 24th Plant and Animal Genome Conference*, San Diego, 9-13 January 2016.
- [55] Chaisson, M., Huddleston, J., Dennis, M., Sudmant, P.H., Malig, M., Hormozdiari, F., *et al.* (2015) Resolving the Complexity of the Human Genome Using Single-Molecule Sequencing. *Nature*, **517**, 608-611. <https://doi.org/10.1038/nature13907>
- [56] Ling, H.-Q. (2016) Genome Sequence of *T. urartu*. *Proceedings of 24th Plant and Animal Genome Conference*, San Diego, 9-13 January 2016, 1-2.
- [57] Clavijo, B.J., Venturini, L., Schudoma, C., Garcia Accinelli, G., Kaithakottil, G., Wright, J., *et al.* (2016) An Improved Assembly and Annotation of the Allohexaploid Wheat Genome Identifies Complete Families of Agronomic Genes and Provides Genomic Evidence for Chromosomal Translocations.
- [58] Ruperao, P., Chan, C., Azam, S., Karafiatova, M., Hayashi, S., Cizkova, J., *et al.* (2014) A Chromosomal Genomics Approach to Assess and Validate the Desi and Kabuli Draft Chickpea Genome Assemblies. *Plant Biotechnology Journal*, **14**, 778-786. <https://doi.org/10.1111/pbi.12182>
- [59] Ganapathy, G., Howard, J., Ward, J., Li, J., Li, B., Li, Y., *et al.* (2014) High-Coverage Sequencing and Annotated Assemblies of the Budgerigar Genome. *GigaScience*, **3**, 11. <https://doi.org/10.1186/2047-217X-3-11>
- [60] Pendleton, M., Sebra, R., Pang, A., Ummat, A., Franzen, O., Rausch, T., *et al.* (2015) Assembly and Diploid Architecture of an Individual Human Genome via Single-Molecule Technologies. *Nature Methods*, **12**, 780-786. <https://doi.org/10.1038/nmeth.3454>
- [61] Callaway, E. (2014) "Platinum" Genome Shapes Up. *Nature*, **515**, 323.
- [62] Shearer, L., Anderson, L., de Jong, H., Smit, S., Goicoechea, J.L., Roe, B.A., *et al.* (2014) Fluorescence *in Situ* Hybridization and Optical Mapping to Correct Scaffold Arrangement in the Tomato Genome. *Genes Genomes Genetics*, **4**, 1395-1405. <https://doi.org/10.1534/g3.114.011197>
- [63] Paux, E., Sourdille, P., Salse, J., Saintenac, C., Choulet, F., Leroy, P., *et al.* (2008) A Physical Map of the 1-Gigabase Bread Wheat Chromosome 3B. *Science*, **322**, 101-104. <https://doi.org/10.1126/science.1161847>
- [64] Luo, M.-C., Gu, Y.Q., You, F.M., Deal, K.R., Ma, Y., Hu, Y., *et al.* (2013) A 4-Gigabase Physical Map Unlocks the Structure and Evolution of the Complex Genome of *Aegilops tauschii*, the Wheat D-Genome Progenitor. *Proceedings of the National Academy of Sciences USA*, **110**, 7940-7945. <https://doi.org/10.1073/pnas.1219082110>
- [65] Staňková, H., Hastie, A.R., Chan, S., Chan, S., Vrána, J., Tulpova, Z., *et al.* (2016) BioNano Genome Mapping of Individual Chromosomes Supports Physical Mapping and Sequence Assembly in Complex Plant Genomes. *Plant Biotechnology Journal*, **14**, 1523-1531. <https://doi.org/10.1111/pbi.12513>
- [66] Mascher, M., Muehlbauer, G., Rokhsar, D., Chapman, J., Schmutz, J., Barry, K., *et al.* (2013) Anchoring and Ordering NGS Contig Assemblies by Population Sequencing (POPSEQ). *The Plant Journal*, **76**, 718-727. <https://doi.org/10.1111/tpj.12319>
- [67] Glenn, T.C. (2011) Field Guide to Next-Generation DNA Sequencers. *Molecular Ecology Resources*, **11**, 759-769. <https://doi.org/10.1111/j.1755-0998.2011.03024.x>
- [68] Cao, H., Hastie, A.R., Cao, D., *et al.* (2014) Rapid Detection of Structural Variation in a Human Genome Using Nanochannel-Based Genome Mapping Technology. *GigaScience*, **3**, 1-11. <https://doi.org/10.1186/2047-217X-3-34>

- [69] Blakesley, R.W., Hansen, N.F., Gupta, J., McDowell, J.C., Maskeri, B., Barnabas, B B., *et al.* (2010) Effort Required to Finish Shotgun-Generated Genome Sequences Differs Significantly Among Vertebrates. *BMC Genomics*, **11**, 21. <https://doi.org/10.1186/1471-2164-11-21>
- [70] Hastie, A.R., Dong, L., Smith, A., Finkelstein, J., Lam, E.T., Huo, N., *et al.* (2013) Rapid Genome Mapping in Nanochannel Arrays for Highly Complete and Accurate *De Novo* Sequence Assembly of the Complex *Aegilops tauschii* Genome. *PLoS ONE*, **8**, e55864. <https://doi.org/10.1371/journal.pone.0055864>
- [71] Molnár, I., Kubaláková, M., Šimková, H., Farkas, A., Cseh, A., Megyeri, M., *et al.* (2014) Flow Cytometric Chromosome Sorting from Diploid Progenitors of Bread Wheat, *T. urartu*, *Ae. speltoides* and *Ae. tauschii*. *Theoretical and Applied Genetics*, **127**, 1091-1104. <https://doi.org/10.1007/s00122-014-2282-2>
- [72] Mayer, K.F.X., Martis, M., Hedley, P.E. Šimková, H., Liu, H., Morris, J.A., *et al.* (2011) Unlocking the Barley Genome by Chromosomal and Comparative Genomics. *Plant Cell*, **23**, 1249-1263. <https://doi.org/10.1105/tpc.110.082537>
- [73] Hernandez, P., Martis, M., Dorado, G., Pfeifer, M., Gálvez, S., Schaaf, S., *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 Journal*, **69**, 377-386. <https://doi.org/10.1111/j.1365-313X.2011.04808.x>
- [74] Vitulo, N., Albiero, A., Forcato, C., Campagna, D., Dal Pero, F., Bagnaresi, P., *et al.* (2011) First Survey of the Wheat Chromosome 5A Composition through a Next Generation Sequencing Approach. *PLoS ONE*, **6**, Article ID: 226421. <https://doi.org/10.1371/journal.pone.0026421>



Scientific Research Publishing

Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact as@scirp.org